

RESEARCH

Open Access

# Structural characterization of the C-terminal domain of SARS-CoV-2 nucleocapsid protein



Renjie Zhou<sup>1</sup>, Rui Zeng<sup>1</sup>, Albrecht von Brunn<sup>2</sup> and Jian Lei<sup>1\*</sup>

## Abstract

The newly emerging severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) has resulted in a global human health crisis. The CoV nucleocapsid (N) protein plays essential roles both in the viral genomic RNA packaging and the regulation of host cellular machinery. Here, to contribute to the structural information of the N protein, we describe the 2.0 Å crystal structure of the SARS-CoV-2 N protein C-terminal domain (N-CTD). The structure indicates an extensive interaction dimer in a domain-swapped manner. The interface of this dimer was first thoroughly illustrated. Also, the SARS-CoV-2 N-CTD dimerization form was verified in solution using size-exclusion chromatography. Based on the structural comparison of the N-CTDs from *alpha*-, *beta*-, and *gamma*-CoVs, we demonstrate the common and specific characteristics of the SARS-CoV-2 N-CTD. Furthermore, we provide evidence that the SARS-CoV-2 N-CTD possesses the binding ability to single-stranded RNA, single-stranded DNA as well as double-stranded DNA in vitro. In conclusion, this study could potentially accelerate research to understand the complete biological functions of the new CoV N protein.

**Keywords:** Coronavirus, SARS, Nucleocapsid, Viral RNA, Structure

## Introduction

Over the past two decades, the emergence of novel human coronaviruses has caused continuous threats to the public health systems worldwide. Severe acute respiratory syndrome coronavirus (SARS-CoV) infected more than 8000 people with ~10% fatality rate during the global outbreak of 2002/2003 [1]. Ten years later, another highly pathogenic agent, the Middle-East respiratory syndrome coronavirus (MERS-CoV) appeared in the Arabian Peninsula and currently infected approximately 2500 cases with a ~35% mortality rate ([www.who.int](http://www.who.int)). At the end of 2019, a new coronavirus (CoV), SARS-CoV-2 (also named 2019-nCoV) was identified (GenBank: MN908947, [2]). This new CoV could cause the so-called coronavirus disease 2019 (COVID-19). COVID-19 includes various symptoms from the general mild respiratory disease

(i.e. sneezing and coughing) to severe pneumonia as well as enteric disease [3]. As of July 08, 2020, more than 11.5 million confirmed SARS-CoV-2 infection cases have been reported globally, with more than 535,000 deaths (<https://covid19.who.int/>). However, no efficient vaccine or treatment is available so far.

CoVs belong to the family *Coronaviridae* of the order *Nidovirales* and infect various species [4], such as human, cattle, pigs, cats, and birds *etc.* They are enveloped positive-sense single-stranded (+ss) RNA viruses with the largest genome of all currently known RNA viruses. CoVs are divided into four genera: *alpha*-, *beta*-, *gamma*-, and *delta-coronaviruses* [5]. SARS-CoV-2, similar to SARS-CoV and MERS-CoV, belongs to the *beta-coronavirus* group. Its genome comprises about 29.9 kb nucleotides and shares ~80% of genome identity with SARS-CoV [6]. The 5'-terminal two-thirds of its genome contain ORF1a and ORF1a/b encoding sixteen nonstructural proteins Nsp1–16. They are mainly responsible for the membrane-associated viral replication/transcription

\* Correspondence: [lejian@scu.edu.cn](mailto:lejian@scu.edu.cn)

<sup>1</sup>National Clinical Research Center for Geriatrics, State Key Laboratory of Biotherapy and Cancer Center, West China Hospital, Sichuan University, No.17, Block 3, Southern Renmin Road, Chengdu 610041, Sichuan, China  
Full list of author information is available at the end of the article



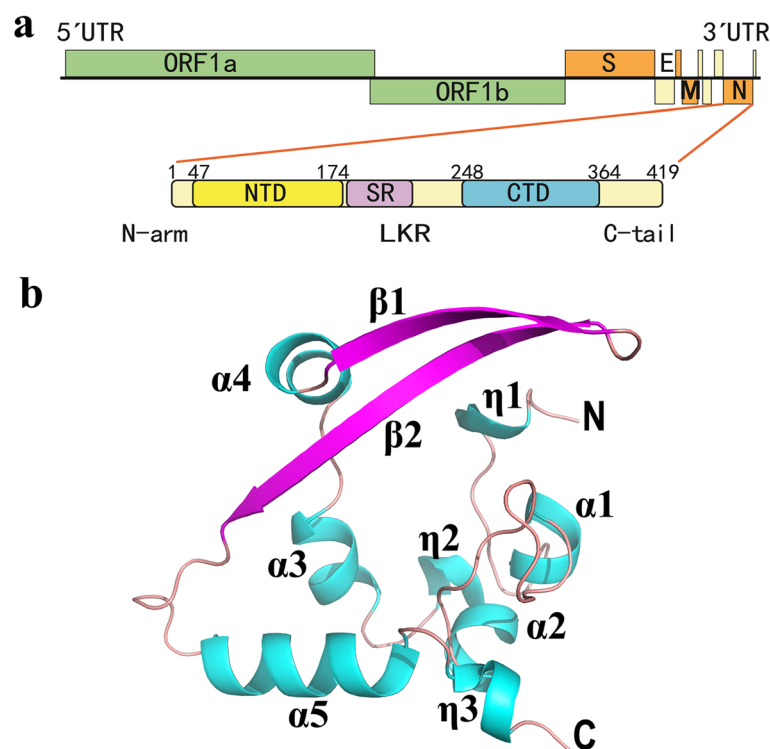
© The Author(s). 2020 **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

complex (RTC) formation [7]. The 3'-proximal one-third encodes four structural proteins: the spike (S), envelope (E), membrane (M, also called matrix protein) and nucleocapsid (N) proteins, as well as several accessory proteins. The S protein mediates the virus recognition of the host cell surface receptor and subsequent membrane fusion [8]. The E and M proteins are mainly involved in viral envelope formation, virus assembly, and viral particle budding [9, 10]. The N protein is a multifunctional protein and plays many roles in the viral life cycle [11, 12].

As a diagnostic marker [13], the coronavirus N protein is present in a large copy number in the CoV-infected host cells. The primary function of the CoV N protein is to bind the viral RNA and form a helical ribonucleoprotein (RNP) complex, in order to protect the viral genome and maintain reliable viral replication [14–16]. Furthermore, the SARS-CoV N regulates the host innate immune response by inhibiting interferon  $\beta$  (IFN- $\beta$ ) production [17, 18]. However, the exact mechanisms of these processes are still unclear. The CoV N is also involved in binding other viral proteins. It interacts with the E and M proteins to facilitate virus envelope

formation and particle assembly [19–21]. MHV (murine hepatitis virus, a *beta*-CoV) N protein binds Nsp3 to tether the viral genome and RTC complex [22, 23]. In addition, the N protein interacts with numerous host cell proteins as well. For example, the SARS-CoV N directly binds to host translation elongation factor 1 $\alpha$  (EF1 $\alpha$ ) to inhibit host cell proliferation, including that of peripheral blood lymphocytes [24]. It also interacts with the complex cyclin-CDK (cyclin dependent kinase) to regulate the cell cycle for facilitating CoV replication [25]. Recently, using affinity-purification mass spectrometry methods to systematically identify all SARS-CoV-2 viral proteins and host protein interactions, the SARS-CoV-2 N was described to potentially interact with 15 human proteins [26]. Most of them are related to RNA processing (i.e. Polyadenylate-binding protein 1 and 4: PABP-1 and PABP-4) or stress granule regulation (e. g., G3BP-1: Ras GTPase-activating protein-binding protein 1, G3BP2). Although these interactions have to be further investigated, they imply the major ability of CoV N protein to interact with other proteins.

All CoV N proteins possess a conserved modular composition comprising two structured domains, the N-



**Fig. 1** Overall structure of the SARS-CoV-2 N-CTD. **a** Genome organization of coronavirus and the domain composition of the SARS-CoV-2 nucleocapsid (N) protein. ORF: open reading frame; S: spike; E: envelope; M: membrane. N protein consists of the N-terminal arm (N-arm), the N-terminal domain (NTD), the central linker region (LKR) within a Ser/Arg (SR)-rich motif, the C-terminal domain (CTD), and the C-terminal tail (C-tail). **b** Ribbon view of the overall structure of the SARS-CoV-2 N-CTD monomer. The structure of N-CTD contains three  $3_{10}$  ( $\eta$ ) helices, five  $\alpha$  helices, and two antiparallel  $\beta$ -strands with the order  $\eta$ 1- $\alpha$ 1- $\alpha$ 2- $\eta$ 2- $\alpha$ 3- $\alpha$ 4- $\beta$ 1- $\beta$ 2- $\alpha$ 5- $\eta$ 3. Helices are shown in cyan, strands in purple. The N and C termini of the N-CTD are marked. Fig. **b** was prepared with the program PyMOL (<https://pymol.org>)

terminal domain (NTD) and the C-terminal domain (CTD), interspersed by three intrinsically disordered regions (IDRs) known as the N-terminal arm (N-arm), the central linker region (LKR) within a Ser/Arg (SR)-rich motif, and the C-terminal tail (C-tail) [11, 27] (Fig. 1a). The N-NTD, N-CTD and C-tail domains all were reported to bind viral RNA in SARS-CoV [28–30]. The SR-rich region in the LKR could regulate the N protein oligomerization upon phosphorylation [31]. The N protein self-association is necessary for viral RNP assembly [32]. The N-CTD has been shown to directly participate in N protein dimerization and oligomerization [29, 30, 32, 33]. Furthermore, the N protein inhibits IFN- $\beta$  and binds to EF1 $\alpha$  (mentioned above) mainly through the N-CTD [18, 24].

To date, the N-CTD structures of SARS-CoV, IBV (infectious bronchitis virus), MHV (mouse hepatitis virus), HCoV-NL63, and MERS-CoV have been determined by NMR [30] and crystallography [29, 33–37]. During preparation of this manuscript, the structure of SARS-CoV-2 N-CTD was reported on the BioRxiv server [38]. These authors intensely discuss the architecture and self-assembly properties of the N protein in SARS-CoV-2, however, the detailed interactions between the N-CTD dimer and the evidence of the N-CTD binding to nucleotides are not reported. In the present study, we determined the crystal structure of SARS-CoV-2 N-CTD at a resolution of 2.0 Å, and we report the interactions of the N-CTD homo-dimer in strong detail. We further identified this dimer in solution. The common and specific characteristics of the SARS-CoV-2 N-CTD are discussed. In addition, we provide direct evidence that the SARS-CoV-2 N-CTD can interact with ssRNA, ssDNA and double-stranded DNA (dsDNA) in vitro. These findings facilitate detailed understanding of the new CoV-2 N protein.

## Results and discussion

### Overall structure of the C-terminal domain of SARS-CoV-2 N protein (N-CTD)

We designed a SARS-CoV-2 (GenBank: MN908947, [2]) N-CTD construct containing the residues from Lys248 to Pro364 of the N protein (Fig. 1a) and determined the three-dimensional structure of the corresponding region at 2.0 Å. The diffraction parameters and refinement statistics are summarized in Table 1. Four molecules (termed as A, B, C and D) formed by two dimers exist in the asymmetric unit (ASU) of the N-CTD crystal. In the final model, residues 256–364 of Chain A, 252–364 of Chain B, 253–364 of Chain C, as well as 257–364 of Chain D were successfully built. The side-chains of the following residues showed the double conformation: Thr329 in Chain A; Thr282, Met317 and 322 in Chain B; and Thr329, 334 in Chain D. The root mean square

**Table 1** Data collection and refinement statistics

	SARS-CoV-2 N-CTD PDB code: 7C22
<b>Data collection statistics</b>	
Space group	$P_1$
Unit-cell dimensions (Å, °)	$a = 43.76, b = 49.46, c = 68.82$ $\alpha = 106.79, \beta = 90.04, \gamma = 97.79$
Wavelength (Å)	0.97851
$V_m$ (Å <sup>3</sup> /Da)	2.69
Solvent content (%)	52.24
Resolution range (Å)	19.81–2.00 (2.05–2.00)
Number of unique reflections	35,328 (2636)
$R_{merge}$	0.107 (0.639)
Completeness (%)	95.6 (96.0)
Mean $I/\sigma(I)$	5.9 (1.4)
Multiplicity	3.5 (3.5)
$CC_{1/2}$	0.995 (0.752)
<b>Refinement statistics</b>	
$R_{factor}$ (%) <sup>a</sup>	18.9
$R_{free}$ (%) <sup>a</sup>	23.8
<b>No. of atoms</b>	
Protein	3568
Ligand	15
Water	395
Clashscore <sup>b</sup>	7
r.m.s.deviation in bond lengths (Å)	0.007
r.m.s.deviation in bond angles (°)	1.437
Average B-factor for all atoms (Å <sup>2</sup> )	26.0
<b>Ramachandran plot</b>	
Residues in preferred regions (%)	96.24
Residues in allowed regions (%)	3.76
Residues in outlier regions (%)	0.00

<sup>a</sup> $R_{factor} = \sum_{hkl} |F_o(hkl) - F_c(hkl)| / \sum_{hkl} F_o(hkl)$ .  $R_{free}$  was calculated for a test set of reflections (4.9%) omitted from the refinement

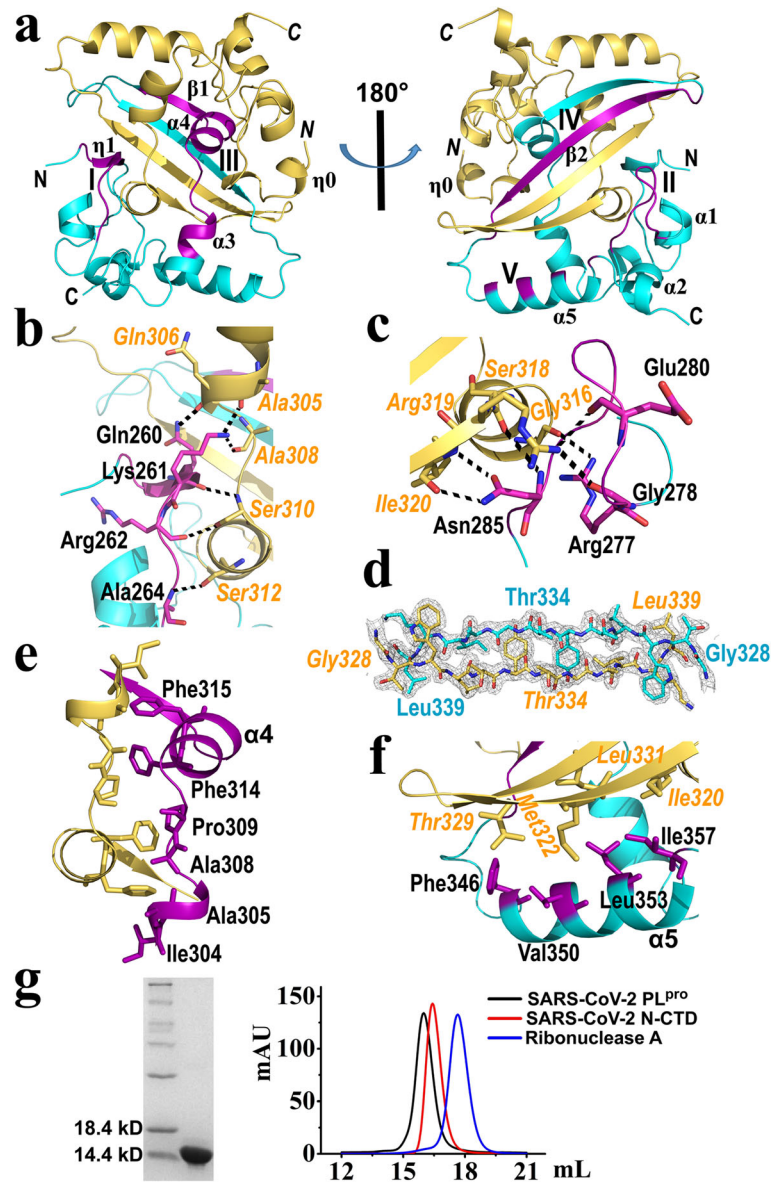
<sup>b</sup>Clashscore is defined as the number of clashes calculated for the model per 1000 atoms (including hydrogens) of the model. Hydrogens were added by MolProbity [39]

deviation (RMSD) values between any two subunits of these four molecules are 0.2–0.9 Å for all C $\alpha$  atoms, as calculated by the DaLi server [40]. Therefore, the overall structures of each molecule are almost identical.

According to the DSSP (assigning secondary structure of protein) server [41], the structure of N-CTD contains three  $3_{10}$  ( $\eta$ ) helices, five  $\alpha$  helices, and two antiparallel  $\beta$ -strands (forming a  $\beta$ -hairpin) in all four subunits, with the order  $\eta 1-\alpha 1-\alpha 2-\eta 2-\alpha 3-\alpha 4-\beta 1-\beta 2-\alpha 5-\eta 3$  (Fig. 1b). The topology shape of N-CTD monomer resembles the letter “C”, forming by the  $\beta$ -hairpin and extending to the rest of the molecule. Interestingly, the structure of subunit B

contains one extra  $3_{10}$  helix (named  $\eta_0$ ) formed by four residues Glu253-Lys256 (Fig. 2a). These residues are partially or completely missing in molecules A and D due to the lack of electron density. Subunit C includes Glu253-Lys256 with poor electron density but these four

residues do not form a  $3_{10}$  helix. These results indicate that the N-termini of the N-CTD are somehow flexible. Takeda *et al.* reported the N-terminal residues 248–259 of the SARS-CoV N-CTD (corresponding to residues 247–258 in SARS-CoV-2), which are disordered in NMR



**Fig. 2** Dimerization of the SARS-CoV-2 N-CTD. **a** Ribbon diagram of the N-CTD dimer. The overall structure displays a rectangular slab with all helices at one side (left image) and the  $\beta$ -sheet at the opposite side (right image). Monomers A and B are depicted in cyan and yellow, respectively. The regions I and III for the N-CTD dimerization are marked in the left image while the regions II, IV and V are indicated in the right one. N- and C- termini of subunit A and B (in italics) are labelled. **b** The hydrogen-bond network in region I for the N-CTD dimerization. Residues in protomer A are marked in black, while residues from subunit B are labelled in italics in orange. **c** The hydrogen-bond network in region II for the N-CTD dimerization. **d** The 2Fo-Fc electron density (gray, 1.5 $\sigma$ ) of the  $\beta_2$  strands (residue Gly328-Leu339; region IV) of both subunits. Residues (N- and C- termini) of the  $\beta_2$  are labelled. The non-conserved residue Thr334 is also marked. **e** and **f** The conserved hydrophobic cores in region III and V for the N-CTD dimerization. **g** Analytical gel filtration assay was performed. The molecular mass of the SARS-CoV-2 N-CTD is ~13.4 kD, which is identified by the SDS-PAGE (left image). In the gel filtration assay, the N-CTD is shown as a dimer in solution (right image). Ribonuclease A (blue line; Sangon Biotech, China): ~13.7 kD. SARS-CoV-2 papain-like protease (PL<sup>pro</sup>; black line): ~35.5 kD. The peak position of N-CTD (red line) is corresponding to ~29.8 kD, representing a dimer of N-CTD (theoretical molecular mass is ~26.7 kD). Figures a-f were made by the program PyMOL (<https://pymol.org>)



structure but more rigid in the crystal structure [30]. These authors suspected that the differences were likely due to crystal packing [30]. However, the N-terminal region of N-CTD is indeed involved in its dimerization/oligomerization (see below). Therefore, future research should investigate whether these conformational changes in the N-termini region could be related to the N protein associations in vivo.

The monomeric N-CTD shows a loose and, – to a certain extent –, an extended conformation with a huge cavity in the center of its structure (Fig. 1b). Thus, the single-molecule conformation is most likely unstable. Dimerization (or oligomerization) is necessary to create a compact and stable conformation for this protein.

### Dimerization of the SARS-CoV-2 N-CTD

Dimerization of N-CTD (molecules A/B, or C/D) exists in our crystal structure. The overall shape of the dimer (here using the A/B dimer as an example) presents a rectangular slab with the two  $\beta$ -hairpins at one side and all helices at the opposite side (Fig. 2a). Approximately 2600 Å<sup>2</sup> of the surface from each N-CTD monomer (A and B) are buried upon dimer formation, as calculated by the PDBePISA server [42]. The predominant dimerization interface of each subunit includes two  $\beta$  strands and  $\eta$ 1,  $\alpha$ 3,  $\alpha$ 4 and  $\alpha$ 5 helices, located in five regions from N-terminus to C-terminus of the N-CTD (Fig. 2a): **I**, residues Pro258-Ala264 ( $\eta$ 1 plus the partial loop between  $\eta$ 1 and  $\alpha$ 1); **II**, residues Arg277-Phe286 (the middle region of the loop between  $\alpha$ 1 and  $\alpha$ 2); **III**, residues Ile304-Met322 (almost the whole region  $\alpha$ 3- $\alpha$ 4- $\beta$ 1); **IV**, residues Gly328-Leu339 ( $\beta$ 2 strand); **V**, residues Phe346-Ile357 ( $\alpha$ 5 helix).

In regions **I**, **II** and **IV**, the interactions between molecules A and B mainly refer to hydrogen bonding. In region **I**, residues Gln260-Ala264 of subunit A form six hydrogen bonds with *Ala305-Ser312* (the residues from subunit B are in italics and as follows) of protomer B (Fig. 2b), and vice versa. Notably, the side-chain of the conserved residue Lys261 is involved in two hydrogen bonds with the main-chain oxygen atoms of *Ala 305* and *308*, thus further fixing the N-terminus and increasing the dimerization formation (Fig. 2b).

In region **II** (Arg277-Phe286, located in the loop between  $\alpha$ 1 and  $\alpha$ 2), a total of six hydrogen bonds were found between A and B molecules (Fig. 2c; only the loop from molecule A is shown here for simplicity). The side-chain of Arg277 of subunit A forms a strong hydrogen bond with the main-chain oxygen of *Gly316* from protomer B (Fig. 2c). The main-chains of Gly278 and Glu280 interact with the side-chain of *Arg319* with two hydrogen bonds. Asn285 with its main-chain and side-chain makes three hydrogen bonds with the main-chains of *Ser318* and *Ile320*. Arg277, Gly278, Glu280, as well as

Asn285 form a hydrogen-bond network with the adjacent protomer, thus reducing the flexibility of the longest loop between  $\alpha$ 1 and  $\alpha$ 2 of the N-CTD. This implies that a single N-CTD does not possess an ideally stable conformation. Furthermore, residues Arg277 and Asn285 are absolutely conserved in the *alpha*- (i.e. HCoV-NL63), *beta*- (i.e. SARS-CoV-2, SARS-CoV, and MERS-CoV), and *gamma* (i.e. IBV)-CoVs (Fig. 4), suggesting that these hydrogen bonds in the dimerization are conserved in various CoVs.

In region **IV** (Gly328-Leu339,  $\beta$ 2 strand region), the  $\beta$ 2 of monomer A interacts with  $\beta$ 2 strand from subunit B to make an intermolecular antiparallel  $\beta$ -sheet, with a total of 10 main-chain hydrogen bonds formed by five pairs *Thr329-Leu339*, *Leu331-Ile337*, *Tyr333-Gly335*, *Gly335-Tyr333*, and *Ile337-Leu331* (Fig. 2d), leading to a significant increase in the stability of the N-CTD homodimer. This observation is in good agreement with previous results [29, 30, 33]. In the  $\beta$ 2 strand region, Thr334 of SARS-CoV-2 N-CTD is replaced by His335 in SARS-CoV (Figs. 2d and 4). The His335Ala mutation of SARS-CoV N-CTD reduces its RNA binding affinity by about 50% [30]. Takeda *et al.* explained that the  $\beta$ -sheet region could be part of the N-CTD nucleotide-binding site [30]. His335 in SARS-CoV is unique compared to the corresponding residue in other CoVs (Fig. 4, indicated by a red arrow). It will be worth investigating the effect of Thr instead of His in SARS-CoV-2 N-CTD on its RNA binding affinity in the future.

In the left two regions, **III** and **V**, the hydrophobic interactions between the two subunits are dominant. In region **III** (Ile304-Met322), residues Ile304, Ala305, Ala308, Pro309, Phe314, and Phe315 from subunits A and B interact with each other to form a strong hydrophobic core (Fig. 2e). In region **V**, residues Phe346, Val350, Leu353 and Ile 357 of the  $\alpha$ 5 helix in protomer A display hydrophobic interactions with *Thr329* (C $\beta$  and C $\gamma$  atoms of the side-chain), *Met322*, *Leu331* and *Ile320* of subunit B, respectively (Fig. 2f). Among all the residues mentioned here, only Pro309 is absolutely conserved in the *alpha*-, *beta*-, and *gamma*-CoVs; however, the hydrophobic property of the left residues is almost retained in different CoV genera (Fig. 4), indicating that the hydrophobic cores in regions **III** and **V** are preserved for N protein dimerization.

With the hydrogen bonds and hydrophobic interactions mentioned above, a very stable N-CTD dimer is formed. This dimer was confirmed in solution by gel filtration assay (Fig. 2g; the sample used for this assay is shown in SDS-PAGE). In addition, a size exclusion chromatography coupled to multi-angle light scattering (SEC-MALS) assay was performed in order to directly detect the molecular mass (M. M.) of the SARS-CoV-2 N-CTD in solution (supplementary Fig. S1). According

to this assay, the M. M. is  $26.8 \pm 0.7$  kD and closely matches with the theoretical value (N-CTD dimer:  $\sim 26.7$  kD, Fig. S1). This intertwined dimer is very likely a functional unit of the SARS-CoV-2 N protein.

The crystal of the SARS-CoV N-CTD consists of an octamer formed by four homo-dimers per ASU (PDB code: 2CJR, [29]). A putative helical oligomer CTDs-RNA model was proposed by packing these octamers [11, 29, 30]. In our case, four molecules formed by two dimers exist in an ASU. The dimeric packing pattern is different from that of SARS-CoV N-CTD [29]. The interface area of the SARS-CoV-2 two dimers is about  $220 \text{ \AA}^2$  (Chain A and Chain C) as determined by the PDBePISA server [42]. The complex (dimer-dimer) formation significance score (scale from 0 to 1) is 0 compared to 0.971 for the intra-dimer (monomer-monomer) formation. In addition, we did not detect tetramers or higher oligomers in the gel filtration assay. Therefore, the tetramer per ASU in our crystal is likely to result from crystal packing.

On the other hand, self-association (dimerization or oligomerization) of CoV N plays a central role during virus replication [15]. The C-terminal tail of the HCoV-229E N protein is important for its oligomerization [44]. Using the C-terminal tail peptide could interrupt the self-association of the N-CTD of HCoV-229E to further attenuate virus replication [44]. Therefore, designing the proper peptides to interfere with the dimerization or oligomerization of the N protein is a potential anti-viral strategy to counter the SARS-CoV-2.

#### Comparison of the N-CTD structures in various coronaviruses

A structural similarity search of SARS-CoV-2 N-CTD subunit A against the Protein Data Bank was performed by the DaLi server [40]. The results indicate that the three-dimensional structures of N-CTDs are very conserved in *alpha*-, *beta*-, and *gamma*-CoVs, despite the lowest sequence identity of 27% between SARS-CoV-2 and IBV (Table 2 and Fig. 3a).

The structure most similar to SARS-CoV-2 N-CTD is that of SARS-CoV N-CTD (95% sequence identity; PDB code: 2CJR, [29]) with an RMSD  $0.4 \text{ \AA}$  over 109 Ca atoms out of 113 residues. The RMSD values between the SARS-CoV-2 N-CTD and the corresponding domains in MERS-CoV [37], HCoV-NL63 [36], as well as IBV [34] are  $1.1 \text{ \AA}$  (PDB code: 6G13),  $2.1 \text{ \AA}$  (PDB code: 5EPW), and  $1.8 \text{ \AA}$  (PDB code: 2GE8), respectively (Table 2). In addition to the differences of the flexible N- and C-termini among these N-CTDs, the loop between  $\beta 1$  and  $\beta 2$  of MERS-CoV is obviously longer than those from other CoVs (Figs. 3a and 4). It remains to be elucidated whether the long insertion into the  $\beta$ -hairpin plays any role in viral replication.

**Table 2** Structural comparisons of the SARS-CoV-2 N-CTD with other homology proteins

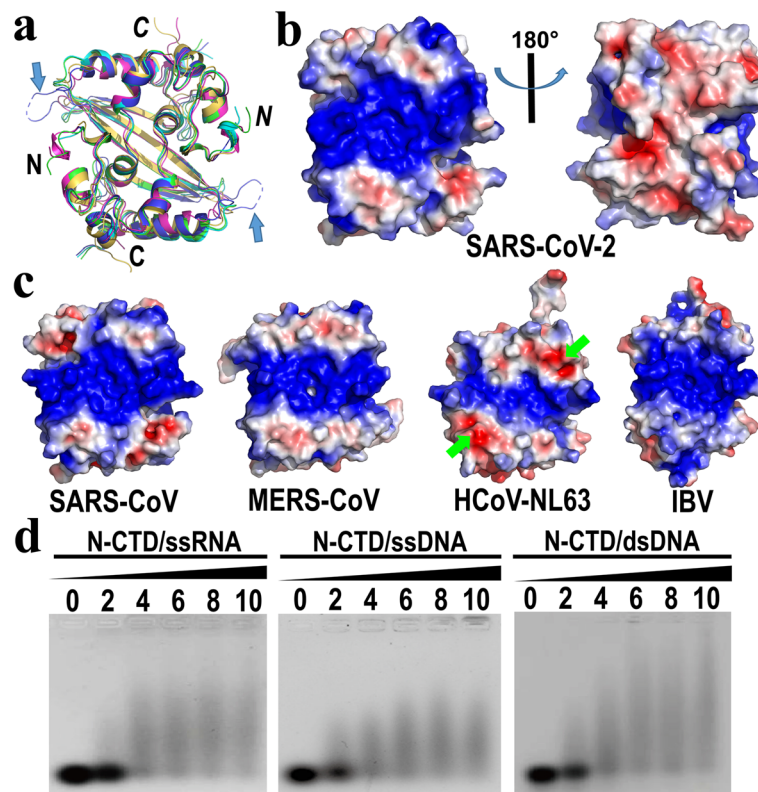
N-CTDs	PDB/Chain	SARS-CoV-2 N-CTD				
		Z-score	RMSD ( $\text{\AA}$ )	Ca <sup>a</sup>	% id <sup>b</sup>	References
SARS-CoV	2CJR/C	19.1	0.4	109/113	95	[29]
MERS-CoV	6G13/D	18	1.1	109/114	54	[37]
HCoV-NL63	5EPW/B	13.8	2.1	102/110	32	[36]
IBV	2GE8/D	13.2	1.8	106/111	27	[34]

<sup>a</sup> aligned Ca atoms/total Ca atoms; <sup>b</sup> Sequence identity  
All values were calculated by the DaLi server [40]

SARS-CoV N-CTD is capable of interacting with ssRNA, ssDNA and dsDNA in vitro without nucleotide sequence specificity [29, 30], suggesting the N-CTD nucleotides binding is most likely due to nonspecific charge interactions. Residues 248–280 in the SARS-CoV N-CTD within a large number of Lys/Arg are essential for RNA binding [29, 30]. The corresponding residues, 247-TKKSAAEASKKPRQKRTATKAYNVTQAFRRGP-279, are conserved in SARS-CoV-2, with only one amino acid Ala267 (underlined) replacement to Gln268 in SARS-CoV (Fig. 4).

The electrostatic surface of the SARS-CoV-2 N-CTD in our structure displays a significantly positively charged region located at the helical side of the dimer (Fig. 3b, left), indicating a potential RNA-binding site. Further analysis revealed the clustering of positive charges due to the N-termini eight basic residues (Lys256, Lys257, Arg259, Lys261, Arg262, Lys266, Arg276 and Arg277) of each subunit (Lys248/249 are invisible in our structure because of the lack of electron density, as mentioned above). On the other hand, the  $\beta$ -sheet side of dimer shows an almost acidic and neutral environment (Fig. 3b, right), suggesting an unideal RNA binding site. Although this  $\beta$ -hairpin site related to RNA binding in vitro was reported in SARS-CoV [30], we need to further confirm whether a similar binding site exists in SARS-CoV-2 or not.

Next, we checked the electrostatic surface of the homology N-CTDs from *alpha*-, *beta*-, and *gamma*-CoVs (Fig. 3c). Most of the eight positive residues (see above) were conserved in SARS-CoV, MERS-CoV, HCoV-NL63, and IBV (Fig. 4). A similar surface with the clustering of positive charges exists in all N-CTDs, suggesting a conserved viral RNA binding pattern throughout the evolution of various CoVs. However, we noticed that the positively charged groove in HCoV-NL63 is much smaller than those of other CoVs. We found two acidic residues Glu243/244 are unique in HCoV-NL63, which could significantly reduce the positive charge around this region (Fig. 3c). These two residues are replaced by neutral and/or hydrophobic residues in other CoVs (Fig. 4). How the different distribution of the electronic surface



**Fig. 3** Structural comparison of N-CTDs from various CoVs and oligonucleotide-binding assay of the SARS-CoV-2 N-CTD. **a** Superposition of N-CTDs from SARS-CoV-2 (cyan), SARS-CoV (green, PDB code: 2CJR), MERS-CoV (blue, PDB code: 6G13), HCoV-NL63 (purple, PDB code: 5EPW) and IBV (yellow, PDB code: 2GE8). The N- and C- termini of subunits A and B (in italics) are labelled. The long insertion region of each  $\beta$ -hairpin in the MERS-CoV N-CTD is marked by a blue arrow. **b** The electrostatic surface of the SARS-CoV-2 N-CTD dimer structure. The contouring level is  $-3$  kBT/e (red) to  $3$  kBT/e (blue). The helices side of the rectangular slab dimer shows a strong positively charged region for the potential RNA binding (left image). The opposite  $\beta$ -sheet side is almost a negatively charged and neutral region (right image). The orientation of the left image is same as in the ribbon representations in (a). **c** The electrostatic surfaces of the N-CTD dimers from SARS-CoV, MERS-CoV, HCoV-NL63 and IBV. The PDB codes used here are same as in (a). Residues Glu243/244 of each subunit in HCoV-NL63 are marked by green arrows. All orientations are same as in the ribbon view in (a). All figures were generated using the program PyMOL (<https://pymol.org>). **d** The EMSA experiment of the SARS-CoV-2 N-CTD. The mobility shifts of ssRNA, ssDNA and dsDNA bound to the N-CTD are shown, respectively. The molar ratios of N-CTD/nucleotides in each lane are indicated

affecting the viral RNA binding affinity in various CoVs is interesting to be investigated later.

#### Nucleotides-binding activity of the SARS-CoV-2 N-CTD

We have analyzed the potential RNA/DNA binding ability of the SARS-CoV-2 N-CTD (see above), however, the direct evidence for the nucleotide-binding activity of this N-CTD is scarce. Therefore, we performed the electrophoretic mobility shift assay (EMSA) (Fig. 3d). The 17-mer ssRNA oligonucleotides 5'-UGUUCUCUAAACGA ACU-3', which are located in the 5' untranslated region of the SARS-CoV-2 genome (GenBank: MN908947, nucleotides 60–76), were used in our assay. This 17-mer RNA also exists in the SARS-CoV genome and is referred to as the stem-loop 3 (SL3, including the transcriptional regulatory sequences (TRS)) [45]. The amount of the free 17-mer RNA decreased with the addition of more N-CTD protein (Fig. 3d, right),

indicating the binding affinity between RNA and the proteins. Furthermore, the N-CTD could interact with this RNA's mimic DNA as well as dsDNA nucleotides (Fig. 3d, middle and left). Taken together, the SARS-CoV-2 N-CTD displays the binding activity of both RNA and DNA, which is similar to that of the SARS-CoV N-CTD [29, 30].

#### Conclusions

CoV nucleocapsid protein, as an essential multifunctional viral protein, plays roles in the viral RNA packaging and the regulation of host cellular processes, such as cell cycle and innate immune response. In an effort to better understand the structure and functions of the SARS-CoV-2 N protein, our study provides the structural data of its C-terminal domain and reports the nucleotide-binding activity of the corresponding region. In addition, CoV N protein is involved in binding







various viral and host proteins. Despite of all recent discoveries, the structure of N protein in complex with viral RNA or its partner proteins is still missing. Future research should focus on demonstrating such structure(s).

## Materials and methods

### Recombinant production of the C-terminal domain of SARS-CoV-2 nucleocapsid protein (N-CTD)

The cDNA plasmids (WH-nCoV in pBluescript II SK (+)) encoding for the full-length SARS-CoV-2 nucleocapsid protein were purchased from General Biosystems, Anhui, China. This cDNA sequence is identical to the sequence of the N gene in SARS-CoV-2 isolated from Wuhan-hu-1 (GenBank: MN908947, [2]). The SARS-CoV-2 N-CTD contains the residues Lys248-Pro364 of the N protein. The corresponding N-CTD gene was amplified by PCR with the forward primer 5'-CTAGCTAGCAAGAAATCTGCTGCTGAGGCTTC-3' and the reverse primer 5'-CCGCTCGAGTTATGGGAA TGTTTTGTATGCGTC-3'. The PCR product was further digested using the *NheI* and *XhoI* restriction endonucleases and ligated into modified pET28a plasmids containing an N-terminal hexa-histidine tag and a Tobacco Etch Virus (TEV) protease cleavage site (the modified plasmids were kindly provided by our colleague Prof. Qiang Chen). The recombinant plasmid DNA was verified by sequencing (Youkang Biology Company, Chengdu, China).

The corrected N-CTD pET28a plasmid was transformed into *E. coli* BL21(DE3) (Novagen). The transformed cells were grown overnight at 37 °C in 50 mL Luria-Broth (LB) medium supplemented with kanamycin at a final concentration 50 µg/mL. The culture was then inoculated into 2 × 1 L LB medium the following day. When the OD600 value of the culture reached 0.6–0.8, overexpression of the N-CTD was induced for 16 h through 0.5 mM isopropyl-D-thiogalactoside (IPTG) supplementation at 18 °C. The 2 × 1 L culture was subsequently harvested by centrifugation for 15 min, 4000 rpm at 4 °C. Pellets were re-suspended in 50 mL buffer A (20 mM Tris-HCl, 10 mM imidazole, 500 mM NaCl, pH 7.5) and lysed by sonication on ice. Debris was removed by centrifugation for 30 min at 20,000 rpm at 4 °C. The supernatant was applied to the Ni Sepharose™ Fast Flow beads (GE Healthcare). The His-tagged N-CTD protein was eluted using buffer B (20 mM Tris-HCl, 500 mM imidazole, 500 mM NaCl, pH 7.5) with a step-gradient method. The target protein was then cleaved by TEV protease (leaving three extra residues Gly-Ala-Ser at the N-terminus of the N-CTD) and dialyzed against buffer C (20 mM Tris-HCl, 150 mM NaCl, pH 7.5) overnight at 4 °C. Next day, the target protein was applied to Nickel beads again to remove the uncleaved His-tag protein. The N-CTD without the His-

tag protein was further purified by gel filtration (Superdex 200 Increase 10/300 GL, GE Healthcare) using buffer C. The quality of the N-CTD protein was checked by sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE, Fig. 2g).

### Crystallization and data collection

Purified SARS-CoV-2 N-CTD was concentrated to ~ 40 mg/mL in buffer C. Crystallization experiments were performed at 291 K by the sitting-drop vapor-diffusion method with 1 µL protein plus 1 µL reservoir. The commercial screen kits Index, SaltRx 1/2, Crystal Screen 1/2, PEG/Ion Screen 1/2, PEGRx 1/2 (Hampton Research) as well as Structure screen 1/2, PACT premier and JCSG plus (Molecular Dimensions) were used. Crystals were observed under Index condition No.57 (0.05 M ammonium sulfate, 0.05 M Bis-Tris pH 6.5, 30% v/v pentaerythritol ethoxylate (15/4 EO/OH) and Structure screen 1 No. 2 (0.2 M ammonium acetate, 0.1 M sodium acetate pH 4.6, 30% PEG4000). Crystals were reproduced under the Index No.57 condition within 3 ~ 5 days. Using 25% PEG400 as a cryo-protected agent, the optimized crystals were shock-cooled in liquid nitrogen. A diffraction dataset to 2.0 Å was collected with the X-ray wavelength 0.97851 Å at Shanghai synchrotron radiation facility (SSRF) beamline BL19U1, Shanghai, China. This dataset was processed by *XDS* [46] and scaled with *Aimless* in *CCP4* [47]. The space group is  $P_1$ , with unit-cell parameters  $a = 43.76$  Å,  $b = 49.46$  Å,  $c = 68.82$  Å,  $\alpha = 106.79^\circ$ ,  $\beta = 90.04^\circ$ ,  $\gamma = 97.79^\circ$ . See Table 1 for the diffraction data statistics.

### Structure determination and refinement

The structure of N-CTD was solved by the molecular replacement method with the program *Phenix. Phaser* [48] using the modified search model (PDB ID: 2CJR, Chain A, [29]). Four molecules (Chain A, B, C and D) within the asymmetric unit (ASU) were identified. The initial model of the N-CTD was circularly rebuilt and refined using programs *Coot* [49] and *Refmac5* [50]. Residues 256–364 of Chain A, 252–364 of Chain B, 253–364 of Chain C and 257–364 of Chain D were successfully built in the final structure model with  $R_{factor}$  and  $R_{free}$  of 0.189 and 0.238, respectively. The final model refinement statistics are listed in Table 1.

### SEC-MALS assay

In order to directly measure the molecular mass of the SARS-CoV-2 N-CTD in solution, the size exclusion chromatography coupled to multi-angle light scattering (SEC-MALS) assay was performed (supplementary Fig. S1). The purified 50 µL N-CTD (~ 2 mg/mL) proteins (totally 100 µg) were injected to the column Superdex 200 Increase 10/300 GL (GE Healthcare) to run with the

flowrate at 0.5 ml/min. The light scattering signals and the refractive index profiles were collected with the miniDAWN and Optilab (Wyatt Technology). The molecular mass was calculated by the software ASTRA (Wyatt Technology).

### Electrophoretic mobility shift assay (EMSA)

The method of this assay was modified according to previous literature [29, 30]. The 17-mer ssRNA oligonucleotides 5'-UGUUCUCUAAACGAACU-3', which are located in the 5' untranslated region of the SARS-CoV-2 genome (GenBank: MN908947, nucleotides 60–76), was purchased from Youkang Biology Company (Chengdu, China). In addition, the corresponding DNA (5'-TGTTCTCTAAACGAACT-3') and its complement oligonucleotides (5'-AGTTCGTTTAGAGAACA-3', for double stranded DNA formation) were purchased from Sangon Biotech Co., Ltd. (Shanghai, China). The dsDNA was further prepared by equal numbers of the ssDNA and its complement DNA after denaturing at 95 °C and renaturing at room temperature. Next, the 1, 2, 3, 4 and 5 µL of the N-CTD (stock concentration: 100 µM) were incubated with 5 µL of dsRNA, ssDNA as well as dsDNA (each stock concentration: 10 µM) at 4 °C for 30 min, respectively. The N-CTD/nucleotides molar ratios are 2:1, 4:1, 6:1, 8:1 and 10:1. The sample of oligonucleotides without protein was used as the control. Reaction in each tube was set up to 10 µL aliquots totally. All the experiments were performed in the buffer: 20 mM HEPES, 150 mM NaCl, pH 7.5. These reaction samples were then separated on 1% agarose gel on ice at 60 V for 1 h. The results were visualized by Tanon-3500B imager (Tanon Co., Ltd., Shanghai, China).

### Supplementary information

**Supplementary information** accompanies this paper at <https://doi.org/10.1186/s43556-020-00001-4>.

#### Additional file 1: Fig. S1. SEC-MALS assay of the SARS-CoV-2 N-CTD.

The size exclusion chromatography coupled to multi-angle light scattering (SEC-MALS) assay was performed to determine the molecular mass (M. M) of the SARSCoV-2 N-CTD in solution. The M. M is 26.8 ± 0.7 kD (indicated by the red arrow; the theoretical N-CTD dimer is ~ 26.7 kD). LS: light scattering; dRI: differential refractive index.

### Acknowledgments

We thank the staff of Shanghai synchrotron radiation facility (SSRF) beamlines for great supports. We also thank Prof. Cong Liu (Shanghai Institute of Organic Chemistry, Chinese Academy of Sciences) to help for the SEC-MALS assay. Funding for this work by “the Fundamental Research Funds for the Central Universities”(20822041D4060) and grants from West China Hospital and Sichuan province, China (HX2019nCoV039; 2020YFS0010), are gratefully acknowledged. The atomic coordinates and structure factors of SARS-CoV-2 N-CTD are available from the PDB with the accession code 7C22.

### Authors' contributions

JL and RJZ designed experiments. RJZ purified and crystallized the protein as well as performed the related assays. RJZ collected the diffraction data. RJZ

and JL prepared the figures. JL wrote the manuscript. AvB and JL revised the manuscript. The authors read and approved the final manuscript.

### Funding

grant from Sichuan province (2020YFS0010).  
grant from West China Hospital (HX2019nCoV039).  
the Fundamental Research Funds for the Central Universities (20822041D4060).

### Availability of data and materials

PDB code: 7C22.

### Ethics approval and consent to participate

Not applicable.

### Consent for publication

Not applicable.

### Competing interests

The authors declare no conflict of interest.

### Author details

<sup>1</sup>National Clinical Research Center for Geriatrics, State Key Laboratory of Biotherapy and Cancer Center, West China Hospital, Sichuan University, No.17, Block 3, Southern Renmin Road, Chengdu 610041, Sichuan, China. <sup>2</sup>Max von Pettenkofer-Institute, Ludwig-Maximilians-University Munich and German Center for Infection Research (DZIF), Partner Site Munich, 80336 Munich, Germany.

Received: 25 June 2020 Accepted: 12 July 2020

Published online: 06 August 2020

### References

- Christian MD, Poutanen SM, Loutfy MR, Muller MP, Low DE. Severe acute respiratory syndrome. *Clin Infect Dis*. 2004;38(10):1420–7. <https://doi.org/10.1086/420743>.
- Wu F, Zhao S, Yu B, Chen Y-M, Wang W, Song Z-G, et al. A new coronavirus associated with human respiratory disease in China. *Nature*. 2020;579(7798):265–9. <https://doi.org/10.1038/s41586-020-2008-3>.
- Huang C, Wang Y, Li X, Ren L, Zhao J, Hu Y, et al. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. *Lancet*. 2020;395(10223):497–506. [https://doi.org/10.1016/S0140-6736\(20\)30183-5](https://doi.org/10.1016/S0140-6736(20)30183-5).
- Fehr AR, Perlman S. Coronaviruses: an overview of their replication and pathogenesis. *Methods Mol Biol*. 2015;1282:1–23. [https://doi.org/10.1007/978-1-4939-2438-7\\_1](https://doi.org/10.1007/978-1-4939-2438-7_1).
- Adams MJ, Lefkowitz EJ, King AMQ, Harrach B, Harrison RL, Knowles NJ, et al. Ratification vote on taxonomic proposals to the international committee on taxonomy of viruses (2016). *Arch Virol*. 2016;161(10):2921–49. <https://doi.org/10.1007/s00705-016-2977-6>.
- Zhou P, Yang X-L, Wang X-G, Hu B, Zhang L, Zhang W, et al. A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature*. 2020;579(7798):270–3. <https://doi.org/10.1038/s41586-020-2012-7>.
- Hagemeyer MC, Verheije MH, Ulasli M, Shaltiel IA, de Vries LA, Reggiori F, et al. Dynamics of coronavirus replication-transcription complexes. *J Virol*. 2010;84(4):2134–49. <https://doi.org/10.1128/JVI.01716-09>.
- Li F. Structure, function, and evolution of coronavirus spike proteins. *Annu Rev Virol*. 2016;3(1):237–61. <https://doi.org/10.1146/annurev-virology-110615-042301>.
- Neuman BW, Kiss G, Kunding AH, Bhella D, Baksh MF, Connelly S, et al. A structural analysis of M protein in coronavirus assembly and morphology. *J Struct Biol*. 2011;174(1):11–22. <https://doi.org/10.1016/j.jsb.2010.11.021>.
- Schoeman D, Fielding BC. Coronavirus envelope protein: current knowledge. *Virol J*. 2019;16(1):69. <https://doi.org/10.1186/s12985-019-1182-0>.
- Chang C-K, Hou M-H, Chang C-F, Hsiao C-D, Huang T-H. The SARS coronavirus nucleocapsid protein—forms and functions. *Antivir Res*. 2014;103:39–50. <https://doi.org/10.1016/j.antiviral.2013.12.009>.
- McBride R, van Zyl M, Fielding BC. The coronavirus nucleocapsid is a multifunctional protein. *Viruses*. 2014;6(8):2991–3018. <https://doi.org/10.3390/v6082991>.

13. Che X-Y, Hao W, Wang Y, Di B, Yin K, Xu Y-C, et al. Nucleocapsid protein as early diagnostic marker for SARS. *Emerging Infect Dis.* 2004;10(11):1947–9. <https://doi.org/10.3201/eid1011.040516>.
14. Masters PS, Parker MM, Ricard CS, Duchala C, Frana MF, Holmes KV, et al. Structure and function studies of the nucleocapsid protein of mouse hepatitis virus. *Adv Exp Med Biol.* 1990;276:239–46. [https://doi.org/10.1007/978-1-4684-5823-7\\_33](https://doi.org/10.1007/978-1-4684-5823-7_33).
15. Narayanan K, Kim KH, Makino S. Characterization of N protein self-association in coronavirus ribonucleoprotein complexes. *Virus Res.* 2003; 98(2):131–40. <https://doi.org/10.1016/j.virusres.2003.08.021>.
16. Chang C-K, Hsu Y-L, Chang Y-H, Chao F-A, Wu M-C, Huang Y-S, et al. Multiple nucleic acid binding sites and intrinsic disorder of severe acute respiratory syndrome coronavirus nucleocapsid protein: implications for ribonucleocapsid protein packaging. *J Virol.* 2009;83(5):2255–64. <https://doi.org/10.1128/JVI.02001-08>.
17. Kopecky-Bromberg SA, Martínez-Sobrido L, Frieman M, Baric RA, Palese P. Severe acute respiratory syndrome coronavirus open reading frame (ORF) 3b, ORF 6, and nucleocapsid proteins function as interferon antagonists. *J Virol.* 2007;81(2):548–57. <https://doi.org/10.1128/JVI.01782-06>.
18. Lu X, Pan J-A, Tao J, Guo D. SARS-CoV nucleocapsid protein antagonizes IFN- $\beta$  response by targeting initial step of IFN- $\beta$  induction pathway, and its C-terminal region is critical for the antagonism. *Virus Genes.* 2011;42(1):37–45. <https://doi.org/10.1007/s11262-010-0544-x>.
19. Simons K, Garoff H. The budding mechanisms of enveloped animal viruses. *J Gen Virol.* 1980;50(1). <https://doi.org/10.1099/0022-1317-50-1-1>.
20. Escors D, Ortego J, Laude H, Enjuanes L. The membrane M protein carboxy terminus binds to transmissible gastroenteritis coronavirus core and contributes to core stability. *J Virol.* 2001;75(3):1312–24. <https://doi.org/10.1128/JVI.75.3.1312-1324.2001>.
21. Kuo L, Masters PS. Genetic evidence for a structural interaction between the carboxy termini of the membrane and nucleocapsid proteins of mouse hepatitis virus. *J Virol.* 2002;76(10):4987–99. <https://doi.org/10.1128/jvi.76.10.4987-4999.2002>.
22. Hurst KR, Koetzner CA, Masters PS. Characterization of a critical interaction between the coronavirus nucleocapsid protein and nonstructural protein 3 of the viral replicase-transcriptase complex. *J Virol.* 2013;87(16):9159–72. <https://doi.org/10.1128/JVI.01275-13>.
23. Lei J, Kusov Y, Hilgenfeld R. Nsp3 of coronaviruses: structures and functions of a large multi-domain protein. *Antivir Res.* 2018;149:58–74. <https://doi.org/10.1016/j.antiviral.2017.11.001>.
24. Zhou B, Liu J, Wang Q, Liu X, Li X, Li P, et al. The nucleocapsid protein of severe acute respiratory syndrome coronavirus inhibits cell cytokinesis and proliferation by interacting with translation elongation factor 1 $\alpha$ . *J Virol.* 2008;82(14):6962–71. <https://doi.org/10.1128/JVI.00133-08>.
25. Surjit M, Liu B, Chow VTK, Lal SK. 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.* 2006;281(16):10669–81. <https://doi.org/10.1074/jbc.M509233200>.
26. Gordon DE, Jang GM, Bouhaddou M, Xu J, Obernier K, White KM, et al. A SARS-CoV-2 protein interaction map reveals targets for drug repurposing. *Nature.* 2020. <https://doi.org/10.1038/s41586-020-2286-9>.
27. Chang C-K, Sue S-C, Yu T-H, Hsieh C-M, Tsai C-K, Chiang Y-C, et al. Modular organization of SARS coronavirus nucleocapsid protein. *J Biomed Sci.* 2006; 13(1):59–72. <https://doi.org/10.1007/s11373-005-9035-9>.
28. Huang Q, Yu L, Petros AM, Gunasekera A, Liu Z, Xu N, et al. Structure of the N-terminal RNA-binding domain of the SARS CoV nucleocapsid protein. *Biochemistry.* 2004;43(20):6059–63. <https://doi.org/10.1021/bi036155b>.
29. Chen C-Y, Chang C-K, Chang Y-W, Sue S-C, Bai H-I, Riang L, et al. Structure of the SARS coronavirus nucleocapsid protein RNA-binding dimerization domain suggests a mechanism for helical packaging of viral RNA. *J Mol Biol.* 2007;368(4):1075–86. <https://doi.org/10.1016/j.jmb.2007.02.069>.
30. Takeda M, Chang C-K, Ikeya T, Güntert P, Chang Y-H, Hsu Y-L, et al. Solution structure of the c-terminal dimerization domain of SARS coronavirus nucleocapsid protein solved by the SAIL-NMR method. *J Mol Biol.* 2008; 380(4):608–22. <https://doi.org/10.1016/j.jmb.2007.11.093>.
31. Peng T-Y, Lee K-R, Tarn W-Y. Phosphorylation of the arginine/serine dipeptide-rich motif of the severe acute respiratory syndrome coronavirus nucleocapsid protein modulates its multimerization, translation inhibitory activity and cellular localization. *FEBS J.* 2008;275(16):4152–63. <https://doi.org/10.1111/j.1742-4658.2008.06564.x>.
32. Luo H, Chen J, Chen K, Shen X, Jiang H. Carboxyl terminus of severe acute respiratory syndrome coronavirus nucleocapsid protein: self-association analysis and nucleic acid binding characterization. *Biochemistry.* 2006;45(39): 11827–35. <https://doi.org/10.1021/bi0609319>.
33. Yu IM, Oldham ML, Zhang J, Chen J. Crystal structure of the severe acute respiratory syndrome (SARS) coronavirus nucleocapsid protein dimerization domain reveals evolutionary linkage between corona- and arterivirus. *J Biol Chem.* 2006;281(25):17134–9. <https://doi.org/10.1074/jbc.M602107200>.
34. Jayaram H, Fan H, Bowman BR, Ooi A, Jayaram J, Collisson EW, et al. X-ray structures of the N- and C-terminal domains of a coronavirus nucleocapsid protein: implications for nucleocapsid formation. *J Virol.* 2006;80(13):6612–20. <https://doi.org/10.1128/JVI.00157-06>.
35. Ma Y, Tong X, Xu X, Li X, Lou Z, Rao Z. Structures of the N- and C-terminal domains of MHV-A59 nucleocapsid protein corroborate a conserved RNA-protein binding mechanism in coronavirus. *Protein Cell.* 2010;1(7):688–97. <https://doi.org/10.1007/s13238-010-0079-x>.
36. Szlezacek B, Kabala W, Kus K, Zdzalik M, Twarda-Clapa A, Golik P, et al. Structural characterization of human coronavirus NL63 N protein. *J Virol.* 2017;91(11):e02503–16. <https://doi.org/10.1128/JVI.02503-16>.
37. Nguyen THV, Lichière J, Canard B, Papageorgiou N, Attoumani S, Ferron F, et al. Structure and oligomerization state of the C-terminal region of the Middle East respiratory syndrome coronavirus nucleoprotein. *Acta Crystallogr D Struct Biol.* 2019;75(Pt 1):8–15. <https://doi.org/10.1107/S2059798318014948>.
38. Ye Q, West AMV, Silletti S, Corbett KD. Architecture and self-assembly of the SARS-CoV-2 nucleocapsid protein. *bioRxiv.* 2020. <https://doi.org/10.1101/2020.05.17.100685>.
39. Chen VB, Arendall WB, Headd JJ, Keedy DA, Immormino RM, Kapral GJ, et al. MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr D Biol Crystallogr.* 2010;66(Pt 1):12–21. <https://doi.org/10.1107/S0907444909042073>.
40. Holm L. Benchmarking fold detection by DaliLite v.5. *Bioinformatics.* 2019; 35(24):5326–7. <https://doi.org/10.1093/bioinformatics/btz536>.
41. Kabsch W, Sander C. Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers.* 1983;22(12):2577–637. <https://doi.org/10.1002/bip.360221211>.
42. Krissinel E, Henrick K. Inference of macromolecular assemblies from crystalline state. *J Mol Biol.* 2007;372(3):774–97. <https://doi.org/10.1016/j.jmb.2007.05.022>.
43. Gouet P, Courcelle E, Stuart DI, Métoz F. ESPript: analysis of multiple sequence alignments in PostScript. *Bioinformatics.* 1999;15(4):305–8. <https://doi.org/10.1093/bioinformatics/15.4.305>.
44. Lo Y-S, Lin S-Y, Wang S-M, Wang C-T, Chiu Y-L, Huang T-H, et al. Oligomerization of the carboxyl terminal domain of the human coronavirus 229E nucleocapsid protein. *FEBS Lett.* 2013;587(2):120–7. <https://doi.org/10.1016/j.febslet.2012.11.016>.
45. Yang D, Leibowitz JL. The structure and functions of coronavirus genomic 3' and 5' ends. *Virus Res.* 2015;206:120–33. <https://doi.org/10.1016/j.virusres.2015.02.025>.
46. Kabsch W. XDS. *Acta Crystallogr D Biol Crystallogr.* 2010;66(Pt 2):125–32. <https://doi.org/10.1107/S0907444909047337>.
47. Evans PR, Murshudov GN. How good are my data and what is the resolution? *Acta Crystallogr D Biol Crystallogr.* 2013;69(Pt 7):1204–14. <https://doi.org/10.1107/S0907444913000061>.
48. Adams PD, Afonine PV, Bunkóczi G, Chen VB, Davis IW, Echols N, et al. PHENIX: a comprehensive python-based system for macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr.* 2010;66(Pt 2):213–21. <https://doi.org/10.1107/S0907444909052925>.
49. Emsley P, Lohkamp B, Scott WG, Cowtan K. Features and development of coot. *Acta Crystallogr D Biol Crystallogr.* 2010;66(Pt 4):486–501. <https://doi.org/10.1107/S0907444910007493>.
50. Murshudov GN, Vagin AA, Dodson EJ. Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr D Biol Crystallogr.* 1997;53(Pt 3):240–55. <https://doi.org/10.1107/S0907444996012255>.

## Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.