



Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.

Biosynthesis, Purification, and Substrate Specificity of Severe Acute Respiratory Syndrome Coronavirus 3C-like Proteinase*

Received for publication, October 2, 2003, and in revised form, October 13, 2003
Published, JBC Papers in Press, October 15, 2003, DOI 10.1074/jbc.M310875200

Keqiang Fan^{‡§}, Ping Wei^{‡§}, Qian Feng[¶], Sidi Chen[¶], Changkang Huang[‡], Liang Ma[‡],
Bing Lai[‡], Jianfeng Pei^{¶||}, Ying Liu[‡], Jianguo Chen^{¶||}, and Luhua Lai^{‡||**}

From the [‡]State Key Laboratory of Structural Chemistry of Stable and Unstable Species, College of Chemistry, Peking University, Beijing 100871, China, [¶]Department of Cell Biology and Genetics, College of Life Science, Peking University, Beijing 100871, China, and ^{||}Center for Theoretical Biology, Peking University, Beijing 100871, China

The 3C-like proteinase of severe acute respiratory syndrome (SARS) coronavirus has been proposed to be a key target for structural-based drug design against SARS. In order to understand the active form and the substrate specificity of the enzyme, we have cloned, expressed, and purified SARS 3C-like proteinase. Analytic gel filtration shows a mixture of monomer and dimer at a protein concentration of 4 mg/ml and mostly monomer at 0.2 mg/ml, which correspond to the concentration used in the enzyme assays. The linear decrease of the enzymatic-specific activity with the decrease of enzyme concentration revealed that only the dimeric form is active and the dimeric interface could be targeted for structural-based drug design against SARS 3C-like proteinase. By using a high pressure liquid chromatography assay, SARS 3C-like proteinase was shown to cut the 11 peptides covering all of the 11 cleavage sites on the viral polyprotein with different efficiency. The two peptides corresponding to the two self-cleavage sites are the two with highest cleavage efficiency, whereas peptides with non-canonical residues at P2 or P1' positions react slower. The P2 position of the substrates seems to favor large hydrophobic residues. Secondary structure studies for the peptide substrates revealed that substrates with more β -sheetlike structure tend to react fast. This study provides a basic understanding of the enzyme catalysis and a full substrate specificity spectrum for SARS 3C-like proteinase, which are helpful for structural-based inhibitor design against SARS and other coronavirus.

The outbreak of a severe atypical pneumonia in early 2003 has caused 8422 cases and 916 related deaths. The World Health Organization has designated the illness as severe acute respiratory syndrome (SARS).¹ A novel form of coronavirus has been identified as the major cause of SARS (1, 2). The genome of SARS coronavirus has been sequenced within a short period

of time after confirmation of the virus (3, 4). Currently, 23 genome sequences of different variations of SARS coronavirus have been released at the NCBI web site (www.ncbi.nlm.nih.gov/). Coronaviruses are members of positive-stranded RNA viruses featuring the largest viral RNA genomes up to date. The SARS coronavirus replicase gene encompasses two overlapping translation products, polyproteins 1a (~450 kDa) and 1ab (~750 kDa), which are conserved both in length and amino acid sequence to other coronavirus replicase proteins. Polyproteins 1a and 1ab are cleaved by the internally encoded 3C-like proteinase to release functional proteins necessary for virus replication. The SARS 3C-like proteinase is fully conserved among all of the released SARS coronavirus genome sequences and is highly homologous with other coronavirus 3C-like proteinase.

Two crystal structures of coronavirus 3C-like proteinase from transmissible gastroenteritis virus (TGEV) (5) and human coronavirus (hCoV) 229E have been solved (6). The structure of coronavirus 3C-like proteinase contains three domains. The first two domains form a chymotrypsin fold, which is responsible for the catalytic reaction, and the third domain is α -helical with unclear biological function. Coronavirus 3C-like proteinase shares the chymotrypsin fold part with the 3C proteinases from other viruses like rhinovirus called picornavirus (7, 8). The 3C proteinase of rhinovirus has been used as a target to develop drugs against the common cold (9–15). Because of the functional importance of SARS 3C-like proteinase in the viral life cycle, it has been proposed to be a key target for structural-based drug design against SARS (6). Homology modeling for the SARS 3C-like proteinase has been performed by various groups (6, 16, 17), and the conformational flexibility of the substrate-binding site has been studied (17). Virtual screening of chemical compounds libraries has given possible inhibitors (16). An 8-mer peptide has been docked on the model of SARS 3C-like proteinase to study the possible interactions of the protein and the substrate (18)

Similar to other coronaviruses, a sequence analysis revealed 11 cleavage sites of the 3C-like proteinase on the SARS polyprotein. The substrate specificity of coronavirus 3C-like proteinase is determined mainly by the P1, P2, and P1' positions (19). The P1 position has a well conserved Gln residue, and the P2 position has a hydrophobic one. Unlike other previously identified coronavirus 3C-like proteinases, which have Leu/Ile at position P2, SARS 3C-like proteinase also tolerates Phe, Val, and Met residues at P2 position. To study the substrate specificity of SARS 3C-like proteinase, we have cloned, expressed, and purified the protein and studied its activity toward 11 peptides covering the 11 cleavage sites on the virus polyprotein. Our results confirm that purified SARS 3C-like proteinase is active toward substrate peptides mapped from

* This work was supported by the Ministry of Education of China, Peking University, the 863 High Technology Project, and National Natural Science Foundation of China. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

§ Both authors contributed equally to this work.

** To whom correspondence should be addressed: College of Chemistry, Peking University, Beijing 100871, China. Tel.: 86-10-62757486; E-mail: lhlai@pku.edu.cn.

¹ The abbreviations used are: SARS, severe acute respiratory syndrome; TGEV, transmissible gastroenteritis virus; hCoV, human coronavirus; RP-HPLC, reversed-phase high performance liquid chromatography; MALDI-TOF MS, matrix-assisted laser desorption/ionization time-of-flight mass spectroscopy.

the cleavage sites on the polyprotein and reveals the substrate requirement of the proteinase-binding site. This study helps to understand the mechanism of SARS polyprotein process and provides clues for drug design.

EXPERIMENTAL PROCEDURES

Cloning of SARS-CoV 3C-like Proteinase—The reverse transcriptional mixture of SARS-CoV RNA from supernatant fluid of the virus-infected Vero cells using random primers was generously supplied by Dr. Y. Lu from Zhejiang Provincial Center for Disease Prevention and Control. For cloning of the cDNA of 3C-like proteinase of the virus, the first stand cDNA mixture was subjected to PCR amplification using a pair of specific primers, comprising F99 (5'-AGT GGT TTT AGG AAA ATG GCA TTC CC-3') and R108 (5'-TTG GAA GGT AAC ACC AGA GC-3') to amplify a 917-bp fragment containing full-length 3C-like proteinase coding sequence. The PCR products were purified by agarose gel electrophoresis and then cloned directly into pGEM T Easy vector (Promega, Madison, WI). The resultant sequence confirmed that the amplified fragment was the same as that of SARS-CoV 3C-like proteinase.

Construction of Plasmid pET 3CLP-21x—The gene of the SARS 3C-like proteinase was amplified by PCR from the cloning vector (pGEM T Easy) described above using primers 3CLP-Nhe (5'-CACTGCTAGCGGTTTTAGGAAAATGGCATTCCC-3') and 3CLP-Xho (5'-CACTCTCGAGTTGGAAGGTAACACCAGAGC-3'). The PCR product was digested with NheI and XhoI and ligated with NdeI/XhoI-digested pET21a DNA. The resulting plasmid pET 3CLP-21x encodes a 35.1 kDa protein containing a C-terminal His₆-Tag.

Expression of 3C-like Proteinase—pET 3CLP-21x was transformed into *Escherichia coli* BL21(DE3) cells. Cultures were grown at 37 °C in 1 liter of LB medium-containing ampicillin (100 µg/ml) until the A₆₀₀ reached 0.8 and then induced with 0.5 mM isopropyl-1-thio-β-D-galactopyranoside at 30 °C for 3 h. The cells were harvested by centrifugation at 5000 × g for 10 min. The pelleted cells were suspended in buffer A (40 mM Tris-HCl, pH 8.0, 100 mM NaCl, 10 mM imidazole, 7.5 mM 2-mercaptoethanol), at 2% of the original culture volume. After cell lysis by ultrasonic, the cell lysate was separated by centrifugation at 24,000 × g for 20 min. The filtrated supernatant was applied to a nickel-nitrilotriacetic acid column (Qiagen) equilibrated by 50 ml of buffer A. After being washed in 100 ml of buffer A, the 3C-like proteinase was eluted with the gradient of 1–100% buffer B (40 mM Tris-HCl, pH 8.0, 100 mM NaCl, 250 mM imidazole, 7.5 mM 2-mercaptoethanol). The eluted enzyme was concentrated and loaded on a gel filtration column Sephacryl S-200 HR (Amersham Biosciences) equilibrated by 180 ml of buffer C (40 mM Tris-HCl, pH 8.0, 100 mM NaCl, 7.5 mM 2-mercaptoethanol). After elution with another 180 ml buffer C, we received over 95% purified 3C-like proteinase.

Analytic Gel Filtration—The aggregation state of the SARS 3C-like proteinase was analyzed using a Superdex 75 HR column (Amersham Biosciences) on ÄKTA fast protein liquid chromatography. Freshly purified protein was diluted to 4 and 0.2 mg/ml and equilibrated at room temperature for 2 h. 400 µl of 4- and 2-ml 0.2 mg/ml samples were injected into the Superdex 75 HR column and eluted with the buffer (40 mM Tris-HCl, pH 8.0, 100 mM NaCl, 7.5 mM 2-mercaptoethanol) at a flow rate of 0.5 ml/min. The eluted peaks were monitored at 280 nm on fast protein liquid chromatography.

CD Spectra—All of the CD spectra of the proteinase and the substrate peptides were recorded on a Jobin Yvon CD 6 spectrometer at 20 °C. The CD spectra of 3C-like proteinase were recorded in 40 mM Tris-HCl buffer, pH 8.0. For near-UV CD spectrum, a cell with a path length of 1 mm was used and the proteinase concentration is 544 µM, whereas a cell with a path length of 0.1 mm and 54.4 µM of proteinase solution was used for far-UV CD spectrum. The substrate peptides were solved in 20 mM Tris-HCl buffer, pH 7.3, and the final concentration was 2 mM. A cell with a path length of 0.1 mm was used. Each spectrum was the average of four scans corrected by subtracting a spectrum of the buffer solution in the absence of proteinase/peptide recorded under identical condition. Each scan in the range of 184–260 nm for far-UV CD and of 250–320 nm for near-UV CD spectra was obtained by taking data points every 0.5 nm with integration time of 1 s and a 2-nm bandwidth. Thermal denaturation spectrum was recorded by CD at 218 nm using the same condition for far-UV CD spectrum from 10 to 90 °C with an interval of 0.5 °C. Secondary structure content was calculated using the program VARSLC1 (20).

Synthesis of Substrate Peptides—The substrate peptide S01 was synthesized by solid-phase peptide synthesis using standard Fmoc (*N*-(9-fluorenyl)methoxycarbonyl)/*tert*-butyl strategy (21). The cleavage of

the peptide from Rink resin and removal of all of the side-chain protecting groups were achieved in trifluoroacetic acid solution. The crude peptide was purified by reversed-phase high performance liquid chromatography (RP-HPLC, LabPrep System, Gilson) on a Vydac C18 semi-preparative column (218TP510, 10 by 250 mm, Vydac) with gradients of water/acetonitrile containing 0.1% trifluoroacetic acid. Peptide homogeneity and identity were analyzed by analytical HPLC and matrix-assisted laser desorption/ionization time-of-flight mass spectroscopy (MALDI-TOF MS), respectively. Other substrate peptides of HPLC purity from S02 to S11 were purchased from GL Biochemistry Ltd. (Shanghai, China).

Peptide Cleavage—The proteolysis activity of the SARS 3C-like proteinase was determined by peptide cleavage assay. Peptide S01 (Table I) was used as substrate and was incubated with the enzyme in Tris-HCl buffer, pH 7.3, at room temperature. The cleavage mixture was analyzed by RP-HPLC. To verify the cleavage site on the substrate peptide, the two products were purified by semi-preparative RP-HPLC using a 15-min 0–50% linear gradient of acetonitrile in 0.1% trifluoroacetic acid and lyophilized. The relative molecular weights of the products were identified by MALDI-TOF MS (BIFLEX III time-of-flight mass spectrometer, Bruker).

The relative enzyme activity at different pH values was determined in citric acid/phosphate buffer (pH 5, 6, 7, and 8) or glycine/NaOH buffer (pH 9 or 10) containing 6.8 mM dithiothreitol, 2 mM S01 as substrate, and 2.14 µM SARS 3C-like proteinase with a final volume of 50 µl. The cleavage reaction was stopped after 20 min by the addition of 50 µl of 0.1% trifluoroacetic acid aqueous solution and analyzed by RP-HPLC (LabPrep System, Gilson) on a Zorbax C18 analytic column (4.6 × 250 mm, Agilent). Cleavage products were resolved using a 15-min 0–50% linear gradient of acetonitrile in 0.1% trifluoroacetic acid.

To determine the k_{cat}/K_m for the substrate, 0.2 mM of substrate peptide was incubated with SARS 3C-like proteinase in 40 mM Tris-HCl buffer, pH 7.3. The concentration of the enzyme varied from 0.90 to 22.5 µM because of the different cleavage activity to different substrates. Reaction aliquots were removed at different times within 7 h and analyzed by RP-HPLC as described above. k_{cat}/K_m was determined by plotting substrate peak area as Equation 1,

$$\ln PA = C - k_{cat}c_E/K_m \quad (\text{Eq. 1})$$

where PA is the peak area of the substrate peptide, c_E is the total concentration of 3C-like proteinase, and C is an experimental constant.

K_m and k_{cat} of the proteinase for selected substrates were determined by incubation of the substrate peptide at different concentration varying from 2 to 0.1 mM with SARS 3C-like proteinase in 40 mM Tris-HCl buffer, pH 7.3, for 20 min and analyzed by RP-HPLC as described above. The concentration of the enzyme varied from 1.07 to 17.1 µM because of the different cleavage activity to different substrate. Peak areas were calculated by integration and converted to absolute units by using peptide standards. The reaction rate was calculated for all of the cleavage products of two experiments and averaged. K_m and k_{cat} were calculated by the Lineweaver-Burk plot.

RESULTS AND DISCUSSION

Biosynthesis, Purification, and Secondary Structures of Recombinant SARS CoV 3CL Proteinase—The C-terminal His tagged SARS 3C-like proteinase has been successfully expressed in *E. coli* and purified. Induction was first done at 37 °C. As a result, the majority of 3C-like proteinase can be found in the insoluble fraction of the cell lysate. Induction with isopropyl-1-thio-β-D-galactopyranoside was then done at 30 °C for 3 h. As a result, most of the 3C-like proteinase was found in the soluble fraction. The protein was purified by nickel column followed by gel filtration on a Sephacryl S-200 HR column (Fig. 1). Approximately 10 mg of purified protein can be obtained from 1-liter cells. The protein can be concentrated to 10 mg/ml in 50 mM Tris-HCl, 0.1 M NaCl, and 1 mM dithiothreitol, pH 7.3.

The CD spectra of SARS 3C-like proteinase were shown in Fig. 2. Far-UV and near-UV CD spectra show that the purified protein has well defined secondary and tertiary structures. Far-UV CD spectrum (Fig. 2a) shows a positive peak at 196 nm and two negative peaks at 209 and 222 nm, respectively, which clearly indicates for a mixed α and β structure. Calculated secondary structure content shows 26% of α-helix and 23% of β-sheet, similar to those in the TGEV 3C-like proteinase crys-

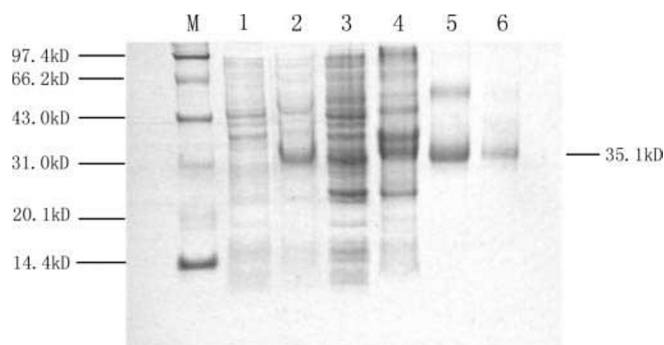


FIG. 1. Expression and purification of the SARS 3C-like proteinase. Samples taken at each step of purification were analyzed on a 12% SDS-polyacrylamide gel, and the protein was stained with Coomassie Brilliant Blue. Lane M, protein molecular mass marker; lanes 1 and 2, cell lysates from non-induced and isopropyl-1-thio- β -D-galactopyranoside-induced *E. coli* BL21(DE3) containing the expression plasmid pET 3CLP-21x; lanes 3 and 4, the supernatant and the deposition separated from the cell lysate by centrifugation; lane 5, collected peak fractions from the nickel-nitrilotriacetic acid column; lane 6, collected peak fractions from the Sephacryl S-200 HR column.

tal structure (22 and 26%, respectively). Near-UV CD spectrum (Fig. 2b) shows a broad positive peak at \sim 280 nm and a small positive shoulder at 291 nm, indicating a well folded tertiary structure. Thermal melting of the protein by monitoring the CD signal at 218 nm (Fig. 2c) gives a sigmoid denaturation curve and a T_m of 61 $^{\circ}$ C, indicating a highly cooperative thermodenaturation.

Aggregation State and Enzyme Activity—Because both the crystal structures of the 3C-like proteinase in TGEV and human coronavirus give dimer structure and the residues at dimeric interface are conserved in coronavirus, it has been proposed that the dimer may be the biological functional form of the protein (5, 6). Dynamic light-scattering experiment shows that both hCoV 229E and TGEV 3C-like proteinases exist as a mixture of monomer (65%) and dimer (35%) at a concentration of 1–2 mg/ml (6). We have performed analytical gel filtration of SARS 3C-like proteinase at different concentrations using a Superdex 75 column (Fig. 3). At the concentration of 4 mg/ml, two peaks corresponding to the monomeric and dimeric form of the protein appeared while only one peak corresponding to the monomeric form was found at a lower concentration of 0.2 mg/ml. The dissociation constant of the dimer was estimated to be around 100 μ M. *In vitro* peptide or protein cleavage assays of coronavirus 3C-like proteinase usually were performed with a protein concentration at micromolar level (22, 23). In this study, we also used a peptide cleavage assay with the SARS 3C-like proteinase concentration not $>$ 22.5 μ M that corresponds mainly to the monomer form of the enzyme. This raises an interesting question whether the minor amount of the dimer plays a major role in catalysis, although we are still not clear about the exact form of the protein in the cell under a molecular crowding environment.

To answer this question, the proteolysis activity of SARS CoV 3C-like proteinase at different enzyme concentrations was studied. Peptide S01 was used as substrate and was incubated with the proteinase in Tris-HCl buffer, pH 7.3, at room temperature. The cleavage mixture was analyzed by RP-HPLC. The peak area of the substrate S01 decreases as reaction time increases, whereas the area of two newly formed peaks increases. The products were collected and identified by MALDI-TOF MS with experimental $(M + H)^+$ of 593 and 618, respectively, that were identical to the theoretical $(M + H)^+$ of the C-terminal pentapeptide and N-terminal hexapeptide. This confirms that the substrate is cleaved by the SARS CoV 3C-like proteinase at the predicted fragile Gln-Ser peptide bond.

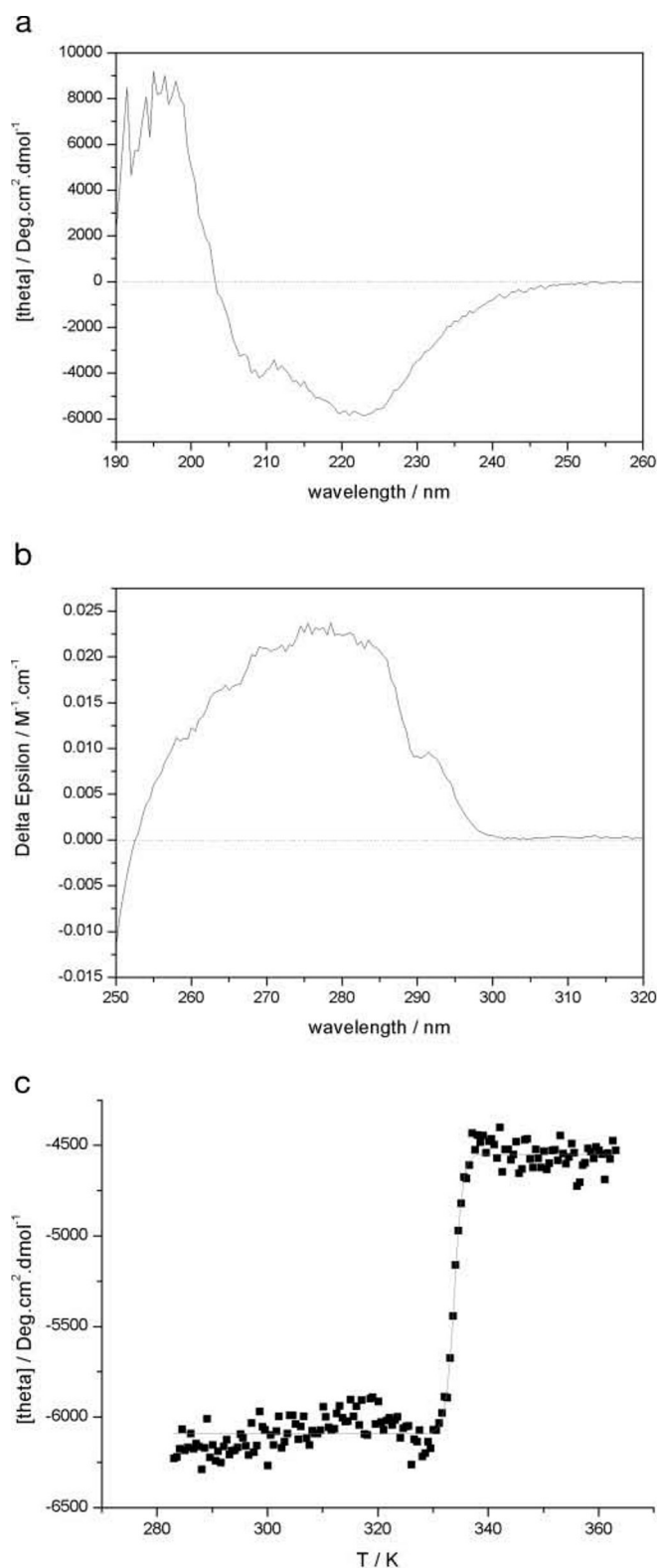


FIG. 2. CD spectra of SARS 3C-like proteinase. The CD spectra of 3C-like proteinase were recorded in 40 mM Tris-HCl buffer, pH 8.0 at 20 $^{\circ}$ C. a, far-UV CD spectrum shows a positive peak at 196 nm and two negative peaks at 209 and 222 nm, respectively, indicating a mixed α and β structure. b, near-UV CD spectrum shows a broad positive peak at \sim 280 nm and a small positive shoulder at 291 nm. c, thermal melting of the protein by monitoring the CD signal at 218 nm gives a sigmoid denaturation curve and a T_m of 61 $^{\circ}$ C

The observed k_{cat}/K_m was determined as described under “Experimental Procedures” at different proteinase concentrations between 4.5 and 0.9 μ M (shown in Fig. 4). The observed

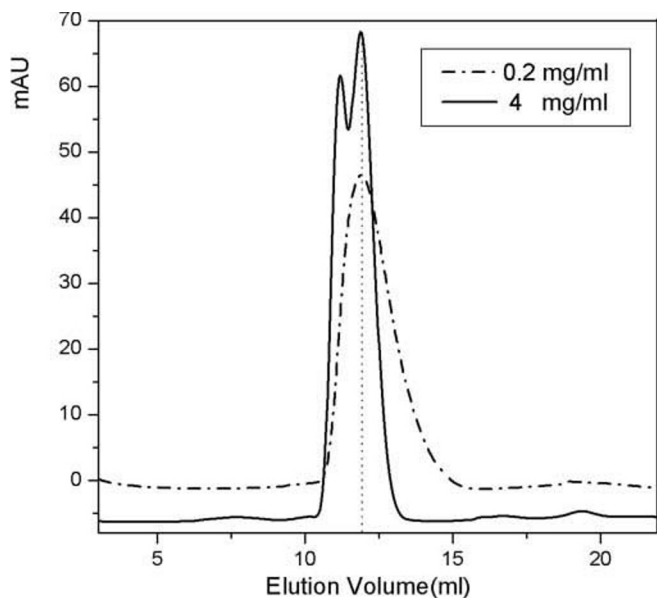


FIG. 3. Aggregation state of SARS 3C-like proteinase on analytical gel filtration column. Superdex 75 HR column was used on ÄKTA fast protein liquid chromatography for the analysis. Solid line, 4 mg/ml; dot-and-dash line, 0.2 mg/ml. Two peaks corresponding to monomeric and dimeric states were detected at 4 mg/ml, whereas at 0.2 mg/ml, only one peak corresponding to the monomeric state appears.

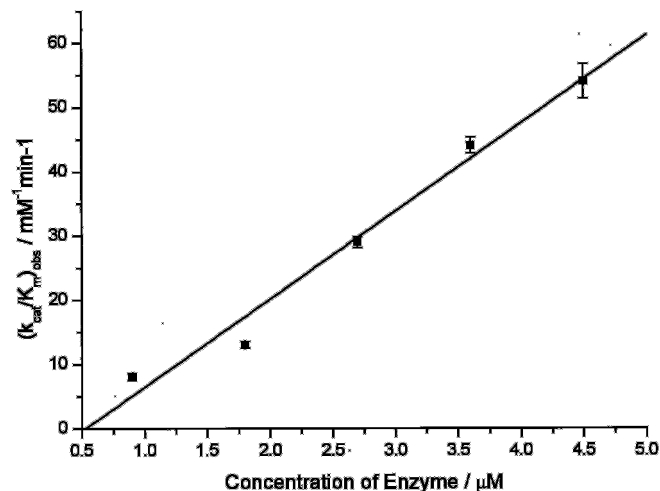


FIG. 4. Enzyme activity of SARS CoV 3C-like proteinase at different concentrations. The proteolysis activity of expressed SARS CoV 3C-like proteinase at different concentrations was determined by incubating 0.2 mM S01 with 3C-like proteinase in 40 mM Tris-HCl buffer, pH 7.3. The observed k_{cat}/K_m increases in a linear manner, whereas the enzyme concentration increases, indicating the dimer form is the active form of the proteinase.

k_{cat}/K_m increases linearly with the increase of the enzyme concentration. Using the estimated dissociation constant of $\sim 100 \mu\text{M}$, we can deduce from the linear dependence that the monomeric form of SARS CoV 3C-like proteinase almost has no catalytic activity, whereas the k_{cat}/K_m of the dimeric form is $\sim 1.4 \times 10^3 \text{ mM}^{-1} \text{ min}^{-1}$. This result supports the previous prediction that the dimer of the 3C-like proteinase is the active form of the enzyme and also explains that the low *in vitro* activity of CoV 3C-like proteinases compared with other monomeric virus 3C proteinases is because of the low concentration of the active dimeric form under the assay conditions.

Substrate Specificity—Eleven 11-mer peptides derived from the 11 cleavage sites of 3C-like proteinase upon the SARS polyprotein have been synthesized and tested for SARS 3C-like proteinase cleavage experiment (Table I). The proteolysis ac-

tivity of expressed SARS CoV 3C-like proteinase at different pH was tested as described under “Experimental Procedures.” When peptide S01 was used as substrate as shown in Fig. 5, the proteinase had the highest activity at around pH 7 and decreased at another pH, resulting a bell-shaped curve. When pH decreased to 5, the proteolytic activity of the enzyme decreased to 10% because of the protonation of the catalytic His-41.

To investigate the relationship between the proteolysis activity and the secondary structure of the substrates, far-UV CD spectra were recorded for all of the peptides at 2 mM (for S09 the peptide concentration was 0.2 mM because of its low solubility) (see Fig. 6a). The secondary structure contents were calculated using VARSLC1 (20) with the CD data in the range of 184–260 nm for each peptide with the exception of S09 (which has very low solubility) and summarized in Table II. Although not obvious in the CD spectra, each peptide forms more or less sheet structures. The calculated sheet content reflects the formation of a β -strandlike extended conformation in equilibrium with other turnlike or random conformations. Fig. 6b shows an interesting tendency in which substrate peptide with small k_{cat}/K_m has relatively less content of sheet but more of other unordered structure, suggesting that the formation of β -strandlike extended structure is favored for binding to the enzyme and proteolysis. As shown in the crystal structure of TGEV 3C-like proteinase, residues P5 to P3 of the substrate form an antiparallel β -sheet with segment 164–167 of the long strand eII of the enzyme on one side and with segment 186–191 on the other side (6). The binding modes of substrate to different coronavirus 3C-like proteinases were predicted to be identical by comparing the crystal structure of the substrate-binding regions of the free proteinases of hCoV and SARS CoV and of TGEV proteinase in complex with a hexapeptidyl chloromethyl ketone inhibitor (6). Substrate peptide that tends to form an extended β -strandlike conformation would be the preferred choice to bind with the enzyme, resulting in a small K_m and a large k_{cat}/K_m . It is also supported by a previous study (25) on 3C proteinase in which helix-destabilizing residues were often found in close proximity to cleavage sites.

Of all of the 11 peptides tested, S01 and S02 were the most suitable substrates for SARS CoV 3C-like proteinase cleavage, which were derived from the N-terminal and C-terminal self-cleavage sites, respectively. K_m of S01 was determined as $1.15 \pm 0.28 \text{ mM}$, which is three times larger than that of hCoV 229E 3C-like proteinase for a 15-mer substrate, $0.39 \pm 0.07 \text{ mM}$. The relative large K_m was counteracted by a large k_{cat} , $12.2 \pm 2.9 \text{ min}^{-1}$, which is an order larger than that of the five other substrates whose k_{cat} and K_m values are also determined by the Lineweaver-Burk plot (Table I).

Most of the cleavage sites of coronavirus polyprotein have a conserved (Leu/Ile)-Gln \downarrow (Ser, Ala, or Gly) core sequence (the cleavage site is indicated by \downarrow). However, the SARS coronavirus polyprotein has three noncanonical cleavage sites with Phe, Val, or Met in the P2 position and one noncanonical cleavage site with Asn in the P1' position. Here, peptides S02, S03, and S07 are derived from the three noncanonical cleavage sites at P2 and peptide S05 is derived from the one noncanonical cleavage site at P1'. With the dipeptide sequence of Asn-Asn at positions P1' and P2', peptide S05 has relatively low catalytic efficiency, which is the second lowest one among all of the 11 peptides. Peptides S02 and S03 have a Phe and a Val, respectively, in the position P2 and have similar cleavage activity as other substrates with Leu in P2 position, indicating that Phe or Val can also be well fitted in the large hydrophobic S2 pocket so that the substitutions are tolerated for 3C-like proteinase cleavage. The substitution by a relative large hydrophobic amino acid Phe did not affect the substrate binding, whereas

TABLE I
Catalytic parameters and relative cleavage efficiencies of eleven peptides that represent all of the eleven cleavage sites in SARS coronavirus polyprotein

Peptide	Cleavage site	Sequence	K_m	k_{cat}	k_{cat}/K_m	$(k_{cat}/K_m)_{rel}$
			<i>mM</i>	<i>min⁻¹</i>	<i>mM⁻¹ min⁻¹</i>	
S01	P1/P2	TSAVLQ/SGFRK-NH ₂	1.15 ± 0.28	12.2 ± 2.9	10.6 ± 0.6 ^a	1.00
S02	P2/P3	SGVTFQ/GKFKK	0.583 ± 0.086	2.55 ± 0.35	4.38 ± 0.26 ^a	0.41
S03	P3/P4	KVATVQ/SKMSD			0.353 ± 0.013 ^b	0.03
S04	P4/P5	NRATLQ/AIASE			0.556 ± 0.126 ^b	0.05
S05	P5/P6	SAVKLQ/NNELS			0.202 ± 0.003 ^b	0.02
S06	P6/P7	ATVRLQ/AGNAT	1.44 ± 0.47	3.29 ± 1.07	2.29 ± 0.09 ^a	0.22
S07	P7/P8	REPLMQ/SADAS			0.0176 ± 0.0022 ^b	0.002
S08	P8/P9	PHTVLQ/AVGAC	1.94 ± 0.66	1.68 ± 0.57	0.865 ± 0.028 ^a	0.08
S09	P9/P10	NVATLQ/AENVT			0.976 ± 0.014 ^b	0.09
S10	P10/P11	TFTRLQ/SLENV	0.286 ± 0.020	0.847 ± 0.051	2.96 ± 0.11 ^a	0.28
S11	P11/P12	FYPKLQ/ASQAW	0.549 ± 0.105	1.57 ± 0.28	2.86 ± 0.21 ^a	0.27

^a Determined by Lineweaver-Burk plot.

^b Determined by fitting reaction kinetic data as in Equation 1 at low substrate concentration.

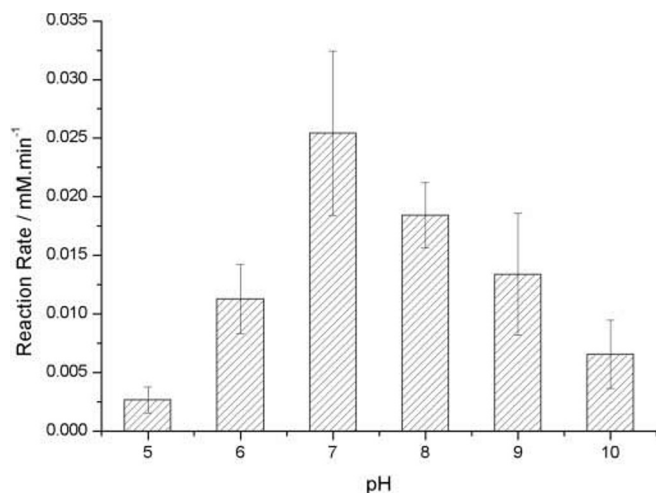


FIG. 5. Enzyme activity of SARS CoV 3C-like proteinase at different pH values. The proteolysis activity of expressed SARS CoV 3C-like proteinase at different pH values was determined in citric acid/phosphate buffer (pH 5, 6, 7, and 8) or glycine/NaOH buffer (pH 9 or 10) containing 6.8 mM dithiothreitol, 2 mM S01 as substrate, and 2.14 μ M 3C-like proteinase.

substitution by a relative small hydrophobic amino acid Val decreased the cleavage activity, indicating an important role of pocket S2 in anti-SARS drug design.

On the other hand, peptide S07, which has a methionine in the P2 position, is the substrate with least cleavage activity among 11 substrates. The low cleavage activity of S07 may be caused by the mutant from Leu to Met, although it may also be affected by other reasons such as different secondary structures. As shown in the far-UV CD spectrum, peptide S07 may adopt an unusual helix-like, type II polyproline conformation, which has a strong negative peak at ~200 nm and a positive peak at ~230 nm. The formation of the polyproline helix will disturb the conformation transition to β -strandlike-extended structure of the substrate peptide and disable the binding of the substrate to the enzyme. It is also supported by the factor that this peptide has the least content of sheet.

Hegyí and Ziebuhr (19) have studied the conservation of substrate specificities among coronavirus 3C-like proteinase from three coronavirus by using peptides corresponding to the four cleavage sites on the viral polyprotein. Our findings that the two peptides S01 (P1/P2) and S02 (P2/P3) corresponding to the two self-cleavage sites of the SARS 3C-like proteinase are the two most reactive ones and S05 (P5/P6) is less reactive are comparable to their results. This implies that SARS 3C-like proteinase follows the same substrate specificity rules governing all of the coronavirus 3C-like proteinase. As we have stud-

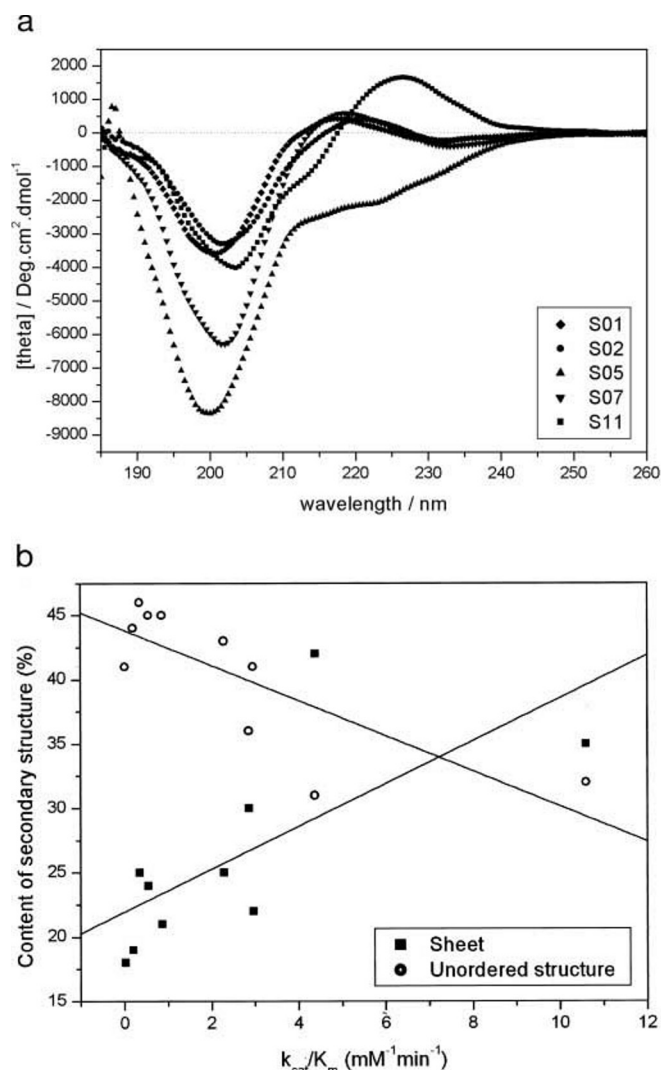


FIG. 6. The relationship of secondary structure contents in substrate peptides with cleavage efficiency. *a*, far-UV CD spectra of peptide substrates. \blacklozenge , S01; \bullet , S02; \blacktriangle , S05; \blacktriangledown , S07; and \blacksquare , S11. *b*, calculated content of secondary structure versus k_{cat}/K_m profile. This figure shows an interesting tendency of which substrate peptide with small k_{cat}/K_m has relatively less content of sheet (\blacksquare) but more content of unordered structure (\circ), suggesting that the formation of β -strandlike extended structure is favored for binding to the enzyme and proteolysis.

ied, the substrate specificities of SARS 3C-like proteinase using peptides covering all of the 11 cleavage sites try to correlate not only the conserved sequences but also secondary structures of

TABLE II
Calculated secondary structure contents for the 11-peptide substrates

	Helix	Sheet	Turn	Other
S01	5	35	27	32
S02	3	42	24	31
S03	2	25	28	46
S04	2	24	30	45
S05	10	19	27	44
S06	2	25	30	43
S07	1	18	40	41
S08	5	21	29	45
S10	4	22	33	41
S11	2	30	32	36

the substrate peptides. The full spectrum of substrate specificity for SARS 3C-like proteinase studied here can be extended to understand other coronavirus 3C-like proteinases.

In summary, the recombinant SARS 3C-like proteinase has been successfully cloned and expressed. The purified protein exists in a mixture of monomer and dimer at a concentration of 4 mg/ml but mostly monomer at 0.2 mg/ml. The specific activity of the enzyme decreases linearly with the decrease of enzyme concentration implying that only the dimeric form is active and that the dimeric interface could be targeted for structural based drug design against SARS 3C-like proteinase. The enzyme can cut the 11 peptides covering all of the 11 cleavage sites on the viral polyprotein with different efficiency. Substrates with a more β -sheetlike structure tend to react fast. The P2 position of the substrates seems to favor large hydrophobic residues. This study provides basic understandings of the enzyme catalysis and substrate specificity for SARS 3C-like proteinase and helpful information for structurally based inhibitor design.

REFERENCES

- Drosten, C., Gunther, S., Preiser, W., van der, W. S., Brodt, H. R., Becker, S., Rabenau, H., Panning, M., Kolesnikova, L., Fouchier, R. A., Berger, A., Burguiere, A. M., Cinatl, J., Eickmann, M., Escriou, N., Grywna, K., Kramme, S., Manuguerra, J. C., Muller, S., Rickerts, V., Sturmer, M., Vieth, S., Klenk, H. D., Osterhaus, A. D., Schmitz, H., and Doerr, H. W. (2003) *N. Engl. J. Med.* **348**, 1967–1976
- Ksiazek, T. G., Erdman, D., Goldsmith, C. S., Zaki, S. R., Peret, T., Emery, S., Tong, S., Urbani, C., Comer, J. A., Lim, W., Rollin, P. E., Dowell, S. F., Ling, A. E., Humphrey, C. D., Shieh, W. J., Guarner, J., Paddock, C. D., Rota, P., Fields, B., DeRisi, J., Yang, J. Y., Cox, N., Hughes, J. M., LeDuc, J. W., Bellini, W. J., and Anderson, L. J. (2003) *N. Engl. J. Med.* **348**, 1953–1966
- 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) *Science* **300**, 1394–1399
- 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., Girm, 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., Drobot, 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., Kraiden, M., Petric, M., Skowronski, D. M., Upton, C., and Roper, R. L. (2003) *Science* **300**, 1399–1404
- Anand, K., Palm, G. J., Mesters, J. R., Siddell, S. G., Ziebuhr, J., and Hilgenfeld, R. (2002) *EMBO J.* **21**, 3213–3224
- Anand, K., Ziebuhr, J., Wadhvani, P., Mesters, J. R., and Hilgenfeld, R. (2003) *Science* **300**, 1763–1767
- Matthews, D. A., Smith, W. W., Ferre, R. A., Condon, B., Budahazi, G., Sisson, W., Villafranca, J. E., Janson, C. A., McElroy, H. E., Gribskov, C. L., and Worland, S. (1994) *Cell* **77**, 761–771
- Seipelt, J., Guarne, A., Bergmann, E., James, M., Sommergruber, W., Fita, I., and Skern, T. (1999) *Virus Res.* **62**, 159–168
- Dragovich, P. S., Prins, T. J., Zhou, R., Brown, E. L., Maldonado, F. C., Fuhrman, S. A., Zalman, L. S., Tuntland, T., Lee, C. A., Patick, A. K., Matthews, D. A., Hendrickson, T. F., Kosa, M. B., Liu, B., Batugo, M. R., Gleeson, J. P., Sakata, S. K., Chen, L., Guzman, M. C., Meador, J. W., III, Ferre, R. A., and Worland, S. T. (2002) *J. Med. Chem.* **45**, 1607–1623
- Dragovich, P. S., Prins, T. J., Zhou, R., Johnson, T. O., Brown, E. L., Maldonado, F. C., Fuhrman, S. A., Zalman, L. S., Patick, A. K., Matthews, D. A., Hou, X., Meador, J. W., Ferre, R. A., and Worland, S. T. (2002) *Bioorg. Med. Chem. Lett.* **12**, 733–738
- Dragovich, P. S., Zhou, R., Webber, S. E., Prins, T. J., Kwok, A. K., Okano, K., Fuhrman, S. A., Zalman, L. S., Maldonado, F. C., Brown, E. L., Meador, J. W., III, Patick, A. K., Ford, C. E., Brothers, M. A., Binford, S. L., Matthews, D. A., Ferre, R. A., and Worland, S. T. (2000) *Bioorg. Med. Chem. Lett.* **10**, 45–48
- Dragovich, P. S., Webber, S. E., Prins, T. J., Zhou, R., Marakovits, J. T., Tikhe, J. G., Fuhrman, S. A., Patick, A. K., Matthews, D. A., Ford, C. E., Brown, E. L., Binford, S. L., Meador, J. W., III, Ferre, R. A., and Worland, S. T. (1999) *Bioorg. Med. Chem. Lett.* **9**, 2189–2194
- Dragovich, P. S., Prins, T. J., Zhou, R., Webber, S. E., Marakovits, J. T., Fuhrman, S. A., Patick, A. K., Matthews, D. A., Lee, C. A., Ford, C. E., Burke, B. J., Rejto, P. A., Hendrickson, T. F., Tuntland, T., Brown, E. L., Meador, J. W., III, Ferre, R. A., Harr, J. E., Kosa, M. B., and Worland, S. T. (1999) *J. Med. Chem.* **42**, 1213–1224
- Dragovich, P. S., Prins, T. J., Zhou, R., Fuhrman, S. A., Patick, A. K., Matthews, D. A., Ford, C. E., Meador, J. W., III, Ferre, R. A., and Worland, S. T. (1999) *J. Med. Chem.* **42**, 1203–1212
- Dragovich, P. S., Webber, S. E., Babine, R. E., Fuhrman, S. A., Patick, A. K., Matthews, D. A., Reich, S. H., Marakovits, J. T., Prins, T. J., Zhou, R., Tikhe, J., Littlefield, E. S., Bleckman, T. M., Wallace, M. B., Little, T. L., Ford, C. E., Meador, J. W., III, Ferre, R. A., Brown, E. L., Binford, S. L., DeLisle, D. M., and Worland, S. T. (1998) *J. Med. Chem.* **41**, 2819–2834
- Xiong, B., Gui, C. S., Xu, X. Y., Luo, C., Chen, J., Luo, H. B., Chen, L. L., Li, G. W., Sun, T., Yu, C. Y., Yue, L. D., Duan, W. H., Shen, J. K., Qin, L., Shi, T. L., Li, Y. X., Chen, K. X., Luo, X. M., Shen, X., Shen, J. H., and Jiang, H. L. (2003) *Acta Pharmacol. Sin.* **24**, 497–504
- Liu, S., Pei, J., Chen, H., Zhu, X., Liu, Z., Ma, W., He, F., and Lai, L. (2003) *Beijing Da. Xue. Xue. Bao.* **35**, (suppl.) 62–65
- Chou, K. C., Wei, D. Q., and Zhong, W. Z. (2003) *Biochem. Biophys. Res. Commun.* **308**, 148–151
- Hegy, A., and Ziebuhr, J. (2002) *J. Gen. Virol.* **83**, 595–599
- Johnson, W. C., Jr. (1988) *Annu. Rev. Biophys. Biophys. Chem.* **17**, 145–166
- Fields, G. B., and Noble, R. L. (1990) *Int. J. Pept. Protein Res.* **35**, 161–214
- Ziebuhr, J., Herold, J., and Siddell, S. G. (1995) *J. Virol.* **69**, 4331–4338
- Ziebuhr, J., Heusipp, G., and Siddell, S. G. (1997) *J. Virol.* **71**, 3992–3997
- Deleted in proof
- Cordingley, M. G., Callahan, P. L., Sardana, V. V., Garsky, V. M., and Colonno, R. J. (1990) *J. Biol. Chem.* **265**, 9062–9065