crystallization papers

Biochemical, biophysical and preliminary X-ray crystallographic analyses of the fusion core of Sendai virus F protein

Xiaojia Wang,[a] Yanhui Xu,[b] David K. Cole,[c] Zhiyong Lou,[b] Yiwei Liu,[b] Zihe Rao,[b] Ming Wang[a]* and George F. Gao[c,d]*

[a]College of Animal Medicine, China Agricultural University, Beijing 100094, People's Republic of China, [b]Laboratory of Structural Biology, Tsinghua University, Beijing 100084, People's Republic of China, [c]Nuffield Department of Clinical Medicine, John Radcliffe Hospital, Oxford University, Oxford OX3 9DU, England, and [d]Institute of Microbiology, Chinese Academy of Sciences, Beijing 100080, People's Republic of China

Correspondence e-mail: vetdean@cau.edu.cn, george.gao@ndm.ox.ac.uk

It is emerging that enveloped viruses may adopt a unique entry/fusion mechanism; in paramyxoviruses, including Sendai virus (SeV), the attachment protein HN (or its homologue H or G) binds a cellular receptor which triggers conformational changes of its fusion protein, F. There are at least three conformations of the F protein in the current fusion model: the pre-fusion native conformation, the pre-hairpin intermediate conformation and the post-fusion coiled-coil conformation. The fusion mechanism of SeV, a member of the *Paramyxoviridae* family, has been well established and several structural and functional domains or modules have been proposed from studies of its F protein. However, biochemical and biophysical studies of the heptad-repeat (HR) regions (HR1 and HR2) have not been systematically carried out. HR1 and HR2 strongly interact with each other to form a stable six-helix coiled-coil bundle as the post-fusion conformation. In this study, a single-chain HR1–linker–HR2 protein of SeV was prepared in an *Escherichia coli* expression system and biochemical and biophysical analyses showed it to form a typical six-helix coiled-coil bundle; its trigonal crystals diffracted X-rays to 2.5 Å resolution. The crystal structure will help to reveal the structural requirements of the post-fusion coiled-coil conformation of SeV F protein.

1. Introduction

Sendai virus (SeV) is the prototype of genus *Respirovirus* of the *Paramyxoviridae* family, which includes five previously defined and two newly designated genera (the genera *Rubulavirus*, *Respirovirus*, *Morbillivirus*, *Pneumovirus*, *Metapnuemovirus* and the new *Avulavirus* and *Henipavirus*; Lamb *et al.*, 2000; Wang *et al.*, 2001; Chang *et al.*, 2001). SeV is also called murine parainfluenza virus 1 and has been extensively studied as a model system for paramyxovirus infection, especially in virus–lipid–membrane fusion (reviewed in Peisajovich & Shai, 2002, 2003). The *Paramyxoviridae* family is a group of enveloped negative-stranded RNA viruses that infect a broad spectrum of species, including humans, domestic (farm and pet) and wild animals (Lamb *et al.*, 2000). Entry of paramyxovirus into the host cells is the first step in virus infection and is mediated by viral and cellular membrane fusion (Lamb, 1993; Lamb *et al.*, 1999; Peisajovich & Shai, 2002, 2003; Morrison, 2003). During this process, the viral envelope proteins embedded on the virion surface play an important role. In SeV, there are two envelope proteins on the surface: haemagglutinin neuraminidase (HN) and fusion protein (F). HN, called the attachment protein, is responsible for initial binding to the sialic acid-containing target-cell receptor and the process triggers conformational changes in the F protein (reviewed in Peisajovich & Shai, 2002, 2003). F protein is initially synthesized as an inactive precursor F_0 and subsequently cleaved by a cellular furin-like enzyme into two disulfide-bond-linked subunits, F_1 and F_2. The covalently linked F_1 and F_2 form a trimer on the virion surface (Lamb, 1993). There are several important functional or structural modules/domains in the SeV F protein, some of which are confirmed to be common to all paramyxoviruses. These include the highly conserved heptad-repeat (HR) regions, F_1 N-terminal fusion peptide (FP), internal fusion peptide (IFP) and an additional internal leucine-zipper (LZ) domain with a virus-fusion inhibition effect (reviewed in Peisajovich & Shai, 2002, 2003; Morrison, 2003). In particular, the HR domains (HR1 and HR2 or HR-A and HR-B) have been shown to possess unique structures in different conformational states of the F protein (Lamb, 1993; Lamb *et al.*, 1999; Peisajovich & Shai, 2002, 2003; Morrison, 2003). It has also been shown that both HR1 and HR2 or HR2 alone have fusion-inhibition effects on virus entry (Lamb *et al.*, 1999; Peisajovich & Shai, 2002, 2003; Morrison, 2003). In the case of SeV, only HR2 shows an inhibitory effect (Rapaport *et al.*, 1995). In the current model, at least three conformational

states of the F protein have been proposed: pre-fusion native conformation, pre-hairpin intermediate conformation and post-fusion coiled-coil conformation (Lamb *et al.*, 1999; Peisajovich & Shai, 2002, 2003; Morrison, 2003). HR1 and HR2 are believed to be separated in the pre-fusion native state and the intermediate state, but may be in different conformations. However, it is clear that they interact with each other to form a six-helix coiled-coil bundle in the post-fusion state as shown by biochemical investigation of a number of paramyxoviruses (Lamb *et al.*, 1999; Peisajovich & Shai, 2002, 2003; Morrison, 2003) and the crystal structures of human respiratory syncytial virus (hRSV) and simian parainfluenza virus 5 (SV5) (Baker *et al.*, 1999; Zhao *et al.*, 2000), both of which are members of the *Paramyxoviridae*. In this six-helix coiled-coil bundle, both HR1 and HR2 are α-helical and three of the HR1 domains interact with each other to form a homotrimeric coiled-coil core surrounded by three HR2 domains in an antiparallel manner (Lamb *et al.*, 1999; Eckert & Kim, 2001). This six-helix coiled coil bundle has been called the fusion core and is believed to be the most stable structure of the F protein. The energy release during the conformational state transitions is believed to be used for viral–cellular membrane fusion.

In an attempt to understand the structural basis and element requirements of the six-helix bundle formation of *Paramyxoviridae* family members, we have carried out a number of paramyxovirus fusion-core analyses by methods including biochemistry, biophysics and X-ray crystallography (*e.g.* Zhu *et al.*, 2002; Yu *et al.*, 2002). Although the SeV HR1 and HR2 regions have previously been studied by Shai and colleagues (Ghosh *et al.*, 1997, 1998), they did not show the formation of the six-helix bundle complex because they used shorter amino-acid sequences for these experiments which did not cover the whole HR regions (see Fig. 1 for details). In their experiments, they only showed that the two regions interact with each other with α-helix structure by labelling the HR2 with fluorescent reagents. Therefore, in this study we have designed a single-chain HR1–linker–HR2 of SeV (2-Helix) and expressed the protein in a prokaryotic system. Subsequently, a series of biochemical, biophysical and X-ray diffraction techniques were employed for analysis. We show here that HR1 and HR2 of SeV F protein interact with each other and form a typical six-helix coiled-coil bundle: the fusion core. This stable protein has been crystallized and the detailed crystal structure

has been solved, showing a typical six-helix coiled-coil structure, which will be reported elsewhere.

2. Materials and methods

2.1. Gene construction, protein expression and purification

The F-protein sequence used in this study was derived from SeV strain T5 (GenBank Accession No. M69046) and its functional/structural modules/domains are depicted in Fig. 1(*a*). To choose the HR1 and HR2 sequences the *LearnCoil-VMF* program (Singh *et al.*, 1999) was used, as it was specifically designed for identification of viral heptad-repeat coiled coils and has been successfully used for a number of virus-fusion proteins. These regions have been studied previously by Shai and coworkers and the peptides were named SV150 or SV163 for HR1 and SV465 or SV473 for HR2 (Rapaport & Shai, 1994; Rapaport *et al.*, 1995; Ghosh *et al.*, 1998). As can be seen in Fig. 1(*b*), the sequences we chose were closer to the prediction than those used in the series of studies carried out by Shai and coworkers. For the 2-Helix gene construction, the HR1 region used was derived from amino acids 142–188 and the HR2 region used was derived from amino acids 462–493. The HR1 region is quite close to the prediction, but for the sake of solubility eight amino acids were omitted. The HR2

region we used is exactly the same as the prediction. The gene for 2-Helix, *i.e.* HR1 and HR2 connected by a six amino-acid linker SGGRGG (single-letter amino-acid abbreviations), was constructed by PCR and cloned into the *Bam*HI–*Xho*I restriction sites of the GST (glutathione *S*-transferase) fusion expression vector pGEX-6P-1 (Pharmacia), in which there is a rhinovirus 3C protease-cleavage site for the fusion protein (the same as the commercial PreScission protease-cleavage site). The positive plasmids (named GEXSeV-2-Helix) were verified by direct DNA sequencing.

For protein expression and purification, the experiments were carried out using the same methods as previously described (Zhu *et al.*, 2002). The purified GST-fusion proteins were cleaved by GST-fusion rhinovirus 3C protease (construct kindly supplied by Drs K. Hudson and J. Heath) and free GST, non-cleaved GST-SeV-2-Helix and GST-3C were removed by passing over a glutathione-Sepharose 4B column. Purified 2-Helix was further purified by gel-filtration on a Superdex G75 column (Pharmacia) with an Äkta Explorer FPLC system (Amersham Pharmacia).

2.2. Biochemical and biophysical analyses of the complex

The purified 2-Helix was further characterized by mass spectroscopy and circular-

(*a*)
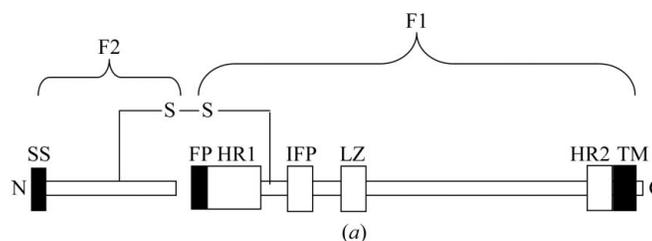
Predicted HR1: AGIALAEAREAKRDIALIKESMTKTHKSIELLQNAVGEQILALKTLQDFVNDEIK

Shai SV150:   DIALIKESMTKTHKSVELLQNAVGEQILALKTLQDFV

Shai SV163:   KSVELLQNAVGEQILALKTLQDFV

Our construct: AEAREAKRDIALIKESMTKTHKSIELLQNAVGEQILALKTLQDFVND

Predicted HR2:   DISLNLADATNFLQDSKAELEKARKILSEVGR

Shai SV465:   LNLADATNFLQDSKAELEKARKILSEVGRWY

Shai SV473:   FLQDSKAELEKARKILSEVGRWY

Our construct: DISLNLADATNFLQDSKAELEKARKILSEVGR

2-Helix: HR1 <u>SGGRGG</u> HR2

*There is a conserved substitution from valine to isoleucine at amino acid 165 in HR1 between Shai's and our strains.

(*b*)

Figure 1
Location of HR regions used in this study in the F-protein gene and their amino-acid sequences. (*a*) Schematic diagram of SeV F protein showing F1 and F2 and the functional/structural modules (domains) in addition to HR1 and HR2. SS, signal sequence; FP, fusion peptide; IFP, internal fusion peptide; LZ, leucine-zipper domain (amino acids 201–229); TM, transmembrane region. 'S—S' represents the disulfide bond linking F1 and F2. (*b*) Sequence alignment of HR1 and HR2 predicted by *LearnCoil-VMF* (Singh *et al.*, 1999) or used by Shai and coworkers and here (our construct is named 2-Helix, in which HR1 and HR2 are linked by a linker, SGGRGG.)

dichroism (CD) spectroscopy as described previously (Zhu *et al.*, 2002).

2.3. Crystallization, X-ray data collection and processing

The 2-Helix protein was buffer-exchanged into 20 m*M* Tris–HCl buffer pH 8.0 and concentrated to 10 mg ml^{-1}. The initial crystallization conditions were screened using the hanging-drop vapour-diffusion method with sparse-matrix crystallization kits (Crystal Screens I and II; Hampton Research, Riverside, CA, USA). 1 µl protein solution was mixed with 1 µl reservoir solution and equilibrated against 200 µl reservoir solution. The crystallization trays were kept at 291 K. 2-Helix protein crystals were obtained and the conditions were optimized to obtain better diffracting crystals.

The X-ray diffraction data for a single crystal were collected at 100 K with a MAR Research (Norderstadt, Germany) image-plate detector using Cu *Kα* radiation from a Rigaku rotating-anode generator. The collected data were processed using the *HKL* programs *DENZO* and *SCALEPACK* (Otwinowski & Minor, 1997).

3. Results and discussion

As described in §2, the SeV HR1 and HR2 regions were predicted using the *LearnCoil-VMF* program (Singh *et al.*, 1999) (Fig. 1) and our constructed HR1 and HR2 regions were subsequently amplified by PCR and cloned into the GST-fusion expression vector pGEX-6P-1. To facilitate crystallization, the two regions were linked by a flexible six-amino-acid linker and the construct (HR1–linker–HR2) was named 2-Helix. This strategy has also been used in our laboratory for other paramyxoviruses (*e.g.* Zhu *et al.*, 2002). Because the *Learn-Coil-VMF* prediction gave longer amino-acid sequences for both regions than those used by Shai and coworkers, we chose the HR1 region (47 amino acids) to be close to the prediction and the HR2 region (32

Table 1
Diffraction data statistics.

Space group	R3
Unit-cell parameters (Å, °)	$a = 42.149$, $b = 42.149$, $c = 182.577$, $\alpha = 90$, $\beta = 90$, $\gamma = 120$
Wavelength (Å)	1.5418
Resolution (Å)	2.5
Observed reflections	24732
Unique reflections	5274
Completeness (%)	100 (100)
$I/\sigma(I)$	5.5 (5.2)
R_{sym} (%)	12.9 (43.6)

amino acids) to be identical to the prediction as shown in Fig. 1. Five amino acids from the N-terminus and three from the C-terminus in the HR1 region were omitted compared with the prediction in order to increase the solubility of the polypeptide.

When 2-Helix (85 amino acids in total) was expressed in *Escherichia coli* as a GST-fusion protein, it was soluble and could subsequently be cleaved from the fusion partner GST (Fig. 2*a*). The protein was purified by affinity column and gel filtration with high homogeneity as shown by MALDI–TOF (Bruker Daltonics Biflex III) mass spectrometry (Fig. 2*b*) (9.718 kDa, the molecular weight of 85 amino acids in 2-Helix protein). A single symmetrical peak of 2-Helix protein was observed on gel filtration (Fig. 2*c*) and estimation of the molecular weight showed the formation of the trimer (approximately 30 kDa, 3 × HR1–HR2, consistent with other paramyxovirus fusion cores, including those of hRSV and SV5, the crystal structures of which have been solved (Baker *et al.*, 1999; Zhao *et al.*, 2000).

To test the secondary structure of the purified 2-Helix protein, CD spectrometry was carried out and the results show that 2-Helix protein forms a typical *α*-helix structure (Fig. 2*d*). The melting temperature of the 2-Helix protein is as high as 370 K (data not shown), indicating that it is extremely stable, and the virus fusion core is characterized by the coiled-coil structure (Lamb *et al.*, 1999; Eckert & Kim, 2001). All these results including the gel-filtration data support the formation of a stable coiled-coil six-helix bundle in the 2-Helix protein.

Crystallization of 2-Helix protein proved to be easy under several conditions (data not shown), indicating its stable tightly packed structure. A crystal grown (within days) from a condition consisting of 0.1 *M* citric acid pH 2.5, 20% PEG 4000 (Fig. 3) diffracted X-rays to 2.5 Å and belonged to space group *R*3, with unit-cell parameters $a = b = 42.149$, $c = 182.57$ Å. Data-collection statistics are given in Table 1. Molecular-replacement calculations were performed
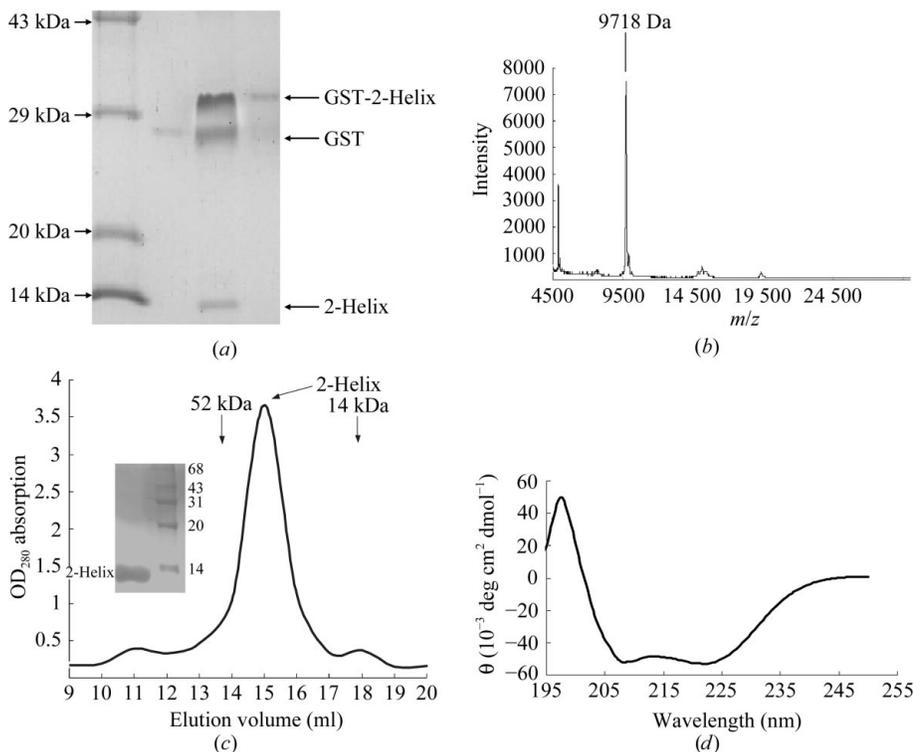
Figure 2
Characterization of the 2-Helix protein. (*a*) SDS–PAGE analysis of the proteins. 2-Helix protein was expressed as a GST-fusion protein and cleaved by GST-3C enzyme. Lane *M*, protein molecular-weight markers in kDa as indicated; lane 2, purified GST control protein; lane 3, 2-Helix protein after GST-3C cleavage; lane 4, purified GST-fusion 2-Helix. (*b*) MALDI–TOF mass-spectrometry profile of the 2-Helix. It is clearly shown that the molecular weight of the major peak is 9.718 kDa, matching the molecular weight of the 2-Helix monomer. (*c*) Gel-filtration analysis of the 2-Helix protein. In the Superdex G75 column elution pattern, a clear peak between the eluted volumes corresponding to 52 and 14.4 kDa standards indicates the 2-Helix protein. It is of about 30 kDa in weight, indicating the formation of a 2-Helix trimer or six-helix bundle. The inset picture is Tris–tricine SDS–PAGE analysis of the peak, indicating the existence of only 2-Helix protein in the peak. The molecular standards run are indicated in kDa. (*d*) CD spectra of the 2-Helix protein. A typical *α*-helix secondary structure with double minima at 208 and 222 nm was seen for the 2-Helix protein sample.
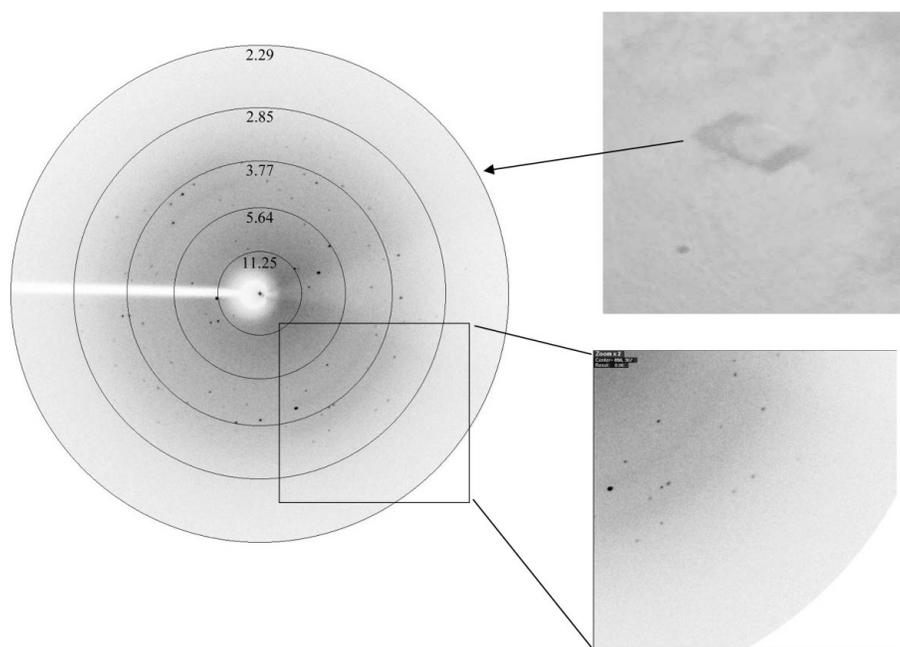
Figure 3
Sample 2-Helix crystal and its X-ray diffraction pattern.

using the program *CNS* (Brünger *et al.*, 1998) with the crystal structure of SV5 N1/C1 trimer as a search model (PDB code 1svf; Baker *et al.*, 1999). The structure has been solved, showing a typical six-helix coiled-coil structure, which will be reported elsewhere. Recently, a cryo-electron microscopic (EM) structure of the SeV F protein was solved at 16 Å and a six-helix bundle structure was proposed in the stalk region (Ludwig *et al.*, 2003), which must be formed of HR1 and HR2 regions. Using the information from the HR1–HR2 fusion-core structure, we should be able to produce a picture of the structural requirement of SeV F protein in virus membrane fusion.

References

Baker, K. A., Dutch, R. E., Lamb, R. A. & Jardetzky, T. S. (1999). *Mol. Cell*, **3**, 309–319.

Brünger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J.-S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T. & Warren, G. L. (1998). *Acta Cryst.* D**54**, 905–921.

Chang, P.-C., Hsieh, M.-L., Shien, J.-H., Graham, D. A., Lee, M.-S. & Shieh, H. K. (2001). *J. Gen. Virol.* **82**, 2157–2168.

Eckert, D. M. & Kim, P. S. (2001). *Annu. Rev. Biochem.* **70**, 777–810.

Ghosh, J. K., Ovadia, M. & Shai, Y. (1997). *Biochemistry*, **36**, 15451–15462.

Ghosh, J. K., Peisajovich, S. G., Ovadia, M. & Shai, Y. (1998). *J. Biol. Chem.* **273**, 27182–27190.

Lamb, R. A. (1993). *Virology*, **197**, 1–11.

Lamb, R. A., Collins, P. L., Kolakofsky, D., Melero, J. A., Nagai, Y., Oldstone, M. B. A., Pringle, C. R. & Rima, B. K. (2000). *Virus Taxonomy: Classicification and Nomenclature of Viruses, 7th Report*, edited by M. H. V. van Regenmortel, C. M. Fauquet, D. H. L. Bishop, E. B. Carstens, M. K. Estes, S. M. Lemon, J. Maniloff, M. A. Mayo, D. J. McGeoch, C. R. Pringle & R. B. Wickner, pp. 549–561. San Diego: Academic Press.

Lamb, R. A., Joshi, S. B. & Dutch, R. E. (1999). *Mol. Membr. Biol.* **16**, 11–19.

Ludwig, K., Baljinnyam, B., Herrmann, A. & Bottcher, C. (2003). *EMBO J.* **22**, 3761–3771.

Morrison, T. G. (2003). *Biochim. Biophys. Acta*, **1614**, 73–84.

Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.

Peisajovich, S. G. & Shai, Y. (2002). *Trends Biochem. Sci.* **27**, 183–190.

Peisajovich, S. G. & Shai, Y. (2003). *Biochim. Biophys. Acta*, **1614**, 122–129.

Rapaport, D., Ovadia, M. & Shai, Y. (1995). *EMBO J.* **14**, 5524–5531.

Rapaport, D. & Shai, Y. (1994). *J. Biol. Chem.* **269**, 15124–15131.

Singh, M., Berger, B. & Kim, P. S. (1999). *J. Mol. Biol.* **290**, 1031–1041.

Wang, L., Harcourt, B. H., Yu, M., Tamin, A., Rota, P. A., Bellini, W. J. & Eaton, B. T. (2001). *Microbes Infect.* **3**, 279–287.

Yu, M., Wang, E., Liu, Y., Cao, D., Jin, N., Zhang, C. W.-H., Bartlam, M., Rao, Z., Tien, P. & Gao, G. F. (2002). *J. Gen. Virol.* **83**, 623–629.

Zhao, X., Singh, M., Malashkevich, V. N. & Kim, P. S. (2000). *Proc. Natl Acad. Sci. USA*, **97**, 14172–14177.

Zhu, J., Zhang, C. W.-H., Qi, Y., Tien, P. & Gao, G. F. (2002). *Biochem. Biophys. Res. Commun.* **299**, 897–902.