(19) Europäisches Patentamt
European Patent Office
Office européen des brevets

(11) **EP 2 898 067 B1**

(12) **EUROPEAN PATENT SPECIFICATION**

(45) Date of publication and mention of the grant of the patent:
15.01.2020 Bulletin 2020/03

(21) Application number: **13801769.4**

(22) Date of filing: **23.09.2013**

(51) Int Cl.:
C12N 7/00 *(2006.01)* ***C07K 14/005*** *(2006.01)*
A61K 39/215 *(2006.01)*

(86) International application number:
PCT/IB2013/058772

(87) International publication number:
WO 2014/045254 (27.03.2014 Gazette 2014/13)

(54) HUMAN BETACORONAVIRUS LINEAGE C AND IDENTIFICATION OF N-TERMINAL DIPEPTIDYL PEPTIDASE AS ITS VIRUS RECEPTOR

MENSCHLICHE BETACORONAVIRUS-LINIE C UND IDENTIFIZIERUNG VON N-TERMINALER DIPEPTIDYLPEPTIDASE ALS VIRUSREZEPTOR DAVON

LIGNÉE C DE CORONAVIRUS BÊTA HUMAINS ET IDENTIFICATION DE LA PEPTIDASE DIPEPTIDYLIQUE N-TERMINALE EN TANT QUE RÉCEPTEUR VIRAL

(84) Designated Contracting States:
AL AT BE BG CH CY CZ DE DK EE ES FI FR GB GR HR HU IE IS IT LI LT LU LV MC MK MT NL NO PL PT RO RS SE SI SK SM TR

(30) Priority: **23.09.2012 US 201261704531 P**
26.11.2012 US 201261730027 P
04.06.2013 US 201361831070 P

(43) Date of publication of application:
29.07.2015 Bulletin 2015/31

(73) Proprietors:
- **Erasmus University Medical Center Rotterdam**
  3015 GE Rotterdam (NL)
- **Dr. Soliman AbdulKhader Fakeeh Hospital Company**
  Jeddah (SA)

(72) Inventors:
- **HAAGMANS, Bartholomeus, Leonardus**
  NL-3015 GE Rotterdam (NL)
- **BESTEBROER, Theodorus, Marinus**
  NL-3015 GE Rotterdam (NL)
- **VAN BOHEEMEN, Sander**
  NL-3015 GE Rotterdam (NL)
- **FOUCHIER, Ronaldus, Adrianus, Maria**
  NL-3015 GE Rotterdam (NL)
- **OSTERHAUS, Albertus, Dominicus, Marcellinus, Erasmus**
  NL-3015 GE Rotterdam (NL)
- **ZAKI, Ali, Moh**
  NL-3015 GE Rotterdam (NL)
- **RAJ, Victor, Stalin**
  NL-3015 GE Rotterdam (NL)
- **BOSCH, Berend, Jan**
  NL-3015 GE Rotterdam (NL)

(74) Representative: **V.O.**
P.O. Box 87930
2508 DH Den Haag (NL)

(56) References cited:
WO-A1-2013/164476

- **Zaki A M: "Novel Coronavirus - Saudi Arabia: Human Isolate", ProMED-mail, Archive Number: 20120920.1302733, 20 September 2012 (2012-09-20), page 1, XP055128128, Retrieved from the Internet: URL:http://www.promedmail.org/direct.php?id=20120920.1302733 [retrieved on 2014-07-10]**
- **DATABASE EMBL [Online] 28 September 2012 (2012-09-28), "Human betacoronavirus 2c EMC/2012, complete genome.", XP002727344, retrieved from EBI accession no. EM_STD:JX869059 Database accession no. JX869059**

Printed by Jouve, 75001 PARIS (FR)

(Cont. next page)

- VAN BOHEEMEN S ET AL: "Genomic Characterization of a Newly Discovered Coronavirus Associated with Acute Respiratory Distress Syndrome in Humans", MBIO, vol. 3, no. 6, 20 November 2012 (2012-11-20), pages e00473-12, XP055128733, ISSN: 2150-7511, DOI: 10.1128/mBio.00473-12
- ZHAO G ET AL: "A safe and convenient pseudovirus-based inhibition assay to detect neutralizing antibodies and screen for viral entry inhibitors against the novel human coronavirus MERS-CoV", VIROLOGY JOURNAL, vol. 10, no. 1, 1 January 2013 (2013-01-01), pages 266-266, XP055129672, ISSN: 1743-422X, DOI: 10.1186/1742-4690-7-37
- COCKRELL A S ET AL: "Mouse Dipeptidyl Peptidase 4 Is Not a Functional Receptor for Middle East Respiratory Syndrome Coronavirus Infection", JOURNAL OF VIROLOGY, THE AMERICAN SOCIETY FOR MICROBIOLOGY, US, vol. 88, no. 9, 1 May 2014 (2014-05-01), pages 5195-5199, XP009178943, ISSN: 0022-538X
- WRENGER S ET AL: "The N-terminal structure of HIV-1 Tat is required for suppression of CD26-dependent T cell growth", JOURNAL OF BIOLOGICAL CHEMISTRY, AMERICAN SOCIETY FOR BIOCHEMISTRY AND MOLECULAR BIOLOGY, US, vol. 272, no. 48, 28 November 1997 (1997-11-28), pages 30283-30288, XP002531563, ISSN: 0021-9258, DOI: 10.1074/JBC.272.48.30283
- P. C. Y. Woo ET AL: "Comparative Analysis of Twelve Genomes of Three Novel Group 2c and Group 2d Coronaviruses Reveals Unique Group and Subgroup Features", JOURNAL OF VIROLOGY., vol. 81, no. 4, 22 November 2006 (2006-11-22), pages 1574-1585, XP055360978, US ISSN: 0022-538X, DOI: 10.1128/JVI.02182-06
- Y. YANG ET AL: "Receptor usage and cell entry of bat coronavirus HKU4 provide insight into bat-to-human transmission of MERS coronavirus", PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF THE UNITED STATES OF AMERICA, vol. 111, no. 34, 11 August 2014 (2014-08-11), pages 12516-12521, XP055537034, US ISSN: 0027-8424, DOI: 10.1073/pnas.1405889111
- WANG QIHUI ET AL: "Bat Origins of MERS-CoV Supported by Bat Coronavirus HKU4 Usage of Human Receptor CD26", CELL HOST & MICROBE, ELSEVIER, NL, vol. 16, no. 3, 11 September 2014 (2014-09-11), pages 328-337, XP029054254, ISSN: 1931-3128, DOI: 10.1016/J.CHOM.2014.08.009

Description

[0001] The invention provides a new previously undescribed Coronavirus isolated from cases of unexplained disease in September 2012 and identified herein as belonging to a newly recognized and previously undescribed species of human Corona Virus (HCoV), herein identified as HCoV-SA1 or HCoV EMC or Middle East Respiratory Syndrome-Coronavirus (MERS-CoV). In particular the nucleic acid and/or amino acid sequences of the MERS-CoV genome and sequences specifically encoding (parts of) viral proteins and antigenic polypeptides are provided. Further, the invention relates to diagnostic means and methods, prophylactic means and methods and therapeutic means and methods to be employed in the diagnosis, prevention and/or treatment of disease, in particular of respiratory disease, in particular of mammals, more in particular in humans. It particularly also relates to an isolated virus and its receptor.

[0002] A fundamental yet unresolved puzzle in virology is how viruses evolve to recognize their receptor proteins on the cells they need to enter in order to replicate. Specifically, how do different viruses recognize the same receptor protein and how do similar viruses recognize different receptor proteins? Do viruses select their receptor proteins by chance or do they target specific virus binding hotspots on these receptor proteins? Structural information of virus-receptor interfaces can potentially answer these questions. To date, although a few studies have obtained structural information for a single virus-receptor interface, even less studies have provided structural information for the interfaces between different viruses and their common receptor protein.

[0003] The invention in particular relates to coronaviruses that are the second leading cause of adult colds. Of the more than 30 kinds, three or four infect humans. The 2003 SARS virus is a coronavirus. Coronaviruses are rather difficult to grow in the laboratory, so they have not been studied to the same extent as other viruses. NL63 coronavirus (NL63 CoV), a prevalent human respiratory virus, is a group I coronavirus known to use angiotensin converting enzyme 2 (ACE2, a cell membrane bound carboxy terminal dipeptidyl peptidase) as its receptor. Incidentally, ACE2 is also used by group II SARS coronavirus (SARS CoV).

[0004] The distribution of coronavirus receptors is critical to the pathogenic outcome of the disease they cause. In this regard, it is notable that coronavirus spikes exhibit a wide range of receptor specificities; human aminopeptidase N (a metalloprotease) is a receptor for human coronavirus 229E, mouse hepatitis virus enters after binding members of a pleiotropic family of carcinoembryonic antigen cell adhesion molecules (CEACAMs); feline and porcine coronaviruses also bind various metalloproteases; and bovine coronaviruses recognize 9 O acetylated sialic acids.

[0005] Coronaviruses enter cells through a large spike protein on their envelopes. The coronavirus spike protein is a membrane anchored trimer and contains two subunits, receptor binding subunit S1 and membrane fusion subunit S2. The S2 subunits from group I and group II coronaviruses share both sequence and structural homology; they contain homologous heptad repeat segments that fold into a conserved trimers of hairpin structure, which is essential for membrane fusion. Surprisingly, the S1 subunits from group I and group II coronaviruses have no obvious sequence homology. Nevertheless, they can be divided approximately into N terminal region, central region, and C terminal region. Coronaviruses are believed to have common ancestors because they share similar replication mechanisms, genomic structures, and overall gene sequences.

[0006] Among all of the coronavirus genes, the one encoding the spike protein is the most variable. Between the spike protein subunits, S1 is more variable than S2. The current structural divergences of the S1 subunits reveal the tremendous evolutionary pressure that coronaviruses face to adapt to different host receptors, and they also reflect on the evolutionary history of coronaviruses and their receptor selections.

[0007] In general, coronaviruses are well known and most of those who are diagnosed with it recover completely with no complications after receiving the needed supportive therapy. However, in some of the patients who are infected, serious complications can develop affecting the respiratory system and the kidneys and can cause death, especially among the elderly and in patients with chronic respiratory and cardiac conditions and among immune compromised patients.

[0008] Coronaviruses (CoVs), a genus of the Coronaviridae family, are positive strand RNA viruses with the largest viral genome of all RNA viruses (27 - 32 Kb). The genomic RNA is capped, polyadenylated and covered with nucleocapsid proteins. The virus is enveloped and carries large spike glycoproteins. All CoVs employ a common genome organization where the replicase gene encompasses the 5' two thirds of the genome and is comprised of two overlapping open reading frames (ORFs), ORF1a and ORF1b.

[0009] The structural gene region, which covers the 3' third of the genome, encodes the canonical set of structural protein genes in the order 5' spike (S) envelope (E) membrane (M) and nucleocapsid (N) - 3'. Some beta CoVs carry an additional structural protein encoding a hemagglutinin esterase (HE). The gene is located between the ORFlb and S gene. Expression of the nonstructural replicase proteins is mediated by translation of the genomic RNA that gives rise to the biosynthesis of two large polyproteins, ppla (encoded by ORF1a) and pplab (encoded by ORF1a and ORF1b) facilitated by a ribosomal frame shift at the ORF1a/1b junction.

[0010] In contrast, the structural proteins are translated from sub genomic (sg) mRNAs. These sg mRNAs are the result of discontinuous transcription, a hallmark of CoV gene expression. The structural gene region also harbors several

ORFs that are interspersed along the structural protein coding genes. The number and location of these accessory ORFs varies between the CoV species.

[0011] Although coronaviruses were first identified nearly 60 years ago, they only received notoriety in 2003 when one of their members was identified as the aetiological agent of severe acute respiratory syndrome (SARS). Previously these viruses were known to be important agents of respiratory and enteric infections of domestic and companion animals and to cause approximately 15% of all cases of the common cold. Coronaviruses (CoVs), a genus of the Coronaviridae family, are positive strand RNA viruses with the largest viral genome of all RNA viruses (27 - 32 Kb). The genomic RNA is capped, polyadenylated and covered with nucleocapsid proteins. The virus is enveloped and carries large spike glycoproteins. All CoVs employ a common genome organization where the replicase gene encompasses the 5'-two thirds of the genome and is comprised of two overlapping open reading frames (ORFs), ORF1a and ORF1b. The structural gene region, which covers the 3'-third of the genome, encodes the canonical set of structural protein genes in the order 5' - spike (S) - envelope (E) - membrane (M) and nucleocapsid (N) - 3'. Some beta-CoVs carry an additional structural protein encoding a heamagglutinin-esterase (HE). The gene is located between the ORFIb and S gene. Expression of the nonstructural replicase proteins is mediated by translation of the genomic RNA that gives rise to the biosynthesis of two large polyproteins, ppla (encoded by ORF1a) and pplab (encoded by ORF1a and ORF1b) facilitated by a ribosomal frame shift at the ORF1a/1b junction. In contrast, the structural proteins are translated from sub genomic (sg) mRNAs. These sg mRNAs are the result of discontinuous transcription, a hallmark of CoV gene expression. The structural gene region also harbors several ORFs that are interspersed along the structural protein coding genes. The number and location of these accessory ORFs varies between the CoV species. In animals CoV infections can lead to a variety of syndromes, e.g. bronchitis, gastroenteritis, progressive demyelinating encephalitis, diarrhea, peritonitis and respiratory tract disease. The first reports on human CoVs (HCoV) appeared in the mid-1960s. The human viruses were isolated from persons with common cold, and two species were detected: HCoV-229E and HCoV-OC43. Almost 40 years later, SARS-CoV was identified as the causative agent of the Severe Acute Respiratory Syndrome (SARS). A highly effective global public health response prevented further spread of this virus, and as a result SARS-CoV was eradicated from the human population. Soon thereafter it became clear that there are more HCoVs. HCoV-NL63 was identified in 2004 and HCoV-HKU1 in 2005. Both viruses are not emerging viruses like SARS-CoV but were previously unidentified. In fact, infections by these viruses are as common and wide spread as HCoV-229E and HCoV-OC43 infections. The SARS outbreak intensified the research on the unknown animal CoVs. As much as 16 new animal CoV species were identified till 2008. There are currently at around 29 complete reference genome sequences available in Genbank of the various viruses. Recently, the Coronavirus Study Group of the International Committee for Taxonomy of Viruses has proposed renaming the traditional group 1, 2, and 3 coronaviruses into the genus Alphacoronavirus, Betacoronavirus, and Gammacoronavirus, respectively (http://talk.ictvonline.org/media/p/1230.aspx). Each genus is subdivided into different species on the basis of sequence identity in the replicase domains of the polyprotein pplab.

[0012] The classification of the family Coronaviridae and the organization of the established subfamily Coronavirinae is based upon rooted phylogeny and pair-wise comparisons using Coronaviridae-wide conserved domains in replicase polyprotein pplab as well as the structural proteins S, E, M and N. In rooted trees, the proposed genera Alpha-, Beta- and Gammacoronavirus consistently form three distinct monophyletic groups and in pair-wise comparisons, they form three robust non-overlapping clusters. The inter-group pair-wise scores for coronaviruses are comparable to those calculated for structural and non-structural proteins of different genera in other RNA virus families (e.g. Potyviridae, Picornaviridae). Based on this defacto criterion phylogroups 1 through 3 are named into genera designated Alpha-, Beta and Gammacoronavirus, respectively. The 90% aa sequence identity threshold now proposed as a species demarcation criterion within each genus has been determined from the analysis of pair-wise aa distances in seven conserved replicase domains (nsp3 ADRP, nsp5 (3CLpro), nsp12 (RdRp), nsp13 (Hell), nsp14 (ExoN), nsp15 (NendoU) and nsp16 (O-MT)) of 156 viruses in the Coronaviridae. In this analysis, 20 distinct groups (17 coronaviruses, 2 toroviruses, 1 bafinivirus) are unambiguously recognized as non-overlapping clusters (with the largest intra-cluster distance being smaller than the smallest inter-cluster distance). Of these clusters, at least 7 fall into the genus Betacoronavirus, each of which represents a distinct betacoronavirus species (Betacoronavirus 1, Murine coronavirus, Human coronavirus HKU1 , Rousettus bat coronavirus HKU9, Tylonycteris bat coronavirus HKU4, Pipistrellus bat coronavirus HKU5, Severe acute respiratory syndrome-related coronavirus (SARS-CoV). The Betacoronavirus genus is additionally considered to contain 4 lineages (A, B, C and D). Human coronaviruses HCoV-HKU1 and HCoV-OC43 belong to lineage A while human coronavirus SARS-CoV belongs to lineage B. Lineage C and D are not known to contain any human representatives. Other human coronaviruses, such as HCoV-NL63 and HCoV-229E, are even more distinct since these two human pathogens belong to a different genus, the Alphacoronavirus genus.

[0013] The invention also relates to so called "pull down" experiments, which are methods for the identification of protein protein interactions based on affinity purification of interacting proteins from complex proteinaceous substances such as cellular extracts. Pull down experiments with, for example, fusion proteins attached to inert beads are a screening technique for isolating proteinaceous substances having specific protein components that bind to each other and thus lead to identification of protein protein interactions.

[0014] Typically, pull down experiments are used to identify interactions between a probe protein and unknown targets and to confirm suspected interactions between a probe protein and a known protein. When coupled with peptide digests of pulled down proteins and with mass spectrometry to sequence those peptides and identify targets, pull downs can be considered as the protein based equivalent of a yeast two hybrid screen.

[0015] To improve the isolation of specific binding partners, pull down methods have been developed involving the use of cross linking, of large scale tissue lysates, and of spin columns. Appropriate methods of sample preparation for mass spectrometry based identification of interacting proteins have been developed as well, including specialized gel staining techniques, band excision, and in gel tryptic digestion. Data interpretation and most commonly encountered problems are, for example, discussed in Current Protocols in Cell Biology, "UNIT 17.5 Protein Protein Interactions Identified by Pull Down Experiments and Mass Spectrometry," Adam Brymora, Valentina A. Valova, and Phillip J. Robinson, Published Online: 1 May 2004 DOI: 10.1002/0471143030.cb1705s22, and included herein by reference.

The invention

[0016] The invention provides an essentially mammalian positive-sense single stranded RNA virus, a nucleic acid, a vector, a cell, a protein, an antigen, an antibody, a diagnostic kit, a pharmaceutical composition a proteineous substance, a container qnd a method as set forth in the claims. cIn particular, the invention provides an essentially mammalian positive-sense single stranded RNA virus Betacoronavirus having a receptor binding domain (RBD) capable of binding to a dipeptidyl peptidase 4 (DPP4). In particular, no such isolates have been deposited or in any other way made available to the art until now. In a preferred embodiment, a virus according to the invention is isolated or isolatable from a human. In particular, the invention provides a new previously undescribed Coronavirus isolated from cases of unexplained disease in September 2012 and identified herein as belonging to a newly recognized and previously undescribed species of human Corona Virus (HCoV), herein identified as HCoV-SA1 or HCoV EMC or Middle East Respiratory Syndrome-Coronavirus (MERS-CoV). In particular the specific nucleic acid and/or amino acid sequences of the MERS-CoV genome and sequences encoding (parts of) viral proteins and antigenic polypeptides are provided. Further, the invention relates to diagnostic means and methods, prophylactic means and methods and therapeutic means and methods to be employed in the diagnosis, prevention and/or treatment of disease, in particular of respiratory disease, in particular of mammals, more in particular in humans, most in particular specific for MERS-CoV. It particularly also relates to an isolated virus and its receptor. The description also provides identification of N-terminal dipeptidyl peptidase as virus receptor and uses thereof, identification of the receptor binding domain of MERS-CoV mapping to a 231-residue region 2 in the spike protein that efficiently elicits neutralizing antibodies identification and uses thereof and dipeptidyl peptidase 4 receptor determinants of respiratory MERS-coronavirus infection, and uses thereof. The description in particular provides specific diagnostics of MERS-CoV, sub-unit compositions of S1-MERS CoV protein for vaccine purposes, screening tests for detecting compounds capable of interfering with MER-CoV-DPP4 binding, and animal models for determining activity of compounds capable of interfering with MERS-CoV-DPP4 binding.

[0017] The description also provides a virus according to the invention having an amino acid sequence of its receptor binding domain that is at least 75%, more preferably at least 80%, more preferably at least 85%, more preferably at least 90%, most preferably at least 95% identical with the amino acid sequence of the receptor binding domain of an isolated essentially mammalian positive-sense single stranded RNA virus classifiable as belonging to the Order: Nidovirales; Family: Coronaviridae; Subfamily: Coronavirinae; Genus: Betacoronavirus; Lineage C isolatable from humans and comprising one or more of the sequences selected from any of figures 3 or 5 to 15, preferably wherein said receptor domain comprises residues 1 747 of the S1 spike protein, preferably residues 358 588 of the S1 spike protein.

[0018] In a one embodiment, a virus is provided in the present description that belongs to the Coronaviruses, genus Betacoronavirus and is identifiable as phylogenetically corresponding or specific to the MERS-CoV thereto by determining a nucleic acid or amino acid sequence of said virus or fragments thereof and testing it in phylogenetic tree analyses wherein maximum likelihood trees are generated and finding it, the virus or fragment, to be more closely phylogenetically corresponding to a virus isolate or fragment thereof having the sequences as depicted in any of figures 3 or 5 to 15 than it is corresponding to a bat coronavirus HKU4 or HKU5, or fragments thereof, in another embodiment, a virus is provided that belongs to the Coronaviruses and is identifiable as phylogenetically corresponding or specific to the MERS-CoV thereto by determining a nucleic acid sequence or amino acid sequence of said virus and testing it in phylogenetic tree analyses wherein maximum likelihood trees are generated and finding it to be more closely phylogenetically corresponding to a virus isolate isolatable from humans having the sequences as depicted in any of figures 3 or 5 to 15 than it is corresponding to a human coronavirus virus isolate HCoV-HKU1 or HCoV-OC43 or SARS-CoV, or fragments thereof.

[0019] In a preferred embodiment, a virus is provided herein that belongs to the Coronaviruses, genus Betacoronavirus and is identifiable as phylogenetically corresponding thereto by determining a nucleic acid or amino acid sequence of said virus and testing it in phylogenetic tree analyses wherein maximum likelihood trees are generated and finding it to be more closely phylogenetically corresponding to a virus isolate having the sequences as depicted in any of figures 3 or 5 to 15 than it is corresponding to a bat coronavirus HKU4 or HKU5 or to a human coronavirus virus isolate HCoV-

HKU1 or HCoV-OC43 or SARS-CoV.

[0020] The invention also provides a cell, preferably a host cell, and a culture of such a cell or host cell, i.e. a cultured cell, comprising a virus according to the invention. Preferred examples of such cells and cell cultures comprise a Vero cell or LLC-MK2 cell and cultures thereof; other preferred examples comprise a Huh-7 cell, a primary nonciliated human airway epithelial cell, a primary human fibroblast, a primary human kidney cell, a primary human alveolar type 2 cell, or a primary kidney cell of Pipistrellus pipistrellu, and cultures of said cells.

[0021] The description also provides a nucleic acid, preferably a cDNA, or MERS-CoV-specific fragment thereof obtainable, derived or obtained from an isolated essentially mammalian positive-sense single stranded RNA virus classifiable as belonging to the Order: Nidovirales; Family: Coronaviridae; Subfamily: Coronavirinae; Genus: Betacoronavirus; and non-Lineage A, non-Lineage B or non-Lineage D, human betacoronavirus. In a preferred embodiment, the invention provides a nucleic acid isolatable from a human virus, preferably isolatable from humans, having a receptor binding domain (RBD) capable of binding to a dipeptidyl peptidase 4 (DPP4), In particular, a nucleic acid is provided by the invention obtainable, derived or obtained from an isolated essentially mammalian positive-sense single stranded RNA virus classifiable as belonging to the Order: Nidovirales; Family: Coronaviridae; Subfamily: Coronavirinae; Genus: Betacoronavirus; Lineage C human betacoronavirus. a nucleic acid is provided by the description obtainable, derived or obtained from an isolated essentially mammalian positive-sense single stranded RNA virus classifiable as belonging to the Order: Nidovirales; Family: Coronaviridae; Subfamily: Coronavirinae; Genus: Betacoronavirus; Lineage: C and isolatable from humans, and components thereof. Until now, no Betacoronavirus isolates have been isolated from humans that were then classified as belonging to Lineage: C of Betacoronavirus. In particular, a nucleic acid is provided by the description obtainable, derived or obtained from a Lineage: C Betacoronavirus having a receptor binding domain (RBD) capable of binding to a dipeptidyl peptidase 4 (DPP4), preferably from a virus having a nucleic acid sequence at least 75%, more preferably at least 80%, more preferably at least 85%, more preferably at least 90%, most preferably at least 95% identical with a nucleic acid sequence provided in figure 3 or figures 5 to 15. In particular, a MERS-CoV specific fragment of a nucleic acid, RNA or DNA or cDNA is provided by the description which comprises one or more of the sequences of MERS-CoV as depicted in figures 3, or 5 to 15 or a nucleic acid sequence which can hybridize with any of these sequences under stringent conditions. The invention also provides a vector comprising a nucleic acid according to the invention, and a host cell comprising a nucleic acid according to the invention or a vector according to the invention.

[0022] The description also provides an isolated or recombinant proteinaceous molecule or MERS-CoV-specific fragment thereof encoded by a nucleic acid according to the invention. In a preferred embodiment, the invention provides a MERS-CoV-specific viral protein encoded by a nucleic acid according to the invention. Useful proteinaceous molecules are for example derived from any of the genes or genomic fragments or open reading frames (ORFs) derivable from a virus according to the invention. Such molecules, or antigenic fragments thereof, as provided herein, are for example useful in diagnostic methods or kits and in pharmaceutical compositions such as sub-unit vaccines and inhibitory peptides. Particularly useful is the viral polymerase protein, the spike protein, the nucleocapsid or antigenic fragments thereof for inclusion as antigen or subunit immunogen in a vaccine, but inactivated whole virus can also be used. Particularly useful are also those proteinaceous substances that are encoded by recombinant nucleic acid fragments that are identified by phylogenetic analyses as being MERS-CoV specific fragments, of course preferred are those that are within the preferred bounds and metes of ORFs useful in phylogenetic analyses, in particular for eliciting MERS-CoV specific antibodies, whether in vivo (e.g. for protective purposes or for providing diagnostic antibodies) or in vitro (e.g. by phage display technology or another technique useful for generating synthetic antibodies).

[0023] In one embodiment, the description provides a viral replicase or MERS-CoV-specific fragment thereof having an amino acid sequence at least 75%, more preferably at least 80%, more preferably at least 85%, more preferably at least 90%, most preferably at least 95% identical with an amino acid sequence provided in figure 13, said viral replicase or MERS-CoV-specific fragment thereof preferably encoded by an RNA or DNA or cDNA sequence or fragments or homologues thereof as provided herein, having a nucleic acid sequence at least 75%, more preferably at least 80%, more preferably at least 85%, more preferably at least 90%, most preferably at least 95% identical with a nucleic acid sequence provided in figure 13.

[0024] In another embodiment, the description provides a viral spike protein or MERS-CoV-specific fragment thereof having an amino acid sequence at least 75%, more preferably at least 80%, more preferably at least 85%, more preferably at least 90%, most preferably at least 95% identical with an amino acid sequence provided in figure 12, said viral spike protein or MERS-CoV-specific fragment thereof preferably encoded by an RNA or DNA or cDNA sequence or fragments or homologues thereof as provided herein, having a nucleic acid sequence at least 75%, more preferably at least 80%, more preferably at least 85%, more preferably at least 90%, most preferably at least 95% identical with a nucleic acid sequence provided in figure 12.

[0025] In another embodiment, the description provides an S1 spike protein or MERS-CoV-specific fragment thereof having an amino acid sequence at least 75%, more preferably at least 80%, more preferably at least 85%, more preferably at least 90%, most preferably at least 95% identical with an amino acid sequence provided for residues 1 588 in figure 17.

[0026] In another embodiment, the description provides an S1 spike protein or MERS-CoV-specific fragment thereof having an amino acid sequence at least 75%, more preferably at least 80%, more preferably at least 85%, more preferably at least 90%, most preferably at least 95% identical with an amino acid sequence provided for residues 1 357 in figure 17.

[0027] In another embodiment, the description provides an S1 spike protein or MERS-CoV-specific fragment thereof having an amino acid sequence at least 75%, more preferably at least 80%, more preferably at least 85%, more preferably at least 90%, most preferably at least 95% identical with an amino acid sequence provided for residues 358 747 in figure 17.

[0028] In another embodiment, the description provides an S1 spike protein or MERS-CoV-specific fragment thereof having an amino acid sequence at least 75%, more preferably at least 80%, more preferably at least 85%, more preferably at least 90%, most preferably at least 95% identical with an amino acid sequence provided for residues 358-588 in figure 17.

[0029] In another embodiment, the description provides an S1 spike protein or MERS-CoV-specific fragment thereof having an amino acid sequence at least 75%, more preferably at least 80%, more preferably at least 85%, more preferably at least 90%, most preferably at least 95% identical with an amino acid sequence provided for residues 589 747 in figure 17.

[0030] In another embodiment, the description provides an S1 spike protein or MERS-CoV-specific fragment thereof having an amino acid sequence at least 75%, more preferably at least 80%, more preferably at least 85%, more preferably at least 90%, most preferably at least 95% identical with an amino acid sequence provided for residues 1 747 in figure 17.

[0031] In another embodiment, the description provides an S1 spike protein or MERS-CoV-specific fragment thereof having an amino acid sequence at least 75%, more preferably at least 80%, more preferably at least 85%, more preferably at least 90%, most preferably at least 95% identical with an amino acid sequence provided for residues 1 747 in figure 17.

[0032] In another embodiment, the description provides a viral non-structural gene protein or MERS-CoV-specific fragment thereof having an amino acid sequence at least 75%, more preferably at least 80%, more preferably at least 85%, more preferably at least 90%, most preferably at least 95% identical with an amino acid sequence provided in figure 8, said viral non-structural gene protein or MERS-CoV-specific fragment preferably encoded by an RNA or DNA or cDNA sequence or fragments or homologues thereof as provided herein, having a nucleic acid sequence at least 75%, more preferably at least 80%, more preferably at least 85%, more preferably at least 90%, most preferably at least 95% identical with a nucleic acid sequence provided in figure 8.

[0033] In another embodiment, the description provides a viral non-structural gene protein or MERS-CoV-specific fragment thereof having an amino acid sequence at least 75%, more preferably at least 80%, more preferably at least 85%, more preferably at least 90%, most preferably at least 95% identical with an amino acid sequence provided in figure 9, said viral non-structural gene protein or MERS-CoV-specific fragment preferably encoded by an RNA or DNA or cDNA sequence or fragments or homologues thereof as provided herein, having a nucleic acid sequence at least 75%, more preferably at least 80%, more preferably at least 85%, more preferably at least 90%, most preferably at least 95% identical with a nucleic acid sequence provided in figure 9.

[0034] In another embodiment, the description provides a viral non-structural gene protein or MERS-CoV-specific fragment thereof having an amino acid sequence at least 75%, more preferably at least 80%, more preferably at least 85%, more preferably at least 90%, most preferably at least 95% identical with an amino acid sequence provided in figure 10, said viral non-structural gene protein or MERS-CoV-specific fragment preferably encoded by an RNA or DNA or cDNA sequence or fragments or homologues thereof as provided herein, having a nucleic acid sequence at least 75%, more preferably at least 80%, more preferably at least 85%, more preferably at least 90%, most preferably at least 95% identical with a nucleic acid sequence provided in figure 10.

[0035] In another embodiment, the description provides a viral non-structural gene protein or MERS-CoV-specific fragment thereof having an amino acid sequence at least 75%, more preferably at least 80%, more preferably at least 85%, more preferably at least 90%, most preferably at least 95% identical with an amino acid sequence provided in figure 11, said viral non-structural gene protein or MERS-CoV-specific fragment preferably encoded by an RNA or DNA or cDNA sequence or fragments or homologues thereof as provided herein, having a nucleic acid sequence at least 75%, more preferably at least 80%, more preferably at least 85%, more preferably at least 90%, most preferably at least 95% identical with a nucleic acid sequence provided in figure 11.

[0036] In another embodiment, the description provides a viral small envelope (E) protein or MERS-CoV-specific fragment thereof having an amino acid sequence at least 75%, more preferably at least 80%, more preferably at least 85%, more preferably at least 90%, most preferably at least 95% identical with an amino acid sequence provided in figure 7, said viral small envelope (E) protein or MERS-CoV-specific fragment preferably encoded by an RNA or DNA or cDNA sequence or fragments or homologues thereof as provided herein, having a nucleic acid sequence at least 75%, more preferably at least 80%, more preferably at least 85%, more preferably at least 90%, most preferably at least 95% identical with a nucleic acid sequence provided in figure 7.

[0037] In another embodiment, the description provides a viral matrix (M) protein or MERS-CoV-specific fragment thereof having an amino acid sequence at least 75%, more preferably at least 80%, more preferably at least 85%, more preferably at least 90%, most preferably at least 95% identical with an amino acid sequence provided in figure 6, said viral matrix (M) protein or MERS-CoV-specific fragment preferably encoded by an RNA or DNA or cDNA sequence or

fragments or homologues thereof as provided herein, having a nucleic acid sequence at least 75%, more preferably at least 80%, more preferably at least 85%, more preferably at least 90%, most preferably at least 95% identical with a nucleic acid sequence provided in figure 6.

[0038] In another embodiment, the description provides a nucleocapsid (N) protein or MERS-CoV-specific fragment thereof having an amino acid sequence at least 75%, more preferably at least 80%, more preferably at least 85%, more preferably at least 90%, most preferably at least 95% identical with an amino acid sequence provided in figure 5, said nucleocapsid (N) protein or MERS-CoV-specific fragment preferably encoded by an RNA or DNA or cDNA sequence or fragments or homologues thereof as provided herein, having a nucleic acid sequence at least 75%, more preferably at least 80%, more preferably at least 85%, more preferably at least 90%, most preferably at least 95% identical with a nucleic acid sequence provided in figure 5.

[0039] The invention also provides an antigen comprising a protein as provided herein. In a preferred embodiment, said proteinaceous molecule comprises or consists of a nucleocapsid (N) protein thereof as provided herein, or a viral matrix (M) protein as provided herein, or a viral small envelope (E) protein as provided herein, or a viral non-structural gene protein o as provided herein, or an S1 spike protein or fragment thereof as provided herein, or a viral replicase as provided herein.

[0040] Also provided herein are antibodies, be it natural polyclonal or monoclonal, or synthetic (e.g. (phage) library-derived binding molecules) antibodies that specifically react with an antigen according to the invention. A person skilled in the art will be able to develop (monoclonal) antibodies using isolated virus material and/or recombinantly expressed viral proteins. In particular the invention provides a rabbit antibody specifically directed against an antigen according to the invention, rabbits being particularly well suited to raise antibodies against an antigen according to the invention. Such antibodies are also useful in a method for identifying a viral isolate as a MERS-CoV comprising reacting said viral isolate or a component thereof with an antibody as provided herein. This can for example be achieved by using purified or non-purified MERS-CoV or parts thereof (proteins, peptides) using ELISA, RIA, FACS or similar formats of antigen detection assays (Current Protocols in Immunology). Alternatively, infected cells or cell cultures may be used to identify viral antigens using classical immunofluorescence or immunohistochemical techniques. Specifically useful in this respect are antibodies raised against MERS-CoV proteins or peptides of the invention which are encoded by a nucleotide sequence comprising one or more of the fragments disclosed in figures 3 and 5 to 15. Antibodies, both monoclonal and polyclonal, or fragments thereof, can also be used for detection purpose in the present invention, for example, in immunoassays in which they can be utilized in liquid phase or bound to a solid phase carrier. In addition, the monoclonal antibodies in these immunoassays can be detectably labeled in various ways. A variety of immunoassay formats may be used to select antibodies specifically reactive with a particular protein (or other analyte). For example, solid-phase ELISA immunoassays are routinely used to select monoclonal antibodies specifically immunoreactive with a protein. See Harlow and Lane, Antibodies, A Laboratory Manual, Cold Spring Harbor Publications, New York (1988), for a description of immunoassay formats and conditions that can be used to determine selective binding. Examples of types of immunoassays that can utilize antibodies of the invention are competitive and non-competitive immunoassays in either a direct or indirect format. Examples of such immunoassays are the radioimmunoassay (RIA) and the sandwich (immunometric) assay. Detection of the antigens using the antibodies of the invention can be done utilizing immunoassays that are run in either the forward, reverse, or simultaneous modes, including immunohistochemical assays on physiological samples. Those of skill in the art will know, or can readily discern, other immunoassay formats without undue experimentation.

[0041] Antibodies can be bound to many different carriers and used to detect the presence of the target molecules. Examples of well-known carriers include glass, polystyrene, polypropylene, polyethylene, dextran, nylon, amylases, natural and modified celluloses, polyacrylamides, agaroses and magnetite. The nature of the carrier can be either soluble or insoluble for purposes of the invention. Those skilled in the art will know of other suitable carriers for binding monoclonal antibodies, or will be able to ascertain such using routine experimentation.

[0042] The invention also provides method for identifying a viral isolate as a MERS-CoV comprising reacting said viral isolate or a component thereof with a nucleic acid according to the invention and/or with an antibody according to the invention. The invention for example provides a method for virologically diagnosing a MERS-CoV infection of an animal, in particular of a mammal, more in particular of a human being, comprising determining in a sample of said animal the presence of a viral isolate or component thereof by reacting said sample with a MERS-CoV specific nucleic acid or antibody according to the invention, and a method for serologically diagnosing a MERS-CoV infection of a mammal comprising determining in a sample of said mammal the presence of an antibody specifically directed against a MERS-CoV or component thereof by reacting said sample with a MERS-CoV-specific proteinaceous molecule or fragment thereof or an antigen according to the invention.

[0043] The invention also provides a diagnostic kit for diagnosing a MERS-CoV infection comprising a MERS-CoV, a MERS-CoV-specific nucleic acid, proteinaceous molecule or fragment thereof, antigen and/or an antibody according to the invention, and preferably a means for detecting said MERS-CoV, MERS-CoV-specific nucleic acid, proteinaceous molecule or fragment thereof, antigen and/or an antibody, said means for example comprising an excitable group such

as a fluorophore or enzymatic detection system used in the art (examples of suitable diagnostic kit format comprise IF, ELISA, neutralization assay, RT-PCR assay). To determine whether an as yet unidentified virus component or synthetic analogue thereof such as nucleic acid, proteinaceous molecule or fragment thereof can be identified as MERS-CoV-specific, it suffices to analyze the nucleic acid or amino acid sequence of said component, for example for a stretch of said nucleic acid or amino acid, preferably of at least 10, more preferably at least 25, more preferably at least 40 nucleotides or amino acids (respectively), by sequence homology comparison with the provided MERS-CoV nucleic acid or amino acid sequences and with known non-MERS-CoV nucleic acid or amino acid sequences using for example phylogenetic analyses as provided herein. Depending on the degree of relationship with said MERS-CoV or non-MERS-CoV viral sequences, the component or synthetic analogue can be identified.

[0044] The invention also provides use of a virus according to the invention, and/or a nucleic acid according to the invention , a vector according to the invention, a host cell according the invention, a proteinaceous molecule or fragment thereof according to the invention, an antigen according to the invention, or an antibody according to the invention for the production of a pharmaceutical composition, preferably for the production of a pharmaceutical composition for therapeutical use, preferably for use in antiviral therapy, preferably for the treatment or prevention of a Betacoronavirus, Lineage C virus infection, preferably a human infection, preferably an infection with a MERS-Cov,. Preferably a peptide comprising part of the amino acid sequence of the spike protein as depicted in figure 17 (residues 358-588, comprising the essential receptor binding domain) is used for the preparation of a therapeutic or prophylactic peptide, preferably for inclusion in said pharmaceutical composition. Also preferably, a protein comprising the amino acid sequence of the spike protein as depicted in figure 17 (residues 358-588) is used for the preparation of a sub-unit vaccine. Furthermore, the nucleocapsid of Coronaviruses, as depicted in figure 5, is known to be particularly useful for eliciting cell-mediated immunity against Coronaviruses and can be used for the preparation of a sub-unit vaccine. The invention also comprises a pharmaceutical composition comprising a virus according to the invention, and/or a nucleic acid according to the invention, a vector according to the invention, a host cell according the invention, a proteinaceous molecule or fragment thereof according to the invention, an antigen according to the invention, or an antibody according to the invention.

[0045] The description also provides a method for the treatment or prevention of a Betacoronavirus, Lineage C virus infection or for the treatment or prevention of atypical pneumonia comprising providing a mammal, preferably a human individual with a pharmaceutical composition according to the invention. Also, the description provides a method for the treatment or prevention of atypical pneumonia and/or renal failure comprising providing an individual with a pharmaceutical composition according to the invention. In a preferred embodiment, a method for the treatment or prevention of a MERS-CoV infection is provided comprising providing a mammal with a pharmaceutical composition according the invention, preferably wherein said mammal is a rabbit. The description also provides a method for in vivo determining of parameters of MERS-CoV infection, preferably for determining parameters of MERS-CoV-DPP4 interaction in an animal experiment, comprising providing a mammal with a pharmaceutical composition according to the invention, and/or with a virus according to the invention, and/or with a nucleic acid according to the invention, and/or with a vector according to the invention, and/or with a host cell according to the invention, and/or with a proteinaceous molecule or fragment thereof according to the invention, and/or with an antigen according to the invention, and/or with an antibody according to the invention, preferably wherein said mammal is a rabbit. It is herein found that rabbits have several advantages over other experimental animals in that they have a remarkably similar target sequence for MERS-CoV-DPP4-receptor interaction, resulting in proficient infection of a rabbit with MERS-CoV and thus ample chance to study various aspects and parameters of MERS-CoV-DPP4-receptor interaction that resemble those in humans, giving the rabbit experimental animal model a distinct advantage over other animal models, such as the ferret animal model. Phylogenetic analysis of the MERS-CoV binding region of DPP4 indicated that human, macaque, horse and rabbit DPP4 cluster together as do DPP4's from cattle, pig and bats, that are somewhat more distantly related. Small animals including ferret, mice and most likely hamsters, shown to resist MERS-CoV infection, are more divergent in the DPP4 virus binding region, which at least in the case of ferrets has consequences for MERS-CoV binding. Besides macaques, rabbits indeed are a potential animal model for MERS-CoV infection; ex vivo inoculation of rabbit lung and kidney tissues revealed susceptibility to MERS-CoV. Similarly, the description provides a method for in vivo determining of parameters of MERS-CoV infection, preferably for determining parameters of protection against MERS-CoV-infection, comprising providing a mammal with a pharmaceutical composition according to the invention, and/or with a virus according to the invention, and/or with a nucleic acid according to the invention, and/or with a vector according to the invention, and/or with a host cell according to the invention, and/or with a proteinaceous molecule or fragment thereof according to the invention, and/or with an antigen according to the invention, and/or with an antibody according to the invention, preferably wherein said mammal is a rabbit. In particular, a rabbit model is a model of choice for testing a pharmaceutical composition comprising a subunit peptide vaccine comprising part of the amino acid sequence of the spike protein as depicted in figure 17 (fragments of residues 358-588, comprising the essential receptor binding domain) which is used for the preparation of a therapeutic or prophylactic peptide for the preparation of a sub-unit vaccine. Vaccinating or immunizing rabbits with variant peptide vaccines and then challenging vaccinated and control rabbits with MERS-CoV that allows rapid infection and measurement of essential parameters such as development of (neutralizing) antibodies in experimental and control rabbits,

development of protection against MERS-CoV infection or against MERS-CoV transmission allows for relatively inexpensive and rapid vaccine development studies, thereby allowing rapid vaccine development against human MERS-CoV infections. Attenuation of the virus by serial passage of MERS-CoV can now preferably achieved in rabbits by established methods developed for this purpose, including but not limited to the use of related viruses of other species, serial passages through other laboratory animals or/and tissue/cell cultures, serial passages through cell cultures at temperatures below 37C (cold-adaption), site directed mutagenesis of molecular clones and exchange of genes or gene fragments between related viruses.

[0046] Now, as herein provided, a new human coronavirus was isolated from a patient with pneumonia. The virus was isolated from sputum of a male patient aged 60 years old presenting with pneumonia associated with acute renal failure. The virus grows readily on Vero cells and LLC-MK2 cells producing CPE in the form of rounding and syncytia formation and uses dipeptidyl peptidase 4 (DPP4) as a viral receptor for entry into cells establishing infection.

[0047] The clinical isolate was initially tested for influenza virus A, influenza virus B, parainfluenza virus, enterovirus and adenovirus, with negative results. Testing with a pancoronavirus RT-PCR yielded a band at a molecular weight appropriate for a coronavirus. The virus RNA was tested and it was confirmed to be a new member of the beta group of coronaviruses, closely related to bat coronaviruses. The invention relates to a new previously undescribed Coronavirus isolated from cases of unexplained disease in September 2012 and identified herein as belonging to a newly recognized and previously undescribed species of human Corona Virus (HCoV), herein identified as HCoV-SA1. In particular the nucleic acid and/or amino acid sequences of the HCoV-SA1 genome and sequences encoding (parts of) viral proteins are provided. Further, the invention relates to diagnostic means and methods, prophylactic means and methods and therapeutic means and methods to be employed in the diagnosis, prevention and/or treatment of disease, in particular of respiratory disease and/or renal failure (atypical pneumonia), in particular of mammals, more in particular in humans.

[0048] In particular diagnostic tests for example useful in PCR and serology with nucleic acids (primers) and antibodies and other reagents that are specifically targeted at the nucleic acid or amino acid sequences of the HCoV-SA1 genome are herein provided. The invention also provides vectors, such as bacterial and viral vectors based on nucleic acid or amino acid sequences of the HCoV-SA1 genome. In addition, the invention also provides antigenic polypeptides based amino acid sequences of the HCoV-SA1 genome are herein provided.

[0049] Also, the invention provides vaccines against HCoV-SA1 (based on nucleic acid or amino acid sequences or antigenic polypeptides of the HCoV-SA1 genome, and the invention provides use of antiviral drugs directed against nucleic acid or amino acid sequences or polypeptides of the HCoV-SA1 genome, as herein provided.

[0050] As for yet it is not known if there is a cure for the disease. Several antiviral therapies have been applied, but with various results. Also, for being able to prevent spread of the disease, it is of great importance to be able to recognize the disease in an early stage. Only then sufficient measures can be taken to isolate patients and initiate quarantine precautions. At this moment there is not yet a diagnostic tool in place. Thus, there is great need in developing diagnostic tools and therapies for this disease.

[0051] As further described in the detailed description herein, the isolated essentially mammalian positive-sense single stranded RNA virus classifiable as belonging to the Order: Nidovirales; Family: Coronaviridae; Subfamily: Coronavirinae; Genus: Betacoronavirus; and non-Lineage A, non-Lineage B or non-Lineage D, human betacoronavirus here provided was isolated from a patient with pneumonia. The virus was isolated from sputum of a male patient aged 60 years old presenting with pneumonia associated with acute renal failure. The virus grows readily on Vero cells and LLC-MK2 cells producing CPE in the form of rounding and syncytia formation. It was classified as an isolated essentially mammalian positive-sense single stranded RNA virus classifiable as belonging to the Order: Nidovirales; Family: Coronaviridae; Subfamily: Coronavirinae; Genus: Betacoronavirus; Lineage C human betacoronavirus by comparison of its RNA sequences. It is remarkable that now, at about 9 years after the isolation of the SARS-virus (also related to bat coronavirus) another betacoronavirus has been isolated from humans.

[0052] The invention also provides an essentially mammalian positive-sense single stranded RNA virus, which is q betacoronavirus, comprising all of the amino acid sequences selected from figure 5 file N.rtf depicting the nucleocapsid (N) protein, figure 6 file M.rtf depicting the matrix (M) protein, figure 7 file E.rtf depicting the small envelope (E) protein, figure 8 file NS3d.rtf depicting the non-structural gene NS3d, figure 9 file NS3c.rtf depicting the non-structural gene NS3c, figure 10 file NS3b.rtf depicting the non-structural gene NS3b, figure 11 file NS3a.rtf depicting the non-structural gene NS3a, figure 12 file S.rtf depicting the spike surface glycoprotein (S), figure 13 file Orf1ab.rtf encoding many enzymatic products among which the replicase, or comprising the nucleic acid sequence of figure 14 file HCoV-SA1.rtf depicting isolate HCoV-SA1.

[0053] In another e,bodiment;, the invention provides essentially mammalian positive-sense single stranded RNA virus which is a betacoronavirus, and identifiable as phylogenetically corresponding thereto by determining the amino acid sequence of the conserved replicase domain of said virus to have at least 90% identity with the Orf1AB amino acid sequence as depicted in Fig. 13..

[0054] The description also provides an isolated positive-sense single stranded RNA virus belonging to the Coronaviruses, genus Betacoronavirus and identifiable as phylogenetically corresponding thereto by determining a nucleic acid

or amino acid sequence of said virus and testing it in phylogenetic tree analyses wherein maximum likelihood trees are generated, preferably with 100 bootstraps and 3 jumbles, and finding it to be more closely phylogenetically corresponding to a virus isolate or nucleic acid having the sequences as depicted in any of the oligonucleotide or amino acid sequences submitted to GENBANK under accession JX869059 (http://www.ncbi.nlm.nih.gov/nuccore/JX869059) than it is corresponding to any of the oligonucleotide or amino acid sequences of bat coronavirus virus HKU4 or HKU5.

[0055] Although phylogenetic analyses provide a convenient method of identifying a virus as a Betacoronavirus; Lineage C virus several other possibly more straightforward albeit somewhat more coarse methods for identifying said virus or viral proteins or nucleic acids from said virus are herein also provided. As a rule of thumb a Betacoronavirus; Lineage C virus can be identified by the percentages of homology of the virus, proteins or nucleic acids to be identified in comparison with viral proteins or nucleic acids identified herein in or in Genbank accession JX869059by sequence. It is generally known that virus species, especially RNA virus species, often constitute a quasi species wherein a cluster of said viruses displays heterogeneity among its members. Thus it is expected that each isolate may have a somewhat different percentage relationship with the sequences of the isolate as provided herein.

[0056] The description in particular provides an isolated positive-sense single stranded RNA virus belonging to the Coronaviruses and identifiable as phylogenetically corresponding thereto by determining a nucleic acid sequence or amino acid sequence of said virus and testing it in phylogenetic tree analyses wherein maximum likelihood trees are generated and finding it to be more closely phylogenetically corresponding to a virus isolate or nucleic acid having the sequences as depicted in any of the oligonucleotide or amino acid sequences submitted to GENBANK under accession JX869059 (http://www.ncbi.nlm.nih.gov/nuccore/JX869059) than it is corresponding to any of the oligonucleotide or amino acid sequences of human coronavirus virus isolate HCoV-HKU1 or HCoV-OC43 or SARS-CoV.

[0057] The description in particular provides an isolated positive-sense single stranded RNA virus belonging to the Coronaviruses and identifiable as phylogenetically corresponding thereto by determining a nucleic acid sequence or amino acid sequence of said virus and testing it in phylogenetic tree analyses wherein maximum likelihood trees are generated and finding it to be more closely phylogenetically corresponding to a virus isolate isolatable from humans having the sequences as depicted in any of figures 3 or 5 to 15 than it is corresponding to a human coronavirus virus isolate HCoV-HKU1 or HCoV-OC43 or SARS-CoV.

[0058] The invention also provides a virus according to the invention wherein its positive-sense single stranded RNA nucleic acid sequence comprises an open reading frame (ORF) encoding a viral protein of said virus, preferably selected from the group of ORFs encoding the spike surface glycoprotein (S), the non-structural genes NS3a, NS3b, NS3c, NS3d, the small envelope (E) protein, the matrix (M) protein, and the nucleocapsid (N) protein. With the provision of the sequence information of this MERS-CoV, the invention provides diagnostic means and methods, prophylactic means and methods and therapeutic means and methods to be employed in the diagnosis, prevention and/or treatment of disease, in particular of respiratory disease (atypical pneumonia), in particular of mammals, more in particular in humans. In virology, it is most advisory that diagnosis, prophylaxis and/or treatment of a specific viral infection is performed with reagents that are most specific for said specific virus causing said infection. In this case this means that it is preferred that said diagnosis, prophylaxis and/or treatment of a Betacoronavirus; Lineage C virus infection is performed with reagents that are most specific for Betacoronavirus; Lineage C virus. This by no means however excludes the possibilities that less specific, but sufficiently cross-reactive reagents are used instead, for example because they are more easily available and sufficiently address the task at hand. The invention for example provides a method for virologically diagnosing a MERS CoV infection of an animal, in particular of a mammal, more in particular of a human being, comprising determining in a sample of said animal the presence of a viral isolate or component thereof by reacting said sample with a MERS CoV specific nucleic acid or antibody according to the invention, and a method for serologically diagnosing a MERS CoV infection of a mammal comprising determining in a sample of said mammal the presence of an antibody specifically directed against a MERS CoV virus or component thereof by reacting said sample with a MERS-CoV-specific proteinaceous molecule or fragment thereof or an antigen according to the invention. The invention also provides a diagnostic kit or other system for diagnosing a MERS CoV infection comprising a MERS-CoV-specific nucleic acid, proteinaceous molecule or fragment thereof, antigen and/or an antibody according to the invention, and preferably a means for detecting said MERS-CoV-specific nucleic acid, proteinaceous molecule or fragment thereof, antigen and/or an antibody, said means for example comprising an excitable group such as a fluorophore or enzymatic detection system used in the art (examples of suitable diagnostic kit format comprise IF, ELISA, neutralization assay, RT-PCR assay). To determine whether an as yet unidentified virus component or synthetic analogue thereof such as nucleic acid, proteinaceous molecule or fragment thereof can be identified as Betacoronavirus; Lineage C -MERS-CoV-specific, it suffices to analyze the nucleic acid or amino acid sequence of said component, for example for a stretch of said nucleic acid or amino acid, preferably of at least 10, more preferably at least 25, more preferably at least 40 nucleotides or amino acids (respectively), by sequence homology comparison with the provided Betacoronavirus; Lineage C viral sequences and with known non-Betacoronavirus; Lineage C viral sequences (SARS is preferably used) using for example phylogenetic analyses as provided herein. Depending on the degree of relationship with said Betacoronavirus; Lineage C or non- Betacoronavirus; Lineage C viral sequences, (herein also called HCoV-SA1 virus-like virus sequences) the component or synthetic ana-

logue can be identified. The description also provides a virus according to the invention that is isolatable from a human with atypical pneumonia. Also, isolated or recombinant nucleic acid or MERS-CoV-specific fragments thereof are obtainable, derived or obtained from a virus according to the invention, as are a vector comprising a nucleic acid according to the invention, and a host cell comprising a nucleic acid or vector according to the invention.

[0059] The description also provides an isolated or recombinant proteinaceous molecule or MERS-CoV-specific fragment thereof encoded by a nucleic acid according to the invention. In a preferred embodiment, the description provides a proteinaceous molecule or MERS-CoV-specific viral protein or fragment thereof encoded by a nucleic acid according to the invention. Useful proteinaceous molecules are for example derived from any of the genes or genomic fragments derivable from a virus according to the invention. Such molecules, or antigenic fragments thereof, as provided herein, are for example useful in diagnostic methods or kits and in pharmaceutical compositions such as sub-unit vaccines and inhibitory peptides. Particularly useful is the viral polymerase protein, the spike protein, the nucleocapsid or antigenic fragments thereof for inclusion as antigen or subunit immunogen, but inactivated whole virus can also be used. Particularly useful are also those proteinaceous substances that are encoded by recombinant nucleic acid fragments that are identified for phylogenetic analyses, of course preferred are those that are within the preferred bounds and metes of ORFs useful in phylogenetic analyses, in particular for eliciting HCoV-SA1 virus-like virus specific antibodies, whether in vivo (e.g. for protective purposes or for providing diagnostic antibodies) or in vitro (e.g. by phage display technology or another technique useful for generating synthetic antibodies). Similarly, the invention provides an antigen comprising a proteinaceous molecule or MERS-CoV-specific fragment thereof according to the invention, or reactive with an antibody according to the invention.

[0060] Also provided herein are antibodies, be it natural polyclonal or monoclonal, or synthetic (e.g. (phage) library-derived binding molecules) antibodies that specifically react with an antigen comprising a proteinaceous molecule or HCoV-virus-like MERS-CoV-specific fragment thereof according to the invention. A person skilled in the art will be able to develop (monoclonal) antibodies using isolated virus material and/or recombinantly expressed viral proteins. Sui et al. (Proc. Natl. Acad. Sci. 101(8), 2536-2541, 2004) have transiently expressed fragments of the spike protein and found several antibodies through phage display methods. Such antibodies are also useful in a method for identifying a viral isolate as a HCoV-SA1 virus-like virus comprising reacting said viral isolate or a component thereof with an antibody as provided herein. This can for example be achieved by using purified or non-purified HCoV-SA1 virus-like virus or parts thereof (proteins, peptides) using ELISA, RIA, FACS or similar formats of antigen detection assays (Current Protocols in Immunology). Alternatively, infected cells or cell cultures may be used to identify viral antigens using classical immunofluorescence or immunohistochemical techniques. Specifically useful in this respect are antibodies raised against HCoV-SA1 virus-like virus proteins which are encoded by a nucleotide sequence comprising one or more of the fragments disclosed herein.

[0061] The invention also provides method for identifying a viral isolate as a MERS CoV comprising reacting said viral isolate or a component thereof with a nucleic acid according to the invention. Other methods for identifying a viral isolate as a HCOV-SA1 virus or MERS-CoV comprise reacting said viral isolate or a component thereof with a virus specific nucleic acid according to the invention

[0062] In this way the invention provides a viral isolate identifiable with a method according to the invention as a mammalian virus taxonomically corresponding to a positive-sense single stranded RNA virus identifiable as likely belonging to the HCOV-SA1 or MERS-CoV virus genus within the family of Coronaviruses.

[0063] The method is useful in a method for virologically diagnosing a HCOV-SA1 or MERS-CoV virus infection of a mammal, said method for example comprising determining in a sample of said mammal the presence of a viral isolate or component thereof by reacting said sample with a nucleic acid or an antibody according to the invention.

[0064] Methods of the invention can in principle be performed by using any nucleic acid amplification method, such as the. Polymerase Chain Reaction (PCR; Mullis 1987, U.S. Pat. No. 4,683,195, 4,683,202, en 4,800,159) or by using amplification reactions such as Ligase Chain Reaction (LCR; Barany 1991, Proc. Natl. Acad. Sci. USA 88:189-193; EP Appl. No., 320,308), Self-Sustained Sequence Replication (3SR; Guatelli et al., 1990, Proc. Natl. Acad. Sci. USA 87:1874-1878), Strand Displacement Amplification (SDA; U.S. Pat. Nos. 5,270,184, en 5,455,166), Transcriptional Amplification System (TAS; Kwoh et al., Proc. Natl. Acad. Sci. USA 86:1173-1177), Q-Beta Replicase (Lizardi et al., 1988, Bio/Technology 6:1197), Rolling Circle Amplification (RCA; U.S. Pat. No. 5,871,921), Nucleic Acid Sequence Based Amplification (NASBA), Cleavage Fragment Length Polymorphism (U.S. Pat. No. 5,719,028), Isothermal and Chimeric Primer-initiated Amplification of Nucleic Acid (ICAN), Ramification-extension Amplification Method (RAM; U.S. Pat. Nos. 5,719,028 and 5,942,391) or other suitable methods for amplification of nucleic acids.

[0065] In order to amplify a nucleic acid with a small number of mismatches to one or more of the amplification primers, an amplification reaction may be performed under conditions of reduced stringency (e.g. a PCR amplification using an annealing temperature of 38.degree. C., or the presence of 3.5 mM MgCl2). The person skilled in the art will be able to select conditions of suitable stringency.

[0066] The primers herein are selected to be "substantially" complementary (i.e. at least 65%, more preferably at least 80% perfectly complementary) to their target regions present on the different strands of each specific sequence to be

amplified. It is possible to use primer sequences containing e.g. inositol residues or ambiguous bases or even primers that contain one or more mismatches when compared to the target sequence. In general, sequences that exhibit at least 65%, more preferably at least 80% homology with the target DNA or RNA oligonucleotide sequences are considered suitable for use in a method of the present invention. Sequence mismatches are also not critical when using low stringency hybridization conditions.

[0067] The detection of the amplification products can in principle be accomplished by any suitable method known in the art. The detection fragments may be directly stained or labeled with radioactive labels, antibodies, luminescent dyes, fluorescent dyes, or enzyme reagents. Direct DNA stains include for example intercalating dyes such as acridine orange, ethidium bromide, ethidium monoazide or Hoechst dyes.

[0068] Alternatively, the DNA or RNA fragments may be detected by incorporation of labeled dNTP bases into the synthesized fragments. Detection labels which may be associated with nucleotide bases include e.g. fluorescein, cyanine dye or BrdUrd. When using a probe-based detection system, a suitable detection procedure for use in the present invention may for example comprise an enzyme immunoassay (EIA) format (Jacobs et al., 1997, J. Clin. Microbiol. 35, 791-795). For performing a detection by manner of the EIA procedure, either the forward or the reverse primer used in the amplification reaction may comprise a capturing group, such as a biotin group for immobilization of target DNA PCR amplicons on e.g. a streptavidin coated microtiter plate wells for subsequent EIA detection of target DNA-amplicons (see below). The skilled person will understand that other groups for immobilization of target DNA PCR amplicons in an EIA format may be employed.

[0069] Probes useful for the detection of the target DNA as disclosed herein preferably bind only to at least a part of the DNA sequence region as amplified by the DNA amplification procedure. Those of skill in the art can prepare suitable probes for detection based on the nucleotide sequence of the target DNA without undue experimentation as set out herein. Also the complementary nucleotide sequences, whether DNA or RNA or chemically synthesized analogs, of the target DNA may suitably be used as type-specific detection probes in a method of the invention, provided that such a complementary strand is amplified in the amplification reaction employed.

[0070] Suitable detection procedures for use herein may for example comprise immobilization of the amplicons and probing the DNA sequences thereof by e.g. southern blotting. Other formats may comprise an EIA format as described above. To facilitate the detection of binding, the specific amplicon detection probes may comprise a label moiety such as a fluorophore, a chromophore, an enzyme or a radio-label, so as to facilitate monitoring of binding of the probes to the reaction product of the amplification reaction. Such labels are well-known to those skilled in the art and include, for example, fluorescein isothiocyanate (FITC), beta-galactosidase, horseradish peroxidase, streptavidin, biotin, digoxigenin, 35S or 1251. Other examples will be apparent to those skilled in the art.

[0071] Detection may also be performed by a so called reverse line blot (RLB) assay, such as for instance described by Van den Brule et al. (2002, J. Clin. Microbiol. 40, 779-787). For this purpose RLB probes are preferably synthesized with a 5' amino group for subsequent immobilization on e.g. carboxyl-coated nylon membranes. The advantage of an RLB format is the ease of the system and its speed, thus allowing for high throughput sample processing.

[0072] The use of nucleic acid probes for the detection of RNA or DNA fragments is well known in the art. Mostly these procedures comprise the hybridization of the target nucleic acid with the probe followed by post-hybridization washings. Specificity is typically the function of post-hybridization washes, the critical factors being the ionic strength and temperature of the final wash solution. For nucleic acid hybrids, the Tm can be approximated from the equation of Meinkoth and Wahl, Anal. Biochem., 138: 267-284 (1984): Tm=81.5.degree. C.+16.6 (log M)+0.41 (% GC)-0.61 (% form)-500/L; where M is the molarity of monovalent cations, % GC is the percentage of guanosine and cytosine nucleotides in the nucleic acid, % form is the percentage of formamide in the hybridization solution, and L is the length of the hybrid in base pairs. The Tm is the temperature (under defined ionic strength and pH) at which 50% of a complementary target sequence hybridizes to a perfectly matched probe. Tm is reduced by about 1.degree. C. for each 1% of mismatching; thus, the hybridization and/or wash conditions can be adjusted to hybridize to sequences of the desired identity. For example, if sequences with >90% identity are sought, the Tm can be decreased 10.degree. C. Generally, stringent conditions are selected to be about 5.degree. C. lower than the thermal melting point (Tm) for the specific sequence and its complement at a defined ionic strength and pH. However, severely stringent conditions can utilize hybridization and/or wash at 1, 2, 3, or 4.degree. C. lower than the thermal melting point (Tm); moderately stringent conditions can utilize a hybridization and/or wash at 6, 7, 8, 9, or 10.degree. C. lower than the thermal melting point (Tm); low stringency conditions can utilize a hybridization and/or wash at 11, 12, 13, 14, 15, or 20.degree. C. lower than the thermal melting point (Tm). Using the equation, hybridization and wash compositions, and desired Tm, those of ordinary skill will understand that variations in the stringency of hybridization and/or wash solutions are inherently described. If the desired degree of mismatching results in a Tm of less than 45.degree. C. (aqueous solution) or 32.degree. C. (formamide solution) it is preferred to increase the SSC concentration so that a higher temperature can be used. An extensive guide to the hybridization of nucleic acids is found in Tijssen, Laboratory Techniques in Biochemistm and Molecular Biology--Hybridization with Nucleic Acid Probes, Part I, Chapter 2 "Overview of principles of hybridization and the strategy of nucleic acid probe assays", Elsevier. New York (1993); and Current Protocols in Molecular Biology, Chapter 2, Ausubel, et al., Eds., Greene

Publishing and Wiley-Interscience, New York (1995).

[0073] In another aspect, the description provides oligonucleotide probes for the generic detection of target RNA or DNA. The detection probes herein are selected to be "substantially" complementary to one of the strands of the double stranded nucleic acids generated by an amplification reaction of the invention. Preferably the probes are substantially complementary to the immobilizable, e.g. biotin labelled, antisense strands of the amplicons generated from the target RNA or DNA.

[0074] It is allowable for detection probes to contain one or more mismatches to their target sequence. In general, sequences that exhibit at least 65%, more preferably at least 80% homology with the target oligonucleotide sequences are considered suitable for use in a method of the present invention. Antibodies, both monoclonal and polyclonal, can also be used for detection purpose in the present invention, for example, in immunoassays in which they can be utilized in liquid phase or bound to a solid phase carrier. In addition, the monoclonal antibodies in these immunoassays can be detectably labeled in various ways. A variety of immunoassay formats may be used to select antibodies specifically reactive with a particular protein (or other analyte). For example, solid-phase ELISA immunoassays are routinely used to select monoclonal antibodies specifically immunoreactive with a protein. See Harlow and Lane, Antibodies, A Laboratory Manual, Cold Spring Harbor Publications, New York (1988), for a description of immunoassay formats and conditions that can be used to determine selective binding. Examples of types of immunoassays that can utilize antibodies of the invention are competitive and non-competitive immunoassays in either a direct or indirect format. Examples of such immunoassays are the radioimmunoassay (RIA) and the sandwich (immunometric) assay. Detection of the antigens using the antibodies of the invention can be done utilizing immunoassays that are run in either the forward, reverse, or simultaneous modes, including immunohistochemical assays on physiological samples. Those of skill in the art will know, or can readily discern, other immunoassay formats Antibodies can be bound to many different carriers and used to detect the presence of the target molecules. Examples of well-known carriers include glass, polystyrene, polypropylene, polyethylene, dextran, nylon, amylases, natural and modified celluloses, polyacrylamides, agaroses and magnetite. The nature of the carrier can be either soluble or insoluble for purposes of the invention. Those skilled in the art will know of other suitable carriers for binding monoclonal antibodies, or will be able to ascertain such using routine experimentation.

[0075] The description also provides a method for serologically diagnosing a MERS-CoV infection of a mammal comprising determining in a sample of said mammal the presence of an antibody specifically directed against a MERS-CoV or component thereof by reacting said sample with a proteinaceous molecule or fragment thereof or an antigen according to the invention.

[0076] Methods and means provided herein are particularly useful in a diagnostic kit for diagnosing a MERS-CoV infection, be it by virological or serological diagnosis. Such kits or assays may for example comprise a virus, a nucleic acid, a proteinaceous molecule or fragment thereof, an antigen and/or an antibody according to the invention.

[0077] Herewith, the invention provides a method for virologically diagnosing a MERS CoV infection of a mammal comprising determining in a sample of said mammal the presence of a viral isolate or component thereof by reacting said sample with a nucleic acid according to the invention or an antibody according to the invention or determining in a sample of said mammal the presence of an antibody specifically directed against MERS CoV virus or component thereof by reacting said sample with a proteinaceous molecule or fragment thereof according to the invention or an antigen according to the invention.

[0078] The invention also provides diagnostic kit for diagnosing a MERS CoV infection comprising a virus according to the invention, a nucleic acid according to the invention, a proteinaceous molecule or fragment thereof according to the invention, an antigen according to the invention and/or an antibody according to the invention.

[0079] The description also provides use of a MERS-CoV according to the invention, a nucleic acid according to the invention, a vector according to the invention, a host cell according to the invention, a proteinaceous molecule or fragment thereof according to the invention, an antigen according to the invention, or an antibody according to the invention for the production of a pharmaceutical composition, preferably for the production of a pharmaceutical composition for the treatment or prevention of a Betacoronavirus, Lineage C virus infection, preferably a human infection, or for the production of a pharmaceutical composition for the treatment or prevention of atypical pneumonia and/or renal failure, preferably wherein said atypical pneumonia and/or renal failure is a human disease.

[0080] The invention also provides pharmaceutical composition comprising a virus according to the invention, a nucleic acid according to the invention, a vector according to the invention, a host cell according to the invention, a proteinaceous molecule or fragment thereof according to the invention, an antigen according to the invention, or an antibody according to the invention.

[0081] A pharmaceutical composition comprising a virus, a nucleic acid, a proteinaceous molecule or fragment thereof, an antigen and/or an antibody according to the invention can for example be used in a method for the treatment or prevention of a MERS-CoV infection comprising providing an individual with a pharmaceutical composition according to the invention. This is most useful when said individual comprises a human. Antibodies against MERS-CoV proteins, especially against the spike protein of MERS-CoV, preferably against the amino acid sequence as depicted herein are also useful for prophylactic or therapeutic purposes, as passive vaccines. It is known from other coronaviruses that the

spike protein is a very strong antigen and that antibodies against spike protein can be used in prophylactic and therapeutic vaccination.

[0082] The description also provides method to obtain a modulator or an antiviral agent useful in the treatment of atypical pneumonia comprising establishing a cell culture or experimental animal comprising a virus according to the invention, treating said culture or animal with a candidate antiviral agent, and determining the effect of said modulator or agent on said virus or its infection of said culture or animal. An example of such an antiviral agent comprises a MERS-CoV virus-neutralizing antibody, or functional component thereof, as provided herein, but antiviral agents of other nature, such as ADA or adenosine are obtained as well. The description also provides use of a modulator or an antiviral agent for the preparation of a pharmaceutical composition, in particular for the preparation of a pharmaceutical composition for the treatment of atypical pneumonia, specifically when caused by a MERS-CoV infection, and provides a pharmaceutical composition comprising an antiviral agent, useful in a method for the treatment or prevention of a MERS-CoV infection or atypical pneumonia, said method comprising providing an individual with such a pharmaceutical composition.

[0083] The invention also provides a method for the treatment or prevention of a MERS CoV infection comprising providing an individual, preferably a human individual with a pharmaceutical composition according to the invention. In particular individual MERS-CoV virus-like polypeptide are provided herein as well, such as the viral replicase encoded by an RNA or DNA or cDNA sequence, as depicted in figure 13. A viral spike protein encoded by an RNA or DNA or cDNA sequence or fragments thereof, as depicted in figure 12, a viral non-structural gene protein encoded by an RNA or DNA or cDNA sequence as depicted in any of figures 8, 9, 10 or 11, a small envelope (E) protein encoded by an RNA or DNA or cDNA sequence as depicted in figure 7, a matrix (M) protein encoded by an RNA or DNA or cDNA sequence depicted in figure 6, a nucleocapsid (N) protein encoded by an RNA or DNA or cDNA sequence as depicted in figure 5, a nucleic acid sequence which comprises one or more of the sequences of HCoV-SA1 as depicted in figures 3, or 5 to 15.

[0084] With the provision of the sequence information of this MERS virus, MERS-CoV, the invention provides diagnostic means and methods, prophylactic means and methods and therapeutic means and methods to be employed in the diagnosis, prevention and/or treatment of disease, in particular of respiratory disease (atypical pneumonia), in particular of mammals, more in particular in humans. In virology, it is most advisory that diagnosis, prophylaxis and/or treatment of a specific viral infection is performed with reagents that are most specific for said specific virus causing said infection. In this case this means that it is preferred that said diagnosis, prophylaxis and/or treatment of a MERS virus infection is performed with reagents that are most specific for MERS virus. This by no means however excludes the possibilities that less specific, but sufficiently cross-reactive reagents are used instead, for example because they are more easily available and sufficiently address the task at hand.

[0085] The invention for example provides a method for virologically diagnosing a MERS infection of an animal, in particular of a mammal, more in particular of a human being, comprising determining in a sample of said animal the presence of a viral isolate or component thereof by reacting said sample with a MERS specific nucleic acid or antibody according to the invention, and a method for serologically diagnosing a MERS infection of a mammal comprising determining in a sample of said mammal the presence of an antibody specifically directed against a MERS virus or component thereof by reacting said sample with a MERS-CoV-specific proteinaceous molecule or fragment thereof or an antigen according to the invention. Suitable MERS-CoV specific nucleic acid for example is provided herein as well, such as the RNA or DNA or cDNA sequence or fragments or homologues thereof, as depicted in figure 13, or as depicted in figure 12, or in any of figures 8, 9, 10 or 11, in figure 7, in figure 6, in figure 5, a nucleic acid sequence which comprises one or more of the sequences of HCoV-SA1 or a MERS-CoV specific nucleic acid sequence which can hybridize with sequences in any of figures as depicted in figures 3, or 5 to 15 under stringent conditions, or a MERS-CoV specific nucleic acid sequence, such as an RNA or a DNA or preferably a cDNA, which has at least 65%, preferably at least 75%, more preferably at least 85%, most preferably at least 95% homology or are substantially, at least 65%, preferably at least 75%, more preferably at least 85%, most preferably at least 95%, complementary with a nucleotide sequence as depicted in figures 3, or 5 to 15. For MERS CoV nucleic acid diagnosis, short nucleotide stretches of 10 to 40, preferably 12 to 30, more preferably 15 to 25 nucleotides long, commonly called "primers" are provided herein that preferably are MERS-CoV specific or at least substantially complementary to MERS virus nucleic acid as depicted in figures 3, or 5-15 and have stretches of at least 10, preferably at least 12, more preferably at least 15, most preferably at least 18 or 19 nucleotides that are 100% complementary to at least a fragment of a nucleotide sequence as depicted in figures 3, or 5 to 15. The term "nucleotide sequence homology" as used herein denotes the presence of homology between two (poly) nucleotides, such as a RNA or a DNA or a cDNA sequence. Polynucleotides have "homologous" sequences if the sequence of nucleotides in the two sequences is the same when aligned for maximum correspondence. Sequence comparison between two or more polynucleotides is generally performed by comparing portions of the two sequences over a comparison window to identify and compare local regions of sequence similarity. The comparison window is generally from about 20 to 200 contiguous nucleotides. The "percentage of sequence homology" for polynucleotides, such as 50, 60, 70, 80, 90, 95, 98, 99 or 100 percent sequence homology may be determined by comparing two optimally aligned sequences over a comparison window, wherein the portion of the polynucleotide sequence in the comparison window may include additions or deletions (i.e. gaps) as compared to the reference sequence (which does

not comprise additions or deletions) for optimal alignment of the two sequences. Nucleotide or base G is homologous to G, C is homologous to C, A is homologous to A and nucleotides T or U are homologous to T or U, to calculate overall homology or complementarities between DNA and RNA. The percentage is calculated by: (a) determining the number of positions at which the identical nucleic acid base occurs in both sequences to yield the number of matched positions; (b) dividing the number of matched positions by the total number of positions in the window of comparison; and (c) multiplying the result by 100 to yield the percentage of sequence homology. Optimal alignment of sequences for comparison may be conducted by computerized implementations of known algorithms, or by inspection. Readily available sequence comparison and multiple sequence alignment algorithms are, respectively, the Basic Local Alignment Search Tool (BLAST) (Altschul, S. F. et al. 1990. J. Mol. Biol. 215:403; Altschul, S. F. et al. 1997. Nucleic Acid Res. 25:3389-3402) and ClustalW programs both available on the internet. Other suitable programs include GAP, BESTFIT and FASTA in the Wisconsin Genetics Software Package (Genetics Computer Group (GCG), Madison, Wis., USA).

[0086] As used herein, "substantially complementary" means that two nucleic acid sequences have at least about 65%, preferably about 70%, more preferably about 80%, even more preferably 90%, and most preferably about 98%, sequence complementarities to each other. This means that the primers and probes must exhibit sufficient complementarity to their template and target nucleic acid, respectively, to hybridize under stringent conditions. Therefore, the primer sequences as disclosed in this specification need not reflect the exact sequence of the binding region on the template and degenerate primers can be used. A substantially complementary primer sequence is one that has sufficient sequence complementarity to the amplification template to result in primer binding and second-strand synthesis.

[0087] The term "hybrid" refers to a double-stranded nucleic acid molecule, or duplex, formed by hydrogen bonding between complementary nucleotides. The terms "hybridize" or "anneal" refer to the process by which single strands of nucleic acid sequences form double-helical segments through hydrogen bonding between complementary nucleotides, according to a strict rule called base-pairing defined by the complementary structures of the nucleotides or bases (b). Typically, in two nucleic acid strands, nucleotide guanine (G) is complementary to nucleotide cytosine (C), G and C pair wise capable of forming three hydrogen bonds, and nucleotide adenine (A) is complementary to nucleotides thymine (T) or uracil (U), A and T or A and U pair wise capable of forming two hydrogen bonds, thus G pairs with C and A pairs with T or U. Conventionally, in depicting a nucleic acid sequence, T is commonly identified as uracil (U) to identify RNA (ribonucleic acid), and as thymine (T) when identifying DNA (deoxyribonucleic acid) or cDNA (complementary or copy DNA). A DNA polymerase is a cellular or viral polymerase enzyme that synthesizes DNA molecules from their nucleotide building blocks. DNA polymerases are essential for DNA replication, and usually function in pairs while copying one double-stranded DNA molecule into two double-stranded DNAs in a process termed DNA replication. RNA viruses commonly use an RNA-dependent RNA-polymerase to replicate their RNA. DNA can be used to produce RNA by the actions of a transcriptase; RNA can be used to produce DNA or cDNA by the actions of a reverse transcriptase. A transcriptase is a polymerase that catalyzes the formation of RNA from a DNA template in the process of transcription. Reverse transcriptase (RT) is a polymerase enzyme used to generate complementary DNA (cDNA) from an RNA template, a process termed reverse transcription.

[0088] The term "oligonucleotide" refers to a short sequence of nucleotide monomers (usually 6 to 100 nucleotides) joined by phosphorous linkages (e.g., phosphodiester, alkyl and aryl-phosphate, phosphorothioate), or non-phosphorous linkages (e.g., peptide, sulfamate and others). An oligonucleotide may contain modified nucleotides having modified bases (e.g., 5-methyl cytosine) and modified sugar groups (e.g., 2'-O-methyl ribosyl 2'-O-methoxyethyl ribosyl, 2'-fluoro ribosyl, 2'-amino ribosyl, and the like). Oligonucleotides may be naturally-occurring or synthetic molecules of double- and single-stranded DNA and double- and single-stranded RNA with circular, branched or linear shapes and optionally including domains capable of forming stable secondary structures (e.g., stem-and-loop and loop-stem-loop structures).

[0089] The term "primer" as used herein also refers to an oligonucleotide which is capable of annealing to the amplification target allowing a DNA polymerase to attach thereby serving as a point of initiation of DNA synthesis when placed under conditions in which synthesis of primer extension product which is complementary to a nucleic acid strand is induced, i.e., in the presence of nucleotides and an agent for polymerization such as DNA polymerase and at a suitable temperature and pH. The (amplification) primer is preferably single stranded for maximum efficiency in amplification. Preferably, the primer is an oligodeoxyribonucleotide. The primer must be sufficiently long to prime the synthesis of extension products in the presence of the agent for polymerization. The exact lengths of the primers will depend on many factors, including temperature and source of primer. A "pair of bi-directional primers" as used herein refers to one forward and one reverse primer as commonly used in the art of RNA or DNA amplification such as in PCR amplification.

[0090] The term "probe" refers to a single-stranded oligonucleotide sequence that will recognize and form a hydrogen-bonded duplex with a complementary sequence in a target nucleic acid sequence analyte or its cDNA derivative.

[0091] The terms "stringency" or "stringent hybridization conditions" refer to hybridization conditions that affect the stability of hybrids, e.g., temperature, salt concentration, pH, formamide concentration and the like. These conditions are empirically optimized to maximize specific binding and minimize non-specific binding of primer or probe to its target nucleic acid sequence. The terms as used include reference to conditions under which a probe or primer will hybridize to its target sequence, to a detectably greater degree than other sequences (e.g. at least 2-fold over background).

Stringent conditions are sequence dependent and will be different in different circumstances. Longer sequences hybridize specifically at higher temperatures. Generally, stringent conditions are selected to be about 5. degree. C. lower than the thermal melting point (Tm) for the specific sequence at a defined ionic strength and pH. The Tm is the temperature (under defined ionic strength and pH) at which 50% of a complementary target sequence hybridizes to a perfectly matched probe or primer. Typically, stringent conditions will be those in which the salt concentration is less than about 1.0 M Na+ ion, typically about 0.01 to 1.0 M Na+ ion concentration (or other salts) at pH 7.0 to 8.3 and the temperature is at least about 30.degree. C. for short probes or primers (e.g. 10 to 50 nucleotides) and at least about 60.degree. C. for long probes or primers (e.g. greater than 50 nucleotides). Stringent conditions may also be achieved with the addition of destabilizing agents such as formamide. Exemplary low stringent conditions or "conditions of reduced stringency" include hybridization with a buffer solution of 30% formamide, 1 M NaCl, 1% SDS at 37.degree. C. and a wash in 2.times.SSC at 40.degree. C. Exemplary high stringency conditions include hybridization in 50% formamide, 1 M NaCl, 1% SDS at 37.degree. C., and a wash in 0.1.times.SSC at 60.degree. C. Hybridization procedures are well known in the art and are described in e.g. Ausubel et al, Current Protocols in Molecular Biology, John Wiley & Sons Inc., 1994.

[0092] The invention for example provides a method for virologically diagnosing a MERS infection of a mammal, more in particular of a human being, comprising determining in a sample of said animal the presence of a viral isolate or component thereof by reacting said sample with a MERS specific nucleic acid or antibody according to the invention, and a method for serologically diagnosing a MERS infection of a mammal comprising determining in a sample of said mammal the presence of an antibody specifically directed against a MERS virus or component thereof by reacting said sample with a MERS MERS-CoV-specific proteinaceous molecule or fragment thereof or an antigen according to the invention. Suitable MERS specific proteinaceous molecules or MERS virus specific fragment thereof is provided herein as well, such as the viral replicase encoded by an RNA or DNA or cDNA sequence or fragments or homologues thereof, as depicted in figure 13, a viral spike protein encoded by an RNA or DNA or cDNA sequence or fragments or homologues thereof, as depicted in figure 12, a viral non-structural gene protein encoded by an RNA or DNA or cDNA sequence or fragments or homologues thereof, as depicted in any of figures 8, 9, 10 or 11, a small envelope (E) protein encoded by an RNA or DNA or cDNA sequence or fragments or homologues thereof, as depicted in figure 7, a matrix (M) protein encoded by an RNA pr DNA sequence or fragments or homologues thereof, as depicted in figure 6, a nucleocapsid (N) protein encoded by an RNA or DNA or cDNA sequence or fragments or homologues thereof, as depicted in figure 5, a nucleic acid sequence which comprises one or more of the sequences of HCoV-SA1 as depicted in figures 3, or 5 to 15 or a nucleic acid sequence which can hybridize with any of these sequences under stringent conditions.

[0093] Suitable MERS CoV specific antibodies directed against MERS CoV specific proteinaceous molecules or MERS CoV specific fragment thereof is provided herein as well, such as antibodies raised against a viral replicase encoded by an RNA sequence or fragments or homologues thereof, as depicted in figure 13, raised against a viral spike protein encoded by an RNA or DNA or cDNA sequence or fragments or homologues thereof, as depicted in figure 12, raised against a viral non-structural gene protein encoded by an RNA or DNA or cDNA sequence or fragments or homologues thereof, as depicted in any of figures 8, 9, 10 or 11, raised against a small envelope (E) protein encoded by an RNA or DNA or cDNA sequence or fragments or homologues thereof, as depicted in figure 7, raised against a matrix (M) protein encoded by an RNA or DNA or cDNA sequence or fragments or homologues thereof, as depicted in figure 6, raised against a nucleocapsid (N) protein encoded by an RNA or DNA or cDNA sequence or fragments or homologues thereof, as depicted in figure 5, a nucleic acid sequence which comprises one or more of the sequences of HCoV-SA1 as depicted in figures 3, or 5 to 15 or a nucleic acid sequence which can hybridize with any of these sequences under stringent conditions.

[0094] The term "antibody" includes reference to antigen binding forms of antibodies (e. g., Fab, F(ab)2). The term "antibody" frequently refers to a polypeptide substantially encoded by an immunoglobulin gene or immunoglobulin genes, or fragments thereof which specifically bind and recognize an analyte (antigen). However, while various antibody fragments can be defined in terms of the digestion of an intact antibody, one of skill will appreciate that such fragments may be synthesized de novo either chemically or by utilizing recombinant DNA methodology. Thus, the term antibody, as used herein, also includes antibody fragments such as single chain Fv, chimeric antibodies (i. e., comprising constant and variable regions from different species), humanized antibodies (i. e., comprising a complementarity determining region (CDR) from a non-human source) and heteroconjugate antibodies (e. g., bispecific antibodies).

[0095] The invention also provides a diagnostic kit for diagnosing a MERS-CoV infection comprising a MERS Corona virus, or a MERS-CoV-specific nucleic acid, or a MERS-CoV-specific proteinaceous molecule or fragment thereof, a MERS-CoV-specific antigen and/or an a MERS-CoV-specific antibody according to the invention, and preferably a means for detecting said MERS-CoV, MERS-CoV-specific nucleic acid, said proteinaceous molecule or fragment thereof, said antigen and/or said antibody, said means for example comprising an excitable group such as a fluorophore or enzymatic detection system used in the art (examples of suitable diagnostic kit format comprise IF, ELISA, neutralization assay, RT-PCR assay). To determine whether an as yet unidentified virus component or synthetic analogue thereof such as nucleic acid, proteinaceous molecule or fragment thereof can be identified as MERS-CoV-specific, it suffices to analyze the nucleic acid or amino acid sequence of said component, for example for a stretch of said nucleic acid or amino acid,

preferably of at least 10, more preferably at least 25, more preferably at least 40 nucleotides or amino acids (respectively), by sequence homology comparison with the herein provided MERS viral sequences and with known non-MERS viral sequences (HUK4 or HUK5 are preferably used) using for example phylogenetic analyses as provided herein. Depending on the degree of relationship with said MERS or non-MERS viral sequences, the component or synthetic analogue can be identified.

[0096] The sequence of the first isolate of MERS-CoV is also deposited in Genbank under:

LOCUS JX869059 30119 bp RNA linear VRL 04-DEC-2012
DEFINITION Human betacoronavirus 2c EMC/2012, complete genome.
ACCESSION JX869059
VERSION JX869059.2 GI:409052551
KEYWORDS
SOURCE Human betacoronavirus 2c EMC/2012
ORGANISM Human betacoronavirus 2c EMC/2012

Viruses; ssRNA positive-strand viruses, no DNA stage; Nidovirales; Coronaviridae; Coronavirinae; Betacoronavirus; unclassified Betacoronavirus.

REFERENCE 1 (bases 1 to 30119)

AUTHORS van Boheemen,S., de Graaf,M., Lauber,C., Bestebroer,T.M., Raj,V.S., Zaki,A.M., Osterhaus,A.D., Haagmans,B.L., Gorbalenya,A.E., Snijder,E.J. and Fouchier,R.A. TITLE Genomic characterization of a newly discovered coronavirus associated with acute respiratory distress syndrome in humans JOURNAL MBio 3 (6), e00473-12 (2012)PUBMED 23170002

[0097] The present invention in particular also relates to the spike (S) protein of a Coronavirus that utilizes DPP4 as a virus receptor and fragments thereof as depicted in figures 16, 17 and 32.

[0098] The invention in a further embodiment also provides a proteinaceous substance comprising a protein according to the invention, wherein said protein is a spike protein from figure 12 and additionally comprising at least a fragment of an N-terminal dipeptidyl peptidase protein wherein said fragment is derived from the ectodomain.

[0099] In describing a proteinaceous substance herein, reference is made to protein containing material, such as an organism or a part thereof, microbial organism, virus, tissue, cell, cell culture, cell culture precipitate, cell culture supernatant, cell content such as cytoplasm, nucleoplasm, nuclei, nucleoli, cell organelles, mitochondria, ribosome, tubuli, plasma, blood, serum, lymph, drainage fluid, and to a protein containing preparation, such as a buffer, dilution, precipitate, extraction, pull down sample, test sample, spray, chromatographic sample, or a crystal.

[0100] Surprisingly, in pull down binding experiments with a fragment of a newly discovered coronavirus, it was found that the first isolated fragment, comprising an ectodomain of the spike protein of the virus, bound to at least the ectodomain of a prolyl peptidase, an N terminal dipeptidyl peptidase, the identity of which was confirmed by mass spectrometric analyses of tryptic peptide digests. No binding interaction between an N terminal dipeptidyl peptidase and a viral protein has been found before, in particular, not wherein the peptidase is acting as a receptor for the virus, allowing viral entry and replication in a cell. Blocking DPP4 with specific anti DPP4 antibodies indeed abolished viral infection.

[0101] The description also relates to ten protease families that are unique to higher organisms (16 protease families can be identified in the genomes of all forms of cellular life). Within this core group of ten protease families, a multitude of proteases evolved to yield intra and extra cellular processes. Dipeptidyl peptidase 4 (DPP4; Dipeptidyl Peptidase IV (DPPIV)) is a member of this large family of proteases (peptidases). DPP4 is a serine protease of family S9. DPP4 is a 240 kDa homodimeric, multi functional type II membrane bound glycoprotein, widely distributed in all mammalian tissues, but highly expressed in kidney, liver and endothelium. DPPIV is also known as DPP4, CD26, adenosine deaminase complexing protein 2 or adenosine deaminase binding protein (ADAbp). DPP4 consists of a short cytoplasmic domain of six amino acids, followed by a hydrophobic transmembrane domain (amino acids 7 28) and an extracellular (ectodomain) sequence of 739 amino acids. DPP4 is a highly specific aminopeptidase and releases dipeptides from the amino terminus of peptides with a Pro or Ala in the penultimate position. N terminal degradation of the substrate peptides may result in the activation, inactivation or modulation of their activity. Besides its well known exopeptidase activity, DPPIV also exhibits endopeptidase activity toward denatured collagen. Expression of DPPIV is associated with cell adhesion and is a co stimulant during T cell activation and proliferation.

[0102] DPPIV (DPP4, CD26) is a member of the class of proteases known as prolyl peptidases, which cleave proteins after proline residues. DPPIV, a serine dipeptidyl peptidase, cleaves the N terminal X Ala or X Pro from target polypeptides, such as chemokines (e.g., CXCL11) and peptide hormones (e.g., GLP 1, PACAP, VIP, BNP). DPPIV possesses a

transmembrane region and a very short cytoplasmic domain, but is often cleaved and released as a soluble, circulating fragment. Serine proteases are grouped into 43 families. Protease family S9 is divided into four subfamilies: S9A (type prolyl oligopeptidase), S9B (DPP4), S9C (acylaminoacyl peptidase), and S9D (glutamyl endopeptidase).

[0103] In humans, members of the subfamily S9B include DPP4, fibroblast activation protein alpha (FAPα), dipeptidyl peptidase 8 (DPP8), and dipeptidyl peptidase 9 (DPP9). DPP4 is also known as adenosine deaminase binding protein (ADBP) or T cell activation antigen CD26. DPP4 is a serine exopeptidase that catalyzes the release of an N terminal dipeptide provided that the next to last residue is proline, hydroxyproline, dehydroproline or alanine.

[0104] Only oligopeptides in the trans conformation are able to bind to the active site of DPP4. It also has non peptidase functions: through its interaction with adenosine deaminase (ADA) and extracellular matrix components, it influences T cell activation and proliferation. It is thought to play roles in diabetes, cancer, and autoimmune diseases, making it a target for drug discovery. In particular, cleavage of GLP 1 (7 36) amide, an incretin hormone that stimulates insulin biosynthesis and secretion, into GLP 1 (9 36) amide by DPPIV reverses the glucoregulatory actions of GLP 1. Therefore, DPPIV inhibitors are attractive targets for stimulating insulin production in type II diabetes. Several specific DPPIV inhibitors have been approved by the FDA for type II diabetes.

[0105] Repeating binding experiments with a recombinant, isolated, fragment of the peptidase indeed confirmed the identification of the peptidase as a receptor of the MERS CoV and of the HKU4 CoV, allowing binding of the virus to a mammalian cell (both bat DPP4 as well as human DPP4 were tested), and entry of the MERS CoV leading to abundant replication of that virus in COS7 cells having been provided with the isolated second fragment, whereas COS7 cells not having been provided with the second fragment remain essentially impervious for infection with the virus.

[0106] This description thus provides a proteinaceous substance having been provided with the isolated first probing fragment, preferably a recombinant fragment, of a viral protein and an isolated second fragment, preferably recombinant fragment, of an N terminal dipeptidyl peptidase protein, establishing a probe identified target pair of binding proteins that may be used for binding or affinity studies and preferably also for methods to identify modulators of the interaction of the binding pair.

[0107] In a further embodiment, the description provides a proteinaceous substance comprising, preferably having been provided with, a first fragment of a viral protein and an isolated second, preferably recombinant, fragment of an N terminal dipeptidyl peptidase protein (a fragment obtained by regular peptide synthesis may also be use as first or second fragment). Such a substance provided, in particular, is useful in identifying further binding sites of viral proteins, and fragments thereof, e.g., for narrowing down of specific binding site sequences.

[0108] In a preferred embodiment, the description provides a proteinaceous substance having been provided with at least a fragment of a viral protein, preferably an isolated fragment, wherein the viral protein comprises an ectodomain of a spike protein or of an envelope protein, the ectodomain being the most preferred site for virus cell receptor interaction.

[0109] It is preferred that a substance according to the invention comprises coronaviral protein, , preferably wherein the first fragment is derived from the S1 region of a coronavirus. In a particular embodiment, the first fragment comprises residues 1 747 of the viral spike protein of HCoV EMC 1 as depicted in figure 16.

[0110] The description also provides a substance according to the invention wherein the first fragment comprises, preferably consists of, at least 10, preferably of at least 50, preferably of at least 100 residues derived from the S1 region of a coronavirus. Using smaller fragments from distinct locations in the viral sequence allows for further identifying minimal essential sequences, and thereby narrowing down on the binding site, necessary for binding with the peptidase.

[0111] In particular, a substance according to the invention is provided wherein the first fragment is derived from the S1 region of a coronavirus, for example, comprising residues 1 747 as depicted in 161. Examples of such selected fragments are also found in figure 17, preferably the invention provides a substance with a first fragment consisting of residues 1 357, or of residues 1 588, or of residues 358 588, or of residues 358 747, or of residues 588 747 as depicted in figure 16 or figure 17, or of residues 363 593 of the spike protein of HKU4 CoV as shown in figure 17.

[0112] Even more in particular, a substance according to the invention is provided wherein the first fragment is derived from the S1 region of a coronavirus, which fragment then is subjected to limited proteolysis after which the protease resistant domains are identified by MS, and the interaction between probe and target is studied further.

[0113] The invention also provides a substance according to the invention wherein the peptidase is a dipeptidyl peptidase 4 (DPP4), preferably human DPP4, and preferably wherein the fragment is derived from the ectodomain of dipeptidyl peptidase. In one embodiment, it is provided that the second fragment comprises residues 39 766 of human DPP4 as depicted in figure 18.

[0114] The description also provides a substance according to the invention wherein the second fragment comprises, preferably consists of, at least 10, preferably of at least 50, preferably of at least 100 residues derived from the ectodomain of dipeptidyl peptidase, such as wherein the second fragment is derived from the ectodomain of human DDP4 comprising residues 39 766 as depicted in figure 18.

[0115] Examples of such selected fragments are also found in figure 3, preferably a substance with a second fragment consisting of residues 1 6, or of residues 1 28, or of residues 29 38, or of residues 39 51, or of residues 506 766 as depicted in figure 18.

[0116] Even more in particular, a substance is provided in the present description wherein the second fragment is derived from the ectodomain of a peptidase, which fragment is then subjected to limited proteolysis after which the protease resistant domains are identified by MS, and the interaction between probe and target is studied further.

[0117] The description also provides a substance according to the invention wherein at least one of the isolated fragments has been provided with an affinity tag, preferably a tag having affinity to binding with Protein A or a tag having affinity for binding with streptavidin.

[0118] The description also provides a substance according to the invention consisting essentially of an isolated first fragment of a viral protein and an isolated second fragment of an N terminal peptidase protein. In a preferred embodiment, the viral protein is a coronaviral protein, preferably derived from a virus capable of infecting a human cell, whereas the peptidase protein is a DPP4 protein, preferably a human DPP4 protein.

[0119] Furthermore, a substance according to the invention is herein provided that has been subjected to crystallization, preferably a substance comprising a crystal consisting essentially of an isolated first fragment of a viral protein and an isolated second fragment of an N terminal peptidase protein. In a preferred embodiment, the viral protein is a coronaviral protein, preferably derived from a virus capable of infecting a human cell, whereas the peptidase protein is a DPP4 protein, preferably a human DPP4 protein.

[0120] The description also provides a method for identifying a binding site comprising subjecting a crystal consisting essentially of an isolated first fragment of a viral protein and an isolated second fragment of an N terminal peptidase protein to X ray or neutron diffraction analysis. This is, for example, in order to determine the three dimensional structure of fragments of DPPIV and coronaviral protein and, in particular, to assist in the identification of its active site where fragments may bind. Knowledge of the binding site region allows rational design and construction of ligands including inhibitors. Crystallization and structural determination of fragments of DPPIV mutants and/or viral protein mutants having altered bioactivity allows the evaluation of whether such changes are caused by general structure deformation or by side chain alterations at the substitution site.

[0121] The invention also provides a container with a substance according to the invention, such as container provided with a virus according to the invention, and/or a nucleic acid according to the invention, and/or a vector according to the invention, and/or a host cell according to the invention, and/or a proteinaceous molecule according the invention, and/or an antigen according to the invention, and/or or an antibody according to the invention and/or a pharmaceutical composition according to the invention. In describing a container herein, reference is made to a test device, test tube (commonly Eppendorf tubes are used), test vessel, pipette, pipette tip, reaction device, cell culture vessel, cell culture well, reaction chamber, cover slip, crystallization chamber, crystallization device, crystallization well, microplate well, crystallization plate well, gel, column wherein, on or under a proteinaceous substance according to the invention may be placed or contained or that are useful for storing, shipping, testing or handling a proteinaceous substance provided herein.

[0122] The invention also provides a method of identifying a candidate modulator as an agent that modulates the function of a dipeptidyl peptidase, the method comprising providing a substance w according to the invention in the presence and absence of the candidate modulator under conditions permitting binding of a protein derived from the virus with the fragment derived from a peptidase protein. Measuring binding of said protein to said fragment, wherein a decrease or increase in binding in the presence of the candidate modulator, relative to binding in the absence of the candidate modulator, identifies the candidate modulator as an agent that modulates the function of a dipeptidyl peptidase, or identifies said antiviral agent as an agent that modulates the function of a dipeptidyl peptidase, preferably wherein said protein and/or said fragment is detectably labeled, preferably with a moiety selected from the group consisting of a radioisotope, a fluorophore, a quencher of fluorescence, an enzyme, and an affinity tag.

[0123] The description further provides a method of detecting, in a sample, the presence of an agent that modulates the function of a dipeptidyl peptidase, said method comprising providing a substance with a first and a second fragment according to the invention in the presence and absence of said sample under conditions permitting binding of said first fragment with said second fragment. Measuring binding of said first fragment to said second fragment, wherein a decrease or increase in binding in the presence of said sample, relative to binding in the absence of said sample, identifies said sample as comprising an agent that modulates the function of a dipeptidyl peptidase.

[0124] The description further provides a method of identifying a candidate modulator as an agent that modulates the function of a dipeptidyl peptidase, said method comprising providing a substance with a first and a second fragment according to the invention in the presence and absence of said candidate modulator under conditions permitting determining enzymatic activity of a peptidase. Measuring enzymatic activity of a peptidase, wherein a decrease or increase in enzymatic in the presence of said candidate modulator, relative to binding in the absence of said candidate modulator, identifies said candidate modulator as an agent that modulates the function of a dipeptidyl peptidase.

[0125] The description further provides a method of detecting, in a sample, the presence of an agent that modulates the function of a dipeptidyl peptidase, said method comprising providing a substance with a first and a second fragment according to the invention in the presence and absence of said sample under conditions permitting determining enzymatic activity of a peptidase. Measuring enzymatic activity of a peptidase, wherein a decrease or increase in enzymatic in the

presence of said sample, relative to binding in the absence of said sample, identifies said sample as comprising an agent that modulates the function of a dipeptidyl peptidase.

[0126] In a preferred embodiment, the description further provides a method of identifying a candidate modulator as an that modulates the function of a dipeptidyl peptidase or a provides a method of detecting, in a sample, the presence of an agent that modulates the function of a dipeptidyl peptidase wherein said first fragment and/or said second fragment is detectably labeled, preferably wherein said first fragment and/or said second fragment is detectably labeled with a moiety selected from the group consisting of a radioisotope, a fluorophore, a quencher of fluorescence, an enzyme, and an affinity tag. It is also provided that said substance comprises a cell expressing said first fragment and/or said second fragment.

[0127] The description also provides use of a substance, a container or a method according to the invention for identifying an agent that modulates the function of a peptidase or a viral protein, use of an isolated fragment of a viral protein as an agent that modulates the function of an N-terminal dipeptidyl peptidase, and use of an isolated fragment of a N-terminal dipeptidyl peptidase as an agent that modulates the function of a viral protein.

[0128] The description further provides use of an inhibitor, preferably adenosine, or a functional equivalent thereof, of N-terminal dipeptidyl peptidase cell-surface expression on a cell, as a modulator or antiviral agent for inhibition of replication of a virus in said cell, in particular wherein said peptidase is DPP4, preferably human DPP4, preferably wherein said virus is a Coronavirus.

[0129] Also, the description provides vaccines against HCoV SA1 (based on nucleic acid or amino acid sequences or antigenic polypeptides of the HCoV SA1 genome, and the invention provides use of antiviral drugs directed against nucleic acid or amino acid sequences or polypeptides of the HCoV SA1 (herein also called MERS HCoV) genome, as herein provided. At this time, it is not known if there is a cure for the disease. Several antiviral therapies have been applied, but with various results. Also, for being able to prevent spread of the disease, it is of great importance to be able to recognize the disease in an early stage. Only then, sufficient measures can be taken to isolate patients and initiate quarantine precautions.

[0130] The invention also provides an isolated or recombinant proteinaceous molecule or MERS-CoV-specific fragment thereof encoded by a nucleic acid according to the invention. In a preferred embodiment, the invention provides a proteinaceous molecule or corona MERS-CoV-specific viral protein or fragment thereof encoded by a nucleic acid according to the invention. Useful proteinaceous molecules are, for example, derived from any of the genes or genomic fragments derivable from a virus according to the invention. Such molecules or antigenic fragments thereof, as provided herein, are, for example, useful in diagnostic methods or kits and in pharmaceutical compositions such as sub unit vaccines and inhibitory peptides. Particularly useful is the viral polymerase protein, the spike protein, the nucleocapsid for inclusion as antigen or subunit immunogen, but inactivated whole virus can also be used.

[0131] Particularly useful are also those proteinaceous substances that are encoded by recombinant nucleic acid fragments that are identified for phylogenetic analyses, of course, preferred are those that are within the preferred bounds and metes of ORFs useful in phylogenetic analyses, in particular, for eliciting MERS-CoV-specific antibodies, whether in vivo (e.g., for protective purposes or for providing diagnostic antibodies) or in vitro (e.g., by phage display technology or another technique useful for generating synthetic antibodies). Similarly, the invention provides an antigen comprising a proteinaceous molecule or MERS-CoV-specific fragment thereof according to the invention, or reactive with an antibody according to the invention.

[0132] Also provided herein are antibodies, be it natural polyclonal or monoclonal, or synthetic (e.g., (phage) library derived binding molecules) antibodies that specifically react with an antigen comprising a proteinaceous molecule or HCoV virus like MERS-CoV-specific fragment thereof according to the invention. A person skilled in the art will be able to develop (monoclonal) antibodies using isolated virus material and/or recombinantly expressed viral proteins. Sui et al. (Proc. Natl. Acad. Sci. 101(8):2536 2541, 2004) have transiently expressed fragments of the spike protein and found several antibodies through phage display methods. Such antibodies are also useful in a method for identifying a viral isolate as a MERS HCoV virus like virus comprising reacting the viral isolate or a component thereof with an antibody as provided herein. This can, for example, be achieved by using purified or non purified HCoV SA1 virus like virus or parts thereof (proteins, peptides) using ELISA, RIA, FACS or similar formats of antigen detection assays (Current Protocols in Immunology). Alternatively, infected cells or cell cultures may be used to identify viral antigens using classical immunofluorescence or immunohistochemical techniques. Specifically useful in this respect are antibodies raised against MERS HCoV virus like virus proteins that are encoded by a nucleotide sequence comprising one or more of the fragments disclosed herein.

[0133] In particular, MERS HCoV virus like polypeptide or fragments are provided herein as well, such as those provided in figure 16 or figure 17, in particular, fragments derived from a viral spike protein, preferably the S1 spike protein, in particular, fragments of the S1 protein, such as fragment 1 357, or fragment 358 747, or fragment 358-588, or homologues thereof, as depicted in figure 17, or fragment 363 593 of the spike protein of HKU4 Co, as shown in figure 32, are herein provided. Also, isolated or recombinant nucleic acid, or MERS-CoV-specific fragments thereof that are obtainable from a MERS HCoV virus are provided, such as nucleic acid encoding fragments of the S1 protein, such as

fragment 1 357, or fragment 358 747, or fragment 358-588, or homologues thereof, as depicted in figure 17, as are a vector or plasmid comprising a nucleic acid according to the invention, and a cell, such as host cell, such as a 293T cell comprising a nucleic acid or vector (vector comprising plasmid herein) according to the invention.

[0134] The invention also provides an isolated or recombinant proteinaceous molecule or MERS-CoV-specific fragment thereof encoded by a nucleic acid according to the invention. In a preferred embodiment, the invention provides a proteinaceous molecule or MERS-CoV-specific viral protein or fragment thereof encoded by a nucleic acid according to the invention for use in a vaccine. Useful proteinaceous molecules are, for example, derived from any of the genes or genomic fragments derivable from a virus or fragment thereof according to the invention. Such molecules, or antigenic fragments thereof, as provided herein, are, for example, useful in diagnostic methods or kits and in pharmaceutical compositions such as sub unit vaccines and inhibitory peptides.

[0135] Particularly useful are the viral polymerase protein, the spike protein, the nucleocapsid or antigenic fragments thereof for inclusion in a vaccine as antigen or subunit immunogen, in particular, fragments derived from a viral spike protein, preferably the S1 spike protein is provide for use in a vaccine, in particular, fragments of the S1 protein, such as fragment 1 357, or fragment 358 747, or fragment 358-588, or homologues thereof, as depicted in figure 17 that were are shown herein to interact with DPP4 and to elicit neutralizing antibodies, or fragment 363 593 of the spike protein of HKU4 Co remarkably interacting with DPP4 as well, as shown in figure 32, but inactivated whole virus can also be used in a vaccine. Particularly useful are those proteinaceous substances that are encoded by recombinant nucleic acid fragments that are identified for phylogenetic analyses, of course, preferred are those that are within the preferred bounds and metes of ORFs useful in phylogenetic analyses, in particular, for eliciting MERS-CoV specific antibodies, whether in vivo (e.g., for protective purposes such as by vaccination or for providing diagnostic antibodies) or in vitro (e.g., by phage display technology or another technique useful for generating synthetic antibodies). Similarly, the invention provides an antigen comprising a proteinaceous molecule or MERS-CoV-specific fragment thereof according to the invention, reactive with an antibody according to the invention. Such an antibody as herein provided is preferably reactive with a fragment of the S1 protein, such as fragment 1 357, or fragment 358 747, preferably fragment 358-588, of MERS-CoV or homologues thereof, as depicted in figure 17 and 32.

[0136] The invention also provides a pharmaceutical composition comprising a virus, a nucleic acid, a proteinaceous molecule or fragment thereof, preferably consisting of the amino acid sequence 358 588 of MERS CoV or of the sequence 363 593 of the spike protein of HKU4 CoV, more preferably having at least a part of the amino acid sequence 358 588 of MERS CoV or of the sequence 363 593 of the spike protein of HKU4 CoV as depicted herein.

[0137] An antigen and/or an antibody according to the invention can, for example, be used in a method for the treatment or prevention of a MERS HCoV infection and/or a respiratory illness comprising providing an individual with a pharmaceutical composition according to the invention, for example as a vaccination against useful against infection with corona viruses that use DPP4 as a virus receptor such as seen with MERS-CoV infection or HKU4-CoV infection. This is most useful when the individual comprises a human. Antibodies directed against MERS HCoV proteins, especially against the spike protein of MERS HCoV, preferably against the amino acid sequence 358 588 or the sequence 363 593 of the spike protein of HKU4 CoV, or more preferably directed against at least a part of the amino acid sequence 358 588 of MERS CoV or of the sequence 363 593 of the spike protein of HKU4 Co are herein also provided and are useful for prophylactic or therapeutic purposes, as passive vaccines or part of an anti-serum useful to protect against infection with corona viruses that use DPP4 as a virus receptor, such as MERS-CoV or HKU4-CoV. It is known from other coronaviruses that the spike protein is a very strong antigen and that antibodies against spike protein can be used in prophylactic and therapeutic treatment.

[0138] The description also provides a proteinaceous substance having been provided with a isolated or recombinant proteinaceous molecule or MERS-CoV-specific fragment thereof encoded by a nucleic acid according to the invention and additionally comprising at least a fragment of an N-terminal dipeptidyl peptidase protein. In a preferred embodiment, the description provides a proteinaceous substance having been provided with a proteinaceous molecule or MERS-CoV-specific viral protein or fragment thereof encoded by a nucleic acid according to the invention. Useful proteinaceous molecules are for example derived from any of the genes or genomic fragments or open reading frames (ORFs) derivable from a virus according to the invention. Particularly useful are the viral polymerase protein, the spike protein, the nucleocapsid or antigenic fragments thereof, but inactivated whole virus can also be used. Particularly useful are also those proteinaceous substances that are encoded by recombinant nucleic acid fragments that are identified for phylogenetic analyses, of course preferred are those that are within the preferred bounds and metes of ORFs useful in phylogenetic analyses, in particular for eliciting MERS-CoV specific antibodies,

[0139] The invention also provides a proteinaceous substance comprising an isolated or recombinant proteinaceous molecule according to the invention or MERS-CoV-specific fragment thereof wherein said proteinaceous molecule comprises an ectodomain of a spike protein, said ectodomain preferably derived from the S1 region of a coronavirus. In another preferred embodiment, the invention also proteinaceous substance having been provided with a isolated or recombinant proteinaceous molecule or MERS-CoV-specific fragment thereof wherein said peptidase protein is a dipeptidyl peptidase 4 (DPP4), preferably human DPP4, or a fragment of DPP4.

[0140] Typically the invention provides a proteinaceous substance comprising an isolated or recombinant proteinaceous molecule or MERS-CoV-specific fragment thereof encoded by a nucleic acid according to the invention and additionally comprising at least a fragment of an N-terminal dipeptidyl peptidase protein, said substance having been subjected to crystallization. The invention also provides a container provided with a proteinaceous substance having been provided with a isolated or recombinant proteinaceous molecule according to the invention or fragment thereof encoded by a nucleic acid according to the invention and additionally having been provided with or comprising at least a fragment of an N-terminal dipeptidyl peptidase protein.

[0141] The description also provides a method of identifying a candidate modulator as an agent that modulates the function of a dipeptidyl peptidase, said method comprising: providing a proteinaceous substance according to the invention in the presence and absence of said candidate modulator under conditions permitting binding of said proteinaceous molecule of first fragment of viral fragment with said fragment of said peptidase protein, measuring binding of said molecule to said fragment, wherein a decrease or increase in binding in the presence of said candidate modulator, relative to binding in the absence of said candidate modulator, identifies said candidate modulator as an agent that modulates the function of a dipeptidyl peptidase. It is preferred that said molecule and/or said fragment is detectably labeled, preferably with a moiety selected from the group consisting of a radioisotope, a fluorophore, a quencher of fluorescence, an enzyme, and an affinity tag.

[0142] The description also provides use of at least a fragment of a viral protein as an agent that modulates the function of an N-terminal dipeptidyl peptidase, for example such use is provided herein in a method according to the invention. Similarly, the description provides use of a fragment of an N-terminal dipeptidyl peptidase as an agent that modulates the function of a viral protein for example such use is provided herein in a method according to the invention. The description also provides use of an inhibitor of N-terminal dipeptidyl peptidase, such as ADA, or a functional equivalent thereof, in a method for detecting inhibition of replication of a virus in a cell, preferably wherein said peptidase is DPP4, more preferably wherein said virus is a Coronavirus. The description also provides use of an inhibitor of N-terminal dipeptidyl peptidase cell-surface expression, such as adenosine, or a functional equivalent thereof, in a method for detecting inhibition of replication of a virus in a cell, preferably wherein said peptidase is DPP4, more preferably wherein said virus is a Coronavirus.

[0143] The description thus further provides a method of detecting, in a sample, the presence of an agent that modulates the function of a dipeptidyl peptidase, the method comprising providing a substance with a first and a second fragment in the presence and absence of the sample under conditions permitting binding of the first fragment with the second fragment. Measuring binding of the first fragment to the second fragment, wherein a decrease or increase in binding in the presence of the sample, relative to binding in the absence of the sample, identifies the sample as comprising an agent that modulates the function of a dipeptidyl peptidase.

[0144] The description further provides a method of identifying a candidate modulator as an agent that modulates the function of a dipeptidyl peptidase, the method comprising providing a substance with a first and a second fragment according to the invention in the presence and absence of the candidate modulator under conditions permitting determining enzymatic activity of a peptidase. Measuring enzymatic activity of a peptidase, wherein a decrease or increase in enzymatic activity in the presence of the candidate modulator, relative to binding in the absence of the candidate modulator, identifies the candidate modulator as an agent that modulates the function of a dipeptidyl peptidase.

[0145] The description further provides a method of detecting, in a sample, the presence of an agent that modulates the function of a dipeptidyl peptidase, the method comprising providing a substance with a first and a second fragment according to the invention in the presence and absence of the sample under conditions permitting determining enzymatic activity of a peptidase. Measuring enzymatic activity of a peptidase, wherein a decrease or increase in enzymatic activity in the presence of the sample, relative to binding in the absence of the sample, identifies the sample as comprising an agent that modulates the function of a dipeptidyl peptidase.

[0146] In a preferred embodiment, the description further provides a method of identifying a candidate modulator as an agent that modulates the function of a dipeptidyl peptidase or provides a method of detecting, in a sample, the presence of an agent that modulates the function of a dipeptidyl peptidase, wherein the first fragment and/or the second fragment is detectably labeled, preferably wherein the first fragment and/or the second fragment is detectably labeled with a moiety selected from the group consisting of a radioisotope, a fluorophore, a quencher of fluorescence, an enzyme, and an affinity tag. It is also provided that the substance comprises a cell expressing the first fragment and/or the second fragment. The description also provides a method of identifying a candidate modulator as an agent that modulates the function of a dipeptidyl peptidase, said method comprising providing a proteinaceous substance according to the invention in the presence and absence of said candidate modulator under conditions permitting binding of a first fragment derived from a virus with a second fragment derived from a peptidase protein, and measuring binding of said first to said second fragment, wherein a decrease or increase in binding in the presence of said candidate modulator, relative to binding in the absence of said candidate modulator, identifies said candidate modulator as an agent that modulates the function of a dipeptidyl peptidase, it is preferred that said first and/or said second fragment is detectably labeled, preferably with a moiety selected from the group consisting of a radioisotope, a fluorophore, a quencher of fluorescence, an enzyme,

and an affinity tag. The description also provides a method of identifying a candidate antiviral agent as an agent that modulates the binding of a virus to dipeptidyl peptidase, said method comprising providing a proteinaceous substance according to the invention in the presence and absence of said candidate antiviral agent under conditions permitting binding of a first fragment derived from a virus with a second fragment derived from a peptidase protein, measuring binding of said first to said second fragment, wherein a decrease or increase in binding in the presence of said antiviral agent, relative to binding in the absence of said candidate modulator, identifies said antiviral agent as an agent that modulates the function of a dipeptidyl peptidase. It is preferred that said first and/or second fragment is detectably labeled, preferably with a moiety selected from the group consisting of a radioisotope, a fluorophore, a quencher of fluorescence, an enzyme, and an affinity tag.

[0147] The description also provides use of a substance, a container or a method according to the invention for identifying an agent that modulates the function of a peptidase or a viral protein, use of an isolated fragment of a viral protein, preferably of a viral spike protein as provided herein, or recombinant or synthetic peptide derived thereof as provided herein, as an agent that modulates the function of an N terminal dipeptidyl peptidase, and use of an isolated fragment of a N terminal dipeptidyl peptidase, preferably of a soluble fragment of said peptidase as provided herein, or recombinant or synthetic peptide derived thereof as provided herein, as an agent that modulates the function of a viral protein. The description further provides use of an inhibitor, preferably ADA, or a functional equivalent thereof, of N terminal dipeptidyl peptidase activity of a cell, for inhibition of replication of a virus in a cell, in particular, wherein the peptidase is DPP4, preferably human DPP4, preferably wherein the virus is a Coronavirus. The description further provides use of an inhibitor, preferably adenosine, or a functional equivalent thereof, of N terminal dipeptidyl peptidase cell surface expression on a cell, for inhibition of replication of a virus in a cell, in particular, wherein the peptidase is DPP4, preferably human DPP4, preferably wherein the virus is a Coronavirus.

[0148] In describing protein or peptide composition, structure and function herein, reference is made to amino acids. In the present specification, amino acid residues are expressed by using the following abbreviations. Also, unless explicitly otherwise indicated, the amino acid sequences of peptides and proteins are identified from N terminal to C terminal, left terminal to right terminal, the N terminal being identified as a first residue. Ala: alanine residue; Asp: aspartate residue; Glu: glutamate residue; Phe: phenylalanine residue; Gly: glycine residue; His: histidine residue; Ile: isoleucine residue; Lys: lysine residue; Leu: leucine residue; Met: methionine residue; Asn: asparagine residue; Pro: proline residue; Gln: glutamine residue; Arg: arginine residue; Ser: serine residue; Thr: threonine residue; Val: valine residue; Trp: tryptophane residue; Tyr: tyrosine residue; Cys: cysteine residue. The amino acids may also be referred to by their conventional one letter code abbreviations; A=Ala; T=Thr; V=Val; C=Cys; L=Leu; Y=Tyr; I=Ile; N=Asn; P=Pro; Q=Gln; F=Phe; D=Asp; W=Trp; E=Glu; M=Met; K=Lys; G=Gly; R=Arg; S=Ser; and H=His.

Figure legends

[0149]

Figure 1. Light microscopy images of LLC-MK2 cells (A, B) and VERO cells (C, D) inoculated with phosphate-buffered saline (A, C) or novel human coronavirus HCoV-SA1 (B, D) 5 days after inoculation.

Figure 2. Results of pan-coronavirus PCR. Various samples (numbered 1-12) of cell culture supernatants infected with HCoV-SA1 reacted with a pair of primers specific for the coronavirus family. A positive control virus (HCoV-NL63) was also reactive.

Figure 3. Partial open reading frame of HCoV-SA1.

Figure 4. Maximum Likelihood tree of partial polymerase gene sequences of representative coronaviruses. HCoC-SA1 is shown in the cluster on the right hand side of the tree, labeled as "New HCoV". The cluster of viruses at the top represents viruses in the genus alphacoronavirus. The Beluga whale coronavirus (BWCoV) represents a gammacoronavirus, while the Bulbul-CoV and IBV represent a proposed new genus of the coronavirinae.

Figure 5 file N.rtf nucleocapsid (N) protein

Figure 6 file M.rtf matrix (M) protein

Figure 7 file E.rtf small envelope (E) protein

Figure 8 file NS3d.rtf non-structural gene NS3d

Figure 9 file NS3c.rtf non-structural gene NS3c

Figure 10 file NS3b.rtf non-structural gene NS3b

Figure 11 file NS3a.rtf non-structural gene NS3a

Figure 12 file S.rtf spike surface glycoprotein (S)

Figure 13 file Orflab.rtf encoding many enzymatic products among which the replicase

Figure 14 file HCoV-SA1.rtf

Figure 15 HCoV-SA1.rtf translation 3 frames

Figure 16 Amino acid sequence of the spike protein of HCoV EMC (HCoV SA1). Panel A, schematic presentation of the HCoV EMC S and S1 Fc fusion protein. Position of the predicted N glycosylation sites (Ψ; predicted by the NetNGlyc

server) and TM domain (yellow bar; predicted by the TMHMM server) are indicated in the full length S protein. The border between the S1 and S2 subunits is marked by the presence of a predicted furin cleavage site (red triangle; predicted by the ProP 1.0 server). Residues 1 747 comprise the N terminal region. Panel B, amino acid sequence of the spike protein with the S1 region indicated in red.

Figure 17 Amino acid sequence and domain structure of residues 1 747 of the S1 spike protein of HCoV EMC (HCoV SA1). RBD = Receptor Binding Region.

Figure 18 Domain structure and amino acid sequence of residues 1 766 of human DPP IV, domain borders based on crystal structure (Rasmussen, Nat. Struct. Biol. 2003, herein included by reference).

Figure 19 Binding of HCoV EMC S1 is correlated to infection with HCoV EMC in vero cells (Panel A), Cos 7 cells (Panel B) Huh7 cells (Panel C) and bat cells (Panel D). Shown on the left is the FACS analysis of HCoV EMC S1 binding (red line), a feline CoV S1 protein as control (blue line) and non stained cells (black line). In the middle panels, HCoV EMC infected cells are visualized using an antiserum that recognizes the NSp4 protein and on the left, supernatants of the infected cells are tested by Taqman for the presence of viral transcripts at 0, 20 and 40 hours after infection.

Figure 20 Immunoprecipitation with S1 on Huh7cells and mass spec analysis reveals cd26 as the interacting protein.

Figure 21 Peptides identified in fraction 2 are indicated in red and relate to the fragment or topological domain involving residues 29 766 comprising the extracellular region (ectodomain) of the membrane bound DPP4 (Uniprot identifier P27487) but do not relate to the cytoplasmic domain (residues 1 6) nor to the helical Signal anchor for type II membrane protein domain (residues 7 28) of membrane bound DPP4. Soluble DPP4 runs from residue 39 to residue 766.

Figure 22 HCoV EMC and SARS CoV S1 Fc proteins (2.5 μg) were mock incubated or incubated with 12.5 μg soluble DPP IV (sDPP IV) or soluble ACE2 (sACE2) in 100 μl PBS. Precipitates were washed thrice with lysis buffer and once with PBS, and subjected to a NOVEX® 4 12% Tris Glycine gradient gel (Invitrogen) under non reducing conditions.

Figure 23 Cells were washed twice with ice cold PBS, scraped off the plastic with a rubber policeman and suspended into single cells by pipetting cells up and down. S1 binding of cells was measured by incubating 2.5 x 105 cells with 15 μg/ml of S1 Fc followed by incubation with the fluorescent dye Alexa488 labeled goat anti human IgG antibody and analyzed by flow cytometry.

Figure 24 Inhibition of HCoV EMC replication in Huh7 cells by antibodies to DPP4. Huh7 cells were incubated with 20 μg/ml goat polyclonal antiserum against DPP4, a goat antiserum against ACE2, normal goat serum or left untreated. After 1 hour incubation, the cells were infected with HCoV EMC at a multiplicity of infection of 0.01 and left for 1 hour. Cells were washed twice and again incubated with medium containing the respective antibodies. Supernatant collected at 2 hours (open bars) and 20 hours (closed bars) was tested for presence of HCoV using a Taqman assay. Results are shown as Δ Ct. HCoV EMC infection of Huh7 cells is inhibited by antibodies to DPP4 but not by the other antibodies tested.

Figure 25 Cos7 cells transfected with plasmids encoding human DPP4 (hDPP4) or bat DPP4 (bDPP4), a control plasmid (pcDNA) or left untreated were infected with HCoV EMC at a multiplicity of infection of 1 and left for 1 hour. Cells were washed twice and supernatant collected at 2 hours (open bars), 20 hours (blue bars) and 40 hours (red bars) was tested for presence of HCoV using a Taqman assay. Results are shown as Δ Ct.

Figure 26 Blocking of DPP4 -S1 binding by antibodies directed against S1 serum from a macaque infected with HCoV EMC inhibits binding of recombinant S1 to Huh7 cells. Serum at a dilution of 1:20, obtained from macaques at day 0 (blue line) and day 14 (red line) after infection with 5 x 107 TCID50 HCoV EMC, was preincubated for 1 hour at room temperature with 1.25 μg/ml recombinant S1 protein that was biotinylated and subsequently incubated on Huh7 cells. After treatment with FITC-labeled streptavidin, cells were analyzed for fluorescence. In gray background, binding using a control biotinylated protein is shown.

Figure 27 Inhibition of HCoV EMC replication in Huh7 cells by soluble adenosine deaminase (ADA). Huh7 cells were incubated with different concentrations of recombinant soluble ADA (closed bars) or recombinant soluble ACE2 (open bars). After 1 hour incubation, the cells were infected with HCoV EMC at a multiplicity of infection of 0.01. After 8 hours, cells were fixed and stained with a rabbit antiserum against HCoV EMC nsp4 and cells were counted. Results are shown as number of infected cells per well. Infection of Huh7 cells is inhibited by recombinant soluble ADA but not by recombinant soluble ACE2.

Figure 28 Inhibition of HCoV EMC replication in Huh7 cells by soluble DPP4. Different concentrations of recombinant soluble DPP4 (open bars) or recombinant soluble ACE2 (closed bars) were incubated with HCoV EMC for 1 hour at 37°C and used to infect Huh7 cells. After 8 hours, cells were fixed and stained with a rabbit antiserum against HCoV EMC nsp4 and cells were counted. Results are shown as number of infected cells per well. Infection of Huh7 cells is inhibited by recombinant soluble DPP4 but not by recombinant soluble ACE2.

Figure 29 Receptor binding domains in betacoronavirus spike proteins and S1 Fc expression constructs. Panel a), schematic representation of the betacoronaviruses SARS CoV, hCoV EMC S and MHV (strain A59) spike (S) protein sequence (drawn to scale) aligned at the S1 S2 junction. The known receptor binding domain in the S1 subunit of MHV and SARS CoV S proteins and their corresponding homologous regions in hCoV EMC S as defined by ClustalW alignment are indicated. Positions of the transmembrane domain (yellow bar; predicted by the TMHMM server) and of the predicted

N glycosylation sites (Ψ; predicted by the NetNGlyc server, only shown for the hCoV EMC S) are indicated. The border between the S1 and S2 subunits of the spike protein is represented by a vertical white line. Panel b), upper panel, schematic presentation of the hCoV EMC S1 subunit (residues 1 751) sequence. Cysteine positions in S1 subunit are indicated by vertical white lines with corresponding amino acid positions on top. Positions of cysteines highly conserved among betacoronaviruses S1 proteins are in bold. Predicted disulfide bond connections inferred from the structure of the SARS CoV receptor binding domain are presented as connecting black lines underneath. Lower panel, domains of the hCoV EMC S1 subunit expressed as Fc chimeras.

Figure 30 The DPP4 binding domain is located within residues 358 588 of the hCoV EMC spike protein and efficiently elicits neutralizing antibodies. Panel a), S1 Fc chimeric proteins and soluble DPP4 (sDPP4) receptor were expressed from HEK 293T cells and purified from the culture supernatant. S1 Fc proteins were mixed with sDPP4 followed by protein A sepharose affinity isolation, analyzed on a NOVEX® 4 12% Tris Glycine gradient gel under non reducing conditions, and stained with GelCodeBlue reagent. Position of the S1 Fc proteins, running as dimers under non reducing conditions due to an Fc interchain disulphide bond, and sDPP4 as well as the sizes of the marker proteins are indicated. Individual proteins were loaded as controls. Panel b), binding of hCoV EMC S1 Fc proteins to DPP4 expressing cells. 2.5 x 105 HEK 293T cells transfected with control pCAGGS (grey shaded area) or with pCAGGS DPP4 (black line) expression plasmid were incubated with 15 μg/ml of the indicated S1 Fc followed by incubation with DyLight488 labeled goat anti human IgG antibody and analysis by flow cytometry. An Fc chimera containing the S1 of infectious bronchitis virus (IBV S1 Fc) was taken along as a negative control. Panel c), neutralization of hCoV EMC infection by rabbit antisera raised against the S1 Fc 1 747, 1 357 and 358 588 variants. Virus (100 pfu) was premixed 1:1 with serial dilutions of sera obtained (open bars) or after immunization (closed bars) prior to inoculation onto VERO cells and virus infection was monitored by the occurrence of CPE at 72 hours post infection. Virus neutralization titers (VNT) were determined in quadruplicate as the highest serum dilutions that completely prevent CPE. The experiment was carried out twice and the data of one representative experiment are shown.

Figure 31 Localization of receptor binding domains in coronavirus spike proteins. Schematic presentation of the spike proteins of the alphacoronaviruses TGEV and hCoV NL63 and of the betacoronaviruses SARS CoV, hCoV EMC and MHV (drawn to scale), aligned at the S1 S2 junction. Blue boxes represent the receptor binding domains (RBD) and indicate the engaged receptor. The RBD of TGEV, hCoV NL63, SARS CoV and MHV have been confirmed by crystallography (12, 15, 22, 26). Grey boxes indicate the transmembrane domain. Sequence IDs: TGEV (ABG89335.1), hCoV NL63 (NC_005831.2), SARS CoV (NP_828851.1), hCoV EMC (AFS88936.1), MHV (NC_001846.1).

Figure 32 Residues 363 593 of the spike protein of HKU4 CoV bind to human DPP4. Shown is the binding ability of different S1 Fc proteins to DPP4 expressing cells. 2.5 x 105 HEK 293T cells transfected with control pCAGGS (grey shaded area) or with pCAGGS DPP4 (black line) expression plasmid were incubated with 15 μg/ml of the hCoV EMC S1 Fc followed by incubation with DyLight488 labeled goat anti human IgG antibody and analysis by flow cytometry. S1 Fc protein chimeras were tested containing the hCoV EMC S1 subunit (residues 1 747), the hCoV EMC spike receptor binding domain (RBD; residues 358 588) or the hCoV EMC RBD homologous regions of the spike proteins of HKU4 CoV (residues 363 593) and HKU5 CoV (residues 366 586). Mock incubated cells (mock) or cells incubated with an Fc chimera containing the S1 of feline infectious peritonitis virus (FIPV S1 Fc) was taken along as negative controls.

HKU4 CoV spike (S) protein ID [YP_001039953.1].
HKU5 CoV spike (S) protein ID [YP_001039962.1].

Region in S homologous to hCoV EMC RBD highlighted in yellow.

Figure 33. Characterization of the functional MERS-CoV DPP4 receptor S1 binding site.

[0150] A, Different plasmids encoding either full length human DPP4, ferret DPP4 or human-ferret DPP4 chimera's (human-ferret-human and ferret-human ferret, HFH and FHF respectively) were constructed. B, DPP4 expression and S1 binding to cells transfected with different DPP4 constructs as analysed by FACS analysis. C, MERS-CoV RNA levels in supernatants of DPP4 transfected cells infected with MERS-CoV at 2 and 20 h after infection
using a TaqMan assay, expressed as genome equivalents (GE; half maximal tissue-culture infectious dose (TCID50) per ml). D, S1 binding to cells transfected with different hDPP4 mutants. E, MERS-CoV infection of cells transfected with different hDPP4 constructs. Data in panel a and b were corrected for DPP4 expression of the different constructs.

Detailed description

Novel human coronavirus HCoV-SA1

Classification:

[0151]

Order: Nidovirales
Family: Coronaviridae
Subfamily: Coronavirinae
Genus: Betacoronavirus
Lineage: C

Example 1

[0152] Virus was isolated from a 60-year old man with acute pneumonia and acute renal failure in Saudi Arabia.

[0153] Virus was isolated from sputum specimen in VERO cells and LLC-MK2 cells.

[0154] Five days after inoculation, cytopathic effects were observed, consisting of rounding of the cells, detachment of cells, and syncytium formation (Figure 1).

[0155] Cells in the original sputum sample and infected cultured cells were also tested with specific antibodies against influenza A and B viruses, parainfluenza viruses types 1-3, respiratory syncytial virus, and adenovirus, but such tests yielded negative results. Sputum specimens and infected cell culture supernatants did not react in PCR-based assays specific for paramyxoviruses, enteroviruses, and adenoviruses. However, these samples did react with PCR-based assays to detect all coronaviruses. A 251 nucleotide fragment was amplified with one such test (Vijgen, L., E. Moes, E. Keyaerts, S. Li, and M. Van Ranst. 2008. A pancoronavirus RT-PCR assay for detection of all known coronaviruses. Methods Mol Biol 454:3-12). A second PCR-based assay to detect all coronaviruses (Drosten C, Günther S, Preiser W, van der Werf S, Brodt HR, Becker S, Rabenau H, Panning M, Kolesnikova L, Fouchier RA, Berger A, Burguière AM, Cinatl J, Eickmann M, Escriou N, Grywna K, Kramme S, Manuguerra JC, Müller S, Rickerts V, Stürmer M, Vieth S, Klenk HD, Osterhaus AD, Schmitz H, Doerr HW. Identification of a novel coronavirus in patients with severe acute respiratory syndrome. N Engl J Med. 348, 1967-76 (2003)) also yielded positive results (Figure 2).

Example 2

[0156] Viral RNA was isolated from infected cell culture supernatants using a High Pure RNA Isolation Kit (Roche). Extracted RNA was copied to cDNA by reverse transcriptase using random hexamers. Pan-coronavirus polymerase chain reaction (PCR) was used to amplify a conserved region of open reading frame 1b (Drosten C, Günther S, Preiser W, van der Werf S, Brodt HR, Becker S, Rabenau H, Panning M, Kolesnikova L, Fouchier RA, Berger A, Burguière AM, Cinatl J, Eickmann M, Escriou N, Grywna K, Kramme S, Manuguerra JC, Müller S, Rickerts V, Stürmer M, Vieth S, Klenk HD, Osterhaus AD, Schmitz H, Doerr HW. Identification of a novel coronavirus in patients with severe acute respiratory syndrome. N Engl J Med. 348, 1967-76 (2003)). The PCR fragments of the pan-coronavirus PCRs were sequenced. To this end, PCR products were purified from the gel and sequenced using a BigDye Terminator v3.1 Cycle sequencing kit (Applied Biosystems, Nieuwerkerk a/d IJssel, the Netherlands) and a 3130XL genetic analyzer (Applied Biosystems), according to the instructions of the manufacturer. The sequence clearly corresponded with conserved region of open reading frame 1b of a coronavirus (Figure 3).

Example 3

[0157] Reference coronavirus genome sequences were downloaded from GenBank and the part of the genomes that corresponded with the amplified fragment of HCoV-SA1 were aligned. A Maximum Likelihood tree was constructed to infer the phylogenetic relationships (Figure 4). This phylogenetic tree showed that the new HCoV-SA1 belongs to lineage C of the genus Betacoronavirus, along with the bat coronaviruses HKU4 and HKU5. The Betacoronavirus genus contains 3 additional lineages (A, B, D). HCoV-HKU1 and HCoV-OC43 belong to lineage A while SARS-CoV belongs to lineage B. Lineage D does not contain any human pathogens, and is represented in the tree by Rousettus bat coronavirus HKU9.

[0158] HCoV-SA1 is thus clearly distinct from previously known human coronaviruses. HCoV-NL63 and HCoV-229E are even more distinct from HCoV-SA1, since these two human pathogens belong to a different genus, the Alphacoronavirus genus.

Example 4

[0159] To further characterize the virus genome, viral RNA was extracted from infected cell culture supernatant using the High Pure RNA Isolation Kit (Roche). RNA was subjected to reverse transcriptase using circular permuted primers (Welsh, J. & McClelland, M. Fingerprinting genomes using PCR with arbitrary primers. Nucleic Acids Res. 18, 7213-7218 (1990)) that were extended with random hexamer sequences. The amount of DNA was amplified by polymerase chain reaction (PCR), using the circular permuted primers. The randomly amplified fragments were sequenced using the 454/Roche GS-FLX sequencing platform. A fragment library was created according to the manufacturer's protocol without DNA fragmentation (GS FLX Titanium Rapid Library Preparation, Roche). The emPCR (Amplification Method Lib-L) and GS junior sequencing run was performed according to instructions of the manufacturer (Roche). The sequence reads were trimmed at 30 nucleotides from the 3' and 5' ends to remove all primer sequences. Sequence reads from the GS-FLX sequencing data were assembled into contigs using CLC Genomics software 4.6.1. Using this "deep-sequencing" approach on the 454-sequencing platform, approximately 80% of the virus genome sequence was obtained. Subsequently, specific primers were designed to amplify 30 overlapping fragments of approximately 1500 basepairs by PCR. Each of these PCR products was sequenced using conventional Sanger sequencing. To this end, PCR products were purified from the gel and sequenced using a BigDye Terminator v3.1 Cycle sequencing kit (Applied Biosystems, Nieuwerkerk a/d IJssel, the Netherlands) and a 3130XL genetic analyzer (Applied Biosystems), according to the instructions of the manufacturer. The nearly full-length sequence is presented in file HCoV-SA1.rtf. This sequence contains some uncertainties within the extreme 50 nucleotides of both ends. However, this information is not required to classify the coronavirus. The same figure also displays the full coding potential of HCoV-SA1. As a minimum, the HCoV-SA1 virus genome encodes the open reading frames common to the virus of the betacoronavirus genus, including orflab that encodes many enzymatic products, the spike surface glycoprotein (S), the non-structural genes NS3a, NS3b, NS3c, NS3d, the small envelope (E) protein, the matrix (M) protein, and the nucleocapsid (N) protein. Open reading frames are presented in files Orflab.rtf, S.rtf, NS3a.rtf, NS3b.rtf, NS3c.rtf, NS3d.rtf, E.rtf, M.rtf, N.rtf. Other open reading frames may be present.

Example 5

[0160] Comparison of the Orflab gene product of HCoV-SA1 with those of the other members of the Betacoronavirus genus, HKU4 and HKU5 was used to test if HCoV-SA1 belongs to one of these known species or represents a new species within the genus. The International Committee on the Taxonomy of Viruses (ICTV) considers viruses that share more than 90% aa sequence identity in the conserved replicase domains to belong to the same species. This 90% identity threshold serves as the sole species demarcation criterion. Since amino acid sequence identity of Orflab between HCoV-SA1 and HKU4 and HKU5 is below 74% (Table 1), we conclude that HCoV-SA1 represents a novel species of the Betacoronavirus genus, although such classification requires ICTV approval.

Table 1. Percentage amino acid sequence identity between ORFlab of HCoV-SA1, HKU4 (Genbank accession numbers EF065505-EF065508) and HKU5 (accession numbers EF065509-EF065512)

	HCoV-SA1	HKU4	HKU5
HCoV-SA1	100%	72%	74%
HKU4	72%	99-100%	77%
HKU5	74%	77%	99-100%

[0161] The present invention in particular also relates to the spike (S) protein of a coronavirus and fragments thereof as depicted in figures 16 and 17.

[0162] The present descrption also relates to a member of the S9 family of human proteases known as dipeptidyl peptidase IV (DPPIV, figure 18), and fragments thereof.

Protein Expression

Example 6

[0163] A plasmid encoding HCoV EMC S1 Fc was generated by ligating a fragment encoding the S1 region (residues 1 747) into the pCAGGS expression vector as an N terminal fusion with the fragment encoding the Fc domain of human IgG (figures 1 and 2). Likewise, an S1 Fc expression plasmid was made for the SARS coronavirus S1 subunit (strain Urbani: residues 1 676) and the FIPV S1 subunit (strain 79 1146; residues 1 788). S1 Fc proteins were expressed by transfection of the expression plasmids into 293T cells and affinity purified from the culture supernatant using protein A

sepharose beads.

Example 7

[0164] A plasmid encoding the ectodomain of human DPP4 (figure 18) was generated by ligating a fragment encoding residues 39 766 of human DPP4 into a pCD5 expression vector encoding the signal sequence of CD5 and a OneSTrEP affinity tag (IBA GmbH). Soluble DPP4 ectodomain was expressed by transfection of the expression plasmid into 293T cells and affinity purified from the culture supernatant using Streptactin sepharose beads (IBA GmbH).

Example 8

[0165] A plasmid encoding HCoV EMC S1 Fc was generated by ligating a fragment encoding the S1 region (residues 1 747) into the pCAGGS expression vector as an N terminal fusion with the fragment encoding the Fc domain of human IgG separated by a thrombin cleavage site. Likewise, an Fc expression plasmid was made for the SARS coronavirus S1 subunit (isolate CUHK W1: residues 1 676), the FIPV S1 subunit (isolate 79 1146; residues 1 788) and the ectodomain of human ACE2 (sACE2; residues 1 614). Fc chimeric proteins were expressed by transfection of the expression plasmids into 293T cells and affinity purified from the culture supernatant using protein A sepharose beads (GE Healthcare). Purified ACE2 Fc was cleaved with thrombin and soluble ACE2 was purified by gel filtration.

Example 9

[0166] A plasmid encoding the ectodomain of human DPP IV (sDPP IV) was generated by ligating a fragment encoding residues 39 766 of human DPP IV into a pCD5 expression vector encoding the signal sequence of CD5 and the OneSTrEP tag (IBA GmbH). Soluble DPP IV ectodomain was expressed by transfection of the expression plasmid into HEK 293T cells and affinity purified from the culture supernatant using Strep Tactin Sepharose beads (IBA GmbH).

[0167] Pull down; immunoprecipitation and detection of DPP4

Example 10

[0168] The immunoprecipitation protocol was essentially carried out as described before with some modifications (Liet al., 2003, Nature 426:450, included herein by reference). In short, Huh 7 cells were washed twice with ice cold PBS, scraped off the plastic with a rubber policeman, pelleted and lysed in ice cold lysis buffer (0.3% DDM in PBS) containing protease inhibitors (Roche Complete Mini) at a final density of ~2.5 x 107 cells/mL. Cell lysates were precleared with protein A sepharose beads after which 10 micrograms of probe S1 Fc was added to 1 ml of cell lysate and incubated for 1 hour at 4°C under rotation. Precipitates were washed thrice with lysis buffer and once with PBS and subjected to NOVEX® 4 12% Tris Glycine gradient gel (Invitrogen) under reducing and non reducing conditions. A distinct 110 kDa band precipitated with EMC S1 Fc was visualized by GelCodeBlue staining, excised from the gel, incubated with trypsin and analyzed by MS. Results are shown in figure 20 and results of target analyses are shown in figure 21.

Example 11

[0169] DPP4 cell surface expression was measured using S1 Fc proteins. Cells were washed twice with ice cold PBS, scraped off the plastic with a rubber policeman and suspended into single cells by pipetting cells up and down. S1 binding of cells was measured by incubating 2.5 x 105 cells with 15 μg/ml of S1 Fc followed by incubation with the fluorescent dye Alexa488 labeled goat anti human IgG antibody and analyzed by flow cytometry. Results are shown in figure 23.

RNA extraction and quantitative RT PCR

Example 12

[0170] RNA from 200 μl of supernatant was isolated with the Magnapure LC total nucleic acid isolation kit (Roche) external lysis protocol and eluted in 100 μl. HCoV EMC RNA was quantified on the ABI prism 7700, with use of the Taqman Reverse Transcription Reagents and Taqman PCR Core Reagent kit (Applied Biosystems), using 20 μl isolated RNA, 1× Taqman buffer, 5.5 mM MgCl2, 1.2 mM dNTPs, 0.25 U Amplitaq gold DNA polymerase, 0.25 U Multiscribe reverse transcriptase, 0.4 U RNAse inhibitor, 200 nM primers, and 100 nM probe. Amplification parameters were 30 minutes at 48°C, 10 minutes at 95°C, and 40 cycles of 15 seconds at 95°C, and 1 minute at 60°C. RNA dilutions isolated from a HCoV EMC stock were used as a standard. Results are shown in figures 17, 24, 25 and 26.

Example 13

[0171] HCoV EMC and SARS CoV S1 Fc proteins (2.5 μg) were mock incubated or incubated with 12.5 μg soluble DPP IV (sDPP IV) or soluble ACE2 (sACE2) in a total volume of 100 μl PBS. Precipitates were washed thrice with lysis buffer and once with PBS, and subjected to a NOVEX® 4 12% Tris Glycine gradient gel (Invitrogen) under non reducing conditions. Results are shown in figure 22.

Identification of DPP4 using mass spec analysis of peptide fragments

Example 14

[0172] 1D SDS PAGE gel lanes were cut into ~1 mm slices (indicated as nr. 2 in figure 3) using an automatic gel slicer and subjected to in gel reduction with dithiothreitol, alkylation with chloroacetamide and digestion with trypsin (Promega, sequencing grade), essentially as described by Van den Berg et al. (Cell Stem Cell 6:369, included herein by reference). Alternatively, immunoprecipitated proteins were reduced and alkylated on beads similarly as described above. Nanoflow LC MS/MS was performed on either an 1100 series capillary LC system (Agilent Technologies) coupled to an LTQ Orbitrap XL mass spectrometer (Thermo), or an EASY nLC coupled to a Q Exactive mass spectrometer (Thermo), operating in positive mode and equipped with a nanospray source. Peptide mixtures were trapped on a ReproSil C18 reversed phase column (Dr Maisch GmbH; column dimensions 1.5 cm × 100 μm, packed in house) at a flow rate of 8 μl/minute. Peptide separation was performed on ReproSil C18 reversed phase column (Dr Maisch GmbH; column dimensions 15 cm × 50 μm, packed in house) using a linear gradient from 0 to 80% B (A = 0.1% formic acid; B = 80% (v/v) acetonitrile, 0.1% formic acid) in 70 or 120 minutes and at a constant flow rate of 200 nl/minute. The column eluent was directly sprayed into the ESI source of the mass spectrometer. Mass spectra were acquired in continuum mode; fragmentation of the peptides was performed in data dependent mode by CID or HCD. Peak lists were automatically created from raw data files using the Mascot Distiller software (version 2.3; MatrixScience) or Proteome Discoverer (version 1.3; Thermo). The Mascot algorithm (version 2.2; MatrixScience, UK) was used for searching against a Uniprot database (release 2012_10.fasta, taxonomy: Homo sapiens, or Macaca mulatta, or Myotis lucifugus, or Chlorocebus sabaeus, or Felis catus, included herein by reference). The peptide tolerance was set to 10 ppm and the fragment ion tolerance was set to 0.8 Da for CID spectra (LTQ Orbitrap) or to 20 mmu for HCD (Q Exactive) spectra). A maximum number of two missed cleavages by trypsin were allowed and carbamidomethylated cysteine and oxidized methionine were set as fixed and variable modifications, respectively. Results are shown in figure 21.

Inhibition of HCoV EMC replication in Huh7 cells by antibodies to DPP4

Example 15

[0173] Huh7 cells were incubated with 20 μg/ml goat polyclonal antiserum against DPP4, a goat antiserum against ACE2, normal goat serum or left untreated. After 1 hour incubation, the cells were infected with HCoV EMC at a multiplicity of infection of 0.01 and left for 1 hour. Cells were washed twice and again incubated with medium containing the respective antibodies. Supernatant collected at 2 hours (open bars) and 20 hours (closed bars) was tested for presence of HcoV using a Taqman assay. Results are shown as Δ Ct in figure 25.

Blocking of DPP4 -S1 binding by antibodies directed against S1

Example 16

[0174] Serum from a macaque infected with HCoV EMC inhibits binding of recombinant S1 to Huh7 cells. Serum at a dilution of 1:20, obtained from macaques at day 0 (blue line) and day 14 (red line) after infection with 5 x 107 TCID50 HCoV EMC, was preincubated for 1 hour at room temperature with 1.25 μg/ml recombinant S1 protein that was biotinylated and subsequently incubated on Huh7 cells. After treatment with FITC labeled streptavidin, cells were analyzed for fluorescence. In gray background, binding using a control biotinylated protein is shown (figure 26).

Crystallization and crystals comprising a DPP fragment and a viral protein fragment

Example 17

[0175] One aspect of the present invention relates to methods for forming crystals comprising fragments of DPP and viral protein as well as crystals comprising fragments of DPP and viral protein. Crystallization of DPP is essentially known

from, for example, U.S. Patent 7,344,852 or U.S. Patent Publication 2005/0260723 that are included herein by reference.

[0176] In one embodiment of the present invention, a method for forming crystals comprising fragments of DPPIV and viral protein is provided comprising forming a crystallization volume comprising fragments of DPPIV and viral protein, one or more precipitants, optionally a buffer, optionally a monovalent and/or divalent salt and optionally an organic solvent; and storing the crystallization volume in a container under conditions suitable for crystal formation.

[0177] In yet another embodiment, a method for forming crystals comprising fragments of DPPIV and viral protein is provided comprising forming a crystallization volume comprising fragments of DPPIV and viral protein in solution comprising PEG precipitant listed hereinbelow; and storing the crystallization volume in a container under conditions suitable for crystal formation. PEG precipitant 5 50% w/v of precipitant, wherein the precipitant comprises one or more members of the group consisting of PEG MME having a molecular weight range between 300 10000, and PEG having a molecular weight range between 100 10000 pH 5 9. Buffers that may be used include, but are not limited to, tris, bicine, cacodylate, acetate, citrate, MES and combinations thereof. Additives optionally 0.05 to 0.8 M additives wherein the additives comprises sarcosine or 0.5% to 25% additives wherein the additives comprises xylitrol; Protein Concentration 1 mg/ml 50 mg/ml; Temperature 1°C to 25°C.

[0178] In yet another embodiment, a method for forming crystals comprising fragments of DPPIV and viral protein is provided comprising forming a crystallization volume comprising fragments of DPPIV and viral protein; introducing crystals comprising fragments of DPPIV and viral protein as nucleation sites, and storing the crystallization volume under conditions suitable for crystal formation.

[0179] Crystallization experiments may optionally be performed in volumes commonly used in the art, for example, typically 15, 10, 5, or 2 microliters or less. It is noted that the crystallization volume optionally has a volume of less than 1 microliter, optionally 500, 250, 150, 100, 50 or less nanoliters.

[0180] It is also noted that crystallization may be performed by any crystallization method including, but not limited to, batch, dialysis and vapor diffusion (e.g., sitting drop and hanging drop) methods. Micro and/or macro seeding of crystals may also be performed to facilitate crystallization.

[0181] In one variation, crystals comprising DPPIV are formed by mixing a substantially pure DPPIV fragment and a substantially pure S1 HCoV EMC fragment with an aqueous buffer containing a precipitant at a concentration just below a concentration necessary to precipitate the proteinaceous substance. One suitable precipitant for crystallizing fragments of DPPIV and viral protein is polyethylene glycol (PEG), which combines some of the characteristics of the salts and other organic precipitants (see, for example, Ward et al., J. Mol. Biol. 98:161, 1975, and McPherson, J. Biol. Chem. 251:6300, 1976.

[0182] During a crystallization experiment, water is removed by diffusion or evaporation to increase the concentration of the precipitant, thus creating precipitating conditions for the protein. In one particular variation, crystals are grown by vapor diffusion in hanging drops or sitting drops. According to these methods, a protein/precipitant solution is formed and then allowed to equilibrate in a closed container with a larger aqueous reservoir having a precipitant concentration for producing crystals. The protein/precipitant solution continues to equilibrate until crystals grow.

[0183] By performing submicroliter volume sized crystallization experiments, as detailed in U.S. Patent No. 6,296,673, effective crystallization conditions for forming crystals of fragments of DPPIV and viral protein complex are obtained. In order to accomplish this, systematic broad screen crystallization trials are performed on a DPPIV/viral protein fragment complex using the sitting drop technique.

[0184] One skilled in the art will recognize that the crystallization conditions provided herein can be varied and still yield protein crystals comprising fragments of DPPIV and viral protein. As the conditions for the crystallization, physical and chemical factors such as a hydrogen ion concentration (pH), a kind of buffer used and a concentration thereof, a kind of a precipitant added and a concentration thereof, protein concentration, salt concentration, temperature and the like can be involved. A method for controlling and investigating the factors includes batch methods, dialysis methods, vapor diffusion methods (hanging drop method, sitting drop method and the like) and the like, described, for instance, in T.L. Blundell et al., PROTEIN CRYSTALLOGRAPHY, 59 82 (1976), published by Academic Press, or the like.

[0185] The method for crystallization includes the batch methods, dialysis methods, vapor diffusion methods and the like. By the above method, physical and chemical factors such as a hydrogen ion concentration (pH), a kind and a concentration of the buffer used, and a kind and a concentration of the precipitant used, and physical and chemical factors such as protein concentration, salt concentration and temperature can be also determined.

[0186] The hydrogen ion concentration (pH) can be adjusted with a buffer. It is desired that the buffer is a buffer having buffering action in a broad range of pH, and being capable of suppressing precipitation of a non proteinous crystal between the co existing ion in the solution used during crystallization and the precipitant or the like. The buffer includes Tris hydrochloric acid buffer, phosphate buffer, cacodylate buffer, acetate buffer, citrate buffer, glycine buffer and the like.

[0187] The precipitant may be a substance capable of elevating an effective concentration of the protein or changing a hydrogen ion concentration (pH) of the protein solution. Generally, the precipitant includes salts such as ammonium sulfate, sodium sulfate, sodium phosphate, potassium phosphate, sodium citrate, ammonium citrate, sodium chloride, potassium chloride and ammonium chloride; polyethylene glycols having various average molecular weights of about

200, about 1000, about 2000, about 4000, about 6000, about 8000, about 20000 or the like; organic solvents such as 2 methyl 2,4 pentadiol, methanol, ethanol, isopropanol, butanol and acetone, and the like.

[0188] The protein concentration may be a concentration suitable for crystallization, and it is desired that the protein concentration is, for example, 1 to 50 mg/ml, preferably 5 to 20 mg/ml, more preferably 7 to 15 mg/ml.

[0189] It is desired that the temperature conditions are 3°C to 25°C., preferably 12°C to 22°C.

[0190] In the case where the crystallization is carried out by the batch method, the crystallization can be carried out by gradually adding a precipitant solution comprising a precipitant, buffer and the like, so as to form a layer on the top layer of the solution containing the dipeptidyl peptidase to give a mixture, or by gradually adding the solution comprising the DPPIV/viral protein fragment complex, so that the solution is an upper layer of the precipitant solution to give a mixture. Here, the mixture is allowed to stand in a tightly closed vessel or container.

[0191] In the case where the crystallization is carried out by the dialysis method, the crystallization can be carried out by placing a solution comprising DPPIV/viral protein fragment complex in a size exclusion semi permeable membrane, and placing a precipitant solution outside of the size exclusion semi permeable membrane as a reservoir solution, thereby diffusing the reservoir solution to the solution comprising the DPPIV/viral protein fragment complex via the semi permeable membrane.

[0192] In the case where the crystallization is carried out by the hanging drop method in the vapor diffusion method, the crystallization can be carried out by placing a mixed solution of a solution comprising the DPPIV/viral protein fragment complex and a precipitant solution in a closed vessel allowing to be hanged at a position above the upper space of a reservoir in which the precipitant solution is contained as a reservoir solution, wherein the vapor pressure of the reservoir solution in the reservoir is set to be lower than that of the mixed solution.

[0193] In the case where the crystallization is carried out by the sitting drop method in the vapor diffusion method, the crystallization can be carried out by placing a mixed solution comprising a solution comprising the DPPIV/viral protein fragment complex and a precipitant solution in a closed vessel at a position higher than the liquid surface of a reservoir in which the precipitant solution is contained as a reservoir solution, wherein the vapor pressure of the reservoir solution in the reservoir is set to be lower than that of the mixed solution.

[0194] The crystallization can be carried out by the sitting drop method from the viewpoint of obtaining excellent quality and large crystals.

[0195] Crystals comprising fragments of DPPIV and viral protein have a wide range of uses. Such crystals may, for example, be used to perform X ray or neutron diffraction analysis in order to determine the three dimensional structure of fragments of DPPIV and viral protein and, in particular, to assist in the identification of its active site where fragments may bind. Knowledge of the binding site region allows rational design and construction of ligands including inhibitors. Crystallization and structural determination of fragments of DPPIV mutants and/or viral protein mutants having altered bioactivity allows the evaluation of whether such changes are caused by general structure deformation or by side chain alterations at the substitution site.

Example 18

[0196] Because DPPIV protein levels may not always accurately reflect the levels of active DPPIV enzyme, it may be useful to measure DPPIV enzymatic activity in proteinaceous substances instead. Use of a test system that is tested for DPPIV assay in proteinaceous substances as diverse as plasma, serum, urine, saliva, tissue, live cells and cell extracts, and exudates is recommended. Such a test system may be the DPPIV/CD26 Activity Assay for Biological Samples provided by ENZO® life sciences (on the World Wide Web at enzolifesciences.com). A known DPPIV inhibitor, such as P32/98 (Ki=130 nM) is preferably included for use as a control.

Example 19

[0197] To examine if cytokines decrease susceptibility to HCoV EMC infection through an effect on cell surface DPP4 expression, we analyzed DPP4 expression after treatment with different cytokines.

[0198] All treatments were done in quadruplets (96 well experiments) or triplicate (6 well and 24 well experiments). Cell cultures were grown for 24 to 48 hours and then changed to medium containing 1% newborn calf serum, and treated with recombinant human (r hu) IL 4 (BD Pharmingen), r hu IFN y, r hu TNF α, r hu IL 13, r hu IL 10, r hu IL 1, r hu TGF beta (Peprotech Inc.) and r hu IFN α (Roche) at a concentration of 10 ng/ml, 48 hours before infection for a further evaluation of changes in DPPIV surface protein expression and changes in susceptibility to HCoV EMC infection. In a first experiment, r hu TGF beta down regulates DPP4 expression and reduces the cells' susceptibility to virus infection and reduces virus replication.

Example 20

[0199] To examine if a compound decreases susceptibility to HCoV EMC infection through an effect on cell surface DPP4 expression, we analyze DPP4 expression after treatment with different compounds. Huh 7 cells are grown in Dulbecco's modified Eagle's medium (DMEM), supplemented with 10% fetal bovine serum (FBS), sodium bicarbonate and 20 mM HEPES buffer. All treatments are done in quadruplets (96 well experiments) or triplicate (6 well and 24 well experiments). Cultures are grown for 24 to 48 hours and then changed to medium containing 1% newborn calf serum, and treated with compound, i.e., adenosine (300 μM) or control vehicle for a further 48 hour evaluation of changes in DPPIV surface protein expression and changes in susceptibility to HCoV EMC infection. In a first experiment, adenosine down regulates DPP4 expression and reduces the cells' susceptibility to virus infection and reduces virus replication.

[0200] In a second experiment, inhibition of HCoV EMC replication in Huh7 cells by soluble adenosine deaminase (ADA) was demonstrated where inhibition with ACE2 was negative. Huh7 cells were incubated with different concentrations of recombinant soluble ADA or recombinant soluble ACE2. After 1 hour incubation, the cells were infected with HCoV EMC at a multiplicity of infection of 0.01. After 8 hours, cells were fixed and stained with a rabbit antiserum against HCoV EMC nsp4 and cells were counted. Results are shown as number of infected cells per well. Infection of Huh7 cells is inhibited by recombinant soluble ADA but not by recombinant soluble ACE2. The results are shown in figure 27.

[0201] In a third experiment, inhibition of HCoV EMC replication in Huh7 cells by soluble DPP4 was demonstrated. Different concentrations of recombinant soluble DPP4 or recombinant soluble ACE2 were incubated with HCoV EMC for 1 hour at 37°C and used to infect Huh7 cells. After 8 hours, cells were fixed and stained with a rabbit antiserum against HCoV EMC nsp4 and cells were counted. Results are shown as number of infected cells per well. Infection of Huh7 cells is inhibited by recombinant soluble DPP4 but not by recombinant soluble ACE2. The results are shown in figure 28.

Example 21

[0202] The spike (S) protein of the recently emerged human coronavirus (MERS CoV) mediates infection by binding to the cellular receptor dipeptidyl peptidase 4 (DPP4). Here, we mapped the receptor binding domain in the S protein to a 231 amino acid fragment (residues 358 588) by evaluating the interaction of spike truncation variants with receptor expressing cells and soluble DPP4.

[0203] Antibodies to this domain much less so to the preceding N terminal region efficiently neutralize MERS CoV infection. It is herein now also shown by co immunoprecipitation and FACS analyses that an internal region of the S1 of hCoV EMC consisting of 231 amino acids is sufficient to bind its receptor, DPP4. It was also shown that the region elicits the most neutralizing antibodies against the virus. Those results identified the receptor binding region of the S protein by convincing methods and the region contains major neuralization epitopes.

[0204] Additionally, the inventors herein further map the receptor binding domain (RBD) in the spike protein of the novel coronavirus EMC (hCoV EMC, now MERS CoV). Based on data obtained with bioinformatic tools they designed truncation variants of the S1 portion of hCoV EMC S (EMC S) and showed that the S1 variant harboring residues 358 588 i) co purifies with recombinant CD26 (the hCoV EMC receptor), binds to cellular CD26 in a FACS based assay and elicits neutralizing antibodies in immunized rabbits with higher efficiency than the wt S1 subunit.

[0205] Just 10 years following the outbreak of the severe respiratory acute syndrome coronavirus (SARS CoV), the world is confronted with yet another deadly human coronavirus. The virus, first provisionally called human coronavirus EMC (hCoV EMC) but now named MERS CoV, referring to its emergence in the Middle East and to the respiratory syndrome it causes, belongs to the betacoronavirus genus lineage 2c. It has thus far been identified in over 50 patients from or linked to the Arabian Peninsula, approximately half of them being fatal. Like with SARS CoV, patients affected by MERS CoV suffer from severe and often lethal lower respiratory tract infection. The epidemiology of MERS CoV is still enigmatic, but the geographical distribution of epidemiologically unlinked individuals points to intermittent, zoonotic transmission from a - so far unknown animal source, whereas a number of reported clusters indicate limited human to human spread.

[0206] The main determinant of coronavirus tropism is the viral spike (S) protein as it mediates binding to a cell surface receptor. The MERS CoV S protein, a 1353 amino acid type I membrane glycoprotein, assembles into trimers that constitute the spikes or peplomers on the surface of the enveloped coronavirus particle. The protein combines the two essential entry functions, namely that of host receptor binding and membrane fusion, which are attributed to the N terminal (S1, residues 1 751) and C terminal (S2, residues 752 1353) half of the S protein, respectively. Recently, we have identified dipeptidyl peptidase 4 (DPP4, also known as CD26), expressed in the human lung, as a functional receptor for MERS CoV. Importantly, MERS CoV can also use the evolutionary conserved DPP4 of other species, most notably that of bats.

[0207] Coronaviruses bind to receptors via independently folded, generally about 150 300 residues long, receptor binding domains (RBD) present in their S1 subunit, of which the location within S1 can vary. Thus, for the betacoronavirus

mouse hepatitis virus (MHV), the binding to its CEACAM receptor has been mapped to the N terminal -300 amino acids of the spike protein whereas for the SARS CoV of the same genus binding to the ACE2 receptor maps to residues 323 502 of S1. Identification of the RBD can hence help the development of monoclonal antibodies and vaccines for the treatment and prevention of infection. The RBD is the most important target for neutralizing antibodies preventing virus receptor interaction.

[0208] We previously used the S1 domain of MERS CoV fused to the Fc region of human IgG to demonstrate the interaction of S1 with DPP4 expressing cells and with soluble, i.e., non membrane anchored DPP4. To identify the receptor binding domain in the MERS CoV S1 subunit, we generated S1 Fc protein chimeras with truncations at the C terminus and N terminus of the S1 domain. We considered a three domain structure of the MERS CoV S1 protein (residues 1 357, 358 588 and 589 747) based on the predicted location and structure of the RBD of two other betacoronaviruses, MHV and SARS CoV, of which the homologous regions for MERS CoV S map to the residues 18 351 and 379 580, respectively. In addition, a soluble form of human DPP4 (residues 39 766) was made, which was C terminally tagged with the Fc region. These proteins were expressed in HEK 293T cells after transfection of the expression plasmids and subsequently affinity purified from the cell culture supernatant using protein A sepharose beads as described. The Fc region of purified sDPP4 Fc was proteolytically removed using trypsin (data not shown). First, we analyzed the S1 Fc proteins and C terminal S1 truncations thereof for their ability to interact with sDPP4 using a co purification assay. sDPP4 was efficiently co purified by the S1 Fc variants encompassing residues 1 588 and 1 747, whereas the 1 357 S1 Fc variant was unable to bind sDPP4. We next generated an S1 Fc variant comprising residues 358 588, a region homologous to the ACE2 receptor binding domain in SARS CoV S1. This S1 Fc truncation variant efficiently bound soluble DPP4, indicating that the DPP4 receptor binding domain is located within the 358 588 residues domain of the MERS CoV spike protein.

[0209] We subsequently tested the ability of these S1 Fc variants to bind to HEK 293T cells transiently expressing DPP4 by using flow cytometry. The S1 Fc variants encompassing residues 1 588 and 358 588 bound to DPP4 expressing HEK 293T cells with efficiencies comparable to the full length S1 protein, whereas no binding was observed with the 1 357 S1 Fc variant. These data show the 358 588 amino acids S1 region to be essential and sufficient for binding to DPP4 expressing cells, consistent with the results of the sDPP4 interaction study.

[0210] Finally polyclonal antibodies were raised in rabbits against the 1 747, 1 357 and 358 588 S1 Fc variants (Davids Biotechnology GmbH, Germany). The sera, which displayed equal ELISA titers towards its antigen (1:300,000, data not shown), were tested for their ability to neutralize virus infectivity. Antibodies raised against the 358 588 S1 Fc variant efficiently neutralized virus infectivity, superior to those raised against the 1 747 and 1 357 S1 Fc variants. This indicates that neutralizing epitopes within S1 are primarily localized to the RBD region. The elicited antibodies are likely to block the interaction of the spike protein with DPP4 thereby neutralizing MERS CoV infectivity. The results demonstrate the preferred potential of S1 protein and of the 358 588 S1 polypeptide or functional fragments thereof reactive with the MERS CoV neutralizing antibody for use as subunit vaccines with a high biosafety profile compared to vaccines based on inactivated viruses or live attenuated virus.

[0211] Except for the betacoronavirus MHV, which binds to its CEACAM receptor through a domain in the N terminal part of its S1 protein, the RBDs of all other coronaviruses that engage protein receptors and that have been mapped occur in the C terminal portion of the S1 subunit. Examples also include the alphacoronaviruses binding to ACE2 (hCoV NL63) and APN (e.g., TGEV, hCoV 229E). In this study, we have experimentally mapped the RBD of MERS CoV to a 231 amino acid fragment (residues 358 588) within the spike protein. This domain nicely corresponds with the S1 region recently anticipated to interact with the DPP4 receptor on the basis of theoretical S1 structure predictions. The RBD in the MERS CoV S1 protein localizes in the same region where the SARS CoV S protein interacts with its ACE2 receptor. The SARS CoV RBD structure displays a five stranded 6 sheet core structure (β1 4 and β7) maintaining the overall domain conformation, and a long extended loop containing two anti parallel 6 sheets (65 and 66) responsible for receptor binding{{}}. Intriguingly, compared to SARS CoV, the RBD of MERS CoV contains a relatively conserved core domain but a highly variable loop region, tentatively explaining the differential receptor usage. Crystallization and structure analysis of this MERS CoV RBD region in complex with DPP4 will give detailed insight into the virus receptor binding interface.

Example 22

Dipeptidyl peptidase 4 receptor determinants of respiratory MERS-coronavirus infection

[0212] Here we show that MERS coronavirus (MERS-CoV) replicates in cells of different species using dipeptidyl peptidase 4 (DPP4) as a functional receptor. This suggests a broad host species tropism allowing zoonotic transmission from many animal species. Here we show contrasting DPP4 receptor functionality in different animal species. Resistance of ferrets to MERS-CoV infection was due to the inability to bind MERS-CoV as a result of amino acid variation in the ferret DPP4 β-propeller region. In contrast, DPP4 expressing respiratory epithelial cells in the lower - but not upper -

respiratory tract of cynomolgus macaques were targeted by MERS-CoV, which resulted in relatively mild disease. Variable DPP4 expression and adenosine deaminase (ADA) - shown to act as a natural antagonist for MERS-CoV infection - may potentially modulate MERS-CoV infection. Our findings illuminate the role of DPP4 sequence and expression variability in host range restriction and outcome of respiratory MERS-CoV infection and lead us to conclude that MERS-CoV receptor sequence and expression variability determine host range restriction of lower respiratory MERS-CoV infection.

[0213] Coronaviruses (CoVs) usually cause common colds in humans but zoonotic transmission occasionally introduces more pathogenic viruses into the human population as was demonstrated by the severe acute respiratory syndrome (SARS) outbreak. In 2012 a previously unknown human coronavirus (CoV), now named Middle East respiratory syndrome CoV (MERS-CoV), was isolated from the sputum of a 60-year-old man in Saudi Arabia who presented with acute pneumonia with a fatal outcome. To date, several infection clusters have been reported over a one-year period with around 50% of the reported human cases being fatal. Although limited human-to-human transmission has been observed, it is feared that by acquiring additional mutations MERS-CoV may spread more easily.

[0214] MERS-CoV represents a novel betacoronavirus species with the closest known relatives being clade 2c bat CoVs, detected in diverse species of bats but not yet in any animal species from the Arabian Peninsula. Although MERS-CoV replicates in cells of different species including bats, pigs and (non-) human primates, its ability to infect different animal species may be restricted given the fact that hamsters were shown to resist MERS-CoV infection. Therefore, a further understanding of factors that determine host restriction and viral transmission need to be revealed by studies in different animal species.

[0215] Herein we identified dipeptidyl peptidase 4 (DPP4) as a functional MERS-CoV receptor in human and bat cells. To further analyse DPP4 usage by MERS-CoV in vivo, ferrets (n = 4), known to be susceptible to several respiratory viruses including SARS-CoV and influenza viruses, were inoculated intratracheally with MERS-CoV. The animals did not seroconvert and only low levels of virus were detected by RT-qPCR in respiratory swabs at 1-2 days post infection (dpi). In vitro, ferret primary kidney cells could not be infected with MERS-CoV despite DPP4 surface expression, while transfection of these cells with human DPP4 (hDPP4) rendered the cells susceptible, suggesting that ferret DPP4 (fDPP4) does not efficiently bind MERS-CoV. Consistently, MDCK cells transfected with fDPP4 did not bind to synthetic MERS-CoV spike (S1) protein and were not infected by the virus (Fig 33B,C). DPP4 is an ectoenzyme that cleaves dipeptides from hormones, chemokines and cytokines by its conserved C-terminal α/β-hydrolase domain of the protein, while its N-terminal eight-blade β-propeller domain contains more sequence variability. By constructing DPP4 chimeras we observed that the blades 4 and 5 containing hDPP4 domain (residues 246-505) could confer to ferret DPP4 the ability to bind to S1 and to mediate MERS-CoV infection when expressed in non-susceptible cells(Fig33B,C). A Quick Change site-directed mutagenesis kit (Stratagene) was used to construct different hDPP4 point mutants. The presence of the correct mutations and absence of undesired mutations was confirmed by sequencing analysis. Plasmids were transfected into MDCK cells in triplicate, after 24 h incubation individual wells were split to determine DPP4 cell surface expression, S1-binding and susceptibility to MERS-CoV infection on the same transfected cell culture. Consistently, substitution of selected solvent exposed residues present in blades 4 and 5 of hDPP4 by those occurring at these positions in fDPP4, abrogated DPP4's capacity to bind to S1 and to mediate MERS-CoV cell susceptibility upon transfection, suggesting that these residues are involved in MERS-CoV binding and entry (Fig33D,E). Reciprocal substitutions of these amino acids in fDPP4 however, did not confer S1 binding, demonstrating the complexity of the interaction in the face the highly polymorphic nature of these two blades. The identified residues also are critical in binding the human enzyme adenosine deaminase (ADA), a natural DPP4 ligand that is involved in the development and maintenance of the immune system. Using recombinant ADA, significant inhibition of MERS-CoV infection and spike protein binding was demonstrated revealing a natural occurring antagonist able to block MERS-CoV infection. The data on the co-crystallization of the receptor binding domain of S1 and DPP4 are in line with the data presented. Phylogenetic analysis of the virus binding region of DPP4 indicated that human, macaque, horse and rabbit DPP4 cluster together as do DPP4's from cattle, pig and bats, that are somewhat more distantly related. Small animals including ferret, mice and most likely hamsters, shown to resist MERS-CoV infection, are more divergent in the DPP4 virus binding region, which at least in the case of ferrets has consequences for MERS-CoV binding.

[0216] Considering the highly conserved virus binding region in macaque DPP4 as compared to hDPP4, we first confirmed the use of cynomolgus macaque DPP4 as a functional MERS-CoV receptor. DPP4 antibodies blocked MERS-CoV infection of macaque primary kidney cells in vitro. Besides macaques, rabbits may be a potential animal model for MERS-CoV infection; ex vivo inoculation of rabbit lung and kidney tissues revealed susceptibility to MERS-CoV. We subsequently inoculated ten young adult cynomolgus macaques intratracheally with MERS-CoV and euthanized them at 1 (n = 4, macaques 1-4), 4 (n = 4, macaques 5-8) and 28 dpi (n = 2, macaques 9 and 10). All animals remained free of severe clinical signs and maintained a rhythmic pattern of body temperatures fluctuating between 35°C (night) and 39°C (day) that seemed slightly elevated after inoculation. Neutralizing antibodies with titers 40-80 were detected in the two MERS-CoV infected macaques that were euthanized at 28 dpi. Upon necropsy, there were a few mild focal red-grey slightly depressed areas affecting less than 5 % of the lung tissue, although one lobe of macaque 7 had a dark red

rim with evidence of suppurative bronchopneumonia, consistent with the detection of Escherichia coli bacteria in this lobe. MERS-CoV mRNA was detected at highly variable levels in pharyngeal and nasal swabs on 1 to 11 dpi and at low levels in rectal swabs on 2 and 3 dpi. In addition, MERS-CoV was detected by RT-qPCR in the lungs, nasal septum, serum and spleen and in one animal - macaque 1 - also in the kidney, liver, colon and urine at 1 dpi. Infectious virus was detected only in one pharyngeal swab sample and to a limited extent in the lungs. Using a probe that targets the MERS-CoV nucleocapsid gene, hybridization was observed in epithelial cells in bronchioles, and in moderate numbers of type 2 and few type 1 pneumocyte-resembling cells in the alveoli at 1 dpi while at 4 dpi very few cells were found positive. Consistent with activation of cytokines like CCL3, the lungs showed mild alveolitis, characterized by thickening of the alveolar septa with infiltration of few neutrophils and macrophages and moderate type 2 hypertrophy and hyperplasia at 4 dpi. In the alveolar lumina there were increased numbers of alveolar macrophages and occasionally small amounts of edematous fluid with fibrin and few neutrophils. Consistent with the capacity of the virus to induce syncytia in vitro, syncytial cells were seen. By applying a technique that enables successive staining of the same tissue section, tropism of MERS-CoV for cells expressing DPP4 in vivo was demonstrated. Thus, the experimental infection of young adult macaques with MERS-CoV revealed that macaque DPP4 positive cells in the lower respiratory tract can be infected with MERS-CoV but the associated pathological changes are relatively mild, indicating that young adult macaques are at best a suboptimal MERS-CoV animal model for the often fatal MERS-CoV infection in humans.

[0217] Abundant ACE2 expression in the respiratory tract has been suggested to facilitate rapid spread of SARS-CoV, a critical factor in the rapid induction of innate immune responses that drive the acute respiratory distress syndrome. In non-infected macaques DPP4 expression was restricted to non-ciliated cells, type 2 cells and endothelial cells whereas no staining was observed in ciliated epithelial cells of the (upper) respiratory tract. The absence of DPP4 on the upper respiratory tract epithelial cells, consistent with the inability to detect viral antigen in these cells, therefore may limit efficient virus transmission through the upper respiratory route. Kidneys, liver, intestine, and sub mucosal glands of the upper respiratory tract were found to contain varying levels of DPP4, which mainly localized to the apical side of the cells. Enhanced DPP4 expression was observed in the lungs of the bacterial co-infected macaque 7, which excreted infectious virus in the pharyngeal swab and displayed a higher body temperature. We observed that LPS stimulation of in vitro differentiated macrophages enhanced DPP4 expression. Attempts to infect these cells were unsuccessful, likely due to ADA production by these cells. Interestingly, DPP4 was virtually absent in the lower respiratory tract epithelium of ferrets but could be visualized in the kidneys of these animals. Contrastingly, relatively strong DPP4 expression was observed on different cell types in human lungs, including a MERS-CoV infected individual. In several pathological conditions such as viral infections and type 2 diabetes increased levels of (soluble) DPP4 have been demonstrated. Thus, relatively low levels of DPP4 expression in the lungs of young adult macaques could partly explain the mild infection observed after MERS-CoV infection but further studies need to reveal the role of varying DPP4 and ADA expression levels in regulating MERS-CoV replication in vivo.

[0218] Our findings demonstrate that the host range potential of the emerging novel human MERS-CoV is primarily determined by the MERS-CoV binding to and tissue distribution of DPP4. The co-localisation of DPP4 with MERS-CoV in the lower respiratory tract of MERS-CoV infected non-human primates (in bronchioles and alveoli), and the inability to infect ferrets further supports the sole involvement of DPP4 as a functional receptor in MERS-CoV entry. Variable levels of DPP4 expression in the lower respiratory tract may impose MERS-CoV host range restriction and explain why studies in rhesus macaques have not been successful to reproduce the severe disease seen in humans. Future studies need to unravel the significance of variable DPP4 expression in MERS-CoV patients, for example as a result of co morbidities like microbial infections, type 2 diabetes or aging.

Material and Methods

[0219] **Cloning of human and ferret DPP4.** The hDPP4 cDNA was obtained as described. Total RNA was isolated from ferret primary kidney cells using RNeasy mini kit (Qiagen) and cDNAs were synthesized by using the Superscript reverse transcriptase (Life Technologies). The complete DPP4 genes were amplified using Pfu Ultra II fusion HS DNA polymerase (Stratagene) and cloned into the pcDNA 3.1 expression vector (Life Technologies). Human to ferret DPP4 mutants of cDNA constructs were made by utilizing unique restriction enzyme sites shared by human and ferret DPP4. Pst I can cut human and ferret DPP4 into three fragments (human, amino acid 1-246, 247-504 and 505-766 and ferret, amino acid 1-245, 246-503 and 504-765). The middle fragment of human and ferret DPP4 was exchanged between human and ferret, the final plasmid constructs contained different combinations of fragments: human-ferret-human (HFH) or ferret-human-ferret (FHF). A Quick Change site-directed mutagenesis kit (Stratagene) was used to construct different hDPP4 point mutants. The presence of the correct mutations and absence of undesired mutations was confirmed by sequencing analysis. Plasmids were transfected into MDCK cells in triplicate, after 24 h incubation individual wells were split to determine DPP4 cell surface expression, S1-binding and susceptibility to MERS-CoV infection on the same transfected cell culture. S1 binding and infection were corrected for DPP4 cell surface expression as determined by the goat polyclonal antiserum against DPP4 (R&D systems), a secondary FITC conjugated rabbit anti goat serum followed

by FACS analysis.

[0220] Phylogenetic analysis of DPP4. Sequence alignment was performed by using ClustalW in the MEGA5.0 software package (www.megasoftware.net), and the trees were constructed by using the neighbor-joining method with p-distance (gap/missing data treatment; complete deletion) and 1,000 bootstrap replicates as in MEGA version 5.0.

[0221] Protein expression and S1 binding assay. A plasmid encoding MERS-CoV S1-Fc was generated by ligating a fragment encoding the S1 domain (residues 1-747) 3'-terminally to a fragment encoding the Fc domain of human IgG into the pCAGGS expression vector. Likewise, an S1-Fc expression plasmid was made the FIPV S1 domain (isolate 79-1146; residues 1-788). Fc chimeric proteins were expressed by transfection of the expression plasmids into HEK-293T cells and affinity purified from the culture supernatant using Protein A Sepharose beads (GE Healthcare). S1 binding of cells was measured by incubating 105 cells with 15 mg/ml of S1-Fc followed by incubation with FITC or DyLight-488-labelled goat-anti-human IgG antibody and analysis by flow cytometry.

[0222] Virus infection experiments. Virus stocks of MERS-CoV (EMC isolate) were prepared. Transfected COS-7 cells, Huh-7 and primary ferret and macaque kidney cells were inoculated with MERS-CoV for 1 h with high MOI. After washing the cells were incubated with medium containing 1% fetal bovine serum. Alternatively we used thin cut slices from the lungs and kidneys of rabbits that were incubated in culture medium with virus for 24h. At 8 or 24 h after infection cells were fixed with formaldehyde and stained using rabbit-anti-SARS-CoV NSP4 antibodies that are cross-reactive for hCoV-EMC, according to standard protocols using a FITC conjugated swine-anti-rabbit antibody as a second step. Primary ferret or macaque kidney cells were preincubated with antibodies to DPP4 (polyclonal goat-anti DPP4 immunoglobulin, R&D systems) at 20 μg/ml to block MERS-CoV infection. Recombinant human ADA (R&D systems) was preincubated with hDPP4 transfected cells or Huh7 cells for 1 h after which the cells were infected with MERS-CoV for 8 h and processed.

[0223] Animal studies. Ten cynomolgus macaques (Macaca fascicularis), 3-5 years old with active temperature transponders in the peritoneal cavity (n = 3), were inoculated with 5 × 106 TCID50 of MERS-CoV via the intranasal and intratracheal route. In addition, four ferrets (Mustello fuoris) were inoculated with 1 × 106 TCID50 of MERS-CoV via the intranasal and intratracheal route. Animals were checked daily for clinical signs. Just before infection and at different dpi, animals were anesthetized with ketamine and nasal, pharyngeal, and rectal swabs were taken, which were placed in 1 ml Dulbecco's modified Eagle's medium supplemented with 100 IU penicillin/ml and 100 □g of streptomycin/ml (virus transport medium) and frozen at -70°C until RT-PCR analysis. The animals were euthanized at different days (Day 1, 4 or 28) p.i. by exsanguination under ketamine anesthesia. Approval for animal experiments was obtained from the Institutional Animal Welfare Committee (nr EMC 2808).

[0224] Necropsies were performed according to a standard protocol. For semi-quantitative assessment of gross pathology, the percentage of affected lung tissue from each lung lobe was determined at necropsy, recorded on a schematic diagram of the lung and the area of affected lung tissue was subsequently calculated (gross pathology score). One lung of each monkey was inflated with 10% neutral-buffered formalin by intrabronchial intubation and suspended in 10% neutral-buffered formalin overnight. Samples were collected in a standard manner (from the cranial, medial and caudal parts of the lung), embedded in paraffin, cut at 3 □m and used for immunohistochemistry (see below) or stained with hematoxylin and eosin (H&E). The lung, liver, spleen, kidney, intestine, trachea, and tracheobronchial lymphnode H&E sections were examined by light microscopy.

[0225] In situ hybridization. The ISH probes targeting the nucleocapsid gene of MERS-CoV were designed by Advanced Cell Diagnostics (Hayward, CA) and ISH was performed according to the manufacturer's instructions and ISH staining was visualized using substrate Fast Red (pink). Controls included probes against SARS-CoV nucleocapsid protein and tissues from non infected animals.

[0226] Imunohistochemistry. Family consent was granted for limited postmortem tissue retrieval from a MERS-CoV patient in the UK, consisting of a 20-cm-long midline incision in lower chest and upper abdomen, from which tissue samples were collected from both lungs. Archival paraffin-embedded human tissue sections were obtained from the Department of Pathology, Erasmus MC. Four historic macaque controls were used as mock (PBS) infected. For histological analysis, samples were placed in 10% neutral-buffered formalin and further processed for routine immunohistochemistry. Serial 3 μm lung sections were stained using according to standard protocols using antibodies to DPP4 (polyclonal goat-anti DPP4 immunoglobulin, R&D systems. For phenotyping to test DDP4 expression of MERS-CoV infected cells, we used a destaining-restaining technique. Briefly, the precipitate used for visualization of MERS-CoV antigen staining was dissolved in graded 100%-70% alcohols. To detach the primary antibody and red immunohistochemistry signals, slides were soaked in eluding buffer (5ml 0.1M HCl, 95 ml 0.1M NaCl containing 1M glycine) for 2 hours. The sections were treated with two 5 min intervals heating in citric acid buffer pH 6.0 to denature any undetached primary antibody. The slides were then incubated with antibodies against DPP4 in PBS/0.1% BSA for 1 hour at RT. After washing, sections were incubated with horseradishperoxidase labeled anti-goat IgG 1/100 in PBS/0.1% BSA for 1 hour at RT. Peroxidase activity was revealed by incubating slides in 3,3'-diaminobenzidine-tetrachlorhydrate (DAB) (Sigma) for 3-5 minutes, resulting in a brown precipitate, followed by counterstaining with hematoxylin.

[0227] RNA-extraction and quantitative RT-PCR. Samples were analysed with the upE PCR and confirmed by a

nucleocapsid specific PCR. RNA from 200 □l of culture supernatant was isolated with the Magnapure LC total nucleic acid isolation kit (Roche) and eluted in 100 □l. MERS-CoV RNA was quantified on the ABI prism 7700, with the TaqMan® Fast Virus 1-Step Master Mix (Applied Biosystems) using 20 □l isolated RNA, 1×Taqman mix, 0.5U uracil-N-glycosylase, 45 pmol forward primer (5'-GGGTGTACCTCTTAATGCCAATTC-3'), 45 pmol reverse primer (5'-TCTGTCCTGTCTC-CGCCAAT-3') and 5 pmol probe (5'-FAM-ACCCCTGCGCAAAATGCTGGG-BHQ1-3'). Amplification parameters were 5 min at 50°C, 20 sec at 95°C, and 45 cycles of 3 s at 95°C, and 30 sec at 60°C. RNA dilutions isolated from an MERS-CoV stock were used as a standard.

[0228] Lung tissue samples (0.3-0.5 gram) were taken for RT-PCR and microarray analysis in RNA-later (Ambion, Inc.). RNA was isolated from homogenized post mortem tissue samples using Trizol Reagent (Invitrogen) and the RNeasy mini kit (Qiagen). cDNA synthesis was performed with ~1 □g total RNA and Superscript III RT (Invitrogen) with oligo(dT), according to the manufacturer's instructions. Semi-quantitative RT-PCR was performed as described previously to detect MERS-CoV and to validate cellular gene expression changes as detected with microarrays of CCL3 (Applied Biosystems). Differences in gene expression are represented as the fold change in gene expression relative to a calibrator and normalized to a reference. GAPDH (glyceraldehydes-3-phosphate dehydrogenase) was used as endogenous control to normalize quantification of the target gene. The samples from the mock-infected macaques were used as a calibrator. Average results (□ s.e.m.) for groups were expressed as fold change compared to PBS-infected animals.

[0229] **Macrophage cultures.** Monocytes isolated from peripheral blood mononuclear cells were cultured with GM-CSF for 5 days to generate macrophages. Subsequently cells were stimulated with LPS at 1μg/ml for 24h and processed for DPP4 staining and FACS analysis.

[0230] **Statistical analysis.** Data were compared using one way ANOVA with post-test Bonferroni. Statistical analysis was performed with Prism 4.0 (Graphpad).

References to example 22

[0231]

1. J. S. Peiris, Y. Guan, K. Y. Yuen, Nat. Med. 10, S88-97 (2004).
2. A. M. Zaki, S. van Boheemen, T. M. Bestebroer, A. D. Osterhaus, R. A. Fouchier, N. Engl. J. Med. 367, 1814-1820 (2012).
3. S. van Boheemen et al., MBio 3, 00473-12 (2012).
4. World Health Organisation. Middle East respiratory syndrome coronavirus (MERS-CoV)-update, http://www.who.int/csr/don/2013_08_01/en/index.html (8 August 2013).
5. A. Assiri et al., N. Engl. J. Med. 369, 407-416 (2013).
6. B. Guery et al., Lancet 381, 2265-2272 (2013).
7. C. B. Reusken et al., Vector Borne Zoonotic Dis 10, 785-791 (2010).
8. A. Annan et al., Emerg Infect Dis 19, 456-459 (2013).
9. M. A. Muller et al., MBio 3, 00515-12 (2012).
10. J. F. Chan et al., J. Infect. Dis. 207, 1743-1752 (2013).
11. E. de Wit et al., PLoS One 8, e69127 (2013).
12. V. S. Raj et al., Nature 495, 251-254 (2013).
13. B. E. Martina et al., Nature 425, 915 (2003).
14. J. A. Belser, J. M. Katz, T. M. Tumpey, Dis. Model. Mech. 4, 575-579 (2011).
15. E. Boonacker, C. J. Van Noorden, Eur. J. Cell. Biol. 82, 53-73 (2003).
16. W. A. Weihofen, J. Liu, W. Reutter, W. Saenger, H. Fan, J. Biol. Chem. 279, 43330-43335 (2004).
17 R.P. Dong et al, J. Immunol. 156, 1349-1355 (1996).
18. G. Lu et al., Nature doi:nature12328 (2013).
19. N. Wang et al., Cell. Res. doi:cr201392 (2013).
20. B. Rockx et al., J. Virol. 83, 7062-7074 (2009).
21. S. L. Smits et al., PLoS Pathog. 6, e1000756 (2010).
22. J. Zhong et al., Diabetes 62, 149-157 (2013).
23. B. A. Conlon, W. R. Law, Clin. Exp. Immunol. 138, 14-20 (2004).
24. S. A. Lee et al., J. Clin. Endocrinol. Metab. 98, 2553-2561 (2013).
25. T. Andrieu et al., J. Clin. Virol. 27, 59-68 (2003).
26. V. J. Munster, E. de Wit, H. Feldmann, N. Engl. J. Med. 368, 1560-1562 (2013).
27. S. L. Smits et al., J. Virol. 85, 4234-4245 (2011).
28. V. M. Corman et al., Euro. Surveil.l 17, pii: 20334 (2012).

Claims

1. An essentially mammalian positive-sense single stranded RNA virus which is a betacoronavirus, comprising all of the amino acid sequences selected from figure 5 file N.rtf depicting the nucleocapsid (N) protein, figure 6 file M.rtf depicting the matrix (M) protein, figure 7 file E.rtf depicting the small envelope (E) protein, figure 8 file NS3d.rtf depicting the non-structural gene NS3d, figure 9 file NS3c.rtf depicting the non-structural gene NS3c, figure 10 file NS3b.rtf depicting the non-structural gene NS3b, figure 11 file NS3a.rtf depicting the non-structural gene NS3a, figure 12, file S.rtf depicting the spike surface glycoprotein (S), figure 13 file Orflab.rtf, encoding many enzymatic products among which the replicase or comprising the nucleic acid sequence of figure 14 file HCoV-SA1.rtf depicting isolate HCoV-SA1,.

2. An essentially mammalian positive-sense single stranded RNA virus which is a betacoronavirus, and identifiable as phylogenetically corresponding thereto by determining the amino acid sequence of the conserved replicase domain of said virus to have at least 90% identity with the Orf1AB amino acid sequence as depicted in Fig. 13.

3. A virus according to claim 1 or claim 2 that comprises the nucleotide sequence as depicted in figure 14.

4. A virus according to any of claims 1-3 isolatable from humans.

5. A nucleic acid, preferably a cDNA, encoding a protein as defined in claim 1 or a protein having at least 90% identity with the Orf1AB amino acid sequence as depicted in Fig. 13.

6. A vector comprising a nucleic acid according to claim 5.

7. A cell comprising a virus according to anyone of claims 1 to 4, a nucleic acid according to claim 5 or a vector according to claim 6.

8. A protein as depicted in any of the figures 5-13, or a protein having at least 90% identity with the Orf1AB amino acid sequence as depicted in Fig. 13.

9. An antigen comprising a protein according to claim 8.

10. An antibody specifically directed against a protein according to claim 8.

11. A method for identifying a viral isolate as a MERS-CoV comprising reacting said viral isolate or a component thereof with a nucleic acid according to claim 5 and/or with an antibody according to claim 10.

12. A method for virologically diagnosing a MERS-CoV infection of a mammal comprising determining in a sample of said mammal the presence of a viral isolate or component thereof by reacting said sample with a nucleic acid according to claim 5 or an antibody according to claim 10.

13. A method for serologically diagnosing a MERS-CoV infection of a mammal comprising determining in a sample of said mammal the presence of an antibody specifically directed against a Betacoronavirus, preferably Lineage C virus or component thereof by reacting said sample with a proteinaceous molecule or fragment thereof according to claim 8 or an antigen according to claim 9.

14. A diagnostic kit for diagnosing a MERS-CoV infection comprising a virus according to anyone of claims 1 to 4, and/or a nucleic acid according to claim 5, and/or a protein according to claim 8, and/or an antigen according to claim 9 and/or an antibody according to claim 10.

15. A virus according to any one claims 1 to 4, and/or a nucleic acid according to claim 5, and/or a vector according to claim 6, and/or a cell according to claim 7, and/or a protein according to claim 8, and/or an antigen according to claim 9 and/or or an antibody according to claim 10 for use in the treatment or prevention of a Betacoronavirus infection with a MERS-CoV.

16. A pharmaceutical composition comprising virus according to any one claims 1 to 4, and/or a nucleic acid according to claim 5, and/or a vector according to claim 6, and/or a protein according to claim 8, and/or an antigen according to claim 9 and/or or an antibody according to claim 10.

17. A pharmaceutical composition according to claim 16 for use in the treatment or prevention of a MERS-CoV infection in a mammal.

18. A pharmaceutical composition for use according to claim 17, wherein said mammal is a human.

19. A proteinaceous substance, comprising a protein according to claim 8, wherein said protein is a spike protein from figure 12 and additionally comprising at least a fragment of an N-terminal dipeptidyl peptidase protein wherein said fragment is derived from the ectodomain.

20. A proteinaceous substance according to claim 19, wherein said substance is crystallized.

21. A proteinaceous substance according to claim 19 or 20, wherein said dipeptidyl peptidase protein is a dipeptidyl peptidase 4 (DPP4) and preferably wherein the fragment comprises residues 39 - 766 of human DPP4.

22. A proteinaceous substance according to any of claims 19 - 21, wherein said proteinaceous molecule comprises an ectodomain of a spike protein.

23. A proteinaceous substance according to claim 22, wherein said ectodomain is derived from the S1 region of a coronavirus.

24. A container provided with a virus according to any one claims 1 to 4, and/or a nucleic acid according to claim 5, and/or a vector according to claim 6, and/or a cell according to claim 7, and/or a protein according to claim 8, and/or an antigen according to claim 9 and/or or an antibody according to claim 10, and/or a pharmaceutical composition according to claim 16 and/or a substance according to any of claims 19 - 23.

25. A method of identifying a candidate modulator or a candidate antiviral agent as an agent that modulates the function of or the binding of a virus to a dipeptidyl peptidase, said method comprising:

a. Providing a substance according to any of claims 19 - 23 in the presence and absence of said candidate modulator or said candidate antiviral agent under conditions permitting binding of a the protein derived from a virus with the fragment derived from a peptidase protein;
b. Measuring binding of said protein to said fragment, wherein a decrease or increase in binding in the presence of said candidate modulator or said antiviral agent, relative to binding in the absence of said candidate modulator, identifies said candidate modulator as an agent that modulates the function of a dipeptidyl peptidase or identifies said antiviral agent as an agent that modulates the function of a dipeptidyl peptidase, preferably wherein said protein and/or said fragment is detectably labeled, preferably with a moiety selected from the group consisting of a radioisotope, a fluorophore, a quencher of fluorescence, an enzyme, and an affinity tag.

Patentansprüche

1. Im Wesentlichen positiv-gerichtetes einzelsträngiges RNA-Virus von einem Säugetier, das ein Betacoronavirus ist, umfassend alle der Aminosäuresequenzen, ausgewählt aus Figur 5 Datei N.rtf, darstellend das Nucleocapsid- (N) Protein, Figur 6 Datei M.rtf, darstellend das Matrix- (M) Protein, Figur 7 Datei E.rtf , darstellend das kleine Hüll- (E) Protein, Figur 8 Datei NS3d.rtf, darstellend das nichtstrukturelle Gen NS3d, Figur 9 Datei NS3c.rtf, darstellend das nichtstrukturelle Gen NS3c, Figur 10 Datei NS3b.rtf, darstellend das nichtstrukturelle Gen NS3b darstellt, Figur 11 Datei NS3a.rtf, darstellend das nichtstrukturelle Gen NS3a, Figur 12 Datei S.rtf, darstellend das Spike-Oberflächenglykoprotein (S), Figur 13 Datei Orf1ab.rtf, kodierend viele enzymatische Produkte, darunter die Replikase oder umfassend die Nukleinsäuresequenz von Figur 14 Datei HCoV-SA1.rtf, darstellend Isolat HCoV-SA1.

2. Im Wesentlichen positiv-gerichtetes einzelsträngiges RNA-Virus von einem Säugetier, das ein Betacoronavirus ist und identifizierbar als phylogenetisch diesem entsprechend, durch Bestimmender Aminosäuresequenz der konservierten Replikasedomäne des Virus, wenigstens 90 % Identität mit der OrflAB-Aminosäuresequenz, wie in Figur 13 dargestellt, zu haben.

3. Virus nach Anspruch 1 oder Anspruch 2, das die Nukleotidsequenz, wie in Figur 14 dargestellt, umfasst.

4. Virus nach einem der Ansprüche 1 - 3, isolierbar aus Menschen.

5. Nucleinsäure, vorzugsweise eine cDNA, kodierend ein Protein wie definiert in Anspruch 1 oder ein Protein mit wenigstens 90 % Identität mit der OrflAB-Aminosäuresequenz, wie dargestellt in Figur 13 .

6. Vektor, umfassend eine Nukleinsäure nach Anspruch 5.

7. Zelle, umfassend ein Virus nach einem der Ansprüche 1 bis 4, eine Nukleinsäure nach Anspruch 5 oder einen Vektor nach Anspruch 6.

8. Ein Protein, wie dargestellt in einer der Figuren 5-13, oder ein Protein mit wenigstens 90 % Identität mit der OrflAB-Aminosäuresequenz, wie dargestellt in Figur 13.

9. Antigen, umfassend ein Protein nach Anspruch 8.

10. Antikörper, spezifisch gerichtet gegen ein Protein nach Anspruch 8.

11. Verfahren zum Identifizieren eines Virusisolats, wie ein MERS-CoV, umfassend Umsetzen des Virusisolats oder einer Komponente davon mit einer Nukleinsäure nach Anspruch 5 und/oder mit einem Antikörper nach Anspruch 10.

12. Verfahren zum virologischen Diagnostizieren einer MERS-CoV-Infektion eines Säugetiers, umfassend Bestimmen in einer Probe des Säugetiers die Anwesenheit eines Virusisolats oder Komponente davon durch Umsetzen der Probe mit einer Nukleinsäure nach Anspruch 5 oder einem Antikörper nach Anspruch 10.

13. Verfahren zum serologischen Diagnostizieren einer MERS-CoV-Infektion eines Säugetiers, umfassend Bestimmen in einer Probe des Säugetiers die Anwesenheit eines Antikörpers, spezifisch gerichtet gegen ein Betacoronavirus, vorzugsweise Lineage C-Virus oder Komponente davon, durch Umsetzen der Probe mit einem proteinhaltigen Molekül oder Fragment davon nach Anspruch 8 oder einem Antigen nach Anspruch 9.

14. Diagnose-Kit zum Diagnostizieren einer MERS-CoV-Infektion, umfassend ein Virus nach einem der Ansprüche 1 bis 4, und/oder eine Nukleinsäure nach Anspruch 5, und/oder ein Protein nach Anspruch 8, und/oder ein Antigen nach Anspruch 9, und/oder ein Antikörper nach Anspruch 10.

15. Virus nach einem der Ansprüche 1 bis 4, und/oder eine Nukleinsäure nach Anspruch 5, und/oder einen Vektor nach Anspruch 6, und/oder eine Zelle nach Anspruch 7, und/oder ein Protein nach Anspruch 8, und/oder ein Antigen nach Anspruch 9, und/oder ein Antikörper nach Anspruch 10 zur Verwendung bei der Behandlung oder Prävention einer Betacoronavirus-Infektion mit einem MERS-CoV.

16. Pharmazeutische Zusammensetzung, umfassend ein Virus nach einem der Ansprüche 1 bis 4, und/oder eine Nukleinsäure nach Anspruch 5, und/oder einen Vektor nach Anspruch 6, und/oder ein Protein nach Anspruch 8, und/oder ein Antigen nach Anspruch 9 und/oder einen Antikörper nach Anspruch 10.

17. Pharmazeutische Zusammensetzung nach Anspruch 16 zur Verwendung bei der Behandlung oder Prävention einer MERS-CoV-Infektion bei einem Säugetier.

18. Pharmazeutische Zusammensetzung zur Verwendung nach Anspruch 17, wobei das Säugetier ein Mensch ist.

19. Proteinhaltige Substanz, umfassend ein Protein nach Anspruch 8, wobei das Protein ein Spike-Protein aus Figur 12 ist und zusätzlich umfassend wenigstens ein Fragment eines N-terminalen Dipeptidylpeptidase-Proteins, wobei das Fragment von der Ektodomäne abgeleitet ist.

20. Proteinhaltige Substanz nach Anspruch 19, wobei die Substanz kristallisiert ist.

21. Proteinhaltige Substanz nach Anspruch 19 oder 20, wobei das Dipeptidylpeptidase-Protein eine Dipeptidylpeptidase 4 (DPP4) ist und vorzugsweise wobei das Fragment Reste 39 - 766 von menschlichem DPP4 umfasst.

22. Proteinhaltige Substanz nach einem der Ansprüche 19 - 21, wobei das proteinhaltige Molekül eine Ektodomäne eines Spike-Proteins umfasst.

23. Proteinhaltige Substanz nach Anspruch 22, wobei die Ektodomäne aus der S1-Region eines Coronavirus abgeleitet

ist.

24. Behälter, bereitgestellt mit einem Virus nach einem der Ansprüche 1 bis 4, und/oder einer Nukleinsäure nach Anspruch 5, und/oder einem Vektor nach Anspruch 6 und/oder einer Zelle nach Anspruch 7, und/oder einem Protein nach Anspruch 8, und/oder einem Antigen nach Anspruch 9, und/oder einem Antikörper nach Anspruch 10, und/oder einer pharmazeutischen Zusammensetzung nach Anspruch 16, und/oder einer Substanz nach einem der Ansprüche 19 - 23.

25. Verfahren zum Identifizierung eines Modulatorkandidaten oder eines antiviralen Wirkstoffkandidaten als ein Wirkstoff, der die Funktion von oder die Bindung von einem Virus an eine Dipeptidylpeptidase moduliert, das Verfahren umfassend:

a. Bereitstellen einer Substanz nach einem der Ansprüche 19 - 23 in der Anwesenheit und Abwesenheit des Modulatorkandidaten oder des antiviralen Wirkstoffkandidaten unter Bedingungen, erlaubend die Bindung eines Proteins, abgeleitet von einem Virus, mit dem Fragment, abgeleitet von einem Peptidase-Protein;
b. Messen des Bindens des Proteins an das Fragment, wobei eine Abnahme oder Zunahme im Binden in der Anwesenheit des Kandidatenmodulators oder des antiviralen Wirkstoffs, relativ zum Binden in Abwesenheit des Modulatorkandidaten, den Modulatorkandidaten als einen Wirkstoff identifiziert, der die Funktion einer Dipeptidylpeptidase moduliert oder den antiviralen Wirkstoff als Wirkstoff identifiziert, der die Funktion einer Dipeptidylpeptidase moduliert, vorzugsweise wobei das Protein und/oder das Fragment nachweisbar markiert ist, vorzugsweise mit einem Teil, ausgewählt aus der Gruppe, bestehend aus einem Radioisotop, einem Fluorophor, einem Quencher der Fluoreszenz, einem Enzym, und einem Affinitäts-Tag.

Revendications

1. Virus à ARN simple brin de sens positif essentiellement de mammifère qui est un bêtacoronavirus, comprenant toutes les séquences d'acides aminés choisies parmi la figure 5 fichier N.rtf représentant la protéine de nucléocapside (N), la figure 6 fichier M.rtf représentant la protéine de matrice (M), la figure 7 fichier E.rtf représentant la petite protéine d'enveloppe (E), la figure 8 fichier NS3d.rtf représentant le gène non structural NS3d, la figure 9 fichier NS3c.rtf représentant le gène non structural NS3c, la figure 10 fichier NS3b.rtf représentant le gène non structural NS3b, la figure 11 fichier NS3a.rtf représentant le gène non structural NS3a, la figure 12, fichier S.rtf représentant la glycoprotéine de surface spike (S), la figure 13 fichier Orf1ab.rtf, codant de nombreux produits enzymatiques parmi lesquels la réplicase ou comprenant la séquence d'acide nucléique de la figure 14 fichier HCoV-SA1.rtf représentant l'isolat HCoV-SA1.

2. Virus à ARN simple brin de sens positif essentiellement de mammifère qui est un bêtacoronavirus, et identifiable comme correspondant du point de vue phylogénétique à celui-ci par détermination de la séquence d'acides aminés du domaine de réplicase conservé dudit virus comme ayant au moins 90 % d'identité avec la séquence d'acides aminés de Orf1AB telle que représentée dans la figure 13.

3. Virus selon la revendication 1 ou la revendication 2 qui comprend la séquence nucléotidique telle que représentée dans la figure 14.

4. Virus selon l'une quelconque des revendications 1-3 pouvant être isolé à partir d'humains.

5. Acide nucléique, de préférence ADNc, codant une protéine telle que définie dans la revendication 1 ou une protéine ayant au moins 90 % d'identité avec la séquence d'acides aminés de Orf1AB telle que représentée dans la figure 13.

6. Vecteur comprenant un acide nucléique selon la revendication 5.

7. Cellule comprenant un virus selon l'une quelconque des revendications 1 à 4, un acide nucléique selon la revendication 5 ou un vecteur selon la revendication 6.

8. Protéine telle que représentée dans l'une quelconque des figures 5-13, ou protéine ayant au moins 90 % d'identité avec la séquence d'acides aminés de Orf1AB telle que représentée dans la figure 13.

9. Antigène comprenant une protéine selon la revendication 8.

10. Anticorps dirigé spécifiquement contre une protéine selon la revendication 8.

11. Procédé pour identifier un isolat viral comme étant un MERS-CoV comprenant la réaction dudit isolat viral ou d'un composant de celui-ci avec un acide nucléique selon la revendication 5 et/ou avec un anticorps selon la revendication 10.

12. Procédé pour diagnostiquer de manière virologique une infection par MERS-CoV d'un mammifère comprenant la détermination dans un échantillon dudit mammifère de la présence d'un isolat viral ou d'un composant de celui-ci par réaction dudit échantillon avec un acide nucléique selon la revendication 5 ou un anticorps selon la revendication 10.

13. Procédé pour diagnostiquer de manière sérologique une infection par MERS-CoV d'un mammifère comprenant la détermination dans un échantillon dudit mammifère de la présence d'un anticorps dirigé spécifiquement contre un bêtacoronavirus, de préférence un virus de la lignée C ou un composant de celui-ci par réaction dudit échantillon avec une molécule protéinique ou un fragment de celle-ci selon la revendication 8 ou un antigène selon la revendication 9.

14. Kit de diagnostic pour diagnostiquer une infection par MERS-CoV comprenant un virus selon l'une quelconque des revendications 1 à 4, et/ou un acide nucléique selon la revendication 5, et/ou une protéine selon la revendication 8, et/ou un antigène selon la revendication 9 et/ou un anticorps selon la revendication 10.

15. Virus selon l'une quelconque des revendications 1 à 4, et/ou acide nucléique selon la revendication 5, et/ou vecteur selon la revendication 6, et/ou cellule selon la revendication 7, et/ou protéine selon la revendication 8, et/ou antigène selon la revendication 9 et/ou anticorps selon la revendication 10 destinés à être utilisés dans le traitement ou la prévention d'une infection à bêtacoronavirus avec un MERS-CoV.

16. Composition pharmaceutique comprenant un virus selon l'une quelconque des revendications 1 à 4, et/ou un acide nucléique selon la revendication 5, et/ou un vecteur selon la revendication 6, et/ou une protéine selon la revendication 8, et/ou un antigène selon la revendication 9 et/ou un anticorps selon la revendication 10.

17. Composition pharmaceutique selon la revendication 16 destinée à être utilisée dans le traitement ou la prévention d'une infection par MERS-CoV chez un mammifère.

18. Composition pharmaceutique destinée à être utilisée selon la revendication 17, où ledit mammifère est un humain.

19. Substance protéinique, comprenant une protéine selon la revendication 8, où ladite protéine est une protéine spike de la figure 12 et comprenant en outre au moins un fragment d'une protéine dipeptidyl peptidase N-terminale où ledit fragment est dérivé de l'ectodomaine.

20. Substance protéinique selon la revendication 19, où ladite substance est cristallisée.

21. Substance protéinique selon la revendication 19 ou 20, où ladite protéine dipeptidyl peptidase est une dipeptidyl peptidase 4 (DPP4) et de préférence où le fragment comprend les résidus 39 - 766 de la DPP4 humaine.

22. Substance protéinique selon l'une quelconque des revendications 19 - 21, où ladite molécule protéinique comprend un ectodomaine d'une protéine spike.

23. Substance protéinique selon la revendication 22, où ledit ectodomaine est dérivé de la région S1 d'un coronavirus.

24. Récipient pourvu d'un virus selon l'une quelconque des revendications 1 à 4, et/ou d'un acide nucléique selon la revendication 5, et/ou d'un vecteur selon la revendication 6, et/ou d'une cellule selon la revendication 7, et/ou d'une protéine selon la revendication 8, et/ou d'un antigène selon la revendication 9 et/ou d'un anticorps selon la revendication 10, et/ou d'une composition pharmaceutique selon la revendication 16 et/ou d'une substance selon l'une quelconque des revendications 19 - 23.

25. Procédé d'identification d'un modulateur candidat ou d'un agent antiviral candidat comme étant un agent qui module la fonction de ou la liaison d'un virus à une dipeptidyl peptidase, ledit procédé comprenant :

a. la fourniture d'une substance selon l'une quelconque des revendications 19 - 23 en présence et en l'absence dudit modulateur candidat ou dudit agent antiviral candidat dans des conditions permettant la liaison d'une protéine dérivée d'un virus avec le fragment dérivé d'une protéine peptidase ;

b. la mesure de la liaison de ladite protéine audit fragment, où une diminution ou une augmentation de la liaison en présence dudit modulateur candidat ou dudit agent antiviral, par rapport à la liaison en l'absence dudit modulateur candidat, identifie ledit modulateur candidat comme étant un agent qui module la fonction d'une dipeptidyl peptidase ou identifie ledit agent antiviral comme étant un agent qui module la fonction d'une dipeptidyl peptidase, de préférence où ladite protéine et/ou ledit fragment est marqué de manière détectable, de préférence avec une entité choisie dans le groupe consistant en un radio-isotope, un fluorophore, un extincteur de fluorescence, une enzyme et une étiquette d'affinité.

Figure 1. Light microscopy images of LLC-MK2 cells (A, B) and VERO cells (C, D) inoculated with phosphate-buffered saline (A, C) or novel human coronavirus HCoV-SA1 (B, D) 5 days after inoculation.

Figure 2. Results of pan-coronavirus PCR. Various samples (numbered 1-12) of cell culture supernatants infected with HCoV-SA1 reacted with a pair of primers specific for the coronavirus family. A positive control virus (HCoV-NL63) was also reactive.

```
                10        20        30        40        50        60        70        80
                |.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|
HCoV-SA1        GGGGTTGGGACTATCCTAAGTGTGATAGAGCTATGCCTAATATGTGTAGAATCTTCGCTTCACTCATATTAGCTCGTAAA

                90        100       110       120       130       140       150       160
                |.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|
HCoV-SA1        CATGGCACTTGTTGTACTACAAGGGACAGATTTTAATCGCTTGGCAAATGAGTGTGCTCAGGTGCTAAGCGAATATGTCT

                170       180       190       200       210       220       230       240
                |.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|
HCoV-SA1        ATGTGGTGGTGGTTACTAACGTCAAACCTGGAGGTACCAGTAGCGGAGATGCCACCACTGCATAGTGCCAATAGTGTCTTTA

                250       260       270       280       290       300       310       320
                |.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|
HCoV-SA1        ACATTTGCAGGCGACAACTGCTAATGTCAGTGCACTTATGGGTGCTAATGGCAACAAGATTGTTGACAAAGAAGTTAAA

                330       340       350       360       370       380       390       400
                |.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|
HCoV-SA1        GACATGCAGTTTGATTTTGTATGTCAATGTTTACAGGAGCACTAGCCCAGACCCCAAATTTGTTGATAAATACTTATGCTTT

                410       420       430       440       450
                |.....|.....|.....|.....|.....|.....|.....|
HCoV-SA1        TCTTAATAAGCACTTTTCTATGATGTACTGTCTGACGACGGCGTTTGTGTGTTAA
```

Figure 3. Partial open reading frame of HCoV-SA1.

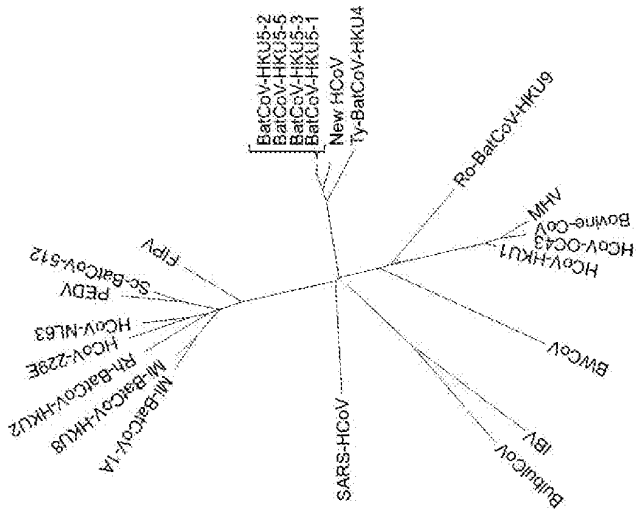


Figure 4. Maximum Likelihood tree of partial polymerase gene sequences of representative coronaviruses. HCoC-SA1 is shown in the cluster on the right hand side of the tree, labeled as "New HCoV". The cluster of viruses at the top represents viruses in the genus alphacoronavirus. The Beluga whale coronavirus (BWCoV) represents a gammacoronavirus, while the Bulbul-CoV and IBV represent a proposed new genus of the coronavirinae.

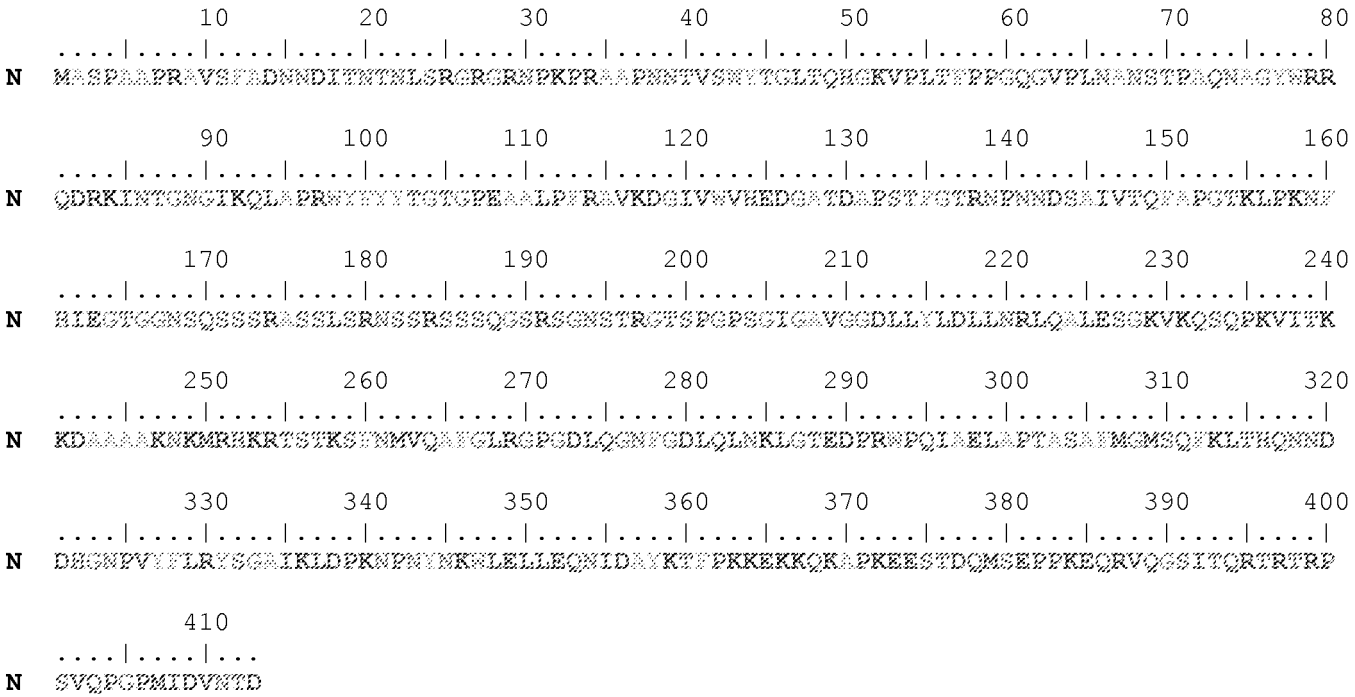


Figure 5 file N.rtf

```
             10        20        30        40        50        60        70        80
    ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
M   MSNMTQLTEAQIIAIIKDWNFAWSLIFLLITIVLQYGYPSRSMTVYVYKMFVLWLLWPSSMALSIFSAVYPIDLASQIIS

             90       100       110       120       130       140       150       160
    ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
M   GIVAAVSAMMWISYFVQSIRLFMRTGSWWSFNPETNCLLNVPFGGTTVVRPLVEDSTSVTAVVTNGHLKMAGMHFGACDY

            170       180       190       200       210
    ....|....|....|....|....|....|....|....|....|....|....|....
M   DRLPNEVTVAKPNVLIALKMVKRQSYGTNSGVAIYHRYKAGNYRSPPITADIELALLRA
```

Figure 6 file M.rtf

```
              10        20        30        40        50        60        70        80
      ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
E     MLPFVQERIGLFIVNFFIFTVVCAITLLVCMAFLTATRLCVQCMTGFNTLLVQPALYLYNTGRSVYVKFQDSKPPLPPDE

      ..
E     WV
```

Figure 7 E.rtf

```
               10        20        30        40        50        60        70        80
       ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
NS3D   MAFSASLFKPVQLVPVSPAFHRIESTDSIVFTYIPASGYVAALAVNVCLIPLLLLLRQDTCRRSIIRTMVLYFLVLYNFL

               90       100
       ....|....|....|....|....|..
NS3D   LAIVLVNGVHYPTGSCLIAFLVILIIL
```

Figure 8 file NS3d.rtf

```
                10        20        30        40        50        60        70        80
       ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
NS3C   MEESLMDVPSTSGTQVYSRKARKRSHSPTKKLRYVKRRFSLLRHEDLSVIVQPTHYVRVTFSDPNMWYLRSGHHLHSVHN

                90       100       110       120       130       140       150       160
       ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
NS3C   WLKPYGGQPVSEYHITLALLNLTDEDLARDYSPIALFLRNVRFELHEYALLRKTLVLNASEIYCANIHRFKPVYRVNTAI

               170       180       190       200       210       220       230       240
       ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
NS3C   PTIKDWLLVQGFSLYHSGLPLHMSISKLHALDDVTRNYIITMPCFRTYPQQMFVTPLAVDVVSIRSSNQGNKQIVHSYPI

       ....|.
NS3C   LHHPGF
```

Figure 9 file NS3c.rtf

```
                10        20        30        40        50        60        70        80
     ....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|
NS3B MDYVSLMQLEQKLMSPYGWCLIPRPTKTPLVGVSLAPVLAMCQLSPAGYTESAVMSTKLAKQDAQRLAELMK

                90        100
     .....|.....|.....|.....|.....|.....
NS3B DGGIPDVCSLVLRSSLMSQSEEEEPYSM
```

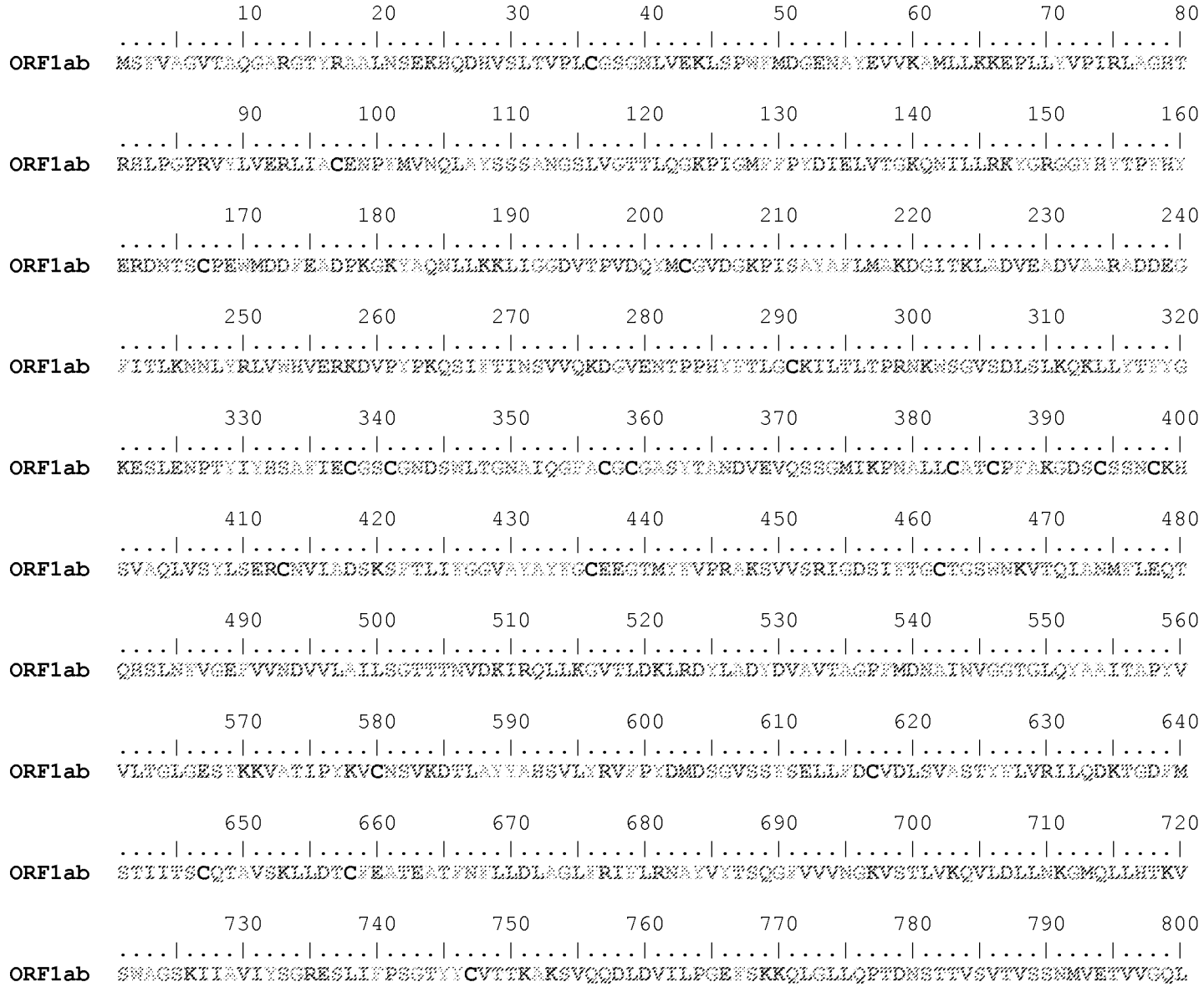
Figure 10 file NS3b.rtf

```
                10        20        30        40        50        60        70        80
                |.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|
NS3A  MRVQRPETLLVESLSLLVEASSKPLKVPEMCQWVSSCMLRCIKMQDTASLKTMRKIDVPSEESMQSVSVDLEST
                90        100
                |.....|.....|.....|....
NS3A  STMDSPTEWVTSVMLEDVGESVM
```

Figure 11 file NS3a.rtf

```
         10        20        30        40        50        60        70        80
 ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
S MIHSVFLLMFLLTPTESYVDVGPDSVKSACIEVDIQQTFFDKTWPRPIDVSKADGIIYPQGRTYSNITITYQGLFPYQGD

         90       100       110       120       130       140       150       160
 ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
S HGDMYVYSAGHATGTTPQKLFVANYSQDVKQFANGFVVRIGAAANSTGTVIISPSTSATIRKIYPAFMLGSSVGNFSDGK

        170       180       190       200       210       220       230       240
 ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
S MGRFFNHTLVLLPDGCGTLLRAFYCILEPRSGNHCPAGNSYTSFATYHTPATDCSDGNYNRNASLNSFKEYFNLRNCTFM

        250       260       270       280       290       300       310       320
 ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
S YTYNITEDEILEWFGITQTAQGVHLFSSRYVDLYGGNMFQFATLPVYDTIKYYSIIPHSIRSIQSDRKAWAAFYVYKLQP

        330       340       350       360       370       380       390       400
 ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
S LTFLLDFSVDGYIRRAIDCGFNDLSQLHCSYESFDVESGVYSVSSFEAKPSGSVVEQAEGVECDFSPLLSGTPPQVYNFK

        410       420       430       440       450       460       470       480
 ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
S RLVFTNCNYNLTKLLSLFSVNDFTCSQISPAAIASNCYSSLILDYFSYPLSMKSDLSVSSAGPISQFNYKQSFSNPTCLI

        490       500       510       520       530       540       550       560
 ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
S LATVPHNLTTITKPLKYSYINKCSRLLSDDRTEVPQLVNANQYSPCVSIVPSTVWEDGDYYRKQLSPLEGGGWLVASGST

        570       580       590       600       610       620       630       640
 ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
S VAMTEQLQMGFGITVQYGTDTNSVCPKLEFANDTKIASQLGNCVEYSLYGVSGRGVFQNCTAVGVRQQRFVYDAYQNLVG

        650       660       670       680       690       700       710       720
 ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
S YYSDDGNYYCLRACVSVPVSVIYDKETKTHATLFGSVACEHISSTMSQYSRSTRSMLKRRDSTYGPLQTPVGCVLGLVNS

        730       740       750       760       770       780       790       800
 ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
S SLFVEDCKLPLGQSLCALPDTPSTLTPRSVRSVPGEMRLASIAFNHPIQVDQLNSSYFKLSIPTNFSFGVTQEYIQTTIQ
```

Figure 12 file S.rtf


10 20 30 40 50 60 70 80
ORF1ab
90 100 110 120 130 140 150 160
ORF1ab
170 180 190 200 210 220 230 240
ORF1ab
250 260 270 280 290 300 310 320
ORF1ab
330 340 350 360 370 380 390 400
ORF1ab
410 420 430 440 450 460 470 480
ORF1ab
490 500 510 520 530 540 550 560
ORF1ab
570 580 590 600 610 620 630 640
ORF1ab
650 660 670 680 690 700 710 720
ORF1ab
730 740 750 760 770 780 790 800
ORF1ab

```
                810       820       830       840       850       860       870       880
        ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
ORF1ab  EQTNMHSPDVIVGDYVIISEKLFVRSKEEDGFAFYPACTNGHAVPTLFRLKGGAPSKKVAFGGDQVHEVAAVRSVTVEYN

                890       900       910       920       930       940       950       960
        ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
ORF1ab  IHAVLDTLLASSSLRTFVVDKSLSIEEFADVVKEQVSDLLVKLLRGMPIPDFDLDDFIDAPCYCFNAEGDASWSSTMIFS

                970       980       990      1000      1010      1020      1030      1040
        ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
ORF1ab  LHPVECDEECSEVEASDLEEGESECISETSTEQVDVSHETSDDEWAAAVDEAFPLDEAEDVTESVQEEAQPVEVPVEDIA

               1050      1060      1070      1080      1090      1100      1110      1120
        ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
ORF1ab  QVVIADTLQETPVVPDTVEVPPQVVKLPSAPQTIQPEVKEVAPVYEADTEQTQNVTVKPKRLRKKRNVDPLSNFEHKVIT

               1130      1140      1150      1160      1170      1180      1190      1200
        ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
ORF1ab  ECVTIVLGDAIQVAKCYGESVLVNAANTHLKHGGGIAGAINAASKGAVQKESDEYILAKGPLQVGDSVLLQGHSLAKNIL

               1210      1220      1230      1240      1250      1260      1270      1280
        ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
ORF1ab  HVVGPDARAKQDVSLLSKCYKAMNAYPLVVTPLVSAGIFGVKPAVSFDYLIREAKTRVLVVVNSQDVYKSLTIVDIPQSL

               1290      1300      1310      1320      1330      1340      1350      1360
        ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
ORF1ab  TFSYDGLRGAIRKAKDYGFTVFVCTDNSANTKVLRNKGVDYTKKFLTVDGVQYYCYTSKDTLDDILQQANKSVGIISMPL

               1370      1380      1390      1400      1410      1420      1430      1440
        ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
ORF1ab  GYVSHGLDLMQAGSVVRRVNVPYVCLLANKEQEAILMSEDVKLNPSEDFIKHVRTNGGYNSWHLVEGELLVQDLRLNKLL

               1450      1460      1470      1480      1490      1500      1510      1520
        ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
ORF1ab  HWSDQTICYKDSVFYVVKNSTAFPFETLSACRAYLDSRTTQQLTIEVLVTVDGVNFRTVVLNNKNTYRSQLGCVFFNGAD

               1530      1540      1550      1560      1570      1580      1590      1600
        ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
ORF1ab  ISDTIPDEKQNGHSLYLADNLTADETKALKELYGPVDPTFLHRFYSLKAAVHGWKMVVCDKVRSLKLSDNNCYLNAVIMT
```

```
              1610      1620      1630      1640      1650      1660      1670      1680
         ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
ORF1ab   LDLLKDIKFVIPALQHAFMKHKGGDSTDFIALIMAYGNCTFGAPDDASRLLHTVLAKAELCCSARMVWREWCNVCGIKDV

              1690      1700      1710      1720      1730      1740      1750      1760
         ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
ORF1ab   VLQGLKACCYVGVQTVEDLRARMTYVCQCGGERHRQLVEHTTPWLLLSGTPNEKLVTTSTAPDFVAFNVFQGIETAVGHY

              1770      1780      1790      1800      1810      1820      1830      1840
         ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
ORF1ab   VHARLKGGLILKFDSGTVSKTSDWKCKVTDVLFPGQKYSSDCNVVRYSLDGNFRTEVDPDLSAFYVKDGKYFTSEPPVTY

              1850      1860      1870      1880      1890      1900      1910      1920
         ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
ORF1ab   SPATILAGSVYTNSCLVSSDGQPGGDAISLSFNNLLGFDSSKPVTKKYTYSFLPKEDGDVLLAEFDTYDPIYKNGAMYKG

              1930      1940      1950      1960      1970      1980      1990      2000
         ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
ORF1ab   KPILWVNKASYDTNLNKFNRASLRQIFDVAPIELENKFTPLSVESTPVEPPTVDVVALQQEMTIVKCKGLNKPFVKDNVS

              2010      2020      2030      2040      2050      2060      2070      2080
         ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
ORF1ab   FVADDSGTPVVEYLSKEDLHTLYVDPKYQVIVLKDNVLSSMLRLHTVESGDINVVAASGSLTRKVKLLFRASFYFKEFAT

              2090      2100      2110      2120      2130      2140      2150      2160
         ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
ORF1ab   RTFTATTAVGSCIKSVVRHLGVTKGILTGCFSFAKMLFMLPLAYFSDSKLGTTEVKVSALKTAGVVTGNVVKQCCTAAVD

              2170      2180      2190      2200      2210      2220      2230      2240
         ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
ORF1ab   LSMDKLRRVDWKSTLRLLLMLCTTMVLLSSVYHLYVFNQVLSSDVMFEDAQGLKKFYKEVRAYLGISSACDGLASAYRAN

              2250      2260      2270      2280      2290      2300      2310      2320
         ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
ORF1ab   SFDVPTFCANRSAMCNWCLISQDSITHYPALKMVQTHLSHYVLNIDWLWFAFETGLAYMLYTSAFNWLLLAGTLHYFFAQ

              2330      2340      2350      2360      2370      2380      2390      2400
         ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
ORF1ab   TSIFVDWRSYNYAVSSAFWLFTHIPMAGLVRMYNLLACLWLLRKFYQHVINGCKDTACLLCYKRNRLTRVEASTVVCGGK
```

```
                 2410      2420      2430      2440      2450      2460      2470      2480
         ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
ORF1ab   RTFYITANGGISFCRRHNWNCVDCDTAGVGNTFICEEVANDLTTALRRPINATDRSHYYVDSVTVKETVVQFNYRRDGQP

                 2490      2500      2510      2520      2530      2540      2550      2560
         ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
ORF1ab   FYERFPLCAFTNLDKLKFKEVCKTTTGIPEYNFIIYDSSDRGQESLARSACVYYSQVLCKSILLVDSSLVTSVGDSSEIA

                 2570      2580      2590      2600      2610      2620      2630      2640
         ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
ORF1ab   TKMFDSYVNSFVSLYNVTRDKLEKLISTARDGVRRGDNFHSVLTTFIDAARGPAGVESDVETNEIVDSVQYAHKHDIQIT

                 2650      2660      2670      2680      2690      2700      2710      2720
         ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
ORF1ab   NESYNNYVPSYVKPDSVSTSDLGSLIDCNAASVNQIVLRNSNGACIWNAAAYMKLSDALKRQIRIACRKCNLAFRLTTSK

                 2730      2740      2750      2760      2770      2780      2790      2800
         ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
ORF1ab   LRANDNILSVRFTANKIVGGAPTWFNALRDFTLKGYVLATIIVFLCAVLMYLCLPTFSMAPVEFYEDRILDFKVLDNGII

                 2810      2820      2830      2840      2850      2860      2870      2880
         ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
ORF1ab   RDVNPDDKCYANKHRSFTQWYHEHVGGVYDNSITCPLTVAVIAGVAGARIPDVPTTLAWVNNQIIFFVSRVFANTGSVCY

                 2890      2900      2910      2920      2930      2940      2950      2960
         ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
ORF1ab   TPIDEIPYKSFSDSGCILPSECTMFRDAEGRMTPYCHDPTVLPGAFAYSQMRPHVRYDLYDGNMFIKFPEVVFESTLRIT

                 2970      2980      2990      3000      3010      3020      3030      3040
         ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
ORF1ab   RTLSTQYCRFGSCEYAQEGVCITTNGSWAIFNDHHLNRPGVYCGSDFIDIVRRLAVSLFQPITYFQLTTSLVLGIGLCAF

                 3050      3060      3070      3080      3090      3100      3110      3120
         ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
ORF1ab   LTLLFYYINKVKRAFADYTQCAVIAVVAAVLNSLCICFVTSIPLCIVPYTALYYYATFYFTNEPAFIMHVSWYIMFGPIV

                 3130      3140      3150      3160      3170      3180      3190      3200
         ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
ORF1ab   PIWMTCVYTVAMCFRHFFWVLAYFSKKHVEVFTDGKLNCSFQDAASNIFVINKDTYAALRNSLTNDAYSRFLGLFNKYKY
```

```
               3210      3220      3230      3240      3250      3260      3270      3280
       ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
ORF1ab YSGAMETAAYREAAACHLAKALQTYSETGSDLLYQPPNCSITSGVLQSGLVKMSHPSGDVEACMVQVTCGSMTLNGLWLD

               3290      3300      3310      3320      3330      3340      3350      3360
       ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
ORF1ab NTVWCPRHVMCPADQLSDPNYDALLISMTNHSFSVQKHIGAPANLRVVGHAMQGTLLKLTVDVANPSTPAYTFTTVKPGA

               3370      3380      3390      3400      3410      3420      3430      3440
       ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
ORF1ab AFSVLACYNGRPTGTFTVVMRPNYTIKGSFLCGSCGSVGYTKEGSVINFCYMHQMELANGTHTGSAFDGTMYGAFMDKQV

               3450      3460      3470      3480      3490      3500      3510      3520
       ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
ORF1ab HQVQLTDKYCSVNVVAWLYAAILNGCAWFVKPNRTSVVSFNEWALANQFTEFVGTQSVDMLAVKTGVAIEQLLYAIQQLY

               3530      3540      3550      3560      3570      3580      3590      3600
       ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
ORF1ab TGFQGKQILGSTMLEDEFTPEDVNMQIMGVVMQSGVRKVTYGTAHWLFATLVSTYVIILQATKFTLWNYLFETIPTQLFP

               3610      3620      3630      3640      3650      3660      3670      3680
       ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
ORF1ab LLFVTMAFVMLLVKHKHTFLTLFLLPVAICLTYANIVYEPTTPISSALIAVANWLAPTNAYMRTTHTDIGVYISMSLVLV

               3690      3700      3710      3720      3730      3740      3750      3760
       ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
ORF1ab IVVKRLYNPSLSNFALALCSGVMWLYTYSIGEASSPIAYLVFVTTLFSDYTITVFVTVNLAKVCTYAIFAYSPQLTLVFP

               3770      3780      3790      3800      3810      3820      3830      3840
       ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
ORF1ab EVKMILLLYTCLGFMCTCYFGVFSLLNLKLRAPMGVYDFKVSTQEFRFMTANNLTAPRNSWEAMALNFKLIGIGGTPCIK

               3850      3860      3870      3880      3890      3900      3910      3920
       ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
ORF1ab VAAMQSKLTDLKCTSVVLLSVLQQLHLEANSRAWAFCVKCHNDILAATDPSEAFEKFVSLFATLMTFSGNVDLDALASDI

               3930      3940      3950      3960      3970      3980      3990      4000
       ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
ORF1ab FDTPSVLQATLSEFSHLATFAELEAAQKAYQEAMDSGDTSPQVLKALQKAVNIAKNAYEKDKAVARKLERMADQAMTSMY
```

```
              4010      4020      4030      4040      4050      4060      4070      4080
          ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
ORF1ab    KQARAEDKKAKIVSAMQTMLFGMIKKLDNDVLNGIISNARNGCIPLSVIPLCASNKLRVVIPDFTVWNQVVTYPSLNYAG

              4090      4100      4110      4120      4130      4140      4150      4160
          ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
ORF1ab    ALWDITVINNVDNEIVKSSDVVDSNENLTWPLVLECTRASTSAVKLQNNEIKPSGLKTMVVSAGQEQTNCNTSSLAYYEP

              4170      4180      4190      4200      4210      4220      4230      4240
          ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
ORF1ab    VQGRKMLMALLSDNAYLKWARVEGKDGFVSVELQPPCKFLIAGPKGPEIRYLYFVKNLNNLHRGQVLGHIAATVRLQAGS

              4250      4260      4270      4280      4290      4300      4310      4320
          ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
ORF1ab    NTEFASNSSVLSLVNFTVDPQKAYLDFVNAGGAPLTNCVKMLTPKTGTGIAISVKPESTADQETYGGASVCLYCRAHIEH

              4330      4340      4350      4360      4370      4380      4390      4400
          ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
ORF1ab    PDVSGVCKYKGKFVQIPAQCVRDPVGFCLSNTPCNVCQYWIGYGCNCDSLRQAALPQSKDSNFLNRVRGSIVNARIEPCS

              4410      4420      4430      4440      4450      4460      4470      4480
          ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
ORF1ab    SGLSTDVVFRAFDICNYKAKVAGIGKYYKTNTCRFVELDDQGHHLDSYFVVKRHTMENYELEKHCYDLLRDCDAVAPHDF

              4490      4500      4510      4520      4530      4540      4550      4560
          ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
ORF1ab    FIFDVDKVKTPHIVRQRLTEYTMMDLVYALRHFDQNSEVLKAILVKYGCCDVTYFENKLWFDFVENPSVIGVYHKLGERV

              4570      4580      4590      4600      4610      4620      4630      4640
          ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
ORF1ab    RQAILNTVKFCDHMVKAGLVGVLTLDNQDLNGKWYDFGDFVITQPGSGVAIVDSYYSYLMPVLSMTDCLAAETHRDCDFN

              4650      4660      4670      4680      4690      4700      4710      4720
          ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
ORF1ab    KPLIEWPLTEYDFTDYKVQLFEKYFKYWDQTYHANCVNCTDDRCVLHCANFNVLFAMTMPKTCFGPIVRKIFVDGVPFVV

              4730      4740      4750      4760      4770      4780      4790      4800
          ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
ORF1ab    SCGYHYKELGLVMNMDVSLHRHRLSLKELMMYAADPAMHIASSNAFLDLRTSCFSVAALTTGLTFQTVRPGNFNQDFYDF
```

```
                4810      4820      4830      4840      4850      4860      4870      4880
        ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
ORF1ab  VVSKGFFKEGSSVTLKHFFFAQDGNAAITDYNYYSYNLPTMCDIKQMLFCMEVVNKYFEIYDGGCLNASEVVVNNLDKSA

                4890      4900      4910      4920      4930      4940      4950      4960
        ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
ORF1ab  GHPFNKFGKARVYYESMSYQEQDELFAMTKRNVIPTMTQMNLKYAISAKNRARTVAGVSILSTMTNRQYHQKMLKSMAAT

                4970      4980      4990      5000      5010      5020      5030      5040
        ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
ORF1ab  RGATCVIGTTKFYGGWDFMLKTLYKDVDNPHLMGWDYPKCDRAMPNMCRIFASLILARKHGTCCTTRDRFYRLANECAQV

                5050      5060      5070      5080      5090      5100      5110      5120
        ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
ORF1ab  LSEYVLCGGGYYVKPGGTSSGDATTAYANSVFNILQATTANVSALMGANGNKIVDKEVKDMQFDLYVNVYRSTSPDPKFV

                5130      5140      5150      5160      5170      5180      5190      5200
        ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
ORF1ab  DKYYAFLNKHFSMMILSDDGVVCYNSDYAAKGYIAGIQNFKETLYYQNNVFMSEAKCWVETDLKKGPHEFCSQHTLYIKD

                5210      5220      5230      5240      5250      5260      5270      5280
        ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
ORF1ab  GDDGYFLPYPDPSRILSAGCFVDDIVKTDGTLMVERFVSLAIDAYPLTKHEDIEYQNVFWVYLQYIEKLYKDLTGHMLDS

                5290      5300      5310      5320      5330      5340      5350      5360
        ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
ORF1ab  YSVMLCGDNSAKFWEEAFYRDLYSSPTTLQAVGSCVVCHSQTSLRCGTCIRRPFLCCKCCYDHVIATPHKMVLSVSPYVC

                5370      5380      5390      5400      5410      5420      5430      5440
        ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
ORF1ab  NAPGCGVSDVTKLYLGGMSYFCVDHRPVCSFPLCANGLVFGLYKNMCTGSPSIVEFNRLATCDWTESGDYTLANTTTEPL

                5450      5460      5470      5480      5490      5500      5510      5520
        ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
ORF1ab  KLFAAETLRATEEASKQSYAIATIKEIVGERQLLLVWEAGKSKPPLNRNYVFTGYHITKNSKVQLGEYIFERIDYSDAVS

                5530      5540      5550      5560      5570      5580      5590      5600
        ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
ORF1ab  YKSSTTYKLTVGDIFVLTSHSVATLTAPTIVNQERYVKITGLYPTITVPEEFASHVANFQKSGYSKYVTVQGPPGTGKSH
```

```
                5610      5620      5630      5640      5650      5660      5670      5680
        ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
ORF1ab  FAIGLALYYPTARVVYTACSHAAVDALCEKAFKYLNIAKCSRIIPAKARVECYDRFKVNETNSQYLFSTINALPETSADI

                5690      5700      5710      5720      5730      5740      5750      5760
        ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
ORF1ab  LVVDEVSMCTNYDLSIINARIKAKHIVYVGDPAQLPAPRTLLTRGTLEPENFNSVTRLMCNLGPDIFLSMCYRCPKEIVS

                5770      5780      5790      5800      5810      5820      5830      5840
        ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
ORF1ab  TVSALVYNNKLLAKKELSGQCFKILYKGNVTHDASSAINRPQLTFVKNFITANPAWSKAVFISPYNSQNAVSRSMLGLTT

                5850      5860      5870      5880      5890      5900      5910      5920
        ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
ORF1ab  QTVDSSQGSEYQYVIFCQTADTAHANNINRFNVAITRAQKGILCVMTSQALFESLEFTELSFTNYKLQSQIVTGLFKDCS

                5930      5940      5950      5960      5970      5980      5990      6000
        ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
ORF1ab  RETSGLSPAYAPTYVSVDDKYKTSDELCVNLNLPANVPYSRVISRMGFKLDATVPGYPKLFITREEAVRQVRSWIGFDVE

                6010      6020      6030      6040      6050      6060      6070      6080
        ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
ORF1ab  GAHASRNACGTNVPLQLGFSTGVNFVVQPVGVVDTEWGNMLTGIAARPPPGEQFKHLVPLMHKGAAWPIVRRRIVQMLSD

                6090      6100      6110      6120      6130      6140      6150      6160
        ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
ORF1ab  TLDKLSDYCTFVCWAHGFELTSASYFCKIGKEQKCCMCNRRAAAYSSPLQSYACWTHSCGYDYVYNPFFVDVQQWGYVGN

                6170      6180      6190      6200      6210      6220      6230      6240
        ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
ORF1ab  LATNHDRYCSVHQGAHVASNDAIMTRCLAIHSCFIERVDWDIEYPYISHEKKLNSCCRIVERNVVRAALLAGSFDKVYDI

                6250      6260      6270      6280      6290      6300      6310      6320
        ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
ORF1ab  GNPKGIPIVDDPVVDWHYFDAQPLTRKVQQLFYTEDMASRFADGLCLFWNCNVPKYPNNAIVCRFDTRVHSEFNLPGCDG

                6330      6340      6350      6360      6370      6380      6390      6400
        ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
ORF1ab  GSLYVNKHAFHTPAYDVSAFRDLKPLPFFYYSTTPCEVHGNGSMIEDIDYVPLKSAVCITACNLGGAVCRKHATEYREYM
```

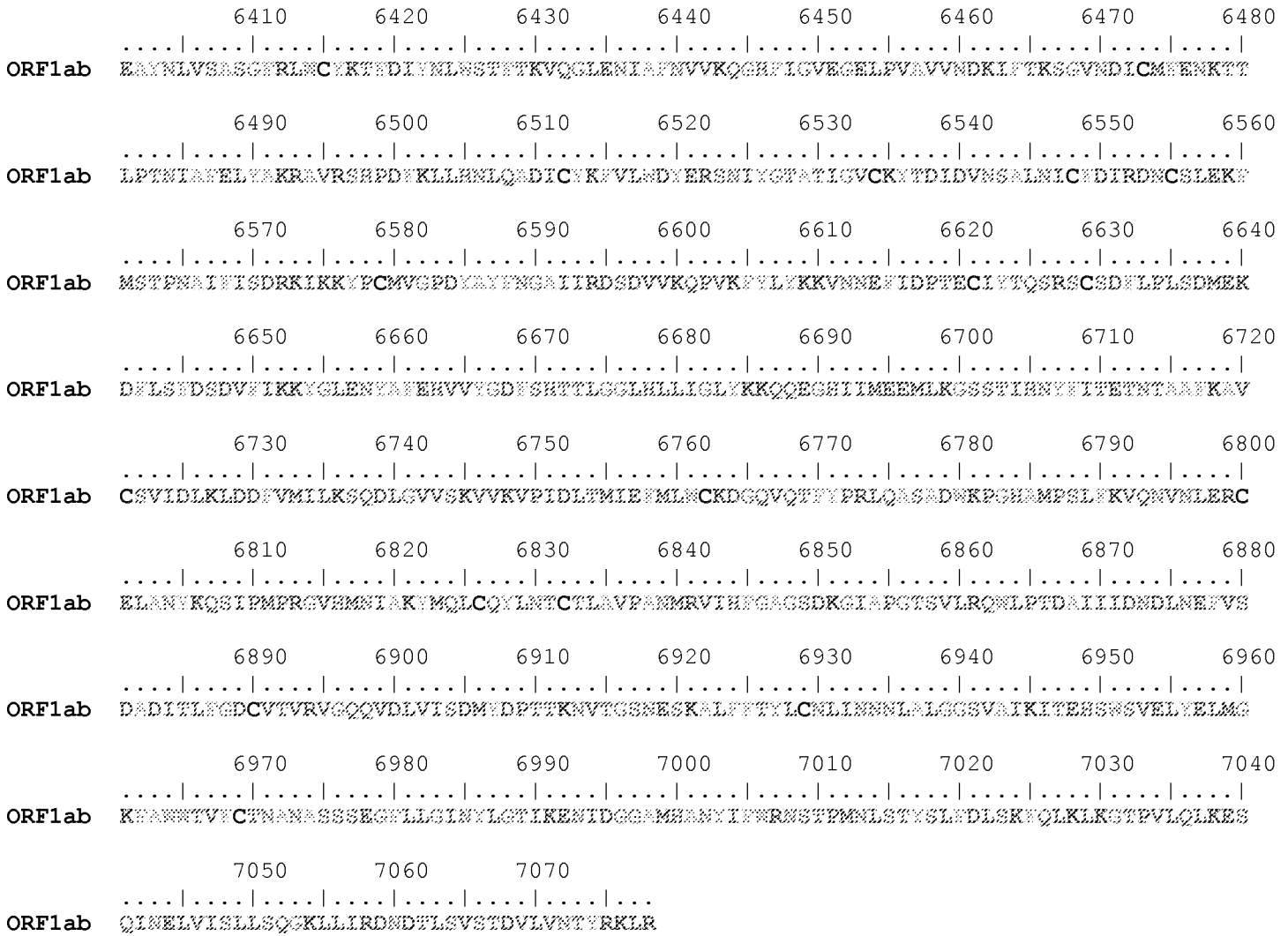
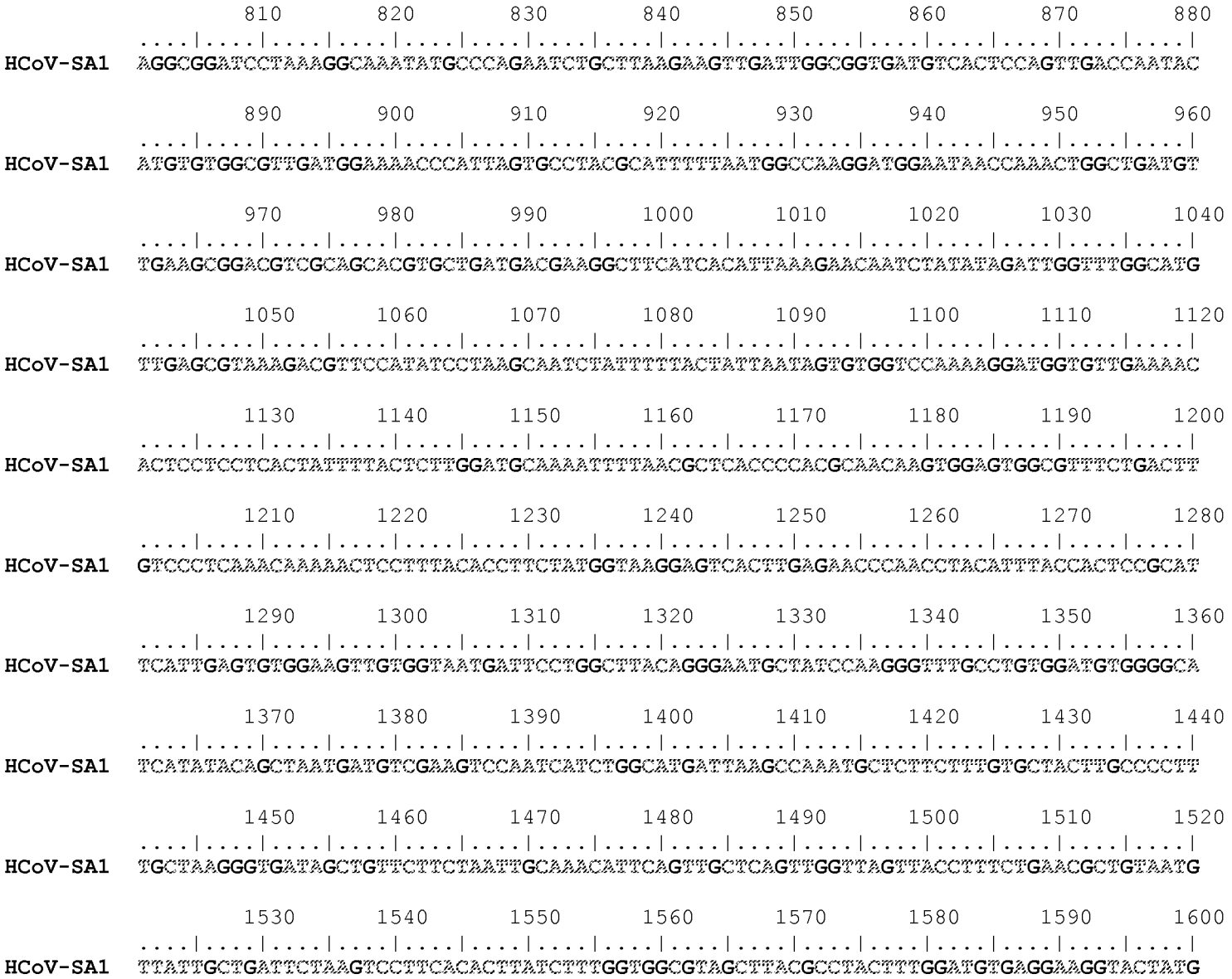
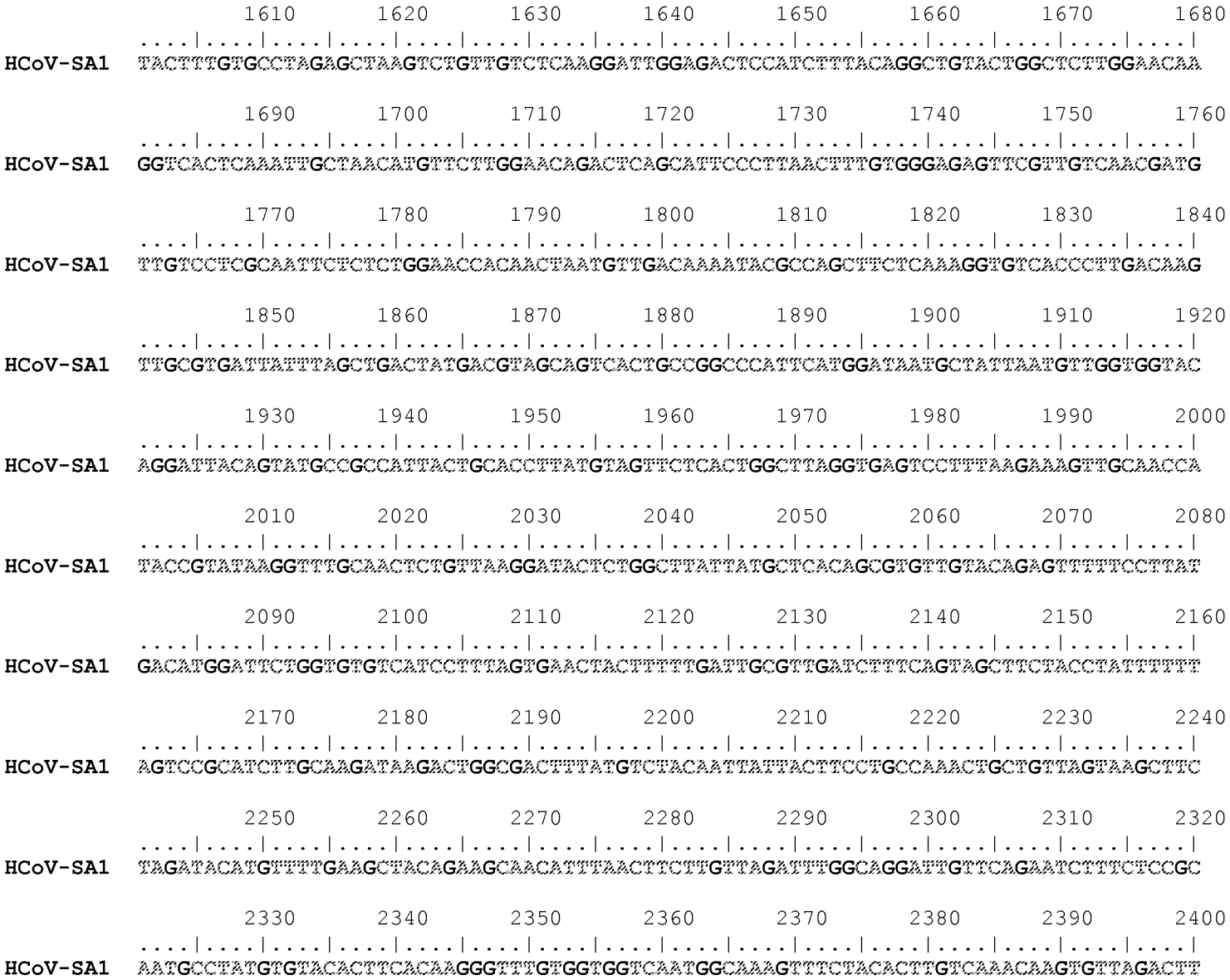
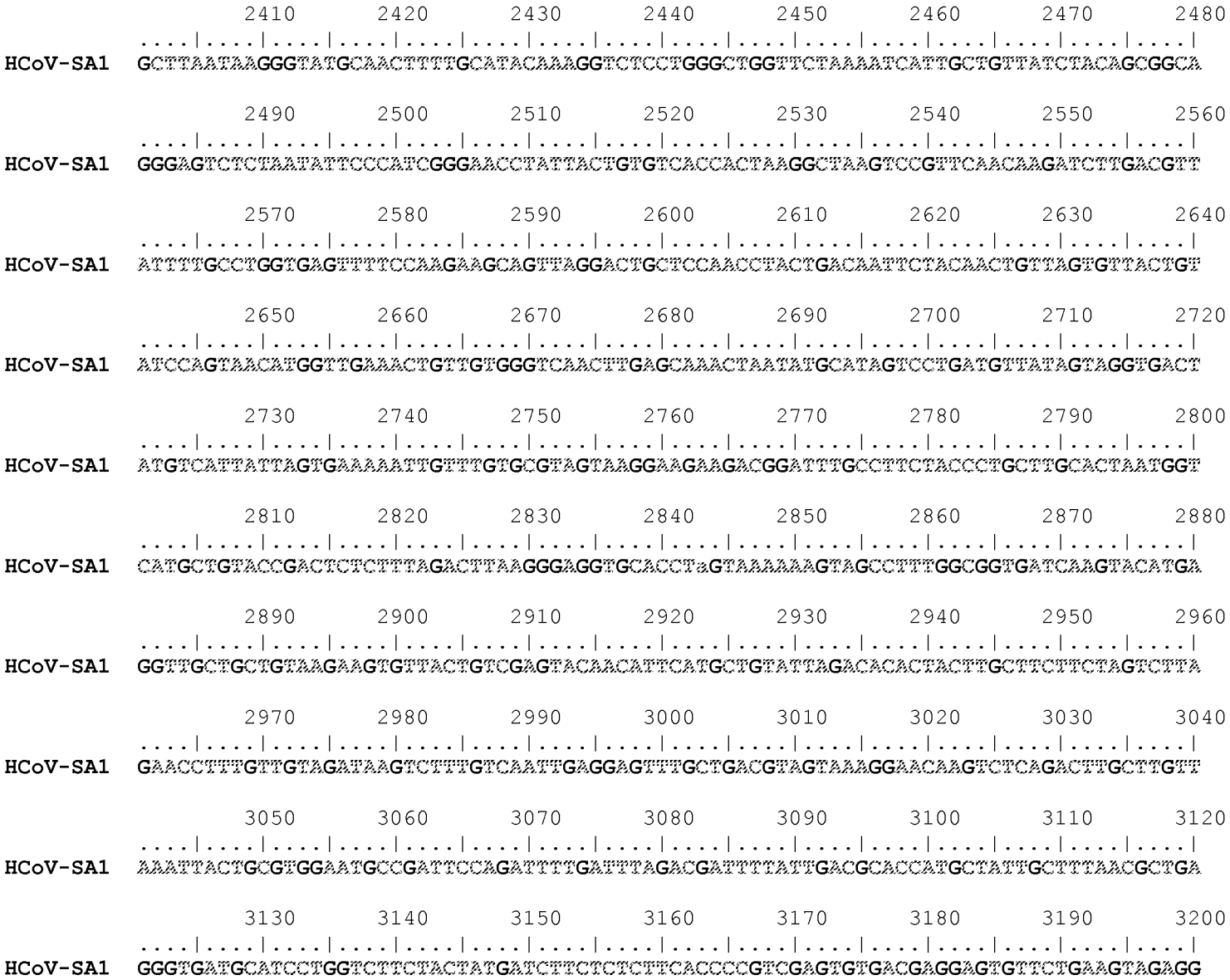


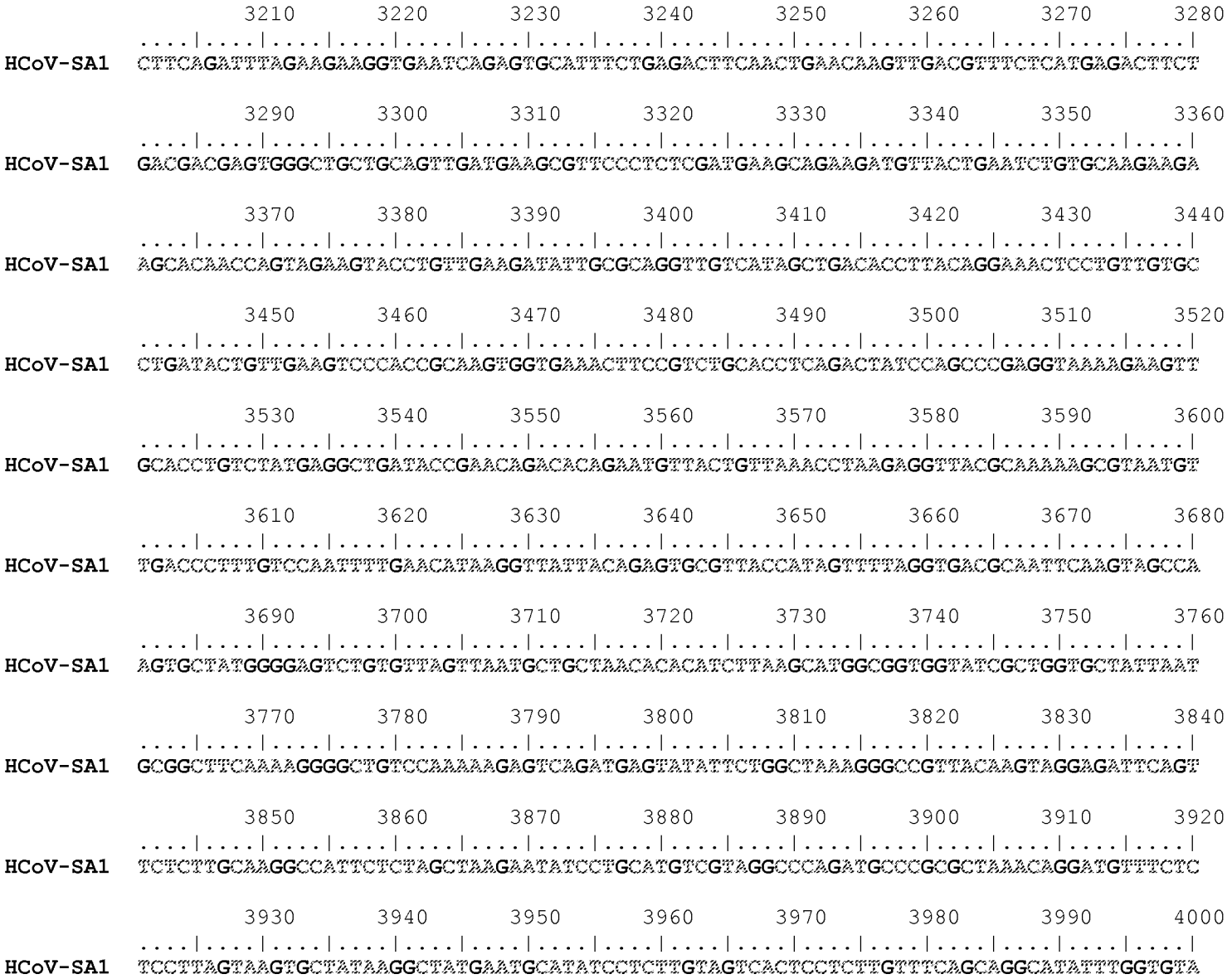
Figure 13 file Orf1ab.rtf

```
                10        20        30        40        50        60        70        80
                |....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
HCoV-SA1        ATTTAAGTGAATAGCTTGGCTATCTCACTTCCCCTCGTTCTCTTGCAGAACTTTGATTTTAACGAAACTTAAATAAAGCC

                90        100       110       120       130       140       150       160
                |....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
HCoV-SA1        CTGTTGTTTAGCGTATCGTTGCACTTGTCTGGTGGGATTGTGGCATTAATTTTGCCTGCTCATCTAGGCAGTGGACAATATG

                170       180       190       200       210       220       230       240
                |....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
HCoV-SA1        CTCAACACTGGGTATAATTCTAAATTGAATACTATTTTTCAGTTAGAGCGTCGTGTCTCTTGTACGTCTCGGTCACAAATAC

                250       260       270       280       290       300       310       320
                |....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
HCoV-SA1        ACGGTTTCGTCCGGTGCGTGGCAATTCGGGGCACATCATGTCTTTTCGTGGCTGGTGTGACCGCGCAAGGTGCGCGGGTA

                330       340       350       360       370       380       390       400
                |....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
HCoV-SA1        CGTATCGAGCAGCGCTCAACTCTGAAAACATCAAGACCATGTGTCTCTAACTGTGCCACTCTGTGGTTCAGGAACCTG

                410       420       430       440       450       460       470       480
                |....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
HCoV-SA1        GTTGAAAAACTTTCACCATGGTTCATGGATGGCGAAAATGCCTATGAAGTGGTGAAGGCCATGTTACTTAAAAAGGAGCC

                490       500       510       520       530       540       550       560
                |....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
HCoV-SA1        ACTTCTCTATGTGCCCATCCGGCTGGCACACTAGACACCTCCCAGGTCCTCGTGTGTACCTGGTTGAGAGGCTCA

                570       580       590       600       610       620       630       640
                |....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
HCoV-SA1        TTGCTTGTGAAAATCCATTCATGGTTAACCAATTGGCTTATAGCTCTAGTGCAAATGGCAGCCTGGTTGGCACAACTTTG

                650       660       670       680       690       700       710       720
                |....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
HCoV-SA1        CAGGGCAAGCCTATTGGTATGTTCTCCTTATGACATCGAACTTGTCACAGGAAAGCAAATATTCTCCTGCGCAAGTA

                730       740       750       760       770       780       790       800
                |....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
HCoV-SA1        TGGCCGTGGTGGTTATCACTACACCCCATTCCACTATGAGCGAGAACAACACCTCTTGCCCTGAGTGGATGGACGATTTTG
```


810 820 830 840 850 860 870 880
HCoV-SA1 AGGCGGATCCTAAAGGCAAATATGCCCAGAATCTGCTTAAGAAGTTGATTGGCGGTGATGTCACTCCAGTTGACCAATAC
890 900 910 920 930 940 950 960
HCoV-SA1 ATGTGTGGCGTTGATGGAAAACCCATTAGTGCCTACGCATTTTTAATGGCCAAGGATGGAATAACCAAACTGGCTGATGT
970 980 990 1000 1010 1020 1030 1040
HCoV-SA1 TGAAGCGGACGTCGCAGCACGTGCTGATGACGAAGGCTTCATCACATTAAAGAACAATCTATATAGATTGGTTTGGCATG
1050 1060 1070 1080 1090 1100 1110 1120
HCoV-SA1 TTGAGCGTAAAGACGTTCCATATCCTAAGCAATCTATTTTTACTATTAATAGTGTGGTCCAAAAGGATGGTGTTGAAAAC
1130 1140 1150 1160 1170 1180 1190 1200
HCoV-SA1 ACTCCTCCTCACTATTTTACTCTTGGATGCAAAATTTTAACGCTCACCCCACGCAACAAGTGGAGTGGCGTTTCTGACTT
1210 1220 1230 1240 1250 1260 1270 1280
HCoV-SA1 GTCCCTCAAACAAAAACTCCTTTACACCTTCTATGGTAAGGAGTCACTTGAGAACCCAACCTACATTTACCACTCCGCAT
1290 1300 1310 1320 1330 1340 1350 1360
HCoV-SA1 TCATTGAGTGTGGAAGTTGTGGTAATGATTCCTGGCTTACAGGGAATGCTATCCAAGGGTTTGCCTGTGGATGTGGGGCA
1370 1380 1390 1400 1410 1420 1430 1440
HCoV-SA1 TCATATACAGCTAATGATGTCGAAGTCCAATCATCTGGCATGATTAAGCCAAATGCTCTTCTTTGTGCTACTTGCCCCTT
1450 1460 1470 1480 1490 1500 1510 1520
HCoV-SA1 TGCTAAGGGTGATAGCTGTTCTTCTAATTGCAAACATTCAGTTGCTCAGTTGGTTAGTTACCTTTCTGAACGCTGTAATG
1530 1540 1550 1560 1570 1580 1590 1600
HCoV-SA1 TTATTGCTGATTCTAAGTCCTTCACACTTATCTTTGGTGGCGTAGCTTACGCCTACTTTGGATGTGAGGAAGGTACTATG


1610 1620 1630 1640 1650 1660 1670 1680
HCoV-SA1 TACTTTGTGCCTAGAGCTAAGTCTGTTGTCTCAAGGATTGGAGACTCCATCTTTACAGGCTGTACTGGCTCTTGGAACAA
1690 1700 1710 1720 1730 1740 1750 1760
HCoV-SA1 GGTCACTCAAATTGCTAACATGTTCTTGGAACAGACTCAGCATTCCCTTAACTTTGTGGGAGAGTTCGTTGTCAACGATG
1770 1780 1790 1800 1810 1820 1830 1840
HCoV-SA1 TTGTCCTCGCAATTCTCTCTGGAACCACAACTAATGTTGACAAAATACGCCAGCTTCTCAAAGGTGTCACCCTTGACAAG
1850 1860 1870 1880 1890 1900 1910 1920
HCoV-SA1 TTGCGTGATTATTTAGCTGACTATGACGTAGCAGTCACTGCCGGCCCATTCATGGATAATGCTATTAATGTTGGTGGTAC
1930 1940 1950 1960 1970 1980 1990 2000
HCoV-SA1 AGGATTACAGTATGCCGCCATTACTGCACCTTATGTAGTTCTCACTGGCTTAGGTGAGTCCTTTAAGAAAGTTGCAACCA
2010 2020 2030 2040 2050 2060 2070 2080
HCoV-SA1 TACCGTATAAGGTTTGCAACTCTGTTAAGGATACTCTGGCTTATTATGCTCACAGCGTGTTGTACAGAGTTTTTCCTTAT
2090 2100 2110 2120 2130 2140 2150 2160
HCoV-SA1 GACATGGATTCTGGTGTGTCATCCTTTAGTGAACTACTTTTTGATTGCGTTGATCTTTCAGTAGCTTCTACCTATTTTTT
2170 2180 2190 2200 2210 2220 2230 2240
HCoV-SA1 AGTCCGCATCTTGCAAGATAAGACTGGCGACTTTATGTCTACAATTATTACTTCCTGCCAAACTGCTGTTAGTAAGCTTC
2250 2260 2270 2280 2290 2300 2310 2320
HCoV-SA1 TAGATACATGTTTTGAAGCTACAGAAGCAACATTTAACTTCTTGTTAGATTTGGCAGGATTGTTCAGAATCTTTCTCCGC
2330 2340 2350 2360 2370 2380 2390 2400
HCoV-SA1 AATGCCTATGTGTACACTTCACAAGGGTTTGTGGTGGTCAATGGCAAAGTTTCTACACTTGTCAAACAAGTGTTAGACTT


2410 2420 2430 2440 2450 2460 2470 2480
HCoV-SA1 GCTTAATAAGGGTATGCAACTTTTGCATACAAAGGTCTCCTGGGCTGGTTCTAAAATCATTGCTGTTATCTACAGCGGCA
2490 2500 2510 2520 2530 2540 2550 2560
HCoV-SA1 GGGAGTCTCTAATATTCCCATCGGGAACCTATTACTGTGTCACCACTAAGGCTAAGTCCGTTCAACAAGATCTTGACGTT
2570 2580 2590 2600 2610 2620 2630 2640
HCoV-SA1 ATTTTGCCTGGTGAGTTTTCCAAGAAGCAGTTAGGACTGCTCCAACCTACTGACAATTCTACAACTGTTAGTGTTACTGT
2650 2660 2670 2680 2690 2700 2710 2720
HCoV-SA1 ATCCAGTAACATGGTTGAAACTGTTGTGGGTCAACTTGAGCAAACTAATATGCATAGTCCTGATGTTATAGTAGGTGACT
2730 2740 2750 2760 2770 2780 2790 2800
HCoV-SA1 ATGTCATTATTAGTGAAAAATTGTTTGTGCGTAGTAAGGAAGAAGACGGATTTGCCTTCTACCCTGCTTGCACTAATGGT
2810 2820 2830 2840 2850 2860 2870 2880
HCoV-SA1 CATGCTGTACCGACTCTCTTTAGACTTAAGGGAGGTGCACCTAGTAAAAAAGTAGCCTTTGGCGGTGATCAAGTACATGA
2890 2900 2910 2920 2930 2940 2950 2960
HCoV-SA1 GGTTGCTGCTGTAAGAAGTGTTACTGTCGAGTACAACATTCATGCTGTATTAGACACACTACTTGCTTCTTCTAGTCTTA
2970 2980 2990 3000 3010 3020 3030 3040
HCoV-SA1 GAACCTTTGTTGTAGATAAGTCTTTGTCAATTGAGGAGTTTGCTGACGTAGTAAAGGAACAAGTCTCAGACTTGCTTGTT
3050 3060 3070 3080 3090 3100 3110 3120
HCoV-SA1 AAATTACTGCGTGGAATGCCGATTCCAGATTTTGATTTAGACGATTTTATTGACGCACCATGCTATTGCTTTAACGCTGA
3130 3140 3150 3160 3170 3180 3190 3200
HCoV-SA1 GGGTGATGCATCCTGGTCTTCTACTATGATCTTCTCTCTTCACCCCGTCGAGTGTGACGAGGAGTGTTCTGAAGTAGAGG


3210 3220 3230 3240 3250 3260 3270 3280
HCoV-SA1 CTTCAGATTTAGAAGAAGGTGAATCAGAGTGCATTTCTGAGACTTCAACTGAACAAGTTGACGTTTCTCATGAGACTTCT
3290 3300 3310 3320 3330 3340 3350 3360
HCoV-SA1 GACGACGAGTGGGCTGCTGCAGTTGATGAAGCGTTCCCTCTCGATGAAGCAGAAGATGTTACTGAATCTGTGCAAGAAGA
3370 3380 3390 3400 3410 3420 3430 3440
HCoV-SA1 AGCACAACCAGTAGAAGTACCTGTTGAAGATATTGCGCAGGTTGTCATAGCTGACACCTTACAGGAAACTCCTGTTGTGC
3450 3460 3470 3480 3490 3500 3510 3520
HCoV-SA1 CTGATACTGTTGAAGTCCCACCGCAAGTGGTGAAACTTCCGTCTGCACCTCAGACTATCCAGCCCGAGGTAAAAGAAGTT
3530 3540 3550 3560 3570 3580 3590 3600
HCoV-SA1 GCACCTGTCTATGAGGCTGATACCGAACAGACACAGAATGTTACTGTTAAACCTAAGAGGTTACGCAAAAAGCGTAATGT
3610 3620 3630 3640 3650 3660 3670 3680
HCoV-SA1 TGACCCTTTGTCCAATTTTGAACATAAGGTTATTACAGAGTGCGTTACCATAGTTTTAGGTGACGCAATTCAAGTAGCCA
3690 3700 3710 3720 3730 3740 3750 3760
HCoV-SA1 AGTGCTATGGGGAGTCTGTGTTAGTTAATGCTGCTAACACACATCTTAAGCATGGCGGTGGTATCGCTGGTGCTATTAAT
3770 3780 3790 3800 3810 3820 3830 3840
HCoV-SA1 GCGGCTTCAAAAGGGGCTGTCCAAAAAGAGTCAGATGAGTATATTCTGGCTAAAGGGCCGTTACAAGTAGGAGATTCAGT
3850 3860 3870 3880 3890 3900 3910 3920
HCoV-SA1 TCTCTTGCAAGGCCATTCTCTAGCTAAGAATATCCTGCATGTCGTAGGCCCAGATGCCCGCGCTAAACAGGATGTTTCTC
3930 3940 3950 3960 3970 3980 3990 4000
HCoV-SA1 TCCTTAGTAAGTGCTATAAGGCTATGAATGCATATCCTCTTGTAGTCACTCCTCTTGTTTCAGCAGGCATATTTGGTGTA

```
                4010      4020      4030      4040      4050      4060      4070      4080
                ....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|
HCoV-SA1        AAACCAGCTGTGTCTTTTGATATCTTATTAGGGAGGCTAAGACTAGAGTTTAGTCGTCGTTAATTCCCAAGATGTCTA

                4090      4100      4110      4120      4130      4140      4150      4160
                ....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|
HCoV-SA1        TAAGAGTCTTACCATAGTTGACATTCCACAGAGTTTGACTTTTTCATATGATGGGTTACGTGGCGCAATACGTAAAGCTA

                4170      4180      4190      4200      4210      4220      4230      4240
                ....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|
HCoV-SA1        AAGATTATGGTTTTACTGTTTTTGTGCACAGACAACTCTGCTAACACTAAAGTTCTTAGGAACAAGGGTGTTGATTAT

                4250      4260      4270      4280      4290      4300      4310      4320
                ....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|
HCoV-SA1        ACTAAGAAGTTTCTTACAGTTGACGGTGTGCAATATTATTGCTACACGTCTAAGGACACTTTAGATGATATCTTACAACA

                4330      4340      4350      4360      4370      4380      4390      4400
                ....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|
HCoV-SA1        GGCTAATAAGTCTGTTGGTAATTATATCTATGCCTTTGGGATATGTGTCTCATGGTTTAGACTTAATGCAAGCAGGGAGTG

                4410      4420      4430      4440      4450      4460      4470      4480
                ....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|
HCoV-SA1        TCGTGCGTAGAGTTAACGTGCCCTACGTGTGTCTCCTAGCTAATAAAGAGCAAGAAGCTATTTTGATGTCTGAAGACGTT

                4490      4500      4510      4520      4530      4540      4550      4560
                ....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|
HCoV-SA1        AAGTTAAACCCTTCAGAAGATTTTATAAAGCACGTCCGCACTAATGGTGGTTACAATTCTTGGCATTTAGTCGAGGGTGA

                4570      4580      4590      4600      4610      4620      4630      4640
                ....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|
HCoV-SA1        ACTATTGGTGCAAGACTTACGCTTAAATAAGCTCCTGCATTGGTCTGAATCAAACCATATGCTACAAGGATAGTGTGTTTT

                4650      4660      4670      4680      4690      4700      4710      4720
                ....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|
HCoV-SA1        ATGTTGTAAAGAATAGTACAGCTTTTCCATTTGAAACACTTTCAGCATGTCGTGCGTATTTGGATTCACGCACGACACAG

                4730      4740      4750      4760      4770      4780      4790      4800
                ....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|
HCoV-SA1        CAGTTAACAATCGAAGTCTTAGTGACTGTCGATGGTGTAAATTTTAGAACAGTCGTTCTAAATAATAAGAACACTTATAG
```

```
          4810      4820      4830      4840      4850      4860      4870      4880
....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|
HCoV-SA1  ATCACAGCTTGGATGCGTTTTCTTTAATGGTGCTGATATTTCTGACACCATTCCTGATGAGAAACAGAATGGTCACAGTT

          4890      4900      4910      4920      4930      4940      4950      4960
....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|
HCoV-SA1  TATATCTAGCAGACAATTTGACTGCTGATGAAACAAAGGCGCTTAAAGAGTTATATGGCCCGTTGATCCTACTTTCTTA

          4970      4980      4990      5000      5010      5020      5030      5040
....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|
HCoV-SA1  CACAGATTCTATTCACTTAAGGCTGCAGTCCATGGGTGGAAGATGGTTGTGTGATAAGGTACGTTCTCTCAAATTGAG

          5050      5060      5070      5080      5090      5100      5110      5120
....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|
HCoV-SA1  TGATAAATAATTGTTATCTTAATGCAGTTTATATGACACTTGATTTATTGAAGGACATTTAAATTTGTTATACCTGCTCTAC

          5130      5140      5150      5160      5170      5180      5190      5200
....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|
HCoV-SA1  AGCATGCATTTATGAAACATAAGGGCGGTGATTCAACTGACTTCATAGCCCTCATTATGGCTTATGGCAATTGCACATTT

          5210      5220      5230      5240      5250      5260      5270      5280
....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|
HCoV-SA1  GGTGCTCCAGATGATGCCTCTCGGTTACTTCATACCGTGCTTGCAAAGGCTGAGTTATGCTGTTCTGCACGCATGGTTTG

          5290      5300      5310      5320      5330      5340      5350      5360
....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|
HCoV-SA1  GAGAGAGTGGTGCAATGTCTGTGGCAAAAAAGATGTTGTTCTACAAGGCTTAAAAGCTTGTTGTTACGTGGGTGTGCAAA

          5370      5380      5390      5400      5410      5420      5430      5440
....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|
HCoV-SA1  CTGTTGAAGATCTGCGTGCTCGCATGACATATGTATGCCAGTGTGGTGGTGAACGTCATCGGCAATTAGTCCGAACACACC

          5450      5460      5470      5480      5490      5500      5510      5520
....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|
HCoV-SA1  ACCCCCTGGTTGCTGCTCTCAGGCACACCAAATGAAAAATTGGTGACAACCTCCACGGCGCCTGATTTTGTAGCATTTAA

          5530      5540      5550      5560      5570      5580      5590      5600
....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|
HCoV-SA1  TGTCTTTTCAGGGCAATTGAAACGGCTGTTGGCCATTTATGTTCATGCTCGCCTGAAGGGTGGTCTTTATTTTTAAAGTTTTGACT
```

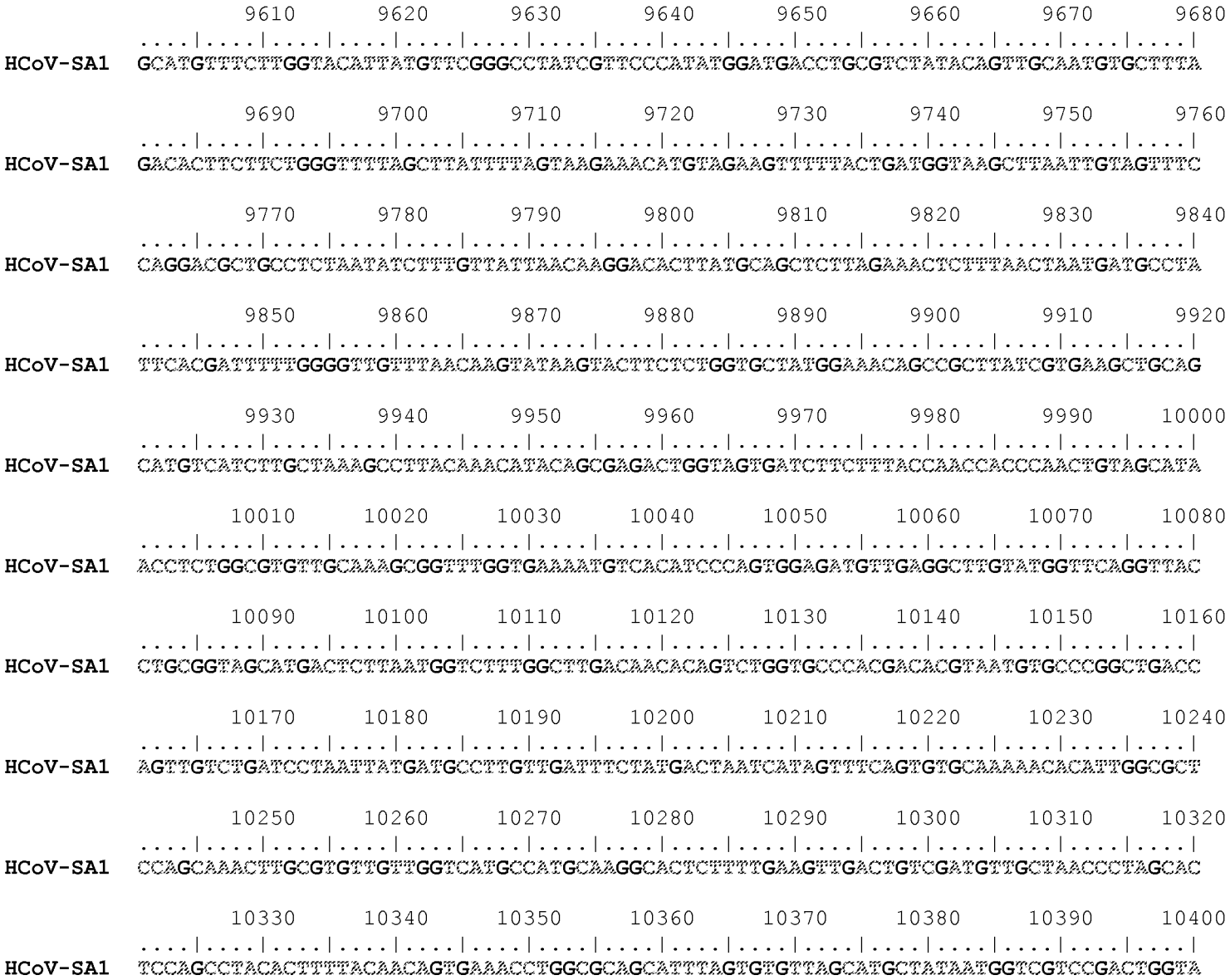
```
                5610      5620      5630      5640      5650      5660      5670      5680
                ....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|
HCoV-SA1        CTGGCACCGTTAGCAAGACTTCAGACTGGAAGTGCAAGGTGACAGATGTACTTTCCCCGGCCAAAAATACAGTAGCGAT

                5690      5700      5710      5720      5730      5740      5750      5760
                ....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|
HCoV-SA1        TGTAATGTCGTACGGTATTCTTTTGGACGGTAATTTCAGAACAGAGGTTGATCCCGACCTAATCTGCTTTCTATGTTAAGGA

                5770      5780      5790      5800      5810      5820      5830      5840
                ....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|
HCoV-SA1        TGGTAAATACTTTACAAGTGAAACCACCCGTAACATATTCACCAGCTACAATTTTAGCTGGTAGTGTCTACACTAATAGCT

                5850      5860      5870      5880      5890      5900      5910      5920
                ....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|
HCoV-SA1        GCCTTGTATCGTCTGATGGCAACCTGGCGGTGATGCTATTAGTTTGAGTTTTAATACACTTTTAGGGTTTGATTCTAGT

                5930      5940      5950      5960      5970      5980      5990      6000
                ....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|
HCoV-SA1        AAACCAGTCACTAAGAAATACACTTACTCCTTCTTGCCTAAAGAGACGGCGATGTGTTGGCTGAGTTTGACACTTA

                6010      6020      6030      6040      6050      6060      6070      6080
                ....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|
HCoV-SA1        TGACCCTATTATAAGAATGGTGCCATGTATAAAAGGCAAACCAATTCTTTGGGTCAATAAAGCATCTTATGATACTAATC

                6090      6100      6110      6120      6130      6140      6150      6160
                ....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|
HCoV-SA1        TTAATAAGTTCAATAGAGCTAGTTTGCGTCAAATTTTTGACGTAGCCCCCATTGAACTCGAAAATAAATTCACACCTTTG

                6170      6180      6190      6200      6210      6220      6230      6240
                ....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|
HCoV-SA1        AGTGTGGAGTCTACACCAGTTGAACCTCCAACTGTAGATGTGGTAGCACTTCAACAGGAAATGACAATTGTCAAATGTAA

                6250      6260      6270      6280      6290      6300      6310      6320
                ....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|
HCoV-SA1        GGGTTTAAATAAACCTTCGTGAAGGACAATGTCAGTTTCGTTGCTGATGATTCAGGTACTCCCGTTGTTGAGTATCTGT

                6330      6340      6350      6360      6370      6380      6390      6400
                ....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|
HCoV-SA1        CTAAAGAGACCTACATACATTGTATGTAGACCCTAAGTATCAAGTCATTGTCTTTAAAAGACAATGTACTTTCTTCTATG
```

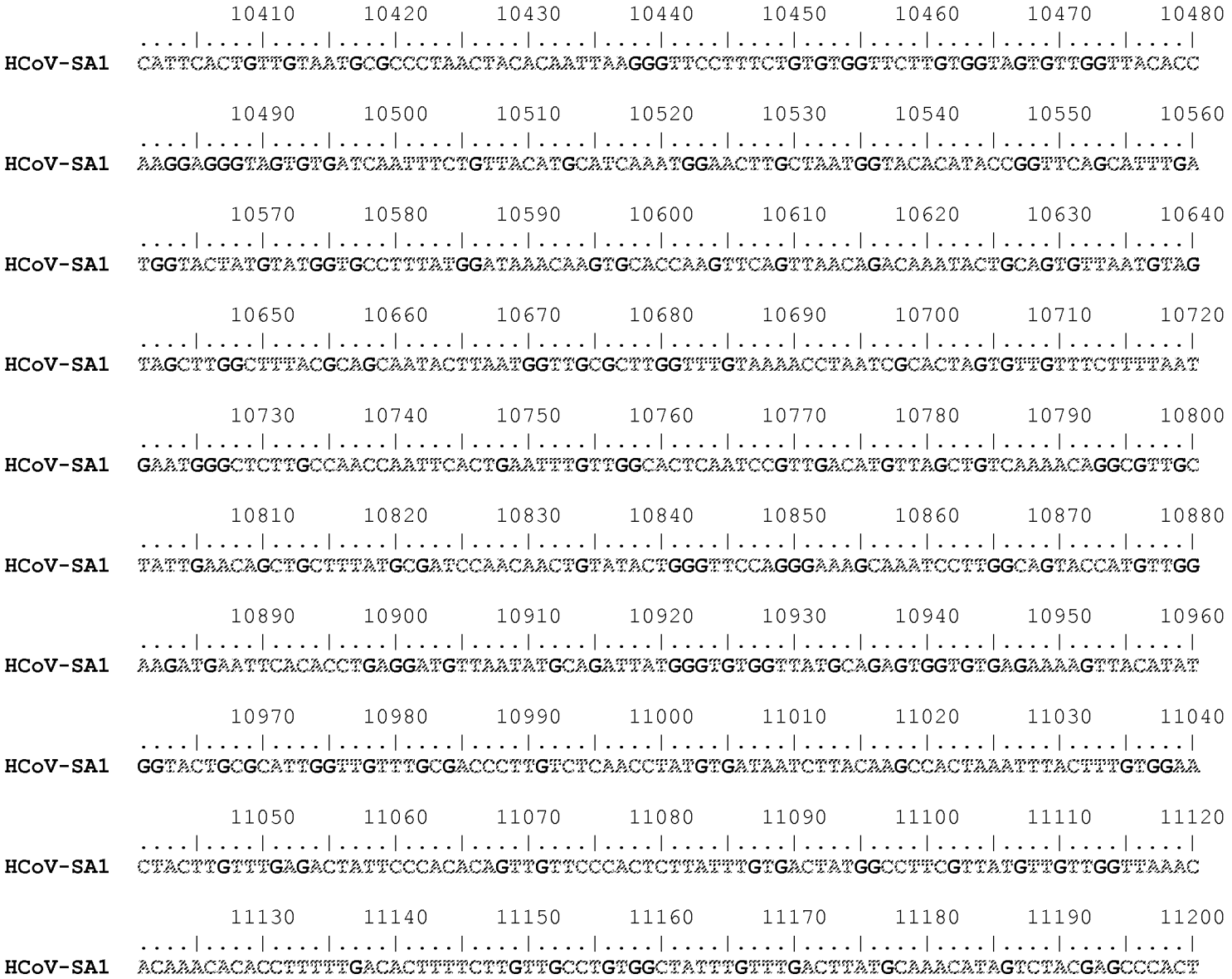
```
                 6410      6420      6430      6440      6450      6460      6470      6480
         ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
HCoV-SA1 CTTAGATTGCACACCGTTGAGTCAGGTGATATTAACGTTGTTGCAGCTTCCGGATCTTTGACACGTAAAGTGAAGTTACT

                 6490      6500      6510      6520      6530      6540      6550      6560
         ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
HCoV-SA1 ATTTAGGGCTTCATTTTATTTCAAAGAATTTGCTACCCGCACTTTCACTGCTACCACTGCTGTAGGTAGTTGTATAAAGA

                 6570      6580      6590      6600      6610      6620      6630      6640
         ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
HCoV-SA1 GTGTAGTGCGGCATCTAGGTGTTACTAAAGGCATATTGACAGGCTGTTTTAGTTTTGCCAAGATGTTATTTATGCTTCCA

                 6650      6660      6670      6680      6690      6700      6710      6720
         ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
HCoV-SA1 CTAGCTTACTTTAGTGATTCAAAACTCGGCACCACAGAGGTTAAAGTGAGTGCTTTGAAAACAGCCGGCGTTGTGACAGG

                 6730      6740      6750      6760      6770      6780      6790      6800
         ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
HCoV-SA1 TAATGTTGTAAAACAGTGTTGCACTGCTGCTGTTGATTTAAGTATGGATAAGTTGCGCCGTGTGGATTGGAAATCAACCC

                 6810      6820      6830      6840      6850      6860      6870      6880
         ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
HCoV-SA1 TACGGTTGTTACTTATGTTATGCACAACTATGGTATTGTTGTCTTCTGTGTATCACTTGTATGTCTTCAATCAGGTCTTA

                 6890      6900      6910      6920      6930      6940      6950      6960
         ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
HCoV-SA1 TCAAGTGATGTTATGTTTGAAGATGCCCAAGGTTTGAAAAAGTTCTACAAAGAAGTTAGAGCTTACCTAGGAATCTCTTC

                 6970      6980      6990      7000      7010      7020      7030      7040
         ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
HCoV-SA1 TGCTTGTGACGGTCTTGCTTCAGCTTATAGGGCGAATTCCTTTGATGTACCTACATTCTGCGCAAACCGTTCTGCAATGT

                 7050      7060      7070      7080      7090      7100      7110      7120
         ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
HCoV-SA1 GTAATTGGTGCTTGATTAGCCAAGATTCCATAACTCACTACCCAGCTCTTAAGATGGTTCAAACACATCTTAGCCACTAT

                 7130      7140      7150      7160      7170      7180      7190      7200
         ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
HCoV-SA1 GTTCTTAACATAGATTGGTTGTGGTTTGCATTTGAGACTGGTTTGGCATACATGCTCTATACCTCGGCCTTCAACTGGTT
```

```
                 7210      7220      7230      7240      7250      7260      7270      7280
             ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
HCoV-SA1     GTTGTTGGCAGGTACATTGCATTATTTCTTTGCACAGACTTCCATATTTGTAGACTGGCGGTCATACAATTATGCTGTGT

                 7290      7300      7310      7320      7330      7340      7350      7360
             ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
HCoV-SA1     CTAGTGCCTTCTGGTTATTCACCCACATTCCAATGGCGGGTTTGGTACGAATGTATAATTTGTTAGCATGCCTTTGGCTT

                 7370      7380      7390      7400      7410      7420      7430      7440
             ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
HCoV-SA1     TTACGCAAGTTTTATCAGCATGTAATCAATGGTTGCAAAGATACGGCATGCTTGCTCTGCTATAAGAGGAACCGACTTAC

                 7450      7460      7470      7480      7490      7500      7510      7520
             ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
HCoV-SA1     TAGAGTTGAAGCTTCTACCGTTGTCTGTGGTGGAAAACGTACGTTTTATATCACAGCAAATGGCGGTATTTCATTCTGTC

                 7530      7540      7550      7560      7570      7580      7590      7600
             ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
HCoV-SA1     GTAGGCATAATTGGAATTGTGTGGATTGTGACACTGCAGGTGTGGGGAATACCTTCATCTGTGAAGAAGTCGCAAATGAC

                 7610      7620      7630      7640      7650      7660      7670      7680
             ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
HCoV-SA1     CTCACTACCGCCCTACGCAGGCCTATTAACGCTACGGATAGATCACATTATTATGTGGATTCCGTTACAGTTAAAGAGAC

                 7690      7700      7710      7720      7730      7740      7750      7760
             ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
HCoV-SA1     TGTTGTTCAGTTTAATTATCGTAGAGACGGTCAACCATTCTACGAGCGGTTTCCCCTCTGCGCTTTTACAAATCTAGATA

                 7770      7780      7790      7800      7810      7820      7830      7840
             ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
HCoV-SA1     AGTTGAAGTTCAAAGAGGTCTGTAAAACTACTACTGGTATACCTGAATACAACTTTATCATCTACGACTCATCAGATCGT

                 7850      7860      7870      7880      7890      7900      7910      7920
             ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
HCoV-SA1     GGCCAGGAAAGTTTAGCTAGGTCTGCATGTGTTTATTATTCTCAAGTCTTGTGTAAATCAATTCTTTTGGTTGACTCAAG

                 7930      7940      7950      7960      7970      7980      7990      8000
             ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
HCoV-SA1     TTTGGTTACTTCTGTTGGTGATTCTAGTGAAATCGCCACTAAAATGTTTGATTCCTTTGTTAATAGTTTCGTCTCGCTGT
```

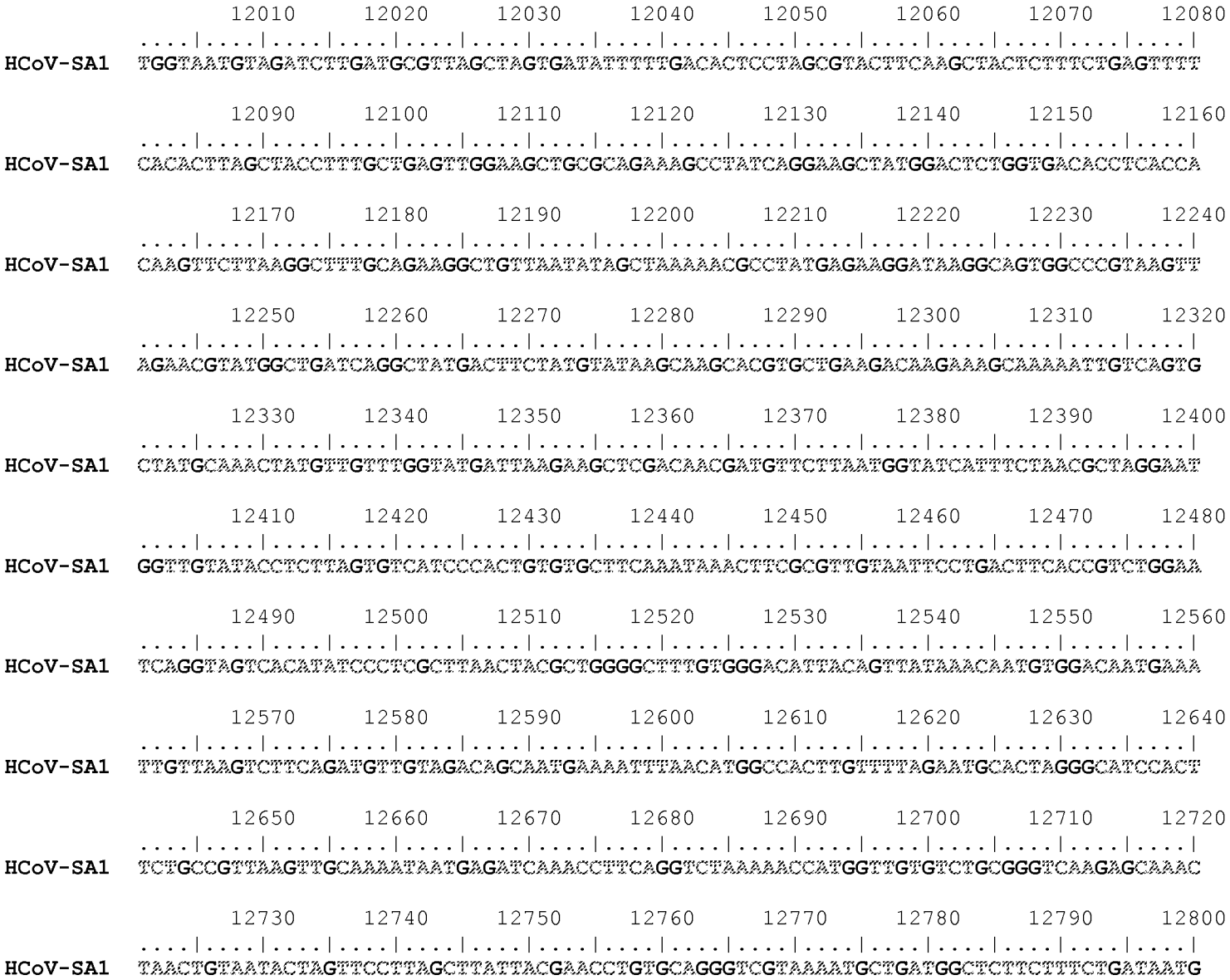
```
                8010      8020      8030      8040      8050      8060      8070      8080
                ....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|
HCoV-SA1        ATAATGTCACACGCGATAAGTTGGAAAAACTTAACTCTACTGCTCGTGATGGCGTAAGGCGAGGCGATAACTTCCATAGT

                8090      8100      8110      8120      8130      8140      8150      8160
                ....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|
HCoV-SA1        GTCTTAACAACATTCATTGACGCAGCACGAGGCCCCGCAGGTGTGGAGTCTGATGTTGAGACCAATGAAATTGTTGACTC

                8170      8180      8190      8200      8210      8220      8230      8240
                ....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|
HCoV-SA1        TGTGCAGTATGCTCATAAACATGACATACAAAATTACTAAAGAGCTACAATAATTATATGTACCCTCATATGTTAAACCCTG

                8250      8260      8270      8280      8290      8300      8310      8320
                ....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|
HCoV-SA1        ATAGTGTGTCTACCAGCGATTTAGGTAGTCTCATTGATTGTAATGCGGCTTCAGTTAACCAAATTGTCTTGCGTAATTCT

                8330      8340      8350      8360      8370      8380      8390      8400
                ....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|
HCoV-SA1        AATGGTGCTTGCATTTTGGAACGCTTGCTGCATATATGAAACTCTCGGAATGCACTTAAACGACAGATTCGCATTGCCG

                8410      8420      8430      8440      8450      8460      8470      8480
                ....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|
HCoV-SA1        TAAGTGTAATTTAGCTTTCCGGTTAACCACCTCAAAGCTACGCGCTAAATGATAATATCTTATCAGTTAGATTCACTGCTA

                8490      8500      8510      8520      8530      8540      8550      8560
                ....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|
HCoV-SA1        ACAAAATTGTTGGTGGTGCTCCTACATGGTTTAAATGCGTTGCGTGACTTTACGTTAAAGGGTTATGTTCTTGCTACCATT

                8570      8580      8590      8600      8610      8620      8630      8640
                ....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|
HCoV-SA1        ATTGTGTTTCTGTGCTGTACTGATATTTGTGTTTACCTACATTTTCCTATGGCACCTGTTGAATTTTATAGAAGACCG

                8650      8660      8670      8680      8690      8700      8710      8720
                ....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|
HCoV-SA1        CATCTTGGACTTTTAAAGTTCTTGATAATGGTATCATTAGGGATGTAAATCCTGATGATAAGTGCTTTGCTAATAAGCACC

                8730      8740      8750      8760      8770      8780      8790      8800
                ....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|
HCoV-SA1        GGTCCTTCACACAATGGTATCAATGAGCAATGTTGGTGGTGTCTATATGACAACTCTATCACATGCCCATTGACAGTTGCAGTA
```

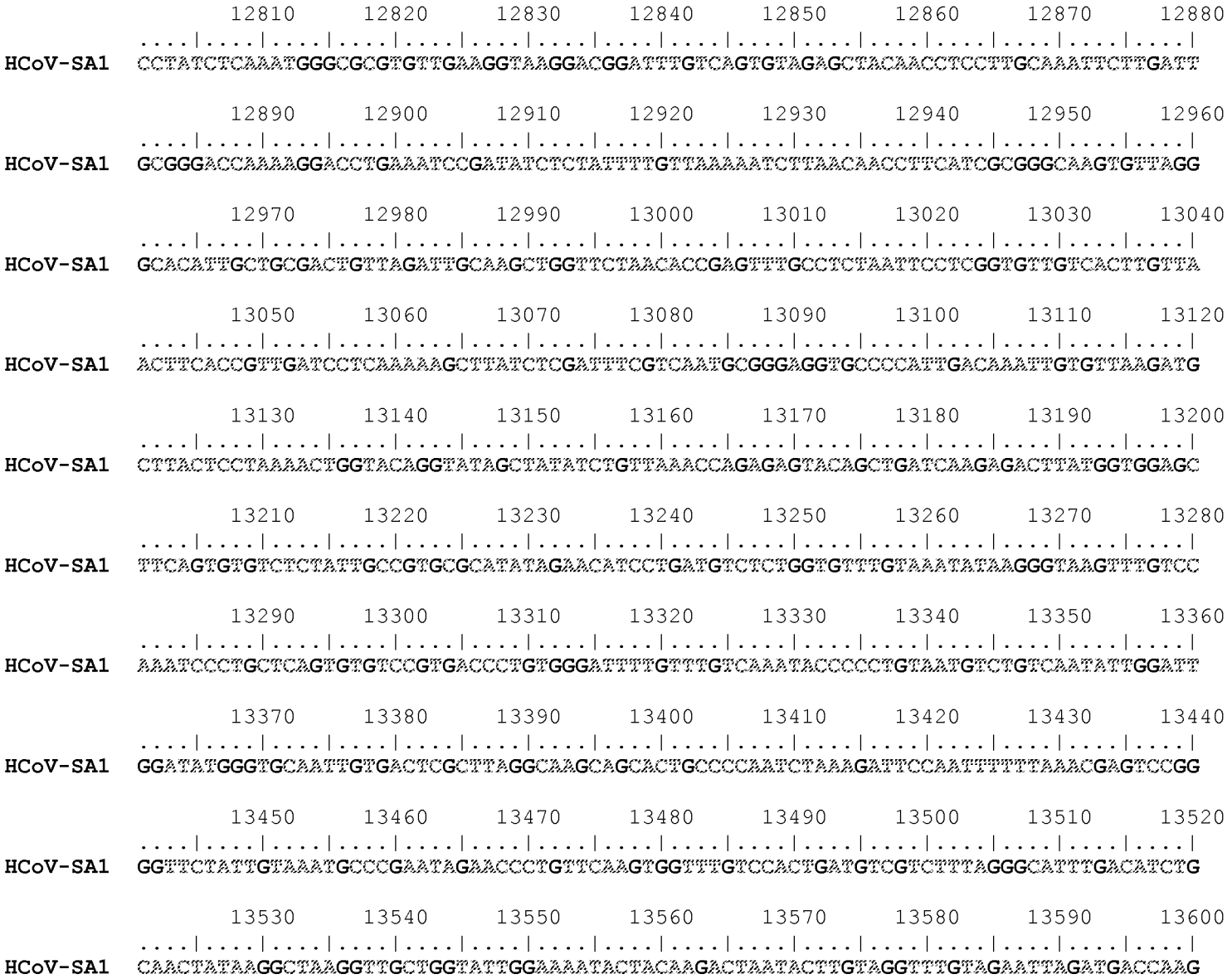
```
                  8810      8820      8830      8840      8850      8860      8870      8880
          ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
HCoV-SA1  ATTGCTGGAGTTGCTGGTGCTCGCATTCCAGACGTACCTACTACATTGGCTTGGGTGAACAATCAGATAATTTTCTTTGT

                  8890      8900      8910      8920      8930      8940      8950      8960
          ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
HCoV-SA1  TTCTCGAGTCTTTGCTAATACAGGCAGTGTTTGCTACACTCCTATAGATGAGATACCCTATAAGAGTTTCTCTGATAGTG

                  8970      8980      8990      9000      9010      9020      9030      9040
          ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
HCoV-SA1  GTTGCATTCTTCCATCTGAGTGCACTATGTTTAGGGATGCAGAGGGCCGTATGACACCATACTGCCATGATCCTACTGTT

                  9050      9060      9070      9080      9090      9100      9110      9120
          ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
HCoV-SA1  TTGCCTGGGGCTTTTGCGTACAGTCAGATGAGGCCTCATGTTCGTTACGACTTGTATGATGGTAACATGTTTATTAAATT

                  9130      9140      9150      9160      9170      9180      9190      9200
          ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
HCoV-SA1  TCCTGAAGTAGTATTTGAAAGTACACTTAGGATTACTAGAACTCTGTCAACTCAGTACTGCCGGTTCGGTAGTTGTGAGT

                  9210      9220      9230      9240      9250      9260      9270      9280
          ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
HCoV-SA1  ATGCACAAGAGGGTGTTTGTATTACCACAAATGGCTCGTGGGCCATTTTTAATGACCACCATCTTAATAGACCTGGTGTC

                  9290      9300      9310      9320      9330      9340      9350      9360
          ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
HCoV-SA1  TATTGTGGCTCTGATTTTATTGACATTGTCAGGCGGTTAGCAGTATCACTGTTCCAGCCTATTACTTATTTCCAATTGAC

                  9370      9380      9390      9400      9410      9420      9430      9440
          ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
HCoV-SA1  TACCTCATTGGTCTTGGGTATAGGTTTGTGTGCGTTCCTGACTTTGCTCTTCTATTATATTAATAAAGTAAAACGTGCTT

                  9450      9460      9470      9480      9490      9500      9510      9520
          ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
HCoV-SA1  TTGCAGATTACACCCAGTGTGCTGTAATTGCTGTTGTTGCTGCTGTTCTTAATAGCTTGTGCATCTGCTTTGTTACCTCT

                  9530      9540      9550      9560      9570      9580      9590      9600
          ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
HCoV-SA1  ATACCATTGTGTATAGTACCTTACACTGCATTGTACTATTATGCTACATTCTATTTTACTAATGAGCCTGCATTTATTAT
```


9610 9620 9630 9640 9650 9660 9670 9680
HCoV-SA1 GCATGTTTCTTGGTACATTATGTTCGGGCCTATCGTTCCCATATGGATGACCTGCGTCTATACAGTTGCAATGTGCTTTA
9690 9700 9710 9720 9730 9740 9750 9760
HCoV-SA1 GACACTTCTTCTGGGTTTTAGCTTATTTTAGTAAGAAACATGTAGAAGTTTTTACTGATGGTAAGCTTAATTGTAGTTTC
9770 9780 9790 9800 9810 9820 9830 9840
HCoV-SA1 CAGGACGCTGCCTCTAATATCTTTGTTATTAACAAGGACACTTATGCAGCTCTTAGAAACTCTTTAACTAATGATGCCTA
9850 9860 9870 9880 9890 9900 9910 9920
HCoV-SA1 TTCACGATTTTTGGGGTTGTTTAACAAGTATAAGTACTTCTCTGGTGCTATGGAAACAGCCGCTTATCGTGAAGCTGCAG
9930 9940 9950 9960 9970 9980 9990 10000
HCoV-SA1 CATGTCATCTTGCTAAAGCCTTACAAACATACAGCGAGACTGGTAGTGATCTTCTTTACCAACCACCCAACTGTAGCATA
10010 10020 10030 10040 10050 10060 10070 10080
HCoV-SA1 ACCTCTGGCGTGTTGCAAAGCGGTTTGGTGAAAATGTCACATCCCAGTGGAGATGTTGAGGCTTGTATGGTTCAGGTTAC
10090 10100 10110 10120 10130 10140 10150 10160
HCoV-SA1 CTGCGGTAGCATGACTCTTAATGGTCTTTGGCTTGACAACACAGTCTGGTGCCCACGACACGTAATGTGCCCGGCTGACC
10170 10180 10190 10200 10210 10220 10230 10240
HCoV-SA1 AGTTGTCTGATCCTAATTATGATGCCTTGTTGATTTCTATGACTAATCATAGTTTCAGTGTGCAAAAACACATTGGCGCT
10250 10260 10270 10280 10290 10300 10310 10320
HCoV-SA1 CCAGCAAACTTGCGTGTTGTTGGTCATGCCATGCAAGGCACTCTTTTGAAGTTGACTGTCGATGTTGCTAACCCTAGCAC
10330 10340 10350 10360 10370 10380 10390 10400
HCoV-SA1 TCCAGCCTACACTTTTACAACAGTGAAACCTGGCGCAGCATTTAGTGTGTTAGCATGCTATAATGGTCGTCCGACTGGTA

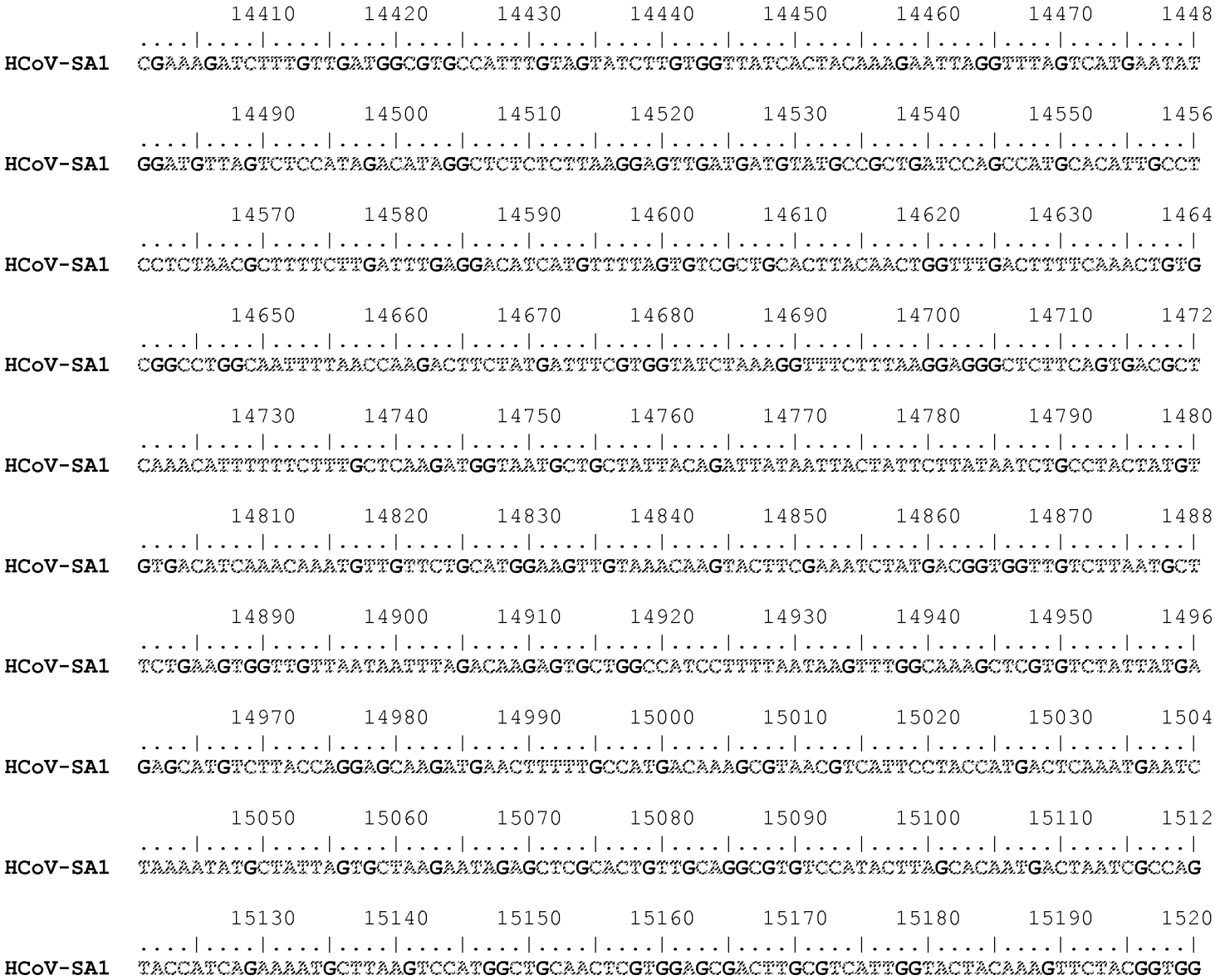

10410 10420 10430 10440 10450 10460 10470 10480
HCoV-SA1 CATTCACTGTTGTAATGCGCCCTAACTACACAATTAAGGGTTCCTTTCTGTGTGGTTCTTGTGGTAGTGTTGGTTACACC
10490 10500 10510 10520 10530 10540 10550 10560
HCoV-SA1 AAGGAGGGTAGTGTGATCAATTTCTGTTACATGCATCAAATGGAACTTGCTAATGGTACACATACCGGTTCAGCATTTGA
10570 10580 10590 10600 10610 10620 10630 10640
HCoV-SA1 TGGTACTATGTATGGTGCCTTTATGGATAAACAAGTGCACCAAGTTCAGTTAACAGACAAATACTGCAGTGTTAATGTAG
10650 10660 10670 10680 10690 10700 10710 10720
HCoV-SA1 TAGCTTGGCTTTACGCAGCAATACTTAATGGTTGCGCTTGGTTTGTAAAACCTAATCGCACTAGTGTTGTTTCTTTTAAT
10730 10740 10750 10760 10770 10780 10790 10800
HCoV-SA1 GAATGGGCTCTTGCCAACCAATTCACTGAATTTGTTGGCACTCAATCCGTTGACATGTTAGCTGTCAAAACAGGCGTTGC
10810 10820 10830 10840 10850 10860 10870 10880
HCoV-SA1 TATTGAACAGCTGCTTTATGCGATCCAACAACTGTATACTGGGTTCCAGGGAAAGCAAATCCTTGGCAGTACCATGTTGG
10890 10900 10910 10920 10930 10940 10950 10960
HCoV-SA1 AAGATGAATTCACACCTGAGGATGTTAATATGCAGATTATGGGTGTGGTTATGCAGAGTGGTGTGAGAAAAGTTACATAT
10970 10980 10990 11000 11010 11020 11030 11040
HCoV-SA1 GGTACTGCGCATTGGTTGTTTGCGACCCTTGTCTCAACCTATGTGATAATCTTACAAGCCACTAAATTTACTTTGTGGAA
11050 11060 11070 11080 11090 11100 11110 11120
HCoV-SA1 CTACTTGTTTGAGACTATTCCCACACAGTTGTTCCCACTCTTATTTGTGACTATGGCCTTCGTTATGTTGTTGGTTAAAC
11130 11140 11150 11160 11170 11180 11190 11200
HCoV-SA1 ACAAACACACCTTTTTGACACTTTTCTTGTTGCCTGTGGCTATTTGTTTGACTTATGCAAACATAGTCTACGAGCCCACT

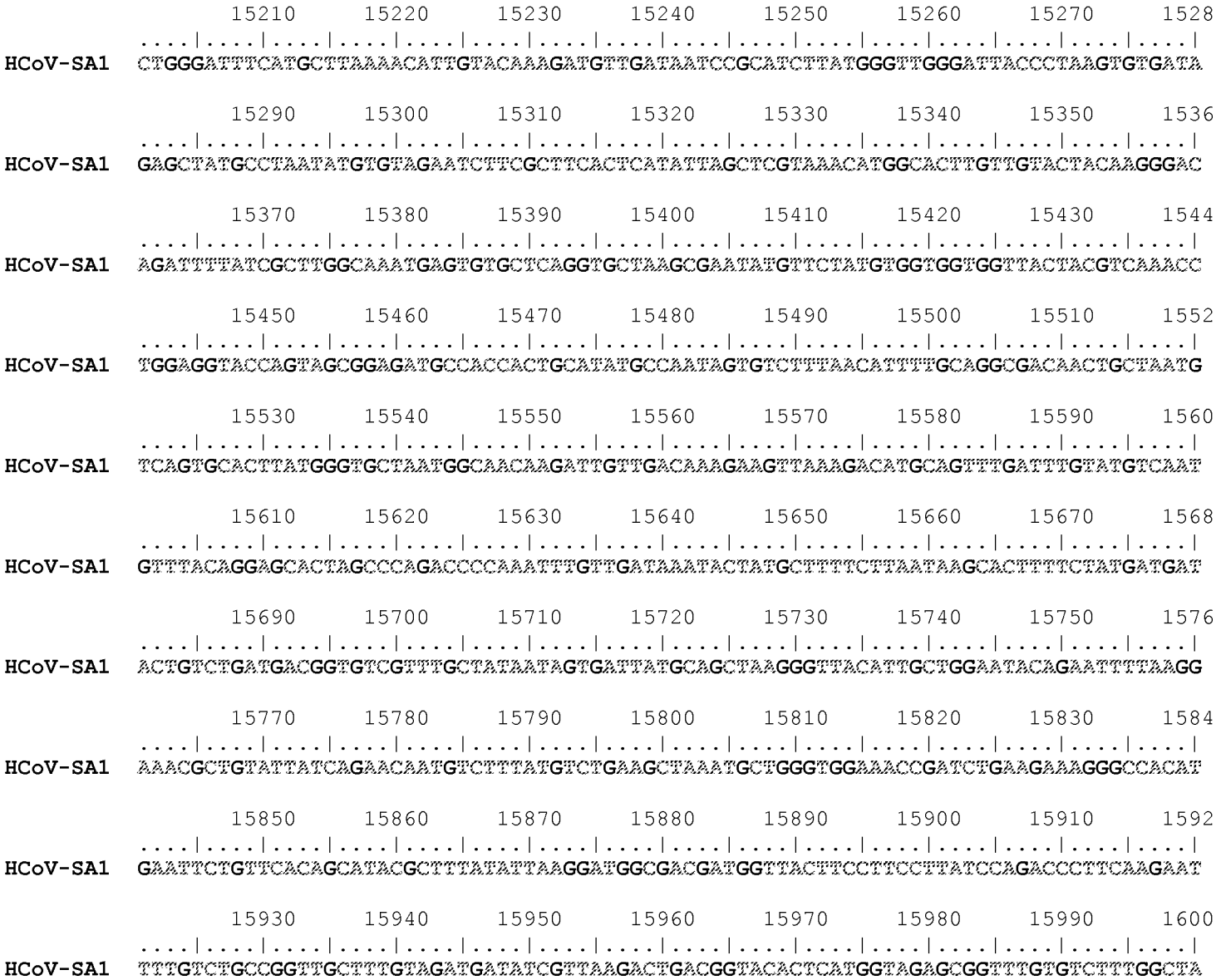
```
      11210     11220     11230     11240     11250     11260     11270     11280
....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|
HCoV-SA1 ACTCCCATTTCGTCAGCGCTGATTGCAGTTGCAAATTGGCTTGCCCCCACTAATGCTTATATGCGCACTACACATACTGA

      11290     11300     11310     11320     11330     11340     11350     11360
....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|
HCoV-SA1 TATTTGGTGTCTACATTAGTATTGTCACTTTGTCATTTAGTCTGAAGAGATTGTACAACCCATCACTTCACTTTCTAACTTTG

      11370     11380     11390     11400     11410     11420     11430     11440
....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|
HCoV-SA1 CGTTAGCATTGTGCAGTGGTGTAATGTGGTTGTACACTTAATAGCATTGGAGAAGCCTCAAGCCCCATTGCCTATCTGGTT

      11450     11460     11470     11480     11490     11500     11510     11520
....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|
HCoV-SA1 TTTGTCACTACACTCACTAGTGATTATACGATTACAGTCTTTGTTACTGTCAACCTTGCAAAAGTTTGCACTTATGCCAT

      11530     11540     11550     11560     11570     11580     11590     11600
....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|
HCoV-SA1 CTTTGCTTTACTCACCACAGCTTTACACTTTGTGTTTCCGGAAGTGAAGATGATACTTTTTATTTACACATGTTTTAGGTTTTCA

      11610     11620     11630     11640     11650     11660     11670     11680
....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|
HCoV-SA1 TGTGTACTTTGCTATTTTTGGTGTCTTCTCTCTTTTGAACCTTAAGCTTAGAGCACCTATGGGTGTCTATGACTTTAAGGTC

      11690     11700     11710     11720     11730     11740     11750     11760
....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|
HCoV-SA1 TCAACACAAGAGTTCAGATTCAATGACTGCTAACAATCTAACTGCACCTAGAAATTCTTGGGAGGCTATGGCTCTGAACTT

      11770     11780     11790     11800     11810     11820     11830     11840
....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|
HCoV-SA1 TAAGTTAATAGGTATTGGCGGTACACCTTGTATAAAGGTTGCTATGCAGTCTAAACTTACAGATCTTAAATGCACAT

      11850     11860     11870     11880     11890     11900     11910     11920
....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|
HCoV-SA1 CTGTGGTTCTCCTCTCTGCTCCAACAGTTACACTTAGAGGCTAATAGTAGGGCCTGGGCTTTCTGTGTTAAATGCCAT

      11930     11940     11950     11960     11970     11980     11990     12000
....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|
HCoV-SA1 AATGATATATTGGCAGCAACAGACCCCAGTGAGGCTTTCGAGAAATTCGTAAGTCTCTTTGCTACTTTAAATGACTTTTTC
```


12010 12020 12030 12040 12050 12060 12070 12080
HCoV-SA1 TGGTAATGTAGATCTTGATGCGTTAGCTAGTGATATTTTTGACACTCCTAGCGTACTTCAAGCTACTCTTTCTGAGTTTT
12090 12100 12110 12120 12130 12140 12150 12160
HCoV-SA1 CACACTTAGCTACCTTTGCTGAGTTGGAAGCTGCGCAGAAAGCCTATCAGGAAGCTATGGACTCTGGTGACACCTCACCA
12170 12180 12190 12200 12210 12220 12230 12240
HCoV-SA1 CAAGTTCTTAAGGCTTTGCAGAAGGCTGTTAATATAGCTAAAAACGCCTATGAGAAGGATAAGGCAGTGGCCCGTAAGTT
12250 12260 12270 12280 12290 12300 12310 12320
HCoV-SA1 AGAACGTATGGCTGATCAGGCTATGACTTCTATGTATAAGCAAGCACGTGCTGAAGACAAGAAAGCAAAAATTGTCAGTG
12330 12340 12350 12360 12370 12380 12390 12400
HCoV-SA1 CTATGCAAACTATGTTGTTTGGTATGATTAAGAAGCTCGACAACGATGTTCTTAATGGTATCATTTCTAACGCTAGGAAT
12410 12420 12430 12440 12450 12460 12470 12480
HCoV-SA1 GGTTGTATACCTCTTAGTGTCATCCCACTGTGTGCTTCAAATAAACTTCGCGTTGTAATTCCTGACTTCACCGTCTGGAA
12490 12500 12510 12520 12530 12540 12550 12560
HCoV-SA1 TCAGGTAGTCACATATCCCTCGCTTAACTACGCTGGGGCTTTGTGGGACATTACAGTTATAAACAATGTGGACAATGAAA
12570 12580 12590 12600 12610 12620 12630 12640
HCoV-SA1 TTGTTAAGTCTTCAGATGTTGTAGACAGCAATGAAAATTTAACATGGCCACTTGTTTTAGAATGCACTAGGGCATCCACT
12650 12660 12670 12680 12690 12700 12710 12720
HCoV-SA1 TCTGCCGTTAAGTTGCAAAATAATGAGATCAAACCTTCAGGTCTAAAAACCATGGTTGTGTCTGCGGGTCAAGAGCAAAC
12730 12740 12750 12760 12770 12780 12790 12800
HCoV-SA1 TAACTGTAATACTAGTTCCTTAGCTTATTACGAACCTGTGCAGGGTCGTAAAATGCTGATGGCTCTTCTTTCTGATAATG

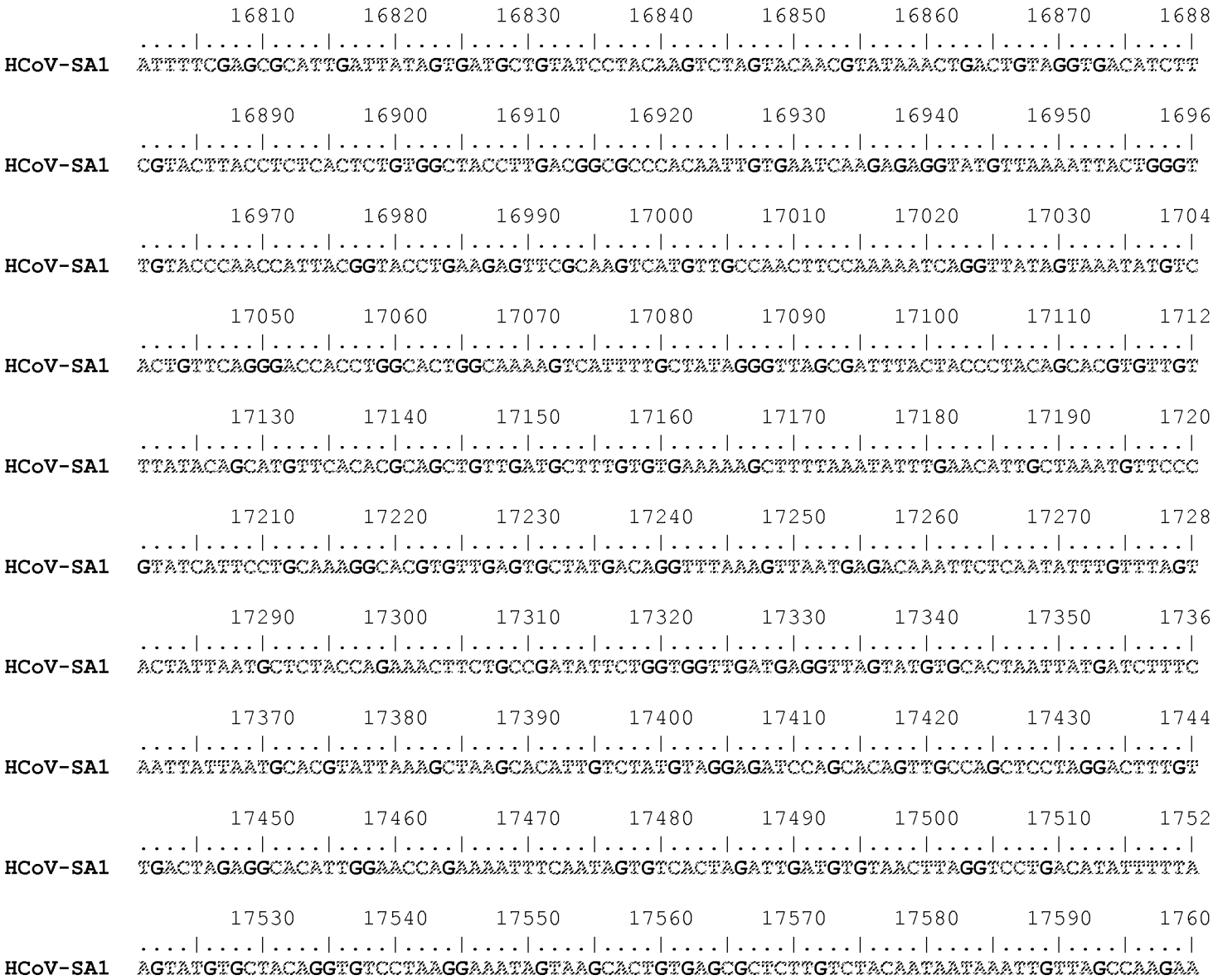

12810 12820 12830 12840 12850 12860 12870 12880
HCoV-SA1 CCTATCTCAAATGGGCGCGTGTTGAAGGTAAGGACGGATTTGTCAGTGTAGAGCTACAACCTCCTTGCAAATTCTTGATT
12890 12900 12910 12920 12930 12940 12950 12960
HCoV-SA1 GCGGGACCAAAAGGACCTGAAATCCGATATCTCTATTTTGTTAAAAATCTTAACAACCTTCATCGCGGGCAAGTGTTAGG
12970 12980 12990 13000 13010 13020 13030 13040
HCoV-SA1 GCACATTGCTGCGACTGTTAGATTGCAAGCTGGTTCTAACACCGAGTTTGCCTCTAATTCCTCGGTGTTGTCACTTGTTA
13050 13060 13070 13080 13090 13100 13110 13120
HCoV-SA1 ACTTCACCGTTGATCCTCAAAAAGCTTATCTCGATTTCGTCAATGCGGGAGGTGCCCCATTGACAAATTGTGTTAAGATG
13130 13140 13150 13160 13170 13180 13190 13200
HCoV-SA1 CTTACTCCTAAAACTGGTACAGGTATAGCTATATCTGTTAAACCAGAGAGTACAGCTGATCAAGAGACTTATGGTGGAGC
13210 13220 13230 13240 13250 13260 13270 13280
HCoV-SA1 TTCAGTGTGTCTCTATTGCCGTGCGCATATAGAACATCCTGATGTCTCTGGTGTTTGTAAATATAAGGGTAAGTTTGTCC
13290 13300 13310 13320 13330 13340 13350 13360
HCoV-SA1 AAATCCCTGCTCAGTGTGTCCGTGACCCTGTGGGATTTTGTTTGTCAAATACCCCCTGTAATGTCTGTCAATATTGGATT
13370 13380 13390 13400 13410 13420 13430 13440
HCoV-SA1 GGATATGGGTGCAATTGTGACTCGCTTAGGCAAGCAGCACTGCCCCAATCTAAAGATTCCAATTTTTTAAACGAGTCCGG
13450 13460 13470 13480 13490 13500 13510 13520
HCoV-SA1 GGTTCTATTGTAAATGCCCGAATAGAACCCTGTTCAAGTGGTTTGTCCACTGATGTCGTCTTTAGGGCATTTGACATCTG
13530 13540 13550 13560 13570 13580 13590 13600
HCoV-SA1 CAACTATAAGGCTAAGGTTGCTGGTATTGGAAAATACTACAAGACTAATACTTGTAGGTTTGTAGAATTAGATGACCAAG

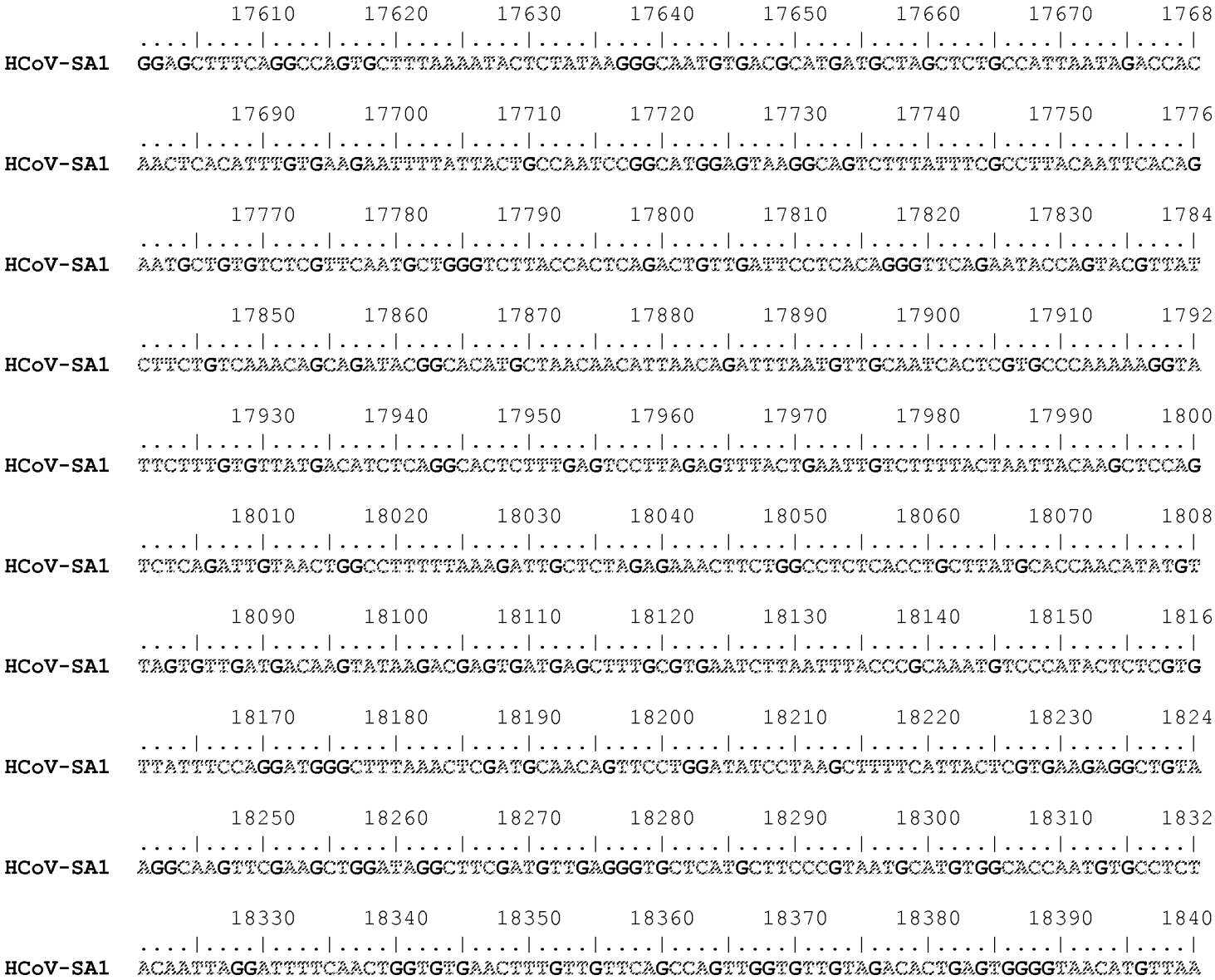
```
      13610     13620     13630     13640     13650     13660     13670     13680
....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|
HCoV-SA1 GGCATCATTTAGACTCCTATTTTGTCGTTAAGAGGCATACTATGGAGAAATTATGAACTAGAGAAGCACTGTTACGACTTG

      13690     13700     13710     13720     13730     13740     13750     13760
....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|
HCoV-SA1 TTACGTGACTGTGATGCTGTTAGCTCCCCATGATTTCTTCATCTTTGATGTAGACAAAGTTAAAACACCTCATATTGTACG

      13770     13780     13790     13800     13810     13820     13830     13840
....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|
HCoV-SA1 TCAGCGTTTAACTGAGTACACTATGATGGATCTTGTATATGCCCTGAGGCACTTTGATCAAAATAGCGAAGTGCTTAAGG

      13850     13860     13870     13880     13890     13900     13910     13920
....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|
HCoV-SA1 CTATCTTAGTGAAGTATGGTTGCTGTGATGTTACCTACTTTGAAAATAAACTCTGGTTTGATTTTGTTGAAAATCCCAGT

      13930     13940     13950     13960     13970     13980     13990     14000
....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|
HCoV-SA1 GTTATTGGTGTTTATCATAAACTTGGAGACGTGTACGCCAAGCTATCTTAAACACTGTTAAATTTTGTGACCACATGGT

      14010     14020     14030     14040     14050     14060     14070     14080
....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|
HCoV-SA1 CAAGGCTGGTTTAGTCGGTGTGCTCACACTAGACAACCAGGACCTTAAATGGCAAGTGGTATGATTTTGGTGACTTCGTAA

      14090     14100     14110     14120     14130     14140     14150     14160
....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|
HCoV-SA1 TCACTCAACCTGGTTCAGGAGTAGCTTAGTTGATAGCTACTATTCTTATTTGATGCCTGTGCTCTCAAATGACCGAATTGT

      14170     14180     14190     14200     14210     14220     14230     14240
....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|
HCoV-SA1 CTGGCCGCTGAGACACATAGGGATTGTGATTTTAATAAACCACTCATTGAGTGGCCACTTACTGAGTATGATTTTACTGA

      14250     14260     14270     14280     14290     14300     14310     14320
....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|
HCoV-SA1 TTATAAGGTACAACTCTTTGAGAAGTACTTTAAATATTGGGATCAGACGTATCACGCAAATTGCGTTAATTGTACTGATG

      14330     14340     14350     14360     14370     14380     14390     14400
....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|
HCoV-SA1 ACCGTTGTGTTACATTGTGCTAATTTCAATGTATTTGCTATGACCATGCCTAAGACTTGTTTCGGACCCATAGTC
```

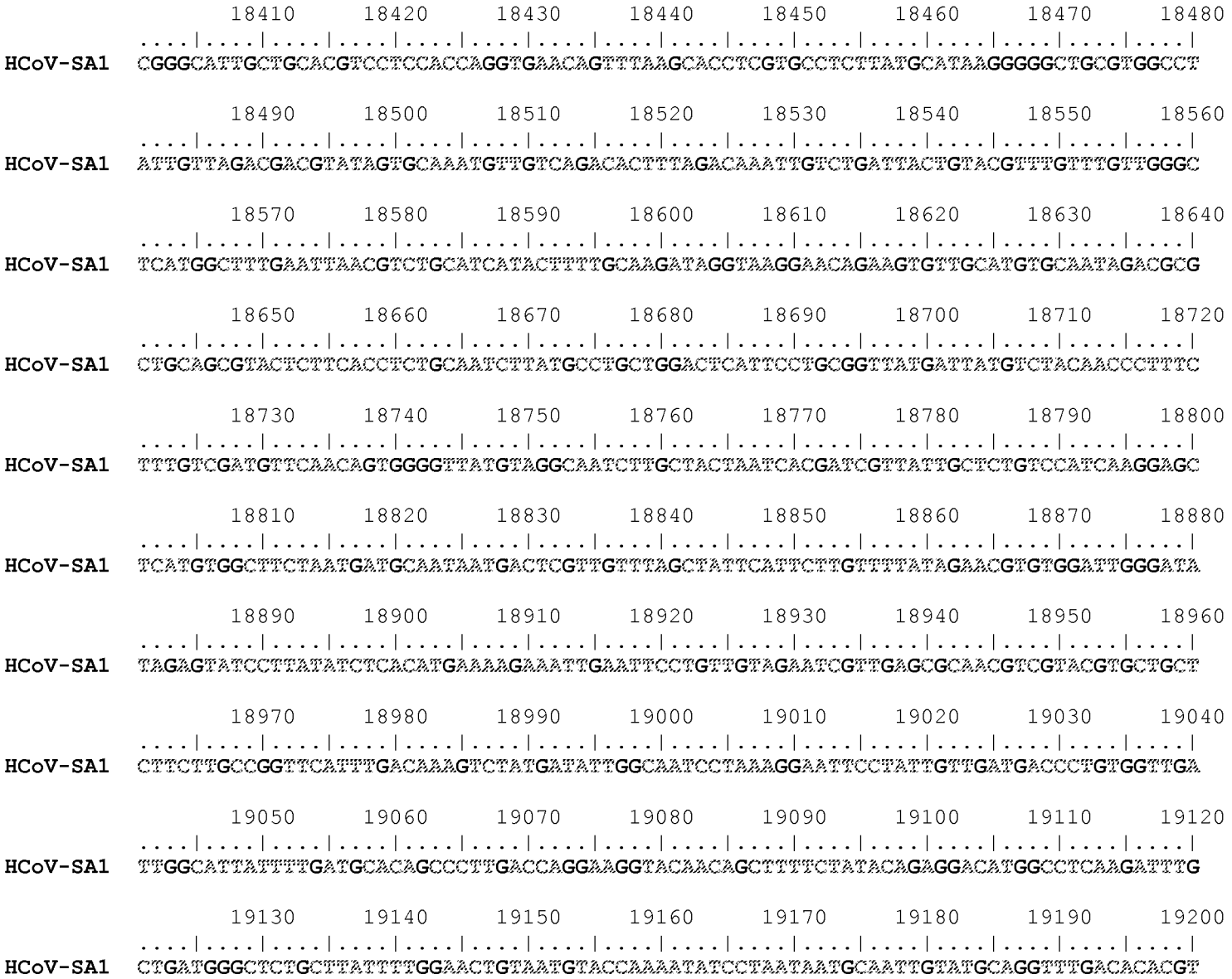

14410 14420 14430 14440 14450 14460 14470 14480
HCoV-SA1 CGAAAGATCTTTGTTGATGGCGTGCCATTTGTAGTATCTTGTGGTTATCACTACAAAGAATTAGGTTTAGTCATGAATAT
14490 14500 14510 14520 14530 14540 14550 14560
HCoV-SA1 GGATGTTAGTCTCCATAGACATAGGCTCTCTCTTAAGGAGTTGATGATGTATGCCGCTGATCCAGCCATGCACATTGCCT
14570 14580 14590 14600 14610 14620 14630 14640
HCoV-SA1 CCTCTAACGCTTTTCTTGATTTGAGGACATCATGTTTTAGTGTCGCTGCACTTACAACTGGTTTGACTTTTCAAACTGTG
14650 14660 14670 14680 14690 14700 14710 14720
HCoV-SA1 CGGCCTGGCAATTTTAACCAAGACTTCTATGATTTCGTGGTATCTAAAGGTTTCTTTAAGGAGGGCTCTTCAGTGACGCT
14730 14740 14750 14760 14770 14780 14790 14800
HCoV-SA1 CAAACATTTTTTCTTTGCTCAAGATGGTAATGCTGCTATTACAGATTATAATTACTATTCTTATAATCTGCCTACTATGT
14810 14820 14830 14840 14850 14860 14870 14880
HCoV-SA1 GTGACATCAAACAAATGTTGTTCTGCATGGAAGTTGTAAACAAGTACTTCGAAATCTATGACGGTGGTTGTCTTAATGCT
14890 14900 14910 14920 14930 14940 14950 14960
HCoV-SA1 TCTGAAGTGGTTGTTAATAATTTAGACAAGAGTGCTGGCCATCCTTTTAATAAGTTTGGCAAAGCTCGTGTCTATTATGA
14970 14980 14990 15000 15010 15020 15030 15040
HCoV-SA1 GAGCATGTCTTACCAGGAGCAAGATGAACTTTTTGCCATGACAAAGCGTAACGTCATTCCTACCATGACTCAAATGAATC
15050 15060 15070 15080 15090 15100 15110 15120
HCoV-SA1 TAAAATATGCTATTAGTGCTAAGAATAGAGCTCGCACTGTTGCAGGCGTGTCCATACTTAGCACAATGACTAATCGCCAG
15130 15140 15150 15160 15170 15180 15190 15200
HCoV-SA1 TACCATCAGAAAATGCTTAAGTCCATGGCTGCAACTCGTGGAGCGACTTGCGTCATTGGTACTACAAAGTTCTACGGTGG

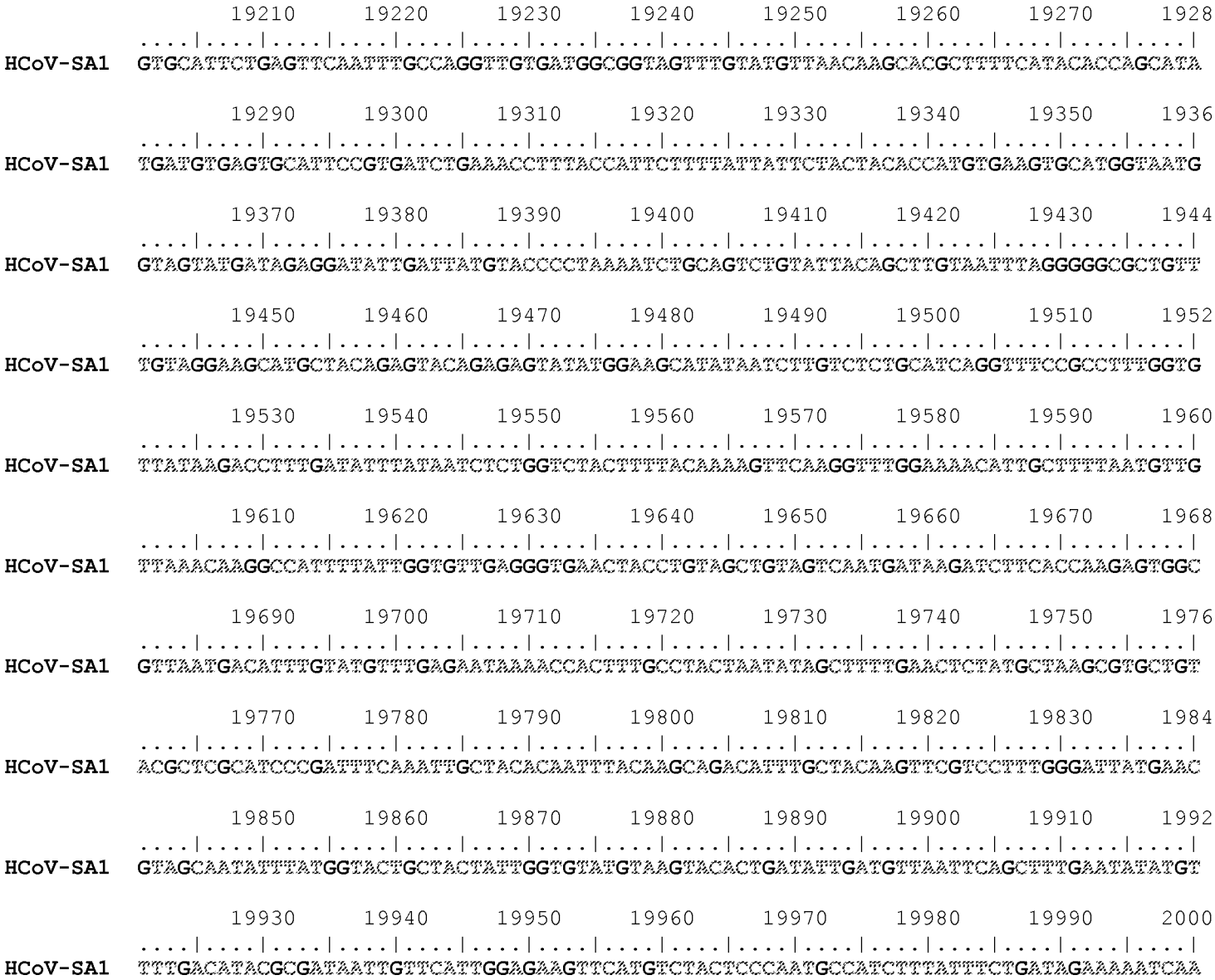

15210 15220 15230 15240 15250 15260 15270 15280
HCoV-SA1 CTGGGATTTCATGCTTAAAACATTGTACAAAGATGTTGATAATCCGCATCTTATGGGTTGGGATTACCCTAAGTGTGATA
15290 15300 15310 15320 15330 15340 15350 15360
HCoV-SA1 GAGCTATGCCTAATATGTGTAGAATCTTCGCTTCACTCATATTAGCTCGTAAACATGGCACTTGTTGTACTACAAGGGAC
15370 15380 15390 15400 15410 15420 15430 15440
HCoV-SA1 AGATTTTATCGCTTGGCAAATGAGTGTGCTCAGGTGCTAAGCGAATATGTTCTATGTGGTGGTGGTTACTACGTCAAACC
15450 15460 15470 15480 15490 15500 15510 15520
HCoV-SA1 TGGAGGTACCAGTAGCGGAGATGCCACCACTGCATATGCCAATAGTGTCTTTAACATTTTGCAGGCGACAACTGCTAATG
15530 15540 15550 15560 15570 15580 15590 15600
HCoV-SA1 TCAGTGCACTTATGGGTGCTAATGGCAACAAGATTGTTGACAAAGAAGTTAAAGACATGCAGTTTGATTTGTATGTCAAT
15610 15620 15630 15640 15650 15660 15670 15680
HCoV-SA1 GTTTACAGGAGCACTAGCCCAGACCCCAAATTTGTTGATAAATACTATGCTTTTCTTAATAAGCACTTTTCTATGATGAT
15690 15700 15710 15720 15730 15740 15750 15760
HCoV-SA1 ACTGTCTGATGACGGTGTCGTTTGCTATAATAGTGATTATGCAGCTAAGGGTTACATTGCTGGAATACAGAATTTTAAGG
15770 15780 15790 15800 15810 15820 15830 15840
HCoV-SA1 AAACGCTGTATTATCAGAACAATGTCTTTATGTCTGAAGCTAAATGCTGGGTGGAAACCGATCTGAAGAAAGGGCCACAT
15850 15860 15870 15880 15890 15900 15910 15920
HCoV-SA1 GAATTCTGTTCACAGCATACGCTTTATATTAAGGATGGCGACGATGGTTACTTCCTTCCTTATCCAGACCCTTCAAGAAT
15930 15940 15950 15960 15970 15980 15990 16000
HCoV-SA1 TTTGTCTGCCGGTTGCTTTGTAGATGATATCGTTAAGACTGACGGTACACTCATGGTAGAGCGGTTTGTGTCTTTGGCTA

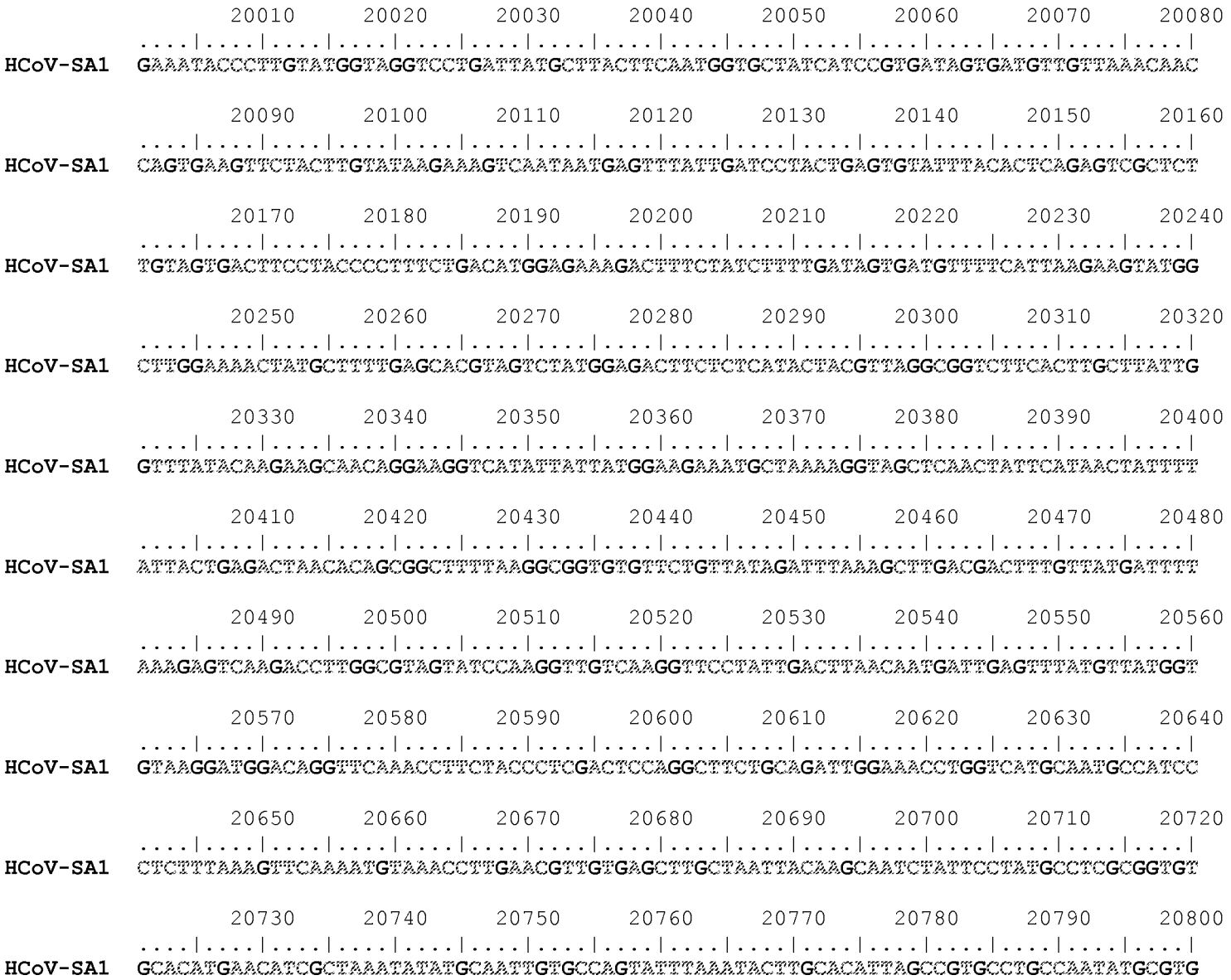
```
          16010     16020     16030     16040     16050     16060     16070     16080
          ....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|
HCoV-SA1  TAGATGCTTACCCTCTCACAAAGCATGAAGATATAGAATACCAGAATGTATTCTGGGTCTACTTACAGTATATAGAAAAA

          16090     16100     16110     16120     16130     16140     16150     16160
          ....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|
HCoV-SA1  CTGTATAAAAGACCTTACAGGACACATGCTTGACAGTTATTCTGTCATGCTATGTGGTGAAAATTCTGCTAAGTTTTTGGGA

          16170     16180     16190     16200     16210     16220     16230     16240
          ....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|
HCoV-SA1  AGAGGCATTCTATAGAGATCTCTATAGTTCGCCTACCACTTTGCAGGCTGTCGGTTCATGCGTTGTATGCCATTCACAGA

          16250     16260     16270     16280     16290     16300     16310     16320
          ....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|
HCoV-SA1  CTTCCCTACGCTGTGGGACATGCATCCGTAGACCATTTCTCTGCTGTAAAATGCTGCTATGATCATGTTTATAGCAACTCCA

          16330     16340     16350     16360     16370     16380     16390     16400
          ....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|
HCoV-SA1  CATAAGATGGTTTTGTCTGTTTTCTCCTTACGTTTGTAATGCCCCTGGTTGTGGCGTTTCAGACGTTACTAAGCTATATTT

          16410     16420     16430     16440     16450     16460     16470     16480
          ....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|
HCoV-SA1  AGGTGGTATGAGCTACTTTTGTGTAGATCATAGACCTGTGTGTAGTTTTCCACTTTGGCTAATGGTCTTGTATTCGGCT

          16490     16500     16510     16520     16530     16540     16550     16560
          ....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|
HCoV-SA1  TATACAGAATATGTGCACAGGTAGTCCTTCTATAGTTGAATTTAATAGGTTGGCTACCTGTGACTGGACTGAAAGTGGT

          16570     16580     16590     16600     16610     16620     16630     16640
          ....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|
HCoV-SA1  GATTACACCCTTGCCAATACTACAACAGAACCACTCAAACTTTTTGCTGCTGAGACTTTACGTGCCACTGAAGAGGCGTC

          16650     16660     16670     16680     16690     16700     16710     16720
          ....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|
HCoV-SA1  TAAGCAGTCTTATGCTATTGCCACCATCAAAGAAATTGTTGGTGAGCGCCAACTATTACTTGTGTGGGAGGCTGGCAAGT

          16730     16740     16750     16760     16770     16780     16790     16800
          ....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|
HCoV-SA1  CCAAACCACACTCAATCGTAATTATGTTTTTACTGGTTTATCATATAACCAAAAAATAGTAAAGTGCAGCTCGGTGAGTAC
```

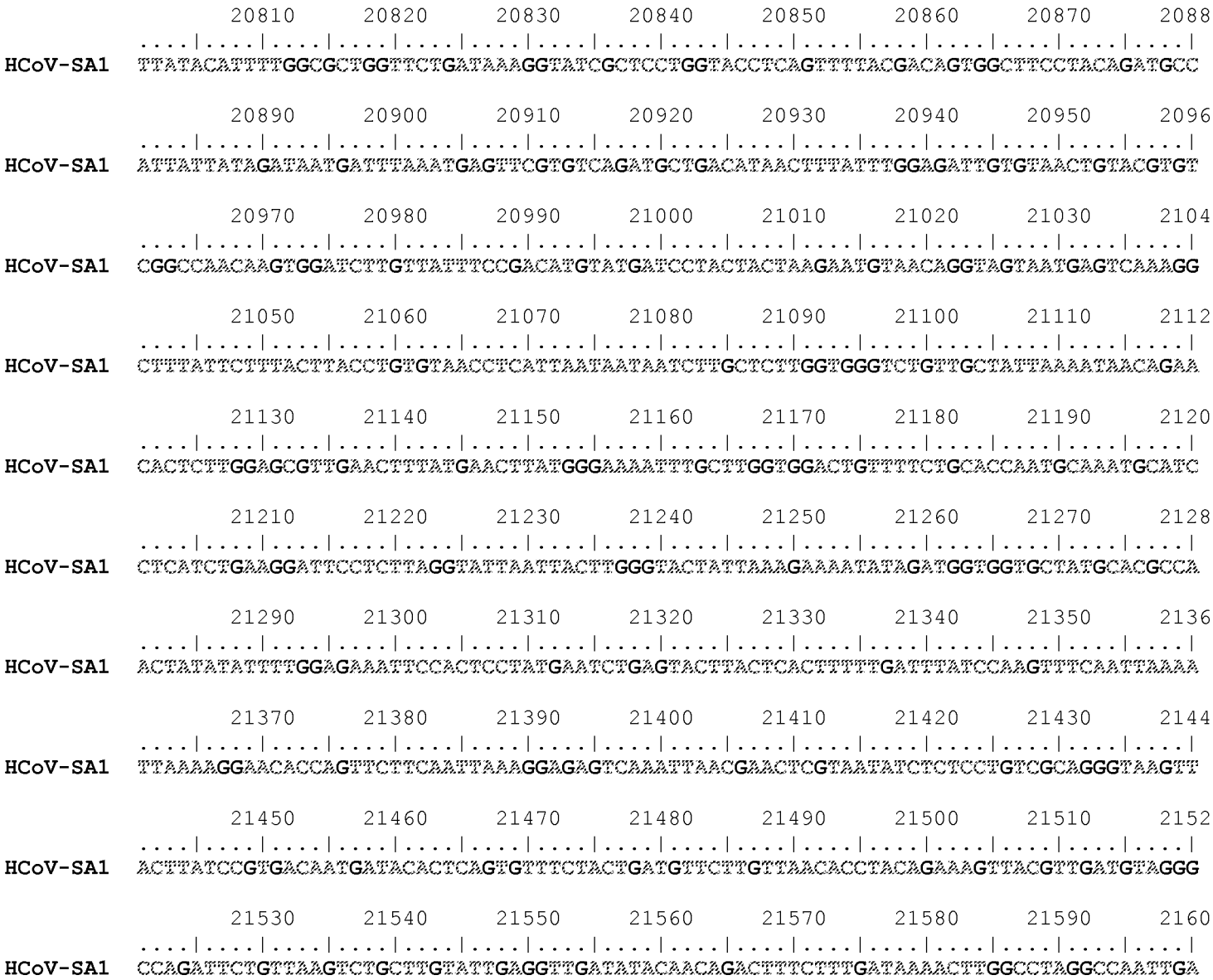

16810 16820 16830 16840 16850 16860 16870 16880
HCoV-SA1 ATTTTCGAGCGCATTGATTATAGTGATGCTGTATCCTACAAGTCTAGTACAACGTATAAACTGACTGTAGGTGACATCTT
16890 16900 16910 16920 16930 16940 16950 16960
HCoV-SA1 CGTACTTACCTCTCACTCTGTGGCTACCTTGACGGCGCCCACAATTGTGAATCAAGAGAGGTATGTTAAAATTACTGGGT
16970 16980 16990 17000 17010 17020 17030 17040
HCoV-SA1 TGTACCCAACCATTACGGTACCTGAAGAGTTCGCAAGTCATGTTGCCAACTTCCAAAAATCAGGTTATAGTAAATATGTC
17050 17060 17070 17080 17090 17100 17110 17120
HCoV-SA1 ACTGTTCAGGGACCACCTGGCACTGGCAAAAGTCATTTTGCTATAGGGTTAGCGATTTACTACCCTACAGCACGTGTTGT
17130 17140 17150 17160 17170 17180 17190 17200
HCoV-SA1 TTATACAGCATGTTCACACGCAGCTGTTGATGCTTTGTGTGAAAAAGCTTTTAAATATTTGAACATTGCTAAATGTTCCC
17210 17220 17230 17240 17250 17260 17270 17280
HCoV-SA1 GTATCATTCCTGCAAAGGCACGTGTTGAGTGCTATGACAGGTTTAAAGTTAATGAGACAAATTCTCAATATTTGTTTAGT
17290 17300 17310 17320 17330 17340 17350 17360
HCoV-SA1 ACTATTAATGCTCTACCAGAAACTTCTGCCGATATTCTGGTGGTTGATGAGGTTAGTATGTGCACTAATTATGATCTTTC
17370 17380 17390 17400 17410 17420 17430 17440
HCoV-SA1 AATTATTAATGCACGTATTAAAGCTAAGCACATTGTCTATGTAGGAGATCCAGCACAGTTGCCAGCTCCTAGGACTTTGT
17450 17460 17470 17480 17490 17500 17510 17520
HCoV-SA1 TGACTAGAGGCACATTGGAACCAGAAAATTTCAATAGTGTCACTAGATTGATGTGTAACTTAGGTCCTGACATATTTTTA
17530 17540 17550 17560 17570 17580 17590 17600
HCoV-SA1 AGTATGTGCTACAGGTGTCCTAAGGAAATAGTAAGCACTGTGAGCGCTCTTGTCTACAATAATAAATTGTTAGCCAAGAA

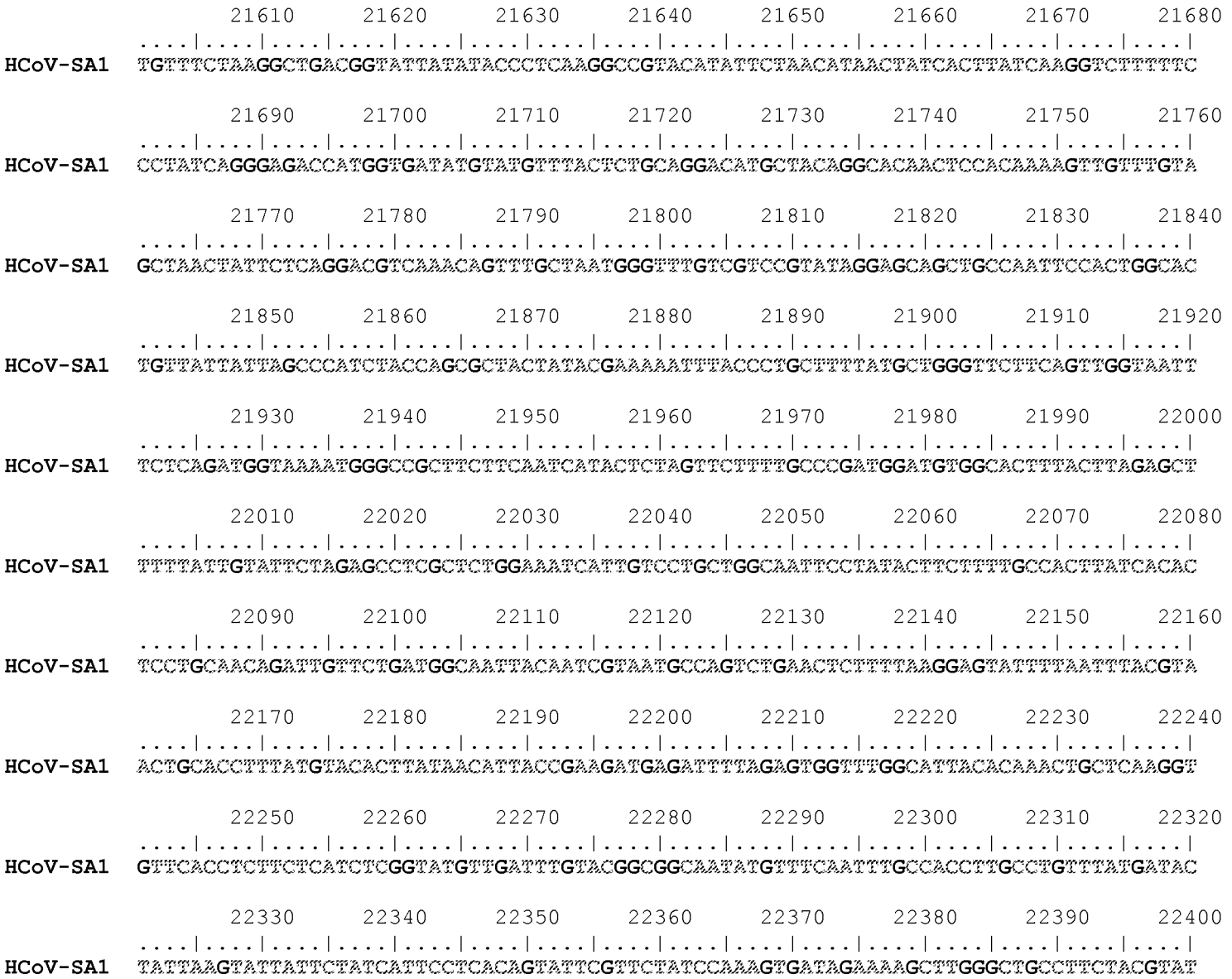

17610 17620 17630 17640 17650 17660 17670 17680
HCoV-SA1 GGAGCTTTCAGGCCAGTGCTTTAAAATACTCTATAAGGGCAATGTGACGCATGATGCTAGCTCTGCCATTAATAGACCAC
17690 17700 17710 17720 17730 17740 17750 17760
HCoV-SA1 AACTCACATTTGTGAAGAATTTTATTACTGCCAATCCGGCATGGAGTAAGGCAGTCTTTATTTCGCCTTACAATTCACAG
17770 17780 17790 17800 17810 17820 17830 17840
HCoV-SA1 AATGCTGTGTCTCGTTCAATGCTGGGTCTTACCACTCAGACTGTTGATTCCTCACAGGGTTCAGAATACCAGTACGTTAT
17850 17860 17870 17880 17890 17900 17910 17920
HCoV-SA1 CTTCTGTCAAACAGCAGATACGGCACATGCTAACAACATTAACAGATTTAATGTTGCAATCACTCGTGCCCAAAAAGGTA
17930 17940 17950 17960 17970 17980 17990 18000
HCoV-SA1 TTCTTTGTGTTATGACATCTCAGGCACTCTTTGAGTCCTTAGAGTTTACTGAATTGTCTTTTACTAATTACAAGCTCCAG
18010 18020 18030 18040 18050 18060 18070 18080
HCoV-SA1 TCTCAGATTGTAACTGGCCTTTTTAAAGATTGCTCTAGAGAAACTTCTGGCCTCTCACCTGCTTATGCACCAACATATGT
18090 18100 18110 18120 18130 18140 18150 18160
HCoV-SA1 TAGTGTTGATGACAAGTATAAGACGAGTGATGAGCTTTGCGTGAATCTTAATTTACCCGCAAATGTCCCATACTCTCGTG
18170 18180 18190 18200 18210 18220 18230 18240
HCoV-SA1 TTATTTCCAGGATGGGCTTTAAACTCGATGCAACAGTTCCTGGATATCCTAAGCTTTTCATTACTCGTGAAGAGGCTGTA
18250 18260 18270 18280 18290 18300 18310 18320
HCoV-SA1 AGGCAAGTTCGAAGCTGGATAGGCTTCGATGTTGAGGGTGCTCATGCTTCCCGTAATGCATGTGGCACCAATGTGCCTCT
18330 18340 18350 18360 18370 18380 18390 18400
HCoV-SA1 ACAATTAGGATTTTCAACTGGTGTGAACTTTGTTGTTCAGCCAGTTGGTGTTGTAGACACTGAGTGGGGTAACATGTTAA

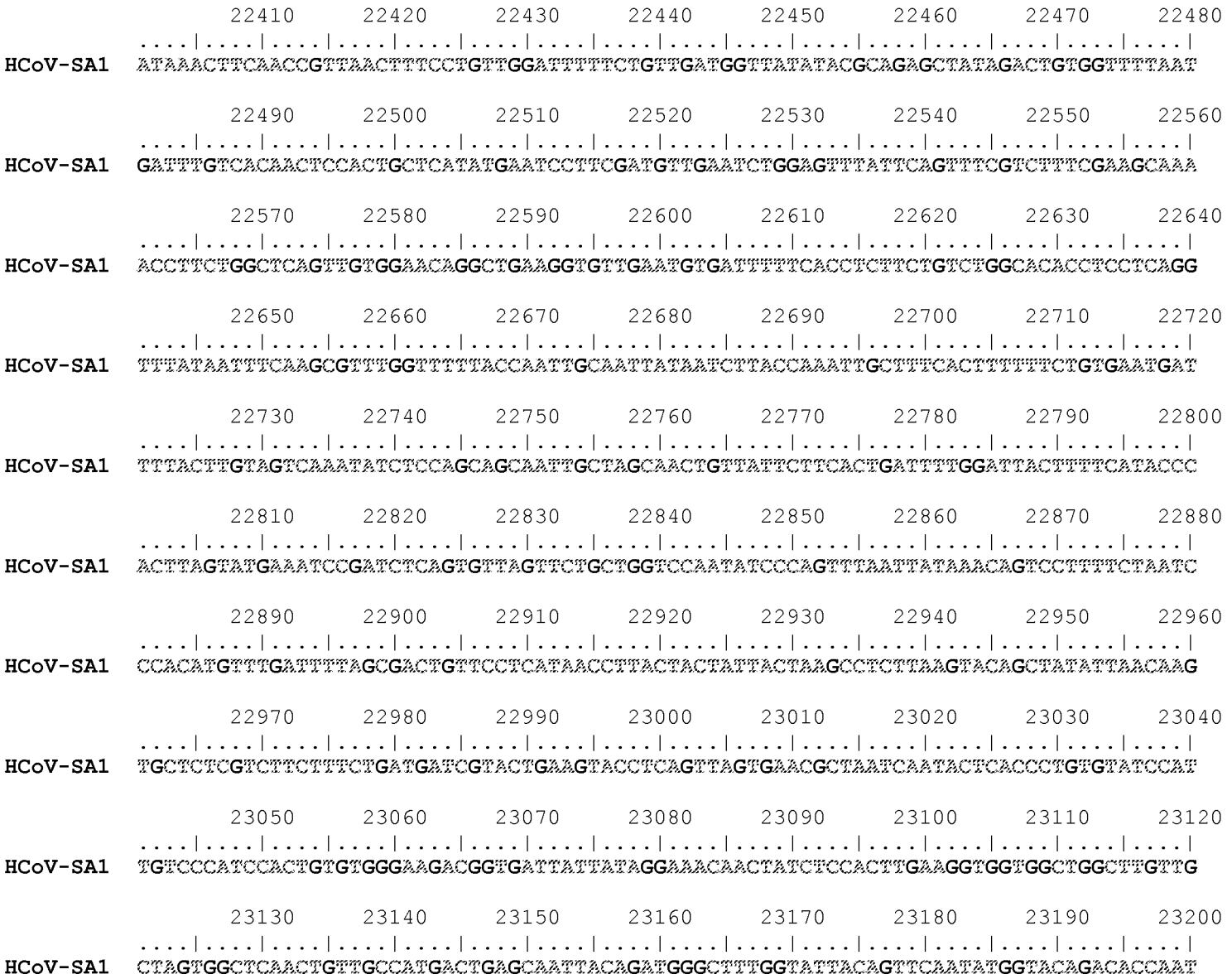

18410 18420 18430 18440 18450 18460 18470 18480
HCoV-SA1 CGGGCATTGCTGCACGTCCTCCACCAGGTGAACAGTTTAAGCACCTCGTGCCTCTTATGCATAAGGGGGCTGCGTGGCCT
18490 18500 18510 18520 18530 18540 18550 18560
HCoV-SA1 ATTGTTAGACGACGTATAGTGCAAATGTTGTCAGACACTTTAGACAAATTGTCTGATTACTGTACGTTTGTTTGTTGGGC
18570 18580 18590 18600 18610 18620 18630 18640
HCoV-SA1 TCATGGCTTTGAATTAACGTCTGCATCATACTTTTGCAAGATAGGTAAGGAACAGAAGTGTTGCATGTGCAATAGACGCG
18650 18660 18670 18680 18690 18700 18710 18720
HCoV-SA1 CTGCAGCGTACTCTTCACCTCTGCAATCTTATGCCTGCTGGACTCATTCCTGCGGTTATGATTATGTCTACAACCCTTTC
18730 18740 18750 18760 18770 18780 18790 18800
HCoV-SA1 TTTGTCGATGTTCAACAGTGGGGTTATGTAGGCAATCTTGCTACTAATCACGATCGTTATTGCTCTGTCCATCAAGGAGC
18810 18820 18830 18840 18850 18860 18870 18880
HCoV-SA1 TCATGTGGCTTCTAATGATGCAATAATGACTCGTTGTTTAGCTATTCATTCTTGTTTTATAGAACGTGTGGATTGGGATA
18890 18900 18910 18920 18930 18940 18950 18960
HCoV-SA1 TAGAGTATCCTTATATCTCACATGAAAAGAAATTGAATTCCTGTTGTAGAATCGTTGAGCGCAACGTCGTACGTGCTGCT
18970 18980 18990 19000 19010 19020 19030 19040
HCoV-SA1 CTTCTTGCCGGTTCATTTGACAAAGTCTATGATATTGGCAATCCTAAAGGAATTCCTATTGTTGATGACCCTGTGGTTGA
19050 19060 19070 19080 19090 19100 19110 19120
HCoV-SA1 TTGGCATTATTTTGATGCACAGCCCTTGACCAGGAAGGTACAACAGCTTTTCTATACAGAGGACATGGCCTCAAGATTTG
19130 19140 19150 19160 19170 19180 19190 19200
HCoV-SA1 CTGATGGGCTCTGCTTATTTTGGAACTGTAATGTACCAAAATATCCTAATAATGCAATTGTATGCAGGTTTGACACACGT

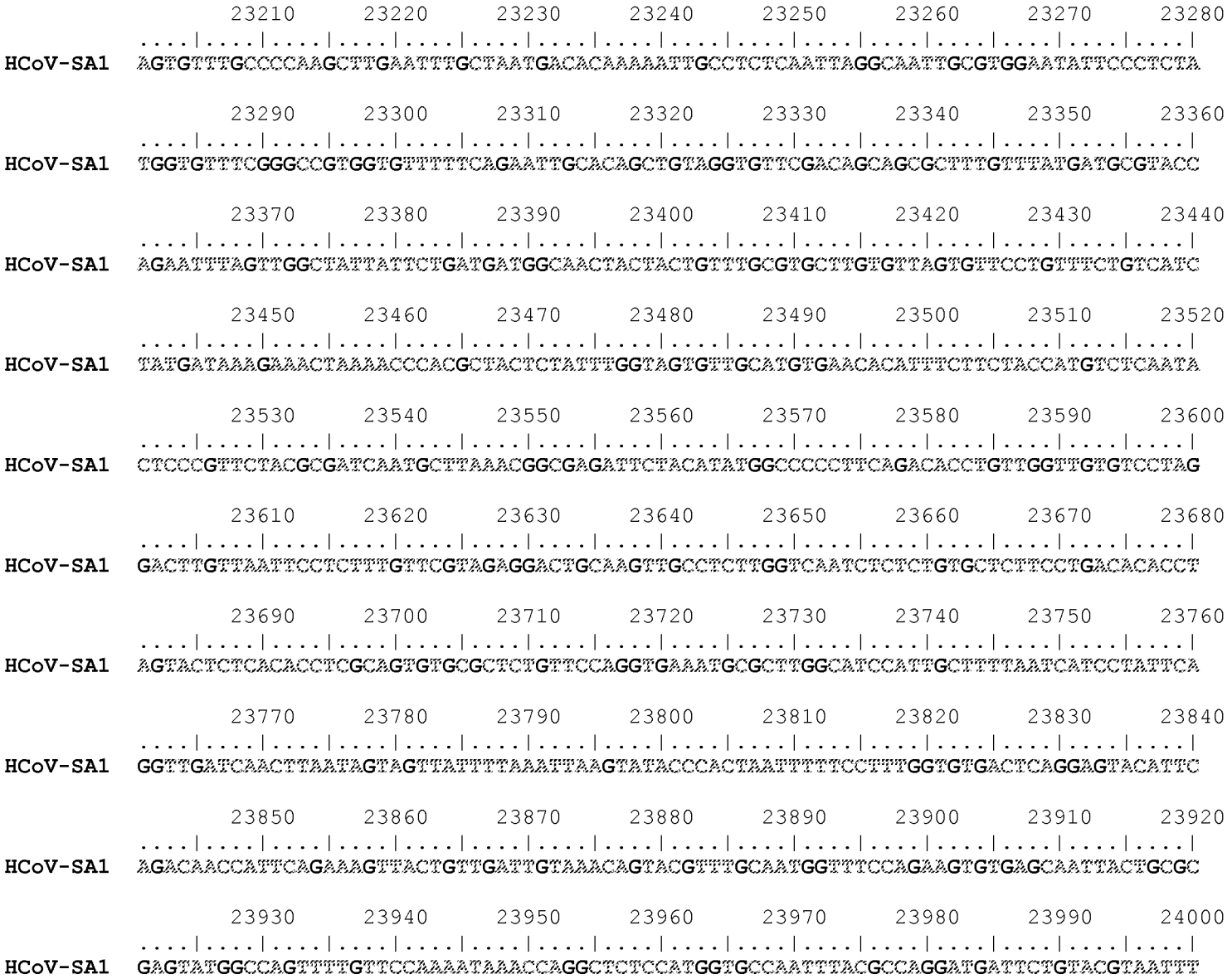

19210 19220 19230 19240 19250 19260 19270 19280
HCoV-SA1 GTGCATTCTGAGTTCAATTTGCCAGGTTGTGATGGCGGTAGTTTGTATGTTAACAAGCACGCTTTTCATACACCAGCATA
19290 19300 19310 19320 19330 19340 19350 19360
HCoV-SA1 TGATGTGAGTGCATTCCGTGATCTGAAACCTTTACCATTCTTTTATTATTCTACTACACCATGTGAAGTGCATGGTAATG
19370 19380 19390 19400 19410 19420 19430 19440
HCoV-SA1 GTAGTATGATAGAGGATATTGATTATGTACCCCTAAAATCTGCAGTCTGTATTACAGCTTGTAATTTAGGGGGCGCTGTT
19450 19460 19470 19480 19490 19500 19510 19520
HCoV-SA1 TGTAGGAAGCATGCTACAGAGTACAGAGAGTATATGGAAGCATATAATCTTGTCTCTGCATCAGGTTTCCGCCTTTGGTG
19530 19540 19550 19560 19570 19580 19590 19600
HCoV-SA1 TTATAAGACCTTTGATATTTATAATCTCTGGTCTACTTTTACAAAAGTTCAAGGTTTGGAAAACATTGCTTTTAATGTTG
19610 19620 19630 19640 19650 19660 19670 19680
HCoV-SA1 TTAAACAAGGCCATTTTATTGGTGTTGAGGGTGAACTACCTGTAGCTGTAGTCAATGATAAGATCTTCACCAAGAGTGGC
19690 19700 19710 19720 19730 19740 19750 19760
HCoV-SA1 GTTAATGACATTTGTATGTTTGAGAATAAAACCACTTTGCCTACTAATATAGCTTTTGAACTCTATGCTAAGCGTGCTGT
19770 19780 19790 19800 19810 19820 19830 19840
HCoV-SA1 ACGCTCGCATCCCGATTTCAAATTGCTACACAATTTACAAGCAGACATTTGCTACAAGTTCGTCCTTTGGGATTATGAAC
19850 19860 19870 19880 19890 19900 19910 19920
HCoV-SA1 GTAGCAATATTTATGGTACTGCTACTATTGGTGTATGTAAGTACACTGATATTGATGTTAATTCAGCTTTGAATATATGT
19930 19940 19950 19960 19970 19980 19990 20000
HCoV-SA1 TTTGACATACGCGATAATTGTTCATTGGAGAAGTTCATGTCTACTCCCAATGCCATCTTTATTTCTGATAGAAAAATCAA


20010 20020 20030 20040 20050 20060 20070 20080
HCoV-SA1 GAAATACCCTTGTATGGTAGGTCCTGATTATGCTTACTTCAATGGTGCTATCATCCGTGATAGTGATGTTGTTAAACAAC
20090 20100 20110 20120 20130 20140 20150 20160
HCoV-SA1 CAGTGAAGTTCTACTTGTATAAGAAAGTCAATAATGAGTTTATTGATCCTACTGAGTGTATTTACACTCAGAGTCGCTCT
20170 20180 20190 20200 20210 20220 20230 20240
HCoV-SA1 TGTAGTGACTTCCTACCCCTTTCTGACATGGAGAAAGACTTTCTATCTTTTGATAGTGATGTTTTCATTAAGAAGTATGG
20250 20260 20270 20280 20290 20300 20310 20320
HCoV-SA1 CTTGGAAAACTATGCTTTTGAGCACGTAGTCTATGGAGACTTCTCTCATACTACGTTAGGCGGTCTTCACTTGCTTATTG
20330 20340 20350 20360 20370 20380 20390 20400
HCoV-SA1 GTTTATACAAGAAGCAACAGGAAGGTCATATTATTATGGAAGAAATGCTAAAAGGTAGCTCAACTATTCATAACTATTTT
20410 20420 20430 20440 20450 20460 20470 20480
HCoV-SA1 ATTACTGAGACTAACACAGCGGCTTTTAAGGCGGTGTGTTCTGTTATAGATTTAAAGCTTGACGACTTTGTTATGATTTT
20490 20500 20510 20520 20530 20540 20550 20560
HCoV-SA1 AAAGAGTCAAGACCTTGGCGTAGTATCCAAGGTTGTCAAGGTTCCTATTGACTTAACAATGATTGAGTTTATGTTATGGT
20570 20580 20590 20600 20610 20620 20630 20640
HCoV-SA1 GTAAGGATGGACAGGTTCAAACCTTCTACCCTCGACTCCAGGCTTCTGCAGATTGGAAACCTGGTCATGCAATGCCATCC
20650 20660 20670 20680 20690 20700 20710 20720
HCoV-SA1 CTCTTTAAAGTTCAAAATGTAAACCTTGAACGTTGTGAGCTTGCTAATTACAAGCAATCTATTCCTATGCCTCGCGGTGT
20730 20740 20750 20760 20770 20780 20790 20800
HCoV-SA1 GCACATGAACATCGCTAAATATATGCAATTGTGCCAGTATTTAAATACTTGCACATTAGCCGTGCCTGCCAATATGCGTG

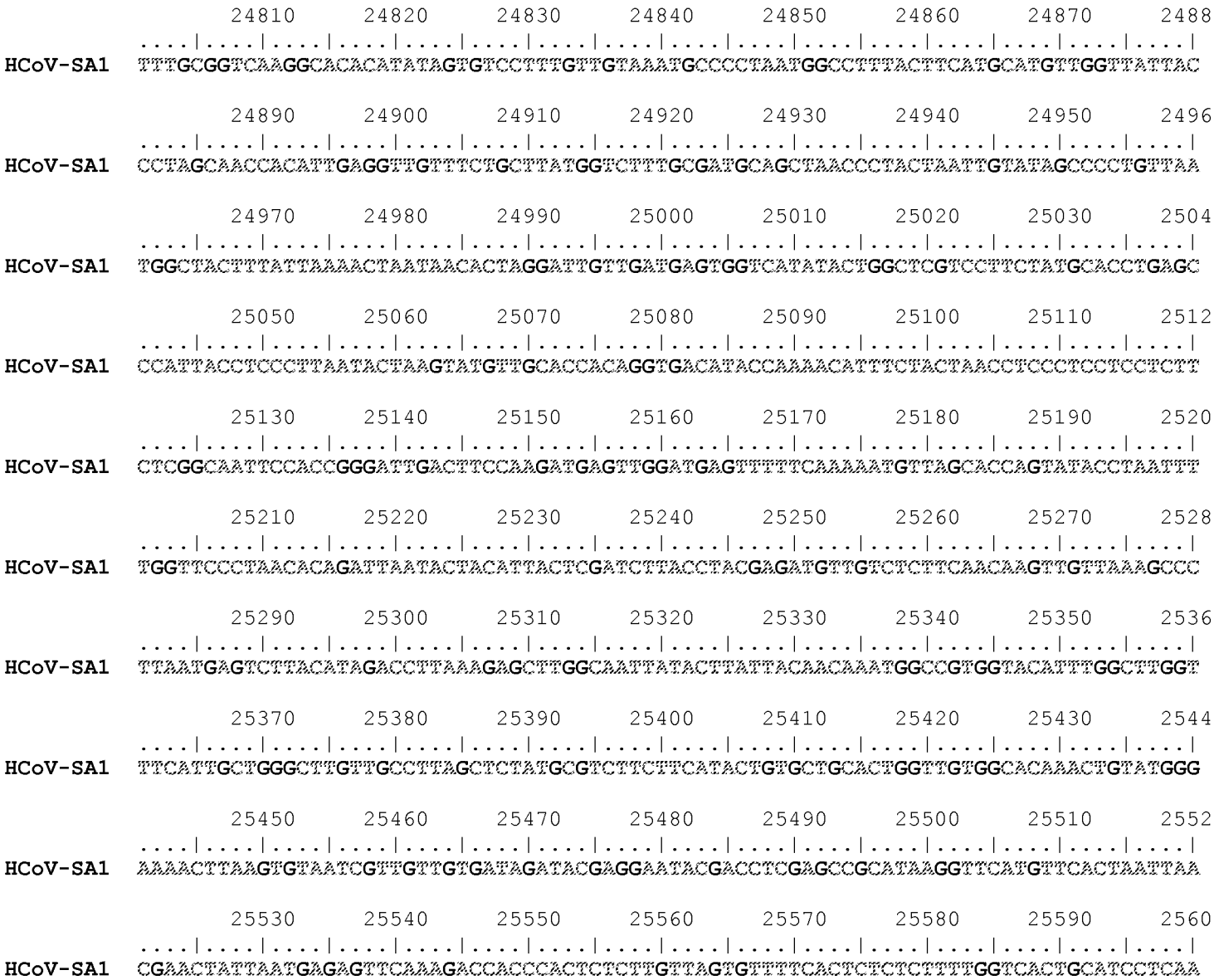

20810 20820 20830 20840 20850 20860 20870 20880
HCoV-SA1 TTATACATTTTGGCGCTGGTTCTGATAAAGGTATCGCTCCTGGTACCTCAGTTTTACGACAGTGGCTTCCTACAGATGCC
20890 20900 20910 20920 20930 20940 20950 20960
HCoV-SA1 ATTATTATAGATAATGATTTAAATGAGTTCGTGTCAGATGCTGACATAACTTTATTTGGAGATTGTGTAACTGTACGTGT
20970 20980 20990 21000 21010 21020 21030 21040
HCoV-SA1 CGGCCAACAAGTGGATCTTGTTATTTCCGACATGTATGATCCTACTACTAAGAATGTAACAGGTAGTAATGAGTCAAAGG
21050 21060 21070 21080 21090 21100 21110 21120
HCoV-SA1 CTTTATTCTTTACTTACCTGTGTAACCTCATTAATAATAATCTTGCTCTTGGTGGGTCTGTTGCTATTAAAATAACAGAA
21130 21140 21150 21160 21170 21180 21190 21200
HCoV-SA1 CACTCTTGGAGCGTTGAACTTTATGAACTTATGGGAAAATTTGCTTGGTGGACTGTTTTCTGCACCAATGCAAATGCATC
21210 21220 21230 21240 21250 21260 21270 21280
HCoV-SA1 CTCATCTGAAGGATTCCTCTTAGGTATTAATTACTTGGGTACTATTAAAGAAAATATAGATGGTGGTGCTATGCACGCCA
21290 21300 21310 21320 21330 21340 21350 21360
HCoV-SA1 ACTATATATTTTGGAGAAATTCCACTCCTATGAATCTGAGTACTTACTCACTTTTTGATTTATCCAAGTTTCAATTAAAA
21370 21380 21390 21400 21410 21420 21430 21440
HCoV-SA1 TTAAAAGGAACACCAGTTCTTCAATTAAAGGAGAGTCAAATTAACGAACTCGTAATATCTCTCCTGTCGCAGGGTAAGTT
21450 21460 21470 21480 21490 21500 21510 21520
HCoV-SA1 ACTTATCCGTGACAATGATACACTCAGTGTTTCTACTGATGTTCTTGTTAACACCTACAGAAAGTTACGTTGATGTAGGG
21530 21540 21550 21560 21570 21580 21590 21600
HCoV-SA1 CCAGATTCTGTTAAGTCTGCTTGTATTGAGGTTGATATACAACAGACTTTCTTTGATAAAACTTGGCCTAGGCCAATTGA

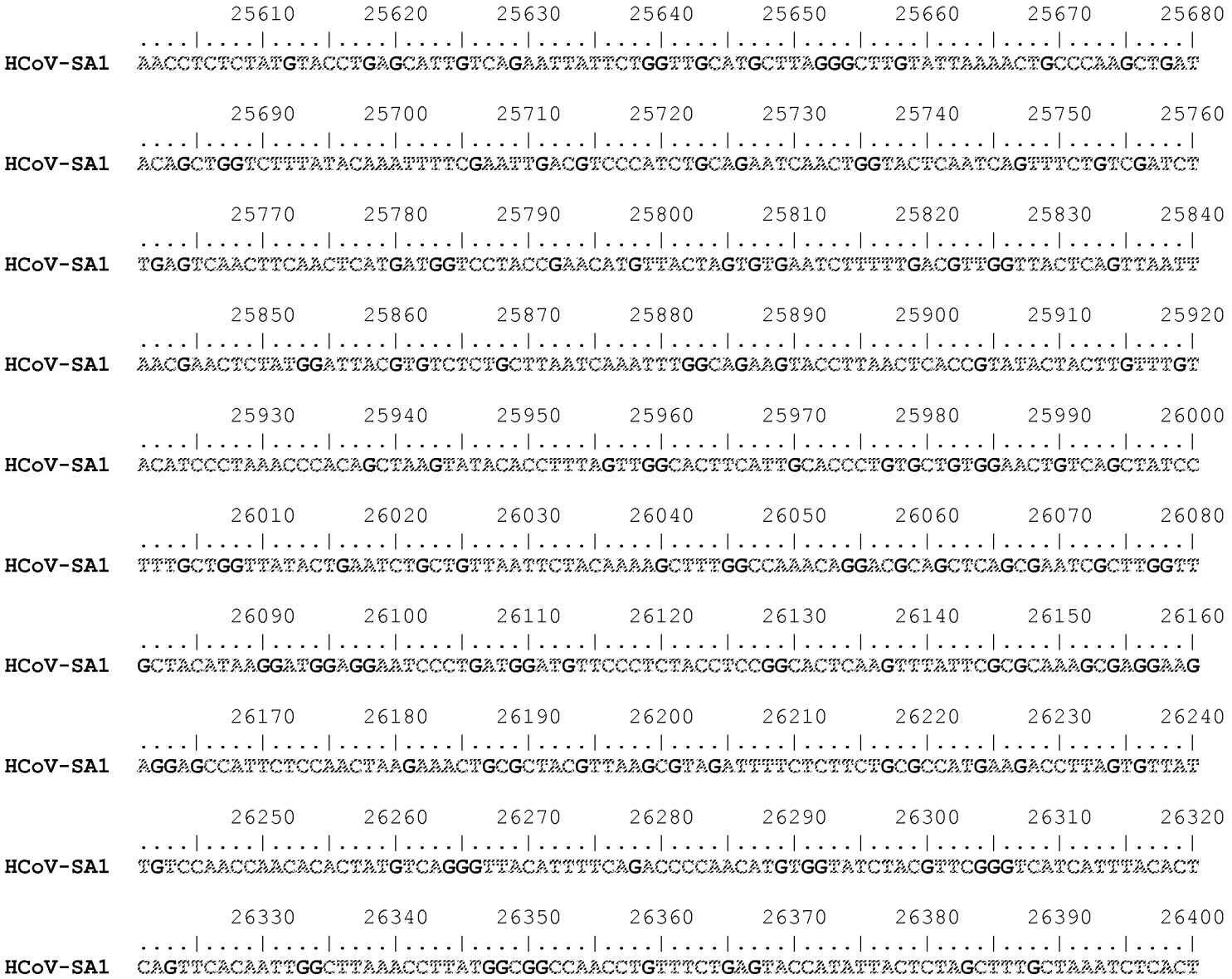

21610 21620 21630 21640 21650 21660 21670 21680
HCoV-SA1 TGTTTCTAAGGCTGACGGTATTATATACCCTCAAGGCCGTACATATTCTAACATAACTATCACTTATCAAGGTCTTTTTC
21690 21700 21710 21720 21730 21740 21750 21760
HCoV-SA1 CCTATCAGGGAGACCATGGTGATATGTATGTTTACTCTGCAGGACATGCTACAGGCACAACTCCACAAAAGTTGTTTGTA
21770 21780 21790 21800 21810 21820 21830 21840
HCoV-SA1 GCTAACTATTCTCAGGACGTCAAACAGTTTGCTAATGGGTTTGTCGTCCGTATAGGAGCAGCTGCCAATTCCACTGGCAC
21850 21860 21870 21880 21890 21900 21910 21920
HCoV-SA1 TGTTATTATTAGCCCATCTACCAGCGCTACTATACGAAAAATTTACCCTGCTTTTATGCTGGGTTCTTCAGTTGGTAATT
21930 21940 21950 21960 21970 21980 21990 22000
HCoV-SA1 TCTCAGATGGTAAAATGGGCCGCTTCTTCAATCATACTCTAGTTCTTTTGCCCGATGGATGTGGCACTTTACTTAGAGCT
22010 22020 22030 22040 22050 22060 22070 22080
HCoV-SA1 TTTTATTGTATTCTAGAGCCTCGCTCTGGAAAATCATTGTCCTGCTGGCAATTCCTATACTTCTTTTGCCACTTATCACAC
22090 22100 22110 22120 22130 22140 22150 22160
HCoV-SA1 TCCTGCAACAGATTGTTCTGATGGCAATTACAATCGTAATGCCAGTCTGAACTCTTTTAAGGAGTATTTTAATTTACGTA
22170 22180 22190 22200 22210 22220 22230 22240
HCoV-SA1 ACTGCACCTTTATGTACACTTATAACATTACCGAAGATGAGATTTTAGAGTGGTTTGGCATTACACAAACTGCTCAAGGT
22250 22260 22270 22280 22290 22300 22310 22320
HCoV-SA1 GTTCACCTCTTCTCATCTCGGTATGTTGATTTGTACGGCGGCAATATGTTTCAATTTGCCACCTTGCCTGTTTATGATAC
22330 22340 22350 22360 22370 22380 22390 22400
HCoV-SA1 TATTAAGTATTATTCTATCATTCCTCACAGTATTCGTTCTATCCAAAGTGATAGAAAAGCTTGGGCTGCCTTCTACGTAT


22410 22420 22430 22440 22450 22460 22470 22480
HCoV-SA1 ATAAACTTCAACCGTTAACTTTCCTGTTGGATTTTTCTGTTGATGGTTATATACGCAGAGCTATAGACTGTGGTTTTAAT
22490 22500 22510 22520 22530 22540 22550 22560
HCoV-SA1 GATTTGTCACAACTCCACTGCTCATATGAATCCTTCGATGTTGAATCTGGAGTTTATTCAGTTTCGTCTTTCGAAGCAAA
22570 22580 22590 22600 22610 22620 22630 22640
HCoV-SA1 ACCTTCTGGCTCAGTTGTGGAACAGGCTGAAGGTGTTGAATGTGATTTTTCACCTCTTCTGTCTGGCACACCTCCTCAGG
22650 22660 22670 22680 22690 22700 22710 22720
HCoV-SA1 TTTATAATTTCAAGCGTTTGGTTTTTACCAATTGCAATTATAATCTTACCAAATTGCTTTCACTTTTTTCTGTGAATGAT
22730 22740 22750 22760 22770 22780 22790 22800
HCoV-SA1 TTTACTTGTAGTCAAATATCTCCAGCAGCAATTGCTAGCAACTGTTATTCTTCACTGATTTTGGATTACTTTTCATACCC
22810 22820 22830 22840 22850 22860 22870 22880
HCoV-SA1 ACTTAGTATGAAATCCGATCTCAGTGTTAGTTCTGCTGGTCCAATATCCCAGTTTAATTATAAACAGTCCTTTTCTAATC
22890 22900 22910 22920 22930 22940 22950 22960
HCoV-SA1 CCACATGTTTGATTTTAGCGACTGTTCCTCATAACCTTACTACTATTACTAAGCCTCTTAAGTACAGCTATATTAACAAG
22970 22980 22990 23000 23010 23020 23030 23040
HCoV-SA1 TGCTCTCGTCTTCTTTCTGATGATCGTACTGAAGTACCTCAGTTAGTGAACGCTAATCAATACTCACCCTGTGTATCCAT
23050 23060 23070 23080 23090 23100 23110 23120
HCoV-SA1 TGTCCCATCCACTGTGTGGGAAGACGGTGATTATTATAGGAAACAACTATCTCCACTTGAAGGTGGTGGCTGGCTTGTTG
23130 23140 23150 23160 23170 23180 23190 23200
HCoV-SA1 CTAGTGGCTCAACTGTTGCCATGACTGAGCAATTACAGATGGGCTTTGGTATTACAGTTCAATATGGTACAGACACCAAT


23210 23220 23230 23240 23250 23260 23270 23280
HCoV-SA1 AGTGTTTGCCCCAAGCTTGAATTTGCTAATGACACAAAAATTGCCTCTCAATTAGGCAATTGCGTGGAATATTCCCTCTA
23290 23300 23310 23320 23330 23340 23350 23360
HCoV-SA1 TGGTGTTTCGGGCCGTGGTGTTTTTCAGAATTGCACAGCTGTAGGTGTTCGACAGCAGCGCTTTGTTTATGATGCGTACC
23370 23380 23390 23400 23410 23420 23430 23440
HCoV-SA1 AGAATTTAGTTGGCTATTATTCTGATGATGGCAACTACTACTGTTTGCGTGCTTGTGTTAGTGTTCCTGTTTCTGTCATC
23450 23460 23470 23480 23490 23500 23510 23520
HCoV-SA1 TATGATAAAGAAACTAAAACCCACGCTACTCTATTTGGTAGTGTTGCATGTGAACACATTTCTTCTACCATGTCTCAATA
23530 23540 23550 23560 23570 23580 23590 23600
HCoV-SA1 CTCCCGTTCTACGCGATCAATGCTTAAACGGCGAGATTCTACATATGGCCCCCTTCAGACACCTGTTGGTTGTGTCCTAG
23610 23620 23630 23640 23650 23660 23670 23680
HCoV-SA1 GACTTGTTAATTCCTCTTTGTTCGTAGAGGACTGCAAGTTGCCTCTTGGTCAATCTCTCTGTGCTCTTCCTGACACACCT
23690 23700 23710 23720 23730 23740 23750 23760
HCoV-SA1 AGTACTCTCACACCTCGCAGTGTGCGCTCTGTTCCAGGTGAAATGCGCTTGGCATCCATTGCTTTTAATCATCCTATTCA
23770 23780 23790 23800 23810 23820 23830 23840
HCoV-SA1 GGTTGATCAACTTAATAGTAGTTATTTTAAATTAAGTATACCCACTAATTTTTCCTTTGGTGTGACTCAGGAGTACATTC
23850 23860 23870 23880 23890 23900 23910 23920
HCoV-SA1 AGACAACCATTCAGAAAGTTACTGTTGATTGTAAACAGTACGTTTGCAATGGTTTCCAGAAGTGTGAGCAATTACTGCGC
23930 23940 23950 23960 23970 23980 23990 24000
HCoV-SA1 GAGTATGGCCAGTTTTGTTCCAAAATAAACCAGGCTCTCCATGGTGCCAATTTACGCCAGGATGATTCTGTACGTAATTT

```
          24010     24020     24030     24040     24050     24060     24070     24080
          ....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|
HCoV-SA1  GTTTGCGAGCGTGAAAAGCTCTCAATCATCCTATACCAGGTTTTGGAGGTGACTTTAATTTGACACTTCTAGAAC

          24090     24100     24110     24120     24130     24140     24150     24160
          ....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|
HCoV-SA1  CTGTTTCTATATCTACTGGCAGTCGTAGTGCACGTAGTGCTATTTGAGGATTTGCTATTTGACAAAGTCACTATAGCTTGAT

          24170     24180     24190     24200     24210     24220     24230     24240
          ....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|
HCoV-SA1  CCTGGTTTATATGCAAGGTTACGATGATTGCAGCAAGGTCCAGCATCAGCTCGTGATCTTATTTGTGCTCAATATGT

          24250     24260     24270     24280     24290     24300     24310     24320
          ....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|
HCoV-SA1  GGCTGGTTACAAAGTATTACCTCCTCTTTATGGATGTTAATATGGAAGCCGCGTATACTTCATCTTTGCTTGGCAGCATAG

          24330     24340     24350     24360     24370     24380     24390     24400
          ....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|
HCoV-SA1  CAGGTGTTGGCTGGACTGCTGGCTTATCCTCCTTTGCTGCTATTCCATTTGCACAGAGTATCTTTTATAGGTTAAACGGT

          24410     24420     24430     24440     24450     24460     24470     24480
          ....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|
HCoV-SA1  GTTGGCATTACTCAACAGGTTCTTTCAGAGAACCAAAAGCTTATTGCCAATAAGTTTTAATCAGGCTCTGGGAGCTATGCA

          24490     24500     24510     24520     24530     24540     24550     24560
          ....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|
HCoV-SA1  AACAGGCTTCACTACAACTAATGAAGCTTTTCAGAAGGTTCAGGATGCTGTGAACAACAATGCACAGGCTCTATCCAAAT

          24570     24580     24590     24600     24610     24620     24630     24640
          ....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|
HCoV-SA1  TAGCTAGCGAGCTATCTAATACTTTTTGGTGCTATTTCCGCCTCTATTGGAGACATCATACAACGTCTTGATGTTCTCGAA

          24650     24660     24670     24680     24690     24700     24710     24720
          ....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|
HCoV-SA1  CAGGACGCCCAAATAGACAGACTTATTAATGGCCGTTTGACAACACTAAATGCTTTTGTTGCACAGCAGCTTGTTCGTTC

          24730     24740     24750     24760     24770     24780     24790     24800
          ....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|
HCoV-SA1  CGAATCAGCTGCTCTTTTCCGTCAATTGGCTAAAGATAAAGTCAATGAGTGTGTCAAGGCACAATCCAAGCGTTCTGGAT
```

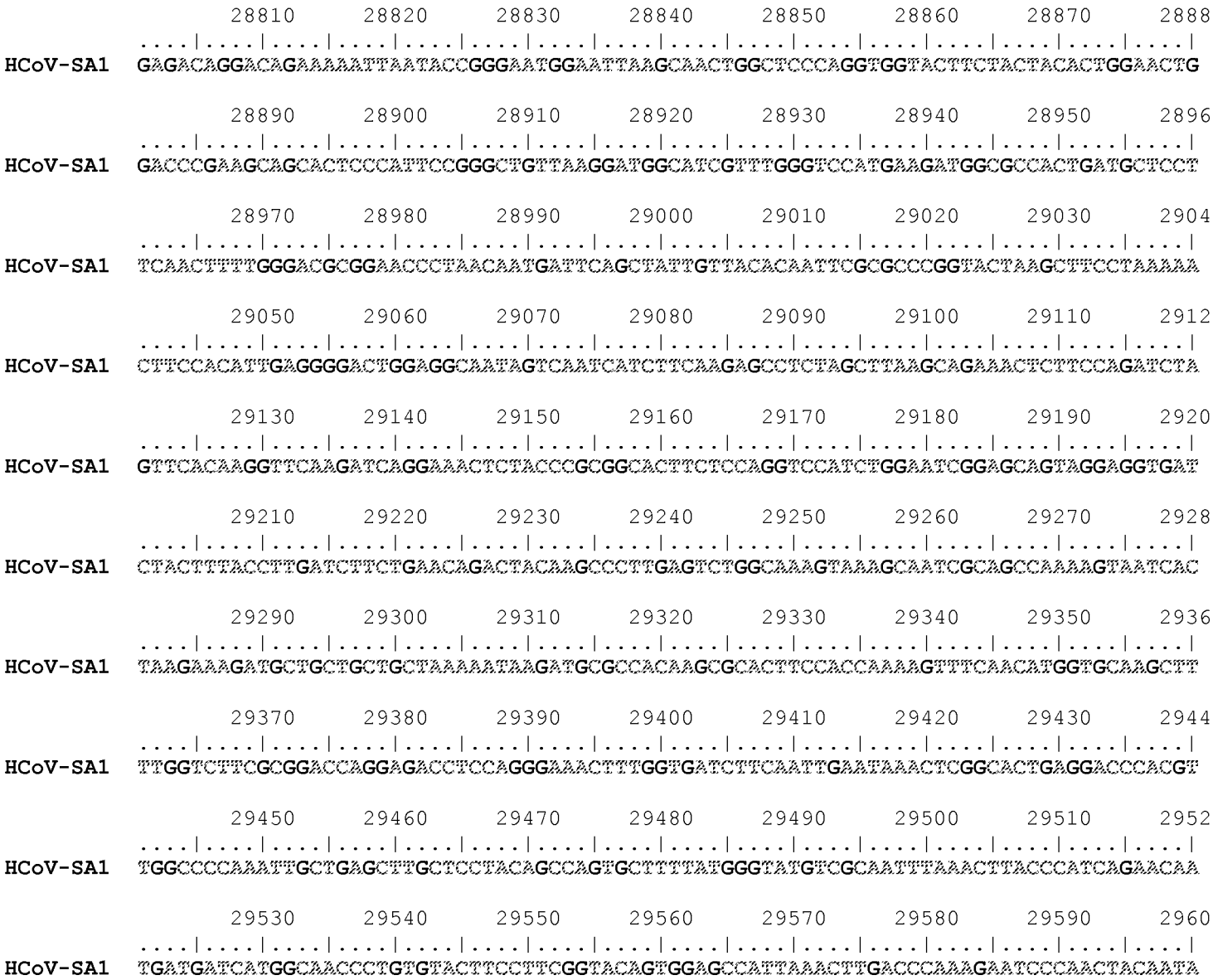

24810 24820 24830 24840 24850 24860 24870 24880
HCoV-SA1 TTTGCGGTCAAGGCACACATATAGTGTCCTTTGTTGTAAATGCCCCTAATGGCCTTTACTTCATGCATGTTGGTTATTAC
24890 24900 24910 24920 24930 24940 24950 24960
HCoV-SA1 CCTAGCAACCACATTGAGGTTGTTTCTGCTTATGGTCTTTGCGATGCAGCTAACCCTACTAATTGTATAGCCCCTGTTAA
24970 24980 24990 25000 25010 25020 25030 25040
HCoV-SA1 TGGCTACTTTATTAAAACTAATAACACTAGGATTGTTGATGAGTGGTCATATACTGGCTCGTCCTTCTATGCACCTGAGC
25050 25060 25070 25080 25090 25100 25110 25120
HCoV-SA1 CCATTACCTCCCTTAATACTAAGTATGTTGCACCACAGGTGACATACCAAAACATTTCTACTAACCTCCCTCCTCCTCTT
25130 25140 25150 25160 25170 25180 25190 25200
HCoV-SA1 CTCGGCAATTCCACCGGGATTGACTTCCAAGATGAGTTGGATGAGTTTTTCAAAAATGTTAGCACCAGTATACCTAATTT
25210 25220 25230 25240 25250 25260 25270 25280
HCoV-SA1 TGGTTCCCTAACACAGATTAATACTACATTACTCGATCTTACCTACGAGATGTTGTCTCTTCAACAAGTTGTTAAAGCCC
25290 25300 25310 25320 25330 25340 25350 25360
HCoV-SA1 TTAATGAGTCTTACATAGACCTTAAAGAGCTTGGCAATTATACTTATTACAACAAATGGCCGTGGTACATTTGGCTTGGT
25370 25380 25390 25400 25410 25420 25430 25440
HCoV-SA1 TTCATTGCTGGGCTTGTTGCCTTAGCTCTATGCGTCTTCTTCATACTGTGCTGCACTGGTTGTGGCACAAACTGTATGGG
25450 25460 25470 25480 25490 25500 25510 25520
HCoV-SA1 AAAACTTAAGTGTAATCGTTGTTGTGATAGATACGAGGAATACGACCTCGAGCCGCATAAGGTTCATGTTCACTAATTAA
25530 25540 25550 25560 25570 25580 25590 25600
HCoV-SA1 CGAACTATTAATGAGAGTTCAAAGACCACCCACTCTCTTGTTAGTGTTTTCACTCTCTCTTTTGGTCACTGCATCCTCAA


25610 25620 25630 25640 25650 25660 25670 25680
HCoV-SA1 AACCTCTCTATGTACCTGAGCATTGTCAGAATTATTCTGGTTGCATGCTTAGGGCTTGTATTAAAACTGCCCAAGCTGAT
25690 25700 25710 25720 25730 25740 25750 25760
HCoV-SA1 ACAGCTGGTCTTTATACAAATTTTCGAATTGACGTCCCATCTGCAGAATCAACTGGTACTCAATCAGTTTCTGTCGATCT
25770 25780 25790 25800 25810 25820 25830 25840
HCoV-SA1 TGAGTCAACTTCAACTCATGATGGTCCTACCGAACATGTTACTAGTGTGAATCTTTTTTGACGTTGGTTACTCAGTTAATT
25850 25860 25870 25880 25890 25900 25910 25920
HCoV-SA1 AACGAACTCTATGGATTACGTGTCTCTGCTTAATCAAATTTGGCAGAAGTACCTTAACTCACCGTATACTACTTGTTTGT
25930 25940 25950 25960 25970 25980 25990 26000
HCoV-SA1 ACATCCCTAAACCCACAGCTAAGTATACACCTTTAGTTGGCACTTCATTGCACCCTGTGCTGTGGAACTGTCAGCTATCC
26010 26020 26030 26040 26050 26060 26070 26080
HCoV-SA1 TTTGCTGGTTATACTGAATCTGCTGTTAATTCTACAAAAGCTTTGGCCAAACAGGACGCAGCTCAGCGAATCGCTTGGTT
26090 26100 26110 26120 26130 26140 26150 26160
HCoV-SA1 GCTACATAAGGATGGAGGAATCCCTGATGGATGTTCCCTCTACCTCCGGCACTCAAGTTTATTCGCGCAAAGCGAGGAAG
26170 26180 26190 26200 26210 26220 26230 26240
HCoV-SA1 AGGAGCCATTCTCCAACTAAGAAACTGCGCTACGTTAAGCGTAGATTTTCTCTTCTGCGCCATGAAGACCTTAGTGTTAT
26250 26260 26270 26280 26290 26300 26310 26320
HCoV-SA1 TGTCCAACCAACACACTATGTCAGGGTTACATTTTCAGACCCCAACATGTGGTATCTACGTTCGGGTCATCATTTACACT
26330 26340 26350 26360 26370 26380 26390 26400
HCoV-SA1 CAGTTCACAATTGGCTTAAACCTTATGGCGGCCAACCTGTTTCTGAGTACCATATTACTCTAGCTTTGCTAAATCTCACT

```
                 26410     26420     26430     26440     26450     26460     26470     26480
           ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
HCoV-SA1   GATGAAGATTTAGCTAGAGATTTTTCACCCATTGCGCTCTTTTTGCGCAATGTCAGATTTGAGCTACATGAGTTCGCCTT

                 26490     26500     26510     26520     26530     26540     26550     26560
           ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
HCoV-SA1   GCTGCGCAAAACTCTTGTTCTTAATGCATCAGAGATCTACTGTGCTAACATACATAGATTTAAGCCTGTGTATAGAGTTA

                 26570     26580     26590     26600     26610     26620     26630     26640
           ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
HCoV-SA1   ACACGGCAATCCCTACTATTAAGGATTGGCTTCTCGTTCAGGGATTTTCCCTTTACCATAGTGGCCTCCCTTTACATATG

                 26650     26660     26670     26680     26690     26700     26710     26720
           ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
HCoV-SA1   TCAATCTCTAAATTGCATGCACTGGATGATGTTACTCGCAATTACATCATTACAATGCCATGCTTTAGAACTTACCCTCA

                 26730     26740     26750     26760     26770     26780     26790     26800
           ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
HCoV-SA1   ACAAATGTTTGTTACTCCTTTGGCCGTAGATGTTGTCTCCATACGGTCTTCCAATCAGGGTAATAAACAAATTGTTCATT

                 26810     26820     26830     26840     26850     26860     26870     26880
           ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
HCoV-SA1   CTTATCCCATTTTACATCATCCAGGATTTTAACGAACTATGGCTTTCTCGGCGTCTTTATTTAAACCCGTCCAGCTAGTC

                 26890     26900     26910     26920     26930     26940     26950     26960
           ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
HCoV-SA1   CCAGTTTCTCCTGCATTTCATCGCATTGAGTCTACTGACTCTATTGTTTTCACATACATTCCTGCTAGCGGCTATGTAGC

                 26970     26980     26990     27000     27010     27020     27030     27040
           ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
HCoV-SA1   TGCTTTAGCTGTCAATGTGTGTCTCATTCCCCTATTATTACTGCTACGTCAAGATACTTGTCGTCGCAGCATTATCAGAA

                 27050     27060     27070     27080     27090     27100     27110     27120
           ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
HCoV-SA1   CTATGGTTCTCTATTTCCTTGTTCTGTATAACTTTTTATTAGCCATTGTACTAGTCAATGGTGTACATTATCCAACTGGA

                 27130     27140     27150     27160     27170     27180     27190     27200
           ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
HCoV-SA1   AGTTGCCTGATAGCCTTCTTAGTTATCCTCATAATACTTTAGTTTGTAGATAGAATTCGTTTCTGTCTCATGCTGAATTC
```

```
        27210     27220     27230     27240     27250     27260     27270     27280
....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|
HCoV-SA1 CTACATTCCACTGTTTGACATGCGTTCCCACTTTATTCGTGTTAGTACAGTTCTTCTCATGGTATGGTCCCTGTAATAC

        27290     27300     27310     27320     27330     27340     27350     27360
....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|
HCoV-SA1 ACACCAAACCATTATTTATTTAGAAACTTCGATCAGCGTTGCAGCTGTTCTCGTTGTTTTTATTTGCACTCTTCCACTTAT

        27370     27380     27390     27400     27410     27420     27430     27440
....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|
HCoV-SA1 ATAGAGTGCACTTATTATTAGCCGTTTAGTAAGATTAGCCTAGTTTCTGTAACTGACTTCTCCTTTAAACGGCAATGTTTC

        27450     27460     27470     27480     27490     27500     27510     27520
....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|
HCoV-SA1 CACTGTTTTCGTGCCTGCAACGCGCGATTTCAGTTCCTCTTCACATAATCGCCCCGAGCTCGCTTATCGTTTAAGCAGCTC

        27530     27540     27550     27560     27570     27580     27590     27600
....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|
HCoV-SA1 TGCGCTACTATGGGTCCCGTGTAGAGGCTAATCCATTAGTCTCTCTTTGGACATATGGAAAACGAACTATGTTACCCTTT

        27610     27620     27630     27640     27650     27660     27670     27680
....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|
HCoV-SA1 GTCCAAGAACGAATAGGGTTGTTCATAGTAAACTTTTTCATTTTTACCGTAGTATGTGCTATAACACTCTTGGTGTGTAT

        27690     27700     27710     27720     27730     27740     27750     27760
....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|
HCoV-SA1 GGCTTTCCTTACGGCTACTAGATTATGTGTGCAATGTATGACAGGCTTCAATACCCTGTTAGTTCAGCCCGCATTAAACT

        27770     27780     27790     27800     27810     27820     27830     27840
....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|
HCoV-SA1 TGTATAATACTGGACGTTCAGTCTATGTAAAATTCCAGGATAGTAAACCCCCTCTACCACCTGACGAGTGGGTTTAACGA

        27850     27860     27870     27880     27890     27900     27910     27920
....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|
HCoV-SA1 ACTCCTTCATAATGTCTAATATGACGCAACTCACTGAGGCGCAGATTATTGCCATTATTAAAGACTGGAACTTTGCATGG

        27930     27940     27950     27960     27970     27980     27990     28000
....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|
HCoV-SA1 TCCCTGATCTTTCTCTTAATTACTATCGTACTACAGTATGGATACCCATCCCGTAGTATGACTGTCTATGTCTTTTAAAAT
```

```
        28010     28020     28030     28040     28050     28060     28070     28080
....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|
HCoV-SA1 GTTTGTTTTTATGGCTCCTATGGCCATCTTCCATGGCGCTATCCAATATTTAGCGCCGTTTATCCAATTGATCTAGCTTCCC

        28090     28100     28110     28120     28130     28140     28150     28160
....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|
HCoV-SA1 AGATAAATCTCTGGCATTGTAGCAGCTGTTTTCAGCTATGATGTGGATTTCCTACTTTGTGCAGAGTATCCGGCTGTTTTATG

        28170     28180     28190     28200     28210     28220     28230     28240
....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|
HCoV-SA1 AGAACTGGATCATGGTGGTCAATTCCAATCCCTGAGACTAATTGCCTTTTGAACGTTCCATTTGGTGGTACAACTGTCGTTACG

        28250     28260     28270     28280     28290     28300     28310     28320
....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|
HCoV-SA1 TCCACTCGTAGAGGACTCTACCAGTGTAACTGCTGTTGTAACCAATGGCCACCTCAAAATGGCTGGCATGCATTTCGGTG

        28330     28340     28350     28360     28370     28380     28390     28400
....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|
HCoV-SA1 CTTGTGACTACGACAGACTTCCTAATGAAGTCACCGTGGCCAAACCCAATGTGCTGATTGCTTTAAAAATGGTGAAGCGG

        28410     28420     28430     28440     28450     28460     28470     28480
....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|
HCoV-SA1 CAAAGCTACGGAACTAATTCCGGCGTTGCCATTACCATAGATATAAGGCAGGTAATTACAGGAGTCCGCCCTATTACGGC

        28490     28500     28510     28520     28530     28540     28550     28560
....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|
HCoV-SA1 GGAAATTGAACTTGCATTGCTTCGAGCTTTAGGCTCTTTAGTAAGAGTATCTTTAATTGATTTTAACGAAATCTCAATTTCAT

        28570     28580     28590     28600     28610     28620     28630     28640
....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|
HCoV-SA1 TGTTTATGGCATCCCCCTGCTGCACCTCGTGCTGTTTCCTTTGCCGATAACAATGATAATAACAAATACAAACCTATCTCGAG

        28650     28660     28670     28680     28690     28700     28710     28720
....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|
HCoV-SA1 GTAGAGGACGTAATCCAAAACCACACGAGCTTGCACCCAAATAACACTGTCTCTTGGTACACTGGGCTTACCCAACACGGGAAA

        28730     28740     28750     28760     28770     28780     28790     28800
....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|
HCoV-SA1 GTCCCTCTTACCTTTCCACCTGGGCAGGGTGTACCTCTTAATGCCAATTCTACCCCTGCGCAAATGCTGGGTATTGGCG
```


28810 28820 28830 28840 28850 28860 28870 28880
HCoV-SA1 GAGACAGGACAGAAAAATTAATACCGGGAATGGAATTAAGCAACTGGCTCCCAGGTGGTACTTCTACTACACTGGAACTG
28890 28900 28910 28920 28930 28940 28950 28960
HCoV-SA1 GACCCGAAGCAGCACTCCCATTCCGGGCTGTTAAGGATGGCATCGTTTGGGTCCATGAAGATGGCGCCACTGATGCTCCT
28970 28980 28990 29000 29010 29020 29030 29040
HCoV-SA1 TCAACTTTTGGGACGCGGAACCCTAACAATGATTCAGCTATTGTTACACAATTCGCGCCCGGTACTAAGCTTCCTAAAAA
29050 29060 29070 29080 29090 29100 29110 29120
HCoV-SA1 CTTCCACATTGAGGGGACTGGAGGCAATAGTCAATCATCTTCAAGAGCCTCTAGCTTAAGCAGAAACTCTTCCAGATCTA
29130 29140 29150 29160 29170 29180 29190 29200
HCoV-SA1 GTTCACAAGGTTCAAGATCAGGAAACTCTACCCGCGGCACTTCTCCAGGTCCATCTGGAATCGGAGCAGTAGGAGGTGAT
29210 29220 29230 29240 29250 29260 29270 29280
HCoV-SA1 CTACTTTACCTTGATCTTCTGAACAGACTACAAGCCCTTGAGTCTGGCAAAGTAAAGCAATCGCAGCCAAAAGTAATCAC
29290 29300 29310 29320 29330 29340 29350 29360
HCoV-SA1 TAAGAAAGATGCTGCTGCTGCTAAAAATAAGATGCGCCACAAGCGCACTTCCACCAAAAGTTTCAACATGGTGCAAGCTT
29370 29380 29390 29400 29410 29420 29430 29440
HCoV-SA1 TTGGTCTTCGCGGACCAGGAGACCTCCAGGGAAACTTTGGTGATCTTCAATTGAATAAACTCGGCACTGAGGACCCACGT
29450 29460 29470 29480 29490 29500 29510 29520
HCoV-SA1 TGGCCCCAAATTGCTGAGCTTGCTCCTACAGCCAGTGCTTTTATGGGTATGTCGCAATTTAAACTTACCCATCAGAACAA
29530 29540 29550 29560 29570 29580 29590 29600
HCoV-SA1 TGATGATCATGGCAACCCTGTGTACTTCCTTCGGTACAGTGGAGCCATTAAACTTGACCCAAAGAATCCCAACTACAATA

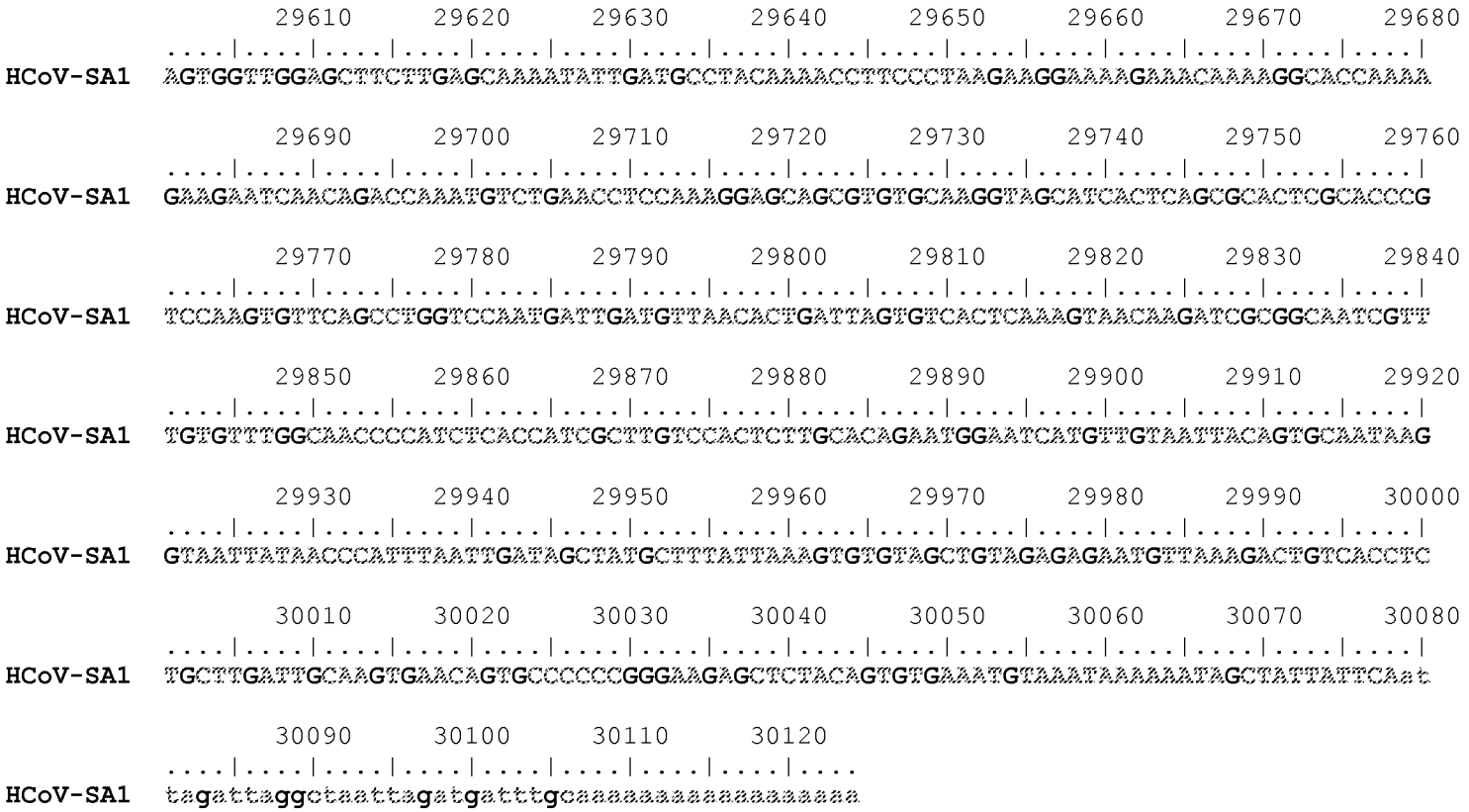
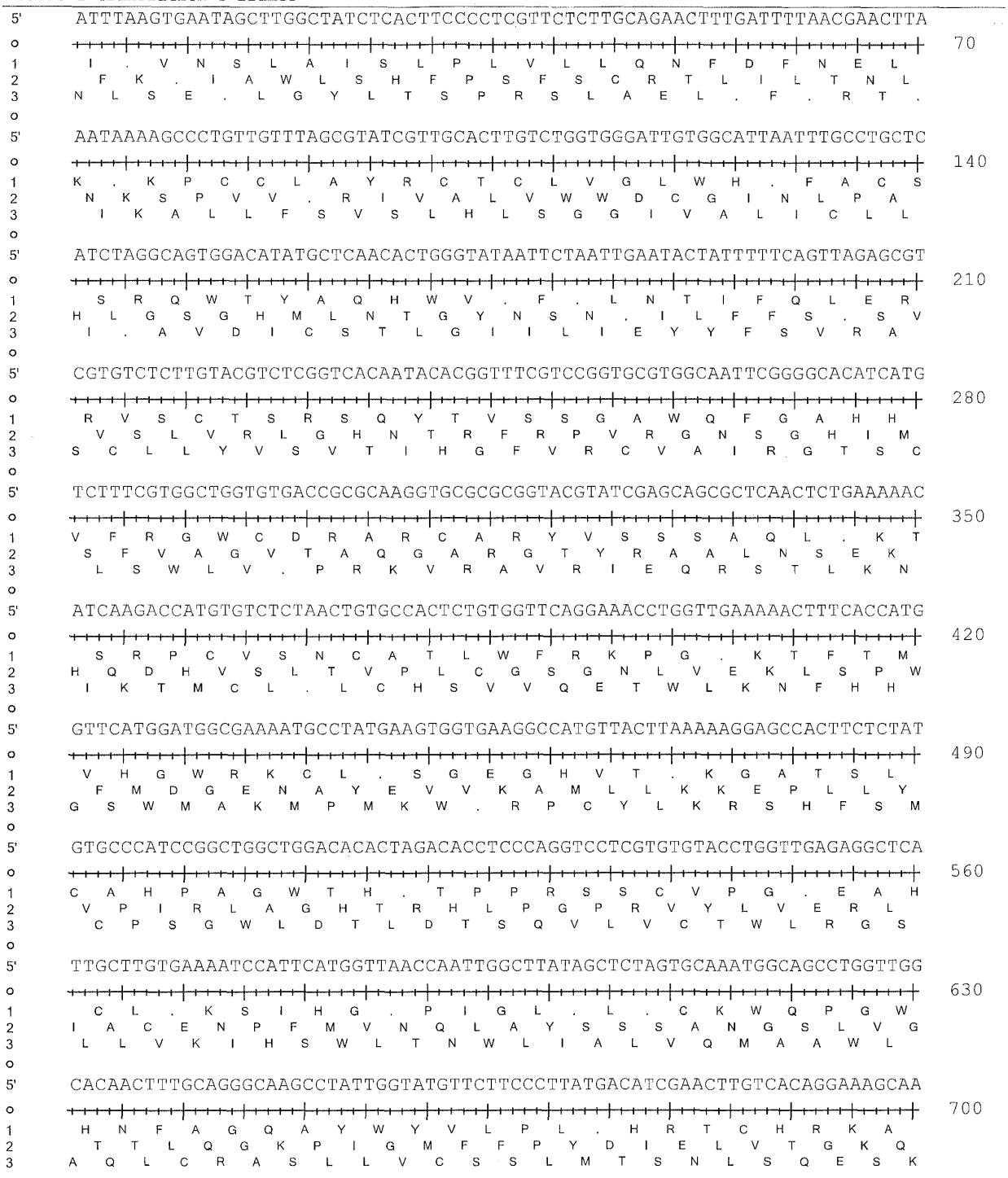


Figure 14 file HCoV-SA1.rtf

Figure 15

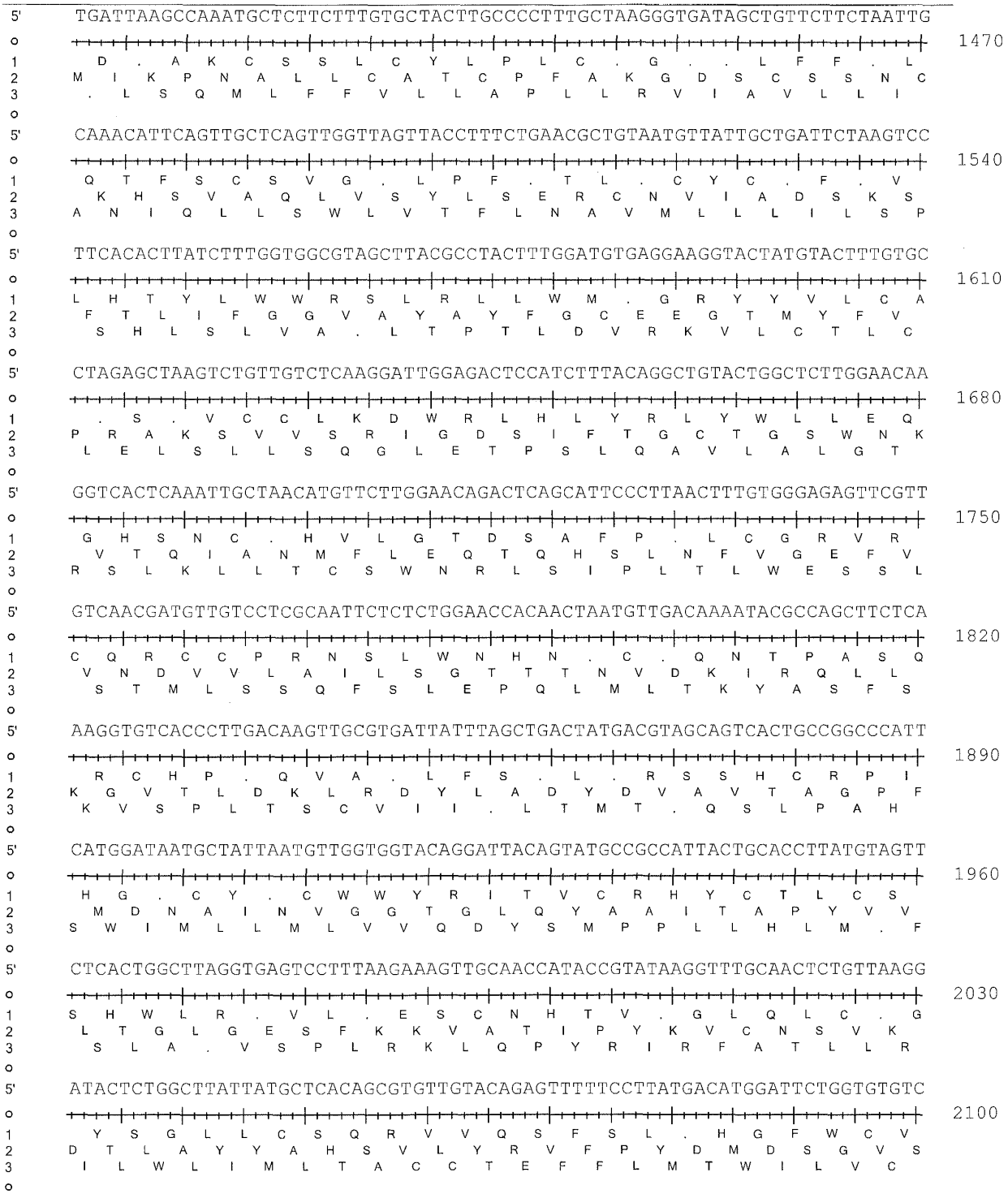
Sunday, September 23, 2012 1:41 PM
HCoV-SA1 translation 3 frames



Sunday, September 23, 2012 1:41 PM
HCoV-SA1 translation 3 frames

```
5'    AATATTCTCCTGCGCAAGTATGGCCGTGGTGGTTATCACTACACCCCATTCCACTATGAGCGAGACAACA
o     +---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+ 770
1     K   Y   S   P   A   Q   V   W   P   W   W   L   S   L   H   P   I   P   L   .   A   R   Q   H
2       N   I   L   L   R   K   Y   G   R   G   G   Y   H   Y   T   P   F   H   Y   E   R   D   N
3         I   F   S   C   A   S   M   A   V   V   V   I   T   T   P   H   S   T   M   S   E   T   T
o
5'    CCTCTTGCCCTGAGTGGATGGACGATTTTGAGGCGGATCCTAAAGGCAAATATGCCCAGAATCTGCTTAA
o     +---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+ 840
1       L   L   P   .   V   D   G   R   F   .   G   G   S   .   R   Q   I   C   P   E   S   A   .
2     T   S   C   P   E   W   M   D   D   F   E   A   D   P   K   G   K   Y   A   Q   N   L   L   K
3       P   L   A   L   S   G   W   T   I   L   R   R   I   L   K   A   N   M   P   R   I   C   L
o
5'    GAAGTTGATTGGCGGTGATGTCACTCCAGTTGACCAATACATGTGTGGCGTTGATGGAAAACCCATTAGT
o     +---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+ 910
1       E   V   D   W   R   .   C   H   S   S   .   P   I   H   V   W   R   .   W   K   T   H   .
2         K   L   I   G   G   D   V   T   P   V   D   Q   Y   M   C   G   V   D   G   K   P   I   S
3       R   S   .   L   A   V   M   S   L   Q   L   T   N   T   C   V   A   L   M   E   N   P   L   V
o
5'    GCCTACGCATTTTTAATGGCCAAGGATGGAATAACCAAACTGGCTGATGTTGAAGCGGACGTCGCAGCAC
o     +---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+ 980
1     C   L   R   I   F   N   G   Q   G   W   N   N   Q   T   G   .   C   .   S   G   R   R   S   T
2       A   Y   A   F   L   M   A   K   D   G   I   T   K   L   A   D   V   E   A   D   V   A   A
3         P   T   H   F   .   W   P   R   M   E   .   P   N   W   L   M   L   K   R   T   S   Q   H
o
5'    GTGCTGATGACGAAGGCTTCATCACATTAAAGAACAATCTATATAGATTGGTTTGGCATGTTGAGCGTAA
o     +---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+ 1050
1       C   .   .   R   R   L   H   H   I   K   E   Q   S   I   .   I   G   L   A   C   .   A   .
2     R   A   D   D   E   G   F   I   T   L   K   N   N   L   Y   R   L   V   W   H   V   E   R   K
3       V   L   M   T   K   A   S   S   H   .   R   T   I   Y   I   D   W   F   G   M   L   S   V
o
5'    AGACGTTCCATATCCTAAGCAATCTATTTTTACTATTAATAGTGTGGTCCAAAAGGATGGTGTTGAAAAC
o     +---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+ 1120
1       R   R   S   I   S   .   A   I   Y   F   Y   Y   .   .   C   G   P   K   G   W   C   .   K
2         D   V   P   Y   Y   P   K   Q   S   I   F   T   I   N   S   V   V   Q   K   D   G   V   E   N
3       K   T   F   H   I   L   S   N   L   F   L   L   L   I   V   W   S   K   R   M   V   L   K   T
o
5'    ACTCCTCCTCACTATTTTACTCTTGGATGCAAAATTTTAACGCTCACCCCACGCAACAAGTGGAGTGGCG
o     +---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+ 1190
1     H   S   S   S   L   F   Y   S   W   M   Q   N   F   N   A   H   P   T   Q   Q   V   E   W   R
2       T   P   P   H   Y   F   T   L   G   C   K   I   L   T   L   T   P   R   N   K   W   S   G
3         L   L   L   T   I   L   L   L   D   A   K   F   .   R   S   P   H   A   T   S   G   V   A
o
5'    TTTCTGACTTGTCCCTCAAACAAAAACTCCTTTACACCTTCTATGGTAAGGAGTCACTTGAGAACCCAAC
o     +---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+ 1260
1       F   .   L   V   P   Q   T   K   T   P   L   H   L   L   W   .   G   V   T   .   E   P   N
2     V   S   D   L   S   L   K   Q   K   L   L   Y   T   F   Y   G   K   E   S   L   E   N   P   T
3       F   L   T   C   P   S   N   K   N   S   F   T   P   S   M   V   R   S   H   L   R   T   Q
o
5'    CTACATTTACCACTCCGCATTCATTGAGTGTGGAAGTTGTGGTAATGATTCCTGGCTTACAGGGAATGCT
o     +---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+ 1330
1       L   H   L   P   L   R   I   H   .   V   W   K   L   W   .   .   F   L   A   Y   R   E   C
2         Y   I   Y   H   S   A   F   I   E   C   G   S   C   G   N   D   S   W   L   T   G   N   A
3     P   T   F   T   T   P   H   S   L   S   V   E   V   V   V   M   I   P   G   L   Q   G   M   L
o
5'    ATCCAAGGGTTTGCCTGTGGATGTGGGGCATCATATACAGCTAATGATGTCGAAGTCCAATCATCTGGCA
o     +---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+ 1400
1     Y   P   R   V   C   L   W   M   W   G   I   I   Y   S   .   .   C   R   S   P   I   I   W   H
2       I   Q   G   F   A   C   G   C   G   A   S   Y   T   A   N   D   V   E   V   Q   S   S   G
3         S   K   G   L   P   V   D   V   G   H   H   I   Q   L   M   M   S   K   S   N   H   L   A
o
```

Sunday, September 23, 2012 1:41 PM
HCoV-SA1 translation 3 frames



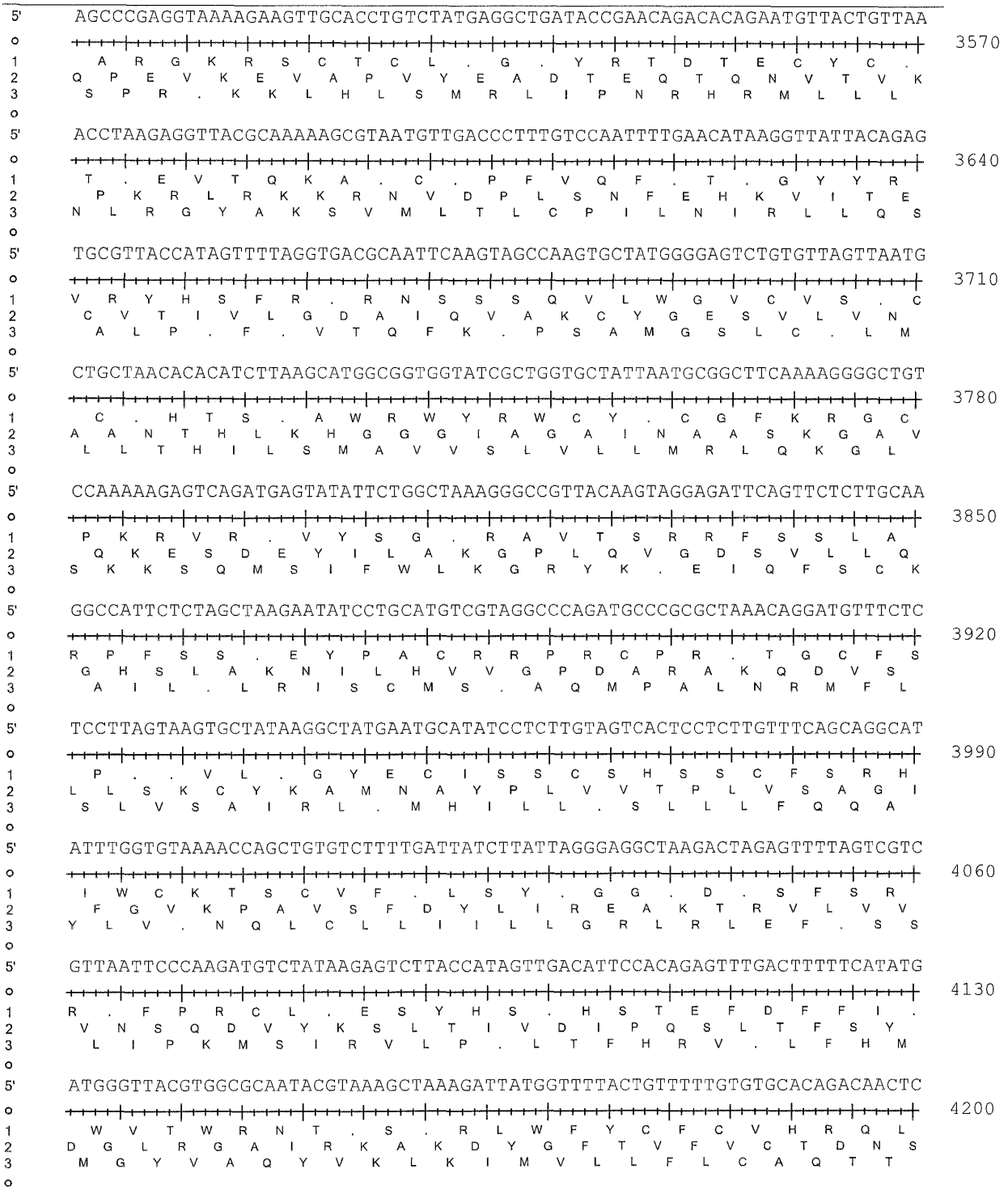
Sunday, September 23, 2012 1:41 PM
HCoV-SA1 translation 3 frames

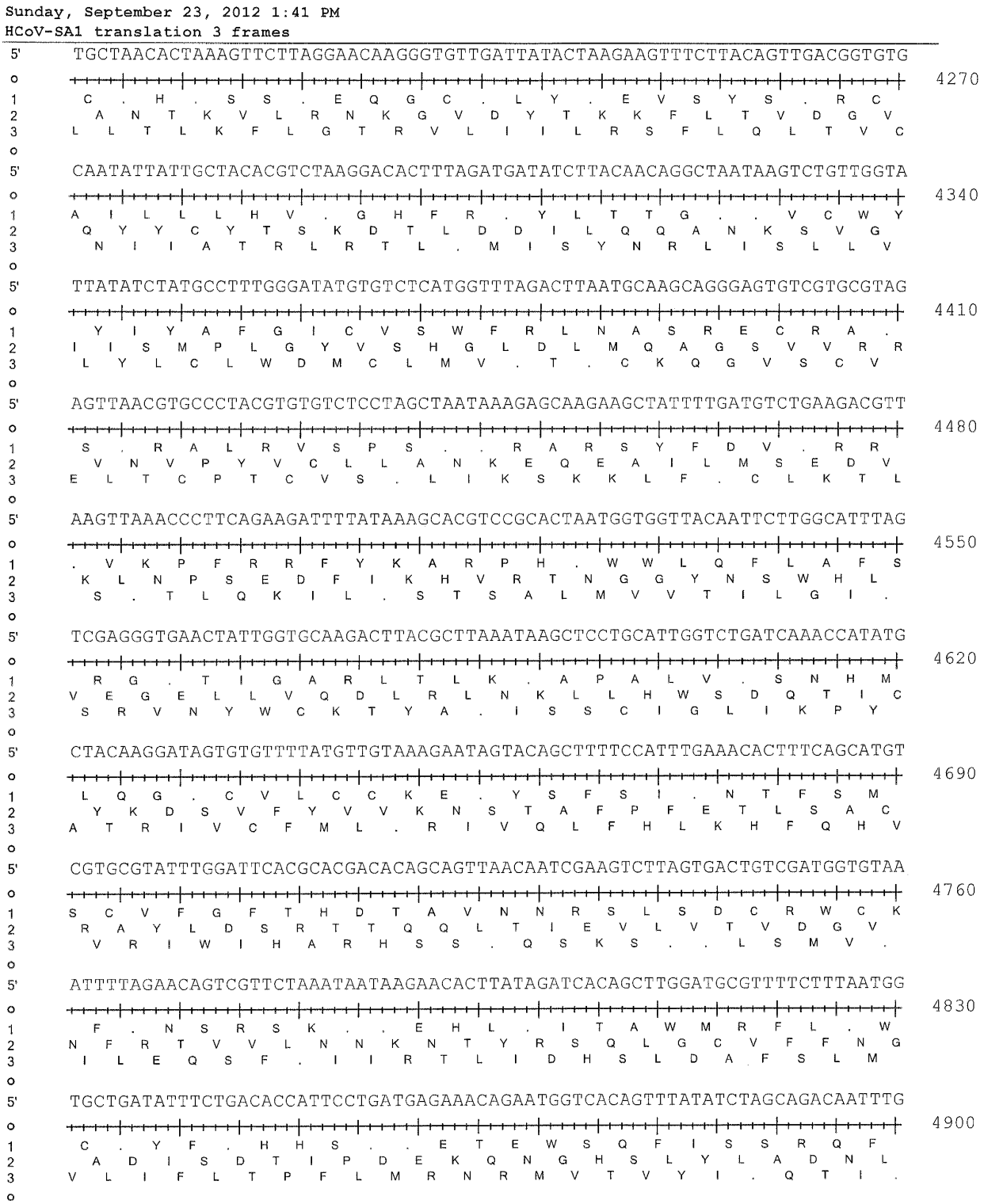
```
5'  ATCCTTTAGTGAACTACTTTTTGATTGCGTTGATCTTTCAGTAGCTTCTACCTATTTTTTAGTCCGCATC
o  +++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++| 2170
1   I  L  .  .  T  T  F  .  L  R  .  S  F  S  S  F  Y  L  F  F  S  P  H
2    S  F  S  E  L  L  F  D  C  V  D  L  S  V  A  S  T  Y  F  L  V  R  I
3   H  P  L  V  N  Y  F  L  I  A  L  I  F  Q  .  L  L  P  I  F  .  S  A  S
o
5'  TTGCAAGATAAGACTGGCGACTTTATGTCTACAATTATTACTTCCTGCCAAACTGCTGTTAGTAAGCTTC
o  +++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++| 2240
1   L  A  R  .  D  W  R  L  Y  V  Y  N  Y  Y  F  L  P  N  C  C  .  .  A  S
2    L  Q  D  K  T  G  D  F  M  S  T  I  I  T  S  C  Q  T  A  V  S  K  L
3     C  K  I  R  L  A  T  L  C  L  Q  L  L  L  P  A  K  L  L  L  V  S  F
o
5'  TAGATACATGTTTTGAAGCTACAGAAGCAACATTTAACTTCTTGTTAGATTTGGCAGGATTGTTCAGAAT
o  +++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++| 2310
1   R  Y  M  F  .  S  Y  R  S  N  I  .  L  L  V  R  F  G  R  I  V  Q  N
2   L  D  T  C  F  E  A  T  E  A  T  F  N  F  L  L  D  L  A  G  L  F  R  I
3    .  I  H  V  L  K  L  Q  K  Q  H  L  T  S  C  .  I  W  Q  D  C  S  E
o
5'  CTTTCTCCGCAATGCCTATGTGTACACTTCACAAGGGTTTGTGGTGGTCAATGGCAAAGTTTCTACACTT
o  +++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++| 2380
1   L  S  P  Q  C  L  C  V  H  F  T  R  V  C  G  G  Q  W  Q  S  F  Y  T
2    F  L  R  N  A  Y  V  Y  T  S  Q  G  F  V  V  V  N  G  K  V  S  T  L
3   S  F  S  A  M  P  M  C  T  L  H  K  G  L  W  W  S  M  A  K  F  L  H  L
o
5'  GTCAAACAAGTGTTAGACTTGCTTAATAAGGGTATGCAACTTTTGCATACAAAGGTCTCCTGGGCTGGTT
o  +++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++| 2450
1   C  Q  T  S  V  R  L  A  .  .  G  Y  A  T  F  A  Y  K  G  L  L  G  W  F
2    V  K  Q  V  L  D  L  L  N  K  G  M  Q  L  L  H  T  K  V  S  W  A  G
3     S  N  K  C  .  T  C  L  I  R  V  C  N  F  C  I  Q  R  S  P  G  L  V
o
5'  CTAAAATCATTGCTGTTATCTACAGCGGCAGGGAGTCTCTAATATTCCCATCGGGAACCTATTACTGTGT
o  +++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++| 2520
1    .  N  H  C  C  Y  L  Q  R  Q  G  V  S  N  I  P  I  G  N  L  L  L  C
2   S  K  I  I  A  V  I  Y  S  G  R  E  S  L  I  F  P  S  G  T  Y  Y  C  V
3    L  K  S  L  L  L  S  T  A  A  G  S  L  .  Y  S  H  R  E  P  I  T  V
o
5'  CACCACTAAGGCTAAGTCCGTTCAACAAGATCTTGACGTTATTTTGCCTGGTGAGTTTTCCAAGAAGCAG
o  +++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++| 2590
1   H  H  .  G  .  V  R  S  T  R  S  .  R  Y  F  A  W  .  V  F  Q  E  A
2    T  T  K  A  K  S  V  Q  Q  D  L  D  V  I  L  P  G  E  F  S  K  K  Q
3   S  P  L  R  L  S  P  F  N  K  I  L  T  L  F  C  L  V  S  F  P  R  S  S
o
5'  TTAGGACTGCTCCAACCTACTGACAATTCTACAACTGTTAGTGTTACTGTATCCAGTAACATGGTTGAAA
o  +++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++| 2660
1   V  R  T  A  P  T  Y  .  Q  F  Y  N  C  .  C  Y  C  I  Q  .  H  G  .  N
2    L  G  L  L  Q  P  T  D  N  S  T  T  V  S  V  T  V  S  S  N  M  V  E
3     .  D  C  S  N  L  L  T  I  L  Q  L  L  V  L  L  Y  P  V  T  W  L  K
o
5'  CTGTTGTGGGTCAACTTGAGCAAACTAATATGCATAGTCCTGATGTTATAGTAGGTGACTATGTCATTAT
o  +++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++| 2730
1    C  C  G  S  T  .  A  N  .  Y  A  .  S  .  C  Y  S  R  .  L  C  H  Y
2   T  V  V  G  Q  L  E  Q  T  N  M  H  S  P  D  V  I  V  G  D  Y  V  I  I
3    L  L  W  V  N  L  S  K  L  I  C  I  V  L  M  L  .  .  V  T  M  S  L
o
5'  TAGTGAAAAATTGTTTGTGCGTAGTAAGGAAGAAGACGGATTTGCCTTCTACCCTGCTTGCACTAATGGT
o  +++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++| 2800
1    .  .  K  I  V  C  A  .  .  G  R  R  R  I  C  L  L  P  C  L  H  .  W
2     S  E  K  L  F  V  R  S  K  E  E  D  G  F  A  F  Y  P  A  C  T  N  G
3   L  V  K  N  C  L  C  V  V  R  K  K  T  D  L  P  S  T  L  L  A  L  M  V
o
```

Sunday, September 23, 2012 1:41 PM
HCoV-SA1 translation 3 frames

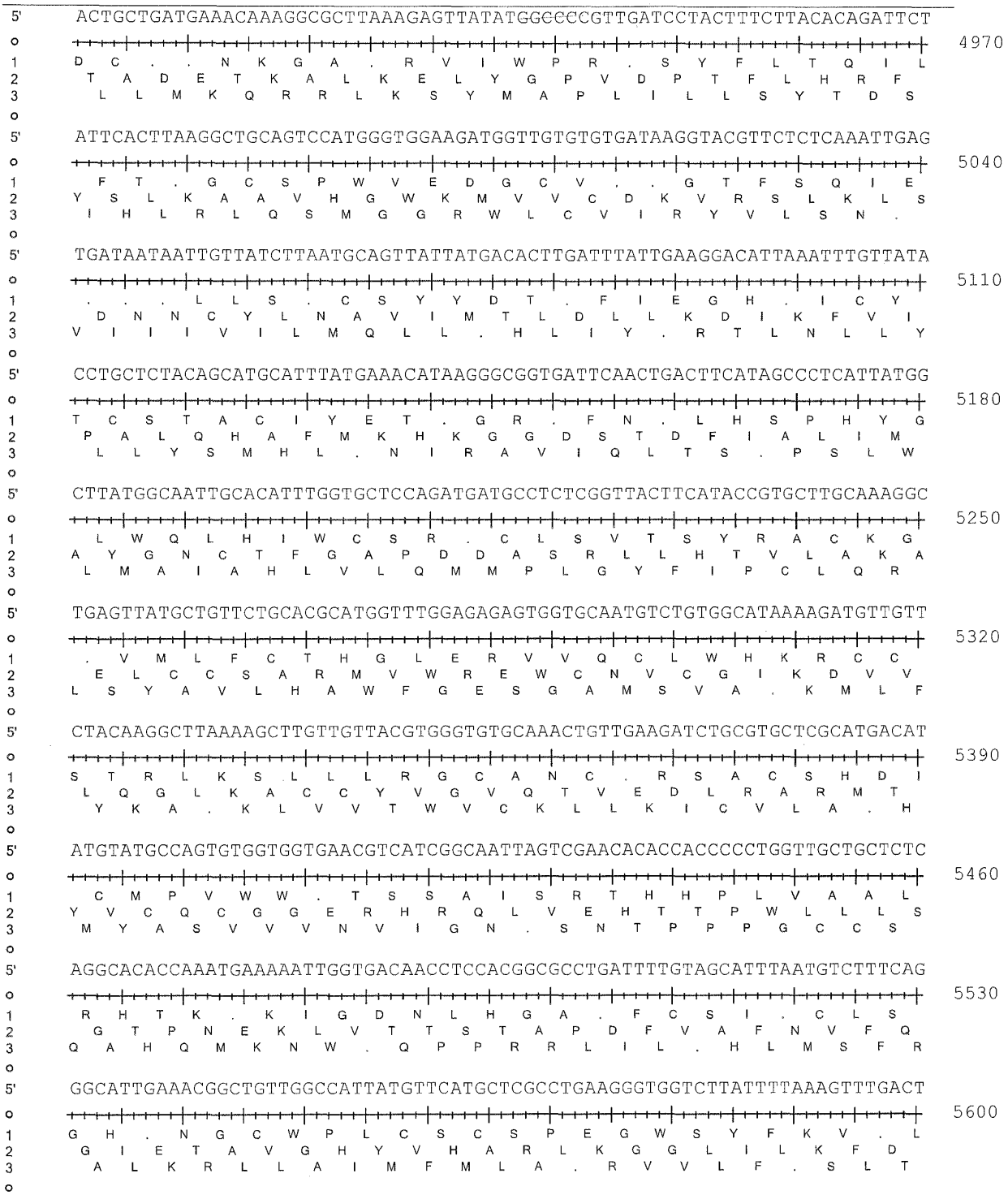
```
5'    CATGCTGTACCGACTCTCTTTAGACTTAAGGGAGGTGCACCTaGTAAAAAAGTAGCCTTTGGCGGTGATC
o     +---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+  2870
1     S   C   C   T   D   S   L   .   T   .   G   R   C   T   .   .   K   S   S   L   W   R   .   S
2       H   A   V   P   T   L   F   R   L   K   G   G   A   P   S   K   K   V   A   F   G   G   D
3         M   L   Y   R   L   S   L   D   L   R   E   V   H   L   V   K   K   .   P   L   A   V   I
o
5'    AAGTACATGAGGTTGCTGCTGTAAGAAGTGTTACTGTCGAGTACAACATTCATGCTGTATTAGACACACT
o     +---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+  2940
1       S   T   .   G   C   C   C   K   K   C   Y   C   R   V   Q   H   S   C   C   I   R   H   T
2     Q   V   H   E   V   A   A   V   R   S   V   T   V   E   Y   N   I   H   A   V   L   D   T   L
3       K   Y   M   R   L   L   L   .   E   V   L   L   S   S   T   T   F   M   L   Y   .   T   H
o
5'    ACTTGCTTCTTCTAGTCTTAGAACCTTTGTTGTAGATAAGTCTTTGTCAATTGAGGAGTTTGCTGACGTA
o     +---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+  3010
1       T   C   F   F   .   S   .   N   L   C   C   R   .   V   F   V   N   .   G   V   C   .   R
2         L   A   S   S   S   L   R   T   F   V   V   D   K   S   L   S   I   E   E   F   A   D   V
3     Y   L   L   L   L   L   V   L   E   P   L   L   .   I   S   L   C   Q   L   R   S   L   L   T   .
o
5'    GTAAAGGAACAAGTCTCAGACTTGCTTGTTAAATTACTGCGTGGAATGCCGATTCCAGATTTTGATTTAG
o     +---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+  3080
1     S   K   G   T   S   L   R   L   A   C   .   I   T   A   W   N   A   D   S   R   F   .   F   R
2       V   K   E   Q   V   S   D   L   L   V   K   L   L   R   G   M   P   I   P   D   F   D   L
3         .   R   N   K   S   Q   T   C   L   L   N   Y   C   V   E   C   R   F   Q   I   L   I   .
o
5'    ACGATTTTATTGACGCACCATGCTATTGCTTTAACGCTGAGGGTGATGCATCCTGGTCTTCTACTATGAT
o     +---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+  3150
1       R   F   Y   .   R   T   M   L   L   L   .   R   .   G   .   C   I   L   V   F   Y   Y   D
2     D   D   F   I   D   A   P   C   Y   C   F   N   A   E   G   D   A   S   W   S   S   T   M   I
3       T   I   L   L   T   H   H   A   I   A   L   T   L   R   V   M   H   P   G   L   L   L   .
o
5'    CTTCTCTCTTCACCCCGTCGAGTGTGACGAGGAGTGTTCTGAAGTAGAGGCTTCAGATTTAGAAGAAGGT
o     +---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+  3220
1       L   L   S   S   P   R   R   V   .   R   G   V   F   .   S   R   G   F   R   F   R   R   R
2         F   S   L   H   P   V   E   C   D   E   E   C   S   E   V   E   A   S   D   L   E   E   G
3     S   S   L   F   T   P   S   S   V   T   R   S   V   L   K   .   R   L   Q   I   .   K   K   V
o
5'    GAATCAGAGTGCATTTCTGAGACTTCAACTGAACAAGTTGACGTTTCTCATGAGACTTCTGACGACGAGT
o     +---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+  3290
1       .   I   R   V   H   F   .   D   F   N   .   T   S   .   R   F   S   .   D   F   .   R   R   V
2         E   S   E   C   I   S   E   T   S   T   E   Q   V   D   V   S   H   E   T   S   D   D   E
3         N   Q   S   A   F   L   R   L   Q   L   N   K   L   T   F   L   M   R   L   L   T   T   S
o
5'    GGGCTGCTGCAGTTGATGAAGCGTTCCCTCTCGATGAAGCAGAAGATGTTACTGAATCTGTGCAAGAAGA
o     +---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+  3360
1       G   C   C   S   .   .   S   V   P   S   R   .   S   R   R   C   Y   .   I   C   A   R   R
2     W   A   A   A   V   D   E   A   F   P   L   D   E   A   E   D   V   T   E   S   V   Q   E   E
3       G   L   L   Q   L   M   K   R   S   L   S   M   K   Q   K   M   L   L   N   L   C   K   K
o
5'    AGCACAACCAGTAGAAGTACCTGTTGAAGATATTGCGCAGGTTGTCATAGCTGACACCTTACAGGAAACT
o     +---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+  3430
1       S   T   T   S   R   S   T   C   .   R   Y   C   A   G   C   H   S   .   H   L   T   G   N
2         A   Q   P   V   E   V   P   V   E   D   I   A   Q   V   V   I   A   D   T   L   Q   E   T
3     K   H   N   Q   .   K   Y   L   L   K   I   L   R   R   L   S   .   L   T   P   Y   R   K   L
o
5'    CCTGTTGTGCCTGATACTGTTGAAGTCCCACCGCAAGTGGTGAAACTTCCGTCTGCACCTCAGACTATCC
o     +---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+  3500
1     S   C   C   A   .   Y   C   .   S   P   T   A   S   G   E   T   S   V   C   T   S   D   Y   P
2       P   V   V   P   D   T   V   E   V   P   P   Q   V   V   K   L   P   S   A   P   Q   T   I
3         L   L   C   L   I   L   L   K   S   H   R   K   W   .   N   F   R   L   H   L   R   L   S
o
```

Sunday, September 23, 2012 1:41 PM
HCoV-SA1 translation 3 frames




Sunday, September 23, 2012 1:41 PM
HCoV-SA1 translation 3 frames
5' TGCTAACACTAAAGTTCTTAGGAACAAGGGTGTTGATTATACTAAGAAGTTTCTTACAGTTGACGGTGTG
4270
C . H . S S . E Q G C . L Y . E V S Y S . R C
A N T K V L R N K G V D Y T K K F L T V D G V
L L T L K F L G T R V L I I L R S F L Q L T V C
5' CAATATTATTGCTACACGTCTAAGGACACTTTAGATGATATCTTACAACAGGCTAATAAGTCTGTTGGTA
4340
A I L L L H V . G H F R . Y L T T G . . V C W Y
Q Y Y C Y T S K D T L D D I L Q Q A N K S V G
N I I A T R L R T L . M I S Y N R L I S L L V
5' TTATATCTATGCCTTTGGGATATGTGTCTCATGGTTTAGACTTAATGCAAGCAGGGAGTGTCGTGCGTAG
4410
Y I Y A F G I C V S W F R L N A S R E C R A .
I I S M P L G Y V S H G L D L M Q A G S V V R R
L Y L C L W D M C L M V . T . C K Q G V S C V
5' AGTTAACGTGCCCTACGTGTGTCTCCTAGCTAATAAAGAGCAAGAAGCTATTTTGATGTCTGAAGACGTT
4480
S . R A L R V S P S . . R A R S Y F D V . R R
V N V P Y V C L L A N K E Q E A I L M S E D V
E L T C P T C V S . L I K S K K L F . C L K T L
5' AAGTTAAACCCTTCAGAAGATTTTATAAAGCACGTCCGCACTAATGGTGGTTACAATTCTTGGCATTTAG
4550
. V K P F R R F Y K A R P H . W W L Q F L A F S
K L N P S E D F I K H V R T N G G Y N S W H L
S . T L Q K I L . S T S A L M V V T I L G I .
5' TCGAGGGTGAACTATTGGTGCAAGACTTACGCTTAAATAAGCTCCTGCATTGGTCTGATCAAACCATATG
4620
R G . T I G A R L T L K . A P A L V . S N H M
V E G E L L V Q D L R L N K L L H W S D Q T I C
S R V N Y W C K T Y A . I S S C I G L I K P Y
5' CTACAAGGATAGTGTGTTTTATGTTGTAAAGAATAGTACAGCTTTTCCATTTGAAACACTTTCAGCATGT
4690
L Q G . C V L C C K E . Y S F S I . N T F S M
Y K D S V F Y V V K N S T A F P F E T L S A C
A T R I V C F M L . R I V Q L F H L K H F Q H V
5' CGTGCGTATTTGGATTCACGCACGACACAGCAGTTAACAATCGAAGTCTTAGTGACTGTCGATGGTGTAA
4760
S C V F G F T H D T A V N N R S L S D C R W C K
R A Y L D S R T T Q Q L T I E V L V T V D G V
V R I W I H A R H S S . Q S K S . . L S M V .
5' ATTTTAGAACAGTCGTTCTAAATAATAAGAACACTTATAGATCACAGCTTGGATGCGTTTTCTTTAATGG
4830
F . N S R S K . . E H L . I T A W M R F L . W
N F R T V V L N N K N T Y R S Q L G C V F F N G
I L E Q S F . I I R T L I D H S L D A F S L M
5' TGCTGATATTTCTGACACCATTCCTGATGAGAAACAGAATGGTCACAGTTTATATCTAGCAGACAATTTG
4900
C . Y F . H H S . . E T E W S Q F I S S R Q F
A D I S D T I P D E K Q N G H S L Y L A D N L
V L I F L T P F L M R N R M V T V Y I . Q T I .

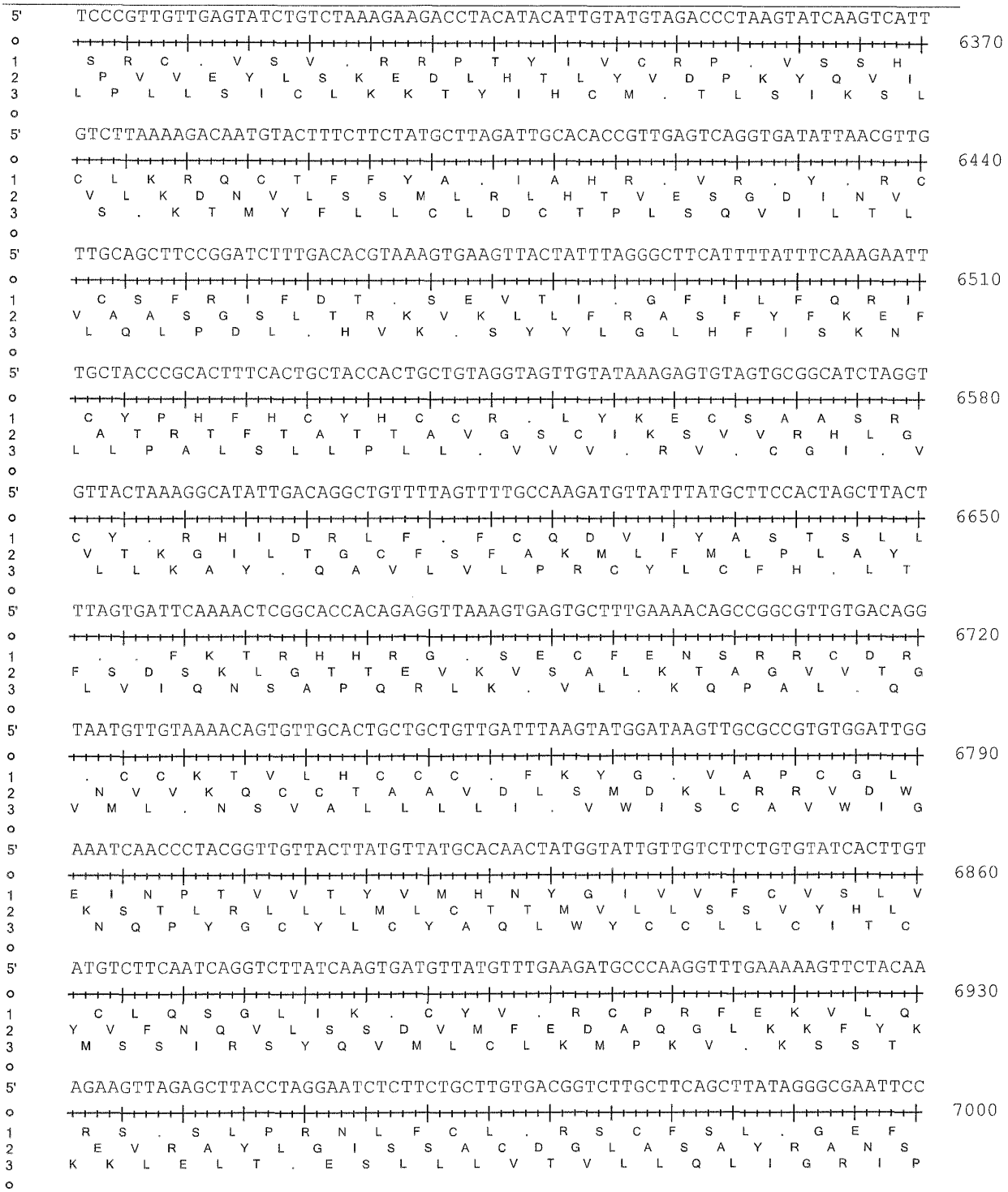
Sunday, September 23, 2012 1:41 PM
HCoV-SA1 translation 3 frames



Sunday, September 23, 2012 1:41 PM
HCoV-SA1 translation 3 frames

```
5'    CTGGCACCGTTAGCAAGACTTCAGACTGGAAGTGCAAGGTGACAGATGTACTTTTCCCCGGCCAAAAATA
o     +----|----|----|----|----|----|----|----|----|----|----|----|----|----|  5670
1       W   H   R   .   Q   D   F   R   L   E   V   Q   G   D   R   C   T   F   P   R   P   K   I
2     S   G   T   V   S   K   T   S   D   W   K   C   K   V   T   D   V   L   F   P   G   Q   K   Y
3       L   A   P   L   A   R   L   Q   T   G   S   A   R   .   Q   M   Y   F   S   P   A   K   N
o
5'    CAGTAGCGATTGTAATGTCGTACGGTATTCTTTGGACGGTAATTTCAGAACAGAGGTTGATCCCGACCTA
o     +----|----|----|----|----|----|----|----|----|----|----|----|----|----|  5740
1       Q   .   R   L   .   C   R   T   V   F   F   G   R   .   F   Q   N   R   G   .   S   R   P
2     S   S   D   C   N   V   V   R   Y   S   L   D   G   N   F   R   T   E   V   D   P   D   L
3     T   V   A   I   V   M   S   Y   G   I   L   W   T   V   I   S   E   Q   R   L   I   P   T   Y
o
5'    TCTGCTTTCTATGTTAAGGATGGTAAATACTTTACAAGTGAACCACCCGTAACATATTCACCAGCTACAA
o     +----|----|----|----|----|----|----|----|----|----|----|----|----|----|  5810
1       I   C   F   L   C   .   G   W   .   I   L   Y   K   .   T   T   R   N   I   F   T   S   Y   N
2     S   A   F   Y   V   K   D   G   K   Y   F   T   S   E   P   P   V   T   Y   S   P   A   T
3       L   L   S   M   L   R   M   V   N   T   L   Q   V   N   H   P   .   H   I   H   Q   L   Q
o
5'    TTTTAGCTGGTAGTGTCTACACTAATAGCTGCCTTGTATCGTCTGATGGACAACCTGGCGGTGATGCTAT
o     +----|----|----|----|----|----|----|----|----|----|----|----|----|----|  5880
1       F   S   W   .   C   L   H   .   .   L   P   C   I   V   .   W   T   T   W   R   .   C   Y
2     I   L   A   G   S   V   Y   T   N   S   C   L   V   S   S   D   G   Q   P   G   G   D   A   I
3       F   .   L   V   V   S   T   L   I   A   A   L   Y   R   L   M   D   N   L   A   V   M   L
o
5'    TAGTTTGAGTTTTAATAACCTTTTAGGGTTTGATTCTAGTAAACCAGTCACTAAGAAATACACTTACTCC
o     +----|----|----|----|----|----|----|----|----|----|----|----|----|----|  5950
1       .   F   E   F   .   .   P   F   R   V   .   F   .   .   T   S   H   .   E   I   H   L   L
2     S   L   S   F   N   N   L   L   G   F   D   S   S   K   P   V   T   K   K   Y   T   Y   S
3     L   V   .   V   L   I   T   F   .   G   L   I   L   V   N   Q   S   L   R   N   T   L   T   P
o
5'    TTCTTGCCTAAAGAAGACGGCGATGTGTTGTTGGCTGAGTTTGACACTTATGACCCTATTTATAAGAATG
o     +----|----|----|----|----|----|----|----|----|----|----|----|----|----|  6020
1     L   L   A   .   R   R   R   R   C   V   V   G   .   V   .   H   L   .   P   Y   L   .   E   W
2       F   L   P   K   E   D   G   D   V   L   L   A   E   F   D   T   Y   D   P   I   Y   K   N
3       S   C   L   K   K   T   A   M   C   C   W   L   S   L   T   L   M   T   L   F   I   R   M
o
5'    GTGCCATGTATAAAGGCAAACCAATTCTTTGGGTCAATAAAGCATCTTATGATACTAATCTTAATAAGTT
o     +----|----|----|----|----|----|----|----|----|----|----|----|----|----|  6090
1       C   H   V   .   R   Q   T   N   S   L   G   Q   .   S   I   L   .   Y   .   S   .   .   V
2     G   A   M   Y   K   G   K   P   I   L   W   V   N   K   A   S   Y   D   T   N   L   N   K   F
3       V   P   C   I   K   A   N   Q   F   F   G   S   I   K   H   L   M   I   L   I   L   I   S
o
5'    CAATAGAGCTAGTTTGCGTCAAATTTTTGACGTAGCCCCCATTGAACTCGAAAATAAATTCACACCTTTG
o     +----|----|----|----|----|----|----|----|----|----|----|----|----|----|  6160
1       Q   .   S   .   F   A   S   N   F   .   R   S   P   H   .   T   R   K   .   I   H   T   F
2     N   R   A   S   L   R   Q   I   F   D   V   A   P   I   E   L   E   N   K   F   T   P   L
3     S   I   E   L   V   C   V   K   F   L   T   .   P   P   L   N   S   K   I   N   S   H   L   .
o
5'    AGTGTGGAGTCTACACCAGTTGAACCTCCAACTGTAGATGTGGTAGCACTTCAACAGGAAATGACAATTG
o     +----|----|----|----|----|----|----|----|----|----|----|----|----|----|  6230
1     E   C   G   V   Y   T   S   .   T   S   N   C   R   C   G   S   T   S   T   G   N   D   N   C
2       S   V   E   S   T   P   V   E   P   P   T   V   D   V   V   A   L   Q   Q   E   M   T   I
3       V   W   S   L   H   Q   L   N   L   Q   L   .   M   W   .   H   F   N   R   K   .   Q   L
o
5'    TCAAATGTAAGGGTTTAAATAAACCTTTCGTGAAGGACAATGTCAGTTTCGTTGCTGATGATTCAGGTAC
o     +----|----|----|----|----|----|----|----|----|----|----|----|----|----|  6300
1       Q   M   .   G   F   K   .   T   F   R   E   G   Q   C   Q   F   R   C   .   .   F   R   Y
2     V   K   C   K   G   L   N   K   P   F   V   K   D   N   V   S   F   V   A   D   D   S   G   T
3       S   N   V   R   V   .   I   N   L   S   .   R   T   M   S   V   S   L   L   M   I   Q   V
o
```

Sunday, September 23, 2012 1:41 PM
HCoV-SA1 translation 3 frames

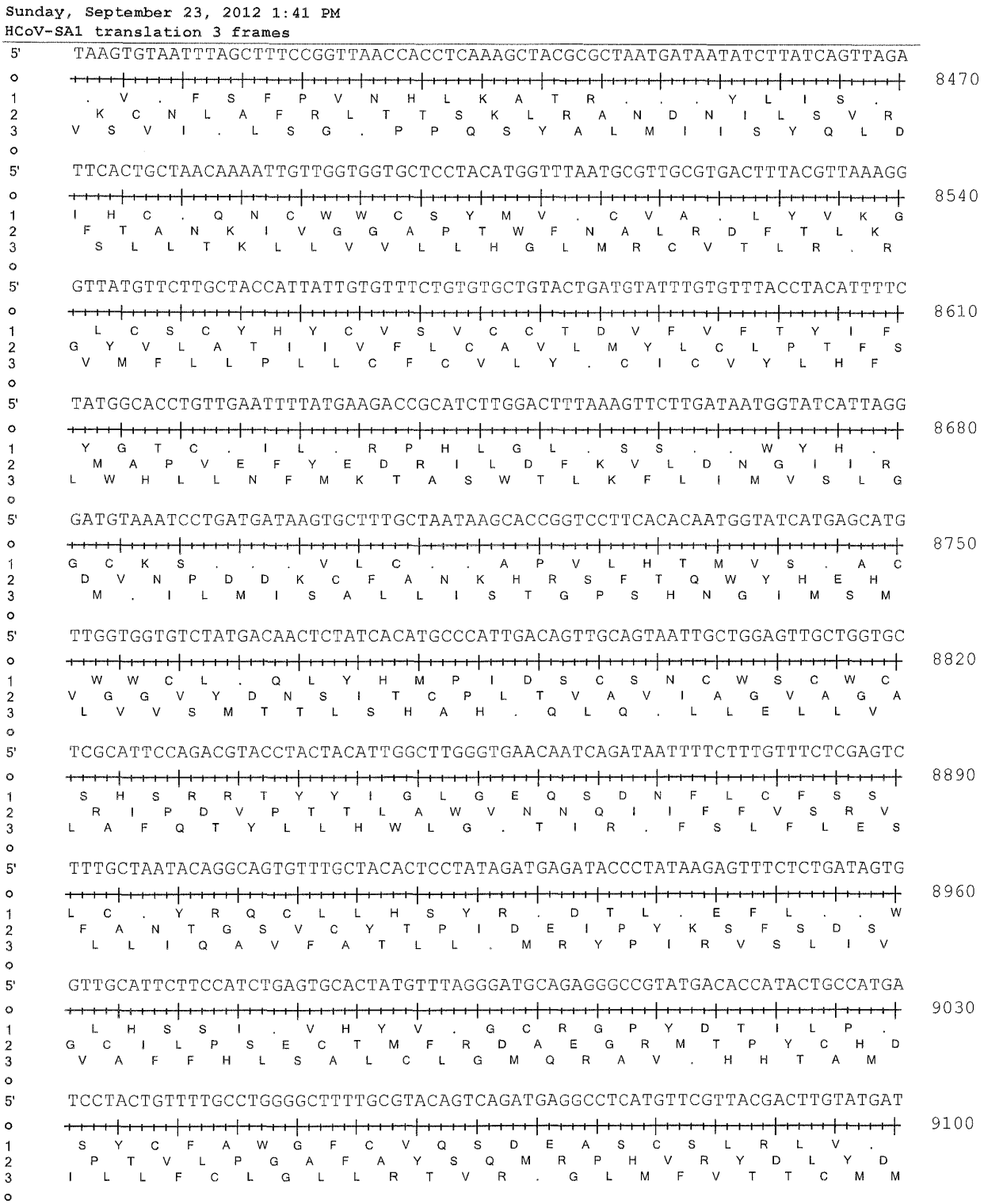


Sunday, September 23, 2012 1:41 PM
HCoV-SA1 translation 3 frames

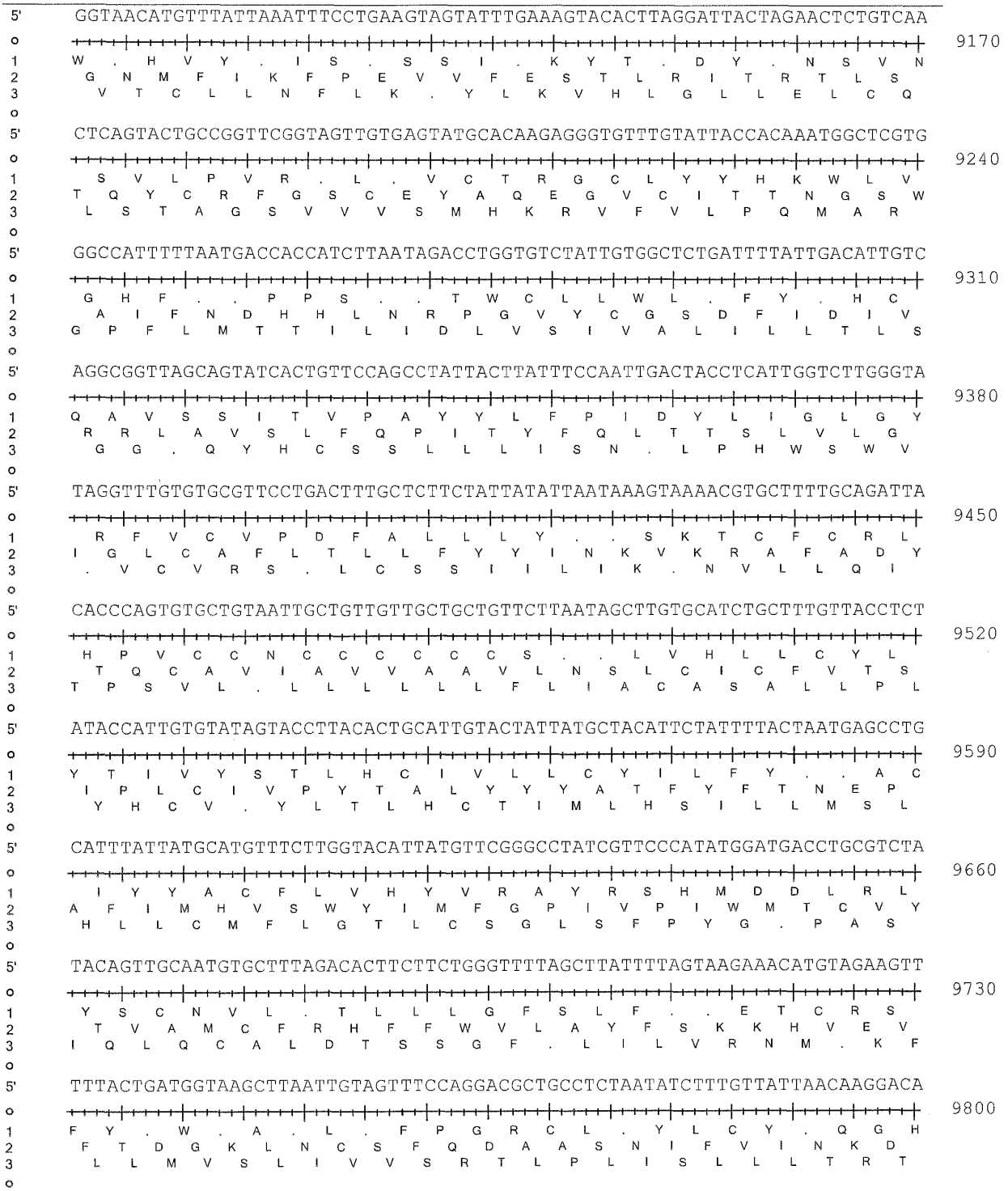
```
5'  TTTGATGTACCTACATTCTGCGCAAACCGTTCTGCAATGTGTAATTGGTGCTTGATTAGCCAAGATTCCA
o  +----|----|----|----|----|----|----|----|----|----|----|----|----|----|  7070
1  L . C T Y I L R K P F C N V . L V L D . P R F H
2   F D V P T F C A N R S A M C N W C L I S Q D S
3    L M Y L H S A Q T V L Q C V I G A . L A K I P
o
5'  TAACTCACTACCCAGCTCTTAAGATGGTTCAAACACATCTTAGCCACTATGTTCTTAACATAGATTGGTT
o  +----|----|----|----|----|----|----|----|----|----|----|----|----|----|  7140
1    N S L P S S . D G S N T S . P L C S . H R L V
2   I T H Y P A L K M V Q T H L S H Y V L N I D W L
3    . L T T Q L L R W F K H I L A T M F L T . I G
o
5'  GTGGTTTGCATTTGAGACTGGTTTGGCATACATGCTCTATACCTCGGCCTTCAACTGGTTGTTGTTGGCA
o  +----|----|----|----|----|----|----|----|----|----|----|----|----|----|  7210
1   V V C I . D W F G I H A L Y L G L Q L V V V G
2    W F A F E T G L A Y M L Y T S A F N W L L L A
3   C G L H L R L V W H T C S I P R P S T G C C W Q
o
5'  GGTACATTGCATTATTTCTTTGCACAGACTTCCATATTTGTAGACTGGCGGTCATACAATTATGCTGTGT
o  +----|----|----|----|----|----|----|----|----|----|----|----|----|----|  7280
1  R Y I A L F L C T D F H I C R L A V I Q L C C V
2   G T L H Y F F A Q T S I F V D W R S Y N Y A V
3    V H C I I S L H R L P Y L . T G G H T I M L C
o
5'  CTAGTGCCTTCTGGTTATTCACCCACATTCCAATGGCGGGTTTGGTACGAATGTATAATTTGTTAGCATG
o  +----|----|----|----|----|----|----|----|----|----|----|----|----|----|  7350
1    . C L L V I H P H S N G G F G T N V . F V S M
2   S S A F W L F T H I P M A G L V R M Y N L L A C
3    L V P S G Y S P T F Q W R V W Y E C I I C . H
o
5'  CCTTTGGCTTTTACGCAAGTTTTATCAGCATGTAATCAATGGTTGCAAAGATACGGCATGCTTGCTCTGC
o  +----|----|----|----|----|----|----|----|----|----|----|----|----|----|  7420
1   P L A F T Q V L S A C N Q W L Q R Y G M L A L
2    L W L L R K F Y Q H V I N G C K D T A C L L C
3   A F G F Y A S F I S M . S M V A K I R H A C S A
o
5'  TATAAGAGGAACCGACTTACTAGAGTTGAAGCTTCTACCGTTGTCTGTGGTGGAAAACGTACGTTTTATA
o  +----|----|----|----|----|----|----|----|----|----|----|----|----|----|  7490
1  L . E E P T Y . S . S F Y R C L W W K T Y V L Y
2   Y K R N R L T R V E A S T V V C G G K R T F Y
3    I R G T D L L E L K L L P L S V V E N V R F I
o
5'  TCACAGCAAATGGCGGTATTTCATTCTGTCGTAGGCATAATTGGAATTGTGTGGATTGTGACACTGCAGG
o  +----|----|----|----|----|----|----|----|----|----|----|----|----|----|  7560
1   H S K W R Y F I L S . A . L E L C G L . H C R
2  I T A N G G I S F C R R H N W N C V D C D T A G
3   S Q Q M A V F H S V V G I I G I V W I V T L Q
o
5'  TGTGGGGAATACCTTCATCTGTGAAGAAGTCGCAAATGACCTCACTACCGCCCTACGCAGGCCTATTAAC
o  +----|----|----|----|----|----|----|----|----|----|----|----|----|----|  7630
1   C G E Y L H L . R S R K . P H Y R P T Q A Y .
2    V G N T F I C E E V A N D L T T A L R R P I N
3  V W G I P S S V K K S Q M T S L P P Y A G L L T
o
5'  GCTACGGATAGATCACATTATTATGTGGATTCCGTTACAGTTAAAGAGACTGTTGTTCAGTTTAATTATC
o  +----|----|----|----|----|----|----|----|----|----|----|----|----|----|  7700
1  R Y G . I T L L C G F R Y S . R D C C S V . L S
2   A T D R S H Y Y V D S V T V K E T V V Q F N Y
3    L R I D H I I M W I P L Q L K R L L F S L I I
o
```

Sunday, September 23, 2012 1:41 PM
HCoV-SA1 translation 3 frames

```
5'    GTAGAGACGGTCAACCATTCTACGAGCGGTTTCCCCTCTGCGCTTTTACAAATCTAGATAAGTTGAAGTT
o     +++++|+++++|+++++|+++++|+++++|+++++|+++++|+++++|+++++|+++++|+++++|+++++| 7770
1       .  R  R  S  T  I  L  R  A  V  S  P  L  R  F  Y  K  S  R  .  V  E  V
2     R  R  D  G  Q  P  F  Y  E  R  F  P  L  C  A  F  T  N  L  D  K  L  K  F
3      V  E  T  V  N  H  S  T  S  G  F  P  S  A  L  L  Q  I  .  I  S  .  S
o
5'    CAAAGAGGTCTGTAAAACTACTACTGGTATACCTGAATACAACTTTATCATCTACGACTCATCAGATCGT
o     +++++|+++++|+++++|+++++|+++++|+++++|+++++|+++++|+++++|+++++|+++++|+++++| 7840
1      Q  R  G  L  .  N  Y  Y  W  Y  T  .  I  Q  L  Y  H  L  R  L  I  R  S
2        K  E  V  C  K  T  T  T  G  I  P  E  Y  N  F  I  I  Y  D  S  S  D  R
3     S  K  R  S  V  K  L  L  L  V  Y  L  N  T  T  L  S  S  T  T  H  Q  I  V
o
5'    GGCCAGGAAAGTTTAGCTAGGTCTGCATGTGTTTATTATTCTCAAGTCTTGTGTAAATCAATTCTTTTGG
o     +++++|+++++|+++++|+++++|+++++|+++++|+++++|+++++|+++++|+++++|+++++|+++++| 7910
1     W  P  G  K  F  S  .  V  C  M  C  L  L  F  S  S  L  V  .  I  N  S  F  G
2       G  Q  E  S  L  A  R  S  A  C  V  Y  Y  S  Q  V  L  C  K  S  I  L  L
3         A  R  K  V  .  L  G  L  H  V  F  I  I  L  K  S  C  V  N  Q  F  F  W
o
5'    TTGACTCAAGTTTGGTTACTTCTGTTGGTGATTCTAGTGAAATCGCCACTAAAATGTTTGATTCCTTTGT
o     +++++|+++++|+++++|+++++|+++++|+++++|+++++|+++++|+++++|+++++|+++++|+++++| 7980
1       .  L  K  F  G  Y  F  C  W  .  F  .  .  N  R  H  .  N  V  .  F  L  C
2     V  D  S  S  L  V  T  S  V  G  D  S  S  E  I  A  T  K  M  F  D  S  F  V
3      L  T  Q  V  W  L  L  L  L  V  I  L  V  K  S  P  L  K  C  L  I  P  L
o
5'    TAATAGTTTCGTCTCGCTGTATAATGTCACACGCGATAAGTTGGAAAAACTTATCTCTACTGCTCGTGAT
o     +++++|+++++|+++++|+++++|+++++|+++++|+++++|+++++|+++++|+++++|+++++|+++++| 8050
1       .  .  F  R  L  A  V  .  C  H  T  R  .  V  G  K  T  Y  L  Y  C  S  .
2         N  S  F  V  S  L  Y  N  V  T  R  D  K  L  E  K  L  I  S  T  A  R  D
3      L  I  V  S  S  R  C  I  M  S  H  A  I  S  W  K  N  L  S  L  L  V  M
o
5'    GGCGTAAGGCGAGGCGATAACTTCCATAGTGTCTTAACAACATTCATTGACGCAGCACGAGGCCCCGCAG
o     +++++|+++++|+++++|+++++|+++++|+++++|+++++|+++++|+++++|+++++|+++++|+++++| 8120
1     W  R  K  A  R  R  .  L  P  .  C  L  N  N  I  H  .  R  S  T  R  P  R  R
2       G  V  R  R  G  D  N  F  H  S  V  L  T  T  F  I  D  A  A  R  G  P  A
3         A  .  G  E  A  I  T  S  I  V  S  .  Q  H  S  L  T  Q  H  E  A  P  Q
o
5'    GTGTGGAGTCTGATGTTGAGACCAATGAAATTGTTGACTCTGTGCAGTATGCTCATAAACATGACATACA
o     +++++|+++++|+++++|+++++|+++++|+++++|+++++|+++++|+++++|+++++|+++++|+++++| 8190
1       C  G  V  .  C  .  D  Q  .  N  C  .  L  C  A  V  C  S  .  T  .  H  T
2     G  V  E  S  D  V  E  T  N  E  I  V  D  S  V  Q  Y  A  H  K  H  D  I  Q
3      V  W  S  L  M  L  R  P  M  K  L  L  T  L  C  S  M  L  I  N  M  T  Y
o
5'    AATTACTAATGAGAGCTACAATAATTATGTACCCTCATATGTTAAACCTGATAGTGTGTCTACCAGCGAT
o     +++++|+++++|+++++|+++++|+++++|+++++|+++++|+++++|+++++|+++++|+++++|+++++| 8260
1     N  Y  .  .  E  L  Q  .  L  C  T  L  I  C  .  T  .  .  C  V  Y  Q  R
2       I  T  N  E  S  Y  N  N  Y  V  P  S  Y  V  K  P  D  S  V  S  T  S  D
3     K  L  L  M  R  A  T  I  I  M  Y  P  H  M  L  N  L  I  V  C  L  P  A  I
o
5'    TTAGGTAGTCTCATTGATTGTAATGCGGCTTCAGTTAACCAAATTGTCTTGCGTAATTCTAATGGTGCTT
o     +++++|+++++|+++++|+++++|+++++|+++++|+++++|+++++|+++++|+++++|+++++|+++++| 8330
1     F  R  .  S  H  .  L  .  C  G  F  S  .  P  N  C  L  A  .  F  .  W  C  L
2       L  G  S  L  I  D  C  N  A  A  S  V  N  Q  I  V  L  R  N  S  N  G  A
3         .  V  V  S  L  I  V  M  R  L  Q  L  T  K  L  S  C  V  I  L  M  V  L
o
5'    GCATTTGGAACGCTGCTGCATATATGAAACTCTCGGATGCACTTAAACGACAGATTCGCATTGCATGCCG
o     +++++|+++++|+++++|+++++|+++++|+++++|+++++|+++++|+++++|+++++|+++++|+++++| 8400
1       H  L  E  R  C  C  I  Y  E  T  L  G  C  T  .  T  T  D  S  H  C  M  P
2     C  I  W  N  A  A  A  Y  M  K  L  S  D  A  L  K  R  Q  I  R  I  A  C  R
3      A  F  G  T  L  L  H  I  .  N  S  R  M  H  L  N  D  R  F  A  L  H  A
o
```


Sunday, September 23, 2012 1:41 PM
HCoV-SA1 translation 3 frames
5' TAAGTGTAATTTAGCTTTCCGGTTAACCACCTCAAAGCTACGCGCTAATGATAATATCTTATCAGTTAGA
8470
5' TTCACTGCTAACAAAATTGTTGGTGGTGCTCCTACATGGTTTAATGCGTTGCGTGACTTTACGTTAAAGG
8540
5' GTTATGTTCTTGCTACCATTATTGTGTTTCTGTGTGCTGTACTGATGTATTTGTGTTTACCTACATTTTC
8610
5' TATGGCACCTGTTGAATTTTATGAAGACCGCATCTTGGACTTTAAAGTTCTTGATAATGGTATCATTAGG
8680
5' GATGTAAATCCTGATGATAAGTGCTTTGCTAATAAGCACCGGTCCTTCACACAATGGTATCATGAGCATG
8750
5' TTGGTGGTGTCTATGACAACTCTATCACATGCCCATTGACAGTTGCAGTAATTGCTGGAGTTGCTGGTGC
8820
5' TCGCATTCCAGACGTACCTACTACATTGGCTTGGGTGAACAATCAGATAATTTTCTTTGTTTCTCGAGTC
8890
5' TTTGCTAATACAGGCAGTGTTTGCTACACTCCTATAGATGAGATACCCTATAAGAGTTTCTCTGATAGTG
8960
5' GTTGCATTCTTCCATCTGAGTGCACTATGTTTAGGGATGCAGAGGGCCGTATGACACCATACTGCCATGA
9030
5' TCCTACTGTTTTGCCTGGGGCTTTTGCGTACAGTCAGATGAGGCCTCATGTTCGTTACGACTTGTATGAT
9100

Sunday, September 23, 2012 1:41 PM
HCoV-SA1 translation 3 frames



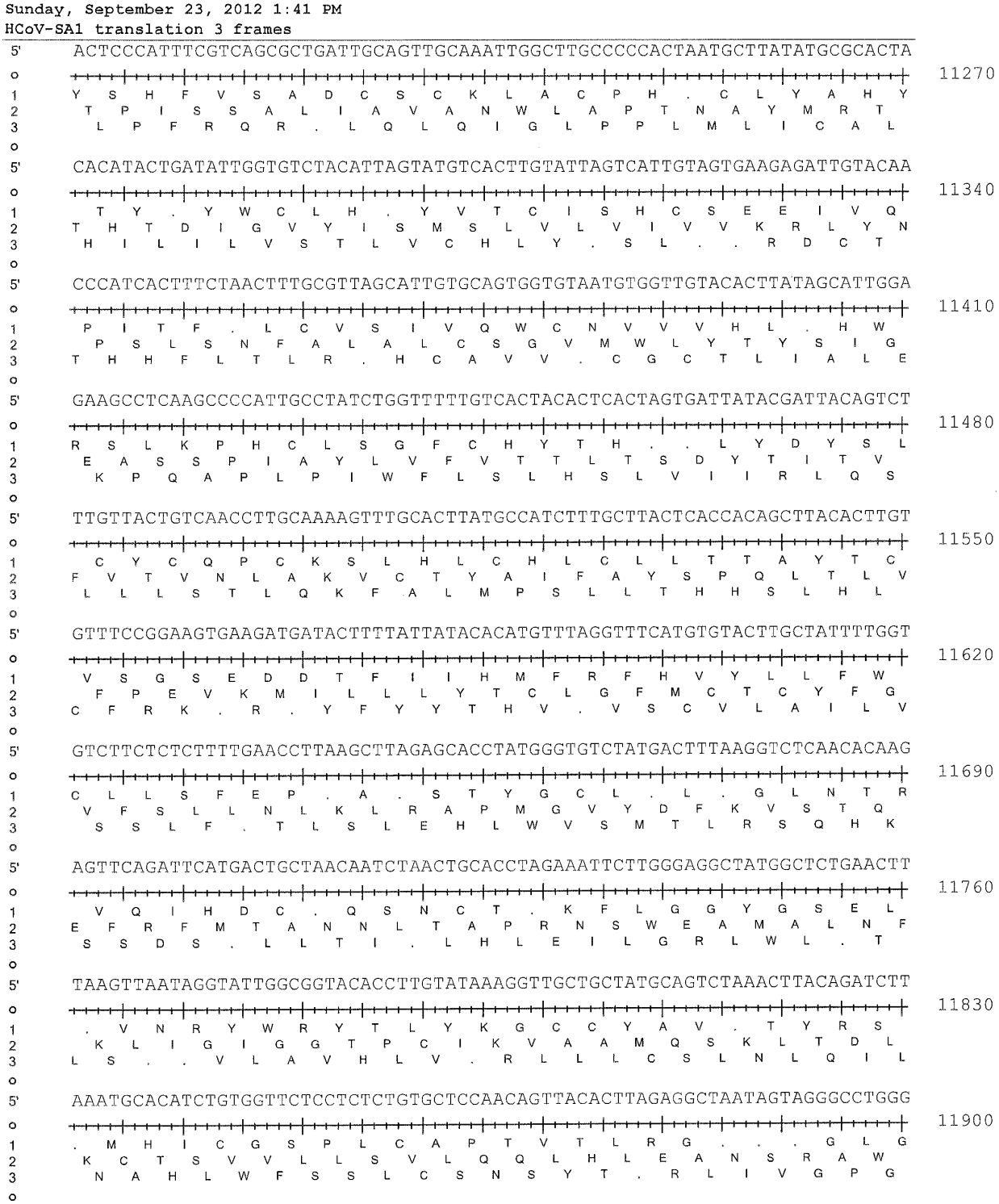
Sunday, September 23, 2012 1:41 PM
HCoV-SA1 translation 3 frames

```
5'    CTTATGCAGCTCTTAGAAACTCTTTAACTAATGATGCCTATTCACGATTTTTGGGGTTGTTTAACAAGTA
o     +----|----|----|----|----|----|----|----|----|----|----|----|----|----|  9870
1       L   C   S   S   .   K   L   F   N   .   .   C   L   F   T   I   F   G   V   V   .   Q   V
2     T   Y   A   A   L   R   N   S   L   T   N   D   A   Y   S   R   F   L   G   L   F   N   K   Y
3       L   M   Q   L   L   E   T   L   .   L   M   M   P   I   H   D   F   W   G   C   L   T   S
o
5'    TAAGTACTTCTCTGGTGCTATGGAAACAGCCGCTTATCGTGAAGCTGCAGCATGTCATCTTGCTAAAGCC
o     +----|----|----|----|----|----|----|----|----|----|----|----|----|----|  9940
1       .   V   L   L   W   C   Y   G   N   S   R   L   S   .   S   C   S   M   S   S   C   .   S
2       K   Y   F   S   G   A   M   E   T   A   A   Y   R   E   A   A   A   C   H   L   A   K   A
3     I   S   T   S   L   V   L   W   K   Q   P   L   I   V   K   L   Q   H   V   I   L   L   K   P
o
5'    TTACAAACATACAGCGAGACTGGTAGTGATCTTCTTTACCAACCACCCAACTGTAGCATAACCTCTGGCG
o     +----|----|----|----|----|----|----|----|----|----|----|----|----|----|  10010
1     L   T   N   I   Q   R   D   W   .   .   S   S   L   P   T   T   Q   L   .   H   N   L   W   R
2       L   Q   T   Y   S   E   T   G   S   D   L   L   Y   Q   P   P   N   C   S   I   T   S   G
3       Y   K   H   T   A   R   L   V   V   I   F   F   T   N   H   P   T   V   A   .   P   L   A
o
5'    TGTTGCAAAGCGGTTTGGTGAAAATGTCACATCCCAGTGGAGATGTTGAGGCTTGTATGGTTCAGGTTAC
o     +----|----|----|----|----|----|----|----|----|----|----|----|----|----|  10080
1       V   A   K   R   F   G   E   N   V   T   S   Q   W   R   C   .   G   L   Y   G   S   G   Y
2     V   L   Q   S   G   L   V   K   M   S   H   P   S   G   D   V   E   A   C   M   V   Q   V   T
3       C   C   K   A   V   W   .   K   C   H   I   P   V   E   M   L   R   L   V   W   F   R   L
o
5'    CTGCGGTAGCATGACTCTTAATGGTCTTTGGCTTGACAACACAGTCTGGTGCCCACGACACGTAATGTGC
o     +----|----|----|----|----|----|----|----|----|----|----|----|----|----|  10150
1       L   R   .   H   D   S   .   W   S   L   A   .   Q   H   S   L   V   P   T   T   R   N   V
2       C   G   S   M   T   L   N   G   L   W   L   D   N   T   V   W   C   P   R   H   V   M   C
3     P   A   V   A   .   L   L   M   V   F   G   L   T   T   Q   S   G   A   H   D   T   .   C   A
o
5'    CCGGCTGACCAGTTGTCTGATCCTAATTATGATGCCTTGTTGATTTCTATGACTAATCATAGTTTCAGTG
o     +----|----|----|----|----|----|----|----|----|----|----|----|----|----|  10220
1     P   G   .   P   V   V   .   S   .   L   .   C   L   V   D   F   Y   D   .   S   .   F   Q   C
2       P   A   D   Q   L   S   D   P   N   Y   D   A   L   L   I   S   M   T   N   H   S   F   S
3       R   L   T   S   C   L   I   L   I   M   M   P   C   .   F   L   .   L   I   I   V   S   V
o
5'    TGCAAAAACACATTGGCGCTCCAGCAAACTTGCGTGTTGTTGGTCATGCCATGCAAGGCACTCTTTTGAA
o     +----|----|----|----|----|----|----|----|----|----|----|----|----|----|  10290
1       A   K   T   H   W   R   S   S   K   L   A   C   C   W   S   C   H   A   R   H   S   F   E
2     V   Q   K   H   I   G   A   P   A   N   L   R   V   V   G   H   A   M   Q   G   T   L   L   K
3       C   K   N   T   L   A   L   Q   Q   T   C   V   L   L   V   M   P   C   K   A   L   F   .
o
5'    GTTGACTGTCGATGTTGCTAACCCTAGCACTCCAGCCTACACTTTTACAACAGTGAAACCTGGCGCAGCA
o     +----|----|----|----|----|----|----|----|----|----|----|----|----|----|  10360
1     V   D   C   R   C   C   .   P   .   H   S   S   L   H   F   Y   N   S   E   T   W   R   S
2       L   T   V   D   V   A   N   P   S   T   P   A   Y   T   F   T   T   V   K   P   G   A   A
3     S   .   L   S   M   L   L   T   L   A   L   Q   P   T   L   L   Q   Q   .   N   L   A   Q   H
o
5'    TTTAGTGTGTTAGCATGCTATAATGGTCGTCCGACTGGTACATTCACTGTTGTAATGCGCCCTAACTACA
o     +----|----|----|----|----|----|----|----|----|----|----|----|----|----|  10430
1     I   .   C   V   S   M   L   .   W   S   S   D   W   Y   I   H   C   C   N   A   P   .   L   H
2       F   S   V   L   A   C   Y   N   G   R   P   T   G   T   F   T   V   V   M   R   P   N   Y
3       L   V   C   .   H   A   I   M   V   V   R   L   V   H   S   L   L   .   C   A   L   T   T
o
5'    CAATTAAGGGTTCCTTTCTGTGTGGTTCTTGTGGTAGTGTTGGTTACACCAAGGAGGGTAGTGTGATCAA
o     +----|----|----|----|----|----|----|----|----|----|----|----|----|----|  10500
1       N   .   G   F   L   S   V   W   F   L   W   .   C   W   L   H   Q   G   G   .   C   D   Q
2     T   I   K   G   S   F   L   C   G   S   C   G   S   V   G   Y   T   K   E   G   S   V   I   N
3       Q   L   R   V   P   F   C   V   V   L   V   V   V   L   V   T   P   R   R   V   V   .   S
o
```

Figure 15 continued

Sunday, September 23, 2012 1:41 PM
HCoV-SA1 translation 3 frames

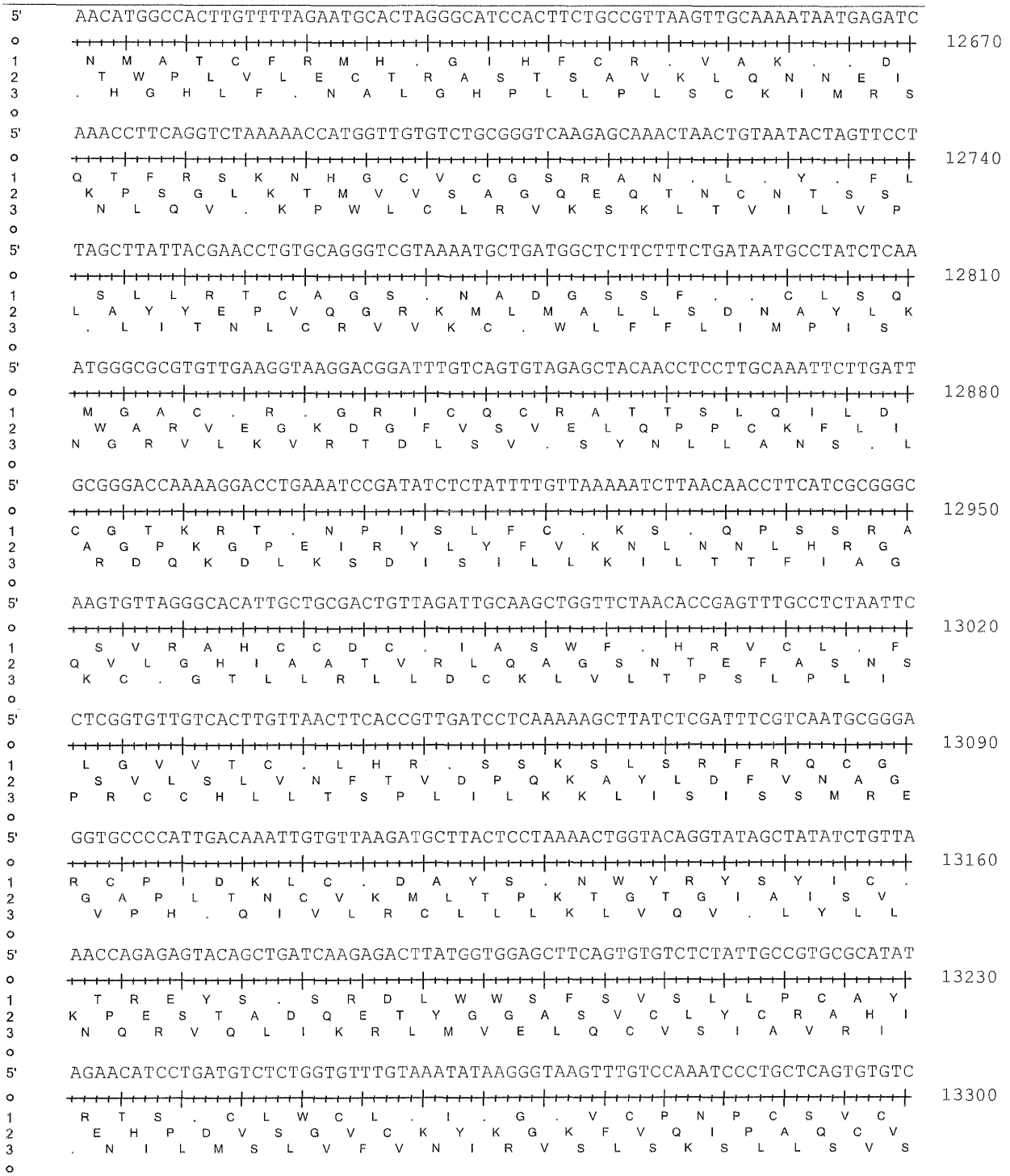
```
5'  TTTCTGTTACATGCATCAAATGGAACTTGCTAATGGTACACATACCGGTTCAGCATTTGATGGTACTATG
o  +----|----|----|----|----|----|----|----|----|----|----|----|----|----|  10570
1    F  L  L  H  A  S  N  G  T  C  .  W  Y  T  Y  R  F  S  I  .  W  Y  Y
2     F  C  Y  M  H  Q  M  E  L  A  N  G  T  H  T  G  S  A  F  D  G  T  M
3    I  S  V  T  C  I  K  W  N  L  L  M  V  H  I  P  V  Q  H  L  M  V  L  C
o
5'  TATGGTGCCTTTATGGATAAACAAGTGCACCAAGTTCAGTTAACAGACAAATACTGCAGTGTTAATGTAG
o  +----|----|----|----|----|----|----|----|----|----|----|----|----|----|  10640
1    V  W  C  L  Y  G  .  T  S  A  P  S  S  V  N  R  Q  I  L  Q  C  .  C  S
2     Y  G  A  F  M  D  K  Q  V  H  Q  V  Q  L  T  D  K  Y  C  S  V  N  V
3     M  V  P  L  W  I  N  K  C  T  K  F  S  .  Q  T  N  T  A  V  L  M  .
o
5'  TAGCTTGGCTTTACGCAGCAATACTTAATGGTTGCGCTTGGTTTGTAAAACCTAATCGCACTAGTGTTGT
o  +----|----|----|----|----|----|----|----|----|----|----|----|----|----|  10710
1    S  L  A  L  R  S  N  T  .  W  L  R  L  V  C  K  T  .  S  H  .  C  C
2    V  A  W  L  Y  A  A  I  L  N  G  C  A  W  F  V  K  P  N  R  T  S  V  V
3     .  L  G  F  T  Q  Q  Y  L  M  V  A  L  G  L  .  N  L  I  A  L  V  L
o
5'  TTCTTTTAATGAATGGGCTCTTGCCAACCAATTCACTGAATTTGTTGGCACTCAATCCGTTGACATGTTA
o  +----|----|----|----|----|----|----|----|----|----|----|----|----|----|  10780
1    F  F  .  .  M  G  S  C  Q  P  I  H  .  I  C  W  H  S  I  R  .  H  V
2     S  F  N  E  W  A  L  A  N  Q  F  T  E  F  V  G  T  Q  S  V  D  M  L
3    F  L  L  M  N  G  L  L  P  T  N  S  L  N  L  L  A  L  N  P  L  T  C  .
o
5'  GCTGTCAAAACAGGCGTTGCTATTGAACAGCTGCTTTATGCGATCCAACAACTGTATACTGGGTTCCAGG
o  +----|----|----|----|----|----|----|----|----|----|----|----|----|----|  10850
1    S  C  Q  N  R  R  C  Y  .  T  A  A  L  C  D  P  T  T  V  Y  W  V  P  G
2     A  V  K  T  G  V  A  I  E  Q  L  L  Y  A  I  Q  Q  L  Y  T  G  F  Q
3     L  S  K  Q  A  L  L  L  N  S  C  F  M  R  S  N  N  C  I  L  G  S  R
o
5'  GAAAGCAAATCCTTGGCAGTACCATGTTGGAAGATGAATTCACACCTGAGGATGTTAATATGCAGATTAT
o  +----|----|----|----|----|----|----|----|----|----|----|----|----|----|  10920
1    K  A  N  P  W  Q  Y  H  V  G  R  .  I  H  T  .  G  C  .  Y  A  D  Y
2    G  K  Q  I  L  G  S  T  M  L  E  D  E  F  T  P  E  D  V  N  M  Q  I  M
3     E  S  K  S  L  A  V  P  C  W  K  M  N  S  H  L  R  M  L  I  C  R  L
o
5'  GGGTGTGGTTATGCAGAGTGGTGTGAGAAAAGTTACATATGGTACTGCGCATTGGTTGTTTGCGACCCTT
o  +----|----|----|----|----|----|----|----|----|----|----|----|----|----|  10990
1    G  C  G  Y  A  E  W  C  E  K  S  Y  I  W  Y  C  A  L  V  V  C  D  P
2     G  V  V  M  Q  S  G  V  R  K  V  T  Y  G  T  A  H  W  L  F  A  T  L
3    W  V  W  L  C  R  V  V  .  E  K  L  H  M  V  L  R  I  G  C  L  R  P  L
o
5'  GTCTCAACCTATGTGATAATCTTACAAGCCACTAAATTTACTTTGTGGAACTACTTGTTTGAGACTATTC
o  +----|----|----|----|----|----|----|----|----|----|----|----|----|----|  11060
1    C  L  N  L  C  D  N  L  T  S  H  .  I  Y  F  V  E  L  L  V  .  D  Y  S
2     V  S  T  Y  V  I  I  L  Q  A  T  K  F  T  L  W  N  Y  L  F  E  T  I
3     S  Q  P  M  .  .  S  Y  K  P  L  N  L  L  C  G  T  T  C  L  R  L  F
o
5'  CCACACAGTTGTTCCCACTCTTATTTGTGACTATGGCCTTCGTTATGTTGTTGGTTAAACACAAACACAC
o  +----|----|----|----|----|----|----|----|----|----|----|----|----|----|  11130
1    H  T  V  V  P  T  L  I  C  D  Y  G  L  R  Y  V  V  G  .  T  Q  T  H
2    P  T  Q  L  F  P  L  L  F  V  T  M  A  F  V  M  L  L  V  K  H  K  H  T
3     P  H  S  C  S  H  S  Y  L  .  L  W  P  S  L  C  C  W  L  N  T  N  T
o
5'  CTTTTTGACACTTTTCTTGTTGCCTGTGGCTATTTGTTTGACTTATGCAAACATAGTCTACGAGCCCACT
o  +----|----|----|----|----|----|----|----|----|----|----|----|----|----|  11200
1    L  F  D  T  F  L  V  A  C  G  Y  L  F  D  L  C  K  H  S  L  R  A  H
2     F  L  T  L  F  L  L  P  V  A  I  C  L  T  Y  A  N  I  V  Y  E  P  T
3    P  F  .  H  F  S  C  C  L  W  L  F  V  .  L  M  Q  T  .  S  T  S  P  L
o
```


Sunday, September 23, 2012 1:41 PM
HCoV-SA1 translation 3 frames
5' ACTCCCATTTCGTCAGCGCTGATTGCAGTTGCAAATTGGCTTGCCCCCACTAATGCTTATATGCGCACTA
11270
Y S H F V S A D C S C K L A C P H . C L Y A H Y
T P I S S A L I A V A N W L A P T N A Y M R T
L P F R Q R . L Q L Q I G L P P L M L I C A L
5' CACATACTGATATTGGTGTCTACATTAGTATGTCACTTGTATTAGTCATTGTAGTGAAGAGATTGTACAA
11340
T Y . Y W C L H . Y V T C I S H C S E E I V Q
T H T D I G V Y I S M S L V L V I V V K R L Y N
H I L I L V S T L V C H L Y . S L . . R D C T
5' CCCATCACTTTCTAACTTTGCGTTAGCATTGTGCAGTGGTGTAATGTGGTTGTACACTTATAGCATTGGA
11410
P I T F . L C V S I V Q W C N V V V H L . H W
P S L S N F A L A L C S G V M W L Y T Y S I G
T H H F L T L R . H C A V V . C G C T L I A L E
5' GAAGCCTCAAGCCCCATTGCCTATCTGGTTTTTGTCACTACACTCACTAGTGATTATACGATTACAGTCT
11480
R S L K P H C L S G F C H Y T H . . L Y D Y S L
E A S S P I A Y L V F V T T L T S D Y T I T V
K P Q A P L P I W F L S L H S L V I I R L Q S
5' TTGTTACTGTCAACCTTGCAAAAGTTTGCACTTATGCCATCTTTGCTTACTCACCACAGCTTACACTTGT
11550
C Y C Q P C K S L H L C H L C L L T T A Y T C
F V T V N L A K V C T Y A I F A Y S P Q L T L V
L L L S T L Q K F A L M P S L L T H H S L H L
5' GTTTCCGGAAGTGAAGATGATACTTTTATTATACACATGTTTAGGTTTCATGTGTACTTGCTATTTTGGT
11620
V S G S E D D T F I I H M F R F H V Y L L F W
F P E V K M I L L L Y T C L G F M C T C Y F G
C F R K . R . Y F Y Y T H V . V S C V L A I L V
5' GTCTTCTCTCTTTTGAACCTTAAGCTTAGAGCACCTATGGGTGTCTATGACTTTAAGGTCTCAACACAAG
11690
C L L S F E P . A . S T Y G C L . L . G L N T R
V F S L L N L K L R A P M G V Y D F K V S T Q
S S L F . T L S L E H L W V S M T L R S Q H K
5' AGTTCAGATTCATGACTGCTAACAATCTAACTGCACCTAGAAATTCTTGGGAGGCTATGGCTCTGAACTT
11760
V Q I H D C . Q S N C T . K F L G G Y G S E L
E F R F M T A N N L T A P R N S W E A M A L N F
S S D S . L L T I . L H L E I L G R L W L . T
5' TAAGTTAATAGGTATTGGCGGTACACCTTGTATAAAGGTTGCTGCTATGCAGTCTAAACTTACAGATCTT
11830
. V N R Y W R Y T L Y K G C C Y A V . T Y R S
K L I G I G G T P C I K V A A M Q S K L T D L
L S . . V L A V H L V . R L L L C S L N L Q I L
5' AAATGCACATCTGTGGTTCTCCTCTCTGTGCTCCAACAGTTACACTTAGAGGCTAATAGTAGGGCCTGGG
11900
. M H I C G S P L C A P T V T L R G . . . G L G
K C T S V V L L S V L Q Q L H L E A N S R A W
N A H L W F S S L C S N S Y T . R L I V G P G

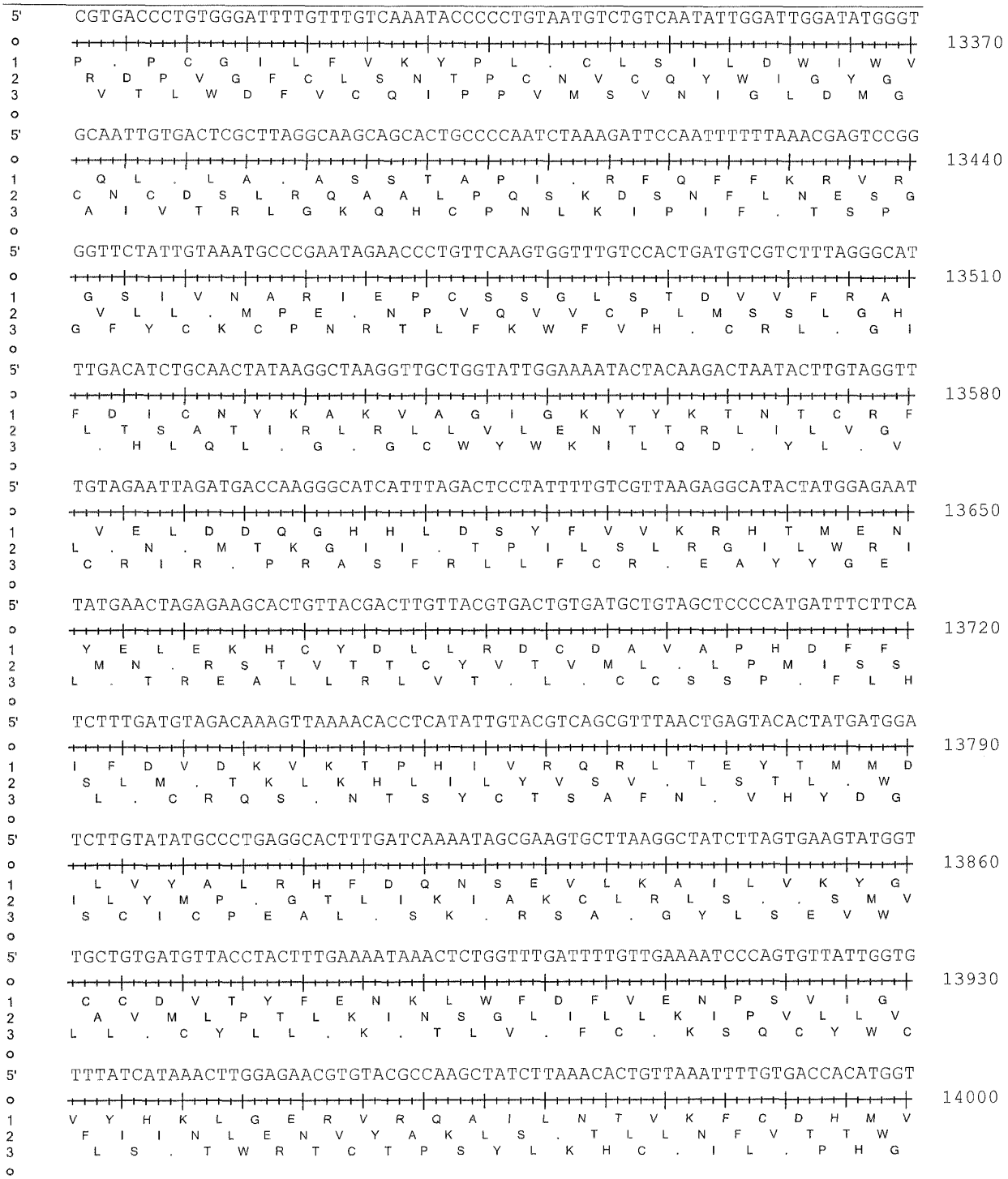
Sunday, September 23, 2012 1:41 PM
HCoV-SA1 translation 3 frames

```
5'  CTTTCTGTGTTAAATGCCATAATGATATATTGGCAGCAACAGACCCCAGTGAGGCTTTCGAGAAATTCGT
o  +----+----+----+----+----+----+----+----+----+----+----+----+----+----+  11970
1    F  L  C  .  M  P  .  .  Y  I  G  S  N  R  P  Q  .  G  F  R  E  I  R
2  A  F  C  V  K  C  H  N  D  I  L  A  A  T  D  P  S  E  A  F  E  K  F  V
3    L  S  V  L  N  A  I  M  I  Y  W  Q  Q  Q  T  P  V  R  L  S  R  N  S
o
5'  AAGTCTCTTTGCTACTTTAATGACTTTTTCTGGTAATGTAGATCTTGATGCGTTAGCTAGTGATATTTTT
o  +----+----+----+----+----+----+----+----+----+----+----+----+----+----+  12040
1    K  S  L  C  Y  F  N  D  F  F  W  .  C  R  S  .  C  V  S  .  .  Y  F
2      S  L  F  A  T  L  M  T  F  S  G  N  V  D  L  D  A  L  A  S  D  I  F
3    .  V  S  L  L  L  .  .  L  F  L  V  M  .  I  L  M  R  .  L  V  I  F  L
o
5'  GACACTCCTAGCGTACTTCAAGCTACTCTTTCTGAGTTTTCACACTTAGCTACCTTTGCTGAGTTGGAAG
o  +----+----+----+----+----+----+----+----+----+----+----+----+----+----+  12110
1    .  H  S  .  R  T  S  S  Y  S  F  .  V  F  T  L  S  Y  L  C  .  V  G  S
2      D  T  P  S  V  L  Q  A  T  L  S  E  F  S  H  L  A  T  F  A  E  L  E
3      T  L  L  A  Y  F  K  L  L  F  L  S  F  H  T  .  L  P  L  L  S  W  K
o
5'  CTGCGCAGAAAGCCTATCAGGAAGCTATGGACTCTGGTGACACCTCACCACAAGTTCTTAAGGCTTTGCA
o  +----+----+----+----+----+----+----+----+----+----+----+----+----+----+  12180
1      C  A  E  S  L  S  G  S  Y  G  L  W  .  H  L  T  T  S  S  .  G  F  A
2    A  A  Q  K  A  Y  Q  E  A  M  D  S  G  D  T  S  P  Q  V  L  K  A  L  Q
3      L  R  R  K  P  I  R  K  L  W  T  L  V  T  P  H  H  K  F  L  R  L  C
o
5'  GAAGGCTGTTAATATAGCTAAAAACGCCTATGAGAAGGATAAGGCAGTGGCCCGTAAGTTAGAACGTATG
o  +----+----+----+----+----+----+----+----+----+----+----+----+----+----+  12250
1      E  G  C  .  Y  S  .  K  R  L  .  E  G  .  G  S  G  P  .  V  R  T  Y
2        K  A  V  N  I  A  K  N  A  Y  E  K  D  K  A  V  A  R  K  L  E  R  M
3    R  R  L  L  I  .  L  K  T  P  M  R  R  I  R  Q  W  P  V  S  .  N  V  W
o
5'  GCTGATCAGGCTATGACTTCTATGTATAAGCAAGCACGTGCTGAAGACAAGAAAGCAAAAATTGTCAGTG
o  +----+----+----+----+----+----+----+----+----+----+----+----+----+----+  12320
1    G  .  S  G  Y  D  F  Y  V  .  A  S  T  C  .  R  Q  E  S  K  N  C  Q  C
2      A  D  Q  A  M  T  S  M  Y  K  Q  A  R  A  E  D  K  K  A  K  I  V  S
3        L  I  R  L  .  L  L  C  I  S  K  H  V  L  K  T  R  K  Q  K  L  S  V
o
5'  CTATGCAAACTATGTTGTTTGGTATGATTAAGAAGCTCGACAACGATGTTCTTAATGGTATCATTTCTAA
o  +----+----+----+----+----+----+----+----+----+----+----+----+----+----+  12390
1      Y  A  N  Y  V  V  W  Y  D  .  E  A  R  Q  R  C  S  .  W  Y  H  F  .
2    A  M  Q  T  M  L  F  G  M  I  K  K  L  D  N  D  V  L  N  G  I  I  S  N
3      L  C  K  L  C  C  L  V  .  L  R  S  S  T  T  M  F  L  M  V  S  F  L
o
5'  CGCTAGGAATGGTTGTATACCTCTTAGTGTCATCCCACTGTGTGCTTCAAATAAACTTCGCGTTGTAATT
o  +----+----+----+----+----+----+----+----+----+----+----+----+----+----+  12460
1    R  .  E  W  L  Y  T  S  .  C  H  P  T  V  C  F  K  .  T  S  R  C  N
2      A  R  N  G  C  I  P  L  S  V  I  P  L  C  A  S  N  K  L  R  V  V  I
3    T  L  G  M  V  V  Y  L  L  V  S  S  H  C  V  L  Q  I  N  F  A  L  .  F
o
5'  CCTGACTTCACCGTCTGGAATCAGGTAGTCACATATCCCTCGCTTAACTACGCTGGGGCTTTGTGGGACA
o  +----+----+----+----+----+----+----+----+----+----+----+----+----+----+  12530
1    S  .  L  H  R  L  E  S  G  S  H  I  S  L  A  .  L  R  W  G  F  V  G  H
2      P  D  F  T  V  W  N  Q  V  V  T  Y  P  S  L  N  Y  A  G  A  L  W  D
3        L  T  S  P  S  G  I  R  .  S  H  I  P  R  L  T  T  L  G  L  C  G  T
o
5'  TTACAGTTATAAACAATGTGGACAATGAAATTGTTAAGTCTTCAGATGTTGTAGACAGCAATGAAAATTT
o  +----+----+----+----+----+----+----+----+----+----+----+----+----+----+  12600
1      Y  S  Y  K  Q  C  G  Q  .  N  C  .  V  F  R  C  C  R  Q  Q  .  K  F
2    I  T  V  I  N  N  V  D  N  E  I  V  K  S  S  D  V  V  D  S  N  E  N  L
3      L  Q  L  .  T  M  W  T  M  K  L  L  S  L  Q  M  L  .  T  A  M  K  I
o
```

Sunday, September 23, 2012 1:41 PM
HCoV-SA1 translation 3 frames



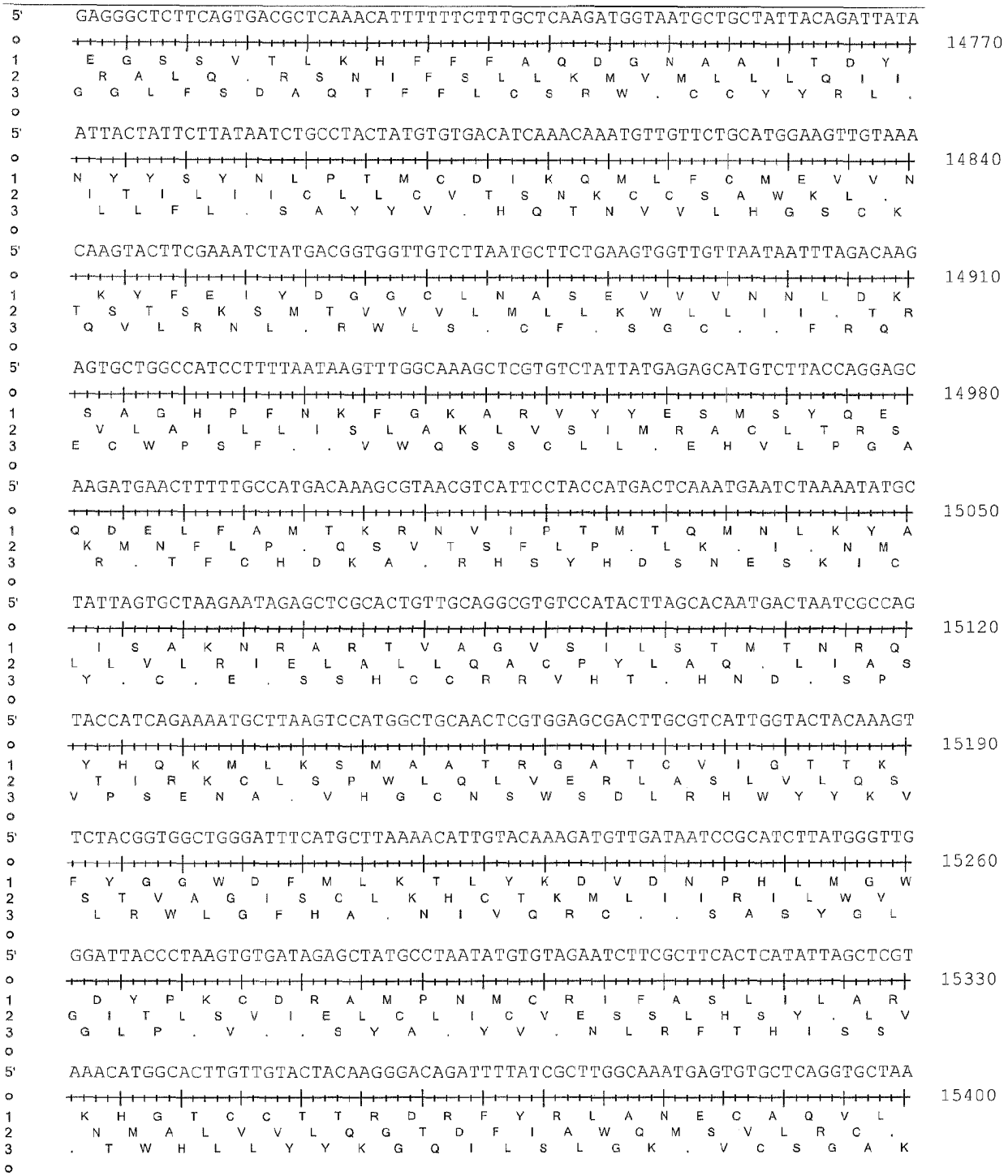
Sunday, September 23, 2012 1:41 PM
HCoV-SA1 translation 3 frames



Sunday, September 23, 2012 1:41 PM
HCoV-SA1 translation 3 frames

```
5'  CAAGGCTGGTTTAGTCGGTGTGCTCACACTAGACAACCAGGACCTTAATGGCAAGTGGTATGATTTTGGT
o   +++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+  14070
1     K  A  G  L  V  G  V  L  T  L  D  N  Q  D  L  N  G  K  W  Y  D  F  G
2   S  R  L  V  .  S  V  C  S  H  .  T  T  R  T  L  M  A  S  G  M  I  L  V
3     Q  G  W  F  S  R  C  A  H  T  R  Q  P  G  P  .  W  Q  V  V  .  F  W
o
5'  GACTTCGTAATCACTCAACCTGGTTCAGGAGTAGCTATAGTTGATAGCTACTATTCTTATTTGATGCCTG
o   +++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+  14140
1     D  F  V  I  T  Q  P  G  S  G  V  A  I  V  D  S  Y  Y  S  Y  L  M  P
2     T  S  .  S  L  N  L  V  Q  E  .  L  .  L  I  A  T  I  L  I  .  C  L
3   .  L  R  N  H  S  T  W  F  R  S  S  Y  S  .  .  L  L  F  L  F  D  A  C
o
5'  TGCTCTCAATGACCGATTGTCTGGCCGCTGAGACACATAGGGATTGTGATTTTAATAAACCACTCATTGA
o   +++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+  14210
1   V  L  S  M  T  D  C  L  A  A  E  T  H  R  D  C  D  F  N  K  P  L  I  E
2     C  S  Q  .  P  I  V  W  P  L  R  H  I  G  I  V  I  L  I  N  H  S  L
3     A  L  N  D  R  L  S  G  R  .  D  T  .  G  L  .  F  .  T  T  H  .
o
5'  GTGGCCACTTACTGAGTATGATTTTACTGATTATAAGGTACAACTCTTTGAGAAGTACTTTAAATATTGG
o   +++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+  14280
1     W  P  L  T  E  Y  D  F  T  D  Y  K  V  Q  L  F  E  K  Y  F  K  Y  W
2   S  G  H  L  L  S  M  I  L  L  I  I  R  Y  N  S  L  R  S  T  L  N  I  G
3     V  A  T  Y  .  V  .  F  Y  .  L  .  G  T  T  L  .  E  V  L  .  I  L
o
5'  GATCAGACGTATCACGCAAATTGCGTTAATTGTACTGATGACCGTTGTGTGTTACATTGTGCTAATTTCA
o   +++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+  14350
1     D  Q  T  Y  H  A  N  C  V  N  C  T  D  D  R  C  V  L  H  C  A  N  F
2     I  R  R  I  T  Q  I  A  L  I  V  L  M  T  V  V  C  Y  I  V  L  I  S
3   G  S  D  V  S  R  K  L  R  .  L  Y  .  P  L  C  V  T  L  C  .  F  Q
o
5'  ATGTATTGTTTGCTATGACCATGCCTAAGACTTGTTTCGGACCCATAGTCCGAAAGATCTTTGTTGATGG
o   +++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+  14420
1   N  V  L  F  A  M  T  M  P  K  T  C  F  G  P  I  V  R  K  I  F  V  D  G
2     M  Y  C  L  L  .  P  C  L  R  L  V  S  D  P  .  S  E  R  S  L  L  M
3     C  I  V  C  Y  D  H  A  .  D  L  F  R  T  H  S  P  K  D  L  C  .  W
o
5'  CGTGCCATTTGTAGTATCTTGTGGTTATCACTACAAAGAATTAGGTTTAGTCATGAATATGGATGTTAGT
o   +++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+  14490
1     V  P  F  V  V  S  C  G  Y  H  Y  K  E  L  G  L  V  M  N  M  D  V  S
2   A  C  H  L  .  Y  L  V  V  I  T  T  K  N  .  V  .  S  .  I  W  M  L  V
3     R  A  I  C  S  I  L  W  L  S  L  Q  R  I  R  F  S  H  E  Y  G  C  .
o
5'  CTCCATAGACATAGGCTCTCTCTTAAGGAGTTGATGATGTATGCCGCTGATCCAGCCATGCACATTGCCT
o   +++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+  14560
1   L  H  R  H  R  L  S  L  K  E  L  M  M  Y  A  A  D  P  A  M  H  I  A
2     S  I  D  I  G  S  L  L  R  S  .  .  C  M  P  L  I  Q  P  C  T  L  P
3   S  P  .  T  .  A  L  S  .  G  V  D  D  V  C  R  .  S  S  H  A  H  C  L
o
5'  CCTCTAACGCTTTTCTTGATTTGAGGACATCATGTTTTAGTGTCGCTGCACTTACAACTGGTTTGACTTT
o   +++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+  14630
1   S  S  N  A  F  L  D  L  R  T  S  C  F  S  V  A  A  L  T  T  G  L  T  F
2     P  L  T  L  F  L  I  .  G  H  H  V  L  V  S  L  H  L  Q  L  V  .  L
3     L  .  R  F  S  .  F  E  D  I  M  F  .  C  R  C  T  Y  N  W  F  D  F
o
5'  TCAAACTGTGCGGCCTGGCAATTTTAACCAAGACTTCTATGATTTCGTGGTATCTAAAGGTTTCTTTAAG
o   +++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+  14700
1     Q  T  V  R  P  G  N  F  N  Q  D  F  Y  D  F  V  V  S  K  G  F  F  K
2   F  K  L  C  G  L  A  I  L  T  K  T  S  M  I  S  W  Y  L  K  V  S  L  R
3     S  N  C  A  A  W  Q  F  .  P  R  L  L  .  F  R  G  I  .  R  F  L  .
o
```

Sunday, September 23, 2012 1:41 PM
HCoV-SA1 translation 3 frames



Sunday, September 23, 2012 1:41 PM
HCoV-SA1 translation 3 frames

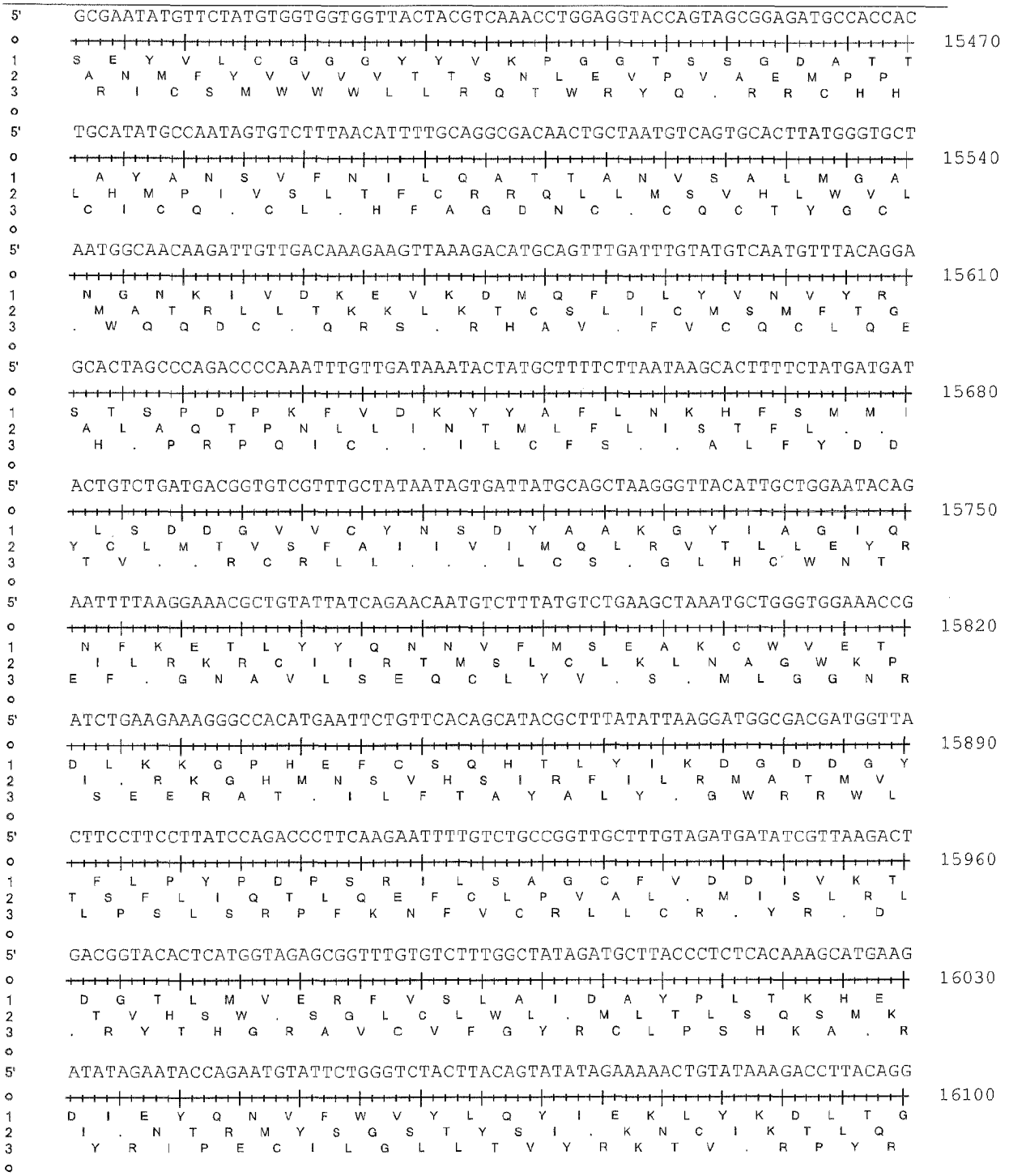
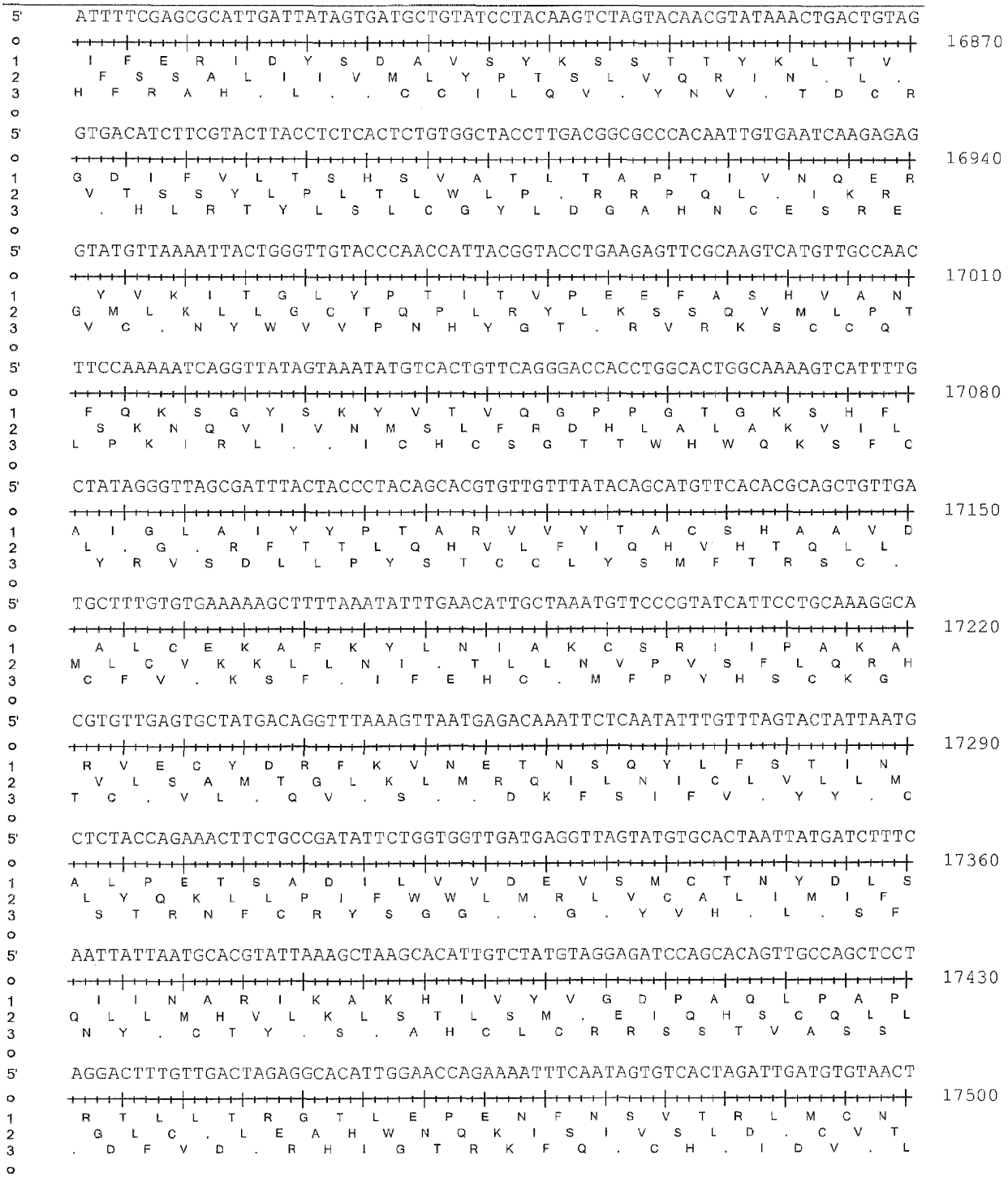


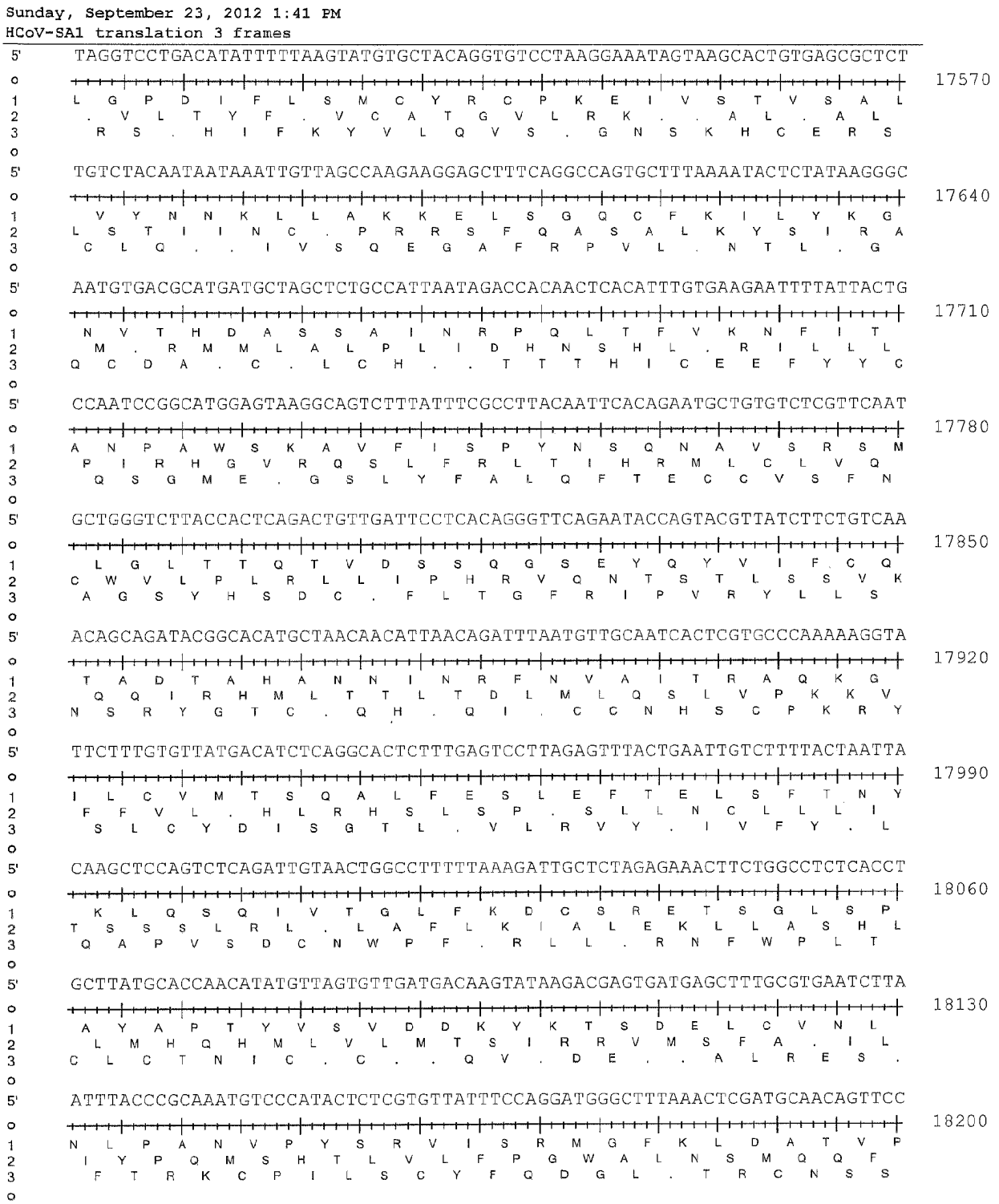
Figure 15 continued

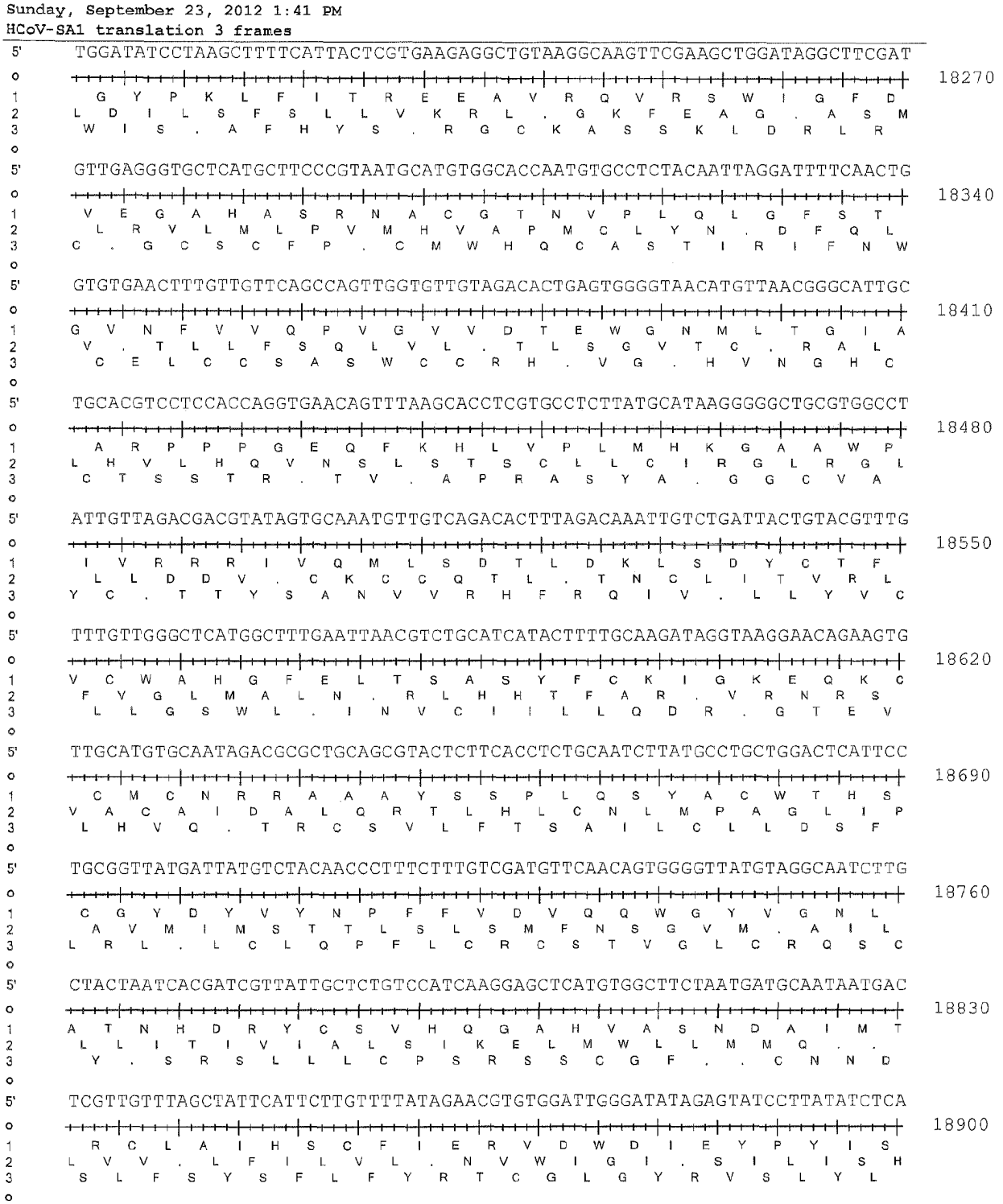
Sunday, September 23, 2012 1:41 PM
HCoV-SA1 translation 3 frames

```
5'  ACACATGCTTGACAGTTATTCTGTCATGCTATGTGGTGATAATTCTGCTAAGTTTTGGGAAGAGGCATTC
o  +---------+---------+---------+---------+---------+---------+---------+  16170
1    H  M  L  D  S  Y  S  V  M  L  C  G  D  N  S  A  K  F  W  E  E  A  F
2  D  T  C  L  T  V  I  L  S  C  Y  V  V  I  I  L  L  S  F  G  K  R  H  S
3    T  H  A  .  Q  L  F  C  H  A  M  W  .  .  F  C  .  V  L  G  R  G  I
o
5'  TATAGAGATCTCTATAGTTCGCCTACCACTTTGCAGGCTGTCGGTTCATGCGTTGTATGCCATTCACAGA
o  +---------+---------+---------+---------+---------+---------+---------+  16240
1    Y  R  D  L  Y  S  S  P  T  T  L  Q  A  V  G  S  C  V  V  C  H  S  Q
2    I  E  I  S  I  V  R  L  P  L  C  R  L  S  V  H  A  L  Y  A  I  H  R
3  L  .  R  S  L  .  F  A  Y  H  F  A  G  C  R  F  M  R  C  M  P  F  T  D
o
5'  CTTCCCTACGCTGTGGGACATGCATCCGTAGACCATTTCTCTGCTGTAAATGCTGCTATGATCATGTTAT
o  +---------+---------+---------+---------+---------+---------+---------+  16310
1  T  S  L  R  C  G  T  C  I  R  R  P  F  L  C  C  K  C  C  Y  D  H  V  I
2    L  P  Y  A  V  G  H  A  S  V  D  H  F  S  A  V  N  A  A  M  I  M  L
3    F  P  T  L  W  D  M  H  P  .  T  I  S  L  L  .  M  L  L  .  S  C  Y
o
5'  AGCAACTCCACATAAGATGGTTTTGTCTGTTTCTCCTTACGTTTGTAATGCCCCTGGTTGTGGCGTTTCA
o  +---------+---------+---------+---------+---------+---------+---------+  16380
1    A  T  P  H  K  M  V  L  S  V  S  P  Y  V  C  N  A  P  G  C  G  V  S
2  .  Q  L  H  I  R  W  F  C  L  F  L  L  T  F  V  M  P  L  V  V  A  F  Q
3    S  N  S  T  .  D  G  F  V  C  F  S  L  R  L  .  C  P  W  L  W  R  F
o
5'  GACGTTACTAAGCTATATTTAGGTGGTATGAGCTACTTTTGTGTAGATCATAGACCTGTGTGTAGTTTTC
o  +---------+---------+---------+---------+---------+---------+---------+  16450
1    D  V  T  K  L  Y  L  G  G  M  S  Y  F  C  V  D  H  R  P  V  C  S  F
2    T  L  L  S  Y  I  .  V  V  .  A  T  F  V  .  I  I  D  L  C  V  V  F
3  R  R  Y  .  A  I  F  R  W  Y  E  L  L  L  C  R  S  .  T  C  V  .  F  S
o
5'  CACTTTGCGCTAATGGTCTTGTATTCGGCTTATACAAGAATATGTGCACAGGTAGTCCTTCTATAGTTGA
o  +---------+---------+---------+---------+---------+---------+---------+  16520
1  P  L  C  A  N  G  L  V  F  G  L  Y  K  N  M  C  T  G  S  P  S  I  V  E
2    H  F  A  L  M  V  L  Y  S  A  Y  T  R  I  C  A  Q  V  V  L  L  .  L
3    T  L  R  .  W  S  C  I  R  L  I  Q  E  Y  V  H  R  .  S  F  Y  S  .
o
5'  ATTTAATAGGTTGGCTACCTGTGACTGGACTGAAAGTGGTGATTACACCCTTGCCAATACTACAACAGAA
o  +---------+---------+---------+---------+---------+---------+---------+  16590
1    F  N  R  L  A  T  C  D  W  T  E  S  G  D  Y  T  L  A  N  T  T  T  E
2  N  L  I  G  W  L  P  V  T  G  L  K  V  V  I  T  P  L  P  I  L  Q  Q  N
3    I  .  .  V  G  Y  L  .  L  D  .  K  W  .  L  H  P  C  Q  Y  Y  N  R
o
5'  CCACTCAAACTTTTTGCTGCTGAGACTTTACGTGCCACTGAAGAGGCGTCTAAGCAGTCTTATGCTATTG
o  +---------+---------+---------+---------+---------+---------+---------+  16660
1    P  L  K  L  F  A  A  E  T  L  R  A  T  E  E  A  S  K  Q  S  Y  A  I
2    H  S  N  F  L  L  L  R  L  Y  V  P  L  K  R  R  L  S  S  L  M  L  L
3  T  T  Q  T  F  C  C  .  D  F  T  C  H  .  R  G  V  .  A  V  L  C  Y  C
o
5'  CCACCATCAAAGAAATTGTTGGTGAGCGCCAACTATTACTTGTGTGGGAGGCTGGCAAGTCCAAACCACC
o  +---------+---------+---------+---------+---------+---------+---------+  16730
1  A  T  I  K  E  I  V  G  E  R  Q  L  L  L  V  W  E  A  G  K  S  K  P  P
2    P  P  S  K  K  L  L  V  S  A  N  Y  Y  L  C  G  R  L  A  S  P  N  H
3    H  H  Q  R  N  C  W  .  A  P  T  I  T  C  V  G  G  W  Q  V  Q  T  T
o
5'  ACTCAATCGTAATTATGTTTTTACTGGTTATCATATAACCAAAAATAGTAAAGTGCAGCTCGGTGAGTAC
o  +---------+---------+---------+---------+---------+---------+---------+  16800
1    L  N  R  N  Y  V  F  T  G  Y  H  I  T  K  N  S  K  V  Q  L  G  E  Y
2  H  S  I  V  I  M  F  L  L  V  I  I  .  P  K  I  V  K  C  S  S  V  S  T
3    T  Q  S  .  L  C  F  Y  W  L  S  Y  N  Q  K  .  .  S  A  A  R  .  V
o
```

Sunday, September 23, 2012 1:41 PM
HCoV-SA1 translation 3 frames




Sunday, September 23, 2012 1:41 PM
HCoV-SA1 translation 3 frames
5' TAGGTCCTGACATATTTTTAAGTATGTGCTACAGGTGTCCTAAGGAAATAGTAAGCACTGTGAGCGCTCT
17570
L G P D I F L S M C Y R C P K E I V S T V S A L
. V L T Y F . V C A T G V L R K . . A L . A L
R S . H I F K Y V L Q V S . G N S K H C E R S
5' TGTCTACAATAATAAATTGTTAGCCAAGAAGGAGCTTTCAGGCCAGTGCTTTAAAATACTCTATAAGGGC
17640
V Y N N K L L A K K E L S G Q C F K I L Y K G
L S T I I N C . P R R S F Q A S A L K Y S I R A
C L Q . . I V S Q E G A F R P V L . N T L . G
5' AATGTGACGCATGATGCTAGCTCTGCCATTAATAGACCACAACTCACATTTGTGAAGAATTTTATTACTG
17710
N V T H D A S S A I N R P Q L T F V K N F I T
M . R M M L A L P L I D H N S H L . R I L L L
Q C D A . C . L C H . . T T T H I C E E F Y Y C
5' CCAATCCGGCATGGAGTAAGGCAGTCTTTATTTCGCCTTACAATTCACAGAATGCTGTGTCTCGTTCAAT
17780
A N P A W S K A V F I S P Y N S Q N A V S R S M
P I R H G V R Q S L F R L T I H R M L C L V Q
Q S G M E . G S L Y F A L Q F T E C C V S F N
5' GCTGGGTCTTACCACTCAGACTGTTGATTCCTCACAGGGTTCAGAATACCAGTACGTTATCTTCTGTCAA
17850
L G L T T Q T V D S S Q G S E Y Q Y V I F C Q
C W V L P L R L L I P H R V Q N T S T L S S V K
A G S Y H S D C . F L T G F R I P V R Y L L S
5' ACAGCAGATACGGCACATGCTAACAACATTAACAGATTTAATGTTGCAATCACTCGTGCCCAAAAAGGTA
17920
T A D T A H A N N I N R F N V A I T R A Q K G
Q Q I R H M L T T L T D L M L Q S L V P K K V
N S R Y G T C . Q H . Q I . C C N H S C P K R Y
5' TTCTTTGTGTTATGACATCTCAGGCACTCTTTGAGTCCTTAGAGTTTACTGAATTGTCTTTTACTAATTA
17990
I L C V M T S Q A L F E S L E F T E L S F T N Y
F F V L . H L R H S L S P . S L L N C L L L I
S L C Y D I S G T L . V L R V Y . I V F Y . L
5' CAAGCTCCAGTCTCAGATTGTAACTGGCCTTTTTAAAGATTGCTCTAGAGAAACTTCTGGCCTCTCACCT
18060
K L Q S Q I V T G L F K D C S R E T S G L S P
T S S S L R L . L A F L K I A L E K L L A S H L
Q A P V S D C N W P F . R L L . R N F W P L T
5' GCTTATGCACCAACATATGTTAGTGTTGATGACAAGTATAAGACGAGTGATGAGCTTTGCGTGAATCTTA
18130
A Y A P T Y V S V D D K Y K T S D E L C V N L
L M H Q H M L V L M T S I R R V M S F A . I L
C L C T N I C . C . . Q V . D E . . A L R E S .
5' ATTTACCCGCAAATGTCCCATACTCTCGTGTTATTTCCAGGATGGGCTTTAAACTCGATGCAACAGTTCC
18200
N L P A N V P Y S R V I S R M G F K L D A T V P
I Y P Q M S H T L V L F P G W A L N S M Q Q F
F T R K C P I L S C Y F Q D G L . T R C N S S


Sunday, September 23, 2012 1:41 PM
HCoV-SA1 translation 3 frames
5' TGGATATCCTAAGCTTTTCATTACTCGTGAAGAGGCTGTAAGGCAAGTTCGAAGCTGGATAGGCTTCGAT
18270
G Y P K L F I T R E E A V R Q V R S W I G F D
L D I L S F S L L V K R L . G K F E A G . A S M
W I S . A F H Y S . R G C K A S S K L D R L R
5' GTTGAGGGTGCTCATGCTTCCCGTAATGCATGTGGCACCAATGTGCCTCTACAATTAGGATTTTCAACTG
18340
V E G A H A S R N A C G T N V P L Q L G F S T
L R V L M L P V M H V A P M C L Y N . D F Q L
C . G C S C F P . C M W H Q C A S T I R I F N W
5' GTGTGAACTTTGTTGTTCAGCCAGTTGGTGTTGTAGACACTGAGTGGGGTAACATGTTAACGGGCATTGC
18410
G V N F V V Q P V G V V D T E W G N M L T G I A
V . T L L F S Q L V L . T L S G V T C . R A L
C E L C C S A S W C C R H . V G . H V N G H C
5' TGCACGTCCTCCACCAGGTGAACAGTTTAAGCACCTCGTGCCTCTTATGCATAAGGGGGCTGCGTGGCCT
18480
A R P P P P G E Q F K H L V P L M H K G A A W P
L H V L H Q V N S L S T S C L L C I R G L R G L
C T S S T R . T V . A P R A S Y A . G G C V A
5' ATTGTTAGACGACGTATAGTGCAAATGTTGTCAGACACTTTAGACAAATTGTCTGATTACTGTACGTTTG
18550
I V R R R I V Q M L S D T L D K L S D Y C T F
L L D D V . C K C C Q T L . T N C L I T V R L
Y C . T T Y S A N V V R H F R Q I V . L L Y V C
5' TTTGTTGGGCTCATGGCTTTGAATTAACGTCTGCATCATACTTTTGCAAGATAGGTAAGGAACAGAAGTG
18620
V C W A H G F E L T S A S Y F C K I G K E Q K C
F Y G L M A L N . R L H H T F A R . V R N R S
L L G S W L . I N V C I I L L Q D R . G T E V
5' TTGCATGTGCAATAGACGCGCTGCAGCGTACTCTTCACCTCTGCAATCTTATGCCTGCTGGACTCATTCC
18690
C M C N R R A A A Y S S P L Q S Y A C W T H S
V A C A I D A L Q R T L H L C N L M P A G L I P
L H V Q . T R C S V L F T S A I L C L L D S F
5' TGCGGTTATGATTATGTCTACAACCCTTTCTTTGTCGATGTTCAACAGTGGGGTTATGTAGGCAATCTTG
18760
C G Y D Y V Y N P F F V D V Q Q W G Y V G N L
A V M I M S T T L S L S M F N S G V M . A I L
L R L . L C L Q P F L C R C S T V G L C R Q S C
5' CTACTAATCACGATCGTTATTGCTCTGTCCATCAAGGAGCTCATGTGGCTTCTAATGATGCAATAATGAC
18830
A T N H D R Y C S V H Q G A H V A S N D A I M T
L L I T I V I A L S I K E L M W L L M M Q . .
Y . S R S L L L C P S R S S C G F . . C N N D
5' TCGTTGTTTAGCTATTCATTCTTGTTTTATAGAACGTGTGGATTGGGATATAGAGTATCCTTATATCTCA
18900
R C L A I H S C F I E R V D W D I E Y P Y I S
L V V . L F I L V L . N V W I G I . S I L I S H
S L F S Y S F L F Y R T C G L G Y R V S L Y L

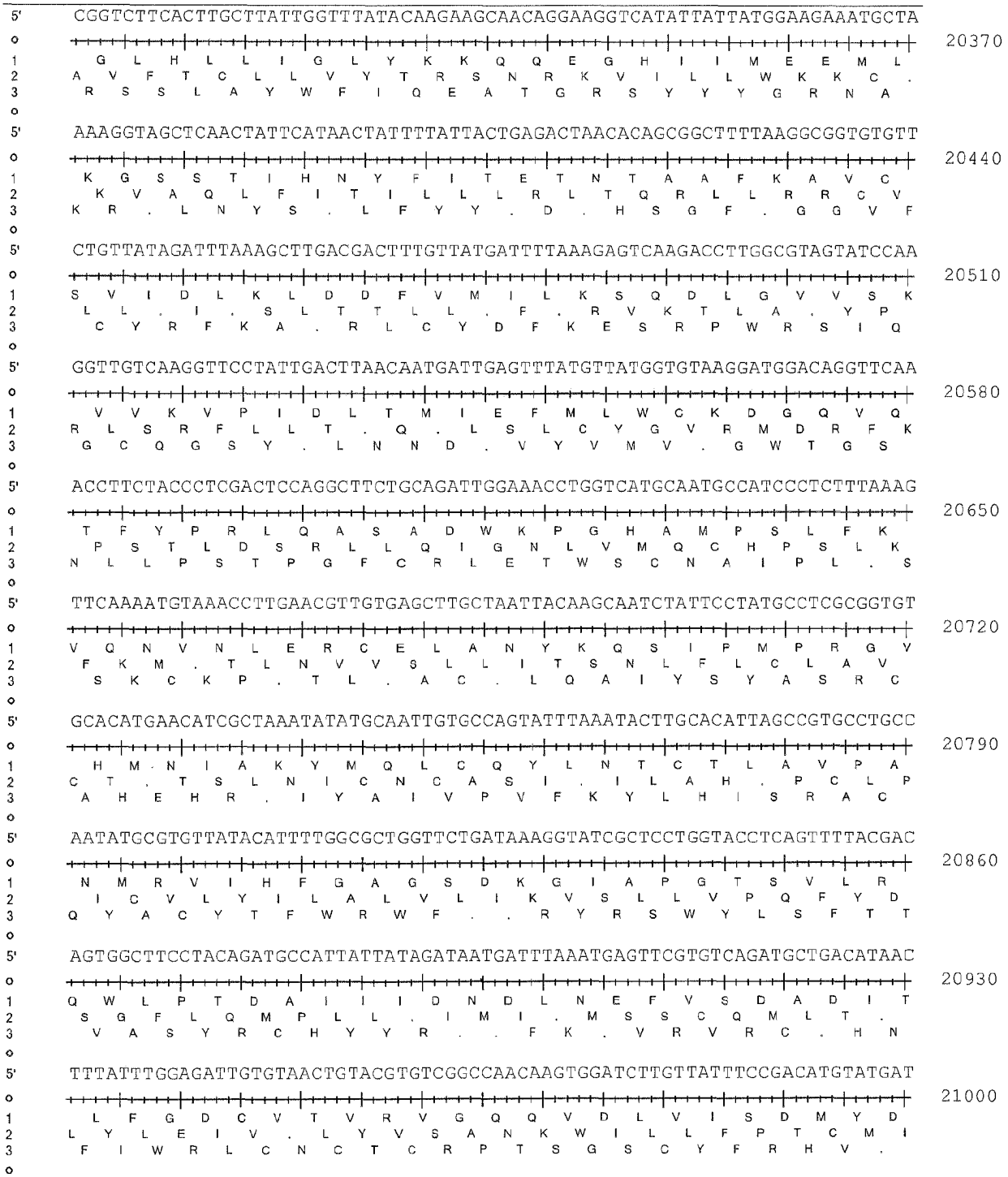
Sunday, September 23, 2012 1:41 PM
HCoV-SA1 translation 3 frames

```
5'  CATGAAAAGAAATTGAATTCCTGTTGTAGAATCGTTGAGCGCAACGTCGTACGTGCTGCTCTTCTTGCCG
o  +---------+---------+---------+---------+---------+---------+---------+ 18970
1    H  E  K  K  L  N  S  C  C  R  I  V  E  R  N  V  V  R  A  A  L  L  A
2     M  K  R  N  .  I  P  V  V  E  S  L  S  A  T  S  Y  V  L  L  F  L  P
3    T  .  K  E  I  E  F  L  L  .  N  R  .  A  Q  R  R  T  C  C  S  S  C  R
o
5'  GTTCATTTGACAAAGTCTATGATATTGGCAATCCTAAAGGAATTCCTATTGTTGATGACCCTGTGGTTGA
o  +---------+---------+---------+---------+---------+---------+---------+ 19040
1    G  S  F  D  K  V  Y  D  I  G  N  P  K  G  I  P  I  V  D  D  P  V  V  D
2     V  H  L  T  K  S  M  I  L  A  I  L  K  E  F  L  L  L  M  T  L  W  L
3     F  I  .  Q  S  L  .  Y  W  Q  S  .  R  N  S  Y  C  .  .  P  C  G  .
o
5'  TTGGCATTATTTTGATGCACAGCCCTTGACCAGGAAGGTACAACAGCTTTTCTATACAGAGGACATGGCC
o  +---------+---------+---------+---------+---------+---------+---------+ 19110
1    W  H  Y  F  D  A  Q  P  L  T  R  K  V  Q  Q  L  F  Y  T  E  D  M  A
2    I  G  I  I  L  M  H  S  P  .  P  G  R  Y  N  S  F  S  I  Q  R  T  W  P
3     L  A  L  F  .  C  T  A  L  D  Q  E  G  T  T  A  F  L  Y  R  G  H  G
o
5'  TCAAGATTTGCTGATGGGCTCTGCTTATTTTGGAACTGTAATGTACCAAAATATCCTAATAATGCAATTG
o  +---------+---------+---------+---------+---------+---------+---------+ 19180
1    S  R  F  A  D  G  L  C  L  F  W  N  C  N  V  P  K  Y  P  N  N  A  I
2     Q  D  L  L  M  G  S  A  Y  F  G  T  V  M  Y  Q  N  I  L  I  M  Q  L
3    L  K  I  C  .  W  A  L  L  I  L  E  L  .  C  T  K  I  S  .  .  C  N  C
o
5'  TATGCAGGTTTGACACACGTGTGCATTCTGAGTTCAATTTGCCAGGTTGTGATGGCGGTAGTTTGTATGT
o  +---------+---------+---------+---------+---------+---------+---------+ 19250
1    V  C  R  F  D  T  R  V  H  S  E  F  N  L  P  G  C  D  G  G  S  L  Y  V
2     Y  A  G  L  T  H  V  C  I  L  S  S  I  C  Q  V  V  M  A  V  V  C  M
3      M  Q  V  .  H  T  C  A  F  .  V  Q  F  A  R  L  .  W  R  .  F  V  C
o
5'  TAACAAGCACGCTTTTCATACACCAGCATATGATGTGAGTGCATTCCGTGATCTGAAACCTTTACCATTC
o  +---------+---------+---------+---------+---------+---------+---------+ 19320
1    N  K  H  A  F  H  T  P  A  Y  D  V  S  A  F  R  D  L  K  P  L  P  F
2    L  T  S  T  L  F  I  H  Q  H  M  M  .  V  H  S  V  I  .  N  L  Y  H  S
3     .  Q  A  R  F  S  Y  T  S  I  .  C  E  C  I  P  .  S  E  T  F  T  I
o
5'  TTTTATTATTCTACTACACCATGTGAAGTGCATGGTAATGGTAGTATGATAGAGGATATTGATTATGTAC
o  +---------+---------+---------+---------+---------+---------+---------+ 19390
1    F  Y  Y  S  T  T  P  C  E  V  H  G  N  G  S  M  I  E  D  I  D  Y  V
2     F  I  I  L  L  H  H  V  K  C  M  V  M  V  V  .  .  R  I  L  I  M  Y
3    L  L  L  F  Y  Y  T  M  .  S  A  W  .  W  .  Y  D  R  G  Y  .  L  C  T
o
5'  CCCTAAAATCTGCAGTCTGTATTACAGCTTGTAATTTAGGGGGCGCTGTTTGTAGGAAGCATGCTACAGA
o  +---------+---------+---------+---------+---------+---------+---------+ 19460
1    P  L  K  S  A  V  C  I  T  A  C  N  L  G  G  A  V  C  R  K  H  A  T  E
2     P  .  N  L  Q  S  V  L  Q  L  V  I  .  G  A  L  F  V  G  S  M  L  Q
3      P  K  I  C  S  L  Y  Y  S  L  .  F  R  G  R  C  L  .  E  A  C  Y  R
o
5'  GTACAGAGAGTATATGGAAGCATATAATCTTGTCTCTGCATCAGGTTTCCGCCTTTGGTGTTATAAGACC
o  +---------+---------+---------+---------+---------+---------+---------+ 19530
1    Y  R  E  Y  M  E  A  Y  N  L  V  S  A  S  G  F  R  L  W  C  Y  K  T
2    S  T  E  S  I  W  K  H  I  I  L  S  L  H  Q  V  S  A  F  G  V  I  R  P
3     V  Q  R  V  Y  G  S  I  .  S  C  L  C  I  R  F  P  P  L  V  L  .  D
o
5'  TTTGATATTTATAATCTCTGGTCTACTTTTACAAAAGTTCAAGGTTTGGAAAACATTGCTTTTAATGTTG
o  +---------+---------+---------+---------+---------+---------+---------+ 19600
1    F  D  I  Y  N  L  W  S  T  F  T  K  V  Q  G  L  E  N  I  A  F  N  V
2     L  I  F  I  I  S  G  L  L  L  Q  K  F  K  V  W  K  T  L  L  L  M  L
3    L  .  Y  L  .  S  L  V  Y  F  Y  K  S  S  R  F  G  K  H  C  F  .  C  C
o
```

Sunday, September 23, 2012 1:41 PM
HCoV-SA1 translation 3 frames

```
5'  TTAAACAAGGCCATTTTATTGGTGTTGAGGGTGAACTACCTGTAGCTGTAGTCAATGATAAGATCTTCAC
o   +++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++ 19670
1   V  K  Q  G  H  F  I  G  V  E  G  E  L  P  V  A  V  V  N  D  K  I  F  T
2    L  N  K  A  I  L  L  V  L  R  V  N  Y  L  .  L  .  S  M  I  R  S  S
3     .  T  R  P  F  Y  W  C  .  G  .  T  T  C  S  C  S  Q  .  .  D  L  H
o
5'  CAAGAGTGGCGTTAATGACATTTGTATGTTTGAGAATAAAACCACTTTGCCTACTAATATAGCTTTTGAA
o   +++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++ 19740
1    K  S  G  V  N  D  I  C  M  F  E  N  K  T  T  L  P  T  N  I  A  F  E
2   P  R  V  A  L  M  T  F  V  C  L  R  I  K  P  L  C  L  L  I  .  L  L  N
3    Q  E  W  R  .  .  H  L  Y  V  .  E  .  N  H  F  A  Y  .  Y  S  F  .
o
5'  CTCTATGCTAAGCGTGCTGTACGCTCGCATCCCGATTTCAAATTGCTACACAATTTACAAGCAGACATTT
o   +++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++ 19810
1   L  Y  A  K  R  A  V  R  S  H  P  D  F  K  L  L  H  N  L  Q  A  D  I
2    S  M  L  S  V  L  Y  A  R  I  P  I  S  N  C  Y  T  I  Y  K  Q  T  F
3   T  L  C  .  A  C  C  T  L  A  S  R  F  Q  I  A  T  Q  F  T  S  R  H  L
o
5'  GCTACAAGTTCGTCCTTTGGGATTATGAACGTAGCAATATTTATGGTACTGCTACTATTGGTGTATGTAA
o   +++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++ 19880
1   C  Y  K  F  V  L  W  D  Y  E  R  S  N  I  Y  G  T  A  T  I  G  V  C  K
2    A  T  S  S  S  F  G  I  M  N  V  A  I  F  M  V  L  L  L  L  V  Y  V
3     L  Q  V  R  P  L  G  L  .  T  .  Q  Y  L  W  Y  C  Y  Y  W  C  M  .
o
5'  GTACACTGATATTGATGTTAATTCAGCTTTGAATATATGTTTTGACATACGCGATAATTGTTCATTGGAG
o   +++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++ 19950
1    Y  T  D  I  D  V  N  S  A  L  N  I  C  F  D  I  R  D  N  C  S  L  E
2   S  T  L  I  L  M  L  I  Q  L  .  I  Y  V  L  T  Y  A  I  I  V  H  W  R
3    V  H  .  Y  .  C  .  F  S  F  E  Y  M  F  .  H  T  R  .  L  F  I  G
o
5'  AAGTTCATGTCTACTCCCAATGCCATCTTTATTTCTGATAGAAAAATCAAGAAATACCCTTGTATGGTAG
o   +++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++ 20020
1    K  F  M  S  T  P  N  A  I  F  I  S  D  R  K  I  K  K  Y  P  C  M  V
2     S  S  C  L  L  P  M  P  S  L  F  L  I  E  K  S  R  N  T  L  V  W  .
3   E  V  H  V  Y  S  Q  C  H  L  Y  F  .  .  K  N  Q  E  I  P  L  Y  G  R
o
5'  GTCCTGATTATGCTTACTTCAATGGTGCTATCATCCGTGATAGTGATGTTGTTAAACAACCAGTGAAGTT
o   +++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++ 20090
1   G  P  D  Y  A  Y  F  N  G  A  I  I  R  D  S  D  V  V  K  Q  P  V  K  F
2    V  L  I  M  L  T  S  M  V  L  S  S  V  I  V  M  L  L  N  N  Q  .  S
3     S  .  L  C  L  L  Q  W  C  Y  H  P  .  .  .  C  C  .  T  T  S  E  V
o
5'  CTACTTGTATAAGAAAGTCAATAATGAGTTTATTGATCCTACTGAGTGTATTTACACTCAGAGTCGCTCT
o   +++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++ 20160
1    .  Y  L  Y  K  K  V  N  N  E  F  I  D  P  T  E  C  I  Y  T  Q  S  R  S
2   S  T  C  I  R  K  S  I  M  S  L  L  I  L  L  S  V  F  T  L  R  V  A  L
3    L  L  V  .  E  S  Q  .  .  V  Y  .  S  Y  .  V  Y  L  H  S  E  S  L
o
5'  TGTAGTGACTTCCTACCCCTTTCTGACATGGAGAAAGACTTTCTATCTTTTGATAGTGATGTTTTCATTA
o   +++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++ 20230
1   C  S  D  F  L  P  L  S  D  M  E  K  D  F  L  S  F  D  S  D  V  F  I
2    V  V  T  S  Y  P  F  L  T  W  R  K  T  F  Y  L  L  I  V  M  F  S  L
3   L  .  .  L  P  T  P  F  .  H  G  E  R  L  S  I  F  .  .  .  C  F  H  .
o
5'  AGAAGTATGGCTTGGAAAACTATGCTTTTGAGCACGTAGTCTATGGAGACTTCTCTCATACTACGTTAGG
o   +++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++ 20300
1   K  K  Y  G  L  E  N  Y  A  F  E  H  V  V  Y  G  D  F  S  H  T  T  L  G
2    R  S  M  A  W  K  T  M  L  L  S  T  .  S  M  E  T  S  L  I  L  R  .
3     E  V  W  L  G  K  L  C  F  .  A  R  S  L  W  R  L  L  S  Y  Y  V  R
o
```

Sunday, September 23, 2012 1:41 PM
HCoV-SA1 translation 3 frames



Sunday, September 23, 2012 1:41 PM
HCoV-SA1 translation 3 frames

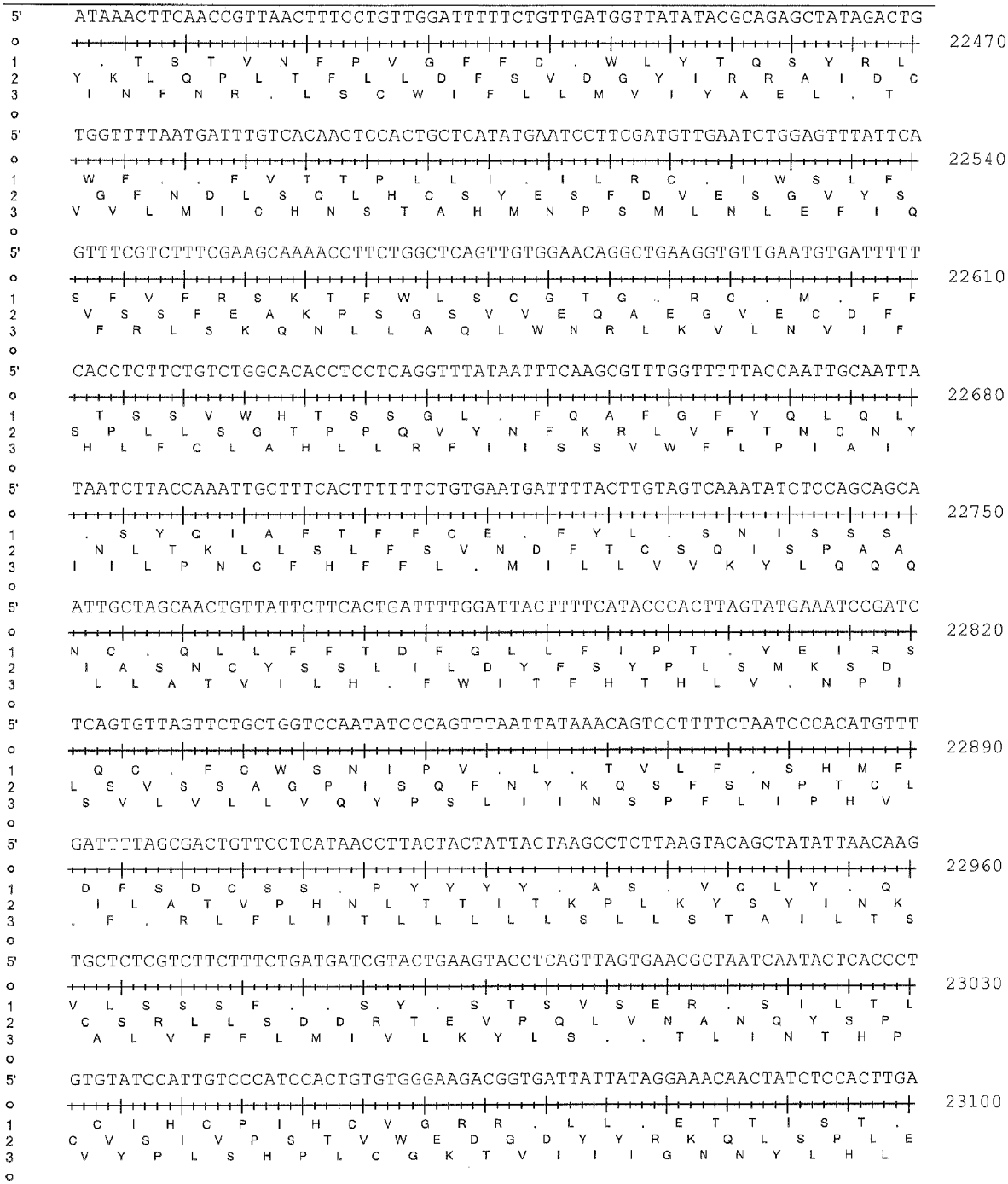
```
5'  CCTACTACTAAGAATGTAACAGGTAGTAATGAGTCAAAGGCTTTATTCTTTACTTACCTGTGTAACCTCA
o   +++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++  21070
1    P   T   T   K   N   V   T   G   S   N   E   S   K   A   L   F   F   T   Y   L   C   N   L
2     L   L   L   R   M   .   Q   V   V   M   S   Q   R   L   Y   S   L   L   T   C   V   T   S
3    S   Y   Y   .   E   C   N   R   .   .   .   V   K   G   F   I   L   Y   L   P   V   .   P   H
o
5'  TTAATAATAATCTTGCTCTTGGTGGGTCTGTTGCTATTAAAATAACAGAACACTCTTGGAGCGTTGAACT
o   +++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++  21140
1    I   N   N   N   L   A   L   G   G   S   V   A   I   K   I   T   E   H   S   W   S   V   E   L
2     L   I   I   I   L   L   L   V   G   L   L   L   L   K   .   Q   N   T   L   G   A   L   N
3     .   .   .   S   C   S   W   W   V   C   C   Y   .   N   N   R   T   L   L   E   R   .   T
o
5'  TTATGAACTTATGGGAAAATTTGCTTGGTGGACTGTTTTCTGCACCAATGCAAATGCATCCTCATCTGAA
o   +++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++  21210
1    Y   E   L   M   G   K   F   A   W   W   T   V   F   C   T   N   A   N   A   S   S   S   E
2    F   M   N   L   W   E   N   L   L   G   G   L   F   S   A   P   M   Q   M   H   P   H   L   K
3     L   .   T   Y   G   K   I   C   L   V   D   C   F   L   H   Q   C   K   C   I   L   I   .
o
5'  GGATTCCTCTTAGGTATTAATTACTTGGGTACTATTAAAGAAAATATAGATGGTGGTGCTATGCACGCCA
o   +++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++  21280
1    G   F   L   L   G   I   N   Y   L   G   T   I   K   E   N   I   D   G   G   A   M   H   A
2     D   S   S   .   V   L   I   T   W   V   L   L   K   K   I   .   M   V   V   L   C   T   P
3    R   I   P   L   R   Y   .   L   L   G   Y   Y   .   R   K   Y   R   W   W   C   Y   A   R   Q
o
5'  ACTATATATTTTGGAGAAATTCCACTCCTATGAATCTGAGTACTTACTCACTTTTTGATTTATCCAAGTT
o   +++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++  21350
1    N   Y   I   F   W   R   N   S   T   P   M   N   L   S   T   Y   S   L   F   D   L   S   K   F
2     T   I   Y   F   G   E   I   P   L   L   .   I   .   V   L   T   H   F   L   I   Y   P   S
3     L   Y   I   L   E   K   F   H   S   Y   E   S   E   Y   L   L   T   F   .   F   I   Q   V
o
5'  TCAATTAAAATTAAAAGGAACACCAGTTCTTCAATTAAAGGAGAGTCAAATTAACGAACTCGTAATATCT
o   +++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++  21420
1    Q   L   K   L   K   G   T   P   V   L   Q   L   K   E   S   Q   I   N   E   L   V   I   S
2    F   N   .   N   .   K   E   H   Q   F   F   N   .   R   R   V   K   L   T   N   S   .   Y   L
3     S   I   K   I   K   R   N   T   S   S   S   I   K   G   E   S   N   .   R   T   R   N   I
o
5'  CTCCTGTCGCAGGGTAAGTTACTTATCCGTGACAATGATACACTCAGTGTTTCTACTGATGTTCTTGTTA
o   +++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++  21490
1    L   L   S   Q   G   K   L   L   I   R   D   N   D   T   L   S   V   S   T   D   V   L   V
2     S   C   R   R   V   S   Y   L   S   V   T   M   I   H   S   V   F   L   L   M   F   L   L
3    S   P   V   A   G   .   V   T   Y   P   .   Q   .   Y   T   Q   C   F   Y   .   C   S   C   .
o
5'  ACACCTACAGAAAGTTACGTTGATGTAGGGCCAGATTCTGTTAAGTCTGCTTGTATTGAGGTTGATATAC
o   +++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++  21560
1    N   T   Y   R   K   L   R   .   C   R   A   R   F   C   .   V   C   L   Y   .   G   .   Y   T
2     T   P   T   E   S   Y   V   D   V   G   P   D   S   V   K   S   A   C   I   E   V   D   I
3     H   L   Q   K   V   T   L   M   .   G   Q   I   L   L   S   L   L   V   L   R   L   I   Y
o
5'  AACAGACTTTCTTTGATAAAACTTGGCCTAGGCCAATTGATGTTTCTAAGGCTGACGGTATTATATACCC
o   +++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++  21630
1    T   D   F   L   .   .   N   L   A   .   A   N   .   C   F   .   G   .   R   Y   Y   I   P
2    Q   Q   T   F   F   D   K   T   W   P   R   P   I   D   V   S   K   A   D   G   I   I   Y   P
3     N   R   L   S   L   I   K   L   G   L   G   Q   L   M   F   L   R   L   T   V   L   Y   T
o
5'  TCAAGGCCGTACATATTCTAACATAACTATCACTTATCAAGGTCTTTTTCCCTATCAGGGAGACCATGGT
o   +++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++  21700
1    S   R   P   Y   I   F   .   H   N   Y   H   L   S   R   S   F   S   L   S   G   R   P   W
2     Q   G   R   T   Y   S   N   I   T   I   T   Y   Q   G   L   F   P   Y   Q   G   D   H   G
3    L   K   A   V   H   I   L   T   .   L   S   L   I   K   V   F   F   P   I   R   E   T   M   V
o
```

Figure 15 continued

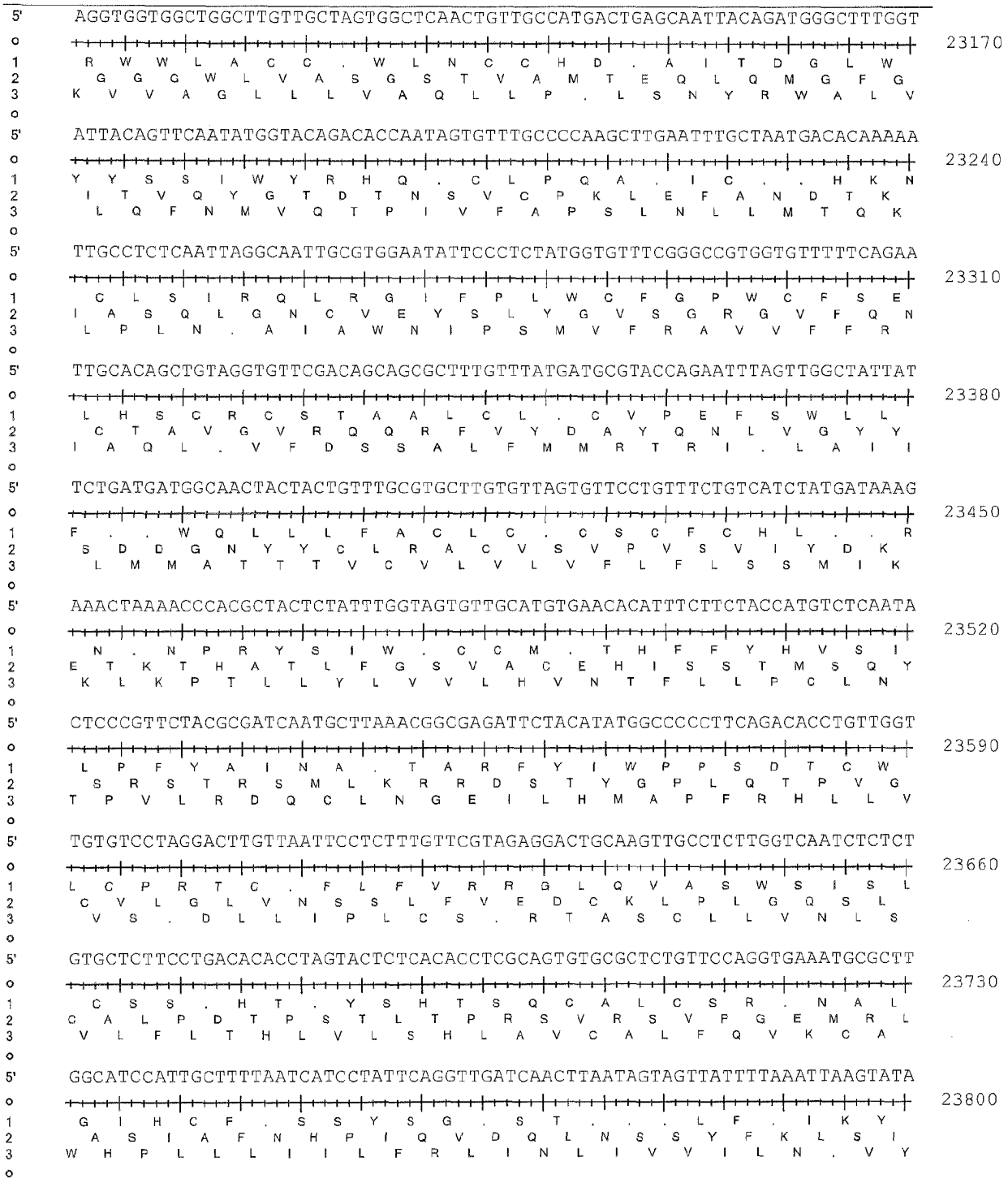
Sunday, September 23, 2012 1:41 PM
HCoV-SA1 translation 3 frames

```
5'  GATATGTATGTTTACTCTGCAGGACATGCTACAGGCACAACTCCACAAAAGTTGTTTGTAGCTAACTATT
0   +---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+  21770
1    .  Y  V  C  L  L  C  R  T  C  Y  R  H  N  S  T  K  V  V  C  S  .  L  F
2     D  M  Y  V  Y  S  A  G  H  A  T  G  T  T  P  Q  K  L  F  V  A  N  Y
3      I  C  M  F  T  L  Q  D  M  L  Q  A  Q  L  H  K  S  C  L  .  L  T  I
0
5'  CTCAGGACGTCAAACAGTTTGCTAATGGGTTTGTCGTCCGTATAGGAGCAGCTGCCAATTCCACTGGCAC
0   +---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+  21840
1     S  G  R  Q  T  V  C  .  W  V  C  R  P  Y  R  S  S  C  Q  F  H  W  H
2    S  Q  D  V  K  Q  F  A  N  G  F  V  V  R  I  G  A  A  A  N  S  T  G  T
3     L  R  T  S  N  S  L  L  M  G  L  S  S  V  .  E  Q  L  P  I  P  L  A
0
5'  TGTTATTATTAGCCCATCTACCAGCGCTACTATACGAAAAATTTACCCTGCTTTTATGCTGGGTTCTTCA
0   +---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+  21910
1     C  Y  Y  .  P  I  Y  Q  R  Y  Y  T  K  N  L  P  C  F  Y  A  G  F  F
2      V  I  I  S  P  S  T  S  A  T  I  R  K  I  Y  P  A  F  M  L  G  S  S
3     L  L  L  L  A  H  L  P  A  L  L  Y  E  K  F  T  L  L  L  C  W  V  L  Q
0
5'  GTTGGTAATTTCTCAGATGGTAAAATGGGCCGCTTCTTCAATCATACTCTAGTTCTTTTGCCCGATGGAT
0   +---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+  21980
1    S  W  .  F  L  R  W  .  N  G  P  L  L  Q  S  Y  S  S  S  F  A  R  W  M
2     V  G  N  F  S  D  G  K  M  G  R  F  F  N  H  T  L  V  L  L  P  D  G
3      L  V  I  S  Q  M  V  K  W  A  A  S  S  I  I  L  .  F  F  C  P  M  D
0
5'  GTGGCACTTTACTTAGAGCTTTTTATTGTATTCTAGAGCCTCGCTCTGGAAATCATTGTCCTGCTGGCAA
0   +---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+  22050
1     W  H  F  T  .  S  F  L  L  Y  S  R  A  S  L  W  K  S  L  S  C  W  Q
2    C  G  T  L  L  R  A  F  Y  C  I  L  E  P  R  S  G  N  H  C  P  A  G  N
3     V  A  L  Y  L  E  L  F  I  V  F  .  S  L  A  L  E  I  I  V  L  L  A
0
5'  TTCCTATACTTCTTTTGCCACTTATCACACTCCTGCAACAGATTGTTCTGATGGCAATTACAATCGTAAT
0   +---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+  22120
1     F  L  Y  F  F  C  H  L  S  H  S  C  N  R  L  F  .  W  Q  L  Q  S  .
2      S  Y  T  S  F  A  T  Y  H  T  P  A  T  D  C  S  D  G  N  Y  N  R  N
3     I  P  I  L  L  L  P  L  I  T  L  L  Q  Q  I  V  L  M  A  I  T  I  V  M
0
5'  GCCAGTCTGAACTCTTTTAAGGAGTATTTTAATTTACGTAACTGCACCTTTATGTACACTTATAACATTA
0   +---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+  22190
1    C  Q  S  E  L  F  .  G  V  F  .  F  T  .  L  H  L  Y  V  H  L  .  H  Y
2     A  S  L  N  S  F  K  E  Y  F  N  L  R  N  C  T  F  M  Y  T  Y  N  I
3      P  V  .  T  L  L  R  S  I  L  I  Y  V  T  A  P  L  C  T  L  I  T  L
0
5'  CCGAAGATGAGATTTTAGAGTGGTTTGGCATTACACAAACTGCTCAAGGTGTTCACCTCTTCTCATCTCG
0   +---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+  22260
1     R  R  .  D  F  R  V  V  W  H  Y  T  N  C  S  R  C  S  P  L  L  I  S
2    T  E  D  E  I  L  E  W  F  G  I  T  Q  T  A  Q  G  V  H  L  F  S  S  R
3     P  K  M  R  F  .  S  G  L  A  L  H  K  L  L  K  V  F  T  S  S  H  L
0
5'  GTATGTTGATTTGTACGGCGGCAATATGTTTCAATTTGCCACCTTGCCTGTTTATGATACTATTAAGTAT
0   +---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+  22330
1     V  C  .  F  V  R  R  Q  Y  V  S  I  C  H  L  A  C  L  .  Y  Y  .  V
2      Y  V  D  L  Y  G  G  N  M  F  Q  F  A  T  L  P  V  Y  D  T  I  K  Y
3     G  M  L  I  C  T  A  A  I  C  F  N  L  P  P  C  L  F  M  I  L  L  S  I
0
5'  TATTCTATCATTCCTCACAGTATTCGTTCTATCCAAAGTGATAGAAAAGCTTGGGCTGCCTTCTACGTAT
0   +---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+  22400
1    L  F  Y  H  S  S  Q  Y  S  F  Y  P  K  .  .  K  S  L  G  C  L  L  R  I
2     Y  S  I  I  P  H  S  I  R  S  I  Q  S  D  R  K  A  W  A  A  F  Y  V
3      I  L  S  F  L  T  V  F  V  L  S  K  V  I  E  K  L  G  L  P  S  T  Y
0
```

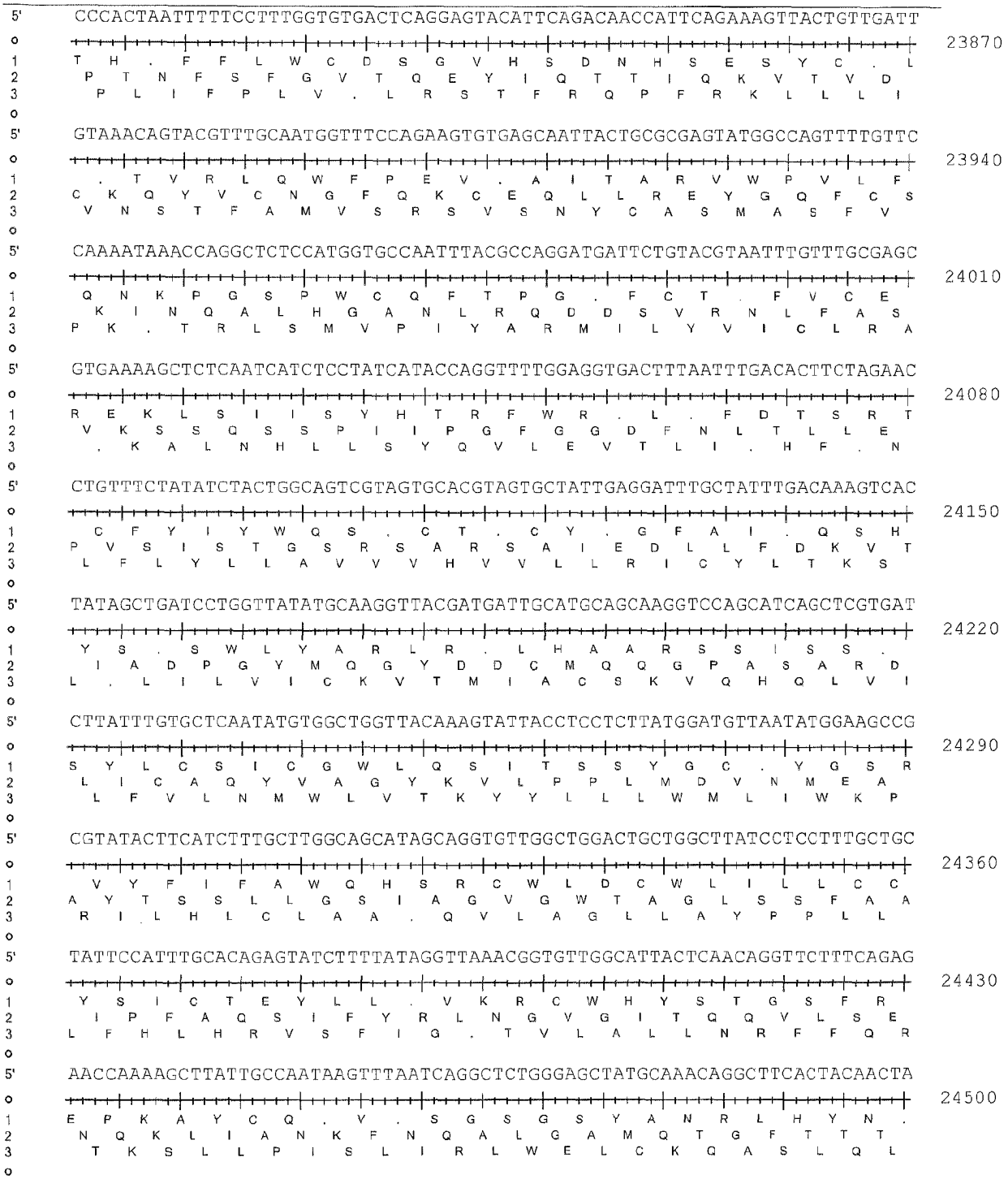
Sunday, September 23, 2012 1:41 PM
HCoV-SA1 translation 3 frames



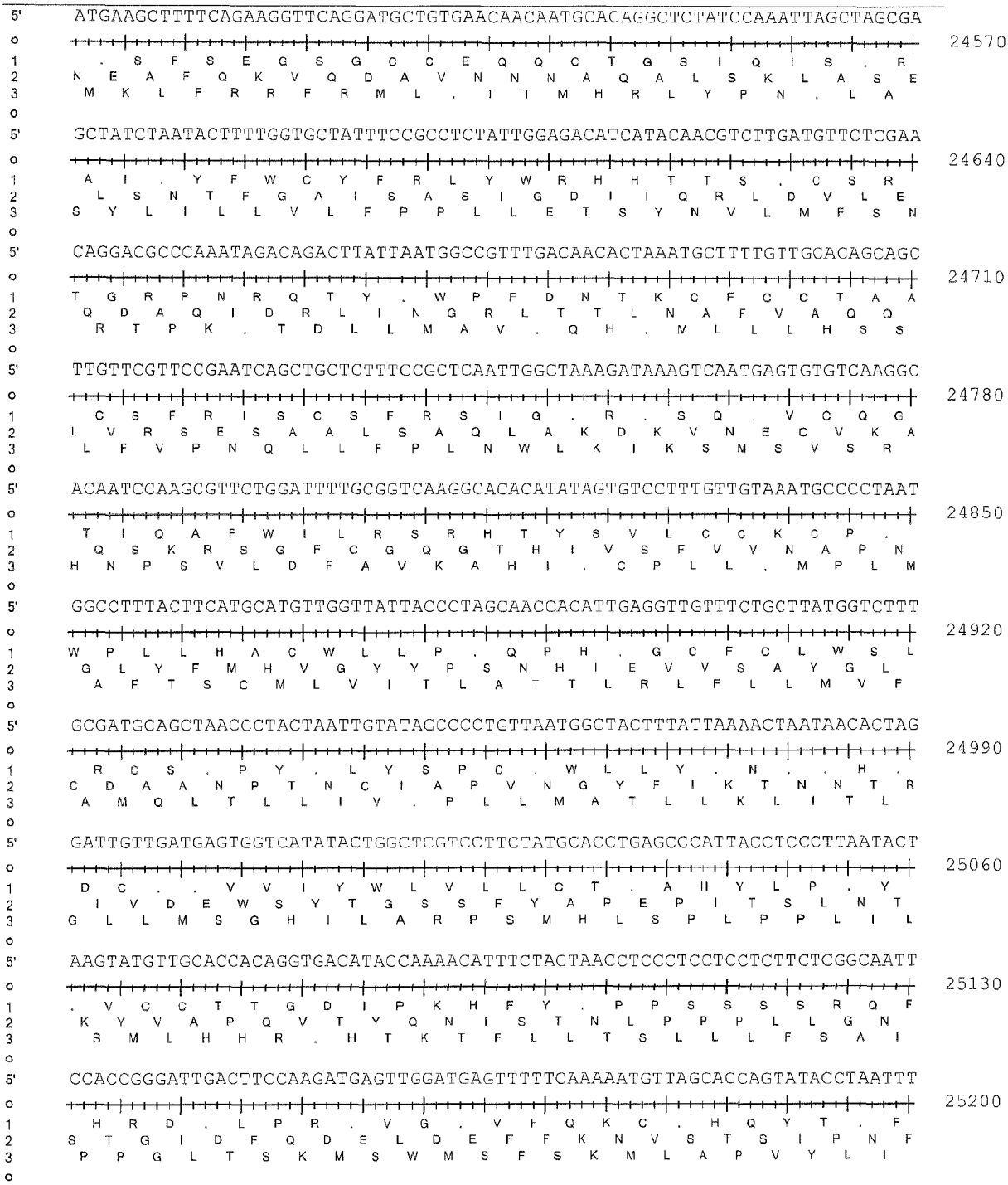
Sunday, September 23, 2012 1:41 PM
HCoV-SA1 translation 3 frames



Sunday, September 23, 2012 1:41 PM
HCoV-SA1 translation 3 frames



Sunday, September 23, 2012 1:41 PM
HCoV-SA1 translation 3 frames



Sunday, September 23, 2012 1:41 PM
HCoV-SA1 translation 3 frames

```
5'  TGGTTCCCTAACACAGATTAATACTACATTACTCGATCTTACCTACGAGATGTTGTCTCTTCAACAAGTT
o  +----|----|----|----|----|----|----|----|----|----|----|----|----|----|  25270
1   W  F  P  N  T  D  .  Y  Y  I  T  R  S  Y  L  R  D  V  V  S  S  T  S
2    G  S  L  T  Q  I  N  T  T  L  L  D  L  T  Y  E  M  L  S  L  Q  Q  V
3   L  V  P  .  H  R  L  I  L  H  Y  S  I  L  P  T  R  C  C  L  F  N  K  L
o
5'  GTTAAAGCCCTTAATGAGTCTTACATAGACCTTAAAGAGCTTGGCAATTATACTTATTACAACAAATGGC
o  +----|----|----|----|----|----|----|----|----|----|----|----|----|----|  25340
1   C  .  S  P  .  .  V  L  H  R  P  .  R  A  W  Q  L  Y  L  L  Q  Q  M  A
2    V  K  A  L  N  E  S  Y  I  D  L  K  E  L  G  N  Y  T  Y  Y  N  K  W
3     L  K  P  L  M  S  L  T  .  T  L  K  S  L  A  I  I  L  I  T  T  N  G
o
5'  CGTGGTACATTTGGCTTGGTTTCATTGCTGGGCTTGTTGCCTTAGCTCTATGCGTCTTCTTCATACTGTG
o  +----|----|----|----|----|----|----|----|----|----|----|----|----|----|  25410
1   V  V  H  L  A  W  F  H  C  W  A  C  C  L  S  S  M  R  L  L  H  T  V
2  P  W  Y  I  W  L  G  F  I  A  G  L  V  A  L  A  L  C  V  F  F  I  L  C
3   R  G  T  F  G  L  V  S  L  L  G  L  L  P  .  L  Y  A  S  S  S  Y  C
o
5'  CTGCACTGGTTGTGGCACAAACTGTATGGGAAAACTTAAGTGTAATCGTTGTTGTGATAGATACGAGGAA
o  +----|----|----|----|----|----|----|----|----|----|----|----|----|----|  25480
1   L  H  W  L  W  H  K  L  Y  G  K  T  .  V  .  S  L  L  .  .  I  R  G
2    C  T  G  C  G  T  N  C  M  G  K  L  K  C  N  R  C  C  D  R  Y  E  E
3  A  A  L  V  V  A  Q  T  V  W  E  N  L  S  V  I  V  V  V  I  D  T  R  N
o
5'  TACGACCTCGAGCCGCATAAGGTTCATGTTCACTAATTAACGAACTATTAATGAGAGTTCAAAGACCACC
o  +----|----|----|----|----|----|----|----|----|----|----|----|----|----|  25550
1  I  R  P  R  A  A  .  G  S  C  S  L  I  N  E  L  L  M  R  V  Q  R  P  P
2   Y  D  L  E  P  H  K  V  H  V  H  .  L  T  N  Y  .  .  E  F  K  D  H
3    T  T  S  S  R  I  R  F  M  F  T  N  .  R  T  I  N  E  S  S  K  T  T
o
5'  CACTCTCTTGTTAGTGTTTTCACTCTCTCTTTTGGTCACTGCATCCTCAAAACCTCTCTATGTACCTGAG
o  +----|----|----|----|----|----|----|----|----|----|----|----|----|----|  25620
1   T  L  L  L  V  F  S  L  S  L  L  V  T  A  S  S  K  P  L  Y  V  P  E
2  P  L  S  C  .  C  F  H  S  L  F  W  S  L  H  P  Q  N  L  S  M  Y  L  S
3   H  S  L  V  S  V  F  T  L  S  F  G  H  C  I  L  K  T  S  L  C  T  .
o
5'  CATTGTCAGAATTATTCTGGTTGCATGCTTAGGGCTTGTATTAAAACTGCCCAAGCTGATACAGCTGGTC
o  +----|----|----|----|----|----|----|----|----|----|----|----|----|----|  25690
1   H  C  Q  N  Y  S  G  C  M  L  R  A  C  I  K  T  A  Q  A  D  T  A  G
2    I  V  R  I  I  L  V  A  C  L  G  L  V  L  K  L  P  K  L  I  Q  L  V
3  A  L  S  E  L  F  W  L  H  A  .  G  L  Y  .  N  C  P  S  .  Y  S  W  S
o
5'  TTTATACAAATTTTCGAATTGACGTCCCATCTGCAGAATCAACTGGTACTCAATCAGTTTCTGTCGATCT
o  +----|----|----|----|----|----|----|----|----|----|----|----|----|----|  25760
1  L  Y  T  N  F  R  I  D  V  P  S  A  E  S  T  G  T  Q  S  V  S  V  D  L
2   F  I  Q  I  F  E  L  T  S  H  L  Q  N  Q  L  V  L  N  Q  F  L  S  I
3    L  Y  K  F  S  N  .  R  P  I  C  R  I  N  W  Y  S  I  S  F  C  R  S
o
5'  TGAGTCAACTTCAACTCATGATGGTCCTACCGAACATGTTACTAGTGTGAATCTTTTTGACGTTGGTTAC
o  +----|----|----|----|----|----|----|----|----|----|----|----|----|----|  25830
1   E  S  T  S  T  H  D  G  P  T  E  H  V  T  S  V  N  L  F  D  V  G  Y
2  L  S  Q  L  Q  L  M  M  V  L  P  N  M  L  L  V  .  I  F  L  T  L  V  T
3   .  V  N  F  N  S  .  W  S  Y  R  T  C  Y  .  C  E  S  F  .  R  W  L
o
5'  TCAGTTAATTAACGAACTCTATGGATTACGTGTCTCTGCTTAATCAAATTTGGCAGAAGTACCTTAACTC
o  +----|----|----|----|----|----|----|----|----|----|----|----|----|----|  25900
1   S  V  N  .  R  T  L  W  I  T  C  L  C  L  I  K  F  G  R  S  T  L  T
2    Q  L  I  N  E  L  Y  G  L  R  V  S  A  .  S  N  L  A  E  V  P  .  L
3  L  S  .  L  T  N  S  M  D  Y  V  S  L  L  N  Q  I  W  Q  K  Y  L  N  S
o
```

Sunday, September 23, 2012 1:41 PM
HCoV-SA1 translation 3 frames

```
5'  ACCGTATACTACTTGTTTGTACATCCCTAAACCCACAGCTAAGTATACACCTTTAGTTGGCACTTCATTG
o   +---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+  25970
1   H R I L L V C T S L N P Q L S I H L . L A L H C
2    T V Y Y L F V H P . T H S . V Y T F S W H F I
3     P Y T T C L Y I P K P T A K Y T P L V G T S L
o
5'  CACCCTGTGCTGTGGAACTGTCAGCTATCCTTTGCTGGTTATACTGAATCTGCTGTTAATTCTACAAAAG
o   +---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+  26040
1    T L C C G T V S Y P L L V I L N L L L I L Q K
2   A P C A V E L S A I L C W L Y . I C C . F Y K S
3    H P V L W N C Q L S F A G Y T E S A V N S T K
o
5'  CTTTGGCCAAACAGGACGCAGCTCAGCGAATCGCTTGGTTGCTACATAAGGATGGAGGAATCCCTGATGG
o   +---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+  26110
1    L W P N R T Q L S E S L G C Y I R M E E S L M
2     F G Q T G R S S A N R L V A T . G W R N P . W
3   A L A K Q D A A Q R I A W L L H K D G G I P D G
o
5'  ATGTTCCCTCTACCTCCGGCACTCAAGTTTATTCGCGCAAAGCGAGGAAGAGGAGCCATTCTCCAACTAA
o   +---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+  26180
1   D V P S T S G T Q V Y S R K A R K R S H S P T K
2    M F P L P P A L K F I R A K R G R G A I L Q L
3     C S L Y L R H S S L F A Q S E E E E P F S N .
o
5'  GAAACTGCGCTACGTTAAGCGTAGATTTTCTCTTCTGCGCCATGAAGACCTTAGTGTTATTGTCCAACCA
o   +---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+  26250
1    K L R Y V K R R F S L L R H E D L S V I V Q P
2   R N C A T L S V D F L F C A M K T L V L L S N Q
3    E T A L R . A . I F S S A P . R P . C Y C P T
o
5'  ACACACTATGTCAGGGTTACATTTTCAGACCCCAACATGTGGTATCTACGTTCGGGTCATCATTTACACT
o   +---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+  26320
1    T H Y V R V T F S D P N M W Y L R S G H H L H
2     H T M S G L H F Q T P T C G I Y V R V I I Y T
3   N T L C Q G Y I F R P Q H V V S T F G S S F T L
o
5'  CAGTTCACAATTGGCTTAAACCTTATGGCGGCCAACCTGTTTCTGAGTACCATATTACTCTAGCTTTGCT
o   +---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+  26390
1   S V H N W L K P Y G G Q P V S E Y H I T L A L L
2    Q F T I G L N L M A A N L F L S T I L L . L C
3     S S Q L A . T L W R P T C F . V P Y Y S S F A
o
5'  AAATCTCACTGATGAAGATTTAGCTAGAGATTTTTCACCCATTGCGCTCTTTTTGCGCAATGTCAGATTT
o   +---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+  26460
1    N L T D E D L A R D F S P I A L F L R N V R F
2   . I S L M K I . L E I F H P L R S F C A M S D L
3    K S H . . R F S . R F F T H C A L F A Q C Q I
o
5'  GAGCTACATGAGTTCGCCTTGCTGCGCAAAACTCTTGTTCTTAATGCATCAGAGATCTACTGTGCTAACA
o   +---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+  26530
1   E L H E F A L L R K T L V L N A S E I Y C A N
2    S Y M S S P C C A K L L F L M H Q R S T V L T
3   . A T . V R L A A Q N S C S . C I R D L L C . H
o
5'  TACATAGATTTAAGCCTGTGTATAGAGTTAACACGGCAATCCCTACTATTAAGGATTGGCTTCTCGTTCA
o   +---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+  26600
1   I H R F K P V Y R V N T A I P T I K D W L L V Q
2   Y I D L S L C I E L T R Q S L L R I G F S F
3    T . I . A C V . S . H G N P Y Y . G L A S R S
o
```

Sunday, September 23, 2012 1:41 PM
HCoV-SA1 translation 3 frames

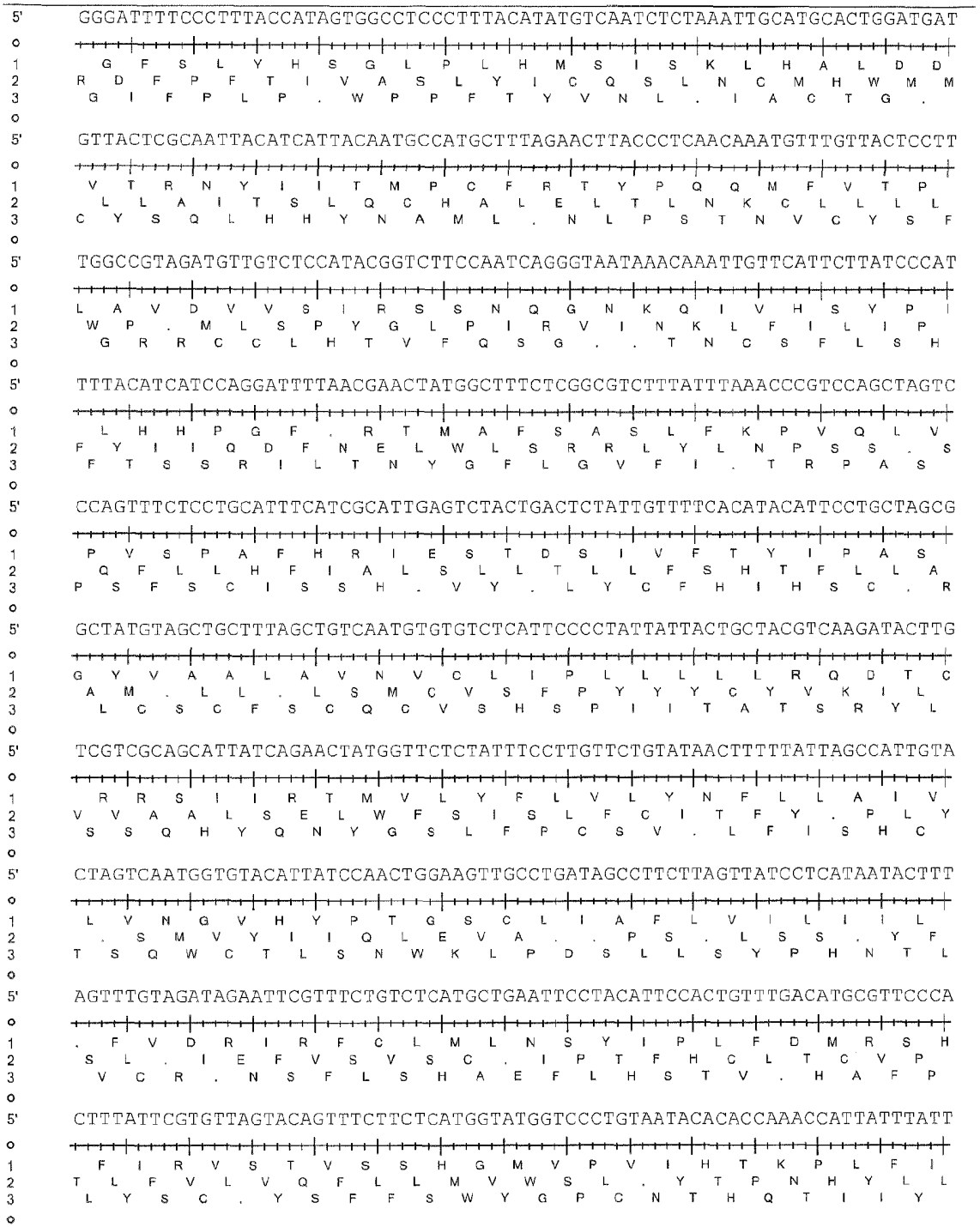


Figure 15 continued

Sunday, September 23, 2012 1:41 PM
HCoV-SA1 translation 3 frames

```
5'  AGAAACTTCGATCAGCGTTGCAGCTGTTCTCGTTGTTTTTATTTGCACTCTTCCACTTATATAGAGTGCA
o  +++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+  27370
1   R  N  F  D  Q  R  C  S  C  S  R  C  F  Y  L  H  S  S  T  Y  I  E  C
2    E  T  S  I  S  V  A  A  V  L  V  V  F  I  C  T  L  P  L  I  .  S  A
3   .  K  L  R  S  A  L  Q  L  F  S  L  F  L  F  A  L  F  H  L  Y  R  V  H
o
5'  CTTATATTAGCCGTTTTAGTAAGATTAGCCTAGTTTCTGTAACTGACTTCTCCTTAAACGGCAATGTTTC
o  +++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+  27440
1   T  Y  I  S  R  F  S  K  I  S  L  V  S  V  T  D  F  S  L  N  G  N  V  S
2    L  I  L  A  V  L  V  R  L  A  .  F  L  .  L  T  S  P  .  T  A  M  F
3     L  Y  .  P  F  .  .  D  .  P  S  F  C  N  .  L  L  L  K  R  Q  C  F
o
5'  CACTGTTTTCGTGCCTGCAACGCGCGATTCAGTTCCTCTTCACATAATCGCCCCGAGCTCGCTTATCGTT
o  +++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+  27510
1   T  V  F  V  P  A  T  R  D  S  V  P  L  H  I  I  A  P  S  S  L  I  V
2  P  L  F  S  C  L  Q  R  A  I  Q  F  L  F  T  .  S  P  R  A  R  L  S  F
3   H  C  F  R  A  C  N  A  R  F  S  S  S  S  H  N  R  P  E  L  A  Y  R
o
5'  TAAGCAGCTCTGCGCTACTATGGGTCCCGTGTAGAGGCTAATCCATTAGTCTCTCTTTGGACATATGGAA
o  +++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+  27580
1   .  A  A  L  R  Y  Y  G  S  R  V  E  A  N  P  L  V  S  L  W  T  Y  G
2    K  Q  L  C  A  T  M  G  P  V  .  R  L  I  H  .  S  L  F  G  H  M  E
3  L  S  S  S  A  L  L  W  V  P  C  R  G  .  S  I  S  L  S  L  D  I  W  K
o
5'  AACGAACTATGTTACCCTTTGTCCAAGAACGAATAGGGTTGTTCATAGTAAACTTTTTCATTTTTACCGT
o  +++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+  27650
1  K  R  T  M  L  P  F  V  Q  E  R  I  G  L  F  I  V  N  F  F  I  F  T  V
2   N  E  L  C  Y  P  L  S  K  N  E  .  G  C  S  .  .  T  F  S  F  L  P
3    T  N  Y  V  T  L  C  P  R  T  N  R  V  V  H  S  K  L  F  H  F  Y  R
o
5'  AGTATGTGCTATAACACTCTTGGTGTGTATGGCTTTCCTTACGGCTACTAGATTATGTGTGCAATGTATG
o  +++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+  27720
1   V  C  A  I  T  L  L  V  C  M  A  F  L  T  A  T  R  L  C  V  Q  C  M
2  .  Y  V  L  .  H  S  W  C  V  W  L  S  L  R  L  L  D  Y  V  C  N  V  .
3   S  M  C  Y  N  T  L  G  V  Y  G  F  P  Y  G  Y  .  I  M  C  A  M  Y
o
5'  ACAGGCTTCAATACCCTGTTAGTTCAGCCCGCATTATACTTGTATAATACTGGACGTTCAGTCTATGTAA
o  +++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+  27790
1   T  G  F  N  T  L  L  V  Q  P  A  L  Y  L  Y  N  T  G  R  S  V  Y  V
2    Q  A  S  I  P  C  .  F  S  P  H  Y  T  C  I  I  L  D  V  Q  S  M  .
3  D  R  L  Q  Y  P  V  S  S  A  R  I  I  L  V  .  Y  W  T  F  S  L  C  K
o
5'  AATTCCAGGATAGTAAACCCCCTCTACCACCTGACGAGTGGGTTTAACGAACTCCTTCATAATGTCTAAT
o  +++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+  27860
1  K  F  Q  D  S  K  P  P  L  P  P  D  E  W  V  .  R  T  P  S  .  C  L  I
2   N  S  R  I  V  N  P  L  Y  H  L  T  S  G  F  N  E  L  L  H  N  V  .
3    I  P  G  .  .  T  P  S  T  T  .  R  V  G  L  T  N  S  F  I  M  S  N
o
5'  ATGACGCAACTCACTGAGGCGCAGATTATTGCCATTATTAAAGACTGGAACTTTGCATGGTCCCTGATCT
o  +++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+  27930
1   .  R  N  S  L  R  R  R  L  L  P  L  L  K  T  G  T  L  H  G  P  .  S
2  Y  D  A  T  H  .  G  A  D  Y  C  H  Y  .  R  L  E  L  C  M  V  P  D  L
3   M  T  Q  L  T  E  A  Q  I  I  A  I  I  K  D  W  N  F  A  W  S  L  I
o
5'  TTCTCTTAATTACTATCGTACTACAGTATGGATACCCATCCCGTAGTATGACTGTCTATGTCTTTAAAAT
o  +++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+++|+  28000
1   F  S  .  L  L  S  Y  Y  S  M  D  T  H  P  V  V  .  L  S  M  S  L  K
2    S  L  N  Y  Y  R  T  T  V  W  I  P  I  P  .  Y  D  C  L  C  L  .  N
3  F  L  L  I  T  I  V  L  Q  Y  G  Y  P  S  R  S  M  T  V  Y  V  F  K  M
o
```

Sunday, September 23, 2012 1:41 PM
HCoV-SA1 translation 3 frames

```
5'  GTTTGTTTTATGGCTCCTATGGCCATCTTCCATGGCGCTATCAATATTTAGCGCCGTTTATCCAATTGAT
o  +---------+---------+---------+---------+---------+---------+---------+  28070
1  C  L  F  Y  G  S  Y  G  H  L  P  W  R  Y  Q  Y  L  A  P  F  I  Q  L  I
2   V  C  F  M  A  P  M  A  I  F  H  G  A  I  N  I  .  R  R  L  S  N  .
3    F  V  L  W  L  L  W  P  S  S  M  A  L  S  I  F  S  A  V  Y  P  I  D
o

5'  CTAGCTTCCCAGATAATCTCTGGCATTGTAGCAGCTGTTTCAGCTATGATGTGGATTTCCTACTTTGTGC
o  +---------+---------+---------+---------+---------+---------+---------+  28140
1   .  L  P  R  .  S  L  A  L  .  Q  L  F  Q  L  .  C  G  F  P  T  L  C
2  S  S  F  P  D  N  L  W  H  C  S  S  C  F  S  Y  D  V  D  F  L  L  C  A
3   L  A  S  Q  I  I  S  G  I  V  A  A  V  S  A  M  M  W  I  S  Y  F  V
o

5'  AGAGTATCCGGCTGTTTATGAGAACTGGATCATGGTGGTCATTCAATCCTGAGACTAATTGCCTTTTGAA
o  +---------+---------+---------+---------+---------+---------+---------+  28210
1   R  V  S  G  C  L  .  E  L  D  H  G  G  H  S  I  L  R  L  I  A  F  .
2    E  Y  P  A  V  Y  E  N  W  I  M  V  V  I  Q  S  .  D  .  L  P  F  E
3  Q  S  I  R  L  F  M  R  T  G  S  W  W  S  F  N  P  E  T  N  C  L  L  N
o

5'  CGTTCCATTTGGTGGTACAACTGTCGTACGTCCACTCGTAGAGGACTCTACCAGTGTAACTGCTGTTGTA
o  +---------+---------+---------+---------+---------+---------+---------+  28280
1  T  F  H  L  V  V  Q  L  S  Y  V  H  S  .  R  T  L  P  V  .  L  L  L  .
2   R  S  I  W  W  Y  N  C  R  T  S  T  R  R  G  L  Y  Q  C  N  C  C  C
3    V  P  F  G  G  T  T  V  V  R  P  L  V  E  D  S  T  S  V  T  A  V  V
o

5'  ACCAATGGCCACCTCAAAATGGCTGGCATGCATTTCGGTGCTTGTGACTACGACAGACTTCCTAATGAAG
o  +---------+---------+---------+---------+---------+---------+---------+  28350
1   P  M  A  T  S  K  W  L  A  C  I  S  V  L  V  T  T  T  D  F  L  M  K
2  N  Q  W  P  P  Q  N  G  W  H  A  F  R  C  L  .  L  R  Q  T  S  .  .  S
3   T  N  G  H  L  K  M  A  G  M  H  F  G  A  C  D  Y  D  R  L  P  N  E
o

5'  TCACCGTGGCCAAACCCAATGTGCTGATTGCTTTAAAAATGGTGAAGCGGCAAAGCTACGGAACTAATTC
o  +---------+---------+---------+---------+---------+---------+---------+  28420
1   S  P  W  P  N  P  M  C  .  L  L  .  K  W  .  S  G  K  A  T  E  L  I
2    H  R  G  Q  T  Q  C  A  D  C  F  K  N  G  E  A  A  K  L  R  N  .  F
3  V  T  V  A  K  P  N  V  L  I  A  L  K  M  V  K  R  Q  S  Y  G  T  N  S
o

5'  CGGCGTTGCCATTTACCATAGATATAAGGCAGGTAATTACAGGAGTCCGCCTATTACGGCGGATATTGAA
o  +---------+---------+---------+---------+---------+---------+---------+  28490
1  P  A  L  P  F  T  I  D  I  R  Q  V  I  T  G  V  R  L  L  R  R  I  L  N
2   R  R  C  H  L  P  .  I  .  G  R  .  L  Q  E  S  A  Y  Y  G  G  Y  .
3    G  V  A  I  Y  H  R  Y  K  A  G  N  Y  R  S  P  P  I  T  A  D  I  E
o

5'  CTTGCATTGCTTCGAGCTTAGGCTCTTTAGTAAGAGTATCTTAATTGATTTTAACGAATCTCAATTTCAT
o  +---------+---------+---------+---------+---------+---------+---------+  28560
1   L  H  C  F  E  L  R  L  F  S  K  S  I  L  I  D  F  N  E  S  Q  F  H
2  T  C  I  A  S  S  L  G  S  L  V  R  V  S  .  L  I  L  T  N  L  N  F  I
3   L  A  L  L  R  A  .  A  L  .  .  E  Y  L  N  .  F  .  R  I  S  I  S
o

5'  TGTTATGGCATCCCCTGCTGCACCTCGTGCTGTTTCCTTTGCCGATAACAATGATATAACAAATACAAAC
o  +---------+---------+---------+---------+---------+---------+---------+  28630
1  C  Y  G  I  P  C  C  T  S  C  C  F  L  C  R  .  Q  .  Y  N  K  Y  K
2   V  M  A  S  P  A  A  P  R  A  V  S  F  A  D  N  N  D  I  T  N  T  N
3  L  L  W  H  P  L  L  H  L  V  L  F  P  L  P  I  T  M  I  .  Q  I  Q  T
o

5'  CTATCTCGAGGTAGAGGACGTAATCCAAAACCACGAGCTGCACCAAATAACACTGTCTCTTGGTACACTG
o  +---------+---------+---------+---------+---------+---------+---------+  28700
1  P  I  S  R  .  R  T  .  S  K  T  T  S  C  T  K  .  H  C  L  L  V  H  W
2   L  S  R  G  R  G  R  N  P  K  P  R  A  A  P  N  N  T  V  S  W  Y  T
3    Y  L  E  V  E  D  V  I  Q  N  H  E  L  H  Q  I  T  L  S  L  G  T  L
o
```

Sunday, September 23, 2012 1:41 PM
HCoV-SA1 translation 3 frames

```
5'  GGCTTACCCAACACGGGAAAGTCCCTCTTACCTTTCCACCTGGGCAGGGTGTACCTCTTAATGCCAATTC
o   +---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+  28770
1     A   Y   P   T   R   E   S   P   S   Y   L   S   T   W   A   G   C   T   S   .   C   Q   F
2   G   L   T   Q   H   G   K   V   P   L   T   F   P   P   G   Q   G   V   P   L   N   A   N   S
3     G   L   P   N   T   G   K   S   L   L   P   F   H   L   G   R   V   Y   L   L   M   P   I
o
5'  TACCCCTGCGCAAAATGCTGGGTATTGGCGGAGACAGGACAGAAAAATTAATACCGGGAATGGAATTAAG
o   +---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+  28840
1   Y   P   C   A   K   C   W   V   L   A   E   T   G   Q   K   N   .   Y   R   E   W   N   .
2     T   P   A   Q   N   A   G   Y   W   R   R   Q   D   R   K   I   N   T   G   N   G   I   K
3   L   P   L   R   K   M   L   G   I   G   G   D   R   T   E   K   L   I   P   G   M   E   L   S
o
5'  CAACTGGCTCCCAGGTGGTACTTCTACTACACTGGAACTGGACCCGAAGCAGCACTCCCATTCCGGGCTG
o   +---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+  28910
1   A   T   G   S   Q   V   V   L   L   L   H   W   N   W   T   R   S   S   T   P   I   P   G   C
2     Q   L   A   P   R   W   Y   F   Y   Y   T   G   T   G   P   E   A   A   L   P   F   R   A
3     N   W   L   P   G   G   T   S   T   T   L   E   L   D   P   K   Q   H   S   H   S   G   L
o
5'  TTAAGGATGGCATCGTTTGGGTCCATGAAGATGGCGCCACTGATGCTCCTTCAACTTTTGGGACGCGGAA
o   +---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+  28980
1     .   G   W   H   R   L   G   P   .   R   W   R   H   .   C   S   F   N   F   W   D   A   E
2   V   K   D   G   I   V   W   V   H   E   D   G   A   T   D   A   P   S   T   F   G   T   R   N
3     L   R   M   A   S   F   G   S   M   K   M   A   P   L   M   L   L   Q   L   L   G   R   G
o
5'  CCCTAACAATGATTCAGCTATTGTTACACAATTCGCGCCCGGTACTAAGCTTCCTAAAAACTTCCACATT
o   +---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+  29050
1     P   .   Q   .   F   S   Y   C   Y   T   I   R   A   R   Y   .   A   S   .   K   L   P   H
2       P   N   N   D   S   A   I   V   T   Q   F   A   P   G   T   K   L   P   K   N   F   H   I
3   T   L   T   M   I   Q   L   L   L   H   N   S   R   P   V   L   S   F   L   K   T   S   T   L
o
5'  GAGGGGACTGGAGGCAATAGTCAATCATCTTCAAGAGCCTCTAGCTTAAGCAGAAACTCTTCCAGATCTA
o   +---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+  29120
1   .   G   D   W   R   Q   .   S   I   I   F   K   S   L   .   L   K   Q   K   L   F   Q   I   .
2     E   G   T   G   G   N   S   Q   S   S   S   R   A   S   S   L   S   R   N   S   S   R   S
3     R   G   L   E   A   I   V   N   H   L   Q   E   P   L   A   .   A   E   T   L   P   D   L
o
5'  GTTCACAAGGTTCAAGATCAGGAAACTCTACCCGCGGCACTTCTCCAGGTCCATCTGGAATCGGAGCAGT
o   +---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+  29190
1     F   T   R   F   K   I   R   K   L   Y   P   R   H   F   S   R   S   I   W   N   R   S   S
2   S   S   Q   G   S   R   S   G   N   S   T   R   G   T   S   P   G   P   S   G   I   G   A   V
3     V   H   K   V   Q   D   Q   E   T   L   P   A   A   L   L   Q   V   H   L   E   S   E   Q
o
5'  AGGAGGTGATCTACTTTACCTTGATCTTCTGAACAGACTACAAGCCCTTGAGTCTGGCAAAGTAAAGCAA
o   +---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+  29260
1     R   R   .   S   T   L   P   .   S   S   E   Q   T   T   S   P   .   V   W   Q   S   K   A
2       G   G   D   L   L   Y   L   D   L   L   N   R   L   Q   A   L   E   S   G   K   V   K   Q
3   .   E   V   I   Y   F   T   L   I   F   .   T   D   Y   K   P   L   S   L   A   K   .   S   N
o
5'  TCGCAGCCAAAAGTAATCACTAAGAAAGATGCTGCTGCTGCTAAAAATAAGATGCGCCACAAGCGCACTT
o   +---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+  29330
1   I   A   A   K   S   N   H   .   E   R   C   C   C   C   .   K   .   D   A   P   Q   A   H   F
2     S   Q   P   K   V   I   T   K   K   D   A   A   A   A   K   N   K   M   R   H   K   R   T
3     R   S   Q   K   .   S   L   R   K   M   L   L   L   L   K   I   R   C   A   T   S   A   L
o
5'  CCACCAAAAGTTTCAACATGGTGCAAGCTTTTGGTCTTCGCGGACCAGGAGACCTCCAGGGAAACTTTGG
o   +---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+  29400
1     H   Q   K   F   Q   H   G   A   S   F   W   S   S   R   T   R   R   P   P   G   K   L   W
2   S   T   K   S   F   N   M   V   Q   A   F   G   L   R   G   P   G   D   L   Q   G   N   F   G
3     P   P   K   V   S   T   W   C   K   L   L   V   F   A   D   Q   E   T   S   R   E   T   L
o
```

Sunday, September 23, 2012 1:41 PM
HCoV-SA1 translation 3 frames

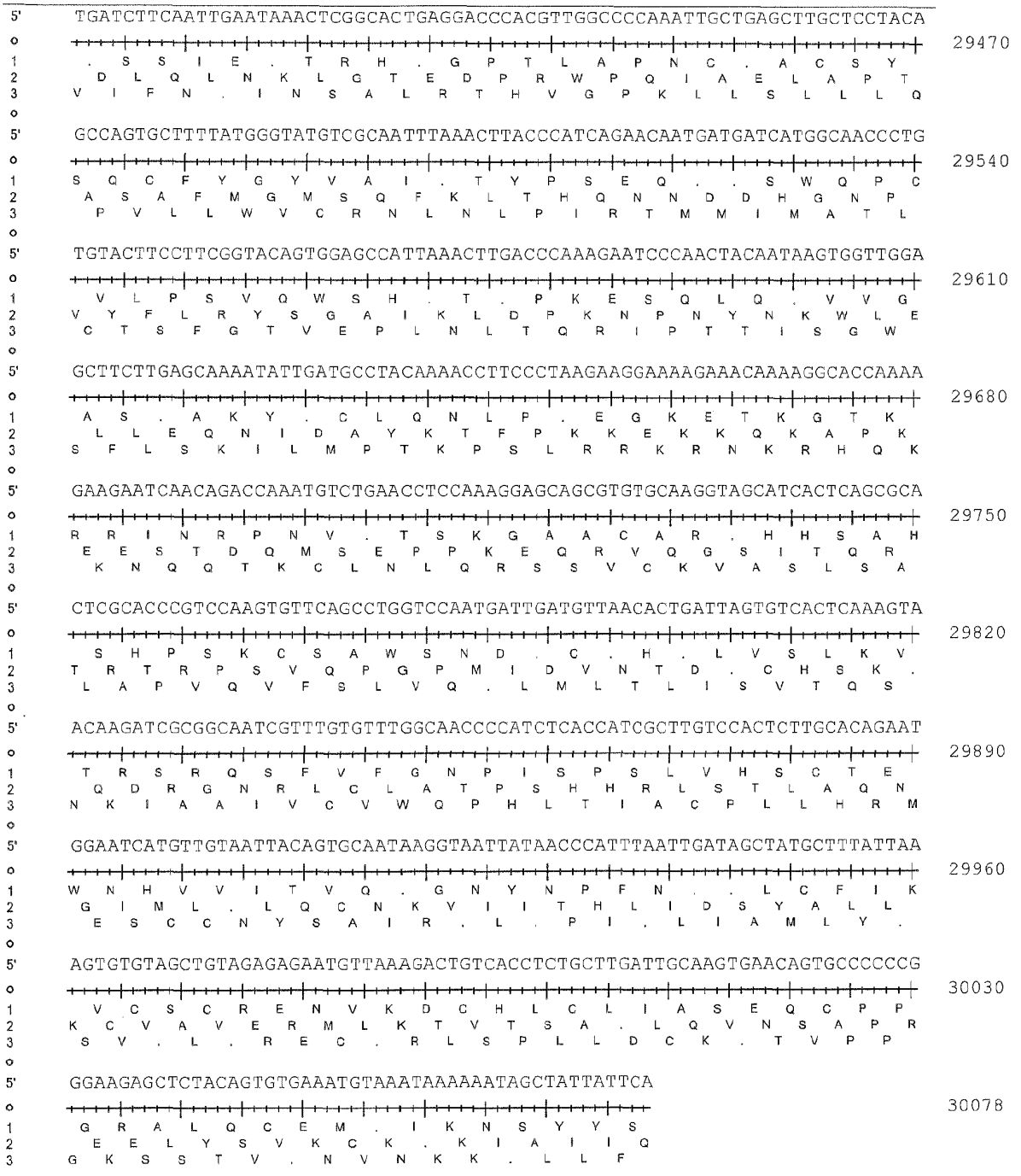


Figure 16

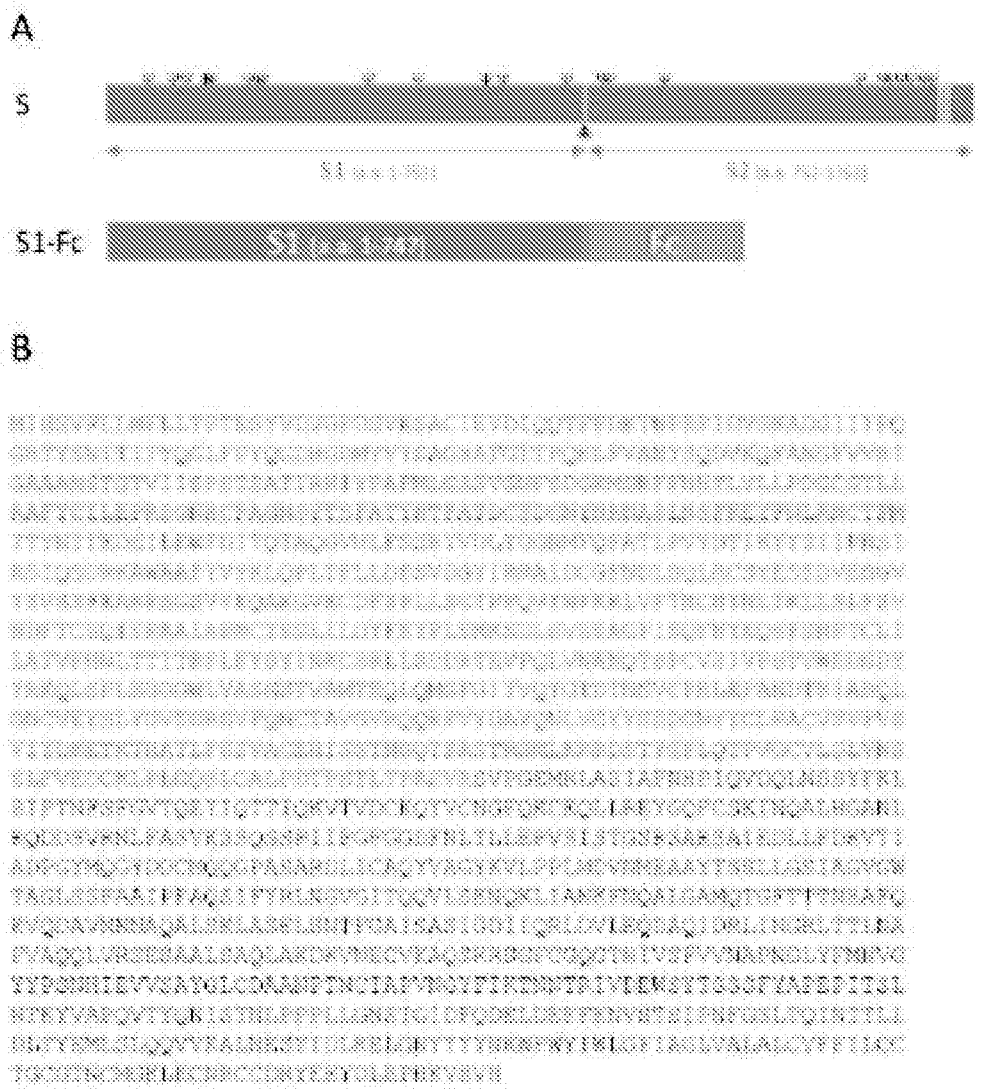


Figure 17

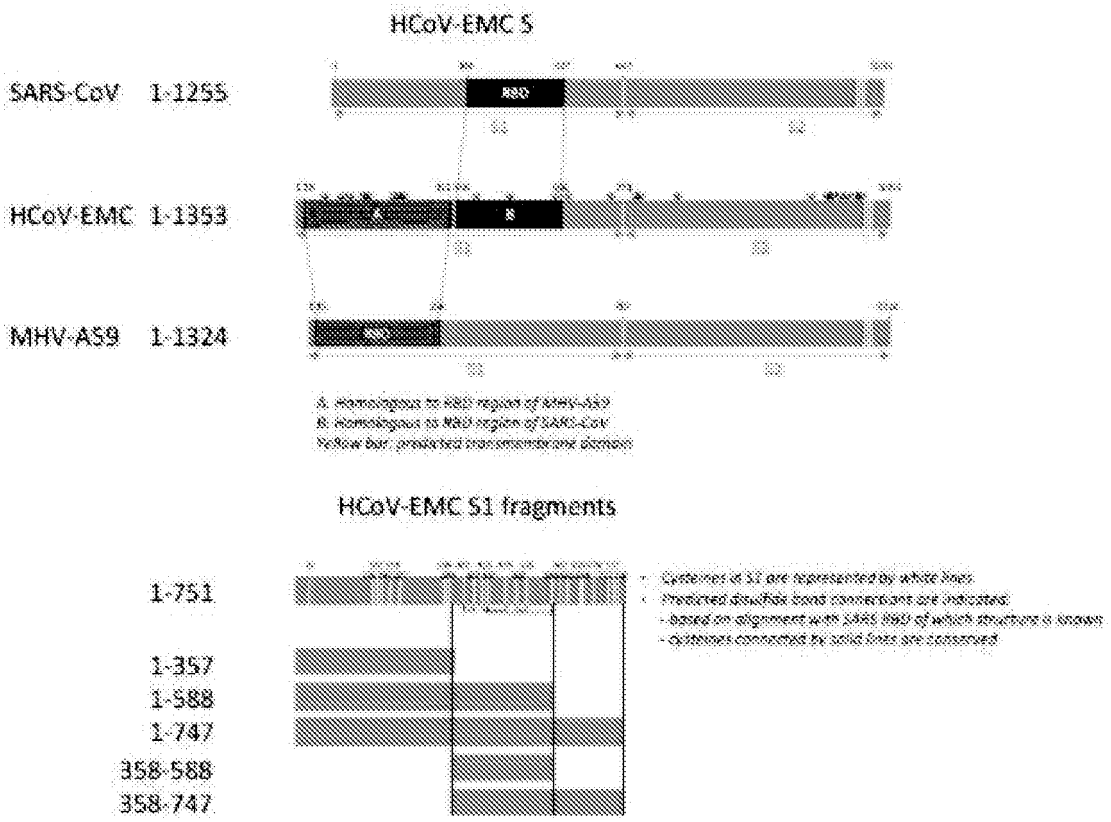

HCoV-EMC S
SARS-CoV 1-1255
RBD
HCoV-EMC 1-1353
A
B
MHV-A59 1-1324
A: homologous to RBD region of MHV-A59
B: homologous to RBD region of SARS-CoV
Yellow bar: predicted transmembrane domain
HCoV-EMC S1 fragments
1-751
1-357
1-588
1-747
358-588
358-747
- Cysteines of S1 are represented by white lines
- Predicted disulfide bond connections are indicated
- based on alignment with SARS RBD of which structure is known
- Cysteines connected by solid lines are conserved

Figure 18


Hum.DPP-IV 1-766
Orange: α/β-hydrolase domain
Green: β-propellor domain
Red bar: predicted transmembrane domain
Domain structure human DPP-IV
Residues 1-766; human DPP-IV
Residues 1-6; cytoplasmic tail
Residues 7-28; transmembrane region human DPP-IV
Residues 29-39; region connecting transmembrane region and α/β-hydrolase domain
Residues 39-766; Generally secreting soluble form of DPP-IV
Domains in the DPP-IV ectodomain:
Residues 55-497; β-propellor domain in green
Residues 39-51 and 506-766; α/β-hydrolase domain in grey
Domain borders based on crystal structure (Rasmussen Nat.Struct.Biol. 2003)

Figure 19

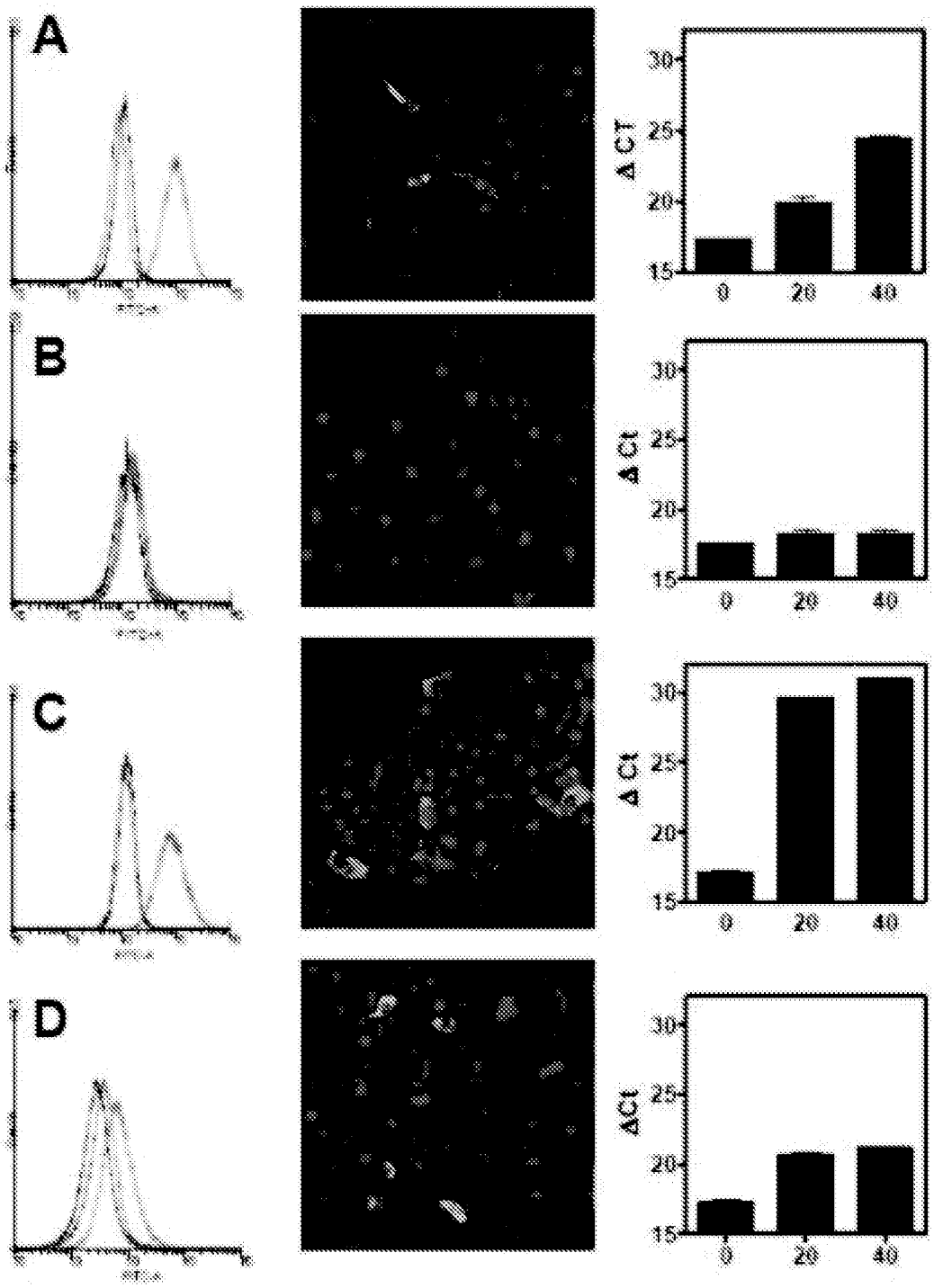


Figure 20

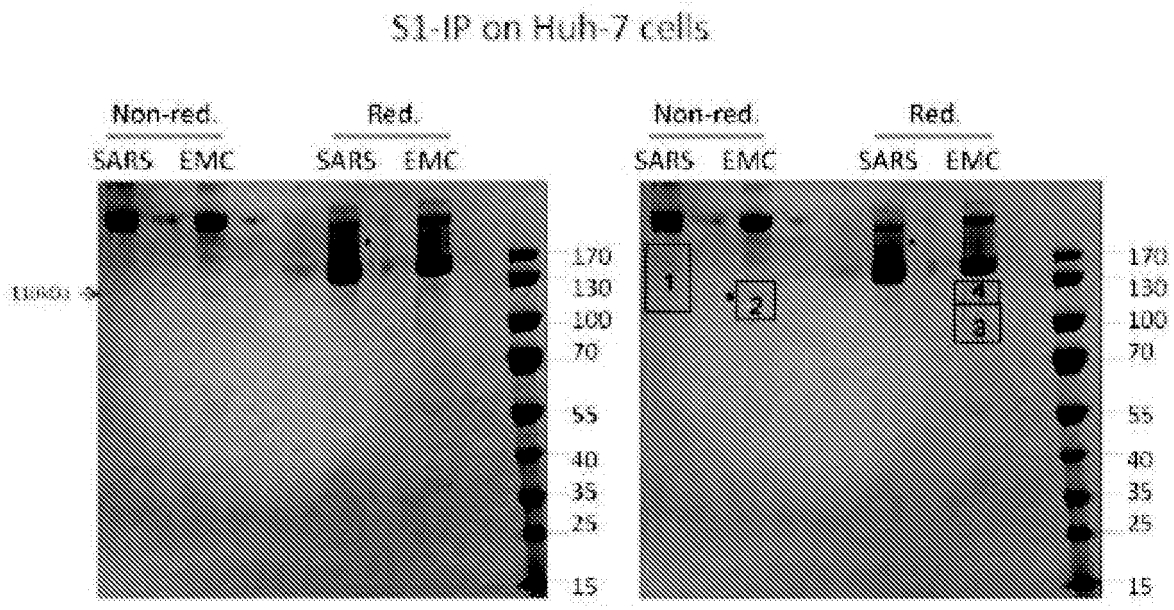

S1-IP on Huh-7 cells
Non-red.
Red.
SARS
EMC
SARS
EMC
Non-red.
Red.
SARS
EMC
SARS
EMC
170
130
100
70
55
40
35
25
15
1
2
3
4
170
130
100
70
55
40
35
25
15

Figure 21

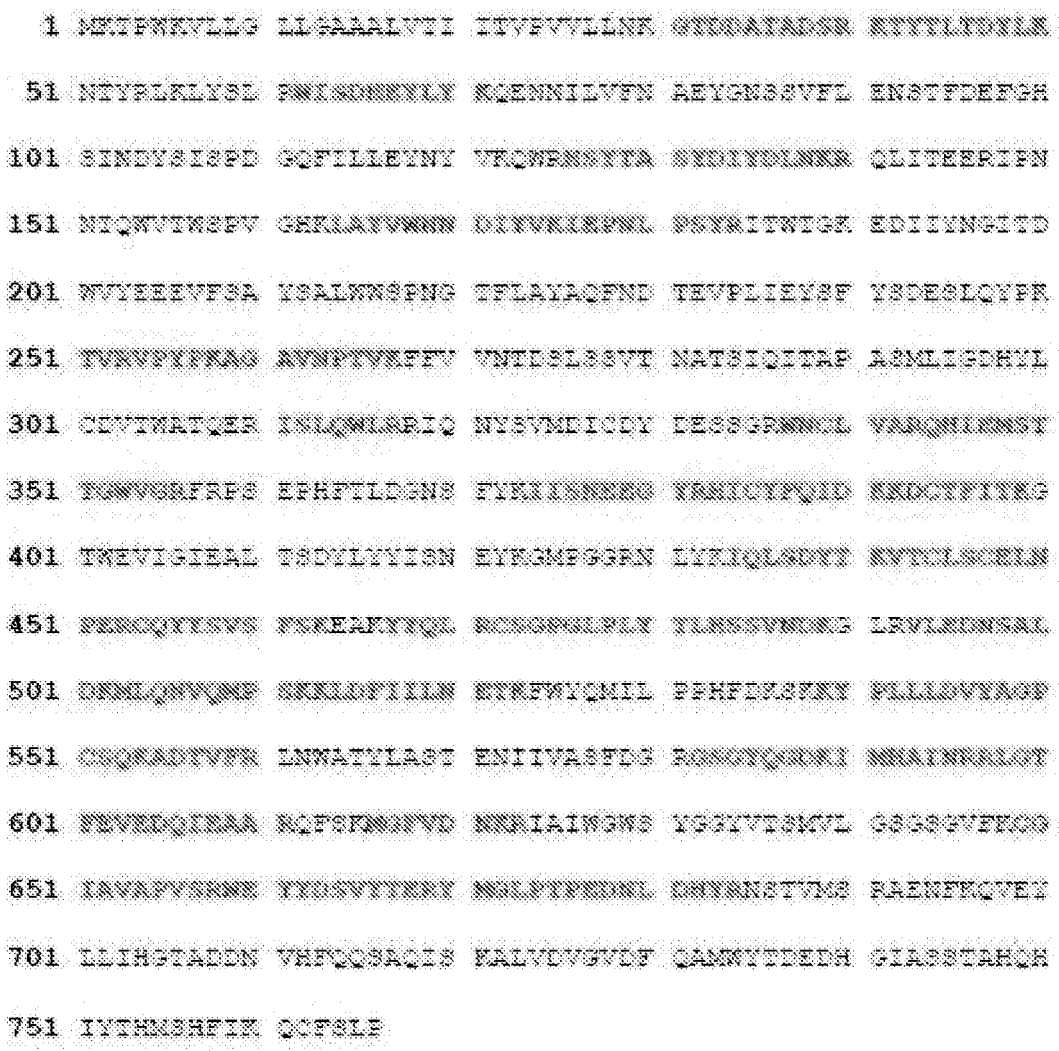


Figure 22

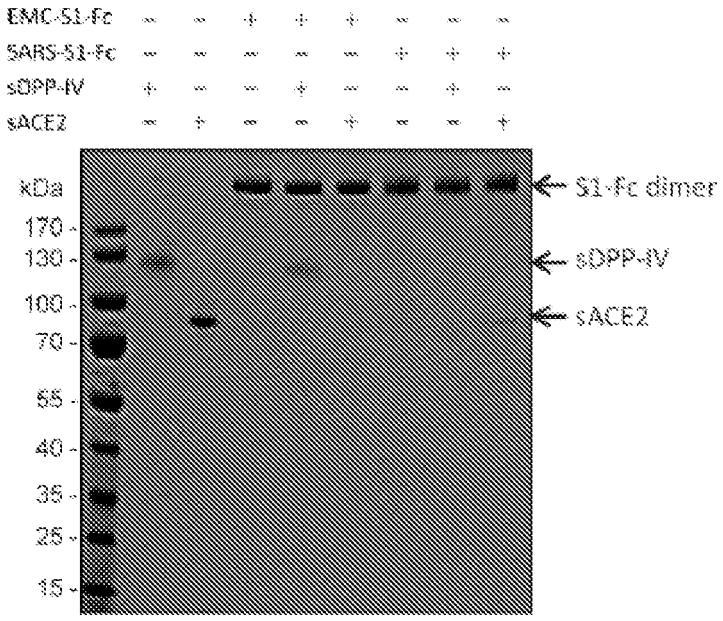

EMC-S1-Fc
SARS-S1-Fc
sDPP-IV
sACE2
kDa
170
130
100
70
55
40
35
25
15
S1-Fc dimer
sDPP-IV
sACE2

Figure 23

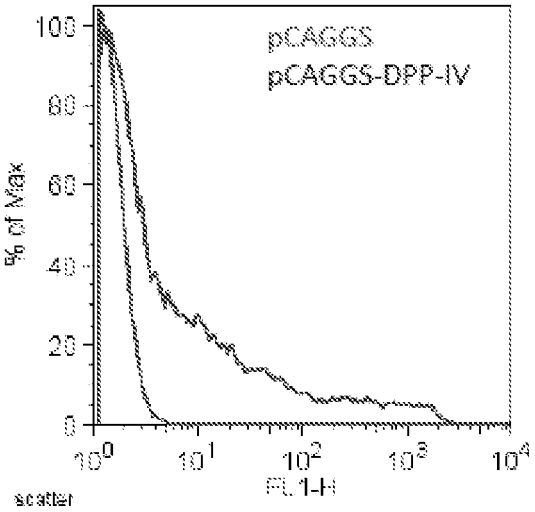

pCAGGS
pCAGGS-DPP-IV
% of Max
100
80
60
40
20
0
10^0
10^1
10^2
10^3
10^4
FL1-H
scatter

Figure 24

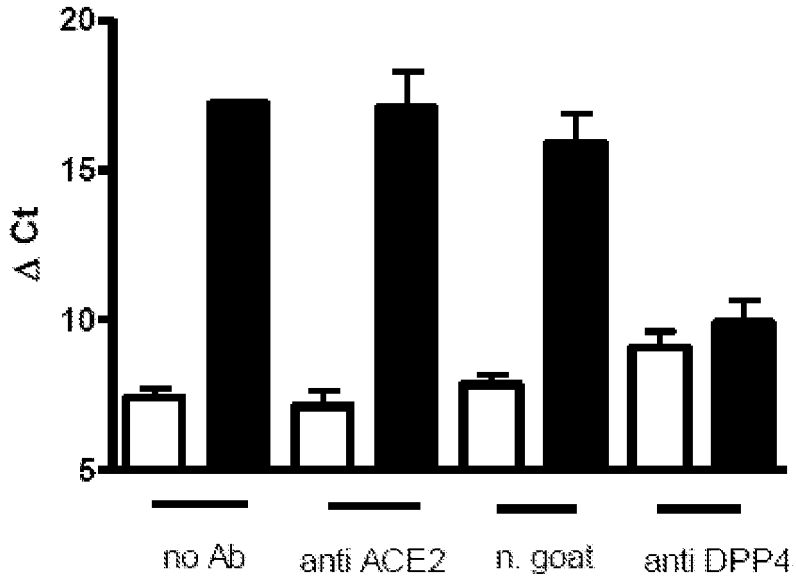


Figure 25

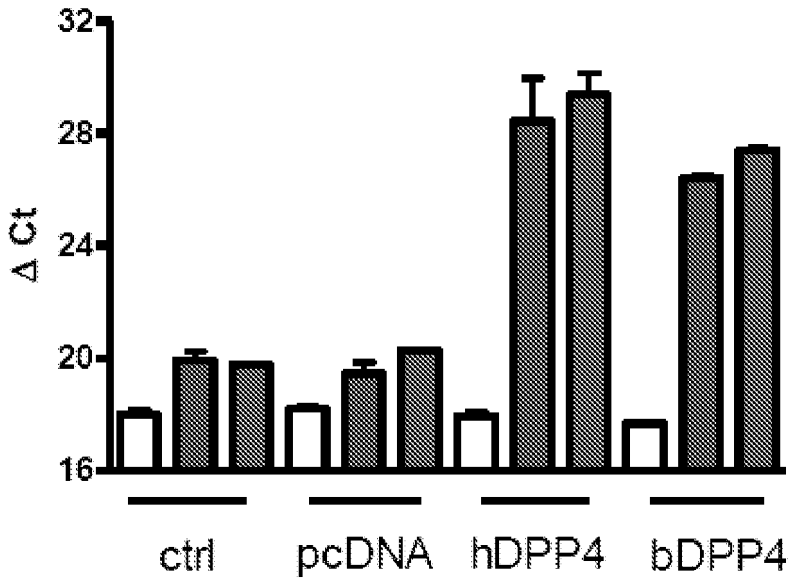


Figure 26

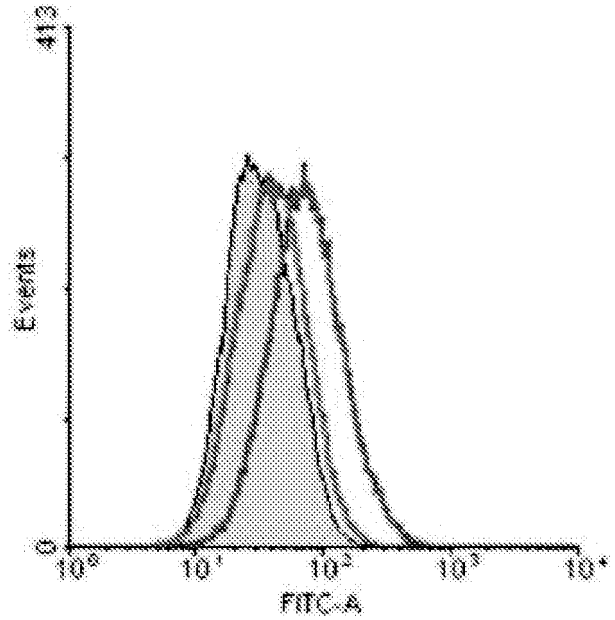


Figure 27

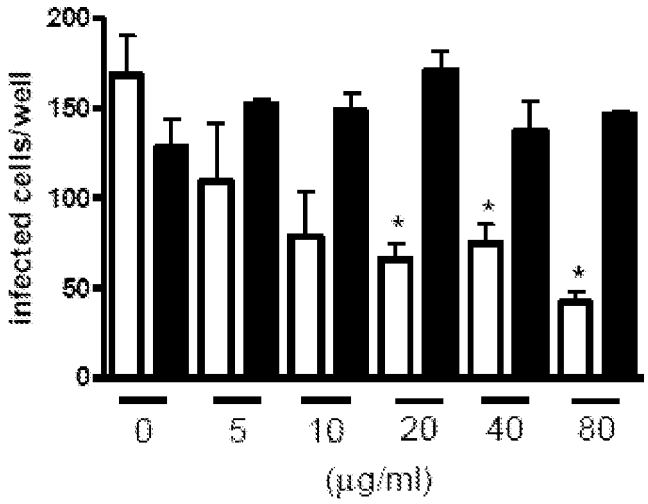


Figure 28

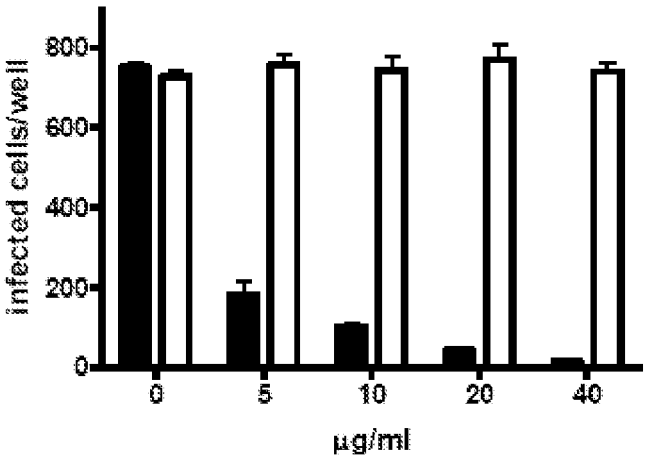


Figure 29

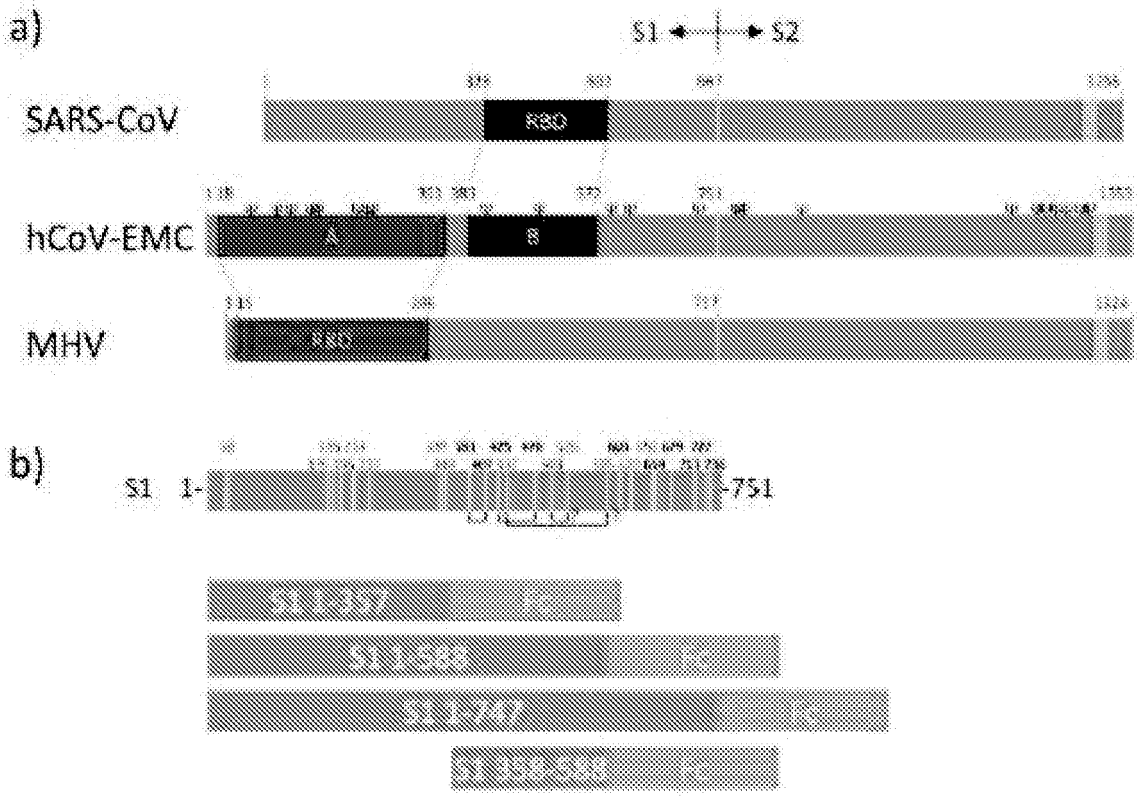


Figure 30

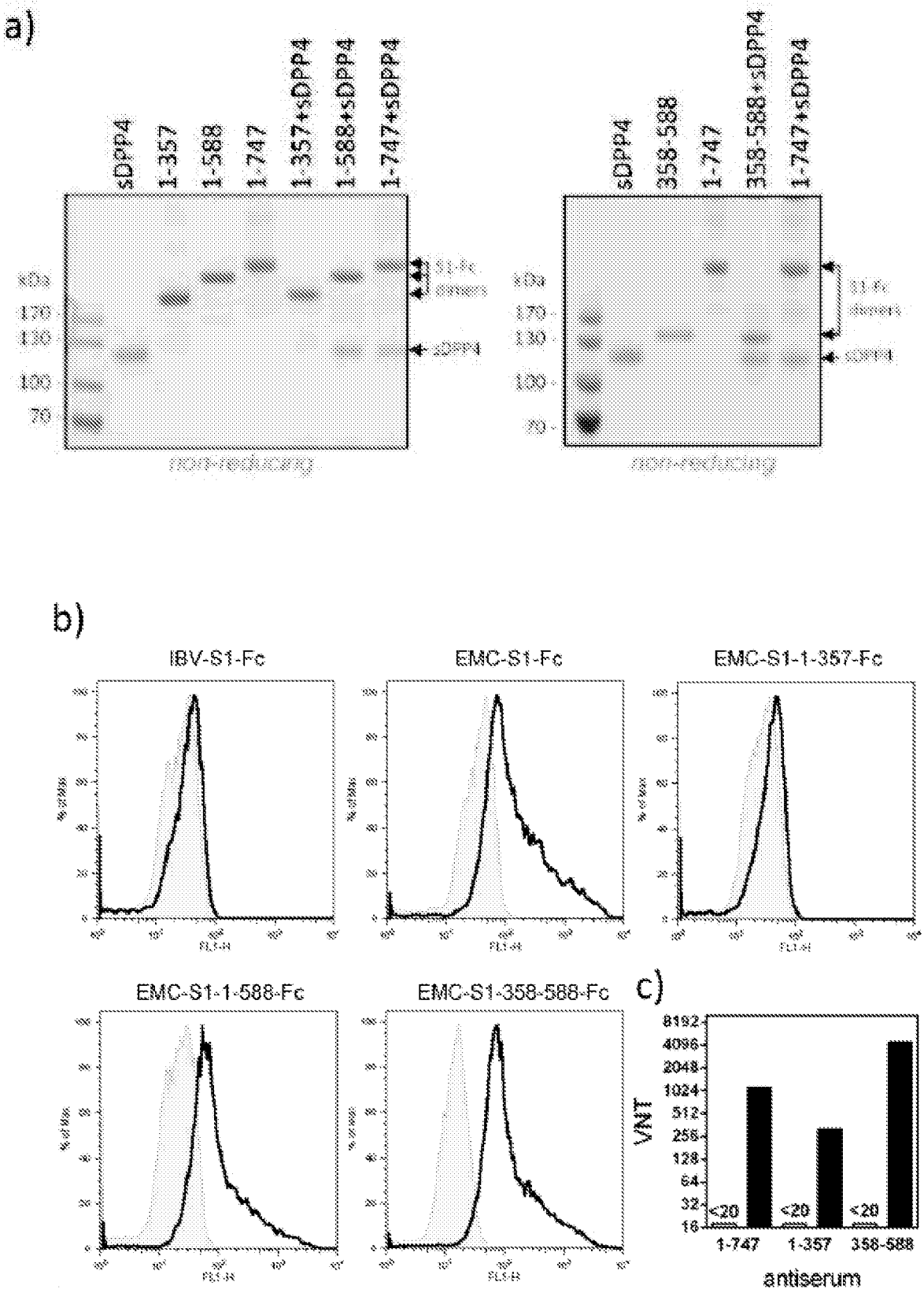

a)
sDPP4
1-357
1-588
1-747
1-357+sDPP4
1-588+sDPP4
1-747+sDPP4
kDa
170
130
100
70
S1-Fc dimers
sDPP4
non-reducing
sDPP4
358-588
1-747
358-588+sDPP4
1-747+sDPP4
kDa
170
130
120
80
S1-Fc dimers
sDPP4
non-reducing
b)
IBV-S1-Fc
EMC-S1-Fc
EMC-S1-1-357-Fc
EMC-S1-1-588-Fc
EMC-S1-358-588-Fc
c)
VNT
8192
4096
2048
1024
512
256
128
64
32
16
<20
<20
<20
1-747
1-357
358-588
antiserum

Figure 31

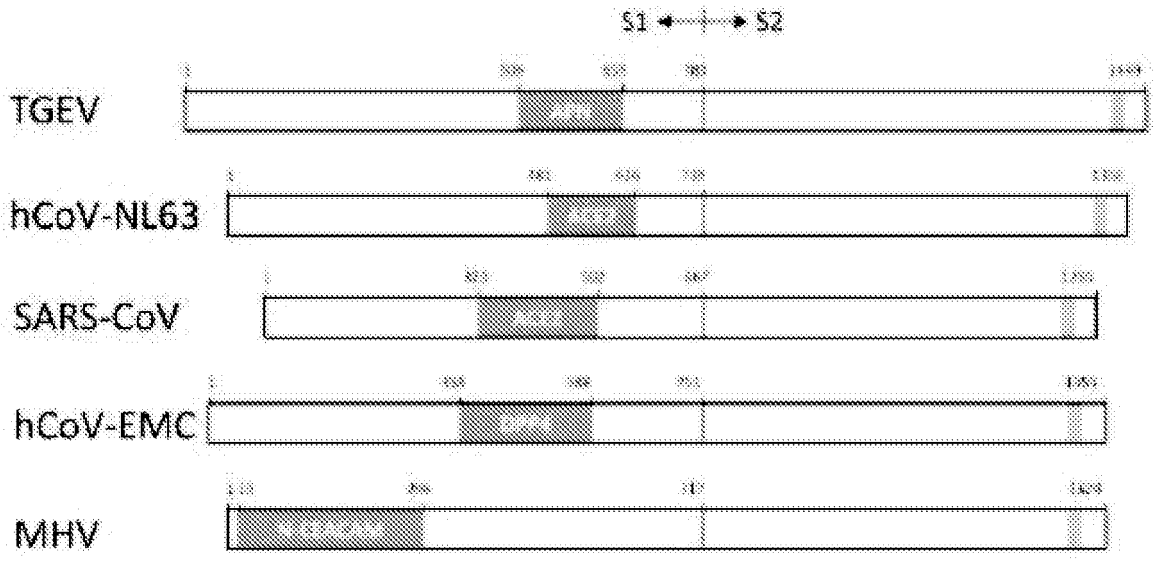
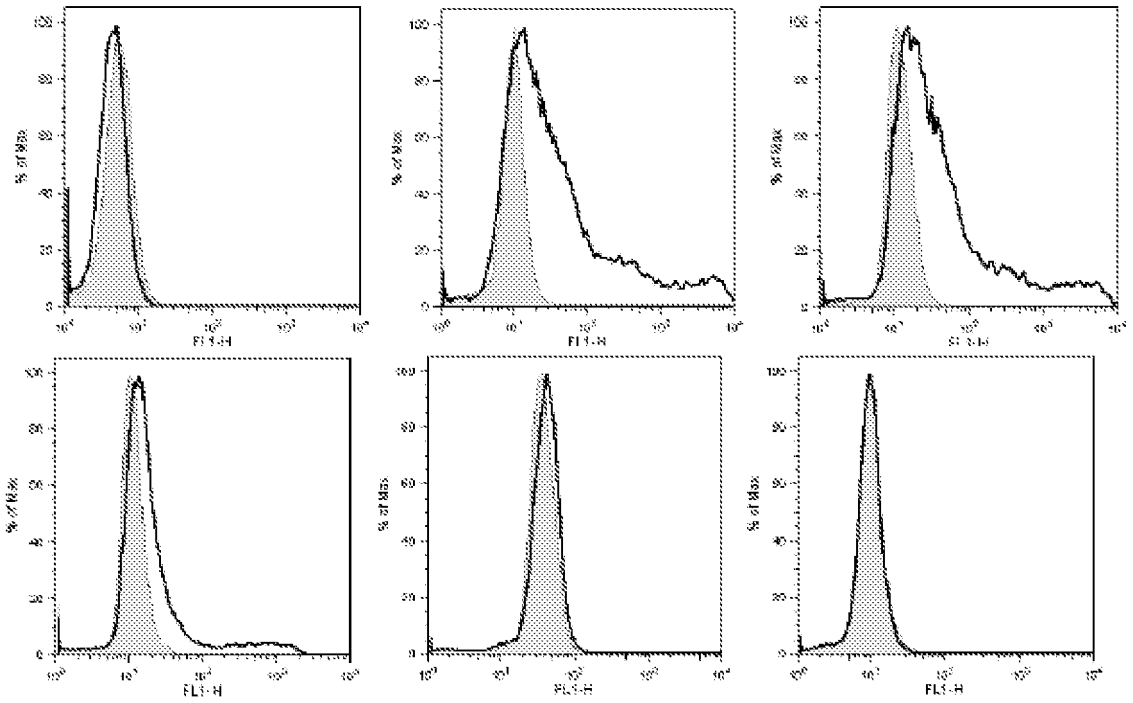


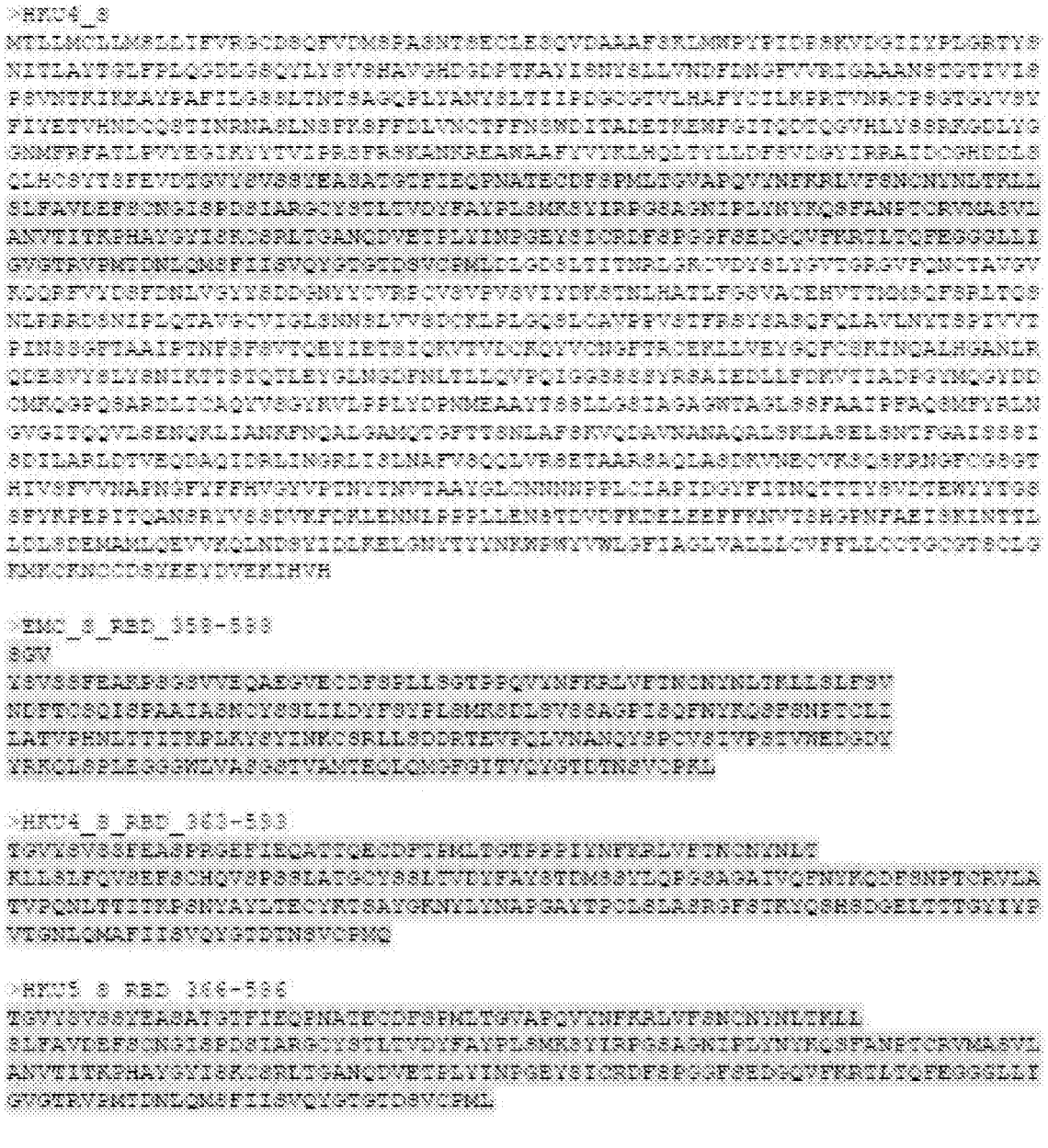
Figure 32



>EMC_S
MIHSVFLLMFLLTPTESYVDVGPDSVKSACIEVDIQQTFFDKTWPRPIDVSKADGIIYPQ
GRTYSNITITYQGLFPYQGDHGDMYVYSAGHATGTTPQKLFVANYSQDVKQFANGFVVRI
GAAANSTGTVIISPSTSATIRKIYPAFMLGSSVGNFSDGKMGRFFNHTLVLLPDGCGTLL
RAFYCILEPRSGNHCPAGNSYTSFATYHTPATDCSDGNYNRNASLNSFKEYFNLRNCTFM
YTYNITEDEILEWFGITQTAQGVHLFSSRYVDLYGGNMFQFATLPVYDTIKYYSIIPHSI
RSIQSDRKAWAAFYVYKLQPLTFLLDFSVDGYIRRAIDCGFNDLSQLHCSYESFDVESGV
YSVSSFEAKPSGSVVEQAEGVECDFSPLLSGTPPQVYNFKRLVFTNCNYNLTKLLSLFSV
NDFTCSQISPAAIASNCYSSLILDYFSYPLSMKSDLSVSSAGPISQFNYKQSFSNPTCLI
LATVPHNLTTITKPLKYSYINKCSRLLSDDRTEVPQLVNANQYSPCVSIVPSTVWEDGDY
YRKQLSPLEGGGWLVASGSTVAMTEQLQMGFGITVQYGTDTNSVCPKLEFANDTKIASQL
GNCVEYSLYGVSGRGVFQNCTAVGVRQQRFVYDAYQNLVGYYSDDGNYYCLRACVSVPVS
VIYDKETKTHATLFGSVACEHISSTMSQYSRSTRSMLKRRDSTYGPLQTPVGCVLGLVNS
SLFVEDCKLPLGQSLCALPDTPSTLTPRSVRSVPGEMRLASIAFNHPIQVDQLNSSYFKL
SIPTNFSFGVTQEYIQTTIQKVTVDCKQYVCNGFQKCEQLLREYGQFCSKINQALHGANL
RQDDSVRNLFASVKSSQSSPIIPGFGGDFNLTLLEPVSISTGSRSARSAIEDLLFDKVTI
ADPGYMQGYDDCMQQGPASARDLICAQYVAGYKVLPPLMDVNMEAAYTSSLLGSIAGVGW
TAGLSSFAAIPFAQSIFYRLNGVGITQQVLSENQKLIANKFNQALGAMQTGFTTTNEAFQ
KVQDAVNNNAQALSKLASELSNTFGAISASIGDIIQRLDVLEQDAQIDRLINGRLTTLNA
FVAQQLVRSESAALSAQLAKDKVNECVKAQSKRSGFCGQGTHIVSFVVNAPNGLYFMHVG
YYPSNHIEVVSAYGLCDAANPTNCIAPVNGYFIKTNNTRIVDEWSYTGSSFYAPEPITSL
NTKYVAPQVTYQNISTNLPPPLLGNSTGIDFQDELDEFFKNVSTSIPNFGSLTQINTTLL
DLTYEMLSLQQVVKALNESYIDLKELGNYTYYNKWPWYIWLGFIAGLVALALCVFFILCC
TGCGTNCMGKLKCNRCCDRYEEYDLEPHKVHVH

>HKU5_S
MIRSVLVLMCSLTFIGNLTRGQSVDMGHNGTGSCLDSQVQPDYFESVHTTWPMPIDTSKAEGVIYPNGRS
YSNITLTYTGLYPKANDLGKQYLFSDGHSAPGRLNNLFVSNYSSQVESFDDGFVVRIGAAANKTGTTVIS
QSTFKPIKKIYPAFLLGHSVGNYTPSNRTGRYLNHTLVILPDGCGTILHAFYCVLHPRTQQNCAGETNFK
SLSLWDTPASDCVSGSYNQEATLGAFKVYFDLINCTFRYNYTITEDENAEWFGITQDTQGVHLYSSRKEN
VFRNNMFHFATLPVYQKILYYTVIPRSIRSPFNDRKAWAAFYIYKLHPLTYLLNFDVEGYITKAVDCGYD
DLAQLQCSYESFEVETGVYSVSSFEASPRGEFIEQATTQECDFTPMLTGTPPPIYNFKRLVFTNCNYNLT
KLLSLFQVSEFSCHQVSPSSLATGCYSSLTVDYFAYSTDMSSYLQPGSAGAIVQFNYKQDFSNPTCRVLA
TVPQNLTTITKPSNYAYLTECYKTSAYGKNYLYNAPGAYTPCLSLASRGFSTKYQSHSDGELTTTGYIYP
VTGNLQMAFIISVQYGTDTNSVCPMQALRNDTSIEDKLDVCVEYSLHGITGRGVFHNCTSVGLRNQRFVY
DTFDNLVGYHSDNGNYYCVRPCVSVPVSVIYDKASNSHATLFGSVACSHVTTMMSQFSRMTKTNLLARTT
PGPLQTTVGCAMGFINSSMVVDECQLPLGQSLCAIPPTTSSRVRRATSGASDVFQIATLNFTSPLTLAPI
NSTGFVVAVPTNFTFGVTQEFIETTIQKITVDCKQYVCNGFKKCEDLLKEYGQFCSKINQALHGANLRQD
ESIANLFSSIKTQNTQPLQAGLNGDFNLTMLQIPQVTTGERKYRSTIEDLLFNKVTIADPGYMQGYDECM
QQGPQSARDLICAQYVAGYKVLPPLYDPYMEAAYTSSLLGSIAGASWTAGLSSFAAIPFAQSIFYRLNGV
GITQQVLSENQRIIANKFNQALGAMQTGFTTTNLAFNKVQDAVNANAMALSKLAAELSNTFGAISSSISD
ILARLDTVEQEAQIDRLINGRLTSLNAFVAQQLVRTEAAARSAQLAQDKVNECVKSQSKRNGFCGTGTHI
VSFAINAPNGLYFFHVGYQPTSHVNATAAYGLCNTENPQKCIAPIDGYFVLNQTTSTVADSDQQWYYTGS
SFFHPEPITEANSKYVSMDVKFENLTNRLPPPLLSNSTDLDFKEELEEFFKNVSSQGPNFQEISKINTTL
LNLNTELMVLSEVVKQLNESYIDLKELGNYTFYQKWPWYIWLGFIAGLVALALCVFFILCCTGCGTSCLG
KLKCNRCCDSYDEYEVEKIHVH

Figure 32 continued



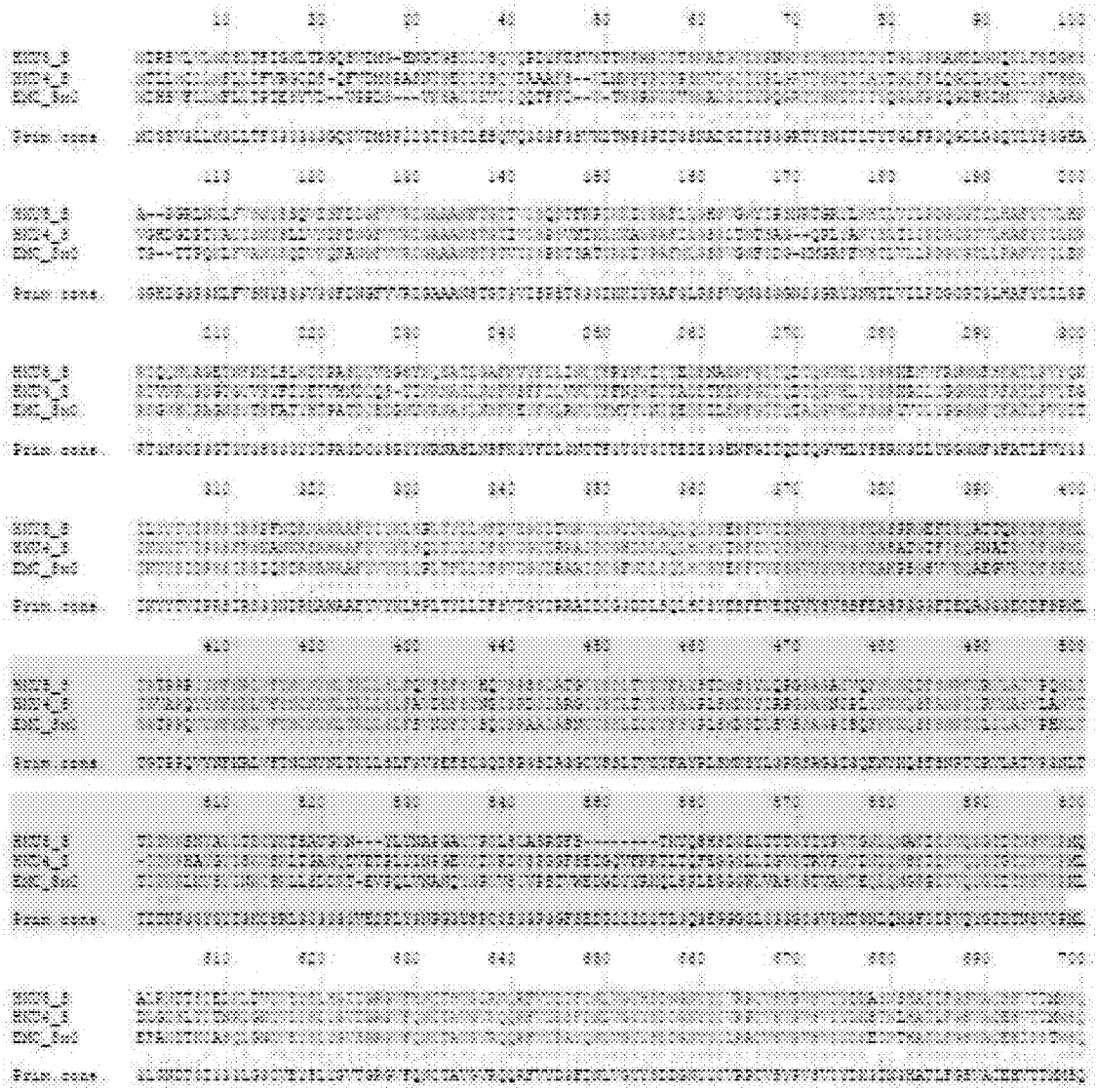
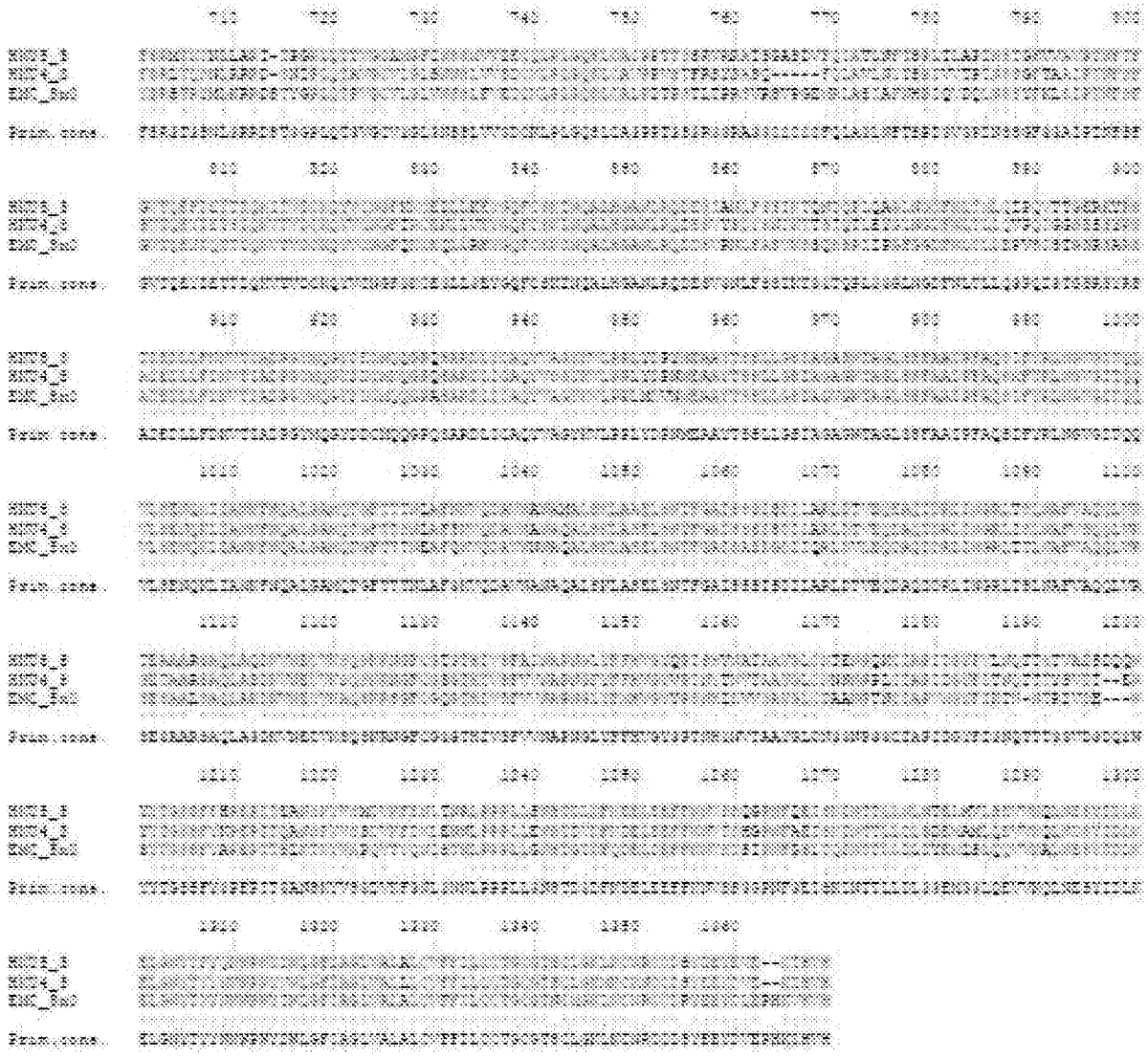
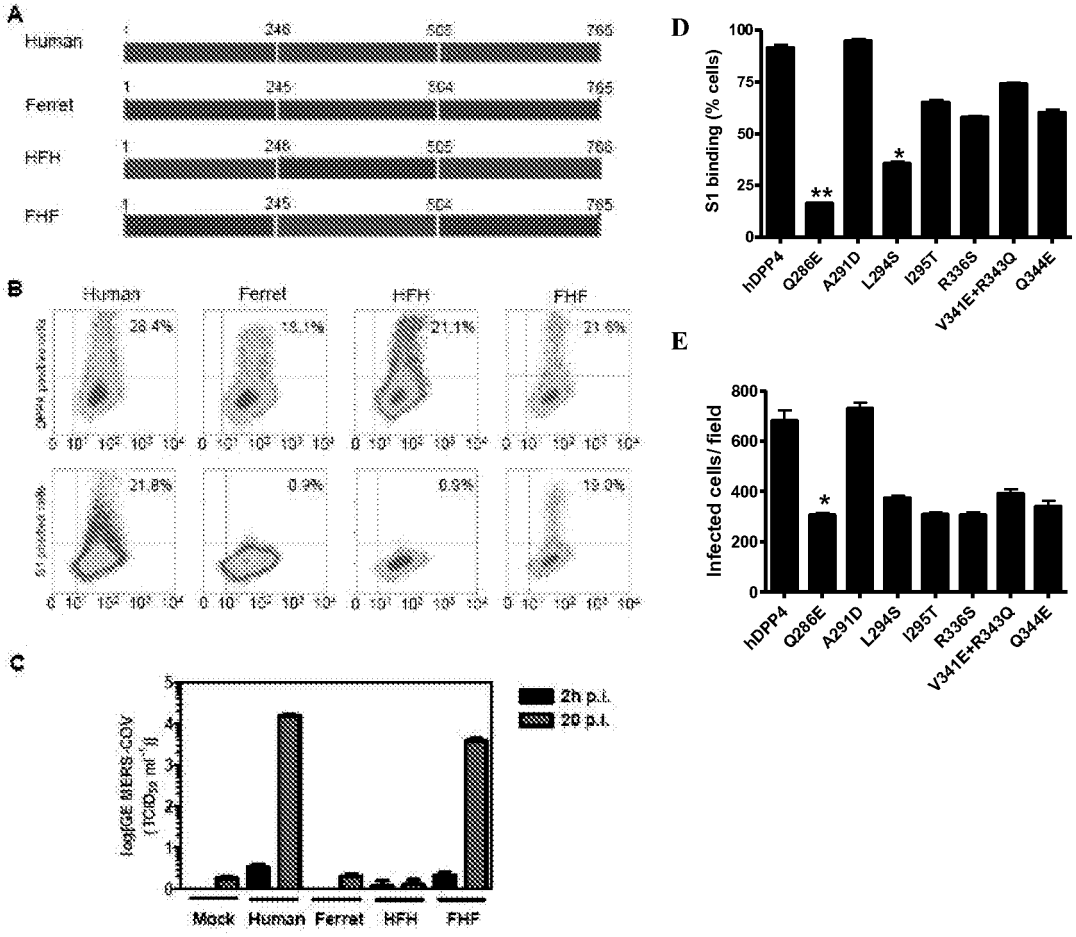


Figure 32 continued



Alignment data :
Alignment length : 1369
Identity (*) : 766 is 55.95 %
Strongly similar (:) : 233 is 17.02 %
Weakly similar (.) : 110 is 8.04 %
Different : 260 is 18.99 %
Sequence 0001 : HXU5_S (1352 residues)
Sequence 0002 : HXU4_S (1352 residues)
Sequence 0003 : EMC_Sx0 (1353 residues)

Figure 33


A
Human
Ferret
HFH
FHF
B
Human
Ferret
HFH
FHF
C
2h p.i.
20 p.i.
Mock
Human
Ferret
HFH
FHF
D
S1 binding (% cells)
hDPP4
Q286E
A291D
L294S
I295T
R336S
V341E+R343Q
Q344E
E
Infected cells/ field
hDPP4
Q286E
A291D
L294S
I295T
R336S
V341E+R343Q
Q344E

REFERENCES CITED IN THE DESCRIPTION

This list of references cited by the applicant is for the reader's convenience only. It does not form part of the European patent document. Even though great care has been taken in compiling the references, errors or omissions cannot be excluded and the EPO disclaims all liability in this regard.

Patent documents cited in the description

- US 4683195 A, Mullis **[0064]**
- US 4683202 A **[0064]**
- US 4800159 A **[0064]**
- EP 320308 A **[0064]**
- US 5270184 A **[0064]**
- US 5455166 A **[0064]**
- US 5871921 A **[0064]**
- US 5719028 A **[0064]**
- US 5942391 A **[0064]**
- US 7344852 B **[0175]**
- US 20050260723 A **[0175]**
- US 6296673 B **[0183]**

Non-patent literature cited in the description

- UNIT 17.5 Protein Protein Interactions Identified by Pull Down Experiments and Mass Spectrometry. **ADAM BRYMORA ; VALENTINA A. VALOVA ; PHILLIP J. ROBINSON.** Cell Biology. Current Protocols, 01 May 2004 **[0015]**
- **HARLOW ; LANE.** Antibodies, A Laboratory Manual. Cold Spring Harbor Publications, 1988 **[0040] [0074]**
- **SUI et al.** *Proc. Natl. Acad. Sci.,* 2004, vol. 101 (8), 2536-2541 **[0060]**
- **BARANY.** *Proc. Natl. Acad. Sci. USA,* 1991, vol. 88, 189-193 **[0064]**
- **GUATELLI et al.** *Proc. Natl. Acad. Sci. USA,* 1990, vol. 87, 1874-1878 **[0064]**
- **KWOH et al.** *Proc. Natl. Acad. Sci. USA,* vol. 86, 1173-1177 **[0064]**
- **LIZARDI et al.** *Bio/Technology,* 1988, vol. 6, 1197 **[0064]**
- **JACOBS et al.** *J. Clin. Microbiol.,* 1997, vol. 35, 791-795 **[0068]**
- **VAN DEN BRULE et al.** *J. Clin. Microbiol.,* 2002, vol. 40, 779-787 **[0071]**
- **MEINKOTH ; WAHL.** *Anal. Biochem.,* 1984, vol. 138, 267-284 **[0072]**
- Overview of principles of hybridization and the strategy of nucleic acid probe assays. **TIJSSEN.** Laboratory Techniques in Biochemistm and Molecular Biology--Hybridization with Nucleic Acid Probes. Elsevier, 1993 **[0072]**
- Current Protocols in Molecular Biology. Greene Publishing and Wiley-Interscience, 1995 **[0072]**
- **ALTSCHUL, S. F. et al.** *J. Mol. Biol.,* 1990, vol. 215, 403 **[0085]**
- **ALTSCHUL, S. F. et al.** *Nucleic Acid Res.,* 1997, vol. 25, 3389-3402 **[0085]**
- **AUSUBEL et al.** Current Protocols in Molecular Biology. John Wiley & Sons Inc, 1994 **[0091]**
- **VAN BOHEEMEN,S. ; DE GRAAF,M. ; LAUBER,C. ; BESTEBROER,T.M. ; RAJ,V.S. ; ZAKI,A.M. ; OSTERHAUS,A.D. ; HAAGMANS,B.L. ; GORBALENYA,A.E. ; SNIJDER,E.J.** Genomic characterization of a newly discovered coronavirus associated with acute respiratory distress syndrome in humans. *JOURNAL MBio,* 2012, vol. 3 (6), e00473-12 **[0096]**
- **SUI et al.** *Proc. Natl. Acad. Sci.,* 2004, vol. 101 (8), 2536-2541 **[0132]**
- **RASMUSSEN.** *Nat. Struct. Biol.,* 2003 **[0149]**
- **VIJGEN, L. ; E. MOES ; E. KEYAERTS ; S. LI ; M. VAN RANST.** A pancoronavirus RT-PCR assay for detection of all known coronaviruses. *Methods Mol Biol,* 2008, vol. 454, 3-12 **[0155]**
- **DROSTEN C ; GÜNTHER S ; PREISER W ; VAN DER WERF S ; BRODT HR ; BECKER S ; RABENAU H ; PANNING M ; KOLESNIKOVA L ; FOUCHIER RA.** Identification of a novel coronavirus in patients with severe acute respiratory syndrome. *N Engl J Med.,* 2003, vol. 348, 1967-76 **[0155] [0156]**
- **WELSH, J. ; MCCLELLAND, M.** Fingerprinting genomes using PCR with arbitrary primers. *Nucleic Acids Res.,* 1990, vol. 18, 7213-7218 **[0159]**
- **LIET ; 2003.** *Nature,* vol. 426, 450 **[0168]**
- **VAN DEN BERG et al.** *Cell Stem Cell,* vol. 6, 369 **[0172]**
- **WARD et al.** *J. Mol. Biol.,* 1975, vol. 98, 161 **[0181]**
- **MCPHERSON.** *J. Biol. Chem.,* 1976, vol. 251, 6300 **[0181]**
- **T.L. BLUNDELL et al.** PROTEIN CRYSTALLOGRAPHY. Academic Press, 1976, vol. 59, 82 **[0184]**
- **J. S. PEIRIS ; Y. GUAN ; K. Y. YUEN.** *Nat. Med.,* 2004, vol. 10, 88-97 **[0231]**
- **A. M. ZAKI ; S. VAN BOHEEMEN ; T. M. BESTEBROER ; A. D. OSTERHAUS ; R. A. FOUCHIER.** *N. Engl. J. Med.,* 2012, vol. 367, 1814-1820 **[0231]**

- **S. VAN BOHEEMEN et al.** *MBio,* 2012, vol. 3, 00473-12 **[0231]**
- Middle East respiratory syndrome coronavirus. MERS-CoV. World Health Organisation, 08 August 2013 **[0231]**
- **A. ASSIRI et al.** *N. Engl. J. Med.,* 2013, vol. 369, 407-416 **[0231]**
- **B. GUERY et al.** *Lancet,* 2013, vol. 381, 2265-2272 **[0231]**
- **C. B. REUSKEN et al.** *Vector Borne Zoonotic Dis,* 2010, vol. 10, 785-791 **[0231]**
- **A. ANNAN et al.** *Emerg Infect Dis,* 2013, vol. 19, 456-459 **[0231]**
- **M. A. MULLER et al.** *MBio,* 2012, vol. 3, 00515-12 **[0231]**
- **J. F. CHAN et al.** *J. Infect. Dis.,* 2013, vol. 207, 1743-1752 **[0231]**
- **E. DE WIT et al.** *PLoS One,* 2013, vol. 8, e69127 **[0231]**
- **V. S. RAJ et al.** *Nature,* 2013, vol. 495, 251-254 **[0231]**
- **B. E. MARTINA et al.** *Nature,* 2003, vol. 425, 915 **[0231]**
- **J. A. BELSER ; J. M. KATZ ; T. M. TUMPEY.** *Dis. Model. Mech.,* 2011, vol. 4, 575-579 **[0231]**
- **E. BOONACKER ; C. J. VAN NOORDEN.** *Eur. J. Cell. Biol.,* 2003, vol. 82, 53-73 **[0231]**
- **W. A. WEIHOFEN ; J. LIU ; W. REUTTER ; W. SAENGER ; H. FAN.** *J. Biol. Chem.,* 2004, vol. 279, 43330-43335 **[0231]**
- **R.P. DONG et al.** *J. Immunol.,* 1996, vol. 156, 1349-1355 **[0231]**
- **G. LU et al.** *Nature,* 2013 **[0231]**
- **N. WANG et al.** *Cell. Res.,* 2013 **[0231]**
- **B. ROCKX et al.** *J. Virol.,* 2009, vol. 83, 7062-7074 **[0231]**
- **S. L. SMITS et al.** *PLoS Pathog.,* 2010, vol. 6, e1000756 **[0231]**
- **J. ZHONG et al.** *Diabetes,* 2013, vol. 62, 149-157 **[0231]**
- **B. A. CONLON ; W. R. LAW.** *Clin. Exp. Immunol.,* 2004, vol. 138, 14-20 **[0231]**
- **S. A. LEE et al.** *J. Clin. Endocrinol. Metab.,* 2013, vol. 98, 2553-2561 **[0231]**
- **T. ANDRIEU et al.** *J. Clin. Virol.,* 2003, vol. 27, 59-68 **[0231]**
- **V. J. MUNSTER ; E. DE WIT ; H. FELDMANN.** *N. Engl. J. Med.,* 2013, vol. 368, 1560-1562 **[0231]**
- **S. L. SMITS et al.** *J. Virol.,* 2011, vol. 85, 4234-4245 **[0231]**
- **V. M. CORMAN et al.** *Euro. Surveil,* 2012, vol. l 17, 20334 **[0231]**