Protein Cell 2010, 1(7): 688–697
DOI 10.1007/s13238-010-0079-x

Protein & Cell

RESEARCH ARTICLE

Structures of the N- and C-terminal domains of MHV-A59 nucleocapsid protein corroborate a conserved RNA-protein binding mechanism in coronavirus

Yanlin Ma[1,2*], Xiaohang Tong[1,2*], Xiaoling Xu[3], Xuemei Li[1], Zhiyong Lou[3], Zihe Rao[1,3] ✉

[1] Life National Laboratory of Biomacromolecules, Institute of Biophysics, Chinese Academy of Sciences, Beijing 100101, China
[2] Graduate University of the Chinese Academy of Sciences, Beijing 100049, China
[3] Laboratory of Structural Biology, Tsinghua University, Beijing 100084, China
✉ Correspondence: raozh@xtal.tsinghua.edu.cn
Received June 1, 2010 Accepted June 7, 2010

ABSTRACT

Coronaviruses are the causative agent of respiratory and enteric diseases in animals and humans. One example is SARS, which caused a worldwide health threat in 2003. In coronaviruses, the structural protein N (nucleocapsid protein) associates with the viral RNA to form the filamentous nucleocapsid and plays a crucial role in genome replication and transcription. The structure of N-terminal domain of MHV N protein also implicated its specific affinity with transcriptional regulatory sequence (TRS) RNA. Here we report the crystal structures of the two proteolytically resistant N- (NTD) and C-terminal (CTD) domains of the N protein from murine hepatitis virus (MHV). The structure of NTD in two different crystal forms was solved to 1.5 Å. The higher resolution provides more detailed structural information than previous reports, showing that the NTD structure from MHV shares a similar overall and topology structure with that of SARS-CoV and IBV, but varies in its potential surface, which indicates a possible difference in RNA-binding module. The structure of CTD was solved to 2.0-Å resolution and revealed a tightly intertwined dimer. This is consistent with analytical ultracentrifugation experiments, suggesting a dimeric assembly of the N protein. The similarity between the structures of these two domains from SARS-CoV, IBV and MHV corroborates a conserved mechanism of nucleocapsid formation for coronaviruses.

KEYWORDS crystal structure, nucleocapsid protein, murine hepatitis virus

INTRODUCTION

Coronaviruses are large, enveloped, positive single-stranded RNA viruses, which belong to *Coronaviridae* family, *Nidovirales* order. Coronatviruses are the causative agent of many animal and human diseases (Rota et al., 2003). Especially, in 2003, SARS-CoV caused a worldwide health threat and accounted for over 8098 infection and 774 death cases (Drosten et al., 2003; Fleischauer and CDC SARS Investigative Team, 2003; Ksiazek et al., 2003). The coronavirus has an extraordinary large genome, ranging from ~ 27 to 31.5 kb. On the basis of antigenic cross-reactivity and sequence similarity, coronaviruses can be assigned to three groups, with HCoV-229E (group I), mouse hepatitis virus (MHV, group II), and avian infectious bronchitis virus (IBV, group III) being the representatives of each group. MHV, which causes liver or neuron infection in mice, is the best-studied coronavirus before the 2003 SARS outbreak.

MHV contains a 31.4-kb positive-sense ssRNA genome (Lai and Stohlman, 1978; Sturman and Holmes, 1983). The genomic RNA is encapsidated by the nucleocapsid (N) protein into a capsid core. The other four structural proteins, including spike (S), membrane (M), envelope (E) and hemagglutinin-esterase (HE), surrounded the capsid core to form the crown-like viral particles (Sturman and Holmes, 1983). Upon infection into a cell, the virus produces two large

*These authors contributed equally to the work.

polyproteins (pp1a and pp1ab). They are cleaved by papain-like proteinase 1 (PLP1) and the poliovirus 3C-like proteinase (3CL M^{pro}) into 16 non-structural proteins, which function as the replication-transcription complexes (RTC) (Sturman and Holmes, 1983).

The MHV-A59 N protein is well-conserved among the various MHV strains. It interacts with genomic RNA to form the helical nucleocapsid (Macneughton and Davies, 1978; Robbins et al., 1986; Baric et al., 1988; Almazán et al., 2004; Sawicki et al., 2005), and associates with the membrane glycoprotein via its C-terminal to stabilize virion assembly (Kuo and Masters, 2002; Hurst et al., 2005; Bednar et al., 2006; Verma et al., 2006). It is also considered as an RNA chaperone (Mir and Panganiban, 2006; Zúñiga et al., 2007). Previous biochemical results indicated that the N protein binds specific RNA sequences, e.g., the leader RNA (Stohlman et al., 1988; Zhang et al., 1994; Nelson et al., 2000) and the packaging signal (Molenkamp and Spaan, 1997). The leader RNA contains 72–76 nucleotides, which consist of two or three copies of penta-nucleotide sequence (UCUAA) that is critical for virus transcription. Nelson et al. (2000) used a RNA ligand binding assay to demonstrate that the N protein had a dissociation constant (K_d) of 14.7 nM when RNA contains UCUAA sequence. They also located the smallest N protein fragment with a significant K_d of 32 nM as residues 177–231. The specific interaction of MHV packaging signal and N proteins was observed *in vitro*, and similar packaging signal or (nucleo)capsid protein interactions have been observed in several other RNA viruses, including alphaviruses and retroviruses (Molenkamp and Spaan, 1997). It has been postulated that the packaging signal functions as a selective encapsidation initiation site by its specific interaction with the N protein (Molenkamp and Spaan, 1997). Recently, Grossoehme et al. (2009) reported that the MHV-N219 (residues 60–219) selectively binds to TRS (transcription regulatory sequence) RNA with high affinity. Moreover, van der Meer et al. (1999) used immunofluorescence microscopy to prove the co-localization of the N protein with 3CL M^{pro}, helicase protein and RNA polymerase protein in early MHV-A59 infected cells. Using the same assay, Bost et al. (2000) reported that pp1ab and N protein could be closely localized *in vivo*. Furthermore, the reverse genetic results showed that the rescue of recombinant coronaviruses (TGEV, IBV, MHV) from cells can be greatly enhanced when the cells express N protein (Almazán et al., 2000; Casais et al., 2001; Coley et al., 2005).

The N protein of MHV-A59 is a highly basic phosphoprotein with the molecular weight of 55 kDa. It could be sub-divided into three conserved domains: domains I (residues M1–A139) and II (residues D163–Q380) are basic, and the C-terminal domain III (residues E406–V454) is acidic. A general RNA binding region was initially located at residues H136–R397 (Masters, 1992; Cologna et al., 2000; You et al., 2007), while the conserved negatively charged amino acids in domain III are believed to play an important role in N-M protein interactions during assembly (Verma et al., 2006).

To gain insight into the precise mechanism of N protein, several crystallographic or NMR structural results were reported, including MHV N-terminal RNA binding domain (residues 60–195) (Grossoehme et al., 2009), two protease-resistant domains of the N protein from SARS-CoV (Huang et al., 2004; Luo et al., 2006; Yu et al., 2006; Chen et al., 2007; Saikatendu et al., 2007; Takeda et al., 2008), and IBV (Beaudette strain and Gray strain) (Fan et al., 2005; Jayaram et al., 2006). The two domains of IBV and SARS-CoV and the flexible linker between them provide a putative binding surface for viral RNA. This is supported by reported structures, which also revealed the dimerization of the C-terminal domain. Thus, a hypothesis for nucleocapsid formation proposes that the N protein self-assembles via its C-terminal dimeric domain, and the viral RNA entwines around the protein (Jayaram et al., 2006). In this work, we report the crystal structures of two proteolytically stable domains of MHV-A59 N protein.

In overall ribbon posture, the high resolution structure of MHV-NTD determined using two forms of crystals with different packing modes is similar to previously reported SARS-CoV and IBV structures, with a remarkable difference in surface electrostatic distribution. The CTD displayed a tightly intertwined dimerization structure as expected, indicating a potential role in self-association of N protein. These results suggest a similar model, but with exceptions in certain details for RNA binding style.

RESULTS

Monomer folding of the MHV-N NTD and CTD

MHV-NTD was crystallized into two different packing forms under various conditions. The rod-shaped NTD1 crystal diffracts to higher resolution (1.5 Å), comparing to the reported 1.75-Å resolution (Grossoehme et al., 2009). There are two NTD1 molecules in one asymmetric unit (ASU), and they are related by twofold axis. The NTD1 molecule consists of five β-sheets and a single short 3/10 helix in the stable core, surrounded by large loops on the periphery (Fig. 1A), which is consistent with the reported structure of MHV-A59 NTD (PDB number: 3HD4) (Grossoehme et al., 2009). It is notable that the loop corresponding to residues Arg110–Gln121 was missed due to the lack of electron density, and another crystal structure (packing form of NTD2) provides a good supplement at this point.

The crystal of NTD2 was obtained from another diamond-shaped crystal and diffracts to 2.9-Å resolution. Its structure was determined by molecular replacement, using NTD1 monomer as a searching model. Comparing to the structure of NTD1, NTD2 has unambiguous density at Arg110–Gln121 loop, especially at the side chain of Lys113, which was

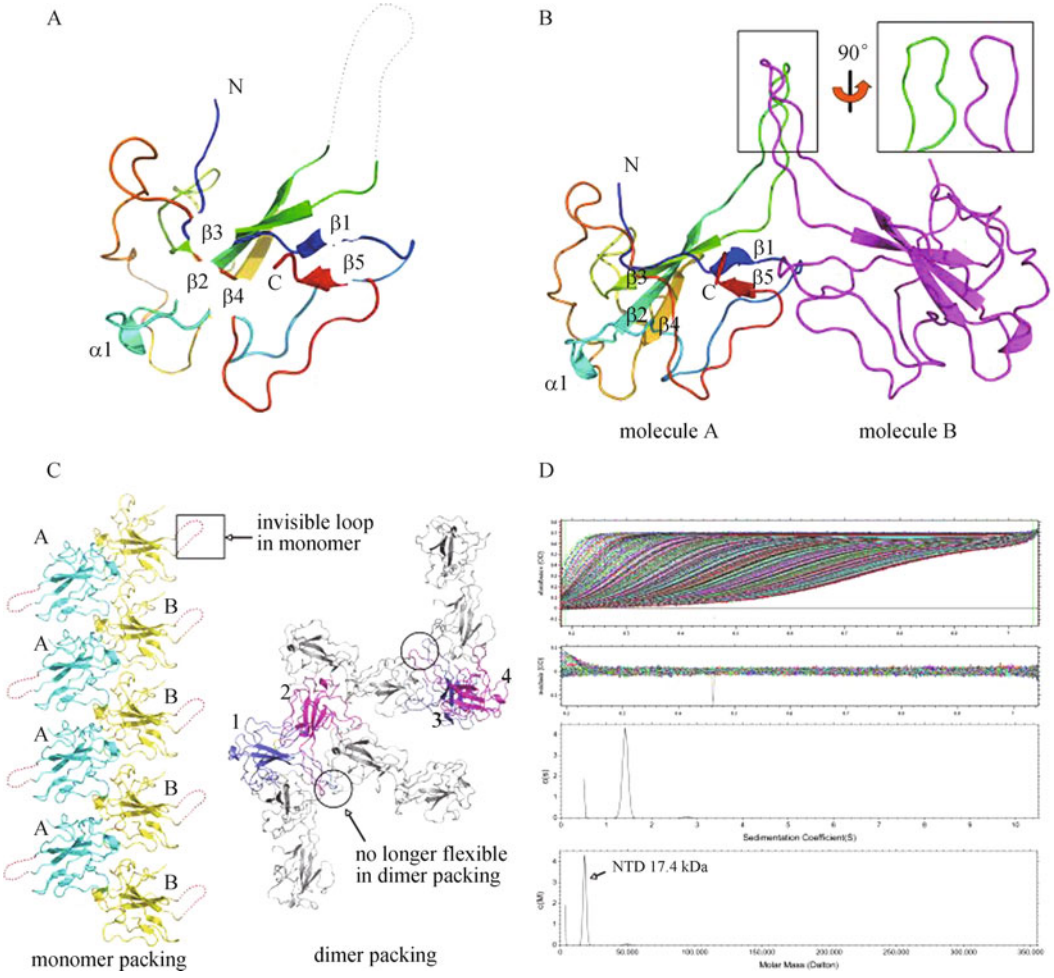

Figure 1. Overall structure of NTD of MHV-A59 N protein. (A) The ribbon diagram of NTD monomer. Secondary structures (helix, strands and loops) are colored in a rainbow fashion, from blue (N terminus) to red (C terminus). A single 3_{10} helix is labeled as α1, and β-strands are numbered from β1 to β5. The disordered loop between strands β2 and β3 is sketched by a dotted line. (B) Overviews of the homodimer, in which molecule A is in rainbow color. (C) Packing mode in the two crystal forms. The comparison clearly explained why the flexible loop in NTD1 is not flexible in NTD2. In NTD1, the dotted loops corresponding to residues Arg110–Gln121 of molecule A and molecule B are exposed to the solvent; while in NTD2, colored molecules 1 and 2 form a dimer, in which the loop is fixed by adjacent molecules. (D) Sedimentation analysis of NTD by analytical ultracentrifugation (AUC). The two curves are the continuous sedimentation coefficient and molar mass distribution of the protein. The molar mass distribution shows a single peak with a molecular mass of 17.4 kDa, which is consistent with the molecular mass of the monomer.

modeled as an Ala in the reported MHV-A59 NTD structure (PDB code: 3HD4). The stabilization of this loop has a straightforward explanation based on the crystal packing (Fig. 1C): the dotted loops, including residues Arg110–Gln121 in NTD1, are exposed to the solvent, but in NTD2, the corresponding loops are fixed at their equilibrium position by the adjacent dimer via hydrogen bonds and hydrophobic interactions between side chains. Moreover, the structures of MHV-NTD molecules in these two different crystal forms are identified to share high similarity with a root-mean-square-deviation (rmsd) of 1.09 Å.

In the 2.0-Å-resolution structure of CTD, two molecules are related by a non-crystallographic twofold axis in one asymmetry unit (Fig. 1B). Each monomeric subunit consists of two anti-parallel β-strands and five α-helices, among which one helix (α3) and two stands (β1, β2) associate tightly with the adjacent monomer. The CTD dimer is a tightly intertwined, domain swapping homo-dimer that looks like a rectangular slab (Fig. 2A). In the final refined structure, several residues of N terminus (Pro282–Cys286), C terminus (Asp382–Arg397), and the part between the two strands could not be observed due to the poor electron density.

Since several homologous structures of NTD and CTD have been reported, we performed a superposition of these structures (Fig. 4A and 5A). The rmsd for two MHV-A59 NTD structures (our structure and the reported 3HD4) is 1.97 Å,

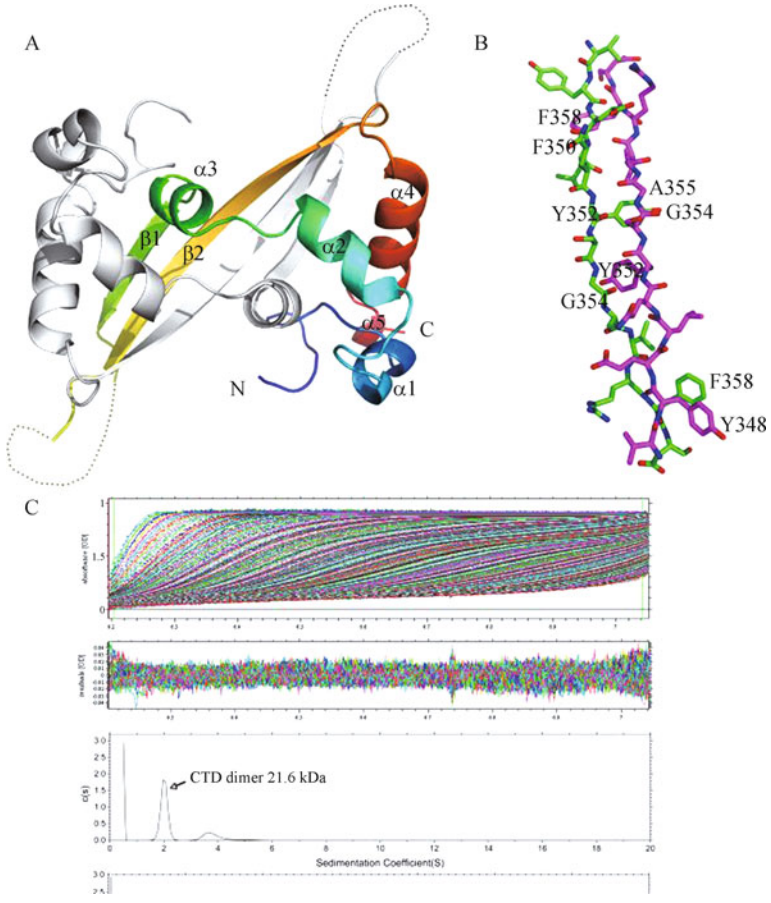

Figure 2. Overall structure of CTD of MHV-A59 N protein. (A) Structure of the CTD dimer. The secondary structures are labeled from α1–α5 and β1–β2. The invisible residues are sketched by a dotted line. (B) Dimeric interface residues within the dimer. Residues belonging to different molecules are differently colored and labeled. All of the N atoms in the sticks are colored blue, and the O atoms are colored red. (C) Sedimentation analysis of CTD by analytical ultracentrifugation (AUC). The continuous sedimentation coefficient and molar mass distribution show two peaks with different molecular masses, which indicate the various association modes of the protein in solution. Here the main peak corresponding 21.6 kDa represents the CTD dimer, and the another peak is meaningless for its too large width and bad symmetry.

while the NTD structures from different coronavirus showed difference, with a total rmsd for MHV-A59 (our structure) vs. SARS-CoV NTDs of 5.39 Å and that vs. IBV of 4.62 Å. The Cα backbones of large loops share less similarity than the helix and strands in the core region. The superposition of CTD structures gives the rmsd of MHV-A59 CTD vs. SARS-CoV CTD is 1.35 Å, and that vs. IBV is 1.51 Å. Amino acid sequence alignment of N proteins from five representative strains of coronavirus also revealed their similarity (Fig. 3). The highly conserved amino acid residues are located in the three strands (β2, β3 and β4) of NTD, and the N-terminal loop in CTD (Fig. 3). These fully conserved residues, in addition to many partially conserved residues, contribute to the majority of the secondary structures (3_{10} helices, α-helices and β-sheets). Some of them also play important roles in RNA binding, which will be discussed in detail.

Oligomerization state of NTD and CTD of MHV N protein

The NTD2 exists as a dimer in the ASU of crystal (Fig. 1B): each monomer looks like a bottle with a narrow neck and big belly. The ends of the "necks" in two subunits cross at an angle of approximately 45 degrees, leaving a gap between the "belly" regions. The flexible loops (Arg110–Gln121) in NTD1 correspond to the crossing necks, which are stabilized by the two bellies from adjacent asymmetric units. It is notable that the two necks seem tightly intertwined, but in fact, they are separated, with a minimum distance of 4.4 Å between two loops.

Since the NTD of MHV N protein exists in two oligomeric forms in the crystals, it is necessary to clarify its oligomerization state, which is monomer, dimer or an equilibrium between the two states. In the dimer structure of NTD2, the calculated

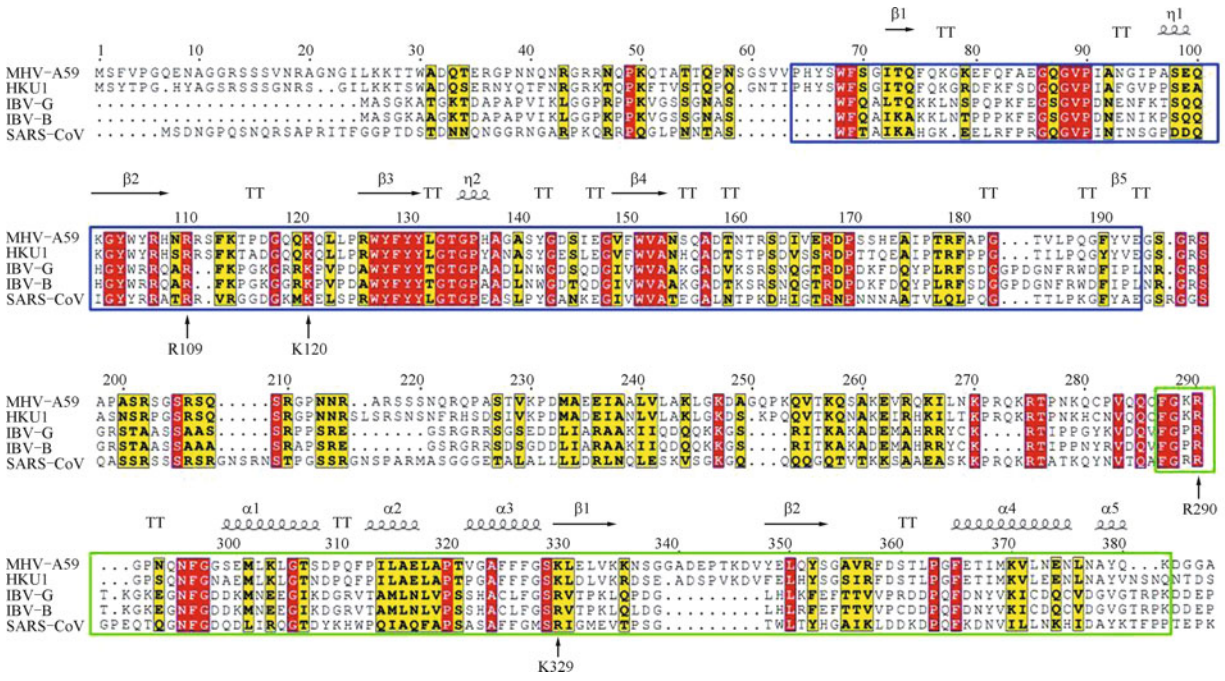

Figure 3. Amino acid sequence alignment of coronavirus N protein. Secondary structure elements of NTD (blue box) and CTD (green box) are labeled above the sequence for the MHV-N. All the sequences for were obtained from Swiss-Prot (MHV [strain A59], NP_045302.1; HKU1, YP_173242.1; SARS-CoV, NP_828858.1; IBV [Gray strain], AAA91856.1; IBV [Beaudette strain], AAA46214.1). 3–10 helices are shown as squiggles and β strands as arrows. Boxes indicate residues that are fully or partially conserved. Fully conserved residues are shaded in red and partially conserved in yellow. The residues labeled with a black arrow are highly conserved in the positive charged surface.

interface area between two molecules is approximately 555 Å^2, with a majority of nonpolar residues (58.21%). These residues associate via hydrophobic interactions and dominate the dimerization. Usually, the protein-protein complexes have a similar structural feature of 17–41 involved residues and a buried surface in the range of 1250–1950 Å^2 (Janin and Chothia, 1990). These suggest a weak interaction between the two molecules inside one homedimer, which is consistent with the sedimentation velocity experiment using analytical ultracentrifugation (AUC). The AUC result proved that the NTD exists as a monomer in solution with a mass of 17.4 kDa (Fig. 1D).

As suggested previously, the dimer of CTD is tightly intertwined and stable. Within the dimer, two subunits are associated through hydrophobic interactions and several salt bridges. These interactions may play an important role in stabilizing the secondary structures of the protein. Area calculations indicate that the buried interface area of each molecule is up to 2338 Å^2 (32.31%, comparing to the total surface area of CTD molecule), formed by a majority of nonpolar residues (45.83% comparing to the complete CTD molecule). Residues located on the β1 strand, including Leu350, Ala355, Tyr352, Gly354, Phe358 and Val356, contribute to strong hydrophobic interactions for dimerization (Fig. 2B). The strong interaction between two subunits in the CTD dimer was also demonstrated by AUC experiment. The molar mass distribution curve showed a main peak of CTD dimer (Fig. 2C). Importantly, the AUC experiment detected the existence of CTD dimers in solution but could not identify other higher-order oligomers.

The potential RNA binding surface of NTD and CTD

Unlike the similarity between NTD secondary structures from the three coronaviruses, there are remarkable difference in their RNA binding surface. The electrostatic distribution on the surface of MHV-N NTD forms a significant positively charged region, which consists of Lys77, Arg109, Arg110, Lys113 and Lys120 (Fig. 4B). All these central residues, including the highly conserved Arg109 and Lys120 (Fig. 3), form a large contiguous surface. Another residue Tyr127 is interpreted to be crucial as the mutant leads to abolish of NTD-TRS binding affinity (Grossoehme et al., 2009), which could be caused by the contribution for the stability of secondary structure. The variation between the three electrostatic surface potentials may result in differences in their RNA binding sites (Fig. 4B).

The electrostatic surface of CTD also appears different. In MHV CTD, the dimer surface looks like a dumb-bell, with a positively charged region (including Lys289, Arg290, Lys303

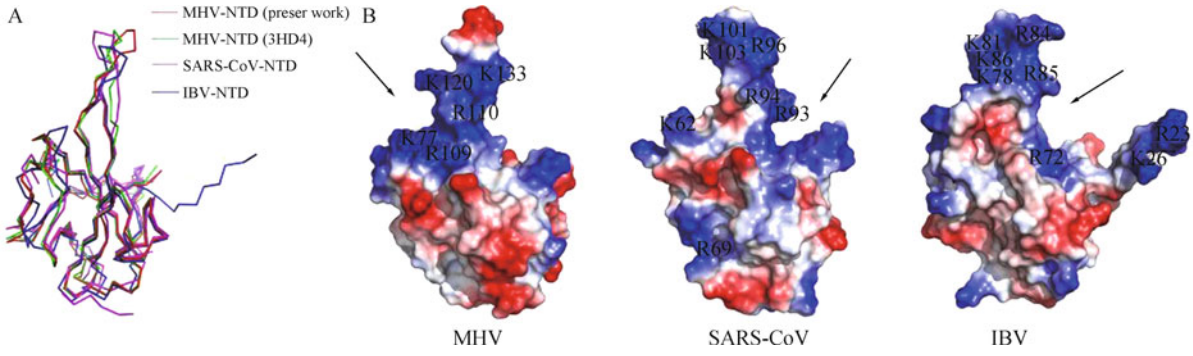

Figure 4. Comparison with other homologous structures of NTD. (A) Superimposed ribbon structures of MHV-A59-N NTD, MHV-A59-N NTD (PDB code: 3HD4), SARS-CoV-N NTD (PDB code: 2OFZ) and IBV-N NTD (PDB code: 2GEC). (B) Distribution of electrostatic potential on the surface of NTDs from MHV, SARS-CoV and IBV. The potential distribution was generated in Pymol (DeLano, 2002). The surface colors are clamped at red (−) or blue (+), which represents the k_T values, where k is the Boltzmann constant and T is the absolute temperature. The variation focuses on the positively charged regions as pointed by arrows.

and Lys329) winding around the middle in a spiral (Fig. 5B). A second positive region consists of Lys334, Lys335 and Arg357 on the other diagonal. On the surface of SARS-CoV and IBV CTDs, the positively charged regions all located at the middle of the dimer in spite of their different shapes and detailed sites. It is expected that this shared pattern might be important for viral nucleocapsid assembly.

DISCUSSION

Models for nucleocapsid formation of related coronaviruses

Because N protein plays an essential role in the formation of viral genome via its self-association, the structural information of N protein from the IBV (group III) and SARS-CoV (closely related to group II) could help propose a possible model for coronavirus nucleocapsid formation. The model is based on two central events: first, both NTD and CTD have multiple putative RNA binding sites. In the N protein of IBV, NTD provides a binding surface for viral RNA through several crucial residues (Lys40, Lys42, Lys43, Arg76, Lys78, Lys81, Arg84 and Arg154) (Fan et al., 2005), while the CTD also provides a positively charged surface to RNA binding (Jayaram et al., 2006). In the N protein of SARS-CoV, the residues (Arg55, Arg59, Arg60, Arg62, Lys67, Arg74, Arg94 and Arg116) of NTD contribute to RNA binding (Saikatendu et al., 2007), residues Thr363–Pro382 of CTD are the responsible interacting partner with RNA (Luo et al., 2006), and the long disordered regions between NTD and CTD was also proved capable of binding RNA (Chang et al., 2009). Moreover, the CTD acts as a dimeric domain to mediate the clustering of N protein. Crystallography and solution structures of IBV-CTD (Jayaram et al., 2006) and SARS-CTD

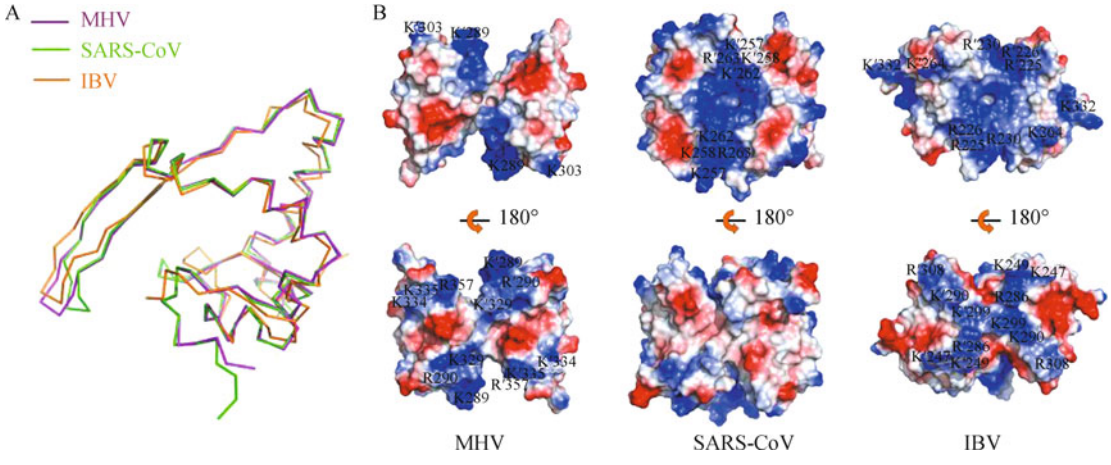

Figure 5. Comparison with other homologous structures of CTD. (A) Superimposed ribbon structures of the CTDs from MHV-A59, SARS-CoV (PDB code: 2CJR) and IBV (PDB code: 2GE8). (B) Distribution of electrostatic surface of the CTD dimers from the MHV, SARS-CoV and IBV N proteins. The potential distribution was generated in Pymol (DeLano, 2002).

(Chen et al., 2007; Takeda et al., 2008) also implicated that the CTD is dimeric characteristic. Therefore, they purposed a model that the dimerization of CTD provides a scaffold, while both the NTD and CTD provide multiple RNA binding sites.

Implications for the function of MHV N protein

The structural alignments show that the overall folding of NTD and CTD domains of MHV N protein were consistent with that of IBV and SARS-CoV. Previous RNA binding assays (Masters, 1992; Cologna et al., 2000; Grossoehme et al., 2009) and the structure surface analysis demonstrated that NTD and CTD both have large positively charged regions for RNA binding. Furthermore, the interface between two CTD molecules in the crystal and sedimentation velocity experiment confirmed a dimeric CTD architecture. Considering the electrostatic distribution (Fig. 5B), positively charged residues (including Lys289, Arg290, Lys303 and Lys329) form a spiral line on the surface, which may provide a helical RNA binding groove.

All the information is consistent with the above models for IBV and SARS-CoV. The conserved model for coronavirus nucleocapsid formation is summarized as following: the N protein dimerizes via its C-terminal domain, providing a platform to recruit viral RNA; the prominent NTD is responsible for recruiting specific or non-specific RNAs; the linkers between NTD and CTD may act as a flexible arm to change the relative position of the two domains (Fig. 6).

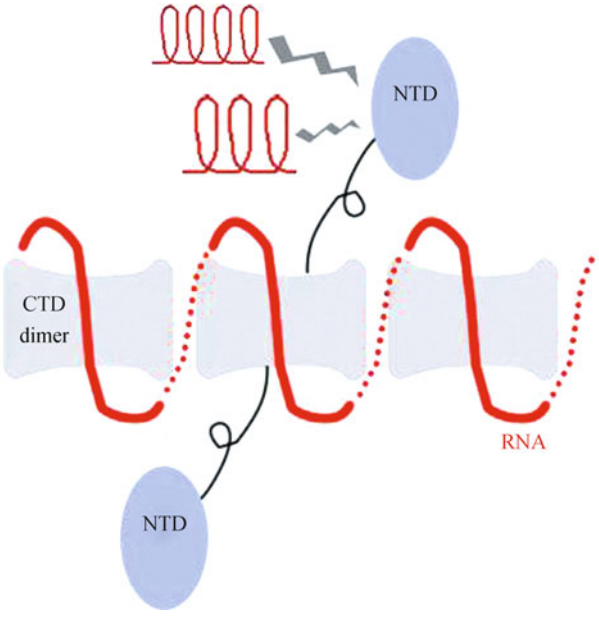

Figure 6. The corroborated conserved RNA-protein binding mechanism in coronavirus. The CTDs dimerize to providing a platform to recruit viral RNA. The prominent NTD is also responsible for recruiting RNA. The linkers between NTD and CTD may act as a flexible arm to change the relative position of the two domains.

This conserved model can explain the fundamental mechanism how coronavirus N protein functions; however, there are still some differences among different coronavirus, e.g., the RNA binding sites in NTD. Although continuous positively charged regions exist in all of the three structures, they clearly show different shapes and locations. This region in IBV protein looks like a clamp to fix RNA, and the positive regions in the SARS-CoV and MHV proteins seem to be a binding groove, but in opposite orientations. The surface structures of different proteins possibly determine the different manners of RNA-NTD binding, including recognition sites, relative position, binding ratios and affinity.

MATERIALS AND METHODS

Cloning, expression and purification of the NTD and CTD

The gene encoding the MHV N protein (MHV-N) was amplified by polymerase chain reaction (PCR) from strain MHV-A59 (located at nucleotides 29,669–31,033 in the genome). Following that, the gene of NTD ($N^{28–195}$) and CTD ($N^{282–397}$) of MHV-N, which are composed by nucleotides 29,752–30,253 and 30,514–30,859, respectively, were sub-cloned for protein expression and crystallization. The NTD was amplified by PCR with the primers: 5′-CGC**GGATCC**ACCACTTGGGCTGACCAAAC-3′ and 5′-CCG**CTCGAG**TTATCCAGAGCCTTCAACAT-3′. The PCR for CTD was performed with the primer pairs: 5′-CGC**GGATCC**CCAGTGCAGCAGTGTTTTGGAAAG-3′ and 5′-CG**CTCGAG**TTAACGCCCTTTTCTTTGGGGCTTTG-3′. The PCR strategy introduced a *Bam*HI site via the forward primer and an *Xho*I site (shown in bold) in the reverse such that the PCR products could be inserted into the pGEX-6p-1 vector (GE Healthcare) using T4 ligase.

The recombinant plasmids were subsequently transformed into *Escherichia coli* strain BL21 (DE3). For each plasmid, a well-isolated colony was transferred into 5 mL LB medium containing 0.1 mg/mL ampicillin and incubated at 37°C overnight. The cell culture was further grown at 37°C in LB medium supplemented with ampicillin (0.1 mg/mL) until the cells reach OD_{600} of 0.8. Protein expression was induced by the addition of 0.4 mM isopropyl-β-D-thiogalactopyranoside (IPTG) for another 16 h at 16°C.

Cells were harvested and lysed by mild sonication in 1 × PBS (phosphate-buffered saline: 140 mM NaCl, 2.7 mM KCl, 10 mM Na_2HPO_4, and 1.8 mM KH_2PO_4, pH 7.3). The supernatants containing the recombinant glutathione S-transferase (GST) fusion proteins, GST-NTD and GST-CTD, were applied to a glutathione sepharose 4B (GE Healthcare) column, followed by on-bead cleavage with PreScission protease (GE Healthcare) to remove the GST tag. Following cleavage, the protein was purified by two chromatography processes: ion exchange chromatography through a pre-packed column Resource S (GE Healthcare), and then gel exclusion chromatography through a Superdex 75 10/30 column (GE Healthcare). SDS-PAGE analysis showed the protein purity over 90%, with expected molar masses. The purified NTD and CTD were concentrated to 5 mg/mL using a spin filter for crystallization. Selenomethionine-labeled NTD and CTD were expressed in *E.coil* strain B834, and purified by the same procedure as the native protein. As there is no methionine in the NTD, we introduced an I72M mutation (numbering

refers to full-length N protein) for Se-Met labeling.

Analytical ultracentrifugation

Sedimentation velocity experiments were performed in a Proteome-lab™XL-1 analytical ultracentrifuge (Beckman coulter). Fresh protein in its own comfortable buffer was centrifuged at 60,000 rpm for 5 h in an An60Ti rotor at 20°C. Protein absorbance was monitored by continuous scans at 280 nm. The protein partial specific volume, buffer viscosity and buffer density were determined using a c(M) distribution model (Schuck, 2000). The protein samples for analytical ultracentrifugation were prepared at a concentration of OD_{280} = 0.75 in the buffer containing 0.2 M HEPES, pH 7.4, 150 mM NaCl.

Crystallization of the NTD and CTD

Crystals of the MHV-N NTD and CTD were both grown at 16°C using the hanging drop diffusion method. One microliter of protein at a concentration of 5 mg/mL was mixed with 1 μL well solution against 200 μL well solution.

Two different crystal forms of the NTD (NTD1 is the I72M mutant and NTD2 is wild type) were obtained. For the native and Se-Met derivation of NTD1, the optimal rod-shaped crystals were obtained in 0.1 M Tris-HCl, pH 8.5 and 8% (*w/v*) PEG8000. The best diamond-shaped crystals of NTD2 were obtained in the condition of 0.2 M ammonium sulfate, 0.1 M MES, pH 6.5, and 30% (*w/v*) PEG-MME 5000 within 10 d. In the case of CTD and its Se-Met derivative, the crystals were obtained in the optimal condition containing 1.3 M sodium citrate (pH 6.5) using crystal seeds initially generated in 1.6 M sodium citrate (pH 6.5).

Prior to data collection, all these crystals were transferred to the reservoir solution (supplemented with 3 M sodium formate) for 5–10 min dehydration before plunged into liquid nitrogen for storage.

Data collection and processing

A 1.5-Å resolution single wavelength desperation (SAD) data set of the Se-Met labeling NTD1 was collected at 100 K using an SBC2 3000 × 3000 CCD detector on beamline BL19-ID at the Advanced Photon Source (APS, Argonne National Laboratory) at the wavelength of

Table 1 Data collection and refinement statistics

data set	NTD1 SeMet	NTD2 native	CTD SeMet
data collection statistics			
cell parameters	*a* = 34.1 Å, *b* = 52.1 Å, *c* = 71.4 Å, *α* = *β* = *γ* = 90°	*a* = 59.9 Å, *b* = 62.1 Å, *c* = 118.9 Å, *α* = *β* = *γ* = 90°	*a* = 66.6 Å, *b* = 66.6 Å, *c* = 50.8 Å, *α* = *β* = *γ* = 90°
space group	$P\ 2_12_12_1$	$P\ 2_12_12_1$	*P422*
resolution (Å)	42.10 (1.53)[c]–1.50	50.00 (2.88)–2.90	66.60 (2.07)–2.00
wavelength (Å)	0.9798	1.0000	0.9800
No. of all reflections	149,451 (3,898)	68,274 (2,334)	101,871 (6,370)
No. of unique reflections	20,680 (886)	11,088 (898)	7,981 (700)
completeness (%)	97.9 (84.8)	97.7 (81.9)	97.3 (87.4)
average I/σ (I)	44.6 (4.5)	16.4 (2.4)	17.6 (3.7)
R_{merge}[a] (%)	7.5 (28.0)	10.7 (51.6)	10.1 (59.0)
refinement statistics			
No. of reflections used (σ(F)> 0)	19,568	10,519	7,606
R_{work}[b] (%)	19.3	23.7	21.9
R_{free}[b] (%)	21.5	28.6	26.6
RMSD bond distance (Å)	0.008	0.017	0.022
RMSD bond angle (°)	1.07	1.68	2.16
average B value ($Å^2$)	13.3	55.9	47.1
ramachandran plot (excluding Pro and Gly)			
Res. in most favored regions	89 (95.7%)	183 (87.6%)	66 (95.7%)
Res. in additionally allowed regions	4 (4.3%)	26 (12.4%)	3 (4.3%)
Res. in generously allowed regions	0 (0%)	0 (0%)	0(0%)

[a] $R_{merge} = \Sigma_h\Sigma_i \mid I_{ih} - \langle I_h \rangle \mid / \Sigma_h\Sigma_i \langle I_h \rangle$, where $\langle I_h \rangle$ is the mean of the observations I_{ih} of reflection h.

[b] $R_{work} = \Sigma(||F_p(obs)| - |F_p(calc)||)/ \Sigma|F_p(obs)|$; R_{free} = R factor for a selected subset (5%) of the reflections that was not included in prior refinement calculations.

[c] Numbers in parentheses are corresponding values for the highest resolution shell.

0.9798 Å. Data for NTD2 was collected to 2.9 Å resolution on beamline BL-17A of the Photon Factory (Japan) using an ADSC Q270 detector. Data of the Se-Met labeling CTD was collected to 2.0-Å resolution on BL-17A of Photon Factory (Japan) at the wavelength of 1.0000 Å. The crystal of NTD1 belongs to the orthorhombic space group $P2_12_12_1$ with the cell parameter of a = 34.1 Å, b = 52.1 Å, c = 71.4 Å, $\alpha = \beta = \gamma = 90°$, while NTD2 belongs to the space group of $P2_12_12_1$ with the cell parameters a = 59.9 Å, b = 62.1 Å, c = 118.9 Å, $\alpha = \beta = \gamma = 90°$. The CTD belongs to the space group $P422$ with cell parameters of $a = b$ = 66.6 Å, c = 50.8 Å, $\alpha = \beta = \gamma = 90°$. Diffraction processing, scaling and integration were performed by using the HKL2000 software package (Otwinowski and Minor, 1997).

Structure determination and refinement

The structure of NTD1 was solved by the single-wavelength anomalous dispersion (SAD) method from a Se-Met derivative. The initial phases were calculated by the program SOLVE (Terwilliger and Berendzen, 1999). Density modification was performed using RESOLVE (Terwilliger, 2000). An initial model of NTD1 was automatically traced using the program ARP/wARP (Perrakis et al., 1999) to approximately 70% of total 138 residues and then further manually built and refined using the programs COOT (Emsley and Cowtan, 2004) and REFMAC5 (Bailey, 1994) at 1.5-Å resolution to a final R_{work} of 19.3% and R_{free} of 21.5%. The residues from Arg110 to Gln121 missed due to lack of electron density. The structure of NTD2 were phased using molecular replacement (MR) in PHASER (McCoy et al., 2007), with the previously solved NTD1 structure as initial searching model and then was manually build using COOT and refined using REFMAC5 (Bailey, 1994) at 2.9-Å resolution to a final R_{work} of 23.7% and R_{free} of 28.6%.

The CTD structure of 2.0-Å resolution was also determined using SAD method. Data was collected and phased following a similar procedure to NTD1 and finally refined to a final R_{work} of 21.9% and R_{free} of 26.6%.

The stereochemistry of all the structures was validated by the program PROCHECK (Laskowski et al., 1993). The statistics of data collection and structure refinement are summarized in Table 1.

ACKNOWLEDGEMENTS

This work was supported by the National Natural Science Foundation of China (Grant No. 30730022), the National Basic Research Program (973 Program) (Grant No. 2006CB806503), the National Programs for High Technology Research and Development Program (863 Program) (Grant Nos. 2006AA02A322 and 2006AA020502), the National Major Project (Grant No. 2009ZX10004-304), and the Chinese Academy of Sciences Grant KSCX2-YW-R-05 to Z.R.

ABBREVIATIONS

ASU, asymmetric unit; AUC, analytical ultracentrifugation; CTD, C-terminal domain; HCoV-229E, human coronavirus; IBV, avian infectious bronchitis virus; MHV, murine hepatitis virus; N protein, nucleocapsid protein; NTD, N-terminal domain; SAD, single-wavelength anomalous dispersion; SARS-CoV, severe acute respiratory syndrome coronavirus; SDS-PAGE, sodium dodecyl sulfate polyacrylamide gel electrophoresis; TGEV, transmissible gastroenteritis virus; TRS, transcriptional regulatory sequence

REFERENCES

Almazán, F., Galán, C., and Enjuanes, L. (2004). The nucleoprotein is required for efficient coronavirus genome replication. J Virol 78, 12683–12688.

Almazán, F., González, J.M., Pénzes, Z., Izeta, A., Calvo, E., Plana-Durán, J., and Enjuanes, L. (2000). Engineering the largest RNA virus genome as an infectious bacterial artificial chromosome. Proc Natl Acad Sci U S A 97, 5516–5521.

Bailey, S., and the Collaborative Computational Project, Number 4. (1994). The CCP4 suite: programs for protein crystallography. Acta Crystallogr D Biol Crystallogr 50, 760–763.

Baric, R.S., Nelson, G.W., Fleming, J.O., Deans, R.J., Keck, J.G., Casteel, N., and Stohlman, S.A. (1988). Interactions between coronavirus nucleocapsid protein and viral RNAs: implications for viral transcription. J Virol 62, 4280–4287.

Bednar, V., Verma, S., Blount, A., and Hogue, B.G. (2006). Importance of MHV-CoV A59 nucleocapsid protein COOH-terminal negative charges. Adv Exp Med Biol 581, 127–132.

Bost, A.G., Carnahan, R.H., Lu, X.T., and Denison, M.R. (2000). Four proteins processed from the replicase gene polyprotein of mouse hepatitis virus colocalize in the cell periphery and adjacent to sites of virion assembly. J Virol 74, 3379–3387.

Casais, R., Thiel, V., Siddell, S.G., Cavanagh, D., and Britton, P. (2001). Reverse genetics system for the avian coronavirus infectious bronchitis virus. J Virol 75, 12359–12369.

Chang, C.K., Hsu, Y.L., Chang, Y.H., Chao, F.A., Wu, M.C., Huang, Y. S., Hu, C.K., and Huang, T.H. (2009). Multiple nucleic acid binding sites and intrinsic disorder of severe acute respiratory syndrome coronavirus nucleocapsid protein: implications for ribonucleocapsid protein packaging. J Virol 83, 2255–2264.

Chen, C.Y., Chang, C.K., Chang, Y.W., Sue, S.C., Bai, H.I., Riang, L., Hsiao, C.D., and Huang, T.H. (2007). Structure of the SARS coronavirus nucleocapsid protein RNA-binding dimerization domain suggests a mechanism for helical packaging of viral RNA. J Mol Biol 368, 1075–1086.

Coley, S.E., Lavi, E., Sawicki, S.G., Fu, L., Schelle, B., Karl, N., Siddell, S.G., and Thiel, V. (2005). Recombinant mouse hepatitis virus strain A59 from cloned, full-length cDNA replicates to high titers in vitro and is fully pathogenic *in vivo*. J Virol 79, 3097–3106.

Cologna, R., Spagnolo, J.F., and Hogue, B.G. (2000). Identification of nucleocapsid binding sites within coronavirus-defective genomes. Virology 277, 235–249.

DeLano, W. (2002). The PyMOL Molecular Graphics System. DeLano Scientific, Palo Alto, CA, USA.

Drosten, C., Günther, S., Preiser, W., van der Werf, S., Brodt, H.R., Becker, S., Rabenau, H., Panning, M., Kolesnikova, L., Fouchier, R.A., *et al*. (2003). Identification of a novel coronavirus in patients with severe acute respiratory syndrome. N Engl J Med 348, 1967–1976.

Emsley, P., and Cowtan, K. (2004). Coot: model-building tools for molecular graphics. Acta Crystallogr D Biol Crystallogr 60, 2126–2132.

Fan, H., Ooi, A., Tan, Y.W., Wang, S., Fang, S., Liu, D.X., and Lescar, J. (2005). The nucleocapsid protein of coronavirus infectious bronchitis virus: crystal structure of its N-terminal domain and multimerization properties. Structure 13, 1859–1868.

Fleischauer, A.T., and the Centers for Disease Control and Prevention

(CDC). (2003). Outbreak of severe acute respiratory syndrome—worldwide, 2003. MMWR Morb Mortal Wkly Rep 52, 226–228.

Grossoehme, N.E., Li, L., Keane, S.C., Liu, P., Dann, C.E. 3rd, Leibowitz, J.L., and Giedroc, D.P. (2009). Coronavirus N protein N-terminal domain (NTD) specifically binds the transcriptional regulatory sequence (TRS) and melts TRS-cTRS RNA duplexes. J Mol Biol 394, 544–557.

Huang, Q., Yu, L., Petros, A., Gunasekera, A., Liu, Z., Xu, N., Hajduk, P., Mack, J., Fesik, S., and Olejniczak, E. (2004). Structure of the N-Terminal RNA-Binding domain of the SARS CoV nucleocapsid protein. Biochemisty 43, 6059–6063.

Hurst, K.R., Kuo, L., Koetzner, C.A., Ye, R., Hsue, B., and Masters, P. S. (2005). A major determinant for membrane protein interaction localizes to the carboxy-terminal domain of the mouse coronavirus nucleocapsid protein. J Virol 79, 13285–13297.

Janin, J., and Chothia, C. (1990). Minireview:The Structure of Protein-Protein Recognition Sites. J Biol Chem 265, 16221–16224.

Jayaram, H., Fan, H., Bowman, B.R., Ooi, A., Jayaram, J., Collisson, E.W., Lescar, J., and Prasad, B.V.V. (2006). X-ray structures of the N- and C-terminal domains of a coronavirus nucleocapsid protein: implications for nucleocapsid formation. J Virol 80, 6612–6620.

Ksiazek, T.G., Erdman, D., Goldsmith, C.S., Zaki, S.R., Peret, T., Emery, S., Tong, S., Urbani, C., Comer, J.A., Lim, W., *et al*, and the SARS Working Group. (2003). A novel coronavirus associated with severe acute respiratory syndrome. N Engl J Med 348, 1953–1966.

Kuo, 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., and Stohlman, S.A. (1978). RNA of mouse hepatitis virus. J Virol 26, 236–242.

Laskowski, R.A., Moss, D.S., and Thornton, J.M. (1993). Main-chain bond lengths and bond angles in protein structures. J Mol Biol 231, 1049–1067.

Luo, H., Chen, J., Chen, K., Shen, X., and Jiang, H. (2006). Carboxyl terminus of severe acute respiratory syndrome coronavirus nucleocapsid protein: self-association analysis and nucleic acid binding characterization. Biochemisty 45, 11827–11835.

Macneughton, M.R., and Davies, H.A. (1978). Ribonucleoprotein-like structures from coronavirus particles. J Gen Virol 39, 545–549.

Masters, P.S. (1992). Localization of an RNA-binding domain in the nucleocapsid protein of the coronavirus mouse hepatitis virus. Arch Virol 125, 141–160.

McCoy, A.J., Grosse-Kunstleve, R.W., Adams, P.D., Winn, M.D., Storoni, L.C., and Read, R.J. (2007). Phaser crystallographic software. J Appl Cryst 40, 658–674.

Mir, M.A., and Panganiban, A.T. (2006). Characterization of the RNA chaperone activity of hantavirus nucleocapsid protein. J Virol 80, 6276–6285.

Molenkamp, R., and Spaan, W.J.M. (1997). Identification of a specific interaction between the coronavirus mouse hepatitis virus A59 nucleocapsid protein and packaging signal. Virology 239, 78–86.

Nelson, G.W., Stohlman, S.A., and Tahara, S.M. (2000). High affinity interaction between nucleocapsid protein and leader/intergenic sequence of mouse hepatitis virus RNA. J Gen Virol 81, 181–188.

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

Perrakis, A., Morris, R.M., and Lamzin, V.S. (1999). Automated protein model building combined with iterative structure refinement. Nat Struct Biol 6, 458–463.

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.

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

Saikatendu, K.S., Joseph, J.S., Subramanian, V., Neuman, B.W., Buchmeier, M.J., Stevens, R.C., and Kuhn, P. (2007). Ribonucleocapsid formation of severe acute respiratory syndrome coronavirus through molecular action of the N-terminal domain of N protein. J Virol 81, 3913–3921.

Sawicki, S.G., Sawicki, D.L., Younker, D., Meyer, Y., Thiel, V., Stokes, H., and Siddell, S.G. (2005). Functional and genetic analysis of coronavirus replicase-transcriptase proteins. PLoS Pathog 1, e39.

Schuck, P. (2000). Size-distribution analysis of macromolecules by sedimentation velocity ultracentrifugation and lamm equation modeling. Biophys J 78, 1606–1619.

Stohlman, S.A., Baric, R.S., Nelson, G.N., Soe, L.H., Welter, L.M., and Deans, R.J. (1988). Specific interaction between coronavirus leader RNA and nucleocapsid protein. J Virol 62, 4288–4295.

Sturman, L.S., and Holmes, K.V. (1983). The molecular biology of coronaviruses. Adv Virus Res 28, 35–112.

Takeda, M., Chang, C.K., Ikeya, T., Güntert, P., Chang, Y.H., Hsu, Y. L., Huang, T.H., and Kainosho, M. (2008). Solution structure of the c-terminal dimerization domain of SARS coronavirus nucleocapsid protein solved by the SAIL-NMR method. J Mol Biol 380, 608–622.

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.

van der Meer, Y., Snijder, E.J., Dobbe, J.C., Schleich, S., Denison, M. R., Spaan, W.J., and Locker, J.K. (1999). Localization of mouse hepatitis virus nonstructural proteins and RNA synthesis indicates a role for late endosomes in viral replication. J Virol 73, 7641–7657.

Verma, S., Bednar, V., Blount, A., and Hogue, B.G. (2006). Identification of functionally important negatively charged residues in the carboxy end of mouse hepatitis coronavirus A59 nucleocapsid protein. J Virol 80, 4344–4355.

You, J.-H., Reed, M.L., and Hiscox, J.A. (2007). Trafficking motifs in the SARS-coronavirus nucleocapsid protein. Biochem Biophys Res Commun 358, 1015–1020.

Yu, I.M., Oldham, M.L., Zhang, J., and Chen, J. (2006). Crystal structure of the severe acute respiratory syndrome (SARS) coronavirus nucleocapsid protein dimerization domain reveals evolutionary linkage between corona- and arteriviridae. J Biol Chem 281, 17134–17139.

Zhang, X., Liao, C.-L., and Lai, M.M. (1994). Coronavirus leader RNA regulates and initiates subgenomic mRNA transcription both *in trans* and *in cis*. J Virol 68, 4738–4746.

Zúñiga, S., Sola, I., Moreno, J.L., Sabella, P., Plana-Durán, J., and Enjuanes, L. (2007). Coronavirus nucleocapsid protein is an RNA chaperone. Virology 357, 215–227.

