

# Crystallization and preliminary crystallographic analysis of the fusion core of the spike protein of the murine coronavirus mouse hepatitis virus (MHV)

Yanhui Xu,<sup>a,b</sup> Zhihong Bai,<sup>a,b</sup> Lan Qin,<sup>a,b</sup> Xu Li,<sup>a,b</sup> George Gao<sup>c</sup> and Zihe Rao<sup>a,b\*</sup>

<sup>a</sup>Laboratory of Structural Biology, Tsinghua University, Beijing 100084, People's Republic of China, <sup>b</sup>National Laboratory of Biomacromolecules, Institute of Biophysics, Beijing 100101, People's Republic of China, and <sup>c</sup>Nuffield Department of Clinical Medicine, John Radcliffe Hospital, Oxford University, Oxford OX3 9DU, England

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

Crystals of a 2-Helix fusion-core construct of MHV spike protein (commonly referred to as E2) 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. Furthermore, a selenomethionine (SeMet) derivative of MHV spike protein fusion core has been overexpressed and purified. The derivative crystals were obtained under similar conditions and three different wavelength data sets were collected to 2.4 Å resolution from a single derivative crystal at BSRF (Beijing Synchrotron Radiation Facility). The crystals have unit-cell parameters  $a = b = 48.3$ ,  $c = 199.6$  Å,  $\alpha = \beta = 90$ ,  $\gamma = 120^\circ$  and belong to space group  $R3$ . Assuming the presence of two molecules in the asymmetric unit, the solvent content is calculated to be about 46%.

Received 25 April 2004  
Accepted 17 August 2004

## 1. Introduction

Mouse hepatitis virus strain A59 (MHV-A59) belongs to the coronaviruses, which comprise a large and diverse family of enveloped viruses with a single-stranded positive-sense RNA genome of approximately 31 000 bp (Lee *et al.*, 1991; Spaan *et al.*, 1988). The *Coronaviridae* exhibit a broad host range, infecting many mammalian and avian species and causing upper respiratory, gastrointestinal, hepatic and central nervous system diseases. In humans and birds coronaviruses primarily cause upper respiratory tract infections, while porcine and bovine coronaviruses establish enteric infections that result in severe economic loss (Siddell, 1995).

The coronavirus spike (S) protein, which is also commonly referred to as E2, forms the peplomer structure on the viral envelope and each spike is thought to be a dimer or trimer (Cavanagh, 1983). The S protein mediates binding of virions to the host-cell receptor (Collins *et al.*, 1982), viral fusion during entry and cell-to-cell fusion at the later stage of post-infection (Vennema *et al.*, 1990). The MHV S protein is cleaved in the Golgi apparatus by a host-cell protease into two similarly sized subunits: the amino-terminal S1 and the carboxyl-terminal S2 (Frana *et al.*, 1985; Luytjes *et al.*, 1987; Sturman *et al.*, 1985). It is believed that the S1 subunit forms the globular head of the spike, whereas the S2 subunit forms the transmembrane stalk portion (de Groot *et al.*, 1987). Sequence analysis suggests that the coronavirus spike protein has the structural features of a type I membrane protein, including two heptad-repeat regions, a

fusion peptide, a transmembrane domain near the carboxyl-terminus of S2 and a hydrophobic signal peptide at the N-terminus of S1 (Spaan *et al.*, 1988). The S2 domain has an internal fusion peptide and two heptad-repeat regions, designated HR1 and HR2, respectively. HR2 is located close to the transmembrane region and HR1 is some 170 amino acids upstream of it.

Typical class I viral fusion proteins contain several notable features. They all have a fusion peptide at the amino-terminus that is thought to insert directly into the target membrane during the viral fusion process. Carboxyl-terminal to the fusion peptide are two regions containing 4,3 hydrophobic (heptad) repeats, sequence motifs that form a coiled-coil structure. This structure forms a stable protease K-resistant core in the ectodomain of the enveloped glycoprotein (or fusion protein). The fusion-core complex consists of two peptides, termed HR1 and HR2 peptides, that correspond to the two regions with hydrophobic heptad repeats. Heptad-repeat (HR) regions are found in the fusion proteins of many different viruses and form an important characteristic of class I viral fusion proteins (Bosch *et al.*, 2003; Eckert & Kim, 2001). Previous biophysical and structural studies show that the HR1 and HR2 peptides of MHV spike protein form a stable helical trimer of heterodimers (Bosch *et al.*, 2003). It is believed that the two HR domains refold into a six-helix bundle during the fusion process, in which the HR1 domain forms a trimeric coiled coil surrounded by three antiparallel helices of the HR2 domain (Eckert & Kim, 2001). Consequently, the transmembrane domain and the fusion peptide, which is known to insert into

the cell membrane, are both transposed into close association. This pulls the cellular and viral membranes into proximity, facilitating the membrane fusion.

Our recent experiments have shown that the two HR domains of MHV spike protein can also form a six-helix bundle (Xu *et al.*, in preparation). Here, we report the preliminary X-ray crystallographic analysis of the first fusion-core complex of a coronavirus spike protein, namely the MHV fusion core.

2. Materials and methods

2.1. Expression and purification of the fusion-core proteins

The fusion-core proteins of MHV were prepared as a single chain by linking the HR1 and HR2 domains *via* an eight-amino-acid linker (GGSGGSGG). The amino-acid sequences of HR1 and HR2 were taken from the murine coronavirus mouse hepatitis virus strain A59 (GeneBank No. M18379). The HR1 region used was derived from amino acids 968–1027 and HR2 from 1216–1254. The constructs and the encoded proteins were called 2-Helix (Fig. 1). The preparation and characterization of the 2-Helix proteins will be reported elsewhere (Xu *et al.*, in preparation).

The construct was cloned into the *Nde*I and *Xho*I sites (introduced by synthetic PCR primers) of pET expression vector pET22b (Novagen). The construct 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<sup>-1</sup> ampicillin. The cells were cultured at 310 K in 2×YT medium containing 100 µg ml<sup>-1</sup> ampicillin. When the culture density (*A*<sub>600</sub>) reached 0.6–0.8, the culture was induced with 0.2 mM IPTG and

grown for an additional 10 h at 289 K before the cells were harvested.

The bacterial cell pellet was resuspended in PBS and homogenized by sonication. The lysate was centrifuged at 18 000g for 20 min at 277 K and the supernatant was loaded onto an Ni<sup>2+</sup>-NTA column (Qiagen). The contaminated protein was washed with washing buffer (1×PBS, 60 mM imidazole) and the target protein was eluted with elution buffer (1×PBS, 500 mM imidazole). The protein purified by affinity chromatography was further purified using a Superdex 75 column (Pharmacia) and analyzed by SDS-PAGE.

The selenomethionine derivative of the MHV 2-Helix was expressed using *E. coli* BL21(DE3) cultured in minimal medium M9 containing 60 mg l<sup>-1</sup> L-SetMet. Six amino acids (lysine, threonine, phenylalanine, leucine, isoleucine and valine) were added to the culture for inhibited Met biosynthesis of the BL21(DE3) expression strain. Purification of the selenomethionene MHV 2-Helix was performed as for the native MHV 2-Helix. The incorporation of selenium was confirmed by mass-spectrometric analysis.

2.2. Crystallization

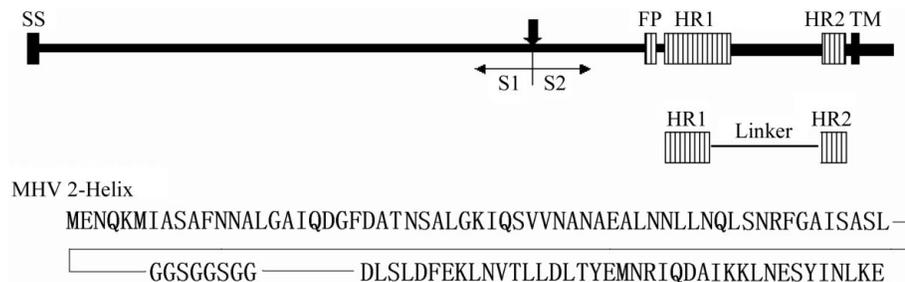
The purified protein was dialyzed against crystallization buffer (10 mM Tris-HCl pH 8.0, 10 mM NaCl) and concentrated to 8–10 mg ml<sup>-1</sup>. Protein concentrations were determined by absorbance at 280 nm, assuming an *A*<sub>280</sub> of 0.22 for a 1.0 mg ml<sup>-1</sup> solution. Initial crystallization conditions were screened using Crystal Screen reagent kits (Hampton Research). The protein could be crystallized under several conditions. Conditions yielding crystals were further optimized by variation of precipitant and protein concentration and additives. Good-quality crystals could be obtained in 0.1 M MES pH 6.5, 10% (v/v) PEG 4000, 8% (v/v)

DMSO and 5 mM hexaminecobalt trichloride (Figs. 2a and 2b). 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. The crystals appeared in 3 d.

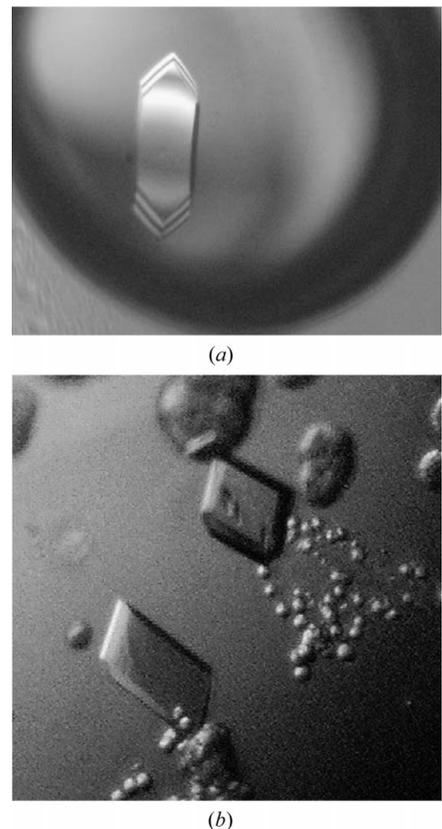
The purified selenomethionine derivative of MHV 2-Helix was concentrated to 8 mg ml<sup>-1</sup>. Crystallization trials were set up based on the optimum conditions used for native protein.

2.3. Data collection and processing

Crystals were soaked for a few seconds in a solution containing 50% (w/v) polyvinylpyrrolidone K59, which served as a cryoprotectant. The crystal was mounted in nylon loops and flashed-cooled in a cold nitrogen-gas stream at 100 K using an Oxford Cryostream. Data collection was performed by the rotation method using a MAR CCD detector with synchrotron radiation at BSRF (beamline 3W1A of



**Figure 1** Schematic representation of coronavirus MHV A59 spike protein and the MHV 2-Helix constructs. S1 and S2 are formed after proteolytic cleavage (vertical arrow) and noncovalently linked. The enveloped protein has an N-terminal signal sequence (SS) and a transmembrane domain (TM) adjacent to the C-terminus. S2 contains two HR (heptad-repeat) regions (hatched bars), termed HR1 and HR2. FP (hatched bars) is a putative fusion peptide followed by an HR1 region. Two HR regions were linked to a single polypeptide with an eight-residue linker (GGSGGSGG).



**Figure 2** (a) Typical crystals of the MHV 2-Helix proteins grown using the hanging-drop method in 0.1 M MES pH 6.5, 6% (v/v) PEG 4000. (b) Typical crystals of the MHV 2-Helix protein (selenomethionine derivative) grown using the hanging-drop method in 0.1 M MES pH 6.5, 10% (v/v) PEG 4000, 8% (v/v) DMSO and 5 mM hexaminecobalt trichloride.

**Table 1**

Data-collection and processing statistics.

Values in parentheses correspond to the highest resolution shell.

Data	Peak	Edge	Remote	Native
Space group	<i>R</i> 3	<i>R</i> 3	<i>R</i> 3	<i>R</i> 3
Unit-cell parameters (Å)				
<i>a</i> (Å)	48.4	48.5	48.3	48.3
<i>b</i> (Å)	48.4	48.5	48.3	48.3
<i>c</i> (Å)	200.0	200.1	199.4	199.6
Wavelength (Å)	0.9799	0.9801	0.9000	0.9000
Resolution range (Å)	35–2.4	35–2.4	35–2.4	35–2.5
Observed reflections	49717	50167	49352	43687
Unique reflections	6842	6867	6804	5901
Completeness (%)	100 (100)	100 (100)	100 (100)	100 (100)
$\langle I/\sigma(I) \rangle$	14.3 (4.1)	14.6 (4.5)	13.7 (3.8)	12.4 (5.6)
$R_{\text{merge}}$ (%)	9.4 (34.0)	8.4 (31.0)	9.3 (36.1)	7.4 (24.2)

†  $R_{\text{merge}} = \sum_h \sum_i |I_{h,i} - I_h| / \sum_h \sum_i I_{h,i}$  for the intensity ( $I$ ) of  $i$  observations of reflection  $h$ . The native data were reprocessed using remote data in non-anomalous form with the resolution range 35–2.5 Å.

Beijing Synchrotron Radiation Facility). Data were indexed, integrated and scaled using *DENZO* and *SCALEPACK* from the *HKL* program suite (Otwinowski & Minor, 1997).

### 3. Results and discussion

The MHV 2-Helix protein construct could be crystallized under several conditions. Large birefringent parallelepiped crystals could be obtained using 0.1 *M* MES pH 6.5, 6% (*v/v*) PEG 4000, although the diffraction pattern of the crystals only extended to 8 Å resolution (Fig. 2*a*). Well diffracting crystals could only be obtained using 0.1 *M* MES pH 6.5, 10% (*v/v*) PEG 4000, 8% (*v/v*) DMSO and 5 *mM* hexaminecobalt trichloride (Fig. 2*b*). The crystals belong to space group *R*3, with unit-cell parameters  $a = b = 48.3$ ,  $c = 199.6$  Å,  $\alpha = \beta = 90$ ,  $\gamma = 120^\circ$ . Assuming the presence of two molecules in the asym-

metric unit, the solvent content is calculated to be about 46%. Selected data statistics are shown in Table 1.

Multiple-wavelength anomalous dispersion (MAD) data were collected from a single selenomethionine-derivative crystal at peak (0.9799 Å), inflection (0.9801 Å) and remote (0.900 Å) wavelengths to 2.4 Å. The structure of MHV 2-Helix has been determined by the MAD phasing method and will be published elsewhere.

Clear structural analysis of the fusion core of MHV would provide a detailed picture of the viral fusion-core structure (the first crystal structure of the fusion core of a coronavirus spike protein) and the viral fusion mechanisms mediated by the spike protein of coronaviruses. This will add to the repertoire of viral fusion-core structures. This will also open a new avenue towards the structure-based fusion-inhibitor design of peptides or peptide analogues, e.g. small

molecules, for these emerging infectious diseases.

This work was supported by Projects 973 and 863 of the Ministry of Science and Technology of China (Grant Nos. 200BA711A12, G199075600 and 2003CB514116) and The Key Project of the Knowledge Innovation Program of Chinese Academy of Sciences (CAS). GFG's stay at Tsinghua University was supported by the Chunhui Project Scheme of the Ministry of Education of China.

### References

- Bosch, B. J., van der Zee, R., de Haan, C. A. & Rottier, P. J. (2003). *J. Virol.* **77**, 8801–8811.
- Cavanagh, D. (1983). *J. Gen. Virol.* **64**, 2577–2583.
- Collins, A. R., Knobler, R. L., Powell, H. & Buchmeier, M. J. (1982). *Virology*, **119**, 358–371.
- Eckert, D. M. & Kim, P. S. (2001). *Annu. Rev. Biochem.* **70**, 777–810.
- Frana, M. F., Behnke, J. N., Sturman, L. S. & Holmes, K. V. (1985). *J. Virol.* **56**, 912–920.
- Groot, R. J. de, Luytjes, W., Horzinek, M. C., van der Zeijst, B. A., Spaan, W. J. & Lenstra, J. A. (1987). *J. Mol. Biol.* **196**, 963–966.
- Lee, H. J., Shieh, C. K., Gorbalenya, A. E., Koonin, E. V., La Monica, N., Tuler, J., Bagdzhadzhyan, A. & Lai, M. M. (1991). *Virology*, **180**, 567–582.
- Luytjes, W., Sturman, L. S., Bredenbeek, P. J., Charite, J., van der Zeijst, B. A., Horzinek, M. C. & Spaan, W. J. (1987). *Virology*, **161**, 479–487.
- Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.
- Siddell, S. G. (1995). *The Coronaviridae: an Introduction*, pp. 1–10. New York: Plenum.
- Spaan, W., Cavanagh, D. & Horzinek, M. C. (1988). *J. Gen. Virol.* **69**, 2939–2952.
- Sturman, L. S., Ricard, C. S. & Holmes, K. V. (1985). *J. Virol.* **56**, 904–911.
- Vennema, H., Heijnen, L., Zijderveld, A., Horzinek, M. C. & Spaan, W. J. (1990). *J. Virol.* **64**, 339–346.