



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.

Critical Assessment of Important Regions in the Subunit Association and Catalytic Action of the Severe Acute Respiratory Syndrome Coronavirus Main Protease*

Received for publication, March 8, 2005, and in revised form, April 4, 2005
Published, JBC Papers in Press, April 14, 2005, DOI 10.1074/jbc.M502556200

Wen-Chi Hsu, Hui-Chuan Chang, Chi-Yuan Chou, Pui-Jen Tsai, Pei-In Lin, and Gu-Gang Chang‡

From the Faculty of Life Sciences, Institute of Biochemistry, and Structural Biology Program,
National Yang-Ming University, Taipei 112, Taiwan

The severe acute respiratory syndrome (SARS) coronavirus (CoV) main protease represents an attractive target for the development of novel anti-SARS agents. The tertiary structure of the protease consists of two distinct folds. One is the N-terminal chymotrypsin-like fold that consists of two structural domains and constitutes the catalytic machinery; the other is the C-terminal helical domain, which has an unclear function and is not found in other RNA virus main proteases. To understand the functional roles of the two structural parts of the SARS-CoV main protease, we generated the full-length of this enzyme as well as several terminally truncated forms, different from each other only by the number of amino acid residues at the C- or N-terminal regions. The quaternary structure and K_d value of the protease were analyzed by analytical ultracentrifugation. The results showed that the N-terminal 1–3 amino acid-truncated protease maintains 76% of enzyme activity and that the major form is a dimer, as in the wild type. However, the amino acids 1–4-truncated protease showed the major form to be a monomer and had little enzyme activity. As a result, the fourth amino acid seemed to have a powerful effect on the quaternary structure and activity of this protease. The last C-terminal helically truncated protease also exhibited a greater tendency to form monomer and showed little activity. We concluded that both the C- and the N-terminal regions influence the dimerization and enzyme activity of the SARS-CoV main protease.

Severe acute respiratory syndrome (SARS)¹ is an acute respiratory illness that causes an atypical and highly contagious pneumonia. It has spread in many countries and areas. The international outbreak of this disease resulted in great damage to public health (1–3). A novel form of coronavirus (CoV), SARS-CoV, has been identified to be the major cause of SARS (3–8).

Coronaviruses are positive-sense, single-stranded RNA viruses. The genome of SARS-CoV is ~30,000 nucleotides in

length, and the organization is similar to that of other coronaviruses. The replicase gene encodes two overlapping polyproteins, polyprotein 1a (~450 kDa) and polyprotein 1ab (~750 kDa). The polyproteins are cleaved by the internally encoded main protease (M^{pro} , 3CL), which is required for the production of new infectious viruses. The main protease represents an attractive target for the development of novel anti-viral agents due to the functional importance of this enzyme in the viral life cycle (9–12).

The crystal structures of reported CoV, including the SARS-CoV main proteases, are homodimers (13–15). Each protomer of the main protease is composed of three structural domains (Fig. 1). The first two domains of the SARS-CoV main protease have an antiparallel β -barrel structure, which is similar to the other CoV proteases, and form a chymotrypsin-like fold responsible for catalytic reactions (15). The active site containing a catalytic dyad defined by His-41 and Cys-145 is located between domains I and II. The third domain contains five α -helices with an unclear biological function. The domain III of one protomer and the domain II of another form a contacting region in the dimer. The N terminus (N-finger containing amino acid residues 1–7) is seated at this region and plays an important role in dimerization (16).

We have demonstrated that the major quaternary structure of SARS-CoV main protease at neutral pH is a dimer, which is the catalytically competent form (17). It is ultimately important to understand the factors that control dimerization, as the dissociated monomer could be enzymatically inactive. The C-terminal helical domain interacts with the active site of another protomer in the dimer and switches the enzyme molecule from the inactive form to the active form (18). The structural and biochemical data also show that the N-terminal residues 1–7 play an important role in the dimerization and formation of the active site of SARS main protease (15). N-terminal truncation of the whole N-finger, $\Delta(1-7)$ results in almost complete loss of enzymatic activity (19).

In this report, we studied critically the functional role of the N and C termini by serial truncations. We report the stability and structure-function relationship of the full-length SARS-CoV main protease in comparison with the various truncated forms. Our results demonstrate that both N- and C-terminal regions are involved in the enzyme activity as well as in dimerization. We have narrowed down the critical amino acid residues to the fourth amino acid residue of the N-terminal and the last helical amino acids of the C-terminal region as those involved in dimerization to give a correct conformation of the active site.

EXPERIMENTAL PROCEDURES

Construction of WT and N- and C-terminally Truncated SARS-CoV Main Proteases—The plasmids containing the full-length main protease

* This work was supported by the National Science Council, Republic 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.

‡ To whom correspondence should be addressed: Faculty of Life Sciences, National Yang-Ming University, 155 Li-Nong St., Section 2, Taipei 112, Taiwan. Tel.: 886-2-2820-1854; Fax: 886-2-2820-1886; E-mail: ggchang@ym.edu.tw.

¹ The abbreviations used are: SARS, severe acute respiratory syndrome; CoV, coronavirus; WT, wild type; CD, circular dichroism; BME, 2-mercaptoethanol; PBS, phosphate-buffered saline.

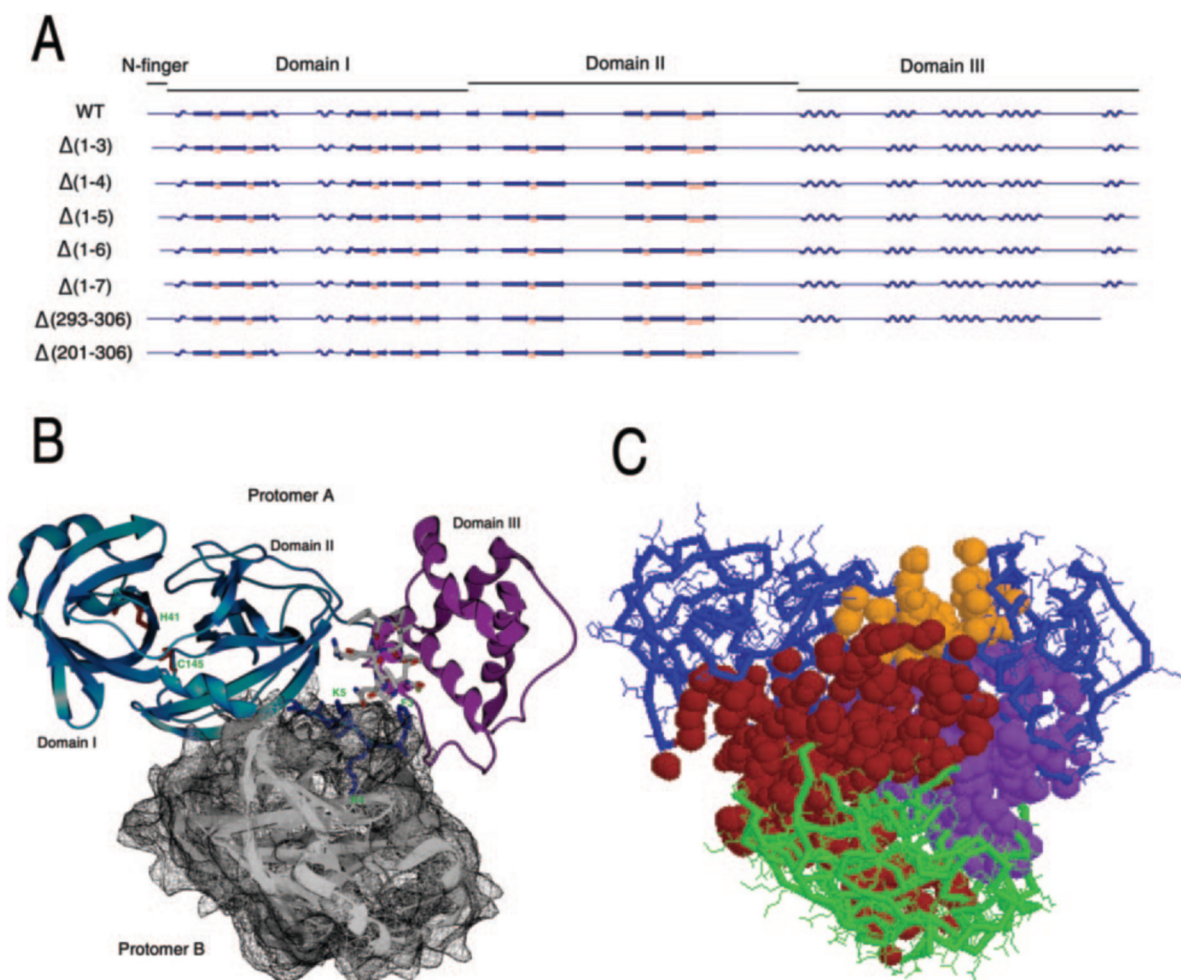


FIG. 1. Structural features of the dimeric SARS-CoV main protease. *A*, secondary structure of the enzyme and the deleted regions of the various deletion mutants. *B*, the enzyme structure (Protein Data Bank code 1uk3) is presented as *ribbons* and labeled as to coincide with the *stick model*. Protomer A is colored with the catalytic domains I and II in *blue*, and the helical domain III in *purple*. The N-finger residues and the last α -helix of protomer A are shown in *stick model* form in *blue* and CPK (oxygen in *red*, nitrogen in *blue*, sulfur in *yellow*, and *gray* for all other atoms), respectively. The catalytic dyad His-41 and Cys-145 are in *red*. The surface of protomer B is shown in *mesh model* form. This figure was generated with Spock software. *C*, the backbone of the enzyme is shown as a *stick model* in *blue* and *green*, and the major pockets are shown as a *space-filling model*. This figure was generated with CASTp (39).

were kindly provided by Dr. Shao-Hung Wang (Genome Research Center, National Yang-Ming University, Taipei, Taiwan). The genes of the full-length SARS-CoV main protease were amplified by polymerase chain reaction with appropriate primers. The forward primer for the full-length WT SARS-CoV main protease was 5'-GGTGGTCATATGAGTGGTTTATAGG-3', and the reverse primer was 5'-AACTCGAGGGTAACACCAGAG-3'. After digestion with BglII and XhoI, the PCR product was cut into two fragments, 168 and 747 bp. The 168-bp fragment was then digested with NdeI. Finally, the 168-bp NdeI-BglII and 747-bp BglII-XhoI fragments were co-ligated to the NdeI and XhoI sites of the vector pET-29a(+) (Novagen, Madison, WI).

The N-terminally truncated proteases were made by PCR. The forward primers were as follows: $\Delta(1-3)$, 5'-GGAGATATACATATGAGAAAATGGCATTC-3'; $\Delta(1-4)$, 5'-GGAGATATACATATGAAAATGGCATTCCCG-3'; $\Delta(1-5)$, 5'-GGAGATATACATATGATGGCATTCCCGTCA-3'; $\Delta(1-6)$, 5'-GGAGATATACATATGGCATTCCCGTCAAGC-3'; and $\Delta(1-7)$, 5'-GGAGATATACATATGTTCCCGTCAAGCAAA-3'. The reverse mutagenic primers were as follows: $\Delta(1-3)$, 5'-GAATGCCATTTCCTCATATGTATATCTCC-3'; $\Delta(1-4)$, 5'-CGGGAATGCCATTTTCATATGTATATCTCC-3'; $\Delta(1-5)$, 5'-TGACGGGAATGCCATCATATGTATATCTCC-3'; $\Delta(1-6)$, 5'-GCCTGACGGGAATGCCATATGTATATCTCC-3'; and $\Delta(1-7)$, 5'-TTTGCCCTGACGGGAACATATGTATATCTCC-3'.

The pET-SARS-CoV main protease vector was used as the template. The DNA polymerase *Pfu* (Promega, Madison, WI) extended and incorporated the mutagenic primers in the process of PCR. After 16–18 temperature cycles, the N-terminally truncated plasmid containing staggered nicks was generated. The PCR products were then treated with DpnI (New England Biolabs, Beverly, MA) to digest the template. Finally, the vector containing the protease cDNA with the desired

mutation was transformed into *Escherichia coli*. The C-terminally truncated proteases were subsequently amplified by PCR using the following sequences. The forward primer, 5'-TGAAGATCTGCTCATTCGCA-A-3'; reverse primer of $\Delta(293-306)$, 5'-AACTCGACTGTAACTCATCTTC-3'; reverse primer of $\Delta(201-306)$, 5'-AACTCGACTATGGTTGTGTCTG-3'. After digestion with BglII and XhoI, the PCR products were inserted into the BglII and XhoI sites of the pET-SARS-CoV main protease. The DNA sequences of the full-length, N- and C-terminally truncated SARS-CoV main proteases were checked by autosequencing. The recombinant SARS-CoV main protease has a His tag at the C terminus. This His tag was not removed, because earlier reports seemed to rule out the possible effect of the His tag on the dimeric structure or enzyme activity (17).

Expression and Purification of WT and N- and C-terminally Truncated SARS-CoV Main Proteases—The modified plasmids of the recombinant proteases were transformed into the *E. coli* strain BL21 (DE3)-competent cells. The cells were grown at 37 °C in Luria-Bertani medium with 50 μ g/ml kanamycin until the absorbance at 600 nm reached 0.8 and were then induced by 1 mM isopropyl-1-thio- β -D-galactoside at 18 °C overnight. The cells were centrifuged at 5,000 \times g, 4 °C for 10 min. The supernatant was removed, and the pelleted cells were then suspended in binding buffer (20 mM Tris-HCl, 300 mM NaCl, and 2 mM BME, pH 7.6). The cells were sonicated for 10 min at 10-s burst cycles at 300 W with a 10-s cooling period between each burst. The cell debris was removed by centrifugation (10,000 \times g at 4 °C for 25 min). One ml of binding buffer-equilibrated nickel-nitrilotriacetic acid slurry (Qiagen, Hilden, Germany) was then added to the soluble lysate, and the solution was mixed gently at 4 °C for 50 min to equilibrium. The lysate-nickel-nitrilotriacetic acid mixture was then loaded into a col-

FIG. 2. CD spectra of the full-length WT and truncated SARS-CoV main proteases. Far-UV CD spectra of all recombinant SARS-CoV main proteases were monitored at a 0.8 mg/ml concentration in 10 mM PBS buffer (pH 7.6) at 25 °C, where *deg* is the ellipticity in degrees.

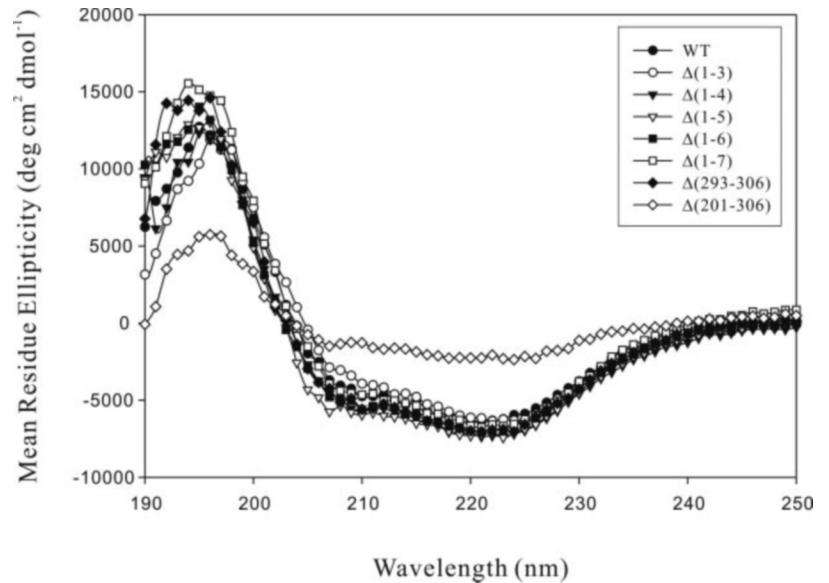


TABLE I
Structural characteristics of the recombinant SARS-CoV main proteases in crystal and solution

Protease	α -Helix	β -Strand	Turns	Unordered	Normalized NRMSD ^b	T_m ^c	$\langle \lambda \rangle^d$
						°C	nm
1uk3 ^a	22	25	20	33			
WT	21	27	22	30	0.030	54.8	342.1
$\Delta(1-3)$	20	25	21	34	0.023	54.8	341.8
$\Delta(1-4)$	22	23	19	35	0.050	52.0	341.8
$\Delta(1-5)$	22	26	19	33	0.028	53.4	341.8
$\Delta(1-6)$	21	26	18	34	0.025	51.5	ND ^e
$\Delta(1-7)$	21	28	20	31	0.024	54.6	342.1
$\Delta(293-306)$	24	27	19	30	0.024	48.4	342.1
$\Delta(201-306)$	4	31	21	43	0.140	55.0	331.1

^a Crystal structure determined by x-ray crystallography (Protein Data Bank code 1uk3).

^b Normalized root mean S.D. of the secondary structure fitting results.

^c Melting temperature determined by the CD spectropolarimeter.

^d Average fluorescence emission wavelength shift.

^e Not determined.

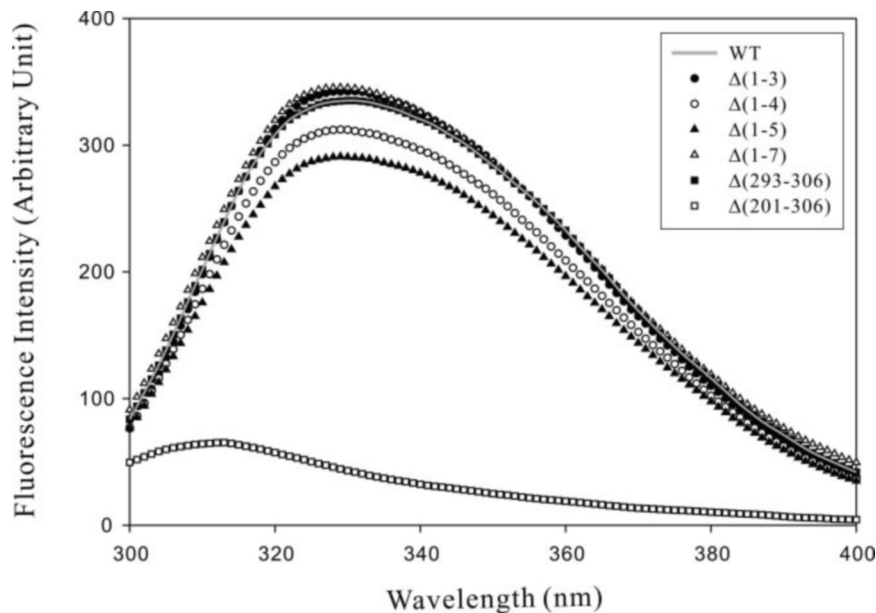


FIG. 3. Fluorescence spectra of the full-length WT and truncated SARS-CoV main proteases. Fluorescence emission spectra of all recombinant proteases were monitored at 8 μ g/ml concentration in 10 mM PBS buffer (pH 7.6) at 25 °C.

umn and washed with the washing buffer (20 mM imidazole, 20 mM Tris-HCl, 300 mM NaCl, and 2 mM BME, pH 7.6). Finally, the protease was eluted with elution buffer (400 mM imidazole, 20 mM Tris-HCl, 300 mM NaCl, and 2 mM BME, pH 7.6). The purified protein was then concentrated at 4 °C using Amicon Ultra-4 centrifugal filter units (Mil-

lipore, Bedford, MA) with molecular mass cutoff at 10 kDa. The purified protein was concentrated to 5–15 mg/ml and then diluted to 0.5–3 mg/ml by 10 mM PBS (containing 10 mM sodium phosphate buffer, 150 mM NaCl, 2 mM BME, pH 7.6), which was used to replace the elution buffer over six concentration-dilution cycles. The sample from the pu-

rification step was separated on a 4–12% gradient sodium dodecyl sulfate polyacrylamide gel to check the homogeneity.

Circular Dichroism (CD) and Fluorescence Analyses—CD experiments were performed in a Jasco J-810 spectropolarimeter (Tokyo, Japan) equipped with a Neslab RTE-111 water-circulated thermal controller. The samples were prepared in 10 mM PBS solution at pH 7.6 with a protein concentration of 0.5 mg/ml. Far-UV CD spectra from 250 to 190 nm were collected using a 0.01-cm path-length cuvette with a 0.1-nm spectral resolution at 25 °C. Ten independent scans were averaged for each sample. All spectra were corrected for buffer contributions and converted to mean residue ellipticity ($[\Theta]$). The $[\Theta]$ at each wavelength was calculated from Equation 1,

$$[\Theta] = \frac{MRW \cdot \theta_\lambda}{10 \cdot l \cdot c} \quad (\text{Eq. 1})$$

where MRW is the mean residue weight (a value of 111.3 was used for the WT), θ_λ is the measured ellipticity in degree at wavelength λ , l is the cuvette path length (0.01 cm), and c is the protein concentration in g/ml. The secondary structure analysis was performed by DICHROWEB (20, 21), which provides an interactive web site server allowing the deconvolution of data from circular dichroism spectroscopy experiments (public-1.cryst.bbk.ac.uk/cdweb/html/). DICHROWEB offers several important types of software, such as CDSSTR (22–24), CONTINLL (25, 26), SELCON3 (27, 28), and K2D (29).

Thermal stability of the WT and various truncated mutants were analyzed with a spectropolarimeter by monitoring the 222-nm circular dichroism at the temperature range between 30 and 90 °C, and the temperature at which half of the protein molecules were unfolded was recorded (T_m).

Fluorescence experiments were performed in a PerkinElmer Life Sciences 50B luminescence spectrometer (Beaconsfield, Buckinghamshire, England). The sample was prepared in 10 mM PBS solution at pH 7.6 with a protein concentration of 8 $\mu\text{g/ml}$. The fluorescence emission spectra from 300 to 400 nm were collected after excitation at 280 nm. Fluorescence spectra of proteins were determined with a 1-cm path quartz cuvette at 25 °C. The spectral bandwidth was 5 nm for excitation and 10 nm for emission. All spectra were corrected for the buffer contribution. The average emission wavelength ($\langle\lambda\rangle$) was calculated from Equation 2 (30),

$$\langle\lambda\rangle = \frac{\sum_{i=\lambda_1}^{\lambda_N} (F_i \cdot \lambda_i)}{\sum_{i=\lambda_1}^{\lambda_N} F_i} \quad (\text{Eq. 2})$$

where F is the fluorescence intensity and λ is the wavelength.

Analytical Ultracentrifugation Analysis—The molar mass and sedimentation coefficient of the proteases were analyzed by a sedimentation velocity experiment. It was performed on a Beckman Optima XL-A analytical Ultracentrifuge (Fullerton, CA). Prior to the experiments, the sample was diluted to various protein concentrations with 10 mM PBS buffer at pH 7.6. Sample (400 μl) and buffer (440 μl) solutions were loaded into the double sector centerpiece separately. All experiments were carried out at 20 °C with an An50 rotor at the speed of 42,000 revolutions/min. The proteins were measured by UV absorbance at 280 nm in a continuous mode with a time interval of 480 s. The recorded scans at different time points were collected and fitted to a continuous size distribution model by the SEDFIT program (31–34). The observed sedimentation profiles of a continuous size distribution ($c(s)$) can be calculated from Equation 3,

$$a(r,t) = \int c(s)L(s,D,r,t)ds + \varepsilon \quad (\text{Eq. 3})$$

where $a(r,t)$ denotes the experimentally observed signal, $L(s,D,r,t)$ denotes the solution of the Lamm equation for a single species (35), and ε is the noise component.

For a precise determination of the monomer-dimer equilibrium of the SARS-CoV main protease, the sedimentation velocity experiment was performed at three different protein concentrations, and all sedimentation data were subjected to the monomer-dimer equilibrium model fitting. The partial specific volume of the protease, solvent density, and viscosity were calculated by the software program SEDNTERP (36). The dissociation constant (K_d) was calculated by the global modeling of the SEDPHAT program (33).

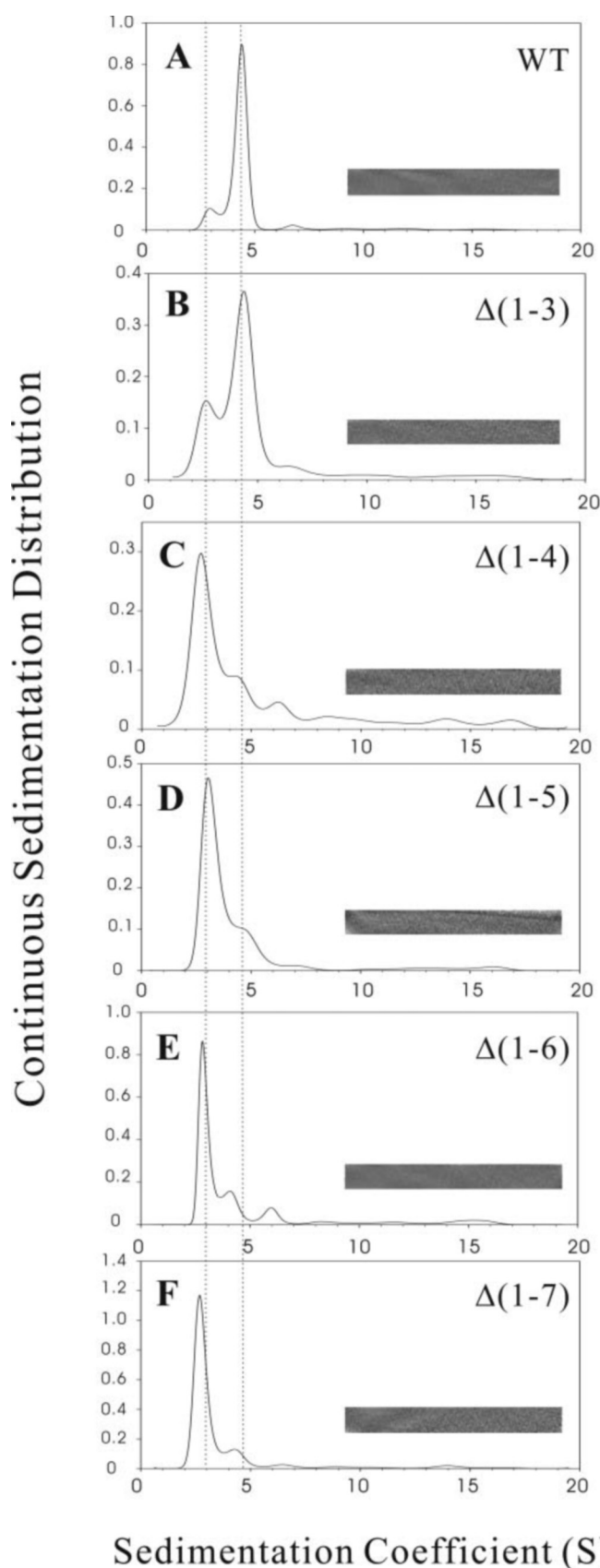


FIG. 4. Continuous sedimentation coefficient distribution of the full-length WT and N-terminally truncated SARS-CoV main protease. The residual bitmaps of various main proteases are shown in the insets. All enzyme preparations used a concentration of 1 mg/ml in 10 mM PBS buffer (pH 7.6) at 20 °C. A, WT; B, $\Delta(1-3)$; C, $\Delta(1-4)$; D, $\Delta(1-5)$; E, $\Delta(1-6)$; F, $\Delta(1-7)$. The left dotted line indicates the monomer, and the right one is the dimer form.

Enzymatic Activity Assay of the SARS-CoV Main Protease Using a Fluorogenic Substrate—The kinetic measurements of the SARS-CoV main protease activity were performed in 10 mM PBS with 2 mM BME at 30 °C. The reaction was initiated by adding 12 μ g of WT and 1.5–2.0 mg of mutants in 1 ml of reaction mixture. Enhanced fluorescence due to cleavage of the internally quenched fluorogenic substrate peptides (*ortho*-aminobenzoic acid-TSAVLQSGFRK-2,4-dinitrophenylamide) by protease was monitored at 420 nm with excitation at 362 nm using a PerkinElmer Life Sciences 50B luminescence spectrometer.

The mixture containing N-terminal peptides (*ortho*-aminobenzoic acid-TSAVLQ) and C-terminal peptides (SQFRK-2,4-dinitrophenylamide) of different concentrations were prepared to monitor the specific fluorogenic intensity. All intensities were corrected for the buffer contributions. The serial intensities at 420 nm were used to construct a standard curve for quantifying the product. In this way, the enzyme activity could be precisely determined.

RESULTS

Expression, Purification, and Characterization of the Recombinant SARS-CoV Main Proteases—The full-length and truncated SARS-CoV main proteases have been successfully expressed in *E. coli* and purified by a single affinity column. All the recombinant SARS-CoV main proteases were found in the soluble fraction of the cell lysate. The expressed SARS-CoV main proteases bound to the nickel column, but other proteins flowed through the column as refuse. SDS-PAGE analysis indicated that the recombinant proteins were almost homogeneous in solution. All purified proteins had M_r in agreement with the theoretical values. After concentration, 5–15 mg/ml recombinant main protease could be obtained from 200 ml of cells. Unfortunately, the other C-terminally truncated mutants were not successfully expressed, probably because of their instabilities.

To determine the secondary structure of the successfully expressed and purified recombinant main proteases, far-UV CD spectra were recorded. The overall CD spectra were shown in Fig. 2. The spectra of all recombinant SARS-CoV main proteases seemed to be similar, except that of $\Delta(201-306)$ as anticipated (Fig. 1A). These results indicated that the proteins have a well defined secondary structure. This is reflected in the secondary structural estimation by the DICHROWEB server (20, 21). The CDSSTR analysis is shown in Table I. The normalized root mean square deviation values of the data fitting for WT and various truncations were all <0.2 and thus showed excellent goodness-of-fit parameters (37). The analysis of the secondary structure of the full-length SARS-CoV main protease is consistent with the data derived from the crystal structure, 1uk3 (Table I). The helical contents of the recombinant WT and crystal structure (1uk3) were 0.21 and 0.22, respectively. All truncations showed similar results with full-length main protease, except that $\Delta(201-306)$ had significant low α -helix content, which was in agreement with the structure in which the whole helical domain III was deleted. The thermal stability of the recombinant SARS-CoV main proteases was also examined (Table I). Among the truncated SARS-CoV main proteases, the C-terminally truncated protease $\Delta(293-306)$ has significantly lower T_m than WT.

The fluorescence emission spectra of the recombinant SARS-CoV main proteases were shown in Fig. 3. Only $\Delta(201-306)$ had significantly low fluorescence intensity. The average emission wavelengths were calculated by the method of Sanchez del Pino and Fersht (30), which accounts for both wavelength shift and the fluorescence intensity attenuation. The average emission wavelength of the full-length SARS-CoV main protease is 342 nm. With the only exception of $\Delta(201-306)$, which has an emission wavelength 11 nm lower than WT, other truncations show minor differences (Table I).

Analytical Ultracentrifugation Analysis—Analytical ultracentrifugation was performed to investigate the association

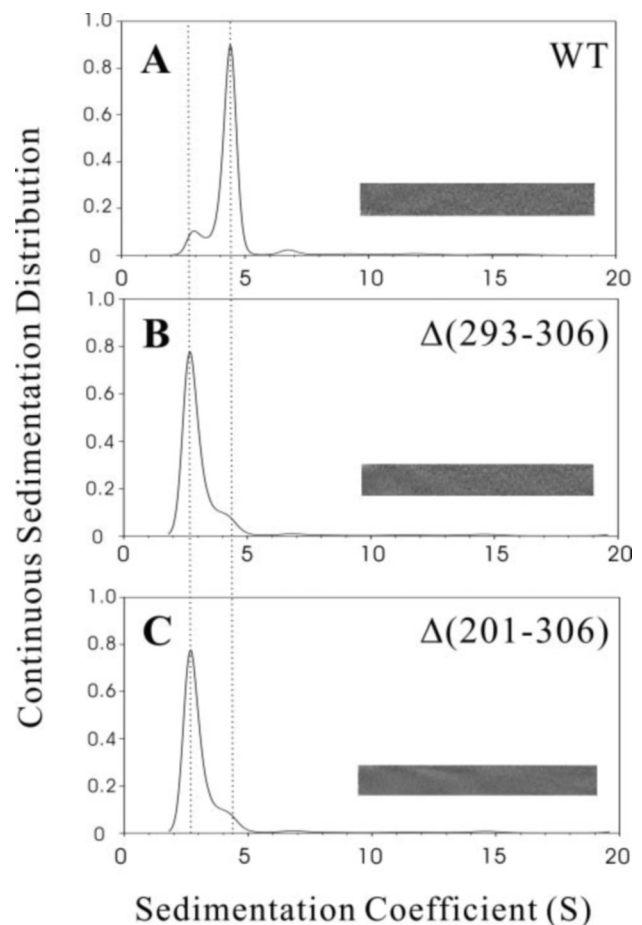


FIG. 5. Continuous sedimentation coefficient distribution of the full-length WT and C-terminally truncated SARS-CoV main protease. The residual bitmaps of various main proteases are shown in the insets. All enzyme preparations used a concentration of 1 mg/ml in 10 mM PBS buffer (pH 7.6) at 20 °C. A, WT; B, $\Delta(293-306)$; C, $\Delta(201-306)$. The left dotted line indicates the monomer, and the right one is the dimer form.

states of WT and truncated SARS-CoV main proteases. This method was successfully used to demonstrate the dimerization of SARS-CoV main protease under various conditions (17). The data were analyzed by continuous size distribution, which implemented a highly reliable model, as indicated by the homogeneous bitmap picture (Figs. 4 and 5, insets). All of these data were derived from an excellent matching curve of the original raw sedimentation data and the randomly distributed residual values (data not shown). WT protease shows a monomer-dimer equilibrium in solution. The sedimentation coefficients of 2.4 S and 4.2 S were monomer and dimer, respectively, corresponding to species with molar mass measurements of 34 and 68 kDa (17). WT and N- and C-terminally truncated proteases also displayed a mixture of monomer and dimer (Figs. 4 and 5). Deletion of three residues from the N terminus showed a similar monomer-dimer distribution with WT. The monomer became the major species when the fourth amino acid residue was deleted from the N terminus (Fig. 4). Further deletion of more residues ($\Delta(1-5)$, $\Delta(1-6)$, and $\Delta(1-7)$) from the N terminus showed a similar pattern with $\Delta(1-4)$. In addition, we have also examined the involvement of domain III in the subunit association. Truncation of the last helix, the $\Delta(293-306)$ mutant, caused the SARS-CoV main protease to become a monomer (Fig. 5). These results clearly indicated that residues 4 and 293–306 are critically involved in stabilizing the dimer structure.

The influence of individual residues or domains in the sub-

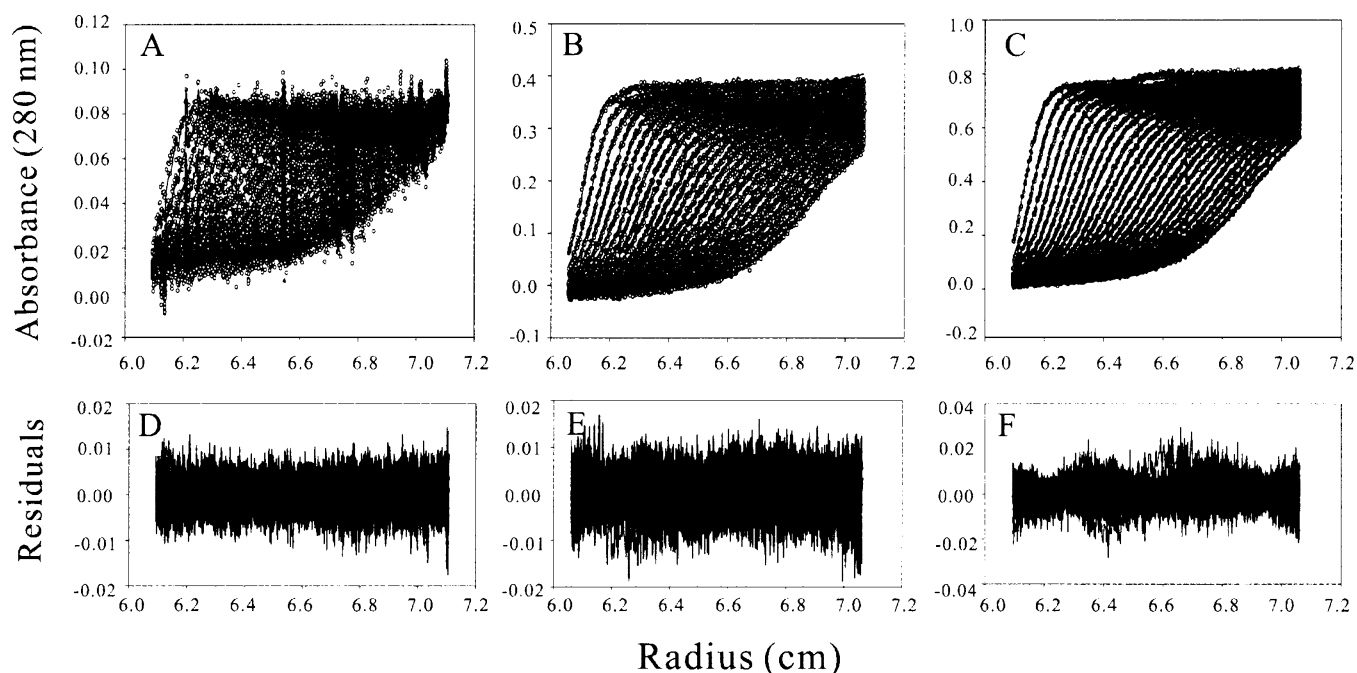


FIG. 6. Global analysis of the sedimentation velocity data of full-length WT SARS-CoV main protease at three protein concentrations. Sedimentation was performed at 20 °C with an An50 rotor and at rotor speeds of 42,000 revolutions/min. A–C, the concentrations of the protein were 0.1 mg/ml, 0.5 mg/ml, and 1 mg/ml, respectively. The sedimentation profiles are from the absorbance optical system at a wavelength of 280 nm. The symbols are the raw sedimentation data, and the lines are the theoretical fitted data to the Lamm equation implemented in the software SEDPHAT. D–F, the randomly distributed residuals of the fitting model from the upper panel provide a credible analysis result for a dissociation constant (K_d) of the monomer-dimer equilibrium.

unit interaction was further quantified by comparing the monomer-dimer dissociation constants. As mentioned above, the global analysis was employed to determine the K_d value of WT and various truncated mutant proteases (Fig. 6). The K_d value for monomer-dimer equilibrium of WT was measured to be 0.28 μM (Table II). Sequential deletion of residues from the N terminus increased the K_d value. The C-terminally truncated proteases have much higher K_d values than the others.

Kinetic Properties of the Protease—The deletion mutant of the related transmissible gastroenteritis virus main protease that lacks residues 1–5 is almost enzymatically inactive (13). The crystal structure of the SARS-CoV main protease reveals that the N-terminal residues 1–7 from subunit A are directly inserted into the active site of subunit B (15). We determined the enzyme activity of the full-length, sequential N- and C-terminally truncated proteases. The enzyme activity was measured by peptide cleavage assay (17). The internally quenched fluorescent substrate is cleaved specifically at the Q–S peptide bond (38). The apparent K_m value of WT measured was 17 ± 1 μM , and the apparent k_{cat} value was 198 ± 22 s^{-1} . The mutant main protease $\Delta(1-3)$ still possesses 76% of enzyme activity as compared with WT (Table II). However, the enzyme activity of other truncated mutants was decreased to only 0.2–1.3% of WT activity.

DISCUSSION

The biophysical analyses of the SARS-CoV main protease and its truncated mutants performed here allowed a detailed structural characterization of the mutants compared with the full-length protease. In our present data, except the domain III-truncated protease, $\Delta(201-306)$, which has lower mean residue ellipticity and fluorescence intensity, the CD and fluorescence emission spectra of other truncated proteases were similar to the full-length protease. With the secondary and tertiary structures as anticipated, we then tried to study the structure and function relationships of the N- and C-terminal regions of the SARS-CoV main protease, especially on the correlation

TABLE II
Quaternary structure and enzymatic activity of the full-length and truncated SARS-CoV main proteases

Protease	K_d^a	Fold	RMSD ^b	Relative enzyme activity ^c
	μM			%
WT	0.28 ± 0.01	1	0.003–0.005	100
$\Delta(1-3)$	3.4 ± 0.1	12	0.007–0.010	76
$\Delta(1-4)$	57.5 ± 6.3	205	0.007–0.017	1.3
$\Delta(1-5)$	35.4 ± 2.2	126	0.007–0.013	0.3
$\Delta(1-6)$	151 ± 0.4	539	0.003–0.007	0.3
$\Delta(1-7)$	357 ± 19	1,275	0.006–0.010	0.4
$\Delta(293-306)$	163 ± 3.2	582	0.005–0.009	1.1
$\Delta(201-306)$	7870 ± 302	28,107	0.005–0.008	0.2

^a Dissociation constant of the dimeric SARS-CoV main proteases.

^b Root mean square deviation of the fitted sedimentation velocity data.

^c Enzyme activity assay performed with the internally quenched substrate *ortho*-aminobenzoic acid-TSAVLQSGFRK-2,4-dinitrophenylamide. The enzyme activities of the various truncated mutants were too small to allow a precise determination of the kinetic parameters. Only the relative enzyme activity was reported.

between dimerization and enzyme activity. We used the CASTp program (39) to analyze the pockets and cavities of the enzyme. Only three of the cavities with mouths were big enough to be significant (Fig. 1C). One of these pockets was located at the interface between domains II and III with a surface area of 2.82 nm^2 and has no contact with protomer B (Fig. 1C, yellow pocket). The two largest pockets were located at the subunit interfacial region. The N-terminal finger and C-terminal tip were sited at the pocket with a solvent-accessible surface area of 18.41 nm^2 and a volume of 2.173 nm^3 (Fig. 1C, purple pocket). The amino acid residue contacts between subunits A and B were analyzed with the Contacts of Structural Units software (40). Phe-3 in subunit A contacts with subunit B with only one destabilizing hydrophobic-hydrophilic contact. However, Arg-4 (A) extends deeply into subunit B (Figs. 1 and 7) involving six destabilizing contacts. It is then clear that re-

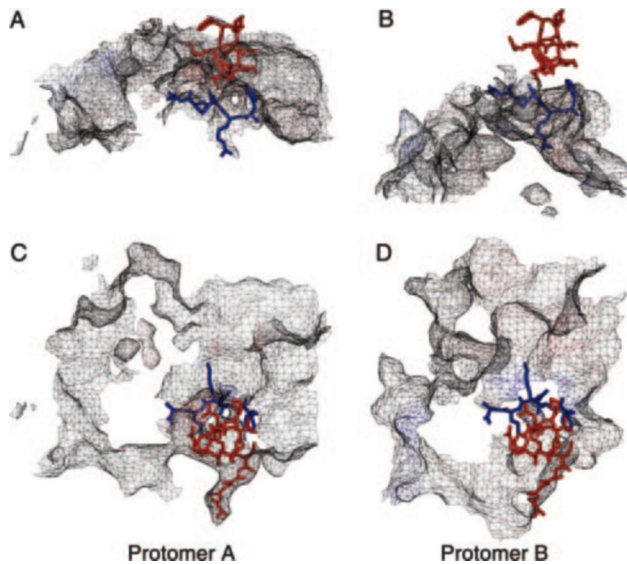


FIG. 7. Interfacial regions of the dimeric SARS-CoV main protease. The interfacial contact surface regions within 1 nm for both protomers A and B are shown in mesh model form. The same surface areas were turned 90° around the x-axis to give C and D, respectively, from A and B. The C-terminal helical region (red) and the N-terminal finger (blue) are shown in stick mode. This figure was generated with Spock software.

removal of Arg-4 has great impact on the structure of the enzyme molecule.

All of the coronavirus conserved an extra large C-terminal α -helical domain that is not found in other RNA virus 3C-like proteases (13, 15). The precise biological role of the α -helical domain is very interesting but still not completely understood. Our experimental data indicate that the C-terminal domain plays an important role in dimerization and enzyme activity. These results are consistent with the previous report that loss of the extra domain III of the SARS-CoV main protease induced monomer formation and loss of enzyme activity. A further step forward, we have narrowed down the region to a single α -helix.

A similar thermal stability curve was observed for most truncations, except for the last α -helically truncated protease, $\Delta(293-306)$, which showed significantly lower T_m than WT (Table I). Deletion of the whole domain III, on the other hand, restores the stability of the molecule. These results imply the functional role of domain III of the protease. The deleted last helical segment is located at a large pocket between domains II and III (Fig. 1C, yellow pocket). This helix is essential for protein stability of the whole molecule. Without this helix, the remaining domain III becomes a burden for domains I and II that renders the protease unstable.

With a complete domain III, the full-length SARS main protease was in monomer-dimer equilibrium with dimer as the major form, even in very low protein concentration (0.1 mg/ml) (17). This result is different from some of the recent reports (18, 41), which showed a major monomer in a protein concentration of <0.2 mg/ml. This discrepancy is due to technical differences used in characterizing the quaternary structure of the protein. It is extremely important to obtain an unequivocal answer to this question, because the dissociated monomer might be enzymatically inactive. To study the protein self-association, we used the rebirth state-of-the-art analytical ultracentrifugation technique. The K_d value determined by analytical ultracentrifugation (0.28 μ M), however, is \sim 360 and 810 times smaller than that previously estimated from the analytical gel filtration experiment and isothermal titration calorimeter, respectively (19, 42). Our K_d value for the monomer-dimer equilibrium of SARS-CoV main protease was obtained from global

analysis of three different protein concentrations. Under the stringent conditions in analytical ultracentrifugation, we believe that the K_d obtained is more reliable (33).

Our data showed that the N-finger of SARS-CoV main protease was indispensable for proteolytic activity, which was consistent with previous findings (19). The contribution of the N-finger in dimerization was clearly demonstrated by analytical ultracentrifugation analysis (Fig. 4). As shown in Fig. 7, the N-finger from protomer A is completely buried into protomer B (Fig. 7, A and B). We have further narrowed down to a single residue, Arg-4, which plays a pivotal role in building the molecular interaction (Fig. 4). The correlation between enzymatic activity and dimerization is also demonstrated. The fourth residue plays a critical role in dimerization of SARS-CoV main protease, which is essential for the enzymatic activity.

In conclusion, both the N- and C-terminal regions play pivotal roles in controlling the dimerization and activity of the enzyme. Our study provides fundamental information for the novel design of inhibitors of the SARS-CoV main protease activity by disrupting its dimerization interface.

REFERENCES

- Chan, H. L., Tsui, S. K., and Sung, J. J. (2003) *Trends Mol. Med.* **9**, 323–325
- Leng, Q., and Bentwich, Z. (2003) *N. Engl. J. Med.* **349**, 709
- Kuiken, T., Fouchier, R. A., Schutten, M., Rimmelzwaan, G. F., van Amerongen, G., van Riel, D., Laman, J. D., de Jong, T., van Doornum, G., Lim, W., Ling, A. E., Chan, P. K., Tam, J. S., Zambon, M. C., Gopal, R., Drosten, C., van der Werf, S., Escriou, N., Manuguerra, J. C., Stöhr, K., Peiris, J. S., and Osterhaus, A. D. (2003) *Lancet* **362**, 263–270
- 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., Khattri, 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., Drebot, M., Fernando, L., Flick, R., Garbutt, M., Gray, M., Grolla, A., Jones, S., Feldmann, H., Meyers, A., Kabani, A., Li, Y., Normand, S., Stroher, U., Tipples, G. A., Tyler, S., Vogrig, R., Ward, D., Watson, B., Brunham, R. C., Krajdenc, M., Petric, M., Skowronski, D. M., Upton, C., and Roper, R. L. (2003) *Science* **300**, 1399–1404
- Tanner, J. A., Watt, R. M., Chai, Y. B., Lu, L. Y., Lin, M. C., Peiris, J. S., Poon, L. L., Kung, H. F., and Huang, J. D. (2003) *J. Biol. Chem.* **278**, 39578–39582
- Stavrinos, J., and Guttman, D. S. (2004) *J. Virol.* **78**, 76–82
- Eickmann, M., Becker, S., Klenk, H. D., Doerr, H. W., Stadler, K., Censini, S., Guidotti, S., Masignani, V., Scarselli, M., Mora, M., Donati, C., Han, J. H., Song, H. C., Abignani, S., Covacci, A., and Rappuoli, R. (2003) *Science* **302**, 1504–1505
- Holmes, K. V. (2003) *J. Clin. Investig.* **111**, 1605–1609
- Kesel, A. J. (2003) *Bioorg. Med. Chem.* **11**, 4599–4613
- Thiel, V., Ivanov, K. A., Putics, A., Hertzog, T., Schelle, B., Bayer, S., Weissbrich, B., Snijder, E. J., Rabenau, H., Doerr, H. W., Gorbalenya, A. E., and Ziebuhr, J. (2003) *J. Gen. Virol.* **84**, 2305–2315
- Tong, L. (2002) *Chem. Rev.* **102**, 4609–4626
- 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
- Yang, H., Yang, M., Ding, Y., Liu, Y., Lou, Z., Zhou, Z., Sun, L., Mo, L., Ye, S., Pang, H., Gao, G. F., Anand, K., Bartlam, M., Hilgenfeld, R., and Rao, Z. (2003) *Proc. Natl. Acad. Sci. U. S. A.* **100**, 13190–13195
- Bacha, U., Barrila, J., Velazquez-Campoy, A., Leavitt, S. A., and Freire, E. (2004) *Biochemistry* **43**, 4906–4912
- Chou, C. Y., Chang, H. C., Hsu, W. C., Lin, T. Z., Lin, C. H., and Chang, G. G. (2004) *Biochemistry* **43**, 14958–14970
- Shi, J., Wei, Z., and Song, J. (2004) *J. Biol. Chem.* **279**, 24765–24773
- Chen, S., Chen, L., Tan, J., Chen, J., Du, L., Sun, T., Shen, J., Chen, K., Jiang, H., and Shen, X. (2005) *J. Biol. Chem.* **280**, 164–173
- Lobley, A., Whitmore, L., and Wallace, B. A. (2002) *Bioinformatics* **18**, 211–212
- Whitmore, L., and Wallace, B. A. (2004) *Nucleic Acids Res.* **32**, W668–673
- Compton, L. A., and Johnson, W. C., Jr. (1986) *Anal. Biochem.* **155**, 155–167
- Manavalan, P., and Johnson, W. C., Jr. (1987) *Anal. Biochem.* **167**, 76–85
- Sreerama, N., and Woody, R. W. (2000) *Anal. Biochem.* **287**, 252–260
- Provencher, S. W., and Glockner, J. (1981) *Biochemistry* **20**, 33–37
- Van Stokkum, I. H., Spoelder, H. J., Bloemendal, M., van Grondelle, R., and Groen, F. C. (1990) *Anal. Biochem.* **191**, 110–118
- Sreerama, N., and Woody, R. W. (1993) *Anal. Biochem.* **209**, 32–44

28. Sreerama, N., Venyaminov, S. Y., and Woody, R. W. (1999) *Protein Sci.* **8**, 370–380
29. Andrade, M. A., Chacon, P., Merelo, J. J., and Moran, F. (1993) *Protein Eng.* **6**, 383–390
30. Sanchez del Pino, M. M., and Fersht, A. R. (1997) *Biochemistry* **36**, 5560–5565
31. Schuck, P. (2000) *Biophys. J.* **78**, 1606–1619
32. Lebowitz, J., Lewis, M. S., and Schuck, P. (2002) *Protein Sci.* **11**, 2067–2079
33. Schuck, P. (2003) *Anal. Biochem.* **320**, 104–124
34. Schuck, P., Perugini, M. A., Gonzales, N. R., Howlett, G. J., and Schubert, D. (2002) *Biophys. J.* **82**, 1096–1111
35. Lamm, O. (1929) *Ark. Mat. Astr. Fys.* **21B**, 1–4
36. Laue, T. M., Shah, B. D., Ridgeway, T. M., and Pelleter, S. L. (1992) *Analytical Ultracentrifugation in Biochemistry and Polymer Science*, The Royal Society of Chemistry, Cambridge, UK
37. Mao, D., Wachter, E., and Wallace, B. A. (1982) *Biochemistry* **21**, 4960–4968
38. Hegyi, A., and Ziebuhr, J. (2002) *J. Gen. Virol.* **83**, 595–599
39. Liang, J., Edelsbrunner, H., and Woodward, C. (1998) *Protein Sci.* **7**, 1884–1897
40. Sobolev, V., Sorokine, A., Prilusky, J., Abola, E. E., and Edelman, M. (1999) *Bioinformatics* **15**, 327–332
41. Fan, K., Wei, P., Feng, Q., Chen, S., Huang, C., Ma, L., Lai, B., Pei, J., Liu, Y., Chen, J., and Lai, L. (2004) *J. Biol. Chem.* **279**, 1637–1642
42. Kuo, C. J., Chi, Y. H., Hsu, J. T., and Liang, P. H. (2004) *Biochem. Biophys. Res. Commun.* **318**, 862–867