Structure, Vol. 12, 341–353, February, 2004, ©2004 Elsevier Science Ltd. All rights reserved. DOI 10.1016/j.str.2004.01.016

The nsp9 Replicase Protein of SARS-Coronavirus, Structure and Functional Insights

Geoff Sutton,[1] Elizabeth Fry,[1] Lester Carter,[1,2] Sarah Sainsbury,[2] Tom Walter,[2] Joanne Nettleship,[2] Nick Berrow,[2] Ray Owens,[2] Robert Gilbert,[1] Andrew Davidson,[3] Stuart Siddell,[3] Leo L.M. Poon,[4] Jonathan Diprose,[2] David Alderton,[2] Martin Walsh,[5] Jonathan M. Grimes,[1,2] and David I. Stuart*[,1,2]

[1]Division of Structural Biology
The Henry Wellcome Building for Genomic Medicine
Oxford University
Roosevelt Drive
Oxford OX3 7BN
United Kingdom

[2]Oxford Protein Production Facility
The Henry Wellcome Building for Genomic Medicine
Oxford University
Roosevelt Drive
Oxford OX3 7BN
United Kingdom

[3]Department of Pathology and Microbiology
School of Medical Sciences
University of Bristol
University Walk
Bristol BS8 1TD
United Kingdom

[4]Department of Microbiology
The University of Hong Kong
Queen Mary Hospital
Pokfulam Road
Hong Kong
SAR ROC

[5]CRG BM14
ESRF
B.P.220
F-38043 Grenoble CEDEX
France

Summary

As part of a high-throughput structural analysis of SARS-coronavirus (SARS-CoV) proteins, we have solved the structure of the non-structural protein 9 (nsp9). This protein, encoded by ORF1a, has no designated function but is most likely involved with viral RNA synthesis. The protein comprises a single β-barrel with a fold previously unseen in single domain proteins. The fold superficially resembles an OB-fold with a C-terminal extension and is related to both of the two subdomains of the SARS-CoV 3C-like protease (which belongs to the serine protease superfamily). nsp9 has, presumably, evolved from a protease. The crystal structure suggests that the protein is dimeric. This is confirmed by analytical ultracentrifugation and dynamic light scattering. We show that nsp9 binds RNA and interacts with nsp8, activities that may be essential for its function(s).

Introduction

Severe acute respiratory syndrome (SARS) is a new disease of humans that emerged in Southern China in late 2002. The first manifestation of SARS is a febrile illness, with respiratory symptoms, headaches, and myalgia, followed by progression to acute respiratory distress and progressive respiratory failure (Peiris et al., 2003). The etiological agent of SARS is a coronavirus (Kuiken et al., 2003). Coronaviruses are enveloped, positive-strand RNA viruses that are commonly associated with enteric and respiratory disease (Ziebuhr and Siddell, 2002). The severity of SARS-CoV infection is unusual and probably reflects the introduction of an animal coronavirus into a susceptible human population. In the first outbreak of SARS in 2003, at least 8000 people were infected and there were over 750 fatalities (Donnelly et al., 2003). To date, SARS has been controlled using conventional measures such as rapid detection, infection control, isolation, quarantine, contact tracing, etc. Clearly, these measures cannot be sustained indefinitely or repeatedly, and there is an urgent need to elucidate the natural history and pathogenesis of SARS-CoV infection, as well as to develop improved diagnostic tests and specific antiviral drugs and vaccines. We have initiated a high-throughput strategy to determine the crystal structures of SARS-CoV proteins, to facilitate functional analyses, and to assist in the design of antiviral compounds. This is a test of the efficacy of focused structural genomics (Burley, 2000) in combating emerging diseases where rapid control measures are vital.

The SARS-CoV genome is positive-strand RNA of approximately 29,700 nucleotides. It is composed of at least 14 functional ORFs that encode three classes of proteins; structural proteins (the S, M, E, and N proteins), non-structural proteins involved in viral RNA synthesis (the nsp or replicase proteins), and proteins that are thought to be non-essential for replication in tissue culture but clearly provide a selective advantage in vivo (the nspX or accessory proteins) (Marra et al., 2003; Rota et al., 2003). In common with other coronaviruses, the expression of the SARS-CoV genome is mediated by translation of the genomic RNA and a set of subgenomic mRNAs (Thiel et al., 2003). These mRNAs are produced by a unique mechanism that involves discontinuous transcription during negative-strand RNA synthesis and involves *cis*-acting elements, known as transcription-associated sequences (Pasternak et al., 2001; Sawicki et al., 2001). Once synthesized, the coronavirus mRNAs are translated by a variety of mechanisms, including programmed (−1) ribosomal frameshifting, stop-start initiation, and leaky scanning. The virus replicase proteins are translated from the genomic RNA and are initially synthesized as large polyproteins that are exten-

*Correspondence: dave@strubi.ox.ac.uk

sively processed by virus-encoded proteinases to produce a functional replicase-transcriptase complex (Ziebuhr et al., 2000). The structural and accessory proteins are translated from the subgenomic mRNAs.

The SARS-CoV replicase gene has been shown, or is predicted, to encode multiple enzymatic functions (Snijder et al., 2003). These include an RNA-dependent RNA polymerase activity (RdRp, nsp12), a 3C-like serine proteinase activity (3CL^{pro}, nsp5, also known as the main proteinase M^{pro}), a papain-like proteinase activity (PL2^{pro}, nsp3), and a superfamily 1-like helicase activity (HEL1, nsp13). These types of proteins are common to the replicative machinery of many positive-strand RNA viruses. In addition, the replicase gene encodes proteins that have domains indicative of 3′-5′ exoribonuclease activity (ExoN homolog, nsp14), endoribonuclease activity (XendoU homolog, nsp15), adenosine diphosphate-ribose 1″-phosphatase activity (ADRP, nsp3), and ribose 2′ -O- methyl transferase activity (2′ -O- MT, nsp16). These functions are less common in positive-strand RNA viruses and may be related to the unique features of coronavirus replication and transcription. Finally, the replicase gene encodes another nine proteins for which there is little or no information on their structure or function. nsps 10, 4, and 16 have been implicated by genetic analysis in the assembly of a functional replicase-transcriptase complex (Siddell et al., 2001; S.S., unpublished data). nsp9 corresponds to a 12 kDa cleavage product (P1a-12) in the related mouse hepatitis virus (MHV) that is most prominent in discrete foci in the perinuclear region of infected cells, colocalized with other components of the viral replication complex (Bost et al., 1999). Crystal structures are available for the 3CL^{pro} of SARS-CoV (Yang et al., 2003), transmissible gastroenteritis virus (Anand et al., 2002), and human coronavirus 229E (Anand et al., 2003). The structure of nsp9 reported here is the first product of our high-throughput analysis. In addition, we have produced a number of other SARS-CoV proteins in pure soluble form and these have been used for the analysis of nsp9 interactions with other replicase components, demonstrating an interaction with nsp9. We have also investigated the possible function of nsp9 and found it to bind RNA.

Results

We have determined the structure of the SARS-CoV nsp9 protein as part of a structural genomics project within the Oxford Protein Production Facility (OPPF) that targets the proteins of SARS-CoV. Table 1 shows that of 21 targets initially selected 16, including nsp9, were successfully expressed as soluble products using a standardized high-throughput approach (see Experimental Procedures). In particular, nsp8 and the 3CL^{pro} (nsp5) were produced in large quantities in pure soluble form. Since these two proteins are implicated in a replicase complex that includes nsp9, we have used them in some of the experiments described below.

Description of the Structure

The *E. coli*-expressed protein product for nsp9 corresponds to residues 4118–4230 (nucleotides 12,616–12,954) of the ORF1a replicase polyprotein (the putative mature nsp9 protein), together with a 6-His tag, Gateway ATT site (a recombination site used in the Gateway cloning strategy), and a rhinovirus 3C protease cleavage sequence (in total an addition of 30 amino acids N-terminal to the 113 of nsp9). The numbering scheme used throughout is relative to the natural cleavage point. Two tetragonal crystal forms of the protein (unrelated to crystals reported by Campanacci et al., [2003]) were solved using MAD and molecular replacement methods (see Experimental Procedures). In the final model of crystal form I, with one molecule in the asymmetric unit, all the residues of nsp9 are well defined in the electron density map together with an additional nine residues that correspond to part of the N-terminal tag (see Figure 1A; Tables 1 and 2). This model was refined at a resolution of 2.8 Å to an R factor of 22.8% with an R free of 31.4%; it possesses reasonable stereochemistry and 77% of residues lie in the most favored region of the Ramachandran diagram (none are in disallowed regions). This structure was used to solve crystal form II (four molecules in the crystallographic asymmetric unit) by molecular replacement methods (see Experimental Procedures). In both crystal forms, there are common associations (via crystallographic or non-crystallographic symmetry) of the molecule that form two distinct types of dimers. The core of the protein is an open 6-stranded β-barrel (see Figure 1B). The barrel comprises two antiparallel β sheets packed orthogonally (Figure 1B), forming a somewhat flattened barrel with shear number $S = 8$. Strands 1, 2, 3, and one half of 7 form one sheet, while a β-bulge extension from strand 1 and strands 4 and 5 form the second sheet. Strand 6 forms a tight β-hairpin with the section of strand 7, which extends out of the β-barrel. The curvature of the β strands combined with the long loops L_{45} and L_{67} gives the molecule the appearance of a boomerang, reminiscent of nucleic acid binding OB-fold proteins (Murzin, 1993; Theobald et al., 2003), although the fold of nsp9 is unrelated to the OB-fold. The first nine residues of the mature protein form, with the nine additional residues contributed by the N-terminal tag, a β-hairpin (Figure 1B). This extended structure has few interactions with the rest of the protein. Residues 96–110 form a C-terminal α helix that folds back antiparallel to strand 7 (Figure 1B).

nsp9 is structurally homologous to subdomains of serine proteases, in particular the second domain of the coronavirus 3CL^{pro}s (PDB codes 1Q2W, 1P9U, and 1P9S [Berman et al., 2000]) and the first domain of picornaviral 3CL^{pro}s (PDB codes 1CQQ and 1L1N [Berman et al., 2000]). Structural superposition of nsp9 (excluding the N-terminal tag residues) with the SARS-CoV 3Cl^{pro} domain II (1Q2W, residues 100–205) using SHP (Stuart et al., 1979) equivalences 71 residues with an rms deviation of 3.2 Å, with no significant insertions in either structure (Figures 2A and 2C). An alignment with domain I of the 3C protease from human rhinovirus 2 (HRV2) (1CQQ) gives 68 residues equivalenced with an rms deviation of 3.1 Å (Figure 2B). In comparison residues 3–184 of SARS-CoV M^{pro} and 1–180 of HRV2 3C protease can be superposed to equivalence 145 residues with an rms deviation of 2.9 Å.

Table 1. SARS-CoV Protein Expression Targets

Target Accession Number	Annotation	Amino Acid Residues in Construct	Soluble Expression in *E. coli* (Vector)	Baculovirus Expression
NSP1 NP_828860.1	Putative leader protein	1–180	+ (H) ++ (HG) +++ (HN)	
NSP2 NP_828861.1	MHV P65 homolog	1–639	+ (HG)	x
NSP3 DOMAIN NP_828862.1		190–340	+ (HN)	
NSP3 DOMAIN NP_828862.1		814–1031		x
NSP4 NP_828862.1	Contains transmembrane domain 2	1923–2422		
NSP5 NP_828863.1	3C-like proteinase	1–306	++ (H) + (HG)	x
NSP7 NP_828865.1		1–83	++ (HG) ++ (HN)	
NSP8 NP_828866.1		1–198	+++ (H) +++ (HG) +++ (HN)	x
NSP9 NP_828867.1		1–113	+++ (H) ++ (HG) +++ (HN)	x
NSP10 NP_828868.1		1–139	+ (HN)	x
NSP12 NP_828869.1	RNA-dependent RNA polymerase	1–932		
NSP12 DOMAIN NP_828869.1	RNA-dependent RNA polymerase	380–932		
NSP13 NP_828870.1	Zinc binding NTPase/ helicase	1–601	+ (HG) ++ (HN)	
NSP14 NP_828871.1	Putative ExoN-like nuclease	1–527		
NSP15 NP_828872.1	Putative XendoU-like endoRNAase	1–346	++ (H) ++ (HG) ++ (HN)	x
NSP16 NP_828873.2	Putative ribose 2′-O-methyltransferase	1–298	++ (HN) ++ (HG)	
NSP16 DOMAIN NP_828873.2	Putative ribose 2′-O-methyltransferase	1–213	+ (HG) + (HN)	x
SARS 3b NP_828853.1		1–155	+ (HN)	
SARS 6 NP_828856.1		1–64	+++ (HN)	
SARS 7a NP_828857.1		1–123	++ (HN)	x
SARS 9b NP_828859.1		1–99	++ (HG) ++ (HN)	x

x, expressed with correct molecular weight; +++, >5 mg/l; ++, 0.5–5 mg/l; +, 0.2–0.5 mg/l; H, pDEST17; HG, pDESTNHIS15 (modification of pDEST15); HN, pET44AGW (Gateway-adapted version of pET-43.1). Strain Rosetta PLysS.

Dimer Formation

Two structurally different dimers are observed in both of the two different crystal forms we have analyzed. In one of the dimers, the interface is principally formed by the parallel association of the C-terminal α helices (Figures 1C and 3A). This dimer has overall dimensions of 70 $\times$ 40 $\times$ 40 Å, and a total surface area of 1240 Å^2 per monomer is buried upon dimer formation. This surface area drops to 990 Å^2 on exclusion of the N-terminal tag (AREAIMOL [CCP4, 1994]). We would expect that in the absence of the N-terminal tag residues 1–3 at the mature N terminus may be poorly ordered but the dimer interface is likely to remain extensive. The two helices pack together at an angle of -28° but unusually closely (the closest approach of the helix axes is 5.4 Å). This close packing is possible because the heart of the dimer interface is formed from two glycines (Gly100 and Gly104) (the closest C$\tilde{\alpha}$-Cα distance between equivalent glycines across the dimer axis is 3.5 Å). The correlation coefficient that measures surface complementarity (Sc [CCP4, 1994; Lawrence and Colman, 1993]) for this dimer interaction surface is 0.71 (0.76 excluding the N-terminal tag) and 0.77 for the helices alone, which corresponds to a better shape matching than is observed in, for example, antibody-antigen interactions (Lawrence and Colman, 1993). Analysis of the sequence conservation across known coronaviruses (Figure 2C) reveals that the N and C termini of the protein are more conserved than the central core region, and the two key glycines are strictly conserved. Further stabilization of this hydrophobic interface arises from Leu4 and Ser5, which form part of the N-terminal extended β chain, clipping onto the edge of the inner β sheet (strand 6) from its dimer partner (see Figure 3A). This interaction effectively forms two six-stranded β sheets that run across the dimer interface, locking the dimer together. The N-terminal tag residues form an association with the end of the C-terminal helix, which may account for a kink in the α helices, bending them away from the dimer interface. This kinking, along with the extended β-hairpin that forms a tower (residues 74–90), results in a long groove that runs along the length of the dimer (Figure 3A). The base of this groove is rather hydrophobic in character, although the walls have some positive charged character. The external sides of the dimer are clearly more charged and present more accessible surfaces for interaction (Figure 3A).

The second dimer observed in the crystals (Figure 3B) is formed by an interaction between β strand 5 (residues 63–68) from both subunits zippering the two β-barrels together (Figure 2C). The surface area buried on dimer formation is only 540 Å^2 per monomer (the N-terminal tag does not participate in this interface), and the surface complementarity is 0.70. There is very little sequence conservation in residues involved in this dimer, but since the interactions involved are primarily main chain atoms this is perhaps not surprising. Although the surface area occluded on formation of the second type of dimer is

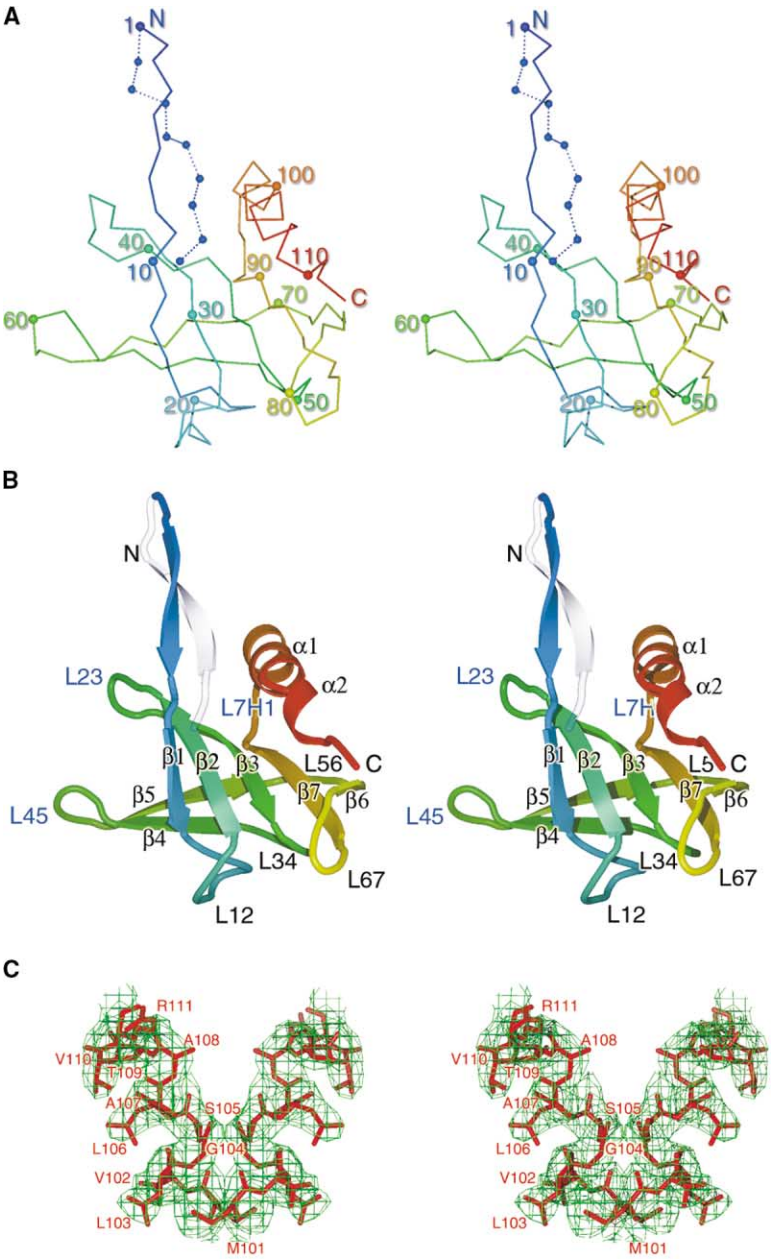

Figure 1. Structure of SARS-CoV nsp9

(A) A stereo α carbon trace colored blue to red from the N to the C terminus. The nine residues of the N-terminal tag are shown dashed. Every tenth residue is labeled.

(B) A stereo ribbon depiction colored as in (A) with the main secondary structure elements labeled according to Figure 2C. The N-terminal tag residues are shown transparent. Figures are produced using BOBSCRIPT (Esnouf, 1997) and RASTER3D (Merrit and Murphy, 1994).

(C) Stereo diagram of the $2F_o - F_c$ electron density for the C-terminal α helix residues 101–111, contoured at 1σ. The electron density is shown as a green mesh with the residues depicted in red ball-and-stick.

less than for the first, this dimer type is strictly maintained in both crystal forms. In contrast, in the larger cell with four copies of the monomer in the asymmetric unit, although one ordered copy of the helix dimer is present, in the second helix dimer one monomer is disordered, reflecting fluidity in the packing of the monomer along the helix axis.

Further Characterization of nsp9

Biophysical and functional experiments have been performed with the crystallized form of nsp9 (including the N-terminal tag) and with protein from which the N-terminal tag has been removed by treatment with human rhinovirus 3C protease (see Experimental Procedures).

Dynamic Light Scattering

Analysis of dynamic light scattering data (see Experimental Procedures) indicates that for concentrations above 1.5mg ml^{-1}, nsp9 is monodisperse with a Stokes' radius of approximately 2.1 nm (Figure 4A), which is in close agreement with the calculated radius of 1.9 nm for a dimer. At concentrations below 1 mg ml^{-1}, the Stokes' radius steadily decreases with decreasing concentration. This suggests that, at these lower concentrations, nsp9 is in a dynamic equilibrium between monomeric and dimeric forms, with the equilibrium favoring the monomeric species at the lowest concentrations.

Analytical Ultracentrifugation

Figure 4B shows the variation of apparent nsp9 weight (M_w) with concentration and centrifugation speed, for

Table 2. SARS-CoV nsp8, nsp9, nsp5 Expression Targets—Clones

Nsp8 fwd 5′-	ggggacaagtttgtacaaaaaagcaggcttc***ctggaagttctgttccagggcccg***GCTATTGCTTCAGAATTTAGTTCTTTACCATC-3′
(F5 clone)	T S L Y K K A G F *L E V L F Q G P* A I A S E F S S L P
Nsp8 rev 5′-	ggggaccactttgtacaagaaagctgggtctcaCTGTAGTTTAACAGCTGAGTTGGCTCTTAG-3′
(F5 clone)	* Q L K V A S N A R L
Nsp9 fwd 5′-	ggggacaagtttgtacaaaaaagcaggcttc***ctggaagttctgttccagggcccg***AATAATGAACTGAGTCCAGTAGCACTACGACAG-3′
(F4 clone)	T S L Y K K A G F *L E V L F Q G P* N N E L S P V A L R Q
Nsp9 rev 5′-	ggggaccactttgtacaagaaagctgggtctcaCTGAAGACGTACTGTAGCAGCTAAACTGCCC-3′
(F5 clone)	* Q L R V T A A L S G
M^{pro} fwd 5′-	ggggacaagtttgtacaaaaaagcaggcttc***ctggaagttctgttccagggcccg***AGTGGTTTTAGGAAAATGGCATTCCCG-3′
(F4 clone)	T S L Y K K A G F *L E V L F Q G P* S G F R K M A F P
M^{pro} rev 5′-	ggggaccactttgtacaagaaagctgggtctcaTTGGAAGGTAACACCAGAGCATTGTC-3′
(F4 clone)	* Q F T V G S C Q

Sequence of the primers used to amplify the coding regions for the nsp8, nsp9, and $3CL^{pro}$ proteins and the SARS-CoV clones used as template. The attB Gateway recombination sites are underlined, the rhinovirus 3C-protease cleavage site is in boldface italics, and the sequences that align to the SARS-CoV genes are in bold capitals (*, stop codon).

both tagged and untagged protein. For the tagged material, the trend in M_w with concentration for the lowest speed studied (12,000 rpm) indicates that self-association is occurring. The reduction in M_w of the tagged material when the speed is raised to 15,000 rpm indicates that there is substantial nonspecific aggregation of the protein. The behavior at low concentration and high speed shows that there is a specific self-association underlying the polydispersity at lower speeds and shows the molecule to be essentially monomeric in this regime. In contrast, at high concentration and high speed, nsp9 behaves as a dimer with approximate K_d of 6.0 $\pm$ 2.0 mg ml^{-1}, or 0.46 mM. The presence of a 6-His tag can lead to nonspecific aggregation. We therefore performed the same experiment using untagged material (Figure 4B). The measured values of M_w were similar at all three speeds for the untagged material, which shows that the polydisperse behavior observed for the tagged protein was due to the presence of the tag. However, the apparent dimerization observed in the tagged material at high speed is also present in the untagged material, with estimated K_d of 2.0 $\pm$ 0.5 mg ml^{-1}, or 0.16 mM. This value is indistinguishable from that obtained for the tagged protein, given the substantial experimental errors.

In order to try to define interaction partners for nsp9, we mixed equimolar proportions of nsp9 with nsp8, nsp5 (the 3C-like protease), and in combination with both. In addition, we examined nsp8 and the protease alone. The overall concentrations of these samples were in the region of 0.5 mg ml^{-1}. We performed sedimentation equilibrium experiments on these samples (Table 3) and analyzed them as described in the methods. As in the experiments reported above, nsp9 could be analyzed as if it were a monodisperse, ideal system with a raised molecular weight, indicating self-association with a time constant rapid on the timescale of the experiment (hence the ideal behavior). nsp8 consistently showed a weight in the region of 50 kDa, suggesting that it is constitutively a dimer. However, it showed non-ideal behavior, which may arise from the presence of an impurity such as a disordered form of the protein. The 3C-like protease had a weight of around 33 kDa, which suggests that at the concentrations used here it is monomeric (in agreement with published data [Yang et al., 2003]). Mixtures of nsp8 and nsp9 showed ideal behavior as opposed to the non-ideal behavior of nsp8 alone. However, binary mixtures of nsp8 and the 3C-like protease, or nsp9 and the 3C-like protease, showed non-ideal behavior, indicative of a mixture of non-interacting species. With all three species together, the data could be treated as ideal, presumably because of the complexity of the mixture. In summary, the nsp9 appeared to change the behavior of nsp8, suggesting that the two proteins interact. To investigate this further, we performed sedimentation velocity experiments and analyzed them using the time derivative g(s*) method (Stafford, 1992), which allows a model-independent analysis. Figure 4C shows g(s*) profiles for nsp9, nsp8, and a mixture of the two. nsp9 shows two peaks, presumably corresponding to the monomeric and dimeric forms of the protein. nsp8 alone shows a polydisperse profile in line with its non-ideal behavior as observed earlier. However, in the presence of nsp9, there is no evidence of the higher molecular weight species.

Membrane Interaction

Viral replication complexes are frequently membrane associated (Brockway et al., 2003; Egger et al., 2000; Sethna and Brian, 1997), although this is in general poorly understood. In order to investigate whether either nsp8 or nsp9 might be responsible for membrane interactions, coflotation experiments were conducted (see Experimental Procedures). These phase partitioning experiments showed that both nsp8 and cleaved nsp9 concentrated exclusively in the aqueous phase.

RNA Binding

As a putative component in the replication complex (Bost et al., 1999), nsp9 may possibly have an RNA binding activity. To investigate this possibility, electrophoretic mobility shift assays (EMSAs) were conducted with untagged nsp9 using both short and long RNA substrates. nsp9 binds to RNA as shown by the decrease in mobility of both a short (20-mer) oligoribonucleotide and longer (538 and 582 base) RNA substrates in a concentration dependent manner (Figure 4D). As observed in Figure 4D lanes 3–8, with a fixed concentration of nsp9 and decreasing amounts of the short oligoribonucleotide, the free RNA band reduced until all the RNA was shifted into an RNA-protein complex. At the higher RNA concentrations, the amount of

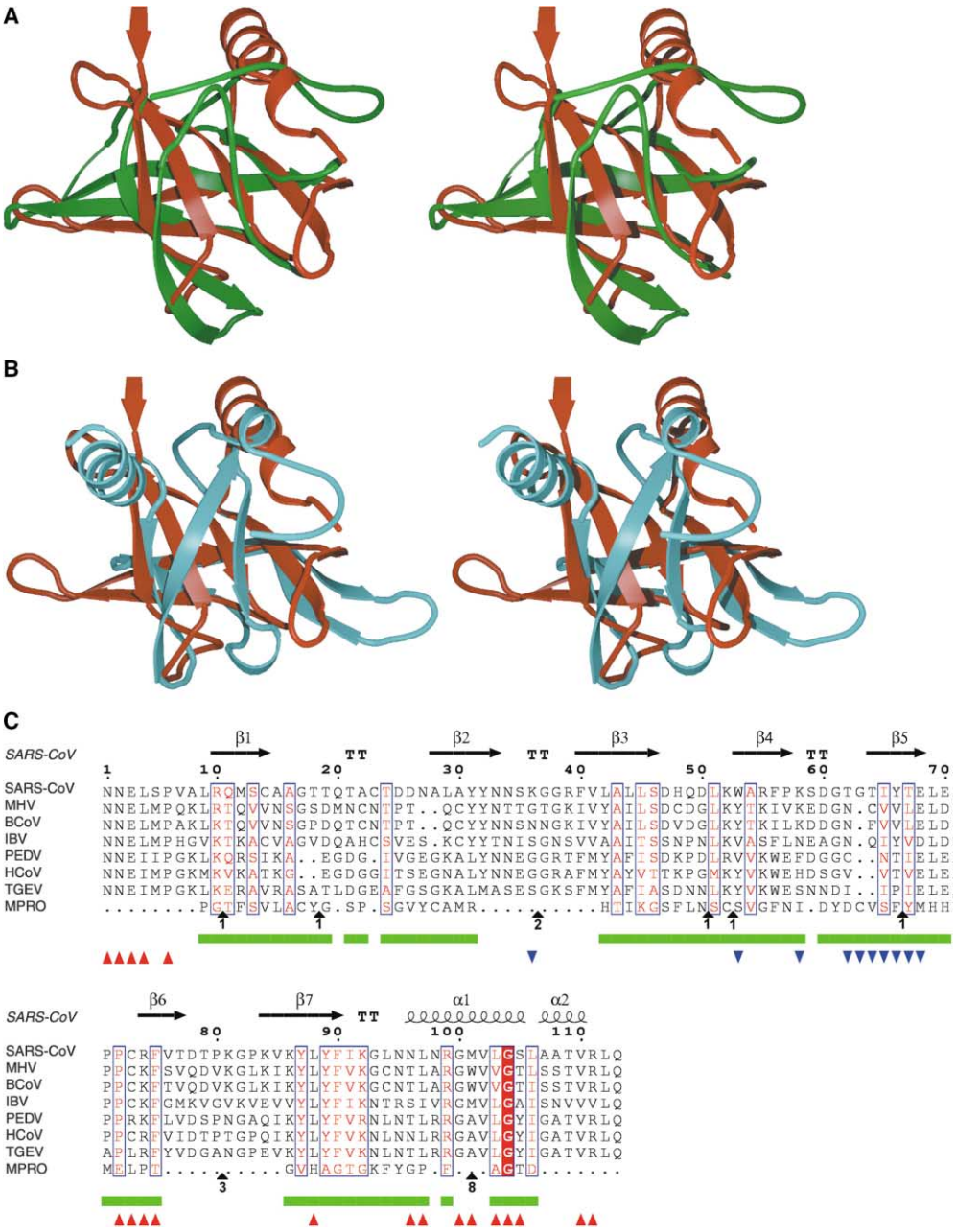

Figure 2. Similarity to Other Structures

(A) A stereo diagram of the SHP (Stuart et al., 1979) superposition of nsp9 (red ribbon) and oriented as in Figure 1, with the SARS CoV 3CL^{pro} domain II (residues 100–200) (PDB code 1Q2W [Berman et al., 2000]), shown in green.

(B) A stereo diagram of the SHP superposition of nsp9 (red ribbon), with the HRV2 3C protease domain I (residues 1–97) (PDB code 1CQQ [Berman et al., 2000]) (in blue).

(C) Sequence alignment (using CLUSTALW [Thompson et al., 1994]) of coronavirus proteins homologous to SARS-CoV nsp9: murine hepatitis virus (MHV) (NP_740614.1), bovine coronavirus (BCoV) (NP_742136.1), avian infectious bronchitis virus (IBV) (NP_740627.1), porcine epidemic diarrhea virus (PEDV) (NP_839963.1), human coronavirus 229E (HCoV) (NP_835350.1), and transmissible gastroenteritis virus (TGEV) (NP_840007). In addition, we have included domain II of SARS 3CL^{pro} as aligned structurally with nsp9 using SHP (Stuart et al., 1979). Aligned residues are marked by green bars, and to avoid breaking up the nsp9 sequence, residues for SARS 3CL^{pro} not matched are omitted and the position and number is indicated under the 3CL^{pro} sequence. The table is produced using ESpript (Gouet et al., 1999) with the secondary structure elements for SARS-CoV nsp9 assigned using DSSP. Residues boxed in red are completely conserved. Helix-dimer contacts are marked as red triangles and sheet-dimer contacts as blue triangles.

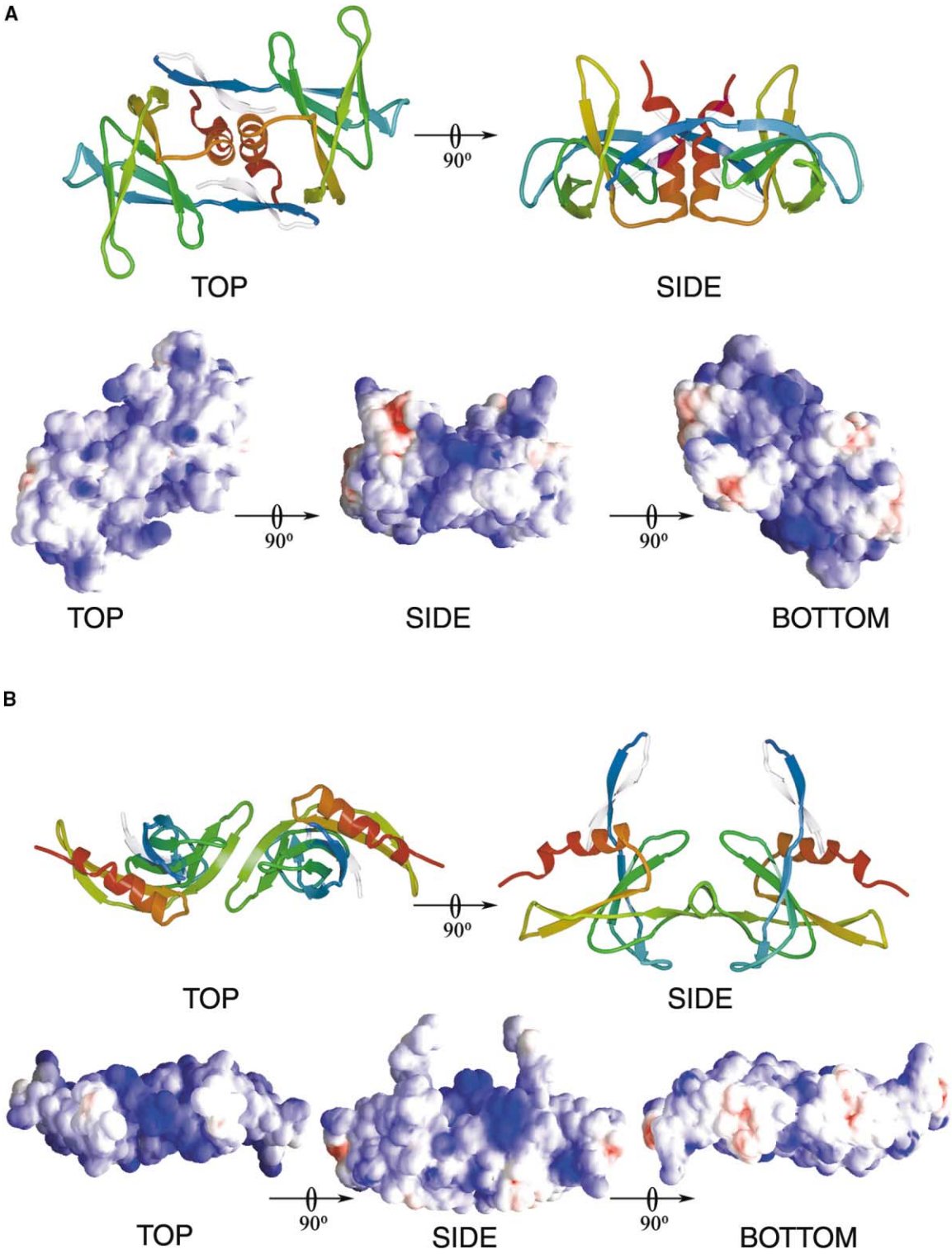

Figure 3. Dimer Structure(s)

(A) Orthogonal views of the helix-dimer, depicted as a ribbon and colored as in Figure 1B. The views are looking along the dimer 2-fold axis (TOP) and perpendicular to this axis (SIDE) following a rotation of 90° about the horizontal axis. Below are shown Grasp (Nicholls et al., 1991) depictions of the electrostatic potential mapped onto the accessible surface, orthogonal views as above plus a further 90° rotation about the horizontal axis (BOTTOM). The scale on which the electrostatic potential was colored was the same in each representation, with positive charge in blue and negative charge in red.

(B) Orthogonal views of the sheet-dimer, depicted as a ribbon and colored as in Figure 1B. The views are in the same relationship to the local 2-fold axis as those in (A). Below are shown Grasp (Nicholls et al., 1991) depictions of the electrostatic potential mapped onto the accessible surface. The scale for the electrostatic potential is the same as that for (A).

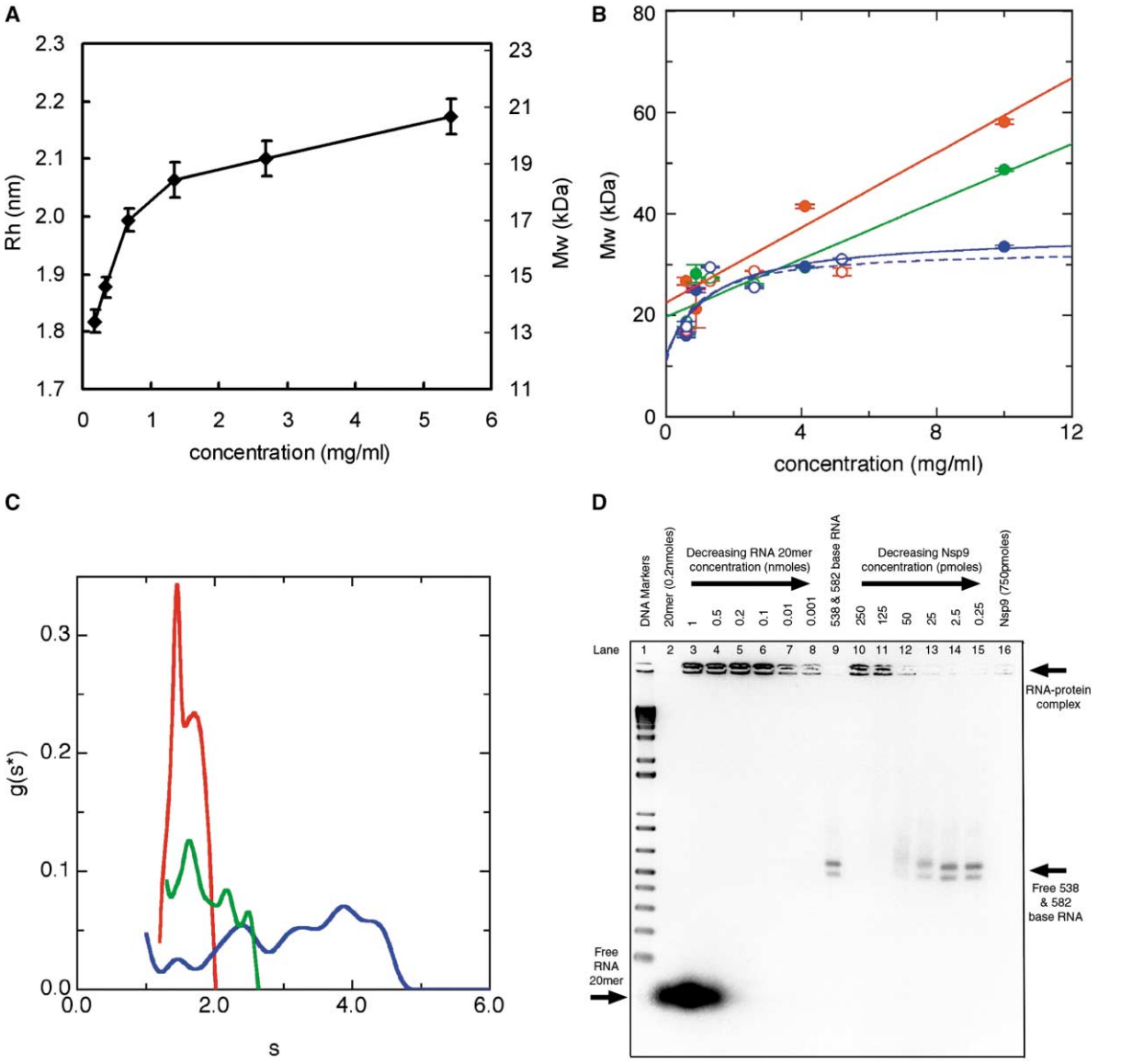

Figure 4. Characterization of nsp9

(A) Dynamic light scattering: the measured radius (and corresponding molecular weight [kDa]) is plotted against the concentration (mg ml^{-1}).

(B) Plots of apparent M_w against concentration for tagged nsp9 (closed symbols, solid lines) at 12,000 rpm (red symbols), 15,000 rpm (green), and 22,000 rpm (blue) derived from analytical AUC experiments. For untagged nsp9 (open symbols, colors as for tagged), only the plot at 22,000 rpm is shown (blue broken line) for clarity as the measured values were similar at all three speeds.

(C) G(s*) profiles of nsp9 (red), nsp8 (green), and an equimolar mixture of the two (blue) showing a change in the behavior of nsp8 on addition of nsp9. These g(s*) profiles were calculated with the same time-relative data for each sample using the second half of the experiment to increase the resolution of the analysis; however, using earlier scans did not alter their interpretation.

(D) Electrophoretic mobility shift of RNA by untagged nsp9. Lanes 2, 9, and 16 are controls of the individual components, RNA 20-mer, 538 and 582 base RNA, and nsp9, respectively. Lanes 3–8 have a constant amount of nsp9 (750 pmoles) with a decreasing concentration of RNA 20-mer. Lanes 10–15 have a constant quantity of 538 and 582 base RNA with a decreasing concentration of nsp9.

RNA-protein complex is constant, indicating that the nsp9 is saturated with an excess of RNA. Similarly, with a fixed amount of the longer RNA substrates, as the concentration of nsp9 was increased, the free RNA band reduced and the intensity of the RNA-protein complex band increased (Figure 4D, lanes 10–15). The RNA binding activity of nsp9 could only be competed out with heparin at heparin concentrations at least 5-fold higher than the protein concentration (data not shown).

Discussion

The structure of SARS-CoV nsp9 has a central core comprised of a six-stranded barrel, flanked by a C-terminal helix and N-terminal extension. The topology of the protein most closely resembles the domains of the chymotrypsin-like proteases (members of the serine protease superfamily), which have two domains comprising a six-stranded barrel motif (coronavirus prote-

Table 3. AUC Analysis of Protein Interactions

Sample	Mw, 15,000 rpm (Da)	Ideal?
Nsp9	254 nm = 24985 ± 2168	Yes
	280 nm = 24245 ± 1053	
	290 nm = 24924 = 1755	
Nsp8	254 nm = 48819 ± 1128	No
	280 nm = 43416 ± 512	
	290 nm = 54051 ± 790	
3C	254 nm = 30825 ± 1092	Yes
	280 nm = 36985 ± 586	
	290 nm = 34648 ± 758	
Nsp9 + Nsp8	254 nm = 39443 ± 490	Yes
	280 nm = 47774 ± 324	
	290 nm = 56738 ± 350	
Nsp8 + 3C	254 nm = 39980 ± 574	No
	280 nm = 40490 ± 387	
	290 nm = 40910 ± 434	
Nsp9 + 3C	254 nm = 35312 ± 320	No
	280 nm = 32085 ± 78	
	290 nm = 36729 ± 313	
Nsp9 + Nsp8 + 3C	254 nm = 30764 ± 981	Yes
	280 nm = 44920 ± 4428	
	290 nm = 27395 ± 1804	

ases have a third α-helical domain). Indeed, nsp9 represents the first example of a protein containing a single copy of this barrel motif. Structural alignments show the best match to domain II of the coronavirus $3CL^{pro}$s and subdomain I of the picornaviral $3CL^{pro}$s (with more distant similarity to the adjacent β-barrel domains in both cases). Thus, it would seem that an evolutionary relationship, based presumably on gene duplication processes within the genome of the SARS-CoV, exists, at least between the $3CL^{pro}$ and nsp9.

Both dynamic light scattering (DLS) and analytical ultracentrifugation (AUC) experiments on nsp9 (with and without the N-terminal tag), indicate that the molecule exists as a dimer in solution at mM concentrations. This agrees with an independent analysis of nsp9 in which a dimer was detected (Campanacci et al., 2003). In crystals of nsp9, we observe two possible dimers, one of which is presumably biologically relevant. The most extensive interaction is that mediated by helix packing. Although there are few specific interactions, there can be little doubt, given the hydrophobic nature of the interacting surface and the striking conservation of the amino acids involved, that this interaction is biologically important. The homophilic nsp9-nsp9 interaction could also be indicative of heterophilic protein-protein interactions. We have searched for GXXXG motifs in other proteins in the replicase complex that may interact with nsp9 but have been unable to identify candidates. However, the fluidity of packing via the hydrophobic surface as seen in the second crystal form suggests that this surface may play a generic role in nsp9 interactions with other proteins in the replicase complex. This dimer contains a narrow groove (defined largely by the "scissors"-like disposition of the interacting helices), which could conceivably accommodate a peptide. The second putative dimer is conserved in both crystal forms and involves an edge-to-edge interaction of β sheets that is frequently used to stabilize oligomers. Nevertheless, the lack of sequence conservation and limited area of interaction argue that this second dimer form may not be biologically relevant.

In cells infected by the related coronavirus MHV, nsp9 is localized in the perinuclear region, together with three other proteins of the replication complex (Bost et al., 1999). Also, for the MHV system, the polymerase (NSP12) has been shown to coimmunoprecipitate with $3CL^{pro}$ (nsp5), nsp8, and nsp9 (Brockway et al., 2003). For the SARS-CoV system, our AUC experiments suggest an interaction between nsp9 and nsp8 that may induce structural ordering of at least part of nsp8. This possibility is in line with PONDR analysis (Dunker at al., 2002) of nsp8 that strongly suggests that residues 43–84 and possibly the C-terminal region are disordered in the native protein. Protein partitioning experiments indicate that neither nsp9 nor nsp8 interact strongly with membranes and are thus unlikely to act as a membrane anchor for the replication-transcription complex.

nsp9 has no sequence motifs that suggest a biochemical function; for instance, it has none of the residues typically associated with the active site of serine proteases. However, in addition to their protease activity, the picornaviral 3C proteases bind RNA, forming a complex with the 5′-terminal 90 nucleotides of their RNA. This binding is mediated by a conserved RNA binding motif KFRDI (residues 82–86 HRV14) on the opposite face of the molecule to that which catalyzes proteolysis (Walker et al., 1995). This motif (which is not conserved in the SARSCoV-$3CL^{pro}$) is located in domain I and corresponds structurally to the beginning of helix 1 (94–97) in nsp9, a region rich in polar and hydrophobic residues. A second region of the picornavirus protease (residues 153–155 in HRV14) has also been implicated in binding RNA (Walker et al., 1995), and this corresponds structurally to β4 and L45 of nsp9, a region rich in basic amino acids. We have shown that nsp9 binds RNA and that this binding is not strongly RNA sequence specific (Figure 4D). RNA recognition motifs are generally rich in not only basic amino acids but also solvent-exposed hydrophobic side chains that make ionic and stacking interactions. Given this information, the most likely site of RNA binding is on the face of nsp9 that presents loops L23, L45, and L7H1 (See Figure 1B, where these loops are labeled in blue); this face is accessible in our preferred helix stabilized dimer but largely occluded in the putative β sheet-stabilized dimer. This presentation is reminiscent of that seen in OB-fold proteins that commonly bind oligonucleotides. Although their folds have different topologies, and are therefore not evolutionarily related, we find this a compelling case of convergent evolution to a similar architecture that reflects similar functions. This type of convergence in overall molecular shape is the other side of the coin to that observed for the classic case of convergent evolution, namely subtilisin and the tryspin-like serine proteases, where only the very local active site environment is reproduced.

In summary, our structural and functional analyses indicate that nsp9 may play multiple roles in the replicative cycle of coronaviruses. Its interaction with other proteins may be essential for the formation of the viral replication complex together with its ability to interact with RNA (in the absence of other viral or cellular proteins). The loops presented by the β-barrel may principally confer the RNA binding capacity via nonspecific interactions while the C-terminal β-hairpin and helix,

which display a greater conservation across coronaviruses are likely to be involved in dimerization and interaction with other proteins.

Finally, the structure of SARS-CoV nsp9 presented here has established that the Oxford Protein Production Facility pipeline for cloning, protein expression, purification, crystallization, structure determination, and functional characterization is in place. As can be seen from Table 1, out of 21 target constructs, 16 were expressed as soluble products in *E. coli* and 10 in a baculovirus system. The rapid progress from genome sequence (deposited at the NCBI on April 14, 2003) to X-ray structure (initial refinement completed on July 31, 2003) demonstrates that such high-throughput activities have the potential to contribute in a timely fashion to global health crises.

Experimental Procedures

RNA Isolation and Cloning

Total RNA was isolated using a QIAamp UltraSens Virus RNA extraction Kit (Qiagen) from 15 ml of tissue culture supernatant taken from SARS-CoV (strain HKU-39849, accession number AY278491, Zeng et al. [2003]) infected Vero E6 cells. Aliquots of the RNA were used as templates in one step RT-PCR reactions (SuperScript One-Step RT-PCR System for Long Templates; Invitrogen) to generate two cDNA products of 4851 bp (F4) and 6207 bp (F5) in size. The RT-PCR primers used to amplify F4 (5′- GTCATTTCATCAGCAATTCT TGGC-3′ [SARS-CoV nucleotides 7262–7285] and 5′-GAATCACC ATTAGCTACAGCCTGC-3′ [reverse primer; SARS-CoV nucleotides 12090-12113]) and F5 (5′-CAACTGAAGCTTTCGAGAAGATGG-3′ [SARS-CoV nucleotides 11906–11929] and 5′-GTCCTTTGGTATGCC TGGTATGTC-3′ [reverse primer; SARS-CoV nucleotides 18090–18113]) were designed from the sequence of the TOR2 strain of SARS-CoV (accession number AY274119; Marra et al., [2003]). The RT-PCR products were blunt end cloned into pBluescript SKII$^+$ to produce the clones SARS F4 and SARS F5, and the sequence verified using SARSCoV specific primers.

Protein Expression, Purification, and Characterization

The coding sequences for nsp8, nsp9, and 3CL^{pro} were amplified by PCR using the primers and clones described in Table 2. The forward primers encode a rhinovirus 3C-protease cleavage site positioned N-terminal to the gene and both forward and reverse primers contain the attB site of the Gateway cloning system (Invitrogen). The PCR fragments were subcloned into the pDEST17 plasmid (Invitrogen), producing clones pD17-Nsp9, pD17-Nsp8, and pD17-3CL^{pro}, which contain the full-length gene product with an N-terminal extension (MSYYHHHHHHLESTSLYKKAGFLEVLFQGP) including a 6-His tag for protein purification and a rhinovirus 3C-protease cleavage site for tag removal.

For expression of native protein, the pD17 plasmids were transformed into *E. coli* strain Rosetta PLysS (Novagen). Cultures were grown in GS-96 media (Qbiogene) with 1% glucose at 310 K until an OD_{620} of 0.6 was reached, and then cooled to 293 K for 30 min. Expression was induced by the addition of 0.5 mM IPTG, and the cultures were grown for a further 20 hr at 293 K. Seleno-methionine derivatized protein was produced by transforming the pD17 plasmid into the auxotrophic strain *E. coli* B834(DE3). Cells were cultured in SelenoMet Media (Molecular Dimensions Limited) according to the manufacturer's instructions up to the point of induction when the cultures were cooled to 293 K for 30 min, induced by the addition of IPTG to 0.5 mM, and grown for a further 20 hr at 293 K. Both native and seleno-methionine derivatized protein were purified as follows. The cells were harvested by centrifugation at 12,000 $\times$ g for 30 min and the bacterial pellets resuspended in 50 mM Tris-HCl, 500 mM NaCl (pH 7.5) (TN). Tween-20 was added to 1% and imidazole to 20 mM, and the cells lysed by sonication. The sample was clarified by centrifugation at 20,000 $\times$ g for 30 min and the supernatant loaded on to a Ni^{2+} charged 5 ml HiTrap-Chelating HP column (Amersham Biosciences). After washing with 20 column volumes of TN plus 20 mM imidazole, the protein was eluted with TN plus 500 mM imidazole. The eluate was applied to a Superdex 200 size-exclusion column preequilibrated in 20 mM Tris-HCl, 200 mM NaCl (pH 7.5). Fractions containing pure protein were pooled and DTT was added to 2 mM. One hundred percent seleno-methionine incorporation was confirmed by mass spectroscopy.

Crystallization

Prior to crystallization, nsp9 was concentrated by ultrafiltration, the buffer was exchanged for 10 mM Tris-HCl, 100 mM NaCl, 2 mM DTT (pH 8.0), and the final protein concentration adjusted to 10 mg ml^{-1}. An initial crystallization screen of 480 conditions was carried out by the sitting drop vapor diffusion method for both native and seleno-methionine derivatized nsp9 with a 200 nl drop size (1:1 protein/precipitant ratio) using a Cartesian robot (Brown et al., 2003; Walter et al., 2003). Based on these results, further fine screens to optimize the crystals were performed on the Cartesian robot using the same drop size or multiples thereof. Crystals of native protein were optimized at 100 mM citrate/phosphate buffer (pH 3.0), 1.5 M ammonium sulfate while the crystallization condition for seleno-methionine derivatized nsp9 was in $\sim$100 mM Citrate/phosphate (pH 3.8), 20% PEG 8000.

Structure Determination and Analysis

Crystals were flash frozen at 100 K in mother liquor containing either 25% or 10% glycerol for the native and seleno-methionine derivatized nsp9, respectively. A MAD experiment was performed at beamline BM14 (ESRF, Grenoble, France) along with native data collection. Data were recorded on a MarCCD detector as described in Table 4 and processed using the HKL2000 suite of programs (Otwinowski and Minor, 1997). Subsequent programs were from the CCP4 suite (CCP4, 1994), unless separately referenced. For the MAD data, two selenium sites were found using SOLVE (Terwilliger and Berendzen, 1999) and SOLVE/RESOLVE (Terwilliger, 2000; Terwilliger and Berendzen, 1999) produced an interpretable map. The structure was built using O (Jones et al., 1991) and refined with CNS (Brunger et al., 1998) using all data to 2.8 Å resolution (Table 4). The data were sharpened to a model with an average main chain B factor of 15 $Å^2$ using XPLOR. The final R factor is 22.8% and the R free is 31.4%. The native crystal structure was subsequently solved by molecular replacement using AMORE (final correlation coefficient and R factor of 66% and 53% after rigid body fitting) (CCP4, 1994). The search model was one of the two possible dimers of nsp9 observed in the other crystal form (where the α helix forms the tight dimeric interface). The two dimers in the asymmetric unit have identical orientations and are related by a translation of (0, 0, 1/2), consistent with the native Patterson map. The four monomers were initially refined as rigid bodies using CNS (Brunger et al., 1998). Electron density maps reveal that the monomers in one dimer are well ordered (and identical to that observed in the SeMet crystal structure), but one of the monomers in the second dimer is disordered. The ordered monomer in this dimer retains the same crystal contacts that form the other putative dimer (via the clipping together of strands 5). No further refinement work was done on this crystal form.

Cleavage of the N-Terminal Tag

The engineered N-terminal tag contains a rhinovirus 3C protease cleavage site, the enzyme cutting after the Q in the LFQGP sequence. Native nsp9 was cleaved by adjusting the protein concentration to 0.5 mg/ml in a 20 mM Tris-HCl (pH 8.5), 500 mM NaCl, 2 mM DTT buffer and incubating with HRV 3C protease (with N-terminal His-tag) for 16 hr at 20°C. Cleaved nsp9 passed straight through a Ni^{2+} charged 1 ml HiTrap-Chelating HP column (Amersham Biosciences) with the cleaved tag, uncleaved nsp9, and rhinovirus 3C protease binding to the beads. As determined by mass spectroscopy, the cleaved nsp9 had a mass of 12600 $\pm$ 10 Da which is 44 Da larger than the mass calculated from the sequence (12556 Da). The difference in mass can be explained by the presence of two tightly bound sodium ions (2 $\times$ 23 Da). Unlike the tagged material, the cleaved form did not readily crystallize.

Table 4. Data Collection and Processing

	Se-Met			
	λ_{PEAK}	λ_{REM}	λ_{INF}	Native
Wavelength (Å)	0.97829	0.8856	0.99988	0.97848
Resolution (Å) (final shell)	2.8 (3.5–3.3)	2.8 (3.5–3.3)	2.9 (3.4–3.3)	2.8 (2.9–2.8)
Completeness (%)	100 (100)	100 (100)	100 (100)	99 (94)
Redundancy	24 (25)	17 (18)	20 (21)	14 (7)
Images processed	322	225	300	280
Unique reflections	3908	3887	3521	21668
I/σ(I)	20 (8)	17 (7)	27 (8)	19 (1)
Anomalous completeness				
($F > 2\sigma$(6–3 Å)) (%)	99	97	98	NA
R merge (%)	19 (48)	16 (40)	15 (42)	9.5
R anom (%)	6.0	5.2	4.6	NA
Structure Refinement Statistics				
Refinement resolution (Å) (final shell)		20.0–2.8 (2.9–2.8)		
R factor (final shell)		23.6 (38.5)		
R free (final shell)		29.4 (40.5)		
Rmsd bonds/angles (Å/°)		0.009/1.7		
Average B factor, main chain/side chain (Å^2)		50/58		

Se-Met data were collected from one crystal at three wavelengths around the selenium absorption edge at station BM14 of the ESRF. The space group is $P4_322$, a = b = 58.0 Å, c = 85.0 Å, with one molecule in the asymmetric unit. Data were processed to 2.8 Å and phase refinement performed to 3.3 Å, hence the final resolution range given. The figure of merit determined using the program SOLVE (Terwilliger and Berendzen, 1999) was 0.51 while that given by RESOLVE (Terwilliger, 2000) was 0.38 initially and 0.58 corrected. Numbers given in brackets are for the appropriate outer shell.
Native data were collected from one crystal on station BM14 of the ESRF. The space group is $P4_32_12$, a = b = 88.6 Å, c = 202.0 Å, with four molecules in the asymmetric unit.

Analytical UltraCentrifugation

Sedimentation equilibrium experiments were performed in Beckman Optima XL-I or XL-A analytical ultracentrifuges as previously described (Ikemizu et al., 2000). Samples of nsp9 (with and without the amino-terminal tag removed) were at a range of concentrations in 20 mM Tris, 75 mM NaCl (pH 8.0) buffer and data were collected using interference optics. The sample distributions were fitted with the program ULTRASPIN (Altamirano et al., 2001) using a single-species equation. Any non-ideal behavior manifests itself as increasing apparent whole-cell weight-average molecular weights (M_w) with increasing concentration. The values for M_w obtained were plotted against sample concentration over the full range studied and fitted with either a straight line or the equation

$$M_w = \frac{2M_1c}{K_d + c} + M_1$$

for dimerization, as appropriate. Here, M_1 was fixed at the known monomeric molecular weight of nsp9. K_d is the equilibrium constant of dissociation. The same procedures were followed for nsp8, the SARS-CoV 3C-like protease, and mixtures of these proteins, except that a concentration range was not covered for these cases. Because the mixtures might represent interacting systems or non-ideal systems of non-interacting species, if a single-species equation did not fit the data sufficiently and there was no evidence of aggregation, then we used a two-species model to fit the data, either using ULTRASPIN, as above, or (with absorbance data) using the curve-fitting package ProFit (QuantumSoft, Uetikon am See, Switzerland). In ProFit the equation

$$A(r) = A(r_F)\exp\left[\left(\frac{(1 - \bar{v}\rho)\omega^2}{2RT}\right)M(r^2 - r_F^2)\right] + E$$

was used for a single species fit (where A is absorbance, r denotes radius in cm, r_F a reference radius, $\bar{v}$ is the partial specific volume [in ml/g], ρ is solvent density [in g/ml], ω is angular momentum [in radians/s], R is the gas constant, T is the absolute temperature, M is the protein weight [in Da], and E is the baseline). For a two-species fit, the equation

$$A(r) = \phi\left\{A(r_F)\exp\left[\left(\frac{(1 - \bar{v}\rho)\omega^2}{2RT}\right)M_1(r^2 - r_F^2)\right]\right\} + (1 - \phi)\left\{A(r_F)\exp\left[\left(\frac{(1 - \bar{v}\rho)\omega^2}{2RT}\right)M_2(r^2 - r_F^2)\right]\right\} + E$$

(φ is the fraction of species 1 and M_1 and M_2 are the weights of species 1 and 2) was used.

Sedimentation velocity experiments were performed using a Beckmann XL-I with interference optics and analyzed using the g(s*) (time derivative) method (Stafford, 1992)

$$g(s^*)_t = \left(\frac{d\{c(r,t)/c_0\}}{dt}\right)\left(\frac{\omega^2t^2}{\ln(r_m/r)}\right)\left(\frac{r}{r_m}\right)^2$$

in the program SEDFIT (Schuck and Rossmanith, 2000). The g(s*) profiles were then fitted with Gaussian curves which describe the distribution of individual species in them using ProFit (see above).

To compare the experimental sedimentation behavior of nsp9 with that predicted from its structure, we followed a previously described protocol (Merry et al., 2003), computing bead models from the atomic coordinates with the program AtoB (Byron, 1997) and calculating the solution behavior of those models with the aid of SOLPRO (Garcia de la Torre et al., 1999). In comparing these values—for a model in vacuo—with those for the experimental data, we corrected the computed parameters with the equation

$$s_\delta = s_0\left[1 + \left(\frac{\delta}{\bar{v}\rho}\right)\right]^{\frac{1}{3}},$$

where δ is the hydration fraction, which was set at 0.3 g/g water (the generally recognized standard value), s_δ is sedimentation coefficient at that hydration, and s_0 is the anhydrous value. This then allowed us to calculate the expected Stokes' radius of the protein, to compare with light scattering measurements, using the equation

$$R_s = \frac{M(1 - \bar{v}\rho)}{N6\pi\eta s},$$

where R_s is Stokes' radius, M is weight (in Da), N is Avogadro's number, and η is viscosity.

Dynamic Light Scattering

A series of dynamic light scattering measurements were taken on nsp9 over a concentration range of 0.16–5.5 mg/ml on a Dynapro Microsampler (Protein Solutions) in 20 mM Tris-HCl (pH 8.0), 100 mM NaCl, and data analyzed using software supplied by Protein Solutions.

Electrophoretic Mobility Shift Assays

SARS-CoV-specific RNA was generated with the AmpliScribe T7 Transcription Kit (Epicentre Technologies) using as template the pDEST17 plasmid coding for the SARS-CoV nsp9, which were specifically cleaved so as to produce single-strand RNA transcripts 538 and 582 bases long. A 20-mer oligoribonucleotide (CGACUCAUG GACCUUGGCAG) was synthesized by Eurogentec. RNA was incubated with protein in 10 mM Tris-HCl (pH 8.0), 100 mM NaCl for 20 min. Heparin competition experiments were carried out as described above but with the addition of varying quantities of low molecular weight heparin (average molecular weight 3000). The samples were run on 2% Agarose gels and the RNA was visualized with Sybr Green II (Molecular Probes).

Protein Partitioning in an Aqueous Micellar Two-Phase System

The partitioning of proteins in an aqueous micellar two-phase system (AMTPS) was used to characterize the amphiphilic nature of nsp8 and nsp9 (Bordier, 1981; Tani et al., 1998). The proteins were partitioned by addition to a 2% solution of precondensed Triton X-114 (in 10 mM Tris-HCl [pH 8.0], 150 mM NaCl) (Bordier, 1981), the aqueous and detergent phases allowed to separate by incubation at 30°C for 10 min and centrifuged for 5 min at 300 × g at the same temperature. Aliquots of the two phases were collected and analyzed by SDS-PAGE. A range of proteins known to partition in either the aqueous or detergent phase were used as controls (Bordier, 1981; Tani et al., 1998).

Acknowledgments

The work was supported by the Medical Research Council (MRC) UK. We thank Prof. J.S.M. Peiris for supply of SARS-CoV. The analytical ultracentrifugation experiments were performed in the facility established by the Biotechnology and Biological Sciences Research Council and the Wellcome Trust in the Glycobiology Institute of the University of Oxford and managed by Russell Wallis. We are grateful to M. Zambon and R. Gopal for help with the isolation of SARS-CoV RNA and E. Mancini for help with data collection. J.M.G. is supported by the Royal Society and D.I.S. by the MRC. BM14 is supported by the UK research councils. This work is supported by the European Commission as SPINE, contract no. QLG2-CT-2002-00988 under the Integrated program "Quality of Life and Management of Living Resources."

Received: November 30, 2003
Revised: December 30, 2003
Accepted: January 6, 2004
Published online: January 16, 2004

References

Altamirano, M.M., Woolfson, A., Donda, A., Shamshiev, A., Briseno-Roa, L., Foster, N.W., Veprintsev, D.B., De Libero, G., Fersht, A.R., and Milstein, C. (2001). Ligand-independent assembly of recombinant human CD1 by using oxidative refolding chromatography. Proc. Natl. Acad. Sci. USA *98*, 2950–2952.

Anand, K., Palm, G.J., Mesters, J.R., Siddell, S.G., Ziebuhr, J., and Hilgenfeld, R. (2002). Structure of coronavirus main proteinase reveals combination of a chymotrypsin fold with an extra α-helical domain. EMBO J. *21*, 3213–3224.

Anand, K., Ziebuhr, J., Wadhwani, P., Mesters, J.R., and Hilgenfeld, R. (2003). Coronavirus main proteinase ($3CL^{pro}$) structure: basis for design of anti-SARS drugs. Science *300*, 1763–1767.

Berman, H.M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T.N., Weissig, H., Shindyalov, I.N., and Bourne, P.E. (2000). The Protein Data Bank. Nucleic Acids Res. *28*, 235–242.

Bordier, C. (1981). Phase separation of integral membrane proteins in Triton X-114 solution. J. Biol. Chem. *256*, 1604–1607.

Bost, A.G., Carnahan, R.H., Tao Lu, X., and Denison, M.R. (1999). Four proteins processed from the replicase gene polyprotein of mouse hepatitis virus colocalize in the cell periphery and adjacent to sites of virion assembly. J. Virol. *74*, 3379–3387.

Brockway, S.M., Clay, C.T., Lu, X.T., and Denison, M.R. (2003). Characterization of the expression, intracellular localization, and replication complex associated with the putative mouse hepatitis virus RNA-dependent RNA polymerase. J. Virol. *77*, 10515–10527.

Brown, J., Walter, T.S., Carter, L., Abrescia, N.G.A., Ariescu, A.R., Batuwangala, T.D., Bird, L.E., Brown, N., Chamberlain, P.P., Davis, S.J., et al. (2003). A procedure for setting up high-throughput nanolitre crystallization experiments. II. Crystallization results. J. Appl. Crystallogr. *36*, 315–318.

Brunger, A.T., Adams, P.D., Clore, G.M., DeLano, W.L., Gros, P., Grosse-Kunstleve, R.W., Jiang, J.S., Kuszewski, J., Nilges, M., Pannu, N.S., et al. (1998). Crystallography and NMR system: a new software suite for macromolecular structure determination. Acta Crystallogr. D Biol. Crystallogr. *54*, 905–921.

Burley, S.K. (2000). An overview of structural genomics. Nat. Struct. Biol. Suppl. *7*, 932–934.

Byron, O. (1997). Construction of hydrodynamic bead models from high resolution X-ray crystallographic or nuclear magnetic resonance data. Biophys. J. *72*, 408–415.

Campanacci, V., Egloff, M.-P., Longhi, S., Ferron, F., Rancurel, C., Salomoni, A., Durousseau, C., Tocque, F., Bremond, N., Dobbe, J.C., et al. (2003). Structural genomics of the SARS coronavirus: cloning, expression, crystallization and preliminary crystallographic study of the Nsp9 protein. Acta Crystallogr. D Biol. Crystallogr. *59*, 1628–1631.

CCP4 (1994). The CCP4 suite: programs for protein crystallography. Acta Crystallogr. D Biol. Crystallogr. *50*, 760–763.

Donnelly, C.A., Ghani, A.C., Leung, G.M., Hedley, A.J., Fraser, C., Riley, S., Abu-Raddad, L.J., Ho, L.M., Thach, T.Q., Chau, P., et al. (2003). Epidemiological determinants of spread of causal agent of severe acute respiratory syndrome in Hong Kong. Lancet *361*, 1761–1766.

Dunker, A.K., Brown, C.J., Lawson, J.D., Iakoucheva, L.M., and Obradovic, Z. (2002). Intrinsic disorder and protein function. Biochemistry *41*, 6573–6582.

Egger, D., Teterina, N., Ehrenfeld, E., and Bienz, K. (2000). Formation of the poliovirus replication complex requires coupled viral translation, vesicle production and viral RNA synthesis. J. Virol. *74*, 6570–6580.

Esnouf, R.M. (1997). An extensively modified version of MolScript which includes greatly enhanced coloring capabilities. J. Mol. Graph. *15*, 132–134.

Garcia de la Torre, J., Harding, S.E., and Carrasco, B. (1999). Calculation of NMR relaxation, covolume, and scattering-related properties of bead models using the SOLPRO computer program. Eur. Biophys. J. *28*, 119–132.

Gouet, P., Courcelle, E., Stuart, D.I., and Metoz, F. (1999). ESPript: multiple sequence alignments in Postscript. Bioinformatics *15*, 305–308.

Ikemizu, S., Gilbert, R.J.C., Fennelly, J.A., Collins, A.V., Harlos, K., Jones, E.Y., Stuart, D.I., and Davis, S.J. (2000). Structure and dimerization of a soluble form of B7-1. Immunity *12*, 51–60.

Jones, T.A., Zou, Y.-J., Cowan, S.W., and Kjeldgaard, M. (1991). Improved methods for building protein models in electron density maps and the location of errors in these models. Acta Crystallogr. A *47*, 110–119.

Kuiken, T., Fouchier, R.A., Schutten, M., Rimmelzwaan, G.F., van Amerongen, G., van Riel, D., Laman, J.D., de Jong, T., van Doornum, G., Lim, W., et al. (2003). Newly discovered coronavirus as the primary cause of severe acute respiratory syndrome. Lancet *362*, 263–270.

Lawrence, M.C., and Colman, P.M. (1993). Shape complementarity at protein/protein interfaces. J. Mol. Biol. *234*, 946–950.

Marra, M.A., Jones, S.J., Astell, C.R., Holt, R.A., Brooks-Wilson, A., Butterfield, Y.S., Khattra, J., Asano, J.K., Barber, S.A., Chan, S.Y., et al. (2003). The genome sequence of the SARS-associated coronavirus. Science *300*, 1399–1404.

Merrit, E.A., and Murphy, M.E.P. (1994). Raster3D version 2.0. A program for photorealistic molecular graphics. Acta Crystallogr. D Biol. Crystallogr. *50*, 869–873.

Merry, A.H., Gilbert, R.J.C., Shore, D.A., Royle, L., Miroshnychenko, M.V., Wormald, M.R., Harvey, D.J., Dwek, R.A., Classon, B.J., Rudd, P.M., et al. (2003). *o*-glycan sialylation and the structure of the stalk-like region of the T cell co-receptor CD8. J. Biol. Chem. *278*, 27119–27128.

Murzin, A.G. (1993). OB (oligonucleotide/oligosaccharide binding)-fold: common structural and functional solution for non-homologous sequences. EMBO J. *12*, 861–867.

Nicholls, A., Sharp, K.A., and Honig, B. (1991). Protein folding and association: insights from the interfacial and thermodynamic properties of hydrocarbons. Proteins *11*, 281–296.

Otwinowski, Z., and Minor, W. (1997). Processing of X-ray diffraction data collected in oscillation mode. Methods Enzymol. *276*, 307–326.

Pasternak, A.O., van den Born, E., Spaan, W.J., and Snijder, E.J. (2001). Sequence requirements for RNA strand transfer during nidovirus discontinuous subgenomic RNA synthesis. EMBO J. *20*, 7220–7228.

Peiris, J.S., Chu, C.M., Cheng, V.C., Chan, K.S., Hung, I.F., Poon, L.L., Law, K.I., Tang, B.S., Hon, T.Y., Chan, C.S., et al. (2003). Clinical progression and viral load in a community outbreak of coronavirus-associated SARS pneumonia: a prospective study. Lancet *361*, 1767–1772.

Rota, P.A., Oberste, M.S., Monroe, S.S., Nix, W.A., Campagnoli, R., Icenogle, J.P., Penaranda, S., Bankamp, B., Maher, K., Chen, M.H., et al. (2003). Characterization of a novel coronavirus associated with severe acute respiratory syndrome. Science *300*, 1394–1399.

Sawicki, D., Wang, T., and Sawicki, S. (2001). The RNA structures engaged in replication and transcription of the A59 strain of mouse hepatitis virus. J. Gen. Virol. *82*, 385–396.

Schuck, P., and Rossmanith, P. (2000). Determination of the sedimentation coefficient distribution by least-squares boundary modeling. Biopolymers *54*, 328–341.

Sethna, P.B., and Brian, D.A. (1997). Coronavirus genomic and subgenomic minus-strand RNAs copartition in membrane-protected replication complexes. J. Virol. *71*, 7744–7749.

Siddell, S., Sawicki, D., Meyer, Y., Thiel, V., and Sawicki, S. (2001). Identification of the mutations responsible for the phenotype of three MHV RNA-negative ts mutants. Adv. Exp. Med. Biol. *494*, 453–458.

Snijder, E.J., Bredenbeek, P.J., Dobbe, J.C., Thiel, V., Ziebuhr, J., Poon, L.L., Guan, Y., Rozanov, M., Spaan, W.J., and Gorbalenya, A.E. (2003). Unique and conserved features of genome and proteome of SARS-coronavirus, an early split-off from the coronavirus group 2 lineage. J. Mol. Biol. *331*, 991–1004.

Stafford, W.F.I. (1992). Boundary analysis in sedimentation transport experiments: a procedure for obtaining sedimentation coefficient distributions using the time derivative of the concentration profile. Anal. Biochem. *203*, 295–301.

Stuart, D.I., Levine, M., Muirhead, H., and Stammers, D.K. (1979). Crystal structure of cat muscle pyruvate kinase at a resolution of 2.6Å. J. Mol. Biol. *134*, 109–142.

Tani, H., Kamidate, T., and Watanabe, H. (1998). Aqueous micellar two-phase systems for protein separation. Anal. Sci. *14*, 875–888.

Terwilliger, T.C. (2000). Maximum likelihood density modification. Acta Crystallogr. D Biol. Crystallogr. *56*, 965–972.

Terwilliger, T.C., and Berendzen, J. (1999). Automated MAD and MIR structure solution. Acta Crystallogr. D Biol. Crystallogr. *55*, 849–861.

Theobald, D.L., Mitton-Fry, R.M., and Wuttke, D.S. (2003). Nucleic acid recognition by OB-fold proteins. Annu. Rev. Biophys. Biomol. Struct. *32*, 115–133.

Thiel, V., Ivanov, K.A., Putics, A., Hertzig, T., Schelle, B., Bayer, S., Weissbrich, B., Snijder, E.J., Rabenau, H., Doerr, H.W., et al. (2003). Mechanisms and enzymes involved in SARS coronavirus genome expression. J. Gen. Virol. *84*, 2305–2315.

Thompson, J.D., Higgins, D.G., and Gibson, T.J. (1994). Clustal W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res. *22*, 4673–4680.

Walker, P.A., Leong, L.E.-C., and Porter, A.G. (1995). Sequence and structural determinants of the interaction between the 5′-noncoding region of picornavirus RNA and rhinovirus protease 3C. J. Biol. Chem. *270*, 14510–14516.

Walter, T.S., Diprose, J., Brown, J., Pickford, M., Owens, R.J., Stuart, D.I., and Harlos, K. (2003). A procedure for setting up high-throughput nanolitre crystallization experiments. I. Protocol design and validation. J. Appl. Crystallogr. *36*, 308–314.

Yang, H., Yang, M., Ding, Y., Liu, Y., Lou, Z., Zhou, Z., Sun, L., Mo, L., Ye, S., Pang, H., et al. (2003). The Crystal structures of severe acute respiratory syndrome virus main protease and its complex with an inhibitor. Proc. Natl. Acad. Sci. USA *100*, 13190–13195.

Zeng, F.Y., Chan, C.W., Chan, M.N., Chen, J.D., Chow, K.Y., Hon, C.C., Hui, K.H., Li, J., Li, V.Y., Wang, C.Y., et al. (2003). The complete genome sequence of severe acute respiratory syndrome coronavirus strain HKU-39849 (HK-39). Exp. Biol. Med. (Maywood) *228*, 866–873.

Ziebuhr, J., and Siddell, S. (2002). *Nidovirales*. In The Encyclopaedia of Life Sciences (London: Stockton Press), pp. 190–198.

Ziebuhr, J., Snijder, E.J., and Gorbalenya, A.E. (2000). Virus-encoded proteinases and proteolytic processing in the Nidovirales. J. Gen. Virol. *81*, 853–879.

Accession Numbers

The coordinates have been deposited in the PDB database (Berman et al., 2000) with accession code 1UW7.