JOURNAL OF VIROLOGY, Aug. 2006, p. 7902–7908 Vol. 80, No. 16
0022-538X/06/$08.00+0 doi:10.1128/JVI.00483-06

Dodecamer Structure of Severe Acute Respiratory Syndrome Coronavirus Nonstructural Protein nsp10

Dan Su,[1,2]† Zhiyong Lou,[1,2]† Fei Sun,[1,2] Yujia Zhai,[1,2] Haitao Yang,[1,2] Rongguang Zhang,[2,3] Andrzej Joachimiak,[3] Xuejun C. Zhang,[2,4] Mark Bartlam,[1,2] and Zihe Rao[1,2]*

"Tsinghua-IBP Joint Research Group for Structural Biology," Tsinghua University, Beijing 100084, China[1]; National Laboratory of Biomacromolecules, Institute of Biophysics (IBP), Chinese Academy of Sciences, Beijing 100101, China[2]; Biosciences Division, Argonne National Laboratory, Argonne, Illinois 60439[3]; and Crystallography Research Program, Oklahoma Medical Research Foundation, Oklahoma City, Oklahoma 73104[4]

Received 8 March 2006/Accepted 9 May 2006

The severe acute respiratory syndrome coronavirus (SARS-CoV) nonstructural proteins nsp1 to nsp16 have been implicated by genetic analysis in the assembly of a functional replication/transcription complex. We report the crystal structure of nsp10 from SARS-CoV at 2.1-Å resolution. The nsp10 structure has a novel fold, and 12 identical subunits assemble to form a unique spherical dodecameric architecture. Two zinc fingers have been identified from the nsp10 monomer structure with the sequence motifs C-$(X)_2$-C-$(X)_5$-H-$(X)_6$-C and C-$(X)_2$-C-$(X)_7$-C-(X)-C. The nsp10 crystal structure is the first of a new class of zinc finger protein three-dimensional structures to be revealed experimentally. The zinc finger sequence motifs are conserved among all three coronavirus antigenic groups, implicating an essential function for nsp10 in all coronaviruses. Based on the structure, we propose that nsp10 is a transcription factor for coronavirus replication/transcription.

Coronaviruses (CoVs) are positive-stranded RNA viruses known to cause highly prevalent respiratory and enteric diseases in humans and animals. About two-thirds of their genomes encode the viral replicase that mediates viral RNA synthesis (19). The replicase gene is comprised of two open reading frames at the 5′ end of the genome, termed ORF1a and ORF1b (26). The upstream ORF1a encodes polyprotein pp1a (450 to 500 kDa), while ORF1a and ORF1b together encode pp1ab (750 to 800 kDa). Expression of the ORF1b-encoded half of pp1ab requires a −1 ribosomal frameshift during translation. The polyproteins pp1a and pp1ab undergo extensive proteolytic processing by viral proteases to produce 16 functional subunits known as nonstructural proteins, which then assemble to form the replicase complex required for viral replication and transcription.

Prior to the global severe acute respiratory syndrome (SARS) outbreak in 2003, scant attention was paid to coronaviruses by researchers, as this genus of viruses causes severe diseases predominantly in animals and only comparatively mild diseases in humans. In the wake of the SARS outbreak, greater attention has been focused on the replicase proteins with a view to understanding the replication/transcription machinery and to identify new therapeutic targets. To date, the three-dimensional structures of a series of nsp proteins have been reported (1). nsp5, also called the main protease or 3C-like protease, was the first SARS protein structure determined in 2003 (24) and has since been the focus of concerted efforts for the design of antiviral inhibitors. Last year, a broad-spectrum inhibitor was reported with efficient in vitro inactivation of multiple coronavirus main proteases, potent antiviral activity, and extremely low cellular toxicity (23). The structure of nsp9 was determined in 2004, and it was found to be a single-stranded RNA binding protein (4, 17). More recently, the complex structure between nsp7 and nsp8 revealed a hexadecameric assembly that should constitute a processivity factor for nsp12, an RNA-dependent RNA polymerase (25). Also determined recently was the ADP ribose 1-phosphatase domain of nsp3 (14).

nsp10 and nsp11 immediately precede nsp12 in the pp1ab sequence and have been implicated in RNA synthesis, while the mechanism remains to be illustrated through both structural and functional studies. To improve our understanding of functions of CoV nsp proteins, we carried out systematic structural studies on this group of proteins. We report the crystal structure of nsp10. This protein appears to have a novel fold, featuring two zinc fingers with C-$(X)_2$-C-$(X)_5$-H-$(X)_6$-C and C-$(X)_2$-C-$(X)_7$-C-(X)-C motifs. Furthermore, 12 identical subunits assemble into a novel spherical dodecameric architecture which is proposed to be a functional form of nsp10.

MATERIALS AND METHODS

Protein expression, purification, and characterization. The coding sequence for SARS-CoV nsp10-nsp11 was amplified by PCR from the SARS-CoV BJ01 strain (corresponding to Ala4231 to Val4382 of the ORF1a polyprotein) and inserted into pGEX-6p-1 plasmid DNA (Amersham Biosciences) using BamHI and XhoI sites. The forward and reverse PCR primers used for amplification were F′-nsp10-nsp11 (5′-CGGGATCCGCTGGAAATGCTACAGAAGT-3′) and R′-nsp10-nsp11 (5′-CGCTCGAGTTACACCGCAAACCCGTTTA-3′), respectively. The protein was expressed in *Escherichia coli* strain BL21(DE3) as a glutathione *S*-transferase (GST) fusion protein and purified using glutathione affinity chromatography. The GST tag was removed by PreScission protease (Amersham Biosciences), leading to five additional residues (GPLGS) at the N terminus. The protein was further purified by a G25 desalting column in 2 mM dithiothreitol (DTT) and concentrated to 10 mg ml^{-1} in preparation for crystallization screening.

* Corresponding author. Mailing address: Laboratory of Structural Biology, Life Sciences Building, Tsinghua University, Beijing 100084, China. Phone: 86 10 62771493. Fax: 86 10 62773145. E-mail: raozh@xtal.tsinghua.edu.cn.

† These authors made equal contributions.

TABLE 1. Data collection and refinement statistics

Statistical element	nsp10	nsp10 in the absence of zinc
Data collection statistics		
Cell parameters	a = 159.1 Å, b = 321.8 Å, c = 161.9 Å, $\alpha = \beta = \gamma = 90°$	a = 159.7 Å, b = 322.7 Å, c = 162.2 Å, $\alpha = \beta = \gamma = 90°$
Space group	$C222_1$	$C222_1$
Resolution (Å)	50 (2.2)[c]–2.1	50 (2.8)–2.7
No. of all reflections	1,545,407	845,526
No. of unique reflections	276,259	114,838
Completeness (%)	99.5 (99.4)	100.0 (99.9)
Average I/σ(I)	10.5 (4.2)	6.6 (3.8)
R_{merge}[a] (%)	6.1 (36.3)	10.4 (40.1)
Refinement statistics		
No. of reflections used (σ(F) > 0)	468,890	249,037
R_{work}[b] (%)	21.9	21.6
R_{free}[b] (%)	24.7	25.1
RMSD bond distance (Å)	0.013	0.008
RMSD bond angle (°)	1.67	1.79
Average B value (Å^2)	31.8	34.5
Ramachandran plot (excluding Pro and Gly)		
Res. in most favored regions	2,039 (87.6%)	1,932 (83%)
Res. in additionally allowed regions	278 (12.0%)	375 (16.1%)
Res. in generously allowed regions	0	18 (0.8%)

[a] $R_{merge} = \Sigma_h\Sigma_i I_{ih} - \langle I_h\rangle/\Sigma_h\Sigma_I \langle I_h\rangle$, where $\langle I_h\rangle$ is the mean of the observations I_{ih} of reflection h.
[b] $R_{work} = \Sigma(\|F_p(\text{observed}) - F_p(\text{calculated})\|)/\Sigma F_p(\text{observed})$. R_{free} indicates the R factor for a selected subset (5%) of the reflections that was not included in prior refinement calculations.
[c] Numbers in parentheses are corresponding values for the highest resolution shell.

Crystallization. SARS nsp10-nsp11 crystals were grown at 291 K by the hanging drop vapor diffusion method from a sodium chloride system in desalting buffer supplemented with 65 mM zinc chloride. The optimum brick-like crystals were obtained using a reservoir solution of 1.8 M NaCl, 0.1 M NaH_2PO_4, 0.1 M KH_2PO_4, 0.1 M DTT, and 2-(*N*-morpholino)ethanesulfonic acid (MES; pH 6.5). A single crystal was transferred to the reservoir solution (supplemented with 5 M sodium formate) for 1 h prior to submersion in a cold nitrogen stream.

Data collection and processing. The "frozen" nsp10-nsp11 crystal diffracted at up to a 2.1-Å resolution. Single-wavelength anomalous dispersion (SAD) data using the bound zinc were collected for the native nsp10-nsp11 crystal at 100 K using an SBC2 (3,000 by 3,000) CCD detector on beamline BL19-ID at the Advanced Photon Source (Argonne National Laboratory). The intensity data were processed and scaled using the software HKL2000 (11). Data collection statistics are summarized in Table 1.

Phasing, model building, and refinement. Forty-eight heavy atoms were located from a search with the program SHELX (16, 21) that were consistent with the solution from a self-rotation function search using the program MOLREP (22). Initial phases were calculated using the single-wavelength anomalous dispersion (SAD) method with the program SOLVE (18). Twenty-two sites were found to have occupancy greater than 50%, and the remaining sites had occupancy greater than 30%. Density modification at 2.8-Å resolution by solvent flipping was performed with the program CNS (3).

Initial tracing of the monomer polypeptide chain was performed manually using the program O (7), and a starting dodecameric model was generated using noncrystallography symmetry (NCS) relations generated from CNS. Initial refinement was performed using simulated annealing in CNS with very tight NCS restraints. During the later stages of positional refinement, restraints were relaxed and a bulk solvent correction was applied under the guidance of R_{free}. Model geometry was verified using the program PROCHECK (9). Solvent molecules were located from stereochemically reasonable peaks in the σ^A-weighted $F_o - F_c$ difference electron density map, where F_o is the observed structure factor amplitude and F_c is the calculated structure factor amplitude. Final refinement statistics are shown in Table 1.

Protein structure accession numbers. Coordinates and structure factors for the SARS-CoV nsp10 crystal structures have been deposited in the RCSB Protein Data Bank (PDB) under accession numbers 2G9T (for the 2.1-Å structure) and 2GA6 (for the 2.7-Å structure).

RESULTS AND DISCUSSION

Overall structure. The cDNA coding for SARS-CoV nsp10-nsp11 was amplified using PCR. The coded protein consists of amino acid residues 4231 to 4382 of pp1a, which are renumbered 1 to 152 hereafter for convenience, and includes both nsp10 and nsp11 regions. The recombinant protein was expressed in *E. coli* as a GST fusion protein, purified, separated from the GST tag, and crystallized in both the presence and absence of additional zinc ion. Crystals grown under these two conditions have the same morphology, and their crystal cell parameters are similar to each other. Unless otherwise mentioned, the higher resolution crystal (with supplementary zinc salt) will be used in the following structural description. The crystal form belongs to an orthogonal space group with 24 copies of nsp10-nsp11 per crystallography asymmetric unit and a V_M of 3.1 Å^3 Da^{-1}, where V_M is the Matthews' coefficient. The crystal structure was determined using SAD phasing from the bound zinc ions. The structure model was refined at 2.1-Å resolution to a final R_{work} of 21.9% and R_{free} of 24.7%. Refinement statistics are detailed in Table 1. All 24 crystallographically independent copies of nsp10-nsp11 have very similar structures, with all pair-wise Cα root mean square deviations (RMSD) being less than 0.5 Å. Residues before Ala9 (including five leading residues left from the tag) and Pro86-Gly88 and after Ser129 could not be traced due to lack of interpretable electron density. Since the latter region includes the entire nsp11 peptide, the crystallized protein is nsp10 per se; it is called such hereafter. The absence of Cys130 and subsequent residues is a result of incidental cleavage, as indicated by (i) the reduced molecular weight band from so-

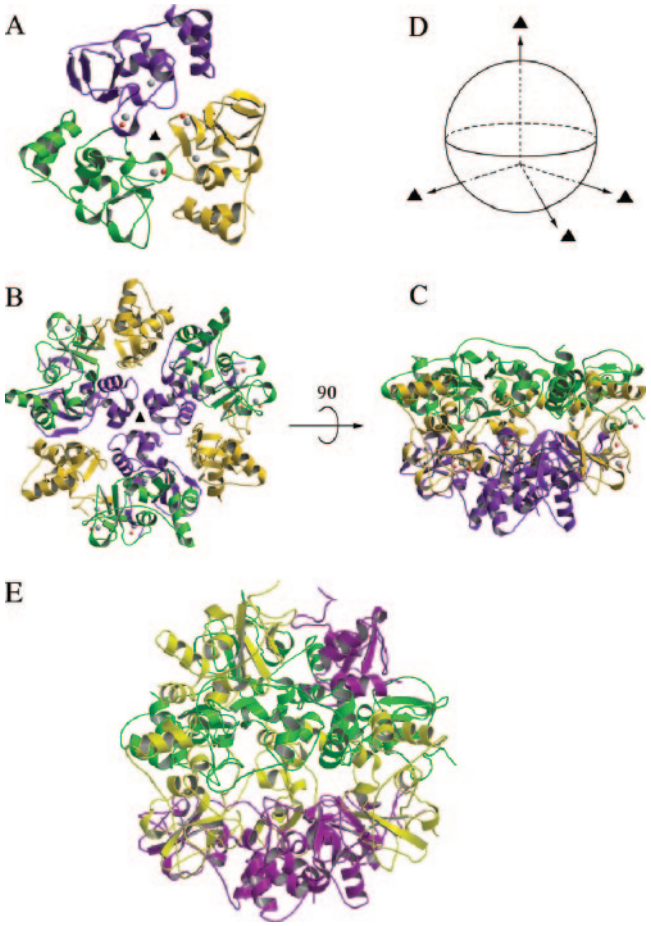

FIG. 1. Ribbon representation of the nsp10 crystal structure. A. One nsp10 trimer is viewed along the noncrystallography threefold axis from the outside of a dodecamer. The three protomers are colored in magenta, gold, and green. B. The remaining three trimers of the dodecamer are shown in the same orientation as that for panel A using the same color scheme. C. The same as panel B but rotated by 90°. D. The relationship between threefold axes. The twist angle between pairs of threefold axes is approximately 108°. E. The complete dodecamer structure of SARS-CoV nsp10. The three protomers in a trimer are colored in magenta, gold, and green. This figure was prepared with the programs Molscript (8), Bobscript (5), and Raster3D (10).

dium dodecyl sulfate polyacrylamide gel electrophoresis analysis compared with that of a freshly prepared protein sample and (ii) matrix-assisted laser desorption ionization–time of flight mass spectrometry analysis (data not shown). The cause of the loss of Cys130 and its C-terminal residues remains unclear. However, analysis of the crystal packing suggests that the C-terminal peptide deletion is required for the formation of the crystal lattice.

The 24 independent copies of nsp10 assemble into two identical spherical dodecamers related by a local twofold symmetry. The novel dodecameric sphere possesses a tetrahedral symmetry and can be viewed as the assembly of four nsp10 trimers (Fig. 1). The dodecamer is hollowed in the center, with an outer radius of 42 Å and an inner radius of 18 Å (Fig. 2). Two zinc binding sites were identified in the three-dimensional structure of each nsp10 monomer. The first one is located in the N-terminal region and on the inner surface of the dodecamer, and the second one is located at the C terminus on the outer surface. The missing C tail of nsp10-nsp11 peptide would also be located on the outer surface of the dodecamer. Both inner and outer surfaces of the dodecamer have a dominantly positive electrostatic potential (Fig. 2), although the calculated pI of nsp10 is 5.8.

The monomer fold. The full-length nsp10 polypeptide has a mixed α/β fold comprised of five α-helices (α1 to α5), one 3_{10}-helix, and three β-strands (β1 to β3) (Fig. 3). The monomer peptide model consists of residues Ala9 to Ser129 and is continuous, except that the region Pro86-Gly88 is missing in the α4-β3 connecting loop. Nevertheless, in the 2.7-Å structure which was solved using crystals grown in the absence of additional zinc, we were able to build the Pro86-Gly88 loop region for some nsp10 monomers, suggesting that this region is flexible in general. The central core of the nsp10 monomer is an antiparallel β-sheet formed by strands β1 (residues 55 and 56), β2 (65 to 69), and β3 (96 to 100). The central β-sheet is flanked on one side by helices α3 (residues 70 to 73) and α4 (75 to 79), while helices α1 (residues 10 to 18), α2 (23 to 32) at the N terminus, helix α5 (107 to 113), and the extended C-terminal coil shy away from the central core. Residues on the α4 helix and α4-β3 loop constitute the N-terminal zinc binding site, and the C-terminal coil contributes to the C-terminal zinc binding site. A DALI (http://www.ebi.ac.uk/dali/) search indicated no similar match to the nsp10 monomer in the current Protein Data Bank (PDB), suggesting a novel fold for nsp10.

nsp10 is cysteine-rich, with 13 cysteines in its sequence, namely Cys17, Cys41, Cys46, Cys73, Cys74, Cys77, Cys79, Cys90, Cys103, Cys117, Cys120, Cys128, and Cys130. Of these, Cys74, Cys77, and Cys90 form the N-terminal zinc binding site, while Cys117, Cys120, and Cys128 chelate the C-terminal zinc ion (Fig. 4). Cys130 is missing from the refined model due to the loss of the C-terminal peptide. Although Cys17-Cys46 and Cys41-Cys73 pairs are close in the three-dimensional structure, with Sγ-Sγ distances of 5.5 Å and 4.7 Å, respectively, they do not form disulfide linkages. This observation is consistent with the presence of high concentrations of reducing agents in the crystallization reservoir.

The trimeric assembly. Three monomers are related by a noncrystallographic threefold symmetry to form a trimeric subunit (Fig. 1A). The trimer is stabilized largely by hydrophobic interactions. The buried solvent-accessible surface (SAS) of an nsp10 monomer within a trimer is about 620 $Å^2$ (or about 9% of the total SAS of an nsp10 monomer), with 65% contributed from hydrophobic atoms. Residue Val21 in the α1-α2 loop of one monomer forms hydrophobic interactions with residues Val57 and Thr58 in strand β1 of an adjacent monomer; residues Thr115 and Thr118 of one monomer also form hydrophobic interactions with Thr118 and Val119 of an adjacent monomer. Additional hydrogen bonds are formed between the Nζ atom of Lys25 in one molecule and the $O_{\varepsilon 2}$ atom of Glu60 in an adjacent molecule, with a distance of 2.6 Å, and between the main chain O atom of Pro84 in one molecule and the Nξ atom of Lys95 in an adjacent molecule, with a distance of 3.3 Å. Furthermore, the three monomers are oriented such that their C-terminal zinc fingers are clustered around the threefold axis, with their zinc ions separated by approximately 14 Å.

Trimer-trimer interactions and the dodecamer architecture. The assembly of the 24 copies of nsp10 into two identical

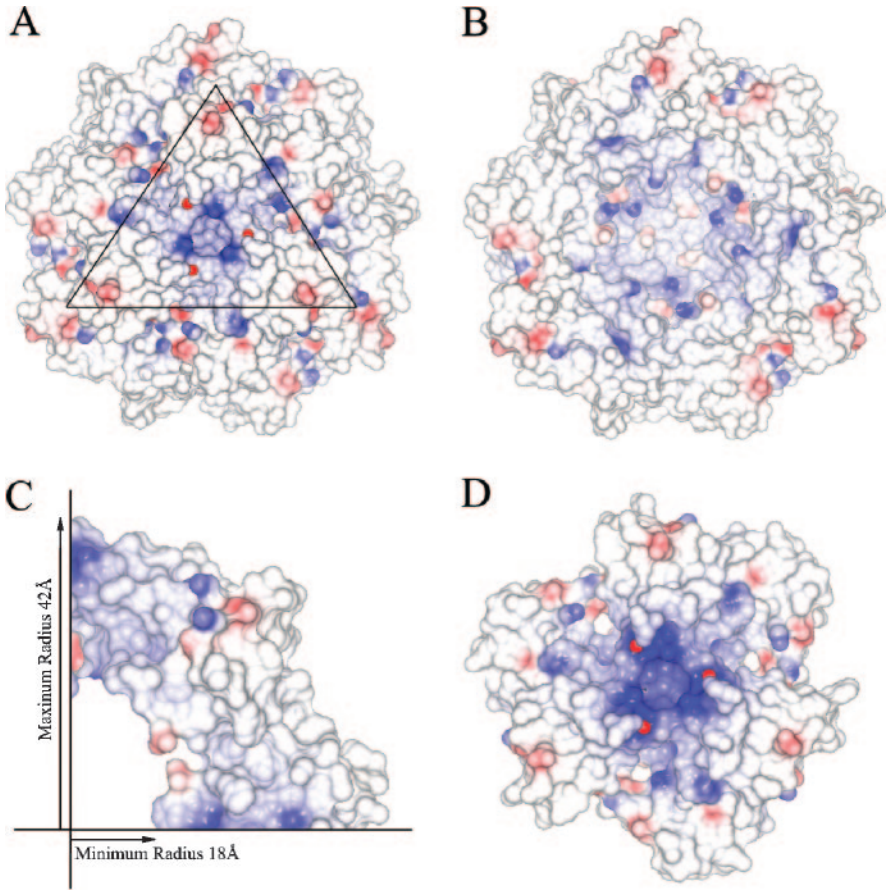

FIG. 2. Electrostatic potential of the nsp10 crystal structure. A. The electrostatic potential of a complete dodecamer is mapped on its outer molecular surface. Negatively charged regions are colored in red, positively charged regions are colored in blue, and neutral regions are colored in gray. The C-terminal-bound zinc ions, which are located on the outer surface, are depicted as red spheres. The position of the front trimer is highlighted by a triangle. B. The inner surface electrostatic potential of the dodecamer is shown in the same orientation as that of panel A. C. A cross-section of the nsp10 dodecamer to illustrate the minimum and maximum radii of the shell. D. The electrostatic potential surface of an isolated trimer. This figure was drawn with the program CCP4 mg (12, 13).

dodecamers indicates that the dodecamer is a stable structural unit of nsp10 under the crystallization conditions. The four nsp10 trimers (named trimers 1 to 4) in a dodecamer are related by a tetrahedral symmetry (Fig. 1). Any combination of three trimers is related by a local threefold symmetry, and so is a combination of two trimers related by a local twofold symmetry. The buried SAS of a trimer in the dodecamer is about 3,040 Å^2 (or about 16% of the total SAS of an nsp10 trimer), with 69% contributed by hydrophobic atom groups. The crystal structure shows that helix α1 and residues in the α2-β1 loop (residues 42 to 46) play a key role in trimer-trimer interactions. In the first (threefold symmetry-related) interaction region, Leu14, Cys17, Ala18, and Cys79 of trimer 1 form a hydrophobic base, which is directed towards Phe19 of trimer 2. At the same time, Phe19 of trimer 1 forms hydrophobic interactions with the equivalent hydrophobic base of trimer 3, and so on. In the second (twofold symmetry-related) interaction region, Met44 and Leu45 of trimer 1 are oriented to interact directly with their counterparts in trimer 2; Val42 and Tyr96 of trimer 1 interact with Tyr96 and Val42 of trimer 2, respectively. Both Cys17 and Cys79 contribute to stability of the dodecamer architecture, although they do not form a disulfide bridge in the reduced conformation state. Following this protocol, the 12 molecules can assemble to form a pseudododecahedron.

The zinc ion binding sites. Zinc binding is a major structural feature of nsp10. Spectral analysis of the nsp10 crystal during synchrotron data collection clearly demonstrated the presence of zinc, evident by a clear peak near the zinc absorption edge. It allows the nsp10 crystal structure to be solved by the SAD method. Furthermore, the 2.7-Å resolution structure of nsp10 determined from crystals prepared in the absence of additional zinc shows the same monomer fold, dodecameric architecture, and conformation in the zinc binding sites, with an RMSD of 0.4 Å for all Cα atoms in a dodecamer.

Two bound zinc ions were identified from the crystal structure of nsp10 with unambiguous electron density: one located in the N-terminal region on the inner surface of the dodecamer and the other one at the C-terminal region on the outer surface (Fig. 2 and 4). The first zinc binding site is formed by residues on helix α4 and the α4-β3 loop. This zinc ion is tetrahedrally chelated by the Sγ atoms of three cysteine residues (Cys74, Cys77, and Cys90) and the $N_{\varepsilon 2}$ atom of His83, which have bond distances of 2.3 Å, 2.3 Å, 2.4 Å, and 2.2 Å, respectively (Fig.

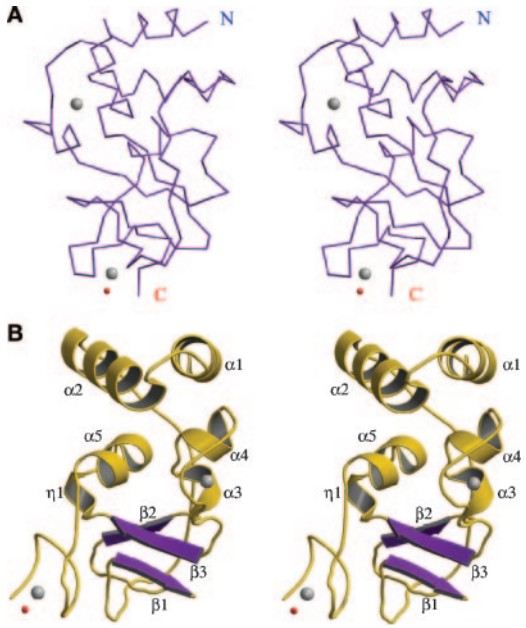

FIG. 3. nsp10 monomer fold. A. Stereo diagram showing a Cα trace of the nsp10 monomer. Two zinc ions are shown as gray spheres, and one chelating water molecule is shown as a smaller red sphere. B. Stereo ribbon diagram of the nsp10 monomer with β-strands in purple and α-helices in gold. Secondary structure elements are labeled. This view is rotated by ~90° relative to that of panel A. This figure was prepared with the programs Molscript (8), Bobscript (5), and Raster3D (10).

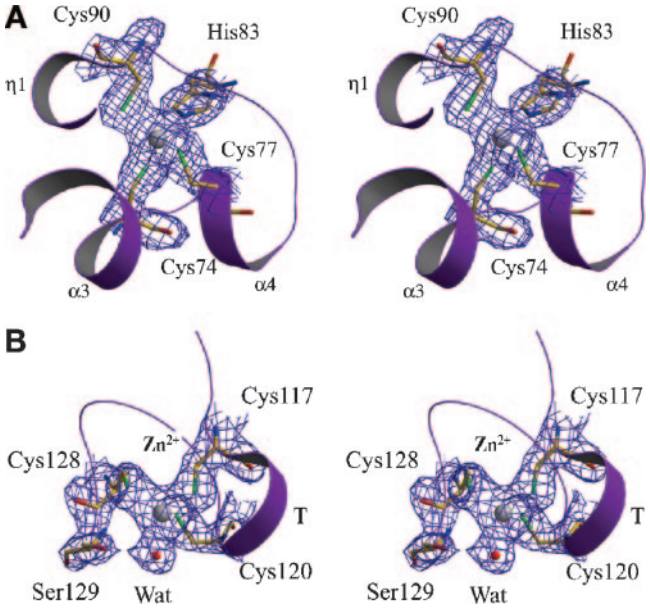

FIG. 4. Stereo views of the two independent zinc binding sites. A. The N-terminal zinc binding site with electron density. Key residues interacting with the zinc ion are labeled and depicted as sticks with carbon, nitrogen, oxygen, and sulfur atoms colored in yellow, blue, red, and green, respectively. The zinc ion is shown as a gray sphere. A refined-model-phased $2F_o$-F_c electron density map is shown for the zinc ion and chelated residues at 1.3 σ. Secondary structure elements are labeled. B. The C-terminal zinc binding site with electron density. The turn between C117 and C120 is labeled. This figure was prepared with the programs Molscript (8), Bobscript (5), and Raster3D (10).

4A; Table 2). This binding site constitutes a CCHC-type zinc finger with a C-$(X)_2$-C-$(X)_5$-H-$(X)_6$-C sequence motif.

The second zinc ion is located at the C terminus and is also tetrahedrally coordinated (Fig. 4B). Three zinc-chelating cysteine residues (Cys117, Cys120, and Cys128) have distances of 2.4 Å from the zinc ion to their side chain Sγ atoms. The fourth ligand is a water molecule, with clearly defined electron density, whose distance to the zinc ion is 2.5 Å. This chelating water molecule also interacts with Ser129 through a perfect hydrogen bond to form a stabilized hydrogen bond network. All bond lengths related to zinc binding are summarized in Table 2. Noteworthy is the observation that Cys130 neighbors Ser129 in the nsp10 sequence but did not exist in the crystallized protein. Otherwise, the presence of Cys130 might suitably position it to chelate the zinc ion, suggesting the second zinc binding site should also be a zinc finger of a C-$(X)_2$-C-$(X)_7$-C-(X)-C sequence motif. This CCCC-type zinc finger is distributed on the surface of the dodecamer and is relatively flexible in our structure compared to the CCHC-type zinc finger (Table 2). However, the flexibility of this region may not reflect the conformation of the full-length protein.

A multiple-sequence alignment of nsp10 from SARS-CoV with coronaviruses from groups I, II, and III of the genus *Coronavirus* indicates that all seven observed zinc-chelating residues, plus Cys130, are strictly conserved (Fig. 5), implying their importance in the functional replicase-transcriptase complex. In contrast, a number of other cysteine residues in SARS nsp10 are not conserved at all.

Functional implications. The finding of zinc binding sites in nsp10 provides a strong implication for nsp10 biological functions. The zinc finger is a common DNA binding motif found as part of transcription-regulatory proteins, which can often include more than one zinc finger in the same peptide chain. Moreover, increasing evidence suggests that zinc fingers can more widely be used to recognize RNA or even in protein-protein recognition.

Although the DALI search did not find any candidate structure similar to SARS-CoV nsp10 from PDB, a PFAM search (http://www.sanger.ac.uk/Software/Pfam) for similar sequence motifs identified several members of the HIT-type zinc finger family as nsp10 homologous candidates. Named after the first protein that originally defined the domain, the yeast HIT1

TABLE 2. Zinc chelation

Zinc ion	Chelating residue	Chelating atom	Bond length[a] (Å)
Zn 1	Cys74	S_γ	2.29 ± 0.04
	Cys77	S_γ	2.27 ± 0.04
	His83	$N_{\varepsilon 2}$	2.18 ± 0.05
	Cys90	S_γ	2.40 ± 0.03
Zn 2	Cys117	S_γ	2.41 ± 0.05
	Cys120	S_γ	2.45 ± 0.15
	Cys128	S_γ	2.38 ± 0.11
	Wat997	H_2O	2.51 ± 0.07
Ser129	Wat997	H_2O	2.76 ± 0.16

[a] The listed bond length is an average value over the 24 monomers with standard deviations.

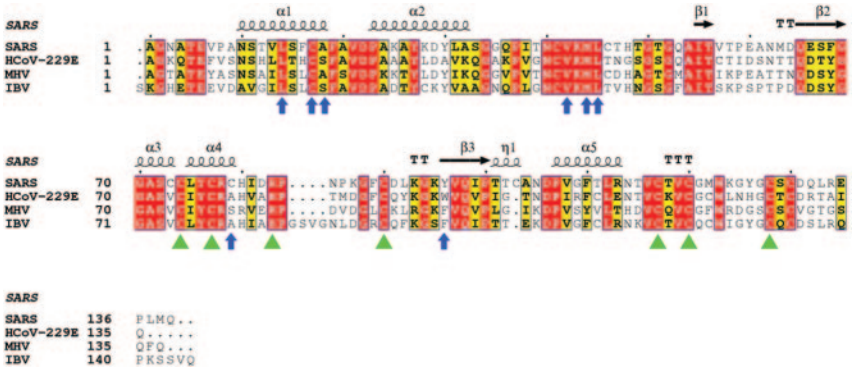

FIG. 5. Multiple-sequence alignment of SARS-CoV nsp10 with representatives from all three groups of the genus *Coronavirus*. HCOV-229E, human coronavirus strain 229E; MHV, mouse hepatitis virus; IBV, avian infectious bronchiolitis virus. The secondary structure for SARS-CoV is shown at the top of the alignment; arrows indicate β-strands, and helical curves denote α- or 3_{10}-helices. Residues highlighted in red are identical among the compared proteins; residues highlighted in yellow are conserved. Residues important for zinc binding are marked with green triangles, and residues important for stability of the dodecamer are marked with blue vertical arrows. The alignment was generated by ClustalX (20) and drawn with ESPript (6).

protein, the HIT-type zinc finger contains seven conserved cysteines and one histidine that can potentially coordinate two zinc atoms. While the function of the HIT-type zinc finger is unknown, this motif is mainly found in nuclear proteins involved in gene regulation and chromatin remodeling. To date, there are no three-dimensional structures of HIT-type zinc finger domains reported in the PDB. Therefore, our nsp10 crystal structure serves as the first example of a three-dimensional structure of this novel class of double zinc finger-containing motif.

This finding provides strong structural evidence that, like other better studied nsp proteins, nsp10 also likely plays a role in RNA synthesis, as suggested by other researchers (26). nsp10 is involved in network interactions with other nsp proteins, and the integrity of its zinc fingers seems important for such interactions. Experiments on mouse hepatitis virus (MHV), a group II coronavirus along with SARS-CoV, demonstrated the colocalization of nsp10 with nsp7, nsp8, and nsp9, providing solid evidence for their interaction in the coronavirus life cycle (2). Our chemical cross-linking experiment further demonstrated that SARS-CoV nsp10 can be cross-linked with nsp9 (data not shown), which itself interacts with nsp8 (17). Furthermore, an MHV *ts* mutant, Alb *ts*6, encoding a mutant form of nsp10 with a Gln65-to-Glu mutation, was shown to have a defect in negative-strand RNA synthesis (15). The Gln65 residue, conserved in all three groups of the genus *Coronavirus*, is located on strand β2 of the SARS nsp10 structure and hydrogen bonds via the $N_{\varepsilon 2}$ atom to the main chain carbonyl oxygen of Gly52. Gln65 is thus important for the conformational stability of nsp10 and particularly for the α4 helix which forms part of the N-terminal CCHC zinc finger. Therefore, mutation of Gln65 might be expected to perturb the folding of pp1a into a less productive conformation that would prevent it from participating in the formation of a replicase-transcriptase complex with negative-strand activity.

Conclusions. The scientific significance of the SARS-CoV nsp10 structure is at least threefold. First, nsp10 has a novel protein fold. A search with the DALI web engine (http://www.ebi.ac.uk/dali/) for structural homologs failed to yield any match to the nsp10 fold, suggesting a novel function for the nsp10 family members. Second, nsp10 possesses two zinc fingers, located in the N-terminal region and at the C-terminal region, with C-$(X)_2$-C-$(X)_5$-H-$(X)_6$-C and C-$(X)_2$-C-$(X)_7$-C-(X)-C sequence motifs, respectively. These motifs are conserved in all three groups of the genus *Coronavirus*, and our crystal structure illustrates for the first time the significance of these conserved residues. Further sequence analysis suggests that nsp10 is related to the HIT-type zinc finger family, which is often found in nuclear proteins involved in gene regulation and chromatin remodeling. Thus, our nsp10 crystal structure becomes the first of a new class of zinc finger protein three-dimensional structures to be revealed experimentally. Third, the molecular assembly of nsp10 is a hollow dodecamer with an outer diameter of 84 Å and an inner diameter of 36 Å. Twelve C-terminal zinc fingers stick out from the outer surface of the sphere, and another 12 zinc fingers are distributed around the inner surface. The strong positive electrostatic potential found on both the inner and outer surfaces of the dodecamer is intriguing, consistent with the probable function of nsp10 in the RNA synthesis machinery.

To date, the structures and functions of several components of the replication/transcription machinery have been determined, including the nsp3 ADP ribose 1-phosphatase domain (14); nsp5 (23, 24), nsp7, and nsp8 in complex (25); and nsp9 (4, 17). nsp5 is the main protease for cleavage of the replicase polyproteins, nsp7 and nsp8 are proposed to function as processivity factors for nsp12 (the RNA-dependent RNA polymerase), and nsp9 is a single-strand RNA binding protein. The crystal structure reported here will help to clarify the function(s) of nsp10, in which the presence of two zinc fingers should enable it to play an important role in RNA synthesis. Elucidation of the nsp10 structure will provide further insights into the sophisticated replication/transcription mechanism of SARS-CoV and other coronaviruses, such as mouse hepatitis virus (MHV), human coronavirus strain 229E, and human coronavirus strain HKU1.

ACKNOWLEDGMENTS

We thank Xiaoling Xu, Yuanyuan Xu, Feng Xu, Sheng Ye, Xiaorui Chen, Jun He, Xuemei Li, and Yi Han for technical assistance. We are

also grateful to Guangxia Gao and Hong Tang for valuable comments and critical discussions.

This work was supported by Project 973 of the Ministry of Science and Technology of China (grant number 2004CB720000), the NSFC (grant number 30221003), the Sino-German Center [grant number GZ236(202/9)], and the "Sino-European Project on SARS Diagnostics and Antivirals" (SEPSDA) of the European Commission (grant number 003831).

REFERENCES

1. **Bartlam, M., H. Yang, and Z. Rao.** 2005. Structural insights into SARS coronavirus proteins. Curr. Opin. Struct. Biol. **15:**664–672.
2. **Bost, A. G., R. H. Carnahan, X. T. Lu, and M. R. Denison.** 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.
3. **Brunger, A. T., P. D. Adams, G. M. Clore, W. L. DeLano, P. Gros, R. W. Grosse-Kunstleve, J. S. Jiang, J. Kuszewski, M. Nilges, N. S. Pannu, R. J. Read, L. M. Rice, T. Simonson, and G. L. Warren.** 1998. Crystallography and NMR system: a new software suite for macromolecular structure determination. Acta Crystallogr. Sect. D Biol. Crystallogr. **54:**905–921.
4. **Egloff, M. P., F. Ferron, V. Campanacci, S. Longhi, C. Rancurel, H. Dutartre, E. J. Snijder, A. E. Gorbalenya, C. Cambillau, and B. Canard.** 2004. The severe acute respiratory syndrome-coronavirus replicative protein nsp9 is a single-stranded RNA-binding subunit unique in the RNA virus world. Proc. Natl. Acad. Sci. USA **101:**3792–3796.
5. **Esnouf, R. M.** 1997. An extensively modified version of MolScript that includes greatly enhanced coloring capabilities. J. Mol. Graph Model **15:** 112–134.
6. **Gouet, P., X. Robert, and E. Courcelle.** 2003. ESPript/ENDscript: extracting and rendering sequence and 3D information from atomic structures of proteins. Nucleic Acids Res. **31:**3320–3323.
7. **Jones, T. A., J. Y. Zou, S. W. Cowan, and L. Kjeldgaard.** 1991. Improved methods for building protein models in electron density maps and the location of errors in these models. Acta Crystallogr. Sect. A. **47:**110–119.
8. **Kraulis, P. J.** 1991. MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures. J. Appl. Crystallogr. **24:**946–950.
9. **Laskowski, R. A., M. W. MacArthur, D. S. Moss, and J. M. Thornton.** 1993. PROCHECK: a program to check the stereochemical quality of protein structures. J. Appl. Crystallogr. **26:**283–291.
10. **Merritt, E. A., and D. J. Bacon.** 1997. Raster3D: photorealistic molecular graphics. Methods Enzymol. **277:**505–524.
11. **Otwinowski, Z., and W. Minor.** 1997. Processing of X-ray diffraction data collected in oscillation mode, p. 307–326. *In* C. W. Carter, Jr., and R. M. Sweet (ed.), Macromolecular crystallography, part A, vol. 276. Academic Press, New York, N.Y.
12. **Potterton, E., S. McNicholas, E. Krissinel, K. Cowtan, and M. Noble.** 2002. The CCP4 molecular-graphics project. Acta Crystallogr. Sect. D Biol. Crystallogr. **58:**1955–1957.
13. **Potterton, L., S. McNicholas, E. Krissinel, J. Gruber, K. Cowtan, P. Emsley, G. N. Murshudov, S. Cohen, A. Perrakis, and M. Noble.** 2004. Developments in the CCP4 molecular-graphics project. Acta Crystallogr. Sect. D Biol. Crystallogr. **60:**2288–2294.
14. **Saikatendu, K. S., J. S. Joseph, V. Subramanian, T. Clayton, M. Griffith, K. Moy, J. Velasquez, B. W. Neuman, M. J. Buchmeier, R. C. Stevens, and P. Kuhn.** 2005. Structural basis of severe acute respiratory syndrome coronavirus ADP-ribose-1″-phosphate dephosphorylation by a conserved domain of nsP3. Structure (Cambridge) **13:**1665–1675.
15. **Sawicki, S. G., D. L. Sawicki, D. Younker, Y. Meyer, V. Thiel, H. Stokes, and S. G. Siddell.** 2005. Functional and genetic analysis of coronavirus replicase-transcriptase proteins. PLoS Pathog. **1:**e39.
16. **Schneider, T. R., and G. M. Sheldrick.** 2002. Substructure solution with SHELXD. Acta Crystallogr. Sect. D Biol. Crystallogr. **58:**1772–1779.
17. **Sutton, G., E. Fry, L. Carter, S. Sainsbury, T. Walter, J. Nettleship, N. Berrow, R. Owens, R. Gilbert, A. Davidson, S. Siddell, L. L. Poon, J. Diprose, D. Alderton, M. Walsh, J. M. Grimes, and D. I. Stuart.** 2004. The nsp9 replicase protein of SARS-coronavirus, structure and functional insights. Structure (Cambridge) **12:**341–353.
18. **Terwilliger, T. C., and J. Berendzen.** 1999. Automated MAD and MIR structure solution. Acta. Crystallogr. Sect. D Biol. Crystallogr. **55:**849–861.
19. **Thiel, V., J. Herold, B. Schelle, and S. G. Siddell.** 2001. Viral replicase gene products suffice for coronavirus discontinuous transcription. J. Virol. **75:** 6676–6681.
20. **Thompson, J. D., T. J. Gibson, F. Plewniak, F. Jeanmougin, and D. G. Higgins.** 1997. The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res. **25:**4876–4882.
21. **Uson, I., and G. M. Sheldrick.** 1999. Advances in direct methods for protein crystallography. Curr. Opin. Struct. Biol. **9:**643–648.
22. **Vagin, A. A., and A. Teplyakov.** 1997. MOLREP: an automated program for molecular replacement. J. Appl. Crystallogr. **30:**1022–1025.
23. **Yang, H., W. Xie, X. Xue, K. Yang, J. Ma, W. Liang, Q. Zhao, Z. Zhou, D. Pei, J. Ziebuhr, R. Hilgenfeld, K. Y. Yuen, L. Wong, G. Gao, S. Chen, Z. Chen, D. Ma, M. Bartlam, and Z. Rao.** 2005. Design of wide-spectrum inhibitors targeting coronavirus main proteases. PLoS Biol. **3:**e324.
24. **Yang, H., M. Yang, Y. Ding, Y. Liu, Z. Lou, Z. Zhou, L. Sun, L. Mo, S. Ye, H. Pang, G. F. Gao, K. Anand, M. Bartlam, R. Hilgenfeld, and Z. Rao.** 2003. The crystal structures of severe acute respiratory syndrome virus main protease and its complex with an inhibitor. Proc. Natl. Acad. Sci. USA **100:** 13190–13195.
25. **Zhai, Y., F. Sun, X. Li, H. Pang, X. Xu, M. Bartlam, and Z. Rao.** 2005. Insights into SARS-CoV transcription and replication from the structure of the nsp7-nsp8 hexadecamer. Nat. Struct. Mol. Biol. **12:**980–986.
26. **Ziebuhr, J.** 2005. The coronavirus replicase. Curr. Top. Microbiol. Immunol. **287:**57–94.