

## Proteomics Analysis Unravels the Functional Repertoire of Coronavirus Nonstructural Protein 3<sup>∇†</sup>

Benjamin W. Neuman,<sup>1,5‡</sup> Jeremiah S. Joseph,<sup>2‡</sup> Kumar S. Saikatendu,<sup>2</sup> Pedro Serrano,<sup>3</sup>  
Amarnath Chatterjee,<sup>3</sup> Margaret A. Johnson,<sup>3</sup> Lujian Liao,<sup>2</sup> Joseph P. Klaus,<sup>1</sup>  
John R. Yates III,<sup>2</sup> Kurt Wüthrich,<sup>3,4</sup> Raymond C. Stevens,<sup>3</sup>  
Michael J. Buchmeier,<sup>1\*</sup> and Peter Kuhn<sup>2,3\*</sup>

*Molecular and Integrative Neurosciences Department,<sup>1</sup> Department of Cell Biology,<sup>2</sup> Department of Molecular Biology,<sup>3</sup>  
and Skaggs Institute of Chemical Biology,<sup>4</sup> The Scripps Research Institute, 10550 N. Torrey Pines Road, La Jolla,  
California 92037, and School of Biological Sciences, University of Reading, Whiteknights,  
RG6 6AJ Reading, United Kingdom<sup>5</sup>*

Received 11 December 2007/Accepted 16 March 2008

Severe acute respiratory syndrome (SARS) coronavirus infection and growth are dependent on initiating signaling and enzyme actions upon viral entry into the host cell. Proteins packaged during virus assembly may subsequently form the first line of attack and host manipulation upon infection. A complete characterization of virion components is therefore important to understanding the dynamics of early stages of infection. Mass spectrometry and kinase profiling techniques identified nearly 200 incorporated host and viral proteins. We used published interaction data to identify hubs of connectivity with potential significance for virion formation. Surprisingly, the hub with the most potential connections was not the viral M protein but the nonstructural protein 3 (nsp3), which is one of the novel virion components identified by mass spectrometry. Based on new experimental data and a bioinformatics analysis across the *Coronaviridae*, we propose a higher-resolution functional domain architecture for nsp3 that determines the interaction capacity of this protein. Using recombinant protein domains expressed in *Escherichia coli*, we identified two additional RNA-binding domains of nsp3. One of these domains is located within the previously described SARS-unique domain, and there is a nucleic acid chaperone-like domain located immediately downstream of the papain-like proteinase domain. We also identified a novel cysteine-coordinated metal ion-binding domain. Analyses of interdomain interactions and provisional functional annotation of the remaining, so-far-uncharacterized domains are presented. Overall, the ensemble of data surveyed here paint a more complete picture of nsp3 as a conserved component of the viral protein processing machinery, which is intimately associated with viral RNA in its role as a virion component.

The severe acute respiratory syndrome coronavirus (SARS-CoV) is an enveloped virus with a 29.7-kb positive-strand RNA genome (35). Replication of this genome and transcription are mediated by a large membrane-anchored RNA processing complex. Components of this complex are derived from the 16 nonstructural proteins (nsp1 to nsp16) that are processed from the open reading frame 1a (ORF1a) and ORF1b. The polyprotein 1a (pp1a) is translated from ORF1a, while the polyprotein 1ab (pp1ab) is formed by a  $-1$  ribosomal frameshift upstream of the ORF1a stop codon, causing read-through into ORF1b. SARS-CoV encodes two proteinases, a “main proteinase” (nsp5) and a papain-like proteinase (PL2<sup>PRO</sup> domain of nsp3). These two proteins proteolytically cleave pp1a and pp1ab into the 16 mature nsp’s (61). Specifically, SARS-CoV PL2<sup>PRO</sup> cleaves pp1a at the three sites <sub>177</sub>LNGG ↓ AVT<sub>183, 815</sub>LKGG

↓ API<sub>821</sub>, and <sub>2737</sub>LKGG ↓ KIV<sub>2743</sub> to release nsp1, nsp2, and nsp3, respectively.

In current coronavirus terminology, the term “nonstructural protein” usually refers to peptides processed from pp1a and pp1ab, while “structural protein” refers to the N, M, S, and E proteins, which interact to coordinate the structure of the virion lipidic envelope (39). The term “accessory protein” refers to group- or subgroup-specific proteins, some of which may be incorporated in virions. A typical virion may contain the viral RNA genome, plus tens to hundreds of copies of N, M, and S proteins; a few E proteins (16); and an unknown but presumably small quantity of accessory proteins such as the SARS-CoV ORF3a (22), ORF6 (21), ORF7a (20), and ORF7b (51) proteins. Incorporation of the accessory ORF9b protein can be inferred from incorporation of the homologous I protein of murine hepatitis virus (MHV) (13). Furthermore, our recent electron cryomicroscopy (cryo-EM) analysis of coronavirus ultrastructure (39) revealed that the viral ribonucleoprotein is sufficiently loosely packed in the virion core to leave ample space for possible additional incorporation of host proteins (56).

Here we used mass spectrometry proteomics and protein kinase profiling techniques to probe the contents of purified SARS-CoV virions. We investigated cellular pathways involved

\* Corresponding author. Mailing address: 10550 N. Torrey Pines Road, SP30-2020, The Scripps Research Institute, La Jolla, CA 92037. Phone: (858) 784-9114. Fax: (858) 784-8996. E-mail for Michael J. Buchmeier: buchm@scripps.edu. E-mail for Peter Kuhn: pkuhn@scripps.edu.

† Supplemental material for this article may be found at <http://jvi.asm.org/>.

‡ These authors contributed equally to this paper.

∇ Published ahead of print on 26 March 2008.

TABLE 1. Recombinant expression of nsp3 domains

Construct	Boundary positions	Vector	Tag	<i>E. coli</i> strain	Reference
UB1	1–112	pET25b	Tagless	BL21(DE3)RIL	53
UB1-AC	1–183	pET25b	Tagless	BL21(DE3)RIL	53
ADRP	184–365	pMH1F	His tag	DL41	47
SUD	389–726	pET28aTEV	Cleavable His tag	Rosetta pLysS	
SUD <sub>451–651</sub>	451–651	pET28aTEV	Cleavable His tag	Rosetta pLysS	
SUD-C	513–651	pET28b	Cleavable His tag	BL21(DE3)	7
UB2-PL2 <sup>pro</sup>	723–1037	pET11a	Tagless	BL21(DE3)	45
NAB	1066–1225	pMH1F	His tag	DL41	

in coronavirus assembly, and we expected our experimental approach to identify novel virus component-host protein interactions important to virogenesis. We attempted to bias the analysis toward identification of biologically significant host proteins by subtracting proteins purified from uninfected cells, proteins identified with only one sample preparation method, and proteins occurring only on the proteolytically sensitive surface of the virion during the analysis. One hundred seventy-two host proteins and eight viral proteins meeting these criteria are described here, including three nsp's. Network analysis (2) based on previously reported biochemical interaction mapping (65) revealed several hubs of connectivity (we use the term "hub of connectivity" or "hub" to refer to molecular species showing an outstandingly large number of intermolecular interactions) among incorporated components of viral origin. Among the hubs with the most connections are the viral M protein, the RNA genome, and nsp3. The M protein links the other major virion components at the site of budding (35), and an integral role for the RNA genome in assembly had been anticipated (15). nsp3, however, which is the protein capable of making the most connections to other virus-encoded components of the virion, had not previously been implicated in coronavirus assembly. We therefore selected nsp3 for further functional and structural characterization.

SARS-CoV nsp3 is a large multidomain protein that includes confirmed proteinase and poly(ADP-ribose) binding domains. We present here an updated nsp3 phylogeny and domain map, including novel validated metal ion-binding and nucleic acid-binding domains. We also describe the use of relative conservation data to infer functional information for the remaining uncharacterized nsp3 domains. We interpret these data in light of recent functional and structural characterizations of nsp3 domains (45, 47, 53), which leads us to suggest an important role for nsp3 in coronavirus RNA synthesis and virogenesis.

#### MATERIALS AND METHODS

**SARS-CoV growth, purification, and treatment.** SARS-CoV Tor2 was cultured in Vero-E6 cells, which are derived from the African green, or vervet, monkey *Cercopithecus aethiops*. Vero-E6 cells were selected for high viral growth rate and reproducibility of infection. Cells were inoculated at a high multiplicity (~1 to 3 PFU/cell), medium was exchanged after 24 h, and high-titer infectious supernatant was collected 48 h after inoculation. Viral supernatants were clarified by centrifugation at 12,000 × g for 30 min, collected by precipitation with 8% polyethylene glycol 8000 and 2% NaCl, and banded at 140,000 × g for 1.5 h on discontinuous five-step 10% to 50% sucrose gradients. Purified native virus was collected by side puncture and pelleted through HEPES-buffered 0.9% saline (pH 7.0). At this point, aliquots representing virus purified from about 1 liter of infectious supernatant were treated with 5,000 U DNase I (New England Biolabs) for 1 h at 37°C in the supplied DNase I buffer to remove any adherent host

chromatin and associated proteins, followed by 60 mg proteinase K (New England Biolabs) for 1 h at 37°C. Proteinase K treatments were not performed in the presence of a detergent in order to preserve the integrity of the viral membrane. Proteinase K was then removed by pelleting virus through a 30% sucrose cushion. Native and enzymatically treated virus preparations were lysed and inactivated with 1% Triton X-100 (for kinase assays), followed by boiling for 5 min (for mass spectrometry). The concentration of detergent was reduced by pelleting denatured protein aggregates through HEPES-buffered saline.

Infectious SARS-CoV in this study was purified by density gradient banding. Banded viruses are expected to be more pure than viruses purified by pelleting through a discontinuous 10 to 30% sucrose cushion, as was done in our previous cryo-EM study of SARS-CoV supramolecular architecture (39). Analysis of a representative portion of that set of cryo-EM images containing 1,018 enveloped particles from pelleted SARS-CoV revealed 42 particles not visibly recognizable as SARS-CoV (4% of the total) and eight apparently empty vesicles (1%), which are not expected to contribute a significant amount of protein to the mass spectrometry analysis. The purity of the SARS-CoV used for mass spectrometry and kinase analysis was therefore estimated to be greater than or equal to 95%.

**Protein construct design, cloning, expression, and purification.** SARS-CoV nsp3 (GenBank accession number NP\_828862) extends from nucleotides 2719 to 8484, corresponding to residues Ala907 to Gly2828 of pp1a. A summary of selected nsp3 expression constructs and conditions is shown in Table 1. Expression of several nsp3 domains has been described previously (3, 47, 53). The UB2-PL2<sup>pro</sup> expression construct was a kind gift from Andrew Mesecar (University of Chicago—Illinois). All other constructs were amplified by PCR from genomic cDNA of the SARS-CoV Tor-2 strain. Amplification primers were designed to produce the constructs listed in Table 1. Amplicons were cloned into the expression vectors pMH1F (N-terminal His<sub>6</sub>Thio<sub>6</sub> tag; derivative of pBAD from Invitrogen), pET25b (tagless construct), pET28b (thrombin-cleavable N-terminal His<sub>6</sub> tag), or pET28aTEV (tobacco etch virus protease-cleavable N-terminal His<sub>6</sub> tag).

For expression of all constructs in Table 1 except SUD-C and UB1, a sequence-verified clone was transformed into *Escherichia coli*, and an overnight culture from a fresh transformant was used to inoculate flasks of LB medium containing antibiotic. Cultures were grown at 37°C with vigorous shaking to an optical density at 600 nm of 0.6 to 0.8, induced as needed, and grown at 14°C overnight. Bacteria were harvested by centrifugation and lysed by sonication in a buffer containing 50 mM potassium phosphate, pH 7.8, 300 mM NaCl, 10% glycerol, 5 mM imidazole, 0.5 mg/ml lysozyme, 100 μl/liter benzonase, and EDTA-free protease inhibitor (Roche; one tablet per 50 ml buffer). The lysate was clarified by ultracentrifugation at 45,000 rpm for 20 min at 4°C, and the soluble fraction was applied onto a metal chelate column (Talon resin charged with cobalt; Clontech). The column was washed with a solution containing 20 mM Tris, pH 7.8, 300 mM NaCl, 10% glycerol, and 5 mM imidazole and eluted in buffer containing 25 mM Tris, pH 7.8, 300 mM NaCl, and 150 mM imidazole. The eluate was then purified by anion exchange on a Poros HQ column using a linear gradient of NaCl (0 to 1 M) in 25 mM Tris-HCl, pH 8.0. Tobacco etch virus protease was added to proteins with cleavable tags in a 1:50 molar ratio. After incubation overnight at 4°C, the cleaved tags and uncleaved proteins were captured by a Talon resin column, and the flowthrough was concentrated and further purified by size-exclusion chromatography on a Superdex 75 column equilibrated with 10 mM Tris, pH 7.8, 150 mM NaCl. Pure fractions were concentrated and either used immediately for assays or flash-frozen in liquid nitrogen. SUD-C was produced as described in reference 7. UB1 was produced as described in reference 53.

**Metal ion-binding assay.** Purified proteins were not actively stripped of metal ions before analysis; rather, proteins were selected that did not measurably strip

CoCl<sub>2</sub> from the Talon affinity matrix at the time of purification. Ten-micromolar SUD-C, SUD<sub>451-651</sub>, or full-length SUD solutions were mixed with CoCl<sub>2</sub> to final concentrations ranging from 0 to 50  $\mu$ M Co(II) in a buffer containing 25 mM Tris at pH 7.8 and 300 mM NaCl. Samples were incubated on ice for 30 min, and absorption spectra from 250 to 800 nm were then recorded on a Cary UV-Vis spectrophotometer. Matched baseline spectra from samples containing only buffer and CoCl<sub>2</sub> were subtracted from the absorption spectra of the protein-containing samples. Zn(II) titration was performed by recording optical spectra after addition of ZnCl<sub>2</sub> following incubation with 50  $\mu$ M CoCl<sub>2</sub>.

**Extraction of viral proteins and digestion.** Native SARS-CoV, enzymatically treated SARS-CoV, and host background protein samples were divided into two identical parts, one used for trichloroacetic acid (TCA) precipitation and the other for methanol delipidation. For TCA precipitation, TCA was added to the sample to a final content of 25% (vol/vol). The sample was then placed on ice for 30 min and centrifuged at 13,000  $\times$  g for 5 min. The pellet was twice washed with cold acetone to ready it for the next step. For methanol delipidation, 2.5 volumes of methanol, 0.25 volume of chloroform, and 0.5 volume of water were added. The sample was then centrifuged at 16,000  $\times$  g for 2 min, and the organic layer was removed. After back extraction with 3 volumes of methanol, the sample was centrifuged at 16,000  $\times$  g for 2 min to obtain a pellet. The resulting pellets from the two extraction conditions were separately solubilized in Invitrosol (Invitrogen, Carlsbad, CA), sonicated for 30 min, and reduced with tris(2-carboxyethyl)phosphine, and the cysteines were alkylated with iodoacetamide. Acetonitrile was then added to a final content of 80% (vol/vol). Finally, the sample was digested with trypsin (enzyme/substrate ratio of 1:50 [wt/wt]) at 37°C overnight.

**Mass spectrometry analysis of viral proteins.** The protein digest from each sample was analyzed by Multidimensional Protein Identification Technology (MudPIT) (69). Briefly, digested proteins were pressure loaded onto a fused silica capillary column packed with a 3-cm, 5- $\mu$ m Partisphere strong cation exchanger (SCX; Whatman, Clifton, NJ) and 3-cm, 5- $\mu$ m Aqua C<sub>18</sub> material (RP; Phenomenex, Ventura, CA), with a 2- $\mu$ m filter union (UpChurch Scientific, Oak Harbor, WA) attached to the SCX end. The column was washed with buffer containing 94.9% water, 5% acetonitrile, and 0.1% formic acid. After desalting, a 100- $\mu$ m-inside-diameter capillary with a 5- $\mu$ m pulled tip packed with 10-cm, 3- $\mu$ m Aqua C<sub>18</sub> material was attached to the filter union, and the entire split column was placed in line with an Agilent 1100 quaternary high-pressure liquid chromatograph (Agilent, Palo Alto, CA) and analyzed using a modified 11-step separation (66). Three buffer solutions were used: 5% acetonitrile-0.1% formic acid (buffer A), 80% acetonitrile-0.1% formic acid (buffer B), and 500 mM ammonium acetate-5% acetonitrile-0.1% formic acid (buffer C). The first step consisted of a 100-min gradient from 0 to 100% buffer B. Steps 2 to 10 had the following profile: 3 min of 100% buffer A, 5 min of X% buffer C, a 10-min gradient from 0 to 15% buffer B, and a 97-min gradient from 15 to 45% buffer B. The 5-min buffer C percentages (X) were 5, 10, 15, 20, 25, 30, 40, 55, and 75%, respectively. In the final step, the gradient contained 3 min of 100% buffer A, 20 min of 100% buffer C, a 10-min gradient from 0 to 15% buffer B, and a 107-min gradient from 15 to 100% buffer B. As peptides were eluted from the microcapillary column, they were electrosprayed directly into an LTQ linear ion trap mass spectrometer (ThermoFinnigan, San Jose, CA) with the application of a distal 2.4-kV spray voltage. A cycle of one full-scan mass spectrum (400 to 1,400 *m/z*) followed by five data-dependent tandem mass spectrometry (MS/MS) spectra at a 35% normalized collision energy was repeated continuously throughout each step of the multidimensional separation.

**Processing of mass spectra.** MS/MS spectra were analyzed using the following software analysis protocol. Poor-quality spectra were removed from the data set using an automated spectrum quality assessment algorithm (4). MS/MS spectra remaining after filtering were searched with the SEQUEST algorithm (12) against a combined human, SARS-CoV, and vervet monkey database from NCBI that was concatenated to a decoy database in which the sequence for each entry in the original database was reversed. SEQUEST results were assembled and filtered using the DTASelect program (60) with a peptide false-positive rate of 5%. To increase the probability of identifying viral proteins while simultaneously maintaining reasonably high filtering criteria, proteins with one peptide hit were accepted, but we required all peptides identified to be fully tryptic.

**Bioinformatics analysis.** An initial multiple sequence alignment was produced using NCBI BLAST (1) to identify homologous regions and then Clustal to align the homologous regions (8). The initial alignment was manually fine tuned to reflect (in hierarchical order) solved coronavirus protein structures, conserved cysteine and histidine residues, TMHMM2 transmembrane region prediction (30), and structure/loop context from PredictProtein analysis (46). Annotations and region boundaries displayed here were derived from published analysis by Gorbalenya et al. (18), de novo SARS-CoV-specific domain structure-prediction

(24), and a combination of domain expression and nuclear magnetic resonance screening for foldedness.

The following sequences were used for nsp3 alignments in Fig. 2A and S2 in the supplemental material: group Ia, HCoV-NL63 (YP\_003766), HCoV-229E (NP\_073549), PEDV (AAK38661), BtCoV 512/2005 (ABG47077); group Ib, transmissible gastroenteritis virus (TGEV) (NP\_058422), PRCoV (ABG89316), FCoV (YP\_239353); group IIa, HCoV-HKU1-A (YP\_173236), HCoV-HKU1-N6 (ABD75567), MHV-JHM (AAA46457), MHV-A59 (NP\_068668), BCoV (NP\_150073), HCoV-OC43 (NP\_937947), HEV (YP\_459949); group IIb, SARS-CoV (AAP41036), BtCoV-HKU3 (AAAY88865), BtCoV-Rf1 (ABD75321); group IIc, BtCoV-HKU5 (ABN10892), BtCoV 133/2005 (YP\_729202); group IId, BtCoV-HKU9-1 (YP\_001039970), BtCoV-HKU9-2 (ABN10918), BtCoV-HKU9-3 (ABN10926), BtCoV-HKU9-4 (ABN10934); group III, IBV-Beaudette (NP\_066134), IBV-Peafowl/GD/KQ6/2003 (AAT70073), IBV-LX4 (AAQ21583), IBV-BJ (AAP2673); torovirus group (aligned from ADP-ribose-1'-phosphatase [ADRP] onward), ETov (ABC26008), BTov (YP\_337905). The alignment presented in Fig. 2C and analysis in Fig. 8 include HCoV-229E, HCoV-NL63, BtCoV 512/2005, FCoV, HCoV-HKU1, MHV-A59, HCoV-OC43, SARS-CoV Tor2, BtCoV 133/2005, BtCoV-HKU5, BtCoV-HKU9-1, BtCoV-HKU9-4, IBV-Beaudette, and IBV-Peafowl/GD/KQ6/2003 sequences listed under or linked from the accession numbers above.

**Kinase array analysis.** A full PepChip protein kinase substrate usage profiling assay (Pepscan Systems, Lelystad, Netherlands) was performed according to the manufacturer's instructions. Briefly, purified native SARS-CoV was lysed by trituration in a protease inhibitor cocktail containing 1% Triton X-100. SARS-CoV lysate was applied to duplicate peptide substrate arrays in the presence of [ $\gamma$ -<sup>33</sup>P]ATP. The labeled substrate array was visualized by autoradiography, digitally scanned, and quantified using ImageJ densitometry software (NIH). Duplicate PepChip arrays incorporated a total of 48 nonsubstrate peptides, which were used as negative controls to determine the background levels in the densitometry analysis. Density values for these spots were used to assess and filter results. Peptides for which both replicate spots exceeded the mean density value plus 1 standard deviation on the controls on the scanned autoradiograph were taken as positive results.

**Protein stoichiometry analysis.** A detailed description and validation of perfluoro-octanoic acid (PFO)-polyacrylamide gel electrophoresis (PAGE) as a tool for protein stoichiometry assessment can be found elsewhere (44). Briefly, purified protein samples were incubated at 37°C for 1 h; mixed 1:1 with PFO loading buffer containing 8% (wt/vol) PFO, 100 mM Tris base, 20% (vol/vol) glycerol, and 0.05% (wt/vol) orange G; and loaded onto precast 4 to 20% Tris-glycine gradient gels. Gel electrophoresis was performed with a standard Tris-glycine running buffer to which 0.5% (wt/vol) PFO was added. Protein was detected by SYPRO-ruby poststain (Invitrogen).

**Electrophoretic mobility shift and unwinding assays.** For electrophoretic mobility shift assay (EMSA), protein samples were mixed with 0.8  $\mu$ g of RNA or DNA substrate and assay buffer containing 150 mM NaCl-50 mM Tris at pH 8.0 to a total reaction volume of 20  $\mu$ l. Sequence-matched RNA and DNA oligomers were designed (substituting T for U as appropriate) with randomized sequences designed to adopt single-stranded conformations: ssRNA1/ssDNA1, 5'-AAAU ACCUCUAAAAUAACACCACCAUAUACCACAU-3', and ssRNA2/ssDNA2, 5'-AGUCAGUCAGUCAGUCAGUCAGUCAGUCAGUCAGUCAGUCAGUCAGUC-3'. Double-stranded RNA (dsRNA) and DNA (dsDNA) were produced by boiling and slowly cooling equimolar mixtures of single-stranded RNA (ssRNA) or DNA (ssDNA) (substituting T for U) oligomers, 5'-GAAAGGAA AAAGGGAGAAGA-3' and 5'-UCUUCUCCUUUUCCUUUC-3'. Protein-nucleic acid mixtures were incubated at 37°C for 1 h and analyzed by native electrophoresis on precast 6% acrylamide DNA retardation gels (Invitrogen). Nucleic acid was detected by SYBR-gold poststain (Invitrogen) and photographed using a UV light source equipped with a digital camera. SYBR-gold was rinsed out and protein was subsequently detected by SYPRO-ruby poststain (Invitrogen). Densitometry analysis was performed using a flatbed scanner with ImageJ software (NIH). The mobility shift of RNA at each protein concentration was calculated relative to the maximum shift observed in each experiment. *K<sub>d</sub>* (dissociation constant) values were measured from the midpoints of the fitted titration data.

For unwinding assays, nucleic acid and protein mixtures were prepared and incubated as described above for the EMSAs. Instead of applying the samples immediately to polyacrylamide gels, samples were incubated at 4°C overnight to allow protein-nucleic acid complexes to dissociate before native PAGE analysis. Results were visualized and recorded as for EMSA.

TABLE 2. Background host proteins excluded from this analysis

No. of times found (of two possible times) in sample type:			No. of peptides <sup>d</sup>		% Coverage <sup>e</sup>	Description <sup>f</sup>	Abbreviation <sup>g</sup>
Background <sup>a</sup>	Nat <sup>b</sup>	PK <sup>c</sup>	Unique	Total			
2	2	2	5	9	22.90	Actin beta ( <i>Cercopithecus aethiops</i> )	ACTB
1	1	0	3	4	9.90	Actin kappa ( <i>Homo sapiens</i> )	FKSG30
1	2	0	2	2	6.10	Adenylyl cyclase-associated protein 1 ( <i>Homo sapiens</i> )	CAP1
1	1	0	2	2	1.20	Agrin ( <i>Homo sapiens</i> )	AGRN
2	1	0	5	7	4.40	Alpha 1 type VI collagen ( <i>Homo sapiens</i> )	COL6A1
1	1	0	2	2	1.00	Alpha 1 type VII collagen ( <i>Homo sapiens</i> )	COL7A1
1	1	0	5	5	4.10	Alpha 1 type XII collagen ( <i>Homo sapiens</i> )	COL12A1
2	1	0	5	11	2.90	Alpha-2-macroglobulin precursor ( <i>Homo sapiens</i> )	A2M
1	1	0	9	16	24.70	Amylase ( <i>Homo sapiens</i> )	AMY
2	1	0	5	5	1.50	Apolipoprotein B precursor ( <i>Homo sapiens</i> )	APOB
2	1	0	5	5	27.70	Apolipoprotein E precursor ( <i>Cercopithecus aethiops</i> )	APOE
2	1	0	3	6	1.20	Chondroitin sulfate proteoglycan 2 ( <i>Homo sapiens</i> )	VCAN
2	2	0	4	11	3.60	Complement component 3 precursor ( <i>Homo sapiens</i> )	C3
2	1	0	7	18	2.60	Complement component 4A/4B ( <i>Homo sapiens</i> )	C4A/C4B
1	0	0	2	3	0.80	Complement component 5 ( <i>Homo sapiens</i> )	C5
1	2	0	2	2	7.10	Enolase 1 ( <i>Homo sapiens</i> )	ENO1
1	1	0	2	2	0.90	Fibrillin 1 ( <i>Homo sapiens</i> )	FBN1
2	2	2	33	107	18.20	Fibronectin 1 ( <i>Cercopithecus aethiops</i> )	FN1
1	1	0	3	3	3.30	Fibulin 1 isoform C ( <i>Cercopithecus aethiops</i> )	FBLN1C
1	1	0	3	7	3.10	Fibulin 1 isoform D ( <i>Homo sapiens</i> )	FBLN1D
2	2	2	8	11	4.90	Filamin 1 ( <i>Homo sapiens</i> )	FLNB
1	1	0	3	3	6.80	Galectin 3 binding protein ( <i>Homo sapiens</i> )	LGALS3BP
2	2	0	7	11	8.20	Gelsolin ( <i>Homo sapiens</i> )	GSN
1	2	1	2	2	5.10	Glyceraldehyde-3-phosphate dehydrogenase, spermatogenic ( <i>Homo sapiens</i> )	GAPDH5
1	2	1	2	2	5.50	Heat shock 70-kDa protein 8 ( <i>Homo sapiens</i> )	HSPA8
1	2	0	2	2	4.50	Heat shock 70-kDa protein 1 or 6 ( <i>Homo sapiens</i> )	HSPA1 or HSPA6
1	2	2	6	29	34.50	Hemoglobin alpha 2 subunit ( <i>Homo sapiens</i> )	HBA2
1	2	1	2	12	6.80	Hemoglobin beta subunit ( <i>Homo sapiens</i> )	HBB
2	2	2	15	21	4.00	Heparan sulfate proteoglycan 2 ( <i>Homo sapiens</i> )	HSPG2
2	2	0	3	5	3.90	Inter-alpha-globulin inhibitor H2 polypeptide ( <i>Homo sapiens</i> )	ITIH2
2	1	0	4	4	2.60	Laminin, alpha 4 precursor ( <i>Homo sapiens</i> )	LAMA4
2	1	0	3	3	1.90	Laminin, beta 1 precursor ( <i>Homo sapiens</i> )	LAMB1
2	1	0	4	5	3.60	Laminin, gamma 1 precursor ( <i>Homo sapiens</i> )	LAMC1
2	1	0	2	2	3.10	Latent transforming growth factor beta binding protein 3 ( <i>Homo sapiens</i> )	LTBP3
1	1	0	2	3	2.10	Latent transforming growth factor beta binding protein 4 ( <i>Homo sapiens</i> )	LTBP4
1	2	2	5	6	3.70	Myosin, heavy polypeptide 9, nonmuscle ( <i>Homo sapiens</i> )	MYH9
1	1	0	2	2	5.60	Neuronal pentraxin I precursor ( <i>Homo sapiens</i> )	NPTX1
1	1	0	2	2	2.00	Nidogen (enactin) ( <i>Homo sapiens</i> )	NID1
1	1	0	2	2	3.50	Olfactory receptor 5, H2 ( <i>Homo sapiens</i> )	OR5H2
1	2	1	2	2	5.70	Plasminogen activator inhibitor 1 ( <i>Cercopithecus aethiops</i> )	SERPINE1
1	1	0	3	7	2.30	Pregnancy-zone protein ( <i>Homo sapiens</i> )	PZP
1	2	2	2	2	5.30	Pyruvate kinase 3 ( <i>Homo sapiens</i> )	PKM2
1	1	0	2	2	3.20	Quiescin Q6 ( <i>Homo sapiens</i> )	QSOX1
2	1	0	5	5	1.80	Reelin ( <i>Homo sapiens</i> )	RELN
1	2	0	2	3	6.50	S-Adenosylhomocysteine hydrolase ( <i>Homo sapiens</i> )	AHCY
1	1	0	3	16	9.30	Serine (or cysteine) proteinase inhibitor F member 1 ( <i>Homo sapiens</i> )	SERPINF1
1	1	0	2	3	6.20	Stem cell growth factor precursor ( <i>Homo sapiens</i> )	CLEC11a
2	2	0	3	5	2.00	Talin 1 ( <i>Homo sapiens</i> )	TLN1
2	1	0	23	58	16.40	Thrombospondin 1 ( <i>Homo sapiens</i> )	THBS1
1	2	2	3	4	11.80	Tubulin alpha ( <i>Homo sapiens</i> )	TUBA
1	0	2	2	6	10.90	Ubiquitin B precursor ( <i>Homo sapiens</i> )	UBB
2	1	0	6	13	10.90	VGF nerve growth factor inducible ( <i>Homo sapiens</i> )	VGF

<sup>a</sup> The term "background" refers to proteins that we were unable to specifically exclude as being copurified with virus.

<sup>b</sup> Nat, native SARS-CoV samples.

<sup>c</sup> PK, DNase I-proteinase K-treated SARS-CoV.

<sup>d</sup> Number of total and nonoverlapping (unique) peptides identified for each protein from the sample yielding the highest percent coverage.

<sup>e</sup> "Coverage" here refers to the percentage of the intact protein length accounted for by unique contributions of the fragments detected by mass spectrometry.

<sup>f</sup> In cases of unambiguous identifications of *Cercopithecus aethiops* proteins and cases in which *C. aethiops* attribution could not be ruled out, the *C. aethiops* sequence is noted. Proteins identified solely from homology to *H. sapiens* homologs are listed as *H. sapiens*. Protein isoforms are noted only where explicitly identified.

<sup>g</sup> We have identified proteins here with standard abbreviations for the corresponding human genes from the NCBI Entrez Gene database (<http://www.ncbi.nlm.nih.gov/sites/entrez>).

TABLE 3. Host proteins identified in purified SARS-CoV grouped by function<sup>a</sup>

Function	Proteins
Proteins related to vesicular trafficking or viral budding.....	CHC1, SNX6, $\alpha$ -COP, $\beta$ -COP, $\gamma$ -COP, ARF4, CDC42
Cytoplasmic and shuttling RNA-binding proteins.....	hnRNP-A1, hnRNP-A2/B1, hnRNP-A3, hnRNP-C, hnRNP-H1, hnRNP-H2, hnRNP-K, hnRNP-L, hnRNP-M, hnRNP-R, hnRNP-U, LRPPRC, PABPC1 or -3, PABPC4, ROD1, PCBP1, PCBP2, PPTB1
Unfolded protein response.....	CCT complex (three subunits), HSP90 (three subunits), HSPB1, HSPD2, PPIA, VCP
Cytoplasmic proteins and proteins of undetermined localization.....	26S proteasome (six subunits), SMIC, ALCY, ADK, AKR1 (two subunits), ASS1, CBR1, CLIC1, CLIC4, CSNK2, EPRS, FASN, GARS, GART, GDI1 or -2, GNB2L1, GSTP1, LDHA, LGALS1, NARS, NT5C2, PAFAH1B, PGAM1 or PGAM2, PTGES3, RRM1, UCH-L1, CALR, CSE1L, KPNB1, NPM1, RAN, STAT1 $\alpha$ , YWHA, 14-3-3, STOM, STOML2, 2'-PDE, ATIC, BLVRB, CTPS1, CYB5R3, GNB1, HEATR2, HKDC1, PPP1C, PPP2 complex (one subunit), PRDX1, PRDX6, USP14
ER-resident proteins.....	CANX, NSF
Membrane-associated proteins.....	C1orf57, SLC25A6, ACSL4, ESD, PHB2, TAGLN2
Mitochondrial proteins.....	DLST, EF-Tu, LONP1, MDH2, MTHFD1, PCK2, F <sub>1</sub> F <sub>0</sub> -ATP synthase, HADHB, MTCH2, VDAC1, VDAC2, VDAC3
Nuclear proteins.....	PRKD, NAP1L1, NME2, SAE1, SFPQ, H2A, H2B, H3, H4, RRP12, MATR3, IGF2BP1, DDX3, DDX5, DDX9, DDX21, DDX39, NONO, PRPF8
Ribosomal subunits and translational cofactors.....	L3, L4, L5, L6, L7, L7a, L9, L10a, L14, L17, L18, L18a, L23, L24, L27, L27a, L30, P0, S2, S3, S3a, S5, S6, S7, S8, S9, S11, S12, S13, S16, S17, S20, S23, SA, eEF1, eEF1, eEF2, eEF3, eEF4, eEF5a, eIF4a-1, eIF4a-3, GCN1L1

<sup>a</sup> For a more detailed listing, see Table S1 in the supplemental material. Proteins listed here were identified in at least one native and one PK SARS-CoV preparation and were not identified in background samples. We have identified proteins here with standard abbreviations for the corresponding human genes from the NCBI Entrez Gene database (<http://www.ncbi.nlm.nih.gov/sites/entrez>).

## RESULTS

**Protein purification and background analysis.** We investigated the protein composition of SARS-CoV released from Vero-E6 cells during the peak growth period from 24 h to 48 h after inoculation. The protein fraction of clarified cell culture supernatant was collected by polyethylene glycol precipitation, and virus particles were purified by banding on a sucrose density gradient. Virus purified in this way is referred to here as “native” SARS-CoV. Purified virus subjected to surface cleansing with DNase I followed by proteinase K is referred to as “PK” virus.

We also attempted a proteomics analysis of Junin-Candid1 arenavirus at the same time as SARS-CoV. However, due to the slow growth of Candid1 in our hands, the resulting samples were essentially virus free but contained numerous high-molecular-weight proteins associated with the cytoskeleton (21% of the proteins identified) and extracellular matrix (60% of the proteins identified [Table 2]). Trace sequences totaling 3.4% of the full-length Candid1 nucleoprotein were identified, but these samples were otherwise free of viral proteins. Nucleoprotein is the most plentiful component of purified arenavirus (58), but characteristic virion components such as the Candid1 SSP, GP-C, Z, and L proteins and host ribosomes (41) were conspicuously absent from these preparations. These samples were used to approximate the spectrum of proteins purified from uninfected Vero-E6 cells and are referred to here as “background” samples.

Background samples also contained several proteins previously identified as components of other enveloped viruses, for example, actin, myosin, and fibronectin (28). Enzymatic treatment in PK samples appeared to eliminate most background

proteins, but a few cytostructural proteins including actin, myosin, filamin, tubulin, and fibronectin were consistently found in PK samples, indicating probable incorporation into the virion. Ubiquitin appeared to be enriched following PK treatment and therefore also likely represents a genuine virion component.

**Proteomics of SARS-CoV.** To determine the protein composition of the purified native, PK, and background samples, we performed two-dimensional liquid chromatography MS/MS analysis of peptide mixture generated by in-solution digestion of the proteins. Two primary extraction techniques were employed: TCA precipitation and methanol delipidation. Peptides extracted by TCA and methanol delipidation were analyzed separately, and the results were combined. Some proteins were identified using only one extraction technique, while others were identified with both. Except where explicitly stated, proteins reported here met three criteria: (i) presence in at least one PK sample, (ii) presence in one native sample, and (iii) absence from both background samples. SARS-CoV grows relatively poorly in most human cell types, and so the virus was grown in Vero-E6 cells derived from the African green monkey *Cercopithecus aethiops*. Because of the limited number of *Cercopithecus aethiops* protein sequences available, peptides were screened against a database including *Cercopithecus aethiops* and *Homo sapiens* sequences in addition to all SARS-CoV protein sequences of at least 9 amino acids. Using this procedure, eight viral proteins and 172 host proteins were identified from SARS-CoV, including the three explicit *Cercopithecus aethiops* sequences cyclophilin A (PPIA), calreticulin (CALR), and STAT-1 $\alpha$  (overview in Table 3; see also detailed descriptions in Table S1 in the supplemental material). Be-

TABLE 4. Viral proteins identified in purified SARS-CoV<sup>a</sup>

Protein	No. of times found (of two possible times) in sample type:		No. of peptides <sup>d</sup>		% Coverage <sup>e</sup>	Length (aa) <sup>f</sup>
	Native <sup>b</sup>	Digested <sup>c</sup>	Total	Unique		
N (ORF9)	2	2	4	4	18.0	423
M (ORF5)	2	2	6	2	19.4	222
S (ORF2)	1	2	16	16	21.6	1,256
nsp3	1	2	16	14	12.5	1,922
nsp5	1	2	2	2	14.1	306
nsp2	1	1	4	3	11.0	639
9b (ORF9b)	1	1	2	1	22.2	99
3a (ORF3a)	1	1	2	1	6.2	275
nsp4 <sup>g</sup>	0	2	5	3	15.2	500
nsp9 <sup>g</sup>	0	2	2	2	23.9	113

<sup>a</sup> Proteins are ranked in relative confidence order as a surrogate measurement for relative copy number according to the following criteria: number of times detected in native samples > PK samples > the product of percent coverage and protein length.

<sup>b</sup> Purified, native virions.

<sup>c</sup> Purified, DNase I-treated, proteinase K-treated, repurified virions.

<sup>d</sup> Number of total and nonoverlapping (unique) peptides identified for each protein.

<sup>e</sup> Coverage refers to the percentage of the intact protein length accounted for from unique contributions of the fragments detected by mass spectrometry.

<sup>f</sup> Length, in amino acids, of each protein or proteolytically processed nsp.

<sup>g</sup> nsp4 and nsp9 were not detected in native samples and thus did not meet the full validation criteria of this study.

cause of the large number of proteins identified, most of the host proteins listed in Table 3 and Table S1 in the supplemental material are presented without regarding their potential relevance to the viral replication cycle.

**Specificity of incorporated protein kinases.** Coronavirus nucleoproteins are phosphorylated by host protein kinases, including cyclin-dependent kinase, glycogen synthase kinase, mitogen-activated protein kinase (MAPK), and casein kinase II (CSNK2) (59). Nucleoprotein phosphorylation has been proposed as the mechanism leading to incorporation of host protein kinases in coronavirus particles, as has been demonstrated for MHV (56). The two host protein kinases identified here by mass spectrometry (CSNK2 and DNA-activated protein kinase [PRKD]) function in host signaling cascades and are therefore of potential importance to SARS-CoV pathogenesis. A function-based screening method was used to further investigate the presence of protein kinases identified by mass spectrometry in that native SARS-CoV lysates were used to radiolabel a microarray containing 1,152 peptides with known phosphorylation sites (see Fig. S1 in the supplemental material).

Substrates that were phosphorylated by at least 1 standard deviation above background levels in each of two replicate arrays are reported here. Of 77 phosphorylated substrates, 29 could be linked with a specific protein kinase. Three kinase activity signatures were detected multiple times in the virion lysate, i.e., CSNK2 (four substrates), protein kinase A (PRKA; 12 substrates), and protein kinase C (PRKC; five substrates). Other kinase signatures represented by a single phosphorylated substrate included CAMK2, CKS1, CSK, epidermal growth factor receptor, GRK1, MAPK1, PHK, and RPS6K. Of these, CSNK2 was detected in both PK and native virion lysates and thus represents a probable virion component. Ribosomal protein S6 kinase (RPS6K) was found in both PK sam-

ples and is probably incorporated, as we conclude from the generally heavy ribosomal protein representation in SARS-CoV as well as the specific presence of the RPS6 substrate in the sample. PRKA, PRKC, and MAPK1 were absent in PK samples, and each was detected in only one native sample (data not shown), and therefore we concluded that they were present through adventitious copurification or entanglement at the virion surface. One protein kinase detected by mass spectrometry, PRKD, was not detected by substrate phosphorylation, possibly due to the presence of only three validated PRKD substrates on the chip.

**Relative abundance of viral proteins.** Protein detection by mass spectrometry proteomics depends on factors including abundance, sensitivity of detection, enzymatic pretreatment, extraction method, proteinase accessibility, and availability of potential proteolytic cleavage products of appropriate molecular weight. Mass spectrometry is therefore not an optimal tool for precise measurement of the absolute stoichiometry of incorporated components but can provide a general idea of ranked abundance within a sample. We used a hierarchy of native detection frequency > PK detection frequency > peptide coverage relative to protein length for a tentative ranking of the relative abundance of viral and host proteins in SARS-CoV (Table 4). SARS-CoV N, M, and S were consistently among the 10 most abundant proteins detected in PK samples. The accessory SARS-CoV ORF3a and ORF9b proteins and nsp2, nsp3, and nsp5 were present in lower relative abundance and were of equal or lesser abundance in PK samples than were some ribosomal proteins, histones, heat shock protein 90, and phosphatase I (Table 4; also see Table S1 in the supplemental material).

**Characterization of novel SARS-CoV virion proteins.** Mass spectrometry proteomics revealed the SARS N, M, S, ORF3a, and ORF9b proteins, as well as three components of pp1a, i.e., nsp2, nsp3 and nsp5, as viral components (Table 4). Two additional replicase components, i.e., nsp4 and nsp9, were enriched in PK virus but were not detected in native SARS-CoV preparations. An interaction network was created incorporating biochemical interaction data (see reference 65 and references therein) and protein-RNA interaction data (11, 38, 53, 54) to illustrate the network of interactions related to virion assembly (Fig. 1). All viral proteins identified in this study or known from other, previously published work can be linked

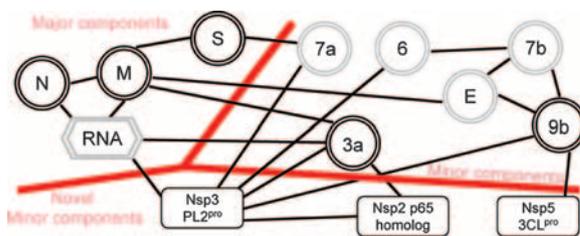


FIG. 1. Interaction map for SARS-CoV-derived components. Double outlines indicate major components, including known high-copy-number virion proteins and the large viral RNA genome, and minor components, including low-copy-number and weakly conserved proteins. Black outlines identify components detected by mass spectrometry proteomics. Gray outlines indicate components identified in other published studies. Solid single outlines denote novel components identified in both native and PK SARS-CoV.

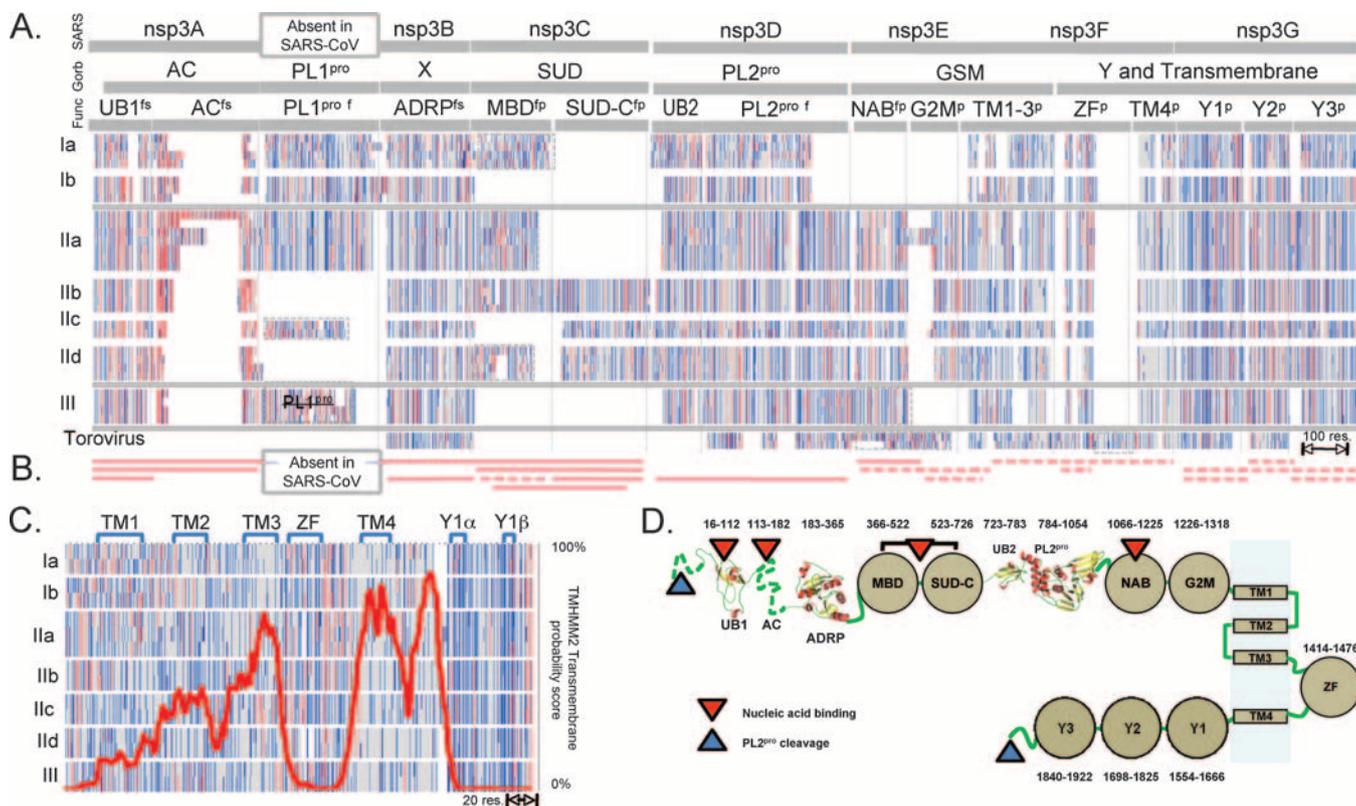


FIG. 2. Overview of nsp3 organization. (A) Multiple sequence alignment of coronavirus and torovirus nsp3 homologs. The 16-component functional annotation presented here (Func) is an extension of our previous SARS-CoV-specific domain boundary prediction (SARS) and the ongoing analysis by Gorbalenya and collaborators (Gorb). It incorporates domain boundaries defined in a hierarchy of functional (f), structural (s), and phylogenetic (p) criteria. The functional annotation was compiled from published data and results presented here. Region designations include the following: ubiquitin-related domains (UB1 and UB2), an acidic hypervariable region (AC), complete (PL1<sup>pro</sup> and PL2<sup>pro</sup>) or partial (<sup>pro</sup>) papain-like cysteine proteinases, ADRP, a SARS-CoV subgroup-specific MBD, the carboxyl-terminal moiety of the “SARS-unique domain” (SUD-C), group II-specific NAB domain and marker domain (G2M), two predicted double-pass transmembrane domains (TM1-2 and TM3-4), a putative metal-binding region (ZF), and three subdomains forming part of the Y region (Y1 to Y3) originally described by Gorbalenya et al. (18). Dotted lines denote additional subgroup-specific domains not included in the annotation above. Amino acid residues are color coded gray (AFGILMPVWY), light blue (KNQRST), blue (CH), or red (DE) to highlight patterns that may mark conserved protein structures. We divide group II into four subgroups following published suggestions (71) and divide group I into two subgroups. Sequences from equine and bovine toroviruses are shown from the domain homologous to ADRP onward. (B) Selected SARS-CoV expression constructs. Solid lines denote expression (also Table 1); dashed lines indicate that no expression has so far been obtained. (C) Enlargement of the ZF and flanking regions, with transmembrane domain predictions. The overlay shows the average transmembrane probability score for 400-amino-acid regions centered on the first conserved cysteine of ZF. A red overlay displays average transmembrane probability scores calculated by TMHMM2 for this region from a set of 15 representative coronaviruses, approximately equally weighted with respect to each subgroup (see Materials and Methods). For display purposes, in this panel the sequences are aligned only with conserved clusters of four cysteine/histidine residues in ZF and Y1 ( $\alpha$  and  $\beta$ ). (D) Structural annotation of SARS-CoV nsp3. Experimentally characterized flexibly disordered regions are indicated with dashed green lines, and predicted flexible regions separating conserved domains are indicated with solid green lines.

directly or indirectly to the four major virion components (defined here as major components with respect to copy number and relative molecular weight), i.e., N, M, S, and the genomic RNA. We were unable to detect the small, hydrophobic E protein in SARS-CoV lysates by mass spectrometry.

Proteins with many interacting partners identified among virion-incorporated proteins included the M protein, which coordinates S, E, N, and possibly RNA incorporation into the virion, and nsp3, which is a novel virion component. Only less than half of the SARS-CoV nsp3 protein has been characterized to date. The characterized regions include a poly(ADP-ribose)-binding ADRP, a papain-like proteinase and deubiquitinase (PL2<sup>pro</sup>), and two domains with ubiquitin-like folds (UB1 and UB2). Therefore, we selected SARS-CoV nsp3 for further characterization.

**Phylogenetic analysis of nsp3.** The most frequently encountered protein globular domains are formed from contiguous polypeptide chain segments of about 100 amino acid residues (68). Previous bioinformatics analyses of nsp3 had identified only a few domains fitting this criterion, but they predicted several large regions likely to include multiple structural domains. We therefore compiled a higher-resolution analysis of nsp3 domain architecture as a tool for novel structural and functional characterization. We performed a phylogenetic analysis of nsp3 (Fig. 2; see also Fig. S2 in the supplemental material) to identify small, conserved regions that might yield expressible protein domains. Protein sequence analysis of coronavirus and torovirus nsp3 homologs revealed a pattern of alternating conserved and nonconserved regions, consistent with a multiple-domain and linker structure (Fig. 2A; see also

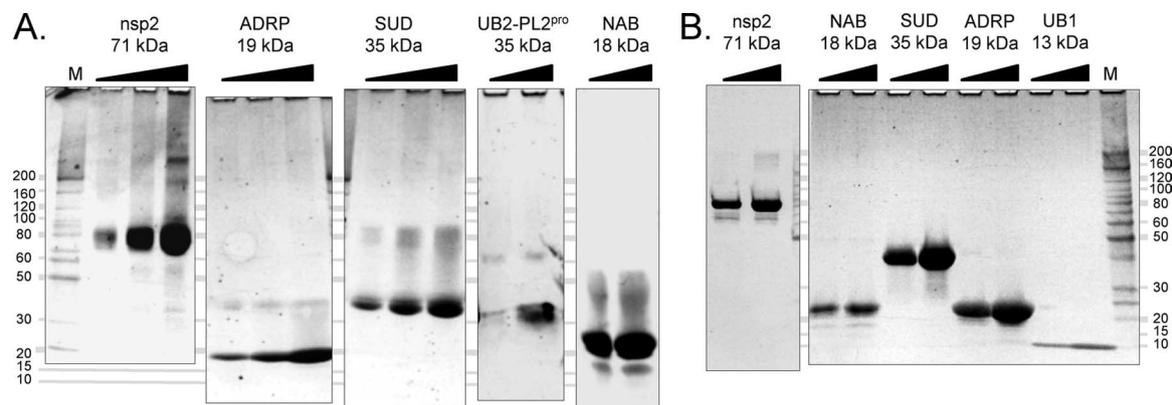


FIG. 3. Oligomerization of SARS-CoV nsp3 domains. (A) PFO-PAGE analysis reveals the oligomeric state of selected nsp3 domains in solution. A Benchmark protein ladder (M) was used to estimate protein and protein complex molecular masses, indicated in kDa at left. Lanes in panel A contain, from left to right, 25  $\mu$ M, 50  $\mu$ M, and 100  $\mu$ M nsp2, ADRP, and SUD; 25  $\mu$ M and 50  $\mu$ M UB2-PL2<sup>Pto</sup>; and 50  $\mu$ M and 100  $\mu$ M NAB, respectively. (B) Reducing sodium dodecyl sulfate-PAGE analysis of selected nsp3 domains. Lanes in panel B contain, from left to right, 50  $\mu$ M and 100  $\mu$ M nsp2, NAB, SUD, ADRP, and UB1, respectively.

Fig. S2 in the supplemental material). Results from previously published studies (17, 61, 73) and fold recognition software (24) were incorporated in this process of construct design. Previous studies showing that the UB1, PL2<sup>Pto</sup>, and ADRP domains of nsp3 were both well folded and functional when expressed separately were taken as support of the domain and linker structure of nsp3 (45, 47, 53).

As shown in Fig. 2B, predicted domains located toward the amino terminus of nsp3 were tested and found to be generally amenable to expression as domains, while all but one region downstream of the PL2<sup>Pto</sup> domain was not efficiently expressed. One possible reason for the expression difficulties may lie in the presence of a long hydrophobic domain predicted to contain four transmembrane spans in this region (Fig. 2C). Based on the expression pattern and the available structural data, a general model of nsp3 structure was proposed (Fig. 2D). In modeling nsp3, we were guided by the assumption that nsp3 topology would be constant among coronaviruses. The proposed structure contains four transmembrane spans and places nearly all of nsp3, including the PL2<sup>Pto</sup> domain, on one face of the membrane. The domain topology of the model of membrane-embedded nsp3 is inferred from the presence of PL2<sup>Pto</sup> cleavage sites at both termini of nsp3 and bioinformatic predictions. While the exact number of transmembrane spans is not certain, any multiple of two could be conducive to post-translational processing of nsp3 by PL2<sup>Pto</sup> and would present the bulk of nsp3 on the same membrane face occupied by nsp5 3CL<sup>Pto</sup> and the pp1b replicase proteins. Our model of TM distribution (Fig. 2C) follows the 3TM + 1TM distribution of transmembrane regions recently proposed for MHV nsp3 (26), which was based in part on observed glycosylation patterns from truncated nsp3 constructs (19, 26) and is consistent with an independent model of nsp4 structure (40). The interpretation presented in Fig. 2C includes all three major phylogenetic groups and the newly sequenced group II bat coronaviruses. Although we note that phylogenetic evidence more consistently suggests a 2TM + 2TM distribution across the coronavirus family (Fig. 2C), the weight of biochemical evidence currently favors the 3TM + 1TM distribution.

Several types of domain designation may be possible for a given set of input sequences, depending on the criteria used for selection. Here we present a working functional annotation based on a hierarchy of functional > structural > phylogeny-based domain identification. Where protein function and structure are known, “functional” domains such as ADRP and PL2<sup>Pto</sup> have been noted. Where only the structure is known, as for ubiquitin-related UB2, “structural” domains are noted. Where only the primary sequence data were available, islands of sequence conservation, termed “phylogenetic” domains such as Y1 to Y3, were designated. Our analysis revealed 16 conserved nsp3 domains—identified here as UB1, AC, PL1<sup>Pto</sup>, ADRP, MBD (metal-binding domain), SUD-C, UB2, PL2<sup>Pto</sup>, NAB, G2M, TM1-2, ZF, TM3-4, Y1, Y2, and Y3—of which between 12 and 15 domain homologs could be identified in any one coronavirus (Fig. 2A). Tryptic peptide fragments of nsp3 identified by mass spectrometry were derived from the ADRP (four peptides), MBD (one peptide), SUD-C (three peptides), PL2<sup>Pto</sup> (two peptides), Y1 (two peptides), and Y2 (three peptides) domains. The multidomain construct SUD, with residues 389 to 726, encompasses the newly annotated MBD and SUD-C domains.

**Stoichiometry of nsp3.** PFO is a nondissociative detergent that can be used with native PAGE to determine the mass of protein complexes (44). We investigated the oligomeric structure of purified nsp3 domains using PFO-PAGE. The expressed domains and multidomain constructs of nsp3 tested here (Fig. 3) and previously (45, 53) appeared to migrate mainly as monomeric species, with trace amounts of dimers visible, while lysozyme and protein molecular weight markers migrated as monomers, as previously reported (44). In contrast, full-length nsp2 was primarily monomeric, with a small concentration of trimeric species and traces of dimeric, tetrameric, and higher-molecular-weight species (compare Fig. 3A and 3B), confirming that monomer > dimer oligomerization is characteristic of nsp3 domains.

PFO-PAGE analysis of mixed nsp3 domains revealed the presence of high-molecular-weight species consistent with the size of 1 $\times$  UB1+SUD, 2 $\times$  UB1+SUD, and 1 $\times$  ADRP+SUD

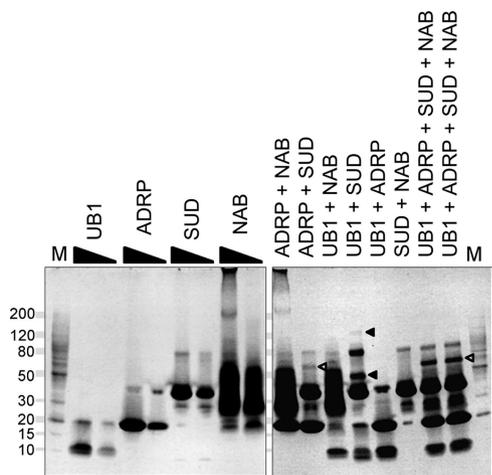


FIG. 4. PFO-PAGE analysis of interdomain oligomerization. Approximately equimolar concentrations of bacterially expressed nsp3 domains were incubated separately (left) or in combination (right) at 37°C for 1 h and analyzed by PFO-PAGE. The panel at left demonstrates the electrophoretic mobility of each protein species and homooligomer; lanes at left contain 2 and 1 nanomole of UB1, ADRP, or SUD or 10 and 5 nanomoles NAB, respectively. Each lane at right depicts mixtures of 2 nanomoles of UB1, ADRP, or SUD and 5 nanomoles NAB as shown. Proteins were visualized with SYPRO-ruby staining. Marked bands correspond to 50-kDa and 110-kDa UB1+SUD complexes (filled triangles) and 60-kDa ADRP+SUD complexes (open triangles). In the presence of additional nsp3 domains, UB1+SUD complexes are not formed, but the amount of ADRP+SUD complex is increased. Duplicate samples are shown for the four-domain mixture. Lanes containing the Benchmark protein ladder are indicated (M), with masses in kilodaltons indicated at left.

(Fig. 4). Both UB1-SUD complexes disappeared when additional nsp3 domains were added prior to incubation, consistent with a weak ionic interaction between acidic UB1 and SUD, whereas the ADRP-SUD complex formation was enhanced in the presence of additional nsp3 domains. Homodimeric forms of nsp3 domains also persisted in the presence of additional

nsp3 domains. As shown in the rightmost lanes of Fig. 4, SUD+ADR and SUD+SUD complexes were present in the same sample. Products of the expected size for a SUD+SUD+ADR complex were not observed, indicating that the SUD and ADR binding sites on an SUD molecule either overlap or are mutually antagonistic. Although the in vitro data show that the individual nsp3 domains are predominantly monomeric, there is also support for the hypothesis that in vivo the macromolecular structure of nsp3 may be constrained by multiple intrachain and homotypic interchain interactions and that nsp3 may exist in multiple alternate macromolecular assemblies.

**Metal binding analysis.** A relatively large number of MBDs have been discovered or predicted among coronavirus replicase proteins. The recent structure of nsp10 revealed two zinc fingers (25), nsp15 utilizes manganese as a cofactor (23), and both the nsp13 helicase and nsp14 exonuclease domains contain conserved clusters of cysteine and histidine residues that are characteristic of metal ion-binding domains. In addition to the validated MBD located within PL2<sup>pro</sup>, at least three other conserved potential metal-binding motifs exist in the carboxyl-terminal region of nsp3 (Fig. 2C; ZF, Y1 $\alpha$ , and Y1 $\beta$ ). During some but not all purifications of bacterially expressed SUD, addition of protein caused a visible “bleaching” effect on the Talon affinity matrix which was interpreted to arise from cobalt stripping activity.

To test for metal-binding activity by SUD, we added additional CoCl<sub>2</sub> and ZnCl<sub>2</sub> to purified SUD, SUD<sub>451-651</sub>, and SUD-C domains and examined the UV-visible spectra (Fig. 5). Zinc binding does not produce a detectable spectral change, but charge transfer between cobalt(II) and sulfur atoms (here, probably cysteine residues) produces a characteristic absorption signal with peaks at ~310 and 340 nm (5). UV-visible spectrum analysis indicated that full-length SUD (389 to 726; Fig. 5A to C) bound cobalt, whereas neither the truncated SUD (451 to 651; Fig. 5F) nor the carboxyl-terminal portion of this region (SUD-C 513 to 651; Fig. 5D and E) showed evi-

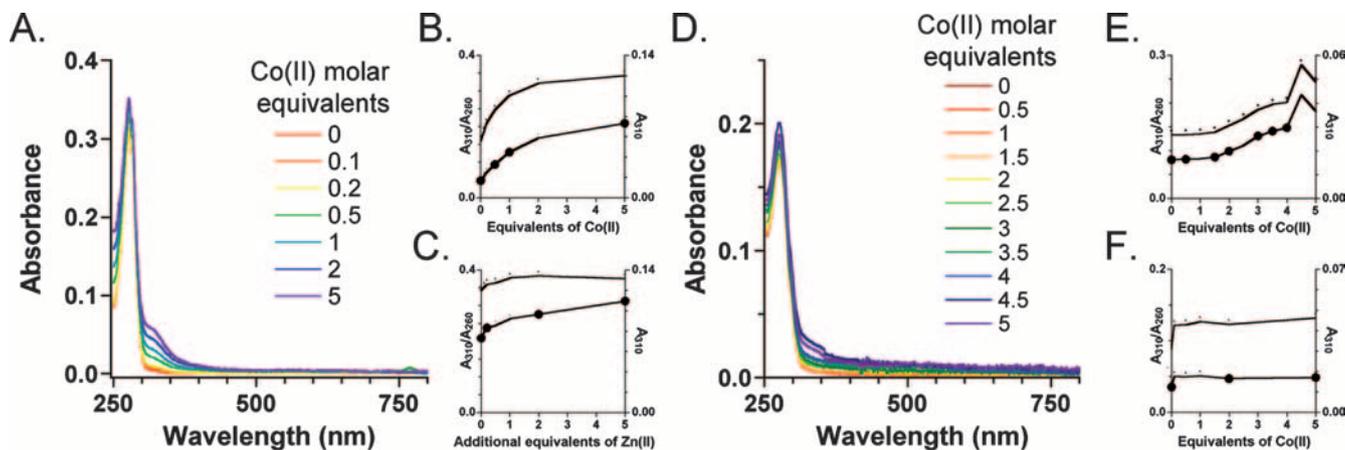


FIG. 5. Titration of cobalt binding by 10  $\mu$ M SUD and SUD-C. UV-visible spectra of 10  $\mu$ M full-length SUD (A to C), SUD-C (D and E), and truncated SUD<sub>451-651</sub> (F) solutions were measured after addition of 0 to 5 molar equivalents of Co(II) in the form of CoCl<sub>2</sub>. Relative Co(II) concentration is indicated with colored lines running from red (0 equivalents) to violet (5 equivalents). Because of the observed metal ion concentration-dependent protein precipitation during these experiments, both the raw absorbance at 310 nm ( $A_{310}$ ; panels B, C, E, and F; black circles) and normalized absorbance ( $A_{310}/A_{250}$ ; open circles) are plotted. (C) Displacement of Co(II) by Zn(II) was investigated by addition of ZnCl<sub>2</sub> to 10  $\mu$ M SUD solutions that had been previously saturated with 5 equivalents of Co(II).

dence of cobalt binding. Addition of zinc(II) to cobalt-complexed SUD did not dampen the S-Co(II) spectral signal but appeared to induce additional protein precipitation, visualized as a general increase in the absorbance in the far-UV range, which was confirmed by visual inspection. The precipitation-corrected 310-nm absorbance curves (Fig. 5B, inset) are most consistent with binding of a single cobalt atom per SUD molecule. Addition of zinc following cobalt saturation did not diminish the spectral signal at 310 nm, indicating that equimolar zinc is unable to displace bound cobalt bound to SUD. These data were interpreted to indicate that a cysteine-coordinated metal ion-binding site with a high affinity for cobalt is localized partly or wholly in the amino-terminal domain of SUD, which we therefore describe as the MBD. SUD contains six conserved cysteines (SARS-CoV nsp3 positions 393, 456, 492, 507, 550, and 623) and two conserved histidine residues (positions 539 and 613), which could participate in a tetrahedral metal ion coordination site. The lack of metal ion binding by truncated SUD<sub>451-651</sub> suggests that Cys393 may have a key role in metal ion coordination.

**Nucleic acid binding analysis.** We previously reported that both the UB1 domain and the glutamic acid-rich acidic (AC) hypervariable domain, collectively known as nsp3a, consistently copurified with nucleic acid, implicating nsp3 as a nucleic acid-binding protein (53). EMSAs were performed to investigate whether nsp3 domains concealed further nucleic acid-binding sites. Two domains, the full-length SUD and the NAB domain, which immediately follows PL2<sup>PRO</sup>, exhibited nucleic acid-binding activity at micromolar concentrations (Fig. 6). SUD and NAB were therefore tentatively annotated as nucleic acid-binding domains pending further functional characterization. Relatively high micromolar concentrations of the SUD-C domain produced a reproducible but indistinct electrophoretic mobility shift in the presence of nucleic acids, which may be attributable to an electrostatic interaction mediated by the net positive charge of SUD-C at neutral pH. The lack of appreciable nucleic acid-binding affinity by SUD-C suggests that MBD may modulate nucleic acid binding by the full-length SUD. RNA binding, rather than DNA binding, is expected to be the native function of nsp3 domains as was previously suggested for SARS-CoV nsp9 (11). SUD and NAB showed an equivalent or slightly higher affinity for ssRNA than for dsRNA (Fig. 6B). Neither nsp2 nor ADRP showed appreciable binding to any of the generic RNA or DNA substrates tested (Fig. 6).

Bacterially expressed NAB displayed similar ATP-independent dsDNA unwinding properties (Fig. 7), consistent with preferential single-stranded nucleic acid binding. We compared the activity of NAB with that of a previously described amino-terminal structured domain of the SARS-CoV N protein (N-NTD [48]). Approximately 20-fold-less NAB was required to generate the same level of unwinding activity on dsDNA as that observed for N-NTD on dsRNA (Fig. 7B). Single-strand binding and double-strand unwinding by NAB are consistent with a nucleic acid chaperone function, which has also been proposed for the coronavirus nucleoprotein (74).

**Annotation of uncharacterized domains.** We were prevented from experimentally characterizing the function of all nsp3 domains, since seven domains were not expressed by *E. coli*: G2M, TM1-2, ZF, TM3-4, Y1, Y2, and Y3. We therefore

used conservation-based statistics to qualitatively profile the function of the unexpressed domains. This method is based on two hypotheses: (i) the extent of protein conservation mirrors the relative importance in the virus replication cycle and (ii) conserved enzymatic activity should place more constraints on protein sequence divergence than nonenzymatic function or species-specific “accessory” enzymatic function. We calculated the maximum percent amino acid identity for 392 pairs of aligned protein or domain homologs belonging to different subgroups within the same group (i.e., group IIa versus group IIb and group IIa versus group IIc). Whole proteins were used for this analysis, except where multiple sequence alignments revealed the presence of multiple conserved domains surrounded by areas of very low conservation, as was observed for the amino- and carboxyl-terminal domains of the N protein. Proteins and domains from 13 representative coronaviruses were included in the comparison. As expected, enzymatic and nonenzymatic functions corresponded to significantly different levels of conservation (Mann-Whitney U test;  $P < 0.001$ ; see Fig. 8). These results reflected the following rank order of conservation: enzymes > enzymatic domains > nonenzymes > nonenzymatic domains. These data demonstrate how a qualitative functional assignment can be inferred from the degree of conservation for coronavirus proteins.

We examined protein conservation for the aforementioned seven uncharacterized nsp3 domains (Fig. 8). Conservation analysis predicted nonenzymatic (or nonconserved enzymatic) function for the four domains G2M, TM1-2, ZF, and TM3-4. All three domains from the Y region (Y1 to Y3) were approximately equally conserved and ranked between enzymes and enzymatic domains. From the consistently high conservation of Y1, Y2, and Y3, we hypothesize that Y1 to Y3 may form a single functional unit with a conserved enzymatic function.

## DISCUSSION

**Limitations in the interpretation of mass spectrometry results.** The results presented here indicate that nsp3 and several other proteins of viral and host origin may be contained in purified virions. Although we have not formally eliminated rare vesicles of a buoyant density similar to that of SARS-CoV as a source for some of the noncanonical proteins detected in this study, we believe that this possibility is remote, based on our earlier observations of the purity of SARS-CoV preparations (see above and reference 39) and considering the fact that other replicase proteins such as the polymerase and helicase were not observed in this study. The biological significance of nsp3 packaging and the implications for other coronaviruses remain to be determined.

DNase and proteinase K treatments were performed to differentiate between proteins entwined or embedded at the virion surface and internal proteins. Data presented in the “PK” column of Table 2 demonstrate that the enzymatic treatment followed by an additional density gradient purification step did reduce detection of most “background” extracellular matrix proteins below the threshold of detection. However, proteinase K treatment did not completely eliminate all viral surface proteins. The spike protein ectodomain was detected after enzymatic treatment of the virions, possibly because of the persistence of a proteinase-resistant core. Thus, we were

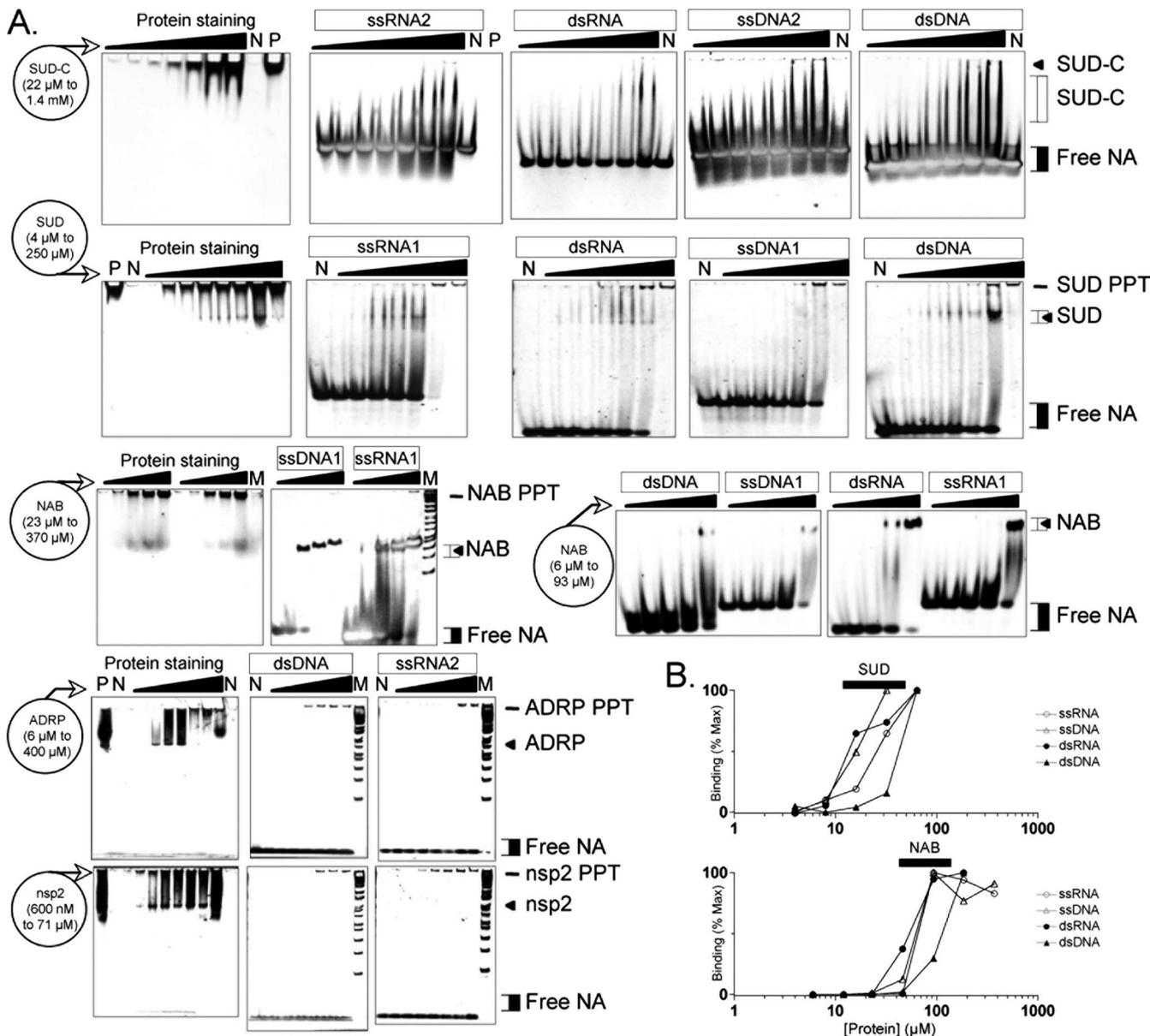


FIG. 6. Generic nucleic acid binding properties of SUD-C, SUD, and NAB domains of nsp3. (A) EMSAs were performed with sequence-matched 20-nucleotide dsDNA or dsRNA or one of two functionally equivalent sets of sequence-matched 40-nucleotide ssDNA or ssRNA oligomers. Gels were stained for protein or nucleic acid as indicated. Lanes containing protein only at the highest listed concentration (P), 800 ng of nucleic acid only (N), dsDNA ladder marker (M), and mixtures of protein with 800 ng nucleic acid are indicated. Protein concentration decreases in twofold increments from left to right within the indicated range. Maximum protein concentrations used here were determined empirically by expression and stability in solution at 4°C. Electrophoretic mobility ranges for nucleic acids (black brackets), protein (small triangles), and protein-nucleic acid complexes (white brackets) are indicated on the right. SUD-C has a small net positive charge at neutral pH and migrated through the gel only in complex with nucleic acid (NA). Results from two single-stranded nucleic acid sequences that behaved equivalently in non-sequence-specific EMSA are shown. (B) Binding curves were constructed from densitometry data calibrated to the maximum and minimum binding in each gel. The range in which increasing nucleic acid binding was observed is indicated with a bold line above each graph to facilitate comparison. SUD binding curves may overestimate affinity since maximal binding overlapped with the limit of protein solubility.

unable to rule out either possible topology for nsp3 in the viral membrane based on proteolytic cleavage by proteinase K.

**Origin and quantification of proteins detected in our analysis.** The diversity of proteinase K-resistant host proteins that we found to be associated with purified virions may best be explained as a manifestation of the internal state of the infected host cells at the time of peak viral release. Between 24

and 48 h postinoculation the infected Vero-E6 cells became rounded and detached, with a granular appearance characteristic of late-stage infection. The extensive collection of histones observed in SARS-CoV virions following sequential treatment with DNase I and proteinase K argues against entwined host chromatin on the virion surface as the source of these proteins, just as the absence of the most common, high-copy-number

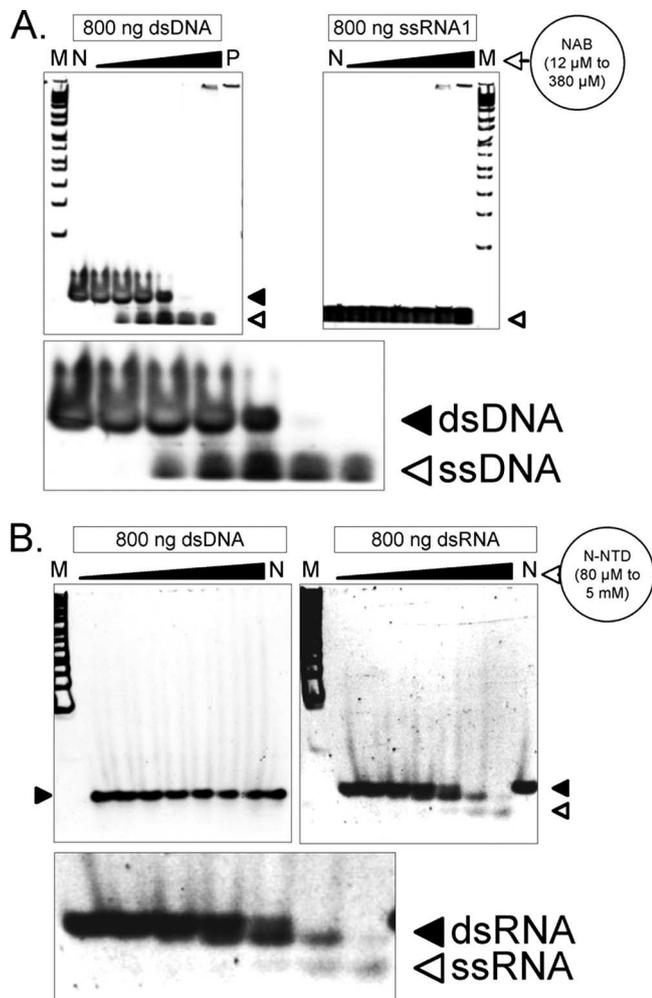


FIG. 7. Duplex unwinding activity of NAB and comparison with SARS-CoV nucleoprotein amino-terminal structured domain (N-NTD). Samples of NAB (A) or N-NTD (B) protein and single-stranded or duplex nucleic acid were mixed and incubated as for EMSA and then chilled overnight to allow the protein-nucleic acid complexes to dissociate before analysis by native PAGE. Lanes containing the highest concentration of protein only (P), nucleic acid only (N), and dsDNA marker (M) are indicated. Double-stranded (filled triangles) and single-stranded (open triangles) nucleic acids were detected with SYBR-gold dye, which stains double-stranded nucleic acid more prominently than single-stranded nucleic acid. Protein concentration decreases in twofold increments within the range shown. Enlargements showing the dose-dependent nucleic acid unwinding activity are included at the bottom of each panel.

nuclear proteins argues against copurification of intact nuclei. Packaged shreds of degraded chromatin from apoptotic cells would seem to be a more likely source, and it has been estimated that 95% of the Vero-E6 cells are apoptotic within 48 h of infection (37). This could also explain the presence of the mitochondrial and ribosomal proteins observed. Analysis of background proteins pelleted from the supernatant of Vero-E6 cells over the same time period did not reveal any histone, ribosome, nuclear, or mitochondrial proteins. When discussing these observations, we need to keep in mind that the presently used collection procedure was designed to maximize virus yield and therefore protein detection. Analysis of virus collected

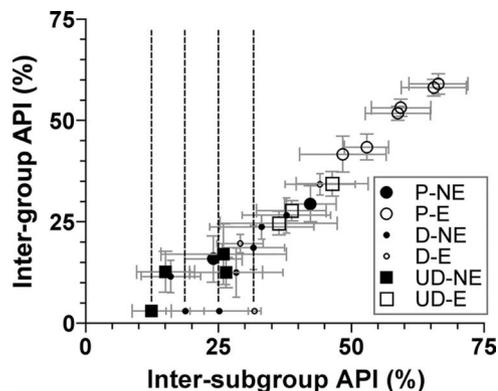


FIG. 8. Use of amino acid conservation to infer function for experimentally uncharacterized nsp3 domains. Average percent identity (API) was measured by pairwise alignment of conserved proteins and domains from different subgroups (Ia versus Ib, IIa versus IIb, etc.) or groups (I versus III, etc.). Conserved coronavirus proteins are grouped by functional class, including enzymes (P-E; nsp5, nsp12, nsp13, nsp14, nsp15, and nsp16), nonenzymatic proteins (P-NE; M and E), enzymatic domains (D-E; ADRP, PL1<sup>pro</sup>, and PL2<sup>pro</sup>), and putative nonenzymatic domains (D-NE; UB1, AC, SUD-C, UB2, NAB, and two nucleoprotein domains). Dotted lines mark intersubgroup API values associated with domains not found in all groups (PLP1, SUD-C, NAB, and G2M). Subgroup-specific markers such as SARS-CoV MBD were not included. Uncharacterized nsp3 domains clustering with enzymatic (UD-E; Y1, Y2, and Y3) and nonenzymatic (UD-NE; TM1-2, ZF, TM3-4, and G2M) classes are indicated.

before the onset of apoptosis might reveal a somewhat different protein profile and will be the focus of a future study.

Proteins previously reported to interact with incorporated coronavirus proteins and genomic RNA were well represented in the proteomics results, including two proteins reported to bind N protein, i.e., cyclophilin A and 14-3-3 (tyrosine 3-monooxygenase/tryptophan 5-monooxygenase activation protein) (34, 59). We did not detect UBE2I, which is an E2 ubiquitin-conjugating enzyme reported to interact with the SARS-CoV nucleoprotein (32), but did identify three other points of contact with the ubiquitin and ubiquitin-like conjugation pathways, i.e., the ubiquitin-specific proteinase 14 (USP14), the ubiquitin-carboxyl-terminal esterase L1 (UCH-L1), and the SUMO-1 activating enzyme 1 (SAE1). Further research may determine whether the presence of these host proteins could be related to the presence of two ubiquitin-like domains (45, 53) or to the ubiquitin-cleaving activity of nsp3 (33). Proteins previously reported to interact with the MHV genome including polypyrimidine tract binding protein (PPTB1), cytoplasmic polyadenosine binding proteins (PABP4 and PABP1/3), and heterogeneous nuclear ribonucleoproteins (hnRNPA1, hnRNPA2/B1, and hnRNPA3) were among the numerous RNA-binding proteins detected (55). However, in this study we were unable to distinguish whether RNA-binding proteins were incorporated bound to viral RNA or host mRNAs or as soluble proteins.

The relative abundance of proteins detected by mass spectrometry can be approximated from the reproducibility of detection and from the variety of peptides found, with more-complete coverage expected for overrepresented proteins than for rare proteins. Our data are consistent with published reports that the SARS-CoV M and N proteins are highly abun-

dant in the virion, closely followed by the S protein (16). Proteins that were detected in fewer experiments and with lower coverage, such as ORF3a, ORF9b, nsp2, nsp3, and nsp5 proteins, are therefore likely to be present in lower relative copy numbers. Each SARS-CoV preparation that was analyzed contained  $\sim 10^{10}$  virions, making it possible that proteins present in single copies on only a small percentage of virions could still be identified.

**Connections to vesicular trafficking pathways.** The host factors involved in coronavirus budding remain largely unknown. Most enveloped viruses bud by coopting host proteins, often from the intracellular ESCRT transport pathways (reviewed in reference 67). Elements of the clathrin and COPI protein complexes were identified in PK SARS-CoV preparations. Clathrin coats assemble at the plasma membrane and the trans-Golgi network, which are quite distant from the endoplasmic reticulum (ER)-Golgi intermediate compartment (ERGIC) where SARS-CoV budding occurs. Other components of assembled clathrin lattices, including clathrin light chain and adaptor proteins, were not detected, suggesting that free clathrin may have been captured from the cytoplasm at the time of budding.

Three COPI components ( $\alpha$ -COP,  $\beta$ -COP, and  $\gamma$ -COP) and a protein involved in coatamer assembly (ARF4) were detected. The COPI coatamer plays a role in transport between the ER and the Golgi apparatus and in transport between Golgi stacks (27). COPI proteins are abundant at ERGIC membranes and have been shown to colocalize with budding MHV (29). A dibasic motif in the cytoplasmic tail of the SARS-CoV and MHV spike proteins of the type K(X)KXX was recently shown to bind COPI through an undetermined mechanism and is required for efficient interaction with the M protein (36). Coronavirus M proteins also possess a conserved dibasic motif in the cytoplasmic tail region, which might function similarly (see Fig. S3 in the supplemental material).

The ADP-ribosylation factor 4 (ARF4) is a small guanine nucleotide-binding protein involved in COPI trafficking. It has been shown that depletion of ARF4, but not of ARF1, ARF3, ARF5, or ARF6 (ARF2 having been lost in mammalian cells), induced tubulation at Golgi membranes (64). A similar phenomenon has been observed in MHV-infected cells and is linked to E protein expression (see reference 43 and references therein). ARF4-binding motifs are found in G-protein-coupled receptors such as rhodopsin, and these generally take the form of conserved NP(X<sub>n</sub>)Y motifs, where X<sub>n</sub> typically denotes the presence of one to three intervening nonconserved residues (10). A similar XP(X<sub>1</sub>)Y, XP(X<sub>2</sub>)Y, or XP(X<sub>4</sub>)Y motif can be found in most coronavirus E proteins (see Fig. S3 in the supplemental material). While we have not mapped the precise amino acid requirements for E-mediated budding in this study, two other mutagenesis and reversion studies have identified a region critical to the function of MHV (14) and TGEV (31) E proteins that maps to residues 47 to 65 in SARS-CoV E (see Fig. S3 in the supplemental material). The SARS-CoV E protein XP(X<sub>2</sub>)Y or XP(X<sub>4</sub>)Y motif, KPTVYVY, is found between residues 53 and 59. We note that the proline and the C-terminal tyrosine of this motif appear to be highly conserved among coronaviruses.

**Anticipated proteins that were not detected in this study.** Some peptides were likely missed in our study because of low

solubility, poor proteinase accessibility, or unreported sequence differences from the *Homo sapiens* homologs. As expected for the presently used technique, hydrophobic transmembrane regions are underrepresented, even among the proteins that were otherwise unambiguously detected. Thus, no peptides were recovered from the transmembrane regions of the S, M, ORF3a, nsp3, and nsp4 proteins. This may also explain the failure to detect hydrophobic low-copy-number virion components such as the E, ORF6, and ORF7b proteins. Proteinase K treatment may have eliminated the detectable regions of some type I integral membrane proteins with very small cytoplasmic tail regions such as the ORF7a protein, which was detected in native but not PK samples (data not shown). We therefore suspect that proteins from ORF2 to ORF9b may have been present in at least limited quantity in purified SARS-CoV preparations and that ORF3b, E, ORF6, ORF7a, ORF7b, ORF8a, and ORF8b proteins were not detected here due to technical limitations and the biochemical properties of these proteins.

Although we detected the ARF5 and COPI proteins, which are abundant at membranes of the ERGIC (29, 64), which is the site of SARS-CoV assembly (57), we did not detect any of the other ERGIC components identified in a recent proteomics survey (6). Overall, integral membrane proteins and membrane-associated proteins are underrepresented in the results of our analysis. While we were unable to exclude the possibility that low membrane protein detection was due to purely technical reasons, such as low solubility, limited protease accessibility, or paucity of trypsin-cleavable fragments of appropriate molecular weight, our results would appear to corroborate a previous observation that M protein networks can exclude host proteins that are present at the site of assembly from the viral membrane (9).

**Novel viral proteins identified.** We identified three new incorporated SARS-CoV proteins (nsp2, nsp3, and nsp5) in PK virus samples. We were also able to confirm that the ORF9b protein is incorporated, as was suspected from previously published results for the MHV I protein (13). We are unable to determine from the present results whether nsp2 and nsp3 were incorporated as a polyprotein. We were also unable to exclude the possibility that the nsp's were associated with other membrane-bound structures that copurified with virus. However, the best available evidence suggests that the viral replicase proteins, with the exception of nsp1, colocalize at the site of replication (42, 63), and no viral structure has yet been described which is specifically enriched in nsp2, nsp3, and nsp5. Therefore, we interpret these results to indicate that the nsp's identified by mass spectrometry proteomics analysis were incorporated in virions.

Two viral proteinases, nsp3 and nsp5, were detected in the virion. Finding the 1,922-amino-acid, multiple-membrane-spanning nsp3 in the virion was especially unexpected. nsp3 is best known for the presence of the highly conserved second papain-like cysteine proteinase (PL2<sup>PP1C</sup>) and ADRP, which together comprise about one-third of the mass of nsp3. But how was it incorporated in the virion? The lack of evidence for incorporated SARS-CoV polymerase, helicase, and nuclease proteins would appear to rule out efficient copurification of replicase complexes or double-membraned replicase vesicles as a source of nsp3. At the time at which this work was started,

no protein-protein interactions involving nsp3 had been reported, except for PL2<sup>PRO</sup>-mediated cleavage of polyubiquitin substrates (33). The relative abundance of nsp3 (as estimated by the frequency of detection and the percent coverage) appeared to be greater than that of nsp2, ORF3a, ORF6, ORF7a, and ORF9b proteins, all of which are reported to interact with nsp3 (65). This observation suggests that viral protein-protein interaction is probably not the primary mechanism of nsp3 incorporation.

We recently reported that the UB1 domain of nsp3 binds a discrete ssRNA species and that the adjacent AC domain binds bacterial dsDNA (53). Here we now report that two additional domains of SARS-CoV nsp3, i.e., SUD and NAB, exhibit distinct nucleic acid-binding characteristics. MBD incorporates a metal ion-binding site which may mediate RNA binding, and NAB exhibits energy-independent double-stranded nucleic acid unwinding properties, which would be consistent with nucleic acid chaperone activity. The presence of conserved cysteine/histidine clusters between the putative transmembrane domains (ZF) and at the amino terminus of the Y domain (Y1) may signal the presence of additional MBDs, which could increase the total number of nucleic acid-binding domains in nsp3 to six, i.e., UB1, AC, SUD, NAB, ZF, and Y1. RNA-binding proteins were also abundant among the detected host proteins and may have been packaged with genomic RNA or with incorporated host mRNAs. However, we also note that several other putative or confirmed SARS-CoV RNA-binding proteins were not detected in this study, perhaps suggesting that nsp3 has a specialized role in virogenesis, as was previously suggested for the PL<sup>PRO</sup>-containing nsp1 of equine arteritis virus (62).

The mechanism of putative nsp3 incorporation in purified SARS-CoV preparations remains unclear. Despite the presence of a host-derived envelope on each virion, host integral membrane and membrane-associated proteins comprised only 1% and 7%, respectively, of the detected proteins. A much higher percentage of viral integral and membrane-associated proteins was detected, including M, S, nsp3, and the ORF3a and ORF9b proteins. Therefore, while nucleic acid binding properties may have contributed to the disproportionate detection of nsp3 in purified virions relative to the adjacent proteolytic products of pp1a, the transmembrane region of nsp3 may also have contributed to incorporation.

The recent structural characterization of twin ubiquitin-related domains near the amino terminus of nsp3 (53) and the present demonstration of the presence of additional RNA-binding domains in nsp3 have implications for reconstructing the path of nidovirus replicase evolution. The *Nidovirales* encompass the coronaviruses, toroviruses, arteriviruses, and roniviruses. Replicase polyproteins from these viruses share conserved domains and common transcriptional mechanisms. As shown in Fig. 2A, coronaviruses from groups I, II, and III contain an initial UB1 homolog, followed by either a functional (group I and IIa) or a vestigial (group III) PL1<sup>PRO</sup> domain lacking the catalytic histidine found in functional PL1<sup>PRO</sup> and PL2<sup>PRO</sup>. We interpret the existence of paired UB and PL<sup>PRO</sup> domains as favoring an evolutionary model in which a prototypic PL<sup>PRO</sup> gene was duplicated in the last common ancestor of coronaviruses and subsequent loss of PL1<sup>PRO</sup> occurred in some coronavirus lineages (73). Two possible mechanisms for the

duplication of the UB and PL<sup>PRO</sup> domains are duplication of a gene cassette by direct repeat, as observed on a smaller scale in the AC domain of HCoV-HKU1 nsp3 from various isolates (Fig. 2) (70), and a recombination event between viruses with distinct UB and PL<sup>PRO</sup> domains prior to the divergence of the known coronavirus lineages.

We identified the novel nucleic acid binding domains SUD and NAB, which are located downstream of UB and PL<sup>PRO</sup> homologs. SUD and NAB share no detectable sequence homology. This does not necessarily preclude a structural relationship, since sequence-based criteria did not predict the structural homology between the two ubiquitin-related domains of nsp3, i.e., UB1 and UB2. Further investigation will be required to determine whether SUD and NAB domains are the result of duplication of a putative UB-PL<sup>PRO</sup>-nucleic acid binding protein gene cassette or became embedded in nsp3 independently.

**Functional implications.** There is growing evidence that the function of nsp3 is closely tied to association with nucleic acids. We would hypothesize that the functions of the UB1, AC, SUD, NAB, ADRP, and PL2<sup>PRO</sup> domains could be coordinated on a complex of protein and single-stranded and double-stranded RNA, such as the viral replicative form RNA (recently reviewed in reference 50). The character of coronavirus RNA replicase activity changes from an early, unstable form associated with discontinuous negative-strand synthesis to a later form associated with positive-strand synthesis (49). These observations suggest that PL<sup>PRO</sup>-mediated cleavage of the coronavirus polyprotein or other substrates such as polyubiquitin or poly(ADP-ribose), may drive the shift from coronavirus negative-sense to positive-sense RNA synthesis. A possible location for this activity would be the template-switching hot spots mapped to complementary sequences near the 5' genomic terminus and 3' antigenomic terminus (72). PL<sup>PRO</sup> involvement in viral RNA synthesis has a precedent in arterivirus, which requires the multifunctional, multidomain papain-related proteinase nsp1 for subgenomic RNA transcription but not for replication (62).

Although the work presented here represents primarily a starting point for detailed exploration of the overall function of coronavirus nsp3, we already note similarities to eukaryotic poly(ADP-ribose) polymerase (PARP) enzymes. Activated PARP consumes NAD<sup>+</sup> to synthesize a polymer of ADP-ribose that is covalently linked to a target protein (reviewed in reference 52). The best-characterized member of the family is PARP-1, which initiates the repair of nicked DNA through two N-terminal nucleic acid binding zinc fingers and is auto-poly-(ADP-ribosyl)ated on a glutamic acid-rich domain. PARPs can contain multiple adaptor domains preceding a conserved C-terminal catalytic domain, which in some cases includes one or more H2A macrodomains homologous to the ADRP of nsp3, and also nucleic acid binding domains. The ADRP domain of SARS-CoV nsp3 has been shown to strongly bind poly(ADP-ribose), and analysis of the nsp3 Y region yielded several Fold and Function Assignment System (FFAS) (24) hits on viral RNA-dependent RNA polymerase and both prokaryotic and eukaryotic DNA-dependent RNA polymerase domains (based on FFAS confidence scores of -9.5 and lower; data not shown). If nsp3 did indeed contain a functional PARP domain, it could obviously function in proofreading, genome repair, or

nidovirus-specific discontinuous subgenomic RNA transcription.

#### ACKNOWLEDGMENTS

We thank Maneesh Yadav for assistance with mass spectrometry, Michael Nelson for assistance with the kinase experiments, Vanitha Subramanian for assistance with cloning and protein expression, and Alexander E. Gorbalenya for helpful discussions of nsp3 phylogeny and domain structure.

This work was supported by NIH grant AI-59799 (M.J.B.), NIH/NIAID contract HHSN266200400058C (P.K.), NIH P41 RR11823 (J.R.Y.) and U01 DE016267 (J.R.Y.), and by the Joint Center for Structural Genomics through the NIH/NIGMS grant U54-GM074898. Additional support was obtained for M.A.J. and P.S. through fellowships from the Canadian Institutes of Health Research and the Spanish Ministry of Science and Education and by the Skaggs Institute for Chemical Biology. Kurt Wüthrich is the Cecil H. and Ida M. Green Professor of Structural Biology at The Scripps Research Institute.

This is TSRI manuscript 19291.

#### REFERENCES

- Altschul, S. F., W. Gish, W. Miller, E. W. Myers, and D. J. Lipman. 1990. Basic local alignment search tool. *J. Mol. Biol.* **215**:403–410.
- Barabási, A. L., and Z. N. Oltvai. 2004. Network biology: understanding the cell's functional organization. *Nat. Rev. Genet.* **5**:101–113.
- Barretto, N., D. Jukneliene, K. Ratia, Z. Chen, A. D. Mesecar, and S. C. Baker. 2005. The papain-like protease of severe acute respiratory syndrome coronavirus has deubiquitinating activity. *J. Virol.* **79**:15189–15198.
- Bern, M., D. Goldberg, W. H. McDonald, and J. R. Yates III. 2004. Automatic quality assessment of peptide tandem mass spectra. *Bioinformatics* **20**(Suppl. 1):i49–i54.
- Bertini, I., and C. Luchinat. 1984. High spin cobalt(II) as a probe for the investigation of metalloproteins. *Adv. Inorg. Biochem.* **6**:71–111.
- Breuzza, L., R. Halbeisen, P. Jenö, S. Otte, C. Barlowe, W. Hong, and H. P. Hauri. 2004. Proteomics of endoplasmic reticulum-Golgi intermediate compartment (ERGIC)-32, a new cycling protein that interacts with human Erv46. *J. Biol. Chem.* **279**:47242–47253.
- Chatterjee, A., M. A. Johnson, P. Serrano, B. Pedrini, and K. Wüthrich. 2007. NMR assignment of the domain 513–651 from the SARS-CoV nonstructural protein nsp3. *Biomol. NMR Assign.* **1**:191–194.
- Chenna, R., H. Sugawara, T. Koike, R. Lopez, T. J. Gibson, D. G. Higgins, and J. D. Thompson. 2003. Multiple sequence alignment with the Clustal series of programs. *Nucleic Acids Res.* **31**:3497–3500.
- de Haan, C. A., H. Vennema, and P. J. Rottier. 2000. Assembly of the coronavirus envelope: homotypic interactions between the M proteins. *J. Virol.* **74**:4967–4978.
- Deretec, D., A. H. Williams, N. Ransom, V. Morel, P. A. Hargrave, and A. Arendt. 2005. Rhodopsin C terminus, the site of mutations causing retinal disease, regulates trafficking by binding to ADP-ribosylation factor 4 (ARF4). *Proc. Natl. Acad. Sci. USA* **102**:3301–3306.
- 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.
- Eng, J., A. L. McCormack, and J. R. Yates III. 1994. An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database. *J. Am. Soc. Mass Spectrom.* **5**:976–989.
- Fischer, F., D. Peng, S. T. Hingley, S. R. Weiss, and P. S. Masters. 1997. The internal open reading frame within the nucleocapsid gene of mouse hepatitis virus encodes a structural protein that is not essential for viral replication. *J. Virol.* **71**:996–1003.
- Fischer, F., C. F. Stegen, P. S. Masters, and W. A. Samsonoff. 1998. Analysis of constructed E gene mutants of mouse hepatitis virus confirms a pivotal role for E protein in coronavirus assembly. *J. Virol.* **72**:7885–7894.
- Fosmire, J. A., K. Hwang, and S. Makino. 1992. Identification and characterization of a coronavirus packaging signal. *J. Virol.* **66**:3522–3530.
- Godet, M., R. L'Haridon, J. F. Vautherot, and H. Laude. 1992. TGEV corona virus ORF4 encodes a membrane protein that is incorporated into virions. *Virology* **188**:666–675.
- Gorbalenya, A. E., L. Enjuanes, J. Ziebuhr, and E. J. Snijder. 2006. Nidovirales: evolving the largest RNA virus genome. *Virus Res.* **117**:17–37.
- Gorbalenya, A. E., E. V. Koonin, and M. M. Lai. 1991. Putative papain-related thiol proteases of positive-strand RNA viruses. Identification of rubi- and aphthovirus proteases and delineation of a novel conserved domain associated with proteases of rubi-, alpha- and coronaviruses. *FEBS Lett.* **288**:201–205.
- Harcourt, B. H., D. Jukneliene, A. Kanjanahaluethai, J. Bechill, K. M. Severson, C. M. Smith, P. A. Rota, and S. C. Baker. 2004. Identification of severe acute respiratory syndrome coronavirus replicase products and characterization of papain-like protease activity. *J. Virol.* **78**:13600–13612.
- Huang, C., N. Ito, C. T. Tseng, and S. Makino. 2006. Severe acute respiratory syndrome coronavirus 7a accessory protein is a viral structural protein. *J. Virol.* **80**:7287–7294.
- Huang, C., C. J. Peters, and S. Makino. 2007. Severe acute respiratory syndrome coronavirus accessory protein 6 is a virion-associated protein and is released from 6 protein-expressing cells. *J. Virol.* **81**:5423–5426.
- Ito, N., E. C. Mossel, K. Narayanan, V. L. Popov, C. Huang, T. Inoue, C. J. Peters, and S. Makino. 2005. Severe acute respiratory syndrome coronavirus 3a protein is a viral structural protein. *J. Virol.* **79**:3182–3186.
- Ivanov, K. A., T. Hertzog, M. Rozanov, S. Bayer, V. Thiel, A. E. Gorbalenya, and J. Ziebuhr. 2004. Major genetic marker of nidoviruses encodes a replicative endoribonuclease. *Proc. Natl. Acad. Sci. USA* **101**:12694–12699.
- Jaroszewski, L., L. Rychlewski, Z. Li, W. Li, and A. Godzik. 2005. FFAS03: a server for profile-profile sequence alignments. *Nucleic Acids Res.* **33**:W284–W288.
- Joseph, J. S., K. S. Saikatendu, V. Subramanian, B. W. Neuman, A. Brooun, M. Griffith, K. Moy, M. K. Yadav, J. Velasquez, M. J. Buchmeier, R. C. Stevens, and P. Kuhn. 2006. Crystal structure of nonstructural protein 10 from the severe acute respiratory syndrome coronavirus reveals a novel fold with two zinc-binding motifs. *J. Virol.* **80**:7894–7901.
- Kanjanahaluethai, A., Z. Chen, D. Jukneliene, and S. C. Baker. 2007. Membrane topology of murine coronavirus replicase nonstructural protein 3. *Virology* **361**:391–401.
- Kirchhausen, T. 2000. Three ways to make a vesicle. *Nat. Rev. Mol. Cell Biol.* **1**:187–198.
- Kolegraaf, K., P. Bostik, and A. A. Ansari. 2006. Characterization and role of lentivirus-associated host proteins. *Exp. Biol. Med.* (Maywood) **231**:252–263.
- Krijnse-Locker, J., M. Ericsson, P. J. Rottier, and G. Griffiths. 1994. Characterization of the budding compartment of mouse hepatitis virus: evidence that transport from the RER to the Golgi complex requires only one vesicular transport step. *J. Cell Biol.* **124**:55–70.
- Krogh, A., B. Larsson, G. von Heijne, and E. L. Sonnhammer. 2001. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *J. Mol. Biol.* **305**:567–580.
- Kuo, L., K. R. Hurst, and P. S. Masters. 2007. Exceptional flexibility in the sequence requirements for coronavirus small envelope protein function. *J. Virol.* **81**:2249–2262.
- Li, F. Q., H. Xiao, J. P. Tam, and D. X. Liu. 2005. Sumoylation of the nucleocapsid protein of severe acute respiratory syndrome coronavirus. *FEBS Lett.* **579**:2387–2396.
- Lindner, H. A., N. Fotouhi-Ardakani, V. Lytvyn, P. Lachance, T. Sulea, and R. Menard. 2005. The papain-like protease from the severe acute respiratory syndrome coronavirus is a deubiquitinating enzyme. *J. Virol.* **79**:15199–15208.
- Luo, C., H. Luo, S. Zheng, C. Gui, L. Yue, C. Yu, T. Sun, P. He, J. Chen, J. Shen, X. Luo, Y. Li, H. Liu, D. Bai, Y. Yang, F. Li, J. Zuo, R. Hilgenfeld, G. Pei, K. Chen, X. Shen, and H. Jiang. 2004. Nucleocapsid protein of SARS coronavirus tightly binds to human cyclophilin A. *Biochem. Biophys. Res. Commun.* **321**:557–565.
- Masters, P. S. 2006. The molecular biology of coronaviruses. *Adv. Virus Res.* **66**:193–292.
- McBride, C. E., J. Li, and C. E. Machamer. 2007. The cytoplasmic tail of the severe acute respiratory syndrome coronavirus spike protein contains a novel endoplasmic reticulum retrieval signal that binds COPI and promotes interaction with membrane protein. *J. Virol.* **81**:2418–2428.
- Mizutani, T., S. Fukushi, K. Ishii, Y. Sasaki, T. Kenri, M. Saijo, Y. Kanaji, K. Shirota, I. Kurane, and S. Morikawa. 2006. Mechanisms of establishment of persistent SARS-CoV-infected cells. *Biochem. Biophys. Res. Commun.* **347**:261–265.
- Narayanan, K., C. J. Chen, J. Maeda, and S. Makino. 2003. Nucleocapsid-independent specific viral RNA packaging via viral envelope protein and viral RNA signal. *J. Virol.* **77**:2922–2927.
- Neuman, B. W., B. D. Adair, C. Yoshioka, J. D. Quispe, G. Orca, P. Kuhn, R. A. Milligan, M. Yeager, and M. J. Buchmeier. 2006. Supramolecular architecture of severe acute respiratory syndrome coronavirus revealed by electron cryomicroscopy. *J. Virol.* **80**:7918–7928.
- Oostra, M., E. G. te Lintelo, M. Deijs, M. H. Verheije, P. J. M. Rottier, and C. A. M. de Haan. 2007. Localization and membrane topology of coronavirus nonstructural protein 4: involvement of the early secretory pathway in replication. *J. Virol.* **81**:12323–12336.
- Pedersen, I. R., and E. P. Konigshofer. 1976. Characterization of ribonucleoproteins and ribosomes isolated from lymphocytic choriomeningitis virus. *J. Virol.* **20**:14–21.
- 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.
- Raamsman, M. J., J. K. Locker, A. de Hooge, A. A. de Vries, G. Griffiths, H. Vennema, and P. J. Rottier. 2000. Characterization of the coronavirus mouse hepatitis virus strain A59 small membrane protein E. *J. Virol.* **74**:2333–2342.

44. Ramjeesingh, M., L. J. Huan, E. Garami, and C. E. Bear. 1999. Novel method for evaluation of the oligomeric structure of membrane proteins. *Biochem. J.* **342**:119–123.
45. Ratia, K., K. S. Saikatendu, B. D. Santarsiero, N. Barretto, S. C. Baker, R. C. Stevens, and A. D. Mesecar. 2006. Severe acute respiratory syndrome coronavirus papain-like protease: structure of a viral deubiquitinating enzyme. *Proc. Natl. Acad. Sci. USA* **103**:5717–5722.
46. Rost, B., G. Yachdav, and J. Liu. 2004. The PredictProtein server. *Nucleic Acids Res.* **32**:W321–W326.
47. 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* **13**:1665–1675.
48. Saikatendu, K. S., J. S. Joseph, V. Subramanian, B. W. Neuman, M. J. Buchmeier, R. C. Stevens, and P. Kuhn. 2007. Ribonucleocapsid formation of severe acute respiratory syndrome coronavirus through molecular action of the N-terminal domain of N protein. *J. Virol.* **81**:3913–3921.
49. Sawicki, S. G., and D. L. Sawicki. 1986. Coronavirus minus-strand RNA synthesis and effect of cycloheximide on coronavirus RNA synthesis. *J. Virol.* **57**:328–334.
50. Sawicki, S. G., D. L. Sawicki, and S. G. Siddell. 2007. A contemporary view of coronavirus transcription. *J. Virol.* **81**:20–29.
51. Schaecher, S. R., J. M. Mackenzie, and A. Pekosz. 2007. The ORF7b protein of severe acute respiratory syndrome coronavirus (SARS-CoV) is expressed in virus-infected cells and incorporated into SARS-CoV particles. *J. Virol.* **81**:718–731.
52. Schreiber, V., F. Dantzer, J. C. Ame, and G. de Murcia. 2006. Poly(ADP-ribose): novel functions for an old molecule. *Nat. Rev. Mol. Cell Biol.* **7**:517–528.
53. Serrano, P., M. A. Johnson, M. S. Almeida, R. Horst, T. Herrmann, J. S. Joseph, B. W. Neuman, V. Subramanian, K. S. Saikatendu, M. J. Buchmeier, R. C. Stevens, P. Kuhn, and K. Wüthrich. 2007. Nuclear magnetic resonance structure of the N-terminal domain of nonstructural protein 3 from the severe acute respiratory syndrome coronavirus. *J. Virol.* **81**:12049–12060.
54. Sharma, K., M. Surjit, N. Satija, B. Liu, V. T. Chow, and S. K. Lal. 2007. The 3a accessory protein of SARS coronavirus specifically interacts with the 5'UTR of its genomic RNA, using a unique 75 amino acid interaction domain. *Biochemistry* **46**:6488–6499.
55. Shi, S. T., and M. M. Lai. 2005. Viral and cellular proteins involved in coronavirus replication. *Curr. Top. Microbiol. Immunol.* **287**:95–131.
56. Siddell, S. G., A. Barthel, and V. ter Meulen. 1981. Coronavirus JHM: a virion-associated protein kinase. *J. Gen. Virol.* **52**:235–243.
57. Stertz, S., M. Reichelt, M. Spiegel, T. Kuri, L. Martinez-Sobrido, A. Garcia-Sastre, F. Weber, and G. Kochs. 2007. The intracellular sites of early replication and budding of SARS-coronavirus. *Virology* **361**:304–315.
58. Strecker, T., R. Eichler, J. Meulen, W. Weissenhorn, H. D. Klenk, W. Garten, and O. Lenz. 2003. Lassa virus Z protein is a matrix protein sufficient for the release of virus-like particles. *J. Virol.* **77**:10700–10705.
59. Surjit, M., R. Kumar, R. N. Mishra, M. K. Reddy, V. T. Chow, and S. K. Lal. 2005. The severe acute respiratory syndrome coronavirus nucleocapsid protein is phosphorylated and localizes in the cytoplasm by 14-3-3-mediated translocation. *J. Virol.* **79**:11476–11486.
60. Tabb, D. L., W. H. McDonald, and J. R. Yates III. 2002. DTASelect and Contrast: tools for assembling and comparing protein identifications from shotgun proteomics. *J. Proteome Res.* **1**:21–26.
61. Thiel, V., K. A. Ivanov, A. Putics, T. Hertzog, 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.
62. Tijms, M. A., D. D. Nedialkova, J. C. Zevenhoven-Dobbe, A. E. Gorbalenya, and E. J. Snijder. 2007. Arterivirus subgenomic mRNA synthesis and virion biogenesis depend on the multifunctional nsp1 autoprotease. *J. Virol.* **81**:10496–10505.
63. van der Meer, Y., E. J. Snijder, J. C. Dobbe, S. Schleich, M. R. Denison, W. J. Spaan, and J. K. Locker. 1999. Localization of mouse hepatitis virus non-structural proteins and RNA synthesis indicates a role for late endosomes in viral replication. *J. Virol.* **73**:7641–7657.
64. Volpicelli-Daley, L. A., Y. Li, C. J. Zhang, and R. A. Kahn. 2005. Isoform-selective effects of the depletion of ADP-ribosylation factors 1–5 on membrane traffic. *Mol. Biol. Cell* **16**:4495–4508.
65. von Brunn, A., C. Teepe, J. C. Simpson, R. Pepperkok, C. C. Friedel, R. Zimmer, R. Roberts, R. Baric, and J. Haas. 2007. Analysis of intraviral protein-protein interactions of the SARS coronavirus ORF6. *PLoS ONE* **2**:e459.
66. Washburn, M. P., D. Wolters, and J. R. Yates III. 2001. Large-scale analysis of the yeast proteome by multidimensional protein identification technology. *Nat. Biotechnol.* **19**:242–247.
67. Welsch, S., B. Muller, and H. G. Krausslich. 2007. More than one door—budding of enveloped viruses through cellular membranes. *FEBS Lett.* **581**:2089–2097.
68. Wheelan, S. J., A. Marchler-Bauer, and S. H. Bryant. 2000. Domain size distributions can predict domain boundaries. *Bioinformatics* **16**:613–618.
69. Wolters, D. A., M. P. Washburn, and J. R. Yates III. 2001. An automated multidimensional protein identification technology for shotgun proteomics. *Anal. Chem.* **73**:5683–5690.
70. Woo, P. C., S. K. Lau, C. C. Yip, Y. Huang, H. W. Tsoi, K. H. Chan, and K. Y. Yuen. 2006. Comparative analysis of 22 coronavirus HKU1 genomes reveals a novel genotype and evidence of natural recombination in coronavirus HKU1. *J. Virol.* **80**:7136–7145.
71. Woo, P. C., M. Wang, S. K. Lau, H. Xu, R. W. Poon, R. Guo, B. H. Wong, K. Gao, H. W. Tsoi, Y. Huang, K. S. Li, C. S. Lam, K. H. Chan, B. J. Zheng, and K. Y. Yuen. 2007. Comparative analysis of twelve genomes of three novel group 2c and group 2d coronaviruses reveals unique group and subgroup features. *J. Virol.* **81**:1574–1585.
72. Wu, H. Y., and D. A. Brian. 2007. 5'-proximal hot spot for an inducible positive-to-negative-strand template switch by coronavirus RNA-dependent RNA polymerase. *J. Virol.* **81**:3206–3215.
73. Ziebuhr, J., V. Thiel, and A. E. Gorbalenya. 2001. The autocatalytic release of a putative RNA virus transcription factor from its polyprotein precursor involves two paralogous papain-like proteases that cleave the same peptide bond. *J. Biol. Chem.* **276**:33220–33232.
74. Zúñiga, S., I. Sola, J. L. Moreno, P. Sabella, J. Plana-Durán, and L. Enjuanes. 2007. Coronavirus nucleocapsid protein is an RNA chaperone. *Virology* **357**:215–227.