Acta Crystallographica Section F
Structural Biology and Crystallization Communications
ISSN 1744-3091

Crystallization and preliminary X-ray diffraction analysis of the N-terminal domain of human coronavirus OC43 nucleocapsid protein

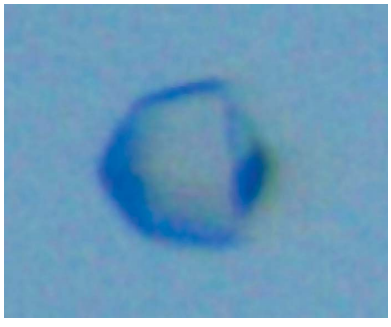
I-Jung Chen,[a]‡ Chia-Cheng Chou,[b]‡ Chia-Ling Liu,[a] Cheng-Chung Lee,[c] Lou-Sing Kan[d,e] and Ming-Hon Hou[a,f]*

[a]Department of Life Sciences, National Chung Hsing University, Taichung 402, Taiwan, [b]National Synchrotron Radiation Research Center, Hsinchu 300, Taiwan, [c]Institute of Biological Chemistry, Academia Sinica, Taipei 115, Taiwan, [d]Institute of Bioengineering, Tatung University, Taipei 104, Taiwan, [e]Institute of Chemistry, Academia Sinica, Taipei 11529, Taiwan, and [f]Institute of Genomics and Bioinformatics, National Chung Hsing University, Taichung 402, Taiwan

‡ These authors contributed equally to this paper.

Correspondence e-mail: mhho@dragon.nchu.edu.tw

Received 13 March 2010
Accepted 13 May 2010



The N-terminal domain of nucleocapsid protein from human coronavirus OC43 (HCoV-OC43 N-NTD) mostly contains positively charged residues and has been identified as being responsible for RNA binding during ribonucleocapsid formation in the coronavirus. In this study, the crystallization and preliminary crystallographic analysis of HCoV-OC43 N-NTD (amino acids 58–195) with a molecular weight of 20 kDa are reported. HCoV-OC43 N-NTD was crystallized at 293 K using PEG 1500 as a precipitant and a 99.9% complete native data set was collected to 1.7 Å resolution at 100 K with an overall $R_{\rm merge}$ of 5.0%. The crystals belonged to the hexagonal space group $P6_5$, with unit-cell parameters $a = 81.57$, $c = 42.87$ Å. Solvent-content calculations suggest that there is likely to be one subunit of N-NTD in the asymmetric unit.

1. Introduction

Human coronavirus OC43 (HCoV-OC43) is responsible for approximately 20% of all colds (Kaye *et al.*, 1971; Lai & Cavanagh, 1997). Although HCoV-OC43 infections are generally mild, more severe upper and lower respiratory-tract infections such as bronchiolitis and pneumonia have been characterized, especially in infants, elderly individuals and immunocompromised patients (El-Sahly *et al.*, 2000; Gagneur *et al.*, 2002; St-Jean *et al.*, 2004). According to serological cross-reactivity, HCoV-OC43 is a representative of the class II coronaviruses. The RNA genomes of coronaviruses comprise of several genes that encode several structural and nonstructural proteins and are arranged in the order 5′-pol (polymerase)–S (spike)–E (envelope)–M (matrix)–N (nucleocapsid)-3′ (Navas-Martin & Weiss, 2004). The virion envelope surrounding the nucleocapsid contains the structural proteins S, M, E and N. Some of them contain a third glycoprotein, HE (haemagglutinin-esterase), which is present in most class II coronaviruses. The primary function of the HCoV N protein is to recognize a stretch of RNA that serves as a packaging signal and leads to the formation of the ribonucleoprotein (RNP) complex during assembly (Lai & Cavanagh, 1997). RNP may be important in keeping the RNA in an ordered conformation suitable for replication and transcription of the viral genome (Lai, 2003; Nelson *et al.*, 2000; Huang *et al.*, 2004; Navas-Martin & Weiss, 2004). In addition, the N protein is also an important diagnostic marker and is the most immunodominant antigen in infected hosts (Chan *et al.*, 2005; Woo *et al.*, 2004).

The N protein of HCoV-OC43 has a molecular weight of 49.5 kDa and a pI of 10.0 and is a highly basic protein with a high hydrophilicity (Hogue & Brian, 1986; Pohl-Koppe *et al.*, 1995). We have previously shown that the N-terminal domain of HCoV-OC43 (N-NTD) contains most of the positively charged residues that are responsible for RNA binding, while the C-terminal domain (N-CTD) mainly acts as an oligomerization module to form a capsid (Huang *et al.*, 2009; Saikatendu *et al.*, 2007; Jayaram *et al.*, 2006; Fan *et al.*, 2005). The central disordered region of the N protein has also been shown to contain an RNA-binding region (Chang *et al.*, 2009). In order to clarify the mechanism by which the N protein of HCoV-OC43 binds to nucleic acids, we have undertaken the determination of the crystal structure of the N-terminal domain of HCoV-OC43 (residues 58–195).

Table 1
Data-collection statistics for HCoV-OC43 N-NTD crystals.

Values in parentheses are for the highest resolution shell.

X-ray source	BL13B1, NSRRC
Wavelength (Å)	1.0
Space group	$P6_5$
Unit-cell parameters (Å)	a = 81.57, c = 42.87
Resolution limits (Å)	30–1.70 (1.76–1.70)
Total reflections	149206
Unique reflections	18040
Completeness (%)	99.9 (99.9)
Redundancy	8.3 (7.7)
$R_{\rm merge}$† (%)	5.0 (76.1)
$\langle I/\sigma(I)\rangle$	27.8 (2.46)

† $R_{\rm merge} = \sum_{hkl}\sum_i |I_i(hkl) - \langle I(hkl)\rangle| / \sum_{hkl}\sum_i I_i(hkl)$.

2. Experimental methods

2.1. Expression and purification of HCoV-OC43 N-NTD

The templates for the HCoV-OC43 N protein were kindly provided by the Institute of Biological Chemistry, Academia Sinica (Taipei, Taiwan). In order to generate a truncated form of the recombinant HCoV-OC43 N-NTD, the gene encoding the N protein was amplified by the polymerase chain reaction (PCR) using *Pfu* polymerase with the forward primer 5′-CGCTAT**GAATTC**AATGTTGTACCCTAC-TATTCTTGGTTC-3′ and the reverse primer 5′-ACAAC**GCTCGA-G**AGCAGACCTTCCTGAGCCTTCAATAT-3′. The PCR products were digested with *Eco*RI and *Xho*I and the resulting DNA fragments were cloned into pET28a (Novagen) using T4 ligase (NEB). The recombinant plasmid was transformed into *Escherichia coli* BL21-RIL using the heat-shock method. The cells were grown at 310 K in Luria–Bertani medium containing 50 mg l^{-1} kanamycin. Protein expression was induced at an OD_{600} of 0.6 by the addition of 1 m*M* isopropyl β-D-1-thiogalactopyranoside (IPTG) at 283 K for 24 h. The expressed N-NTD contains N-terminal (MGSSHHHHH-HSSGLVPRGSHMASMTGGQQMGRGSEF) and C-terminal (LE-HHHHHH) tags. After harvesting the bacteria by centrifugation (6000 rev min^{-1}, 30 min, 277 K), the bacterial pellets were resuspended in sonication buffer (50 m*M* Tris-buffered solution pH 7.5, 150 m*M* NaCl and 15 m*M* imidazole) and lysed by sonication for 30 min (3 s pulse on and 6 s pulse off). Soluble proteins were obtained from the supernatant by centrifugation (13 000 rev min^{-1}, 30 min, 277 K). N-NTD carrying His_6 tags at the N-terminus and C-terminus was purified using an Ni–NTA column (Novagen) with an elution gradient from 15 to 250 m*M* imidazole in buffer solution (50 m*M* Tris-buffered solution pH 7.5 and 150 m*M* NaCl). Fractions containing N-NTD were pooled and collected at 250 m*M* imidazole and dialyzed against 50 m*M* Tris-buffered solution pH 7.5 containing 150 m*M* NaCl (Fig. 1). The purified N-NTD was concentrated to 8 mg ml^{-1} in 50 m*M* Tris–HCl pH 7.5 containing 255 m*M* NaCl prior to crystallization.

2.2. Crystallization

Initial crystallization experiments were set up using the Qiagen JCSG+ Suite and PACT Suite crystal screens (Newman, 2005) using the sitting-drop vapour-diffusion method in accordance with our previously described protocol (Chou & Hou, 2008). Each of the crystallization conditions (2 µl) from the screening kits was mixed with 1.5 µl purified protein solution (8 mg ml^{-1}) and 0.5 µl 40% hexanediol at room temperature (∼298 K) and equilibrated against 400 µl solution in the well of a Cryschem plate. The conditions were refined over seven cycles and crystals were grown in solution containing 0.25 *M* succinic acid–phosphate–glycine (SPG) buffer pH 6.0, 25% PEG 1500 equilibrated at 293 K against 400 µl 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 was adjusted to pH 6.0 with sodium hydroxide (Newman, 2004). The crystals appeared within two weeks and the largest crystal grew to dimensions of approximately 200 × 100 × 100 µm (Fig. 2).

2.3. X-ray data collection and processing

Crystals were soaked in reservoir solution containing 30%(*v*/*v*) glycerol as a cryoprotectant prior to flash-cooling in a nitrogen-gas stream at 100 K. High-resolution X-ray data were collected using a synchrotron-radiation source. A preliminary diffraction image was obtained on NSRRC (National Synchrotron Radiation Research

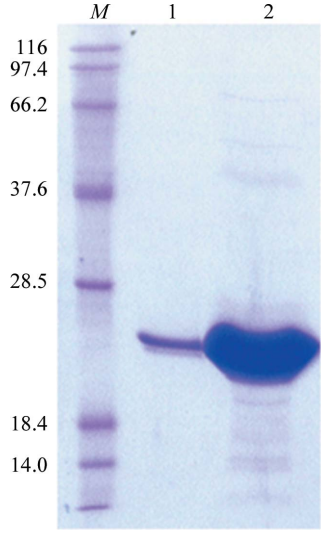


Figure 1
SDS–PAGE analysis of HCoV-OC43 N-NTD stained with Coomassie Brilliant Blue. Lane *M*, protein markers (kDa); lane 1, purified HCoV-OC43 N-NTD; lane 2, concentrated HCoV-OC43 N-NTD after dialysis.

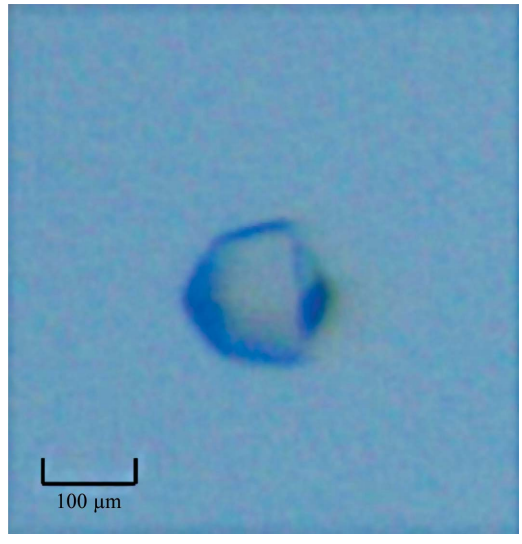


Figure 2
Crystals of HCoV-OC43 N-NTD obtained with 25% PEG 1500 as a precipitant at pH 6.0 by the sitting-drop vapour-diffusion method. The approximate dimensions of the crystal are 200 × 100 × 100 µm.

Center, Hsinchu, Taiwan) contract beamline BL12B2 at SPring-8 (Hyogo, Japan) using an ADSC Q210r detector. The complete data set was collected on beamline BL13B1 at NSRRC using an ADSC Q315r detector. The crystal-to-detector distance was 150 mm. The oscillation width and exposure time for each frame were 1° and 10 s, respectively. Crystallographic data integration and reduction were performed using the *HKL*-2000 program package (Otwinowski & Minor, 1997). The crystallographic statistics of data collection for N-NTD are listed in Table 1.

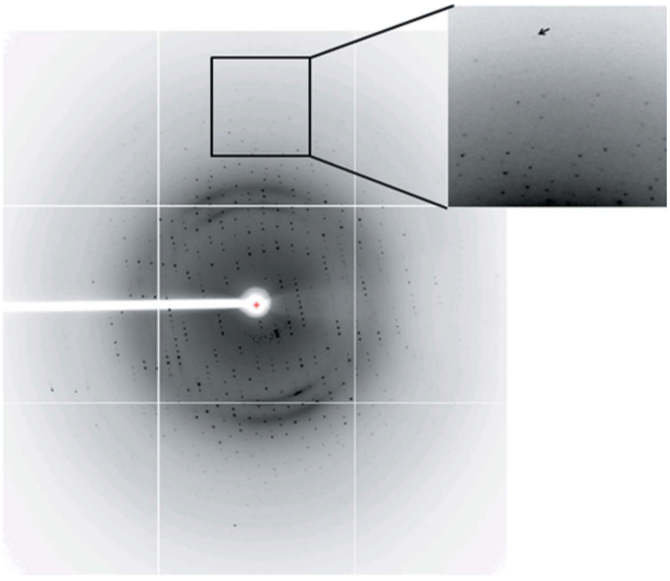


Figure 3
Typical X-ray diffraction pattern of HCoV-OC43 N-NTD. The arrow shows the data at the resolution limit (1.7 Å).

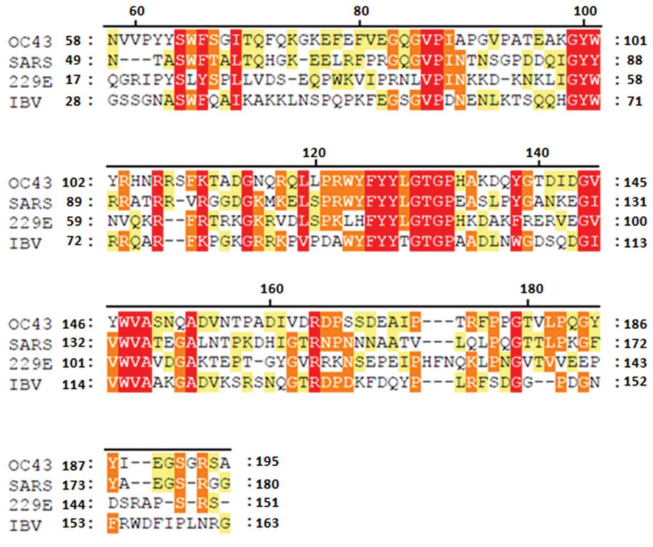


Figure 4
Multiple sequence alignment of CoV N-NTDs; *T-Coffee* (Notredame *et al.*, 2000) was used to define conserved residues. The numbering above the sequence is for the amino-acid sequence from OC43. Conserved residues are shaded. Fully conserved residues are shaded red. Residues that were partially conserved at levels of 75% and 50% are shaded orange and yellow, respectively. The amino-acid sequences of OC43 (HCoV-OC43; NP_937954), SARS (SARS-CoV; ABI96968), 229E (HCoV-229E; AAG48597) and IBV (infectious bronchitis virus; AAB24054) N-NTD were obtained from GenBank.

3. Results and discussion

The HCoV-OC43 N-NTD crystal chosen for this study was shown to diffract X-rays to 1.7 Å resolution (Fig. 3) and belonged to space group $P6_5$, with unit-cell parameters a = 81.57, c = 42.87 Å. The Matthews coefficient of 2.06 Å^3 Da^{-1} calculated using *MATTHEWS_COEF* (Collaborative Computational Project, Number 4, 1994; Matthews, 1968) suggested there was likely to be one molecule in the asymmetric unit, with a solvent content of 40.26%. A homology search for the HCoV-OC43 N-NTD structure was performed using the *BLAST* server (Altschul *et al.*, 1997; http://blast.ncbi.nlm.nih.gov/Blast.cgi) when the synchrotron data were collected in December 2008. The sequence-alignment search indicated that HCoV-OC43 N-NTD shares 30–40% sequence identity with other N-NTDs of coronaviruses (Fig. 4) and the N-terminal domain from SARS coronavirus (PDB code 2ofz; Saikatendu *et al.*, 2007) was chosen as an initial search model owing to the low E value of 1×10^{-23}. The first molecular-replacement trial was performed using the automated interface *PERON* at the Protein Tectonics Platform (PTP), RIKEN SPring-8 Center, Japan (Sugahara *et al.*, 2008). The best result was obtained using the *MOLREP* program (Vagin & Teplyakov, 2010). A single and unambiguous solution for the rotation and translation function was obtained with reflections in the resolution range 30–3.0 Å, and yielded a final correlation coefficient of 0.79 and an R factor of 0.44. The core of the model consists of a tightly packed β-sheet of five antiparallel strands surrounded by large loops. Structural refinement of the HCoV-OC43 N-NTD is currently in progress.

This work was supported by NCS 97-2311-B-005-003-MY3. We thank the beamline staff at beamline BL13B1 at the National Synchrotron Radiation Research Center (NSRRC; Hsinchu, Taiwan) and at the NSRRC contract beamline BL12B2 at SPring-8 (Hyogo, Japan) for assistance. We are also grateful to have the convenient computing environment provided by the Protein Tectonics Platform (PTP), which is a collaborative structural biology facility between NSRRC and RIKEN SPring-8 Center (Hyogo, Japan).

References

Altschul, S. F., Madden, T. L., Schäffer, A. A., Zhang, J., Zhang, Z., Miller, W. & Lipman, D. J. (1997). *Nucleic Acids Res.* **25**, 3389–3402.
Chan, K. H., Cheng, V. C., Woo, P. C., Lau, S. K., Poon, L. L., Guan, Y., Seto, W. H., Yuen, K. Y. & Peiris, J. S. (2005). *Clin. Diagn. Lab. Immunol.* **12**, 1317–1321.
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.
Chou, C.-C. & Hou, M.-H. (2008). *Acta Cryst.* F**64**, 1118–1120.
Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* D**50**, 760–763.
El-Sahly, H. M., Atmar, R. L., Glezen, W. P. & Greenberg, S. B. (2000). *Clin. Infect. Dis.* **31**, 96–100.
Fan, H., Ooi, A., Tan, Y. W., Wang, S., Fang, S., Liu, D. X. & Lescar, J. (2005). *Structure*, **13**, 1859–1868.
Gagneur, A., Sizun, J., Vallet, S., Legr, M. C., Picard, B. & Talbot, P. J. (2002). *J. Hosp. Infect.* **51**, 59–64.
Hogue, B. G. & Brian, D. A. (1986). *Virus Res.* **5**, 131–144.
Huang, C.-Y., Hsu, Y.-L., Chiang, W.-L. & Hou, M.-H. (2009). *Protein Sci.* **18**, 2209–2218.
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.
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.
Kaye, H. S., Marsh, H. B. & Dowdle, W. R. (1971). *Am. J. Epidemiol.* **94**, 43–49.
Lai, M. M. (2003). *J. Biomed. Sci.* **10**, 664–675.
Lai, M. M. & Cavanagh, D. (1997). *Adv. Virus Res.* **48**, 1–100.

Matthews, B. W. (1968). *J. Mol. Biol.* **33**, 491–497.
Navas-Martin, S. R. & Weiss, S. (2004). *J. Neurovirol.* **10**, 75–85.
Nelson, G. W., Stohlman, S. A. & Tahara, S. M. (2000). *J. Gen. Virol.* **81**, 181–188.
Newman, J. (2004). *Acta Cryst.* D**60**, 610–612.
Newman, J. (2005). *Acta Cryst.* D**61**, 490–493.
Notredame, C., Higgins, D. G. & Heringa, J. (2000). *J. Mol. Biol.* **302**, 205–217.
Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.
Pohl-Koppe, A., Raabe, T., Siddell, S. G. & ter Meulen, V. (1995). *J. Virol. Methods*, **55**, 175–183.
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.
St-Jean, J. R., Jacomy, H., Desforges, M., Vabret, A., Freymuth, F. & Talbot, P. J. (2004). *J. Virol.* **78**, 8824–8834.
Sugahara, M. *et al.* (2008). *J. Struct. Funct. Genomics*, **9**, 21–28.
Vagin, A. & Teplyakov, A. (2010). *Acta Cryst.* D**66**, 22–25.
Woo, P. C., Lau, S. K., Wong, B. H., Chan, K.-H., Hui, W.-T., Kwan, G. S., Peiris, J. S., Couch, R. B. & Yuen, K.-Y. (2004). *J. Clin. Microbiol.* **42**, 5885–5888.