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

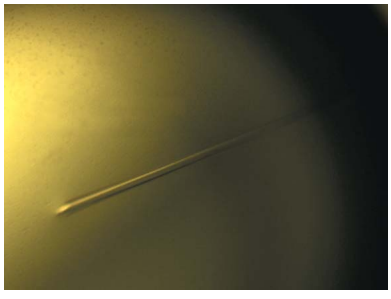
Expression, crystallization and preliminary crystallographic study of the C-terminal half of nsp2 from SARS coronavirus

Yuanyuan Li,[a,b,c] Zhilin Ren,[d] Zehua Bao,[c] Zhenhua Ming[c] and Xuemei Li[a]*

[a]National Laboratory of Macromolecules, Institute of Biophysics, Chinese Academy of Sciences, Beijing 100101, People's Republic of China, [b]Graduate University of Chinese Academy of Sciences, Beijing 100049, People's Republic of China, [c]Structural Biology Laboratory, Tsinghua University, Beijing 100084, People's Republic of China, and [d]College of Life Sciences and Tianjin State Laboratory of Protein Science, Nankai University, Tianjin 300071, People's Republic of China

Correspondence e-mail: lixm@sun5.ibp.ac.cn

Received 19 April 2011
Accepted 11 May 2011



SARS coronavirus (SARS-CoV) is the aetiological agent of the highly infectious severe acute respiratory syndrome (SARS). To gain a better understanding of SARS-CoV replication and transcription proteins, a preliminary X-ray crystallographic study of the C-terminal domain of SARS-CoV nonstructural protein 2 (nsp2) is reported here. The C-terminal domain of SARS-CoV nsp2 was cloned, overexpressed, purified and crystallized using polyethylene glycol 5000 monomethyl ether as the precipitant; the crystals diffracted to 2.5 Å resolution. The crystals belonged to space group $P6_5$, with unit-cell parameters $a = b = 112.8$, $c = 91.1$ Å, $\alpha = \beta = 90$, $\gamma = 120^\circ$. One molecule is assumed to be present per asymmetric unit, which gives a Matthews coefficient of 2.89 Å^3 Da^{-1} and a solvent content of 56.2%.

1. Introduction

The 16 replicase/transcriptase proteins are crucial for the life cycle of SARS coronavirus (SARS-CoV), which is the causative agent of the highly infectious severe acute respiratory syndrome (SARS) that first appeared in late 2002 (Stadler *et al.*, 2003) in Asia and subsequently spread worldwide. SARS-CoV belongs to the Coronaviridae family, with a single positive-stranded RNA genome of approximately 30 kb in length (Snijder *et al.*, 2003).

Since the SARS pandemic, the three-dimensional structures and functions of most of the replicase/transcriptase components, the non-structural proteins (nsps), of SARS-CoV have been determined (Zhang *et al.*, 2010; Yang *et al.*, 2003; Almeida *et al.*, 2006; Su *et al.*, 2006; Xue *et al.*, 2008; Xu *et al.*, 2009). However, the structure and function of nsp2 from either SARS-CoV or related coronaviruses have remained uncharacterized. To date, the only reported crystal structure of a coronavirus nsp2 is that of the N-terminal domain of nsp2 from avian infectious bronchitis virus (IBV; Yang *et al.*, 2009), a representive of the group 3 coronaviruses which is quite different from SARS-CoV nsp2 in primary sequence (with a sequence identity of less than 20%) and gene locus (with no nsp1 prior to nsp2 in IBV). Meanwhile, owing to a lack of research on nsp2-deficient viruses in animal models or on changes in host environment resulting from nsp2 deficiency, it has been widely accepted that nsp2 replicase proteins are dispensable for replication of mouse hepatitis virus (MHV) and SARS-CoV in cell culture owing to the detection of genomic and subgenomic RNA production of nsp2-deficient viruses in cell culture (Graham *et al.*, 2005).

In order to help to elucidate the function(s) of this relatively large protein (70 and 65 kDa for SARS-CoV and MHV, respectively), we now report the expression, purification and crystallization of the C-terminal half of nsp2 from SARS-CoV (referred to here as nsp2C; 58 kDa, corresponding to Lys112–Gly638 of full-length SARS-CoV nsp2) as well as its preliminary structure determination by single-wavelength anomalous dispersion (SAD). Further analysis of the structure and function of nsp2 from SARS-CoV and other members of the coronavirus family should allow the elucidation of its precise function(s) involved in coronavirus pathogenesis on the basis of our crystal structure of SARS-CoV nsp2C.

2. Materials and methods

2.1. Cloning and expression

The coding sequence for SARS-CoV nsp2C was amplified by a standard PCR-based approach from the cDNA of SARS-CoV BJ01 strain (corresponding to Lys292–Gly818 of pp1a replicative polyprotein, renumbered as Lys112–Gly638). The PCR product was digested by *Bam*HI and *Not*I and ligated into a pGEX-6p-1 expression vector (Pharmacia, New York, USA). The integrity of the construct was confirmed by DNA sequencing. SARS-CoV nsp2C was overexpressed in *Escherichia coli* strain BL21 (DE3) (Novagen, Merck, USA) as a GST (glutathione *S*-transferase) fusion protein. A selenomethionyl (SeMet) derivative of nsp2C was prepared using the method of methionine-biosynthesis pathway inhibition. Expression of native and SeMet-derivative nsp2C was performed in 0.8 l Luria–Bertani medium and M9 medium, respectively, which was incubated at 310 K until the OD_{600} reached about 0.6. SARS-CoV nsp2C expression was induced by the addition of 0.5 m*M* isopropyl β-D-1-thiogalactopyranoside (IPTG) for an additional 16 h at 289 K. For the preparation of soluble protein fractions, cells from the 0.8 l culture were pelleted, resuspended in 50 ml cold PBS as a lysis buffer containing 137 m*M* NaCl, 2.7 m*M* KCl, 4.3 m*M* Na_2HPO_4, 1.4 m*M* KH_2PO_4, 1 m*M* DTT, 1 m*M* EDTA and 0.01% NP-40 pH 8.5 and lysed using a JN-3000 PLUS low-temperature ultrahigh-pressure cell disrupter (JNBIO, Guangzhou, People's Republic of China).

2.2. Purification

The supernatant after centrifugation was collected and the fusion protein was purified by GST-glutathione affinity chromatography. The native and SeMet-derivative nsp2C-GST fusion proteins were further purified using the same procedure as follows: briefly, the bacterial cell lysate containing nsp2C protein was incubated with Glutathione-Sepharose 4B resin (GE Healthcare, USA) at 277 K. The GST tag was removed by overnight digestion at 277 K using GST-tagged PreScission protease (Amersham Biosciences) in 1×PBS (137 m*M* NaCl, 2.7 m*M* KCl, 4.3 m*M* Na_2HPO_4, 1.4 m*M* KH_2PO_4, 10% glycerol pH 8.5), leaving five additional residues (GPLGS) at the N-terminus. SARS-CoV nsp2C was further purified by Resource Q ion-exchange column chromatography (Amersham Biosciences, USA) at 291 K in 25 m*M* Tris–HCl, 1 m*M* EDTA, 1 m*M* DTT, 0.01% NP-40 pH 8.5 with a 50–250 m*M* NaCl gradient and achieved high homogeneity; the protein purity was estimated to be about 90% by inspection of Coomassie-stained Tris–glycine SDS–PAGE gels (Fig. 1). All proteins were further characterized by MALDI–TOF mass spectrometry and incorporation of selenomethionine was also confirmed by mass spectroscopy. Fractions containing pure protein were pooled and concentrated to 8 mg ml^{-1} for crystallization.

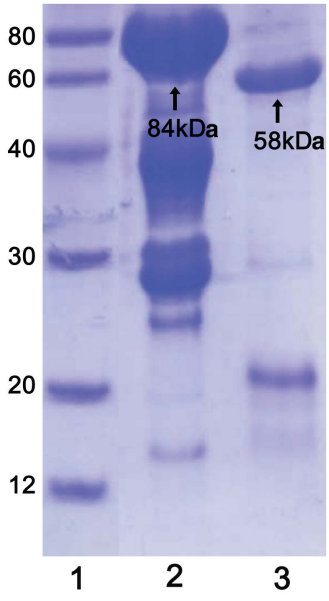
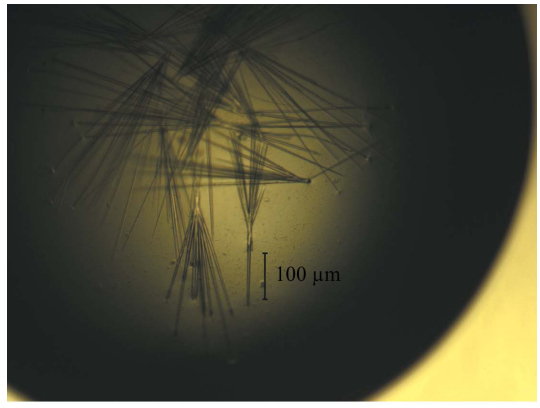


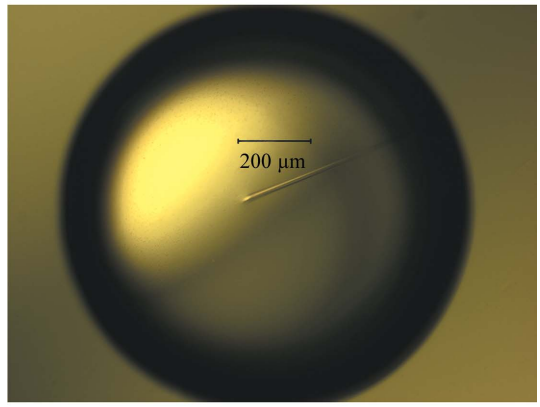
Figure 1
SDS–PAGE analysis of SARS-CoV nsp2C during purification. Proteins were analysed on 10% SDS–PAGE and stained with Coomassie Blue. Lane 1, molecular-weight markers (labelled in kDa). Lane 2, purified nsp2C after GST (glutathione *S*-transferase) affinity column chromatography. The molecular weight of GST-nsp2C is 84 kDa, as indicated by a black arrow. Lane 3, purified nsp2C with approximately 90% purity after Resource Q ion-exchange column chromatography. The molecular weight of nsp2C is 58 kDa, as indicated by a black arrow.

2.3. Crystal growth, data collection and processing

Initial crystals were obtained *via* the hanging-drop vapour-diffusion method at 291 K by mixing 1 µl protein solution and 1 µl reservoir solution. Initial needle-shaped crystals of SARS-CoV nsp2C with extensive twinning appeared after one week in a condition from Hampton Research Crystal Screen. Crystals suitable for data collection with a length of up to 200 µm and variable thickness were grown using a reservoir solution consisting of 0.1 *M* Bis-Tris pH 6.5, 0.2 *M* NaCl, 20%(*w*/*v*) polyethylene glycol 5000 monomethyl ether (Fig. 2*a*). Crystals of the SeMet derivative of nsp2C were obtained using the same conditions but grew to larger size (Fig. 2*b*). Prior to data collection, the crystals were dehydrated for 2 h with reservoir solution plus 10% glycerol and were immediately soaked in



(*a*)



(*b*)

Figure 2
SARS-CoV nsp2C crystals. (*a*) Native; (*b*) SeMet derivative.

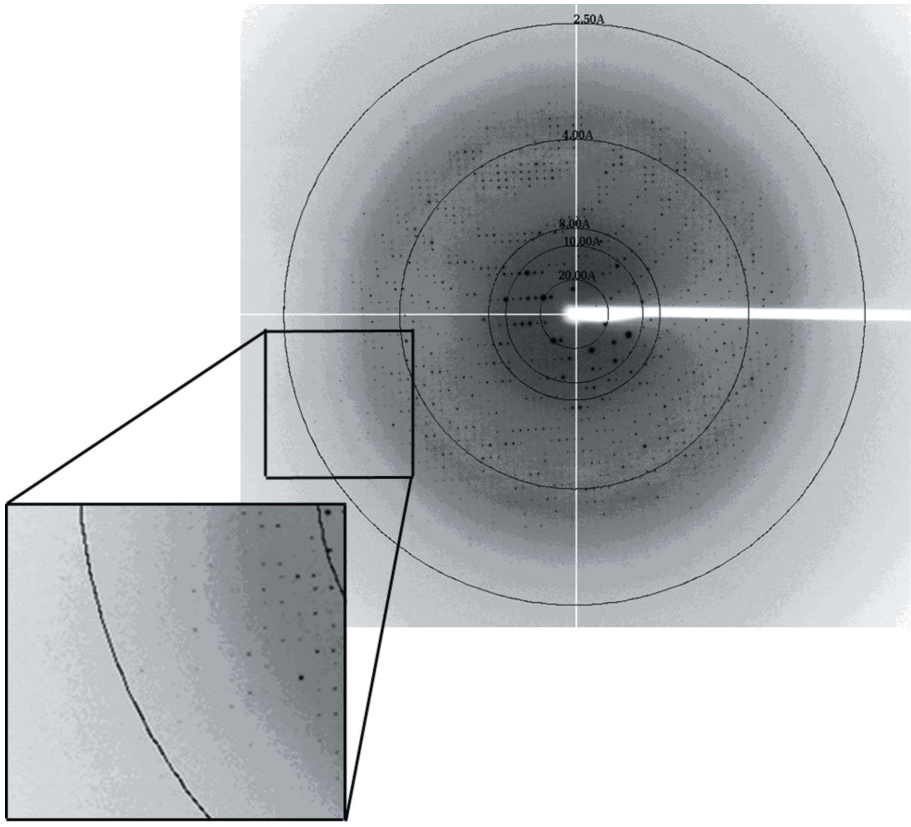


Figure 3
X-ray diffraction pattern from a crystal of native SARS-CoV nsp2C.

cryoprotectant solution constituted of reservoir solution with 20% ethylene glycol followed by flash-freezing in liquid nitrogen.

The crystal-to-detector distance was set to 273.5 and 362.5 mm for native and SeMet-derivative nsp2C, respectively. All frames were collected at 93 K using a 0.8° oscillation angle with an exposure time of 2 s per frame. A total of 300 frames and 450 frames were collected for native and SeMet nsp2C, respectively. A native data set for nsp2C was collected to 2.7 Å resolution at a wavelength of 1.0 Å using a MAR 165 CCD detector on beamline BL17U at Shanghai Synchrotron Radiation Facility (SSRF; People's Republic of China) and a single-wavelength anomalous dispersion (SAD; Terwilliger & Berendzen, 1999) data set for the SeMet derivative of nsp2C was collected to 3.5 Å resolution (Fig. 3) at a wavelength of 0.9787 Å using an ADSC Q270 CCD detector on beamline BL17A at Photon Factory (PF; Tsukuba, Japan). Data were processed, integrated and scaled using the *HKL*-2000 program package (Otwinowski & Minor, 1997). The initial phases were obtained using *SHELX* (Sheldrick, 2008) and *PHENIX* (Adams *et al.*, 2002). The selenomethionine sites were located and interpretable maps were obtained. The phases were greatly improved after density-modification procedures using *RESOLVE* (Terwilliger, 2000, 2001) and *DM* in *CCP*4 (Winn *et al.*, 2011). A summary of data collection and processing is shown in Table 1.

3. Results and discussion

The crystal of nsp2C belonged to space group $P6_5$, which was identified after we had obtained the initial phases using *SHELX* and the SeMet-nsp2C data set. The unit-cell parameters were $a = b = 112.8$, $c = 91.1$ Å, $\alpha = \beta = 90$, $\gamma = 120^\circ$ and there was only one molecule per asymmetric unit, corresponding to a calculated Matthews coefficient of 2.89 Å^3 Da^{-1} and a solvent content of 57.4% (Matthews, 1968). The anomalous data provided four clear selenium sites. The initial

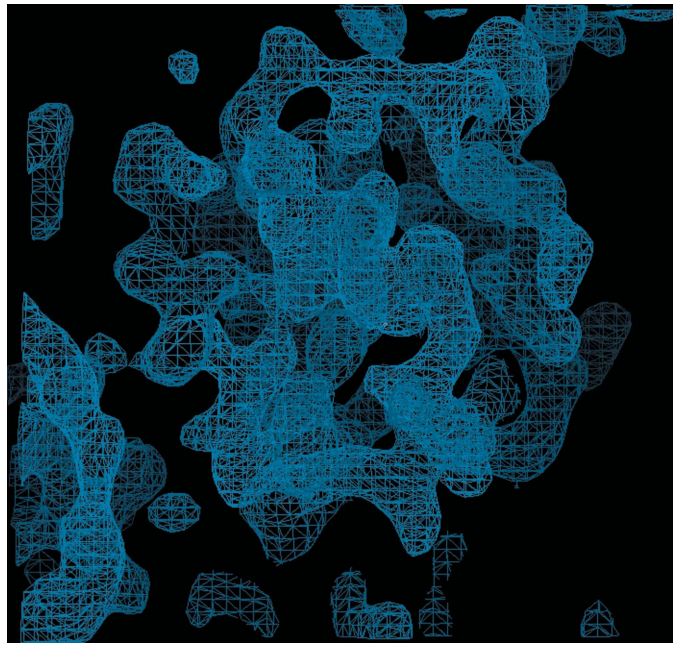


Figure 4
A 1.5σ-weighted $2F_o - F_c$ electron-density map of SARS-CoV nsp2C.

Table 1
Data-collection and processing statistics.

Values in parentheses are for the highest resolution shell.

	Native nsp2C	SeMet-derivative nsp2C
Data-collection statistics		
Unit-cell parameters (Å, °)	$a = b = 112.8$, $c = 91.1$, $\alpha = \beta = 90$, $\gamma = 120$	$a = b = 113.8$, $c = 91.2$, $\alpha = \beta = 90$, $\gamma = 120$
Space group	$P6_5$	$P6_5$
Wavelength (Å)	1.0	0.9787
Resolution range (Å)	50.0–2.7 (2.75–2.70)	50.0–3.5 (3.56–3.50)
Mosaicity (°)	0.7	1.1
Solvent content (%)	57.4	58.2
Multiplicity	12.2 (5.1)	16.2 (7.1)
Data processing		
No. of observed reflections	220979	138180
No. of unique reflections	18044	8556
Completeness (%)	99.5 (98.5)	100.0 (99.5)
$\langle I/\sigma(I)\rangle$	32.6 (2.6)	21.8 (2.7)
$R_{\rm merge}$† (%)	13.1 (65.8)	16.0 (61.3)

† $R_{\rm merge} = \sum_{hkl}\sum_i |I_i(hkl) - \langle I(hkl)\rangle| / \sum_{hkl}\sum_i I_i(hkl)$, where $\langle I(hkl)\rangle$ is the mean of the observations $I_i(hkl)$ of reflection hkl.

phases were improved by solvent flattening (*DM*; Winn *et al.*, 2011) and the electron-density map allowed us to trace most of the main-chain residues of SARS-CoV nsp2C. After several iterations of density modification using both the *DM* and *SHARP* programs, the phases were greatly improved, with most of the amino-acid residues being clearly interpretable (Fig. 4); exceptions were the C-terminal end, which consists mostly of loop and coil, as well as electron density in some regions owing to flexibility. Further model building and refinement of the structure of SARS-CoV nsp2C is in progress. The successful crystallization of nsp2C from SARS-CoV to give crystals that were suitable for structure determination should allow us to answer many of the fundamental questions that remain unclear about the role(s) of nsp2 in the regulation of coronavirus pathogenesis.

This work was supported by the Ministry of Science and Technology 973 Project (grant Nos. 2006CB806503 and 2007CB914301) and the National Natural Science Foundation of China (grant Nos. 30221003 and 30730022).

References

Adams, P. D., Grosse-Kunstleve, R. W., Hung, L.-W., Ioerger, T. R., McCoy, A. J., Moriarty, N. W., Read, R. J., Sacchettini, J. C., Sauter, N. K. & Terwilliger, T. C. (2002). *Acta Cryst.* D**58**, 1948–1954.
Almeida, M. S., Johnson, M. A. & Wüthrich, K. (2006). *J. Biomol. NMR*, **36**, Suppl. 1, 46.
Graham, R. L., Sims, A. C., Brockway, S. M., Baric, R. S. & Denison, M. R. (2005). *J. Virol.* **79**, 13399–13411.
Matthews, B. W. (1968). *J. Mol. Biol.* **33**, 491–497.
Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.
Sheldrick, G. M. (2008). *Acta Cryst.* A**64**, 112–122.
Snijder, E. J., Bredenbeek, P. J., Dobbe, J. C., Thiel, V., Ziebuhr, J., Poon, L. L., Guan, Y., Rozanov, M., Spaan, W. J. & Gorbalenya, A. E. (2003). *J. Mol. Biol.* **331**, 991–1004.
Stadler, K., Masignani, V., Eickmann, M., Becker, S., Abrignani, S., Klenk, H. D. & Rappuoli, R. (2003). *Nature Rev. Microbiol.* **1**, 209–218.
Su, D., Lou, Z., Sun, F., Zhai, Y., Yang, H., Zhang, R., Joachimiak, A., Zhang, X. C., Bartlam, M. & Rao, Z. (2006). *J. Virol.* **80**, 7902–7908.
Terwilliger, T. C. (2000). *Acta Cryst.* D**56**, 965–972.
Terwilliger, T. C. (2001). *Acta Cryst.* D**57**, 1763–1775.
Terwilliger, T. C. & Berendzen, J. (1999). *Acta Cryst.* D**55**, 849–861.
Winn, M. D. *et al.* (2011). *Acta Cryst.* D**67**, 235–242.
Xu, Y., Cong, L., Chen, C., Wei, L., Zhao, Q., Xu, X., Ma, Y., Bartlam, M. & Rao, Z. (2009). *J. Virol.* **83**, 1083–1092.
Xue, X., Yu, H., Yang, H., Xue, F., Wu, Z., Shen, W., Li, J., Zhou, Z., Ding, Y., Zhao, Q., Zhang, X. C., Liao, M., Bartlam, M. & Rao, Z. (2008). *J. Virol.* **82**, 2515–2527.
Yang, A., Wei, L., Zhao, W., Xu, Y. & Rao, Z. (2009). *Acta Cryst.* F**65**, 788–790.
Yang, H., Yang, M., Ding, Y., Liu, Y., Lou, Z., Zhou, Z., Sun, L., Mo, L., Ye, S., Pang, H., Gao, G. F., Anand, K., Bartlam, M., Hilgenfeld, R. & Rao, Z. (2003). *Proc. Natl Acad. Sci. USA*, **100**, 13190–13195.
Zhang, S., Zhong, N., Xue, F., Kang, X., Ren, X., Chen, J., Jin, C., Lou, Z. & Xia, B. (2010). *Protein Cell*, **1**, 371–383.