Microbiol. Immunol., **49**(10), 899–908, 2005

In Silico Analysis of ORF1ab in Coronavirus HKU1 Genome Reveals a Unique Putative Cleavage Site of Coronavirus HKU1 3C-Like Protease

Patrick C.Y. Woo[1, 2, 3, †], Yi Huang[1, †], Susanna K.P. Lau[1, 2, 3], Hoi-wah Tsoi[1], and Kwok-yung Yuen*[, 1, 2, 3]

[1]*Department of Microbiology, The University of Hong Kong, University Pathology Building, Queen Mary Hospital, Hong Kong,* [2]*Research Centre of Infection and Immunology, Faculty of Medicine, The University of Hong Kong, and* [3]*State Key Laboratory of Emerging Infectious Diseases (The University of Hong Kong), Hong Kong*

Received March 23, 2005; in revised form, July 12, 2005. Accepted July 28, 2005

***Abstract*: Recently we have described the discovery and complete genome sequence of a novel coronavirus associated with pneumonia, coronavirus HKU1 (CoV-HKU1). In this study, a detailed *in silico* analysis of the ORF1ab, encoding the 7,182-amino acid replicase polyprotein in the CoV-HKU1 genome showed that the replicase polyprotein of CoV-HKU1 is cleaved by its papain-like proteases and 3C-like protease (3CL^{pro}) into 16 polypeptides homologous to the corresponding polypeptides in other coronaviruses. Surprisingly, analysis of the putative cleavage sites of the 3CL^{pro} revealed a unique putative cleavage site. In all known coronaviruses, the P1 positions at the cleavage sites of the 3CL^{pro} are occupied by glutamine. This is also observed in CoV-HKU1, except for one site at the junction between nsp10 (helicase) and nsp11 (member of exonuclease family), where the P1 position is occupied by histidine. This amino acid substitution is due to a single nucleotide mutation in the CoV-HKU1 genome, CAG/A to CAT. This probably represents a novel cleavage site because the same mutation was consistently observed in CoV-HKU1 sequences from multiple specimens of different patients; the P2 and P1′-P12′ positions of this cleavage site are consistent between CoV-HKU1 and other coronaviruses; and as the helicase is one of the most conserved proteins in coronaviruses, cleavage between nsp10 and nsp11 should be an essential step for the generation of the mature functional helicase. Experiments, including purification and C-terminal amino acid sequencing of the CoV-HKU1 helicase and trans-cleavage assays of the CoV-HKU1 3CL^{pro} will confirm the presence of this novel cleavage site.**

***Key words*: Coronavirus HKU1, ORF1ab, 3C-like protease**

Coronaviruses, a genus of the family *Coronaviridae*, are large, enveloped, positive-stranded RNA viruses that cause respiratory, enteric, hepatic, and neurological diseases in humans and domestic animals. In humans, it has been estimated that coronaviruses [human coronaviruses 229E (HCoV-229E) and OC43 (HCoV-OC43)] are responsible for causing about 5–30% of respiratory tract infections. In late 2002 and 2003, a novel clinical syndrome, Severe Acute Respiratory Syndrome (SARS), affected 30 countries on five continents with more than 8,000 cases and 750 deaths. A novel coronavirus, the SARS coronavirus (SARS-CoV), has been confirmed to be the aetiological agent, and its genome has been completely sequenced (13, 28, 32, 36, 37, 39,

**Address correspondence to* Dr. Kwok-yung Yuen, Department of Microbiology, The University of Hong Kong, Room 423 University Pathology Building, Queen Mary Hospital Compound, Pokfulam Road, Hong Kong. Fax: +852–2855–1241. E-mail: hkumicro@hkucc.hku.hk

[†]These authors contributed the same to the manuscript.

Abbreviations: Appr>p, ADP-ribose 1″,2″-cyclic phosphate; Appr-1″p, ADP-ribose 1″-phosphate; ADRP, ADP-ribose 1″-phosphatase; BCoV, bovine coronavirus; 3CL^{pro}, 3C-like protease; CoV-HKU1, coronavirus HKU1; CPDase, cyclic nucleotide phosphodiesterase; E, envelope; ExoN, 3′-to-5′ exonuclease; HCoV-229E, human coronavirus 229E; HCoV-NL63, human coronavirus NL63; HCoV-OC43, human coronavirus OC43; IBV, infectious bronchitis virus; M, membrane; MHV, murine hepatitis virus; N, nucleocapsid; 2′-O-MT, *S*-adenosylmethionine-dependent ribose 2′-*O*-methyltransferase; ORF, open reading frame; PL^{pro}, papain-like proteases; *pol*, RNA-dependent RNA polymerase; RT-PCR, reverse transcriptase polymerase chain reaction; S, spike; SARS, Severe Acute Respiratory Syndrome; SARS-CoV, SARS coronavirus; XendoU, poly(U)-specific endoribonuclease.

51–53). In early 2004, a novel human coronavirus associated with upper respiratory tract infections, human coronavirus NL63 (HCoV-NL63), was also discovered (11, 47). Phenotypically, the envelopes of all coronaviruses are studded with long, petal-shaped spikes, resulting in the appearance of a crown under the electron microscope. Genotypically, coronaviruses possess the largest genomes of all RNA viruses of about 30 kb. As a result of a unique mechanism of viral replication, coronaviruses have a high frequency of recombination (24, 25, 29, 31). The order of the genes in the genomes of all coronaviruses that encode the RNA-dependent RNA polymerase (*pol*) and the four structural proteins present in all coronaviruses is 5′-*pol*, spike (S), envelope (E), membrane (M), nucleocapsid (N)-3′. Based on serological and genotypic characterization, coronaviruses were divided into three distinct groups, with HCoV-229E and HCoV-NL63 being group 1 coronaviruses and HCoV-OC43 a group 2 coronavirus respectively (11, 26, 47).

Recently, we have described the discovery of a group 2 novel coronavirus associated with pneumonia, coronavirus HKU1 (CoV-HKU1), and its complete genome sequence (48–50). In the genome of CoV-HKU1, more than two thirds of the genome was made up of a single open reading frame (ORF), ORF1ab, which encodes the putative replicase polyprotein. In other coronaviruses, this polyprotein, after cleavage by proteases encoded by this ORF, gives rise to more than 10 proteins, of which many are important enzymes essential for the survival of the coronaviruses. In this article, we describe a detailed *in silico* analysis of this ORF in CoV-HKU1, which reveals a unique putative cleavage site of the 3C-like protease ($3CL^{pro}$) in coronavirus HKU1. Other similarities and differences of this ORF in CoV-HKU1 as compared to other coronaviruses are also discussed.

Materials and Methods

Viral sequences. The predicted amino acid sequences of ORF1ab in CoV-HKU1 were extracted from the CoV-HKU1 genome sequence (GenBank accession no. NC_006577) (48). The corresponding amino acid sequences of ORF1ab in other coronaviruses were extracted from GenBank (HCoV-OC43, GenBank accession no. NP_937947; BCoV, GenBank accession no. NP_150073; MHV A59, GenBank accession no. P16342; HCoV-229E, GenBank accession no. NP_073549; IBV, GenBank accession no. NP_066134; and SARS-CoV, GenBank accession no. NP_828849). The amino acid sequences of the RNA-dependent RNA polymerases in hepatitis C virus, rabbit hemorrhagic disease virus and poliovirus 1 and their secondary structures were retrieved from Protein DataBank (PDB ID 1QUV, 1KHV and 1RDR respectively).

In silico *analysis*. Multiple alignment was performed using ClustalX 1.83 (46). Protein family analysis was performed using PFAM (2). Secondary structure prediction was performed using PROFsec (38). Three-dimensional modeling of $3CL^{pro}$ of CoV-HKU1 was performed using $3CL^{pro}$ of SARS-CoV as the template (54). Manually corrected sequence alignment and homology modeling requirement were submitted via Deep View – spdbv 3.7 (14) to Swiss-model Protein Modeling web server (http://swissmodel.expasy.org/) (40). The 3D structure was displayed using MOLMOL 2K.2 (22). Phylogenetic tree construction was performed using neighbour joining method with ClustalX 1.83.

Results and Discussion

General Overview

The replicase polyprotein (7,182 amino acids) of CoV-HKU1 is translated from ORF1a (13,395 nt) and ORF1b (8,154 nt). Similar to other coronaviruses, a slippery sequence (UUUAAAC), followed by sequences that form a putative pseudoknot structure, are present (Fig. 1) (5, 6). Translation will presumably occur by a -1 RNA-mediated ribosomal frameshift at the end of the slippery sequence. Therefore, instead of reading the transcript as UUUAAACGGG, it will be read as UUUAAACCGGG. In infectious bronchitis virus (IBV) and SARS-CoV, site directed mutagenesis experiments have confirmed the importance of the slippery sequence as well as the pseudoknot structure for the frameshift to occur (7, 45).

Proteolytic Cleavage of Replicase Polyprotein

Multiple alignments between the replicase polyprotein of CoV-HKU1 and those of other coronaviruses reveal the orthologs of two types of proteases, papain-like proteases (PL^{pro}) and $3CL^{pro}$. As in other coronaviruses and by multiple alignment and analysis of sequences around consensus cleavage sites, the replicase polyprotein of CoV-HKU1 is cleaved by its PL^{pro} and $3CL^{pro}$ into 16 polypeptides (Table 1). The three N-terminal cleavage sites are putatively cleaved by the PL^{pro}, whereas the others are putatively cleaved by the $3CL^{pro}$.

The putative PL^{pro} cleavage sites in the replicase polyprotein of CoV-HKU1 follow the general rules determined by site directed mutagenesis experiments in other group 2 coronaviruses. In MHV, site directed mutagenesis studies have shown that at the p28/p65 cleavage site, a basic residue (arginine or lysine) in the

P5 position, arginine in the P2 position, and glycine in the P1 position are the critical amino acids for PL1^{pro} recognition and processing (17). In CoV-HKU1, arginine, arginine and glycine are present in these three positions respectively (Table 2). In MHV, site directed mutagenesis studies have shown that at the p65/nsp1 cleavage site, arginine in the P5 position, alanine in the P1 position, and glycine in the P1′ position are the critical amino acids for PL1^{pro} recognition and processing (4). In CoV-HKU1, arginine, alanine and glycine are present in these three positions respectively (Table 2). In MHV, site directed mutagenesis studies have shown that at the nsp1/p44 cleavage site, phenylalanine in the P6 position, glycine in the P2 position, and glycine in the P1 position are the critical amino acids for PL2^{pro} recognition and processing (20). In CoV-HKU1, phenylalanine, glycine and glycine are present in these three positions respectively (Table 2).

In contrast to the putative PL^{pro} cleavage sites, analysis of the putative cleavage sites of the 3CL^{pro} in the replicase polyprotein of CoV-HKU1 revealed a unique putative cleavage site by the 3CL^{pro} of CoV-HKU1. The 3CL^{pro} of coronaviruses are named as such because their structures and substrate specificities resemble those of the 3C proteases in picornaviruses. In all known coronaviruses, the P1 positions at the cleavage sites of the 3CL^{pro} are occupied by glutamine, whereas the P1′ positions are occupied by aliphatic amino acids, including alanine, glycine, cysteine, asparagine and serine (3, 9,

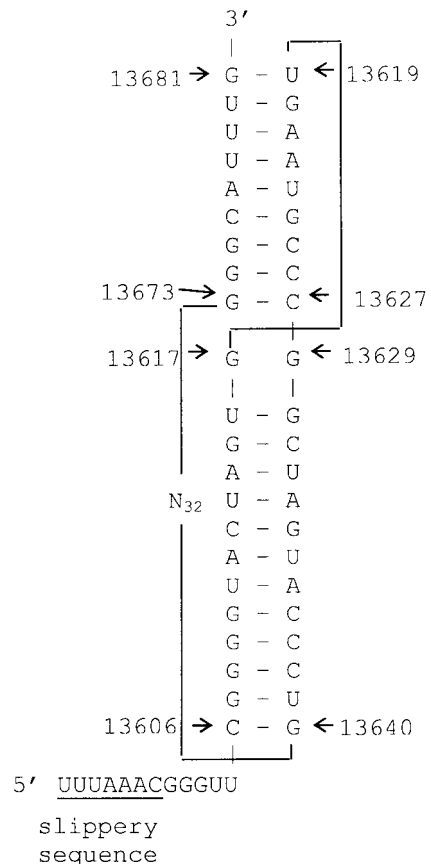

Fig. 1. Model of ribosomal frameshift element in CoV-HKU1. It consists of a slippery sequence (UUUAAAC) and a pseudoknot structure made up by two stems and two loops.

Table 1. Characteristics of the 16 putative polypeptides derived from replicase polyprotein in CoV-HKU1

Polypeptides	First amino acid$^{\text{Position in replicase polyprotein}}$ –Last amino acid$^{\text{Position in replicase polyprotein}}$	Number of amino acid residues	Characteristics
p28	Met^{1}–Gly^{222}	222	Unknown
p65	Ile^{223}–Ala^{809}	587	Unknown
nsp1	Gly^{810}–Gly^{2838}	2,029	ATR, AC, ADRP, PL1^{pro}, PL2^{pro}, Y domain
p44	Val^{2839}–Gln^{3334}	496	Hydrophobic domain
nsp2	Ser^{3335}–Gln^{3637}	303	3CL^{pro}
nsp3	Ser^{3638}–Gln^{3924}	287	Hydrophobic domain
nsp4	Ser^{3925}–Gln^{4013}	92	Unknown
nsp5	Ala^{4014}–Gln^{4210}	194	Unknown
nsp6	Asn^{4211}–Gln^{4320}	110	Unknown
nsp7	Ala^{4321}–Gln^{4457}	137	Unknown
nsp8	Ser^{4458}–Val^{4471}	14	Unknown (short peptide at the end of ORF1a)
nsp9	Ser^{4458}–Gln^{5385}	928	Pol
nsp10	Ser^{5386}–His^{5988}	603	Hel
nsp11	Cys^{5989}–Gln^{6509}	521	ExoN
nsp12	Ser^{6510}–Gln^{6883}	374	XendoU
nsp13	Ala^{6884}–Cys^{7182}	299	2′-O-MT

ATR, acidic tandem repeats; AC, acidic domain; ADRP, ADP-ribose 1″-phosphatase; PL1^{pro}, papain-like protease 1; PL2^{pro}, papain-like protease 2; 3CL^{pro}, 3C-like protease; Pol, RNA-dependent RNA polymerase; Hel, helicase; ExoN, 3′-to-5′ exonuclease; XendoU, poly(U)-specific endoribonuclease and 2′-O-MT, *S*-adenosylmethionine-dependent 2′-*O*-ribose methyltransferase.

Table 2. Analysis of amino acid residues around putative cleavage sites of PL^{pro} and 3CL^{pro} in replicase polyprotein of CoV-HKU1

Putative polypeptides	Amino acid residues around putative cleavage sites														
	CoV-HKU1													MHV, BCoV, HCoV-OC43 (1|1′)	Putatively cleaved by
	Positions (P1|P1′)	6	5	4	3	2	1	1′	2′	3′	4′	5′	6′		
p28|p65	222|223	L	R	Q	Y	R	G	I	K	P	V	L	F	G|V	PL1^{pro} PL2^{pro}
p65|nsp1	809|810	W	R	F	P	C	A	G	R	K	V	N	F	—[a)]	
nsp1|p44	2838|2839	F	S	L	K	G	G	V	V	L	S	N	L	G|A	
p44|nsp2	3334|3335	S	T	S	F	L	Q	S	G	I	V	K	M	—	
nsp2|nsp3	3637|3638	A	G	V	K	L	Q	S	K	T	K	R	F	—	
nsp3|nsp4	3924|3925	E	V	S	Q	I	Q	S	K	L	T	D	V	—	
nsp4|nsp5	4013|4014	D	S	T	V	L	Q	A	L	Q	S	E	F	—	
nsp5|nsp6	4210|4211	A	N	A	V	M	Q	N	N	E	L	M	P	—	
nsp6|nsp7	4320|4321	S	T	I	R	L	Q	A	G	V	A	T	E	—	3CL^{pro}
nsp7|nsp9	4457|4458	S	S	V	A	V	Q	S	K	D	L	N	F	—	
nsp9|nsp10	5385|5386	K	S	A	V	M	Q	S	V	G	A	C	V	—	
nsp10|nsp11	5988|5989	T	L	P	R	L	H	C	T	T	N	L	F	Q|C	
nsp11|nsp12	6509|6510	T	F	T	T	L	Q	S	L	E	N	V	I	—	
nsp12|nsp13	6883|6884	F	Y	P	K	M	Q	A	T	N	D	W	K	—	

[a)] 1|1′ amino acid residues in CoV-HKU1 are the same as those in MHV, BCoV and HCoV-OC43.

The amino acid residues critical for cleavage by PL1^{pro} in MHV as determined by site directed mutagenesis studies are highlighted.

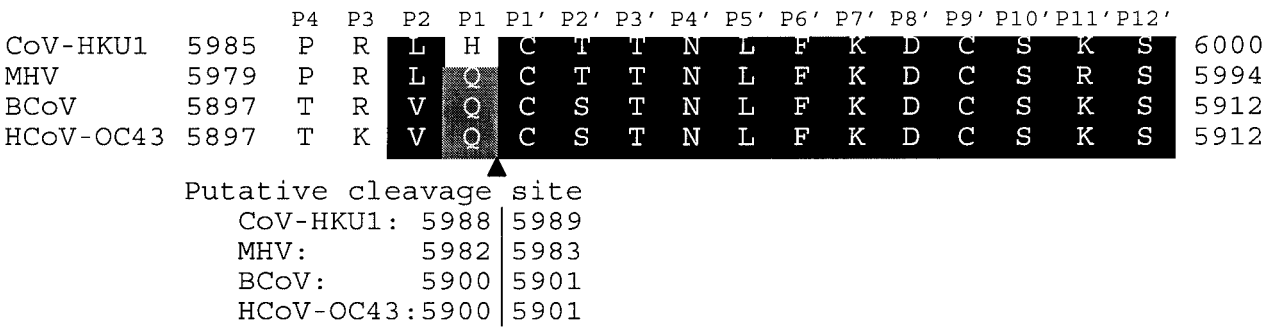

Fig. 2. Multiple alignment of the amino acid sequences around the putative cleavage site between nsp10 and nsp11 in CoV-HKU1, MHV, BCoV and HCoV-OC43. The conserved amino acids from positions P2 to P12′ in the four coronaviruses are highlighted.

16, 23). In most cases, the P1′ positions are occupied by alanine, glycine or serine. Occasionally, in group 2 coronaviruses, the P1′ positions are occupied by cysteine (30); and in rhinoviruses, by asparagines (3). In CoV-HKU1, multiple alignment with the genome sequences of other coronaviruses revealed that all, except one, P1 positions of putative cleavage sites of 3CL^{pro} are occupied by glutamine. The exception lies at the junction between nsp10 (helicase) and nsp11 (member of exonuclease family), where the P1 position is occupied by histidine. This probably represents a novel cleavage site due to the following reasons. First, the sequence is genuine instead of a result of RT-PCR or sequencing errors because the same position has been amplified and sequenced four times in different clinical specimens of the index patient (48). Second, the same mutation was consistently found in the genome sequence of the CoV-HKU1 in the second patient with pneumonia (data not shown). Therefore, it is unlikely that it is due to an occasionally occurring mutation in just one patient. Third, in all other group 2 coronaviruses with genome sequences available (MHV, BCoV and HCoV-OC43), the P2, P1′, P2′, P3′, P4′, P5′, P6′, P7′, P8′, P9′, P10′, P11′ and P12′ positions of this cleavage site are all occupied by leucine/valine [shown to be important for 3CL^{pro} cleavage (35)], cysteine, serine/threonine, threonine, asparagines, leucine, phenylalanine, lysine, aspartic acid, cysteine, serine, lysine/arginine and serine respectively, which are also present at the corresponding positions in CoV-HKU1

(Fig. 2). Fourth, this amino acid substitution is probably due to a single nucleotide mutation in the CoV-HKU1 genome, CAG (as in MHV) or CAA (as in BCoV and HCoV-OC43) (which encodes glutamine) to CAT (which encodes histidine). Finally, as the helicase is one of the most conserved proteins in coronaviruses, cleavage between nsp10 and nsp11 should be an essential step for the generation of the mature functional helicase. The present novel putative cleavage site, as well as some other atypical cleavage sites by $3CL^{pro}$, were not predicted by a program for prediction of $3CL^{pro}$ cleavage sites (21). Experiments, including purification of the CoV-HKU1 helicase in its native form and C-terminal amino acid sequencing and trans-cleavage assays of the CoV-HKU1 $3CL^{pro}$ will confirm the presence of this novel cleavage site.

Putative Polypeptides of the CoV-HKU1 Replicase Polyprotein

The replicase polyprotein of CoV-HKU1 is putatively cleaved by its PL^{pro} and $3CL^{pro}$ into 16 polypeptides (Table 1). Among these 16 putative polypeptides, p28, p65, nsp3, nsp4, nsp5, nsp6, nsp7 and nsp8 have no homologues with known functions by BLAST search.

Similar to other coronaviruses, the N terminal of the putative nsp1 consists of an acidic domain (amino acids 1–333 of nsp1). However, unlike other coronaviruses, there are 16 tandem copies of a 10-amino acid repeat near the C terminal of the acidic domain. The first 14 tandem copies are perfect repeats of NDDENVVTGD, with four of the 10 amino acids being either glutamic acid or aspartic acid. The last two copies are imperfect repeats of NNDEEIVTGD and NDDQIVVTGD respectively. In picornaviruses, it has been found that there were three tandem copies of 24 imperfect amino acid repeats, named VPg1, VPg2 and VPg3 (10). However, no specific functions have been found for these tandem repeats. Downstream to the acidic domain is the $PL1^{pro}$ (amino acids 334–549 of nsp1), with the characteristic catalytic dyad of cysteine and a downstream histidine. Moreover, similar to the $PL1^{pro}$ of other coronaviruses, four conserved cysteine residues (Cys^{429}, Cys^{432}, Cys^{455} and Cys^{457}) that may contain a Zn^{2+} binding motif are present. Downstream to $PL1^{pro}$ is the X domain (amino acids 550–709 of nsp1) that contains putative ADP-ribose 1″-phosphatase (ADRP) activity [amino acids 575–685 of nsp1 belonging to Appr-1-p processing enzyme family (Pfam accession no. PF01661)]. The putative ADRP activity was first described for this domain by Snijder et al. (43). In other microorganisms, such as *Saccharomyces cerevisiae* and other eukaryotes, ADRP and its functionally related enzyme cyclic nucleotide phosphodiesterase (CPDase), were important for tRNA processing (33). ADP-ribose 1″,2″-cyclic phosphate (Appr>p) is produced as a result of tRNA splicing. Appr>p is converted to ADP-ribose 1″-phosphate (Appr-1″p) by CPDase. Appr-1″p is then further processed by ADRP. In other group 2 coronaviruses, both putative ADRP and CPDase (in NS2a) homologues are present (43). Interestingly, in CoV-HKU1 and also SARS-CoV, only ADRP, but not CPDase [NS2a is not present in CoV-HKU1 (48)], is present. Downstream to the X domain, there is a segment of 134 amino acids, also present in other coronaviruses, with no matches of known functions found by BLAST search. Downstream to this segment of unknown function is the putative $PL2^{pro}$ (amino acids 844–1140 of nsp1), with the characteristic catalytic dyad of cysteine and a downstream histidine. Similar to the $PL2^{pro}$ of other coronaviruses, four conserved cysteine residues (Cys^{1026}, Cys^{1028}, Cys^{1060} and Cys^{1062}) that may contain a Zn^{2+} binding motif are also present. At the C terminal of nsp1, similar to other coronaviruses, there is a Y domain with two hydrophobic stretches and 11 conserved Cys/His residues present in the N terminal 889 amino acids (30). Ziebuhr et al. speculated that this Y domain may be responsible for anchoring nsp1 into membranes and binding cations (55). Downstream to nsp1, p44 (496 amino acid residues), putatively cleaved from nsp1 by $PL2^{pro}$, with a hydrophobic transmembrane domain, is present.

Three-dimensional structure modeling of the putative $3CL^{pro}$ (nsp2) of CoV-HKU1 using the three-dimensional structure of SARS-CoV as the template revealed a similar structure for the monomer of the CoV-HKU1 $3CL^{pro}$ as compared to the $3CL^{pro}$ of SARS-CoV and HCoV-229E (Fig. 3). Three domains were present, domain I (8–101 amino acid residues), domain II (102–184 amino acid residues) and domain III (201–303 amino acid residues). Domains I and II consist of six strands of antiparallel β barrels and domain III is a globular structure with five α helices. The catalytic dyad of the protease is located in a cleft between domains I (His^{41}) and II (Cys^{165}).

As in other coronaviruses, the putative Pol (nsp9) of CoV-HKU1 consists of the N-terminal domain, the fingers subdomain, the palm subdomain and the thumb subdomain (Fig. 4). The N-terminal domain (residues 1–371) of the Pol of CoV-HKU1 shares 52–87% amino acid identities with those of other coronaviruses but <31% amino acid identities with those of other RNA viruses. The function of the N-terminal domain is unclear. On the other hand, the fingers (motifs F and G), palm (motifs A–E) and thumb subdomains of the Pol of CoV-HKU1 are homologous to those of other coronaviruses and some positive-stranded RNA virus-

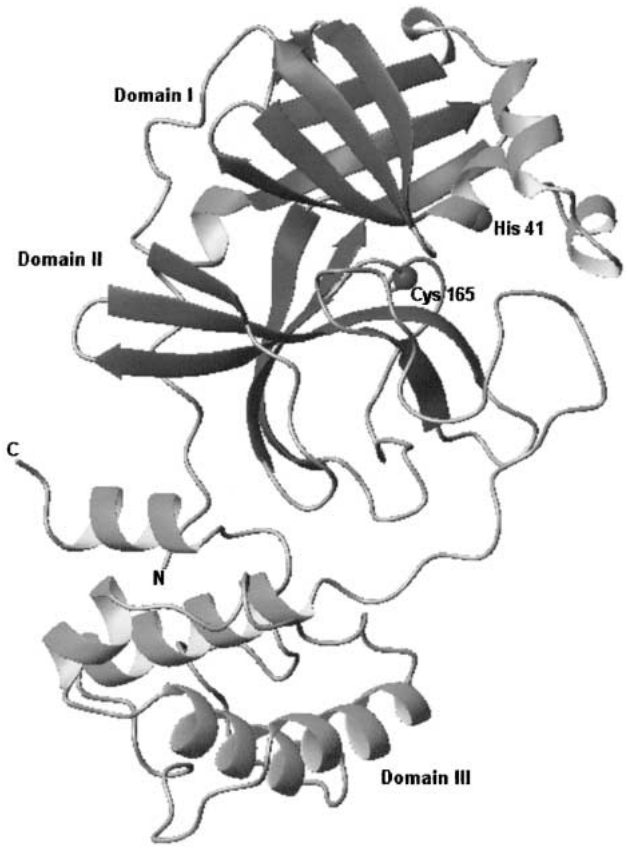

Fig. 3. Three-dimensional structure modeling for monomer of CoV-HKU1 $3CL^{pro}$. The α helices are shown in red and β barrels in blue. The N and C termini are labeled N and C respectively. The catalytic dyad (His^{41} and Cys^{165}) are indicated.

es, such as hepatitis C virus (of family *Flaviviridae*) (1), rabbit hemorrhagic disease virus (of family *Caliciviridae*) (34) and poliovirus 1 (of family *Picornaviridae*) (15) (Fig. 4). The critical amino acids, aspartic acids in motif A and XDD sequence in motif C, which have been demonstrated to be essential for the function of Pol in other coronaviruses and positive-stranded RNA viruses, are also present in Pol of CoV-HKU1.

Similar to the helicase of SARS-CoV, the putative helicase (nsp10) of CoV-HKU1, which unwinds duplex RNA by nucleoside triphosphate hydrolysis activity, possesses six conserved motifs (18) (Fig. 5). Similar to other coronaviruses, based on the presence of conserved motifs, the putative helicase of CoV-HKU1 should belong to superfamily 1 of the three superfamilies of RNA helicases (12, 19). Furthermore, all nine cysteine residues and three histidine residues that are conserved in the N termini of the helicases in nidoviruses are present in the putative helicase of CoV-HKU1 (Fig. 5). It has been suggested that this cysteine/histidine region constitutes a Zn^{2+}-binding domain, which controls the activities of the catalytic domain (42). Similar to the helicases of other nidoviruses and most other helicases of superfamily 1, the helicase of CoV-HKU1 probably unwinds the RNA duplex in a 5′ to 3′ fashion (41, 44).

The putative nsp11 possesses a putative 3′-to-5′ exonuclease (ExoN) domain of the DEDD superfamily (56). The putative nsp12 possesses a putative poly(U)-specific endoribonuclease (XendoU) domain (27). The

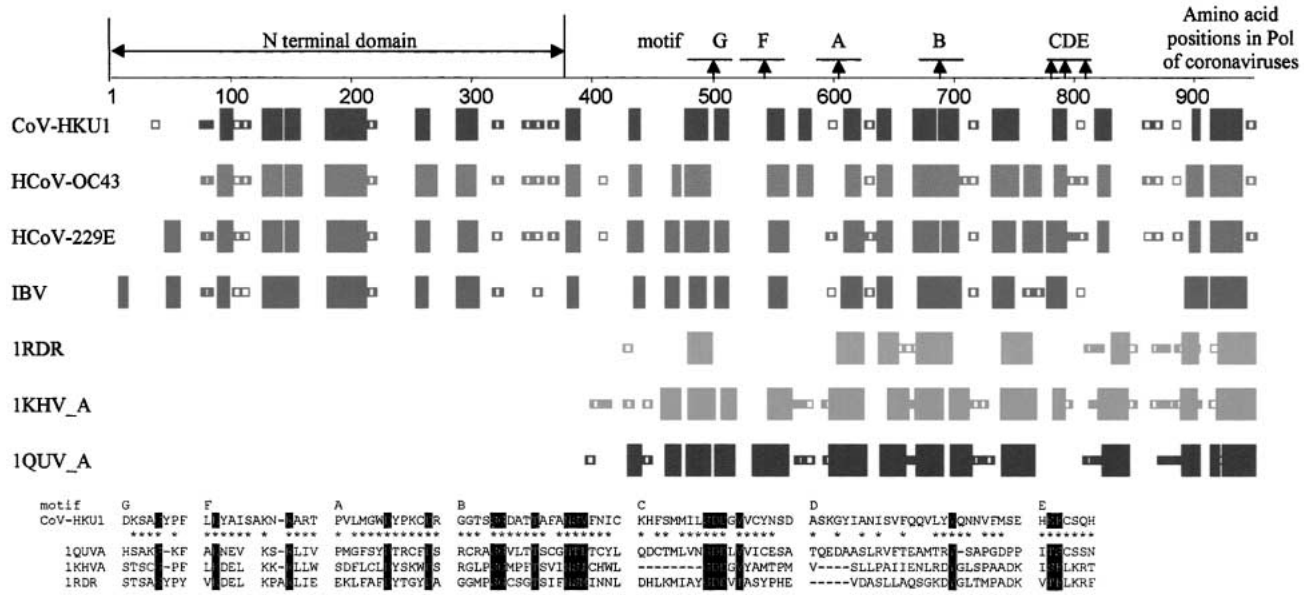

Fig. 4. Schematic representation of the predicted secondary structures of the RNA-dependent RNA polymerases in CoV-HKU1, HCoV-OC43, HCoV-229E and IBV and the secondary structures of the homologous parts of the RNA-dependent RNA polymerases in hepatitis C virus, rabbit hemorrhagic disease virus and poliovirus 1 (large solid bar, α-helix; small empty bar, β-sheet). In the multiple alignment of conserved motifs, amino acid residues identical in the four coronaviruses are marked by stars and conserved amino acid residues in all seven viruses are shaded.

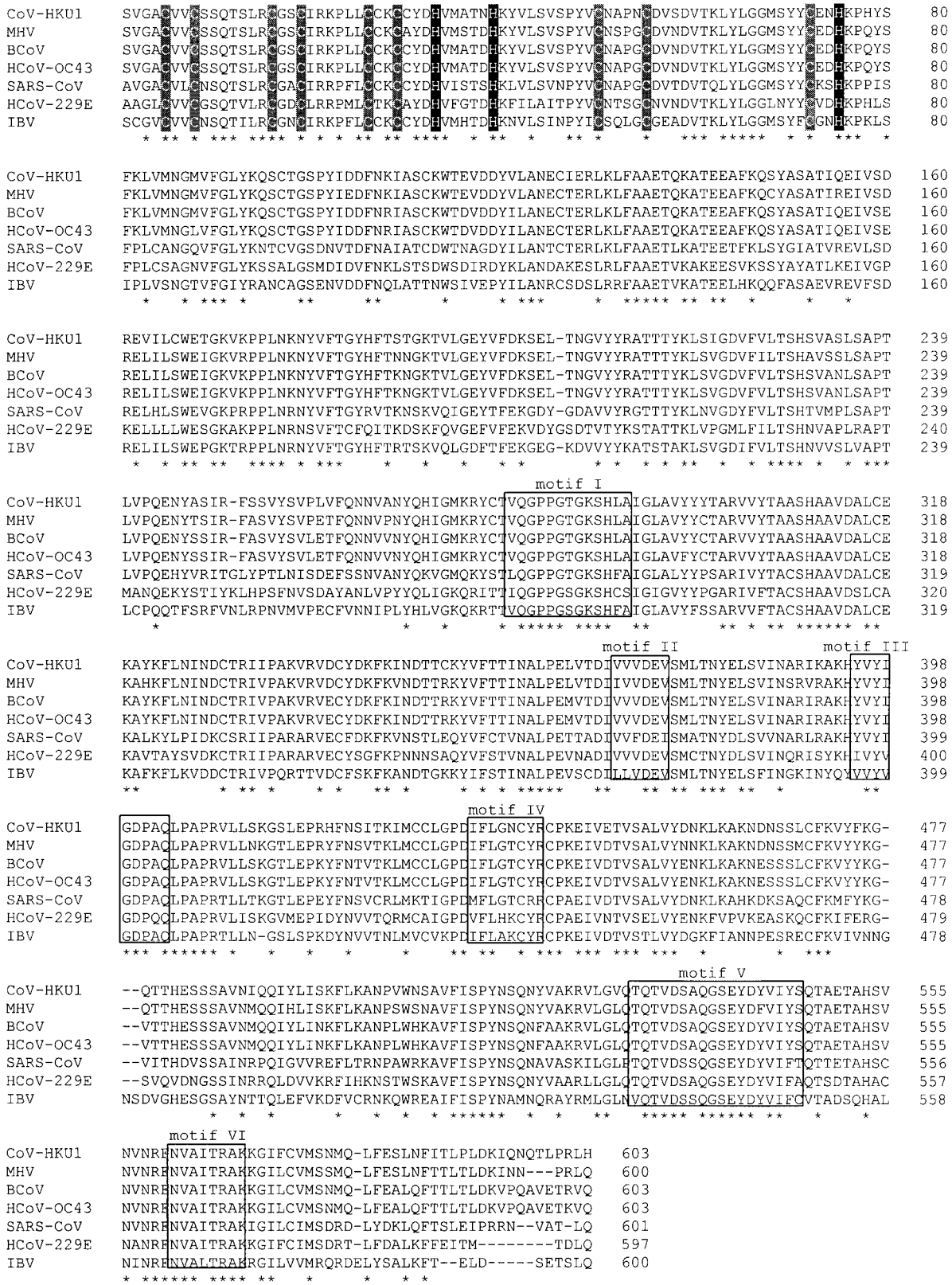

Fig. 5. Multiple alignment of amino acid sequences of the helicases of CoV-HKU1, MHV, BCoV, HCoV-OC43, SARS-CoV, HCoV-229E and IBV. The nine cysteine residues and three histidine residues that are conserved in the N termini of the helicases in nidoviruses are highlighted. The six conserved motifs are also depicted.

putative nsp13 possesses a putative *S*-adenosylmethionine-dependent ribose 2′-*O*-methyltransferase (2′-O-MT) domain of the RrmJ family (8). These three putative enzymes, as well as ADRP and CPDase, are enzymes in RNA processing pathways. ExoN, XendoU and 2′-O-MT are enzymes in a small nucleolar RNA processing and utilization pathway, in contrast to the pre-tRNA splicing pathway that ADRP and CPDase belong to. At the moment, no experiments have been performed to confirm the activities of these putative RNA processing enzymes. Further studies are required to elucidate the exact roles of these putative viral enzymes in the corresponding viruses.

This work is partly supported by the Research Grant Council Grant (7616/05M), Research Fund for the Control of Infectious Diseases of the Health, Welfare and Food Bureau of the Hong Kong SAR Government, Commercial Radio's Fund, Suen Chi Sun Charitable Foundation and Infectious Diseases and William Benter Infectious Disease Fund. We thank Professor Zihe Rao of Tsinghua University for his collaboration on the crystallography study and critical comments of this manuscript.

References

1) Ago, H., Adachi, T., Yoshida, A., Yamamoto, M., Habuka, N., Yatsunami, K., and Miyano, M. 1999. Crystal structure of the RNA-dependent RNA polymerase of hepatitis C virus. Struct. Fold. Des. **7**: 1417–1426.
2) Bateman, A., Coin, L., Durbin, R., Finn, R.D., Hollich, V., Griffiths-Jones, S., Khanna, A., Marshall, M., Moxon, S., Sonnhammer, E.L., Studholme, D.J., Yeats, C., and Eddy, S.R. 2004. The Pfam protein families database. Nucleic Acids Res. **32** (Database issue): D138–D141.
3) Blom, N., Hansen, J., Blaas, D., and Brunak, S. 1996. Cleavage site analysis in picornaviral polyproteins: discovering cellular targets by neural networks. Protein Sci. **5**: 2203–2216.
4) Bonilla, P.J., Hughes, S.A., and Weiss, S.R. 1997. Characterization of a second cleavage site and demonstration of activity in trans by the papain-like proteinase of the murine coronavirus mouse hepatitis virus strain A59. J. Virol. **71**: 900–909.
5) Boursnell, M.E., Brown, T.D., Foulds, I.J., Green, P.F., Tomley, F.M., and Binns, M.M. 1987. Completion of the sequence of the genome of the coronavirus avian infectious bronchitis virus. J. Gen. Virol. **68**: 57–77.
6) Bredenbeek, P.J., Pachuk, C.J., Noten, A.F., Charite, J., Luytjes, W., Weiss, S.R., and Spaan, W.J. 1990. The primary structure and expression of the second open reading frame of the polymerase gene of the coronavirus MHV-A59; a highly conserved polymerase is expressed by an efficient ribosomal frameshifting mechanism. Nucleic Acids Res. **18**: 1825–1832.
7) Brierley, I., Digard, P., and Inglis, S.C. 1989. Characterization of an efficient coronavirus ribosomal frameshifting signal: requirement for an RNA pseudoknot. Cell **57**: 537–547.
8) Bugl, H., Fauman, E.B., Staker, B.L., Zheng, F., Kushner, S.R., Saper, M.A., Bardwell, J.C., and Jakob, U. 2000. RNA methylation under heat shock control. Mol. Cell **6**: 349–360.
9) Dougherty, W.G., and Semler, B.L. 1993. Expression of virus-encoded proteinases: functional and structural similarities with cellular enzymes. Microbiol. Rev. **57**: 781–822.
10) Forss, S., and Schaller, H. 1982. A tandem repeat gene in a picornavirus. Nucleic Acids Res. **10**: 6441–6450.
11) Fouchier, R.A., Hartwig, N.G., Bestebroer, T.M., Niemever, B., de Jong, J.C., Simon, J.H., and Osterhaus, A.D. 2004. A previously undescribed coronavirus associated with respiratory disease in humans. Proc. Natl. Acad. Sci. U.S.A. **101**: 6212–6216.
12) Gorbalenya, A.E., and Koonin, E.V. 1993. Helicases: amino acid sequence comparisons and structure—function relationships. Curr. Opin. Struct. Biol. **3**: 419–429.
13) Guan, Y., Zheng, B.J., He, Y.Q., Liu, X.L., Zhuang, Z.X., Cheung, C.L., Luo, S.W., Li, P.H., Zhang, L.J., Guan, Y.J., Butt, K.M., Wong, K.L., Chan, K.W., Lim, W., Shortridge, K.F., Yuen, K.Y., Peiris, J.S., and Poon, L.L. 2003. Isolation and characterization of viruses related to the SARS coronavirus from animals in southern China. Science **302**: 276–278.
14) Guex, N., and Peitsch, M.C. 1997. SWISS-MODEL and the Swiss-PdbViewer: an environment for comparative protein modelling. Electrophoresis **18**: 2714–2723.
15) Hansen, J.L., Long, A.M., and Schultz, S.C. 1997. Structure of the RNA-dependent RNA polymerase of poliovirus. Structure **5**: 1109–1122.
16) Hegyi, A., and Ziebuhr, J. 2002. Conservation of substrate specificities among coronavirus main proteases. J. Gen. Virol. **83**: 595–599.
17) Hughes, S.A., Bonilla, P.J., and Weiss, S.R. 1995. Identification of the murine coronavirus p28 cleavage site. J. Virol. **69**: 809–813.
18) Ivanov, K.A., Thiel, V., Dobbe, J.C., van der Meer, Y., Snijder, E.J., and Ziebuhr, J. 2004. Multiple enzymatic activities associated with severe acute respiratory syndrome coronavirus helicase. J. Virol. **78**: 5619–5632.
19) Kadare, G., and Haenni, A.L. 1997. Virus-encoded RNA helicases. J. Virol. **71**: 2583–2590.
20) Kanjanahaluethai, A., Jukneliene, D., and Baker, S.C. 2003. Identification of the murine coronavirus MP1 cleavage site recognized by papain-like proteinase 2. J. Virol. **77**: 7376–7382.
21) Kiemer, L., Lund, O., Brunak, S., and Blom, N. 2004. Coronavirus $3CL^{pro}$ proteinase cleavage sites: possible relevance to SARS virus pathology. BMC Bioinformatics **5**: 72.
22) Koradi, R., Billeter, M., and Wüthrich, K. 1996. MOLMOL: a program for display and analysis of macromolecular structures. J. Mol. Grap. **14**: 51–55.
23) Krausslich, H.G., and Wimmer, E. 1988. Viral proteinases. Annu. Rev. Biochem. **57**: 701–754.
24) Kusters, J.G., Jager, E.J., Niesters, H.G., and van der Zeijst, B.A. 1990. Sequence evidence for RNA recombination in field isolates of avian coronavirus infectious bronchitis

virus. Vaccine **8**: 605–608.
25) Lai, M.M., Baric, R.S., Makino, S., Keck, J.G., Egbert, J., Leibowitz, J.L., and Stohlman, S.A. 1985. Recombination between nonsegmented RNA genomes of murine coronaviruses. J. Virol. **56**: 449–456.
26) Lai, M.M., and Cavanagh, D. 1997. The molecular biology of coronaviruses. Adv. Virus Res. **48**: 1–100.
27) Laneve, P., Altieri, F., Fiori, M.E., Scaloni, A., Bozzoni, I., and Caffarelli, E. 2003. Purification, cloning, and characterization of XendoU, a novel endoribonuclease involved in processing of intron-encoded small nucleolar RNAs in *Xenopus laevis*. J. Biol. Chem. **278**: 13026–13032.
28) Lau, S.K., Woo, P.C., Wong, B.H., Tsoi, H.W., Woo, G.K., Poon, R.W., Chan, K.H., Wei, W.I., Peiris, J.S., and Yuen, K.Y. 2004. Detection of severe acute respiratory syndrome (SARS) coronavirus nucleocapsid protein in SARS patients by enzyme-linked immunosorbent assay. J. Clin. Microbiol. **42**: 2884–2889.
29) Lee, C.W., and Jackwood, M.W. 2000. Evidence of genetic diversity generated by recombination among avian coronavirus IBV. Arch. Virol. **145**: 2135–2148.
30) Lee, H.J., Shieh, C.K., Gorbalenya, A.E., Koonin, E.V., La Monica, N., Tuler, J., Bagdzhadzhyan, A., and Lai, M.M. 1991. The complete sequence (22 kilobases) of murine coronavirus gene 1 encoding the putative proteases and RNA polymerase. Virology **180**: 567–582.
31) Makino, S., Keck, J.G., Stohlman, S.A., and Lai, M.M. 1986. High-frequency RNA recombination of murine coronaviruses. J. Virol. **57**: 729–737.
32) Marra, M.A., Jones, S.J., Astell, C.R., Holt, R.A., Brooks-Wilson, A., Butterfield, Y.S., Khattra, J., Asano, J.K., Barber, S.A., Chan, S.Y., Cloutier, A., Coughlin, S.M., Freeman, D., Girn, N., Griffith, O.L., Leach, S.R., Mayo, M., McDonald, H., Montgomery, S.B., Pandoh, P.K., Petrescu, A.S., Robertson, A.G., Schein, J.E., Siddiqui, A., Smailus, D.E., Stott, J.M., Yang, G.S., Plummer, F., Andonov, A., Artsob, H., Bastien, N., Bernard, K., Booth, T.F., Bowness, D., Czub, M., Drebot, M., Fernando, L., Flick, R., Garbutt, M., Gray, M., Grolla, A., Jones, S., Feldmann, H., Meyers, A., Kabani, A., Li, Y., Normand, S., Stroher, U., Tipples, G.A., Tyler, S., Vogrig, R., Ward, D., Watson, B., Brunham, R.C., Krajden, M., Petric, M., Skowronski, D.M., Upton, C., and Roper, R.L. 2003. The genome sequence of the SARS-associated coronavirus. Science **300**: 1399–1404.
33) Nasr, F., and Filipowicz, W. 2000. Characterization of the *Saccharomyces cerevisiae* cyclic nucleotide phosphodiesterase involved in the metabolism of ADP-ribose 1″,2″-cyclic phosphate. Nucleic Acids Res. **28**: 1676–1683.
34) Ng, K.K., Cherney, M.M., Vazquez, A.L., Machin, A., Alonso, J.M., Parra, F., and James, M.N. 2002. Crystal structures of active and inactive conformations of a calicivirus RNA-dependent RNA polymerase. J. Biol. Chem. **277**: 1381–1387.
35) Palmenberg, A.C. 1987. Picornaviral processing: some new ideas. J. Cell Biochem. **33**: 191–198.
36) Peiris, J.S., Lai, S.T., Poon, L.L., Guan, Y., Yam, L.Y., Lim, W., Nicholls, J., Yee, W.K., Yan, W.W., Cheung, M.T., Cheng, V.C., Chan, K.H., Tsang, D.N., Yung, R.W., Ng, T.K., Yuen, K.Y., and SARS Study Group. 2003. Coronavirus as a possible cause of severe acute respiratory syndrome. Lancet **361**: 1319–1325.
37) Peiris, J.S., Chu, C.M., Cheng, V.C., Chan, K.S., Hung, I.F., Poon, L.L., Law, K.I., Tang, B.S., Hon, T.Y., Chan, C.S., Chan, K.H., Ng, J.S., Zheng, B.J., Ng, W.L., Lai, R.W., Guan, Y., Yuen, K.Y., and HKU/UCH SARS Study Group. 2003. Clinical progression and viral load in a community outbreak of coronavirus-associated SARS pneumonia: a prospective study. Lancet **361**: 1767–1772.
38) Rost, B., and Liu, J. 2003. The PredictProtein server. Nucleic Acids Res. **31**: 3300–3304.
39) Rota, P.A., Oberste, M.S., Monroe, S.S., Nix, W.A., Campagnoli, R., Icenogle, J.P., Penaranda, S., Bankamp, B., Maher, K., Chen, M.H., Tong, S., Tamin, A., Lowe, L., Frace, M., DeRisi, J.L., Chen, Q., Wang, D., Erdman, D.D., Peret, T.C., Burns, C., Ksiazek, T.G., Rollin, P.E., Sanchez, A., Liffick, S., Holloway, B., Limor, J., McCaustland, K., Olsen-Rasmussen, M., Fouchier, R., Gunther, S., Osterhaus, A.D., Drosten, C., Pallansch, M.A., Anderson, L.J., and Bellini, W.J. 2003. Characterization of a novel coronavirus associated with severe acute respiratory syndrome. Science **300**: 1394–1399.
40) Schwede, T., Kopp, J., Guex, N., and Peitsch, M.C. 2003. SWISS-MODEL: an automated protein homology-modeling server. Nucleic Acids Res. **31**: 3381–3385.
41) Seybert, A., Hegyi, A., Siddell, S.G., and Ziebuhr, J. 2000. The human coronavirus 229E superfamily 1 helicase has RNA and DNA duplex-unwinding activities with 5′-to-3′ polarity. RNA **6**: 1056–1068.
42) Seybert, A., Posthuma, C.C., van Dinten, L.C., Snijder, E.J., Gorbalenya, A.E., and Ziebuhr, J. 2005. A complex zinc finger controls the enzymatic activities of nidovirus helicases. J. Virol. **79**: 696–704.
43) Snijder, E.J., Bredenbeek, P.J., Dobbe, J.C., Thiel, V., Ziebuhr, J., Poon, L.L., Guan, Y., Rozanov, M., Spaan, W.J., and Gorbalenya, A.E. 2003. Unique and conserved features of genome and proteome of SARS-coronavirus, an early split-off from the coronavirus group 2 lineage. J. Mol. Biol. **331**: 991–1004.
44) Tanner, J.A., Watt, R.M., Chai, Y.B., Lu, L.Y., Lin, M.C., Peiris, J.S., Poon, L.L., Kung, H.F., and Huang, J.D. 2003. The severe acute respiratory syndrome (SARS) coronavirus NTPase/helicase belongs to a distinct class of 5′ to 3′ viral helicases. J. Biol. Chem. **278**: 39578–39582.
45) Thiel, V., Ivanov, K.A., Putics, A., Hertzig, T., Schelle, B., Bayer, S., Weissbrich, B., Snijder, E.J., Rabenau, H., Doerr, H.W., Gorbalenya, A.E., and Ziebuhr, J. 2003. Mechanisms and enzymes involved in SARS coronavirus genome expression. J. Gen. Virol. **84**: 2305–2315.
46) Thompson, J.D., Gibson, T.J., Plewniak, F., Jeanmougin, F., and Higgins, D.G. 1997. The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res. **25**: 4876–4882.
47) van der Hoek, L., Pyrc, K., Jebbink, M.F., Vermeulen-Oost, W., Berkhout, R.J., Wolthers, K.C., Wertheim-Van Dillen, P.M., Kaandorp, J., Spaargaren, J., and Berkhout, B. 2004. Identification of a new human coronavirus. Nat. Med. **10**: 368–373.

48) Woo, P.C., Lau, S.K., Chu, C.M., Chan, K.H., Tsoi, H.W., Huang, Y., Wong, B.H., Poon, R.W., Cai, J.J., Luk, W.K., Poon, L.L., Wong, S.S., Guan, Y., Peiris, J.S., and Yuen, K.Y. 2005. Characterization and complete genome sequence of a novel coronavirus, coronavirus HKU1, from patients with pneumonia. J. Virol. **79**: 884–895.
49) Woo, P.C., Lau, S.K., Tsoi, H.W., Huang, Y., Poon, R.W., Chu, C.M., Lee, R.A., Luk, W.K., Wong, G.K., Wong, B.H., Cheng, V.C., Tang, B.S., Wu, A.K., Yung, R.W., Chen, H., Guan, Y., Chan, K.H., and Yuen, K.Y. 2005. Clinical features and molecular epidemiology of coronavirus HKU1 associated community-acquired pneumonia. J. Infect. Dis. (in press)
50) Woo, P.C., Lau, S.K., Huang, Y., Tsoi, H.W., Chan, K.H., and Yuen, K.Y. 2005. Phylogenetic and recombination analysis of coronavirus HKU1, a novel coronavirus from patients with pneumonia. Arch. Virol. (in press)
51) Woo, P.C., Lau, S.K., Tsoi, H.W., Chan, K.H., Wong, B.H., Che, X.Y., Tam, V.K., Tam, S.C., Cheng, V.C., Hung, I.F., Wong, S.S., Zheng, B.J., Guan, Y., and Yuen, K.Y. 2004. Relative rates of non-pneumonic SARS coronavirus infection and SARS coronavirus pneumonia. Lancet **363**: 841–845.
52) Woo, P.C., Lau, S.K., Wong, B.H., Tsoi, H.W., Fung, A.M., Chan, K.H., Tam, V.K., Peiris, J.S., and Yuen, K.Y. 2004. Detection of specific antibodies to severe acute respiratory syndrome (SARS) coronavirus nucleocapsid protein for serodiagnosis of SARS coronavirus pneumonia. J. Clin. Microbiol. **42**: 2306–2309.
53) Woo, P.C., Lau, S.K., Wong, B.H., Chan, K.H., Chu, C.M., Tsoi, H.W., Huang, Y., Peiris, J.S., and Yuen, K.Y. 2004. Longitudinal profile of immunoglobulin G (IgG), IgM, and IgA antibodies against the severe acute respiratory syndrome (SARS) coronavirus nucleocapsid protein in patients with pneumonia due to the SARS coronavirus. Clin. Diagn. Lab. Immunol. **11**: 665–668.
54) Yang, H., Yang, M., Ding, Y., Liu, Y., Lou, Z., Zhou, Z., Sun, L., Mo, L., Ye, S., Pang, H., Gao, G.F., Anand, K., Bartlam, M., Hilgenfeld, R., and Rao, Z. 2003. The crystal structures of severe acute respiratory syndrome virus main protease and its complex with an inhibitor. Proc. Natl. Acad. Sci. U.S.A. **100**: 13190–13195.
55) Ziebuhr, J., Thiel, V., and Gorbalenya, A.E. 2001. The autocatalytic release of a putative RNA virus transcription factor from its polyprotein precursor involves two paralogous papain-like proteases that cleave the same peptide bond. J. Biol. Chem. **276**: 33220–33232.
56) Zuo, Y., and Deutscher, M.P. 2001. Exoribonuclease superfamilies: structural analysis and phylogenetic distribution. Nucleic Acids Res. **29**: 1017–1026.