

The Nucleocapsid Protein of Coronavirus Infectious Bronchitis Virus: Crystal Structure of Its N-Terminal Domain and Multimerization Properties

Hui Fan,^{1,3,4} Amy Ooi,^{1,3,4} Yong Wah Tan,²
Sifang Wang,² Shouguo Fang,² Ding Xiang Liu,^{2,*}
and Julien Lescar^{1,3,*}

¹School of Biological Sciences
Nanyang Technological University
60 Nanyang Drive
Singapore 637551

²Institute of Molecular and Cell Biology
61 Biopolis Drive
Proteos
Singapore 138673

Summary

The coronavirus nucleocapsid (N) protein packages viral genomic RNA into a ribonucleoprotein complex. Interactions between N proteins and RNA are thus crucial for the assembly of infectious virus particles. The 45 kDa recombinant nucleocapsid N protein of coronavirus infectious bronchitis virus (IBV) is highly sensitive to proteolysis. We obtained a stable fragment of 14.7 kDa spanning its N-terminal residues 29–160 (IBV-N29-160). Like the N-terminal RNA binding domain (SARS-N45-181) of the severe acute respiratory syndrome virus (SARS-CoV) N protein, the crystal structure of the IBV-N29-160 fragment at 1.85 Å resolution reveals a protein core composed of a five-stranded antiparallel β sheet with a positively charged β hairpin extension and a hydrophobic platform that are probably involved in RNA binding. Crosslinking studies demonstrate the formation of dimers, tetramers, and higher multimers of IBV-N. A model for coronavirus shell formation is proposed in which dimerization of the C-terminal domain of IBV-N leads to oligomerization of the IBV-nucleocapsid protein and viral RNA condensation.

Introduction

Coronaviruses are large enveloped single-stranded RNA viruses of positive polarity which cause a wide spectrum of diseases affecting humans and animals (reviewed in Lai and Holmes, 2001). In 2003, the causative agent for the outbreak of atypical pneumonia with a high fatality rate was identified as the severe acute respiratory syndrome (SARS) coronavirus (SARS-CoV) (Peiris et al., 2003), and its genome was rapidly sequenced and characterized (Marra et al., 2003; Rota et al., 2003). The potential risks for public health posed by SARS-CoV and the current lack of specific antiviral agents or vaccines against this emerging pathogen have triggered a global research effort in order to characterize this fam-

ily of viruses at the molecular level. Coronavirus infectious bronchitis virus (IBV) causes an acute and contagious disease in chickens, with a significant impact on the poultry industry worldwide.

In structural terms, coronavirus virions are roughly spherical, with an approximate diameter of 120 nm. Their detailed in vivo morphology is still a matter of debate but might be composed of three structural layers: a lipid envelope with three or four glycoproteins, a protein core, and a tubular or helicoidal nucleocapsid, as shown for the porcine transmissible gastroenteritis virus (TGEV) (Escors et al., 2001). Low-resolution electron micrographs have highlighted the crown-like structure that surrounds the coronavirus envelope (Sturman et al., 1980). These spikes contain the S protein, a class I fusion glycoprotein (Bosch et al., 2004; Lescar et al., 2001) which is also responsible for binding to the receptor (Lai and Holmes, 2001). Two integral membrane proteins, M (about 230 amino acids) and E (about 100 amino acids), are essential for the maturation of newly formed virions, and are sufficient for the formation of a closed viral particle (Vennema et al., 1996). The M protein is thought to possess three transmembrane segments and a large C-terminal endodomain that interacts with the nucleocapsid and possibly also with the RNA genome (Sturman et al., 1980; Kou and Masters, 2002; Narayanan et al., 2003). The nucleocapsid protein of IBV (IBV-N) is a phosphoprotein of 409 amino acids that is well-conserved across various IBV strains (Williams et al., 1992) and is also important for cell-mediated immunity. It forms a protective shell that packages the viral genomic RNA of 27.6 kb and is also thought to participate in viral RNA replication and transcription. Specific packaging of viral genetic material is usually performed via the recognition of a particular nucleotide sequence by a nucleocapsid protein. Such “packaging signals” have been identified at the 3′ end of the viral genomes of mouse hepatitis virus (MHV) (Fosmire et al., 1992) and bovine coronavirus (BCV) (Cologna and Hogue, 2000) and at the 5′ end of the TGEV genome (Escors et al., 2003), but not unambiguously for the IBV genome. In elegant structural studies performed in other viral families with RNA genomes, such as HIV (De Guzman et al., 1998) and the MS2 bacteriophage (Valegard et al., 1997), the packaging signals were seen to form a stem-loop structure that is recognized by the nucleocapsid protein. In the case of the IBV genome, this special RNA structure has not been determined with certainty, although previous studies demonstrated that the IBV-N protein interacts specifically with RNA sequences located at the 3′ noncoding region of the viral genome (Zhou et al., 1996). Both the N- and C-terminal domains of IBV-N, but not its middle region, bind to an oligoribonucleotide of 155 nucleotides, located at the 3′ end of the viral genome nontranslated region, but little is known about the details of this interaction and how it relates to virus assembly (Zhou and Collisson, 2000).

In an attempt to define how the viral genome is incorporated into newly formed viral particles and how this process is coupled with nucleocapsid assembly, we

*Correspondence: julien@ntu.edu.sg (J.L.); dxliu@imcb.a-star.edu.sg (D.X.L.)

³Lab address: <http://www.ntu.edu.sg/sbs/staff/julien/labsite/Home.htm>.

⁴These authors contributed equally to this work.

have undertaken functional and structural studies using the full-length N protein from IBV expressed in bacteria. The full-length recombinant IBV-N protein expressed in *Escherichia coli* is unstable. Through cleavage by *E. coli* proteases, a stable fragment of 14.7 kDa comprising its N-terminal residues 29–160 can be obtained. We report here the crystal structure of this N-terminal fragment refined at 1.85 Å resolution and compare it to the N-terminal RNA binding domain of the SARS-CoV N protein (SARS-N45–181), which was solved recently using NMR (Huang et al., 2004). We demonstrate the formation of multimers of the IBV-N protein in vitro and propose that dimers and possibly tetramers of IBV-N, which are stabilized predominantly via dimerization of their C-terminal domains, act as elementary building blocks for RNA genome condensation and nucleocapsid assembly.

Results and Discussion

Identification of a Stable Proteolytic Fragment of IBV-N

The full-length IBV-N protein comprising 409 residues was expressed in *E. coli* in a soluble form and purified as described in the [Experimental Procedures](#). Crystallization trials with the full-length protein produced crystals that grew from a precipitate after about 3 months. Analysis of dissolved crystals using SDS-PAGE reveals that they contain a fragment of the full-length protein of about 14.7 kDa (Figure 1). A domain of similar size could be obtained by incubating the IBV-N protein at room temperature for the same period (Figure 1). Thus, this polypeptide fragment presumably derives from slow proteolysis of IBV-N by traces of *E. coli* proteases present in the crystallization solution. In order to identify its nature, this proteolytically stable fragment was subjected to mass spectrometry, which revealed a mass of 14,692 Da. N-terminal amino acid sequencing identified residues Ser-Ser-Gly-Asn-Ala-Ser-Trp, which are located at positions 29–35 of the IBV-N amino acid sequence. Given that Ser-29 is the first amino acid of the fragment, the closest mapping onto the sequence gives Leu-160 as the C-terminal residue (calculated mass 14,691 Da). The IBV-N29-160 protein shares 37% amino acid sequence identity with the N-terminal RNA binding domain of a comparable domain from the SARS-CoV N protein (SARS-N45–181), whose structure was reported recently (Figure 2).

Structure Determination and Quality of the Model

Overexpression of the recombinant N-terminal IBV-N29-160 fragment readily gave crystals diffracting beyond 2.0 Å (Figure 1). Attempts to solve the structure by molecular replacement using the averaged NMR structure of a SARS-CoV nucleocapsid N-terminal domain deposited in the Protein Data Bank (PDB) (Huang et al., 2004) were unsuccessful, even though the two structures turned out to adopt a related fold (Figure 2). The IBV-N29-160 protein is devoid of methionine and cysteine residues. Thus, in order to assist structure determination using the multiwavelength anomalous dispersion (MAD) method, Ile-62, Leu-104, and Val-116 were mutated to methionine. These hydrophobic amino acid residues have been shown to introduce little perturbation in the native protein structure when substituted by methionine

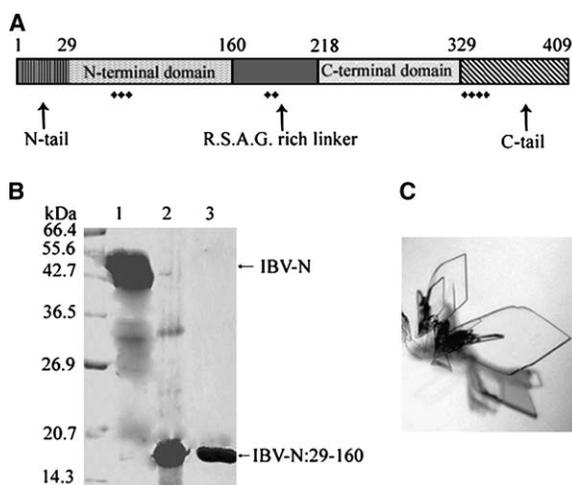


Figure 1. Structural Domains of the IBV-N Protein

(A) Schematic representation of the IBV-N protein depicting its various domains and clustering of positive charges, as inferred from the present and other studies.

(B) SDS-PAGE analysis of the full-length recombinant IBV-N protein of 44.9 kDa (lane 1, arrow) and the N-terminal proteolytically stable fragment of 14.7 kDa spanning residues 29–160 of the sequence which was crystallized (lane 2). The recombinant IBV-N29-160 is shown in lane 3.

(C) Typical plate-shaped crystals of the recombinant IBV-N29-160 protein.

residues (Gassner and Matthews, 1999). In addition, the presumably exposed residue Lys-85 (as suggested by an amino acid sequence alignment with the SARS-CoV N protein) was mutated to Cys in order to introduce a potential binding site for mercury compounds. This mutated fragment of IBV-N29-160 was used for structure determination using the MAD method with crystals containing the selenomethionyl protein. Data collection, phasing, and refinement statistics are summarized in [Tables 1 and 2](#) for the selenomethionine-derivatized crystal (SeMet) and for the native protein crystal. Overall, the path of the main chain is unambiguously defined in clear electron density for the two IBV-N29-160 molecules present in the asymmetric unit in each crystal form. A total of 134 protein residues per molecule (two extra residues at the N terminus derive from the cloning procedure) were included in the final models, which have excellent stereochemical parameters as well as 182 and 188 well-defined water molecules, respectively ([Table 2](#)). Electron density is absent for the Lys-81 side chain which is exposed to the solvent.

Overall Structure

The two monomers present in the asymmetric unit can be superimposed with a root mean square (rms) deviation of 0.5 Å for their main chain atoms. The IBV-N29-160 monomer has approximate overall dimensions of 35 Å × 35 Å × 30 Å and consists of a core formed by a five-stranded antiparallel β sheet with the topology β₄-β₂-β₃-β₁-β₅, which faces a smaller antiparallel sheet composed of only two strands, β₁-β₄, which are absent in the SARS-N protein (Figure 2). A long flexible hairpin loop β₂-β₃, which is inserted between the β₂ and β₃ strands, protrudes largely from the protein core. This

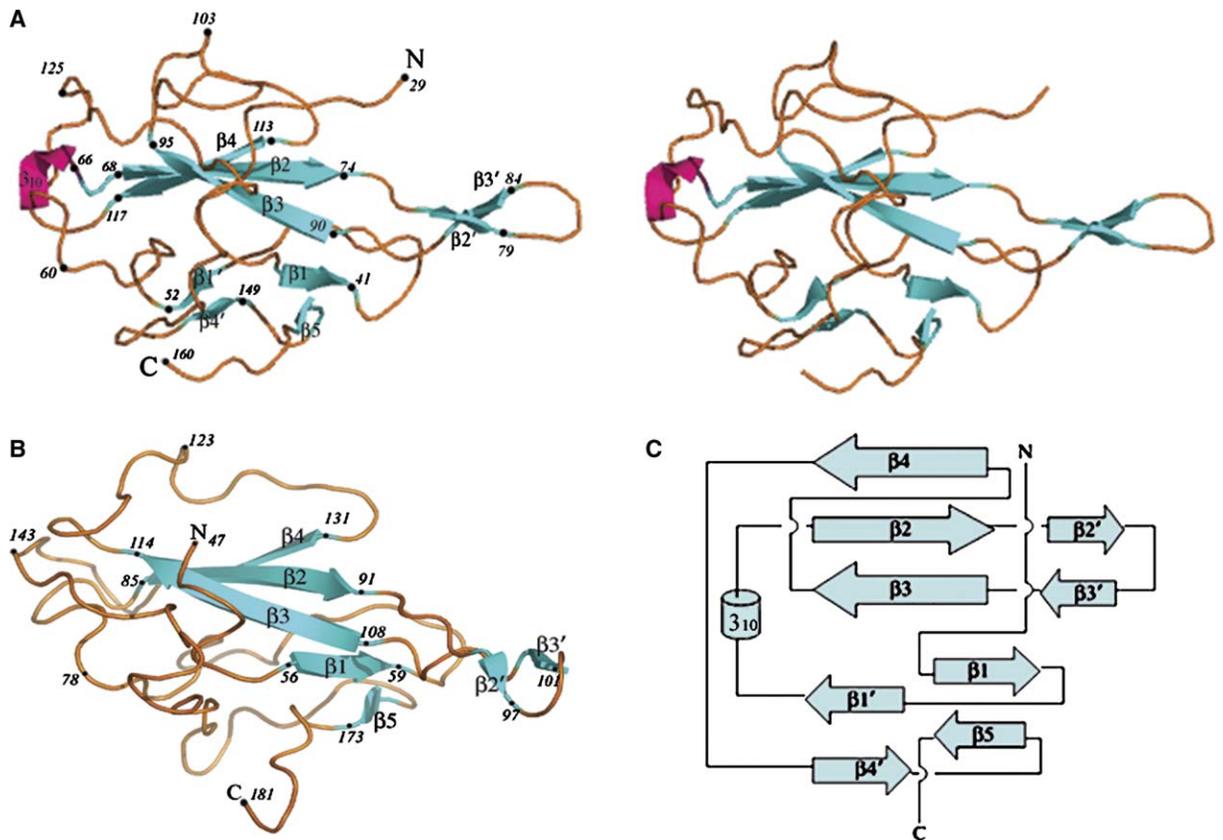


Figure 2. Overall Fold of the IBV-N Protein

(A and B) Comparison of the folds adopted by IBV-N29-160 (A; shown as a stereoview, top) and the N-terminal domain of the SARS-CoV nucleocapsid protein (B) (Huang et al., 2004). The two proteins are displayed in the same orientation. Secondary structure elements and some residue numbers are indicated.

(C) Topology diagram of the IBV-N29-160 protein. Its N- and C-terminal ends are labeled.

extension is mobile, as shown by higher than average temperature factors, and contains several basic residues which are conserved across various coronavirus N protein sequences (Figure 3). Extended loops spanning up to 30 residues connect the various secondary structure elements, presumably introducing flexibility to the overall architecture. This potential adaptability to various structural contexts might be important for assembly and disassembly of the nucleocapsid during the virus life cycle. The overall fold is similar to the SARS-N45-181 protein (Figure 2) with a few structural differences, such as the presence of a short 3_{10} helix connecting strands β_1 and β_2 . Overall, a three-dimensional structural alignment between the SARS-CoV and IBV nucleocapsid N-terminal domains using the program DALI (Holm and Sander, 1993) shows that a total of 124 equivalent $C\alpha$ atoms can be superimposed, with an rms deviation of 3.0 Å. The Z score is 10.4, confirming the global similarity of the two folds. The rather large difference between the SARS-CoV and IBV nucleocapsid N-terminal domain structures accounts for the failure of molecular replacement procedures to solve the latter structure using the former as a model. The important structural differences we observe between the SARS-CoV and IBV nucleocapsid N-terminal domain structures may stem from an inherent mobility of the coronavirus nucleocapsid structure or from a large uncertainty of the atomic

positions determined by NMR, or both. A search through the PDB did not return any other protein with a statistically significant Z score, emphasizing the uniqueness of this fold as noted by Huang et al. (2004).

Dimer Formation

In our crystal structure, the two monomers assemble into a butterfly-shaped dimer related by a 180° rotation, burying in this interaction an accessible surface area of 560 \AA^2 . The transformation is not a pure rotation, as a residual translation is needed to bring the two monomers into coincidence. The relatively small surface area suggests a rather weak binding affinity, an observation in agreement with the fact that, using size exclusion chromatography, the recombinant IBV-N29-160 protein predominantly elutes as a monomer (see below). This is also consistent with our findings of a different dimeric interface adopted by the same recombinant IBV-N29-160 protein in a nonrelated crystal form (with space group C2) that diffracts only to medium resolution.

Nucleic Acid Binding

In order to package the viral genome of 27.6 kb, the IBV-N protein must provide extended surfaces to bind the viral RNA genome both specifically and nonspecifically (without a requirement for a special base sequence). N- and C-terminal regions of IBV-N

Table 1. Crystallographic Data Collection and Phasing Statistics

Data Set	Native	SeMet		
	IBV-N: 29–160	IBV-N: 29–160 (Three Residues Mutated to Met)		
		Peak	Inflection	Remote
Wavelength	1.5418	0.97943	0.97956	0.98729
Cell parameters (Å, °), P1	a = 35.48 b = 35.72 c = 56.11 α = 99.05 β = 93.93 γ = 109.53	a = 34.77 b = 35.37 c = 55.95 α = 100.51 β = 95.48 γ = 110.16		
Resolution (Å)	20–1.85	20–1.95		
Total number of reflections	75,798	76,265	64,999	72,832
No. of unique reflection	20,031	20,083	17,032	19,684
Completeness (%) ^a	92.4 (88.8)	96.6 (95.0)	96.5 (95.2)	95.6 (87.8)
Multiplicity ^b	3.8 (3.7)	3.8 (3.7)	3.8 (3.6)	3.7 (3.5)
R _{merge} ^c	0.064 (0.625)	0.05 (0.118)	0.05 (0.131)	0.06 (0.177)
I/ σ (I)	7.4 (1.1)	8.6 (3.7)	8.3 (4.4)	9.1 (6.0)
Solvent content (%)	43.3	40.6		
No. of Se sites	—	6		
Phasing power ^d	—	0.7/0.6	0.6/0.4	0.2/1.1
f'/f'' ^e	—	−8.1/5.7	−10.5/3.3	−4.3/0.5
Figure of merit ^f	—	0.61/0.793		
20–2.5 Å				

^aThe numbers in parentheses refers to the last (highest) resolution shell.

^bFor the SeMet crystal, Friedel pairs are treated as different reflections.

^c $R_{\text{merge}} = \sum_h \sum_i |I_{hi} - \langle I_h \rangle| / \sum_h \sum_i I_{hi}$, where I_{hi} is the i th observation of the reflection h , while $\langle I_h \rangle$ is its mean intensity.

^dAnomalous phasing power/dispersive phasing power, where anomalous phasing power is $|^{\lambda}F_h| - |^{\lambda}F_{-h}|$ /anomalous lack of closure and dispersive phasing power is $|^{\lambda}F_h| - |^{\lambda}F_{-h}|$ /dispersive lack of closure.

^eValues of f' and f'' where estimated from a scan of the absorption edge using the program CHOOCH (Evans and Pettifer, 2001).

^fFigures of merit are given before and after real space density modification, respectively.

encompassing residues 1–171 and 268–407, respectively, interact with noncoding regions of the viral genomic RNA located at its 3' end (Zhou and Collisson, 2000).

Table 2. Refinement Statistics

	Native	SeMet
Resolution range (Å)	19.92–1.85	20.0–1.95
Intensity cutoff (F/ σ (F))	none	none
No. of reflections: completeness (%)	100.0	96.1
Used for refinement	18,921	16,077
Used for R _{free} calculation	1,026	881
No. of nonhydrogen atoms		
Protein	2130	2128
Water molecules	188	182
R factor (%) ^a	22.96	22.73
R _{free} (%) ^b	27.03	27.59
Rms deviations from ideality		
Bond lengths (Å)	0.007	0.008
Bond angles (°)	1.05	1.14
Ramachandran Plot		
Residues in most favored regions (%)	88.8	90.3
Residues in additional allowed regions (%)	10.2	9.2
Residues in generously allowed regions (%)	1.0	0.5
Overall G factor ^c	0.10	0.04
PDB accession code	2BXX	2BTL

^aR factor = $\sum |F_{\text{obs}}| - |F_{\text{calc}}| / \sum |F_{\text{obs}}|$.

^bR_{free} was calculated with 5% of reflections excluded from the whole refinement procedure.

^cG factor is the overall measure of structure quality from PROCHECK (Laskowski et al., 1993).

As the fragment 1–91 does not bind RNA, residues between 91 and 171 were proposed to either make direct contacts with RNA or be necessary for the integrity of the protein structure (Zhou and Collisson, 2000). Because the segment 92–95 includes strictly conserved hydrophobic residues which are buried in the protein core in our structure, we propose that the fragment 1–91 studied by Zhou and Collisson (2000) was probably poorly folded and thus nonactive. We tested nucleic acid binding by IBV-N29-160 and found that the recombinant fragment was able to bind an oligoribonucleotide from the 3' end of the viral genome (Figure 4). This result is in agreement with studies by Huang et al. (2004), who used NMR to demonstrate that SARS-CoV N45-181 could bind a 32-mer oligoribonucleotide located at the 3' end of the SARS-CoV genome. Interestingly, this oligoribonucleotide has a highly conserved sequence across various coronaviruses including IBV, and adopts a unique tertiary structure (Robertson et al., 2005). A surface representation of electrostatic charges of the IBV-N29-160 protein shown in Figure 5 reveals a striking segregation in the charge distribution on the protein surface. The β_2 - β_3 hairpin forms a basic patch at the thumb, whereas the base is acidic (Figure 5). These two charged patches are separated by a neutral and rather hydrophobic platform contributed by residues projecting from strands β_4 - β_2 - β_3 that form a palm-like structure. An alignment of nucleocapsid protein amino acid sequences from various coronaviruses highlights the conservation of several residues exposed at the protein surface, suggesting that some might play a role in nucleic acid recognition (Figures 3 and 5). The topology of the protein and its charge distribution

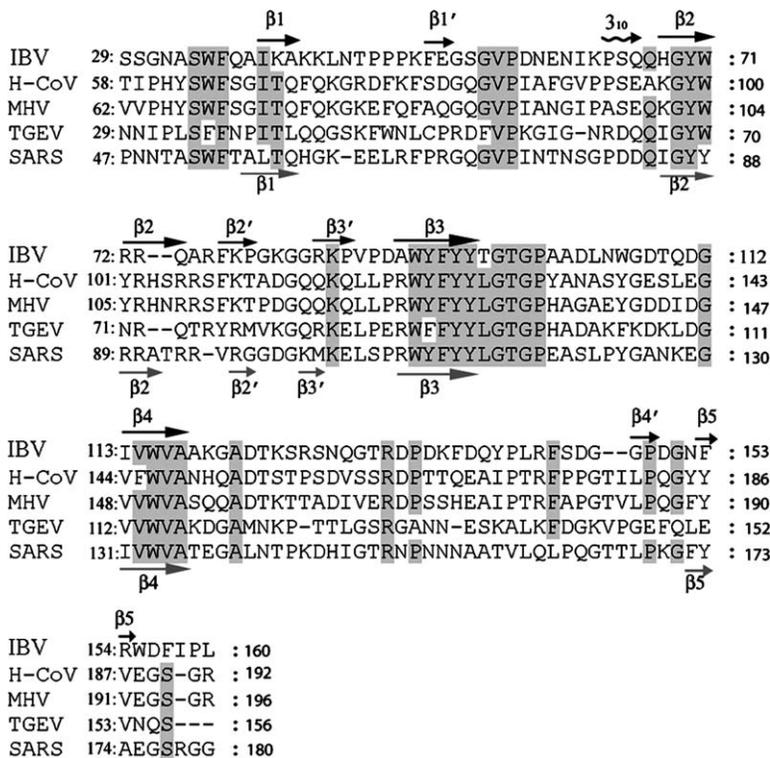


Figure 3. Structure-Based Alignment of Coronavirus Nucleocapsid Amino Acid Sequences Corresponding to the Proteolytically Stable N-Terminal Fragment

Secondary structure elements are labeled above the sequence for IBV-N29-160 and below for the SARS-CoV N-terminal fragment (Huang et al., 2004). Sequences of IBV (infectious bronchitis virus, strain Beaudette, NP_040838); H-CoV (human coronavirus, strain HKU1, YP_173242); MHV (murine hepatitis virus, strain 1, AAA46439); TGEV (porcine transmissible gastroenteritis virus, strain RM4, AAG30228); and SARS (SARS-CoV, 1SSK_A) were obtained from GenBank. Conserved residues are shaded.

suggest a mode of RNA binding in which its phosphate groups would project toward the basic β_2 - β_3 hairpin, possibly making electrostatic interactions with the conserved positively charged Arg-76 and Lys-78 residues, while the sugar and base moieties would contact the hydrophobic platform. In this model, the exposed hydrophobic residues Tyr-92 and Tyr-94 (strand β_3) could form stacking interactions with the bases, as was observed, for instance, in complexes between the vaccinia virus protein VP39 and mRNA (Hu et al., 1999) or between the matrix protein VP40 from Ebola virus and a tri-ribonucleotide (Gomis-Ruth et al., 2003). As suggested by Huang et al. (2004), additional favorable interactions might be formed upon closure of the flexible β_2 - β_3 hairpin onto the incoming RNA ligand.

In Vitro Oligomerization of the IBV Nucleocapsid Protein

Oligomerization of N protein has been studied in MHV (Robbins et al., 1986) and SARS-CoV (He et al., 2004). Trimers of N subunits linked by intermolecular disulfide bonds were identified in MHV. Using mutational analysis, the Ser/Arg-rich motif spanning residues 184-196 (immediately downstream from our crystallized fragment) was shown to be essential for the multimerization of the N protein from SARS-CoV (He et al., 2004). We analyzed the oligomerization states of the full-length IBV-N protein, IBV-N29-160, and IBV-N218-329 in solution. Crosslinking experiments were performed using glutaraldehyde, a short self-polymerizing reagent mostly reacting with the amino and amine groups of lysine and histidine, respectively (Buehler et al., 2005), and suberic acid bis *N*-hydroxy-succinamide ester (SAB), a reagent which only crosslinks lysine residues at larger distances. Concentrations of crosslinking agent higher than

0.1 mM led to the formation of dimers, tetramers (but not trimers), and larger oligomers of IBV-N, along with the disappearance of monomeric species (Figure 6). By contrast, an approximately 20-fold higher concentration of crosslinking agent (2 mM glutaraldehyde or 1 mM SAB; see Figure 6) was required to obtain equal amounts

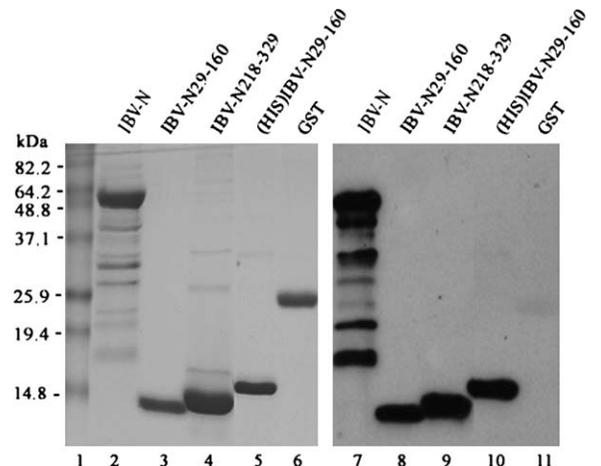


Figure 4. Analysis of the RNA Binding Activity of the Full-Length IBV-N Protein and IBV-N29-160 and IBV-N218-329 Fragments

The purified IBV-N (lanes 2 and 7), IBV-N29-160 (lanes 3 and 8), IBV-N218-329 (lanes 4 and 9), His-tagged IBV-N29-160 (lanes 5 and 10), and GST (negative control, lanes 6 and 11) were separated on a 15% SDS-PAGE gel. The proteins were either visualized by Coomassie brilliant blue staining (lanes 1-6) or transferred to Hybond C extra membrane (Amersham) and detected by Northwestern blot with a digoxin-labeled RNA probe corresponding to the IBV genome sequence from nucleotides 26,539-27,608 (lanes 7-11). Molecular masses of standard proteins are indicated.

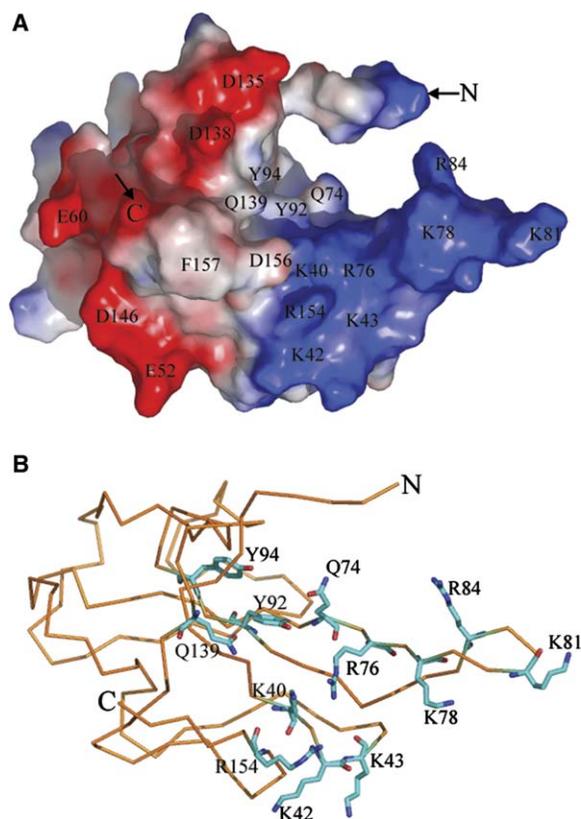


Figure 5. Proposed RNA Binding Site of IBV-N
(A) Surface representation of the IBV-N29-160 fragment with electrostatic potentials colored in blue (positive) and red (negative). Residues which are suggested to participate in RNA binding are labeled. The N- and C-terminal ends of the polypeptide chains are indicated. (B) Close-up view of the proposed RNA binding site of the IBV-N29-160 fragment. The C α trace of IBV-N29-160 is displayed. Side chains which are likely to participate in nucleic acid binding are shown as sticks.

of monomers and dimers of IBV-N29-160. This suggests that regions within the C-terminal domain of IBV-N make a predominant contribution to the multimerization of IBV-N. Secondary structure predictions and limited proteolysis studies of the IBV-N protein suggest the presence of a structured—possibly α -helical—C-terminal domain of about 12 kDa, which is connected to IBV-N29-160 by a Ser/Arg/Ala/Gly-rich loop of approximately 50 amino acid residues (Figure 1). We expressed such a stable recombinant C-terminal domain encompassing residues 218–329 of the N protein in a soluble form. This C-terminal domain can bind RNA (Figure 4), fold independently, and was recently crystallized (H.F., D.X.L., and J.L., unpublished data). Crosslinking experiments show that IBV-N218-329 forms dimers, trimers, tetramers, and higher oligomers for concentrations of crosslinking agent higher than 1 mM with a concomitant decrease in monomer species, thus confirming the important contribution of the C-terminal domain of IBV to the formation of IBV-N multimers (Figure 6). As an independent confirmation, we subjected the IBV-N29-160 and IBV-N218-329 domains to size exclusion chromatography (Figure 7). Under these conditions, the C-terminal domain IBV-N218-329 elutes faster than the

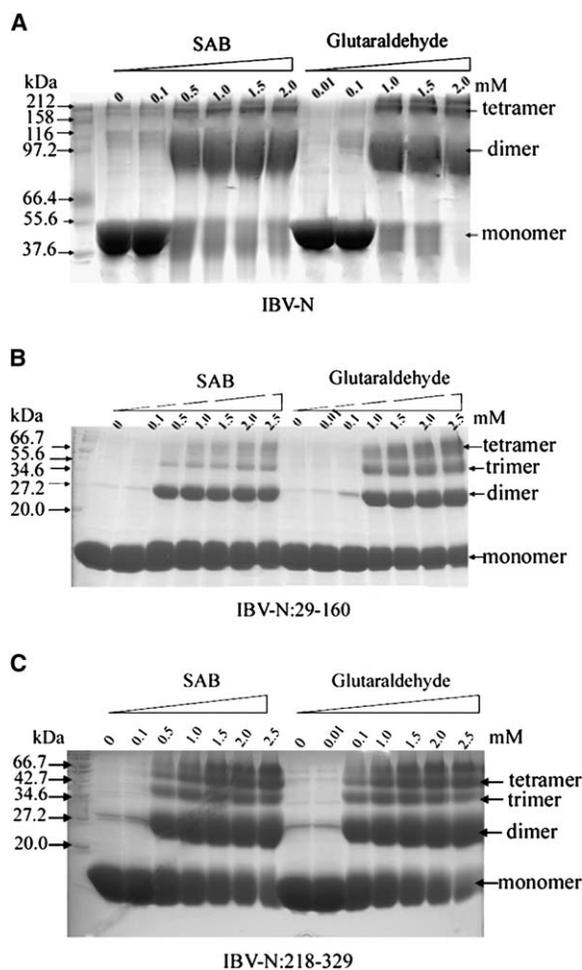


Figure 6. Crosslinking Experiments
(A) Full-length IBV-N protein.
(B) IBV-N29-160 protein, which was crystallized.
(C) C-terminal fragment, IBV-N218-329.
The nature and concentrations of crosslinking agent are shown. Monomer, dimer, trimer, and tetramer species of the recombinant proteins are indicated.

N-terminal domain as a sharp symmetric peak corresponding to a dimer. The N-terminal domain elutes at a position intermediate between a monomer and a dimer (with an estimated size corresponding to a protein of molecular weight 18.1 kDa). This pattern of migration could be due to the asymmetric shape of the IBV N-terminal domain or could be indicative of the presence of a mixture of monomer and dimer of the N-terminal domain in solution.

Implications for Coronavirus Nucleocapsid Assembly

Our data suggest that residues 218–329 at the C-terminal end of the IBV-N protein play a major role for its multimerization. This is consistent with results reported by Surjit et al. (2004), who studied SARS-CoV nucleocapsid dimerization using the yeast two-hybrid system, and points to conserved assembly properties between the SARS-CoV and IBV in spite of significant amino acid differences between their two nucleocapsid

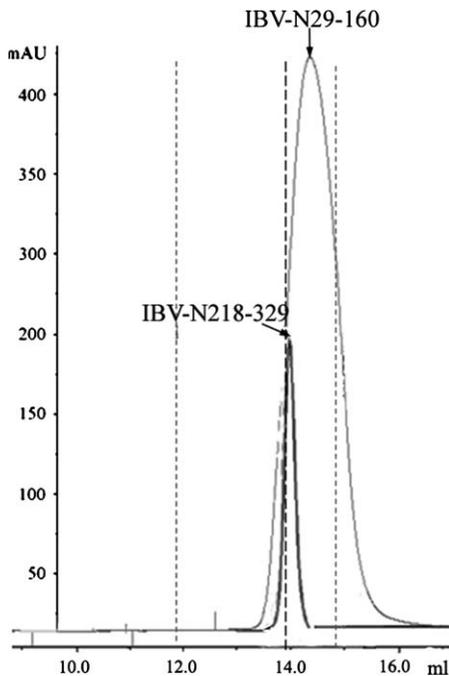


Figure 7. Size Exclusion Chromatography Elution Profiles of IBV-N29-160 and IBV-N218-329

The vertical axis shows absorbance at 280 nm. The horizontal axis indicates the elution volume in milliliters. Three thin vertical lines indicate the positions of molecular weight of protein standards (from left to right: ovalbumin, 43 kDa; chymotrypsinogen A, 25.0 kDa; and ribonuclease A, 13.7 kDa). The large difference in absorbance stems from the different individual molar absorptance coefficients at 280 nm of IBV-N29-160 ($40,540 \text{ M}^{-1}\text{cm}^{-1}$) and IBV-N218-329 ($4,080 \text{ M}^{-1}\text{cm}^{-1}$).

proteins. Can we ascribe a function to multimer formation by the N protein? One obvious explanation is that multimerization increases the protein surface area accessible for binding the viral genomic RNA, thus provid-

ing the elementary building block for nucleocapsid assembly. Indeed, several crystal structures of capsid proteins have revealed the presence of multimers that present continuous patches of basic residues at their surface: the capsid proteins of West Nile virus and Borna disease virus form tetrameric assemblies (Dokland et al., 2004; Rudolph et al., 2003) and the nucleocapsid protein of porcine respiratory syndrome virus, an arterivirus, forms dimers (Doan and Dokland, 2003). Unfortunately, because these structures were determined in the absence of an RNA ligand, it is difficult to evaluate to what extent multimer formation is coupled with nucleic acid recognition. In the *Arteriviridae*, a viral family genomically related to the coronaviruses, the basic N-terminal half of the nucleocapsid protein is involved in RNA binding while its C-terminal domain forms a tight dimer (Doan and Dokland, 2003).

Further complexity for the study of coronavirus nucleocapsid assembly stems from its interaction with the M protein endodomain (Kou and Masters, 2002; Narayanan et al., 2000, 2003) and from the fact that several coronavirus proteins can interact with single-stranded RNA, including the nsp9 replicase protein from SARS-CoV (Egloff et al., 2004; Sutton et al., 2004). In the absence of a nucleic acid ligand, the N protein appears to be composed of two main globular domains loosely connected by Arg/Ser/Ala/Gly-rich loops that are highly sensitive to proteolysis. These connecting regions may undergo modifications (e.g., phosphorylation) that could influence the multimerization state of the protein and control its interaction with RNA. In a recent report, sumoylation of Lys-62 of SARS-CoV N protein expressed in mammalian cells was proposed to promote dimerization of the protein (Li et al., 2005). It is not known whether similar modifications of the IBV-N occur in virus-infected cells. Nevertheless, a working hypothesis for coronavirus nucleocapsid formation can be proposed (Figure 8). In this model, viral genomic RNA binding by both the N- and C-terminal domains would lead to a clustering

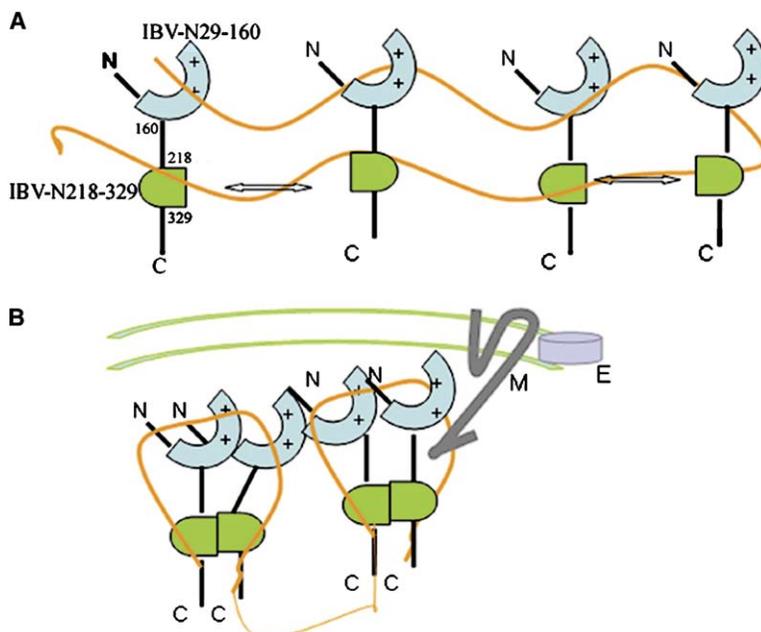


Figure 8. Hypothetical Model for the Assembly of the IBV Ribonucleoprotein Complex
(A) Both the N- (cyan) and C-terminal (green) domains of the IBV-N protein can bind RNA (represented as a thin orange line). The basic patch in IBV-N29-160 is depicted with plus signs. Dimerization of the C-terminal domains (arrows) leads to a clustering of IBV-N proteins and to their oligomerization.
(B) The endodomain of the integral membrane protein M can provide further contacts to the ribonucleocapsid (see text). However, the precise coupling between RNA recognition and IBV-N multimerization remains uncertain.

of N proteins. Dimerization of the C-terminal domains would trigger oligomerization of the N-terminal domains by increasing their local concentration above a certain threshold. In turn, this would trigger condensation of viral RNA. Interdomain flexibility we have defined in the linker regions could facilitate the necessary conformational changes during the transition to a more compact form of the ribonucleocapsid (Figure 8).

Further studies are underway to elucidate the three-dimensional structure of the globular C-terminal domain of IBV-N, to define the interactions between the IBV-N protein and viral RNA, and to characterize the morphology of the ribonucleocapsid.

Experimental Procedures

Cloning and Expression

The gene encoding the IBV-N protein was amplified by PCR using the *Pfu* polymerase (Stratagene, Singapore) with the forward (5'-AT TATT CAT ATG GCA AGC GGT AAA GCA GC-3') and reverse primer (5'-ATTATT CTC GAG TCA AAG TTC ATT CTC TCC TA-3') and cloned into the pET 29b vector using T4 ligase (Research Biolabs, Singapore). The underlined sequences correspond to NdeI and XhoI sites, respectively. Proteins (lacking the His₆ tag due to the insertion of a stop codon in the reverse primer) were expressed in *E. coli* BL21(DE3). The cells were grown at 37°C in Luria-Bertani medium containing 100 µg/ml ampicillin until the culture reached an OD₆₀₀ of 0.7. Protein expression was induced by the addition of 1 mM isopropyl-β-D-thiogalactopyranoside for 3 hr at 30°C. Cells harvested and resuspended at 4°C in a buffer containing 20 mM Na₃PO₄ (pH 7.8) were lysed by sonication and the remaining insoluble material was removed by centrifugation at 20,000 × g for 20 min at 4°C. N- and C-terminal fragments of the IBV-N gene coding for residues 29–160 and 218–329, respectively, were cloned into pET-16b using the following primers: 5'-AATA CAT ATG TCT TCT GGA AAT GCA TCT TG-3'; 5'-AATA CTC GAG TCA CAG GGG AAT GAA GTC CC-3' and 5'-A AATA CAT ATG AAG GCA GAT GAA ATG GC-3'; 5'-AA ATA CTC GAG TCA CGT TCC TAC ACC ATC GAC-3'. These two proteins (hereafter named IBV-N29-160 and IBV-N218-329, respectively) were expressed as described above for IBV-N, yielding truncated fragments having a His₁₀ tag at their N terminus followed by a Factor Xa cleavage site. The His₁₀ tags were cleaved during purification. Expression of the selenomethionylated protein IBV-N29-160 was carried out as described in Doublé (1997).

Analysis of the Proteolytically Stable Fragment Derived from IBV-N

Automated N-terminal amino acid sequence determination of the proteolytic fragment obtained by degradation of IBV-N was performed using an Applied Biosystems (Singapore) Procise sequencer. The molecular mass of purified proteins was analyzed using a MALDI-TOF mass spectrometer (API 300 MS/MS; Applied Biosystems).

Protein Purification

The IBV-N protein precipitated with ammonium sulfate at 30% saturation, was centrifuged and resuspended in PBS, dialyzed against buffer A (20 mM HEPES, 1 mM EDTA, 1 mM DTT [pH 6.8]), and loaded onto a cation exchange chromatography column (Mono S HR 5/5; GE Biosciences, Singapore) preequilibrated with buffer A. Elution was carried out using a NaCl gradient of buffer B (20 mM HEPES, 1 mM EDTA, 1 mM DTT, 1 M NaCl [pH 6.8]). Fractions containing the protein—as shown by SDS-PAGE—were pooled and concentrated to 10–15 mg/ml by ultrafiltration using a Centrprep device (Millipore, Singapore) with a molecular weight cutoff of 10 kDa. Size exclusion chromatography (Superdex 75; Amersham) was carried out in a buffer containing 20 mM Tris-HCl (pH 8.0), 150 mM NaCl, 1 mM DTT, 0.1% NaN₃. The protein was concentrated to 10 mg/ml as determined by the Bradford assay (Bio-Rad, Singapore), using BSA as a standard. The truncated recombinant IBV-N29-160 was resuspended in PBS and loaded onto a Ni-NTA column (Qiagen, Singapore) preequilibrated with 20 mM KH₂PO₄, 50 mM NaCl (pH 7.8). Af-

ter washing with 20 mM KH₂PO₄, 1 M NaCl, 10 mM imidazole (pH 7.2), IBV-N29-160 was eluted using a buffer containing 20 mM KH₂PO₄, 0.5 M NaCl, 0.5 M imidazole (pH 6.0). The His₁₀ tag was removed by proteolysis with Factor Xa in a buffer containing 100 mM NaCl, 2 mM CaCl₂, 10 mM Tris (pH 8.0) using a substrate enzyme molar ratio of 50:1 for 4 hr at room temperature. The cleavage mixture was loaded onto a benzamidine column to eliminate Factor Xa, and the IBV-N29-160 protein recovered in the flow through was subjected to two final steps of purification as described above for the full-length IBV-N protein. Purification of the recombinant IBV-N218-329 was carried out using a similar protocol. Purification of the selenomethionine-substituted protein was performed using the same protocol as the native protein.

Crystallization of IBV-N29-160

Crystals of the recombinant IBV-N29-160 were grown at 18°C by vapor diffusion using the hanging drop method. Two microliters of the protein at a concentration of 10 mg/ml was mixed with an equal volume of the precipitating solution from the well (0.1 M sodium sulfate, 20% PEG 3350), yielding plate-shaped crystals growing to maximum dimensions of about 0.3 × 0.3 × 0.05 mm³ in about 2 weeks (Figure 1). Crystals of the selenomethionine protein were obtained under the same conditions.

Data Collection, Structure Determination, and Refinement

For data collection, crystals were soaked in a cryoprotecting solution (25% glycerol, 0.1 M sodium sulfate, 20% PEG 3350 [pH 6.5]) before being mounted and cooled to 100 K in a nitrogen gas stream (Oxford Cryosystems, Oxford, UK). Diffraction intensities at three wavelengths (Table 1) were recorded from a selenomethionine (SeMet)-derivatized IBV-N29-160 crystal on beamline NW12 at the Photon Factory (Tsukuba, Japan) on an ADSC charge-coupled device (CCD) detector (ADSC Corporation, Poway, CA) using an attenuated beam of dimensions 0.1 × 0.1 mm² (Table 1). Integration, scaling, and merging of the intensities were carried out using programs from the CCP4 (1994). The six selenium atoms present within the two molecules of the asymmetric unit were located using the program SOLVE (Terwilliger, 2003). An initial electron density map was calculated and modified using the program RESOLVE (Terwilliger, 2003), using these selenium atom positions to locate the noncrystallographic symmetry (ncs) axis relating the two molecules in the asymmetric unit, and model building was first carried out in this map using the program O (Jones et al., 1991). For subsequent cycles, electron density maps were calculated using partial model phases combined with experimental MAD phases with the program REFMAC5 from the CCP4 (1994), which was used for the initial refinement of the structure, that included ncs restraints. A few cycles of refinement using molecular dynamics with a slow cooling protocol using a maximum likelihood target incorporating phase probability distribution encoded in the form of Hendrickson Lattman coefficients were subsequently carried out using the program CNS (Brunger et al., 1998), with ncs restraints. A data set for the native protein was collected on an R axis IV++ image plate detector using CuK α radiation from a Micromax-007 rotating anode (Rigaku/MSK, The Woodlands, TX) operating at 20 mA and 40 kV (Table 1). The SeMet model was placed in the native crystal form and adjustments to the model were carried out using difference Fourier maps calculated with REFMAC5, which was used for refinement. Superposition of structures and rms deviation calculations were carried out using the program LSQKAB from the CCP4 (1994). Figures 2 and 5 were produced with the program PyMOL (DeLano, 2002).

Crosslinking Experiments

The purified recombinant proteins IBV-N, IBV-N29-160, and IBV-N218-329 were incubated with either glutaraldehyde or SAB (Sigma-Aldrich, St. Louis, MO) for 2 hr at 20°C using a constant amount of protein (5 µg) with increasing amounts of the crosslinking agent. The samples were submitted to electrophoresis on an 8%–15% SDS-PAGE gel and stained with Coomassie blue.

Size Exclusion Chromatography

A Superdex 75 10/300 GL size exclusion chromatographic column (Amersham) mounted on an AKTA FPLC (GE Biosciences, Singapore) was used to analyze the homogeneity and apparent

multimerization states of IBV-N29-160 and IBV-N218-329, respectively. Protein concentrations used were 10 mg/ml and the loaded sample volume was 0.1 ml. The buffer was 10 mM Tris-HCl, 0.2 M NaCl, 3 mM β -mercapto-ethanol (pH 7.5) and the flow rate was 0.5 ml/min. Standard protein markers (Amersham) used for calibration were ribonuclease A, 13.7 kDa, elution 14.88 ml; chymotrypsinogen A, 25.0 kDa, 13.81 ml; ovalbumin, 43.0 kDa, 11.81 ml; BSA, 67.0 kDa, 10.89 ml. Apparent size/molecular weights were deduced by plotting K_{av} versus log (MW) with $K_{av} = (V_e - V_0)/(V_t - V_0)$, where V_e is the elution volume of the protein, V_t is the total column bed volume, and V_0 is the void volume.

RNA Binding Assay

The full-length IBV-N protein and the IBV-N29-160 and IBV-N218-329 fragments were expressed in *E. coli* BL21 cells and purified as described above. The polyhistidine tags of the truncated proteins were removed by digestion with Factor Xa. The purified proteins were separated on 15% SDS-PAGE, transferred to Hybond C extra membrane (Amersham), and probed with digoxin-labeled RNA representing the negative sense of the IBV genome from nucleotides 25,873–27,608. The probe was made by *in vitro* transcription using SP6 polymerase in the presence of digoxin according to the manufacturer's instructions (Roche, Singapore).

Acknowledgments

We thank Jacques d'Alayer (Institut Pasteur) for performing N-terminal amino acid sequencing, and Terje Dokland and Soichi Wakatsuki for help and very useful discussions. Financial support via grants from NTU (SUG 14/02), the Singapore Biomedical Research Council (03/1/21/20/291 and 02/1/22/17/043), and the Singapore National Medical Research Council (NMRC/SRG/001/2003) to the Lescair laboratory is acknowledged as well as provision of excellent beam time by the Photon Factory (Japan).

Received: May 30, 2005

Revised: August 9, 2005

Accepted: August 30, 2005

Published: December 13, 2005

References

- Bosch, B.J., Martina, B.E., Van Der Zee, R., Lepault, J., Haijema, B.J., Versluis, C., Heck, A.J., De Groot, R., Osterhaus, A.D., and Rotter, P.J. (2004). Severe acute respiratory syndrome coronavirus (SARS-CoV) infection inhibition using spike protein heptad repeat-derived peptides. *Proc. Natl. Acad. Sci. USA* **101**, 8455–8460.
- 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 & NMR system: a new software suite for macromolecular structure determination. *Acta Crystallogr. D Biol. Crystallogr.* **54**, 905–921.
- Buehler, P.W., Boykins, R.A., Jia, Y., Norris, S., Freedberg, D.I., and Alayash, A.I. (2005). Structural and functional characterization of glutaraldehyde polymerized bovine hemoglobin and its isolated fractions. *Anal. Chem.* **77**, 3466–3478.
- CCP4 (Collaborative Computational Project, Number 4) (1994). The CCP4 suite: programs for protein crystallography. *Acta Crystallogr. D Biol. Crystallogr.* **50**, 760–763.
- Cologna, R., and Hogue, B.G. (2000). Identification of a bovine coronavirus packaging signal. *J. Virol.* **74**, 580–583.
- De Guzman, R.N., Wu, Z.R., Stalling, C.C., Pappalardo, L., Borer, P.N., and Summers, M.F. (1998). Structure of the HIV-1 nucleocapsid protein bound to the SL3 ψ -RNA recognition element. *Science* **279**, 384–388.
- DeLano, W.L. (2002). *The PyMOL User's Manual* (San Carlos, CA: DeLano Scientific).
- Doan, D.N.P., and Dokland, T. (2003). Structure of the nucleocapsid protein of porcine reproductive and respiratory syndrome virus. *Structure* **11**, 1445–1451.

Dokland, T., Walsh, M., Mackenzie, J.M., Khromykh, A.A., Ee, K.H., and Wang, S. (2004). West Nile virus core protein; tetramer structure and ribbon formation. *Structure* **12**, 1157–1163.

Doublé, S. (1997). Preparation of selenomethionyl proteins for phase determination. *Methods Enzymol.* **276**, 523–530.

Egloff, M.-P., Ferron, F., Campanacci, V., Longhi, S., Rancurel, C., Dutartre, H., Snijder, E.J., Gorbalenya, A.E., Cambillau, C., and Carnard, B. (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.

Escors, D., Ortego, J., Laude, H., and Enjuanes, L. (2001). The membrane M protein carboxy terminus binds to transmissible gastroenteritis coronavirus core and contributes to core stability. *J. Virol.* **75**, 1312–1324.

Escors, D., Izeta, A., Capiscol, C., and Enjuanes, L. (2003). Transmissible gastroenteritis coronavirus packaging signal is located at the 5' end of the virus genome. *J. Virol.* **77**, 7890–7902.

Evans, G., and Pettifer, R.F. (2001). CHOOCH: a program for deriving anomalous-scattering factors from X-ray fluorescence spectra. *J. Appl. Crystallogr.* **34**, 82–86.

Fosmire, J.A., Hwang, K., and Makino, S. (1992). Identification and characterization of a coronavirus packaging signal. *J. Virol.* **66**, 3522–3530.

Gassner, N.C., and Matthews, B.W. (1999). Use of differentially substituted selenomethionine proteins in X-ray structure determination. *Acta Crystallogr. D Biol. Crystallogr.* **55**, 1967–1970.

Gomis-Ruth, F.X., Dessen, A., Timmins, J., Bracher, A., Kolesnikowa, L., Becker, S., Klenk, H.D., and Weissenhorn, W. (2003). The matrix protein VP40 from Ebola virus octamerizes into pore-like structures with specific RNA binding properties. *Structure* **11**, 423–433.

He, R., Dobie, F., Ballantine, M., Leeson, A., Li, Y., Bastien, N., Cutts, T., Andonov, A., Cao, J., Booth, T.F., et al. (2004). Analysis of multimerization of the SARS coronavirus nucleocapsid protein. *Biochem. Biophys. Res. Commun.* **316**, 476–483.

Holm, L., and Sander, C. (1993). Protein structure comparison by alignment of distance matrices. *J. Mol. Biol.* **233**, 123–138.

Hu, G., Gershon, P.D., Hodel, A.E., and Quijoch, F.A. (1999). mRNA cap recognition: dominant role of enhanced stacking interactions between methylated bases and protein aromatic side chains. *Proc. Natl. Acad. Sci. USA* **96**, 7149–7154.

Huang, Q., Yu, L., Petros, A.M., Gunasekera, A., Liu, Z., Xu, N., Hajduk, P., Mack, J., Fesik, S.W., and Olejniczak, E.T. (2004). Structure of the N-terminal RNA-binding domain of the SARS CoV nucleocapsid protein. *Biochemistry* **43**, 6059–6063.

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

Kou, L., and Masters, P.S. (2002). Genetic evidence for a structural interaction between the carboxy termini of the membrane and nucleocapsid proteins of mouse hepatitis virus. *J. Virol.* **76**, 4987–4999.

Lai, M.M.C., and Holmes, K.V. (2001). *Coronaviridae: The Viruses and Their Replication in Fundamental Virology*, 4th Edition (Philadelphia: Lippincott Raven).

Laskowski, R.A., MacArthur, M.W., Moss, D.S., and Thornton, J.M. (1993). PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Crystallogr.* **26**, 283–291.

Lescair, J., Roussel, A., Wien, M.W., Navaza, J., Fuller, S.D., Wengler, G., Wengler, G., and Rey, F.A. (2001). The fusion glycoprotein shell of Semliki Forest virus: an icosahedral assembly primed for fusogenic activation at endosomal pH. *Cell* **105**, 137–148.

Li, F.Q., Xiao, H., Tam, J.P., and Liu, D.X. (2005). Sumoylation of the nucleocapsid protein of severe acute respiratory syndrome associated coronavirus. *FEBS Lett.* **579**, 2387–2396.

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 SARS-associated coronavirus. *Science* **300**, 1399–1403.

- Narayanan, K., Maeda, A., Maeda, J., and Makino, S. (2000). Characterization of the coronavirus M protein and nucleocapsid in infected cells. *J. Virol.* *74*, 8127–8134.
- Narayanan, K., Chen, C.J., Maeda, J., and Makino, S. (2003). Nucleocapsid-independent specific viral RNA packing via viral envelope protein and viral RNA signal. *J. Virol.* *77*, 2922–2927.
- 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.
- Robbins, S.G., Frana, M.F., McGowan, J.J., Boyle, J.F., and Holmes, K.V. (1986). RNA-binding proteins of coronavirus MHV: detection of monomeric and multimeric N protein with an RNA overlay-protein blot assay. *Virology* *150*, 402–410.
- Robertson, M.P., Igel, H., Baertsch, R., Haussler, D., Ares, M., Jr., and Scott, W.G. (2005). The structure of a rigorously conserved RNA element within the SARS virus genome. *PLoS Biol.* *3*, e5.
- Rota, P.A., Oberste, M.S., Monroe, S.S., Nix, W.A., Campagnoli, R., Icenogle, J.P., Peñaranda, S., Bankamp, B., Maher, K., Chen, M.-h., et al. (2003). Characterization of a novel coronavirus associated with severe acute respiratory syndrome. *Science* *300*, 1394–1398.
- Rudolph, M.G., Kraus, I., Dickmanns, A., Eickmann, M., Garten, W., and Ficner, R. (2003). Crystal structure of the borna disease virus nucleoprotein. *Structure* *11*, 1219–1226.
- Sturman, L.S., Holmes, K.V., and Behnke, J. (1980). Isolation of coronavirus envelope glycoproteins and interaction with the viral nucleocapsid. *J. Virol.* *33*, 449–462.
- Surjit, M., Liu, B., Kumar, P., Chow, V.T.K., and Lal, S.K. (2004). The nucleocapsid protein of the SARS coronavirus is capable of self association through a C-terminal 209 amino acid interaction domain. *Biochem. Biophys. Res. Commun.* *317*, 1030–1036.
- Sutton, G., Fry, E., Carter, L., Sainsbury, S., Walter, T., Nettleship, J., Berrow, N., Owens, R., Gilbert, R., Davidson, A., et al. (2004). The nsp9 replicase protein of SARS-coronavirus, structure and functional insights. *Structure* *12*, 341–353.
- Terwilliger, T.C. (2003). SOLVE and RESOLVE: automated structure solution and density modification. *Methods Enzymol.* *374*, 22–37.
- Valegard, K., Murray, J.B., Stonehouse, N.J., van den Worm, S., Stockley, P.G., and Liljas, L. (1997). The three-dimensional structures of two complexes between recombinant MS2 capsids and RNA operator fragments reveal sequence-specific protein-RNA interactions. *J. Mol. Biol.* *270*, 724–738.
- Vennema, H., Godeke, G.J., Rossen, J.W.A., Voorhout, W.F., Horzinek, M.C., Opstelten, D.J., and Rottier, P.J.M. (1996). Nucleocapsid-independent assembly of coronavirus-like particles by co-expression of viral envelope protein gene. *EMBO J.* *15*, 2020–2029.
- Williams, A.K., Wang, L., Sneed, L.W., and Collisson, E.W. (1992). Comparative analysis of the nucleocapsid genes of several strains of infectious bronchitis virus and other coronaviruses. *Virus Res.* *25*, 213–222.
- Zhou, M., and Collisson, E.W. (2000). The amino and carboxyl domains of the infectious bronchitis virus nucleocapsid protein interact with 3' genomic RNA. *Virus Res.* *67*, 31–39.
- Zhou, M., Williams, A.K., Chung, S.I., Wang, L., and Collisson, E.W. (1996). The infectious bronchitis virus nucleocapsid protein binds RNA sequences in the 3' terminus of the genome. *Virology* *217*, 191–199.