Crystal Structure of the Severe Acute Respiratory Syndrome (SARS) Coronavirus Nucleocapsid Protein Dimerization Domain Reveals Evolutionary Linkage between *Corona*- and *Arteriviridae**

I-Mei Yu[‡1]**, Michael L. Oldham**[‡1]**, Jingqiang Zhang**[§]**, and Jue Chen**[‡2]

From the ‡*Department of Biological Sciences and the Cancer Center, Purdue University, West Lafayette, Indiana 47907
and the* §*State Key Laboratory for Biocontrol, Zhongshan University, Guangzhou 510275, China*

The causative agent of severe acute respiratory syndrome (SARS) is the SARS-associated coronavirus, SARS-CoV. The nucleocapsid (N) protein plays an essential role in SARS-CoV genome packaging and virion assembly. We have previously shown that SARS-CoV N protein forms a dimer in solution through its C-terminal domain. In this study, the crystal structure of the dimerization domain, consisting of residues 270–370, is determined to 1.75 Å resolution. The structure shows a dimer with extensive interactions between the two subunits, suggesting that the dimeric form of the N protein is the functional unit *in vivo*. Although lacking significant sequence similarity, the dimerization domain of SARS-CoV N protein has a fold similar to that of the nucleocapsid protein of the porcine reproductive and respiratory syndrome virus. This finding provides structural evidence of the evolutionary link between *Coronaviridae* and *Arteriviridae*, suggesting that the N proteins of both viruses have a common origin.

Coronaviruses are enveloped, single-stranded, positive-sense RNA viruses that infect a variety of mammals and birds. Although previously identified human coronaviruses cause only mild respiratory infections, in the 2003 outbreak of severe acute respiratory syndrome (SARS),[3] a disease caused by a new type of coronavirus (SARS-CoV), there were more than 8,000 cases resulting in 774 deaths (~10% mortality). Phylogenetic analysis suggests that SARS-CoV diverged early from group 2 coronaviruses and has evolved independently for a long period of time (1).

The coronavirus genome, containing ~30,000 bases, is the largest among positive-sense RNA viruses (2, 3). It encodes non-structural proteins including the RNA polymerase and helicase, as well as the spike (S), envelope (E), membrane (M), and nucleocapsid (N) structural proteins. The coronavirus virion is about 120 nm in diameter and consists of a lipid envelope containing three or four anchored glycoproteins and a helical ribonucleoprotein core (4). The surface projections forming the crown-like structure observed via electron microscopy are made up of the S protein, which is responsible for receptor recognition and membrane fusion (5–9). The integral membrane proteins M and E are essential for virus budding. When co-expressed in animal cells, the M and E proteins are sufficient to form virus-like particles (10). The N protein interacts with the viral genome to form the ribonucleoprotein core and has been shown to be involved in viral RNA synthesis, transcriptional regulation of genomic RNA, translation of viral proteins, and budding (2, 11, 12).

Coronaviruses are related to arteriviruses by their similar genome organizations and viral replication mechanisms (3, 13). Recently, *Coronaviridae* and *Arteriviridae* were united to form the new order *Nidovirales*. The name of the order comes from a property common to both viruses, a nested set of subgenomic mRNAs for structural protein expression (Latin *nidus*, meaning nest). The replicase genes of arteriviruses and coronaviruses are thought to have a common origin since they both have conserved domains that are present at the same relative positions (2, 14, 15). The structural proteins, however, are thought to be unrelated due to differences in protein size and lack of sequence similarity (15).

Coronavirus N proteins are composed of three distinct domains. High resolution structures of the N-terminal domain (~130 residues) were determined by NMR (16) and crystallography (17). This region folds similarly to the U1A RNA-binding protein and is suggested to bind RNA (16, 17). The central region of the N protein has also been shown by several laboratories to be an RNA-binding domain (18–21). No structural information is available for the central domain, possibly due to its high positive charge and flexible nature. Recently, secondary structure elements of residues 248–365 of SARS-CoV N protein have been reported from NMR studies; however, the tertiary structure was not resolved (22). We have previously shown that the full-length SARS N protein, consisting of 422 amino acids, forms a dimer in solution through its C-terminal domain. In this study, we report the crystal structure of the dimerization domain of SARS-CoV N protein consisting of residues 270–370. Consistent with biochemical studies showing that this domain mediates dimer formation of the N protein (17, 23), the structure consists of a dimer with extensive interactions between subunits, suggesting that the N protein is not stable in the monomeric form and that the dimeric structure represents the functional unit of the N protein. Furthermore, the dimerization domain of SARS-CoV N protein shows a similar fold to that of the N protein of porcine reproductive and respiratory syndrome virus (PRRSV), a member of *Arteriviridae*. The structural similarity between these two N proteins, in the absence

TABLE 1

Data collection and refinement statistics

Erroneous I/sigma values have been omitted, since overlapping reflections due to twinning render proper background estimates impossible.

Data Collection	Refinement set	Se-Met (phasing set)		
		Peak	Inflection	Remote
Space group C2				
a, Å	124.2	124.3		
b, Å	50.5	50.6		
c, Å	41.5	41.6		
β, °	108.9	108.8		
Resolution range, Å	30-1.75	50-1.9	50-2.2	50-2.2
Wavelength, Å	0.97929	0.97929	0.97940	0.94285
Unique reflections	24479	27506	24045	24131
Completeness[a]	99.2 (99.6)	100.0 (100.0)	99.2 (100.0)	99.2 (100.0)
Redundancy	2.9 (2.9)	8.2 (7.8)	9.4 (9.2)	9.4 (9.2)
R_{sym}[b]	7.4 (15.0)	10.6 (16.5)	12.8 (15.5)	15.1 (21.2)
Refinement				
R_{work}[c], %	18.7			
R_{free}, %	23.5			
Resolution range, Å	30-1.75			
No. of reflections used	23088			
No. of molecules in asymmetric unit	2			
No. of non-hydrogen atoms in protein	1511			
No of water molecules	212			
No of sulfate ions	1			
r.m.s.d.[d] from ideality				
Bond lengths, Å	0.02			
Bond angles, °	1.67			
Estimated coordinate error (Luzzati), Å	0.16			
Average B-factor Å^2	18.0			
Ramachandran statistics (non glycine)				
Most favored, %	97.5			
Allowed, %	2.5			
Generously allowed, %	0.0			
Disallowed, %	0.0			

[a] Values in parentheses are for the highest resolution shell.
[b] $R_{sym} = \Sigma|I_i - \langle I \rangle| / \Sigma I_i$, where I_i is the intensity of the i*th* observation and $\langle I \rangle$ is the mean intensity of the reflection.
[c] $R = \Sigma\|F_o| - |F_o\|/\Sigma|F_o|$, where F_o and F_c are the observed and calculated structure factors amplitudes. R_{free} is calculated using 5% of the total reflections.
[d] r.m.s.d., root mean square deviation.

of significant sequence similarity, adds further evidence that *Coronaviridae* and *Arteriviridae* have evolved from a common ancestor.

EXPERIMENTAL PROCEDURES

Cloning, Expression, and Purification—DNAs encoding various fragments of SARS-CoV (strain GD01) nucleocapsid protein, SARS N, were cloned into a pMCSG7 plasmid with an N-terminal His$_6$ tag and a tobacco etch virus protease cleavage site. For selenomethionine (SeMet) derivatized protein, the methionine auxotroph strain B834(DE3) cells (Invitrogen) were transformed with the expression plasmid. The cells were grown to log phase at 37 °C in minimal medium supplemented with SeMet (Sigma) and cooled to 16 °C, and protein expression was induced by the addition of 0.1 mM isopropyl-1-thio-β-D-galactopyranoside. Cells were harvested by centrifugation at 4000 × *g* for 20 min. Cell pellets were resuspended in lysis buffer (20 mM Tris, pH 8.5, 200 mM NaCl) and lysed by sonication. Cell lysate was centrifuged at 90,000 × *g* for 45 min at 4 °C, and the supernatant was loaded onto a cobalt column (Clontech; TALON metal affinity resin). The column was then washed with lysis buffer followed by lysis buffer plus 5 mM imidazole. The protein was eluted with lysis buffer plus 100 mM imidazole. Fractions containing SARS N proteins were pooled and dialyzed against 20 mM Tris, pH 8.5, 150 mM NaCl to remove the imidazole. Tobacco etch virus protease was added to 10% (w/w), and the reaction mixture was incubated at 30 °C for 10 h. Complete removal of the affinity His tag was monitored by SDS-PAGE. Further purification was achieved by size exclusion chromatography (Amersham Biosciences; Superdex 75) in lysis buffer plus 10 mM dithiothreitol to keep all SeMet in the reduced form. Incorporation of SeMet was confirmed by mass spectrometry and fluorescence scans around the selenium potassium absorption edge.

The native N protein fragments were expressed in BL21star(DE3) cells and purified similarly except that no dithiothreitol was used in the buffer.

Crystallization and Data Processing—Protein samples were dialyzed against 10 mM Tris at pH 8.5 and 70 mM NaCl and concentrated to 45 mg/ml. Crystals were grown by mixing protein solution with the reservoir solution containing 30–33% pentaerythritol ethoxylate, 15/4 ethoxylate/hydroxyl, 50 mM (NH$_4$)$_2$SO$_4$, and 50 mM Bis-Tris, pH 6.5, at 1:1 ratio in sitting drops by vapor diffusion at 4 °C. Crystals were looped out of the drop and flash-frozen in liquid nitrogen. Data were collected at 100 K with a Quantum-Q315 CCD (ADSC) detector on beamline 19-ID at the Advanced Photon Source (Structural Biology Center, Argonne National Laboratory). X-ray diffraction data for the initial structure determination were collected to 2.2 Å resolution from crystals containing SeMet protein (see Table 1, phasing set). Data from both lattices of the twin domains were processed separately with HKL2000 (24) and then merged together in Scalepack. Phases were obtained from the four selenium sites found by multiple wavelength anomalous diffraction using the program SOLVE (25) and improved by density modification using CNS (26). The initial model was built automatically using Arp/Warp in CCP4 (27). A higher resolution data set (1.75 Å) from a SeMet protein crystal (see Table 1, refinement set) was used for refinement. The resolution was extended over several cycles of model building in O (28) and refinement in REFMAC5 in CCP4 (27).

RESULTS

Crystallization and Structure Determination—The full-length SARS-CoV N protein (SARS N) and 18 truncated constructs were expressed in

bacteria and purified for crystallization. Although most constructs yielded either no crystals or merely clusters of small needles, a C-terminal construct containing residues 270–370 (cSARS-N) was readily crystallized to give well diffracting crystals. Although the crystals appear to be composed of a single lattice as viewed under a microscope fitted with a light polarizer, the diffraction pattern indicated that the crystals were unusually twinned, resulting in two overlapping monoclinic lattices of similar unit cell dimensions. The two lattices are related by a 180° rotation around the *y* axis followed by a 50° rotation around the *z* axis. As a result, nearly half of the observable reflections overlapped to some extent. Both native and SeMet derivatized proteins formed twinned crystals regardless of changes in temperature, precipitant, pH, and salt concentration. Because of the large number of partially overlapped reflections, only the central four pixels of each reflection were measured, these being proportional to the integrated intensity presuming the profile was fairly constant (details of the data processing procedure will be published elsewhere).

The structure of cSARS-N was determined by multiwavelength anomalous diffraction using a SeMet derivative to 1.75 Å resolution. The crystal belongs to space group C2, and the asymmetric unit contains a dimer of two subunits related by a 180° rotation. The final model, containing residues 270–366 of one subunit and residues 274–369 of the other, was refined to R_{work}/R_{free} of 18.7/23.5 (Table 1). A representative $2F_o - F_c$ electron density map is shown in Fig. 1. The model, analyzed with PROCHECK (29), has 97.5% of non-glycine residues in the most favored regions of the Ramachandran plot, with no disallowed residues.

Overall Structure—Sequence comparison shows that the dimerization domain of the nucleocapsid protein is conserved among the three groups of coronaviruses, suggesting a common structural and functional role of this domain (Fig. 2*A*). The monomer of cSARS-N contains five short α-helices, one 3₁₀ helix, and two β-strands. The overall shape of the monomer resembles the letter C, with one edge formed by a β-hairpin extending away from the rest of the molecule (Fig. 2*B*). In contrast to many small proteins that are remarkably compact, the monomeric cSARS-N domain folds into an extended conformation with a large cavity in its center. It is therefore likely that the N protein is not stable in the monomeric form and that higher oligomerization of the polypeptide is necessary to produce a stable conformation. Consistent with previous biochemical studies suggesting that the full-length SARS N protein forms a dimer in solution through its C-terminal domain (17, 23), a compact dimeric form of cSARS-N was observed in the crystal (Fig. 2*C*). The dimer has a flat structure of approximate dimensions 48 × 42 × 25 Å. The two subunits in the dimer are almost identical, with a root mean square deviation of 0.36 Å over 92 C$_α$ atoms out of 101 residues and the largest displacement of 0.56 Å. The dimer interface is largely formed by insertion of the β-hairpin of one subunit into the cavity of the opposite subunit (Fig. 2*C*). As a result, the four β-strands of the two subunits form an anti-parallel β-sheet, with 10 hydrogen bonds formed across the dimer interface by the main chain atoms of residues 330–340. Two large, pyramidal hydrophobic cores form a bow tie-shaped pocket that further stabilizes the dimer. Each core consists of base residues Phe-287, Leu-292, Trp-302, Ile-305, Phe-308, Tyr-334, Ile-338, Leu-340, Leu-354, Ile-358 from one subunit and apex residues Phe-315, Phe-316, Ile-321, and Leu-332 from the other subunit. The buried area in the dimer interface is 2093 Å² per subunit, which accounts for 28% of the total solvent-accessible surface area of each subunit. The extended structure of the monomeric subunit and the extensive interactions between the subunits within the dimer suggest that the dimeric arrangement observed crystallographically represents
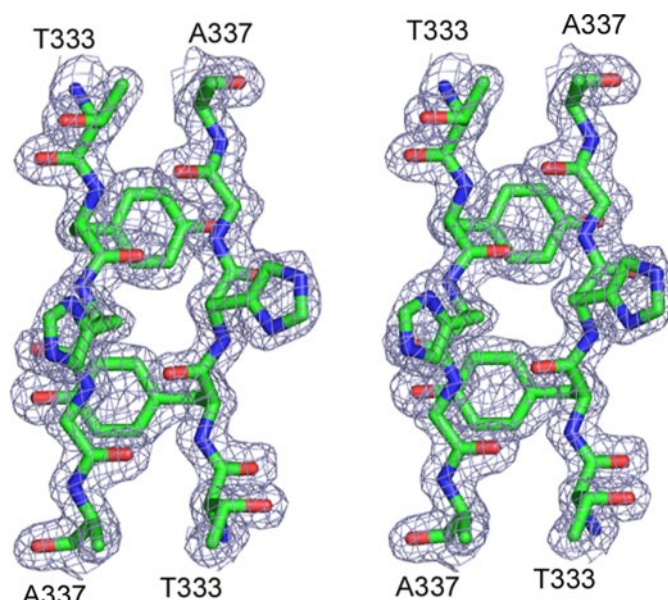

FIGURE 1. **Electron density map.** Stereo view of electron density map ($2F_o - F_c$, 1.5σ) around β-strand residues 333–337 of both subunits at the dimer interface.

the biological architecture of the coronavirus N protein. The structure of cSARS-N is consistent with biochemical studies showing that residues 1–284 of the SARS N protein are dispensable for dimer formation (23). The strong interactions between the two subunits explain the previous observation that a significant amount of denaturant (4 M urea) was necessary to disrupt the dimer in solution (23).

A recent NMR study reported secondary structural assignments of a SARS N protein construct that is similar to cSARS-N (22). Despite the difficulty in resolving the three-dimensional structure, a dimeric interface consisting of a four-stranded anti-parallel β-sheet and two α-helices was proposed (22). Comparing the crystal structure with the proposed NMR structure, one of the most notable differences is the location of the longest helix, helix E (residues 346–357). In the crystal structure, helices E of both subunits are located at the edge of the dimer molecule, away from each other (Fig. 2*C*), whereas in the structure proposed by NMR, they are placed at the interface to stabilize the dimer (22).
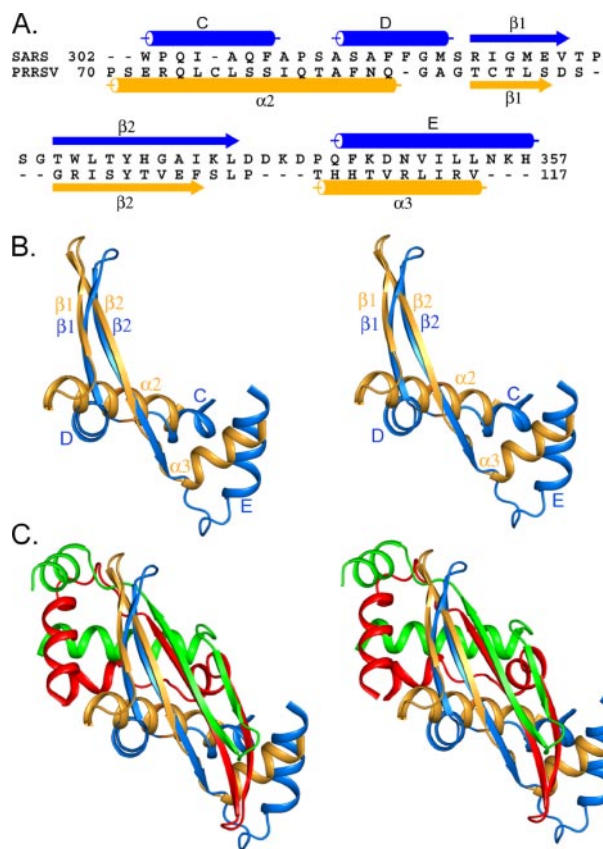
Structural Similarity to the Nucleocapsid Protein of Arterivirus—A DALI search of the Protein Data Bank did not identify any other proteins with similar folds to that of cSARS-N. However, common structural features can be found between cSARS-N and the nucleocapsid protein of PRRSV (Fig. 2). PRRSV belongs to the recently recognized family *Arteriviridae,* which is in the same order as *Coronaviridae.* The crystal structure of the C-terminal 65 residues of PRRSV N protein (30) is a homodimer. Each subunit consists of two central anti-parallel β-strands flanked by three α-helices. Although the sequences of PRRSV N and cSARS-N have no significant similarity (Fig. 3*A*), superposition of the monomeric N proteins of PRRSV and SARS-CoV shows that the two have similar folds in a core region of ∼50 residues (Fig. 3, *B* and *C*). The central β-strands are superimposable, and the helices following the β-strands are of similar length and position (Fig. 3*B*). In addition, as proposed in the recent NMR study (22), the dimer interface of the two N proteins are very similar, noted by the intertwining of the central β-strands to form a four-stranded sheet (Fig. 3*C*). This is evident by a root mean square deviation of 2.24 Å for the C$_α$ positions among 38 residues at the dimer interface.
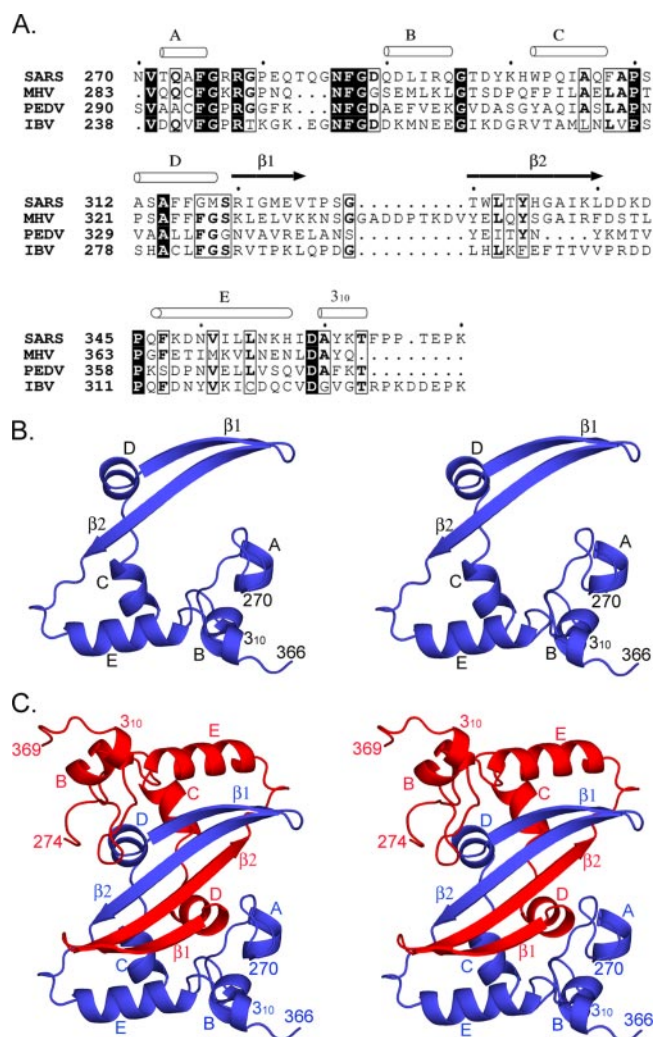
FIGURE 2. **Structure of the SARS-CoV nucleocapsid dimerization domain.** *A*, sequence alignment of the nucleocapsid protein dimerization domains of SARS-CoV (strain GD01) and representative coronaviruses from the three previously defined groups: mouse hepatitis virus (*MHV*, group 2), porcine epidemic diarrhea virus (*PEDV*, group 1), and avian infectious bronchitis virus (*IBV*, group 3). *Solid boxes* indicate strictly conserved residues, whereas *open boxes* indicate partially conserved residues (shown in *bold letters*). The secondary structure elements of SARS N protein are represented by *cylinders* (α-helices) and *arrows* (β-strands) above the sequences. *B*, cSARS-N monomer in stereo. *C*, stereo view of cSARS-N dimer with the two subunits colored in *blue* and *red*. Secondary structure elements and terminal residue numbers are labeled in the ribbon diagrams. The figures were prepared with ESPript (48) or PyMOL.

FIGURE 3. **Structural similarity between *Coronavirus* and *Arterivirus* nucleocapsid proteins.** *A*, structure-based alignment of the dimerization domains of SARS N and PRRSV N. The secondary structure elements of SARS N (*blue*) and PRRSV N (*gold*) are shown as *cylinders* (α-helices) and *arrows* (β-strands). *B*, superposition of SARS N residues 302–357 (*blue*) with PRRSV N residues 70–117 (*gold*) monomers in stereo. This view is obtained by a 90° counterclockwise rotation of the structure shown in Fig. 2*B* about a vertical axis. *C*, superposition of the dimeric SARS N (*blue* and *red*) and PRRSV N (*gold* and *green*) in stereo. The figures were prepared using Alscript (49) or PyMOL.

DISCUSSION

Specific packaging of the viral genome into the virion is a critical step in the life cycle of an infectious virus. The N protein plays an essential role in this process through self-association and interactions with viral RNA and other viral proteins. In an effort to understand the mechanism of how SARS N protein functions, we determined the crystal structure of its C-terminal dimerization domain. Extensive hydrogen-bonding and hydrophobic interactions are observed between the two subunits within the dimer found in the crystal, suggesting that the functional unit of the N protein is dimeric. Strong protein-protein interactions at the dimerization region may be essential to hold the putatively monomeric, highly charged RNA-binding domains in close proximity (16), thereby facilitating the formation of a large helical nucleocapsid core. Association of the N protein dimers is necessary for further assembly of the core. The full-length dimeric N protein has a propensity to form

tetramers and higher molecular weight oligomers *in vitro* (23). A serine/arginine-rich motif (residues 184–196) was shown to be important for N protein oligomerization (31). Since constructs containing residues 211–422 or 285–422 of SARS N do not form oligomers larger than dimers (23), it is possible that the serine/arginine-rich motif located outside the dimerization domain is necessary to mediate further association of N protein dimers. Recently, the C-terminal 45 residues of the mouse hepatitis virus N protein were shown to be the major determinant for interaction with the M protein (32). Association of the N protein with the M protein may also play a role in the assembly of the nucleocapsid core into a progeny virion.

RNA viruses have high rates of sequence divergence and genome recombination, and thus it is often a challenge to study the evolutionary relationships among viruses. Structural studies have become an important method for revealing distant relationships among viruses. Among all lipid-enveloped RNA viruses, high resolution structures of the proteins that package the viral genome into virions, often termed the nucleocapsid or capsid proteins, are currently available from five virus families (Fig. 4) (*Retroviridae* are not included in this discussion because their replication pathway is through a DNA intermediate (33)). Although functionally equivalent, the size and structure of nucleocapsid proteins are remarkably diverse among these five virus families. The nucleoprotein of borna disease virus (BDV), a single-stranded RNA virus in the family *Bornaviridae*, con-
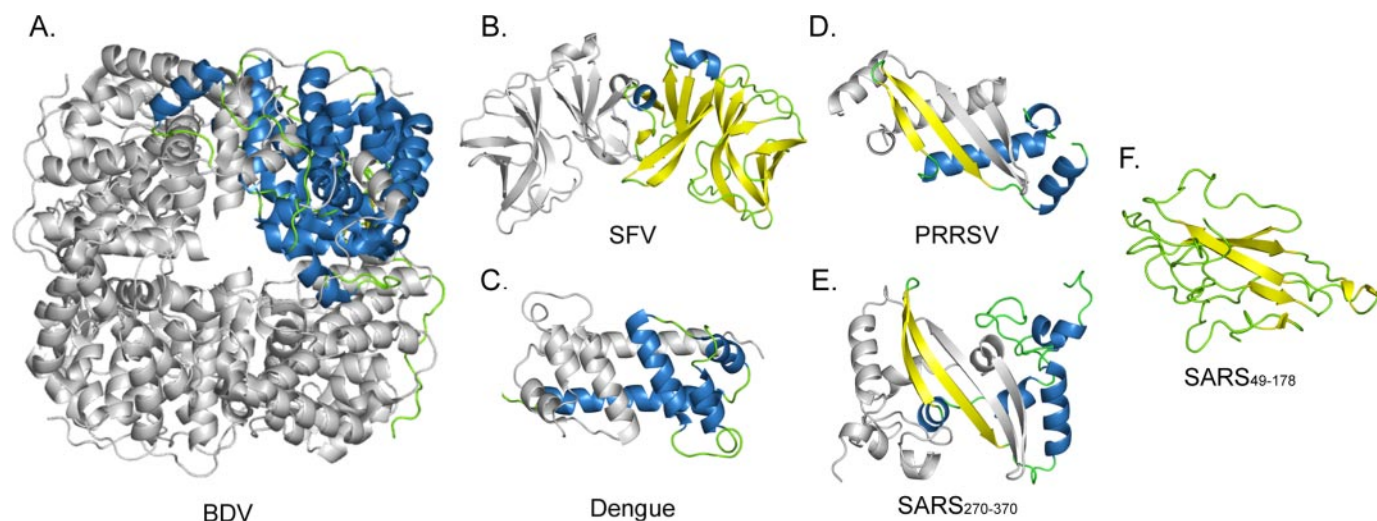
FIGURE 4. **Structural diversity of enveloped RNA virus nucleocapsid proteins.** Representative known structural domains of nucleocapsid proteins of *Bornaviridae* (BDV), *Togaviridae* (*SFV*), *Flaviviridae* (dengue), *Arteriviridae* (PRRSV), and *Coronaviridae* (SARS-CoV) are shown. One subunit in the oligomer is colored based on its secondary structure elements: *blue* (α-helices), *yellow* (β-strands), and *green* (coils); the other subunits are colored in *gray*. *A*, the crystallographic tetramer of BDV nucleocapsid protein (1N93) (34). *B*, Semliki Forest virus nucleocapsid protein (1VCP). Only one of the two crystallographic dimers is shown (40). *C*, NMR structure of dimeric dengue capsid protein (1R6R) (43). *D*, PRRSV nucleocapsid protein (1P65) crystallographic dimer (30). *E*, structure of the C-terminal dimerization domain, residues 270–370, of SARS-CoV nucleocapsid protein (2GIB). *F*, NMR structure of the N-terminal domain of SARS-CoV nucleocapsid protein (1SSK), residues 49–178 (16). The figures were prepared with PyMOL.

tains 370 residues and folds into an S-shaped molecule consisting of 16 helices and two short β-strands (34). In the crystal lattice, four subunits of BDV N interact extensively to form a homotetramer, suggesting that the functional unit of BDV N is a tetramer (Fig. 4*A*). Both Semliki Forest virus (genus *Alphavirus*, family *Togaviridae*) and dengue virus (genus *Flavivirus*, family *Flaviviridae*) are small single-stranded RNA viruses with icosahedral symmetry. It has been suggested that they have a common ancestor because the structures of one of their surface glycoproteins are remarkably similar (35–37). Inside the lipid bilayer, however, the nucleocapsid cores of these two viruses are quite different. An alphavirus core consists of 240 copies of the capsid protein arranged in a T = 4 icosahedral lattice (38). Each capsid protein has ~270 residues, of which the N-terminal ~120 residues are flexible. The remaining 150 residues adopt a chymotrypsin-like fold and form homodimers in the crystal (39, 40) (Fig. 4*B*). In contrast, cryoelectron microscopy reconstructions of flaviviruses did not show a defined structure of the capsid core (41, 42), suggesting that the core either is disordered or has a different symmetry from the surface icosahedral lattice. In addition, the exact copy number of the capsid protein in the core is unclear. The C-terminal 100 residues of flavivirus capsid proteins fold into a four-helical structure with no similarity to that of an alphavirus capsid protein (43, 44) (Fig. 4*C*).

It is striking that, within the group of enveloped RNA viruses, no two nucleocapsid proteins are known to be structurally similar except those of *Arteriviridae* and *Coronaviridae* (Fig. 4). The structural similarity between the N proteins of SARS-CoV and PRRSV provides valuable information for understanding the evolutionary links between corona- and arteriviruses. Although these viruses were grouped in the same order, *Nidovirales*, their structural proteins were previously thought to be unrelated due to the marked difference in their sizes and lack of sequence similarity (15). In this study, a common fold of the dimerization domains of *Nidovirales* N proteins is observed, suggesting a possible common origin of these two proteins. The amino acid sequence and gene size diversity may have resulted from extensive mutation and RNA recombination during evolution. It is also important to note that the assembled cores of corona- and arteriviruses have completely different structures. PRRSV N protein interacts with the RNA genome to form a spherical, possibly icosahedral core (45), whereas

coronaviruses are known to have helical nucleocapsids (46). It was previously suggested that the ancestral virus had an icosahedral nucleocapsid core, and the larger N protein of coronavirus, freed from its icosahedral package constraints, allowed the coronavirus genome to become larger during evolution (47).

Addendum—While this work was in progress, the structure of the nucleocapsid protein of another coronavirus, the avian infectious bronchitis virus (IBV), was determined by Drs. H. Jayaram and B. V. V. Prasad at Baylor College of Medicine (unpublished data). The coordinates of the IBV N protein were kindly provided to us by them for phasing via molecular replacement. However, they were unnecessary since the experimental multiwavelength anomalous dispersion phasing was sufficient to build the initial model.

REFERENCES

1. Stadler, K., Masignani, V., Eickmann, M., Becker, S., Abrignani, S., Klenk, H. D., and Rappuoli, R. (2003) *Nat. Rev. Microbiol.* **1,** 209–218
2. Lai, M. M., and Cavanagh, D. (1997) *Adv. Virus. Res.* **48,** 1–100
3. Lai, M. M. C., and Holmes, K. V. (2001) in *Fields Virology* (Fields, B. N., Knipe, D. M., and Howley, P. M., eds) Vol. 1, 4th Ed., pp. 1163–1203, Lippincott-Raven Publishers, Philadelphia and New York
4. Sturman, L. S., Holmes, K. V., and Behnke, J. (1980) *J. Virol.* **33,** 449–462
5. Babcock, G. J., Esshaki, D. J., Thomas, W. D., Jr., and Ambrosino, D. M. (2004) *J. Virol.* **78,** 4552–4560
6. Kubo, H., Yamada, Y. K., and Taguchi, F. (1994) *J. Virol.* **68,** 5403–5410
7. Bos, E. C., Heijnen, L., and Spaan, W. J. (1995) *Adv. Exp. Med. Biol.* **380,** 283–286
8. Krueger, D. K., Kelly, S. M., Lewicki, D. N., Ruffolo, R., and Gallagher, T. M. (2001) *J. Virol.* **75,** 2792–2802
9. Bosch, B. J., van der Zee, R., de Haan, C. A., and Rottier, P. J. (2003) *J. Virol.* **77,** 8801–8811
10. Vennema, H., Godeke, G. J., Rossen, J. W., Voorhout, W. F., Horzinek, M. C., Opstelten, D. J., and Rottier, P. J. (1996) *EMBO J.* **15,** 2020–2028
11. He, R., Leeson, A., Andonov, A., Li, Y., Bastien, N., Cao, J., Osiowy, C., Dobie, F., Cutts, T., Ballantine, M., and Li, X. (2003) *Biochem. Biophys. Res. Commun.* **311,** 870–876
12. Tahara, S. M., Dietlin, T. A., Nelson, G. W., Stohlman, S. A., and Manno, D. J. (1998) *Adv. Exp. Med. Biol.* **440,** 313–318

13. Snijder, E. J., and Meulenberg, J. J. M. (2001) in *Fields Virology* (Fields, B. N., Knipe, D. M., and Howley, P. M., eds) Vol. 1, 4th Ed., pp. 1205–1220, Lippincott-Raven Publishers, Philadelphia and New York
14. Snijder, E. J., and Meulenberg, J. J. (1998) *J. Gen. Virol.* **79,** 961–979
15. Cavanagh, D. (1997) *Arch. Virol.* **142,** 629–633
16. 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) *Biochemistry* **43,** 6059–6063
17. Fan, H., Ooi, A., Tan, Y. W., Wang, S., Fang, S., Liu, D. X., and Lescar, J. (2005) *Structure* (*Camb.*) **13,** 1859–1868
18. Masters, P. S. (1992) *Arch. Virol.* **125,** 141–160
19. Nelson, G. W., and Stohlman, S. A. (1993) *J. Gen. Virol.* **74,** 1975–1979
20. Peng, D., Koetzner, C. A., McMahon, T., Zhu, Y., and Masters, P. S. (1995) *J. Virol.* **69,** 5475–5484
21. Nelson, G. W., Stohlman, S. A., and Tahara, S. M. (2000) *J. Gen. Virol.* **81,** 181–188
22. Chang, C. K., Sue, S. C., Yu, T. H., Hsieh, C. M., Tsai, C. K., Chiang, Y. C., Lee, S. J., Hsiao, H. H., Wu, W. J., Chang, C. F., and Huang, T. H. (2005) *FEBS Lett.* **579,** 5663–5668
23. Yu, I. M., Gustafson, C. L., Diao, J., Burgner, J. W., II, Li, Z., Zhang, J., and Chen, J. (2005) *J. Biol. Chem.* **280,** 23280–23286
24. Otwinowski, Z., Minor, W. (1997) *Methods Enzymol.* **276,** 307–326
25. Terwilliger, T. C. (2003) *Methods Enzymol.* **374,** 22–37
26. 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., Read, R. J., Rice, L. M., Simonson, T., and Warren, G. L. (1998) *Acta Crystallogr. Sect. D Biol. Crystallogr.* **54,** 905–921
27. Collaborative Computational Project, N. (1994) *Acta Crystallogr. Sect. D Biol. Crystallogr.* 760–763
28. Jones, T. A., Zou, J. Y., Cowan, S. W., and Kjeldgaard. (1991) *Acta Crystallogr. Sect. A* **47,** 110–119
29. Laskowski, R. A., Macarthur, M. W., Moss, D. S., and Thornton, J. M. (1993) *J. Appl. Crystallogr.* **26,** 283–291
30. Doan, D. N., and Dokland, T. (2003) *Structure* (*Camb.*) **11,** 1445–1451
31. He, R., Dobie, F., Ballantine, M., Leeson, A., Li, Y., Bastien, N., Cutts, T., Andonov, A., Cao, J., Booth, T. F., Plummer, F. A., Tyler, S., Baker, L., and Li, X. (2004) *Biochem. Biophys. Res. Commun.* **316,** 476–483
32. Hurst, K. R., Kuo, L., Koetzner, C. A., Ye, R., Hsue, B., and Masters, P. S. (2005) *J. Virol.* **79,** 13285–13297
33. Goff, S. P. (2001) in *Retroviridae: the Viruses and Their Replication* (Fields, B. N., Knipe, D. M., and Howley, P. M., eds) Vol. 2, 4th Ed., pp. 1871–1940, Lippincott Williams and Wilkins, Philadelphia and New York
34. Rudolph, M. G., Kraus, I., Dickmanns, A., Eickmann, M., Garten, W., and Ficner, R. (2003) *Structure* (*Camb.*) **11,** 1219–1226
35. Pletnev, S. V., Zhang, W., Mukhopadhyay, S., Fisher, B. R., Hernandez, R., Brown, D. T., Baker, T. S., Rossmann, M. G., and Kuhn, R. J. (2001) *Cell* **105,** 127–136
36. Rey, F. A., Heinz, F. X., Mandl, C., Kunz, C., and Harrison, S. C. (1995) *Nature* **375,** 291–298
37. Lescar, J., Roussel, A., Wien, M. W., Navaza, J., Fuller, S. D., Wengler, G., and Rey, F. A. (2001) *Cell* **105,** 137–148
38. Zhang, W., Mukhopadhyay, S., Pletnev, S. V., Baker, T. S., Kuhn, R. J., and Rossmann, M. G. (2002) *J. Virol.* **76,** 11645–11658
39. Choi, H. K., Tong, L., Minor, W., Dumas, P., Boege, U., Rossmann, M. G., and Wengler, G. (1991) *Nature* **354,** 37–43
40. Choi, H. K., Lu, G., Lee, S., Wengler, G., and Rossmann, M. G. (1997) *Proteins* **27,** 345–359
41. Kuhn, R. J., Zhang, W., Rossmann, M. G., Pletnev, S. V., Corver, J., Lenches, E., Jones, C. T., Mukhopadhyay, S., Chipman, P. R., Strauss, E. G., Baker, T. S., and Strauss, J. H. (2002) *Cell* **108,** 717–725
42. Zhang, Y., Corver, J., Chipman, P. R., Zhang, W., Pletnev, S. V., Sedlak, D., Baker, T. S., Strauss, J. H., Kuhn, R. J., and Rossmann, M. G. (2003) *EMBO J.* **22,** 2604–2613
43. Ma, L., Jones, C. T., Groesch, T. D., Kuhn, R. J., and Post, C. B. (2004) *Proc. Natl. Acad. Sci. U. S. A.* **101,** 3414–3419
44. Dokland, T., Walsh, M., Mackenzie, J. M., Khromykh, A. A., Ee, K. H., and Wang, S. (2004) *Structure* (*Camb.*) **12,** 1157–1163
45. Brinton-Darnell, M., and Plagemann, P. G. (1975) *J. Virol.* **16,** 420–433
46. Macneughton, M. R., and Davies, H. A. (1978) *J. Gen. Virol.* **39,** 545–549
47. Godeny, E. K., Chen, L., Kumar, S. N., Methven, S. L., Koonin, E. V., and Brinton, M. A. (1993) *Virology* **194,** 585–596
48. Gouet, P., Robert, X., and Courcelle, E. (2003) *Nucleic Acids Res.* **31,** 3320–3323
49. Barton, G. J. (1993) *Protein Eng.* **6,** 37–40
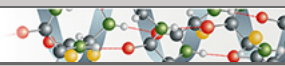
I-Mei Yu, Michael L. Oldham, Jingqiang Zhang and Jue Chen

Access the most updated version of this article at doi: 10.1074/jbc.M602107200

Find articles, minireviews, Reflections and Classics on similar topics on the JBC Affinity Sites.

Alerts:

- When this article is cited
- When a correction for this article is posted

Click here to choose from all of JBC's e-mail alerts

This article cites 48 references, 18 of which can be accessed free at
http://www.jbc.org/content/281/25/17134.full.html#ref-list-1