JOURNAL OF VIROLOGY, Oct. 2005, p. 12905–12913 Vol. 79, No. 20
0022-538X/05/$08.00+0 doi:10.1128/JVI.79.20.12905–12913.2005

Structural Genomics of the Severe Acute Respiratory Syndrome Coronavirus: Nuclear Magnetic Resonance Structure of the Protein nsP7

Wolfgang Peti,[1,2]† Margaret A. Johnson,[1,2] Torsten Herrmann,[1]‡ Benjamin W. Neuman,[2,3] Michael J. Buchmeier,[2,3] Mike Nelson,[1,2] Jeremiah Joseph,[2,4] Rebecca Page,[1,5]§ Raymond C. Stevens,[1,2,5] Peter Kuhn,[2,4,5]* and Kurt Wüthrich[1,2,5]*

Department of Molecular Biology,[1] Consortium for Functional and Structural Proteomics of SARS-CoV Related Proteins,[2] Department of Neuropharmacology,[3] Department of Cell Biology,[4] and Joint Center for Structural Genomics,[5] The Scripps Research Institute, 10550 North Torrey Pines Rd., La Jolla, California 92037

Received 9 June 2005/Accepted 22 July 2005

Here, we report the three-dimensional structure of severe acute respiratory syndrome coronavirus (SARS-CoV) nsP7, a component of the SARS-CoV replicase polyprotein. The coronavirus replicase carries out regulatory tasks involved in the maintenance, transcription, and replication of the coronavirus genome. nsP7 was found to assume a compact architecture in solution, which is comprised primarily of helical secondary structures. Three helices (α2 to α4) form a flat up-down-up antiparallel α-helix sheet. The N-terminal segment of residues 1 to 22, containing two turns of α-helix and one turn of 3_{10}-helix, is packed across the surface of α2 and α3 in the helix sheet, with the α-helical region oriented at a 60° angle relative to α2 and α3. The surface charge distribution is pronouncedly asymmetrical, with the flat surface of the helical sheet showing a large negatively charged region adjacent to a large hydrophobic patch and the opposite side containing a positively charged groove that extends along the helix α1. Each of these three areas is thus implicated as a potential site for protein-protein interactions.

The severe acute respiratory syndrome coronavirus (SARS-CoV) most closely resembles the group II coronaviruses, which infect mice, rats, pigs, and humans (34). Upon viral entry, the ~30-kb SARS-CoV genome is translated to produce a predicted 486-kDa polyprotein (PP1a) as well as a longer form of the polyprotein containing a predicted 304-kDa carboxyl-terminal extension (PP1ab) that is generated via a ribosomal frameshift event (38). The PP1ab form of the replicase polyprotein contains enzymatic signatures likely involved in RNA replication and processing. The short and long forms of the polyprotein are proteolytically processed into about 16 mature polypeptides (nonstructural proteins; nsP) by proteinases encoded in PP1a (31, 34). These polypeptides form the subunits of a replicase complex that associates with intracellular membranes and is responsible for replication of the viral genome at defined intracellular sites.

* Corresponding author. Mailing address for Peter Kuhn: Department of Cell Biology, The Scripps Research Institute, 10550 North Torrey Pines Rd., CB-265, La Jolla, CA 92037. Phone: (858) 784-9114. Fax: (858) 784-8996. E-mail: pkuhn@scripps.edu. Mailing address for Kurt Wüthrich: Department of Molecular Biology, The Scripps Research Institute, 10550 North Torrey Pines Rd., MB-44, La Jolla, CA 92037. Phone: (858) 784-8011. Fax: (858) 784-8014. E-mail: wuthrich@scripps.edu.

† Present address: Brown University, Department of Molecular Pharmacology, Physiology and Biotechnology, 70 Ship Street, GE-3, Providence, RI 02912.

‡ Present address: Institut für Molekularbiologie und Biophysik, ETH Zürich, CH-8093 Zürich, Switzerland.

§ Present address: Brown University, Department of Molecular Biology, Cell Biology and Biochemistry, 70 Ship Street, GE-4, Providence, RI 02912.

The functions of the nonstructural proteins of the replicase complexes of coronaviruses are, as yet, poorly defined. Several proteins are predicted to be involved in RNA processing. For example, nsP13 has confirmed helicase activity (13, 33) and nsP12 is predicted to be an RNA-dependent RNA polymerase (34), while several other proteins have, as yet, no functional annotation. In addition to their implicated roles in viral genome replication and subgenomic RNA synthesis, new roles for nsP as determinants of pathogenicity have been postulated (22). For example, certain viruses modulate host cell signaling pathways to down-regulate the immune response, modify cytokine secretion, and allow greater viral proliferation (1). Host cell apoptotic pathways may also be up- or down-regulated (3). Thus, the overexpression of a protein unique to SARS-CoV, the accessory protein ORF7a, has been shown to stimulate apoptosis via a caspase-dependent pathway (37). Interactions between nsP1 of equine arteritis virus (*Nidovirales* order) and host cell transcription regulatory factors have been demonstrated (39).

The consortium for Functional and Structural Proteomics of SARS-CoV-related proteins (http://sars.scripps.edu) was established to provide structural information for SARS-CoV proteins and to characterize their protein-protein and protein-nucleic acid interactions. A structural genomics approach originally developed for *Thermotoga maritima* (reference 17 and http://www.jcsg.org) was adapted for the 28 proteins encoded by the SARS genome. A bioinformatics approach to domain identification and definition within the genome was used to design 163 constructs, most of which were cloned and tested for soluble expression in a small-scale structural genomics ex-

pression system (28). Several expressing constructs were biophysically analyzed using one-dimensional (1D) ^{1}H nuclear magnetic resonance (NMR) screening (29, 30), and the first target selected from this screen for structure determination was the nonstructural protein 7 (nsP7).

Here, we describe the structure of the predicted nsP7 domain of the polyprotein 1ab (PP1ab). nsP7 is conserved within the *Coronaviridae* and has no detectable orthologs outside this family of viruses. nsP7 is contained within the portion of *pp1ab* thought to comprise a replication complex and is an 83-amino-acid polypeptide predicted to include four α-helices. The expression of nsP7 in infected cells, and its localization to cytoplasmic, membrane-containing foci thought to be sites of viral RNA replication, have been demonstrated for cells infected with mouse hepatitis virus (MHV) (2), human coronavirus 229E (44), and avian infectious bronchitis virus (23). Furthermore, the MHV nsP7 has been shown to interact specifically with nsP10, a protein also specific to coronaviruses, and with nsP1, a protein thought to be involved in viral replication and assembly (4). These data indicate a function of nsP7 in coronavirus-specific RNA replication mechanisms.

MATERIALS AND METHODS

Cloning of the SARS-CoV proteome. Vero-E6 cells were inoculated with SARS-CoV strain Tor2 (GenBank accession number NC_004718) at a multiplicity of ~10 PFU/cell. Twenty-four hours after inoculation, cells were lysed with TRIzol (Invitrogen) and RNA was extracted according to the manufacturer's protocol. First-strand cDNA synthesis using murine leukemia virus reverse transcriptase (Invitrogen) was primed with random hexamer, oligonucleotide T, or SARS-CoV-specific oligonucleotides. SARS-specific primers used for first-strand synthesis were as follows: SARS-1r, CTTCAGGTGTAGGTTCTGG; SARS-4r, CAGTCTTTAATAATGATTGGC; SARS-7r, GAGTTAAATAAAGAGTGT CTG; and SARS-10r, TTTTTTTTTTGTCATTCTCC. Full-length nsP7 amplicons were obtained by PCR using the following primers: SARS070f, ATGTCT AAAATGTCTGACGTAAAGTGCACATCTG; and SARS070r, CCCGGCCG GCCCTACTGAAGAGTAGCACGGTTATCG. SARS-CoV cDNA was cloned into pMH1F, which is a customized expression vector derived from pBAD (Invitrogen). Expression in pMH1F is driven by the *araBAD* promoter, and the recombinant protein is produced with a $Thio_6His_6$ tag (MGSDKIHHHHHH) at its N terminus. Four designed nsP7 constructs were transformed into the methionine auxotrophic *Escherichia coli* strain DL41, and microexpression trials were conducted as described previously (28), using 2XYT as the growth medium and 0.2% (wt/vol, final concentration) L-arabinose (Sigma, St. Louis, MO) as the inducer at different temperatures. Cell pellets were lysed by resuspension and incubation at room temperature for 15 min in a solution of lysozyme (1 mg/ml) in 50 mM Tris-HCl, pH 7.5, with 50 mM sucrose, 1 mM EDTA, and 0.25 μl/ml Benzonase endonuclease. Equal volumes of 10 mM Tris-HCl [pH 7.5] with 50 mM KCl, 10 mM $MgCl_2$, and 1 mM EDTA were added, and the suspensions were incubated for a further 15 min at room temperature. Cell debris was collected by centrifugation, and the soluble protein fractions were evaluated by sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE). Soluble protein constructs were selected for larger-scale fermentation and further evaluation. Larger-scale fermentation was carried out in a fermenter equipped for 65-ml cultures (17) with terrific broth as the growth medium; otherwise, growth conditions were as described for microexpression. Proteins were purified by IMAC Co^{2+}-affinity chromatography (Talon resin; Clontech) and by a second ion exchange step. Purified proteins were subjected to 1D ^{1}H NMR screening (29, 30).

Production of nsP7 for NMR spectroscopy. A construct representing nsP7 with an extra N-terminal dipeptide Gly-His (residues 1 and 2) was subcloned into a vector derived from pET-28 (Novagen), which encodes a $Thio_6His_6$ expression/purification tag (MGSDKIHHHHHH) and a tobacco etch virus (TEV) cleavage site (ENLYFQGH). This plasmid was transformed into the *E. coli* strain BL21-CodonPlus (DE3)-RIL (Stratagene). The expression of uniformly ^{15}N-labeled and ^{13}C,^{15}N-labeled nsP7 was carried out by growing freshly transformed cells in M9 minimal medium containing 1 g/liter $^{15}NH_4Cl$ and 4 g/liter [$^{13}C_6$]-D-glucose as the sole nitrogen and carbon sources, respectively. Cell cultures were grown at 37°C with vigorous shaking to an optical density at 600 nm of 0.6 to 0.7. The temperature was slowly lowered to 18°C, and after induction with 1 mM isopropyl-β-D-thiogalactopyranoside, the cell cultures were grown for 18 h. The cells were harvested by centrifugation, resuspended in extraction buffer (50 mM Tris-HCl at pH 8.0, 5 mM imidazole, 500 mM NaCl, 0.1% Triton X-100, and Complete protease inhibitor tablets [Roche]), and lysed by sonication. The cell debris was removed by centrifugation. For the first purification step, the soluble protein was loaded onto a HisTrap HP column (Pharmacia), equilibrated with 50 mM Tris-HCl at pH 8.0, 5 mM imidazole, and 500 mM NaCl. The protein was eluted with a 0 to 250 mM imidazole gradient. Fractions containing nsP7 were pooled and buffer exchanged against 50 mM sodium phosphate at pH 7.5 with 300 mM NaCl and concentrated to a total volume of about 10 ml. After the addition of TEV NIa protease, the solution was vigorously shaken for 3 to 6 h at room temperature. The progression of the TEV cleavage was tested by SDS-PAGE analysis. After cleavage was at least 95% complete, the sample was again concentrated and loaded onto a precalibrated (50 mM sodium phosphate at pH 7.5, 300 mM NaCl) IMAC column (Talon resin; Clontech). The fractions containing nsP7 were again pooled, the homogeneity of the purified protein was evaluated by SDS-PAGE, and the solution was concentrated to a final volume of 550 μl and dithiothreitol-d_{10} was added at a concentration of 5 mM. The final concentrations of nsP7 in the different NMR samples were between 1.0 and 3.5 mM.

Data collection. NMR measurements were performed at 298 K with Bruker Avance600 and Avance900 spectrometers, using TXI-HCN-z or TXI-HCN-xyz gradient probe heads. Proton chemical shifts were referenced to internal 3-(trimethylsilyl)-1-propanesulfonic acid sodium salt (DSS). The ^{13}C and ^{15}N chemical shifts were referenced indirectly to DSS, using the absolute frequency ratios.

Chemical shift assignment and structure calculation. The determination of the 3D structure of a protein (41, 42) requires sequence-specific resonance assignment (assignment of each ^{1}H, ^{13}C, and ^{15}N resonance frequency to a specific atom within the protein), obtained by combining the results of several 2D and 3D NMR experiments recorded on uniformly ^{13}C- and ^{15}N-enriched protein samples. A separate set of experiments based on the nuclear Overhauser effect (NOE) measures interatomic distances within the protein. These distances are applied as restraints during molecular dynamics simulations, which also include restraints on bond lengths and angles, chirality, planarity, and torsional angles, to enforce correct geometry. A molecular dynamics protocol implemented in the program DYANA, in which torsion angles rather than Cartesian coordinates are the degrees of freedom, offers good sampling and convergence properties for biomolecular structures (7–9). Calculations are repeated several times with different randomized starting structures to yield an ensemble of conformers that are representative of the conformation space sampled by the protein in solution.

The following spectra (32) were used to obtain sequence-specific backbone and side chain assignments: 2D [^{1}H,^{15}N]-heteronuclear single quantum coherence (HSQC), 3D HNCA, 3D HNCACB, 3D CBCA(CO)NH, 3D HNCO, 3D HBHA(CO)NH, 3D ^{15}N-resolved [^{1}H,^{1}H]–total-correlation spectroscopy (TOCSY), and 3D HC(C)H-TOCSY. 2D [^{1}H,^{1}H]-NOE spectroscopy (NOESY), 2D [^{1}H,^{1}H]-TOCSY, and 2D [^{1}H,^{1}H]-correlation spectroscopy (COSY) of a nsP7 sample in D_2O solution after complete H/D exchange of the labile protons were used to assign the aromatic side chains (41). The NMR spectra were processed with XWINNMR3.5 (Bruker, Billerica, Mass.) and analyzed with CARA (R. Keller et al., unpublished data).

The input for the structure calculation was collected from 3D ^{15}N-resolved [^{1}H,^{1}H]-NOESY and 3D ^{13}C-resolved [^{1}H,^{1}H]-NOESY spectra recorded in H_2O solution and a 2D [^{1}H,^{1}H]-NOESY spectrum recorded in D_2O solution. All three NOESY spectra were measured at 900 MHz with a mixing time of 90 ms and were automatically analyzed with a standalone version of the new software package ATNOS/CANDID (version 0.9), which incorporates the functionalities of the two algorithms ATNOS (11) for automated peak picking and NOE identification in 2D homonuclear- and 3D heteronuclear-resolved [^{1}H,^{1}H]-NOESY spectra, and CANDID (10) for automated NOE assignment. ATNOS/CANDID was combined with the program DYANA (9), which was used to perform the structure calculation with simulated annealing in torsion angle space. The ATNOS/CANDID input consisted of the chemical shift lists obtained from the sequence-specific resonance assignment and the three aforementioned NOESY spectra. The standard protocol with seven cycles of peak picking, NOE assignment, and 3D structure calculation was applied. At the outset of the spectral analysis, ATNOS/CANDID used highly permissive criteria to identify a comprehensive set of peaks in the NOESY spectra. Only the knowledge of the covalent polypeptide structure and the chemical shift lists were exploited to guide NOE cross peak identification, and ambiguous constraints (24) were used for the NOE assignment. In the second and subsequent cycles, the intermediate protein three-dimensional structures served as an additional guide for the interpretation of the NOESY data. The ATNOS/CANDID output consisted of assigned NOE

peak lists for each input spectrum, and a final set of meaningful upper limit distance constraints which constituted the input for the DYANA three-dimensional structure calculation algorithm. For each cycle of structure calculation, constraints on the backbone dihedral angles angles ϕ and ψ derived from the C^{α} chemical shifts (19, 35) were added to the ATNOS/CANDID output. For the final structure calculation in cycle 7, ATNOS/CANDID retained only distance constraints that could be unambiguously assigned based on the protein three-dimensional structure from cycle 6. The 20 conformers with the lowest residual DYANA target function values obtained from cycle 7 were energy refined in a water shell with the program OPALp (14, 18), using the AMBER force field (5). The program MOLMOL (15) was used to analyze the protein structure and to prepare the figures of the NMR structures.

Validation and data deposition. Analysis of the stereochemical quality of the models was accomplished using the Joint Center for Structural Genomics (JCSG) Validation Central suite (http://www.jcsg.org), which integrates seven validation tools: Procheck 3.5.4, SFcheck 4.0, Prove 2.5.1, ERRAT, WASP, DDQ 2.0, and Whatcheck. The ^{1}H, ^{13}C, and ^{15}N chemical shifts have been deposited in the BioMagResBank (BMRB; http://www.bmrb.wisc.edu) under BMRB accession number 6513.

Protein structure accession number. The atomic coordinates of the bundle of 20 conformers used to represent the nsP7 structure have been deposited in the Protein Data Bank (http://www.rcsb.org/pdb/) with the code 1YSY.

RESULTS AND DISCUSSION

Structural genomics strategy. The SARS-CoV genome is predicted to encode 28 proteins (34). Bioinformatics analyses of these protein sequences involving prediction of secondary structure, domain boundaries, and disordered regions (details to be published elsewhere) yielded a set of 163 constructs, which had a reasonable probability of yielding soluble recombinant proteins while providing redundant coverage of the proteome. These constructs were amenable to processing via a high-throughput structural genomics pipeline adapted from the JCSG (reference 17 and http://www.jcsg.org) that included (i) PCR amplification of the constructs from a SARS cDNA library and the cloning of these constructs into multiple expression vectors, (ii) microexpression trials of the cloned constructs in multiple *E. coli* strains using different induction temperatures, (iii) large-scale fermentation of the expressing clones, and (iv) purification of the expressed proteins by Co^{2+}-affinity purification and ion exchange chromatography.

Expression and solubility screening of nsP7 constructs. Four alternate nsP7 constructs were designed based on the prediction that the first seven and the last five residues of the sequence would form disordered random coil segments. These constructs were prepared as described in Materials and Methods and tested for expression in *E. coli*. Soluble expression was observed for full-length nsP7 and for the individually N- or C-terminally truncated constructs, but not for the combined N- and C-terminally truncated constructs. Since in this case no improvement in expression or solubility resulted from truncation, the full-length protein was selected for further studies.

1D ^{1}H NMR fold screening. A subset of the SARS-CoV protein domain constructs that were successfully expressed in *E. coli* was screened for globular folding in solution, using a 1D ^{1}H NMR screening approach developed for structural genomics (29, 30). nsP7 was chosen as a promising target for NMR structure determination based on a 1D spectrum indicative of a well-folded protein.

Structure determination. Using an input consisting of the chemical shift lists from the sequence-specific resonance assignments of nsP7 and of the three NOESY spectra described in Materials and Methods, the program package ATNOS/CANDID yielded a total of 2,413 assigned NOE cross peaks in the final cycle 7. These yielded 1,066 meaningful NOE upper distance limits as input for the final structure calculation with the program DYANA (Table 1) (Fig. 1). The low residual DYANA target function value of 1.73 ± 0.30 Å^2 (Table 1) and the average global root-mean-square deviation (RMSD) value relative to the mean coordinates of 0.89 ± 0.19 Å calculated for the backbone atoms of residues 6 to 83 in the bundle of Fig. 1a (Table 1) represent a high-quality NMR structure determination.

NMR structure of nsP7. The most striking feature of nsP7 is that three helices, α2 (29 to 42), α3 (47 to 65), and α4 (71 to 81), form a flat up-down-up three-α-helix sheet linked by two short, well-defined loops with residues 43 to 46 and 66 to 70. The stabilization of this unusual structural motif by side chain-side chain interactions is discussed below. Overall, the solution structure of nsP7 (Fig. 1) consists of a single domain that contains a total of five helical secondary structures. Helix α1 spans the residues 11 to 17 and is connected via a two-amino-acid linker with a 3_{10}-helix of residues 20 to 22. The 3_{10}-helix leads, via a somewhat disordered loop of residues 23 to 28, to α2 (Fig. 1a). The helix α1 is packed at a 60° angle against the surface of α2 and α3 in the flat three-helix sheet, and the 3_{10}-helix runs parallel to α3.

The structure comparison programs DALI (12) and FATCAT (43) at first indicated apparent statistically significant structural similarity to several previously described folds. However, the apparent similarities are all to helix bundles, where three or four helices contribute nearly equally to form a tight core. In nsP7, the arrangement of the helices α2 to α4 into a flat sheet is unique in that no such arrangement of three sequentially adjoining α-helices could be found in the SCOP (21) or CATH (26) databases, indicating that nsP7 actually represents a novel fold. The nsP7 fold may alternatively be viewed as comprising an antiparallel three-helix bundle (helices α1 to α3) with an additional helix (α4) added at the C terminus of the three-helix bundle in an antiparallel orientation relative to α3.

The term "helical sheet" has previously been used to describe helix packing in large proteins, such as annexins, but in these structures the helices nonetheless tend to form tightly packed bundles. Considering the apparent novelty of the nsP7 fold, we further investigated the role of the side chain packing in stabilizing this molecular architecture, which revealed distinctly different interhelical side chain-side chain interactions in the individual pairs of helices. The tightest association is seen between α2 and α3. These antiparallel helices are closely packed together, and their axes are oriented at an angle of 10°. They associate by virtue of the interdigitation of side chains that are separated by three or four residues in the sequence and are therefore positioned in two ridges along each helix. These residues are predominantly hydrophobic and form two interdigitated layers that hold the helices together (Fig. 2). The residues forming the layers are, on the one hand, Leu 30, Cys 34, and Ile 41 in the helix α2 and Ala 50, Leu 57, and Leu 61 in α3 and, on the other hand, Trp 31, His 38, and Leu 42 in α2 and Thr 47, Met 54, Leu 58, and Gln 65 in α3. These interactions are shown explicitly in Fig. 2a and b and schematically in the helical wheel plot of Fig. 2c, where the two layers are indicated by the two green boxes. This type of interaction is reminiscent of that observed in coiled coils, where the super-

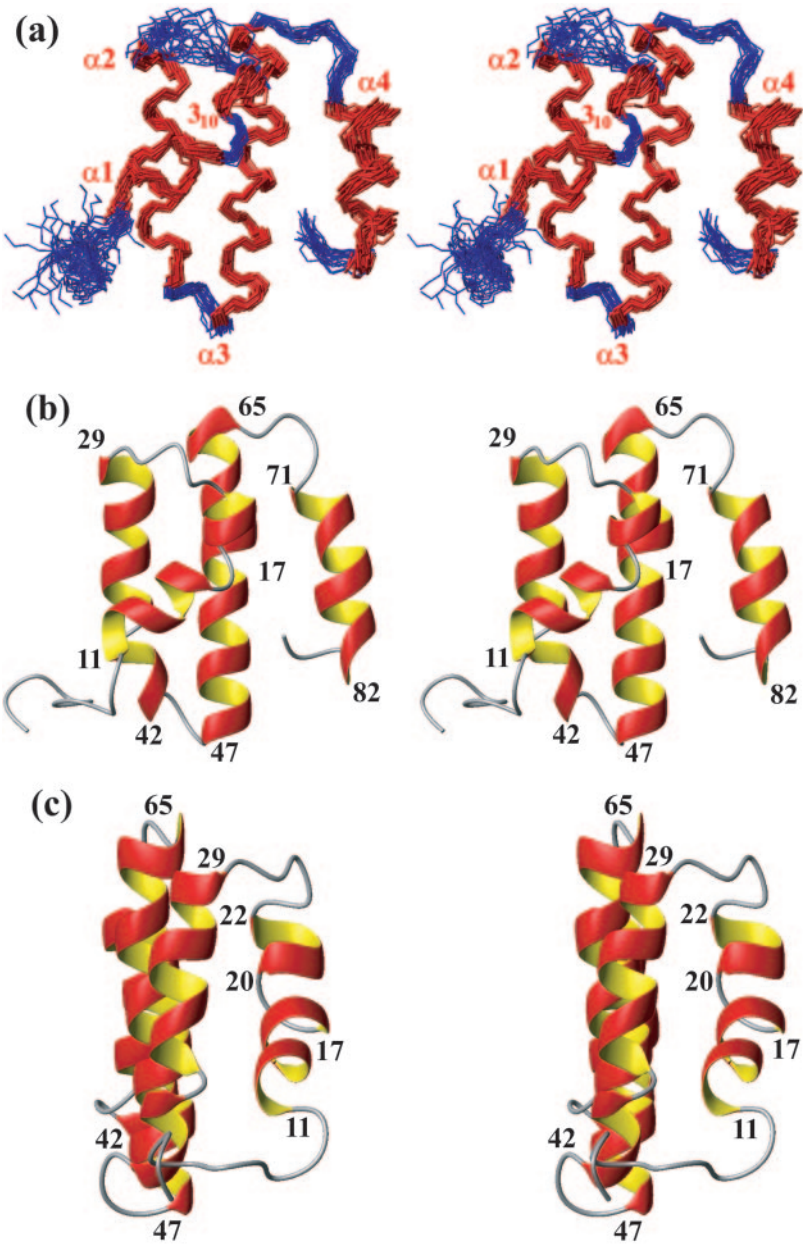

FIG. 1. Wall-eye stereo views of the solution structure of nsP7. (a) Bundle of the best 20 DYANA conformers of nsP7 after energy minimization. Only the polypeptide backbone is displayed. The 20 conformers have been superimposed for minimal RMSD of the backbone heavy atoms of residues 6 to 83. The four major helices, α1 to α4, and the 3_{10}-helix are colored red and labeled at their N termini. (b) Ribbon presentation of the closest conformer of nsP7 to the mean coordinates of the bundle shown in panel a, shown at the same viewing angle as for panel a. The sequence positions at both ends of the four α-helices are identified. (c) Same as panel b after rotation about a vertical axis, such that one looks at the edge of the up-down-up antiparallel three-α-helix sheet. The sequence positions at both ends of the helices α1, 3_{10}, α2, and α3 are identified.

coiling of helices results from the coiling of the ridges around the helix axis (25, 27). However, in nsP7 the helices are relatively short, a characteristic which, combined with slight distortions of the helices at each end, allows the interhelix angle to remain small and the helices to remain antiparallel rather than coiling around each other.

Helix α1 (11 to 17) and the 3_{10}-helical turn following it (20 to 22) associate closely with α2 and α3, and these three helices are arranged similarly to other known three-helix bundles. Interhelical side chain-side chain hydrophobic interactions appear to be responsible for this association. Thus, Leu 15 and Leu 19 of α1 associate with Val 35, His 38, Met 54, Val 55, and Leu 58 of α2 and α3, while Leu 16 of α1 and Val 24 of the first loop contact the indole ring of Trp 31 in α2 (Fig. 2a and c). Val 14 and Val 18 of α1 also contact the surface of helix α3 but are partly solvent exposed. The observation that several of the aforementioned residues involved in interactions between α1, α2, and α3 are highly conserved in coronavirus nsP7 sequences (Fig. 3) is consistent with the conclusion that they have a key role in stabilizing the fold.

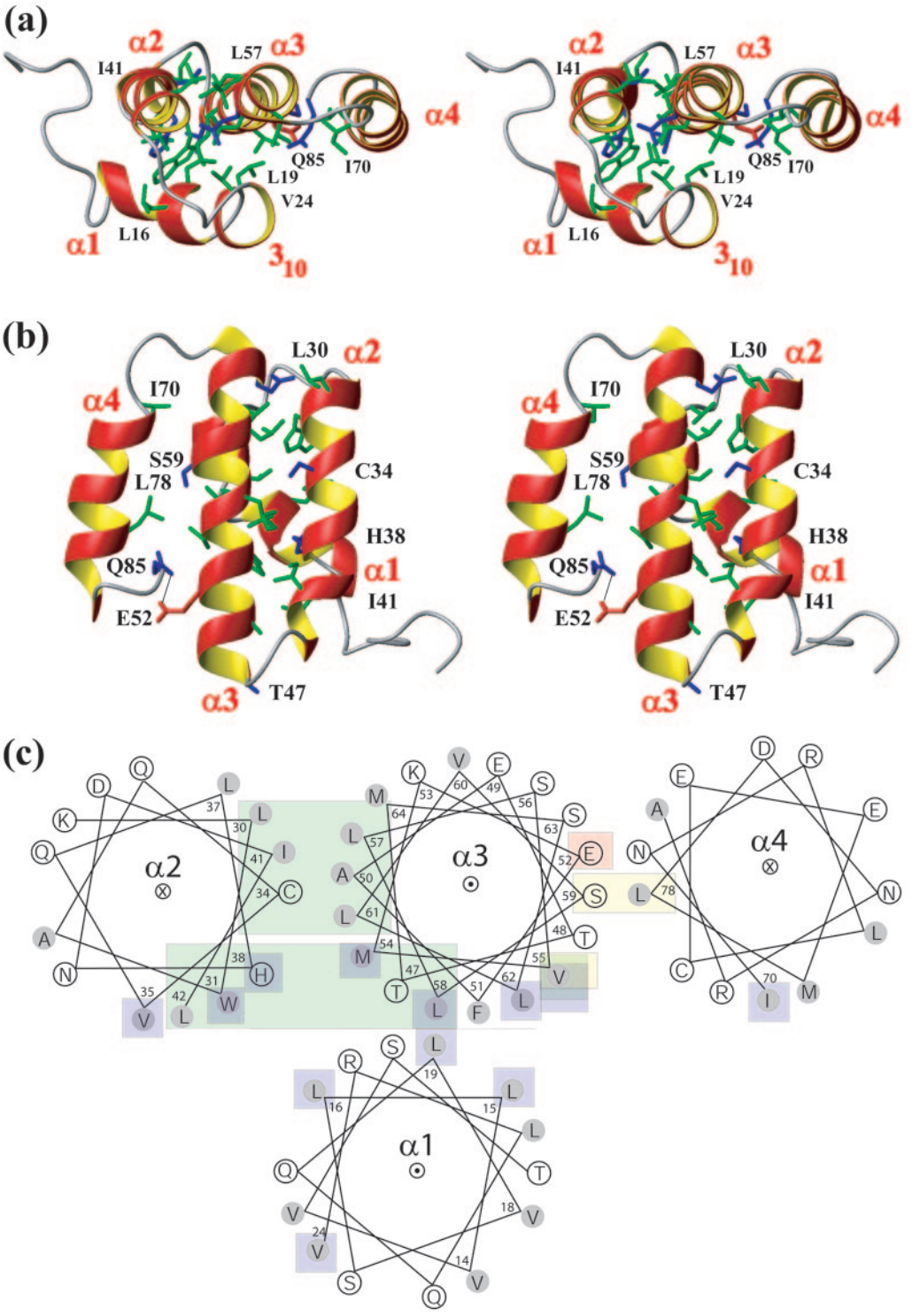

FIG. 2. (a) Top view of nsP7, generated from Fig. 1b by a 90° rotation about a horizontal axis in the projection plane. The helices are shown in a ribbon representation and are labeled at their N termini. Side chains involved in interhelix interactions (see the text) are shown as stick representations using the following color code: green, Ala, Leu, Val, Ile, Tyr, Trp, and Phe; blue, Ser, Thr, Cys, His, Asn, and Gln; red, Arg, Lys, Asp, and Glu. Some of the side chains discussed in the text are identified with the one-letter amino acid code and the residue number. (b) View of nsP7, generated from Fig. 1b by a 180° rotation about a vertical axis; the presentation is the same as for panel a. Some of the side chains discussed in the text are labeled, and a hydrogen bond between Glu 52 and Gln 85 is shown as a thin black line. Panels a and b show wall-eye stereo views. (c) Schematic top view of nsP7 (same as for panel a) with the helices represented as helical wheels. The helices α2 and α4 are directed from N to C into the page, and α3 runs out of the plane toward the viewer. α1 and the 3_{10}-helical turn following it are represented as one helical wheel running out of the page; this presentation does not show the tilt of α1 relative to the other three helices (Fig. 1). The side chains are represented as circles, with the hydrophobic side chains shaded. The side chains involved in α2-α3 interactions are shown on a green background, α3-α4 interactions are on a yellow background, and α1-α2, α3 interactions are on a blue background. Glu 52 is shown on a red background to indicate the hydrogen bonding interaction with the C-terminal residue, Gln 85. The helical wheel plot was prepared with the Web-based tool at http://kael.net.

TABLE 1. Input for the structure calculation and characterization of the energy-minimized bundle of 20 DYANA conformers of nsP7

Parameter	Value[a]
NOE upper distance limits	1,066
Dihedral angle constraints	96
Residual target function (Å^2)	1.73 ± 0.30
Residual NOE violations	
No. >0.1 Å	25 ± 4
Maximum (Å)	0.17 ± 0.10
Residual dihedral angle violations	
No. >2.5°	1 ± 1
Maximum (°)	3.21 ± 1.02
Amber energies (kcal/mol)	
Total	−2,995.65 ± 157.74
van der Waals	−172.00 ± 31.55
Electrostatic	−3,566.95 ± 103.42
RMSD from ideal geometry[b]	
Bond lengths (Å)	0.0086 ± 0.0022
Bond angles (°)	2.257 ± 0.171
RMSD to the mean coordinates (Å)[b]	
bb (6–83)	0.89 ± 0.19
ha (6–83)	1.35 ± 0.17
Ramachandran plot statistics (%)[c]	
Most favored regions	65
Additional allowed regions	30
Generously allowed regions	3
Disallowed regions	2

[a] Except for the top two entries, the average values for the 20 energy-minimized conformers with the lowest residual DYANA target function values and the standard deviation among them are listed, with the ranges indicating minimum and maximum values.

[b] bb indicates the backbone atoms N, Cα, C′; ha stands for "all heavy atoms." The numbers in parentheses indicate the residues for which the RMSD was calculated.

[c] Determined by PROCHECK (16, 20).

Helix α4 appears to be only weakly associated with α3. These two helices are connected by a five-residue loop, in which Ala 67 and Ile 70 associate with Leu 62 of α3. The N-terminal end of α4 is therefore in close proximity to α3, while the C-terminal end of α4 is displaced away from the N terminus of α3. The C-terminal residue, Gln 85, is inserted between α3 and α4 and forms a hydrogen bond to the Glu 52 side chain of α3 (Fig. 2b). Helix α4 is predominantly polar, with Leu 78 being the only hydrophobic residue pointing toward α3 (Fig. 2b). Polar and electrostatic interactions involving Glu 52, Lys 53, Ser 59, Cys 74, Asp 79, Arg 81, and the C-terminal carboxylate group may also contribute to stabilizing the α3-α4 packing arrangement. The lack of sequence conservation in helix α4 (Fig. 3) would be consistent with the assumption that the α3-α4 association can be supported by a variety of different interactions.

The positioning of the C-terminal tripeptide segment of nsP7 between the helices α3 and α4 is intriguing, since it might actually serve to stabilize this part of the fold. The residues Leu 84 and Gln 85 form part of a conserved proteolytic cleavage site and are stringently required for cleavage of the polyprotein by the 3C-like protease (3CLpro). The crystal structure of 3CLpro with a bound inhibitor (40) shows that the side chain of the glutamine residue, which is absolutely required for proteolysis, inserts into a conserved pocket in the protease active site where it forms hydrogen bonds to histidine and glutamate residues. The Leu residue also interacts with the protease, although it is partially solvent exposed, and SARS-CoV 3CLpro has relaxed specificity at this position in that it also tolerates other hydrophobic residues (38). In the solution structure of nsP7, the glutamine residue would be nearly inac-

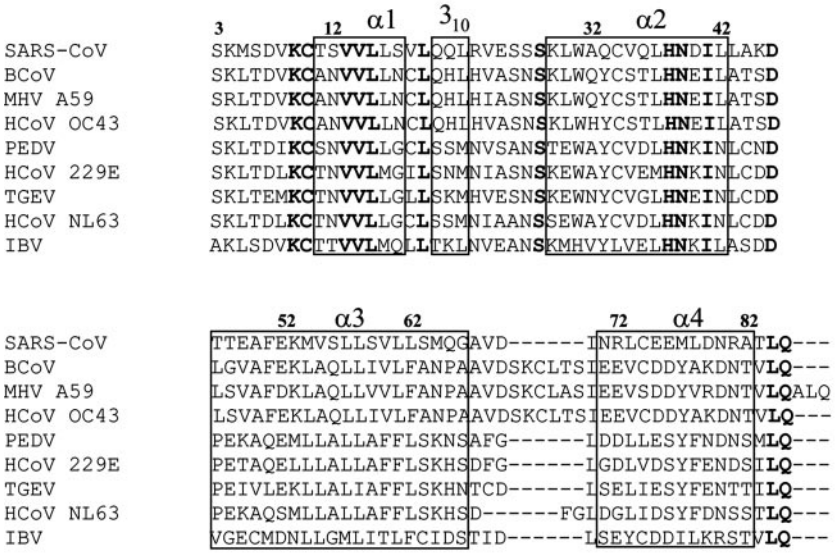

FIG. 3. Multiple alignment of coronavirus nsP7 sequences. Strictly conserved positions are indicated in boldface type. The positions of the helices α1 to α4 and 3_{10} are indicated by boxes. Abbreviations and GenBank accession codes for the nsP7 sequences used are as follows: SARS-CoV, SARS coronavirus, strain Tor2, NP_828865; PEDV, porcine epidemic diarrhea virus, strain CV777, NP_839961; HCoV 229E, human coronavirus 229E, NP_835348; TGEV, transmissible gastroenteritis virus, strain Purdue, NP_840005; BCoV, bovine coronavirus ENT, NP_742134; MHV A59, murine hepatitis virus, strain A59, NP_740612; HCoV OC43, human coronavirus OC43, strain ATCC VR-759, NP_937947; HCoV NL63, human coronavirus NL63 strain Amsterdam I, YP_003766; IBV, avian infectious bronchitis virus, strain Beaudette, NP_740625. The residue numbering of nsP7 is according to the construct used in this study (see text), with the nsP7 sequence in positions 3 to 85. The sequence positions 3, 12, 22, 32, etc. are labeled (corresponding to residues 1, 10, 20, 30, etc. of the nsP7 sequence).

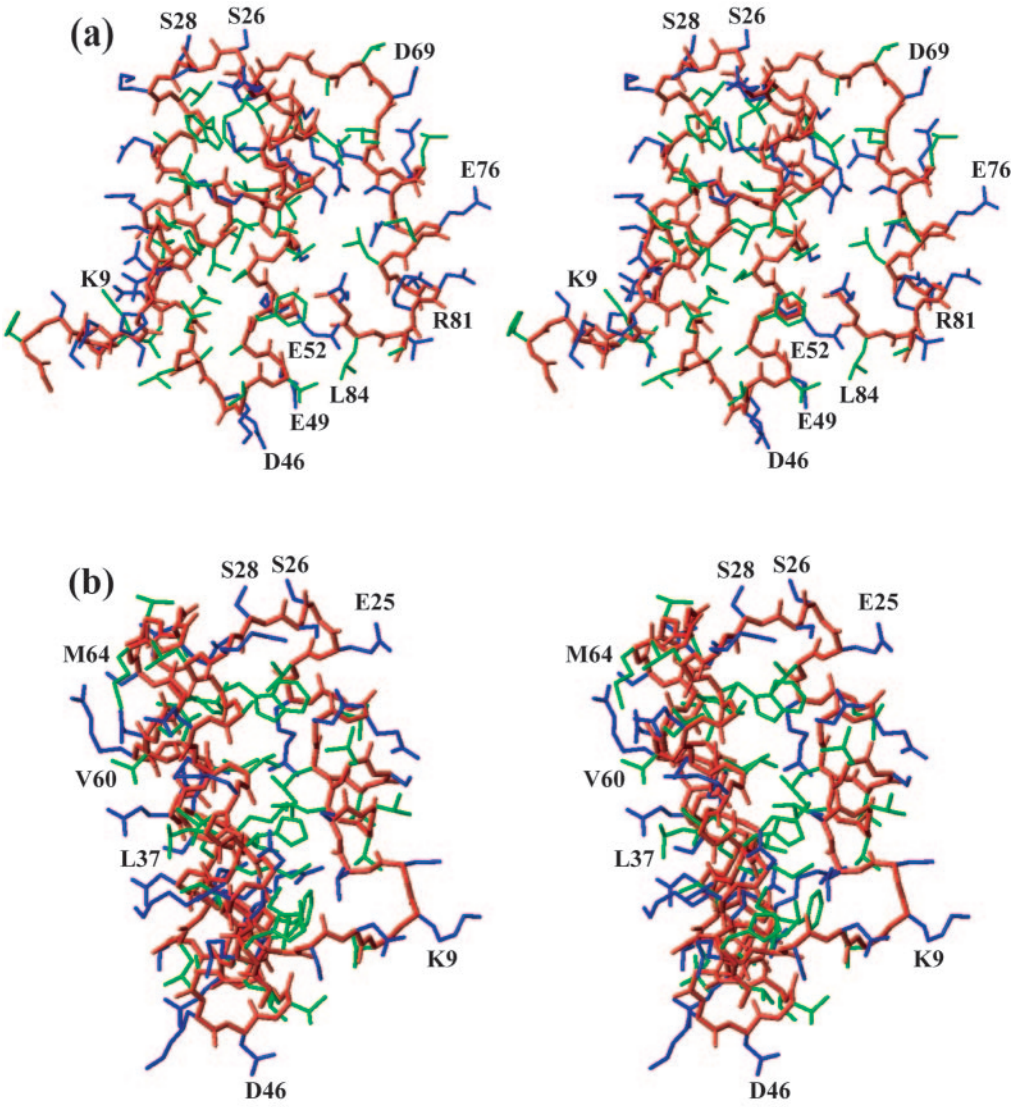

FIG. 4. Wall-eye stereo views of all-heavy-atom presentations of the same conformer of nsP7 as that shown in Fig. 1b. (a) Viewing angle as in Fig. 1b. (b) Viewing angle as in Fig. 1c. Color code: green, hydrophobic side chains; blue, all other side chains; red, polypeptide backbone. Some of the side chains contributing to surface features discussed in the text are identified with the one-letter amino acid code and the residue number.

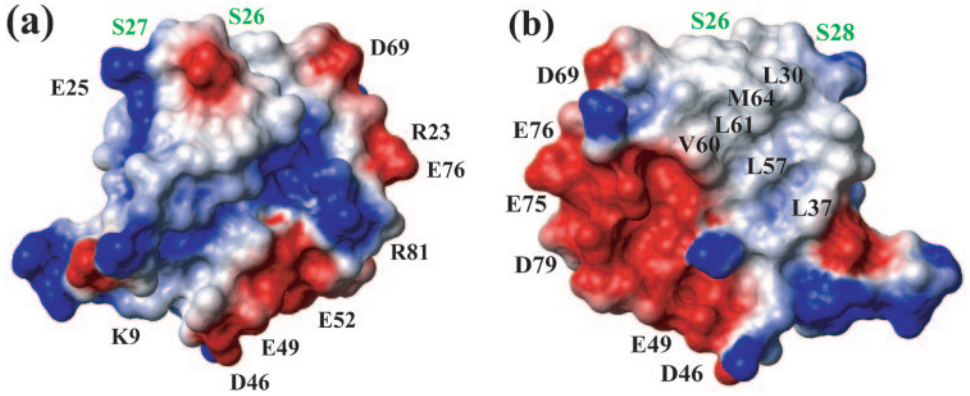

FIG. 5. Surface views of nsP7 in a space-filling presentation. (a) Same orientation as in Fig. 1b. (b) View after a 180° rotation about a vertical axis, showing the surface formed by the flat three-helix sheet. Color code: gray, hydrophobic and polar residues; red, negatively charged; blue, positively charged. Some of the surface side chains discussed in the text are identified with the one-letter amino acid code and the residue number. The residues 26 to 28, which are discussed in the text as a putative functional site, are identified in green.

cessible to interactions with other proteins, and therefore the observed conformation is likely to be assumed only after release from the polyprotein.

In an inspection of the presentation of nsP7 shown in Fig. 4, the aforementioned helix-helix interactions can quite readily be observed. Furthermore, a space-filling surface presentation reveals that the residues Lys 9, Arg 23, and Arg 81 lend positive charge to one face of the protein, including the groove between helix α1 and the helical sheet (Fig. 5a). On the opposite side, the flat surface formed by the three-helix sheet of α2 to α4 is divided nearly evenly into negatively charged and hydrophobic patches (Fig. 5b). The negatively charged surface areas contain the side chains of the residues Asp 46, Glu 49, Asp 69, Glu 75, Glu 76, and Asp 79. A large hydrophobic patch is formed by the partly or completely surface-exposed side chains of the residues Leu 30, Leu 37, Ile 41, Leu 57, Val 60, Leu 61, and Met 64 in the helices α2 and α3, some of which also participate in the interhelix interactions described above. Both areas would seem to be potential sites for protein-protein interactions.

Studies of other coronaviruses have demonstrated the involvement of nsP7 in viral replicase complexes and in specific interactions with other nonstructural proteins (2, 4, 44). Given the lack of functional information regarding nsP7, we employed bioinformatics techniques to search for possible functional sites. The serine residues 26 to 28 in the exposed loop connecting the helices α1 and α2 were thus identified as a likely functional site by the ConSurf algorithm (6), based on strong sequence conservation and surface exposure. These three residues were also identified as part of known active sites by a search with the PINTS server (36); however, other surface features were not similar enough to infer a unique function. This apparent failure to relate nsP7 with functional properties of related proteins leaves us at present with the possibility that the unique sequence and structure of nsP7 are the basis for an as-yet-unrecognized, novel functional role unique to the *Coronaviridae*.

ACKNOWLEDGMENTS

This study was supported by the NIAID/NIH contract HHSN 266200400058C, "Functional and Structural Proteomics of the SARS-CoV," to P.K. and M.J.B. W.P. was a Max-Kade Foundation scholar; M.A.J. is supported by a Canadian Institutes of Health Research postdoctoral fellowship and by the Skaggs Institute for Chemical Biology at TSRI. K.W. is the Cecil H. and Ida M. Green Professor of Structural Biology at TSRI and a member of the Skaggs Institute for Chemical Biology.

The Joint Center for Structural Genomics is supported by National Institute of General Medical Sciences (NIGMS) grant GM062411 as part of the Protein Structure Initiative of the National Institutes of Health.

We thank Kin Moy and Jeffrey Velasquez of the Joint Center for Structural Genomics for help with the cloning of the SARS-CoV proteome and Ian A. Wilson for a critical reading of the manuscript.

REFERENCES

1. **Alcami, A., and U. H. Koszinowski.** 2000. Viral mechanisms of immune evasion. Immunol. Today **21:**447–455.
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. **Bowie, A. G., J. Zhan, and W. L. Marshall.** 2004. Viral appropriation of apoptotic and NF-κB signaling pathways. J. Cell. Biochem. **91:**1099–1108.
4. **Brockway, S. M., X. T. Lu, T. R. Peters, T. S. Dermody, and M. R. Denison.** 2004. Intracellular localization and protein interactions of the gene 1 protein p28 during mouse hepatitis virus replication. J. Virol. **78:**11551–11562.
5. **Cornell, W. D., P. Cieplak, C. I. Bayly, I. R. Gould, K. M. Merz, Jr., D. M. Ferguson, D. C. Spellmeyer, T. Fox, J. W. Caldwell, and P. A. Kollman.** 1995. A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. J. Am. Chem. Soc. **117:**5179–5197.
6. **Glaser, F., T. Pupko, I. Paz, R. E. Bell, D. Bechor-Shental, E. Martz, and N. Ben-Tal.** 2003. ConSurf: identification of functional regions in proteins by surface-mapping of phylogenetic information. Bioinformatics **19:**163–164.
7. **Güntert, P.** 1998. Structure calculation of biological macromolecules from NMR data. Q. Rev. Biophys. **31:**145–237.
8. **Güntert, P., and K. Wüthrich.** 2001. Sampling of conformation space in torsion angle dynamics calculations. Comp. Phys. Commun. **138:**155–169.
9. **Güntert, P., C. Mumenthaler, and K. Wüthrich.** 1997. Torsion angle dynamics for NMR structure calculation with the new program DYANA. J. Mol. Biol. **273:**283–298.
10. **Herrmann, T., P. Güntert, and K. Wüthrich.** 2002. Protein NMR structure determination with automated NOE assignment using the new software CANDID and the torsion angle dynamics algorithm DYANA. J. Mol. Biol. **319:**209–227.
11. **Herrmann, T., P. Güntert, and K. Wüthrich.** 2002. Protein NMR structure determination with automated NOE-identification in NOESY spectra using the new software ATNOS. J. Biomol. NMR **24:**171–189.
12. **Holm, L., and C. Sander.** 1993. Protein structure comparison by alignment of distance matrices. J. Mol. Biol. **233:**123–138.
13. **Ivanov, K. A., V. Thiel, J. C. Dobbe, Y. van der Meer, E. J. Snijder, and J. Ziebuhr.** 2004. Multiple enzymatic activities associated with severe acute respiratory syndrome coronavirus helicase. J. Virol. **78:**5619–5632.
14. **Koradi, R., M. Billeter, and P. Güntert.** 2000. Point-centered domain decomposition for parallel molecular dynamics simulation. Comp. Phys. Commun. **124:**139–147.
15. **Koradi, R., M. Billeter, and K. Wüthrich.** 1996. MOLMOL: a program for display and analysis of macromolecular structures. J. Mol. Graphics **14:**51–55.
16. **Laskowski, R. A., M. W. MacArthur, D. S. Moss, and J. M. Thornton.** 1993. Main-chain bond lengths and bond angles in protein structures. J. Appl. Crystallogr. **26:**283–291.
17. **Lesley, S. A., P. Kuhn, A. Godzik, A. M. Deacon, I. Mathews, A. Kreusch, G. Spraggon, H. E. Klock, D. McMullan, T. Shin, J. Vincent, A. Robb, L. S. Brinen, M. D. Miller, T. M. McPhillips, M. A. Miller, D. Scheibe, J. M. Canaves, C. Guda, L. Jaroszewski, T. L. Selby, M. A. Elsliger, J. Wooley, S. S. Taylor, K. O. Hodgson, I. A. Wilson, P. G. Schultz, and R. C. Stevens.** 2002. Structural genomics of the *Thermotoga maritima* proteome implemented in a high-throughput structure determination pipeline. Proc. Natl. Acad. Sci. USA **99:**11664–11669.
18. **Luginbühl, P., P. Güntert, M. Billeter, and K. Wüthrich.** 1996. The new program OPAL for molecular dynamics simulations and energy refinements of biological macromolecules. J. Biomol. NMR **8:**136–146.
19. **Luginbühl, P., T. Szyperski, and K. Wüthrich.** 1995. Statistical basis for the use of $^{13}C^{\alpha}$ chemical shifts in protein structure determination. J. Magn. Reson. **109:**229–233.
20. **Morris, A. L., M. W. MacArthur, E. G. Hutchinson, and J. M. Thornton.** 1992. Stereochemical quality of protein structure coordinates. Proteins **12:**345–364.
21. **Murzin, A. G., S. E. Brenner, T. Hubbard, and C. Chothia.** 1995. SCOP: a structural classification of proteins database for the investigation of sequences and structures. J. Mol. Biol. **247:**536–540.
22. **Navas, S., and S. R. Weiss.** 2003. Murine coronavirus-induced hepatitis: JHM genetic background eliminates A59 spike-determined hepatotropism. J. Virol. **77:**4972–4978.
23. **Ng, L. F. P., H. Y. Xu, and D. X. Liu.** 2001. Further identification and characterization of products processed from the coronavirus avian infectious bronchitis virus (IBV) 1a polyprotein by the 3C-like proteinase. Adv. Exp. Med. Biol. **494:**291–298.
24. **Nilges, M.** 1997. Ambiguous distance data in the calculation of NMR structures. Folding Design **2:**S53–S57.
25. **Oakley, M. G., and J. J. Hollenbeck.** 2001. The design of antiparallel coiled coils. Curr. Opin. Struct. Biol. **11:**450–457.
26. **Orengo, C. A., A. D. Michie, S. Jones, D. T. Jones, M. B. Swindells, and J. M. Thornton.** 1997. CATH—a hierarchic classification of protein domain structures. Structure **5:**1093–1108.
27. **O'Shea, E. K., J. D. Klemm, P. S. Kim, and T. Alber.** 1991. X-ray structure of the GCN4 leucine zipper, a two-stranded, parallel coiled coil. Science **254:**539–544.
28. **Page, R., K. Moy, E. C. Sims, J. Velasquez, B. McManus, C. Grittini, T. L. Clayton, and R. C. Stevens.** 2004. Scalable high-throughput micro-expression device for recombinant proteins. BioTechniques **37:**364–368.
29. **Page, R., W. Peti, I. A. Wilson, R. C. Stevens, and K. Wüthrich.** 2005. NMR screening and crystal quality of bacterially expressed prokaryotic and eukaryotic proteins in a structural genomics pipeline. Proc. Natl. Acad. Sci. USA **102:**1901–1905.
30. **Peti, W., T. Etezady-Esfarjani, T. Herrmann, H. E. Klock, S. A. Lesley, and**

K. Wüthrich. 2004. NMR for structural proteomics of *Thermotoga maritima*: screening and structure determination. J. Struct. Funct. Genomics **5:**205–215.

31. **Prentice, E., J. McAuliffe, X. Lu, K. Subbarao, and M. R. Denison.** 2004. Identification and characterization of severe acute respiratory syndrome coronavirus replicase proteins. J. Virol. **78:**9977–9986.
32. **Sattler, J., J. Schleucher, and C. Griesinger.** 1999. Heteronuclear multidimensional NMR experiments for the structure determination of proteins in solution employing pulsed field gradients. Prog. NMR Spectr. **34:**93–158.
33. **Seybert, A., C. C. Posthuma, L. C. van Dinten, E. J. Snijder, A. E. Gorbalenya, and J. Ziebuhr.** 2005. A complex zinc finger controls the enzymatic activities of nidovirus helicases. J. Virol. **79:**696–704.
34. **Snijder, E. J., P. J. Bredenbeek, J. C. Dobbe, V. Thiel, J. Ziebuhr, L. L. Poon, Y. Guan, M. Rozanov, W. J. Spaan, and A. E. Gorbalenya.** 2003. Unique and conserved features of genome and proteome of SARS-coronavirus, an early split-off from the coronavirus group 2 lineage. J. Mol. Biol. **331:**991–1004.
35. **Spera, S., and A. Bax.** 1991. Empirical correlation between protein backbone conformation and C^{α} and C^{β} ^{13}C nuclear magnetic resonance chemical shifts. J. Am. Chem. Soc. **113:**5490–5492.
36. **Stark, A., S. Sunyaev, and R. B. Russell.** 2003. A model for statistical significance of local similarities in structure. J. Mol. Biol. **326:**1307–1316.
37. **Tan, Y.-J., B. C. Fielding, P.-Y. Goh, S. Shen, T. H. P. Tan, S. G. Lim, and W. Hong.** 2004. Overexpression of 7a, a protein specifically encoded by the severe acute respiratory syndrome coronavirus, induces apoptosis via a caspase-dependent pathway. J. Virol. **78:**14043–14047.
38. **Thiel, V., K. A. Ivanov, A. Putics, T. Hertzig, B. Schelle, S. Bayer, B. Weissbrich, E. J. Snijder, H. Rabenau, H. W. Doerr, A. E. Gorbalenya, and J. Ziebuhr.** 2003. Mechanisms and enzymes involved in SARS coronavirus genome expression. J. Gen. Virol. **84:**2305–2315.
39. **Tijms, M. A., and E. J. Snijder.** 2003. Equine arteritis virus non-structural protein 1, an essential factor for viral subgenomic mRNA synthesis, interacts with the cellular transcription co-factor p100. J. Gen. Virol. **84:**2317–2322.
40. **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.
41. **Wüthrich, K.** 1986. NMR of proteins and nucleic acids. Wiley, New York, N.Y.
42. **Wüthrich, K.** 2003. NMR studies of structure and function of biological macromolecules. J. Biomol. NMR **27:**13–39.
43. **Ye, Y., and A. Godzik.** 2003. Flexible structure alignment by chaining aligned fragment pairs allowing twists. Bioinformatics **19**(Suppl. 2)**:**II246–II255.
44. **Ziebuhr, J., and S. G. Siddell.** 1999. Processing of the human coronavirus 229E replicase polyproteins by the virus-encoded 3C-like proteinase: identification of proteolytic products and cleavage sites common to pp1a and pp1ab. J. Virol. **73:**177–185.