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

Cloning, expression and preliminary crystallographic analysis of the equine infectious anaemia virus (EIAV) gp45 ectodomain

Pei-Long Sun,[a,b,c] Shu-Xia Lv,[b] Jian-Hua Zhou[c] and Xin-Qi Liu[a]*

[a]College of Life Science, Nankai University, Tianjin 300071, People's Republic of China, [b]Institute of Biological Science and Technology, Shenyang Agricultural University, Shenyang 110866, People's Republic of China, and [c]Harbin Veterinary Research Institute, Chinese Academy of Agricultural Sciences, Harbin 150001, People's Republic of China

Correspondence e-mail: liu2008@nankai.edu.cn

Received 2 December 2010
Accepted 7 February 2011

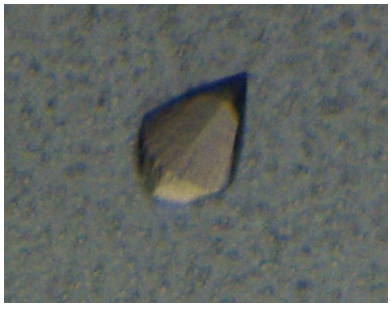
Like human immunodeficiency virus (HIV), equine infectious anaemia virus (EIAV) belongs to the lentivirus genus. The first successful lentiviral vaccine was developed for EIAV. Thus, EIAV may serve as a valuable model for HIV vaccine research. EIAV glycoprotein 45 (gp45) plays a similar role to gp41 in HIV by mediating virus–host membrane fusion. The gp45 ectodomain was constructed according to the structure of HIV gp41, with removal of the disulfide-bond loop region. The protein was expressed in *Escherichia coli* and crystallized following purification. However, most of the crystals grew as aggregates and could not be used for data collection. By extensively screening hundreds of crystals, a 2.7 Å resolution data set was collected from a single crystal. The crystal belonged to space group $P6_3$, with unit-cell parameters $a = b = 46.84$, $c = 101.61$ Å, $\alpha = \beta = 90$, $\gamma = 120^\circ$. Molecular replacement was performed using the coordinates of various lengths of HIV gp41 as search models. A long bent helix was identified and a well defined electron-density map around the long helix was obtained. This primary model provided the starting point for further refinement.

1. Introduction

Since HIV was implicated in acquired immunodeficiency syndrome (AIDS) in 1981, the disease has claimed over 25 million lives. Even today, HIV still causes 2.5 million new infections worldwide each year (Council of the Global HIV Vaccine Enterprise, 2010). Despite the success of antiretroviral therapy, its high cost and severe side effects and the emergence of viral resistance have posed substantial challenges in the treatment of HIV-infected individuals. An efficient HIV vaccine that could prevent all subtypes of HIV is urgently needed. Unfortunately, all HIV vaccine candidates tested to date apparently give no appreciable protection. Only recently has the relative protective effect observed in a Thai trial (RV144) shed some light on HIV vaccine development (Haynes *et al.*, 2010).

Research on equine infectious anaemia virus (EIAV), which is related to HIV, has long been neglected. EIAV only infects equids and causes a persistent infection characterized by recurring febrile episodes associated with viraemia, fever, thrombocytopaenia and wasting symptoms (Leroux *et al.*, 2004). An EIAV vaccine was developed more than 30 years ago (Shen *et al.*, 1979). Unlike HIV-1, EIAV only generates periodic peaks of viraemia and fever within asymptomatic stages and is nonlethal in most infected horses. The EIAV RNA genome is the smallest of the lentiviruses. In addition to the structural proteins encoded by *gag*, *pol* and *env*, which are comparable to those of HIV, EIAV lacks most of the accessory proteins that exist in HIV, with the exceptions of Tat and Rev (Montelaro *et al.*, 1993). EIAV has a life cycle similar to that of HIV and thus could be used as a reference model in HIV vaccine studies.

The gp45 protein mediates virus–host-cell membrane fusion during EIAV infection, similar to gp41 in HIV (Gonzalez-Scarano *et al.*, 1987). gp45 has a core structure similar to that of gp41, forming a six-helical bundle *via* N-terminal heptad repeats (HR1) and C-terminal heptad repeats (HR2) as the postfusion conformation. The sequence homology between EIAV gp45 and different strains of HIV gp41 is approximately 20–25% according to *BLAST* (Johnson *et al.*, 2008).



No structural information is available for the prefusion conformation of gp41/gp45 from any lentivirus. All of the published gp41 structures are in the postfusion conformation consisting of a six-helix bundle that corresponds to the conformation of gp41 after fusion between membranes of HIV and the host cell. Although the length of HR1/HR2 varies from the reported HIV gp41 structures, the overall architecture of the six-helix bundle is retained well. The crystal structure of simian immunodeficiency virus (SIV) gp41 (PDB code 1qbz; Roux & Taylor, 2007) contains the longest helices. To solve the structure of EIAV gp45, we cloned, expressed and crystallized the ectodomain of EIAV gp45 from the $EIAV_{LN40}$ strain (Qi *et al.*, 2010). To facilitate purification, a short stretch of the disulfide-bond loop region was omitted and most of the sequences predicted within the helix region were included in our construct (Fig. 1).

2. Materials and methods

2.1. Cloning

The ectodomain of EIAV gp45 (henceforth termed gp45-ecto) was constructed according to a previous HIV gp41 crystallographic report (Chan *et al.*, 1997). Briefly, the HR1 and HR2 regions were connected by a GGSGG (GGAGGTTCCGGCGGG) linker and the disulfide-bond loop between HR1 and HR2 was replaced by this linker. HR1 and HR2 were amplified by polymerase chain reaction (PCR) using the gene encoding full-length EIAV gp140 (gp90 plus gp45) from the $EIAV_{LN40}$ strain (Qi *et al.*, 2010) as a template. The forward primer for gp45-HR1 was 5′-TACTTCCAATCCAATGCCGATAGTGTA-CAAAAT-3′ and the reverse primer was 5′-CCCGCCGGAACCT-CCAATCAAATTAAATGT-3′. The forward primer for gp45-HR2 was 5′-GGAGGTTCCGGCGGGTCTACACAGTGGGAT-3′ and the reverse primer was 5′-TTATCCACTTCCAATGCTATTTTCC-AAATTGTGC-3′. The ligation-independent cloning (LIC) method was used in cloning. Firstly, the two fragments of HR1 and HR2 were amplified by PCR; the two fragments were then connected by overlap PCR through the overlap region of the GGSGG linker. The pET30-TEV/LIC vector was digested with the *Ssp*I restriction enzyme and was then extracted using a Gel Extraction Kit (Axygen). Cleaved plasmids and the purified PCR products were digested with T4 DNA polymerase (Promega) in the presence of dGTP or dCTP. The annealed mixture was transformed into *Escherichia coli* DH5α competent cells for plasmid propagation. Using this method, a 6×His tag and a TEV cleavage site originating from the LIC vector were attached to the N-terminus of gp45-ecto. The correct construct was confirmed by sequencing. The detailed sequence of the expression construct is shown in Fig. 1.

2.2. Expression and purification

EIAV gp45-ecto was expressed as an N-terminally 6×His-tagged protein. The expression construct was transformed into *E. coli* strain BL21 (DE3). A single clone was chosen from the agar plate and inoculated into Luria broth (LB) medium containing 50 µg ml^{-1} kanamycin. After the optical density at 600 nm of the culture reached 0.6–0.8 at 310 K, protein expression was induced with 0.2 m*M* isopropyl β-D-1-thiogalactopyranoside (IPTG) and growth of the culture continued for 12 h at 291 K. The cells were harvested by centrifugation (6000*g*, 20 min, 291 K) and suspended in 50 m*M* Tris–HCl, 500 m*M* NaCl and 4 m*M* imidazole pH 8.0. The cell suspension was lysed by sonication using a sonifier (SEIENTZ-II) for 10 min on ice. The crude cell extract was clarified by centrifugation at 38 000*g* for 40 min.

The supernatant was loaded onto a 3 ml gravity-flow column of Ni–NTA resin (Qiagen) after equilibration with four column volumes of binding buffer (50 m*M* Tris–HCl, 500 m*M* NaCl, 4 m*M* imidazole pH 8.0). Nonspecifically binding proteins were washed away with four column volumes of washing buffer (50 m*M* Tris–HCl, 500 m*M* NaCl, 20 m*M* imidazole pH 8.0). The target protein was eluted with four column volumes of elution buffer (50 m*M* Tris–HCl, 500 m*M* NaCl, 500 m*M* imidazole pH 8.0). After transfer to a lower salt buffer (50 m*M* Tris–HCl, 10 m*M* NaCl pH 8.0) by ultrafiltration with a 10 kDa cutoff filter (Amicon), the protein was further purified on a HiTrap Q HP column (GE Healthcare) with a linear gradient of NaCl from 100 to 600 m*M*. The target protein was then polished on a Superdex-200 gel-filtration column (GE Healthcare) with a solution consisting of 50 m*M* Tris–HCl and 300 m*M* NaCl pH 8.0. The purity of the eluted protein was verified by SDS–PAGE analysis (Fig. 2).

2.3. Crystallization

The 6×His-tagged protein was concentrated to 8 mg ml^{-1} in 50 m*M* Tris–HCl, 300 m*M* NaCl pH 8.0 using an Amicon Ultra filter (Millipore). Protein crystallization screens were performed using the sitting-drop vapour-diffusion technique (1 µl protein solution was mixed with 1 µl reservoir solution and equilibrated against 100 µl mother liquor at 293 K). Crystal screening was performed manually in 48-well plates (BioXtal). Protein crystals appeared in two different conditions overnight: solution Nos. 21 (0.1 *M* sodium phosphate monobasic monohydrate, 0.1 *M* potassium phosphate monobasic, 0.1 *M* MES monohydrate pH 6.5, 2.0 *M* sodium chloride) and 25 [0.01 *M* cobalt(II) chloride hexahydrate, 0.1 *M* MES monohydrate pH 6.5, 1.8 *M* ammonium sulfate] of Crystal Screen 2 (Hampton Research). Further optimization of these conditions was performed by adjusting the pH and the concentration of the precipitants. The

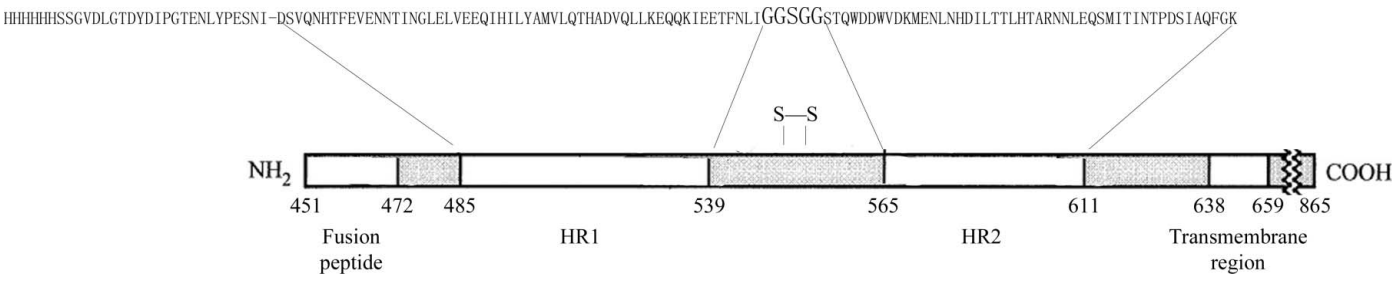


Figure 1
Sequence of gp45-ecto. The HR1 (485–539) and HR2 (565–611) regions used in the expression construct are labelled and the corresponding sequence is listed at the top. A GGSGG linker was used to replace the disulfide-bond loop. An additional N-terminal sequence containing a 6×His tag and a TEV cleavage site was included in our construct. The numbers under the bar correspond to the amino-acid positions in the gp90–gp45 precursor.

crystallization condition that used sodium and potassium phosphate as precipitants based on solution No. 21 of Crystal Screen 2 produced single crystals that were suitable for data collection (Fig. 3). The final optimized solution for crystal growth was a mixture consisting of 0.15 *M* sodium phosphate monobasic monohydrate, 0.1 *M* potassium phosphate monobasic, 0.1 *M* MES monohydrate pH 6.5 and 1.4 *M* sodium chloride.

2.4. Data collection

The mother liquor supplemented with 20% glycerol was used as the cryoprotectant for data collection at 100 K. The gp45-ecto crystals were soaked in the cryoprotectant for a few seconds and then mounted directly onto the goniometer head. Most of the crystals grew as aggregates. Some crystals that appeared to be smooth on the surface were shown to be clusters of single crystals by their diffraction image. Hundreds of crystals were screened and finally a single crystal (0.1 × 0.1 × 0.15 mm) with acceptable mosaicity was used for data collection (Fig. 4). One data set was collected to 2.7 Å resolution at 100 K using 180° rotation in increments of 1° per frame. The data were processed using the *HKL*-2000 package (Otwinowski & Minor, 1997) and converted to amplitudes using the *CCP*4 suite (Collaborative Computational Project, Number 4, 1994). A summary of the data-collection statistics is given in Table 1.

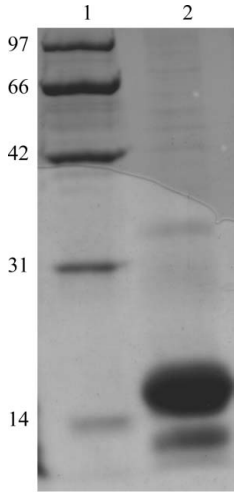


Figure 2
SDS–PAGE analysis of gp45-ecto after purification. Lane 1, molecular-mass markers (kDa); lane 2, purified gp45-ecto protein with a molecular mass of approximately 15 kDa.

Table 1
Data-collection and processing statistics.

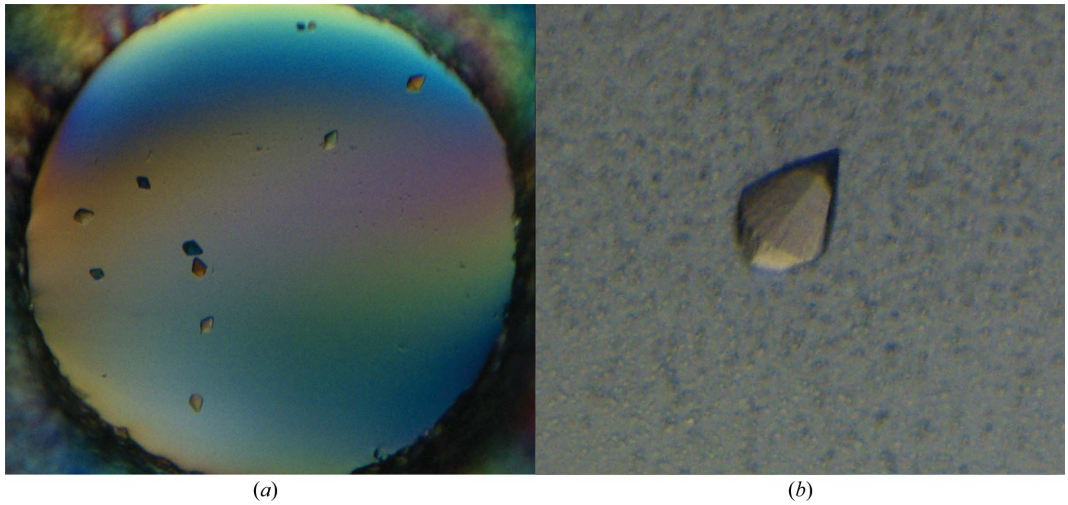
Values in parentheses are for the outer resolution shell.

X-ray source	Beijing Synchrotron Radiation Facility, Beijing, China
Wavelength (Å)	0.9798
Detector	MAR 165 mm charge-coupled device
Crystal parameters	
Space group	$P6_3$
Unit-cell parameters	
a (Å)	46.84
b (Å)	46.84
c (Å)	101.61
α (°)	90
β (°)	90
γ (°)	120
Resolution (Å)	2.70
Mosaicity (°)	0.86
Matthews coefficient (Å^3 Da^{-1})	2.09
Solvent content (%)	41.12
Molecules per asymmetric unit	1
Data processing	
No. of observed reflections	37404 (1506)
No. of unique reflections	3981 (193)
Completeness (%)	99.6 (97.3)
Multiplicity	9.4 (7.6)
$R_{\rm merge}$† (%)	5.5 (24.4)
Average $I/\sigma(I)$	16.2 (2.4)
Wilson B factor (Å^2)	55.45

† $R_{\rm merge} = \sum_{hkl}\sum_i |I_i(hkl) - \langle I(hkl)\rangle| / \sum_{hkl}\sum_i I_i(hkl)$, where $I_i(hkl)$ is the observed intensity and $\langle I(hkl)\rangle$ is the average intensity of multiple observations of symmetry-related reflections.

3. Results and discussion

Lentiviruses, including HIV, are notoriously difficult targets for vaccine development. Only two vaccines, those against EIAV and feline immunodeficiency virus, have successfully been developed (Shen *et al.*, 1979; Andersen & Tyrrell, 2004). 30 years ago, Chinese scientists developed an EIAV vaccine using a live-attenuated EIAV strain produced by serial passage through donkey leukocyte cells (Shen *et al.*, 1979). Both the wild-type and vaccine strains of EIAV are available, which provides invaluable information for HIV vaccine development. EIAV gp45 has low sequence homology to HIV gp41,



(*a*) (*b*)

Figure 3
Crystals of gp45-ecto. (*a*) Crystals appeared in solution No. 21 of Crystal Screen 2 (Hampton Research). (*b*) Crystal used for X-ray data collection.

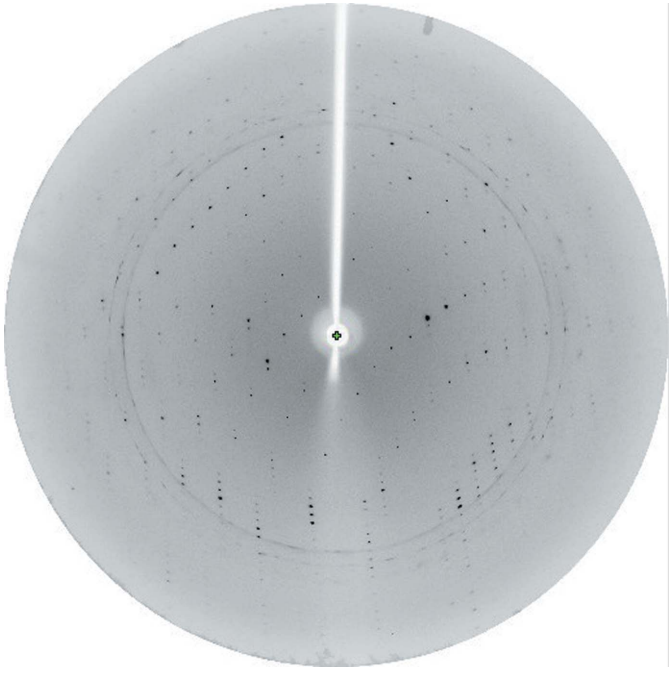


Figure 4
One representative diffraction image of the gp45-ecto crystal. The resolution at the edge of the frame is 2.6 Å.

but the two proteins are believed to play similar roles in driving virus–host-cell membrane fusion (Liang *et al.*, 2006).

On SDS–PAGE analysis of gp45-ecto, a weak band smaller than gp45-ecto was observed. It appeared during protein purification and became more apparent when the protein was stored for a few days. We speculate that these extra bands were degradation products of the protein. The purified protein exists as a trimer in solution, which was confirmed by analytical ultracentrifugation (results not shown). Each monomer includes two helices, HR1 and HR2, so that the trimer includes six helices. We believe that this oligomeric state corresponds to the postfusion conformation of the protein found in the crystal, but this will be confirmed when the structure is solved.

An interesting issue is the contribution of the N-terminal 6×His tag to crystallization. As part of an attempt to refine the crystallization conditions, we removed the 6×His tag. No crystals appeared under the previous crystallization conditions or in new cycles of crystal screening. We speculate that the N-terminal 6×His tag may therefore be involved in crystal packing.

As the gp45-ecto crystals easily grew as aggregates of multiple crystals, obtaining a single crystal for data collection was difficult. Hundreds of crystals were tried before the data set was finally collected. We used *phenix.xtriage* (Adams *et al.*, 2010) to check for merohedral twinning problems in our data. To a resolution of 2.7 Å, no merohedral twinning could be identified. Therefore, the indexing problem in data collection was attributable to the presence of multiple lattices resulting from multiple crystals rather than twinning.

EIAV gp45-ecto has low sequence homology to HIV gp41 (20–25% identity depending on the HIV subtype). We used several published HIV gp41 structures (PDB entry 2x7r, Buzon *et al.*, 2010; PDB entry 1aik, Chan *et al.*, 1997; PDB entry 3k9a, Shi *et al.*, 2010; PDB entry 1env, Weissenhorn *et al.*, 1997) as starting models to perform molecular replacement against our EIAV gp45-ecto data set using the *Phaser* program (McCoy *et al.*, 2007) from the *CCP*4 suite (Collaborative Computational Project, Number 4, 1994). Initially, no reliable solution was found. When the single helices from these structures were tried, the long helix from PDB entry 3k9a (Shi *et al.*, 2010), corresponding to HR2 and MPER (the membrane-proximal external region) of HIV gp41, generated a solution with a log-likelihood gain of 94. By checking the packing, this long helix formed a three-helix bundle through the threefold noncrystallographic symmetry axis. The electron density of the backbone of this 50-amino-acid helix was clearly discernible. Further refinement confirmed that this solution was correct.

This preliminary study showed that this substructure corresponded to the inner trihelix: the HR1 region of gp45. Given the low sequence similarity between gp45 and gp41, the backbone of this long helix made the greatest contribution to this solution. Further refinement is needed to identify the HR2 region of gp45 and a detailed crystal structure will be available following refinement of this starting model.

We are grateful to Professor Yiming Shao at the Chinese Centre for Disease Control and Prevention. We are grateful to the staff at the Beijing Synchrotron Radiation Facility (Beijing, China) for help with data collection. This research was supported by the National S&T Major Project on Major Infectious Diseases (grant 2008ZX10001-010) and the National Basic Research Program of China (973 Program; grant 2010CB911800) from the Ministry of Science and Technology of the People's Republic of China.

References

Adams, P. D. *et al.* (2010). *Acta Cryst.* D**66**, 213–221.
Andersen, P. R. & Tyrrell, P. (2004). *Anim. Health Res. Rev.* **5**, 327–330.
Buzon, V., Natrajan, G., Schibli, D., Campelo, F., Kozlov, M. M. & Weissenhorn, W. (2010). *PLoS Pathog.* **6**, e1000880.
Chan, D. C., Fass, D., Berger, J. M. & Kim, P. S. (1997). *Cell*, **89**, 263–273.
Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* D**50**, 760–763.
Council of the Global HIV Vaccine Enterprise (2010). *Nature Med.* **16**, 981–989.
Gonzalez-Scarano, F., Waxham, M. N., Ross, A. M. & Hoxie, J. A. (1987). *AIDS Res. Hum. Retroviruses*, **3**, 245–252.
Haynes, B. F., Liao, H.-X. & Tomaras, G. D. (2010). *Curr. Opin. HIV AIDS*, **5**, 362–367.
Johnson, M., Zaretskaya, I., Raytselis, Y., Merezhuk, Y., McGinnis, S. & Madden, T. L. (2008). *Nucleic Acids Res.* **36**, W5–W9.
Leroux, C., Cadoré, J. L. & Montelaro, R. C. (2004). *Vet. Res.* **35**, 485–512.
Liang, H., He, X., Shen, R. X., Shen, T., Tong, X., Ma, Y., Xiang, W. H., Zhang, X. Y. & Shao, Y. M. (2006). *Arch. Virol.* **151**, 1387–1403.
McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C. & Read, R. J. (2007). *J. Appl. Cryst.* **40**, 658–674.
Montelaro, R. C., Ball, J. M. & Rushlow, K. E. (1993). *Retroviridae*, edited by J. A. Levy, pp. 257–360. New York: Plenum.
Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.
Qi, X., Wang, X., Wang, S., Lin, Y., Jiang, C., Ma, J., Zhao, L., Lv, X., Shen, R., Wang, F., Kong, X., Su, Z. & Zhou, J. (2010). *Virus Genes*, **41**, 86–98.
Roux, K. H. & Taylor, K. A. (2007). *Curr. Opin. Struct. Biol.* **17**, 244–252.
Shen, R. X., Xu, Z. D., He, X. S. & Zhang, S. X. (1979). *Chin. Agric. Sci.* **4**, 1–15.
Shi, W., Bohon, J., Han, D. P., Habte, H., Qin, Y., Cho, M. W. & Chance, M. R. (2010). *J. Biol. Chem.* **285**, 24290–24298.
Weissenhorn, W., Dessen, A., Harrison, S. C., Skehel, J. J. & Wiley, D. C. (1997). *Nature (London)*, **387**, 426–430.