STRUCTURAL BIOLOGY

ISSN 2059-7983

Structural characterization of the N-terminal part of the MERS-CoV nucleocapsid by X-ray diffraction and small-angle X-ray scattering

Nicolas Papageorgiou,[a,b]* Julie Lichière,[a,b] Amal Baklouti,[a,b] François Ferron,[a,b] Marion Sévajol,[a,b] Bruno Canard[a,b] and Bruno Coutard[a,b]*

[a]CNRS, AFMB UMR 7257, 13288 Marseille, France, and [b]Aix-Marseille Université, AFMB UMR 7257, 13288 Marseille, France. *Correspondence e-mail: nicolas.papageorgiou@afmb.univ-mrs.fr, bruno.coutard@afmb.univ-mrs.fr

Received 5 October 2015
Accepted 17 December 2015

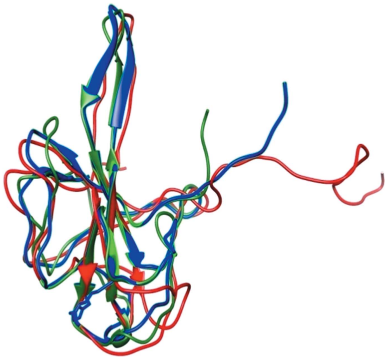
Edited by R. McKenna, University of Florida, USA

Keywords: MERS-CoV; nucleocapsid; structure; SAXS; RNA-binding domain.

PDB reference: N-terminal domain of the MERS-CoV nucleocapsid, 4ud1

Supporting information: this article has supporting information at journals.iucr.org/d

The N protein of coronaviruses is a multifunctional protein that is organized into several domains. The N-terminal part is composed of an intrinsically disordered region (IDR) followed by a structured domain called the N-terminal domain (NTD). In this study, the structure determination of the N-terminal region of the MERS-CoV N protein *via* X-ray diffraction measurements is reported at a resolution of 2.4 Å. Since the first 30 amino acids were not resolved by X-ray diffraction, the structural study was completed by a SAXS experiment to propose a structural model including the IDR. This model presents the N-terminal region of the MERS-CoV as a monomer that displays structural features in common with other coronavirus NTDs.



1. Introduction

Middle East respiratory syndrome coronavirus (MERS-CoV) is the aetiological agent responsible for an outbreak of a severe respiratory syndrome disease in the Middle East with sporadic spread to Europe, Africa, Asia and America. Since June 2012, the date of the first clinical case observed in Saudi Arabia, more than 1600 other cases have been confirmed, with a fatality rate of 36%, as reported by the World Health Organization in December 2015. In humans, MERS-CoV can result in asymptomatic viraemia to severe respiratory distress associated with organ failure. The animal source of the virus is still under investigation, but to date scenarios involving bats and/or camels have been proposed (Azhar *et al.*, 2014; Corman *et al.*, 2014; Müller *et al.*, 2015; Raj *et al.*, 2014).

MERS-CoV is a positive-strand RNA virus belonging to the *Betacoronavirus* genus within the *Coronaviridae* family of the *Nidovirales* order. The 5′ two-thirds of the MERS-CoV genome contains open reading frames 1a and 1b, which are translated into the viral polyproteins pp1a and pp1ab, respectively. The processing of these two proteins leads to 16 nonstructural proteins which are involved in the replication complex. The 3′-proximal third encodes the viral structural proteins and several accessory proteins, which are translated from a set of subgenomic mRNAs. Among the structural proteins, the nucleocapsid (N) protein is a key protein displaying several functions. Indeed, it is not only involved in genome packaging through the formation of a ribonucleoprotein complex (RNP; reviewed in Chang *et al.*, 2014), but it also regulates viral RNA synthesis during replication and transcription (Chang *et al.*, 2014; Nelson *et al.*, 2000; Stohlman *et al.*, 1988; Wu *et al.*, 2014). Furthermore, the N protein participates in deregulation of the cell cycle, host translational shutoff and innate immunity disruption (reviewed in McBride *et al.*, 2014).

The coronavirus N protein is a >40 kDa protein organized into two main folded domains called the N-terminal domain (NTD) and the C-terminal domain (CTD) separated by a flexible linker (Fig. 1*a*). The NTD is subdivided into an RNA-binding module preceded by an intrinsically disordered region (IDR). This IDR contributes to the RNA-binding properties of the NTD (Chang *et al.*, 2009). The linker region contains a Ser/Arg stretch that could correspond to phosphorylation sites involved in the regulation of the functions of the N protein (Chang *et al.*, 2009; He *et al.*, 2004; Peng *et al.*, 2008; Surjit *et al.*, 2005; Wu *et al.*, 2009). The CTD is involved in N-protein dimerization and RNA binding (Chen *et al.*, 2007; Takeda *et al.*, 2008; Yu *et al.*, 2006). It is followed by another IDR. Based on electron microscopy or small-angle X-ray scattering data, the proposed structural models of the N-protein organization suggest that the NTD and CTD may be independent domains with no interaction between each other (Chang *et al.*, 2006). Since the full-length N protein contains a large fraction (>35%) of disordered regions that are recalcitrant to crystallization, it is therefore relevant to study the structural and functional features of each domain independently.

The three-dimensional structures of the NTDs from several coronaviruses have previously been determined: those from Severe acute respiratory syndrome coronavirus (SARS-CoV; Huang *et al.*, 2004; Saikatendu *et al.*, 2007), infectious bronchitis virus Beaudette (IBV; Fan *et al.*, 2005; Jayaram *et al.*, 2006), murine hepatitis virus (MHV; Grossoehme *et al.*, 2009) and human coronavirus OC43 (HCoV-OC43; Chen *et al.*, 2013). An alignment between the NTD sequences of MERS-CoV, IBV and SARS-CoV shows 37% sequence identity, while all reported coronavirus NTD structures are organized as a five-stranded antiparallel β-sheet core domain with disordered loops of varying sizes connecting the strands. A positively charged β-hairpin protrudes out of the core domain. Structural analysis enabled this region to be proposed as a possible RNA-binding domain. Indeed, mutational analysis and NMR studies led to the conclusion that this region is a major actor in RNA recruitment (Chen *et al.*, 2013; Clarkson *et al.*, 2009; Grossoehme *et al.*, 2009; Saikatendu *et al.*, 2007; Tan *et al.*, 2006). Recently, an inhibitor of the RNA–N interaction has been demonstrated to have an antiviral effect on HCoV-OC43 (Lin *et al.*, 2014). The coronavirus N protein is thus an attractive target for antiviral research. Because of the worldwide spread of MERS-CoV, with potential risks for healthcare, it is important to accumulate structural and functional information on MERS-CoV proteins in order to provide tools for antiviral development.

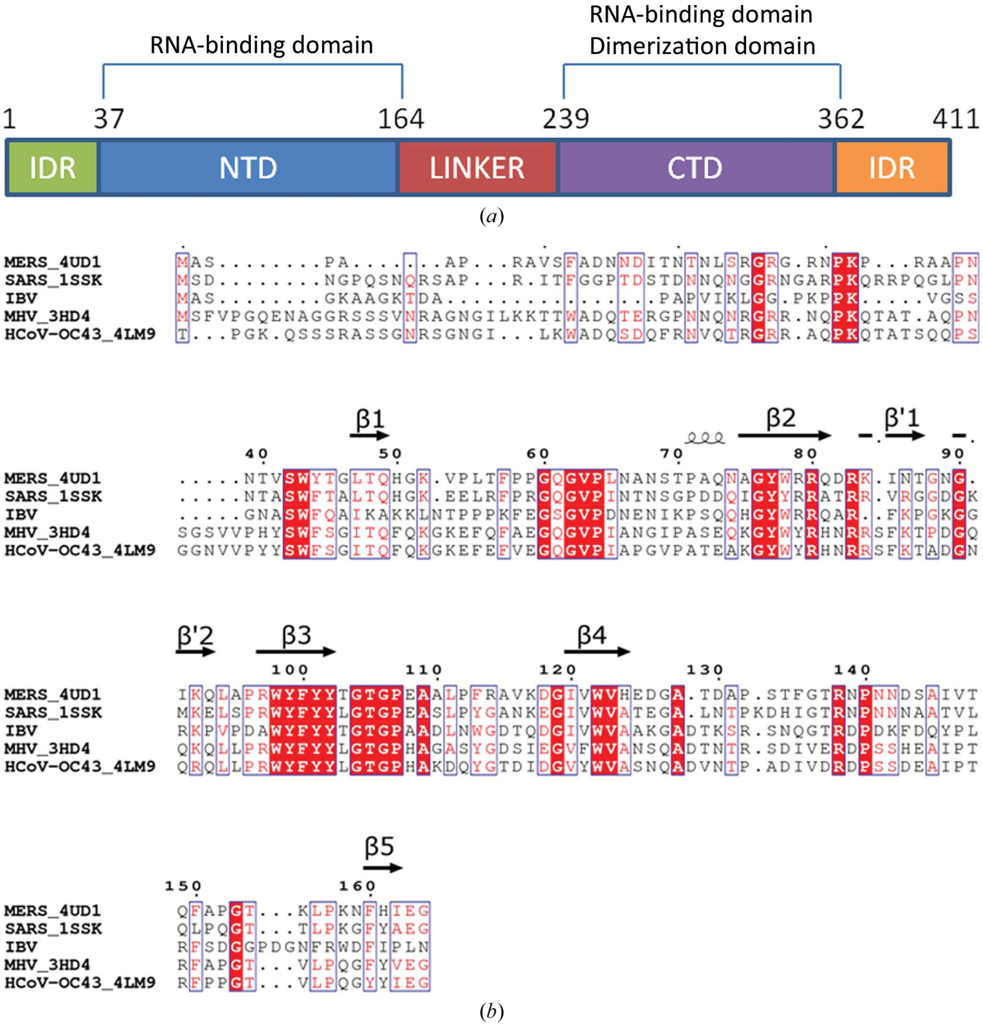


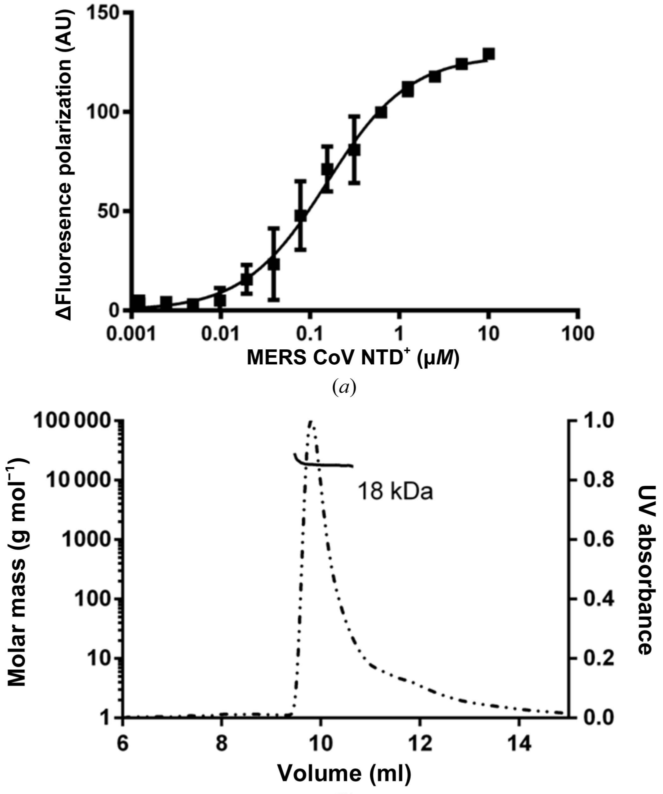
Figure 1
Sequence analysis of MERS-CoV N protein. (*a*) Modular organization of the coronavirus N protein. The domain boundaries correspond to the MERS-CoV N protein numbering. IDR, intrinsically disordered region; NTD, N-terminal domain; CTD, C-terminal domain. (*b*) Sequence alignment of the N-terminal region of several N proteins for which the structure of the NTD has been documented. The sequences are named as follows: virus acronym_PDB code. The depicted secondary structure was retrieved from the MERS-CoV NTD$^+$ structure described in this study. The alignment was obtained using *T-Coffee* (Notredame *et al.*, 2000) and *ESPript* 3.0 (Robert & Gouet, 2014).

In this study, we present the three-dimensional structure of the MERS-CoV NTD with its upstream IDR region (hereafter referred to as NTD$^+$) at a resolution of 2.4 Å obtained by X-ray diffraction measurements as well as a small-angle X-ray scattering (SAXS) study of the protein in solution in order to assess the oligomerization state of MERS-CoV NTD$^+$, and propose a structural model of NTD$^+$ including its disordered region.

2. Materials and methods

2.1. Protein production, purification and characterization

The codon-optimized DNA encoding the N-terminal region (amino acids 1–164; NTD$^+$) of the MERS-CoV nucleocapsid (strain Betacoronavirus England 1, accession No. KC164505) was synthesized by GenScript. The DNA sequence was amplified by a two-step PCR before cloning into the expression vector pMCOX20A to enable the fusion of a cleavable thioredoxin/6×His tag at the N-terminus of NTD$^+$ (Lantez *et al.*, 2011). The protein was then produced in *Escherichia coli* T7 Express (DE3) cells (New England Biolabs). Expression was induced overnight at 290 K in TB medium with 0.5 m*M* IPTG when the OD$_{600\,\mathrm{nm}}$ of the culture had reached 0.6. Purification of the protein and tag removal was performed in non-denaturing conditions as described previously (Lantez *et al.*, 2011). The final size-exclusion chromatography (SEC) step was performed in 10 m*M* HEPES, 300 m*M* NaCl pH 7.5. The protein folding and stability were assessed by thermal shift assays (Geerlof *et al.*, 2006) following a previously described protocol (Malet *et al.*, 2009). The nonspecific RNA-binding properties of the MERS-CoV NTD$^+$ were assessed by fluorescence polarization-based assays using 5′-CCAGGCGACAUCAGCG-3′ RNA labelled at the 3′ end with a Cy3 probe. The binding assay was performed in 50 m*M* Tris buffer pH 7.5, 1 m*M* DTT, 2 m*M* MgCl$_2$. The Cy3-labelled RNA was used at 50 n*M*. Fluorescence polarization measurements were performed in a microplate reader (PHERAstar FS, BMG Labtech) with an optical module equipped with polarizers and using excitation and emission wavelengths of 546 and 563 nm, respectively.



(*a*)

(*b*)

Figure 2
Biophysical characterization of MERS-CoV NTD$^+$. (*a*) Binding of RNA to MERS-CoV NTD$^+$ assessed by fluorescence polarization. 3′-Cy5-labelled RNA was incubated with various concentrations of NTD$^+$. (*b*) MALLS/UV/refractometry/SEC analysis of NTD$^+$. The right *y* axis represents the OD$_{280\,\mathrm{nm}}$ and the left *y* axis represents the molar mass in g mol^{-1} (Da). The absorption peak is shown as a noncontinuous line, with the continuous trace indicating the molar mass. The value of the measured mass at the volume corresponding to the top of the peak is reported.

In order to confirm the oligomerization state, analytical SEC with online multi-angle laser light scattering, absorbance and refractive index (MALLS/UV/RI) measurements were carried out on an Alliance 2695 HPLC system (Waters) using a Silica Gel KW802.5 column (Shodex) equilibrated in 10 m*M* HEPES, 150 m*M* NaCl pH 7.5. Detection was performed using a triple-angle light-scattering detector (miniDAWN TREOS, Wyatt Technology), a quasi-elastic light-scattering instrument (DynaPro, Wyatt Technology) and a differential refractometer (Optilab rEX, Wyatt Technology). Molecular-weight and hydrodynamic radius determination was performed by the *ASTRA V* software (Wyatt Technology) using a d*n*/d*c* value of 0.185 ml g^{-1}. Proteins were loaded at a final concentration of 5 mg ml^{-1}.

2.2. Protein crystallization, data collection and processing

Crystallization trials were performed by the hanging-drop vapour-diffusion method at 293 K using a nanodrop-dispensing robot (Mosquito, TTP Labtech). Optimal crystallization conditions were obtained by mixing 100 nl protein solution at 25 mg ml^{-1} with 100 nl reservoir solution (0.1 *M* sodium acetate pH 5, 2 *M* ammonium sulfate). The crystals of MERS-CoV NTD$^+$ were flash-cooled in liquid nitrogen at 110 K. Glycerol was added to the mother liquor as a cryoprotectant to a final concentration of 20%.

Diffraction data for the native enzyme were collected on the ID30A-1 beamline at the European Synchrotron Radiation Facility (ESRF), Grenoble, France using a Pilatus3 2M detector at a wavelength of 0.965 Å and a temperature of 100 K.

The data were integrated by *XDS* (Kabsch, 2010), indexed in space group *C*121 (unit-cell parameters a = 208.648, b = 67.021, c = 60.94, $\alpha = \beta = \gamma = 90$), scaled and truncated to a resolution of 2.4 Å assuming a minimum accepted signal-to-noise ratio equal to 2. The data were subsequently used for molecular replacement with *Phaser* for phasing and were refined and validated by *PHENIX* (Adams *et al.*, 2010). The molecular-graphics software *Coot* (Emsley & Cowtan, 2004) was used for model building, while *Chimera* (Pettersen *et al.*, 2004) was used to visualize and draw the models.

2.3. Small-angle X-ray scattering (SAXS) measurements and data processing

All SAXS measurements were carried out on beamline BM29 at the ESRF at a working energy of 12.5 keV, leading to

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

Data collection	
Wavelength (Å)	0.965
Resolution range (Å)	48.26–2.48 (2.486–2.480)
Space group	$C121$
Unit-cell parameters (Å, °)	a = 208.648, b = 67.021, c = 60.94, α = 90, β = 89.98, γ = 90
Total reflections	101472 (9834)
Unique reflections	32509 (3088)
Multiplicity	3.2 (3.2)
Completeness (%)	98.04 (97.87)
Mean $I/\sigma(I)$	5.23 (1.90)
Wilson B factor (Å^2)	55.81
$R_{\rm merge}$	0.143 (0.6257)
$R_{\rm meas}$	0.1708
$\rm CC_{1/2}$	0.967 (0.831)
CC*	0.991 (0.953)
Refinement	
$R_{\rm work}$	0.2281 (0.4445)
$R_{\rm free}$	0.2738 (0.4771)
No. of non-H atoms	
Total	4922
Macromolecules	4839
Ligands	32
Water	51
No. of protein residues	621
R.m.s.d., bonds (Å)	0.005
R.m.s.d., angles (°)	0.87
Ramachandran favoured (%)	98.2
Ramachandran outliers (%)	0
Rotamer outliers (%)	0
C^{β} outliers (%)	0
Clashscore	5.08
Average B factor (protein) (Å^2)	64.70
Overall *MolProbity* score	1.27

a scattering vector s ranging from 0.025 to 5 nm^{-1}. Data were recorded using a Pilatus 1M detector at a sample-to-detector distance of 2.43 m. The scattering vector s is defined as

$$s = \frac{4\pi}{\lambda} \sin\theta, \qquad (1)$$

where λ is the wavelength of the incident radiation in nanometres and θ is half of the angle between the incident and the scattered radiation.

A range of seven protein concentrations was measured (0.21, 0.41, 0.77, 1.55, 3.06, 5.85 and 11.66 mg ml^{-1}) in 10 mM HEPES pH 7.5, 300 mM NaCl. All samples were centrifuged for 15 min at 15 000g before the experiment in order to minimize contributions from aggregated particles.

The measurements proceeded with 45 µl protein sample injected into a 1.8 mm capillary with a flow to minimize radiation damage; data were collected at 277 K.

Ten exposures of 1 s each were made for each protein concentration and were combined to give the average scattering curve for each measurement. A buffer reference was performed before and after measurement for each protein sample in the same conditions. The forward scattering intensity was calibrated using BSA as a reference at 5 mg ml^{-1}.

Data were processed with the *ATSAS* package (Petoukhov & Svergun, 2007) according to the standard procedure. Ten frames of 1 s per protein were averaged using *PRIMUS* (Konarev *et al.*, 2003). Frames affected by radiation damage were excluded and the buffer background was subtracted from the sample.

3. Results

3.1. Protein production and purification

NTD$^+$ of the MERS-CoV N protein was partly designed based on available coronavirus N-protein crystal structures (Chen *et al.*, 2007; Fan *et al.*, 2005; Grossoehme *et al.*, 2009; Saikatendu *et al.*, 2007). Sequence alignments coupled with disorder predictions revealed that the N-terminal intrinsically disordered region (IDR) starts at position 1 and ends at position 36, whereas the structured RNA-binding domain (RBD) encompasses amino acids 37–164 (Fig. 1*a*). To date, the crystal structures of NTDs from different coronaviruses have been obtained after the removal of the IDR. In this study, the domain encompassing amino acid 1 to amino acid 164 was used to produce NTD$^+$ of the MERS-CoV N, thus including the N-terminal IDR. The recombinant wild-type (wt) protein was produced in *E. coli* and was purified for crystallization, SAXS experiments and oligomerization studies. The quality of the recombinant protein was assessed by a thermal shift assay (TSA) and an RNA-binding assay using a nonspecific RNA sequence. TSA experiments revealed a transition curve from a folded to an unfolded state for MERS-CoV NTD$^+$, suggesting that the protein is structured. Moreover, the observed melting temperature ($T_{\rm m}$) is 313 K. For comparison, the $T_{\rm m}$ of the MERS-CoV NTD$^+$ is 8 K lower than that obtained for the SARS-CoV NTD (Fang *et al.*, 2009). An RNA-binding assay using fluorescence polarization was performed as an additional quality assessment. The protein displays an affinity constant in the hundred-nanomolar range for a nonspecific

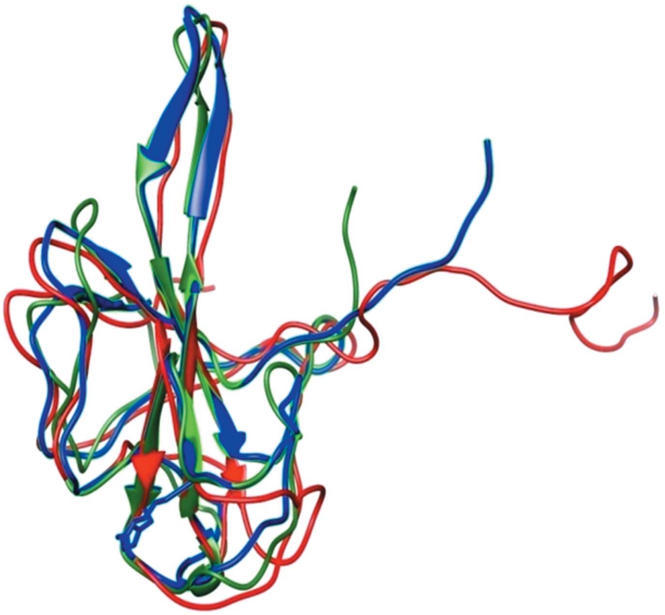


Figure 3
Ribbon representation of the coronavirus NTDs. Superposition of the MERS-CoV NTD$^+$ structure (blue) with those of the IBV (green) and the SARS-CoV (red) NTDs, showing the relatively good superposition of the main core of the protein and the high flexibility of the interconnecting random-coil loops.

RNA, with a K_d of 153 nM. The results are shown in Fig. 2(*a*). Together, these results show that MERS-CoV NTD$^+$ is well folded and suitable for crystallization.

3.2. Structure determination by X-ray diffraction

Molecular replacement (MR) was carried out using the IBV NTD (PDB entry 2btl; Fan *et al.*, 2005). A unique solution was found in space group *C*121 with five molecules per asymmetric unit. Subsequent refinement converged with no particular problems to an $R_{\rm work}$ of 0.23 and an $R_{\rm free}$ of 0.27. Details of data collection and refinement are shown in Table 1. The structure was deposited as PDB entry 4ud1.

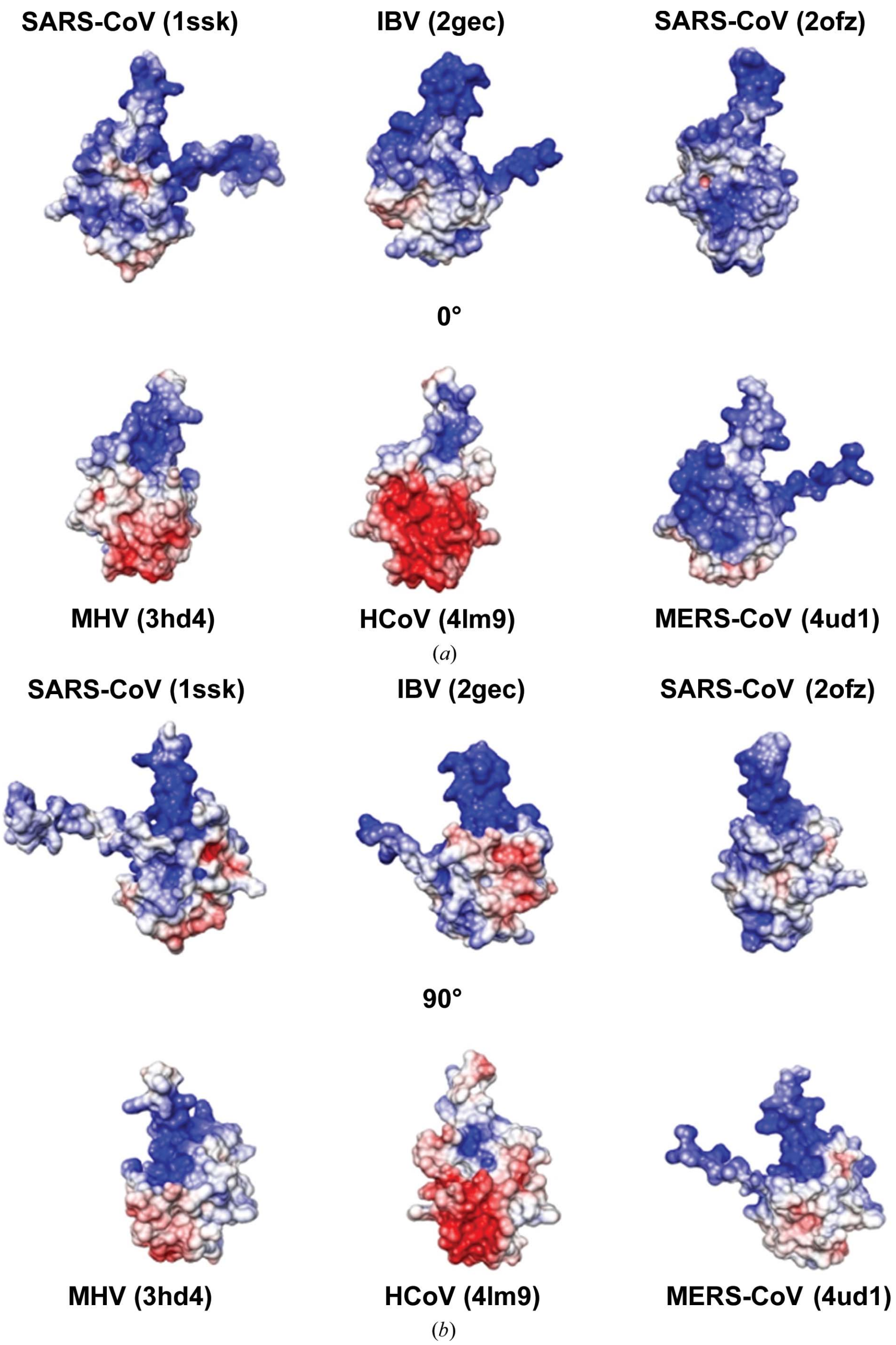


Figure 4
Front (*a*) and back (*b*) views of electrostatic surfaces from several NTDs originating from closely related viruses. The structure of the SARS-CoV NTD was measured in solution by NMR, while the other structures were all obtained by single-crystal X-ray diffraction. Surfaces were calculated by *APBS* and truncated from −5 mV (red) to +5 mV (blue).

The NTD$^+$ domain is globular overall (fully β), presenting a kidney-shaped surface organized as a main core featuring three antiparallel β-strands forming a central β-sheet from which a flexible β-hairpin structure extrudes. These features are interconnected with flexible coil loops. The flexibility of these coils is reflected by elevated *B* factors (overall *versus* local) and is also illustrated in Fig. 3, where the MERS-CoV NTD$^+$ structure (PDB entry 4ud1, blue) is superimposed with the IBV (PDB entry 2btl; Fan *et al.*, 2005) and SARS-CoV (PDB entry 1ssk; Huang *et al.*, 2004) NTD structures (green and red, respectively). The first 30 residues of the N-terminal region of MERS-CoV NTD$^+$, corresponding to a disordered region, could not be resolved. This region was previously scarcely resolved in all structures obtained using X-ray diffraction techniques, while it was resolved as far as amino acid 24 in the 1ssk structure measured in solution by NMR.

Fig. 4 shows the electrostatic surfaces of several N-terminal RNA-binding domains of NTD$^+$ originating from closely related viruses. Surfaces were calculated by *APBS* and truncated from −5 mV (red) to +5 mV (blue). Figs. 4(*a*) and 4(*b*) show front and back views of the protein, respectively. The common feature of all structures is the positively charged hairpin (oriented towards the top side in the figure) and a hydrophobic core domain (Fig. 5*b*). By contrast, the bottom side presents some differences in the charge distribution, especially for the proteins originating from murine hepatitis virus and human coronavirus OC43, which show a more important negatively charged region compared with the other structures (Ma *et al.*, 2010).

Analyses of the molecular arrangement within the crystal show stacking of Pro33 in front of the Trp43 residue belonging to the nearby molecule. The same stacking is also found within the crystal packing of the IBV NTD structure as presented in Fig. 5. In these figures hydrophobic residues (Val, Ile, Leu, Phe, Trp, Cys, Ala, Tyr, His, Thr, Ser, Gly and Lys) following the classification of Livingstone & Barton (1993) are labelled in green and highlight the conserved hydrophobic propensity of the core domain.

3.3. Oligomerization assays

Preparative SEC experiments that were carried out for protein crystallization and RNA binding revealed that wt NTD$^+$ eluted predominantly as monomers. However, this result is not sufficient to claim that the protein cannot behave as a transient multimer. SEC-MALLS experiments indicated that the observed molecular weight was 18 kDa, which is close to the theoretical molecular weight (17.8 kDa), confirming the SEC results (Fig. 2*b*). The hydrodynamic radius (R_h) was equal to 1.85 ± 0.45 nm. Although the MALLS data exclude the possibility that multimers are formed after the protein elutes from SEC, cross-linking experiments were also performed with NTD$^+$. The results revealed that even in the presence of glutaraldehyde NTD$^+$ has a very poor ability to produce cross-linked multimers (data not shown). Altogether, the oligomerization studies suggest that NTD$^+$ of the MERS-CoV nucleocapsid behaves as a monomer in solution and that, as suggested by the small interface between NTD$^+$ molecules in the crystal, the interaction observed in the crystal does not reflect an interaction in solution. In order to characterize the behaviour of MERS-CoV NTD$^+$ in solution as well as to obtain structural information about the N-terminal part of NTD$^+$, SAXS experiments were also carried out.

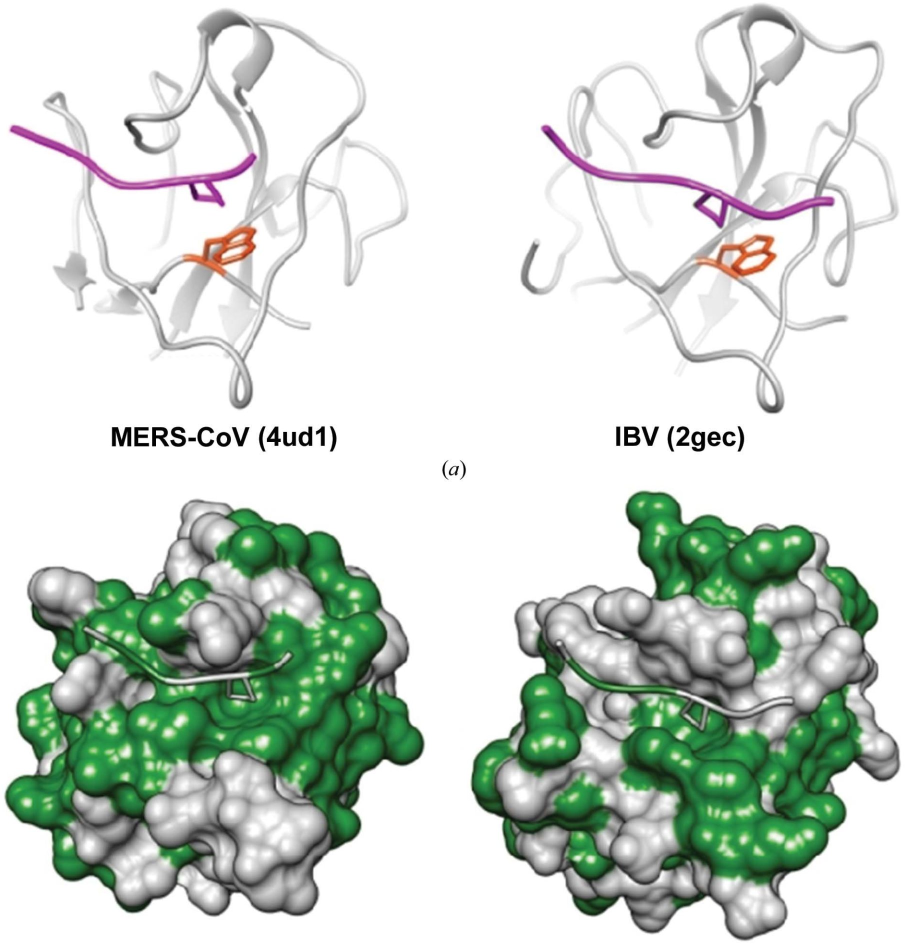


Figure 5
Representations of the interaction between NTD molecules in the crystal. (*a*) Ribbon representation of the structures at the interface between NTD$^+$ of MERS-CoV (left panel) and the NTD of IBV (right panel) in which Pro33 of the nearby molecule (in purple) is stacked with Trp43 (highlighted in red). (*b*) Surface representation of the same structures as in (*a*), where hydrophobic residues are labelled in green and the others in grey.

3.4. Small-angle X-ray scattering (SAXS) measurements

3.4.1. Guinier analysis. Fig. 6 shows the Guinier analysis in the range sR_g < 1.3, where R_g is the radius of gyration expressed in nanometres (R_g = 2 nm). The data corresponding to concentrations (conc) from 1.55 to 11.66 mg ml^{-1} present increasing nonlinearity with concentration. Data are represented as open circles, while red lines correspond to a linear regression. The values of concentration in mg ml^{-1}, R_g in nanometres as well as the normalized forward scattering intensity $I(0)$/conc are shown. Both R_g and $I(0)$/conc increase with increasing concentration, as is characteristic of an interparticle interference term originating from attractive intermolecular forces. However, for the lower concentration data (0.2, 0.41 and 0.77 mg ml^{-1}) the above values stabilize at approximately R_g = 2 nm and $I(0)$/conc = 26.5.

Taking into account the calibration SAXS profile of BSA, we estimate a molecular weight (MW) of 25 kDa, which is approximately 30% higher than the expected protein MW of 17.8 kDa.

3.4.2. *GNOM* analysis. All concentrations were processed separately by *GNOM* (Svergun, 1992) using the *PRIMUS* interface. The results are shown in Table 2. All parameters increase together with the concentration, as observed by Guinier analysis in the low-*s* region. Taking into account a mean value of the Porod volume (V) calculated for the three lower concentrations, $\langle V \rangle$ = 27.7 nm^3, we estimate an MW in the range 13.8 < MW < 18.4 kDa ($V/2$ < MW < $V/1.5$), which is in accordance with the MW of the protein monomer (17.8 kDa).

Considering the three lower concentrations, the *GNOM* analysis gives a radius of gyration R_g of 2 nm, which is equal to that found by the Guinier analysis. The ratio of R_g and R_h (R_g/R_h) provides shape information about proteins. Comparison with the hydrodynamic radius R_h of 1.85 nm measured

Table 2
Results of *GNOM* and *AUTOPOROD* analysis of all concentrations separately.

Concentration (mg ml^{-1})	R_g (nm)	$I(0)$/conc	Porod volume (nm^3)	D_{max} (nm)
0.20	2.04	26.8	27.1	8.0
0.41	2.06	26.8	28.8	8.5
0.77	2.13	29.5	27.0	8.5
1.55	2.21	30.7	30.2	8.5
3.06	2.40	32.3	29.7	10.9
5.85	2.65	38.0	32.7	11.6
11.66	3.01	32.2	36.1	14.0

by MALLS gives an R_g/R_h ratio of 1.08. The usual R_g/R_h value for a globular protein is ~0.775, which means that R_g is smaller than R_h. However, when molecules deviate from globular to nonspherical or elongated structures then R_g/R_h tends to values above 0.775 as R_g becomes larger than R_h. In the present case the R_g/R_h value is probably owing to a deviation from a globular structure owing to the additional N-terminal region (amino acids 1–30) of the protein which protrudes out of the main globular domain of the protein as discussed below.

3.4.3. *EOM* analysis (ensemble-optimization method). A preliminary analysis with *DAMMIF* (Franke & Svergun, 2009) of all four lower concentrations separately showed almost identical particles that nicely fit the experimental data (data not shown). The calculated particles feature a globular region with a protuberance (Fig. 10*b*). The structure 4ud1 described above nicely fits the globular part. In order to model the supplementary protuberant part of the *DAMMIF* particle we considered an additional 30-amino-acid N-terminal region and used the *EOM* 2.0 program (Bernadó *et al.*, 2007; Tria *et al.*, 2015). Model PDB structures with an N-terminal region containing the missing sequence were created. Two different kinds of *EOM* strategy were used: (i) the N-terminal region (amino acids 1–30) was modelled as a fully random region and (ii) the N-terminal region was divided into three equally long fully random parts spanning amino acids 1–30. The globular part of the protein taken from the 4ud1 structure was kept fixed in all cases. Both strategies gave the same results with respect to the final model structures as well as the corresponding R_g, size distributions, R_{flex} and R_σ statistics given by *EOM*.

EOM created a pool of 10 000 possible particles with R_g and size distributed randomly over the Gaussian-like distribution profiles shown in red bold lines in Fig. 7(*a*). For each specific concentration data set *EOM* chooses particle ensembles

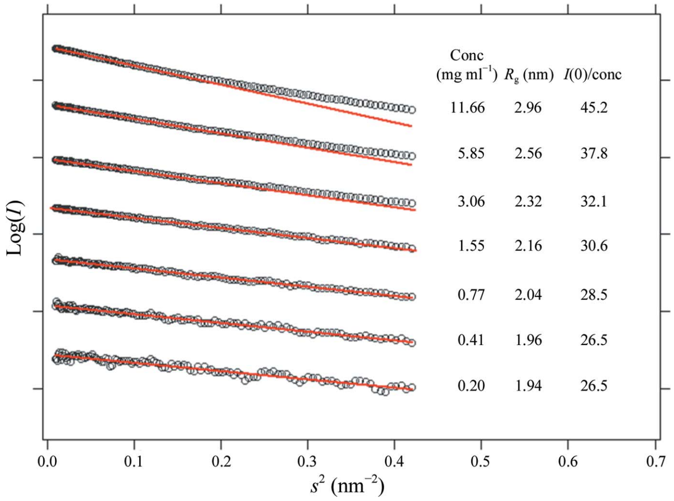
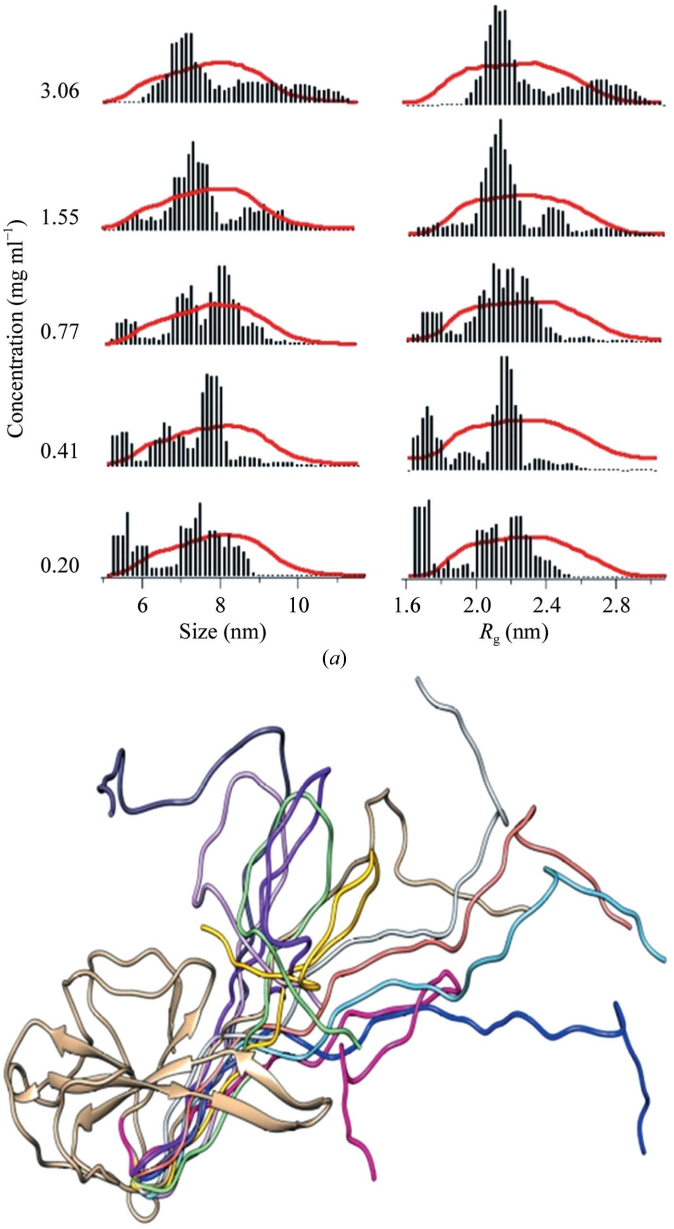


Figure 6
Guinier analysis in the range $sR_g < 1.3$ ($R_g = 2$ nm). Experimental data are represented by open circles, while red lines correspond to a linear regression. The values of concentration (conc), R_g in nanometres as well as the normalized forward scattering intensity $I(0)$/conc are shown. Both R_g and $I(0)$/conc increase with increasing concentration, which is characteristic of an interparticle interference term originating from attractive intermolecular forces.



(*a*)

(*b*)

Figure 7
EOM analysis. (*a*) R_g and size distributions for each concentration (black sticks) calculated by *EOM*. Pool distributions are shown as red continuous lines. (*b*) Superposition of several representative ensemble models calculated by *EOM*.

Table 3
$R_{\rm flex}$, $R_{\rm flex\text{-}pool}$ and R_σ metrics as a function of the concentration calculated by *EOM* for the five lower concentrations separately.

Concentration (mg ml^{-1})	$R_{\rm flex}$ (%)	$R_{\rm flex\text{-}pool}$ (%)	R_σ
0.20	76.8	91.1	0.99
0.41	75.1	90.4	1.30
0.77	78.8	90.5	1.20
1.55	84.7	90.1	0.95
3.06	85.2	89.4	1.32

Table 4
SAXS data-collection and scattering-derived parameters.

The scattering and molecular-mass parameters are related to the merged data profile discussed in the text.

Data-collection parameters	
Instrument	BM29, ESRF
Detector	Pilatus 1M
Beam geometry (mm)	700 × 700
Wavelength (Å)	0.992
q range (nm^{-1})	0.025–5
Exposure time (s)	1
Concentration range (mg ml^{-1})	0.2–11.66
Temperature (°C)	4
Structural parameters	
$I(0)$/conc from $P(r)$ (arbitrary units)	26.8
$I(0)$/conc from Guinier (arbitrary units)	26.5
R_g from $P(r)$ (nm)	2.04
R_g from Guinier (nm)	1.95
$D_{\rm max}$ (nm)	8
Porod volume (nm^3)	27.5
Molecular-mass determination	
Partial specific volume† (nm^3)	21.4
Molecular mass from $I(0)$‡ (kDa)	24.1
Molecular mass from sequence (kDa)	17.731
Software employed	
Primary data reduction	*FIT2D*
Data processing	*ATSAS*
Ab initio analysis	*DAMMIF*
Validation and averaging	*DAMAVER*
Computation of intensities	*CRYSOL*
Three-dimensional graphics	*CHIMERA*

† The partial specific volume was taken to be (1.21 × MW) Å^3 per molecule (Harpaz *et al.*, 1994). ‡ The molecular weight was calculated using calibrated data for BSA (66 kDa) at a concentration of 5 mg ml^{-1}.

containing all possible conformations that fit the data. The R_g and size distributions of these ensembles are shown as black sticks. Fig. 7(*b*) depicts the typical ensemble models calculated. Models with high R_g (>2.2 nm) and large size (>8 nm) correspond to molecules with a spatially extended N-terminal region, while lower values of R_g and size correspond to molecules with an N-terminal region folded closely to the main core.

In order to enable a quantitative characterization of the size or R_g distributions, *EOM* uses two metrics: $R_{\rm flex}$ and R_σ (Tria *et al.*, 2015). The $R_{\rm flex}$ metric is given as a percentage in the range 0–100%, with $R_{\rm flex} = 100\%$ indicating maximum flexibility. The R_σ metric indicates the variance of the ensemble distribution with respect to the original pool and is defined as the ratio $\sigma_{\rm e}/\sigma_{\rm p}$, where $\sigma_{\rm e}$ and $\sigma_{\rm p}$ are the standard deviations for the distributions of the selected ensemble and the pool, respectively. Values close to 1 describe a fully flexible system.

In Table 3 we report the $R_{\rm flex}$ and R_σ metrics of the ensembles, as well as $R_{\rm flex\text{-}pool}$ showing the $R_{\rm flex}$ metric of the pool, as a function of concentration between 0.2 and 3.06 mg ml^{-1}. For each concentration the $R_{\rm flex}$ of the ensemble ranges between 75 and 85% with an $R_{\rm flex\text{-}pool}$ of 90% and a R_σ in the range between 0.99 and 1.3. These values suggest the presence of important flexibility as discussed by Tria *et al.* (2015). However, visual inspection of the R_g and size distributions in Fig. 7 shows the presence of a main ensemble population centred around R_g = 2.2 ± 0.2 nm as well as the presence of a secondary ensemble population centred around R_g = 1.7 ± 0.2 nm which is clearly evident for the three lower concentrations. The presence of these distinct ensembles might indicate two preferential conformations of the N-terminal region, which could be open or closed.

3.4.4. Analysis of a merged profile. Taking into account all the above analysis, the data at the five lower concentrations (0.2, 0.41, 0.77, 1.55 and 3.06 mg ml^{-1}) were merged in order to retrieve a single data file with good signal to noise. The low-*s* region of the merged file is an average of the 0.2 and 0.4 mg ml^{-1} data, while the high-*s* region is an average of the 0.77, 1.55 and 3.06 mg ml^{-1} data.

A full analysis of the merged data with *GNOM*, *DAMMIF* and *EOM* was undertaken. Details of this analysis are shown in Table 4. A three-dimensional *ab initio* model was constructed from the merged SAXS data using *DAMMIF* (*s* range from 0.1 to 5 nm^{-1}).

No predefined shape or symmetry was used. 20 different models were calculated and subsequently averaged and filtered by *DAMAVER* (Volkov & Svergun, 2003). Fig. 8 presents the resulting numerical fits of *GNOM*, *DAMMIF* and *EOM* analysis. The *GNOM* fit curve is completely superimposed on the *DAMMIF* fit curve shown by the red continuous line, while the *EOM* fit curve is shown by the blue

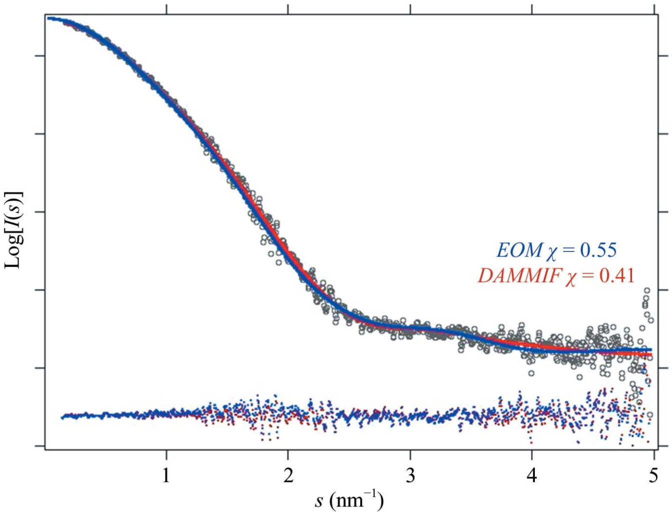


Figure 8
GNOM, *DAMMIF* and *EOM* fits. The *GNOM* fit is completely superimposed on the *DAMMIF* fit shown by the red continuous line, while the *EOM* fit is shown by the blue continuous line together with the merged data (open circles). The error signals, red dots for *GNOM* and *DAMMIF* and blue dots for *EOM*, are shown at the bottom of the figure and they have been almost superimposed for spatial economy.

continuous line along with the merged data (open circles). The error signals, red dots for *GNOM* and *DAMMIF* and blue dots for *EOM*, are shown at the bottom of the figure and have been almost superimposed for spatial economy.

Fig. 9(*a*) shows the $P(r)$ distribution of the merged data with $D_{\rm max}$ = 8 nm. Fig. 9(*b*) shows a Kratky plot of the merged data, featuring an increase in the signal in the high-s region (>2.5 nm^{-1}), probably owing to the flexible N-terminal region.

In Fig. 10(*a*) we show the three ensemble structures resulting from the *EOM* analysis of the merged data together with the corresponding $R_{\rm g}$ and size distributions considered by *EOM* (Fig. 10*b*). The contribution of each ensemble to the measured SAXS profile is shown as a percentage under each ensemble structure.

The corresponding metrics $R_{\rm flex}$ = 84% ($R_{\rm flex\text{-}pool}$ = 91%) and R_σ = 1.12 correspond to a predominantly flexible N-terminal region, while the size and $R_{\rm g}$ distributions shown in Fig. 10(*b*) provide evidence for two distinct populations (around $R_{\rm g}$ = 2.2 nm and $R_{\rm g}$ = 1.6 nm) as calculated previously for each concentration separately. As a conclusion from the *EOM* analysis, we can stress the presence of a flexible N-terminal region together with the presence of two conformations for the molecule: open and closed. In all cases the first 20 amino acids form a loop without being randomly unfolded.

The three ensemble model structures calculated by *EOM* fit within the filtered particle calculated by *DAMAVER* as shown in Fig. 10(*c*). The protuberant part of the particle mainly corresponds to the open conformation with the N-terminal region pointing out from the globular part.

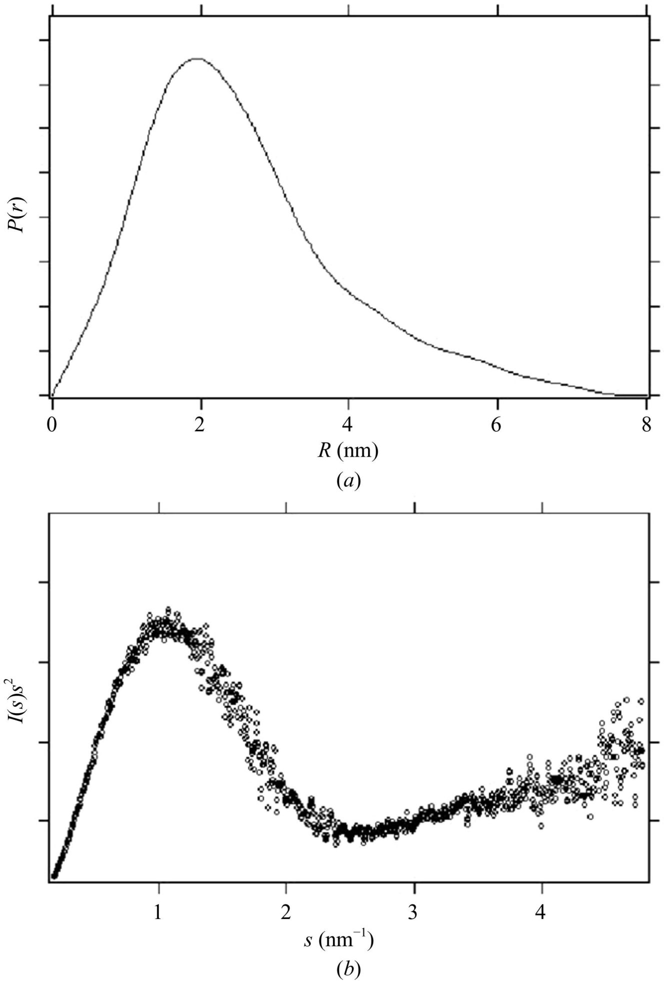


Figure 9
Distance distribution and Kratky plot. (*a*) Distance distribution $P(r)$ of the merged data with $D_{\rm max}$ = 8 nm. (*b*) Kratky plot of the merged data showing the characteristic increase of the signal in the high-s region (greater than 2.5 nm^{-1}) owing to partial disorder of the molecule.

4. Discussion and conclusion

MERS NTD$^+$ encompassing amino acids 1–164 was crystallized even though sequence prediction revealed the presence of a disordered N-terminal region from amino acid 1 to amino acid 37. Attempts to crystallize the NTD domain (from positions 37 to 164) did not lead to suitable crystals from commercial screens. It should be noted that during the writing of this study a preliminary crystallographic analysis of the NTD region encompassing amino acids 39–165 has been published (Wang *et al.*, 2015).

Our structural study therefore focused on the NTD$^+$ domain, which could provide additional structural information by combining crystal structure and SAXS analysis. Interestingly, the asymmetric unit of the crystal was made up of five NTD$^+$ molecules organized with an interaction between Pro33 and Trp43 of the neighbouring molecule (Fig. 5). This interaction highlights the role of the intrinsically disordered region (IDR) in the crystal organization and illustrates the need to design several constructs for a given domain of interest in order to improve the success rate at the level of protein expression or crystallization (Bignon *et al.*, 2013; Gräslund *et al.*, 2008).

Although divergent in size and sequence, the NTDs for which structures have been determined to date share a common fold with a central β-sheet and a β-hairpin extension. As expected, the three-dimensional structure of MERS-CoV NTD$^+$ displays the same structural organization. Its β-extension is positively charged (Fig. 4) and in several coronaviruses it has been proposed to contribute to RNA binding through electrostatic interactions with the phosphodiester groups of the RNA (Fan *et al.*, 2005). In addition to these interactions, the conserved hydrophobic core domain could interact with bases of the RNA through several aromatic residues (Huang *et al.*, 2004; McBride *et al.*, 2014; Spencer & Hiscox, 2006). Interestingly, in the crystal structure of MERS-CoV NTD$^+$, Trp43 located in the core domain was found to stack with a proline (Pro33) of the adjacent NTD$^+$, as in the crystal packing of IBV NTD (Huang *et al.*, 2004).

We therefore wanted to assess the oligomerization state of MERS-CoV NTD$^+$. SEC-MALLS analysis together with SAXS data demonstrated that, at least under our experimental conditions at pH 7.5, MERS NTD$^+$ behaves mainly as a monomer in solution. As previously reported, the oligomerization of the MHV N protein is partly mediated by an interaction between the NTD modules (Hurst *et al.*, 2009). Moreover, the NTD of the IBV N protein also has a weak propensity for dimerization (Fan *et al.*, 2005; Jayaram *et al.*, 2006). We can thus propose that, by itself, MERS-CoV NTD$^+$ cannot promote MERS-CoV N-protein assembly, but could contribute by a mechanism that still has to be unravelled to the formation of the RNP through NTD–NTD interaction after the CTD has already promoted the oligomerization.

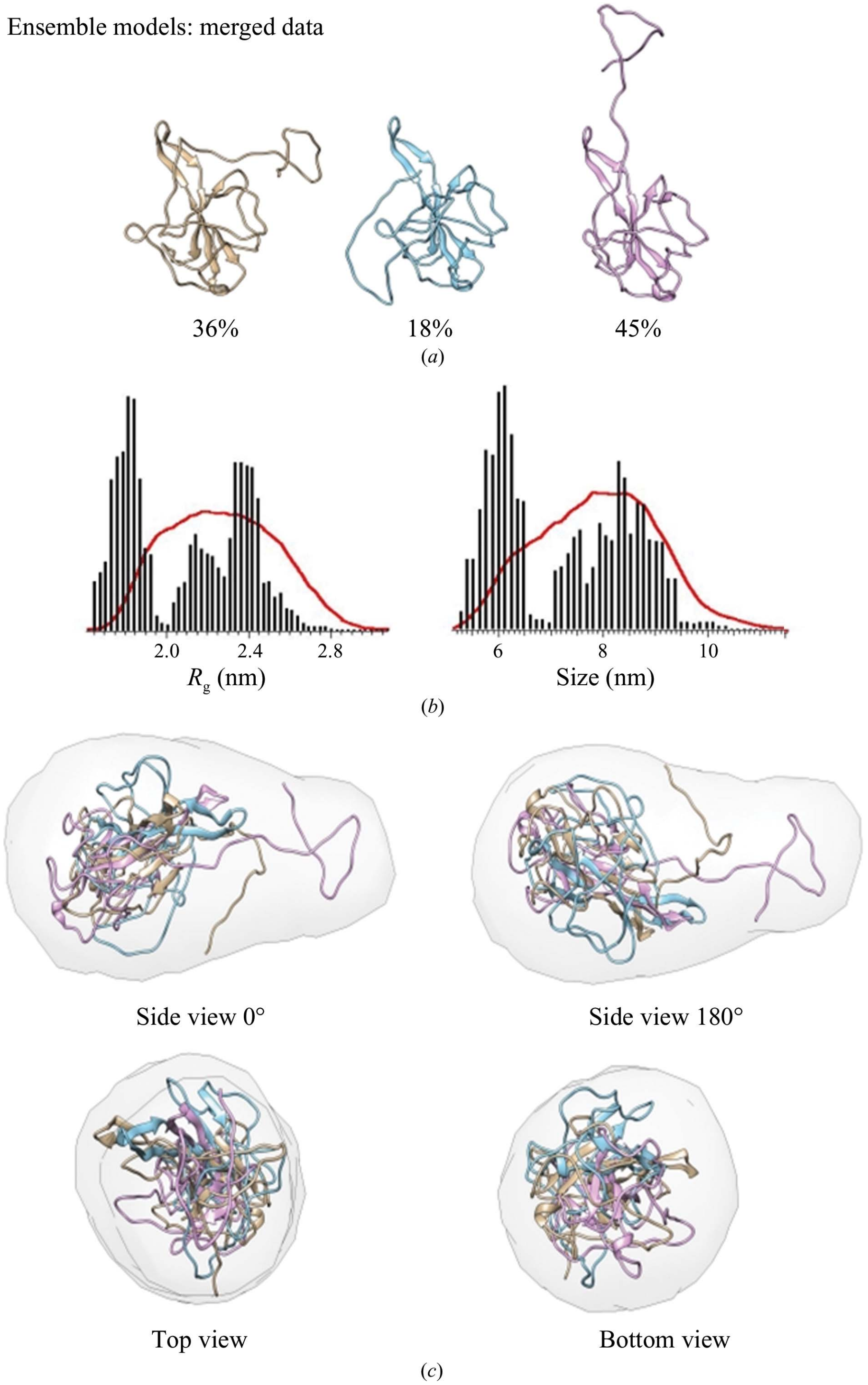


Figure 10
EOM analysis of a merged data set. (*a*) Three ensemble model structures resulting from *EOM* analysis of the merged data together with the corresponding percentage presence. (*b*) R_g and size distributions of the merged data. (*c*) Fit of the three structures within the filtered *DAMAVER* particle.

The role of Trp43 in RNA binding has also been addressed. MERS-CoV NTD$^+$ binds to a nonspecific RNA in the hundred-nanomolar range, and a W43A substitution did not significantly alter the affinity constant (data not shown). RNA-binding studies on MHV NTD revealed that aromatic residues located in $\beta3$ and $\beta5$ are strong determinants for sequence-specific RNA binding (Grossoehme *et al.*, 2009), whereas Trp43 precedes the $\beta1$ strand. However, analysis of RNA binding to the SARS-CoV NTD by NMR showed that even for a poly-adenine RNA the core region is still involved in the binding (Huang *et al.*, 2004). It is therefore possible that Trp43 could be involved in RNA binding, but not as a main contributor, and its single substitution does not allow perturbation of the function. At this stage, the role of Trp43, a conserved residue in the betacoronaviruses, still remains to be investigated.

SAXS experiments were undertaken to characterize MERS-CoV NTD$^+$ in solution. The SAXS results showed that the protein is monomeric and composed of a globular part and a flexible N-terminal region of 30 residues. This region is not completely disordered but points out from the globular part with two preferential conformations: an open conformation and a closed conformation. In all cases the first 20 amino acids form a loop without being randomly unfolded. Let us recall that this N-terminus is involved in RNA binding, working cooperatively with the other RNA-binding sites of the N protein (Chang *et al.*, 2009). However, it is possible that it plays other roles since viral IDRs tend to be involved in various interactions with host-cell components other than RNA, such as host membranes or proteins (Xue *et al.*, 2014).

Acknowledgements

This work was supported by the SILVER Large Scale Collaborative Project (grant agreement No. 260644) of the European Union Seventh Framework and the French Infrastructure for Integrated Structural Biology (FRISBI) ANR-10-INSB-05-01. The authors also thank the ESRF for provision of synchrotron beam time and the MASSIF and BM29 staff for data collection, as well as Stéphanie Blangy and Silvia Spinelli for their contribution to the experiments.

References

Adams, P. D. *et al.* (2010). *Acta Cryst.* D**66**, 213–221.
Azhar, E. I., Hashem, A. M., El-Kafrawy, S. A., Sohrab, S. S., Aburizaiza, A. S., Farraj, S. A., Hassan, A. M., Al-Saeed, M. S., Jamjoom, G. A. & Madani, T. A. (2014). *MBio*, **5**, e01450-14.
Bernadó, P., Mylonas, E., Petoukhov, M. V., Blackledge, M. & Svergun, D. I. (2007). *J. Am. Chem. Soc.* **129**, 5656–5664.
Bignon, C., Li, C., Lichière, J., Canard, B. & Coutard, B. (2013). *Acta Cryst.* D**69**, 2580–2582.
Chang, C.-K., Hou, M.-H., Chang, C.-F., Hsiao, C.-D. & Huang, T.-H. (2014). *Antiviral Res.* **103**, 39–50.
Chang, C.-K., Hsu, Y.-L., Chang, Y.-H., Chao, F.-A., Wu, M.-C., Huang, Y.-S., Hu, C.-K. & Huang, T.-H. (2009). *J. Virol.* **83**, 2255–2264.
Chang, C.-K., Sue, S.-C., Yu, T.-H., Hsieh, C.-M., Tsai, C.-K., Chiang, Y.-C., Lee, S.-J., Hsiao, H.-H., Wu, W.-J., Chang, W.-L., Lin, C.-H. & Huang, T.-H. (2006). *J. Biomed. Sci.* **13**, 59–72.
Chen, C.-Y., Chang, C.-K., Chang, Y.-W., Sue, S.-C., Bai, H.-I., Riang, L., Hsiao, C.-D. & Huang, T.-H. (2007). *J. Mol. Biol.* **368**, 1075–1086.
Chen, I.-J., Yuann, J.-M. P., Chang, Y.-M., Lin, S.-Y., Zhao, J., Perlman, S., Shen, Y.-Y., Huang, T.-H. & Hou, M.-H. (2013). *Biochim. Biophys. Acta*, **1834**, 1054–1062.
Clarkson, M. W., Lei, M., Eisenmesser, E. Z., Labeikovsky, W., Redfield, A. & Kern, D. (2009). *J. Biomol. NMR*, **45**, 217–225.
Corman, V. M., Ithete, N. L., Richards, L. R., Schoeman, M. C., Preiser, W., Drosten, C. & Drexler, J. F. (2014). *J. Virol.* **88**, 11297–11303.
Emsley, P. & Cowtan, K. (2004). *Acta Cryst.* D**60**, 2126–2132.
Fan, H., Ooi, A., Tan, Y. W., Wang, S., Fang, S., Liu, D. X. & Lescar, J. (2005). *Structure*, **13**, 1859–1868.
Fang, H.-J., Chen, Y.-Z., Li, M. S., Wu, M.-C., Chang, C.-L., Chang, C.-K., Hsu, Y.-L., Huang, T.-H., Chen, H.-M., Tsong, T.-Y. & Hu, C.-K. (2009). *Biophys. J.* **96**, 1892–1901.
Franke, D. & Svergun, D. I. (2009). *J. Appl. Cryst.* **42**, 342–346.
Geerlof, A. *et al.* (2006). *Acta Cryst.* D**62**, 1125–1136.
Gräslund, S., Sagemark, J., Berglund, H., Dahlgren, L. G., Flores, A., Hammarström, M., Johansson, I., Kotenyova, T., Nilsson, M., Nordlund, P. & Weigelt, J. (2008). *Protein Expr. Purif.* **58**, 210–221.
Grossoehme, N. E., Li, L., Keane, S. C., Liu, P., Dann, C. E., Leibowitz, J. L. & Giedroc, D. P. (2009). *J. Mol. Biol.* **394**, 544–557.
Harpaz, Y., Gerstein, M. & Chothia, C. (1994). *Structure*, **2**, 641–649.
He, R. *et al.* (2004). *Virus Res.* **105**, 121–125.
Huang, Q., Yu, L., Petros, A. M., Gunasekera, A., Liu, Z., Xu, N., Hajduk, P., Mack, J., Fesik, S. W. & Olejniczak, E. T. (2004). *Biochemistry*, **43**, 6059–6063.
Hurst, K. R., Koetzner, C. A. & Masters, P. S. (2009). *J. Virol.* **83**, 7221–7234.
Jayaram, H., Fan, H., Bowman, B. R., Ooi, A., Jayaram, J., Collisson, E. W., Lescar, J. & Prasad, B. V. (2006). *J. Virol.* **80**, 6612–6620.
Kabsch, W. (2010). *Acta Cryst.* D**66**, 125–132.
Konarev, P. V., Volkov, V. V., Sokolova, A. V., Koch, M. H. J. & Svergun, D. I. (2003). *J. Appl. Cryst.* **36**, 1277–1282.
Lantez, V., Dalle, K., Charrel, R., Baronti, C., Canard, B. & Coutard, B. (2011). *PLoS Negl. Trop. Dis.* **5**, e936.
Lin, S.-Y., Liu, C.-L., Chang, Y.-M., Zhao, J., Perlman, S. & Hou, M.-H. (2014). *J. Med. Chem.* **57**, 2247–2257.
Livingstone, C. D. & Barton, G. J. (1993). *Comput. Appl. Biosci.* **9**, 745–756.
Ma, Y., Tong, X., Xu, X., Li, X., Lou, Z. & Rao, Z. (2010). *Protein Cell*, **1**, 688–697.
Malet, H., Coutard, B., Jamal, S., Dutartre, H., Papageorgiou, N., Neuvonen, M., Ahola, T., Forrester, N., Gould, E. A., Lafitte, D., Ferron, F., Lescar, J., Gorbalenya, A. E., de Lamballerie, X. & Canard, B. (2009). *J. Virol.* **83**, 6534–6545.
McBride, R., van Zyl, M. & Fielding, B. C. (2014). *Viruses*, **6**, 2991–3018.
Müller, M. A. *et al.* (2015). *Lancet Infect. Dis.* **15**, 559–564.
Nelson, G. W., Stohlman, S. A. & Tahara, S. M. (2000). *J. Gen. Virol.* **81**, 181–188.
Notredame, C., Higgins, D. G. & Heringa, J. (2000). *J. Mol. Biol.* **302**, 205–217.
Peng, T.-Y., Lee, K.-R. & Tarn, W.-Y. (2008). *FEBS J.* **275**, 4152–4163.
Petoukhov, M. V. & Svergun, D. I. (2007). *Curr. Opin. Struct. Biol.* **17**, 562–571.
Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C. & Ferrin, T. E. (2004). *J. Comput. Chem.* **25**, 1605–1612.
Raj, V. S. *et al.* (2014). *Emerg. Infect. Dis.* **20**, 1339–1342.
Robert, X. & Gouet, P. (2014). *Nucleic Acids Res.* **42**, W320–W324.
Saikatendu, K. S., Joseph, J. S., Subramanian, V., Neuman, B. W., Buchmeier, M. J., Stevens, R. C. & Kuhn, P. (2007). *J. Virol.* **81**, 3913–3921.
Spencer, K. A. & Hiscox, J. A. (2006). *FEBS Lett.* **580**, 5993–5998.
Stohlman, S. A., Baric, R. S., Nelson, G. N., Soe, L. H., Welter, L. M. & Deans, R. J. (1988). *J. Virol.* **62**, 4288–4295.
Surjit, M., Kumar, R., Mishra, R. N., Reddy, M. K., Chow, V. T. K. & Lal, S. K. (2005). *J. Virol.* **79**, 11476–11486.
Svergun, D. I. (1992). *J. Appl. Cryst.* **25**, 495–503.
Takeda, M., Chang, C.-K., Ikeya, T., Güntert, P., Chang, Y.-H., Hsu, Y.-L., Huang, T.-H. & Kainosho, M. (2008). *J. Mol. Biol.* **380**, 608–622.
Tan, Y. W., Fang, S., Fan, H., Lescar, J. & Liu, D. X. (2006). *Nucleic Acids Res.* **34**, 4816–4825.
Tria, G., Mertens, H. D. T., Kachala, M. & Svergun, D. I. (2015). *IUCrJ*, **2**, 207–217.
Volkov, V. V. & Svergun, D. I. (2003). *J. Appl. Cryst.* **36**, 860–864.
Wang, Y.-S., Chang, C. & Hou, M.-H. (2015). *Acta Cryst.* F**71**, 977–980.
Wu, C.-H., Chen, P.-J. & Yeh, S.-H. (2014). *Cell Host Microbe*, **16**, 462–472.
Wu, C.-H., Yeh, S.-H., Tsay, Y.-G., Shieh, Y.-H., Kao, C.-L., Chen, Y.-S., Wang, S.-H., Kuo, T.-J., Chen, D.-S. & Chen, P.-J. (2009). *J. Biol. Chem.* **284**, 5229–5239.
Xue, B., Blocquel, D., Habchi, J., Uversky, A. V., Kurgan, L., Uversky, V. N. & Longhi, S. (2014). *Chem. Rev.* **114**, 6880–6911.
Yu, I.-M., Oldham, M. L., Zhang, J. & Chen, J. (2006). *J. Biol. Chem.* **281**, 17134–17139.