Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

Crystallization and preliminary crystallographic analysis of the heptad-repeat complex of SARS coronavirus spike protein

Yanhui Xu,[a,b] Nan Su,[a,b] Lan Qin,[a,b] Zhihong Bai,[a,b] George F. Gao[c] and Zihe Rao[a,b]*

[a]Laboratory of Structural Biology, Tsinghua University, Beijing 100084, People's Republic of China, [b]National Laboratory of Biomacromolecules, Institute of Biophysics, Beijing 100101, People's Republic of China, and [c]Institute of Microbiology, Chinese Academy of Sciences, Beijing 100080, People's Republic of China

Correspondence e-mail: raozh@xtal.tsinghua.edu.cn

Received 2 September 2004
Accepted 25 October 2004

The aetiological agent of an emergent outbreak of atypical pneumonia, severe acute respiratory syndrome (SARS), is a positive-stranded RNA virus (SARS-CoV) belonging to the Coronaviridae family with a genome that differs substantially from those of other known coronaviruses. Highly conserved heptad-repeat (HR1 and HR2) regions in class I viral fusion proteins, including spike protein from SARS coronavirus, interact with each other to form a six-helix bundle, which is called a fusion core. The crystal structure of the fusion core is expected to greatly facilitate drug design. Crystals of the fusion core of SARS-CoV spike protein have been grown at 291 K using PEG 4000 as precipitant. The diffraction pattern of the crystal extends to 2.8 Å resolution at 100 K in-house. The crystals have unit-cell parameters $a = 121.2$, $b = 66.3$, $c = 70.0$ Å, $\alpha = \gamma = 90$, $\beta = 107.4^\circ$ and belong to space group $C2$. Assuming the presence of six molecules per asymmetric unit, the solvent content is estimated to be about 28%.

1. Introduction

An emergent outbreak of atypical pneumonia, referred to as severe acute respiratory syndrome (SARS) and first identified in Guangdong Province, China, spread to several countries in the spring of 2003 (Marra *et al.*, 2003). The National Microbiology Laboratory in Canada obtained the Tor2 isolate from a patient in Toronto and succeeded in growing a coronavirus-like agent in African green monkey kidney (Vero E6) cells (Marra *et al.*, 2003). This coronavirus has been named publicly by the World Health Organization as 'SARS coronavirus' (Marra *et al.*, 2003). This coronavirus has been identified as the distinct pathological entity responsible for severe acute respiratory syndrome (SARS), which infected more than 8000 people and killed 774 worldwide in 2003.

The coronaviruses are single-stranded positive-sense RNA viruses and are members of a family of enveloped viruses that include a large number of viruses that infect different animal species (Siddell *et al.*, 1983). The predominant diseases associated with these viruses are respiratory and enteric infections, although hepatic and neurological diseases also occur (Siddell, 1995). Although morphologically a coronavirus, the SARS coronavirus (SARS-CoV) is not closely related to any of the three known classes of coronavirus and has been proposed to define a fourth class of coronavirus (group 4) (Rota *et al.*, 2003).

The surface glycoproteins of enveloped viruses are essential for viral entry and replication. These envelope proteins mediate both the initial virion attachment to the cell and the subsequent fusion of viral and cellular membranes, resulting in the release of the viral contents into the host cell (Eckert & Kim, 2001*b*). As with other enveloped RNA viruses including the coronaviruses, genome sequencing reveals that SARS-CoV envelope spike (S) protein contains highly conserved heptad-repeat regions (HR1 and HR2), which have been shown to be important in virus membrane fusion. Such heptad-repeat regions have been successfully used as targets for virus entry/fusion inhibitors in a number of viruses (Chan & Kim, 1998; Eckert & Kim, 2001*b*). It is generally believed that the envelope protein undergoes a series of conformational changes during the virus-fusion process. The HR1 and HR2 regions are believed to be important modules/domains in this process and show different conformations in different fusion states (Eckert & Kim, 2001*b*). In the current model, there are at least three conformational states of the envelope fusion protein: the pre-fusion native state, the pre-hairpin intermediate state and the post-fusion hairpin state (Chan *et al.*, 1997; Eckert & Kim, 2001*b*). During these state transitions, HR1 and HR2 are exposed in an intermediate conformational state but bind to each other to form a coiled-coil structure in the post-fusion state. Therefore, the *in vitro* introduced HR peptides compete with their endogenous HR counterparts in the intermediate state, preventing the transition into the formation of HR1/HR2 coiled-coil bundle in the post-fusion state.

Some HR peptides or analogues have been shown to be potent inhibitors of virus fusion (Eckert & Kim, 2001*a*,*b*; Root *et al.*, 2001).

Membrane fusion is the key step in enveloped virus infection. The coiled-coil bundle conformation is believed to be important for bringing two lipid membranes (cellular and viral) into proximity with each other, thus enabling the membrane fusion for virus entry. The HR1/HR2 coiled-coil bundle is also called the fusion core of the virus fusion protein (Eckert & Kim, 2001*b*). Several crystal structures of fusion cores have been determined to date, including HIV gp41, HRSV F, influenza virus HA, Ebola virus GP and MHV (Bullough *et al.*, 1994; Chan *et al.*, 1997; Malashkevich *et al.*, 1999; Xu, Liu *et al.*, 2004; Zhao *et al.*, 2000). We have solved the crystal structure of the fusion core of MHV spike protein and proposed a molecular mechanism for coronavirus-mediated viral fusion (Xu, Liu *et al.*, 2004). However, the detailed structure of the SARS-CoV spike protein fusion core is still unknown.

Our previous studies have shown that the complex of HR1 and HR2 from the SARS-CoV spike protein also forms a typical stable trimeric α-helical bundle (Xu, Zhu *et al.*, 2004). Here, we report the purification, crystallization and preliminary crystallographic studies of the fusion core of the SARS-CoV spike protein.

2. Materials and methods

2.1. Expression and purification of SARS fusion core (2-Helix) proteins

The SARS spike gene was cloned from SARS coronavirus GZ02 clone 1 (Gene Bank No. AY390556). The HR1 region used was derived from residues 900–948 and HR2 from residues 1145–1184. The SARS 2-Helix construct was made by linking the HR1 region and HR2 region with a 22 amino-acid linker (LVPRGSGGSGGSGGLEVLFQGP, single-letter amino-acid codes). The linker includes a thrombin-recognition site, a prescission recognition site and a typical flexible linker (GGSGGSGG). The constructs were cloned into the *Nde*I and *Xho*I sites (introduced by synthetic PCR primers) of the pET22b expression vector (Pharmacia). The construction was verified by sequencing, the expected plasmid was transformed into *Escherichia coli* strain BL21(DE3) competent cells and the transformants were selected on LB agar plates containing 100 µg ml^{-1} ampicillin. The cells were cultured at 310 K in 2×YT medium containing 100 µg ml^{-1} ampicillin. When the culture density (OD_{600}) reached 0.8, the culture was induced with 0.2 m*M* IPTG and grown for an additional 10 h at 289 K before the cells were harvested.

The bacterial cell pellet was resuspended in 1×PBS (140 m*M* NaCl, 2.7 m*M* KCl, 10 m*M* Na_2HPO_4, 1.8 m*M* KH_2PO_4) and homogenized by sonication. The lysate was centrifuged at 18 000*g* for 20 min at 277 K, the supernatant was loaded onto an Ni^{2+}–NTA column (Qiagen), the contaminated unbound protein was washed off with 50 ml washing buffer (1×PBS, 60 m*M* imidazole) and the target protein was eluted with elution buffer (1×PBS, 500 m*M* imidazole). The sample from affinity chromatography was concentrated to 0.5 ml and subsequently loaded onto a Superdex 75 column HR10/30 (Pharmacia Biotech) column with an ÄKTA Purifier System (Pharmacia Biotech) in buffer containing 25 m*M* Tris–HCl pH 8.0, 150 m*M* NaCl. Protein purity was determined by 10% tricine SDS–PAGE.

2.2. Crystallization

The purified SARS 2-Helix protein was dialyzed against crystallization buffer (10 m*M* Tris–HCl pH 8.0, 10 m*M* NaCl) and concentrated to 10–15 mg ml^{-1}. Crystallization was performed by the hanging-drop vapour-diffusion method at 291 K. 1 µl protein solution was mixed with 1 µl reservoir solution and the mixture was equilibrated against 200 µl reservoir solution at 291 K. Initial crystallization conditions were screened using Crystal Screen reagent kits (Hampton Research) and a PEG 4000 Screening kit prepared in-house. The protein could be crystallized under several conditions containing PEG 4000. Conditions yielding small crystals were further optimized by variation of precipitant, protein concentration, buffer pH and additives. Good-quality well diffracting crystals could be obtained in 0.1 *M* sodium acetate pH 2.5, 10–15%(*v*/*v*) PEG 4000, 0.02 *M* spermidine.4HCl. The crystals appeared in two weeks.

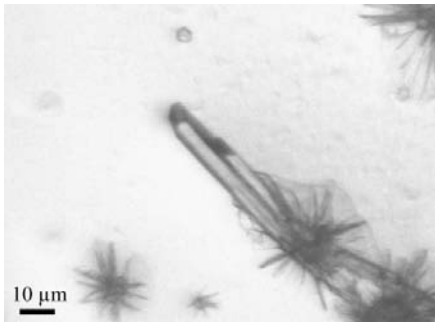

Figure 1
Typical crystals of the SARS 2-Helix protein grown by the hanging-drop method in 0.1 *M* sodium acetate pH 2.5, 13%(*v*/*v*) PEG 4000, 0.02 *M* spermidine.4HCl.

Table 1
Data-collection and processing statistics for SARS 2-Helix.

Space group	*C*2
Unit-cell parameters (Å, °)	a = 121.2, b = 66.3, c = 70.0, β = 107.4
Wavelength (Å)	1.5418
Resolution limit (Å)	2.8
Observed reflections	27312
Unique reflections	11974
Completeness (%)	91.4 (80.8)
$I/\sigma(I)$	5.7 (1.2)
$R_{\rm merge}$ (%)	13.9 (51.1)

† $R_{\rm merge} = \sum_h \sum_i |I_{ih} - \langle I_h \rangle| / \sum_h \sum_i \langle I_h \rangle$, where $\langle I_h \rangle$ is the mean of the observations I_{ih} of reflection h.

2.3. Data collection and processing

The SARS 2-Helix crystal was mounted in a nylon loop and flash-cooled in a cold nitrogen-gas stream at 100 K using an Oxford Cryosystems coldstream with 0.1 *M*

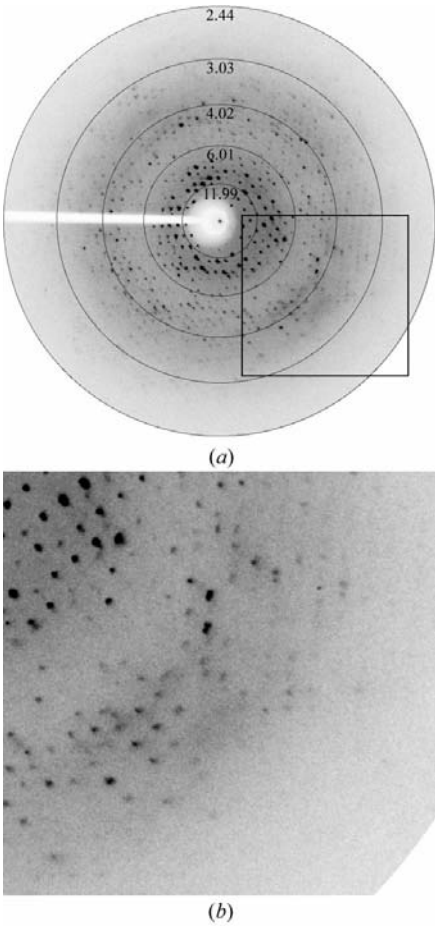

Figure 2
(*a*) A typical X-ray diffraction pattern from a crystal of SARS 2-Helix. The diffraction image was collected on a MAR 345 image-plate detector. The oscillation range is 1°. (*b*) An enlarged image of the area indicated in Fig. 2(*a*).

sodium acetate pH 2.5, 25% PEG 4000 as cryoprotectant. Data collection was performed in-house on a Rigaku RU2000 rotating copper-anode X-ray generator operated at 48 kV and 98 mA (Cu $K\alpha$; λ = 1.5418 Å) with a MAR 345 image-plate detector. Data were indexed, integrated and scaled using *DENZO* and *SCALEPACK* (Otwinowski & Minor, 1997).

3. Results and discussion

The SARS-CoV 2-Helix protein construct could be crystallized easily under several conditions, while good-quality crystals could only be obtained using 0.1 *M* sodium acetate pH 2.5, 13%(*v*/*v*) PEG 4000, 0.02 *M* spermidine.4HCl (Fig. 1). Diffraction data were collected to 2.8 Å from a single crystal. The crystals belong to space group *C*2, with unit-cell parameters $a = 121.2$, $b = 66.3$, $c = 70.0$ Å, $\alpha = \gamma = 90$, $\beta = 107.4^\circ$. Assuming the presence of six molecules (two stable fusion-core complexes) in the asymmetric unit, the Matthews number (V_M) is about 1.7 Å^3 Da^{-1} and the solvent content is calculated to be about 28%. Selected data statistics are shown in Table 1.

X-ray diffraction data of 2-Helix of SARS-CoV were collected from a single crystal using a Rigaku RU2000 rotating copper-anode X-ray generator and a MAR 345 image-plate detector (Figs. 2*a* and 2*b*). The crystal-to-detector distance was 120 mm, the exposure time was 600 s, the oscillation angle was 1° and the wavelength was 1.5418 Å.

The collected data were indexed and reduced with *DENZO* and *SCALEPACK* (Otwinowski & Minor, 1997). The crystal structure of the SARS 2-Helix has been determined by molecular replacement with the MHV 2-Helix structure (PDB code 1wdf) as a search model. Clearly, the crystal structure of the fusion core of SARS-CoV spike protein will provide a detailed picture of the viral fusion-core structure and the viral fusion mechanisms mediated by the SARS coronavirus spike protein. This will open a new avenue towards the structure-based fusion inhibitor design of peptides or peptide analogues, *e.g.* small molecules, for this emerging infectious disease.

This work was supported by Project 973 and 863 of the Ministry of Science and Technology of China [grant Nos. 200BA711A12, G199075600, GZ236(202/9) and 2003CB514103].

References

Bullough, P. A., Hughson, F. M., Skehel, J. J. & Wiley, D. C. (1994). *Nature (London)*, **371**, 37–43.
Chan, D. C., Fass, D., Berger, J. M. & Kim, P. S. (1997). *Cell*, **89**, 263–273.
Chan, D. C. & Kim, P. S. (1998). *Cell*, **93**, 681–684.
Eckert, D. M. & Kim, P. S. (2001*a*). *Proc. Natl Acad. Sci. USA*, **98**, 11187–11192.
Eckert, D. M. & Kim, P. S. (2001*b*). *Annu. Rev. Biochem.* **70**, 777–810.
Malashkevich, V. N., Schneider, B. J., McNally, M. L., Milhollen, M. A., Pang, J. X. & Kim, P. S. (1999). *Proc. Natl Acad. Sci. USA*, **96**, 2662–2667.
Marra, M. A. *et al.* (2003). *Science*, **300**, 1399–1404.
Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.
Root, M. J., Kay, M. S. & Kim, P. S. (2001). *Science*, **291**, 884–888.
Rota, P. A. *et al.* (2003). *Science*, **300**, 1394–1399.
Siddell, S. G. (1995). *The Coronaviridae: an Introduction*, pp. 1–10. New York: Plenum.
Siddell, S., Wege, H. & Ter Meulen, V. (1983). *J. Gen. Virol.* **64**, 761–776.
Xu, Y., Liu, Y., Lou, Z., Qin, L., Li, X., Bai, Z., Pang, H., Tien, P., Gao, G. F. & Rao, Z. (2004). *J. Biol. Chem.* **279**, 30514–50322.
Xu, Y., Zhu, J., Liu, Y., Lou, Z., Yuan, F., Liu, Y., Cole, D. K., Ni, L., Su, N., Qin, L., Li, X., Bai, Z., Bell, J. I., Pang, H., Tien, P., Rao, Z. & Gao, G. F. (2004). In the press.
Zhao, X., Singh, M., Malashkevich, V. N. & Kim, P. S. (2000). *Proc. Natl Acad. Sci. USA*, **97**, 14172–14177.