^1H , ^{15}N and ^{13}C resonance assignments of the N-terminal domain of the nucleocapsid protein from the endemic human coronavirus HKU1

Aline de Luna Marques^{1,2,4} · Icaro Putinhon Caruso^{2,3,4} · Marcos Caique Santana-Silva^{1,2,4} · Peter Reis Bezerra^{2,4} · Gabriela Rocha Araujo^{2,4} · Fabio Ceneviva Lacerda Almeida^{2,4} · Gisele Cardoso Amorim^{1,2,4}

Received: 12 October 2020 / Accepted: 11 December 2020 / Published online: 3 January 2021
© The Author(s), under exclusive licence to Springer Nature B.V. part of Springer Nature 2021

Abstract

Coronaviruses have become of great medical and scientific interest because of the Covid-19 pandemic. The hCoV-HKU1 is an endemic betacoronavirus that causes mild respiratory symptoms, although the infection can progress to severe lung disease and death. During viral replication, a discontinuous transcription of the genome takes place, producing the sub-genomic messenger RNAs. The nucleocapsid protein (N) plays a pivotal role in the regulation of this process, acting as an RNA chaperone and participating in the nucleocapsid assembly. The isolated N-terminal domain of protein N (N-NTD) specifically binds to the transcriptional regulatory sequences and control the melting of the double-stranded RNA. Here, we report the resonance assignments of the N-NTD of HKU1-CoV.

Keywords Human betacoronavirus · HKU1 · Nucleocapsid protein · N-terminal domain · NMR assignment

Biological context

Coronaviruses (CoV) are positive single-stranded RNA viruses, which are among the largest viral genomes, containing approximately 30 kb (Ou et al. 2017). CoVs are classified into four genera, based on their genomic organization and similarity, replication strategies, structural features, and

pathogenicity: alpha, beta, gamma, and delta coronavirus (Taskin Tok et al. 2017).

Since 2002, three betacoronaviruses have been causing concerns and gained scientific notoriety because of their high mortality and economic impacts: severe acute respiratory syndrome coronavirus (SARS-CoV), Middle East respiratory syndrome coronavirus (MERS-CoV), and SARS-CoV-2, which is causing the Covid-19 pandemic (Raoult et al. 2020). In addition to these more aggressive, there are endemic human CoVs (non-SARS-like CoV), such as hCoV-OC43 and hCoV-HKU1, associated with mild upper respiratory tract diseases (Ou et al. 2017).

Although the symptoms of endemic coronavirus infection are, in general, similar to those of common colds, hCoV-HKU1 and hCoV-OC43 can cause severe lung damage in more debilitated people (Hulswit et al. 2019). Thus, these endemic infections result in loss of quality of life, negatively impacting the economy (due to the temporary removal of symptomatic individuals) and cause preventable deaths (Raoult et al. 2020).

A unique and essential feature of CoVs is the discontinuous transcription of the genomic RNA, generating subgenomic messenger RNAs (sgmRNA). This process is regulated by the leader transcriptional regulatory sequence

Aline de Luna Marques and Icaro Putinhon Caruso contributed equally to this work.

✉ Gisele Cardoso Amorim
gisele.amorim@caxias.ufrj.br

- ¹ NUMPEX-BIO, Campus Duque de Caxias, Federal University of Rio de Janeiro (UFRJ), Duque de Caxias, RJ 25245-390, Brazil
- ² Institute of Medical Biochemistry Leopoldo de Meis (IBqM) and National Center for Structural Biology and Bioimaging (CENABIO), Federal University of Rio de Janeiro (UFRJ), Rio de Janeiro, RJ 21941-590, Brazil
- ³ Multiuser Center for Biomolecular Innovation (CMIB) and Department of Physics, Institute of Biosciences, Letters and Exact Sciences (IBILCE), São Paulo State University (UNESP), São José do Rio Preto, SP 15054-000, Brazil
- ⁴ Rio BioNMR Network, Rio de Janeiro, Brazil

(TRS-L), located at the 5' end of the genome, and the body TRSs (TRS-B), at the 5' end of each open reading frame. During transcription, once the TRS-B sequence is copied, the nascent negative-stranded RNA is transferred to the TRS-L and the transcription is then terminated. The presence of TRS allows the switching of templates by the replicase complex, during the negative-stranded RNA synthesis (Baker 2008).

The nucleocapsid protein (N) participates in the regulation of the discontinuous transcription, acting as an RNA chaperone, and also in the nucleocapsid assembly. It is a flexible-structured phosphoprotein that interacts with positive, negative, and double-stranded RNAs (Grossoehme et al. 2009). It has two structurally independent domains: the dimerization C-terminal domain (N-CTD) and the N-terminal RNA binding domain (N-NTD). N-CTD and N-NTD are linked by an intrinsically disordered region, which contains a phosphorylation site and a serine and arginine-rich motif (Taskin Tok et al. 2017). The isolated N-NTD specifically interacts with TRS and has been shown to display dsTRS melting activity (Grossoehme et al. 2009). The structure of the N-NTD of SARS-CoV, SARS-CoV-2, MERS-CoV, and hCoV-OC43 have already been elucidated (Chang et al. 2014; Kang et al. 2020).

In this work, we report the ^1H , ^{15}N , and ^{13}C backbone and side-chain resonance assignments of the N-NTD of hCoV-HKU1. To date, there is no structural information available for this protein and these data are important for the comparison with the hCoVs that cause severe acute respiratory symptoms, contributing to the understanding of diseases such as Covid-19.

Methods and experiments

Cloning, expression and purification

The DNA encoding 138 amino acid residues (56 to 193) of the N-terminal region of hCoV-HKU1 protein N was codon-optimized, synthesized, and subcloned between NdeI and BamHI sites in pET28a by GenScript®. *Escherichia coli* strain BL21(DE3) was transformed with this plasmid and selected colonies were grown in LB media for 6 h at 37 °C. The pre-culture was transferred to 100 mL of fresh minimal media (M9) and then expanded to 1 L of M9, supplemented with $^{15}\text{NH}_4\text{Cl}$ (1 g/L) and ^{13}C -glucose (3 g/L). All media contained 50 µg/mL of kanamycin. Protein expression was induced with 0.2 mM IPTG when optical density (O.D.) reached 0.6–0.8 at 600 nm. After 16 h of induction at 16 °C, the cells were centrifuged and the cell pellet was resuspended in 35 mL of lysis buffer: 50 mM Tris–HCl (pH 8), 500 mM NaCl, 20 mM imidazole, 10% glycerol, 0.01 mg/mL of DNAase and 0.1 mM PMSF. The cells were disrupted

by sonication. After centrifugation (10,000 rpm, 4 °C, 15 min), the supernatant was applied to HisTrap FF column (GE Healthcare Life Sciences) and purified by nickel affinity chromatography. The washing buffer contained 50 mM Tris–HCl (pH 8), 500 mM NaCl, 20 mM imidazole, and 10% glycerol. The bound proteins were eluted with a linear 20–500 mM imidazole gradient. The fractions containing the protein of interest were dialyzed against 2 L of 50 mM Tris–HCl (pH 8), 500 mM NaCl and 1 mM DTT overnight at 4 °C. Simultaneously, the protein was cleaved with TEV protease (molar ratio TEV:protein equals to 1:30), to remove the His-tag. After dialysis, the sample was reapplied to the HisTrap FF column, using the same purification buffers. The fractions containing the protein were concentrated with Millipore Centrprep (3000 MWCO) and the buffer composition was changed to 20 mM sodium phosphate (pH 6.8), 50 mM NaCl, 500 µM PMSF, 3 mM sodium azide, and 3 mM EDTA.

NMR spectroscopy

The NMR sample was supplied with 5% (v/v) D_2O . All NMR spectra were acquired at 298 K on a Bruker 800 MHz AVANCE III spectrometer. Data were processed using NMRPipe (Delaglio et al. 1995) software and analyzed with CCPNMR Analysis (Vranken et al. 2005). Backbone resonance assignments were determined using the ^{15}N – ^1H -HSQC spectrum (Fig. 1) together with a set of triple resonance experiments: HNCO, HN(CA)CO, HNCA, CBCA(CO)NH, HNCACB, and HBHA(CO)NH. Side-chain resonance assignments were performed using ^{13}C -HSQC, (H)CCH-TOCSY, HCCH-TOCSY, and ^{15}N and ^{13}C -NOESY-HSQC experiments. The chemical shift of water proton was used as an internal reference for ^1H while ^{13}C and ^{15}N chemical shifts were referenced indirectly to water (Wishart et al. 1995).

Assignments and data deposition

Chemical shift assignments (^1H , ^{15}N , and ^{13}C) have been deposited in the BioMagResBank (<http://www.bmrb.wisc.edu>) under the ID 50507. We almost fully assigned the ^1H , ^{15}N , and ^{13}C resonances, except for N2 Nε/He, S8 N/HN, F16 Hδ, D21 N/HN, K23 N/HN, Y82 Hδ/He, T111 Hγ2/Cγ2, E133 Hβ/Hγ, and G123 N/HN. We did not assign histidine Ce and some phenylalanine's He/Cε. We assigned 96% of the backbone NHs, all Cα and Cβ, 93.6% of C', 70.8% of side-chain carbons, and 86.6% of side-chain hydrogens. G83 has unusual HN and HA chemical shifts (3.04 and 2.57 ppm, respectively), probably due to interaction to the aromatic ring of W91 (He), as confirmed by the presence of a NOE crosspeak in the ^{15}N -NOESY spectrum.

Fig. 1 ^{15}N - ^1H -HSQC spectrum of the N-terminal domain (residues 56 to 193) of the nucleocapsid protein of hCoV-HKU1. Peak assignments are indicated by residue number and the one-letter amino acid code

The chemical shifts were analyzed with TalosN software (Fig. 2) (Shen and Bax 2013). The random coil index (RCI- S^2) values (Berjanskii and Wishart 2005) (Fig. 2a) indicated 4 dynamical regions: N-terminal, C-terminal, the loop between residues 49 to 64, known as the finger, and the loop between residues 102 and 115. These flexible regions confer the necessary plasticity for the protein to interact

with cellular partners such as proteins or RNA (Caruso et al. 2020). TalosN predicted 5 β -strands with high confidence, named $\beta 1$, $\beta 2$, $\beta 5$, $\beta 6$, and $\beta 7$. It also predicted two small β -strands within the finger, named $\beta 3$ and $\beta 4$. For these two β -strands, the confidence is small, probably due to the flexibility of the loop, but they are present in the structures of homologous proteins (Kang et al. 2020).

Fig. 2 Secondary Structure Prediction using TalosN software.

a RCI-S2 values indicate a flexible region between residues 49 to 64; **b** Predicted Secondary Structures confirmed that beta strands are predominant in hCoV N-NTD protein

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 author IPC gratefully acknowledges the post-doctoral fellowship and financial support from FAPERJ (202.279/2018) and the PROPe UNESP. We also acknowledge the Covid-19 NMR Consortium (<https://covid19-nmr.de/>) for providing an excellent environment for scientific discussions.

Compliance with ethical standards

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

References

- Baker SC (2008) Coronaviruses: molecular biology. *Encycl Virol*. <https://doi.org/10.1016/B978-012374410-4.00769-X>
- 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>
- Caruso ÍP, 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>
- Chang CK, Hou MH, Chang CF et al (2014) The SARS coronavirus nucleocapsid protein—forms and functions. *Antiviral Res* 103:39–50
- 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
- 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>
- Hulswit RJG, Lang Y, Bakkers MJG et al (2019) Human coronaviruses OC43 and HKU1 bind to 9-O-acetylated sialic acids via a conserved receptor-binding site in spike protein domain A. *Proc Natl Acad Sci USA* 116:2681–2690. <https://doi.org/10.1073/pnas.1809667116>
- Kang S, Yang M, Hong Z et al (2020) Crystal structure of SARS-CoV-2 nucleocapsid protein RNA binding domain reveals potential unique drug targeting sites. *Acta Pharm Sin B*. <https://doi.org/10.1016/j.apsb.2020.04.009>
- Ou X, Guan H, Qin B et al (2017) Crystal structure of the receptor binding domain of the spike glycoprotein of human betacoronavirus

- HKU1. *Nat Commun* 8:1–10. <https://doi.org/10.1038/ncomms15216>
- Raoult D, Zumla A, Locatelli F et al (2020) Coronavirus infections: epidemiological, clinical and immunological features and hypotheses. *Cell Stress* 4:66–75. <https://doi.org/10.15698/cst2020.04.216>
- 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>
- Taskin Tok T, Tatar G, Tugba TT (2017) Structures and Functions of Coronavirus Proteins: Molecular Modeling of Viral Nucleo-protein-*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
- 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.