High-resolution structure and biophysical characterization of the nucleocapsid phosphoprotein dimerization domain from the Covid-19 severe acute respiratory syndrome coronavirus 2

Luca Zinzula [a, *], Jerome Basquin [b], Stefan Bohn [a, c], Florian Beck [a], Sven Klumpe [a], Günter Pfeifer [a], István Nagy [a], Andreas Bracher [d], F. Ulrich Hartl [d], Wolfgang Baumeister [a]

[a] The Max-Planck Institute of Biochemistry, Department of Molecular Structural Biology, Am Klopferspitz 18, 82152, Martinsried, Germany
[b] The Max-Planck Institute of Biochemistry, Department of Structural Cell Biology, Am Klopferspitz 18, 82152, Martinsried, Germany
[c] The Max-Planck Institute of Biochemistry, Department of Cellular Machines and Signaling, Am Klopferspitz 18, 82152, Martinsried, Germany
[d] The Max-Planck Institute of Biochemistry, Department of Cellular Biochemistry, Am Klopferspitz 18, 82152, Martinsried, Germany

ARTICLE INFO

ABSTRACT

Unprecedented by number of casualties and socio-economic burden occurring worldwide, the coronavirus disease 2019 (Covid-19) pandemic caused by the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) is the worst health crisis of this century. In order to develop adequate countermeasures against Covid-19, identification and structural characterization of suitable antiviral targets within the SARS-CoV-2 protein repertoire is urgently needed. The nucleocapsid phosphoprotein (N) is a multifunctional and highly immunogenic determinant of virulence and pathogenicity, whose main functions consist in oligomerizing and packaging the single-stranded RNA (ssRNA) viral genome. Here we report the structural and biophysical characterization of the SARS-CoV-2 N C-terminal domain (CTD), on which both N homo-oligomerization and ssRNA binding depend. Crystal structures solved at 1.44 Å and 1.36 Å resolution describe a rhombus-shape N CTD dimer, which stably exists in solution as validated by size-exclusion chromatography coupled to multi-angle light scattering and analytical ultracentrifugation. Differential scanning fluorimetry revealed moderate thermal stability and a tendency towards conformational change. Microscale thermophoresis demonstrated binding to a 7-bp SARS-CoV-2 genomic ssRNA fragment at micromolar affinity. Furthermore, a low-resolution preliminary model of the full-length SARS-CoV N in complex with ssRNA, obtained by cryo-electron microscopy, provides an initial understanding of self-associating and RNA binding functions exerted by the SARS-CoV-2 N.

© 2020 Elsevier Inc. All rights reserved.

1. Introduction

In December 2019, a cluster of pneumonia cases was reported in Wuhan, capital of the Hubei province of China, showing clinical manifestations that resembled those observed during severe acute respiratory syndrome coronavirus (SARS-CoV) and Middle East respiratory syndrome (MERS)-CoV infections [1,2]. The etiologic agent responsible for the outbreak was soon proven capable of human-to-human transmission [3,4] and identified as a new member of the genus *Betacoronavirus* in the family *Coronaviridae* of the order *Nidovirales* (realm *Riboviria*) [5,6]. The virus was named

SARS-CoV-2 because of its genetic relatedness to SARS-CoV, MERS-CoV and other SARS-related bat coronaviruses (CoVs), whereas the associated respiratory illness was termed as CoV disease 2019 (Covid-19) [7,8]. Since then, Covid-19 spread worldwide and has turned into a pandemic that at present caused more than twenty-three million infected humans and more than eight hundred thousand deaths [9]. This global health threat has created an urgency for development of therapeutic countermeasures, triggering a race to obtain a Covid-19 vaccine and to evaluate the effectiveness and safety of several antiviral drug candidates [10,11]. In turn, this race has made understanding the SARS-CoV-2 life cycle at the molecular level and solving the structures of its protein repertoire two needs of utmost importance. Among the four structural proteins and the (at least) sixteen non-structural ones encoded by the ≈ 30 kb positive-sense, single-stranded RNA (ssRNA) genome of

* Corresponding authors.
E-mail addresses: zinzula@biochem.mpg.de (L. Zinzula), baumeist@biochem.mpg.de (W. Baumeister).
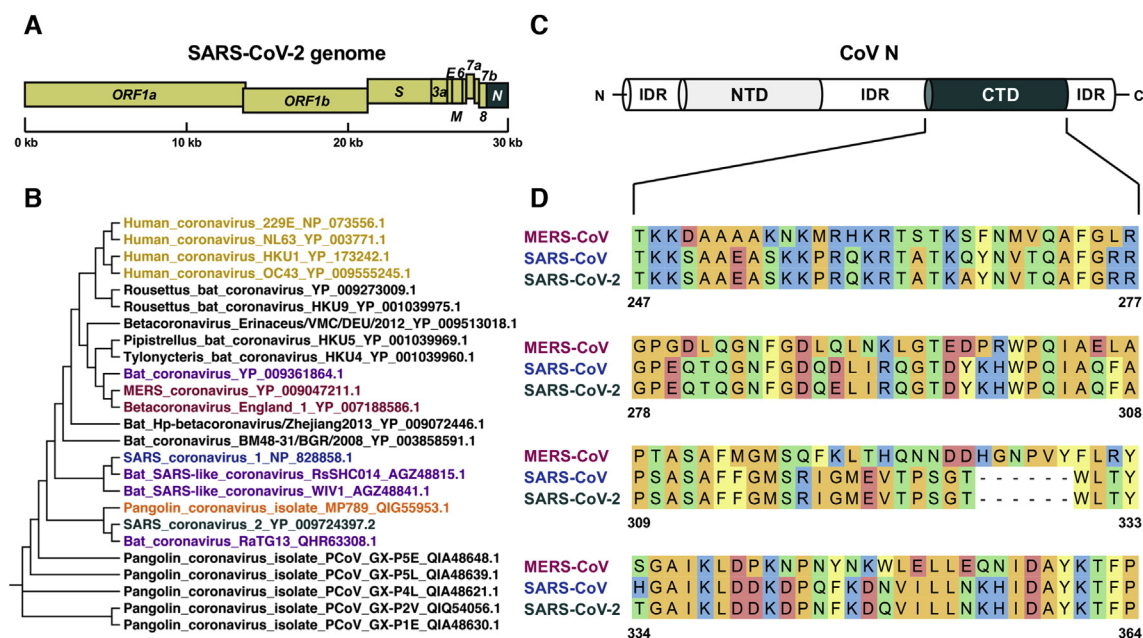
Fig. 1. CTD boundary delineation in the SARS-CoV-2 N. (A) Schematic organization of the SARS-CoV-2 genome; the N gene is highlighted in dark green. (B) Molecular evolutionary analysis of CoV N proteins from SARS-related members of the *Betacoronavirus* genus; human CoVs were included in the maximum-likelihood phylogenetic tree as non-SARS-related outgroup. (C) Structural organization of CoV N protein; the CTD is highlighted in dark green. (D) Amino acid sequence alignment of the solved MERS-CoV (PDB: 6G13) and SARS-CoV (PDB: 2CJR) to the putative SARS-CoV-2 N CTD.

SARS-related CoVs, the nucleocapsid (N) phosphoprotein is a particularly attractive antiviral target [5,6]. In fact, not only is this protein fundamental for the viral RNA genome packaging into a ribo-nucleocapsid (RNP) complex and for its assembly into the viral particle [12], but N is also the most abundant protein in the virion, a highly immunogenic antigen [14,15] and a determinant of virulence and pathogenesis [16,17]. N residues number and molecular weight vary among coronaviral species between 342 and 468 amino acids and 37.7–51.5 kDa, respectively. In order to exert its multifunctional properties, CoV N adopts a modular organization where two globular, independently folded units, namely the N-terminal domain (NTD) and the C-terminal domain (CTD), are respectively preceded, separated and followed by flexible and intrinsically disordered regions (IDRs) [12]. Both the NTD and the CTD participate in the packaging of RNA viral genome, whereas the IDRs are involved in modulating the RNA binding activity. Furthermore, assembly into a functional nucleocapsid is a result of the homo-oligomerization properties of its CTD [13]. Here we report crystal structures of the SARS-CoV-2 N CTD solved at high resolution (1.44 Å and 1.36 Å) and provide a biophysical characterization of its functional properties regarding homo-oligomerization and RNA binding. Also, we report a preliminary low-resolution 3D volume obtained by cryo-electron micrococpy (cryo-EM) of the full-length SARS-CoV-2 N in complex with ssRNA.

2. Material and methods

2.1. Cloning, expression, purification and validation

The cDNA encoding for the SARS-CoV-2 N (GenBank: NC_045512.2) was obtained by synthetic preparation (BioCat) and cloned into a pET41b (Novagen) plasmid vector between the NdeI and XhoI restriction sites. Boundaries of the CTD (residues 247–364) were assigned after comparative bioinformatic analysis and the resulting construct was cloned into a pRSF-Duet (Novagen) plasmid vector between the SacI and AvrII restriction sites.

Recombinant hexahistidine (His 6) - tagged SARS-CoV-2 N and N CTD proteins were expressed in *E. coli* BL21-DE3 (New England Biolabs) cells grown in LB medium supplemented with 50 mg mL-1 Kanamycin at 37 °C to an OD 600 nm of 0.8, upon induction with 0.65 mM isopropyl-b-D-1-thiogalactopyranoside overnight at 24 °C. Harvested cells were lysed in buffer A (50 mM sodium phosphate, pH 8.0; 500 mM NaCl; 10% (v/v) glycerol; 15 mM imidazole) supplemented with 1 mg mL-1 Lysozyme, cOmplete EDTA-free Protease Inhibitor Cocktail (Roche), ~ 2000 Units Endonuclease from *S. marcescens*, then sonicated and centrifuged at 30,000 g at 4 °C for 30 min. The supernatant was loaded on a 1 mL HisTrap FF crude (GE Healthcare) column equilibrated in buffer A for affinity chromatography purification. Washing and elution of bound protein were performed with buffer A containing 30 mM and 600 mM imidazole, respectively. Subsequent size-exclusion chromatography (SEC) step was performed on a Superose 12 10/300 GL (GE Healthcare) column in buffer B (10 mM HEPES, pH 7.2; 150 mM NaCl). Purity and homogeneity were assessed by 4–12% NuPAGE SDS-PAGE (Thermo-Fisher). Identity and integrity of purified proteins were confirmed by WB and LC/MS (data not shown).

2.2. Miniaturized differential scanning fluorimetry (Nano-DSF)

Thermal unfolding profiles were acquired by measuring the temperature-dependent shift in protein intrinsic fluorescence at emission wavelengths of 330 and 350 nm using a Prometheus NT.48 Nano-DSF instrument (NanoTemper Technologies). SARS-CoV-2 N CTD was diluted to 3.0 mg mL^{-1} in buffer B, loaded on standard capillaries and subjected to a linear 20–95 °C thermal gradient at 0.5 °C min^{-1} rate. The inflection point fluorescence transition corresponding to the melting temperature (Tm) value was determined as the first derivative maximum of the fluorescence intensities ratio at the measured wavelengths (F330/F350). Data from three independent measurements were processed using the PR. ThermControl software (NanoTemper Technologies).
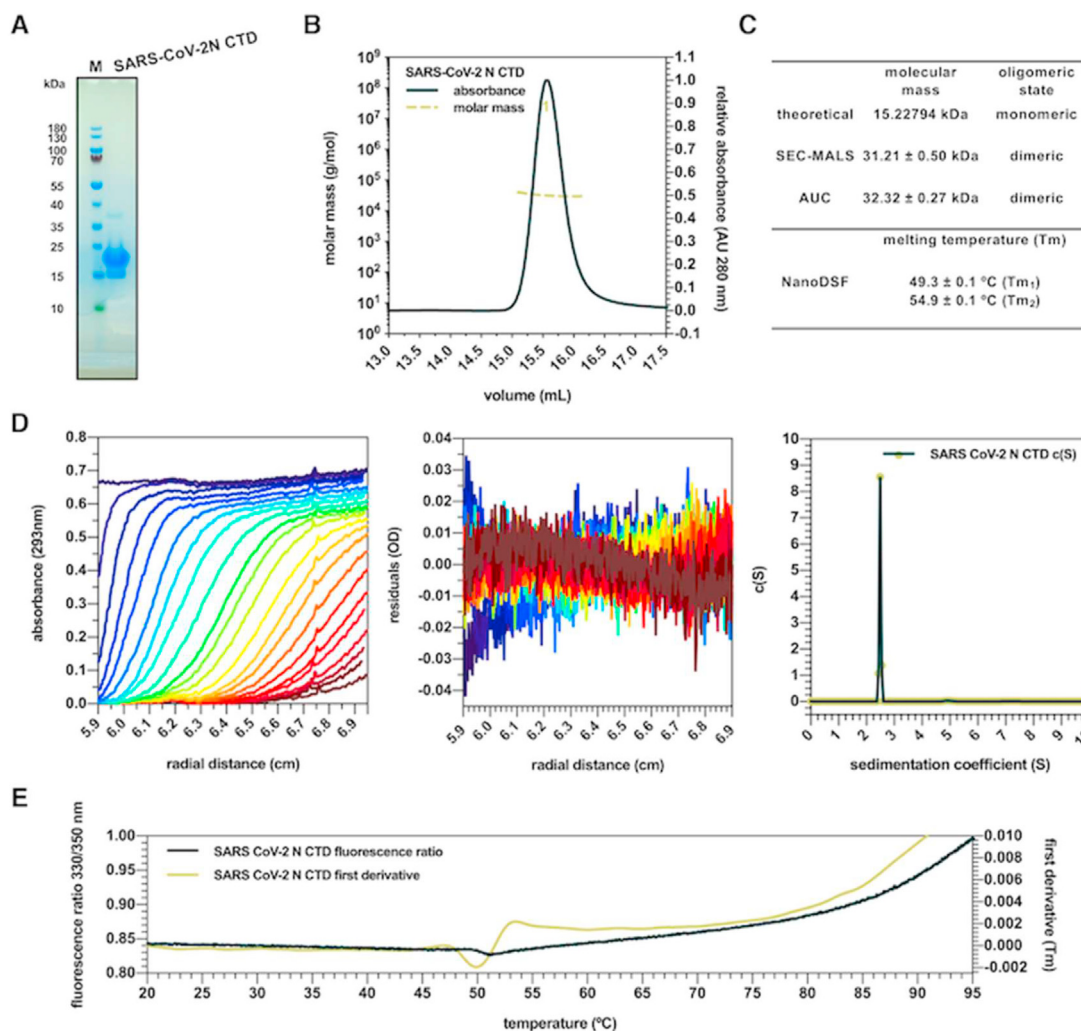
Fig. 2. Biophysical characterization of the SARS-CoV-2 N CTD. (A) SDS-PAGE analysis showing the migration of the SARS-CoV-2 N CTD as a monomeric band (M, molecular weight marker). (B) SEC-MALS analysis showing elution of the SARS-CoV-2 N CTD as a dimer. (C). Summary of the homo-oligomeric in solution profile and thermal stability values for the SARS-CoV-2 N CTD. (D) Raw data, residuals of the fits and sedimentation coefficient distribution of AUC analysis showing sedimentation of the SARS-CoV-2 N CTD as a dimer. (E) Raw data (dark green) and first derivative of 330/350 nm fluorescence ratio (light green) of Nano-DSF analysis showing inflection points and Tm of the SARS-CoV-2 N CTD as two conformational states.

2.3. SEC - multi-angle light scattering (SEC-MALS)

Purified SARS-CoV-2 N CTD (0.3 mg) was loaded on a Superdex 200 10/300 GL (GE Healthcare) column connected to a HPLC (Agilent Technologies, 1100 series) system with variable UV absorbance detector set at 280 nm, coupled in line with a mini DAWN TREOS MALS detector followed by an Optilab rEX refractive-index detector (Wyatt Technology, 690 nm laser). Analysis was performed at 20 °C and 0.75 mL min^{-1} flow rate in buffer B. Bovine serum albumin was used as calibration standard. Absolute molecular mass was calculated with ASTRA 6 software (Wyatt Technology) with the dn/dc value set to 0.185 mL/g.

2.4. Analytical ultracentrifugation (AUC)

Purified SARS-CoV-2 N CTD (1 and 5 mg mL^{-1} in buffer B) was subjected to sedimentation velocity analysis on an Optima XL-I analytical centrifuge (Beckman-Coulter) using an 60 Ti rotor and double-sector 12 mm centerpieces. Buffer density was measured to 1.0052 kg/L using a DMA 5000 densitometer (Anton Paar). Protein concentration distribution was monitored at 293 nm, using

50.000 r.p.m. rotor speed. Time-derivative analysis was computed using the SEDFIT software package [18], resulting in a c(s) distribution and an estimate for the molecular weight from sedimentation and diffusion coefficients, inferred from the peak width.

2.5. Crystallization and data collection

The 60 mg mL^{-1} SARS-CoV-2 N CTD was used to set up initial screens using the sitting-drop vapor-diffusion method at 4 °C. Two initial crystal forms were found from precipitant solution containing 31% PEG 4 K, 0.2 M Lithium sulfate, 50 mM Tris pH 7.8 and 34% ethanol, 5% PEG 1 K, 0.1 M Tri-sodium citrate pH 4.5, respectively. Crystals were fished using nylon loops, soaked in mother liquor containing 30% (v/v) ethylene glycol, and flash-frozen in liquid nitrogen. Diffraction data were collected on the PXIII (06 DA) beamline at the Swiss Light Source and processed using software in the XDS suite [19].

2.6. Structure determination and refinement

Diffraction data were integrated and scaled with XDS [19]. The

Table 1
X-ray crystallography data collection and refinement statistics.

	Crystal form I	Crystal form II
PDB ID	6YUN	6ZCO
Wavelength (Å)	1.000	1.000
Resolution range (Å)	41.25−1.445 (1.496−1.445)[a]	44.10−1.36 (1.43−1.36)
Space group	$P\,2_1\,2_1\,2_1$	$I\,4_1$
Unit cell (Å, °)	43.43 46.86,131.95 90 90 90	88.19 88.19 42.76 90 90 90
Total reflections	411,001 (28,010)	264,134 (15,504)
Unique reflections	48,359 (4160)	34,588 (4414)
Multiplicity	8.5 (6.7)	7.6 (3.5)
Completeness (%)	98.22 (85.74)	97.8 (86.6)
Mean I/sigma(I)	21.16 (2.18)	20.2 (0.9)
Wilson B-factor	12	19
R-merge	0.1265 (0.9028)	0.041 (1.093)
R-meas	0.1344 (0.9751)	0.047 (1.446)
R-pim	0.0448 (0.3599)	0.016 (0.698)
CC1/2	0.999 (0.786)	1.000 (0.593)
Reflections used in refinement	48,329 (4160)	32,473 (1794)
Reflections used for R-free	2415 (183)	1709 (97)
R-work	0.1476 (0.2399)	0.1495 (0.50)
R-free	0.1891 (0.2912)	0.1962 (0.49)
CC(work)	0.969 (0.897)	0.981 (0.638)
CC(free)	0.951 (0.812)	0.968 (0.717)
Number of non-hydrogen atoms	2382	1119
macromolecules	1990	966
solvent	392	153
Protein residues	251	118
RMS(bonds) (Å)	0.011	0.016
RMS(angles) (°)	1.10	1.65
Ramachandran favored (%)	98.8	99.2
Ramachandran allowed (%)	1.2	0.8
Ramachandran outliers (%)	0.0	0.0
Rotamer outliers (%)	0.0	2.0
Clashscore	3.58	0.52
Average B-factor (Å^2)	17.9	28.9
Macromolecules (Å^2)	15.3	27.6
Solvent (Å^2)	31.1	41.5

[a] Statistics for the highest-resolution shell are shown in parentheses.

structure of crystal form I was solved by molecular replacement using Phaser [20] with the coordinates of the crystal structure of the SARS-CoV N CTD (PDB: 2JCR) [21] as a search model. Interactive model building was done with Coot [22]. The final restrained refinement cycles were done with Phenix [23] and Refmac5 [24].

2.7. Micro-scale thermophoresis (MST) RNA binding assay

Purified SARS-CoV-2 N CTD and a synthetic, fluorophore-labeled ssRNA oligomer of 7 bp in length resembling the SARS-CoV-2 genome initiation sequence (5′-Cy5-AUUAAAG-3′, Metabion) were tested for interaction, both diluted in MST buffer (buffer B supplemented with 0.05% Tween-20) at the final concentrations of 2.4−80 μM and 40 nM, respectively. After 60 min incubation at RT, reaction samples were loaded into standard capillaries for measurement at 22 °C temperature, 20% LED-power and high MST-power using a Monolith NT.115 instrument (NanoTemper Technologies). Data from three independent measurements of the signal corresponding to 1.5 s MST-On time were analyzed using the MO. Affinity Analysis software (NanoTemper Technologies).

## 2.8. Cryo-EM specimen preparation, data collection and image processing

Purified full-length SARS-CoV-2 N (4.5 μL; 0.6 mg mL^{-1}) was applied to glow-discharged Quantifoil 1.2/1.3 grids, blotted for 3.5 s with force 4 at 100% humidity and 4 °C, and plunge-frozen in liquid ethane cooled by liquid nitrogen in a Vitrobot Mark III (Thermo Fisher). Data were acquired at 0° and 30° tilt angle in a Titan Krios

transmission electron microscope (Thermo Fisher) operated by using Latitude in Digital Micrograph (Gatan) and SerialEM [25], respectively. Movie frames were recorded at a nominal magnification of 22,500 X using a K3 direct electron detector (Gatan). An electron dose of ~55 e$^-$/Å^2 was distributed over 30 frames at a calibrated physical pixel size of 1.09 Å, and a defocus range of −0.5 to −3.0 μm was applied. Micrographs were processed on-the-fly using Focus, if they passed the selection criteria (iciness < 1.05, drift 0.4 Å < x < 70 Å, defocus 0.5 μm < x < 5.5 μm, estimated CTF resolution < 5 Å) [26]. Frames were aligned using MotionCor2 [27] and the contrast transfer function (CTF) for aligned frames was determined using GCTF [28]. Particles (11, 453, 521 and 1,108,479 from 8657 to 955 micrographs acquired at 0° and 30° tilt, respectively) were template-free picked using Gautomatch [29] and extracted with a 192 rescaled to 96 pixels box size using RELION 3.1 [30]. Reference-free, 2D-classified particles (1,536,618 and 450,685 at 0 and 30° tilt, respectively) were re-extracted with a 192 pixels box size, re-centred and imported into Cryosparc 2.14 [31] for initial model ab initio reconstruction and then subjected to multiple rounds of 3D-classification. Selected ones (1,085,151 particles) were refined by applying C2-symmetry and optimizing CTF-parameters.

3. Results

3.1. Domain boundaries assignment and purification of the SARS-CoV-2 N CTD

SARS-CoV-2 N is encoded by the last open reading frame along
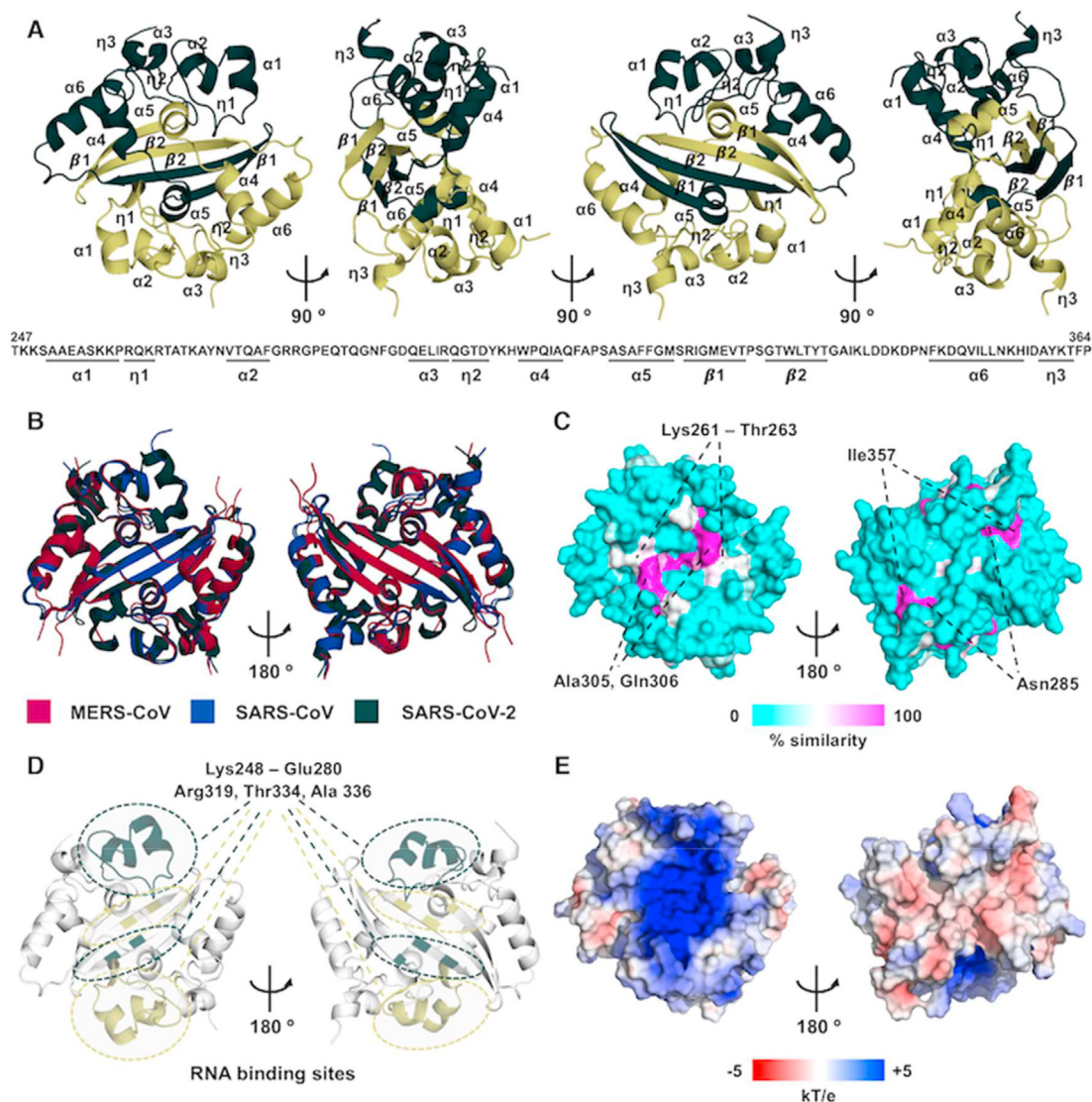
Fig. 3. Structure of the SARS-CoV-2 N CTD. (A) Ribbon diagram and topological organization of the SARS-CoV-2 N CTD dimer, shown in four orientations; the two CTD monomers are depicted in dark and light green, respectively. (B) Superimposition of MERS-CoV (PDB: 6G13, purple) and SARS-CoV (PDB: 2CJR, blue) to the SARS-CoV-2 N CTD (dark green; r.m.s.d.s. of 0.4—1.3 Å for matching Cα atoms). (C) Surface residue conservation profile of the SARS-CoV-2 N CTD among CoVs (scale from 0 to 100% similarity score). (D) Putative RNA binding sites in the SARS-CoV-2 N CTD dimer; involved residues of each CTD monomer are highlighted in dark and light green, respectively. (E) Electrostatic surface potential analysis of the dimeric SARS-CoV-2 N CTD; red, white and blue regions represent areas with negative, neutral and positive electrostatic potential, respectively (scale from −5 to +5 kT e $^{-1}$).

the 30 kb viral genome (Fig. 1A). It is active in the formation of the CoV nucleocapsid and in providing a protective coat for the long viral genomic ssRNA. This implies N to be capable of binding the nucleic acid on one hand, and of self-associating into homo-oligomers on the other. According to the currently accepted model for CoV N, the NTD binds bases of ribonucleotide moieties and the CTD oligomerizes and coats the nucleic acid backbone [12,13]. Hence, in order to solve the structure of the CTD from the SARS-CoV-2 N and to characterize its biochemical properties in vitro, we first defined its exact boundaries within the 419 amino acid long sequence of the SARS-CoV-2 N. Whole genome sequence analysis of SARS-CoV-2 and previously identified members of the family *Coronaviridae*, groups the novel SARS-CoV-2 within the genus *Betacoronavirus* (subgenus *Sarbecovirus*), which includes the highly pathogenic SARS-CoV and MERS-CoV as well as other SARS-related coronaviral species isolated mainly from bats [6]. Moreover, due to the tendency of these viruses to undergo genetic

recombination, the topological position of SARS-CoV-2 in the phylogenetic tree may change depending on which gene is considered for comparison [5]. In particular, analysis of the N gene shows that SARS-CoV-2 N has the highest amino acid sequence identity with that of bat-CoV RaTG13 (99.0%) and pangolin-CoV MP789 (98.9% %) (Fig. 1B and Fig. S1). Within the same phylogenetic cluster, structures have been obtained for SARS-CoV and MERS-CoV N CTD [21,32—34], which - albeit displaying lower amino acid sequence identity - fully match at the domain boundaries with those predicted for SARS-CoV-2. Therefore, by combining information from the amino acid sequence alignments and the available crystallographic structures, we could assign the boundaries of the SARS-CoV-2 N CTD to residues Thr247 and Pro 374 (Fig. 1C and D). The corresponding construct was therefore cloned for bacterial expression as a N-terminally His-tagged protein, then purified to homogeneity for subsequent structural and biophysical investigation. In SDS-PAGE, SARS-CoV-2 N CTD migrated as one
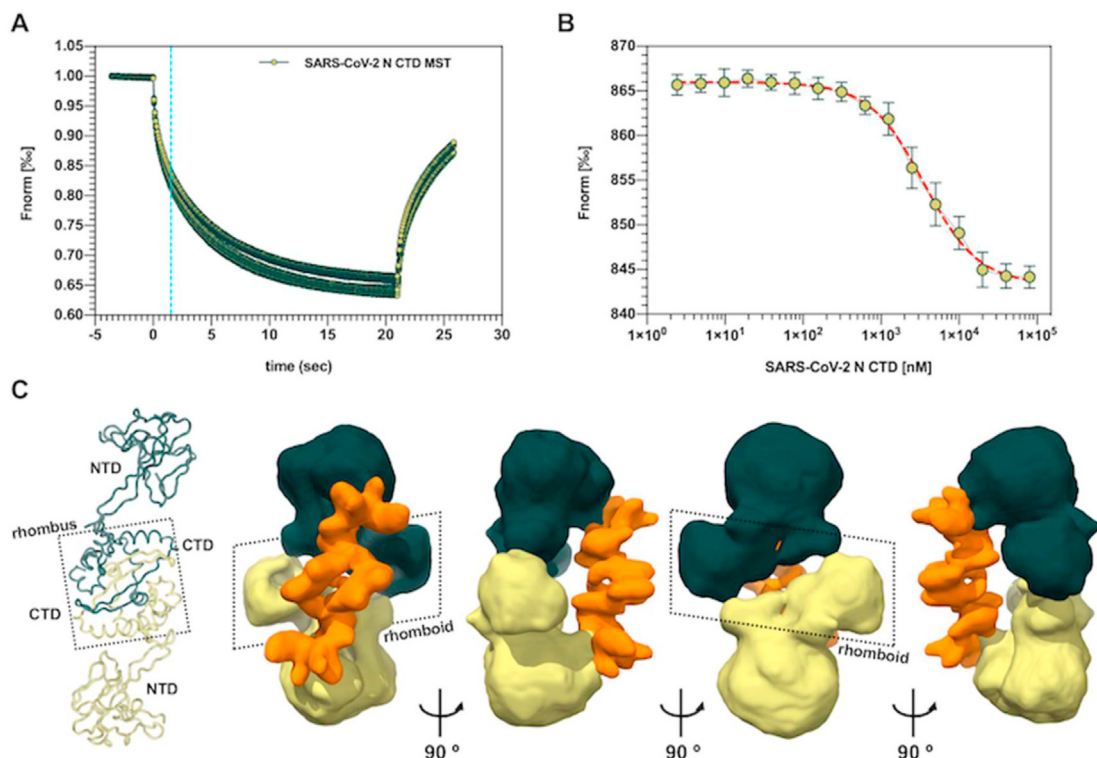
Fig. 4. Interaction of the SARS-CoV-2 N CTD with an RNA heptamer. (A) Raw data of MST analysis showing binding of the SARS-CoV-2 N CTD to a Cy-5-labeled 7 bp ssRNA (B) Curve fit of MST traces for determination of the SARS-CoV-2 N CTD-ssRNA binding affinity ($K_d \approx 4$ μM). (C) Current model for the SARS-CoV N structural organization based on reported NMR and SAXS data (left), and Cryo-EM 3D volume of the SARS-CoV-2 N complexed to a 7 bp ssRNA, shown in four orientations; the two N monomers and the putative ssRNA densities are depicted in dark green, light green and orange, respectively.

major band at an apparent molecular weight consistent with its expected monomeric mass of 15.2 kDa (Fig. 2A).

3.2. SARS-CoV-2 N CTD oligomerization profile and thermal stability

Previous structural studies on the isolated N CTD from CoV species such as the avian infectious bronchitis virus (IBV), the mouse hepatitis virus A59 (MHV-A59), the human CoV NL63 (hCoV-NL63), the SARS-CoV and the MERS-CoV, reported the self-association of this domain into a stable dimer [21,32–37]. Moreover, crystal packing analysis suggested that SARS-CoV N CTD dimers are able to form octameric superhelical complexes, and transient interactions between CTD dimers leading to the formation of higher-order oligomers were reported to occur in solution, depending on protein concentration [21,38]. Furthermore, a work published during the preparation of this manuscript described full-length SARS-CoV-2 N in either dimeric and oligomeric state, as detected by static light scattering and chemical cross-linking [39]. We therefore assessed the oligomerization profile in solution of our SARS-CoV-2 N CTD construct by SEC-MALS and AUC orthogonal analysis. As shown in the SEC-MALS chromatogram, SARS-CoV-2 N CTD eluted as a single, uniform peak with an apparent molecular mass of 31.2 ± 0.5 kDa (Fig. 2B and C). Similarly, the sedimentation curves obtained by AUC suggested presence of a single species with an apparent molecular mass of 32.3 ± 0.3 kDa (Fig. 2D and C). With both techniques, only this dimeric protein species was detected at every concentration tested (Fig. 2C). Next, we assessed the thermal stability of the SARS-CoV N CTD dimer by NanoDSF. Early studies described the full-length N protein of SARS-CoV as thermally unstable, starting to unfold at 37 °C and denaturing at 55 °C, whereas

for the protein from hCoV-OC43 a melting temperature (Tm) value of 55 °C was reported [40,41]. By monitoring the SARS-CoV-2 N CTD intrinsic fluorescence during a 20–95 °C thermal gradient, we observed two transitions, corresponding to Tm values of 49.3 ± 0.1 °C and 54.9 ± 0.1 °C, respectively (Fig. 2E and C). Taken together, these data show that the SARS-CoV-2 N CTD exists in a stable dimeric state in solution and suggest that the overall stability of the protein depends on the self-association of this domain.

3.3. Crystal structure of the SARS-CoV-2 N CTD

The SARS-CoV-2 N CTD crystallized in two different crystal forms belonging to the $P2_12_12_1$ and $I4_1$ space groups, from which we determined high resolution structures at 1.44 Å (form I, PDB: 6YUN) and 1.36 Å (form II, PDB: 6ZCO), respectively (Table 1). Both structures consist of a dimeric, rhombus-shaped tile made up by the association of two CTD monomers, each comprised of six α-helices (α1- α6), two 3_{10}-helices (η_1, η_2) and two β-strands (β_1, β_2) in a $\alpha_1\eta_1\alpha_2\alpha_3\eta_2\alpha_4\alpha_5\beta_1\beta_2\alpha_6\eta_3$ topology (Fig. 3A and S2). The three crystallographically independent subunits in the two structures diverge at the N-terminus until residue Ala 252. In one copy, the N-terminal His 6-tag is resolved (Fig. S3A). The core structures, encompassing residues Ala252–Phe363, are highly similar with r.m.s.d.s of approximately 0.3 Å (Fig. S3B). Dimerization takes place via domain swapping of the two β_1-β_2 hairpins, resulting in a central four-stranded β-sheet, which is further stabilized at its edges by the α6 helix of the opposite monomer, and in the middle by the interaction between the α4 of one CTD and the α5 of the opposite one (Fig. 3A). Overall, our structures resemble those of previously solved analogues from other CoVs, with r.m.s.d.s of 0.4–1.3 Å for the matching Cα atoms (Fig. 3B) [21,32–37]. The

sequence differs from that of the closely-related SARS-CoV in only five positions, namely 267, 290, 334, 345 and 349 (Fig. 3B and S2), whereas surface conservation among CoVs is found at positions 261–263, 285, 305, 306 and 357 (Fig. 3C). Based on this near-identity, the putative RNA binding site that mapped to residues 248–280 in the SARS-CoV N CTD [21,33] would be conserved in SARS-CoV-2 N. Similarly, residues identified as the secondary RNA binding site in the SARS-CoV N CTD [33] correspond to Arg 319, Thr334 and Ala 336 along the β_1-β_2 hairpin (Fig. 3D). Reflecting these locations, analysis of the surface electrostatic potential shows that positively charged residues belonging to these amino acid tracts all cluster at opposite edges of a basic groove running transversely with respect to the dimeric interface line (Fig. 3E).

3.4. SARS-CoV-2 N CTD RNA binding

Previous studies revealed that interactions with nucleic acids occur at regions both in the NTD and CTD of SARS-CoV N, and that the connecting IDRs play a role in enhancing binding affinity [42–44]. The SARS-CoV N CTD is able to bind surrogates of its natural genomic substrate such as ssRNA, ssDNA, dsRNA or dsDNA, and displays micromolar binding affinities for nucleic acid moieties as short as 10 bp irrespective of the sequence [21,33,44]. Moreover, according to the currently proposed model for the CoV RNP complex, N CTD protomers pack into a helical core around which genomic ssRNA twists, so that each CTD accommodates in its positively charged groove a ssRNA tract of seven bases [12]. With this model in mind, we assessed the RNA binding properties of the SARS-CoV-2 N CTD by means of MST, using a 7 bp-long Cy5-fluorolabeled ssRNA probe whose sequence corresponds to the initiation of the SARS-CoV-2 genome (5' − AUUAAAG − 3'). The SARS-CoV-2 N CTD was able to bind this probe with micromolar affinity ($K_d = 4.2 \pm 0.1$ μM), which is tighter than the interaction of the isolated N CTD from the closely related SARS-CoV with similar substrates (Fig. 4A and B). This result indicates that a short oligo-nucleotide of biologically relevant viral sequence can be efficiently bound by the CTD of SARS-CoV-2 N in the absence of other domains, and suggests that ssRNA molecules as short as 7 bp may well be cradled by the N CTD basic groove.

3.5. Proposed model for full length SARS-CoV-2 N bound to ssRNA

Small-angle X-ray scattering (SAXS) data have provided low resolution models for the SARS-CoV and recently also for SARS-CoV-2 full length N, showing that NTDs are flexibly tethered to CTD dimers in solution and do not stably interact with each other or with the CTDs, and that the protein is largely disordered at physiological temperature due to the dynamic extension of its IDRs [39,44]. While the isolated N CTD failed to co-crystallize with an unlabeled version of the 7 bp long ssRNA in our hands, we were able to obtain a preliminary cryo-EM single particle image dataset from full-length SARS-CoV-2 N in presence of this substrate, resulting in a low-resolution 3D reconstruction of the putative SARS-CoV-2 N-RNA complex (Fig. S4). The 2-fold-symmetric density map consists of a bow tie made by two P-shaped monomers interacting back-to-back. Interaction leads to the formation of a central rhomboid with an open cleft in the middle, from which globular densities axially depart to form the bow tie ends. Bound to them, a density reminiscent of two short RNA helices runs diagonally along the bow tie axis (Fig. 4C). Even though the NTD and CTD of N could not be unequivocally fitted into the density at the current resolution, a model where the CTDs form the central rhomboid and the NTDs make up the bow tie ends agrees with the modular organization suggested by previously reported SAXS data. It is worth noting that, if so, re-arrangement of the CTD dimer conformation

displacing the two β-hairpins apart must be considered in order to justify the hole seen in the middle of the rhomboid (Fig. 4C).

4. Discussion

We solved the structure of the CTD in SARS-CoV-2 N protein, a domain indispensable for the nucleocapsid formation and viral RNA packaging. The two crystal forms show a rhombus-shaped particle formed by two interlaced CTD monomers which, in accordance with the high amino acid sequence conservation among corona-viral N proteins, recapitulates the structural organization of previously reported N CTD dimers from SARS-CoV, MERS-CoV, MHV-A59, IBV and hCoV-NL63 [21,32–37]. This similarity suggests that strong selective pressure exists to maintain this domain organization and this - given the immunogenicity and importance for virulence and pathogenicity displayed by the CoV N - renders the CTD a very attractive target for diagnostic purposes as well as for vaccine development and antiviral drug design. The SEC-MALS chromatogram and the AUC sedimentogram both demonstrate the existence of stable SARS-CoV-2 N CTD dimers in solution, supporting the idea that the SARS-CoV-2 RNP complex likely forms upon self-association between dimeric N protomers. However, the twin transitions and relatively low Tm values observed in DSF also suggest that the N CTD dimer is more dynamic in solution and prone to undergo conformational changes with respect to the compactness shown by the crystal structures. Conformational dynamism of the N CTD dimer in aqueous environment has been suggested by NMR data for the same domain from SARS-CoV, which showed disordered and protruding N-termini from each CTD monomer and a more relaxed dimer interface due to significant perturbations at the β-sheet interlace [21]. Hence, since regions affected by disorder and perturbation are those involved in RNA binding, a triggering factor for the CTD to undergo extensive conformational remodeling could be the interaction with nucleic acid. Our MST experiments revealed that the SARS-CoV-2 N CTD binds to a ssRNA corresponding to the first seven ribonucleotides of the SARS-CoV-2 genome, showing micromolar affinity, which is higher than those reported for longer ssRNA or ssDNA. Therefore, even though the CoV N CTD are known for being non-specifically avid of any kind of nucleic acid moiety, this biologically relevant ssRNA may represent the length for a poly-ribonucleotide back-bone to be accommodated along the basic groove of a N CTD pro-tomer. In our preliminary cryo-EM 3D map of a full-length SARS-CoV-2 N sample pre-incubated with the same nucleic acid, densities attributable to two ssRNA strands are bound to a dimeric N. The bow tie-like shape with two globular volumes departing from a rhomboid tile in the center is compatible with reported SAXS-based models of two floating NTD that extend from a central CTD dimeric core [39,44]. However, given the empty space in the rhomboid, our SARS-CoV-2 N CTD crystal structures and our cryo-EM 3D model would reconcile only by admitting conformational changes at the CTD dimer interface. At the current resolution, limited by strong preferential orientation of the single particles in the cryo-EM image dataset, further speculation is not possible. Moreover, since the proposed 3D model originates from a 2D classification sub-dataset from which non-averageable branched and heterogeneous oligo-meric forms were excluded, it may represent only a nucleocapsid building block rather than being descriptive of any genome pack-aging function. During the preparation of this manuscript, four additional crystal structures of the SARS-CoV-2 CTD were determined independently by other groups (PDB deposition codes 6WJI, 6WZO, 6WZQ and 7C22) [45]. This emphasizes the massive research effort undertaken during the Covid-19 pandemic to provide structure-based platforms for drug development, also high-lighting the attractiveness of this protein as antiviral target. The

high-resolution structures of the SARS-CoV-2 N CTD herein described and the accompanying biophysical characterization improve the framework with new fundamental tools to counter the ongoing global health emergency.

Author contributions

L.Z. conceived the study. L.Z. and I.N. designed and performed molecular cloning. L.Z. performed protein expression and purification, carried out biochemical experiments, obtained the crystals and performed EM specimen preparation. J.B. optimized crystal screenings, collected diffraction data, performed model building and crystal structure refinement. L.Z., J.B. and A.B. performed crystal structures analysis and data interpretation. S.B. and G.P. performed EM dataset acquisition. S.B., F.B. and S.K. performed EM image processing. L.Z., S·B., F.B. and S.K. performed EM structural analysis and data interpretation. F.U.H. and W.B. supervised experimental design and data interpretation. L.Z. drafted the manuscript. All authors contributed to experimental design, data analyses and manuscript review.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.bbrc.2020.09.131.

References

[1] C. Huang, Y. Wang, X. Li, L. Ren, J. Zhao, Y. Hu, L. Zhang, G. Fan, J. Xu, X. Gu, Z. Cheng, T. Yu, J. Xia, Y. Wei, W. Wu, X. Xie, W. Yin, H. Li, M. Liu, Y. Xiao, H. Gao, L. Guo, J. Xie, G. Wang, R. Jiang, Z. Gao, Q. Jin, J. Wang, B. Cao, Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China, Lancet 395 (2020) 497−506.

[2] N. Chen, M. Zhou, X. Dong, J. Qu, F. Gong, Y. Han, Y. Qiu, J. Wang, Y. Liu, Y. Wei, J. Xia, T. Yu, X. Zhang, L. Zhang, Epidemiological and clinical characteristics of 99 cases of 2019 novel coronavirus pneumonia in Wuhan, China: a descriptive study, Lancet 395 (2020) 507−513.

[3] J.F. Chan, S. Yuan, K.H. Kok, K.K. To, H. Chu, J. Yang, F. Xing, J. Liu, C.C. Yip, R.W. Poon, H.W. Tsoi, S.K. Lo, K.H. Chan, V.K. Poon, W.M. Chan, J.D. Ip, J.P. Cai, V.C. Cheng, H. Chen, C.K. Hui, K.Y. Yuen, A familial cluster of pneumonia associated with the 2019 novel coronavirus indicating person-to-person transmission: a study of a family cluster, Lancet 395 (2020) 514−523.

[4] Q. Li, X. Guan, P. Wu, X. Wang, L. Zhou, Y. Tong, R. Ren, K.S.M. Leung, E.H.Y. Lau, J.Y. Wong, X. Xing, N. Xiang, Y. Wu, C. Li, Q. Chen, D. Li, T. Liu, J. Zhao, M. Liu, W. Tu, C. Chen, L. Jin, R. Yang, Q. Wang, S. Zhou, R. Wang, H. Liu, Y. Luo, Y. Liu, G. Shao, H. Li, Z. Tao, Y. Yang, Z. Deng, B. Liu, Z. Ma, Y. Zhang, G. Shi, T.T.Y. Lam, J.T. Wu, G.F. Gao, B.J. Cowling, B. Yang, G.M. Leung, Z. Feng, Early transmission dynamics in wuhan, China, of novel coronavirus-infected pneumonia, N. Engl. J. Med. 382 (2020) 1199−1207.

[5] F. Wu, S. Zhao, B. Yu, Y.M. Chen, W. Wang, Z.G. Song, Y. Hu, Z.W. Tao, J.H. Tian, Y.Y. Pei, M.L. Yuan, Y.L. Zhang, F.H. Dai, Y. Liu, Q.M. Wang, J.J. Zheng, L. Xu, E.C. Holmes, Y.Z. Zhang, A new coronavirus associated with human respiratory disease in China, Nature 579 (2020) 265−269.

[6] P. Zhou, X.L. Yang, X.G. Wang, B. Hu, L. Zhang, W. Zhang, H.R. Si, Y. Zhu, B. Li, C.L. Huang, H.D. Chen, J. Chen, Y. Luo, H. Guo, R.D. Jiang, M.Q. Liu, Y. Chen, X.R. Shen, X. Wang, X.S. Zheng, K. Zhao, Q.J. Chen, F. Deng, L.L. Liu, B. Yan, F.X. Zhan, Y.Y. Wang, G.F. Xiao, Z.L. Shi, A pneumonia outbreak associated with a new coronavirus of probable bat origin, Nature 579 (2020) 270−273.

[7] Coronaviridae Study Group of the International Committee on Taxonomy of Viruses, The species Severe acute respiratory syndrome-related coronavirus: classifying 2019-nCoV and naming it SARS-CoV-2. Version 2, Nat Microbiol 5 (2020) 536−544.

[8] C. Wang, P.W. Horby, F.G. Hayden, G.F. Gao, A novel coronavirus outbreak of global health concern, Lancet 395 (2020) 470−473.

[9] (CSSE) at Johns Hopkins University (JHU), COVID-19 Dashboard by the Center for Systems Science and Engineering, 2020. https://coronavirus.jhu.edu/map.html/. (Accessed 24 August 2020).

[10] T. Thanh Le, Z. Andreadakis, A. Kumar, R. Gómez Román, S. Tollefsen, M. Saville, S. Mayhew, The COVID-19 vaccine development landscape, Nat. Rev. Drug Discov. 19 (2020) 305−306.

[11] H. Li, Y. Zhou, M. Zhang, H. Wang, Q. Zhao, J. Liu, Updated approaches against SARS-CoV-2, Antimicrob. Agents Chemother. 64 (6) (2020), e00483, 20.

[12] C.K. Chang, M.H. Hou, C.F. Chang, C.D. Hsiao, T.H. Huang, The SARS coronavirus nucleocapsid protein - forms and functions, Antivir. Res. 103 (2014) 39−50.

[13] R. McBride, M. van Zyl, B.C. Fielding, The coronavirus nucleocapsid is a multifunctional protein, Viruses 6 (2014) 2991−3018.

[14] P.D. Burbelo, F.X. Riedo, C. Morishima, S. Rawlings, D. Smith, S. Das, J.R. Strich, D.S. Chertow, R.T. Davey, J.I. Cohen, Detection of nucleocapsid antibody to SARS-CoV-2 is more sensitive than antibody to spike protein in COVID-19 patients, J. Infect. Dis. 19 (2020), jiaa273.

[15] P.R. Randad, N. Pisanic, K. Kruczynski, Y.C. Manabe, D. Thomas, A. Pekosz, S. Klein, M.J. Betenbaugh, W.A. Clarke, O. Laeyendecker, P.P. Caturegli, H.B. Larman, B. Detrick, J.K. Fairley, A.C. Sherman, N. Rouphael, S. Edupuganti, D.A. Granger, S.W. Granger, M. Collins, C.D. Heaney, COVID-19 Serology at Population Scale: SARS-CoV-2-specific Antibody Responses in Saliva, medRxiv [Preprint], 2020, https://doi.org/10.1101/2020.05.24.20112300.

[16] F. Yasui, C. Kai, M. Kitabatake, S. Inoue, M. Yoneda, S. Yokochi, R. Kase, S. Sekiguchi, K. Morita, T. Hishima, H. Suzuki, K. Karamatsu, Y. Yasutomi, H. Shida, M. Kidokoro, K. Mizuno, K. Matsushima, M. Kohara, Prior immunization with severe acute respiratory syndrome (SARS)-associated coronavirus (SARS-CoV) nucleocapsid protein causes severe pneumonia in mice infected with SARS-CoV, J. Immunol. 181 (2008) 6337−6348.

[17] T. Gao, M. Hu, X. Zhang, H. Li, L. Zhu, H. Liu, Q. Dong, Z. Zhang, Z. Wang, Y. Hu, Y. Fu, Y. Jin, K. Li, S. Zhao, Y. Xiao, S. Luo, L. Li, L. Zhao, J. Liu, H. Zhao, Y. Liu, W. Yang, J. Peng, X. Chen, P. Li, Y. Liu, Y. Xie, J. Song, L. Zhang, Q. Ma, X. Bian, W. Chen, X. Liu, Q. Mao, C. Cao, Highly Pathogenic Coronavirus N Protein Aggravates Lung Injury by MASP-2-Mediated Complement Over-activation, medRxiv [Preprint, 2020, https://doi.org/10.1101/2020.03.29.20041962.

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

[19] W. Kabsch, XDS, Acta Crystallogr D Biol Crystallogr 66 (2010) 125−132.

[20] A.J. McCoy, R.W. Grosse-Kunstleve, P.D. Adams, M.D. Winn, L.C. Storoni, R.J. Read, Phaser crystallographic software, J. Appl. Crystallogr. 40 (2007) 658−674.

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

[22] P. Emsley, B. Lohkamp, W.G. Scott, K. Cowtan, Features and development of Coot, Acta Crystallogr D Biol Crystallogr 66 (2010) 486−501.

[23] P.D. Adams, P.V. Afonine, G. Bunkóczi, V.B. Chen, I.W. Davis, N. Echols, J.J. Headd, L.W. Hung, G.J. Kapral, R.W. Grosse-Kunstleve, A.J. McCoy, N.W. Moriarty, R. Oeffner, R.J. Read, D.C. Richardson, J.S. Richardson, T.C. Terwilliger, P.H. Zwart, PHENIX: a comprehensive Python-based system for macromolecular structure solution, Acta Crystallogr D Biol Crystallogr 66 (2010) 213−221.

[24] G.N. Murshudov, P. Skubák, A.A. Lebedev, N.S. Pannu, R.A. Steiner, R.A. Nicholls, M.D. Winn, F. Long, A.A. Vagin, REFMAC5 for the refinement of macromolecular crystal structures, Acta Crystallogr D Biol Crystallogr 67 (2011) 355−367.

[25] D.N. Mastronarde, Automated electron microscope tomography using robust prediction of specimen movements, J. Struct. Biol. 152 (2005) 36−51.

[26] N. Biyani, R.D. Righetto, R. McLeod, D. Caujolle-Bert, D. Castano-Diez, K.N. Goldie, H. Stahlberg, Focus: the interface between data collection and data processing in cryo-EM, J. Struct. Biol. 198 (2017) 124−133.

[27] S.Q. Zheng, E. Palovcak, J.P. Armache, K.A. Verba, Y. Cheng, D.A. Agard, MotionCor2: anisotropic correction of beam-induced motion for improved cryo-electron microscopy, Nat. Methods 14 (2017) 331−332.

[28] K. Zhang, Gctf: real-time CTF determination and correction, J. Struct. Biol. 193 (2016) 1−12.

[29] K. Zhang, M. Li, F. Sun, Gautomatch: an Efficient and Convenient Gpu-Based Automatic Particle Selection Program, 2011. https://www.mrc-lmb.cam.ac.uk/kzhang/.

[30] S.H. Scheres, RELION: implementation of a Bayesian approach to cryo-EM

structure determination, J. Struct. Biol. 180 (2012) 519–530.

[31] A. Punjani, J.L. Rubinstein, D.J. Fleet, M.A. Brubaker, cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination, Nat. Methods 14 (2017) 290–296.

[32] I.M. Yu, M.L. Oldham, J. Zhang, J. Chen, 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 (2006) 17134–17139.

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

[34] T.H.V. Nguyen, J. Lichière, B. Canard, N. Papageorgiou, S. Attoumani, F. Ferron, B. Coutard, Structure and oligomerization state of the C-terminal region of the Middle East respiratory syndrome coronavirus nucleoprotein, Acta Crystallogr D Struct Biol (2019) 8–15.

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

[36] Y. Ma, X. Tong, X. Xu, X. Li, Z. Lou, Z. Rao, 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) 688–697.

[37] B. Szelazek, W. Kabala, K. Kus, M. Zdzalik, A. Twarda-Clapa, P. Golik, M. Burmistrz, D. Florek, B. Wladyka, K. Pyrc, G. Dubin, Structural characterization of human coronavirus NL63 N protein, J. Virol. 91 (11) (2017) e02503–e02516.

[38] C.K. Chang, C.M. Chen, M.H. Chiang, Y.L. Hsu, T.H. Huang, Transient oligomerization of the SARS-CoV N protein-implication for virus ribonucleo-protein packaging, PloS One 8 (5) (2013), e65045, https://doi.org/10.1371/journal.pone.0065045.

[39] W. Zeng, G. Liu, H. Ma, D. Zhao, Y. Yang, M. Liu, A. Mohammed, C. Zhao, Y. Yang, J. Xie, C. Ding, X. Ma, J. Weng, Y. Gao, H. He, T. Jin, Biochemical characterization of SARS-CoV-2 nucleocapsid protein, Biochem. Biophys. Res. Commun. 527 (2020) 618–623.

[40] Y. Wang, X. Wu, Y. Wang, B. Li, H. Zhou, G. Yuan, Y. Fu, Y. Luo, Low stability of nucleocapsid protein in SARS virus, Biochemistry 43 (2004) 11103–11108.

[41] C.Y. Huang, Y.L. Hsu, W.L. Chiang, M.H. Hou, Elucidation of the stability and functional regions of the human coronavirus OC43 nucleocapsid protein, Protein Sci. 18 (2009) 2209–2218.

[42] P.K. Hsieh, S.C. Chang, C.C. Huang, T.T. Lee, C.W. Hsiao, Y.H. Kou, I.Y. Chen, C.K. Chang, T.H. Huang, M.F. Chang, Assembly of severe acute respiratory syndrome coronavirus RNA packaging signal into virus-like particles is nucleocapsid dependent, J. Virol. 79 (2005) 13848–13855.

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

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

[45] Q. Ye, A.M.V. West, S. Silletti, K.D. Corbett, Architecture and Self-Assembly of the SARS-CoV-2 Nucleocapsid Protein, bioRxiv [Preprint], 2020, https://doi.org/10.1101/2020.05.17.100685.