Mechanisms and enzymes involved in SARS coronavirus genome expression

Volker Thiel,[1]† Konstantin A. Ivanov,[1] Ákos Putics,[1] Tobias Hertzig,[1] Barbara Schelle,[1] Sonja Bayer,[1] Benedikt Weißbrich,[1] Eric J. Snijder,[2] Holger Rabenau,[3] Hans Wilhelm Doerr,[3] Alexander E. Gorbalenya[2] and John Ziebuhr[1]

Correspondence
John Ziebuhr
j.ziebuhr@mail.uni-wuerzburg.de

[1]Institute of Virology and Immunology, University of Würzburg, Versbacher Str. 7, 97078 Würzburg, Germany

[2]Molecular Virology Laboratory, Department of Medical Microbiology, Leiden University Medical Center, Leiden, The Netherlands

[3]Institute for Medical Virology, Johann Wolfgang Goethe University, Frankfurt (Main), Germany

A novel coronavirus is the causative agent of the current epidemic of severe acute respiratory syndrome (SARS). Coronaviruses are exceptionally large RNA viruses and employ complex regulatory mechanisms to express their genomes. Here, we determined the sequence of SARS coronavirus (SARS-CoV), isolate Frankfurt 1, and characterized key RNA elements and protein functions involved in viral genome expression. Important regulatory mechanisms, such as the (discontinuous) synthesis of eight subgenomic mRNAs, ribosomal frameshifting and post-translational proteolytic processing, were addressed. Activities of three SARS coronavirus enzymes, the helicase and two cysteine proteinases, which are known to be critically involved in replication, transcription and/or post-translational polyprotein processing, were characterized. The availability of recombinant forms of key replicative enzymes of SARS coronavirus should pave the way for high-throughput screening approaches to identify candidate inhibitors in compound libraries.

Received 13 June 2003
Accepted 19 June 2003

INTRODUCTION

Severe acute respiratory syndrome (SARS) is a life-threatening form of pneumonia (Peiris *et al.*, 2003a). In the course of a few months in 2003, an epidemic emerged that has spread from its likely origin in Guangdong Province, China, to 32 countries. By 11 June 2003 more than 8400 cases and 789 deaths had been recorded by the World Health Organization. The rapid transmission by aerosols (and probably also the faecal–oral route) and the high mortality rate make SARS a global threat for which no efficacious therapy is available. There is now clear evidence that SARS is caused by a previously unknown coronavirus, provisionally termed SARS coronavirus (SARS-CoV) (Peiris *et al.*, 2003b; Drosten *et al.*, 2003; Ksiazek *et al.*, 2003; Fouchier *et al.*, 2003). Genome sequences of SARS-CoV isolates obtained from a number of index patients have been published recently and provide important information on the organization, phylogeny and variability of the 29·7 kb positive-strand RNA genome of SARS-CoV (Rota *et al.*, 2003; Marra *et al.*, 2003; Ruan *et al.*, 2003). By analogy with other coronaviruses (Lai & Holmes, 2001; Siddell, 1995; Gorbalenya, 2001), SARS-CoV gene expression is expected to involve complex transcriptional, translational and post-translational regulatory mechanisms, whose molecular details remain to be determined. SARS-CoV genome expression starts with the translation of two large replicative polyproteins, pp1a (486 kDa) and pp1ab (790 kDa), which are encoded by the viral replicase gene (21 221 nt) that comprises ORFs 1a and 1b (Fig. 1). Expression of the ORF1b-encoded region of pp1ab is predicted to involve ribosomal frameshifting into the −1 frame just upstream of the ORF1a translation termination codon (Brierley *et al.*, 1989). The pp1a and pp1ab polyproteins are processed by viral proteinases to yield the functional components of the membrane-bound replicase complex (Ziebuhr *et al.*, 2000). In contrast to most other coronaviruses, which use three proteinase activities for replicase polyprotein processing (Ziebuhr *et al.*, 2000; Gorbalenya, 2001), SARS-CoV is predicted to encode only two proteinases (Rota *et al.*, 2003; Snijder *et al.*, 2003). The replicase complex mediates both genome replication and transcription of a 'nested' set of subgenomic mRNAs. These mRNAs encode the structural

†Present address: Research Department, Cantonal Hospital, St Gallen, Switzerland.

The nucleotide sequence of SARS-CoV, isolate Frankfurt 1, has been deposited in GenBank, accession no. AY291315.

Published ahead of print on 19 June 2003 as DOI 10.1099/vir.0.19424-0.

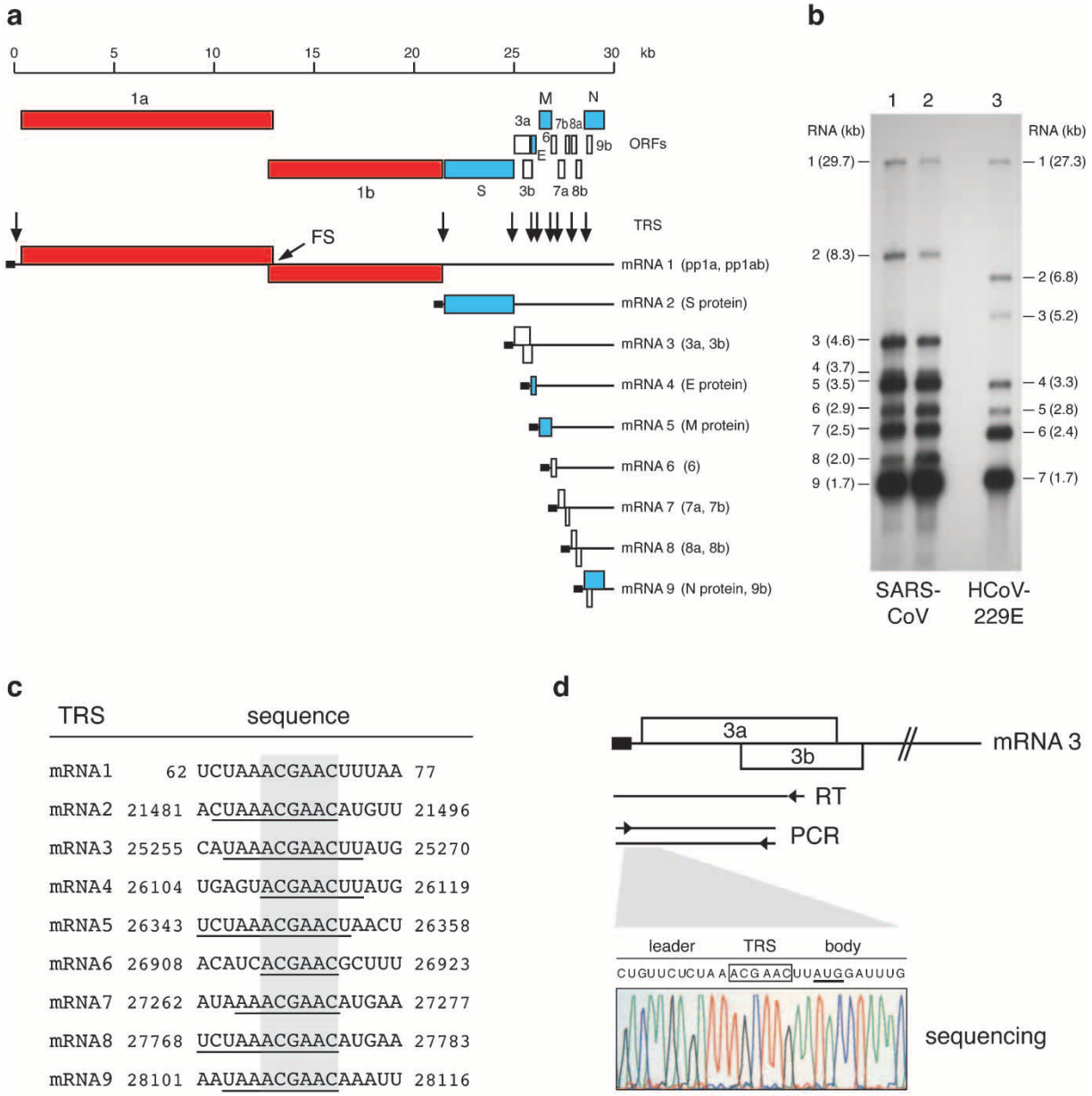

Fig. 1. SARS-CoV genome organization and expression. (a) The SARS-CoV ORFs, frameshift (FS) and TRS elements, and genomic and subgenomic mRNAs, are shown. Black boxes represent the 72 nt leader RNA sequence located at the 5′ end of each viral mRNA. Also indicated are the viral proteins predicted to be expressed from a given mRNA's 'unique' region (i.e. the region not present on smaller mRNAs). (b) Northern blot analysis of poly(A)-containing RNA from SARS-CoV infected Vero (lane 1) and Vero E6 (lane 2) cells. A ^{32}P-labelled probe corresponding to the 3′-terminal 794 nt of the SARS-CoV genome was used to detect genomic and subgenomic SARS-CoV mRNAs. Poly(A)-containing RNA from HCoV-infected MRC-5 cells, which was hybridized with a ^{32}P-labelled probe specific for HCoV nt 26297–27273, was used as size marker (lane 3). (c) Alignment of SARS-CoV TRS elements identified by RT-PCR amplification and sequencing. Nucleotides matching the leader TRS are underlined. These sequences represent the leader-to-body fusion sites of subgenomic RNAs. The minimal TRS (5′-ACGAAC-3′) present in all functional TRSs is highlighted in grey. (d) Strategy used to identify the leader-to-body fusion sites of SARS-CoV subgenomic mRNAs. As an example, the RT-PCR amplification and sequence analysis of the fusion site of mRNA 3 is shown. The reverse transcription reaction was primed using an oligonucleotide specific for the body sequence. For the subsequent PCR, a leader-specific oligonucleotide was used in combination with a second, body-specific oligonucleotide. Sequence analysis of the resulting PCR product is shown in the bottom panel, with the TRS core sequence and translation initiation codon indicated.

proteins, S, E, M and N, and a set of accessory proteins whose number and sequence vary among different coronavirus species (Siddell, 1995). The extraordinary size of the coronavirus replicase (poly)proteins, their generally large phylogenetic distance from those of other RNA viruses, and the presence of several predicted RNA processing activities which are not found in other positive-strand RNA viruses (Gorbalenya, 2001; Snijder *et al.*, 2003), indicate that coronavirus replicases are of an unparalleled complexity. The underlying biological mechanisms and functional constraints that determine the evolution and conservation of these unique activities remain to be elucidated.

In this study, we used the Frankfurt 1 isolate, the sequence of which we report here, to characterize critical steps of SARS-CoV gene expression, such as synthesis of subgenomic mRNAs, translation of key replicative proteins by frameshifting and post-translational processing. Proteolytic processing by viral proteinases provides the active components of the viral replicase complex. The results we obtained were used to rank possible targets for therapeutic intervention in SARS and other coronavirus infections.

METHODS

Preparation of SARS-CoV RNA. Vero or Vero E6 cells (1×10^7) were infected with SARS-CoV (Drosten *et al.*, 2003) (isolate Frankfurt 1, fifth passage in cell culture) at an m.o.i. of 0·01. Two days after infection, intracellular poly(A)-containing RNA was prepared as described by Thiel *et al.* (2001). RNA isolated from respiratory tract and stool specimens was prepared using the QIAamp and QIAamp stool kit (Qiagen), respectively, according to the manufacturer's instructions.

Sequencing of the SARS-CoV (Frankfurt 1) RNA genome. To determine the SARS-CoV genomic sequence, a set of overlapping RT-PCR products with an average size of 2 kb encompassing the entire genome was generated as described by Thiel *et al.* (1997). To generate RT-PCR products containing the exact 3′-terminal sequence of SARS-CoV genomic RNA, reverse transcription was primed using oligonucleotide OLV1/57 (5′-GCCGGCGCCAGCGAGGAGGCTGGGACCATGCCGGCCTTTTTTTTTTTTTTTTTT-3′) and PCR was done using oligonucleotides PCR-L (5′-$_{23}$GGAAAAGCCAACCAACCTCGATCTC$_{47}$-3′) and OLV1/58 (5′-ACGTTCTAGAGCCCAGCCGGCGCCAGCGAGGAGGCT-3′). To generate RT-PCR products containing the exact 5′-terminal sequence the FirstChoice RLM-RACE Kit (Ambion) was used according to the manufacturer's instructions with the following modifications. A synthetic RNA corresponding to human coronavirus 229E (HCoV-229E) nt 1–600 was used as RNA adapter and reverse transcription was primed using oligonucleotide S65 (5′-$_{439}$CTTTTTCCAGCTCTACTAGACCAC$_{416}$-3′). Outer PCR was done using oligonucleotides 240up (5′-CCTTACTCGAGGTTCCGTCTCGTG-3′) and S132 (5′-$_{341}$ACGTCTCTAACCTGAAGGACAGGC$_{318}$-3′); inner PCR was done using oligonucleotides Oli5 (5′-GCGAGGCCGCTAGCAATGG-3′) and S133 (5′-$_{235}$CTAGGTATGCTGATGATCGACTGC$_{212}$-3′). All RT-PCR products served as template for sequencing analysis using a total number of 149 sequencing primers and the BigDye Terminator v3.1 Cycle Sequencing Kit. Sequencing products were detected using an ABI PRISM 3100 Genetic Analyser (Applied Biosystems) and computer-assisted analysis of sequencing data was facilitated by the Lasergene bio-computing software (DNASTAR).

Analysis of SARS-CoV mRNAs. Poly(A)-containing RNA from SARS-CoV- and HCoV-infected cells was separated on a 2·2 M formaldehyde/1 % agarose gel, blotted on nylon membrane and hybridized with ^{32}P-multiprime-labelled DNA probes corresponding to the 3′-terminal 794 nt of the SARS-CoV genome and HCoV nt 26297–27273, respectively. RNAs were analysed by autoradiography. To determine the leader-to-body fusion sites of SARS-CoV subgenomic mRNAs, reverse transcription of poly(A)-containing RNA from SARS-CoV-infected cells was primed using oligonucleotides RT-S (5′-$_{22247}$ATAGGCTGCAGCTGACGTGCCCCA$_{22224}$-3′), RT-3 (5′-$_{25805}$GTTTTGGTGTTGAAATGCCGTCACC$_{25781}$-3′), RT-E (5′-$_{26304}$TTAACACGCGAGTAGACGTAAACCG$_{26280}$-3′), RT-M (5′-$_{26964}$ATCAGTGCCTACACGCTGCGACGC$_{26941}$-3′), RT 6-8 (5′-$_{28043}$ACACCTAGCTATAAGCGCACCACC$_{28020}$-3′) and RT-N (5′-$_{28798}$TGTCTAGCAGCAATAGCGCGAGGGC$_{28774}$-3′). PCR amplification was done using the SARS-CoV leader-specific oligonucleotide PCR-L in combination with body-specific oligonucleotides PCR-S (5′-$_{22108}$ACTACATCTATAGGTTGATAGCCCT$_{22084}$-3′), PCR-3 (5′-$_{25734}$TAGTCATAGTTATGTGTGTGCCAGC$_{25710}$-3′), PCR-E (5′-$_{26243}$AGTACGCACACAATCGAAGCGCAG$_{26220}$-3′), PCR-M (5′-$_{26781}$CACAATTGTCCCCCGGAGAGGCAC$_{26758}$-3′), PCR-6-8 (5′-$_{27937}$CCTAGAGCACAAAGCCAAGCAGTGC$_{27913}$-3′) and PCR-N (5′-$_{28608}$AGGAAGTTGTAGCACGGTGGCAGC$_{28585}$-3′). Sequence analysis of PCR products was done using primers SEQ-S (5′-$_{21738}$AAGGTATGACAGGGTTGCCAAACG$_{21715}$-3′), SEQ-3 (5′-$_{25515}$AAATTGCAAATGAACTGGAAGCCC$_{25492}$-3′), SEQ-E (5′-$_{26243}$AGTACGCACACAATCGAAGCGCAG$_{26220}$-3′), SEQ-M (5′-$_{26640}$AATCGCAATCCCGCCAGTCACCC$_{26618}$-3′), SEQ-6 (5′-$_{27244}$AGGTTCTTCATCATCTAACTCCGA$_{27221}$-3′), SEQ-7 (5′-$_{27439}$AGTGCAAATTTATTGTCAGCAAGA$_{27416}$-3′), SEQ-8 (5′-$_{27937}$CCTAGAGCACAAAGCCAAGCAGTGC$_{27913}$-3′) and SEQ-N (5′-$_{28398}$TCGGGTAGCTCTTCGGTAGTAGCC$_{28375}$-3′).

***In vitro* transcription and translation.** *In vitro* transcription reactions were done using the RiboMAX Large Scale RNA Production System–T7 (Promega) and m^{7}G(5′)ppp(5′)G cap structure analogue as described by Thiel *et al.* (2001). *In vitro* translation reactions were done in rabbit reticulocyte lysate (Promega) using *in vitro*-transcribed RNA (Ziebuhr & Siddell, 1999). Alternatively, DNA templates containing a T7 promoter were transcribed and translated using the TNT T7-Coupled Reticulocyte Lysate System (Promega). To analyse the SARS-CoV frameshifting element, a DNA fragment corresponding to SARS-CoV nt 12955–13961 was amplified by RT-PCR using oligonucleotides S25 (5′-$_{17511}$CAATTTCAGCAGGACAACGGCGAC$_{17488}$-3′; RT reaction), JZ464 (5′-AATAATACGACTCACTATAGGGAACCATGGCTGGAAATGCTACAGAAGTACCTGC-3′, PCR sense primer) and JZ465 (5′-AAAGAATTCTTAACGCATAGCATCGCAGAATTGTAC-3′; PCR antisense primer). To generate a DNA fragment with mutated 'slippery' sequence ($_{13392}$UGUAGCC$_{13398}$), two PCRs were done using oligonucleotides S59 (5′-$_{12916}$ATGGTGCTGGGCAGTTTAGCTGCT$_{12939}$-3′; PCR 1, sense primer), JZ467 (5′-AAAAGGTCTCCGGCTAGAAACGTTGATGCATCCGCAGAC-3′; PCR 1, antisense primer), JZ466 (5′-AAAAGGTCTCTAGCCGGGTTTGCGGTGTAAGTGCAGCCC-3′; PCR 2, sense primer) and S91 (5′-$_{14039}$TACGAAATCACCGAAATCGTACCA$_{14016}$-3′; PCR 2, antisense primer). PCR products 1 and 2 were cleaved with *Bsa*I restriction endonuclease and ligated using T4 DNA ligase. The ligation product was used as template to amplify a DNA fragment corresponding to SARS-CoV nt 12955–13961 (with mutated 'slippery' sequence) using oligonucleotides JZ464 and JZ465.

For *in vitro* transcription–translation of the 3CL^{pro} substrate representing nsp7–nsp10 of SARS-CoV (Ser-3837–Gln-4369), the corresponding coding sequence of SARS-CoV was amplified using primers JZ433 (5′-AATAATACGACTCACTATAGGGCGAACCATGTCTAAAATGTCTGACGTAAAGTGCA-3′) and JZ434 (5′-AAAGAATTCTTACTGCATCAAGGGTTCGCGGAGTTG-3′). The upstream primer contained a T7 RNA polymerase promoter. For *in vitro* transcription–translation

of the pp1a/pp1ab sequence Lys-737–Ser-1858, containing the $PL2^{pro}$ domain and the presumed nsp2|3 cleavage site, the corresponding coding sequence was amplified by RT-PCR using primers AP91 (5′-TAATACGACTCACTATAGGGACGGGAACACCATGGCAAAAGA-AGTAACCTTTCTTGAAGGT-3′) and S77 (5′-$_{5839}$ACGACACAGGC-TTGATGGTTGTAG$_{5816}$-3′). To introduce a $PL2^{pro}$ active-site mutation (Cys-1651 to Ala) in this sequence, two PCRs were done using (1) primers AP91 and AP94 (5′-ATAGCTCTTCATGCATTGTTATCAG-CCCATTTAATTGA-3′) and (2) AP 95 (5′-ATAGCTCTTCAGCATA-TTTGTCTAGTGTTTTATTAGCA-3′) and S77. Following digestion of the PCR products obtained with *Sap*I and ligation with T4 DNA ligase, the pp1a/pp1ab coding sequence Lys-737–Ser-1858 was re-amplified using the ligation product as a template and primers AP91 and S77. The sequences of PCR products used as templates for *in vitro* transcription were confirmed by nucleotide sequencing.

Protein expression, purification, and activities. Plasmid construction, expression in *Escherichia coli* and purification of the maltose-binding protein (MBP) fusion proteins MBP–HCoV $3CL^{pro}$ and MBP–TGEV (porcine transmissible gastroenteritis virus) $3CL^{pro}$, and the corresponding active-site mutants, MBP–HCoV $3CL^{pro}$_C3109V and MBP–TGEV $3CL^{pro}$_C3022A, have been described previously (Ziebuhr *et al.*, 1995, 1997; Hegyi *et al.*, 2002). The same approach was taken to express the SARS-CoV pp1a/pp1ab amino acids 3241–3545 (i.e. the SARS-CoV $3CL^{pro}$ domain lacking the two C-terminal residues, Phe-3545 and Gln-3546) and the corresponding active-site Cys-3385-to-Ala mutant. The coronavirus proteinases were released from MBP by factor Xa cleavage and used, according to previously published protocols (Ziebuhr & Siddell, 1999), in *trans*-cleavage assays with *in vitro*-translated substrate or 0·5 mM of synthetic 15-mer peptides whose sequences were derived from the N-terminal TGEV and mouse hepatitis virus (MHV) $3CL^{pro}$ autoprocessing sites (Seybert *et al.*, 1997; Hegyi & Ziebuhr, 2002). The SARS-CoV helicase (SARS-CoV HEL, pp1ab residues Ala-5302–Gln-5902) and a control protein, SARS-CoV HEL_KA, in which the conserved Lys of the Walker A box (SARS-CoV pp1ab K_{5589}) was replaced by Ala, were expressed and purified in a similar way. Briefly, the helicase-coding region was amplified by RT-PCR using primers JZ425 (5′-GCTGTAGGTGCTTGTGTATTGTGC-3′) and JZ426 (5′-AAAACTGCAGTTATTGTAATGTAGCCACATTGCGACGTGG-3′). The PCR product was digested with *Pst*I and inserted in *Xmn*I- and *Pst*I-digested pMal-c2 DNA (New England Biolabs). The mutation was introduced using a PCR-*in vivo* recombination method (Yao *et al.*, 1992). Expression and purification of MBP–HEL and MBP–HEL_KA were done essentially as described for the HCoV-229E helicase (Heusipp *et al.*, 1997). The partially double-stranded DNA substrate used in the unwinding assay was produced by annealing oligonucleotides D2 [5′-GGTGCAGCCGCAGCGGTGCTCG-d(pT)$_{30}$-3′] and [α-^{32}P]ATP-labelled D3 [5′-d(pT)$_{30}$-CGAGCACCGCTGCGGC-TGCACC-3′] as described by Seybert *et al.* (2000a). The unwinding reaction was done for 30 min at 25 °C in buffer A (HEPES/KOH, pH 7·4, 10 % glycerol, 5 mM magnesium acetate, 2 mM dithiothreitol and 0·1 mg BSA ml^{-1}) using 10 nM of substrate and various concentrations of MBP–HEL (8, 80 and 800 nM) and MBP–HEL_KA (800 nM), respectively. The reaction products were analysed on polyacrylamide/TBE gels, which were exposed to X-ray film. ATPase reactions were done in buffer A for 5 min at 25 °C using the following concentrations: MBP–HEL and MBP–HEL_KA each at 0·8 μM, 10 μM [α-^{32}P]ATP, 1 μM poly(U)$_{250}$ (when included). The samples were analysed by polyethyleneimine–cellulose thin-layer chromatography with 0·25 M potassium phosphate, pH 4·0, as the liquid phase. The reaction products were quantified by phosphorimaging of the dried chromatographic plates (ImageQuant software, Molecular Dynamics).

RESULTS AND DISCUSSION

Genomic sequence of SARS-CoV Frankfurt 1

Here we report the complete genome sequence of a SARS-CoV isolate, Frankfurt 1, obtained from a 32-year-old male physician who was admitted with typical symptoms of SARS to the isolation ward of the Frankfurt University hospital on 15 March 2003 (Drosten *et al.*, 2003). The virus was propagated in African green monkey kidney (Vero and Vero E6) cells, poly(A)-containing RNA was isolated and used for reverse transcription and PCR amplification with primers derived from the SARS-CoV TOR2 sequence. The exact genome termini were determined by 5′ and 3′ RACE methods. The complete sequence encompasses 29 727 nt [excluding the 3′ poly(A) tail] and has been deposited with the GenBank database (accession no. AY291315). Comparison of the genomic sequence of SARS-CoV Frankfurt 1 with 17 previously sequenced SARS-CoV isolates (TOR2, AY274119.3; Urbani, AY278741.1; CUHK-W1, AY278554.2; BJ01, AY278488.2; HKU-39849, AY278491.2; BJ02, AY278487.3; BJ03, AY278490.3; BJ04, AY279354.2; GZ01, AY278489.2; SIN2679, AY283796.1; SIN2500, AY283794.1; ZJ01, AY297028.1; SIN2677, AY283795.1; TW1, AY291451.1; CUHK-Su10, AY282752.1; SIN2748, AY283797.1; SIN2774, AY283798.1) revealed four nucleotide exchanges that have not been described previously for other SARS-CoV isolates and thus seem to be specifically linked to the Frankfurt 1 isolate (A2557, U11448, U24933, U28268). Three of these (may) lead to amino acid substitutions (assuming all predicted ORFs are expressed). Interestingly, the sequence analysis also revealed that, upon SARS-CoV propagation in cell culture (starting with passage 3), a virus variant emerged in which a 45 nt, in-frame sequence (27670–27714) was deleted from ORF7b, thus reducing the genome size to 29 682 nt. The deletion provides evidence that SARS-CoV may undergo rapid adaptation in cell culture. The biological significance of this observation remains to be investigated in detail.

Subgenomic mRNA synthesis

Coronaviruses (and arteriviruses) use a unique strategy to synthesize a set of subgenomic RNAs with common 5′ and 3′ sequences (Fig. 1) (Lai & Holmes, 2001; Siddell, 1995; Pasternak *et al.*, 2001; Sawicki & Sawicki, 1998). Each mRNA contains a short 5′-terminal 'leader' sequence derived from the 5′ end of the genome. The fusion of the noncontiguous sequences is currently believed to be achieved by a discontinuous step during minus-strand synthesis and involves transcription regulatory sequences (TRSs). In addition to the TRS at the 3′ end of the leader sequence (leader TRS), TRSs are located upstream of the genes in the 3′-proximal part of the genome (body TRSs) (Lai & Holmes, 2001; Siddell, 1995; Sawicki & Sawicki, 1998). To confirm that SARS-CoV also uses this discontinuous transcription strategy and to elucidate the molecular details of the resulting subgenomic RNAs, we analysed intracellular RNA synthesis

by Northern blotting and determined the SARS-CoV mRNA sequences at the sites where the common 5′ leader is fused to the various 3′ 'body' sequences, thus generating the subgenomic RNAs identified in SARS-CoV-infected cells. A Northern blot analysis using poly(A)-containing RNA isolated from SARS-CoV-infected cells and a probe specific for the 3′-proximal 794 nt revealed the synthesis of as many as nine RNAs, with RNA 1 representing the viral genome of 29·7 kb. The sizes of the subgenomic mRNAs were assessed using the previously characterized HCoV-229E RNAs (Thiel *et al.*, 2001) as markers. To provide conclusive evidence for the presence of common 5′ leader sequences in each of the SARS-CoV mRNAs and to determine the leader-to-body fusion sites precisely, the 5′-proximal regions of mRNAs 2 to 9 were amplified by RT-PCR and sequenced. The amplification strategy used in these experiments is illustrated for subgenomic mRNA 3 in Fig. 1(d). In some cases we obtained, in addition to the expected RT-PCR product for a given mRNA, larger PCR products that corresponded to the expected RT-PCR products for the next largest subgenomic mRNAs. Sequence analysis of these products confirmed their identity unambiguously. The data obtained by RT-PCR amplification, sequence analysis and Northern blotting consistently suggest that SARS-CoV produces eight subgenomic mRNAs. Furthermore, the study revealed that a minimal consensus sequence, 5′-ACGAAC-3′, is sufficient to direct the synthesis of SARS-CoV subgenomic mRNAs, most probably by base-pairing of its negative-stranded counterpart to the leader TRS during minus-strand synthesis. The number of identical nucleotides in leader TRS and body TRS regions varies between 6 and 11 (Fig. 1c), but there is no clear correlation between the extent of sequence complementarity and abundance of a given mRNA (Fig. 1b and 1c), indicating that additional factors (such as sequence elements, RNA structures, proteins) are involved in regulating the relative abundance of viral mRNAs. It is tempting to speculate that the transcription mechanism used by coronaviruses, arteriviruses and (in part) toroviruses (van Vliet *et al.*, 2002) has evolved to allow the production of a large set of structural and nonstructural (some of them probably virulence-associated) proteins (de Haan *et al.*, 2002), whose abundance can be regulated at the transcriptional level. Regulation of coronavirus gene expression can be even further extended by the presence of additional, downstream ORFs in the 5′ unique regions of some of the subgenomic mRNAs. These are generally expressed by leaky scanning of ribosomes or internal ribosomal entry (Lai & Holmes, 2001; Siddell, 1995; Thiel *et al.*, 1994). As shown in Fig. 1, SARS-CoV also produces four subgenomic RNAs (mRNA 3, 7, 8, 9) with downstream ORFs in their unique regions. The functions of the corresponding SARS-CoV gene products remain to be characterized. The observed 45 nt deletion in the putative ORF7b (see above) appears to suggest that at least one of these gene products is dispensable in cell culture. However, it cannot be excluded that the 15 aa deletion from the ORF7b gene product gives rise to an active (or partially active) protein.

Translation

The structures of SARS-CoV mRNAs (Fig. 1) lead us to suggest that five of the nine SARS-CoV RNAs are functionally bicistronic. For most of them, the mechanisms used to express the downstream ORFs remain to be determined. On the basis of the available data for other coronaviruses (Brierley *et al.*, 1989; Eleouet *et al.*, 1995; Herold *et al.*, 1993; Kocherhans *et al.*, 2001), it seemed likely that ORF1b expression from the genomic RNA would involve −1 ribosomal frameshifting, a process that essentially depends on two elements, known as the 'slippery' sequence, i.e. the site where the ribosomes shift into the −1 reading frame, and a complex RNA pseudoknot structure (Brierley *et al.*, 1989, 1995). Analysis of the SARS-CoV sequence flanking the ORF1a termination codon revealed a putative SARS-CoV frameshifting element comprised of a putative 'slippery' sequence ($_{13392}$UUUAAAC$_{13398}$) and, further downstream, stretches of complementary sequences that can be modelled to form a typical pseudoknot structure (Fig. 2a) (Brierley *et al.*, 1995). To confirm that these elements mediate frameshifting in SARS-CoV and to define the frameshift site precisely, we synthesized RNAs containing the SARS-CoV frameshift region and produced a mutant version of the presumed 'slippery' sequence ($_{13392}$UUUAAAC$_{13398}$→ $_{13392}$UGUAGCC$_{13398}$). As shown in Fig. 2(b), efficient ribosomal frameshifting depended on an essentially unmodified $_{13392}$UUUAAAC$_{13398}$ sequence, supporting the prediction that this sequence constitutes the actual slippage site.

Polyprotein processing by PL2^{pro}

Translation of pp1a and its C-terminally extended version, pp1ab (see above), is coupled with extensive proteolytic processing by viral proteinases (Fig. 3). Except for infectious bronchitis virus (IBV) (Liu *et al.*, 1995; Lim *et al.*, 2000), all previously characterized coronaviruses encode two (paralogous) papain-like cysteine proteinases (PL1^{pro} and PL2^{pro}), which cleave the N-proximal polyprotein regions at three sites (Gorbalenya *et al.*, 1991; Bonilla *et al.*, 1997; Baker *et al.*, 1989; Kanjanahaluethai & Baker, 2000; Ziebuhr *et al.*, 2001; Herold *et al.*, 1998), while the 3C-like cysteine proteinase (3CL^{pro}; also called main proteinase, M^{pro}) cleaves the central and C-proximal regions at 11 conserved sites (Ziebuhr *et al.*, 2000). The conservation of both the positions and sequences of coronavirus pp1a/pp1ab cleavage sites allows the substrate specificities of proteinases of newly identified coronaviruses to be predicted (Fig. 3). Thus, it has been proposed that SARS-CoV PL2^{pro} cleaves three sites in the N-proximal region of pp1a/pp1ab (Snijder *et al.*, 2003). This is similar to IBV, where one proteinase, PL2^{pro}, cleaves two sites in this region, but stands in contrast to other coronaviruses (e.g. HCoV-229E and MHV), which cleave three sites by using two PL^{pro} activities, PL1^{pro} and PL2^{pro}. To establish the proteolytic activity of the SARS-CoV PL2^{pro} domain at one of the predicted sites, we expressed, by *in vitro* translation, the SARS-CoV pp1a/pp1ab amino acids 737–1858 along with a mutant construct in which the

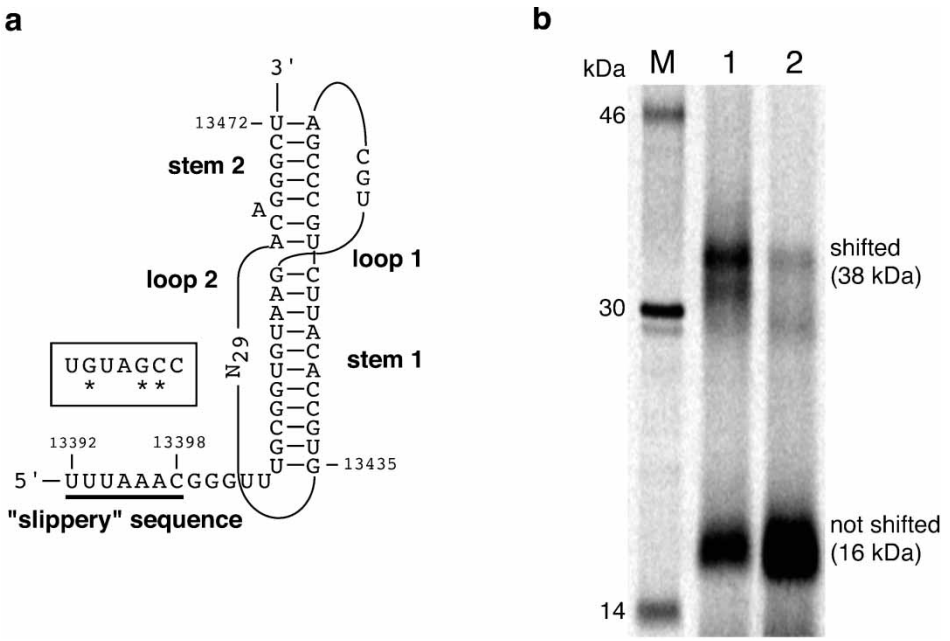

Fig. 2. SARS-CoV RNA-mediated ribosomal frameshifting. (a) Model of the SARS-CoV ribosomal frameshifting element which, by analogy with other coronaviruses, is proposed to consist of a putative pseudoknot structure comprising two stems and two loops and a 'slippery' sequence ($_{13392}$UUUAAAC$_{13398}$, underlined). Also shown are nucleotide changes that were introduced into the 'slippery' sequence to test its role in SARS-CoV RNA-mediated ribosomal frameshifting. (b) Functional analysis of ribosomal frameshifting. Synthetic RNAs corresponding to SARS-CoV nt 12955–13961 with the authentic ($_{13392}$UUUAAAC$_{13398}$, lane 1) or mutagenized ($_{13392}$UGUAGCC$_{13398}$, lane 2) putative SARS-CoV 'slippery' sequence were translated *in vitro* using rabbit reticulocyte lysate. The sizes of translation products calculated for non-shifted ORF1a-encoded and shifted ORF1a/1b-encoded translation products are given.

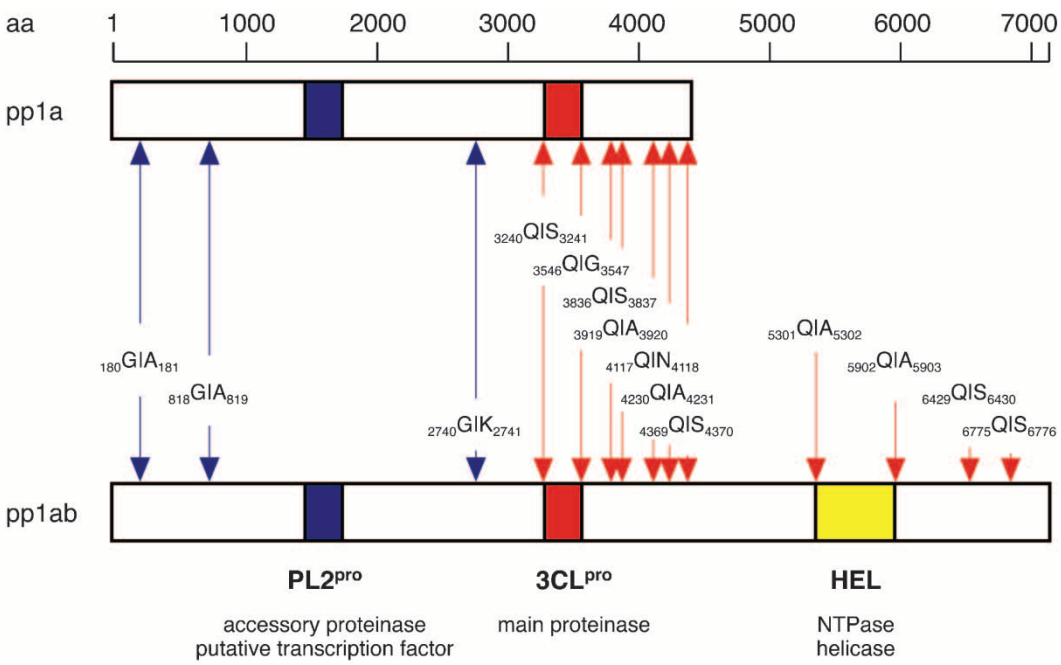

Fig. 3. SARS-CoV enzymatic activities characterized in this study. The two SARS-CoV replicase polyproteins, pp1a and pp1ab, are shown together with the papain-like proteinase 2 (PL2^{pro}), 3C-like proteinase (3CL^{pro}) and helicase (HEL) domains characterized in this study. The positions of cleavage sites predicted to be processed by PL2^{pro} (blue) and 3CL^{pro} (red) are indicated.

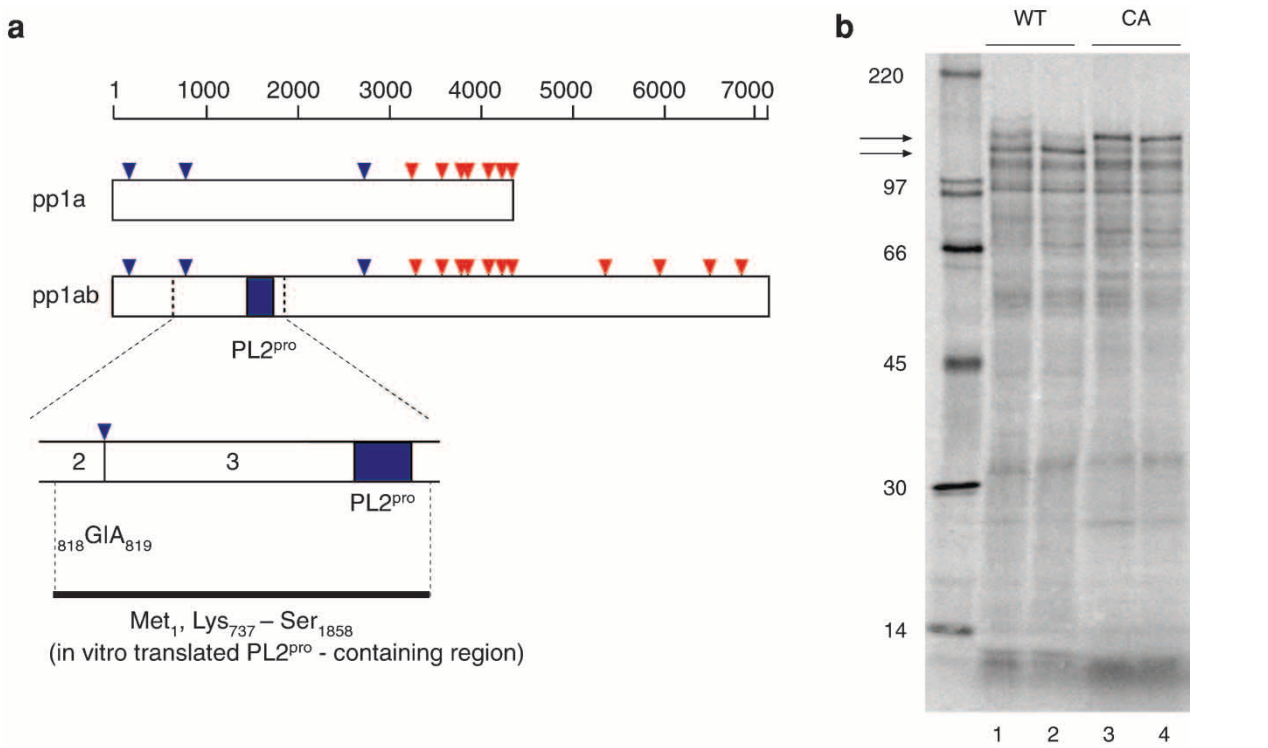

Fig. 4. Proteolytic activity of the SARS-CoV papain-like cysteine proteinase 2. (a) Representation of the SARS-CoV pp1a/pp1ab cleavage sites predicted to by cleaved by $PL2^{pro}$ (shown in blue) or $3CL^{pro}$ (red). Also shown is the structure of the *in vitro*-translated protein that contains the putative $PL2^{pro}$ domain and the predicted nsp2|nsp3 site. (b) *In vitro* translation of the SARS-CoV pp1a/pp1ab 737–1858 residues. Lane 1, 737–1858 reaction after 40 min; lane 2, 737–1858 reaction after 160 min; lane 3, 737–1858_CA (Cys-1651→Ala) reaction after 40 min; lane 4, 737–1858_CA reaction after 160 min. Full-length proteins (substrates) and the larger cleavage product are indicated by arrows.

presumed active-site Cys nucleophile of $PL2^{pro}$ was replaced by Ala (C-1651–A). The data shown in Fig. 4 strongly suggest that rapid, probably co-translational autoprocessing of the wild-type protein occurred. The apparent sizes of the major products of ~125 (C-1651–A) and ~115 kDa (wild-type sequence) are consistent with cleavage at the predicted site, ^{818}Gly|Ala^{819} (Snijder *et al.*, 2003). Probably due to the presence of only two Met residues in the N-terminal cleavage product, the expected N-terminal 9 kDa protein could not be detected convincingly. Comparative analysis of coronavirus PL^{pro} cleavage sites revealed that the (predicted) SARS-CoV $PL2^{pro}$ sites are more conserved than the equivalent HCoV-229E $PL1^{pro}$ and $PL2^{pro}$ sites (Fig. 5). The poor conservation of HCoV-229E PL^{pro} cleavage sites probably results from the overlapping substrate specificities of $PL1^{pro}$ and $PL2^{pro}$, which we reported previously (Ziebuhr *et al.*, 2001). Consistent with the hypothesis that conservation of only one PL^{pro} domain is linked to an increase in specificity, we also find that the IBV $PL2^{pro}$ cleavage sites are much better conserved than the corresponding HCoV-229E $PL1^{pro}$/$PL2^{pro}$ sites (Fig. 5). The observed narrow substrate specificity of the SARS-CoV $PL2^{pro}$ will certainly facilitate the development of selective proteinase inhibitors.

```
                                  P|P'
Virus        Site          987654321|123456789   Protease
HCoV-229E    nsp1|nsp2   : VVFGKRGGG|NVTYTDQYL : PL1/PL2?
HCoV-229E    nsp2|nsp3   : PVAFTKAAG|GKVSFSDDV : PL1/PL2
HCoV-229E    nsp3|nsp4   : TSIVAKQGA|GDAGHSLTW : PL1?/PL2?

SCoV-FRA1    nsp1|nsp2   : ELTRELNGG|AVTRYVDNN : PL2?
SCoV-FRA1    nsp2|nsp3   : NNVFRLKGG|APIKGVTFG : PL2?
SCoV-FRA1    nsp3|nsp4   : TTKISLKGG|KIVSTCFKL : PL2?

IBV-B        nsp2|nsp3   : AINVVCKAG|GKTVTFGET : PL2
IBV-B        nsp3|nsp4   : KLLVEKKAG|GIVSGTFKC : PL2
```

Fig. 5. Alignment of established and predicted HCoV-229E $PL1^{pro}$ and $PL2^{pro}$, and SARS-CoV (SCoV) and IBV $PL2^{pro}$ cleavage sites. Note the overlapping substrate specificity of HCoV-229E $PL1^{pro}$ and $PL2^{pro}$ towards (at least) one cleavage site. IBV-B, infectious bronchitis virus strain Beaudette, NC_001451; HCoV-229E, human coronavirus 229E, NC_002645. Positions with absolute conservation are highlighted in red. Green and yellow mark descending levels of conservation using the following amino acid similarity groups: (i) D, E, N, Q; (ii) S, T; (iii) K, R; (iv) F, W, Y; and (v) I, L, M, V. The (predicted) locations of the scissile bonds (|) in the cleavage sites along with the numbers of pp1a/pp1ab residues are indicated.

Polyprotein processing by 3CL^{pro}

At present the viral main proteinase, 3CL^{pro}, is the best-characterized coronavirus enzyme. Its essential function is reflected by the fact that it cleaves as many as 11 sites in the replicase polyproteins and also releases the key replicative functions, such as the RNA-dependent RNA polymerase and the helicase, from the polyprotein precursors (Gorbalenya *et al.*, 1989a; Ziebuhr *et al.*, 2000). Furthermore, 3CL^{pro} is the only coronavirus protein for which structure information is available (Anand *et al.*, 2002, 2003). Coronavirus 3CL^{pro}s are cysteine proteinases with interesting properties. They feature a serine proteinase (chymotrypsin)-like two-β-barrel fold but employ a catalytic Cys–His dyad instead of the classical Ser–His–Asp triad. Furthermore, they possess a third, C-terminal domain composed of five α-helices. Our previous work has shown that the 3CL^{pro} substrate specificities are conserved among the three established groups of coronaviruses (Hegyi & Ziebuhr, 2002) and comparison of the previously known coronavirus cleavage sites with those identified in the SARS-CoV polyproteins (Fig. 3; Fig. 6a

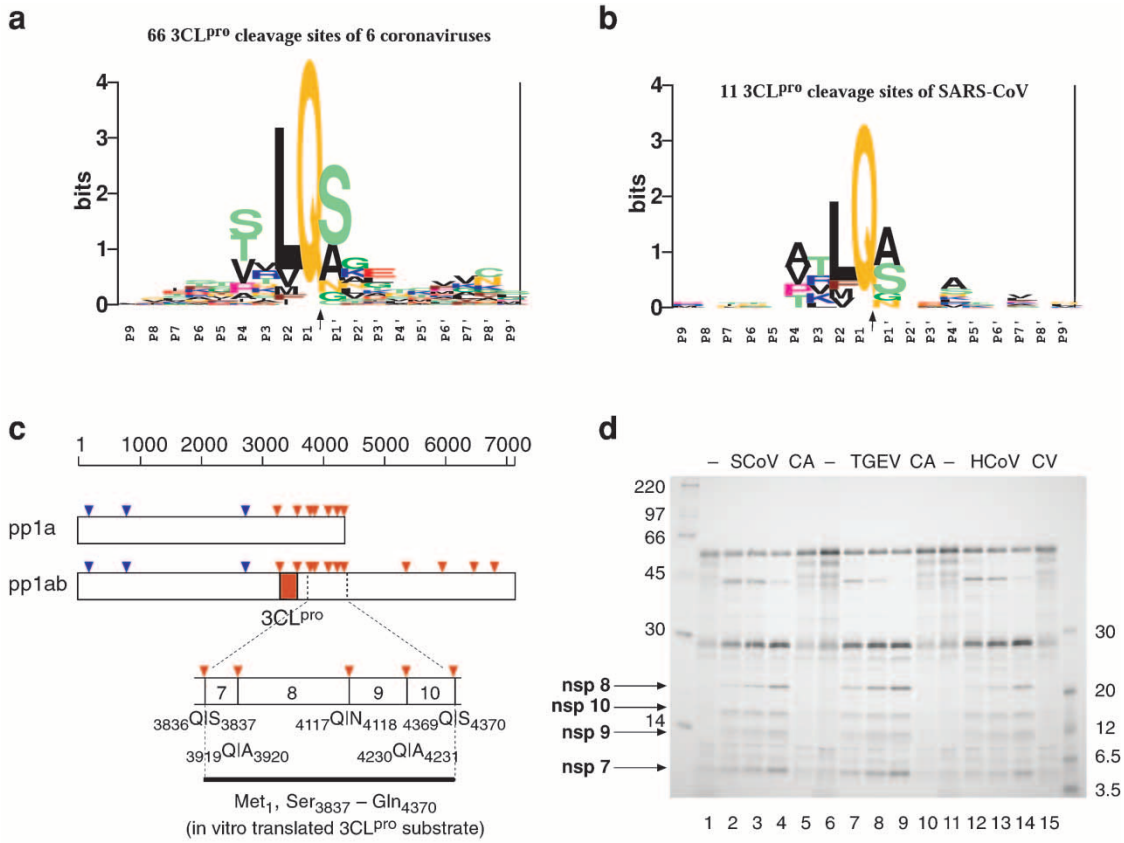

Fig. 6. Conservation of coronavirus main proteinase substrate specificities. Conservation of 3CL^{pro} sites in polyproteins of SARS-CoV (b) and six other coronaviruses (a). Two separate multiple, gap-free 18-aa-long alignments including the P9–P9′ positions of the sites (presumably) cleaved by the 3CL^{pro} domains of six coronaviruses (transmissible gastroenteritis virus, NC_002306; HCoV-229E; porcine epidemic diarrhoea virus, NC_003436; mouse hepatitis virus A59, NC_001846; bovine coronavirus, AF391542; IBV-B) and SARS-CoV were converted into two sequence logo (Schneider & Stevens, 1990) presentations. In the logos, the height of each letter (amino acid residue) is proportional to its frequency at the specific position, with the highest-frequency residue being on top of the stack. The height of the entire stack is proportional to the information at this position which is measured in bits, with the upper limit of information at any position being equal to 4·32 bits. Amino acid residues are coloured in the following groups: light-green – S, T, C; orange – N, Q; red – D, E; blue – K, R, H; brown – W, F, Y; black – A, L, I, V, M; pink – P; green – G. The most conserved and important positions have relatively high letter heights and are easily recognized, along with the individual and group residues occupying them. (c) Representation of the SARS-CoV pp1a/pp1ab cleavage sites predicted to be cleaved by PL2^{pro} (shown in blue) or 3CL^{pro} (red). Also shown is the structure of the substrate used in the *trans*-cleavage assay. The P1 and P1′ residues of predicted cleavage sites are given. SARS-CoV pp1a/pp1ab residues were translated *in vitro* and incubated for 180 min at 25 °C (lanes 1, 6, 11); or were incubated for 30 min (lanes 2, 7, 12), 60 min (lanes 3, 8, 13) and 180 min (lanes 4, 9, 14), respectively, with bacterially expressed SARS-CoV (lanes 2–4), TGEV (lanes 7–9) and HCoV (lanes 12–14) main proteinases. As controls, the corresponding active-site Cys-substituted SARS-CoV (lane 5), TGEV (lane 10) and HCoV (lane 15) 3CL^{pro}s were incubated for 180 min with the same substrate.

and b) suggests conservation of the P4 (Ser, Thr, Val, Pro, Ala), P2 (Leu, Ile, Val, Phe, Met), P1 (Gln) and P1′ (Ser, Ala, Gly, Asn, Cys) residues, all of which have previously been proposed to define the coronavirus 3CL^{pro} substrate specificity (Ziebuhr *et al.*, 2000). In SARS-CoV, the strong preference for Leu at the P2 positions of substrates is less prominent and the more frequent use of Phe may indicate a slightly larger S2 subsite. There are also conservative changes in the P4 and P1′ positions of SARS-CoV 3CL^{pro} substrates. To determine whether these subtle variations result in differences in substrate specificity, we characterized the SARS-CoV 3CL^{pro} specificity in more detail and compared it with that of the well-characterized HCoV-229E and TGEV enzymes. SARS-CoV 3CL^{pro} was expressed as an MBP fusion protein, an approach that had proven suitable for the production of active coronavirus 3CL^{pro}s (Seybert *et al.*, 1997; Ziebuhr *et al.*, 1997; Hegyi *et al.*, 2002). Following factor Xa cleavage of the purified fusion protein, the released SARS-CoV 3CL^{pro} was used in *trans*-cleavage experiments with *in vitro*-translated substrates (Fig. 6d) or synthetic peptides derived from group 1 (TGEV) and group 2 (MHV) N-terminal 3CL^{pro} autoprocessing sites (Hegyi & Ziebuhr, 2002; Seybert *et al.*, 1997) (data not shown). The substrate derived from the C-terminal pp1a sequence was chosen because it contained as many as three cleavage sites, allowing side-by-side comparison of cleavage efficiencies at different sites. Furthermore, the fact that the proteolytic processing of this region has already been studied extensively in other coronaviruses (Ziebuhr & Siddell, 1999; Bost *et al.*, 2000) allowed safe predictions on SARS-CoV 3CL^{pro} cleavage products in this region. As shown in Fig. 6(d), the cleavage products (and several intermediate products) predicted to be released by 3CL^{pro} cleavage (Fig. 3; Snijder *et al.*, 2003) were readily detectable upon incubation of the translation product with purified SARS-CoV 3CL^{pro}, but not with the active-site Cys-to-Ala mutant. Significantly, nearly identical cleavage kinetics were observed when purified TGEV and HCoV-229E 3CL^{pro}s (instead of SARS-CoV 3CL^{pro}) were incubated with the SARS-CoV-derived substrate. Again, as expected, no cleavage was observed with the active-site Cys-to-Ala (Val) mutants. Consistent data were also obtained in peptide cleavage assays, where the SARS-CoV, HCoV-229E and TGEV enzymes were shown to cleave TGEV and MHV substrates with equal efficiencies (data not shown). The data conclusively demonstrate the proteolytic activity of the expressed SARS-CoV protein and, even more importantly, suggest conservation of coronavirus 3CL^{pro} substrate specificities.

Essential proteins involved in replication and transcription

Characterization of individual protein functions involved in coronavirus replication and transcription is still at a very early stage. Thus, for example, there is essentially no experimental data on coronavirus RNA polymerase activities. Only the helicase has been characterized to some extent (Seybert *et al.*, 2000a). On the basis of seven conserved sequence motifs, the coronavirus enzyme has been classified as belonging to the helicase superfamily 1 (Gorbalenya *et al.*, 1989b). We previously reported that helicases of HCoV-229E and the arterivirus equine arteritis virus (EAV) possess polynucleotide-stimulated NTPase activity and 5′-to-3′ RNA and DNA duplex-unwinding activities (Seybert *et al.*, 2000a, b). Both coronavirus and arterivirus helicases require an N-terminal (putative) metal-binding domain, consisting of at least 12 conserved Cys and His residues, for activity. The domain is essential for replication, transcription and virion morphogenesis (as demonstrated for EAV) (van Dinten *et al.*, 2000). The SARS-CoV helicase is predicted to be released from pp1ab by 3CL^{pro}-mediated cleavages of the ^{5301}Gln|Ala^{5302} and ^{5902}Gln|Ala^{5903} dipeptide bonds (Fig. 3; Snijder *et al.*, 2003). We therefore expressed SARS-CoV pp1ab residues 5302–5902, representing the 67 kDa helicase together with its N-terminal metal-binding domain, in a bacterial expression system. A fusion protein, consisting of MBP and the SARS-CoV helicase domain, was purified by amylose affinity chromatography and used in ATPase and duplex-unwinding assays. Fig. 7(a) shows that this fusion protein (MBP–HEL), but not a control protein (MBP–HEL_KA) in which the conserved Lys residue of the Walker A box (Walker *et al.*, 1982) was replaced by Ala, had ATPase activity that could be stimulated by poly(U). The protein also had DNA duplex-unwinding activity (Fig. 7b). This is consistent with the previously characterized HCoV-229E helicase (Seybert *et al.*, 2000a), which also had revealed a high promiscuity with respect to the substrates used. Thus, both DNA and RNA duplexes were found to be unwound by the HCoV-229E helicase with similar efficiency and all types of nucleotide and ribonucleotide cofactors were used (Seybert *et al.*, 2000a; K. A. Ivanov, V. Thiel & J. Ziebuhr, unpublished data).

Conclusion

The study provides the first experimental data on both mechanisms and enzymes involved in SARS coronavirus genome expression and extends our understanding of the phylogenetic relationship between SARS-CoV and other coronaviruses. Given the unresolved global threat of the SARS-CoV epidemic, the rapid development of efficacious antiviral drugs is urgently needed. From this perspective, our results provide insights into key replicative enzymes which represent attractive targets for antiviral therapy. The methods established here for the large-scale production of recombinantly expressed, active SARS-CoV enzymes pave the way for high-throughput screening approaches to identify candidate inhibitors in large compound libraries. Thus, for example, the DNA duplex-unwinding activity of the SARS-CoV helicase should allow the use of high-throughput DNA (instead of RNA)-based helicase assays, which will significantly facilitate the search for (currently not evident) inhibitors. Furthermore, the SARS-CoV PL2^{pro} is shown to possess an (among coronaviruses PL^{pro}s) unusually narrow substrate specificity, making this protein another suitable target for rational drug design. At present,

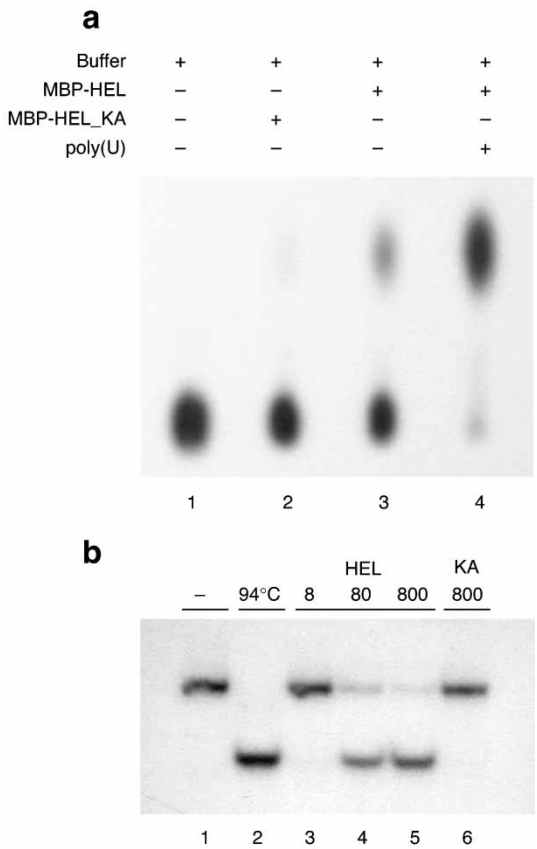

Fig. 7. ATPase and duplex-unwinding activities of the SARS-CoV superfamily 1 helicase. (a) ATPase activity was analysed by thin-layer chromatography using [α-^{32}P]ATP as a substrate. Lanes: 1, reaction without protein; 2, with MBP-HEL_KA; 3, with MBP-HEL; 4, with MBP-HEL and 1 μM poly(U)$_{250}$. (b) Duplex-unwinding activity was analysed using a twin-tailed ('forked') DNA substrate consisting of a 22 bp DNA duplex with 30-nt-long, single-stranded oligo(dT) tails (see Methods). Lanes: 1, substrate incubated with buffer; 2, heat-denatured substrate; 3–5, reactions with 8 nM (lane 4), 80 nM (lane 5), 800 nM (lane 6) MBP–HEL; 6, reaction with 800 nM MBP–HEL_KA.

the SARS-CoV 3CL^{pro} represents the most promising target. Both the availability of crystal structures for TGEV and HCoV-229E coronavirus 3CL^{pro} enzymes, which share their substrate specificities with that of SARS-CoV 3CL^{pro}, and the availability of active SARS-CoV 3CL^{pro} for high-throughput screening assays, should provide an excellent basis for the rapid identification of candidate drugs suitable for SARS therapy.

ACKNOWLEDGEMENTS

J. Z. and V. T. gratefully acknowledge the continuous support by Stuart Siddell (University of Bristol, UK) over many years. The authors are grateful to Barbara Scheiner, Hanne Weinand, Martin Heinkelein and Axel Rethwilm for excellent technical assistance, reagents and/or helpful discussions. The work was supported by the Deutsche Forschungsgemeinschaft.

REFERENCES

Anand, K., Palm, G. J., Mesters, J. R., Siddell, S. G., Ziebuhr, J. & Hilgenfeld, R. (2002). Structure of coronavirus main proteinase reveals combination of a chymotrypsin fold with an extra alpha-helical domain. *EMBO J* **21**, 3213–3224.

Anand, K., Ziebuhr, J., Wadhwani, P., Mesters, J. R. & Hilgenfeld, R. (2003). Coronavirus main proteinase (3CL^{pro}) structure: basis for design of anti-SARS drugs. *Science* **300**, 1763–1767.

Baker, S. C., Shieh, C. K., Soe, L. H., Chang, M. F., Vannier, D. M. & Lai, M. M. (1989). Identification of a domain required for autoproteolytic cleavage of murine coronavirus gene A polyprotein. *J Virol* **63**, 3693–3699.

Bonilla, P. J., Hughes, S. A. & 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.

Bost, A. G., Carnahan, R. H., Lu, X. T. & Denison, M. R. (2000). Four proteins processed from the replicase gene polyprotein of mouse hepatitis virus colocalize in the cell periphery and adjacent to sites of virion assembly. *J Virol* **74**, 3379–3387.

Brierley, I. (1995). Ribosomal frameshifting viral RNAs. *J Gen Virol* **76**, 1885–1892.

Brierley, I., Digard, P. & Inglis, S. C. (1989). Characterization of an efficient coronavirus ribosomal frameshifting signal: requirement for an RNA pseudoknot. *Cell* **57**, 537–547.

de Haan, C. A., Masters, P. S., Shen, X., Weiss, S. & Rottier, P. J. (2002). The group-specific murine coronavirus genes are not essential, but their deletion, by reverse genetics, is attenuating in the natural host. *Virology* **296**, 177–189.

Drosten, C., Gunther, S., Preiser, W. & 23 other authors (2003). Identification of a novel coronavirus in patients with severe acute respiratory syndrome. *N Engl J Med* **348**, 1967–1976.

Eleouet, J. F., Rasschaert, D., Lambert, P., Levy, L., Vende, P. & Laude, H. (1995). Complete sequence (20 kilobases) of the polyprotein-encoding gene 1 of transmissible gastroenteritis virus. *Virology* **206**, 817–822.

Fouchier, R. A., Kuiken, T., Schutten, M. & 7 other authors (2003). Aetiology: Koch's postulates fulfilled for SARS virus. *Nature* **423**, 240.

Gorbalenya, A. E. (2001). Big nidovirus genome. When count and order of domains matter. *Adv Exp Med Biol* **494**, 1–17.

Gorbalenya, A. E., Koonin, E. V., Donchenko, A. P. & Blinov, V. M. (1989a). Coronavirus genome: prediction of putative functional domains in the non-structural polyprotein by comparative amino acid sequence analysis. *Nucleic Acids Res* **17**, 4847–4861.

Gorbalenya, A. E., Koonin, E. V., Donchenko, A. P. & Blinov, V. M. (1989b). Two related superfamilies of putative helicases involved in replication, recombination, repair and expression of DNA and RNA genomes. *Nucleic Acids Res* **17**, 4713–4730.

Gorbalenya, A. E., Koonin, E. V. & Lai, M. M. (1991). Putative papain-related thiol proteases of positive-strand RNA viruses. Identification of rubi- and aphthovirus proteases and delineation of a novel conserved domain associated with proteases of rubi-, alpha- and coronaviruses. *FEBS Lett* **288**, 201–205.

Hegyi, A. & Ziebuhr, J. (2002). Conservation of substrate specificities among coronavirus main proteases. *J Gen Virol* **83**, 595–599.

Hegyi, A., Friebe, A., Gorbalenya, A. E. & Ziebuhr, J. (2002). Mutational analysis of the active centre of coronavirus 3C-like proteases. *J Gen Virol* **83**, 581–593.

Herold, J., Raabe, T., Schelle-Prinz, B. & Siddell, S. G. (1993). Nucleotide sequence of the human coronavirus 229E RNA polymerase locus. *Virology* **195**, 680–691.

Herold, J., Gorbalenya, A. E., Thiel, V., Schelle, B. & Siddell, S. G. (1998). Proteolytic processing at the amino terminus of human coronavirus 229E gene 1-encoded polyproteins: identification of a papain-like proteinase and its substrate. *J Virol* **72**, 910–918.

Heusipp, G., Harms, U., Siddell, S. G. & Ziebuhr, J. (1997). Identification of an ATPase activity associated with a 71-kilodalton polypeptide encoded in gene 1 of the human coronavirus 229E. *J Virol* **71**, 5631–5634.

Kanjanahaluethai, A. & Baker, S. C. (2000). Identification of mouse hepatitis virus papain-like proteinase 2 activity. *J Virol* **74**, 7911–7921.

Kocherhans, R., Bridgen, A., Ackermann, M. & Tobler, K. (2001). Completion of the porcine epidemic diarrhoea coronavirus (PEDV) genome sequence. *Virus Genes* **23**, 137–144.

Ksiazek, T. G., Erdman, D., Goldsmith, C. S. & 23 other authors (2003). A novel coronavirus associated with severe acute respiratory syndrome. *N Engl J Med* **348**, 1953–1966.

Lai, M. M. C. & Holmes, K. V. (2001). *Coronaviridae*: the viruses and their replication. In *Fields Virology*, 4th edn, pp. 1163–1185. Edited by D. M. Knipe & P. M. Howley. Philadelphia: Lippincott Williams & Wilkins.

Lim, K. P., Ng, L. F. & Liu, D. X. (2000). Identification of a novel cleavage activity of the first papain-like proteinase domain encoded by open reading frame 1a of the coronavirus avian infectious bronchitis virus and characterization of the cleavage products. *J Virol* **74**, 1674–1685.

Liu, D. X., Tibbles, K. W., Cavanagh, D., Brown, T. D. & Brierley, I. (1995). Involvement of viral and cellular factors in processing of polyprotein encoded by ORF1a of the coronavirus IBV. *Adv Exp Med Biol* **380**, 413–421.

Marra, M. A., Jones, S. J., Astell, C. R. & 56 other authors (2003). The genome sequence of the SARS-associated coronavirus. *Science* **300**, 1399–1404.

Pasternak, A. O., van den Born, E., Spaan, W. J. & Snijder, E. J. (2001). Sequence requirements for RNA strand transfer during nidovirus discontinuous subgenomic RNA synthesis. *EMBO J* **20**, 7220–7228.

Peiris, J. S., Chu, C. M., Cheng, V. C. & 14 other authors (2003a). Clinical progression and viral load in a community outbreak of coronavirus-associated SARS pneumonia: a prospective study. *Lancet* **361**, 1767–1772.

Peiris, J. S., Lai, S. T., Poon, L. L. & 13 other authors (2003b). Coronavirus as a possible cause of severe acute respiratory syndrome. *Lancet* **361**, 1319–1325.

Rota, P. A., Oberste, M. S., Monroe, S. S. & 32 other authors (2003). Characterization of a novel coronavirus associated with severe acute respiratory syndrome. *Science* **300**, 1394–1399.

Ruan, Y. J., Wei, C. L., Ee, A. L. & 17 other authors (2003). Comparative full-length genome sequence analysis of 14 SARS coronavirus isolates and common mutations associated with putative origins of infection. *Lancet* **361**, 1779–1785.

Sawicki, S. G. & Sawicki, D. L. (1998). A new model for coronavirus transcription. *Adv Exp Med Biol* **440**, 215–219.

Schneider, T. D. & Stephens, R. M. (1990). Sequence logos: a new way to display consensus sequences. *Nucleic Acids Res* **18**, 6097–6100.

Seybert, A., Ziebuhr, J. & Siddell, S. G. (1997). Expression and characterization of a recombinant murine coronavirus 3C-like proteinase. *J Gen Virol* **78**, 71–75.

Seybert, A., Hegyi, A., Siddell, S. G. & Ziebuhr, J. (2000a). The human coronavirus 229E superfamily 1 helicase has RNA and DNA duplex-unwinding activities with 5′-to-3′ polarity. *RNA* **6**, 1056–1068.

Seybert, A., van Dinten, L. C., Snijder, E. J. & Ziebuhr, J. (2000b). Biochemical characterization of the equine arteritis virus helicase suggests a close functional relationship between arterivirus and coronavirus helicases. *J Virol* **74**, 9586–9593.

Siddell, S. G. (1995). *The Coronaviridae*. New York: Plenum Press.

Snijder, E. J., Bredenbeek, P. J., Dobbe, J. C. & 7 other authors (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* (in press).

Thiel, V. & Siddell, S. G. (1994). Internal ribosome entry in the coding region of murine hepatitis virus mRNA 5. *J Gen Virol* **75**, 3041–3046.

Thiel, V., Rashtchian, A., Herold, J., Schuster, D. M., Guan, N. & Siddell, S. G. (1997). Effective amplification of 20-kb DNA by reverse transcription PCR. *Anal Biochem* **252**, 62–70.

Thiel, V., Herold, J., Schelle, B. & Siddell, S. G. (2001). Infectious RNA transcribed *in vitro* from a cDNA copy of the human coronavirus genome cloned in vaccinia virus. *J Gen Virol* **82**, 1273–1281.

van Dinten, L. C., van Tol, H., Gorbalenya, A. E. & Snijder, E. J. (2000). The predicted metal-binding region of the arterivirus helicase protein is involved in subgenomic mRNA synthesis, genome replication, and virion biogenesis. *J Virol* **74**, 5213–5223.

van Vliet, A. L., Smits, S. L., Rottier, P. J. & de Groot, R. J. (2002). Discontinuous and non-discontinuous subgenomic RNA transcription in a nidovirus. *EMBO J* **21**, 6571–6580.

Walker, J. E., Saraste, M., Runswick, M. J. & Gay, N. J. (1982). Distantly related sequences in the alpha- and beta-subunits of ATP synthase, myosin, kinases and other ATP-requiring enzymes and a common nucleotide binding fold. *EMBO J* **1**, 945–951.

Yao, Z., Jones, D. H. & Grose, C. (1992). Site-directed mutagenesis of herpesvirus glycoprotein phosphorylation sites by recombination polymerase chain reaction. *PCR Methods Appl* **1**, 205–207.

Ziebuhr, J. & Siddell, S. G. (1999). Processing of the human coronavirus 229E replicase polyproteins by the virus-encoded 3C-like proteinase: identification of proteolytic products and cleavage sites common to pp1a and pp1ab. *J Virol* **73**, 177–185.

Ziebuhr, J., Herold, J. & Siddell, S. G. (1995). Characterization of a human coronavirus (strain 229E) 3C-like proteinase activity. *J Virol* **69**, 4331–4338.

Ziebuhr, J., Heusipp, G. & Siddell, S. G. (1997). Biosynthesis, purification, and characterization of the human coronavirus 229E 3C-like proteinase. *J Virol* **71**, 3992–3997.

Ziebuhr, J., Snijder, E. J. & Gorbalenya, A. E. (2000). Virus-encoded proteinases and proteolytic processing in the *Nidovirales*. *J Gen Virol* **81**, 853–879.

Ziebuhr, J., Thiel, V. & 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.