Biochimica et Biophysica Acta 1834 (2013) 1054–1062

Contents lists available at SciVerse ScienceDirect

Biochimica et Biophysica Acta

journal homepage: www.elsevier.com/locate/bbapap

Crystal structure-based exploration of the important role of Arg106 in the RNA-binding domain of human coronavirus OC43 nucleocapsid protein

I.-Jung Chen [a,1], Jeu-Ming P. Yuann [b,1], Yu-Ming Chang [c,1], Shing-Yen Lin [d,1], Jincun Zhao [e], Stanley Perlman [e], Yo-Yu Shen [a], Tai-Huang Huang [f], Ming-Hon Hou [a,d,g,*]

[a] Department of Life Science, National Chung Hsing University, Taichung, 40254, Taiwan
[b] Department of Biotechnology, Ming Chuan University, Taoyuan County, 33348, Taiwan
[c] Institute of Biological Chemistry, Academia Sinica, Taipei, 11529, Taiwan
[d] Institute of Genomics and Bioinformatics, National Chung Hsing University, Taichung, 40254, Taiwan
[e] Department of Microbiology, The University of Iowa, Iowa City, IA 52240, USA
[f] Institute of Biomedical Science, Academia Sinica, Taipei, 115, 11529, Taiwan
[g] Biotechnology Center, National Chung Hsing University, Taichung, 40254, Taiwan

ARTICLE INFO

Article history:
Received 1 October 2012
Received in revised form 6 February 2013
Accepted 5 March 2013
Available online 15 March 2013

Keywords:
Coronaviruses
Nucleocapsid protein
RNA-binding
X-ray crystallography
Arginine106
Virus replication

ABSTRACT

Human coronavirus OC43 (HCoV-OC43) is a causative agent of the common cold. The nucleocapsid (N) protein, which is a major structural protein of CoVs, binds to the viral RNA genome to form the virion core and results in the formation of the ribonucleoprotein (RNP) complex. We have solved the crystal structure of the N-terminal domain of HCoV-OC43 N protein (N-NTD) (residues 58 to 195) to a resolution of 2.0 Å. The HCoV-OC43 N-NTD is a single domain protein composed of a five-stranded β-sheet core and a long extended loop, similar to that observed in the structures of N-NTDs from other coronaviruses. The positively charged loop of the HCoV-OC43 N-NTD contains a structurally well-conserved positively charged residue, R106. To assess the role of R106 in RNA binding, we undertook a series of site-directed mutagenesis experiments and docking simulations to characterize the interaction between R106 and RNA. The results show that R106 plays an important role in the interaction between the N protein and RNA. In addition, we showed that, in cells transfected with plasmids that encoded the mutant (R106A) N protein and infected with virus, the level of the matrix protein gene was decreased by 7-fold compared to cells that were transfected with the wild-type N protein. This finding suggests that R106, by enhancing binding of the N protein to viral RNA plays a critical role in the viral replication. The results also indicate that the strength of N protein/RNA interactions is critical for HCoV-OC43 replication.

1. Introduction

The OC43 strain of the beta coronavirus family (HCoV-OC43), which was first identified in the 1960s, is responsible for ~20% of all "common colds" in humans [1,2]. Although HCoV-OC43 infections are generally mild, more severe upper and lower respiratory tract infections, such as bronchiolitis and pneumonia, have been documented, particularly in infants, elderly individuals, and immunocompromised patients [1,3,4]. Moreover, there have been reports that clusters of HCoV-OC43 infections cause pneumonia in otherwise healthy adults [2,5]. Several studies have also reported that both neurotropism and neuroinvasion of HCoV, particularly the OC43 strain, are associated with multiple sclerosis [6].

CoV particles have an irregular shape defined by an outer envelope with a distinctive club-like shape. Peplomers on the outer envelope give the virus a crown-like (coronal) appearance [7]. The viral genomes of coronaviruses consist of approximately 30 kb of positive sense, single-stranded RNA. These genomes contain several genes encoding structural and nonstructural proteins that are required for the production of progeny virions [1]. The virion envelope that surrounds the nucleocapsid contains the S (spike), M (matrix), and E (envelope) structural proteins. A third glycoprotein, HE (hemagglutinin-esterase), is present in most betacoronaviruses [8,9]. The virion contains a helical nucleocapsid, which consists of the N protein bound to viral RNA. The N protein is the major structural protein of CoVs [10–12]. The formation of the RNP is important for maintaining the RNA in an ordered conformation that is suitable for viral genome replication and transcription [13–16]. Previous studies have shown that the CoV N protein is involved in the regulation of cellular processes, such as gene transcription, actin reorganization, host cell cycle progression, and apoptosis [17–20]. It has also been shown to act as an RNA chaperone [21]. Moreover, the

* Corresponding author at: Biotechnology Center, National Chung Hsing University, Taichung, 40254, Taiwan. Tel.: +886 4 22840338x7011; fax: +886 4 22859329.
E-mail address: mhho@dragon.nchu.edu.tw (M.-H. Hou).
[1] These authors contributed equally to this work.

1570-9639/$ – see front matter.
http://dx.doi.org/10.1016/j.bbapap.2013.03.003

N protein is an important diagnostic marker and is an antigen for the host antibody and T cell immune responses [16,22–24].

Previous studies have revealed that the N- and C-terminal domains of the CoV N proteins, including those of the SARS-CoV, murine hepatitis virus (MHV), and avian infectious bronchitis virus (IBV), are responsible for RNA binding and oligomerization, respectively [25–31]. The central region of the N protein has also been shown to contain an RNA-binding region and primary phosphorylation sites [32–34]. Phosphorylation of the N protein has been shown to play an important role in virus biology [35,36]. To clarify the molecular mechanism of RNP formation in CoVs, the structures of truncated fragments of the N protein, including the N-terminal and C-terminal domains, were investigated [27,37–39]. Despite the conservation of some motifs, the CoV N proteins from various different strains often exhibit different properties, due primarily to their low sequence homology [25].

The N protein of HCoV-OC43, which has a molecular weight of ~50 kDa, is highly basic (pI, 10.0) and exhibits strong hydrophilicity [40]. It also shows only 26–30% amino acid identity to N proteins from other CoV strains [25]. In this study, because of its importance for RNA binding, we chose to characterize the structure of the N-terminus of the HCoV-OC43 N protein using X-ray crystallography. Using the crystal structure and the surface charge distribution of the N-terminus, we were able to further investigate the interactions between the HCoV-OC43 N protein and the RNA molecule. Furthermore, we identified an important role of R106 in the binding of the HCoV-OC43 N protein to RNA using SPR analysis and site-directed mutagenesis. Finally, we present a structural model of the HCoV-OC43 N protein in complex with RNA, which clearly demonstrates the critical role of R106.

2. Materials and methods

All drugs and reagents were purchased from Sigma Chemical Co. (St. Louis, MO) unless otherwise specified. All oligoribonucleotides and oligodeoxyribonucleotides were synthesized using an automated DNA synthesizer and purified by gel electrophoresis. Biotin-linked oligomers were synthesized by incorporating the biotin synthon at the 5′-end; the oligomers were then immobilized to streptavidin-coated biosensor chips for the SPR experiments.

2.1. Expression and purification of the full-length and truncated N proteins

The templates for the HCoV-OC43 N protein were kindly provided by the Institute of Biological Chemistry, Academia Sinica (Taipei, Taiwan). To generate both a full-length versions and the N-terminal domain of the recombinant N protein, the N protein gene was amplified by polymerase chain reaction (PCR) from a plasmid (pGENT) using various primers. The PCR products were digested with *Nde*I and *Xho*I, and the DNA fragments were cloned into pET28a (Novagen) using T4 ligase (NEB). The bacteria transformed with the resultant plasmid were grown in culture. The expression of the recombinant N proteins was induced by supplementing the culture medium with 1 mM IPTG, followed by incubation at 10 °C for 24 h. After harvesting by centrifugation (8,000 g, 10 min, 4 °C), the bacterial pellets were lysed (50 mM Tris-buffered solution at pH 7.3, 150 mM NaCl, 0.1% CHAPS, and 15 mM imidazole). The soluble proteins were obtained from the supernatant following centrifugation (13,000 rpm, 40 min, 4 °C). The methods used for the protein purification have been described previously [41]. Full-length and truncated N proteins carrying a His_6-tag at the N-termini were purified using a Ni-NTA column (Novagen) with an elution gradient ranging from 15 to 300 mM imidazole. Fractions were collected and dialyzed against low-salt buffer. The protein concentrations of the resulting samples were determined using the Bradford method with Bio-Rad protein assay reagents.

2.2. Crystallization

The initial crystallization experiments were set up using Qiagen crystal screens JCSG+ Suite and PACT Suite [42] using the sitting-drop vapor-diffusion method in accordance with our previously described protocol [43]. Each of the crystallization solutions (2 μl) obtained from the screen was mixed with 1.5 μl of purified protein solution (8 mg/ml) and 0.5 μl of 40% hexanediol at room temperature (~298 K) against 400 μl solution in each well of a Cryschem plate. The conditions were refined through seven cycles, and the crystals were grown in a solution containing 0.25 M SPG buffer (pH 6.0) and 25% PEG1500 and then equilibrated at 293 K against 400 μl of the precipitation solution. The SPG buffer was prepared by mixing succinic acid (Sigma), sodium dihydrogen phosphate, (Merck) and glycine (Merck) in a 2:7:7 molar ratio and then adjusting with sodium hydroxide to obtain a pH of 6.0 [44]. The crystals appeared within two weeks, and the largest crystal grew to dimensions of approximately 200 × 100 × 100 μm. The crystals were then soaked in reservoir solution containing 30% (v/v) glycerol as the cryoprotectant prior to being flash-cooled in a nitrogen-gas stream at 100 K. High resolution X-ray data were collected using a synchrotron radiation source. The complete dataset was collected at the beamline BL13B1 in the NSRRC using a ADSC Q315r detector. The crystallographic data integration and reduction were performed using the software package HKL2000 [45]. The crystallographic statistics are listed in Table 1. The Matthews coefficient of 2.06 $Å^3$/Da, which was calculated using *Matthews* (Collaborative Computational Project, 1994) [46], suggested that this structure is likely to represent one molecule in an asymmetric unit. The solvent content was 40.26%. The N-terminal domain of the N protein obtained from the SARS-CoV (PDB ID: 2ofz) was chosen as the initial search model due to its low E-value of 1×10^{-23}. The first molecular replacement trial was performed using the PERON automated interface at the Protein Tectonics Platform (PTP), RIKEN SPring-8 Center, Japan [47]. The best results were obtained using the *MOLREP* program [48]. A single and unambiguous solution for the rotation and translation function was found with the reflections in the resolution range of 3.0–30 Å, a final correlation coefficient of 0.79 and an R factor of 0.44. The structure was refined further using the Crystallography & NMR system (CNS) [49] and deposited in the Protein Data Bank (PDB ID: 4j3k).

2.3. Site-directed mutagenesis

The single mutants (R106A, R106K, R106Q, R106E, R107A, K110A, and R117A) were constructed using a QuikChange™ kit (Stratagene) with a plasmid containing an open reading frame that encodes the full-length HCoV-OC43 N protein as the template for mutagenesis.

Table 1
Crystallographic and refinement data for the HCoV-OC43 N-NTD.

Data collection	
Wavelength (Å)	1.0
Space group	$P6_5$
Unit cell parameter (Å, °)	a = b = 81.57, c = 42.55 α = β = 90°, γ = 120°
Resolution limit (Å)	30–2.0 (2.07–2.00)
Completeness (%)	99.8 (100)
Unique reflections	11,131 (1086)
Redundancy	8.2 (8.5)
R_{merge}	0.037 (0.166)
(I/σ (I))	49.4 (13.77)
Refinement statistics	
R_{cryst} (%)	0.200
R_{free} (%)	0.214
RMSD bond lengths (Å)	0.013
RMSD bond angles (°)	1.978
Most favored region (%)	97.0
Generally allowed region (%)	1.5
Others (%)	1.5
Average B-factor	43.8