Structural Characterization of Human Coronavirus NL63 N Protein

Bożena Szelazek,^{a,e} Wojciech Kabala,^b Krzysztof Kus,^c Michał Zdzalik,^b
Aleksandra Twarda-Clapa,^{d,e} Przemysław Golik,^{b,e} Michał Burmistrz,^b
Dominik Florek,^b Benedykt Władyka,^{a,e} Krzysztof Pyrc,^{b,e} Grzegorz Dubin^{b,e}

Department of Analytical Biochemistry, Faculty of Biochemistry, Biophysics and Biotechnology, Jagiellonian University, Krakow, Poland^a; Department of Microbiology, Faculty of Biochemistry, Biophysics and Biotechnology, Jagiellonian University, Krakow, Poland^b; Instituto Gulbenkian de Ciencia, Oeiras, Portugal^c; Faculty of Chemistry, Jagiellonian University, Krakow, Poland^d; Malopolska Centre of Biotechnology, Jagiellonian University, Krakow, Poland^e

ABSTRACT Coronaviruses are responsible for upper and lower respiratory tract infections in humans. It is estimated that 1 to 10% of the population suffers annually from cold-like symptoms related to infection with human coronavirus NL63 (HCoV-NL63), an alphacoronavirus. The nucleocapsid (N) protein, the major structural component of the capsid, facilitates RNA packing, links the capsid to the envelope, and is also involved in multiple other processes, including viral replication and evasion of the immune system. Although the role of N protein in viral replication is relatively well described, no structural data are currently available regarding the N proteins of alphacoronaviruses. Moreover, our understanding of the mechanisms of RNA binding and nucleocapsid formation remains incomplete. In this study, we solved the crystal structures of the N- and C-terminal domains (NTD, residues 10 to 140, and CTD, residues 221 to 340, respectively) of the N protein of HCoV-NL63, both at a 1.5-Å resolution. Based on our structure of NTD solved here, we proposed and experimentally evaluated a model of RNA binding. The structure of the CTD reveals the mode of N protein dimerization. Overall, this study expands our understanding of the initial steps of N protein-nucleic acid interaction and may facilitate future efforts to control the associated infections.

IMPORTANCE Coronaviruses are responsible for the common cold and other respiratory tract infections in humans. According to multiple studies, 1 to 10% of the population is infected each year with HCoV-NL63. Viruses are relatively simple organisms composed of a few proteins and the nucleic acids that carry the information determining their composition. The nucleocapsid (N) protein studied in this work protects the nucleic acid from the environmental factors during virus transmission. This study investigated the structural arrangement of N protein, explaining the first steps of its interaction with nucleic acid at the initial stages of virus structure assembly. The results expand our understanding of coronavirus physiology and may facilitate future efforts to control the associated infections.

KEYWORDS CTD, N protein, NL63, NTD, coronavirus, nucleocapsid, structure

Coronaviruses are single-stranded (+) RNA viruses named after the corona of the sun, which they resemble in electron micrographs (1). The subfamily *Coronavirinae* contains four genera, but only alpha- and betacoronaviruses are known to infect humans. Coronaviruses cause a variety of diseases in animals, but in humans, they are associated only with respiratory tract infections of various severities. Low-pathogenic species (LPCs) are estimated to be responsible for up to 30% of all respiratory tract infections, causing severe disease mostly in infants, young children, and elderly and

Received 28 December 2016 Accepted 7 February 2017

Accepted manuscript posted online 22 March 2017

Citation Szelazek B, Kabala W, Kus K, Zdzalik M, Twarda-Clapa A, Golik P, Burmistrz M, Florek D, Władyka B, Pyrc K, Dubin G. 2017. Structural characterization of human coronavirus NL63 N protein. *J Virol* 91:e02503-16. <https://doi.org/10.1128/JVI.02503-16>.

Editor Stanley Perlman, University of Iowa

Copyright © 2017 American Society for Microbiology. All Rights Reserved.

Address correspondence to Grzegorz Dubin, grzegorz.dubin@uj.edu.pl.

immunocompromised individuals. Currently, no specific agents or vaccines are available to target coronaviruses.

The first human coronaviruses (HCoVs) to be identified, HCoV-229E and -OC43, were characterized in the 1960s, and for almost 40 years they were believed to be the sole representatives of this family in humans (2, 3). Studies prompted by the 2002/2003 emergence of a novel coronavirus responsible for atypical pneumonia with a high fatality rate (severe acute respiratory syndrome coronavirus [SARS-CoV]) (4–6) resulted in the identification of two further human LPCs (HCoV-NL63 and HCoV-HKU1) (7, 8). Most recently, Middle East respiratory syndrome coronavirus (MERS-CoV), with a mortality rate exceeding 30%, emerged in the Middle East (9, 10).

Although the prevalence of high pathogenic species (HPCs) is limited (SARS-CoV disappeared from the human population shortly after its emergence more than 10 years ago, and MERS-CoV persists at very low levels in the Arabian Peninsula), LPCs remain relevant human pathogens. Of these, HCoV-OC43 and HCoV-NL63 are most frequently diagnosed (11).

HCoV-NL63 has existed in human populations for at least the last 30 years, but molecular dating suggests that it diverged with HCoV-229E from the most recent common ancestor roughly a thousand years ago (12, 13). Infections with HCoV-NL63 are most often characterized by self-limiting upper and lower respiratory tract disease, with symptoms like cough, rhinorrhea, and fever of various severities (14). However, sporadic fatal cases have also been reported (15, 16). Annually, 1 to 10% of the population suffers from upper or lower respiratory tract infection with HCoV-NL63, with the frequency increasing in winter and spring. Moreover, HCoV-NL63 has been revealed as the most important etiologic factor for croup in children, dethroning the parainfluenza viruses, which were previously considered to be the major causative agents of this illness (12, 17–19).

Because coronaviruses enter the cell by a receptor-mediated route, their species and tissue specificity is determined by the repertoire of host cell surface receptors. Although most alphacoronaviruses use aminopeptidase N (CD13) to enter the cell, HCoV-NL63 employs angiotensin-converting enzyme 2 (ACE-2) for this purpose (20). The same receptor is also utilized by SARS-CoV, despite the significant differences in the epidemiologic characteristics of the two viruses (21). HCoV-NL63 also requires heparin sulfate as a cell attachment receptor (22).

The coronavirus particle is composed of a core formed by the nucleocapsid (N) protein bound to genomic RNA. The N protein interacts with transmembrane M (membrane) protein, which stabilizes the lipid bilayer envelope encasing the core. The envelope is decorated with viral structural proteins, further stabilizing the particle and participating in virus entry and egress (23). Aside from its structural role, several other functions have been assigned to the N protein, including roles in viral mRNA transcription (24) and replication, cytoskeleton organization, induction of apoptosis, and avoidance of detection by pattern recognition molecules (25–28).

Coronaviral N proteins contain two structured domains: the N-terminal domain (NTD; ~15 kDa), which is responsible for RNA binding, and the C-terminal domain (CTD; ~14 kDa), which mediates oligomerization and RNA binding (29, 30). These domains are connected by a relatively long linker that is likely to be unstructured, and additional unstructured regions are located at the N and C termini. Together, the structured domains and unstructured regions of N protein participate in nucleic acid binding and subsequent steps of nucleocapsid formation (31), but our understanding of the mechanistic and structural bases of these processes remains limited. Due to the relative flexibility of the full-length N protein, which hinders its structural characterization, to date only structures of NTDs and CTDs derived from beta- and gammacoronaviruses have been reported, and no structural information on alphacoronaviral N protein is available. Models of CTD-mediated N protein oligomerization and assembly of N protein into higher-order structures (31) have been proposed, but, at the same time, the structural basis of the interaction of the NTD with RNA remains poorly defined. In an attempt to better understand certain aspects of the interaction between nucleic acid

A (NTD)

B (CTD)

FIG 1 Structure-guided alignment of the NTD (A) and CTD (B) of N protein. Secondary structures are indicated above the alignment. The regions of highest sequence conservation are highlighted. The residues which involvement in nucleic acid binding was tested by mutagenesis in this study are marked with asterisks. TT, turn.

and N protein, we solved the crystal structure of the NTD of alphacoronavirus HCoV-NL63 and, on the basis of this structure, proposed and validated a model of NTD-RNA interaction. Moreover, the crystal structure of the CTD is presented, casting additional light on N protein dimerization during nucleocapsid assembly.

RESULTS AND DISCUSSION

Characterization of structured domains of NL63 N protein. Recombinant N protein of HCoV-NL63 was obtained in an *Escherichia coli* expression system with a yield of 6 mg per liter of starting culture and then purified to homogeneity. Multiple crystallization trials of the full-length protein were unsuccessful. Limited proteolysis was employed to identify stable domains suitable for crystallization (if any), but the full-length protein was either entirely stable or rapidly degraded into short peptides, depending on the protease used (data not shown), preventing a conclusive definition of the domain boundaries. Therefore, NTD- and CTD-expressing constructs were designed based on sequence analysis and the available literature (Fig. 1). Both domains

TABLE 1 Data collection and refinement statistics

Parameter	Value for domain ^a	
	NTD (PDB code 5N4K)	CTD (PDB code 5EPW)
Data collection		
Space group	P3 ₂	P6 ₁
Cell dimensions		
<i>a</i> , <i>b</i> , <i>c</i> (Å)	48.09, 48.09, 101.22	68.29, 68.29, 92.49
α , β , γ (°)	90, 90, 120	90, 90, 120
Resolution (Å)	1.49	1.50
<i>R</i> _{merge}	0.034 (0.095)	0.085 (0.841)
<i>I</i> / σ <i>I</i>	16.1 (7.4)	15.3 (3.0)
Completeness (%)	99.1 (99.5)	100 (100)
Redundancy	2.9 (2.9)	10.2 (9.7)
Refinement		
Resolution (Å)	1.49	1.5
No. of reflections	42,693	39,119
<i>R</i> _{work} / <i>R</i> _{free}	0.151/0.197	0.169/0.198
No. of atoms	2,645	1,955
Protein	2,205	1,736
Ligand/ion	75	0
Water	365	219
Average B-factors		
Protein	21.06	21.75
Ligand/ion	34.59	20.91
Water	30.65	28.46
RMSD		
Bond length (Å)	0.024	0.027
Bond angle (°)	2.53	2.40
Ramachandran favored (%)	97.77	98
Ramachandran allowed (%)	1.79	2.3

^aValues in parentheses are for the highest-resolution shell.

were obtained with high efficiency (~30 mg/liter of culture) and purified to homogeneity. Size exclusion chromatography revealed that the NTD was present in a monomeric form in solution, whereas the CTD had eluted as a dimer (and also as high-molecular-weight aggregates that were unstable at higher concentrations), consistent with previously published data (32). The NTD crystallized readily, but we failed to obtain CTD crystals from the initial construct. Limited proteolysis using V8 protease revealed that our CTD construct could be further truncated into a lower-molecular-weight form. The cleavage site was identified by microsequencing within the linker connecting the purification tag and CTD. The processed, tag-free form crystallized readily under multiple different conditions.

Crystal structure of NTD of HCoV-NL63 nucleocapsid protein. The crystal structure of the NTD, solved and refined at a 1.5-Å resolution, had an *R*_{factor}/*R*_{free} of 15.1%/19.7% and satisfactory stereochemical parameters (Table 1). The crystal belongs to space group P3₂. The asymmetric unit contains two NTD molecules encompassing residues 9 to 140 and 10 to 140 for molecules A and B, respectively. Each molecule is well defined by electron density except for a single partially unstructured, solvent-exposed loop (Arg65-Gln68). The monomers are virtually identical and superpose with a root mean square deviation (RMSD) of 0.2 Å (C α atoms). Each monomer has a palm-like shape with a single protruding finger (Fig. 2A). In this representation, the finger is composed of a long β -hairpin with a partially unstructured loop connecting the strands (β 2- β 3) and relatively higher B-factors characterizing the structured part distal from the core of the molecule. The other side of the hairpin is embedded in the center of the globular part of the molecule. The approximate dimensions of the single monomer are 50 by 30 by 29 Å, with the longest dimension corresponding to the protruding finger-like β 2- β 3 hairpin.

Despite the variation at the level of primary sequence, the overall fold of alphacoronavirus NTD is similar to those of previously determined NTD structures from beta-

FIG 2 Crystal structure of the NTD of HCoV-NL63 nucleocapsid protein. (A) Overall structure of the NTD monomer. β -Strands are depicted in pink, helices in red, and loops in blue. (B) Topology of the NTD. Color coding is as in panel A. (C) NTD dimer in the asymmetric unit. Sulfate ions from the crystallization buffer are shown in red. (D) Overlay of the NL63 NTD structure determined in this study (blue) and structures of NTDs of other coronaviruses, determined previously: infectious bronchitis virus (PDB code [2BXX](#); salmon), murine hepatitis virus (PDB code [3HD4](#); red), and severe acute respiratory syndrome coronavirus (PDB code [2OFZ](#); pink). The additional helical region characteristic of the NL63 NTD is highlighted in green.

and gammacoronaviruses (Table 2). The fold consists of a five-stranded, antiparallel β -sheet with the topology β_4 - β_2 - β_3 - β_1 - β_5 (Fig. 2B), partially encapsulated within extended, intertwining loops and turns connecting the β -strands. Comparison of the previously determined structures of beta- and gammacoronavirus NTDs with the structure of alphacoronavirus NTD determined in this study (Fig. 2D) revealed that the largest differences in the main-chain arrangement are in the N-terminal part immediately C terminal to the β_1 strand (Ser29-Asn40) and in an adjacent fragment just N terminal of the β_5 strand (Ser128-Glu134). In the former fragment, a short α -helix

TABLE 2 Structural similarity between NTDs and CTDs of coronaviral N proteins determined previously and corresponding domains of HCoV-NL63 N protein determined in this study, shown as RMSD of equivalent C_α atoms

Virus	PDB code		Family	RMSD, Å (no. of aligned C_α atoms)	
	NTD	CTD		NTD	CTD
SARS	2OFZ	2CJR	β	1.9 (105)	2.2 (196)
MHV	3HD4		β	1.9 (109)	
OC43	4J3K		β	1.2 (101)	
IBV	2BXX	2CA1	γ	1.8 (109)	2.7 (187)

encompassing residues Ala32 and Ile37 is characteristic of the NL63 NTD but is not present in any of the structures of beta- or gammacoronaviruses determined to date (Fig. 2D). The second region of significant structural variation is located within the protruding finger motif (Gln52-Leu81). In HCoV-NL63, this region assumes a fully β -sheet conformation similar to that of infectious bronchitis virus (IBV) but distinct from those of SARS-CoV and murine hepatitis virus (MHV), in which the β -sheet is interrupted in its center. In HCoV-NL63 N protein, the rigid β -structure within this region restricts its plasticity in comparison with those of the interrupted β -structures in SARS-CoV and MHV. The significance of the aforementioned differences for nucleic acid binding remains unknown.

The two NL63 NTD monomers in the asymmetric unit are arranged in an entwining dimer characterized by a pseudo-2-fold rotational symmetry (Fig. 2C). Although the buried accessible surface area is relatively large ($\sim 2,000 \text{ \AA}^2$), the interaction surface within the globular part of each monomer is slightly concave; therefore, direct contacts between monomers are relatively few in number and located only at the rim of the interaction surface. These include hydrogen bonds between $_A\text{Asp30}$ and $_B\text{Lys15}$ and of $_A/_B\text{Asp88}$ and $_B/_A\text{Gln93}$ and a salt bridge between $_A/_B\text{Glu60}$ and $_B/_A\text{Arg118}$. The central part of the dimer is composed of a large, solvent-filled cavity, and tightly coordinated water molecules mediate the majority of intermolecular interactions. The protruding finger-like β -hairpins of each monomer are located around $13 \text{ \AA}$ from each other and do not directly interact within the dimer. The only indirect contacts within this part of the dimer are provided through solvent-derived sulfate ions. Overall, the interactions cementing the crystallographic dimer are weak, and the dimer is not stable in solution, as demonstrated by size exclusion chromatography in which recombinant NTD was eluted as a monomer (32) (data not shown). Therefore, it is likely that the pseudodimerization is an effect of crystal packing alone. This conclusion is corroborated by crystal structures of NTDs derived from other coronaviruses that do not exhibit comparable dimerization (30, 33, 34).

The electrostatic potential surface of the NL63 NTD contains a large, positively charged groove where the N terminus and $\beta 2$ - $\beta 3$ finger protrude from the globular part of the molecule. An extended positively charged patch is characteristic of several NTDs characterized to date and has been implicated in RNA binding.

Each monomer within the structure coordinates several sulfate ions derived from the crystallization buffer, three of which are located within the groove described above and are of special interest concerning the presumed RNA binding (see below). The first of these sulfate ions, which is located in the middle part of a finger-like $\beta 2$ - $\beta 3$ hairpin, is coordinated by the side chains of Arg61 and Arg63. The second sulfate ion, which is located at the site where the hairpin protrudes from the globular part of the molecule, is coordinated by the side chains of Gln59, Arg61, and Lys75. The third sulfate ion forms a salt bridge with the side chain of Arg116 and is further stabilized within a network of surrounding water molecules.

Model of NTD-RNA binding. Previous work showed that the NL63-NTD is the primary RNA-binding moiety within the nucleocapsid protein (32). To elucidate the molecular basis of the interaction, we modeled binding of a short RNA to the NL63 NTD. We used some indirect information to guide our modeling. Specifically, the overall fold of the NTD distantly resembles that of the U1A RNA-binding protein, whose mode of interaction with nucleic acid has been determined experimentally. Nevertheless, detailed analysis excluded the possibility that the NTD binds RNA in a manner similar to that of U1A. Lin et al. crystallographically characterized a single nucleotide-binding site at the surface of the HCoV-OC43 NTD (35). However, the NL63 NTD structure we solved in this study was not compatible with this nucleotide-binding mode. In the OC43 structure, the binding involves formation of a hydrogen bond between the phosphate group and Gly68 backbone amide, whereas in NL63, proline is present at the equivalent position, precluding hydrogen bond formation. Moreover, the proline side chain sterically occludes the site equivalent to the site of the phosphate group in the HCoV-OC43

FIG 3 Model of RNA interaction with the NL63 NTD. The protein model is shown in gray. Interacting residues are highlighted blue (His77 is located at the opposite side of the molecule and not visible in this view). The RNA backbone is shown in red, and each base is shown in yellow. Hydrogen bonds are depicted as black dotted lines. The inset shows the overall orientation of the RNA-binding site depicted in the main panel.

NTD structure. Therefore, the details of nucleic acid interaction must differ between the NL63 NTD (alphacoronavirus) and HCoV-OC43 NTD (betacoronavirus). Differences between coronaviruses in the mode of interaction of NTDs with nucleic acid were also suggested by a previous analysis (31). No other experimental structures of proteins homologous to NTDs from coronaviruses or structures of the NTD/nucleic acid complexes have been determined to date. However, the nucleic acid interaction surface within SARS-CoV NTD was roughly delineated experimentally by monitoring nuclear magnetic resonance (NMR) chemical shift perturbation upon RNA titration (36). Abundant mutagenesis data from homologous NTDs corroborate the extent of the binding surface and highlight particular residues involved in the interaction (35–39). The crystal structure of the NL63 NTD revealed a network of sulfate ions, mostly coordinated by positively charged residues. We based our modeling on the fact that sulfate ions chemically resemble phosphate moieties present within the structure of RNA and are therefore likely to be located at the same binding sites. Further, the interactions of the sulfate ions were used to guide those of the ribose hydroxyls within the nucleic acid. In our modeling, we considered only sulfate ions in the region of the NL63 NTD previously shown to be responsible for nucleic acid binding in homologous domains. Analysis of distance constraints revealed that a four-nucleotide polymer is long enough to extend between distal sulfate ions. We used a five-nucleotide RNA to allow additional flexibility while at the same time keeping the required computation time within reasonable limits. A library of low-energy RNA structures was generated and docked to the structure of the NTD using weights strongly favoring overlap between phosphates and ribose moieties within the RNA and sulfate ions within the NTD structure. The best solutions were energy minimized and inspected for clashes and geometry problems. The model characterized by the lowest potential energy within which the RNA molecule could be easily extended from both ends was chosen as the optimal representative of the NL63 NTD-RNA interaction (Fig. 3).

Our model predicts that at least five consecutive nucleotides within RNA polymer interact with the NTD. Two successive phosphate moieties at the 3' end form salt bridges with positively charged residues (Arg61/Arg63 and Gln59/Arg61/Lys75, respectively) at the protruding finger-like structure within the NTD. The third nucleotide does not form any interactions with the NTD but instead engages in intermolecular interactions within RNA. The fourth nucleotide is located roughly at the binding pocket previously characterized in OC43 (35). This is especially interesting because we made no

TABLE 3 Affinity of the NTD and its mutants for nucleic acid^a

NTD variants	FP		MST	
	K_d (μ M)	SD	K_d (μ M)	SD
WT	2.3	0.2	6.0	0.9
R63A	6.5	0.4	11.4	1.9
K75A	3.3	2.7	13.8	2.2
K121A	5.6	0.8	17.6	2.7
R61A	7.0	0.6	24.0	4.0
R116A	6.7	0.8	27.5	4.7
H77A	5.7	1.0	40.2	5.4
Q59A	7.3	0.1	43.1	5.5
Q59A/H77A	9.7	1.1	47.3	8.9
R61A/K75A	11.8	2.8	55.4	8.0
H77A/R116A	11.8	1.9	72.4	11.7

^aThe affinities were determined by fluorescence polarization (FP) and microscale thermophoresis (MST). WT, wild type.

assumptions concerning the positions of the bases; therefore, their optimization was guided solely by the parameters of the force field. The ribose moiety of the nucleotide forms a hydrogen bond with the side chain of His77, while its phosphate moiety forms a salt bridge with the side chain of Arg116. This suggests that although the exact site differs significantly between the NL63 NTD and OC43 NTD, as described above, the overall capability of this region to interact with nucleic acid is conserved. The fifth nucleotide forms two hydrogen bonds with surface residues (Lys121 and Glu139).

In keeping with the model assumptions, the interaction is guided primarily by electrostatic forces and mainly involves the backbone phosphates and sugars. Therefore, according to our model, sequence specificity of NTD-RNA interaction in the modeled region, if any, would be defined primarily by intermolecular interactions within the RNA (secondary structures). This conclusion is consistent with experimental results demonstrating that the NL63 NTD has little or no specificity for particular sequences, or even the type of nucleic acid (i.e., RNA versus DNA) (32). Direct base interactions with the NTD would have a less prominent effect on binding, although our model accounts for a base-binding pocket at the finger/palm groove. The involvement of aromatic residues in the palm region was previously suggested in RNA binding by the SARS-CoV (31) and MHV (33) NTDs, but these features were not analyzed in this study. Our model does not extend to the palm of the NL63 NTD, as no sulfate ions were present in that region to guide modeling. However, we cannot exclude the possibility of a more extended interaction involving the palm region.

Experimental evaluation of model assumptions. Next, we tested the mode of interaction of the NTD with nucleic acid predicted by our model using site-directed mutagenesis. We evaluated the influence of residues presumably involved in interaction with the phosphate moieties within the 3' portion of the nucleic acid (Gln59, Arg61, Arg63, and Lys75) and the residue likely involved in similar interactions with the 5' end (Arg116). Additionally, we tested Lys121, predicted to be involved in the interaction with the 5' end residue, as well as His77, which is likely to stabilize binding in the middle of the ligand. Thus, all the most prominent interactions predicted by our model were covered by mutagenesis. In all cases, the replacement of a charged side chain with alanine resulted in reduced affinity of the mutant toward nucleic acid (Table 3). Replacement of H77 with alanine also resulted in a comparable drop in affinity. These results support the relevance of our model in predicting the role of specific residues in determining the affinity of the NTD for nucleic acids.

If a large number of interactions influence the affinity of the NTD, individual mutations should have relatively small effects. To further test the assumptions of our model, we also evaluated double mutants of the above residues. The affinities of the tested double mutants were lower than the affinities of the respective single mutants, further supporting the predictions of our model.

The results described above are consistent with prior mutagenesis data obtained for beta- and gammacoronavirus NTDs. In IBV, residues equivalent to positions 61, 63, and 77 in NL63 were experimentally demonstrated to participate in the nucleic acid interaction (37), and the residue equivalent to position 61 is involved in nucleic acid binding by OC43 (38). Further, residues at positions equivalent to positions 59 and 75 (39), and 77 of NL63 (33), are important for nucleic acid binding by MHV. Structural analysis of OC43 NTD complexes with nucleotides and mutagenesis data (35) revealed the importance of residues equivalent to positions 75, 77, and 116 of NL63. Collectively, the close agreement of mutagenesis data obtained for NTDs from different coronaviruses suggests that despite their marked differences in primary structure, the overall modes of alpha-, beta-, and gammacoronavirus NTD binding to nucleic acid are comparable.

Although our model depicts how nucleic acid interacts with the NTD, the overall interaction of N protein with nucleic acid is likely to be much more complex. Along these lines, previous work showed that intrinsically disordered regions within N protein and its CTD significantly contribute to affinity. We have not systematically studied these phenomena, but we did observe that N protein copurifies with nucleic acid that is hard to remove, in contrast to the NTD, from which nucleic acid easily dissociates during purification (data not shown). Moreover, the disordered and ordered regions are involved in self-assembly of the N protein/RNA complex into higher-order quaternary structures, which cannot be accounted for by our characterization of isolated domains. Although the entire interaction remains poorly understood, several explanatory models have been proposed. For a recent review, we direct the reader to the comprehensive work of Chang and collaborators (31).

Crystal structure of the CTD of HCoV-NL63 nucleocapsid protein. The crystal structure of the NL63 CTD was solved at a resolution of 1.5 Å in the $P6_1$ space group. Two CTD molecules are present in the asymmetric unit. The final model contains residues 228 to 331 (molecule A) and 228 to 337 (molecule B) and was refined to $R_{\text{factor}}/R_{\text{free}}$ of 16.9%/19.8% (Table 1). The majority of the protein is well defined by electron density and characterized by uniform, low B-factors. Each monomer contains five α -helices, three 3_{10} helices, and two β -strands and assumes a $\alpha_1\text{--}\alpha_4\beta_1\beta_2\alpha_5$ topology (Fig. 4B). The two C-shaped monomers in the asymmetric unit are almost identical and superpose with a C_α RMSD of 0.32 Å. Essentially, one arm in each monomer is constituted by the helical regions, whereas the other arm consists of a β -hairpin (Fig. 4A).

Each monomer in the asymmetric unit inserts its β -hairpin arm into the concavity within the second monomer, forming a tightly interwoven homodimer (Fig. 4C). This arrangement generates the pseudo-2-fold rotational symmetry that characterizes the dimer and creates a four-stranded, antiparallel β -sheet at its core. The dimer has a flat, roughly square shape with approximate dimensions of 38 by 38 by 25 Å. One face is formed by the β -sheet, whereas the opposite face is composed of α -helices. The C termini are located at the apices of the β -sheet face, and the N termini protrude from the centers of long edges. Apart from the 14 intermonomer hydrogen bonds within the β -sheet, the dimer is further stabilized by hydrophobic interactions between α 4 helices (primarily involving Phe288 and Phe289 within each monomer). Altogether, the interaction buries a surface area of $\sim 2,900$ Å², which accounts for 36% of the total surface of each separated monomer. The extended contact area and large number of mediating intermolecular interactions suggest that the dimeric arrangement of the CTD represents its true biological architecture. This is corroborated by our previous biochemical characterization showing that the CTD exists as a dimer in solution (32). Therefore, the basic organizational unit within the nucleocapsid is the N protein dimer, rather than the monomer, consistent with data obtained previously for different coronaviruses. Moreover, the CTD exhibits the tendency to aggregate in solution, which may reflect further steps of nucleocapsid assembly.

FIG 4 Crystal structure of the CTD of HCoV-NL63 nucleocapsid protein. (A) Overall structure of the CTD monomer. β -Strands are depicted in pink, helices in red, and loops in blue. (B) Topology of the CTD. Color coding is as in panel A. (C) CTD protein dimer as found in the asymmetric unit. (D) Overlay of the NL63 CTD dimer structure determined in this study (yellow) and structures of CTD dimers from other coronaviruses, determined previously: SARS-CoV (PDB code [2CJR](#); pink) and IBV (PDB code [2CA1](#); salmon).

Analysis of the molecular surface of the CTD revealed a characteristic groove encircling one side of the dimer and forming a belt of positively charged and polar residues that uniformly exposed only the atoms capable of donating hydrogen bonds. This patch could provide a site for nucleic acid interaction, consistent with previous studies suggesting that the CTD interacts with RNA (29). Nevertheless, in the case of the NL63 CTD, the presumed interaction is relatively weak, if present at all, since we could not detect nucleic acid/CTD binding in any of the assays we performed previously (32).

The structure of the CTD of the HCoV-NL63 nucleocapsid protein reveals, for the first time, the fold and structural details of the CTD in alphacoronaviruses. Despite differences among species in the primary sequences of CTDs, the overall fold of the NL63 CTD resembles those determined to date for SARS-CoV (betacoronavirus) and IBV (gammacoronavirus), the only structures of CTDs solved previously (Fig. 4D). The NL63 CTD dimer superposes with the CTD dimers of SARS-CoV and IBV with RMSD of 2.2 Å and 2.7 Å, respectively. However, detailed analysis of the intermonomer interactions

within the dimer revealed a single notable difference between the alpha- and gamma-coronaviruses and the betacoronaviruses: the intermolecular β -sheet is longer in the NL63 CTD and IBV CTD (40) (14 hydrogen bonds) than in the corresponding structure of the SARS-CoV CTD dimer (10 hydrogen bonds) (38). Certain other differences are present in solvent-exposed regions, but because those regions are characterized by relatively high B-factors, these differences may not be significant. Together, these observations confirm the high structural conservation of the CTD among coronaviruses of different families. As noted previously (41, 42), the overall fold of the CTD also resembles that of the nucleocapsid protein of arterivirus, a porcine reproductive and respiratory syndrome virus, suggesting analogies in nucleocapsid formation.

In summary, our structures of functional domains of NL63 nucleocapsid protein and the proposed model of RNA binding provide structural insight into the initial steps of N protein interaction with nucleic acid, as well as N protein dimerization, within the alphacoronavirus family. Future studies should seek to determine the structure of full-length nucleocapsid protein in complex with nucleic acid, and these efforts are under way in our laboratory.

MATERIALS AND METHODS

Cloning and expression. The constructs used for expression of the NTD (residues 10 to 140) and CTD (residues 221 to 340) fragments of HCoV-NL63 N protein (Uniprot [Q6Q1R8](#)) were designed based on sequence comparisons and available literature. The coding fragments were amplified by PCR from viral cDNA and cloned into vector pETDuet-1 at the BamHI/HindIII restriction sites, yielding constructs encoding six histidine-tagged fusion proteins. Mutants were generated by site-directed mutagenesis. All constructs were verified by sequencing.

Proteins were expressed in *Escherichia coli* BL21(DE3). Bacteria were cultured in lysogeny broth (LB) supplemented with ampicillin at 37°C with thorough aeration. When the culture optical density at 600 nm (OD_{600}) reached 0.5, protein production was induced by addition of 0.5 mM isopropyl- β -D-thiogalactopyranoside (IPTG); the temperature was lowered to 18°C, and the culture was continued overnight.

Protein purification. Identical purification protocols were used for the NTD and CTD constructs and all mutants. After protein expression, the cells were harvested by centrifugation. Pellets were resuspended in lysis buffer (20 mM Tris-HCl [pH 8.0] containing 0.25 M NaCl, 20 mM imidazole, and 20% glycerol) and disrupted by sonication. Debris was removed by centrifugation, and the protein of interest was recovered using immobilized-metal affinity chromatography (IMAC Sepharose; GE Healthcare). The proteins were further purified, and the buffer was exchanged to 5 mM Tris-HCl (pH 8.0) containing 20 mM NaCl by gel filtration using Superdex 75 (GE Healthcare).

Crystallization and structure determination. The purified proteins were concentrated up to 105 mg/ml (NTD) and 50 mg/ml (CTD), and crystallization screening was performed at 22°C using the sitting drop vapor-diffusion method. The CTD preparation was supplemented with V8 protease at a 1,000:1 molar ratio immediately before screening to facilitate limited proteolysis *in situ*. The best crystals were obtained from 0.1 M Tris (pH 8.5) containing 2 M ammonium sulfate (NTD) and 0.1 M PCTP buffer (pH 8.0; PCTP buffer contains sodium propionate, sodium cacodylate trihydrate, and Bis-Tris propane) containing 25% polyethylene glycol 1500 (PEG 1500) (CTD). NTD crystals were cryopreserved in 25% (vol/vol) glycerol and flash-cooled in liquid nitrogen. No additional cryoprotection was performed for CTD crystals. Diffraction data were collected at 100 K at SLS (Villigen, Switzerland) on beamline X10SA (NTD) and at DESY (Hamburg, Germany) on beamline P11 (CTD). Data were indexed and integrated using XDS software (43). Subsequent computational steps were performed using programs in the CCP4 software package (44). Data were scaled using SCALA (45). Molecular replacement was performed with Phaser software (46), using the infectious bronchitis virus NTD structure (Protein Data Bank [PDB] code [2BXX](#)) and SARS virus CTD structure (PDB code [2CJR](#)) as search models. The structures were refined in multiple rounds of manual model building and restrained refinement, which were performed using Coot (47) and Refmac 5.0 (48), respectively. Throughout the refinement, 5% of the reflections were used for cross-validation analysis (49), and the behavior of R_{free} was used to monitor the refinement strategy. Data collection and refinement statistics are summarized in Table 1.

Molecular modeling of NTD-RNA interaction. Rosetta software (50) was used to obtain a library of 8,250 energy-minimized RNA structures of randomly selected sequence. A custom Python script was used to align phosphate and/or hydroxyl groups within ribose moieties of the RNAs within the library and the positions of sulfate ions present within the crystal structure. The protein-RNA complexes characterized by best alignment (RMSD) were energy minimized using GROMACS (51) with an AMBER force field (52). Models were scored according to potential energy and inspected for clashes and geometry problems using PHENIX (53). The highest-scoring model with satisfactory geometry and in which the RNA could be further extended at both ends was chosen as the optimal representation of the NTD-RNA interaction.

Nucleic acid affinity assays. The affinity of nucleic acid for NTD and its mutants was tested by fluorescence polarization (FP) and microscale thermophoresis (MST) in 50 mM phosphate buffer (pH 6.0) at room temperature. For FP, 5 nM 6-carboxyfluorescein (FAM)-TCAACTAAAC was titrated with increasing

concentrations of the tested protein, and changes in FP were monitored. Affinity (K_d [dissociation constant]) was determined by nonlinear regression. An identical probe was used for MST. The probe (100 nM) was titrated with increasing concentrations of tested protein, and changes in mobility in a temperature gradient were monitored. To determine affinity, data were fitted using algorithms implemented in software provided by the manufacturer (NanoTemper Technologies GmbH, Germany).

Miscellaneous. Protein oligomerization status was determined by gel filtration on a precalibrated Superdex 75 gel filtration column (GE Healthcare).

NL63 N, NTD, and CTD constructs were probed for unstructured regions by limited proteolysis. Staphylococcal V8 protease, staphopains A and B, YaT protease, papain, subtilisin, chymotrypsin, trypsin, thrombin, tobacco etch virus (TEV) protease, papain, and PreScission were incubated for 60 min at room temperature with the tested protein at various molar ratios. Reactions were stopped by addition of loading buffer and boiling. Reaction products were analyzed by SDS-PAGE.

Structures were analyzed and the figures were prepared using Coot, ENDscript (54), DSSP (55), and PyMOL (<http://www.pymol.org>).

Accession number(s). The final models were deposited in PDB under accession numbers [5N4K](#) and [5EPW](#) for the NTD and CTD, respectively.

ACKNOWLEDGMENTS

This work was supported in part by grants UMO-2011/01/D/NZ1/01169 (to G.D.) and UMO-2012/07/E/NZ6/01712 (to K.P.) from the National Science Center, grant Lider/27/55/L-2/10/2011 (to K.P.) from the National Centre for Research and Development, and grants for the promotion of young scientists K/DSC/003266 (to B.S.) and K/DSC/001060 (to M.Z.) from FBBB UJ. G.D. was supported by grant UMO-2012/07/E/NZ1/01907 from the National Science Center. Parts of this research were carried out at the light source Petra III at the German Electron Synchrotron (DESY, Hamburg, Germany), at Swiss Light Source (Villingen, Switzerland), and at BESSY II at Helmholtz-Zentrum Berlin für Materialien und Energie (HZB; Berlin, Germany) and with equipment purchased thanks to the financial support of the European Union structural funds (grants POIG.02.01.00-12-064/08 and POIG.02.01.00-12-167/08). Faculty of Biochemistry, Biophysics and Biotechnology of the Jagiellonian University is a partner of the Leading National Research Center supported by the Ministry of Science and Higher Education of the Republic of Poland.

REFERENCES

1. Sturman LS, Holmes KV, Behnke J. 1980. Isolation of coronavirus envelope glycoproteins and interaction with the viral nucleocapsid. *J Virol* 33:449–462.
2. Hamre D, Procknow JJ. 1966. A new virus isolated from the human respiratory tract. *Proc Soc Exp Biol Med* 121:190–193. <https://doi.org/10.3181/00379727-121-30734>.
3. Tyrrell DA, Bynoe ML. 1965. Cultivation of a novel type of common-cold virus in organ cultures. *Br Med J* 1:1467–1470. <https://doi.org/10.1136/bmj.1.5448.1467>.
4. Drosten C, Gunther S, Preiser W, van der Werf S, Brodt HR, Becker S, Rabenau H, Panning M, Kolesnikova L, Fouchier RAM, Berger A, Burguiere AM, Cinatl J, Eickmann M, Esciou N, Grywna K, Kramme S, Manuguerra JC, Muller S, Rickerts V, Sturmer M, Vieth S, Klenk HD, Osterhaus ADME, Schmitz H, Doerr HW. 2003. Identification of a novel coronavirus in patients with severe acute respiratory syndrome. *N Engl J Med* 348:1967–1976. <https://doi.org/10.1056/NEJMoa030747>.
5. Ksiazek TG, Erdman D, Goldsmith CS, Zaki SR, Peret T, Emery S, Tong S, Urbani C, Comer JA, Lim W, Rollin PE, Dowell SF, Ling AE, Humphrey CD, Shieh WJ, Guarner J, Paddock CD, Rota P, Fields B, DeRisi J, Yang JY, Cox N, Hughes JM, LeDuc JW, Bellini WJ, Anderson LJ, SARS Working Group. 2003. A novel coronavirus associated with severe acute respiratory syndrome. *N Engl J Med* 348:1953–1966. <https://doi.org/10.1056/NEJMoa030781>.
6. Peiris JSM, Lai ST, Poon LLM, Guan Y, Yam LYC, Lim W, Nicholls J, Yee WKS, Yan WW, Cheung MT, Cheng VCC, Chan KH, Tsang DNC, Yung RWH, Ng TK, Yuen KY. 2003. Coronavirus as a possible cause of severe acute respiratory syndrome. *Lancet* 361:1319–1325. [https://doi.org/10.1016/S0140-6736\(03\)13077-2](https://doi.org/10.1016/S0140-6736(03)13077-2).
7. van der Hoek L, Pyrc K, Jebbink MF, Vermeulen-Oost W, Berkhout RJM, Wolthers KC, Wertheim-van Dillen PME, Kaandorp J, Spaargaren J, Berkhout B. 2004. Identification of a new human coronavirus. *Nat Med* 10:368–373. <https://doi.org/10.1038/nm1024>.
8. Woo PCY, Lau SKP, Chu C, Chan K, Tsoi H, Huang Y, Wong BHL, Poon RWS, Cai JJ, Luk W, Poon LLM, Wong SSY, Guan Y, Peiris JSM, Yuen K. 2005. Characterization and complete genome sequence of a novel coronavirus, coronavirus HKU1, from patients with pneumonia. *J Virol* 79:884–895. <https://doi.org/10.1128/JVI.79.2.884-895.2005>.
9. Zaki A, van Boheemen S, Bestebroer TM, Osterhaus AD, Fouchier RA. 2012. Isolation of a novel coronavirus from a man with pneumonia in Saudi Arabia. *N Engl J Med* 367:1814–1820. <https://doi.org/10.1056/NEJMoa1211721>.
10. Zumla A, Hui DS, Perlman S. 2015. Middle East respiratory syndrome. *Lancet* 386:995–1007. [https://doi.org/10.1016/S0140-6736\(15\)60454-8](https://doi.org/10.1016/S0140-6736(15)60454-8).
11. Dijkman R, Jebbink MF, Gaunt E, Rossen JWA, Templeton KE, Kuijpers TW, van der Hoek L. 2012. The dominance of human coronavirus OC43 and NL63 infections in infants. *J Clin Virol* 53:135–139. <https://doi.org/10.1016/j.jcv.2011.11.011>.
12. Fouchier RAM, Hartwig NG, Bestebroer TM, Niemeyer B, de Jong JC, Simon JH, Osterhaus ADME. 2004. A previously undescribed coronavirus associated with respiratory disease in humans. *Proc Natl Acad Sci U S A* 101:6212–6216. <https://doi.org/10.1073/pnas.0400762101>.
13. Pyrc K, Dijkman R, Deng L, Jebbink MF, Ross HA, Berkhout B, van der Hoek L. 2006. Mosaic structure of human coronavirus NL63, one thousand years of evolution. *J Mol Biol* 364:964–973. <https://doi.org/10.1016/j.jmb.2006.09.074>.
14. Cabeça TK, Granato C, Bellei N. 2013. Epidemiological and clinical features of human coronavirus infections among different subsets of patients. *Influenza Other Respir Viruses* 7:1040–1047. <https://doi.org/10.1111/irv.12101>.
15. Cabeça TK, Bellei N. 2012. Human coronavirus NL-63 infection in a Brazilian patient suspected of H1N1 2009 influenza infection: description of a fatal case. *J Clin Virol* 53:82–84. <https://doi.org/10.1016/j.jcv.2011.09.006>.
16. Oosterhof L, Christensen CB, Sengeløv H. 2010. Fatal lower respiratory

- tract disease with human corona virus NL63 in an adult haematopoietic cell transplant recipient. *Bone Marrow Transplant* 45:1115–1116. <https://doi.org/10.1038/bmt.2009.292>.
17. Esper F, Weibel C, Ferguson D, Landry ML, Kahn JS. 2005. Evidence of a novel human coronavirus that is associated with respiratory tract disease in infants and young children. *J Infect Dis* 191:492–498. <https://doi.org/10.1086/428138>.
 18. Lee J, Storch GA. 2014. Characterization of human coronavirus-OC43 and HCoV-NL63 infections among hospitalized children < 5 years of age. *Pediatr Infect Dis J* 33:2014. <https://doi.org/10.1097/INF.0000000000000292>.
 19. Van Der Hoek L, Sure K, Ihorst G, Stang A, Pyrc K, Jebbink MF, Petersen G, Forster J, Berkhout B, Überla K. 2005. Croup is associated with the novel coronavirus NL63. *PLoS Med* 2:0764–0770. <https://doi.org/10.1371/journal.pmed.0020240>.
 20. Hofmann H, Pyrc K, van der Hoek L, Geier M, Berkhout B, Pöhlmann S. 2005. Human coronavirus NL63 employs the severe acute respiratory syndrome coronavirus receptor for cellular entry. *Proc Natl Acad Sci U S A* 102:7988–7993. <https://doi.org/10.1073/pnas.0409465102>.
 21. Pyrc K, Berkhout B, van der Hoek L. 2007. Identification of new human coronaviruses. *Expert Rev Anti Infect Ther* 5:245–253. <https://doi.org/10.1586/14787210.5.2.245>.
 22. Milewska A, Zarebski M, Nowak P, Stozek K, Potempa J, Pyrc K. 2014. Human coronavirus NL63 utilizes heparan sulfate proteoglycans for attachment to target cells. *J Virol* 88:13221–13230. <https://doi.org/10.1128/JVI.02078-14>.
 23. Kuo L, Masters P. 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. <https://doi.org/10.1128/JVI.76.10.4987-4999.2002>.
 24. Zúñiga S, Cruz JLG, Sola I, Mateos-Gómez PA, Palacio L, Enjuanes L. 2010. Coronavirus nucleocapsid protein facilitates template switching and is required for efficient transcription. *J Virol* 84:2169–2175. <https://doi.org/10.1128/JVI.02011-09>.
 25. Zhou B, Liu J, Wang Q, Liu X, Li X, Li P, Ma Q, Cao C. 2008. The nucleocapsid protein of severe acute respiratory syndrome coronavirus inhibits cell cytokinesis and proliferation by interacting with translation elongation factor 1alpha. *J Virol* 82:6962–6971. <https://doi.org/10.1128/JVI.00133-08>.
 26. Surjit M, Liu B, Chow VTK, Lal SK. 2006. The nucleocapsid protein of severe acute respiratory syndrome-coronavirus inhibits the activity of cyclin-cyclin-dependent kinase complex and blocks S phase progression in mammalian cells. *J Biol Chem* 281:10669–10681. <https://doi.org/10.1074/jbc.M509233200>.
 27. Lu X, Pan J, Tao J, Guo D. 2011. SARS-CoV nucleocapsid protein antagonizes IFN- β response by targeting initial step of IFN- β induction pathway, and its C-terminal region is critical for the antagonism. *Virus Genes* 42:37–45. <https://doi.org/10.1007/s11262-010-0544-x>.
 28. Zhao G, Shi SQ, Yang Y, Peng JP. 2006. M and N proteins of SARS coronavirus induce apoptosis in HPF cells. *Cell Biol Toxicol* 22:313–322. <https://doi.org/10.1007/s10565-006-0077-1>.
 29. Chen CY, Chang CK, Chang YW, Sue SC, Bai HI, Riag L, Hsiao CD, Huang TH. 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. <https://doi.org/10.1016/j.jmb.2007.02.069>.
 30. Fan H, Ooi A, Tan YW, Wang S, Fang S, Liu DX, 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. <https://doi.org/10.1016/j.str.2005.08.021>.
 31. Chang CK, Hou MH, Chang CF, Hsiao CD, Huang TH. 2014. The SARS coronavirus nucleocapsid protein—forms and functions. *Antiviral Res* 103:39–50. <https://doi.org/10.1016/j.antiviral.2013.12.009>.
 32. Zuwała K, Golda A, Kabala W, Burmistrz M, Zdzalik M, Nowak P, Kedracka-Krok S, Zarebski M, Dobrucki J, Florek D, Zeglen S, Wojarski J, Potempa J, Dubin G, Pyrc K. 2015. The nucleocapsid protein of human coronavirus NL63. *PLoS One* 10:e0117833. <https://doi.org/10.1371/journal.pone.0117833>.
 33. Grosseohme NE, Li L, Keane SC, Liu P, Dann CE, Leibowitz JL, Giedroc DP. 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. <https://doi.org/10.1016/j.jmb.2009.09.040>.
 34. Saikatendu KS, Joseph JS, Subramanian V, Neuman BW, Buchmeier MJ, Stevens RC, 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. <https://doi.org/10.1128/JVI.02236-06>.
 35. Lin SY, Liu CL, Chang YM, Zhao J, Perlman S, Hou MH. 2014. Structural basis for the identification of the N-terminal domain of coronavirus nucleocapsid protein as an antiviral target. *J Med Chem* 57:2247–2257. <https://doi.org/10.1021/jm500089r>.
 36. Huang Q, Yu L, Petros AM, Gunasekera A, Liu Z, Xu N, Hajduk P, Mack J, Fesik SW, Olejniczak ET. 2004. Structure of the N-terminal RNA-binding domain of the SARS CoV nucleocapsid protein. *Biochemistry* 43: 6059–6063. <https://doi.org/10.1021/bi036155b>.
 37. Tan YW, Fang S, Fan H, Lescar J, Liu DX. 2006. Amino acid residues critical for RNA-binding in the N-terminal domain of the nucleocapsid protein are essential determinants for the infectivity of coronavirus in cultured cells. *Nucleic Acids Res* 34:4816–4825. <https://doi.org/10.1093/nar/gkl650>.
 38. Chen IJ, Yuann JMP, Chang YM, Lin SY, Zhao J, Perlman S, Shen YY, Huang TH, Hou MH. 2013. Crystal structure-based exploration of the important role of Arg106 in the RNA-binding domain of human coronavirus OC43 nucleocapsid protein. *Biochim Biophys Acta* 1834: 1054–1062. <https://doi.org/10.1016/j.bbapap.2013.03.003>.
 39. Keane SC, Lius P, Leibowitz JL, Giedroc DP. 2012. Functional transcriptional regulatory sequence (TRS) RNA binding and helix destabilizing determinants of murine hepatitis virus (MHV) nucleocapsid (N) protein. *J Biol Chem* 287:7063–7073. <https://doi.org/10.1074/jbc.M111.287763>.
 40. Jayaram H, Fan H, Bowman BR, Ooi A, Jayaram J, Collisson EW, Lescar J, Prasad BVV. 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. <https://doi.org/10.1128/JVI.00157-06>.
 41. Yu IM, Oldham ML, Zhang J, 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. <https://doi.org/10.1074/jbc.M602172200>.
 42. Takeda M, Chang C-K, Ikeya T, Güntert P, Chang Y-H, Hsu Y-I, Huang T-H, 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. <https://doi.org/10.1016/j.jmb.2007.11.093>.
 43. Kabsch W. 2010. Xds. *Acta Crystallogr D Biol Crystallogr* 66:125–132. <https://doi.org/10.1107/S0907444909047337>.
 44. Winn MD, Ballard CC, Cowtan KD, Dodson EJ, Emsley P, Evans PR, Keegan RM, Krissinel EB, Leslie AGW, McCoy A, McNicholas SJ, Murshudov GN, Pannu NS, Potterton EA, Powell HR, Read RJ, Vagin A, Wilson KS. 2011. Overview of the CCP4 suite and current developments. *Acta Crystallogr D Biol Crystallogr* 67(Part 4):235–242. <https://doi.org/10.1107/S0907444910045749>.
 45. Evans P. 2006. Scaling and assessment of data quality. *Acta Crystallogr D Biol Crystallogr* 62:72–82. <https://doi.org/10.1107/S0907444905036693>.
 46. McCoy AJ, Grosse-Kunstleve RW, Storoni LC, Read RJ. 2005. Likelihood-enhanced fast translation functions. *Acta Crystallogr D Biol Crystallogr* 61:458–464. <https://doi.org/10.1107/S0907444905001617>.
 47. Emsley P, Cowtan K. 2004. Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr* 60:2126–2132. <https://doi.org/10.1107/S0907444904019158>.
 48. Vagin AA, Steiner RA, Lebedev AA, Pottterton L, McNicholas S, Long F, Murshudov GN. 2004. REFMAC5 dictionary: organization of prior chemical knowledge and guidelines for its use. *Acta Crystallogr D Biol Crystallogr* 60:2184–2195. <https://doi.org/10.1107/S0907444904023510>.
 49. Brünger AT. 1992. Free R value: a novel statistical quantity for assessing the accuracy of crystal structures. *Nature* 355:472–475. <https://doi.org/10.1038/355472a0>.
 50. Cheng CY, Chou FC, Das R. 2015. Modeling complex RNA tertiary folds with Rosetta. *Methods Enzymol* 553:35–64. <https://doi.org/10.1016/bs.mie.2014.10.051>.
 51. Pronk S, Páll S, Schulz R, Larsson P, Bjelkmar P, Apostolov R, Shirts MR, Smith JC, Kasson PM, Van Der Spoel D, Hess B, Lindahl E. 2013. GROMACS 4.5: a high-throughput and highly parallel open source molecular simulation toolkit. *Bioinformatics* 29:845–854. <https://doi.org/10.1093/bioinformatics/btt055>.
 52. Lindorff-Larsen K, Piana S, Palmo K, Maragakis P, Klepeis JL, Dror RO, Shaw DE. 2010. Improved side-chain torsion potentials for the Amber ff99SB protein force field. *Proteins* 78:1950–1958.

53. Adams PD, Afonine PV, Bunkóczi G, Chen VB, Davis IW, Echols N, Headd JJ, Hung LW, Kapral GJ, Grosse-Kunstleve RW, McCoy AJ, Moriarty NW, Oeffner R, Read RJ, Richardson DC, Richardson JS, Terwilliger TC, Zwart PH. 2010. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr* 66:213–221. <https://doi.org/10.1107/S0907444909052925>.
54. Robert X, Gouet P. 2014. Deciphering key features in protein structures with the new ENDscript server. *Nucleic Acids Res* 42:W320–W324. <https://doi.org/10.1093/nar/gku316>.
55. Kabsch W, Sander C. 1983. Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers* 22:2577–2637. <https://doi.org/10.1002/bip.360221211>.