The ^1H , ^{15}N , and ^{13}C resonance assignments of the N-terminal domain of the nucleocapsid protein from the Middle East respiratory syndrome coronavirus

Talita Stelling de Araujo^{2,3} · Glaucé Moreno Barbosa^{1,3} · Karoline Sanches^{1,3} · Jéssica M. Azevedo^{2,3} · Katia Maria dos Santos Cabral^{2,3} · Marcius S. Almeida^{2,3} · Fabio C. L. Almeida^{1,3}

Received: 21 February 2021 / Accepted: 20 April 2021 / Published online: 29 April 2021
© The Author(s), under exclusive licence to Springer Nature B.V. 2021

Abstract

During the past 17 years, the coronaviruses have become a global public emergency, with the first appearance in 2012 in Saudi Arabia of the Middle East respiratory syndrome. Among the structural proteins encoded in the viral genome, the nucleocapsid protein is the most abundant in infected cells. It is a multifunctional phosphoprotein involved in the capsid formation, in the modulation and regulation of the viral life cycle. The N-terminal domain of N protein specifically interacts with transcriptional regulatory sequence (TRS) and is involved in the discontinuous transcription through the melting activity of double-stranded TRS (dsTRS).

Keywords MERS-CoV · Coronavirus · NMR · Nucleoprotein · Nucleocapsid · N protein

Biological context

During the past 17 years the three coronaviruses, severe acute respiratory syndrome (SARS) in November 2002, Middle East respiratory syndrome (MERS) in April 2012, and more recently the coronavirus disease (COVID-19) in December 2019, have become a global public emergency (Jiang et al. 2020; Singhal 2020).

The MERS-CoV was first identified in Saudi Arabia when isolated from an adult patient lung who was diagnosed with severe pneumonia and died of multiorgan failure (Nguyen et al. 2019). MERS-CoV, like SARS-CoV and SARS-CoV-2, is a member of the *Coronaviridae* family of the order Nidovirales. It is a large single-strand positive-sense RNA

with 30 kb which encodes four structural proteins: spike (S), envelope (E), membrane (M), and nucleocapsid (N). The N protein is the most abundant in infected cells (Carlson et al. 2020). It is composed of two domains: the dimerization C-terminal domain (N-CTD) and the RNA binding N-terminal domain (N-NTD). The N-CTD and N-NTD are linked by an intrinsically disordered region, which contains the Arg-Ser-rich region (SR region) and the phosphorylation site (Taskin Tok et al. 2017; Nguyen et al. 2019). This multifunctional phosphoprotein is involved in the capsid formation, in the modulation and regulation of the viral life cycle. N protein is directly involved in the discontinuous transcription process, acting as an RNA chaperone (Huang et al. 2004).

Coronaviruses are among the largest RNA viruses and they undergo a unique discontinuous transcription of the viral RNA into subgenomic mRNAs (sgmRNAs). At the 5' end of the genome is found the leader transcriptional regulatory sequence (TRS-L) and at the 5' end of each subgenomic RNA, the body transcriptional regulatory sequence (TRS-B). When the TRS-B is copied during the transcription process, the nascent negative-strand RNA is transferred to the TRS-L portion through a template switch, finalizing the transcription process. The N-NTD domain of N protein has been reported to specifically interact with TRS and catalyse

✉ Fabio C. L. Almeida
falmeida@bioqmed.ufrj.br

¹ National Center of Nuclear Magnetic Resonance (CNRMN), Institute of Medical Biochemistry (IBqM)—Center for Structural Biology and Bioimaging (CENABIO), Federal University of Rio de Janeiro, Rio de Janeiro, Brazil

² Protein Advanced Biochemistry (PAB), Institute of Medical Biochemistry (IBqM)—National Center for Structural Biology and Bioimaging (CENABIO), Federal University of Rio de Janeiro, Rio de Janeiro, RJ, Brazil

³ Rio BioNMR Network (RMNRio), Rio de Janeiro, Brazil

the template switch acting in the melting activity of dsTRS (Grossoehme et al. 2009).

Our group is involved in the study of the mechanism of specific recognition and melting activity of the N-terminal domain of human betacoronaviruses (Caruso et al. 2020; de Luna Marques et al. 2021). This work is part of an international effort to combat the Covid-19 pandemic (<https://covid19-nmr.de/>, (Altincekic et al. 2021). Here we report the ^1H , ^{15}N , and ^{13}C backbone and side-chain resonance assignments of the N-NTD domain of MERS-CoV without the SR region (N-NTD) and containing the SR region (N-NTD-SR). These assignments are fundamental to obtain structural information on the N-NTD and its Ser-Arg-rich region which in turn will contribute to the better understanding of coronavirus diseases.

Methods and experiments

Protein expression and purification

Two distinct constructs of MERS-CoV protein N were synthesized. The first one contained only the N-terminal domain of MERS-CoV protein N comprising residues 35 to 169 (N-NTD domain), and the other one including, besides the N-NTD domain, the Arg-Ser-rich sequence from residue 170 to 202 (N-NTD-SR domain). Both proteins were subcloned between NdeI and BamHI restriction sites in plasmid pET28a by Genscript Company.

Escherichia coli BL21 (DE3) was transformed with pET28a. One colony was picked and transferred to Luria Bertani (LB) medium. The bacteria were grown in minimal medium (M9) containing $^{15}\text{NH}_4\text{Cl}$ (1 g/L) and ^{13}C -glucose (3 g/L) for isotopic labeling and kanamycin (30 $\mu\text{g}/\text{mL}$) for bacterial selection. The protein expression was induced with 0.2 mM IPTG (isopropyl β -D-thiogalactoside), overnight at 18°C. Cells were centrifuged and the pellet was disrupted by ultrasonication in lysis buffer (50 mM Tris-HCl pH 8.0 containing 500 mM NaCl, 20 mM imidazole, 5% glycerol, 0.01 mg/ml DNase and 5mL SigmaFast protease inhibitor cocktail tablet 1x diluted). The lysate was centrifuged, and the supernatant was applied to a HisTrap FF column (GE Healthcare Life Sciences). The N-terminal domains of N protein were purified by nickel affinity chromatography, using washing buffer A (50 mM Tris-HCl, 500 mM NaCl, 20 mM imidazole, pH 8.0) and buffer B (50 mM Tris-HCl, 500 mM NaCl, 500 mM imidazole, pH 8.0). For His-tag removal the protein was cleaved overnight with TEV protease (TEV:protein 1:30 molar ratio) and the mixture was dialyzed against dialysis buffer (50 mM Tris-HCl pH 7.5, 0.5 mM EDTA and 1 mM DTT). After dialysis, a new cycle of nickel affinity chromatography was performed to improve the purity of the N-terminal domain of MERS protein N

and remove the tag cleaved by TEV. The sample containing the protein was concentrated at 5000 g, 10 min, in Amicon Ultra 15 10,000 MWCO, in the presence of PMSF 0.5 mM. The buffer of fractions containing the purified protein was changed by gel filtration chromatography (Superdex 75 column) using the buffer 20 mM sodium phosphate, 50 mM NaCl, 500 μM PMSF, 3 mM sodium azide, and 3 mM EDTA, pH 5.5. At the end of gel filtration, the sample was concentrated using an Amicon and 0.5 mM PMSF, 3 mM EDTA, and 3 mM azide was added to the sample. The samples for NMR were in 20 mM sodium phosphate, 50 mM NaCl, 500 μM PMSF, and 3 mM sodium azide.

NMR experiments

For all NMR experiments, we added 5% (v/v) D_2O to the sample. The triple resonance NMR spectra were acquired at 298 K on a Bruker 800 MHz AVANCE III spectrometer equipped with a pulse-field Z-axis gradient triple-resonance probe. We assigned the backbone resonances of ^{15}N - ^1H -HSQC spectrum (Fig. 1) through the triple resonance experiments HNCO, HNCA, CBCA(CO)NH, HNCACB, and HBHA(CO)NH (Whitehead et al. 1997). We assigned the side-chain resonance through ^{13}C -HSQC, (H)CCH-TOCSY, HCCH-TOCSY (Kay et al. 1993), and ^{15}N and ^{13}C -NOESY-HSQC (for both aliphatic and aromatic regions) experiments. The NOESY spectra were acquired at 298 K on a Bruker 900 MHz AVANCE IIIHD spectrometer equipped with pulse-field Z-axis gradient triple-resonance probes. For all experiments, we used the chemical shift of water proton as an internal reference for ^1H while ^{13}C and ^{15}N chemical shifts were referenced indirectly to water (Wishart et al. 1995). For the triple resonance measurements, we used non-uniform sampling (NUS) of the NMR data based on a 13% Poisson gap sampling schedule (Hyberts et al. 2012). The iterative soft threshold method was used for the spectral reconstruction (Hyberts et al. 2012). We processed the data using the NMRPipe software (Delaglio et al. 1995) and analysed it with CCPNMR Analysis (Vranken et al. 2005) both available on NMRbox (Maciejewski et al. 2017).

Assignment and data deposition

Chemical shift assignments ^1H , ^{15}N , and ^{13}C have been deposited in Biological Magnetic Resonance Bank (BMRB) under IDs 50,772 and 50,771 for MERS-CoV N-NTD and N-NTD-SR, respectively. Figure 1 shows the assigned 2D ^1H - ^{15}N HSQC spectrum of the MERS-CoV N-NTD domain and N-NTD-SR domain.

For the MERS-CoV N-NTD domain, we assigned 93.8% of the backbone nuclei ($^{13}\text{C}\alpha$, ^{13}CO , $\text{H}\alpha$, amide HN, and ^{15}N). We have a total of 96.3% $^{13}\text{C}\alpha$ and 92.8% $\text{H}\alpha$. For the ^{13}CHn aliphatic side chain moieties of the protein, 67.1% of

Fig. 1 Two-dimensional ^1H , ^{15}N -HSQC spectrum of uniformly $^{15}\text{N}/^{13}\text{C}$ -labelled MERS-CoV N-NTD domains at 298 K in 20 mM NaPi pH 5.5, 50 mM NaCl and 5% (v/v) D_2O . The labels show the

assigned backbone amino acid residues. MERS-CoV N-NTD spectrum is shown in red and N-NTD-SR domain in green

^{13}C and 70% of ^1H were assigned. For the ^{13}CHn aromatic side chain moieties of the protein, 40% of ^{13}C and 80.9% of ^1H were assigned. We assigned 98.3% $\text{C}\beta$ and 94.5% $\text{H}\beta$. We assigned 123 amide ^1HN (95.1%), 136 ^{15}N (86%) and 136 ^{13}CO (86%).

For the MERS-CoV N-NTD-SR domain, we assigned 87.5% of the backbone nuclei ($^{13}\text{C}\alpha$, ^{13}CO , $\text{H}\alpha$, amide HN , and ^{15}N). We have a total of 95.3% $^{13}\text{C}\alpha$ and 86.9% $\text{H}\alpha$. For the ^{13}CHn aliphatic side chain moieties of the protein, 57.7% of ^{13}C and 60.8% of ^1H were assigned. For the ^{13}CHn aromatic side chain moieties of the protein, 27.5% of ^{13}C and 57.1% of ^1H were assigned. We assigned 96% $\text{C}\beta$ and

84.9% $\text{H}\beta$. We assigned 156 amide ^1HN (89.7%), 171 ^{15}N (81.9%) and 171 ^{13}CO (80.1%).

From the resonance assignment we could compare the chemical shift derived order parameter (S^2), from the random coil index (Berjanskii and Wishart 2005) and secondary structure prediction using TalosN (Shen and Bax 2013). It is interesting to note subtle differences of backbone flexibility when the N-NTD and N-NTD-SR are compared. The Ser-Arg-rich region is flexible but contains a more ordered region around residue 183 (Fig. 2a). We observed secondary structure elements compatible with

Fig. 2 Protein dynamics and secondary structure prediction. **a** Random-coil index order parameter as a function of the residue number for MERS-CoV N-NTD (red) and N-NTD-SR (black). **b** TalosN secondary structure prediction of MERS-CoV N-NTD-SR as a function of residue number. **c** TalosN secondary structure prediction of MERS-CoV N-NTD as a function of residue number. In blue the

predicted probabilities for helix and in green for extended structure (β -strand) as a function of the residue number. In the top, the green rectangles represent the β -strands and the blue the helices, corresponding to the secondary structure in the crystal structure [PDB 6KL2 (Lin et al. 2020)]

the crystal structure (Papageorgiou et al. 2016) and a one-residue shift for β -strand 1 when N-NTD and N-NTD-SR are compared (Fig. 2b, c). Further studies are necessary to understand these structural and dynamical features.

Acknowledgements This work was supported by Fundação Carlos Chagas Filho de Amparo à Pesquisa do Estado do Rio de Janeiro (FAPERJ) and Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq). The authors KMSC and TSA gratefully acknowledge the post-doctoral fellowship and financial support from FAPERJ (grants E-26/260.005/2020 to TSA and E-26/260.002/2020 to KMSC). We acknowledge the National Center of Nuclear Magnetic Resonance (CNRMN) and the Protein Advance Biochemistry platform (PAB). We also acknowledge the Covid-19 NMR Consortium (<https://covid19-nmr.de/>) for providing an excellent environment for scientific discussions.

Declarations

Conflict of interest The authors declare that there are no conflicts of interest.

References

- Altincekic N, Korn SM, Qureshi NS et al (2021) Large-scale recombinant production of the SARS-CoV-2 proteome for high-throughput and structural biology applications. *Front Mol Biosci* 8:89. <https://doi.org/10.3389/FMOLB.2021.653148>
- Berjanskii MV, Wishart DS (2005) A simple method to predict protein flexibility using secondary chemical shifts. *J Am Chem Soc* 127:14970–14971. <https://doi.org/10.1021/ja054842f>
- Carlson CR, Asfaha JB, Ghent CM et al (2020) Phosphoregulation of phase separation by the SARS-CoV-2 N protein suggests

- a biophysical basis for its dual functions. *Mol Cell* 80:1092–1103. e4. <https://doi.org/10.1016/j.molcel.2020.11.025>
- Caruso IP, Sanches K, Da Poian AT et al (2020) Dynamics of the N-terminal domain of SARS-CoV-2 nucleocapsid protein drives dsRNA melting in a counterintuitive tweezer-like mechanism. *BioRxiv*. <https://doi.org/10.1101/2020.08.24.264465>
- de Luna Marques A, Caruso IP, Santana-Silva MC et al (2021) ^1H , ^{15}N and ^{13}C resonance assignments of the N-terminal domain of the nucleocapsid protein from the endemic human coronavirus HKU1. *Biomol NMR Assign*. <https://doi.org/10.1007/s12104-020-09998-9>
- Delaglio F, Grzesiek S, Vuister GW et al (1995) NMRPipe: a multi-dimensional spectral processing system based on UNIX pipes. *J Biomol NMR* 6:277–293. <https://doi.org/10.1007/BF00197809>
- Grossoehme NE, Li L, Keane SC et al (2009) Coronavirus N protein N-terminal domain (NTD) specifically binds the transcriptional regulatory sequence (TRS) and melts TRS-cTRS RNA duplexes. *J Mol Biol* 394:544–557. <https://doi.org/10.1016/j.jmb.2009.09.040>
- Huang Q, Yu L, Petros AM et al (2004) Structure of the N-terminal RNA-binding domain of the SARS CoV nucleocapsid protein. *Biochemistry* 43:6059–6063. <https://doi.org/10.1021/bi036155b>
- Hyberts SG, Milbradt AG, Wagner AB et al (2012) Application of iterative soft thresholding for fast reconstruction of NMR data non-uniformly sampled with multidimensional Poisson Gap scheduling. *J Biomol NMR* 52:315–327. <https://doi.org/10.1007/s10858-012-9611-z>
- Jiang F, Deng L, Zhang L et al (2020) Review of the clinical characteristics of Coronavirus disease 2019 (COVID-19). *J Gen Intern Med* 35:1545
- Kay LE, Xu GY, Singer AU et al (1993) A gradient-enhanced HCCH-TOCSY experiment for recording side-chain ^1H and ^{13}C correlations in H_2O samples of proteins. *J Magn Reson Ser B* 101:333–337. <https://doi.org/10.1006/jmrb.1993.1053>
- Lin SM, Lin SC, Hsu JN et al (2020) Structure-based stabilization of non-native protein–protein interactions of coronavirus nucleocapsid proteins in antiviral drug design. *J Med Chem* 63:3131–3141. <https://doi.org/10.1021/acs.jmedchem.9b01913>
- Maciejewski MW, Schuyler AD, Gryk MR et al (2017) NMRbox: a resource for biomolecular NMR computation. *Biophys J* 112:1529–1534. <https://doi.org/10.1016/j.bpj.2017.03.011>
- Nguyen TH, Van, Lichère J, Canard B et al (2019) Structure and oligomerization state of the C-terminal region of the Middle East respiratory syndrome coronavirus nucleoprotein. *Acta Crystallogr Sect D Struct Biol* 75:8–15. <https://doi.org/10.1107/S2059798318014948>
- Papageorgiou N, Lichère J, Baklouti A et al (2016) Structural characterization of the N-terminal part of the MERS-CoV nucleocapsid by X-ray diffraction and small-angle X-ray scattering. *Acta Crystallogr Sect D Struct Biol* 72:192–202. <https://doi.org/10.1107/S2059798315024328>
- Shen Y, Bax A (2013) Protein backbone and sidechain torsion angles predicted from NMR chemical shifts using artificial neural networks. *J Biomol NMR* 56:227–241. <https://doi.org/10.1007/s10858-013-9741-y>
- Singhal T (2020) A review of coronavirus disease-2019 (COVID-19). *Indian J Pediatr* 87:281–286
- Taskin Tok T, Tatar G, Tugba TT (2017) Structures and functions of coronavirus proteins: molecular modeling of viral nucleoprotein-international journal of virology & infectious diseases international journal of virology & infectious diseases. *Int J Virol Infect Dis* 2:1–7
- Vranken WF, Boucher W, Stevens TJ et al (2005) The CCPN data model for NMR spectroscopy: development of a software pipeline. *Proteins* 59:687–696. <https://doi.org/10.1002/prot.20449>
- Whitehead B, Craven CJ, Waltho JP (1997) Double and triple resonance NMR methods for protein assignment. *Methods Mol Biol* 60:29–52. <https://doi.org/10.1385/0-89603-309-0:29>
- Wishart DS, Bigam CG, Yao J et al (1995) ^1H , ^{13}C and ^{15}N chemical shift referencing in biomolecular NMR. *J Biomol NMR* 6:135–140. <https://doi.org/10.1007/BF00211777>

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