Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

Crystallization and preliminary X-ray analysis of the matrix protein from Ebola virus

Andréa Dessen,[a] Eric Forest,[b] Viktor Volchkov,[c] Olga Dolnik,[c] Hans-Dieter Klenk[c] and Winfried Weissenhorn[a]*

[a]European Molecular Biology Laboratory (EMBL), 6 Rue Jules Horowitz, 38000 Grenoble, France, [b]Institut de Biologie Structurale Jean-Pierre Ebel, 41 Avenue des Martyrs, 38027 Grenoble, France, and [c]Institute for Virology, Robert-Koch-Strasse 17, 35037 Marburg, Germany

Correspondence e-mail: weissen@embl-grenoble.fr

Received 12 January 2000
Accepted 21 March 2000

The matrix protein from Ebola virus is a membrane-associated molecule that plays a role in viral budding. Despite its functional similarity to other viral matrix proteins, it displays no sequence similarity and hence may have a distinct fold. X-ray diffraction quality crystals of the Ebola VP40 matrix protein were grown by the hanging-drop vapour-diffusion method. The crystals belong to the monoclinic space group $C2$, with unit-cell parameters $a = 64.4$, $b = 91.1$, $c = 47.9$ Å, $\beta = 96.3^\circ$. A data set to 1.9 Å resolution has been collected using synchrotron radiation. The unit cell contains one molecule of molecular weight 35 kDa per asymmetric unit, with a corresponding volume solvent content of 35%.

1. Introduction

Ebola virus, which together with Marburg virus constitutes the filovirus family, is one of the most dangerous agents causing severe haemorrhagic fever in man and non-human primates (Feldmann & Klenk, 1996). Filoviruses are enveloped filamentous particles up to 14 µm long containing a non-segmented negative-strand RNA genome and seven structural proteins. Four polypeptides, NP, VP30, VP35 and L, are associated with the genomic RNA in the RNP (ribonucleoprotein) complex, while the glycoprotein Gp is anchored in the membrane. In addition, VP40 and VP24 are believed to be associated with the viral membrane as so-called matrix proteins. VP40 forms a layer internal to the viral lipid bilayer and has been suggested not to interact with the RNP complex (Elliott *et al.*, 1985; Kiley *et al.*, 1988). In contrast, VP24 is not completely removed from the RNP complex under isotonic conditions (Elliott *et al.*, 1985; Kiley *et al.*, 1988). Therefore, VP24 may link VP40 to the RNP complex.

Viral matrix proteins play an important role in the assembly and budding of viral particles and have been implicated in the sorting of the viral glycoproteins to budding virions. The cytoplasmic tails of viral glycoproteins are thought to interact with the matrix protein in the assembly process (Garoff *et al.*, 1998). There is also increasing evidence that the viral assembly machinery hijacks cellular proteins to perform this task. A WW binding-domain motif (Macias *et al.*, 1996; PY motif) was first noted in the retroviral gag polyprotein of Rous sarcoma virus (Garnier *et al.*, 1996); it has been shown to play an important role in a late stage of the budding process by interacting with the WW domain of the cellular YES-kinase associated protein (YAP; Garnier *et al.*, 1996; Xiang *et al.*, 1996). VP40 also contains a potential WW binding-domain motif (PY motif) at amino-acid positions 10–13 (PPEY) and filoviruses may therefore also involve cellular proteins in their assembly process.

X-ray and NMR structures of retroviral matrix proteins (Hill *et al.*, 1996; Conte *et al.*, 1997; Matthews *et al.*, 1996; Rao *et al.*, 1995; Massiah *et al.*, 1994) have revealed a highly similar fold consisting of two domains. The monomer consists of three α-helices and two short 3_{10} helical stretches, which pack tightly together into a globular domain, and a three-stranded mixed β-sheet that caps the N-terminal helical region. The N-terminal domain has been postulated to be involved in protein assembly and membrane binding, while the C-terminal region may play a part in a post-entry step (Hill *et al.*, 1996; Rao *et al.*, 1995). Large basic patches on the proposed membrane-binding face as well as an N-terminal myristoyl group (Gottlinger *et al.*, 1989) have been suggested as being crucial for bilayer targeting (Hill *et al.*, 1996; Rao *et al.*, 1995). The structure of a proteolytic fragment of influenza virus matrix protein (M) that forms a stable dimer with a mostly α-helical fold revealed its bifunctional mechanism of membrane association and RNP binding (Sha & Luo, 1997). Although presumably performing a similar function, the matrix protein from Ebola virus displays no sequence similarity to any of the known viral matrix protein structures, suggesting that a novel different matrix scaffold may be used by filoviruses. In order to understand the structural basis for the function of the Ebola virus matrix protein in light of structures of other viral matrix molecules, we prepared diffraction-

Table 1
Crystallographic statistics.

Data collection.

	Native	Pt† remote	Pt peak	Pt inflection
ESRF beamline	ID14-EH3	BM-14	BM-14	BM-14
Wavelength (Å)	$\lambda = 0.931$	$\lambda_1 = 1.0505$	$\lambda_2 = 1.0715$	$\lambda_3 = 1.0720$
Max. resolution‡ (Å)	1.9 (1.97–1.9)	3.0 (3.07–3.0)	3.0 (3.07–3.0)	3.0 (3.07–3.0)
$R_{\rm merge}$§	3.8 (30.0)	10.2 (17.9)	10.6 (18.0)	10.2 (19.9)
Completeness (%)	97.6 (97.3)	99.9 (99.4)	99.8 (97.6)	99.7 (97.1)
No. of reflections	148825	58657	60974	61176
No. of unique reflections	21500	5378	5610	5634
f' (e^-)¶		−5.23	−5.6	−5.82
f'' (e^-)		11.32	14.39	9.95

MAD phasing (*SHARP*).

Phasing power†† overall	1.01	$R_{\rm Cullis}$‡‡ overall	0.59
Figure of merit§§ overall	0.35	Number of sites	1

† Potassium tetrachloroplatinate (II) soaks. ‡ Values in parentheses are for the final resolution shell. § $R_{\rm merge} = \sum_h \sum_i |I - i(h) - \langle I(h)\rangle| / \sum I_i(h)$, where $I_i(h)$ is the ith measurement and $\langle I(h)\rangle$ is the weighted mean of all measurements of $I(h)$. ¶ f' and f'' reported values were refined by *SHARP*. †† Phasing power = $\langle |F_H| \rangle / E$, where $\langle |F_H| \rangle$ is the r.m.s. structure-factor amplitude for the heavy atom and E is the estimated lack-of-closure error. ‡‡ $R_{\rm Cullis}$ is the mean lack of closure divided by the mean isomorphous difference. §§ Figure of merit = $\langle |\sum P(\alpha) \exp(i\alpha) / \sum |P(\alpha)| \rangle$, where α is the phase and $P(\alpha)$ is the phase probability distribution.

quality crystals of Ebola virus VP40 and collected a high-resolution native data set.

2. Materials and methods

2.1. Expression and purification

The VP40 gene sequence corresponding to residues 1–326 (Ebola virus Zaire strain) was amplified with synthetic oligonucleotides by standard PCR methods and was cloned into expression vector pMAL-C2 (New England Biolabs) and transformed into *Escherichia coli* cells BL21 (DE3/pUBS) (Brinkmann *et al.*, 1989). Protein expression was induced with 1 m*M* IPTG and cells were harvested 3 h post-induction. Cells were lysed by sonication in 50 m*M* Tris pH 8.8, 100 m*M* NaCl, 10 m*M* DTT and the supernatant was cleared by centrifugation at 20 000 rev min^{-1}. Subsequently, the lysate was loaded onto an amylose column (New England Biolabs) and the purification of the fusion protein was performed according to the manufacturer's protocol (New England Biolabs). The final purification step included gel filtration in 20 m*M* bicine pH 9.3, 100 m*M* NaCl, 10 m*M* DTT. Factor Xa (Roche Biochemicals) cleavage was performed in the same buffer containing 2 m*M* $CaCl_2$ at room temperature for 2 h. N-terminal sequencing was performed after factor Xa cleavage.

A DNA fragment encoding the N-terminal factor Xa-specific digestion product (residues 31–326), generated by standard PCR procedures, was cloned into vector pRSET (Invitrogen) by introducing an additional methione at the N-terminus and expressed in *E. coli* BL21 (DE3/pUBS). Protein expression was induced with 1 m*M* IPTG for 3 h. Cells were harvested and centrifuged and pellets were resuspended and lysed by sonication in 50 m*M* Tris pH 8.8, 100 m*M* NaCl, 10 m*M* DTT. Insoluble material was pelleted at 20 000 rev min^{-1} for 1 h and the supernatant was loaded onto a Q-Sepharose column pre-equilibrated in lysis buffer. The VP40-containing flow-through was collected and VP40(31–326) was precipitated with 50% $(NH_4)_2SO_4$ in a 2:1(v/v) ratio. The pellet was resuspended in buffer containing 50 m*M* Tris pH 8.8, 100 m*M* NaCl, 10 m*M* DTT. A final gel-filtration step was performed on a Superose 6 column (Pharmacia) in buffer containing 20 m*M* bicine pH 9.3, 100 m*M* NaCl, 10 m*M* DTT. Purified VP40(31–326) was concentrated to 10 mg ml^{-1} for crystallization experiments. For derivatization of methionine residues with selenium, BL21/pUBS *E. coli* cells expressing VP40(31–326) were grown in minimal medium M9 supplemented with selenomethionine (75 mg l^{-1}) and other essential amino acids (lysine, phenylalanine, threonine at 100 mg l^{-1} each and isoleucine, leucine and valine at 50 mg l^{-1} each) before induction with 1 m*M* IPTG for 9 h. SeMet-VP40(31–326) was then purified as described above for wild-type VP40(31–326). Mass spectrometry of VP4(31–326) was performed on a Perseptive Voyager XL time-of-flight mass spectrometer.

Figure 1
Crystals of VP40(31–326). The dimensions of the largest crystals are approximately 0.2 × 0.05 × 0.03 mm.

2.2. Crystallization and X-ray analysis

Crystallization trials were set up at room temperature as hanging-drop vapour-diffusion experiments on Linbro crystallization plates. Initial screening was performed using the sparse-matrix method (Jancarik & Kim, 1991) with commercial crystal screening kits (Hampton Research). Protein samples were ultracentrifuged for 10 min at 0.2 MPa (Beckman air-driven ultracentrifuge) prior to crystallization. Optimization of crystallization parameters resulted in diffraction-quality crystals grown in drops of a mixture of 2 µl of protein with 1 µl of a reservoir solution containing 100 m*M* Tris pH 8.5, 25% PEG 750 monomethylether, 100 m*M* $MgCl_2$ and 0.2% β-octylglucoside (Sigma), suspended over 1 ml of reservoir solution. Rod-shaped crystals (Fig. 1) appeared overnight and grew to maximum dimensions of 0.2 × 0.05 × 0.03 mm over a 2–3 d period. SeMet-VP40(31–326) crystals grew under the same conditions. Cryo treatment of both native and selenomethionyl crystals was performed by quick transfer into a solution containing 100 m*M* Tris pH 8.5, 20% PEG 750 monomethylether, 100 m*M* $MgCl_2$, 12% glycerol, 0.2% β-octylglucoside followed by flash-cooling in a gaseous nitrogen stream. Native data were collected at the European Synchrotron Radiation Facility (ESRF, Grenoble, France) ID14 EH3 beamline set at 0.93 Å; 95 frames of data were collected in 1° oscillations at a crystal-to-detector distance of 100 mm and were recorded on a MAR CCD detector. A data set from a crystal derivatized with 200 m*M* potassium tetrachloroplatinate (II) (incubated in cryo-solution without glycerol for 12 h) was also collected at ID14 EH3 to a completeness of 80% and a resolution of 2.7 Å (data not shown). To perform a MAD experiment at the platinum $L_{\rm III}$ absorption edge, data sets at three wavelengths (Table 1) were collected at ESRF beamline BM-14. All data

were processed and scaled with *DENZO* and *SCALEPACK* (Otwinowski, 1993; Otwinowski & Minor, 1997). The *CCP*4 suite of programs was used for further data processing and analysis (Collaborative Computational Project, Number 4, 1994).

3. Results and discussion

Factor Xa treatment of VP40(1–326) expressed as a maltose fusion protein resulted in cleavage at the factor Xa site present in the vector as well as at Arg28, as determined by N-terminal sequencing. In order to produce a more compact VP40 for crystallization trials, we expressed the shorter fragment (31–326). The N-terminal sequence cleaved off with factor Xa and subsequently not included in the clone contains the putative PY WW domain-binding motif, indicating that this region may be flexible in an unbound state. VP40(31–326) was overexpressed in *E. coli* with a final yield of 20 mg l^{-1} after purification by anion-exchange chromatography, ammonium sulfate fractionation and gel filtration. Initial crystallization trials displayed a large amount of nucleation, a problem that was overcome by the addition of small amounts of detergent. Eventually, the inclusion of 0.2% β-octylglucoside in crystallization drops produced the highest quality crystals. Small crystals diffract to 1.9 Å and belong to space group *C*2, with unit-cell parameters a = 64.4, b = 91.1, c = 47.9 Å, β = 96.3°. The crystals have a Matthews coefficient of 1.9 $Å^3$ Da^{-1}, which is consistent with one molecule per asymmetric unit and a solvent content of 35%. Data-collection statistics are included in Table 1.

In spite of slight non-isomorphism, scaling between a platinum-derivatized and a native data set revealed a single 6σ platinum site in a difference Patterson map. In order to optimize the signal collected from this single site, a MAD experiment at the platinum L_{III} edge was performed. Potassium tetrachloroplatinate (II) soaks were extremely difficult to reproduce and the best derivatized crystals diffracted only to 3 Å resolution, with a high degree of anisotropy in one direction. Refinement of the highly anisotropic single site with *SHARP* (Fortelle & Bricogne, 1997) revealed additional spurious peaks of positive density around the original platinum site, which were not refineable and therefore did not contribute to phasing. Solvent flattening of the electron-density map produced from the MAD Pt phases allowed the identification of short stretches of secondary-structure elements and the positioning of the platinum site along the protein–solvent boundary, but the experimental map was of insufficient quality to permit an unambiguous tracing of the chain.

In order to avoid problems with non-isomorphism and anisotropy observed for heavy-metal soaks of native crystals, selenomethionine-derivatized VP40(31–326) was prepared and the protein was crystallized under the same conditions as wild-type VP40(31–326). Mass spectrometry confirmed the incorporation of selenium into seven methionine residues [SeMet-VP40(31–326) observed mass, 32 297 Da] compared with native VP40(31–326) (observed mass, 31 967 Da). Preliminary X-ray analysis on a rotating-anode generator coupled to a MAR image-plate detector showed that SeMet-VP40(31–326) crystals diffract to approximately 3.3 Å with the same unit-cell parameters as the native crystals and display better spot shape and lower mosaicity. These crystals will be employed in the structure-solution process by MAD.

The Ebola virus matrix protein does not show any primary- or secondary-structure homology to any known viral matrix protein. Interestingly, secondary-structure prediction analyses suggest a high β-sheet content, in contrast with the highly α-helical structures of matrix proteins from retroviruses (Conte & Matthews, 1998) and influenza virus (M protein; Sha & Luo, 1997). The crystal structure of the Ebola virus matrix protein will give insight into the function of a new matrix protein from a different viral family.

We thank Dr Gordon Leonard of the ESRF, Grenoble, for help with data collection, S. Sciannimanico for technical assistance, Dr K. Ashman of EMBL Heidelberg for N-terminal sequencing and Y. Pétillot for mass spectrometry.

References

Brinkmann, U., Mattes, R. E. & Buckel, P. (1989). *Gene*, **85**, 109–114.
Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* D**50**, 760–763.
Conte, M. R., Klikova, M., Hunter, E., Ruml, T. & Matthews, S. (1997). *EMBO J.* **16**, 5819–5826.
Conte, M. R. & Matthews, S. (1998). *Virology*, **5**, 191–198.
Elliott, L. H., Kiley, M. P. & McCormick, J. B. (1985). *Virology*, **147**, 169–176.
Feldmann, H. & Klenk, H.-D. (1996). *Adv. Virus Res.* **47**, 1–52.
Fortelle, E. de la & Bricogne, G. (1997). *Methods Enzymol.* **276**, 472–494.
Garnier, L., Wills, J. W., Verderame, M. F. & Sudol, M. (1996). *Nature (London)*, **381**, 744–745.
Garoff, H., Hewson, R. & Opstelten, D.-J. E. (1998). *Microbiol. Mol. Biol. Rev.* **62**, 1171–1190.
Gottlinger, H. G., Sodroski, J. G. & Haseltine, W. A. (1989). *Proc. Natl Acad. Sci. USA*, **86**, 5781–5785.
Hill, C. P., Worthylake, D., Bancroft, D. P., Christensen, A. M. & Sundquist, W. I. (1996). *Proc. Natl Acad. Sci. USA*, **93**, 3099–3104.
Jancarik, J. & Kim, S.-H. (1991). *J. Appl. Cryst.* **24**, 409–411.
Kiley, M. P., Cox, N. J., Elliot, L. H., Sanchez, A., DeFries, R., Buchmeier, M. J., Richman, D. D. & McCormick, J. (1988). *J. Gen. Virol.* **69**, 1957–1967.
Macias, M. J., Hyvonem, M., Baraldi, E., Schultz, J., Sudol, M., Saraste, M. & Oschkinat, H. (1996). *Nature (London)*, **382**, 646–649.
Massiah, M., Starich, M. R., Paschall, C. M., Summers, M. F., Christensen, A. M. & Sundquist, W. I. (1994). *J. Mol. Biol.* **244**, 198–223.
Matthews, S., Barlow, P., Boyd, J., Barton, G., Russel, R., Mills, S., Cunningham, M., Meyers, N., Burns, N., Clark, N., Kingsman, S., Kingsman, A. & Campbell, I. (1996). *Nature (London)*, **370**, 666–668.
Otwinowski, Z. (1993). *Proceedings of the CCP4 Study Weekend. Data Collection and Processing*, edited by L. Sawyer, N. Isaacs & S. Bailey, pp. 56–62. Warrington: Daresbury Laboratory.
Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.
Rao, Z., Belyaev, A. S., Fry, E., Roy, P., Jones, I. A. & Stuart, D. I. (1995). *Nature (London)*, **378**, 743–747.
Sha, B. & Luo, M. (1997). *Nature Struct. Biol.* **4**, 239–244.
Xiang, Y., Cameron, C. E., Wills, J. W. & Leis, J. (1996). *J. Virol.* **70**, 5695–5700.