Patents on Coronavirus

A.B. Solunke **R.S. Awasthi**

V.S. Hamde **P.S. Wakte**

US007776521B1

(12) **United States Patent**
Rota et al.

(10) **Patent No.: US 7,776,521 B1**
(45) **Date of Patent: Aug. 17, 2010**

(54) **CORONAVIRUS ISOLATED FROM HUMANS**

(75) Inventors: **Paul A. Rota**, Decatur, GA (US); **Larry J. Anderson**, Atlanta, GA (US); **William J. Bellini**, Lilburn, GA (US); **Michael D. Bowen**, Decatur, GA (US); **Cara Carthel Burns**, Avondale Estates, GA (US); **Raymond Campagnoli**, Decatur, GA (US); **Qi Chen**, Marietta, GA (US); **James A. Comer**, Decatur, GA (US); **Byron T. Cook**, Augusta, GA (US); **Shannon L. Emery**, Lusaka (ZM); **Dean D. Erdman**, Decatur, GA (US); **Cynthia S. Goldsmith**, Lilburn, GA (US); **Jeanette Guarner**, Decatur, GA (US); **Charles D. Humphrey**, Lilburn, GA (US); **Joseph P. Icenogle**, Atlanta, GA (US); **Thomas G. Ksiazek**, Lilburn, GA (US); **Richard F. Meyer**, Roswell, GA (US); **Stephan S. Monroe**, Decatur, GA (US); **William Allan Nix**, Bethlehem, GA (US); **M. Steven Oberste**, Lilburn, GA (US); **Christopher D. Paddock**, Atlanta, GA (US); **Teresa C. T. Peret**, Atlanta, GA (US); **Pierre E. Rollin**, Lilburn, GA (US); **Mark A. Pallansch**, Lilburn, GA (US); **Anthony Sanchez**, Lilburn, GA (US); **Wun-Ju Shieh**, Norcross, GA (US); **Suxiang Tong**, Alpharetta, GA (US); **Sherif R. Zaki**, Atlanta, GA (US)

(73) Assignee: **The United States of America as represented by the Secretary of the Department of Health and Human Services, Centers for Disease Control and Prevention**, Washington, DC (US)

(*) Notice: Subject to any disclaimer, the term of this patent is extended or adjusted under 35 U.S.C. 154(b) by 96 days.

(21) Appl. No.: **11/748,359**

(22) Filed: **May 14, 2007**

Related U.S. Application Data

(62) Division of application No. 10/822,904, filed on Apr. 12, 2004, now Pat. No. 7,220,852.

(60) Provisional application No. 60/465,927, filed on Apr. 25, 2003.

(51) **Int. Cl.**
C12Q 1/70 (2006.01)
C07H 21/00 (2006.01)
C12N 15/50 (2006.01)

(52) **U.S. Cl.** **435/5**; 536/24.32; 536/24.33

(58) **Field of Classification Search** ..... None
See application file for complete search history.

(56) **References Cited**

U.S. PATENT DOCUMENTS

2005/0181357 A1* 8/2005 Peiris et al. 435/5

FOREIGN PATENT DOCUMENTS

WO WO 2004/092360 10/2004

OTHER PUBLICATIONS

Ksiazek et al (New England Journal of Medicine 348:1953-1966, published online Apr. 10, 2003).*
Gut et al (Journal of Virological Methods 77:37-46, 1999).*
Neilan et al (Nucleic Acids Research 25:2938-9, 1997).*
Peiris et al (Lancet, 361:1319-1325, published online Apr. 8, 2003).*
Genbank Accession No. AY274119.1 GI:29826276 (Apr. 14, 2003).*
SARS-associated Coronavirus. Genomic Sequence Availability. [online] [retreived on Aug. 8, 2005]. Retreived from the Internet <URL: http://www.bcgsc.ca/bioinfo/SARS>.*
"Update: Outbreak of Severe Acute Respiratory Syndrome—Worldwide, 2003," *MMWR Weekly* 52:241-248, 2003.
Emery et al., "Real-Time Reverse Transcription-Polymerase Chain Reaction Assay for SARS-Associated Coronavirus," *Emerg. Infect. Diseases* 10:311-316, 2004.
Goldsmith et al., "Ultrastructural Characterization of SARS Coronavirus," *Emerg. Infect. Diseases* 10:320-326, 2004.
Ksiazek et al., "A Novel Coronavirus Associated with Severe Acute Respiratory Syndrome," *N. Engl. J. Med.* 348:1953-1966, 2003.
Luo and Luo, "Initial SARS Coronavirus Genome Sequence Analysis Using a Bioinformatics Platform," *2nd Asia-Pacific Bioinformatics Conference* (*APBC2004*), Dunedin, New Zealand, 2004.
Marra et al., "The Genome Sequence of the SARS-Associated Coronavirus," *Science* 300:1393-1404, 2003.
Supplementary Online Material for Marra et al., www.sciencemag.org/cgi/content/full/1085953/DC1, (2003).
Rota et al., "Characterization of a Novel Coronavirus Associated with Severe Acute Respiratory Syndrome," *Science* 300:1394-1399, 2003.
Supplementary Online Material for Rota et al., www.sciencemag.org/cgi/content/full/1085953/DC1, (2003).
GenBank Accession No. AY274119, Apr. 14, 2003.
GenBank Accession No. AY278741, Apr. 21, 2003.
GenBank Accession No. AY278554.1, Apr. 18, 2003.
GenBank Accession No. AY278491.1, Apr. 18, 2003.
GenBank Accession No. AY278487, Apr. 21, 2003.

* cited by examiner

Primary Examiner—Mary E Mosher
(74) *Attorney, Agent, or Firm*—Klarquist Sparkman, LLP

(57) **ABSTRACT**

Disclosed herein is a newly isolated human coronavirus (SARS-CoV), the causative agent of severe acute respiratory syndrome (SARS). Also provided are the nucleic acid sequence of the SARS-CoV genome and the amino acid sequences of the SARS-CoV open reading frames, as well as methods of using these molecules to detect a SARS-CoV and detect infections therewith. Immune stimulatory compositions are also provided, along with methods of their use.

7 Claims, 7 Drawing Sheets

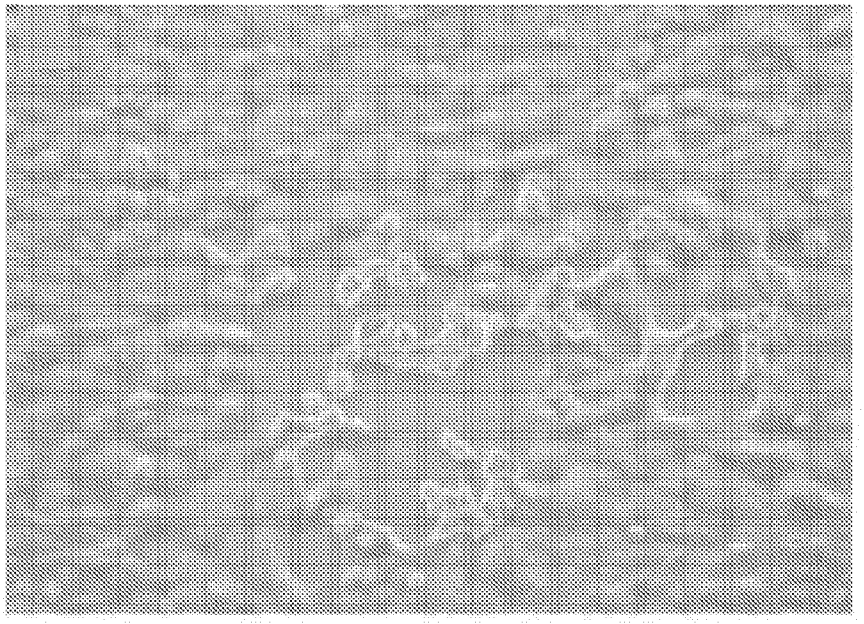


FIG. 1A

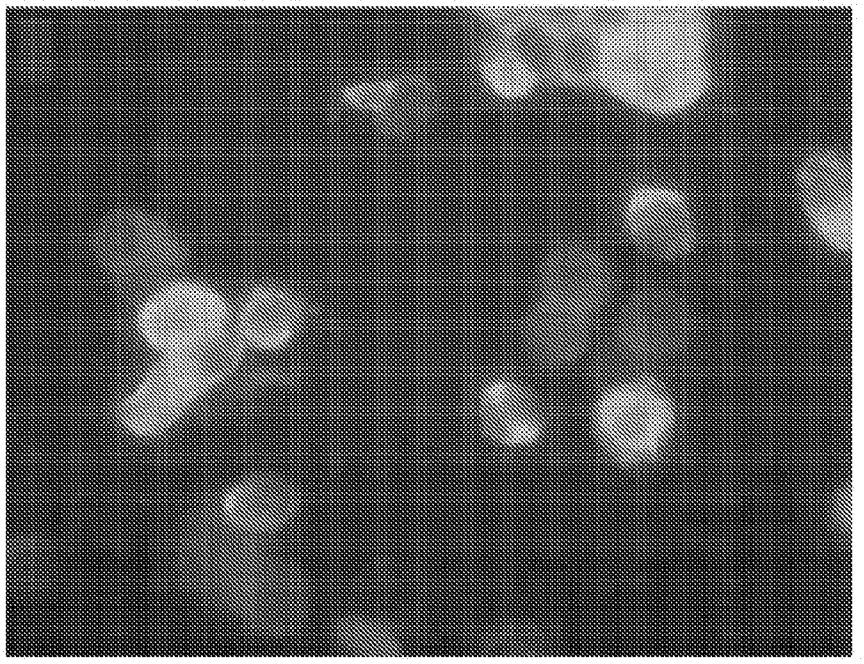


FIG. 1B

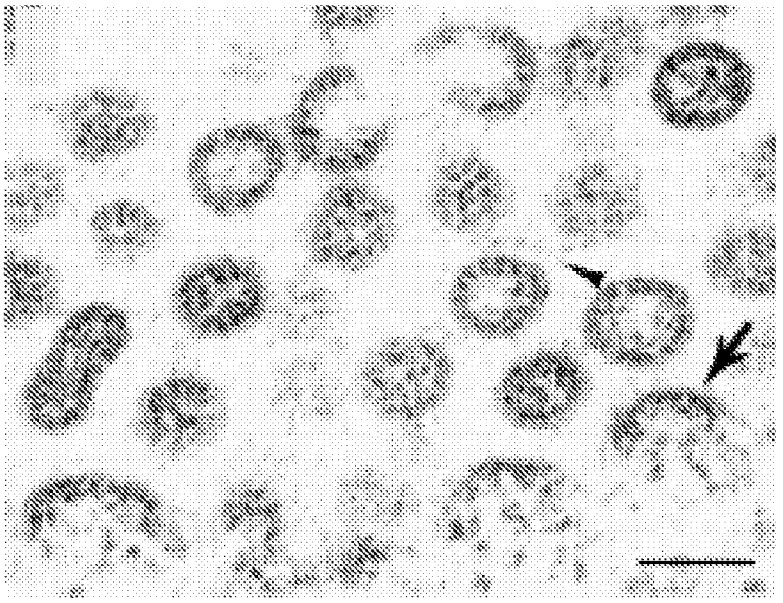


FIG. 2A

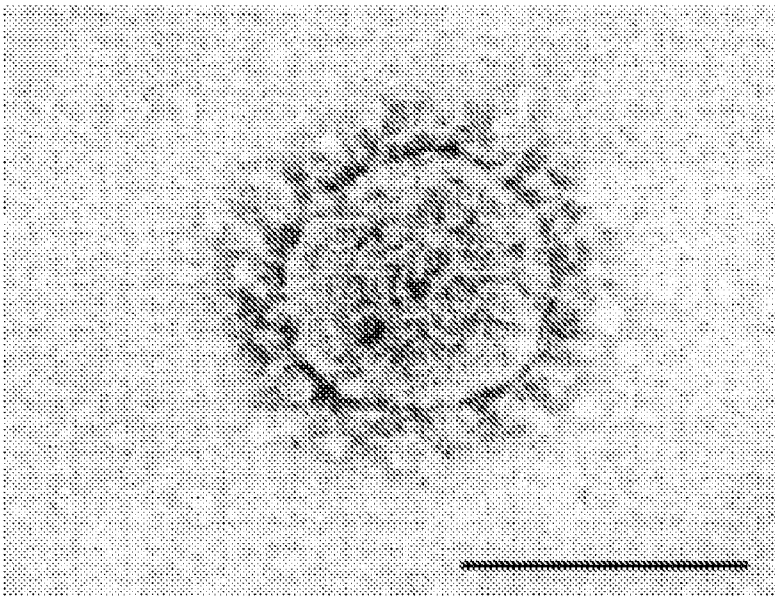


FIG. 2B

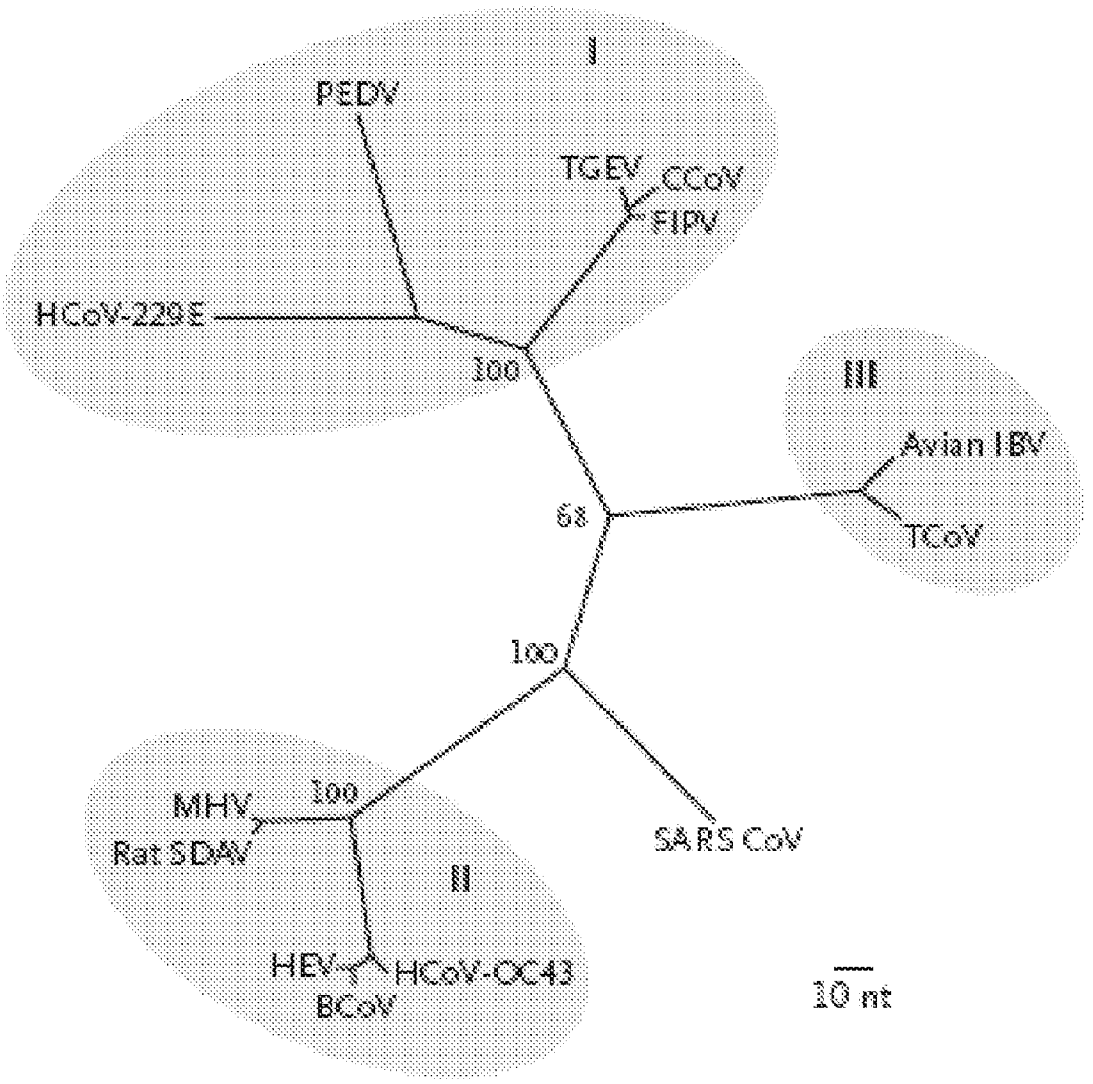

I
PEDV
TGEV
CCoV
FIPV
HCoV-229E
100
III
Avian IBV
68
TCoV
100
100
MHV
Rat SDAV
II
SARS CoV
HEV
BCoV
HCoV-OC43
10 nt

FIG. 3

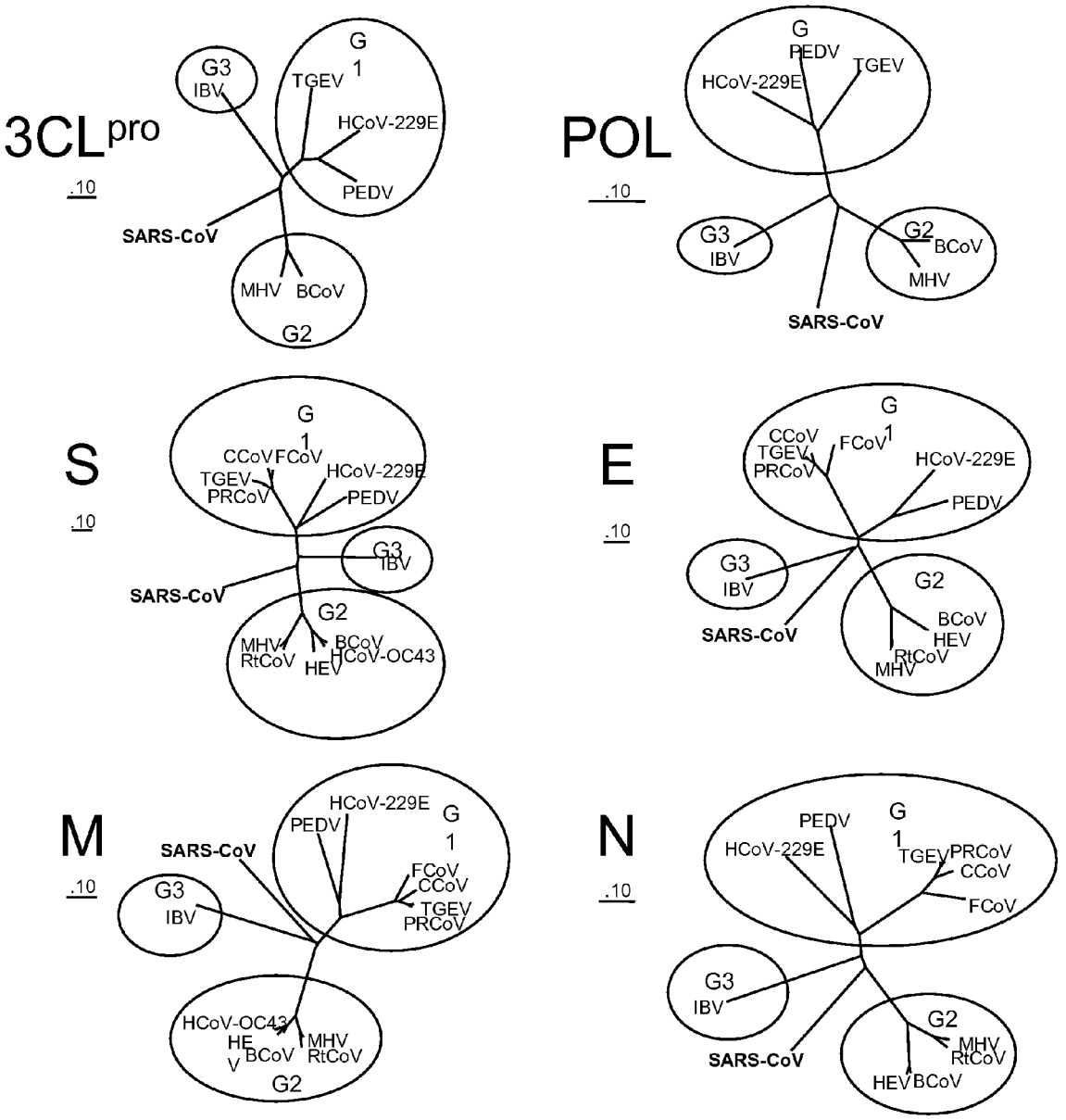

3CLpro
.10
G3
IBV
G
1
TGEV
HCoV-229E
PEDV
SARS-CoV
MHV
BCoV
G2
POL
.10
G
PEDV
TGEV
HCoV-229E
G3
IBV
G2
BCoV
MHV
SARS-CoV
S
.10
G
1
CCoV
FCoV
TGEV
PRCoV
HCoV-229E
PEDV
G3
IBV
SARS-CoV
G2
MHV
RtCoV
BCoV
HEV
HCoV-OC43
E
.10
G
1
CCoV
TGEV
PRCoV
FCoV
HCoV-229E
PEDV
G3
IBV
SARS-CoV
G2
BCoV
HEV
RtCoV
MHV
M
.10
HCoV-229E
PEDV
G
1
SARS-CoV
FCoV
CCoV
TGEV
PRCoV
G3
IBV
HCoV-OC43
HE
V
MHV
BCoV
RtCoV
G2
N
.10
PEDV
G
1
HCoV-229E
TGEV
PRCoV
CCoV
FCoV
G3
IBV
SARS-CoV
G2
MHV
RtCoV
HEV
BCoV

FIG. 4

FIG. 5A

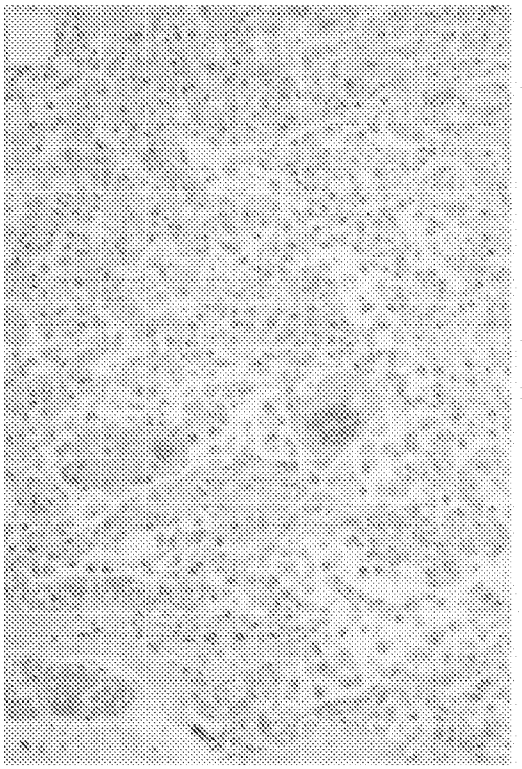
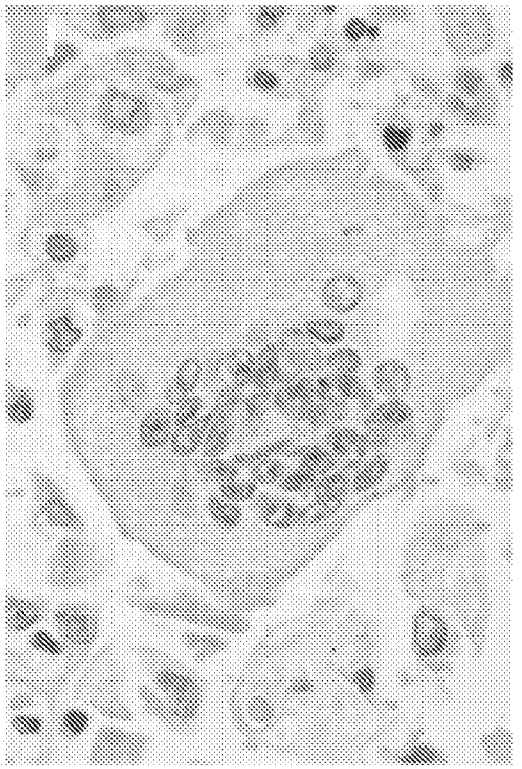


FIG. 5B



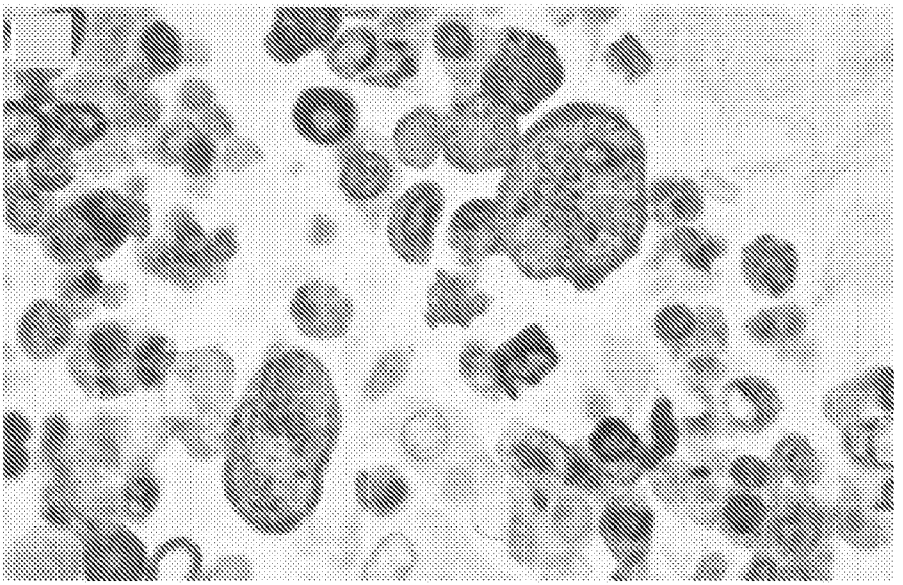


FIG. 5C

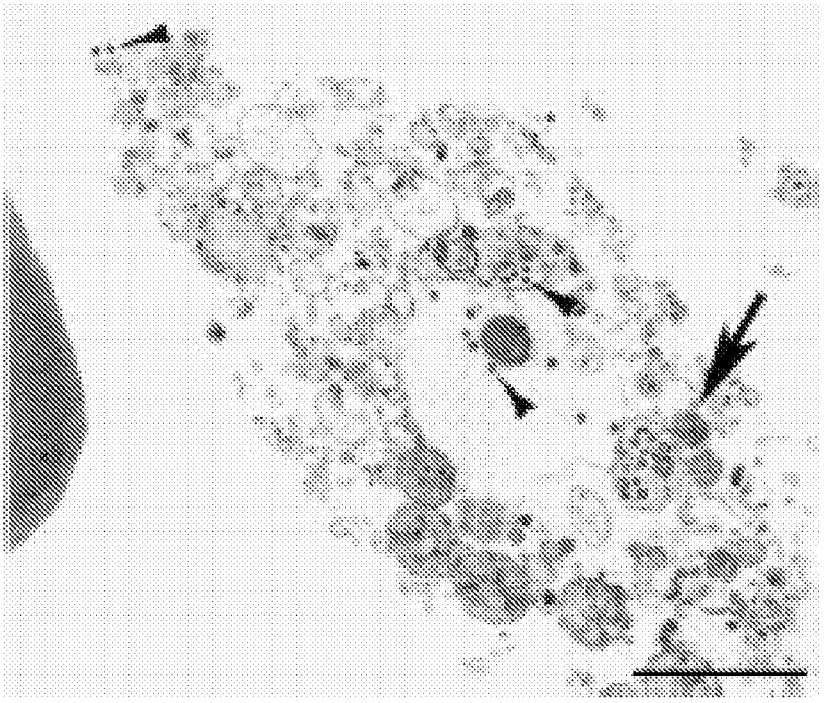


FIG. 6A

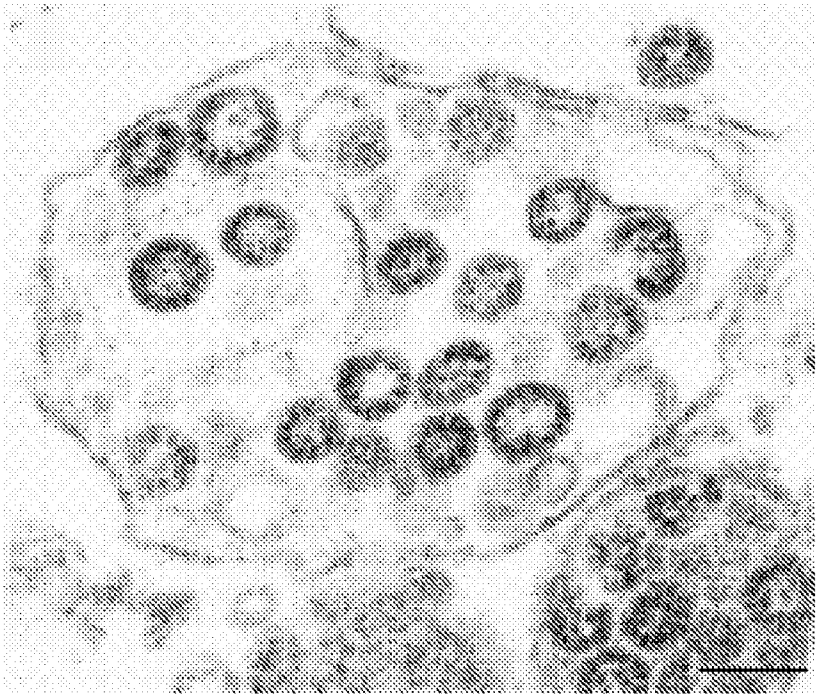


FIG. 6B

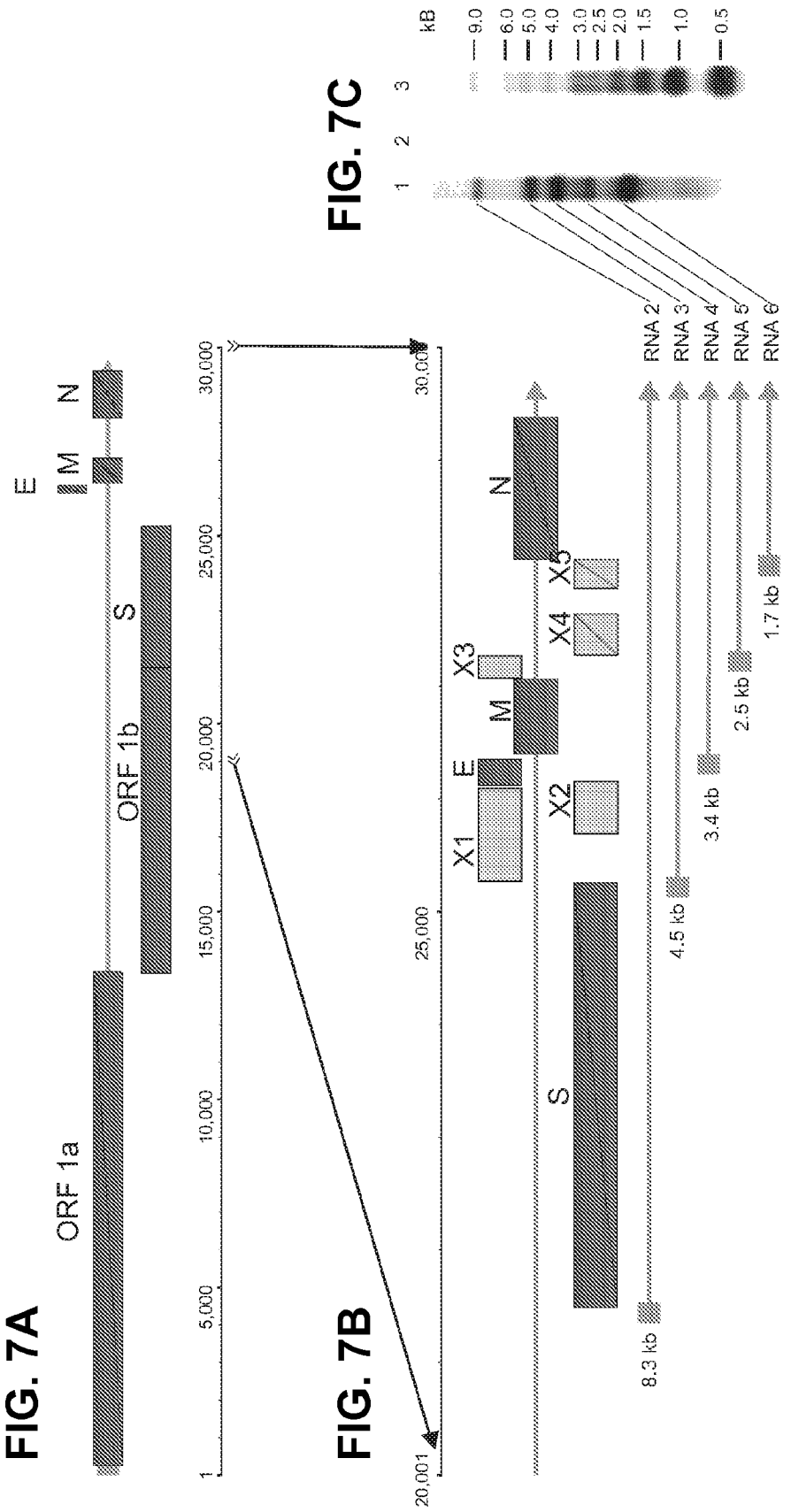


FIG. 7A
ORF 1a
ORF 1b
S
M
N
E
1
5,000
10,000
15,000
20,000
25,000
30,000
FIG. 7B
20,001
25,000
30,000
X1
X2
X3
X4
X5
E
M
N
S
8.3 kb
4.5 kb
3.4 kb
2.5 kb
1.7 kb
FIG. 7C
kB
9.0
6.0
5.0
4.0
3.0
2.5
2.0
1.5
1.0
0.5
1
2
3
RNA 2
RNA 3
RNA 4
RNA 5
RNA 6

CORONAVIRUS ISOLATED FROM HUMANS

PRIORITY CLAIM

This is a division of co-pending U.S. patent application Ser. No. 10/822,904, filed Apr. 12, 2004, and issued as U.S. Pat. No. 7,220,852 on May 22, 2007, which in turn claims the benefit of U.S. Provisional Patent Application No. 60/465,927 filed Apr. 25, 2003. Both applications are incorporated herein by reference in their entirety.

STATEMENT OF GOVERNMENT SUPPORT

This invention was made by the Centers for Disease Control and Prevention, an agency of the United States Government. Therefore, the U.S. Government has certain rights in this invention.

FIELD OF THE DISCLOSURE

This invention relates to a newly isolated human coronavirus. More particularly, it relates to an isolated coronavirus genome, isolated coronavirus proteins, and isolated nucleic acid molecules encoding the same. The disclosure further relates to methods of detecting a severe acute respiratory syndrome-associated coronavirus and compositions comprising immunogenic coronavirus compounds.

BACKGROUND

The coronaviruses (order Nidovirales, family Coronaviridae, genus *Coronavirus*) are a diverse group of large, enveloped, positive-stranded RNA viruses that cause respiratory and enteric diseases in humans and other animals. At approximately 30,000 nucleotides (nt), their genome is the largest found in any of the RNA viruses. Coronaviruses are spherical, 100-160 nm in diameter with 20-40 nm complex club shaped surface projections surrounding the periphery. Coronaviruses share common structural proteins including a spike protein (S), membrane protein (M), envelope protein (E), and, in a subset of coronaviruses, a hemagglutinin-esterase protein (HE). The S protein, a glycoprotein which protrudes from the virus membrane, is involved in host cell receptor binding and is a target for neutralizing antibodies. The E and M proteins are involved in virion formation and release from the host cell. Coronavirus particles are found within the cisternae of the rough endoplasmic reticulum and in vesicles of infected host cells where virions are assembled. The coronavirus genome consists of two open reading frames (ORF1a and ORF1b) yielding an RNA polymerase and a nested set of subgenomic mRNAs encoding structural and nonstructural proteins, including the S, E, M, and nucleocapsid (N) proteins. The genus *Coronavirus* includes at least 13 species which have been subdivided into at least three groups (groups I, II, and III) on the basis of serological and genetic properties (deVries et al., *Sem. Virol.* 8:33-47, 1997; Fields et al. eds. *Fields Virology,* 3rd edition, Raven Press, Philadelphia, 1323-1341, 1996; Mahey and Collier eds. *Microbiology and Microbial Infections*, Volume 1 Virology, 9th edition, Oxford University Press, 463-479, 1998).

The three known groups of coronavirus are associated with a variety of diseases of humans and domestic animals (for example, cattle, pigs, cats, dogs, rodents, and birds), including gastroenteritis and upper and lower respiratory tract disease. Known coronaviruses include human Coronavirus 229E (HCoV-229E), canine coronavirus (CCoV), feline infectious peritonitis virus (FIPV), porcine transmissible gastroenteritis virus (TGEV), porcine epidemic diarrhea virus (PEDV), human coronavirus OC43 (HCoV-OC43), bovine coronavirus (BCoV), porcine hemagglutinating encephalomyelitis virus (HEV), rat sialodacryoadenitis virus (SDAV), mouse hepatitis virus (MHV), turkey coronavirus (TCoV), and avian infectious bronchitis virus (IBV-Avian) (Fields et al. eds. *Fields Virology,* 3rd edition, Raven Press, Philadelphia, 1323-1341, 1996; Mahey and Collier eds. *Microbiology and Microbial Infections*, Volume 1 Virology, 9th edition, Oxford University Press, 463-479, 1998).

Coronavirus infections are generally host specific with respect to infectivity and clinical symptoms. Coronaviruses further exhibit marked tissue tropism; infection in the incorrect host species or tissue type may result in an abortive infection, mutant virus production and altered virulence. Coronaviruses generally do not grow well in cell culture, and animal models for human *coronavirus* infection are lacking. Therefore, little is known about them (Fields et al. eds. Fields Virology, 3rd edition, Raven Press, Philadelphia, 1323-1341, 1996). The known human coronaviruses are notably fastidious in cell culture, preferring select cell lines, organ culture, or suckling mice for propagation. Coronaviruses grown in cell culture exhibit varying degrees of virulence and/or cytopathic effect (CPE) depending on the host cell type and culture conditions. The only human or animal coronavirus which has been shown to grow in Vero E6 cells is PEDV, and it requires the addition of trypsin to culture medium for growth in Vero E6 cells. Moreover, PEDV adapted to Vero E6 cell culture results in a strikingly different CPE, with cytoplasmic vacuoles and the formation of large syncytia (Hofmann and Wyler, *J. Clin. Micro.* 26:2235-39, 1988; Kusanagi et el., *J. Vet. Med. Sci.* 554:313-18, 1991).

Coronavirus have not previously been known to cause severe disease in humans, but have been identified as a major cause of upper respiratory tract illness, including the common cold. Repeat infections in humans are common within and across serotype, suggesting that immune response to coronavirus infection in humans is either incomplete or short lived. Coronavirus infection in animals can cause severe enteric or respiratory disease. Vaccination has been used successfully to prevent and control some coronavirus infections in animals. The ability of animal-specific coronaviruses to cause severe disease raises the possibility that coronavirus could also cause more severe disease in humans (Fields et al. eds. *Fields Virology,* 3rd edition, Raven Press, Philadelphia, 1323-1341, 1996; Mahey and Collier eds. *Microbiology and Microbial Infections*, Volume 1 Virology, 9th edition, Oxford University Press, 463-479, 1998).

In late 2002, cases of life-threatening respiratory disease with no identifiable etiology were reported from Guangdong Province, China, followed by reports from Vietnam, Canada, and Hong Kong of severe febrile respiratory illness that spread to household members and health care workers. The syndrome was designated "severe acute respiratory syndrome" (SARS) in February 2003 by the Centers for Disease Control and Prevention (*MMWR,* 52:241-48, 2003).

Past efforts to develop rapid diagnostics and vaccines for coronavirus infection in humans have been hampered by a lack of appropriate research models and the moderate course of disease in humans. Therefore, a need for rapid diagnostic tests and vaccines exists.

SUMMARY OF THE DISCLOSURE

A newly isolated human coronavirus has been identified as the causative agent of SARS, and is termed SARS-CoV. The

nucleic acid sequence of the SARS-CoV genome and the amino acid sequences of the SARS-CoV open reading frames are provided herein.

This disclosure provides methods and compositions useful in detecting the presence of a SARS-CoV nucleic acid in a sample and/or diagnosing a SARS-CoV infection in a subject. Also provided are methods and compositions useful in detecting the presence of a SARS-CoV antigen or antibody in a sample and/or diagnosing a SARS-CoV infection in a subject.

The foregoing and other features and advantages will become more apparent from the following detailed description of several embodiments, which proceeds with reference to the accompanying figures.

BRIEF DESCRIPTION OF THE FIGURES

FIGS. **1**A-B are photomicrographs illustrating typical early cytopathic effects seen with coronavirus isolates and serum from SARS patients. FIG. **1**A is a photomicrograph of Vero E6 cells inoculated with an oropharyngeal specimen from a SARS patient (x40). FIG. **1**B is a photomicrograph of infected Vero E6 cells reacting with the serum of a convalescent SARS patient in an indirect fluorescent antibody (IFA) assay (x400).

FIGS. **2**A-B are electronmicrographs illustrating ultrastructural characteristics of the SARS-associated coronavirus (SARS-CoV). FIG. **2**A is a thin-section electron-microscopical view of viral nucleocapsids aligned along the membrane of the rough endoplasmic reticulum (arrow) as particles bud into the cisternae. Enveloped virions have surface projections (arrowhead) and an electron-lucent center. Directly under the viral envelope lies a characteristic ring formed by the helical nucleocapsid, often seen in cross-section. FIG. **2**B is a negative stain (methylamine tungstate) electronmicrograph showing stain-penetrated coronavirus particle with the typical internal helical nucleocapsid-like structure and club-shaped surface projections surrounding the periphery of the particle. Bars: 100 nm.

FIG. **3** is an estimated maximum parsimony tree illustrating putative phylogenetic relationships between SARS-CoV and other human and animal coronaviruses. Phylogenetic relationships are based on sequence alignment of 405 nucleotides of the coronavirus polymerase gene ORF1b (nucleic acid 15,173 to 15,578 of SEQ ID NO: 1). The three major coronavirus antigenic groups (I, II and III), represented by HCoV-229E, CCoV, FIPV, TGEV, PEDV, HCoV-OC43, BCoV, HEV, SDAV, MHV, TCoV, and IBV-Avian, are shown shaded. Bootstrap values (100 replicates) obtained from a 50% majority rule consensus tree are plotted at the main internal branches of the phylogram. Branch lengths are proportionate to nucleotide differences.

FIG. **4** is a pictorial representation of neighbor joining trees illustrating putative phylogenetic relationships between SARS-CoV and other human and animal coronaviruses. Amino acid sequences of the indicated SARS-CoV proteins were compared with those from reference viruses representing each species in the three groups of coronaviruses for which complete genomic sequence information was available [group 1: HCoV-229E (AF304460); PEDV (AF353511); TGEV (AJ271965); group 2: BCoV (AF220295); MHV (AF201929); group 3: infectious bronchitis virus (M95169)]. Sequences for representative strains of other coronavirus species, for which partial sequence information was available, were included for some of the structural protein comparisons [group 1: CCoV (D13096); FCoV (AY204704); porcine respiratory coronavirus (Z24675); group 2: HCoV-OC43 (M76373, L14643, M93390); HEV (AY078417); rat coronavirus (AF207551)]. Sequence alignments and neighbor joining trees were generated by using Clustalx 1.83 with the Gonnet protein comparison matrix. The resulting trees were adjusted for final output using treetool 2.0.1.

FIGS. **5**A-C are photomicrographs illustrating diffuse alveolar damage in a patient with SARS (FIGS. **5**A-B), and immunohistochemical staining of SARS-CoV-infected Vero E6 cells (FIG. **5**C). FIG. **5**A is a photomicrograph of lung tissue from a SARS patient (x50). Diffuse alveolar damage, abundant foamy macrophages and multinucleated syncytial cells are present; hematoxylin and eosin stain was used. FIG. **5**B is a higher magnification photomicrograph of lung tissue from the same SARS patient (x250). Syncytial cells show no conspicuous viral inclusions. FIG. **5**C is a photomicrograph of immunohistochemically stained SARS-CoV-infected cells (x250). Membranous and cytoplasmic immunostaining of individual and syncytial Vero E6 cells was achieved using feline anti-FIPV-1 ascitic fluid. Immunoalkaline phosphatase with naphthol-fast red substrate and hematoxylin counter stain was used.

FIGS. **6**A-B are electronmicrographs illustrating ultrastructural characteristics of a coronavirus-infected cell in bronchoalveolar lavage (BAL) from a SARS patient. FIG. **6**A is an electron micrograph of a coronavirus-infected cell. Numerous intracellular and extracellular particles are present; virions are indicated by the arrowheads. FIG. **6**B is a higher magnification electronmicrograph of the area seen at the arrow in FIG. **6**A (rotated clockwise approximately) 90°. Bars: FIG. **6**A, 1 μm; FIG. **6**B, 100 nm.

FIGS. **7**A-C illustrate the organization of the SARS-CoV genome. FIG. **7**A is a diagram of the overall organization of the 29,727-nt SARS-CoV genomic RNA. The 72-nt leader sequence is represented as a small rectangle at the left-most end. ORFs1a and 1b, encoding the nonstructural polyproteins, and those ORFs encoding the S, E, M, and N structural proteins, are indicated. Vertical position of the boxes indicates the phase of the reading frame (phase 1 for proteins above the line, phase two for proteins on the line and phase three for proteins below the line). FIG. **7**B is an expanded view of the structural protein encoding region and predicted mRNA transcripts. Known structural protein encoding regions (dark grey boxes) and regions and reading frames for potential products X1-X5 (light gray boxes) are indicated. Lengths and map locations of the 3'-coterminal mRNAs expressed by the SARS-CoV are indicated, as predicted by identification of conserved transcriptional regulatory sequences. FIG. **7**C is a digitized image of a nylon membrane showing Northern blot analysis of SARS-CoV mRNAs. Poly (A)+ RNA from infected Vero E6 cells was separated on a formaldehyde-agarose gel, transferred to a nylon membrane, and hybridized with a digoxigenin-labeled riboprobe overlapping the 3'-untranslated region. Signals were visualized by chemiluminescence. Sizes of the SARS-CoV mRNAs were calculated by extrapolation from a log-linear fit of the molecular mass marker. Lane 1, SARS-CoV mRNA; lane 2, Vero E6 cell mRNA; lane 3, molecular mass marker, sizes in kB.

SEQUENCE LISTING

The nucleic and amino acid sequences listed in the accompanying sequence listing are shown using standard letter abbreviations for nucleotide bases, and three letter code for amino acids, as defined in 37 C.F.R. 1.822. Only one strand of each nucleic acid sequence is shown, but the complementary strand is understood as included by any reference to the displayed strand. In the accompanying sequence listing:

SEQ ID NO: 1 shows the nucleic acid sequence of the SARS-CoV genome.

SEQ ID NO: 2 shows the amino acid sequence of the SARS-CoV polyprotein 1a (encoded by nucleic acid 265 to nucleic acid 13,398 of SEQ ID NO: 1).

SEQ ID NO: 3 shows the amino acid sequence of the SARS-CoV polyprotein 1b (encoded by nucleic acid 13,398 to 21,482 of SEQ ID NO: 1).

SEQ ID NO: 4 shows the amino acid sequence of the SARS-CoV S protein (encoded by nucleic acid 21,492 to 25,256 of SEQ ID NO: 1).

SEQ ID NO: 5 shows the amino acid sequence of the SARS-CoV X1 protein (encoded by nucleic acid 25,268 to 26,089 of SEQ ID NO: 1).

SEQ ID NO: 6 shows the amino acid sequence of the SARS-CoV X2 protein (encoded by nucleic acid 25,689 to 26,150 of SEQ ID NO: 1).

SEQ ID NO: 7 shows the amino acid sequence of the SARS-CoV E protein (encoded by nucleic acid 26,117 to 26,344 of SEQ ID NO: 1).

SEQ ID NO: 8 shows the amino acid sequence of the SARS-CoV M protein (encoded by nucleic acid 26,398 to 27,060 of SEQ ID NO: 1).

SEQ ID NO: 9 shows the amino acid sequence of the SARS-CoV X3 protein (encoded by nucleic acid 27,074 to 27,262 of SEQ ID NO: 1).

SEQ ID NO: 10 shows the amino acid sequence of the SARS-CoV X4 protein (encoded by nucleic acid 27,273 to 27,638 of SEQ ID NO: 1).

SEQ ID NO: 11 shows the amino acid sequence of the SARS-CoV X5 protein (encoded by nucleic acid 27,864 to 28,115 of SEQ ID NO: 1).

SEQ ID NO: 12 shows the amino acid sequence of the SARS-CoV N protein (encoded by nucleic acid 28,120 to 29,385 of SEQ ID NO: 1).

SEQ ID NOs: 13-15 show the nucleic acid sequence of several SARS-CoV-specific oligonucleotide primers.

SEQ ID NOs: 16-33 show the nucleic acid sequence of several oligonucleotide primers/probes used for real-time reverse transcription-polymerase chain reaction (RT-PCR) SARS-CoV assays.

SEQ ID NOs: 34-35 show the nucleic acid sequence of two degenerate primers designed to anneal to sites encoding conserved coronavirus amino acid motifs.

SEQ ID NOs: 36-38 show the nucleic acid sequence of several oligonucleotide primers/probes used as controls in real-time RT-PCR assays.

DETAILED DESCRIPTION OF SEVERAL EMBODIMENTS

I. Abbreviations

BAL: bronchoalveolar lavage
CPE: cytopathic effect
E: coronavirus transmembrane protein
ELISA: enzyme-linked immunosorbent assay
HE: coronavirus hemagglutinin-esterase protein
IFA: indirect fluorescent antibody
M: coronavirus membrane protein
N: coronavirus nucleoprotein
ORF: open reading frame
PCR polymerase chain reaction
RACE: 5' rapid amplification of cDNA ends
RT-PCR: reverse transcription-polymerase chain reaction
S: coronavirus spike protein
SARS: severe acute respiratory syndrome
SARS-CoV: severe acute respiratory syndrome-associated coronavirus
TRS: transcriptional regulatory sequence

II. Terms

Unless otherwise noted, technical terms are used according to conventional usage. Definitions of common terms in molecular biology may be found in Benjamin Lewin, Genes VII, published by Oxford University Press, 2000 (ISBN 019879276X); Kendrew et al. (eds.), *The Encyclopedia of Molecular Biology*, published by Blackwell Publishers, 1994 (ISBN 0632021829); and Robert A. Meyers (ed.), *Molecular Biology and Biotechnology: a Comprehensive Desk Reference*, published by Wiley, John & Sons, Inc., 1995 (ISBN 0471186341); and other similar references.

As used herein, the singular terms "a," "an," and "the" include plural referents unless context clearly indicates otherwise. Similarly, the word "or" is intended to include "and" unless the context clearly indicates otherwise. Also, as used herein, the term "comprises" means "includes." Hence "comprising A or B" means including A, B, or A and B. It is further to be understood that all nucleotide sizes or amino acid sizes, and all molecular weight or molecular mass values, given for nucleic acids or polypeptides are approximate, and are provided for description. Although methods and materials similar or equivalent to those described herein can be used in the practice or testing of the present invention, suitable methods and materials are described below. All publications, patent applications, patents, and other references mentioned herein are incorporated by reference in their entirety. In case of conflict, the present specification, including explanations of terms, will control. In addition, the materials, methods, and examples are illustrative only and not intended to be limiting.

In order to facilitate review of the various embodiments of this disclosure, the following explanations of specific terms are provided:

Adjuvant: A substance that non-specifically enhances the immune response to an antigen. Development of vaccine adjuvants for use in humans is reviewed in Singh et al. (*Nat. Biotechnol.* 17:1075-1081, 1999), which discloses that, at the time of its publication, aluminum salts and the MF59 microemulsion are the only vaccine adjuvants approved for human use.

Amplification: Amplification of a nucleic acid molecule (e.g., a DNA or RNA molecule) refers to use of a laboratory technique that increases the number of copies of a nucleic acid molecule in a sample. An example of amplification is the polymerase chain reaction (PCR), in which a sample is contacted with a pair of oligonucleotide primers under conditions that allow for the hybridization of the primers to a nucleic acid template in the sample. The primers are extended under suitable conditions, dissociated from the template, re-annealed, extended, and dissociated to amplify the number of copies of the nucleic acid. The product of amplification can be characterized by such techniques as electrophoresis, restriction endonuclease cleavage patterns, oligonucleotide hybridization or ligation, and/or nucleic acid sequencing.

Other examples of amplification methods include strand displacement amplification, as disclosed in U.S. Pat. No. 5,744,311; transcription-free isothermal amplification, as disclosed in U.S. Pat. No. 6,033,881; repair chain reaction amplification, as disclosed in WO 90/01069; ligase chain reaction amplification, as disclosed in EP-A-320,308; gap filling ligase chain reaction amplification, as disclosed in U.S. Pat. No. 5,427,930; and NASBA™ RNA transcription-free amplification, as disclosed in U.S. Pat. No. 6,025,134. An

amplification method can be modified, including for example by additional steps or coupling the amplification with another protocol.

Animal: Living multi-cellular vertebrate organisms, a category that includes, for example, mammals and birds. The term mammal includes both human and non-human mammals. Similarly, the term "subject" includes both human and veterinary subjects, for example, humans, non-human primates, dogs, cats, horses, and cows.

Antibody: A protein (or protein complex) that includes one or more polypeptides substantially encoded by immunoglobulin genes or fragments of immunoglobulin genes. The recognized immunoglobulin genes include the kappa, lambda, alpha, gamma, delta, epsilon, and mu constant region genes, as well as the myriad immunoglobulin variable region genes. Light chains are classified as either kappa or lambda. Heavy chains are classified as gamma, mu, alpha, delta, or epsilon, which in turn define the immunoglobulin classes, IgG, IgM, IgA, IgD and IgE, respectively.

The basic immunoglobulin (antibody) structural unit is generally a tetramer. Each tetramer is composed of two identical pairs of polypeptide chains, each pair having one "light" (about 25 kDa) and one "heavy" (about 50-70 kDa) chain. The N-terminus of each chain defines a variable region of about 100 to 110 or more amino acids primarily responsible for antigen recognition. The terms "variable light chain" (V_L) and "variable heavy chain" (V_H) refer, respectively, to these light and heavy chains

As used herein, the term "antibodies" includes intact immunoglobulins as well as a number of well-characterized fragments. For instance, Fabs, Fvs, and single-chain Fvs (SCFvs) that bind to target protein (or epitope within a protein or fusion protein) would also be specific binding agents for that protein (or epitope). These antibody fragments are defined as follows: (1) Fab, the fragment which contains a monovalent antigen-binding fragment of an antibody molecule produced by digestion of whole antibody with the enzyme papain to yield an intact light chain and a portion of one heavy chain; (2) Fab', the fragment of an antibody molecule obtained by treating whole antibody with pepsin, followed by reduction, to yield an intact light chain and a portion of the heavy chain; two Fab' fragments are obtained per antibody molecule; (3) $(Fab')_2$, the fragment of the antibody obtained by treating whole antibody with the enzyme pepsin without subsequent reduction; (4) $F(ab')_2$, a dimer of two Fab' fragments held together by two disulfide bonds; (5) Fv, a genetically engineered fragment containing the variable region of the light chain and the variable region of the heavy chain expressed as two chains; and (6) single chain antibody, a genetically engineered molecule containing the variable region of the light chain, the variable region of the heavy chain, linked by a suitable polypeptide linker as a genetically fused single chain molecule. Methods of making these fragments are routine (see, for example, Harlow and Lane, *Using Antibodies: A Laboratory Manual*, CSHL, New York, 1999).

Antibodies for use in the methods and devices of this disclosure can be monoclonal or polyclonal. Merely by way of example, monoclonal antibodies can be prepared from murine hybridomas according to the classical method of Kohler and Milstein (*Nature* 256:495-97, 1975) or derivative methods thereof. Detailed procedures for monoclonal antibody production are described in Harlow and Lane, *Using Antibodies: A Laboratory Manual*, CSHL, New York, 1999.

Antigen: A compound, composition, or substance that can stimulate the production of antibodies or a T-cell response in an animal, including compositions that are injected or absorbed into an animal. An antigen reacts with the products of specific humoral or cellular immunity, including those induced by heterologous immunogens. In one embodiment, an antigen is a coronavirus antigen.

Binding or Stable Binding: An oligonucleotide binds or stably binds to a target nucleic acid if a sufficient amount of the oligonucleotide forms base pairs or is hybridized to its target nucleic acid, to permit detection of that binding. Binding can be detected by either physical or functional properties of the target:oligonucleotide complex. Binding between a target and an oligonucleotide can be detected by any procedure known to one skilled in the art, including functional or physical binding assays. Binding can be detected functionally by determining whether binding has an observable effect upon a biosynthetic process such as expression of a gene, DNA replication, transcription, translation, and the like.

Physical methods of detecting the binding of complementary strands of DNA or RNA are well known in the art, and include such methods as DNase I or chemical footprinting, gel shift and affinity cleavage assays, Northern blotting, Southern blotting, dot blotting, and light absorption detection procedures. For example, a method which is widely used, because it is so simple and reliable, involves observing a change in light absorption of a solution containing an oligonucleotide (or an analog) and a target nucleic acid at 220 to 300 nm as the temperature is slowly increased. If the oligonucleotide or analog has bound to its target, there is a sudden increase in absorption at a characteristic temperature as the oligonucleotide (or analog) and target dissociate or melt.

The binding between an oligomer and its target nucleic acid is frequently characterized by the temperature (T_m) at which 50% of the oligomer is melted from its target. A higher T_m means a stronger or more stable complex relative to a complex with a lower T_m.

cDNA (complementary DNA): A piece of DNA lacking internal, non-coding segments (introns) and regulatory sequences that determine transcription. cDNA is synthesized in the laboratory by reverse transcription from messenger RNA extracted from cells.

Electrophoresis: Electrophoresis refers to the migration of charged solutes or particles in a liquid medium under the influence of an electric field. Electrophoretic separations are widely used for analysis of macromolecules. Of particular importance is the identification of proteins and nucleic acid sequences. Such separations can be based on differences in size and/or charge. Nucleotide sequences have a uniform charge and are therefore separated based on differences in size. Electrophoresis can be performed in an unsupported liquid medium (for example, capillary electrophoresis), but more commonly the liquid medium travels through a solid supporting medium. The most widely used supporting media are gels, for example, polyacrylamide and agarose gels.

Sieving gels (for example, agarose) impede the flow of molecules. The pore size of the gel determines the size of a molecule that can flow freely through the gel. The amount of time to travel through the gel increases as the size of the molecule increases. As a result, small molecules travel through the gel more quickly than large molecules and thus progress further from the sample application area than larger molecules, in a given time period. Such gels are used for size-based separations of nucleotide sequences.

Fragments of linear DNA migrate through agarose gels with a mobility that is inversely proportional to the $\log_{10}$ of their molecular weight. By using gels with different concentrations of agarose, different sizes of DNA fragments can be resolved. Higher concentrations of agarose facilitate separation of small DNAs, while low agarose concentrations allow resolution of larger DNAs.

Hybridization: Oligonucleotides and their analogs hybridize by hydrogen bonding, which includes Watson-Crick, Hoogsteen or reversed Hoogsteen hydrogen bonding, between complementary bases. Generally, nucleic acid consists of nitrogenous bases that are either pyrimidines (cytosine (C), uracil (U), and thymine (T)) or purines (adenine (A) and guanine (G)). These nitrogenous bases form hydrogen bonds between a pyrimidine and a purine, and the bonding of the pyrimidine to the purine is referred to as "base pairing." More specifically, A will hydrogen bond to T or U, and G will bond to C. "Complementary" refers to the base pairing that occurs between two distinct nucleic acid sequences or two distinct regions of the same nucleic acid sequence.

"Specifically hybridizable" and "specifically complementary" are terms that indicate a sufficient degree of complementarity such that stable and specific binding occurs between the oligonucleotide (or its analog) and the DNA or RNA target. The oligonucleotide or oligonucleotide analog need not be 100% complementary to its target sequence to be specifically hybridizable. An oligonucleotide or analog is specifically hybridizable when binding of the oligonucleotide or analog to the target DNA or RNA molecule interferes with the normal function of the target DNA or RNA, and there is a sufficient degree of complementarity to avoid non-specific binding of the oligonucleotide or analog to non-target sequences under conditions where specific binding is desired, for example under physiological conditions in the case of in vivo assays or systems. Such binding is referred to as specific hybridization.

Hybridization conditions resulting in particular degrees of stringency will vary depending upon the nature of the hybridization method of choice and the composition and length of the hybridizing nucleic acid sequences. Generally, the temperature of hybridization and the ionic strength (especially the Na^+ and/or Mg^{++} concentration) of the hybridization buffer will determine the stringency of hybridization, though wash times also influence stringency. Calculations regarding hybridization conditions required for attaining particular degrees of stringency are discussed by Sambrook et al. (ed.), *Molecular Cloning: A Laboratory Manual,* 2nd ed., vol. 1-3, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., 1989, chapters 9 and 11; and Ausubel et al. *Short Protocols in Molecular Biology,* 4th ed., John Wiley & Sons, Inc., 1999.

For purposes of the present disclosure, "stringent conditions" encompass conditions under which hybridization will only occur if there is less than 25% mismatch between the hybridization molecule and the target sequence. "Stringent conditions" may be broken down into particular levels of stringency for more precise definition. Thus, as used herein, "moderate stringency" conditions are those under which molecules with more than 25% sequence mismatch will not hybridize; conditions of "medium stringency" are those under which molecules with more than 15% mismatch will not hybridize, and conditions of "high stringency" are those under which sequences with more than 10% mismatch will not hybridize. Conditions of "very high stringency" are those under which sequences with more than 6% mismatch will not hybridize.

Immune Stimulatory Composition: A term used herein to mean a composition useful for stimulating or eliciting a specific immune response (or immunogenic response) in a vertebrate. In some embodiments, the immunogenic response is protective or provides protective immunity, in that it enables the vertebrate animal to better resist infection with or disease progression from the organism against which the vaccine is directed.

Without wishing to be bound by a specific theory, it is believed that an immunogenic response may arise from the generation of an antibody specific to one or more of the epitopes provided in the immune stimulatory composition. Alternatively, the response may comprise a T-helper or cytotoxic cell-based response to one or more of the epitopes provided in the immune stimulatory composition. All three of these responses may originate from naïve or memory cells. One specific example of a type of immune stimulatory composition is a vaccine.

In some embodiments, an "effective amount" or "immune-stimulatory amount" of an immune stimulatory composition is an amount which, when administered to a subject, is sufficient to engender a detectable immune response. Such a response may comprise, for instance, generation of an antibody specific to one or more of the epitopes provided in the immune stimulatory composition. Alternatively, the response may comprise a T-helper or CTL-based response to one or more of the epitopes provided in the immune stimulatory composition. All three of these responses may originate from naïve or memory cells. In other embodiments, a "protective effective amount" of an immune stimulatory composition is an amount which, when administered to a subject, is sufficient to confer protective immunity upon the subject.

Inhibiting or Treating a Disease: Inhibiting the full development of a disease or condition, for example, in a subject who is at risk for a disease such as SARS. "Treatment" refers to a therapeutic intervention that ameliorates a sign or symptom of a disease or pathological condition after it has begun to develop. As used herein, the term "ameliorating," with reference to a disease, pathological condition or symptom, refers to any observable beneficial effect of the treatment. The beneficial effect can be evidenced, for example, by a delayed onset of clinical symptoms of the disease in a susceptible subject, a reduction in severity of some or all clinical symptoms of the disease, a slower progression of the disease, a reduction in the number of relapses of the disease, an improvement in the overall health or well-being of the subject, or by other parameters well known in the art that are specific to the particular disease.

Isolated: An "isolated" microorganism (such as a virus, bacterium, fungus, or protozoan) has been substantially separated or purified away from microorganisms of different types, strains, or species. Microorganisms can be isolated by a variety of techniques, including serial dilution and culturing.

An "isolated" biological component (such as a nucleic acid molecule, protein or organelle) has been substantially separated or purified away from other biological components in the cell of the organism in which the component naturally occurs, such as other chromosomal and extra-chromosomal DNA and RNA, proteins, and organelles. Nucleic acids and proteins that have been "isolated" include nucleic acids and proteins purified by standard purification methods. The term also embraces nucleic acids and proteins prepared by recombinant expression in a host cell, as well as chemically synthesized nucleic acids or proteins, or fragments thereof.

Label: A detectable compound or composition that is conjugated directly or indirectly to another molecule to facilitate detection of that molecule. Specific, non-limiting examples of labels include fluorescent tags, enzymatic linkages, and radioactive isotopes.

Nucleic Acid Molecule: A polymeric form of nucleotides, which may include both sense and anti-sense strands of RNA,

cDNA, genomic DNA, and synthetic forms and mixed polymers of the above. A nucleotide refers to a ribonucleotide, deoxynucleotide or a modified form of either type of nucleotide. A "nucleic acid molecule" as used herein is synonymous with "nucleic acid" and "polynucleotide." A nucleic acid molecule is usually at least 10 bases in length, unless otherwise specified. The term includes single- and double-stranded forms of DNA. A polynucleotide may include either or both naturally occurring and modified nucleotides linked together by naturally occurring and/or non-naturally occurring nucleotide linkages.

Oligonucleotide: A nucleic acid molecule generally comprising a length of 300 bases or fewer. The term often refers to single-stranded deoxyribonucleotides, but it can refer as well to single- or double-stranded ribonucleotides, RNA:DNA hybrids and double-stranded DNAs, among others. The term "oligonucleotide" also includes oligonucleosides (that is, an oligonucleotide minus the phosphate) and any other organic base polymer. In some examples, oligonucleotides are about 10 to about 90 bases in length, for example, 12, 13, 14, 15, 16, 17, 18, 19 or 20 bases in length. Other oligonucleotides are about 25, about 30, about 35, about 40, about 45, about 50, about 55, about 60 bases, about 65 bases, about 70 bases, about 75 bases or about 80 bases in length. Oligonucleotides may be single-stranded, for example, for use as probes or primers, or may be double-stranded, for example, for use in the construction of a mutant gene. Oligonucleotides can be either sense or anti-sense oligonucleotides. An oligonucleotide can be modified as discussed above in reference to nucleic acid molecules. Oligonucleotides can be obtained from existing nucleic acid sources (for example, genomic or cDNA), but can also be synthetic (for example, produced by laboratory or in vitro oligonucleotide synthesis).

Open Reading Frame (ORF): A series of nucleotide triplets (codons) coding for amino acids without any internal termination codons. These sequences are usually translatable into a peptide/polypeptide/protein/polyprotein.

Operably Linked: A first nucleic acid sequence is operably linked with a second nucleic acid sequence when the first nucleic acid sequence is placed in a functional relationship with the second nucleic acid sequence. For instance, a promoter is operably linked to a coding sequence if the promoter affects the transcription or expression of the coding sequence. Generally, operably linked DNA sequences are contiguous and, where necessary to join two protein coding regions, in the same reading frame. If introns are present, the operably linked DNA sequences may not be contiguous.

Pharmaceutically Acceptable Carriers: The pharmaceutically acceptable carriers useful in this disclosure are conventional. *Remington's Pharmaceutical Sciences*, by E. W. Martin, Mack Publishing Co., Easton, Pa., 15th Edition (1975), describes compositions and formulations suitable for pharmaceutical delivery of one or more therapeutic compounds or molecules, such as one or more SARS-CoV nucleic acid molecules, proteins or antibodies that bind these proteins, and additional pharmaceutical agents.

In general, the nature of the carrier will depend on the particular mode of administration being employed. For instance, parenteral formulations usually comprise injectable fluids that include pharmaceutically and physiologically acceptable fluids such as water, physiological saline, balanced salt solutions, aqueous dextrose, glycerol or the like as a vehicle. For solid compositions (for example, powder, pill, tablet, or capsule forms), conventional non-toxic solid carriers can include, for example, pharmaceutical grades of mannitol, lactose, starch, or magnesium stearate. In addition to biologically-neutral carriers, pharmaceutical compositions to be administered can contain minor amounts of non-toxic auxiliary substances, such as wetting or emulsifying agents, preservatives, and pH buffering agents and the like, for example sodium acetate or sorbitan monolaurate.

Polypeptide: A polymer in which the monomers are amino acid residues which are joined together through amide bonds. When the amino acids are alpha-amino acids, either the L-optical isomer or the D-optical isomer can be used, the L-isomers being preferred. The terms "polypeptide" or "protein" as used herein are intended to encompass any amino acid sequence and include modified sequences such as glycoproteins. The term "polypeptide" is specifically intended to cover naturally occurring proteins, as well as those which are recombinantly or synthetically produced.

Conservative amino acid substitutions are those substitutions that, when made, least interfere with the properties of the original protein, that is, the structure and especially the function of the protein is conserved and not significantly changed by such substitutions. Examples of conservative substitutions are shown below.

Original Residue	Conservative Substitutions
Ala	Ser
Arg	Lys
Asn	Gln, His
Asp	Glu
Cys	Ser
Gln	Asn
Glu	Asp
His	Asn; Gln
Ile	Leu, Val
Leu	Ile; Val
Lys	Arg; Gln; Glu
Met	Leu; Ile
Phe	Met; Leu; Tyr
Ser	Thr
Thr	Ser
Trp	Tyr
Tyr	Trp; Phe
Val	Ile; Leu

Conservative substitutions generally maintain (a) the structure of the polypeptide backbone in the area of the substitution, for example, as a sheet or helical conformation, (b) the charge or hydrophobicity of the molecule at the target site, or (c) the bulk of the side chain.

The substitutions which in general are expected to produce the greatest changes in protein properties will be non-conservative, for instance changes in which (a) a hydrophilic residue, for example, seryl or threonyl, is substituted for (or by) a hydrophobic residue, for example, leucyl, isoleucyl, phenylalanyl, valyl or alanyl; (b) a cysteine or proline is substituted for (or by) any other residue; (c) a residue having an electropositive side chain, for example, lysyl, arginyl, or histadyl, is substituted for (or by) an electronegative residue, for example, glutamyl or aspartyl; or (d) a residue having a bulky side chain, for example, phenylalanine, is substituted for (or by) one not having a side chain, for example, glycine.

Probes and Primers: A probe comprises an isolated nucleic acid attached to a detectable label or other reporter molecule. Typical labels include radioactive isotopes, enzyme substrates, co-factors, ligands, chemiluminescent or fluorescent agents, haptens, and enzymes. Methods for labeling and guidance in the choice of labels appropriate for various purposes are discussed, for example, in Sambrook et al. (ed.), *Molecular Cloning: A Laboratory Manual, 2nd* ed., vol. 1-3, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.,

1989 and Ausubel et al. *Short Protocols in Molecular Biology,* 4th ed., John Wiley & Sons, Inc., 1999.

Primers are short nucleic acid molecules, for instance DNA oligonucleotides 10 nucleotides or more in length, for example that hybridize to contiguous complementary nucleotides or a sequence to be amplified. Longer DNA oligonucleotides may be about 15, 20, 25, 30 or 50 nucleotides or more in length. Primers can be annealed to a complementary target DNA strand by nucleic acid hybridization to form a hybrid between the primer and the target DNA strand, and then the primer extended along the target DNA strand by a DNA polymerase enzyme. Primer pairs can be used for amplification of a nucleic acid sequence, for example, by the PCR or other nucleic-acid amplification methods known in the art, as described above.

Methods for preparing and using nucleic acid probes and primers are described, for example, in Sambrook et al. (ed.), *Molecular Cloning: A Laboratory Manual,* 2nd ed., vol. 1-3, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., 1989; Ausubel et al. *Short Protocols in Molecular Biology,* 4th ed., John Wiley & Sons, Inc., 1999; and Innis et al. *PCR Protocols, A Guide to Methods and Applications*, Academic Press, Inc., San Diego, Calif., 1990. Amplification primer pairs can be derived from a known sequence, for example, by using computer programs intended for that purpose such as Primer (Version 0.5, © 1991, Whitehead Institute for Biomedical Research, Cambridge, Mass.). One of ordinary skill in the art will appreciate that the specificity of a particular probe or primer increases with its length. Thus, in order to obtain greater specificity, probes and primers can be selected that comprise at least 20, 25, 30, 35, 40, 45, 50 or more consecutive nucleotides of a target nucleotide sequence.

Protein: A biological molecule, particularly a polypeptide, expressed by a gene and comprised of amino acids. A "polyprotein" is a protein that, after synthesis, is cleaved to produce several functionally distinct polypeptides.

Purified: The term "purified" does not require absolute purity; rather, it is intended as a relative term. Thus, for example, a purified protein preparation is one in which the subject protein is more pure than in its natural environment within a cell. Generally, a protein preparation is purified such that the protein represents at least 50% of the total protein content of the preparation.

Recombinant Nucleic Acid: A sequence that is not naturally occurring or has a sequence that is made by an artificial combination of two otherwise separated segments of sequence. This artificial combination is often accomplished by chemical synthesis or, more commonly, by the artificial manipulation of isolated segments of nucleic acids, for example, by genetic engineering techniques such as those described in Sambrook et al. (ed.), *Molecular Cloning: A Laboratory Manual,* 2nd ed., vol. 1-3, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., 1989. The term recombinant includes nucleic acids that have been altered solely by addition, substitution, or deletion of a portion of the nucleic acid.

Sample: A portion, piece, or segment that is representative of a whole. This term encompasses any material, including for instance samples obtained from an animal, a plant, or the environment.

An "environmental sample" includes a sample obtained from inanimate objects or reservoirs within an indoor or outdoor environment. Environmental samples include, but are not limited to: soil, water, dust, and air samples; bulk samples, including building materials, furniture, and landfill contents; and other reservoir samples, such as animal refuse, harvested grains, and foodstuffs.

A "biological sample" is a sample obtained from a plant or animal subject. As used herein, biological samples include all samples useful for detection of viral infection in subjects, including, but not limited to: cells, tissues, and bodily fluids, such as blood; derivatives and fractions of blood (such as serum); extracted galls; biopsied or surgically removed tissue, including tissues that are, for example, unfixed, frozen, fixed in formalin and/or embedded in paraffin; tears; milk; skin scrapes; surface washings; urine; sputum; cerebrospinal fluid; prostate fluid; pus; bone marrow aspirates; BAL; saliva; cervical swabs; vaginal swabs; and oropharyngeal wash.

Sequence Identity The similarity between two nucleic acid sequences, or two amino acid sequences, is expressed in terms of the similarity between the sequences, otherwise referred to as sequence identity. Sequence identity is frequently measured in terms of percentage identity (or similarity or homology); the higher the percentage, the more similar the two sequences are.

Methods of alignment of sequences for comparison are well known in the art. Various programs and alignment algorithms are described in: Smith and Waterman (*Adv. Appl. Math.,* 2:482, 1981); Needleman and Wunsch (*J. Mol. Biol.,* 48:443, 1970); Pearson and Lipman (*Proc. Natl. Acad. Sci.,* 85:2444, 1988); Higgins and Sharp (*Gene,* 73:237-44, 1988); Higgins and Sharp (*CABIOS,* 5:151-53, 1989); Corpet et al. (*Nuc. Acids Res.,* 16:10881-90, 1988); Huang et al. (*Comp. Appls Biosci.,* 8:155-65, 1992); and Pearson et al. (*Meth. Mol. Biol.,* 24:307-31, 1994). Altschul et al. (*Nature Genet.,* 6:119-29, 1994) presents a detailed consideration of sequence alignment methods and homology calculations.

The alignment tools ALIGN (Myers and Miller, *CABIOS* 4:11-17, 1989) or LFASTA (Pearson and Lipman, 1988) may be used to perform sequence comparisons (Internet Program© 1996, W. R. Pearson and the University of Virginia, "fasta20u63" version 2.0u63, release date December 1996). ALIGN compares entire sequences against one another, while LFASTA compares regions of local similarity. These alignment tools and their respective tutorials are available on the Internet at the NCSA website. Alternatively, for comparisons of amino acid sequences of greater than about 30 amino acids, the "Blast 2 sequences" function can be employed using the default BLOSUM62 matrix set to default parameters, (gap existence cost of 11, and a per residue gap cost of 1). When aligning short peptides (fewer than around 30 amino acids), the alignment should be performed using the "Blast 2 sequences" function, employing the PAM30 matrix set to default parameters (open gap 9, extension gap 1 penalties). The BLAST sequence comparison system is available, for instance, from the NCBI web site; see also Altschul et al., *J. Mol. Biol.,* 215:403-10, 1990; Gish and States, *Nature Genet.,* 3:266-72, 1993; Madden et al., *Meth. Enzymol.,* 266:131-41, 1996; Altschul et al., *Nucleic Acids Res.,* 25:3389-402, 1997; and Zhang and Madden, *Genome Res.,* 7:649-56, 1997.

Orthologs (equivalent to proteins of other species) of proteins are in some instances characterized by possession of greater than 75% sequence identity counted over the full-length alignment with the amino acid sequence of specific protein using ALIGN set to default parameters. Proteins with even greater similarity to a reference sequence will show increasing percentage identities when assessed by this method, such as at least 80%, at least 85%, at least 90%, at least 92%, at least 95%, or at least 98% sequence identity. In addition, sequence identity can be compared over the full length of one or both binding domains of the disclosed fusion proteins.

When significantly less than the entire sequence is being compared for sequence identity, homologous sequences will

typically possess at least 80% sequence identity over short windows of 10-20, and may possess sequence identities of at least 85%, at least 90%, at least 95%, or at least 99% depending on their similarity to the reference sequence. Sequence identity over such short windows can be determined using LFASTA; methods are described at the NCSA website. One of skill in the art will appreciate that these sequence identity ranges are provided for guidance only; it is entirely possible that strongly significant homologs could be obtained that fall outside of the ranges provided. Similar homology concepts apply for nucleic acids as are described for protein.

An alternative indication that two nucleic acid molecules are closely related is that the two molecules hybridize to each other under stringent conditions. Representative hybridization conditions are discussed above.

Nucleic acid sequences that do not show a high degree of identity may nevertheless encode similar amino acid sequences, due to the degeneracy of the genetic code. It is understood that changes in nucleic acid sequence can be made using this degeneracy to produce multiple nucleic acid sequences that each encode substantially the same protein.

Specific Binding Agent: An agent that binds substantially only to a defined target. Thus a protein-specific binding agent binds substantially only the defined protein, or to a specific region within the protein. As used herein, a protein-specific binding agent includes antibodies and other agents that bind substantially to a specified polypeptide. The antibodies may be monoclonal or polyclonal antibodies that are specific for the polypeptide, as well as immunologically effective portions ("fragments") thereof.

The determination that a particular agent binds substantially only to a specific polypeptide may readily be made by using or adapting routine procedures. One suitable in vitro assay makes use of the Western blotting procedure (described in many standard texts, including Harlow and Lane, *Using Antibodies: A Laboratory Manual*, CSHL, New York, 1999).

Transformed: A "transformed" cell is a cell into which has been introduced a nucleic acid molecule by molecular biology techniques. The term encompasses all techniques by which a nucleic acid molecule might be introduced into such a cell, including transfection with viral vectors, transformation with plasmid vectors, and introduction of naked DNA by electroporation, lipofection, and particle gun acceleration.

Vector: A nucleic acid molecule as introduced into a host cell, thereby producing a transformed host cell. A vector may include nucleic acid sequences that permit it to replicate in a host cell, such as an origin of replication. A vector may also include one or more selectable marker genes and other genetic elements known in the art.

Virus: Microscopic infectious organism that reproduces inside living cells. A virus typically consists essentially of a core of a single nucleic acid surrounded by a protein coat, and has the ability to replicate only inside a living cell. "Viral replication" is the production of additional virus by the occurrence of at least one viral life cycle. A virus may subvert the host cells' normal functions, causing the cell to behave in a manner determined by the virus. For example, a viral infection may result in a cell producing a cytokine, or responding to a cytokine, when the uninfected cell does not normally do so.

"Coronaviruses" are large, enveloped, RNA viruses that cause respiratory and enteric diseases in humans and other animals. Coronavirus genomes are non-segmented, single-stranded, positive-sense RNA, approximately 27-31 kb in length. Genomes have a 5' methylated cap and 3' poly-A tail, and function directly as mRNA. Host cell entry occurs via endocytosis and membrane fusion, and replication occurs in the cytoplasm. Initially, the 5' 20 kb of the positive-sense genome is translated to produce a viral polymerase, which then produces a full-length negative-sense strand used as a template to produce subgenomic mRNA as a "nested set" of transcripts. Assembly occurs by budding into the golgi apparatus, and particles are transported to the surface of the cell and released.

III. Overview of Several Embodiments

A newly isolated human coronavirus (SARS-CoV) is disclosed herein. The entire genomic nucleic acid sequence of this virus is also provided herein. Also disclosed are the nucleic acid sequences of the SARS-CoV ORFs, and the polypeptide sequences encoded by these ORFs. Pharmaceutical and immune stimulatory compositions are also disclosed that include one or more SARS-CoV viral nucleic acids, polypeptides encoded by these viral nucleic acids and antibodies that bind to a SARS-CoV polypeptide or SARS-CoV polypeptide fragment.

In one embodiment, a method is provided for detecting the presence of SARS-CoV in a sample. This method includes contacting the sample with a pair of nucleic acid primers that hybridize to a SARS-CoV nucleic acid, wherein at least one primer is 5'-end labeled with a reporter dye, amplifying the SARS-CoV nucleic acid or a fragment thereof from the sample utilizing the pair of nucleic acid primers, electrophoresing the amplified products, and detecting the 5'-end labeled reporter dye, thereby detecting a SARS-CoV. In a specific, non-limiting example, the amplification utilizes RT-PCR. In a further specific example of the provided method, at least one of the nucleic acid primers that hybridize to a SARS-CoV nucleic acid includes a sequence as set forth in any one of SEQ ID NOs: 13-15.

In another example of the provided method, detecting the presence of SARS-CoV in a sample includes contacting the sample with a pair of nucleic acid primers that hybridize to a SARS-CoV nucleic acid, amplifying the SARS-CoV nucleic acid or a fragment thereof from the sample utilizing the pair of nucleic acid primers, adding to the amplified SARS-CoV nucleic acid or the fragment thereof a TaqMan SARS-CoV probe that hybridizes to the SARS-CoV nucleic acid, wherein the TaqMan SARS-CoV probe is labeled with a 5'-reporter dye and a 3'-quencher dye, performing one or more additional rounds of amplification, and detecting fluorescence of the 5'-reporter dye, thereby detecting a SARS-CoV. In a specific, non-limiting example, the amplification utilizes RT-PCR. In a further specific example of the provided method, at least one of the nucleic acid primers that hybridize to a SARS-CoV nucleic acid and/or the TaqMan SARS-CoV probe that hybridizes to the SARS-CoV nucleic acid includes a sequence as set forth in any one of SEQ ID NOs: 16-33.

In another embodiment, a method is provided for detecting a SARS-CoV in a biological sample that contains antibodies. This method includes contacting the biological sample with a SARS-CoV-specific antigen, wherein the antigen includes a SARS-CoV organism and determining whether a binding reaction occurs between the SARS-CoV-specific antigen and an antibody in the biological sample if such is present, thereby detecting SARS-CoV.

In a further embodiment, a method is provided for detecting a SARS-CoV in a biological sample that contains polypeptides and/or fragments thereof. This method includes contacting the biological sample with a SARS-CoV-specific antibody and determining whether a binding reaction occurs between the SARS-CoV-specific antibody and a SARS-CoV polypeptide or fragment thereof in the biological sample if

such is present, thereby detecting SARS-CoV. In a specific, non-limiting example, determining whether a binding reaction occurs between the SARS-CoV-specific antibody and a SARS-CoV polypeptide or fragment thereof is carried out in situ or in a tissue sample. In a further specific example, determining whether a binding reaction occurs between the SARS-CoV-specific antibody and a SARS-CoV polypeptide or fragment thereof includes an immunohistochemical assay.

An additional embodiment includes a kit for detecting a SARS-CoV in a sample, including a pair of nucleic acid primers that hybridize under stringent conditions to a SARS-CoV nucleic acid, wherein one primer is 5'-end labeled with a reporter dye. In a specific, non-limiting example, at least one of the nucleic acid primers that hybridize to a SARS-CoV nucleic acid includes a sequence as set forth in any one of SEQ ID NOs: 13-15.

Another example of the provided kit includes a pair of nucleic acid primers that hybridize under high stringency conditions to a SARS-CoV nucleic acid and a TaqMan SARS-CoV probe that hybridizes to the SARS-CoV nucleic acid, wherein the TaqMan SARS-CoV probe is labeled with a 5'-reporter dye and a 3'-quencher dye. In a specific, non-limiting example, at least one of the nucleic acid primers that hybridize to a SARS-CoV nucleic acid and/or the TaqMan SARS-CoV probe that hybridizes to the SARS-CoV nucleic acid includes a sequence as set forth in any one of SEQ ID NOs: 16-33.

Also disclosed herein is a composition including an isolated SARS-CoV organism. In one embodiment, the isolated SARS-CoV organism is an inactive isolated SARS-CoV organism. In another embodiment, the composition includes at least one component selected from the group consisting of pharmaceutically acceptable carriers, adjuvants and combinations of two or more thereof. In yet another embodiment, the composition is introduced into a subject, thereby eliciting an immune response against a SARS-CoV antigenic epitope in a subject.

IV. SARS-CoV Nucleotide and Amino Acid Sequences

The current disclosure provides an isolated SARS-CoV genome, isolated SARS-CoV polypeptides, and isolated nucleic acid molecules encoding the same. In one embodiment, the isolated SARS-CoV genome has a sequence as shown in SEQ ID NO: 1 or an equivalent thereof. Polynucleotides encoding a SARS-CoV polypeptide (encoded by an ORF from within the genome) are also provided, and are termed SARS-CoV nucleic acid molecules. These nucleic acid molecules include DNA, cDNA and RNA sequences which encode a SARS-CoV polypeptide. Specific, non-limiting examples of a SARS-CoV nucleic acid molecule encoding an ORF are nucleic acid 265 to nucleic acid 13,398 of SEQ ID NO: 1 (encoding SARS-CoV 1a, SEQ ID NO: 2), nucleic acid 13,398 to 21,482 of SEQ ID NO: 1 (encoding SARS-CoV 1b, SEQ ID NO: 3), nucleic acid 21,492 to 25,256 of SEQ ID NO: 1 (encoding SARS-CoV S, SEQ ID NO: 4), nucleic acid 25,268 to 26,089 of SEQ ID NO: 1 (encoding SARS-CoV X1, SEQ ID NO: 5), nucleic acid 25,689 to 26,150 of SEQ ID NO: 1 (encoding SARS-CoV X2, SEQ ID NO: 6), nucleic acid 26,117 to 26,344 of SEQ ID NO: 1 (encoding SARS-CoV E, SEQ ID NO: 7), nucleic acid 26,398 to 27,060 of SEQ ID NO: 1 (encoding SARS-CoV M, SEQ ID NO: 8), nucleic acid 27,074 to 27,262 of SEQ ID NO: 1 (encoding SARS-CoV X3, SEQ ID NO: 9), nucleic acid 27,273 to 27,638 of SEQ ID NO: 1 (encoding SARS-CoV X4, SEQ ID NO: 10), nucleic acid 27,864 to 28,115 of SEQ ID NO: 1 (encoding SARS-CoV X5, SEQ ID NO: 11), and nucleic acid 28,120 to 29,385 of SEQ ID NO: 1 (encoding SARS-CoV N, SEQ ID NO: 12).

Oligonucleotide primers and probes derived from the SARS-CoV genome (SEQ ID NO: 1) are also encompassed within the scope of the present disclosure. Such oligonucleotide primers and probes may comprise a sequence of at least about 15 consecutive nucleotides of the SARS-CoV nucleic acid sequence, such as at least about 20, 25, 30, 35, 40, 45, or 50 or more consecutive nucleotides. These primers and probes may be obtained from any region of the disclosed SARS-CoV genome (SEQ ID NO: 1), including particularly from any of the ORFs disclosed herein. Specific, non-limiting examples of oligonucleotide primers derived from the SARS-CoV genome (SEQ ID NO: 1) include: Cor-p-F2 (SEQ ID NO: 13), Cor-p-F3 (SEQ ID NO: 14), Cor-p-R1 (SEQ ID NO: 15), SARS1-F (SEQ ID NO: 16), SARS1—R (SEQ ID NO: 17), SARS2-F (SEQ ID NO: 19), SARS2—R (SEQ ID NO: 20), SARS3-F (SEQ ID NO: 22), SARS3—R (SEQ ID NO: 23), N3-F (SEQ ID NO: 25), N3-R (SEQ ID NO: 26), 3'NTR-F (SEQ ID NO: 28), 3'NTR-R (SEQ ID NO: 29), M-F (SEQ ID NO: 31), and M-R (SEQ ID NO: 32). Specific, non-limiting examples of oligonucleotide probes derived from the SARS-CoV genome (SEQ ID NO: 1) include: SARS1-P (SEQ ID NO: 18), SARS2-P (SEQ ID NO: 21), SARS3-P (SEQ ID NO: 24), N3-P (SEQ ID NO: 27), 3'NTR-P (SEQ ID NO: 30), and M-P (SEQ ID NO: 33).

Nucleic acid molecules encoding a SARS-CoV polypeptide can be operatively linked to regulatory sequences or elements. Regulatory sequences or elements include, but are not limited to promoters, enhancers, transcription terminators, a start codon (for example, ATG), stop codons, and the like.

Additionally, nucleic acid molecules encoding a SARS-CoV polypeptide can be inserted into an expression vector. Specific, non-limiting examples of vectors include, plasmids, bacteriophages, cosmids, animal viruses and yeast artificial chromosomes (YACs) (Burke et al., *Science* 236:806-12, 1987). Such vectors may then be introduced into a variety of hosts including somatic cells, and simple or complex organisms, such as bacteria, fungi (Timberlake and Marshall, *Science* 244:1313-17, 1989), invertebrates, plants (Gasser et al., *Plant Cell* 1:15-24, 1989), and animals (Pursel et al., *Science* 244:1281-88, 1989).

Transformation of a host cell with an expression vector carrying a nucleic acid molecule encoding a SARS-CoV polypeptide may be carried out by conventional techniques, as are well known to those skilled in the art. By way of example, where the host is prokaryotic, such as *E. coli*, competent cells that are capable of DNA uptake can be prepared from cells harvested after exponential growth phase and subsequently treated by the $CaCl_2$ method using procedures well known in the art. Alternatively, $MgCl_2$ or RbCl can be used. Transformation can also be performed after forming a protoplast of the host cell if desired, or by electroporation.

When the host is a eukaryote, methods of transfection of DNA, such as calcium phosphate coprecipitates, conventional mechanical procedures such as microinjection, electroporation, insertion of a plasmid encased in liposomes, or virus vectors, may be used. Eukaryotic cells can also be cotransformed with SARS-CoV nucleic acid molecules, and a second foreign DNA molecule encoding a selectable phenotype, such as the herpes simplex thymidine kinase gene. Another method is to use a eukaryotic viral vector, such as simian virus 40 or bovine papilloma virus, to transiently infect or transform eukaryotic cells and express the protein

(see, for example, *Eukaryotic Viral Vectors*, Cold Spring Harbor Laboratory, Gluzman ed., 1982).

The SARS-CoV polypeptides of this disclosure include proteins encoded by any of the ORFs disclosed herein, and equivalents thereof. Specific, non-limiting examples of SARS-CoV proteins are provided in SEQ ID NOs: 2-12. Fusion proteins are also contemplated that include a heterologous amino acid sequence chemically linked to a SARS-CoV polypeptide. Exemplary heterologous sequences include short amino acid sequence tags (such as six histidine residues), as well as a fusion of other proteins (such as c-myc or green fluorescent protein fusions). Epitopes of the SARS-CoV proteins, that are recognized by an antibody or that bind the major histocompatibility complex, and can be used to induce a SARS-CoV-specific immune response, are also encompassed by this disclosure.

Methods for expressing large amounts of protein from a cloned gene introduced into *E. coli* may be utilized for the purification and functional analysis of proteins. For example, fusion proteins consisting of amino terminal peptides encoded by a portion of the *E. coli* lacZ or trpE gene linked to SARS-CoV proteins may be used to prepare polyclonal and monoclonal antibodies against these proteins.

Intact native protein may also be produced in *E. coli* in large amounts for functional studies. Methods and plasmid vectors for producing fusion proteins and intact native proteins in bacteria are described by Sambrook et al. (ed.), *Molecular Cloning: A Laboratory Manual,* 2nd ed., vol. 1-3, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., 1989. Such fusion proteins may be made in large amounts, are easy to purify, and can be used to elicit antibody response. Native proteins can be produced in bacteria by placing a strong, regulated promoter and an efficient ribosome-binding site upstream of the cloned gene. If low levels of protein are produced, additional steps may be taken to increase protein production; if high levels of protein are produced, purification is relatively easy. Suitable methods are presented by Sambrook et al. (ed.), *Molecular Cloning: A Laboratory Manual,* 2nd ed., vol. 1-3, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., 1989, and are well known in the art. Often, proteins expressed at high levels are found in insoluble inclusion bodies. Methods for extracting proteins from these aggregates are described by Sambrook et al. (ed.), *Molecular Cloning: A Laboratory Manual,* 2nd ed., vol. 1-3, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., 1989.

Isolation and purification of recombinantly expressed proteins may be carried out by conventional means including preparative chromatography and immunological separations. Additionally, the proteins can be chemically synthesized by any of a number of manual or automated methods of synthesis known in the art.

V. Specific Binding Agents

The disclosure provides specific binding agents that bind to SARS-CoV polypeptides disclosed herein. The binding agent may be useful for purifying and detecting the polypeptides, as well as for detection and diagnosis of SARS-CoV. Examples of the binding agents are a polyclonal or monoclonal antibody, and fragments thereof, that bind to any of the SARS-CoV polypeptides disclosed herein.

Monoclonal or polyclonal antibodies may be raised to recognize a SARS-CoV polypeptide described herein, or a fragment or variant thereof. Optimally, antibodies raised against these polypeptides would specifically detect the polypeptide with which the antibodies are generated. That is, antibodies raised against a specific SARS-CoV polypeptide will recognize and bind that polypeptide, and will not substantially recognize or bind to other polypeptides or antigens. The determination that an antibody specifically binds to a target polypeptide is made by any one of a number of standard immunoassay methods; for instance, the Western blotting technique (Sambrook et al. (ed.), *Molecular Cloning: A Laboratory Manual,* 2nd ed., vol. 1-3, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., 1989).

Substantially pure SARS-CoV recombinant polypeptide antigens suitable for use as immunogen may be isolated from the transformed cells described above, using methods well known in the art. Monoclonal or polyclonal antibodies to the antigens may then be prepared.

Monoclonal antibodies to the polypeptides can be prepared from murine hybridomas according to the classic method of Kohler & Milstein (*Nature* 256:495-97, 1975), or a derivative method thereof. Briefly, a mouse is repetitively inoculated with a few micrograms of the selected protein immunogen (for example, a polypeptide comprising at least one SARS-CoV-specific epitope, a portion of a polypeptide comprising at least one SARS-CoV-specific epitope, or a synthetic peptide comprising at least one SARS-CoV-specific epitope) over a period of a few weeks. The mouse is then sacrificed, and the antibody-producing cells of the spleen isolated. The spleen cells are fused by means of polyethylene glycol with mouse myeloma cells, and the excess unfused cells destroyed by growth of the system on selective media comprising aminopterin (HAT media). The successfully fused cells are diluted and aliquots of the dilution placed in wells of a microtiter plate where growth of the culture is continued. Antibody-producing clones are identified by detection of antibody in the supernatant fluid of the wells by immunoassay procedures, such as ELISA, as originally described by Engvall (*Meth. Enzymol.,* 70:419-39, 1980), or a derivative method thereof. Selected positive clones can be expanded and their monoclonal antibody product harvested for use. Detailed procedures for monoclonal antibody production are described in Harlow and Lane, *Using Antibodies: A Laboratory Manual*, CSHL, New York, 1999.

Polyclonal antiserum containing antibodies can be prepared by immunizing suitable animals with a polypeptide comprising at least one SARS-CoV-specific epitope, a portion of a polypeptide comprising at least one SARS-CoV-specific epitope, or a synthetic peptide comprising at least one SARS-CoV-specific epitope, which can be unmodified or modified, to enhance immunogenicity.

Effective antibody production (whether monoclonal or polyclonal) is affected by many factors related both to the antigen and the host species. For example, small molecules tend to be less immunogenic than others and may require the use of carriers and adjuvant. Also, host animals vary in response to site of inoculations and dose, with either inadequate or excessive doses of antigen resulting in low titer antisera. Small doses (ng level) of antigen administered at multiple intradermal sites appear to be most reliable. An effective immunization protocol for rabbits can be found in Vaitukaitis et al. (*J. Clin. Endocrinol. Metab.,* 33:988-91, 1971).

Booster injections can be given at regular intervals, and antiserum harvested when the antibody titer thereof, as determined semi-quantitatively, for example, by double immunodiffusion in agar against known concentrations of the antigen, begins to fall. See, for example, Ouchterlony et al., *Handbook of Experimental Immunology*, Wier, D. (ed.), Chapter 19, Blackwell, 1973. A plateau concentration of antibody is usually in the range of 0.1 to 0.2 mg/ml of serum

(about 12 μM). Affinity of the antisera for the antigen is determined by preparing competitive binding curves, as described, for example, by Fisher (*Manual of Clinical Immunology*, Ch. 42, 1980).

Antibody fragments may be used in place of whole antibodies and may be readily expressed in prokaryotic host cells. Methods of making and using immunologically effective portions of monoclonal antibodies, also referred to as "antibody fragments," are well known and include those described in Better & Horowitz, *Methods Enzymol.* 178:476-96, 1989; Glockshuber et al., *Biochemistry* 29:1362-67, 1990; and U.S. Pat. Nos. 5,648,237 (Expression of Functional Antibody Fragments); 4,946,778 (Single Polypeptide Chain Binding Molecules); and 5,455,030 (Immunotherapy Using Single Chain Polypeptide Binding Molecules), and references cited therein. Conditions whereby a polypeptide/binding agent complex can form, as well as assays for the detection of the formation of a polypeptide/binding agent complex and quantitation of binding affinities of the binding agent and polypeptide, are standard in the art. Such assays can include, but are not limited to, Western blotting, immunoprecipitation, immunofluorescence, immunocytochemistry, immunohistochemistry, fluorescence activated cell sorting (FACS), fluorescence in situ hybridization (FISH), immunomagnetic assays, ELISA, ELISPOT (Coligan et al., *Current Protocols in Immunology*, Wiley, NY, 1995), agglutination assays, flocculation assays, cell panning, and the like, as are well known to one of skill in the art.

Binding agents of this disclosure can be bound to a substrate (for example, beads, tubes, slides, plates, nitrocellulose sheets, and the like) or conjugated with a detectable moiety, or both bound and conjugated. The detectable moieties contemplated for the present disclosure can include, but are not limited to, an immunofluorescent moiety (for example, fluorescein, rhodamine), a radioactive moiety (for example, ^{32}P, ^{125}I, ^{35}S), an enzyme moiety (for example, horseradish peroxidase, alkaline phosphatase), a colloidal gold moiety, and a biotin moiety. Such conjugation techniques are standard in the art (for example, see Harlow and Lane, *Using Antibodies: A Laboratory Manual*, CSHL, New York, 1999; Yang et al., *Nature,* 382:319-24, 1996).

VI. Detection and Diagnosis of SARS-CoV

A. Nucleic Acid Based Methods of Detection and Diagnosis

A major application of the SARS-CoV sequence information presented herein is in the area of detection and diagnostic testing for SARS-CoV infection. Methods for screening a subject to determine if the subject has been or is currently infected with SARS-CoV are disclosed herein.

One such method includes providing a sample, which sample includes a nucleic acid such as DNA or RNA, and providing an assay for detecting in the sample the presence of a SARS-CoV nucleic acid molecule. Suitable samples include all biological samples useful for detection of viral infection in subjects, including, but not limited to, cells, tissues (for example, lung and kidney), bodily fluids (for example, blood, serum, urine, saliva, sputum, and cerebrospinal fluid), bone marrow aspirates, BAL, and oropharyngeal wash. Additional suitable samples include all environmental samples useful for detection of viral presence in the environment, including, but not limited to, a sample obtained from inanimate objects or reservoirs within an indoor or outdoor environment. The detection in the sample of a SARS-CoV nucleic acid molecule may be performed by a number of methodologies, non-limiting examples of which are outlined below.

In one embodiment, detecting in the sample the presence of a SARS-CoV nucleic acid molecule includes the amplification of a SARS-CoV nucleic acid sequence (or a fragment thereof). Any nucleic acid amplification method can be used. In one specific, non-limiting example, PCR is used to amplify the SARS-CoV nucleic acid sequence(s). In another non-limiting example, RT-PCR can be used to amplify the SARS-CoV nucleic acid sequences. In an additional non-limiting example, transcription-mediated amplification can be used to amplify the SARS-CoV nucleic acid sequences.

In some embodiments, a pair of SARS-CoV-specific primers are utilized in the amplification reaction. One or both of the primers can be end-labeled (for example, radiolabeled, fluoresceinated, or biotinylated). In one specific, non-limiting example, at least one of the primers is 5'-end labeled with the reporter dye 6-carboxyfluorescein (6-FAM). The pair of primers includes an upstream primer (which binds 5' to the downstream primer) and a downstream primer (which binds 3' to the upstream primer). In one embodiment, either the upstream primer or the downstream primer is labeled. Specific, non-limiting examples of SARS-CoV-specific primers include, but are not limited to: Cor-p-F2 (SEQ ID NO: 13), Cor-p-F3 (SEQ ID NO: 14), Cor-p-R1 (SEQ ID NO: 15), SARS1-F (SEQ ID NO: 16), SARS1-R (SEQ ID NO: 17), SARS2-F (SEQ ID NO: 19), SARS2-R (SEQ ID NO: 20), SARS3-F (SEQ ID NO: 22), SARS3R (SEQ ID NO: 23), N3-F (SEQ ID NO: 25), N3-R (SEQ ID NO: 26), 3'NTR-F (SEQ ID NO: 28), 3'NTR-R (SEQ ID NO: 29), M-F (SEQ ID NO: 31), and M-R (SEQ ID NO: 32). Additional primer pairs can be generated, for instance, to amplify any of the specific ORFs described herein, using well known primer design principles and methods.

In one specific, non-limiting example, electrophoresis is used to detect amplified SARS-CoV-specific sequences. Electrophoresis can be automated using many methods well know in the art. In one embodiment, a genetic analyzer is used, such as an ABI 3100 Prism Genetic Analyzer (PE Applied Biosystems, Foster City, Calif.), wherein the bands are analyzed using GeneScan software (PE Applied Biosystems, Foster City, Calif.).

In another specific, non-limiting example, hybridization assays are used to detect amplified SARS-CoV-specific sequences using distinguishing oligonucleotide probes. Such probes include "TaqMan" probes. TaqMan probes consist of an oligonucleotide with a reporter at the 5'-end and a quencher at the 3'-end. In one specific, non-limiting example, the reporter is 6-FAM and the quencher is Blackhole Quencher (Biosearch Tech., Inc., Novato, Calif.). When the probe is intact, the proximity of the reporter to the quencher results in suppression of reporter fluorescence, primarily by fluorescence resonance energy transfer. If the target of interest is present, the TaqMan probe specifically hybridizes between the forward and reverse primer sites during the PCR annealing step. In the process of PCR elongation, the 5'-3' nucleolytic activity of the Taq DNA polymerase cleaves the hybridized probe between the reporter and the quencher. The probe fragments are then displaced from the target, and polymerization of the strand continues. Taq DNA polymerase does not cleave non-hybridized probe, and cleaves the hybridized probe only during polymerization. The 3'-end of the probe is blocked to prevent extension of the probe during PCR. The 5'-3' nuclease cleavage of the hybridized probe occurs in every cycle and does not interfere with the exponential accumulation of PCR product. Accumulation of PCR products is detected directly by monitoring the increase in fluorescence of the released reporter. The increase in fluorescence signal is detected only if the target sequence is complementary to the

probe and is amplified during PCR. Therefore, non-specific amplification is not detected. SARS-CoV-specific TaqMan probes of the present disclosure include, but are not limited to: SARS1-P (SEQ ID NO: 18), SARS2-P (SEQ ID NO: 21), SARS3-P (SEQ ID NO: 24), N3-P (SEQ ID NO: 27), 3'NTR-P (SEQ ID NO: 30), and M-P (SEQ ID NO: 33), and hybridization assays include, but are not limited to, a real-time RT-PCR assay.

B. Protein Based Methods of Detection and Diagnosis

The present disclosure further provides methods of detecting a SARS-CoV antigen in a sample, and/or diagnosing SARS-CoV infection in a subject by detecting a SARS-CoV antigen. Examples of such methods comprise contacting the sample with a SARS-CoV-specific binding agent under conditions whereby an antigen/binding agent complex can form; and detecting formation of the complex, thereby detecting SARS-CoV antigen in a sample and/or diagnosing SARS-CoV infection in a subject. It is contemplated that at least certain antigens will be on an intact SARS-CoV virion, will be a SARS-CoV-encoded protein displayed on the surface of a SARS-CoV-infected cell expressing the antigen, or will be a fragment of the antigen. Contemplated samples subject to analysis by these methods can comprise any sample, such as a clinical sample, useful for detection of viral infection in a subject.

Methods for detecting antigens in a sample are discussed, for example, in Ausubel et al. *Short Protocols in Molecular Biology,* 4th ed., John Wiley & Sons, Inc., 1999. Enzyme immunoassays such as IFA, ELISA and immunoblotting can be readily adapted to accomplish the detection of SARS-CoV antigens according to the methods of this disclosure. An ELISA method effective for the detection of soluble SARS-CoV antigens is the direct competitive ELISA. This method is most useful when a specific SARS-CoV antibody and purified SARS-CoV antigen are available. Briefly: 1) coat a substrate (for example, a microtiter plate) with a sample suspected of containing a SARS-CoV antigen; 2) contact the bound SARS-CoV antigen with a SARS-CoV-specific antibody bound to a detectable moiety (for example, horseradish peroxidase enzyme or alkaline phosphatase enzyme); 3) add purified inhibitor SARS-CoV antigen; 4) contact the above with the substrate for the enzyme; and 5) observe/measure inhibition of color change or fluorescence and quantitate antigen concentration (for example, using a microtiter plate reader).

An additional ELISA method effective for the detection of soluble SARS-CoV antigens is the antibody-sandwich ELISA. This method is frequently more sensitive in detecting antigen than the direct competitive ELISA method. Briefly: 1) coat a substrate (for example, a microtiter plate) with a SARS-CoV-specific antibody; 2) contact the bound SARS-CoV antibody with a sample suspected of containing a SARS-CoV antigen; 3) contact the above with SARS-CoV-specific antibody bound to a detectable moiety (for example, horseradish peroxidase enzyme or alkaline phosphatase enzyme); 4) contact the above with the substrate for the enzyme; and 5) observe/measure color change or fluorescence and quantitate antigen concentration (for example, using a microtiter plate reader).

An ELISA method effective for the detection of cell-surface SARS-CoV antigens is the direct cellular ELISA. Briefly, cells suspected of exhibiting a cell-surface SARS-CoV antigen are fixed (for example, using glutaraldehyde) and incubated with a SARS-CoV-specific antibody bound to a detectable moiety (for example, horseradish peroxidase enzyme or alkaline phosphatase enzyme). Following a wash to remove unbound antibody, substrate for the enzyme is added and color change or fluorescence is observed/measured.

The present disclosure further provides methods of detecting a SARS-CoV-reactive antibody in a sample, and/or diagnosing SARS-CoV infection in a subject by detecting a SARS-CoV-reactive antibody. Examples of such methods comprise contacting the sample with a SARS-CoV polypeptide of this disclosure under conditions whereby a polypeptide/antibody complex can form; and detecting formation of the complex, thereby detecting SARS-CoV antibody in a sample and/or diagnosing SARS-CoV infection in a subject. Contemplated samples subject to analysis by these methods can comprise any sample, such as a clinical sample, as described herein as being useful for detection of viral infection in a subject.

Methods for detecting antibodies in a sample are discussed, for example, in Ausubel et al. *Short Protocols in Molecular Biology,* 4th ed., John Wiley & Sons, Inc., 1999. Enzyme immunoassays such as IFA, ELISA and immunoblotting can be readily adapted to accomplish the detection of SARS-CoV antibodies according to the methods of this disclosure. An ELISA method effective for the detection of specific SARS-CoV antibodies is the indirect ELISA method. Briefly: 1) bind a SARS-CoV polypeptide to a substrate (for example, a microtiter plate); 2) contact the bound polypeptide with a sample suspected of containing SARS-CoV antibody; 3) contact the above with a secondary antibody bound to a detectable moiety which is reactive with the bound antibody (for example, horseradish peroxidase enzyme or alkaline phosphatase enzyme); 4) contact the above with the substrate for the enzyme; and 5) observe/measure color change or fluorescence.

Another immunologic technique that can be useful in the detection of SARS-CoV antibodies uses monoclonal antibodies for detection of antibodies specifically reactive with SARS-CoV polypeptides in a competitive inhibition assay. Briefly, a sample suspected of containing SARS-CoV antibodies is contacted with a SARS-CoV polypeptide of this disclosure which is bound to a substrate (for example, a microtiter plate). Excess sample is thoroughly washed away. A labeled (for example, enzyme-linked, fluorescent, radioactive, and the like) monoclonal antibody specific for the SARS-CoV polypeptide is then contacted with any previously formed polypeptide-antibody complexes and the amount of monoclonal antibody binding is measured. The amount of inhibition of monoclonal antibody binding is measured relative to a control (no monoclonal antibody), allowing for detection and measurement of antibody in the sample. The degree of monoclonal antibody inhibition can be a very specific assay for detecting SARS-CoV. Monoclonal antibodies can also be used for direct detection of SARS-CoV in cells or tissue samples by, for example, IFA analysis according to standard methods.

As a further example, a micro-agglutination test can be used to detect the presence of SARS-CoV antibodies in a sample. Briefly, latex beads, red blood cells or other agglutinable particles are coated with a SARS-CoV polypeptide of this disclosure and mixed with a sample, such that antibodies in the sample that are specifically reactive with the antigen crosslink with the antigen, causing agglutination. The agglutinated polypeptide-antibody complexes form a precipitate, visible with the naked eye or measurable by spectrophotometer. In a modification of the above test, SARS-CoV-specific antibodies of this disclosure can be bound to the agglutinable particles and SARS-CoV antigen in the sample thereby detected.

VII. Pharmaceutical and Immune Stimulatory Compositions and Uses Thereof

Pharmaceutical compositions including SARS-CoV nucleic acid sequences, SARS-CoV polypeptides, or antibodies that bind these polypeptides, are also encompassed by the present disclosure. These pharmaceutical compositions include a therapeutically effective amount of one or more SARS-CoV polypeptides, one or more nucleic acid molecules encoding a SARS-CoV polypeptide, or an antibody that binds a SARS-CoV polypeptide, in conjunction with a pharmaceutically acceptable carrier.

Disclosed herein are substances suitable for use as immune stimulatory compositions for the inhibition or treatment of SARS. Particular immune stimulatory compositions are directed against SARS-CoV, and include antigens obtained from SARS-CoV. In one embodiment, an immune stimulatory composition contains attenuated SARS-CoV. Methods of viral attenuation are well known in the art, and include, but are not limited to, high serial passage (for example, in susceptible host cells under specific environmental conditions to select for attenuated virions), exposure to a mutagenic agent (for example, a chemical mutagen or radiation), genetic engineering using recombinant DNA technology (for example, using gene replacement or gene knockout to disable one or more viral genes), or some combination thereof.

In another embodiment, the immune stimulatory composition contains inactivated SARS-CoV. Methods of viral inactivation are well known in the art, and include, but are not limited to, heat and chemicals (for example, formalin, β-propiolactone, and ethylenimines).

In yet another embodiment, the immune stimulatory composition contains a nucleic acid vector that includes SARS-CoV nucleic acid molecules described herein, or that includes a nucleic acid sequence encoding an immunogenic polypeptide or polypeptide fragment of SARS-CoV or derived from SARS-CoV, such as a polypeptide that encodes a surface protein of SARS-CoV.

In a further embodiment, the immune stimulatory composition contains a SARS-CoV subunit, such as glycoprotein, major capsid protein, or other gene products found to elicit humoral and/or cell mediated immune responses.

The provided immune stimulatory SARS-CoV polypeptides, constructs or vectors encoding such polypeptides, are combined with a pharmaceutically acceptable carrier or vehicle for administration as an immune stimulatory composition to human or animal subjects. In some embodiments, more than one immune stimulatory SARS-CoV polypeptide may be combined to form a single preparation.

The immunogenic formulations may be conveniently presented in unit dosage form and prepared using conventional pharmaceutical techniques. Such techniques include the step of bringing into association the active ingredient and the pharmaceutical carrier(s) or excipient(s). In general, the formulations are prepared by uniformly and intimately bringing into association the active ingredient with liquid carriers. Formulations suitable for parenteral administration include aqueous and non-aqueous sterile injection solutions which may contain anti-oxidants, buffers, bacteriostats and solutes which render the formulation isotonic with the blood of the intended recipient; and aqueous and non-aqueous sterile suspensions which may include suspending agents and thickening agents. The formulations may be presented in unit-dose or multi-dose containers, for example, sealed ampules and vials, and may be stored in a freeze-dried (lyophilized) condition requiring only the addition of a sterile liquid carrier, for example, water for injections, immediately prior to use. Extemporaneous injection solutions and suspensions may be prepared from sterile powders, granules and tablets commonly used by one of ordinary skill in the art.

In certain embodiments, unit dosage formulations are those containing a dose or unit, or an appropriate fraction thereof, of the administered ingredient. It should be understood that in addition to the ingredients particularly mentioned above, formulations encompassed herein may include other agents commonly used by one of ordinary skill in the art.

The compositions provided herein, including those for use as immune stimulatory compositions, may be administered through different routes, such as oral, including buccal and sublingual, rectal, parenteral, aerosol, nasal, intramuscular, subcutaneous, intradermal, and topical. They may be administered in different forms, including but not limited to solutions, emulsions and suspensions, microspheres, particles, microparticles, nanoparticles, and liposomes.

The volume of administration will vary depending on the route of administration. By way of example, intramuscular injections may range from about 0.1 ml to about 1.0 ml. Those of ordinary skill in the art will know appropriate volumes for different routes of administration.

A relatively recent development in the field of immune stimulatory compounds (for example, vaccines) is the direct injection of nucleic acid molecules encoding peptide antigens (broadly described in Janeway & Travers, *Immunobiology: The Immune System In Health and Disease*, page 13.25, Garland Publishing, Inc., New York, 1997; and McDonnell & Askari, N. *Engl. J. Med.* 334:42-45, 1996). Vectors that include nucleic acid molecules described herein, or that include a nucleic acid sequence encoding an immunogenic SARS-CoV polypeptide may be utilized in such DNA vaccination methods.

Thus, the term "immune stimulatory composition" as used herein also includes nucleic acid vaccines in which a nucleic acid molecule encoding a SARS-CoV polypeptide is administered to a subject in a pharmaceutical composition. For genetic immunization, suitable delivery methods known to those skilled in the art include direct injection of plasmid DNA into muscles (Wolff et al., *Hum. Mol. Genet.* 1:363, 1992), delivery of DNA complexed with specific protein carriers (Wu et al., *J. Biol. Chem.* 264:16985, 1989), co-precipitation of DNA with calcium phosphate (Benvenisty and Reshef, *Proc. Natl. Acad. Sci.* 83:9551, 1986), encapsulation of DNA in liposomes (Kaneda et al., *Science* 243:375, 1989), particle bombardment (Tang et al., *Nature* 356:152, 1992; Eisenbraun et al., *DNA Cell Biol.* 12:791, 1993), and in vivo infection using cloned retroviral vectors (Seeger et al., *Proc. Natl. Acad. Sci.* 81:5849, 1984). Similarly, nucleic acid vaccine preparations can be administered via viral carrier.

The amount of immunostimulatory compound in each dose of an immune stimulatory composition is selected as an amount that induces an immunostimulatory or immunoprotective response without significant, adverse side effects. Such amount will vary depending upon which specific immunogen is employed and how it is presented. Initial injections may range from about 1 μg to about 1 mg, with some embodiments having a range of about 10 μg to about 800 μg, and still other embodiments a range of from about 25 μg to about 500 μg. Following an initial administration of the immune stimulatory composition, subjects may receive one or several booster administrations, adequately spaced. Booster administrations may range from about 1 μg to about 1 mg, with other embodiments having a range of about 10 μg to about 750 μg, and still others a range of about 50 μg to about 500 μg.

Periodic boosters at intervals of 1-5 years, for instance three years, may be desirable to maintain the desired levels of protective immunity.

It is also contemplated that the provided immunostimulatory molecules and compositions can be administered to a subject indirectly, by first stimulating a cell in vitro, which stimulated cell is thereafter administered to the subject to elicit an immune response. Additionally, the pharmaceutical or immune stimulatory compositions or methods of treatment may be administered in combination with other therapeutic treatments.

VIII. Kits

Also provided herein are kits useful in the detection and/or diagnosis of SARS-CoV. This includes kits for use with nucleic acid and protein detection methods, such as those disclosed herein.

The SARS-CoV-specific oligonucleotide primers and probes described herein can be supplied in the form of a kit for use in detection of SARS-CoV. In such a kit, an appropriate amount of one or more of the oligonucleotides is provided in one or more containers, or held on a substrate. An oligonucleotide primer or probe can be provided in an aqueous solution or as a freeze-dried or lyophilized powder, for instance. The container(s) in which the oligonucleotide(s) are supplied can be any conventional container that is capable of holding the supplied form, for instance, microfuge tubes, ampoules, or bottles. In some applications, pairs of primers are provided in pre-measured single use amounts in individual (typically disposable) tubes or equivalent containers. With such an arrangement, the sample to be tested for the presence of a SARS-CoV nucleic acid can be added to the individual tubes and amplification carried out directly.

The amount of each oligonucleotide supplied in the kit can be any appropriate amount, and can depend on the market to which the product is directed. For instance, if the kit is adapted for research or clinical use, the amount of each oligonucleotide primer provided would likely be an amount sufficient to prime several PCR amplification reactions. General guidelines for determining appropriate amounts can be found, for example, in Sambrook et al., *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., 2001; Ausubel et al. (eds.), *Short Protocols in Molecular Biology*, John Wiley and Sons, New York, N.Y., 1999; and Innis et al., *PCR Applications, Protocols for Functional Genomics*, Academic Press, Inc., San Diego, Calif., 1999. A kit can include more than two primers, in order to facilitate the amplification of a larger number of SARS-CoV nucleotide sequences.

In some embodiments, kits also include one or more reagents necessary to carry out in vitro amplification reactions, including DNA sample preparation reagents, appropriate buffers (for example, polymerase buffer), salts (for example, magnesium chloride), and deoxyribonucleotides (dNTP5).

Kits can include either labeled or unlabeled oligonucleotide primers and/or probes for use in detection of SARS-CoV nucleotide sequences. The appropriate sequences for such a probe will be any sequence that falls between the annealing sites of the two provided oligonucleotide primers, such that the sequence that the probe is complementary to is amplified during the amplification reaction.

One or more control sequences for use in the amplification reactions also can be supplied in the kit. In other particular embodiments, the kit includes equipment, reagents, and instructions for extracting and/or purifying nucleotides from a sample.

Kits for the detection of SARS-CoV antigen include for instance at least one SARS-CoV antigen-specific binding agent (for example, a polyclonal or monoclonal antibody or antibody fragment). The kits may also include means for detecting antigen:specific binding agent complexes, for instance the specific binding agent may be detectably labeled. If the specific binding agent is not labeled, it may be detected by second antibodies or protein A, for example, which may also be provided in some kits in one or more separate containers. Such techniques are well known.

Another example of an assay kit provided herein is a recombinant SARS-CoV-specific polypeptide (or fragment thereof) as an antigen and an enzyme-conjugated anti-human antibody as a second antibody. Examples of such kits also can include one or more enzymatic substrates. Such kits can be used to test if a sample from a subject contains antibodies against a SARS-CoV-specific protein.

The subject matter of the present disclosure is further illustrated by the following non-limiting Examples.

EXAMPLES

Example 1

Isolation and Characterization of SARS-CoV

Virus Isolation and Ultrastructural Characterization

This example describes the original isolation and characterization of a new human coronavirus from patients with SARS.

A variety of clinical specimens (blood, serum, material from oropharyngeal swabs or washings, material from nasopharyngeal swabs, and tissues of major organs collected at autopsy) from patients meeting the case definition of SARS were sent to the Centers for Disease Control and Prevention (CDC) as part of the etiologic investigation of SARS. These samples were inoculated onto a number of continuous cell lines, including Vero E6, NCI-H292, MDCK, LLC-MK2, and B95-8 cells, and into suckling ICR mice by the intracranial and intraperitoneal routes. All cultures were observed daily for CPE. Maintenance medium was replenished at day seven, and cultures were terminated fourteen days after inoculation. Any cultures exhibiting identifiable CPE were subjected to several procedures to identify the cause of the effect. Suckling mice were observed daily for fourteen days, and any sick or dead mice were further tested by preparing a brain suspension that was filtered and subcultured. Mice that remained well after fourteen days were killed, and their test results were recorded as negative.

Two cell lines, Vero E6 cells and NCI-H292 cells, inoculated with oropharyngeal specimens from Patient 16 (a 46 year old male physician with an epidemiologic link to a hospital with multiple SARS patients) initially showed CPE (Table 1)

TABLE 1

Specimens from patients with SARS that were positive for SARS-CoV by one or more methods*.

Patient No	Exposure and Setting	Age/Sex	Findings on Chest Radiograph	Hospital-ization	Serologic Results	Specimen	Isolation	RT-PCR
1	Singapore, hospital	53 yr/F	Pneumonia	Yes	+	Nasal, oropharyngeal swabs	–	Not done
2†	Hong Kong, hotel	36 yr/F	Pneumonia	Yes	+	Nasal, swab	–	Not done
3	Hong Kong, hotel	22 yr/M	Pneumonia	Yes	+	Swab	–	–
4†	Hong Kong, hotel	39 yr/M	Pneumonia	Yes	+	Nasal, pharyngeal swab	–	–
5	Hong Kong, hotel	49 Yr/M	Pneumonia	Yes	Not done	Sputum	+	+
6‡	Hong Kong, hotel	46 yr/M	Pneumonia	Yes	+	Kidney, lung, broncho-alveolar lavage	+§	+
7	Vietnam, hospital	Adult/ unknown	Pneumonia	Yes	–	Oropharyngeal wash	+	+
8	Vietnam, hospital	Adult/ unknown	Pneumonia	Yes	–	Oropharyngeal wash	–	+
9	Vietnam, hospital	Adult/ unknown	Pneumonia	Yes	–	Oropharyngeal wash	–	+
10	Vietnam, hospital	Adult/ unknown	Pneumonia	Yes	–	Oropharyngeal wash	–	+
11	Vietnam, hospital	Adult/ unknown	Pneumonia	Yes	–	Oropharyngeal wash	–	+
12	Vietnam, hospital	Adult/ unknown	Pneumonia	Yes	–	Oropharyngeal wash	–	+
13	Vietnam, hospital	Adult/ unknown	Pneumonia	Yes	–	Oropharyngeal wash	+	+
14	Vietnam, hospital	Adult/ unknown	Pneumonia	Yes	–	Oropharyngeal wash	–	+
15	Vietnam, hospital	Adult/ unknown	Pneumonia	Yes	–	Oropharyngeal wash	–	+
16	Vietnam, hospital	46 yr/M	Pneumonia	Yes	+	Nasal, oropharyngeal swabs	+¶	+
17	Canada, family	43 yr/M	Pneumonia	Yes	Not done	Lung, bone marrow	–	
18	Taiwan, family	51 yr/F	Pneumonia	Yes	–	Sputum	–	+
19	Hong Kong, hotel	Adult/F	Pneumonia	Yes	+	Oropharyngeal wash	–	+

*Plus signs denote positive results, and minus signs negative results. The serologic and RT-PCR assays were not necessarily performed on samples obtained at the same time.

†This was a late specimen, antibody positive at first sample.

‡Travel included China, Hong Kong (hotel), and Hanoi (the patient was the index patient in the French Hospital).

§Isolation was from the kidney only.

¶Isolation was from the oropharyngeal only.

The CPE in the Vero E6 cells was first noted on the fifth day post-inoculation; it was focal, with cell rounding and a refractive appearance in the affected cells that was soon followed by cell detachment (FIG. **1**A). The CPE spread quickly to involve the entire cell monolayer within 24 to 48 hours. Subculture of material after preparation of a master seed stock (used for subsequent antigen production) resulted in the rapid appearance of CPE, as noted above, and in complete destruction of the monolayer in the inoculated flasks within 48 hours. Similar CPE was also noted in four additional cultures: three cultures of respiratory specimens (two oropharyngeal washes and one sputum specimen) and one culture of a suspension of kidney tissue obtained at autopsy. In these specimens, the initial CPE was observed between day two and day four and, as noted above, the CPE rapidly progressed to involve the entire cell monolayer.

Tissue culture samples showing CPE were prepared for electron-microscopical examination. Negative-stain electron-microscopical specimens were prepared by drying culture supernatant, mixed 1:1 with 2.5% paraformaldehyde, onto Formvarcarbon-coated grids and staining with 2% methylamine tungstate. Thin-section electron-microscopical specimens were prepared by fixing a washed cell pellet with 2.5% glutaraldehyde and embedding the cell pellet in epoxy resin. In addition, a master seed stock was prepared from the remaining culture supernatant and cells by freeze-thawing the culture flask, clarifying the thawed contents by centrifugation at 1000×g, and dispensing the supernatant into aliquots stored in gas phase over liquid nitrogen. The master seed stock was subcultured into 850-cm^2 roller bottles of Vero E6 cells for the preparation of formalin-fixed positive control cells for immunohistochemical analysis, mixed with normal Vero E6 cells, and gamma-irradiated for preparation of spot slides for IFA tests or extracted with detergent and gamma-irradiated for use as an ELISA antigen for antibody tests.

Examination of CPE-positive Vero E6 cells by thin-section electron microscopy revealed characteristic coronavirus particles within the cisternae of the rough endoplasmic reticulum and in vesicles (FIG. **2**A) (Becker et al., *J. Virol.* 1:1019-27, 1967; Oshiro et al. *J. Gen. Virol.* 12:161-8, 1971). Extracellular particles were found in large clusters and adhering to the surface of the plasma membrane. Negative-stain electron microscopy identified coronavirus particles, 80 to 140 nm in diameter, with 20- to 40-nm complex surface projections

surrounding the periphery (FIG. 2B). Hemagglutinin esterase-type glycoprotein projections were not seen.

The isolation and growth of a human-derived coronavirus in Vero E6 cells were unexpected. The previously known human coronaviruses are notably fastidious, preferring select cell lines, organ culture, or suckling mice for propagation. The only human or animal coronavirus which has been shown to grow in Vero E6 cells is PEDV, and it requires the addition of trypsin to culture medium for growth in the cells. Moreover, PEDV adapted to growth in Vero E6 cells results in a strikingly different CPE, with cytoplasmic vacuoles and the formation of large syncytia. Syncytial cells were only observed occasionally in monolayers of Vero E6 cells infected with the SARS-CoV; they clearly do not represent the dominant CPE.

Reverse Transcription-Polymerase Chain Reaction and Sequencing

For RT-PCR assays, cell-culture supernatants were placed in lysis buffer. RNA extracts were prepared from 100 μl of each specimen (or culture supernatant) with the automated NucliSens extraction system (bioMérieux, Durham, N.C.). Initially, degenerate, inosine-containing primers IN-2 (+) 5'GGGTTGGGACTA TCCTAAGTGTGA3' (SEQ ID NO: 34) and IN-4(−) 5'TAACACACAACICCATCA TCA3' (SEQ ID NO: 35) were designed to anneal to sites encoding conserved amino acid motifs that were identified on the basis of alignments of available coronavirus ORF1a, ORF1b, S, HE, M, and N gene sequences. Additional, SARS-specific, primers Cor-p-F2 (+) 5'CTAACATGCTTAGGATAATGG3' (SEQ ID NO: 13), Cor-p-F3 (+) 5'GCCTCTCTTGTTCTTGCTCGC3' (SEQ ID NO: 14), and Cor-p-R1 (−) 5' CAGGTAAGCGTAAAACTCATC3' (SEQ ID NO: 15) were designed as sequences were generated from RT-PCR products amplified with the degenerate primers. These SARS-specific primers were used to test patient specimens for SARS (see below). Primers used for specific amplification of human metapneumovirus have been described by Falsey et al. (*J. Infect. Dis.* 87:785-90, 2003).

For RT-PCR products of less than 3 kb, cDNA was synthesized in a 20 μl reaction mixture containing 500 ng of RNA, 200 U of Superscript™ II reverse transcriptase (Invitrogen Life Technologies, Carlsbad, Calif.), 40 U of RNasin (Promega Corp., Madison, Wis.), 100 mM each dNTP (Roche Molecular Biochemicals, Indianapolis, Ind.), 4 μl of 5× reaction buffer (Invitrogen Life Technologies, Carlsbad, Calif.), and 200 pmol of the reverse primer. The reaction mixture, except for the reverse transcriptase, was heated to 70° C. for 2 minutes, cooled to 4° C. for 5 minutes and then heated to 42° C. in a thermocycler. The mixture was held at 42° C. for 4 minutes, and then the reverse transcriptase was added, and the reactions were incubated at 42° C. for 45 minutes. Two microliters of the cDNA reaction was used in a 50 μl PCR reaction containing 67 mM Tris-HCl (pH 8.8), 1 mM each primer, 17 mM ammonium sulfate, 6 mM EDTA, 2 mM $MgCl_2$, 200 mM each dNTP, and 2.5 U of Taq DNA polymerase (Roche Molecular Biochemicals, Indianapolis, Ind.). The thermocycler program for the PCR consisted of 40 cycles of denaturation at 95° C. for 30 seconds, annealing at 42° C. for 30 seconds, and extension at 65° C. for 30 seconds. For SARS-CoV-specific primers, the annealing temperature was increased to 55° C.

For amplification of fragments longer than 3 kb, regions of the genome between sections of known sequence were amplified by means of a long RT-PCR protocol and SARS-CoV-specific primers. First-strand cDNA synthesis was performed at 42° C. or 50° C. using Superscript™ II RNase H reverse transcriptase (Invitrogen Life Technologies, Carlsbad, Calif.) according to the manufacturer's instructions with minor modifications. Coronavirus-specific primers (500 ng) and SARS-CoV RNA (350 ng) were combined with the PCR Nucleotide Mix (Roche Molecular Biochemicals, Indianapolis, Ind.), heated for 1 minute at 94° C., and cooled to 4° C. in a thermocycler. The 5× first-strand buffer, dithiothreitol (Invitrogen Life Technologies, Carlsbad, Calif.), and Protector RNase Inhibitor (Roche Molecular Biochemicals, Indianapolis, Ind.) were added, and the samples were incubated at 42° C. or 50° C. for 2 minutes. After reverse transcriptase (200 U) was added, the samples were incubated at 42° C. or 50° C. for 1.5 to 2 hours. Samples were inactivated at 70° C. for 15 minutes and subsequently treated with 2 U of RNase H (Roche Molecular Biochemicals, Indianapolis, Ind.) at 37° C. for 30 minutes. Long RT-PCR amplification of 5- to 8-kb fragments was performed using Taq Plus Precision (Stratagene, La Jolla, Calif.) and AmpliWax PCR Gem 100 beads (Applied Biosystems; Foster City, Calif.) for "hot start" PCR with the following thermocycling parameters: denaturation at 94° C. for 1 minute followed by 35 cycles of 94° C. for 30 seconds, 55° C. for 30 seconds, an increase of 0.4 degrees per second up to 72° C., and 72° C. for 7 to 10 minutes, with a final extension at 72° C. for 10 minutes. RT-PCR products were separated by electrophoresis on 0.9% agarose TAE gels and purified by use of a QIAquick Gel Extraction Kit (Qiagen, Inc., Santa Clarita, Calif.).

In all cases, the RT-PCR products were gel-isolated and purified for sequencing by means of a QIAquick Gel Extraction kit (Qiagen, Inc., Santa Clarita, Calif.). Both strands were sequenced by automated methods, using fluorescent dideoxy-chain terminators (Applied Biosystems; Foster City, Calif.).

The sequence of the leader was obtained from the subgenomic mRNA coding for the N gene and from the 5' terminus of genomic RNA. The 5' rapid amplification of cDNA ends (RACE) technique (Harcourt et al., *Virology* 271:334-49, 2000) was used with reverse primers specific for the N gene or for the 5' untranslated region. RACE products were either sequenced directly or were cloned into a plasmid vector before sequencing. A primer that was specific for the leader of SARS-CoV was used to amplify the region between the 5'-terminus of the genome and known sequences in the rep gene. The 3'-terminus of the genome was amplified for sequencing by use of an oligo-(dT) primer and primers specific for the N gene.

Once the complete SARS-CoV genomic sequence had been determined, it was confirmed by sequencing a series of independently amplified RT-PCR products spanning the entire genome. Positive- and negative-sense sequencing primers, at intervals of approximately 300 nt, were used to generate a confirmatory sequence with an average redundancy of 9.1. The confirmatory sequence was identical to the original sequence. The genomic sequence (SEQ ID NO: 1) was published in the GenBank sequence database (Accession No. AY278741) on Apr. 21, 2003.

Sequence Analysis

Predicted amino acid sequences were compared with those from reference viruses representing each species for which complete genomic sequence information was available: group 1 representatives included human coronavirus 229E (GenBank Accession No. AF304460), porcine epidemic diarrhea virus (GenBank Accession No. AF353511), and transmissible gastroenteritis virus (GenBank Accession No. AF271965); group 2 representatives included bovine coronavirus (GenBank Accession No. AF220295) and mouse hepatitis virus (GenBank Accession No. AF201929); group 3

was represented by infectious bronchitis virus (GenBank Accession No. M95169). Sequences for representative strains of other coronavirus species for which partial sequence information was available were included for some of the structural protein comparisons: group 1 representative strains included canine coronavirus (GenBank Accession No. D13096), feline coronavirus (GenBank Accession No. AY204704), and porcine respiratory coronavirus (GenBank Accession No. Z24675); and group 2 representatives included three strains of human coronavirus OC43 (GenBank Accession Nos. M76373, L14643 and M93390), porcine hemagglutinating encephalomyelitis virus (GenBank Accession No. AY078417), and rat coronavirus (GenBank Accession No. AF207551).

Partial nucleotide sequences of the polymerase gene were aligned with published coronavirus sequences, using CLUSTAL W for Unix (version 1.7; Thompson et al., *Nucleic Acids Res.* 22:4673-80, 1994). Phylogenetic trees were computed by maximum parsimony, distance, and maximum likelihood-based criteria analysis with PAUP (version 4.0.d10; Swofford ed., *Phylogenetic Analysis using Parsimony and other Methods*, Sinauer Associates, Sunderland, Mass.). When compared with other human and animal coronaviruses, the nucleotide and deduced amino acid sequence from this region had similarity scores ranging from 0.56 to 0.63 and from 0.57 to 0.74, respectively. The highest sequence similarity was obtained with group II coronaviruses. The maximum-parsimony tree obtained from the nucleotide-sequence alignment is shown in FIG. **3**. Bootstrap analyses of the internal nodes at the internal branches of the tree provided strong evidence that the SARS-CoV is genetically distinct from other known coronaviruses.

Microarray analyses (using a long oligonucleotide DNA microarray with array elements derived from highly conserved regions within viral families) of samples from infected and uninfected cell cultures gave a positive signal for a group of eight oligonucleotides derived from two virus families: Coronaviridae and Astroviridae (Wang et al., PNAS 99:15687-92, 2002). All of the astroviruses and two of the coronavirus oligonucleotides share a consensus sequence motif that maps to the extreme 3'-end of astroviruses and two members of the coronavirus family: avian infectious bronchitis and turkey coronavirus (Jonassen et al., *J. Gen. Virol.* 79:715-8, 1998). Results were consistent with the identity of the isolate as a coronavirus.

Additional sequence alignments and neighbor-joining trees were generated by using ClustalX (Thompson et al., *Nucleic Acids Res.* 25:4876-82, 1997), version 1.83, with the Gonnet protein comparison matrix. The resulting trees were adjusted for final output by using treetool version 2.0.1. Uncorrected pairwise distances were calculated from the aligned sequences by using the Distances program from the Wisconsin Sequence Analysis Package, version 10.2 (Accelrys, Burlington, Mass.). Distances were converted to percent identity by subtracting from 100. The amino acid sequences for three well-defined enzymatic proteins encoded by the rep gene and the four major structural proteins of SARS-CoV were compared with those from representative viruses for each of the species of coronavirus for which complete genomic sequence information was available (FIG. **4**, Table 2). The topologies of the resulting phylograms are remarkably similar (FIG. **4**). For each protein analyzed, the species formed monophyletic clusters consistent with the established taxonomic groups. In all cases, SARS-CoV sequences segregated into a fourth, well-resolved branch. These clusters were supported by bootstrap values above 90% (1000 replicates). Consistent with pairwise comparisons between the previously characterized coronavirus species (Table 2), there was greater sequence conservation in the enzymatic proteins ($3CL^{pro}$, polymerase (POL), and helicase (HEL)) than among the structural proteins (S, E, M, and N). These results indicate that SARS-CoV is not closely related to any of the previously characterized coronaviruses and forms a distinct group within the genus *Coronavirus*.

TABLE 2

Pairwise amino acid identities of coronavirus proteins.

Group	Virus	3CLPRO	POL	HEL	S	E	M	N
		Pairwise Amino Acid Identity (Percent)						
G1	HCoV-229E	40.1	58.8	59.7	23.9	22.7	28.8	23.0
	PEDV	44.4	59.5	61.7	21.7	17.6	31.8	22.6
	TGEV	44.0	59.4	61.2	20.6	22.4	30.0	25.6
G2	BCoV	48.8	66.3	68.3	27.1	20.0	39.7	31.9
	MHV	49.2	66.5	67.3	26.5	21.1	39.0	33.0
G3	IBV	41.3	62.5	58.6	21.8	18.4	27.2	24.0
		Predicted Protein Length (aa)						
	SARS-CoV	306	932	601	1255	76	221	422
	CoV Range	302-307	923-940	506-600	1173-1452	76-108	225-262	377-454

Example 2

Detection of SARS-CoV in a Subject

This example demonstrates the detection of SARS-CoV in patient specimens using SARS-CoV-specific primers.

The SARS-specific primers Cor-p-F2 (SEQ ID NO: 13), Cor-p-F3 (SEQ ID NO: 14) and Cor-p-R1 (SEQ ID NO: 15) were used to test patient specimens for SARS. One primer for each set was 5'-end-labeled with 6-FAM to facilitate GeneScan analysis. One-step amplification reactions were performed with the Access RT-PCR System (Promega, Madison, Wis.) as described by Falsey et al., *J. Infect. Dis.* 87:785-90, 2003. Positive and negative RT-PCR controls, containing standardized viral RNA extracts, and nuclease-free water were included in each run. Amplified 6-FAM-labeled products were analyzed by capillary electrophoresis on an ABI 3100 Prism Genetic Analyzer with GeneScan software (version 3.1.2; Applied Biosystems; Foster City, Calif.). Specimens were considered positive for SARS-CoV if the amplification products were within one nucleotide of the expected product size (368 nucleotides for Cor-p-F2 or Cor-p-R1 and

348 nucleotides for Cor-p-F3 or Cor-p-R1) for both specific primer sets, as confirmed by a second PCR reaction from another aliquot of RNA extract in a separate laboratory. Where DNA yield was sufficient, the amplified products were also sequenced. Additionally, as described above, microarray-based detection of SARS-CoV in patient specimens was carried out (Wang et al., *PNAS* 99:15687-92, 2002 and Bohlander et al., *Genomics* 13:1322-24, 1992).

Example 3

Immunohistochemical and Histopathological Analysis, and Electron-Microscopical Analysis of Bronchoalveolar Lavage Fluid

This example illustrates immunohistochemical, histopathological and electron-microscopical analysis of Vero E6 cells infected with the SARS-CoV and tissue samples from SARS patients.

Formalin-fixed, paraffin-embedded Vero E6 cells infected with the SARS-CoV and tissues obtained from patients with SARS were stained with hematoxylin and eosin and various immunohistochemical stains. Immunohistochemical assays were based on a method described previously for hantavirus (Zaki et al., *Amer. J. Pathol.* 146:552-79, 1995). Briefly, 4-μm sections were deparaffinized, rehydrated, and digested in Proteinase K for 15 minutes. Slides were then incubated for 60 minutes at room temperature with monoclonal antibodies, polyclonal antiserum or ascitic fluids derived from animal species with reactivities to various known coronaviruses, and with a convalescent-phase serum specimen from a patient with SARS.

Optimal dilutions of the primary antibodies were determined by titration experiments with coronavirus-infected cells from patients with SARS and with noninfected cells or, when available, with concentrations recommended by the manufacturers. After sequential application of the appropriate biotinylated link antibody, avidin-alkaline phosphatase complex, and naphthol-fast red substrate, sections were counterstained in Mayer's hematoxylin and mounted with aqueous mounting medium. The following antibody and tissue controls were used: serum specimens from noninfected animals, various coronavirus-infected cell cultures and animal tissues, noninfected cell cultures, and normal human and animal tissues. Tissues from patients were also tested by immunohistochemical assays for various other viral and bacterial pulmonary pathogens. In addition, a BAL specimen was available from one patient for thin-section electron-microscopical evaluation.

Lung tissues were obtained from the autopsy of three patients and by open lung biopsy of one patient, 14-19 days following onset of SARS symptoms. Confirmatory laboratory evidence of infection with coronavirus was available for two patients (patients 6 and 17) and included PCR amplification of coronavirus nucleic acids from tissues, viral isolation from BAL fluid or detection of serum antibodies reactive with coronavirus (Table 1). For two patients, no samples were available for molecular, cell culture, or serological analysis; however, both patients met the CDC definition for probable SARS cases and had strong epidemiologic links with laboratory-confirmed SARS cases. Histopathologic evaluation of lung tissues of the four patients showed diffuse alveolar damage at various levels of progression and severity. Changes included hyaline membrane formation, interstitial mononuclear inflammatory infiltrates, and desquamation of pneumocytes in alveolar spaces (FIG. **5**A). Other findings identified in some patients included focal intraalveolar hemorrhage, necrotic inflammatory debris in small airways, and organizing pneumonia. Multinucleated syncytial cells were identified in the intraalveolar spaces of two patients who died 14 and 17 days, respectively, after onset of illness. These cells contained abundant vacuolated cytoplasm with cleaved and convoluted nuclei. No obvious intranuclear or intracytoplasmic viral inclusions were identified (FIG. **5**B), and electron-microscopical examination of a limited number of these syncytial cells revealed no coronavirus particles. No definitive immunostaining was identified in tissues from SARS patients with the use of a battery of immunohistochemical stains reactive with coronaviruses from antigenic groups I, II, and III. In addition, no staining of patient tissues was identified with the use of immunohistochemical stains for influenzaviruses A and B, adenoviruses, Hendra and Nipah viruses, human metapneumovirus, respiratory syncytial virus, measles virus, *Mycoplasma pneumoniae*, and *Chlamydia pneumoniae*.

Evaluation of Vero E6 cells infected with coronavirus isolated from a patient with SARS revealed viral CPE that included occasional multinucleated syncytial cells but no obvious viral inclusions (FIG. **5**C). Immunohistochemical assays with various antibodies reactive with coronaviruses from antigenic group I, including HCoV-229E, FIPV and TGEV, and with an immune serum specimen from a patient with SARS, demonstrated strong cytoplasmic and membranous staining of infected cells (FIG. **5**C and Table 3); however, cross-reactivity with the same immune human serum sample and FIPV antigen was not observed. No staining was identified with any of several monoclonal or polyclonal antibodies reactive with coronaviruses in antigenic group II (HCoV-OC43, BCoV and MHV) or group III (TCoV and IBV-Avian). Electron microscopical examination of a BAL fluid from one patient revealed many coronavirus-infected cells (FIGS. **6**A-B).

TABLE 3

Immunohistochemical reactivities of various polyclonal group I anti-coronavirus reference antiserum samples with a coronavirus isolated from a patient with SARS and with selected antigenic group I coronaviruses.

	Immunohistochemical reactivity of antiserum with coronavirus-infected culture cells		
Antiserum	SARS-CoV (Vero E6)	HCoV-229E (mouse 3T3-hAPN)	FIPV-1 (BHK-fAPN)
Convalescent-phase SARS (patient 3)	+	+	–
Guinea pig anti-HCoV-229E	+	+	–
Rabbit anti HCoV-229E	+	+	+
Feline anti-FIPV-1	+	+	+
Porcine anti-TGEV	+	–	+

Example 4

SARS-CoV Serologic Analysis

This example illustrates representative methods of performing serological analysis of SARS-CoV.

Spot slides were prepared by applying 15 μl of the suspension of gamma-irradiated mixed infected and noninfected cells onto 12-well Teflon-coated slides. Slides were allowed

to air dry before being fixed in acetone. Slides were then stored at −70° C. until used for IFA tests (Wulff and Lange, *Bull. WHO* 52:429-36, 1975). An ELISA antigen was prepared by detergent extraction and subsequent gamma irradiation of infected Vero E6 cells (Ksiazek et al., *J. Infect. Dis.* 179 (suppl. 1):S191-8, 1999). The optimal dilution (1:1000) for the use of this antigen was determined by checkerboard titration against SARS patient serum from the convalescent phase; a control antigen, similarly prepared from uninfected Vero E6 cells, was used to control for specific reactivity of tested sera. The conjugates used were goat antihuman IgG, IgA, and IgM conjugated to fluorescein isothiocyanate and horseradish peroxidase (Kirkegaard and Perry, Gaithersburg, Md.), for the IFA test and ELISA, respectively. Specificity and cross-reactivity of a variety of serum samples to the newly identified virus were evaluated by using the tests described herein. For this evaluation, serum from SARS patients in Singapore, Bangkok and Hong Kong was used, along with serum from healthy blood donors from the CDC serum bank and from persons infected with known human coronavirus (human coronaviruses OC43 and 229E) (samples provided by E. Walsh and A. Falsey, University of Rochester School of Medicine and Dentistry, Rochester, N.Y.).

Spot slides with infected cells reacted with serum from patients with probable SARS in the convalescent phase (FIG. **1**B). Screening of a panel of serum from patients with suspected SARS from Hong Kong, Bangkok, Singapore as well as the United States showed a high level of specific reaction with infected cells, and conversion from negative to positive reactivity or diagnostic rises in the IFA test by a factor of four. Similarly, tests of these same serum samples with the ELISA antigen showed high specific signal in the convalescent-phase samples and conversion from negative to positive antibody reactivity or diagnostic increases in titer (Table 4).

TABLE 4

Results of serological testing with both IFA assay and ELISA in SARS patients tested against the newly isolated human coronavirus.

Source	Serum No.	Days After Onset	ELISA Titer*	IFA Titer*
Hong Kong	1.1	4	<100	<25
Hong Kong	1.2	13	≧6400	1600
Hong Kong	2.1	11	400	100
Hong Kong	2.2	16	1600	200
Hong Kong	3.1	7	<100	<25
Hong Kong	3.2	17	≧6400	800
Hong Kong	4.1	8	<100	<25
Hong Kong	4.2	13	1600	50
Hong Kong	5.1	10	100	<25
Hong Kong	5.2	17	≧6400	1600
Hong Kong	6.1	12	1600	200
Hong Kong	6.2	20	≧6400	6400
Hong Kong	7.1	17	400	50
Hong Kong	7.2	24	≧6400	3200
Hong Kong	8.1	3	<100	<25
Hong Kong	8.2	15	≧6400	200
Hong Kong (Hanoi)	9.1	5	<100	<25
Hong Kong	9.2	11	≧6400	1600
Bangkok	1.1	2	<100	<25
Bangkok	1.2	4	<100	<25
Bangkok	1.3	7	<100	<25
Bangkok	1.4	15	1600	200
United States	1.1	2	<100	<25
United States	1.2	6	400	50
United States	1.3	13	≧6400	800
Singapore	1.1	2	100	<25
Singapore	1.2	11	≧6400	800
Singapore	2.1	6	100	<25
Singapore	2.2	25	≧6400	400

TABLE 4-continued

Results of serological testing with both IFA assay and ELISA in SARS patients tested against the newly isolated human coronavirus.

Source	Serum No.	Days After Onset	ELISA Titer*	IFA Titer*
Singapore	3.1	6	100	<25
Singapore	3.2	14	≧6400	400
Singapore	4.1	5	100	<25
Singapore	4.2	16	1600	400

*Reciprocal of dilution

Information from the limited numbers of samples tested suggests that antibody is first detectable by IFA assay and ELISA between one and two weeks after the onset of symptoms in the patient. IFA testing and ELISA of a panel of 384 randomly selected serum samples (from U.S. blood donors) were negative for antibodies to the new coronavirus, with the exception of one specimen that had minimal reactivity on ELISA. A panel of paired human serum samples with diagnostic increases (by a factor of four or more) in antibody (with very high titers to the homologous viral antigen in the convalescent-phase serum) to the two known human coronaviruses, OC43 (13 pairs) and 229E (14 pairs), showed no reactivity in either acute- or convalescent-phase serum with the newly isolated coronavirus by either the IFA test or the ELISA.

Example 5

Poly(A)$^+$RNA Isolation and Northern Hybridization

This example illustrates a representative method of Northern hybrididization to detect SARS-CoV messages in Vero E6 cells.

Total RNA from infected or uninfected Vero E6 cells was isolated with Trizol reagent (Invitrogen Life Technologies, Carlsbad, Calif.) used according to the manufacturer's recommendations. Poly(A)$^+$RNA was isolated from total RNA by use of the Oligotex Direct mRNA Kit (Qiagen, Inc., Santa Clarita, Calif.), following the instructions for the batch protocol, followed by ethanol precipitation. RNA isolated from 1 cm^2 of cells was separated by electrophoresis on a 0.9% agarose gel containing 3.7% formaldehyde, followed by partial alkaline hydrolysis (Ausubel et al. eds. *Current Protocols in Molecular Biology*, vol. 1, John Wiley & Sons, Inc., NY, N.Y., Ch. 4.9, 1996). RNA was transferred to a nylon membrane (Roche Molecular Biochemicals, Indianapolis, Ind.) by vacuum blotting (Bio-Rad, Hercules, Calif.) and fixed by UV cross-linking. The DNA template for probe synthesis was generated by RT-PCR amplification of SARS-CoV nt 29,083 to 29,608 (SEQ ID NO: 1), by using a reverse primer containing a T7 RNA polymerase promoter to facilitate generation of a negative-sense riboprobe. In vitro transcription of the digoxigenin-labeled riboprobe, hybridization, and detection of the bands were carried out with the digoxigenin system by using manufacturer's recommended procedures (Roche Molecular Biochemicals, Indianapolis, Ind.). Signals were visualized by chemiluminescence and detected with x-ray film.

Example 6

SARS-CoV Genome Organization

This example illustrates the genomic organization of the SARS-CoV genome, including the location of SARS-CoV ORFs.

The genome of SARS-CoV is a 29,727-nucleotide, polyadenylated RNA, and 41% of the residues are G or C (range for published coronavirus complete genome sequences, 37% to 42%). The genomic organization is typical of coronaviruses, having the characteristic gene order [5'-replicase (rep), spike (S), envelope (E), membrane (M), nucleocapsid (N)-3'] and short untranslated regions at both termini (FIG. 7A, Table 5). The SARS-CoV rep gene, which comprises approximately two-thirds of the genome, encodes two polyproteins (encoded by ORF1a and ORF1b) that undergo co-translational proteolytic processing. There are four ORFs downstream of rep that encode the structural proteins, S, E, M, and N, which are common to all known coronaviruses. The hemagglutinin-esterase gene, which is present between ORF1b and S in group 2 and some group 3 coronaviruses (Lai and Holmes, in *Fields Virology*, eds. Knipe and Howley, Lippincott Williams and Wilkins, New York, 4th edition, 2001, Ch. 35), was not found in SARS-CoV.

Coronaviruses also encode a number of non-structural proteins that are located between S and E, between M and N, or downstream of N. These non-structural proteins, which vary widely among the different coronavirus species, are of unknown function and are dispensable for virus replication (Lai and Holmes, in *Fields Virology*, eds. Knipe and Howley, Lippincott Williams and Wilkins, New York, 4th edition, 2001, Ch. 35). The genome of SARS-CoV contains ORFs for five non-structural proteins of greater than 50 amino acids (FIG. 7B, Table 5). Two overlapping ORFs encoding proteins of 274 and 154 amino acids (termed X1 (SEQ ID NO: 5) and X2 (SEQ ID NO: 6), respectively) are located between S (SEQ ID NO: 4) and E (SEQ ID NO: 7). Three additional non-structural genes, X3 (SEQ ID NO: 9), X4 (SEQ ID NO: 10), and X5 (SEQ ID NO: 11) (encoding proteins of 63, 122, and 84 amino acids, respectively), are located between M (SEQ ID NO: 8) and N (SEQ ID NO: 12). In addition to the five ORFs encoding the non-structural proteins described above, there are also two smaller ORFs between M and N, encoding proteins of less than 50 amino acids. Searches of the GenBank database (BLAST and FastA) indicated that there is no significant sequence similarity between these non-structural proteins of SARS-CoV and any other proteins.

The coronavirus rep gene products are translated from genomic RNA, but the remaining viral proteins are translated from subgenomic mRNAs that form a 3'-coterminal nested set, each with a 5'-end derived from the genomic 5'-leader sequence. The coronavirus subgenomic mRNAs are synthesized through a discontinuous transcription process, the mechanism of which has not been unequivocally established (Lai and Holmes, in *Fields Virology*, eds. Knipe and Howley, Lippincott Williams and Wilkins, New York, 4th edition, 2001, Ch. 35; Sawicki and Sawicki, *Adv. Exp. Med. Biol.* 440:215-19, 1998). The SARS-CoV leader sequence was mapped by comparing the sequence of 5'-RACE products synthesized from the N gene mRNA with those synthesized from genomic RNA. A sequence, AAACGAAC (nucleotides 65-72 of SEQ ID NO: 1), was identified immediately upstream of the site where the N gene mRNA and genomic sequences diverged. This sequence was also present upstream of ORF1a and immediately upstream of five other ORFs (Table 5), suggesting that it functions as the conserved core of the transcriptional regulatory sequence (TRS).

In addition to the site at the 5'-terminus of the genome, the TRS conserved core sequence appears six times in the remainder of the genome. The positions of the TRS in the genome of SARS-CoV predict that subgenomic mRNAs of 8.3, 4.5, 3.4, 2.5, 2.0, and 1.7 kb, not including the poly(A) tail, should be produced (FIGS. 7A-B, Table 5). At least five subgenomic mRNAs were detected by Northern hybridization of RNA from SARS-CoV-infected cells, using a probe derived from the 3'-untranslated region (FIG. 7C). The calculated sizes of the five predominant bands correspond to the sizes of five of the predicted subgenomic mRNAs of SARS-CoV; the possibility that other, low-abundance mRNAs are present cannot be excluded. By analogy with other coronaviruses (Lai and Holmes, in *Fields Virology*, eds. Knipe and Howley, Lippincott Williams and Wilkins, New York, 4th edition, 2001, Ch. 35), the 8.3-kb and 1.7-kb subgenomic mRNAs are monocistronic, directing translation of S and N, respectively, whereas multiple proteins are translated from the 4.5-kb (X1, X2, and E), 3.4-kb (M and X3), and 2.5-kb (X4 and X5) mRNAs. A consensus TRS is not found directly upstream of the ORF encoding the predicted E protein, and a monocistronic mRNA that would be predicted to code for E could not be clearly identified by Northern blot analysis. It is possible that the 3.6-kb band contains more than one mRNA species or that the monocistronic mRNA for E is a low-abundance message.

TABLE 5

Locations of SARS-CoV ORFs and sizes of proteins and mRNAs

	Genome Location			Predicted Size	
ORF	TRS[a]	ORF Start	ORF End	Protein (aa)	mRNA (nt)[b]
1a	72	265	13,398	4,378	29,727
1b		13,398	21,482	2,695	
S	21,491	21,492	25,256	1,255	8,308[c]
X1	25,265	25,268	26,089	274	4,534[c]
X2		25,689	26,150	154	
E		26,117	26,344	76	
M	26,353	26,398	27,060	221	3,446[c]
X3		27,074	27,262	63	
X4	27,272	27,273	27,638	122	2,527[c]
X5	27,778	27,864	28,115	84	2,021[d]
N	28,111	28,120	29,385	422	1,688[c]

[a]The location is the 3'-most nucleotide in the consensus TRS, AAACGAAC.
[b]Not including poly(A). Predicted size is based on the position of the conserved TRS.
[c]Corresponding mRNA detected by Northern blot analysis (FIG. 7C)
[d]No mRNA corresponding to utilization of this consensus TRS was detected by Northern blot analysis (FIG. 7C)

Example 7

Real-Time RT-PCR Assay for SARS-CoV Detection

This example demonstrates the use of SARS-CoV-specific primers and probes in a real-time RT-PCR assay to detect SARS-CoV in patient specimens.

A variant of the real-time format, based on TaqMan probe hydrolysis technology (Applied Biosystems, Foster City, Calif.), was used to analyze a total of 340 clinical specimens collected from 246 persons with confirmed or suspected SARS-CoV infection. Specimens included oro- and nasopharyngeal swabs (dry and in viral transport media), sputa, nasal aspirates and washes, BAL, and lung tissue specimens collected at autopsy.

Nucleic Acid Extraction

SARS-CoV nucleic acids were recovered from clinical specimens using the automated NucliSens extraction system (bioMérieux, Durham, N.C.). Following manufacturer's instructions, specimens received in NucliSens lysis buffer were incubated at 37° C. for 30 min with intermittent mixing, and 50 μL of silica suspension, provided in the extraction kit, was added and mixed. The contents of the tube were then transferred to a nucleic acid extraction cartridge and processed on an extractor workstation. Approximately 40-50 μL of total nucleic acid eluate was recovered into nuclease-free vials and either tested immediately or stored at −70° C.

Primers and Probes

Multiple primer and probe sets were designed from the SARS-CoV polymerase 1b (nucleic acid 13,398 to 21,482 of SEQ ID NO: 1) and nucleocapsid gene (nucleic acid 28,120 to 29,385 of SEQ ID NO: 1) sequences by using Primer Express software version 1.5 or 2.0.0 (Applied Biosystems, Foster City, Calif.) with the following default settings: primer melting temperature (T_M) set at 60° C.; probe T_M set at 10° C. greater than the primers at approximately 70° C.; and no guanidine residues permitted at the 5' probe termini. All primers and probes were synthesized by standard phosphoramidite chemistry techniques. TaqMan probes were labeled at the 5'-end with the reporter 6-FAM and at the 3'-end with the quencher Blackhole Quencher 1 (Biosearch Technologies, Inc., Novato, Calif.). Optimal primer and probe concentrations were determined by cross-titration of serial twofold dilutions of each primer against a constant amount of purified SARS-CoV RNA. Primer and probe concentrations that gave the highest amplification efficiencies were selected for further study (Table 6).

TABLE 6

Primers and probes used for real-time RT-PCR assays[a]

Assay ID	Primer/probe	Sequence	Genomic Region
		Primary diagnostic assay	
SARS1	F	CATGTGTGGCGGCTCACTATAT (SEQ ID NO: 16)	RNA Pol
	R	GACACTATTAGCATAAGCAGTTGTAGCA (SEQ ID NO: 17)	
	P	TTAAACCAGGTGGAACATCATCCGGTG (SEQ ID NO: 18)	
SARS2	F	GGAGCCTTGAATACACCCAAAG (SEQ ID NO: 19)	Nucleocapsid
	R	GCACGGTGGCAGCATTG (SEQ ID NO: 20)	
	P	CCACATTGGCACCCGCAATCC (SEQ ID NO: 21)	
SARS3	F	CAAACATTGGCCGCAAATT (SEQ ID NO: 22)	Nucleocapsid
	R	CAATGCGTGACATTCCAAAGA (SEQ ID NO: 23)	
	P	CACAATTTGCTCCAAGTGCCTCTGCA (SEQ ID NO: 24)	
		To confirm positive results	
N3	F	GAAGTACCATCTGGGGCTGAG (SEQ ID NO: 25)	Nucleocapsid
	R	CCGAAGAGCTACCCGACG (SEQ ID NO: 26)	
	P	CTCTTTCATTTTGCCGTCACCACCAC (SEQ ID NO: 27)	
3'-NTR	F	AGCTCTCCCTAGCATTATTCACTG (SEQ ID NO: 28)	3'-NTR
	R	CACCACATTTTCATCGAGGC (SEQ ID NO: 29)	
	P	TACCCTCGATCGTACTCCGCGT (SEQ ID NO: 30)	
M	F	TGTAGGCACTGATTCAGGTTTTG (SEQ ID NO: 31)	M protein
	R	CGGCGTGGTCTGTATTTAATTTA (SEQ ID NO: 32)	
	P	CTGCATACAACCGCTACCGTATTGGAA (SEQ ID NO: 33)	

[a]RT-PCR, reverse transcription-polymerase chain reaction; F, forward primer; R, reverse primer; P, probe; NTR, nontranslated region.

Real-Time RT-PCR Assay

The real-time RT-PCR assay was performed by using the Real-Time One-Step RT-PCR Master Mix (Applied Biosystems, Foster City, Calif.). Each 25-μL reaction mixture contained 12.5 μL of 2× Master Mix, 0.625 μL of the 40× MultiScribe and RNase Inhibitor mix, 0.25 μL of 10 μM probe, 0.25 μL each of 50 μM forward and reverse primers, 6.125 μL of nuclease-free water, and 5 μL of nucleic acid extract. Amplification was carried out in 96-well plates on an iCycler iQ Real-Time Detection System (Bio-Rad, Hercules, Calif.). Thermocycling conditions consisted of 30 minutes at 48° C. for reverse transcription, 10 minutes at 95° C. for activation of the AmpliTaq Gold DNA polymerase, and 45 cycles of 15 seconds at 95° C. and 1 minute at 60° C. Each run included one SARS-CoV genomic template control and at least two no-template controls for the extraction (to check for contamination during sample processing) and one no-template control for the PCR-amplification step. As a control for PCR inhibitors, and to monitor nucleic acid extraction efficiency, each sample was tested by real-time RT-PCR for the presence of the human ribonuclease (RNase) P gene (GenBank Accession No. NM_006413) by using the following primers and probe: forward primer 5'-AGATTTGGACCTGCGAGCG-3' (SEQ ID NO: 36); reverse primer 5'-GAGCGGCTGTCTCCACAAGT-3' (SEQ ID NO: 37); probe 5'-TTCTGACC TGAAGGCTCTGCGCG-3' (SEQ ID NO: 38). The assay reaction was performed identically to that described above except that primer concentrations used were 30 μM each. Fluorescence measurements were taken and the threshold cycle (C_T) value for each sample was calculated by determining the point at which fluorescence exceeded a threshold limit set at the mean plus 10 standard deviations above the baseline. A test result was considered positive if two or more of the SARS genomic targets showed positive results ($C_T \leqq 45$ cycles) and all positive and negative control reactions gave expected values.

While this disclosure has been described with an emphasis upon preferred embodiments, it will be obvious to those of ordinary skill in the art that variations and equivalents of the preferred embodiments may be used and it is intended that the disclosure may be practiced otherwise than as specifically described herein. Accordingly, this disclosure includes all modifications encompassed within the spirit and scope of the disclosure as defined by the claims below.

SEQUENCE LISTING

<160> NUMBER OF SEQ ID NOS: 38

<210> SEQ ID NO 1
<211> LENGTH: 29727
<212> TYPE: DNA
<213> ORGANISM: Coronavirus
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (265)..(13398)
<223> OTHER INFORMATION: ORF 1a
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (13398)..(21482)
<223> OTHER INFORMATION: ORF 1b
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (21492)..(25256)
<223> OTHER INFORMATION: ORF S
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (25268)..(26089)
<223> OTHER INFORMATION: ORF X1
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (25689)..(26150)
<223> OTHER INFORMATION: ORF X2
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (26117)..(26344)
<223> OTHER INFORMATION: ORF E
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (26398)..(27060)
<223> OTHER INFORMATION: ORF M
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (27074)..(27262)
<223> OTHER INFORMATION: ORF X3
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (27273)..(27638)
<223> OTHER INFORMATION: ORF X4
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (27864)..(28115)
<223> OTHER INFORMATION: ORF X5
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (28120)..(29385)
<223> OTHER INFORMATION: ORF N

<400> SEQUENCE: 1

ttattaggtt tttacctacc caggaaaagc caaccaacct cgatctcttg tagatctgtt 60

ctctaaacga actttaaaat ctgtgtagct gtcgctcggc tgcatgccta gtgcacctac 120

gcagtataaa caataataaa ttttactgtc gttgacaaga aacgagtaac tcgtccctct 180

tctgcagact gcttacggtt tcgtccgtgt tgcagtcgat catcagcata cctaggtttc 240

gtccgggtgt gaccgaaagg taagatggag agccttgttc ttggtgtcaa cgagaaaaca 300

cacgtccaac tcagtttgcc tgtccttcag gttagagacg tgctagtgcg tggcttcggg 360

gactctgtgg aagaggccct atcggaggca cgtgaacacc tcaaaaatgg cacttgtggt 420

ctagtagagc tggaaaaagg cgtactgccc cagcttgaac agccctatgt gttcattaaa 480

cgttctgatg ccttaagcac caatcacggc cacaaggtcg ttgagctggt tgcagaaatg 540

gacggcattc agtacggtcg tagcggtata acactgggag tactcgtgcc acatgtgggc 600

gaaaccccaa ttgcataccg caatgttctt cttcgtaaga acggtaataa gggagccggt 660

ggtcatagct atggcatcga tctaaagtct tatgacttag gtgacgagct tggcactgat 720

-continued

cccattgaag attatgaaca aaactggaac actaagcatg gcagtggtgc actccgtgaa 780

ctcactcgtg agctcaatgg aggtgcagtc actcgctatg tcgacaacaa tttctgtggc 840

ccagatgggt accctcttga ttgcatcaaa gattttctcg cacgcgcggg caagtcaatg 900

tgcactcttt ccgaacaact tgattacatc gagtcgaaga gaggtgtcta ctgctgccgt 960

gaccatgagc atgaaattgc ctggttcact gagcgctctg ataagagcta cgagcaccag 1020

acacccttcg aaattaagag tgccaagaaa tttgacactt tcaaagggga atgcccaaag 1080

tttgtgtttc ctcttaactc aaaagtcaaa gtcattcaac cacgtgttga aaagaaaaag 1140

actgagggtt tcatggggcg tatacgctct gtgtaccctg ttgcatctcc acaggagtgt 1200

aacaatatgc acttgtctac cttgatgaaa tgtaatcatt gcgatgaagt ttcatggcag 1260

acgtgcgact ttctgaaagc cacttgtgaa cattgtggca ctgaaaattt agttattgaa 1320

ggacctacta catgtgggta cctacctact aatgctgtag tgaaaatgcc atgtcctgcc 1380

tgtcaagacc cagagattgg acctgagcat agtgttgcag attatcacaa ccactcaaac 1440

attgaaactc gactccgcaa gggaggtagg actagatgtt ttggaggctg tgtgtttgcc 1500

tatgttggct gctataataa gcgtgcctac tgggttcctc gtgctagtgc tgatattggc 1560

tcaggccata ctggcattac tggtgacaat gtggagacct tgaatgagga tctccttgag 1620

atactgagtc gtgaacgtgt taacattaac attgttggcg attttcattt gaatgaagag 1680

gttgccatca ttttggcatc tttctctgct tctacaagtg cctttattga cactataaag 1740

agtcttgatt acaagtcttt caaaaccatt gttgagtcct gcggtaacta taaagttacc 1800

aagggaaagc ccgtaaaagg tgcttggaac attggacaac agagatcagt tttaacacca 1860

ctgtgtggtt ttccctcaca ggctgctggt gttatcagat caatttttgc gcgcacactt 1920

gatgcagcaa accactcaat tcctgatttg caaagagcag ctgtcaccat acttgatggt 1980

atttctgaac agtcattacg tcttgtcgac gccatggttt atacttcaga cctgctcacc 2040

aacagtgtca ttattatggc atatgtaact ggtggtcttg tacaacagac ttctcagtgg 2100

ttgtctaatc ttttgggcac tactgttgaa aaactcaggc ctatctttga atggattgag 2160

gcgaaactta gtgcaggagt tgaatttctc aaggatgctt gggagattct caaatttctc 2220

attacaggtg tttttgacat cgtcaagggt caaatacagg ttgcttcaga taacatcaag 2280

gattgtgtaa aatgcttcat tgatgttgtt aacaaggcac tcgaaatgtg cattgatcaa 2340

gtcactatcg ctggcgcaaa gttgcgatca ctcaacttag gtgaagtctt catcgctcaa 2400

agcaagggac tttaccgtca gtgtatacgt ggcaaggagc agctgcaact actcatgcct 2460

cttaaggcac caaaagaagt aacctttctt gaaggtgatt cacatgacac agtacttacc 2520

tctgaggagg ttgttctcaa gaacggtgaa ctcgaagcac tcgagacgcc cgttgatagc 2580

ttcacaaatg gagctatcgt tggcacacca gtctgtgtaa atggcctcat gctcttagag 2640

attaaggaca aagaacaata ctgcgcattg tctcctggtt tactggctac aaacaatgtc 2700

tttcgcttaa aaggggggtgc accaattaaa ggtgtaacct ttggagaaga tactgtttgg 2760

gaagttcaag gttacaagaa tgtgagaatc acatttgagc ttgatgaacg tgttgacaaa 2820

gtgcttaatg aaaagtgctc tgtctacact gttgaatccg gtaccgaagt tactgagttt 2880

gcatgtgttg tagcagaggc tgttgtgaag actttacaac cagtttctga tctccttacc 2940

aacatgggta ttgatcttga tgagtggagt gtagctacat tctacttatt tgatgatgct 3000

ggtgaagaaa acttttcatc acgtatgtat tgttcctttt accctccaga tgaggaagaa 3060

-continued

gaggacgatg cagagtgtga ggaagaagaa attgatgaaa cctgtgaaca tgagtacggt 3120

acagaggatg attatcaagg tctccctctg gaatttggtg cctcagctga aacagttcga 3180

gttgaggaag aagaagagga agactggctg gatgatacta ctgagcaatc agagattgag 3240

ccagaaccag aacctacacc tgaagaacca gttaatcagt ttactggtta tttaaaactt 3300

actgacaatg ttgccattaa atgtgttgac atcgttaagg aggcacaaag tgctaatcct 3360

atggtgattg taaatgctgc taacatacac ctgaaacatg gtggtggtgt agcaggtgca 3420

ctcaacaagg caaccaatgg tgccatgcaa aaggagagtg atgattacat taagctaaat 3480

ggccctctta cagtaggagg gtcttgtttg ctttctggac ataatcttgc taagaagtgt 3540

ctgcatgttg ttggacctaa cctaaatgca ggtgaggaca tccagcttct taaggcagca 3600

tatgaaaatt tcaattcaca ggacatctta cttgcaccat tgttgtcagc aggcatattt 3660

ggtgctaaac cacttcagtc tttacaagtg tgcgtgcaga cggttcgtac acaggtttat 3720

attgcagtca atgacaaagc tctttatgag caggttgtca tggattatct tgataacctg 3780

aagcctagag tggaagcacc taaacaagag gagccaccaa acacagaaga ttccaaaact 3840

gaggagaaat ctgtcgtaca gaagcctgtc gatgtgaagc caaaaattaa ggcctgcatt 3900

gatgaggtta ccacaacact ggaagaaact aagtttctta ccaataagtt actcttgttt 3960

gctgatatca atggtaagct ttaccatgat tctcagaaca tgcttagagg tgaagatatg 4020

tctttccttg agaaggatgc accttacatg gtaggtgatg ttatcactag tggtgatatc 4080

acttgtgttg taataccctc caaaaaggct ggtggcacta ctgagatgct ctcaagagct 4140

ttgaagaaag tgccagttga tgagtatata accacgtacc ctggacaagg atgtgctggt 4200

tatacacttg aggaagctaa gactgctctt aagaaatgca aatctgcatt ttatgtacta 4260

ccttcagaag cacctaatgc taaggaagag attctaggaa ctgtatcctg gaatttgaga 4320

gaaatgcttg ctcatgctga agagacaaga aaattaatgc ctatatgcat ggatgttaga 4380

gccataatgg caaccatcca acgtaagtat aaaggaatta aaattcaaga gggcatcgtt 4440

gactatggtg tccgattctt cttttatact agtaaagagc ctgtagcttc tattattacg 4500

aagctgaact ctctaaatga gccgcttgtc acaatgccaa ttggttatgt gacacatggt 4560

tttaatcttg aagaggctgc gcgctgtatg cgttctctta aagctcctgc cgtagtgtca 4620

gtatcatcac cagatgctgt tactacatat aatggatacc tcacttcgtc atcaaagaca 4680

tctgaggagc actttgtaga aacagtttct ttggctggct cttacagaga ttggtcctat 4740

tcaggacagc gtacagagtt aggtgttgaa tttcttaagc gtggtgacaa aattgtgtac 4800

cacactctgg agagccccgt cgagtttcat cttgacggtg aggttctttc acttgacaaa 4860

ctaaagagtc tcttatccct gcgggaggtt aagactataa aagtgttcac aactgtggac 4920

aacactaatc tccacacaca gcttgtggat atgtctatga catatggaca gcagtttggt 4980

ccaacatact tggatggtgc tgatgttaca aaaattaaac ctcatgtaaa tcatgagggt 5040

aagactttct ttgtactacc tagtgatgac acactacgta gtgaagcttt cgagtactac 5100

catactcttg atgagagttt tcttggtagg tacatgtctg ctttaaacca cacaaagaaa 5160

tggaaatttc ctcaagttgg tggtttaact tcaattaaat gggctgataa caattgttat 5220

ttgtctagtg ttttattagc acttcaacag cttgaagtca aattcaatgc accagcactt 5280

caagaggctt attatagagc ccgtgctggt gatgctgcta acttttgtgc actcatactc 5340

gcttacagta ataaaactgt tggcgagctt ggtgatgtca gagaaactat gacccatctt 5400

ctacagcatg ctaatttgga atctgcaaag cgagttctta atgtggtgtg taaacattgt 5460

-continued

ggtcagaaaa ctactacctt aacgggtgta gaagctgtga tgtatatggg tactctatct 5520

tatgataatc ttaagacagg tgtttccatt ccatgtgtgt gtggtcgtga tgctacacaa 5580

tatctagtac aacaagagtc ttcttttgtt atgatgtctg caccacctgc tgagtataaa 5640

ttacagcaag gtacattctt atgtgcgaat gagtacactg gtaactatca gtgtggtcat 5700

tacactcata taactgctaa ggagaccctc tatcgtattg acggagctca ccttacaaag 5760

atgtcagagt acaaaggacc agtgactgat gttttctaca aggaaacatc ttacactaca 5820

accatcaagc ctgtgtcgta taaactcgat ggagttactt acacagagat tgaaccaaaa 5880

ttggatgggt attataaaaa ggataatgct tactatacag agcagcctat agaccttgta 5940

ccaactcaac cattaccaaa tgcgagtttt gataatttca aactcacatg ttctaacaca 6000

aaatttgctg atgatttaaa tcaaatgaca ggcttcacaa agccagcttc acgagagcta 6060

tctgtcacat tcttcccaga cttgaatggc gatgtagtgg ctattgacta tagacactat 6120

tcagcgagtt tcaagaaagg tgctaaatta ctgcataagc caattgtttg gcacattaac 6180

caggctacaa ccaagacaac gttcaaacca aacacttggt gtttacgttg tctttggagt 6240

acaaagccag tagatacttc aaattcattt gaagttctgg cagtagaaga cacacaagga 6300

atggacaatc ttgcttgtga aagtcaacaa cccacctctg aagaagtagt ggaaaatcct 6360

accatacaga aggaagtcat agagtgtgac gtgaaaacta ccgaagttgt aggcaatgtc 6420

atacttaaac catcagatga aggtgttaaa gtaacacaag agttaggtca tgaggatctt 6480

atggctgctt atgtggaaaa cacaagcatt accattaaga aacctaatga gctttcacta 6540

gccttaggtt taaaaacaat tgccactcat ggtattgctg caattaatag tgttccttgg 6600

agtaaaattt tggcttatgt caaaccattc ttaggacaag cagcaattac aacatcaaat 6660

tgcgctaaga gattagcaca acgtgtgttt aacaattata tgccttatgt gtttacatta 6720

ttgttccaat tgtgtacttt tactaaaagt accaattcta gaattagagc ttcactacct 6780

acaactattg ctaaaaatag tgttaagagt gttgctaaat tatgtttgga tgccggcatt 6840

aattatgtga agtcacccaa attttctaaa ttgttcacaa tcgctatgtg gctattgttg 6900

ttaagtattt gcttaggttc tctaatctgt gtaactgctg cttttggtgt actcttatct 6960

aattttggtg ctccttctta ttgtaatggc gttagagaat tgtatcttaa ttcgtctaac 7020

gttactacta tggatttctg tgaaggttct tttccttgca gcatttgttt aagtggatta 7080

gactcccttg attcttatcc agctcttgaa accattcagg tgacgatttc atcgtacaag 7140

ctagacttga caattttagg tctggccgct gagtgggttt tggcatatat gttgttcaca 7200

aaattctttt atttattagg tctttcagct ataatgcagg tgttctttgg ctattttgct 7260

agtcatttca tcagcaattc ttggctcatg tggtttatca ttagtattgt acaaatggca 7320

cccgtttctg caatggttag gatgtacatc ttctttgctt ctttctacta catatggaag 7380

agctatgttc atatcatgga tggttgcacc tcttcgactt gcatgatgtg ctataagcgc 7440

aatcgtgcca cacgcgttga gtgtacaact attgttaatg gcatgaagag atctttctat 7500

gtctatgcaa atggaggccg tggcttctgc aagactcaca attggaattg tctcaattgt 7560

gacacatttt gcactggtag tacattcatt agtgatgaag ttgctcgtga tttgtcactc 7620

cagtttaaaa gaccaatcaa ccctactgac cagtcatcgt atattgttga tagtgttgct 7680

gtgaaaaatg gcgcgcttca cctctacttt gacaaggctg gtcaaaagac ctatgagaga 7740

catccgctct cccattttgt caatttagac aatttgagag ctaacaacac taaaggttca 7800

-continued

ctgcctatta atgtcatagt ttttgatggc aagtccaaat gcgacgagtc tgcttctaag 7860

tctgcttctg tgtactacag tcagctgatg tgccaaccta ttctgttgct tgaccaagtt 7920

cttgtatcag acgttggaga tagtactgaa gtttccgtta agatgtttga tgcttatgtc 7980

gacacctttt cagcaacttt tagtgttcct atggaaaaac ttaaggcact tgttgctaca 8040

gctcacagcg agttagcaaa gggtgtagct ttagatggtg tcctttctac attcgtgtca 8100

gctgcccgac aaggtgttgt tgataccgat gttgacacaa aggatgttat tgaatgtctc 8160

aaactttcac atcactctga cttagaagtg acaggtgaca gttgtaacaa tttcatgctc 8220

acctataata aggttgaaaa catgacgccc agagatcttg gcgcatgtat tgactgtaat 8280

gcaaggcata tcaatgccca agtagcaaaa agtcacaatg tttcactcat ctggaatgta 8340

aaagactaca tgtctttatc tgaacagctg cgtaaacaaa ttcgtagtgc tgccaagaag 8400

aacaacatac cttttagact aacttgtgct acaactagac aggttgtcaa tgtcataact 8460

actaaaatct cactcaaggg tggtaagatt gttagtactt gttttaaact tatgcttaag 8520

gccacattat tgtgcgttct tgctgcattg gtttgttata tcgttatgcc agtacataca 8580

ttgtcaatcc atgatggtta cacaaatgaa atcattggtt acaaagccat tcaggatggt 8640

gtcactcgtg acatcatttc tactgatgat tgttttgcaa ataaacatgc tggttttgac 8700

gcatggttta gccagcgtgg tggttcatac aaaaatgaca aaagctgccc tgtagtagct 8760

gctatcatta caagagagat tggtttcata gtgcctggct taccgggtac tgtgctgaga 8820

gcaatcaatg gtgacttctt gcattttcta cctcgtgttt ttagtgctgt tggcaacatt 8880

tgctacacac cttccaaact cattgagtat agtgattttg ctacctctgc ttgcgttctt 8940

gctgctgagt gtacaatttt taaggatgct atgggcaaac ctgtgccata ttgttatgac 9000

actaatttgc tagagggttc tatttcttat agtgagcttc gtccagacac tcgttatgtg 9060

cttatggatg gttccatcat acagtttcct aacacttacc tggagggttc tgttagagta 9120

gtaacaactt ttgatgctga gtactgtaga catggtacat gcgaaaggtc agaagtaggt 9180

atttgcctat ctaccagtgg tagatgggtt cttaataatg agcattacag agctctatca 9240

ggagttttct gtggtgttga tgcgatgaat ctcatagcta acatctttac tcctcttgtg 9300

caacctgtgg gtgctttaga tgtgtctgct tcagtagtgg ctggtggtat tattgccata 9360

ttggtgactt gtgctgccta ctactttatg aaattcagac gtgtttttgg tgagtacaac 9420

catgttgttg ctgctaatgc acttttgttt ttgatgtctt tcactatact ctgtctggta 9480

ccagcttaca gctttctgcc gggagtctac tcagtctttt acttgtactt gacattctat 9540

ttcaccaatg atgtttcatt cttggctcac cttcaatggt ttgccatgtt ttctcctatt 9600

gtgccttttt ggataacagc aatctatgta ttctgtattt ctctgaagca ctgccattgg 9660

ttctttaaca actatcttag gaaaagagtc atgtttaatg gagttacatt tagtaccttc 9720

gaggaggctg ctttgtgtac ctttttgctc aacaaggaaa tgtacctaaa attgcgtagc 9780

gagacactgt tgccacttac acagtataac aggtatcttg ctctatataa caagtacaag 9840

tatttcagtg gagccttaga tactaccagc tatcgtgaag cagcttgctg ccacttagca 9900

aaggctctaa atgactttag caactcaggt gctgatgttc tctaccaacc accacagaca 9960

tcaatcactt ctgctgttct gcagagtggt tttaggaaaa tggcattccc gtcaggcaaa 10020

gttgaagggt gcatggtaca agtaacctgt ggaactacaa ctcttaatgg attgtggttg 10080

gatgacacag tatactgtcc aagacatgtc atttgcacag cagaagacat gcttaatcct 10140

aactatgaag atctgctcat tcgcaaatcc aaccatagct ttcttgttca ggctggcaat 10200

-continued

gttcaacttc gtgttattgg ccattctatg caaaattgtc tgcttaggct taaagttgat 10260

acttctaacc ctaagacacc caagtataaa tttgtccgta tccaacctgg tcaaacattt 10320

tcagttctag catgctacaa tggttcacca tctggtgttt atcagtgtgc catgagacct 10380

aatcatacca ttaaaggttc tttccttaat ggatcatgtg gtagtgttgg ttttaacatt 10440

gattatgatt gcgtgtcttt ctgctatatg catcatatgg agcttccaac aggagtacac 10500

gctggtactg acttagaagg taaattctat ggtccatttg ttgacagaca aactgcacag 10560

gctgcaggta cagacacaac cataacatta aatgttttgg catggctgta tgctgctgtt 10620

atcaatggtg ataggtggtt tcttaataga ttcaccacta ctttgaatga ctttaacctt 10680

gtggcaatga agtacaacta tgaaccttg acacaagatc atgttgacat attgggacct 10740

ctttctgctc aaacaggaat tgccgtctta gatatgtgtg ctgctttgaa agagctgctg 10800

cagaatggta tgaatggtcg tactatcctt ggtagcacta ttttagaaga tgagtttaca 10860

ccatttgatg ttgttagaca atgctctggt gttaccttcc aaggtaagtt caagaaaatt 10920

gttaagggca ctcatcattg gatgctttta actttcttga catcactatt gattcttgtt 10980

caaagtacac agtggtcact gtttttcttt gtttacgaga atgctttctt gccatttact 11040

cttggtatta tggcaattgc tgcatgtgct atgctgcttg ttaagcataa gcacgcattc 11100

ttgtgcttgt ttctgttacc ttctcttgca acagttgctt actttaatat ggtctacatg 11160

cctgctagct gggtgatgcg tatcatgaca tggcttgaat tggctgacac tagcttgtct 11220

ggttataggc ttaaggattg tgttatgtat gcttcagctt tagttttgct tattctcatg 11280

acagctcgca ctgtttatga tgatgctgct agacgtgttt ggacactgat gaatgtcatt 11340

acacttgttt acaaagtcta ctatggtaat gctttagatc aagctatttc catgtgggcc 11400

ttagttattt ctgtaacctc taactattct ggtgtcgtta cgactatcat gtttttagct 11460

agagctatag tgtttgtgtg tgttgagtat tacccattgt tatttattac tggcaacacc 11520

ttacagtgta tcatgcttgt ttattgtttc ttaggctatt gttgctgctg ctactttggc 11580

cttttctgtt tactcaaccg ttacttcagg cttactcttg gtgtttatga ctacttggtc 11640

tctacacaag aatttaggta tatgaactcc caggggcttt tgcctcctaa gagtagtatt 11700

gatgctttca agcttaacat taagttgttg ggtattggag gtaaaccatg tatcaaggtt 11760

gctactgtac agtctaaaat gtctgacgta aagtgcacat ctgtggtact gctctcggtt 11820

cttcaacaac ttagagtaga gtcatcttct aaattgtggg cacaatgtgt acaactccac 11880

aatgatattc ttcttgcaaa agacacaact gaagctttcg agaagatggt ttctcttttg 11940

tctgttttgc tatccatgca gggtgctgta gacattaata ggttgtgcga ggaaatgctc 12000

gataaccgtg ctactcttca ggctattgct tcagaattta gttctttacc atcatatgcc 12060

gcttatgcca ctgcccagga ggcctatgag caggctgtag ctaatggtga ttctgaagtc 12120

gttctcaaaa agttaaagaa atctttgaat gtggctaaat ctgagtttga ccgtgatgct 12180

gccatgcaac gcaagttgga aaagatggca gatcaggcta tgacccaaat gtacaaacag 12240

gcaagatctg aggacaagag ggcaaaagta actagtgcta tgcaaacaat gctcttcact 12300

atgcttagga agcttgataa tgatgcactt aacaacatta tcaacaatgc gcgtgatggt 12360

tgtgttccac tcaacatcat accattgact acagcagcca aactcatggt tgttgtccct 12420

gattatggta cctacaagaa cacttgtgat ggtaacacct ttacatatgc atctgcactc 12480

tgggaaatcc agcaagttgt tgatgcggat agcaagattg ttcaacttag tgaaattaac 12540

-continued

atggacaatt caccaaattt ggcttggcct cttattgtta cagctctaag agccaactca 12600

gctgttaaac tacagaataa tgaactgagt ccagtagcac tacgacagat gtcctgtgcg 12660

gctggtacca cacaaacagc ttgtactgat gacaatgcac ttgcctacta taacaattcg 12720

aagggaggta ggtttgtgct ggcattacta tcagaccacc aagatctcaa atgggctaga 12780

ttccctaaga gtgatggtac aggtacaatt tacacagaac tggaaccacc ttgtaggttt 12840

gttacagaca caccaaaagg gcctaaagtg aaatacttgt acttcatcaa aggcttaaac 12900

aacctaaata gaggtatggt gctgggcagt ttagctgcta cagtacgtct tcaggctgga 12960

aatgctacag aagtacctgc caattcaact gtgctttcct tctgtgcttt tgcagtagac 13020

cctgctaaag catataagga ttacctagca agtggaggac aaccaatcac caactgtgtg 13080

aagatgttgt gtacacacac tggtacagga caggcaatta ctgtaacacc agaagctaac 13140

atggaccaag agtcctttgg tggtgcttca tgttgtctgt attgtagatg ccacattgac 13200

catccaaatc ctaaaggatt ctgtgacttg aaaggtaagt acgtccaaat acctaccact 13260

tgtgctaatg acccagtggg ttttacactt agaaacacag tctgtaccgt ctgcggaatg 13320

tggaaaggtt atggctgtag ttgtgaccaa ctccgcgaac ccttgatgca gtctgcggat 13380

gcatcaacgt ttttaaacgg gtttgcggtg taagtgcagc ccgtcttaca ccgtgcggca 13440

caggcactag tactgatgtc gtctacaggg cttttgatat ttacaacgaa aaagttgctg 13500

gttttgcaaa gttcctaaaa actaattgct gtcgcttcca ggagaaggat gaggaaggca 13560

atttattaga ctcttacttt gtagttaaga ggcatactat gtctaactac caacatgaag 13620

agactattta taacttggtt aaagattgtc cagcggttgc tgtccatgac tttttcaagt 13680

ttagagtaga tggtgacatg gtaccacata tatcacgtca gcgtctaact aaatacacaa 13740

tggctgattt agtctatgct ctacgtcatt ttgatgaggg taattgtgat acattaaaag 13800

aaatactcgt cacatacaat tgctgtgatg atgattattt caataagaag gattggtatg 13860

acttcgtaga gaatcctgac atcttacgcg tatatgctaa cttaggtgag cgtgtacgcc 13920

aatcattatt aaagactgta caattctgcg atgctatgcg tgatgcaggc attgtaggcg 13980

tactgacatt agataatcag gatcttaatg ggaactggta cgatttcggt gatttcgtac 14040

aagtagcacc aggctgcgga gttcctattg tggattcata ttactcattg ctgatgccca 14100

tcctcacttt gactagggca ttggctgctg agtcccatat ggatgctgat ctcgcaaaac 14160

cacttattaa gtgggatttg ctgaaatatg attttacgga agagagactt tgtctcttcg 14220

accgttattt taaatattgg gaccagacat accatcccaa ttgtattaac tgtttggatg 14280

ataggtgtat ccttcattgt gcaaacttta atgtgttatt ttctactgtg tttccaccta 14340

caagttttgg accactagta agaaaaatat ttgtagatgg tgttccttttt gttgtttcaa 14400

ctggatacca ttttcgtgag ttaggagtcg tacataatca ggatgtaaac ttacatagct 14460

cgcgtctcag tttcaaggaa cttttagtgt atgctgctga tccagctatg catgcagctt 14520

ctggcaattt attgctagat aaacgcacta catgcttttc agtagctgca ctaacaaaca 14580

atgttgcttt tcaaactgtc aaacccggta attttaataa agacttttat gactttgctg 14640

tgtctaaagg tttctttaag gaaggaagtt ctgttgaact aaaacacttc ttctttgctc 14700

aggatggcaa cgctgctatc agtgattatg actattatcg ttataatctg ccaacaatgt 14760

gtgatatcag acaactccta ttcgtagttg aagttgttga taaatacttt gattgttacg 14820

atggtggctg tattaatgcc aaccaagtaa tcgttaacaa tctggataaa tcagctggtt 14880

tcccatttaa taaatggggt aaggctagac tttattatga ctcaatgagt tatgaggatc 14940

-continued

```
aagatgcact tttcgcgtat actaagcgta atgtcatccc tactataact caaatgaatc  15000
ttaagtatgc cattagtgca aagaatagag ctcgcaccgt agctggtgtc tctatctgta  15060
gtactatgac aaatagacag tttcatcaga aattattgaa gtcaatagcc gccactagag  15120
gagctactgt ggtaattgga acaagcaagt tttacggtgg ctggcataat atgttaaaaa  15180
ctgtttacag tgatgtagaa actccacacc ttatgggttg ggattatcca aaatgtgaca  15240
gagccatgcc taacatgctt aggataatgg cctctcttgt tcttgctcgc aaacataaca  15300
cttgctgtaa cttatcacac cgtttctaca ggttagctaa cgagtgtgcg caagtattaa  15360
gtgagatggt catgtgtggc ggctcactat atgttaaacc aggtggaaca tcatccggtg  15420
atgctacaac tgcttatgct aatagtgtct ttaacatttg tcaagctgtt acagccaatg  15480
taaatgcact tctttcaact gatggtaata agatagctga caagtatgtc cgcaatctac  15540
aacacaggct ctatgagtgt ctctatagaa atagggatgt tgatcatgaa ttcgtggatg  15600
agttttacgc ttacctgcgt aaacatttct ccatgatgat tctttctgat gatgccgttg  15660
tgtgctataa cagtaactat gcggctcaag gtttagtagc tagcattaag aactttaagg  15720
cagttcttta ttatcaaaat aatgtgttca tgtctgaggc aaaatgttgg actgagactg  15780
accttactaa aggacctcac gaattttgct cacagcatac aatgctagtt aaacaaggag  15840
atgattacgt gtacctgcct tacccagatc catcaagaat attaggcgca ggctgttttg  15900
tcgatgatat tgtcaaaaca gatggtacac ttatgattga aaggttcgtg tcactggcta  15960
ttgatgctta cccacttaca aaacatccta atcaggagta tgctgatgtc tttcacttgt  16020
atttacaata cattagaaag ttacatgatg agcttactgg ccacatgttg gacatgtatt  16080
ccgtaatgct aactaatgat aacacctcac ggtactggga acctgagttt tatgaggcta  16140
tgtacacacc acatacagtc ttgcaggctg taggtgcttg tgtattgtgc aattcacaga  16200
cttcacttcg ttgcggtgcc tgtattagga gaccattcct atgttgcaag tgctgctatg  16260
accatgtcat ttcaacatca cacaaattag tgttgtctgt taatccctat gtttgcaatg  16320
ccccaggttg tgatgtcact gatgtgacac aactgtatct aggaggtatg agctattatt  16380
gcaagtcaca taagcctccc attagttttc cattatgtgc taatggtcag gtttttggtt  16440
tatacaaaaa cacatgtgta ggcagtgaca atgtcactga cttcaatgcg atagcaacat  16500
gtgattggac taatgctggc gattacatac ttgccaacac ttgtactgag agactcaagc  16560
ttttcgcagc agaaacgctc aaagccactg aggaaacatt taagctgtca tatggtattg  16620
ctactgtacg cgaagtactc tctgacagag aattgcatct ttcatgggag gttggaaaac  16680
ctagaccacc attgaacaga aactatgtct ttactggtta ccgtgtaact aaaaatagta  16740
aagtacagat tggagagtac acctttgaaa aaggtgacta tggtgatgct gttgtgtaca  16800
gaggtactac gacatacaag ttgaatgttg gtgattactt tgtgttgaca tctcacactg  16860
taatgccact tagtgcacct actctagtgc cacaagagca ctatgtgaga attactggct  16920
tgtacccaac actcaacatc tcagatgagt tttctagcaa tgttgcaaat tatcaaaagg  16980
tcggcatgca aaagtactct acactccaag gaccacctgg tactggtaag agtcattttg  17040
ccatcggact tgctctctat tacccatctg ctcgcatagt gtatacggca tgctctcatg  17100
cagctgttga tgccctatgt gaaaaggcat taaaatattt gcccatagat aaatgtagta  17160
gaatcatacc tgcgcgtgcg cgcgtagagt gttttgataa attcaaagtg aattcaacac  17220
tagaacagta tgttttctgc actgtaaatg cattgccaga aacaactgct gacattgtag  17280
```

-continued

tctttgatga aatctctatg gctactaatt atgacttgag tgttgtcaat gctagacttc 17340

gtgcaaaaca ctacgtctat attggcgatc ctgctcaatt accagccccc cgcacattgc 17400

tgactaaagg cacactagaa ccagaatatt ttaattcagt gtgcagactt atgaaaacaa 17460

taggtccaga catgttcctt ggaacttgtc gccgttgtcc tgctgaaatt gttgacactg 17520

tgagtgcttt agtttatgac aataagctaa aagcacacaa ggataagtca gctcaatgct 17580

tcaaaatgtt ctacaaaggt gttattacac atgatgtttc atctgcaatc aacagacctc 17640

aaataggcgt tgtaagagaa tttcttacac gcaatcctgc ttggagaaaa gctgttttta 17700

tctcaccttа taattcacag aacgctgtag cttcaaaaat cttaggattg cctacgcaga 17760

ctgttgattc atcacagggt tctgaatatg actatgtcat attcacacaa actactgaaa 17820

cagcacactc ttgtaatgtc aaccgcttca atgtggctat cacaagggca aaaattggca 17880

ttttgtgcat aatgtctgat agagatcttt atgacaaact gcaatttaca agtctagaaa 17940

taccacgtcg caatgtggct acattacaag cagaaaatgt aactggactt tttaaggact 18000

gtagtaagat cattactggt cttcatccta cacaggcacc tacacacctc agcgttgata 18060

taaagttcaa gactgaagga ttatgtgttg acataccagg cataccaaag gacatgacct 18120

accgtagact catctctatg atgggtttca aaatgaatta ccaagtcaat ggttacccta 18180

atatgtttat cacccgcgaa gaagctattc gtcacgttcg tgcgtggatt ggctttgatg 18240

tagagggctg tcatgcaact agagatgctg tgggtactaa cctacctctc cagctaggat 18300

tttctacagg tgttaactta gtagctgtac cgactggtta tgttgacact gaaaataaca 18360

cagaattcac cagagttaat gcaaaacctc caccaggtga ccagtttaaa catcttatac 18420

cactcatgta taaaggcttg ccctggaatg tagtgcgtat taagatagta caaatgctca 18480

gtgatacact gaaaggattg tcagacagag tcgtgttcgt cctttgggcg catggctttg 18540

agcttacatc aatgaagtac tttgtcaaga ttggacctga aagaacgtgt tgtctgtgtg 18600

acaaacgtgc aacttgcttt tctacttcat cagatactta tgcctgctgg aatcattctg 18660

tgggttttga ctatgtctat aacccattta tgattgatgt tcagcagtgg ggctttacgg 18720

gtaaccttca gagtaaccat gaccaacatt gccaggtaca tggaaatgca catgtggcta 18780

gttgtgatgc tatcatgact agatgtttag cagtccatga gtgctttgtt aagcgcgttg 18840

attggtctgt tgaataccct attataggag atgaactgag ggttaattct gcttgcagaa 18900

aagtacaaca catggttgtg aagtctgcat tgcttgctga taagtttcca gttcttcatg 18960

acattggaaa tccaaaggct atcaagtgtg tgcctcaggc tgaagtagaa tggaagttct 19020

acgatgctca gccatgtagt gacaaagctt acaaaataga ggagctcttc tattcttatg 19080

ctacacatca cgataaattc actgatggtg tttgtttgtt ttggaattgt aacgttgatc 19140

gttacccagc caatgcaatt gtgtgtaggt ttgacacaag agccttgtca aacttgaact 19200

taccaggctg tgatggtggt agtttgtatg tgaataagca tgcattccac actccagctt 19260

tcgataaaag tgcatttact aatttaaagc aattgccttt cttttactat tctgatagtc 19320

cttgtgagtc tcatggcaaa caagtagtgt cggatattga ttatgttcca ctcaaatctg 19380

ctacgtgtat tacacgatgc aatttaggtg gtgctgtttg cagacaccat gcaaatgagt 19440

accgacagta cttggatgca tataatatga tgatttctgc tggatttagc ctatggattt 19500

acaaacaatt tgatacttat aacctgtgga atacatttac caggttacag agtttagaaa 19560

atgtggctta taatgttgtt aataaaggac actttgatgg acacgccggc gaagcacctg 19620

tttccatcat taataatgct gtttacacaa aggtagatgg tattgatgtg gagatctttg 19680

-continued

aaaataagac aacacttcct gttaatgttg catttgagct ttgggctaag cgtaacatta 19740

aaccagtgcc agagattaag atactcaata atttgggtgt tgatatcgct gctaatactg 19800

taatctggga ctacaaaaga gaagccccag cacatgtatc tacaataggt gtctgcacaa 19860

tgactgacat tgccaagaaa cctactgaga gtgcttgttc ttcacttact gtcttgtttg 19920

atggtagagt ggaaggacag gtagaccttt ttagaaacgc ccgtaatggt gttttaataa 19980

cagaaggttc agtcaaaggt ctaacacctt caaagggacc agcacaagct agcgtcaatg 20040

gagtcacatt aattggagaa tcagtaaaaa cacagtttaa ctactttaag aaagtagacg 20100

gcattattca acagttgcct gaaacctact ttactcagag cagagactta gaggatttta 20160

agcccagatc acaaatggaa actgactttc tcgagctcgc tatggatgaa ttcatacagc 20220

gatataagct cgagggctat gccttcgaac acatcgttta tggagatttc agtcatggac 20280

aacttggcgg tcttcattta atgataggct tagccaagcg ctcacaagat tcaccactta 20340

aattagagga ttttatccct atggacagca cagtgaaaaa ttacttcata acagatgcgc 20400

aaacaggttc atcaaaatgt gtgtgttctg tgattgatct tttacttgat gactttgtcg 20460

agataataaa gtcacaagat ttgtcagtga tttcaaaagt ggtcaaggtt acaattgact 20520

atgctgaaat ttcattcatg ctttggtgta aggatggaca tgttgaaacc ttctacccaa 20580

aactacaagc aagtcaagcg tggcaaccag gtgttgcgat gcctaacttg tacaagatgc 20640

aaagaatgct tcttgaaaag tgtgaccttc agaattatgg tgaaaatgct gttataccaa 20700

aaggaataat gatgaatgtc gcaaagtata ctcaactgtg tcaatactta aatacactta 20760

ctttagctgt accctacaac atgagagtta ttcactttgg tgctggctct gataaaggag 20820

ttgcaccagg tacagctgtg ctcagacaat ggttgccaac tggcacacta cttgtcgatt 20880

cagatcttaa tgacttcgtc tccgacgcag attctacttt aattggagac tgtgcaacag 20940

tacatacggc taataaatgg gaccttatta ttagcgatat gtatgaccct aggaccaaac 21000

atgtgacaaa agagaatgac tctaaagaag ggtttttcac ttatctgtgt ggatttataa 21060

agcaaaaact agccctgggt ggttctatag ctgtaaagat aacagagcat tcttggaatg 21120

ctgaccttta caagcttatg ggccatttct catggtggac agcttttgtt acaaatgtaa 21180

atgcatcatc atcggaagca tttttaattg gggctaacta tcttggcaag ccgaaggaac 21240

aaattgatgg ctataccatg catgctaact acattttctg gaggaacaca aatcctatcc 21300

agttgtcttc ctattcactc tttgacatga gcaaatttcc tcttaaatta agaggaactg 21360

ctgtaatgtc tcttaaggag aatcaaatca atgatatgat ttattctctt ctggaaaaag 21420

gtaggcttat cattagagaa aacaacagag ttgtggtttc aagtgatatt cttgttaaca 21480

actaaacgaa catgtttatt ttcttattat ttcttactct cactagtggt agtgaccttg 21540

accggtgcac cacttttgat gatgttcaag ctcctaatta cactcaacat acttcatcta 21600

tgaggggggt ttactatcct gatgaaattt ttagatcaga cactctttat ttaactcagg 21660

atttatttct tccattttat tctaatgtta cagggtttca tactattaat catacgtttg 21720

gcaaccctgt catacctttt aaggatggta tttattttgc tgccacagag aaatcaaatg 21780

ttgtccgtgg ttgggttttt ggttctacca tgaacaacaa gtcacagtcg gtgattatta 21840

ttaacaattc tactaatgtt gttatacgag catgtaactt tgaattgtgt gacaaccctt 21900

tctttgctgt ttctaaaccc atgggtacac agacacatac tatgatattc gataatgcat 21960

ttaattgcac tttcgagtac atatctgatg ccttttcgct tgatgtttca gaaaagtcag 22020

-continued

gtaattttaa acacttacga gagtttgtgt ttaaaaataa agatgggttt ctctatgttt 22080

ataagggcta tcaacctata gatgtagttc gtgatctacc ttctggtttt aacactttga 22140

aacctatttt taagttgcct cttggtatta acattacaaa ttttagagcc attcttacag 22200

ccttttcacc tgctcaagac atttggggca cgtcagctgc agcctatttt gttggctatt 22260

taaagccaac tacatttatg ctcaagtatg atgaaaatgg tacaatcaca gatgctgttg 22320

attgttctca aaatccactt gctgaactca aatgctctgt taagagcttt gagattgaca 22380

aaggaattta ccagacctct aatttcaggg ttgttccctc aggagatgtt gtgagattcc 22440

ctaatattac aaacttgtgt ccttttggag aggtttttaa tgctactaaa ttcccttctg 22500

tctatgcatg ggagagaaaa aaaatttcta attgtgttgc tgattactct gtgctctaca 22560

actcaacatt tttttcaacc tttaagtgct atggcgtttc tgccactaag ttgaatgatc 22620

tttgcttctc caatgtctat gcagattctt ttgtagtcaa gggagatgat gtaagacaaa 22680

tagcgccagg acaaactggt gttattgctg attataatta taaattgcca gatgatttca 22740

tgggttgtgt ccttgcttgg aatactagga acattgatgc tacttcaact ggtaattata 22800

attataaata taggtatctt agacatggca agcttaggcc ctttgagaga gacatatcta 22860

atgtgccttt ctcccctgat ggcaaacctt gcaccccacc tgctcttaat tgttattggc 22920

cattaaatga ttatggtttt tacaccacta ctggcattgg ctaccaacct tacagagttg 22980

tagtactttc ttttgaactt ttaaatgcac cggccacggt ttgtggacca aaattatcca 23040

ctgaccttat taagaaccag tgtgtcaatt ttaattttaa tggactcact ggtactggtg 23100

tgttaactcc ttcttcaaag agatttcaac catttcaaca atttggccgt gatgtttctg 23160

atttcactga ttccgttcga gatcctaaaa catctgaaat attagacatt tcaccttgct 23220

cttttggggg tgtaagtgta attacacctg gaacaaatgc ttcatctgaa gttgctgttc 23280

tatatcaaga tgttaactgc actgatgttt ctacagcaat tcatgcagat caactcacac 23340

cagcttggcg catatattct actggaaaca atgtattcca gactcaagca ggctgtctta 23400

taggagctga gcatgtcgac acttcttatg agtgcgacat tcctattgga gctggcattt 23460

gtgctagtta ccatacagtt tctttattac gtagtactag ccaaaaatct attgtggctt 23520

atactatgtc tttaggtgct gatagttcaa ttgcttactc taataacacc attgctatac 23580

ctactaactt ttcaattagc attactacag aagtaatgcc tgtttctatg gctaaaacct 23640

ccgtagattg taatatgtac atctgcggag attctactga atgtgctaat ttgcttctcc 23700

aatatggtag cttttgcaca caactaaatc gtgcactctc aggtattgct gctgaacagg 23760

atcgcaacac acgtgaagtg ttcgctcaag tcaaacaaat gtacaaaacc ccaactttga 23820

aatattttgg tggttttaat ttttcacaaa tattacctga ccctctaaag ccaactaaga 23880

ggtcttttat tgaggacttg ctctttaata aggtgacact cgctgatgct ggcttcatga 23940

agcaatatgg cgaatgccta ggtgatatta atgctagaga tctcatttgt gcgcagaagt 24000

tcaatggact tacagtgttg ccacctctgc tcactgatga tatgattgct gcctacactg 24060

ctgctctagt tagtggtact gccactgctg gatggacatt tggtgctggc gctgctcttc 24120

aaataccttt tgctatgcaa atggcatata ggttcaatgg cattggagtt acccaaaatg 24180

ttctctatga gaaccaaaaa caaatcgcca accaatttaa caaggcgatt agtcaaattc 24240

aagaatcact tacaacaaca tcaactgcat tgggcaagct gcaagacgtt gttaaccaga 24300

atgctcaagc attaaacaca cttgttaaac aacttagctc taattttggt gcaatttcaa 24360

gtgtgctaaa tgatatcctt tcgcgacttg ataaagtcga ggcggaggta caaattgaca 24420

-continued

ggttaattac aggcagactt caaagccttc aaacctatgt aacacaacaa ctaatcaggg 24480

ctgctgaaat cagggcttct gctaatcttg ctgctactaa aatgtctgag tgtgttcttg 24540

gacaatcaaa aagagttgac ttttgtggaa agggctacca ccttatgtcc ttcccacaag 24600

cagccccgca tggtgttgtc ttcctacatg tcacgtatgt gccatcccag gagaggaact 24660

tcaccacagc gccagcaatt tgtcatgaag gcaaagcata cttccctcgt gaaggtgttt 24720

ttgtgtttaa tggcacttct tggtttatta cacagaggaa cttcttttct ccacaaataa 24780

ttactacaga caatacattt gtctcaggaa attgtgatgt cgttattggc atcattaaca 24840

acacagttta tgatcctctg caacctgagc tcgactcatt caaagaagag ctggacaagt 24900

acttcaaaaa tcatacatca ccagatgttg atcttggcga catttcaggc attaacgctt 24960

ctgtcgtcaa cattcaaaaa gaaattgacc gcctcaatga ggtcgctaaa aatttaaatg 25020

aatcactcat tgaccttcaa gaattgggaa aatatgagca atatattaaa tggccttggt 25080

atgtttggct cggcttcatt gctggactaa ttgccatcgt catggttaca atcttgcttt 25140

gttgcatgac tagttgttgc agttgcctca agggtgcatg ctcttgtggt tcttgctgca 25200

agtttgatga ggatgactct gagccagttc tcaagggtgt caaattacat tacacataaa 25260

cgaacttatg gatttgttta tgagattttt tactcttgga tcaattactg cacagccagt 25320

aaaaattgac aatgcttctc ctgcaagtac tgttcatgct acagcaacga taccgctaca 25380

agcctcactc cctttcggat ggcttgttat tggcgttgca tttcttgctg tttttcagag 25440

cgctaccaaa ataattgcgc tcaataaaag atggcagcta gccctttata agggcttcca 25500

gttcatttgc aatttactgc tgctatttgt taccatctat tcacatcttt tgcttgtcgc 25560

tgcaggtatg gaggcgcaat ttttgtacct ctatgccttg atatattttc tacaatgcat 25620

caacgcatgt agaattatta tgagatgttg gctttgttgg aagtgcaaat ccaagaaccc 25680

attactttat gatgccaact actttgtttg ctggcacaca cataactatg actactgtat 25740

accatataac agtgtcacag atacaattgt cgttactgaa ggtgacggca tttcaacacc 25800

aaaactcaaa gaagactacc aaattggtgg ttattctgag gataggcact caggtgttaa 25860

agactatgtc gttgtacatg gctatttcac cgaagtttac taccagcttg agtctacaca 25920

aattactaca gacactggta ttgaaaatgc tacattcttc atctttaaca agcttgttaa 25980

agacccaccg aatgtgcaaa tacacacaat cgacggctct tcaggagttg ctaatccagc 26040

aatggatcca atttatgatg agccgacgac gactactagc gtgcctttgt aagcacaaga 26100

aagtgagtac gaacttatgt actcattcgt ttcggaagaa acaggtacgt taatagttaa 26160

tagcgtactt ctttttcttg ctttcgtggt attcttgcta gtcacactag ccatccttac 26220

tgcgcttcga ttgtgtgcgt actgctgcaa tattgttaac gtgagtttag taaaaccaac 26280

ggtttacgtc tactcgcgtg ttaaaaatct gaactcttct gaaggagttc ctgatcttct 26340

ggtctaaacg aactaactat tattattatt ctgtttggaa ctttaacatt gcttatcatg 26400

gcagacaacg gtactattac cgttgaggag cttaaacaac tcctggaaca atggaaccta 26460

gtaataggtt tcctattcct agcctggatt atgttactac aatttgccta ttctaatcgg 26520

aacaggtttt tgtacataat aaagcttgtt ttcctctggc tcttgtggcc agtaacactt 26580

gcttgttttg tgcttgctgc tgtctacaga attaattggg tgactggcgg gattgcgatt 26640

gcaatggctt gtattgtagg cttgatgtgg cttagctact tcgttgcttc cttcaggctg 26700

tttgctcgta cccgctcaat gtggtcattc aacccagaaa caaacattct tctcaatgtg 26760

-continued

```
cctctccggg ggacaattgt gaccagaccg ctcatggaaa gtgaacttgt cattggtgct  26820
gtgatcattc gtggtcactt gcgaatggcc ggacaccccc tagggcgctg tgacattaag  26880
gacctgccaa aagagatcac tgtggctaca tcacgaacgc tttcttatta caaattagga  26940
gcgtcgcagc gtgtaggcac tgattcaggt tttgctgcat acaaccgcta ccgtattgga  27000
aactataaat taaatacaga ccacgccggt agcaacgaca atattgcttt gctagtacag  27060
taagtgacaa cagatgtttc atcttgttga cttccaggtt acaatagcag agatattgat  27120
tatcattatg aggactttca ggattgctat ttggaatctt gacgttataa taagttcaat  27180
agtgagacaa ttatttaagc ctctaactaa gaagaattat tcggagttag atgatgaaga  27240
acctatggag ttagattatc cataaaacga acatgaaaat tattctcttc ctgacattga  27300
ttgtatttac atcttgcgag ctatatcact atcaggagtg tgttagaggt acgactgtac  27360
tactaaaaga accttgccca tcaggaacat acgagggcaa ttcaccattt caccctcttg  27420
ctgacaataa atttgcacta acttgcacta gcacacactt tgcttttgct tgtgctgacg  27480
gtactcgaca tacctatcag ctgcgtgcaa gatcagtttc accaaaactt ttcatcagac  27540
aagaggaggt tcaacaagag ctctactcgc cacttttttct cattgttgct gctctagtat  27600
ttttaatact ttgcttcacc attaagagaa agacagaatg aatgagctca ctttaattga  27660
cttctatttg tgctttttag cctttctgct attccttgtt ttaataatgc ttattatatt  27720
ttggttttca ctcgaaatcc aggatctaga agaaccttgt accaaagtct aaacgaacat  27780
gaaacttctc attgttttga cttgtatttc tctatgcagt tgcatatgca ctgtagtaca  27840
gcgctgtgca tctaataaac ctcatgtgct tgaagatcct tgtaaggtac aacactaggg  27900
gtaatactta tagcactgct tggctttgtg ctctaggaaa ggttttacct tttcatagat  27960
ggcacactat ggttcaaaca tgcacaccta atgttactat caactgtcaa gatccagctg  28020
gtggtgcgct tatagctagg tgttggtacc ttcatgaagg tcaccaaact gctgcattta  28080
gagacgtact tgttgtttta aataaacgaa caaattaaaa tgtctgataa tggaccccaa  28140
tcaaaccaac gtagtgcccc ccgcattaca tttggtggac ccacagattc aactgacaat  28200
aaccagaatg gaggacgcaa tggggcaagg ccaaaacagc gccgacccca aggtttaccc  28260
aataatactg cgtcttggtt cacagctctc actcagcatg gcaaggagga acttagattc  28320
cctcgaggcc agggcgttcc aatcaacacc aatagtggtc cagatgacca aattggctac  28380
taccgaagag ctacccgacg agttcgtggt ggtgacggca aaatgaaaga gctcagcccc  28440
agatggtact tctattacct aggaactggc ccagaagctt cacttcccta cggcgctaac  28500
aaagaaggca tcgtatgggt tgcaactgag ggagccttga atacacccaa agaccacatt  28560
ggcacccgca atcctaataa caatgctgcc accgtgctac aacttcctca aggaacaaca  28620
ttgccaaaag gcttctacgc agagggaagc agaggcggca gtcaagcctc ttctcgctcc  28680
tcatcacgta gtcgcggtaa ttcaagaaat tcaactcctg gcagcagtag gggaaattct  28740
cctgctcgaa tggctagcgg aggtggtgaa actgccctcg cgctattgct gctagacaga  28800
ttgaaccagc ttgagagcaa agtttctggt aaaggccaac aacaacaagg ccaaactgtc  28860
actaagaaat ctgctgctga ggcatctaaa aagcctcgcc aaaaacgtac tgccacaaaa  28920
cagtacaacg tcactcaagc atttgggaga cgtggtccag aacaaaccca aggaaatttc  28980
ggggaccaag acctaatcag acaaggaact gattacaaac attggccgca aattgcacaa  29040
tttgctccaa gtgcctctgc attctttgga atgtcacgca ttggcatgga agtcacacct  29100
tcgggaacat ggctgactta tcatggagcc attaaattgg atgacaaaga tccacaattc  29160
```

-continued

```
aaagacaacg tcatactgct gaacaagcac attgacgcat acaaaacatt cccaccaaca  29220

gagcctaaaa aggacaaaaa gaaaaagact gatgaagctc agcctttgcc gcagagacaa  29280

aagaagcagc ccactgtgac tcttcttcct gcggctgaca tggatgattt ctccagacaa  29340

cttcaaaatt ccatgagtgg agcttctgct gattcaactc aggcataaac actcatgatg  29400

accacacaag gcagatgggc tatgtaaacg ttttcgcaat tccgtttacg atacatagtc  29460

tactcttgtg cagaatgaat tctcgtaact aaacagcaca agtaggttta gttaacttta  29520

atctcacata gcaatcttta atcaatgtgt aacattaggg aggacttgaa agagccacca  29580

cattttcatc gaggccacgc ggagtacgat cgagggtaca gtgaataatg ctagggagag  29640

ctgcctatat ggaagagccc taatgtgtaa aattaatttt agtagtgcta tccccatgtg  29700

attttaatag cttcttagga gaatgac                                     29727
```

<210> SEQ ID NO 2
<211> LENGTH: 4382
<212> TYPE: PRT
<213> ORGANISM: Coronavirus

<400> SEQUENCE: 2

```
Met Glu Ser Leu Val Leu Gly Val Asn Glu Lys Thr His Val Gln Leu
1               5                   10                  15

Ser Leu Pro Val Leu Gln Val Arg Asp Val Leu Val Arg Gly Phe Gly
            20                  25                  30

Asp Ser Val Glu Glu Ala Leu Ser Glu Ala Arg Glu His Leu Lys Asn
        35                  40                  45

Gly Thr Cys Gly Leu Val Glu Leu Glu Lys Gly Val Leu Pro Gln Leu
    50                  55                  60

Glu Gln Pro Tyr Val Phe Ile Lys Arg Ser Asp Ala Leu Ser Thr Asn
65                  70                  75                  80

His Gly His Lys Val Val Glu Leu Val Ala Glu Met Asp Gly Ile Gln
                85                  90                  95

Tyr Gly Arg Ser Gly Ile Thr Leu Gly Val Leu Val Pro His Val Gly
            100                 105                 110

Glu Thr Pro Ile Ala Tyr Arg Asn Val Leu Leu Arg Lys Asn Gly Asn
        115                 120                 125

Lys Gly Ala Gly Gly His Ser Tyr Gly Ile Asp Leu Lys Ser Tyr Asp
    130                 135                 140

Leu Gly Asp Glu Leu Gly Thr Asp Pro Ile Glu Asp Tyr Glu Gln Asn
145                 150                 155                 160

Trp Asn Thr Lys His Gly Ser Gly Ala Leu Arg Glu Leu Thr Arg Glu
                165                 170                 175

Leu Asn Gly Gly Ala Val Thr Arg Tyr Val Asp Asn Asn Phe Cys Gly
            180                 185                 190

Pro Asp Gly Tyr Pro Leu Asp Cys Ile Lys Asp Phe Leu Ala Arg Ala
        195                 200                 205

Gly Lys Ser Met Cys Thr Leu Ser Glu Gln Leu Asp Tyr Ile Glu Ser
    210                 215                 220

Lys Arg Gly Val Tyr Cys Cys Arg Asp His Glu His Glu Ile Ala Trp
225                 230                 235                 240

Phe Thr Glu Arg Ser Asp Lys Ser Tyr Glu His Gln Thr Pro Phe Glu
                245                 250                 255

Ile Lys Ser Ala Lys Lys Phe Asp Thr Phe Lys Gly Glu Cys Pro Lys
            260                 265                 270
```

-continued

```
Phe Val Phe Pro Leu Asn Ser Lys Val Lys Val Ile Gln Pro Arg Val
        275                 280                 285

Glu Lys Lys Lys Thr Glu Gly Phe Met Gly Arg Ile Arg Ser Val Tyr
    290                 295                 300

Pro Val Ala Ser Pro Gln Glu Cys Asn Asn Met His Leu Ser Thr Leu
305                 310                 315                 320

Met Lys Cys Asn His Cys Asp Glu Val Ser Trp Gln Thr Cys Asp Phe
                325                 330                 335

Leu Lys Ala Thr Cys Glu His Cys Gly Thr Glu Asn Leu Val Ile Glu
            340                 345                 350

Gly Pro Thr Thr Cys Gly Tyr Leu Pro Thr Asn Ala Val Val Lys Met
        355                 360                 365

Pro Cys Pro Ala Cys Gln Asp Pro Glu Ile Gly Pro Glu His Ser Val
    370                 375                 380

Ala Asp Tyr His Asn His Ser Asn Ile Glu Thr Arg Leu Arg Lys Gly
385                 390                 395                 400

Gly Arg Thr Arg Cys Phe Gly Gly Cys Val Phe Ala Tyr Val Gly Cys
                405                 410                 415

Tyr Asn Lys Arg Ala Tyr Trp Val Pro Arg Ala Ser Ala Asp Ile Gly
            420                 425                 430

Ser Gly His Thr Gly Ile Thr Gly Asp Asn Val Glu Thr Leu Asn Glu
        435                 440                 445

Asp Leu Leu Glu Ile Leu Ser Arg Glu Arg Val Asn Ile Asn Ile Val
    450                 455                 460

Gly Asp Phe His Leu Asn Glu Glu Val Ala Ile Ile Leu Ala Ser Phe
465                 470                 475                 480

Ser Ala Ser Thr Ser Ala Phe Ile Asp Thr Ile Lys Ser Leu Asp Tyr
                485                 490                 495

Lys Ser Phe Lys Thr Ile Val Glu Ser Cys Gly Asn Tyr Lys Val Thr
            500                 505                 510

Lys Gly Lys Pro Val Lys Gly Ala Trp Asn Ile Gly Gln Gln Arg Ser
        515                 520                 525

Val Leu Thr Pro Leu Cys Gly Phe Pro Ser Gln Ala Ala Gly Val Ile
    530                 535                 540

Arg Ser Ile Phe Ala Arg Thr Leu Asp Ala Ala Asn His Ser Ile Pro
545                 550                 555                 560

Asp Leu Gln Arg Ala Ala Val Thr Ile Leu Asp Gly Ile Ser Glu Gln
                565                 570                 575

Ser Leu Arg Leu Val Asp Ala Met Val Tyr Thr Ser Asp Leu Leu Thr
            580                 585                 590

Asn Ser Val Ile Ile Met Ala Tyr Val Thr Gly Gly Leu Val Gln Gln
        595                 600                 605

Thr Ser Gln Trp Leu Ser Asn Leu Leu Gly Thr Thr Val Glu Lys Leu
    610                 615                 620

Arg Pro Ile Phe Glu Trp Ile Glu Ala Lys Leu Ser Ala Gly Val Glu
625                 630                 635                 640

Phe Leu Lys Asp Ala Trp Glu Ile Leu Lys Phe Leu Ile Thr Gly Val
                645                 650                 655

Phe Asp Ile Val Lys Gly Gln Ile Gln Val Ala Ser Asp Asn Ile Lys
            660                 665                 670

Asp Cys Val Lys Cys Phe Ile Asp Val Val Asn Lys Ala Leu Glu Met
        675                 680                 685
```

-continued

```
Cys Ile Asp Gln Val Thr Ile Ala Gly Ala Lys Leu Arg Ser Leu Asn
    690                 695                 700

Leu Gly Glu Val Phe Ile Ala Gln Ser Lys Gly Leu Tyr Arg Gln Cys
705                 710                 715                 720

Ile Arg Gly Lys Glu Gln Leu Gln Leu Leu Met Pro Leu Lys Ala Pro
                725                 730                 735

Lys Glu Val Thr Phe Leu Glu Gly Asp Ser His Asp Thr Val Leu Thr
            740                 745                 750

Ser Glu Glu Val Val Leu Lys Asn Gly Glu Leu Glu Ala Leu Glu Thr
        755                 760                 765

Pro Val Asp Ser Phe Thr Asn Gly Ala Ile Val Gly Thr Pro Val Cys
    770                 775                 780

Val Asn Gly Leu Met Leu Leu Glu Ile Lys Asp Lys Glu Gln Tyr Cys
785                 790                 795                 800

Ala Leu Ser Pro Gly Leu Leu Ala Thr Asn Asn Val Phe Arg Leu Lys
                805                 810                 815

Gly Gly Ala Pro Ile Lys Gly Val Thr Phe Gly Glu Asp Thr Val Trp
            820                 825                 830

Glu Val Gln Gly Tyr Lys Asn Val Arg Ile Thr Phe Glu Leu Asp Glu
        835                 840                 845

Arg Val Asp Lys Val Leu Asn Glu Lys Cys Ser Val Tyr Thr Val Glu
    850                 855                 860

Ser Gly Thr Glu Val Thr Glu Phe Ala Cys Val Val Ala Glu Ala Val
865                 870                 875                 880

Val Lys Thr Leu Gln Pro Val Ser Asp Leu Leu Thr Asn Met Gly Ile
                885                 890                 895

Asp Leu Asp Glu Trp Ser Val Ala Thr Phe Tyr Leu Phe Asp Asp Ala
            900                 905                 910

Gly Glu Glu Asn Phe Ser Ser Arg Met Tyr Cys Ser Phe Tyr Pro Pro
        915                 920                 925

Asp Glu Glu Glu Glu Asp Asp Ala Glu Cys Glu Glu Glu Glu Ile Asp
    930                 935                 940

Glu Thr Cys Glu His Glu Tyr Gly Thr Glu Asp Asp Tyr Gln Gly Leu
945                 950                 955                 960

Pro Leu Glu Phe Gly Ala Ser Ala Glu Thr Val Arg Val Glu Glu Glu
                965                 970                 975

Glu Glu Glu Asp Trp Leu Asp Asp Thr Thr Glu Gln Ser Glu Ile Glu
            980                 985                 990

Pro Glu Pro Glu Pro Thr Pro Glu  Glu Pro Val Asn Gln  Phe Thr Gly
        995                 1000                1005

Tyr Leu  Lys Leu Thr Asp Asn  Val Ala Ile Lys Cys  Val Asp Ile
    1010                1015                1020

Val Lys  Glu Ala Gln Ser Ala  Asn Pro Met Val Ile  Val Asn Ala
    1025                1030                1035

Ala Asn  Ile His Leu Lys His  Gly Gly Gly Val Ala  Gly Ala Leu
    1040                1045                1050

Asn Lys  Ala Thr Asn Gly Ala  Met Gln Lys Glu Ser  Asp Asp Tyr
    1055                1060                1065

Ile Lys  Leu Asn Gly Pro Leu  Thr Val Gly Gly Ser  Cys Leu Leu
    1070                1075                1080

Ser Gly  His Asn Leu Ala Lys  Lys Cys Leu His Val  Val Gly Pro
    1085                1090                1095

Asn Leu  Asn Ala Gly Glu Asp  Ile Gln Leu Leu Lys  Ala Ala Tyr
```

-continued

```
    1100                1105                1110

Glu Asn  Phe Asn Ser Gln Asp  Ile Leu Leu Ala Pro  Leu Leu Ser
    1115                1120                1125

Ala Gly  Ile Phe Gly Ala Lys  Pro Leu Gln Ser Leu  Gln Val Cys
    1130                1135                1140

Val Gln  Thr Val Arg Thr Gln  Val Tyr Ile Ala Val  Asn Asp Lys
    1145                1150                1155

Ala Leu  Tyr Glu Gln Val Val  Met Asp Tyr Leu Asp  Asn Leu Lys
    1160                1165                1170

Pro Arg  Val Glu Ala Pro Lys  Gln Glu Glu Pro Pro  Asn Thr Glu
    1175                1180                1185

Asp Ser  Lys Thr Glu Glu Lys  Ser Val Val Gln Lys  Pro Val Asp
    1190                1195                1200

Val Lys  Pro Lys Ile Lys Ala  Cys Ile Asp Glu Val  Thr Thr Thr
    1205                1210                1215

Leu Glu  Glu Thr Lys Phe Leu  Thr Asn Lys Leu Leu  Leu Phe Ala
    1220                1225                1230

Asp Ile  Asn Gly Lys Leu Tyr  His Asp Ser Gln Asn  Met Leu Arg
    1235                1240                1245

Gly Glu  Asp Met Ser Phe Leu  Glu Lys Asp Ala Pro  Tyr Met Val
    1250                1255                1260

Gly Asp  Val Ile Thr Ser Gly  Asp Ile Thr Cys Val  Val Ile Pro
    1265                1270                1275

Ser Lys  Lys Ala Gly Gly Thr  Thr Glu Met Leu Ser  Arg Ala Leu
    1280                1285                1290

Lys Lys  Val Pro Val Asp Glu  Tyr Ile Thr Thr Tyr  Pro Gly Gln
    1295                1300                1305

Gly Cys  Ala Gly Tyr Thr Leu  Glu Glu Ala Lys Thr  Ala Leu Lys
    1310                1315                1320

Lys Cys  Lys Ser Ala Phe Tyr  Val Leu Pro Ser Glu  Ala Pro Asn
    1325                1330                1335

Ala Lys  Glu Glu Ile Leu Gly  Thr Val Ser Trp Asn  Leu Arg Glu
    1340                1345                1350

Met Leu  Ala His Ala Glu Glu  Thr Arg Lys Leu Met  Pro Ile Cys
    1355                1360                1365

Met Asp  Val Arg Ala Ile Met  Ala Thr Ile Gln Arg  Lys Tyr Lys
    1370                1375                1380

Gly Ile  Lys Ile Gln Glu Gly  Ile Val Asp Tyr Gly  Val Arg Phe
    1385                1390                1395

Phe Phe  Tyr Thr Ser Lys Glu  Pro Val Ala Ser Ile  Ile Thr Lys
    1400                1405                1410

Leu Asn  Ser Leu Asn Glu Pro  Leu Val Thr Met Pro  Ile Gly Tyr
    1415                1420                1425

Val Thr  His Gly Phe Asn Leu  Glu Glu Ala Ala Arg  Cys Met Arg
    1430                1435                1440

Ser Leu  Lys Ala Pro Ala Val  Val Ser Val Ser Ser  Pro Asp Ala
    1445                1450                1455

Val Thr  Thr Tyr Asn Gly Tyr  Leu Thr Ser Ser Ser  Lys Thr Ser
    1460                1465                1470

Glu Glu  His Phe Val Glu Thr  Val Ser Leu Ala Gly  Ser Tyr Arg
    1475                1480                1485

Asp Trp  Ser Tyr Ser Gly Gln  Arg Thr Glu Leu Gly  Val Glu Phe
    1490                1495                1500
```

-continued

```
Leu Lys  Arg Gly Asp Lys Ile  Val Tyr His Thr Leu  Glu Ser Pro
    1505                1510                1515

Val Glu  Phe His Leu Asp Gly  Glu Val Leu Ser Leu  Asp Lys Leu
    1520                1525                1530

Lys Ser  Leu Leu Ser Leu Arg  Glu Val Lys Thr Ile  Lys Val Phe
    1535                1540                1545

Thr Thr  Val Asp Asn Thr Asn  Leu His Thr Gln Leu  Val Asp Met
    1550                1555                1560

Ser Met  Thr Tyr Gly Gln Gln  Phe Gly Pro Thr Tyr  Leu Asp Gly
    1565                1570                1575

Ala Asp  Val Thr Lys Ile Lys  Pro His Val Asn His  Glu Gly Lys
    1580                1585                1590

Thr Phe  Phe Val Leu Pro Ser  Asp Asp Thr Leu Arg  Ser Glu Ala
    1595                1600                1605

Phe Glu  Tyr Tyr His Thr Leu  Asp Glu Ser Phe Leu  Gly Arg Tyr
    1610                1615                1620

Met Ser  Ala Leu Asn His Thr  Lys Lys Trp Lys Phe  Pro Gln Val
    1625                1630                1635

Gly Gly  Leu Thr Ser Ile Lys  Trp Ala Asp Asn Asn  Cys Tyr Leu
    1640                1645                1650

Ser Ser  Val Leu Leu Ala Leu  Gln Gln Leu Glu Val  Lys Phe Asn
    1655                1660                1665

Ala Pro  Ala Leu Gln Glu Ala  Tyr Tyr Arg Ala Arg  Ala Gly Asp
    1670                1675                1680

Ala Ala  Asn Phe Cys Ala Leu  Ile Leu Ala Tyr Ser  Asn Lys Thr
    1685                1690                1695

Val Gly  Glu Leu Gly Asp Val  Arg Glu Thr Met Thr  His Leu Leu
    1700                1705                1710

Gln His  Ala Asn Leu Glu Ser  Ala Lys Arg Val Leu  Asn Val Val
    1715                1720                1725

Cys Lys  His Cys Gly Gln Lys  Thr Thr Thr Leu Thr  Gly Val Glu
    1730                1735                1740

Ala Val  Met Tyr Met Gly Thr  Leu Ser Tyr Asp Asn  Leu Lys Thr
    1745                1750                1755

Gly Val  Ser Ile Pro Cys Val  Cys Gly Arg Asp Ala  Thr Gln Tyr
    1760                1765                1770

Leu Val  Gln Gln Glu Ser Ser  Phe Val Met Met Ser  Ala Pro Pro
    1775                1780                1785

Ala Glu  Tyr Lys Leu Gln Gln  Gly Thr Phe Leu Cys  Ala Asn Glu
    1790                1795                1800

Tyr Thr  Gly Asn Tyr Gln Cys  Gly His Tyr Thr His  Ile Thr Ala
    1805                1810                1815

Lys Glu  Thr Leu Tyr Arg Ile  Asp Gly Ala His Leu  Thr Lys Met
    1820                1825                1830

Ser Glu  Tyr Lys Gly Pro Val  Thr Asp Val Phe Tyr  Lys Glu Thr
    1835                1840                1845

Ser Tyr  Thr Thr Thr Ile Lys  Pro Val Ser Tyr Lys  Leu Asp Gly
    1850                1855                1860

Val Thr  Tyr Thr Glu Ile Glu  Pro Lys Leu Asp Gly  Tyr Tyr Lys
    1865                1870                1875

Lys Asp  Asn Ala Tyr Tyr Thr  Glu Gln Pro Ile Asp  Leu Val Pro
    1880                1885                1890
```

-continued

```
Thr Gln  Pro Leu Pro Asn Ala  Ser Phe Asp Asn Phe  Lys Leu Thr
    1895                1900                1905

Cys Ser  Asn Thr Lys Phe Ala  Asp Asp Leu Asn Gln  Met Thr Gly
    1910                1915                1920

Phe Thr  Lys Pro Ala Ser Arg  Glu Leu Ser Val Thr  Phe Phe Pro
    1925                1930                1935

Asp Leu  Asn Gly Asp Val Val  Ala Ile Asp Tyr Arg  His Tyr Ser
    1940                1945                1950

Ala Ser  Phe Lys Lys Gly Ala  Lys Leu Leu His Lys  Pro Ile Val
    1955                1960                1965

Trp His  Ile Asn Gln Ala Thr  Thr Lys Thr Thr Phe  Lys Pro Asn
    1970                1975                1980

Thr Trp  Cys Leu Arg Cys Leu  Trp Ser Thr Lys Pro  Val Asp Thr
    1985                1990                1995

Ser Asn  Ser Phe Glu Val Leu  Ala Val Glu Asp Thr  Gln Gly Met
    2000                2005                2010

Asp Asn  Leu Ala Cys Glu Ser  Gln Gln Pro Thr Ser  Glu Glu Val
    2015                2020                2025

Val Glu  Asn Pro Thr Ile Gln  Lys Glu Val Ile Glu  Cys Asp Val
    2030                2035                2040

Lys Thr  Thr Glu Val Val Gly  Asn Val Ile Leu Lys  Pro Ser Asp
    2045                2050                2055

Glu Gly  Val Lys Val Thr Gln  Glu Leu Gly His Glu  Asp Leu Met
    2060                2065                2070

Ala Ala  Tyr Val Glu Asn Thr  Ser Ile Thr Ile Lys  Lys Pro Asn
    2075                2080                2085

Glu Leu  Ser Leu Ala Leu Gly  Leu Lys Thr Ile Ala  Thr His Gly
    2090                2095                2100

Ile Ala  Ala Ile Asn Ser Val  Pro Trp Ser Lys Ile  Leu Ala Tyr
    2105                2110                2115

Val Lys  Pro Phe Leu Gly Gln  Ala Ala Ile Thr Thr  Ser Asn Cys
    2120                2125                2130

Ala Lys  Arg Leu Ala Gln Arg  Val Phe Asn Asn Tyr  Met Pro Tyr
    2135                2140                2145

Val Phe  Thr Leu Leu Phe Gln  Leu Cys Thr Phe Thr  Lys Ser Thr
    2150                2155                2160

Asn Ser  Arg Ile Arg Ala Ser  Leu Pro Thr Thr Ile  Ala Lys Asn
    2165                2170                2175

Ser Val  Lys Ser Val Ala Lys  Leu Cys Leu Asp Ala  Gly Ile Asn
    2180                2185                2190

Tyr Val  Lys Ser Pro Lys Phe  Ser Lys Leu Phe Thr  Ile Ala Met
    2195                2200                2205

Trp Leu  Leu Leu Leu Ser Ile  Cys Leu Gly Ser Leu  Ile Cys Val
    2210                2215                2220

Thr Ala  Ala Phe Gly Val Leu  Leu Ser Asn Phe Gly  Ala Pro Ser
    2225                2230                2235

Tyr Cys  Asn Gly Val Arg Glu  Leu Tyr Leu Asn Ser  Ser Asn Val
    2240                2245                2250

Thr Thr  Met Asp Phe Cys Glu  Gly Ser Phe Pro Cys  Ser Ile Cys
    2255                2260                2265

Leu Ser  Gly Leu Asp Ser Leu  Asp Ser Tyr Pro Ala  Leu Glu Thr
    2270                2275                2280

Ile Gln  Val Thr Ile Ser Ser  Tyr Lys Leu Asp Leu  Thr Ile Leu
```

-continued

```
    2285                2290                2295

Gly Leu  Ala Ala Glu Trp Val  Leu Ala Tyr Met Leu  Phe Thr Lys
    2300                2305                2310

Phe Phe  Tyr Leu Leu Gly Leu  Ser Ala Ile Met Gln  Val Phe Phe
    2315                2320                2325

Gly Tyr  Phe Ala Ser His Phe  Ile Ser Asn Ser Trp  Leu Met Trp
    2330                2335                2340

Phe Ile  Ile Ser Ile Val Gln  Met Ala Pro Val Ser  Ala Met Val
    2345                2350                2355

Arg Met  Tyr Ile Phe Phe Ala  Ser Phe Tyr Tyr Ile  Trp Lys Ser
    2360                2365                2370

Tyr Val  His Ile Met Asp Gly  Cys Thr Ser Ser Thr  Cys Met Met
    2375                2380                2385

Cys Tyr  Lys Arg Asn Arg Ala  Thr Arg Val Glu Cys  Thr Thr Ile
    2390                2395                2400

Val Asn  Gly Met Lys Arg Ser  Phe Tyr Val Tyr Ala  Asn Gly Gly
    2405                2410                2415

Arg Gly  Phe Cys Lys Thr His  Asn Trp Asn Cys Leu  Asn Cys Asp
    2420                2425                2430

Thr Phe  Cys Thr Gly Ser Thr  Phe Ile Ser Asp Glu  Val Ala Arg
    2435                2440                2445

Asp Leu  Ser Leu Gln Phe Lys  Arg Pro Ile Asn Pro  Thr Asp Gln
    2450                2455                2460

Ser Ser  Tyr Ile Val Asp Ser  Val Ala Val Lys Asn  Gly Ala Leu
    2465                2470                2475

His Leu  Tyr Phe Asp Lys Ala  Gly Gln Lys Thr Tyr  Glu Arg His
    2480                2485                2490

Pro Leu  Ser His Phe Val Asn  Leu Asp Asn Leu Arg  Ala Asn Asn
    2495                2500                2505

Thr Lys  Gly Ser Leu Pro Ile  Asn Val Ile Val Phe  Asp Gly Lys
    2510                2515                2520

Ser Lys  Cys Asp Glu Ser Ala  Ser Lys Ser Ala Ser  Val Tyr Tyr
    2525                2530                2535

Ser Gln  Leu Met Cys Gln Pro  Ile Leu Leu Leu Asp  Gln Val Leu
    2540                2545                2550

Val Ser  Asp Val Gly Asp Ser  Thr Glu Val Ser Val  Lys Met Phe
    2555                2560                2565

Asp Ala  Tyr Val Asp Thr Phe  Ser Ala Thr Phe Ser  Val Pro Met
    2570                2575                2580

Glu Lys  Leu Lys Ala Leu Val  Ala Thr Ala His Ser  Glu Leu Ala
    2585                2590                2595

Lys Gly  Val Ala Leu Asp Gly  Val Leu Ser Thr Phe  Val Ser Ala
    2600                2605                2610

Ala Arg  Gln Gly Val Val Asp  Thr Asp Val Asp Thr  Lys Asp Val
    2615                2620                2625

Ile Glu  Cys Leu Lys Leu Ser  His His Ser Asp Leu  Glu Val Thr
    2630                2635                2640

Gly Asp  Ser Cys Asn Asn Phe  Met Leu Thr Tyr Asn  Lys Val Glu
    2645                2650                2655

Asn Met  Thr Pro Arg Asp Leu  Gly Ala Cys Ile Asp  Cys Asn Ala
    2660                2665                2670

Arg His  Ile Asn Ala Gln Val  Ala Lys Ser His Asn  Val Ser Leu
    2675                2680                2685
```

-continued

```
Ile Trp  Asn Val Lys Asp Tyr  Met Ser Leu Ser Glu  Gln Leu Arg
    2690                 2695                 2700

Lys Gln  Ile Arg Ser Ala Ala  Lys Lys Asn Asn Ile  Pro Phe Arg
    2705                 2710                 2715

Leu Thr  Cys Ala Thr Thr Arg  Gln Val Val Asn Val  Ile Thr Thr
    2720                 2725                 2730

Lys Ile  Ser Leu Lys Gly Gly  Lys Ile Val Ser Thr  Cys Phe Lys
    2735                 2740                 2745

Leu Met  Leu Lys Ala Thr Leu  Leu Cys Val Leu Ala  Ala Leu Val
    2750                 2755                 2760

Cys Tyr  Ile Val Met Pro Val  His Thr Leu Ser Ile  His Asp Gly
    2765                 2770                 2775

Tyr Thr  Asn Glu Ile Ile Gly  Tyr Lys Ala Ile Gln  Asp Gly Val
    2780                 2785                 2790

Thr Arg  Asp Ile Ile Ser Thr  Asp Asp Cys Phe Ala  Asn Lys His
    2795                 2800                 2805

Ala Gly  Phe Asp Ala Trp Phe  Ser Gln Arg Gly Gly  Ser Tyr Lys
    2810                 2815                 2820

Asn Asp  Lys Ser Cys Pro Val  Val Ala Ala Ile Ile  Thr Arg Glu
    2825                 2830                 2835

Ile Gly  Phe Ile Val Pro Gly  Leu Pro Gly Thr Val  Leu Arg Ala
    2840                 2845                 2850

Ile Asn  Gly Asp Phe Leu His  Phe Leu Pro Arg Val  Phe Ser Ala
    2855                 2860                 2865

Val Gly  Asn Ile Cys Tyr Thr  Pro Ser Lys Leu Ile  Glu Tyr Ser
    2870                 2875                 2880

Asp Phe  Ala Thr Ser Ala Cys  Val Leu Ala Ala Glu  Cys Thr Ile
    2885                 2890                 2895

Phe Lys  Asp Ala Met Gly Lys  Pro Val Pro Tyr Cys  Tyr Asp Thr
    2900                 2905                 2910

Asn Leu  Leu Glu Gly Ser Ile  Ser Tyr Ser Glu Leu  Arg Pro Asp
    2915                 2920                 2925

Thr Arg  Tyr Val Leu Met Asp  Gly Ser Ile Ile Gln  Phe Pro Asn
    2930                 2935                 2940

Thr Tyr  Leu Glu Gly Ser Val  Arg Val Val Thr Thr  Phe Asp Ala
    2945                 2950                 2955

Glu Tyr  Cys Arg His Gly Thr  Cys Glu Arg Ser Glu  Val Gly Ile
    2960                 2965                 2970

Cys Leu  Ser Thr Ser Gly Arg  Trp Val Leu Asn Asn  Glu His Tyr
    2975                 2980                 2985

Arg Ala  Leu Ser Gly Val Phe  Cys Gly Val Asp Ala  Met Asn Leu
    2990                 2995                 3000

Ile Ala  Asn Ile Phe Thr Pro  Leu Val Gln Pro Val  Gly Ala Leu
    3005                 3010                 3015

Asp Val  Ser Ala Ser Val Val  Ala Gly Gly Ile Ile  Ala Ile Leu
    3020                 3025                 3030

Val Thr  Cys Ala Ala Tyr Tyr  Phe Met Lys Phe Arg  Arg Val Phe
    3035                 3040                 3045

Gly Glu  Tyr Asn His Val Val  Ala Ala Asn Ala Leu  Leu Phe Leu
    3050                 3055                 3060

Met Ser  Phe Thr Ile Leu Cys  Leu Val Pro Ala Tyr  Ser Phe Leu
    3065                 3070                 3075
```

-continued

```
Pro Gly  Val Tyr Ser Val Phe  Tyr Leu Tyr Leu Thr  Phe Tyr Phe
    3080                3085                3090

Thr Asn  Asp Val Ser Phe Leu  Ala His Leu Gln Trp  Phe Ala Met
    3095                3100                3105

Phe Ser  Pro Ile Val Pro Phe  Trp Ile Thr Ala Ile  Tyr Val Phe
    3110                3115                3120

Cys Ile  Ser Leu Lys His Cys  His Trp Phe Phe Asn  Asn Tyr Leu
    3125                3130                3135

Arg Lys  Arg Val Met Phe Asn  Gly Val Thr Phe Ser  Thr Phe Glu
    3140                3145                3150

Glu Ala  Ala Leu Cys Thr Phe  Leu Leu Asn Lys Glu  Met Tyr Leu
    3155                3160                3165

Lys Leu  Arg Ser Glu Thr Leu  Leu Pro Leu Thr Gln  Tyr Asn Arg
    3170                3175                3180

Tyr Leu  Ala Leu Tyr Asn Lys  Tyr Lys Tyr Phe Ser  Gly Ala Leu
    3185                3190                3195

Asp Thr  Thr Ser Tyr Arg Glu  Ala Ala Cys Cys His  Leu Ala Lys
    3200                3205                3210

Ala Leu  Asn Asp Phe Ser Asn  Ser Gly Ala Asp Val  Leu Tyr Gln
    3215                3220                3225

Pro Pro  Gln Thr Ser Ile Thr  Ser Ala Val Leu Gln  Ser Gly Phe
    3230                3235                3240

Arg Lys  Met Ala Phe Pro Ser  Gly Lys Val Glu Gly  Cys Met Val
    3245                3250                3255

Gln Val  Thr Cys Gly Thr Thr  Thr Leu Asn Gly Leu  Trp Leu Asp
    3260                3265                3270

Asp Thr  Val Tyr Cys Pro Arg  His Val Ile Cys Thr  Ala Glu Asp
    3275                3280                3285

Met Leu  Asn Pro Asn Tyr Glu  Asp Leu Leu Ile Arg  Lys Ser Asn
    3290                3295                3300

His Ser  Phe Leu Val Gln Ala  Gly Asn Val Gln Leu  Arg Val Ile
    3305                3310                3315

Gly His  Ser Met Gln Asn Cys  Leu Leu Arg Leu Lys  Val Asp Thr
    3320                3325                3330

Ser Asn  Pro Lys Thr Pro Lys  Tyr Lys Phe Val Arg  Ile Gln Pro
    3335                3340                3345

Gly Gln  Thr Phe Ser Val Leu  Ala Cys Tyr Asn Gly  Ser Pro Ser
    3350                3355                3360

Gly Val  Tyr Gln Cys Ala Met  Arg Pro Asn His Thr  Ile Lys Gly
    3365                3370                3375

Ser Phe  Leu Asn Gly Ser Cys  Gly Ser Val Gly Phe  Asn Ile Asp
    3380                3385                3390

Tyr Asp  Cys Val Ser Phe Cys  Tyr Met His His Met  Glu Leu Pro
    3395                3400                3405

Thr Gly  Val His Ala Gly Thr  Asp Leu Glu Gly Lys  Phe Tyr Gly
    3410                3415                3420

Pro Phe  Val Asp Arg Gln Thr  Ala Gln Ala Ala Gly  Thr Asp Thr
    3425                3430                3435

Thr Ile  Thr Leu Asn Val Leu  Ala Trp Leu Tyr Ala  Ala Val Ile
    3440                3445                3450

Asn Gly  Asp Arg Trp Phe Leu  Asn Arg Phe Thr Thr  Thr Leu Asn
    3455                3460                3465

Asp Phe  Asn Leu Val Ala Met  Lys Tyr Asn Tyr Glu  Pro Leu Thr
```

-continued

```
    3470                3475                3480

Gln Asp  His Val Asp Ile Leu  Gly Pro Leu Ser Ala  Gln Thr Gly
    3485                3490                3495

Ile Ala  Val Leu Asp Met Cys  Ala Ala Leu Lys Glu  Leu Leu Gln
    3500                3505                3510

Asn Gly  Met Asn Gly Arg Thr  Ile Leu Gly Ser Thr  Ile Leu Glu
    3515                3520                3525

Asp Glu  Phe Thr Pro Phe Asp  Val Val Arg Gln Cys  Ser Gly Val
    3530                3535                3540

Thr Phe  Gln Gly Lys Phe Lys  Lys Ile Val Lys Gly  Thr His His
    3545                3550                3555

Trp Met  Leu Leu Thr Phe Leu  Thr Ser Leu Leu Ile  Leu Val Gln
    3560                3565                3570

Ser Thr  Gln Trp Ser Leu Phe  Phe Phe Val Tyr Glu  Asn Ala Phe
    3575                3580                3585

Leu Pro  Phe Thr Leu Gly Ile  Met Ala Ile Ala Ala  Cys Ala Met
    3590                3595                3600

Leu Leu  Val Lys His Lys His  Ala Phe Leu Cys Leu  Phe Leu Leu
    3605                3610                3615

Pro Ser  Leu Ala Thr Val Ala  Tyr Phe Asn Met Val  Tyr Met Pro
    3620                3625                3630

Ala Ser  Trp Val Met Arg Ile  Met Thr Trp Leu Glu  Leu Ala Asp
    3635                3640                3645

Thr Ser  Leu Ser Gly Tyr Arg  Leu Lys Asp Cys Val  Met Tyr Ala
    3650                3655                3660

Ser Ala  Leu Val Leu Leu Ile  Leu Met Thr Ala Arg  Thr Val Tyr
    3665                3670                3675

Asp Asp  Ala Ala Arg Arg Val  Trp Thr Leu Met Asn  Val Ile Thr
    3680                3685                3690

Leu Val  Tyr Lys Val Tyr Tyr  Gly Asn Ala Leu Asp  Gln Ala Ile
    3695                3700                3705

Ser Met  Trp Ala Leu Val Ile  Ser Val Thr Ser Asn  Tyr Ser Gly
    3710                3715                3720

Val Val  Thr Thr Ile Met Phe  Leu Ala Arg Ala Ile  Val Phe Val
    3725                3730                3735

Cys Val  Glu Tyr Tyr Pro Leu  Leu Phe Ile Thr Gly  Asn Thr Leu
    3740                3745                3750

Gln Cys  Ile Met Leu Val Tyr  Cys Phe Leu Gly Tyr  Cys Cys Cys
    3755                3760                3765

Cys Tyr  Phe Gly Leu Phe Cys  Leu Leu Asn Arg Tyr  Phe Arg Leu
    3770                3775                3780

Thr Leu  Gly Val Tyr Asp Tyr  Leu Val Ser Thr Gln  Glu Phe Arg
    3785                3790                3795

Tyr Met  Asn Ser Gln Gly Leu  Leu Pro Pro Lys Ser  Ser Ile Asp
    3800                3805                3810

Ala Phe  Lys Leu Asn Ile Lys  Leu Leu Gly Ile Gly  Gly Lys Pro
    3815                3820                3825

Cys Ile  Lys Val Ala Thr Val  Gln Ser Lys Met Ser  Asp Val Lys
    3830                3835                3840

Cys Thr  Ser Val Val Leu Leu  Ser Val Leu Gln Gln  Leu Arg Val
    3845                3850                3855

Glu Ser  Ser Ser Lys Leu Trp  Ala Gln Cys Val Gln  Leu His Asn
    3860                3865                3870
```

-continued

```
Asp Ile  Leu Leu Ala Lys Asp  Thr Thr Glu Ala Phe  Glu Lys Met
    3875                 3880                 3885

Val Ser  Leu Leu Ser Val Leu  Leu Ser Met Gln Gly  Ala Val Asp
    3890                 3895                 3900

Ile Asn  Arg Leu Cys Glu Glu  Met Leu Asp Asn Arg  Ala Thr Leu
    3905                 3910                 3915

Gln Ala  Ile Ala Ser Glu Phe  Ser Ser Leu Pro Ser  Tyr Ala Ala
    3920                 3925                 3930

Tyr Ala  Thr Ala Gln Glu Ala  Tyr Glu Gln Ala Val  Ala Asn Gly
    3935                 3940                 3945

Asp Ser  Glu Val Val Leu Lys  Lys Leu Lys Lys Ser  Leu Asn Val
    3950                 3955                 3960

Ala Lys  Ser Glu Phe Asp Arg  Asp Ala Ala Met Gln  Arg Lys Leu
    3965                 3970                 3975

Glu Lys  Met Ala Asp Gln Ala  Met Thr Gln Met Tyr  Lys Gln Ala
    3980                 3985                 3990

Arg Ser  Glu Asp Lys Arg Ala  Lys Val Thr Ser Ala  Met Gln Thr
    3995                 4000                 4005

Met Leu  Phe Thr Met Leu Arg  Lys Leu Asp Asn Asp  Ala Leu Asn
    4010                 4015                 4020

Asn Ile  Ile Asn Asn Ala Arg  Asp Gly Cys Val Pro  Leu Asn Ile
    4025                 4030                 4035

Ile Pro  Leu Thr Thr Ala Ala  Lys Leu Met Val Val  Val Pro Asp
    4040                 4045                 4050

Tyr Gly  Thr Tyr Lys Asn Thr  Cys Asp Gly Asn Thr  Phe Thr Tyr
    4055                 4060                 4065

Ala Ser  Ala Leu Trp Glu Ile  Gln Gln Val Val Asp  Ala Asp Ser
    4070                 4075                 4080

Lys Ile  Val Gln Leu Ser Glu  Ile Asn Met Asp Asn  Ser Pro Asn
    4085                 4090                 4095

Leu Ala  Trp Pro Leu Ile Val  Thr Ala Leu Arg Ala  Asn Ser Ala
    4100                 4105                 4110

Val Lys  Leu Gln Asn Asn Glu  Leu Ser Pro Val Ala  Leu Arg Gln
    4115                 4120                 4125

Met Ser  Cys Ala Ala Gly Thr  Thr Gln Thr Ala Cys  Thr Asp Asp
    4130                 4135                 4140

Asn Ala  Leu Ala Tyr Tyr Asn  Asn Ser Lys Gly Gly  Arg Phe Val
    4145                 4150                 4155

Leu Ala  Leu Leu Ser Asp His  Gln Asp Leu Lys Trp  Ala Arg Phe
    4160                 4165                 4170

Pro Lys  Ser Asp Gly Thr Gly  Thr Ile Tyr Thr Glu  Leu Glu Pro
    4175                 4180                 4185

Pro Cys  Arg Phe Val Thr Asp  Thr Pro Lys Gly Pro  Lys Val Lys
    4190                 4195                 4200

Tyr Leu  Tyr Phe Ile Lys Gly  Leu Asn Asn Leu Asn  Arg Gly Met
    4205                 4210                 4215

Val Leu  Gly Ser Leu Ala Ala  Thr Val Arg Leu Gln  Ala Gly Asn
    4220                 4225                 4230

Ala Thr  Glu Val Pro Ala Asn  Ser Thr Val Leu Ser  Phe Cys Ala
    4235                 4240                 4245

Phe Ala  Val Asp Pro Ala Lys  Ala Tyr Lys Asp Tyr  Leu Ala Ser
    4250                 4255                 4260
```

-continued

```
Gly Gly  Gln Pro Ile Thr Asn  Cys Val Lys Met Leu  Cys Thr His
    4265                 4270                 4275

Thr Gly  Thr Gly Gln Ala Ile  Thr Val Thr Pro Glu  Ala Asn Met
    4280                 4285                 4290

Asp Gln  Glu Ser Phe Gly Gly  Ala Ser Cys Cys Leu  Tyr Cys Arg
    4295                 4300                 4305

Cys His  Ile Asp His Pro Asn  Pro Lys Gly Phe Cys  Asp Leu Lys
    4310                 4315                 4320

Gly Lys  Tyr Val Gln Ile Pro  Thr Thr Cys Ala Asn  Asp Pro Val
    4325                 4330                 4335

Gly Phe  Thr Leu Arg Asn Thr  Val Cys Thr Val Cys  Gly Met Trp
    4340                 4345                 4350

Lys Gly  Tyr Gly Cys Ser Cys  Asp Gln Leu Arg Glu  Pro Leu Met
    4355                 4360                 4365

Gln Ser  Ala Asp Ala Ser Thr  Phe Leu Asn Gly Phe  Ala Val
    4370                 4375                 4380


<210> SEQ ID NO 3
<211> LENGTH: 2695
<212> TYPE: PRT
<213> ORGANISM: Coronavirus

<400> SEQUENCE: 3

Arg Val Cys Gly Val Ser Ala Ala Arg Leu Thr Pro Cys Gly Thr Gly
1               5                   10                  15

Thr Ser Thr Asp Val Val Tyr Arg Ala Phe Asp Ile Tyr Asn Glu Lys
            20                  25                  30

Val Ala Gly Phe Ala Lys Phe Leu Lys Thr Asn Cys Cys Arg Phe Gln
        35                  40                  45

Glu Lys Asp Glu Glu Gly Asn Leu Leu Asp Ser Tyr Phe Val Val Lys
    50                  55                  60

Arg His Thr Met Ser Asn Tyr Gln His Glu Glu Thr Ile Tyr Asn Leu
65                  70                  75                  80

Val Lys Asp Cys Pro Ala Val Ala Val His Asp Phe Phe Lys Phe Arg
                85                  90                  95

Val Asp Gly Asp Met Val Pro His Ile Ser Arg Gln Arg Leu Thr Lys
            100                 105                 110

Tyr Thr Met Ala Asp Leu Val Tyr Ala Leu Arg His Phe Asp Glu Gly
        115                 120                 125

Asn Cys Asp Thr Leu Lys Glu Ile Leu Val Thr Tyr Asn Cys Cys Asp
    130                 135                 140

Asp Asp Tyr Phe Asn Lys Lys Asp Trp Tyr Asp Phe Val Glu Asn Pro
145                 150                 155                 160

Asp Ile Leu Arg Val Tyr Ala Asn Leu Gly Glu Arg Val Arg Gln Ser
                165                 170                 175

Leu Leu Lys Thr Val Gln Phe Cys Asp Ala Met Arg Asp Ala Gly Ile
            180                 185                 190

Val Gly Val Leu Thr Leu Asp Asn Gln Asp Leu Asn Gly Asn Trp Tyr
        195                 200                 205

Asp Phe Gly Asp Phe Val Gln Val Ala Pro Gly Cys Gly Val Pro Ile
    210                 215                 220

Val Asp Ser Tyr Tyr Ser Leu Leu Met Pro Ile Leu Thr Leu Thr Arg
225                 230                 235                 240

Ala Leu Ala Ala Glu Ser His Met Asp Ala Asp Leu Ala Lys Pro Leu
                245                 250                 255
```

-continued

```
Ile Lys Trp Asp Leu Leu Lys Tyr Asp Phe Thr Glu Glu Arg Leu Cys
            260                 265                 270

Leu Phe Asp Arg Tyr Phe Lys Tyr Trp Asp Gln Thr Tyr His Pro Asn
        275                 280                 285

Cys Ile Asn Cys Leu Asp Asp Arg Cys Ile Leu His Cys Ala Asn Phe
    290                 295                 300

Asn Val Leu Phe Ser Thr Val Phe Pro Pro Thr Ser Phe Gly Pro Leu
305                 310                 315                 320

Val Arg Lys Ile Phe Val Asp Gly Val Pro Phe Val Val Ser Thr Gly
                325                 330                 335

Tyr His Phe Arg Glu Leu Gly Val Val His Asn Gln Asp Val Asn Leu
            340                 345                 350

His Ser Ser Arg Leu Ser Phe Lys Glu Leu Leu Val Tyr Ala Ala Asp
        355                 360                 365

Pro Ala Met His Ala Ala Ser Gly Asn Leu Leu Leu Asp Lys Arg Thr
    370                 375                 380

Thr Cys Phe Ser Val Ala Ala Leu Thr Asn Asn Val Ala Phe Gln Thr
385                 390                 395                 400

Val Lys Pro Gly Asn Phe Asn Lys Asp Phe Tyr Asp Phe Ala Val Ser
                405                 410                 415

Lys Gly Phe Phe Lys Glu Gly Ser Ser Val Glu Leu Lys His Phe Phe
            420                 425                 430

Phe Ala Gln Asp Gly Asn Ala Ala Ile Ser Asp Tyr Asp Tyr Tyr Arg
        435                 440                 445

Tyr Asn Leu Pro Thr Met Cys Asp Ile Arg Gln Leu Leu Phe Val Val
    450                 455                 460

Glu Val Val Asp Lys Tyr Phe Asp Cys Tyr Asp Gly Gly Cys Ile Asn
465                 470                 475                 480

Ala Asn Gln Val Ile Val Asn Asn Leu Asp Lys Ser Ala Gly Phe Pro
                485                 490                 495

Phe Asn Lys Trp Gly Lys Ala Arg Leu Tyr Tyr Asp Ser Met Ser Tyr
            500                 505                 510

Glu Asp Gln Asp Ala Leu Phe Ala Tyr Thr Lys Arg Asn Val Ile Pro
        515                 520                 525

Thr Ile Thr Gln Met Asn Leu Lys Tyr Ala Ile Ser Ala Lys Asn Arg
    530                 535                 540

Ala Arg Thr Val Ala Gly Val Ser Ile Cys Ser Thr Met Thr Asn Arg
545                 550                 555                 560

Gln Phe His Gln Lys Leu Leu Lys Ser Ile Ala Ala Thr Arg Gly Ala
                565                 570                 575

Thr Val Val Ile Gly Thr Ser Lys Phe Tyr Gly Gly Trp His Asn Met
            580                 585                 590

Leu Lys Thr Val Tyr Ser Asp Val Glu Thr Pro His Leu Met Gly Trp
        595                 600                 605

Asp Tyr Pro Lys Cys Asp Arg Ala Met Pro Asn Met Leu Arg Ile Met
    610                 615                 620

Ala Ser Leu Val Leu Ala Arg Lys His Asn Thr Cys Cys Asn Leu Ser
625                 630                 635                 640

His Arg Phe Tyr Arg Leu Ala Asn Glu Cys Ala Gln Val Leu Ser Glu
                645                 650                 655

Met Val Met Cys Gly Gly Ser Leu Tyr Val Lys Pro Gly Gly Thr Ser
            660                 665                 670
```

-continued

```
Ser Gly Asp Ala Thr Thr Ala Tyr Ala Asn Ser Val Phe Asn Ile Cys
        675                 680                 685

Gln Ala Val Thr Ala Asn Val Asn Ala Leu Leu Ser Thr Asp Gly Asn
    690                 695                 700

Lys Ile Ala Asp Lys Tyr Val Arg Asn Leu Gln His Arg Leu Tyr Glu
705                 710                 715                 720

Cys Leu Tyr Arg Asn Arg Asp Val Asp His Glu Phe Val Asp Glu Phe
                725                 730                 735

Tyr Ala Tyr Leu Arg Lys His Phe Ser Met Met Ile Leu Ser Asp Asp
            740                 745                 750

Ala Val Val Cys Tyr Asn Ser Asn Tyr Ala Ala Gln Gly Leu Val Ala
        755                 760                 765

Ser Ile Lys Asn Phe Lys Ala Val Leu Tyr Tyr Gln Asn Asn Val Phe
    770                 775                 780

Met Ser Glu Ala Lys Cys Trp Thr Glu Thr Asp Leu Thr Lys Gly Pro
785                 790                 795                 800

His Glu Phe Cys Ser Gln His Thr Met Leu Val Lys Gln Gly Asp Asp
                805                 810                 815

Tyr Val Tyr Leu Pro Tyr Pro Asp Pro Ser Arg Ile Leu Gly Ala Gly
            820                 825                 830

Cys Phe Val Asp Asp Ile Val Lys Thr Asp Gly Thr Leu Met Ile Glu
        835                 840                 845

Arg Phe Val Ser Leu Ala Ile Asp Ala Tyr Pro Leu Thr Lys His Pro
    850                 855                 860

Asn Gln Glu Tyr Ala Asp Val Phe His Leu Tyr Leu Gln Tyr Ile Arg
865                 870                 875                 880

Lys Leu His Asp Glu Leu Thr Gly His Met Leu Asp Met Tyr Ser Val
                885                 890                 895

Met Leu Thr Asn Asp Asn Thr Ser Arg Tyr Trp Glu Pro Glu Phe Tyr
            900                 905                 910

Glu Ala Met Tyr Thr Pro His Thr Val Leu Gln Ala Val Gly Ala Cys
        915                 920                 925

Val Leu Cys Asn Ser Gln Thr Ser Leu Arg Cys Gly Ala Cys Ile Arg
    930                 935                 940

Arg Pro Phe Leu Cys Cys Lys Cys Cys Tyr Asp His Val Ile Ser Thr
945                 950                 955                 960

Ser His Lys Leu Val Leu Ser Val Asn Pro Tyr Val Cys Asn Ala Pro
                965                 970                 975

Gly Cys Asp Val Thr Asp Val Thr Gln Leu Tyr Leu Gly Gly Met Ser
            980                 985                 990

Tyr Tyr Cys Lys Ser His Lys Pro  Pro Ile Ser Phe Pro  Leu Cys Ala
        995                 1000                1005

Asn Gly  Gln Val Phe Gly Leu  Tyr Lys Asn Thr Cys  Val Gly Ser
    1010                1015                1020

Asp Asn  Val Thr Asp Phe Asn  Ala Ile Ala Thr Cys  Asp Trp Thr
    1025                1030                1035

Asn Ala  Gly Asp Tyr Ile Leu  Ala Asn Thr Cys Thr  Glu Arg Leu
    1040                1045                1050

Lys Leu  Phe Ala Ala Glu Thr  Leu Lys Ala Thr Glu  Glu Thr Phe
    1055                1060                1065

Lys Leu  Ser Tyr Gly Ile Ala  Thr Val Arg Glu Val  Leu Ser Asp
    1070                1075                1080

Arg Glu  Leu His Leu Ser Trp  Glu Val Gly Lys Pro  Arg Pro Pro
```

-continued

```
    1085                1090                1095

Leu Asn  Arg Asn Tyr Val Phe  Thr Gly Tyr Arg Val  Thr Lys Asn
    1100                1105                1110

Ser Lys  Val Gln Ile Gly Glu  Tyr Thr Phe Glu Lys  Gly Asp Tyr
    1115                1120                1125

Gly Asp  Ala Val Val Tyr Arg  Gly Thr Thr Thr Tyr  Lys Leu Asn
    1130                1135                1140

Val Gly  Asp Tyr Phe Val Leu  Thr Ser His Thr Val  Met Pro Leu
    1145                1150                1155

Ser Ala  Pro Thr Leu Val Pro  Gln Glu His Tyr Val  Arg Ile Thr
    1160                1165                1170

Gly Leu  Tyr Pro Thr Leu Asn  Ile Ser Asp Glu Phe  Ser Ser Asn
    1175                1180                1185

Val Ala  Asn Tyr Gln Lys Val  Gly Met Gln Lys Tyr  Ser Thr Leu
    1190                1195                1200

Gln Gly  Pro Pro Gly Thr Gly  Lys Ser His Phe Ala  Ile Gly Leu
    1205                1210                1215

Ala Leu  Tyr Tyr Pro Ser Ala  Arg Ile Val Tyr Thr  Ala Cys Ser
    1220                1225                1230

His Ala  Ala Val Asp Ala Leu  Cys Glu Lys Ala Leu  Lys Tyr Leu
    1235                1240                1245

Pro Ile  Asp Lys Cys Ser Arg  Ile Ile Pro Ala Arg  Ala Arg Val
    1250                1255                1260

Glu Cys  Phe Asp Lys Phe Lys  Val Asn Ser Thr Leu  Glu Gln Tyr
    1265                1270                1275

Val Phe  Cys Thr Val Asn Ala  Leu Pro Glu Thr Thr  Ala Asp Ile
    1280                1285                1290

Val Val  Phe Asp Glu Ile Ser  Met Ala Thr Asn Tyr  Asp Leu Ser
    1295                1300                1305

Val Val  Asn Ala Arg Leu Arg  Ala Lys His Tyr Val  Tyr Ile Gly
    1310                1315                1320

Asp Pro  Ala Gln Leu Pro Ala  Pro Arg Thr Leu Leu  Thr Lys Gly
    1325                1330                1335

Thr Leu  Glu Pro Glu Tyr Phe  Asn Ser Val Cys Arg  Leu Met Lys
    1340                1345                1350

Thr Ile  Gly Pro Asp Met Phe  Leu Gly Thr Cys Arg  Arg Cys Pro
    1355                1360                1365

Ala Glu  Ile Val Asp Thr Val  Ser Ala Leu Val Tyr  Asp Asn Lys
    1370                1375                1380

Leu Lys  Ala His Lys Asp Lys  Ser Ala Gln Cys Phe  Lys Met Phe
    1385                1390                1395

Tyr Lys  Gly Val Ile Thr His  Asp Val Ser Ser Ala  Ile Asn Arg
    1400                1405                1410

Pro Gln  Ile Gly Val Val Arg  Glu Phe Leu Thr Arg  Asn Pro Ala
    1415                1420                1425

Trp Arg  Lys Ala Val Phe Ile  Ser Pro Tyr Asn Ser  Gln Asn Ala
    1430                1435                1440

Val Ala  Ser Lys Ile Leu Gly  Leu Pro Thr Gln Thr  Val Asp Ser
    1445                1450                1455

Ser Gln  Gly Ser Glu Tyr Asp  Tyr Val Ile Phe Thr  Gln Thr Thr
    1460                1465                1470

Glu Thr  Ala His Ser Cys Asn  Val Asn Arg Phe Asn  Val Ala Ile
    1475                1480                1485
```

-continued

```
Thr Arg  Ala Lys Ile Gly Ile  Leu Cys Ile Met Ser  Asp Arg Asp
    1490                1495                1500

Leu Tyr  Asp Lys Leu Gln Phe  Thr Ser Leu Glu Ile  Pro Arg Arg
    1505                1510                1515

Asn Val  Ala Thr Leu Gln Ala  Glu Asn Val Thr Gly  Leu Phe Lys
    1520                1525                1530

Asp Cys  Ser Lys Ile Ile Thr  Gly Leu His Pro Thr  Gln Ala Pro
    1535                1540                1545

Thr His  Leu Ser Val Asp Ile  Lys Phe Lys Thr Glu  Gly Leu Cys
    1550                1555                1560

Val Asp  Ile Pro Gly Ile Pro  Lys Asp Met Thr Tyr  Arg Arg Leu
    1565                1570                1575

Ile Ser  Met Met Gly Phe Lys  Met Asn Tyr Gln Val  Asn Gly Tyr
    1580                1585                1590

Pro Asn  Met Phe Ile Thr Arg  Glu Glu Ala Ile Arg  His Val Arg
    1595                1600                1605

Ala Trp  Ile Gly Phe Asp Val  Glu Gly Cys His Ala  Thr Arg Asp
    1610                1615                1620

Ala Val  Gly Thr Asn Leu Pro  Leu Gln Leu Gly Phe  Ser Thr Gly
    1625                1630                1635

Val Asn  Leu Val Ala Val Pro  Thr Gly Tyr Val Asp  Thr Glu Asn
    1640                1645                1650

Asn Thr  Glu Phe Thr Arg Val  Asn Ala Lys Pro Pro  Pro Gly Asp
    1655                1660                1665

Gln Phe  Lys His Leu Ile Pro  Leu Met Tyr Lys Gly  Leu Pro Trp
    1670                1675                1680

Asn Val  Val Arg Ile Lys Ile  Val Gln Met Leu Ser  Asp Thr Leu
    1685                1690                1695

Lys Gly  Leu Ser Asp Arg Val  Val Phe Val Leu Trp  Ala His Gly
    1700                1705                1710

Phe Glu  Leu Thr Ser Met Lys  Tyr Phe Val Lys Ile  Gly Pro Glu
    1715                1720                1725

Arg Thr  Cys Cys Leu Cys Asp  Lys Arg Ala Thr Cys  Phe Ser Thr
    1730                1735                1740

Ser Ser  Asp Thr Tyr Ala Cys  Trp Asn His Ser Val  Gly Phe Asp
    1745                1750                1755

Tyr Val  Tyr Asn Pro Phe Met  Ile Asp Val Gln Gln  Trp Gly Phe
    1760                1765                1770

Thr Gly  Asn Leu Gln Ser Asn  His Asp Gln His Cys  Gln Val His
    1775                1780                1785

Gly Asn  Ala His Val Ala Ser  Cys Asp Ala Ile Met  Thr Arg Cys
    1790                1795                1800

Leu Ala  Val His Glu Cys Phe  Val Lys Arg Val Asp  Trp Ser Val
    1805                1810                1815

Glu Tyr  Pro Ile Ile Gly Asp  Glu Leu Arg Val Asn  Ser Ala Cys
    1820                1825                1830

Arg Lys  Val Gln His Met Val  Val Lys Ser Ala Leu  Leu Ala Asp
    1835                1840                1845

Lys Phe  Pro Val Leu His Asp  Ile Gly Asn Pro Lys  Ala Ile Lys
    1850                1855                1860

Cys Val  Pro Gln Ala Glu Val  Glu Trp Lys Phe Tyr  Asp Ala Gln
    1865                1870                1875
```

-continued

```
Pro Cys  Ser Asp Lys Ala Tyr  Lys Ile Glu Glu Leu  Phe Tyr Ser
    1880                1885                1890

Tyr Ala  Thr His His Asp Lys  Phe Thr Asp Gly Val  Cys Leu Phe
    1895                1900                1905

Trp Asn  Cys Asn Val Asp Arg  Tyr Pro Ala Asn Ala  Ile Val Cys
    1910                1915                1920

Arg Phe  Asp Thr Arg Val Leu  Ser Asn Leu Asn Leu  Pro Gly Cys
    1925                1930                1935

Asp Gly  Gly Ser Leu Tyr Val  Asn Lys His Ala Phe  His Thr Pro
    1940                1945                1950

Ala Phe  Asp Lys Ser Ala Phe  Thr Asn Leu Lys Gln  Leu Pro Phe
    1955                1960                1965

Phe Tyr  Tyr Ser Asp Ser Pro  Cys Glu Ser His Gly  Lys Gln Val
    1970                1975                1980

Val Ser  Asp Ile Asp Tyr Val  Pro Leu Lys Ser Ala  Thr Cys Ile
    1985                1990                1995

Thr Arg  Cys Asn Leu Gly Gly  Ala Val Cys Arg His  His Ala Asn
    2000                2005                2010

Glu Tyr  Arg Gln Tyr Leu Asp  Ala Tyr Asn Met Met  Ile Ser Ala
    2015                2020                2025

Gly Phe  Ser Leu Trp Ile Tyr  Lys Gln Phe Asp Thr  Tyr Asn Leu
    2030                2035                2040

Trp Asn  Thr Phe Thr Arg Leu  Gln Ser Leu Glu Asn  Val Ala Tyr
    2045                2050                2055

Asn Val  Val Asn Lys Gly His  Phe Asp Gly His Ala  Gly Glu Ala
    2060                2065                2070

Pro Val  Ser Ile Ile Asn Asn  Ala Val Tyr Thr Lys  Val Asp Gly
    2075                2080                2085

Ile Asp  Val Glu Ile Phe Glu  Asn Lys Thr Thr Leu  Pro Val Asn
    2090                2095                2100

Val Ala  Phe Glu Leu Trp Ala  Lys Arg Asn Ile Lys  Pro Val Pro
    2105                2110                2115

Glu Ile  Lys Ile Leu Asn Asn  Leu Gly Val Asp Ile  Ala Ala Asn
    2120                2125                2130

Thr Val  Ile Trp Asp Tyr Lys  Arg Glu Ala Pro Ala  His Val Ser
    2135                2140                2145

Thr Ile  Gly Val Cys Thr Met  Thr Asp Ile Ala Lys  Lys Pro Thr
    2150                2155                2160

Glu Ser  Ala Cys Ser Ser Leu  Thr Val Leu Phe Asp  Gly Arg Val
    2165                2170                2175

Glu Gly  Gln Val Asp Leu Phe  Arg Asn Ala Arg Asn  Gly Val Leu
    2180                2185                2190

Ile Thr  Glu Gly Ser Val Lys  Gly Leu Thr Pro Ser  Lys Gly Pro
    2195                2200                2205

Ala Gln  Ala Ser Val Asn Gly  Val Thr Leu Ile Gly  Glu Ser Val
    2210                2215                2220

Lys Thr  Gln Phe Asn Tyr Phe  Lys Lys Val Asp Gly  Ile Ile Gln
    2225                2230                2235

Gln Leu  Pro Glu Thr Tyr Phe  Thr Gln Ser Arg Asp  Leu Glu Asp
    2240                2245                2250

Phe Lys  Pro Arg Ser Gln Met  Glu Thr Asp Phe Leu  Glu Leu Ala
    2255                2260                2265

Met Asp  Glu Phe Ile Gln Arg  Tyr Lys Leu Glu Gly  Tyr Ala Phe
```

-continued

```
    2270                2275                2280

Glu His  Ile Val Tyr Gly Asp  Phe Ser His Gly Gln  Leu Gly Gly
    2285                2290                2295

Leu His  Leu Met Ile Gly Leu  Ala Lys Arg Ser Gln  Asp Ser Pro
    2300                2305                2310

Leu Lys  Leu Glu Asp Phe Ile  Pro Met Asp Ser Thr  Val Lys Asn
    2315                2320                2325

Tyr Phe  Ile Thr Asp Ala Gln  Thr Gly Ser Ser Lys  Cys Val Cys
    2330                2335                2340

Ser Val  Ile Asp Leu Leu Leu  Asp Asp Phe Val Glu  Ile Ile Lys
    2345                2350                2355

Ser Gln  Asp Leu Ser Val Ile  Ser Lys Val Val Lys  Val Thr Ile
    2360                2365                2370

Asp Tyr  Ala Glu Ile Ser Phe  Met Leu Trp Cys Lys  Asp Gly His
    2375                2380                2385

Val Glu  Thr Phe Tyr Pro Lys  Leu Gln Ala Ser Gln  Ala Trp Gln
    2390                2395                2400

Pro Gly  Val Ala Met Pro Asn  Leu Tyr Lys Met Gln  Arg Met Leu
    2405                2410                2415

Leu Glu  Lys Cys Asp Leu Gln  Asn Tyr Gly Glu Asn  Ala Val Ile
    2420                2425                2430

Pro Lys  Gly Ile Met Met Asn  Val Ala Lys Tyr Thr  Gln Leu Cys
    2435                2440                2445

Gln Tyr  Leu Asn Thr Leu Thr  Leu Ala Val Pro Tyr  Asn Met Arg
    2450                2455                2460

Val Ile  His Phe Gly Ala Gly  Ser Asp Lys Gly Val  Ala Pro Gly
    2465                2470                2475

Thr Ala  Val Leu Arg Gln Trp  Leu Pro Thr Gly Thr  Leu Leu Val
    2480                2485                2490

Asp Ser  Asp Leu Asn Asp Phe  Val Ser Asp Ala Asp  Ser Thr Leu
    2495                2500                2505

Ile Gly  Asp Cys Ala Thr Val  His Thr Ala Asn Lys  Trp Asp Leu
    2510                2515                2520

Ile Ile  Ser Asp Met Tyr Asp  Pro Arg Thr Lys His  Val Thr Lys
    2525                2530                2535

Glu Asn  Asp Ser Lys Glu Gly  Phe Phe Thr Tyr Leu  Cys Gly Phe
    2540                2545                2550

Ile Lys  Gln Lys Leu Ala Leu  Gly Gly Ser Ile Ala  Val Lys Ile
    2555                2560                2565

Thr Glu  His Ser Trp Asn Ala  Asp Leu Tyr Lys Leu  Met Gly His
    2570                2575                2580

Phe Ser  Trp Trp Thr Ala Phe  Val Thr Asn Val Asn  Ala Ser Ser
    2585                2590                2595

Ser Glu  Ala Phe Leu Ile Gly  Ala Asn Tyr Leu Gly  Lys Pro Lys
    2600                2605                2610

Glu Gln  Ile Asp Gly Tyr Thr  Met His Ala Asn Tyr  Ile Phe Trp
    2615                2620                2625

Arg Asn  Thr Asn Pro Ile Gln  Leu Ser Ser Tyr Ser  Leu Phe Asp
    2630                2635                2640

Met Ser  Lys Phe Pro Leu Lys  Leu Arg Gly Thr Ala  Val Met Ser
    2645                2650                2655

Leu Lys  Glu Asn Gln Ile Asn  Asp Met Ile Tyr Ser  Leu Leu Glu
    2660                2665                2670
```

-continued

```
Lys Gly  Arg Leu Ile Ile Arg  Glu Asn Asn Arg Val  Val Val Ser
    2675                2680                2685

Ser Asp  Ile Leu Val Asn Asn
    2690                2695


<210> SEQ ID NO 4
<211> LENGTH: 1255
<212> TYPE: PRT
<213> ORGANISM: Coronavirus

<400> SEQUENCE: 4

Met Phe Ile Phe Leu Leu Phe Leu Thr Leu Thr Ser Gly Ser Asp Leu
1               5                  10                  15

Asp Arg Cys Thr Thr Phe Asp Asp Val Gln Ala Pro Asn Tyr Thr Gln
            20                  25                  30

His Thr Ser Ser Met Arg Gly Val Tyr Tyr Pro Asp Glu Ile Phe Arg
        35                  40                  45

Ser Asp Thr Leu Tyr Leu Thr Gln Asp Leu Phe Leu Pro Phe Tyr Ser
    50                  55                  60

Asn Val Thr Gly Phe His Thr Ile Asn His Thr Phe Gly Asn Pro Val
65                  70                  75                  80

Ile Pro Phe Lys Asp Gly Ile Tyr Phe Ala Ala Thr Glu Lys Ser Asn
                85                  90                  95

Val Val Arg Gly Trp Val Phe Gly Ser Thr Met Asn Asn Lys Ser Gln
            100                 105                 110

Ser Val Ile Ile Ile Asn Asn Ser Thr Asn Val Val Ile Arg Ala Cys
        115                 120                 125

Asn Phe Glu Leu Cys Asp Asn Pro Phe Phe Ala Val Ser Lys Pro Met
    130                 135                 140

Gly Thr Gln Thr His Thr Met Ile Phe Asp Asn Ala Phe Asn Cys Thr
145                 150                 155                 160

Phe Glu Tyr Ile Ser Asp Ala Phe Ser Leu Asp Val Ser Glu Lys Ser
                165                 170                 175

Gly Asn Phe Lys His Leu Arg Glu Phe Val Phe Lys Asn Lys Asp Gly
            180                 185                 190

Phe Leu Tyr Val Tyr Lys Gly Tyr Gln Pro Ile Asp Val Val Arg Asp
        195                 200                 205

Leu Pro Ser Gly Phe Asn Thr Leu Lys Pro Ile Phe Lys Leu Pro Leu
    210                 215                 220

Gly Ile Asn Ile Thr Asn Phe Arg Ala Ile Leu Thr Ala Phe Ser Pro
225                 230                 235                 240

Ala Gln Asp Ile Trp Gly Thr Ser Ala Ala Ala Tyr Phe Val Gly Tyr
                245                 250                 255

Leu Lys Pro Thr Thr Phe Met Leu Lys Tyr Asp Glu Asn Gly Thr Ile
            260                 265                 270

Thr Asp Ala Val Asp Cys Ser Gln Asn Pro Leu Ala Glu Leu Lys Cys
        275                 280                 285

Ser Val Lys Ser Phe Glu Ile Asp Lys Gly Ile Tyr Gln Thr Ser Asn
    290                 295                 300

Phe Arg Val Val Pro Ser Gly Asp Val Val Arg Phe Pro Asn Ile Thr
305                 310                 315                 320

Asn Leu Cys Pro Phe Gly Glu Val Phe Asn Ala Thr Lys Phe Pro Ser
                325                 330                 335

Val Tyr Ala Trp Glu Arg Lys Lys Ile Ser Asn Cys Val Ala Asp Tyr
```

-continued

```
            340                 345                 350

Ser Val Leu Tyr Asn Ser Thr Phe Phe Ser Thr Phe Lys Cys Tyr Gly
        355                 360                 365

Val Ser Ala Thr Lys Leu Asn Asp Leu Cys Phe Ser Asn Val Tyr Ala
    370                 375                 380

Asp Ser Phe Val Val Lys Gly Asp Asp Val Arg Gln Ile Ala Pro Gly
385                 390                 395                 400

Gln Thr Gly Val Ile Ala Asp Tyr Asn Tyr Lys Leu Pro Asp Asp Phe
                405                 410                 415

Met Gly Cys Val Leu Ala Trp Asn Thr Arg Asn Ile Asp Ala Thr Ser
            420                 425                 430

Thr Gly Asn Tyr Asn Tyr Lys Tyr Arg Tyr Leu Arg His Gly Lys Leu
        435                 440                 445

Arg Pro Phe Glu Arg Asp Ile Ser Asn Val Pro Phe Ser Pro Asp Gly
    450                 455                 460

Lys Pro Cys Thr Pro Pro Ala Leu Asn Cys Tyr Trp Pro Leu Asn Asp
465                 470                 475                 480

Tyr Gly Phe Tyr Thr Thr Thr Gly Ile Gly Tyr Gln Pro Tyr Arg Val
                485                 490                 495

Val Val Leu Ser Phe Glu Leu Leu Asn Ala Pro Ala Thr Val Cys Gly
            500                 505                 510

Pro Lys Leu Ser Thr Asp Leu Ile Lys Asn Gln Cys Val Asn Phe Asn
        515                 520                 525

Phe Asn Gly Leu Thr Gly Thr Gly Val Leu Thr Pro Ser Ser Lys Arg
    530                 535                 540

Phe Gln Pro Phe Gln Gln Phe Gly Arg Asp Val Ser Asp Phe Thr Asp
545                 550                 555                 560

Ser Val Arg Asp Pro Lys Thr Ser Glu Ile Leu Asp Ile Ser Pro Cys
                565                 570                 575

Ser Phe Gly Gly Val Ser Val Ile Thr Pro Gly Thr Asn Ala Ser Ser
            580                 585                 590

Glu Val Ala Val Leu Tyr Gln Asp Val Asn Cys Thr Asp Val Ser Thr
        595                 600                 605

Ala Ile His Ala Asp Gln Leu Thr Pro Ala Trp Arg Ile Tyr Ser Thr
    610                 615                 620

Gly Asn Asn Val Phe Gln Thr Gln Ala Gly Cys Leu Ile Gly Ala Glu
625                 630                 635                 640

His Val Asp Thr Ser Tyr Glu Cys Asp Ile Pro Ile Gly Ala Gly Ile
                645                 650                 655

Cys Ala Ser Tyr His Thr Val Ser Leu Leu Arg Ser Thr Ser Gln Lys
            660                 665                 670

Ser Ile Val Ala Tyr Thr Met Ser Leu Gly Ala Asp Ser Ser Ile Ala
        675                 680                 685

Tyr Ser Asn Asn Thr Ile Ala Ile Pro Thr Asn Phe Ser Ile Ser Ile
    690                 695                 700

Thr Thr Glu Val Met Pro Val Ser Met Ala Lys Thr Ser Val Asp Cys
705                 710                 715                 720

Asn Met Tyr Ile Cys Gly Asp Ser Thr Glu Cys Ala Asn Leu Leu Leu
                725                 730                 735

Gln Tyr Gly Ser Phe Cys Thr Gln Leu Asn Arg Ala Leu Ser Gly Ile
            740                 745                 750

Ala Ala Glu Gln Asp Arg Asn Thr Arg Glu Val Phe Ala Gln Val Lys
        755                 760                 765
```

-continued

```
Gln Met Tyr Lys Thr Pro Thr Leu Lys Tyr Phe Gly Gly Phe Asn Phe
    770                 775                 780

Ser Gln Ile Leu Pro Asp Pro Leu Lys Pro Thr Lys Arg Ser Phe Ile
785                 790                 795                 800

Glu Asp Leu Leu Phe Asn Lys Val Thr Leu Ala Asp Ala Gly Phe Met
                805                 810                 815

Lys Gln Tyr Gly Glu Cys Leu Gly Asp Ile Asn Ala Arg Asp Leu Ile
            820                 825                 830

Cys Ala Gln Lys Phe Asn Gly Leu Thr Val Leu Pro Pro Leu Leu Thr
        835                 840                 845

Asp Asp Met Ile Ala Ala Tyr Thr Ala Ala Leu Val Ser Gly Thr Ala
    850                 855                 860

Thr Ala Gly Trp Thr Phe Gly Ala Gly Ala Ala Leu Gln Ile Pro Phe
865                 870                 875                 880

Ala Met Gln Met Ala Tyr Arg Phe Asn Gly Ile Gly Val Thr Gln Asn
                885                 890                 895

Val Leu Tyr Glu Asn Gln Lys Gln Ile Ala Asn Gln Phe Asn Lys Ala
            900                 905                 910

Ile Ser Gln Ile Gln Glu Ser Leu Thr Thr Thr Ser Thr Ala Leu Gly
        915                 920                 925

Lys Leu Gln Asp Val Val Asn Gln Asn Ala Gln Ala Leu Asn Thr Leu
    930                 935                 940

Val Lys Gln Leu Ser Ser Asn Phe Gly Ala Ile Ser Ser Val Leu Asn
945                 950                 955                 960

Asp Ile Leu Ser Arg Leu Asp Lys Val Glu Ala Glu Val Gln Ile Asp
                965                 970                 975

Arg Leu Ile Thr Gly Arg Leu Gln Ser Leu Gln Thr Tyr Val Thr Gln
            980                 985                 990

Gln Leu Ile Arg Ala Ala Glu Ile  Arg Ala Ser Ala Asn  Leu Ala Ala
        995                 1000                1005

Thr Lys  Met Ser Glu Cys Val  Leu Gly Gln Ser Lys  Arg Val Asp
    1010                1015                1020

Phe Cys  Gly Lys Gly Tyr His  Leu Met Ser Phe Pro  Gln Ala Ala
    1025                1030                1035

Pro His  Gly Val Val Phe Leu  His Val Thr Tyr Val  Pro Ser Gln
    1040                1045                1050

Glu Arg  Asn Phe Thr Thr Ala  Pro Ala Ile Cys His  Glu Gly Lys
    1055                1060                1065

Ala Tyr  Phe Pro Arg Glu Gly  Val Phe Val Phe Asn  Gly Thr Ser
    1070                1075                1080

Trp Phe  Ile Thr Gln Arg Asn  Phe Phe Ser Pro Gln  Ile Ile Thr
    1085                1090                1095

Thr Asp  Asn Thr Phe Val Ser  Gly Asn Cys Asp Val  Val Ile Gly
    1100                1105                1110

Ile Ile  Asn Asn Thr Val Tyr  Asp Pro Leu Gln Pro  Glu Leu Asp
    1115                1120                1125

Ser Phe  Lys Glu Glu Leu Asp  Lys Tyr Phe Lys Asn  His Thr Ser
    1130                1135                1140

Pro Asp  Val Asp Leu Gly Asp  Ile Ser Gly Ile Asn  Ala Ser Val
    1145                1150                1155

Val Asn  Ile Gln Lys Glu Ile  Asp Arg Leu Asn Glu  Val Ala Lys
    1160                1165                1170
```

-continued

```
Asn Leu  Asn Glu Ser Leu Ile  Asp Leu Gln Glu Leu  Gly Lys Tyr
    1175                1180                1185

Glu Gln  Tyr Ile Lys Trp Pro  Trp Tyr Val Trp Leu  Gly Phe Ile
    1190                1195                1200

Ala Gly  Leu Ile Ala Ile Val  Met Val Thr Ile Leu  Leu Cys Cys
    1205                1210                1215

Met Thr  Ser Cys Cys Ser Cys  Leu Lys Gly Ala Cys  Ser Cys Gly
    1220                1225                1230

Ser Cys  Cys Lys Phe Asp Glu  Asp Asp Ser Glu Pro  Val Leu Lys
    1235                1240                1245

Gly Val  Lys Leu His Tyr Thr
    1250                1255


<210> SEQ ID NO 5
<211> LENGTH: 274
<212> TYPE: PRT
<213> ORGANISM: Coronavirus

<400> SEQUENCE: 5

Met Asp Leu Phe Met Arg Phe Phe Thr Leu Gly Ser Ile Thr Ala Gln
1               5                   10                  15

Pro Val Lys Ile Asp Asn Ala Ser Pro Ala Ser Thr Val His Ala Thr
            20                  25                  30

Ala Thr Ile Pro Leu Gln Ala Ser Leu Pro Phe Gly Trp Leu Val Ile
        35                  40                  45

Gly Val Ala Phe Leu Ala Val Phe Gln Ser Ala Thr Lys Ile Ile Ala
    50                  55                  60

Leu Asn Lys Arg Trp Gln Leu Ala Leu Tyr Lys Gly Phe Gln Phe Ile
65                  70                  75                  80

Cys Asn Leu Leu Leu Leu Phe Val Thr Ile Tyr Ser His Leu Leu Leu
                85                  90                  95

Val Ala Ala Gly Met Glu Ala Gln Phe Leu Tyr Leu Tyr Ala Leu Ile
            100                 105                 110

Tyr Phe Leu Gln Cys Ile Asn Ala Cys Arg Ile Ile Met Arg Cys Trp
        115                 120                 125

Leu Cys Trp Lys Cys Lys Ser Lys Asn Pro Leu Leu Tyr Asp Ala Asn
    130                 135                 140

Tyr Phe Val Cys Trp His Thr His Asn Tyr Asp Tyr Cys Ile Pro Tyr
145                 150                 155                 160

Asn Ser Val Thr Asp Thr Ile Val Val Thr Glu Gly Asp Gly Ile Ser
                165                 170                 175

Thr Pro Lys Leu Lys Glu Asp Tyr Gln Ile Gly Gly Tyr Ser Glu Asp
            180                 185                 190

Arg His Ser Gly Val Lys Asp Tyr Val Val Val His Gly Tyr Phe Thr
        195                 200                 205

Glu Val Tyr Tyr Gln Leu Glu Ser Thr Gln Ile Thr Thr Asp Thr Gly
    210                 215                 220

Ile Glu Asn Ala Thr Phe Phe Ile Phe Asn Lys Leu Val Lys Asp Pro
225                 230                 235                 240

Pro Asn Val Gln Ile His Thr Ile Asp Gly Ser Ser Gly Val Ala Asn
                245                 250                 255

Pro Ala Met Asp Pro Ile Tyr Asp Glu Pro Thr Thr Thr Thr Ser Val
            260                 265                 270

Pro Leu
```

-continued

<210> SEQ ID NO 6
<211> LENGTH: 154
<212> TYPE: PRT
<213> ORGANISM: Coronavirus

<400> SEQUENCE: 6

Met Met Pro Thr Thr Leu Phe Ala Gly Thr His Ile Thr Met Thr Thr
1 5 10 15

Val Tyr His Ile Thr Val Ser Gln Ile Gln Leu Ser Leu Leu Lys Val
20 25 30

Thr Ala Phe Gln His Gln Asn Ser Lys Lys Thr Thr Lys Leu Val Val
35 40 45

Ile Leu Arg Ile Gly Thr Gln Val Leu Lys Thr Met Ser Leu Tyr Met
50 55 60

Ala Ile Ser Pro Lys Phe Thr Thr Ser Leu Ser Leu His Lys Leu Leu
65 70 75 80

Gln Thr Leu Val Leu Lys Met Leu His Ser Ser Ser Leu Thr Ser Leu
85 90 95

Leu Lys Thr His Arg Met Cys Lys Tyr Thr Gln Ser Thr Ala Leu Gln
100 105 110

Glu Leu Leu Ile Gln Gln Trp Ile Gln Phe Met Met Ser Arg Arg Arg
115 120 125

Leu Leu Ala Cys Leu Cys Lys His Lys Lys Val Ser Thr Asn Leu Cys
130 135 140

Thr His Ser Phe Arg Lys Lys Gln Val Arg
145 150

<210> SEQ ID NO 7
<211> LENGTH: 76
<212> TYPE: PRT
<213> ORGANISM: Coronavirus

<400> SEQUENCE: 7

Met Tyr Ser Phe Val Ser Glu Glu Thr Gly Thr Leu Ile Val Asn Ser
1 5 10 15

Val Leu Leu Phe Leu Ala Phe Val Val Phe Leu Leu Val Thr Leu Ala
20 25 30

Ile Leu Thr Ala Leu Arg Leu Cys Ala Tyr Cys Cys Asn Ile Val Asn
35 40 45

Val Ser Leu Val Lys Pro Thr Val Tyr Val Tyr Ser Arg Val Lys Asn
50 55 60

Leu Asn Ser Ser Glu Gly Val Pro Asp Leu Leu Val
65 70 75

<210> SEQ ID NO 8
<211> LENGTH: 221
<212> TYPE: PRT
<213> ORGANISM: Coronavirus

<400> SEQUENCE: 8

Met Ala Asp Asn Gly Thr Ile Thr Val Glu Glu Leu Lys Gln Leu Leu
1 5 10 15

Glu Gln Trp Asn Leu Val Ile Gly Phe Leu Phe Leu Ala Trp Ile Met
20 25 30

Leu Leu Gln Phe Ala Tyr Ser Asn Arg Asn Arg Phe Leu Tyr Ile Ile
35 40 45

Lys Leu Val Phe Leu Trp Leu Leu Trp Pro Val Thr Leu Ala Cys Phe

-continued

```
    50                  55                  60

Val Leu Ala Ala Val Tyr Arg Ile Asn Trp Val Thr Gly Gly Ile Ala
65                  70                  75                  80

Ile Ala Met Ala Cys Ile Val Gly Leu Met Trp Leu Ser Tyr Phe Val
                85                  90                  95

Ala Ser Phe Arg Leu Phe Ala Arg Thr Arg Ser Met Trp Ser Phe Asn
            100                 105                 110

Pro Glu Thr Asn Ile Leu Leu Asn Val Pro Leu Arg Gly Thr Ile Val
        115                 120                 125

Thr Arg Pro Leu Met Glu Ser Glu Leu Val Ile Gly Ala Val Ile Ile
    130                 135                 140

Arg Gly His Leu Arg Met Ala Gly His Pro Leu Gly Arg Cys Asp Ile
145                 150                 155                 160

Lys Asp Leu Pro Lys Glu Ile Thr Val Ala Thr Ser Arg Thr Leu Ser
                165                 170                 175

Tyr Tyr Lys Leu Gly Ala Ser Gln Arg Val Gly Thr Asp Ser Gly Phe
            180                 185                 190

Ala Ala Tyr Asn Arg Tyr Arg Ile Gly Asn Tyr Lys Leu Asn Thr Asp
        195                 200                 205

His Ala Gly Ser Asn Asp Asn Ile Ala Leu Leu Val Gln
    210                 215                 220


<210> SEQ ID NO 9
<211> LENGTH: 63
<212> TYPE: PRT
<213> ORGANISM: Coronavirus

<400> SEQUENCE: 9

Met Phe His Leu Val Asp Phe Gln Val Thr Ile Ala Glu Ile Leu Ile
1               5                   10                  15

Ile Ile Met Arg Thr Phe Arg Ile Ala Ile Trp Asn Leu Asp Val Ile
            20                  25                  30

Ile Ser Ser Ile Val Arg Gln Leu Phe Lys Pro Leu Thr Lys Lys Asn
        35                  40                  45

Tyr Ser Glu Leu Asp Asp Glu Glu Pro Met Glu Leu Asp Tyr Pro
    50                  55                  60


<210> SEQ ID NO 10
<211> LENGTH: 122
<212> TYPE: PRT
<213> ORGANISM: Coronavirus

<400> SEQUENCE: 10

Met Lys Ile Ile Leu Phe Leu Thr Leu Ile Val Phe Thr Ser Cys Glu
1               5                   10                  15

Leu Tyr His Tyr Gln Glu Cys Val Arg Gly Thr Thr Val Leu Leu Lys
            20                  25                  30

Glu Pro Cys Pro Ser Gly Thr Tyr Glu Gly Asn Ser Pro Phe His Pro
        35                  40                  45

Leu Ala Asp Asn Lys Phe Ala Leu Thr Cys Thr Ser Thr His Phe Ala
    50                  55                  60

Phe Ala Cys Ala Asp Gly Thr Arg His Thr Tyr Gln Leu Arg Ala Arg
65                  70                  75                  80

Ser Val Ser Pro Lys Leu Phe Ile Arg Gln Glu Glu Val Gln Gln Glu
                85                  90                  95

Leu Tyr Ser Pro Leu Phe Leu Ile Val Ala Ala Leu Val Phe Leu Ile
```

-continued

```
            100                 105                 110

Leu Cys Phe Thr Ile Lys Arg Lys Thr Glu
        115                 120


<210> SEQ ID NO 11
<211> LENGTH: 84
<212> TYPE: PRT
<213> ORGANISM: Coronavirus

<400> SEQUENCE: 11

Met Cys Leu Lys Ile Leu Val Arg Tyr Asn Thr Arg Gly Asn Thr Tyr
1               5                   10                  15

Ser Thr Ala Trp Leu Cys Ala Leu Gly Lys Val Leu Pro Phe His Arg
            20                  25                  30

Trp His Thr Met Val Gln Thr Cys Thr Pro Asn Val Thr Ile Asn Cys
        35                  40                  45

Gln Asp Pro Ala Gly Gly Ala Leu Ile Ala Arg Cys Trp Tyr Leu His
    50                  55                  60

Glu Gly His Gln Thr Ala Ala Phe Arg Asp Val Leu Val Val Leu Asn
65                  70                  75                  80

Lys Arg Thr Asn


<210> SEQ ID NO 12
<211> LENGTH: 422
<212> TYPE: PRT
<213> ORGANISM: Coronavirus

<400> SEQUENCE: 12

Met Ser Asp Asn Gly Pro Gln Ser Asn Gln Arg Ser Ala Pro Arg Ile
1               5                   10                  15

Thr Phe Gly Gly Pro Thr Asp Ser Thr Asp Asn Asn Gln Asn Gly Gly
            20                  25                  30

Arg Asn Gly Ala Arg Pro Lys Gln Arg Arg Pro Gln Gly Leu Pro Asn
        35                  40                  45

Asn Thr Ala Ser Trp Phe Thr Ala Leu Thr Gln His Gly Lys Glu Glu
    50                  55                  60

Leu Arg Phe Pro Arg Gly Gln Gly Val Pro Ile Asn Thr Asn Ser Gly
65                  70                  75                  80

Pro Asp Asp Gln Ile Gly Tyr Tyr Arg Arg Ala Thr Arg Arg Val Arg
                85                  90                  95

Gly Gly Asp Gly Lys Met Lys Glu Leu Ser Pro Arg Trp Tyr Phe Tyr
            100                 105                 110

Tyr Leu Gly Thr Gly Pro Glu Ala Ser Leu Pro Tyr Gly Ala Asn Lys
        115                 120                 125

Glu Gly Ile Val Trp Val Ala Thr Glu Gly Ala Leu Asn Thr Pro Lys
    130                 135                 140

Asp His Ile Gly Thr Arg Asn Pro Asn Asn Asn Ala Ala Thr Val Leu
145                 150                 155                 160

Gln Leu Pro Gln Gly Thr Thr Leu Pro Lys Gly Phe Tyr Ala Glu Gly
                165                 170                 175

Ser Arg Gly Gly Ser Gln Ala Ser Ser Arg Ser Ser Ser Arg Ser Arg
            180                 185                 190

Gly Asn Ser Arg Asn Ser Thr Pro Gly Ser Ser Arg Gly Asn Ser Pro
        195                 200                 205

Ala Arg Met Ala Ser Gly Gly Gly Glu Thr Ala Leu Ala Leu Leu Leu
    210                 215                 220
```

-continued

```
Leu Asp Arg Leu Asn Gln Leu Glu Ser Lys Val Ser Gly Lys Gly Gln
225                 230                 235                 240

Gln Gln Gln Gly Gln Thr Val Thr Lys Lys Ser Ala Ala Glu Ala Ser
                245                 250                 255

Lys Lys Pro Arg Gln Lys Arg Thr Ala Thr Lys Gln Tyr Asn Val Thr
            260                 265                 270

Gln Ala Phe Gly Arg Arg Gly Pro Glu Gln Thr Gln Gly Asn Phe Gly
        275                 280                 285

Asp Gln Asp Leu Ile Arg Gln Gly Thr Asp Tyr Lys His Trp Pro Gln
    290                 295                 300

Ile Ala Gln Phe Ala Pro Ser Ala Ser Ala Phe Phe Gly Met Ser Arg
305                 310                 315                 320

Ile Gly Met Glu Val Thr Pro Ser Gly Thr Trp Leu Thr Tyr His Gly
                325                 330                 335

Ala Ile Lys Leu Asp Asp Lys Asp Pro Gln Phe Lys Asp Asn Val Ile
            340                 345                 350

Leu Leu Asn Lys His Ile Asp Ala Tyr Lys Thr Phe Pro Pro Thr Glu
        355                 360                 365

Pro Lys Lys Asp Lys Lys Lys Lys Thr Asp Glu Ala Gln Pro Leu Pro
    370                 375                 380

Gln Arg Gln Lys Lys Gln Pro Thr Val Thr Leu Leu Pro Ala Ala Asp
385                 390                 395                 400

Met Asp Asp Phe Ser Arg Gln Leu Gln Asn Ser Met Ser Gly Ala Ser
                405                 410                 415

Ala Asp Ser Thr Gln Ala
            420
```

<210> SEQ ID NO 13
<211> LENGTH: 21
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic oligonucleotide.

<400> SEQUENCE: 13

ctaacatgct taggataatg g 21

<210> SEQ ID NO 14
<211> LENGTH: 21
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic oligonucleotide.

<400> SEQUENCE: 14

gcctctcttg ttcttgctcg c 21

<210> SEQ ID NO 15
<211> LENGTH: 21
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic oligonucleotide.

<400> SEQUENCE: 15

caggtaagcg taaaactcat c 21

<210> SEQ ID NO 16
<211> LENGTH: 22

-continued

<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic oligonucleotide.

<400> SEQUENCE: 16

catgtgtggc ggctcactat at 22

<210> SEQ ID NO 17
<211> LENGTH: 28
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic oligonucleotide.

<400> SEQUENCE: 17

gacactatta gcataagcag ttgtagca 28

<210> SEQ ID NO 18
<211> LENGTH: 27
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic oligonucleotide.

<400> SEQUENCE: 18

ttaaaccagg tggaacatca tccggtg 27

<210> SEQ ID NO 19
<211> LENGTH: 22
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic oligonucleotide.

<400> SEQUENCE: 19

ggagccttga atacacccaa ag 22

<210> SEQ ID NO 20
<211> LENGTH: 17
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic oligonucleotide.

<400> SEQUENCE: 20

gcacggtggc agcattg 17

<210> SEQ ID NO 21
<211> LENGTH: 21
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic oligonucleotide.

<400> SEQUENCE: 21

ccacattggc acccgcaatc c 21

<210> SEQ ID NO 22
<211> LENGTH: 19
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic oligonucleotide.

<400> SEQUENCE: 22

-continued

caaacattgg ccgcaaatt 19

<210> SEQ ID NO 23
<211> LENGTH: 21
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic oligonucleotide.

<400> SEQUENCE: 23

caatgcgtga cattccaaag a 21

<210> SEQ ID NO 24
<211> LENGTH: 26
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic oligonucleotide.

<400> SEQUENCE: 24

cacaatttgc tccaagtgcc tctgca 26

<210> SEQ ID NO 25
<211> LENGTH: 21
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic oligonucleotide.

<400> SEQUENCE: 25

gaagtaccat ctggggctga g 21

<210> SEQ ID NO 26
<211> LENGTH: 18
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic oligonucleotide.

<400> SEQUENCE: 26

ccgaagagct acccgacg 18

<210> SEQ ID NO 27
<211> LENGTH: 26
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic oligonucleotide.

<400> SEQUENCE: 27

ctctttcatt ttgccgtcac caccac 26

<210> SEQ ID NO 28
<211> LENGTH: 24
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic oligonucleotide.

<400> SEQUENCE: 28

agctctccct agcattattc actg 24

<210> SEQ ID NO 29
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence

-continued

<220> FEATURE:
<223> OTHER INFORMATION: Synthetic oligonucleotide.

<400> SEQUENCE: 29

caccacattt tcatcgaggc 20

<210> SEQ ID NO 30
<211> LENGTH: 22
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic oligonucleotide.

<400> SEQUENCE: 30

taccctcgat cgtactccgc gt 22

<210> SEQ ID NO 31
<211> LENGTH: 23
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic oligonucleotide.

<400> SEQUENCE: 31

tgtaggcact gattcaggtt ttg 23

<210> SEQ ID NO 32
<211> LENGTH: 23
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic oligonucleotide.

<400> SEQUENCE: 32

cggcgtggtc tgtatttaat tta 23

<210> SEQ ID NO 33
<211> LENGTH: 27
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic oligonucleotide.

<400> SEQUENCE: 33

ctgcatacaa ccgctaccgt attggaa 27

<210> SEQ ID NO 34
<211> LENGTH: 24
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic oligonucleotide.

<400> SEQUENCE: 34

gggttgggac tatcctaagt gtga 24

<210> SEQ ID NO 35
<211> LENGTH: 21
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic oligonucleotide.
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (12)..(12)
<223> OTHER INFORMATION: "n" equals inosine.

-continued

<400> SEQUENCE: 35

taacacacaa cnccatcatc a 21

<210> SEQ ID NO 36
<211> LENGTH: 19
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic oligonucleotide.

<400> SEQUENCE: 36

agatttggac ctgcgagcg 19

<210> SEQ ID NO 37
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic oligonucleotide.

<400> SEQUENCE: 37

gagcggctgt ctccacaagt 20

<210> SEQ ID NO 38
<211> LENGTH: 23
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic oligonucleotide.

<400> SEQUENCE: 38

ttctgacctg aaggctctgc gcg 23

The invention claimed is:

1. A method of detecting a severe acute respiratory syndrome-associated coronavirus (SARS-CoV) in a sample comprising:

contacting the sample with a pair of nucleic acid primers that hybridize to a SARS-CoV nucleic acid, wherein at least one primer is 5'-end labeled with a reporter dye, and wherein at least one of the primers comprises the sequence as set forth in any one of SEQ ID NOs: 13-15;

amplifying the SARS-CoV nucleic acid or a fragment thereof from the sample utilizing the pair of nucleic acid primers;

electrophoresing the amplified products; and

detecting the 5'-end labeled reporter dye, thereby detecting a SARS-CoV.

2. The method of claim **1**, wherein the amplification utilizes reverse transcriptase-polymerase chain reaction.

3. A method of detecting a severe acute respiratory syndrome-associated coronavirus (SARS-CoV) in a sample, comprising:

contacting the sample with a pair of nucleic acid primers that hybridize to a SARS-CoV nucleic acid, wherein at least one of the nucleic acid primers comprises the sequence as set forth in any one of SEQ ID NOs: 13-15;

amplifying the SARS-CoV nucleic acid or a fragment thereof from the sample utilizing the pair of nucleic acid primers;

adding to the amplified SARS-CoV nucleic acid or the fragment thereof a SARS-CoV probe that hybridizes to the SARS-CoV nucleic acid, wherein the SARS-CoV probe is labeled with a 5'-reporter dye and a 3'-quencher dye;

performing one or more additional rounds of amplification with Taq DNA polymerase; and

detecting fluorescence of the 5'-reporter dye, thereby detecting a SARS-CoV.

4. A kit for detecting a severe acute respiratory syndrome-associated coronavirus (SARS-CoV) in a sample, comprising a pair of nucleic acid primers that hybridize under stringent conditions to a SARS-CoV nucleic acid, wherein at least one of the primers comprises the sequence as set forth in any one of SEQ ID NOs: 13-15.

5. The kit of claim **4**, wherein one primer is 5'-end labeled with a reporter dye.

6. The kit of claim **4**, further comprising a SARS-CoV probe that hybridizes to the SARS-CoV nucleic acid amplified by the pair of primers, wherein the SARS-CoV probe is labeled with a 5'-reporter dye and a 3'-quencher dye.

7. The kit of claim **4**, further comprising an isolated SARS-CoV organism.

* * * * *

(19) Europäisches Patentamt
European Patent Office
Office européen des brevets

(11) EP 1 508 615 A1

(12) EUROPEAN PATENT APPLICATION

(43) Date of publication:
23.02.2005 Bulletin 2005/08

(21) Application number: **03077602.5**

(22) Date of filing: **18.08.2003**

(51) Int Cl.[7]: **C12N 7/00**, C07K 14/165, A61K 39/215, C12Q 1/68, G01N 33/569, C07K 16/10

(84) Designated Contracting States:
AT BE BG CH CY CZ DE DK EE ES FI FR GB GR HU IE IT LI LU MC NL PT RO SE SI SK TR
Designated Extension States:
AL LT LV MK

(71) Applicant: **Amsterdam Institute of Viral Genomics B.V.**
1105 BA Amsterdam (NL)

(72) Inventor: **van der Hoek, Cornelia**
1111 HK Diemen (NL)

(74) Representative:
Prins, Adrianus Willem, Mr. Ir. et al
Vereenigde,
Nieuwe Parklaan 97
2587 BN Den Haag (NL)

Remarks:
The sequence listing, which is published as annex to the application documents, was filed after the date of filing. The applicant has declared that it does not include matter which goes beyond the content of the application as filed.

(54) Coronavirus, nucleic acid, protein, and methods for the generation of vaccine, medicaments and diagnostics

(57) A new coronavirus (human coronavirus NL63 (HCoV-NL63)) is disclosed herein with a tropism that includes humans. Means and methods are provided for diagnosing subjects (previously) infected with the virus. Also provided are among others vaccines, medicaments, nucleic acids and specific binding members.

EP 1 508 615 A1

Description

[0001] The invention relates to the fields of virology and medicine. More in particular the invention relates to the identification of a new coronavirus and to means and methods associated with a virus such as means and methods for typing the virus in various samples and diagnosing of disease, means and methods for developing vaccines and medicaments for the treatment of infected subjects or of subjects at risk thereof.

[0002] To date there is a range of human diseases with unknown etiology. For many of these a viral origin has been suggested, emphasizing the importance of a continuous search for new viruses[22, 23, 24]. Major difficulties are encountered when searching for new viruses. First, some viruses do not replicate in vitro, at least not in the cells that are commonly used in viral diagnostics. Second, for those viruses that do replicate in vitro and that cause a cytopathic effect (CPE), the subsequent virus-identification methods may fail. Antibodies raised against known viruses may not recognize the cultured virus and virus specific PCR methods may not amplify the new viral genome. We have developed a method for virus discovery based on the cDNA amplified restriction fragment length polymorphism technique (cDNA-AFLP). With this technique, RNA or DNA is reproducibly amplified. There is no need to have prior knowledge of the sequence of the target gene[1]. Generally the cDNA-AFLP method is used to monitor differential gene expression, however, we modified this method such that it can amplify viral sequences either directly from patient blood-plasma/serum samples or indirectly from CPE-positive virus culture (Figure 1). In the modified Virus-Discovery-cDNA-AFLP (VIDISCA) method the mRNA isolation step prior to amplification is replaced by a treatment to selectively enrich for viral nucleic acid. Of relevance to the purification is a centrifugation step to remove residual cells and mitochondria. In addition, a DNAse treatment can be used to remove interfering chromosomal and mitochondrial DNA from degraded cells whereas viral nucleic acid is protected within the viral particle. Finally, by choosing frequently cutting restriction enzymes, the method can be fine-tuned such that most viruses will be amplified.

[0003] In January 2003 a 7-month-old child appeared in the hospital with coryza, conjunctivitis and fever. Chest radiography showed typical features of bronchiolitis and a nasopharyngeal aspirate specimen was collected (sample nr: NL63) five days after the onset of disease. All diagnostic tests on this sample for respiratory syncytial virus (RSV), adenovirus, influenza A and B virus, parainfluenza virus type 1, 2 and 3, rhinovirus, enterovirus, HCoV-229E and HCoV-OC43 were negative. Immunofluorescent assays to detect RSV, adenovirus, influenza A and B virus, and parainfluenza virus type 1, 2 and 3 in cultures of the virus remained negative. Acid lability and chloroform sensitivity tests demonstrated that the virus was most likely enveloped and not a member of the Picornavirus group.

[0004] VIDISCA analysis and subsequent sequencing of cloned fragments revealed that the child was infected by a previously unknown member of the Coronavirus family. Coronaviruses are characterized by a very long non-segmented, single-stranded, (+) sense RNA of approximately 27-31 kb. This is the longest genome of any known RNA virus. The genome has a 5' methylated cap and 3' poly-A and functions directly as mRNA.

The newly found coronavirus, (designated HCoV-NL63) was characterized and several of the isolated fragments were sequenced.. The sequences of a number of the fragments are depicted in table 3. The location of the fragments in the large genomic RNA is depicted in figure 5. The invention therefore, in one aspect, provides an isolated or recombinant virus comprising a nucleic acid sequence as depicted in table 3, or a functional part, derivative or analogue of said virus. With the aid of the identifying prototype fragments it is possible to further sequence the genome. One way of doing this by primer walking on the genome. A primer is directed to a region of which the sequence is known and this primer is used to sequence a flanking region that is as yet unknown. A subsequent primer can be generated against the newly identified sequence and a further region can be sequenced. This procedure can be repeated until the entire sequence of the virus is elucidated. As a source of the virus one may turn Dr. C van der Hoek, Department of Human Retrovirology, Academic Medical Center, University of Amsterdam, Amsterdam, The Netherlands

[0005] Alignments of the determined nucleic acid sequences revealed the reading frame used in the sequences found, accordingly the invention further provides an isolated or recombinant virus comprising an amino acid sequence as depicted in (table 3). or a functional part, derivative or analogue of said virus. A particular amino acid sequence can be produced from a variety of nucleic acids depending on the codons used. Thus the invention further provides a nucleic acid encoding an amino acid sequence as depicted in (table 3). Further provided is an isolated or recombinant virus comprising a nucleic acid sequence encoding an amino acid sequence as depicted in (table 3), or a functional part, derivative or analogue of said virus.

[0006] Coronaviruses as many other types of viruses acquire a plurality of spontaneous and selected mutations upon spreading of the virus through the subject population and/or during culturing ex vivo. Moreover, artificial mutations having no recognized counterpart in nature can be introduced into the sequence of the prototype virus or a derivative thereof, without altering the viral- and/or disease causing properties of the virus. Having characterized the prototype of the newly discovered subtype gives access to this group of viruses belonging to the same subtype. Thus the invention further provides an isolated or recombinant virus comprising a nucleic acid sequence that is approximately 80% homologous to a sequence as depicted in table 3, or 80 % homologous to an amino acid sequence depicted in Table 3 (. Preferably the homology is at least 90%, more preferably at least 95% and even more preferably at least 99%.

[0007] The respective prototype fragments were compared with a database of viral sequences and hits having a particularly high homology are mentioned in the tables 5 and 6. It may be noted that the compared fragments do not share extensive homology with any of the currently known Coronaviruses. The invention thus provides an isolated and/or recombinant virus comprising an amino acid sequence which is more than 89% homologous to 163- 2 amino acid sequence as depicted in Table 3. Preferably said homology is at least 90%, more preferably at least 95% and even more preferably at least 99%.

Further provided is an isolated or recombinant virus comprising an amino acid sequence which is more than 60 % homologous to 163- 4 amino acid sequence as depicted in Table 3. Preferably said homology is at least 90%, more preferably at least 95% and even more preferably at least 99%.

Further provided is an isolated or recombinant virus comprising a nucleic acid sequence which is more than 85 % homologous to 163- 9 nucleic acid sequence as depicted in Table 3. Preferably said homology is at least 90%, more preferably at least 95% and even more preferably at least 99%.

Further provided is an isolated or recombinant virus comprising an amino acid sequence which is more than 94 % homologous to 163- 10 amino acid sequence as depicted in Table 3. Preferably said homology is at least 90%, more preferably at least 95% and even more preferably at least 99%.

Further provided is an isolated or recombinant virus comprising an amino acid sequence which is more than 50 % homologous to 163- 11 amino acid sequence as depicted in Table 3. Preferably said homology is at least 90%, more preferably at least 95% and even more preferably at least 99%.

Further provided is an isolated or recombinant virus comprising an amino acid sequence which is more than 87 % homologous to 163- 14 amino acid sequence as depicted in Table 3. Preferably said homology is at least 90%, more preferably at least 95% and even more preferably at least 99%.

Further provided is an isolated or recombinant virus comprising an amino acid sequence which is more than 83 % homologous to 163- 15 amino acid sequence as depicted in Table 3. Preferably said homology is at least 90%, more preferably at least 95% and even more preferably at least 99%.

Further provided is an isolated or recombinant virus comprising an amino acid sequence which is more than 78 % homologous to 163- 18 amino acid sequence as depicted in Table 3. Preferably said homology is at least 90%, more preferably at least 95% and even more preferably at least 99%.

Further provided is an isolated or recombinant virus comprising a nucleic acid sequence which is at least 50 % homologous to a nucleic acid sequence as depicted in Table 3. Preferably said homology is at least 80%, more preferably at least 90%, more preferably at least 95% and even more preferably at least 99%.

[0008] The invention also provides a functional part, derivative and/or analogue of an isolated and/or recombinant HCoV-NL63 virus. A part of a virus can be a membrane containing part, a nucleocapsid containing part, a proteinaceous fragment and/or a nucleic acid containing part. The functionality of the part varies with the application chosen for the part, for instance, part of the virus may be used for immunization purposes. In this embodiment the functionality comprises similar immunogenic properties in kind as the entire virus not necessarily in amount. Another use of the virus is the infectivity of the virus, for instance, for in vitro (or in vivo) culture, in this embodiment the functionality comprises a similar infectivity in kind not necessarily in amount. Many other functionalities may be defined, as there are many different uses for viruses, non-limiting examples are the generation of chimeric viruses, (i.e. with one or more other (corona) viruses, and the generation of viral vectors for vaccination and/or gene therapeutic purposes.

[0009] Such viruses and/or vectors also contain a functional part of HCoV-NL63 and are thus also encompassed in the present invention. A functional derivative of a virus of the invention is defined as a virus that has been altered such that the properties of said compound are essentially the same in kind, not necessarily in amount. A derivative can be provided in many ways, for instance through nucleotide substitution (preferably "wobble" based), through (conservative) amino acid substitution, subsequent modification, etcetera.

Analogous compounds of a virus can also be generated using methods in the art. For instance, a chimeric virus can be produced, or an HCoV-NL63 virus having a chimeric protein. For instance, HCoV-NL63 can be rendered more immunogenic by generating a cell surface associated fusion protein comprising at least part of an HCoV-NL63 surface protein and a non-HCoV-NL63 immunogenic part. HCoV-NL63 virus comprising such chimeric protein can be used for inducing an enhanced immune response in a host, for instance for vaccination purposes.

As used herein, the term "a virus of the invention" is meant to also comprise a functional part, derivative and/or analogue of said virus.

[0010] The three groups of coronaviruses are associated with a variety of diseases of humans and domestic animals, including gastroenteritis and upper and lower respiratory tract disease. The human coronaviruses HCoV-229E and HCoV-OC43 are associated with mild disease (the common cold) but more severe disease is observed in children [16], albeit at a very low incidence. Several coronaviruses cause a severe disease in animals and SARS-CoV is the first example of a coronavirus that causes severe disease in humans. However, it should be emphasized that a substantial part of respiratory disease cases in humans remains undiagnosed. For instance, a recent survey of respiratory viruses in hospitalized children with bronchiolitis in Canada could not reveal a viral pathogen in about 20% of the patients[17].

The fact that we identified the new coronavirus in a child with bronchiolitis shows that HCoV-NL63 is a pathogenic respiratory virus.

When considering that the HCoV-NL63 is a pathogenic respiratory virus able to cause bronchiolitis in infected children, the interesting question remains why HCoV-NL63 was not recognized previously by cell culture. We found that the virus can be cultured in monkey kidney cells (tMK or LLC-MK2 cells), cells that are often used in a routine diagnostic setting and one might therefore speculate that HCoV-NL63, like SARS-CoV, was newly introduced from an animal reservoir into the human population or that this is a human virus that recently broadened its host cell range. Clearly it is of importance to study the prevalence of HCoV-NL63 infection, and screening specimens from patients with respiratory tract disease using the HCoV-NL63 diagnostic RT-PCR will shed light on this issue.

It is remarkable that the new human coronavirus was harvested from tMK cells and LLC-MK2 cells since coronaviruses are typically fastidious in cell culture with a narrow host range. However, both SARS-CoV and HCoV-NL63 seem to replicate efficiently in monkey kidney cells (Vero-E6 cells and NCI-H292 cells for SARS-CoV). The recently described genome of SARS-CoV has several exclusive features, including some unique open reading frames that are probably of biological significance[15,18]. We will therefore analyze the complete genome sequence of HCoV-NL63 to screen for similarities and differences with SARS-CoV that may determine the expanded host cell range and enhanced pathogenicity of these viruses.

[0011] HCoV-NL63 is associated with a particular phenotype in infected subjects. The phenotype can encompass bronchiolitis, coryza, conjunctivitis and fever and may further encompass other respiratory problems and diarrhea. In one embodiment the invention thus further provides an isolated and or recombinant virus of the invention (having one or more of the above mentioned homology) wherein said virus or functional part, derivative and/or analogue further comprises the capability to induce an HCoV-NL63 related disease or symptom in a subject. In another embodiment the invention provides an isolated and/or recombinant virus of the invention further comprising the property to cause CPE in tertiary monkey kidney cells (tMK; Cynomolgus monkey[37]) and/or upon passage onto the monkey cell line LLC-MK2 (ECCAC 85062804, ATCC CCL-7). In a preferred embodiment said virus does not produce CPE in Vero-cells **(ATCC CRL-1586)[34].**

[0012] The invention further provides a nucleic acid as depicted in table 3, and an amino acid sequence as depicted in Table 3, or a functional part and/or equivalent of such a nucleic acid and/or amino acid sequence. A functional equivalent of said nucleic acid comprises the same hybridization properties in kind, not necessarily in amount, as said nucleic acid (or part thereof). A functional equivalent of an amino acid sequence of the invention comprises the same immunogenic properties in kind, not necessarily in amount, as said amino acid sequence (or part thereof). A part of a nucleic acid of the invention comprises at least 15 nucleotides, preferably at least 20, more preferably at least 30 nucleotides. A part of an amino acid sequence comprises at least 5 amino acids in peptidic linkage with each other, more preferably at least 8, and more preferably at least 12, more preferably at least 16 amino acids. In a preferred embodiment said nucleotides and/or amino acids are at least semi-consecutive, more preferably, said nucleotides and/or amino acids are consecutive. An equivalent of a nucleic acid and/or amino acid sequence of the invention or part thereof comprises at least 80% homology to a nucleic acid and/or amino acid sequence of the invention, preferably at least 90% homology, more preferably at least 95% and even more preferably at least 99% homology to a nucleic acid and/or amino acid sequence of the invention or a part thereof.

[0013] The invention further provides a primer and/or probe, capable of specifically hybridizing to a nucleic acid of a virus or functional part, derivative or analogue according to the invention, preferably a primer and/or probe, capable of specifically hybridizing to a nucleic acid sequence as depicted in Table 3. More preferably, a primer and/or probe, which is capable of hybridizing to said nucleic acid under stringent conditions. In a particular preferred embodiment is provided a primer and/or probe, comprising a sequence as depicted in Table 7.

[0014] The art knows many ways in which a specific binding member can be generated against an identified nucleic acid, lipid and/or amino acid sequence. Such specific binding members may be of any nature but are typically of a nucleic acid and/or proteinaceous nature. The invention thus further provides an isolated molecule capable of specifically binding a virus, nucleic acid and/or amino acid or functional part, derivative or analogue thereof according to the invention. Said isolated molecule is also referred to as specific binding member. Preferably said specific binding member is capable of specifically binding at least part of a nucleic acid sequence as depicted in table 3 and/or at least part of an amino acid sequence as depicted in Table 3. In a preferred embodiment said binding member is a proteinaceous molecule. Preferably an antibody or a functional part, derivative and/or analogue thereof. A specific binding member preferably comprises a significantly better binding property for the HCoV-NL63 virus compared to unrelated control. However, for instance for antibodies, it is possible that the epitope specifically recognized in HCoV-NL63 is also present in a limited number of other molecules. Thus though the binding of the binding member may be specific, it may recognize also other molecules than those present in HCoV-NL63. This cross-reactivity is to be separated from a-specific binding and is a general property of antibodies. Cross-reactivity does not usually hinder the selection of suitable specific binding members for particular purposes. For instance a specific binding member that also recognized a protein in liver cells can be used in many applications even in the presence of liver cells, where additional information such as location in

the cell can often be used to discriminate.

[0015] One source of an antibody of the invention is the blood of the infected subjects screened for the virus of the present invention. One may further characterize B-cells obtained from said subject. A suitable B-cell may be cultured and the antibody collected. Alternatively, the antibody may be sequenced from this B-cell and generated artificially. Another source of an antibody of the invention can be generated by immuninisation of test animals or using artificial libraries to screen a purified fraction of virus. A functional part of an antibody has essentially the same properties of said antibody in kind, not necessarily in amount. Said functional part is preferably capable of specifically binding an antigen of HCoV-NL63. However, said functional part may bind such antigen to a different extend as compared to said whole antibody. A functional part or derivative of an antibody for instance comprises a FAB fragment or a single chain antibody. An analogue of an antibody for instance comprises a chimeric antibody. As used herein, the term "antibody" is also meant to comprise a functional part, derivative and/or analogue of said antibody.

[0016] Once antibody of the invention is obtained, a desired property, such as its binding capacity, can be improved. This can for instance be done by an Ala-scan and/or replacement net mapping method. With these methods, many different proteinaceous molecules are generated, based on an original amino acid sequence but each molecule containing a substitution of at least one amino acid residue. Said amino acid residue may either be replaced by Alanine (Ala-scan) or by any other amino acid residue (replacement net mapping). Each variant is subsequently screened for said desired property. Generated data are used to design an improved proteinaceous molecule.

[0017] There are many different ways in which a specific binding member can be generated. In a preferred embodiment the invention provides a method for producing a specific proteinaceous binding member comprising producing proteinaceous molecules capable of binding a virus according to the invention or to a functional part, derivative or analogue, and selecting a proteinaceous molecule that is specific for said virus. If need be, the method may be used to generate a collection of proteinaceous molecules capable of binding to said virus or functional part, derivative and/or analogue thereof and selecting from said collection one or more binding members capable of specifically binding said virus or functional part, derivative and/or analogue thereof.

[0018] Any specific binding member is characteristic for the HCoV-NL63virus of the invention. Thus a virus that is specifically reactive with such binding member is an HCoV-NL63 virus and thus provided by the invention. Thus the invention provides an isolated and/or recombinant virus that is immunoreactive with specific binding member of the invention, preferably a proteinaceous binding member. The invention further provides a composition of matter comprising isolated HCoV-NL63 virus, and/or a virus essentially corresponding to HCoV-NL63. The term, a virus "essentially corresponding to HCoV-NL63" refers to HCoV-NL63 viruses which are either identical to the HCoV-NL63 strain described hereinabove, or which comprises one or more mutations compared to the said HCoV-NL63strain. These mutations may include natural mutations or artificial mutations. Said mutations of course should allow detection with a specific binding member of HCoV-NL63, not necessarily with all of the specific binding members). Said mutations should allow the detection of the variants using common detection methods such as antibody interaction, amplification and/or hybridization.

[0019] Considering that specific binding members are important molecules for instance for diagnostic purposes, the invention further provides the use of a virus of the invention or functional part, derivative and/or analogue thereof, for detecting a molecule capable of specifically binding said virus in a sample. Further provided is the use of a nucleic acid and/or amino acid sequence of a virus or functional part, derivative or analogue as defined by the invention, for detecting a molecule capable of specifically binding said virus or functional part, derivative and/or analogue in a sample. Preferably said nucleic acid and/or amino acid sequence comprises a sequence as depicted in table 3 or Table 3 or a functional part, derivative or analogue thereof. Preferably said part is at least 30 nucleotides and/or amino acids long wherein said part preferably comprises more than 95% sequence identity, preferably more than 99%. In a preferred aspect said specific binding member comprises a specific ligand and/or antibody of said virus.

[0020] Further provided is a primer and/or probe according to the invention, a specific binding member of the invention, and/or a nucleic acid of a virus or functional part, derivative or analogue according to the invention, for detecting and/or identifying a HCoV-NL63 coronavirus or part thereof in a sample. Preferably, said nucleic acid comprises a sequence as depicted in table 3.

[0021] HCoV-NL63 virus may be used to generate an immune response in a subject. This can be useful for instance in vaccination strategies. Thus the invention further HCoV-NL63 provides HCoV-NL63 virus or functional part, derivative or analogue thereof for use as a vaccine or medicament. The medicament use is typically when the subject is already infected with the virus and the immunogen is used to augment the immune response against the virus. The invention further provides a specific binding member of the invention for use as a vaccine or medicament. This use is particularly favorable for when the specific binding member comprises a proteinaceous molecule, preferably an antibody or functional part, derivative and/or analogue thereof. Such an antibody can provide passive immunity but may also have active components such as proteases attached to it. The medicament use may again be the case wherein a subject infected with an HCoV-NL63 virus is treated with the specific binding member.

[0022] Vaccines may be generated in a variety of ways. One way is to culture the HCoV-NL63 virus for example on

the mentioned monkey cell line(s) and to use inactivated virus harvested from the culture. Alternatively, attenuated virus may be used either inactivated or as a live vaccine. Methods for the generation of coronavirus vaccines may be adapted to produce vaccines for the HCoV-NL63 of the invention. The invention thus further provides the use of an HCoV-NL63 virus or functional part, derivative or analogue thereof for the preparation of a vaccine against a coronaviral genus related disease. The invention further provides the use of a specific binding member of the invention for the preparation of a vaccine or medicament against a coronaviral genus related disease. Further provided is the use of an HCoV-NL63 virus or functional part, derivative or analogue thereof, a specific binding member of the invention, a nucleic acid of the invention or a primer and/or probe of the invention for diagnosis of a coronaviral genus related disease. Preferably said coronaviral genus related disease comprises a HCoV-NL63coronavirus related disease.

[0023] Further provided is a vaccine comprising an HCoV-NL63 virus or functional part, derivative or analogue thereof and/or a specific binding member of the invention. Also provided is a medicament comprising an HCoV-NL63virus or functional part, derivative or analogue thereof and/or a specific binding member of the invention. Preferably said vaccine or medicament is used for at least in part preventing and/or treating a HCoV-NL63 related disease.

[0024] An important use of the present invention is the generation of a diagnostic tool for determining whether a subject is suffering from an HCoV-NL63 virus infection or has been exposed to an HCoV-NL63 virus infection. Many different diagnostic applications can be envisioned. They typically contain an identifying component allowing the typing of the virus that is or was present in the subject. One diagnostic tool for HCoV-NL63 makes use of the particular proliferation characteristics of the virus in various cell lines. It replicates in the mentioned preferred monkey cell lines but does not replicate in Vero- cells. This property can be used to discriminate HCoV-NL63 from other known coronaviruses. Thus in one aspect the invention provides a diagnostic kit comprising at least one of the preferred monkey cell lines, preferably the tertiary monkey kidney cells (tMK; Cynomolgus monkey or the monkey cell line LLC-MK2. Many modern diagnostic kits comprise a specific binding member (to detect the virus or virus infected cells) and/or an HCoV-NL63 virus or a functional part, derivative and/or analogue thereof and/or amino acid of the invention or a functional part, derivative and/or analogue thereof (for detecting antibodies in blood components of the diagnosed subject). Many other current diagnostic kits rely on identification of HCoV-NL63 virus specific nucleic acid in a sample. There are various ways in which such an assay may be implemented one is a method for detecting an HCoV-NL63 virus or functional part, derivative or analogue thereof in a sample, comprising hybridizing and/or amplifying a nucleic acid of said virus or functional part, derivative or analogue with a primer and/or probe according to the invention and detecting hybridized and/or amplified product. The invention thus also provides a diagnostic kit comprising an HCoV-NL63 virus or functional part, derivative or analogue thereof, a specific binding member according to the invention and/or a primer/probe according to the invention.

[0025] Further provided is a method for treating an individual suffering from, or at risk of suffering from, a HCoV-NL63 related disease, comprising administering to said individual a vaccine or medicament according to the invention. Also provided is a method for determining whether an individual suffers from a HCoV-NL63 related disease, comprising obtaining a sample from said individual and detecting a HCoV-NL63 virus or functional part, derivative or analogue thereof in said sample with a method and/or diagnostic kit of the invention.

[0026] Further provided is an isolated or recombinant nucleic acid encoding a virus or functional part, derivative and/or analogue according to the invention and a nucleic acid according to the invention, comprising at least a functional part of a sequence as depicted in Table 3. Further provided is an amino acid sequence encoded by a nucleic acid according to the invention, and an amino acid sequence according to the invention, comprising at least a functional part of a sequence as depicted in Table 3. Further provided is a proteinaceous molecule capable of specifically binding HCoV-NL63, obtainable by a method according to the invention and, the use of such a proteinaceous molecule in a vaccine or a diagnostic method for the detection of HCoV-NL63.

EXAMPLES

cDNA-AFLP for virus discovery

[0027] We modified the cDNA-AFLP technique such that it can amplify viral sequences from blood-plasma/serum samples or from CPE-positive culture supernatants (Figure 1). In the adjusted method the mRNA isolation step prior to amplification is replaced by a treatment to purify viral nucleic acid. Of importance to the purification is a centrifugation step to remove residual cells and mitochondria. In addition, a single DNAse treatment is sufficient to get rid of interfering chromosomal DNA and mitochondrial DNA from broken down cells and finally, by choosing frequent cutting restriction enzymes, the method is fine-tuned such that the majority of viruses will be amplified. With this so-called Virus Discovery cDNA-AFLP (VIDISCA) we were able to amplify viral nucleic acids from EDTA-plasma of a person with hepatitis B virus infection and a person with an acute Parvo B19 infection (results not shown). The technique can also detect HIV-1 in a positive culture supernatant demonstrating its capacity to identify both RNA and DNA viruses (results not shown).

[0028] To eliminate residual cells, 110 μl of virus culture supernatant was spun down for 10 min at maximum speed

in an Eppendorf microcentrifuge (13500 rpm). One hundred µl was transferred to a fresh tube and DNAse treated for 45 minutes at 37°C using 15 µl of DNAse buffer and 20 Units of DNAse I (Ambion). The DNAse treatment was included to get rid of chromosomal DNA from broken down cells. After this 900 µl of L6 lysis buffer and 40 µl of silica suspension was added and nucleic acids were extracted as described by Boom[4]. The viral nucleic acids were eluted in 40 µl H_2O. With 20 µl eluate the reverse transcription was performed using 2.5 µg random hexamers (Amersham Bioscience), 200 U MMLV-RT (Invitrogen) in a buffer containing 10 mM Tris-HCl pH 8.3, 50 mM KCl, 0.1% Triton X-100, 4.8 mM MgCl2, and 0.4 mM of each dNTP. The sample was incubated at 37°C for 90 minutes. Subsequently the second strand DNA synthesis was performed using 26 U Sequenase II (Amersham Bioscience), 7.5 U RNAse H (Amersham Bioscience) in 0.25 mM dNTPs each, 17.5 mM MgCl2 and 35 mM Tris-HCl pH 7.5. After the incubation at 37°C for 90 minutes a phenol/chloroform extraction was performed followed by an ethanol precipitation. The pellet was dissolved in 30 µl of H_2O. The cDNA-AFLP was performed essentially as described by Bachem[1] with some modifications. The dsDNA was digested with the HinP I and MseI restriction enzymes (New England Biolabs) according to the manufacturers protocol. After the digestion, MseI adaptor and HinP I adaptor (see below) are added together with 5U ligase enzyme (Invitrogen) and ligase buffer, followed by an additional incubation of 2 hrs at 37°C. The MseI adaptor and HinP I adaptor were prepared previously by mixing a top strand oligo for the MSE and the HinP1 adaptors (Table 1) with a bottom strand oligo for the MSE adaptor and for the HinP1 adaptor, incubate at 65° C. followed by cooling down to room temperature in the presence of a 1:40 dilution of ligase buffer.

[0029] The first PCR was performed with 10 µl of ligation mixture as input, 2.5 U of AmpliTaq polymerase (Perkin-Elmer), 100 ng of HinPI standard primer and 100 ng of MseI standard primer. The PCR reaction was performed according to the profile 5min 95C; 20 cycles of: 1min 95°C-1min 55°C-2min 72°C; 10 min 72 °C. Five µl of first PCR product was used as input in the second "selective" amplification step containing 100 ng of HinPI-N primer and 100 ng MseI-N (sequence of the standard primers extended with one nucleotide) and 2 U AmpliTaq polymerase. The selective PCRs were amplified according to the profile of the "touch down PCR": 10 cycles of 60 sec 94° C-30 sec 65° C-1 min 72°C over which the annealing temperature was reduced from 65°C with 1 °C with each cycle, followed by 23 cycles: 30 sec 94°C.30 sec 56 °C-1 min 72 °C. Finally the sample was incubated for 10 min at 72°C. The PCR products were evaluated on 4% Metaphor® gels (Cambrex, Rockland, USA). If the bands on the gel were very faint the PCR products were concentrated by vacuum drying using 60 µl of the PCR product. The PCR fragments of interest were cut out of gel and DNA was eluted from the gel using the Qiagen gel purification kit according to the manufacturer's protocol. The PCR products were cloned using pCR® 2.1-TOPO plasmid (Invitrogen) and chemically competent One Shot E. coli (Invitrogen). A PCR on the colony was performed and this PCR product was input for sequencing the insert using Big Dye terminator chemistry (Applied Biosystems). The reverse transcription step was excluded, only HinP I digestion and adaptor ligation was performed, the first PCR was performed with 35 cycles instead of 20 and those first PCR fragments were visualized on agarose gel electrophoresis.

DNA sequencing and analysis.

[0030] Coronavirus-PCR product containing plasmids were sequenced with the BigDyeTM Terminator Cycle Sequencing Kit (Applied Biosystems, Foster City, Calif.), using the -21 M13RP and T7 primers. Electrophoresis of sequencing reaction mixtures was performed with an Applied Biosystems 377 automated sequencer, following the manufacturer's protocols. The Sequence Navigator (version 1.01) and Auto Assembler (version 2.1) software packages (ABI, California, USA) were used to analyze all sequencing data. Sequences were compared to all sequences in the Genbank database using the BLAST tool of the NCBI webpage: http://www.ncbi.nlm.nih.gov/blast. For phylogenetic analysis the sequences were aligned using the ClustalX software package[34] with the following settings: Gap opening penalties: 10.00; Gap extension penalty 0.20, Delay divergent sequences switch at 30% and transition weight 0.59. Phylogenetic analysis was carried out using the neighbor-joining method of the MEGA program[35]. The nucleotide distance matrix was generated either by Kimura's 2-parameter estimation or by the p-distance estimation[36]. Bootstrap resampling (500 replications) was employed to place approximate confidence limits on individual branches.

Determining the nucleotide sequence of the complete HCoV-NL63 genome.

[0031] Using a combination of specific primers, located in the already sequenced domains of the HCoV-NL63 genome, and the proprietary PALM-method (WO 0151661) we are in the process of cloning and determining the full-length genomic sequence for this new coronavirus. Using a combination of 5'-oligonucleotides located in the analyzed part of the HCoV-NL63 genome and a 3' tagged random primer (JZH2R) additional fragments were amplified using a nested RT-PCR protocol similar to the one mentioned previously.

Isolation of SZ 163

[0032] In January 2003 a 7-month-old child appeared in hospital with coryza, conjunctivitis and fever. Chest radiography showed typical features of bronchiolotis and four days after the onset of disease a nasopharyngeal aspirate specimen was collected (sample nr: HCoV-NL63). All routinely used tests on this sample for adenovirus, respiratory syncytial virus (RSV), influenza A and B, parainfluenza 1, 2 and 3, rhinovirus, HCoV-229E and HCoV-OC43 were negative. The clinical sample was subsequently inoculated onto a variety of cells including human fibroblast lung (HFL) cells, tertiary monkey kidney cells (tMK; Cynomolgus) and R-HeLa cells. A CPE was detected exclusively on tMK cells and first noted at eight days post-inoculation. The CPE was diffuse with a refractive appearance in the affected cells followed by cell detachment after 7 days. More pronounced CPE was observed upon passage onto LLC-MK2 cells. Besides overall cell rounding, moderate cell enlargement was observed. Additional subcultering on human endothelial lung cells, HFL, Rhabdomyosarcoma cells and Vero cells remained negative for CPE. Immunofluorescent assays to detect influenzavirus A and B, RSV, adenoviruses or parainfluenza virus types 1, 2 or 3 in the culture remained negative The culture supernatant of infected LLC-MK2 cells was subsequently analyzed by VIDISCA. As control we used the supernatant of uninfected LLC-MK2 cells. After the second PCR amplification step, several DNA fragments were present in the test sample but not in the control. These fragments were cloned and sequenced. A Blast search in GenBank revealed that 8 of 16 fragments had sequence similarity to the family of corona viruses with the highest homology the human corona virus 229E (Tables 4 and 5).

[0033] Phylogenetic analysis of a 270 nt fragment of the replicase 1B region indicated that we identified a distinct new member of the coronavirus group 1. With the VIDISCA technique, 8 HCOV-163-specific fragments, named 163-2, 163-4, 163-9, 163-10, 163-11, 163-14, 163-15 and 163-18 were isolated, cloned, sequenced and aligned with the relevant sequences from GenBank. The Genbank accession number of the used sequences are: MHV (mouse hepatitis virus): AF201929; HCoV-229E: AF304460; PEDV (porcine epidemic diarrhea virus): AF353511; TGEV (transmissible gastroenteritis virus): AJ271965; SARS-CoV: AY278554; IBV (avian infectious bronchitis virus): NC_001451; BCoV (bovine coronavirus): NC_003045; FCoV (feline coronavirus): Y13921 and X80799; CCoV (canine coronavirus): AB105373 and A22732; PRCoV (porcine respiratory coronavirus): M94097; FIPV (feline infectious peritonitis virus): D32044. Position of the HCoV-NL63 fragments compared to HCoV-229E (AF304460): Replicase 1AB gene: 15155-15361, 16049-16182, 16190-16315, 18444-18550, Spike gene: 22124-22266; Nucleocapsid gene: 25667-25882 and 25887-25957; 3'UTR: 27052-27123. Branch lengths indicate the number of substitutions per sequence. From the most closely related species sequence identity scores were calculated (Tables 5 and 6).

[0034] Also the deduced amino acid sequence were aligned to the corresponding domains in the open reading frames of related corona (-like) viruses (Table 6).

[0035] The human corona viruses account for 10 to 30% of the common colds in man[7], and it is not unusual to find a coronavirus in a child with a respiratory illness. However, it is striking that the virus HCoV-NL63 was harvested from LLC-MK cells. Human Corona virus 229E and OC-43 are known for there inability to replicate on monkey kidney cells. Intriguingly, the newly identified human corona virus that is responsible for SARS is also able to replicate in monkey kidney cells 30.

Propagation of HCoV-NL63 in cell culture

[0036] A nasopharyngeal aspirate was collected 4 days after the onset of symptoms. The specimen was tested for the presence of adenovirus, RSV, influenza A, influenza B, and parainfluenza type 1, 2 an 3 using the Virus Respiratory Kit (Bartels: Trinity Biotech plc, Wicklow Ireland). In addition, PCR diagnosis for rhinoviruses, meta-pneumovirus and HCoV-OC43 and HCoV-229E were performed[2, 10]. The original nasopharyngeal aspirate was subsequently inoculated onto a variety of cell cultures including HFL cells, tMK cells and R-HeLa cells. The tubes were kept in a roller drum at 34°C and observed every 3 to 4 days. Maintenance medium was replenished every 3 to 4 days. Two different types of medium were implemented: Optimem 1 (Gibco) without bovine fetal serum was used for the tMK cells and MEM Hanks' /Earle's medium (Gibco) with 3% bovine fetal serum was used for the remaining cell types. On the virus culture direct staining was performed with pools of fluorescent-labeled mouse antibodies against influenzavirus A and B, RSV and adenoviruses (Imagen, DAKO). Indirect staining was performed for parainfluenza virus types 1, 2 or 3 with mouse antibodies (Chemicon, Brunschwig, Amsterdam Netherlands) and subsequent staining with labeled rabbit anti-mouse antibodies (Imagen, DAKO).

Method to detect HCoV-NL63 in nasopharyngeal swabs.

[0037] For the diagnostic RT-PCR, nucleic acids were extracted by the Boom method[4] 4 from 50 µl virus supernatant or 50 µl suspended nasopharyngeal swab. The reverse transcription was performed as described above with the exception that 10 ng of reverse transcription primer repSZ-RT (Table 7) was used. The entire RT mixture was added to

the first PCR mixture containing 100 ng of primer repSZ-1 and 100 ng of primer repSZ-3. The PCR reaction was performed according to the profile 5 min 95 °C; 20 cycles of: 1 min 95°C - 1min 55°C - 2 min 72°C; 10 min 72°C. A nested PCR was started using 5 μl of the first PCR with 100 ng of primer repSZ-2 and 100 ng of primer repSZ-4. Twenty-five PCR cycles were performed of the same profile as the first PCR. Ten μl of the first and 10 μl of the nested PCR was analyzed by agarose gel electrophoresis (Fig. 2). Cloning and sequencing of the fragments was performed essentially as described above.

Method of raising polyclonal antibodies

[0038] Appropriate domains within the HCoV-NL63 surface proteins (e.g. S-glycoprotein or HE- glycoprotein) can be selected and amplified with suitable oligonucleotides and RT-PCR. The corresponding purified viral antigens can be obtained by expression in a suitable host *(e.g. Yarrowia lipolytica* as previously described[38]). Female NZW rabbits (approx 4 kg) are primed with 0.5 to 5.0 mg of viral protein antigen preparation. The antigen is suspended in 0.5 ml. of phosphate buffered saline (pH 7.3) and emulsified in an equal volume of complete Freund's adjuvant (CFA). Freund's Adjuvant is a well-established adjuvant system that is appropriate for use in these experiments where small amounts of antigen are used, and where immunogenicity of the antigen (although likely) is unknown. Published guidelines for use will be followed, including limiting injection to 0.1 ml at each site, using CFA only for initial immunization dose. This antigen preparation (1 ml total volume) is injected subdermally in the loose skin on the backside of the rabbit's neck. This injection route is immunologically effective and minimizes the possibility of local inflammation associated with unilateral or bilateral flank injection (such ensuing flank inflammation can impair animal mobility). After resting for 3 weeks, one ml of blood will be removed from the ear artery for a test bleed. Antibodies will be boosted if titers of the desirable antibodies are judged to be too low. Rabbits with adequate antibody levels will be boosted subdermally 1.0 mg of antigen contained in CFA. Boosted animals will be bled after two weeks; i.e., 15 ml of blood will be taken from the ear artery using a heat lamp to dilate the blood vessel. The rabbit will be placed in a commercial restraint, tranquillized with xylazine not more than seven times in total after which the rabbit will be exsanguinated by cardiac puncture following anesthesia using xylazinelketamine.

Method for Vaccine production

[0039] For the production of a subunit vaccine the S-glycoprotein perhaps combined with the HE, M and N proteins, could be expressed in a suitable eukaryotic host (e.g. *Y. lipolytica* or LLC-MK2 cells) and purified using preferentially two small affinity tags (e.g. His-tag or the StrepII tag). After appropriate purification, the resulting viral proteins can be used as a subunit vaccine.

[0040] Alternatively the HCoV-NL63 virus can be propagated in a suitable cell line as described above and subsequently treated as described by Wu [11]. Briefly the virus is precipitated from culture medium with 20% polyethylene glycol 6000 and purified by ultracentrifugation at 80.000 × g for 4 hours through a discontinuous 40-65% sucrose gradient followed by a linear 5 to 40 % CsCl gradient for 4 hours at 120.000 × g. The resulting virus preparation can be inactivated by heating for 30 minutes at 65° C as described by Blondel[3].

[0041] *Analysis of S glycoprotein* or any *of the HCOV-NL63 viral proteins binding to an immobilized ligand (e.g. antibody) in* an *optical biosensor.*

Binding reactions were carried out in an IAsys two-channel resonant mirror biosensor at 20°C (Affinity Sensors, Saxon Hill, Cambridge, United Kingdom) with minor modifications. Planar biotin surfaces, with which a signal of 600 arc s corresponds to 1 ng of bound protein/mm2, were derivatized with streptavidin according to the manufacturer's instructions. Controls showed that the viral proteins did not bind to streptavidin-derivatized biotin surfaces (result not shown). Biotinylated antibody was immobilized on planar streptavidin-derivatized surfaces, which were then washed with PBS. The distribution of the immobilized ligand and of the bound S-glycoprotein on the surface of the biosensor cuvette was inspected by the resonance scan, which showed that at all times these molecules were distributed uniformly on the sensor surface and therefore were not micro-aggregated. Binding assays were conducted in a final volume of 30 μl of PBS at 20 ± 0.1°C. The ligate was added at a known concentration in 1 μl to 5 μl of PBS to the cuvette to give a final concentration of S-glycoprotein ranging from 14 to 70 nM. To remove residual bound ligate after the dissociation phase, and thus regenerate the immobilized ligand, the cuvette was washed three times with 50 μl of 2 M NaCl-10 mM Na2HPO4, pH 7.2, and three times with 50 μl of 20 mM HCl. Data were pooled from experiments carried out with different amounts of immobilized antibody (0.2, 0.6, and 1.2 ng/mm2). For the calculation of k_{on}, low concentrations of ligate (S-glycoprotein) were used, whereas for the measurement of k_{off}, higher concentrations of ligate were employed (1 μM) to avoid any rebinding artefacts. The binding parameters k_{on} and k_{off} were calculated from the association and dissociation phases of the binding reactions, respectively, using the non-linear curve-fitting FastFit software (Affinity Sensors) provided with the instrument. The dissociation constant (K_d) was calculated from the association and dissociation rate constants and from the extent of binding observed near equilibrium.

Tables

[0042]

Table 1: cDNA- AFLP oligonucleotides for virus discovery

Oligo	Sequence
Top strand MSE adaptor	CTCGTAGACTGCGTACC
Top strand for HinP1 adaptor	GACGATGAGTCCTGAC
Bottom strand oligo for MSE adaptor	TAGGTACGCAGTC
Bottom strand oligo for HinP1 adaptor	CGGTCAGGACTCAT
HinPI standard primer	GACGATGAGTCCTGACCGC
MseI standard primer	CTCGTAGACTGCGTACCTAA

Table 2: Oligonucleotide for PALM extension of the HCOV-NL63 Sequence

Oligonucleotide name,	*Application*,	Sequence 5'- 3'
JZH2R	1st PCR	GCTATCATCACAATGGACNNNNNG

Table 3 Nucleotide- and corresponding deduced amino acid sequences

Fragment	Sequence
163-2	GTATTGTTTTTGTTGCTTGTGCCCATGCTGCTGTTGATTCCTTATGTGCAAAAGCTATGA CTGTTTATAGCATTGATAAGTGTACTAGGATTATACCTGCAAGAGCTCGGGTTGAGTGTT ATAGTGGCT
163-2 Translation	Replicase polyprotein 1a IVFVACAHAAVDSLCAKAMTVYSIDKCTRIIPARARVECYSG
163-4	ATGGGTCTAGATATGGCTTGCAAAACTTACTACAGTTACCTAACTTTTATTATGTTAGTA ATGGTGGTAACAATTGCACTACGGCCGTTATGACCTATTCTAATTTTGGTATTTGTGCTG ATGGTTCTTTGATTCCTGTTCGTCC
163-4 Translation	Spike protein GSRYGLQNLLQLPNFYYVSNGGNNCTTAVMTYSNFGICADGSLIPVR
163-9 (3'-UTR)	ATGATAAGGGTTTAGTCTTACACACAATGGTAGGCCAGTGATAGTAAAGTGTAAGTAATT TGCTATCATAT
163-10	ATGTCAGTGATGCATATGCTAATTTGGTTCCATATTACCAACTTATTGGTAAACAAAAGA TAACTACAATACAGGGTCCTCCTGGTAGTGGTAAGTCACATTGTTCCATTGGACTTGGAT TGTACTACCCAGGT
163-10 Translation	Replicase polyprotein 1ab VSDAYANLVPYYQLIGKQKITTIQGPPGSGKSHCSIGLGLYYPG
163-11	ATCTAAACTAAACAAAATGGCTAGTGTAAATTGGGCCGATGACAGAGCTGCTAGGAAGAA ATTTCCTCCTCCTTCATTTTACATGCCTCTTTTGGTTAGTTCTGATAAGGCACCATATAG GGTCATTCCCAGGAATCTTGTCCCTATTGGTAAGGGTAATAAAGATGAGCAGATTGGTTA TTGGAATGTTCAAGAGCGTTGGCGTAT
163-11 Translation	Nucleocapsid protein SKLNKMASVNWADDRAARKKFPPPSFYMPLLVSSDKAPYRVIPRNLVPIGKGNKDEQIGY WNVQERWR
163-14	ACAAAAATTTGAATGAGGGTGTTCTTGAATCTTTTTCTGTTACACTTCTTGATAATCAAG AAGATAAGTTTTGGTGTGAAGATTTTTATGCTAGTATGTATGAAAATTCTACAATATTGC AAGCTGCTGGTTTATGTGTTGTTTGTGGTTCACAAACTGTACTTCGTTGTGGTGATTGTC TGCGTAAGCCTATGTTGTGCACTAAAT
163-14 Translation	Replicase polyprotein 1ab KNLNEGVLESFSVTLLDNQEDKFWCEDFYASMYENSTILQAAGLCVVCGSQTVLRCGDCL RKPMLCTK
163-15	AGGGGGCAACGTGTTGATTTGCCTCCTAAAGTTCATTTTTATTACCTAGGTACTGGACCT CATAAGGACCT

163-15 Translation	**Nucleocapsid protein** RGQRVDLPPKVHFYYLGTGPHKD
163-18	TAGTAGTTGTGTTACTCGTTGTAATATAGGTGGTGCTGTTTGTTCAAAACATGCAAATTT GTATCAAAAATACGTTGAGGCATATAATACATTTACACAGGCAGGTT
163-18 Translation	**Replicase polyprotein 1ab** SSCVTRCNIGGAVCSKHANLYQKYVEAYNTFTQAG

Table 4

Identification of cDNA-AFLP fragments	
Fragment	Identification best Blast hit
163-2	replicase polyprotein 1ab [Human coronavirus 229E]
163-4	spike protein [Human coronavirus 229E]
163-9	3'UTR Human coronavirus 229E
163-10	replicase polyprotein 1ab [Human coronavirus 229E]
163-11	replicase polyprotein 1ab [Human coronavirus 229E]
163- 14	replicase polyprotein 1ab [Human coronavirus 229E]
163-15	nucleocapsid protein [Human coronavirus 229E]
163-18	replicase polyprotein 1ab [Human coronavirus 229E]

Table 5:

Pairwise nucleotide sequence homologies between the virus of the present invention and different corona (like) viruses in percentages sequence identity (%)							
Fragment	BcoV	MHV	HcoV	PEDV	TGE	SARS	IBV
Replicase 1AB 163-2	59.6	61.2	76.7	70.5	64.3	65.8	64.3
Spike gene 163-4	31.7	26.5	64.6	48.9	45.4	33.7	25.9
3'UTR 163-9	29.5	34	81.9	53.6	50	31.5	38
Replicase 1AB 163-10	55.2	57.4	82	73.8	69.4	64.1	65.1
Nucleocapsid 163-11	25.5	23.8	54.9	51.5	44.6	23.3	27.6
Replicase 1AB 163-14	52.1	52.1	78.7	72.9	76.3	52.6	58.4
Nucleocapsid 163-15	29.5	35.2	71.8	63.3	60.5	25.3	45
Replicase 1AB 163-18	67.2	65.4	72.8	65.4	61.6	68.2	57

Table 6:

Pairwise deduced amino acid sequence homologies between different corona (like) viruses in percentages sequence identity (%)							
Fragment	BCoV	MHV	HcoV	PEDV	TGE	SARS	IBV
Replicase 1AB 163-2	55.8	53.4	88.3	79	60.4	67.4	55.8
Spike gene 163-4	ND	ND	56.2	ND	ND	ND	ND
Replicase 1AB 163-10	51.1	53.3	93.3	86.6	80	57.7	55.5
Nucleocapsid 163-11	ND	ND	48.4	ND	ND	ND	ND
Replicase 1AB 163-14	50.7	50.7	86.9	78.2	78.2	46.3	47.8
Nucleocapsid 163-15	ND	ND	82.6	ND	ND	ND	ND
Nucleocapsid 163-18	63.8	63.8	77.7	69.4	69.4	58.3	55.5
ND = Not Determined							

Table 7: Oligos for specific detection of HcoV-163

Primer	Sequence
repSZ-RT	CCACTATAAC
repSZ-1	GTGATGCATATGCTAATTTG
repSZ-2	TTGGTAAACAAAAGATAACT
repSZ-3	CTCTTGCAGGTATAATCCTA
repSZ-4	TCAATGCTATAAACAGTCAT

Brief description of the drawings

[0043]

Figure 1: cDNA-AFLP allows amplification of nucleic acids without any prior sequence information. Culture supernatants from CPE-positive and uninfected cells are subjected to the cDNA-AFLP procedure. Amplification products derived from the CPE-positive culture which are not present in the uninfected control sample are cloned and sequenced.

Figure 2, LLC-MK2 cells infected with HCoV-SLA163. Panel A and B are unstained cells while panel C and D are stained with haematoxilin eosin. The typical CPE of HCoV-SLA163 is shown in panel A and C. The control uninfected LLC-MK cells are shown in panel B and D.

Figure 3, VD-cDNA-AFLP PCR products visualized by Metaphor® agarose gel electrophoreses. The PCR products of 1 (HinP I-G and Mse I-A) of 16 primer pair combinations used during the selective amplification step. Lanes 1 and 2: duplicate PCR product of virus culture SLA163; lanes 5 and 6 control supernatant of LLC-MK2 cells and in lane 7 and 8 the negative PCR control. Lanes M: 25bp molecular weight marker (InVitrogen). The arrow indicates a new coronavirus fragment that was excised out of gel and sequenced.

Figure 4, Phylogenetic analysis of the HCoV-163 sequences. G1, G2 and G3 denote the group 1, group 2 and group 3 coronavirus clusters. The Genbank accession number of the used sequences are: MHV (mouse hepatitis

virus): AF201929; HCoV-229E: AF304460; PEDV (porcine epidemic diarrhea virus): AF353511; TGEV (transmissible gastroenteritis virus): AJ271965; SARS-CoV: AY278554; IBV (avian infectious bronchitis virus): NC_001451; BCoV (bovine coronavirus): NC_003045; FCoV (feline coronavirus): Y13921 and X80799; CCoV (canine coronavirus): AB105373 and A22732; PRCoV (porcine respiratory coronavirus) : M94097; FIPV (feline infectious peritonitis virus): D32044. Position of the HCoV-163 fragments compared to HCoV-229E (AF304460): Replicase 1AB gene: 15155-15361, 16049-16182, 16190-16315, 18444-18550, Spike gene: 22124-22266; Nucleocapsid gene: 25667-25882 and 25887-25957; 3'UTR: 27052-27123. Branch lengths indicate the number of substitutions per sequence.

Figure 5: Schematic representation of Coronavirus and the location of the 163-fragments listed in table 3.

Reference List

[0044]

1. Bachem, C.W., R.S. van der Hoeven, S.M. de Bruijn, D. Vreugdenhil, M. Zabeau, and R.G. Visser. 1996. Visualization of differential gene expression using a novel method of RNA fingerprinting based on AFLP: analysis of gene expression during potato tuber development. Plant J. 9:745-753.

2. Bestebroer, T.M., A.I.M. Bartelds, A.M. van Loon, H. Boswijk, K. Bijlsma, E.C.J. Claas, J.A.F.W. Kleijne, C. Verweij, M.W. Verweij-Uijterwaal, A.G. Wermenbol, and J. de Jong,. Virological NIVEL/RIVM-surveillance of respiratory virus infection in the season 1994/95. 245607002, 1-38. 1995. Bilthoven, RIVM. Virologische NIVEL/RIVM-surveillance van respiratoire virusinfecties in het seizoen 1994/95 RIVM. Ref Type: Report

3. Blondel, B., O. Akacem, R. Crainic, P. Couillin, and F. Horodniceanu. 1983. Detection by monoclonal antibodies of an antigenic determinant critical for poliovirus neutralization present on VP1 and on heat-inactivated virions. Virology 126:707-710.

4. Boom, R., C. J. Sol, M. M. Salimans, C. L. Jansen, P. M. Wertheim-van Dillen, and van der Noordaa J. 1990. Rapid and simple method for purification of nucleic acids. J. Clin. Microbiol. 28:495-503.

5. Kamur, S., Tamura, K., and Wei, M. Molecular Evolutionary Genetics Analysis (MEGA 2.0). 1993. Institute of Molecular Evolutionary Genetics, Pennsylvania State University, University Park. Ref Type: Computer Program

6. Kimura, M. 1980. A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. J. Mol. Evol. 16:111-120.

7. Kunkel, F. and G. Herrler. 1993. Structural and functional analysis of the surface protein of human coronavirus OC43. Virology 195:195-202.

8. Mounir, S., P. Labonte, and P. J. Talbot. 1993. Characterization of the nonstructural and spike proteins of the human respiratory coronavirus OC43: comparison with bovine enteric coronavirus. Adv. Exp. Med. Biol. 342:61-67.

9. Thompson, J. D., T. J. Gibson, F. Plewniak, F. Jeanmougin, and D. G. Higgins. 1997. The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res. 25:4876-4882.

10. Van Den Hoogen, B. G., J. C. de Jong, J. Groen, T. Kuiken, R. de Groot, R. A. Fouchier, and A. D. Osterhaus. 2001. A newly discovered human pneumovirus isolated from young children with respiratory tract disease. Nat. Med. 7:719-724.

11. Wu, C. N., Y. C. Lin, C. Fann, N. S. Liao, S. R. Shih, and M. S. Ho. 2001. Protection against lethal enterovirus 71 infection in newborn mice by passive immunization with subunit VP1 vaccines and inactivated virus. Vaccine 20:895-904.

13. Almeida, J.D. and D.A. Tyrrell, The morphology of three previously uncharacterized human respiratory viruses that grow in organ culture. J Gen Virol 1, 175-178 (1967).

14. Thiel, V., J.Herold, , B. Schelle, and S.G. Siddell, Infectious RNA transcribed in vitro from a cDNA copy of the human coronavirus genome cloned in vaccinia virus. J Gen Virol 82, 1273-1281 (2001).

15. Hendley, J.O., H.B. Fishburne, and J.M. Gwaltney, Jr. Coronavirus infections in working adults. Eight-year study with 229 E and OC 43. Am Rev. Respir. Dis. 105, 805-811 (1972).

16. Mounir, S., P. Labonte, and P.J. Talbot, Characterization of the nonstructural and spike proteins of the human respiratory coronavirus OC43: comparison with bovine enteric coronavirus. Adv. Exp Med Biol 342, 61-67 (1993).

17. Kunkel, F. and G. Herrler, Structural and functional analysis of the surface protein of human coronavirus OC43. Virol. 195, 195-202 (1993).

18. Tyrrell, D.A.J. and M.L. Bynoe, Cultivation of novel type of common-cold virus in organ cultures. Br. Med J 1, 1467-1470 (1965).

19. Bradburne, A.F., M.L. Bynoe, and D.A. Tyrrell, Effects of a "new" human respiratory virus in volunteers. Br. Med J 3, 767-769 (1967).

20. Kapikian, A.Z. et al. Isolation from man of "avian infectious bronchitis virus-like" viruses (coronaviruses) similar to 229E virus, with some epidemiological observations. J Infect. Dis. 119, 282-290 (1969).
21. Ksiazek, T.G. et al. A novel coronavirus associated with severe acute respiratory syndrome. N Engl J Med 2003. May 15. ;348. (20):1953. -66. 348, 1953-1966 (2003).
22. Stohlman, S.A. and D.R. Hinton, Viral induced demyelination. Brain Pathol. 11, 92-106 (2001).
23. Jubelt, B. and J.R. Berger, Does viral disease underlie ALS? Lessons from the AIDS pandemic. Neurology 57, 945-946 (2001).
24. Shingadia, D., A. Bose, and R. Booy, Could a herpesvirus be the cause of Kawasaki disease? Lancet Infect. Dis. 2, 310-313 (2002).
25. Bachem, C.W. et al. Visualization of differential gene expression using a novel method of RNA fingerprinting based on AFLP: analysis of gene expression during potato tuber development. Plant J 9, 745-753 (1996).
26. Hamparian, V.V. Diagnostic procedures for viral, rickettsial and chlamydial infection. E.H. Lennette, and N.J. Schmidt, (eds.), pp. 5621979).
27. Marra, M.A. et al. The Genome sequence of the SARS-associated coronavirus. Science 2003. May 30. ;300. (5624.):1399. -404. 300, 1399-1404 (2003).
28. McIntosh, K. et al. Coronavirus infection in acute lower respiratory tract disease of infants. J Infect. Dis. 130, 502-507 (1974).
29. Boivin, G. et al. Human metapneumovirus infections in hospitalized children. Emerg. Infect. Dis. 9, 634-640 (2003).
30. Rota, P.A. et al. Characterization of a novel coronavirus associated with severe acute respiratory syndrome. Science 2003. May 30. ;300. (5624.):1394. - 9. 300, 1394-1399 (2003).
31. Bestebroer, T.M. et al. Virological NIVEL/RIVM-surveillance of respiratory virus infection in the season 1994/95. 245607002, 1-38. 1995. Ref Type: Report
32. van den Hoogen, B.G. et al. A newly discovered human pneumovirus isolated from young children with respiratory tract disease. Nat. Med 7, 719-724 (2001).
34. Earley, E. M. and K. M. Johnson. 1988. The lineage of Vero, Vero 76 and its clone C1008 in the United States., p. 26-29. In B. Simizu and T. Terasima (eds.), Vero cells: origin, properties and biomedical applications. Chiba Univ, Tokyo. 35. Kamur, S., K. Tamura, and M. Wei, Molecular Evolutionary Genetics Analysis (MEGA). (2.0). 1993. Institute of Molecular Evolutionary Genetics, Pennsylvania State University, University Park. Ref Type: Computer Program
36. Kimura, M. A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. J Mol Evol. 16, 111-120 (1980).
37 . Fouchier, R. A., T. M. Bestebroer, S. Herfst, K. L. Van Der, G. F. Rimmelzwaan, and A. D. Osterhaus. 2000. Detection of influenza A viruses from different species by PCR amplification of conserved sequences in the matrix gene. J. Clin. Microbiol. 38:4096-4101.
38. Nicaud, J. M., C. Madzak, B. P. van den, C. Gysler, P. Duboc, P. Niederberger, and C. Gaillardin. 2002. Protein expression and secretion in the yeast Yarrowia lipolytica. FEM. Yeast Res. 2:371-379

SEQUENCE LISTING

<110> PrimaGen Holding B.V.

<120> Coronavirus, nucleic acid, protein, and methods for the generation of vaccine, medicaments and diagnostics.

<130> P65647EP00

<140> 03077602.5
<141> 2003-08-18

<160> 27

<170> PatentIn Ver. 2.1

<210> 1
<211> 17
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: top strand MSE adaptor

<220>
<221> misc_feature
<222> (1)..(17)

<400> 1
ctcgtagact gcgtacc 17

<210> 2
<211> 16
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: top strand for HinP1 adaptor

<220>
<221> misc_feature
<222> (1)..(16)

<400> 2
gacgatgagt cctgac 16

<210> 3
<211> 13
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: bottom strand oligo for MSE adaptor

<220>
<221> misc_feature

<222> (1)..(13)

<400> 3
taggtacgca gtc 13

<210> 4
<211> 14
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: bottom strand oligo for HinP1 adaptor

<220>
<221> misc_feature
<222> (1)..(14)

<400> 4
cggtcaggac tcat 14

<210> 5
<211> 19
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: HinPI standard primer

<220>
<221> misc_feature
<222> (1)..(19)

<400> 5
gacgatgagt cctgaccgc 19

<210> 6
<211> 20
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: MseI standard primer

<220>
<221> misc_feature
<222> (1)..(20)

<400> 6
ctcgtagact gcgtacctaa 20

<210> 7
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:
oligonucleotide J2H2R

<220>
<221> misc_feature
<222> (1)..(24)

<400> 7
gctatcatca caatggacnn nnng 24

<210> 8
<211> 129
<212> DNA
<213> Human coronavirus

<220>
<223> Description of Combined DNA/RNA Molecule: part of
replicase polyprotein 1ab

<220>
<221> CDS
<222> (3)..(128)

<400> 8
gt att gtt ttt gtt gct tgt gcc cat gct gct gtt gat tcc tta tgt 47
Ile Val Phe Val Ala Cys Ala His Ala Ala Val Asp Ser Leu Cys
1 5 10 15

gca aaa gct atg act gtt tat agc att gat aag tgt act agg att ata 95
Ala Lys Ala Met Thr Val Tyr Ser Ile Asp Lys Cys Thr Arg Ile Ile
20 25 30

cct gca aga gct cgg gtt gag tgt tat agt ggc t 129
Pro Ala Arg Ala Arg Val Glu Cys Tyr Ser Gly
35 40

<210> 9
<211> 42
<212> PRT
<213> Human coronavirus
<223> Description of Combined DNA/RNA Molecule: part of
replicase polyprotein 1ab

<400> 9
Ile Val Phe Val Ala Cys Ala His Ala Ala Val Asp Ser Leu Cys Ala
1 5 10 15

Lys Ala Met Thr Val Tyr Ser Ile Asp Lys Cys Thr Arg Ile Ile Pro
20 25 30

Ala Arg Ala Arg Val Glu Cys Tyr Ser Gly
35 40

<210> 10
<211> 145

<212> DNA
<213> Human coronavirus

<220>
<223> Description of Combined DNA/RNA Molecule: part of
spike protein

<220>
<221> CDS
<222> (3)..(143)

<400> 10
at ggg tct aga tat ggc ttg caa aac tta cta cag tta cct aac ttt 47
Gly Ser Arg Tyr Gly Leu Gln Asn Leu Leu Gln Leu Pro Asn Phe
1 5 10 15

tat tat gtt agt aat ggt ggt aac aat tgc act acg gcc gtt atg acc 95
Tyr Tyr Val Ser Asn Gly Gly Asn Asn Cys Thr Thr Ala Val Met Thr
20 25 30

tat tct aat ttt ggt att tgt gct gat ggt tct ttg att cct gtt cgt 143
Tyr Ser Asn Phe Gly Ile Cys Ala Asp Gly Ser Leu Ile Pro Val Arg
35 40 45

cc 145

<210> 11
<211> 47
<212> PRT
<213> Human coronavirus
<223> Description of Combined DNA/RNA Molecule: part of
spike protein

<400> 11
Gly Ser Arg Tyr Gly Leu Gln Asn Leu Leu Gln Leu Pro Asn Phe Tyr
1 5 10 15

Tyr Val Ser Asn Gly Gly Asn Asn Cys Thr Thr Ala Val Met Thr Tyr
20 25 30

Ser Asn Phe Gly Ile Cys Ala Asp Gly Ser Leu Ile Pro Val Arg
35 40 45

<210> 12
<211> 71
<212> DNA
<213> Human coronavirus

<220>
<221> misc_feature
<222> (1)..(71)
<223> /note="Part of 3'-UTR"

<400> 12
atgataaggg tttagtctta cacacaatgg taggccagtg atagtaaagt gtaagtaatt 60
tgctatcata t 71

<210> 13
<211> 134
<212> DNA
<213> Human coronavirus

<220>
<223> Description of Combined DNA/RNA Molecule: part of replicase polyprotein 1ab

<220>
<221> CDS
<222> (3)..(134)

<400> 13
at gtc agt gat gca tat gct aat ttg gtt cca tat tac caa ctt att 47
Val Ser Asp Ala Tyr Ala Asn Leu Val Pro Tyr Tyr Gln Leu Ile
1 5 10 15

ggt aaa caa aag ata act aca ata cag ggt cct cct ggt agt ggt aag 95
Gly Lys Gln Lys Ile Thr Thr Ile Gln Gly Pro Pro Gly Ser Gly Lys
20 25 30

tca cat tgt tcc att gga ctt gga ttg tac tac cca ggt 134
Ser His Cys Ser Ile Gly Leu Gly Leu Tyr Tyr Pro Gly
35 40

<210> 14
<211> 44
<212> PRT
<213> Human coronavirus
<223> Description of Combined DNA/RNA Molecule: part of replicase polyprotein 1ab

<400> 14
Val Ser Asp Ala Tyr Ala Asn Leu Val Pro Tyr Tyr Gln Leu Ile Gly
1 5 10 15

Lys Gln Lys Ile Thr Thr Ile Gln Gly Pro Pro Gly Ser Gly Lys Ser
20 25 30

His Cys Ser Ile Gly Leu Gly Leu Tyr Tyr Pro Gly
35 40

<210> 15
<211> 207
<212> DNA
<213> Human coronavirus

<220>
<223> Description of Combined DNA/RNA Molecule: part of replicase polyprotein 1ab

<220>
<221> CDS
<222> (2)..(205)

<400> 15
a tct aaa cta aac aaa atg gct agt gta aat tgg gcc gat gac aga gct 49

```
  Ser Lys Leu Asn Lys Met Ala Ser Val Asn Trp Ala Asp Asp Arg Ala
    1               5                  10                  15

gct agg aag aaa ttt cct cct cct tca ttt tac atg cct ctt ttg gtt   97
Ala Arg Lys Lys Phe Pro Pro Pro Ser Phe Tyr Met Pro Leu Leu Val
             20                  25                  30

agt tct gat aag gca cca tat agg gtc att ccc agg aat ctt gtc cct   145
Ser Ser Asp Lys Ala Pro Tyr Arg Val Ile Pro Arg Asn Leu Val Pro
         35                  40                  45

att ggt aag ggt aat aaa gat gag cag att ggt tat tgg aat gtt caa   193
Ile Gly Lys Gly Asn Lys Asp Glu Gln Ile Gly Tyr Trp Asn Val Gln
     50                  55                  60

gag cgt tgg cgt at                                               207
Glu Arg Trp Arg
 65
```

<210> 16
<211> 68
<212> PRT
<213> Human coronavirus
<223> Description of Combined DNA/RNA Molecule: part of replicase polyprotein lab

<400> 16

```
Ser Lys Leu Asn Lys Met Ala Ser Val Asn Trp Ala Asp Asp Arg Ala
  1               5                  10                  15

Ala Arg Lys Lys Phe Pro Pro Pro Ser Phe Tyr Met Pro Leu Leu Val
             20                  25                  30

Ser Ser Asp Lys Ala Pro Tyr Arg Val Ile Pro Arg Asn Leu Val Pro
         35                  40                  45

Ile Gly Lys Gly Asn Lys Asp Glu Gln Ile Gly Tyr Trp Asn Val Gln
     50                  55                  60

Glu Arg Trp Arg
 65
```

<210> 17
<211> 207
<212> DNA
<213> Human coronavirus

<220>
<223> Description of Combined DNA/RNA Molecule: part of replicase polyprotein lab

<220>
<221> CDS
<222> (3)..(206)

<400> 17

```
ac aaa aat ttg aat gag ggt gtt ctt gaa tct ttt tct gtt aca ctt    47
   Lys Asn Leu Asn Glu Gly Val Leu Glu Ser Phe Ser Val Thr Leu
```

```
     1               5                  10                  15

ctt gat aat caa gaa gat aag ttt tgg tgt gaa gat ttt tat gct agt   95
Leu Asp Asn Gln Glu Asp Lys Phe Trp Cys Glu Asp Phe Tyr Ala Ser
                 20                  25                  30

atg tat gaa aat tct aca ata ttg caa gct gct ggt tta tgt gtt gtt   143
Met Tyr Glu Asn Ser Thr Ile Leu Gln Ala Ala Gly Leu Cys Val Val
             35                  40                  45

tgt ggt tca caa act gta ctt cgt tgt ggt gat tgt ctg cgt aag cct   191
Cys Gly Ser Gln Thr Val Leu Arg Cys Gly Asp Cys Leu Arg Lys Pro
         50                  55                  60

atg ttg tgc act aaa t                                            207
Met Leu Cys Thr Lys
     65
```

<210> 18
<211> 68
<212> PRT
<213> Human coronavirus
<223> Description of Combined DNA/RNA Molecule: part of replicase polyprotein 1ab

<400> 18

```
Lys Asn Leu Asn Glu Gly Val Leu Glu Ser Phe Ser Val Thr Leu Leu
  1               5                  10                  15

Asp Asn Gln Glu Asp Lys Phe Trp Cys Glu Asp Phe Tyr Ala Ser Met
             20                  25                  30

Tyr Glu Asn Ser Thr Ile Leu Gln Ala Ala Gly Leu Cys Val Val Cys
         35                  40                  45

Gly Ser Gln Thr Val Leu Arg Cys Gly Asp Cys Leu Arg Lys Pro Met
     50                  55                  60

Leu Cys Thr Lys
 65
```

<210> 19
<211> 71
<212> DNA
<213> Human coronavirus

<220>
<223> Description of Combined DNA/RNA Molecule: part of nucleocapsid protein

<220>
<221> CDS
<222> (1)..(69)

<400> 19

```
agg ggg caa cgt gtt gat ttg cct cct aaa gtt cat ttt tat tac cta   48
Arg Gly Gln Arg Val Asp Leu Pro Pro Lys Val His Phe Tyr Tyr Leu
  1               5                  10                  15
```

ggt act gga cct cat aag gac ct 71
Gly Thr Gly Pro His Lys Asp
20

<210> 20
<211> 23
<212> PRT
<213> Human coronavirus
<223> Description of Combined DNA/RNA Molecule: part of nucleocapsid protein

<400> 20
Arg Gly Gln Arg Val Asp Leu Pro Pro Lys Val His Phe Tyr Tyr Leu
1 5 10 15

Gly Thr Gly Pro His Lys Asp
20

<210> 21
<211> 107
<212> DNA
<213> Human coronavirus

<220>
<223> Description of Combined DNA/RNA Molecule: part of replicase polyprotein 1ab

<220>
<221> CDS
<222> (2)..(106)

<400> 21
t agt agt tgt gtt act cgt tgt aat ata ggt ggt gct gtt tgt tca aaa 49
Ser Ser Cys Val Thr Arg Cys Asn Ile Gly Gly Ala Val Cys Ser Lys
1 5 10 15

cat gca aat ttg tat caa aaa tac gtt gag gca tat aat aca ttt aca 97
His Ala Asn Leu Tyr Gln Lys Tyr Val Glu Ala Tyr Asn Thr Phe Thr
20 25 30

cag gca ggt t 107
Gln Ala Gly
35

<210> 22
<211> 35
<212> PRT
<213> Human coronavirus
<223> Description of Combined DNA/RNA Molecule: part of replicase polyprotein 1ab

<400> 22
Ser Ser Cys Val Thr Arg Cys Asn Ile Gly Gly Ala Val Cys Ser Lys
1 5 10 15

His Ala Asn Leu Tyr Gln Lys Tyr Val Glu Ala Tyr Asn Thr Phe Thr

20 25 30

Gln Ala Gly
35

<210> 23
<211> 10
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: primer repSZ-RT

<220>
<221> misc_feature
<222> (1)..(10)

<400> 23
ccactataac 10

<210> 24
<211> 20
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: primer repSZ-1

<220>
<221> misc_feature
<222> (1)..(20)

<400> 24
gtgatgcata tgctaatttg 20

<210> 25
<211> 20
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: primer repSZ-2

<220>
<221> misc_feature
<222> (1)..(20)

<400> 25
ttggtaaaca aaagataact 20

<210> 26
<211> 20
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: primer repSZ-3

<220>
<221> misc_feature
<222> (1)..(20)

<400> 26
ctcttgcagg tataatccta 20

<210> 27
<211> 20
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: primer repSZ-4

<220>
<221> misc_feature
<222> (1)..(20)

<400> 27
tcaatgctat aaacagtcat 20

Claims

1. An isolated or recombinant virus comprising a nucleic acid sequence as depicted in Table 3, or a functional part, derivative or analogue of said virus.

2. An isolated or recombinant virus comprising an amino acid sequence as depicted in Figure 1, or a functional part, derivative or analogue of said virus.

3. An isolated or recombinant virus comprising a nucleic acid sequence encoding an amino acid sequence as depicted in Figure 1, or a functional part, derivative or analogue of said virus.

4. An isolated or recombinant virus comprising an amino acid sequence which is more than 84 % homologous to at least part of a 163-2 amino acid sequence as depicted in Figure 1, said part having at least 30 amino acid residues.

5. An isolated or recombinant virus comprising an amino acid sequence which is more than 51 % homologous to at least part of a 163-4 amino acid sequence as depicted in Figure 1, said part having at least 30 amino acid residues.

6. An isolated or recombinant virus comprising a nucleic acid sequence which is more than 81 % homologous to at least part of a 163-9 nucleic acid sequence as depicted in Figure 1, said part having at least 30 nucleic acid residues.

7. An isolated or recombinant virus comprising an amino acid sequence which is more than 91 % homologous to at least part of a 163-10 amino acid sequence as depicted in Figure 1, said part having at least 30 amino acid residues.

8. An isolated or recombinant virus comprising an amino acid sequence which is more than 33 % homologous to at least part of a 163-11 amino acid sequence as depicted in Figure 1, said part having at least 30 amino acid residues.

9. An isolated or recombinant virus comprising an amino acid sequence which is more than 81 % homologous to at least part of a 163-14 amino acid sequence as depicted in Figure 1, said part having at least 30 amino acid residues.

10. An isolated or recombinant virus comprising an amino acid sequence which is more than 81 % homologous to at

least part of a 163-15 amino acid sequence as depicted in Figure 1, said part having at least 20 amino acid residues.

11. An isolated or recombinant virus comprising an amino acid sequence which is more than 74 % homologous to at least part of a 163-18 amino acid sequence as depicted in Figure 1, said part having at least 30 amino acid residues.

12. An isolated or recombinant virus comprising a nucleic acid sequence which is at least 50 % homologous to at least part of a nucleic acid sequence as depicted in Table 3, said part having at least 50 nucleotides.

13. An isolated or recombinant virus or a functional part, derivative or analogue according to any one of claims 1-12, capable of inducing a HCoV-NL63-related disease.

14. A primer and/or probe, capable of specifically hybridizing to a nucleic acid of a virus or functional part, derivative or analogue according to any one of claims 1-13.

15. A primer and/or probe, capable of specifically hybridizing to a nucleic acid sequence as depicted in Table 3.

16. A primer and/or probe according to claim 14 or 15, which is capable of hybridizing to said nucleic acid under stringent conditions.

17. A primer and/or probe according to any one of claims 14-16, comprising a sequence as depicted in Table 7.

18. An isolated molecule capable of specifically binding a virus or functional part, derivative or analogue according to any one of claims 1-13.

19. An isolated molecule capable of specifically binding an amino acid sequence of a virus or functional part, derivative or analogue according to any one of claims 1-13.

20. An isolated molecule capable of specifically binding a nucleic acid sequence of a virus or functional part, derivative or analogue according to any one of claims 1-13.

21. An isolated molecule capable of specifically binding at least part of a nucleic acid sequence as depicted in Table 3.

22. An isolated molecule capable of specifically binding at least part of an amino acid sequence as depicted in Figure 1.

23. A molecule according to any one of claims 18-22 which is a proteinaceous molecule.

24. A method for producing a proteinaceous molecule capable of specifically binding a virus or functional part, derivative or analogue according to any one of claims 1-13 comprising:

- producing proteinaceous molecules capable of binding a virus or functional part, derivative or analogue according to any one of claims 1-13, and
- selecting a proteinaceous molecule that is specific for said virus.

25. An isolated or recombinant virus which is immunoreactive with a molecule according to any one of claims 18-23.

26. Use of a virus or functional part, derivative and/or analogue according to any one of claims 1-13 or 25 for detecting a molecule capable of specifically binding said virus in a sample.

27. Use of an amino acid sequence of a virus or functional part, derivative or analogue according to any one of claims 1-13 or 25, for detecting a molecule capable of specifically binding said virus or functional part, derivative and/or analogue in a sample.

28. Use according to claim 27, wherein said amino acid sequence comprises a sequence as depicted in Figure 1 or a functional part, derivative or analogue thereof.

29. Use according to any one of claims 26-28, wherein said molecule comprises a specific ligand and/or antibody of said virus.

30. Use of a primer and/or probe according to any one of claims 14-17, a molecule according to any one of claims 18-23, and/or a nucleic acid of a virus or functional part, derivative or analogue according to any one of claims 1-13 or 25, for detecting and/or identifying a HCoV-NL63 coronavirus in a sample.

31. Use according to claim 30 wherein said nucleic acid comprises a sequence as depicted in Table 3.

32. A virus or functional part, derivative or analogue according to any one of claims 1-13 or 25 for use as a vaccine or medicament.

33. A molecule according to any one of claims 18-23 for use as a vaccine or medicament.

34. Use of a virus or functional part, derivative or analogue according to any one of claims 1-13 or 25 for the preparation of a vaccine against a coronaviral genus related disease.

35. Use of a molecule according to any one of claims 18-23 for the preparation of a vaccine or medicament against a coronaviral genus related disease.

36. Use of a virus or functional part, derivative or analogue according to any one of claims 1-13 or 25, a molecule according to any one of claims 18-23, or a primer and/or probe according to any one of claims 14-17 for diagnosis of a coronaviral genus related disease.

37. Use according to any one of claims 34-36 wherein said coronaviral genus related disease comprises an HCoV-NL63 coronavirus related disease.

38. A vaccine comprising a virus or functional part, derivative or analogue according to any one of claims 1-13 or 25 and/or a molecule according to any one of claims 18-23.

39. A medicament comprising a virus or functional part, derivative or analogue according to any one of claims 1-13 or 25 and/or a molecule according to any one of claims 18-23.

40. A vaccine or medicament according to claim 38 or 39 for at least in part preventing and/or treating an HCoV-NL63-related disease.

41. A method for detecting a virus or functional part, derivative or analogue according to any one of claims 1-13 or 25 in a sample, comprising hybridizing and/or amplifying a nucleic acid of said virus or functional part, derivative or analogue with a primer and/or probe according to any one of claims 14-17 and detecting hybridized and/or amplified product.

42. A diagnostic kit comprising a virus or functional part, derivative or analogue according to any one of claims 1-13 or 25, a molecule according to any one of claims 18-23, and/or a primer/probe according to any one of claims 14-17.

43. A method for treating an individual suffering from, or at risk of suffering from, an HCoV-NL63 related disease, comprising administering to said individual a vaccine or medicament according to any one of claims 38-40.

44. A method for determining whether an individual suffers from an HCoV-NL63 related disease, comprising obtaining a sample from said individual and detecting a HCoV-NL63 virus or functional part, derivative or analogue thereof in said sample with a method according to claim 41 or a diagnostic kit according to claim 42.

45. An isolated or recombinant nucleic acid encoding a virus or functional part, derivative and/or analogue according to any one of claims 1-13 or 25.

46. A nucleic acid according to claim 45, comprising at least a functional part of a sequence as depicted in Table 3.

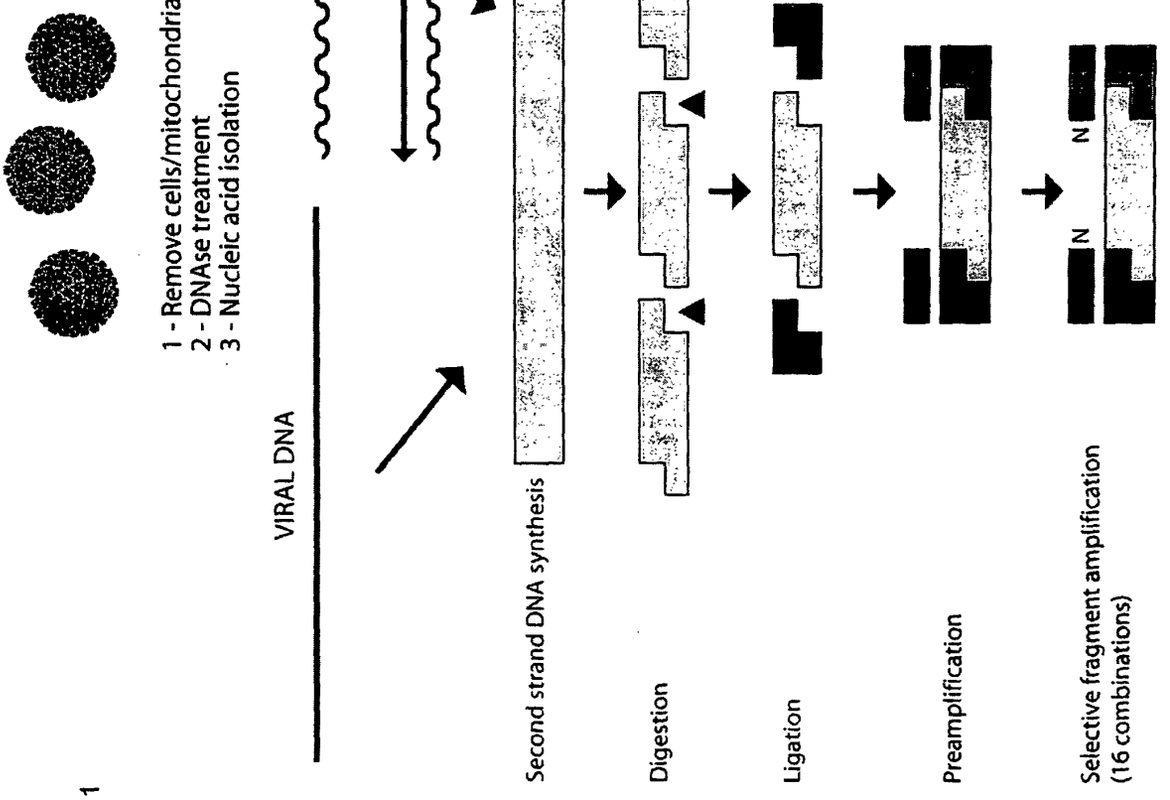
47. An amino acid sequence encoded by a nucleic acid according to claim 45 or 46.

48. An amino acid sequence according to claim 47, comprising at least a functional part of a sequence as depicted in Figure 1.

49. A proteinaceous molecule capable of specifically binding HCoV-NL63, obtainable by a method according to claim 24.

50. Use of a proteinaceous molecule according to claim 49 in a vaccine or a diagnostic method for the detection of HCoV-NL63.

Figure 1


1 - Remove cells/mitochondria
2 - DNAse treatment
3 - Nucleic acid isolation
VIRAL DNA
VIRAL RNA
cDNA synthesis
Second strand DNA synthesis
Digestion
Ligation
Preamplification
Selective fragment amplification
(16 combinations)
Restriction Enzymes
Anchors
Primers
Selective primers
N
N
N
N

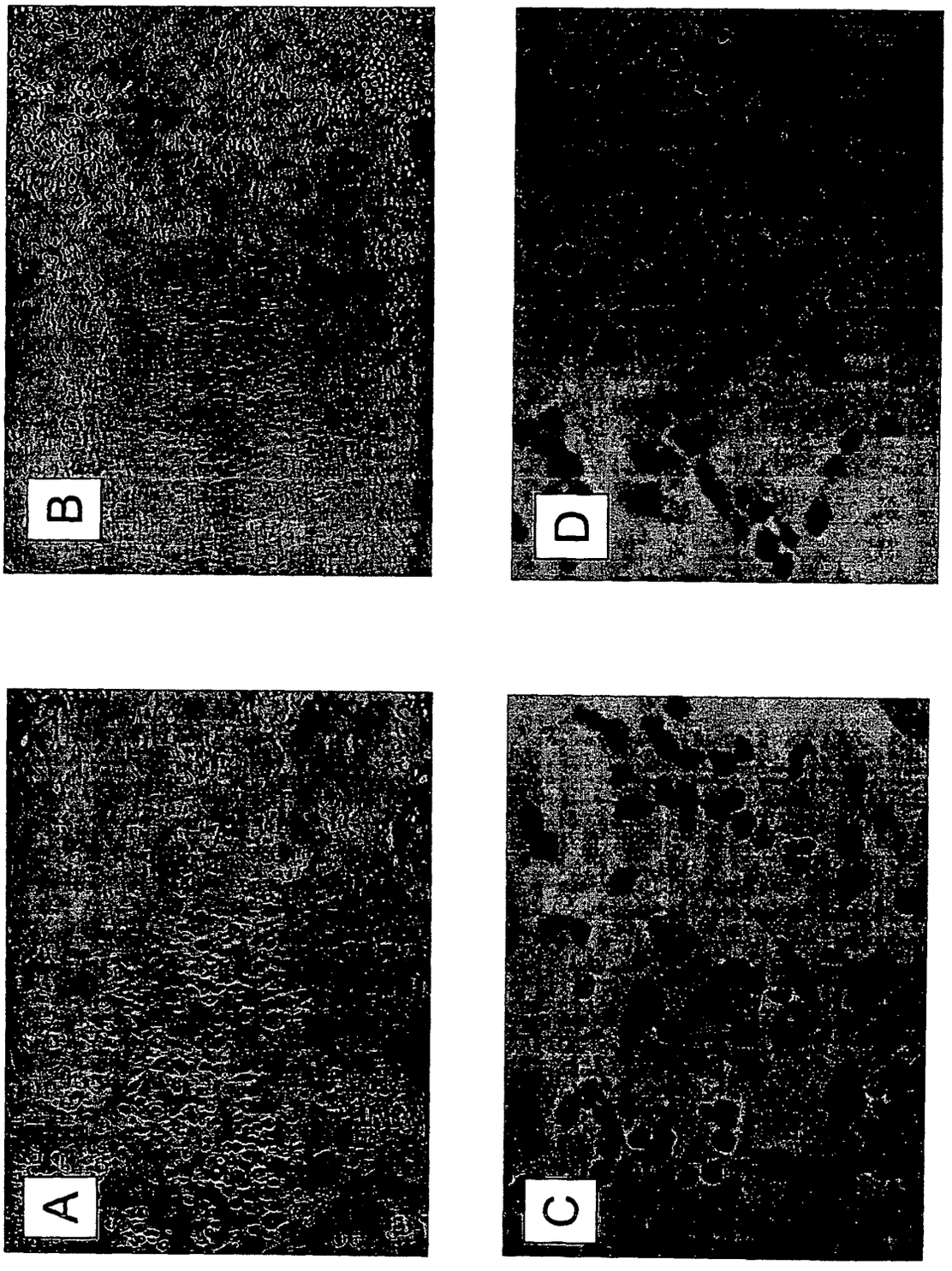


Figure 2

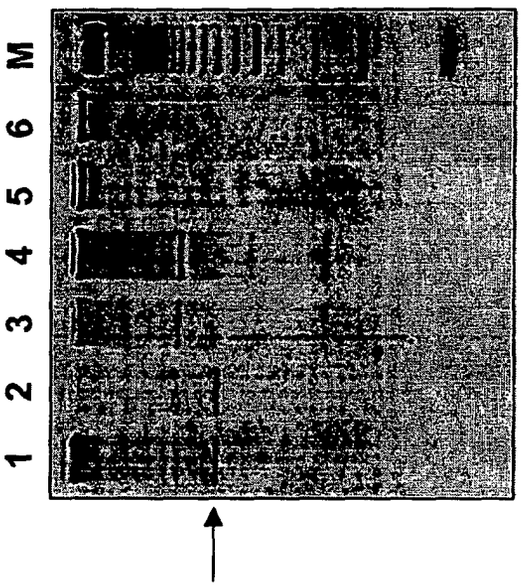
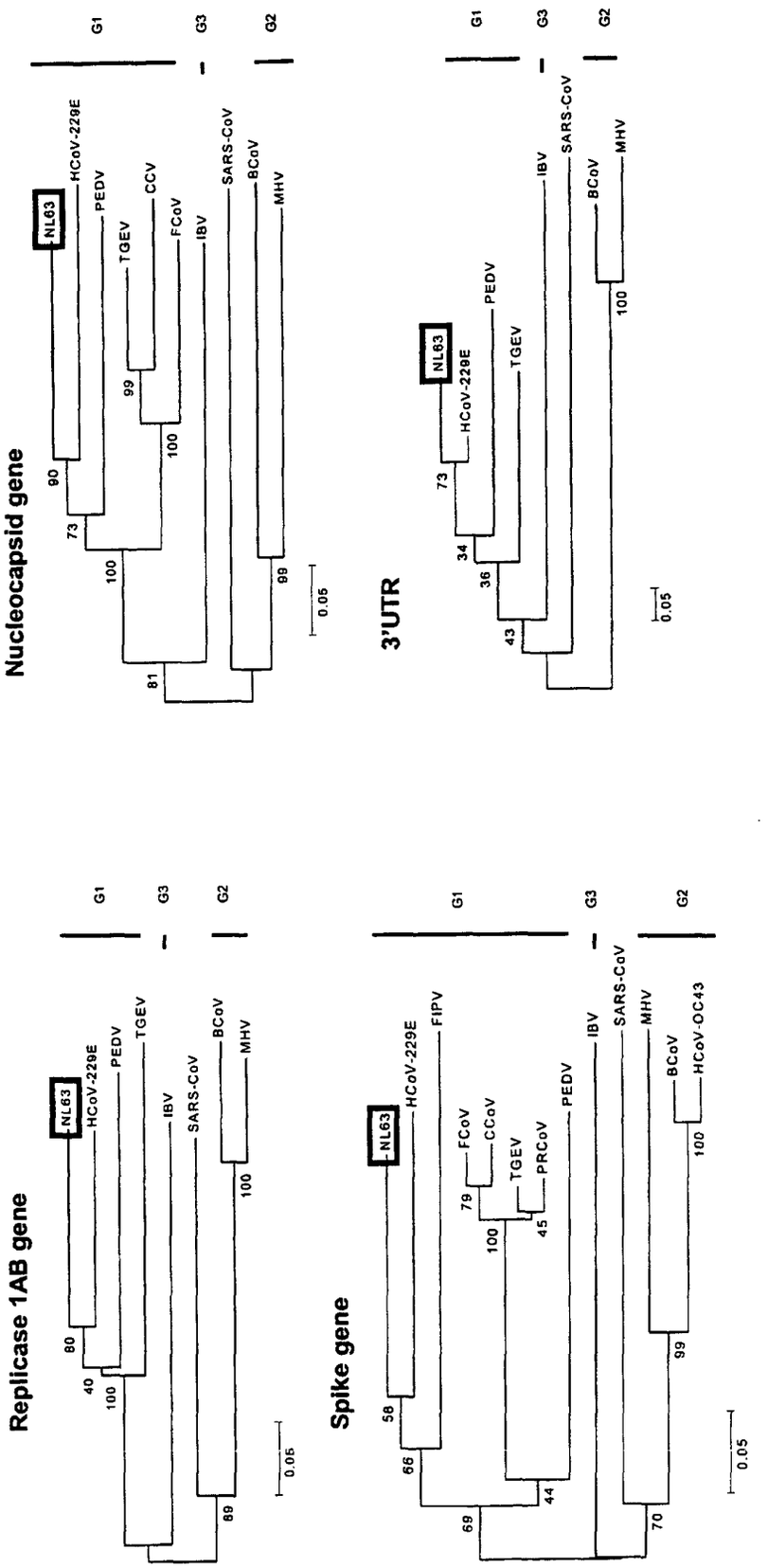


Figure 3

Figure 4


Nucleocapsid gene
G1
G3
G2
NL63
HCoV-229E
PEDV
TGEV
CCV
FCoV
IBV
SARS-CoV
BCoV
MHV
90
73
100
99
100
81
99
0.05
3'UTR
G1
G3
G2
NL63
HCoV-229E
PEDV
TGEV
IBV
SARS-CoV
BCoV
MHV
73
34
36
43
100
0.05
Replicase 1AB gene
G1
G3
G2
NL63
HCoV-229E
PEDV
TGEV
IBV
SARS-CoV
BCoV
MHV
80
40
100
89
100
0.06
Spike gene
G1
G3
G2
NL63
HCoV-229E
FIPV
FCoV
CCoV
TGEV
PRCoV
PEDV
IBV
SARS-CoV
MHV
BCoV
HCoV-OC43
58
66
79
100
45
44
69
70
99
100
0.05

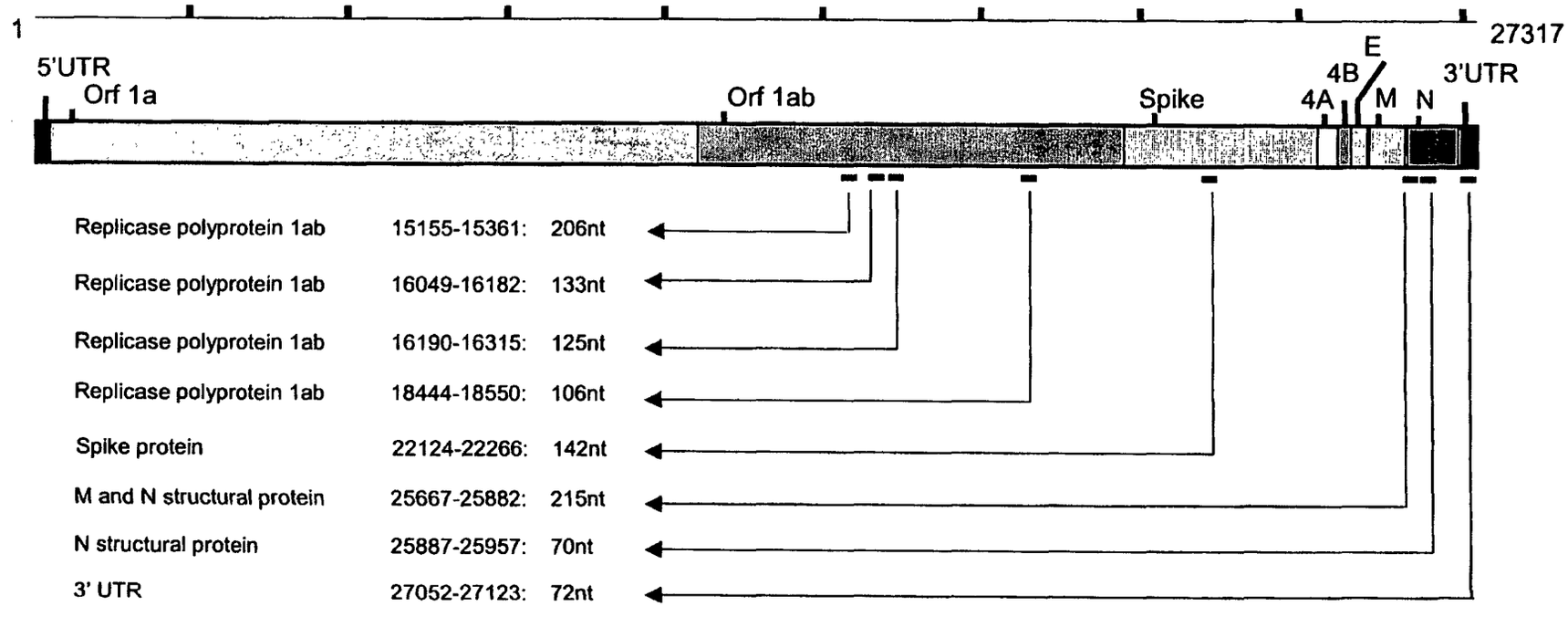


Figure 5, a schematic representation of the coronavirus genome organization (with HCoV-229E accession number AF304460 as an example). Red lines indicate the genome location of the sequence fragments of the new coronavirus HCoV-SLA163. UTR denotes unranslated region, orf: open reading frame, E: envelope protein, M: membrane protein, N: nucleocapsid protein.

European Patent Office

PARTIAL EUROPEAN SEARCH REPORT

which under Rule 45 of the European Patent Convention shall be considered, for the purposes of subsequent proceedings, as the European search report

Application Number EP 03 07 7602

DOCUMENTS CONSIDERED TO BE RELEVANT

Category	Citation of document with indication, where appropriate, of relevant passages	Relevant to claim	CLASSIFICATION OF THE APPLICATION (Int.Cl.7)
X	DATABASE SWISS-PROT 'Online! EBI; 11 August 2001 (2001-08-11) THIEL V ET AL.: "Replicase polyprotein 1ab, Human coronavirus 229E" Database accession no. AAG48591 XP002275245 88% homologous to SEQ ID NO:9; 93% homologous to SEQ ID NO:14; 87% homologous to SEQ ID NO:18; 77% homologous to SEQ ID NO:22 * abstract * ---	1-4,7,9, 11,13, 25,32	C12N7/00 C07K14/165 A61K39/215 C12Q1/68 G01N33/569 C07K16/10
X	DATABASE SWISS-PROT 'Online! EBI; 15 August 1999 (1999-08-15) RAABE T ET AL.: "E2 glycoprotein precursor (spike glycoprotein) (peplomer protein)" Database accession no. P15423 XP002275246 58% homologous to SEQ ID NO:11 * abstract * --- -/--	1-3,5, 13,25,32	

TECHNICAL FIELDS SEARCHED (Int.Cl.7)

C12N
C07K
A61K

INCOMPLETE SEARCH

The Search Division considers that the present application, or one or more of its claims, does/do not comply with the EPC to such an extent that a meaningful search into the state of the art cannot be carried out, or can only be carried out partially, for these claims.

Claims searched completely :

Claims searched incompletely :

Claims not searched :

Reason for the limitation of the search:

see sheet C

Place of search	Date of completion of the search	Examiner
THE HAGUE	8 April 2004	Brouns, G

CATEGORY OF CITED DOCUMENTS

X : particularly relevant if taken alone
Y : particularly relevant if combined with another document of the same category
A : technological background
O : non-written disclosure
P : intermediate document

T : theory or principle underlying the invention
E : earlier patent document, but published on, or after the filing date
D : document cited in the application
L : document cited for other reasons

& : member of the same patent family, corresponding document

EPO FORM 1503 03.82 (P04C07)

European Patent Office

INCOMPLETE SEARCH SHEET C

Application Number
EP 03 07 7602

Although claim 44 is directed to a method of treatment of the human/animal body (Article 52(4) EPC), the search has been carried out and based on the alleged effects of the compound/composition.
Although claim 43 is directed to a diagnostic method practised on the human/animal body (Article 52(4) EPC), the search has been carried out and based on the alleged effects of the compound/composition.

Reason for the limitation of the search:

The claims refer to amino acid and nucleotide sequences as depicted in Figure 1 whereas said figure does not depict amino acid or nucleotide sequences. Therefore the search has been based on the SEQ ID NOs 9, 11, 12, 14, 16, 18, 20, and 22, that seem to define the amino acid and nucleotide sequences depicted in table 3, which correspond to the numbers mentioned in claims 4-11. In addition the sequences defined by SEQ ID NOs 23-27, corresponding to the sequences in table 7, have been searched.

In addition, claims 18-23 refer to an isolated molecule binding to the virus of the invention, or a part thereof. No specific examples for said molecules are provided and the search has been limited to antibodies, antisense nucleic acid molecules and ribozymes binding to the virus, viral proteins or viral nucleic acids of the invention. Since the essential technical and structural features of said molecules are well known to the skilled person, said molecules are considered to be sufficiently disclosed and supported.

European Patent Office

PARTIAL EUROPEAN SEARCH REPORT

Application Number EP 03 07 7602

DOCUMENTS CONSIDERED TO BE RELEVANT			CLASSIFICATION OF THE APPLICATION (Int.Cl.7)
Category	Citation of document with indication, where appropriate, of relevant passages	Relevant to claim	
X	DATABASE EMBL 'Online! EBI; 11 July 2001 (2001-07-11) SCHREIBER S ET AL.: "Human coronavirus 229E nucleocapsid protein mRNA" Database accession no. M22052 XP002275247 84% homologous to SEQ ID NO:12; 54% homologous to SEQ ID NO:16; 82% homologous to SEQ ID NO:20 * abstract * ---	1-3,6,8, 10, 12-16, 25,32	
X	VABRET ASTRID ET AL: "Direct diagnosis of human respiratory coronaviruses 229E and OC43 by the polymerase chain reaction" JOURNAL OF VIROLOGICAL METHODS, vol. 97, no. 1-2, September 2001 (2001-09), pages 59-66, XP002275244 ISSN: 0166-0934	30,31, 36-39, 41,42,44	TECHNICAL FIELDS SEARCHED (Int.Cl.7)
Y	* the whole document * ---	24, 26-29, 34,40,43	
X	DATABASE ENTREZ NUCLEOTIDES 'Online! NCBI; 24 June 2003 (2003-06-24) ZIEBUHR J ET AL.: "Avian infectious bronchitis virus" Database accession no. NC_001451 XP002275248 100% identical over 10 nucleotides:13644-13654 * abstract * ---	17	
Y	WO 01 09290 A (LOZANO DUBERNARD BERNARDO ;SARFATI MISRAHI DAVID (MX); ARANDA MERL) 8 February 2001 (2001-02-08) * claims 17-36,39-42,44 * -----	24, 26-29, 34,40,43	

EPO FORM 1503 03.82 (P04C10)

ANNEX TO THE EUROPEAN SEARCH REPORT ON EUROPEAN PATENT APPLICATION NO. EP 03 07 7602

This annex lists the patent family members relating to the patent documents cited in the above-mentioned European search report. The members are as contained in the European Patent Office EDP file on
The European Patent Office is in no way liable for these particulars which are merely given for the purpose of information.

08-04-2004

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 0109290 A	08-02-2001	WO 0109290 A2	08-02-2001

EPO FORM P0459

For more details about this annex : see Official Journal of the European Patent Office, No. 12/82

PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 4 :
C12N 7/00, G01N 33/569
A61K 39/42, 39/215

A1

(11) International Publication Number: WO 88/ 01292

(43) International Publication Date: 25 February 1988 (25.02.88)

(21) International Application Number: PCT/US87/02034

(22) International Filing Date: 18 August 1987 (18.08.87)

(31) Priority Application Number: 899,534

(32) Priority Date: 22 August 1986 (22.08.86)

(33) Priority Country: US

(71) Applicant: BOARD OF REGENTS, THE UNIVERSITY OF TEXAS SYSTEM [US/US]; 201 West Seventh Street, Austin, TX 78701 (US).

(72) Inventors: RESTA, Silvia (NMI) ; 8159 Southwestern, #1-259, Dallas, TX 75206 (US). LUBY, James, P. ; 4312 Fairfax Avenue, Dallas, TX 75205 (US).

(74) Agent: HODGINS, Daniel, S.; Arnold, White & Durkee, P.O. Box 4433, Houston, TX 77210 (US).

(81) Designated States: AT, AT (European patent), AU, BB, BE (European patent), BG, BJ (OAPI patent), BR, CF (OAPI patent), CG (OAPI patent), CH, CH (European patent), CM (OAPI patent), DE, DE (European patent), DK, FI, FR (European patent), GA (OAPI patent), GB, GB (European patent), HU, IT (European patent), JP, KP, KR, LK, LU, LU (European patent), MC, MG, ML (OAPI patent), MR (OAPI patent), MW, NL, NL (European patent), NO, RO, SD, SE, SE (European patent), SN (OAPI patent), SU, TD (OAPI patent), TG (OAPI patent).

Published
With international search report.
Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.

(54) Title: HUMAN ENTERIC CORONAVIRUS

(57) Abstract

Biologically pure cultures of human enteric coronaviruses A-14 and C-14 have been prepared. The serial cultivation of viruses from fecal samples of infant patients suffering from necrotizing enterocolitis resulted in the obtainment of the biologically pure cultures. This serial cultivation was in an antibiotic-supplemented medium for mammalian cell culture preferably Leibovitz L-15 medium (pH 6.8), which contained fetal human intestinal pieces. Infections by human enteric coronavirus may be immunochemically diagnosed by use of antibodies against the A-14 and C-14 coronaviruses. Such antibodies have been prepared and found to bind strongly to coronavirus A-14 or C-14 and not to OC43, 229E, MHV-A59 and Breda 1 and 2 viruses. Fecal samples from patients suffering from enteric disease can be interacted with antibody and the amount of binding between the antibody and antigens of the sample measured by any of several means well-known in the art. When the binding between the antibody and antigens of a patient's fecal sample is determined to be significantly greater than that found between the antibody and antigens of a fecal sample from a control uninfected individual which usually demonstrate no interaction, a positive diagnosis of infection by human enteric coronavirus is obtained.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AT	Austria	**FR**	France	**ML**	Mali
AU	Australia	**GA**	Gabon	**MR**	Mauritania
BB	Barbados	**GB**	United Kingdom	**MW**	Malawi
BE	Belgium	**HU**	Hungary	**NL**	Netherlands
BG	Bulgaria	**IT**	Italy	**NO**	Norway
BJ	Benin	**JP**	Japan	**RO**	Romania
BR	Brazil	**KP**	Democratic People's Republic of Korea	**SD**	Sudan
CF	Central African Republic			**SE**	Sweden
CG	Congo	**KR**	Republic of Korea	**SN**	Senegal
CH	Switzerland	**LI**	Liechtenstein	**SU**	Soviet Union
CM	Cameroon	**LK**	Sri Lanka	**TD**	Chad
DE	Germany, Federal Republic of	**LU**	Luxembourg	**TG**	Togo
DK	Denmark	**MC**	Monaco	**US**	United States of America
FI	Finland	**MG**	Madagascar		

HUMAN ENTERIC CORONAVIRUS

The present invention relates to the detection of infection by human enteric coronavirus and to the first isolation and propagation of a human enteric coronavirus.

Coronaviruses are single-stranded RNA viruses with distinctive surface projections from their envelopes. The members of this group cause a variety of infections in humans and in animals. Because human strains are difficult to isolate and cultivate, they are not well studied. Coronaviruses most commonly cause minor upper respiratory infections (colds) and are sometimes capable of affecting the lower respiratory tract. Of the many coronaviruses affecting animals, the best known are those causing avian infectious bronchitis of chickens, mouse hepatitis, and transmissible gastroenteritis of piglets. Other coronaviruses infect rats, calves, dogs, cats, and foals. A number of the animal strains infect the intestinal tract of the host species; analogous human enteric coronaviruses have also been described (Cukor, et al., Human viral gastroenteritis, Microbiol. Rev. (1984) 48: 157-179; Kapikan, et al., J. Infect. Dis. (1969) 119: 282-290; and, Monto, Medical reviews: Coronaviruses, Yale

J. Biol. Med. (1974) 47: 234-251), incorporated by reference herein).

Coronaviruses are about 100-150 nm in diameter and tend to be pleomorphic. The coronaviruses are surrounded by a halo of club-shaped projections resembling spikes or petals which suggest the solar corona. These projections are longer (up to 20 nm) and usually more widely spaced than the spikes of orthomyxoviruses, to which the coronaviruses bear a slight morphologic similarity.

All coronaviruses probably share the same basic biophysical and biochemical structure, but much of the information about them derives from studies of animal strains, particularly mouse hepatitis. The internal component of coronavirus consists of RNA and protein. The nucleoprotein is coiled and forms a helical structure 9-16 nm wide that seems fragile and easily disrupted. The membrane that encloses the coronavirus particle has protein and lipid components; its club-shaped projections, called peplomers, are proteins. Biochemical analysis of the virions indicates that they contain polypeptides of three main classes. The first, which is phosphorylated but nonglycosylated (mol. wt. about 50,000 daltons), is associated with the internal ribonucleoprotein. The second polypeptide is glycosylated and is associated with the membrane; the third, which is also glycosylated, is associated with the peplomers. The coronaviruses of humans and of various animal species seem to differ in the number, molecular weight, and extent of glycosylation of these polypeptides. The virus particles can be destroyed by lipid solvents (Siddell, et al., J. Gen. Virol. (1983) 64: 761-776, incorporated by reference herein).

Coronavirus replication occurs in the cytoplasm of infected cells. The virus matures by budding into vesicles in the cell cytoplasm and is released when the cell lyses.

Although some of the animal coronaviruses can be grown relatively easily in the laboratory, human strains have very strict host cell specificities, and are notoriously difficult to isolate from patients and to adapt to cell cultures. As a result, coronaviruses still pose many unanswered questions. The prototype human respiratory coronavirus, 229E, and strains related to it can be grown in certain tissue cultures. Other human strains, the OC strains, are so called because most of them can be grown only in organ cultures (explants) of differentiated human respiratory epithelium originating from nasal or tracheal tissue of aborted human fetuses; the presence of virus in the culture requires confirmation by electron microscopy.

Of the human coronaviruses, strains 229E and OC43 have been studied most (Table 1). These two viruses are antigenically distinct. As yet, there is no definition of an antigenic type or group, and it may be difficult to make such a definition if partial sharing of antigens proves to be as common among the human coronaviruses as among coronaviruses causing infections bronchitis of chickens. Subtypes or variants may prove to be epidemiologically important. Human coronavirus strains isolated in tissue culture all appear to be antigenically related to 229E and may thus constitute one broad type or group, and OC strains constitute at least one other such group. None of the human respiratory coronaviruses tested so far is related antigenically to the infectious bronchitis of chicken, although human strain 229E has been found to be

related to certain coronaviruses of pigs, dogs, and cats, and human strain OC43 to coronaviruses of mice (mouse hepatitis), rats, and cattle. The relationship of virus structure to antigenic composition is being investigated. The surface spikes probably elicit neutralizing activity and, when it occurs, hemagglutinating activity, but definite information is sparse.

TABLE 1

TYPES OF HUMAN CORONAVIRUS: IN VITRO CULTIVATION AND SEROLOGIC TESTS

Types of Human Coronavirus	Cultivation in Vitro			Serologic Tests Available*					
	Organ Culture	Tissue Culture	Baby Mice	N	CF	IF	HI	IEM	ELISA
229E types	+	+	–	+	+	+	–	+	+
OC43	+	±†	+†	+	+	+	+	+	+
Other OC types	+	–	–	+*	–	+**	–	+**	?

* N, neutralization; CF, complement-fixation, IF, immunofluorescence; HI, hemagglutination inhibition IEM, immunoelectron microscopy; ELISA, enzyme-linked immunosorbent assay.

† After adaptation.

** Using organ cultures.

Type OC43 coronavirus has been adapted to grow in brains of infant mice and also (although somewhat poorly) in tissue culture. The mouse-adapted virus agglutinates red blood cells and this property can be used in serologic tests. Serologic properties of the strains that grow in tissue culture may be studied by neutralization or complement fixation; neutralization tests can also be done with some difficulty in organ culture. ELISAs with 229E and OC43 antigens have been used successfully and have confirmed the antigenic difference between the two strains. All strains may be examined by the rather cumbersome techniques of fluorescent staining of viral antigen in infected cells or by immune electron microscopy.

Inoculation of human volunteers with coronaviruses has provided valuable information. Volunteers develop colds, often with profuse coryza, 2-3 days after intranasal inoculation of a coronavirus. The virus evidently multiplies in nasal epithelium and can be recovered form nasal secretions by appropriate tissue cultures or organ cultures. In adult volunteers, infections generally are brief and without pyrexia; asymptomatic infections also may occur. Natural infections with coronaviruses occur in all age groups. Lower respiratory involvement is presently not thought to be a common feature of the infection, although coronaviruses, like rhinoviruses, have been implicated in exacerbations of asthma and chronic bronchitis.

Coronaviruses can cause chronic or persistent infections in some animals and in tissue cultures, but persistent infection in humans has not been demonstrated conclusively. Coronaviruses or coranaviruslike particles have been observed by electron microscopy in human feces,

but the association of these particles with disease has not previously been established. However, coronavirus strains have been associated with diarrheal diseases in lower animals, and there is evidence that these viruses may be involved in human enteric diseases (Stair et al., Am. J. Vet. Res., 33: 1147 (1972); Tajima et al., Arch. Gesamte Virusforshung, 29, 105 (1975); Garwes et al., J. Gen. Virol. 29: 25 (1975); Hierholzer et al., Infect. Immun. 24: 508 (1979). Most of the data in support of the latter hypothesis result from electron microscopic observations of coronavirus-like particles in stool samples obtained from patients with gastroenteritis or necrotizing enterocolitis (NEC) (Chany et al., Pediatrics, 69, 209 (1982); Vaucher et al., J. Infect. Dis., 145, 27 (1982); Gerna et al., J. Infect. Dis., 150, 618 (1984); Laporte et al., Perspect. Virol, 11, 189 (1981) and; MacNaughton et al., Arch. Virol., 70, 301 (1981). Attempts to cultivate these particles for antigenic or biochemical analysis have been unrewarding to date.

Coronavirus infections cannot be distinguished clinically from other viral respiratory infections. Laboratory diagnosis usually is made by examining paired sera to detect rising antibody titers. Viral strain 229E is used in complement-fixation or neutralization tests and strain OC43 in complement-fixation or hemagglutination-inhibition or neutralization tests. ELISA methods may supplement these. Rising antibody titers against 229E or OC43 may also be produced by infection with heterologous but related viruses.

Virus may be isolated from nasal swabs or washings by using human embryo nasal or tracheal organ cultures. Strain 229E and related strains sometimes can be isolated directly in human embryo kidney cells, diploid human

embryo fibroblast lines, or other sensitive cells, but isolation may be easier after preliminary passage in organ culture. The cytopathic effect in tissue culture is not distinctive; the presence of coronavirus in the cultures is confirmed by electron microscopy and use of reference antisera. Identification of coronavirus in nasal secretions by ELISA also has been described. Neither vaccination nor chemoprophylaxis is available for human secretions by ELSIA also has been described. Neither vaccination nor chemoprophylaxis is available for human coronavirus infections.

Other workers have reported an association between coronaviruses and NEC or with serious gastrointestinal disease in nursery infants. These reports were based on observations of virus particles in stools by electron microscopy of immune electron microscopy (Chany et al., Pediatrics 49, 209 (1982); Vaucher et al., J. Infect. Dis. 145, 27 (1982); and Gerna et al., J. Infect Dis. 150, 618 (1984). However, their attempts to cultivate these particles were not successful. The results described herein establish the existence of a human enteric coronavirus (HEC) and indicate an association between these virus particles and cases of NEC observed in a Dallas epidemic.

Biologically pure cultures of human enteric coronaviruses A-14 and C-14 have been prepared. The serial cultivation of viruses from fecal samples of infant patients suffering from necrotizing enterocolitis resulted in the obtainment of the biologically pure cultures. This serial cultivation was in an antibiotic-supplemented medium for mammalian cell culture preferably Leibovitz L-15 medium (pH 6.8), which contained fetal human intestinal pieces.

Infections by human enteric coronavirus may be immunochemically diagnosed by use of antibodies against the A-14 and C-14 coronaviruses. Such antibodies have been prepared and found to bind strongly to coronavirus A-14 or C-14 and not to OC43, 229E, MHV-A59 and Breda 1 and 2 viruses. Fecal samples from patients suffering from enteric disease can be interacted with antibody and the amount of binding between the antibody and antigens of the sample measured by any of several means well-known in the art. When the binding between the antibody and antigens of a patient's fecal sample is determined to be significantly greater than that found between the antibody and antigens of a fecal sample from a control uninfected individual which usually demonstrate no interaction, a positive diagnosis of infection by human enteric coronavirus is obtained..

An immunological response of a patient to human enteric coronavirus infection may be determined by a process of the present invention. Antigens obtained from coronavirus A-15 or C-15, when brought into contact with a patient's serum containing antibody against those coronaviruses, result in well-known and recognized patterns characteristic of antigen-antibody interactions. The presence of such antibodies are indicative of a present or past immunological response of the patient to human enteric coronavirus infection.

The present invention comprises a human enteric coronavirus (HEC) associated with necrotizing enterocolitis (NEC) in ailing infants which has been isolated, identified and characterized as described herein. Pure cultures of the virus (A-14 and C-14) have been obtained from fecal samples of two infants. The A-14 and C-14 viruses of these cultures appeared to be morphologically

and antigenically identical to each other and different from any previously described coronavirus.

An immunological response to HEC was noted in infants suffering from NEC and was generated in guinea pigs by injection of whole HEC. The diagnosis of NEC when involving HEC may be facilitated by a determination of the presence of an antibody against HEC in serum samples from a patient. This presence may be determined with many standard immunoassay techniques by using the unique HEC described herein or characteristic antigens thereof.

Immunoprophylaxis has been a successful approach to the control of viral diseases. Such immunoprophylaxis may involve passive or active immunization. Passive immunization with preparations comprising antibodies specific for HEC may be used for the short-term treatment of acute conditions involving HEC. Polyclonal or monoclonal antibodies specific for HEC A-14 or HEC C-14 or antigens thereof may be prepared by contemporary techniques of immunology. Currently pooled plasma preparations containing antibodies against Varicella zoster and the viruses causing rabies and hepatitis B are commercially available for clinical use.

Immunity to HEC may be generated by an active immunization process. Currently available for general use are numerous anti-viral vaccines containing live attenuated virus, killed virus or viral antigens. These vaccines include those to generate immunity against smallpox, rabies, yellow fever, influenza, poliomyelitis, measles, mumps and rubella. Active immunization against HEC may be accomplished by the administration of an attenuated A-14 or C-14 virus. The development of a viable non-toxic attenuated HEC will require extensive

experimentation and testing. Immunization with killed HEC or HEC antigens should be more readily accomplished than the development of attenuated virus strains. The isolated HEC A-14 or C-14 described herein may be killed, fragmented or modified by chemical or physical means and used as non-infectious antigens by methods well known in the immunological arts. Parenteral administration of such antigens mixed with a pharmaceutically acceptable carrier such as a sterile isotonic salt solution, with or without an immunological adjuvant, comprises use as a vaccine to preclude the onset of infectious NEC or any other infection by a coronavirus having antigenic commonality with HEC C-14 or HEC A-14.

These examples are presented to describe preferred embodiments and utilities of the present invention and are not meant to limit the present invention unless otherwise stated in the claims appended hereto.

EXAMPLE 1

Isolation of Human Enteric Coronavirus

An epidemic of NEC occurred in a hospital special care nursery in Dallas, Texas in 1982-83. All of the patients showed established criteria for NEC (Bell et al., Arch. Int. Med. 198, 1 (1978)) -intolerance to food, abdominal distension, occult or gross blood in stool, and radiologic evidence of pneumatosis intestinalis. Stool samples from patients revealed coronavirus-like particles. During the epidemic, stool specimens and sera were obtained from controls and from patients with NEC and diarrhea at acute and convalescent stages of disease.

All stool specimens were screened for the presence of coronavirus-like particles by electron microscopy before inoculation in cultures of human fetal intestinal organ. The cultures were prepared as earlier described (Horn et al., J. Exp. Pathol., 46, 109 (1965); Rubenstein et al., J. Exp. Pathol., 51, 210 (1970); Dolin et al., J. Infect. Dis., 122, 227 (1970); and Antrup et al., Gastroenterology, 74, 1248 (1978)), incorporated by reference herein). Briefly, the intestines were opened longitudinally with microscissors, and 2 by 2 mm pieces, with the intestinal villi oriented upward, were placed in tissue culture dishes. At each passage, the organ cultures were incubated in Leiboviz L-15 medium (pH 6.8) supplemented with antibiotics in a humidified environment at 37°C with a 5 percent CO_2 atmosphere. Uninoculated organ cultures, prepared as controls, were maintained under the same conditions.

Seven stool samples, showing positive results by electron microscopy, and eight showing negative results were cultured. None of the negative inocula resulted in the growth of coronavirus-like particles after being passaged five times and assayed by electron microscopy and sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE). Two of seven positive specimens led to the isolation and propagation of coronavirus-like particles and these have been maintained for 14 passages (HEC A-14 and HEC C-14). After the first blind passages (3 to 5), the two strains produced destruction of the brush border of the intestinal epithelium and degeneration of the villi. For passages 8 to 14, trypsin, which is thought to have an enhancing effect on coronavirus replication, was added to the growth medium at 5 ug/ml. Cultures infected in the absence of trypsin were maintained in parallel. Treatment of the inocula with chloroform or heat (56°C for

15 minutes) nullified the effect of the inocula on the organ cultures--that is, there was no cytopathic effect, virus-like particles were not seen by electron microscopy, and the protein band profile characteristic of viral growth was not revealed by SDS-PAGE. Control infected cultures not treated with chloroform or heat continued to show evidence of the presence of HEC. Filtration of the inocula through filters of 0.45 um and 0.22 um pore did not prevent infection of the organ cultures.

The supernatants and tissue extracts of the infected cultures from passages 6 to 14 were purified on a glycerol-potassium tartrate gradient (Hierholzer, Virology 75, 155 (1976); and Obijeski et al., J. Gen. Virol. 22, 21 (1974)). The peak of spectrophotometric activity at 280nm wavelength corresponded to a density of 1.18 g/cm^3. Observation of the band collected from the gradient revealed particles with morphology typical of coronaviruses--namely, club-shaped spikes, a diameter of 100 to 150 nm, a pleomorphic appearance, and an erythrocyte-like profile. The purified particles were tested for hemagglutination with goose, chicken, rat, guinea pig, rabbit, and human O erythrocytes at 4°, 25°, and 37°C. No hemagglutinating activity was observed.

EXAMPLE 2

Characteristics of the Isolated Coronavirus Particles

During single radial hemolysis (SRH) assay, the purified particles of Example 1 were reacted with convalescent-stage sera from six patients with NEC; four of the patients showed seroconversion. Sera from control infants showed no reactivity against the antigens (Table

2). Neither the purified HEC antigens nor the infants' sera revealed cross-reactivity with antisera to OC43 and 229E and with OC43 and 229E antigens, respectively.

Single radial hemolysis (SRH) assay was performed according to a previously described technique (Hierholzer et al., J. Clin. Microbiol. 5, 613 (1977). Before being used, all sera were adsorbed overnight (4°C) with human fetal intestinal homogenate. All sera were tested after heating at 56°C for 30 minutes (1, first serum; 2, second serum). A serum was considered positive when the reaction caused a halo of hemolysis at least 3 mm in diameter (well diameter, 1mm). The first or acute-stage serum was collected 4 to 8 weeks after the onset of illness. Controls were infants in the nursery without necrotizing enterocolitis or diarrhea. A6, A9, A-14 and C6, C9, C-14 designate the passage level of the isolates. Abbreviations: NCS, newborn calf serum; FCS, fetal calf serum; BSA, bovine serum albumin; SRBC; sheep red blood cells; ND, not done.

TABLE 2

SINGLE RADIAL HEMOLYSIS ASSAY

Patients	Diameter of HEC sample (mm)						Controls	Diameter of HEC sample (mm)					
	A6	A9	A14	C6	C9	C14		A6	A9	A14	C6	C9	C14
H-1	0	0	0	0	0	0	R	0	0	0	0	0	0
H-2	4.3	3.4	4.5*	4.2	3.5	4.0*	McM-1	0	0	0	0	0	0
M-1	0	0	0	0	0	0	McM-2	0	0	0	0	0	0
McC-1	0	0	0	0	0	0*	P	0	0	0	0	0	0
McC-2	0	0	5.0*	0	0	4.8*	Pc	0	0	0	0	0	0
E-2	ND	ND	3.5	ND	ND	3.5	T	0	0	0	0	0	0
D-1	0	0	0*	0	0	0*	Lo	0	0	0	0	0	0
D-2	3.5	3.5	4.2*	3.5	3.5	4.0*	Hi	0	0	0	0	0	0
Hn-1	0	0	0*	0	0	0*	NCS	0	0	0	0	0	0
Hn-2	3.0	3.5	4.5*	3.0	3.0	4.3*	FCS	0	0	0	0	0	0
Mo-1	0	3.5	3.5*	0	3.5	3.5*	Anti-OC43	0	0	0	0	0	0
Mo-2	4.0	4.0	4.5*	3.8	4.0	4.3*	Anti-229E	0	0	0	0	0	0
							Anti-BSA	0	0	0	0	0	0
							Anti-SRBC	10	10	10	10	10	10

* Sample diluted 1:5; all other sera were used undiluted except the NCS, FCS, anti-BSA, and anti-SRBC, which were diluted 1:10.

Several in vitro systems, such as primary human embryonic kidney cells, human embryonic kidney cells human embryonic lung fibroblasts, HEP-2, Vero, and BHK cells, did not support the growth of the viral particles. Although attempts have been made to adapt the virus to a cellular substrate that can be more easily managed, human fetal intestinal organ culture is the only reproducible system discovered at present. Treatment of the cultures, with trypsin appeared to facilitate the infection, since the treated cultures gave rise to higher yields of viral particles, as seen on electron microscopy, than did untrypsinized cultures.

In tests to date, the two strains isolated appear to be identical. Immunologic tests with specific antisera should allow verification of this finding and enable the establishment of possible antigenic relationships with other coronaviruses. Both strains: 1.) Were obtained from infants in the same epidemic, 2.) Had identical physical characteristics (density, sensitivity to heat and chloroform, 3.) Had identical appearances by electron microscopy, 4.) Had identical protein patterns by SDS-PAGE, and 5.) Had antisera to each virus interact in the same way with the other virus, i.e., antiserum to A-14 reacts with A-14 proteins in an identical manner to C-14 proteins in a Western blot and vice versa.

EXAMPLE 3

Immunological HEC Antigen Characterization

Antigens (from purified HEC A-14 and C-14, OC43, 229E, and human fetal intestinal homogenates) were tested by enzyme-linked immunosorbent assay (ELISA) against the infants' sera and against antisera to OC43, 229E, MHV-A59,

and Breda 1 and 2 viruses. For sources of other viruses see Woode et al., Vet. Microbiol. 7,221 (1982); Weiss et al., J. Gen. Virol. 64 1849 (1983); Horzineck et al., J. Gen Virol., 65, 1849 (1983); and Beards et al., Lancet 1984-1, 1050 (1984), incorporated by reference herein. These interactions were also tested by standard techniques (Voller et al. in Manual of Clinical Immunology, King et al., eds., Amer. Soc. for Microbiol., Wash. D.C. (1976) p 506; and Yolken et al., Lancet 1977-11, 261 (1977). The working dilutions of serum were 1:10 and 1:100, and each determination was made in triplicate. Appropriate controls for all of the reagents used were included in each assay. All of the sera were adsorbed overnight (4°C) with human fetal intestinal homogenates before being used. Briefly, both the infants' sera and the other antisera were assayed by binding the antigen directly to the wells of the microtiter plates. After incubation of the test serum samples, alkaline phosphatase-conjugated immunoglobulins (anti-human, anti-rabbit, and anti-guinea pig) were incubated in the wells. The enzyme substrate (p-nitrophenyl phosphate in diethanolamine buffer) was added after washing, and the reaction was stopped and evaluated for absorbance (AB) in a spectrophotometer. Antisera to OC43, 229E. MHV-A59, and Breda 1 and 2 viruses were also assayed by means of a capturing antibody-coated well assay, with specific guinea pig antisera to HEC A-14 and C-14, OC43, and human intestinal homogenates, test antisera, conjugated immunoglobulins, and enzyme substrate were consecutively incubated in the wells. Each incubation step was followed by washing three times with standard buffer solutions. The AB value of each serum-control antigen reaction was subtracted from the corresponding AB value of the serum-viral antigen reaction to obtain the value of the test sample. A threshold cutoff point was determined on the basis of the highest values

obtained on a group of negative controls. Positive sera always had an AB value greater than 3 standard deviations above the mean of a group of negative control sera. The ELISA test confirmed the results obtained by the SRH assay. Convalescent-stage sera from five of the infants with NEC showed titers of 1:100 or more; two infants showed seroconversion; and six control sera had titers less than 1:20. The infants' sera did not react with OC43 and 229E antigens. No reactions were demonstrable between A-14 and C-14 antigens and antisera to OC43, 229E, MHV-A59, and Breda 1 and 2 viruses.

Gradient (5 to 17 percent) SDS-PAGE of the purified particles revealed the presence of at least five major bands corresponding to molecular sizes ranging from 190 to 23 kilodaltons. Electrophoretically separated proteins (HEC A-14 and C-14, OC43 and 229E viruses, and human fetal intestinal homogenates) were blotted onto nitrocellulose paper for Western immunoblotting. The blotted proteins were allowed to react with dilutions (1:50 or 1:100) of the acute-and convalescent-stage sera from patients and serum samples from controls obtained during the Dallas NEC epidemic. Seven of seven convalescent-stage sera and two of 11 control sera reacted against HEC A-14 and C-14. Two patients showed seroconversion. The convalescent-stage sera did not react against OC43 and 229E viruses. Reactions occurred mainly with proteins corresponding to molecular sizes of 190, 120, and 50kD. A reaction was seen with the 23-kD protein for some of the samples.

EXAMPLE 4

Preparation of Guinea Pig Antisera to HEC

Each of two guinea pigs were given two intramuscular injections of whole HEC A-14 virus or whole HEC C-14 virus in complete Freund's adjuvant at 4 week intervals. At two week intervals each guinea pig was administered four additional doses of the virus (in normal saline), the first two doses by intramuscular injection and the last two by subcutaneous injection. As a later final booster each guinea pig was given, one week before blood was withdrawn, a subcutaneous injection of virus in normal saline. Blood was drawn, allowed to clot, and antisera stored in a conventional manner.

X X X X X X X X

Changes may be made in the components described herein or in the steps or the sequence of steps of the methds described herein without departing from the concept and scope of the invention as defined in the following claims.

CLAIMS:

1. A biologically pure culture of HEC A-14.

2. A biologically pure culture of HEC C-14.

3. A biologically pure culture of human enteric coronavirus, said coronavirus being obtainable by serial cultivation of viruses from fecal samples of patients having necrotizing enteric coronavirus infection, said cultivation being in an antibiotic-supplemented medium for mammalian cell culture with human fetal intestinal pieces.

4. The biologically pure culture of claim 3 wherein the medium is pH 6.8 Liebowitz L-15 medium.

5. A method for the immunochemical diagnosis of infection with human enteric coronavirus in a patient, the method comprising:

contacting a fecal sample from the patient with an antibody binding more strongly to HEC A-14; or HEC C-14 than to OC43, 229E, MHV-A59 and Breda 1 and 2 viruses;

measuring the amount of binding between antigens of the sample and said antibody; and

determining, for a positive diagnosis, whether the amount of binding is significantly greater than

that found between the antibody and antigens of a fecal sample from a control uninfected individual.

6. A method for the detection of an immunological response of a patient to infection by human enteric coronavirus, the method comprising:

obtaining a serum sample from the patient; and

contacting the serum sample with antigen from HEC A-14 or HEC C-14 to ascertain whether an antibody specific for said antigen is present in the serum, presence of such antibody being indicative of a present or past immunological response of the patient to human enteric coronavirus infection.

7. A composition comprising an antibody, the antibody being characterized by binding more strongly to HEC A-14 or HEC C-14 than to OC43, 229E, MHV-A59 and Breda 1 and 2 viruses.

8. A method for producing a vaccine capable of inducing an immunological response against HEC, the method comprising preparing an antigen from HEC A-14 or HEC C-14 and mixing said antigen with a pharmaceutical acceptable carrier or an immunological adjuvant and a pharmaceutically acceptable carrier.

9. A method for inducing immunity to HEC in an individual, the method comprising parenteral administration of antigens from HEC A-14 or HEC C-14 to said individual.

10. A composition comprising a pharmaceutically acceptable carrier and an antigen from HEC A-14 or HEC C-14.

11. A method for passive immunization to protect an individual from HEC infection, the method comprising parenteral administration to said individual of a preparation comprising antibodies specific for HEC A-14 or HEC C-14.

12. The method of claim 11 wherein the antibodies are defined further as binding more strongly to HEC A-14 or HEC C-14 than to OC43, 229E, MHV-A59 and Breda 1 and 2 viruses.

INTERNATIONAL SEARCH REPORT

International Application No PCT/US 87/02034

I. CLASSIFICATION OF SUBJECT MATTER (if several classification symbols apply, indicate all) [6]

According to International Patent Classification (IPC) or to both National Classification and IPC

IPC[4]: C 12 N 7/00; G 01 N 33/569; A 61 K 39/42; 39/215

II. FIELDS SEARCHED

Minimum Documentation Searched [7]

Classification System	Classification Symbols
IPC[4]	A 61 K; C 12 N

Documentation Searched other than Minimum Documentation to the Extent that such Documents are Included in the Fields Searched [8]

III. DOCUMENTS CONSIDERED TO BE RELEVANT [9]

Category [10]	Citation of Document, [11] with indication, where appropriate, of the relevant passages [12]	Relevant to Claim No. [13]
X	Science, vol. 229, 6 September 1985 S. Resta et al.: "Isolation and propagation of a human enteric coronavirus", pages 978-981 see the whole document	1-8,10

---

* Special categories of cited documents: [10]

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

IV. CERTIFICATION

Date of the Actual Completion of the International Search	Date of Mailing of this International Search Report
20th November 1987	08 JAN 1988
International Searching Authority	**Signature of Authorized Officer**
EUROPEAN PATENT OFFICE	M. VAN MOL

FURTHER INFORMATION CONTINUED FROM THE SECOND SHEET

V. [X] OBSERVATIONS WHERE CERTAIN CLAIMS WERE FOUND UNSEARCHABLE [1]

This international search report has not been established in respect of certain claims under Article 17(2) (a) for the following reasons:

1. [X] Claim numbers **), because they relate to subject matter not required to be searched by this Authority, namely:

**) 9,11-12 See PCT Rule 39.1(iv) Methods for treatment of the human or animal body by surgery or therapy as well as diagnostic methods

2. [ ] Claim numbers....., because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:

3. [ ] Claim numbers....., because they are dependent claims and are not drafted in accordance with the second and third sentences of PCT Rule 6.4(a).

VI. [ ] OBSERVATIONS WHERE UNITY OF INVENTION IS LACKING [2]

This International Searching Authority found multiple inventions in this international application as follows:

1. [ ] As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims of the international application.

2. [ ] As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims of the international application for which fees were paid, specifically claims:

3. [ ] No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claim numbers:

4. [ ] As all searchable claims could be searched without effort justifying an additional fee, the International Searching Authority did not invite payment of any additional fee.

Remark on Protest

[ ] The additional search fees were accompanied by applicant's protest.

[ ] No protest accompanied the payment of additional search fees.

US007452542B2

(12) **United States Patent**
Denison

(10) **Patent No.: US 7,452,542 B2**
(45) **Date of Patent: Nov. 18, 2008**

(54) **LIVE ATTENUATED CORONAVIRUS VACCINES**

(75) Inventor: **Mark Denison**, Nashville, TN (US)

(73) Assignee: **Vanderbilt University**, Nashville, TN (US)

(*) Notice: Subject to any disclaimer, the term of this patent is extended or adjusted under 35 U.S.C. 154(b) by 275 days.

(21) Appl. No.: **11/135,603**

(22) Filed: **May 23, 2005**

(65) **Prior Publication Data**

US 2006/0039926 A1 Feb. 23, 2006

Related U.S. Application Data

(60) Provisional application No. 60/573,587, filed on May 21, 2004.

(51) **Int. Cl.**

A61K 39/215	(2006.01)
C12N 7/00	(2006.01)
C12N 7/04	(2006.01)
C12N 15/50	(2006.01)

(52) **U.S. Cl.** **424/221.1**; 435/235.1; 435/236; 536/23.72

(58) **Field of Classification Search** ..... None
See application file for complete search history.

(56) **References Cited**

OTHER PUBLICATIONS

Enjuanes et al. Current Topics in Microbiology and Immunology 287: 161-197, 2005.*

Thiel et al. Current Topics in Microbiology and Immunology 287: 199-227, 2005.*

Masters et al. Current Topics in Microbiology and Immunology 287: 133-159. 2005.*

Gorbalenya et al. Virus Research 117:17-37, 2006.*

Cavanagh. Avian Pathology 32(6):567-582, 2003.*

Saif. "Animal Coronavirus Vaccines: Lessons for SARS." Developments in Biologicals 119:129-140, 2004.*

Index of Viruses—Coronaviridae (2006). In: ICTVdB—The Universal Virus Database, version 4. Büchen-Osmond, C (Ed), Columbia University, New York, USA. <<http://www.ncbi.nlm.nih.gov/ICTVdb/Ictv/fs_index.htm>> Downloaded Aug. 1, 2008.*

Baker et al., "Identification of the catalytic sites of a papain-like cysteine proteinase of murine coronavirus," *J. Virol.*, 67:6056-6063, 1993.

Bonilla et al., "Characterization of a second cleavage site and demonstration of activity in trans by the papain-like proteinase of the murine coronavirus mouse hepatitis virus strain A59," *J. Virol.*, 71:900-909, 1997.

Bonilla et al., "Characterization of the Leader Papain-like Proteinase of MHV-A59: Identification of a New in Vitro Cleavage Site," *Virology*, 209:489-497, 1995.

Boursnell et al., "Completion of the sequence of the genome of the coronavirus avian infectious bronchitis virus," *J. Gen. Virol.*, 68:57-77, 1987.

Chouljenko et al., "Comparison of genomic and predicted amino acid sequences of respiratory and enteric bovine coronaviruses isolated from the same animal with fatal shipping pneumonia," *J. Gen. Virol.*, 82:2927-2933, 2001.

Curtis et al., "Heterologous gene expression from transmissible gastroenteritis virus replicon particles," *J. Virol.*, 76:1422-1434, 2002.

Dong and Baker, "Determinants of the p28 Cleavage Site Recognized by the First Papain-like Cysteine Proteinase of Murin Coronavirus," *Virology*, 204:541-549, 1994.

Eleouet et al., "Complete Sequence (20 Kilobases) of the Polyprotein-Encoding Gene 1 of Transmissible Gastroenteritis Virus," *Virology*, 206:817-822, 1995.

Yoo and Pei, "Full-length genomic sequence of bovine coronavirus (31 kb). Completion of the open reading frame 1a/1b sequences," *Adv. Exp. Med. Biol.*, 494:73-76, 2001.

Yount et al., "Strategy for systematic assembly of large RNA and DNA genomes: transmissible gastroenteritis virus model," *J. Virol.*, 74:10600-10611, 2000.

Yount et al., "Systematic assembly of a full-length infectious cDNA of mouse hepatitis virus strain A59," *J. Virol.*, 76:11065-11078, 2002.

Ziebuhr et al., "The autocatalytic release of a putative RNA virus transcription factor from its polyprotein precursor involves two paralogous papain-like proteases that cleave the same peptide bond," *J. Biol. Chem.*, 276:33220-33232, 2001.

Genbank Accession No. AY274119, (2004).

Genbank Accession No. AY278491, (2003).

Genbank Accession No. AY278554, (2003).

Genbank Accession No. NC_001451, (2005).

Genbank Accession No. NC_001846, (2006).

Genbank Accession No. NC_002306, (2006).

Genbank Accession No. NC_002645, (2004).

Genbank Accession No. NC_003045, (2005).

Herold et al., "A human RNA viral cysteine proteinase that depends upon a unique Zn2+−binding finger connecting the two domains of a papain-like fold," *J. Biol. Chem.*, 274:14918-14925, 1999.

Herold et al., "Nucleotide Sequence of the Human Coronavirus 229E RNA Polymerase Locus," *Virology*, 195:680-691, 1993.

Hughes et al., "Identification of the murine coronavirus p. 28 cleavage site," *J. Virol.*, 69:809-813, 1995.

Kanjanahaluethai and Baker, "Identification of mouse hepatitis virus papain-like proteinase 2 activity," *J. Virol.*, 74:7911-7921, 2000.

Kanjanahaluethai and Baker., "Processing of the replicase of murine coronavirus: papain-like proteinase 2 (PLP2) acts to generate p. 150 and p. 44," *Adv. Exp. Med. Biol.*, 494:267-273, 2001.

Lee et al., "The Complete Sequence (22 Kilobases) of Murine Coronavirus Gene 1 Encoding the Putative Proteases and RNA Polymerase," *Virology*, 180:567-582, 1991.

Schaad et al., "Genetics of Mouse Hepatitis Virus Transcription: Identification of Cistrons which May Function in Positive and Negative Strand RNA Synthesis," *Virology*, 177:634-645, 1990.

Tijms et al., "A zinc finger-containing papain-like protease couples subgenomic mRNA synthesis to genome translation in a positive-stranded RNA virus," *Proc. Natl. Acad. Sci. USA*, 98:1889-1894, 2001.

* cited by examiner

Primary Examiner—Mary E Mosher
(74) *Attorney, Agent, or Firm*—Fulbright & Jaworski

(57) **ABSTRACT**

The present invention is directed live, attenuated coronavirus vaccines. The vaccine comprises a viral genome encoding a p59 protein having at mutation at a specific tyrosine residue, and may include other attenuating mutations. Such viruses show reduced growth and pathogenicity in vivo.

32 Claims, 5 Drawing Sheets

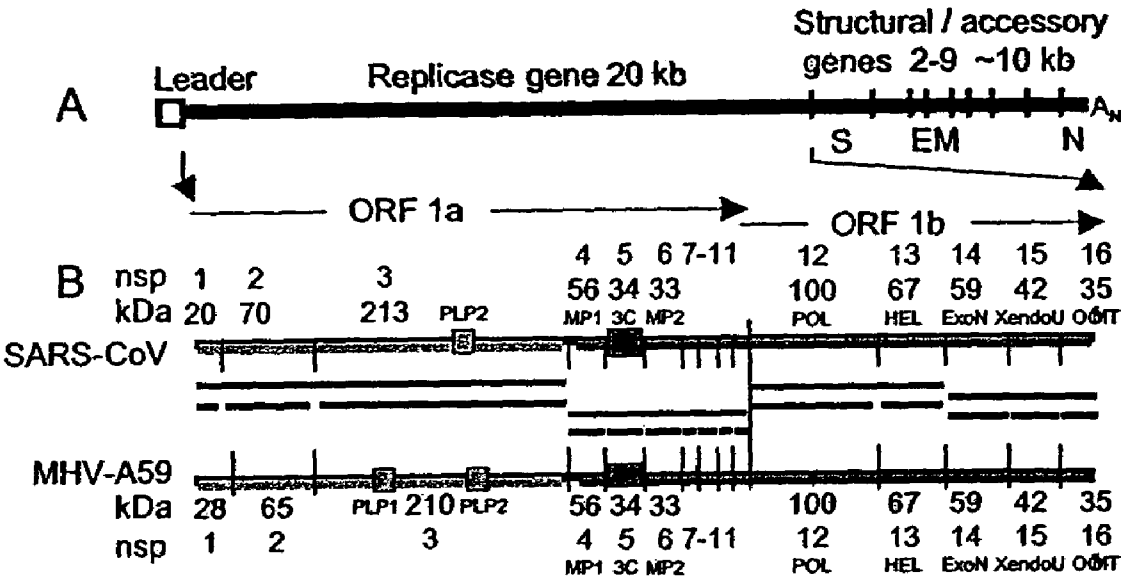

A
Leader
Replicase gene 20 kb
Structural / accessory
genes 2-9 ~10 kb
S
EM
N
ORF 1a
ORF 1b
B
nsp 1 2 3 4 5 6 7-11 12 13 14 15 16
kDa 20 70 213 PLP2 56 34 33 MP1 3C MP2 100 POL 67 HEL 59 ExoN 42 XendoU 35 O-MT
SARS-CoV
MHV-A59
kDa 28 65 PLP1 210 PLP2 56 34 33 100 67 59 42 35
nsp 1 2 3 4 5 6 7-11 12 13 14 15 16
MP1 3C MP2 POL HEL ExoN XendoU O-MT

FIG. 1A-B

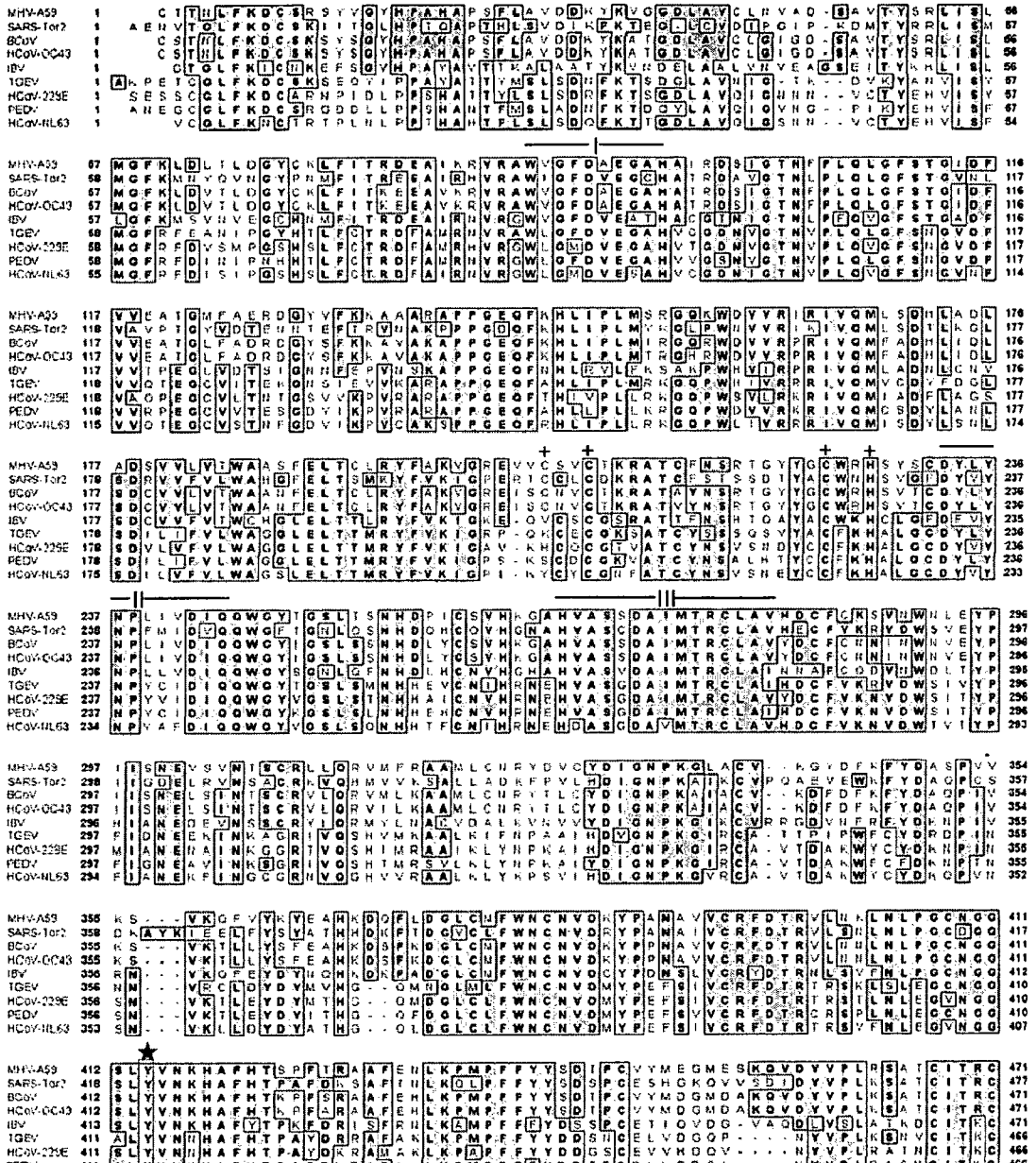
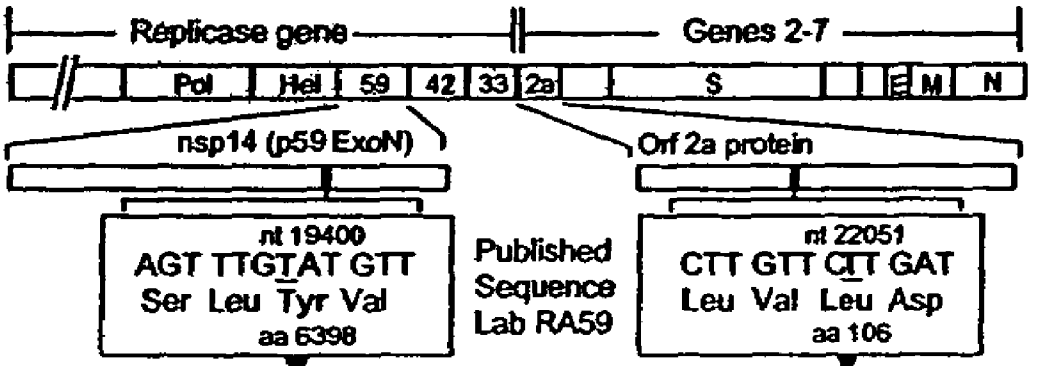

MHV-A59
SARS-Tor2
BCoV
HCoV-OC43
IBV
TGEV
HCoV-229E
PEDV
HCoV-NL63

FIG. 2

A



B

Virus	Amino Acids	LD_{50} (PFU)	LD_{50} Log_{10}
RA59	wtA59 Tyr_{6398} Leu_{106}	$6.32*10^3$	3.8
icwt "uncorrected"	His_{6398} Pro_{106}	$>2*10^5$	>5.3
VUSS-1	Tyr_{6398} Pro_{106}	$4.74*10^4$	4.67
VUSS-2	His_{6398} Leu_{106}	$>2*10^5$	>5.3
VUSS-3	Tyr_{6398} Leu_{106}	$6.32*10^3$	3.8

FIG. 3A-B

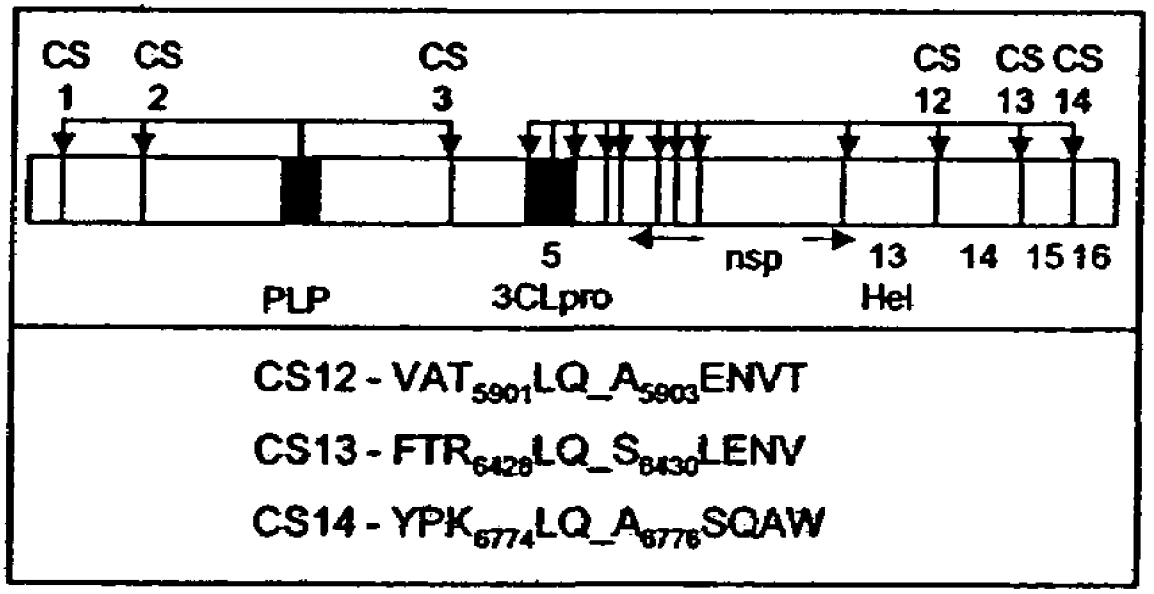

CS 1
CS 2
CS 3
CS 12
CS 13
CS 14
5
nsp
13
14
15
16
PLP
3CLpro
Hel
CS12 - VAT5901LQ_A5903ENVT
CS13 - FTR6428LQ_S6430LENV
CS14 - YPK6774LQ_A6776SQAW

FIG. 4

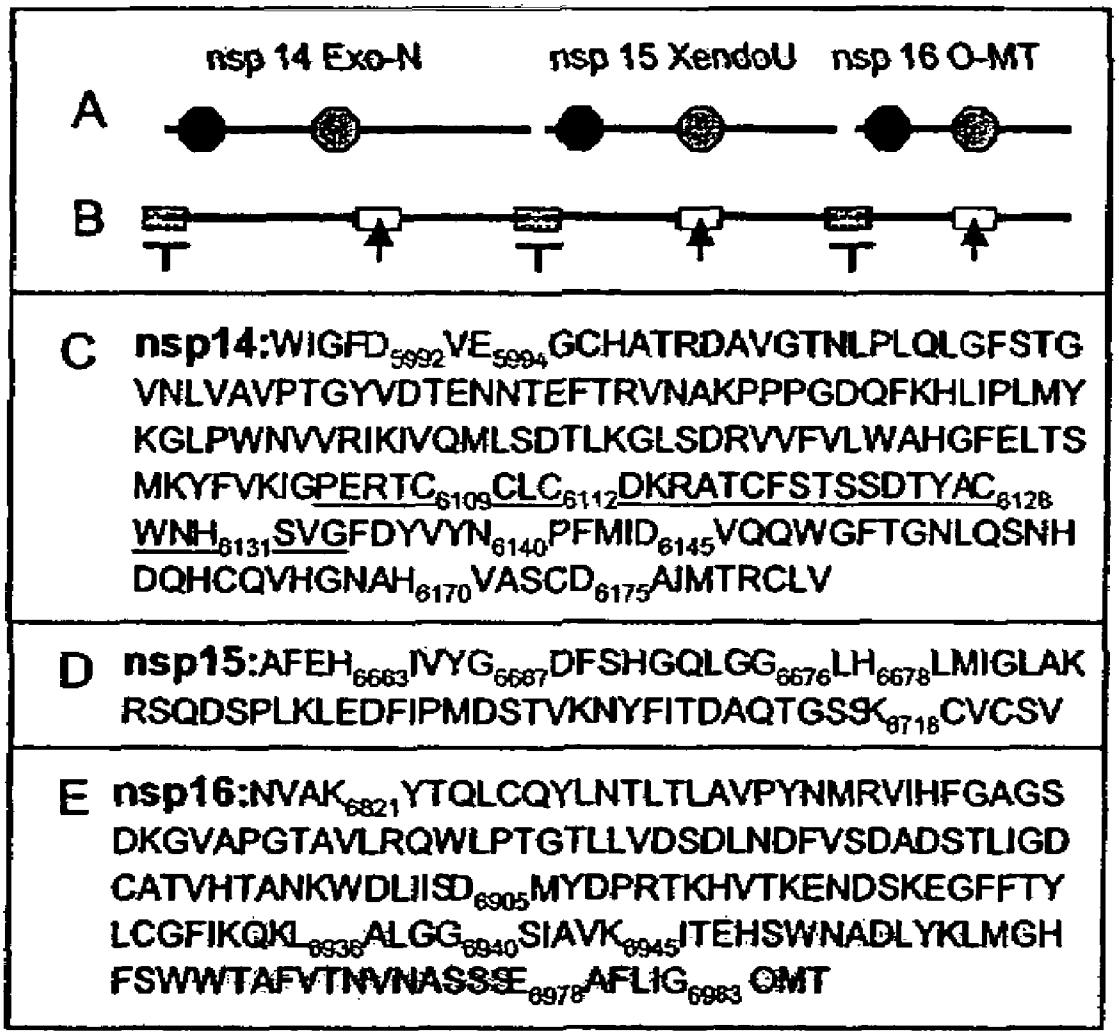

nsp 14 Exo-N
nsp 15 XendoU
nsp 16 O-MT
A
B
C nsp14:WIGFD5992VE5994GCHATRDAVGTNLPLQLGFSTG
VNLVAVPTGYVDTENNTEFTRVNAKPPPGDQFKHLIPLMY
KGLPWNVVRIKIVQMLSDTLKGLSDRVVFVLWAHGFELTS
MKYFVKIGPERTC6109CLC6112DKRATCFSTSSDTYAC6128
WNH6131SVGFDYVYN6140PFMID6145VQQWGFTGNLQSNH
DQHCQVHGNAH6170VASCD6175AIMTRCLV
D nsp15:AFEH6663IVYG6667DFSHGQLGG6676LH6678LMIGLAK
RSQDSPLKLEDFIPMDSTVKNYFITDAQTGSSK6718CVCSV
E nsp16:NVAK6821YTQLCQYLNTLTLAVPYNMRVIHFGAGS
DKGVAPGTAVLRQWLPTGTLLVDSDLNDFVSDADSTLIGD
CATVHTANKWDLIISD6905MYDPRTKHVTKENDSKEGFFTY
LCGFIKQKL6936ALGG6940SIAVK6945ITEHSWNADLYKLMGH
FSWWTAFVTNVNASSSE6978AFLIG6983 OMT

FIG. 5A-E

LIVE ATTENUATED CORONAVIRUS VACCINES

This application claims benefit of priority to U.S. Provisional Application Ser. No. 60/573,587, filed May 21, 2004, the entire contents of which are hereby incorporated by reference.

The government owns rights in the present invention pursuant to grant number 5RO1 A126603-15 of the National Institutes of Health and National Institute Allergy Infectious Disease.

BACKGROUND OF THE INVENTION

1. Field of the Invention

The present invention relates generally to the fields of microbiology, immunology and virology. More particularly, it concerns live, attenuated Coronaviridae vaccines and methods for preventing or limiting Coronaviridae infections.

2. Description of Related Art

Coronaviruses have been long known to cause important diseases in a wide variety of animal species, including humans, cattle, swine, chickens, dogs, cats and mice. Coronavirus diseases in non-human species may be quite severe, and devastating in domestic livestock such as pigs, cattle and chickens. The characterized human coronaviruses—HCoV-229E and HCoV OC43—are significant causes of upper respiratory infections, responsible for 10-35% of human colds. Studies of human coronaviruses have been limited by their lack of growth in culture from primary isolates, and by the lack, until recently, of reverse genetic approaches for their study. Thus, while the human coronaviruses are arguably two of the most economically important viruses in humans, ongoing research has been pursued only by a handful of dedicated investigators.

The emergence of a new human coronavirus associated with "severe acute respiratory syndrome" (SARS) surprised many scientists and public health officials, but has highlighted characteristics of coronaviruses well known to investigators. The coronaviruses have high rates of mutagenesis and homologous RNA recombination. In fact, template switching and recombination are essential to the normal life cycle of the viruses. In addition, the species barrier for coronaviruses has been predicted to be tenuous. Studies of coronaviruses in culture have demonstrated the ability of coronaviruses to adapt for replication in cells of different species. In addition, some studies have demonstrated that the murine coronaviruses may cause disease in primates following direct inoculation into brain. Finally, coronaviruses have been proposed, based on evolutionary studies, to have acquired genes from other viruses or cells, probably by recombination events. The emergence of a new coronavirus pathogenic for humans, by either adaptation of an animal virus, or by recombination of two coronaviruses during a coinfection, is consistent with these features of coronavirus evolution, replication and maintenance in populations.

Vaccine approaches for important domestic animal coronaviruses diseases, specifically the chicken avian infectious bronchitis virus (E3V), porcine transmissible gastroenteritis virus (TGEV), canine coronavirus (CCV), bovine coronavirus (BCV) and feline infectious peritonitis virus (FIPV), have been developed or attempted over the past 20 years. The approaches to vaccine development have been based on non-targeted natural attenuation, virus expression vectors, virus inactivation, recombinant viral structural proteins, and novel approaches to deliver or adjuvant vaccines. Responses and protectivity of these vaccines have varied widely, but no vaccine has been shown to possess all of the characteristics of safety, stability and efficacy.

For FIPV, live-attenuated, inactivated, and subunit vaccines based on recombinant or purified spike protein, have not only failed to protect against FIPV disease, but have resulted in immune enhancement of infection and disease, a response disturbingly reminiscent of the result following vaccination of humans with inactivated vaccines for measles and respiratory syncytial virus. The most useful animal coronavirus vaccine has been the live-attenuated vaccine for IBV. However, efficacy is still clearly less than optimal. In addition, reversion to virulence may occur, and recombination of the vaccine strain with wild-type viruses has occurred, with disease in chickens caused by the recombinant vaccine-wild-type viruses.

For the most part, vaccines have not been pursued in the past for human coronaviruses, likely because the frequency and severity of infections could not be well defined, and the determinants for protection have not been identified. It is also known that 229E and OC43 can re-infect humans, possibly as often as every other year, suggesting that vaccine strategies may need to be targeted toward limitation of disease severity, since prevention may not be possible.

Together, the known biological properties of coronaviruses, as well as the concerns with limited protection or immune enhancement of disease by coronavirus vaccines, are compelling arguments for a new approach in the development of live, attenuated vaccines that are less subject to reversion and recombination, but possess normal pathways for infection and immune response. This need is all the more critical in light of the emerging human SARS situation.

SUMMARY OF THE INVENTION

Thus, in accordance with the present invention, there is provided a live, attenuated virus of the family Coronaviridae, wherein the virus is characterized as comprising a genome encoding an ExoN comprising a substitution at tyrosine^{6398} of MHV-A59, or an analogous position thereof. The virus may be a group 2 coronavirus, and the genome further encodes an Orf2a polypeptide comprising a substitution at leu^{106} of MHV-A59, or an analogous position thereof. The virus may be a coronavirus or a torovirus, including coronaviruses such as avian infectious bronchitis virus, bovine coronavirus, canine coronavirus, feline infectious peritonitis virus, human coronavirus 229E, human coronavirus OC43, murine hepatitis virus, porcine epidemic diarrhea virus, porcine hemagglutinating encephalomyelitis virus, porcine transmissible gastroenteritis virus, rat coronavirus, turkey coronavirus, severe acute respiratory syndrome virus, or rabbit coronavirus, and toroviruses such as Berne virus or Breda virus.

The virus may further comprise a mutation in least one polyprotein proteinase cleavage site that exhibits reduced as compared to wild-type or no cleavage, such as a C1-C14 cleavage site, or a MHV p28-p65 or p65-p210 cleavage site or analogous position thereof. The cleavage site may comprise an amino acid deletion, an amino acid insertion or an amino acid substitution. Alternatively, the cleavage site may be wild-type, but cleavage may be reduced or eliminated by an allosteric mutation. The tyrosine^{6398} substitution may be a non-conservative substitution, or a histidine in particular. The leu^{106} substitution may be a non-conservative substitution, or a proline in particular. The virus genome may further encodes a

mutation in one or more of nsp1, nsp2, nsp3, nsp4, nsp5, nsp6, nsp7, nsp8, nsp9, nsp10, nsp11, nsp12, nsp13, nsp15 or nsp 16 coding region.

In another embodiment, there is provided a method of inducing an anti-viral immune response in a host comprising administering to the host a live, attentuated virus vaccine of the family Coronaviridae, wherein the virus is characterized as comprising a genome encoding an ExoN comprising a substitution at tyrosine^{6398} of MHV-A59, or an analogous position thereof. The virus may be a group 2 coronavirus, and the genome further encodes an Orf2a polypeptide comprising a substitution at leu106 of MHV-A59, or an analogous position thereof. The virus may be a coronavirus or a torovirus, including coronaviruses such as avian infectious bronchitis virus, bovine coronavirus, canine coronavirus, feline infectious peritonitis virus, human coronavirus 229E, human coronavirus OC43, murine hepatitis virus, porcine epidemic diarrhea virus, porcine hemagglutinating encephalomyelitis virus, porcine transmissible gastroenteritis virus, rat coronavirus, turkey coronavirus, severe acute respiratory syndrome virus, or rabbit coronavirus, and toroviruses such as Berne virus or Breda virus. The vaccine may be administered intravenously or subcutaneously, and/or co-administered with an immunostimulant. The host may be a dog, a cow, a pig, a cat, a mouse, a rat, a horse, a chicken, a turkey, a monkey or a human.

The method may comprise a virus that further comprises a mutation in least one polyprotein proteinase cleavage site that exhibits reduced as compared to wild-type or no cleavage, such as a C1-C14 cleavage site, or a MHV p28-p65 or p65-p210 cleavage site or analogous position thereof. The cleavage site may comprise an amino acid deletion, an amino acid insertion or an amino acid substitution. Alternatively, the cleavage site may be wild-type, but cleavage may be reduced or eliminated by an allosteric mutation. The tyrosine^{6398} substitution may be a non-conservative substitution, or a histidine in particular. The leu^{106} substitution may be a non-conservative substitution, or a proline in particular. The virus genome may further encodes a mutation in one or more of nsp1, nsp2, nsp3, nsp4, nsp5, nsp6, nsp7, nsp8, nsp9, nsp10, nsp11, nsp12, nsp13, nsp15 or nsp16 coding region.

In yet another embodiment, there is provided a coronavirus genome, the genome encoding an ExoN polypeptide comprising a substitution at tyrosine^{6398} of MHV-A59, or an analogous position thereof. Also provided is a coronavirus ExoN polypeptide comprising comprising a substitution at tyrosine^{6398} of MHV-A59, or an analogous position thereof.

In still yet another embodiment, there is provided a vaccine comprising (a) a live, attentuated virus of the family Coronaviridae, the virus characterized as comprising a genome encoding an ExoN polypeptide comprising comprising a substitution at tyrosine^{6398} of MHV-A59, or an analogous position thereof, and (b) a pharmaceutically acceptable diluent. The vaccine may be formulated as a unit dose of 10^6 to 10^{14} infectious particles. The vaccine may be provided in unit dose is provided in a 100 ml aliquot. The vaccine may further comprise a preservative. The vaccine may be lyophilized.

It is contemplated that any method or composition described herein can be implemented with respect to any other method or composition described herein.

The use of the word "a" or "an" when used in conjunction with the term "comprising" in the claims and/or the specification may mean "one," but it is also consistent with the meaning of "one or more," "at least one," and "one or more than one."

It is contemplated that any embodiment discussed in this specification can be implemented with respect to any method or composition of the invention, and vice versa. Furthermore, compositions and kits of the invention can be used to achieve methods of the invention.

Throughout this application, the term "about" is used to indicate that a value includes the inherent variation of error for the device, the method being employed to determine the value, or the variation that exists among the study subjects.

BRIEF DESCRIPTION OF THE DRAWINGS

The following drawings form part of the present specification and are included to further demonstrate certain aspects of the present invention. The invention may be better understood by reference to one or more of these drawings in combination with the detailed description of specific embodiments presented herein.

FIGS. **1**A-B - Genome organization, replicase proteins, and processing of SARS-CoV and MHV. FIG. **1**A. Schematic of the SARS-CoV genome. 30 kb, single-stranded, plus-strand RNA genome is shown, including leader RNA, 20 kb replicase gene, and general organization of structural and accessory genes, including those encoding spike (S), envelope protein, (E), membrane (M), and nucleocapsid proteins (N). Vertical black bars are intergenic regions and transcriptional regulatory sequences. FIG. **1**B. Replicase gene organization, protein domains, and mature and intermediate processing products. The nonstructural protein number (nsp), predicted size in kDa, and names or putative functions of proteins are shown for SARS-CoV: PLP2—papain-like proteinase orthologous to MHV PLP2; MP1 and MP2—hydrophobic membrane proteins; 3C-3C-like proteinase; POL—putative RNA-dependent RNA polymerase; HEL—RNA ATPase/helicase; ExoN—putative exonuclease; XendoU—predicted poly(U)-specific endoribonuclease; 2'-O-MT—predicted 2'-O-methyltransferase. For MHV, nsp numbers have not been assigned or are controversial. For this proposal, nsp numbers will correspond to those of SARS-CoV. PLP and regions of the polyprotein cleaved by PLP are shaded in yellow. 3Clpro and regions cleaved by 3Clpro in green. Red bars between genomes—possible intermediate precursor proteins. Black bars,—mature replicase proteins.

FIG. **2** - Alignment of p59 (nspl14, Exo N) proteins of coronaviruses and conservation of Tvr6398 residue. Alignments were perfonned of available group 1, 2 and 3 coronaviruses using a Clustal W protein alignment (implemented in MacVector 7.1- Accelerys). MHV-A59 - mouse hepatitis virus; SARS-Tor2 - Tor 2 strain of Severe Acute Respiratory Syndrome Coronavirus; BCoV - bovine coronavirus; HCoV - 0C43 -human coronavirus 0C43; JBV - infectious bronchitis virus; TGEV - transmissible gastroenteritis virus of pigs; HCoV-229E - human coronavirus 229E; PEDV - porcine enteric diarrhea virus; HCoV-NL63 - human coronavirus NL63. The location of ORFib Tyr6398 (p59 Tyr4l4) is indicated by box and star. Potential active residues are indicated by regions (horizontal bars) and zinc fingers by +(SEQ. ID NOS: 1-9).

FIGS. **3**A-B - Mutations in nsp14 (ExoN) attenuate virulence in mice. (FIG. **3**A) Schematic of genome is shown, with enlargement of nsp14 (p59-ExoN) and ORF2a protein (Pro_{106}). Nt and aa sequence of published MHV sequences, and those of sequenced virulent A59 strains is shown in the boxes. Changes identified in clones and virus sequence of "uncorrected" icwt are shown by nt and aa sequence. (FIG. **3**B). LD_{50} data on mice infected ic with viruses as shown for "uncorrected icwt" and corrected for nsp14 and ORF2A proteins, alone and together.

FIG. **4** - Location and sequence of 3CLpro cleavage sites. CS11, 12, and 13. Schematic of 1a/b polyprotein shows PLP and 3CLpro cleavage sites, with location of CS11, 12, and 13. Amino acid sequence flanking cleavage sites is shown, with LQ_(S,A) indicating cleavage between Q and S or A.

FIGS. **5**A-E - Mutations approaches and sequences of nsp14, nsp15 and nsp16. (FIG. **5**A) Organization of proteins and introduction of stop codons beginning in nsp16 toward nsp14: yellow-"midprotein" allowing partial translation and cleavage; red -change initial residue to stop codon. (FIG. **5**B) Cleavage site mutations from Aim 1 (grey boxes) and introduction of inactivation cleavage sites (white boxes with arrows). (FIGS. **5**C-E) Core sequences with conserved residues with subscript residue numbers in pp1ab. For nsp 14, putative metal finger is underlined (SEQ ID NOS:10-12).

DESCRIPTION OF ILLUSTRATIVE EMBODIMENTS

Coronaviruses are common pathogens for respiratory, enteric, and neurologic diseases in humans and animals. It has been estimated that coronaviruses are responsible for up to 30% of common colds in humans, and they cause economically significant diseases in domestic animals including cows, chickens, swine, and cats. Group 2 viruses include mouse hepatitis virus (MHV), human coronavirus OC-43, and bovine coronavirus, and the recently identified severe acute respiratory syndrome-associated coronavirus (SARS-CoV) appears to be closely related to the group 2 coronaviruses. Thus, there is considerable interest in the development of vaccines against coronaviruses generally, and against SARS-CoV in particular.

I. The Present Invention

The present inventors isolated an attenuated infectious clone virus produced using an MHV reverse genetics system, and identified two unique mutations in different coding regions of the viral genome. By assembling several infectious clone viruses with each mutation corrected individually and simultaneously, they demonstrated that these mutations both independently and in combination attenuate virulence in mice, while not displaying any distinctive phenotypes, and specifically no inhibition of growth, in DBT cell culture relative to lab strain MHV-A59. One of the mutations, leading to a leucine to proline change at amino acid 106 in the Orf2a protein, was only partially attenuating in the absence of any other changes. The second mutation, a tyrosine to histidine change at amino acid 6398 in the Orf1a/b polyprotein (p59), was completely attenuating independently or in combination with the first mutation.

Neither of these mutations affected virus growth in DBT cells, though there was evidence of replication defects in mice for the attenuated viruses. Interestingly, both of the attenuating mutation sites are located outside of the known virulence- or tropism-associated loci of MHV: S, M, N, and HE. These attenuating mutations are also not located in known structural proteins. There have been previous suggestions that genes outside of S and the structural proteins might be involved in mediating pathogenic properties, but this is the first demonstrations of specific lesions of such. Significantly, the discovery of these attenuating mutations should be applicable to other coronaviruses, as all coronaviruses conserve the tyrosine residue in their p59 homologous protein in Orf1b.

The discovery of two randomly generated mutations in the cDNA of the infectious clone, which were independently and simultaneously attenuating in mice but caused no observed phenotype in cell culture, was very surprising. The inventor's hypothesis is that these mutations were selected for during passage of the F clone plasmid in *E. coli* cells. Experience during cloning with this plasmid suggested that it was not stable in bacteria. Following correction of nt 19400 and growth of the plasmid in *E. coli* cells, the mutation would occasionally spontaneously revert. To stabilize the F fragment, it was subcloned it into pSMART-LCAmp vector (Lucigen), which resulted in increased stability of the F clone.

The attenuated viruses described herein have the key features of (a) replicating to high titers in culture (>10^8 pfu/ml), (b) diminished replication in animals, and (c) failure to cause disease or illness. The main potential drawback for any live-attenuated vaccine development in coronaviruses is the possibility of reversion or recombination that might restore a virulent phenotype or alter host cell tropism and disease. One possible approach to avoid or overcome this problem is the identification and introduction of multiple attenuating mutations spaced throughout the genome of the virus. Multiple mutations could be introduced singly or in combinations by introducing them on different cDNA fragments. The presence of the multiple mutations would guard against the risk of losing attenuation due to single recombination events or single-site reversions. It would also maintain or introduce attenuating mutations in any recombinant viruses that might be generated by recombination between different coronaviruses.

For human viruses acquired by the respiratory route, live-attenuated vaccines impaired in protein processing would have several potential advantages. Because there is no alteration in the viral structural glycoproteins, it is predicted that replicase protein cleavage mutants would have normal "wild-type" transmission, tropism, attachment, entry and uncoating, and thus could theoretically be administered by oral, intranasal or inhaled approaches. The initial replication and spread from the respiratory epithelium and lymphoid organs also allows for the development of both systemic and mucosal inmmunity.

Furthermore, studies with other animal coronavirus vaccines suggest that viral replication may be necessary for protection from virus challenges. The use of a virus with multiple virulence-attenuating mutations avoids concerns about atypical infections with wild-type viruses following vaccination with inactivated viruses or purified viral proteins, such as occurred with measles virus and respiratory syncytial virus, and also seen with the vaccines for the feline coronavirus, FIPV. Most importantly, the use of a live-attenuated virus allows for both humoral and cellular immunity.

II. Coronaviridae

Viruses in this family infect hosts in the Domain Eucarya, Kingdom Animalia, Phylum Chordata, Subphylum Vertebrata, Classes Mammalia and Aves, Orders Primates, Camivora, Perissodactyla, Artiodactyla, Rodentia, and Lagomorphia. It is transmitted by means not involving a vector. Worldwide distribution is likely.

Virions are enveloped, slightly pleomorphic, spherical or kidney shaped, and about 120-160 nm in diameter. Surface projections of envelope are distinct, club-shaped, spaced widely apart and dispersed evenly over all the surface. Nucleocapsids are rod-shaped (straight or bent), about 9-13 nm in diameter. Virions associated RNA nucleocapsids exhibit helical or tubular symmetry.

Molecular mass (Mr) of the virion 400×106. Buoyant density is 1.23-1.24 g cm-3 in CsCl, and 1.15-1.19 g cm-3 in sucrose. The sedimentation coefficient is 300-500S. Under in

vitro conditions, virions are stable in acid environment (pH 3), relatively stable in presence of Mg^{++}. Virions are sensitive to heat, lipid solvents, non-iomc detergents, formaldehyde, and oxidizing agents.

Virions contain one molecule of linear positive-sense single stranded RNA with a total genome length is 20,000-33,000 nt. The 5' end of the genome has a cap, and the 3' end has a poly(A) tract. Subgenomic mRNA is found in infected cells.

Five structural virion proteins found ranging is size between 18,0000 and 220,000 Da. The first is the surface glycoprotein or spike (S) protein. The S protein is responsible for attachment to cells, hemagglutination and membrane fusion. It has a carboxy-terminal half with a coiled-coil structure. The second largest protein (30,000-35,000 Da) is the integral membrane protein (M) which spans the virus envelope three times, with only 10% protruding at the virion surface. The third largest protein (50,000-60,000 Da) is the nucleocapsid protein (N). The fourth largest protein (65,000 Da) is the hemagglutinine-esterase protein (HE), which forms short surface projections, and can have receptor binding, hemagglutination and receptor destroying activities. The fifth largest protein (10,000-12,000 Da) is tentatively designated as the small membrane protein (sM), detected in avian infectious bronchitis virus (IBV) and porcine transmissible gastroenteritis virus (TGEV).

The virus exhibits distinct antigen determinants on envelope and spikes, those corresponding to each of the major structural glycoproteins—S, HE, M, and N. Antigenic specificity of virion can be determined by neutralization tests (S and HE), or complement fixation tests (M). Protective immunity is induced in form of complement independent neutralizing antibodies.

The Coronaviridae family is split into two groups—coronavirus and torovirus. Coronaviruses include avian infectious bronchitis virus, bovine coronavirus, canine coronavirus, feline infectious peritonitis virus, human coronavirus 229E, human coronavirus OC43, murine hepatitis virus, porcine epidemic diarrhea virus, porcine hemagglutinating encephalomyelitis virus, porcine transmissible gastroenteritis virus, rat coronavirus, turkey coronavirus, severe acute respiratory syndrome virus, rabbit coronavirus, or the recently identified SARS associated human coronavirus. Toroviruses include Berne virus or Breda virus.

III. Targets for Attenuating Mutations

In accordance with the present invention, attenuating mutations have been identified in the p59 and Orf2a proteins of MHV-A59. The p59 residue (Tyr6398) (also referred to as nsp14 and Exo N) is one of many residues in the replicase that are 100% conserved across all coronaviruses, and therefore provide the basis for making similar changes in analogous residues of the other coronaviruses. In addition, such mutations may be combined into a single virus, and with other mutations in structural and non-structural genes. The p59 and Orf2a proteins, as well as other targets, are described below.

A. Nsp14/p59 (ExoN)

The p59 protein if part of the complex polyprotein, described in greater detail below. No function has been demonstrated for this protein, though this highly conserved protein is predicted to be an exoribonuclease (ExoN) of the DEDD superfamily of exonucleases. This prediction is based on primary amino acid sequence identity of three motifs containing the putative catalytic aspartic and glutamic acid residues necessary for the exonuclease activity.

The virulence-associated tyrosine/histidine residue is located in a region 140 amino acids carboxy-terminal to the last predicted ExoN catalytic motif. Analysis of deduced amino acid sequences of all genome-sequenced coronaviurses shows increasing identity across the carboxy-terminal half of the ORF 1a polyprotein and all of ORF1b polyrpotein. Specifically, up to 60% to 80% identity is observed across putative fuictional proteins such as the RNA-dependent RNA polymerase (RdRp, pol, nsp12), the ATPase/helicase (hel, nsp14, p67). Further, the amino acid identity within proteins is focused in regions or motifs. However for proteins with predicted functions, the predictions were based on organization or sequences discrete regions, and thus the identity across the remainder of the proteins is of unknown significance in protein function, viral replication or viral pathogenesis. This suggests that many of these regions of identity may play identical conserved roles, but distinct from the predicted functions.

B. Orf2a

The Orf2a protein also has no known function. It has been reported to be a non-structural protein that is cytosolic and non-membrane associated. The Orf2a protein was shown to be non-essential for MHV replication in cell culture, and its deletion did not affect growth, RNA synthesis, or protein expression, leading to the suggestion that its functional role may only be manifest in vivo. It is predicted to be a cyclic phosphodiesterase (CPD), another RNA processing enzyme associated with tRNA maturation.

The mutated leucine residue is located in the Orf2a protein, of group 2 coronaviruses including, including MHV, OC43, and BCoV. The leucine residue is conserved among those viruses that possess the Orf2a protein. However, the mutation is not associated with the predicted catalytic residues of the putative enzyme. While this ORF and protein is not conserved among all coronaviruses, there are several medically and agriculturally important viruses in this group. In addition, the fact the ORF 2a mutation is independently attenuating in animals but has no independent or synergistic impact on replication in culture suggests that mutations in both conserved and non-conserved ORFS of different group coronaviruses will aid efforts to introduce multiple mutation across the genome that stabilize against or eliminate reversion to virulence or recombination.

C. Nsp15/XendoU

The prediction of nsp15 as a XendoU ortholog (poly(U)-specific endoribonuclease) is based on alignment of a small number of identical and similar residues within a region of the predicted protein sequence. Since active site residues for XendoU proteins have not been defined, the prediction is both tenuous and difficult to prove biochemically. However, the predictions do identify residues that are highly conserved in the coronavirus proteins and relatively with other non-viral proteins known to have XendoU function. These glycine, histidine, and lysine residues may be substituted with alanine and with conservative and non-conservative residues.

D. Nsp16/2' O-methyltransferase.

The 2'-O-MT activity of nsp16 was predicted based on alignment of residues with those of known proteins with methyltransferase function, specifically the RrmJ family. However, this is only a subset of the critical residues of the known enzymes and thus establishment of O-MT activity will require experimental confirmation of function. At this time, it is unclear if this protein actually methylates the RNA cap, the penultimate nucleotide, or both within the genome. For SARS-CoV nsp16, lys^{6821}, lys^{6945}, and asp^{6905} have been proposed to be the site of a catalytic triad. Mutations may be introduced at lys^{6821}, lys^{6945}, and asp^{6905}, and will include

both conservative and non-conservative changes; for Lys these will include alanine, histidine, arginine, and proline; for asp, substitutions include alanine, asparagine, glutamic acid, histidine, and proline. Other potential mutatable residues are conserved in the RrmJ family and between coronaviruses and include leu^{6936}, gly^{6940}, glu^{6978}, and gly^{6983}. Since coronaviruses appear to mediate all stages of their mRNA synthesis in the host cell cytoplasm, presumably including addition of methylguanosine caps to mRNAs, nsp16 is an attractive target for a specific function in viral replication. If, in fact, nsp16 is involved in mRNA capping, then it would possess a presumably critical activity, and alteration of residues known to be required for MT activity of known enzymes would likely reduce or abolish this activity.

E. Replicase Polyprotein Cleavages Sites

1. Gene and Protein Structure

The coronavirus replicase gene (also known as gene 1 or the polymerase gene) comprises 22 kB of the coronavirus genome, corresponding to some 7800 amino acids, and is composed of two overlapping open reading frames—ORF1a and 1b. Following uncoating of the RNA genome in the cell cytoplasm, the replicase gene is translated as either an ORF la polyprotein (495 kD) or as an ORF1ab fusion polyprotein (803 kD), with translation of ORF 1b requiring a ribosomal frameshift event at the end of ORF 1 a. The intact replicase polyproteins are not detected during natural infection, since maturation proteolytic cleavages occur cotranslationally by three proteinase functions encoded in ORF la polyprotein. The proteolytic processing results in 15 mature proteins, including the proteinases, an RNA helicase, and a putative RNA-dependent RNA polymerase. The MHV proteins are set forth in Table 1.

TABLE 1

MHV Replicase Polyproteins

Protein Designation	Residues	Function
p28	1-247	Unknown; localizes to replication complexes early in infection where it associates with membranes by easily disrupted peripheral mechanisms
p65	248-834	Unknown; shown to associate throughout infection with membranes of replication complexes at sites of viral RNA synthesis, likely by interactions with other proteins
P210	833-2837	Encodes a protein with two papain-like proteinase domains that cleave the first three (C1-C3) cleavage sites
MP1	2838-3332	Highly hydrophobic, membrane associated, found in replication complexes
3CLpro	3333-3633	Picornain like proteinase responsible for cleavage at C4-C14
MP2		Highly hydrophobic, membrane associated, localization in cells unknown
p10		Associates with p22, p12, p15 in replication complexes
p22		Associates with p10, p12, p15 in replication complexes
p12		Associates with p22, p10, p15 in replication complexes
p15		Associates with p22, p12, p10 in replication complexes
Polymerase (pol, p100)		Putative RNA dependent RNA polymerase, localizes to replication complexes
Helicase (hel, p67)		RNA unwinding and NTPase activities
p57		Unknown
p42		Unknown
p33		Unknown

TABLE 1-continued

MHV Replicase Polyproteins

Protein Designation	Residues	Function
Polymerase (pol, p100)		Putative RNA dependent RNA polymerase, localizes to replication complexes
Helicase (hel, p67)		RNA unwinding and NTPase activities
p57		Unknown
p42		Unknown
p33		Unknown

The replicase gene expresses all of the viral factors required for all stages of MHV mRNA synthesis and replication. In addition, it has been shown that inhibition of polyprotein processing at any time during infection results in rapid shutoff of viral RNA synthesis, indicating that at least some of the proteolytic processing events are required for RNA synthesis. However there are differences in mature replicase proteins among different coronaviruses, particularly in the amino-terminal 100 kD of the polyproteins.

2. Cleavage Sites

By convention, the present invention identifies the polyprotein cleavage sites as C1-C14. These sites are defined as cleaving between adjacent products. Examples for MHV are set forth in Table 2, below.

TABLE 2

Replicase Polyprotein Cleavage Sites

Cleavage Site	Upstream Protein	Downstream Protein
C1	p28	p65
C2	p65	p210
C3	P210	MP1
C4	MP1	3C
C5	3C	MP2
C6	MP2	p10
C7	p10	p22
C8	p22	p12
C9	p12	p15
C10	p15	Polymerase
C11	Polymerase	Helicase
C12	Helicase	p57
C13	p57	p42
C14	p42	p33

The third protein processed from the replicase polyprotein is p210 (Schiller et al., 1998). The p210 protein incorporates amino acids 833 to a predicted carboxy-terminus at amino acid 2837, with a predicted mass of 221 kD. p210 contains the two papain-like proteinase domains (PLP1 and PLP2) that have been shown to cleave the first three cleavage sites (CS1, CS2 and CS3) at the carboxy-termini of p28, p65 and p210, respectively. The apparent difference between coronaviruses in the predicted number of proteinases, and the differences in the size and number of proteins in the amino-terminal half of the polyprotein, was interpreted to indicate a lack of common critical functions in this region of gene. A recent study used sequence comparisons, parsimony analyses, and studies of the cleavage sites and proteinase functions to compare the coronavirus p210 and the corresponding p195 proteins of the human coronavirus 229E (HCoV-229E) and infectious bronchitis virus (IBV) (Ziebuhr et al., 2001). The analyses identified common domains of the coronavirus p210/p195 proteins (FIG. **2**), several of which had previously been predicted or confirmed for MHV (Lee et al., 1991). The amino-terminal

domain of p210 was referred to as the "acidic domain" (Ac) based on the concentration of acidic residues. The PLP1 domain consists of the sequence required for proteinase activity during in vitro cleavage reactions (Bonilla et al. 1995). The X domain is a region of increased conservation among the different coronavirus p210/p195 proteins with no known or predicted functions (Lee et al., 1991). The functional PLP2 domains are a variable distance from the X domains, and have been less completely characterized as to their functional requirements. Both PLP1 and PLP2 have been demonstrated to function with a catalytic dyad of Cys and His residues (Baker et al., 1993; Bonilla et al., 1995; Kanjanahaluethai and Baker, 2000). Finally, a Y domain consists of a region incorporating two stretches of predominantly hydrophobic residues that predict membrane-spanning helices (Lee et al., 1991).

Coronavirus PLPs have a zinc finger motif in the predicted papain-like fold of the enzymes, with predicted similarities to the human transcription elongation factor TFIIS (Herold et al., 1999). The zinc finger has been shown to bind zinc, which is required for PLP function in vitro. Mutations in this motif abolish proteolytic activity. It has been suggested based on these features and demonstrated contributions of the zinc finger to RNA synthesis in the arterivirus, equine arteritis virus (EAV) (Tijms et al., 2001), that the zinc finger may serve functions in addition to PLP proteolytic activity.

Studies of PLP1 and PLP2, as well as identification and detailed mutagenesis of replicase polyprotein cleavage sites, have been performed in vitro. PLP1 has been shown to proteolytically process the first two cleavage sites in the MHV replicase polyprotein: between p28 and p65 at 247G/V248 (referred to as CS1) and between p65 and p210 at 832A/G833 (CS2) (Dong and Baker, 1994; Hughes et al., 1995; Bonilla et al., 1997; Baker et al., 1993). PLP2 has been shown to cleave at the carboxy-terminus of p210 (CS3), likely in a cis autocat (Kanjanahaluethai and Baker, 2000; Kanjanahaluethai et al., 2001). Although the MHV CS3 cleavage site has not been reported, by direct comparison with identified IBV PLP2 cleavage site the MHV-A59 p210 carboxy-terminal cleavage (CS3) would be predicted to be 2837G/A2838. Analysis of the MHV CS1 and CS2 in comparison with other group 1 coronaviruses (TGEV, HCoV-229E) (Elcouet et al., 1995; Herold et al., 1993), group 2 coronaviruses (MHV-JHM, BCV) (Yoo and Pei, 2001; Chouljenko et al., 2001), and group 3 coronaviruses (IBV) (Boursnell et al., 1987), has demonstrated similarities at the P1/P1' cleavage dipeptides; Gly or Ala at P1 of all coronavirus PLP CS, and Val, Ala or Gly at P1'. HCoV is the exception, using Asn in the P1' position. Overall, P5, P2, P1 and P1' have been most intolerant of changes, with mutations at these sites disrupting cleavage in vitro.

Analysis of the coronavirus PLPs and their cognate cleavage sites suggests that PLP1 and PLP2 are paralogous proteinases, originating from a common coronavirus or pre-coronavirus ancestor, and that they have diverged over time (Ziebuhr et al., 2001) (FIG. **3**). For example, all coronaviruses except IBV express both PLP1 and PLP2 activities and share the common feature that PLP1 cleaves CS1 and CS2. IBV only expresses a PLP2 that cleaves at a single site equivalent to CS2. In IBV, a functional PLP1 is not detected, whereas a residual, highly altered and inactive PLP 1 domain has recently been identified by sequence comparison (Ziebuhr et al., 2001). These observations have led to the hypothesis that there may be overlap of cleavage site specificity and PLP activity, and possible redundancy of cleavage activity, with PLP 2 able to mediate cleavages at PLP1 cognate sites. This has been demonstrated to be true for HCoV, with both PLP1 and PLP2 able to cleave CS2 in vitro (Ziebuhr et al., 2001). In fact, the data suggest that the "normal" CS2 cleavage event may involve the cooperative activity of PLP1 and PLP2. However, it was also demonstrated that when PLP 1 was catalytically inactivated, PLP2 was able to independently mediate CS2 cleavage in vitro.

IV. Engineering of Coronaviridae Genomes

Thus, in accordance with the present invention, it will be desirable to create a variety of different mutants in Coronaviridae proteins. Mutagenesis is the process whereby changes occur in the structure of a genome. Mutation can involve modification of the nucleotide sequence of a single gene, blocks of genes or a whole chromosome. Changes in single genes may be the consequence of point mutations which involve the removal, addition or substitution of a single nucleotide base within a DNA sequence, or they may be the consequence of changes involving the insertion or deletion of large numbers of nucleotides.

Mutations can arise spontaneously as a result of events such as errors in the fidelity of DNA replication or the movement of transposable genetic elements (transposons) within the genome. They also are induced following exposure to chemical or physical mutagens. Such mutation-inducing agents include ionizing radiations, ultraviolet light and a diverse array of chemical such as alkylating agents and polycyclic aromatic hydrocarbons all of which are capable of interacting either directly or indirectly (generally following some metabolic biotransformations) with nucleic acids. The DNA lesions induced by such environmental agents may lead to modifications of base sequence when the affected DNA is replicated or repaired and thus to a mutation. Mutation also can be site-directed through the use of particular targeting methods. Any number of different mutagenic approaches may be taken, as described below.

A. Coronaviridae Genomes

One of skill in the art may use various Coronaviridae sequences to design specific mutations that create attenuated viruses. The following constitute non-limiting examples of accession nos., each of which are incorporated by reference: human coronavirus 229E (NC002645), SARS TOR2 (AY274119), SARS HKU-39849 (AY278491), SARS CUHK-W1 (AY278554), bovine coronavirus (BCV) (NC003045), avian infectious bronchitis virus (IBV) (NC001451), transmissible gastroenteritis virus (TGEV), (NC002306), mouse hepatitis virus (MHV) (NC001846).

B. Random Mutagenesis

In one embodiment, random mutagenesis may be applied. This will, of course, require an additional step of screening for the desired mutations. Screening will typically be accomplished by nucleic acid hybridization (Southern or Northern blotting), sequencing, or SnP analysis, methods of which are well known to those of skill in the art.

1. Insertional Mutagenesis

Insertional mutagenesis is based on the inactivation of a gene via insertion of a known DNA fragment. Because it involves the insertion of some type of DNA fragment, the mutations generated are generally loss-of-function, rather than gain-of-function mutations. However, there are several examples of insertions generating gain-of-function mutations (Oppenheimer et al. 1991). Insertion mutagenesis has been very successful in bacteria and *Drosophila* (Cooley et al. 1988) and recently has become a powerful tool in corn (Schmidt et al. 1987); *Arabidopsis;* (Marks et al., 1991; Koncz et al. 1990); and *Antirrhinum* (Sommer et al. 1990).

Transposable genetic elements are DNA sequences that can move (transpose) from one place to another in the genome of a cell. The first transposable elements to be recognized were the Activator/Dissociation elements of *Zea mays* (McClintock, 1957). Since then, they have been identified in a wide range of organisms, both prokaryotic and eukaryotic.

Transposable elements in the genome are characterized by being flanked by direct repeats of a short sequence of DNA that has been duplicated during transposition and is called a target site duplication. Virtually all transposable elements whatever their type, and mechanism of transposition, make such duplications at the site of their insertion. In some cases the number of bases duplicated is constant, in other cases it may vary with each transposition event. Most transposable elements have inverted repeat sequences at their termini. These terminal inverted repeats may be anything from a few bases to a few hundred bases long and in many cases they are known to be necessary for transposition.

Prokaryotic transposable elements have been most studied in *E. coli* and Gram negative bacteria, but also are present in Gram positive bacteria. They are generally termed insertion sequences if they are less than about 2 kB long, or transposons if they are longer. Bacteriophages such as mu and D108, which replicate by transposition, make up a third type of transposable element. Elements of each type encode at least one polypeptide a transposase, required for their own transposition. Transposons often further include genes coding for function unrelated to transposition, for example, antibiotic resistance genes.

Transposons can be divided into two classes according to their structure. First, compound or composite transposons have copies of an insertion sequence element at each end, usually in an inverted orientation. These transposons require transposases encoded by one of their terminal IS elements. The second class of transposon have terminal repeats of about 30 base pairs and do not contain sequences from IS elements.

Transposition usually is either conservative or replicative, although in some cases it can be both. In replicative transposition, one copy of the transposing element remains at the donor site, and another is inserted at the target site. In conservative transposition, the transposing element is excised from one site and inserted at another.

Eukaryotic elements also can be classified according to their structure and mechanism of transportation. The primary distinction is between elements that transpose via an RNA intermediate, and elements that transpose directly from DNA to DNA.

Elements that transpose via an RNA intermediate often are referred to as retrotransposons, and their most characteristic feature is that they encode polypeptides that are believed to have reverse transcriptionase activity. There are two types of retrotransposon. Some resemble the integrated proviral DNA of a retrovirus in that they have long direct repeat sequences, long terminal repeats (LTRs), at each end. The similarity between these retrotransposons and proviruses extends to their coding capacity. They contain sequences related to the gag and pol genes of a retrovirus, suggesting that they transpose by a mechanism related to a retroviral life cycle. Retrotransposons of the second type have no terminal repeats. They also code for gag- and pol-like polypeptides and transpose by reverse transcription of RNA intermediates, but do so by a mechanism that differs from that or retrovirus-like elements. Transposition by reverse transcription is a replicative process and does not require excision of an element from a donor site.

Transposable elements are an important source of spontaneous mutations, and have influenced the ways in which genes and genomes have evolved. They can inactivate genes by inserting within them, and can cause gross chromosomal rearrangements either directly, through the activity of their transposases, or indirectly, as a result of recombination between copies of an element scattered around the genome. Transposable elements that excise often do so imprecisely and may produce alleles coding for altered gene products if the number of bases added or deleted is a multiple of three.

Transposable elements themselves may evolve in unusual ways. If they were inherited like other DNA sequences, then copies of an element in one species would be more like copies in closely related species than copies in more distant species. This is not always the case, suggesting that transposable elements are occasionally transmitted horizontally from one species to another.

2. Chemical Mutagenesis

Chemical mutagenesis offers certain advantages, such as the ability to find a full range of mutant alleles with degrees of phenotypic severity, and is facile and inexpensive to perform. The majority of chemical carcinogens produce mutations in DNA. Benzo[a]pyrene, N-acetoxy-2-acetyl aminofluorene and aflotoxin B1 cause GC to TA transversions in bacteria and mammalian cells. Benzo[a]pyrene also can produce base substitutions such as AT to TA. N-nitroso compounds produce GC to AT transitions. Alkylation of the 04 position of thymine induced by exposure to n-nitrosoureas results in TA to CG transitions.

A high correlation between mutagenicity and carcinogenity is the underlying assumption behind the Ames test (McCann et al., 1975) which speedily assays for mutants in a bacterial system, together with an added rat liver homogenate, which contains the microsomal cytochrome P450, to provide the metabolic activation of the mutagens where needed.

In vertebrates, several carcinogens have been found to produce mutation in the ras proto-oncogene. N-nitroso-N-methyl urea induces mammary, prostate and other carcinomas in rats with the majority of the tumors showing a G to A transition at the second position in codon 12 of the Ha-ras oncogene. Benzo[a]pyrene-induced skin tumors contain A to T transformation in the second codon of the Ha-ras gene.

3. Radiation Mutagenesis

The integrity of biological molecules is degraded by the ionizing radiation. Adsorption of the incident energy leads to the formation of ions and free radicals, and breakage of some covalent bonds. Susceptibility to radiation damage appears quite variable between molecules, and between different crystalline forms of the same molecule. It depends on the total accumulated dose, and also on the dose rate (as once free radicals are present, the molecular damage they cause depends on their natural diffusion rate and thus upon real time). Damage is reduced and controlled by making the sample as cold as possible.

Ionizing radiation causes DNA damage and cell killing, generally proportional to the dose rate. Ionizing radiation has been postulated to induce multiple biological effects by direct interaction with DNA, or through the formation of free radical species leading to DNA damage (Hall, 1988). These effects include gene mutations, malignant transformation, and cell killing. Although ionizing radiation has been demonstrated to induce expression of certain DNA repair genes in some prokaryotic and lower eukaryotic cells, little is known about the effects of ionizing radiation on the regulation of mammalian gene expression (Borek, 1985). Several studies have described changes in the pattern of protein synthesis observed after irradiation of mammalian cells. For example, ionizing radiation treatment of human malignant melanoma cells is

associated with induction of several unidentified proteins (Boothman et al., 1989). Synthesis of cyclin and co-regulated polypeptides is suppressed by ionizing radiation in rat REF52 cells, but not in oncogene-transformed REF52 cell lines (Lambert and Borek, 1988). Other studies have demonstrated that certain growth factors or cytokines may be involved in x-ray-induced DNA damage. In this regard, platelet-derived growth factor is released from endothelial cells after irradiation (Witte et al., 1989).

In the present invention, the term "ionizing radiation" means radiation comprising particles or photons that have sufficient energy or can produce sufficient energy via nuclear interactions to produce ionization (gain or loss of electrons). An exemplary and preferred ionizing radiation is an x-radiation. The amount of ionizing radiation needed in a given cell generally depends upon the nature of that cell. Typically, an effective expression-inducing dose is less than a dose of ionizing radiation that causes cell damage or death directly. Means for determining an effective amount of radiation are well known in the art.

In a certain embodiments, an effective expression inducing amount is from about 2 to about 30 Gray (Gy) administered at a rate of from about 0.5 to about 2 Gy/minute. Even more preferably, an effective expression inducing amount of ionizing radiation is from about 5 to about 15 Gy. In other embodiments, doses of 2-9 Gy are used in single doses. An effective dose of ionizing radiation may be from 10 to 100 Gy, with 15 to 75 Gy being preferred, and 20 to 50 Gy being more preferred.

Any suitable means for delivering radiation to a tissue may be employed in the present invention in addition to external means. For example, radiation may be delivered by first providing a radiolabeled antibody that immunoreacts with an antigen of the tumor, followed by delivering an effective amount of the radiolabeled antibody to the tumor. In addition, radioisotopes may be used to deliver ionizing radiation to a tissue or cell.

4. In vitro Scanning Mutagenesis

Random mutagenesis also may be introduced using error prone PCR (Cadwell and Joyce, 1992). The rate of mutagenesis may be increased by performing PCR in multiple tubes with dilutions of templates.

One particularly useful mutagenesis technique is alanine scanning mutagenesis in which a number of residues are substituted individually with the amino acid alanine so that the effects of losing side-chain interactions can be determined, while minimizing the risk of large-scale perturbations in protein conformation (Cunningham and Wells, 1989).

In recent years, techniques for estimating the equilibrium constant for ligand binding using minuscule amounts of protein have been developed (Blackburn et al., 1991; U.S. Pat. Nos. 5,221,605 and 5,238,808). The ability to perform functional assays with small amounts of material can be exploited to develop highly efficient, in vitro methodologies for the saturation mutagenesis of antibodies. The inventor bypassed cloning steps by combining PCR mutagenesis with coupled in vitro transcription/translation for the high throughput generation of protein mutants. Here, the PCR products are used directly as the template for the in vitro transcription/translation of the mutant single chain antibodies. Because of the high efficiency with which all 19 amino acid substitutions can be generated and analyzed in this way, it is now possible to perform saturation mutagenesis on numerous residues of interest, a process that can be described as in vitro scanning saturation mutagenesis (Burks et al., 1997).

In vitro scanning saturation mutagenesis provides a rapid method for obtaining a large amount of structure-function information including: (I) identification of residues that modulate ligand binding specificity, (ii) a better understanding of ligand binding based on the identification of those amino acids that retain activity and those that abolish activity at a given location, (iii) an evaluation of the overall plasticity of an active site or protein subdomain, (iv) identification of amino acid substitutions that result in increased binding.

5. Random Mutagenesis by Fragmentation and Reassembly

A method for generating libraries of displayed polypeptides is described in U.S. Pat. No. 5,380,721. The method comprises obtaining polynucleotide library members, pooling and fragmenting the polynucleotides, and reforming fragments there from, performing PCR amplification, thereby homologously recombining the fragments to form a shuffled pool of recombined polynucleotides.

C. Site-Directed Mutagenesis

Structure-guided site-specific mutagenesis represents a powerful tool for the dissection and engineering of protein-ligand interactions (Braisted and Wells, 1996), especially in the context of the present invention where specific mutations in cleavage sites are sought. The technique provides for the preparation of sequence variants by introducing one or more discrete nucleotide sequence changes into a selected nucleic acid.

Site-specific mutagenesis uses specific oligonucleotide sequences which encode the DNA sequence of the desired mutation, as well as a sufficient number of adjacent, unmodified nucleotides. In this way, a primer sequence is provided with sufficient size and complexity to form a stable duplex on both sides of the deletion junction being traversed. A primer of about 17 to 25 nucleotides in length is preferred, with about 5 to 10 residues on both sides of the junction of the sequence being altered.

The technique typically employs a bacteriophage vector that exists in both a single-stranded and double-stranded form. Vectors useful in site-directed mutagenesis include vectors such as the M13 phage. These phage vectors are commercially available and their use is generally well known to those skilled in the art. Double-stranded plasmids are also routinely employed in site-directed mutagenesis, which eliminates the step of transferring the gene of interest from a phage to a plasmid.

In general, one first obtains a single-stranded vector, or melts two strands of a double-stranded vector, which includes within its sequence a DNA sequence encoding the desired protein or genetic element. An oligonucleotide primer bearing the desired mutated sequence, synthetically prepared, is then annealed with the single-stranded DNA preparation, taking into account the degree of mismatch when selecting hybridization conditions. The hybridized product is subjected to DNA polymerizing enzymes such as *E. coli* polymerase I (Klenow fragment) in order to complete the synthesis of the mutation-bearing strand. Thus, a heteroduplex is formed, wherein one strand encodes the original non-mutated sequence, and the second strand bears the desired mutation. This heteroduplex vector is then used to transform appropriate host cells, such as *E. coli* cells, and clones are selected that include recombinant vectors bearing the mutated sequence arrangement.

Comprehensive information on the functional significance and information content of a given residue of protein can best be obtained by saturation mutagenesis in which all 19 amino acid substitutions are examined. The shortcoming of this approach is that the logistics of multiresidue saturation mutagenesis are daunting (Warren et al., 1996, Zeng et al., 1996; Barbas et al., 1994; Yelton et al., 1995; Wong et al.,

1996; Hilton et al., 1996). Hundreds, and possibly even thousands, of site specific mutants must be studied. However, improved techniques make production and rapid screening of mutants much more straightforward. See also, U.S. Pat. Nos. 5,798,208 and 5,830,650, for a description of "walk-through" mutagenesis.

Other methods of site-directed mutagenesis are disclosed in U.S. Pat. Nos. 5,220,007; 5,284,760; 5,354,670; 5,366,878; 5,389,514; 5,635,377; and 5,789,166.

D. Virus Transformation and Propagation

Targeted recombination has become a powerful tool to introduce mutations into the genome and determine their effects on protein function, virus replication and virus pathogenesis (Koetzner et al., 1992; Masters et al., 1994; Fischer et al., 1997; Lavi et al., 1998; Leparc-Goffart et al., 1998; Phillips et al., 1999; Sanchez et al., 1999; Phillips et al., 2001; de haan et al., 2002; Sarma et al., 2002). However, the available recombination constructs and methodologies have thus far limited the use of targeted recombination, and have not been employed to examine mutations in the replicase gene.

The inventor has previously collaborated in the development of a system for assembly of full-length MHV genome cDNA, generation of genome length RNA, and recovery of virus from transfected cells (Schaad et al., 1990; Yount et al., 2002). In this process, seven contiguous cDNA clones that spanned the 31.5-kb genome of mouse hepatitis virus strain A59 (MHV-A59) were isolated. The ends of the cDNAs were engineered with unique junctions and assembled with only the adjacent cDNA subclones, resulting in an intact MHV-A59 cDNA construct of about 31.5 kb in length. The interconnecting restriction site junctions that are located at the ends of each cDNA are systematically removed during the assembly of the complete full-length cDNA product, allowing reassembly without the introduction of nucleotide changes.

RNA transcripts derived from the full-length MHV-A59 construct were infectious, although transfection frequencies were enhanced 10- to 15-fold in the presence of transcripts encoding the nucleocapsid protein N. Plaque-purified virus derived from the infectious construct replicated efficiently and displayed similar growth kinetics, plaque morphology, and cytopathology in murine cells as did wild-type MHV-A59. Molecularly cloned viruses recognized the MHV receptor (MHVR) for docking and entry, and pretreatment of cells with monoclonal antibodies against MHVR blocked virus entry and replication. Cells infected with molecularly cloned MHV-A59 virus expressed replicase (gene 1) proteins identical to those of laboratory MHV-A59. Importantly, the molecularly cloned viruses contained three marker mutations that had been derived from the engineered component clones.

Using this process, full-length infectious constructs of MHV-A59 and other coronaviruses with genetic modifications of may be created. In fact, the method has the potential to be used to construct viral, microbial, or eukaryotic genomes approaching several million base pairs in length and used to insert restriction sites at any given nucleotide in a microbial genome. A similar system approach was used previously with TGEV, including the insertion of heterologous genes into the TGEV genome (Yount, 2000; Curtis et al., 2002). The inventor described herein the use of this same assembly approach to introduce five different mutations into the MHV p28/p65 cleavage site (CS 1). While the approaches are similar, it was not usually necessary with MHV to introduce mutations and new restriction sites into the wild-type virus genome to direct the assembly cascade. Rather, type IIS restriction endonuclease Esp3I sites can be used to create the unique interconnecting junctions, and yet be subsequently removed from the final assembly product, allowing for the reconstruction of an intact wild-type sequence. This approach avoids the introduction of nucleotide changes that are normally associated with building a full-length cDNA product of a viral genome.

The use of non-palindromic restriction sites also provides other novel recombinant DNA applications. For example, by PCR, it is be possible to insert Esp3I or a related non-palindromic restriction site at any given nucleotide in a viral genome and use the variable domain for simple and rapid site-specific mutagenesis. By orienting the restriction sites as No See'm, the sites are removed during reassembly, leaving only the desired mutation in the final DNA product. The dual properties of strand specificity and a variable end overhang that can be tailored to match any sequence allow for Esp3I sites to be engineered as universal connectors that can be joined with any other 4-nucleotide restriction site overhang (e.g., EcoRI, PstXI, and BamHI). Alternatively, No See'm sites can be used to insert foreign genes into viral, eukaryotic, or microbial genomes or vectors, simultaneously removing all evidence of the restriction sites that were used in the recombinant DNA manipulation.

In order to remove preexisting Esp3I sites that resided within the MHV-A59 genome sequence, silent mutations were created. This helped to distinguish between molecularly cloned and wild-type viruses. In one instance, the Esp3I site at position 4875 was removed because it left a TTAA overhang that would have prevented the directionality of assembly. The other Esp3I sites were removed to minimize the total number of MHV-A59 subdlones used in the assembly cascade. In two instances, silent mutations were inserted into the Esp3I overhang to maximize sequence specificity and directionality at a particular junction, but this could be circumvented by choosing slightly different junction sites. Clearly, each virus sequence will need to be evaluated for the need for similar changes.

cDNA cassettes can be ligated systematically as previously described for TGEV, or simultaneously as described herein. Although numerous incomplete assembly intermediates occur were evident, the inventor has found that simultaneous ligation of seven cDNAs will result in full-length cDNA, thereby simplifying the complexity of the assembly strategy. There is no evidence to indicate that this approach might introduce spurious mutations or genome rearrangements from aberrant assembly cascades. And while it is possible that such variants might arise following RNA transfection (as a consequence of high-frequency MHV RNA recombination between incomplete and genome-length transcripts), it is highly likely that such variants would be replication impaired and rapidly outcompeted by wild-type virus. A second limitation is that the yield of full-length cDNA product is reduced, resulting in less robust transfection efficiencies than those of the more traditional systematic assembly method. This downside is more than compensated by the reduced complexity in many cases.

V. Vaccines

A. Formulations and Administration

The present invention provides for Nidovirus vaccine formulations. Such compositions will generally be dissolved or dispersed in a pharmaceutically acceptable carrier or aqueous medium. There are numerous examples of vaccine formulations in the literature, and one of skill in the art will be capable of formulating such vaccines.

The phrases "pharmaceutically acceptable" or "pharmacologically acceptable" refer to molecular entities and compo-

sitions that do not produce an adverse, allergic, or other untoward reaction when administered to an animal, or human, as appropriate. As used herein, "pharmaceutically acceptable carrier" includes any and all solvents, dispersion media, coatings, antibacterial and antifungal agents, isotonic and absorption delaying agents, and the like. The use of such media and agents for pharmaceutical active substances is well known in the art.

The vaccines of the present invention can be formulated for parenteral administration, e.g., formulated for injection via the intravenous, intramuscular, subcutaneous, or even intraperitoneal routes. Typically, such compositions can be prepared as injectables, either as liquid solutions or suspensions; solid forms suitable for use to prepare solutions or suspensions upon the addition of a liquid prior to injection can also be prepared; and, the preparations can also be emulsified. The pharmaceutical forms suitable for injectable use include sterile aqueous solutions or dispersions; formulations including sesame oil, peanut oil, or aqueous propylene glycol; and sterile powders for the extemporaneous preparation of sterile injectable solutions or dispersions. In all cases the form must be sterile and must be fluid to the extent that it may be easily injected. It also should be stable under the conditions of manufacture and storage and must be preserved against the contaminating action of microorganisms, such as bacteria and fungi.

The carrier also can be a solvent or dispersion medium containing, for example, water, ethanol, polyol (for example, glycerol, propylene glycol, and liquid polyethylene glycol, and the like), suitable mixtures thereof, and vegetable oils. The proper fluidity can be maintained, for example, by the use of a coating, such as lecithin, by the maintenance of the required particle size in the case of dispersion, and by the use of surfactants. The prevention of the action of microorganisms can be brought about by various antibacterial and antifungal agents, for example, parabens, chlorobutanol, phenol, sorbic acid, thimerosal, and the like. In many cases, it will be preferable to include isotonic agents, for example, sugars or sodium chloride. Prolonged absorption of the injectable compositions can be brought about by the use in the compositions of agents delaying absorption, for example, aluminum monostearate and gelatin.

Generally, dispersions are prepared by incorporating the various sterilized active ingredients into a sterile vehicle which contains the basic dispersion medium and the required other ingredients from those enumerated above. In the case of sterile powders for the preparation of sterile injectable solutions, the preferred methods of preparation are vacuum-drying and freeze-drying techniques. In certain cases, the therapeutic formulations of the invention also may be prepared in forms suitable for oral or intranasal administration.

An effective amount of the vaccine is determined based on the intended goal. The term "unit dose" or "dosage" refers to physically discrete units suitable for use in a subject, each unit containing a predetermined-quantity of the composition calculated to produce the desired responses, discussed above, in association with its administration, i.e., the appropriate route and treatment regimen. Precise amounts of the vaccine composition also depend on the judgment of the practitioner and are peculiar to each individual. Factors affecting dose include physical and clinical state of the patient, the route of administration, the intended goal of treatment (alleviation of symptoms versus cure) and the potency, stability, and toxicity of the particular substance.

The following is a listing of references describing various live vaccines, the relevant contents of which (formulations and administration) are hereby incorporated by reference:

U.S. Pat. Nos. 6,479,056 ; 6,444,445; 6,306,400; 6,296,854; 6,231,871; 6,217,882; 6,159,477; 6,153,199; 6,136,325; 6,077,516; 6,051,237; 6,045,803; 6,039,958; 6,039,941; 6,033,670; 5,993,822; 5,980,906; 5,958,423; 5,948,411; 5,871,742; 5,869,036; 5,792,452; 5,733,555; 5,733,554; 5,651,972; 5,632,989; 5,626,850; 5,580,557; 5,436,001; 5,310,668; 5,149,531; 5,068,104; 5,037,650; 5,024,836; 5,006,335; 4,985,244; 4,980,162; 4,808,404; 4,770,875; 4,762,711; 4,752,474; 4,673,572; 4,645,665; 4,624,850; 4,590,072; 4,555,401; 4,554,158; 4,472,378; 4,456,588; 4,324,861; 4,311,797; 4,235,876; 4,004,974

B. Additional Agents

In addition to the inactive agents discussed above, the vaccine may comprise, or may be given in conjunction with, a supplemental agent. One example is an immunostimulant.

VI. EXAMPLES

The following examples are included to further illustrate various aspects of the invention. It should be appreciated by those of skill in the art that the techniques disclosed in the examples which follow represent techniques and/or compositions discovered by the inventor to function well in the practice of the invention, and thus can be considered to constitute preferred modes for its practice. However, those of skill in the art should, in light of the present disclosure, appreciate that many changes can be made in the specific embodiments which are disclosed and still obtain a like or similar result without departing from the spirit and scope of the invention.

Example 1

Materials and Methods

Viruses and cells. Mouse hepatitis virus lab strain RA59 was the wildtype control in all experiments. Delayed brain tumor (DBT) cells and baby hamster kidney cells expressing the MHV receptor (BHK-MHVR) were grown in Dulbecco's modified Eagle medium (DMEM) containing 10% fetal calf serum (FCS). The BHK-MHVR cells were grown under G418 selection (0.8 mg/ml).

Sequencing MHV cDNA fragments. The infectious clone MHV cDNA fragments (A-G) were sequenced with a set of primers designed from the MHV genome (NCBI accession: NC001846). Sequencing primers were created every 600 bp, beginning with the 5' end, of both the sense and antisense strands. These primers generated overlapping sequences covering the cDNA clones at least twice, and up to four times in certain regions. M13 forward and reverse primers were also used to sequence the insert-vector junction of the pCR-XL-Topo (Invitrogen) or pSMART-LCAmp (Lucigen) parent vectors. The sequencing was performed with an ABIPrism automated sequencer. The sequences were aligned and compared with MHV sequences from the NCBI database: NC001846, AF029248, AF201929, AF207902, AF208066, AF208067, M18379, M55148, X00509, X51939, X57302, and X73559.

Site-directed mutagenesis of fragment F. Site-directed mutations were made in the fragment F plasmid, which consists of genomic nucleotides (nt) 15755 to 22740, at nt 19400 and 22051 using PCR. A 5' BstBI and 3' AflII site or a 5' AflII and 3' SpeI site were used to clone the PCR product with the C19400T correction or C22051T correction, respectively, into the pCR-XL-Topo-F. The resultant F plasmids are (pCR-XL-Topo-F 19400T, 22051C), (pCR-XL-Topo-F 19400C, 2205 1T), and (pCR-XL-Topo-F 19400T, 2205 1T).

Virus assembly. The reverse genetics system for MHV-A59 was used to create viruses with the engineered mutations. Briefly, plasmids A through G containing cDNA cassettes of the MHV genome were digested. The digested fragments were gel purified and ligated together, following which full-length transcripts were generated in vitro using the mMessage mMachine T7 Transcription Kit (Ambion), following the manufacturer's protocol with modifications. Breifly, a 50 μL reaction was supplemented with 7.5 μL of 30 mM GTP and the transcription reaction was performed at 40.5° C. for 25 minutes, 37.5° C. for 50 minutes, and 40.5° C. for 25 minutes. The icMHV transcripts were then combined with N transcripts generated in vitro and 600 μL of BHK-MHVR cells in PBS (10^7 cells/ml) in a 4 mm gap cuvette and three electrical pulses of 850 V at 25 μF were delivered to the mixture with a Bio-Rad Gene Pulser Xcell electroporator. Transfected cells were seeded over a layer of 10^6 uninfected DBT cells in a 75 cm^2 flask and incubated at 37° C. for 30 hours. Virus viability was determined by synctia formation, and progeny virus was passaged and purified by plaque assay.

RT-PCR and sequencing of viral RNA. Plaque isolated virus was used to infect DBT cells in 25 cm^2 flasks (~3×10^6 cells) at a high MOI. With ~80% of the monolayer involved in synctitia, the cells were lysed with TRIzol reagent (Invitrogen) and total RNA was isolated according to the manufacturer's protocol. An antisense primer complimentary to nt 22126 through 22144 of the MHV genome was used to generate viral cDNA from the total RNA using reverse transcription. The RT product was amplified by PCR with the same antisense primer and a sense primer complimentary to nt 19321 through nt 19338. This PCR product was then sequenced using an ABIPrism automated sequencer over nt 19400 and nt 22051 to confirm the mutations present for each virus.

Determination of viral titer from brain of infected mice. Mice were infected with each virus. Five mice per virus were sacrificed 5 d p.i. and 5 mice per virus 7 d p.i. The brains were weighed and placed in 2 ml saline, homogenized, and stored at −70° C. The number of infectious virus in the brain was titered on DBT cells. Plaque isolates of each virus were harvested and sequenced using RT-PCR as described in the previous section. One plaque isolate from two different mice per virus was sequenced over nt 19400 and nt 22051.

Growth experiments. Duplicate DBT cells (10^7) in 75 cm^2 flasks were infected at an MOI of 5 PFU/cell with plaque isolates from the first passage of each virus. The cells were rocked for 30 minutes and then washed three times with PBS (10 minutes per wash). Media was added (10 mL) and the cells were incubated at 37° C. Samples of media (0.5 mL) were collected at 1 h, 4 h, 8 h, 12 h, 16 h, and 25 h p.i., and viral titers were determined by plaque assay.

Example 2

Results

icMHV virulence attenuated in mice. The inventor's lab has previously shown that a wild-type MHV infectious clone virus (icMHV) demonstrated a wild-type phenotype in growth, protein processing, and RNA synthesis assays in DBT cell culture. To determine whether icMHV is virulent in mice, 10^5 pfu of both wild-type A59 and icMHV were injected into mice. 100% of mice inoculated with wild-type lab strain A59 eventually died. Surprisingly, none of the mice inoculated with icMHV died or became ill, suggesting that icMHV was attenuated. It was hypothesized that an unaccounted mutation in the icMHV was responsible for attenuating the virus.

Identification of possible virulence attenuating mutations in icMHV. To identify all possible mutations in the icMHV that could cause this lack of virulence, the inventor sequenced the entire cDNA genome of the infectious clone. The cDNA constructs (A-G) of the MHV infectious clone were sequenced bi-directionally using overlapping sets of primers. This strategy generated sequences covering the genome at least twice, and up to four times, confirming the sequence of the infectious clone. The sequences were initially compared with the published MHV-A59 genome sequence of the C12 mutant virus,.(NCBI accession #NC001846). Compared to this sequence, the inventor identified 17 silent and 5 coding differences in clones A-F, and an additional 11 silent and 6 coding differences in the G clone. The G clone was derived from the pMH54 plasmid, and the coding differences found within the G clone have previously been reported. Furthermore, studies have shown viruses generated by homologous recombination using a pMH54 derivative to be virulent. Therefore our analysis focused on the five coding differences residing in the first ~22.7 kb of the genome: two in fragment B (nt T5304C and A6796T) and three in fragment F (nt T17533G, T19400C, and T22051C). The first two mutations (nt T5304C and A6796T) were previously reported for the infectious clone. The first three coding differences between the icMHV and the C12 mutant virus at nt 5304, 6796, and 17533 coincided with the positions of nucleotide mutations reported for the C12 mutant virus. Upon comparison to other MHV sequences available for both strains A59 and JHM, the inventor found that the nucleotides at the first three positions (nt C5304, T6796, and G17533) were identical to these other sequences. The final two mutations, T19400C and T22051C, were not described in any of the available MHV-A59 and MHV-JHM sequences. The first of these mutations coded a tyr6398 his amino acid change in the Orf1ab polyprotein. The latter mutation was a leu^{106}pro amino acid change in the Orf2a protein. Among group 2 coronaviruses that possess the Orf2a protein, the leucine residue is conserved. The former tyr^{6398}his mutation lies in MHV nsp14 (p59) and the tyrosine is conserved for all coronaviruses. Thus, sequencing of the icMHV cDNA identified two novel candidate mutations at conserved residues that were possibly responsible for attenuating the icMHV virus.

Correction of candidate attenuating mutations in icMHV. To determine whether either, both, or neither of the identified mutations was the cause of the attenuating phenotype in icMHV, the inventor corrected each mutation individually and simultaneously. The nt T19400C and T22051C mutations are both found in the F fragment of the MHV infectious clone. Site-directed PCR mutagenesis was used to individually correct the two mutations, changing them back to wild-type sequence at that nucleotide. Hence, the inventor corrected the amino acid at 6398 in Orf1ab from a histidine to the wild-type tyrosine and the amino acid at 106 in Orf2a protein from a proline to the wild-type leucine. Using the reverse genetics system, the inventor assembled viruses with each possible combination of corrected and uncorrected residues. VUSS 0 was made with the original, uncorrected histidine and proline residues. VUSS 1 corrected the Orf1ab mutation, with the combination of tyrosine at position 6398 and proline at position 106. VUSS 2 corrected the Orf2a protein, with a histidine at position 6398 and leucine at position 106. VUSS 3 corrected both sites with a tyrosine and leucine combination. The

viruses were sequenced following three successive passages in cell culture, confirming the mutations were present and maintained (data not shown).

Uncorrected and corrected viruses exhibit wild-type growth in DBT cells. The assembled viruses grew in DBT cells to titers approximately the same as lab strain MHV-A59 virus. There was no observed difference in plaque morphology. Single-cycle growth curve experiments performed at an MOI of 5 PFU/cell demonstrated that there was no distinguishable difference in the growth kinetics of these viruses in culture. Peak titers >10^6 PFU/ml for each virus were achieved at 12 h p.i. Thus, there was no difference between lab strain MHV-A59 and the corrected and uncorrected viruses, in agreement with our previous results with the original icMHV.

Both Orf1ab and Orf2a mutations attenuate MHV-A59 in mice. The virulence of the four infectious clone viruses were examined alongside lab strain MHV-A59 in mice. The inventors calculated LD_{50} values for each virus. Wild-type A59 and VUSS 3 both had an $\log_{10}$ LD_{50} of 3.8. VUSS 0 and VUSS 2 both had an $\log_{10}$ LD_{50}>5.3. This data indicated that VUSS 0, with both uncorrected histidine and proline mutations, and VUSS 2, with just the histidine mutation in Orf1 ab uncorrected, were completely attenuated in mice. Correcting both mutations to the conserved tyrosine and leucine, as in VUSS 3, restores wild-type virulence. VUSS 1, with just the uncorrected proline mutation in Orf2a, appeared to be partially attenuated. Thus, these results demonstrated that the single tyrosine to histidine mutation in the p59 protein of the Orf1ab polyprotein completely eliminated the virulence of MHV-A59 in mice.

Attenuated MHV viruses have reduced replication in mice at day five following IC inoculation. To determine whether the completely attenuated viruses were replicating in vivo, mice were inoculated intra-cranially with VUSS 0, VUSS 2, VUSS 3, and lab strain MHV-A59. Five mice per virus were sacrificed on day 5 and again on day 7 and their organs were collected. Supernatant from the homogenized brain of each mouse was used in plaque assays to determine virus titer at each day. At day five p.i., the virulent viruses wtA59 and VUSS 3 had viral titers of approximately 10^4 PFU/gram brain tissue, while the attenuated viruses VUSS 0 and VUSS 2 had titers of approximately 10^2 PFU/gram brain tissue. At 7 d p.i., each of the viruses had been cleared from the mice brains. These results suggest that the avirulent viruses VUSS 0 and VUSS 2 were able to replicate in the brains of the mice, although the avirulent viruses had reduced replication compared to the virulent viruses.

All of the compositions and methods disclosed and claimed herein can be made and executed without undue experimentation in light of the present disclosure. While the compositions and methods of this invention have been described in terms of preferred embodiments, it will be apparent to those of skill in the art that variations may be applied to the compositions and methods, and in the steps or in the sequence of steps of the methods described herein without departing from the concept, spirit and scope of the invention. More specifically, it will be apparent that certain agents which are both chemically and physiologically related may be substituted for the agents described herein while the same or similar results would be achieved. All such similar substitutes and modifications apparent to those skilled in the art are deemed to be within the scope of the invention as defined by the appended claims.

VII. REFERENCES

The following references, to the extent that they provide exemplary procedural or other details supplementary to those set forth herein, are specifically incorporated herein by reference:

U.S. Pat. No. 4,004,974
U.S. Pat. No. 4,235,876
U.S. Pat. No. 4,311,797
U.S. Pat. No. 4,324,861
U.S. Pat. No. 4,456,588
U.S. Pat. No. 4,472,378
U.S. Pat. No. 4,554,158
U.S. Pat. No. 4,555,401
U.S. Pat. No. 4,590,072
U.S. Pat. No. 4,624,850
U.S. Pat. No. 4,645,665
U.S. Pat. No. 4,673,572
U.S. Pat. No. 4,752,474
U.S. Pat. No. 4,762,711
U.S. Pat. No. 4,770,875
U.S. Pat. No. 4,808,404
U.S. Pat. No. 4,980,162
U.S. Pat. No. 4,985,244
U.S. Pat. No. 5,006,335
U.S. Pat. No. 5,024,836
U.S. Pat. No. 5,037,650
U.S. Pat. No. 5,068,104
U.S. Pat. No. 5,149,531
U.S. Pat. No. 5,220,007
U.S. Pat. No. 5,221,605
U.S. Pat. No. 5,238,808
U.S. Pat. No. 5,284,760
U.S. Pat. No. 5,310,668
U.S. Pat. No. 5,354,670
U.S. Pat. No. 5,366,878
U.S. Pat. No. 5,380,721
U.S. Pat. No. 5,389,514
U.S. Pat. No. 5,436,001
U.S. Pat. No. 5,580,557
U.S. Pat. No. 5,626,850
U.S. Pat. No. 5,632,989
U.S. Pat. No. 5,635,377
U.S. Pat. No. 5,651,972
U.S. Pat. No. 5,733,554
U.S. Pat. No. 5,733,555
U.S. Pat. No. 5,789,166
U.S. Pat. No. 5,792,452
U.S. Pat. No. 5,798,208
U.S. Pat. No. 5,830,650
U.S. Pat. No. 5,869,036
U.S. Pat. No. 5,871,742
U.S. Pat. No. 5,948,411
U.S. Pat. No. 5,958,423
U.S. Pat. No. 5,980,906
U.S. Pat. No. 5,993,822
U.S. Pat. No. 6,033,670
U.S. Pat. No. 6,039,941
U.S. Pat. No. 6,039,958
U.S. Pat. No. 6,045,803
U.S. Pat. No. 6,051,237
U.S. Pat. No. 6,077,516
U.S. Pat. No. 6,136,325
U.S. Pat. No. 6,153,199
U.S. Pat. No. 6,159,477
U.S. Pat. No. 6,217,882
U.S. Pat. No. 6,231,871

U.S. Pat. No. 6,296,854
U.S. Pat. No. 6,306,400
U.S. Pat. No. 6,444,445
U.S. Pat. No. 6,479,056
Baker et al., *J Virol.,* 67:6056-6063, 1993.
Barbas et al., *Proc. Natl. Acad. Sci. USA,* 91(9):3809-3813, 1994.
Blackburn et al., *J. Lipid. Res.,* 32(12):1911-1918, 1991.
Bonilla et al., *J. Virol.,* 71:900-909, 1997.
Bonilla et al., *Virology,* 209:489-497, 1995.
Boothman et al., *Cancer Res.,* 49(11):2871-2878, 1989.
Borek, *Carcinog. Compr. Surv.,* 10:303-316, 1985.
Boursnell et al., *J. Gen. Virol.,* 68:57-77, 1987.
Braisted and Wells, *Proc. Natl. Acad. Sci. USA,* 93(12):5688-5692, 1996.
Burks et al., *Proc. Natl. Acad. Sci. USA,* 94(2):412-417, 1997.
Caldwell and Joyce, *PCR Methods Appl.,* 2(1):28-33, 1992.
Chouljenko et al., *J. Gen. Virol.,* 82:2927-2933, 2001.
Cooleyetal., *Science,* 239(4844):1121-1128, 1988.
Cunningham and Wells, *Science,* 244(4908):1081-1085, 1989.
Curtis et al., *J. Virol.,* 76:1422-1434, 2002.
de Haan et al., *Virology,* 296:177-189, 2002.
Dong and Baker, *Virology,* 204:541-549, 1994.
Eleouet et al., *Virology,* 206:817-822, 1995.
Fischer et al., *J. Virol.,* 71:5148-5146, 1997.
Hall, *Genetics,* 120(4):887-897, 1988.
Herold et al., *J. Biol. Chem.,* 274:14918-14925, 1999.
Herold et al., *Virology,* 195:680-691, 1993.
Hilton et al., *J. Biol. Chem.,* 271(9):4699-4708, 1996.
Hughes et al., *J. Virol.,* 69:809-813, 1995.
Kanjanahaluethai and Baker, *J. Virol.,* 74:7911-7921, 2000.
Kanjanahaluethai et al., *Adv. Exp. Med. Biol.,* 494:267-273, 2001.
Koetzner et al., *J. Virol.,* 66:1841-1848, 1992.
Koncz et al., *EMBO J.,* 9(5):1337-1346, 1990.
Lambert and Borek, *J. Natl. Cancer Inst.,* 80(18):1492-1497, 1988.
Lavi et al., *Adv. Exp. Med. Biol.,* 440:543-547, 1998.
Lee et al., *Virology,* 180:567-582, 1991.
Leparc-Goffart et al., *J. Virol.,* 72:9628-9636, 1998.
Marks et al., *Symp. Soc. Exp. Biol.,* 45:77-87, 1991.
Masters et al., *J. Virol.,* 68:328-337, 1994.
McCann et al., *Proc. Natl. Acad. Sci. USA,* 72(3):979-983, 1975.
McClintok et al, *Am. J. Physiol.,* 189(3):463-469, 1957.
Oppenheimer et al., *Cell,* 67(3):483-493, 1991.
Phillips et al., *J. Neurovirol.,* 7:421-431, 2001.
Phillips et al., *J. Virol.,* 73:7752-7760, 1999.
Sanchez et al., J. Virol., 73:7607-7618, 1999.
Sarma et al., *J. Neurovirol.,* 8:381-391, 2002.
Schaad et al., *Virology,* 177:634-645, 1990.
Schiller et al., *Virology,* 242:288-302, 1998.
Schmidt et al., *Science,* 238(4829):960-963, 1987.
Sommer et al. *EMBO J.,* 9(3):605-613, 1990.
Tijms et al., *Proc. Natl. Acad. Sci. USA,* 98:1889-1894, 2001.
Warren et al., *Biochemistry,* 35(27):8855-8862, 1996.
Witte et al., *Cancer Res.,* 49(18):5066-5072, 1989.
Wong et al., *J Bacteriol.,* 178(8):2334-2342, 1996.
Yelton et al., *J. Immunol.,* 155(4):1994-2004, 1995.
Yoo and Pei, *Adv. Exp. Med. Biol.,* 494:73-76, 2001.
Yount et al., *J. Virol.,* 76:11065-11078, 2002.
Yount, *J. Virol.,* 74:10600-10611, 2000.
Zeng et al., *Biochemistry,* 35(40):13157-13164, 1996.
Ziebuhr et al., *J. Biol. Chem.,* 276:33220-33232, 2001.

SEQUENCE LISTING

<160> NUMBER OF SEQ ID NOS: 12

<210> SEQ ID NO 1
<211> LENGTH: 521
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic Peptide

<400> SEQUENCE: 1

```
Cys Thr Thr Asn Leu Phe Lys Asp Cys Ser Arg Ser Tyr Val Gly Tyr
  1               5                  10                  15

His Pro Ala His Ala Pro Ser Phe Leu Ala Val Asp Asp Lys Tyr Lys
             20                  25                  30

Val Gly Gly Asp Leu Ala Val Cys Leu Asn Val Ala Asp Ser Ala Val
         35                  40                  45

Thr Tyr Ser Arg Leu Ile Ser Leu Met Gly Phe Lys Leu Asp Leu Thr
     50                  55                  60

Leu Asp Gly Tyr Cys Lys Leu Phe Ile Thr Arg Asp Glu Ala Ile Lys
 65                  70                  75                  80

Arg Val Arg Ala Trp Val Gly Phe Asp Ala Glu Gly Ala His Ala Ile
                 85                  90                  95

Arg Asp Ser Ile Gly Thr Asn Phe Pro Leu Gln Leu Gly Phe Ser Thr
            100                 105                 110

Gly Ile Asp Phe Val Val Glu Ala Thr Gly Met Phe Ala Glu Arg Asp
```

-continued

```
        115                 120                 125

Gly Tyr Val Phe Lys Lys Ala Ala Ala Arg Ala Pro Pro Gly Glu Gln
    130                 135                 140

Gly Lys His Leu Ile Pro Leu Met Ser Arg Gly Gln Lys Trp Asp Val
145                 150                 155                 160

Val Arg Ile Arg Ile Val Gln Met Leu Ser Asp His Leu Ala Asp Leu
                165                 170                 175

Ala Asp Ser Val Val Leu Val Thr Trp Ala Ala Ser Phe Glu Leu Thr
            180                 185                 190

Cys Leu Arg Tyr Phe Ala Lys Val Gly Arg Glu Val Val Cys Ser Val
        195                 200                 205

Cys Thr Lys Arg Ala Thr Cys Phe Asn Ser Arg Thr Gly Tyr Tyr Gly
    210                 215                 220

Cys Trp Arg His Ser Tyr Ser Cys Asp Tyr Leu Tyr Asn Pro Leu Ile
225                 230                 235                 240

Val Asp Ile Gln Gln Trp Gly Tyr Thr Gly Ser Leu Thr Ser Asn His
                245                 250                 255

Asp Pro Ile Cys Ser Val His Lys Gly Ala His Val Ala Ser Ser Asp
            260                 265                 270

Ala Ile Met Thr Arg Cys Leu Ala Val His Asp Cys Phe Cys Lys Ser
        275                 280                 285

Val His Trp Asn Leu Glu Tyr Pro Ile Ile Ser Asn Glu Val Ser Val
    290                 295                 300

Asn Thr Ser Cys Arg Leu Leu Gln Arg Val Met Phe Arg Ala Ala Met
305                 310                 315                 320

Leu Cys Asn Arg Tyr Asp Val Cys Tyr Asp Ile Gly Asn Pro Lys Gly
                325                 330                 335

Leu Ala Cys Val Lys Gly Tyr Asp Phe Lys Phe Tyr Asp Ala Ser Pro
            340                 345                 350

Val Val Lys Ser Val Lys Gln Phe Val Tyr Lys Tyr Glu Ala Asn Lys
        355                 360                 365

Asp Gln Phe Leu Asp Gly Leu Cys Met Phe Trp Asn Cys Asn Val Asp
    370                 375                 380

Lys Tyr Pro Ala Asn Ala Val Val Cys Arg Phe Asp Thr Arg Val Leu
385                 390                 395                 400

Asn Lys Leu Asn Leu Pro Gly Cys Asn Gly Gly Ser Leu Tyr Val Asn
                405                 410                 415

Lys His Ala Phe His Thr Ser Pro Phe Thr Arg Ala Ala Phe Glu Asn
            420                 425                 430

Leu Lys Pro Met Pro Phe Phe Tyr Tyr Ser Asp Thr Pro Cys Val Tyr
        435                 440                 445

Met Glu Gly Met Glu Ser Lys Gln Val Asp Tyr Val Pro Leu Arg Ser
    450                 455                 460

Ala Thr Cys Ile Thr Arg Cys Asn Leu Gly Gly Ala Val Cys Leu Lys
465                 470                 475                 480

His Ala Glu Glu Tyr Arg Glu Tyr Leu Glu Ser Tyr Asn Thr Ala Thr
                485                 490                 495

Thr Ala Gly Phe Thr Phe Trp Val Tyr Lys Thr Phe Asp Phe Tyr Asn
            500                 505                 510

Leu Trp Asn Thr Phe Thr Arg Leu Gln
        515                 520
```

<210> SEQ ID NO 2

-continued

<211> LENGTH: 527
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic Peptide

<400> SEQUENCE: 2

```
Ala Glu Asn Val Thr Gly Leu Phe Lys Asp Cys Ser Lys Ile Ile Thr
  1               5                  10                  15

Gly Leu His Pro Thr Gln Ala Pro Thr His Leu Ser Val Asp Ile Lys
             20                  25                  30

Phe Lys Thr Glu Gly Leu Cys Val Asp Ile Pro Gly Ile Pro Lys Asp
         35                  40                  45

Met Thr Tyr Arg Arg Leu Ile Ser Met Met Gly Phe Lys Met Asn Tyr
     50                  55                  60

Gln Val Asn Gly Tyr Pro Asn Lys Phe Ile Thr Arg Glu Glu Ala Ile
 65                  70                  75                  80

Arg His Val Arg Ala Trp Ile Gly Phe Asp Val Glu Gly Cys His Ala
                 85                  90                  95

Thr Arg Asp Ala Val Gly Thr Asn Leu Pro Leu Gln Leu Gly Phe Ser
            100                 105                 110

Thr Gly Val Asn Leu Val Ala Val Pro Thr Gly Tyr Val Asp Thr Glu
        115                 120                 125

Asn Asn Thr Glu Phe Thr Arg Val Asn Ala Lys Pro Pro Pro Gly Asp
    130                 135                 140

Gln Phe Lys His Leu Ile Pro Leu Met Tyr Lys Gly Leu Pro Trp Asn
145                 150                 155                 160

Val Val Arg Ile Lys Ile Val Gln Met Leu Ser Asp Thr Leu Lys Gly
                165                 170                 175

Leu Ser Asp Arg Val Val Phe Val Leu Trp Ala His Gly Phe Glu Leu
            180                 185                 190

Thr Ser Met Lys Tyr Phe Val Lys Ile Gly Pro Glu Arg Thr Cys Cys
        195                 200                 205

Leu Cys Asp Lys Arg Ala Thr Cys Phe Ser Thr Ser Ser Asp Thr Tyr
    210                 215                 220

Ala Cys Trp Asn His Ser Val Gly Phe Asp Tyr Val Tyr Asn Pro Phe
225                 230                 235                 240

Met Ile Asp Val Gln Gln Trp Gly Phe Thr Gly Asn Leu Gln Ser Asn
                245                 250                 255

His Asp Gln His Cys Gln Val His Gly Asn Ala His Val Ala Ser Cys
            260                 265                 270

Asp Ala Ile Met Thr Arg Cys Leu Ala Val His Glu Cys Phe Val Lys
        275                 280                 285

Arg Val Asp Trp Ser Val Glu Tyr Pro Ile Ile Gly Asp Glu Leu Arg
    290                 295                 300

Val Asn Ser Ala Cys Arg Lys Val Gln His Met Val Val Lys Ser Ala
305                 310                 315                 320

Leu Leu Ala Asp Lys Phe Pro Val Leu His Asp Ile Gly Asn Pro Lys
                325                 330                 335

Ala Ile Lys Cys Val Pro Gln Ala Glu Val Glu Trp Lys Phe Tyr Asp
            340                 345                 350

Ala Gln Pro Cys Ser Asp Lys Ala Tyr Lys Ile Glu Glu Leu Phe Tyr
        355                 360                 365

Ser Tyr Ala Thr His His Asp Lys Phe Thr Asp Gly Val Cys Leu Phe
```

-continued

```
    370                 375                 380

Trp Asn Cys Asn Val Asp Arg Tyr Pro Ala Asn Ala Ile Val Cys Arg
385                 390                 395                 400

Phe Asp Thr Arg Val Leu Ser Asn Leu Asn Leu Pro Gly Cys Asp Gly
                405                 410                 415

Gln Ser Leu Tyr Val Asn Lys His Ala Phe His Thr Pro Ala Phe Asp
            420                 425                 430

Lys Ser Ala Phe Thr Asn Leu Lys Gln Leu Pro Phe Phe Tyr Tyr Ser
        435                 440                 445

Asp Ser Pro Cys Glu Ser His Gly Lys Gln Val Val Ser Asp Ile Asp
    450                 455                 460

Tyr Val Pro Leu Lys Ser Ala Thr Cys Ile Thr Arg Cys Asn Leu Gly
465                 470                 475                 480

Gly Ala Val Cys Arg His His Ala Asn Glu Tyr Arg Gln Tyr Leu Asp
                485                 490                 495

Ala Tyr Asn Met Met Ile Ser Ala Gly Phe Ser Leu Trp Ile Tyr Lys
            500                 505                 510

Gln Phe Asp Thr Tyr Asn Leu Trp Asn Thr Phe Thr Arg Leu Gln
        515                 520                 525


<210> SEQ ID NO 3
<211> LENGTH: 521
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence:  Synthetic
      Peptide

<400> SEQUENCE: 3

Cys Ser Thr Asn Leu Phe Lys Asp Cys Ser Lys Ser Tyr Ser Gly Tyr
  1               5                  10                  15

His Pro Ala His Ala Pro Ser Phe Leu Ala Val Asp Asp Lys Tyr Lys
             20                  25                  30

Ala Thr Gly Asp Leu Ala Val Cys Leu Gly Ile Gly Asp Ser Ala Val
         35                  40                  45

Thr Tyr Ser Arg Leu Ile Ser Leu Met Gly Phe Lys Leu Asp Val Thr
     50                  55                  60

Leu Asp Gly Tyr Cys Lys Leu Phe Ile Thr Lys Glu Glu Ala Val Lys
 65                  70                  75                  80

Arg Val Arg Ala Trp Val Gly Phe Asp Ala Glu Gly Ala His Ala Thr
                 85                  90                  95

Arg Asp Ser Ile Gly Thr Asn Phe Pro Leu Gln Leu Gly Phe Ser Thr
            100                 105                 110

Gly Ile Asp Phe Val Val Glu Ala Thr Gly Leu Phe Ala Asp Arg Asp
        115                 120                 125

Gly Tyr Ser Phe Lys Lys Ala Val Ala Lys Ala Pro Pro Gly Glu Gln
    130                 135                 140

Phe Lys His Leu Ile Pro Leu Met Thr Arg Gly Gln Arg Trp Asp Val
145                 150                 155                 160

Val Arg Pro Arg Ile Val Gln Met Phe Ala Asp His Leu Ile Asp Leu
                165                 170                 175

Ser Asp Cys Val Val Leu Val Thr Trp Ala Ala Asn Phe Glu Leu Thr
            180                 185                 190

Cys Leu Arg Tyr Phe Ala Lys Val Gly Arg Glu Ile Ser Cys Asn Val
        195                 200                 205
```

-continued

```
Cys Thr Lys Arg Ala Thr Ala Tyr Asn Ser Arg Thr Gly Tyr Tyr Gly
    210                 215                 220

Cys Trp Arg His Ser Val Thr Cys Asp Tyr Leu Tyr Asn Pro Leu Ile
225                 230                 235                 240

Val Asp Ile Gln Gln Trp Gly Tyr Ile Gly Ser Leu Ser Ser Asn His
                245                 250                 255

Asp Leu Tyr Cys Ser Val His Lys Gly Ala His Val Ala Ser Ser Asp
            260                 265                 270

Ala Ile Met Thr Arg Cys Leu Ala Val Tyr Asp Cys Phe Cys Asn Asn
        275                 280                 285

Ile Asn Trp Asn Val Glu Tyr Pro Ile Ile Ser Asn Glu Leu Ser Ile
    290                 295                 300

Asn Thr Ser Cys Arg Val Leu Gln Arg Val Met Leu Lys Ala Ala Met
305                 310                 315                 320

Leu Cys Asn Arg Tyr Thr Leu Cys Tyr Asp Ile Gly Asn Pro Lys Ala
                325                 330                 335

Ile Ala Cys Val Lys Asp Phe Asp Phe Lys Phe Tyr Asp Ala Gln Pro
            340                 345                 350

Ile Val Lys Ser Val Lys Thr Leu Leu Tyr Ser Phe Glu Ala His Lys
        355                 360                 365

Asp Ser Phe Lys Asp Gly Leu Cys Met Phe Trp Asn Cys Asn Val Asp
    370                 375                 380

Lys Tyr Pro Pro Asn Ala Val Val Cys Arg Phe Asp Thr Arg Val Leu
385                 390                 395                 400

Asn Asn Leu Asn Leu Pro Gly Cys Asn Gly Gln Ser Leu Tyr Val Asn
                405                 410                 415

Lys His Ala Phe His Thr Lys Pro Phe Ser Arg Ala Ala Phe Glu His
            420                 425                 430

Leu Lys Pro Met Pro Phe Phe Tyr Tyr Ser Asp Thr Pro Cys Val Tyr
        435                 440                 445

Met Asp Gly Met Asp Ala Lys Gln Val Asp Tyr Val Pro Leu Lys Ser
    450                 455                 460

Ala Thr Cys Ile Thr Arg Cys Asn Leu Gly Gly Ala Val Cys Leu Lys
465                 470                 475                 480

His Ala Glu Glu Tyr Arg Glu Tyr Leu Glu Ser Tyr Asn Thr Ala Thr
                485                 490                 495

Thr Ala Gly Phe Thr Phe Trp Val Tyr Lys Thr Phe Asp Phe Tyr Asn
            500                 505                 510

Leu Trp Asn Thr Phe Thr Lys Leu Gln
        515                 520


<210> SEQ ID NO 4
<211> LENGTH: 521
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence:  Synthetic
      Peptide

<400> SEQUENCE: 4

Cys Ser Thr Asn Leu Phe Lys Asp Cys Ser Lys Ser Tyr Ser Gly Tyr
  1               5                  10                  15

His Pro Ala His Ala Pro Ser Phe Leu Ala Val Asp Asp Lys Tyr Lys
             20                  25                  30

Ala Thr Gly Asp Leu Ala Val Cys Leu Gly Ile Gly Asp Ser Ala Val
         35                  40                  45
```

-continued

```
Thr Tyr Ser Arg Leu Ile Ser Leu Met Gly Phe Lys Leu Asp Val Thr
     50                  55                  60

Leu Asp Gly Tyr Cys Lys Leu Phe Ile Thr Lys Glu Glu Ala Val Lys
 65                  70                  75                  80

Arg Val Arg Ala Trp Val Gly Phe Asp Ala Glu Gly Ala His Ala Thr
                 85                  90                  95

Arg Asp Ser Ile Gly Thr Asn Phe Pro Leu Gln Leu Gly Phe Ser Thr
            100                 105                 110

Gly Ile Asp Phe Val Val Glu Ala Thr Gly Leu Phe Ala Asp Arg Asp
        115                 120                 125

Gly Tyr Ser Phe Lys Lys Ala Val Ala Lys Ala Pro Pro Gly Glu Gln
    130                 135                 140

Phe Lys His Leu Ile Pro Leu Met Thr Arg Gly His Arg Trp Asp Val
145                 150                 155                 160

Val Arg Pro Arg Ile Val Gln Met Phe Ala Asp His Leu Ile Asp Leu
                165                 170                 175

Ser Asp Cys Val Val Leu Val Thr Trp Ala Ala Asn Phe Glu Leu Thr
            180                 185                 190

Cys Leu Arg Tyr Phe Ala Lys Val Gln Arg Glu Ile Ser Cys Asn Val
        195                 200                 205

Cys Thr Lys Arg Ala Thr Val Tyr Asn Ser Arg Thr Gly Tyr Tyr Gly
    210                 215                 220

Cys Trp Arg His Ser Val Thr Cys Asp Tyr Leu Tyr Asn Pro Leu Ile
225                 230                 235                 240

Val Asp Ile Gln Gln Trp Gly Tyr Ile Gly Ser Leu Ser Ser Asn His
                245                 250                 255

Asp Leu Tyr Cys Ser Val His Lys Gly Ala His Val Ala Ser Ser Asp
            260                 265                 270

Ala Ile Met Thr Arg Cys Leu Ala Val Tyr Asp Cys Phe Cys Asn Asn
        275                 280                 285

Ile Asn Trp Asn Val Glu Tyr Pro Ile Ile Ser Asn Glu Leu Ser Ile
    290                 295                 300

Asn Thr Ser Cys Arg Val Leu Gln Arg Val Ile Leu Lys Ala Ala Met
305                 310                 315                 320

Leu Cys Asn Arg Tyr Thr Leu Cys Tyr Asp Ile Gly Asn Pro Lys Ala
                325                 330                 335

Ile Ala Cys Val Lys Asp Phe Asp Phe Lys Phe Tyr Asp Ala Gln Pro
            340                 345                 350

Ile Val Lys Ser Val Lys Thr Leu Leu Tyr Ser Phe Glu Ala His Lys
        355                 360                 365

Asp Ser Phe Lys Asp Gly Leu Cys Met Phe Trp Asn Cys Asn Val Asp
    370                 375                 380

Lys Tyr Pro Pro Asn Ala Val Val Cys Arg Phe Asp Thr Arg Val Leu
385                 390                 395                 400

Asn Asn Leu Asn Leu Pro Gly Cys Asn Gly Gly Ser Leu Tyr Val Asn
                405                 410                 415

Lys His Ala Phe His Thr Lys Pro Phe Ala Arg Ala Ala Phe Glu His
            420                 425                 430

Leu Lys Pro Met Pro Phe Phe Tyr Tyr Ser Asp Thr Pro Cys Val Tyr
        435                 440                 445

Met Asp Gly Met Asp Ala Lys Gln Val Asp Tyr Val Pro Leu Lys Ser
    450                 455                 460
```

-continued

```
Ala Thr Cys Ile Thr Lys Cys Asn Leu Gly Gly Ala Val Cys Leu Lys
465                 470                 475                 480

His Ala Glu Glu Tyr Arg Glu Tyr Leu Glu Ser Tyr Asn Thr Ala Thr
                485                 490                 495

Thr Ala Gly Phe Thr Phe Trp Val Tyr Lys Thr Phe Asp Phe Tyr Asn
            500                 505                 510

Leu Trp Asn Thr Phe Thr Lys Leu Gln
        515                 520


<210> SEQ ID NO 5
<211> LENGTH: 521
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence:  Synthetic
      Peptide

<400> SEQUENCE: 5

Gly Thr Gly Leu Phe Lys Ile Cys Asn Lys Glu Phe Ser Gly Val His
  1               5                  10                  15

Pro Ala Tyr Ala Val Thr Thr Lys Ala Leu Ala Ala Thr Tyr Lys Val
             20                  25                  30

Asn Asp Glu Leu Ala Ala Leu Val Asn Val Glu Ala Gly Ser Glu Ile
         35                  40                  45

Thr Tyr Lys His Leu Ile Ser Leu Leu Gly Phe Lys Met Ser Val Asn
     50                  55                  60

Val Glu Gly Cys His Asn Met Phe Ile Thr Arg Asp Glu Ala Ile Arg
 65                  70                  75                  80

Asn Val Arg Gly Trp Val Gly Phe Asp Val Glu Ala Thr His Ala Cys
                 85                  90                  95

Gly Thr Asn Ile Gly Thr Asn Leu Pro Phe Gln Val Gly Phe Ser Thr
            100                 105                 110

Gly Ala Asp Phe Val Val Thr Pro Glu Gly Leu Val Asp Thr Ser Ile
        115                 120                 125

Gly Asn Asn Phe Glu Pro Val Asn Ser Lys Ala Pro Pro Gly Glu Gln
    130                 135                 140

Phe Asn His Leu Arg Val Leu Phe Lys Ser Ala Lys Pro Trp His Val
145                 150                 155                 160

Ile Arg Pro Arg Ile Val Gln Met Leu Ala Asp Asn Leu Cys Asn Val
                165                 170                 175

Ser Asp Cys Val Val Phe Val Thr Trp Cys His Gly Leu Glu Leu Thr
            180                 185                 190

Thr Leu Arg Tyr Phe Val Lys Ile Gly Lys Glu Gln Val Cys Ser Cys
        195                 200                 205

Gly Ser Arg Ala Thr Thr Phe Asn Ser His Thr Gln Ala Tyr Ala Cys
    210                 215                 220

Trp Lys His Cys Leu Gly Phe Asp Phe Val Tyr Asn Pro Leu Leu Val
225                 230                 235                 240

Asp Ile Gln Gln Trp Gly Tyr Ser Gly Asn Leu Gln Phe Asn His Asp
                245                 250                 255

Leu His Cys Asn Val His Gly His Ala His Val Ala Ser Val Asp Ala
            260                 265                 270

Ile Met Thr Arg Cys Leu Ala Ile Asn Asn Ala Phe Cys Gln Asp Val
        275                 280                 285

Asn Trp Asp Leu Thr Tyr Phe His Ile Ala Asn Glu Asp Glu Val Asn
    290                 295                 300
```

-continued

```
Ser Ser Cys Arg Tyr Leu Gln Arg Met Tyr Leu Asn Ala Cys Val Asp
305                 310                 315                 320

Ala Leu Lys Val Asn Val Val Tyr Asp Ile Gly Asn Pro Lys Gly Ile
                325                 330                 335

Lys Cys Val Arg Arg Gly Asp Val Asn Phe Arg Phe Tyr Asp Lys Asn
            340                 345                 350

Pro Ile Val Arg Asn Val Lys Gln Phe Glu Tyr Asp Tyr Asn Gln His
        355                 360                 365

Lys Asp Lys Phe Ala Asp Gly Leu Cys Met Phe Trp Asn Cys Asn Val
    370                 375                 380

Asp Cys Tyr Pro Asp Asn Ser Leu Val Cys Arg Tyr Asp Thr Arg Asn
385                 390                 395                 400

Leu Ser Val Phe Asn Leu Pro Gly Cys Asn Gly Gly Ser Leu Tyr Val
                405                 410                 415

Asn Lys His Ala Phe Tyr Thr Pro Lys Phe Asp Arg Ile Ser Phe Arg
            420                 425                 430

Asn Leu Lys Ala Met Pro Phe Phe Phe Tyr Asp Ser Ser Pro Cys Glu
        435                 440                 445

Thr Ile Gln Val Asp Gly Val Ala Gln Asp Leu Val Ser Leu Ala Thr
    450                 455                 460

Lys Asp Cys Ile Thr Lys Cys Asn Ile Gly Gly Ala Val Cys Lys Lys
465                 470                 475                 480

His Ala Gln Met Tyr Ala Glu Phe Val Thr Ser Tyr Asn Ala Ala Val
                485                 490                 495

Thr Ala Gly Phe Thr Phe Trp Val Thr Asn Lys Leu Asn Pro Tyr Asn
            500                 505                 510

Leu Trp Lys Ser Phe Ser Ala Leu Gln
        515                 520


<210> SEQ ID NO 6
<211> LENGTH: 519
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence:  Synthetic
      Peptide

<400> SEQUENCE: 6

Ala Lys Pro Glu Thr Cys Gly Leu Phe Lys Asp Cys Ser Lys Ser Glu
  1               5                  10                  15

Gln Tyr Ile Pro Pro Ala Tyr Ala Thr Thr Tyr Met Ser Leu Ser Asp
             20                  25                  30

Asn Phe Lys Thr Ser Asp Gly Leu Ala Val Asn Ile Gly Thr Lys Asp
         35                  40                  45

Val Lys Tyr Ala Asn Val Ile Ser Tyr Met Gly Phe Arg Phe Glu Ala
     50                  55                  60

Asn Ile Pro Gly Tyr His Thr Leu Phe Cys Thr Arg Asp Phe Ala Met
 65                  70                  75                  80

Arg Asn Val Arg Ala Trp Leu Gly Phe Asp Val Glu Gly Ala Asn Val
                 85                  90                  95

Cys Gly Asp Asn Val Gly Thr Asn Val Pro Leu Gly Leu Gly Phe Ser
            100                 105                 110

Asn Gly Val Asp Phe Val Val Gln Thr Glu Gly Cys Val Ile Thr Glu
        115                 120                 125

Lys Gly Asn Ser Ile Glu Val Val Lys Ala Arg Ala Pro Pro Gly Glu
```

-continued

```
    130                 135                 140

Gln Phe Ala His Leu Ile Pro Leu Met Arg Lys Gly Gln Pro Trp His
145                 150                 155                 160

Ile Val Arg Arg Arg Ile Val Gln Met Val Cys Asp Tyr Phe Asp Gly
                165                 170                 175

Leu Ser Asp Ile Leu Ile Phe Val Leu Trp Ala Gly Gly Leu Glu Leu
            180                 185                 190

Thr Thr Met Arg Tyr Phe Val Lys Ile Gly Arg Pro Gln Lys Cys Glu
        195                 200                 205

Cys Gly Lys Ser Ala Thr Cys Tyr Ser Ser Ser Gln Ser Val Tyr Ala
    210                 215                 220

Cys Phe Lys His Ala Leu Gly Cys Asp Tyr Leu Tyr Asn Pro Tyr Cys
225                 230                 235                 240

Ile Asp Ile Gln Gln Trp Gly Tyr Thr Gly Ser Leu Ser Met Met His
                245                 250                 255

His Glu Val Cys Asn Ile His Arg Asn Glu His Val Ala Ser Gly Asp
            260                 265                 270

Ala Ile Met Thr Arg Cys Leu Ala Ile His Asp Cys Phe Val Lys Arg
        275                 280                 285

Val Asp Trp Ser Ile Val Tyr Pro Phe Ile Asp Asn Glu Glu Lys Ile
    290                 295                 300

Met Lys Ala Gly Arg Ile Val Gly Ser His Val Met Lys Ala Ala Leu
305                 310                 315                 320

Lys Ile Phe Asn Pro Ala Ala Ile Asn Asp Val Gly Asn Pro Lys Gly
                325                 330                 335

Ile Arg Cys Ala Thr Thr Pro Ile Pro Trp Phe Cys Tyr Asp Arg Asp
            340                 345                 350

Pro Ile Asn Asn Asn Val Arg Cys Leu Asp Tyr Asp Tyr Met Val Met
        355                 360                 365

Gly Gln Met Asn Gly Leu Met Leu Phe Trp Asn Cys Asn Val Asp Met
    370                 375                 380

Tyr Pro Glu Phe Ser Ile Val Cys Arg Phe Asp Thr Arg Thr Arg Ser
385                 390                 395                 400

Lys Leu Ser Leu Glu Gly Cys Asn Gly Gly Ala Leu Tyr Val Asn Asn
                405                 410                 415

His Ala Phe Met Thr Pro Ala Tyr Asp Arg Arg Ala Phe Ala Lys Leu
            420                 425                 430

Lys Pro Met Pro Phe Phe Tyr Tyr Asp Asp Ser Asn Cys Glu Leu Val
        435                 440                 445

Asp Gly Gln Pro Asn Tyr Val Pro Leu Lys Ser Asn Val Cys Ile Thr
    450                 455                 460

Lys Cys Asn Ile Gly Gly Ala Val Cys Lys Lys His Ala Ala Leu Tyr
465                 470                 475                 480

Arg Ala Tyr Val Glu Asp Tyr Asn Ile Phe Met Gln Ala Gly Phe Thr
                485                 490                 495

Ile Trp Cys Pro Gln Asn Phe Asp Thr Tyr Met Leu Trp His Gly Phe
            500                 505                 510

Val Asn Ser Lys Ala Leu Gln
        515


<210> SEQ ID NO 7
<211> LENGTH: 518
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
```

-continued

<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic Peptide

<400> SEQUENCE: 7

```
Ser Glu Ser Ser Cys Gly Leu Phe Lys Asp Cys Ala Arg Asn Pro Ile
  1               5                  10                  15

Asp Leu Pro Pro Ser His Ala Thr Thr Tyr Leu Ser Leu Ser Asp Arg
             20                  25                  30

Phe Lys Thr Ser Gly Asp Leu Ala Val Gln Ile Gly Asn Asn Asn Val
         35                  40                  45

Cys Thr Tyr Glu His Val Ile Ser Tyr Met Gly Phe Arg Phe Asp Val
     50                  55                  60

Ser Met Pro Gly Ser His Ser Leu Phe Cys Thr Arg Asp Phe Ala Met
 65                  70                  75                  80

Arg His Val Arg Gly Trp Leu Gly Met Asp Val Glu Gly Ala His Val
                 85                  90                  95

Thr Cys Asp Asn Val Gly Thr Asn Val Pro Leu Gln Val Gly Phe Ser
            100                 105                 110

Asn Gly Val Asp Phe Val Ala Gln Pro Glu Gly Cys Val Leu Thr Asn
        115                 120                 125

Thr Gly Ser Val Val Lys Pro Val Arg Ala Arg Ala Pro Pro Gly Glu
    130                 135                 140

Gln Phe Thr His Ile Val Pro Leu Leu Arg Lys Gly Gln Pro Trp Ser
145                 150                 155                 160

Val Leu Arg Lys Arg Ile Val Gln Met Ile Ala Asp Phe Leu Ala Gly
                165                 170                 175

Ser Ser Asp Val Leu Val Phe Val Leu Trp Ala Gly Gly Leu Glu Leu
            180                 185                 190

Thr Thr Met Arg Tyr Phe Val Lys Ile Gly Ala Val Lys His Cys Gln
        195                 200                 205

Cys Gly Thr Val Ala Thr Cys Tyr Asn Ser Val Ser Asn Asp Tyr Cys
    210                 215                 220

Cys Phe Lys His Ala Leu Gly Cys Asp Tyr Val Tyr Asn Pro Tyr Val
225                 230                 235                 240

Ile Asp Ile Gln Gln Trp Gly Tyr Val Gly Ser Leu Ser Thr Asn His
                245                 250                 255

His Ala Ile Cys Asn Val His Arg Asn Glu His Val Ala Ser Gly Asp
            260                 265                 270

Ala Ile Met Thr Arg Cys Leu Ala Val Tyr Asp Cys Phe Val Lys Asn
        275                 280                 285

Val Asp Trp Ser Ile Thr Tyr Pro Met Ile Ala Asn Glu Asn Ala Ile
    290                 295                 300

Asn Lys Gly Gly Arg Thr Val Gln Ser His Ile Met Arg Ala Ala Ile
305                 310                 315                 320

Lys Leu Tyr Asn Pro Lys Ala Ile His Asp Ile Gly Asn Pro Lys Gly
                325                 330                 335

Ile Arg Cys Ala Val Thr Asp Ala Lys Trp Tyr Cys Tyr Asp Lys Asn
            340                 345                 350

Pro Ile Asn Ser Asn Val Lys Thr Leu Glu Tyr Asp Tyr Met Thr Asn
        355                 360                 365

Gly Gln Met Asp Gly Leu Cys Leu Phe Trp Asn Cys Asn Val Asp Met
    370                 375                 380

Tyr Pro Glu Phe Ser Ile Val Cys Arg Phe Asp Thr Arg Thr Arg Ser
```

-continued

```
385                 390                 395                 400

Thr Leu Asn Leu Glu Gly Val Asn Gly Gly Ser Leu Tyr Val Asn Asn
                405                 410                 415

His Ala Phe His Thr Pro Ala Tyr Asp Lys Arg Ala Met Ala Lys Leu
            420                 425                 430

Lys Pro Ala Pro Phe Phe Tyr Tyr Asp Asp Gly Ser Cys Glu Val Val
        435                 440                 445

His Asp Gln Val Asn Tyr Val Pro Leu Arg Ala Thr Asn Cys Ile Thr
    450                 455                 460

Lys Cys Asn Ile Gly Gly Ala Val Cys Ser Lys His Ala Asn Leu Tyr
465                 470                 475                 480

Arg Ala Tyr Val Glu Ser Tyr Asn Ile Phe Thr Gln Ala Gly Phe Asn
                485                 490                 495

Ile Trp Val Pro Thr Thr Phe Asp Cys Tyr Asn Leu Trp Gln Thr Phe
            500                 505                 510

Thr Glu Val Asn Leu Gln
        515


<210> SEQ ID NO 8
<211> LENGTH: 517
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence:  Synthetic
      Peptide

<400> SEQUENCE: 8

Ala Asn Glu Gly Cys Gly Leu Phe Lys Asp Cys Ser Arg Gly Asp Asp
  1               5                  10                  15

Leu Leu Pro Pro Ser His Ala Asn Thr Phe Met Ser Leu Ala Asp Asn
             20                  25                  30

Phe Lys Thr Asp Gln Tyr Leu Ala Val Gln Ile Gly Val Asn Gly Pro
         35                  40                  45

Ile Lys Tyr Glu His Val Ile Ser Phe Met Gly Phe Arg Phe Asp Ile
     50                  55                  60

Asn Ile Pro Asn His His Thr Leu Phe Cys Thr Arg Asp Phe Ala Met
 65                  70                  75                  80

Arg Asn Val Arg Gly Trp Leu Gly Phe Asp Val Glu Gly Ala His Val
                85                  90                  95

Val Gly Ser Asn Val Gly Thr Asn Val Pro Leu Gln Leu Gly Phe Ser
            100                 105                 110

Asn Gly Val Asp Phe Val Val Arg Pro Glu Gly Cys Val Val Thr Glu
        115                 120                 125

Ser Gly Asp Tyr Ile Lys Pro Val Arg Ala Arg Ala Pro Pro Gly Glu
    130                 135                 140

Gln Phe Ala His Leu Leu Pro Leu Leu Lys Arg Gly Gln Pro Trp Asp
145                 150                 155                 160

Val Val Arg Lys Arg Ile Val Gln Met Cys Ser Asp Tyr Leu Ala Asn
                165                 170                 175

Leu Ser Asp Ile Leu Ile Phe Val Leu Trp Ala Gly Gly Leu Glu Leu
            180                 185                 190

Thr Thr Met Arg Tyr Phe Val Lys Ile Gly Pro Ser Lys Ser Cys Asp
        195                 200                 205

Cys Gly Lys Val Ala Thr Cys Tyr Asn Ser Ala Leu His Thr Tyr Cys
    210                 215                 220
```

-continued

```
Cys Phe Lys Met Ala Leu Gly Cys Asp Tyr Leu Tyr Asn Pro Tyr Cys
225                 230                 235                 240

Ile Asp Ile Gln Gln Trp Gly Tyr Lys Gly Ser Leu Ser Leu Asn His
                245                 250                 255

His Glu His Cys Asn Val His Arg Asn Glu His Val Ala Ser Gly Asp
            260                 265                 270

Ala Ile Met Thr Arg Cys Leu Ala Ile His Asp Cys Phe Val Lys Asn
        275                 280                 285

Val Asp Trp Ser Ile Thr Tyr Pro Phe Ile Gly Asn Glu Ala Val Ile
    290                 295                 300

Asn Lys Ser Gly Arg Ile Val Gln Ser His Thr Met Arg Ser Val Leu
305                 310                 315                 320

Lys Leu Tyr Asn Pro Lys Ala Ile Tyr Asp Ile Cys Asn Pro Lys Gly
                325                 330                 335

Ile Arg Cys Ala Val Thr Asp Ala Lys Trp Phe Cys Phe Asp Lys Asn
            340                 345                 350

Pro Thr Asn Ser Asn Val Lys Thr Leu Glu Tyr Asp Tyr Ile Thr His
        355                 360                 365

Gly Gln Phe Asp Gly Leu Cys Leu Phe Trp Asn Cys Asn Val Asp Met
    370                 375                 380

Tyr Pro Glu Phe Ser Val Val Cys Arg Phe Asp Thr Arg Cys Arg Ser
385                 390                 395                 400

Pro Leu Asn Leu Glu Gly Cys Asn Gly Gly Ser Leu Tyr Val Lys Asn
                405                 410                 415

His Ala Phe His Thr Pro Ala Phe Asp Lys Arg Ala Phe Ala Lys Leu
            420                 425                 430

Lys Pro Met Pro Phe Phe Phe Tyr Asp Asp Thr Glu Cys Asp Lys Leu
        435                 440                 445

Gln Asp Ser Ile Asn Tyr Val Pro Leu Arg Ala Ser Asn Cys Ile Thr
    450                 455                 460

Lys Cys Asn Val Gly Gly Ala Val Cys Ser Lys His Cys Ala Met Tyr
465                 470                 475                 480

His Ser Tyr Val Asn Ala Tyr Asn Thr Phe Thr Ser Ala Gly Phe Thr
                485                 490                 495

Ile Trp Val Pro Thr Ser Phe Asp Thr Tyr Asn Leu Trp Gln Thr Phe
            500                 505                 510

Ser Asn Asn Leu Gln
        515


<210> SEQ ID NO 9
<211> LENGTH: 515
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence:  Synthetic
      Peptide

<400> SEQUENCE: 9

Val Cys Gly Leu Phe Lys Asn Cys Thr Arg Thr Pro Leu Asn Leu Pro
  1               5                  10                  15

Pro Thr His Ala His Thr Phe Leu Ser Leu Ser Asp Gln Phe Lys Thr
             20                  25                  30

Thr Gly Asp Leu Ala Val Gln Ile Gly Ser Asn Asn Val Cys Thr Tyr
         35                  40                  45

Glu Met Val Ile Ser Phe Met Gly Phe Arg Phe Asp Ile Ser Ile Pro
     50                  55                  60
```

-continued

```
Gly Ser His Ser Leu Phe Cys Thr Arg Asp Phe Ala Ile Arg Asn Val
 65                  70                  75                  80

Arg Gly Trp Leu Gly Met Asp Val Glu Ser Ala His Val Cys Gly Asp
                 85                  90                  95

Asn Ile Gly Thr Asn Val Pro Leu Gln Val Gly Phe Ser Asn Gly Val
            100                 105                 110

Asn Phe Val Val Gln Thr Glu Gly Cys Val Ser Thr Asn Phe Gly Asp
        115                 120                 125

Val Ile Lys Pro Val Cys Ala Lys Ser Pro Pro Gly Glu Gln Phe Arg
    130                 135                 140

His Leu Ile Pro Leu Leu Arg Lys Gly Gln Pro Trp Leu Ile Val Arg
145                 150                 155                 160

Arg Arg Ile Val Gln Met Ile Ser Asp Tyr Leu Ser Asn Leu Ser Asp
                165                 170                 175

Ile Leu Val Phe Val Leu Trp Ala Gly Ser Leu Glu Leu Thr Thr Met
            180                 185                 190

Arg Tyr Phe Val Lys Ile Gly Pro Ile Lys Tyr Cys Tyr Cys Gly Asn
        195                 200                 205

Phe Ala Thr Cys Tyr Asn Ser Val Ser Asn Glu Tyr Cys Cys Phe Lys
    210                 215                 220

His Ala Leu Gly Cys Asp Tyr Val Tyr Asn Pro Tyr Ala Phe Asp Ile
225                 230                 235                 240

Gln Gln Trp Gly Tyr Val Gly Ser Leu Ser Gln Asn His His Thr Phe
                245                 250                 255

Cys Asn Ile His Arg Asn Glu His Asp Ala Ser Gly Asp Ala Val Met
            260                 265                 270

Thr Arg Cys Leu Ala Val His Asp Cys Phe Val Lys Asn Val Asp Trp
        275                 280                 285

Thr Val Thr Tyr Pro Phe Ile Ala Asn Glu Lys Phe Ile Asn Gly Cys
    290                 295                 300

Gly Arg Asn Val Gln Gly His Val Val Arg Ala Ala Leu Lys Leu Tyr
305                 310                 315                 320

Lys Pro Ser Val Ile His Asp Ile Gly Asn Pro Lys Gly Val Arg Cys
                325                 330                 335

Ala Val Thr Asp Ala Lys Trp Tyr Cys Tyr Asp Lys Gln Pro Val Asn
            340                 345                 350

Ser Asn Val Lys Leu Leu Asp Tyr Asp Tyr Ala Thr His Gly Gln Leu
        355                 360                 365

Asp Gly Leu Cys Leu Phe Trp Asn Cys Asn Tyr Asp Met Tyr Phe Glu
    370                 375                 380

Phe Ser Ile Val Cys Arg Phe Asp Thr Arg Thr Arg Ser Val Phe Asn
385                 390                 395                 400

Leu Glu Gly Val Asn Gly Gly Ser Leu Tyr Val Asn Lys His Ala Phe
                405                 410                 415

His Thr Pro Ala Tyr Asp Lys Arg Ala Phe Val Lys Leu Lys Pro Met
            420                 425                 430

Pro Phe Phe Tyr Phe Asp Asp Ser Asp Cys Asp Val Val Gln Glu Gln
        435                 440                 445

Val Asn Tyr Val Pro Leu Arg Ala Ser Ser Cys Val Thr Arg Cys Asn
    450                 455                 460

Ile Gly Gly Ala Val Cys Ser Lys His Ala Asn Leu Tyr Gln Lys Tyr
465                 470                 475                 480
```

-continued

```
Val Glu Ala Tyr Asn Thr Phe Thr Gln Ala Gly Phe Asn Ile Trp Val
                485                 490                 495

Pro His Ser Phe Asp Val Tyr Asn Leu Trp Gln Ile Phe Ile Glu Thr
            500                 505                 510

Asn Leu Gln
        515
```

<210> SEQ ID NO 10
<211> LENGTH: 196
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic Peptide

<400> SEQUENCE: 10

```
Trp Ile Gly Phe Asp Val Glu Gly Cys His Ala Thr Arg Asp Ala Val
  1               5                  10                  15

Gly Thr Asn Leu Pro Leu Gln Leu Gly Phe Ser Thr Gly Val Asn Leu
             20                  25                  30

Val Ala Val Pro Thr Gly Tyr Val Asp Thr Glu Asn Asn Thr Glu Phe
         35                  40                  45

Thr Arg Val Asn Ala Lys Pro Pro Pro Gly Asp Gln Phe Lys His Leu
     50                  55                  60

Ile Pro Leu Met Tyr Lys Gly Leu Pro Trp Asn Val Val Arg Ile Lys
 65                  70                  75                  80

Ile Val Gln Met Leu Ser Asp Thr Leu Lys Gly Leu Ser Asp Arg Val
                 85                  90                  95

Val Phe Val Leu Trp Ala His Gly Phe Glu Leu Thr Ser Met Lys Tyr
            100                 105                 110

Phe Val Lys Ile Gly Pro Glu Arg Thr Cys Cys Leu Cys Asp Lys Arg
        115                 120                 125

Ala Thr Cys Phe Ser Thr Ser Ser Asp Thr Tyr Ala Cys Trp Asn His
    130                 135                 140

Ser Val Gly Phe Asp Tyr Val Tyr Asn Pro Phe Met Ile Asp Val Gln
145                 150                 155                 160

Gln Trp Gly Phe Thr Gly Asn Leu Gln Ser Asn His Asp Gln His Cys
                165                 170                 175

Gln Val His Gly Asn Ala His Val Ala Ser Cys Asp Ala Ile Met Thr
            180                 185                 190

Arg Cys Leu Val
        195
```

<210> SEQ ID NO 11
<211> LENGTH: 64
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic Peptide

<400> SEQUENCE: 11

```
Ala Phe Glu His Ile Val Tyr Gly Asp Phe Ser His Gly Gln Leu Gly
  1               5                  10                  15

Gly Leu His Leu Met Ile Gly Leu Ala Lys Arg Ser Gln Asp Ser Pro
             20                  25                  30

Leu Lys Leu Glu Asp Phe Ile Arg Met Asp Ser Thr Val Lys Asn Tyr
         35                  40                  45
```

-continued

```
Phe Ile Thr Asp Ala Gln Thr Gly Ser Ser Lys Cys Val Cys Ser Val
     50                  55                  60


<210> SEQ ID NO 12
<211> LENGTH: 168
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence:  Synthetic
      Peptide

<400> SEQUENCE: 12

Asn Val Ala Lys Tyr Thr Gln Leu Cys Gln Tyr Leu Asn Thr Leu Thr
  1               5                  10                  15

Leu Ala Val Pro Tyr Asn Met Arg Val Ile His Phe Gly Ala Gly Ser
             20                  25                  30

Asp Lys Gly Val Ala Pro Gly Thr Ala Val Leu Arg Gln Trp Leu Pro
         35                  40                  45

Thr Gly Thr Leu Leu Val Asp Ser Asp Leu Asn Asp Phe Val Ser Asp
     50                  55                  60

Ala Asp Ser Thr Leu Ile Gly Asp Cys Ala Thr Val His Thr Ala Asn
 65                  70                  75                  80

Lys Trp Asp Leu Ile Ile Ser Asp Met Tyr Asp Pro Arg Thr Lys His
                 85                  90                  95

Val Thr Lys Glu Asn Asp Ser Lys Glu Gly Phe Phe Thr Tyr Leu Cys
            100                 105                 110

Gly Phe Ile Lys Gln Lys Leu Ala Leu Gly Gly Ser Ile Ala Val Lys
        115                 120                 125

Ile Thr Glu His Ser Trp Asn Ala Asp Leu Tyr Lys Leu Met Gly His
    130                 135                 140

Phe Ser Trp Trp Thr Ala Phe Val Thr Asn Val Asn Ala Ser Ser Ser
145                 150                 155                 160

Glu Ala Phe Leu Ile Gly Met Thr
                165
```

What is claimed is:

1. A live, attenuated virus groups 2 coronavirus wherein said virus is characterized as comprising a genome encoding (i) an ExoN comprising a substitution at tyrosine^{6398} of MHV-A59, or an analogous position thereof, an (ii) an Orf2a polypeptide comprising a substitution at leu^{106} of MHV-A59, or an analogous position thereof.

2. The virus of claim **1**, wherein said coronavirus is bovine coronavirus, human coronavirus OC43, murine hepatitis virus, porcine hemagglutinating encephalomyelitis virus, rat coronavirus, or severe acute respiratory syndrome virus.

3. The virus of claim **1**, wherein said virus further comprises a mutation in least one polyprotein proteinase cleavage site that exhibits reduced as compared to wild-type or no cleavage.

4. The virus of claim **3**, wherein the cleavage site is a C1-C 14 cleavage site.

5. The virus of claim **3**, wherein the cleavage site is a MHV p28-p65 or p65-p210 cleavage site or analogous position thereof.

6. The virus of claim **3**, wherein the cleavage site comprises an amino acid deletion, an amino acid insertion or an amino acid substitution.

7. The virus of claim **1**, wherein said tyrosine^{6398} substitution is a non-conservative substitution.

8. The virus of claim **1**, wherein said tyrosine^{6398} substitution is a histidine.

9. The virus of claim **1**, wherein said leu^{106} substitution is a non-conservative substitution.

10. The virus of claim **1**, wherein said leu^{106} substitution is a proline.

11. The virus of claim **1**, wherein said virus genome further encodes a mutation in one or more of nsp1, nsp2, nsp3, nsp4, nsp5, nsp6, nsp7, nsp8, nsp9, nsp10, nsp11, nsp12, nsp13, nsp15 or nsp16 coding region.

12. A method of inducing an anti-viral immune response in a host comprising administering to said host a live, attenuated group 2 coronavirus wherein said virus is characterized as comprising a genome encoding (i) an ExoN comprising a substitution at tyrosine^{6398} of MHV-A59, or an analogous position thereof, and (ii) an Orf2a polypeptide comprising a substitution at leu^{106} of MHV-A59, or an analogous position thereof.

13. The method of claim **12**, wherein the coronavirus is bovine coronavirus, human coronavirus OC43, murine hepatitis virus, porcine hemagglutinating encephalomyelitis virus, rat coronavirus, severe acute respiratory syndrome virus, human coronavirus NL63 or human coronavirus NL.

14. The method of claim 12, wherein said virus further comprises a mutation in least one polyprotein proteinase cleavage site that exhibits wild-type or reduced cleavage as opposed no cleavage.

15. The method of claim 14, wherein the cleavage site is a C1-C14 cleavage site.

16. The method of claim 14, wherein the cleavage site is a MHV p28-p65 or p65-p210 cleavage site or analogous position thereof.

17. The method of claim 14, wherein the cleavage site contains an amino acid deletion, an amino acid insertion or an amino acid substitution.

18. The method of claim 12, wherein said tyrosine^{6398} substitution is a non-conservative substitution.

19. The method of claim 12, wherein said tyrosine^{6398} substitution is a histidine.

20. The method of claim 12, wherein said leu106 substitution is a non-conservative substitution.

21. The method of claim 12, wherein said leu^{106} substitution is a proline.

22. The method of claim 12, wherein said virus genome further encodes a mutation in one or more of nsp1, nsp2, nsp3, nsp4, nsp5, nsp6, nsp7, nsp8, nsp9, nsp10, nsp11, nsp12, nsp13, nsp15 or nsp16 coding region.

23. The method of claim 12, wherein said vaccine is administered intravenously or subcutaneously.

24. The method of claim 12, further comprising administering an immunostimulant.

25. The method of claim 12, wherein said host is a dog, a cow, a pig, a mouse, a rat, or a human.

26. A coronavirus genome, said genome encoding an ExoN polypeptide comprising a substitution at tyrosine6398 of MHV-A59, or an analogous position thereof.

27. An isolated and purified nucleic acid segment encoding coronavirus ExoN polypeptide comprising comprising a substitution at tyrosine^{6398} of MHV-A59, or an analogous position thereof.

28. A composition comprising (a) a live, attentuated group 2 coronavirus, said virus characterized as comprising a genome encoding (i) an ExoN polypeptide comprising comprising a substitution at tyrosine^{6398} of MHV-A59, or an analogous position thereof, and (ii) an Orf2a polypeptide comprising a substitution at leu^{106} of MHV-A59, or an analogous position thereof, and (b) a pharmaceutically acceptable diluent.

29. The composition of claim 28, wherein said composition is formulated as a unit dose of 10^6 to 10^{14} infectious particles.

30. The composition of claim 28, wherein said unit dose is provided in a 100 ml aliquot.

31. The composition of claim 28, further comprising a preservative.

32. The composition of claim 28, wherein said composition is lyophilized.

* * * * *

UNITED STATES PATENT AND TRADEMARK OFFICE

CERTIFICATE OF CORRECTION

PATENT NO. : 7,452,542 B2
APPLICATION NO. : 11/135603
DATED : November 18, 2008
INVENTOR(S) : Mark Denison

It is certified that error appears in the above-identified patent and that said Letters Patent is hereby corrected as shown below:

In claim 1, column 53, line 44, delete "groups" and insert --group-- therefor.

In claim 1, column 53, line 44, insert comma after "coronavirus".

In claim 1, column 53, line 47, delete "an (ii)" and insert --and (ii)-- therefor.

In claim 12, column 54, lines 55-56, delete "attentuated" and insert --attenuated-- therefor.

In claim 20, column 55, line 17, delete "leu106" and insert --leu^{106}-- therefor.

In claim 26, column 56, line 4, delete "tyrosine6398" and insert --tyrosine^{6398}-- therefor.

In claim 27, column 56, line 7, delete "comprising comprising" and insert --comprising-- therefor.

In claim 28, column 56, line 10, delete "attentuated" and insert --attenuated-- therefor.

In claim 28, column 56, lines 12-13, delete "comprising comprising" and insert --comprising-- therefor.

Signed and Sealed this

Tenth Day of February, 2009

John Doll

JOHN DOLL
Acting Director of the United States Patent and Trademark Office

US 20040229219A1

(19) United States

(12) Patent Application Publication (10) Pub. No.: US 2004/0229219 A1

Gallaher et al. (43) Pub. Date: Nov. 18, 2004

(54) METHOD OF INHIBITING HUMAN METAPNEUMOVIRUS AND HUMAN CORONAVIRUS IN THE PREVENTION AND TREATMENT OF SEVERE ACUTE RESPIRATORY SYNDROME (SARS)

(76) Inventors: **William R. Gallaher**, Pearl River, LA (US); **Robert F. Garry**, New Orleans, LA (US)

Correspondence Address:
Mark E. Mahaffey
8555 United Plaza Blvd., 5th Floor
Baton Rouge, LA 70809 (US)

(21) Appl. No.: **10/834,666**

(22) Filed: **Apr. 29, 2004**

Related U.S. Application Data

(60) Provisional application No. 60/466,978, filed on Apr. 30, 2003.

Publication Classification

(51) **Int. Cl.**[7] **C07K 14/165**; C12Q 1/70

(52) **U.S. Cl.** **435/5**; 530/395

(57) **ABSTRACT**

The present invention relates to peptides that show significant antiviral activity against viral respiratory disease. More particularly, the invention relates to the use of peptides to inhibit membrane fusion and infection by human metapneumovirus and/or human coronavirus in the prevention and treatment of Severe Acute Respiratory Syndrome (SARS) or other severe respiratory diseases caused by theses agents. The peptides are derived from the known amino acid sequence of the fusion glycoproteins of each virus.

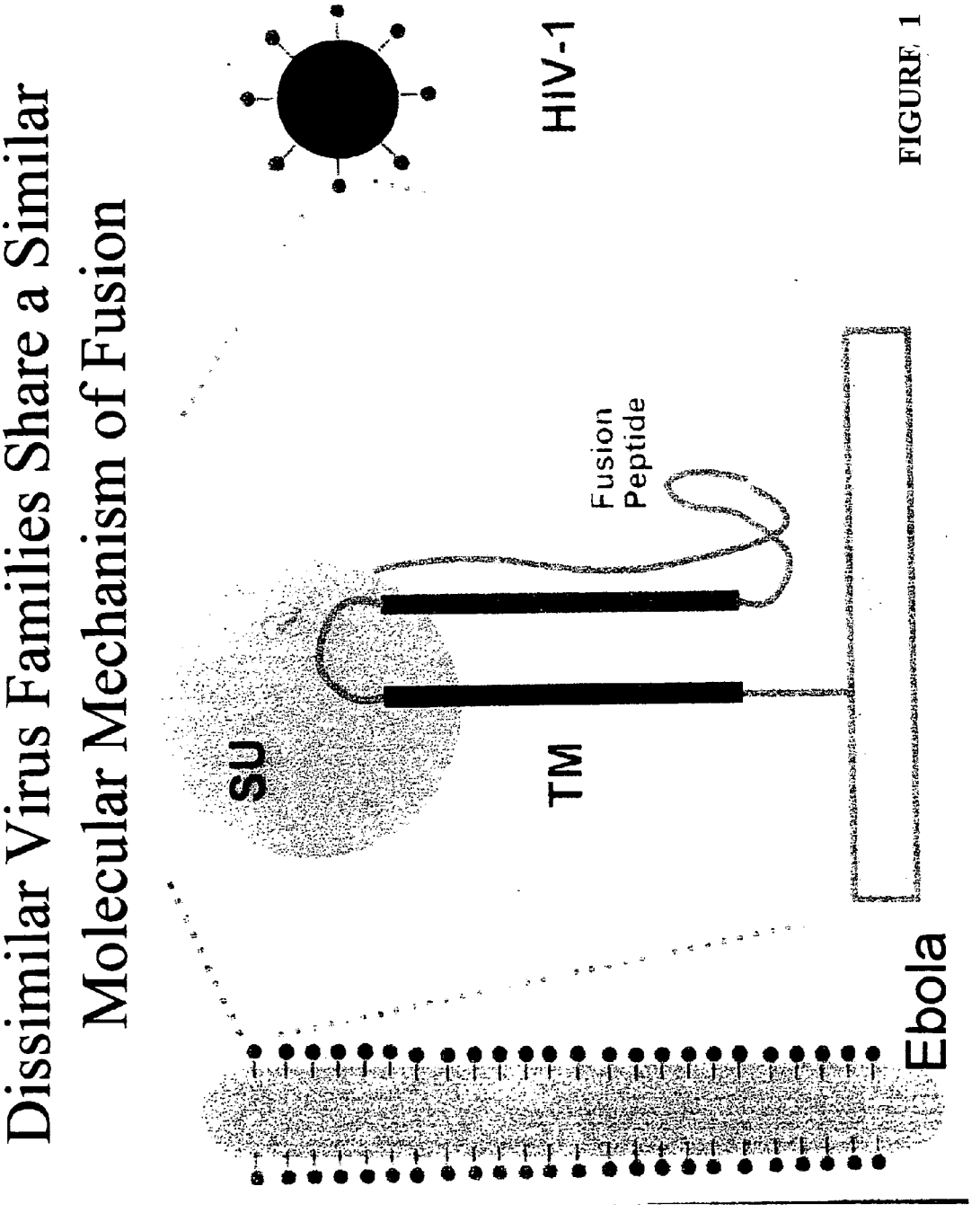

Dissimilar Virus Families Share a Similar Molecular Mechanism of Fusion
HIV-1
Fusion Peptide
SU
TM
Ebola

FIGURE 1

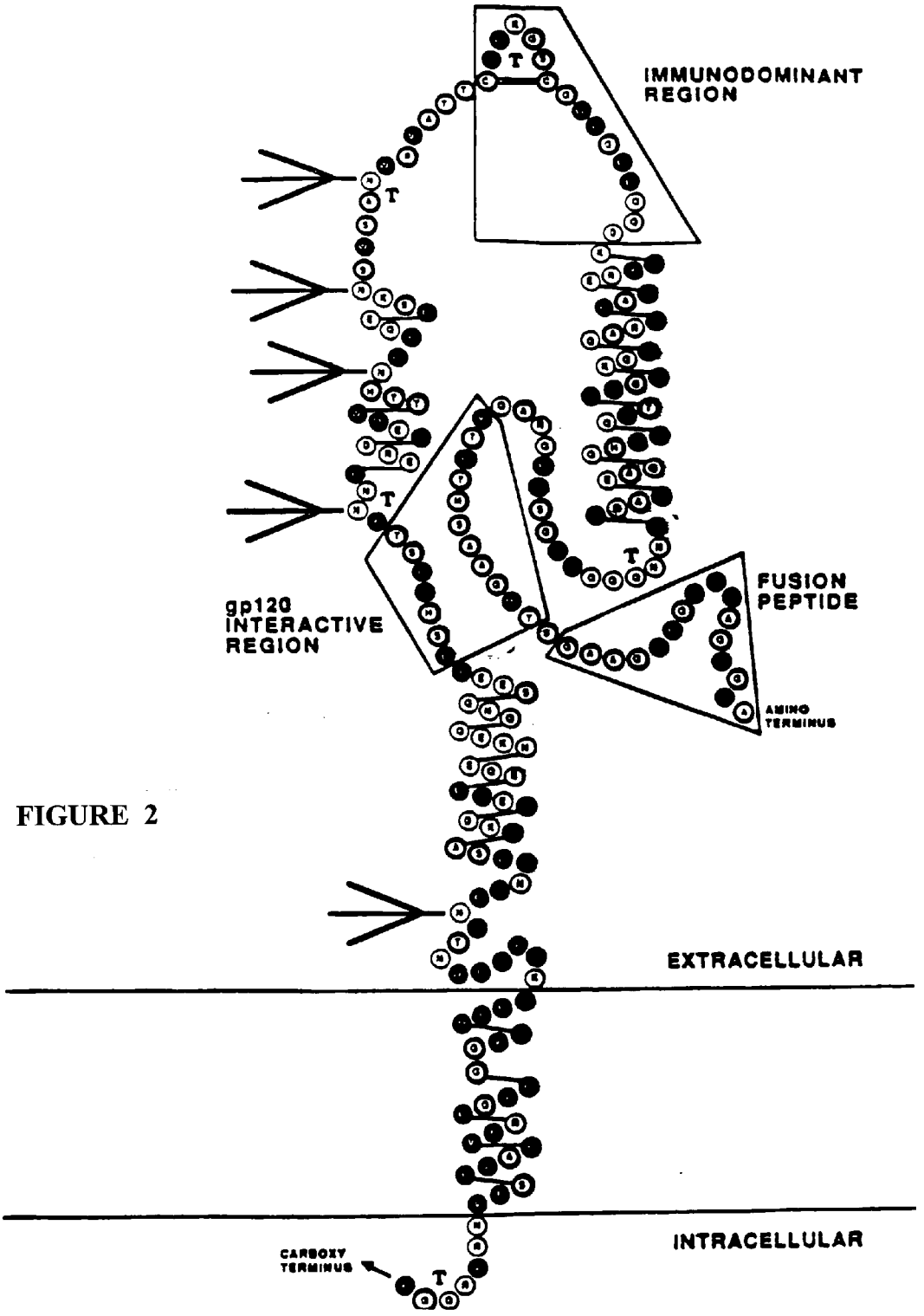

IMMUNODOMINANT REGION
gp120 INTERACTIVE REGION
FUSION PEPTIDE
AMINO TERMINUS
EXTRACELLULAR
INTRACELLULAR
CARBOXY TERMINUS

FIGURE 2

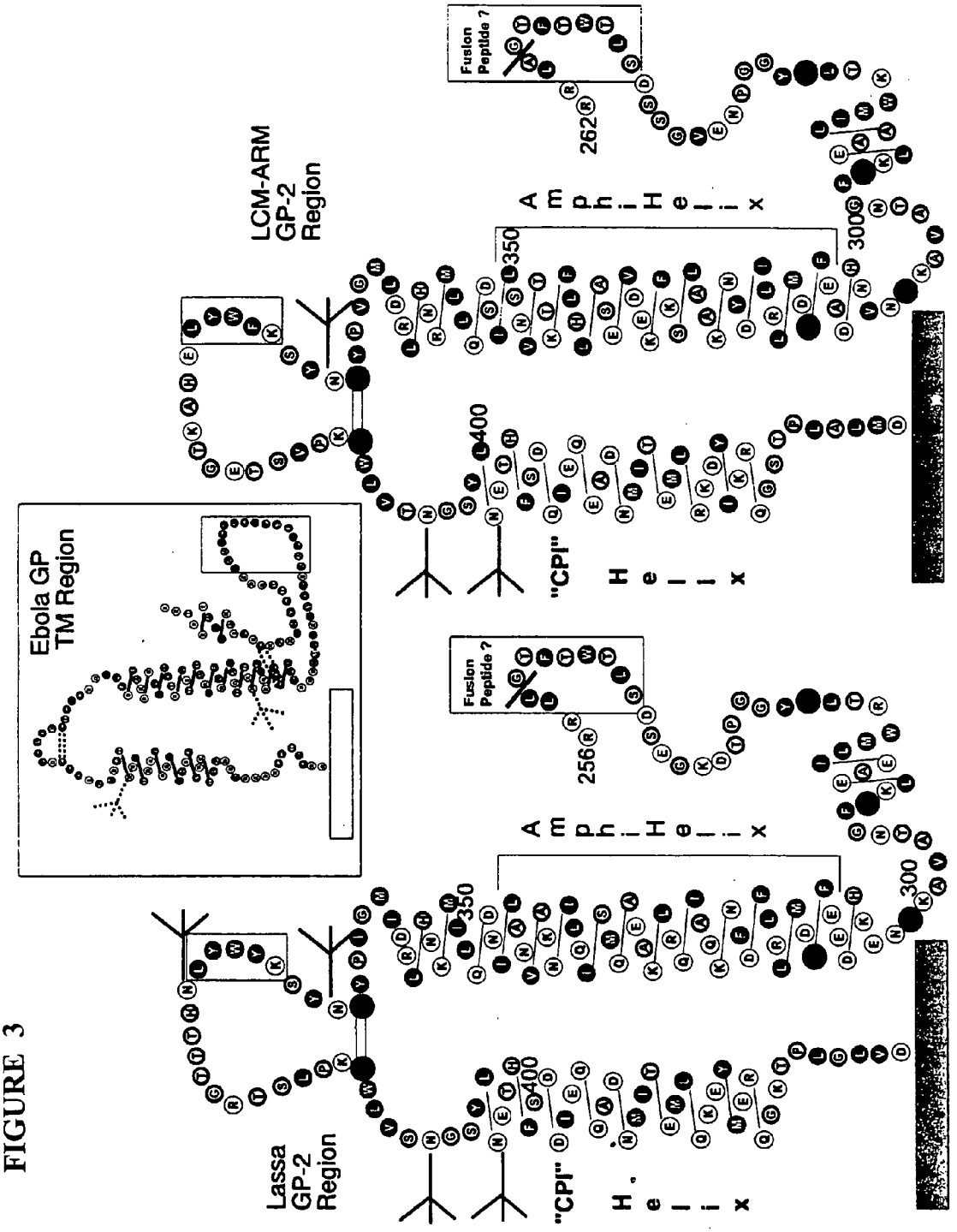
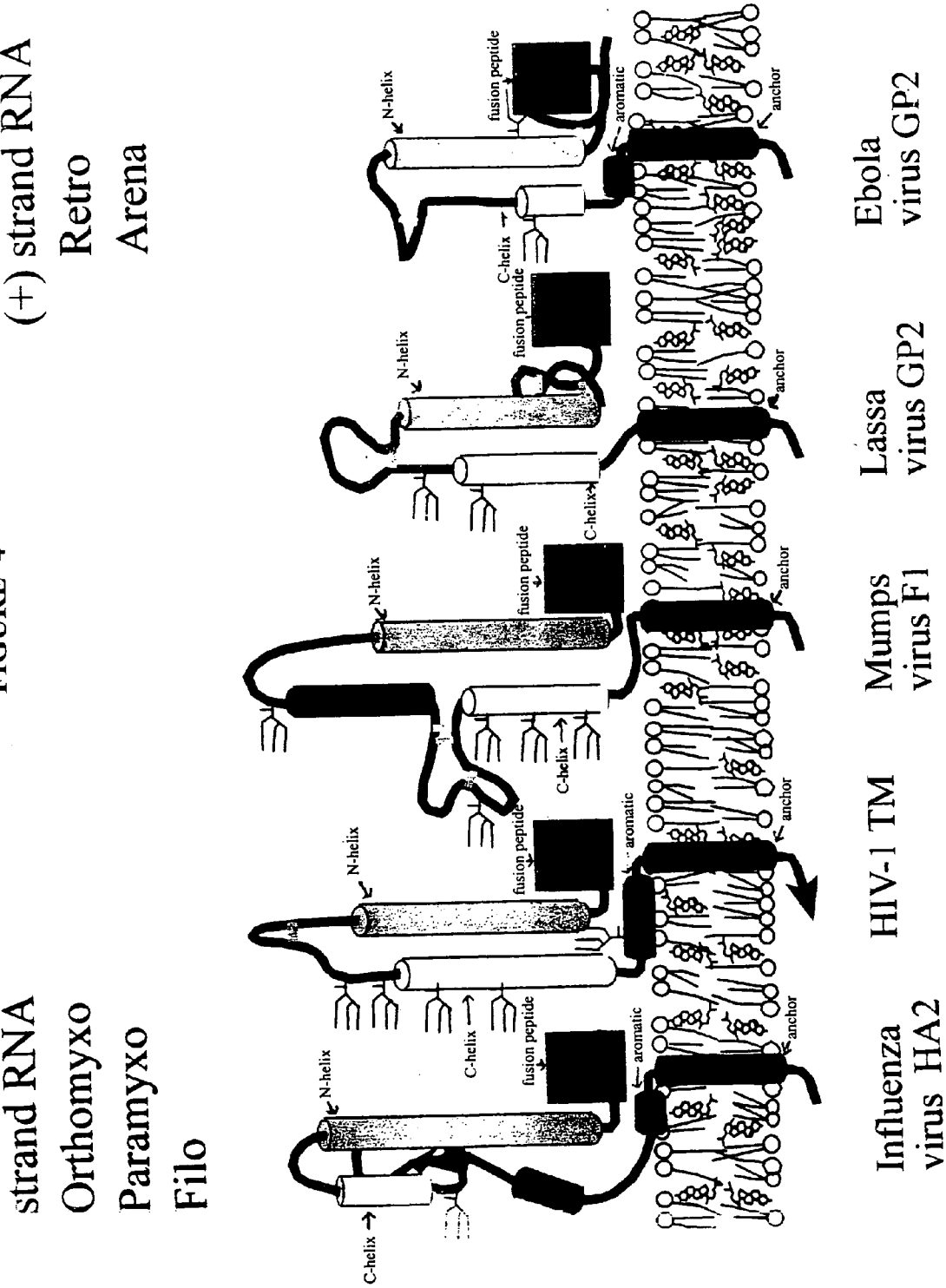
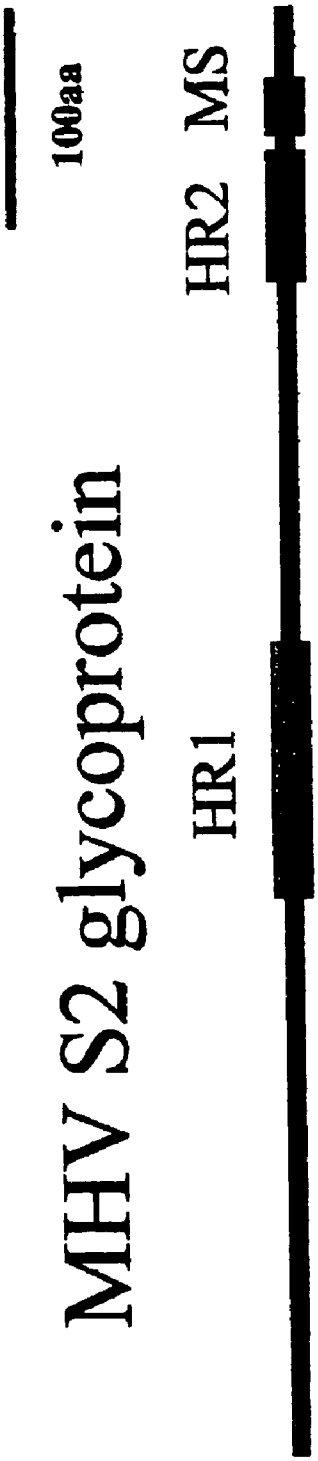

LCM-ARM
GP-2
Region
Fusion Peptide ?
262
Amph-Helix
300
350
400
"CPI"
Helix
Ebola GP
TM Region
Fusion Peptide ?
256
Amph-Helix
300
350
400
"CPI"
Helix
Lassa
GP-2
Region

FIGURE 3

FIGURE 4


(+) strand RNA
Retro
Arena
(-) strand RNA
Orthomyxo
Paramyxo
Filo
C-helix
N-helix
fusion peptide
aromatic
anchor
Influenza virus HA2
HIV-1 TM
Mumps virus F1
Lassa virus GP2
Ebola virus GP2



Modified from:
Luo, Matthews,and Weiss
1999J.Virol. 73:8152.

FIGURE 5

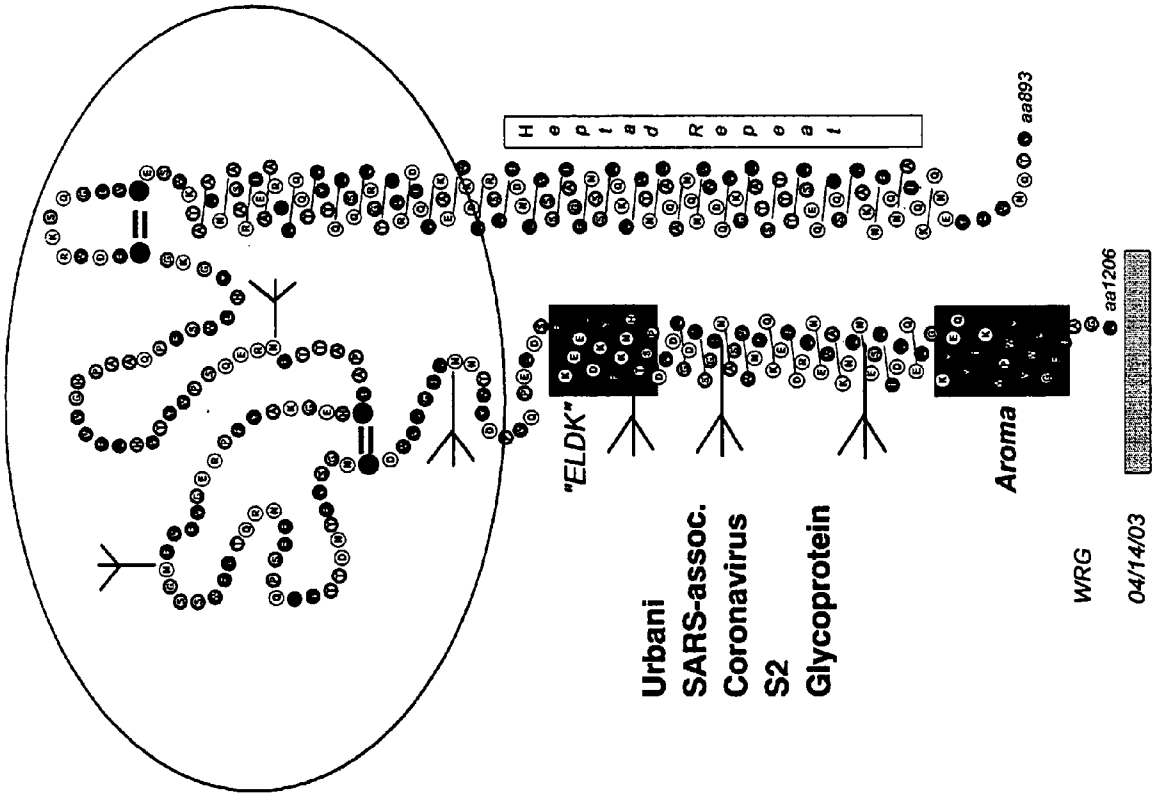

Heptad Repeat
aa893
"ELDK"
Aroma
aa1206
Urbani
SARS-assoc.
Coronavirus
S2
Glycoprotein
WRG
04/14/03

FIGURE 6

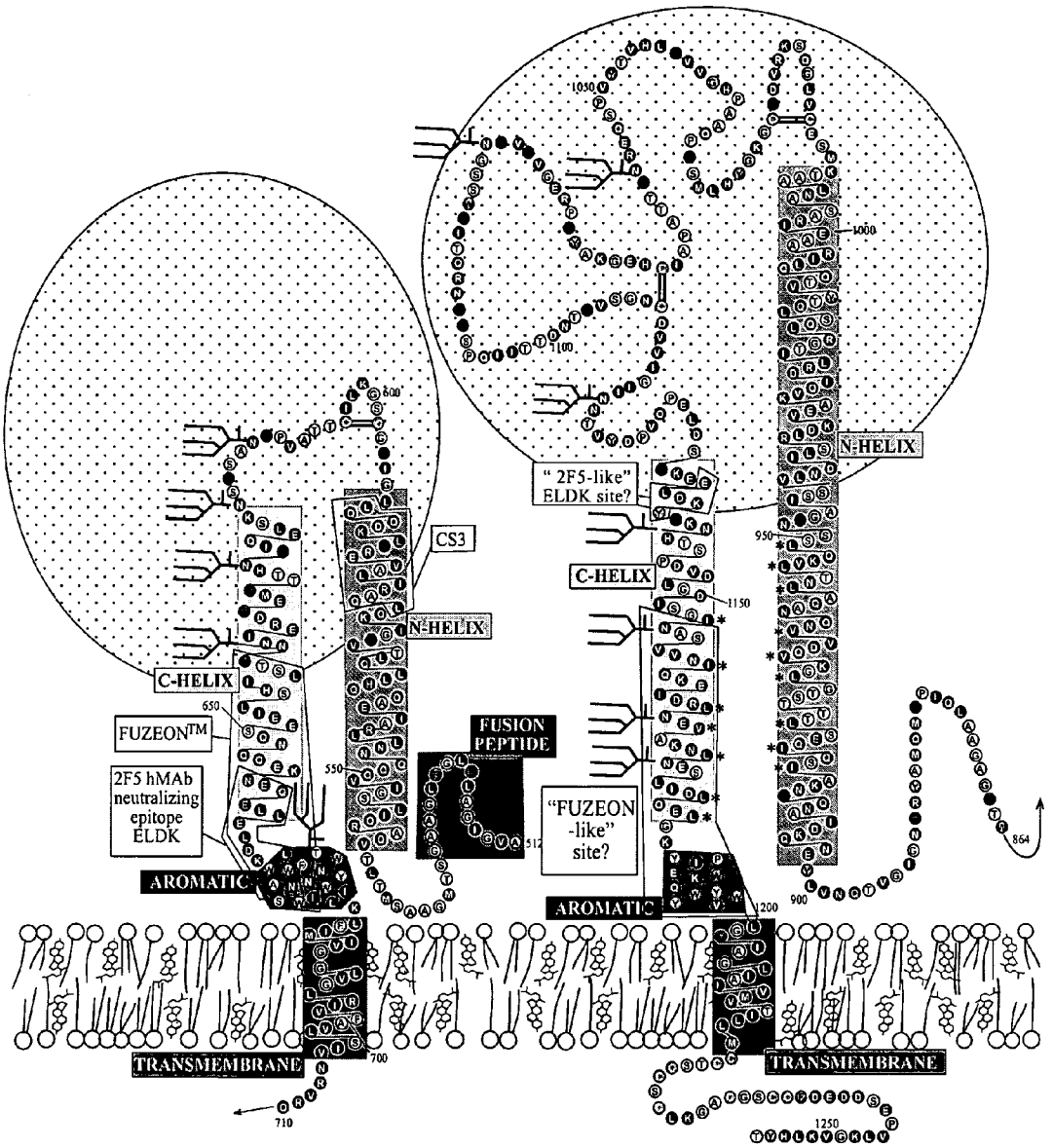
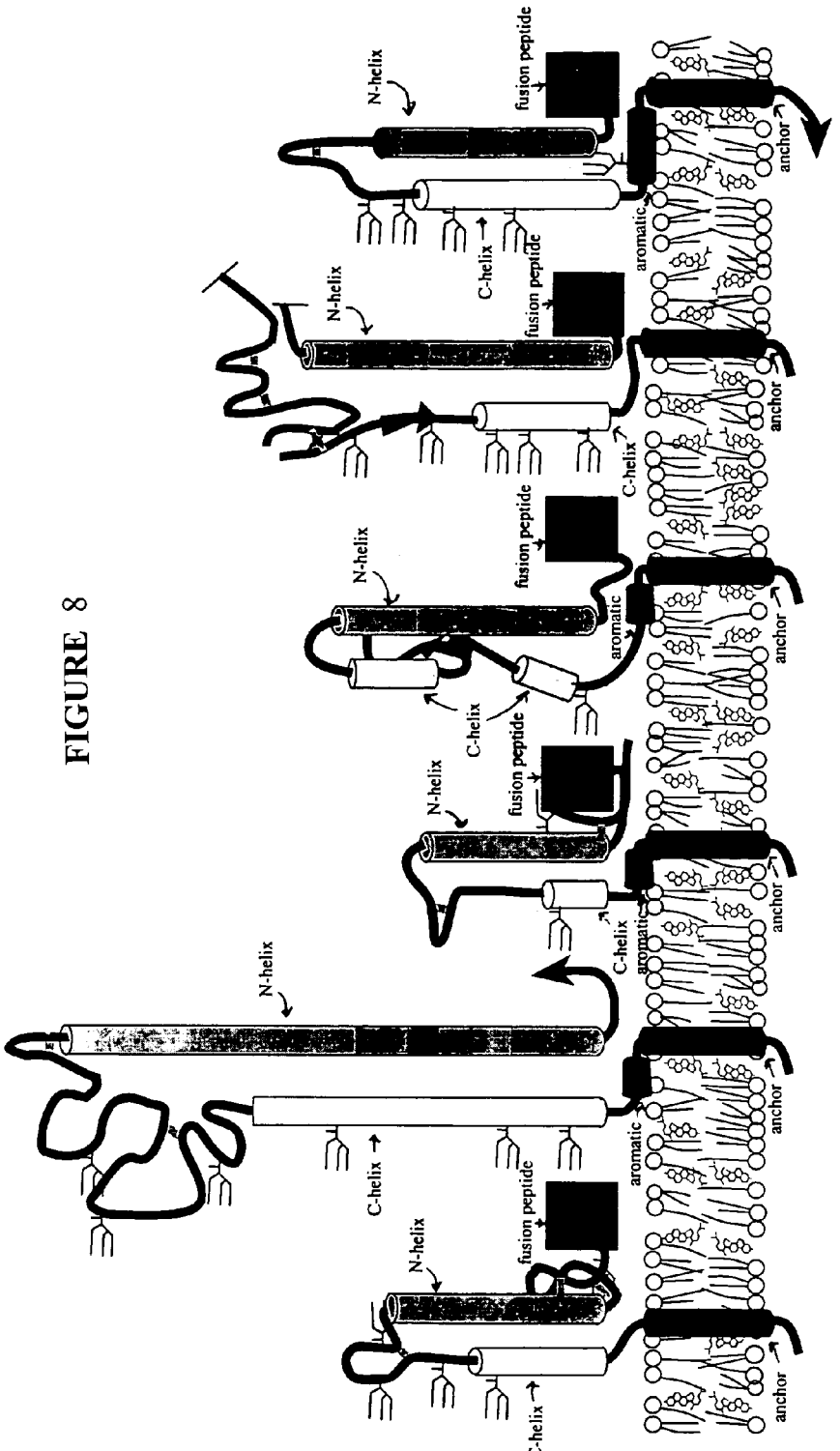

" 2F5-like" ELDK site?
N-HELIX
CS3
C-HELIX
N-HELIX
C-HELIX
FUZEON™
2F5 hMAb neutralizing epitope ELDK
FUSION PEPTIDE
"FUZEON -like" site?
AROMATIC
AROMATIC
TRANSMEMBRANE
TRANSMEMBRANE
HIV-1 TM
SARS CoV S

FIGURE 7

FIGURE 8



Family:	Arenavirus	Coronavirus	Filovirus	Orthomyxovirus	Paramyxovirus	Retrovirus
Example:	Lassa virus GP2	SARS coronavirus S2	Ebola virus GP2	Influenza virus HA2	Measles virus F1	HIV-1 TM

Fig. 1. Common structural features of RNA virus fusion proteins.
R.F. Garry, SARS coronavirus fusion inhibitors

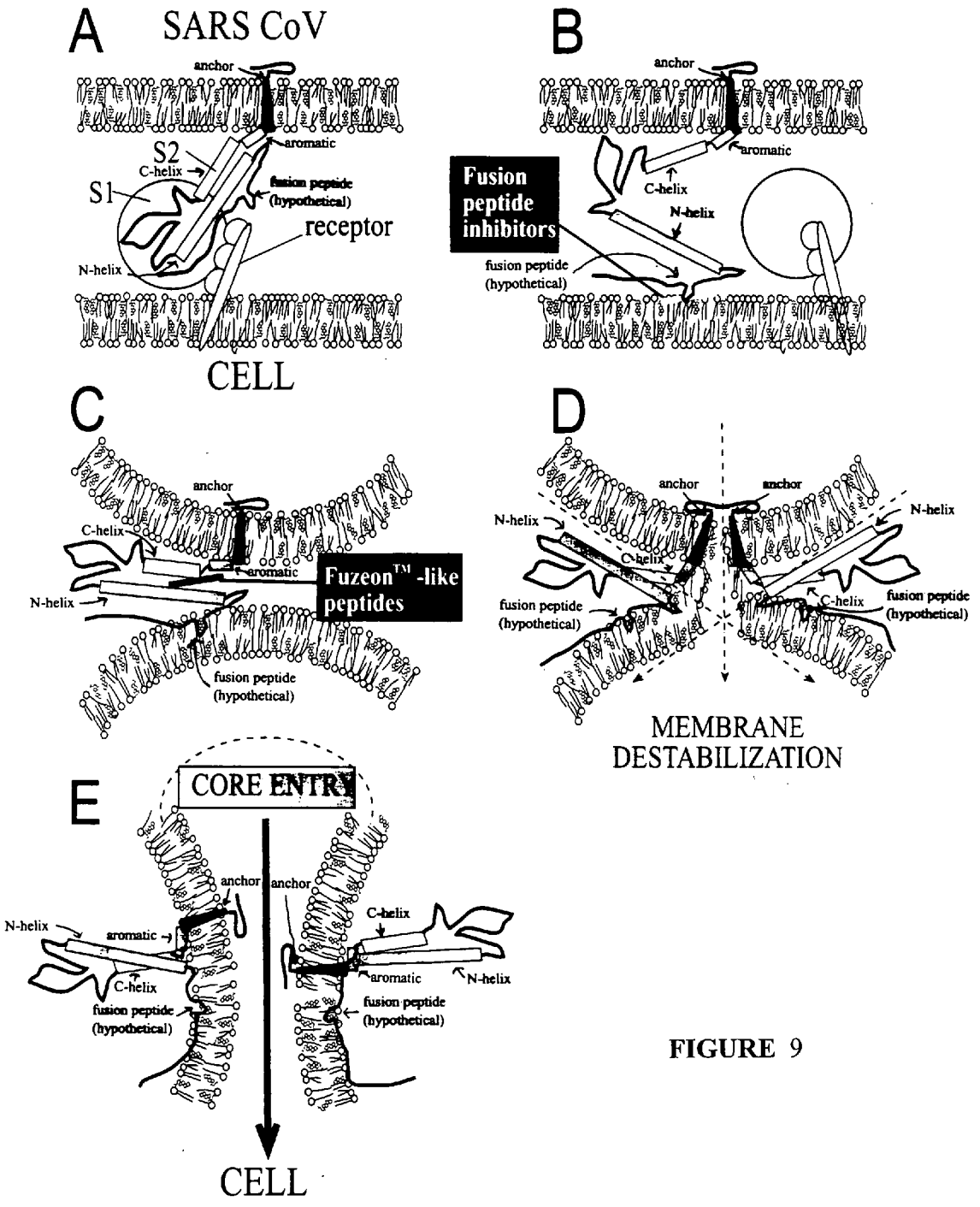

A
SARS CoV
anchor
aromatic
S2
C-helix
S1
fusion peptide
(hypothetical)
receptor
N-helix
CELL
B
anchor
aromatic
Fusion peptide inhibitors
C-helix
N-helix
fusion peptide
(hypothetical)
C
anchor
C-helix
aromatic
Fuzeon™ -like peptides
N-helix
fusion peptide
(hypothetical)
D
anchor
anchor
N-helix
N-helix
C-helix
C-helix
fusion peptide
(hypothetical)
fusion peptide
(hypothetical)
MEMBRANE DESTABILIZATION
E
CORE ENTRY
anchor
anchor
N-helix
aromatic
C-helix
fusion peptide
(hypothetical)
C-helix
aromatic
N-helix
fusion peptide
(hypothetical)
CELL

FIGURE 9

Localizing conserved regions of amino acid sequence
Alignment of Charged Pre-Insertion Helix
Among human and murine coronavirus strains

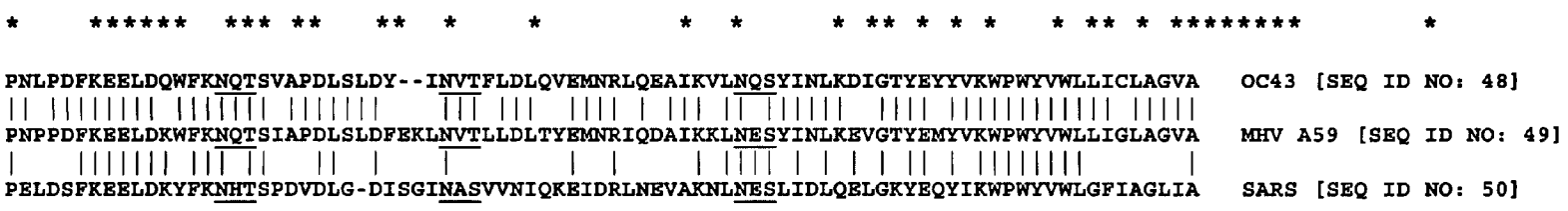


FIGURE 10

Metapneumovirus Fusion glycoprotein
Charged Pre-Insertion Helix

Example from AY145300 Human metapneumovirus isolate CAN00-15 fusion protein (F) gene
Amino acid residues 425-491 inclusive.

YQLSKVEGEQHVIKGRPVSSSFDPIKFPEDQFNVALDQVFESIENSQALVDQSNKILNSAEKGNTGF	[Seq ID 01]
YQLSKVEGEQHVIKGRPVSSSFDPIKFPEDQFNV	[Seq ID 03]
PVSSSFDPIKFPEDQFNVALDQVFESIENSQAL	[Seq ID 04]
ALDQVFESIENSQALVDQSNKILNSAEKGNTGF	[Seq ID 05]
YQLSKVEGEQHVIKGRPVSSSFDPIKFP	[Seq ID 06]
EDQFNVALDQVFESIENSQALVDQSNKILNSAEKGNTGF	[Seq ID 07]
KFPEDQFNVALDQVFESIENSQALVDQSNKILNSAEK	[Seq ID 08]
EDQFNVALDQVFESIENSQALVDQSNKILNSAEK	[Seq ID 09]
QALVDQ	[Seq ID 36]

FIGURE 11

Human SARS Coronavirus surface (S) glycoprotein
Charged Pre-insertion Helix

Example from NC_004718 Human SARS coronavirus, TOR2 strain, surface glycoprotein S gene
Amino acids 1125-1202 inclusive

```
PELDSFKEELDKYFKNHTSPDVDLGDISGINASVVNIQKEIDRLNEVAKNLNESLIDLQELGKYEQYIKWPWYVWLGF   [Seq ID 02]
PELDSFKEELDKYFKNHTSPDVDLGDISGINASVVN                                              [Seq ID 10]
                    DVDLGDISGINASVVNIQKEIDRLNEVAKNLNESLID                          [Seq ID 11]
                                         RLNEVAKNLNESLIDLQELGKYEQYIKWPWYVWLGF   [Seq ID 12]
                                  NIQKEIDRLNEVAKNLNESLIDLQEL                       [Seq ID 13]
                                                LNESLIDLQELGKYEQYIKWPWYVWLGF   [Seq ID 14]
                                                        QELGKYEQYIKWPWYVWLGF   [Seq ID 15]
                                                             YEQYIKWPWYVWLGF   [Seq ID 16]
                                                             YEQYIKWPWYVWLG    [Seq ID 17]
                                                                YIKWPWYVWLGF   [Seq ID 18]
PELDSFKEELDKYFKNHTSP                                                           [Seq ID 19]
     FKEELDK                                                                   [Seq ID 34]
                                                                  KWPWYVWL     [Seq ID 35]
```

FIGURE 12

Human Coronavirus OC43 surface (S) glycoprotein
Charged Pre-insertion helix

Example from S62886 Human coronavirus OC43 surface glycoprotein S gene
Amino acids 1220-1309 inclusive

PNLPDFKEELDQWFKNQTSVAPDLSLDYINVTFLDLQVEMNRLQEAIKVLNQSYINLKDIGTYEYYVKWPWYVW	[Seq ID 20]
PNLPDFKEELDQWFKNQTSVAPDLSLD	**[Seq ID 21]**
YINVTFLDLQVEMNRLQEAIKVLNQSYINLKDIGTYEYYVKW	**[Seq ID 22]**
QVEMNRLQEAIKVLNQSYINLKDIGTYEYYVKW	**[Seq ID 23]**
QEAIKVLNQSYINLKDIGTYEYYVKWPW	**[Seq ID 24]**
QSYINLKDIGTYEYYVKWPW	**[Seq ID 25]**
YEYYVKWPWYVW	**[Seq ID 26]**

FIGURE 13

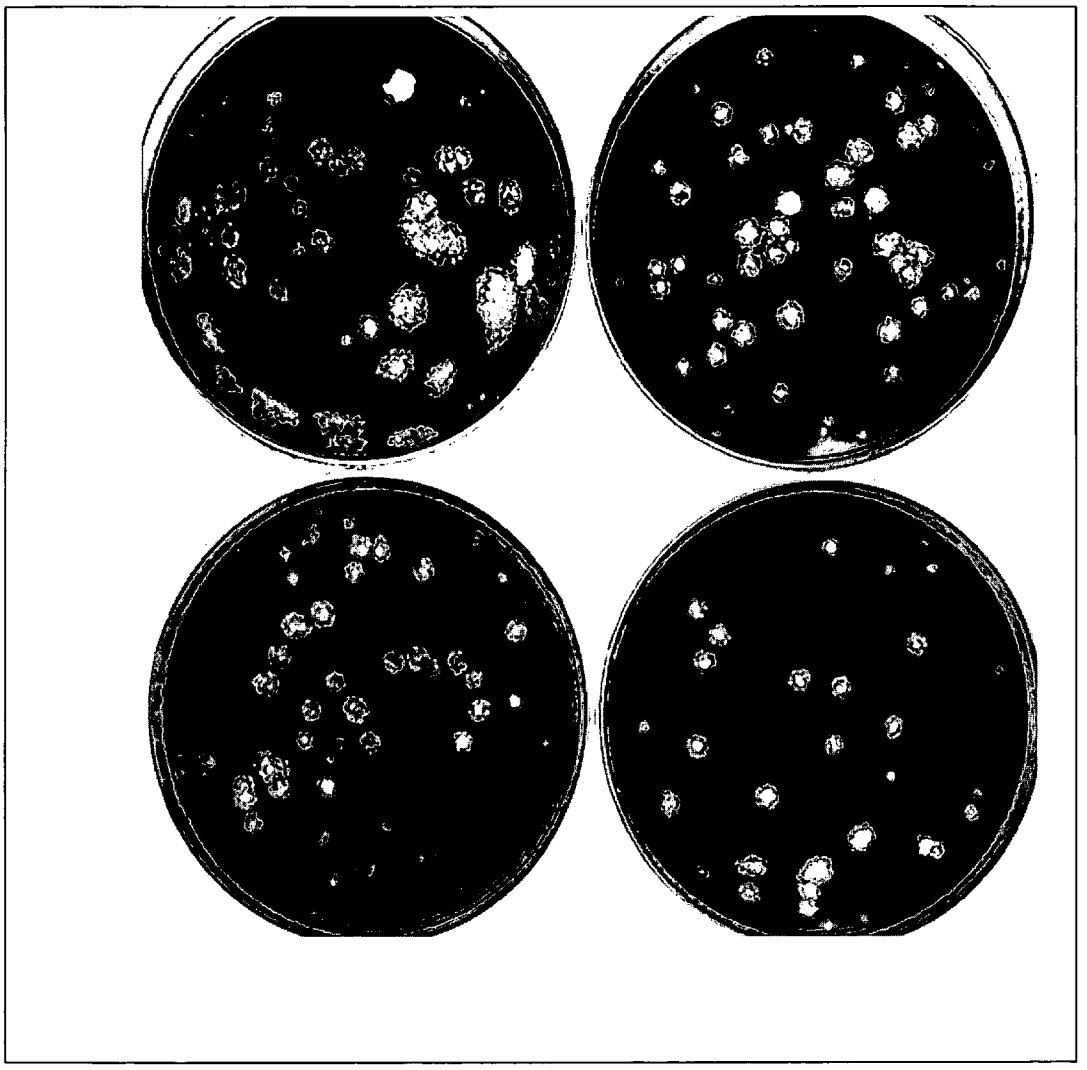


FIGURE 14

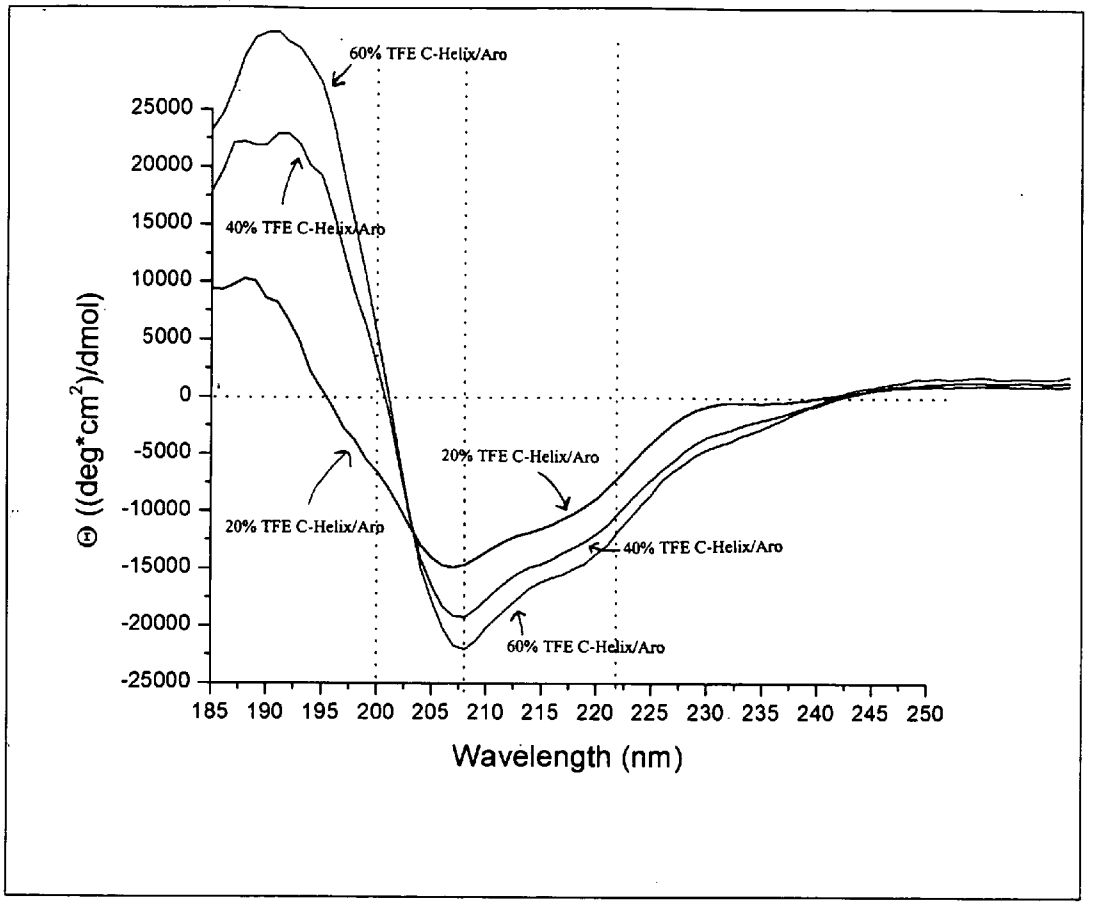

60% TFE C-Helix/Aro
40% TFE C-Helix/Aro
20% TFE C-Helix/Aro
20% TFE C-Helix/Aro
40% TFE C-Helix/Aro
60% TFE C-Helix/Aro
Θ ((deg*cm²)/dmol)
Wavelength (nm)
25000
20000
15000
10000
5000
0
-5000
-10000
-15000
-20000
-25000
185 190 195 200 205 210 215 220 225 230 235 240 245 250

FIGURE 15

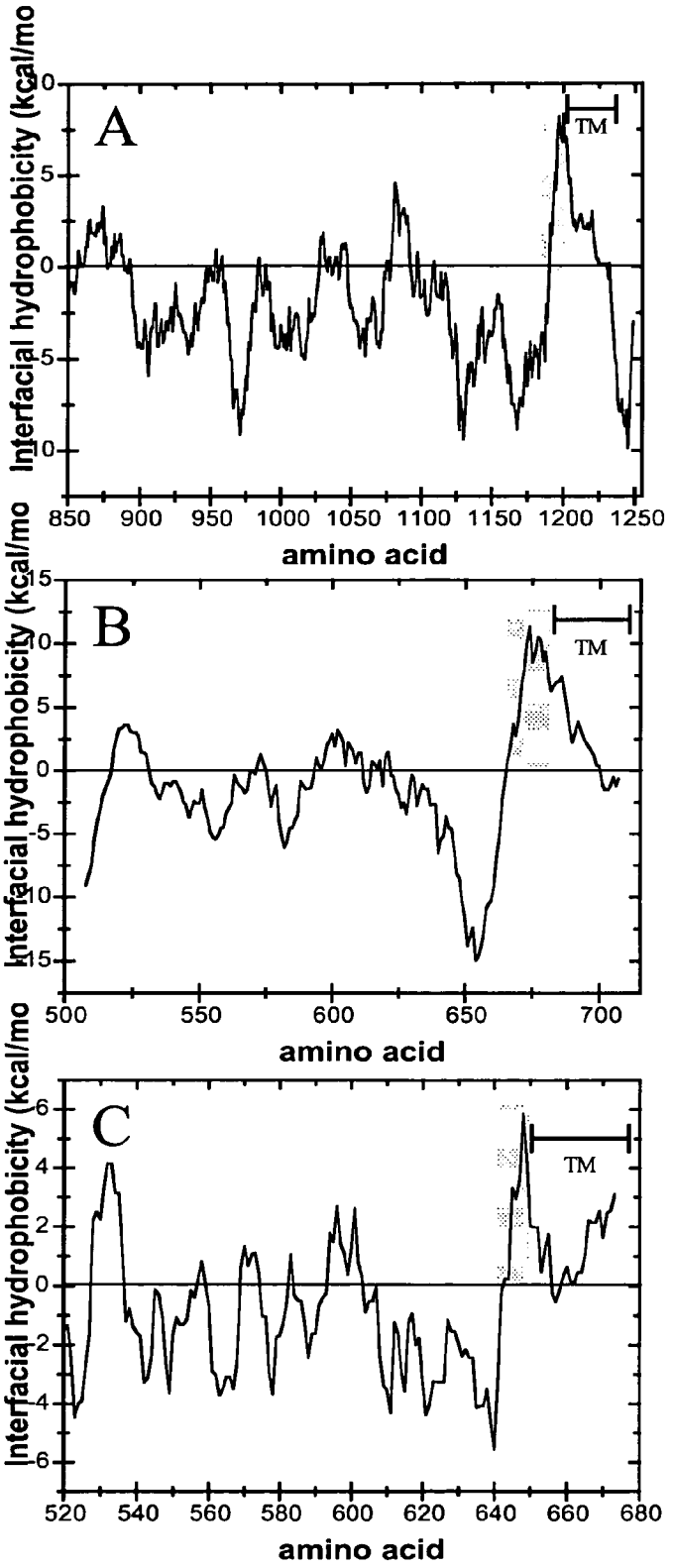

A
TM
Interfacial hydrophobicity (kcal/mo
10
5
0
-5
-10
850 900 950 1000 1050 1100 1150 1200 1250
amino acid
B
TM
Interfacial hydrophobicity (kcal/mo
15
10
5
0
-5
-10
-15
500 550 600 650 700
amino acid
C
TM
Interfacial hydrophobicity (kcal/mo
6
4
2
0
-2
-4
-6
520 540 560 580 600 620 640 660 680
amino acid

FIGURE 16

Amino Acid Sequences and WW Hydropathy Scores of CoV Aromatic Peptides

Peptide	Amino acid sequence	Position within S2	WW interfacial hydrophobicityd
$SARS_{Aro}$	**K**YE**Q**Y**I**KWPWYVW [SEQ ID NO: 44]	1187-1199	3.58
MHV_{Aro}	**T**YE**M**Y**Y**KWPWYVW [SEQ ID NO: 45]	1264-1276	4.86
$OC43_{Aro}$	**T**YE**Y**Y**Y**KWPWYVW [SEQ ID NO: 46]	1289-1301	5.57
$SARS_{Scrb}$	YEWKWIYWYPVKQ [SEQ ID NO: 47]	1187-1199	3.58

FIGURE 17

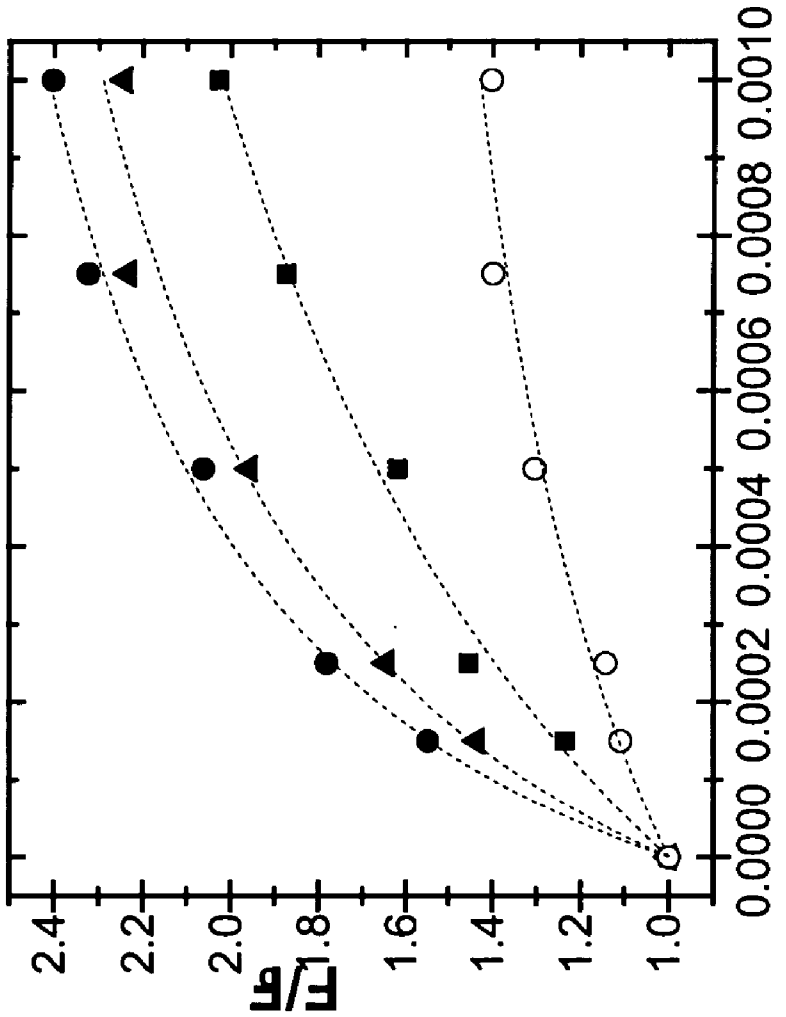

F/F
2.4
2.2
2.0
1.8
1.6
1.4
1.2
1.0
0.0000
0.0002
0.0004
0.0006
0.0008
0.0010
[Lipid] (M)

FIGURE 18

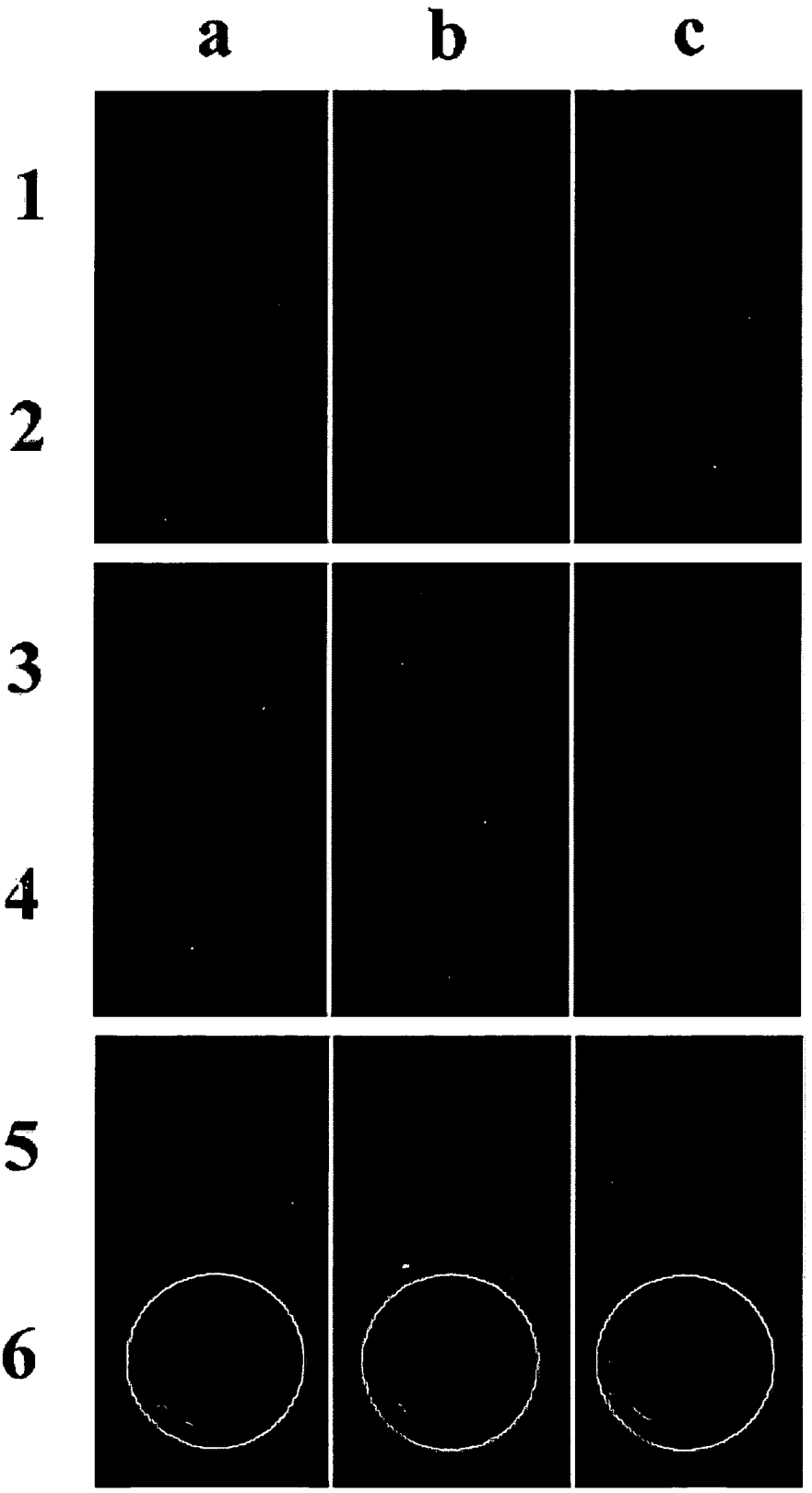


FIGURE 19

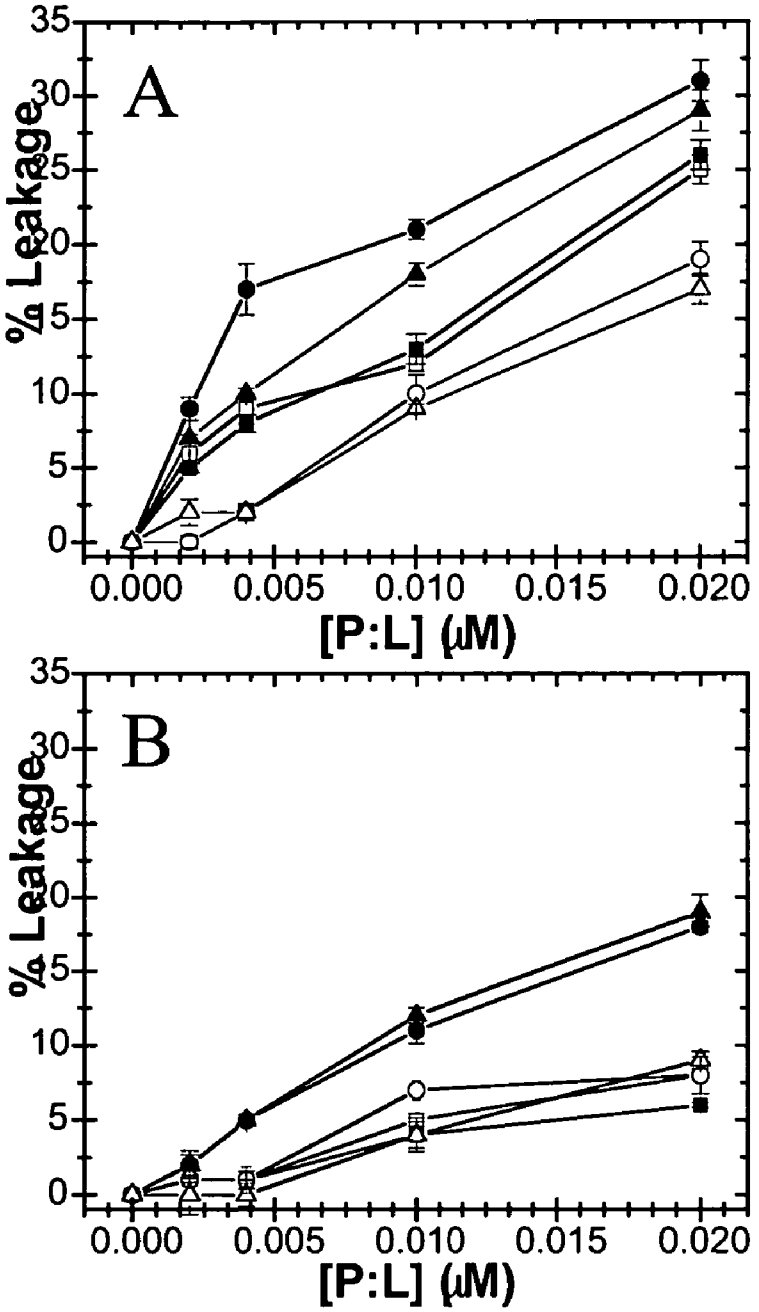

A
% Leakage
[P:L] (μM)
0.000
0.005
0.010
0.015
0.020
0
5
10
15
20
25
30
35
B
% Leakage
[P:L] (μM)

FIGURE 20

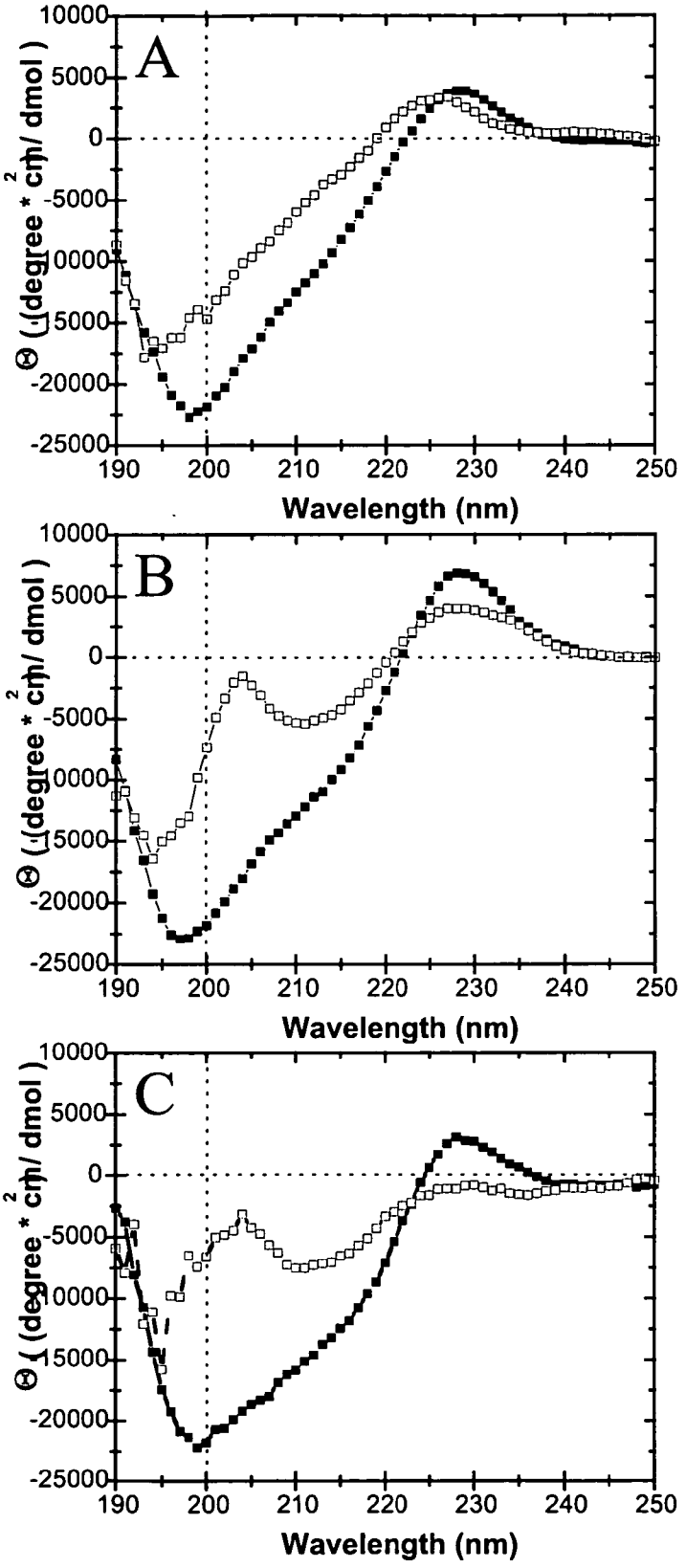

A
B
C
10000
5000
0
-5000
-10000
-15000
-20000
-25000
190
200
210
220
230
240
250
Wavelength (nm)
Θ (degree * cm2/ dmol)

FIGURE 21

METHOD OF INHIBITING HUMAN METAPNEUMOVIRUS AND HUMAN CORONAVIRUS IN THE PREVENTION AND TREATMENT OF SEVERE ACUTE RESPIRATORY SYNDROME (SARS)

CROSS-REFERENCE TO RELATED APPLICATION

[0001] This application claims the benefit of U.S. Provisional Application No. 60/466,978, filed Apr. 30, 2003, which is hereby incorporated by reference herein in its entirety.

STATEMENT REGARDING FEDERALLY SPONSORED RESEARCH AND DEVELOPMENT

[0002] This invention was made with Government support under Grants No. AI-54238, No. AI-34764 and No. CA-08921 awarded by the National Institutes of Health, and Grant No. BC990847 awarded by the Department of Defense. The Government has certain rights in the invention.

BACKGROUND OF THE INVENTION

[0003] A. Field of the Invention

[0004] The present invention relates to peptides that show significant antiviral activity. In certain embodiments, the invention relates to the design and use of peptides to inhibit membrane fusion and infection by human metapneumovirus and human coronavirus in the prevention and treatment of Severe Acute Respiratory Syndrome (SARS) or other severe respiratory diseases caused by theses agents.

[0005] B. Description of Related Art

[0006] SARS is a newly emerging infection which was first identified during an outbreak in southern China in March of 2003 (Drosten et al. "Identification of a Novel Coronavirus in Patients with Severe Acute Respiratory Syndrome,"*N Engl J Med* 2003 Apr. 10; Ksiazek et al. "A Novel Coronavirus Associated with Severe Acute Respiratory Syndrome,"*N Engl J Med* 2003 Apr. 10; Poutanen S M, et al. "Identification of Severe Acute Respiratory Syndrome in Canada."*N Engl J Med* 2003 Mar. 31). By Apr. 16, 2003, over 3,235 individuals had been diagnosed with SARS, and over 161 SARS-related deaths had been recorded in 22 countries, on every continent except Antarctica. Air travel by infected individuals during the incubation period prior to the onset of symptoms greatly facilitated the spread of the infection to many countries, including the United States. The nature of the epidemic and the exact etiologic agent(s) of SARS are still under investigation. However, the illness does not appear to involve bacterial, fungal, or previously identified viral agents of human disease. Molecular amplification of nucleic acid sequences from patient samples revealed that a number of patients were infected contemporaneously by human metapneumovirus (MPV) and a new human coronavirus (CoV).

[0007] Through the end of 2003, the initial overall outbreak totaled approximately 8,098 cases of SARS worldwide, with an overall mortality of 9.6% (http://www.who.int/csr/sars/en/). The previously unrecognized SARS CoV has been demonstrated to have been the principal cause of the new disease (Drosten et al., 2003; Poutanen et al., 2003; Peiris et al. (2003). Coronavirus as a possible cause of severe acute respiratory syndrome. Lancet 361, 1319-25). (It is important to note that metapneumovirus was found in a substantial percentage of cases in China, to an extent greater than pure coincidence would indicate; thus, its role in increasing the severity of SARS cannot be ruled out.) In a remarkably short period of time, the entire genetic sequences of several strains of the novel SARS CoV were determined (Ksiazek et al., 2003; Marra et al. (2003). The genome sequence of the SARS associated coronavirus. *Science* 300, 1399-1404; Rota et al. Characterization of a novel coronavirus associated with severe acute respiratory syndrome. *Science.* 2003 May 30; Wang et al. (2003). Gene sequence analysis of SARS-associated coronavirus by nested RT-PCR. *Di Yi Jun Yi Da Xue Xue Bao* 23, 421-3, each of which is hereby incorporated by reference herein in its entirety) and the cellular receptor, ACE-2, for the virion identified (Li et al. (2003). Angiotensin-converting enzyme 2 is a functional receptor for the SARS coronavirus. *Nature* 426, 450-454). Isolates of a similar CoV were obtained from civets and other animals that are trapped for food or medicine at live animal markets in Guangdong mainland China, the presumed epicenter of the SARS outbreak (Guan et al. (2004). Molecular epidemiology of the novel coronavirus that causes severe acute respiratory syndrome. Lancet 363, 99-104). SARS CoV or a closely related CoV also infects animals in the wild (Guan, Y., Zheng, B. J., He, Y. Q., Liu, X. L., Zhuang, Z. X., Cheung, C. L., Luo, S. W., Li, P. H., Zhang, L. J., Guan, Y. J., Butt, K. M., Wong, K. L., Chan, K. W., Lim, W., Shortridge, K. F., Yuen, K. Y., Peiris, J. S., and Poon, L. L. (2003). Isolation and characterization of viruses related to the SARS coronavirus from animals in southern China. *Science* 302, 276-278.) and appears to have entered the human population in the past (Zheng et al. (2004). SARS-related virus predating SARS outbreak, Hong Kong. *Emerging Infectious Diseases.* e-pub Jan. 16, 2004). Public health interventions such as surveillance, travel restrictions, and quarantines contained the spread of SARS in 2003 and appear to have stopped the spread of SARS after the appearance of a few new cases in 2004. It is unknown, however, whether these draconian containment measures can be sustained with each appearance of SARS in humans. Furthermore, this new and sometimes lethal CoV has potential as a bioterrorism threat. There are no antiviral agents which are known to be effective in the treatment of SARS, and no antiviral agents are known to be effective against either metapneumovirus or coronavirus in humans or animals.

[0008] Human MPV is a recently characterized agent of human respiratory infection that appears to be a member of the Paramyxoviridae family of viruses (van den Hoogen, B. G. et al. "Analysis of the genomic sequence of a human metapneumovirus,"*Virology* 2002 Mar. 30, 295(1): 119-32; Peret, T. C. et al. "Characterization of human metapneumoviruses isolated from patients in North America,"*J Infect Dis* 2002 Jun. 1, 185(11):1660-3; van den Hoogen, B. G. et al. "A newly discovered human pneumovirus isolated from young children with respiratory tract disease,"*Nat Med* 2001 Jun., 7(6):719-24). Other members of this virus family include historically significant human pathogens such as measles virus, mumps virus, parainfluenza virus, and respiratory syncytial virus. Prior to the identification of SARS, human MPV was generally associated with mild respiratory

infection in humans, except for a small number of cases in individuals with serious pulmonary or immunological compromise such as leukemia. The molecular sequence of the nucleic acid genome of human MPV has recently been determined, confirming the similarity of its genome sequence to other Paramyxoviruses and indicating that human MPV is distantly related to other Paramyxovirus agents of human disease such as measles virus, mumps virus, parainfluenza virus, and respiratory syncytial virus. The molecular sequence of human MPV, which is hereby incorporated by reference in its entirety, can be accessed at the National Center for Biotechnology Information's (NCBI) web site at http://www.ncbi.nlm.nih.gov/ as Genbank reference sequence AY145301. Human MPV appears to be most similar at the molecular level to avian metapneumovirus, perhaps reflecting an introduction of the virus into the human population from an avian source at some undetermined time in the past (Njenga, M. K. et al. Metapneumoviruses in birds and humans. *Virus Res.* 2003 February;91(2):163-9). While there is extensive literature concerning the molecular and cell biology of Paramyxoviruses generally because of their overall significance in human disease, there is relatively little known specifically concerning human MPV. The antiviral drug ribavirin has been used to treat severe cases of human respiratory syncytial virus, which is distantly related to human MPV, and there is experimental evidence in mice that anti-inflammatory cytokines may augment ribavirin therapy (Bonville, et al., 2003 "Altered Pathogenesis of Severe Pneumovirus Infection in Response to Combined Antiviral and Specific Immunomodulatory Agents," J. Virol. 77:1237-1244, which is hereby incorporated by reference herein in its entirety), but there is no evidence that such a therapeutic regimen is effective against SARS or human MPV infection.

[**0009**] Human coronavirus (human CoV) is a member of the Coronaviridae family of viruses. Various strains of human CoV have been isolated from mild outbreaks of human respiratory infection for many years, and these viruses are generally known to be a part of the diverse group of "common cold" viruses. There has been little direct characterization of human CoV and the specific aspects of its molecular and cell biology. However, there has been written a significant amount of literature regarding a murine coronvirus known as mouse hepatitis virus (MHV), which is much more severe in the mouse than human CoV has been heretofore in humans (see Luo et al., 1999, "Amino Acid Substitutions within the Leucine Zipper Domain of the Murine Coronavirus Spike Protein Cause Defects in Oligomerization and the Ability to Induce Cell-to-Cell Fusion,"*J. Virol.* 73: 8152-8159, which is incorporated herein by reference in its entirety). The entire molecular genome sequence of the human CoV strain involved in SARS has been determined both in Canada and at the Centers for Disease Control and Prevention (CDCP), where this new strain of human CoV was recently isolated. The genome sequence has been made available on the Internet at http://www.ncbi.nlm.nih.gov/ in advance of publication as sequence number NC_004718 by the NCBI, which genome sequence is incorporated herein by reference in its entirety. Preliminary analysis of a conserved region of the genome indicates that this strain constitutes a new group within the Coronaviridae family, not closely related to any previously identified strain of the virus (Marra, M. A. et al. The Genome sequence of the SARS-associated coronavirus Science 300 (5624), 1399-1404 (2003); Rota, P. A. et al. Characterization of a novel coronavirus associated with severe acute respiratory syndrome. Science. 2003 May 30;300 (5624):1394-9. Epub 2003 May 1, each of which is hereby incorporated by reference herein in its entirety).

[**0010**] Human MPV and human CoV are thus very different viruses, both from one another and from other human viral pathogens. Nevertheless, drawing from our knowledge of the viruses' families, we have identified common features that may be used to design antiviral drugs effective at inhibiting infection at the cellular level.

[**0011**] Human MPV and human CoV are both members of a subset of all viruses known as "enveloped" viruses. Their outer layer is composed of a membranous envelope which is derived from the cellular membranes of host cells during infection. The envelope is studded with proteins encoded by the viral genome. These proteins are modified by the addition of sugar side groups at specific positions in the linear sequence of amino acids that comprise the protein and are thus termed "viral membrane glycoproteins." Such viral membrane glycoproteins are quite variable and individual in their amino acid sequences (even sometimes from strain to strain of the same virus) and may serve a variety of functions in infection. Some of these viral membrane glycoproteins are directly anchored to the membrane because part of the protein spans the membrane—they are generally known as "viral transmembrane glycoproteins" or sometimes "spike" glycoproteins because of their shape. Other viral membrane glycoproteins, termed "viral peripheral glycoproteins," are indirectly anchored to the viral membrane by specific association with such viral transmembrane glycoproteins, even though they do not themselves have a membrane anchor sequence. It has been discovered that a number of subcategories of viral membrane glycoproteins have general features that may be exploited for the development of specific antiviral drugs.

[**0012**] One subcategory includes viral membrane glycoproteins responsible for the entry of the virus into the host cell via specific binding to the host cell followed by fusion of the viral membrane with a host cell membrane, either the plasma membrane or an internal membrane (see White, J. M., 1992, "Membrane Fusion," Science 258:917-924, which is hereby incorporated by reference herein in its entirety). The binding and fusion functions are performed by separate regions of the glycoprotein complex. Attachment is usually mediated by a viral peripheral glycoprotein, and membrane fusion or entry, is usually mediated by a viral transmembrane glycoprotein (those viral transmembrane glycoproteins that mediate fusion are known as "fusion glycoproteins" or "transmembrane fusion glycoproteins"). In many cases, viral glycoproteins responsible for binding and fusion are made together as one complex, which is later divided by a polypeptide cleavage event into two functional subunits; this happens with influenza and HIV, for instance. In other cases, such as measles, the binding and fusion functions are always separated on two different glycoproteins.

[**0013**] Work over the last 25 years has shown that dissimilar virus families share a similar molecular machinery and mechanism of viral entry (see **FIG. 1**). This similarity was first detailed by the structural studies of the viral membrane glycoprotein of influenza virus, known as the hemagglutinin (Wilson, I. A., et al. 1981. "Structure of the

hemagglutinin membrane glycoprotein of influenza virus at 3 A resolution."*Nature* 289: 366-373, which is hereby incorporated by reference herein in its entirety). High resolution x-ray crystallography allowed visualization of the globular head group of the hemagglutinin, which binds to the cell receptor for influenza, and the fibrous leg region of the protein complex, which anchors the protein complex to the viral membrane via a transmembrane spanning domain and induces fusion of the viral and cellular membranes. As noted, these two functional regions are activated by the proteolytic cleavage of a hemagglutinin precursor into two glycoprotein subunits that correspond to each functional region—the receptor binding glycoprotein is known as HA1 and the fusion glycoprotein is known as HA2. It was recognized in influenza virus (and measles virus) that the new amino terminus of the fusion gylycoprotein generated by this cleavage event was highly hydrophobic and of conserved sequence (White, J. M., "Membrane Fusion-."*Science, vol.* 258 (Nov. 6, 1992), pp. 917-924; Eckert D. M., and Kim P. S., "Mechanisms of viral membrane fusion and its inhibition."*Annu Rev Biochem.* 2001, 70:777-810, each of which is hereby incorporated by reference herein in its entirety). This hydrophobic segment of amino acids is thought to be a critical functional element in the viral fusion transmembrane glycoprotein; it is thought to interact with and insert into the target membrane, inducing membrane perturbation and thereby membrane fusion. This segment of amino acids, first identified in measles virus by Choppin's lab in the early 1980s, and immediately found also in influenza virus, became known as the "fusion peptide." The hypothesis that the fusion peptide is a critical element in fusion became known as the "fusion peptide hypothesis."

[**0014**] Such structural studies converged with early efforts to use peptides as antivirals in controlling infection. Much earlier, Parke-Davis researchers had tested a series of random small peptides against a variety of viral infections and discovered that a carbobenzoxy derivative of phenylalanine-phenylalanine-glycine (z-FFG) was effective against measles virus (Miller, F. A., et al. (1968), "Antiviral activity of carbobenzoxy di- and tripeptides on Measles virus,"*Applied Microbiology* 16: 1489-1496; Nicolaides, E., et al. 1968 "Potential antiviral agents. Carbobenzoxy di- and tri-peptides active against Measles and herpes viruses,"*Journal of Medicinal Chemistry* 11: 74-79, each of which is hereby incorporated by reference herein in its entirety). These results were confirmed by more standard virological techniques in 1971 (Norrby, E. 1971, "The effect of a carbobenzoxy tripeptides on the biological activities of measles virus,"*Virology* 44: 599-608, which is hereby incorporated by reference herein in its entirety). Subsequent structural studies showed that z-FFG was a peptide analogue of the fusion peptide sequences at the amino termini of the measles virus and influenza virus fusion glycoproteins (Richardson, C. D., et al. (1980) "Specific inhibition of Paramyxovirus and myxovirus replication by oligopeptides and amino acid sequences similar to those at the N-termini of the F1 or HA2 viral polypeptides,"*Virology* 105: 205-222; Hsu, M. C. et al. (1981) "Activation of the Sendai virus Fusion protein (F) involves a conformational change with exposure of a new amino terminus,"*Virology* 104: 294-302; Richardson, C. D., and Choppin, P. W. (1983) "Oligopeptides that specifically inhibit membrane fusion by paramyxoviruses: studies on the site of action,"*Virology* 131: 518-532, each of which is hereby incorporated by reference herein in its entirety). However, the highly hydrophobic nature of the peptide and the existence of potent vaccines for each of these viruses precluded the development of z-FFG and similar peptides as clinically useful antiviral drugs.

[**0015**] Beginning in 1987, Gallaher and co-workers extended these studies to human immuno-deficiency virus type 1 (HIV-1), providing the first evidence that the structure of the influenza virus fusion glycoprotein and, more generally, the fusion peptide hypothesis could be extended to a superfamily of viral entry glycoproteins that crossed the lines delineating a number of otherwise dissimilar virus families. The tandem repeat of a fusion peptide motif (FLGFLG [SEQ ID NO: 28]) was located in the amino terminus of the transmembrane fusion glycoprotein subunit of HIV-1 known as gp41 (Gallaher, W. R. (1987) "Detection of a fusion peptide sequence in the transmembrane protein of human immunodeficiency virus,"*Cell* 50: 327-328, which is hereby incorporated by reference herein in its entirety). It was found that: "First, the gp41 transmembrane protein is likely to be the fusion glycoprotein of HIV and may be responsible for infection of cells as well as for the cytopathic effects of fusion and cytolysis. Second, as in the case of paramyxoviruses, small peptides such as Phe-Leu-Gly, its derivatives, or drugs targeted to this peptide region, may have direct inhibitory effects on HIV infection and cytopathlogy with high specific activity." These findings were embodied in U.S. Pat. No. 4,880,779 to Gallaher, which is hereby incorporated by reference herein in its entirety, and confirmed by genetic studies of HIV (Kowalski, M., et al. 1987 "Functional Regions of the envelope glycoprotein of human immunodeficiency virus Type 1". *Science* 237: 1351-1355, which is hereby incorporated by reference herein in its entirety).

[**0016**] By mid-1988, Gallaher and co-workers determined that the remaining structure of gp41 could be fitted to approximate the scaffold of the structure of the fusion glycoprotein of influenza virus (see Gallaher, W. R., et al. (1989) "A general model for the transmembrane proteins of HIV and other retroviruses,"*AIDS Research and Human Retroviruses* 5: 431-440, which is hereby incorporated by reference in its entirety). The structure of the transmembrane fusion glycoproteins of a number of viruses in the Retrovirus family were found to have similar overall structures to the transmembrane fusion glycoprotein of influenza virus, in spite of wide amino acid sequence variation. This overall structure was found to define a transmembrane fusion glycoprotein superfamily containing at its core a "coiled coil" structure. The elements of the overall structure were identified as an amino terminal fusion peptide region, followed by an extended "sided" helix termed "amphipathic" (or "N-helix," which is hydrophobic on one side, hydrophilic on the other), a disulfide cross-linked central region, and a "charged pre-insertion helix" (or "C-helix") just prior to membrane insertion (see **FIG. 2**) (Gallaher, W. R., et al. (1989) "A general model for the transmembrane proteins of HIV and other retroviruses,"*AIDS Research and Human Retroviruses* 5: 431-440, which is hereby incorporated by reference herein in its entirety). The helices were designated N-helix and C-helix, depending on which end of the fusion glycoprotein, N-terminus or C-terminus, it is closer to relative to the other helix. The two antiparallel helices partly wrap around two other pairs in a trimeric structure to form

the coiled coil. This superfamily of viral fusion glycoproteins has come to be known as the "class I" superfamily of fusion glycoproteins.

[0017] In 1989, Gallaher extended the concept of utilizing peptide analogues of the sequence of gp41 to include analogues of the two major helical regions of HIV-1 and described this approach in a series of grant applications to the National Institutes of Health from 1989 through 1990. The applications were not funded, and, thus, the extended study of inhibitory peptides was deferred indefinitely.

[0018] In 1990, Delwart introduced the term "leucine zipper-like" to describe the helical regions in gp41 (Delwart, E. L. et al., 1990 "Retroviral envelope glycoproteins contain a 'leucine zipper'-like repeat," AIDS Research and Human Retroviruses 6:703-706, which is hereby incorporated by reference herein in its entirety). Although not entirely accurate, this characterization has since been widely applied to helical structural elements in viral transmembrane fusion glycoproteins. Also in 1990, Pringle's laboratory discovered that helical structural motifs could also be found in the transmembrane fusion glycoproteins of members of the Paramyxovirus family (Chambers et al., 1990, "Heptad repeat sequences are located adjacent to hydrophobic regions in several types of virus fusion glycoproteins,"*J. Gen. Virology* 71:3075-3080, which is hereby incorporated by reference herein in its entirety), providing additional impetus for the concept that a superfamily of viral entry proteins extended over the Orthomyxoviridae, Paramyxoviridae and Retroviridae virus families, despite significant differences in amino acid sequence and genome structure (reviewed in Gallaher, W., Henderson, L., Fermin, C., Montelaro, R., Martin, A., Qureshi, M., Ball, J., Sattentau, Q., Luo-Zhang, H., and Garry, R. (1992a). Membrane interactions of human immunodeficiency virus: Attachment, fusion and cytopathology. In "Membrane Interactions of HIV" (R. Aloia, Ed.), Vol. 6, pp. 113-142. Wiley-Liss, Inc., NY., and in Gallaher, W., Fermin, C., Henderson, L., Montelaro, R., Martin, A., Qureshi, M., Ball, J., Luo-Zhang, H., and Garry, R. (1992b). Membrane interactions of HIV: Attachment, fusion and cytopathology. *Adv Membrane Fluidity* 6, 113-42., each of which is hereby incorporated by reference herein in their entirety).

[0019] In 1992 and 1993, Matthews and co-workers used peptides derived from the HIV-1 gp41 sequence in an assay to determine the potential of the peptides to inhibit fusion induced by HIV-1 (Wild et al., 1992 "A synthetic peptide inhibitor of human immunodeficiency virus replication: Correlation between solution structure and viral inhibition, "*Proc Natl Acad. Sci. USA.* 89:10537-10541; Wild et al., 1994 "Propensity for a Leucine Zipper-Like Domain of Human Immunodeficiency Virus Type 1 gp41 to Form Oligomers Correlates With a Role in Virus-Induced Fusion Rather Than Assembly of the Glycoprotein Complex,"*Proc Natl Acad. Sci USA* 91:12676-30, each of which is hereby incorporated by reference herein in its entirety). These findings were embodied in U.S. Pat. No. 5,464,933 to Bolognesi, et al. and U.S. Pat. No. 5,656,480 to Wild, et al. The length and location of the inhibitory peptides, including the drug Fuzeon™ recently licensed for use against HIV by the Food and Drug Administration, was set by the length of the amphipathic helix first described by Gallaher (Rimsky et al., 1998, "Determinants of Human Immunodeficiency Virus type 1 Resistance to gp41-derived Inhibitory Peptides", *J. Virol.* 72:986-993, which is hereby incorporated by reference herein in its entirety).

[0020] In 1993, Carr and Kim demonstrated that the fusion glycoprotein of influenza virus undergoes a "spring-loaded" conformational change in the course of being activated into a fusogenic form, triggering the action of the fusion peptide. (Carr, C. M. and Kim, P. S. "A spring-loaded mechanism for the conformational change of influenza hemagglutinin, "*Cell.* 1993 May 21;73(4):823-32, which is hereby incorporated by reference herein in its entirety). It has since been theorized that peptide inhibitors of fusion modeled after the sequence of fusion glycoproteins may function by interfering with this essential conformational change, possibly by preventing the firing of the fusion peptide towards its target cellular membrane. (Chen, et al. 1994, "Functional role of the zipper motif region of human immunodeficiency virus type 1 transmembrane protein gp41,"*J. Virology* 68:2002-2010, which is hereby incorporated by reference herein in its entirety).

[0021] In 1996, Lambert and co-workers extended this rationale to the Paramyxoviruses, which had been suggested five years earlier by Chambers to also contain coiled coil structures (see Chambers, et al., 1990 "Heptad repeat sequences are located adjacent to hydrophobic regions in several types of virus fusion glycoproteins,"*J Gen Virology* 71:3075-3080, which is hereby incorporated by reference in its entirety). Specific peptide analogues were designed of the amino acid sequences of the fusion glycoproteins for measles virus, respiratory syncytial virus, and humar parainfluenza virus (Lambert, D. M., et al. "Peptides from conserved regions of paramyxovirus fusion (F) proteins are potent inhibitors of viral fusion."*Proc Natl Acad Sci USA.* 1996 Mar. 5;93(5):2186-91, which is hereby incorporated by reference herein in its entirety). These discoveries were embodied in U.S. Pat. No. 6,479,055 to Bolognesi, et al. This patent also teaches that a computer-assisted search technology may be used to identify coiled coil motifs as candidates for fusion inhibitors. The computer technology described is neither necessary nor sufficient, with a specificity so low that the tables included in the patent enumerate approximately 13,000 peptides, varying in length from 15 to hundreds of amino acids in length, the vast bulk of which are likely irrelevant to the process of membrane fusion. The peptides described in the present invention are not included even in that extensive enumeration of peptides.

[0022] In 1996 and 2001, respectively, the viral entry glycoprotein superfamily was extended by Gallaher and co-workers to Ebola virus of the Filoviridae family (Gallaher, W. R. (1996) "Similar structural models of the transmembrane proteins of Ebola and Avian sarcoma viruses, "*Cell.* 85: 477-478, which is hereby incorporated by reference herein in its entirety) and to Lassa fever virus of the Arenaviridae family (Gallaher, W. R., et al. "The viral transmembrane superfamily: possible divergence of Arenavirus and Filovirus glycoproteins from a common RNA virus ancestor."*BMC Microbiol.* 2001;1(1):1, which is hereby incorporated by reference herein in its entirety), both of which are agents of hemorrhagic fevers (see **FIG. 3**) (see also U.S. Pat. No. 6,713,069 to Gallaher, which is hereby incorporated by reference herein in its entirety). Potentially inhibitor effective peptides have also been identified from the amino acid sequences of fusion glycoproteins from the

Filoviridae, from other retroviruses, such as human T-cell leukemia virus (Pinon et al., 2003, "An Antiviral Peptide Targets a Coiled-Coil Domain of the Human T-Cell Leukemia Virus Envelope Glycoprotein,"*J. Virol.* 77:3281-3290, which is hereby incorporated by reference herein in its entirety) and from feline immunodeficiency virus (Medinas, R. J., et al. "C-Terminal gp40 peptide analogs inhibit feline immunodeficiency virus: cell fusion and virus spread."*J. Virol.* 2002 September;76(18):9079-86, which is hereby incorporated by reference herein in its entirety).

[**0023**] Most of the fusion glycoprotein peptide analogues which have been proposed as antivirals thus far are "long" (about 30 amino acids or more). A defect of such long peptide analogues is that, for example, in the case of HIV-1 inhibition, the peptides must be administered by subcutaneous injection. More recently, Kim and co-workers have created variants of the HIV-1 peptide analogues which are intended to provide small molecules as inhibitors that may ultimately be orally administered (Sia, S. K., et al. "Short constrained peptides that inhibit HIV-1 entry."*Proc Natl Acad Sci USA.* 2002 Nov. 12;99(23):14664-9; Eckert, D. M., Kim, P. S. "Design of potent inhibitors of HIV-1 entry from the gp41 N-peptide region,"*Proc Natl Acad Sci USA.* 2001 Sep. 25;98(20):11187-92; Root, M. J., et al. "Protein design of an HIV-1 entry inhibitor,"*Science* 2001 Feb. 2;291(5505):884-8; Eckert, D. M., et al. "Inhibiting HIV-1 entry: discovery of D-peptide inhibitors that target the gp41 coiled-coil pocket,"*Cell* 1999 Oct. 1;99(1):103-15, each of which is hereby incorporated by reference herein in its entirety). Thus far, however, these variants have not exhibited the potency of the long peptides, and they have not been tested in humans.

[**0024**] In sum, work in several laboratories established that there is a superfamily of viral transmembrane fusion glycoproteins, quite variable in size and amino acid sequence, which extends to a superfamily of viruses which includes Orthomyxoviruses, Paramyxoviruses, Filoviruses, Arenaviruses and Retroviruses, five virus families that differ widely in genome structure and replication strategy (see **FIG. 4**). By alignment of the fusion peptide and membrane-spanning regions of the corresponding fusion glycoproteins, and by examining the vicinal sequences for possible alpha helical sequence motifs, potential inhibitory effective peptides may be designed for each individual member of these virus families.

[**0025**] Coronaviruses have long been considered unique and very distant outliers from the viruses which contain the superfamily of fusion glycoproteins discussed above. The genome structure and replication strategy of Coronaviruses is markedly different, and the entry proteins themselves are more complex and of a different overall structure. A coiled coil model of a MHV membrane glycoprotein was presented as long ago as 1987 (deGroot et al 1987, "Evidence for a Coiled-coil Structure in the Spike Proteins of Coronaviruses"*J Mol Biol* 196:963-6, which is hereby incorporated by reference herein in its entirety), and the identification of the fusion glycoprotein of MHV and the identification of extended "leucine zipper" heptad repeat motifs **(FIG. 5)** was achieved by mutational analyses (Luo et al., 1999 "Amino Acid Substitutions within the Leucine Zipper Domain of the Murine Coronavirus Spike Protein Cause Defects in Oligomerization and the Ability to Induce Cell-to-Cell Fusion,"*J. Virol* 73: 8152-8159, which is incorporated herein by reference in its entirety). Such analyses establish a latter carboxy-terminal half of the fusion glycoprotein, which is generally cleaved from the first half by an endoproteolytic enzyme, as the glycoprotein most responsible for coronavirus-induced membrane fusion. However, three factors prevented the CoV fusion glycoprotein (also known as the "S2 glycoprotein, ""S2," or "the S2 subunit") from being included in the superfamily of fusion glycoproteins discussed above. First, there is no canonical fusion peptide motif. Second, there is a great deal of amino acid sequence in the S2 glycoprotein prior to the first heptad repeat motif, which is unprecedented in the other virus families. Third, there is an extensive disulfide cross-linked region between the two heptad repeat motifs.

[**0026**] We recently examined these apparent dissimilarities in the context of the more unusual members of the virus families already included in the viral entry glycoprotein superfamily, such as the spumaretroviruses which contain large inserts of extra amino acid sequence and lack clearly defined fusion peptides. We found that the amino acids of the SARS CoV fusion glycoprotein located prior to membrane insertion can in fact be modeled as a similar structure to the viral fusion glycoproteins seen in HIV-1 and the other Retroviruses and Filoviruses, with an approximately 100 amino acid disulfide cross-linked region between the two heptad repeat regions (see **FIG. 6**). Also, the charged pre-insertion helix (with 16 charged amino acids out of 56 total) of the SARS CoV fusion glycoprotein is followed by a region rich in aromatic amino acids highly similar to corresponding regions in HIV-1 and Ebola virus.

[**0027**] The peptide sequence of the fusion glycoprotein of the SARS CoV (Urbani strain AY278741) can be fitted to the Gallaher et al. (1989) general scaffold of the gp41 fusion glycoprotein (also known as "TM") of HIV-1 (see **FIG. 6**). While lacking x-ray crystallographic or other biophysical data needed for confirmation, this model is consistent with the proven structures of other viral fusion glycoproteins, beginning with the influenza virus hemagglutinin in 1981 (Wilson, I. A., et al. 1981. "Structure of the haemagglutinin Membrane glycoprotein of influenza virus at 3 A resolution, "*Nature* 289, 366-373, which is hereby incorporated by reference herein in its entirety), as well as with similar suggestions and experimental data in other coronavirus systems from other laboratories (e.g., see Luo, et al., 1999 "Amino Acid Substitutions within the Leucine Zipper Domain of the Murine Coronavirus Spike Protein Cause Defects in Oligomerization and the Ability to Induce Cell-to-Cell Fusion,"*J Virol* 73: 8152-8159; Zelus et al. (2003). Conformational changes in the spike glycoprotein of murine Coronaviruses are induced at 37 degrees C. either by soluble murine CEACAM1 receptors or by pH 8. *J Virol* 77, 830-40, each of which is hereby incorporated by reference herein in its entirety). While cartoon models of the CoV fusion glycoprotein as a coiled coil were proposed as early as 1987 (de Groot et al. (1987). Sequence and structure of the coronavirus peplomer protein. *Adv Exp Med Biol* 218, 31-8, which is hereby incorporated by reference herein in its entirety), previous models have not been presented in this detail or demonstrated such close parallels with the other fusion glycoproteins. The detailed model presented here **(FIG. 7)**, shown in comparison to the known features and structure of HIV-1 TM glycoprotein, has significant implications for avenues to develop antiviral drugs that function as fusion inhibitors of the SARS CoV.

[0028] First, there is a minimum furin-like cleavage site located at amino acids 758-762. Beginning about amino acid 900 there is an extended heptad repeat region similar to the N-helix of the HIV-1 transmembrane glycoprotein. This region differs from the N-helices of the fusion glycoproteins of retroviruses, filoviruses (Ebola virus) and arenaviruses (Lassa fever virus) principally in the extraordinarily length of the helix (see Gallaher, W. R. 1996 "Similar structural models of the transmembrane proteins of Ebola and Avian sarcoma viruses,"*Cell* 85: 477-478; Gallaher, W. R., et al. 1989 "A general model for the transmembrane proteins of HIV and other retroviruses,"*AIDS Research and Human Retroviruses* 5, 431-440; Gallaher et al. (2001). The viral transmembrane superfamily: possible divergence of arenavirus and filovirus glycoproteins from a common RNA virus ancestor. *BMC Microbiol* 1, 1, each of which is incorporated by reference herein in its entirety) (see **FIG. 8**). While there are helix-breaking motifs present (e.g., TTTS [SEQ ID NO: 29]), the helix may be stabilized in such areas by the very strong heptad repeat of hydrophobic amino acids along the left side of the helix projection. At 17 nm, this helix is overly long for the known dimensions of the coronavirus surface spike, but may reflect an extension that occurs upon binding or configurational alteration of the protein while in the process of becoming a fusion-active form.

[0029] Second, there is a short region bounded by cysteines, (see **FIG. 7**) which is so similar to that of the fusion glycoproteins of the retroviruses and Ebola virus to prompt us to model it as a similar disulfide-stabilized apex.

[0030] Third, there is a region with several sites (shown by stick figures in **FIG. 7**) for possible N-linked glycosylation that, like HIV-1, are only found after the disulfide-linked apex. This region is highly variable among Coronavirus membrane glycoproteins proteins, not unlike the variability among the retrovirus transmembrane glycoproteins.

[0031] Fourth, there is a region prior to the point the SARS CoV fusion glycoprotein is anchored in the viral envelope membrane, which has a high percentage of charged amino acids, a strong propensity to form an α helix, and a heptad repeat, so that it is comparable to the C-helix (known as HR2) of the HIV-1 transmembrane glycoprotein. SARS CoV and other CoV have well-conserved "leucine-zipper-like" motifs in the C-helix with leucine or isoleucines spaced such that they would form a highly hydrophobic face along the helix (Luo et al. (1999). Amino Acid Substitutions within the Leucine Zipper Domain of the Murine Coronavirus Spike Protein Cause Defects in Oligomerization and the Ability to Induce Cell-to-Cell Fusion. *J. Virol.* 73: 8152-8159, which is hereby incorporated by reference herein in its entirety). It has been demonstrated that mutations in this region of the C-helix of the MHV fusion glycorprotein cause defects in oligomerization and the ability to induce cell:cell fusion. (Luo et al. (1999). Amino Acid Substitutions within the Leucine Zipper Domain of the Murine Coronavirus Spike Protein Cause Defects in Oligomerization and the Ability to Induce Cell-to-Cell Fusion. *J. Virol.* 73: 8152-8159). The N-helix of the SARS CoV also has a readily identifiable "leucine-zipper-like" motif. Although the "leucine-zipper" is not as evident in the N-helices of other CoV, the N- and C-helices may nevertheless interact to form a "hydrophobic groove" or other non-covalent interactions (see Bosch, B. J., van der Zee, R., de Haan, C. A., and Rottier, P. J. (2003). The coronavirus spike protein is a class I virus fusion protein: structural and functional characterization of the fusion core complex. *J Virol* 77, 8801-8811, which is hereby incorporated by reference herein in its entirety). The "hydrophobic groove" is a groove or slot in the antiparallel helical structure that is lined with hydrophobic amino acids. The "leucine-zipper-like" motifs, with amino acids in the predicted hydrophobic grove of the SARS CoV fusion glycoprotein, marked by asterisks, are depicted in **FIG. 7**.

[0032] The amino terminal end of this charged pre-insertion helix shows a peptide motif ELDKY [SEQ ID NO: 30] highly conserved among Coronaviruses, which is very similar to a biologically significant peptide, ELDKW [SEQ ID NO: 31], in the C-helix of HIV-1 gp41. In HIV-1 this peptide is recognized as a neutralization epitope, for which a human monoclonal antibody has been developed (Muster et al. (1993). A conserved neutralizing epitope on gp41 of human immunodeficiency virus type 1. *J Viol* 67, 6642-7; Muster et al. (1994). Cross-neutralizing activity against divergent human immunodeficiency virus type 1 isolates induced by the gp41 sequence ELDKWAS. *J Virol* 68, 4031-4, each of which is hereby incorporated by reference herein in its entirety) and is in human clinical trials (Stiegler et al. (2002). Antiviral activity of the neutralizing antibodies 2F5 and 2G12 in asymptomatic HIV-1 infected humans: a phase I evaluation. *AIDS* 16, 2019-25, which in hereby incorporated by reference herein in its entirety) The ELDKW [SEQ ID NO: 31] motif is also represented in the recently licensed peptide fusion inhibitor, Fuzeon™, that suppresses HIV-1 infection in the nanomolar range (Kilby et al. (1998). Potent suppression of HIV-1 replication in humans by t-20, a peptide inhibitor of gp41-mediated virus entry. *Nat Med* 4, 1302-7, which is hereby incorporated by reference herein in its entirety).

[0033] Finally, just prior to membrane insertion (the membrane spanning domain was predicted by TMpred (http://www.ch.embnet.org/software/TMPRED_form.html). there is a region enriched in aromatic amino acids and extraordinarily highly conserved throughout the Coronaviridae. Because of its high interfacial propensity (Yau et al., 1998) it is unlikely that the tryptophan (W) rich aromatic domain is part of the transmembrane anchor in contrast to the prediction of Rota et al. (2003). This region lies in an identical location to comparable aromatic rich regions in the fusion glycoproteins of HIV-1 and Ebola virus, which have been shown to be fusogenic in liposome systems (Suarez, et al., 2000 "Membrane Interface-Interacting Sequences within the Ectodomain of the Human Immunodeficiency Virus type 1 Envelope Glycoprotein: Putative Role During Viral Fusion,"*J. Virol.* 74:8038-8047, which is hereby incorporated by reference herein in its entirety). An experimental octapeptide mimicking this region of the feline immunodeficiency virus (FUV) transmembrane glycoprotein has been found to inhibit fusion by that retrovirus in cell culture (Giannecchini et al., 2003 "Antiviral Activity and Conoformational Features of an Octapeptide Derived from the Membrane-Proximal Ectodomain of the Feline Immunodeficiency Virus Transmembrane Ectodomain,"*J. Virol.* 77:3724-3733, which is hereby incorporated by reference herein in its entirety).

[0034] We have not modeled further toward the amino terminus of the SARS CoV fusion glycoprotein, since there are no parallels established among other viruses for the

structure of the fusion glycoprotein prior to the N-helix. This region, including the receptor-binding domain, is only shown schematically in **FIG. 7** as a large ellipse corresponding to the large globular head group that forms the top of the characteristic "lollipop" spike seen in electron micrographs of the Coronavirus, giving it the "crown-like" appearance from which the virus family derives its name.

[**0035**] **FIG. 9** illustrates our hypothetical mechanism for SARS CoV virion-cell fusion. PANEL A shows binding of the SARS CoV membrane glycoprotein to the cell receptor. Class I viral fusion proteins have a fusion peptide at the amino terminus, two extended α helices (N-helix and C-helix) and most have an aromatic rich domain proximal to the transmembrane anchor. Although it has been proposed that the viral entry glycoprotein of SARS CoV is not cleaved into S1 and S2 (also known as the "fusion glycoprotein") subunits (see Rota et al., 2003), the presence of a minimal furin cleavage site suggests that the viral entry glycoprotein is cleaved. PANEL B shows rearrangement of the helical domains of the viral entry glycoprotein. The rearrangement allows the putative fusion peptide to interact with the cell plasma membrane. S1 is released from S2 in CoV when cleavage occurs. The fusion peptide may also reside between the N and C helical domains (Luo et al., 1999). PANEL C shows the helical domains of S2 "snap back" bringing the viral and cell membrane in closer proximity, and resulting in membrane deformation or "nipple" formation. Alternatively, the rearrangement of the S2 protein into the six-helix bundle confirmation does not result in nipple formation, but rather the virion itself is drawn closer to the cell surface. The fusion peptide, aromatic domain, and transmembrane anchor then constitute a contiguous track of sequences that can facilitate the flow of lipid between the two membranes. PANEL D shows the six helix bundle formation driving the cellular and viral membrane closer together resulting in spontaneous hemifusion. Peptide mimics (e.g. Fuzeon™-like peptides) of the paired helices and/or the aromatic domain are expected to block 6-helix formation in this step or in the alternative arrangement of PANEL C. PANEL E shows the fusion pore permitting cytoplasmic entry of the SARS CoV core.

[**0036**] The structural parallel of the helical fibrous region of the SARS CoV fusion glycoprotein to the HIV-1 transmembrane glycoprotein and other members of the same superfamily of viral transmembrane glycoproteins offers considerable support for the predicted fusion inhibitory effects of antiviral peptides modeled from the amino acid sequence of the SARS CoV fusion glycoprotein. Structural evidence has recently been provided that is consistent with this model, further suggesting that the coronavirus fusion glycoprotein is a class I fusion glycoprotein (Bosch, B. J., van der Zee, R., de Haan, C. A., and Rottier, P. J. (2003). The coronavirus spike protein is a class I virus fusion protein: structural and functional characterization of the fusion core complex. *J Virol* 77, 8801-8811, which is hereby incorporated by reference herein in its entirety). Furthermore, it has been demonstrated that amino acid substitutions in the N-helix (HR1) affect MHV spread in the central nervous system, and also confirmed the role of this domain in defining pH requirements for cell:cell, fusion and entry (Tsai, J. C. et al. (2003). Amino acid substitutions within the heptad repeat domain 1 of murine coronavirus spike protein restrict viral antigen spread in the central nervous system. *Virology* 312, 369-380, which is hereby incorporated by reference herein in its entirety). Dutch researchers have demonstrated that long synthetic peptides corresponding to the N-helix (HR1) and C-helix (HR2) of the MHV fusion glycoprotein form stable antiparallel helical complexes (Bosch, B. J. et al. (2003). The coronavirus spike protein is a class I virus fusion protein: structural and functional characterization of the fusion core complex. *J Virol* 77, 8801-8811, which is hereby incorporated by reference herein in its entirety). These researchers also demonstrated that a C-helix peptide could inhibit virus entry and cell:cell fusion mediated by the MHV fusion glycoprotein.

[**0037**] This latter study confirmed our earlier hypothesis. We had previously predicted the detailed SARS CoV S glycoprotein model (**FIG. 6**) that fusion inhibitory peptides may be designed from the amino acid sequence of the fusion glycoprotein of the SARS CoV by the methods disclosed herein. No such peptides had been previously proposed to inhibit coronavirus infection.

BRIEF DESCRIPTION OF THE INVENTION

[**0038**] In one embodiment, the present invention relates to a method of inhibiting human metapneumoviral infection and/or human coronavirus infection which comprises administering to a host an inhibitory effective amount of a peptide or peptides comprising an inhibitory effective sequence derived from the sequence of the fusion glycoproteins of human metapneumovirus or human coronavirus, respectively. While the invention may be used in any case of human infection by these respiratory viruses, the principal target of inhibition is to prevent or reduce the severity of SARS. Reference to SARS is intended to encompass any condition meeting the case definition of SARS established by the CDCP or by the World Health Organization (WHO).

[**0039**] The inhibitory peptides are designed as analogues to the amino acid sequence of the metapneumovirus and coronavirus fusion glycoproteins corresponding to regions of those proteins within the linear sequence of about 100 amino acids which lie just prior to the membrane spanning sequence that anchors the glycoprotein complex to the viral membrane. In one aspect, the relevant amino acid sequences for peptides derived from metapneumovirus are:

YQLSKVEGEQHVIKGRPVSSSFDPIKFPEDQFNV [SEQ ID NO: 01]

ALDQVFESIENSQALVDQSNKILNSAEKGNTGF,

[**0040**] and a selection of discreet sub-sequences and derivatives thereof, as defined below. In one aspect, the relevant sequences for peptides derived from human coronavirus are:

PELDSFKEELDKYFKNHTSPDVDLGDISGINASV [SEQ ID NO: 02]

VNIQKEIDRLNEVAKNLNESLIDLQELGKYEQYI

KWPWYVWLGF and

PNLPDFKEELDQWFKNQTSVAPDLSLDYINVTFL [SEQ ID NO: 20],

DLQVEMNRLQEAIKVLNQSYINLKDIGTYEYYVK

WPWYVW,

[**0041**] and a selection of discreet sub-sequences and derivatives thereof, as defined below. For each sequence

discussed herein, amino acids are defined by standard single letter code, defined by convention as follows:

A = Alanine	C = Cysteine	D = Aspartate
E = Glutamate	F = Phenylalanine	G = Glycine
H = Histidine	I = Isoleucine	K = Lysine
L = Leucine	M = Methionine	N = Asparagine
P = Proline	Q = Glutamine	R = Arginine
S = Serine	T = Threonine	V = Valine
W = Tryptophane	Y = Tyrosine	

[0042] In each case, the peptide or peptides to be administered may be given singly or in combination, and either naturally occurring or synthetic amino acids may be used for synthetic generation of peptides, or the peptides may be generated by translation in vivo or in vitro from a DNA plasmid coding for the sequence.

[0043] The overall region from which these peptides are derived has been shown in several viral systems, including the Paramyxoviruses and Coronaviruses that are the subject of his invention, to be critical in the fusion and entry mechanisms leading to infection of human cells.

BRIEF DESCRIPTION OF THE DRAWINGS

[0044] **FIG. 1** illustrates the different morphological forms of enveloped viruses and the common overall structure of the fusion machinery (i.e., the fusion peptide(s) in concert with the antiparallel N-helix and C-helix) used for cell entry, in this case for Ebola virus and HIV-1.

[0045] **FIG. 2** illustrates the 1988 Gallaher model of the viral transmembrane fusion glycoprotein of HIV-1, gp41, which provided the basis for identifying functional helices in such proteins and the design of antiviral drugs based on those helical structures.

[0046] **FIG. 3** illustrates the published models by Gallaher and co-workers for the fusion glycoproteins of Ebola virus and Lassa fever virus, agents of African hemorrhagic fevers, that show a striking similarity to the Gallaher model of HIV-1 gp41.

[0047] **FIG. 4** shows in cartoon form the overall structural similarity of models for the viral fusion glycoproteins from five separate virus families, with significant differences in genome structure and replication strategy.

[0048] **FIG. 5** shows a linear cartoon of the amino acid sequence of MHV, with the heptad repeats (HR1 and HR2) and the membrane-spanning (MS) region annotated, showing the large amount of amino acid sequence both prior to the first heptad repeat and between the heptad repeats.

[0049] **FIG. 6** illustrates a model of the SARS Coronavirus fusion glycoprotein by Garry and Gallaher illustrating the structure of the 350 amino acids prior to membrane insertion, and showing the commonality of structure with the other members of the superfamily of viral entry glycoproteins.

[0050] **FIG. 7** shows a comparison of HIV-1 TM with SARS CoV fusion glycoprotein. At the left of **FIG. 7** is an updated model of HIV-1 TM from Gallaher et al. (1989). At the right of **FIG. 7** is our hypothetical model of the SARS CoV fusion glycoprotein showing motifs shared with HIV-1 TM.

[0051] **FIG. 8** illustrates the common structural features of RNA virus fusion glycoproteins. Similar motifs found in representatives of diverse virus families are depicted in order from amino terminus to carboxyl terminus. These models are based on Gallaher (1987), Gallaher et al. (2001), Gallaher et al. (1989), other references noted in the text, and our preliminary experimental results. Truncations: HIV TM C-term; measles virus F1 after N-helix; SARS CoV S N-term.

[0052] **FIG. 9** illustrates our hypothetical mechanism for SARS CoV virion-cell fusion. PANEL A shows binding of the SARS CoV membrane glycoprotein to the cell receptor. Class I viral fusion proteins have a fusion peptide at the amino terminus, two extended a helices (N-helix and C-helix) and most have an aromatic rich domain proximal to the transmembrane anchor. Although it has been proposed that the viral entry glycoprotein of SARS CoV is not cleaved into S1 and S2 (also known as the "fusion glycoprotein") subunits (see Rota et al., 2003), the presence of a minimal furin cleavage site suggests that the viral entry glycoprotein is cleaved. PANEL B shows rearrangement of the helical domains of the viral entry glycoprotein. The rearrangement allows the putative fusion peptide to interact with the cell plasma membrane. S1 is released from S2 in CoV when cleavage occurs. The fusion peptide may also reside between the N and C helical domains (Luo et al., 1999). PANEL C shows the helical domains of S2 "snap back" bringing the viral and cell membrane in closer proximity, and resulting in membrane deformation or "nipple" foration. Alternatively, the rearrangement of the S2 protein into the six-helix bundle confirmation does not result in nipple formation, but rather the virion itself is drawn closer to the cell surface. The fusion peptide, aromatic domain, and transmembrane anchor then constitute a contiguous track of sequences that can facilitate the flow of lipid between the two membranes. PANEL D shows the six helix bundle formation driving the cellular and viral membrane closer together resulting in spontaneous hemifusion. Peptide mimics (e.g. Fuzeon™-like peptides) of the paired helices and/or the aromatic domain are expected to block 6-helix formation in this step or in the alternative arrangement of PANEL C. PANEL E shows the fusion pore permiting cytoplasmic entry of the SARS CoV core.

[0053] **FIG. 10** contains a comparison of the amino acid sequences of the CPI helices of human coronavirus OC43, MHV A59, and SARS CoV.

[0054] **FIG. 11** is a listing of peptide analogues of the CPI helix of human MPV which are predicted to be inhibitory effective.

[0055] **FIG. 12** is a listing of peptide analogues of the CPI helix of SARS CoV which are predicted to be inhibitory effective.

[0056] **FIG. 13** is a listing of peptide analogues of OC43 corresponding to peptide analogues of human SARS CoV; the figure also illustrates the relationship of those analogues to SEQ ID NO: 20.

[0057] **FIG. 14** illustrates the results of a MHV plaque reduction assay. Approximately 70 PFU of MHV were added to monolayers of L2 target cells in duplicate wells. The upper wells are controls exposed to vehicle and the lower wells exposed to MHV pretreated with a peptide having the amino acid sequence in SEQ ID NO: 52 at a

nominal concentration of 251 m. Plaques were visualized after 3 days by staining cells with crystal violet.

[0058] **FIG. 15** illustrates the results of Circular dichroism (CD) spectroscopy used to delineate the structural properties of a peptide corresponding to a region of the S2 protein of MHV encompassing a portion of the C-helix and the aromatic domain (SEQ ID NO: 52). The Results show that this peptide has a domain or domains with the propensity to form an α-helix.

[0059] **FIG. 16** illustrates interfacial hydrophobicity plots corresponding to sequences of SARS CoV S2, HIV-1 gp41, and EboV GP2. Interfacial hydrophobicity plots (mean values for a window of 13 residues) were generated using the Wimley and White (WW) interfacial hydrophobicity scales for individual residues (Wimley, W. C., and White, S. H. (1996) *Nat Struct Biol* 3, 842-848) of (PANEL A) SARS CoV strain Urbani S2 subunit (amino acids 850-1255), (PAENEL B) HIV-1 strain HXB2 gp41 (amino acids 502-710), and (PANEL C) Ebola virus strain Zaire GP2 (amino acids 520-676).

[0060] **FIG. 17** shows the amino acid sequences and WW hydropathy scores of the CoV aromatic peptides. The SARS aromatic ($SARS_{Aro}$), MHV aromatic (MHV_{Aro}) and OC43 aromatic ($OC43_{Aro}$) were synthesized based on their amino acid sequence determined from GenBank accession no. AY278741 (SARS-COV strain Urbani), AY497331 (MHV strain A59), and NP_937950 (Human CoV OC43). The $SARS_{Aro}$ sequence was arbitrarily scrambled to generate the peptide $SARS_{Scr}$. Amino acid differences among the three CoV aromatic peptides are shown in bold and underlined text. Hydropathy scores were determined according to the Wimley and White (WW) interfacial hydrophobicity scale using a window of 13 residues.

[0061] **FIG. 18** illustrates the $SARS_{Aro}$ peptide partitions into membranes of LUV. Change in tryptophan fluorescence of $SARS_{Aro}$ peptide as a function of increasing concentrations of LUV composed of (closed square) POPC, (closed circle) POPC:PI (9:1), (closed triangle) POPC:POPG (9:1) or (open circle) POPC:PI:CHOL (6.5:1:2.5). LUV were titrated at concentrations of 100, 250, 500, 750 and 1000 μM lipid with 2.5 M peptide. Tryptophan fluorescence values at each lipid titration (F) were normalized to tryptophan fluorescence values in potassium phosphate buffer alone (F_o).

[0062] **FIG. 19** illustrates the results of the Tb3+/DPA microwell assays, showing that the $SARS_{Aro}$ peptide induces leakage of LUV. Each well contained 250 μl of 50 μM DPA and 500 μM Tb3+-entrapped LUV composed of (a) POPC, (b) POPC:PI (9:1), or (c) POPC:POPG (9:1). Wells were treated with $SARS_{Aro}$ peptide at peptide:lipid molar ratios of 1:250 or 1:100 (rows 1-2), $SARS_{Scr}$ peptide at peptide:lipid molar ratios of 1:250 or 1:100 (rows 3-4), 20 μl of DMSO (row 4), or 20 μl of Triton-X-100 (row 5). Plates were incubated for 1 h at room temperature, and membrane permeabilization was determined by visual detection of Tb3+/DPA fluorescence.

[0063] **FIG. 20** illustrates the extent of leakage from ANTS-DPX LUV induced by the $SARS_{Aro}$ and $SARS_{Scr}$ peptides. $SARS_{Aro}$ peptide (PANEL A) and $SARS_{Scr}$ peptide (PANEL B) were added to LUV composed of (closed square) POPC, (closed circle) POPC:PI (9:1), (closed triangle) POPC:POPG (9:1), (open square) POPC:CHOL (7.5:2.5). (open circle) POPC:PI:CHOL (6.5:1:2.5), or (open triangle) POPC:POPG:CHOL (6.5:1:2.5) at different peptide:lipid (P:L) molar ratios. Samples were incubated at room temperature for 24 h before measuring the extent of leakage fluorometrically.

[0064] **FIG. 21** shows CD spectra (mean residue ellipticity θ) of the CoV aromatic peptides for $SARS_{Aro}$ (PANEL A), MHV_{Aro} (PANEL B), and $OC43_{Aro}$ (PANEL C) in 10 mM potassium phosphate buffer pH 7.0 alone (closed square) or with 1 mM LUV composed of POPC:PI (9:1) (open square) at room temperature.

REFERENCES CITED

[0065]

4,880,779	November 1989	Gallaher
4,943,627	July 1990	Gilbert et al.
5,017,688	May 1991	Gilbert et al.
5,464,933	November 1995	Bolognesi et al.
5,656,480	August 1997	Wild et al.
6,479,055	November 2002	Bolognesi et al.
6,713,069	March 2004	Gallaher

OTHER PUBLICATIONS

[0066] Arttamangkul S., Alvarez-Maubecin V., Thomas G., Williams J. T., Grandy D. K. Binding and internalization of fluorescent opioid peptide conjugates in living cells. Mol Pharmacol. 2000 December; 58(6):1570-80.

[0067] Bodansky, M., Bodansky, A., The practice of peptide synthesis (2nd edn.), Springer Verlag, Berlin (1995).

[0068] Bonville, et al., 2003 "Altered Pathogenesis of Severe Pneumovirus Infection in Response to Combined Antiviral and Specific Immunomodulatory Agents,"*J. Virol.* 77:1237-1244.

[0069] Bosch, B. J., van der Zee, R., de Haan, C. A., and Rottier, P. J. (2003). The coronavirus spike protein is a class I virus fusion protein: structural and functional characterization of the fusion core complex. *J Virol* 77, 8801-8811.

[0070] Carr and Kim, 1993 "A spring loaded mechanism for the conformational change of influenza hemagglutinin,"*Cell* 73:823-832.

[0071] Chambers, et al., 1990 "Heptad repeat sequences are located adjacent to hydrophobic regions in several types of virus fusion glycoproteins,"*J. Gen. Virology* 71:3075-3080.

[0072] Chen et al., 1994 "Functional role of the zipper motif region of human immunodeficiency virus type 1 transmembrane protein gp41,"*J. Virology* 68:2002-2010.

[0073] Compton S. R., Winograd D. F., Gaertner D. J. Optimization of in vitro growth conditions for enterotropic murine coronavirus strains. *J Virol Methods.* 1995 April; 52(3): 301-7.

[0074] de Groot et al 1987, "Evidence for a Coiled-coil Structure in the Spike Proteins of Coronaviruses"*J Mol Biol* 196:963-6.

[0075] de Groot, R. J., Lenstra, J. A., Luytjes, W., Niesters, H. G., Horzinek, M. C., van der Zeisjst, B. A., and Spaan, W. J. (1987). Sequence and structure of the coronavirus peplomer protein. *Adv Exp Med Biol* 218, 31-8.

[0076] Delwart, E. L. et al., 1990 "Retroviral envelope glycoproteins contain a 'leucine zipper'-like repeat," AIDS Research and Human Retroviruses 6:703-706.

[0077] Drosten, et al. "Identification of a Novel Coronavirus in Patients with Severe Acute Respiratory Syndrome,"*N Engl J Med* 2003 Apr. 10.

[0078] Eckert, D. M. and Kim, P. S. "Mechanisms of viral membrane fusion and its inhibition,"*Annu Rev Biochem.* 2001;70:777-810.

[0079] Eckert, D. M. and Kim, P. S. "Design of potent inhibitors of HIV-1 entry from the gp41 N-peptide region,"*Proc Natl Acad Sci USA.* 2001 Sep. 25;98(20):11187-92.

[0080] Eckert, D. M., et al. "Inhibiting HIV-1 entry: discovery of D-peptide inhibitors that target the gp41 coiled-coil pocket,"*Cell* 1999 Oct. 1;99(1):103-15.

[0081] Gallaher, W., Henderson, L., Fermin, C., Montelaro, R., Martin, A., Qureshi, M., Ball, J., Sattentau, Q., Luo-Zhang, H., and Garry, R. (1992a). Membrane interactions of human immunodeficiency virus: Attachment, fusion and cytopathology. In "Membrane Interactions of HIV" (R. Aloia, Ed.), Vol. 6, pp. 113-142. Wiley-Liss, Inc., NY.

[0082] Gallaher, W. R. 1987 "Detection of a fusion peptide sequence in the transmembrane protein of human immunodeficiency virus,"*Cell* 50, 327-328.

[0083] Gallaher, W. R. 1996 "Similar structural models of the transmembrane proteins of Ebola and Avian sarcoma viruses,"*Cell* 85: 477-478.

[0084] Gallaher, W. R., Ball, J. M., Garry, R. F., Griffin, M. C., and Montelaro, R. C 1989 "A general model for the transmembrane proteins of H1V and other retroviruses,"*AIDS Research and Human Retroviruses* 5, 431-440.

[0085] Gallaher, W. R., de Simone, C. and Buchmeier, M. (2001). The viral transmembrane superfamily: possible divergence of arenavirus and filovirus glycoproteins from a common RNA virus ancestor. *BMC Microbiol* 1, 1.

[0086] Gallaher, W., Fermin, C., Henderson, L., Montelaro, R., Martin, A., Qureshi, M., Ball, J., Luo-Zhang, H., and Garry, R. (1992b). Membrane interactions of HIV: Attachment, fusion and cytopathology. *Adv Membrane Fluidity* 6, 113-42.

[0087] Garry, R. F. and Dash, S. (2003) Proteomics computational analyses suggest that hepatitis C virus E1 and pestivirus E2 envelope glycoproteins are truncated class II fusion proteins. *Virology* 307, 255-65.

[0088] Giannecchini, et al., 2003 "Antiviral Activity and Conoformational Features of an Octapeptide Derived from the Membrane-Proximal Ectodomain of the Feline Immunodeficiency Virus Transmembrane Ectodomain,"*J. Virol.* 77:3724-3733.

[0089] Guan, Y., Peiris, J. S., Zheng, B., Poon, L. L., Chan, K. H., Zeng, F. Y., Chan, C. W., Chan, M. N., Chen, J. D., Chow, K. Y., Hon, C. C., Hui, K. H., Li, J., Li, V. Y., Wang, Y., Leung, S. W., Yuen, K. Y., and Leung, F. C. (2004). Molecular epidemiology of the novel coronavirus that causes severe acute respiratory syndrome. *Lancet* 363, 99-104.

[0090] Guan, Y., Zheng, B. J., He, Y. Q., Liu, X. L., Zhuang, Z. X., Cheung, C. L., Luo, S. W., Li, P. H., Zhang, L. J., Guan, Y. J., Butt, K. M., Wong, K. L., Chan, K. W., Lim, W., Shortridge, K. F., Yuen, K. Y., Peiris, J. S., and Poon, L. L. (2003). Isolation and characterization of viruses related to the SARS coronavirus from animals in southern China. *Science* 302, 276-278.

[0091] Gutte, B. (ed.), Peptides: Synthesis, Structure and Application, Academic Press, San Diego (1995).

[0092] Haff, R. F. (1962) Plaque formation by a mouse hepatitis virus. Virology 18, 507-508.

[0093] Hoesl C. E., Nefzi A., Ostresh J. M., Yu Y., and Houghten, R. A. Mixture-based combinatorial libraries: from peptides and peptidomimetics to small molecule acyclic and heterocyclic compounds. *Methods Enzymol.* 2003;369:496-517.

[0094] Hsu, M. C. et al. 1981 "Activation of the Sendai virus Fusion protein (F) involves a conformational change with exposure of a new amino terminus,"*Virology* 104, 294-302.

[0095] Jemmerson "Effects of Conformation, Amino Acid Sequence, and Carrier Coupling on the Immunological Recognition of Peptide and Protein Antigens" in: Zegers et al., *Immunological Recognition of Peptides in Medicine and Biology* (New York, CRC, 1995), pp. 213-225.

[0096] Kay, B., Winter, J., and McCafferty, J. Phage Display of Peptides and Proteins: A Laboratory Manual, Academic Press; 1st edition (Jan. 15, 1996).

[0097] Kilby, J. M., Hopkins, S., Venetta, T. M., DiMassimo, B., Cloud, G. A., Lee, J. Y., Alldredge, L., Hunter, E., Lambert, D., Bolognesi, D., Matthews, T., Johnson, M. R., Nowak, M. A., Shaw, G. M., and Saag, M. S. (1998). Potent suppression of HIV-1 replication in humans by t-20, a peptide inhibitor of gp41-mediated virus entry. *Nat Med* 4, 1302-7

[0098] Kowalski, M., et al. 1987 "Functional Regions of the envelope glycoprotein of human immunodeficiency virus Type 1,"*Science* 237, 1351-1355.

[0099] Ksiazek, et al. "A Novel Coronavirus Associated with Severe Acute Respiratory Syndrome,"*N Engl J Med* 2003 Apr. 10.

[0100] Lambert, D. M., et al. "Peptides from conserved regions of paramyxovirus fusion (F) proteins are potent

inhibitors of viral fusion,"*Proc Natl Acad Sci USA.* 1996 Mar. 5;93(5): 2186-91.

[0101] Li, W., Moore, M. J., Vasilieva, N., Sui, J., Wong, S. K., and Berne, M. A., Somasundaran, M., Sullivan, J. L., Luzuriaga, K., Greenough, T. C., Choe, H., and Farzan, M. (2003). Angiotensin-converting enzyme 2 is a functional receptor for the SARS coronavirus. *Nature* 426, 450-454.

[0102] Liu, S., Xiao, G., Chen, Y., He, Y., Niu, J., Escalante, C. R., Xiong, H., Farmar, J., Debnath, A. K., Tien, P., and Jiang, S. (2004). Interaction between heptad repeat 1 and 2 regions in spike protein of SARS-associated coronavirus: implications for virus fusogenic mechanism and identification of fusion inhibitors. *The Lancet* Vol. 363, pp. 938-940.

[0103] Luo, Z., Matthews, A. M., and Weiss, S. R. (1999). Amino Acid Substitutions within the Leucine Zipper Domain of the Murine Coronavirus Spike Protein Cause Defects in Oligomerization and the Ability to Induce Cell-to-Cell Fusion. *J. Virol.* 73: 8152-8159.

[0104] Marra, M. A., Jones, S. J., Astell, C. R., Holt, R. A., Brooks-Wilson, A., Butterfield, Y. S., Khattra, J., Asano, J. K., Barber, S. A., Chan, S. Y., Cloutier, A., Coughlin, S. M., Freeman, D., Girn, N., Griffith, O. L., Leach, S. R., Mayo, M., McDonald, H., Montgomery, S. B., Pandoh, P. K., Petrescu, A. S., Robertson, A. G., Schein, J. E., Siddiqui, A., Smailus, D. E., Stott, J. M., Yang, G. S., Plummer, F., Andonov, A., Artsob, H., Bastien, N., Bernard, K., Booth, T. F., Bowness, D., Drebot, M., Fernando, L., Flick, R., Garbutt, M., Gray, M., Grolla, A., Jones, S., Feldmann, H., Meyers, A., Kabani, A., Li, Y., Normand, S., Stroher, U., Tipples, G. A., Tyler, S., Vogrig, R., Ward, D., Watson, B., Brunham, R. C., Krajden, M., Petric, M., Skowronski, D. M., Upton, C., and Roper, R. L. (2003). The genome sequence of the SARS associated coronavirus. *Science* 300, 1399-1404.

[0105] Mayer L. D., Hope M. J., Cullis P. R. Vesicles of variable sizes produced by a rapid extrusion procedure. *Biochim Biophys Acta.* 1986 Jun. 13;858(1): 161-8.

[0106] Medinas, R. J., et al. "C-Terminal gp40 peptide analogs inhibit feline immunodeficiency virus: cell fusion and virus spread."*J. Virol.* 2002 September; 76(18): 9079-86.

[0107] Miller, F. A., et al. 1968 "Antiviral activity of carbobenzoxy di- and tripeptides on Measles virus, "*Applied Microbiology* 16, 1489-1496.

[0108] Muster, T., Steindl, F., Purtscher, M., Trkola, A., Klima, A., Himmler, G., Ruker, F., and Katinger, H. (1993). A conserved neutralizing epitope on gp41 of human immunodeficiency virus type 1. *J Virol* 67, 6642-7.

[0109] Muster, T., Guinea, R., Trkola, A., Purtscher, M., Klima, A., Steindl, F., Palese, P., and Katinger, H. (1994). Cross-neutralizing activity against divergent human immunodeficiency virus type 1 isolates induced by the gp41 sequence ELDKWAS. *J Virol* 68, 4031-4.

[0110] Nash, T. C. and Buchmeier M. J. (1997). Entry of mouse hepatitis virus into cells by endosomal and nonendosomal pathways. *Virology* 233, 1-8.

[0111] Nicolaides, E., et al. 1968. "Potential antiviral agents. Carbobenzoxy di- and tripeptides active against Measles and herpes viruses,"*Journal of Medicinal Chemistry* 11, 74-79.

[0112] Nieva, J. L., Nir, S., Muga, A., Goni, F. M., and Wilschut, J. (1994). Interaction of the HIV-1 fusion peptide with phospholipid vesicles: different structural requirements for fusion and leakage. *Biochemistry* 33, 3201-9

[0113] Njenga M. K., Lwamba H. M., and Seal B. S. *Virus Res.* Metapneumoviruses in birds and humans. 2003 February; 91(2): 163-9.

[0114] Norrby, E. 1971 "The effect of a carbobenzoxy tripeptides on the biological Activities of measles virus, "*Virology* 44, 599-608.

[0115] Peiris, J. S., Lai, S. T., Poon, L. L., Guan, Y., Yam, L. Y., Lim, W., Nicholls, J., Yee, W. K., Yan, W. W., Cheung, M. T., Cheng, V. C., Chan, K. H., Tsang, D. N., Yung, R. W., Ng, T. K., and Yuen, K. Y. (2003). Coronavirus as a possible cause of severe acute respiratory syndrome. Lancet 361, 1319-25.

[0116] Peret, T. C., et al. "Characterization of human metapneumoviruses isolated from patients in North America,"*J Infect Dis.* 2002 Jun. 1;185(11): 1660-3.

[0117] Pinon, et al., 2003 "An Antiviral Peptide Targets a Coiled-Coil Domain of the Human T-Cell Leukemia Virus Envelope Glycoprotein,"*J. Virol.* 77:3281-3290.

[0118] Poutanen, S. M., et al. "Identification of Severe Acute Respiratory Syndrome in Canada,"*N Engl J Med* 2003 Mar31.

[0119] Rausch J. M. and Wimley W. C. (2001). A high-throughput screen for transmembrane pore-forming peptides. *Analytical Biochemistry* 293:258-63.

[0120] Raussens V, Slupsky C M, Sykes B D, Ryan R O. Lipid-bound structure of an apolipoprotein E-derived peptide. J. Biol. Chem. 2003 Jul. 11;278(28): 25998-6006. Epub 2003 Apr. 22.

[0121] Richard, J. P., Melikov, K., Vives, E., Ramos, C., Verbeure, B., Gait, M. J., Chemomordik, L. V., and Lebleu, B. (2003). Cell-penetrating Peptides A Reevaluation of the Mechanism of Cellular Uptake. *The Journal of Biological Chemistry* Vol. 278, No. 1, Issue of Jan. 3, pp. 585-590.

[0122] Richardson, C. D. and Choppin, P. W. 1983 "Oligopeptides that specifically Inhibit membrane fusion by paramyxoviruses: studies on the site of action,"*Virology* 131, 518-532.

[0123] Richardson, C. D., et al. 1980 "Specific inhibition of Paramyxovirus and myxovirus replication by oligopeptides and amino acid Sequences similar to those at the N-termini of the Fl or HA2 viral polpeptides,"*Virology* 105, 205-222.

[0124] Rimsky, et al., 1998 "Determinants of Human Immunodeficiency Virus type 1 Resistance to gp41-derived Inhibitory Peptides,"*J. Virol.* 72:986-993.

[0125] Root, M. J., et al. "Protein design of an HIV-1 entry inhibitor,"*Science.* 2001 Feb. 2;291(5505): 884-8.

[0126] Rota P. A., Oberste M. S., Monroe S. S., Nix W. A., Campagnoli R., Icenogle J. P., Penaranda S., Bankamp B., Maher K., Chen M. H., Tong S., Tamin A., Lowe L., Frace M., DeRisi J. L., Chen Q., Wang D., Erdman D. D., Peret T. C., Burns C., Ksiazek T. G., Rollin P. E., Sanchez A., Liffick S., Holloway B., Limor J., McCaustland K., Olsen-Rasmussen M., Fouchier R., Gunther S., Osterhaus A. D., Drosten C., Pallansch M. A., Anderson L. J., Bellini W. J. Characterization of a novel coronavirus associated with severe acute respiratory syndrome. Science. 2003 May 30.

[0127] Sia, S. K., et al. "Short constrained peptides that inhibit HIV-1 entry,"*Proc Natl Acad Sci USA*. 2002 Nov. 12;99(23): 14664-9.

[0128] Simmons, G., Amberg, S., Rennekamp, A. J., Piefer, A., and Bates, P. (2004). Keystone Symposium on Bioterrorism and Emerging Infectious Diseases. Abstract 215, p120.

[0129] Spiegelberg H. L., et al. (1997) "DNA immunization: a novel approach to allergen-specific immunotherapy", *Allergy* 52:964-70.

[0130] Stiegler, G., Armbruster, C., Vcelar, B., Stoiber, H., Kunert, R., Michael, N. L., Jagodzinski, L. L., Ammann, C., Jager, W., Jacobson, J., Vetter, N., and Katinger, H. (2002). Antiviral activity of the neutralizing antibodies 2F5 and 2G12 in asymptomatic HIV-1 infected humans: a phase I evaluation. *AIDS* 16, 2019-25.

[0131] Suarez, et al., 2000 "Membrane Interface-Interacting Sequences within the Ectodomain of the Human Immunodeficiency Virus type 1 Envelope Glycoprotein: Putative Role During Viral Fusion,"*J. Virol.* 74:8038-8047.

[0132] Tripet, B., Howard, M. W., Jobling, M., Holmes, R. K., Holmes, K. V., and Hodges, R. S. (2004). Structural Characterization of the SARS-Coronavirus Spike S Fusion Protein Core. JBC Papers in Press. Manuscript M400759200.

[0133] Tsai, J. C., de Groot, L., Pinon, J. D., Iacono, K. T., Phillips, J. J., Seo, S. H., Lavi, E., and Weiss, S. R. (2003). Amino acid substitutions within the heptad repeat domain 1 of murine coronavirus spike protein restrict viral antigen spread in the central nervous system. *Virology* 312, 369-380.

[0134] van den Hoogen, B. G., et al. "A newly discovered human pneumovirus isolated from young children with respiratory tract disease,"*Nat Med.* 2001 June; 7(6): 719-24.

[0135] van den Hoogen, B. G., et al. "Analysis of the genomic sequence of a human metapneumovirus,"*Virology* 2002 Mar. 30;295(1): 119-32.

[0136] Wang, Y., Ma, W. L., Song, Y. B., Xiao, W. W., Zhang, B., Huang, H., Wang, H. M., Ma, X. D., and Zheng, W. L. (2003). Gene sequence analysis of SARS-associated coronavirus by nested RT-PCR. *Di Yi Jun Yi Da Xue Xue Bao* 23, 421-3

[0137] White, J. M., "Membrane Fusion,"*Science*, vol. 258 (Nov. 6, 1992), pp. 917-924.

[0138] White, S. H., Wimley, W. C., Ladokhin, A. S., and Hristova, K. (1998) Protein folding in membranes: determining energetics of peptide-bilayer interactions. Methods Enzymol 295, 62-87.

[0139] Wild, et al. 1992 "A synthetic peptide inhibitor of human immunodeficiency virus replication: Correlation between solution structure and viral inhibition, "*Proc. Natl. Acad. Sci. USA* 89:10537-10541.

[0140] Wild, et al., 1994 "Propensity for a Leucine Zipper-Like Domain of Human Immunodeficiency Virus Type 1 gp41 to Form Oligomers Correlates With a Role in Virus-Induced Fusion Rather Than Assembly of the Glycoprotein Complex,"*Proc. Natl. Acad. Sci. USA* 91:12676-80.

[0141] Wimley, W. C., Selsted, M. E., and White, S. H. (1994). Interactions between human defensins and lipid bilayers; evidence for formation of multimeric pores. *Protein Sci* 3, 1362-73.

[0142] Wimley, W. C., and White, S. H. (2000a). Designing transmembrane alpha-helices that insert spontaneously. Biochemistry 39, 4432-42.

[0143] Wimley, W. C. and White, S. H. (2000b). Determining the membrane topology of peptides by fluorescence quenching. *Biochemistry* 39, 161-70.

[0144] Wimley, W. C., and White, S. H. (1996) *Nat Struct Biol* 3, 842-848.

[0145] Wilson, I. A., et al. 1981. "Structure of the haemagglutinin Membrane glycoprotein of influenza virus at 3 A resolution,"*Nature* 289, 366-373.

[0146] Yao, W. M., Wimley, W. C., Gaurisch, K and White S. H. (1998) The preference of tryptophan for membrane interfaces. *Biochemistry.* 37(42): 14713-8.

[0147] Zelus, B. D., Schickli, J. H., Blau, D. M., Weiss, S. R., and Holmes, K. V. (2003). Conformational changes in the spike glycoprotein of murine Coronaviruses are induced at 37 degrees C. either by soluble murine CEACAMI receptors or by pH 8. *J Virol* 77, 830-40.

[0148] Zheng, B. J., Guan, Y., Wong, K. H., Zhou, J., Wong, K. L., Young, B. W. Y., Lu, L. W., and Lee, S. S. (2004). SARS-related virus predating SARS outbreak, Hong Kong. *Emerging Infectious Diseases*. e-pub Jan. 16, 2004.

DETAILED DESCRIPTION OF THE INVENTION

[0149] For convenience in the ensuing description, the following explanations of terms are adopted. However, these explanations are intended to be exemplary only. They are not intended to limit the terms as they are described or referred to throughout the specification. Rather, these explanations are meant to include any additional aspects and/or examples of the terms as described and claimed herein.

[0150] As used herein, the terms "inhibiting,""inhibition, ""inhibitory," and any variants thereof are to be understood as meaning (with respect to the activity of the peptides) inhibition both in a prophylactic sense (i.e., prevention of the initial transmission of the virus to an individual), as well as

in the sense of preventing the infection from becoming established or ameliorating its effects once the virus has been introduced into the body.

[0151] As used herein, the term "analogue" means a peptide or peptidomimetic compound that has the same amino acid sequence as a segment of the viral membrane glycoprotein, or is designed to mimic the stereochemical shape of a portion of the viral membrane glycoprotein.

[0152] Also, in this regard, it is contemplated that the term "amino acid" as used herein refers to both naturally occurring forms, as well as synthetic forms which have been modified by the addition of side chains or other moieties to increase solubility, biological half-life or uptake and delivery to body tissues. Both D- and L-forms of all amino acids are also contemplated, in any form including their pharmacologically acceptable salts.

[0153] In one embodiment of the present invention, analogues of a portion of the fusion glycoproteins of human CoV and human MPV are employed to inhibit the normal fusion process of the viruses in vivo. In certain aspects, the portion of the fusion glycoprotein for which these analogues have been designed is the "charged pre-insertion helix" (CPI helix). The CPI helix is that portion of the fusion glycoprotein which lies within about 100 amino acids from the point at which the fusion glycoprotein is anchored within the lipid membrane of the virus and which is characterized by a high percentage of hydrophilic amino acids that may be acidic or basic in nature and that have a recognizable propensity to form an alpha helix. As discussed above, CPI helices have been shown in a number of virus systems to be involved in the induction of cell fusion, and, in some cases, analogues of those portions have been shown to inhibit fusion. The CPI helix of a virus fusion glycoprotein may be located using the following method: First, the primary amino acid sequence of the virus entry glycoprotein, toward the carboxy terminus of the virus entry glycoprotein, is examined for a uniformly hydrophobic (i.e., consisting entirely of hydrophobic amino acids, to the exclusion of hydrophilic amino acids) sequence of about 20-25 amino acids, which uniformly hydrophobic sequence has a propensity to span the lipid envelope membrane. The membrane-spanning portion has been found to be composed of more than about 60% aliphatic and aromatic amino acids in virtually all membrane spanning glycoproteins. The 100 amino acid region preceding this membrane-spanning portion is examined for charged amino acids as well as for amino acids such as glutamine (Q), glutamate (E), alanine (A), tryptophane (W), lysine (K) and leucine (L), which have a known propensity to form an alpha helix. While the core of the CPI helix is evident by finding a concentration of such amino acids as have a strong helical propensity, the beginning of the helix is found by locating a di- or tri-peptide motif that has a propensity to "nucleate" or start the helix formation. Generally, this constitutes a pair of amino acids together which each strongly favor a helix, such as glutamate (E), glutamine (Q), phenylalanine (F), lysine (K), alanine (A), or leucine (L). This is even more strongly favored when preceded by a proline (P), particularly when no more than 2 or 3 amino acids separate the P from the di- or tri-peptide motif. For example, in the CPI helix of the SARS CoV, the sequence PEL [SEQ ID NO: 32] comprises such a nucleation motif. In MHV, a comparable nucleation motif is PDFKE [SEQ ID NO: 33]. Once the CPI helix is identified, peptide analogues of the sequence of the CPI helix can be tested for their ability to inhibit virus-induced cell fusion and viral infectivity.

[0154] In one embodiment, the present invention comprises peptides which represent analogues of the CPI helix from human metapneumovirus and the CPI helix from human coronavirus. Overall, the CPI helix of each virus entry glycoprotein is between 50 and 80 amino acids in length. Synthesis and production of peptides of this length are impracticable, due to limitations in efficiency of synthesis or purity. Therefore, peptide analogues are generally limited in practice to shorter peptides over a shorter span of the glycoproteins which are effectively inhibitory at a concentration useful for human administration. This necessarily varies with each virus system and protein portion due to variations in amino acid sequence.

[0155] Peptides of as few as 6 amino acids or as many as 40 may provide the optimal combination of factors in development of an inhibitory peptide into a human drug. When the CPI helix has been located, it is desirable to delineate subsets of the amino acid sequence of the CPI helix which will represent inhibitory-effective peptides themselves, and together represent the best set of such peptides from the entire CPI helix. One method is to divide the entire CPI helix sequence into three segments representing about the first, second, and last third of the amino acid sequence of the CPI helix, while initiating and ending each segment with certain preferred amino acids. In general, alanine (A), glutamate (E), glutamine (Q), tyrosine (Y), phenylalanine (F), lysine (K) and proline (P) are favored as termini, and longer chain aliphatic amino acids such as valine (V), isoleucine (I) and leucine (L) are disfavored. A second, complimentary method involves centering peptides on those areas which are highly conserved in sequence among class I viral fusion glycoproteins. An example is shown in **FIG. 10**, which contains a comparison of the amino acid sequence of the CPI helices of human coronavirus OC43, MHV A59, and SARS CoV. Asterisks denote the identical amino acids in all three viruses, indicating a strong presumption of constancy in structure and function for those regions with a concentration of asterisks. Inhibitory effective peptides may be constructed which center on those sequences and are of decreasing lengths. Using human CoV as an example, the minimum peptide length is likely to be FKEELDK [SEQ ID NO: 34] or KWPWYVWL [SEQ ID NO: 35], the heptamer and octamer that coincide to the constant sequences at either end of the CPI helical region in HIV, MHV, and human CoV. Additional amino acids may be added to either the amino- or carboxy-termini of these conserved peptide sequences to enhance the biological and pharmacological properties of peptides used for treatment of humans using methods known to those practiced in the pharmaceutical arts. It will be apparent to those skilled in the art that other methods may be used to locate inhibitory effective peptide analogues of the amino acid CPI helix, such as screening of overlapping peptides, molecular modeling, and algorithms that utilize the Wimley-White interfacial hydrophobicity scale.

[0156] In one embodiment, inhibitory peptides are stipulated for human MPV and human CoV that range in length from 6 to 40 amino acids in length. Peptides are constructed to represent different segments of the CPI helices of these viruses that may be efficiently synthesized and inhibitory effective when used either alone or in combination.

[0157] In the case of human MPV, the CPI helix comprises the following 67 amino acids:

YQLSKVEGEQHVIKGRPVSSSFDPIKFPEDQFNV [SEQ ID NO: 01]
ALDQVFESIENSQALVDQSNKILNSAEKGNTGF.

[0158] This sequence has been subdivided into 8 peptides [SEQ ID NOS: 3-9 and 36] that overlap different portions of the CPI helix amino acid sequence, as shown in **FIG. 11**. Any one peptide, or combination of peptides, may be used as an analogue(s) of this virus fusion glycoprotein so as to inhibit the natural interactions of this protein portion in inducing membrane fusion.

[0159] In one embodiment, the present invention comprises the following peptide analogue of the CPI helix of human MPV:

EDQFNVALDQVFESIENSQA [SEQ ID NO: 07]
LVDQSNKILNSAEKGNTGF.

[0160] This embodiment contains the maximum percentage of those amino acids, as discussed above, that define the CPI helix (i.e., Q, E, A, W, K and L), and, therefore, this analogue is predicted to be maximally active in competitively inhibiting fusion.

[0161] The minimum inhibitory effective peptide in the case of human MPV is the following hexapeptide.

QALVDQ. [SEQ ID NO: 36]

[0162] Addition of any number of amino acids to either the amino terminus or carboxy terminus of this minimum peptide should not affect its inhibitory potential, but should have the effect of rendering the peptide more desirable for pharmaceutical use in humans.

[0163] In the case of the human SARS CoV, the CPI helix comprises the following 78 amino acid sequence:

PELDSFKEELDKYFKNHTSPDVDLGDISGINASV [SEQ ID NO: 02]
VNIQKEIDRLNEVAKNLNESLIDLQELGKYEQYI
KWPWYVWLGF.

[0164] This region overlies two separate regions that meet the definition of a CPI helix, bridged by a region of lower charge density which is predicted to have a lower helicity. As in the case of human MPV, it is impracticable to synthesize or purify a peptide of this length. Therefore, 12 peptides [SEQ ID NOS: 10-19, 34 and 35] derived from this overall sequence (as shown in **FIG. 12**) are presented. These embodiments are to be used singly or in combination to be maximally inhibitory effective.

[0165] The following embodiment comprises a 36 amino acid peptide derived from the carboxy-terminal region of the amino acid sequence of the CPI helix which overlaps the abnormally high concentration of aromatic amino acids such as tyrosine (Y) and tryptophane (W), which have been shown to be especially active in viral fusion proteins to induce membrane fusion:

RLNEVAKNLNESLIDLQEL [SEQ ID NO: 12]
GKYEQYIKWPWYVWLGF.

[0166] Fragments of this peptide are predicted to have inhibitory effective activity, such that a sequence of as few as 10 amino acids, i.e.:

YIKWPWYVWL, [SEQ ID NO: 18]

[0167] is predicted to yield sufficient inhibition to be effective and, at the same time, enhance ease of preparation and purification. However, the minimum effectively inhibitory peptide in the case of human CoV is either the conserved heptapeptide:

FKEELDK [SEQ ID NO: 34]

[0168] or the conserved octapeptide:

KWPWYVWL, [SEQ ID NO: 35]

[0169] or a combination of the two.

[0170] Each peptide has a unique and relatively poorly predictable behavior in solution. This behavior is dependent not only on the choice of the amino acid sequence, but also on the selection of molecular adducts (which could be added to the amino-terminal end and/or the carboxy-terminal end) such as any of several known to those practiced in the art useful for rendering peptides increasingly soluble, resistant to proteases, or otherwise improving their bioavailability and appropriate configuration. Desirable properties may be imparted, or undesirable properties ameliorated, by addition of adducts at either end of the proposed amino acid sequences in a manner known to those practiced in the peptide synthetic or pharmaceutical arts for development of peptide reagents for use in humans. For example, in certain embodiments of the present invention, the peptide acetyl PEQLK [SEQ ID NO: 37] is used as one of the adducts at the beginning of the peptide sequences. This addition is designed to begin the forming of (i.e., "nucleate") the helix structure-which, once begun, will continue. The use of this additive will produce the proper helical configuration even in shorter sequences. For example, as in the following sequence:

P
E
Q
L
K--- [SEQ ID NO: 37]

[0171] The P starts with a kink due to its ring structure, the E and Q are of high helical propensity, the L interacts with

P, and the E reacts with K—all of which contribute to helix formation (Bodansky, M., Bodansky, A., The practice of peptide synthesis (2nd edn.), Springer Verlag, Berlin (1995); Gutte, B. (ed.), Peptides: Synthesis, Structure and Application, Academic Press, San Diego (1995), each of which is hereby incorporated by reference herein in its entirety).

[0172] The peptides of the present invention may be readily prepared by any of a wide range of methods known in the art, either manually or automated, while the synthetic peptide is immobilized on a solid substrate (examples can be seen in Eckert, D. M. and Kim, P. S. "Design of potent inhibitors of HIV-1 entry from the gp41 N-peptide region-."*Proc Natl Acad Sci U S A.* 2001 Sep. 25;98(20): 11187-92.; Giannecchini et al., 2003, "Antiviral Activity and Conformational Features of an Octapeptide Derived from the Membrane-Proximal Ectodomain of the Feline Immunodeficiency Virus Transmembrane Ectodomain", *J. Virol.* 77:3724-3733; Jemmerson "Effects of Conformation, Amino Acid Sequence, and Carrier Coupling on the Immunological Recognition of Peptide and Protein Antigens" in: Zegers et al., Immunological Recognition of Peptides in Medicine and Biology (New York, CRC, 1995), pp. 213-225, each of which is hereby incorporated by reference herein in its entirety). It is anticipated that reactive side groups of the amino acids will be blocked chemically during synthesis and unblocked when synthesis is completed using methods well known to the skilled artisan. Typically, the final peptide products will be acetylated at the amino-terminal end, and amidated at the carboxy-terminal end, to increase biological half-life. Further, a D-amino acid may be interposed or added at the termini to further reduce susceptibility of the peptide to exoprotease activity in biological fluids. Any of such known methods is suitable for the present purpose.

[0173] Alternately, certain of the peptides of the present invention, especially the longer sequences (such as SEQ ID NO: 07 and SEQ ID NO: 12) may be synthesized from a genetic construct of deoxyribonucleic acid (DNA) (either synthetic or derived by duplication from the respective viral genome) that is linked to a DNA "expression vector" suitable for production of the peptide by natural or in vitro protein synthesis in a prokaryotic or eukaryotic system. A variety of expression vectors are known to those practiced in the genetic arts, and many are under continual development for a variety of genetic production methods (Kay, B., Winter, J., and McCafferty, J. Phage Display of Peptides and Proteins: A Laboratory Manual, Academic Press; 1st edition (Jan. 15, 1996), which is incorporated by reference herein in its entirety). In addition to use of DNA constructs for synthetic purposes, a contemplated application of this invention is expression of inhibitory effective peptides as a form of "gene therapy" through the administration of DNA to a human patient in lieu of the peptide itself. One example of the use of expressible DNA constructs in lieu of proteins or peptides is in immunization by injection of DNA currently under development (see Spiegelberg H L, et al. (1997) "DNA immunization: a novel approach to allergen-specific immunotherapy", *Allergy* 52:964-70, which is incorporated herein by reference in its entirety).

[0174] It is contemplated that the peptides may be used singly or in combination, either with one another or with other pharmaceuticals as may be found to be compatible or synergistic. Examples of such pharmaceuticals include, but are not limited to, immune modulators such as interferon, anti-inflammatory drugs such as corticosteroids, other classes of antiviral drugs such as nucleoside analogues, or antibiotics such as erythromycin.

[0175] The peptides of the present invention may also be covalently linked, either via disulfide bridges or other chemical linkages, to each other or to macromolecular carrier molecules of desirable specificity. For example, the peptides may be linked or adsorbed to lipoproteins to facilitate their uptake into endosomal vesicles within cells as a form of biological targeting that may positively affect their efficacy. (See generally, Richard et al. (2003). Cell-penetrating Peptides A Reevaluation of the Mechanism of Cellular Uptake. *The Journal of Biological Chemistry Vol.* 278, No. 1, Issue of Jan. 3, pp. 585-590, which is hereby incorporated by reference herein in its entirety). Coronaviruses are known to enter cells either through direct fusion at the cell surface or via the process of endocytosis (Nash, T. C. and Buchmeier M. J. (1997). Entry of mouse hepatitis virus into cells by endosomal and nonendosomal pathways. *Virology* 233, 1-8; Tsai et al. (2003). Amino acid substitutions within the heptad repeat domain 1 of murine coronavirus spike protein restrict viral antigen spread in the central nervous system. *Virology* 312, 369-380, each of which is hereby incorporated by reference herein in its entirety). Recent studies presented at the Keystone Symposium on Bioterrorism and Emerging Infectious Diseases indicate that SARS CoV enters via endocytosis (Simmons et al. (2004). Keystone Symposium on Bioterrorism and Emerging Infectious Diseases. Abstract 215, p120, which is hereby incorporated by reference herein in its entirety) or perhaps by utilizing both cell surface and endocytic pathways as is the case with certain strains of MHV (Nash, T. C. and Buchmeier M. J. (1997). Entry of mouse hepatitis virus into cells by endosomal and nonendosomal pathways. *Virology* 233, 1-8; Tsai et al. (2003). Amino acid substitutions within the heptad repeat domain 1 of murine coronavirus spike protein restrict viral antigen spread in the central nervous system. *Virology* 312, 369-380, each of which is hereby incorporated by reference herein in its entirety). Therefore, modifications that enhance uptake of inhibitor peptides into endosomal vesicles may further increase effectiveness of the SARS CoV fusion inhibitory peptides. Certain peptides, such as Antennapedia and pestivirus E_{rns} (Garry, R. F. and Dash, S. (2003) Proteomics computational analyses suggest that hepatitis C virus E1 and pestivirus E2 envelope glycoproteins are truncated class II fusion proteins. *Virology* 307, 255-65, which is hereby incorporated by reference herein in its entirety) can enter cells by direct penetration of the plasma membrane. However, convincing evidence has been presented that peptides containing HIV-1 Tat amino acids 48-60 (GRKKRRQRRRP [SEQ ID NO: 38]) or polyarginine (7-9 arginines) enter cells primarily via the endocytic route (see Richard, J. P. et al. (2003). Cell-penetrating Peptides A Reevaluation of the Mechanism of Cellular Uptake. *The Journal of Biological Chemistry* Vol. 278, No. 1, Issue of Jan. 3, pp. 585-590, which is hereby incorporated by reference herein in its entirety). Similarly, the endosomal targeting peptide region of apolipoprotein E has been identified (Raussens, V. et al. Lipid-bound structure of an apolipoprotein E-derived peptide. *J. Biol. Chem.* 2003 Jul. 11;278(28): 25998-6006. Epub 2003 Apr. 22, which is hereby incorporated by reference herein in its entirety). These studies have caused a stronger consideration of the use of targeting peptide sequences, such

as those found in HIV-1 Tat and human apolipoprotein E for endosomal targeting of peptide inhibitors of SARS CoV infection (see ibid.).

[0176] To determine if these endosome targeting sequences improve the efficiency of fusion inhibition of the peptides of the present invention, lead peptides will be synthesized with Tat48-60, (Arg)9 or an apolipoprotein E-derived endosomal targeting peptide (ELRVRLASHLRKLRKRLLRDADD [SEQ ID NO: 39]) at the amino or carboxyl terminus. Distribution of the modified and unmodified peptides after conjugation to Alexa Fluor 488 (spectral characteristics similar to fluorescein—excitation at 495 nm and emission at 519 nm—but produces conjugates that are brighter, more photostable, and insensitive to pH from 4 to 10) may be assessed by confocal microscopy using appropriate cell compartment tags, such as Lysotracker Red (Molecular Probes) ("Probes for Following Receptor Binding, Endocytosis and Exocytosis." Molecular Probes Handbook., Molecular Probes, Inc., Eugene Oreg. <http://www.probes.com/handbook/sections/1601.html>; "Alexa Fluor Dyes: Simply the Best." Molecular Probes Handbook. Molecular Probes, Inc., Eugene Oreg. <http://www.probes.com/handbook/sections/0103.html>; New Probes for Cell Tracing." Molecular Probes Handbook. Molecular Probes, Inc., Eugene Oreg. <http://www.probes.com/lit/bioprobes25/part10.html>; Arttamangkul S, Alvarez-Maubecin V, Thomas G, Williams J T, Grandy D K. Binding and internalization of fluorescent opioid peptide conjugates in living cells. Mol Pharmacol. 2000 December; 58(6): 1570-80, each of which is hereby incorporated by reference herein in its entirety). It is predicted that the endosome targeted peptides also may inhibit CoV fusion at reduced concentrations because of increased potency.

[0177] The peptides may be suspended in any of a number of appropriate vehicles, aqueous or non-aqueous, that are pharmaceutically acceptable for human use, such as sterile solution containing other solutes (for example, sufficient saline or glucose to make the solution isotonic and compatible with human administration).

[0178] The peptides may be administered in a number of forms, to some extent depending upon the therapeutic intent. For example, one of the more useful aspects of certain embodiments of the present invention is their use prophylactically to prevent infection in those exposed or likely to be exposed to SARS-infected individuals. The peptides may be applied for either preventive or therapeutic use topically or transdermally, or by inhalation, in the form of ointments, aqueous compositions, including solutions and suspensions, creams, lotions, aerosol sprays, or dusting powders. The peptides may also be prepared and used in suppository form. The methods and applicability of such formulations is well known in the pharmaceutical art. Application of the therapeutic preparations may be to any area of the body through which the virus may be found to transmit the infection on any internal or external surface of the body, as appropriate.

[0179] The peptides may be prepared for oral or parenteral administration. In oral administration, where practicable, capsules or tablets may be prepared with stabilizers, carriers, preservatives or flavors, as is common in pharmaceutical practice. For parenteral administration, i.e., intravenous, intramuscular, subcutaneous or intraperitoneal, the peptides are administered with a pharmaceutically acceptable carrier such as a sterile solution containing other solutes or drugs.

[0180] The required dosage varies with the mode of administration. Based on our preliminary data, it appears that inhibitory effective peptides must achieve a localized concentration of 10-20 nanomolar at the site of infection. In practice, this requires administration of concentrations of peptide in micromolar quantities. Modification of the dosage range may also be dependent on whether the intent is prevention of infection or treatment of an already established infection. Such embodiments are achievable by practice of those skilled in medical arts of prevention and treatment of infectious disease. For example, clinical scientists may determine the concentration of a drug which is attained in a particular bodily fluid, such as serum, when a certain quantity of drug is administered in a certain manner and thereby adjust the dosage to attain a concentration which has been shown to be inhibitory effective in vitro.

[0181] As is known in the art, variations of the designated peptide drugs may be obtained which have superior pharmacological properties, or greater ability to inhibit evolving strains of each virus, by substituting one or more amino acids within the peptide sequence with closely related amino acids. For example, substitutions may be made within the following series of amino acids, grouped by their biochemical character:

[0182] Short side chain—Glycine (G) or Proline (P) or Alanine (A)

[0183] Hydroxylated side chain—Serine (S) or Threonine (T) or Tyrosine (Y)

[0184] Aliphatic side chain—Alanine (A) or Valine (V) or Leucine (L) or Isoleucine (1) or Methionine (M) or Cysteine (C)

[0185] Sulphur-containing side chain—Cysteine (C) or Methionine (M)

[0186] Aromatic side chain—Phenylalanine (F) or Tyrosine (Y) or Tryptophane (W)

[0187] Neutral side chain—Glutamine (Q) or Asparagine (N) or Histidine (H)

[0188] Acidic side chain—Glutamate (E) or Aspartate (D)

[0189] Basic side chain—Lysine (K) or Arginine (R)

[0190] Certain amino acids are in multiple series because they share properties with two groups of amino acids, for example, alanine is a short side chain amino acid, but also in the aliphatic series of hydrophobic side chains. The substitutions listed above are merely examples. It will be readily apparent to those skilled in the art that other substitutions are known which could be used to alter the properties of a peptide.

[0191] As an example, the amino acid sequence RIQDAIK [SEQ ID NO: 40] found in MHV is equivalent in character to the sequence RLNEVAK [SEQ ID NO: 41] in the SARS CoV, with which it may be aligned within the charged pre-insertion helix of the S2 fusion glycoprotein.

[0192] In the case of the shortest peptides of constant sequence, the shape of these peptides is critical for their activity. Such a shape can be mimicked by small organic compounds with covalent bonds that can reproduce the three dimensional shape of the natural peptide. The classic case of

such a compound is penicillin, which mimics the structure of D-alanyl-D-alanine, and thus inhibits the use of that dipeptide in crosslinking bacterial cell walls as its mode of antibacterial action. While not a peptide at all, or manufactured from peptides, such compounds function as antimicrobials by mimicking the structure of peptides. Such compounds, known as peptidomimetics, may be constructed by several methods well known to those practiced in the pharmaceutical art (Hoesl C. E., Nefzi A., Ostresh J. M., Yu Y., and Houghten, R. A. Mixture-based combinatorial libraries: from peptides and peptidomimetics to small molecule acyclic and heterocyclic compounds. *Methods Enzymol.* 2003;369:496-517, which is hereby incorporated by reference herein in its entirety). Peptidomimetics designed or found to reproduce the structure of peptides described herein are intended to be within the scope of this invention.

[**0193**] The sequence and shape of the peptides defined herein can also be used to design mirror images of the peptide that would reproduce the structure of any natural ligand of the peptide. Such mirror image compounds would include peptides that complement the shape with high affinity, or antibodies directed against the peptide sequence and thus reactive with it. An example of a mirror image peptide would be regions within the antiparallel heptad repeat helix (or N-helix) of the SARS CoV, for example: ENQKQIANQFNKAISQIQESL [SEQ ID NO: 42] or KVQDVVNQNAQALNTLVKQL [SEQ ID NO: 43]. These helical sequences are similar in character to the charged pre-insertion helix, such that they would be expected to react and bind with the peptide sequences defined in the invention. Such peptides are intended to be within the scope of this invention.

[**0194**] An example of an antibody defined by an amino acid sequence would be an antibody designed or selected to interact with the highly conserved ELDKY [SEQ ID NO: 30] motif in the coronavirus CPI helix. Such an antibody specificity is known, the human monoclonal antibody 2F5 originally generated in the immune response to human immunodeficiency virus, type 1, which contains a highly similar ELDKW [SEQ ID NO: 31] motif in its CPI helix region. Use of such an antibody, that reacts with CPI helix peptides and is used in lieu of such peptides, is also intended to be within the scope of this invention.

[**0195**] It is a contemplated application of the present invention that peptides be tested initially by testing comparable peptides of animal viruses or less virulent strains of human viruses, and that permanent lines of animal and human cells in culture be used both as host cells for experimental infections, as well as for toxicity testing. Such testing systems prevent the endangerment of personnel by exposure to virulent human pathogenic viruses such as the SARS CoV. Combinations of such testing systems include the OC43 strain of the human CoV in infection of the Vero E2 permanent cell line of African green monkey kidney cells (American Type Culture Collection, Manassas, Va.). Peptides from the comparable CPI helix of OC43 are derived from the region:

PNLPDFKEELDQWFKNQTSVAPDLSLDYINVTLD [SEQ ID NO: 20]

LQVEMNRLQEAIKVLNQSYINLKDIGTYEYYVKW

PWYVW.

[**0196**] Peptide analogues of OC43 corresponding to peptide analogues of human SARS CoV include SEQ ID NOS: 21-26, the relationship of which to SEQ ID NO: 20 is shown in **FIG. 13**.

[**0197**] Briefly, Vero E2 cells are treated with an inhibitory effective concentration of peptide to equilibrate the culture system with solution containing peptide. A solution containing OC43 human coronavirus is then added, in the continued presence of the peptide solution. Comparable mock-treated controls are allowed to be infected normally as a positive control, and uninfected controls are treated with peptide continuously in the absence of virus, as a control for toxicity. Other control cultures are continuously treated with solution containing neither peptide nor virus, as a negative control. The effects of infection are measured both by observation of cellular cytopathology as a result of virus multiplication, as well as by noting the yield of progeny virus by any of a variety of molecular and virological means well known to virologists practiced in the art. Such studies generally follow the prototype of peptide inhibition studies established in studies of influenza and measles viruses (Richardson, C. D. et al. 1980. Specific inhibition of Paramyxovirus and myxovirus replication by oligopeptides and amino acid sequences similar to those at the N-termini of the Fl or HA2 viral polypeptides. Virology 105, 205-222.; Hsu, M. C. et al. 1981. Activation of the Sendai virus Fusion protein (F) involves a conformational change with exposure of a new amino terminus. Virology 104, 294-302.; Richardson, C. D. and Choppin, P. W. 1983. Oligopeptides that specifically Inhibit membrane fusion by paramyxoviruses: studies on the site of action. Virology 131, 518-532, each of which is hereby incorporated by reference herein in its entirety).

[**0198**] Given the reduced cytopathology inherent in the OC43 virus, and the general observation of only limited human disease due to OC43, testing of peptides for human use may include the use of experimental infections of humans with OC43, and its prevention or treatment by inhibitory effective dosages of peptides targeted to the OC43 CPI helix sequence of amino acids. Such testing may yield critical information preparatory to clinical trials utilizing peptide drugs targeted against the more virulent and cytopathogenic SARS CoV. Insofar as viruses similar to or identical with OC43 are responsible for human illness such as the common cold, the peptides of this invention may be useful for prevention or treatment of such mild respiratory infections, either alone or in combination with other antiviral drugs or other medications. It is contemplated that the same variations in formulation or delivery may be utilized as described above for the formulations involving peptides targeted against human metapneumovirus or human SARS coronavirus.

[**0199**] Prior to, in lieu of, or to supplement testing with OC43 coronavirus, animal testing is typically performed in vitro, using an appropriate combination of animal virus and animal cell line, or in vivo, using an appropriate animal host. In the case of coronaviruses, a widely established and useful system is that of the MHV in an established permanent line of mouse cells, L2 (American Type Culture Collection, Manassas, Va.), or in experimental infection of mice. Particularly useful is a cytopathogenic strain of MHV, A59, which has been used to study coronavirus induced cell fusion. The peptide region of the S2 glycoprotein to MHV A59 that is similar to the comparable portion from the

human SARS CoV is the following peptide, which was taken from the CPI helix of MHV A59 S2 glycoprotein:

QDAIKKLNESYINLKEVGTYEMYVKWPWYVW. [SEQ ID NO: 27]

[0200] This model peptide is useful as a "proof of concept" peptide, due to its similarity to the comparable region of the human SARS CoV S2 glycoprotein, and due to the fact that MHV A59 is comparably cytopathic in mouse L2 cells, as the SARS CoV is in human cells. This peptide provides a close parallel system that is innocuous to humans but may be utilized to test the full spectrum of toxicity, bioavailability, stability and optimal dosage of the present invention, without endangerment of humans or restriction of studies to specialized biological safety environments.

[0201] To test the unique properties of each inhibitory effective peptide, additional controls to be tested include peptides of equal length and composition to the peptides of this invention, but with the order of amino acids scrambled in random order. The specificity of each peptide is also contemplated to be tested by testing peptides derived from one virus sequence on other viruses with different sequences. Each sequence is unique to each virus, with considerable variation even among closely related viruses in the same family. Optimal peptides for each virus system vary in their position within the CPI helix sequence motif relative to the membrane-spanning domain. Nevertheless, specificity will be demonstrated by testing irrelevant peptide compositions and sequences.

Examples

[0202] A. Inhibitory Peptides

[0203] Preliminary Studies indicate that peptide inhibitors can be developed for members of the Coronaviridae family of viruses. We have tested synthetic peptides for their ability to inhibit plaque formation by MHV. We have observed that certain peptides inhibit plaque formation by MHV, and we have confirmed these results for selected inhibitory and non-inhibitory peptides. We found that a peptide corresponding to the MHV C-helix having the following sequence:

[0204] RIQDAIKKLNESYINLKEVGTYEMYVKWPWYVWLLI (SEQ ID NO: 52)

[0205] reduced plaque formation by about 40% at a nominal concentration of about 25 μM (see **FIG. 14**). There was also a significant reduction (about 50%) in the average diameter of the plaques. These results suggest that this peptide inhibits both entry and spread of MHV. Similar results with this inhibitory peptide were obtained in two additional independent experiments, with significant plaque inhibition observed at concentrations of as low as 1 μM. These results are unlikely to be explained by non-specific cytotoxic effects of the peptide. Killing the cells would inhibit fusion, but the cells in these studies have normal morphology, indicating they are unlikely to be damaged to an extent that would inhibit them through any toxic effect. Except for the plaques, cells in the monolayers were intact and viable, and the low number of plaques that did grow were similar in size to control plaques. Comparable results, with inhibitory activities in the μM range have been reported with a C-helix peptide (Bosch, B. J. et al. (2003). The coronavirus spike protein is a class I virus fusion protein: structural and functional characterization of the fusion core complex. *J Virol* 77, 8801-8811, which is hereby incorporated by reference herein in its entirety). Preliminary experiments also indicate that these peptides form helical structures in aqueous solution which are responsible for their biological function as inhibitors. **FIG. 15** shows the results of Circular dichroism (CD) spectroscopy used to delineate the structural properties of a peptide corresponding to a region of the S2 protein of MHV encompassing a portion of the C-helix and the aromatic domain (SEQ ID NO: 52). Collectively, these results suggest that our approaches can identify synthetic peptides that inhibit fusion/infectivity by members of the Coronaviridae family (see also Tripet, B. et al. (2004). Structural Characterization of the SARS-Coronavirus Spike S Fusion Protein Core. JBC Papers in Press. Manuscript M400759200; Liu, S. et al. (2004). Interaction between heptad repeat 1 and 2 regions in spike protein of SARS-associated coronavirus: implications for virus fusogenic mechanism and identification of fusion inhibitors. *The Lancet* Vol. 363, pp. 938-940, each of which is hereby incorporated by reference herein in its entirety).

[0206] 1. Procedures

[0207] a. CD Spectroscopy

[0208] As noted, to examine the potential for the formation of secondary structures upon interaction with lipid membranes, peptides were examined by CD spectroscopy. Circular dichroism (CD) spectra were recorded on a Jasco J-810 spectrapolarimeter (Jasco Inc., Easton, Md.), using a 1 mm path length, 1 nm bandwith, 16 second response time and a scan speed of 10 nm/min. All CD runs were performed at room temperature with peptide dissolved in 10M potassium phosphate buffer at pH 7.0. LUV were added at a lipid concentration of 1 mM from a stock in 10 mM potassium phosphate buffer pH 7.0. Three successive scans between 190-250 nm were collected and the CD data (see **FIG. 15**) are expressed as the mean residue ellipticity, derived from the formula θ=(deg*cm2)/dmol (see Wimley, W. C., and White, S. H. (2000). Designing transmembrane alpha-helices that insert spontaneously. Biochemistry 39, 4432-42, which is hereby incorporated by reference herein in its entirety).

[0209] 2. Viral Plaque Assays

[0210] L2 cells were maintained as monolayers in complete Dulbecco's modified Eagle's medium (DMEM) containing 0.15% HCO3-supplemented with 10% fetal bovine serum (FBS), penicillin G (100 U/ml), streptomycin (100 mg/ml), and 2 mM L-glutamine at 37° C. in a 5% CO_2 incubator. MHV strain A59 (ATCC, VR764) was propagated on L2 cells as described in Compton S. R., Winograd D. F., Gaertner D. J. Optimization of in vitro growth conditions for enterotropic murine coronavirus strains. *J Virol Methods.* 1995 April; 52(3): 301-7, which is hereby incorporated by reference herein in its entirety. For plaque assays, L2 cells were seeded at a density of 1×10^6 cells in each well of a 6-well plate. Approximately 100-plaque forming units (p.f.u.) of MHV were pre-incubated with or without 100 μg/ml of inhibitory peptide (SEQ ID NO: 52) in serum-free DMEM for 1 h. L2 cells were then infected with peptide-treated inoculum or vehicle control inoculum. After 1 hour adsorption, the inoculum was removed, cells were washed twice with 1× phosphate buffered saline, and the cells were

overlaid with 10% FBS/DMEM containing 0.5% SeaPlaque Agarose (Cambrex Bio Science Rockland, Inc., Rockland, Me.). Monolayers were fixed with 3.7% formalin and stained with 1× crystal violet 2 days post-infection, and plaque numbers were determined by light microscopy (Haff, R. F. (1962) Plaque formation by a mouse hepatitis virus. Virology 18, 507-508, which is hereby incorporated by reference herein in its entirety.)

[0211] Results of the viral plaque assay using the peptide having the sequence of SEQ ID NO 52 are illustrated in **FIG. 14**. The upper wells are controls exposed to vehicle and the lower wells exposed to the peptide at a nominal concentration of 25 μm. Plaques were visualized after 3 days by staining cells with crystal violet. The results show that the peptide reduced plaque formation by about 40%. There was also significant reduction (about 50%) in the average diameter of the plaques. These results suggest that this peptide inhibits both entry and spread of MHV.

[0212] B. Biophysical Experiments

[0213] 1. Interfacial Hydrophobicity Analysis

[0214] The Wimley and White hydrophobicity-at-interface scale was used to identify regions of the CoV fusion glycoprotein with high propensity to partition into lipid membranes. This scale is based on the free energies of transfer DG (kcal/mol) of amino acid sequences from water into bilayer interfaces and n-octanol, taking into consideration the contribution from the peptide bond (Wimley, W. C., Selsted, M. E., and White, S. H. (1994). Interactions between human defensins and lipid bilayers; evidence for formation of multimeric pores. *Protein Sci* 3, 1362-73; Wimley, W. C. and White, S. H. (2000a). Designing transmembrane alpha-helices that insert spontaneously. *Biochemistry* 39, 4432-42; Wimley, W. C. and White, S. H. (2000b). Determining the membrane topology of peptides by fluorescence quenching. Biochemistry 39, 161-70, each of which is hereby incorporated by reference herein in its entirety). Due to the salient similarities between the CoV fusion glycoprotein and the class I fusion glycoproteins of other RNA viruses, we compared the interfacial hydrophobicity plots of SARS CoV fusion glycoprotein to the fusion glycoproteins of HIV-1 gp41 and Ebola virus. When average interfacial hydrophobicity was plotted for the fusion proteins of these three viruses, similar regions with high propensity for membrane partitioning were detected. At the N-terminal region of all three fusion glycoproteins, a region of high interfacial hydrophobicity was detected. For HIV-1 and Ebola virus, this region corresponds to the viral fusion peptide (see **FIG. 16B** and **FIG. 16C**). Although no putative fusion peptide has been determined for the SARS CoV fusion glycoprotein, a stretch of 19 hydrophobic amino acids (WTFGAGAALQIPFAMQMAY [SEQ ID NO 51]) with an average interfacial hydrophobicity score of 2.42 kcal/mol was detected as the N-terminal region of the fusion glycoprotein. The location of this region is almost coincident with that of the HIV-1 and Ebola virus fusion peptides, and should therefore be considered as a possible fusion protein of the SARS-CoV S protein.

[0215] A second region of high interfacial hydrophobicity was detected at the C-terminal end of the fusion glycoproteins, correlating to the putative transmembrane domain of the SARS CoV fusion glycoprotein (residues 1190-1225 of **FIG. 16A**), and the experimentally determined membrane spanning anchors of HIV-1 gp41 and Ebola virus GP2 (residues 665-700 of **FIG. 16B** and residues 644-672 of **FIG. 16C**, respectively). Nieva and colleagues have shown that for HIV-1 and Ebola virus, this large region of high interfacial hydrophobicity is segmented into two-independent domains: one aromatic amino acid rich domain lying within the C-terminal end of the fusion protein and a second domain comprising the membrane-spanning anchor of the fusion protein (Nieva, J. L. et al. (1994). Interaction of the HIV-1 fusion peptide with phospholipid vesicles: different structural requirements for fusion and leakage. *Biochemistry* 33, 3201-9, which is hereby incorporated by reference herein in its entirety). The hydrophobic region at the C-terminal end of the SARS CoV fusion glycoprotein shows a remarkable similarity to that of the HIV-1 gp41 and Ebola virus GP2 in that a region of aromatic amino acids is also present and proximal to the transmembrane domain. Due to the high interfacial propensity of the aromatic region alone (3.58 kcal/mol), it is unlikely that this region is part of the transmembrane anchor as previously predicted by Rota et al. (Rota P. A. et al. Characterization of a novel coronavirus associated with severe acute respiratory syndrome. Science. 2003 May 30, which is hereby incorporated by reference herein in its entirety). Rather, like the aromatic domains of HIV-1 and Ebola virus, this region is most likely an independent domain proximal to the transmembrane anchor of the fusion glycoprotein.

[0216] Sequence analysis of the fusion glycoprotein of MHV and the human CoV OC43 showed coinciding interfacial hydrophobicity plots to that of the SARS CoV fusion glycoprotein. In addition, the presence of highly-conserved aromatic domains, differing in only 3 amino acids to the SARS aromatic domain, were identified (see **FIG. 17**). Interfacial hydrophobicity scores of 3.58, 4.86 and 5.57 kcal/mol were predicted for the aromatic domains of SARS CoV, MHV, and OC43, respectively. Based on the these analyses, peptides of 13 amino acids in length were synthesized and used throughout this study to determine the functional importance of this region within the CoV fusion glycoprotein.

[0217] 2. Peptide Synthesis

[0218] The following peptides were synthesized by solid-phase methodology using a semi-automated peptide synthesizer and conventional N-alpha-9-fluorenylmethyloxycarbonyl (Fmoc) chemistry by Genemed Synthesis, Inc. (San Francisco, Calif.):

($SARS_{Aro}$)	
KYEQYIKWPWYVW	[SEQ ID NO: 44]
(MHV_{Aro})	
TYEMYVKWPWYVW	[SEQ ID NO: 45]
($OC43_{Aro}$)	
TYEYYVKWPWYVW	[SEQ ID NO: 46]

[0219] SARS-CoV scrambled peptide ($SARS_{Scr}$) YEWK-WIYWYPVKQ [SEQ ID NO: 47] The SARS aromatic ($SARS_{Aro}$), MHV aromatic (MHV_{Aro}) and OC43 aromatic ($OC43_{Aro}$) (collectively referred to sometimes as the "CoV aromatic peptides") were synthesized based on their amino acid sequence determined from GenBank accession no. AY278741 (SARS-CoV strain Urbani), AY497331 (MHV

strain A59), and NP_937950 (Human CoV OC43). The $SARS_{Aro}$ sequence was arbitrarily scrambled to generate the peptide $SARS_{Scr}$. Hydropathy scores were determined according to methods known in the art using the Wimley and White (WW) interfacial hydrophobicity scale using a window of 13 residues (see **FIG. 17**). Peptides were purified by reversed-phase high performance liquid chromatography, and their purity confirmed by amino acid analysis and electrospray mass spectrometry. Peptide stock solutions were prepared in DMSO (spectroscopy grade), and concentrations determined spectroscopically (SmartSpec™ 3000, BiORad, Hercules, Calif.).

[0220] 3. CoV Aromatic Domains Interact with Lipid Membranes

[0221] We first assessed the ability of the CoV aromatic peptides to interact with membranes of large unilamellar vesicles (LUV) composed of different lipid compositions. LUV composed of 1-palmitoyl-2-oleyl-sn-glycero-3-phosphocholine (POPC) with phosphatidylinositol (PI), 1-palmitoyl-2-oleyl-sn-glycero-3-[phosphor-rac-(1-glycerol)] (POPG) and/or cholesterol (CHOL) were used as targets in partitioning experiments with the CoV aromatic peptides. The degree to which a peptide partitions into a vesicle can be determined fluorometrically by observing the change in tryptophan fluorescence (F) as a function of increasing lipid titration. The fluorescence of tryptophan increases in the presence of a low-polarity environment, such as the lipid membrane interface. Based on the average interfacial hydrophobicity scores of each CoV aromatic peptide alone, we predicted that all of the CoV aromatic peptides would partition into the membranes of the target vesicles.

[0222] a. LUV Preparation

[0223] Large unilamellar vesicles (LUV) consisting of POPC with POPG, PI (Avanti Polar Lipids, Birmigham, Ala.) and/or cholesterol (Sigma, St. Louis, Mich.) were prepared according to the extrusion method of Mayer, et al (Mayer L. D., Hope M. J., Cullis P. R. Vesicles of variable sizes produced by a rapid extrusion procedure. *Biochim Biophys Acta.* 1986 Jun. 13;858(1):161-8, which is hereby incorporated by reference herein in its entirety). Briefly, lipids were dried from chloroform solution with nitrogen gas stream and high vacuum overnight. Lipid vesicles used in peptide binding assays and CD experiments were resuspended in 10 mM potassium phosphate buffer to bring the concentration to 100 mM total lipid. Samples were subjected to repeated freeze and thaw for 15 cycles followed by extrusion through 0.1 m polycarbonate membranes in a Lipex Biomembranes extruder (Lipex Biomembranes, Vancouver BC). To prepare Tb3+LUV, lipids were resuspended to 100 mM concentration in 50 mM Tb3+, 100 mM sodium citrate, and 10 mM TES pH 7.2. Gel filtration on Sephadex G-200 was used to remove unencapsulated terbium in a buffer of 10 mM TES and 325 mM NaCl (56). LUV were eluted from a Sepadex G-200 gel column using 10 mM potassium phosphate pH 7.0. Final lipid concentrations were determined by phosphate analysis.

[0224] b. Interaction Assay

[0225] Partitioning of peptides into lipid bilayer was monitored by the fluorescence enhancement of tryptophan (White, S. H., Wimley, W. C., Ladokhin, A. S., and Hristova, K. (1998) Protein folding in membranes: determining energetics of peptide-bilayer interactions. Methods Enzymol 295, 62-87, which is hereby incorporated by reference herein in its entirety). Fluorescence was recorded at excitation and emission wavelengths of 280 nm and 340 nm, respectively, and 8 nm bandwidths using an SML Aminco 8100 spectrofluorometer (Rochester, N.Y.). Quartz cuvettes were used with excitation and emission path lengths of 4 mm and 10 mm. Measurements were carried out in 10 mM potassium phosphate pH 7.0. Peptides were added from DMSO stock solutions to 250 μl of buffer and mixed by inversion. LUV at a final lipid concentration of 1 mM were titrated into solution and mixed by inversion. Intensity values (I) were adjusted for lipid scattering and normalized to peptide in buffer (Io). Partitioning coefficients were obtained by fitting the formula:

$$I/Io=1+(((Kx*[L])/([W]+(Kx*[L])))*((Imax/Io)-1)$$

[0226] to the normalized data using 55.3M for water ([W]) and where Imax is equal to peptide signal at 1 mM lipid (Wimley, W. C., and White, S. H. (2000). Designing transmembrane alpha-helices that insert spontaneously. Biochemistry 39, 4432-42, which is hereby incorporated by reference herein in its entirety).

[0227] **FIG. 18** shows the normalized tryptophan fluorescence (F/Fo) for the $SARS_{Aro}$ peptide as a function of increasing lipid concentration of different LUV (mM). $SARS_{Aro}$ fluorescence increased as a direct function of increasing lipid concentrations of LUV composed of POPC. A more significant increase in tryptophan fluorescence was observed when LUV composed of POPC and either PI or POPG were titrated with the peptide, suggesting an intrinsic role for anionic lipids as a part of the membrane composition. This effect, however, was reduced when cholesterol was included as part of the membrane composition of POPC:PI LUV, perhaps due to its ability to rigidify lipid membranes. As predicted, all four CoV peptides examined partition into lipid membranes. The degree of partitioning for all four peptides was similar, and the presence of anionic lipids in the membrane composition enhanced peptide partitioning, as seen in **FIG. 18**. The addition of CHOL, however, inhibited peptide partitioning, most notably with POPC:CHOL LUV and to a lesser extent with POPC:PI:CHOL and POPC:POPG:CHOL LUV.

[0228] 4. Tb3+/DPA Microwell Assay

[0229] To test the potential of the CoV aromatic peptides to perturb membrane integrity, a high-throughput leakage assay was used. The Tb3+/DPA microwell assay is a sensitive visual screening assay known in the art to rapidly identify peptides capable of permeabilizing lipid membranes (see Rausch, J. M., and Wimley, W. C. (2001) *Anal Biochem* 293, 258-263, which is hereby incorporated by reference herein in its entirety). The detectability is based on the strong fluorescence emission of the lanthanide metal Tb3+ when it interacts with the aromatic chelator DPA. In the experimental assay, CoV aromatic peptides were incubated at peptide:lipid molar ratios of 1:100 and 1:50 with 500 mM lipid. After 2 h incubation at room temperature, the extent of Tb3+ leakage from lipid vesicles was visually determined by the detection of a bright green fluorescence upon irradiation with UV light. An example plate is shown in **FIG. 19** in which the $SARS_{Aro}$ (rows 1 and 2) and $SARS_{Scr}$ (rows 3 and 4) peptides were tested for their potential to permeabilize LUV composed of POPC, POPC:PI (9:1) or POPC:POPG

(9:1). The $SARS_{Aro}$ peptide at peptide:lipid ratios of 1:100 and 1:50 permeabilized all three LUV tested, with the greatest degree of fluorescence detected in wells with POPC or POPC:PI (9:1) LUV. In contrast, the $SARS_{Scr}$ peptide did not induce leakage of any of the three LUV tested, as detectable by this assay. The extent of leakage induced by $SARS_{Aro}$ was less than the observed leakage in the detergent solubilized wells (row 6). Comparable results were achieved with the MHV_{Aro} and $OC43_{Aro}$ peptides at peptide:lipid ratios of 1:100 and 1:50, with $OC43_{Aro}$ exhibiting the slightly lower levels of leakage (data not shown).

[0230] 5. ANTS-DPX Leakage Assay

[0231] We employed the use of the ANTS/DPX leakage assay as a second means of determining the membrane permeabilization capacity of the CoV aromatic peptides. The ability of the $SARS_{Aro}$ and $SARS_{Scr}$ peptides to release the fluorescent probe ANTS encapsulated within LUV was examined at peptide to lipid ratios of 1:500, 1:250, 1:100 and 1:50. As with the Tb3+/DPA microwell assay, the $SARS_{Aro}$ peptide induced leakage of ANTS from LUV to a greater degree than its scrambled counterpart, $SARS_{Scr}$ (see **FIG. 20**) On average, the percent leakage detected at all peptide-:lipid ratios was approximately 2 to 3 times greater for the $SARS_{Aro}$ peptide as compared to the $SARS_{Scr}$ peptide (**FIG. 20**). The degree of leakage induced by $SARS_{Aro}$ varied based on the lipid composition of the LUV tested. The percent leakage detected from LUV composed of either POPC:PI or POPC:POPG was 25% and 22%, respectively, as compared to 15% leakage observed in POPC LUV at peptide:lipid ratios of 1:100 (**FIG. 20**).

[0232] 6. CD Spectroscopy

[0233] To examine the potential for the formation of secondary structures upon interaction with lipid membranes, the CoV aromatic peptides were examined by CD spectroscopy. Circular dichroism (CD) spectra were recorded on a Jasco J-810 spectrapolarimeter (Jasco Inc., Easton, Md.), using a 1 mm path length, 1 nm bandwith, 16 second response time and a scan speed of 10 nm/min. All CD runs were performed at room temperature with peptide dissolved in 10 mM potassium phosphate buffer at pH 7.0. LUV were added at a lipid concentration of 1 mM from a stock in 10 mM potassium phosphate buffer pH 7.0. Three successive scans between 190-250 run were collected and the CD data (see **FIG. 21**) are expressed as the mean residue ellipticity, derived from the formula θ=(deg*cm2)/dmol (Wimley, W. C., and White, S. H. (2000). Designing transmembrane alpha-helices that insert spontaneously. *Biochemistry* 39, 4432-42, which is hereby incorporated by reference herein in its entirety).

[0234] The results of the CD spectroscopy study are shown in **FIG. 21**, which illustrates representative far UV CD spectra of the CoV aromatic peptides in buffer and with LUV. Analysis of the CoV peptides in 10 mM PO4 buffer pH 7.0 showed a random coil spectrum with single minima at 200 nm. No defined α-helical or β-sheet structure was apparent for any of the three CoV peptides in buffer alone. We next analyzed the potential of the CoV aromatic peptides to adopt a secondary structure in the presence of lipids. Results from our peptide partitioning and vesicle leakage assays suggested that the CoV aromatic peptides preferentially interacted with LUV composed of POPC and anionic lipids. We therefore analyzed the UV CD spectra of the CoV aromatic peptides with LUV composed of POPC:PI at lipid concentrations of 1 mM. Again, no defined secondary structure was apparent for any of the three CoV peptides in the presence of lipid. For the MHV_{Aro} and $OC43_{Aro}$ peptides, however, there was a distinct change in the observed CD spectra as compared to buffer alone (see **FIGS. 21B and 21C**). Although not indicative of a defined secondary structure due to the lack of minima at 208 nm and 222 nm for α-helical structures or 218 nm for β-sheet structures, it appears that the peptides may be assuming a more ordered structure above that of a random coil. These results are not surprising as the CoV aromatic peptides are only 13 amino acids long, a length not sufficient to cross a lipid membrane (see Rausch J. M. and Wimley W. C. (2001). A high-throughput screen for transmembrane pore-forming peptides. *Analytical Biochemistry* 293:258-63, which is hereby incorporated by reference herein in its entirety).

SEQUENCE LISTING

<160> NUMBER OF SEQ ID NOS: 52

<210> SEQ ID NO 1
<211> LENGTH: 67
<212> TYPE: PRT
<213> ORGANISM: Human metapneumovirus

<400> SEQUENCE: 1

```
Tyr Gln Leu Ser Lys Val Glu Gly Glu Gln His Val Ile Lys Gly Arg
1               5                   10                  15

Pro Val Ser Ser Ser Phe Asp Pro Ile Lys Phe Pro Glu Asp Gln Phe
            20                  25                  30

Asn Val Ala Leu Asp Gln Val Phe Glu Ser Ile Glu Asn Ser Gln Ala
        35                  40                  45

Leu Val Asp Gln Ser Asn Lys Ile Leu Asn Ser Ala Glu Lys Gly Asn
    50                  55                  60
```

-continued

Thr Gly Phe
65

<210> SEQ ID NO 2
<211> LENGTH: 78
<212> TYPE: PRT
<213> ORGANISM: Human coronavirus

<400> SEQUENCE: 2

Pro Glu Leu Asp Ser Phe Lys Glu Glu Leu Asp Lys Tyr Phe Lys Asn
1 5 10 15

His Thr Ser Pro Asp Val Asp Leu Gly Asp Ile Ser Gly Ile Asn Ala
20 25 30

Ser Val Val Asn Ile Gln Lys Glu Ile Asp Arg Leu Asn Glu Val Ala
35 40 45

Lys Asn Leu Asn Glu Ser Leu Ile Asp Leu Gln Glu Leu Gly Lys Tyr
50 55 60

Glu Gln Tyr Ile Lys Trp Pro Trp Tyr Val Trp Leu Gly Phe
65 70 75

<210> SEQ ID NO 3
<211> LENGTH: 34
<212> TYPE: PRT
<213> ORGANISM: Human metapneumovirus

<400> SEQUENCE: 3

Tyr Gln Leu Ser Lys Val Glu Gly Glu Gln His Val Ile Lys Gly Arg
1 5 10 15

Pro Val Ser Ser Ser Phe Asp Pro Ile Lys Phe Pro Glu Asp Gln Phe
20 25 30

Asn Val

<210> SEQ ID NO 4
<211> LENGTH: 33
<212> TYPE: PRT
<213> ORGANISM: Human metapneumovirus

<400> SEQUENCE: 4

Pro Val Ser Ser Ser Phe Asp Pro Ile Lys Phe Pro Glu Asp Gln Phe
1 5 10 15

Asn Val Ala Leu Asp Gln Val Phe Glu Ser Ile Glu Asn Ser Gln Ala
20 25 30

Leu

<210> SEQ ID NO 5
<211> LENGTH: 33
<212> TYPE: PRT
<213> ORGANISM: Human metapneumovirus

<400> SEQUENCE: 5

Ala Leu Asp Gln Val Phe Glu Ser Ile Glu Asn Ser Gln Ala Leu Val
1 5 10 15

Asp Gln Ser Asn Lys Ile Leu Asn Ser Ala Glu Lys Gly Asn Thr Gly
20 25 30

Phe

<210> SEQ ID NO 6

-continued

<211> LENGTH: 28
<212> TYPE: PRT
<213> ORGANISM: Human metapneumovirus

<400> SEQUENCE: 6

Tyr Gln Leu Ser Lys Val Glu Gly Glu Gln His Val Ile Lys Gly Arg
1 5 10 15

Pro Val Ser Ser Ser Phe Asp Pro Ile Lys Phe Pro
20 25

<210> SEQ ID NO 7
<211> LENGTH: 39
<212> TYPE: PRT
<213> ORGANISM: Human metapneumovirus

<400> SEQUENCE: 7

Glu Asp Gln Phe Asn Val Ala Leu Asp Gln Val Phe Glu Ser Ile Glu
1 5 10 15

Asn Ser Gln Ala Leu Val Asp Gln Ser Asn Lys Ile Leu Asn Ser Ala
20 25 30

Glu Lys Gly Asn Thr Gly Phe
35

<210> SEQ ID NO 8
<211> LENGTH: 37
<212> TYPE: PRT
<213> ORGANISM: Human metapneumovirus

<400> SEQUENCE: 8

Lys Phe Pro Glu Asp Gln Phe Asn Val Ala Leu Asp Gln Val Phe Glu
1 5 10 15

Ser Ile Glu Asn Ser Gln Ala Leu Val Asp Gln Ser Asn Lys Ile Leu
20 25 30

Asn Ser Ala Glu Lys
35

<210> SEQ ID NO 9
<211> LENGTH: 34
<212> TYPE: PRT
<213> ORGANISM: Human metapneumovirus

<400> SEQUENCE: 9

Glu Asp Gln Phe Asn Val Ala Leu Asp Gln Val Phe Glu Ser Ile Glu
1 5 10 15

Asn Ser Gln Ala Leu Val Asp Gln Ser Asn Lys Ile Leu Asn Ser Ala
20 25 30

Glu Lys

<210> SEQ ID NO 10
<211> LENGTH: 36
<212> TYPE: PRT
<213> ORGANISM: Human coronavirus

<400> SEQUENCE: 10

Pro Glu Leu Asp Ser Phe Lys Glu Glu Leu Asp Lys Tyr Phe Lys Asn
1 5 10 15

His Thr Ser Pro Asp Val Asp Leu Gly Asp Ile Ser Gly Ile Asn Ala
20 25 30

-continued

Ser Val Val Asn
35

<210> SEQ ID NO 11
<211> LENGTH: 37
<212> TYPE: PRT
<213> ORGANISM: Human coronavirus

<400> SEQUENCE: 11

Asp Val Asp Leu Gly Asp Ile Ser Gly Ile Asn Ala Ser Val Val Asn
1 5 10 15

Ile Gln Lys Glu Ile Asp Arg Leu Asn Glu Val Ala Lys Asn Leu Asn
20 25 30

Glu Ser Leu Ile Asp
35

<210> SEQ ID NO 12
<211> LENGTH: 36
<212> TYPE: PRT
<213> ORGANISM: Human coronavirus

<400> SEQUENCE: 12

Arg Leu Asn Glu Val Ala Lys Asn Leu Asn Glu Ser Leu Ile Asp Leu
1 5 10 15

Gln Glu Leu Gly Lys Tyr Glu Gln Tyr Ile Lys Trp Pro Trp Tyr Val
20 25 30

Trp Leu Gly Phe
35

<210> SEQ ID NO 13
<211> LENGTH: 26
<212> TYPE: PRT
<213> ORGANISM: Human coronavirus

<400> SEQUENCE: 13

Asn Ile Gln Lys Glu Ile Asp Arg Leu Asn Glu Val Ala Lys Asn Leu
1 5 10 15

Asn Glu Ser Leu Ile Asp Leu Gln Glu Leu
20 25

<210> SEQ ID NO 14
<211> LENGTH: 28
<212> TYPE: PRT
<213> ORGANISM: Human coronavirus

<400> SEQUENCE: 14

Leu Asn Glu Ser Leu Ile Asp Leu Gln Glu Leu Gly Lys Tyr Glu Gln
1 5 10 15

Tyr Ile Lys Trp Pro Trp Tyr Val Trp Leu Gly Phe
20 25

<210> SEQ ID NO 15
<211> LENGTH: 20
<212> TYPE: PRT
<213> ORGANISM: Human coronavirus

<400> SEQUENCE: 15

Gln Glu Leu Gly Lys Tyr Glu Gln Tyr Ile Lys Trp Pro Trp Tyr Val
1 5 10 15

-continued

Trp Leu Gly Phe
20

<210> SEQ ID NO 16
<211> LENGTH: 15
<212> TYPE: PRT
<213> ORGANISM: Human coronavirus

<400> SEQUENCE: 16

Tyr Glu Gln Tyr Ile Lys Trp Pro Trp Tyr Val Trp Leu Gly Phe
1 5 10 15

<210> SEQ ID NO 17
<211> LENGTH: 14
<212> TYPE: PRT
<213> ORGANISM: Human coronavirus

<400> SEQUENCE: 17

Tyr Glu Gln Tyr Ile Lys Trp Pro Trp Tyr Val Trp Leu Gly
1 5 10

<210> SEQ ID NO 18
<211> LENGTH: 10
<212> TYPE: PRT
<213> ORGANISM: Human coronavirus

<400> SEQUENCE: 18

Tyr Ile Lys Trp Pro Trp Tyr Val Trp Leu
1 5 10

<210> SEQ ID NO 19
<211> LENGTH: 20
<212> TYPE: PRT
<213> ORGANISM: Human coronavirus

<400> SEQUENCE: 19

Pro Glu Leu Asp Ser Phe Lys Glu Glu Leu Asp Lys Tyr Phe Lys Asn
1 5 10 15

His Thr Ser Pro
20

<210> SEQ ID NO 20
<211> LENGTH: 73
<212> TYPE: PRT
<213> ORGANISM: Human coronavirus

<400> SEQUENCE: 20

Pro Asn Leu Pro Asp Phe Lys Glu Glu Leu Asp Gln Trp Phe Lys Asn
1 5 10 15

Gln Thr Ser Val Ala Pro Asp Leu Ser Leu Asp Tyr Ile Asn Val Thr
20 25 30

Leu Asp Leu Gln Val Glu Met Asn Arg Leu Gln Glu Ala Ile Lys Val
35 40 45

Leu Asn Gln Ser Tyr Ile Asn Leu Lys Asp Ile Gly Thr Tyr Glu Tyr
50 55 60

Tyr Val Lys Trp Pro Trp Tyr Val Trp
65 70

<210> SEQ ID NO 21
<211> LENGTH: 27

-continued

<212> TYPE: PRT
<213> ORGANISM: Human coronavirus

<400> SEQUENCE: 21

Pro Asn Leu Pro Asp Phe Lys Glu Glu Leu Asp Gln Trp Phe Lys Asn
1 5 10 15

Gln Thr Ser Val Ala Pro Asp Leu Ser Leu Asp
20 25

<210> SEQ ID NO 22
<211> LENGTH: 42
<212> TYPE: PRT
<213> ORGANISM: Human coronavirus

<400> SEQUENCE: 22

Tyr Ile Asn Val Thr Phe Leu Asp Leu Gln Val Glu Met Asn Arg Leu
1 5 10 15

Gln Glu Ala Ile Lys Val Leu Asn Gln Ser Tyr Ile Asn Leu Lys Asp
20 25 30

Ile Gly Thr Tyr Glu Tyr Tyr Val Lys Trp
35 40

<210> SEQ ID NO 23
<211> LENGTH: 33
<212> TYPE: PRT
<213> ORGANISM: Human coronavirus

<400> SEQUENCE: 23

Gln Val Glu Met Asn Arg Leu Gln Glu Ala Ile Lys Val Leu Asn Gln
1 5 10 15

Ser Tyr Ile Asn Leu Lys Asp Ile Gly Thr Tyr Glu Tyr Tyr Val Lys
20 25 30

Trp

<210> SEQ ID NO 24
<211> LENGTH: 28
<212> TYPE: PRT
<213> ORGANISM: Human coronavirus

<400> SEQUENCE: 24

Gln Glu Ala Ile Lys Val Leu Asn Gln Ser Tyr Ile Asn Leu Lys Asp
1 5 10 15

Ile Gly Thr Tyr Glu Tyr Tyr Val Lys Trp Pro Trp
20 25

<210> SEQ ID NO 25
<211> LENGTH: 20
<212> TYPE: PRT
<213> ORGANISM: Human coronavirus

<400> SEQUENCE: 25

Gln Ser Tyr Ile Asn Leu Lys Asp Ile Gly Thr Tyr Glu Tyr Tyr Val
1 5 10 15

Lys Trp Pro Trp
20

<210> SEQ ID NO 26
<211> LENGTH: 12
<212> TYPE: PRT

-continued

<213> ORGANISM: Human coronavirus

<400> SEQUENCE: 26

Tyr Glu Tyr Tyr Val Lys Trp Pro Trp Tyr Val Trp
1 5 10

<210> SEQ ID NO 27
<211> LENGTH: 31
<212> TYPE: PRT
<213> ORGANISM: Mouse hepatitis virus

<400> SEQUENCE: 27

Gln Asp Ala Ile Lys Lys Leu Asn Glu Ser Tyr Ile Asn Leu Lys Glu
1 5 10 15

Val Gly Thr Tyr Glu Met Tyr Val Lys Trp Pro Trp Tyr Val Trp
20 25 30

<210> SEQ ID NO 28
<211> LENGTH: 6
<212> TYPE: PRT
<213> ORGANISM: Human immunodeficiency virus type 1

<400> SEQUENCE: 28

Phe Leu Gly Phe Leu Gly
1 5

<210> SEQ ID NO 29
<211> LENGTH: 4
<212> TYPE: PRT
<213> ORGANISM: Human coronavirus

<400> SEQUENCE: 29

Thr Thr Thr Ser
1

<210> SEQ ID NO 30
<211> LENGTH: 5
<212> TYPE: PRT
<213> ORGANISM: Human coronavirus

<400> SEQUENCE: 30

Glu Leu Asp Lys Tyr
1 5

<210> SEQ ID NO 31
<211> LENGTH: 5
<212> TYPE: PRT
<213> ORGANISM: Human immunodeficiency virus type 1

<400> SEQUENCE: 31

Glu Leu Asp Lys Trp
1 5

<210> SEQ ID NO 32
<211> LENGTH: 3
<212> TYPE: PRT
<213> ORGANISM: Human coronavirus

<400> SEQUENCE: 32

Pro Glu Leu
1

-continued

<210> SEQ ID NO 33
<211> LENGTH: 5
<212> TYPE: PRT
<213> ORGANISM: Mouse hepatitis virus

<400> SEQUENCE: 33

Pro Asp Phe Lys Glu
1 5

<210> SEQ ID NO 34
<211> LENGTH: 7
<212> TYPE: PRT
<213> ORGANISM: Human coronavirus

<400> SEQUENCE: 34

Phe Lys Glu Glu Leu Asp Lys
1 5

<210> SEQ ID NO 35
<211> LENGTH: 8
<212> TYPE: PRT
<213> ORGANISM: Human coronavirus

<400> SEQUENCE: 35

Lys Trp Pro Trp Tyr Val Trp Leu
1 5

<210> SEQ ID NO 36
<211> LENGTH: 6
<212> TYPE: PRT
<213> ORGANISM: Human metapneumovirus

<400> SEQUENCE: 36

Gln Ala Leu Val Asp Gln
1 5

<210> SEQ ID NO 37
<211> LENGTH: 5
<212> TYPE: PRT
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: This is an artificially created peptide acetyl

<400> SEQUENCE: 37

Pro Glu Gln Leu Lys
1 5

<210> SEQ ID NO 38
<211> LENGTH: 11
<212> TYPE: PRT
<213> ORGANISM: Human immunodeficiency virus type 1

<400> SEQUENCE: 38

Gly Arg Lys Lys Arg Arg Gln Arg Arg Arg Pro
1 5 10

<210> SEQ ID NO 39
<211> LENGTH: 23
<212> TYPE: PRT
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 39

-continued

Glu Leu Arg Val Arg Leu Ala Ser His Leu Arg Lys Leu Arg Lys Arg
1 5 10 15

Leu Leu Arg Asp Ala Asp Asp
20

<210> SEQ ID NO 40
<211> LENGTH: 7
<212> TYPE: PRT
<213> ORGANISM: Mouse hepatitis virus

<400> SEQUENCE: 40

Arg Ile Gln Asp Ala Ile Lys
1 5

<210> SEQ ID NO 41
<211> LENGTH: 7
<212> TYPE: PRT
<213> ORGANISM: Human coronavirus

<400> SEQUENCE: 41

Arg Leu Asn Glu Val Ala Lys
1 5

<210> SEQ ID NO 42
<211> LENGTH: 21
<212> TYPE: PRT
<213> ORGANISM: Human coronavirus

<400> SEQUENCE: 42

Glu Asn Gln Lys Gln Ile Ala Asn Gln Phe Asn Lys Ala Ile Ser Gln
1 5 10 15

Ile Gln Glu Ser Leu
20

<210> SEQ ID NO 43
<211> LENGTH: 20
<212> TYPE: PRT
<213> ORGANISM: Human coronavirus

<400> SEQUENCE: 43

Lys Val Gln Asp Val Val Asn Gln Asn Ala Gln Ala Leu Asn Thr Leu
1 5 10 15

Val Lys Gln Leu
20

<210> SEQ ID NO 44
<211> LENGTH: 13
<212> TYPE: PRT
<213> ORGANISM: Human coronavirus

<400> SEQUENCE: 44

Lys Tyr Glu Gln Tyr Ile Lys Trp Pro Trp Tyr Val Trp
1 5 10

<210> SEQ ID NO 45
<211> LENGTH: 13
<212> TYPE: PRT
<213> ORGANISM: Mouse hepatitis virus

<400> SEQUENCE: 45

Thr Tyr Glu Met Tyr Val Lys Trp Pro Trp Tyr Val Trp

-continued

1 5 10

<210> SEQ ID NO 46
<211> LENGTH: 13
<212> TYPE: PRT
<213> ORGANISM: Human coronavirus

<400> SEQUENCE: 46

Thr Tyr Glu Tyr Tyr Val Lys Trp Pro Trp Tyr Val Trp
1 5 10

<210> SEQ ID NO 47
<211> LENGTH: 13
<212> TYPE: PRT
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: SEQ ID NO: 44 was arbitrarily scrambled to generate this sequence

<400> SEQUENCE: 47

Tyr Glu Trp Lys Trp Ile Tyr Trp Tyr Pro Val Lys Gln
1 5 10

<210> SEQ ID NO 48
<211> LENGTH: 83
<212> TYPE: PRT
<213> ORGANISM: Human coronavirus

<400> SEQUENCE: 48

Pro Asn Leu Pro Asp Phe Lys Glu Glu Leu Asp Gln Trp Phe Lys Asn
1 5 10 15

Gln Thr Ser Val Ala Pro Asp Leu Ser Leu Asp Tyr Ile Asn Val Thr
20 25 30

Phe Leu Asp Leu Gln Val Glu Met Asn Arg Leu Gln Glu Ala Ile Lys
35 40 45

Val Leu Asn Gln Ser Tyr Ile Asn Leu Lys Asp Ile Gly Thr Tyr Glu
50 55 60

Tyr Tyr Val Lys Trp Pro Trp Tyr Val Trp Leu Leu Ile Cys Leu Ala
65 70 75 80

Gly Val Ala

<210> SEQ ID NO 49
<211> LENGTH: 85
<212> TYPE: PRT
<213> ORGANISM: Mouse hepatitis virus

<400> SEQUENCE: 49

Pro Asn Pro Pro Asp Phe Lys Glu Glu Leu Asp Lys Trp Phe Lys Asn
1 5 10 15

Gln Thr Ser Ile Ala Pro Asp Leu Ser Leu Asp Phe Glu Lys Leu Asn
20 25 30

Val Thr Leu Leu Asp Leu Thr Tyr Glu Met Asn Arg Ile Gln Asp Ala
35 40 45

Ile Lys Lys Leu Asn Glu Ser Tyr Ile Asn Leu Lys Glu Val Gly Thr
50 55 60

Tyr Glu Met Tyr Val Lys Trp Pro Trp Tyr Val Trp Leu Leu Ile Gly
65 70 75 80

Leu Ala Gly Val Ala

-continued

```
               85


<210> SEQ ID NO 50
<211> LENGTH: 84
<212> TYPE: PRT
<213> ORGANISM: Human coronavirus

<400> SEQUENCE: 50

Pro Glu Leu Asp Ser Phe Lys Glu Glu Leu Asp Lys Tyr Phe Lys Asn
1               5                   10                  15

His Thr Ser Pro Asp Val Asp Leu Gly Asp Ile Ser Gly Ile Asn Ala
            20                  25                  30

Ser Val Val Asn Ile Gln Lys Glu Ile Asp Arg Leu Asn Glu Val Ala
        35                  40                  45

Lys Asn Leu Asn Glu Ser Leu Ile Asp Leu Gln Glu Leu Gly Lys Tyr
    50                  55                  60

Glu Gln Tyr Ile Lys Trp Pro Trp Tyr Val Trp Leu Gly Phe Ile Ala
65                  70                  75                  80

Gly Leu Ile Ala


<210> SEQ ID NO 51
<211> LENGTH: 19
<212> TYPE: PRT
<213> ORGANISM: Human coronavirus

<400> SEQUENCE: 51

Trp Thr Phe Gly Ala Gly Ala Ala Leu Gln Ile Pro Phe Ala Met Gln
1               5                   10                  15

Met Ala Tyr


<210> SEQ ID NO 52
<211> LENGTH: 36
<212> TYPE: PRT
<213> ORGANISM: Mouse hepatitis virus

<400> SEQUENCE: 52

Arg Ile Gln Asp Ala Ile Lys Lys Leu Asn Glu Ser Tyr Ile Asn Leu
1               5                   10                  15

Lys Glu Val Gly Thr Tyr Glu Met Tyr Val Lys Trp Pro Trp Tyr Val
            20                  25                  30

Trp Leu Leu Ile
        35
```

What is claimed is:

1. A peptide derived from an enveloped virus having a fusion glycoprotein amino acid residue sequence and a CPI helix amino acid residue subsequence in said fusion glycoprotein amino acid residue sequence, said peptide corresponding to a segment of said CPI helix amino acid residue subsequence.

2. A peptide according to claim 1 wherein said enveloped virus is human metapneumovirus.

3. A peptide according to claim 1 wherein said enveloped virus is human coronavirus.

4. A peptide according to claim 1 wherein said peptide is derived from a segment of said CPI helix amino acid residue subsequence that is conserved among class I viral fusion glycoproteins.

5. A peptide according to claim 1, wherein said peptide comprises an amino acid residue sequence chosen from the group consisting of SEQ ID NO: 1 through SEQ ID NO: 26, SEQ ID NO: 34 and SEQ ID NO: 35.

6. A peptide according to claim 1, wherein said peptide comprises an amino acid residue sequence comprising SEQ ID NO: 52.

7. The peptide of claim 5, wherein said peptide further comprises an adduct (x) at the amino terminus of said peptide or an adduct (x') at the carboxy terminus of said peptide.

8. The peptide of claim 6, wherein said adduct x is selected from the group consisting of an acetyl group, a carbobenzoxy group, a 9-fluorenylmethoxy group, a

D-amino acid, a hydrophobic adduct, a carrier macromolecule, a lipid, and the peptide acetyl PEQLK [SEQ ID NO: 37].

9. The peptides of claim 6 wherein the adduct x' is selected from the group consisting of an amido group, a hydrophobic adduct, a carrier macromolecule, and a lipid.

10. A method for inhibiting infection of a human cell by human metapneumovirus comprising administering to a human host an inhibitory effective concentration of a peptide, wherein said peptide comprises a segment from a CPI helix amino acid residue subsequence of a fusion glycoprotein amino acid residue sequence of said human metapneumovirus.

11. The method according to claim 9, wherein said peptide is derived from SEQ ID NO. 01.

12. A method for inhibiting infection of a human cell by human coronavirus comprising administering to a human host an inhibitory effective concentration of a peptide, wherein said peptide comprises a segment from a CPI helix amino acid residue subsequence of a fusion glycoprotein amino acid residue sequence of said human coronavirus.

13. The method according to claim 11 wherein said peptide comprises a segment from SEQ.ID NO. 02 or SEQ ID NO. 20.

14. A method for inhibiting infection of a human cell by human metapneumovirus comprising administering to a human host an inhibitory effective concentration of a combination of peptides, wherein each of said peptides in said combination of peptides comprises a segment from a CPI helix amino acid residue subsequence of a fusion glycoprotein amino acid residue sequence of said human metapneumovirus.

15. The method according to claim 13 wherein each of said peptides in said combination of peptides comprises a segment from SEQ ID NO. 01.

16. A method for inhibiting infection of a human cell by human coronavirus comprising administering to a human host an inhibitory effective concentration of a combination of peptides wherein each of said peptides in said combination of peptides comprises a segment from a CPI helix amino acid residue subsequence of a fusion glycoprotein amino acid residue sequence of said human coronavirus.

17. The method according to claim 15 wherein each of said peptides in said combination of peptides comprises a segment from SEQ ID NO: 02 or SEQ ID NO: 20.

18. A method for inhibiting infection of a human cell by human coronavirus comprising administering to a human host an inhibitory effective concentration of a combination of peptides wherein each of said peptides in said combination of peptides comprises a segment from the RNA of said virus corresponding to a CPI helix amino acid residue subsequence of a fusion glycoprotein amino acid residue sequence of said human coronavirus.

19. A process for selecting a peptide as a candidate for inhibiting infection of a human cell by an enveloped virus having a fusion glycoprotein amino acid residue sequence and a CPI helix amino acid residue subsequence in said fusion glycoprotein amino acid residue sequence, comprising:

(1) searching the primary amino acid residue sequence of said virus for an amino acid subsequence 1 of about 20-25 amino acid residues containing more than about 60 percent of hydrophobic amino acid residues (Phenylalanine (F), Tyrosine (Y), Tryptophane (W), Alanine (A), Valine (V), Leucine (L), Isoleucine (I), Methionine (M) or Cysteine (C));

(2) searching within a range of about 100 amino acid residues from the amino end of said subsequence 1 for a subsequence 2 containing more than about 60% of: (a) charged amino acid residues (Glutamate (E) or Aspartate (D), Lysine (K) or Arginine (R)) and (b) Helix amino acids residues (glutamine (Q), glutamate (E), alanine (A), phenylalanine (F), tryptophane (W), lysine (K) or leucine (L));

(3) selecting a subsequence 3 of said subsequence 2.

20. The process of claim 18 further including the step of testing said subsequence 3 for inhibitory effectiveness.

21. The process of claim 18 wherein said wherein amino acid residues in said subsequence 3 are further substituted with alternate amino acid residues having similar biological properties, as follows:

Short side chain—Glycine (G) or Proline (P) or Alanine (A)

Hydroxylated side chain—Serine (S) or Threonine (T) or Tyrosine (Y)

Aliphatic side chain—Alanine (A) or Valine (V) or Leucine (L) or Isoleucine (I) or Methionine (M) or Cysteine (C)

Sulphur-containing side chain —Cysteine (C) or Methionine (M)

Aromatic side chain —Phenylalanine (F) or Tyrosine (Y) or Tryptophane (W)

Neutral side chain—Glutamine (Q) or Asparagine (N) or Histidine (H)

Acidic side chain —Glutamate (E) or Aspartate (D), or Histidine (H)

Basic side chain —Lysine (K) or Arginine (R).

22. The process of claim 18 wherein said subsequence 3 begins with a di- or tri-peptide motif of amino acid residues comprising glutamate (E) or glutamine (Q) or phenylalanine (F) or lysine (K) or alanine (A) or leucine (L).

23. The process of claim 18 wherein said subsequence 3 begins with a proline (P) positioned within three amino acid residues of a di- or tri-peptide motif of amino acid residues comprising glutamate (E) or glutamine (Q) or phenylalanine (F) or lysine (K) or alanine (A) or leucine (L).

24. A peptide produced according to the process of claim 18 wherein said subsequence 3 contains a concentration of said Helical amino acids in excess of about 40%.

25. A peptide produced according to the process of claim 18 wherein said subsequence 3 is uniformly constructed from said Helical amino acids.

26. A peptide produced according to the process of claim 18 wherein said subsequence 3 terminates with alanine (A), glutamate (E), glutamine (Q), tyrosine (Y), phenylalanine (F), lysine (K) or proline (P) residues.

27. A peptide produced according to the process of claim 18 wherein said subsequence 3 is of minimum length of 6 amino acid residues which subsequence is conserved across related viral family members.

28. A peptide produced according to the process of claim 18 further comprising one or more adducts at either the amino- or carboxy-termini of said peptide.

* * * * *

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
10 March 2005 (10.03.2005)

PCT

(10) International Publication Number
WO 2005/021798 A1

(51) International Patent Classification: C12Q 1/68

(21) International Application Number:
PCT/US2004/026380

(22) International Filing Date: 13 August 2004 (13.08.2004)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:

60/496,995	22 August 2003 (22.08.2003)	US
60/576,314	3 June 2004 (03.06.2004)	US

(71) Applicant *(for all designated States except US)*: **BIRCH BIOMEDICAL RESEARCH LLC** [US/US]; 226 East 54th Street, New York, NY 10022 (US).

(71) Applicant and
(72) Inventor: **KOSTRIKIS, Leondios, G.** [CY/CY]; 5A Melis Nicolaides Street, Kapsalos Limassol (CY).

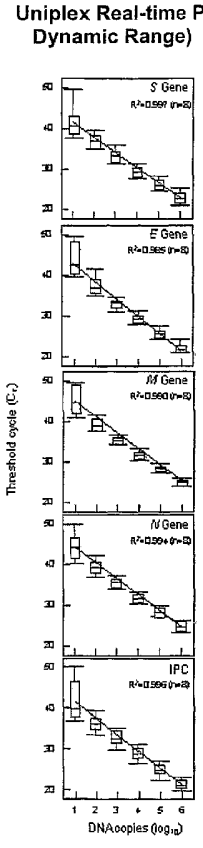
(74) Agent: **WISEMAN, Thomas, G.**; P.O. Box 34385, Washington, D.C., District of Columbia 20043-9998 (US).

(81) Designated States *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BW, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NA, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, YU, ZA, ZM, ZW.

(84) Designated States *(unless otherwise indicated, for every kind of regional protection available)*: ARIPO (BW, GH, GM, KE, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IT, LU, MC, NL, PL, PT, RO, SE, SI,

[Continued on next page]

(54) Title: MULTI-ALLELIC MOLECULAR DETECTION OF SARS-ASSOCIATED CORONAVIRUS



(57) Abstract: The subject invention relates to a multiple-allelic RT-real-time polymerase chain reaction (PCR) assay for coronaviruses including the SARS virus. Multiple target sequences within the SARS-CoV, S, E, M and N genes are identified. The use of the four different targets enhances the likelihood that the fundamental genetic drift of the virus will not lead to a false negative result. Multiplex assays format for the assay are envisioned. Thus, the present invention allows for early diagnosis of a SARS infection. The assay would be useful in the context of monitoring treatment regimens, screening potential anti SARS agents, and similar applications requiring qualitative and quantitative determinations.

WO 2005/021798 A1

SK, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:

— *with international search report*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

Multi-allelic Molecular Detection of SARS-associated Coronavirus

BACKGROUND OF THE INVENTION

Field of Invention

This invention is directed to methods for the detection and/or quantitation of the SARS virus, reagents and test kits containing the same for use in the method.

Background of the Invention

Severe acute respiratory syndrome (SARS) is one of the most recent emerging infectious diseases. The cause of SARS has been identified as a new coronavirus – a virus within the family Coronaviridae – designated as the "SARS coronavirus" (SARS-CoV) [1, 2] by the World Health Organization, following assessment of causation according to Koch's postulates, including monkey inoculation [3]. The coronaviruses are enveloped positive single-stranded RNA viruses with genomes approximately 30 kb in length – the largest of any of the RNA viruses – that replicate in the cytoplasm of host cells without going through DNA intermediates. Coronaviruses have been reported to cause common colds in humans, and to cause respiratory, enteric, and neurological diseases, as well as hepatitis, in animals. Human coronaviruses are usually difficult to culture in vitro, whereas most animal coronaviruses and SARS–CoV can easily be cultured in Vero E6 cells [4]. There are three groups of coronaviruses: Groups 1 and 2 encompass mammalian viruses, whereas Group 3 encompasses avian viruses. Within each group, the coronaviruses are classified into distinct species according to host range, antigenic relationships, and genomic organization. Human coronaviruses (HCoVs) were previously reported to belong in Group 1 (HCoV–229E) and Group 2 (HCoV–OC43), and are responsible for mild respiratory illnesses.

Recently, two independent groups, one at the British Columbia Cancer Agency (BCCA) in Canada [5] (Tor2 isolate), and the other at the Centers for Disease Control and Prevention (CDCP) in the United States [6] (Urbani isolate), were first to obtain full genomic sequences of SARS–CoV. Phylogenetic analyses, based on the genome sequences, revealed that both isolates were distantly related to previously characterized coronaviruses, including the two previously isolated nonpathogenic human coronaviruses strains, HCoV–OC43 and HCoV–229E. The genome of the Tor2 CoV isolate is 29,751 nucleotides long, and the genome of the Urbani CoV isolate is 29,727 nucleotides long, and their sequences differ at only 24 nucleotide positions. The genomic organization of both isolates is characteristic of coronaviruses having the

following typical gene order: 5'-replicase (rep), spike (S), envelope (E), membrane (M), and nucleocapsid (N). The SARS-CoV rep gene, which is approximately 20,000 nucleotides long, is predicted to encode two polyproteins (ORF1a and ORF1b) that undergo proteolytic processing, resulting in several nonstructural proteins. There are four genes downstream of rep that encode the structural proteins S, E, M, and N.

The genome of SARS–CoV has several distinct genomic characteristics that distinguish it from other coronavirus isolates and that could be of biological significance. The gene encoding hemagglutinin-esterase, which is present between ORF1a and S in Group 2 coronaviruses (and in some Group 3 coronaviruses) is absent, and so is the short anchor of the S protein. Furthermore, the short anchor of the S protein, the specific number and location of the small ORFs, and the presence of only one copy of PLP^{PRO} provide a combination of genetic features that readily distinguish SARS–CoV isolates from previously the described coronaviruses [5, 6]. There are several publications that describe reverse-transcriptase polymerase chain reaction assays (RT-PCR assays) for the detection of SARS–CoV.

Perris et al. [2] developed an RT-PCR assay that identifies the virus from nasopharyngeal aspiration samples obtained from patients infected with SARS–CoV. Total RNA from clinical samples is reverse transcribed in the presence of random hexamers, and the resulting cDNA is amplified with primers 5'-TACACACCTCAGCGTTG-3' and 5'-CACGAACGTGACGAAT-3'. To determine the genetic sequence of an unknown RNA virus, they perform a random RT-PCR assay. Total RNA from virus-infected and virus-uninfected fetal rhesus kidney cells were isolated, reverse transcribed with primer 5'-GCCGGAGCTCTGCAGAATTCNNNNNN-3', and the resulting cDNA was amplified with primer 5'-GCCGGAGCTCTGCAGAATTC-3'.

Ksiazek et al. [1] developed a reverse transcription and real-time PCR assay to identify SARS–CoV. Oligonucleotide primers used for amplification and sequencing of the SARS-related coronavirus were designed from alignments in open reading frame 1b of the coronavirus polymerase gene sequences. They used the primer pair IN-2 (+) 5'-GGGTTGGGACTATCCTAAGTGTGA-3' and IN-4 (–) 5'-TAACACACAACICCATC ATCA-3', which was previously designed to hybridize to conserved regions of open reading frame 1b (ORF1b), in order to achieve broad reactivity with the coronavirus/genus. These primers were used to amplify DNA from SARS isolates, and the amplicon sequences obtained were used to design SARS-specific primers Cor-p-F2 (+) 5'-CTAACATGCTTAGGATAATGG-

3', Cor-p-F3 (+) 5'-GCCTCTCTTGTTCTTGCTCGC-3', and Cor-p-R1 (–) 5'-CAGGTAAGCG TAAAACTCATC-3', which were used in turn to test patient specimens. Drosten et al. [4] used a PCR-based random-amplification procedure to genetically characterize a 300-nucleotide-long SARS–CoV genomic segment. On the basis of the sequence that was obtained, conventional and real-time PCR assays for specific detection SARS–CoV ORF1b were established. Poon et al. [7] developed an RT-real-time-PCR assay. Total RNA isolated from stool specimens from SARS–CoV-infected individuals is reverse transcribed with random hexamers and the resulting cDNA is amplified with primers coro3 5'-TACACACCTCAGCGTTG-3' and coro4 5'-CACGAACGTGACGAAT-3', which recognize a region of the viral polymerase gene. It is important to note that these authors acknowledged in their publication that the primers that they use in their assay can cross-react with the nonpathogenic human coronavirus strain HCoV–OC43.

SARS-specific PCR primers and diagnostic procedures were developed in several World Health Organization network laboratories for the amplification of a region of the open reading frame 1b (ORF1b) of the SARS–CoV polymerase gene sequence [8]. These primers are currently being assessed to determine their relative performance and sensitivity with different specimens obtained at different times over the course of illness. Lipkin and Briese have announced they develop a PCR-based SARS diagnostic that detects a SARS–CoV gene that is present in multiple copies, but no further information is available in the literature.

Problems with the prior art that the current invention is designed to solve. The main problems with current molecular diagnostic assays are: a) failure to consider the intrinsically polymorphic nature of coronaviruses, including the current SARS–CoV strains originated from the Tor2 and Urbani isolates – the ability of the virus to mutate and recombine during the period of time it is within the infected individual, and during horizontal transmission; and b) failure to account for the possibility of continuous and/or multiple introduction of non non-genetically identical SARS–CoV strains into the human population.

A characteristic of RNA viruses is their high rate of genetic mutation, which leads to evolution of new viral strains, and is a well-established mechanism by which viruses escape the immune system. Coronaviruses, including SARS–CoV, are quite sloppy when it comes to replicating their genetic material, producing one error for every 10,000 nucleotides that they copy, which is roughly the same error rate as occurs during the replication of the human immunodeficiency virus, HIV-1. Coronavirus RNA polymerase sometimes jumps between

multiple copies of the viral genome that are present in an infected cell. Therefore, each new genome is actually copied from several templates, reducing the chance that any given mutation will become well established in the viral population. Moreover, if one of these jumps is imprecise, a whole chunk of genome can get skipped, resulting in the deletion of part of an important gene. The consequences can be dramatic, particularly if the change affects the protein spikes that enable the virus to bind to the surface of the host's cells. For example, in 1984 a new respiratory sickness appeared on European pig farms. It turned out to be a deletion mutant of a coronavirus that previously had infected piglets' stomachs [9]. It possessed an altered spike protein that enabled the virus to infect a different cell type. Although the new disease was not generally lethal, it has since spread worldwide, and it has complicated the diagnosis of the gut disease. Another example is the recent introduction of SARS–CoV into the human population. It is likely that a genetic deletion may have helped the SARS virus to make the transition from its animal reservoir to humans. Genetics analyses of the viral strains found in animals for sale in Southern Chinese markets indicated that these SARS virus strains lack 29 nucleotides in the gene encoding a protein of unknown function, and the protein product of this gene is attached to the inside of the virus' coat protein. Furthermore, in a recent publication, full genome sequences of 14 isolates from SARS-CoV-infected patients in Singapore, Toronto, China, and Hong Kong were compared, and 14 mutations were revealed [10]. In one respect, this finding may be viewed as indicating that the SARS virus fails to mutate; however, this virus has so far encountered little resistance from its new human hosts, and there has, therefore, been little selective pressure to cause new mutants to be retained. SARS–CoV will probably not remain as stable as it has been so far. Our immune systems could force changes, similar to the changes that frequently occur in flu viruses. In summary, we deemed it prudent to develop a new SARS–CoV diagnostic assay that accounts for the genetically polymorphic nature of coronaviruses, including SARS–CoV.

SUMMARY OF THE INVENTION

The present invention includes a molecular-beacon-based multi-allelic RT-real-time-PCR assay for the detection of and discrimination between SARS-associated and other coronavirus isolates in clinical samples. The main elements of the assay design are: a) mismatch-tolerant molecular beacons; b) four sets of PCR primers for four different viral genes, and four different molecular beacons (each labeled with the same fluorophore, and each specific for a different SARS–CoV gene); c) an exogenous RNA standard that is added to the sample that can be reverse-transcribed and amplified by one of the primer sets; and d) a fifth molecular beacon that is labeled with a different fluorophore that is specific for the exogenous RNA standard. The assay further includes RNA isolation from clinical samples (blood, tissue, sputum, nasopharyngeal aspiration samples, and others), reverse transcription, PCR amplification and simultaneous automated amplicon detection in a spectrofluorometric thermal cycler that measures the fluorescence intensity of each color during the annealing phase of each thermal cycle.

Multiple target sequences within the SARS–CoV S, E, M and N genes (Urbani and Tor2 strains) were identified. The S, M, and E genes encode structural proteins that are present on the outside of the virus, whereas the N gene encodes a structural protein that is required for viral RNA packaging inside the virion. The principle underlying the selection of four target sequences that uniquely identify SARS–CoV (rather than only one target sequence) is that the use of four different targets enhances the likelihood that the fundamental genetic drift of the virus will not lead to a false negative result – that is, one has better chance of hitting a moving target with a shotgun than with a rifle. Thus, by detecting four different target alleles in the same assay tube, and by using a single-fluorophore detection system, the design of the assay significantly minimizes the likelihood of missing the presence of the SARS–CoV in a clinical sample due to the continuous viral evolution of the viral sequence. Moreover, by simultaneously detecting four different target sequences in the same assay tube, the intrinsic sensitivity of the assay is enhanced.

In order to identify the best target sequences within each viral gene that discriminate the SARS-Urbani and SARS-Tor2 strains from other nonpathogenic human and animal coronavirus strains, we used DNA alignments and phylogenetic analysis of available coronavirus gene sequences deposited in GenBank. DNA sequences of SARS–CoV genes were compared with those from reference viruses representing each species in the three known groups of

coronaviruses [group 1 (G1): human coronavirus 229E (HCoV–229E), af304460; porcine epidemic diarrhea virus (PEDV), af353511; transmissible gastroenteritis virus (TGEV), aj271965; canine coronavirus (CCoV), d13096; feline coronavirus (FCoV), ay204704; porcine respiratory coronavirus (PRCoV), z24675; – Group 2 (G2): bovine coronavirus (BCoV), af220295; murine hepatitis virus (MHV), af201929; human coronavirus OC43 (HCoV–OC43), m76373; porcine hemagglutinating encephalomyelitis virus (HEV), ay078417; rat coronavirus (RtCoV), af207551; and – Group 3 (G3): infectious bronchitis virus (IBV), m95169]. Sequence alignments were performed by CLUSTALW, which is a multiple sequence alignment tool that is commonly used in the bioinformatics community. It produces global multiple sequence alignments through three major phases: a) pairwise alignment, b) guide-tree construction, and c) multiple alignment. The guide tree generated by CLUSTALW is an estimate of relationships between sequences that are much like those shown by phylogenetic trees.

The criteria for selecting SARS–CoV gene-specific PCR primers were based on: a) the identification of genomic regions in SARS–CoV that, as a result of an examination of the sequence alignments, showed the highest genetic distance between SARS–CoV and other coronavirus strains; b) selection of primer sequences for amplification of the SARS–CoV targets that form primer-target hybrids whose theoretical melting temperature maximizes the ability of the primer to bind to the target even if nucleotide substitutions are present (mismatch tolerance), and yet enable all of the primers to hybridize to their targets at the same temperature in a multiplex assay (T_m approximately 60°C); and c) selection of primer sequences that enable the amplicons containing each of the four target sequences to be approximately the same (relatively short) length (approximately 100 nucleotides long).

The criteria for selecting the molecular beacon probe sequences, and their arm sequences, were based on: a) the identification of approximately 30-nucleotide-long regions in SARS–CoV (within the amplicons to be generated) that, as a result of an examination of the sequence alignments, showed the highest genetic distance between SARS–CoV and other coronavirus strains (with special emphasis on probe target sequences that encompass gaps or deletions in the SARS-CoV sequence compared to the sequence of other coronaviruses); b) selection of probe sequences that form probe-target hybrids whose theoretical melting temperature maximizes the ability of the probe to bind to the target sequence even if nucleotide substitutions are present (mismatch tolerance), and yet enable all of the probes to hybridize to their targets at the same temperature in a multiplex assay (Tm approximately 63°C); and c)

selection of arm sequences that provide the same degree of stability for the stem hybrids of all of the molecular beacon probes (stem T_m of approximately 70°C).

The assays described herein can be performed in either a heterogeneous or homogeneous format. The reagents needed for performance of the assay can be supplied in a kit format. The kit contains the detectants necessary for measuring two or more of the coronavirus genes S, E, M and N. It is recommended that the kit also contain an internal standard IPC. The reagents including the detectants can be separately packaged in individual ccontainers. The kits may also contain a substrate including reaction tubes for performing an assay for a given sample. The kit may also contain additional reagents for performing amplification reactions including PCR and also for saple pretreatment including those reagent necessary to release and /or purify the coronavirus.

When there is a need to perform two or more different assays on the same sample, most of the time in a single vessel and at about the same time, a multiplex format can be utilized. Such formats are known in the art. Multiplex assays are typically used to determine simultaneously the presence or concentratin of more than one molecule in the samle being analyzed, or alternatively, several characteristics of a single molecule.

BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1 shows the DNA Sequence Alignment of Coronavirus S Genes isolated from different species.

Figure 2 shows the Phylogenetic Analysis of S Gene.

Figure 3 shows molecular designs for SARS-Associated S Gene.

Figure 4 lists Molecular Designs for S Gene.

Figure 5 shows the DNA Sequence Alignment of Coronavirus E Genes isolated from different Species.

Figure 6 shows the phylogenetic analysis of E Gene.

Figure 7 shows molecular designs for SARS-Associated E Gene.

Figure 8 shows molecular designs for SARS-Associated E Gene

Figure 9 shows the DNA Sequence Alignment of Coronavirus M Genes isolate from different species.

Figure 10 shows the phylogenetic analysis of M Gene

Figure 11 shows the molecular designs for SARS-Associated M Gene.

Figure 12 lists molecular designs for M Gene.

Figure 13 shows the DNA Sequence Alignment of Coronavirus N Genes isolate from different species.

Figure 14 shows the phylogenetic analysis of N Gene.

Figure 15 shows the molecular designs for SARS-Associated N Gene.

Figure 16 list molecular designs for N Gene.

Figure 17 shows molecular designs for Internal Positive Control (IPC).

Figure 18 lists molecular designs for IPC.

Figure 19 shows Molecular Beacon Melting Curves.

Figure 20 shows Uniplex Real-time PCR Amplifications (Serial Dilutions-Dynamic Range).

DETAIL DESCRIPTION OF THE INVENTION

The general steps for the assay involve the performance of the following steps:

Step 1. An RNA standard that can be reverse-transcribed and amplified by one of the primer sets is added to each sample prior to isolation of RNA from the sample. RNA is then extracted from each clinical sample (nasopharyngeal aspirations, stool samples, or whole blood, obtained from patients suspected of being infected with SARS–CoV). RNA is then purified, using a QIAamp Viral RNA Kit (Qiagen, Inc., Valencia, CA), according to the manufacturer's instructions.

Step 2. The isolated RNA obtained from each clinical sample is then reverse transcribed with four target-specific primers (S Gene, S-RT 5'-AGGCTGTAAGAA-3'; E-Gene, E-RT 5'-TATTGCAGCAGTAC-3'; M Gene, M-RT 5'-AAGCAACGAAGTAG-3'; N Gene, N-RT 5'-GCCTTCTTTGTTAG-3'; Internal Positive Control, E-RT 5'-TATTGCAGCAGTAC-3'). Reverse transcription with SARS-CoV viral RNA from patient samples is performed by adding 20 μL viral RNA to a mixture of 0.125 μM of (each) gene-specific primer and incubating at 80°C for 5 minutes to denature secondary structures. The tubes are then immediately placed on ice for at least 2 minutes. Reverse transcription reactions consist of PCR buffer II (Applied Biosystems), 3 mM $MgCl_2$, 0.5 mM (each) dNTP, 10 mM dithiothreitol, 20 Units ribonuclease inhibitor (Roche Molecular Biochemicals, Indianapolis, IN) and 80 Units SuperScript II Ribonuclease H-reverse transcriptase (GibcoBRL) in a final volume of 40 μL. The reactions are then incubated at 42°C for 50 minutes, followed by inactivation at 70°C for 15 minutes.

Step 3. Real-time PCR amplification of SARS–CoV cDNA is performed using four primer pairs and five molecular beacons in the same reaction, one primer pair for each viral gene (amplification of the internal positive control is enabled by one of the primer pairs that was designed to enable the amplification of one of the for the viral gene targets). Each PCR reaction consists of 20 μL of cDNA products, 1 x PCR buffer II (Applied Biosystems), 3.5 mM $MgCl_2$, 0.5 mM (each) dNTP, 0.4 μM of each primer, and 2.5 Units of AmpliTaq Gold DNA polymerase (Applied Biosystems) in a final volume of 50 μL. Fifty cycles of amplification (94°C for 15 seconds, 53°C for 30 seconds, and 72°C for 30 seconds) are performed in an 7700

Prism spectrofluorometric thermal cycler (Applied Biosystems). For quantitative measurements, duplicates of six-fold serial dilutions (10^6 to 10 copies) of RNA standards are used as quantitative controls along with the samples being tested for a given experimental run. Viral RNA copy number for each clinical sample is calculated by interpolation of the experimentally determined threshold cycle for the test specimen onto a standard regression curve obtained from the control RNA standards (the logarithm of the number of genomic copies present in the clinical sample is inversely proportional to the observed threshold cycle).

Explanation of how the invention solves the problems of the prior art.
The main problem with the prior art is its failure to consider the intrinsically polymorphic nature of coronaviruses and to account for the possibility of continuous and/or multiple introductions of non non-genetically identical SARS–CoV strains into the human population. The present invention solves these problems by using multiple genetic sequences as targets (the S, E, M, and N genes). The primary principle of using four different genetic targets to identify SARS-CoV is to evade the fundamental genetic drift of the virus. Thus, by using four molecular beacons, each specific for a different amplified SARS–CoV target, and each labeled with same colored fluorophore, the likelihood of not detecting SARS–CoV due to continuous evolutionary changes in the virus is minimized. Furthermore, the assay design includes the presence of an internal positive control RNA, the reverse transcription and PCR amplification of which generates a signal in a different color, thus assuring that if there is an unexpected problem with RNA isolation, reverse transcription, or target amplification, the absence of the control signal will indicate that a problem occurred.

Summary of the features of other SARS nucleic acid assays and description of the differences between those assays and the current invention. As outlined in the previous sections, there are five SARS nucleic acid assays: an RT-PCR assay published by Perris et al. [2], an RT-real-time-PCR assay published by Ksiazek et al. [1], a similar RT-real-time-PCR assay published by Drosten et al. [4], another RT-real-time-PCR assay published by Poon et al. [7], and a PCR-based SARS diagnostic assay developed by Lipkin and Briese – no publication is available describing this assay. All of the published assays use RNA extracted from clinical samples. In addition, all of the published assays initiate reverse transcription of SARS–CoV RNA with random primers, and all of the published assays generate cDNA with primers previously designed to enable the amplification of conserved regions of open reading frame 1b

(ORF1b), in order to achieve broad reactivity with the coronavirus genus. The present invention differs from these inventions in many different ways. The present assay uses four different targets instead of one and it has an internal positive RNA control (artificial RNA molecule) that can be reverse-transcribed and amplified with primers designed for the viral genes (it possesses a unique target recognition sequence for detection by a unique molecular beacon). This RNA molecule serves as a control for RNA isolation, reverse transcription, and PCR amplification. The present uses real-time PCR for nucleic acid amplification and molecular beacons for real-time detection. The present employs a dual-color detection scheme, a yellow signal (tetrachlorofluorescein) for all four SARS–CoV targets and a green signal (fluorescein) indicating that the internal positive control has been isolated, reverse transcribed, and PCR amplified.

Another important aspect of our invention, which is not an aspect of the published assays, is the construction of four viral RNA targets that contain gene-specific reverse transcription sequences built in to their 3' ends. Collectively, these molecules serve as SARS-specific positive controls. All of the present designs are thermodynamically compatible to work together in a five-amplicon multiplex assay.

Also multiplex polymerase chain reaction (mPCR) are envisioned. It is a procedure for simultaneously performing PCR on greater than two different sequences. A mPCR reaction comprises: treating said extracted DNA to form single stranded complementary strands, adding a plurality of labelled paired oligonucleotide primers, each paired primer specific for a different short tandem repeat sequence, one primer of each pair substantially complementary to a part of the sequence in the sense strand and the other primer of each pair substantially complementary to a different part of the same sequence in the complementary antisense strand, annealing the plurality of paird primers to their complementary sequences, simultaneously extending said plurality of annealed primers from the 3' terminus of each primer to synthesize an extension product complementary to the strands annealed to each primer, said extension products, after separation from their complement, serving as templates for the synthesis of an extension product for the other primer of each pair, separating said extension products from said templates to produce single stranded molecules, amplifying said single stranded molecules by repeating at least once said annealing, extending and separating steps.

Lower stringent conditions are routinely used to accommodate the capture of multiple target sequences that contain variations in their nucleic acid sequences. The stringency is

reduced by either powering the temperature of hybridization and wash or by modification of the buffer. When the stringent conditions are reduced and the target nucleic acid sequence is very similar to nucleic acid sequences of another genus specificity of the capture probe for the target genus can be lost.

When multiple capture probes are used and are selected to compatible to variations in the target nucleic acid sequences, the specificity under high stringent conditions can be regained. The blending of multiple probes permits a single positive response for the presence of a grou pf target organisms.

In a multiplex assay, numberous conditions of interest are simultaneously examined. Multiplex analysis relies on the ability to sort sample components or the data associated therewith, during or after the assay is completed.

EXAMPLES

EXAMPLE 1-Design of SARS–CoV-specific molecular beacons, and primers for reverse transcription and for PCR

Purpose: The overall rationale in the design of molecular beacons and oligonuclotides for our SARS assay is to construct mismatch-tolerant molecular beacons that are thermodynamically compatible to work in a five-amplicon multiplex assay.

Design: The molecular beacons were designed so that they are able to hybridize to their targets at the annealing temperature of the PCR, while unbound molecular beacons remain in the closed conformation. These basic aspects were achieved by using coronavirus gene-specific multiple alignments and thermodynamic considerations to select the target sequences, the identity and length of the PCR primers, the identity and length of the probe sequences (target recognition sequences), and the length of the arm sequences.

Materials: In order to theoretically calculate the melting temperatures of the PCR primers and the probe-target hybrids, was used the Oligo Toolkit that is available on the internet. The melting temperatures and secondary structure predictions of the molecular beacons were

calculated by using the DNA folding program developed by Michael Zuker that is available on the internet.

Results: The theoretical melting temperatures of the PCR primers was about 60°C; the T_m of the reverse transcription primers was 47°C; the T_m of the probe- target hybrid was about 63°C; and the T_m of the stem hybrids of the molecular beacons was about 70°C.

Conclusion: The mismatch-tolerant molecular beacons that are thermodynamically compatible are designed to work in a final five-amplicon multiplex SARS–CoV assay.

See the following Figures: The molecular designs for the S gene target of the SARS–CoV assay are described in detail in Figures 1-4; for the E gene target see Figures 5-8; for the M gene target, see Figures 9-12; for the N gene target, see Figures 13-16; and for the Internal Positive Control (IPC), see Figures 17 and 18.

Example 2- Experimental characterization of the molecular beacons and the molecular beacon-target complexes

Purpose: The overall rationale of these experiments is to evaluate the thermodynamic properties of the constructed molecular beacons prior to carrying out real-time PCR experiments.

Design: For each molecular beacon, we have determined two melting curves – one for beacon alone, and one for the beacon-target complex – by using the ABI Prism 7700 spectrofluorometric thermal cycler.

Materials: For each molecular beacon, melting curves were obtained by preparing two tubes containing 50 µL of 200 nM molecular beacon dissolved in 3.5 mM $MgCl_2$ and 10 mM Tris-HCl, pH 8.0, and by adding a complementary oligonucleotide target to one of the tubes at a final concentration of 400 nM.

The fluorescence of each solution was determined as a function of temperature, using a thermal cycler with the capacity to monitor fluorescence. Temperature was decreased linearly with time from 80°C to 10°C in 1°C steps; with each holding period lasting one minute, and fluorescence intensity was measured during each hold.

Results: The theoretical melting temperature of the PCR primers was 60°C ± 2°C; the theoretical T_m of the reverse transcription primers was 47 ± 2°C; the theoretical T_m of the probe target-hybrids was about 63 ± 3°C; and the theoretical T_m of the stem hybrid of the molecular beacons was 70 ± 2°C.

Conclusion: The designed mismatch-tolerant molecular beacons was determined to correctly recognize their DNA targets, and to be thermodynamically compatible to work together in a five-amplicon multiplex SARS–CoV assay.

Figures: The melting curves of the molecular beacons and the molecular beacon-target hybrids are shown in Figure 19.

Example 3 - Uniplex SARS–CoV viral and IPC PCR amplifications, using SYBR Green to detect the amplicaons

Purpose: The overall rationale of these experiments is to evaluate the PCR primers and PCR conditions.

Design: For each SARS–CoV gene-specific and IPC amplification, a synthetic target DNA was used. PCR reactions were performed using a spectrofluorometric thermal cycler (Cepheid).

Materials: The PCR protocols are shown in the following exhibits:

for *S* Gene:

(A) SYBR Green-based Detection of *S* Gene Amplicon (LK250) of SARS-associated CoV

Mixture	Per reaction
dH_20	15 µl
10X PCR Buffer (10X)	*2.5 µl*
$MgCl_2$ (25 mM)	4.0 µl
Plat Taq DNA Pol (5 $U\mu l^{-1}$)	0.3 µl
dNTP (25 mM)	0.3 µl
LK251 (10 pmole/µl)	0.5 µl
LK252 (10 pmole/µl)	0.5 µl
Sybr Green DNA (25X)	1.0 µl
Target DNA	1.0 µl
TOTAL	25.0 µl

Smart Cycler (Cepheid)

DNA Denaturation & Enzyme Activation

Cycles: 1		
Target Temperature (° C):	**95**	**120 sec**

Primary Amplification

Cycles: 35		
Denaturation :	**95 ° C**	**15 sec**
Annealing :	**53 ° C**	**15 sec Spectra ON**
Extension:	**72 ° C**	**15 sec**

(B) Molecular-beacon-based Detection of *S* Gene Amplicon (LK250) of SARS-associated CoV

Mixture	Per reaction
dH_20	15.75 µl
10X PCR Buffer (10X)	*2.5 µl*
$MgCl_2$ (25 mM)	4.0 µl
Plat Taq DNA Pol (5 $U\mu l^{-1}$)	0.3 µl
dNTP (25 mM)	0.3 µl
LK251 (10 pmole/µl)	0.5 µl
LK252 (10 pmole/µl)	0.5 µl
LK249 Beacon (10 pmole/µl)	0.25 µl
Target DNA	1.0 µl
TOTAL	25.0 µl

Smart Cycler (Cepheid)

DNA Denaturation & Enzyme Activation

Cycles: 1		
Target Temperature (° C):	**95**	**120 sec**

Primary Amplification

Cycles: 35		
Denaturation :	**95 ° C**	**15 sec**
Annealing :	**53 ° C**	**15 sec Spectra ON**
Extension:	**72 ° C**	**15 sec**

(C) SYBR Green-based Detection of *S* Gene Amplicon (T7- and RT-Amplicon)

Mixture	Per reaction
dH_20	15 µl
10X PCR Buffer (10X)	*2.5 µl*
$MgCl_2$ (25 mM)	4.0 µl
Plat Taq DNA Pol (5 $U\mu l^{-1}$)	0.3 µl
dNTP (25 mM)	0.3 µl
LK251-T7 (10 pmole/µl)	0.5 µl
LK252-RT (10 pmole/µl)	0.5 µl
Sybr Green DNA (25X)	1.0 µl
Target DNA	1.0 µl
TOTAL	25.0 µl

Smart Cycler (Cepheid)

DNA Denaturation & Enzyme Activation

Cycles: 1		
Target Temperature (° C):	**95**	**120 sec**

Primary Amplification

Cycles: 35		
Denaturation :	**95 ° C**	**15 sec**
Annealing :	**53 ° C**	**15 sec Spectra ON**
Extension:	**72 ° C**	**15 sec**

for the *E* Gene:

(A) SYBR Green-based Detection of *E* Gene Amplicon (LK254) of SARS-associated CoV

Mixture	Per reaction
dH_20	15 µl
10X PCR Buffer (10X)	*2.5 µl*
$MgCl_2$ (25 mM)	4.0 µl
Plat Taq DNA Pol (5 $U\mu l^{-1}$)	0.3 µl
dNTP (25 mM)	0.3 µl
LK255 (10 pmole/µl)	0.5 µl
LK256 (10 pmole/µl)	0.5 µl
Sybr Green DNA (25X)	1.0 µl
Target DNA	1.0 µl
TOTAL	25.0 µl

Smart Cycler (Cepheid)

DNA Denaturation & Enzyme Activation

Cycles: 1		
Target Temperature (° C):	**95**	**120 sec**

Primary Amplification

Cycles: 35		
Denaturation :	**95 ° C**	**15 sec**
Annealing :	**53 ° C**	**15 sec Spectra ON**
Extension:	**72 ° C**	**15 sec**

(B) Molecular-beacon-based Detection of *E* Gene Amplicon (LK254) of SARS-associated CoV

Mixture	Per reaction
dH_20	15.75 µl
10X PCR Buffer (10X)	*2.5 µl*
$MgCl_2$ (25 mM)	4.0 µl
Plat Taq DNA Pol (5 $U\mu l^{-1}$)	0.3 µl
dNTP (25 mM)	0.3 µl
LK255 (10 pmole/µl)	0.5 µl
LK256 (10 pmole/µl)	0.5 µl
LK253 Beacon (10 pmole/µl)	0.25 µl
Target DNA	1.0 µl
TOTAL	25.0 µl

Smart Cycler (Cepheid)

DNA Denaturation & Enzyme Activation

Cycles: 1		
Target Temperature (° C):	**95**	**120 sec**

Primary Amplification

Cycles: 35		
Denaturation :	**95 ° C**	**15 sec**
Annealing :	**53 ° C**	**15 sec Spectra ON**
Extension:	**72 ° C**	**15 sec**

(C) SYBR Green-based Detection of *E* Gene Amplicon (T7- and RT-Amplicon)

Mixture	Per reaction
dH_20	15 µl
10X PCR Buffer (10X)	*2.5 µl*
$MgCl_2$ (25 mM)	4.0 µl
Plat Taq DNA Pol (5 $U\mu l^{-1}$)	0.3 µl
dNTP (25 mM)	0.3 µl
LK255-T7 (10 pmole/µl)	0.5 µl
LK256-RT (10 pmole/µl)	0.5 µl
Sybr Green DNA (25X)	1.0 µl
Target DNA	1.0 µl
TOTAL	25.0 µl

Smart Cycler (Cepheid)

DNA Denaturation & Enzyme Activation

Cycles: 1		
Target Temperature (° C):	**95**	**120 sec**

Primary Amplification

Cycles: 35		
Denaturation :	**95** ° C	**15 sec**
Annealing :	**53** ° C	**15 sec Spectra ON**
Extension:	**72** ° C	**15 sec**

for the *M* Gene

(A) SYBR Green-based Detection of *M* Gene Amplicon (LK258) of SARS-associated CoV

Mixture	Per reaction
dH_20	15 µl
10X PCR Buffer (10X)	*2.5 µl*
$MgCl_2$ (25 mM)	4.0 µl
Plat Taq DNA Pol (5 $U\mu l^{-1}$)	0.3 µl
dNTP (25 mM)	0.3 µl
LK259 (10 pmole/µl)	0.5 µl
LK260 (10 pmole/µl)	0.5 µl
Sybr Green DNA (25X)	1.0 µl
Target DNA	1.0 µl
TOTAL	25.0 µl

Smart Cycler (Cepheid)

DNA Denaturation & Enzyme Activation

Cycles: 1		
Target Temperature (° C):	**95**	**120 sec**

Primary Amplification

Cycles: 35		
Denaturation :	**95 ° C**	**15 sec**
Annealing :	**53 ° C**	**15 sec Spectra ON**
Extension:	**72 ° C**	**15 sec**

(B) Molecular-beacon-based Detection of *M* Gene Amplicon (LK258) of SARS-associated CoV

Mixture	Per reaction
dH_20	15.75 µl
10X PCR Buffer (10X)	*2.5 µl*
$MgCl_2$ (25 mM)	4.0 µl
Plat Taq DNA Pol (5 $U\mu l^{-1}$)	0.3 µl
dNTP (25 mM)	0.3 µl
LK259 (10 pmole/µl)	0.5 µl
LK260 (10 pmole/µl)	0.5 µl
LK257 Beacon (10 pmole/µl)	0.25 µl
Target DNA	1.0 µl
TOTAL	25.0 µl

Smart Cycler (Cepheid)

DNA Denaturation & Enzyme Activation

Cycles: 1		
Target Temperature (° C):	**95**	**120 sec**

Primary Amplification

Cycles: 35		
Denaturation :	**95 ° C**	**15 sec**
Annealing :	**53 ° C**	**15 sec Spectra ON**
Extension:	**72 ° C**	**15 sec**

(C) SYBR Green-based Detection of *M* Gene Amplicon (T7- and RT-Amplicon)

Mixture	Per reaction
dH_20	15 µl
10X PCR Buffer (10X)	*2.5 µl*
$MgCl_2$ (25 mM)	4.0 µl
Plat Taq DNA Pol (5 $U\mu l^{-1}$)	0.3 µl
dNTP (25 mM)	0.3 µl
LK259-T7 (10 pmole/µl)	0.5 µl
LK260-RT (10 pmole/µl)	0.5 µl
Sybr Green DNA (25X)	1.0 µl
Target DNA	1.0 µl
TOTAL	25.0 µl

Smart Cycler (Cepheid)

DNA Denaturation & Enzyme Activation

Cycles: 1		
Target Temperature (° C):	**95**	**120 sec**

Primary Amplification

Cycles: 35		
Denaturation :	**95 ° C**	**15 sec**
Annealing :	**53 ° C**	**15 sec Spectra ON**
Extension:	**72 ° C**	**15 sec**

for the *N* Gene:

(A) SYBR Green-based Detection of *N* Gene Amplicon (LK262) of SARS-associated CoV

Mixture	Per reaction
dH_20	15 µl
10X PCR Buffer (10X)	*2.5 µl*
$MgCl_2$ (25 mM)	4.0 µl
Plat Taq DNA Pol (5 $U\mu l^{-1}$)	0.3 µl
dNTP (25 mM)	0.3 µl
LK263 (10 pmole/µl)	0.5 µl
LK264 (10 pmole/µl)	0.5 µl
Sybr Green DNA (25X)	1.0 µl
Target DNA	1.0 µl
TOTAL	25.0 µl

Smart Cycler (Cepheid)

DNA Denaturation & Enzyme Activation

Cycles: 1		
Target Temperature (° C):	**95**	**120 sec**

Primary Amplification

Cycles: 35		
Denaturation :	**95 ° C**	**15 sec**
Annealing :	**53 ° C**	**15 sec Spectra ON**
Extension:	**72 ° C**	**15 sec**

(B) Molecular-beacon-based Detection of *N* Gene Amplicon (LK262) of SARS-associated CoV

Mixture	Per reaction
dH_20	15.75 µl
10X PCR Buffer (10X)	*2.5 µl*
$MgCl_2$ (25 mM)	4.0 µl
Plat Taq DNA Pol (5 $U\mu l^{-1}$)	0.3 µl
dNTP (25 mM)	0.3 µl
LK263 (10 pmole/µl)	0.5 µl
LK264 (10 pmole/µl)	0.5 µl
LK261 Beacon (10 pmole/µl)	0.25 µl
Target DNA	1.0 µl
TOTAL	25.0 µl

Smart Cycler (Cepheid)

DNA Denaturation & Enzyme Activation

Cycles: 1		
Target Temperature (° C):	**95**	**120 sec**

Primary Amplification

Cycles: 35		
Denaturation :	**95 ° C**	**15 sec**
Annealing :	**53 ° C**	**15 sec Spectra ON**
Extension:	**72 ° C**	**15 sec**

(C) SYBR Green-based Detection of *N* Gene Amplicon (T7- and RT-Amplicon)

Mixture	Per reaction
dH_20	15 µl
10X PCR Buffer (10X)	*2.5 µl*
$MgCl_2$ (25 mM)	4.0 µl
Plat Taq DNA Pol (5 $U\mu l^{-1}$)	0.3 µl
dNTP (25 mM)	0.3 µl
LK263-T7 (10 pmole/µl)	0.5 µl
LK264-RT (10 pmole/µl)	0.5 µl
Sybr Green DNA (25X)	1.0 µl
Target DNA	1.0 µl
TOTAL	25.0 µl

Smart Cycler (Cepheid)

DNA Denaturation & Enzyme Activation

Cycles: 1		
Target Temperature (° C):	**95**	**120 sec**

Primary Amplification

Cycles: 35		
Denaturation :	**95 ° C**	**15 sec**
Annealing :	**53 ° C**	**15 sec Spectra ON**
Extension:	**72 ° C**	**15 sec**

and for Internal Positive Control (**IPC**)

(A) SYBR Green-based Detection of *Internal Positive Control* (LK266)

Mixture	Per reaction
dH_20	15 µl
10X PCR Buffer (10X)	*2.5 µl*
$MgCl_2$ (25 mM)	4.0 µl
Plat Taq DNA Pol (5 $U\mu l^{-1}$)	0.3 µl
dNTP (25 mM)	0.3 µl
LK251 (10 pmole/µl)	0.5 µl
LK256 (10 pmole/µl)	0.5 µl
Sybr Green DNA (25X)	1.0 µl
Target DNA	1.0 µl
TOTAL	25.0 µl

Smart Cycler (Cepheid)

DNA Denaturation & Enzyme Activation

Cycles: 1		
Target Temperature (° C):	**95**	**120 sec**

Primary Amplification

Cycles: 35		
Denaturation :	**95 ° C**	**15 sec**
Annealing :	**53 ° C**	**15 sec Spectra ON**
Extension:	**72 ° C**	**15 sec**

(B) Molecular-beacon-based Detection of *Internal Positive Control* (LK266)

Mixture	Per reaction
dH_20	15.75 µl
10X PCR Buffer (10X)	*2.5 µl*
$MgCl_2$ (25 mM)	4.0 µl
Plat Taq DNA Pol (5 $U\mu l^{-1}$)	0.3 µl
dNTP (25 mM)	0.3 µl
LK251 (10 pmole/µl)	0.5 µl
LK256 (10 pmole/µl)	0.5 µl
LK265 Beacon (10 pmole/µl)	0.25 µl
Target DNA	1.0 µl
TOTAL	25.0 µl

Smart Cycler (Cepheid)

DNA Denaturation & Enzyme Activation

Cycles: 1		
Target Temperature (° C):	**95**	**120 sec**

Primary Amplification

Cycles: 35		
Denaturation :	**95 ° C**	**15 sec**
Annealing :	**53 ° C**	**15 sec Spectra ON**
Extension:	**72 ° C**	**15 sec**

(C) SYBR Green-based Detection of *Internal Positive Control* (T7- and RT-Amplicon)

Mixture	Per reaction
dH_20	15 µl
10X PCR Buffer (10X)	*2.5 µl*
$MgCl_2$ (25 mM)	4.0 µl
Plat Taq DNA Pol (5 $U\mu l^{-1}$)	0.3 µl
dNTP (25 mM)	0.3 µl
LK251-T7 (10 pmole/µl)	0.5 µl
LK256-RT (10 pmole/µl)	0.5 µl
Sybr Green DNA (25X)	1.0 µl
Target DNA	1.0 µl
TOTAL	25.0 µl

Smart Cycler (Cepheid)

DNA Denaturation & Enzyme Activation

Cycles: 1		
Target Temperature (° C):	**95**	**120 sec**

Primary Amplification

Cycles: 35		
Denaturation :	**95 ° C**	**15 sec**
Annealing :	**53 ° C**	**15 sec Spectra ON**
Extension:	**72 ° C**	**15 sec**

Results: Real-time PCR amplification curves were consistent with the presence or absence of target DNA and the amplicons that were synthesized in the PCR reactions had the correct lengths.

Conclusion: The synthesized primers work well under uniform PCR conditions.

Example 4- Uniplex SARS–CoV viral and IPC PCR amplifications, using gene-specific molecular beacons

Purpose: The overall rationale of these experiments is to evaluate the molecular beacons, using the uniform PCR condition established in the tests with the PCR primers.

Design: For each SARS–CoV gene-specific and IPC amplification, a synthetic target DNA and molecular beacon are used. PCR reactions were performed using the spectrofluorometric thermal cycler (Cepheid). For each assay, utilized target dilutions are peformed to establish the linearity and the dynamic range of the molecular beacon-based real-time PCR assays.

Materials: The PCR protocols are shown above in Example 3 for the S Gene, E gene, M Gene, N Gene, and the Internal Positive Control (IPC).

Results: Real-time PCR amplification curves were consistent with the presence or absence of target DNA.

Conclusion: The SARS–CoV-specific and IPC-specific molecular beacons work well in under uniform PCR conditions. See Figure 20 (Exhibit 25).

Example 5- Uniplex SARS–CoV viral and IPC PCR amplifications using bacteriophage T7 RNA polymerase and reverse transcriptase target PCR primers

Purpose: The overall rationale of these experiments is for each of the five amplicons (SARS–CoV S, E, M, and N genes and IPC) to produce dsDNA molecules that contain a

bacteriophage T7 promoter target recognition sequence at their 5' ends and the gene-specific reverse transcriptase primer-binding site at their 3' ends. The synthesized amplicons are used for in vitro RNA production. The synthesized RNA molecules have the specific reverse transcriptase primer-binding site at their 3' ends.

Design: For each SARS–CoV gene-specific and IPC amplification, a synthetic target DNA and bacteriophage T7-RT-PCR primers are used. PCR reactions were performed using the spectrofluorometric thermal cycler (Cepheid), using SYBR green to detect the amplicons.

Materials: The PCR protocols are shown in Example 3.

Results: Real-time PCR amplification curves are consistent with the presence or absence of target DNA and the amplicons synthesized in the PCR reactions had the correct lengths.

Conclusion: SARS–CoV-specific bacteriophage T7-RT amplicons were generated for use as templates for in vitro RNA transcription, in order to produce the target RNAs that will be used to qualify the PCR assays. See Figure 20.

Example 6- In vitro RNA transcription of SARS–CoV- and IPC-specific RNA molecules

Purpose: The overall rationale for these experiments is to produce RNA molecules containing the specific reverse transction primer-binding site at their 3' ends.

Design: For each SARS–CoV gene-specific and IPC amplification, a T7-RT-PCR dsDNA amplicon (generated from Experiment 5) is used.

Materials: RNA transcripts corresponding to the four SARS–CoV-specific and IPC alleles were prepared by in vitro transcription of PCR products that contain the T7 RNA polymerase promoter site, by using a MEGAscript T7 kit (Ambion, Houston, TX).

Results: RNA molecules had the correct lengths, based on 8% polyacrylamide gel electrophoretic analysis of the product strands.

Conclusion: SARS–CoV-specific T7-RT amplicons were generated to be used for in vitro RNA transcription.

Example 7- Uniplex SARS–CoV-specific and IPC-specific RNA reverse transcription and molecular beacon-based real-time PCR

Purpose: The overall rationale of these experiments is to test whether the RNA molecules containing the specific reverse transcription primer-binding site at their 3' ends can be reverse transcribed and the generated cDNA can be amplified and detected by real-time PCR.

Design: For each SARS–CoV gene-specific and IPC amplification, a T7-RT-PCR dsDNA amplicon (generated from Experiment 5) is used.

Materials: Reverse transcription with viral and IPC RNA was performed by adding 20 µL of SARS–CoV-specific and IPC-specific RNA to 0.125 µM of SARS–CoV-specific and IPC-specific primers and incubated at 80°C for 5 minutes. The PCR-tubes were then immediately placed on ice for at least 2 minutes. the reverse transcriptase reactions contained PCR buffer II (Applied Biosystems), 3 mM $MgCl_2$, 0.5 mM of each dNTP (GibcoBRL), 10 mM dithiothreitol, 20 Units ribonuclease inhibitor (Roche Molecular Biochemicals) and 80 Units SuperScript II RNase H-reverse transcriptase (GibcoBRL) in a final volume of 40 µL. The reactions were incubated at 42°C for 50 minutes, followed by inactivation at 70°C for 15 minutes. Real-time PCR of cDNA, using 10 µL cDNA, 1 x PCR buffer, 3.5 mM $MgCl_2$, 0.5 mM of each dNTP, 0.4 µM of each molecular beacon, 0.4 µM of each PCR primer, and 2.5 Units AmpliTaq Gold DNA polymerase in a final volume of 50 µL. 35 cycles of amplification (94°C for 15 sec, 53°C for 30 sec, and 72°C for 30 sec) were performed in a 7700 Prism spectrofluorometric thermal cycler (Applied Biosystems).

Results: SARS–CoV-specific and IPC-specific RNA molecules can be reverse transcribed, and the generated cDNA can be amplified and detected by SARS–CoV-specific and IPC-specific molecular-beacons in real-time PCR assays.

Conclusion: SARS–CoV-specific RNA molecules can be used as controls in the final assay.

Example 8- Multiplex SARS–CoV-specific and IPC-specific RNA reverse transcription and molecular beacon-based real-time PCR

Purpose: The overall rationale of this experiment is to multiplex the uniplex SARS–CoV-specific and IPC-specific RT-PCR reactions in a multiplex RT-PCR reaction.

Design: All five RT-PCR reactions are incorporated into one.

Materials: Identical to Experiment 7, with the exception that all five RT primers, four sets of PCR primers, and five molecular beacons are used in a single RT-PCR reaction

Results: SARS–CoV-specific and IPC-specific RNA molecules can be reverse transcribed and generate cDNA, and can then be amplified and detected by SARS–CoV-specific and IPC-specific molecular beacons in a real-time PCR assay.

Conclusion: SARS–CoV-specific RNA molecules can be used as controls in a multiplex assay.

Example 9- Evaluate the complete SARS–CoV assay using clinical samples

Purpose: The overall rationale of these experiments is to evaluate the SARS–CoV assay using real samples (isolated SARS–CoV from cultures, primary patient isolates from saliva, and whole blood and stool specimens).

Design: Identical to Experiment 8.

Materials: Identical to Experiment 8.

Results: SARS–CoV assay detects SARS–CoV and discriminates between SARS–CoV and other non-pathogenic coronaviruses.

Conclusion: SARS–CoV assay detects SARS RNA extracted from cultured SARS strains and primary isolates.

Example 10-Clinical Evaluation

64 samples collected from 23 individuals who were clinically diagnosed to have SARS based on CDC, WHO and Health Canada case definitions.

The sample types included: bronchial lavage and sputnum (pre-mortem samples) as well as lung, liver, small and large bowel, and spleen tissue (post mortem).

65 samples from 15 patients served as controls. 26 were bone marrow samples sent for routing pathogen screening. 39 were post mortem organ samples from 10 patitnes who died during the outbreak but whose deathes were attributed to other causes including congestive heart failure, cerebrovascular accidents, atherosclerotic heart disease, chornic obstructive pulmonary disease, invasive Group A streptococcal infection, amiodorone pulmonary toxicity, and pulmonary fibrosis.

All samples were blindl assayed by usng the developed assay (usng all four genes: S, E, M, N) and the specificity was 100%.

1. Ksiazek, T.G., et al., A novel coronavirus associated with severe acute respiratory syndrome. N Engl J Med, 2003. 348(20): p. 1953-66.
2. Peiris, J.S., et al., Coronavirus as a possible cause of severe acute respiratory syndrome. Lancet, 2003. 361(9366): p. 1319-25.
3. Munch, R., Robert Koch. Microbes Infect, 2003. 5(1): p. 69-74.
4. Drosten, C., et al., Identification of a novel coronavirus in patients with severe acute respiratory syndrome. N Engl J Med, 2003. 348(20): p. 1967-76.
5. Marra, M.A., et al., The Genome sequence of the SARS-associated coronavirus. Science, 2003. 300(5624): p. 1399-404.
6. Rota, P.A., et al., Characterization of a novel coronavirus associated with severe acute respiratory syndrome. Science, 2003. 300(5624): p. 1394-9.
7. Poon, L.L., et al., Rapid diagnosis of a coronavirus associated with severe acute respiratory syndrome (SARS). Clin Chem, 2003. 49(6 Pt 1): p. 953-5.
8. Organization, W.H., http://www.who.int/csr/sars/primers/en/.
9. Pensaert, M., P. Callebaut, and J. Vergote, Isolation of a porcine respiratory, non-enteric coronavirus related to transmissible gastroenteritis. Vet Q, 1986. 8(3): p. 257-61.
10. Ruan, Y.J., et al., Comparative full-length genome sequence analysis of 14 SARS coronavirus isolates and common mutations associated with putative origins of infection. Lancet, 2003. 361(9371): p. 1779-85.

The content of each of the document identified above and cited in the specification documents is expressly incorporated herein by reference.

Claims:

1. In an assay for coronaviruses wherein the improvement comprises detectants specific for at least two genes selected from the group consisting of spike (S), envelope (E), membrane (M), and nucleocapsid (N).

2. The assay of claim 1 wherein the corona virus is a SARS virus.

3. The assay of claim 2 wherein the SARS virus is SARS–CoV.

4. The assay of claim 1 wherein the assay is a molecular-beacon-based multi-allelic RT-real-time-PCR assay.

5. The assay of claim 4 wherein the detectant is a molecular-beacon.

6. The assay of claim 5 wherein the molecular beacon is a mismatch-tolerant molecular beacons.

7. The assay of claim 1 wherein the assay is a PCR assay.

8. The assay of claim 7 wherein there are four sets of PCR primers for four different viral genes, and four different molecular beacons.

9. The assay of claim 8 wherein each molecular beacon is labeled with the same fluorophore.

10. The assay of claim 8 wherein each molecular beacon is specific for a different SARS–CoV gene.

11. The assay of claim 7 wherein the improvement further comprises an exogenous RNA standard that is added to the sample that can be reverse-transcribed and amplified by one of the primer sets.

12. The assay of claim 11 wherein the improvement further comprises the inclusion of a fifth molecular beacon that is labeled with a different fluorophore that is specific for the exogenous RNA standard.

13. The assay of claim 1 wherein the improvement further comprises RNA isolation from clinical samples

14. The assay of claim 13 wherein the sample includes blood, tissue, sputum, or nasopharyngeal aspiration samples.

15. The assay of claim 1 or 12 wherein the improvement further includes reverse transcription, PCR amplification and simultaneous automated amplicon detection in a spectrofluorometric thermal cycler that measures the fluorescence intensity of each color during the annealing phase of each thermal cycle.

16. The assay of claim 1 wherein the improvement further comprises the selection and detection of four target sequences that uniquely identify SARS–CoV.

17. The assay of claim 1 wherein the improvement further comprises including specific amplicon, primers and IPC standard along with the detectant for each gene.

18. The assay of claim 17 wherein the primers, amplicon and detectant include those having the formulas shown in figures 3, 7, 11 and 15 for each gene and an IPC standard selected from those shown in figure 17.

19. The assay of claim 17 wherein the primers, amplicon and detectant include those having the formulas shown in figures 4, 8, 12 and 16 for each gene and an IPC standard selected from those shown in figure 18.

20. The assay of claim 1 wherein the improvement further comprises a PCR multiplex format.

21. A kit for performing the assay of any one of claims 1-20.

22. The kit of claim 21 wherein the reagents are contained in individual containers.

23. The kit of claim 22 wherein the reagents are sufficient in amount to perform an assay of single sample.

24. The kit of claim 21 wherein the kit contains written instructions.

25. A reagent composition comprising one or more of the following reagent groups

S Gene

LK249 5'-FAM-CCCACGCCAGAAGGTAGATCACGAACTACACGTGGG-3'-Dubcyl

LK249.N 5'-FAM-GCCCACGCCAGAAGGTAGATCACGAACTACACGTGGGC-3'-Dubcyl

LK250

5'-CTCTATGTTTATAAGGGCTATCAACCTATAGATGTAGTTCGTGATCTA
CCTTCTGGTTTTAACACTTTGAAACCTATTTTTAAGTTGCCTCTTGG -3'

LK251 5'-CTCTATGTTTATAAGGGCTATCAACC-3'

LK251-T7 5'-TAATACGACTCACTATAGGCTCTATGTTTATAAGGGCTATCAACC-3'

LK252 5'-CCAAGAGGCAACTTAAAAATAGGTTTC- 3'

LK252-RT 5'- AGGCTGTAAGAACCAAGAGGCAACTTAAAAATAGGTTTC- 3'

S-RT 5'-AGGCTGTAAGAA-3';

E Gene

LK253 5'-FAM-CCTCCGCACGAAAGCAAGAAAAAGAAGTACGCCGGAGG-3'-Dubcyl

LK253.N 5'-FAM-GCCTCCGCACGAAAGCAAGAAAAAGAAGTACGCCGGAGGC-3'-Dubcyl

LK254 5'-CGGAAGAAACAGGTACGTTAATAGTTAATAGCGTACTTCTTTTTC
TTGCTTTCGTGGTATTCTTGCTAGTCACACTAGCCATCCTTACTGCGCTT -3'

LK255 5'- CGGAAGAAACAGGTACGTTAATAG -3'

LK255-T7 5'-TAATACGACTCACTATAGGCGGAAGAAACAGGTACGTTAATAG -3'

LK256 5'- AAGCGCAGTAAGGATGGCTA - 3'

LK256-RT5'- TATTGCAGCAGTACAAGCGCAGTAAGGATGGCTA - 3'

E-RT 5' –TATTGCAGCAGTAC-3';

M Gene

LK257 5'-FAM-CCTCCGACCCAATTAATTCTGTAGACAGCAGCCGGAGG-3'-Dubcyl

LK257.N 5'-FAM-GCCTCCGACCCAATTAATTCTGTAGACAGCAGCCGGAGGC-3'-Dubcyl

LK258

5'-CTTGTTTTCCTCTGGCTCTTGTGGCCAGTAACACTTGCTTGTTTTGTGCT TGCTGCTGTCTACAGAATTAATTGGGTGACTGGCGGGATTGCGATTGCAAT GGCTTG-3'

LK259 5'- CTTGTTTTCCTCTGGCTCTTG -3'

5'- TAATACGACTCACTATAGGCTTGTTTTCCTCTGGCTCTTG -3'

LK260 5'- CAAGCCATTGCAATCGCAATC- 3'

LK260-RT5'- AAGCAACGAAGTAGCAAGCCATTGCAATCGCAATC- 3'

M-RT 5'-AAGCAACGAAGTAG -3' ;or

IPC

LK265 5'-FAM-GCCCACGTACCATCTGGGGCTGTAGACAGCAGCCGTGGGC-3'-Dubcyl

LK266

5'-CTCTATGTTTATAAGGGCTATCAACCTATAGATGCTGCTGTCTACAGC

CCCAGATGGTAGTATTCTTGCTAGTCACACTAGCCATCCTTACTGCGCTT-3'

LK251 5'-CTCTATGTTTATAAGGGCTATCAACC-3'

LK251-T7 5'-TAATACGACTCACTATAGGCTCTATGTTTATAAGGGCTATCAACC-3'

LK256 5'- AAGCGCAGTAAGGATGGCTA - 3'

LK256-RT 5'- TATTGCAGCAGTACAAGCGCAGTAAGGATGGCTA - 3'

E-RT 5' –TATTGCAGCAGTAC-3'.

26. A kit containing the reagent composition of claim 25.

Figure 1: CLUSTAL W (1.81) multiple DNA sequence alignment of coronavirus *S* genes from strains isolated from different species

```
CcoV          ATGATTGTGCTTACATTGTGCCTTTTCTTGTT---TTTGTACAGTAGTGTGAGCTGTACA
FcoV          ATGATTGTGCTCGTAACTTGCCTCTTGTTGTTATGTTCATACCACACAGTTTTGAGTACA
TGE           ATGAAAAAACTATTTGTGGTTTTGGTCGTAATGCCATTGATTTATGGAGACAATTTTCCT
PRCoV         ------------------------------------------------------------
HCoVOC43      ------------------------------------------------------------
PEDV          ------------------------------------------------------------
SARSUrba      ------------------------------------------------------------
SARSTor2      ------------------------------------------------------------
BcoV          ------------------------------------------------------------
HEV           ------------------------------------------------------------
MHV           ------------------------------------------------------------
RtCoV         ------------------------------------------------------------
IBV           ------------------------------------------------------------

CcoV          TCAAACAATGACTGTGTACAAGTTAATGTGACACAACTGCCTGGCAATGAAAATATTATC
FcoV          ACAAATAATGAATGCATACAAGTTAACGTAACACAATTGGCTGGCAATGAAAACCTTATC
TGE           TGTTCTAAATTGACTAATAGAACTATAGGCAACCAGTGGAATCTCATTGAAACCTTCCTT
PRCoV         ------------------------------------------------------------
HCoVOC43      ------------------------------------------------------------
PEDV          ----------------------------ATGAGGTCTTTAATTTACTTCTGGTTGCTCTT
SARSUrba      ------------------------------------------------------ATGTTT
SARSTor2      ------------------------------------------------------ATGTTT
BcoV          ------------------------------------------------ATGTTTTTGATACTT
HEV           ------------------------------------------------ATGTTTTTTATACTT
MHV           ------------------------------------------------ATGCTATTCGTGTTT
RtCoV         ------------------------------------------------ATGCTATTCGTGTTT
IBV           ------------------------------------------------------------

CcoV          AAAGATTTTCTATTTCAGAACTTTAAAGAAGAAGGAAGTTTAGTTGTTGGTGGTTATTAC
FcoV          AGAGATTTTCTGTTTAGTAACTTTAAAGAAGAAGGAAGTGTAGTTGTTGGTGGTTATTAC
TGE           CTAAACTATAGTAGTAGGTTACCACCTAATTCAGATGTGGTGTTAGGTGATTATTTTCCT
PRCoV         ------------------------------------------------------------
HCoVOC43      ------------------------------------------------------------
PEDV          ACCAGTACTTCCAACACTCAGCCTACCACAAGATGTCACTAGGTGCCAGTCTACTACTAA
SARSUrba      ATTTTCTTATTATTTCTTACTCTCACTAGTGGTAGTGACCTTGACCGGTGCACCACTTTT
SARSTor2      ATTTTCTTATTATTTCTTACTCTCACTAGTGGTAGTGACCTTGACCGGTGCACCACTTTT
BcoV          TTAATTTCCTTACCAATGGCTTTTGCTGTTATAGGAGATTTAAAGTGTACTACGGTTTCC
HEV           TTAATCTCCCTGCCTTCTGCTTTTGCAGTTATAGGGGATTTAAAGTGTACTACTTCATTA
MHV           TTAACCTTGTTGCCCTCTTCTCTAGGGTATATTGGTGATTTTAGATGTATCCAACTTGTA
RtCoV         TTAACCCTATTGCCCTCTTGTCTAGGGTATATTGGTGATTTTAGATGTATCAACCTTGTA
IBV           ------------------------------------------------------------

CcoV          CCCACAGAGGTGTGGTATAACTGTTCCACAACTCAACAAACTACCGCTTATAAGTATTTT
FcoV          CCTACAGAGGTGTGGTACAACTGCTCTAGAACAGCTCGAACTACTGCCTTTCAGTATTTT
TGE           ACTGTACAACCTTGGTTTAATTGCATTCGCAATGATAGTAATGACCTTTATGTTACACTG
PRCoV         ------------------------------------------------------------
HCoVOC43      ------------------------------------------------------------
PEDV          CTTTAGGCGGTTCTTTTCAA--AATTTAATGTTCAGGCACCTGCCGTCGTCGTTTTGGGT
SARSUrba      GA-----------TGATGTTCAAGCTCCTA--ATTACACTCAACATACTTCAT----CT
SARSTor2      GA-----------TGATGTTCAAGCTCCTA--ATTACACTCAACATACTTCAT----CT
BcoV          ATTAATGATGT---TGACACCGGTGCTCCCTCTATTAGCACTGATATTGTCGATGTTACT
HEV           ATTAATGACGT---TGACACTGGTGTGCCATCTATTAGCTCTGAAGTTGTTGATGTCACT
MHV           AATACCGACACCTCTAATGCCAGCGCTCCAAGCGTTAGTACAGAGGTAGTTGATGTTTCC
```

```
RtCoV       AACACCCGCATTTCTAATGCGCGCGCACCCAGTGTTAGCACAGAGGTAGTTGATGTTTCT
IBV         ------------------------------------------------------------

CcoV        AGTAATATACATGCATTTTATTTTGATATGGAAGCCATGGAGAATAGTACTGGCAATGCA
FcoV        AATAATATACATGCCTTTTATTTTGTTATGGAAGCCATGGAAAATAGCACTGGTAATGCA
TGE         GAAAATCTTAAAGCATTGTATTGGGATT---ATGCTACAGAAAATATCACTTGGAAT---
PRCoV       ------------------------------------------------------------
HCoVOC43    ------------------------------------------------------------
PEDV        GGTTACCTACCTAGTATGAACTCTTCTAGCTGGTACTGTGGCACAGGCATTGAAACTGCT
SARSUrba    ATGAG---GGGGGTTTACTATCCT---GATGAAATT--TTTAGATCAGACACTCTT----
SARSTor2    ATGAG---GGGGGTTTACTATCCT---GATGAAATT--TTTAGATCAGACACTCTT----
BcoV        AATGGTTTAGGTACTTATTATGTTTTAGATCGTGTGTATTTAAATACTACGTTGTTG---
HEV         AATGGTTTGGGGACTTTCTATGTTTTAGATCGTGTCTATTTAAATACCACATTGTTG---
MHV         AAAGGGATTGGTACTTATTATGTTTTAGATCGAGTCTATTTAAATGCCACACTATTG---
RtCoV       AAAGGTCTTGGTACATATTACGTTTTAGATCGTGTTTATTTAAATGCCACGTTATTG---
IBV         ------------------------------------------------------------

CcoV        CGTGGTAAACCTTTACTAGTACATGTTCATGGTAATCCTGTTAGTATCATTGTTTACATA
FcoV        CGTGGTAAACCATTATTATTTCATGTGCATGGTGAGCCTGTTAGTGTTATT------ATA
TGE         CACAGACAACGGTTAAACGTAGTCGTTAATGGATACCCATACTCCATCACAGTT---ACA
PRCoV       ------------------------------------------------------------
HCoVOC43    ------------------------------------------------------------
PEDV        AGTGGCGTTCATGGTATTTTTCTCAGCTACATCGATTCTGGTCAGGGCTTTGAGA--TTG
SARSUrba    -----TATTTA--ACTCAGGATTTATTTCTTC-CATTTTATTCTAATGTTACAGG--GTT
SARSTor2    -----TATTTA--ACTCAGGATTTATTTCTTC-CATTTTATTCTAATGTTACAGG--GTT
BcoV        CTTAATGGTTACTACCCTACTTCAGGTTCTACATATCGTAATATGGCACTGAAGG--GAA
HEV         CTCAATGGTTATTACCCAATTTCAGGTGCTACATTTCGTAATATGGCTCTGAAAG--GAA
MHV         CTTACTGGTTATTACCCTGTAGATGGGTCCATGTATAGAAACATGGCTCTAACGG--GAA
RtCoV       CTTACTGGTTACTACCCTGTAGATGGGTCCATGTATCGTAACATGGCTCTAATGG--GTA
IBV         ------------------------------------------------------------

CcoV        TCAGCTTATAGAGATGATGTGCAATTTAGGCCGCTTTTAAAGCATGGTTTATTGTGTATA
FcoV        TCGGCTTATAGGGATGATGTGCAACAAAGGCCCCTTTTAAAACATGGGTTAGTGTGCATA
TGE         ACAACCCGCAATTTTAAT-TCTGCTGAAGGTGCTATTATATGCATTTGTAAGGGCTCACC
PRCoV       ------------------------------------------------------------
HCoVOC43    ------------------------------------------------------------
PEDV        GCATTTCGCAAGAGCCGT-TTGATCCTAGTGGTTACCAGCTTTATTTACATAAGGCCACT
SARSUrba    TCATACTATTAATCATAC----GTTTGG--CAACCCTGTCAT-ACCTTTTAAGGATGGTA
SARSTor2    TCATACTATTAATCATAC----GTTTGG--CAACCCTGTCAT-ACCTTTTAAGGATGGTA
BcoV        CTTTACTATTGAGCAGACTATGGTTTAAACCACCTTTTCTTTCTGATTTTATTAATGGTA
HEV         CTCGATTATTGAGCACCTTGTGGTTTAAGCCGCCTTTTTTATCACCTTTTAATGATGGTA
MHV         TTAATACCATAAGCCTTAATTGGTACAAACCACCCTTTTTATCAGAGTTTAATGATGGCA
RtCoV       CTAATACCTTAAGCCTTAATTGGTTTGAACCGCCCTTTTTATCAGAGTTTAACGATGGCA
IBV         ------------------------------------------------------------

CcoV        ACTAAAAATGACACCGTTGACTATAATAGCTTTACAATTAACCAATGGCGAGACATATGT
FcoV        ACTAAAAATCGCCATATTAACTATGAACAATTCACCTCCAACCAGTGGAATTCCACATGT
TGE         ACCTACTACCACCACAGAATCTA-----GTTTGACTTGCAATTGGGGTAGTGAGTGCAGG
PRCoV       ------------------------------------------------------------
HCoVOC43    ------------------------------------------------------------
PEDV        AATGGTAACACTAATGCTATTGCACGACTGCGCATTTGCCAGTTTCCCGATAATAAAACA
SARSUrba    TTTA-TTTTGCTGCCACAGAGAAATCAAATGTTGTCCGTGGTTGGGTTTTTGGTTCTACC
SARSTor2    TTTA-TTTTGCTGCCACAGAGAAATCAAATGTTGTCCGTGGTTGGGTTTTTGGTTCTACC
BcoV        TTT----TTGCTAAGGTCAAAAATACCAAGGTTATTAAAAAGGGTGTAATGTATAGTGAG
HEV         TTT----TTGCCAAGGTTAAAAACAGCAGATTTTCTAAAGATGGTGTTATTTATAGTGAG
MHV         TAT----TTGCTAAGGTAAAGAACCTTAAAGCATCTTTGCCCAAAGATTCTATTTCATAT
RtCoV       TAT----ATGCTAAGGTAAAGAACCTCAAAGCATCTTTGCCCATAGGCTCGGCTTCATAC
IBV         ------------------------------------------------------------
```

```
CcoV          TTGGGTGACGACAGAAAAATACCATTCTCTGTAGTACCCACAGATAATGGTACGAAATTA
FcoV          ACGGGTGCTGACAGAAAAATTCCTTTCTCTGTCATACCCACGGACAATGGAACAAAAATC
TGE           TTAAACCATAAGTTCCCTATATGTCCTTCTAATTCAGAGGCAAATTGTGGTAATATGCTG
PRCoV         ------------------------------------------------------------
HCoVOC43      ------------------------------------------------------------
PEDV          TTGGGCCCTACTGTTAATGATGTTACAACAGGTCGTAACTGCCTATTCAACAAAGCCATT
SARSUrba      ATGAACAACA-AGTCACAGTCGGTGATTA--TTATTAACAATTCTACTAATGTTGTTATA
SARSTor2      ATGAACAACA-AGTCACAGTCGGTGATTA--TTATTAACAATTCTACTAATGTTGTTATA
BcoV          TTTCCTGCTATAACTATAGGTAGTACTT---TTGTAAATACATCCTATAGTGTGGTAGTA
HEV           TTTCCTGCTATTACTATAGGTAGTACTT---TTGTAAATACTTCCTATAGCATAGTAGTA
MHV           TTCCCTACTATAATTATAGGTAGTAATT---TTGTCACCACTTCCTATACTGTAGTATTG
RtCoV         TTTCCTACTATAATTATAGGTAGTAATT---TTGTTAATACTTCCTATACTGTAGTATTG
IBV           ------------------------------------------------------------

CcoV          TTTGGTCTTGAGTGGAATGATGACTATGTTACAGCCTATATTAGTGATGAGTCTCACCGT
FcoV          TATGGTCTTGAGTGGAATGATGACTTTGTTACAGCTTATATTAGTGGTCGTTCTTATCAC
TGE           TATGGCCTACAATGGTTTGCAGATGAGGTTGTTGCTTATTTACATGGTGCTAGTTACCGT
PRCoV         ------------------------------------------------------------
HCoVOC43      ------------------------------------------------------------
PEDV          CCAGCTTATATGCGTGATGGAAAAGATATTGTTGTCGGCATAACATGGGATAATGATCGT
SARSUrba      CGAGCATGTA------------ACTTTGAAT-------TGTGTGACAACCCTTTCTTTGC
SARSTor2      CGAGCATGTA------------ACTTTGAAT-------TGTGTGACAACCCTTTCTTTGC
BcoV          CAACCACATACTACCAATTTGGATAATAAATTACAAGGTCTCTTAGAGATCTCTGTTTGC
HEV           GAGCCTCATACCTCACTTATTAATGGTAATTTACAAGGTTTGTTGCAAATTTCTGTTTGT
MHV           GAACCGTATA------------ATGGTA-----------TAATTAT-GGCATCCATTTGC
RtCoV         GAACCATACA------------ATGGTA-----------TTATTAT-GGCATCTATTTGC
IBV           ------------------------------------------------------------

CcoV          TTGAATATCAATAATAATTGGTTTAACAATGTTACACTC--CTATACTCACGTACAAGCA
FcoV          TTGAACATCAATACTAATTGGTTTAACAATGTCACACTT--TTGTATTCACGCTCAAGCA
TGE           ATTAGTTTTGAAAATCAATGGTCTGGCACTGTCACATTTGGTGATATGCGTGCGACAACA
PRCoV         ------------------------------------------------------------
HCoVOC43      ------------------------------------------------------------
PEDV          GTCACTGTTTTTGCTGACAAGATCTATCATTTTTATCTT--AAAAATGATTGGTCCCGCG
SARSUrba      TGTTTCTAAACCCATGGGTACACAGACACATACTATGAT----ATTCGA----TAATGCA
SARSTor2      TGTTTCTAAACCCATGGGTACACAGACACATACTATGAT----ATTCGA----TAATGCA
BcoV          CAGTATACTATGTGCGAGTACCCACATACGATTTGTCATCCTAAGCTGGG---TAATAAA
HEV           CAATACACTATGTGTGAATACCCACATACTATTTGTCATCCTAATTTGGG---TAATCAA
MHV           CAGTATACCATTTGTCAACTACCGTACACGGATTGCAAACCGAATACGGGCGGTAATAAG
RtCoV         CAGTATACCATTTGTCAATTACCGCACACGGATTGCAAACCTAACACGGGCGGTAACACG
IBV           ------------------------------------------------------------

CcoV          CCGCCACGTGGCAACACA-GTGCTGCATATGTTTA---TCAAGGTGTTTCAAATTTTACT
FcoV          CTGCTACCTGGGAATACA-GTGCTGCATATGCTTA---CCAAGGTGTTTCTAACTTCACT
TGE           TTAGAAGTCGCTGGCACGCTTGTAGACCTTTGGTGGTTTAATCCTGTTTATGATGTCAGT
PRCoV         -------------------ATGAAAAAATTATTTG---TGGTCTTGGTTGTAATGCCATT
HCoVOC43      ------------------------ATGTT-TGTT-TTGCTTGTTGC--ATATGCCTTGT
PEDV          TTGCGACAAGATGTTACAATCGCAGAAGTTGTGCT-ATGCAATATGTTTATACACCTACC
SARSUrba      TTTAATTGCACTTTCGAGTACATATCTGATGCCTTTTCGCTTGATGTTTCAGAA-----A
SARSTor2      TTTAATTGCACTTTCGAGTACATATCTGATGCCTTTTCGCTTGATGTTTCAGAA-----A
BcoV          CG-CGTAGAACTATGGCATTGGGATACAGGTGTTGTTTCCTGTTTATATAAGCG-----T
HEV           CG-CATAGAATTATGGCATTATGACACAGATGTTGTTTCTTGTTTATACAGGCG-----T
MHV           TT-AATTGGCTTTTGGCACACAGAGCTAAAATCCCCTGTGTGCATTTTAAAGCG-----T
RtCoV         CT-AATTGGTTTTTGGCACACAGATTTAAGGCCTCCGGTGTGCATTTTAAAGCG-----T
IBV           ------------ATGTTGGTAACACCTCTTTTACTAGTGACTCTTTTGTGTGCACTATGT
                                                                  *

CcoV          TATTACAAGTTAAATAAAACCGCTGGCTTAAAAAGCTATGAATTGTGTGAAGATTATG-A
FcoV          TATTACAAGTTAAATAACACCAATGGTCTAAAAACCTATGAATTATGTGAAGATTATG-A
```

```
TGE          TATTATAGGGTTAATAATAAAAATGGT------ACTACCGTAGTTTCCAATTGCACTG-A
PRCoV        GATTTATGGA-----GACAAGTTTCCT------ACTTCCGTAGTTTCCAATTGCACTG-A
HCoVOC43     TGCATATTGCTGGTTGTCAA-ACTACAAATGGGCTGAACACTAGTTACT--CTGTTTGCA
PEDV         TACTACATGCTTAATGTTACTAGTGCAGGTGAGGATGGCATTTATTATGAACCCTGTACA
SARSUrba     AGT----CAGGTAATTTTAAAC---ACTTACGAGAGTTTGTGTTTAAAAATAAAGATGGG
SARSTor2     AGT----CAGGTAATTTTAAAC---ACTTACGAGAGTTTGTGTTTAAAAATAAAGATGGG
BcoV         AATTTCACATATGATGTGAATGCTGATTACTTGTATTTCCATTTTTATCAAGAAGGTGGT
HEV          AATTTCACATATGATGTGAATGCTGATTATTTATATTTTCACTTTTATCAGGAAGGTGGC
MHV          AATTTTACGTTTAATGTTAATGCCGAATGGCTTTATTTTCATTTTTACCAGCAGGGTGGT
RtCoV        AATTTTACGTTTAATGTTAATGCCGAATGGCTTTATTTTCATTTTTACCAGCAGGGTGGT
IBV          AGTGCTGTTTTGTATGACAGTAGTTCTTACGTTTACTACTACCAAAGTGCCTTCAGACCA
                                  *

CcoV         AT--ACTGCACTGGCTATGCAACCAATGTGTTTGCTCCGACATCAGGTGGTTATATACCT
FcoV         AC--ATTGCACTGGCTATGCTACCAATGTATTTGCTCCGACATCAGGTGGTTACATACCT
TGE          TC--AATGTGCTAGTTATGTGGCTAATGTTTTTACTACACAGCCAGGAGGTTTTATACCA
PRCoV        TC--AATGTGCTAGTTATGTGGCTAATGTTTTTACTATACTACCAGGAGGCTTTATACCA
HCoVOC43     AC-GGCTGTGTTGGTTATTCAGAAAATGTATTTGCTGTTGAGAGTGGTGGTTATATACCC
PEDV         GCTAATTGCACTGGTTACGCTGCCAATGTATTTGCCACTGATTCCAATGGCCATATACCA
SARSUrba     TTTCTCTATG-----TTTATAAGGGCTAT--------CAACCTATAGATGTAGTTCGTGA
SARSTor2     TTTCTCTATG-----TTTATAAGGGCTAT--------CAACCTATAGATGTAGTTCGTGA
BcoV         ACTTTTTATGCATATTTTACAGACACTGGTGTTGTTACTAAGTTTCTGTTTAATGTTTAT
HEV          ACTTTTTATGCATACTTTACAGATACTGGTTTTGTGACCAAGTTTCTGTTTAAGTTGTAT
MHV          ACTTTTTATGCGTATTATGCGGATGTTTCTTCTGCTACTACGTTTTTGTTTAGTATGTAT
RtCoV        ACTTTTTATGCGTATTATGCAGATGTTTCTTCTGCCACTACGTTTTTGTTTAGTTCGTAT
IBV          CCTAGTGGTTGGCATTTACAAGGGGGTGCTTATGCGGTAGTTAACATTTCTAGCGAATTT
                              *          *

CcoV         GATGGATTCAGTTTTAACAATTGGTTTATGCTTACAAACAGCTCCACTTTTGTTAGTGGC
FcoV         GATGGATTTAGTTTTAACAATTGGTTCTTGCTTACAAATAGTTCCACTTTTGTTAGTGGC
TGE          TCAGATTTTAGTTTTAATAATTGGTTCCTTCTAACTAATAGCTCCACGTTGGTTAGTGGT
PRCoV        TCAGATTTTAGTTTTAATAATTGGTTCCTCCTAACTAATAGCTCCACGTTGGTTAGTGGT
HCoVOC43     TCCGACTTTGCATTCAATAATTGGTTCCTTCTAACTAATACCTCATCTGTTGTAGATGGT
PEDV         GAAGGTTTTAGTTTTAATAATTGGTTTCTTTTATCCAATGACTCCACTTTGTTGCATGGT
SARSUrba     TCTACCTTCTGGTTTTAACACT-TTGAAACCTATTTTTAAGTTGCCTCTT----------
SARSTor2     TCTACCTTCTGGTTTTAACACT-TTGAAACCTATTTTTAAGTTGCCTCTT----------
BcoV         TTAGGCACGGTGCTTTCACATTATTATGTCCTGCCTTTGACTTGTTCT------------
HEV          TTAGGCACTGTGCTGTCACATTATTATGTTATGCCATTGACTTGTAAT------------
MHV          ATTGGTGATGTGTTAACACAATATTTTGTGTTGCCTTATATGTGTACTCTCACTACAACA
RtCoV        ATTGGTGCTGTGTTAACACAGTATTTTGTGTTGCCTTATATGTGTAGTCCCACTACCTCA
IBV          AATAATGCAGGCTCTTCATCAGGGTGTACTGTTGGTATTATTCATGGTGGTCGTGTTGTT
                                    *      *

CcoV         AGATTTGTAACAAATCAACCGCTGCTAGT--TAATTGCTTGTGGCCAGTGC--CCAGTTT
FcoV         AGGTTTGTAACAAATCAACCATTATTGAT--TAATTGCTTGTGGCCAGTGC--CCAGTTT
TGE          AAATTAGTTACCAAACAGCCGTTATTAGT--TAATTGCTTATGGCCAGTCC--CTAGCTT
PRCoV        AAATTAGTTACCAAACAGCCTCTATTAGT--TAATTGCTTATGGCCAGTCC--CTAGCTT
HCoVOC43     GTTGTGAGGAGTTTTCAGCCTTTGTTGCT--TAATTGCTTATGGTCTGTTT--CT--GGC
PEDV         AAAGTGGTTTCCAACCAACCCTTGTTGGT--CAATTGTCTTTTGGCCATTC--CTAAGAT
SARSUrba     GGTATTAACATTACAAATTTTAGAGCCATTCTTACAGCCTTTTCACC-TGC--TCAAGAC
SARSTor2     GGTATTAACATTACAAATTTTAGAGCCATTCTTACAGCCTTTTCACC-TGC--TCAAGAC
BcoV         AGTGCTATGACTTTAGAATATTGGGTTACACCTCTCACTTCTAAACAATAT--TTACTAG
HEV          AGCGCTTTATCTTTAGAATACTGGGTTACACCTCTCACTACTAGACAATTT--CTTCTAG
MHV          GGTGTCTTTTCACCGCAGTATTGGGTTACACCTCTTGTCAAGCGCCAATAT--TTATTTA
RtCoV        GGTGTTTCCTCACCGCAGTATTGGGTTACACCACTTGTTAAGCGCCAATAT--TTATTTA
IBV          AATGCTTCTTCTATAGCTATGACGGCACCGTCATCAGGTATGGCTTGGTCTAGCAGTCAG
                                                              *

CcoV         TGGCGTCGCAGCACAAGAATT--TTGTTTTGAAGGTGCTCAGTTTAGCCAATGTAACGGT
FcoV         TGGTGTAGCAGCACAAGAATT--TTGTTTTGAAGGTGCACAGTTTAGCCAATGTAATGGT
TGE          TGAAGAAGCAGCTTCTACATT--TTGTTTTGAGGGTGCTGGCTTTGATCAATGTAATGGT
PRCoV        TGAAGAAGTAGCTTCTACATT--TTGTTTTGAAGGTGCTGACTTTGATCAATGTAATGGT
HCoVOC43     TTGCGGTTTACTACTGGTTTTGTCTATTTTAATGGTACTGGGAGAGGTGATTGTAAAGGT
```

```
PEDV          TTATGGACTAG-GCCAATTTT-TCTCATTCAATCACACGATGGATGGCGTTTGTAATGGA
SARSUrba      ATTTGGGGCAC-GTCAGC--TGCAGCCTATTTTGTTGGCTATTTAAAGCCAACTACATTT
SARSTor2      ATTTGGGGCAC-GTCAGC--TGCAGCCTATTTTGTTGGCTATTTAAAGCCAACTACATTT
BcoV          CTTTCAATCAA-GATGGTGTTATTTTTAATGCTGTTGATTGTAAGAGTGATTTTATGAGT
HEV           CCTTTGACCAG-GATGGTGTTTTATACCATGCTGTTGATTGTGCTAGTGATTTTATGAGT
MHV           ATTTTAATCAA-AAGGGTATTATTACTAGTGCTGTTGATTGTGCTAGTAGTTATACCAGC
RtCoV         ATTTTAACCAA-AAGGGTATTATTACTAGCGCTGTTGATTGTGCTAGTAGTTATACCAGT
IBV           TTTTGTACTGCACACTGTAATTTTTCAGATACTACAGTGTTTGTTACACATTGTTATAAA
                                    *                                   *

CcoV          GTTTCTTTAAATAATACAGTAGATGTTATTAGATTTAACCTTAATTTCACTACAGATGTA
FcoV          GTGTCTTTAAATAACACAGTGGATGTTATTAGATTCAACCTTAATTTCACTGCAGATGTA
TGE           GCTGTTTTAAATAATACTGTAGACGTCATTAGGTTCAACCTTAATTTTACTACAAATGTA
PRCoV         GCTGTTTTAAATAACACTGTAGACGTCATTAGGTTTAACCTTAATTTTACTACAAATGTA
HCoVOC43      TTTTCCTCAGATGTTTTGTCTGATGTCATACGTTACAACCTCAATTTTGAA------GAA
PEDV          GCTGCTGTGGATCGTGCCCCAGAGGCTCTGAGGTTTAATATTAATGACACCTC---CGTC
SARSUrba      ATGCTCAAGTATGATGAAAATGGTACAATCACAGATGCTGTTGATTGT-----------T
SARSTor2      ATGCTCAAGTATGATGAAAATGGTACAATCACAGATGCTGTTGATTGT-----------T
BcoV          GAGATTAAGTGTAAAACACTATCTATAGCACCATCTACTGGTGTTTATGAATTAAACGGT
HEV           GAGATTATGTGTAAAACTTCTTCAATTACACCACCTACTGGTGTTTATGAACTAAACGGT
MHV           GAAATAAAGTGTAAGACCCAAAGTATGAATCCCAATACGGGAGTTTATGATTTATCCGGT
RtCoV         GAAATAAAGTGTAAGACTCAAAGTATGAATCCCAATACGGGAGTCTATGATTTATCCGGT
IBV           CATGGTGGGTGTCCTTTAACTGGCATGCTTCAACAGAATCTTATACGT---------GTT
                    *

CcoV          CAATCTGGCATGGGTGCTACAGTATTTTCACTGAATACAACAGGCGGTGTCATTCTTGAG
FcoV          CAATCTGGTATGGGTGCTACAGTATTTTCACTGAATACAACAGGTGGTGTCATTCTTGAA
TGE           CAATCAGGTAAGGGTGCCACAGTGTTTTCATTGAACACAACGGGTGGTGTCACTCTTGAA
PRCoV         CAATCAGGTAAGGGTGCTACAGTGTTTTCATTGAACACAACGGGTGGTGTCACTCTTGAA
HCoVOC43      AACCTTAGACGTGGAACCATTTTGTTT------AAAACATCTTATGGTGTTGTTGTGTTT
PEDV          ATTCTTGCTGAAGGCTCAATTGTACTT------CATACTGCTTTAGGAACAAATCTTTCT
SARSUrba      CTCAAAATCCA-CTTGCTGAACTCAAATGCT---CTGTTAAGAGCTT---TGAGATTGAC
SARSTor2      CTCAAAATCCA-CTTGCTGAACTCAAATGCT---CTGTTAAGAGCTT---TGAGATTGAC
BcoV          TACACTGTTCAGCCAATTGCAGATGTTTACCGACGTATACCTAATCTTCCCGATTGTAAT
HEV           TACACAGTTCAACCTGTTGCCACTGTATATCGTAGAATACCTGATTTACCCAATTGCGAT
MHV           TACACCGTCCAACCTGTAGGATTAGTGTACCGGCGTGTTAGAAATTTGCCTGATTGTAAA
RtCoV         TACACCGTCCAACCTGTAGGACTAGTGTACCGGCGTGTTAGAAATTTGCCTGATTGTAAA
IBV           TCTGCTATGAAAAATGGCCAGCTTTTCTATAATTTAACAGTTAGTGTAGCTAAGTACCCT

CcoV          ATTTCTTGTTATAATGACACAGTGAGTGAGTCGAGT-TTCTACAGTTATGGTGAAATTCC
FcoV          ATTTCATGTTATAGTGACACAGTGAGTGAGTCTAGT-TCTTACAGTTATGGTGAAATCCC
TGE           ATTTCATGTTATA------CAGTGAGTGACTCGAGC-TTTTTCAGTTACGGTGAAATTCC
PRCoV         ATCTCATGTTATAATGATACAGTGAGTGATTCGAGC-TTTTCCAGTTACGGTGAAATTCC
HCoVOC43      TATTGTACCAACAACACTTTAGTTTC------AGGTGATGCTCACATACCATTTGGTACA
PEDV          TTTGTTTGCAGTAATTCCTCAGATCCTCATTTAGCCATCTTTGCCATACCTCTGGGTGCT
SARSUrba      AAAGGAATTTACCAGACCTCTAATTTCAGGGTTGTTCC-CTCAGGAGATGTTGTGAGATT
SARSTor2      AAAGGAATTTACCAGACCTCTAATTTCAGGGTTGTTCC-CTCAGGAGATGTTGTGAGATT
BcoV          ATAGAGGCTTGGCTTAATGATAAGTCGGTGCCCTCTCCATTAAATTGGGAACGTAAGACC
HEV           ATCGAAGCTTGGCTTAATTCTAAGACCGTTTCTTCGCCTCTTAATTGGGAACGTAAAATT
MHV           ATTGAGGAATGGCTAACTGCTAAGTCTGTACCTTCTCCTCTCAATTGGGAGCGCAAAACA
RtCoV         ATTGAGGAATGGTTGGCTGCTAACACAGTACCCTCTCCTCTCAATTGGGAGCGCAAAACA
IBV           ACTTTTAGATCATTTCAGTGTGTTAATAATTTAACATCCGTATATTTAAATGGTGATCTT

CcoV          ATTCGGCGTAACTGATG-GACCACGTTACTGTTATGTACTCT-ACAATGGCACAGCTCTT
FcoV          GTTCGGCATAACTGACG-GACCACGATACTGTTATGTACTTT-ACAATGGCACAGCTCTT
TGE           GTTCGGCGTAACTGATG-GACCACGGTACTGTTACGTACACT-ATAATGGCACAGCTCTT
PRCoV         GTTCGGCGTAACTAATG-GACCACGGTACTGTTACGTACTCT-ATAATGGCACAGCTCTT
HCoVOC43      GTTTTGGGCAATTTTT---ATTGCTTTGTAAATACTACTATTGGCAATGAAACTACGTCT
PEDV          ACTGAAGTACCCTACT---ATTGCTTTCTTAAAGTGGATACTTACAACTCCACTGTTTAT
SARSUrba      CCCTAATATTACAA-----ACTTGTGTCCTTTTGGAGAGGTTTTTAATGCTACTAAATTC
SARSTor2      CCCTAATATTACAA-----ACTTGTGTCCTTTTGGAGAGGTTTTTAATGCTACTAAATTC
```

```
BcoV          TTTTCAAATTGTAATTTTAATATGAGCAGCCTGATGTCTTTTATTCAGGCAGACTCATTT
HEV           TTTTCTAATTGTAATTTTAACATGGGCAGGCTGATGTCTTTTATTCAGGCTGACTCTTTT
MHV           TTTCAAAATTGTAACTTCGACCTGAGCAGTCTATTAAGATTTGTTCAGGCTGAGTCACTC
RtCoV         TTTCAAAATTGTAACTTCAACCTGAGCAGTCTATTAAGATTTGTTCAGGCTGAGTCACTC
IBV           GTTTACACCTCTAATGAGACCATAGATGTTACATCTGCAGGTGTTTATTTTAAAGCTGGT
                                                        *     *

CcoV          AAGTATT-TAGGAACATTACCACCTAGTGTCAAGGAAATTGCTATTAGTAAGTGG-----
FcoV          AAATATT-TAGGAACATTACCACCCAGTGTAAAGGAAATTGCTATTAGTAAGTGG-----
TGE           AAGTATT-TAGGAACATTACCACCTAGTGTCAAGGAGATTGCTATTAGTAAGTGG-----
PRCoV         AAGTATC-TAGGAACATTACCACCTAGTGTCAAGGAGATTGCTATTAGTAAGTGG-----
HCoVOC43      GCTTTTG-TGGGTGCACTACCTAAGACAGTTCGTGAGTTTGTTATT-TCACGCACA----
PEDV          AAATTCT-TGGCTGTTTTACCTCCTACTGTCAGGGAAATTGTCATCACCAAGTAT-----
SARSUrba      CCTTCTG------TCTATGCATGGGAGAGAAAAAAAAT---TTCTAATTGTGTTGCT---
SARSTor2      CCTTCTG------TCTATGCATGGGAGAGAAAAAAAAT---TTCTAATTGTGTTGCT---
BcoV          ACTTGTAATAATATTGATGCTGCTAAGATATATGGTATGTGTTTTTCCAGCATAACTATA
HEV           GGTTGTAACAATATTGATGCTTCTCGCTTATATGGTATGTGTTTTGGTAGCATTACTATT
MHV           TCATGTAGTAATATAGATGCTTCCAAGGTTTATGGTATGTGCTTTGGTAGTATATCTATA
RtCoV         TCATGTAGTAATATAGATGCTTCCAAGGTTTATGGAATGTGCTTTGGTAGCATATCTATA
IBV           GGACCTA-TAACTTATAAAGTTATGAGAGAAGTTAAA---GCCCTGGCTTATTTT-----
                                                      *

CcoV          GGACATTTTTATATTAATGGTTACAATTTCTTTAGCACGTTTCCTATTGATTGTATAGCT
FcoV          GGCCATTTTTATATTAATGGTTACAATTTCTTTAGCACATTTCCTATTGGTTGTATATCT
TGE           GGCCATTTTTATATTAATGGTTACAATTTCTTTAGCACATTTCCTATTGATTGTATATCT
PRCoV         GGCCATTTTTATATTAATGGTTACAATTTCTTTAGCACATTTCCTATTGATTGTATATCT
HCoVOC43      GGACATTTTTATATTAATGGCTATCGCTATTTCACTTTAGGTAATGTAGAAGCCGT----
PEDV          GGTGATGTTTATGTCAATGGGTTTGGCTATTTGCATCTCGGTTTGTTGGATGCTGTCACA
SARSUrba      GATTACTCTGTGCTCTACAAC--TCAACATTTTTTTCAACCTTTAAGTGCTATGGCGTTT
SARSTor2      GATTACTCTGTGCTCTACAAC--TCAACATTTTTTTCAACCTTTAAGTGCTATGGCGTTT
BcoV          GATAAGTTTGCTATACCCAATGGTAGGAAGGTTGACCTACAATTGGGCAATTTGGGCTAT
HEV           GATAAGTTTGCTATACCCAATAGTAGAAAGGTTGATCTGCAAGTGGGTAAATCTGGTTAT
MHV           GACAAGTTTGCGATACCCAATAGACGCCGAGTTGATTTGCAGCTAGGCAACTCTGGGTTT
RtCoV         GATAAATTTGCAATACCCAACAGTCGCCGTGTTGATCTTCAGCTAGGTAAATCGGGTCTT
IBV           GTTAATGGTACTGCACAAGATGTTATTTTGTGTGATGGATCACCTAGAGGCTTGTTAGCA
              *   *   *

CcoV          TTTAATTTAA-----------CCACTGGTGCTAGTGGAGCAT-TTTGGACAATTGCTTA
FcoV          TTTAATTTAA-----------CCACTGGTGTTAGTGGAGCTT-TTTGGACAATTGCTTA
TGE           TTTAATTTGA-----------CCACTGGTGATAGTGACGTTT-TCTGGACAATAGCTTA
PRCoV         TTTAATTTGA-----------CTACTGGTGATAGTGACGTCT-TCTGGACAATAGCTTA
HCoVOC43      --TAATTTCA------ATGTCACTACTGCAGAAACCACTGATT-TTTGTACTGTTGCGTT
PEDV          ATTAATTTCACTGGTCATGGCACTGACGATGACGTTTCAGGTT-TCTGGACCATAGCATC
SARSUrba      CTGC---CAC-----------TAAGTTGAATGATCTTTGCTT-CTCCAATGTCTATGCA
SARSTor2      CTGC---CAC-----------TAAGTTGAATGATCTTTGCTT-CTCCAATGTCTATGCA
BcoV          TTGCAGTCTT-----------TTAACTATAGAATTGATACTA-CTGCTACAAGT-TGTC
HEV           TTACAATCTT-----------TTAATTATAAGATTGACACTG-CTGTTAGCAGT-TGTC
MHV           TTGCAATCCT-----------TTAATTACAAAATAGATACAA-GAGCTACTTCG-TGTC
RtCoV         TTGCAATCTT-----------TTAATTATAAAATTGATACAA-GAGCGACCTCG-TGTC
IBV           TGCCAGTATA-----------ATACTGGCAATTTTTCAGATGGCTTTTATCCTTTTACT
                                                          *

CcoV          TACG-TCGTACACAGAAGCATTAGTACAA-GTTGAAAACACAGCTATTAAAAAGGTGACG
FcoV          CACA-TCGTATACTGAAGCATTAGTACAA-GTTGAAAACACAGCTATTAAAAATGTGACG
TGE           CACA-TCGTACACTGAAGCATTAGTACAA-GTTGAAAACACAGCTATTACAAAGGTGACG
PRCoV         CACA-TCGTACACTGAAGCATTAGTACAA-GTTGAAAACACAGCTATTACAAATGTGACG
HCoVOC43      AGCT-TCTTATGCTGACGTTTTGGTTAAT-GTGTCACAAACCTCTATTGCTAATATAATT
PEDV          GACT-AATTTTGTTGATGCACTCATCGAG-GTTCAAGGAACTTCCATTCAGCGTATTCTT
SARSUrba      GATTCTTTTGTAGTCAAG--GGAGATGAT-GTAAG-ACAAATAGCGCCAGGACAAACTGG
SARSTor2      GATTCTTTTGTAGTCAAG--GGAGATGAT-GTAAG-ACAAATAGCGCCAGGACAAACTGG
BcoV          AGTTGTATTATAATTTACCTGCTGCTAAT-GTTTCTGTTAGCAGGTTTAATCCTTCTACT
HEV           AACTCTATTATAGTTTGCCTGCAGCAAAC-GTATCTGTCACTCATTATAATCCTTCATCT
MHV           AGCTCTATTATAGTCTTGCAAAAAATAAT-GTCACTGTCAATAACCATAACCCGTCCTCT
```

```
RtCoV          AGCTCTATTACAGTCTTGCACAAGATAAT-GTCACTGTCATTAACCACAACCCATCCTCC
IBV            AATAGTAGTTTAGTTAAGCAGAAGTTTATTGTCTATCGTGAAAATAGTGTTAATACTACT
                       *                    *  **

CcoV           TATT-GTAACAGTCAC--ATTAATAACATCAAATGTTCTCAACTTACTG----CTAATTT
FcoV           TATT-GTAACAGTCAC--ATTAATAACATTAAATGTTCTCAACTTACTG----CTAATTT
TGE            TATT-GTAATAGTCAC--GTTAATAACATTAAATGCTCTCAAATTACTG----CTAATTT
PRCoV          TATT-GTAATAGTTAT--GTTAATAACATTAAATGCTCTCAACTTACTG----CTAATTT
HCoVOC43       TATT-GCAACTCTGTT--ATTAACAGACTGAGATGTGACCAGTTGTCCT----TTGATGT
PEDV           TATT-GTGATGATCCT--GTTAGCCAACTCAAGTGTTCTCAGGTTGCTT----TTGACCT
SARSUrba       TGTT-ATTGCTGATTATAATT------ATAAATTGCCAGATGATTTCAT------GG---
SARSTor2       TGTT-ATTGCTGATTATAATT------ATAAATTGCCAGATGATTTCAT------GG---
BcoV           TGGA-ATAGGAGATTTGGTTTTACAGAACAATTTGTTTTTAAGCCTCAACCTGTAGGTGT
HEV            TGGA-ATAGAAGGTATGGGTTT----AATAATCAGAGTTTTGGTTCCAG-----AGG---
MHV            TGGA-ACAGGCGTTATGGGTTC----AATGATGTGGCTACATTTGGAAC-----TGG---
RtCoV          TGGA-ATAGGCGTTATGGATTT----AATGACGTGGCTACATTTCATAG-----TGG---
IBV            TGTACGTTACACAATTTCATTTTT--CATAATGAGACTGGCGCCAACCC-----------
               *                   **         *   *

CcoV           GCAAAATGGTTTTTACCCTGTTG---CTTCAAGTGAAGTTGGTCTTGTCAATAAGAGTGT
FcoV           GAATAATGGATTTTATCCTGTTG---CTTCAAGTGAAGTAGGTTTCGTTAATAAGAGTGT
TGE            GAATAATGGATTTTATCCTGTTT---CTTCAAGTGAAGTTGGTCTTGTCAATAAGAGTGT
PRCoV          GAATAATGGATTTTATCCTGTTT---CTTCAAGTGAAGTTGGTTCTGTCAATAAGAGTGT
HCoVOC43       ACCAGATGGTTTTTATTCTACAA---GCCCTATTCAATCCGTTGAGCTACCTGTGTCTAT
PEDV           TGACGATGGTTTTTACCCCATCT---CTTCTAGA-AACCTTCTGAGTCACGAACAGCCAA
SARSUrba       --------GTTGTGTCCTTGCTTGGAATACTAGGAACATTGATGCTACTTCAACTGGTAA
SARSTor2       --------GTTGTGTCCTTGCTTGGAATACTAGGAACATTGATGCTACTTCAACTGGTAA
BcoV           TTTTACTCATCATGATGTTGTTT---ATGCACAACATTGTTTTAAAGCTCCCAAAAATTT
HEV            ------CCTTCATGATGCTGTTT---ATTCACAGCAATGTTTTAATACACCTAACACATA
MHV            ------TAAACATGACGTTGCTT---ATGCTGAGGCTTGTTTTACCGTGGGAGCATCATA
RtCoV          ------TGAACATGACGTTGCTT---ATGCAGAGGCATGTTTCACTGTTGGAGCTTCATA
IBV            ------TAATCCTAGTGGTGTTC---AGAATATTCAAACTTACCAAACAAAAACAGCTCA
                     *

CcoV           TGTGTTACTACCTAGTTTCTATTC---ACATACCAGTGTTAATATAACTATTGATCTTGG
FcoV           TGTGTTATTACCTAGCTTTTTCAC---ATACACCGCTGTCAATATAACCATTGATCTTGG
TGE            TGTGTTACTACCTAGCTTTTACAC---ACATACCATTGTTAACATAACTATTGGTCTTGG
PRCoV          TGTGTTACTACCTAGCTTTCTGAC---ACATACCATTGTTAACATAACTATTGGTCTTGG
HCoVOC43       TGTGTCGCTACCTGTTTATCATAA---ACATACGTTTATTGTGTTGTACGTTGACTTCAA
PEDV           TTTCTT-TTGTTACTTTGCCATCA---TTTAATGATCATTCTTTTGTTAAT--ATTACTG
SARSUrba       T-TATAATTATAAAT---ATAG-------------GTATCTTAGACATGGCAAGCTTAG
SARSTor2       T-TATAATTATAAAT---ATAG-------------GTATCTTAGACATGGCAAGCTTAG
BcoV           C-TGTCCGTGTAAATTGGATGG------GTCTTTGTGTGTAGGTAATGGTCCTGGTATAG
HEV            T-TGTCCTTGTAGA----ACAA------GTCA--ATGCATAGGTGGTG---CAGGCACAG
MHV            T-TGCCCTTGCGCGAACCCCAGCATAGTGTCGCCATGTACCACTGGAAAACCTAACTTTG
RtCoV          T-TGCCCTTGTGCGAAGCCCAGCACAGTCTATTCATGTGTCACAGGTAAACCTAAGTCTG
IBV            G-AGTGGTTATTATAATTTTAA------TTTTTCCTTTCTGAGTAGTTTTGTTTATAAGG
                      *

CcoV           TATGAAGCGTAGTG-TTACGGTCA--CCATAGCCTCACCATTAAGTAACATCACACTACC
FcoV           TATGAAGCTTAGTGGTTATGGTCAACCCATAGCCTCGACACTAAGTAACATCACACTACC
TGE            TATGAAGCGTAGTGGTTATGGTCAACCCATAGCCTCAACATTAAGTAACATCACACTACC
PRCoV          TATGAAGCGTAGTGGTTATGGTCAACCCATAGCCTCAACGCTAAGTAACATTACACTACC
HCoVOC43       ACCTCAGAGTGGCGGTGGCAAGTG-------------CTTTAACTGTTATCCTGCT---
PEDV           TCTCT------GCGGCTTTTGGTG-------------GTCTTAGTAGTGCCAATCT---
SARSUrba       GCCCTT-TGAGAGAGACATATCTAAT-----GTGCCTTTCT----------CCCCTGA--
SARSTor2       GCCCTT-TGAGAGAGACATATCTAAT-----GTGCCTTTCT----------CCCCTGA--
BcoV           ATGCTGGTTATAAAAATAGTGGTATAGGCACTTGTCCTGCA--GGTACTAATTATTTAAC
HEV            GAACTTGTCCTGTAGGCACCACTGTGCGCAAGTGTTTTGCT--G----CAGTTACAAAA-
MHV            CCAATTGCCCTACAGGCACCTCGAATCGTGAGTGCACTGTTATGCCATTGGCTAATAAT-
RtCoV          CTAATTGCCCAACAGGTACCTCGAATCGTGAGTGTAATGTTCAGGCTTCAGGTTTTAA--
IBV            AGTCTAATTTTATGTATGGATCTTAT----------CACCCAAGTTGTAAATTTAGAC-
```

```
CcoV        AATGCAGGATAATAACATAGACGTGTACTGTATTCGTTCTAACCAATTCTCAGTTTATGT
FcoV        AATGCAGGATAACAATACTGATGTGTACTGTATTCGTTCTAACCAATTCTCAGTTTATGT
TGE         AATGCAGGATCACAACACCGATGTGTACTGTATTCGTTCTGACCAATTTTCAGTTTATGT
PRCoV       AATGCAGGATAACAACAACGATGTGTACTGTGTTCGTTCTGACCAATTTTCAGTTTATGT
HCoVOC43    --GGTGTTAATATTACACTGGCCAATTTT--AATGAAACTAAAGGGCCTT---TGTGTGT
PEDV        --CGT-TGCATCTGACACT----ACTATC--AATGGGTTTA---GTTCTT---TCTGTGT
SARSUrba    ------------TGGCAAACCTTGCACCCCACCTGCTCTTA---ATTGT------TATTG
SARSTor2    ------------TGGCAAACCTTGCACCCCACCTGCTCTTA---ATTGT------TATTG
BcoV        TTGCCATAATGCTGCCCAATGTGATTGTTTGTGCACTCCCG---ACCCCATTACATCTAA
HEV         ----------GCTACTAAGTGTACTTGCTGGTGTCAACCAG---ATCCTTCCACATATAA
MHV         ----------CAATTTAAGTGTGATTGCACTTGTAACCCTA---GTCC-------TCTAA
RtCoV       ------------GTCTAAGTGCGATTGCACATGTAACCCTA---GTCC-------TCTAA
IBV         ------------TAGAAACTATTAATAATGGCTTGTGGTTTA--ATTC------ACTTTC
                                                                    !   *

CcoV        TCATTCCACTTGCAAAAGTTCTTTATGGGATAACAATTTTAATTCAGCATGTACCGACGT
FcoV        TCATTCCACTTGCAAAAGTTCTTTATGGGACAATATTTTTAATCAAGACTGCACGGATGT
TGE         TCATTCTACTTGCAAAAGTGCTTTATGGGACAATATTTTTAAGCGAAACTGCACGGACGT
PRCoV       TCATTCTACTTGCAAAAGTGTTTTATGGGACAATGTTTTTAAGCGAAACTGCACGGACGT
HCoVOC43    TGACACATC-----ACACTTCACTACCAAATACGTTGCT-------GTTTATGCCAATGT
PEDV        TGACACTAG-----ACAATTTACCATTACACTGTTTTAT--------AATGTTACAAACAG
SARSUrba    GCCATTAAATGATTATG---GTTTTTACAC------CACTACTGGCATTGGCTACCAACC
SARSTor2    GCCATTAAATGATTATG---GTTTTTACAC------CACTACTGGCATTGGCTACCAACC
BcoV        ATCTACAGGGCCTTACAAGTGCCCCCAAACTAAATACTTAGTTGGCATAGGTGAGCACTG
HEV         AGGTGTAAACGCCTGGACTTGTCCGCAATCTAAAGTTTCTATACAACCAGGTCAGCATTG
MHV         CCACCTATGATCTTAGA--TGTCTCCAAGCAAGAAGCATGCTTGGCGTAGGTGATCATTG
RtCoV       CCACCTATGATCCTAGA--TGTCTTCAAGCGCGGAGCATGCTTGGCGTAGGTGATCATTG
IBV         AGTTTCAATTGCTTACGGTCCTCTTCAAGGTGGTTGCAA----GCAATCTGTCTTTAAAG
                                                                    *

CcoV        TT-TAGACGCCACAGCTGTTATAAAAACTGGTACTTGTC---CTTTCTCATTTGATAAAT
FcoV        TT-TAGAGGCTACAGCTGTTATAAAAACTGGTACTTGTC---CTTTCTCATTTGATAAAT
TGE         TT-TAGATGCCACAGCTGTTATAAAAACTGGTACTTGTC---CTTTCTCATTTGATAAAT
PRCoV       TT-TAGATGCCACAGCTGTTATAAAAACTGGTACTTGTC---CTTTCTCATTTGATAAAT
HCoVOC43    TGGTAGGTGGAGTGCTAGTATTA-ACACGGGAAATTGCC---CTTTTTCTTTTGGCAAAG
PEDV        TTAT-GGTTATGTGTCTAAATCACAGGATAGTAATTGTC---CTTTCACCTTGCAATCTG
SARSUrba    TT-ACAG------AGTTGTAGTACTTTCTTTTGA--------ACTTTTAAATGCACCGGC
SARSTor2    TT-ACAG------AGTTGTAGTACTTTCTTTTGA--------ACTTTTAAATGCACCGGC
BcoV        TT-CGGGTCTTGCTATTAAAAGTGATTATTGTGGAGGTA---ATCCTTGTACTTGCCAAC
HEV         CC-CTGGTTTGGGTCTTGTGGAGGATGATTGCTCTGGCA---ACCCTTGCACTTGTAAAC
MHV         TG-AAGGTCTAGGAGTTTTAGAAGATAAATGTGGTGGCAGCAACACCTGCAATTGTTCTG
RtCoV       TG-AAGGTCTAGGTATTTTAGAAGATAAATGTGGTGGCAGCAACATATGCAATTGTTCGG
IBV         GT--AGAGCAACTTGTTGTTATGCTTATTCATATGGAGG---------TCCTTCGCTGTG

CcoV        TGAATAATTACTTAACTTTTAACAAGTTCTGTTTGTCGTTGAATCCCGTTGGTGCCAACT
FcoV        TGAACAATTACTTGACTTTTAACAAGTTCTGTTTGTCGTTGAGTCCTGTTGGTGCTAATT
TGE         TGAACAATTACTTAACTTTTAACAAGTTCTGTTTGTCGTTGAGTCCTGTTGGTGCTAATT
PRCoV       TGAACAATTACTTAACTTTTAACAAGTTCTGTTTGTCGTTGAGTCCCGTTGGTGCTAATT
HCoVOC43    TTAATAACTTTGTTAAATTTGGCAGTGTATGTTTTTCGCTAAAGGATATACCCGGTGGTT
PEDV        TTAATGATTACCTGTCTTTTAGCAAATTTTGTGTTTCAACCAGCCTTTTGGCTGGTGCTT
SARSUrba    CACGG---TTTGTGGACCAAAAT-TATCCACTGACCTTATTAAGAACCAGTGTGTCAATT
SARSTor2    CACGG---TTTGTGGACCAAAAT-TATCCACTGACCTTATTAAGAACCAGTGTGTCAATT
BcoV        CACAAGCATTTTTGGGTTGGTCTGTTGACTCTTGTTTACAAGGGGATAGGTGTAATATTT
HEV         CACAGGCTTTCATAGGCTGGAGTTCAGAAACTTGTTTGCAAAATGGTAGGTGTAATATTT
MHV         CTCATGCCTTTGTTGGCTGGGCTAAGGATAGTTGCTTGGCTAATGGCCGCTGTCACATTT
RtCoV       CTGATGCCTTTGTTGGCTGGGCTATGGACAGCTGTCTATCTAATGCCCGCTGCCATATTT
IBV         TAAAGGTGTTTATTCAGGTGAGTTAGATCATAATTTTGAATGTGGACTGTTA-GTTTATG
                    *   *

CcoV        GTAAGTTAGATGTTGCCGCCCGTACAAGAACCAATGAGCAGGTTTTTGGAAGTT---TAT
FcoV        GCAAGTTTGATGTTGCTGCACGTACAAGAACCAATGAGCAGGTTGTTAGAAGTC---TAT
```

```
TGE          GTAAGTTTGATGTAGCTGCCCGTACAAGAACCAATGAGCAGGTTGTTAGAAGTT---TGT
PRCoV        GTAAGTTTGATGTAGCTGCCCGTACAAGAACCAATGATCAGGTTGTTAGAAGTT---TGT
HCoVOC43     GCGCAATGCCTATAGTGGCTAATTGGGCTTATAGTAAGTACTATACTATAGGCTCATTGT
PEDV         GTACCATAGATCTTTTTGGTTACCCTGCGTTCGGTAGTGGTGTTAAGTTGACGTCCCTTT
SARSUrba     TTAATTTTAATGGACTCACTGGTACTGGTGTGTTAACTCCTTCTTCAAAGAGAT---TTC
SARSTor2     TTAATTTTAATGGACTCACTGGTACTGGTGTGTTAACTCCTTCTTCAAAGAGAT---TTC
BcoV         TTGCTAATTTTATTTTTCATGATGTTAATAGTGGTACTACTTGTTCTACTGATT---TAC
HEV          TTGCTAATTTTATTCTGAATGATGTTAATAGCGGTACAACCTGTTCTACTGATT---TAC
MHV          TTAGTAATTTGATGTTAAATGGCATTAATAGTGGTACTACATGTTCCATGGATT---TGC
RtCoV        TTAGTAATTTGATGTTAAATGGCATTAATAGTGGTACTACATGTTCCACGGATT---TTC
IBV          TTACTAAGAGCGGTGGCTCTCGTATACAAACAGCCACTGAACCGCCAGTTATAA---CTC

CcoV         ATGTAATATATGAAGAAGGAGACAACATAGTGGGTGTACCGTCTGATAATAGTGGTTT-G
FcoV         ATGTAATATATGAAGAAGGAGACAACATAGTGGGTGTACCGTCTGATAATAGCGGTCT-G
TGE          ATGTAATATATGAAGAAGGAGACAACATAGTGGGTGTACCGTCTGATAATAGTGGTGT-G
PRCoV        ATGTAATATATGAAGAAGGAGACAGCATAGTTGGTGTACCGTCTGACAATAGTGGTTT-G
HCoVOC43     ATGTTTCTTGGAGTGATGGTGATGGAATTACTGGCGTCCCACAACCTGTTGAGGGTGTTA
PEDV         ATTTTCAATTCACAAAAGGTGAGTTGATTACTGGCACGCCTAAACCACTTGAAGGTATCA
SARSUrba     AACCATTT--CAACAATTTGGCCG-TGATGTTTCTG----ATTTCACTGATTCCGTTCGA
SARSTor2     AACCATTT--CAACAATTTGGCCG-TGATGTTTCTG----ATTTCACTGATTCCGTTCGA
BcoV         AAAAATCAAACACAGACATAATTCTTGGTGTTTGTGTTAATTATGATCTTTATGGTATTA
HEV          AACAGGGTAATACTATTATTACTACTGATGTTTGTGTTAATTATGACCTATATGGCATTA
MHV          AATTGCCTAATACTGAAGTGGTCACTGGCGTCTGCGTCAAATATGACCTCTACGGTATAA
RtCoV        AATTGCCTAATACGGAAGTGGTCACTGGCGTTTGTGTCAAGTATGACCTCTACGGTAGTA
IBV          AAAACAATTATAATAATATTACTTTAAATACTTGTGTTGATTATAATATATATGGCAGAA
             *                                                       *

CcoV         CACGATTTGTCAGTGTTGCACTTAGACTCTTGT-ACAGATTACAATATATATGGTAGAAC
FcoV         CACGATTTGTCTGTGCTACACCTAGACTCCTGT-ACAGATTACAATATATATGGTAGAAC
TGE          CACGATTTGTCAGTGCTACACCTAGATTCCTGC-ACAGATTACAATATATATGGTAGAAC
PRCoV        CACGATTTGTCAGTGCTACACCTAGATTCGTGC-ACAGATTACAATATATATGGTAGAAC
HCoVOC43     GTTCCTTTATGAATGTTACAT-TGGACAAATGT-ACTAAATATAATATTTATGATGTATC
PEDV         CAGACGTTTCTTTTATGACTC-TGGATGTGTGT-ACCAAGTATACTATCTATGGCTTTAA
SARSUrba     GATCCTAAAACATCT-----G--AAAT----------ATTA-GACATTTCACCTTGCTC
SARSTor2     GATCCTAAAACATCT-----G--AAAT----------ATTA-GACATTTCACCTTGCGC
BcoV         CAGGCCAAGGTATTTTTGTTG--AGGTTAATGCGACTTATTATAATAGTTGGCAGAACCT
HEV          CAGGCCAGGGCATACTTATAG--AAGTTAATGCCACTTATTATAATAGTTGGCAGAACCT
MHV          CAGGCCAAGGTATTTTTAAGG--AGGTTAAGGCTGACTATTATCATAGTTGGCAAAACCT
RtCoV        CAGGCCAAGGTGTTTTTAAGG--AGGTTAAGGCTGATTATTACAATAGTTGGCAGAACCT
IBV          CTGGCCAAGGTTTTATT--------ACTAATGTGACCGACTCAGCTGTTAGTTATAATTA
                                                  * *

CcoV         TGGTGTTGGTATTATTAGAAAAACTAACAGCACACTACTTAGTGGCTTATATT-ACACAT
FcoV         TGGTGTTGGTATTATTAGACGAACTAACAGTACGCTACTTAGTGGCTTATATT-ACACAT
TGE          TGGTGTTGGTATTATTAGACAAACTAACAGGACGCTACTTAGTGGCTTATATT-ACACAT
PRCoV        TGGTGTTGGTATTATTAGACAAACTAACAGGACGATACTTAGTGGCTTATATT-ACACAT
HCoVOC43     TGGTGTGGGTGTTATTCGCGTTAGCAATGACACCTTTCTTAATGGAATTACGT-ACACAT
PEDV         AGGTGAGGGTATTATTACCCTTACAAATTCTAGCATTTTGGCAGGTGTTTATT-ATACAT
SARSUrba     TTTTGGGGGTGTAAGTG-TAATTACACCTGGAACAAATGCTTCATCTGAAGTT-GCTGTT
SARSTor2     TTTTGGGGGTGTAAGTG-TAATTACACCTGGAACAAATGCTTCATCTGAAGTT-GCTGTT
BcoV         TTTATATGATTCTAATGGTAATCTCTATGGTTTTAGAGACTACTTAACAAACA-GAACTT
HEV          TCTTTATGATTCTAGTGGTAATCTCTATGGCTTTAGAGATTATTTATCAAATA-GAACTT
MHV          CTTATATGATGTTAATGGCAACTTAATCGGATTTCGCGATTTTGTTGCTAATA-AGAGTT
RtCoV        CTTATATGATGTTAATGGTAACTTAAATGGTTTCCGTGACATTGTTACCAATA-AGACTT
IBV          TCTAGCAGACGCAGGT-------TTGGCTATTTTAGATACATCTGGTTCCATAGACATCT
                    *       *                                          *

CcoV         CACTATCAGGTGATTTGTTAGGTTTTAAAAATGTTAGTGATGGTGTTGTCTACTCTGTAA
FcoV         CACTATCAGGTGATTTGTTAGGCTTTAAAAATGTTAGTGATGGTGTCATTTATTCTGTGA
TGE          CACTATCAGGTGATTTGTTAGGTTTTAAAAATGTTAGTGATGGTGTCATCTACTCTGTAA
PRCoV        CACTATCTGGTGATTTGTTAGGTTTTACAAATGTTAGTGATGGTGTTATCTACTCTGTAA
HCoVOC43     CAACTTCAGGTAACCTTCTGGGTTTTAAAGATGTTACTAAGGGCACCATCTACTCTATCA
```

```
PEDV        CTGATTCTGGACAGTTGTTAGCCTTTAAGAATGTCACTAGTGGTGCTGTTTATTCTGTCA
SARSUrba    CTATATCAAGATGTTA--ACTGCACTGATGTTTCTACAGCAATTCATGCAGATCAACTCA
SARSTor2    CTATATCAAGATGTTA--ACTGCACTGATGTTTCTACAGCAATTCATGCAGATCAACTCA
BcoV        TTATGATTCGTAGTTGCTATAGCGGTCGTGTTTCAGCGGCCTTTCATGCTAACTCTTCCG
HEV         TTCTTATTCGTAGCTGCTATAGTGGAAGAGTTTCAGCAGTTTTTCATGCTAACTCATCTG
MHV         ATACTATTCGAAGTTGCTATAGTGGGCGGGTCTCGGCTGCATATCATCAAGATGCACCAG
RtCoV       ATTTATTAAGAAGTTGCTATAGTGGGCGCGTTTCGGCTGCATATCATCAAGATGCACCTG
IBV         TTGTTGTACAAGGTGAATATGGTCTTAATTATTATAAGGTTAACCCTTGCGAAGATGTCA
                                                                    *

CcoV        CGCCATGTGATGTAAGTGCACAAGCTGCTGT-TATTGATGGTGCCATAGTTGGAGCTATG
FcoV        CGCCATGTGATGTAAGCGCACAAGCGGCTGT-TATTGATGGTGCCATAGTTGGAGCTATG
TGE         CGCCATGTGATGTAAGCGCACAAGCAGCTGT-TATTGATGGTACCATAGTTGGGGCTATC
PRCoV       CGCCATGTGATGTTAGCGCACAAGCAGCTAT-TATTGATGGTACCATAGTTGGGGCTATC
HCoVOC43    CTCCTTGTAACCCACCAGATCAGCTTGTTGTTTATCAGCAA-GCTGTTGTTGGTGCTATG
PEDV        CGCCATGTTCTTTTTCAGAGCAGGCTGC-ATATGTTAATGATGATATAGTGGGTGTTATT
SARSUrba    CACCA-GCTT-----------------GGCGCATATATTCTACTGGAAACAATG-TATT
SARSTor2    CACCA-GCTT-----------------GGCGCATATATTCTACTGGAAACAATG-TATT
BcoV        AACCA-GCATTGCTATTTCGGAATATTAAATGCAATTACGTTTTTAATAATACTC-TTTC
HEV         AACCA-GCTTTGATGTTTCGTAATCTTAAATGCAGCCACGTTTTTAATAATACCA-TTTT
MHV         AACCA-GCGCTACTATATCGCAATTTAAAATGTGACTATGTCTTTAACAACAACA-TATC
RtCoV       AACCA-GCGCTACTATATCGCAATTTAAAATGTGATTATGTGTTTAATAACAACA-TATC
IBV         -ACCAGCAGTTTGTAGTTTCTGGTGGTAAATTAGTAGGTATTCTTACTTCACGTAATGAG
              **                                                    *

CcoV        ACTTCCATTAATAGTGAACTGTTAGGT--CTAACTCATTGGACAACAACACCTAATTTTT
FcoV        ACTTCCATTAACAGTGAACTGTTAGGT--CTAACACATTGGACAACGACACCTAATTTTT
TGE         ACTTCCATTAACAGTGAACTGTTAGGT--CTAACACATTGGACAACAACACCTAATTTTT
PRCoV       ACTTCCATTAACAGTGAATTGTTAGGT--CTAACACATTGGACAACAACACCTAATTTTT
HCoVOC43    TTGTCTGAAAATT-----TTACTAGTT--ACGGC---TTTTCTAATGTTGTAGAACTGCC
PEDV        TCT------AGTT-----TGTCTAACT--CCACT---TTTAACAATACTAGGGAGTTGCC
SARSUrba    CC--------------------------AGACTCAAGCAGGCTGTCTTAT-AGGAGCTG
SARSTor2    CC--------------------------AGACTCAAGCAGGCTGTCTTAT-AGGAGCTG
BcoV        ACGACAGCTGCAACCTATTAACTATTTTGATAGTTATCTTGGTTGTGTTGT-CAATGCTG
HEV         AAGACAAATACAGCTTGTTAACTATTTTGATAGTTACCTTGGTTGTGTTGT-TAATGCTT
MHV         CCGTGAGGAGACACCACTTAACTATTTCGATAGTTATCTTGGTTGTGTTGT-TAATGCTG
RtCoV       CCGTGAGGAGACACCACTTAACTATTTTGATAGTTATTTGGGTTGTGTTAT-TAATGCTG
IBV         AC---------TGGTTCTCAGCTTCTT-GAGAACCAGTTTTACATCAAAATCACTAATGG

CcoV        ATTACTACTCCATATATAATTATACAAATGTGATGAATCGTGGCACGGCAATTGA---TA
FcoV        ATTACTACTCTATATATAATTACACAAGTGAGAGGACTCGTGGCACTGCAATTGACAGTA
TGE         ATTACTACTCTATATATAATTACACAAATGATAGGACTCGTGGCACTGCAATTGACAGTA
PRCoV       ATTACTACTCTATATATAATTACACAAATGATAAGACTCGTGGCACTCCAATTGGCAGTA
HCoVOC43    GAAATTTTTCTATGCGT-------CCAATGGCAC-----------TTATAATTGC-----
PEDV        TGGTTTCTTCTACCATT-------CTAATGACGG-----------CTCCAATTGT-----
SARSUrba    AGCAT------GTCGACACTT---CTTATGAGTGCGACATTCCTATTGGAGCTGGCATTT
SARSTor2    AGCAT------GTCGACACTT---CTTATGAGTGCGACATTCCTATTGGAGCTGGCATTT
BcoV        ATAATAGTACTTCTAGTGTTG---TTCAAACATGTGATCTCACAGTAGGTAGTGGTTACT
HEV         ATAATAATACAGCTAGTGCTG---TAAGTACTTGTGATTTAACCGTTGGTAGCGGCTATT
MHV         ACAACTCAACTGAAGAAGCTG---TTGACGCGTGTGATTTGCGTATGGGTAGTGGGCTTT
RtCoV       ATAACTCAACTGAGCAGTCTG---TTGACGCGTGTGATTTGCGTATGGGTAGTGGGCTTT
IBV         AACACGTCGTTTTAGACGTTC-TATTACTGAAAATG---------TTGCAAATTGCCCTT

CcoV        ATGATATTGATTGTGAACCTATCATAACATATTCTAATATAGGTGTTTGTAAAAATGGAG
FcoV        ACGATGTTGATTGTGAACCTGTCATAACCTATTCTAATATAGGTGTTTGTAAAAATGGTG
TGE         ATGATGTTGATTGTGAACCTGTCATAACCTATTCTAACATAGGTGTTTGTAAAAATGGTG
PRCoV       ATGACGTTGATTGTGAACCTGTCATAACCTATTCTAACATAGGTGTTTGTAAAAATGGTG
HCoVOC43    ---ACA--------GACGCTGTTTTAACTTATTCTAGTTTTGGCGTTTGTGCAGATGGTT
PEDV        ---ACA--------GAGCCTGTGTTGGTGTATAGTAACATAGGTGTTTGTAAATCTGGCA
SARSUrba    GTGCTAGTTACCATACAGTTTCTTTATTACGTAG---------TACTAGCCAAAAATCTA
SARSTor2    GTGCTAGTTACCATACAGTTTCTTTATTACGTAG---------TACTAGCCAAAAATCTA
```

```
BcoV          GTGTGGATTACTCTACAAAAAGACGAAGTCGTAGAGCGATTACCACTGGTTATCGGTTTA
HEV           GTGTTGATTATGTTACAGCACTTAGATCACGTAGATCTTTTACTACAGGTTATCGCTTTA
MHV           GTGTCAACTATTCAACGTCTCACCGCGCTCGCAGCTCTGTCAGCACGGGTTATAAATTAA
RtCoV         GTGTCAACTATTCAATCGCTCACCGTGCGCGCAGGTCTGTCAGTACGGGTTATAAATTAA
IBV           ATGTTAGTTATGGTAAGTTTTGTATAAAACCTGATGGCT----CAATTGCCACAATAGTA
                                                                      *

CcoV          CTTTGGTTTTTATT--AACGTCACACATTCTG-ATGGAGACGTTCAACCAATTAGCACCG
FcoV          CTTTGGTTTTTATT--AACGTCACACATTCTG-ACGGAGACGTGCAACCAATTAGCACTG
TGE           CTTTTGTTTTTATT--AACGTCACACATTCTG-ATGGAGACGTGCAACCAATTAGCACTG
PRCoV         CTTTGGTTTTTATT--AACGTCACACATTCTG-ATGGAGACGTGCAACCAATTAGCACTG
HCoVOC43      CTATAATTGCTGTTCAACCACGTAATGTTTCATATGATAGTGTTTCAGCTATCGTCACAG
PEDV          GTATTGGCTATGTTCCATCTCAGTATGGCC---AAGTCAAGATTGCACCCACGGTTACTG
SARSUrba      ------TTGTGGCTTATACTATGTCTTTAGGTGCTGATAGTTC--AATTGCTTACTCTAA
SARSTor2      ------TTGTGGCTTATACTATGTCTTTAGGTGCTGATAGTTC--AATTGCTTACTCTAA
BcoV          CTAATTTTGAGCCATTTACTGTTAATTCAGTAAATGATAGTTTAGAACCTGTAGGTGGTT
HEV           CTAATTTTGAACCATTTGCCGCTAATTTGGTAAATGATAGTATAGAACCTGTTGGTGGTT
MHV           CTACTTTTGAACCATTTACAGTCCGCATTGTCAATGATAGTGTTGAGTCTGTTGATGGGT
RtCoV         CTACTTTTGAACCATTTACAGTCAGCATTGTCAATGATAGTGTTGAGTCTGTTGGTGGAT
IBV           ------------CCAAAACAATTGGAACAGTTTGTGGCACCTTT--ATTTAATGTTACTG
                           *                   *

CcoV          ---GTAATGTCACGATACCCACAAATTTTACTATATCTGTGCAAGTCGAATATATTCAGG
FcoV          ---GTAATGTCACGATACCTACAAATTTTACTATATCTGTGCAAGTTGAATACATGCAGG
TGE           ---GTAATGTCACGATACCTACAAACTTTACCATATCCGTGCAAGTCGAATATATTCAGG
PRCoV         ---GTAACGTCACGATACCTACTAACTTTACTATATCCGTGCAAGTCGAATACATTCAGG
HCoVOC43      ---CTAATTTGTCTATACCTTCCAATTGGACCACTTCGGTCCAGGTTGAGTATTTACAAA
PEDV          ---GGAATATTAGTATTCCCACCAACTTTAGTATGAGTATTAGAACAGAATATTTACAGC
SARSUrba      TA-ACACCATTGCTATACCTACTAACTTTTCAATTAGCATTACTACAGAAGTAATGCCTG
SARSTor2      TA-ACACCATTGCTATACCTACTAACTTTTCAATTAGCATTACTACAGAAGTAATGCCTG
BcoV          TGTATGAAATTCAAATACCTTCAGAGTTTACTATAGGTAATATGGAGGAGTTTATTCAAA
HEV           TGTATGAAATACAGATACCTTCAGAGTTTACCATTGGTAATTTAGAAGAGTTCATTCAAA
MHV           TATATGAGCTGCAAATACCAACCAACTTTACTATAGCTAGCCATCAGGAGTTCGTTCAAA
RtCoV         TATATGAGATGCAAATACCTACTAATTTTACTATAGCTAGCCATCAGGAGTTCATTCAAA
IBV           ---AAAATGTGCTCATACCTAACAGTTTCAACTTAACTGTTACAGATGAGTACATACAAA
                    *    ** **       *                    **     * *

CcoV          TTTACACTACACCAGTTTCAATAGACTGTGCAAGATACGTTTGCAATGGTAACCCAAGAT
FcoV          TTTACACTACACCAGTATCAATAGATTGTGCAAGATACGTTTGTAATGGTAACCCTAGAT
TGE           TTTACACTACACCAGTGTCAATAGACTGTTCAAGATATGTTTGTAATGGTAACCCTAGGT
PRCoV         TTTACACTACACCAGTGTCAATAGACTGTTCAAGATATGTTTGTAATGGTAACCCTAGGT
HCoVOC43      TTACAAGTACACCTATCGTAGTTGATTGCTCCACTTATGTTTGCAATGGTAATGTGCGCT
PEDV          TTTACAACACGCCTGTTAGTGTTGATTGTGCTACATATGTTTGTAATGGTAACTCTCGTT
SARSUrba      TTTCTATGGCTAAAACCTCCGTAGATTGTAATATGTACATCTGCGGAGATTCTACTGAAT
SARSTor2      TTTCTATGGCTAAAACCTCCGTAGATTGTAATATGTACATCTGCGGAGATTCTACTGAAT
BcoV          CAAGCTCTCCTAAAGTTACTATTGATTGTTCTGCTTTTGTCTGTGGTGATTATGCAGCAT
HEV           CGAGATCCCCTAAGGTTACTATAGACTGTGCTACATTTGTTTGTGGTGACTATGCAGCAT
MHV           CGAGGTCTCCAAAGGTTACTATAGACTGTGCTGCATTTGTCTGTGGTGGCCACACAGCAT
RtCoV         CGAGGTCTCCGAAGGTTACTATAGATTGTGCTGCATTTGTCTGTGGTGATTATACAGCGT
IBV           CGCGTATGGATAAGGTCCAAATTAATTGCCTGCAGTATGTTTGTGGCAGTTCTCTGGATT
                           *  * **       *   * **                 *

CcoV          GCAATAAGTTATTAACACAATACGTTTCTGCATGTCAAACTATTGAGCAAGCGCTTGCAA
FcoV          GTAACAAATTGTTAACACAATATGTGTCTGCATGTCAAACTATTGAACAAGCACTTGCAA
TGE           GTAACAAATTGTTAACACAATACGTTTCTGCATGTCAAACTATTGAGCAAGCACTTGCAA
PRCoV         GTAACAAACTGTTAACACAATACGTTTCTGCATGTCAAACTATTGAGCAAGCACTTGCAA
HCoVOC43      GTGTTGAATTGCTTAAGCAGTATACTTCTGCTTGTAAAACTATTGAAGACGCCTTAAGAA
PEDV          GTAAACAATTACTCACCCAGTACACTGCAGCATGTAAGACCATAGAGTCAGCATTACAAC
SARSUrba      GTGCTAATTTGCTTCTCCAATATGGTAGCTTTTGCACACAACTAAATCGTGCACTCTCAG
SARSTor2      GTGCTAATTTGCTTCTCCAATATGGTAGCTTTTGCACACAACTAAATCGTGCACTCTCAG
BcoV          GTAAATCACAGTTGGTTGAATATGGTAGCTTCTGTGACAATATTAATGCTATACTCACAG
HEV           GTAGACAACAGTTAGCTGAGTATGGTAGTTTTTGTGAGAACATTAATGCTATACTCACAG
MHV           GCCGTCAGCAGTTGGTTGAGTACGGCTCATTCTGTGATAATATTAATGCCATTCTTGGCG
```

```
RtCoV          GTAGACAACAGTTGGTTGATTATGGCTCTTTTTGTGATAATATTAATGCCATTCTTGGCG
IBV            GTAGAAAGTTGTTTCAACAATATGGGCCTGTTTGCGACAACATATTGTCTGTAGTAAATA
               *           *     * **          **        *          *

CcoV           TGGGTGCCAG---ACTTGAAAACATGGAGATTGATTCCATGTTATTTGTTTCGGAAAATG
FcoV           TGGGTGCCAG---ACTTGAAAACATGGAGGTTGATTCCATGTTGTTTGTCTCGGAAAATG
TGE            TGGGTGCCAG---ACTTGAAAACATGGAGGTTGATTCCATGTTGTTTGTTTCTGAAAATG
PRCoV          TGGGTGCCAG---ACTTGAAAACATGGAAGTTGATTCCATGTTATTTGTTTCTGAAAATG
HCoVOC43       ATAGCGCCAG---GCTGGAGTCTGCAGATGTTAGTGAGATGCTCACTTTTGACAAGAAAG
PEDV           TCAGCGCTAG---GCTTGAGTCTGTTGAAGTTAACTCTATGCTTACCATTTCTGAAGAGG
SARSUrba       GTATTG-----CTGCTGAACAGGATCGCAACACACGTGAAGTGTTCGCTCAAGTCAAACA
SARSTor2       GTATTG-----CTGCTGAACAGGATCGCAACACACGTGAAGTGTTCGCTCAAGTCAAACA
BcoV           AAGTAAATGAACTACTTGACACTACACAGTTGCAAGT-AGCTAATAGTTTAATGAATGGT
HEV            AAGTAAATGAACTACTTGACACTACACAGTTGCAAGT-AGCTAATAGTTTAATGAATGGA
MHV            AGGTAAATAACCTCATAGATACTATGCAACTTCAAGT-TGCAAGTGCTTTAATCCAAGGT
RtCoV          AGGTGAATAACCTCATAGATACTATGCAATTACAAGT-TGCTAGTGCTCTGATCCAAGGT
IBV            GTGTTGGTCA---AAAAGAAGATATGGAACTTTTGAATTTCTATTCTTCTACTAAACCGG
                          *

CcoV           CCCTTA-AATTGGCATCTGTTGAAGCATTCAATAGTACGGAAAATTTAGACCCTATTTAT
FcoV           CCCTTA-AATTGGCATCTGTTGAGGCGTTCAATAGTACAGAAAATTTAGATCCTATTTAC
TGE            CCCTTA-AATTGGCATCTGTTGAAGCATTCAATAGTTCAGAAACTTTAGACCCTATTTAC
PRCoV          CCCTTA-AATTGGCTTCTGTCGAAGCATTCAATAGTTCAGAAACTTTAGATCCTATTTAC
HCoVOC43       CGTTTACACTT-GCTAATGTTAGT------AGTTTT---GGTGACTACAACCTTAGC---
PEDV           C-TTTACAGTTAGCTACCATCAGTTCGTTTAATGGT---GATGGATATAACTTTACT---
SARSUrba       AATGTACAAAACCCCAACTTTGAAA-------------TATTTTGGTGGTTTTAATT--
SARSTor2       AATGTACAAAACCCCAACTTTGAAA-------------TATTTTGGTGGTTTTAATT--
BcoV           GTCACTCTTAGCACTAAGCTTAAAGATGGCGTTAATTTCAATGTAGACGACATCAATT--
HEV            GTCACCCTTAGTACCAAGATTAAGGATGGCATTAATTTCAATGTTGACGATATCAACT--
MHV            GTCACGTTAAGCTCACGCTTATCGGATGGCATTGGTGGTCAAATAGATGATATTAATT--
RtCoV          GTCACGCTAAGTTCACGCTTGGCAGATGGCATCTCAGGTCAGATTGATGATATTAATT--
IBV            CTGGTTTTAATACACCAGTTCTTAG-------TAATGTTAGCACTGGTGAGTTTAATA--
                              *                                        *

CcoV           AAAGAATGGCCTAACATTGGTGGTTCTTGGCTAGGAGGTTTAAAAGATATATTGCCATCT
FcoV           AAAGAATGGCCTAGCATAGGTGGTTCTTGGCTAGGAGGTCTAAAAGATATACTACCGTCC
TGE            AAAGAATGGCCTAATATAGGTGGTTCTTGGCTAGAAGGTCTAAAATACATACTTCCGTCC
PRCoV          AAAGAATGGCCTAATATAGGTGGCTTTTGGCTAGAAGGTCTAAAATACATACTTCCGTCC
HCoVOC43       ----AGCGTCATA-----CCTAGCTTG-------------------------CCCACA
PEDV           ----AATGTGCTG-----GGTGCTTCCGTGTACGA-----------------TCCTGCA
SARSUrba       ----TTTCACAAATATTACCTGACCCT-------------------CTAAAGCCAA---
SARSTor2       ----TTTCACAAATATTACCTGACCCT-------------------CTAAAGCCAA---
BcoV           ----TTTCCCCTGTATTAGGTTGTTTAGGAAGCGC-----------TTGTAATAAAGTT
HEV            ----TCTCCCCTGTATTAGGTTGTTTAGGAAGCGA-----------ATGTAATAGAGCT
MHV            ----TTAGTCCTCTGCTTGGTTGTTTAGGTTCTGA-----------CTGTGGCGAAGTT
RtCoV          ----TTAGTCCTCTTCTAGGTTGCCTTGGCTCAGA-----------TTGTAGCGAAGGC
IBV            ----TTTCTCTTCTGTTAACAAATCCT-------------------------AGTAGT

CcoV           CATAATAGCAAACGTAAGTACCGCTCGGCTATAGAAGACTTGCTTTTTGATAAGGTTGTA
FcoV           CATAATAGCAAACGTAAGTATGGTTCTGCTATAGAAGATTTGCTTTTTGATAAAGTTGTA
TGE            CATAATAGCAAACGTAAGTATCGTTCAGCTATAGAGGACTTGCTTTTTGATAAGGTTGTA
PRCoV          GATAATAGCAAACGTAAGTATCGTTCAGCTATAGAGGACTTGCTTTTTTCTAAGGTTGTA
HCoVOC43       AGTGGTAGTAGAGTGGCTGGTCGCAGTGCCATAGAAGACATACTTTTTAGCAAACTTGTT
PEDV           AGTGGCAGGGTGGTACAAAAAAGGTCTGTTATTGAAGACTTGCTTTTTAATAAAGTGGTT
SARSUrba       -CTAAGA--------------GGTCTTTTATTGAGGACTTGCTCTTTAATAAGGTGACA
SARSTor2       -CTAAGA--------------GGTCTTTTATTGAGGACTTGCTCTTTAATAAGGTGACA
BcoV           TCCAGCA----------G----ATCTGCTATAGAGGATTTACTTTTTTCTAAAGTAAAG
HEV            TCCACTA----------G----ATCTGCTATAGAGGATTTACTTTTTGATAAAGTAAAA
MHV            ACCATGGCAGCTCAAACCGGACGATCTGCTATAGAGGATGTATTATTTGACAAAGTCAAA
RtCoV          ACCAAGGCAGCGCAA---GGGCGATCTGCTATAGAGGATGTATTATTTGATAAGGTCAAA
IBV            CGTAGAA----------AGCGTTCTCTTATTGAAGACCTTCTATTTACAAGCGTTGAA
                                          ** ** **  *  * ***   *   *
```

```
CcoV          ACATCTGGCTTAGGTACAGTTGACGAAGATTACAAACGTTCTGCAGGTGGTTATGACA--
FcoV          ACATCTGGTTTAGGTACAGTTGATGAAGATTATAAACGTTGTACTGGTGGTTACGACA--
TGE           ACATCTGGTTTAGGTACAGTTGATGAAGATTATAAACGTTGTACAGGTGGTTATGACA--
PRCoV         ACATCTGGTTTAGGTACAGTTGATGAAGATTACAAACGTTGTACAGGTGGTTATGACA--
HCoVOC43      ACTTCTGGACTTGGCACTGTGGACGCAGACTACAAAAAGTGCACTAAGGGTCTTTCCA--
PEDV          ACTAATGGCCTTGGTACTGTTGATGAAGACTATAAGCGCTGTTCTAATGGTCGCTCTG--
SARSUrba      CTCGCTGATGCTGGCTTCATGAA---GCAATATGGCGAATGCCTAGGTGATATTAATG--
SARSTor2      CTCGCTGATGCTGGCTTCATGAA---GCAATATGGCGAATGCCTAGGTGATATTAATG--
BcoV          TTATCTGATGTCGGTTTCGTTGA---GGCTTATAATAATTGTACTGGAGGTGCCGAAA--
HEV           TTGTCTGATGTCGGCTTTGTACA---GGCCTATAATAACTGCACTGGAGGTGCCGAAA--
MHV           CTCTCTGATGTTGGCTTTGTCGA---AGCATATAACAATTGCACTGGAGGCCAAGAAG--
RtCoV         CTCTCTGATGTTGGCTTTGTCGA---ATCATATAATAATTGCACTGGAGGTCAAGAAG--
IBV           TCTGTTGGACTACCAACAAATGA---CGCATATAAAAATTGCACTGCAGGACCTTTAGGC
                   **                *       **       *        *

CcoV          ----TAGCTGACTTAGTGTGTGCACGATATTACAATGGCATCATGGTGCTACCTGGTGTA
FcoV          ----TAGCAGACTTGGTGTGTGCTCAATATTACAATGGCATCATGGTTCTACCAGGTGTA
TGE           ----TAGCTGACTTAGTATGTGCTCAATACTATAATGGCATCATGGTGCTACCTGGTGTG
PRCoV         ----TAGCTGACTTAGTATGTGCTCAATACTATAATGGCATTATGGTGCTACCTGGTGTG
HCoVOC43      ----TTGCTGACTTGGCTTGTGCTCAATATTATAATGGCATTATGGTTTTGCCTGGCGTC
PEDV          ----TGGCTGATCTAGTCTGTGCGCAGTATTACTCTGGTGTCATGGTACTACCTGGCGTT
SARSUrba      ----CTAGAGATCTCATTTGTGCGCAGAAGTTCAATGGACTTACAGTGTTGCCACCTCTG
SARSTor2      ----CTAGAGATCTCATTTGTGCGCAGAAGTTCAATGGACTTACAGTGTTGCCACCTCTG
BcoV          ----TTAGGGACCTCATTTGTGTGCAAAGTTATAATGGTATCAAAGTGTTGCCTCCACTG
HEV           ----TTAGGGATCTCATTTGTGTGCAAAGTTATAATGGTATCAAAGTGTTGCCTCCATTG
MHV           ----TTAGAGACCTACTTTGTGTGCAATCTTTTAATGGCATCAAAGTGCTACCGCCTGTG
RtCoV         ----TTAGAGACCTACTTTGTGTGCAATCTTTTAATGGCATTAAAGTGCTACCGCCTGTA
IBV           TTTTTTAAGGACCTTGCGTGTGCTCGTGAATATAATGGTTTGCTTGTGTTGCCTCCTATC
                    **  *    ****  *     *    ***  *    **  * **     *

CcoV          GCTAATGATGACAAGATGACTATGTACACTGCATCTCTTACAGGTGGTATAACATTAGGT
FcoV          GCTAATGCTGACAAGATGACTATGTACACAGCATCACTTGCAGGTGGTATAACATTAGGT
TGE           GCTAATGCTGACAAAATGACTATGTACACAGCATCCCTTGCAGGTGGTATAACATTAGGT
PRCoV         GCTAATGCTGACAAAATGACTATGTACACAGCATCCCTCGCAGGTGGTATAACATTAGGT
HCoVOC43      GCTGATGCTGAACGAATGGCCATGTATACAGGTTCTTTAATTGGTGGAATTGCTTTAGGA
PEDV          GTTGACGCTGAGAAGCTTCACATGTACAGTGCGTCTCTCATAGGTGGTATGGCGCTAGGA
SARSUrba      CTCACTGATGATATGATTGCTGCCTACACTGCTGCTCTAGTTAGTGGTACTGCCACTGCT
SARSTor2      CTCACTGATGATATGATTGCTGCCTACACTGCTGCTCTAGTTAGTGGTACTGCCACTGCT
BcoV          CTCTCAGTAAATCAGATCAGTGGATACACTTTGGCTGCCACCTCTGCTAGTCTGTTTCCT
HEV           TTATCTGAAAATCAGATCAGTGGCTACACTTTGGCAGCCACCGCTGCTAGCTTATTCCCT
MHV           TTGTCTGAGAATCAAATTTCTGGTTATACAGCGGGAGCTACTGTATCTGCTATGTTCCC-
RtCoV         TTATCCGAGAGTCAAATCTCTGGTTATACAGCGGGAGCTACTGCATCTGCTATGTTCCCT
IBV           ATAACAGCAGAAATGCAAGCTTTGTATACTAGTTCTCTAGTAGCTTCTATGGCTTTTGGT
                    *                ** *

CcoV          GCACTTAGTGGTGGCGCAGTGGC---------TATACCTTTTGCAGTAGCAGTTCAGGCT
FcoV          GCACTTGGTGGTGGCGCCGTGGC---------TATACCTTTTGCAGTAGCAGTACAGGCT
TGE           GCACTTGGTGGAGGCGCCGTGGC---------TATACCTTTTGCAGTAGCAGTTCAGGCT
PRCoV         GCACTTGGTGGAGGCGCCGTGGC---------TATACCTTTTGCAGTAGCAGTTCAGGCT
HCoVOC43      GGTCTAACATCAG---CCGTTTC---------AATACCATTTTCATTAGCAATTCAGGCA
PEDV          GGTATAACTGCTG---CAGCGGC---------ATTGCCTTTTAGCTATGCTGTTCAAGCG
SARSUrba      GGATGGACATTTGGTGCTGGCGCTGCTCTTCAAATACCTTTTGCTATGCAAATGGCATAT
SARSTor2      GGATGGACATTTGGTGCTGGCGCTGCTCTTCAAATACCTTTTGCTATGCAAATGGCATAT
BcoV          CCTTTGTCAGCAGCAGTAGGTG----------TACCATTTTATTTAAATGTTCAGTAT
HEV           CCTTGGACAGCTGCAGCAGGTG----------TACCATTTTATTTAAATGTTCAGTAT
MHV           --ATGGTCTGCAGCTGCAGGTG----------TGCCATTTTCTTTAAGTGTTCAATAT
RtCoV         CCATGGTCTGCAGCTGCAGGTG----------TGCCATTTGCTTTAAGTGTTCAATAT
IBV           GGTATTACTGCAG---CTGGTGC---------TATACCTTTTGCCACACAACTGCAGGCT
                        *     *               * ** ***          *

CcoV          AGACTTAATTATGTTGCTCTACAAACTGATGTATTGAACAAAAACCAACAAATCTTGGCT
FcoV          AGACTTAATTATGTTGCTCTACAAACTGATGTATTGAATAAAAACCAACAGATCCTGGCT
```

```
TGE          AGACTTAATTATGTTGCTCTACAAACTGATGTATTGAACAAAAACCAGCAGATTCTGGCT
PRCoV        AGACTTAATTATGTTGCTCTACAAACTGATGTATTGAACAAAAACCAGCAGATCCTGGCT
HCoVOC43     CGTTTAAATTATGTTGCATTGCAGACTGATGTTTTACAAGAAAATCAGAAAATTCTTGCT
PEDV         AGACTCAATTATCTTGCTTTACAGACGGATGTTCTACAGCGGAACCAGCAATTGCTTGCT
SARSUrba     AGGTTCAATGGCATTGGAGTTACCCAAAATGTTCTCTATGAGAACCAAAAACAAATCGCC
SARSTor2     AGGTTCAATGGCATTGGAGTTACCCAAAATGTTCTCTATGAGAACCAAAAACAAATCGCC
BcoV         CGTATTAATGGGATTGGTGTTACCATGGATGTGTTAAGTCAAAATCAAAAGCTTATTGCT
HEV          CGTATAAATGGGCTTGGCGTCACTATGGATGTGCTAAGTCAAAACCAAAAGCTTATTGCT
MHV          AGAATTAATGGTCTTGGTGTCACTATGAATGTTCTTAGTGAAAATCAGAAAATGATAGCA
RtCoV        AGAATTAATGGTCTTGGTGTCACTATGAATGTTCTTAGTGAAAACCAGAAAATGATAGCT
IBV          AGAATTAATCACTTGGGTATTACCCAGTCACTTTTGTTGAAGAATCAAGAAAAAATTGCT
                 *  * ***    * *   *           *  *       ** **  *     * **

CcoV         AATGCTTTCAATCAAGCTATTGGTAACATTACACAGGCATTTGGTAAGGTTAATGACGCT
FcoV         AATGCTTTCAATCAAGCTATTGGTAACATTACACAGGCTTTTGGTAAGGTTAATGATGCT
TGE          AGTGCTTTCAATCAAGCTATTGGTAACATTACACAGTCATTTGGTAAGGTTAATGATGCT
PRCoV        AGTGCTTTTAATCAAGCTATTGGTAACATTACACAGTCATTTGGTAAGGTTAATGATGCT
HCoVOC43     GCATCTTTTAACAAAGCAATGACCAACATAGTAGATGCCTTTACTGGTGTTAATGATGCT
PEDV         GAGTCTTTTAACTCTGCTATTGGTAATATAACTTCAGCCTTTGAGAGTGTTAAAGAGGCT
SARSUrba     AACCAATTTAACAAGGCGATTAGTCAAAT----------------------------
SARSTor2     AACCAATTTAACAAGGCGATTAGTCAAAT----------------------------
BcoV         AATGCATTTAACAATGCTCTTGATGCTAT----------------------------
HEV          AGTGCATTTAACAATGCTCTTGATGCTAT----------------------------
MHV          AGTGCATTTAACAACGCGATAGGTGCTAT----------------------------
RtCoV        AGTTCATTTAACAACGCGATAGGTGCTAT----------------------------
IBV          GCTTCCTTTAATAAGGCCATTGGTCATAT----------------------------
                   ** **    **  *       **

CcoV         ATACATCAAACATCAAAAGGTCTTGCTACTGTTGCTAAAGCATTGGCAAAGGTGCAAGAT
FcoV         ATACATCAAACATCACAAGGTCTTGCCACTGTTGCTAAAGCGTTGGCAAAAGTGCAAGAT
TGE          ATACATCAAACATCACGAGGTCTTGCTACTGTTGCTAAAGCATTGGCAAAAGTGCAAGAT
PRCoV        ATACATCAAACATCACGAGGTCTTACAACTGTTGCTAAAGCATTGGCAAAAGTGCAAGAT
HCoVOC43     ATTACACAAACTTCACAAGCCCTACAAACAGTTGCTACTGCACTTAACAAGATCCAGGAT
PEDV         ATTAGTCAAACTTCCAAGGGTTTGAACACTGTGGCTCATGCGCTTACTAAGGTTCAAGAG
SARSUrba     -----TCAAG------AATCACTTACAACAACATCAACTGCATTGGGCAAGCTGCAAGAC
SARSTor2     -----TCAAG------AATCACTTACAACAACATCAACTGCATTGGGCAAGCTGCAAGAC
BcoV         -----TCAGG------AAGGGTTTGATGCTACCAATTCTGCTTTAGTTAAAATTCAAGCT
HEV          -----CCAGG------AAGGGTTCGACGCAACCAATTCTGCTTTAGTTAAAATTCAGGCT
MHV          -----ACAGG------AAGGGTTTGCTGCAACCAATTCTGCCTTAGCAAAAATGCAGTTC
RtCoV        -----ACAGG------AAGGGTTCGATGCAACCAATTCTGCTTTAGCGAAAATTCAGTCC
IBV          -----GCAGG------AAGGTTTTAGAAGTACATCTCTAGCATTACAACAAATTCAAGAT
                   **              *                  **  *     *  * **

CcoV         GTTGTTAACACGCAAGGTCAAGCTTTAAGCCACCTAACAGTACAATTGCAAAACAATTTT
FcoV         GTTGTCAACACACAAGGGCAAGCTTTAAGTCACCTTACAGTACAATTGCAAAATAATTTT
TGE          GTTGTCAACATACAAGGGCAAGCTTTAAGCCACCTAACAGTACAATTGCAAAATAATTTC
PRCoV        GTTGTCAACACACAAGGTCAAGCTTTAAGACACCTAACAGTACAATTGCAAAATAATTTC
HCoVOC43     GTTGTTAATCAACAAGGCAACTCATTGAACCATTTAACTTCTCAGTTGAGGCAGAATTTT
PEDV         GTTGTTAATTCGCAGGGTTCAGCTTTGAACCAACTTACCGTACAGCTGCAACACAACTTC
SARSUrba     GTTGTTAACCAGAATGCTCAAGCATTAAACACACTTGTTAAACAACTTAGCTCTAATTTT
SARSTor2     GTTGTTAACCAGAATGCTCAAGCATTAAACACACTTGTTAAACAACTTAGCTCTAATTTT
BcoV         GTTGTTAATGCAAATGCTGAAGCTCTTAATAACTTATTGCAACAACTCTCTAATAGATTT
HEV          GTTGTTAATGCAAATGCTGAAGCACTTAATAACTTATTGCAGCAACTCTCTAACAGATTT
MHV          GTTGTCAATGCAAATGCGGAAGCACTCAATAATTTATTAAACCAGCTTTCCAATAGGTTT
RtCoV        GTTGTCAACGCAAATGCAGAAGCACTCAATAACCTTTTGAATCAGCTTTCCAATAGGTTT
IBV          GTTGTTAGTAAACAGAGTGCTATTCTTACTGAGACTATGGCATCACTTAATAAAAATTTT
             ***** *      *          * *                   *       *  **

CcoV         CAAGCCATTAGCAGTTCTATTAGTGACATTTATAACAGGCTTGATGAATTGAGTGCTGAT
FcoV         CAAGCCATTAGTAGTTCTATTAGTGATATTTATAACAGGCTTGACGAACTGAGTGCTGAT
TGE          CAAGCCATTAGTAGTTCTATTAGTGACATTTATAATAGGCTTGACGAATTGAGTGCTGAT
PRCoV        CAAGCCATTAGTAGTTCTATTAGTGACATTTATAATAGGCTTGATGAATTGAGTGCTGAT
HCoVOC43     CAAGCTATCTCTAGCTCTATTCAGGCTATCTATGACAGACTTGACACTATTCAGGCTGAT
```

```
PEDV          CAAGCCATTTCTAGTTCTATTGATGACATTTATTCCCGACTGGACATTCTTTCAGCCGAT
SARSUrba      GGTGCAATTTCAAGTGTGCTAAATGATATCCTTTCGCGACTTGATAAAGTCGAGGCGGAG
SARSTor2      GGTGCAATTTCAAGTGTGCTAAATGATATCCTTTCGCGACTTGATAAAGTCGAGGCGGAG
BcoV          GGTGCTATAAGTTCTTCTTTACAAGAAATTCTATCTAGACTGGATGCTCTTGAAGCGCAA
HEV           GGTGCCATAAGTGCCTCTTTACAAGAAATTTTATCCAGGCTCGATGCTCTTGAAGCTAAA
MHV           GGTGCAATTAGTGCTTCTTTACAAGAAATTCTATCTCGCCTAGATGCTCTTGAAGCGCAG
RtCoV         GGTGCAATTAGTGCTTCTTTACAGGAAATTCTATCTCGCCTCGATGCTCTTGAAGCTCAG
IBV           GGTGCTATTTCTTCTGTGATTCAAGAAATCTACCAGCAATTTGACGCCATACAAGCAAAT
                 ** **           *    *  **          * **     *    **  *

CcoV          GCACAAGTTGACAGGCTGATTACAGGACGACTTACAGCACTTAATGCATTTGTGTCTCAG
FcoV          GCACAAGTTGATAGGCTGATTACAGGTAGACTTACAGCACTTAATGCATTTGTGTCTCAG
TGE           GCACAAGTTGACAGGCTGATCACAGGAAGACTTACAGCACTTAATGCATTTGTGTCTCAG
PRCoV         GCACAAGTCGACAGGCTGATCACAGGAAGACTTACAGCACTTAATGCATTTGTGTCTCAG
HCoVOC43      CAACAAGTAGATAGGCTGATTACTGGTAGATTGGCTGCTTTGAATGTATTCGTTTCTCAT
PEDV          GTTCAGGTTGATCGTCTCATCACCGGCAGATTATCAGCACTTAATGCTTTTGTTGCCCAA
SARSUrba      GTACAAATTGACAGGTTAATTACAGGCAGACTTCAAAGCCTTCAAACCTATGTAACACAA
SARSTor2      GTACAAATTGACAGGTTAATTACAGGCAGACTTCAAAGCCTTCAAACCTATGTAACACAA
BcoV          GCTCAGATAGACAGACTTATTAATGGGCGTCTTACCGCTCTTAATGTTTATGTTTCTCAA
HEV           GCTCAGATAGACAGACTTATCAATGGGCGTCTCACCGCTCTTAATGCTTATGTTTCTCAG
MHV           GCTCAGATAGACCGTCTTATTAATGGCAGATTAACTGCACTTAATGCATATGTCTCTAAG
RtCoV         GCTCAGATAGACCGTCTTATTAATGGCAGATTAACTGCACTTAATGCATATGTCTCTAAG
IBV           GCTCAAGTGGATCGTCTTATAACTGGTAGATTGTCATCACTTTCTGTTTTAGCATCTGCT
                 **  * **  *  * ** *  **  *  *        *       * *  *

CcoV          ACTTTAACCAGACAAGCAGAGGTTAGGGCTAGTAGACAACTTGCTAAAGACAAGGTTAAT
FcoV          ACTCTAACCAGACAAGCAGAGGTTAGGGCTAGTAGACAACTTGCCAAAGACAAGGTTAAT
TGE           ACTCTAACCAGACAAGCGGAGGTTAGGGCTAGTAGACAACTTGCCAAAGACAAGGTTAAT
PRCoV         ACTCTAACCAGACAAGCCGAGGTTAGGGCTAGTAGACAACTTGCTAAAGACAAGGTTAAT
HCoVOC43      ACATTGACTAAGTACACTGAAGTTCGTGCTTCCAGACAGCTTGCACAACAAAAAGTGAAT
PEDV          ACCCTCACTAAGTATACTGAGGTTCAGGCTAGCAGGAAGCTAGCACAGCAAAAGGTTAAT
SARSUrba      CAACTAATCAGGGCTGCTGAAATCAGGGCTTCTGCTAATCTTGCTGCTACTAAAATGTCT
SARSTor2      CAACTAATCAGGGCTGCTGAAATCAGGGCTTCTGCTAATCTTGCTGCTACTAAAATGTCT
BcoV          CAGCTTAGTGATTCTACACTAGTAAAATTTAGTGCAGCACAAGCTATGGAGAAGGTTAAT
HEV           CAGCTTAGTGATTCTACACTAGTAAAATTTAGTGCAGCACAAGCTATTGAGAAAGTTAAT
MHV           CAGCTGAGTGACATGACCCTTGTTAAGGTGAGTGCAGCCCAGGCTATAGAGAAAGTTAAT
RtCoV         CAGCTGAGCGACATGACCCTTATTAAGGTGAGTGCTGCCCAGGCTATAGAGAAAGTTAAT
IBV           AAGCAGGCGGAGTATATTAGAGTGTCACAACAGCGTGAGTTAGCTACTCAGAAAATTAAT
                                 *                    **       **  *   *

CcoV          GAATGCGTTAGGTCTCAATCCCAGAGATTTGGATTCTGTGGTA---ATGGTACACATTTG
FcoV          GAATGTGTTAGGTCTCAGTCTCAGAGATTCGGATTCTGTGGTA---ATGGTACACATTTG
TGE           GAATGCGTTAGGTCTCAGTCTCAGAGATTCGGATTCTGTGGTA---ATGGTACACATTTG
PRCoV         GAATGCGTTAGGTCTCAGTCTCAGAGATTCGGCTTCTGTGGTA---ATGGTACACATTTG
HCoVOC43      GAGTGTGTCAAATCCCAGTCTAAGCGTTATGGCTTCTGTGGAA---ATGGCACTCACATT
PEDV          GAGTGCGTCAAATCGCAATCTCAGCGTTACGGTTTTTGTGGTGGTGATGGCGAGCACATT
SARSUrba      GAGTGTGTTCTTGGACAATCAAAAAGAGTTGACTTTTGTGGAA---AGGGCTACCACCTT
SARSTor2      GAGTGTGTTCTTGGACAATCAAAAAGAGTTGACTTTTGTGGAA---AGGGCTACCACCTT
BcoV          GAATGTGTCAAAAGCCAATCATCTAGGATAAATTTTTGTGGTA---ATGGTAATCATATT
HEV           GAATGTGTTAAAAGCCAATCATCTAGGATAAATTTCTGTGGTA---ATGGTAATCATATT
MHV           GAGTGTGTTAAAAGCCAATCATCTAGGATAAATTTCTGTGGCA---ATGGCAATCATATA
RtCoV         GAGTGTGTTAAAAGCCAATCACCTAGGATAAATTTCTGTGGCA---ATGGCAATCATATA
IBV           GAGTGTGTTAAGTCACAGTCTATTAGGTACTCCTTTTGTGGTA---ATGGACGACATGTT
              ** ** **        ** **     *        ** *****     * **    **  *

CcoV          TTTTCACTTGCAAATGCGGCACCAAATGGCATGATTTTCTTTCACACAGTGCTATTACCA
FcoV          TTTTCACTAGCAAATGCAGCACCAAATGGCATGATTTTCTTTCATACAGTACTATTACCA
TGE           TTTTCACTCGCAAATGCAGCACCAAATGGCATGATTTTCTTTCACACAGTGCTATTACCA
PRCoV         TTTTCACTCGCAAATGCAGCACCAAATGGCATGATTTTCTTTCACACAGTGCTATTACCA
HCoVOC43      TTCTCAATTGTTAATGCTGCTCCTGAGGGGCTTGTTTTTCTCCACACTGTCTTGTTGCCG
PEDV          TTCTCTCTGGTACAGGCCGCACCTCAGGGCCTGCTGTTCTTACATACAGTACTTGTACCG
SARSUrba      ATGTCCTTCCCACAAGCAGCCCCGCATGGTGTTGTCTTCCTACATGTCACGTATGTGCCA
SARSTor2      ATGTCCTTCCCACAAGCAGCCCCGCATGGTGTTGTCTTCCTACATGTCACGTATGTGCCA
```

```
BcoV        ATATCATTAGTGCAGAATGCTCCATATGGTTTGTATTTTATCCACTTTAGCTATGTCCCT
HEV         ATATCATTAGTACAGAATGCTCCATATGGTTTGTATTTTATCCATTTTAGCTATGTCCCC
MHV         TTGTCATTAGTCCAGAATGCGCCTTATGGTTTATATTTTATTCATTTCAGCTATGTGCCT
RtCoV       TTGTCATTAGTCCAGAATGCGCCTTACGGTTTATATTTTATTCATTTCAGCTATGTGCCT
IBV         CTAACCATACCGCAAAATGCACCTAATGGTATAGTGTTTATACACTTTTCTTATACTCCA
             *  *  *     *    ** **  * **  *    **  * **           **

CcoV        ACAGCTTATGAAACTGTGACGGCCTGGTCAGGTATTTGTGCGT---CAGATGGCAGTCGC
FcoV        ACAGCTTATGAAACTGTAACAGCTTGGTCAGGTATTTGTGCTT---CAGATGGCGATCGC
TGE         ACGGCTTATGAAACTGTGACTGCTTGGCCAGGTATTTGTGCTT---CAGATGGTGATCGC
PRCoV       ACGGCGTATGAAACTGTGACTGCTTGGTCAGGTATTTGTGCTT---TAGATGTTGATCGC
HCoVOC43    ACACAATATAAGGATGTTGAAGCGTGGTCTGGGTTGTGCGTTG---ATGGTACAAACGGT
PEDV        GGTGATTTTGTAAATGTTCTTGCCATCGCTGGCTTATGCGTTA---ATGGTGAAATTGCC
SARSUrba    TCCCAGGAGAGGAACTTCACCACAGCGCCAGCAATTTGTCATG---AAGGCAAAGCATAC
SARSTor2    TCCCAGGAGAGGAACTTCACCACAGCGCCAGCAATTTGTCATG---AAGGCAAAGCATAC
BcoV        ACTAAGTATGTCACTGCGAAGGTTAGTCCCGGTCTGTGCATTG---CTGGTGATAGAGGT
HEV         ACCAAGTATGTTACAGCAAAGGTTAGTCCTGGTTTGTGCATTG---CTGGCGATATAGGA
MHV         ACTTCCTTTACAACGGCAAATGTGAGTCCTGGGCTATGCATTT---CTGGTGATAGAGGA
RtCoV       ACATCCTTTACAACGGTAAATGTGAGTCCTGGACTATGCATTT---CTGGTGATAGAGGA
IBV         GATAGTTTTGTTAATGTTACTGCAATAGTGGGTTTTTGTGTAAAGCCAGCTAATGCTAGT
                                          *   * **          *

CcoV        ACTTTTGGACTTGTTGTTGAGGATGTCCAGCTGACGC-TATTTCGCAA-----TTTAGAT
FcoV        ACTTTCGGACTTGTCGTTAAAGATGTGCAGTTGACGT-TGTTTCGTAA-----TCTAGAT
TGE         ACTTTTGGACTTGTCGTTAAAGATGTCCAGTTGACTT-TGTTTCGTAA-----TCTAGAT
PRCoV       ACTTTTGGACTTGTCGTTAAAGATGTCCAGTTGACTT-TATTTCGTAA-----TCTAGAT
HCoVOC43    TATGTGTTGCGACAACCTAATCTTGCTCT--TTACAAAGAAGGCA---------ATTATT
PEDV        TTGACTCTACGTGAGCCTGGCTTAGTCTTGTTTACGCATGAACTTCAAACTTATACTGCG
SARSUrba    TTCCCT---------CGTGAAGGTGTTTTTGTGTTTAATG--------------GCACTT
SARSTor2    TTCCCT---------CGTGAAGGTGTTTTTGTGTTTAATG--------------GCACTT
BcoV        ATAGCC---------CCTAAGAGTGGTTATTTTGTTAATGTAA-----------ATAATA
HEV         ATATCG---------CCTAAGAGTGGTTATTTTATTAATGTAA-----------ATAATT
MHV         TTAGCA---------CCTAAAGCTGGATATTTTGTTCAAGATG------------ATGGAG
RtCoV       TTAGCA---------CCTAAAGCTGGATATTTTGTTCAAGATC-----------ATGGAG
IBV         CAGTATGCAATAGTGCCCGCTAATGGTAGGGGTATTTTTATACAA---------GTTAAT
                                     *

CcoV        GAAAAATTTTATTTGACGCCCAGAACTATGTATCAGCCCAGAGTTGCAACTAGTTCTGAT
FcoV        GACAAGTTCTATTTGACCCCCAGAACTATGTATCAGCCTAGAGTTGCAACTAGTTCTGAT
TGE         GACAAGTTCTATTTGACCCCCAGAACTATGTATCAGCCTAGAGTTGCAACTAGTTCTGAC
PRCoV       GACAAGTTCTATTTGACACCCAGAACTATGTATCAGCCTAGAGTGGCAACTAGTTCTGAT
HCoVOC43    ATAGAAT-------CACATCTCGCATAATGTTTGAACCACGTATTCCTACCATGGCAGAT
PEDV        ACGGAATATTTTGTTTCATCGCGACGTATGTTTGAACCTAGAAAACCTACCGTTAGTGAT
SARSUrba    CTTGGTTTAT----TACACAGAGGAACTTCTTTTCTCCACAAATAATTACTACAGACAAT
SARSTor2    CTTGGTTTAT----TACACAGAGGAACTTCTTTTCTCCACAAATAATTACTACAGACAAT
BcoV        CTTGGATGTT----CACTGGTAGTGGTTATTACTACCCTGAACCCATAACTGGAAATAAT
HEV         CTTGGATGTT----CACTGGTAGTAGCTATTACTACCCTGAACCTATAACCCAAAATAAT
MHV         AGTGGAAGTT----CACAGGTAGTAATTATTATTACCCTGAACCCATTACAGATAAAAAT
RtCoV       AATGGAAGTT----CACAGGTAGCAATTATTACTACCCTGAATCCATTACAGATAAAAAC
IBV         GGTAGTTACTACATCACTGCACGAGATATGTATATGCCAAGAGCTATTACTGCAGGAGAT
                          *     *       *     **          **       *

CcoV        TTTGTTCAAATAGAAGGCTGTGATGTGTTGTTTGTTAATGGAACTGTAATTGAATTGCCT
FcoV        TTTGTTCAAATTGAAGGGTGTGATGTGTTGTTTGTCAACGCGACTGTAATTGATTTGCCT
TGE         TTTGTTCAAATTGAAGGGTGCGATGTGCTGTTTGTTAATGCAACTGTAAGTGATTTGCCT
PRCoV       TTTGTTCAAATTGAAGGGTGCGATGTGCTGTTTGTTAATACAACTGTAAGTGATTTGCCT
HCoVOC43    TTTGTTCAAATTGAAAATTGCAATGTCACATTTGTTAACATTTCTCGCTCTGAGTTGCAA
PEDV        TTTGTTCAAATTGAGAGTTGTGTGGTCACCTATGTCAATCTGACTAGCGACCAGCTACCA
SARSUrba    ACATTTGTCTCAGGAAATTGTGATGTCGTTATTGGCATCATTAACAACACAGTTTATGAT
SARSTor2    ACATTTGTCTCAGGAAATTGTGATGTCGTTATTGGCATCATTAACAACACAGTTTATGAT
BcoV        GTTGTTGTTATGAGTACCTGTGCTGTTAACTATACTAAAGCGCCGGATGTAATGCTGAAC
HEV         GTTGTTGTGATGAGTACCTGTGCTGTTAATTATACTAAAGCACCGGATCTAATGCTGAAC
MHV         AGTGTCGTGATGAGTAGTTGCGCAGCAAACTACACAAAGGCACCTGAAGTTTTCTTGAAC
```

```
RtCoV       AGTGTCGTGATGAGTAGTTGCGCAGTAAACTACACAAAGGCACCTGAAGTTTTCTTGAAC
IBV         GTAGTTACGCTTACTTCTTGTCAAGCAAATTATGTAAGTGTAAATAAGACCGTCATTACT
                *             **    *          *

CcoV        AGTATCATA---CCTGACTATATCGATATTAATCAAACTGTTCAGGACATATTAGAAAAT
FcoV        AGTATTATA---CCTGACTATATTGACATTAATCAAACTGTTCAAGACATATTAGAAAAT
TGE         AGTATTATA---CCTGATTATATTGATATTAATCAGACTGTTCAAGACATATTAGAAAAT
PRCoV       AGTATTATA---CCTGATTATATTGATATTAATCAGACTGTTCAAGACATATTAGAAAAT
HCoVOC43    ACCATTGTG---CCAGAGTATATTGATGTTAATAAGACGCT--GCAAGAATTAAGTTACA
PEDV        GATGTAATC---CCAGATTACATCGATGTTAACAAAACACTTGATGAGATTTTAGCTTCT
SARSUrba    CCTCTGCAA---CCTGAGCTCGACTCATTCAAAGAAGAGCTGGACAAGTACTTCAAAAAT
SARSTor2    CCTCTGCAA---CCTGAGCTTGACTCATTCAAAGAAGAGCTGGACAAGTACTTCAAAAAT
BcoV        ATTTCAACA---CCCAACCTCCATGATTTTAAGGAAGAGTTGGATCAATGGTTTAAAAAC
HEV         ACATCGACA---CCCAACCTTCCTGACTTCAAGGAAGAATTGTATCAATGGTTTAAAAAC
MHV         ACTTCAATA---CCTAATCTACCCGACTTTAAGGAGGAGTTAGATAAATGGTTTAAAAAT
RtCoV       ACTTCAATA---ACTAATCTACCCGACTTTAAGGAGGAGTTAGATAAATGGTTTAAGAAT
IBV         ACATTCGTAGACAATGATGATTTTGATTTTAATGACGAATTGTCAAAATGGTGGAATGAT
                          *           * **  *     *     *    *

CcoV        TTCAGACCAAATTGGACTGTACCCGAGTTGCCACTTGACATTTTTCATGCAACCTACTTA
FcoV        TACAGACCAAACTGGACTGTACCTGAATTTACACTTGATATTTTCAACGCAACCTATTTA
TGE         TTTAGACCAAATTGGACTGTACCTGAGTTGACATTTGACATTTTTAACGCAACCTATTTA
PRCoV       TTTAGACCAAATTGGACTGTACCTGAGCTGACATTGGACGTTTTTAACGCAACCTATTTA
HCoVOC43    AATTG-CCAAATTACACTGTTCCAGACCTAGTTGTCGAACAGTACAACCAGACTATTTTG
PEDV        --CTG-CCCAATAGAACTGGTCCAAGTCTTCCCCTAGATGTTTTTAATGCCACTTATCTT
SARSUrba    CATACATCACCAGATGTTGATCT-TGGCGACATTTCAGGCA--TTAACGCTTCTGTCGTC
SARSTor2    CATACATCACCAGATGTTGATCT-TGGCGACATTTCAGGCA--TTAACGCTTCTGTCGTC
BcoV        CAAACATCAGTGGCACCAGATTTGTCACTTGATTATA------TAAATGTTACATTCTTG
HEV         CAATCTTCAGTGGCACCAGATTTGTCACTTGATTATA------TTAATGTTACGTTCTTG
MHV         CAGACGTCTATTGCGCCTGATTTATCTCTCGATTTCGAGAAATTAAACGTTACCCTCCTG
RtCoV       CAGACGTCTATTGTGCCTGATTTATCTTTCGATATCGGGAAATTAAATGTTACATTCCTT
IBV         -ACTAAGCATGAGCTACCAGACTTTGAC--AAATTCAATTA---CACAGTACCTATACTT
                  *                                                *    *

CcoV        AACCTGACTGGTGAAATTAATGACTTAGAATTTAGGTCAGAAAAGTTACATAACACCACA
FcoV        AATCTGACTGGTGAAATTGATGACTTAGAGTTTAGGTCAGAAAAGCTACATAACACTACA
TGE         AACCTGACTGGTGAAATTGATGACTTAGAATTTAGGTCAGAAAAGCTACATAACACTACA
PRCoV       AACCTGACTGGTGAAATTGATGACTTAGAGTTTAGGTCAGAAAAGCTACATAACACCACT
HCoVOC43    AATTTGACCAGTGAAATTAGCACCCTTGAAAATAAATCTGCGGAGCTTAATTACACTGTT
PEDV        AATCTTACTGGTGAAATTGCAGATCTAGAGCAGCGTTCAGAGTCTCTCCGTAATACTACA
SARSUrba    AACATTCAAAAAGAAATTGACCGCCTCAATGAGG-------------------------
SARSTor2    AACATTCAAAAAGAAATTGACCGCCTCAATGAGG-------------------------
BcoV        GACCTACAAGATGAAATGAATAGGTTACAGGAGG-------------------------
HEV         GACCTACAAGATGAAATGAATAGGTTACAAGAGG-------------------------
MHV         GACCTGACTGATGAGATGAACAGGATTCAGGATG-------------------------
RtCoV       GACCTGTCCTATGAGATGAACAGGATTCAGGATG-------------------------
IBV         GACATTGATAGTGAAATTGATCGTATTCAAGGCG-------------------------
             *  *        ** **         *  *

CcoV        GTAGAACTTGCTATTCTCATTGATAATATTAATAACACATTAGTCAATCTTGAATGGCTC
FcoV        GTAGAACTTGCCATTCTCATTGATAACATTAATAATACATTAGTCAATCTTGAATGGCTC
TGE         GTAGAACTTGCCATTCTCATTGACAACATTAACAATACATTAGTCAATCTTGAATGGCTC
PRCoV       GTAGAACTTGCCATTCTCATTGACAACATTAACAATACAGTAGTCAATCTTGAATGGCTT
HCoVOC43    CAAAAATTGCAAACTCTGATTGACAACATAAATAGCACATTAGTCGACTTAAAGTGGCTC
PEDV        GAAGAGCTCCGAAGTCTCATTAACAACATCAACAACACACTTGTTGACCTTGAGTGGCTC
SARSUrba    ----------------TCGCTAAAAATTTAAATGAATCACTCATTGACCTTCAAGAATTG
SARSTor2    ----------------TCGCTAAAAATTTAAATGAATCACTCATTGACCTTCAAGAATTG
BcoV        ----------------CAATAAAAGTTTTAAATCAGAGCTACATCAATCTCAAGGACATT
HEV         ----------------CTATAAAAGTTTTAAATCAGAGCTACATCAATCTCAAGGACATT
MHV         ----------------CAATTAAGAAGTTAAATGAGAGTTACATCAACCTCAAGGACGTT
RtCoV       ----------------CAATTAAGAATTTAAATGAGAGTTACATCAACCTCAAGGAAATT
IBV         ----------------TTATACAGGGTCTTAATGACTCTCTAATAGACCTTGAAAAACTT
                              *     * **          *  *  *  *     *
```

```
CcoV          AACAGAATTGAAACTTATGTAAAATGGCCTTGGTATGTTTGGCTACTAATTGGATTAGTA
FcoV          AATAGAATTGAAACTTATGTAAAATGGCCTTGGTATGTGTGGCTACTGATAGGTTTAGTA
TGE           AATAGAATTGAAACCTATGTAAAATGGCCTTGGTATGTGTGGCTACTAATAGGCTTAGTA
PRCoV         AATAGAATTGAAACTTATGTAAAATGGCCTTGGTATGTGTGGCTACTAATAGGCTTAGTA
HCoVOC43      AACCGGGTTGAGACTTACATCAAGTGGCCGTGGTGGGTGTGGTTGTGCATTTCAGTCGTG
PEDV          AACCGAGTTGAGACATACATCAAGTGGCCGTGGTGGGTTTGGTTGATCATTGTTATTGTT
SARSUrba      GGAAAATATGAGCAATATATTAAATGGCCTTGGTATGTTTGGCTCGGCTTCA---TTGCT
SARSTor2      GGAAAATATGAGCAATATATTAAATGGCCTTGGTATGTTTGGCTCGGCTTCA---TTGCT
BcoV          GGTACATATGAGTATTATGTAAAATGGCCTTGGTATGTATGGCTTTTAATTGGCTTTGCT
HEV           GGTACATATGAGTATTATGTGAAATGGCCTTGGTATGTATGGCTTTTAATTGGCTTTGCT
MHV           GGCACATATGAAATGTATGTGAAATGGCCTTGGTATGTATGGCTTTTAATTGGCCTTGCT
RtCoV         GGCACATATGAGATGTATGTGAAATGGCCTTGGTATGTTTGGCTGCTAATTGGATTAGCT
IBV           TCAATACTCAAAACTTATATTAAGTGGCCTTGGTATGTGTGGTTAGCCATAGCTTTTGCC
                     *    **  * ** ***** ****  ** *** *     *     * *

CcoV          GTAATATTCTGCATACCCATATTGCTATTTTGTTGTTGTAGTACTGGTTGTTGTGGATGT
FcoV          GTAGTATTTTGCATACCATTACTGCTATTTTGCTGTTTTAGCACAGGTTGTTGTGGATGC
TGE           GTAATATTTTGCATACCATTACTGCTATTTTGCTGTTGTAGTACAGGTTGCTGTGGATGC
PRCoV         GTAATATTTTGCATACCATTACTGCTATTTTGCTGTTGTAGTACAGGTTGCTGTGGATGC
HCoVOC43      CTCATCTTTGTGGTGAGTATGTTGCTATTATGTTGTTGTTCTACTGGTTGCTGTGGCTTC
PEDV          CTCATCTTTGTTGTGTCATTACTAGTGTTCTGCTGCATTTCCACGGGTTGTTGTGGATGC
SARSUrba      GGACTAATTGCCATCGTCATGGTTACAATCTTGCTTTGTTGCATGACTAGTTGTTGCAGT
SARSTor2      GGACTAATTGCCATCGTCATGGTTACAATCTTGCTTTGTTGCATGACTAGTTGTTGCAGT
BcoV          GGTGTAGCTATGCTTGTTTTACTATTCTTCATATGCTGTTGTACAGGATGTGGGACTAGT
HEV           GGTGTAGCTATGCTTGTTTTACTATTCTTCATATGCTGCTGTACAGGATGTGGGACTAGT
MHV           GGTGTAGCTGTTTGTGTGTTGTTATTTTTCATATGTTGCTGCACGGGTTGTGGCTCATGT
RtCoV         GGTGTAGCTGTTTGTGTTTTGTTATTTTTTATATGTTGCTGCACAGGTTGTGGCTCTTGT
IBV           ACTATTATCTTCATCTTAATACTAGGATGGGTTTTCTTCATGACTGGTTGTTGTGGTTGT
                  *                *  *                    *      *  *

CcoV          ATCGGGTGTTTAGGAAGCTGTTGTCATTCCATAT-GTAGTAGAGGCCA---ATTTGAAAG
FcoV          ATAGGTTGTTTAGGAAGTTGTTGTCACTCTATAT-GTAGTAGAAGACA---ATTTGAAAA
TGE           ATAGGTTGTTTAGGAAGTTGTTGTCACTCTATAT-GTAGTAGAAGACA---ATTTGAAAA
PRCoV         ATAGGTTGTTTAGGAAGTTGTTGTCACTCTATAT-TCAGTAGAAGACA---ATTTGAAAA
HCoVOC43      TTTAGTTGTTTTGCATCTTCTATTAGAGGTTGTT-GTGAATCAACTAA---ACTTCCTTA
PEDV          TGCGGTTGCTGCGGTGCTTGTTTTTCAGGTTGTT-GTAGGGGTCCTAG---ACTTCAACC
SARSUrba      TGCCTCAAGGGTGCATGCTCTTGTGGTTCTTGCT-GCAAGTTTGATGA---GGATGACTC
SARSTor2      TGCCTCAAGGGTGCATGCTCTTGTGGTTCTTGCT-GCAAGTTTGATGA---GGATGACTC
BcoV          TGTTTTAAGAT---ATG---TGGTGGTTGTTGTG-ATGATTATACTGG---ACACCAGG-
HEV           TGTTTTAAGAA---ATG---TGGCGGTTGTTGTG-ATGATTATACTGG---ACACCAGG-
MHV           TGTTTCAAGAA---GTG---TGGAAATTGTTGTG-ATGAGTGTGGAGG---ACACCAGGA
RtCoV         TGTTTTAAGAA---ATG---TGGAAATTGTTGTG-ATGAGTATGGAGG---ACGTCAGGC
IBV           TGTTGTGGATGC-TTTGGCATTATGCCTCTAATGAGTAAGTGTGGTAAGAAATCTTCTTA
                           *

CcoV          TTATGAACCTATTGAAAAAGTTCATGTTCACTGA-------------------------
FcoV          TTATGAACCAATTGAAAAAGTGCATGTCCACTAA-------------------------
TGE           TTACGAACCAATTGAAAAAGTGCACGTCCATTAA-------------------------
PRCoV         TTATGAACCTATTGAAAAAGTGCACGTCCATTAA-------------------------
HCoVOC43      TTACGACG---TTGAAAAGATCCACATACAGTAA-------------------------
PEDV          TTACGAAGCTTTTGAAAAGGTCCACGTGCAGTGA-------------------------
SARSUrba      TGAGCCAGTTCTCAAGGGTGTCAAATTACATTACACATAA-------------------
SARSTor2      TGAGCCAGTTCTCAAGGGTGTCAAATTACATTACACATAA-------------------
BcoV          -AGT-TAGTAAT-TAAAA-----CATCACATGACGACTAA-------------------
HEV           -AGT-TTGTAAT-CAAAA-----CTTCACATGACGATTAA-------------------
MHV           CAGTATTGTGATACATAATATTTCCTCTCATGAGGATTGA-------------------
RtCoV         AGGTATTGTGATACATAATATTTCCTCTCATGAGGATTGA-------------------
IBV           TTACACGACTTTTGATAACGATGTGGTAACTGAACAATACAGACCTAAAAAGTCTGTTTG
                        *  *

CcoV          -
FcoV          -
```

TGE	-
PRCoV	-
HCoVOC43	-
PEDV	-
SARSUrba	-
SARSTor2	-
BcoV	-
HEV	-
MHV	-
RtCoV	-
IBV	A

Figure 2: **Phylogenetic Analysis of *S* Gene**

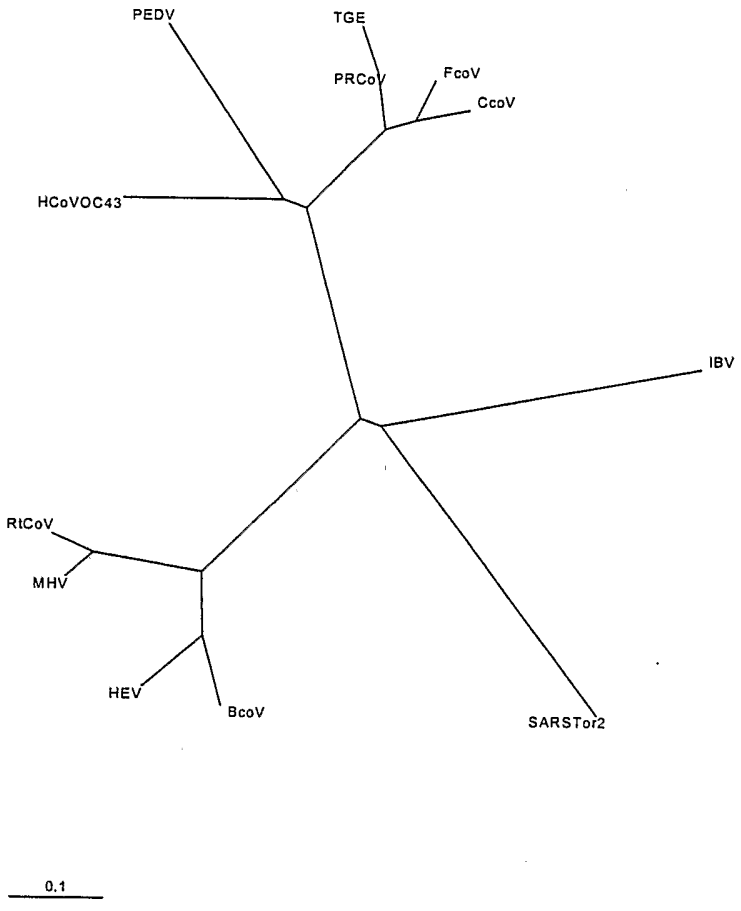
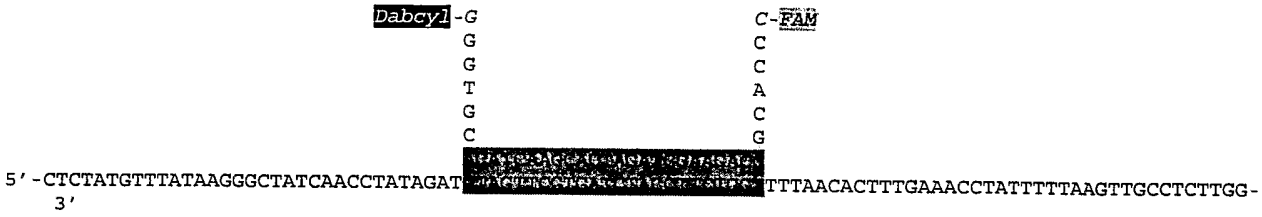


Figure 3 : Molecular Designs for SARS-Associated *S* Gene LK249 (Molecular Beacon) LK250 (Amplicon) and LK251 & LK252 (PCR Primers)

NOTE:	LK249 is a TET-molecular beacon, which recognizes the ***S* gene of coronavirus (SARS Tor2 and SARS urbani human pathogenic strains)**. It will be used in a real-time PCR diagnostic for the identification of **SARS-associated coronavirus RNA/DNA.**

Beacon **LK249**



Amplicon **LK250**

LK249

A.	Target recognition sequence:	**24 nucleotides (11 G/C)**
B.	Length of the arms:	**6 nucleotides (5 G/C)**
C.	Melting temperature of the beacon:	**dG = -2.37 dH = -46.3 dS = -135.9 Tm = 67.7 °C**
D.	Melting temperature of target:	**°C**

LK251 **5'-CTCTATGTTTATAAGGGCTATCAACC-3'**

Length: 26 nucleotides (10 G/C)
Tm: °C
Position: (see alignment below)

LK252 **5' –CCAAGAGGCAACTTAAAAATAGGTTTC- 3'**

Length: 27 nucleotides (10 G/C)
Tm: °C
Position: (see alignment below)

S-RT **5'-AGGCTGTAAGAA-3'**

Length: 14 nucleotides (6 G/C)
Tm: 47 °C
Position: (see alignment below)

LK250 Amplicon (95 nucleotides)

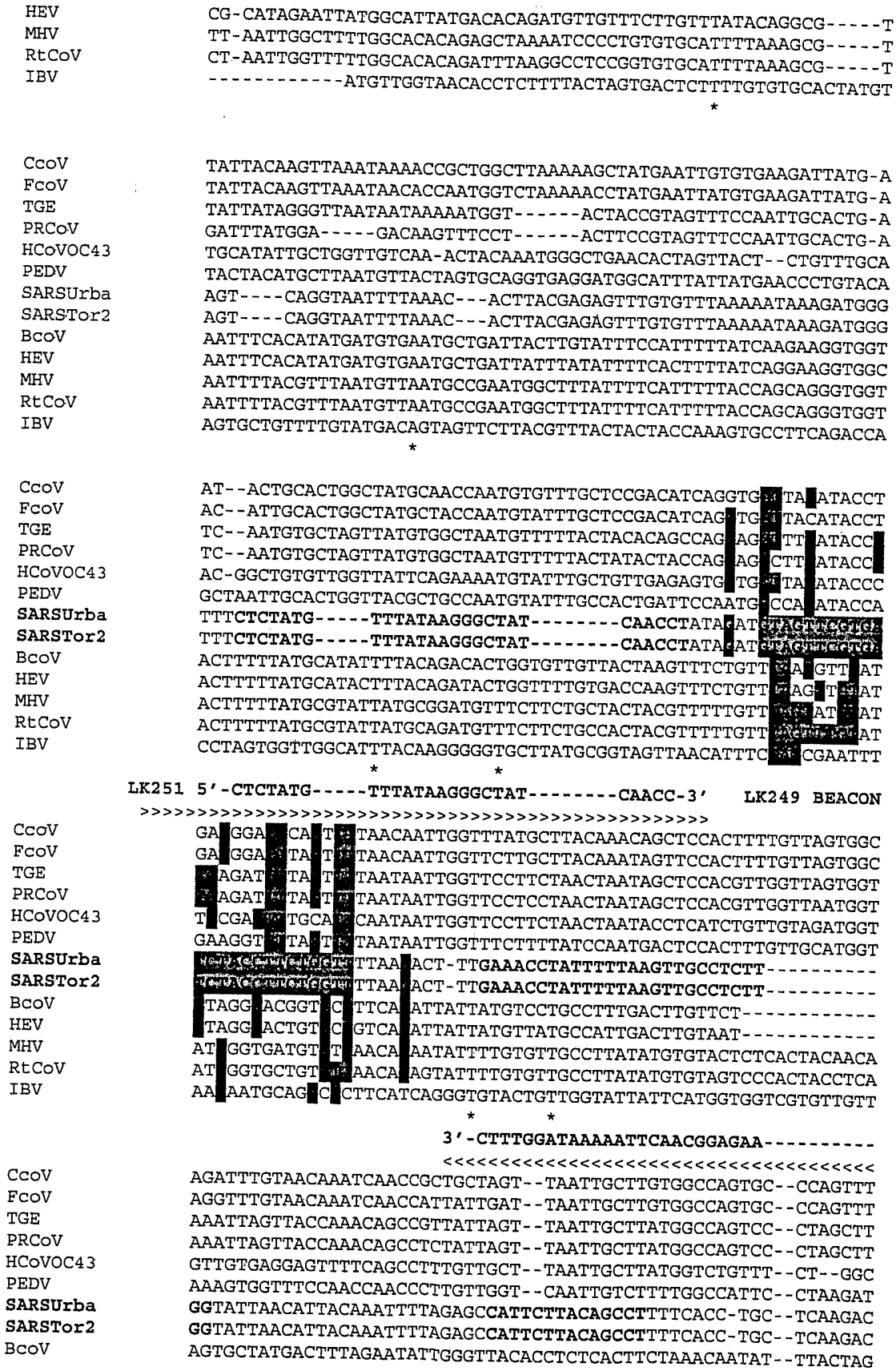
5' – CTCTATGTTTATAAGGGCTATCAACCTATAGATGTAGTTCGTGATCTA CCTTCTGGTTTTAACACTTTGAAACCTATTTTTAAGTTGCCTCTTGG -3'

DNA sequence alignment of complete S genes from corona virus strains

```
CcoV        ATGATTGTGCTTACATTGTGCCTTTTCTTGTT---TTTGTACAGTAGTGTGAGCTGTACA
FcoV        ATGATTGTGCTCGTAACTTGCCTCTTGTTGTTATGTTCATACCACACAGTTTTGAGTACA
TGE         ATGAAAAAACTATTTGTGGTTTTGGTCGTAATGCCATTGATTTATGGAGACAATTTTCCT
PRCoV       ------------------------------------------------------------
HCoVOC43    ------------------------------------------------------------
PEDV        ------------------------------------------------------------
SARSUrba    ------------------------------------------------------------
SARSTor2    ------------------------------------------------------------
BcoV        ------------------------------------------------------------
HEV         ------------------------------------------------------------
MHV         ------------------------------------------------------------
RtCoV       ------------------------------------------------------------
IBV         ------------------------------------------------------------

CcoV        TCAAACAATGACTGTGTACAAGTTAATGTGACACAACTGCCTGGCAATGAAAATATTATC
FcoV        ACAAATAATGAATGCATACAAGTTAACGTAACACAATTGGCTGGCAATGAAAACCTTATC
TGE         TGTTCTAAATTGACTAATAGAACTATAGGCAACCAGTGGAATCTCATTGAAACCTTCCTT
PRCoV       ------------------------------------------------------------
HCoVOC43    ------------------------------------------------------------
PEDV        ----------------------------ATGAGGTCTTTAATTTACTTCTGGTTGCTCTT
SARSUrba    ------------------------------------------------------ATGTTT
SARSTor2    ------------------------------------------------------ATGTTT
BcoV        ------------------------------------------------ATGTTTTTGATACTT
HEV         ----------------------------------------------ATGTTTTTTATACTT
MHV         ----------------------------------------------ATGCTATTCGTGTTT
RtCoV       ----------------------------------------------ATGCTATTCGTGTTT
IBV         ------------------------------------------------------------

CcoV        AAAGATTTTCTATTTCAGAACTTTAAAGAAGAAGGAAGTTTAGTTGTTGGTGGTTATTAC
FcoV        AGAGATTTTCTGTTTAGTAACTTTAAAGAAGAAGGAAGTGTAGTTGTTGGTGGTTATTAC
TGE         CTAAACTATAGTAGTAGGTTACCACCTAATTCAGATGTGGTGTTAGGTGATTATTTTCCT
PRCoV       ------------------------------------------------------------
HCoVOC43    ------------------------------------------------------------
PEDV        ACCAGTACTTCCAACACTCAGCCTACCACAAGATGTCACTAGGTGCCAGTCTACTACTAA
SARSUrba    ATTTTCTTATTATTTCTTACTCTCACTAGTGGTAGTGACCTTGACCGGTGCACCACTTTT
SARSTor2    ATTTTCTTATTATTTCTTACTCTCACTAGTGGTAGTGACCTTGACCGGTGCACCACTTTT
BcoV        TTAATTTCCTTACCAATGGCTTTTGCTGTTATAGGAGATTTAAAGTGTACTACGGTTTCC
HEV         TTAATCTCCCTGCCTTCTGCTTTTGCAGTTATAGGGGATTTAAAGTGTACTACTTCATTA
MHV         TTAACCTTGTTGCCCTCTTCTCTAGGGTATATTGGTGATTTTAGATGTATCCAACTTGTA
RtCoV       TTAACCCTATTGCCCTCTTGTCTAGGGTATATTGGTGATTTTAGATGTATCAACCTTGTA
IBV         ------------------------------------------------------------

CcoV        CCCACAGAGGTGTGGTATAACTGTTCCACAACTCAACAAACTACCGCTTATAAGTATTTT
FcoV        CCTACAGAGGTGTGGTACAACTGCTCTAGAACAGCTCGAACTACTGCCTTTCAGTATTTT
TGE         ACTGTACAACCTTGGTTTAATTGCATTCGCAATGATAGTAATGACCTTTATGTTACACTG
```

```
PRCoV       ------------------------------------------------------------
HCoVOC43    ------------------------------------------------------------
PEDV        CTTTAGGCGGTTCTTTTCAA--AATTTAATGTTCAGGCACCTGCCGTCGTCGTTTTGGGT
SARSUrba    GA-----------TGATGTTCAAGCTCCTA--ATTACACTCAACATACTTCAT----CT
SARSTor2    GA-----------TGATGTTCAAGCTCCTA--ATTACACTCAACATACTTCAT----CT
BcoV        ATTAATGATGT---TGACACCGGTGCTCCCTCTATTAGCACTGATATTGTCGATGTTACT
HEV         ATTAATGACGT---TGACACTGGTGTGCCATCTATTAGCTCTGAAGTTGTTGATGTCACT
MHV         AATACCGACACCTCTAATGCCAGCGCTCCAAGCGTTAGTACAGAGGTAGTTGATGTTTCC
RtCoV       AACACCCGCATTTCTAATGCGCGCGCACCCAGTGTTAGCACAGAGGTAGTTGATGTTTCT
IBV         ------------------------------------------------------------

CcoV        AGTAATATACATGCATTTTATTTTGATATGGAAGCCATGGAGAATAGTACTGGCAATGCA
FcoV        AATAATATACATGCCTTTTATTTTGTTATGGAAGCCATGGAAAATAGCACTGGTAATGCA
TGE         GAAAATCTTAAAGCATTGTATTGGGATT---ATGCTACAGAAAATATCACTTGGAAT---
PRCoV       ------------------------------------------------------------
HCoVOC43    ------------------------------------------------------------
PEDV        GGTTACCTACCTAGTATGAACTCTTCTAGCTGGTACTGTGGCACAGGCATTGAAACTGCT
SARSUrba    ATGAG---GGGGGTTTACTATCCT---GATGAAATT--TTTAGATCAGACACTCTT----
SARSTor2    ATGAG---GGGGGTTTACTATCCT---GATGAAATT--TTTAGATCAGACACTCTT----
BcoV        AATGGTTTAGGTACTTATTATGTTTTAGATCGTGTGTATTTAAATACTACGTTGTTG---
HEV         AATGGTTTGGGGACTTTCTATGTTTTAGATCGTGTCTATTTAAATACCACATTGTTG---
MHV         AAAGGGATTGGTACTTATTATGTTTTAGATCGAGTCTATTTAAATGCCACACTATTG---
RtCoV       AAAGGTCTTGGTACATATTACGTTTTAGATCGTGTTTATTTAAATGCCACGTTATTG---
IBV         ------------------------------------------------------------

CcoV        CGTGGTAAACCTTTACTAGTACATGTTCATGGTAATCCTGTTAGTATCATTGTTTACATA
FcoV        CGTGGTAAACCATTATTATTTCATGTGCATGGTGAGCCTGTTAGTGTTATT------ATA
TGE         CACAGACAACGGTTAAACGTAGTCGTTAATGGATACCCATACTCCATCACAGTT---ACA
PRCoV       ------------------------------------------------------------
HCoVOC43    ------------------------------------------------------------
PEDV        AGTGGCGTTCATGGTATTTTTCTCAGCTACATCGATTCTGGTCAGGGCTTTGAGA--TTG
SARSUrba    -----TATTTA--ACTCAGGATTTATTTCTTC-CATTTTATTCTAATGTTACAGG--GTT
SARSTor2    -----TATTTA--ACTCAGGATTTATTTCTTC-CATTTTATTCTAATGTTACAGG--GTT
BcoV        CTTAATGGTTACTACCCTACTTCAGGTTCTACATATCGTAATATGGCACTGAAGG--GAA
HEV         CTCAATGGTTATTACCCAATTTCAGGTGCTACATTTCGTAATATGGCTCTGAAAG--GAA
MHV         CTTACTGGTTATTACCCTGTAGATGGGTCCATGTATAGAAACATGGCTCTAACGG--GAA
RtCoV       CTTACTGGTTACTACCCTGTAGATGGGTCCATGTATCGTAACATGGCTCTAATGG--GTA
IBV         ------------------------------------------------------------

CcoV        TCAGCTTATAGAGATGATGTGCAATTTAGGCCGCTTTTAAAGCATGGTTTATTGTGTATA
FcoV        TCGGCTTATAGGGATGATGTGCAACAAAGGCCCCTTTTAAAACATGGGTTAGTGTGCATA
TGE         ACAACCCGCAATTTTAAT-TCTGCTGAAGGTGCTATTATATGCATTTGTAAGGGCTCACC
PRCoV       ------------------------------------------------------------
HCoVOC43    ------------------------------------------------------------
PEDV        GCATTTCGCAAGAGCCGT-TTGATCCTAGTGGTTACCAGCTTTATTTACATAAGGCCACT
SARSUrba    TCATACTATTAATCATAC----GTTTGG--CAACCCTGTCAT-ACCTTTTAAGGATGGTA
SARSTor2    TCATACTATTAATCATAC----GTTTGG--CAACCCTGTCAT-ACCTTTTAAGGATGGTA
BcoV        CTTTACTATTGAGCAGACTATGGTTTAAACCACCTTTTCTTTCTGATTTTATTAATGGTA
HEV         CTCGATTATTGAGCACCTTGTGGTTTAAGCCGCCTTTTTTATCACCTTTTAATGATGGTA
MHV         TTAATACCATAAGCCTTAATTGGTACAAACCACCCTTTTTATCAGAGTTTAATGATGGCA
RtCoV       CTAATACCTTAAGCCTTAATTGGTTTGAACCGCCCTTTTTATCAGAGTTTAACGATGGCA
IBV         ------------------------------------------------------------

CcoV        ACTAAAAATGACACCGTTGACTATAATAGCTTTACAATTAACCAATGGCGAGACATATGT
FcoV        ACTAAAAATCGCCATATTAACTATGAACAATTCACCTCCAACCAGTGGAATTCCACATGT
TGE         ACCTACTACCACCACAGAATCTA-----GTTTGACTTGCAATTGGGGTAGTGAGTGCAGG
PRCoV       ------------------------------------------------------------
HCoVOC43    ------------------------------------------------------------
PEDV        AATGGTAACACTAATGCTATTGCACGACTGCGCATTTGCCAGTTTCCCGATAATAAAACA
```

```
SARSUrba        TTTA-TTTTGCTGCCACAGAGAAATCAAATGTTGTCCGTGGTTGGGTTTTTGGTTCTACC
SARSTor2        TTTA-TTTTGCTGCCACAGAGAAATCAAATGTTGTCCGTGGTTGGGTTTTTGGTTCTACC
BcoV            TTT----TTGCTAAGGTCAAAAATACCAAGGTTATTAAAAAGGGTGTAATGTATAGTGAG
HEV             TTT----TTGCCAAGGTTAAAAACAGCAGATTTTCTAAAGATGGTGTTATTTATAGTGAG
MHV             TAT----TTGCTAAGGTAAAGAACCTTAAAGCATCTTTGCCCAAAGATTCTATTTCATAT
RtCoV           TAT----ATGCTAAGGTAAAGAACCTCAAAGCATCTTTGCCCATAGGCTCGGCTTCATAC
IBV             ------------------------------------------------------------

CcoV            TTGGGTGACGACAGAAAAATACCATTCTCTGTAGTACCCACAGATAATGGTACGAAATTA
FcoV            ACGGGTGCTGACAGAAAAATTCCTTTCTCTGTCATACCCACGGACAATGGAACAAAAATC
TGE             TTAAACCATAAGTTCCCTATATGTCCTTCTAATTCAGAGGCAAATTGTGGTAATATGCTG
PRCoV           ------------------------------------------------------------
HCoVOC43        ------------------------------------------------------------
PEDV            TTGGGCCCTACTGTTAATGATGTTACAACAGGTCGTAACTGCCTATTCAACAAAGCCATT
SARSUrba        ATGAACAACA-AGTCACAGTCGGTGATTA--TTATTAACAATTCTACTAATGTTGTTATA
SARSTor2        ATGAACAACA-AGTCACAGTCGGTGATTA--TTATTAACAATTCTACTAATGTTGTTATA
BcoV            TTTCCTGCTATAACTATAGGTAGTACTT---TTGTAAATACATCCTATAGTGTGGTAGTA
HEV             TTTCCTGCTATTACTATAGGTAGTACTT---TTGTAAATACTTCCTATAGCATAGTAGTA
MHV             TTCCCTACTATAATTATAGGTAGTAATT---TTGTCACCACTTCCTATACTGTAGTATTG
RtCoV           TTTCCTACTATAATTATAGGTAGTAATT---TTGTTAATACTTCCTATACTGTAGTATTG
IBV             ------------------------------------------------------------

CcoV            TTTGGTCTTGAGTGGAATGATGACTATGTTACAGCCTATATTAGTGATGAGTCTCACCGT
FcoV            TATGGTCTTGAGTGGAATGATGACTTTGTTACAGCTTATATTAGTGGTCGTTCTTATCAC
TGE             TATGGCCTACAATGGTTTGCAGATGAGGTTGTTGCTTATTTACATGGTGCTAGTTACCGT
PRCoV           ------------------------------------------------------------
HCoVOC43        ------------------------------------------------------------
PEDV            CCAGCTTATATGCGTGATGGAAAAGATATTGTTGTCGGCATAACATGGGATAATGATCGT
SARSUrba        CGAGCATGTA------------ACTTTGAAT-------TGTGTGACAACCCTTTCTTTGC
SARSTor2        CGAGCATGTA------------ACTTTGAAT-------TGTGTGACAACCCTTTCTTTGC
BcoV            CAACCACATACTACCAATTTGGATAATAAATTACAAGGTCTCTTAGAGATCTCTGTTTGC
HEV             GAGCCTCATACCTCACTTATTAATGGTAATTTACAAGGTTTGTTGCAAATTTCTGTTTGT
MHV             GAACCGTATA------------ATGGTA-----------TAATTAT-GGCATCCATTTGC
RtCoV           GAACCATACA------------ATGGTA-----------TTATTAT-GGCATCTATTTGC
IBV             ------------------------------------------------------------

CcoV            TTGAATATCAATAATAATTGGTTTAACAATGTTACACTC--CTATACTCACGTACAAGCA
FcoV            TTGAACATCAATACTAATTGGTTTAACAATGTCACACTT--TTGTATTCACGCTCAAGCA
TGE             ATTAGTTTTGAAAATCAATGGTCTGGCACTGTCACATTTGGTGATATGCGTGCGACAACA
PRCoV           ------------------------------------------------------------
HCoVOC43        ------------------------------------------------------------
PEDV            GTCACTGTTTTTGCTGACAAGATCTATCATTTTTATCTT--AAAAATGATTGGTCCCGCG
SARSUrba        TGTTTCTAAACCCATGGGTACACAGACACATACTATGAT----ATTCGA----TAATGCA
SARSTor2        TGTTTCTAAACCCATGGGTACACAGACACATACTATGAT----ATTCGA----TAATGCA
BcoV            CAGTATACTATGTGCGAGTACCCACATACGATTTGTCATCCTAAGCTGGG---TAATAAA
HEV             CAATACACTATGTGTGAATACCCACATACTATTTGTCATCCTAATTTGGG---TAATCAA
MHV             CAGTATACCATTTGTCAACTACCGTACACGGATTGCAAACCGAATACGGGCGGTAATAAG
RtCoV           CAGTATACCATTTGTCAATTACCGCACACGGATTGCAAACCTAACACGGGCGGTAACACG
IBV             ------------------------------------------------------------

CcoV            CCGCCACGTGGCAACACA-GTGCTGCATATGTTTA---TCAAGGTGTTTCAAATTTTACT
FcoV            CTGCTACCTGGGAATACA-GTGCTGCATATGCTTA---CCAAGGTGTTTCTAACTTCACT
TGE             TTAGAAGTCGCTGGCACGCTTGTAGACCTTTGGTGGTTTAATCCTGTTTATGATGTCAGT
PRCoV           -------------------ATGAAAAAATTATTTG---TGGTCTTGGTTGTAATGCCATT
HCoVOC43        ----------------------ATGTT-TGTT-TTGCTTGTTGC--ATATGCCTTGT
PEDV            TTGCGACAAGATGTTACAATCGCAGAAGTTGTGCT-ATGCAATATGTTTATACACCTACC
SARSUrba        TTTAATTGCACTTTCGAGTACATATCTGATGCCTTTTCGCTTGATGTTTCAGAA-----A
SARSTor2        TTTAATTGCACTTTCGAGTACATATCTGATGCCTTTTCGCTTGATGTTTCAGAA-----A
BcoV            CG-CGTAGAACTATGGCATTGGGATACAGGTGTTGTTTCCTGTTTATATAAGCG-----T
```


HEV CG-CATAGAATTATGGCATTATGACACAGATGTTGTTTCTTGTTTATACAGGCG-----T
MHV TT-AATTGGCTTTTGGCACACAGAGCTAAAATCCCCTGTGTGCATTTTAAAGCG-----T
RtCoV CT-AATTGGTTTTTGGCACACAGATTTAAGGCCTCCGGTGTGCATTTTAAAGCG-----T
IBV ----------ATGTTGGTAACACCTCTTTTACTAGTGACTCTTTTGTGTGCACTATGT
CcoV TATTACAAGTTAAATAAAACCGCTGGCTTAAAAAGCTATGAATTGTGTGAAGATTATG-A
FcoV TATTACAAGTTAAATAACACCAATGGTCTAAAAACCTATGAATTATGTGAAGATTATG-A
TGE TATTATAGGGTTAATAATAAAAATGGT------ACTACCGTAGTTTCCAATTGCACTG-A
PRCoV GATTTATGGA-----GACAAGTTTCCT------ACTTCCGTAGTTTCCAATTGCACTG-A
HCoVOC43 TGCATATTGCTGGTTGTCAA-ACTACAAATGGGCTGAACACTAGTTACT--CTGTTTGCA
PEDV TACTACATGCTTAATGTTACTAGTGCAGGTGAGGATGGCATTTATTATGAACCCTGTACA
SARSUrba AGT----CAGGTAATTTTAAAC---ACTTACGAGAGTTTGTGTTTAAAAATAAAGATGGG
SARSTor2 AGT----CAGGTAATTTTAAAC---ACTTACGAGAGTTTGTGTTTAAAAATAAAGATGGG
BcoV AATTTCACATATGATGTGAATGCTGATTACTTGTATTTCCATTTTTATCAAGAAGGTGGT
HEV AATTTCACATATGATGTGAATGCTGATTATTTATATTTTCACTTTTATCAGGAAGGTGGC
MHV AATTTTACGTTTAATGTTAATGCCGAATGGCTTTATTTTCATTTTTACCAGCAGGGTGGT
RtCoV AATTTTACGTTTAATGTTAATGCCGAATGGCTTTATTTTCATTTTTACCAGCAGGGTGGT
IBV AGTGCTGTTTTGTATGACAGTAGTTCTTACGTTTACTACTACCAAAGTGCCTTCAGACCA
CcoV AT--ACTGCACTGGCTATGCAACCAATGTGTTTGCTCCGACATCAGGTG TA ATACCT
FcoV AC--ATTGCACTGGCTATGCTACCAATGTATTTGCTCCGACATCAG TG TACATACCT
TGE TC--AATGTGCTAGTTATGTGGCTAATGTTTTTACTACACAGCCAG AG TT ATACC
PRCoV TC--AATGTGCTAGTTATGTGGCTAATGTTTTTACTATACTACCAG AG CTT ATACC
HCoVOC43 AC-GGCTGTGTTGGTTATTCAGAAAATGTATTTGCTGTTGAGAGTG TG TA ATACCC
PEDV GCTAATTGCACTGGTTACGCTGCCAATGTATTTGCCACTGATTCCAATG CCA ATACCA
SARSUrba TTTCTCTATG-----TTTATAAGGGCTAT--------CAACCTATA AT
SARSTor2 TTTCTCTATG-----TTTATAAGGGCTAT--------CAACCTATA AT
BcoV ACTTTTTATGCATATTTTACAGACACTGGTGTTGTTACTAAGTTTCTGTT A GTT AT
HEV ACTTTTTATGCATACTTTACAGATACTGGTTTTGTGACCAAGTTTCTGTT AG T AT
MHV ACTTTTTATGCGTATTATGCGGATGTTTCTTCTGCTACTACGTTTTTGTT AT AT
RtCoV ACTTTTTATGCGTATTATGCAGATGTTTCTTCTGCCACTACGTTTTTGTT AT
IBV CCTAGTGGTTGGCATTTACAAGGGGGTGCTTATGCGGTAGTTAACATTTC CGAATTT
LK251 5'-CTCTATG-----TTTATAAGGGCTAT--------CAACC-3' LK249 BEACON
>>
CcoV GA GGA CA T TAACAATTGGTTTATGCTTACAAACAGCTCCACTTTTGTTAGTGGC
FcoV GA GGA TA T TAACAATTGGTTCTTGCTTACAAATAGTTCCACTTTTGTTAGTGGC
TGE AGAT TA T TAATAATTGGTTCCTTCTAACTAATAGCTCCACGTTGGTTAGTGGT
PRCoV AGAT TA T TAATAATTGGTTCCTCCTAACTAATAGCTCCACGTTGGTTAATGGT
HCoVOC43 T CGA TGCA CAATAATTGGTTCCTTCTAACTAATACCTCATCTGTTGTAGATGGT
PEDV GAAGGT TA T TAATAATTGGTTTCTTTTATCCAATGACTCCACTTTGTTGCATGGT
SARSUrba TTAA ACT-TTGAAACCTATTTTTAAGTTGCCTCTT----------
SARSTor2 TTAA ACT-TTGAAACCTATTTTTAAGTTGCCTCTT----------
BcoV TAGG ACGGT C TTCA ATTATTATGTCCTGCCTTTGACTTGTTCT----------
HEV TAGG ACTGT C GTCA ATTATTATGTTATGCCATTGACTTGTAAT----------
MHV AT GGTGATGT T AACA AATATTTTGTGTTGCCTTATATGTGTACTCTCACTACAACA
RtCoV AT GGTGCTGT AACA AGTATTTTGTGTTGCCTTATATGTGTAGTCCCACTACCTCA
IBV AA AATGCAG C CTTCATCAGGGTGTACTGTTGGTATTATTCATGGTGGTCGTGTTGTT
3'-CTTTGGATAAAAATTCAACGGAGAA----------
<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<
CcoV AGATTTGTAACAAATCAACCGCTGCTAGT--TAATTGCTTGTGGCCAGTGC--CCAGTTT
FcoV AGGTTTGTAACAAATCAACCATTATTGAT--TAATTGCTTGTGGCCAGTGC--CCAGTTT
TGE AAATTAGTTACCAAACAGCCGTTATTAGT--TAATTGCTTATGGCCAGTCC--CTAGCTT
PRCoV AAATTAGTTACCAAACAGCCTCTATTAGT--TAATTGCTTATGGCCAGTCC--CTAGCTT
HCoVOC43 GTTGTGAGGAGTTTTCAGCCTTTGTTGCT--TAATTGCTTATGGTCTGTTT--CT--GGC
PEDV AAAGTGGTTTCCAACCAACCCTTGTTGGT--CAATTGTCTTTTGGCCATTC--CTAAGAT
SARSUrba GGTATTAACATTACAAATTTTAGAGCCATTCTTACAGCCTTTTCACC-TGC--TCAAGAC
SARSTor2 GGTATTAACATTACAAATTTTAGAGCCATTCTTACAGCCTTTTCACC-TGC--TCAAGAC
BcoV AGTGCTATGACTTTAGAATATTGGGTTACACCTCTCACTTCTAAACAATAT--TTACTAG

```
HEV          AGCGCTTTATCTTTAGAATACTGGGTTACACCTCTCACTACTAGACAATTT--CTTCTAG
MHV          GGTGTCTTTTCACCGCAGTATTGGGTTACACCTCTTGTCAAGCGCCAATAT--TTATTTA
RtCoV        GGTGTTTCCTCACCGCAGTATTGGGTTACACCACTTGTTAAGCGCCAATAT--TTATTTA
IBV          AATGCTTCTTCTATAGCTATGACGGCACCGTCATCAGGTATGGCTTGGTCTAGCAGTCAG
                                                                *
             CC-5' LK252                3'-GTAAGAATGTCGGA-5' S-RT
             <<<<<<<<<<<

CcoV         TGGCGTCGCAGCACAAGAATT--TTGTTTTGAAGGTGCTCAGTTTAGCCAATGTAACGGT
FcoV         TGGTGTAGCAGCACAAGAATT--TTGTTTTGAAGGTGCACAGTTTAGCCAATGTAATGGT
TGE          TGAAGAAGCAGCTTCTACATT--TTGTTTTGAGGGTGCTGGCTTTGATCAATGTAATGGT
PRCoV        TGAAGAAGTAGCTTCTACATT--TTGTTTTGAAGGTGCTGACTTTGATCAATGTAATGGT
HCoVOC43     TTGCGGTTTACTACTGGTTTTGTCTATTTTAATGGTACTGGGAGAGGTGATTGTAAAGGT
PEDV         TTATGGACTAG-GCCAATTTT-TCTCATTCAATCACACGATGGATGGCGTTTGTAATGGA
SARSUrba     ATTTGGGGCAC-GTCAGC--TGCAGCCTATTTTGTTGGCTATTTAAAGCCAACTACATTT
SARSTor2     ATTTGGGGCAC-GTCAGC--TGCAGCCTATTTTGTTGGCTATTTAAAGCCAACTACATTT
BcoV         CTTTCAATCAA-GATGGTGTTATTTTTAATGCTGTTGATTGTAAGAGTGATTTTATGAGT
HEV          CCTTTGACCAG-GATGGTGTTTTATACCATGCTGTTGATTGTGCTAGTGATTTTATGAGT
MHV          ATTTTAATCAA-AAGGGTATTATTACTAGTGCTGTTGATTGTGCTAGTAGTTATACCAGC
RtCoV        ATTTTAACCAA-AAGGGTATTATTACTAGCGCTGTTGATTGTGCTAGTAGTTATACCAGT
IBV          TTTTGTACTGCACACTGTAATTTTTCAGATACTACAGTGTTTGTTACACATTGTTATAAA
                                  *                                   *

CcoV         GTTTCTTTAAATAATACAGTAGATGTTATTAGATTTAACCTTAATTTCACTACAGATGTA
FcoV         GTGTCTTTAAATAACACAGTGGATGTTATTAGATTCAACCTTAATTTCACTGCAGATGTA
TGE          GCTGTTTTAAATAATACTGTAGACGTCATTAGGTTCAACCTTAATTTTACTACAAATGTA
PRCoV        GCTGTTTTAAATAACACTGTAGACGTCATTAGGTTTAACCTTAATTTTACTACAAATGTA
HCoVOC43     TTTTCCTCAGATGTTTTGTCTGATGTCATACGTTACAACCTCAATTTTGAA------GAA
PEDV         GCTGCTGTGGATCGTGCCCCAGAGGCTCTGAGGTTTAATATTAATGACACCTC---CGTC
SARSUrba     ATGCTCAAGTATGATGAAAATGGTACAATCACAGATGCTGTTGATTGT----------T
SARSTor2     ATGCTCAAGTATGATGAAAATGGTACAATCACAGATGCTGTTGATTGT----------T
BcoV         GAGATTAAGTGTAAAACACTATCTATAGCACCATCTACTGGTGTTTATGAATTAAACGGT
HEV          GAGATTATGTGTAAAACTTCTTCAATTACACCACCTACTGGTGTTTATGAACTAAACGGT
MHV          GAAATAAAGTGTAAGACCCAAAGTATGAATCCCAATACGGGAGTTTATGATTTATCCGGT
RtCoV        GAAATAAAGTGTAAGACTCAAAGTATGAATCCCAATACGGGAGTCTATGATTTATCCGGT
IBV          CATGGTGGGTGTCCTTTAACTGGCATGCTTCAACAGAATCTTATACGT---------GTT
                        *

CcoV         CAATCTGGCATGGGTGCTACAGTATTTTCACTGAATACAACAGGCGGTGTCATTCTTGAG
FcoV         CAATCTGGTATGGGTGCTACAGTATTTTCACTGAATACAACAGGTGGTGTCATTCTTGAA
TGE          CAATCAGGTAAGGGTGCCACAGTGTTTTCATTGAACACAACGGGTGGTGTCACTCTTGAA
PRCoV        CAATCAGGTAAGGGTGCTACAGTGTTTTCATTGAACACAACGGGTGGTGTCACTCTTGAA
HCoVOC43     AACCTTAGACGTGGAACCATTTTGTTT------AAAACATCTTATGGTGTTGTTGTGTTT
PEDV         ATTCTTGCTGAAGGCTCAATTGTACTT------CATACTGCTTTAGGAACAAATCTTTCT
SARSUrba     CTCAAAATCCA-CTTGCTGAACTCAAATGCT---CTGTTAAGAGCTT---TGAGATTGAC
SARSTor2     CTCAAAATCCA-CTTGCTGAACTCAAATGCT---CTGTTAAGAGCTT---TGAGATTGAC
BcoV         TACACTGTTCAGCCAATTGCAGATGTTTACCGACGTATACCTAATCTTCCCGATTGTAAT
HEV          TACACAGTTCAACCTGTTGCCACTGTATATCGTAGAATACCTGATTTACCCAATTGCGAT
MHV          TACACCGTCCAACCTGTAGGATTAGTGTACCGGCGTGTTAGAAATTTGCCTGATTGTAAA
RtCoV        TACACCGTCCAACCTGTAGGACTAGTGTACCGGCGTGTTAGAAATTTGCCTGATTGTAAA
IBV          TCTGCTATGAAAAATGGCCAGCTTTTCTATAATTTAACAGTTAGTGTAGCTAAGTACCCT

CcoV         ATTTCTTGTTATAATGACACAGTGAGTGAGTCGAGT-TTCTACAGTTATGGTGAAATTCC
FcoV         ATTTCATGTTATAGTGACACAGTGAGTGAGTCTAGT-TCTTACAGTTATGGTGAAATCCC
TGE          ATTTCATGTTATA------CAGTGAGTGACTCGAGC-TTTTTCAGTTACGGTGAAATTCC
PRCoV        ATCTCATGTTATAATGATACAGTGAGTGATTCGAGC-TTTTCCAGTTACGGTGAAATTCC
HCoVOC43     TATTGTACCAACAACACTTTAGTTTC------AGGTGATGCTCACATACCATTTGGTACA
PEDV         TTTGTTTGCAGTAATTCCTCAGATCCTCATTTAGCCATCTTTGCCATACCTCTGGGTGCT
SARSUrba     AAAGGAATTTACCAGACCTCTAATTTCAGGGTTGTTCC-CTCAGGAGATGTTGTGAGATT
SARSTor2     AAAGGAATTTACCAGACCTCTAATTTCAGGGTTGTTCC-CTCAGGAGATGTTGTGAGATT
BcoV         ATAGAGGCTTGGCTTAATGATAAGTCGGTGCCCTCTCCATTAAATTGGGAACGTAAGACC
HEV          ATCGAAGCTTGGCTTAATTCTAAGACCGTTTCTTCGCCTCTTAATTGGGAACGTAAAATT
```

```
MHV          ATTGAGGAATGGCTAACTGCTAAGTCTGTACCTTCTCCTCTCAATTGGGAGCGCAAAACA
RtCoV        ATTGAGGAATGGTTGGCTGCTAACACAGTACCCTCTCCTCTCAATTGGGAGCGCAAAACA
IBV          ACTTTTAGATCATTTCAGTGTGTTAATAATTTAACATCCGTATATTTAAATGGTGATCTT

CcoV         ATTCGGCGTAACTGATG-GACCACGTTACTGTTATGTACTCT-ACAATGGCACAGCTCTT
FcoV         GTTCGGCATAACTGACG-GACCACGATACTGTTATGTACTTT-ACAATGGCACAGCTCTT
TGE          GTTCGGCGTAACTGATG-GACCACGGTACTGTTACGTACACT-ATAATGGCACAGCTCTT
PRCoV        GTTCGGCGTAACTAATG-GACCACGGTACTGTTACGTACTCT-ATAATGGCACAGCTCTT
HCoVOC43     GTTTTGGGCAATTTTT---ATTGCTTTGTAAATACTACTATTGGCAATGAAACTACGTCT
PEDV         ACTGAAGTACCCTACT---ATTGCTTTCTTAAAGTGGATACTTACAACTCCACTGTTTAT
SARSUrba     CCCTAATATTACAA-----ACTTGTGTCCTTTTGGAGAGGTTTTTAATGCTACTAAATTC
SARSTor2     CCCTAATATTACAA-----ACTTGTGTCCTTTTGGAGAGGTTTTTAATGCTACTAAATTC
BcoV         TTTTCAAATTGTAATTTTAATATGAGCAGCCTGATGTCTTTTATTCAGGCAGACTCATTT
HEV          TTTTCTAATTGTAATTTTAACATGGGCAGGCTGATGTCTTTTATTCAGGCTGACTCTTTT
MHV          TTTCAAAATTGTAACTTCGACCTGAGCAGTCTATTAAGATTTGTTCAGGCTGAGTCACTC
RtCoV        TTTCAAAATTGTAACTTCAACCTGAGCAGTCTATTAAGATTTGTTCAGGCTGAGTCACTC
IBV          GTTTACACCTCTAATGAGACCATAGATGTTACATCTGCAGGTGTTTATTTTAAAGCTGGT
                                                      *     *

CcoV         AAGTATT-TAGGAACATTACCACCTAGTGTCAAGGAAATTGCTATTAGTAAGTGG-----
FcoV         AAATATT-TAGGAACATTACCACCCAGTGTAAAGGAAATTGCTATTAGTAAGTGG-----
TGE          AAGTATT-TAGGAACATTACCACCTAGTGTCAAGGAGATTGCTATTAGTAAGTGG-----
PRCoV        AAGTATC-TAGGAACATTACCACCTAGTGTCAAGGAGATTGCTATTAGTAAGTGG-----
HCoVOC43     GCTTTTG-TGGGTGCACTACCTAAGACAGTTCGTGAGTTTGTTATT-TCACGCACA----
PEDV         AAATTCT-TGGCTGTTTTACCTCCTACTGTCAGGGAAATTGTCATCACCAAGTAT-----
SARSUrba     CCTTCTG------TCTATGCATGGGAGAGAAAAAAAAT---TTCTAATTGTGTTGCT---
SARSTor2     CCTTCTG------TCTATGCATGGGAGAGAAAAAAAAT---TTCTAATTGTGTTGCT---
BcoV         ACTTGTAATAATATTGATGCTGCTAAGATATATGGTATGTGTTTTTCCAGCATAACTATA
HEV          GGTTGTAACAATATTGATGCTTCTCGCTTATATGGTATGTGTTTTGGTAGCATTACTATT
MHV          TCATGTAGTAATATAGATGCTTCCAAGGTTTATGGTATGTGCTTTGGTAGTATATCTATA
RtCoV        TCATGTAGTAATATAGATGCTTCCAAGGTTTATGGAATGTGCTTTGGTAGCATATCTATA
IBV          GGACCTA-TAACTTATAAAGTTATGAGAGAAGTTAAA---GCCCTGGCTTATTTT-----
                                                        *

CcoV         GGACATTTTTATATTAATGGTTACAATTTCTTTAGCACGTTTCCTATTGATTGTATAGCT
FcoV         GGCCATTTTTATATTAATGGTTACAATTTCTTTAGCACATTTCCTATTGGTTGTATATCT
TGE          GGCCATTTTTATATTAATGGTTACAATTTCTTTAGCACATTTCCTATTGATTGTATATCT
PRCoV        GGCCATTTTTATATTAATGGTTACAATTTCTTTAGCACATTTCCTATTGATTGTATATCT
HCoVOC43     GGACATTTTTATATTAATGGCTATCGCTATTTCACTTTAGGTAATGTAGAAGCCGT----
PEDV         GGTGATGTTTATGTCAATGGGTTTGGCTATTTGCATCTCGGTTTGTTGGATGCTGTCACA
SARSUrba     GATTACTCTGTGCTCTACAAC--TCAACATTTTTTTCAACCTTTAAGTGCTATGGCGTTT
SARSTor2     GATTACTCTGTGCTCTACAAC--TCAACATTTTTTTCAACCTTTAAGTGCTATGGCGTTT
BcoV         GATAAGTTTGCTATACCCAATGGTAGGAAGGTTGACCTACAATTGGGCAATTTGGGCTAT
HEV          GATAAGTTTGCTATACCCAATAGTAGAAAGGTTGATCTGCAAGTGGGTAAATCTGGTTAT
MHV          GACAAGTTTGCGATACCCAATAGACGCCGAGTTGATTTGCAGCTAGGCAACTCTGGGTTT
RtCoV        GATAAATTTGCAATACCCAACAGTCGCCGTGTTGATCTTCAGCTAGGTAAATCGGGTCTT
IBV          GTTAATGGTACTGCACAAGATGTTATTTTGTGTGATGGATCACCTAGAGGCTTGTTAGCA
             *   *   *

CcoV         TTTAATTTAA----------CCACTGGTGCTAGTGGAGCAT-TTTGGACAATTGCTTA
FcoV         TTTAATTTAA----------CCACTGGTGTTAGTGGAGCTT-TTTGGACAATTGCTTA
TGE          TTTAATTTGA----------CCACTGGTGATAGTGACGTTT-TCTGGACAATAGCTTA
PRCoV        TTTAATTTGA----------CTACTGGTGATAGTGACGTCT-TCTGGACAATAGCTTA
HCoVOC43     --TAATTTCA------ATGTCACTACTGCAGAAACCACTGATT-TTTGTACTGTTGCGTT
PEDV         ATTAATTTCACTGGTCATGGCACTGACGATGACGTTTCAGGTT-TCTGGACCATAGCATC
SARSUrba     CTGC---CAC----------TAAGTTGAATGATCTTTGCTT-CTCCAATGTCTATGCA
SARSTor2     CTGC---CAC----------TAAGTTGAATGATCTTTGCTT-CTCCAATGTCTATGCA
BcoV         TTGCAGTCTT----------TTAACTATAGAATTGATACTA-CTGCTACAAGT-TGTC
HEV          TTACAATCTT----------TTAATTATAAGATTGACACTG-CTGTTAGCAGT-TGTC
MHV          TTGCAATCCT----------TTAATTACAAAATAGATACAA-GAGCTACTTCG-TGTC
RtCoV        TTGCAATCTT----------TTAATTATAAAATTGATACAA-GAGCGACCTCG-TGTC
IBV          TGCCAGTATA----------ATACTGGCAATTTTTCAGATGGCTTTTATCCTTTTACT
```

```
                                                                    *
CcoV            TACG-TCGTACACAGAAGCATTAGTACAA-GTTGAAAACACAGCTATTAAAAAGGTGACG
FcoV            CACA-TCGTATACTGAAGCATTAGTACAA-GTTGAAAACACAGCTATTAAAAATGTGACG
TGE             CACA-TCGTACACTGAAGCATTAGTACAA-GTTGAAAACACAGCTATTACAAAGGTGACG
PRCoV           CACA-TCGTACACTGAAGCATTAGTACAA-GTTGAAAACACAGCTATTACAAATGTGACG
HCoVOC43        AGCT-TCTTATGCTGACGTTTTGGTTAAT-GTGTCACAAACCTCTATTGCTAATATAATT
PEDV            GACT-AATTTTGTTGATGCACTCATCGAG-GTTCAAGGAACTTCCATTCAGCGTATTCTT
SARSUrba        GATTCTTTTGTAGTCAAG--GGAGATGAT-GTAAG-ACAAATAGCGCCAGGACAAACTGG
SARSTor2        GATTCTTTTGTAGTCAAG--GGAGATGAT-GTAAG-ACAAATAGCGCCAGGACAAACTGG
BcoV            AGTTGTATTATAATTTACCTGCTGCTAAT-GTTTCTGTTAGCAGGTTTAATCCTTCTACT
HEV             AACTCTATTATAGTTTGCCTGCAGCAAAC-GTATCTGTCACTCATTATAATCCTTCATCT
MHV             AGCTCTATTATAGTCTTGCAAAAAATAAT-GTCACTGTCAATAACCATAACCCGTCCTCT
RtCoV           AGCTCTATTACAGTCTTGCACAAGATAAT-GTCACTGTCATTAACCACAACCCATCCTCC
IBV             AATAGTAGTTTAGTTAAGCAGAAGTTTATTGTCTATCGTGAAAATAGTGTTAATACTACT
                         *                    *  **

CcoV            TATT-GTAACAGTCAC--ATTAATAACATCAAATGTTCTCAACTTACTG----CTAATTT
FcoV            TATT-GTAACAGTCAC--ATTAATAACATTAAATGTTCTCAACTTACTG----CTAATTT
TGE             TATT-GTAATAGTCAC--GTTAATAACATTAAATGCTCTCAAATTACTG----CTAATTT
PRCoV           TATT-GTAATAGTTAT--GTTAATAACATTAAATGCTCTCAACTTACTG----CTAATTT
HCoVOC43        TATT-GCAACTCTGTT--ATTAACAGACTGAGATGTGACCAGTTGTCCT----TTGATGT
PEDV            TATT-GTGATGATCCT--GTTAGCCAACTCAAGTGTTCTCAGGTTGCTT----TTGACCT
SARSUrba        TGTT-ATTGCTGATTATAATT------ATAAATTGCCAGATGATTTCAT------GG---
SARSTor2        TGTT-ATTGCTGATTATAATT------ATAAATTGCCAGATGATTTCAT------GG---
BcoV            TGGA-ATAGGAGATTTGGTTTTACAGAACAATTTGTTTTTAAGCCTCAACCTGTAGGTGT
HEV             TGGA-ATAGAAGGTATGGGTTT----AATAATCAGAGTTTTGGTTCCAG-----AGG---
MHV             TGGA-ACAGGCGTTATGGGTTC----AATGATGTGGCTACATTTGGAAC-----TGG---
RtCoV           TGGA-ATAGGCGTTATGGATTT----AATGACGTGGCTACATTTCATAG-----TGG---
IBV             TGTACGTTACACAATTTCATTTTT--CATAATGAGACTGGCGCCAACCC-----------
                *                   **          *   *

CcoV            GCAAAATGGTTTTTACCCTGTTG---CTTCAAGTGAAGTTGGTCTTGTCAATAAGAGTGT
FcoV            GAATAATGGATTTTATCCTGTTG---CTTCAAGTGAAGTAGGTTTCGTTAATAAGAGTGT
TGE             GAATAATGGATTTTATCCTGTTT---CTTCAAGTGAAGTTGGTCTTGTCAATAAGAGTGT
PRCoV           GAATAATGGATTTTATCCTGTTT---CTTCAAGTGAAGTTGGTTCTGTCAATAAGAGTGT
HCoVOC43        ACCAGATGGTTTTTTATTCTACAA---GCCCTATTCAATCCGTTGAGCTACCTGTGTCTAT
PEDV            TGACGATGGTTTTTTACCCCATCT---CTTCTAGA-AACCTTCTGAGTCACGAACAGCCAA
SARSUrba        --------GTTGTGTCCTTGCTTGGAATACTAGGAACATTGATGCTACTTCAACTGGTAA
SARSTor2        --------GTTGTGTCCTTGCTTGGAATACTAGGAACATTGATGCTACTTCAACTGGTAA
BcoV            TTTTACTCATCATGATGTTGTTT---ATGCACAACATTGTTTTAAAGCTCCCAAAAATTT
HEV             ------CCTTCATGATGCTGTTT---ATTCACAGCAATGTTTTAATACACCTAACACATA
MHV             ------TAAACATGACGTTGCTT---ATGCTGAGGCTTGTTTTACCGTGGGAGCATCATA
RtCoV           ------TGAACATGACGTTGCTT---ATGCAGAGGCATGTTTCACTGTTGGAGCTTCATA
IBV             ------TAATCCTAGTGGTGTTC---AGAATATTCAAACTTACCAAACAAAAACAGCTCA
                            *

CcoV            TGTGTTACTACCTAGTTTCTATTC---ACATACCAGTGTTAATATAACTATTGATCTTGG
FcoV            TGTGTTATTACCTAGCTTTTTCAC---ATACACCGCTGTCAATATAACCATTGATCTTGG
TGE             TGTGTTACTACCTAGCTTTTACAC---ACATACCATTGTTAACATAACTATTGGTCTTGG
PRCoV           TGTGTTACTACCTAGCTTTCTGAC---ACATACCATTGTTAACATAACTATTGGTCTTGG
HCoVOC43        TGTGTCGCTACCTGTTTATCATAA---ACATACGTTTATTGTGTTGTACGTTGACTTCAA
PEDV            TTTCTT-TTGTTACTTTGCCATCA---TTTAATGATCATTCTTTTGTTAAT--ATTACTG
SARSUrba        T-TATAATTATAAAT---ATAG--------------GTATCTTAGACATGGCAAGCTTAG
SARSTor2        T-TATAATTATAAAT---ATAG--------------GTATCTTAGACATGGCAAGCTTAG
BcoV            C-TGTCCGTGTAAATTGGATGG------GTCTTTGTGTGTAGGTAATGGTCCTGGTATAG
HEV             T-TGTCCTTGTAGA----ACAA------GTCA--ATGCATAGGTGGTG---CAGGCACAG
MHV             T-TGCCCTTGCGCGAACCCCAGCATAGTGTCGCCATGTACCACTGGAAAACCTAACTTTG
RtCoV           T-TGCCCTTGTGCGAAGCCCAGCACAGTCTATTCATGTGTCACAGGTAAACCTAAGTCTG
IBV             G-AGTGGTTATTATAATTTTAA------TTTTTCCTTTCTGAGTAGTTTTGTTTATAAGG
                          *

CcoV            TATGAAGCGTAGTG-TTACGGTCA--CCATAGCCTCACCATTAAGTAACATCACACTACC
```

```
FcoV        TATGAAGCTTAGTGGTTATGGTCAACCCATAGCCTCGACACTAAGTAACATCACACTACC
TGE         TATGAAGCGTAGTGGTTATGGTCAACCCATAGCCTCAACATTAAGTAACATCACACTACC
PRCoV       TATGAAGCGTAGTGGTTATGGTCAACCCATAGCCTCAACGCTAAGTAACATTACACTACC
HCoVOC43    ACCTCAGAGTGGCGGTGGCAAGTG--------------CTTTAACTGTTATCCTGCT---
PEDV        TCTCT------GCGGCTTTTGGTG--------------GTCTTAGTAGTGCCAATCT---
SARSUrba    GCCCTT-TGAGAGAGACATATCTAAT-----GTGCCTTTCT----------CCCCTGA--
SARSTor2    GCCCTT-TGAGAGAGACATATCTAAT-----GTGCCTTTCT----------CCCCTGA--
BcoV        ATGCTGGTTATAAAAATAGTGGTATAGGCACTTGTCCTGCA--GGTACTAATTATTTAAC
HEV         GAACTTGTCCTGTAGGCACCACTGTGCGCAAGTGTTTTGCT--G----CAGTTACAAAA-
MHV         CCAATTGCCCTACAGGCACCTCGAATCGTGAGTGCACTGTTATGCCATTGGCTAATAAT-
RtCoV       CTAATTGCCCAACAGGTACCTCGAATCGTGAGTGTAATGTTCAGGCTTCAGGTTTTAA--
IBV         AGTCTAATTTTATGTATGGATCTTAT-----------CACCCAAGTTGTAAATTTAGAC-

CcoV        AATGCAGGATAATAACATAGACGTGTACTGTATTCGTTCTAACCAATTCTCAGTTTATGT
FcoV        AATGCAGGATAACAATACTGATGTGTACTGTATTCGTTCTAACCAATTCTCAGTTTATGT
TGE         AATGCAGGATCACAACACCGATGTGTACTGTATTCGTTCTGACCAATTTTCAGTTTATGT
PRCoV       AATGCAGGATAACAACAACGATGTGTACTGTGTTCGTTCTGACCAATTTTCAGTTTATGT
HCoVOC43    --GGTGTTAATATTACACTGGCCAATTTT--AATGAAACTAAAGGGCCTT---TGTGTGT
PEDV        --CGT-TGCATCTGACACT----ACTATC--AATGGGTTTA---GTTCTT---TCTGTGT
SARSUrba    ------------TGGCAAACCTTGCACCCCACCTGCTCTTA---ATTGT------TATTG
SARSTor2    ------------TGGCAAACCTTGCACCCCACCTGCTCTTA---ATTGT------TATTG
BcoV        TTGCCATAATGCTGCCCAATGTGATTGTTTGTGCACTCCCG---ACCCCATTACATCTAA
HEV         ----------GCTACTAAGTGTACTTGCTGGTGTCAACCAG---ATCCTTCCACATATAA
MHV         ----------CAATTTAAGTGTGATTGCACTTGTAACCCTA---GTCC-------TCTAA
RtCoV       ------------GTCTAAGTGCGATTGCACATGTAACCCTA---GTCC-------TCTAA
IBV         ------------TAGAAACTATTAATAATGGCTTGTGGTTTA--ATTC------ACTTTC
                                                                      *

CcoV        TCATTCCACTTGCAAAAGTTCTTTATGGGATAACAATTTTAATTCAGCATGTACCGACGT
FcoV        TCATTCCACTTGCAAAAGTTCTTTATGGGACAATATTTTTAATCAAGACTGCACGGATGT
TGE         TCATTCTACTTGCAAAAGTGCTTTATGGGACAATATTTTTAAGCGAAACTGCACGGACGT
PRCoV       TCATTCTACTTGCAAAAGTGTTTTATGGGACAATGTTTTTAAGCGAAACTGCACGGACGT
HCoVOC43    TGACACATC-----ACACTTCACTACCAAATACGTTGCT-------GTTTATGCCAATGT
PEDV        TGACACTAG-----ACAATTTACCATTACACTGTTTTAT-------AATGTTACAAACAG
SARSUrba    GCCATTAAATGATTATG---GTTTTTACAC------CACTACTGGCATTGGCTACCAACC
SARSTor2    GCCATTAAATGATTATG---GTTTTTACAC------CACTACTGGCATTGGCTACCAACC
BcoV        ATCTACAGGGCCTTACAAGTGCCCCCAAACTAAATACTTAGTTGGCATAGGTGAGCACTG
HEV         AGGTGTAAACGCCTGGACTTGTCCGCAATCTAAAGTTTCTATACAACCAGGTCAGCATTG
MHV         CCACCTATGATCTTAGA--TGTCTCCAAGCAAGAAGCATGCTTGGCGTAGGTGATCATTG
RtCoV       CCACCTATGATCCTAGA--TGTCTTCAAGCGCGGAGCATGCTTGGCGTAGGTGATCATTG
IBV         AGTTTCAATTGCTTACGGTCCTCTTCAAGGTGGTTGCAA----GCAATCTGTCTTTAAAG
                                                                    *

CcoV        TT-TAGACGCCACAGCTGTTATAAAAACTGGTACTTGTC---CTTTCTCATTTGATAAAT
FcoV        TT-TAGAGGCTACAGCTGTTATAAAAACTGGTACTTGTC---CTTTCTCATTTGATAAAT
TGE         TT-TAGATGCCACAGCTGTTATAAAAACTGGTACTTGTC---CTTTCTCATTTGATAAAT
PRCoV       TT-TAGATGCCACAGCTGTTATAAAAACTGGTACTTGTC---CTTTCTCATTTGATAAAT
HCoVOC43    TGGTAGGTGGAGTGCTAGTATTA-ACACGGGAAATTGCC---CTTTTTCTTTTGGCAAAG
PEDV        TTAT-GGTTATGTGTCTAAATCACAGGATAGTAATTGTC---CTTTCACCTTGCAATCTG
SARSUrba    TT-ACAG------AGTTGTAGTACTTTCTTTTGA--------ACTTTTAAATGCACCGGC
SARSTor2    TT-ACAG------AGTTGTAGTACTTTCTTTTGA--------ACTTTTAAATGCACCGGC
BcoV        TT-CGGGTCTTGCTATTAAAAGTGATTATTGTGGAGGTA---ATCCTTGTACTTGCCAAC
HEV         CC-CTGGTTTGGGTCTTGTGGAGGATGATTGCTCTGGCA---ACCCTTGCACTTGTAAAC
MHV         TG-AAGGTCTAGGAGTTTTAGAAGATAAATGTGGTGGCAGCAACACCTGCAATTGTTCTG
RtCoV       TG-AAGGTCTAGGTATTTTAGAAGATAAATGTGGTGGCAGCAACATATGCAATTGTTCGG
IBV         GT--AGAGCAACTTGTTGTTATGCTTATTCATATGGAGG---------TCCTTCGCTGTG

CcoV        TGAATAATTACTTAACTTTTAACAAGTTCTGTTTGTCGTTGAATCCCGTTGGTGCCAACT
FcoV        TGAACAATTACTTGACTTTTAACAAGTTCTGTTTGTCGTTGAGTCCTGTTGGTGCTAATT
TGE         TGAACAATTACTTAACTTTTAACAAGTTCTGTTTGTCGTTGAGTCCTGTTGGTGCTAATT
PRCoV       TGAACAATTACTTAACTTTTAACAAGTTCTGTTTGTCGTTGAGTCCCGTTGGTGCTAATT
```

```
HCoVOC43    TTAATAACTTTGTTAAATTTGGCAGTGTATGTTTTTCGCTAAAGGATATACCCGGTGGTT
PEDV        TTAATGATTACCTGTCTTTTAGCAAATTTTGTGTTTCAACCAGCCTTTTGGCTGGTGCTT
SARSUrba    CACGG---TTTGTGGACCAAAAT-TATCCACTGACCTTATTAAGAACCAGTGTGTCAATT
SARSTor2    CACGG---TTTGTGGACCAAAAT-TATCCACTGACCTTATTAAGAACCAGTGTGTCAATT
BcoV        CACAAGCATTTTTGGGTTGGTCTGTTGACTCTTGTTTACAAGGGGATAGGTGTAATATTT
HEV         CACAGGCTTTCATAGGCTGGAGTTCAGAAACTTGTTTGCAAAATGGTAGGTGTAATATTT
MHV         CTCATGCCTTTGTTGGCTGGGCTAAGGATAGTTGCTTGGCTAATGGCCGCTGTCACATTT
RtCoV       CTGATGCCTTTGTTGGCTGGGCTATGGACAGCTGTCTATCTAATGCCCGCTGCCATATTT
IBV         TAAAGGTGTTTATTCAGGTGAGTTAGATCATAATTTTGAATGTGGACTGTTA-GTTTATG
                      *   *

CcoV        GTAAGTTAGATGTTGCCGCCCGTACAAGAACCAATGAGCAGGTTTTTGGAAGTT---TAT
FcoV        GCAAGTTTGATGTTGCTGCACGTACAAGAACCAATGAGCAGGTTGTTAGAAGTC---TAT
TGE         GTAAGTTTGATGTAGCTGCCCGTACAAGAACCAATGAGCAGGTTGTTAGAAGTT---TGT
PRCoV       GTAAGTTTGATGTAGCTGCCCGTACAAGAACCAATGATCAGGTTGTTAGAAGTT---TGT
HCoVOC43    GCGCAATGCCTATAGTGGCTAATTGGGCTTATAGTAAGTACTATACTATAGGCTCATTGT
PEDV        GTACCATAGATCTTTTTGGTTACCCTGCGTTCGGTAGTGGTGTTAAGTTGACGTCCCTTT
SARSUrba    TTAATTTTAATGGACTCACTGGTACTGGTGTGTTAACTCCTTCTTCAAAGAGAT---TTC
SARSTor2    TTAATTTTAATGGACTCACTGGTACTGGTGTGTTAACTCCTTCTTCAAAGAGAT---TTC
BcoV        TTGCTAATTTTATTTTTCATGATGTTAATAGTGGTACTACTTGTTCTACTGATT---TAC
HEV         TTGCTAATTTTATTCTGAATGATGTTAATAGCGGTACAACCTGTTCTACTGATT---TAC
MHV         TTAGTAATTTGATGTTAAATGGCATTAATAGTGGTACTACATGTTCCATGGATT---TGC
RtCoV       TTAGTAATTTGATGTTAAATGGCATTAATAGTGGTACTACATGTTCCACGGATT---TTC
IBV         TTACTAAGAGCGGTGGCTCTCGTATACAAACAGCCACTGAACCGCCAGTTATAA---CTC

CcoV        ATGTAATATATGAAGAAGGAGACAACATAGTGGGTGTACCGTCTGATAATAGTGGTTT-G
FcoV        ATGTAATATATGAAGAAGGAGACAACATAGTGGGTGTACCGTCTGATAATAGCGGTCT-G
TGE         ATGTAATATATGAAGAAGGAGACAACATAGTGGGTGTACCGTCTGATAATAGTGGTGT-G
PRCoV       ATGTAATATATGAAGAAGGAGACAGCATAGTTGGTGTACCGTCTGACAATAGTGGTTT-G
HCoVOC43    ATGTTTCTTGGAGTGATGGTGATGGAATTACTGGCGTCCCACAACCTGTTGAGGGTGTTA
PEDV        ATTTTCAATTCACAAAAGGTGAGTTGATTACTGGCACGCCTAAACCACTTGAAGGTATCA
SARSUrba    AACCATTT--CAACAATTTGGCCG-TGATGTTTCTG----ATTTCACTGATTCCGTTCGA
SARSTor2    AACCATTT--CAACAATTTGGCCG-TGATGTTTCTG----ATTTCACTGATTCCGTTCGA
BcoV        AAAAATCAAACACAGACATAATTCTTGGTGTTTGTGTTAATTATGATCTTTATGGTATTA
HEV         AACAGGGTAATACTATTATTACTACTGATGTTTGTGTTAATTATGACCTATATGGCATTA
MHV         AATTGCCTAATACTGAAGTGGTCACTGGCGTCTGCGTCAAATATGACCTCTATGGCATTA
RtCoV       AATTGCCTAATACGGAAGTGGTCACTGGCGTTTGTGTCAAGTATGACCTCTACGGTAGTA
IBV         AAAACAATTATAATAATATTACTTTAAATACTTGTGTTGATTATAATATATATGGCAGAA
            *                                                       *

CcoV        CACGATTTGTCAGTGTTGCACTTAGACTCTTGT-ACAGATTACAATATATATGGTAGAAC
FcoV        CACGATTTGTCTGTGCTACACCTAGACTCCTGT-ACAGATTACAATATATATGGTAGAAC
TGE         CACGATTTGTCAGTGCTACACCTAGATTCCTGC-ACAGATTACAATATATATGGTAGAAC
PRCoV       CACGATTTGTCAGTGCTACACCTAGATTCGTGC-ACAGATTACAATATATATGGTAGAAC
HCoVOC43    GTTCCTTTATGAATGTTACAT-TGGACAAATGT-ACTAAATATAATATTTATGATGTATC
PEDV        CAGACGTTTCTTTTATGACTC-TGGATGTGTGT-ACCAAGTATACTATCTATGGCTTTAA
SARSUrba    GATCCTAAAACATCT-----G--AAAT-----------ATTA-GACATTTCACCTTGCTC
SARSTor2    GATCCTAAAACATCT-----G--AAAT-----------ATTA-GACATTTCACCTTGCGC
BcoV        CAGGCCAAGGTATTTTTGTTG--AGGTTAATGCGACTTATTATAATAGTTGGCAGAACCT
HEV         CAGGCCAGGGCATACTTATAG--AAGTTAATGCCACTTATTATAATAGTTGGCAGAATCT
MHV         CAGGCCAAGGTATTTTTAAGG--AGGTTAAGGCTGACTATTATCATAGTTGGCAAAACCT
RtCoV       CAGGCCAAGGTGTTTTTAAGG--AGGTTAAGGCTGATTATTACAATAGTTGGCAGAACCT
IBV         CTGGCCAAGGTTTTATT--------ACTAATGTGACCGACTCAGCTGTTAGTTATAATTA
                                                    * *

CcoV        TGGTGTTGGTATTATTAGAAAAACTAACAGCACACTACTTAGTGGCTTATATT-ACACAT
FcoV        TGGTGTTGGTATTATTAGACGAACTAACAGTACGCTACTTAGTGGCTTATATT-ACACAT
TGE         TGGTGTTGGTATTATTAGACAAACTAACAGGACGCTACTTAGTGGCTTATATT-ACACAT
PRCoV       TGGTGTTGGTATTATTAGACAAACTAACAGGACGATACTTAGTGGCTTATATT-ACACAT
HCoVOC43    TGGTGTGGGTGTTATTCGCGTTAGCAATGACACCTTTCTTAATGGAATTACGT-ACACAT
PEDV        AGGTGAGGGTATTATTACCCTTACAAATTCTAGCATTTTGGCAGGTGTTTATT-ATACAT
SARSUrba    TTTTGGGGGTGTAAGTG-TAATTACACCTGGAACAAATGCTTCATCTGAAGTT-GCTGTT
```

```
SARSTor2        TTTTGGGGGTGTAAGTG-TAATTACACCTGGAACAAATGCTTCATCTGAAGTT-GCTGTT
BcoV            TTTATATGATTCTAATGGTAATCTCTATGGTTTTAGAGACTACTTAACAAACA-GAACTT
HEV             TCTTTATGATTCTAGTGGTAATCTCTATGGCTTTAGAGATTATTTATCAAATA-GAACTT
MHV             CTTATATGATGTTAATGGCAACTTAATCGGATTTCGCGATTTTGTTGCTAATA-AGAGTT
RtCoV           CTTATATGATGTTAATGGTAACTTAAATGGTTTCCGTGACATTGTTACCAATA-AGACTT
IBV             TCTAGCAGACGCAGGT-------TTGGCTATTTTAGATACATCTGGTTCCATAGACATCT
                       *       *                                          *

CcoV            CACTATCAGGTGATTTGTTAGGTTTTAAAAATGTTAGTGATGGTGTTGTCTACTCTGTAA
FcoV            CACTATCAGGTGATTTGTTAGGCTTTAAAAATGTTAGTGATGGTGTCATTTATTCTGTGA
TGE             CACTATCAGGTGATTTGTTAGGTTTTAAAAATGTTAGTGATGGTGTCATCTACTCTGTAA
PRCoV           CACTATCTGGTGATTTGTTAGGTTTTACAAATGTTAGTGATGGTGTTATCTACTCTGTAA
HCoVOC43        CAACTTCAGGTAACCTTCTGGGTTTTAAAGATGTTACTAAGGGCACCATCTACTCTATCA
PEDV            CTGATTCTGGACAGTTGTTAGCCTTTAAGAATGTCACTAGTGGTGCTGTTTATTCTGTCA
SARSUrba        CTATATCAAGATGTTA--ACTGCACTGATGTTTCTACAGCAATTCATGCAGATCAACTCA
SARSTor2        CTATATCAAGATGTTA--ACTGCACTGATGTTTCTACAGCAATTCATGCAGATCAACTCA
BcoV            TTATGATTCGTAGTTGCTATAGCGGTCGTGTTTCAGCGGCCTTTCATGCTAACTCTTCCG
HEV             TTCTTATTCGTAGCTGCTATAGTGGAAGAGTTTCAGCAGTTTTTCATGCTAACTCATCTG
MHV             ATACTATTCGAAGTTGCTATAGTGGGCGGGTCTCGGCTGCATATCATCAAGATGCACCAG
RtCoV           ATTTATTAAGAAGTTGCTATAGTGGGCGCGTTTCGGCTGCATATCATCAAGATGCACCTG
IBV             TTGTTGTACAAGGTGAATATGGTCTTAATTATTATAAGGTTAACCCTTGCGAAGATGTCA
                                                                        *

CcoV            CGCCATGTGATGTAAGTGCACAAGCTGCTGT-TATTGATGGTGCCATAGTTGGAGCTATG
FcoV            CGCCATGTGATGTAAGCGCACAAGCGGCTGT-TATTGATGGTGCCATAGTTGGAGCTATG
TGE             CGCCATGTGATGTAAGCGCACAAGCAGCTGT-TATTGATGGTACCATAGTTGGGGCTATC
PRCoV           CGCCATGTGATGTTAGCGCACAAGCAGCTAT-TATTGATGGTACCATAGTTGGGGCTATC
HCoVOC43        CTCCTTGTAACCCACCAGATCAGCTTGTTGTTTATCAGCAA-GCTGTTGTTGGTGCTATG
PEDV            CGCCATGTTCTTTTTCAGAGCAGGCTGC-ATATGTTAATGATGATATAGTGGGTGTTATT
SARSUrba        CACCA-GCTT----------------GGCGCATATATTCTACTGGAAACAATG-TATT
SARSTor2        CACCA-GCTT----------------GGCGCATATATTCTACTGGAAACAATG-TATT
BcoV            AACCA-GCATTGCTATTTCGGAATATTAAATGCAATTACGTTTTTAATAATACTC-TTTC
HEV             AACCA-GCTTTGATGTTTCGTAATCTTAAATGCAGCCACGTTTTTAATAATACTC-TTTC
MHV             AACCA-GCGCTACTATATCGCAATTTAAAATGTGACTATGTCTTTAACAACAACA-TATC
RtCoV           AACCA-GCGCTACTATATCGCAATTTAAAATGTGATTATGTGTTTAATAACAACA-TATC
IBV             -ACCAGCAGTTTGTAGTTTCTGGTGGTAAATTAGTAGGTATTCTTACTTCACGTAATGAG
                  **                                                    *

CcoV            ACTTCCATTAATAGTGAACTGTTAGGT--CTAACTCATTGGACAACAACACCTAATTTTT
FcoV            ACTTCCATTAACAGTGAACTGTTAGGT--CTAACACATTGGACAACGACACCTAATTTTT
TGE             ACTTCCATTAACAGTGAACTGTTAGGT--CTAACACATTGGACAACAACACCTAATTTTT
PRCoV           ACTTCCATTAACAGTGAATTGTTAGGT--CTAACACATTGGACAACAACACCTAATTTTT
HCoVOC43        TTGTCTGAAAATT-----TTACTAGTT--ACGGC---TTTTCTAATGTTGTAGAACTGCC
PEDV            TCT------AGTT-----TGTCTAACT--CCACT---TTTAACAATACTAGGGAGTTGCC
SARSUrba        CC--------------------------AGACTCAAGCAGGCTGTCTTAT-AGGAGCTG
SARSTor2        CC--------------------------AGACTCAAGCAGGCTGTCTTAT-AGGAGCTG
BcoV            ACGACAGCTGCAACCTATTAACTATTTTGATAGTTATCTTGGTTGTGTTGT-CAATGCTG
HEV             AAGACAAATACAGCTTGTTAACTATTTTGATAGTTACCTTGGTTGTGTTGT-TAATGCTT
MHV             CCGTGAGGAGACACCACTTAACTATTTCGATAGTTATCTTGGTTGTGTTGT-TAATGCTG
RtCoV           CCGTGAGGAGACACCACTTAACTATTTTGATAGTTATTTGGGTTGTGTTAT-TAATGCTG
IBV             AC---------TGGTTCTCAGCTTCTT-GAGAACCAGTTTTACATCAAAATCACTAATGG

CcoV            ATTACTACTCCATATATAATTATACAAATGTGATGAATCGTGGCACGGCAATTGA---TA
FcoV            ATTACTACTCTATATATAATTACACAAGTGAGAGGACTCGTGGCACTGCAATTGACAGTA
TGE             ATTACTACTCTATATATAATTACACAAATGATAGGACTCGTGGCACTGCAATTGACAGTA
PRCoV           ATTACTACTCTATATATAATTACACAAATGATAAGACTCGTGGCACTCCAATTGGCAGTA
HCoVOC43        GAAATTTTTCTATGCGT-------CCAATGGCAC-----------TTATAATTGC-----
PEDV            TGGTTTCTTCTACCATT-------CTAATGACGG-----------CTCCAATTGT-----
SARSUrba        AGCAT------GTCGACACTT---CTTATGAGTGCGACATTCCTATTGGAGCTGGCATTT
SARSTor2        AGCAT------GTCGACACTT---CTTATGAGTGCGACATTCCTATTGGAGCTGGCATTT
BcoV            ATAATAGTACTTCTAGTGTTG---TTCAAACATGTGATCTCACAGTAGGTAGTGGTTACT
HEV             ATAATAATACAGCTAGTGCTG---TAAGTACTTGTGATTTAACCGTTGGTAGCGGCTATT
```

```
MHV         ACAACTCAACTGAAGAAGCTG---TTGACGCGTGTGATTTGCGTATGGGTAGTGGGCTTT
RtCoV       ATAACTCAACTGAGCAGTCTG---TTGACGCGTGTGATTTGCGTATGGGTAGTGGGCTTT
IBV         AACACGTCGTTTTAGACGTTC-TATTACTGAAAATG---------TTGCAAATTGCCCTT

CcoV        ATGATATTGATTGTGAACCTATCATAACATATTCTAATATAGGTGTTTGTAAAAATGGAG
FcoV        ACGATGTTGATTGTGAACCTGTCATAACCTATTCTAATATAGGTGTTTGTAAAAATGGTG
TGE         ATGATGTTGATTGTGAACCTGTCATAACCTATTCTAACATAGGTGTTTGTAAAAATGGTG
PRCoV       ATGACGTTGATTGTGAACCTGTCATAACCTATTCTAACATAGGTGTTTGTAAAAATGGTG
HCoVOC43    ---ACA--------GACGCTGTTTTAACTTATTCTAGTTTTGGCGTTTGTGCAGATGGTT
PEDV        ---ACA--------GAGCCTGTGTTGGTGTATAGTAACATAGGTGTTTGTAAATCTGGCA
SARSUrba    GTGCTAGTTACCATACAGTTTCTTTATTACGTAG---------TACTAGCCAAAAATCTA
SARSTor2    GTGCTAGTTACCATACAGTTTCTTTATTACGTAG---------TACTAGCCAAAAATCTA
BcoV        GTGTGGATTACTCTACAAAAAGACGAAGTCGTAGAGCGATTACCACTGGTTATCGGTTTA
HEV         GTGTTGATTATGTTACAGCACTTAGATCACGTAGATCTTTTACTACAGGTTATCGCTTTA
MHV         GTGTCAACTATTCAACGTCTCACCGCGCTCGCAGCTCTGTCAGCACGGGTTATAAATTAA
RtCoV       GTGTCAACTATTCAATCGCTCACCGTGCGCGCAGGTCTGTCAGTACGGGTTATAAATTAA
IBV         ATGTTAGTTATGGTAAGTTTTGTATAAAACCTGATGGCT----CAATTGCCACAATAGTA
                                                                    *

CcoV        CTTTGGTTTTTATT--AACGTCACACATTCTG-ATGGAGACGTTCAACCAATTAGCACCG
FcoV        CTTTGGTTTTTATT--AACGTCACACATTCTG-ACGGAGACGTGCAACCAATTAGCACTG
TGE         CTTTTGTTTTTATT--AACGTCACACATTCTG-ATGGAGACGTGCAACCAATTAGCACTG
PRCoV       CTTTGGTTTTTATT--AACGTCACACATTCTG-ATGGAGACGTGCAACCAATTAGCACTG
HCoVOC43    CTATAATTGCTGTTCAACCACGTAATGTTTCATATGATAGTGTTTCAGCTATCGTCACAG
PEDV        GTATTGGCTATGTTCCATCTCAGTATGGCC---AAGTCAAGATTGCACCCACGGTTACTG
SARSUrba    ------TTGTGGCTTATACTATGTCTTTAGGTGCTGATAGTTC--AATTGCTTACTCTAA
SARSTor2    ------TTGTGGCTTATACTATGTCTTTAGGTGCTGATAGTTC--AATTGCTTACTCTAA
BcoV        CTAATTTTGAGCCATTTACTGTTAATTCAGTAAATGATAGTTTAGAACCTGTAGGTGGTT
HEV         CTAATTTTGAACCATTTGCCGCTAATTTGGTAAATGATAGTATAGAACCTGTTGGTGGTT
MHV         CTACTTTTGAACCATTTACAGTCCGCATTGTCAATGATAGTGTTGAGTCTGTTGATGGGT
RtCoV       CTACTTTTGAACCATTTACAGTCAGCATTGTCAATGATAGTGTTGAGTCTGTTGGTGGAT
IBV         ------------CCAAAACAATTGGAACAGTTTGTGGCACCTTT--ATTTAATGTTACTG
                         *                   *

CcoV        ---GTAATGTCACGATACCCACAAATTTTACTATATCTGTGCAAGTCGAATATATTCAGG
FcoV        ---GTAATGTCACGATACCTACAAATTTTACTATATCTGTGCAAGTTGAATACATGCAGG
TGE         ---GTAATGTCACGATACCTACAAACTTTACCATATCCGTGCAAGTCGAATACATTCAGG
PRCoV       ---GTAACGTCACGATACCTACTAACTTTACTATATCCGTGCAAGTCGAATATATTCAGG
HCoVOC43    ---CTAATTTGTCTATACCTTCCAATTGGACCACTTCGGTCCAGGTTGAGTATTTACAAA
PEDV        ---GGAATATTAGTATTCCCACCAACTTTAGTATGAGTATTAGAACAGAATATTTACAGC
SARSUrba    TA-ACACCATTGCTATACCTACTAACTTTTCAATTAGCATTACTACAGAAGTAATGCCTG
SARSTor2    TA-ACACCATTGCTATACCTACTAACTTTTCAATTAGCATTACTACAGAAGTAATGCCTG
BcoV        TGTATGAAATTCAAATACCTTCAGAGTTTACTATAGGTAATATGGAGGAGTTTATTCAAA
HEV         TGTATGAAATACAGATACCTTCAGAGTTTACCATTGGTAATTTAGAAGAGTTCATTCAAA
MHV         TATATGAGCTGCAAATACCAACCAACTTTACTATAGCTAGCCATCAGGAGTTCGTTCAAA
RtCoV       TATATGAGATGCAAATACCTACTAATTTTACTATAGCTAGCCATCAGGAGTTCATTCAAA
IBV         ---AAAATGTGCTCATACCTAACAGTTTCAACTTAACTGTTACAGATGAGTACATACAAA
                  *    ** **       *                          **     * *

CcoV        TTTACACTACACCAGTTTCAATAGACTGTGCAAGATACGTTTGCAATGGTAACCCAAGAT
FcoV        TTTACACTACACCAGTATCAATAGATTGTGCAAGATACGTTTGTAATGGTAACCCTAGAT
TGE         TTTACACTACACCAGTGTCAATAGACTGTTCAAGATATGTTTGTAATGGTAACCCTAGGT
PRCoV       TTTACACTACACCAGTGTCAATAGACTGTTCAAGATATGTTTGTAATGGTAACCCTAGGT
HCoVOC43    TTACAAGTACACCTATCGTAGTTGATTGCTCCACTTATGTTTGCAATGGTAATGTGCGCT
PEDV        TTTACAACACGCCTGTTAGTGTTGATTGTGCTACATATGTTTGTAATGGTAACTCTCGTT
SARSUrba    TTTCTATGGCTAAAACCTCCGTAGATTGTAATATGTACATCTGCGGAGATTCTACTGAAT
SARSTor2    TTTCTATGGCTAAAACCTCCGTAGATTGTAATATGTACATCTGCGGAGATTCTACTGAAT
BcoV        CAAGCTCTCCTAAAGTTACTATTGATTGTTCTGCTTTTGTCTGTGGTGATTATGCAGCAT
HEV         CGAGATCCCCTAAGGTTACTATAGACTGTGCTACATTTGTTTGTGGTGACTATGCAGCAT
MHV         CGAGGTCTCCAAAGGTTACTATAGACTGTGCTGCATTTGTCTGTGGTGGCCACACAGCAT
RtCoV       CGAGGTCTCCGAAGGTTACTATAGATTGTGCTGCATTTGTCTGTGGTGATTATACAGCGT
IBV         CGCGTATGGATAAGGTCCAAATTAATTGCCTGCAGTATGTTTGTGGCAGTTCTCTGGATT
```

```
                                    *  * **       *   * **                 *
CcoV            GCAATAAGTTATTAACACAATACGTTTCTGCATGTCAAACTATTGAGCAAGCGCTTGCAA
FcoV            GTAACAAATTGTTAACACAATATGTGTCTGCATGTCAAACTATTGAACAAGCACTTGCAA
TGE             GTAACAAATTGTTAACACAATACGTTTCTGCATGTCAAACTATTGAGCAAGCACTTGCAA
PRCoV           GTAACAAACTGTTAACACAATACGTTTCTGCATGTCAAACTATTGAGCAAGCACTTGCAA
HCoVOC43        GTGTTGAATTGCTTAAGCAGTATACTTCTGCTTGTAAAACTATTGAAGACGCCTTAAGAA
PEDV            GTAAACAATTACTCACCCAGTACACTGCAGCATGTAAGACCATAGAGTCAGCATTACAAC
SARSUrba        GTGCTAATTTGCTTCTCCAATATGGTAGCTTTTGCACACAACTAAATCGTGCACTCTCAG
SARSTor2        GTGCTAATTTGCTTCTCCAATATGGTAGCTTTTGCACACAACTAAATCGTGCACTCTCAG
BcoV            GTAAATCACAGTTGGTTGAATATGGTAGCTTCTGTGACAATATTAATGCTATACTCACAG
HEV             GTAGACAACAGTTAGCTGAGTATGGTAGTTTTTGTGAGAACATTAATGCTATACTCACAG
MHV             GCCGTCAGCAGTTGGTTGAGTACGGCTCATTCTGTGATAATATTAATGCCATTCTTGGCG
RtCoV           GTAGACAACAGTTGGTTGATTATGGCTCTTTTTGTGATAATATTAATGCCATTCTTGGCG
IBV             GTAGAAAGTTGTTTCAACAATATGGGCCTGTTTGCGACAACATATTGTCTGTAGTAAATA
                *          *     * **          **        *          *

CcoV            TGGGTGCCAG---ACTTGAAAACATGGAGATTGATTCCATGTTATTTGTTTCGGAAAATG
FcoV            TGGGTGCCAG---ACTTGAAAACATGGAGGTTGATTCCATGTTGTTTGTCTCGGAAAATG
TGE             TGGGTGCCAG---ACTTGAAAACATGGAGGTTGATTCCATGTTGTTTGTTTCTGAAAATG
PRCoV           TGGGTGCCAG---ACTTGAAAACATGGAAGTTGATTCCATGTTATTTGTTTCTGAAAATG
HCoVOC43        ATAGCGCCAG---GCTGGAGTCTGCAGATGTTAGTGAGATGCTCACTTTTGACAAGAAAG
PEDV            TCAGCGCTAG---GCTTGAGTCTGTTGAAGTTAACTCTATGCTTACCATTTCTGAAGAGG
SARSUrba        GTATTG-----CTGCTGAACAGGATCGCAACACACGTGAAGTGTTCGCTCAAGTCAAACA
SARSTor2        GTATTG-----CTGCTGAACAGGATCGCAACACACGTGAAGTGTTCGCTCAAGTCAAACA
BcoV            AAGTAAATGAACTACTTGACACTACACAGTTGCAAGT-AGCTAATAGTTTAATGAATGGT
HEV             AAGTAAATGAACTACTTGACACTACACAGTTGCAAGT-AGCTAATAGTTTAATGAATGGA
MHV             AGGTAAATAACCTCATAGATACTATGCAACTTCAAGT-TGCAAGTGCTTTAATCCAAGGT
RtCoV           AGGTGAATAACCTCATAGATACTATGCAATTACAAGT-TGCTAGTGCTCTGATCCAAGGT
IBV             GTGTTGGTCA---AAAAGAAGATATGGAACTTTTGAATTTCTATTCTTCTACTAAACCGG
                           *

CcoV            CCCTTA-AATTGGCATCTGTTGAAGCATTCAATAGTACGGAAAATTTAGACCCTATTTAT
FcoV            CCCTTA-AATTGGCATCTGTTGAGGCGTTCAATAGTACAGAAAATTTAGATCCTATTTAC
TGE             CCCTTA-AATTGGCATCTGTTGAAGCATTCAATAGTTCAGAAACTTTAGACCCTATTTAC
PRCoV           CCCTTA-AATTGGCTTCTGTCGAAGCATTCAATAGTTCAGAAACTTTAGACCCTATTTAC
HCoVOC43        CGTTTACACTT-GCTAATGTTAGT------AGTTTT---GGTGACTACAACCTTAGC---
PEDV            C-TTTACAGTTAGCTACCATCAGTTCGTTTAATGGT---GATGGATATAACTTTACT---
SARSUrba        AATGTACAAAACCCCAACTTTGAAA-------------TATTTTGGTGGTTTTAATT--
SARSTor2        AATGTACAAAACCCCAACTTTGAAA-------------TATTTTGGTGGTTTTAATT--
BcoV            GTCACTCTTAGCACTAAGCTTAAAGATGGCGTTAATTTCAATGTAGACGACATCAATT--
HEV             GTCACCCTTAGTACCAAGATTAAGGATGGCATTAATTTCAATGTTGACGATATCAACT--
MHV             GTCACGTTAAGCTCACGCTTATCGGATGGCATTGGTGGTCAAATAGATGATATTAATT--
RtCoV           GTCACGCTAAGTTCACGCTTGGCAGATGGCATCTCAGGTCAGATTGATGATATTAATT--
IBV             CTGGTTTTAATACACCAGTTCTTAG-------TAATGTTAGCACTGGTGAGTTTAATA--
                                  *                                    *

CcoV            AAAGAATGGCCTAACATTGGTGGTTCTTGGCTAGGAGGTTTAAAAGATATATTGCCATCT
FcoV            AAAGAATGGCCTAGCATAGGTGGTTCTTGGCTAGGAGGTCTAAAAGATATACTACCGTCC
TGE             AAAGAATGGCCTAATATAGGTGGTTCTTGGCTAGAAGGTCTAAAATACATACTTCCGTCC
PRCoV           AAAGAATGGCCTAATATAGGTGGCTTTTGGCTAGAAGGTCTAAAATACATACTTCCGTCC
HCoVOC43        ----AGCGTCATA-----CCTAGCTTG--------------------------CCCACA
PEDV            ----AATGTGCTG-----GGTGCTTCCGTGTACGA-----------------TCCTGCA
SARSUrba        ----TTTCACAAATATTACCTGACCCT-------------------CTAAAGCCAA---
SARSTor2        ----TTTCACAAATATTACCTGACCCT-------------------CTAAAGCCAA---
BcoV            ----TTTCCCCTGTATTAGGTTGTTTAGGAAGCGC-----------TTGTAATAAAGTT
HEV             ----TCTCCCCTGTATTAGGTTGTTTAGGAAGCGA-----------ATGTAATAGAGCT
MHV             ----TTAGTCCTCTGCTTGGTTGTTTAGGTTCTGA-----------CTGTGGCGAAGTT
RtCoV           ----TTAGTCCTCTTCTAGGTTGCCTTGGCTCAGA-----------TTGTAGCGAAGGC
IBV             ----TTTCTCTTCTGTTAACAAATCCT--------------------------AGTAGT

CcoV            CATAATAGCAAACGTAAGTACCGCTCGGCTATAGAAGACTTGCTTTTTGATAAGGTTGTA
```

```
FcoV        CATAATAGCAAACGTAAGTATGGTTCTGCTATAGAAGATTTGCTTTTTGATAAAGTTGTA
TGE         CATAATAGCAAACGTAAGTATCGTTCAGCTATAGAGGACTTGCTTTTTGATAAGGTTGTA
PRCoV       GATAATAGCAAACGTAAGTATCGTTCAGCTATAGAGGACTTGCTTTTTGATAAGGTTGTA
HCoVOC43    AGTGGTAGTAGAGTGGCTGGTCGCAGTGCCATAGAAGACATACTTTTTAGCAAACTTGTT
PEDV        AGTGGCAGGGTGGTACAAAAAAGGTCTGTTATTGAAGACTTGCTTTTTAATAAAGTGGTT
SARSUrba    -CTAAGA---------------GGTCTTTTATTGAGGACTTGCTCTTTAATAAGGTGACA
SARSTor2    -CTAAGA---------------GGTCTTTTATTGAGGACTTGCTCTTTAATAAGGTGACA
BcoV        TCCAGCA-----------G----ATCTGCTATAGAGGATTTACTTTTTTCTAAAGTAAAG
HEV         TCCACTA-----------G----ATCTGCTATAGAGGATTTACTTTTTGATAAAGTAAAA
MHV         ACCATGGCAGCTCAAACCGGACGATCTGCTATAGAGGATGTATTATTTGACAAAGTCAAA
RtCoV       ACCAAGGCAGCGCAA---GGGCGATCTGCTATAGAGGATGTATTATTTGATAAGGTCAAA
IBV         CGTAGAA------------AGCGTTCTCTTATTGAAGACCTTCTATTTACAAGCGTTGAA
                                       ** ** **  *  * ***   *   *

CcoV        ACATCTGGCTTAGGTACAGTTGACGAAGATTACAAACGTTCTGCAGGTGGTTATGACA--
FcoV        ACATCTGGTTTAGGTACAGTTGATGAAGATTATAAACGTTGTACTGGTGGTTACGACA--
TGE         ACATCTGGTTTAGGTACAGTTGATGAAGATTATAAACGTTGTACAGGTGGTTATGACA--
PRCoV       ACATCTGGTTTAGGTACAGTTGATGAAGATTACAAACGTTGTACAGGTGGTTATGACA--
HCoVOC43    ACTTCTGGACTTGGCACTGTGGACGCAGACTACAAAAAGTGCACTAAGGGTCTTTCCA--
PEDV        ACTAATGGCCTTGGTACTGTTGATGAAGACTATAAGCGCTGTTCTAATGGTCGCTCTG--
SARSUrba    CTCGCTGATGCTGGCTTCATGAA---GCAATATGGCGAATGCCTAGGTGATATTAATG--
SARSTor2    CTCGCTGATGCTGGCTTCATGAA---GCAATATGGCGAATGCCTAGGTGATATTAATG--
BcoV        TTATCTGATGTCGGTTTCGTTGA---GGCTTATAATAATTGTACTGGAGGTGCCGAAA--
HEV         TTGTCTGATGTCGGCTTTGTACA---GGCCTATAATAACTGCACTGGAGGTGCCGAAA--
MHV         CTCTCTGATGTTGGCTTTGTCGA---AGCATATAACAATTGCACTGGAGGCCAAGAAG--
RtCoV       CTCTCTGATGTTGGCTTTGTCGA---ATCATATAATAATTGCACTGGAGGTCAAGAAG--
IBV         TCTGTTGGACTACCAACAAATGA---CGCATATAAAAATTGCACTGCAGGACCTTTAGGC
                 **                *       **       *        *

CcoV        ----TAGCTGACTTAGTGTGTGCACGATATTACAATGGCATCATGGTGCTACCTGGTGTA
FcoV        ----TAGCAGACTTGGTGTGTGCTCAATATTACAATGGCATCATGGTTCTACCAGGTGTA
TGE         ----TAGCTGACTTAGTATGTGCTCAATACTATAATGGCATCATGGTGCTACCTGGTGTG
PRCoV       ----TAGCTGACTTAGTATGTGCTCAATACTATAATGGCATTATGGTGCTACCTGGTGTG
HCoVOC43    ----TTGCTGACTTGGCTTGTGCTCAATATTATAATGGCATTATGGTTTTGCCTGGCGTC
PEDV        ----TGGCTGATCTAGTCTGTGCGCAGTATTACTCTGGTGTCATGGTACTACCTGGCGTT
SARSUrba    ----CTAGAGATCTCATTTGTGCGCAGAAGTTCAATGGACTTACAGTGTTGCCACCTCTG
SARSTor2    ----CTAGAGATCTCATTTGTGCGCAGAAGTTCAATGGACTTACAGTGTTGCCACCTCTG
BcoV        ----TTAGGGACCTCATTTGTGTGCAAAGTTATAATGGTATCAAAGTGTTGCCTCCACTG
HEV         ----TTAGGGATCTCATTTGTGTGCAAAGTTATAATGGTATCAAAGTGTTGCCTCCATTG
MHV         ----TTAGAGACCTACTTTGTGTGCAATCTTTTAATGGCATCAAAGTGCTACCGCCTGTG
RtCoV       ----TTAGAGACCTACTTTGTGTGCAATCTTTTAATGGCATTAAAGTGCTACCGCCTGTA
IBV         TTTTTTAAGGACCTTGCGTGTGCTCGTGAATATAATGGTTTGCTTGTGTTGCCTCCTATC
                    **  *    ****  *     *    ***  *    **  * **     *

CcoV        GCTAATGATGACAAGATGACTATGTACACTGCATCTCTTACAGGTGGTATAACATTAGGT
FcoV        GCTAATGCTGACAAGATGACTATGTACACAGCATCACTTGCAGGTGGTATAACATTAGGT
TGE         GCTAATGCTGACAAAATGACTATGTACACAGCATCCCTTGCAGGTGGTATAACATTAGGT
PRCoV       GCTAATGCTGACAAAATGACTATGTACACAGCATCCCTCGCAGGTGGTATAACATTAGGT
HCoVOC43    GCTGATGCTGAACGAATGGCCATGTATACAGGTTCTTTAATTGGTGGAATTGCTTTAGGA
PEDV        GTTGACGCTGAGAAGCTTCACATGTACAGTGCGTCTCTCATAGGTGGTATGGCGCTAGGA
SARSUrba    CTCACTGATGATATGATTGCTGCCTACACTGCTGCTCTAGTTAGTGGTACTGCCACTGCT
SARSTor2    CTCACTGATGATATGATTGCTGCCTACACTGCTGCTCTAGTTAGTGGTACTGCCACTGCT
BcoV        CTCTCAGTAAATCAGATCAGTGGATACACTTTGGCTGCCACCTCTGCTAGTCTGTTTCCT
HEV         TTATCTGAAAATCAGATCAGTGGCTACACTTTGGCAGCCACCGCTGCTAGCTTATTCCCT
MHV         TTGTCTGAGAATCAAATTTCTGGTTATACAGCGGGAGCTACTGTATCTGCTATGTTCCC-
RtCoV       TTATCCGAGAGTCAAATCTCTGGTTATACAGCGGGAGCTACTGCATCTGCTATGTTCCCT
IBV         ATAACAGCAGAAATGCAAGCTTTGTATACTAGTTCTCTAGTAGCTTCTATGGCTTTTGGT
                  *                  ** *

CcoV        GCACTTAGTGGTGGCGCAGTGGC---------TATACCTTTTGCAGTAGCAGTTCAGGCT
FcoV        GCACTTGGTGGTGGCGCCGTGGC---------TATACCTTTTGCAGTAGCAGTACAGGCT
TGE         GCACTTGGTGGAGGCGCCGTGGC---------TATACCTTTTGCAGTAGCAGTTCAGGCT
PRCoV       GCACTTGGTGGAGGCGCCGTGGC---------TATACCTTTTGCAGTAGCAGTTCAGGCT
```

```
HCoVOC43    GGTCTAACATCAG---CCGTTTC---------AATACCATTTTCATTAGCAATTCAGGCA
PEDV        GGTATAACTGCTG---CAGCGGC---------ATTGCCTTTTAGCTATGCTGTTCAAGCG
SARSUrba    GGATGGACATTTGGTGCTGGCGCTGCTCTTCAAATACCTTTTGCTATGCAAATGGCATAT
SARSTor2    GGATGGACATTTGGTGCTGGCGCTGCTCTTCAAATACCTTTTGCTATGCAAATGGCATAT
BcoV        CCTTTGTCAGCAGCAGTAGGTG------------TACCATTTTATTTAAATGTTCAGTAT
HEV         CCTTGGACAGCTGCAGCAGGTG------------TACCATTTTATTTAAATGTTCAGTAT
MHV         --ATGGTCTGCAGCTGCAGGTG------------TGCCATTTTCTTTAAGTGTTCAATAT
RtCoV       CCATGGTCTGCAGCTGCAGGTG------------TGCCATTTGCTTTAAGTGTTCAATAT
IBV         GGTATTACTGCAG---CTGGTGC---------TATACCTTTTGCCACACAACTGCAGGCT
                        *     *                * ** ***          *

CcoV        AGACTTAATTATGTTGCTCTACAAACTGATGTATTGAACAAAAACCAACAAATCTTGGCT
FcoV        AGACTTAATTATGTTGCTCTACAAACTGATGTATTGAATAAAAACCAACAGATCCTGGCT
TGE         AGACTTAATTATGTTGCTCTACAAACTGATGTATTGAACAAAAACCAGCAGATTCTGGCT
PRCoV       AGACTTAATTATGTTGCTCTACAAACTGATGTATTGAACAAAAACCAGCAGATCCTGGCT
HCoVOC43    CGTTTAAATTATGTTGCATTGCAGACTGATGTTTTACAAGAAAATCAGAAAATTCTTGCT
PEDV        AGACTCAATTATCTTGCTTTACAGACGGATGTTCTACAGCGGAACCAGCAATTGCTTGCT
SARSUrba    AGGTTCAATGGCATTGGAGTTACCCAAAATGTTCTCTATGAGAACCAAAAACAAATCGCC
SARSTor2    AGGTTCAATGGCATTGGAGTTACCCAAAATGTTCTCTATGAGAACCAAAAACAAATCGCC
BcoV        CGTATTAATGGGATTGGTGTTACCATGGATGTGTTAAGTCAAAATCAAAAGCTTATTGCT
HEV         CGTATAAATGGGCTTGGCGTCACTATGGATGTGCTAAGTCAAAACCAAAAGCTTATTGCT
MHV         AGAATTAATGGTCTTGGTGTCACTATGAATGTTCTTAGTGAAAATCAGAAAATGATAGCA
RtCoV       AGAATTAATGGTCTTGGTGTCACTATGAATGTTCTTAGTGAAAACCAGAAAATGATAGCT
IBV         AGAATTAATCACTTGGGTATTACCCAGTCACTTTGTTGAAGAATCAAGAAAAAATTGCT
             *  * ***    * *   *           *  *       ** **  *     * **

CcoV        AATGCTTTCAATCAAGCTATTGGTAACATTACACAGGCATTTGGTAAGGTTAATGACGCT
FcoV        AATGCTTTCAATCAAGCTATTGGTAACATTACACAGGCTTTTGGTAAGGTTAATGATGCT
TGE         AGTGCTTTCAATCAAGCTATTGGTAACATTACACAGTCATTTGGTAAGGTTAATGATGCT
PRCoV       AGTGCTTTTAATCAAGCTATTGGTAACATTACACAGTCATTTGGTAAGGTTAATGATGCT
HCoVOC43    GCATCTTTTAACAAAGCAATGACCAACATAGTAGATGCCTTTACTGGTGTTAATGATGCT
PEDV        GAGTCTTTTAACTCTGCTATTGGTAATATAACTTCAGCCTTTGAGAGTGTTAAAGAGGCT
SARSUrba    AACCAATTTAACAAGGCGATTAGTCAAAT-------------------------------
SARSTor2    AACCAATTTAACAAGGCGATTAGTCAAAT-------------------------------
BcoV        AATGCATTTAACAATGCTCTTGATGCTAT-------------------------------
HEV         AGTGCATTTAACAATGCTCTTGATGCTAT-------------------------------
MHV         AGTGCATTTAACAACGCGATAGGTGCTAT-------------------------------
RtCoV       AGTTCATTTAACAACGCGATAGGTGCTAT-------------------------------
IBV         GCTTCCTTTAATAAGGCCATTGGTCATAT-------------------------------
                  ** **    **  *       **

CcoV        ATACATCAAACATCAAAAGGTCTTGCTACTGTTGCTAAAGCATTGGCAAAGGTGCAAGAT
FcoV        ATACATCAAACATCACAAGGTCTTGCCACTGTTGCTAAAGCGTTGGCAAAAGTGCAAGAT
TGE         ATACATCAAACATCACGAGGTCTTGCTACTGTTGCTAAAGCATTGGCAAAAGTGCAAGAT
PRCoV       ATACATCAAACATCACGAGGTCTTACAACTGTTGCTAAAGCATTGGCAAAAGTGCAAGAT
HCoVOC43    ATTACACAAACTTCACAAGCCCTACAAACAGTTGCTACTGCACTTAACAAGATCCAGGAT
PEDV        ATTAGTCAAACTTCCAAGGGTTTGAACACTGTGGCTCATGCGCTTACTAAGGTTCAAGAG
SARSUrba    -----TCAAG------AATCACTTACAACAACATCAACTGCATTGGGCAAGCTGCAAGAC
SARSTor2    -----TCAAG------AATCACTTACAACAACATCAACTGCATTGGGCAAGCTGCAAGAC
BcoV        -----TCAGG------AAGGGTTTGATGCTACCAATTCTGCTTTAGTTAAAATTCAAGCT
HEV         -----CCAGG------AAGGGTTCGACGCAACCAATTCTGCTTTAGTTAAAATTCAGGCT
MHV         -----ACAGG------AAGGGTTTGCTGCAACCAATTCTGCCTTAGCAAAAATGCAGTTC
RtCoV       -----ACAGG------AAGGGTTCGATGCAACCAATTCTGCTTTAGCGAAAATTCAGTCC
IBV         -----GCAGG------AAGGTTTTAGAAGTACATCTCTAGCATTACAACAAATTCAAGAT
                 **             *                ** *     *  * **

CcoV        GTTGTTAACACGCAAGGTCAAGCTTTAAGCCACCTAACAGTACAATTGCAAAACAATTTT
FcoV        GTTGTCAACACACAAGGGCAAGCTTTAAGTCACCTTACAGTACAATTGCAAAATAATTTT
TGE         GTTGTCAACATACAAGGGCAAGCTTTAAGCCACCTAACAGTACAATTGCAAAATAATTTC
PRCoV       GTTGTCAACACACAAGGTCAAGCTTTAAGACACCTAACAGTACAATTGCAAAATAATTTC
HCoVOC43    GTTGTTAATCAACAAGGCAACTCATTGAACCATTTAACTTCTCAGTTGAGGCAGAATTTT
PEDV        GTTGTTAATTCGCAGGGTTCAGCTTTGAACCAACTTACCGTACAGCTGCAACACAACTTC
SARSUrba    GTTGTTAACCAGAATGCTCAAGCATTAAACACACTTGTTAAACAACTTAGCTCTAATTTT
```

```
SARSTor2        GTTGTTAACCAGAATGCTCAAGCATTAAACACACTTGTTAAACAACTTAGCTCTAATTTT
BcoV            GTTGTTAATGCAAATGCTGAAGCTCTTAATAACTTATTGCAACAACTCTCTAATAGATTT
HEV             GTTGTTAATGCAAATGCTGAAGCACTTAATAACTTATTGCAGCAACTCTCTAACAGATTT
MHV             GTTGTCAATGCAAATGCGGAAGCACTCAATAATTTATTAAACCAGCTTTCCAATAGGTTT
RtCoV           GTTGTCAACGCAAATGCAGAAGCACTCAATAACCTTTTGAATCAGCTTTCCAATAGGTTT
IBV             GTTGTTAGTAAACAGAGTGCTATTCTTACTGAGACTATGGCATCACTTAATAAAAATTTT
                ***** *      *           * *                 *       *  **

CcoV            CAAGCCATTAGCAGTTCTATTAGTGACATTTATAACAGGCTTGATGAATTGAGTGCTGAT
FcoV            CAAGCCATTAGTAGTTCTATTAGTGATATTTATAACAGGCTTGACGAACTGAGTGCTGAT
TGE             CAAGCCATTAGTAGTTCTATTAGTGACATTTATAATAGGCTTGACGAATTGAGTGCTGAT
PRCoV           CAAGCCATTAGTAGTTCTATTAGTGACATTTATAATAGGCTTGATGAATTGAGTGCTGAT
HCoVOC43        CAAGCTATCTCTAGCTCTATTCAGGCTATCTATGACAGACTTGACACTATTCAGGCTGAT
PEDV            CAAGCCATTTCTAGTTCTATTGATGACATTTATTCCCGACTGGACATTCTTTCAGCCGAT
SARSUrba        GGTGCAATTTCAAGTGTGCTAAATGATATCCTTTCGCGACTTGATAAAGTCGAGGCGGAG
SARSTor2        GGTGCAATTTCAAGTGTGCTAAATGATATCCTTTCGCGACTTGATAAAGTCGAGGCGGAG
BcoV            GGTGCTATAAGTTCTTCTTTACAAGAAATTCTATCTAGACTGGATGCTCTTGAAGCGCAA
HEV             GGTGCCATAAGTGCCTCTTTACAAGAAATTTTATCCAGGCTCGATGCTCTTGAAGCTAAA
MHV             GGTGCAATTAGTGCTTCTTTACAAGAAATTCTATCTCGCCTAGATGCTCTTGAAGCGCAG
RtCoV           GGTGCAATTAGTGCTTCTTTACAGGAAATTCTATCTCGCCTCGATGCTCTTGAAGCTCAG
IBV             GGTGCTATTTCTTCTGTGATTCAAGAAATCTACCAGCAATTTGACGCCATACAAGCAAAT
                   ** **          *    *  **          * **     *    **  *

CcoV            GCACAAGTTGACAGGCTGATTACAGGACGACTTACAGCACTTAATGCATTTGTGTCTCAG
FcoV            GCACAAGTTGATAGGCTGATTACAGGTAGACTTACAGCACTTAATGCATTTGTGTCTCAG
TGE             GCACAAGTTGACAGGCTGATCACAGGAAGACTTACAGCACTTAATGCATTTGTGTCTCAG
PRCoV           GCACAAGTCGACAGGCTGATCACAGGAAGACTTACAGCACTTAATGCATTTGTGTCTCAG
HCoVOC43        CAACAAGTAGATAGGCTGATTACTGGTAGATTGGCTGCTTTGAATGTATTCGTTTCTCAT
PEDV            GTTCAGGTTGATCGTCTCATCACCGGCAGATTATCAGCACTTAATGCTTTTGTTGCCCAA
SARSUrba        GTACAAATTGACAGGTTAATTACAGGCAGACTTCAAAGCCTTCAAACCTATGTAACACAA
SARSTor2        GTACAAATTGACAGGTTAATTACAGGCAGACTTCAAAGCCTTCAAACCTATGTAACACAA
BcoV            GCTCAGATAGACAGACTTATTAATGGGCGTCTTACCGCTCTTAATGTTTATGTTTCTCAA
HEV             GCTCAGATAGACAGACTTATCAATGGGCGTCTCACCGCTCTTAATGCTTATGTTTCTCAG
MHV             GCTCAGATAGACCGTCTTATTAATGGCAGATTAACTGCACTTAATGCATATGTCTCTAAG
RtCoV           GCTCAGATAGACCGTCTTATTAATGGCAGATTAACTGCACTTAATGCATATGTCTCTAAG
IBV             GCTCAAGTGGATCGTCTTATAACTGGTAGATTGTCATCACTTTCTGTTTTAGCATCTGCT
                   **  * **  *  * ** *  **  *  *       *       *  *  *

CcoV            ACTTTAACCAGACAAGCAGAGGTTAGGGCTAGTAGACAACTTGCTAAAGACAAGGTTAAT
FcoV            ACTCTAACCAGACAAGCAGAGGTTAGGGCTAGTAGACAACTTGCCAAAGACAAGGTTAAT
TGE             ACTCTAACCAGACAAGCGGAGGTTAGGGCTAGTAGACAACTTGCCAAAGACAAGGTTAAT
PRCoV           ACTCTAACCAGACAAGCCGAGGTTAGGGCTAGTAGACAACTTGCTAAAGACAAGGTTAAT
HCoVOC43        ACATTGACTAAGTACACTGAAGTTCGTGCTTCCAGACAGCTTGCACAACAAAAAGTGAAT
PEDV            ACCCTCACTAAGTATACTGAGGTTCAGGCTAGCAGGAAGCTAGCACAGCAAAAGGTTAAT
SARSUrba        CAACTAATCAGGGCTGCTGAAATCAGGGCTTCTGCTAATCTTGCTGCTACTAAAATGTCT
SARSTor2        CAACTAATCAGGGCTGCTGAAATCAGGGCTTCTGCTAATCTTGCTGCTACTAAAATGTCT
BcoV            CAGCTTAGTGATTCTACACTAGTAAAATTTAGTGCAGCACAAGCTATGGAGAAGGTTAAT
HEV             CAGCTTAGTGATTCTACACTAGTAAAATTTAGTGCAGCACAAGCTATTGAGAAAGTTAAT
MHV             CAGCTGAGTGACATGACCCTTGTTAAGGTGAGTGCAGCCCAGGCTATAGAGAAAGTTAAT
RtCoV           CAGCTGAGCGACATGACCCTTATTAAGGTGAGTGCTGCCCAGGCTATAGAGAAAGTTAAT
IBV             AAGCAGGCGGAGTATATTAGAGTGTCACAACAGCGTGAGTTAGCTACTCAGAAAATTAAT
                                  *                    **       **  *   *

CcoV            GAATGCGTTAGGTCTCAATCCCAGAGATTTGGATTCTGTGGTA---ATGGTACACATTTG
FcoV            GAATGTGTTAGGTCTCAGTCTCAGAGATTCGGATTCTGTGGTA---ATGGTACACATTTG
TGE             GAATGCGTTAGGTCTCAGTCTCAGAGATTCGGATTCTGTGGTA---ATGGTACACATTTG
PRCoV           GAATGCGTTAGGTCTCAGTCTCAGAGATTCGGCTTCTGTGGTA---ATGGTACACATTTG
HCoVOC43        GAGTGTGTCAAATCCCAGTCTAAGCGTTATGGCTTCTGTGGAA---ATGGCACTCACATT
PEDV            GAGTGCGTCAAATCGCAATCTCAGCGTTACGGTTTTTGTGGTGGTGATGGCGAGCACATT
SARSUrba        GAGTGTGTTCTTGGACAATCAAAAAGAGTTGACTTTTGTGGAA---AGGGCTACCACCTT
SARSTor2        GAGTGTGTTCTTGGACAATCAAAAAGAGTTGACTTTTGTGGAA---AGGGCTACCACCTT
BcoV            GAATGTGTCAAAAGCCAATCATCTAGGATAAATTTTTGTGGTA---ATGGTAATCATATT
HEV             GAATGTGTTAAAAGCCAATCATCTAGGATAAATTTCTGTGGTA---ATGGTAATCATATT
```

```
MHV          GAGTGTGTTAAAAGCCAATCATCTAGGATAAATTTCTGTGGCA---ATGGCAATCATATA
RtCoV        GAGTGTGTTAAAAGCCAATCACCTAGGATAAATTTCTGTGGCA---ATGGCAATCATATA
IBV          GAGTGTGTTAAGTCACAGTCTATTAGGTACTCCTTTTGTGGTA---ATGGACGACATGTT
             ** ** **       ** **     *       ** *****     * **    **  *

CcoV         TTTTCACTTGCAAATGCGGCACCAAATGGCATGATTTTCTTTCACACAGTGCTATTACCA
FcoV         TTTTCACTAGCAAATGCAGCACCAAATGGCATGATTTTCTTTCATACAGTACTATTACCA
TGE          TTTTCACTCGCAAATGCAGCACCAAATGGCATGATTTTCTTTCACACAGTGCTATTACCA
PRCoV        TTTTCACTCGCAAATGCAGCACCAAATGGCATGATCTTCTTTCACACAGTGCTATTACCA
HCoVOC43     TTCTCAATTGTTAATGCTGCTCCTGAGGGGCTTGTTTTTCTCCACACTGTCTTGTTGCCG
PEDV         TTCTCTCTGGTACAGGCCGCACCTCAGGGCCTGCTGTTCTTACATACAGTACTTGTACCG
SARSUrba     ATGTCCTTCCCACAAGCAGCCCCGCATGGTGTTGTCTTCCTACATGTCACGTATGTGCCA
SARSTor2     ATGTCCTTCCCACAAGCAGCCCCGCATGGTGTTGTCTTCCTACATGTCACGTATGTGCCA
BcoV         ATATCATTAGTGCAGAATGCTCCATATGGTTTGTATTTTATCCACTTTAGCTATGTCCCT
HEV          ATATCATTAGTACAGAATGCTCCATATGGTTTGTATTTTATCCATTTTAGCTATGTCCCC
MHV          TTGTCATTAGTCCAGAATGCGCCTTATGGTTTGTATTTTATCCATTTTAGCTATGTCCCC
RtCoV        TTGTCATTAGTCCAGAATGCGCCTTACGGTTTATATTTTATTCATTTCAGCTATGTGCCT
IBV          CTAACCATACCGCAAAATGCACCTAATGGTATAGTGTTTATACACTTTTCTTATACTCCA
              *  *  *     *    ** **  * **  *    **  * **              **

CcoV         ACAGCTTATGAAACTGTGACGGCCTGGTCAGGTATTTGTGCGT---CAGATGGCAGTCGC
FcoV         ACAGCTTATGAAACTGTAACAGCTTGGTCAGGTATTTGTGCTT---CAGATGGCGATCGC
TGE          ACGGCTTATGAAACTGTGACTGCTTGGCCAGGTATTTGTGCTT---CAGATGGTGATCGC
PRCoV        ACGGCGTATGAAACTGTGACTGCTTGGTCAGGTATTTGTGCTT---TAGATGTTGATCGC
HCoVOC43     ACACAATATAAGGATGTTGAAGCGTGGTCTGGGTTGTGCGTTG---ATGGTACAAACGGT
PEDV         GGTGATTTTGTAAATGTTCTTGCCATCGCTGGCTTATGCGTTA---ATGGTGAAATTGCC
SARSUrba     TCCCAGGAGAGGAACTTCACCACAGCGCCAGCAATTTGTCATG---AAGGCAAAGCATAC
SARSTor2     TCCCAGGAGAGGAACTTCACCACAGCGCCAGCAATTTGTCATG---AAGGCAAAGCATAC
BcoV         ACTAAGTATGTCACTGCGAAGGTTAGTCCCGGTCTGTGCATTG---CTGGTGATAGAGGT
HEV          ACCAAGTATGTTACAGCAAAGGTTAGTCCTGGTTTGTGCATTG---CTGGCGATATAGGA
MHV          ACTTCCTTTACAACGGCAAATGTGAGTCCTGGGCTATGCATTT---CTGGTGATAGAGGA
RtCoV        ACATCCTTTACAACGGTAAATGTGAGTCCTGGACTATGCATTT---CTGGTGATAGAGGA
IBV          GATAGTTTTGTTAATGTTACTGCAATAGTGGGTTTTTGTGTAAAGCCAGCTAATGCTAGT
                                          *   * **          *

CcoV         ACTTTTGGACTTGTTGTTGAGGATGTCCAGCTGACGC-TATTTCGCAA-----TTTAGAT
FcoV         ACTTTCGGACTTGTCGTTAAAGATGTGCAGTTGACGT-TGTTTCGTAA-----TCTAGAT
TGE          ACTTTTGGACTTGTCGTTAAAGATGTCCAGTTGACTT-TGTTTCGTAA-----TCTAGAT
PRCoV        ACTTTTGGACTTGTCGTTAAAGATGTCCAGTTGACTT-TGTTTCGTAA-----TCTAGAT
HCoVOC43     TATGTGTTGCGACAACCTAATCTTGCTCT--TTACAAAGAAGGCA---------ATTATT
PEDV         TTGACTCTACGTGAGCCTGGCTTAGTCTTGTTTACGCATGAACTTCAAACTTATACTGCG
SARSUrba     TTCCCT---------CGTGAAGGTGTTTTTGTGTTTAATG--------------GCACTT
SARSTor2     TTCCCT---------CGTGAAGGTGTTTTTGTGTTTAATG--------------GCACTT
BcoV         ATAGCC---------CCTAAGAGTGGTTATTTTGTTAATGTAA-----------ATAATA
HEV          ATATCG---------CCTAAGAGTGGTTATTTTATTAATGTAA-----------ATAATT
MHV          TTAGCA---------CCTAAAGCTGGATATTTTGTTCAAGATG-----------ATGGAG
RtCoV        TTAGCA---------CCTAAAGCTGGATATTTTGTTCAAGATC-----------ATGGAG
IBV          CAGTATGCAATAGTGCCCGCTAATGGTAGGGGTATTTTTATACAA---------GTTAAT
                                  *

CcoV         GAAAAATTTTATTTGACGCCCAGAACTATGTATCAGCCCAGAGTTGCAACTAGTTCTGAT
FcoV         GACAAGTTCTATTTGACCCCCAGAACTATGTATCAGCCTAGAGTTGCAACTAGTTCTGAT
TGE          GACAAGTTCTATTTGACCCCCAGAACTATGTATCAGCCTAGAGTTGCAACTAGTTCTGAC
PRCoV        GACAAGTTCTATTTGACACCCAGAACTATGTATCAGCCTAGAGTGGCAACTAGTTCTGAT
HCoVOC43     ATAGAAT-------CACATCTCGCATAATGTTTGAACCACGTATTCCTACCATGGCAGAT
PEDV         ACGGAATATTTTGTTTCATCGCGACGTATGTTTGAACCTAGAAAACCTACCGTTAGTGAT
SARSUrba     CTTGGTTTAT----TACACAGAGGAACTTCTTTTCTCCACAAATAATTACTACAGACAAT
SARSTor2     CTTGGTTTAT----TACACAGAGGAACTTCTTTTCTCCACAAATAATTACTACAGACAAT
BcoV         CTTGGATGTT----CACTGGTAGTGGTTATTACTACCCTGAACCCATAACTGGAAATAAT
HEV          CTTGGATGTT----CACTGGTAGTAGCTATTACTACCCTGAACCTATAACCCAAAATAAT
MHV          AGTGGAAGTT----CACAGGTAGTAATTATTATTACCCTGAACCCATTACAGATAAAAAT
RtCoV        AATGGAAGTT----CACAGGTAGCAATTATTACTACCCTGAATCCATTACAGATAAAAAC
IBV          GGTAGTTACTACATCACTGCACGAGATATGTATATGCCAAGAGCTATTACTGCAGGAGAT
```

```
                                 *     *       *     **          **        *

CcoV             TTTGTTCAAATAGAAGGCTGTGATGTGTTGTTTGTTAATGGAACTGTAATTGAATTGCCT
FcoV             TTTGTTCAAATTGAAGGGTGTGATGTGTTGTTTGTCAACGCGACTGTAATTGATTTGCCT
TGE              TTTGTTCAAATTGAAGGGTGCGATGTGCTGTTTGTTAATGCAACTGTAAGTGATTTGCCT
PRCoV            TTTGTTCAAATTGAAGGGTGCGATGTGCTGTTTGTTAATACAACTGTAAGTGATTTGCCT
HCoVOC43         TTTGTTCAAATTGAAAATTGCAATGTCACATTTGTTAACATTTCTCGCTCTGAGTTGCAA
PEDV             TTTGTTCAAATTGAGAGTTGTGTGGTCACCTATGTCAATCTGACTAGCGACCAGCTACCA
SARSUrba         ACATTTGTCTCAGGAAATTGTGATGTCGTTATTGGCATCATTAACAACACAGTTTATGAT
SARSTor2         ACATTTGTCTCAGGAAATTGTGATGTCGTTATTGGCATCATTAACAACACAGTTTATGAT
BcoV             GTTGTTGTTATGAGTACCTGTGCTGTTAACTATACTAAAGCGCCGGATGTAATGCTGAAC
HEV              GTTGTTGTGATGAGTACCTGTGCTGTTAATTATACTAAAGCACCGGATCTAATGCTGAAC
MHV              AGTGTCGTGATGAGTAGTTGCGCAGCAAACTACACAAAGGCACCTGAAGTTTTCTTGAAC
RtCoV            AGTGTCGTGATGAGTAGTTGCGCAGTAAACTACACAAAGGCACCTGAAGTTTTCTTGAAC
IBV              GTAGTTACGCTTACTTCTTGTCAAGCAAATTATGTAAGTGTAAATAAGACCGTCATTACT
                      *              **    *           *

CcoV             AGTATCATA---CCTGACTATATCGATATTAATCAAACTGTTCAGGACATATTAGAAAAT
FcoV             AGTATTATA---CCTGACTATATTGACATTAATCAAACTGTTCAAGACATATTAGAAAAT
TGE              AGTATTATA---CCTGATTATATTGATATTAATCAGACTGTTCAAGACATATTAGAAAAT
PRCoV            AGTATTATA---CCTGATTATATTGATATTAATCAGACTGTTCAAGACATATTAGAAAAT
HCoVOC43         ACCATTGTG---CCAGAGTATATTGATGTTAATAAGACGCT--GCAAGAATTAAGTTACA
PEDV             GATGTAATC---CCAGATTACATCGATGTTAACAAAACACTTGATGAGATTTTAGCTTCT
SARSUrba         CCTCTGCAA---CCTGAGCTCGACTCATTCAAAGAAGAGCTGGACAAGTACTTCAAAAAT
SARSTor2         CCTCTGCAA---CCTGAGCTTGACTCATTCAAAGAAGAGCTGGACAAGTACTTCAAAAAT
BcoV             ATTTCAACA---CCCAACCTCCATGATTTTAAGGAAGAGTTGGATCAATGGTTTAAAAAC
HEV              ACATCGACA---CCCAACCTTCCTGACTTCAAGGAAGAATTGTATCAATGGTTTAAAAAC
MHV              ACTTCAATA---CCTAATCTACCCGACTTTAAGGAGGAGTTAGATAAATGGTTTAAAAAT
RtCoV            ACTTCAATA---ACTAATCTACCCGACTTTAAGGAGGAGTTAGATAAATGGTTTAAGAAT
IBV              ACATTCGTAGACAATGATGATTTTGATTTTAATGACGAATTGTCAAAATGGTGGAATGAT
                              *              * **  *     *     *     *

CcoV             TTCAGACCAAATTGGACTGTACCCGAGTTGCCACTTGACATTTTTCATGCAACCTACTTA
FcoV             TACAGACCAAACTGGACTGTACCTGAATTTACACTTGATATTTTCAACGCAACCTATTTA
TGE              TTTAGACCAAATTGGACTGTACCTGAGTTGACATTTGACATTTTTAACGCAACCTATTTA
PRCoV            TTTAGACCAAATTGGACTGTACCTGAGCTGACATTGGACGTTTTTAACGCAACCTATTTA
HCoVOC43         AATTG-CCAAATTACACTGTTCCAGACCTAGTTGTCGAACAGTACAACCAGACTATTTTG
PEDV             --CTG-CCCAATAGAACTGGTCCAAGTCTTCCCCTAGATGTTTTTAATGCCACTTATCTT
SARSUrba         CATACATCACCAGATGTTGATCT-TGGCGACATTTCAGGCA--TTAACGCTTCTGTCGTC
SARSTor2         CATACATCACCAGATGTTGATCT-TGGCGACATTTCAGGCA--TTAACGCTTCTGTCGTC
BcoV             CAAACATCAGTGGCACCAGATTTGTCACTTGATTATA------TAAATGTTACATTCTTG
HEV              CAATCTTCAGTGGCACCAGATTTGTCACTTGATTATA------TTAATGTTACGTTCTTG
MHV              CAGACGTCTATTGCGCCTGATTTATCTCTCGATTTCGAGAAATTAAACGTTACCCTCCTG
RtCoV            CAGACGTCTATTGTGCCTGATTTATCTTTCGATATCGGGAAATTAAATGTTACATTCCTT
IBV              -ACTAAGCATGAGCTACCAGACTTTGAC--AAATTCAATTA---CACAGTACCTATACTT
                        *                                                *     *

CcoV             AACCTGACTGGTGAAATTAATGACTTAGAATTTAGGTCAGAAAAGTTACATAACACCACA
FcoV             AATCTGACTGGTGAAATTGATGACTTAGAGTTTAGGTCAGAAAAGCTACATAACACTACA
TGE              AACCTGACTGGTGAAATTGATGACTTAGAATTTAGGTCAGAAAAGCTACATAACACCACT
PRCoV            AACCTGACTGGTGAAATTGATGACTTAGAGTTTAGGTCAGAAAAGCTACATAACACTACT
HCoVOC43         AATTTGACCAGTGAAATTAGCACCCTTGAAAATAAATCTGCGGAGCTTAATTACACTGTT
PEDV             AATCTTACTGGTGAAATTGCAGATCTAGAGCAGCGTTCAGAGTCTCTCCGTAATACTACA
SARSUrba         AACATTCAAAAAGAAATTGACCGCCTCAATGAGG--------------------------
SARSTor2         AACATTCAAAAAGAAATTGACCGCCTCAATGAGG--------------------------
BcoV             GACCTACAAGATGAAATGAATAGGTTACAGGAGG--------------------------
HEV              GACCTACAAGATGAAATGAATAGGTTACAAGAGG--------------------------
MHV              GACCTGACTGATGAGATGAACAGGATTCAGGATG--------------------------
RtCoV            GACCTGTCCTATGAGATGAACAGGATTCAGGATG--------------------------
IBV              GACATTGATAGTGAAATTGATCGTATTCAAGGCG--------------------------
                  *  *        ** **        *  *

CcoV             GTAGAACTTGCTATTCTCATTGATAATATTAATAACACATTAGTCAATCTTGAATGGCTC
```

```
FcoV          GTAGAACTTGCCATTCTCATTGATAACATTAATAATACATTAGTCAATCTTGAATGGCTC
TGE           GTAGAACTTGCCATTCTCATTGACAACATTAACAATACATTAGTCAATCTTGAATGGCTC
PRCoV         GTAGAACTTGCCATTCTCATTGACAACATTAACAATACAGTAGTCAATCTTGAATGGCTT
HCoVOC43      CAAAAATTGCAAACTCTGATTGACAACATAAATAGCACATTAGTCGACTTAAAGTGGCTC
PEDV          GAAGAGCTCCGAAGTCTCATTAACAACATCAACAACACACTTGTTGACCTTGAGTGGCTC
SARSUrba      ----------------TCGCTAAAAATTTAAATGAATCACTCATTGACCTTCAAGAATTG
SARSTor2      ----------------TCGCTAAAAATTTAAATGAATCACTCATTGACCTTCAAGAATTG
BcoV          ----------------CAATAAAAGTTTTAAATCAGAGCTACATCAATCTCAAGGACATT
HEV           ----------------CTATAAAAGTTTTAAATCAGAGCTACATCAATCTCAAGGACATT
MHV           ----------------CAATTAAGAAGTTAAATGAGAGTTACATCAACCTCAAGGACGTT
RtCoV         ----------------CAATTAAGAATTTAAATGAGAGTTACATCAACCTCAAGGAAATT
IBV           ----------------TTATACAGGGTCTTAATGACTCTCTAATAGACCTTGAAAAACTT
                               *     * **           *  *  *  *     *

CcoV          AACAGAATTGAAACTTATGTAAAATGGCCTTGGTATGTTTGGCTACTAATTGGATTAGTA
FcoV          AATAGAATTGAAACTTATGTAAAATGGCCTTGGTATGTGTGGCTACTGATAGGTTTAGTA
TGE           AATAGAATTGAAACCTATGTAAAATGGCCTTGGTATGTGTGGCTACTAATAGGCTTAGTA
PRCoV         AATAGAATTGAAACTTATGTAAAATGGCCTTGGTATGTGTGGCTACTAATAGGCTTAGTA
HCoVOC43      AACCGGGTTGAGACTTACATCAAGTGGCCGTGGTGGGTGTGGTTGTGCATTTCAGTCGTG
PEDV          AACCGAGTTGAGACATACATCAAGTGGCCGTGGTGGGTTTGGTTGATCATTGTTATTGTT
SARSUrba      GGAAAATATGAGCAATATATTAAATGGCCTTGGTATGTTTGGCTCGGCTTCA---TTGCT
SARSTor2      GGAAAATATGAGCAATATATTAAATGGCCTTGGTATGTTTGGCTCGGCTTCA---TTGCT
BcoV          GGTACATATGAGTATTATGTAAAATGGCCTTGGTATGTATGGCTTTTAATTGGCTTTGCT
HEV           GGTACATATGAGTATTATGTGAAATGGCCTTGGTATGTATGGCTTTTAATTGGCCTTGCT
MHV           GGCACATATGAAATGTATGTGAAATGGCCTTGGTATGTGTGGTTGCTAATTGGATTAGCT
RtCoV         GGCACATATGAGATGTATGTGAAATGGCCTTGGTATGTTTGGCTGCTAATTGGATTAGCT
IBV           TCAATACTCAAAACTTATATTAAGTGGCCTTGGTATGTGTGGTTAGCCATAGCTTTTGCC
                        *    **  * ** ***** ****  ** *** *     *     * *

CcoV          GTAATATTCTGCATACCCATATTGCTATTTTGTTGTTGTAGTACTGGTTGTTGTGGATGT
FcoV          GTAGTATTTTGCATACCATTACTGCTATTTTGCTGTTTTAGCACAGGTTGTTGTGGATGC
TGE           GTAATATTTTGCATACCATTACTGCTATTTTGCTGTTGTAGTACAGGTTGCTGTGGATGC
PRCoV         GTAATATTTTGCATACCATTACTGCTATTTTGCTGTTGTAGTACAGGTTGCTGTGGATGC
HCoVOC43      CTCATCTTTGTGGTGAGTATGTTGCTATTATGTTGTTGTTCTACTGGTTGCTGTGGCTTC
PEDV          CTCATCTTTGTTGTGTCATTACTAGTGTTCTGCTGCATTTCCACGGGTTGTTGTGGATGC
SARSUrba      GGACTAATTGCCATCGTCATGGTTACAATCTTGCTTTGTTGCATGACTAGTTGTTGCAGT
SARSTor2      GGACTAATTGCCATCGTCATGGTTACAATCTTGCTTTGTTGCATGACTAGTTGTTGCAGT
BcoV          GGTGTAGCTATGCTTGTTTTACTATTCTTCATATGCTGTTGTACAGGATGTGGGACTAGT
HEV           GGTGTAGCTATGCTTGTTTTACTATTCTTCATATGCTGCTGTACAGGATGTGGGACTAGT
MHV           GGTGTAGCTGTTTGTGTGTTGTTATTTTTCATATGTTGCTGCACGGGTTGTGGCTCATGT
RtCoV         GGTGTAGCTGTTTGTGTTTTGTTATTTTTTATATGTTGCTGCACAGGTTGTGGCTCTTGT
IBV           ACTATTATCTTCATCTTAATACTAGGATGGGTTTTCTTCATGACTGGTTGTTGTGGTTGT
                 *                 *  *                      *      *  *

CcoV          ATCGGGTGTTTAGGAAGCTGTTGTCATTCCATAT-GTAGTAGAGGCCA---ATTTGAAAG
FcoV          ATAGGTTGTTTAGGAAGTTGTTGTCACTCTATAT-GTAGTAGAAGACA---ATTTGAAAA
TGE           ATAGGTTGTTTAGGAAGTTGTTGTCACTCTATAT-GTAGTAGAAGACA---ATTTGAAAA
PRCoV         ATAGGTTGTTTAGGAAGTTGTTGTCACTCTATAT-TCAGTAGAAGACA---ATTTGAAAA
HCoVOC43      TTTAGTTGTTTTGCATCTTCTATTAGAGGTTGTT-GTGAATCAACTAA---ACTTCCTTA
PEDV          TGCGGTTGCTGCGGTGCTTGTTTTTCAGGTTGTT-GTAGGGGTCCTAG---ACTTCAACC
SARSUrba      TGCCTCAAGGGTGCATGCTCTTGTGGTTCTTGCT-GCAAGTTTGATGA---GGATGACTC
SARSTor2      TGCCTCAAGGGTGCATGCTCTTGTGGTTCTTGCT-GCAAGTTTGATGA---GGATGACTC
BcoV          TGTTTTAAGAT---ATG---TGGTGGTTGTTGTG-ATGATTATACTGG---ACACCAGG-
HEV           TGTTTTAAGAA---ATG---TGGCGGTTGTTGTG-ATGATTATACTGG---ACACCAGG-
MHV           TGTTTCAAGAA---GTG---TGGAAATTGTTGTG-ATGAGTGTGGAGG---ACACCAGGA
RtCoV         TGTTTTAAGAA---ATG---TGGAAATTGTTGTG-ATGAGTATGGAGG---ACGTCAGGC
IBV           TGTTGTGGATGC-TTTGGCATTATGCCTCTAATGAGTAAGTGTGGTAAGAAATCTTCTTA
                                 *

CcoV          TTATGAACCTATTGAAAAAGTTCATGTTCACTGA-------------------------
FcoV          TTATGAACCAATTGAAAAAGTGCATGTCCACTAA-------------------------
TGE           TTACGAACCAATTGAAAAAGTGCACGTCCATTAA-------------------------
PRCoV         TTATGAACCTATTGAAAAAGTGCACGTCCATTAA-------------------------
```

```
HCoVOC43        TTACGACG---TTGAAAAGATCCACATACAGTAA-------------------------
PEDV            TTACGAAGCTTTTGAAAAGGTCCACGTGCAGTGA-------------------------
SARSUrba        TGAGCCAGTTCTCAAGGGTGTCAAATTACATTACACATAA-------------------
SARSTor2        TGAGCCAGTTCTCAAGGGTGTCAAATTACATTACACATAA-------------------
BcoV            -AGT-TAGTAAT-TAAAA-----CATCACATGACGACTAA-------------------
HEV             -AGT-TTGTAAT-CAAAA-----CTTCACATGACGATTAA-------------------
MHV             CAGTATTGTGATACATAATATTTCCTCTCATGAGGATTGA-------------------
RtCoV           AGGTATTGTGATACATAATATTTCCTCTCATGAGGATTGA-------------------
IBV             TTACACGACTTTTGATAACGATGTGGTAACTGAACAATACAGACCTAAAAAGTCTGTTTG
                          *  *

CcoV            -
FcoV            -
TGE             -
PRCoV           -
HCoVOC43        -
PEDV            -
SARSUrba        -
SARSTor2        -
BcoV            -
HEV             -
MHV             -
RtCoV           -
IBV             A
```

Figure 4: List of Molecular Designs for S Gene

LK249 5'-FAM-CCCACGCCAGAAGGTAGATCACGAACTACACGTGGG-3'-Dubcyl

LK249.N Dubcyl 5'-FAM-GCCCACGCCAGAAGGTAGATCACGAACTACACGTGGGC-3'-

LK250

5'-CTCTATGTTTATAAGGGCTATCAACCTATAGATGTAGTTCGTGATCTA CCTTCTGGTTTTAACACTTTGAAACCTATTTTTAAGTTGCCTCTTGG -3'

LK251 5'-CTCTATGTTTATAAGGGCTATCAACC-3'

LK251-T7 5'-TAATACGACTCACTATAGGCTCTATGTTTATAAGGGCTATCAACC-3'

LK252 5'-CCAAGAGGCAACTTAAAAATAGGTTTC- 3'

LK252-RT 5'- AGGCTGTAAGAACCAAGAGGCAACTTAAAAATAGGTTTC- 3'

S-RT 5'-AGGCTGTAAGAA-3'

Figure 5: CLUSTAL W (1.81) multiple DNA sequence alignment of coronavirus *E* genes from strains isolated from different species

```
SARS-Urb    ------------------------------------------------------------
SARS-Tor    ------------------------------------------------------------
HCoOC43     ------------------------------------------------------------
PEDV        ------------------------------------------------------------
TGE         ------------------------------------------------------------
PRCoV       ------------------------------------------------------------
CcoV        ------------------------------------------------------------
FcoV        ------------------------------------------------------------
MHV         ------------------------------------------------------------
RtCoV       ------------------------------------------------------------
BcoV        ------------------------------------------------------------
IBV         ATGAATTTATTGAATAAGTCGCTAGAGGAGAATGGAAGTTTTCTAACAGCGCTTTACATA

SARS-Urb    ----------ATGTACTCATTCGTTTCGGAAGAAACAGGTACGTTAATAGTTAATAGCGT
SARS-Tor    ----------ATGTACTCATTCGTTTCGGAAGAAACAGGTACGTTAATAGTTAATAGCGT
HCoOC43     ------------ATGTTCCTTAAGCTAGTGGATGATCA-TGCTTTGGTTGTTAATGTACT
PEDV        ---------------ATGCTACAATTAGTGAATGATAA-TGGTCTAGTAGTTAATGTTAT
TGE         ATGACGTTTCCTAGGGCATTGACTGTCATAGATGACAA-TGGAATGGTCATTAACATCAT
PRCoV       ATGACGTTTCCTAGGGCATTGACTGTCATAGATGACAA-CGGAATGGTCATTAGCATCAT
CcoV        ATGACGTTCCCTCGGGCATTGACTGTCATAGATGACAA-TGGAATGGTCATTAGTATCAT
FcoV        ATGACGTTCCCTAGGGCATTTACTATCATAGATGACCA-TGGCATGGTTGTTAGCGTCTT
MHV         ---------ATGTTTAATTTATTCCTTACAGATACAGTATGGTATGTGGGGCAG-ATTAT
RtCoV       ---------ATGTTTAATTTATTCCTTATAGACACAGTATGGTACGTGGGGCAG-ATTAT
BcoV        ---ATGTTTATGGCTGATGCTTATTTTGCAGACACTGTGTGGTATGTGGGGCAA-ATAAT
IBV         ATTGTAGGATTTTTAGCACTTTATCTTCTAGGTA--GAGCACTTCAAGCATTTGTACAGG
                                        *

SARS-Urb    ACTTCTTTTTCTTGCTTTCGTGGTATT-------CTTGCTAGTCACACTAGCCATCCTTA
SARS-Tor    ACTTCTTTTTCTTGCTTTCGTGGTATT-------CTTGCTAGTCACACTAGCCATCCTTA
HCoOC43     ACTCTGGTGTGTGGTGCTTATAGTGAT-------ACTACTAGTGTGTATTACAATAATTA
PEDV        ACTTTGGCTTTTCGTACTCTTTTTCCT-------GCTTATTATAAGCATTACCTTCGTCC
TGE         TTTCTGGTTCCTGTTGATAATTATATT-------GATATTACTTTCAATAGCATTGCTAA
PRCoV       TTTTTGGTTCCTGTTGATAATTATATT-------GATATTACTTTCAATAGCATTGCTAA
CcoV        TTTCTGGTTCCTGTTGATAATTATATT-------GATATTATTTTCAATAGCATTGCTAA
FcoV        CTTCTGGCTCCTGTTGATAATTATATT-------GATATTGTTTTCAATAGCATTGCTAA
MHV         TTTTATAGTCGCAGTGTGTTTGATGGT-------CACCATAATAGTGGTTGCCTTCCTTG
RtCoV       TTTTATAGTCGCAGTGTGTTTGATGGT-------CACCATAATTGTGGTTGCCTTCCTTG
BcoV        TTTTATAGTTGCCATTTGTTTATTGGT-------TATAATAGTTGTAGTGGCATTTTTGG
IBV         CTGCTGATGCTTGTTGTTTATTTTGGTATACATGGGTAGTAATTCCAGGAGCTAAGGGTA
                              *  *  *           *  *        *

SARS-Urb    CTGCGCTTCGATT---GTGTGCGTACTGCTGCAATATTGTTAACGTGAGTTTAGTAAAAC
SARS-Tor    CTGCGCTTCGATT---GTGTGCGTACTGCTGCAATATTGTTAACGTGAGTTTAGTAAAAC
HCoOC43     AACTAATTAAGCT---TTGTTTCACTTGCCATATGTTTTGTAATAGAACAGTTTATGGCC
PEDV        AATTGGTTAATCT---GTGCTTCACTTGTCACCGGTTGTGTAATAGCGCAGTTTATACAC
TGE         ATATAATTAAGCT---ATGCATGGTGTGTTGCAATTTAGGAAGGACAGTTATTATTGTTC
PRCoV       ATATAATTAAGCT---ATGCATGGTGTGTTGCAATTTAGGAAGGACAGTTATTATTGTTC
CcoV        ATATAATTAAGCT---ATGCATGGTATGTTGCAATTTAGGAAGAACAGTTATTATTGTTC
FcoV        ATGTTATTAAATT---GTGCATGGTATGTTGCAATTTGGGTAAGACTATTATAGTACTAC
MHV         CGTCTATCAAACT---TTGTATTCAACTTTGCGGTTTGTGTAATACTTTGTTGCTGTCTC
RtCoV       CGTCTATTAAACT---TTGTATTCAACTTTGCGGTTTGTGTAATACTTTGTTGCTGTCTC
BcoV        CAACTTTTAAATT---GTGTATTCAACTTTGCGGTATGTGTAATACCTTAGTACTGTCCC
IBV         CAGCCTTTGTATACAAGTATACATATGGTAGAAAACTTAACAATCCGGAATTAGAAGCAG
                  *          *                  *    *        *

SARS-Urb    CAACGGTTTACGTCTAC--TCGCGTG--TTAAAAATCTGAACTCTTCTGAAGGA------
SARS-Tor    CAACGGTTTACGTCTAC--TCGCGTG--TTAAAAATCTGAACTCTTCTGAAGGA------
```

```
HCoOC43         CCATTAAAAATGTGTA---CCACATT--TACCAATCATATATGCACATAGACCC-----T
PEDV            CTATAGGGCGCCTGTA---TAGAGTT--TATAAGTCTTACATGCGAATTGACCC-----C
TGE             CAGCGCAACATGCTTA---CGATGCC--TATAAGAATTTTATGCGAATTAAAGCAT--AC
PRCoV           CAGTGCAACATGCTTA---CGATGCC--TATAAGAATTTTATGCGAATTAAAGCAT--AC
CcoV            CAGCTCGACATGCCTA---TGATGCC--TATAAGAATTTTATGCAAATTAGAGCAT--AC
FcoV            CTGCACGCCATGCATA---TGATGCC--TATAAGACCTTTATGCAAACCAAGGCAT--AT
MHV             CTTCTATTTGTGTGTATAATAGGAGT--AAGCAGCTTTATAAGTATTATAATGAAGAAGT
RtCoV           CTTCTATTTATGTGTATAATAGGAGT--AAGCAGCTTTATAAGTATTATAATGAAGAAGT
BcoV            CTTCTATTTATGTGTTTAATAGAGGT--AGGCAGTTTTATGAGTTTTACAACGAT---GT
IBV             TTATTGTTAACGAGTTTCCTAAGAACGGTTGGAATAATAAAAATCCAGCAAATTTTCAAG
                             *                   *    *

SARS-Urb        GTTCCTGATCTTCTGGTCTAA--------------------------
SARS-Tor        GTTCCTGATCTTCTGGTCTAA--------------------------
HCoOC43         TTCCCTAAACGAGTTATTGATTTCTAA---------------------
PEDV            CTCCCCAGTACTGTTATTGACGTATAA---------------------
TGE             AACCCCGATGGAGCACTCCTTGCTTGA---------------------
PRCoV           AACCCTGATGGAGCACTCCTTGTTTGA---------------------
CcoV            AACCCTGATGAAGCACTCCTTGTTTGA---------------------
FcoV            AATCCCGACGAAGCATTTTTGGTTTGA---------------------
MHV             GAGACCGCCCCCGTTAGAGGTGGATGATATAATAATCCAAACATTATGA
RtCoV           GAGACCGCCCCCGTTAGAGGTGGATGATATAATAATCCAAACATTATGA
BcoV            AAAACCACCAGTTCTTGATGTGGATGACGTT-TAG-------------
IBV             ATGCCCAACGAGACAAATTGTACTCTTGA-------------------
          *
```

Figure 6: Phylogenetic Analysis of *E* Gene

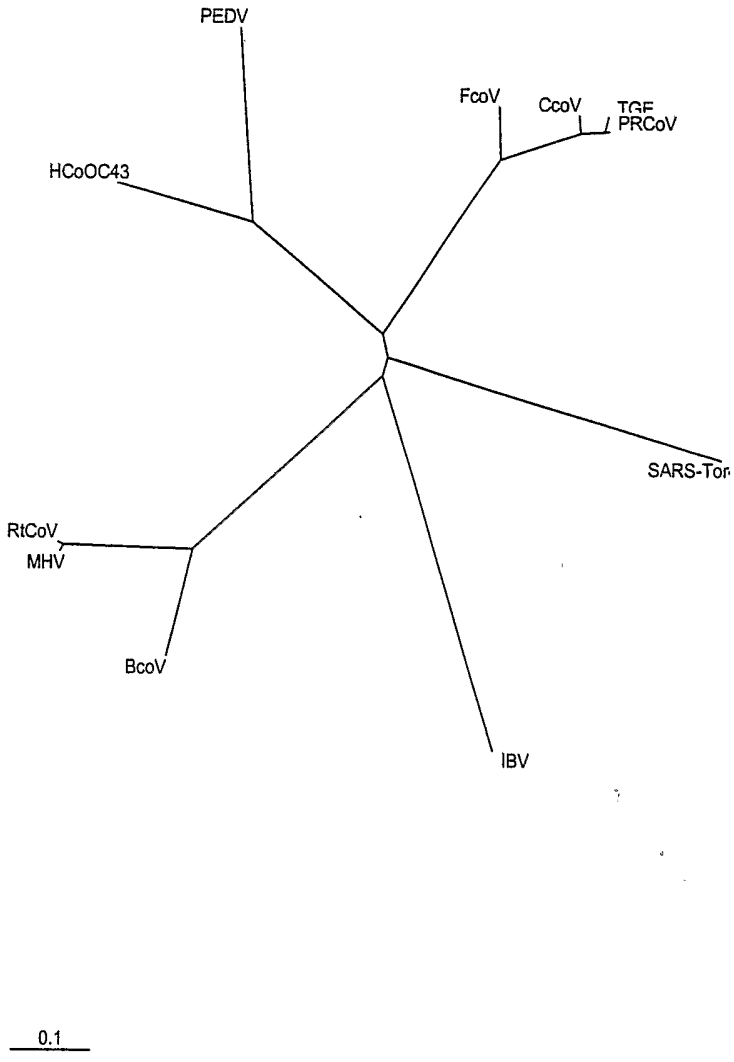
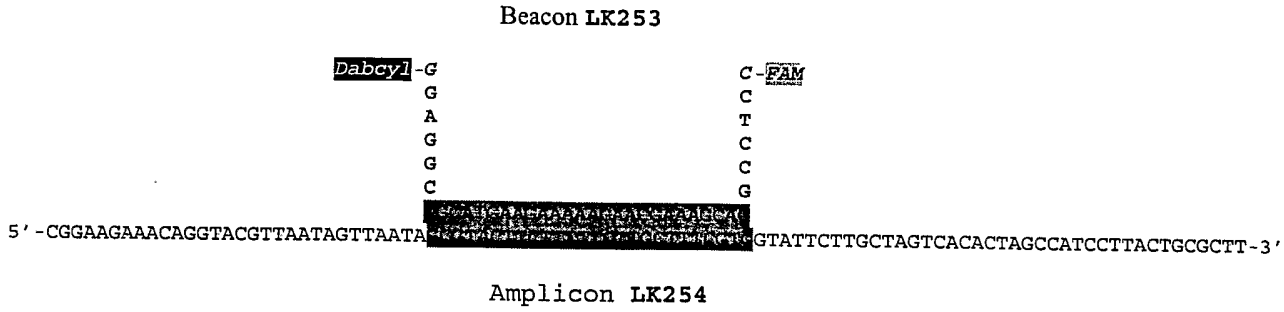


Figure 7: Molecular Designs for SARS-Associated *E* Gene

NOTE:	LK253 is a TET-molecular beacon, which recognizes the ***E* gene of coronavirus (SARS Tor2 and SARS urbani human pathogenic strains)**. It will be used in a real-time PCR diagnostic for the identification of **SARS-associated coronavirus RNA/DNA.**



LK253

E.	Target recognition sequence:	**26 nucleotides (11 G/C)**
F.	Length of the arms:	**6 nucleotides (5 G/C)**
G.	Melting temperature of the beacon:	**dG = -2.29 dH = -67.2 dS = -200.8 Tm = 61.5 °C**
H.	Melting temperature of target:	**°C**

LK255 **5'-CGGAAGAAACAGGTACGTTAATAG-3'**

Length: 24 nucleotides (10 G/C)
Tm: °C
Position: (see alignment below)

LK256 **5' –AAGCGCAGTAAGGATGGCTA-3'**

Length: 20 nucleotides (10 G/C)
Tm: °C
Position: (see alignment below)

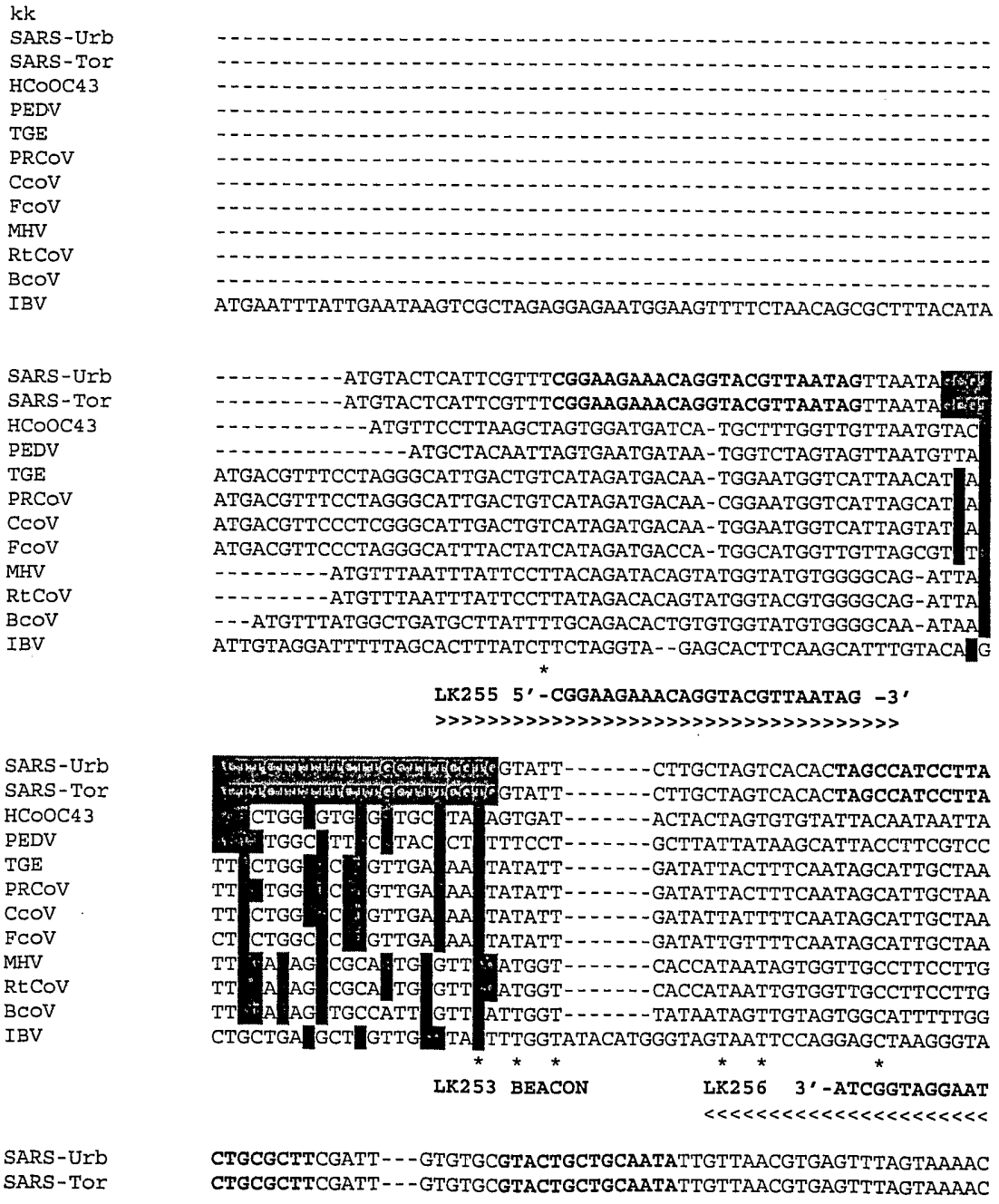
E-RT **5' –TATTGCAGCAGTAC-3'**

Length: 14 nucleotides (10 G/C)
Tm: 47 °C
Position: (see alignment below)

LK254 Amplicon (95 nucleotides)

5'CGGAAGAAACAGGTACGTTAATAGTTAATAGCGTACTTCTTTTTCTTGCTTTCGT
GGTATTCTTGCTAGTCACACTAGCCATCCTTACTGCGCTT-3'

DNA sequence alignment of complete
E genes from corona virus strains



```
HCoOC43     AACTAATTAAGCT---TTGTTTCACTTGCCATATGTTTTGTAATAGAACAGTTTATGGCC
PEDV        AATTGGTTAATCT---GTGCTTCACTTGTCACCGGTTGTGTAATAGCGCAGTTTATACAC
TGE         ATATAATTAAGCT---ATGCATGGTGTGTTGCAATTTAGGAAGGACAGTTATTATTGTTC
PRCoV       ATATAATTAAGCT---ATGCATGGTGTGTTGCAATTTAGGAAGGACAGTTATTATTGTTC
CcoV        ATATAATTAAGCT---ATGCATGGTATGTTGCAATTTAGGAAGAACAGTTATTATTGTTC
FcoV        ATGTTATTAAAATT---GTGCATGGTATGTTGCAATTTGGGTAAGACTATTATAGTACTAC
MHV         CGTCTATCAAACT---TTGTATTCAACTTTGCGGTTTGTGTAATACTTTGTTGCTGTCTC
RtCoV       CGTCTATTAAACT---TTGTATTCAACTTTGCGGTTTGTGTAATACTTTGTTGCTGTCTC
BcoV        CAACTTTTAAATT---GTGTATTCAACTTTGCGGTATGTGTAATACCTTAGTACTGTCCC
IBV         CAGCCTTTGTATACAAGTATACATATGGTAGAAAACTTAACAATCCGGAATTAGAAGCAG
                  *           *                     *    *          *
            GACGCGAA-5'         3'-CATGACGACGTTAT-5' E-RT
            <<<<<<<<<<
```

```
SARS-Urb    CAACGGTTTACGTCTAC--TCGCGTG--TTAAAAATCTGAACTCTTCTGAAGGA------
SARS-Tor    CAACGGTTTACGTCTAC--TCGCGTG--TTAAAAATCTGAACTCTTCTGAAGGA------
HCoOC43     CCATTAAAAATGTGTA---CCACATT--TACCAATCATATATGCACATAGACCC-----T
PEDV        CTATAGGGCGCCTGTA---TAGAGTT--TATAAGTCTTACATGCGAATTGACCC-----C
TGE         CAGCGCAACATGCTTA---CGATGCC--TATAAGAATTTTATGCGAATTAAAGCAT--AC
PRCoV       CAGTGCAACATGCTTA---CGATGCC--TATAAGAATTTTATGCGAATTAAAGCAT--AC
CcoV        CAGCTCGACATGCCTA---TGATGCC--TATAAGAATTTTATGCAAATTAGAGCAT--AC
FcoV        CTGCACGCCATGCATA---TGATGCC--TATAAGACCTTTATGCAAACCAAGGCAT--AT
MHV         CTTCTATTTGTGTGTATAATAGGAGT--AAGCAGCTTTATAAGTATTATAATGAAGAAGT
RtCoV       CTTCTATTTATGTGTATAATAGGAGT--AAGCAGCTTTATAAGTATTATAATGAAGAAGT
BcoV        CTTCTATTTATGTGTTTAATAGAGGT--AGGCAGTTTTATGAGTTTTACAACGAT---GT
IBV         TTATTGTTAACGAGTTTCCTAAGAACGGTTGGAATAATAAAAATCCAGCAAATTTTCAAG
                        *                   *    *
```

```
SARS-Urb    GTTCCTGATCTTCTGGTCTAA----------------------------
SARS-Tor    GTTCCTGATCTTCTGGTCTAA----------------------------
HCoOC43     TTCCCTAAACGAGTTATTGATTTCTAA----------------------
PEDV        CTCCCCAGTACTGTTATTGACGTATAA----------------------
TGE         AACCCCGATGGAGCACTCCTTGCTTGA----------------------
PRCoV       AACCCTGATGGAGCACTCCTTGTTTGA----------------------
CcoV        AACCCTGATGAAGCACTCCTTGTTTGA----------------------
FcoV        AATCCCGACGAAGCATTTTTGGTTTGA----------------------
MHV         GAGACCGCCCCCGTTAGAGGTGGATGATATAATAATCCAAACATTATGA
RtCoV       GAGACCGCCCCCGTTAGAGGTGGATGATATAATAATCCAAACATTATGA
BcoV        AAAACCACCAGTTCTTGATGTGGATGACGTT-TAG--------------
IBV         ATGCCCAACGAGACAAATTGTACTCTTGA--------------------
                *
```

Figure 8: Molecular Designs for SARS-Associated *E* Gene

LK253 Dubcyl 5'-FAM-CCTCCGCACGAAAGCAAGAAAAAGAAGTACGCCGGAGG-3'-

LK253.N Dubcyl 5'-FAM-GCCTCCGCACGAAAGCAAGAAAAAGAAGTACGCCGGAGGC-3'-

LK254 5'-CGGAAGAAACAGGTACGTTAATAGTTAATAGCGTACTTCTTTTTC TTGCTTTCGTGGTATTCTTGCTAGTCACACTAGCCATCCTTACTGCGCTT -3'

LK255 5'- CGGAAGAAACAGGTACGTTAATAG -3'

LK255-T7 5'-TAATACGACTCACTATAGGCGGAAGAAACAGGTACGTTAATAG -3'

LK256 5'- AAGCGCAGTAAGGATGGCTA - 3'

LK256-RT 5'- TATTGCAGCAGTACAAGCGCAGTAAGGATGGCTA - 3'

E-RT 5' –TATTGCAGCAGTAC-3'

Figure 9: DNA Sequence Alignment of Coronavirus *M* Genes Isolated from different Species

```
TGE          ------------------------------------------------------------
PRCoV        ------------------------------------------------------------
CcoV         ------------------------------------------------------------
FCoV         ATGCATATGATGCCTATAAGACCTTTATGCAAACCAAGGCATATAATCCCGACGAAGCAT
HCoVOC43     ------------------------------------------------------------
PEDV         ------------------------------------------------------------
BCoV         ------------------------------------------------------------
HEV          ------------------------------------------------------------
MHV          ------------------------------------------------------------
RtCoV        ------------------------------------------------------------
SARS-Urb     ------------------------------------------------------------
SARS-Tor     ------------------------------------------------------------
IBV          ------------------------------------------------------------

TGE          ------------------------ATGAAGATTTTGTTAATATTAGCGTGTGTGATTGCA
PRCoV        ------------------------ATGAAGATTTTGTTGATATTAGCGTGTGCGATTGCA
CcoV         ----------------------ATGAAGAAAATTTTGTTTTTACTAGCGTGTGCAATTGCA
FCoV         TTTTGGTTTGAACTAAACAAAATGAAGTACATTTTGCTAATACTCGCGTGCATAATTGCA
HCoVOC43     ------------------------------------------------------------
PEDV         ------------------------------------------------------------
BCoV         ------------------------------------------------------------
HEV          ------------------------------------------------------------
MHV          ------------------------------------------------------------
RtCoV        ------------------------------------------------------------
SARS-Urb     ------------------------------------------------------------
SARS-Tor     ------------------------------------------------------------
IBV          ------------------------------------------------------------

TGE          TGCGCATGTGGAGAACGCTATTGTGCTATGAAATCCGATACAGATTTGTCATGTCGCAAT
PRCoV        TGCACATGTGGAGAACGCTATTGTGCTATGAAAGACGATACAGGTTTGTCATGTCGCAAT
CcoV         TGCGTCTATGGAGAACGCTATTGTGCCATGA---CTGAAAGTTCTACGTCATGTCGTAAT
FCoV         TGCGTTTATGGTGAACGCTACTGTGCCATGCAA---GACAGTGGCTTGCAGTGTATTAAT
HCoVOC43     ------------------------------------------------------------
PEDV         ------------------------------------------------------------
BCoV         ----------------------------------------------------ATGAGTAGTG
HEV          ----------------------------------------------------ATGAGTAGTC
MHV          ----------------------------------------------------ATGACTAGTA
RtCoV        ----------------------------------------------------ATGAGTAGTA
SARS-Urb     ------------------------------------------------------------
SARS-Tor     ------------------------------------------------------------
IBV          ------------------------------------------------------------

TGE          AGTACAGCGTCTGATTGTGAGTCA--TGCTTCAACGGAGGCGATCTTATTTGGCATCTTG
PRCoV        GGCACGGCGTCTGATTGTGAGTCA--TGCTTCAACAGAGGCGATCTTATTTGGCTTCTTG
CcoV         AGCACGGCTGGCAACTGTGCTTCA--TGCTTCGAAACAGGTGATCTTATTTGGCATCTTG
FCoV         GGCACAAATTCAAGATGTCAAACC--TGCTTTGAACGTGGTGATCTTATTTGGCATCTTG
HCoVOC43     --------------ATGTCAAATG---ACAATTGTACGGGTGACATTGTCACCCATTTGA
PEDV         --------------ATGTCTAACGGTTCTATTCCCGTTGATGAGGTGATTGAACACCTTA
BCoV         TAACT---ACACCAGCACCAGTTTACACCTGGACTGCTGATGAAGCTATTAAATTCCTAA
HEV          CAACT---ACACCAGTACCAGTTATTAGCTGGACTGCTGATGAAGCTATTAAATTCCTAA
MHV          CCACTCAGGCTCCACAGCCTGTTTATCAGTGGACGGCTGATGAGGCAATTCGATTCCTTA
RtCoV        CCACTCCAGCCCCCCAGACTGTCTATCAATGGACGGCCGATGTGGCAGTTCGATTCCTTA
SARS-Urb     --TTGCTTATCATGGCAGACAACGGTACTATTACCGTTGAGGAGCTTAAACAACTCCTGG
SARS-Tor     --TTGCTTATCATGGCAGACAACGGTACTATTACCGTTGAGGAGCTTAAACAACTCCTGG
```

```
IBV             -----ATGCCCAACGAGACAAATTGTACTCTTGACTTTGAACAGTCAGTTCAGCTTTTTA
                                                     *                *

TGE             CAAACTGGAACTTCAGCTGGTCTATAATATTGATCGTTTTTTATAACTGTGCTACAATATG
PRCoV           CAAACTGGAACTTCAGCTGGTCTATAATATTGATCATTTTTATTACTGTGCTACAATATG
CcoV            CAAACTGGAACTTCAGCTGGTCTGTAATATTGATCATTTTTATAACAGTGTTACAATATG
FCoV            CTAACTGGAACTTCAGCTGGTCTGTAATATTGATTGTTTTTATAACAGTGTTACAATATG
HCoVOC43        AGAATTGGAATTTTGGTTGGAATGTTATTCTAACCATATTCATTGTTATTCTTCAGTTTG
PEDV            GAAACTGGAATTTCACATGGAATATCATACTGACGATACTACTTGTAGTGCTTCAGTATG
BCoV            AGGAATGGAACTTTTCTTTGGGTATTATACTACTTTTTATTACAATCATATTGCAATTTG
HEV             AGGAATGGAATTTTTCTTTGGGTATAATAGTACTCTTTTATCACAATCATACTTCAATTTG
MHV             AGGAATGGAATTTCTCTCTCGGCATTATACTACTTTTTGTTACTATCATACTACAGTTCG
RtCoV           AGGAATGGAACTTCTTGTTGGGCATTATACTACTCTTTATTACTATCATACTACAGTTCG
SARS-Urb        AACAATGGAACCTAGTAATAGGTTTCCTATTCCTAGCCTGGATTATGTTACTACAATTTG
SARS-Tor        AACAATGGAACCTAGTAATAGGTTTCCTATTCCTAGCCTGGATTATGTTACTACAATTTG
IBV             AAGAGTATAATTTATTTATAACTGCATTCTTGTTGTTCTTAACCATAATACTTCAGTATG
                   * *  **  *                *  *                  *  * ** *  *

TGE             GAAGACCTCAATTCAGCTGGTTCGTGTATGGCATTAAAATGCTTATAATGTGGCTATTAT
PRCoV           GAAGACCTCAATTCAGCTGGTTCGTGTATGGCATTAAAATGCTTATAATGTGGCTATTAT
CcoV            GAAGACCTCAATTTAGCTGGTTCGTGTGTGGCATTAAAATGCTTATTATGTGGCTGTTAT
FCoV            GCAGACCACAATTTAGCTGGCTCGTTTATGGCATTAAAATGCTGATCATGTGGCTATTAT
HCoVOC43        GACACTATAAATACTCCAGATTGTTTTATGGTTTGAAGATGCTTGTACTGTGGCTTCTTT
PEDV            GCCATTACAAGTACTCTGTGTTCTTGTATGGTGTCAAGATGGCTATTCTATGGATACTTT
BCoV            GATATACAAGTCGCAGTATGTTTGTTTATGTTATTAAGATGATCATTTTGTGGCTTATGT
HEV             GATATACAAGTCGCAGTATGTTTGTTTATGTTATTAAGATGGTTATTCTGTGGCTCATGT
MHV             GTTACACGAGCCGTAGCATGTTTGTTTATGTTGTGAAAATGATACTTTTGTGGCTTATGT
RtCoV           GTTACACGAGCCGTAGCATGTTTATATATGTTGTGAAAATGATAATCTTGTGGTTAATGT
SARS-Urb        CCTATTCTAATCGGAACAGGTTTTTGTACATAATAAAGCTTGTTTTCCTCTGGCTCTTGT
SARS-Tor        CCTATTCTAATCGGAACAGGTTTTTGTACATAATAAAGCTTGTTTTCCTCTGGCTCTTGT
IBV             GCTATGCAACAAGAAGTAAGGTTATTTATACACTGAAAATGATAGTGTTATGGTGCTTTT
                                  *  * *      * **  *     *  * ***    * *

TGE             GGCCCGTTGTTTTGGCTCTTACGATTTTTAATGCATACTCGGAATACCAAGTGTCCAGAT
PRCoV           GGCCGATTGTTTTGGCTCTTACGATTTTTAATGCATACTCGGAATACCAAGTGTCCAGAT
CcoV            GGCCCATTGTTTTAGCTCTTACGATTTTTAATGCATACCTGGAATACCAAGTGTCCAGGT
FCoV            GGCCTATTGTTCTAGCGCTTACGATTTTTAATGCATACTCTGAGTACCAAGTTTCCAGAT
HCoVOC43        GGCCACTCGTACTTGCTTTGTCAATCTTTGACAC---CTGGGCTAATTGGGATTCTAATT
PEDV            GGCCTCTTGTGTTGGCACTGTCACTTTTTGACGCATGGGCTAGCTTCCAGGT---CAACT
BCoV            GGCCCCTTACTATCATCTTAACTATTTTCAAT------TGCGTGTAT---GCGTTGAATA
HEV             GGCCTCTTACTATAATTTTAACTATCTTCAAC------TGCGTATAC---GCGTTGAATA
MHV             GGCCACTAACTATTGTTTTGTGTATTTTTAAC------TGCGTCTAT---GCGCTAAATA
RtCoV           GGCCACTGACTATTGTTTTGTGTATTTTTAAT------TGCGTGTAT---GCGCTAAATA
SARS-Urb        GGCCAGTAACACTTGCTTGTTTTGTGCTTGCTG------CTGTCTACAGAAT---TAATT
SARS-Tor        GGCCAGTAACACTTGCTTGTTTTGTGCTTGCTG------CTGTCTACAGAAT---TAATT
IBV             GGCCCCTTAACATTGCAGTAGGTGTAATTTCA------TGTACATACC---CACCAAACA
                ****  *     *           *  *                              *

TGE             ATGTAATGTTCGGCTTTAGTATTGCAGGTGCAATTGTTACATTTGTACTCTGGATTATGT
PRCoV           ATGTAATGTTCGGCTTTAGTATTGCAGGTGCAATTGTTACATTTGTACTCTGGATTATGT
CcoV            ATGTAATGTTCGGCTTTAGTGTTGCAGGTGCAACTGTTACATTTATACTTTGGATTATGT
FCoV            ATGTAATGTTCGGCTTTAGTGTTGCAGGTGCAGTTGTAACGTTTGCACTTTGGATGATGT
HCoVOC43        GGGCCTTTGTTGCATTTAGCTTTTTTATGGCCGTATCAACACTCGTTATGTGGGTGATGT
PEDV            GGGTCTTTTTTCGCTTTCAGCATCCTTATGGCTTGCATCACTCTTATGCTGTGGATAATGT
BCoV            ATGTGTATCTTGGCTTTTCTATAGTTTTCACTATAGTGGCCATTATCATGTGGATTGTGT
HEV             ATGTGTACCTTGGCTTCTCTATAGTTTTTACTATAGTGGCCATTATTATGTGGGTTGTTT
MHV             ATGTGTATCTTGGATTTTCTATAGTGTTTACTATAGTGTCCATTATAATGTGGATTATGT
RtCoV           ATGTGTATCTTGGATTTTCTATAGTGTTTACTATAGTGTCCATTGTAATGTGGATTATGT
SARS-Urb        GGGTGACTGGCGGGATTGCGATTGCAATGGCTTGTATTGTAGGCTTGATGTGGCTTAGCT
SARS-Tor        GGGTGACTGGCGGGATTGCGATTGCAATGGCTTGTATTGTAGGCTTGATGTGGCTTAGCT
IBV             CAGGAGGTCTTGTCGCAGCGATAATACTTACAGTGTTTGCGTGTCTGTCTTTTGTAGGTT
```

```
                    *         *         *         *                       *   *   *

TGE         ATTTTGTAAGATCCATTCAGTTGTACAGAAGGACTAAGTCTTGGTGGTCTTTCAACCCTG
PRCoV       ATTTTGTAAGATCCATTCAGTTGTACAGAAGGACTAAGTCTTGGTGGTCCTTCAACCCTG
CcoV        ATTTTGTTAGATCCATTCAGTTATACAGAAGGACTAAGTCTTGGTGGTCTTTCAACCCTG
FCoV        ATTTTGTGAGATCTGTTCAGCTATATAGAAGAACCAAATCATGGTGGTCTTTTAATCCTG
HCoVOC43    ACTTCGCAAACAGTTTCAGACTTTTCCGACGTGCTCGAACTTTTTGGGCATGGAATCCTG
PEDV        ATTTTGTCAATAGCATTCGGTTGTGGCGCAGGACACATTCTTGGTGGTCTTTCAATCCTG
BCoV        ATTTTGTGAATAGTATCAGGTTGTTTATTAGAACTGGAAGTTGGTGGAGTTTCAACCCAG
HEV         ATTTTGTGAATAGTATCAGGTTGTTTATTAGAACTGGAAGTTGGTGGAGTTTCAACCCAG
MHV         ATTTTGTTAATAGCATCAGGTTGTTTATTAGAACTGGAAGTTGGTGGAGTTTCAACCCAG
RtCoV       ATTTTGTTAATAGCATAAGGTTGTTTATCAGGACTGGCAGCTGGTGGAGCTTCAACCCCG
SARS-Urb    ACTTCGTTGCTTCCTTCAGGCTGTTTGCTCGTACCCGCTCAATGTGGTCATTCAACCCTG
SARS-Tor    ACTTCGTTGCTTCCTTCAGGCTGTTTGCTCGTACCCGCTCAATGTGGTCATTCAACCCAG
IBV         ATTGGATCCAGAGTATTAGACTCTTTAAGCGGTGTAGGTCATGGTGGTCATTTAATCCAG
            * *             *     * *     *              ***   *  ** ** *

TGE         AAACTAAAGCAATTCTTTGCGTTAGTGCATTAGGAAGAAGCTATGTGCTTCCTCTCGAAG
PRCoV       AAACTAACGCAATTCTTTGCGTTAGTGCATTAGGAAGAAGCTATGTGCTTCCTCTCGAAG
CcoV        AAACTAGCGCAATTCTTTGCGTTAGTGCGTTAGGAAGAAGCTATGTGCTTCCTCTTGAAG
FCoV        AGACTAATGCAATTCTTTGTGTTAATGCATTGGGTAGAAGTTATGTGCTTCCCTTAGATG
HCoVOC43    AGGTTAATGCAATCACTGTCACAACCGTGTTGGGACAGACATACTATCAACCCATTCAAC
PEDV        AAACTGACGCGCTTCTCACTACTTCTGTGATGGGCCGACAGGTCTGCATTCCAGTGCTTG
BCoV        AAACAAACAACTTGATGTGTATAGATATGAAGGGAAGGATGTATGTTAGGCCGATAATTG
HEV         AAACAAACAACTTGATGTGTATAGATATGAAGGGAAGAATGTATGTTAGGCCGATTATTG
MHV         AAACAAACAACCTAATGTGTATAGATATGAAAGGTACTGTGTATGTTAGACCCATTATAG
RtCoV       AAACAAACAACCTAATGTGTATAGATGTGAAAGGTACTGTGTATGTTAGACCCATTATTG
SARS-Urb    AAACAAACATTCTTCTCAATGTGCCTCTCCGGGGGACAATTGTGACCAGACCGCTCATGG
SARS-Tor    AAACAAACATTCTTCTCAATGTGCCTCTCCGGGGGACAATTGTGACCAGACCGCTCATGG
IBV         AATCTAATGCCGTAGGTTCAATACTCCTAACTAATGGTCAACAATGTAATTTTGCTATAG
            *           *

TGE         GTGTG---CCAACTGGTGTCACTCTAACTTTGCTTTCAGGGAATTTGTACGCTGAAGGGT
PRCoV       GTGTG---CCAACTGGTGTCACTCTAACTTTGCTTTCAGGGAATTTGTACGCTGAAGGGT
CcoV        GTGTG---CCAACTGGTGTCACTCTAACATTGCTTTCAGGGAATTTGTGTGCTGAAGGGT
FCoV        GTACT---CCTACAGGTGTTACCCTTACTCTACTTTCAGGAAATCTATATGCTGAAGGTT
HCoVOC43    AAGCT---CCAACAGGCATTACTGTGACCTTGCTGAGCGGCGTGCTTTACGTTGACGGAC
PEDV        GAGCA---CCAACTGGTGTAACGCTAACACTCCTTAGTGGTACATTGCTTGTAGAGGGCT
BCoV        AGGAC---TACCATACCCTTACGGTCACAATAATACGTGGTCATCTTTACATGCAAGGTA
HEV         AGGAC---TACCACACCCTTACTGCCACAATAATACGTGGCCACCTCTACATCCAAGGTA
MHV         AGGAT---TACCATACACTAACAGCCACTATCATTCGTGGTCACCTCTATATGCAAGGTG
RtCoV       AAGAT---TACCATACACTAACAGCCACAAATGTACGTGGCCACCTTTATATGCAAGGTG
SARS-Urb    AAAGT---GAACTTGTCATTGGTGCTGTGATCATTCGTGGTCACTTGCGAATGGCCGGAC
SARS-Tor    AAAGT---GAACTTGTCATTGGTGCTGTGATCATTCGTGGTCACTTGCGAATGGCCGGAC
IBV         AGAGTGTGCCAATGGTGCTTTCTCCAATTATAAAGAATGGTGTTCTTTATTGTGAGGGTC
                              *                    **     *          **

TGE         TCAAAATTGCAGGTGGTATGAACATCGACAATTTACCAAAATACGTAATGGTTGCATTAC
PRCoV       TCAAAATTGCAGGTGGTATGACCATCGACAATTTGCCAAAATACGTAATGGTTGCATTAC
CcoV        TCAAAATTGCAGGTGGTATGAACATCGACAATTTACCAAAATATGTAATGGTTGCATTAC
FCoV        TCAAAATGGCTGGTGGTTTAACCATCGAGCATTTGCCTAAATACGTCATGATTGCTACAC
HCoVOC43    ATAGATTGGCTTCAGGTGTTCAGGTTCATAACCTACCTGAATACATGACAGTTGCCGTGC
PEDV        ATAAGGTTGCTACTGGCGTACAGGTAAGTCAATTACCTAATTTCGTCACAGTCGCCAAGG
BCoV        TAAAACTAGGTACTGGCTATTCTTTGTCAGATTTGCCAGCTTATGTGACTGTTGCTAAGG
HEV         TAAAACTAGGTACTGGCTATTCTTTGTCAGATTTGCCTGCTTATGTGACCGTTGCTAAGG
MHV         TTAAGCTAGGCACTGGCTTCTCTTTGTCTGATTTGCCTGCTTATGTTACAGTTGCTAAGG
RtCoV       TTAAGCTAGGCACTGGCTTCTCTTTGTCTGATTTGCCCGCTTATGTTACAGTTGCTAAGG
SARS-Urb    ACCCCCTAGGGCGCTGTG---ACATTAAGGACCTGCCAAAAGAGATCACTGTGGCTACAT
SARS-Tor    ACTCCCTAGGGCGCTGTG---ACATTAAGGACCTGCCAAAAGAGATCACTGTGGCTACAT
IBV         AGTGGCTTGCTAAGTGTG---AACCAGACCACTTGCCTAAAGATATATTTGTTTGTACAC
                  * *       *                 *  * **         *      *
```

```
TGE             CTAGCAGGACTATTGTCTACACAC--TTGTTGGCAAGAAGTTGAAAGCAAGTAGTGCGAC
PRCoV           CCAGCAGGACTATTGTTTACACAC--TTGTTGGCAAGAAGTTGAAAGCAAGTAGTGCGAC
CcoV            CTGTCAGAACCATAGTCTACACAC--TTGTTGGCAAGAAATTGAAAGCAAGTAGTGCAAC
FCoV            CTAGTAGAACCATCGTTTATACAT--TAGTTGGAAAACAATTAAAAGCAACTACTGCCAC
HCoVOC43        CGAGCACTACTATAATTTATAGTA--GAGTCGGAAGGTCCGTAAATTCACAAAATAGCAC
PEDV            CCACTACAACAATTGTCTACGGAC--GTGTTGGTCGTTCAGTCAATGCTTCATCTGGCAC
BCoV            T--CTCACACCTGCTCACGTATAA--GCGT-GGTTTTCTTGACAAGATAGGCGATACTAG
HEV             T--TACACACCTGTGCACATATAA--GCGT-GGTTTTCTTGATAGGATAGGCGATACTAG
MHV             T--GTCTCACCTTTGCACTTATAA--GCGC-GCATTCTTAGACAAGGTAGACGGTGTTAG
RtCoV           T--GTCGCACCTTTGCACTTATAA--GCGC-GCATTTTTAGACAAGGTTGACGGTGTTAG
SARS-Urb        C---ACGAACGCTTTCTTATTACA--AATTAGGAGCGTCGCAGCGTGTAGGCACTGATTC
SARS-Tor        C---ACGAACGCTTTCTTATTACA--AATTAGGAGCGTCGCAGCGTGTAGGCACTGATTC
IBV             CGGATAGACGTAATATCTACCGTATGGTGCAGAAATATACTGGTGACCAAAGCGGAAATA
                                                *

TGE             TGGATGGGCTTACTATGTAAAATCTAAAGCTGGTGATTACTCAACAGAGG---CAAGAAC
PRCoV           TGGATGGGCTTACTATGTAAAATCTAAAGCTGGTGATTACTCAACAGAGG---CAAGAAC
CcoV            AGGATGGGCTTACTATGTAAAGTCTAAAGCTGGTGATTACTCAACAGATG---CACGAAC
FCoV            AGGATGGGCTTACTACGTAAAATCTAAAGCTGGTGATTACTCAACAGAAG---CACGTAC
HCoVOC43        AGGCTGGGTTTTCTACGTACGAGTAAAACACGGTGATTTTTCTGCAGTGAGCTCTCCCAT
PEDV            TGGTTGGGCTTTCTATGTCCGGTCAAAACACGGCGACTATTCAGCTGTGAGTAATCCGAG
BCoV            TGGTTTTGCTGTTTATGTTAAGTCCAAAGTCGGTAATTACCGACTGCCATCAACCCAAAA
HEV             TGGTTTTGCTGTTTATGTTAAGTCCAAAGTCGGTAATTATCGATTGCCTTCAACCCATAA
MHV             CGGTTTTGCTGTTTATGTGAAGTCCAAGGTCGGAAATTACCGACTGCCCTCAAATAAACC
RtCoV           CGGTTTTGCTGTTTATGTGAAGTCCAAGGTCGGTAATTACCGACTGCCCTCAAATAAACC
SARS-Urb        AGGTTTTGCTGCATACAACCGCTACCGTATTGGAAACTATAAATTAAATACAGACCACGC
SARS-Tor        AGGTTTTGCTGCATACAACCGCTACCGTATTGGAAACTATAAATTAAATACAGACCACGC
IBV             AGAAAAGGTTTGCTACGTTTG-TCTATGCAAAGCAGTCAGTAGATACTGGCGAGCTAGAA
                 *     * *   **                  *

TGE             TGATAATTTGAGTGAGCAAGAAAAATTATTACATATGGTATAA
PRCoV           TGATAATTTGAGTGAGCAAGAAAAATTATTACATATGGTATAA
CcoV            TGATAATTTGAGTGAGCATGAAAAATTATTACATATGGTATAA
FCoV            TGACAATTTGAGTGAACATGAAAAATTATTACATATGGTGTAA
HCoVOC43        GAGCAACATGACAGAAAACGAAAGATTGCTTCATTTTTTCTAA
PEDV            TGCGGTTCTCACAGATAGTGAGAAAGTGCTTCATTTAGTCTAA
BCoV            GGGTTCTGGCATGGACACCGCATTGTTGAGAAATAATATCTAA
HEV             GGGCTCAGGCATGGACACCGCATTGTTGAGAAATAATATCTAA
MHV             GAGT---GGCATGGACACCGCATTGTTGAGAATCTAA------
RtCoV           GAGT---GGCGCGGACACCGCATTGTTGAGAATCTAA------
SARS-Urb        CGGTAGCAACGACAATATTGCTTTGCTAGTACAGTAA------
SARS-Tor        CGGTAGCAACGACAATATTGCTTTGCTAGTACAGTAA------
IBV             AGTGTAGCAACAGGAGGAAGTAGTCTTTACACATAA-------
                          *    *      *
```

Figure 10: Phylogenetic Analysis of *M* Gene

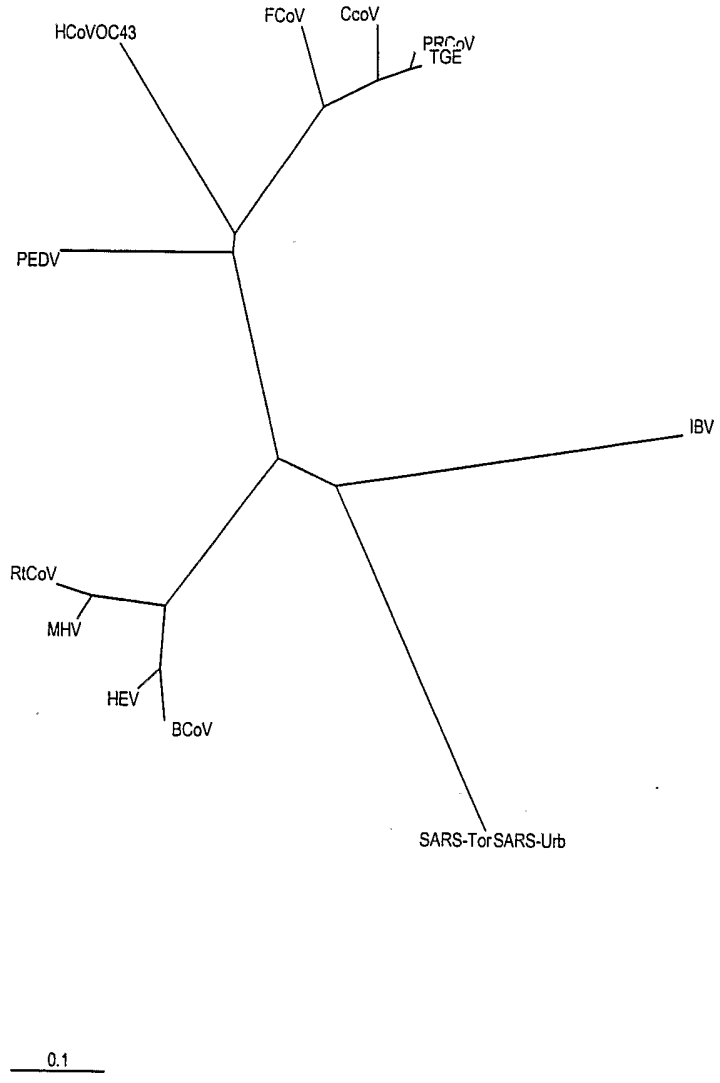
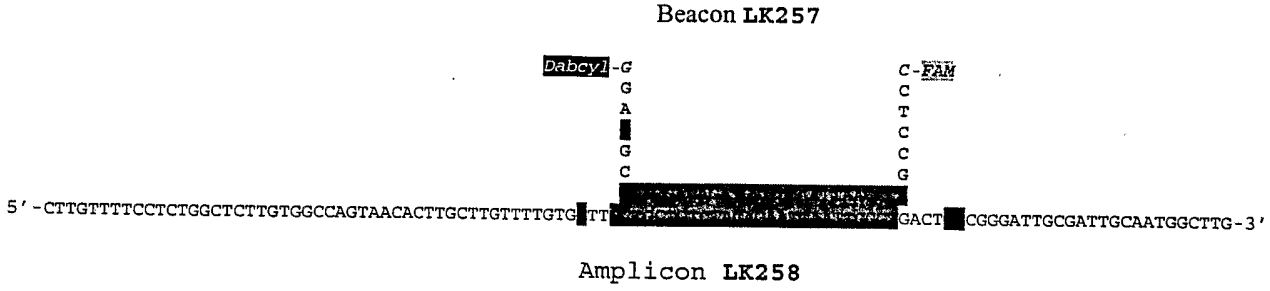


Figure 11: Molecular Designs for SARS-Associated *M* Gene

NOTE:	LK257 is a TET-molecular beacon, which recognizes the ***M* gene of coronavirus (SARS Tor2 and SARS urbani human pathogenic strains)**. It will be used in a real-time PCR diagnostic for the identification of **SARS-associated coronavirus RNA/DNA.**



LK257

I.	Target recognition sequence:	**26 nucleotides (11 G/C) + (3 nts arms G/C)**
J.	Length of the arms:	**6 nucleotides (5 G/C)**
K.	Melting temperature of the beacon:	**dG = -1.52 dH = -46.6 dS = -139.6 Tm = 60.7 °C**
L.	Melting temperature of target:	**°C (including 3 nts from the arms)**

LK259 **5'-CTTGTTTTCCTCTGGCTCTTG-3'**

Length: 21 nucleotides (10 G/C)
Tm: °C
Position: (see alignment below)

LK260 **5' –CAAGCCATTGCAATCGCAATC-3'**

Length: 21 nucleotides (10 G/C)
Tm: °C
Position: (see alignment below)

M-RT **5'-AAGCAACGAAGTAG -3'**

Length: 14 nucleotides (6 G/C)
Tm: 47 °C
Position: (see alignment below)

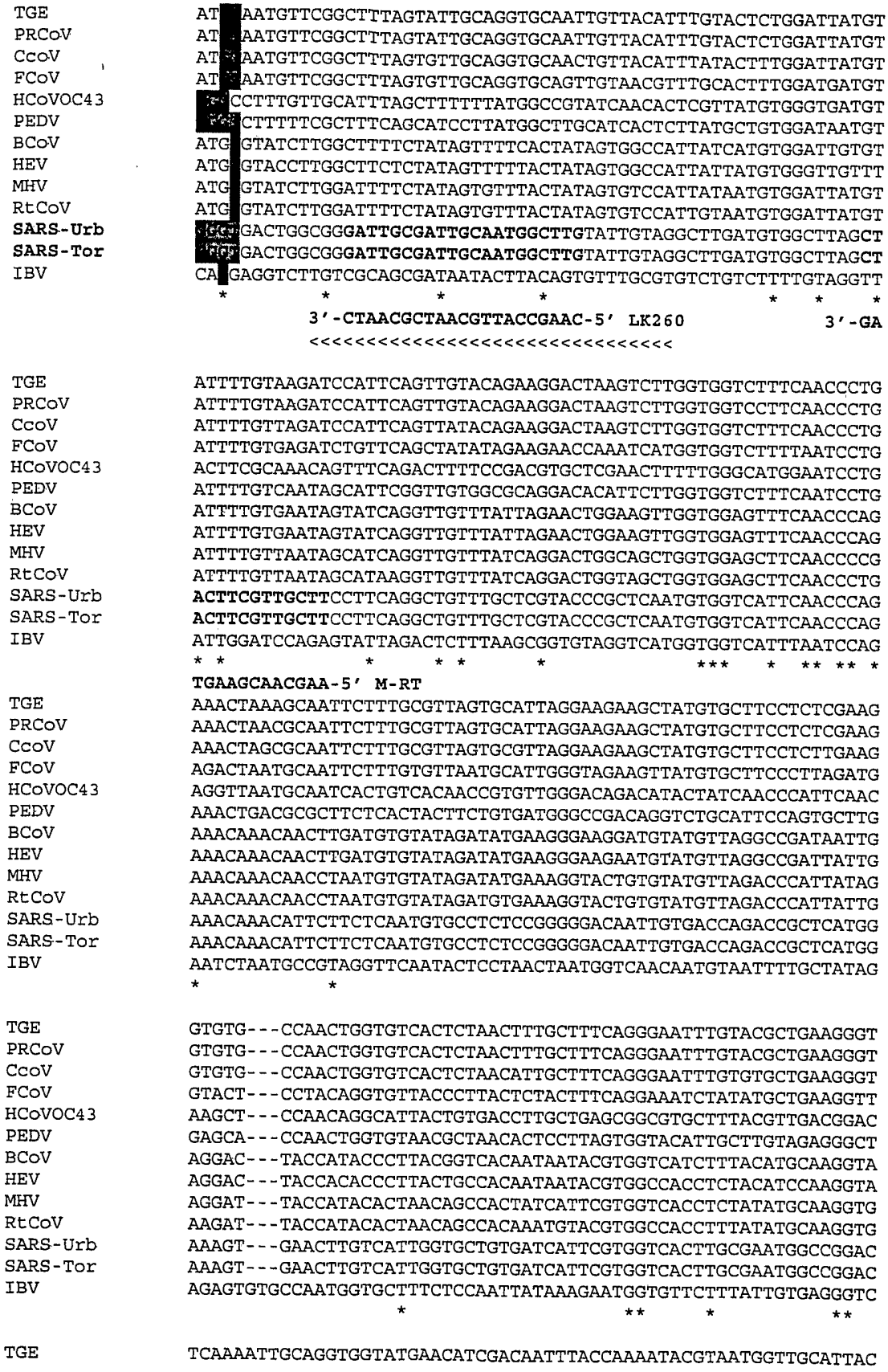
LK258 Amplicon (107 nucleotides)

5'–
CTTGTTTTCCTCTGGCTCTTGTGGCCAGTAACACTTGCTTGTTTTGTGCTTGCTGCTG
TCTACAGAATTAATTGGGTGACTGGCGGGATTGCGATTGCAATGGCTTG -3'

DNA sequence alignment of complete M genes from corona virus strains

```
TGE          ------------------------------------------------------------
PRCoV        ------------------------------------------------------------
CcoV         ------------------------------------------------------------
FCoV         ATGCATATGATGCCTATAAGACCTTTATGCAAACCAAGGCATATAATCCCGACGAAGCAT
HCoVOC43     ------------------------------------------------------------
PEDV         ------------------------------------------------------------
BCoV         ------------------------------------------------------------
HEV          ------------------------------------------------------------
MHV          ------------------------------------------------------------
RtCoV        ------------------------------------------------------------
SARS-Urb     ------------------------------------------------------------
SARS-Tor     ------------------------------------------------------------
IBV          ------------------------------------------------------------

TGE          ------------------------ATGAAGATTTTGTTAATATTAGCGTGTGTGATTGCA
PRCoV        ------------------------ATGAAGATTTTGTTGATATTAGCGTGTGCGATTGCA
CcoV         ----------------------ATGAAGAAAATTTTGTTTTTACTAGCGTGTGCAATTGCA
FCoV         TTTTGGTTTGAACTAAACAAAATGAAGTACATTTTGCTAATACTCGCGTGCATAATTGCA
HCoVOC43     ------------------------------------------------------------
PEDV         ------------------------------------------------------------
BCoV         ------------------------------------------------------------
HEV          ------------------------------------------------------------
MHV          ------------------------------------------------------------
RtCoV        ------------------------------------------------------------
SARS-Urb     ------------------------------------------------------------
SARS-Tor     ------------------------------------------------------------
IBV          ------------------------------------------------------------

TGE          TGCGCATGTGGAGAACGCTATTGTGCTATGAAATCCGATACAGATTTGTCATGTCGCAAT
PRCoV        TGCACATGTGGAGAACGCTATTGTGCTATGAAAGACGATACAGGTTTGTCATGTCGCAAT
CcoV         TGCGTCTATGGAGAACGCTATTGTGCCATGA---CTGAAAGTTCTACGTCATGTCGTAAT
FCoV         TGCGTTTATGGTGAACGCTACTGTGCCATGCAA---GACAGTGGCTTGCAGTGTATTAAT
HCoVOC43     ------------------------------------------------------------
PEDV         ------------------------------------------------------------
BCoV         ----------------------------------------------------ATGAGTAGTG
HEV          ----------------------------------------------------ATGAGTAGTC
MHV          ----------------------------------------------------ATGACTAGTA
RtCoV        ----------------------------------------------------ATGAGTAGTA
SARS-Urb     ------------------------------------------------------------
SARS-Tor     ------------------------------------------------------------
IBV          ------------------------------------------------------------

TGE          AGTACAGCGTCTGATTGTGAGTCA--TGCTTCAACGGAGGCGATCTTATTTGGCATCTTG
PRCoV        GGCACGGCGTCTGATTGTGAGTCA--TGCTTCAACAGAGGCGATCTTATTTGGCTTCTTG
```

```
CcoV          AGCACGGCTGGCAACTGTGCTTCA--TGCTTCGAAACAGGTGATCTTATTTGGCATCTTG
FCoV          GGCACAAATTCAAGATGTCAAACC--TGCTTTGAACGTGGTGATCTTATTTGGCATCTTG
HCoVOC43      --------------ATGTCAAATG---ACAATTGTACGGGTGACATTGTCACCCATTTGA
PEDV          --------------ATGTCTAACGGTTCTATTCCCGTTGATGAGGTGATTGAACACCTTA
BCoV          TAACT---ACACCAGCACCAGTTTACACCTGGACTGCTGATGAAGCTATTAAATTCCTAA
HEV           CAACT---ACACCAGTACCAGTTATTAGCTGGACTGCTGATGAAGCTATTAAATTCCTAA
MHV           CCACTCAGGCTCCACAGCCTGTTTATCAGTGGACGGCTGATGAGGCAATTCGATTCCTTA
RtCoV         CCACTCCAGCCCCCCAGACTGTCTATCAATGGACGGCCGATGTGGCAGTTCGATTCCTTA
SARS-Urb      --TTGCTTATCATGGCAGACAACGGTACTATTACCGTTGAGGAGCTTAAACAACTCCTGG
SARS-Tor      --TTGCTTATCATGGCAGACAACGGTACTATTACCGTTGAGGAGCTTAAACAACTCCTGG
IBV           -----ATGCCCAACGAGACAAATTGTACTCTTGACTTTGAACAGTCAGTTCAGCTTTTTA
                                                    *                 *

TGE           CAAACTGGAACTTCAGCTGGTCTATAATATTGATCGTTTTTTATAACTGTGCTACAATATG
PRCoV         CAAACTGGAACTTCAGCTGGTCTATAATATTGATCATTTTTTATTACTGTGCTACAATATG
CcoV          CAAACTGGAACTTCAGCTGGTCTGTAATATTGATCATTTTTTATAACAGTGTTACAATATG
FCoV          CTAACTGGAACTTCAGCTGGTCTGTAATATTGATTGTTTTTTATAACAGTGTTACAATATG
HCoVOC43      AGAATTGGAATTTTGGTTGGAATGTTATTCTAACCATATTCATTGTTATTCTTCAGTTTG
PEDV          GAAACTGGAATTTCACATGGAATATCATACTGACGATACTACTTGTAGTGCTTCAGTATG
BCoV          AGGAATGGAACTTTTCTTTGGGTATTATACTACTTTTTATTACAATCATATTGCAATTTG
HEV           AGGAATGGAATTTTTCTTTGGGTATAATAGTACTCTTTATCACAATCATACTTCAATTTG
MHV           AGGAATGGAATTTCTCTCTCGGCATTATACTACTTTTTGTTACTATCATACTACAGTTCG
RtCoV         AGGAATGGAACTTCTTGTTGGGCATTATACTACTCTTTATTACTATCATACTACAGTTCG
SARS-Urb      AACAATGGAACCTAGTAATAGGTTTCCTATTCCTAGCCTGGATTATGTTACTACAATTTG
SARS-Tor      AACAATGGAACCTAGTAATAGGTTTCCTATTCCTAGCCTGGATTATGTTACTACAATTTG
IBV           AAGAGTATAATTTATTTATAACTGCATTCTTGTTGTTCTTAACCATAATACTTCAGTATG
                 * *  **  *              *  *                 *  * ** *  *

TGE           GAAGACCTCAATTCAGCTGGTTCGTGTATGGCATTAAAATGCTTATAATGTGGCTATTAT
PRCoV         GAAGACCTCAATTCAGCTGGTTCGTGTATGGCATTAAAATGCTTATAATGTGGCTATTAT
CcoV          GAAGACCTCAATTTAGCTGGTTCGTGTGTGGCATTAAAATGCTTATTATGTGGCTGTTAT
FCoV          GCAGACCACAATTTAGCTGGCTCGTTTATGGCATTAAAATGCTGATCATGTGGCTATTAT
HCoVOC43      GACACTATAAATACTCCAGATTGTTTTATGGTTTGAAGATGCTTGTACTGTGGCTTCTTT
PEDV          GCCATTACAAGTACTCTGTGTTCTTGTATGGTGTCAAGATGGCTATTCTATGGATACTTT
BCoV          GATATACAAGTCGCAGTATGTTTGTTTATGTTATTAAGATGATCATTTTGTGGCTTATGT
HEV           GATATACAAGTCGCAGTATGTTTGTTTATGTTATTAAGATGGTTATTCTGTGGCTCATGT
MHV           GTTACACGAGCCGTAGCATGTTTGTTTATGTTGTGAAAATGATACTTTTGTGGCTTATGT
RtCoV         GTTACACGAGCCGTAGCATGTTTATATATGTTGTGAAAATGATAATCTTGTGGTTAATGT
SARS-Urb      CCTATTCTAATCGGAACAGGTTTTTGTACATAATAAAGCTTGTTTTCCTCTGGCTCTTGT
SARS-Tor      CCTATTCTAATCGGAACAGGTTTTTGTACATAATAAAGCTTGTTTTCCTCTGGCTCTTGT
IBV           GCTATGCAACAAGAAGTAAGGTTATTTATACACTGAAAATGATAGTGTTATGGTGCTTTT
                              *  * *      * **  *      *  * ***    * *
                                                  LK259 5'-CTTGTTTTCCTCTGGCTCTTG-3'
                                                  >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>

TGE           GGCCCGTTGTTTTGGCTCTTACGATTTTTAA GCATACT G AA  CA G GTCC GA
PRCoV         GGCCGATTGTTTTGGCTCTTACGATTTTTAA GCATACTCG AA  CA G GTCC GG
CcoV          GGCCCATTGTTTTAGCTCTTACGATTTTTAA GCATACCTG AA  C  G TTCC GA
FCoV          GGCCTATTGTTCTAGCGCTTACGATTTTTAA GCATACT  AG  CA G TTCC GA
HCoVOC43      GGCCACTCGTACTTGCTTTGTCAATCTTT ACAC---CTGG CTAATT GGATTC
PEDV          GGCCTCTTGTGTTGGCACTGTCACTTTTT ACGCATGGG AG  T CAGG ---C  C
BCoV          GGCCCCTTACTATCATCTTAACTATTTTCAA------TGC G  T---GCGTTG  A
HEV           GGCCTCTTACTATAATTTTAACTATCTTCAAC------TGC A  ---GCGTTG  A
MHV           GGCCACTAACTATTGTTTTGTGTATTTTTAAC------TGC  T---GCGCTA  A
RtCoV         GGCCACTGACTATTGTTTTGTGTATTTTTAA------TGC G  T---GCGCTA  A
SARS-Urb      GGCCAGTAACACTTGCTTGTTTTGTGCTT
SARS-Tor      GGCCAGTAACACTTGCTTGTTTTGTGCTT ------
IBV           GGCCCCTTAACATTGCAGTAGGTGTAATTT A------TG ACA  C---CACCA  CA
              ****  *     *          * *                                *
                                                  LK258 BEACON
```


TGE
PRCoV
CcoV
FCoV
HCoVOC43
PEDV
BCoV
HEV
MHV
RtCoV
SARS-Urb
SARS-Tor
IBV
3'-CTAACGCTAACGTTACCGAAC-5' LK260
3'-GA
TGAAGCAACGAA-5' M-RT

```
PRCoV          TCAAAATTGCAGGTGGTATGACCATCGACAATTTGCCAAAATACGTAATGGTTGCATTAC
CcoV           TCAAAATTGCAGGTGGTATGAACATCGACAATTTACCAAAATATGTAATGGTTGCATTAC
FCoV           TCAAAATGGCTGGTGGTTTAACCATCGAGCATTTGCCTAAATACGTCATGATTGCTACAC
HCoVOC43       ATAGATTGGCTTCAGGTGTTCAGGTTCATAACCTACCTGAATACATGACAGTTGCCGTGC
PEDV           ATAAGGTTGCTACTGGCGTACAGGTAAGTCAATTACCTAATTTCGTCACAGTCGCCAAGG
BCoV           TAAAACTAGGTACTGGCTATTCTTTGTCAGATTTGCCAGCTTATGTGACTGTTGCTAAGG
HEV            TAAAACTAGGTACTGGCTATTCTTTGTCAGATTTGCCTGCTTATGTGACCGTTGCTAAGG
MHV            TTAAGCTAGGCACTGGCTTCTCTTTGTCTGATTTGCCTGCTTATGTTACAGTTGCTAAGG
RtCoV          TTAAGCTAGGCACTGGCTTCTCTTTGTCTGATTTGCCCGCTTATGTTACAGTTGCTAAGG
SARS-Urb       ACCCCCTAGGGCGCTGTG---ACATTAAGGACCTGCCAAAAGAGATCACTGTGGCTACAT
SARS-Tor       ACTCCCTAGGGCGCTGTG---ACATTAAGGACCTGCCAAAAGAGATCACTGTGGCTACAT
IBV            AGTGGCTTGCTAAGTGTG---AACCAGACCACTTGCCTAAAGATATATTTGTTTGTACAC
                     * *      *               *  * **        *     *

TGE            CTAGCAGGACTATTGTCTACACAC--TTGTTGGCAAGAAGTTGAAAGCAAGTAGTGCGAC
PRCoV          CCAGCAGGACTATTGTTTACACAC--TTGTTGGCAAGAAGTTGAAAGCAAGTAGTGCGAC
CcoV           CTGTCAGAACCATAGTCTACACAC--TTGTTGGCAAGAAATTGAAAGCAAGTAGTGCAAC
FCoV           CTAGTAGAACCATCGTTTATACAT--TAGTTGGAAAACAATTAAAAGCAACTACTGCCAC
HCoVOC43       CGAGCACTACTATAATTTATAGTA--GAGTCGGAAGGTCCGTAAATTCACAAAATAGCAC
PEDV           CCACTACAACAATTGTCTACGGAC--GTGTTGGTCGTTCAGTCAATGCTTCATCTGGCAC
BCoV           T--CTCACACCTGCTCACGTATAA--GCGT-GGTTTTCTTGACAAGATAGGCGATACTAG
HEV            T--TACACACCTGTGCACATATAA--GCGT-GGTTTTCTTGATAGGATAGGCGATACTAG
MHV            T--GTCTCACCTTTGCACTTATAA--GCGC-GCATTCTTAGACAAGGTAGACGGTGTTAG
RtCoV          T--GTCGCACCTTTGCACTTATAA--GCGC-GCATTTTTAGACAAGGTTGACGGTGTTAG
SARS-Urb       C---ACGAACGCTTTCTTATTACA--AATTAGGAGCGTCGCAGCGTGTAGGCACTGATTC
SARS-Tor       C---ACGAACGCTTTCTTATTACA--AATTAGGAGCGTCGCAGCGTGTAGGCACTGATTC
IBV            CGGATAGACGTAATATCTACCGTATGGTGCAGAAATATACTGGTGACCAAAGCGGAAATA
                                                 *

TGE            TGGATGGGCTTACTATGTAAAATCTAAAGCTGGTGATTACTCAACAGAGG---CAAGAAC
PRCoV          TGGATGGGCTTACTATGTAAAATCTAAAGCTGGTGATTACTCAACAGAGG---CAAGAAC
CcoV           AGGATGGGCTTACTATGTAAAGTCTAAAGCTGGTGATTACTCAACAGATG---CACGAAC
FCoV           AGGATGGGCTTACTACGTAAAATCTAAAGCTGGTGATTACTCAACAGAAG---CACGTAC
HCoVOC43       AGGCTGGGTTTTCTACGTACGAGTAAAACACGGTGATTTTTCTGCAGTGAGCTCTCCCAT
PEDV           TGGTTGGGCTTTCTATGTCCGGTCAAAACACGGCGACTATTCAGCTGTGAGTAATCCGAG
BCoV           TGGTTTTGCTGTTTATGTTAAGTCCAAAGTCGGTAATTACCGACTGCCATCAACCCAAAA
HEV            TGGTTTTGCTGTTTATGTTAAGTCCAAAGTCGGTAATTATCGATTGCCTTCAACCCATAA
MHV            CGGTTTTGCTGTTTATGTGAAGTCCAAGGTCGGAAATTACCGACTGCCCTCAAATAAACC
RtCoV          CGGTTTTGCTGTTTATGTGAAGTCCAAGGTCGGTAATTACCGACTGCCCTCAAATAAACC
SARS-Urb       AGGTTTTGCTGCATACAACCGCTACCGTATTGGAAACTATAAATTAAATACAGACCACGC
SARS-Tor       AGGTTTTGCTGCATACAACCGCTACCGTATTGGAAACTATAAATTAAATACAGACCACGC
IBV            AGAAAAGGTTTGCTACGTTTG-TCTATGCAAAGCAGTCAGTAGATACTGGCGAGCTAGAA
                *     * *   **                  *

TGE            TGATAATTTGAGTGAGCAAGAAAAATTATTACATATGGTATAA
PRCoV          TGATAATTTGAGTGAGCAAGAAAAATTATTACATATGGTATAA
CcoV           TGATAATTTGAGTGAGCATGAAAAATTATTACATATGGTATAA
FCoV           TGACAATTTGAGTGAACATGAAAAATTATTACATATGGTGTAA
HCoVOC43       GAGCAACATGACAGAAAACGAAAGATTGCTTCATTTTTTCTAA
PEDV           TGCGGTTCTCACAGATAGTGAGAAAGTGCTTCATTTAGTCTAA
BCoV           GGGTTCTGGCATGGACACCGCATTGTTGAGAAATAATATCTAA
HEV            GGGCTCAGGCATGGACACCGCATTGTTGAGAAATAATATCTAA
MHV            GAGT---GGCATGGACACCGCATTGTTGAGAATCTAA------
RtCoV          GAGT---GGCGCGGACACCGCATTGTTGAGAATCTAA------
SARS-Urb       CGGTAGCAACGACAATATTGCTTTGCTAGTACAGTAA------
SARS-Tor       CGGTAGCAACGACAATATTGCTTTGCTAGTACAGTAA------
IBV            AGTGTAGCAACAGGAGGAAGTAGTCTTTACACATAA-------
             * *
```

Figure 12: List of Molecular Designs for *M* Gene

LK257 Dubcyl 5'-FAM-CCTCCGACCCAATTAATTCTGTAGACAGCAGCCGGAGG-3'-

LK257.N Dubcyl 5'-FAM-GCCTCCGACCCAATTAATTCTGTAGACAGCAGCCGGAGGC-3'-

LK258

5'-CTTGTTTTCCTCTGGCTCTTGTGGCCAGTAACACTTGCTTGTTTTGTGCT
TGCTGCTGTCTACAGAATTAATTGGGTGACTGGCGGGATTGCGATTGCA
AT
GGCTTG-3'

LK259 5'- CTTGTTTTCCTCTGGCTCTTG -3'

5'- TAATACGACTCACTATAGGCTTGTTTTCCTCTGGCTCTTG -3'

LK260 5'- CAAGCCATTGCAATCGCAATC- 3'

LK260-RT 5'- AAGCAACGAAGTAGCAAGCCATTGCAATCGCAATC- 3'

M-RT 5'-AAGCAACGAAGTAG -3'

Figure 13: DNA Sequence Alignment of Coronavirus *N* Genes isolated from different Species

```
TGE          ------------------------------------------------------------
PRCoV        ------------------------------------------------------------
CcoV         ------------------------------------------------------------
FCoV         ------------------------------------------------------------
SARS-Urb     ----------------------ATGTCTGATAATGGACCCCAATCAAACCAACGTAGT
SARS-Tor     ----------------------ATGTCTGATAATGGACCCCAATCAAACCAACGTAGT
BcoV         ATGTCTTTTACTCCTGG-TAAGCAAT--CCAGTAGTAGAGCGTCCTTTGGAAATCGTTCT
HEV          ATGTCTTTCACTCCTGG-CAAGCAGT--CCAGCAGTAGAGCGTCCTCTGGAAATCGTTCT
MHV          ATGTCTTTTGTTCCTGGGCAAGAAAATGCCGGTAGCAGAAGCTCCTCTGGAAACCGCGCT
RtCoV        ATGTCTTTTGTTCCCGGACAAGAAAACGCCGGTAGCAGAAGCTCCTCTGGAAACCGCGCT
HCoVOC43     ------------------------------------------------------------
PEDV         ------------------------------------------------------------
IBV          ------------------------------------------------------------

TGE          ---------------------------------ATGGCCAA-CCAGGGACAA--------
PRCoV        ---------------------------------ATGGCCAA-CCAGGGACAA--------
CcoV         ---------------------------------ATGGCCTC-TCAGGGACAA--------
FCoV         ---------------------------------ATGGCCAC-ACAGGGACAA--------
SARS-Urb     GCCCCCCGCATTACATTTGGTGGACCCACAGATTCAACTGA--CAATAAC----------
SARS-Tor     GCCCCCCGCATTACATTTGGTGGACCCACAGATTCAACTGA--CAATAAC----------
BcoV         GGTAATGGCAT--CCTTAAGTGGGCCGATCAGTCCGACCAATCTAGAAATGT--------
HEV          GGTAATGGCAT--CCTTAAGTGGGCCGATCAGTCCGACCAGTCTAGAAATGT--------
MHV          GGTAATGGCAT--CCTCAAGAAGACCACTTGGGCTGACCAAACCGAGCG----------
RtCoV        GGTAATGGAAT--CCTCAAGAAGACCACTTGGGCTGACCAAACCGAGCGCGGACAAAATA
HCoVOC43     ------------------------------------------------------------
PEDV         ------------------------------------------------------------
IBV          ---------------------------------ATGGCAAGCGGTAAAGCAGC-------

TGE          -CGTGTCAGTTGGGGAGATGAATCTACCAAAACACGTGGTCGT-TCCAATTC---CCGTG
PRCoV        -CGTGTCAGTTGGGGGGATGAATCCACCAAAATACGTGGTCGC-TCCAATTC---CCGTG
CcoV         -CGTGTCAGTTGGGGAGATGAATCCACCAAGAGACGCGGTCGT-TCTAATTC---TCGTG
FCoV         -CGCGTCAACTGGGGAGATGAACCTTCCAAAAGACGTGGTCGT-TCTAACTC---TCGTG
SARS-Urb     -CAGAATGGAGGACGCAATGGGGCAAGGCCAAAACAGCGCCGA-CCCCAAGG----TTTA
SARS-Tor     -CAGAATGGAGGACGCAATGGGGCAAGGCCAAAACAGCGCCGA-CCCCAAGG----TTTA
BcoV         -TCAAACCAGGGGTAGAAGAGCTCAACCCAAGCAAACTGCTAC-TTCTCAGCTACCATCA
HEV          -TCAAACCAGGGGTAGAAGAGTTCAATCCAAGCAAACTGCTAC-TTCTCAGCAACCATCA
MHV          -TGGAAATAGAGGCAGAAGGAACCATCCCAAGCAGACTGCAAC-TACTCAGC--CCAATG
RtCoV        ATGGAAATAGAGGCAGAAGGAATCAGCCCAAGCAGACTGCAAC-TACTCAGC--CCAATA
HCoVOC43     -----------------ATGGCTACAGTCAAATGGGCTGATGCATCTGAACCACAACGTG
PEDV         ------------------------------------ATGGCTTCTGTCAGCTTTCAGG
IBV          -TGGAAAAACAGACGCCCCAGCGCCAGTCATTAAACTAGGAGGACCAAAACCACCTAAAG

TGE          GTCGGAAGAATAATAACATACC-TCTTTCATTCTTCAACCCCATAACCCTCCAACAAGGT
PRCoV        GTCGGAAGATTAATAACATACC-TCTTTCATTCTTCAACCCCATAACCCTCCAACAAGGT
CcoV         GCCGGAAGAATAATGATATACC-TCTTTCATTCTTCAACCCCATTACCCTCCAGCAAGGT
FCoV         GTCGGAAGAATAATGATATACC-TTTGTCATTCTACAACCCCATTACCCTCGAGCAAGGA
SARS-Urb     CCCAATAATACTGCG---------TCTTGGTTCACAGCTCTCACTCAGCATGGCAAGG--
SARS-Tor     CCCAATAATACTGCG---------TCTTGGTTCACAGCTCTCACTCAGCATGGCAAGG--
BcoV         GGAGGGAATGTTGTACCCTACTATTCTTGGTTCTCTGGAATTACTCAGTTTCAAAAAGGA
HEV          GGAGGGACTGTTGTACCCTACTATTCTTGGTTCTCTGGAATTACTCAGTTTCAAAAGGGA
MHV          CC-GGGAGTGTGGTTCCCCATTACTCTTGGTTTTCGGGCATCACCCAGTTTCAAAAGGGA
RtCoV        CC-GGGAGTGTGGTTCCCCATTACTCTTGGTTTTCGGGCATTACCCAATTCCAGAAGGGA
HCoVOC43     GTCGTCAGGGTAGAA---TACC-TTATTCTCTTTATAGCCCTTTGCTTGTTGATAGTG--
```

```
PEDV        ATCGTGGCCGCAAACGGGTGCCATTA-TCTCTCTATGCCCCTCTTAGGGTTACTAATGAC
IBV         TCGGTTCTTCTGGAAATGCA----TCTTGGTTTCAAGCAATAAAAGCCAAGAAGTTAAAT
                                    *   *

TGE         TCAAAATTTTGGAACTTATGTCCGAGAGACTTTGTACCCAAAGGAATAGGTAACAGGGAT
PRCoV       GCAAAATTTTGGAACTCATGTCCGAGAGATTTTGTACCCAAAGGAATAGGTAATAGGGAT
CcoV        TCAAAGTTTTGGGACTTATGTCCGAGAGACTTTGTACCCAAAGGAATAGGTAATAAGGAT
FCoV        TCTAAATTTTGGAATTTATGTCCGAGAGACCTTGTTCCCAAAGGAATAGGTAATAAGGAT
SARS-Urb    -AGGAACTTAGATTCCCTCGAGGCCAGGGCGTTCCAATCAACACCAATAGTGGTCCAGAT
SARS-Tor    -AGGAACTTAGATTCCCTCGAGGCCAGGGCGTTCCAATCAACACCAATAGTGGTCCAGAT
BcoV        AAGGAGTTTGAATTTGCAGAGGGACAAGGTGTGCCTATTGCACCAGGAGTCCCAGCTACT
HEV         AAGGAGTTTGAATTTGCAGAGGGACAAGGTGTGCCTATTGCACCAGGAGTCCCATCTACT
MHV         AAGGAGTTCCAGTTTGCACAAGGACAGGGAGTGCCTATTGCCAGTGGAATCCCCGCTTCA
RtCoV       AAAGAGTTCCAGTTTGCAGGTGGACAAGGAGTGCCTATTGCCAATGGAATCCCACCTTCT
HCoVOC43    -AACAACCTTGGAAGGTGATACCTCGTAATTTGGTACCCATCAACAAGAAAGACAAAAAT
PEDV        AAGCCCCTTTCTAAGGTACTTGCAAACAACGCTGTACCCACTAACAAGGGGAATAAGGAC
IBV         ACACCTCCGCCCAAGTTTGAAGGTAGCGGTGTTCCTGATAACGAAAACATTAAGCCAAGC

TGE         CAACAGATTGGTTATTGGAATAGACAAACTCG------CTATCGCATGGTGAAGGGCCAA
PRCoV       CAACAGATTGGTTATTGGAATAGACAAACTCG------CTATCGCATGGTGAAGGGCCAA
CcoV        CAACAAATTGGTTATTGGAACAGGCAAACCCG------TTATCGCATGGTGAAGGGTCGA
FCoV        CAACAAATTGGTTATTGGAATAGACAGATTCG------TTATCGTATTGTAAAAGGCCAG
SARS-Urb    GACCAAATTGGCTACTACCGAAGAGCTACCCGACGAGTTCGTGGT---GGTGACGGCAAA
SARS-Tor    GACCAAATTGGCTACTACCGAAGAGCTACCCGACGAGTTCGTGGT---GGTGACGGCAAA
BcoV        GAAGCTAAGGGGTACTGGTACAGACACAACAGACGTTCTTTTAAAACAGCCGATGGCAAC
HEV         GAAGCTAAGGGGTACTGGTACAGACACAACAGACGTTCTTTTAAAACAGCCGACGGCAAT
MHV         GAGCAAAAGGGATATTGGTATAGACACAACCGACGTTCTTTTAAAACACCTGATGGCCAG
RtCoV       GAGCAAAAGGGATATTGGTATAGACACAACCGTCGTTCTTTTAAAACACCTGATGGGCAG
HCoVOC43    AAGCTTATAGGCTATTGGAATGTTCAAAAACG------TTTCAGAACTAGAAAGGGCAAA
PEDV        CAGCAAATTGGGTACTGGAATGAGCAAATTCG------CTGGCGCATGCGCCGTGGTGAG
IBV         CAGCAACATGGATACTGGAGACGCCAAGCCAG------GTTTAAGCCAGGCAAAGGTGGA
             *       ** ** *                *                        **

TGE         CGTAAAGAGCTTCCTGAAAGGTGGTTCTTCTACTACTTAGGTACTGGACCTCATGCAGAT
PRCoV       CGTAAAGAGCTTCCTGAAAGGTGGTTCTTTTACTACTTAGGCACTGGACCTCATGCAGAT
CcoV        CGTAAAAATCTTCCTGAAAAGTGGTTCTTCTACTATTTAGGAACTGGACCTCATGCTGAT
FCoV        CGTAAGGAACTCGCTGAGAGGTGGTTCTTTTACTTCTTAGGTACAGGACCTCATGCTGAT
SARS-Urb    ATGAAAGAGCTCAGCCCCAGATGGTACTTCTATTACCTAGGAACTGGCCCAGAAGCTTCA
SARS-Tor    ATGAAAGAGCTCAGCCCCAGATGGTACTTCTATTACCTAGGAACTGGCCCAGAAGCTTCA
BcoV        CAGCGTCAACTGCTGCCACGATGGTATTTTTACTATCTTGGAACAGGACCGCATGCCAAA
HEV         CAGCGTCAACTGCTGCCACGATGGTACTTTTACTACCTGGGAACAGGACCGCATGCCAAA
MHV         CACAAGCAGCTACTGCCCAGATGGTATTTTTACTATCTTGGAACAGGGCCCCATGCTGGC
RtCoV       CAGAAGCAACTACTCCCCAGATGGTATTTTTACTATCTTGGGACGGGCCCCCATGCTGGA
HCoVOC43    CGGGTGGATTTGTCACCCAAGCTGCATTTTTATTATCTTGGCACAGGACCCCATAAAGAT
PEDV        CGAATTGAACAACCTTCCAATTGGCATTTCTACTACCTCGGAACAGGACCTCACGGCGAC
IBV         AGAAAACCAGTCCCAGATGCTTGGTACTTTTACTATACTGGAACAGGACCTGCCGCTGAC
                                   *   ** ** *     ** ** ** **

TGE         GCCAAATTTAAAGATAAATTAGATGGAGTTGTCTGGGTTGCCAAGGATGGTGCCATGAAC
PRCoV       GCCAAATTTAAAGATAAATTAGATGGAGTTGTCTGGGTTGCCAAGGATGGTGCCATGAAC
CcoV        GCCAAATTTAAGCAAAAATTAGATGGAGTTGTCTGGGTTGCTAGGGGAGATTCCATGACT
FCoV        GCTAAATTCAAAGACAAGATTGATGGAGTCTTCTGGGTTGCAAGGGATGGTGCCATGAAC
SARS-Urb    CTTCCCTACGGCGCTAACAAAGAAGGCATCGTATGGGTTGCAACTGAGGGAGCCTTGAAT
SARS-Tor    CTTCCCTACGGCGCTAACAAAGAAGGCATCGTATGGGTTGCAACTGAGGGAGCCTTGAAT
BcoV        GACCAGTATGGCACCGATATTGACGGAGTCTTCTGGGTCGCTAGTAACCAGGCTGATGTC
HEV         GACCAGTACGGCACCGACATTGACGGAGTCTTCTGGGTCGCTAGTAACCAGGCTGATATT
MHV         GCAGAGTATGGCGACGATATCGAAGGAGTTGTCTGGGTCGCAAGCCAACAGGCCGACACT
RtCoV       GCCAGTTTCGGAGACAGCATTGAGGGAGTCTTCTGGGTTGCAAATAGTCAGGCGGATACC
HCoVOC43    GCAAAATTTAGAGAGCGTGTTGAAGGTGTCGTCTGGGTTGCTGTTGATGGTGCTAAAACT
PEDV        CTCCGTTATAGGACTCGTACTGAGGGTGTTTTCTGGGTTGCTAAAGAAGGCGCAAAGACT
IBV         CTGAACTGGGGTGATACTCAAGATGGTATAGTGTGGGTTGCTGCTAAGGGTGCTGATACT
                 *              ** **  *  * ***** **           *
```

```
TGE          AAACCAACCACGC---TTGGTAGTCGTGGTGCTAATA---ATGAATCCAAAGCTTTGAAA
PRCoV        AAACCAACCACGC---TTGGTAGTCGTGGTGCTAATA---ATGAATCCAAAGCTTTGAAA
CcoV         AAGCCAACAACTC---TTGGTACTCGTGGCACTAATA---ATGAATCAAAGGCTTTGAAA
FCoV         AAGCCCACAACGC---TTGGCACTCGTGGAACCAATA---ACGAATCCAAACCACTGAGA
SARS-Urb     ACACCCAAAGACCACATTGGCACCCGCAATCCTAATAACAATGCTGCCACCGTGCTACAA
SARS-Tor     ACACCCAAAGACCACATTGGCACCCGCAATCCTAATAACAATGCTGCCACCGTGCTACAA
BcoV         AATACCCCGGCTGACATTCTCGATCGGGACCCAAGTAGCGATGAGGCTATTCCGACTAGG
HEV          AATACCCCGGCTGACATTGTCGATCGGGATCCAAGTAGCGATGAGGCTATTCCGACTAGG
MHV          AAGACCACTGCCGATGTTGTTGAAAGGGACCCAAGCAGTCATGAGGCTATTCCTACTAGG
RtCoV        AACACCTCTGCTGACATTGTTGAAAGGGACCCAAGTAGCCATGAGGCTATTCCTACTAGG
HCoVOC43     GAACCTACAGGTTA---CGGTGTTAGGCGCAAGAATT----CAGAACCAGAGATACCACA
PEDV         GAACCCACTAATT---TGGGTGTCAGAAAGGCGTCTG----AAAAGCCAATCATTCCAAA
IBV          AAATCTAGATCCAATCAGGGTACAAGAGATCCTGATAA-GTTTGACCAATACCCACTACG
                 *                   *                    * *

TGE          TTCG---ATGGTAAAGTGCCAGGCGAA---TTTCAACTTGAAGTTAATCAATCAAGAGAC
PRCoV        TTCG---ATGGTAAAGTGCCAGGCGAA---TTTCAACTTGAAGTTAACCAGTCTAGGGAC
CcoV         TTCG---ATGTCAAAGTACCATCAGAA---TTTCACCTTGAAGTGAACCAATTAAGGGAC
FCoV         TTTG---ATGGTAAGATACCGCCACAG---TTTCAGCTTGAAGTGAACCGTTCTAGGAAC
SARS-Urb     CTTCCTCAAGGAACAACATTGCCAAAAGGCTTCTACGCAGAGGGAAGCAGAGGCGGCAGT
SARS-Tor     CTTCCTCAAGGAACAACATTGCCAAAAGGCTTCTACGCAGAGGGAAGCAGAGGCGGCAGT
BcoV         TTTCCGCCTGGCACGGTACTCCCTCAGGGTTACTATATTGAAGG---CTCAGGAAGGTCT
HEV          TTTCCGCCTGGCACGGTACTCCCTCAAGGTTACTATATTGAAGG---CTCAGGAAGGTCT
MHV          TTTGCGCCCGGCACGGTATTGCCTCAGGGCTTTTATGTAGAAGG---CTCGGGAAGGTCT
RtCoV        TTTGCGCCCGGTACGGTATTGCCTCAGGGTTTCTATGTTGAAGG---CTCGGGAAGGTCT
HCoVOC43     CTTC--AATCAAAAGCTCCCAAATGGTGTTACTGTTGTTGAAGAACCTGACTCCCGTGCT
PEDV         ATTC--TCTCAACAGCTCCCCAGTGTA---GTTGAGATTGTTGAACCTAACACACCTCCT
IBV          ATTC----TCGGATGGCGGACCTGATGGTAATTTCCGTTGGGACTTCATTCCCCTGAACC
              *                                   *

TGE          AATTCAAGGTCACGCTCTCAATCTAGATCTC------GGTCTAGAAATAGATCTCAATCT
PRCoV        AACTCAAGGTCACGCTCTCAATCTAGATCGC------GGTCTAGAAACAGATCTCAATCT
CcoV         AATTCAAGGTCTAGGTCTCAATCTAGATCTC------AGTCCAGAAATAGGTCTCAATCT
FCoV         AATTCAAGGTCTGGTTCTCAGTCTAGATCTG------TTTCAAGAAACAGATCTCAATCT
SARS-Urb     CAAGCCTCTTCTCGCTCCTCATCACGTAGTCGCGGTAATTCAAGAAATTCAACTCCTGGC
SARS-Tor     CAAGCCTCTTCTCGCTCCTCATCACGTAGTCGCGGTAATTCAAGAAATTCAACTCCTGGC
BcoV         GCTCCTAATTCCAGATCTACTTCACGCGCATCCAGTAGAGCCTCTAGTGCAGGATCGCGT
HEV          GCTCCTAATTCCAGATCTACTTCGCGTGCACCCAATAGAGCCCCTAGTGCAGGATCGCGT
MHV          GCACCTGCTAGTCGATCTGGTTCGCGGTCAC------AATCCCGTGGGCCAAATAATCGC
RtCoV        GCACCTGCTAGTCGATCTGGTTCGCGGTCAC------AATCCCGTGGGCCAAATAATCGC
HCoVOC43     CCTTCCCGGTCTCAGTCGAGGTCGCAGAGTCGCGGTCGTGGTGAATCCAAACCTCAATCT
PEDV         GCTTCACGTGCAAATTCGCGTAGCAGGAGTCGTGGCAATGGCAACAATAGGTCTAGATCT
IBV          GTGGTAGGAGTGGAAGATCAACAGCAGCTTCATCA-GCAGCAGCTAGTAGAGCACCATCA

TGE          AGAGGCAGGCAACAATTCAATAACAAGAAGGAT-------GACAGTGTAGAACAAGCTGT
PRCoV        AGAGGTAGGCAACAATCCAATAACAAGAAGGAT-------GACAGTGTAGAACAAGCTGT
CcoV         AGAGGAAGGCAACTATCCAATAATAAGAAGGAT-------GACAATGTTGAACAAGCTGT
FCoV         AGAGGAAGACACCATTCCAATAACCAGAA---T-------AATAATGTTGAGGATACAAT
SARS-Urb     AGCAGTAGGGGAAATTCTCCTGCTCGAATGGCT-------AGCGGAGGTGGTGAAACTGC
SARS-Tor     AGCAGTAGGGGAAATTCTCCTGCTCGAATGGCT-------AGCGGAGGTGGTGAAACTGC
BcoV         AG---TAGAGCCAATTCTGGCAACAGAACCCCT-------ACCTCTGGTGTAACACCTGA
HEV          AG---TAGAGCCAATTCTGGCAATAGAACCTCT-------ACCCCTGGTGTAACACCTGA
MHV          GC---TAGAAGCAGTTCCAACCAGCGCCAGCCT-------GCCTCTGCTGTAAAACCTGA
RtCoV        GC---TAGAAGCAGTTCCAACCAGCGCCAGCCT-------GCCTCTACTGTAAAACCTGA
HCoVOC43     CGGAATCCTTCAAGTGACAGAAACCATAACAGTCAG----GATGACATCATGAAGGCAGT
PEDV         CCAAGTAACAACAGAGGCAATAACCAGTCCCGTGGTAATTCACAGAATCGTGGAAATAAC
IBV          CG------TGAAGGTTCGCGTGGTCGTAGAAGT-------GATTCTGGAGATGACCTTAT
                                                     *

TGE          T---CTTGCCGCACTT-----AAAAAGTTAGGTGTTGACACAGAAAAACAACAGCA-ACG
```

```
PRCoV       T---CTTGCCGCACTT-----AAAAAGTTAGGTGTTTACACAGAAAAACAACAGCA-ACG
CcoV        T---CTTGCTGCACTC-----AAAAAGTTAGGTGTTGACACAGAAAAACAA---CA-AAG
FCoV        T---GTAGCCGTGCTT-----GAAAAATTAGGTGTT---ACTGACAAACAA-------AG
SARS-Urb    C---CTCGC-G--CTA-----TTGCTGCTAGAC-AGATTGAACCAGCTTGAGAGCA-AAG
SARS-Tor    C---CTCGC-G--CTA-----TTGCTGCTAGAC-AGATTGAACCAGCTTGAGAGCA-AAG
BcoV        T---ATGGCTG--ATC-----AAATTGCTAGTC-TTGTTCTGGCAAAACTTG-GCA-AGG
HEV         C---ATGGCTG--ATC-----AAATTGCTAGTC-TTGTTCTGGCAAAACTTG-GCA-AGG
MHV         C---ATGGCCG--AAG-----AAATTGCTGCTC-TTGTTTTGGCTAAGCTTG-GTA-AAG
RtCoV       T---ATGGCCG--AAG-----AAATTGCTGCTC-TTGTTTTGGCTAATCTAG-GCA-AAG
HCoVOC43    T---GCTGCGGCTCTT-----AAATCTTTAGGTTTTGACAAGCCTCAGGAAAAAGATAAA
PEDV        CAGGGTCGTGGAGCTTCTCAGAACAGAGGAGGCAATAATAATAACAATAACAAGTCTCGT
IBV         TGCTCGTGCAGCAAAG------ATAATCCAGGATCAGCAGAAAAAGGGCTCTCGCA--TT
                   *  *
```

```
TGE         CTCTCGTTCTAAATCTAAAG--AACGTAGTAACTCTAAGAC-AAGA--------GATACT
PRCoV       CTCTCGTTCTAAATCTAAAG--AACGTAGTAACTCTAAAAC-AAGA--------GATACT
CcoV        ATCTCGTTCCAAATCTAAGG--AACGTAGCAGCTCTAAGAC-AAGA--------GATACT
FCoV        GTCACGTTCTAAACCTAGAG--AACGTAGTGATTCCAAACC-TAGG--------GACACA
SARS-Urb    TTTCTGGTAAAGGCCAACAA-CAACAAGGCCAAACTGTCACTAAGA--------AATCTG
SARS-Tor    TTTCTGGTAAAGGCCAACAA-CAACAAGGCCAAACTGTCACTAAGA--------AATCTG
BcoV        ATGCCACTAAGCCACAGCAAGTAACTAAGC-AGACTGCCA--AAGA--------AATCAG
HEV         ATGCCACTAAGCCTCAGCAAGTAACTAAGC-AGACTGCCA--AAGA--------GGTCAG
MHV         ATGCCGGCCAGCCCAAGCAGGTAACTAAGC-AAAGCGCCA--AAGA--------AGTCAG
RtCoV       ATGCCGGACAGCCTAAGCAAGTAACTAAGC-AAAGTGCCA--AAGA--------AGTCAG
HCoVOC43    AAGTCAGCGAAAACGGGTAC--TCCTAAGCCTTCTCGTAATCAGAGT----CCTGCTTCT
PEDV        AACCAGTCCAATAACAGGAACCAGTCAAATGACCGTGGTGGTGTAACATCACGCGATGAT
IBV         ACCAAGGCAAAGGCAGATGA--AATGGCTCATCGCCGGTATTGCAAG----CGCACTATC
```

```
TGE         ACACCTAAGAATGAAAACAAACACACCTGGA---AGAGAACTGCAGGTAAA---GGTG--
PRCoV       ACGCCTAAGAATGAAAACAAACACACCTGGA---AGAGAACTGCAGGTAAA---GGTG--
CcoV        ACACCTAAGAATGAAAACAAACACACCTGGA---AGAGAACTGCAGGTAAA---GGTG--
FCoV        ACACCTAAGAATGCCAACAAACACACCTGGA---AGAAAACTGCAGGCAAG---GGAG--
SARS-Urb    CTGCTGAGGCATCTAAA-AAGCCTCGCCAAA---AACGTACTGCCACAAAAC--AGTACA
SARS-Tor    CTGCTGAGGCATCTAAA-AAGCCTCGCCAAA---AACGTACTGCCACAAAAC--AGTACA
BcoV        --ACAGAAAATTTTGAATAAGCCCCGCCAGA---AGAGGAGCCCCAATAAAC--AATGCA
HEV         --ACAGAAAATCTTGAATAAGCCCCGCCAGA---AGAGGAGCCCCAACAAAC--AATGCA
MHV         --GCAGAAAATTTTAACTAAGCCTCGTCAAA---AGAGGACTCCAAACAAGC--AGTGCC
RtCoV       --GCAGAAAATTTTAAATAAGCCTCGCCAAA---AGAGGACTCCAAACAAGC--AGTGCC
HCoVOC43    TCTCAAACTTCTGCCAAGAGTCTTGCTCGTTCTCAGAGTTCTGAAACAAAAG--AACAAA
PEDV        CTGGTGGCTGCTGTCAAGGATGCACTTAAATCTTTGGGTATTGGAGAAAATCCTGACAGG
IBV         CCACCTAATTATAGGGTTGATCAAGTGTTTGGTCCCCGTACTAAAGGTAAGG--AGGGGA
                                                          **
```

```
TGE         ---ATGTGACA---AGATTTTATGGAGCTAGAAGCAGTT--CAGCC-------AAT-TTT
PRCoV       ---ATGTGACA---AGATTTTATGGAGCTAGAAGCAGCT--CAGCC-------AAT-TTT
CcoV        ---ATGTGACA---AAATTTTATGGAGCTAGAAGTAGTT--CAGCC-------AAT-TTT
FCoV        ---ATGTGACA---ACTTTCTATGGTGCTAGAAGTAGTT--CAGCT-------AAC-TTT
SARS-Urb    A-CGTCACTCA---AGCATTTGGGAGACGTGGTCCAGAA--CAAACCCAAGGAAAT-TTC
SARS-Tor    A-CGTCACTCA---AGCATTTGGGAGACGTGGTCCAGAA--CAAACCCAAGGAAAT-TTC
BcoV        C-TGTTCAGCA---GTGTTTTGGGAAGAGAGGCCCCAAT--CAGA---------AT-TTT
HEV         C-TGTTCAGCA---GTGTTTTGGGAAGAGAGGCCCCAAT--CAGA---------AT-TTT
MHV         C-AGTGCAGCA---GTGTTTTGGGAAGAGAGGCCCTAAT--CAGA---------AC-TTT
RtCoV       C-AGTGCAGCA---GTGTTTTGGAAAGAGAGGCCCCAAT--CAGA---------AT-TTT
HCoVOC43    AGCATGAAATGC--AAAAGCCACGGTGGAAAAGACAGCC--TAATGATGATGTGACATCT
PEDV        CATAAGCAACAGCAGAAGCCTAAGCAGGAAAAGTCTGACAACAGCGGCAAAAATACACCT
IBV         A--TTTTGGTG-----ATGACAAGATGAATGAGGAAGGTATTAAGGATG----GGCGTGT
                                                          *
```

```
TGE          GGTGACACTGACCTCGTTGCCAAT----GGGAGCAGTGCCAAGCATTACCCACAACTGGC
PRCoV        GGTGACAGTGACCTCGTTGCCAAT----GGGAGCAGTGCCAAGCATTACCCACAATTGGC
CcoV         GGTGACAGCGATCTTGTTGCCAAT----GGGAACGGTGCCAAGCATTACCCACAACTGGC
FCoV         GGTGATAGTGATCTCGTTGCCAAT----GGTAACGCTGCCAAATGCTACCCTCAGATAGC
SARS-Urb     GGGGACCAAGACCTAATCAGACAA----GGAACTGATTACAAACATTGGCCGCAAATTGC
SARS-Tor     GGGGACCAAGACCTAATCAGACAA----GGAACTGATTACAAACATTGGCCGCAAATTGC
BcoV         GGTGGTGGAGAAATGTTAAAACTT----GGAACTAGTGACCCACAGTTCCCCATTCTTGC
HEV          GGTGGTGGAGAAATGTTAAAACTT----GGAACTAGTGACCCACAGTTCCCCATTCTTGC
MHV          GGAGGCTCTGAAATGTTAAAACTT----GGAACTAGTGATCCGCAGTTCCCCATTCTTGC
RtCoV        GGAGGCCCTGAAATGTTAAAACTT----GGAACTAGTGATCCACAGTTCCCCATTCTTGC
HCoVOC43     AATGTCACACAATGTTTTGGCCCC----AGAGACCTTGAC---CACAACTTTGGAAGTGC
PEDV         AAGAAGAACAAATCCAGGGCCACTTCGAAGGAACGTGACCTCAAAGACATCCCAGAGTGG
IBV          TACAGCAATGCTCAACCTAGTCCCT---AGCAGCCATGCT-----TGTCTTTTTGGAAGT
                                          *                          *

TGE          TGAATGTGTT-CCATCTGTGTC--TAGCATT----CT--GTTTGGAA-GCTATTGGACTT
PRCoV        TGAATGTGTT-CCATCTGTGTC--TAGCATT----TT--GTTTGGAA-GCTATTGGACTT
CcoV         TGAATGTGTT-CCATCTGTATC--TAGCATT----CT--GTTTGGAA-GCCATTGGACTG
FCoV         TGAATGTGTT-CCATCAGTGTC--TAGCATA----AT--CTTTGGCA-GTCAATGGTCTG
SARS-Urb     ACAATTTGCT-CCAAGTGCCTC--TGCA-------TT--CTTTGGAATGTCAC--GCATT
SARS-Tor     ACAATTTGCT-CCAAGTGCCTC--TGCA-------TT--CTTTGGAATGTCAC--GCATT
BcoV         AGAACTCGCA-CCCACAGCTGG--TGCGTTT----TT--CTTTGGATCAAGATTAGAGTT
HEV          AGAACTCGCA-CCCACAGCTGG--TGCGTTT----TT--CTTTGGATCAAGATTAGAGTT
MHV          AGAGTTGGCT-CCAACACCTAG--TGCCTTC----TT--CTTTGGATCTAAATTAGAATT
RtCoV        AGAGTTGGCC-CCAACACCTGG--TGCCTTC----TT--CTTTGGATCTAAATTAGAATT
HCoVOC43     AGGTGTTGTGGCCAATGGTGTT--AAAGCTAAAGGCT--ATCCACAATTTGCTGAGCTTG
PEDV         AGGAGAATTC-CCAAGGGCGAAAATAGCGTAGCAGCTTGCTTCGGACCCAGAGGGGGCTT
IBV          AGAGTGACACCCAAACTTCAAC--TAGATGG--GCTTCACTTGAGATTTGAATTTACTAC
                       *                              *   *

TGE          CAAAGGAA--GATGGCG----ACCAGATAGAAGTCAC----GTTCACACACAAATACC--
PRCoV        CAAAGGAA--GATGGCG----ACCAGATAGAAGTCAC----GTTCACACACAAATACC--
CcoV         CTAAGGAA--GATGGTG----ACCAGATTGAAGTCAC----ATTCACACACAAATACC--
FCoV         CTGAAGAA--GCTGGTG----ATCAAGTGAAAGTCAC----GCTCACTCACACCTACT--
SARS-Urb     GGCA----------------------TGGAAGTCAC----ACCTTCGGGAACATG----
SARS-Tor     GGCA----------------------TGGAAGTCAC----ACCTTCGGGAACATG----
BcoV         GGCCAAAG--TGCAGAA----TTTGTCTGGGAATCTTGACGAGCCCCAGAAGGATGTTTA
HEV          GGCCAAAG--TGCAGAA----TTTGTCTGGGAATCCTGACGAGCCCCAGAAGGATGTTTA
MHV          GGTCAAAA-----AGAA----CT---CTGGTGGTGCTGATGAACCCACCAAAGATGTTTA
RtCoV        GGTCAAAA-----AGAA----TT---CTGGTGGCGTTGATGAACCCACCAAAGATGTGTA
HCoVOC43     TGCCGTCA--ACAGCTGCTATGCTGTTTGATAGTCAC-ATTGTTTCCAAAGAGTCAGGCA
PEDV         CAAAAACTTTGGAGATGCGGAATTTGTCGAAAAAGGTGTTGATGCGTCAGGCTATGCTCA
IBV          TGTGGTCCCATGTGATGACCCGCAGTTTGATAATTATGTGAAAATTTGTGATCAGTGT--

TGE          --ACT-TGCCAAAGGATGA--TCCTA--AGACTGGACAATTCCTTCAGCAGATTAATGCC
PRCoV        --ACT-TGCCAAAGGATCA--TCCTA--AAACTGAACAATTCCTTCAGCAGATTAATGCC
CcoV         --ACT-TGCCAAAGGATGA--TCCTA--AGACTGGACAATTCCTTCAGCAGATTAATGCA
FCoV         --ACC-TGCCAAAGGATGA--TGCCA--AAACTAGTCAATTCCTAGAACAGATTGACGCT
SARS-Urb     ---GC-TGACTTATCATGG--AGCCATTAAATTGGATGACAAAGATCCACAATTCAAAGA
SARS-Tor     ---GC-TGACTTATCATGG--AGCCATTAAATTGGATGACAAAGATCCACAATTCAAAGA
BcoV         TGAAT-TGCGCTATAATGG--TGCAATTAGATTTGACAGTACACTTTCAGGTTTTGAGAC
HEV          TGAAT-TGCGCTATAATGG--CGCGATTAGATTTGACAGCACACTCTCAGGTTTTGAAAC
MHV          TGAAT-TGCAGTATTCAGG--TGCAATTAGATTTGATAGTACTCTACCCGGTTTTGAGAC
RtCoV        TGAGC-TGCAATATTCAGG--TGCAGTCAGATTTGATAGTACTCTACCTGGTTTTGAGAC
HCoVOC43     ACACTGTGGTCTTGACTTT--CACTACTAGAGT-GACTGTGCCCAAAGACCATCCACACT
PEDV         GATCGCCAGTTTAGCACCAAATGTTGCAGCATTGCTCTTTGGTGGTAATGTGGCTGTTCG
IBV          -GTCGATGGTGTAGGAACGCGTCCAA---AAGATGACGA-ACCAAAACCAAAGTCACGCT
                                          *

TGE          TATGCTCG--TCCATCAGAAGTGGCAAAAGAA--CAGAGA-AAAAGA-----AAATCTCG
PRCoV        TATGCTAG--CCCATCAGAATTGGCAAAAGAA--CAGAGA-AAAAGA-----AAGTCTCG
CcoV         TACGCCCG--TCCATCAGAGGTGGCTAAAGAA--CAGAGA-CAACGC-----AAAGCTCG
```

```
FCoV        TACAAGCG--ACCTTCTGAAGTGGCTAAGGAT--CAGAGG-CAAAGA-----AGATCCCG
SARS-Urb    CAACGTCA--TACTGCTGAACAAGCACATTGA--CGCATA-CAAAAC-----ATTCCCAC
SARS-Tor    CAACGTCA--TACTGCTGAACAAGCACATTGA--CGCATA-CAAAAC-----ATTCCCAC
BcoV        CATAATGA--AGGTGTTGAATGAGAATTTGAA--TGCATATCAACAA-----CA---AGA
HEV         CATTATGA--AGGTGCTTAACCAGAATTTGAA--TGCCTATCAACAT-----CAGGAAGA
MHV         TATCATGA--AAGTGTTGACTGAGAATTTGAA--TGCCTACCAGGAC-----CAAGCTGG
RtCoV       TATCATGA--AAGTGTTGAATGAGAATTTGAA--TGCCTACCAGAAT-----CAAGCTGG
HCoVOC43    TG-GGTAA--GTTTCTTGAGGAGTTAAATGCATTCACTAGAGAAATG-----CAACAACA
PEDV        TGAGCTAGCGGACTCTTACGAGATTACATACAACTATAAAATGACTGTGCCAAAGTCAGA
IBV         CAAGTTCAAGACCTGCTACAAGAGGAAATTCT--CCAGCGCCAAGACAACAGCGCCCAAA

TGE         TTCTAAATCTGCAGAAAGGTCAGAGCAAGATGTGGTACCTGATGCATTAATAGAA-AATT
PRCoV       TTCTAAATCTGCAGAAAGGTCAGAGCAAGAGGTGGTACCTGATTCATTAATAGAA-AACT
CcoV        TTCTAAATCTGTAGAAAGGGTAGAGCAAGAGGTTGTACCTGATGCATTAACAGAA-AATT
FCoV        TTCTAAGTCTGCTGATAAG---AAGCCTGAGGAGTTGTCTGTAACTCTTGTGGAG-GCAT
SARS-Urb    CAACAGAGCCTAAAAAGGACAAAAAGAAAAAGACTGATGAAGCTCAGCCTTTGCC-GCAG
SARS-Tor    CAACAGAGCCTAAAAAGGACAAAAAGAAAAAGACTGATGAAGCTCAGCCTTTGCC-GCAG
BcoV        TGGTATGATGAATATGAGTCCAAAACCACAGCGTCAGCGTGGTCAGAAG-------AATG
HEV         TGGGATGATGAATATTAGTCCTAAACCACAGCGGCAGCGTGGTCAGAAG-------AATG
MHV         TAGTGTAGATCTAGTGAGCCCAAAGCCTCCAAGAAGAGGTCGTAGACAGGCTCAA-GAAA
RtCoV       TGGTGCAGATGTAGTGAGCCCAAAGCCCCAAAGAAAGAGAGGGACGAAACAAACG-GCTC
HCoVOC43    TCCTCTTCTTAACCCTAGTGCACTAGAATTCAACCCATCTCAAACTTCACCTGCA-ACTG
PEDV        TCCAAATGTTGAGCTTCTTGTTTCACAGGTGGATGCATTTAAAACTGGGAATGCAAAACT
IBV         GAAGGAGAAAAAGCTAAAGAAGCAGGATGATGAAGCAGATAAAGCATTGACCTCAGATGA

TGE         ATACAGATGTGTTTGAT-GACACACAGGT-TGAGATAATTGATGAGGTA-ACGAACTAA-
PRCoV       ATACAGATGTGTTTGAT-GACACACAGGT-TGAGATGATTGACGAGGTA-ACGAACTAA-
CcoV        ACACAGATGTGTTTGAT-GACACACAGGT-TGAGATTATTGATGAGGTA-ACGAACTAA-
FCoV        ACACAGATGTGTTTGAT-GACACACAGGT-TGAGATGATTGATGAGGTT-ACGAACTAA-
SARS-Urb    AGACAAAAGAAGCAGCCCACTGTGACTCT-TCTT---CCTGCGGCTGAC-ATG-GATGAT
SARS-Tor    AGACAAAAGAAGCAGCCCACTGTGACTCT-TCTT---CCTGCGGCTGAC-ATG-GATGAT
BcoV        GACAAGGAGAAAATGAT-AATATAAGTGT-TGCAGCGCCTAAAAGCCGT-GTGCAGCAAA
HEV         GACAAGTAGAAAATGAT-AATGTAAGTGT-TGCAGCGCCTAAAAGCCGT-GTGCAGCAAA
MHV         AGAAAGATGAAGTAGAT-AATGTAAGCGT-TGCAAAGCCCAAAAGCTTG-GTGCAGCGAA
RtCoV       AGAAAGAAGAATTAGAT-AGTATAAGCGT-TGCAAAGCCCAAAAGTGCC-GTGCAGCGAA
HCoVOC43    CTGAACCAGTGCGTGAT-GAAGTTTCTAT-TGAAA--CTGACATAATTG-ATGAAGTAAA
PEDV        CCAGAGAAAGAAGGAAAAGAAGAACAAGCGTGAAACCACGCTGCAGCAGCATGAAGAGGC
IBV         GGAGAGGAACAATGCACAGCTGGAATTTTATGATGAGCCCAAGGTAATTAACTGGGGGGA
                *                       *

TGE         ------------------------------------------------------------
PRCoV       ------------------------------------------------------------
CcoV        ------------------------------------------------------------
FCoV        ------------------------------------------------------------
SARS-Urb    TTCTCCAGACAACTTCAAAATTCCATGAGTGGAGCTTCTGCTGATTCAACTCAGGCATAA
SARS-Tor    TTCTCCAGACAACTTCAAAATTCCATGAGTGGAGCTTCTGCTGATTCAACTCAGGCATAA
BcoV        ATAAGAGTAGAGAGTTGACTGCAGAGGACATCAGC--CTTCTTAAGAAGAT---GGATGA
HEV         ATAAGAGTAGAGAGTTGACAGCAGAGGACATCAGC--CTTCTTAAGAAGAT---GGATGA
MHV         ATGTAAGTAGAGAATTAACCCCCGAGGATCGTAGC--CTGCTGGCTCAGATCCTAGACGA
RtCoV       ATGTAAGCAGAGAATTAACCCCAGAGGATAGAAGC--CTGTTGGCGCAGATCCTAGATGA
HCoVOC43    CTAA--------------------------------------------------------
PEDV        CATCTACGATGATGTGGGTGCGCCATCTGATGTGACCCATGCCAATCTGGAATGGGACAC
IBV         TGCAGCTCTAGGAGAGAATGAACTTTGA----------------------------------

TGE         ------------------------------------------------------------
PRCoV       ------------------------------------------------------------
CcoV        ------------------------------------------------------------
FCoV        ------------------------------------------------------------
SARS-Urb    ------------------------------------------------------------
```

```
SARS-Tor        ------------------------------------------------------------
BcoV            --GCCCTATACTGAA----------GACACCTCAGAAATATAA-----------------
HEV             --GCCCTATACTGAA----------GATACCTCAGAAATATAA-----------------
MHV             TGGCGTTGTGCCAGATGGGTTGGAAGATGACTCTAATGTGTAA-----------------
RtCoV           TGGCGTTGTGCCTGATGGGTT---AGATGACTCTAATGTGTAA-----------------
HCoVOC43        ------------------------------------------------------------
PEDV            AGCTGTTGATGGTGGTGATACGGCCGTTGAAATTATCAACGAGATCTTCGATACAGGAAA
IBV             ------------------------------------------------------------

TGE             ----
PRCoV           ----
CcoV            ----
FCoV            ----
SARS-Urb        ----
SARS-Tor        ----
BcoV            ----
HEV             ----
MHV             ----
RtCoV           ----
HCoVOC43        ----
PEDV            TTAA
IBV             ----
```

Figure 14: Phylogenetic Analysis of *N* Gene

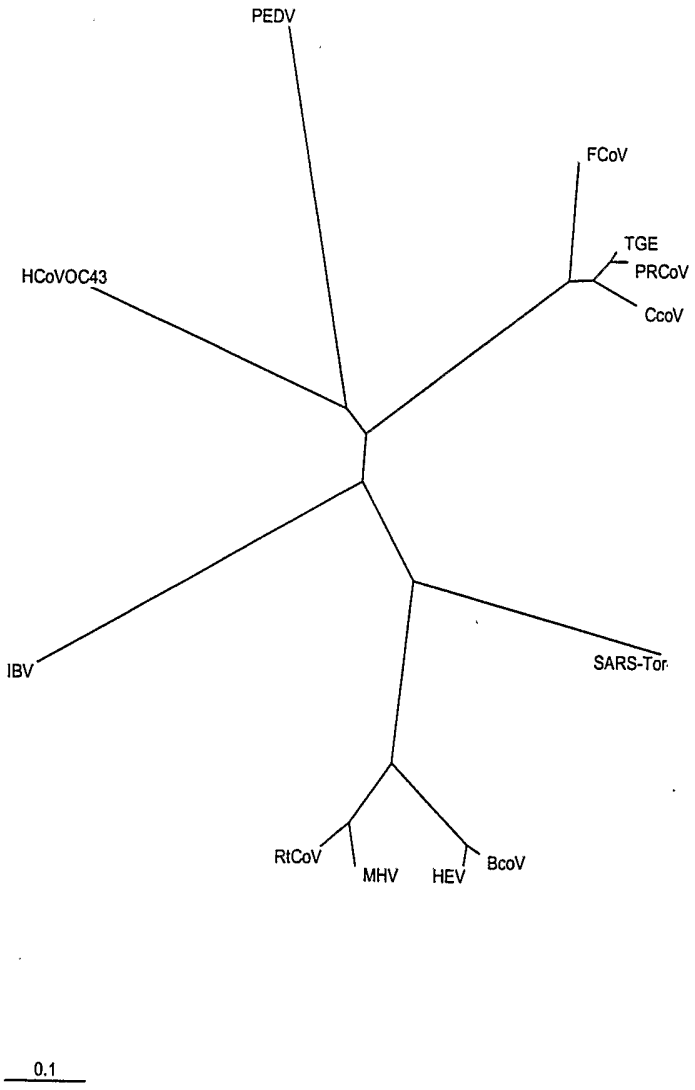
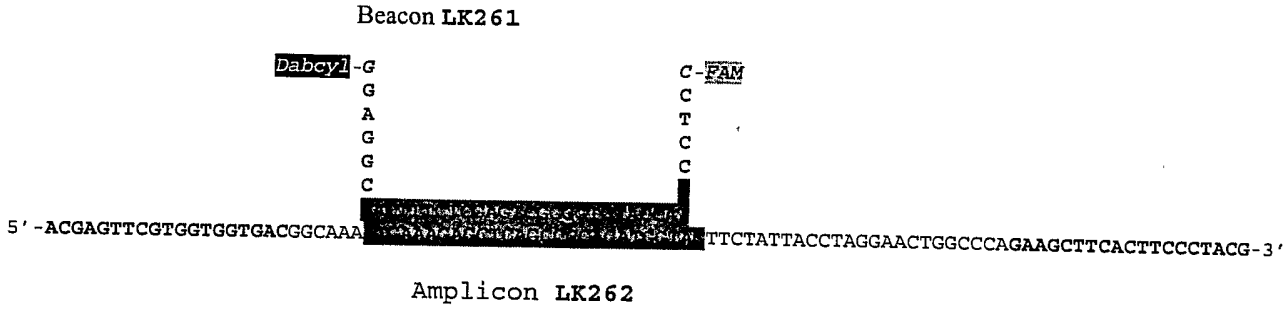


Figure 15: Molecular Designs for SARS-Associated *N* Gene

NOTE:	LK261 is a TET-molecular beacon, which recognizes the ***N* gene of coronavirus (SARS Tor2 and SARS urbani human pathogenic strains)**. It will be used in a real-time PCR diagnostic for the identification of **SARS-associated coronavirus RNA/DNA.**



LK261

M. Target recognition sequence:	**26 nucleotides (11 G/C) + (1 G/C arm)**
N. Length of the arms:	**6 nucleotides (5 G/C)**
O. Melting temperature of the beacon:	**dG = -1.64 dH = -48.7 dS = -145.8 Tm = 61.0 °C**
P. Melting temperature of target:	**°C (including 1 nucleotide from the arms)**

LK263 **5'-ACGAGTTCGTGGTGGTGAC-3'**

Length: 19 nucleotides (11 G/C)
Tm: °C
Position: (see alignment below)

LK264 **5' –CGTAGGGAAGTGAAGCTTC-3'**

Length: 19 nucleotides (10 G/C)
Tm: °C
Position: (see alignment below)

N-RT **5' –GCCTTCTTTGTTAG-3'**

Length: 14 nucleotides (6 G/C)
Tm: 47 °C
Position: (see alignment below)

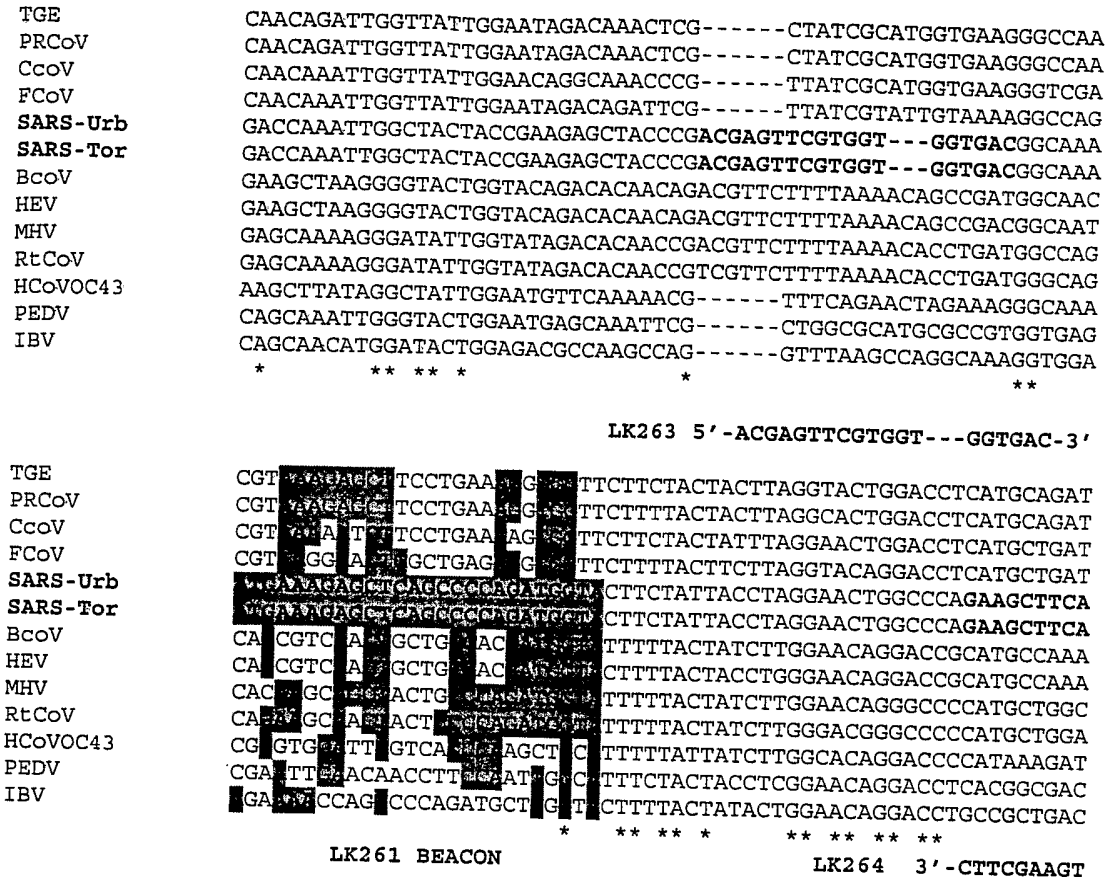
LK262 Amplicon (95 nucleotides)

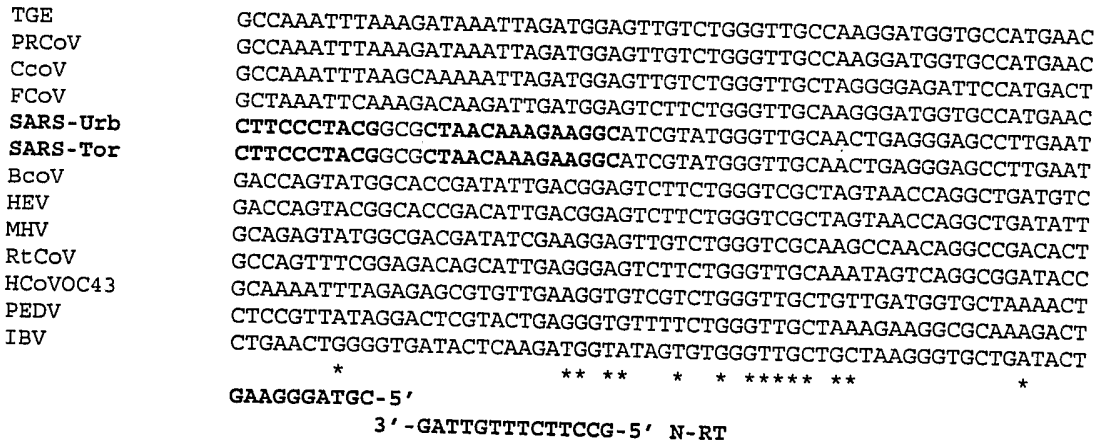
5'–
ACGAGTTCGTGGTGGTGACGGCAAAATGAAAGAGCTCAGCCCCAGATGGTAC
TTCTATTACCTAGGAACTGGCCCAGAAGCTTCACTTCCCTACG -3'

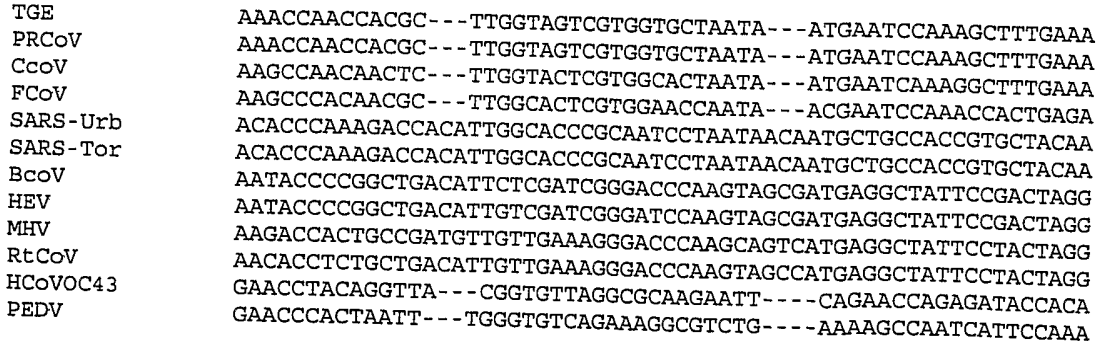
DNA sequence alignment of complete N genes from corona virus strains

```
TGE          ------------------------------------------------------------
PRCoV        ------------------------------------------------------------
CcoV         ------------------------------------------------------------
FCoV         ------------------------------------------------------------
SARS-Urb     ------------------------ATGTCTGATAATGGACCCCAATCAAACCAACGTAGT
SARS-Tor     ------------------------ATGTCTGATAATGGACCCCAATCAAACCAACGTAGT
BcoV         ATGTCTTTTACTCCTGG-TAAGCAAT--CCAGTAGTAGAGCGTCCTTTGGAAATCGTTCT
HEV          ATGTCTTTCACTCCTGG-CAAGCAGT--CCAGCAGTAGAGCGTCCTCTGGAAATCGTTCT
MHV          ATGTCTTTTGTTCCTGGGCAAGAAAATGCCGGTAGCAGAAGCTCCTCTGGAAACCGCGCT
RtCoV        ATGTCTTTTGTTCCCGGACAAGAAAACGCCGGTAGCAGAAGCTCCTCTGGAAACCGCGCT
HCoVOC43     ------------------------------------------------------------
PEDV         ------------------------------------------------------------
```

```
IBV         ------------------------------------------------------------

TGE         ---------------------------------ATGGCCAA-CCAGGGACAA--------
PRCoV       ---------------------------------ATGGCCAA-CCAGGGACAA--------
CcoV        ---------------------------------ATGGCCTC-TCAGGGACAA--------
FCoV        ---------------------------------ATGGCCAC-ACAGGGACAA--------
SARS-Urb    GCCCCCCGCATTACATTTGGTGGACCCACAGATTCAACTGA--CAATAAC----------
SARS-Tor    GCCCCCCGCATTACATTTGGTGGACCCACAGATTCAACTGA--CAATAAC----------
BcoV        GGTAATGGCAT--CCTTAAGTGGGCCGATCAGTCCGACCAATCTAGAAATGT--------
HEV         GGTAATGGCAT--CCTTAAGTGGGCCGATCAGTCCGACCAGTCTAGAAATGT--------
MHV         GGTAATGGCAT--CCTCAAGAAGACCACTTGGGCTGACCAAACCGAGCG-----------
RtCoV       GGTAATGGAAT--CCTCAAGAAGACCACTTGGGCTGACCAAACCGAGCGCGGACAAAATA
HCoVOC43    ------------------------------------------------------------
PEDV        ------------------------------------------------------------
IBV         ---------------------------------ATGGCAAGCGGTAAAGCAGC-------

TGE         -CGTGTCAGTTGGGGAGATGAATCTACCAAAACACGTGGTCGT-TCCAATTC---CCGTG
PRCoV       -CGTGTCAGTTGGGGGGATGAATCCACCAAAATACGTGGTCGC-TCCAATTC---CCGTG
CcoV        -CGTGTCAGTTGGGGAGATGAATCCACCAAGAGACGCGGTCGT-TCTAATTC---TCGTG
FCoV        -CGCGTCAACTGGGGAGATGAACCTTCCAAAAGACGTGGTCGT-TCTAACTC---TCGTG
SARS-Urb    -CAGAATGGAGGACGCAATGGGGCAAGGCCAAAACAGCGCCGA-CCCCAAGG----TTTA
SARS-Tor    -CAGAATGGAGGACGCAATGGGGCAAGGCCAAAACAGCGCCGA-CCCCAAGG----TTTA
BcoV        -TCAAACCAGGGGTAGAAGAGCTCAACCCAAGCAAACTGCTAC-TTCTCAGCTACCATCA
HEV         -TCAAACCAGGGGTAGAAGAGTTCAATCCAAGCAAACTGCTAC-TTCTCAGCTACCATCA
MHV         -TGGAAATAGAGGCAGAAGGAACCATCCCAAGCAGACTGCAAC-TACTCAGC--CCAATG
RtCoV       ATGGAAATAGAGGCAGAAGGAATCAGCCCAAGCAGACTGCAAC-TACTCAGC--CCAATA
HCoVOC43    ----------------ATGGCTACAGTCAAATGGGCTGATGCATCTGAACCACAACGTG
PEDV        ------------------------------------ATGGCTTCTGTCAGCTTTCAGG
IBV         -TGGAAAAACAGACGCCCCAGCGCCAGTCATTAAACTAGGAGGACCAAAACCACCTAAAG

TGE         GTCGGAAGAATAATAACATACC-TCTTTCATTCTTCAACCCCATAACCCTCCAACAAGGT
PRCoV       GTCGGAAGATTAATAACATACC-TCTTTCATTCTTCAACCCCATAACCCTCCAGCAAGGT
CcoV        GCCGGAAGAATAATGATATACC-TCTTTCATTCTTCAACCCCATTACCCTCGAGCAAGGA
FCoV        GTCGGAAGAATAATGATATACC-TTTGTCATTCTACAACCCCATTACCCTCGAACAAGGA
SARS-Urb    CCCAATAATACTGCG---------TCTTGGTTCACAGCTCTCACTCAGCATGGCAAGG--
SARS-Tor    CCCAATAATACTGCG---------TCTTGGTTCACAGCTCTCACTCAGCATGGCAAGG--
BcoV        GGAGGGAATGTTGTACCCTACTATTCTTGGTTCTCTGGAATTACTCAGTTTCAAAAAGGA
HEV         GGAGGGACTGTTGTACCCTACTATTCTTGGTTCTCTGGAATTACTCAGTTTCAAAAGGGA
MHV         CC-GGGAGTGTGGTTCCCCATTACTCTTGGTTTTCGGGCATCACCCAGTTTCAAAAGGGA
RtCoV       CC-GGGAGTGTGGTTCCCCATTACTCTTGGTTTTCGGGCATTACCCAATTCCAGAAGGGA
HCoVOC43    GTCGTCAGGGTAGAA---TACC-TTATTCTCTTTATAGCCCTTTGCTTGTTGATAGTG--
PEDV        ATCGTGGCCGCAAACGGGTGCCATTA-TCTCTCTATGCCCCTCTTAGGGTTACTAATGAC
IBV         TCGGTTCTTCTGGAAATGCA----TCTTGGTTTCAAGCAATAAAAGCCAAGAAGTTAAAT
                                          *   *

TGE         TCAAAATTTTGGAACTTATGTCCGAGAGACTTTGTACCCAAAGGAATAGGTAACAGGGAT
PRCoV       GCAAAATTTTGGAACTCATGTCCGAGAGATTTTGTACCCAAAGGAATAGGTAATAGGGAT
CcoV        TCAAAGTTTTGGGACTTATGTCCGAGAGACTTTGTACCCAAAGGAATAGGTAATAAGGAT
FCoV        TCTAAATTTTGGAATTTATGTCCGAGAGACCTTGTTCCCAAAGGAATAGGTAATAAGGAT
SARS-Urb    -AGGAACTTAGATTCCCTCGAGGCCAGGGCGTTCCAATCAACACCAATAGTGGTCCAGAT
SARS-Tor    -AGGAACTTAGATTCCCTCGAGGCCAGGGCGTTCCAATCAACACCAATAGTGGTCCAGAT
BcoV        AAGGAGTTTGAATTTGCAGAGGGACAAGGTGTGCCTATTGCACCAGGAGTCCCAGCTACT
HEV         AAGGAGTTTGAATTTGCAGAGGGACAAGGTGTGCCTATTGCACCAGGAGTCCCATCTACT
MHV         AAGGAGTTCCAGTTTGCACAAGGACAGGGAGTGCCTATTGCCAGTGGAATCCCCGCTTCA
RtCoV       AAAGAGTTCCAGTTTGCAGGTGGACAAGGAGTGCCTATTGCCAATGGAATCCCACCTTCT
HCoVOC43    -AACAACCTTGGAAGGTGATACCTCGTAATTTGGTACCCATCAACAAGAAAGACAAAAAT
PEDV        AAGCCCCTTTCTAAGGTACTTGCAAACAACGCTGTACCCACTAACAAGGGGAATAAGGAC
IBV         ACACCTCCGCCCAAGTTTGAAGGTAGCGGTGTTCCTGATAACGAAAACATTAAGCCAAGC
```


TGE
PRCoV
CcoV
FCoV
SARS-Urb
SARS-Tor
BcoV
HEV
MHV
RtCoV
HCoVOC43
PEDV
IBV
LK263 5'-ACGAGTTCGTGGT---GGTGAC-3'
LK261 BEACON
LK264 3'-CTTCGAAGT


TGE
PRCoV
CcoV
FCoV
SARS-Urb
SARS-Tor
BcoV
HEV
MHV
RtCoV
HCoVOC43
PEDV
IBV
GAAGGGATGC-5'
3'-GATTGTTTCTTCCG-5' N-RT


TGE
PRCoV
CcoV
FCoV
SARS-Urb
SARS-Tor
BcoV
HEV
MHV
RtCoV
HCoVOC43
PEDV

```
IBV         AAATCTAGATCCAATCAGGGTACAAGAGATCCTGATAA-GTTTGACCAATACCCACTACG
                 *                   *                    * *

TGE         TTCG---ATGGTAAAGTGCCAGGCGAA---TTTCAACTTGAAGTTAATCAATCAAGAGAC
PRCoV       TTCG---ATGGTAAAGTGCCAGGCGAA---TTTCAACTTGAAGTTAACCAGTCTAGGGAC
CcoV        TTCG---ATGTCAAAGTACCATCAGAA---TTTCACCTTGAAGTGAACCAATTAAGGGAC
FCoV        TTTG---ATGGTAAGATACCGCCACAG---TTTCAGCTTGAAGTGAACCGTTCTAGGAAC
SARS-Urb    CTTCCTCAAGGAACAACATTGCCAAAAGGCTTCTACGCAGAGGGAAGCAGAGGCGGCAGT
SARS-Tor    CTTCCTCAAGGAACAACATTGCCAAAAGGCTTCTACGCAGAGGGAAGCAGAGGCGGCAGT
BcoV        TTTCCGCCTGGCACGGTACTCCCTCAGGGTTACTATATTGAAGG---CTCAGGAAGGTCT
HEV         TTTCCGCCTGGCACGGTACTCCCTCAAGGTTACTATATTGAAGG---CTCAGGAAGGTCT
MHV         TTTGCGCCCGGCACGGTATTGCCTCAGGGCTTTTATGTAGAAGG---CTCGGGAAGGTCT
RtCoV       TTTGCGCCCGGTACGGTATTGCCTCAGGGTTTCTATGTTGAAGG---CTCGGGAAGGTCT
HCoVOC43    CTTC--AATCAAAAGCTCCCAAATGGTGTTACTGTTGTTGAAGAACCTGACTCCCGTGCT
PEDV        ATTC--TCTCAACAGCTCCCCAGTGTA---GTTGAGATTGTTGAACCTAACACACCTCCT
IBV         ATTC----TCGGATGGCGGACCTGATGGTAATTTCCGTTGGGACTTCATTCCCCTGAACC
             *                                        *

TGE         AATTCAAGGTCACGCTCTCAATCTAGATCTC------GGTCTAGAAATAGATCTCAATCT
PRCoV       AACTCAAGGTCACGCTCTCAATCTAGATCGC------GGTCTAGAAACAGATCTCAATCT
CcoV        AATTCAAGGTCTAGGTCTCAATCTAGATCTC------AGTCCAGAAATAGGTCTCAATCT
FCoV        AATTCAAGGTCTGGTTCTCAGTCTAGATCTG------TTTCAAGAAACAGATCTCAATCT
SARS-Urb    CAAGCCTCTTCTCGCTCCTCATCACGTAGTCGCGGTAATTCAAGAAATTCAACTCCTGGC
SARS-Tor    CAAGCCTCTTCTCGCTCCTCATCACGTAGTCGCGGTAATTCAAGAAATTCAACTCCTGGC
BcoV        GCTCCTAATTCCAGATCTACTTCACGCGCATCCAGTAGAGCCTCTAGTGCAGGATCGCGT
HEV         GCTCCTAATTCCAGATCTACTTCGCGTGCACCCAATAGAGCCCCTAGTGCAGGATCGCGT
MHV         GCACCTGCTAGTCGATCTGGTTCGCGGTCAC------AATCCCGTGGGCCAAATAATCGC
RtCoV       GCACCTGCTAGTCGATCTGGTTCGCGGTCAC------AATCCCGTGGGCCAAATAATCGC
HCoVOC43    CCTTCCCGGTCTCAGTCGAGGTCGCAGAGTCGCGGTCGTGGTGAATCCAAACCTCAATCT
PEDV        GCTTCACGTGCAAATTCGCGTAGCAGGAGTCGTGGCAATGGCAACAATAGGTCTAGATCT
IBV         GTGGTAGGAGTGGAAGATCAACAGCAGCTTCATCA-GCAGCAGCTAGTAGAGCACCATCA

TGE         AGAGGCAGGCAACAATTCAATAACAAGAAGGAT-------GACAGTGTAGAACAAGCTGT
PRCoV       AGAGGTAGGCAACAATCCAATAACAAGAAGGAT-------GACAGTGTAGAACAAGCTGT
CcoV        AGAGGAAGGCAACTATCCAATAATAAGAAGGAT-------GACAATGTTGAACAAGCTGT
FCoV        AGAGGAAGACACCATTCCAATAACCAGAA---T-------AATAATGTTGAGGATACAAT
SARS-Urb    AGCAGTAGGGGAAATTCTCCTGCTCGAATGGCT-------AGCGGAGGTGGTGAAACTGC
SARS-Tor    AGCAGTAGGGGAAATTCTCCTGCTCGAATGGCT-------AGCGGAGGTGGTGAAACTGC
BcoV        AG---TAGAGCCAATTCTGGCAACAGAACCCCT-------ACCTCTGGTGTAACACCTGA
HEV         AG---TAGAGCCAATTCTGGCAATAGAACCTCT-------ACCCCTGGTGTAACACCTGA
MHV         GC---TAGAAGCAGTTCCAACCAGCGCCAGCCT-------GCCTCTGCTGTAAAACCTGA
RtCoV       GC---TAGAAGCAGTTCCAACCAGCGCCAGCCT-------GCCTCTACTGTAAAACCTGA
HCoVOC43    CGGAATCCTTCAAGTGACAGAAACCATAACAGTCAG----GATGACATCATGAAGGCAGT
PEDV        CCAAGTAACAACAGAGGCAATAACCAGTCCCGTGGTAATTCACAGAATCGTGGAAATAAC
IBV         CG------TGAAGGTTCGCGTGGTCGTAGAAGT-------GATTCTGGAGATGACCTTAT
                                                     *

TGE         T---CTTGCCGCACTT-----AAAAAGTTAGGTGTTGACACAGAAAAACAACAGCA-ACG
PRCoV       T---CTTGCCGCACTT-----AAAAAGTTAGGTGTTTACACAGAAAAACAACAGCA-ACG
CcoV        T---CTTGCTGCACTC-----AAAAAGTTAGGTGTTGACACAGAAAAACAA---CA-AAG
FCoV        T---GTAGCCGTGCTT-----GAAAAATTAGGTGTT---ACTGACAAACAA-------AG
SARS-Urb    C---CTCGC-G--CTA-----TTGCTGCTAGAC-AGATTGAACCAGCTTGAGAGCA-AAG
SARS-Tor    C---CTCGC-G--CTA-----TTGCTGCTAGAC-AGATTGAACCAGCTTGAGAGCA-AAG
BcoV        T---ATGGCTG--ATC-----AAATTGCTAGTC-TTGTTCTGGCAAAACTTG-GCA-AGG
HEV         C---ATGGCTG--ATC-----AAATTGCTAGTC-TTGTTCTGGCAAAACTTG-GCA-AGG
MHV         C---ATGGCCG--AAG-----AAATTGCTGCTC-TTGTTTTGGCTAAGCTTG-GTA-AAG
RtCoV       T---ATGGCCG--AAG-----AAATTGCTGCTC-TTGTTTTGGCTAATCTAG-GCA-AAG
HCoVOC43    T---GCTGCGGCTCTT-----AAATCTTTAGGTTTTGACAAGCCTCAGGAAAAAGATAAA
PEDV        CAGGGTCGTGGAGCTTCTCAGAACAGAGGAGGCAATAATAATAACAATAACAAGTCTCGT
IBV         TGCTCGTGCAGCAAAG------ATAATCCAGGATCAGCAGAAAAAGGGCTCTCGCA--TT
                   *  *
```

```
TGE             CTCTCGTTCTAAATCTAAAG--AACGTAGTAACTCTAAGAC-AAGA--------GATACT
PRCoV           CTCTCGTTCTAAATCTAAAG--AACGTAGTAACTCTAAAAC-AAGA--------GATACT
CcoV            ATCTCGTTCCAAATCTAAGG--AACGTAGCAGCTCTAAGAC-AAGA--------GATACT
FCoV            GTCACGTTCTAAACCTAGAG--AACGTAGTGATTCCAAACC-TAGG--------GACACA
SARS-Urb        TTTCTGGTAAAGGCCAACAA-CAACAAGGCCAAACTGTCACTAAGA--------AATCTG
SARS-Tor        TTTCTGGTAAAGGCCAACAA-CAACAAGGCCAAACTGTCACTAAGA--------AATCTG
BcoV            ATGCCACTAAGCCACAGCAAGTAACTAAGC-AGACTGCCA--AAGA--------AATCAG
HEV             ATGCCACTAAGCCTCAGCAAGTAACTAAGC-AGACTGCCA--AAGA--------GGTCAG
MHV             ATGCCGGCCAGCCCAAGCAGGTAACTAAGC-AAAGCGCCA--AAGA--------AGTCAG
RtCoV           ATGCCGGACAGCCTAAGCAAGTAACTAAGC-AAAGTGCCA--AAGA--------AGTCAG
HCoVOC43        AAGTCAGCGAAAACGGGTAC--TCCTAAGCCTTCTCGTAATCAGAGT----CCTGCTTCT
PEDV            AACCAGTCCAATAACAGGAACCAGTCAAATGACCGTGGTGGTGTAACATCACGCGATGAT
IBV             ACCAAGGCAAAGGCAGATGA--AATGGCTCATCGCCGGTATTGCAAG----CGCACTATC

TGE             ACACCTAAGAATGAAAACAAACACACCTGGA---AGAGAACTGCAGGTAAA---GGTG--
PRCoV           ACGCCTAAGAATGAAAACAAACACACCTGGA---AGAGAACTGCAGGTAAA---GGTG--
CcoV            ACACCTAAGAATGAAAACAAACACACCTGGA---AGAGAACTGCAGGTAAA---GGTG--
FCoV            ACACCTAAGAATGCCAACAAACACACCTGGA---AGAAAACTGCAGGCAAG---GGAG--
SARS-Urb        CTGCTGAGGCATCTAAA-AAGCCTCGCCAAA---AACGTACTGCCACAAAAC--AGTACA
SARS-Tor        CTGCTGAGGCATCTAAA-AAGCCTCGCCAAA---AACGTACTGCCACAAAAC--AGTACA
BcoV            --ACAGAAAATTTTGAATAAGCCCCGCCAGA---AGAGGAGCCCCAATAAAC--AATGCA
HEV             --ACAGAAAATCTTGAATAAGCCCCGCCAGA---AGAGGAGCCCCAACAAAC--AATGCA
MHV             --GCAGAAAATTTTAACTAAGCCTCGTCAAA---AGAGGACTCCAAACAAGC--AGTGCC
RtCoV           --GCAGAAAATTTTAAATAAGCCTCGCCAAA---AGAGGACTCCAAACAAGC--AGTGCC
HCoVOC43        TCTCAAACTTCTGCCAAGAGTCTTGCTCGTTCTCAGAGTTCTGAAACAAAAG--AACAAA
PEDV            CTGGTGGCTGCTGTCAAGGATGCACTTAAATCTTTGGGTATTGGAGAAAATCCTGACAGG
IBV             CCACCTAATTATAGGGTTGATCAAGTGTTTGGTCCCCGTACTAAAGGTAAGG--AGGGGA
                                                                **

TGE             ---ATGTGACA---AGATTTTATGGAGCTAGAAGCAGTT--CAGCC-------AAT-TTT
PRCoV           ---ATGTGACA---AGATTTTATGGAGCTAGAAGCAGCT--CAGCC-------AAT-TTT
CcoV            ---ATGTGACA---AAATTTTATGGAGCTAGAAGTAGTT--CAGCC-------AAT-TTT
FCoV            ---ATGTGACA---ACTTTCTATGGTGCTAGAAGTAGTT--CAGCT-------AAC-TTT
SARS-Urb        A-CGTCACTCA---AGCATTTGGGAGACGTGGTCCAGAA--CAAACCCAAGGAAAT-TTC
SARS-Tor        A-CGTCACTCA---AGCATTTGGGAGACGTGGTCCAGAA--CAAACCCAAGGAAAT-TTC
BcoV            C-TGTTCAGCA---GTGTTTTGGGAAGAGAGGCCCCAAT--CAGA---------AT-TTT
HEV             C-TGTTCAGCA---GTGTTTTGGGAAGAGAGGCCCCAAT--CAGA---------AT-TTT
MHV             C-AGTGCAGCA---GTGTTTTGGGAAGAGAGGCCCTAAT--CAGA---------AC-TTT
RtCoV           C-AGTGCAGCA---GTGTTTTGGAAAGAGAGGCCCCAAT--CAGA---------AT-TTT
HCoVOC43        AGCATGAAATGC--AAAAGCCACGGTGGAAAAGACAGCC--TAATGATGATGTGACATCT
PEDV            CATAAGCAACAGCAGAAGCCTAAGCAGGAAAAGTCTGACAACAGCGGCAAAAATACACCT
IBV             A--TTTTGGTG-----ATGACAAGATGAATGAGGAAGGTATTAAGGATG----GGCGTGT
                                                      *

TGE             GGTGACACTGACCTCGTTGCCAAT----GGGAGCAGTGCCAAGCATTACCCACAACTGGC
PRCoV           GGTGACAGTGACCTCGTTGCCAAT----GGGAGCAGTGCCAAGCATTACCCACAATTGGC
CcoV            GGTGACAGCGATCTTGTTGCCAAT----GGGAACGGTGCCAAGCATTACCCACAACTGGC
FCoV            GGTGATAGTGATCTCGTTGCCAAT----GGTAACGCTGCCAAATGCTACCCTCAGATAGC
SARS-Urb        GGGGACCAAGACCTAATCAGACAA----GGAACTGATTACAAACATTGGCCGCAAATTGC
SARS-Tor        GGGGACCAAGACCTAATCAGACAA----GGAACTGATTACAAACATTGGCCGCAAATTGC
BcoV            GGTGGTGGAGAAATGTTAAAACTT----GGAACTAGTGACCCACAGTTCCCCATTCTTGC
HEV             GGTGGTGGAGAAATGTTAAAACTT----GGAACTAGTGACCCACAGTTCCCCATTCTTGC
MHV             GGAGGCTCTGAAATGTTAAAACTT----GGAACTAGTGATCCGCAGTTCCCCATTCTTGC
RtCoV           GGAGGCCCTGAAATGTTAAAACTT----GGAACTAGTGATCCACAGTTCCCCATTCTTGC
HCoVOC43        AATGTCACACAATGTTTTGGCCCC----AGAGACCTTGAC---CACAACTTTGGAAGTGC
PEDV            AAGAAGAACAAATCCAGGGCCACTTCGAAGGAACGTGACCTCAAAGACATCCCAGAGTGG
IBV             TACAGCAATGCTCAACCTAGTCCCT---AGCAGCCATGCT-----TGTCTTTTTGGAAGT
                                          *                             *

TGE             TGAATGTGTT-CCATCTGTGTC--TAGCATT----CT--GTTTGGAA-GCTATTGGACTT
```

```
PRCoV          TGAATGTGTT-CCATCTGTGTC--TAGCATT----TT--GTTTGGAA-GCTATTGGACTT
CcoV           TGAATGTGTT-CCATCTGTATC--TAGCATT----CT--GTTTGGAA-GCCATTGGACTG
FCoV           TGAATGTGTT-CCATCAGTGTC--TAGCATA----AT--CTTTGGCA-GTCAATGGTCTG
SARS-Urb       ACAATTTGCT-CCAAGTGCCTC--TGCA-------TT--CTTTGGAATGTCAC--GCATT
SARS-Tor       ACAATTTGCT-CCAAGTGCCTC--TGCA-------TT--CTTTGGAATGTCAC--GCATT
BcoV           AGAACTCGCA-CCCACAGCTGG--TGCGTTT----TT--CTTTGGATCAAGATTAGAGTT
HEV            AGAACTCGCA-CCCACAGCTGG--TGCGTTT----TT--CTTTGGATCAAGATTAGAGTT
MHV            AGAGTTGGCT-CCAACACCTAG--TGCCTTC----TT--CTTTGGATCTAAATTAGAATT
RtCoV          AGAGTTGGCC-CCAACACCTGG--TGCCTTC----TT--CTTTGGATCTAAATTAGAATT
HCoVOC43       AGGTGTTGTGGCCAATGGTGTT--AAAGCTAAAGGCT--ATCCACAATTTGCTGAGCTTG
PEDV           AGGAGAATTC-CCAAGGGCGAAAATAGCGTAGCAGCTTGCTTCGGACCCAGAGGGGGCTT
IBV            AGAGTGACACCCAAACTTCAAC--TAGATGG--GCTTCACTTGAGATTTGAATTTACTAC
                          *                          *   *

TGE            CAAAGGAA--GATGGCG----ACCAGATAGAAGTCAC----GTTCACACACAAATACC--
PRCoV          CAAAGGAA--GATGGCG----ACCAGATAGAAGTCAC----GTTCACACACAAATACC--
CcoV           CTAAGGAA--GATGGTG----ACCAGATTGAAGTCAC----ATTCACACACAAATACC--
FCoV           CTGAAGAA--GCTGGTG----ATCAAGTGAAAGTCAC----GCTCACTCACACCTACT--
SARS-Urb       GGCA----------------------TGGAAGTCAC----ACCTTCGGGAACATG----
SARS-Tor       GGCA----------------------TGGAAGTCAC----ACCTTCGGGAACATG----
BcoV           GGCCAAAG--TGCAGAA----TTTGTCTGGGAATCTTGACGAGCCCCAGAAGGATGTTTA
HEV            GGCCAAAG--TGCAGAA----TTTGTCTGGGAATCCTGACGAGCCCCAGAAGGATGTTTA
MHV            GGTCAAAA-----AGAA----CT---CTGGTGGTGCTGATGAACCCACCAAAGATGTTTA
RtCoV          GGTCAAAA-----AGAA----TT---CTGGTGGCGTTGATGAACCCACCAAAGATGTGTA
HCoVOC43       TGCCGTCA--ACAGCTGCTATGCTGTTTGATAGTCAC-ATTGTTTCCAAAGAGTCAGGCA
PEDV           CAAAAACTTTGGAGATGCGGAATTTGTCGAAAAAGGTGTTGATGCGTCAGGCTATGCTCA
IBV            TGTGGTCCCATGTGATGACCCGCAGTTTGATAATTATGTGAAAATTTGTGATCAGTGT--

TGE            --ACT-TGCCAAAGGATGA--TCCTA--AGACTGGACAATTCCTTCAGCAGATTAATGCC
PRCoV          --ACT-TGCCAAAGGATCA--TCCTA--AAACTGAACAATTCCTTCAGCAGATTAATGCC
CcoV           --ACT-TGCCAAAGGATGA--TCCTA--AGACTGGACAATTCCTTCAGCAGATTAATGCA
FCoV           --ACC-TGCCAAAGGATGA--TGCCA--AAACTAGTCAATTCCTAGAACAGATTGACGCT
SARS-Urb       ---GC-TGACTTATCATGG--AGCCATTAAATTGGATGACAAAGATCCACAATTCAAAGA
SARS-Tor       ---GC-TGACTTATCATGG--AGCCATTAAATTGGATGACAAAGATCCACAATTCAAAGA
BcoV           TGAAT-TGCGCTATAATGG--TGCAATTAGATTTGACAGTACACTTTCAGGTTTTGAGAC
HEV            TGAAT-TGCGCTATAATGG--CGCGATTAGATTTGACAGCACACTCTCAGGTTTTGAAAC
MHV            TGAAT-TGCAGTATTCAGG--TGCAATTAGATTTGATAGTACTCTACCCGGTTTTGAGAC
RtCoV          TGAGC-TGCAATATTCAGG--TGCAGTCAGATTTGATAGTACTCTACCTGGTTTTGAGAC
HCoVOC43       ACACTGTGGTCTTGACTTT--CACTACTAGAGT-GACTGTGCCCAAAGACCATCCACACT
PEDV           GATCGCCAGTTTAGCACCAAATGTTGCAGCATTGCTCTTTGGTGGTAATGTGGCTGTTCG
IBV            -GTCGATGGTGTAGGAACGCGTCCAA---AAGATGACGA-ACCAAAACCAAAGTCACGCT
                                           *

TGE            TATGCTCG--TCCATCAGAAGTGGCAAAAGAA--CAGAGA-AAAAGA-----AAATCTCG
PRCoV          TATGCTAG--CCCATCAGAATTGGCAAAAGAA--CAGAGA-AAAAGA-----AAGTCTCG
CcoV           TACGCCCG--TCCATCAGAGGTGGCTAAAGAA--CAGAGA-CAACGC-----AAAGCTCG
FCoV           TACAAGCG--ACCTTCTGAAGTGGCTAAGGAT--CAGAGG-CAAAGA-----AGATCCCG
SARS-Urb       CAACGTCA--TACTGCTGAACAAGCACATTGA--CGCATA-CAAAAC-----ATTCCCAC
SARS-Tor       CAACGTCA--TACTGCTGAACAAGCACATTGA--CGCATA-CAAAAC-----ATTCCCAC
BcoV           CATAATGA--AGGTGTTGAATGAGAATTTGAA--TGCATATCAACAA-----CA---AGA
HEV            CATTATGA--AGGTGCTTAACCAGAATTTGAA--TGCCTATCAACAT-----CAGGAAGA
MHV            TATCATGA--AAGTGTTGACTGAGAATTTGAA--TGCCTACCAGGAC-----CAAGCTGG
RtCoV          TATCATGA--AAGTGTTGAATGAGAATTTGAA--TGCCTACCAGAAT-----CAAGCTGG
HCoVOC43       TG-GGTAA--GTTTCTTGAGGAGTTAAATGCATTCACTAGAGAAATG-----CAACAACA
PEDV           TGAGCTAGCGGACTCTTACGAGATTACATACAACTATAAAATGACTGTGCCAAAGTCAGA
IBV            CAAGTTCAAGACCTGCTACAAGAGGAAATTCT--CCAGCGCCAAGACAACAGCGCCCAAA

TGE            TTCTAAATCTGCAGAAAGGTCAGAGCAAGATGTGGTACCTGATGCATTAATAGAA-AATT
PRCoV          TTCTAAATCTGCAGAAAGGTCAGAGCAAGAGGTGGTACCTGATTCATTAATAGAA-AACT
CcoV           TTCTAAATCTGTAGAAAGGGTAGAGCAAGAGGTTGTACCTGATGCATTAACAGAA-AATT
```

```
FCoV          TTCTAAGTCTGCTGATAAG---AAGCCTGAGGAGTTGTCTGTAACTCTTGTGGAG-GCAT
SARS-Urb      CAACAGAGCCTAAAAAGGACAAAAAGAAAAAGACTGATGAAGCTCAGCCTTTGCC-GCAG
SARS-Tor      CAACAGAGCCTAAAAAGGACAAAAAGAAAAAGACTGATGAAGCTCAGCCTTTGCC-GCAG
BcoV          TGGTATGATGAATATGAGTCCAAAACCACAGCGTCAGCGTGGTCAGAAG-------AATG
HEV           TGGGATGATGAATATTAGTCCTAAACCACAGCGGCAGCGTGGTCAGAAG-------AATG
MHV           TAGTGTAGATCTAGTGAGCCCAAAGCCTCCAAGAAGAGGTCGTAGACAGGCTCAA-GAAA
RtCoV         TGGTGCAGATGTAGTGAGCCCAAAGCCCCAAAGAAAGAGAGGGACGAAACAAACG-GCTC
HCoVOC43      TCCTCTTCTTAACCCTAGTGCACTAGAATTCAACCCATCTCAAACTTCACCTGCA-ACTG
PEDV          TCCAAATGTTGAGCTTCTTGTTTCACAGGTGGATGCATTTAAAACTGGGAATGCAAAACT
IBV           GAAGGAGAAAAAGCTAAAGAAGCAGGATGATGAAGCAGATAAAGCATTGACCTCAGATGA

TGE           ATACAGATGTGTTTGAT-GACACACAGGT-TGAGATAATTGATGAGGTA-ACGAACTAA-
PRCoV         ATACAGATGTGTTTGAT-GACACACAGGT-TGAGATGATTGACGAGGTA-ACGAACTAA-
CcoV          ACACAGATGTGTTTGAT-GACACACAGGT-TGAGATTATTGATGAGGTA-ACGAACTAA-
FCoV          ACACAGATGTGTTTGAT-GACACACAGGT-TGAGATGATTGATGAGGTT-ACGAACTAA-
SARS-Urb      AGACAAAAGAAGCAGCCCACTGTGACTCT-TCTT---CCTGCGGCTGAC-ATG-GATGAT
SARS-Tor      AGACAAAAGAAGCAGCCCACTGTGACTCT-TCTT---CCTGCGGCTGAC-ATG-GATGAT
BcoV          GACAAGGAGAAAATGAT-AATATAAGTGT-TGCAGCGCCTAAAAGCCGT-GTGCAGCAAA
HEV           GACAAGTAGAAAATGAT-AATGTAAGTGT-TGCAGCGCCTAAAAGCCGT-GTGCAGCAAA
MHV           AGAAAGATGAAGTAGAT-AATGTAAGCGT-TGCAAAGCCCAAAAGCTTG-GTGCAGCGAA
RtCoV         AGAAAGAAGAATTAGAT-AGTATAAGCGT-TGCAAAGCCCAAAAGTGCC-GTGCAGCGAA
HCoVOC43      CTGAACCAGTGCGTGAT-GAAGTTTCTAT-TGAAA--CTGACATAATTG-ATGAAGTAAA
PEDV          CCAGAGAAAGAAGGAAAAGAAGAACAAGCGTGAAACCACGCTGCAGCAGCATGAAGAGGC
IBV           GGAGAGGAACAATGCACAGCTGGAATTTTATGATGAGCCCAAGGTAATTAACTGGGGGGA
                   *                          *

TGE           ------------------------------------------------------------
PRCoV         ------------------------------------------------------------
CcoV          ------------------------------------------------------------
FCoV          ------------------------------------------------------------
SARS-Urb      TTCTCCAGACAACTTCAAAATTCCATGAGTGGAGCTTCTGCTGATTCAACTCAGGCATAA
SARS-Tor      TTCTCCAGACAACTTCAAAATTCCATGAGTGGAGCTTCTGCTGATTCAACTCAGGCATAA
BcoV          ATAAGAGTAGAGAGTTGACTGCAGAGGACATCAGC--CTTCTTAAGAAGAT---GGATGA
HEV           ATAAGAGTAGAGAGTTGACAGCAGAGGACATCAGC--CTTCTTAAGAAGAT---GGATGA
MHV           ATGTAAGTAGAGAATTAACCCCCGAGGATCGTAGC--CTGCTGGCTCAGATCCTAGACGA
RtCoV         ATGTAAGCAGAGAATTAACCCCAGAGGATAGAAGC--CTGTTGGCGCAGATCCTAGATGA
HCoVOC43      CTAA------------------------------------------------------
PEDV          CATCTACGATGATGTGGGTGCGCCATCTGATGTGACCCATGCCAATCTGGAATGGGACAC
IBV           TGCAGCTCTAGGAGAGAATGAACTTTGA--------------------------------

TGE           ------------------------------------------------------------
PRCoV         ------------------------------------------------------------
CcoV          ------------------------------------------------------------
FCoV          ------------------------------------------------------------
SARS-Urb      ------------------------------------------------------------
SARS-Tor      ------------------------------------------------------------
BcoV          --GCCCTATACTGAA----------GACACCTCAGAAATATAA-----------------
HEV           --GCCCTATACTGAA----------GATACCTCAGAAATATAA-----------------
MHV           TGGCGTTGTGCCAGATGGGTTGGAAGATGACTCTAATGTGTAA-----------------
RtCoV         TGGCGTTGTGCCTGATGGGTT---AGATGACTCTAATGTGTAA-----------------
HCoVOC43      ------------------------------------------------------------
PEDV          AGCTGTTGATGGTGGTGATACGGCCGTTGAAATTATCAACGAGATCTTCGATACAGGAAA
IBV           ------------------------------------------------------------

TGE           ----
PRCoV         ----
CcoV          ----
FCoV          ----
SARS-Urb      ----
SARS-Tor      ----
```

BcoV	----
HEV	----
MHV	----
RtCoV	----
HCoVOC43	----
PEDV	TTAA
IBV	----

Figure 16: List of Molecular Designs for *N* Gene

LK261 Dubcyl 5'-FAM-CCTCCGTACCATCTGGGGCTGAGCTCTTTCATCGGAGG-3'-

LK261.N Dubcyl 5'-FAM-GCCTCCGTACCATCTGGGGCTGAGCTCTTTCATCGGAGGC-3'-

LK262

5'-ACGAGTTCGTGGTGGTGACGGCAAAATGAAAGAGCTCAGCCCCAGAT GGTACTTCTATTACCTAGGAACTGGCCCAGAAGCTTCACTTCCCTACG-3'

LK263 5'- ACGAGTTCGTGGTGGTGAC -3'

LK263-T7 5'-TAATACGACTCACTATAGGACGAGTTCGTGGTGGTGAC -3'

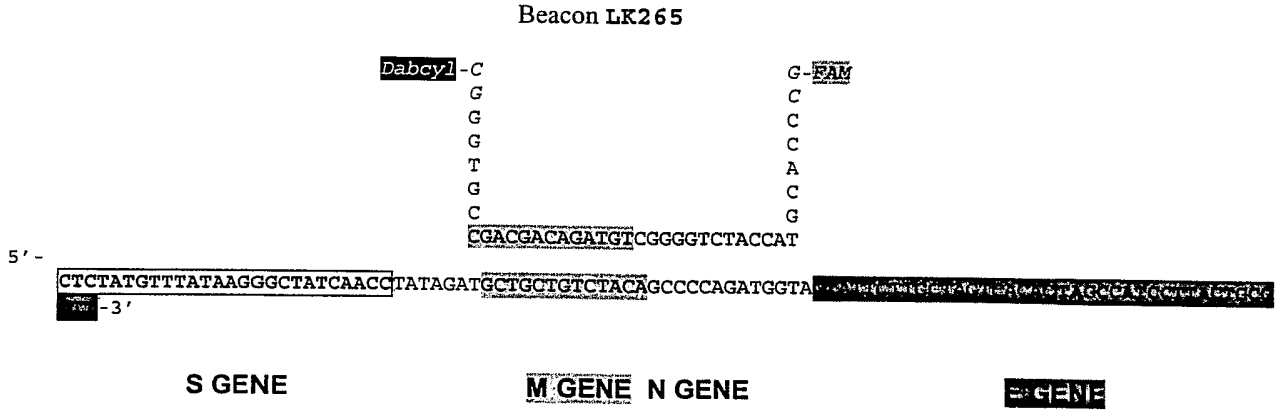
LK264 5'- CGTAGGGAAGTGAAGCTTC- 3'

LK264-RT 5'- GCCTTCTTTGTTAGCGTAGGGAAGTGAAGCTTC- 3'

N-RT 5' –GCCTTCTTTGTTAG-3'

Figure 17: Molecular Designs for Internal Positive Control (IPC)

NOTE:	LK265 is a TET-molecular beacon, which recognizes an artificial target (scrambled amplicon from S, E M and N genes of **coronavirus (SARS-associated coronavirus Tor2 and SARS strains)**. It will be used as an internal positive control (IPC) in a real-time PCR diagnostic for the identification of **SARS-associated coronavirus RNA/DNA.**



Amplicon LK266 (S, M, N and E hybrid)

LK249

Q.	Target recognition sequence:	**24 nucleotides (11 G/C)**
R.	Length of the arms:	**6 nucleotides (5 G/C)**
S.	Melting temperature of the beacon:	**dG = -2.37 dH = -46.3 dS = -135.9 Tm = 67.7 °C**
T.	Melting temperature of target:	**°C**

LK251 **5'-CTCTATGTTTATAAGGGCTATCAACC-3'**

Length: 26 nucleotides (10 G/C)
Tm: °C
Position: (see alignment below)

LK256 **5' –AAGCGCAGTAAGGATGGCTA-3'**

Length: 20 nucleotides (10 G/C)
Tm: °C
Position: (see alignment below)

LK266 Amplicon (98 nucleotides)

5'-
CTCTATGTTTATAAGGGCTATCAACCTATAGATGCTGCTGTCTACAGCCCCAGA
TGGTAGTATTCTTGCTAGTCACACTAGCCATCCTTACTGCGCTT-3'

Figure 18: List of Molecular Designs for IPC

LK265 5'-FAM-GCCCACGTACCATCTGGGGCTGTAGACAGCAGCCGTGGGC-3'-Dubcyl

LK266

5'-CTCTATGTTTATAAGGGCTATCAACCTATAGATGCTGCTGTCTACAGC CCCAGATGGTAGTATTCTTGCTAGTCACACTAGCCATCCTTACTGCGCTT -3'

LK251 5'-CTCTATGTTTATAAGGGCTATCAACC-3'

LK251-T7 5'-TAATACGACTCACTATAGGCTCTATGTTTATAAGGGCTATCAACC-3'

LK256 5'- AAGCGCAGTAAGGATGGCTA - 3'

LK256-RT 5'- TATTGCAGCAGTACAAGCGCAGTAAGGATGGCTA - 3'

E-RT 5' –TATTGCAGCAGTAC-3'

Figure 19: Molecular Beacon Melting Curves

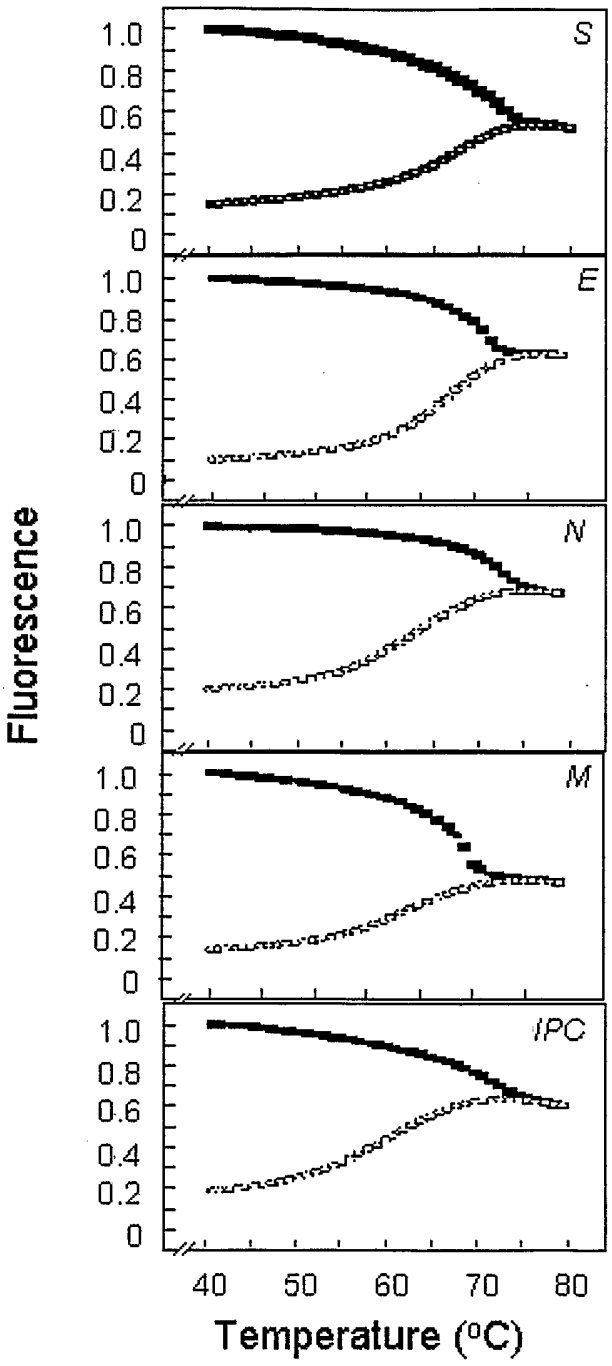
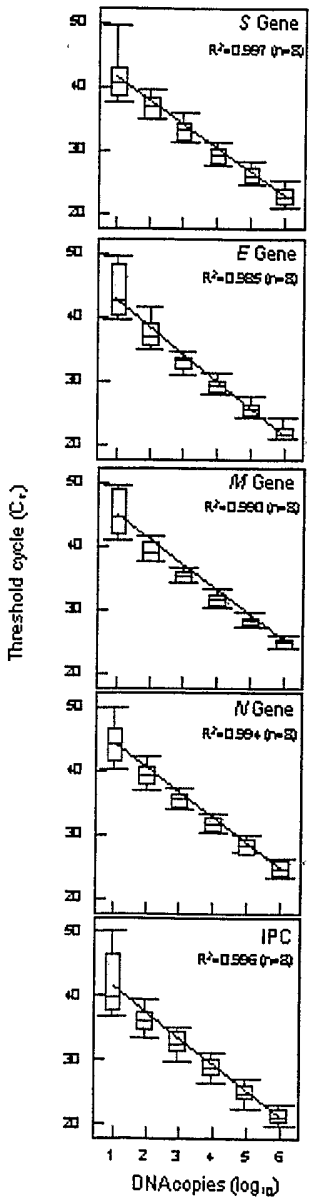


Figure 20: Uniplex Real-time PCR Amplifications (Serial Dilutions-Dynamic Range)



SEQUENCE LISTING

<110> BIRCH BIOMEDICAL RESEARCH, INC.

<120> MULTI-ALLELIC MOLECULAR DETECTION OF SARS-ASSOCIATED CORONAVIRUS

<130> 42392-203686

<140>
<141>

<150> 60/496,995
<151> 2003-08-22

<160> 94

<170> PatentIn Ver. 3.2

<210> 1
<211> 4356
<212> DNA
<213> Canine coronavirus

<400> 1
atgattgtgc ttacattgtg ccttttcttg tttttgtaca gtagtgtgag ctgtacatca 60
aacaatgact gtgtacaagt taatgtgaca caactgcctg gcaatgaaaa tattatcaaa 120
gattttctat ttcagaactt taaagaagaa ggaagtttag ttgttggtgg ttattacccc 180
acagaggtgt ggtataactg ttccacaact caacaaacta ccgcttataa gtattttagt 240
aatatacatg cattttattt tgatatggaa gccatggaga atagtactgg caatgcacgt 300
ggtaaacctt tactagtaca tgttcatggt aatcctgtta gtatcattgt ttacatatca 360
gcttatagag atgatgtgca atttaggccg cttttaaagc atggtttatt gtgtataact 420
aaaaatgaca ccgttgacta taatagcttt acaattaacc aatggcgaga catatgtttg 480
ggtgacgaca gaaaaatacc attctctgta gtacccacag ataatggtac gaaattattt 540
ggtcttgagt ggaatgatga ctatgttaca gcctatatta gtgatgagtc tcaccgtttg 600
aatatcaata ataattggtt taacaatgtt acactcctat actcacgtac aagcaccgcc 660
acgtggcaac acagtgctgc atatgtttat caaggtgttt caaattttac ttattacaag 720
ttaaataaaa ccgctggctt aaaaagctat gaattgtgtg aagattatga atactgcact 780
ggctatgcaa ccaatgtgtt tgctccgaca tcaggtggtt atatacctga tggattcagt 840
tttaacaatt ggtttatgct tacaaacagc tccacttttg ttagtggcag atttgtaaca 900
aatcaaccgc tgctagttaa ttgcttgtgg ccagtgccca gttttggcgt cgcagcacaa 960
gaattttgtt ttgaaggtgc tcagtttagc caatgtaacg gtgtttcttt aaataataca 1020
gtagatgtta ttagatttaa ccttaatttc actacagatg tacaatctgg catgggtgct 1080
acagtatttt cactgaatac aacaggcggt gtcattcttg agatttcttg ttataatgac 1140
acagtgagtg agtcgagttt ctacagttat ggtgaaattc cattcggcgt aactgatgga 1200
ccacgttact gttatgtact ctacaatggc acagctctta agtatttagg aacattacca 1260
cctagtgtca aggaaattgc tattagtaag tggggacatt tttatattaa tggttacaat 1320
ttctttagca cgtttcctat tgattgtata gcttttaatt taaccactgg tgctagtgga 1380
gcattttgga caattgctta tacgtcgtac acagaagcat tagtacaagt tgaaaacaca 1440
gctattaaaa aggtgacgta ttgtaacagt cacattaata acatcaaatg ttctcaactt 1500
actgctaatt tgcaaaatgg tttttaccct gttgcttcaa gtgaagttgg tcttgtcaat 1560
aagagtgttg tgttactacc tagtttctat tcacatacca gtgttaatat aactattgat 1620
cttggtatga agcgtagtgt tacggtcacc atagcctcac cattaagtaa catcacacta 1680
ccaatgcagg ataataacat agacgtgtac tgtattcgtt ctaaccaatt ctcagtttat 1740
gttcattcca cttgcaaaag ttctttatgg gataacaatt ttaattcagc atgtaccgac 1800
gttttagacg ccacagctgt tataaaaact ggtacttgtc ctttctcatt tgataaattg 1860
aataattact taacttttaa caagttctgt ttgtcgttga atcccgttgg tgccaactgt 1920
aagttagatg ttgccgcccg tacaagaacc aatgagcagg tttttggaag tttatatgta 1980

atatatgaag aaggagacaa catagtgggt gtaccgtctg ataatagtgg tttgcacgat 2040
ttgtcagtgt tgcacttaga ctcttgtaca gattacaata tatatggtag aactggtgtt 2100
ggtattatta gaaaaactaa cagcacacta cttagtggct tatattacac atcactatca 2160
ggtgatttgt taggttttaa aaatgttagt gatggtgttg tctactctgt aacgccatgt 2220
gatgtaagtg cacaagctgc tgttattgat ggtgccatag ttggagctat gacttccatt 2280
aatagtgaac tgttaggtct aactcattgg acaacaacac ctaattttta ttactactcc 2340
atatataatt atacaaatgt gatgaatcgt ggcacggcaa ttgataatga tattgattgt 2400
gaacctatca taacatattc taatataggt gtttgtaaaa atggagcttt ggtttttatt 2460
aacgtcacac attctgatgg agacgttcaa ccaattagca ccggtaatgt cacgataccc 2520
acaaatttta ctatatctgt gcaagtcgaa tatattcagg tttacactac accagtttca 2580
atagactgtg caagatacgt ttgcaatggt aacccaagat gcaataagtt attaacacaa 2640
tacgtttctg catgtcaaac tattgagcaa gcgcttgcaa tgggtgccag acttgaaaac 2700
atggagattg attccatgtt atttgtttcg gaaaatgccc ttaaattggc atctgttgaa 2760
gcattcaata gtacggaaaa tttagaccct atttataaag aatggcctaa cattggtggt 2820
tcttggctag gaggtttaaa agatatattg ccatctcata atagcaaacg taagtaccgc 2880
tcggctatag aagacttgct ttttgataag gttgtaacat ctggcttagg tacagttgac 2940
gaagattaca aacgttctgc aggtggttat gacatagctg acttagtgtg tgcacgatat 3000
tacaatggca tcatggtgct acctggtgta gctaatgatg acaagatgac tatgtacact 3060
gcatctctta caggtggtat aacattaggt gcacttagtg gtggcgcagt ggctatacct 3120
tttgcagtag cagttcaggc tagacttaat tatgttgctc tacaaactga tgtattgaac 3180
aaaaaccaac aaatcttggc taatgctttc aatcaagcta ttggtaacat tacacaggca 3240
tttggtaagg ttaatgacgc tatacatcaa acatcaaaag gtcttgctac tgttgctaaa 3300
gcattggcaa aggtgcaaga tgttgttaac acgcaaggtc aagctttaag ccacctaaca 3360
gtacaattgc aaaacaattt tcaagccatt agcagttcta ttagtgacat ttataacagg 3420
cttgatgaat tgagtgctga tgcacaagtt gacaggctga ttacaggacg acttacagca 3480
cttaatgcat ttgtgtctca gactttaacc agacaagcag aggttagggc tagtagacaa 3540
cttgctaaag acaaggttaa tgaatgcgtt aggtctcaat cccagagatt tggattctgt 3600
ggtaatggta cacatttgtt ttcacttgca aatgcggcac caaatggcat gattttcttt 3660
cacacagtgc tattaccaac agcttatgaa actgtgacgg cctggtcagg tatttgtgcg 3720
tcagatggca gtcgcacttt tggacttgtt gttgaggatg tccagctgac gctatttcgc 3780
aatttagatg aaaaatttta tttgacgccc agaactatgt atcagcccag agttgcaact 3840
agttctgatt ttgttcaaat agaaggctgt gatgtgttgt ttgttaatgg aactgtaatt 3900
gaattgccta gtatcatacc tgactatatc gatattaatc aaactgttca ggacatatta 3960
gaaaatttca gaccaaattg gactgtaccc gagttgccac ttgacatttt tcatgcaacc 4020
tacttaaacc tgactggtga aattaatgac ttagaattta ggtcagaaaa gttacataac 4080
accacagtag aacttgctat tctcattgat aatattaata acacattagt caatcttgaa 4140
tggctcaaca gaattgaaac ttatgtaaaa tggccttggt atgtttggct actaattgga 4200
ttagtagtaa tattctgcat acccatattg ctattttgtt gttgtagtac tggttgttgt 4260
ggatgtatcg ggtgtttagg aagctgttgt cattccatat gtagtagagg ccaatttgaa 4320
agttatgaac ctattgaaaa agttcatgtt cactga 4356

<210> 2
<211> 4359
<212> DNA
<213> Feline coronavirus

<400> 2
atgattgtgc tcgtaacttg cctcttgttg ttatgttcat accacacagt tttgagtaca 60
acaaataatg aatgcataca agttaacgta acacaattgg ctggcaatga aaaccttatc 120
agagattttc tgtttagtaa ctttaaagaa gaaggaagtg tagttgttgg tggttattac 180
cctacagagg tgtggtacaa ctgctctaga acagctcgaa ctactgcctt tcagtatttt 240
aataatatac atgcctttta ttttgttatg gaagccatgg aaaatagcac tggtaatgca 300
cgtggtaaac cattattatt tcatgtgcat ggtgagcctg ttagtgttat tatatcggct 360
tatagggatg atgtgcaaca aaggcccctt ttaaaacatg ggttagtgtg cataactaaa 420
aatcgccata ttaactatga acaattcacc tccaaccagt ggaattccac atgtacgggt 480
gctgacagaa aaattccttt ctctgtcata cccacggaca atggaacaaa aatctatggt 540
cttgagtgga atgatgactt tgttacagct tatattagtg gtcgttctta tcacttgaac 600

atcaatacta attggtttaa caatgtcaca cttttgtatt cacgctcaag cactgctacc 660
tgggaataca gtgctgcata tgcttaccaa ggtgtttcta acttcactta ttacaagtta 720
aataacacca atggtctaaa aacctatgaa ttatgtgaag attatgaaca ttgcactggc 780
tatgctacca atgtatttgc tccgacatca ggtggttaca tacctgatgg atttagtttt 840
aacaattggt tcttgcttac aaatagttcc acttttgtta gtggcaggtt tgtaacaaat 900
caaccattat tgattaattg cttgtggcca gtgcccagtt ttggtgtagc agcacaagaa 960
ttttgttttg aaggtgcaca gtttagccaa tgtaatggtg tgtctttaaa taacacagtg 1020
gatgttatta gattcaacct taatttcact gcagatgtac aatctggtat gggtgctaca 1080
gtattttcac tgaatacaac aggtggtgtc attcttgaaa tttcatgtta tagtgacaca 1140
gtgagtgagt ctagttctta cagttatggt gaaatcccgt tcggcataac tgacggacca 1200
cgatactgtt atgtacttta caatggcaca gctcttaaat atttaggaac attaccaccc 1260
agtgtaaagg aaattgctat tagtaagtgg ggccattttt atattaatgg ttacaatttc 1320
tttagcacat ttcctattgg ttgtatatct tttaatttaa ccactggtgt tagtggagct 1380
ttttggacaa ttgcttacac atcgtatact gaagcattag tacaagttga aaacacagct 1440
attaaaaatg tgacgtattg taacagtcac attaataaca ttaaatgttc tcaacttact 1500
gctaatttga ataatggatt ttatcctgtt gcttcaagtg aagtaggttt cgttaataag 1560
agtgttgtgt tattacctag ctttttcaca tacaccgctg tcaatataac cattgatctt 1620
ggtatgaagc ttagtggtta tggtcaaccc atagcctcga cactaagtaa catcacacta 1680
ccaatgcagg ataacaatac tgatgtgtac tgtattcgtt ctaaccaatt ctcagtttat 1740
gttcattcca cttgcaaaag ttctttatgg gacaatattt ttaatcaaga ctgcacggat 1800
gttttagagg ctacagctgt tataaaaact ggtacttgtc ctttctcatt tgataaattg 1860
aacaattact tgacttttaa caagttctgt ttgtcgttga gtcctgttgg tgctaattgc 1920
aagtttgatg ttgctgcacg tacaagaacc aatgagcagg ttgttagaag tctatatgta 1980
atatatgaag aaggagacaa catagtgggt gtaccgtctg ataatagcgg tctgcacgat 2040
ttgtctgtgc tacacctaga ctcctgtaca gattacaata tatatggtag aactggtgtt 2100
ggtattatta gacgaactaa cagtacgcta cttagtggct tatattacac atcactatca 2160
ggtgatttgt taggctttaa aaatgttagt gatggtgtca tttattctgt gacgccatgt 2220
gatgtaagcg cacaagcggc tgttattgat ggtgccatag ttggagctat gacttccatt 2280
aacagtgaac tgttaggtct aacacattgg acaacgacac ctaattttta ttactactct 2340
atatataatt acacaagtga gaggactcgt ggcactgcaa ttgacagtaa cgatgttgat 2400
tgtgaacctg tcataaccta ttctaatata ggtgtttgta aaaatggtgc tttggttttt 2460
attaacgtca cacattctga cggagacgtg caaccaatta gcactggtaa tgtcacgata 2520
cctacaaatt ttactatatc tgtgcaagtt gaatacatgc aggtttacac tacaccagta 2580
tcaatagatt gtgcaagata cgtttgtaat ggtaacccta gatgtaacaa attgttaaca 2640
caatatgtgt ctgcatgtca aactattgaa caagcacttg caatgggtgc cagacttgaa 2700
aacatggagg ttgattccat gttgtttgtc tcggaaaatg cccttaaatt ggcatctgtt 2760
gaggcgttca atagtacaga aaatttagat cctatttaca aagaatggcc tagcataggt 2820
ggttcttggc taggaggtct aaaagatata ctaccgtccc ataatagcaa acgtaagtat 2880
ggttctgcta tagaagattt gctttttgat aaagttgtaa catctggttt aggtacagtt 2940
gatgaagatt ataaacgttg tactggtggt tacgacatag cagacttggt gtgtgctcaa 3000
tattacaatg gcatcatggt tctaccaggt gtagctaatg ctgacaagat gactatgtac 3060
acagcatcac ttgcaggtgg tataacatta ggtgcacttg gtggtggcgc cgtggctata 3120
ccttttgcag tagcagtaca ggctagactt aattatgttg ctctacaaac tgatgtattg 3180
aataaaaacc aacagatcct ggctaatgct ttcaatcaag ctattggtaa cattacacag 3240
gcttttggta aggttaatga tgctatacat caaacatcac aaggtcttgc cactgttgct 3300
aaagcgttgg caaaagtgca agatgttgtc aacacacaag ggcaagcttt aagtcacctt 3360
acagtacaat tgcaaaataa ttttcaagcc attagtagtt ctattagtga tatttataac 3420
aggcttgacg aactgagtgc tgatgcacaa gttgataggc tgattacagg tagacttaca 3480
gcacttaatg catttgtgtc tcagactcta accagacaag cagaggttag ggctagtaga 3540
caacttgcca aagacaaggt taatgaatgt gttaggtctc agtctcagag attcggattc 3600
tgtggtaatg gtacacattt gttttcacta gcaaatgcag caccaaatgg catgattttc 3660
tttcatacag tactattacc aacagcttat gaaactgtaa cagcttggtc aggtatttgt 3720
gcttcagatg gcgatcgcac tttcggactt gtcgttaaag atgtgcagtt gacgttgttt 3780
cgtaatctag atgacaagtt ctatttgacc cccagaacta tgtatcagcc tagagttgca 3840
actagttctg attttgttca aattgaaggg tgtgatgtgt tgtttgtcaa cgcgactgta 3900
attgatttgc ctagtattat acctgactat attgacatta atcaaactgt tcaagacata 3960
ttagaaaatt acagaccaaa ctggactgta cctgaattta cacttgatat tttcaacgca 4020
acctatttaa atctgactgg tgaaattgat gacttagagt ttaggtcaga aaagctacat 4080

aacactacag tagaacttgc cattctcatt gataacatta ataatacatt agtcaatctt 4140
gaatggctca atagaattga aacttatgta aaatggcctt ggtatgtgtg gctactgata 4200
ggtttagtag tagtattttg cataccatta ctgctatttt gctgttttag cacaggttgt 4260
tgtggatgca taggttgttt aggaagttgt tgtcactcta tatgtagtag aagacaattt 4320
gaaaattatg aaccaattga aaaagtgcat gtccactaa 4359

<210> 3
<211> 4344
<212> DNA
<213> Transmissible gastroenteritis virus

<400> 3
atgaaaaaac tatttgtggt tttggtcgta atgccattga tttatggaga caattttcct 60
tgttctaaat tgactaatag aactataggc aaccagtgga atctcattga aaccttcctt 120
ctaaactata gtagtaggtt accacctaat tcagatgtgg tgttaggtga ttatttttct 180
actgtacaac cttggtttaa ttgcattcgc aatgatagta atgaccttta tgttacactg 240
gaaaatctta aagcattgta ttgggattat gctacagaaa atatcacttg gaatcacaga 300
caacggttaa acgtagtcgt taatggatac ccatactcca tcacagttac aacaacccgc 360
aattttaatt ctgctgaagg tgctattata tgcatttgta agggctcacc acctactacc 420
accacagaat ctagtttgac ttgcaattgg ggtagtgagt gcaggttaaa ccataagttc 480
cctatatgtc cttctaattc agaggcaaat tgtggtaata tgctgtatgg cctacaatgg 540
tttgcagatg aggttgttgc ttatttacat ggtgctagtt accgtattag ttttgaaaat 600
caatggtctg gcactgtcac atttggtgat atgcgtgcga caacattaga agtcgctggc 660
acgcttgtag acctttggtg gtttaatcct gtttatgatg tcagttatta tagggttaat 720
aataaaaatg gtactaccgt agtttccaat tgcactgatc aatgtgctag ttatgtggct 780
aatgttttta ctacacagcc aggaggtttt ataccatcag attttagttt taataattgg 840
ttccttctaa ctaatagctc cacgttggtt agtggtaaat tagttaccaa acagccgtta 900
ttagttaatt gcttatggcc agtccctagc tttgaagaag cagcttctac attttgtttt 960
gagggtgctg gctttgatca atgtaatggt gctgttttaa ataatactgt agacgtcatt 1020
aggttcaacc ttaattttac tacaaatgta caatcaggta agggtgccac agtgttttca 1080
ttgaacacaa cgggtggtgt cactcttgaa atttcatgtt atacagtgag tgactcgagc 1140
tttttcagtt acggtgaaat tccgttcggc gtaactgatg gaccacggta ctgttacgta 1200
cactataatg gcacagctct taagtattta ggaacattac cacctagtgt caaggagatt 1260
gctattagta agtggggcca tttttatatt aatggttaca atttctttag cacatttcct 1320
attgattgta tatcttttaa tttgaccact ggtgatagtg acgttttctg gacaatagct 1380
tacacatcgt acactgaagc attagtacaa gttgaaaaca cagctattac aaaggtgacg 1440
tattgtaata gtcacgttaa taacattaaa tgctctcaaa ttactgctaa tttgaataat 1500
ggattttatc ctgtttcttc aagtgaagtt ggtcttgtca ataagagtgt tgtgttacta 1560
cctagctttt acacacatac cattgttaac ataactattg gtcttggtat gaagcgtagt 1620
ggttatggtc aacccatagc ctcaacatta agtaacatca cactaccaat gcaggatcac 1680
aacaccgatg tgtactgtat tcgttctgac caattttcag tttatgttca ttctacttgc 1740
aaaagtgctt tatgggacaa tatttttaag cgaaactgca cggacgtttt agatgccaca 1800
gctgttataa aaactggtac ttgtcctttc tcatttgata aattgaacaa ttacttaact 1860
tttaacaagt tctgtttgtc gttgagtcct gttggtgcta attgtaagtt tgatgtagct 1920
gcccgtacaa gaaccaatga gcaggttgtt agaagtttgt atgtaatata tgaagaagga 1980
gacaacatag tgggtgtacc gtctgataat agtggtgtgc acgatttgtc agtgctacac 2040
ctagattcct gcacagatta caatatatat ggtagaactg gtgttggtat tattagacaa 2100
actaacagga cgctacttag tggcttatat tacacatcac tatcaggtga tttgttaggt 2160
tttaaaaatg ttagtgatgg tgtcatctac tctgtaacgc catgtgatgt aagcgcacaa 2220
gcagctgtta ttgatggtac catagttggg gctatcactt ccattaacag tgaactgtta 2280
ggtctaacac attggacaac aacacctaat ttttattact actctatata taattacaca 2340
aatgatagga ctcgtggcac tgcaattgac agtaatgatg ttgattgtga acctgtcata 2400
acctattcta acataggtgt ttgtaaaaat ggtgcttttg tttttattaa cgtcacacat 2460
tctgatggag acgtgcaacc aattagcact ggtaatgtca cgatacctac aaactttacc 2520
atatccgtgc aagtcgaata tattcaggtt tacactacac cagtgtcaat agactgttca 2580
agatatgttt gtaatggtaa ccctaggtgt aacaaattgt taacacaata cgtttctgca 2640
tgtcaaacta ttgagcaagc acttgcaatg ggtgccagac ttgaaaacat ggaggttgat 2700

tccatgttgt ttgtttctga aaatgccctt aaattggcat ctgttgaagc attcaatagt 2760
tcagaaactt tagaccctat ttacaaagaa tggcctaata taggtggttc ttggctagaa 2820
ggtctaaaat acatacttcc gtcccataat agcaaacgta agtatcgttc agctatagag 2880
gacttgcttt ttgataaggt tgtaacatct ggtttaggta cagttgatga agattataaa 2940
cgttgtacag gtggttatga catagctgac ttagtatgtg ctcaatacta taatggcatc 3000
atggtgctac ctggtgtggc taatgctgac aaaatgacta tgtacacagc atcccttgca 3060
ggtggtataa cattaggtgc acttggtgga ggcgccgtgg ctataccttt tgcagtagca 3120
gttcaggcta gacttaatta tgttgctcta caaactgatg tattgaacaa aaaccagcag 3180
attctggcta gtgctttcaa tcaagctatt ggtaacatta cacagtcatt tggtaaggtt 3240
aatgatgcta tacatcaaac atcacgaggt cttgctactg ttgctaaagc attggcaaaa 3300
gtgcaagatg ttgtcaacat acaagggcaa gctttaagcc acctaacagt acaattgcaa 3360
aataatttcc aagccattag tagttctatt agtgacattt ataataggct tgacgaattg 3420
agtgctgatg cacaagttga caggctgatc acaggaagac ttacagcact taatgcattt 3480
gtgtctcaga ctctaaccag acaagcggag gttagggcta gtagacaact tgccaaagac 3540
aaggttaatg aatgcgttag gtctcagtct cagagattcg gattctgtgg taatggtaca 3600
catttgtttt cactcgcaaa tgcagcacca aatggcatga ttttctttca cacagtgcta 3660
ttaccaacgg cttatgaaac tgtgactgct tggccaggta tttgtgcttc agatggtgat 3720
cgcacttttg gacttgtcgt taaagatgtc cagttgactt tgtttcgtaa tctagatgac 3780
aagttctatt tgacccccag aactatgtat cagcctagag ttgcaactag ttctgacttt 3840
gttcaaattg aagggtgcga tgtgctgttt gttaatgcaa ctgtaagtga tttgcctagt 3900
attatacctg attatattga tattaatcag actgttcaag acatattaga aaatttaga 3960
ccaaattgga ctgtacctga gttgacattt gacattttta acgcaaccta tttaaacctg 4020
actggtgaaa ttgatgactt agaatttagg tcagaaaagc tacataacac cactgtagaa 4080
cttgccattc tcattgacaa cattaacaat acattagtca atcttgaatg gctcaataga 4140
attgaaacct atgtaaaatg gccttggtat gtgtggctac taataggctt agtagtaata 4200
ttttgcatac cattactgct attttgctgt tgtagtacag gttgctgtgg atgcataggt 4260
tgtttaggaa gttgttgtca ctctatatgt agtagaagac aatttgaaaa ttacgaacca 4320
attgaaaaag tgcacgtcca ttaa 4344

<210> 4
<211> 3678
<212> DNA
<213> Porcine respiratory coronavirus

<400> 4
atgaaaaaat tatttgtggt cttggttgta atgccattga tttatggaga caagtttcct 60
acttccgtag tttccaattg cactgatcaa tgtgctagtt atgtggctaa tgtttttact 120
atactaccag gaggctttat accatcagat tttagtttta ataattggtt cctcctaact 180
aatagctcca cgttggttaa tggtaaatta gttaccaaac agcctctatt agttaattgc 240
ttatggccag tccctagctt tgaagaagta gcttctacat tttgttttga aggtgctgac 300
tttgatcaat gtaatggtgc tgttttaaat aacactgtag acgtcattag gtttaacctt 360
aattttacta caaatgtaca atcaggtaag ggtgctacag tgttttcatt gaacacaacg 420
ggtggtgtca ctcttgaaat ctcatgttat aatgatacag tgagtgattc gagcttttcc 480
agttacggtg aaattccgtt cggcgtaact aatggaccac ggtactgtta cgtactctat 540
aatggcacag ctcttaagta tctaggaaca ttaccaccta gtgtcaagga gattgctatt 600
agtaagtggg gccatttta tattaatggt tacaatttct ttagcacatt tcctattgat 660
tgtatatctt ttaatttgac tactggtgat agtgacgtct tctggacaat agcttacaca 720
tcgtacactg aagcattagt acaagttgaa aacacagcta ttacaaatgt gacgtattgt 780
aatagttatg ttaataacat taaatgctct caacttactg ctaatttgaa taatggattt 840
tatcctgttt cttcaagtga agttggttct gtcaataaga gtgttgtgtt actacctagc 900
tttctgacac ataccattgt taacataact attggtcttg gtatgaagcg tagtggttat 960
ggtcaaccca tagcctcaac gctaagtaac attacactac caatgcagga taacaacaac 1020
gatgtgtact gtgttcgttc tgaccaattt tcagtttatg ttcattctac ttgcaaaagt 1080
gttttatggg acaatgtttt taagcgaaac tgcacggacg ttttagatgc cacagctgtt 1140
ataaaaactg gtacttgtcc tttctcattt gataaattga acaattactt aacttttaac 1200
aagttctgtt tgtcgttgag tcccgttggt gctaattgta agtttgatgt agctgcccgt 1260
acaagaacca atgatcaggt tgttagaagt ttgtatgtaa tatatgaaga aggagacagc 1320

atagttggtg taccgtctga caatagtggt ttgcacgatt tgtcagtgct acacctagat 1380
tcgtgcacag attacaatat atatggtaga actggtgttg gtattattag acaaactaac 1440
aggacgatac ttagtggctt atattacaca tcactatctg gtgatttgtt aggttttaca 1500
aatgttagtg atggtgttat ctactctgta acgccatgtg atgttagcgc acaagcagct 1560
attattgatg gtaccatagt tggggctatc acttccatta acagtgaatt gttaggtcta 1620
acacattgga caacaacacc taatttttat tactactcta tatataatta cacaaatgat 1680
aagactcgtg gcactccaat tggcagtaat gacgttgatt gtgaacctgt cataacctat 1740
tctaacatag gtgtttgtaa aaatggtgct ttggttttta ttaacgtcac acattctgat 1800
ggagacgtgc aaccaattag cactggtaac gtcacgatac ctactaactt tactatatcc 1860
gtgcaagtcg aatacattca ggtttacact acaccagtgt caatagactg ttcaagatat 1920
gtttgtaatg gcaaccctag gtgtaacaaa ctgttaacac aatacgtttc tgcatgtcaa 1980
actattgagc aagcacttgc aatgggtgcc agacttgaaa acatggaagt tgattccatg 2040
ttatttgttt ctgaaaatgc ccttaaattg gcttctgtcg aagcattcaa tagttcagaa 2100
actttagatc ctatttacaa agaatggcct aatataggtg gcttttggct agaaggtcta 2160
aaatacatac ttccgtccga taatagcaaa cgtaagtatc gttcagctat agaggacttg 2220
ctttttcta aggttgtaac atctggttta ggtacagttg atgaagatta caaacgttgt 2280
acaggtggtt atgacatagc tgacttagta tgtgctcaat actataatgg cattatggtg 2340
ctacctggtg tggctaatgc tgacaaaatg actatgtaca cagcatccct cgcaggtggt 2400
ataacattag gtgcacttgg tggaggcgcc gtggctatac cttttgcagt agcagttcag 2460
gctagactta attatgttgc tctacaaact gatgtattga acaaaaacca gcagatcctg 2520
gctagtgctt ttaatcaagc tattggtaac attacacagt catttggtaa ggttaatgat 2580
gctatacatc aaacatcacg aggtcttaca actgttgcta aagcattggc aaaagtgcaa 2640
gatgttgtca acacacaagg tcaagcttta agacacctaa cagtacaatt gcaaaataat 2700
ttccaagcca ttagtagttc tattagtgac atttataata ggcttgatga attgagtgct 2760
gatgcacaag tcgacaggct gatcacagga agacttacag cacttaatgc atttgtgtct 2820
cagactctaa ccagacaagc cgaggttagg gctagtagac aacttgctaa agacaaggtt 2880
aatgaatgcg ttaggtctca gtctcagaga ttcggcttct gtggtaatgg tacacatttg 2940
ttttcactcg caaatgcagc accaaatggc atgatcttct ttcacacagt gctattacca 3000
acggcgtatg aaactgtgac tgcttggtca ggtatttgtg ctttagatgt tgatcgcact 3060
tttggacttg tcgttaaaga tgtccagttg actttatttc gtaatctaga tgacaagttc 3120
tatttgacac ccagaactat gtatcagcct agagtggcaa ctagttctga ttttgttcaa 3180
attgaagggt gcgatgtgct gtttgttaat acaactgtaa gtgatttgcc tagtattata 3240
cctgattata ttgatattaa tcagactgtt caagacatat tagaaaattt tagaccaaat 3300
tggactgtac ctgagctgac attggacgtt tttaacgcaa cctatttaaa cctgactggt 3360
gaaattgatg acttagagtt taggtcagaa aagctacata acactactgt agaacttgcc 3420
attctcattg acaacattaa caatacagta gtcaatcttg aatggcttaa tagaattgaa 3480
acttatgtaa aatggccttg gtatgtgtgg ctactaatag gcttagtagt aatatttttgc 3540
ataccattac tgctattttg ctgttgtagt acaggttgct gtggatgcat aggttgttta 3600
ggaagttgtt gtcactctat attcagtaga agacaatttg aaaattatga acctattgaa 3660
aaagtgcacg tccattaa 3678

<210> 5
<211> 3522
<212> DNA
<213> Human coronavirus OC43

<400> 5
atgtttgttt tgcttgttgc atatgccttg ttgcatattg ctggttgtca aactacaaat 60
gggctgaaca ctagttactc tgtttgcaac ggctgtgttg gttattcaga aaatgtattt 120
gctgttgaga gtggtggtta tataccctcc gactttgcat tcaataattg gttccttcta 180
actaatacct catctgttgt agatggtgtt gtgaggagtt ttcagccttt gttgcttaat 240
tgcttatggt ctgtttctgg cttgcggttt actactggtt ttgtctattt taatggtact 300
gggagaggtg attgtaaagg tttttcctca gatgttttgt ctgatgtcat acgttacaac 360
ctcaattttg aagaaaacct tagacgtgga accattttgt ttaaaacatc ttatggtgtt 420
gttgtgtttt attgtaccaa caacacttta gtttcaggtg atgctcacat accatttggt 480
acagttttgg gcaattttta ttgctttgta aatactacta ttggcaatga aactacgtct 540
gcttttgtgg gtgcactacc taagacagtt cgtgagtttg ttatttcacg cacaggacat 600

ttttatatta atggctatcg ctatttcact ttaggtaatg tagaagccgt taatttcaat 660
gtcactactg cagaaaccac tgatttttgt actgttgcgt tagcttctta tgctgacgtt 720
ttggttaatg tgtcacaaac ctctattgct aatataattt attgcaactc tgttattaac 780
agactgagat gtgaccagtt gtcctttgat gtaccagatg gtttttattc tacaagccct 840
attcaatccg ttgagctacc tgtgtctatt gtgtcgctac ctgtttatca taaacatacg 900
tttattgtgt tgtacgttga cttcaaacct cagagtggcg gtggcaagtg ctttaactgt 960
tatcctgctg gtgttaatat tacactggcc aattttaatg aaactaaagg gcctttgtgt 1020
gttgacacat cacacttcac taccaaatac gttgctgttt atgccaatgt tggtaggtgg 1080
agtgctagta ttaacacggg aaattgccct ttttcttttg gcaaagttaa taactttgtt 1140
aaatttggca gtgtatgttt ttcgctaaag gatatacccg gtggttgcgc aatgcctata 1200
gtggctaatt gggcttatag taagtactat actataggct cattgtatgt ttcttggagt 1260
gatggtgatg gaattactgg cgtcccacaa cctgttgagg gtgttagttc ctttatgaat 1320
gttacattgg acaaatgtac taaatataat atttatgatg tatctggtgt gggtgttatt 1380
cgcgttagca atgacacctt tcttaatgga attacgtaca catcaacttc aggtaacctt 1440
ctgggtttta aagatgttac taagggcacc atctactcta tcactccttg taacccacca 1500
gatcagcttg ttgtttatca gcaagctgtt gttggtgcta tgttgtctga aaattttact 1560
agttacggct tttctaatgt tgtagaactg ccgaaatttt tctatgcgtc caatggcact 1620
tataattgca cagacgctgt tttaacttat tctagttttg gcgtttgtgc agatggttct 1680
ataattgctg ttcaaccacg taatgtttca tatgatagtg tttcagctat cgtcacagct 1740
aatttgtcta taccttccaa ttggaccact tcggtccagg ttgagtattt acaaattaca 1800
agtacaccta tcgtagttga ttgctccact tatgtttgca atggtaatgt gcgctgtgtt 1860
gaattgctta agcagtatac ttctgcttgt aaaactattg aagacgcctt aagaaatagc 1920
gccaggctgg agtctgcaga tgttagtgag atgctcactt ttgacaagaa agcgtttaca 1980
cttgctaatg ttagtagttt tggtgactac aaccttagca gcgtcatacc tagcttgccc 2040
acaagtggta gtagagtggc tggtcgcagt gccatagaag acatactttt tagcaaactt 2100
gttacttctg gacttggcac tgtggacgca gactacaaaa agtgcactaa gggtctttcc 2160
attgctgact tggcttgtgc tcaatattat aatggcatta tggttttgcc tggcgtcgct 2220
gatgctgaac gaatggccat gtatacaggt tctttaattg gtggaattgc tttaggaggt 2280
ctaacatcag ccgtttcaat accattttca ttagcaattc aggcacgttt aaattatgtt 2340
gcattgcaga ctgatgtttt acaagaaaat cagaaaattc ttgctgcatc ttttaacaaa 2400
gcaatgacca acatagtaga tgcctttact ggtgttaatg atgctattac acaaacttca 2460
caagccctac aaacagttgc tactgcactt aacaagatcc aggatgttgt taatcaacaa 2520
ggcaactcat tgaaccattt aacttctcag ttgaggcaga attttcaagc tatctctagc 2580
tctattcagg ctatctatga cagacttgac actattcagg ctgatcaaca agtagatagg 2640
ctgattactg gtagattggc tgctttgaat gtattcgttt ctcatacatt gactaagtac 2700
actgaagttc gtgcttccag acagcttgca caacaaaaag tgaatgagtg tgtcaaatcc 2760
cagtctaagc gttatggctt ctgtggaaat ggcactcaca ttttctcaat tgttaatgct 2820
gctcctgagg ggcttgtttt tctccacact gtcttgttgc cgacacaata taaggatgtt 2880
gaagcgtggt ctgggttgtg cgttgatggt acaaacggtt atgtgttgcg acaacctaat 2940
cttgctcttt acaaagaagg caattattat agaatcacat ctcgcataat gtttgaacca 3000
cgtattccta ccatggcaga ttttgttcaa attgaaaatt gcaatgtcac atttgttaac 3060
atttctcgct ctgagttgca aaccattgtg ccagagtata ttgatgttaa taagacgctg 3120
caagaattaa gttacaaatt gccaaattac actgttccag acctagttgt cgaacagtac 3180
aaccagacta ttttgaattt gaccagtgaa attagcaccc ttgaaaataa atctgcggag 3240
cttaattaca ctgttcaaaa attgcaaact ctgattgaca acataaatag cacattagtc 3300
gacttaaagt ggctcaaccg ggttgagact tacatcaagt ggccgtggtg ggtgtggttg 3360
tgcatttcag tcgtgctcat ctttgtggtg agtatgttgc tattatgttg ttgttctact 3420
ggttgctgtg gcttctttag ttgttttgca tcttctatta gaggttgttg tgaatcaact 3480
aaacttcctt attacgacgt tgaaaagatc cacatacagt aa 3522

<210> 6
<211> 4152
<212> DNA
<213> Porcine epidemic diarrhea virus

<400> 6
atgaggtctt taatttactt ctggttgctc ttaccagtac ttccaacact cagcctacca 60

caagatgtca ctaggtgcca gtctactact aactttaggc ggttcttttc aaaatttaat 120
gttcaggcac ctgccgtcgt cgttttgggt ggttacctac ctagtatgaa ctcttctagc 180
tggtactgtg gcacaggcat tgaaactgct agtggcgttc atggtatttt tctcagctac 240
atcgattctg gtcagggctt tgagattggc atttcgcaag agccgtttga tcctagtggt 300
taccagcttt atttacataa ggccactaat ggtaacacta atgctattgc acgactgcgc 360
atttgccagt ttcccgataa taaaacattg ggccctactg ttaatgatgt tacaacaggt 420
cgtaactgcc tattcaacaa agccattcca gcttatatgc gtgatggaaa agatattgtt 480
gtcggcataa catgggataa tgatcgtgtc actgtttttg ctgacaagat ctatcatttt 540
tatcttaaaa atgattggtc ccgcgttgcg acaagatgtt acaatcgcag aagttgtgct 600
atgcaatatg tttatacacc tacctactac atgcttaatg ttactagtgc aggtgaggat 660
ggcatttatt atgaaccctg tacagctaat tgcactggtt acgctgccaa tgtatttgcc 720
actgattcca atggccatat accagaaggt tttagtttta ataattggtt tcttttatcc 780
aatgactcca ctttgttgca tggtaaagtg gtttccaacc aacccttgtt ggtcaattgt 840
cttttggcca ttcctaagat ttatggacta ggccaatttt tctcattcaa tcacacgatg 900
gatggcgttt gtaatggagc tgctgtggat cgtgccccag aggctctgag gtttaatatt 960
aatgacacct ccgtcattct tgctgaaggc tcaattgtac ttcatactgc tttaggaaca 1020
aatctttctt ttgtttgcag taattcctca gatcctcatt tagccatctt tgccatacct 1080
ctgggtgcta ctgaagtacc ctactattgc tttcttaaag tggatactta caactccact 1140
gtttataaat tcttggctgt tttacctcct actgtcaggg aaattgtcat caccaagtat 1200
ggtgatgttt atgtcaatgg gtttggctat ttgcatctcg gtttgttgga tgctgtcaca 1260
attaatttca ctggtcatgg cactgacgat gacgtttcag gtttctggac catagcatcg 1320
actaattttg ttgatgcact catcgaggtt caaggaactt ccattcagcg tattctttat 1380
tgtgatgatc ctgttagcca actcaagtgt tctcaggttg cttttgacct tgacgatggt 1440
ttttacccca tctcttctag aaaccttctg agtcacgaac agccaatttc ttttgttact 1500
ttgccatcat ttaatgatca ttcttttgtt aatattactg tctctgcggc ttttggtggt 1560
cttagtagtg ccaatctcgt tgcatctgac actactatca atgggtttag ttctttctgt 1620
gttgacacta gacaatttac cattacactg ttttataatg ttacaaacag ttatggttat 1680
gtgtctaaat cacaggatag taattgtcct ttcaccttgc aatctgttaa tgattacctg 1740
tcttttagca aattttgtgt ttcaaccagc cttttggctg gtgcttgtac catagatctt 1800
tttggttacc ctgcgttcgg tagtggtgtt aagttgacgt ccctttattt tcaattcaca 1860
aaaggtgagt tgattactgg cacgcctaaa ccacttgaag gtatcacaga cgtttctttt 1920
atgactctgg atgtgtgtac caagtatact atctatggct ttaaaggtga gggtattatt 1980
acccttacaa attctagcat tttggcaggt gtttattata catctgattc tggacagttg 2040
ttagccttta agaatgtcac tagtggtgct gtttattctg tcacgccatg ttctttttca 2100
gagcaggctg catatgttaa tgatgatata gtgggtgtta tttctagttt gtctaactcc 2160
acttttaaca atactaggga gttgcctggt ttcttctacc attctaatga cggctccaat 2220
tgtacagagc ctgtgttggt gtatagtaac ataggtgttt gtaaatctgg cagtattggc 2280
tatgttccat ctcagtatgg ccaagtcaag attgcaccca cggttactgg gaatattagt 2340
attcccacca actttagtat gagtattaga acagaatatt tacagcttta caacacgcct 2400
gttagtgttg attgtgctac atatgtttgt aatggtaact ctcgttgtaa acaattactc 2460
acccagtaca ctgcagcatg taagaccata gagtcagcat tacaactcag cgctaggctt 2520
gagtctgttg aagttaactc tatgcttacc atttctgaag aggctttaca gttagctacc 2580
atcagttcgt ttaatggtga tggatataac tttactaatg tgctgggtgc ttccgtgtac 2640
gatcctgcaa gtggcagggt ggtacaaaaa aggtctgtta ttgaagactt gctttttaat 2700
aaagtggtta ctaatggcct tggtactgtt gatgaagact ataagcgctg ttctaatggt 2760
cgctctgtgg ctgatctagt ctgtgcgcag tattactctg gtgtcatggt actacctggc 2820
gttgttgacg ctgagaagct tcacatgtac agtgcgtctc tcataggtgg tatggcgcta 2880
ggaggtataa ctgctgcagc ggcattgcct tttagctatg ctgttcaagc gagactcaat 2940
tatcttgctt tacagacgga tgttctacag cggaaccagc aattgcttgc tgagtctttt 3000
aactctgcta ttggtaatat aacttcagcc tttgagagtg ttaaagaggc tattagtcaa 3060
acttccaagg gtttgaacac tgtggctcat gcgcttacta aggttcaaga ggttgttaat 3120
tcgcagggtt cagctttgaa ccaacttacc gtacagctgc aacacaactt ccaagccatt 3180
tctagttcta ttgatgacat ttattcccga ctggacattc tttcagccga tgttcaggtt 3240
gatcgtctca tcaccggcag attatcagca cttaatgctt ttgttgccca aaccctcact 3300
aagtatactg aggttcaggc tagcaggaag ctagcacagc aaaaggttaa tgagtgcgtc 3360
aaatcgcaat ctcagcgtta cggtttttgt ggtggtgatg gcgagcacat tttctctctg 3420
gtacaggccg cacctcaggg cctgctgttc ttacatacag tacttgtacc gggtgatttt 3480
gtaaatgttc ttgccatcgc tggcttatgc gttaatggtg aaattgcctt gactctacgt 3540

gagcctggct tagtcttgtt tacgcatgaa cttcaaactt atactgcgac ggaatatttt 3600
gtttcatcgc gacgtatgtt tgaacctaga aaacctaccg ttagtgattt tgttcaaatt 3660
gagagttgtg tggtcaccta tgtcaatctg actagcgacc agctaccaga tgtaatccca 3720
gattacatcg atgttaacaa aacacttgat gagattttag cttctctgcc caatagaact 3780
ggtccaagtc ttcccctaga tgtttttaat gccacttatc ttaatcttac tggtgaaatt 3840
gcagatctag agcagcgttc agagtctctc cgtaatacta cagaagagct ccgaagtctc 3900
attaacaaca tcaacaacac acttgttgac cttgagtggc tcaaccgagt tgagacatac 3960
atcaagtggc cgtggtgggt ttggttgatc attgttattg ttctcatctt tgttgtgtca 4020
ttactagtgt tctgctgcat ttccacgggt tgttgtggat gctgcggttg ctgcggtgct 4080
tgtttttcag gttgttgtag gggtcctaga cttcaacctt acgaagcttt tgaaaaggtc 4140
cacgtgcagt ga 4152

<210> 7
<211> 3768
<212> DNA
<213> SARS coronavirus urbani

<400> 7
atgtttattt tcttattatt tcttactctc actagtggta gtgaccttga ccggtgcacc 60
acttttgatg atgttcaagc tcctaattac actcaacata cttcatctat gaggggggtt 120
tactatcctg atgaaatttt tagatcagac actctttatt taactcagga tttatttctt 180
ccattttatt ctaatgttac agggtttcat actattaatc atacgtttgg caaccctgtc 240
ataccttta aggatggtat ttatttgct gccacagaga aatcaaatgt tgtccgtggt 300
tgggtttttg gttctaccat gaacaacaag tcacagtcgg tgattattat taacaattct 360
actaatgttg ttatacgagc atgtaacttt gaattgtgtg acaacccttt ctttgctgtt 420
tctaaaccca tgggtacaca gacacatact atgatattcg ataatgcatt taattgcact 480
ttcgagtaca tatctgatgc cttttcgctt gatgtttcag aaaagtcagg taattttaaa 540
cacttacgag agtttgtgtt taaaaataaa gatgggtttc tctatgttta taagggctat 600
caacctatag atgtagttcg tgatctacct tctggtttta acactttgaa acctatttt 660
aagttgcctc ttggtattaa cattacaaat tttagagcca ttcttacagc cttttcacct 720
gctcaagaca tttggggcac gtcagctgca gcctattttg ttggctattt aaagccaact 780
acatttatgc tcaagtatga tgaaaatggt acaatcacag atgctgttga ttgttctcaa 840
aatccacttg ctgaactcaa atgctctgtt aagagctttg agattgacaa aggaatttac 900
cagacctcta atttcagggt tgttccctca ggagatgttg tgagattccc taatattaca 960
aacttgtgtc cttttggaga ggtttttaat gctactaaat tcccttctgt ctatgcatgg 1020
gagagaaaaa aaatttctaa ttgtgttgct gattactctg tgctctacaa ctcaacattt 1080
ttttcaacct ttaagtgcta tggcgtttct gccactaagt tgaatgatct ttgcttctcc 1140
aatgtctatg cagattcttt tgtagtcaag ggagatgatg taagacaaat agcgccagga 1200
caaactggtg ttattgctga ttataattat aaattgccag atgatttcat gggttgtgtc 1260
cttgcttgga atactaggaa cattgatgct acttcaactg gtaattataa ttataaatat 1320
aggtatctta gacatggcaa gcttaggccc tttgagagag acatatctaa tgtgcctttc 1380
tcccctgatg gcaaaccttg caccccacct gctcttaatt gttattggcc attaaatgat 1440
tatggttttt acaccactac tggcattggc taccaacctt acagagttgt agtactttct 1500
tttgaacttt taaatgcacc ggccacggtt tgtggaccaa aattatccac tgaccttatt 1560
aagaaccagt gtgtcaattt taatttttaat ggactcactg gtactggtgt gttaactcct 1620
tcttcaaaga gatttcaacc atttcaacaa tttggccgtg atgtttctga tttcactgat 1680
tccgttcgag atcctaaaac atctgaaata ttagacattt caccttgctc ttttgggggt 1740
gtaagtgtaa ttacacctgg aacaaatgct tcatctgaag ttgctgttct atatcaagat 1800
gttaactgca ctgatgtttc tacagcaatt catgcagatc aactcacacc agcttggcgc 1860
atatattcta ctggaaacaa tgtattccag actcaagcag gctgtcttat aggagctgag 1920
catgtcgaca cttcttatga gtgcgacatt cctattggag ctggcatttg tgctagttac 1980
catacagttt ctttattacg tagtactagc caaaaatcta ttgtggctta tactatgtct 2040
ttaggtgctg atagttcaat tgcttactct aataacacca ttgctatacc tactaacttt 2100
tcaattagca ttactacaga agtaatgcct gtttctatgg ctaaaacctc cgtagattgt 2160
aatatgtaca tctgcggaga ttctactgaa tgtgctaatt tgcttctcca atatggtagc 2220
ttttgcacac aactaaatcg tgcactctca ggtattgctg ctgaacagga tcgcaacaca 2280
cgtgaagtgt tcgctcaagt caaacaaatg tacaaaaccc caactttgaa atattttggt 2340

ggttttaatt tttcacaaat attacctgac cctctaaagc caactaagag gtcttttatt 2400
gaggacttgc tctttaataa ggtgacactc gctgatgctg gcttcatgaa gcaatatggc 2460
gaatgcctag gtgatattaa tgctagagat ctcatttgtg cgcagaagtt caatggactt 2520
acagtgttgc cacctctgct cactgatgat atgattgctg cctacactgc tgctctagtt 2580
agtggtactg ccactgctgg atggacattt ggtgctggcg ctgctcttca aataccttt 2640
gctatgcaaa tggcatatag gttcaatggc attggagtta cccaaaatgt tctctatgag 2700
aaccaaaaac aaatcgccaa ccaatttaac aaggcgatta gtcaaattca agaatcactt 2760
acaacaacat caactgcatt gggcaagctg caagacgttg ttaaccagaa tgctcaagca 2820
ttaaacacac ttgttaaaca acttagctct aattttggtg caatttcaag tgtgctaaat 2880
gatatccttt cgcgacttga taaagtcgag gcggaggtac aaattgacag gttaattaca 2940
ggcagacttc aaagccttca aacctatgta acacaacaac taatcagggc tgctgaaatc 3000
agggcttctg ctaatcttgc tgctactaaa atgtctgagt gtgttcttgg acaatcaaaa 3060
agagttgact tttgtggaaa gggctaccac cttatgtcct tcccacaagc agccccgcat 3120
ggtgttgtct tcctacatgt cacgtatgtg ccatcccagg agaggaactt caccacagcg 3180
ccagcaattt gtcatgaagg caaagcatac ttccctcgtg aaggtgtttt tgtgtttaat 3240
ggcacttctt ggtttattac acagaggaac ttcttttctc cacaaataat tactacagac 3300
aatacatttg tctcaggaaa ttgtgatgtc gttattggca tcattaacaa cacagtttat 3360
gatcctctgc aacctgagct cgactcattc aaagaagagc tggacaagta cttcaaaaat 3420
catacatcac cagatgttga tcttggcgac atttcaggca ttaacgcttc tgtcgtcaac 3480
attcaaaaag aaattgaccg cctcaatgag gtcgctaaaa atttaaatga atcactcatt 3540
gaccttcaag aattgggaaa atatgagcaa tatattaaat ggccttggta tgtttggctc 3600
ggcttcattg ctggactaat tgccatcgtc atggttacaa tcttgctttg ttgcatgact 3660
agttgttgca gttgcctcaa gggtgcatgc tcttgtggtt cttgctgcaa gtttgatgag 3720
gatgactctg agccagttct caagggtgtc aaattacatt acacataa 3768

<210> 8
<211> 3768
<212> DNA
<213> SARS coronavirus Tor2

<400> 8
atgtttattt tcttattatt tcttactctc actagtggta gtgaccttga ccggtgcacc 60
acttttgatg atgttcaagc tcctaattac actcaacata cttcatctat gaggggggtt 120
tactatcctg atgaaatttt tagatcagac actctttatt taactcagga tttatttctt 180
ccattttatt ctaatgttac agggtttcat actattaatc atacgtttgg caaccctgtc 240
atacctttta aggatggtat ttattttgct gccacagaga aatcaaatgt tgtccgtggt 300
tgggtttttg gttctaccat gaacaacaag tcacagtcgg tgattattat taacaattct 360
actaatgttg ttatacgagc atgtaacttt gaattgtgtg acaacccttt ctttgctgtt 420
tctaaaccca tgggtacaca gacacatact atgatattcg ataatgcatt taattgcact 480
ttcgagtaca tatctgatgc cttttcgctt gatgtttcag aaaagtcagg taattttaaa 540
cacttacgag agtttgtgtt taaaaataaa gatgggtttc tctatgttta taagggctat 600
caacctatag atgtagttcg tgatctacct tctggtttta acactttgaa acctatttt 660
aagttgcctc ttggtattaa cattacaaat tttagagcca ttcttacagc cttttcacct 720
gctcaagaca tttggggcac gtcagctgca gcctattttg ttggctattt aaagccaact 780
acatttatgc tcaagtatga tgaaaatggt acaatcacag atgctgttga ttgttctcaa 840
aatccacttg ctgaactcaa atgctctgtt aagagctttg agattgacaa aggaatttac 900
cagacctcta atttcagggt tgttccctca ggagatgttg tgagattccc taatattaca 960
aacttgtgtc cttttggaga ggtttttaat gctactaaat tcccttctgt ctatgcatgg 1020
gagagaaaaa aaatttctaa ttgtgttgct gattactctg tgctctacaa ctcaacattt 1080
ttttcaacct ttaagtgcta tggcgtttct gccactaagt tgaatgatct ttgcttctcc 1140
aatgtctatg cagattcttt tgtagtcaag ggagatgatg taagacaaat agcgccagga 1200
caaactggtg ttattgctga ttataattat aaattgccag atgatttcat gggttgtgtc 1260
cttgcttgga atactaggaa cattgatgct acttcaactg gtaattataa ttataaatat 1320
aggtatctta gacatggcaa gcttaggccc tttgagagag acatatctaa tgtgcctttc 1380
tcccctgatg gcaaaccttg caccccacct gctcttaatt gttattggcc attaaatgat 1440
tatggttttt acaccactac tggcattggc taccaacctt acagagttgt agtactttct 1500
tttgaacttt taaatgcacc ggccacggtt tgtggaccaa aattatccac tgaccttatt 1560

aagaaccagt gtgtcaattt taatttaat ggactcactg gtactggtgt gttaactcct 1620
tcttcaaaga gatttcaacc atttcaacaa tttggccgtg atgtttctga tttcactgat 1680
tccgttcgag atcctaaaac atctgaaata ttagacattt caccttgcgc ttttgggggt 1740
gtaagtgtaa ttacacctgg aacaaatgct tcatctgaag ttgctgttct atatcaagat 1800
gttaactgca ctgatgtttc tacagcaatt catgcagatc aactcacacc agcttggcgc 1860
atatattcta ctggaaacaa tgtattccag actcaagcag gctgtcttat aggagctgag 1920
catgtcgaca cttcttatga gtgcgacatt cctattggag ctggcatttg tgctagttac 1980
catacagttt ctttattacg tagtactagc caaaaatcta ttgtggctta tactatgtct 2040
ttaggtgctg atagttcaat tgcttactct aataacacca ttgctatacc tactaacttt 2100
tcaattagca ttactacaga agtaatgcct gtttctatgg ctaaaacctc cgtagattgt 2160
aatatgtaca tctgcggaga ttctactgaa tgtgctaatt tgcttctcca atatggtagc 2220
ttttgcacac aactaaatcg tgcactctca ggtattgctg ctgaacagga tcgcaacaca 2280
cgtgaagtgt tcgctcaagt caaacaaatg tacaaaaccc caactttgaa atattttggt 2340
ggttttaatt tttcacaaat attacctgac cctctaaagc caactaagag gtcttttatt 2400
gaggacttgc tctttaataa ggtgacactc gctgatgctg gcttcatgaa gcaatatggc 2460
gaatgcctag gtgatattaa tgctagagat ctcatttgtg cgcagaagtt caatggactt 2520
acagtgttgc cacctctgct cactgatgat atgattgctg cctacactgc tgctctagtt 2580
agtggtactg ccactgctgg atggacattt ggtgctggcg ctgctcttca aatacctttt 2640
gctatgcaaa tggcatatag gttcaatggc attggagtta cccaaaatgt tctctatgag 2700
aaccaaaaac aaatcgccaa ccaatttaac aaggcgatta gtcaaattca agaatcactt 2760
acaacaacat caactgcatt gggcaagctg caagacgttg ttaaccagaa tgctcaagca 2820
ttaaacacac ttgttaaaca acttagctct aattttggtg caatttcaag tgtgctaaat 2880
gatatccttt cgcgacttga taaagtcgag gcggaggtac aaattgacag gttaattaca 2940
ggcagacttc aaagccttca aacctatgta acacaacaac taatcagggc tgctgaaatc 3000
agggcttctg ctaatcttgc tgctactaaa atgtctgagt gtgttcttgg acaatcaaaa 3060
agagttgact tttgtggaaa gggctaccac cttatgtcct tcccacaagc agccccgcat 3120
ggtgttgtct tcctacatgt cacgtatgtg ccatcccagg agaggaactt caccacagcg 3180
ccagcaattt gtcatgaagg caaagcatac ttccctcgtg aaggtgtttt tgtgtttaat 3240
ggcacttctt ggtttattac acagaggaac ttcttttctc cacaaataat tactacagac 3300
aatacatttg tctcaggaaa ttgtgatgtc gttattggca tcattaacaa cacagtttat 3360
gatcctctgc aacctgagct tgactcattc aaagaagagc tggacaagta cttcaaaaat 3420
catacatcac cagatgttga tcttggcgac atttcaggca ttaacgcttc tgtcgtcaac 3480
attcaaaaag aaattgaccg cctcaatgag gtcgctaaaa atttaaatga atcactcatt 3540
gaccttcaag aattgggaaa atatgagcaa tatattaaat ggccttggta tgtttggctc 3600
ggcttcattg ctggactaat tgccatcgtc atggttacaa tcttgctttg ttgcatgact 3660
agttgttgca gttgcctcaa gggtgcatgc tcttgtggtt cttgctgcaa gtttgatgag 3720
gatgactctg agccagttct caagggtgtc aaattacatt acacataa 3768

<210> 9
<211> 4092
<212> DNA
<213> Bovine coronavirus

<400> 9
atgtttttga tacttttaat ttccttacca atggcttttg ctgttatagg agatttaaag 60
tgtactacgg tttccattaa tgatgttgac accggtgctc cctctattag cactgatatt 120
gtcgatgtta ctaatggttt aggtacttat tatgttttag atcgtgtgta tttaaatact 180
acgttgttgc ttaatggtta ctaccctact tcaggttcta catatcgtaa tatggcactg 240
aagggaactt tactattgag cagactatgg tttaaaccac cttttctttc tgattttatt 300
aatggtattt ttgctaaggt caaaaatacc aaggttatta aaaagggtgt aatgtatagt 360
gagtttcctg ctataactat aggtagtact tttgtaaata catcctatag tgtggtagta 420
caaccacata ctaccaattt ggataataaa ttacaaggtc tcttagagat ctctgtttgc 480
cagtatacta tgtgcgagta cccacatacg atttgtcatc ctaagctggg taataaacgc 540
gtagaactat ggcattggga tacaggtgtt gtttcctgtt tatataagcg taatttcaca 600
tatgatgtga atgctgatta cttgtatttc catttttatc aagaaggtgg tacttttttat 660
gcatatttta cagacactgg tgttgttact aagtttctgt ttaatgttta tttaggcacg 720
gtgctttcac attattatgt cctgcctttg acttgttcta gtgctatgac tttagaatat 780

tgggttacac ctctcacttc taaacaatat ttactagctt tcaatcaaga tggtgttatt 840
tttaatgctg ttgattgtaa gagtgatttt atgagtgaga ttaagtgtaa aacactatct 900
atagcaccat ctactggtgt ttatgaatta aacggttaca ctgttcagcc aattgcagat 960
gtttaccgac gtatacctaa tcttcccgat tgtaatatag aggcttggct taatgataag 1020
tcggtgccct ctccattaaa ttgggaacgt aagacctttt caaattgtaa ttttaatatg 1080
agcagcctga tgtcttttat tcaggcagac tcatttactt gtaataatat tgatgctgct 1140
aagatatatg gtatgtgttt ttccagcata actatagata agtttgctat acccaatggt 1200
aggaaggttg acctacaatt gggcaatttg ggctatttgc agtcttttaa ctatagaatt 1260
gatactactg ctacaagttg tcagttgtat tataatttac ctgctgctaa tgtttctgtt 1320
agcaggttta atccttctac ttggaatagg agatttggtt ttacagaaca atttgttttt 1380
aagcctcaac ctgtaggtgt ttttactcat catgatgttg tttatgcaca acattgtttt 1440
aaagctccca aaaatttctg tccgtgtaaa ttggatgggt ctttgtgtgt aggtaatggt 1500
cctggtatag atgctggtta taaaaatagt ggtataggca cttgtcctgc aggtactaat 1560
tatttaactt gccataatgc tgcccaatgt gattgtttgt gcactcccga ccccattaca 1620
tctaaatcta cagggcctta caagtgcccc caaactaaat acttagttgg cataggtgag 1680
cactgttcgg gtcttgctat taaaagtgat tattgtggag gtaatccttg tacttgccaa 1740
ccacaagcat ttttgggttg gtctgttgac tcttgtttac aaggggatag gtgtaatatt 1800
tttgctaatt ttatttttca tgatgttaat agtggtacta cttgttctac tgatttacaa 1860
aaatcaaaca cagacataat tcttggtgtt tgtgttaatt atgatcttta tggtattaca 1920
ggccaaggta tttttgttga ggttaatgcg acttattata atagttggca gaacctttta 1980
tatgattcta atggtaatct ctatggtttt agagactact taacaaacag aacttttatg 2040
attcgtagtt gctatagcgg tcgtgtttca gcggcctttc atgctaactc ttccgaacca 2100
gcattgctat ttcggaatat taaatgcaat tacgttttta ataatactct ttcacgacag 2160
ctgcaaccta ttaactattt tgatagttat cttggttgtg ttgtcaatgc tgataatagt 2220
acttctagtg ttgttcaaac atgtgatctc acagtaggta gtggttactg tgtggattac 2280
tctacaaaaa gacgaagtcg tagagcgatt accactggtt atcggtttac taattttgag 2340
ccatttactg ttaattcagt aaatgatagt ttagaacctg taggtggttt gtatgaaatt 2400
caaatacctt cagagtttac tataggtaat atggaggagt ttattcaaac aagctctcct 2460
aaagttacta ttgattgttc tgcttttgtc tgtggtgatt atgcagcatg taaatcacag 2520
ttggttgaat atggtagctt ctgtgacaat attaatgcta tactcacaga agtaaatgaa 2580
ctacttgaca ctacacagtt gcaagtagct aatagtttaa tgaatggtgt cactcttagc 2640
actaagctta aagatggcgt taatttcaat gtagacgaca tcaatttttc ccctgtatta 2700
ggttgtttag gaagcgcttg taataaagtt tccagcagat ctgctataga ggatttactt 2760
ttttctaaag taaagttatc tgatgtcggt ttcgttgagg cttataataa ttgtactgga 2820
ggtgccgaaa ttagggacct catttgtgtg caaagttata atggtatcaa agtgttgcct 2880
ccactgctct cagtaaatca gatcagtgga tacactttgg ctgccacctc tgctagtctg 2940
tttcctcctt tgtcagcagc agtaggtgta ccattttatt taaatgttca gtatcgtatt 3000
aatgggattg gtgttaccat ggatgtgtta agtcaaaatc aaaagcttat tgctaatgca 3060
tttaacaatg ctcttgatgc tattcaggaa gggtttgatg ctaccaattc tgctttagtt 3120
aaaattcaag ctgttgttaa tgcaaatgct gaagctctta ataacttatt gcaacaactc 3180
tctaatagat ttggtgctat aagttcttct ttacaagaaa ttctatctag actggatgct 3240
cttgaagcgc aagctcagat agacagactt attaatgggc gtcttaccgc tcttaatgtt 3300
tatgtttctc aacagcttag tgattctaca ctagtaaaat ttagtgcagc acaagctatg 3360
gagaaggtta atgaatgtgt caaaagccaa tcatctagga taaatttttg tggtaatggt 3420
aatcatatta tatcattagt gcagaatgct ccatatggtt tgtattttat ccactttagc 3480
tatgtcccta ctaagtatgt cactgcgaag gttagtcccg gtctgtgcat tgctggtgat 3540
agaggtatag cccctaagag tggttatttt gttaatgtaa ataatacttg gatgttcact 3600
ggtagtggtt attactaccc tgaacccata actggaaata atgttgttgt tatgagtacc 3660
tgtgctgtta actatactaa agcgccggat gtaatgctga acatttcaac acccaacctc 3720
catgatttta aggaagagtt ggatcaatgg tttaaaaacc aaacatcagt ggcaccagat 3780
ttgtcacttg attatataaa tgttacattc ttggacctac aagatgaaat gaataggtta 3840
caggaggcaa taaaagtttt aaatcagagc tacatcaatc tcaaggacat tggtacatat 3900
gagtattatg taaaatggcc ttggtatgta tggcttttaa ttggctttgc tggtgtagct 3960
atgcttgttt tactattctt catatgctgt tgtacaggat gtgggactag ttgttttaag 4020
atatgtggtg gttgttgtga tgattatact ggacaccagg agttagtaat taaaacatca 4080
catgacgact aa 4092

<210> 10
<211> 4050
<212> DNA
<213> Porcine hemagglutinating encephalomyelitis coronavirus

<400> 10
atgtttttta tacttttaat ctccctgcct tctgcttttg cagttatagg ggatttaaag 60
tgtactactt cattaattaa tgacgttgac actggtgtgc catctattag ctctgaagtt 120
gttgatgtca ctaatggttt ggggactttc tatgttttag atcgtgtcta tttaaatacc 180
acattgttgc tcaatggtta ttacccaatt tcaggtgcta catttcgtaa tatggctctg 240
aaaggaactc gattattgag caccttgtgg tttaagccgc ctttttttatc acctttttaat 300
gatggtattt ttgccaaggt taaaaacagc agattttcta aagatggtgt tatttatagt 360
gagtttcctg ctattactat aggtagtact tttgtaaata cttcctatag catagtagta 420
gagcctcata cctcacttat taatggtaat ttacaaggtt tgttgcaaat ttctgtttgt 480
caatacacta tgtgtgaata cccacatact atttgtcatc ctaatttggg taatcaacgc 540
atagaattat ggcattatga cacagatgtt gtttcttgtt tatacaggcg taatttcaca 600
tatgatgtga atgctgatta tttatatttt cacttttatc aggaaggtgg cacttttttat 660
gcatacttta cagatactgg ttttgtgacc aagtttctgt ttaagttgta tttaggcact 720
gtgctgtcac attattatgt tatgccattg acttgtaata gcgctttatc tttagaatac 780
tgggttacac ctctcactac tagacaattt cttctagcct ttgaccagga tggtgtttta 840
taccatgctg ttgattgtgc tagtgatttt atgagtgaga ttatgtgtaa aacttcttca 900
attacaccac ctactggtgt ttatgaacta aacggttaca cagttcaacc tgttgccact 960
gtatatcgta gaatacctga tttacccaat tgcgatatcg aagcttggct taattctaag 1020
accgtttctt cgcctcttaa ttgggaacgt aaaatttttt ctaattgtaa ttttaacatg 1080
ggcaggctga tgtcttttat tcaggctgac tcttttggtt gtaacaatat tgatgcttct 1140
cgcttatatg gtatgtgttt tggtagcatt actattgata agtttgctat acccaatagt 1200
agaaaggttg atctgcaagt gggtaaatct ggttatttac aatcttttaa ttataagatt 1260
gacactgctg ttagcagttg tcaactctat tatagtttgc ctgcagcaaa cgtatctgtc 1320
actcattata atccttcatc ttggaataga aggtatgggt ttaataatca gagttttggt 1380
tccagaggcc ttcatgatgc tgtttattca cagcaatgtt ttaatacacc taacacatat 1440
tgtccttgta gaacaagtca atgcataggt ggtgcaggca caggaacttg tcctgtaggc 1500
accactgtgc gcaagtgttt tgctgcagtt acaaaagcta ctaagtgtac ttgctggtgt 1560
caaccagatc cttccacata taaaggtgta aacgcctgga cttgtccgca atctaaagtt 1620
tctatacaac caggtcagca ttgccctggt ttgggtcttg tggaggatga ttgctctggc 1680
aacccttgca cttgtaaacc acaggctttc ataggctgga gttcagaaac ttgtttgcaa 1740
aatggtaggt gtaatatttt tgctaatttt attctgaatg atgttaatag cggtacaacc 1800
tgttctactg atttacaaca gggtaatact attattacta ctgatgtttg tgttaattat 1860
gacctatatg gcattacagg ccagggcata cttatagaag ttaatgccac ttattataat 1920
agttggcaga atcttcttta tgattctagt ggtaatctct atggctttag agattattta 1980
tcaaatagaa cttttcttat tcgtagctgc tatagtggaa gagtttcagc agtttttcat 2040
gctaactcat ctgaaccagc tttgatgttt cgtaatctta aatgcagcca cgtttttaat 2100
aataccattt taagacaaat acagcttgtt aactattttg atagttacct tggttgtgtt 2160
gttaatgctt ataataatac agctagtgct gtaagtactt gtgatttaac cgttggtagc 2220
ggctattgtg ttgattatgt tacagcactt agatcacgta gatcttttac tacaggttat 2280
cgctttacta attttgaacc atttgccgct aatttggtaa atgatagtat agaacctgtt 2340
ggtggtttgt atgaaataca gataccttca gagtttacca ttggtaattt agaagagttc 2400
attcaaacga gatcccctaa ggttactata gactgtgcta catttgtttg tggtgactat 2460
gcagcatgta gacaacagtt agctgagtat ggtagttttt gtgagaacat taatgctata 2520
ctcacagaag taaatgaact acttgacact acacagttgc aagtagctaa tagtttaatg 2580
aatggagtca cccttagtac caagattaag gatggcatta atttcaatgt tgacgatatc 2640
aacttctccc ctgtattagg ttgtttagga agcgaatgta atagagcttc cactagatct 2700
gctatagagg atttactttt tgataaagta aaattgtctg atgtcggctt tgtacaggcc 2760
tataataact gcactggagg tgccgaaatt agggatctca tttgtgtgca aagttataat 2820
ggtatcaaag tgttgcctcc attgttatct gaaaatcaga tcagtggcta cactttggca 2880
gccaccgctg ctagcttatt ccctccttgg acagctgcag caggtgtacc attttattta 2940
aatgttcagt atcgtataaa tgggcttggc gtcactatgg atgtgctaag tcaaaaccaa 3000
aagcttattg ctagtgcatt taacaatgct cttgatgcta tccaggaagg gttcgacgca 3060
accaattctg ctttagttaa aattcaggct gttgttaatg caaatgctga agcacttaat 3120

aacttattgc agcaactctc taacagattt ggtgccataa gtgcctcttt acaagaaatt 3180
ttatccaggc tcgatgctct tgaagctaaa gctcagatag acagacttat caatgggcgt 3240
ctcaccgctc ttaatgctta tgtttctcag cagcttagtg attctacact agtaaaattt 3300
agtgcagcac aagctattga gaaagttaat gaatgtgtta aaagccaatc atctaggata 3360
aatttctgtg gtaatggtaa tcatattata tcattagtac agaatgctcc atatggtttg 3420
tattttatcc attttagcta tgtccccacc aagtatgtta cagcaaaggt tagtcctggt 3480
ttgtgcattg ctggcgatat aggaatatcg cctaagagtg gttattttat taatgtaaat 3540
aattcttgga tgttcactgg tagtagctat tactaccctg aacctataac ccaaaataat 3600
gttgttgtga tgagtacctg tgctgttaat tatactaaag caccggatct aatgctgaac 3660
acatcgacac ccaaccttcc tgacttcaag gaagaattgt atcaatggtt taaaaaccaa 3720
tcttcagtgg caccagattt gtcacttgat tatattaatg ttacgttctt ggacctacaa 3780
gatgaaatga ataggttaca agaggctata aaagttttaa atcagagcta catcaatctc 3840
aaggacattg gtacatatga gtattatgtg aaatggcctt ggtatgtatg gcttttaatt 3900
ggccttgctg gtgtagctat gcttgtttta ctattcttca tatgctgctg tacaggatgt 3960
gggactagtt gttttaagaa atgtggcggt tgttgtgatg attatactgg acaccaggag 4020
tttgtaatca aaacttcaca tgacgattaa 4050

<210> 11
<211> 4086
<212> DNA
<213> Murine hepatitis virus

<400> 11
atgctattcg tgtttttaac cttgttgccc tcttctctag ggtatattgg tgattttaga 60
tgtatccaac ttgtaaatac cgacacctct aatgccagcg ctccaagcgt tagtacagag 120
gtagttgatg tttccaaagg gattggtact tattatgttt tagatcgagt ctatttaaat 180
gccacactat tgcttactgg ttattaccct gtagatgggt ccatgtatag aaacatggct 240
ctaacgggaa ttaataccat aagccttaat tggtacaaac cacccttttt atcagagttt 300
aatgatggca tatttgctaa ggtaaagaac cttaaagcat ctttgcccaa agattctatt 360
tcatatttcc ctactataat tataggtagt aattttgtca ccacttccta tactgtagta 420
ttggaaccgt ataatggtat aattatggca tccatttgcc agtataccat ttgtcaacta 480
ccgtacacgg attgcaaacc gaatacgggc ggtaataagt taattggctt ttggcacaca 540
gagctaaaat cccctgtgtg cattttaaag cgtaatttta cgtttaatgt taatgccgaa 600
tggctttatt ttcattttta ccagcagggt ggtacttttt atgcgtatta tgcggatgtt 660
tcttctgcta ctacgttttt gtttagtatg tatattggtg atgtgttaac acaatatttt 720
gtgttgcctt atatgtgtac tctcactaca acaggtgtct tttcaccgca gtattgggtt 780
acacctcttg tcaagcgcca atatttattt aattttaatc aaaagggtat tattactagt 840
gctgttgatt gtgctagtag ttataccagc gaaataaagt gtaagaccca aagtatgaat 900
cccaatacgg gagtttatga tttatccggt tacaccgtcc aacctgtagg attagtgtac 960
cggcgtgtta gaaatttgcc tgattgtaaa attgaggaat ggctaactgc taagtctgta 1020
ccttctcctc tcaattggga gcgcaaaaca tttcaaaatt gtaacttcga cctgagcagt 1080
ctattaagat ttgttcaggc tgagtcactc tcatgtagta atatagatgc ttccaaggtt 1140
tatggtatgt gctttggtag tatatctata gacaagtttg cgatacccaa tagacgccga 1200
gttgatttgc agctaggcaa ctctgggttt ttgcaatcct ttaattacaa aatagataca 1260
agagctactt cgtgtcagct ctattatagt cttgcaaaaa ataatgtcac tgtcaataac 1320
cataacccgt cctcttggaa caggcgttat gggttcaatg atgtggctac atttggaact 1380
ggtaaacatg acgttgctta tgctgaggct tgttttaccg tgggagcatc atattgccct 1440
tgcgcgaacc ccagcatagt gtcgccatgt accactggaa aacctaactt tgccaattgc 1500
cctacaggca cctcgaatcg tgagtgcact gttatgccat tggctaataa tcaatttaag 1560
tgtgattgca cttgtaaccc tagtcctcta accacctatg atcttagatg tctccaagca 1620
agaagcatgc ttggcgtagg tgatcattgt gaaggtctag gagttttaga agataaatgt 1680
ggtggcagca acacctgcaa ttgttctgct catgcctttg ttggctgggc taaggatagt 1740
tgcttggcta atggccgctg tcacattttt agtaatttga tgttaaatgg cattaatagt 1800
ggtactacat gttccatgga tttgcaattg cctaatactg aagtggtcac tggcgtctgc 1860
gtcaaatatg acctctacgg tataacaggc caaggtattt ttaaggaggt taaggctgac 1920
tattatcata gttggcaaaa cctcttatat gatgttaatg gcaacttaat cggatttcgc 1980
gattttgttg ctaataagag ttatactatt cgaagttgct atagtgggcg ggtctcggct 2040

gcatatcatc aagatgcacc agaaccagcg ctactatatc gcaatttaaa atgtgactat 2100
gtctttaaca acaacatatc ccgtgaggag acaccactta actatttcga tagttatctt 2160
ggttgtgttg ttaatgctga caactcaact gaagaagctg ttgacgcgtg tgatttgcgt 2220
atgggtagtg ggctttgtgt caactattca acgtctcacc gcgctcgcag ctctgtcagc 2280
acgggttata aattaactac ttttgaacca tttacagtcc gcattgtcaa tgatagtgtt 2340
gagtctgttg atgggttata tgagctgcaa ataccaacca actttactat agctagccat 2400
caggagttcg ttcaaacgag gtctccaaag gttactatag actgtgctgc atttgtctgt 2460
ggtggccaca cagcatgccg tcagcagttg gttgagtacg gctcattctg tgataatatt 2520
aatgccattc ttggcgaggt aaataacctc atagatacta tgcaacttca agttgcaagt 2580
gctttaatcc aaggtgtcac gttaagctca cgcttatcgg atggcattgg tggtcaaata 2640
gatgatatta attttagtcc tctgcttggt tgtttaggtt ctgactgtgg cgaagttacc 2700
atggcagctc aaaccggacg atctgctata gaggatgtat tatttgacaa agtcaaactc 2760
tctgatgttg gctttgtcga agcatataac aattgcactg gaggccaaga agttagagac 2820
ctactttgtg tgcaatcttt taatggcatc aaagtgctac cgcctgtgtt gtctgagaat 2880
caaatttctg gttatacagc gggagctact gtatctgcta tgttcccatg gtctgcagct 2940
gcaggtgtgc cattttcttt aagtgttcaa tatagaatta atggtcttgg tgtcactatg 3000
aatgttctta gtgaaaatca gaaaatgata gcaagtgcat ttaacaacgc gataggtgct 3060
atacaggaag ggtttgctgc aaccaattct gccttagcaa aaatgcagtt cgttgtcaat 3120
gcaaatgcgg aagcactcaa taatttatta aaccagcttt ccaataggtt tggtgcaatt 3180
agtgcttctt tacaagaaat tctatctcgc ctagatgctc ttgaagcgca ggctcagata 3240
gaccgtctta ttaatggcag attaactgca cttaatgcat atgtctctaa gcagctgagt 3300
gacatgaccc ttgttaaggt gagtgcagcc caggctatag agaaagttaa tgagtgtgtt 3360
aaaagccaat catctaggat aaatttctgt ggcaatggca atcatatatt gtcattagtc 3420
cagaatgcgc cttatggttt atattttatt catttcagct atgtgcctac ttcctttaca 3480
acggcaaatg tgagtcctgg gctatgcatt tctggtgata gaggattagc acctaaagct 3540
ggatattttg ttcaagatga tggagagtgg aagttcacag gtagtaatta ttattaccct 3600
gaacccatta cagataaaaa tagtgtcgtg atgagtagtt gcgcagcaaa ctacacaaag 3660
gcacctgaag ttttcttgaa cacttcaata cctaatctac ccgactttaa ggaggagtta 3720
gataaatggt ttaaaaatca gacgtctatt gcgcctgatt tatctctcga tttcgagaaa 3780
ttaaacgtta ccctcctgga cctgactgat gagatgaaca ggattcagga tgcaattaag 3840
aagttaaatg agagttacat caacctcaag gacgttggca catatgaaat gtatgtgaaa 3900
tggccttggt atgtgtggtt gctaattgga ttagctggtg tagctgtttg tgtgttgtta 3960
tttttcatat gttgctgcac gggttgtggc tcatgttgtt tcaagaagtg tggaaattgt 4020
tgtgatgagt gtggaggaca ccaggacagt attgtgatac ataatatttc ctctcatgag 4080
gattga 4086

<210> 12
<211> 4083
<212> DNA
<213> Rat coronavirus

<400> 12
atgctattcg tgtttttaac cctattgccc tcttgtctag ggtatattgg tgattttaga 60
tgtatcaacc ttgtaaacac ccgcatttct aatgcgcgcg cacccagtgt tagcacagag 120
gtagttgatg tttctaaagg tcttggtaca tattacgttt tagatcgtgt ttatttaaat 180
gccacgttat tgcttactgg ttactaccct gtagatgggt ccatgtatcg taacatggct 240
ctaatgggta ctaatacctt aagccttaat tggtttgaac cgccctttt atcagagttt 300
aacgatggca tatatgctaa ggtaaagaac ctcaaagcat ctttgcccat aggctcggct 360
tcatactttc ctactataat tataggtagt aattttgtta atacttccta tactgtagta 420
ttggaaccat acaatggtat tattatggca tctatttgcc agtataccat ttgtcaatta 480
ccgcacacgg attgcaaacc taacacgggc ggtaacacgc taattggttt ttggcacaca 540
gatttaaggc ctccggtgtg cattttaaag cgtaatttta cgtttaatgt taatgccgaa 600
tggctttatt ttcattttta ccagcagggt ggtacttttt atgcgtatta tgcagatgtt 660
tcttctgcca ctacgttttt gtttagttcg tatattggtg ctgtgttaac acagtatttt 720
gtgttgcctt atatgtgtag tcccactacc tcaggtgttt cctcaccgca gtattgggtt 780
acaccacttg ttaagcgcca atatttattt aattttaacc aaaagggtat tattactagc 840
gctgttgatt gtgctagtag ttataccagt gaaataaagt gtaagactca aagtatgaat 900

cccaatacgg gagtctatga tttatccggt tacaccgtcc aacctgtagg actagtgtac 960
cggcgtgtta gaaatttgcc tgattgtaaa attgaggaat ggttggctgc taacacagta 1020
ccctctcctc tcaattggga gcgcaaaaca tttcaaaatt gtaacttcaa cctgagcagt 1080
ctattaagat ttgttcaggc tgagtcactc tcatgtagta atatagatgc ttccaaggtt 1140
tatggaatgt gctttggtag catatctata gataaatttg caatacccaa cagtcgccgt 1200
gttgatcttc agctaggtaa atcgggtctt ttgcaatctt ttaattataa aattgataca 1260
agagcgacct cgtgtcagct ctattacagt cttgcacaag ataatgtcac tgtcattaac 1320
cacaacccat cctcctggaa taggcgttat ggatttaatg acgtggctac atttcatagt 1380
ggtgaacatg acgttgctta tgcagaggca tgtttcactg ttggagcttc atattgccct 1440
tgtgcgaagc ccagcacagt ctattcatgt gtcacaggta aacctaagtc tgctaattgc 1500
ccaacaggta cctcgaatcg tgagtgtaat gttcaggctt caggttttaa gtctaagtgc 1560
gattgcacat gtaaccctag tcctctaacc acctatgatc ctagatgtct tcaagcgcgg 1620
agcatgcttg gcgtaggtga tcattgtgaa ggtctaggta ttttagaaga taaatgtggt 1680
ggcagcaaca tatgcaattg ttcggctgat gcctttgttg gctgggctat ggacagctgt 1740
ctatctaatg cccgctgcca tatttttagt aatttgatgt taaatggcat taatagtggt 1800
actacatgtt ccacggattt tcaattgcct aatacggaag tggtcactgg cgtttgtgtc 1860
aagtatgacc tctacggtag tacaggccaa ggtgttttta aggaggttaa ggctgattat 1920
tacaatagtt ggcagaacct cttatatgat gttaatggta acttaaatgg tttccgtgac 1980
attgttacca ataagactta tttattaaga agttgctata gtgggcgcgt ttcggctgca 2040
tatcatcaag atgcacctga accagcgcta ctatatcgca atttaaaatg tgattatgtg 2100
tttaataaca acatatcccg tgaggagaca ccacttaact attttgatag ttatttgggt 2160
tgtgttatta atgctgataa ctcaactgag cagtctgttg acgcgtgtga tttgcgtatg 2220
ggtagtgggc tttgtgtcaa ctattcaatc gctcaccgtg cgcgcaggtc tgtcagtacg 2280
ggttataaat taactacttt tgaaccattt acagtcagca ttgtcaatga tagtgttgag 2340
tctgttggtg gattatatga gatgcaaata cctactaatt ttactatagc tagccatcag 2400
gagttcattc aaacgaggtc tccgaaggtt actatagatt gtgctgcatt tgtctgtggt 2460
gattatacag cgtgtagaca acagttggtt gattatggct ctttttgtga taatattaat 2520
gccattcttg gcgaggtgaa taacctcata gatactatgc aattacaagt tgctagtgct 2580
ctgatccaag gtgtcacgct aagttcacgc ttggcagatg gcatctcagg tcagattgat 2640
gatattaatt ttagtcctct tctaggttgc cttggctcag attgtagcga aggcaccaag 2700
gcagcgcaag ggcgatctgc tatagaggat gtattatttg ataaggtcaa actctctgat 2760
gttggctttg tcgaatcata taataattgc actggaggtc aagaagttag agacctactt 2820
tgtgtgcaat cttttaatgg cattaaagtg ctaccgcctg tattatccga gagtcaaatc 2880
tctggttata cagcgggagc tactgcatct gctatgttcc ctccatggtc tgcagctgca 2940
ggtgtgccat ttgctttaag tgttcaatat agaattaatg gtcttggtgt cactatgaat 3000
gttcttagtg aaaaccagaa aatgatagct agttcattta acaacgcgat aggtgctata 3060
caggaagggt tcgatgcaac caattctgct ttagcgaaaa ttcagtccgt tgtcaacgca 3120
aatgcagaag cactcaataa ccttttgaat cagctttcca ataggtttgg tgcaattagt 3180
gcttctttac aggaaattct atctcgcctc gatgctcttg aagctcaggc tcagatagac 3240
cgtcttatta atggcagatt aactgcactt aatgcatatg tctctaagca gctgagcgac 3300
atgaccctta ttaaggtgag tgctgcccag gctatagaga aagttaatga gtgttgttaaa 3360
agccaatcac ctaggataaa tttctgtggc aatggcaatc atatattgtc attagtccag 3420
aatgcgcctt acggtttata ttttattcat ttcagctatg tgcctacatc ctttacaacg 3480
gtaaatgtga gtcctggact atgcatttct ggtgatagag gattagcacc taaagctgga 3540
tattttgttc aagatcatgg agaatggaag ttcacaggta gcaattatta ctaccctgaa 3600
tccattacag ataaaaacag tgtcgtgatg agtagttgcg cagtaaacta cacaaaggca 3660
cctgaagttt tcttgaacac ttcaataact aatctacccg actttaagga ggagttagat 3720
aaatggttta agaatcagac gtctattgtg cctgatttat ctttcgatat cgggaaatta 3780
aatgttacat tccttgacct gtcctatgag atgaacagga ttcaggatgc aattaagaat 3840
ttaaatgaga gttacatcaa cctcaaggaa attggcacat atgagatgta tgtgaaatgg 3900
ccttggtatg tttggctgct aattggatta gctggtgtag ctgtttgtgt tttgttattt 3960
tttatatgtt gctgcacagg ttgtggctct tgttgtttta agaaatgtgg aaattgttgt 4020
gatgagtatg gaggacgtca ggcaggtatt gtgatacata atatttcctc tcatgaggat 4080
tga 4083

<210> 13
<211> 3489
<212> DNA
<213> Infectious bronchitis virus

<400> 13
atgttggtaa cacctctttt actagtgact cttttgtgtg cactatgtag tgctgttttg 60
tatgacagta gttcttacgt ttactactac caaagtgcct tcagaccacc tagtggttgg 120
catttacaag ggggtgctta tgcggtagtt aacatttcta gcgaatttaa taatgcaggc 180
tcttcatcag ggtgtactgt tggtattatt catggtggtc gtgttgttaa tgcttcttct 240
atagctatga cggcaccgtc atcaggtatg gcttggtcta gcagtcagtt ttgtactgca 300
cactgtaatt tttcagatac tacagtgttt gttacacatt gttataaaca tggtgggtgt 360
cctttaactg gcatgcttca acagaatctt atacgtgttt ctgctatgaa aaatggccag 420
cttttctata atttaacagt tagtgtagct aagtacccta cttttagatc atttcagtgt 480
gttaataatt taacatccgt atatttaaat ggtgatcttg tttacacctc taatgagacc 540
atagatgtta catctgcagg tgtttatttt aaagctggtg gacctataac ttataaagtt 600
atgagagaag ttaaagccct ggcttatttt gttaatggta ctgcacaaga tgttatttg 660
tgtgatggat cacctagagg cttgttagca tgccagtata atactggcaa tttttcagat 720
ggcttttatc cttttactaa tagtagttta gttaagcaga agtttattgt ctatcgtgaa 780
aatagtgtta atactacttg tacgttacac aatttcattt ttcataatga gactggcgcc 840
aaccctaatc ctagtggtgt tcagaatatt caaacttacc aaacaaaaac agctcagagt 900
ggttattata attttaattt ttccttttctg agtagttttg tttataagga gtctaatttt 960
atgtatggat cttatcaccc aagttgtaaa tttagactag aaactattaa taatggcttg 1020
tggtttaatt cactttcagt ttcaattgct tacggtcctc ttcaaggtgg ttgcaagcaa 1080
tctgtcttta aaggtagagc aacttgttgt tatgcttatt catatggagg tccttcgctg 1140
tgtaaaggtg tttattcagg tgagttagat cataattttg aatgtggact gttagtttat 1200
gttactaaga gcggtggctc tcgtatacaa acagccactg aaccgccagt tataactcaa 1260
aacaattata ataatattac tttaaatact tgtgttgatt ataatatata tggcagaact 1320
ggccaaggtt ttattactaa tgtgaccgac tcagctgtta gttataatta tctagcagac 1380
gcaggtttgg ctattttaga tacatctggt tccatagaca tctttgttgt acaaggtgaa 1440
tatggtctta attattataa ggttaaccct tgcgaagatg tcaaccagca gtttgtagtt 1500
tctggtggta aattagtagg tattcttact tcacgtaatg agactggttc tcagcttctt 1560
gagaaccagt tttacatcaa aatcactaat ggaacacgtc gttttagacg ttctattact 1620
gaaaatgttg caaattgccc ttatgttagt tatggtaagt tttgtataaa acctgatggc 1680
tcaattgcca caatagtacc aaaacaattg gaacagtttg tggcaccttt atttaatgtt 1740
actgaaaatg tgctcatacc taacagtttc aacttaactg ttacagatga gtacatacaa 1800
acgcgtatgg ataaggtcca aattaattgc ctgcagtatg tttgtggcag ttctctggat 1860
tgtagaaagt tgtttcaaca atatgggcct gtttgcgaca acatattgtc tgtagtaaat 1920
agtgttggtc aaaaagaaga tatggaactt ttgaatttct attcttctac taaaccggct 1980
ggttttaata caccagttct tagtaatgtt agcactggtg agtttaatat ttctcttctg 2040
ttaacaaatc ctagtagtcg tagaaagcgt tctcttattg aagaccttct atttacaagc 2100
gttgaatctg ttggactacc aacaaatgac gcatataaaa attgcactgc aggaccttta 2160
ggctttttta aggaccttgc gtgtgctcgt gaatataatg gtttgcttgt gttgcctcct 2220
atcataacag cagaaatgca agctttgtat actagttctc tagtagcttc tatggctttt 2280
ggtggtatta ctgcagctgg tgctatacct tttgccacac aactgcaggc tagaattaat 2340
cacttgggta ttacccagtc acttttgttg aagaatcaag aaaaaattgc tgcttccttt 2400
aataaggcca ttggtcatat gcaggaaggt tttagaagta catctctagc attacaacaa 2460
attcaagatg ttgttagtaa acagagtgct attcttactg agactatggc atcacttaat 2520
aaaaattttg gtgctatttc ttctgtgatt caagaaatct accagcaatt tgacgccata 2580
caagcaaatg ctcaagtgga tcgtcttata actggtagat tgtcatcact ttctgtttta 2640
gcatctgcta agcaggcgga gtatattaga gtgtcacaac agcgtgagtt agctactcag 2700
aaaattaatg agtgtgttaa gtcacagtct attaggtact ccttttgtgg taatggacga 2760
catgttctaa ccataccgca aaatgcacct aatggtatag tgtttataca cttttcttat 2820
actccagata gttttgttaa tgttactgca atagtgggtt tttgtgtaaa gccagctaat 2880
gctagtcagt atgcaatagt gcccgctaat ggtaggggta tttttataca agttaatggt 2940
agttactaca tcactgcacg agatatgtat atgccaagag ctattactgc aggagatgta 3000
gttacgctta cttcttgtca agcaaattat gtaagtgtaa ataagaccgt cattactaca 3060
ttcgtagaca atgatgattt tgattttaat gacgaattgt caaaatggtg gaatgatact 3120

aagcatgagc taccagactt tgacaaattc aattacacag tacctatact tgacattgat 3180
agtgaaattg atcgtattca aggcgttata cagggtctta atgactctct aatagacctt 3240
gaaaaacttt caatactcaa aacttatatt aagtggcctt ggtatgtgtg gttagccata 3300
gcttttgcca ctattatctt catcttaata ctaggatggg ttttcttcat gactggttgt 3360
tgtggttgtt gttgtggatg ctttggcatt atgcctctaa tgagtaagtg tggtaagaaa 3420
tcttcttatt acacgacttt tgataacgat gtggtaactg aacaatacag acctaaaaag 3480
tctgtttga 3489

<210> 14
<211> 36
<212> DNA
<213> SARS coronavirus GD69

<400> 14
cccacgccag aaggtagatc acgaactaca cgtggg 36

<210> 15
<211> 95
<212> DNA
<213> SARS coronavirus GD69

<400> 15
ctctatgttt ataagggcta tcaacctata gatgtagttc gtgatctacc ttctggtttt 60
aacactttga aacctatttt taagttgcct cttgg 95

<210> 16
<211> 26
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic primer

<400> 16
ctctatgttt ataagggcta tcaacc 26

<210> 17
<211> 27
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic primer

<400> 17
ccaagaggca acttaaaaat aggtttc 27

<210> 18
<211> 12
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic primer

<400> 18
aggctgtaag aa 12

<210> 19
<211> 38
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic primer

<400> 19
gcccacgcca gaaggtagat cacgaactac acgtgggc 38

<210> 20
<211> 45
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic primer

<400> 20
taatacgact cactataggc tctatgttta taagggctat caacc 45

<210> 21
<211> 39
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic primer

<400> 21
aggctgtaag aaccaagagg caacttaaaa ataggtttc 39

<210> 22
<211> 231
<212> DNA
<213> SARS coronavirus urbani

<400> 22
atgtactcat tcgtttcgga agaaacaggt acgttaatag ttaatagcgt acttcttttt 60
cttgctttcg tggtattctt gctagtcaca ctagccatcc ttactgcgct tcgattgtgt 120
gcgtactgct gcaatattgt taacgtgagt ttagtaaaac caacggttta cgtctactcg 180
cgtgttaaaa atctgaactc ttctgaagga gttcctgatc ttctggtcta a 231

<210> 23
<211> 231
<212> DNA
<213> SARS coronavirus Tor2

<400> 23
atgtactcat tcgtttcgga agaaacaggt acgttaatag ttaatagcgt acttcttttt 60
cttgctttcg tggtattctt gctagtcaca ctagccatcc ttactgcgct tcgattgtgt 120
gcgtactgct gcaatattgt taacgtgagt ttagtaaaac caacggttta cgtctactcg 180
cgtgttaaaa atctgaactc ttctgaagga gttcctgatc ttctggtcta a 231

<210> 24
<211> 234
<212> DNA
<213> Human coronavirus OC43

<400> 24
atgttcctta agctagtgga tgatcatgct ttggttgtta atgtactact ctggtgtgtg 60
gtgcttatag tgatactact agtgtgtatt acaataatta aactaattaa gctttgtttc 120
acttgccata tgttttgtaa tagaacagtt tatggcccca ttaaaaatgt gtaccacatt 180
taccaatcat atatgcacat agacccttcc cctaaacgag ttattgattt ctaa 234

<210> 25
<211> 231
<212> DNA
<213> Porcine epidemic diarrhea virus

<400> 25
atgctacaat tagtgaatga taatggtcta gtagttaatg ttatactttg gcttttcgta 60
ctctttttcc tgcttattat aagcattacc ttcgtccaat tggttaatct gtgcttcact 120
tgtcaccggt tgtgtaatag cgcagtttat acacctatag ggcgcctgta tagagtttat 180
aagtcttaca tgcgaattga ccccctcccc agtactgtta ttgacgtata a 231

<210> 26
<211> 249
<212> DNA
<213> Transmissible gastroenteritis virus

<400> 26
atgacgtttc ctagggcatt gactgtcata gatgacaatg gaatggtcat taacatcatt 60
ttctggttcc tgttgataat tatattgata ttactttcaa tagcattgct aaatataatt 120
aagctatgca tggtgtgttg caatttagga aggacagtta ttattgttcc agcgcaacat 180
gcttacgatg cctataagaa ttttatgcga attaaagcat acaaccccga tggagcactc 240
cttgcttga 249

<210> 27
<211> 249
<212> DNA
<213> Porcine respiratory coronavirus

<400> 27
atgacgtttc ctagggcatt gactgtcata gatgacaacg gaatggtcat tagcatcatt 60

ttttggttcc tgttgataat tatattgata ttactttcaa tagcattgct aaatataatt 120
aagctatgca tggtgtgttg caatttagga aggacagtta ttattgttcc agtgcaacat 180
gcttacgatg cctataagaa ttttatgcga attaaagcat acaaccctga tggagcactc 240
cttgtttga 249

<210> 28
<211> 249
<212> DNA
<213> Canine coronavirus

<400> 28
atgacgttcc ctcgggcatt gactgtcata gatgacaatg gaatggtcat tagtatcatt 60
ttctggttcc tgttgataat tatattgata ttattttcaa tagcattgct aaatataatt 120
aagctatgca tggtatgttg caatttagga agaacagtta ttattgttcc agctcgacat 180
gcctatgatg cctataagaa ttttatgcaa attagagcat acaaccctga tgaagcactc 240
cttgtttga 249

<210> 29
<211> 249
<212> DNA
<213> Feline coronavirus

<400> 29
atgacgttcc ctagggcatt tactatcata gatgaccatg gcatggttgt tagcgtcttc 60
ttctggctcc tgttgataat tatattgata ttgttttcaa tagcattgct aaatgttatt 120
aaattgtgca tggtatgttg caatttgggt aagactatta tagtactacc tgcacgccat 180
gcatatgatg cctataagac ctttatgcaa accaaggcat ataatcccga cgaagcattt 240
ttggtttga 249

<210> 30
<211> 267
<212> DNA
<213> Murine hepatitis virus

<400> 30
atgtttaatt tattccttac agatacagta tggtatgtgg ggcagattat ttttatagtc 60
gcagtgtgtt tgatggtcac cataatagtg gttgccttcc ttgcgtctat caaactttgt 120
attcaacttt gcggtttgtg taatactttg ttgctgtctc cttctatttg tgtgtataat 180
aggagtaagc agctttataa gtattataat gaagaagtga gaccgccccc gttagaggtg 240
gatgatataa taatccaaac attatga 267

<210> 31
<211> 267
<212> DNA
<213> Rat coronavirus

<400> 31
atgtttaatt tattccttat agacacagta tggtacgtgg ggcagattat ttttatagtc 60
gcagtgtgtt tgatggtcac cataattgtg gttgccttcc ttgcgtctat taaactttgt 120
attcaacttt gcggtttgtg taatactttg ttgctgtctc cttctattta tgtgtataat 180
aggagtaagc agctttataa gtattataat gaagaagtga gaccgccccc gttagaggtg 240
gatgatataa taatccaaac attatga 267

<210> 32
<211> 255
<212> DNA
<213> Bovine coronavirus

<400> 32
atgtttatgg ctgatgctta ttttgcagac actgtgtggt atgtggggca aataattttt 60
atagttgcca tttgtttatt ggttataata gttgtagtgg catttttggc aacttttaaa 120
ttgtgtattc aactttgcgg tatgtgtaat accttagtac tgtccccttc tatttatgtg 180
tttaatagag gtaggcagtt ttatgagttt tacaacgatg taaaaccacc agttcttgat 240
gtggatgacg tttag 255

<210> 33
<211> 327
<212> DNA
<213> Infectious bronchitis virus

<400> 33
atgaatttat tgaataagtc gctagaggag aatggaagtt ttctaacagc gctttacata 60
attgtaggat ttttagcact ttatcttcta ggtagagcac ttcaagcatt tgtacaggct 120
gctgatgctt gttgtttatt ttggtataca tgggtagtaa ttccaggagc taagggtaca 180
gcctttgtat acaagtatac atatggtaga aaacttaaca atccggaatt agaagcagtt 240
attgttaacg agtttcctaa gaacggttgg aataataaaa atccagcaaa ttttcaagat 300
gcccaacgag acaaattgta ctcttga 327

<210> 34
<211> 38
<212> DNA
<213> SARS coronavirus GD69

<400> 34
cctccgcacg aaagcaagaa aaagaagtac gccggagg 38

<210> 35
<211> 95
<212> DNA
<213> SARS coronavirus GD69

<400> 35
cggaagaaac aggtacgtta atagttaata gcgtacttct ttttcttgct ttcgtggtat 60
tcttgctagt cacactagcc atccttactg cgctt 95

<210> 36
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic primer

<400> 36
cggaagaaac aggtacgtta atag 24

<210> 37
<211> 20
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic primer

<400> 37
aagcgcagta aggatggcta 20

<210> 38
<211> 14
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic primer

<400> 38
tattgcagca gtac 14

<210> 39
<211> 40
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic primer

<400> 39
gcctccgcac gaaagcaaga aaaagaagta cgccggaggc 40

<210> 40
<211> 43
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic primer

<400> 40
taatacgact cactataggc ggaagaaaca ggtacgttaa tag 43

<210> 41
<211> 34
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic primer

<400> 41
tattgcagca gtacaagcgc agtaaggatg gcta 34

<210> 42
<211> 789
<212> DNA
<213> Transmissible gastroenteritis virus

<400> 42
atgaagattt tgttaatatt agcgtgtgtg attgcatgcg catgtggaga acgctattgt 60
gctatgaaat ccgatacaga tttgtcatgt cgcaatagta cagcgtctga ttgtgagtca 120
tgcttcaacg gaggcgatct tatttggcat cttgcaaact ggaacttcag ctggtctata 180
atattgatcg tttttataac tgtgctacaa tatggaagac ctcaattcag ctggttcgtg 240
tatggcatta aaatgcttat aatgtggcta ttatggcccg ttgttttggc tcttacgatt 300
tttaatgcat actcggaata ccaagtgtcc agatatgtaa tgttcggctt tagtattgca 360
ggtgcaattg ttacatttgt actctggatt atgtattttg taagatccat tcagttgtac 420
agaaggacta agtcttggtg gtctttcaac cctgaaacta aagcaattct ttgcgttagt 480
gcattaggaa gaagctatgt gcttcctctc gaaggtgtgc caactggtgt cactctaact 540
ttgctttcag ggaatttgta cgctgaaggg ttcaaaattg caggtggtat gaacatcgac 600
aatttaccaa aatacgtaat ggttgcatta cctagcagga ctattgtcta cacacttgtt 660
ggcaagaagt tgaaagcaag tagtgcgact ggatgggctt actatgtaaa atctaaagct 720
ggtgattact caacagaggc aagaactgat aatttgagtg agcaagaaaa attattacat 780
atggtataa 789

<210> 43
<211> 789
<212> DNA
<213> Porcine respiratory coronavirus

<400> 43
atgaagattt tgttgatatt agcgtgtgcg attgcatgca catgtggaga acgctattgt 60
gctatgaaag acgatacagg tttgtcatgt cgcaatggca cggcgtctga ttgtgagtca 120
tgcttcaaca gaggcgatct tatttggctt cttgcaaact ggaacttcag ctggtctata 180
atattgatca tttttattac tgtgctacaa tatggaagac ctcaattcag ctggttcgtg 240
tatggcatta aaatgcttat aatgtggcta ttatggccga ttgttttggc tcttacgatt 300
tttaatgcat actcggaata ccaagtgtcc aggtatgtaa tgttcggctt tagtattgca 360
ggtgcaattg ttacatttgt actctggatt atgtattttg taagatccat tcagttgtac 420
agaaggacta agtcttggtg gtccttcaac cctgaaacta acgcaattct ttgcgttagt 480
gcattaggaa gaagctatgt gcttcctctc gaaggtgtgc caactggtgt cactctaact 540
ttgctttcag ggaatttgta cgctgaaggg ttcaaaattg caggtggtat gaccatcgac 600
aatttgccaa aatacgtaat ggttgcatta cccagcagga ctattgttta cacacttgtt 660
ggcaagaagt tgaaagcaag tagtgcgact ggatgggctt actatgtaaa atctaaagct 720
ggtgattact caacagaggc aagaactgat aatttgagtg agcaagaaaa attattacat 780
atggtataa 789

<210> 44
<211> 789
<212> DNA
<213> Canine coronavirus

<400> 44
atgaagaaaa ttttgttttt actagcgtgt gcaattgcat gcgtctatgg agaacgctat 60
tgtgccatga ctgaaagttc tacgtcatgt cgtaatagca cggctggcaa ctgtgcttca 120
tgcttcgaaa caggtgatct tatttggcat cttgcaaact ggaacttcag ctggtctgta 180
atattgatca tttttataac agtgttacaa tatggaagac ctcaatttag ctggttcgtg 240
tgtggcatta aaatgcttat tatgtggctg ttatggccca ttgttttagc tcttacgatt 300
tttaatgcat acctggaata ccgagtttcc agatatgtaa tgttcggctt tagtgttgca 360
ggtgcaactg ttacatttat actttggatt atgtattttg ttagatccat tcagttatac 420
agaaggacta agtcttggtg gtctttcaac cctgaaacta gcgcaattct ttgcgttagt 480
gcgttaggaa gaagctatgt gcttcctctt gaaggtgtgc caactggtgt cactctaaca 540
ttgctttcag ggaatttgtg tgctgaaggg ttcaaaattg caggtggtat gaacatcgac 600
aatttaccaa aatatgtaat ggttgcatta cctgtcagaa ccatagtcta cacacttgtt 660
ggcaagaaat tgaaagcaag tagtgcaaca ggatgggctt actatgtaaa gtctaaagct 720
ggtgattact caacagatgc acgaactgat aatttgagtg agcatgaaaa attattacat 780
atggtataa 789

<210> 45
<211> 870
<212> DNA
<213> Feline coronavirus

<400> 45
atgcatatga tgcctataag acctttatgc aaaccaaggc atataatccc gacgaagcat 60
ttttggtttg aactaaacaa aatgaagtac attttgctaa tactcgcgtg cataattgca 120
tgcgtttatg gtgaacgcta ctgtgccatg caagacagtg gcttgcagtg tattaatggc 180
acaaattcaa gatgtcaaac ctgctttgaa cgtggtgatc ttatttggca tcttgctaac 240
tggaacttca gctggtctgt aatattgatt gtttttataa cagtgttaca atatggcaga 300
ccacaattta gctggctcgt ttatggcatt aaaatgctga tcatgtggct attatggcct 360
attgttctag cgcttacgat ttttaatgca tactctgagt accaagtttc cagatatgta 420
atgttcggct ttagtgttgc aggtgcagtt gtaacgtttg cactttggat gatgtatttt 480
gtgagatctg ttcagctata tagaagaacc aaatcatggt ggtcttttaa tcctgagact 540
aatgcaattc tttgtgttaa tgcattgggt agaagttatg tgcttccctt agatggtact 600
cctacaggtg ttacccttac tctactttca ggaaatctat atgctgaagg tttcaaaatg 660
gctggtggtt taaccatcga gcatttgcct aaatacgtca tgattgctac acctagtaga 720
accatcgttt atacattagt tggaaaacaa ttaaaagcaa ctactgccac aggatgggct 780
tactacgtaa aatctaaagc tggtgattac tcaacagaag cacgtactga caatttgagt 840
gaacatgaaa aattattaca tatggtgtaa 870

<210> 46
<211> 678
<212> DNA
<213> Human coronavirus OC43

<400> 46
atgtcaaatg acaattgtac gggtgacatt gtcacccatt tgaagaattg gaattttggt 60
tggaatgtta ttctaaccat attcattgtt attcttcagt ttggacacta taaatactcc 120
agattgtttt atggtttgaa gatgcttgta ctgtggcttc tttggccact cgtacttgct 180
ttgtcaatct ttgacacctg ggctaattgg gattctaatt gggcctttgt tgcatttagc 240
ttttttatgg ccgtatcaac actcgttatg tgggtgatgt acttcgcaaa cagtttcaga 300
cttttccgac gtgctcgaac tttttgggca tggaatcctg aggttaatgc aatcactgtc 360
acaaccgtgt tgggacagac atactatcaa cccattcaac aagctccaac aggcattact 420
gtgaccttgc tgagcggcgt gctttacgtt gacggacata gattggcttc aggtgttcag 480
gttcataacc tacctgaata catgacagtt gccgtgccga gcactactat aatttatagt 540
agagtcggaa ggtccgtaaa ttcacaaaat agcacaggct gggttttcta cgtacgagta 600
aaacacggtg atttttctgc agtgagctct cccatgagca acatgacaga aaacgaaaga 660
ttgcttcatt ttttctaa 678

<210> 47
<211> 681
<212> DNA
<213> Porcine epidemic diarrhea virus

<400> 47
atgtctaacg gttctattcc cgttgatgag gtgattgaac accttagaaa ctggaatttc 60
acatggaata tcatactgac gatactactt gtagtgcttc agtatggcca ttacaagtac 120
tctgtgttct tgtatggtgt caagatggct attctatgga tactttggcc tcttgtgttg 180
gcactgtcac tttttgacgc atgggctagc ttccaggtca actgggtctt tttcgctttc 240
agcatcctta tggcttgcat cactcttatg ctgtggataa tgtattttgt caatagcatt 300
cggttgtggc gcaggacaca ttcttggtgg tctttcaatc ctgaaactga cgcgcttctc 360
actacttctg tgatgggccg acaggtctgc attccagtgc ttggagcacc aactggtgta 420
acgctaacac tccttagtgg tacattgctt gtagagggct ataaggttgc tactggcgta 480
caggtaagtc aattacctaa tttcgtcaca gtcgccaagg ccactacaac aattgtctac 540
ggacgtgttg gtcgttcagt caatgcttca tctggcactg gttgggcttt ctatgtccgg 600
tcaaaacacg gcgactattc agctgtgagt aatccgagtg cggttctcac agatagtgag 660
aaagtgcttc atttagtcta a 681

<210> 48
<211> 693
<212> DNA
<213> Bovine coronavirus

<400> 48
atgagtagtg taactacacc agcaccagtt tacacctgga ctgctgatga agctattaaa 60
ttcctaaagg aatggaactt ttctttgggt attatactac tttttattac aatcatattg 120
caatttggat atacaagtcg cagtatgttt gtttatgtta ttaagatgat cattttgtgg 180
cttatgtggc cccttactat catcttaact attttcaatt gcgtgtatgc gttgaataat 240
gtgtatcttg gcttttctat agttttcact atagtggcca ttatcatgtg gattgtgtat 300
tttgtgaata gtatcaggtt gtttattaga actggaagtt ggtggagttt caacccagaa 360
acaaacaact tgatgtgtat agatatgaag ggaaggatgt atgttaggcc gataattgag 420
gactaccata cccttacggt cacaataata cgtggtcatc tttacatgca aggtataaaa 480
ctaggtactg gctattcttt gtcagatttg ccagcttatg tgactgttgc taaggtctca 540
cacctgctca cgtataagcg tggttttctt gacaagatag gcgatactag tggttttgct 600
gtttatgtta agtccaaagt cggtaattac cgactgccat caacccaaaa gggttctggc 660
atggacaccg cattgttgag aaataatatc taa 693

<210> 49
<211> 693
<212> DNA
<213> Porcine hemagglutinating encephalomyelitis coronavirus

<400> 49
atgagtagtc caactacacc agtaccagtt attagctgga ctgctgatga agctattaaa 60
ttcctaaagg aatggaattt ttctttgggt ataatagtac tctttatcac aatcatactt 120
caatttggat atacaagtcg cagtatgttt gtttatgtta ttaagatggt tattctgtgg 180
ctcatgtggc ctcttactat aattttaact atcttcaact gcgtatacgc gttgaataat 240
gtgtaccttg gcttctctat agtttttact atagtggcca ttattatgtg ggttgtttat 300
tttgtgaata gtatcaggtt gtttattaga actggaagtt ggtggagttt caacccagaa 360
acaaacaact tgatgtgtat agatatgaag ggaagaatgt atgttaggcc gattattgag 420
gactaccaca cccttactgc cacaataata cgtggccacc tctacatcca aggtataaaa 480
ctaggtactg gctattcttt gtcagatttg cctgcttatg tgaccgttgc taaggttaca 540
cacctgtgca catataagcg tggttttctt gataggatag gcgatactag tggttttgct 600

gtttatgtta agtccaaagt cggtaattat cgattgcctt caacccataa gggctcaggc 660
atggacaccg cattgttgag aaataatatc taa 693

<210> 50
<211> 687
<212> DNA
<213> Murine hepatitis virus

<400> 50
atgactagta ccactcaggc tccacagcct gtttatcagt ggacggctga tgaggcaatt 60
cgattcctta aggaatggaa tttctctctc ggcattatac tacttttgt tactatcata 120
ctacagttcg gttacacgag ccgtagcatg tttgtttatg ttgtgaaaat gatacttttg 180
tggcttatgt ggccactaac tattgttttg tgtattttta actgcgtcta tgcgctaaat 240
aatgtgtatc ttggattttc tatagtgttt actatagtgt ccattataat gtggattatg 300
tattttgtta atagcatcag gttgtttatc aggactggca gctggtggag cttcaacccc 360
gaaacaaaca acctaatgtg tatagatatg aaaggtactg tgtatgttag acccattata 420
gaggattacc atacactaac agccactatc attcgtggtc acctctatat gcaaggtgtt 480
aagctaggca ctggcttctc tttgtctgat ttgcctgctt atgttacagt tgctaaggtg 540
tctcaccttt gcacttataa gcgcgcattc ttagacaagg tagacggtgt tagcggtttt 600
gctgtttatg tgaagtccaa ggtcggaaat taccgactgc cctcaaataa accgagtggc 660
atggacaccg cattgttgag aatctaa 687

<210> 51
<211> 687
<212> DNA
<213> Rat coronavirus

<400> 51
atgagtagta ccactccagc cccccagact gtctatcaat ggacggccga tgtggcagtt 60
cgattcctta aggaatggaa cttcttgttg ggcattatac tactctttat tactatcata 120
ctacagttcg gttacacgag ccgtagcatg tttatatatg ttgtgaaaat gataatcttg 180
tggttaatgt ggccactgac tattgttttg tgtattttta attgcgtgta tgcgctaaat 240
aatgtgtatc ttggattttc tatagtgttt actatagtgt ccattgtaat gtggattatg 300
tattttgtta atagcataag gttgtttatc aggactggta gctggtggag cttcaaccct 360
gaaacaaaca acctaatgtg tatagatgtg aaaggtactg tgtatgttag acccattatt 420
gaagattacc atacactaac agccacaaat gtacgtggcc acctttatat gcaaggtgtt 480
aagctaggca ctggcttctc tttgtctgat ttgcccgctt atgttacagt tgctaaggtg 540
tcgcaccttt gcacttataa gcgcgcattt ttagacaagg ttgacggtgt tagcggtttt 600
gctgtttatg tgaagtccaa ggtcggtaat taccgactgc cctcaaataa accgagtggc 660
gcggacaccg cattgttgag aatctaa 687

<210> 52
<211> 675
<212> DNA
<213> SARS coronavirus urbani

<400> 52
ttgcttatca tggcagacaa cggtactatt accgttgagg agcttaaaca actcctggaa 60
caatggaacc tagtaatagg tttcctattc ctagcctgga ttatgttact acaatttgcc 120
tattctaatc ggaacaggtt tttgtacata ataaagcttg ttttcctctg gctcttgtgg 180
ccagtaacac ttgcttgttt tgtgcttgct gctgtctaca gaattaattg ggtgactggc 240
gggattgcga ttgcaatggc ttgtattgta ggcttgatgt ggcttagcta cttcgttgct 300
tccttcaggc tgtttgctcg tacccgctca atgtggtcat tcaacccaga aacaaacatt 360
cttctcaatg tgcctctccg ggggacaatt gtgaccagac cgctcatgga aagtgaactt 420
gtcattggtg ctgtgatcat tcgtggtcac ttgcgaatgg ccggacaccc cctagggcgc 480

tgtgacatta aggacctgcc aaaagagatc actgtggcta catcacgaac gctttcttat 540
tacaaattag gagcgtcgca gcgtgtaggc actgattcag gttttgctgc atacaaccgc 600
taccgtattg gaaactataa attaaataca gaccacgccg gtagcaacga caatattgct 660
ttgctagtac agtaa 675

<210> 53
<211> 675
<212> DNA
<213> SARS coronavirus Tor2

<400> 53
ttgcttatca tggcagacaa cggtactatt accgttgagg agcttaaaca actcctggaa 60
caatggaacc tagtaatagg tttcctattc ctagcctgga ttatgttact acaatttgcc 120
tattctaatc ggaacaggtt tttgtacata ataaagcttg ttttcctctg gctcttgtgg 180
ccagtaacac ttgcttgttt tgtgcttgct gctgtctaca gaattaattg ggtgactggc 240
gggattgcga ttgcaatggc ttgtattgta ggcttgatgt ggcttagcta cttcgttgct 300
tccttcaggc tgtttgctcg tacccgctca atgtggtcat tcaacccaga aacaaacatt 360
cttctcaatg tgcctctccg ggggacaatt gtgaccagac cgctcatgga aagtgaactt 420
gtcattggtg ctgtgatcat tcgtggtcac ttgcgaatgg ccggacactc cctagggcgc 480
tgtgacatta aggacctgcc aaaagagatc actgtggcta catcacgaac gctttcttat 540
tacaaattag gagcgtcgca gcgtgtaggc actgattcag gttttgctgc atacaaccgc 600
taccgtattg gaaactataa attaaataca gaccacgccg gtagcaacga caatattgct 660
ttgctagtac agtaa 675

<210> 54
<211> 678
<212> DNA
<213> Infectious bronchitis virus

<400> 54
atgcccaacg agacaaattg tactcttgac tttgaacagt cagttcagct ttttaaagag 60
tataatttat ttataactgc attcttgttg ttcttaacca taatacttca gtatggctat 120
gcaacaagaa gtaaggttat ttatacactg aaaatgatag tgttatggtg cttttggccc 180
cttaacattg cagtaggtgt aatttcatgt acatacccac caaacacagg aggtcttgtc 240
gcagcgataa tacttacagt gtttgcgtgt ctgtcttttg taggttattg gatccagagt 300
attagactct ttaagcggtg taggtcatgg tggtcattta atccagaatc taatgccgta 360
ggttcaatac tcctaactaa tggtcaacaa tgtaattttg ctatagagag tgtgccaatg 420
gtgctttctc caattataaa gaatggtgtt ctttattgtg agggtcagtg gcttgctaag 480
tgtgaaccag accacttgcc taaagatata tttgtttgta caccggatag acgtaatatc 540
taccgtatgg tgcagaaata tactggtgac caaagcggaa ataagaaaag gtttgctacg 600
tttgtctatg caaagcagtc agtagatact ggcgagctag aaagtgtagc aacaggagga 660
agtagtcttt acacataa 678

<210> 55
<211> 38
<212> DNA
<213> SARS coronavirus GD69

<400> 55
cctccgaccc aattaattct gtagacagca gccggagg 38

<210> 56
<211> 107
<212> DNA
<213> SARS coronavirus GD69

<400> 56
cttgttttcc tctggctctt gtggccagta acacttgctt gttttgtgct tgctgctgtc 60
tacagaatta attgggtgac tggcgggatt gcgattgcaa tggcttg 107

<210> 57
<211> 21
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic primer

<400> 57
cttgttttcc tctggctctt g 21

<210> 58
<211> 21
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic primer

<400> 58
caagccattg caatcgcaat c 21

<210> 59
<211> 14
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic primer

<400> 59
aagcaacgaa gtag 14

<210> 60
<211> 40
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic primer

<400> 60
gcctccgacc caattaattc tgtagacagc agccggaggc 40

<210> 61
<211> 40
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic primer

<400> 61
taatacgact cactataggc ttgttttcct ctggctcttg 40

<210> 62
<211> 35
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic primer

<400> 62
aagcaacgaa gtagcaagcc attgcaatcg caatc 35

<210> 63
<211> 1149
<212> DNA
<213> Transmissible gastroenteritis virus

<400> 63
atggccaacc agggacaacg tgtcagttgg ggagatgaat ctaccaaaac acgtggtcgt 60
tccaattccc gtggtcggaa gaataataac atacctcttt cattcttcaa ccccataacc 120
ctccaacaag gttcaaaatt ttggaactta tgtccgagag actttgtacc caaaggaata 180
ggtaacaggg atcaacagat tggttattgg aatagacaaa ctcgctatcg catggtgaag 240
ggccaacgta aagagcttcc tgaaaggtgg ttcttctact acttaggtac tggacctcat 300
gcagatgcca aatttaaaga taaattagat ggagttgtct gggttgccaa ggatggtgcc 360
atgaacaaac caaccacgct tggtagtcgt ggtgctaata atgaatccaa agctttgaaa 420
ttcgatggta aagtgccagg cgaatttcaa cttgaagtta atcaatcaag agacaattca 480
aggtcacgct ctcaatctag atctcggtct agaaatagat ctcaatctag aggcaggcaa 540
caattcaata acaagaagga tgacagtgta gaacaagctg ttcttgccgc acttaaaaag 600
ttaggtgttg acacagaaaa acaacagcaa cgctctcgtt ctaaatctaa agaacgtagt 660
aactctaaga caagagatac tacacctaag aatgaaaaca aacacacctg gaagagaact 720
gcaggtaaag gtgatgtgac aagattttat ggagctagaa gcagttcagc caattttggt 780
gacactgacc tcgttgccaa tgggagcagt gccaagcatt acccacaact ggctgaatgt 840
gttccatctg tgtctagcat tctgtttgga agctattgga cttcaaagga agatggcgac 900
cagatagaag tcacgttcac acacaaatac cacttgccaa aggatgatcc taagactgga 960
caattccttc agcagattaa tgcctatgct cgtccatcag aagtggcaaa agaacagaga 1020
aaaagaaaat ctcgttctaa atctgcagaa aggtcagagc aagatgtggt acctgatgca 1080
ttaatagaaa attatacaga tgtgtttgat gacacacagg ttgagataat tgatgaggta 1140
acgaactaa 1149

<210> 64
<211> 1149
<212> DNA
<213> Porcine respiratory coronavirus

<400> 64
atggccaacc agggacaacg tgtcagttgg ggggatgaat ccaccaaaat acgtggtcgc 60
tccaattccc gtggtcggaa gattaataac atacctcttt cattcttcaa ccccataacc 120
ctccagcaag gtgcaaaatt ttggaactca tgtccgagag attttgtacc caaaggaata 180
ggtaataggg atcaacagat tggttattgg aatagacaaa ctcgctatcg catggtgaag 240
ggccaacgta aagagcttcc tgaaaggtgg ttcttttact acttaggcac tggacctcat 300
gcagatgcca aatttaaaga taaattagat ggagttgtct gggttgccaa ggatggtgcc 360
atgaacaaac caaccacgct tggtagtcgt ggtgctaata atgaatccaa agctttgaaa 420
ttcgatggta aagtgccagg cgaatttcaa cttgaagtta accagtctag ggacaactca 480
aggtcacgct ctcaatctag atcgcggtct agaaacagat ctcaatctag aggtaggcaa 540
caatccaata acaagaagga tgacagtgta gaacaagctg ttcttgccgc acttaaaaag 600
ttaggtgttt acacagaaaa acaacagcaa cgctctcgtt ctaaatctaa agaacgtagt 660
aactctaaaa caagagatac tacgcctaag aatgaaaaca aacacacctg gaagagaact 720
gcaggtaaag gtgatgtgac aagattttat ggagctagaa gcagctcagc caattttggt 780
gacagtgacc tcgttgccaa tgggagcagt gccaagcatt acccacaatt ggctgaatgt 840
gttccatctg tgtctagcat tttgtttgga agctattgga cttcaaagga agatggcgac 900
cagatagaag tcacgttcac acacaaatac cacttgccaa aggatcatcc taaaactgaa 960
caattccttc agcagattaa tgcctatgct agcccatcag aattggcaaa agaacagaga 1020
aaaagaaagt ctcgttctaa atctgcagaa aggtcagagc aagaggtggt acctgattca 1080
ttaatagaaa actatacaga tgtgtttgat gacacacagg ttgagatgat tgacgaggta 1140
acgaactaa 1149

<210> 65
<211> 1146
<212> DNA
<213> Canine coronavirus

<400> 65
atggcctctc agggacaacg tgtcagttgg ggagatgaat ccaccaagag acgcggtcgt 60
tctaattctc gtggccggaa gaataatgat atacctcttt cattcttcaa ccccattacc 120
ctcgagcaag gatcaaagtt ttgggactta tgtccgagag actttgtacc caaaggaata 180
ggtaataagg atcaacaaat tggttattgg aacaggcaaa cccgttatcg catggtgaag 240
ggtcgacgta aaaatcttcc tgaaaagtgg ttcttctact atttaggaac tggacctcat 300
gctgatgcca aatttaagca aaaattagat ggagttgtct gggttgctag gggagattcc 360
atgactaagc caacaactct tggtactcgt ggcactaata atgaatcaaa ggctttgaaa 420
ttcgatgtca aagtaccatc agaatttcac cttgaagtga accaattaag ggacaattca 480
aggtctaggt ctcaatctag atctcagtcc agaaataggt ctcaatctag aggaaggcaa 540
ctatccaata ataagaagga tgacaatgtt gaacaagctg ttcttgctgc actcaaaaag 600
ttaggtgttg acacagaaaa acaacaaaga tctcgttcca aatctaagga acgtagcagc 660
tctaagacaa gagatactac acctaagaat gaaaacaaac acacctggaa gagaactgca 720
ggtaaaggtg atgtgacaaa attttatgga gctagaagta gttcagccaa ttttggtgac 780
agcgatcttg ttgccaatgg gaacggtgcc aagcattacc cacaactggc tgaatgtgtt 840
ccatctgtat ctagcattct gtttggaagc cattggactg ctaaggaaga tggtgaccag 900
attgaagtca cattcacaca caaataccac ttgccaaagg atgatcctaa gactggacaa 960
ttccttcagc agattaatgc atacgcccgt ccatcagagg tggctaaaga acagagacaa 1020
cgcaaagctc gttctaaatc tgtagaaagg gtagagcaag aggttgtacc tgatgcatta 1080
acagaaaatt acacagatgt gtttgatgac acacaggttg agattattga tgaggtaacg 1140
aactaa 1146

<210> 66
<211> 1134
<212> DNA
<213> Feline coronavirus

<400> 66
atggccacac agggacaacg cgtcaactgg ggagatgaac cttccaaaag acgtggtcgt 60
tctaactctc gtggtcggaa gaataatgat ataccttttgt cattctacaa ccccattacc 120
ctcgaacaag gatctaaatt ttggaattta tgtccgagag accttgttcc caaaggaata 180
ggtaataagg atcaacaaat tggttattgg aatagacaga ttcgttatcg tattgtaaaa 240
ggccagcgta aggaactcgc tgagaggtgg ttcttttact tcttaggtac aggacctcat 300
gctgatgcta aattcaaaga caagattgat ggagtcttct gggttgcaag ggatggtgcc 360
atgaacaagc ccacaacgct tggcactcgt ggaaccaata acgaatccaa accactgaga 420
tttgatggta agataccgcc acagtttcag cttgaagtga accgttctag gaacaattca 480
aggtctggtt ctcagtctag atctgtttca agaaacagat ctcaatctag aggaagacac 540
cattccaata accagaataa taatgttgag gatacaattg tagccgtgct tgaaaaatta 600
ggtgttactg acaaacaaag gtcacgttct aaacctagag aacgtagtga ttccaaacct 660
agggacacaa cacctaagaa tgccaacaaa cacacctgga agaaaactgc aggcaaggga 720
gatgtgacaa ctttctatgg tgctagaagt agttcagcta actttggtga tagtgatctc 780
gttgccaatg gtaacgctgc caaatgctac cctcagatag ctgaatgtgt tccatcagtg 840
tctagcataa tctttggcag tcaatggtct gctgaagaag ctggtgatca agtgaaagtc 900
acgctcactc acacctacta cctgccaaag gatgatgcca aaactagtca attcctagaa 960
cagattgacg cttacaagcg accttctgaa gtggctaagg atcagaggca aagaagatcc 1020
cgttctaagt ctgctgataa gaagcctgag gagttgtctg taactcttgt ggaggcatac 1080
acagatgtgt ttgatgacac acaggttgag atgattgatg aggttacgaa ctaa 1134

<210> 67
<211> 1269
<212> DNA
<213> SARS coronavirus urbani

<400> 67
atgtctgata atggacccca atcaaaccaa cgtagtgccc cccgcattac atttggtgga 60
cccacagatt caactgacaa taaccagaat ggaggacgca atggggcaag gccaaaacag 120
cgccgacccc aaggtttacc caataatact gcgtcttggt tcacagctct cactcagcat 180
ggcaaggagg aacttagatt ccctcgaggc cagggcgttc caatcaacac caatagtggt 240
ccagatgacc aaattggcta ctaccgaaga gctacccgac gagttcgtgg tggtgacggc 300
aaaatgaaag agctcagccc cagatggtac ttctattacc taggaactgg cccagaagct 360
tcacttccct acggcgctaa caaagaaggc atcgtatggg ttgcaactga gggagccttg 420
aatacaccca aagaccacat tggcacccgc aatcctaata acaatgctgc caccgtgcta 480
caacttcctc aaggaacaac attgccaaaa ggcttctacg cagagggaag cagaggcggc 540
agtcaagcct cttctcgctc ctcatcacgt agtcgcggta attcaagaaa ttcaactcct 600
ggcagcagta ggggaaattc tcctgctcga atggctagcg gaggtggtga aactgccctc 660
gcgctattgc tgctagacag attgaaccag cttgagagca aagtttctgg taaaggccaa 720
caacaacaag gccaaactgt cactaagaaa tctgctgctg aggcatctaa aaagcctcgc 780
caaaaacgta ctgccacaaa acagtacaac gtcactcaag catttgggag acgtggtcca 840
gaacaaaccc aaggaaattt cggggaccaa gacctaatca gacaaggaac tgattacaaa 900
cattggccgc aaattgcaca atttgctcca agtgcctctg cattctttgg aatgtcacgc 960
attggcatgg aagtcacacc ttcgggaaca tggctgactt atcatggagc cattaaattg 1020
gatgacaaag atccacaatt caaagacaac gtcatactgc tgaacaagca cattgacgca 1080
tacaaaacat tcccaccaac agagcctaaa aaggacaaaa agaaaaagac tgatgaagct 1140
cagcctttgc cgcagagaca aaagaagcag cccactgtga ctcttcttcc tgcggctgac 1200
atggatgatt tctccagaca acttcaaaat tccatgagtg gagcttctgc tgattcaact 1260
caggcataa 1269

<210> 68
<211> 1269
<212> DNA
<213> SARS coronavirus Tor2

<400> 68
atgtctgata atggacccca atcaaaccaa cgtagtgccc cccgcattac atttggtgga 60
cccacagatt caactgacaa taaccagaat ggaggacgca atggggcaag gccaaaacag 120
cgccgacccc aaggtttacc caataatact gcgtcttggt tcacagctct cactcagcat 180
ggcaaggagg aacttagatt ccctcgaggc cagggcgttc caatcaacac caatagtggt 240
ccagatgacc aaattggcta ctaccgaaga gctacccgac gagttcgtgg tggtgacggc 300
aaaatgaaag agctcagccc cagatggtac ttctattacc taggaactgg cccagaagct 360
tcacttccct acggcgctaa caaagaaggc atcgtatggg ttgcaactga gggagccttg 420
aatacaccca aagaccacat tggcacccgc aatcctaata acaatgctgc caccgtgcta 480
caacttcctc aaggaacaac attgccaaaa ggcttctacg cagagggaag cagaggcggc 540
agtcaagcct cttctcgctc ctcatcacgt agtcgcggta attcaagaaa ttcaactcct 600
ggcagcagta ggggaaattc tcctgctcga atggctagcg gaggtggtga aactgccctc 660
gcgctattgc tgctagacag attgaaccag cttgagagca aagtttctgg taaaggccaa 720
caacaacaag gccaaactgt cactaagaaa tctgctgctg aggcatctaa aaagcctcgc 780
caaaaacgta ctgccacaaa acagtacaac gtcactcaag catttgggag acgtggtcca 840
gaacaaaccc aaggaaattt cggggaccaa gacctaatca gacaaggaac tgattacaaa 900
cattggccgc aaattgcaca atttgctcca agtgcctctg cattctttgg aatgtcacgc 960
attggcatgg aagtcacacc ttcgggaaca tggctgactt atcatggagc cattaaattg 1020
gatgacaaag atccacaatt caaagacaac gtcatactgc tgaacaagca cattgacgca 1080
tacaaaacat tcccaccaac agagcctaaa aaggacaaaa agaaaaagac tgatgaagct 1140
cagcctttgc cgcagagaca aaagaagcag cccactgtga ctcttcttcc tgcggctgac 1200
atggatgatt tctccagaca acttcaaaat tccatgagtg gagcttctgc tgattcaact 1260
caggcataa 1269

<210> 69
<211> 1347
<212> DNA
<213> Bovine coronavirus

<400> 69
atgtctttta ctcctggtaa gcaatccagt agtagagcgt cctttggaaa tcgttctggt 60
aatggcatcc ttaagtgggc cgatcagtcc gaccaatcta gaaatgttca aaccaggggt 120
agaagagctc aacccaagca aactgctact tctcagctac catcaggagg gaatgttgta 180
ccctactatt cttggttctc tggaattact cagtttcaaa aaggaaagga gtttgaattt 240
gcagagggac aaggtgtgcc tattgcacca ggagtcccag ctactgaagc taaggggtac 300
tggtacagac acaacagacg ttcttttaaa acagccgatg gcaaccagcg tcaactgctg 360
ccacgatggt atttttacta tcttggaaca ggaccgcatg ccaaagacca gtatggcacc 420
gatattgacg gagtcttctg ggtcgctagt aaccaggctg atgtcaatac cccggctgac 480
attctcgatc gggacccaag tagcgatgag gctattccga ctaggtttcc gcctggcacg 540
gtactccctc agggttacta tattgaaggc tcaggaaggt ctgctcctaa ttccagatct 600
acttcacgcg catccagtag agcctctagt gcaggatcgc gtagtagagc caattctggc 660
aacagaaccc ctacctctgg tgtaacacct gatatggctg atcaaattgc tagtcttgtt 720
ctggcaaaac ttggcaagga tgccactaag ccacagcaag taactaagca gactgccaaa 780
gaaatcagac agaaaatttt gaataagccc cgccagaaga ggagccccaa taaacaatgc 840
actgttcagc agtgttttgg gaagagaggc cccaatcaga attttggtgg tggagaaatg 900
ttaaaacttg gaactagtga cccacagttc cccattcttg cagaactcgc acccacagct 960
ggtgcgtttt tctttggatc aagattagag ttggccaaag tgcaggaattt gtctgggaat 1020
cttgacgagc cccagaagga tgtttatgaa ttgcgctata atggtgcaat tagatttgac 1080
agtacacttt caggttttga gaccataatg aaggtgttga atgagaattt gaatgcatat 1140
caacaacaag atggtatgat gaatatgagt ccaaaaccac agcgtcagcg tggtcagaag 1200
aatggacaag gagaaaatga taatataagt gttgcagcgc ctaaaagccg tgtgcagcaa 1260
aataagagta gagagttgac tgcagaggac atcagccttc ttaagaagat ggatgagccc 1320

tatactgaag acacctcaga aatataa 1347

<210> 70
<211> 1350
<212> DNA
<213> Porcine hemagglutinating encephalomyelitis coronavirus

<400> 70
atgtctttca ctcctggcaa gcagtccagc agtagagcgt cctctggaaa tcgttctggt 60
aatggcatcc ttaagtgggc cgatcagtcc gaccagtcta gaaatgttca aaccaggggt 120
agaagagttc aatccaagca aactgctact tctcagcaac catcaggagg gactgttgta 180
ccctactatt cttggttctc tggaattact cagtttcaaa agggaaagga gtttgaattt 240
gcagagggac aaggtgtgcc tattgcacca ggagtcccat ctactgaagc taaggggtac 300
tggtacagac acaacagacg ttcttttaaa acagccgacg gcaatcagcg tcaactgctg 360
ccacgatggt acttttacta cctgggaaca ggaccgcatg ccaaagacca gtacggcacc 420
gacattgacg gagtcttctg ggtcgctagt aaccaggctg atattaatac cccggctgac 480
attgtcgatc gggatccaag tagcgatgag gctattccga ctaggtttcc gcctggcacg 540
gtactccctc aaggttacta tattgaaggc tcaggaaggt ctgctcctaa ttccagatct 600
acttcgcgtg cacccaatag agcccctagt gcaggatcgc gtagtagagc caattctggc 660
aatagaacct ctacccctgg tgtaacacct gacatggctg atcaaattgc tagtcttgtt 720
ctggcaaaac ttggcaagga tgccactaag cctcagcaag taactaagca gactgccaaa 780
gaggtcagac agaaaatctt gaataagccc cgccagaaga ggagccccaa caaacaatgc 840
actgttcagc agtgttttgg gaagagaggc cccaatcaga attttggtgg tggagaaatg 900
ttaaaacttg gaactagtga cccacagttc cccattcttg cagaactcgc acccacagct 960
ggtgcgtttt tctttggatc aagattagag ttggccaaag tgcagaattt gtctgggaat 1020
cctgacgagc cccagaagga tgtttatgaa ttgcgctata atggcgcgat tagatttgac 1080
agcacactct caggttttga aaccattatg aaggtgctta accagaattt gaatgcctat 1140
caacatcagg aagatgggat gatgaatatt agtcctaaac cacagcggca gcgtggtcag 1200
aagaatggac aagtagaaaa tgataatgta agtgttgcag cgcctaaaag ccgtgtgcag 1260
caaaataaga gtagagagtt gacagcagag gacatcagcc ttcttaagaa gatggatgag 1320
ccctatactg aagatacctc agaaatataa 1350

<210> 71
<211> 1356
<212> DNA
<213> Murine hepatitis virus

<400> 71
atgtcttttg ttcctgggca agaaaatgcc ggtagcagaa gctcctctgg aaaccgcgct 60
ggtaatggca tcctcaagaa gaccacttgg gctgaccaaa ccgagcgtgg aaatagaggc 120
agaaggaacc atcccaagca gactgcaact actcagccca atgccgggag tgtggttccc 180
cattactctt ggttttcggg catcacccag tttcaaaagg gaaaggagtt ccagtttgca 240
caaggacagg gagtgcctat tgccagtgga atccccgctt cagagcaaaa gggatattgg 300
tatagacaca accgacgttc ttttaaaaca cctgatggcc agcacaagca gctactgccc 360
agatggtatt tttactatct tggaacaggg ccccatgctg gcgcagagta tggcgacgat 420
atcgaaggag ttgtctgggt cgcaagccaa caggccgaca ctaagaccac tgccgatgtt 480
gttgaaaggg acccaagcag tcatgaggct attcctacta ggtttgcgcc cggcacggta 540
ttgcctcagg gcttttatgt agaaggctcg ggaaggtctg cacctgctag tcgatctggt 600
tcgcggtcac aatcccgtgg gccaaataat cgcgctagaa gcagttccaa ccagcgccag 660
cctgcctctg ctgtaaaacc tgacatggcc gaagaaattg ctgctcttgt tttggctaag 720
cttggtaaag atgccggcca gcccaagcag gtaactaagc aaagcgccaa agaagtcagg 780
cagaaaattt taactaagcc tcgtcaaaag aggactccaa acaagcagtg cccagtgcag 840
cagtgttttg ggaagagagg ccctaatcag aactttggag gctctgaaat gttaaaactt 900
ggaactagtg atccgcagtt ccccattctt gcagagttgg ctccaacacc tagtgccttc 960
ttctttggat ctaaattaga attggtcaaa aagaactctg gtggtgctga tgaacccacc 1020
aaagatgttt atgaattgca gtattcaggt gcaattagat ttgatagtac tctacccggt 1080

tttgagacta tcatgaaagt gttgactgag aatttgaatg cctaccagga ccaagctggt 1140
agtgtagatc tagtgagccc aaagcctcca agaagaggtc gtagacaggc tcaagaaaag 1200
aaagatgaag tagataatgt aagcgttgca aagcccaaaa gcttggtgca gcgaaatgta 1260
agtagagaat taacccccga ggatcgtagc ctgctggctc agatcctaga cgatggcgtt 1320
gtgccagatg ggttggaaga tgactctaat gtgtaa 1356

<210> 72
<211> 1365
<212> DNA
<213> Rat coronavirus

<400> 72
atgtcttttg ttcccggaca agaaaacgcc ggtagcagaa gctcctctgg aaaccgcgct 60
ggtaatggaa tcctcaagaa gaccacttgg gctgaccaaa ccgagcgcgg acaaaataat 120
ggaaatagag gcagaaggaa tcagcccaag cagactgcaa ctactcagcc caataccggg 180
agtgtggttc cccattactc ttggttttcg ggcattaccc aattccagaa gggaaaagag 240
ttccagtttg caggtggaca aggagtgcct attgccaatg gaatcccacc ttctgagcaa 300
aagggatatt ggtatagaca caaccgtcgt tcttttaaaa cacctgatgg gcagcagaag 360
caactactcc ccagatggta tttttactat cttgggacgg gcccccatgc tggagccagt 420
ttcggagaca gcattgaggg agtcttctgg gttgcaaata gtcaggcgga taccaacacc 480
tctgctgaca ttgttgaaag ggacccaagt agccatgagg ctattcctac taggtttgcg 540
cccggtacgg tattgcctca gggtttctat gttgaaggct cgggaaggtc tgcacctgct 600
agtcgatctg gttcgcggtc acaatcccgt gggccaaata atcgcgctag aagcagttcc 660
aaccagcgcc agcctgcctc tactgtaaaa cctgatatgg ccgaagaaat tgctgctctt 720
gttttggcta atctaggcaa agatgccgga cagcctaagc aagtaactaa gcaaagtgcc 780
aaagaagtca ggcagaaaat tttaaataag cctcgccaaa agaggactcc aaacaagcag 840
tgcccagtgc agcagtgttt tggaaagaga ggccccaatc agaattttgg aggccctgaa 900
atgttaaaac ttggaactag tgatccacag ttccccattc ttgcagagtt ggccccaaca 960
cctggtgcct tcttctttgg atctaaatta gaattggtca aaaagaattc tggtggcgtt 1020
gatgaaccca ccaaagatgt gtatgagctg caatattcag gtgcagtcag atttgatagt 1080
actctacctg gttttgagac tatcatgaaa gtgttgaatg agaatttgaa tgcctaccag 1140
aatcaagctg gtggtgcaga tgtagtgagc ccaaagcccc aaagaaagag agggacgaaa 1200
caaacggctc agaaagaaga attagatagt ataagcgttg caaagcccaa aagtgccgtg 1260
cagcgaaatg taagcagaga attaacccca gaggatagaa gcctgttggc gcagatccta 1320
gatgatggcg ttgtgcctga tgggttagat gactctaatg tgtaa 1365

<210> 73
<211> 1111
<212> DNA
<213> Human coronavirus OC43

<400> 73
atggctacag tcaaatgggc tgatgcatct gaaccacaac gtggtcgtca gggtagaata 60
ccttattctc tttatagccc tttgcttgtt gatagtgaac aaccttggaa ggtgatacct 120
cgtaatttgg tacccatcaa caagaaagac aaaaataagc ttataggcta ttggaatgtt 180
caaaaacgtt tcagaactag aaagggcaaa cgggtggatt tgtcacccaa gctgcatttt 240
tattatcttg gcacaggacc ccataaagat gcaaaattta gagagcgtgt tgaaggtgtc 300
gtctgggttg ctgttgatgg tgctaaaact gaacctacag gttacggtgt taggcgcaag 360
aattcagaac cagagatacc acacttcaat caaaagctcc caaatggtgt tactgttgtt 420
gaagaacctg actcccgtgc tccttcccgg tctcagtcga ggtcgcagag tcgcggtcgt 480
ggtgaatcca aacctcaatc tcggaatcct tcaagtgaca gaaaccataa cagtcaggat 540
gacatcatga aggcagttgc tgcggctctt aaatctttag gttttgacaa gcctcaggaa 600
aaagataaaa agtcagcgaa aacgggtact cctaagcctt ctcgtaatca gagtcctgct 660
tcttctcaaa cttctgccaa gagtcttgct cgttctcaga gttctgaaac aaaagaacaa 720
aagcatgaaa tgcaaaagcc acggtggaaa agacagccta atgatgatgt gacatctaat 780
gtcacacaat gttttggccc cagagacctt gaccacaact ttggaagtgc aggtgttgtg 840

gccaatggtg ttaaagctaa aggctatcca caatttgctg agcttgtgcc gtcaacagct 900
gctatgctgt ttgatagtca cattgtttcc aaagagtcag gcaacactgt ggtcttgact 960
ttcactacta gagtgactgt gcccaaagac catccacact tgggtaagtt tcttgaggag 1020
ttaaatgcat tcactagaga aatgcaacaa cactgaacca gtgcgtgatg aagtttctat 1080
tgaaactgac ataattgatg aagtaaacta a 1111

<210> 74
<211> 1326
<212> DNA
<213> Porcine epidemic diarrhea virus

<400> 74
atggcttctg tcagctttca ggatcgtggc cgcaaacggg tgccattatc tctctatgcc 60
cctcttaggg ttactaatga caagcccctt tctaaggtac ttgcaaacaa cgctgtaccc 120
actaacaagg ggaataagga ccagcaaatt gggtactgga atgagcaaat tcgctggcgc 180
atgcgccgtg gtgagcgaat tgaacaacct tccaattggc atttctacta cctcggaaca 240
ggacctcacg gcgacctccg ttataggact cgtactgagg gtgttttctg ggttgctaaa 300
gaaggcgcaa agactgaacc cactaatttg ggtgtcagaa aggcgtctga aaagccaatc 360
attccaaaat tctctcaaca gctccccagt gtagttgaga ttgttgaacc taacacacct 420
cctgcttcac gtgcaaattc gcgtagcagg agtcgtggca atggcaacaa taggtctaga 480
tctccaagta acaacagagg caataaccag tcccgtggta attcacagaa tcgtggaaat 540
aaccagggtc gtggagcttc tcagaacaga ggaggcaata ataataacaa taacaagtct 600
cgtaaccagt ccaataacag gaaccagtca aatgaccgtg gtggtgtaac atcacgcgat 660
gatctggtgg ctgctgtcaa ggatgcactt aaatctttgg gtattggaga aaatcctgac 720
aggcataagc aacagcagaa gcctaagcag gaaaagtctg acaacagcgg caaaaataca 780
cctaagaaga acaaatccag ggccacttcg aaggaacgtg acctcaaaga catcccagag 840
tggaggagaa ttcccaaggg cgaaaatagc gtagcagctt gcttcggacc cagaggggc 900
ttcaaaaact ttggagatgc ggaatttgtc gaaaaaggtg ttgatgcgtc aggctatgct 960
cagatcgcca gtttagcacc aaatgttgca gcattgctct ttggtggtaa tgtggctgtt 1020
cgtgagctag cggactctta cgagattaca tacaactata aaatgactgt gccaaagtca 1080
gatccaaatg ttgagcttct tgtttcacag gtggatgcat ttaaaactgg gaatgcaaaa 1140
ctccagagaa agaaggaaaa gaagaacaag cgtgaaacca cgctgcagca gcatgaagag 1200
gccatctacg atgatgtggg tgcgccatct gatgtgaccc atgccaatct ggaatgggac 1260
acagctgttg atggtggtga tacggccgtt gaaattatca acgagatctt cgatacagga 1320
aattaa 1326

<210> 75
<211> 1230
<212> DNA
<213> Infectious bronchitis virus

<400> 75
atggcaagcg gtaaagcagc tggaaaaaca gacgccccag cgccagtcat taaactagga 60
ggaccaaaac cacctaaagt cggttcttct ggaaatgcat cttggtttca agcaataaaa 120
gccaagaagt taaatacacc tccgcccaag tttgaaggta gcggtgttcc tgataacgaa 180
aacattaagc caagccagca acatggatac tggagacgcc aagccaggtt taagccaggc 240
aaaggtggaa gaaaaccagt cccagatgct tggtactttt actatactgg aacaggacct 300
gccgctgacc tgaactgggg tgatactcaa gatggtatag tgtgggttgc tgctaagggt 360
gctgatacta aatctagatc caatcagggt acaagagatc ctgataagtt tgaccaatac 420
ccactacgat tctcggatgg cggacctgat ggtaatttcc gttgggactt cattcccctg 480
aaccgtggta ggagtggaag atcaacagca gcttcatcag cagcagctag tagagcacca 540
tcacgtgaag gttcgcgtgg tcgtagaagt gattctggag atgaccttat tgctcgtgca 600
gcaaagataa tccaggatca gcagaaaaag ggctctcgca ttaccaaggc aaaggcagat 660
gaaatggctc atcgccggta ttgcaagcgc actatcccac ctaattatag ggttgatcaa 720
gtgtttggtc cccgtactaa aggtaaggag gggaattttg gtgatgacaa gatgaatgag 780
gaaggtatta aggatgggcg tgttacagca atgctcaacc tagtccctag cagccatgct 840

tgtctttttg gaagtagagt gacacccaaa cttcaactag atgggcttca cttgagattt 900
gaatttacta ctgtggtccc atgtgatgac ccgcagtttg ataattatgt gaaaatttgt 960
gatcagtgtg tcgatggtgt aggaacgcgt ccaaaagatg acgaaccaaa accaaagtca 1020
cgctcaagtt caagacctgc tacaagagga aattctccag cgccaagaca acagcgccca 1080
aagaaggaga aaaagctaaa gaagcaggat gatgaagcag ataaagcatt gacctcagat 1140
gaggagagga acaatgcaca gctggaattt tatgatgagc ccaaggtaat taactggggg 1200
gatgcagctc taggagagaa tgaactttga 1230

<210> 76
<211> 38
<212> DNA
<213> SARS coronavirus GD69

<400> 76
cctccgtacc atctggggct gagctctttc atcggagg 38

<210> 77
<211> 95
<212> DNA
<213> SARS coronavirus GD69

<400> 77
acgagttcgt ggtggtgacg gcaaaatgaa agagctcagc cccagatggt acttctatta 60
cctaggaact ggcccagaag cttcacttcc ctacg 95

<210> 78
<211> 19
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic primer

<400> 78
acgagttcgt ggtggtgac 19

<210> 79
<211> 19
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic primer

<400> 79
cgtagggaag tgaagcttc 19

<210> 80
<211> 14
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic primer

<400> 80
gccttctttg ttag 14

<210> 81
<211> 40
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic primer

<400> 81
gcctccgtac catctggggc tgagctcttt catcggaggc 40

<210> 82
<211> 38
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic primer

<400> 82
taatacgact cactatagga cgagttcgtg gtggtgac 38

<210> 83
<211> 33
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic primer

<400> 83
gccttctttg ttagcgtagg gaagtgaagc ttc 33

<210> 84
<211> 40
<212> DNA
<213> SARS coronavirus GD69

<400> 84
gcccacgtac catctggggc tgtagacagc agccgtgggc 40

<210> 85
<211> 98
<212> DNA
<213> SARS coronavirus GD69

<400> 85
ctctatgttt ataagggcta tcaacctata gatgctgctg tctacagccc cagatggtag 60
tattcttgct agtcacacta gccatcctta ctgcgctt 98

<210> 86
<211> 17
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic primer

<400> 86
tacacacctc agcgttg 17

<210> 87
<211> 16
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic primer

<400> 87
cacgaacgtg acgaat 16

<210> 88
<211> 26
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic primer

<220>
<221> modified_base
<222> (21)..(26)
<223> a, c, g, t, unknown, or other

<400> 88
gccggagctc tgcagaattc nnnnnn 26

<210> 89
<211> 20
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic primer

<400> 89
gccggagctc tgcagaattc 20

<210> 90
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic primer

<400> 90
gggttgggac tatcctaagt gtga 24

<210> 91
<211> 21
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic primer

<220>
<221> modified_base
<222> (11)
<223> i

<400> 91
taacacacaa ncccatcatc a 21

<210> 92
<211> 21
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic primer

<400> 92
ctaacatgct taggataatg g 21

<210> 93
<211> 21
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic primer

<400> 93
gcctctcttg ttcttgctcg c 21

<210> 94
<211> 21
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic primer

<400> 94
caggtaagcg taaaactcat c 21

INTERNATIONAL SEARCH REPORT

International Application No
PCT/US2004/026380

A. CLASSIFICATION OF SUBJECT MATTER
IPC 7 C12Q1/68

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC 7 C12Q

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)
EPO-Internal, BIOSIS, EMBASE, FSTA, WPI Data, PAJ

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	THIEL V ET AL: "Mechanisms and enzymes involved in SARS coronavirus genome expression" JOURNAL OF GENERAL VIROLOGY, SOCIETY FOR GENERAL MICROBIOLOGY, READING, GB, vol. 84, no. 9, 19 June 2003 (2003-06-19), pages 2305-2315, XP002301827 ISSN: 0022-1317 page 2207, left-hand column, last paragraph - page 2207, right-hand column, paragraph 1 ----- -/--	1-3,7, 13,16

[X] Further documents are listed in the continuation of box C.

[X] Patent family members are listed in annex.

° Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
26 November 2004	07/12/2004

Name and mailing address of the ISA	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL – 2280 HV Rijswijk Tel. (+31-70) 340-2040, Tx. 31 651 epo nl, Fax: (+31-70) 340-3016	Knudsen, H

INTERNATIONAL SEARCH REPORT

International Application No
PCT/US2004/026380

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	ADZHAR AZRI ET AL: "Universal oligonucleotides for the detection of infectious bronchitis virus by the polymerase chain reaction" AVIAN PATHOLOGY, vol. 25, no. 4, 1996, pages 817-836, XP009040626 ISSN: 0307-9457 page 831, paragraph 2 page 817	1
Y	WITTWER C T ET AL: "Real-time multiplex PCR assays" METHODS : A COMPANION TO METHODS IN ENZYMOLOGY, ACADEMIC PRESS INC., NEW YORK, NY, US, vol. 25, no. 4, December 2001 (2001-12), pages 430-442, XP002265718 ISSN: 1046-2023 abstract	1-26
Y	ZHOU D ET AL: "ONE-STEP DUPLEX RT-PCR ASSAY FOR DETECTION SARS ASSOCIATED CORONAVIRUS" BINGDUXUE ZAZHI - VIROLOGICA SINICA, KEXUE CHUBANSHE, BEIJING,, CN, vol. 18, no. 3, June 2003 (2003-06), pages 232-236, XP008033949 ISSN: 1000-3223 abstract	1-26
Y	ROTA P A ET AL: "Characterization of a novel coronavirus associated with severe acute respiratory syndrome" SCIENCE, AMERICAN ASSOCIATION FOR THE ADVANCEMENT OF SCIENCE,, US, vol. 300, no. 5624, 30 May 2003 (2003-05-30), pages 1394-1399, XP002269482 ISSN: 0036-8075 page 1394, right-hand column, paragraph 4	1-26
P,X	WU HO-SHENG ET AL: "Early detection of antibodies against various structural proteins of the SARS-associated coronavirus in SARS patients." JOURNAL OF BIOMEDICAL SCIENCE, vol. 11, no. 1, February 2004 (2004-02), pages 117-126, XP009040395 ISSN: 1021-7770 page 120, left-hand column, last paragraph; table 1 page 125, left-hand column, lines 7-9	1-3

-/--

INTERNATIONAL SEARCH REPORT

International Application No
PCT/US2004/026380

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
P,X	HUANG LI-RUNG ET AL: "Evaluation of antibody responses against SARS coronaviral nucleocapsid or spike proteins by immunoblotting or ELISA" JOURNAL OF MEDICAL VIROLOGY, vol. 73, no. 3, July 2004 (2004-07), pages 338-346, XP009040396 ISSN: 0146-6615 abstract	1-3
P,Y	EMERY S L ET AL: "REAL-TIME REVERSE TRANSCRIPTION-POLYMERASE CHAIN REACTION ASSAY FOR SARS-ASSOCIATED CORONAVIRUS" EMERGING INFECTIOUS DISEASES, EID, ATLANTA, GA, US, vol. 10, no. 2, February 2004 (2004-02), pages 311-316, XP008033948 ISSN: 1080-6040 table 1	1-26
P,Y	ZHAI J ET AL: "REAL-TIME POLYMERASE CHAIN REACTION FOR DETECTING SARS CORONAVIRUS, BEIJING, 2003" EMERGING INFECTIOUS DISEASES, EID, ATLANTA, GA, US, vol. 10, no. 2, February 2004 (2004-02), pages 300-303, XP008033953 ISSN: 1080-6040 page 300, right-hand column	1-26
E	WO 2004/094675 A (GEN PROBE INC ; GETMAN DAMON K (US); KACIAN DANIEL L (US); NELSON NORM) 4 November 2004 (2004-11-04) page 12	1-26

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US2004/026380

Box No. I Nucleotide and/or amino acid sequence(s) (Continuation of item 1.b of the first sheet)

1. With regard to any nucleotide and/or amino acid sequence disclosed in the international application and necessary to the claimed invention, the international search was carried out on the basis of:

   a. type of material

   [X] a sequence listing

   [ ] table(s) related to the sequence listing

   b. format of material

   [X] in written format

   [X] in computer readable form

   c. time of filing/furnishing

   [X] contained in the international application as filed

   [ ] filed together with the international application in computer readable form

   [X] furnished subsequently to this Authority for the purpose of search

2. [X] In addition, in the case that more than one version or copy of a sequence listing and/or table relating thereto has been filed or furnished, the required statements that the information in the subsequent or additional copies is identical to that in the application as filed or does not go beyond the application as filed, as appropriate, were furnished.

3. Additional comments:

INTERNATIONAL SEARCH REPORT

Information on patent family members

International Application No
PCT/US2004/026380

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 2004094675 A	04-11-2004	WO 2004094675 A2	04-11-2004