crystallization communications

Crystallization and preliminary X-ray diffraction analysis of protein 14 from *Sulfolobus islandicus* filamentous virus (SIFV)
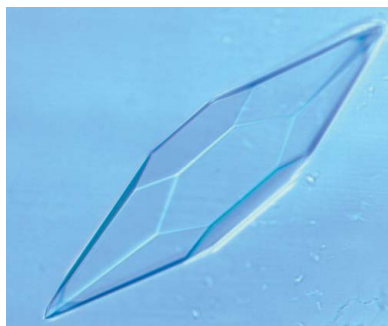
Adeline Goulet,[a] Silvia Spinelli,[a] Valérie Campanacci,[a] Sophie Porciero,[a] Stéphanie Blangy,[a] Roger A. Garrett,[b] Herman van Tilbeurgh,[c] Nicolas Leulliot,[c] Tamara Basta,[d] David Prangishvili[d] and Christian Cambillau[a]*

[a]Architecture et Fonction des Macromolécules Biologiques, CNRS and Universités d'Aix-Marseille I et II, UMR 6098, Case 932, 163 Avenue de Luminy, 13288 Marseille CEDEX 9, France, [b]Danish Archaea Centre, Institute of Molecular Biology, Copenhagen University, Soelvgade 83H, DK1307 Copenhagen K, Denmark, [c]Institut de Biochimie et de Biophysique Moléculaire et Cellulaire (CNRS-UMR 8619), Université Paris 11, Bâtiment 430, 91405 Orsay, France, and [d]Institut Pasteur, Unité de Biologie Moléculaire du Gène chez les Extrêmophiles, 25 Rue du Dr Roux, 75724 Paris CEDEX 15, France

Correspondence e-mail: cambillau@afmb.univ-mrs.fr

A large-scale programme has been embarked upon aiming towards the structural determination of conserved proteins from viruses infecting hyperthermophilic archaea. Here, the crystallization of protein 14 from the archaeal virus SIFV is reported. This protein, which contains 111 residues (MW 13 465 Da), was cloned and expressed in *Escherichia coli* with an N-terminal His$_6$ tag and purified to homogeneity. The tag was subsequently cleaved and the protein was crystallized using PEG 1000 or PEG 4000 as a precipitant. Large crystals were obtained of the native and the selenomethionine-labelled protein using sitting drops of 100–300 nl. Crystals belong to space group $P6_2 22$ or $P6_4 22$, with unit-cell parameters $a = b = 68.1$, $c = 132.4$ Å. Diffraction data were collected to a maximum acceptable resolution of 2.95 and 3.20 Å for the SeMet-labelled and native protein, respectively.

1. Introduction

Linear double-stranded (ds) DNA viruses that infect hyperthermophilic members of the third domain of life, the archaea, are diverse in their morphological and genomic properties and have been assigned to several families, including the *Lipothrixviridae* (Fauquet *et al.*, 2005). Lipothrixviruses are covered by a lipid-containing envelope and their terminal structures involved in adsorption to host cells are distinct and diverse.

In common with members of other viral families that infect hyperthermophilic archaea, lipothrixviruses have an exceptionally high proportion of genes which have neither homologues in public databases nor any predictable functions (Prangishvili *et al.*, 2006). However, members of this family carry genes that are orthologous to those of rudiviridae. Their number is comparable to the number of orthologous genes between members of the same family and suggests a common ancestry of the viral families lipothrixviridae and rudiviridae. For example, the lipothrixvirus SIFV (Arnold *et al.*, 2000) shares the same number of inferred orthologues, nine, with another lipothrixvirus, AFV1, and with the rudivirus SIRV1 (Prangishvili *et al.*, 2006). Identification of these putatively highly conserved functions, none of which is presently clear, will not only contribute to our understanding of viral biology, but will also shed light on evolutionary relationships amongst linear dsDNA viruses.

The power of structural analysis in establishing evolutionary relationships amongst viruses has recently been demonstrated by the structure determination of the major capsid protein of STIV (Khayat *et al.*, 2005). In the absence of sequence similarity to proteins of known function, determination of the three-dimensional structure often provides an efficient approach to deriving strong hypotheses about protein function. With these observations in mind, we decided to embark upon the large-scale structure determination of selected crenarchaeal viral proteins. In light of the small sequence resemblance to other proteins, we considered that these organisms might be enriched in novel folds.

One of the highly conserved proteins of the linear dsDNA viruses is exemplified by protein P14 (SIFV0014) of the lipothrixvirus *Sulfolobus islandicus* filamentous virus (SIFV), which infects the hyperthermophilic and extremely acidophilic archaeaon *Sulfolobus*

(Arnold *et al.*, 2000). This protein is 111 residues in length (MW 13 465 Da) and has a single inferred orthologue in the lipothrixvirus AFV1 and multiple homologues in the rudiviruses SIRV1 and SIRV2 (Prangishvili *et al.*, 2006). In order to obtain further clues to the possible function of this protein, it was cloned, expressed and crystallized. Here, we report these results as well as its biophysical characterization.

2. Cloning, expression and purification

The coding sequence of SIFV ORF14 (SIFV0014) was amplified by PCR from cDNA using two primers containing the *att*B sites of the Gateway recombination system (Invitrogen). The cDNA was cloned in the pDest17 plasmid according to the Gateway cloning technology (Invitrogen; Walhout *et al.*, 2000). A TEV protease cleavage site was inserted between the *att*B1 and the gene of interest. Automated DNA sequencing (GenomeExpress, France) confirmed the identity of the product.

Multiparameter expression screening in *Escherichia coli* was performed using a sparse matrix followed by a dot-blot detection procedure (Vincentelli *et al.*, 2005). The optimized expression condition was observed with *E. coli* strain Rosetta (DE3) transformed with the pLysS plasmid (Novagen) and grown in Superior Broth media (Athena Enzymes) at 301 K. Overnight induction was performed with 0.5 mM isopropyl 1-thio-D-galactopyranoside (IPTG) when OD$_{600\,nm}$ reached 0.6. The cells were harvested by centrifugation at 8000g for 10 min and the pellet was resuspended in 10 ml of buffer *A* [50 mM Tris pH 8, 300 mM NaCl, 10 mM imidazole, 5 mM EDTA, 0.25 mg ml^{-1} lysozyme and protease-cocktail inhibitor (Complete EDTA-free, Roche)] per OD$_{600\,nm}$ unit and per litre of culture and then frozen at 193 K. Cellular resuspensions were thawed at room temperature, supplemented with 0.1 µg ml^{-1} DNase, 20 mM MgSO$_4$, lysed by sonication and centrifuged at 48 500g for 30 min to produce cell-free extract. All purification steps were performed at 293 K. The purification consisted of two steps. Firstly, IMAC (immobilized metal-ion affinity chromatography) was performed on a HiTrap Ni column (Amersham Biosciences) connected to an FPLC system (Amersham Biosciences). The protein was eluted with 50 mM Tris–HCl, 300 mM NaCl, 250 mM imidazole pH 8.0. Size-exclusion chromatography was then performed using a Superdex 200 HiLoad 26/60 column (Amersham Biosciences) in 10 mM Tris pH 8, 300 mM NaCl. SIFV0014 was eluted with a retention volume corresponding to a molecular weight of about 30 kDa, in accord with a dimeric association. Cleavage of the N-terminal hexahistidine tag was carried out in the fractions pooled from the gel filtration, adding TEV protease to a TEV:SIFV0014 ratio of 1:10 at 310 K for 1 h followed by 2 h at room temperature. The cleaved protein only contains an extra glycine at the N-terminus compared with the native sequence. The protein was purified in the flowthrough of a nickel-immobilized metal-ion affinity chromatography and its final buffer was changed by dialysis in 10 mM Tris pH 8, 300 mM NaCl. Cleavage efficiency and protein purity were analyzed by SDS–PAGE. The recombinant protein was concentrated to 11 mg ml^{-1} using a Vivaspin 10 kDa molecular-weight cutoff centrifugal concentrator (Vivascience). The protein concentration was determined from the molar extinction coefficient of the protein at 280 nm (22 920 M^{-1} cm^{-1}) as calculated by the *ExPASy* server (http://www.expasy.org/tools/protparam.html). The concentrated protein was stored frozen at 253 K. Selenomethionine (SeMet) labelled protein was prepared following standard procedures (Doublié, 1997) and in order to minimize the formation of inclusion bodies, induction was performed at 301 K for 3 h and

growth was performed at 290 K overnight. The labelled protein was purified using the same protocol as for the native protein except that, owing to lower solubility, the final buffer of the cleaved protein was changed by dialysis into 10 mM sodium acetate pH 6, 500 mM NaCl. The labelled protein was concentrated to 5.8 mg ml^{-1}. Expression yielded about 35 and 17 mg per litre of culture of soluble native and SeMet-labelled SIFV0014, respectively.

3. Biophysical characterization

The accurate molecular weight of a monomer was determined by MALDI–TOF on a Brucker Autoflex1 mass spectrometer with 2 µl protein (4.1 µM) mixed with an equal volume of DiHydroxy-AcetoPhenol matrix (DHAP) solution and spotted on the spectrometer target plate and then dried at room temperature for 10 min. The mass standards were insulin, ubiquitin I, cytochrome *c* and myoglobin. The mass spectrum shows clearly a peak at 13 461.4 (theoretical mass 13 464.8). Identity of SIFV0014 was checked by Peptide Mass Fingerprinting after in-gel trypsin digestion (Shevchenko *et al.*, 1996). Dynamic light-scattering (DLS) analyses carried out with a Dynapro-MS800 (Protein Solution) confirmed the dimeric association of the cleaved protein in solution according to the apparent molecular weight of 35 kDa deduced from the measurements. Circular-dichroism (CD) spectra were recorded in 10 mM Na/Na$_2$ phosphate buffer pH 8, with a protein concentration of 0.1 mg ml^{-1}, at 293 and 353 K on a J-810 Jasco spectrometer (Easton, MD, USA) in the range 190–260 nm. These spectra show a minimum at 215 nm, typical of a folding mainly composed of β-sheet secondary structure. Their deconvolution (CDNN method; Bohm *et al.*, 1992) yields a prediction of 20% α-helix, 35% β-sheet and 45% random coil, which is in agreement with the secondary-structure prediction carried out using the Protein Prediction Structure Server *PSIPRED* at an expected average accuracy of ∼77% for the three states (Jones, 1999). Moreover, the thermostability of SIFV0014 was confirmed owing to the exact superimposition of the spectra recorded at 293 and 353 K.

4. Crystallization

Crystallization trials were performed using the sitting-drop vapour-diffusion method at 293 K using a nanodrop-dispensing robot
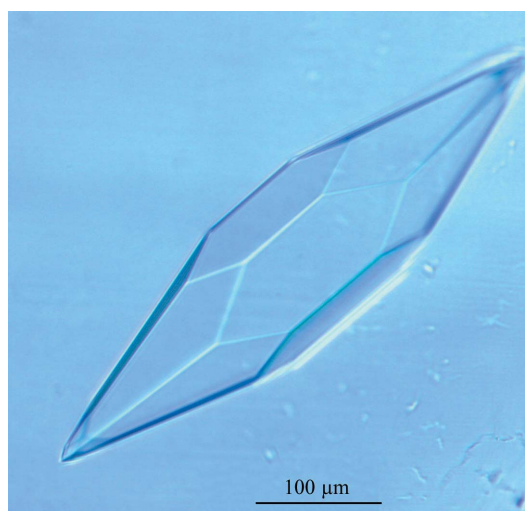
Figure 1
Crystals of SeMet-labelled protein 14 from SIFV.

crystallization communications

Table 1
Data-collection statistics of the SeMet-labelled and native crystals.

Values in parentheses are for the outer resolution shell.

	SeMet	Native
Low-resolution limit (Å)	23.00 (3.11)	30.00 (3.37)
High-resolution limit (Å)	2.95 (2.95)	3.20 (3.20)
R_{sym}†	0.070 (0.386)	0.103 (0.569)
R_{meas} (within I^+/I^-)	0.071 (0.392)	0.108 (0.591)
R_{meas} (all I^+ and I^-)	0.071 (0.392)	0.108 (0.591)
Total no. of observations	222695 (32868)	43554 (6329)
No. of unique observations	4170 (592)	3418 (470)
$\langle I/\sigma(I)\rangle$	74.9 (17.3)	21.0 (6.1)
Completeness (%)	99.3 (99.4)	99.1 (99.3)
Multiplicity	53.4 (55.5)	12.7 (13.5)
Anomalous completeness (%)	100.0 (100.0)	
Anomalous multiplicity	30.9 (30.6)	

† $R_{sym} = \sum_{\mathbf{h}} \sum_{l} |I_{\mathbf{h}l} - \langle I_{\mathbf{h}}\rangle| / \sum_{\mathbf{h}} \sum_{l}\langle I_{\mathbf{h}}\rangle|$, where I_l is the lth observation of reflection **h** and $\langle I_{\mathbf{h}}\rangle$ is the weighted average intensity for all observations l of reflection **h**.

(HoneyBee 961, Cartesian Inc.) in 96-well Greiner crystallization plates (Sulzenbacher *et al.*, 2002). SM1 (Nextal Qiagen), Structure Screens 1 and 2 and Stura Footprint Screen (Molecular Dimensions Ltd) were used for initial screening for the native protein, whereas for the SeMet-labelled protein only the Stura Footprint Screen was tested owing to a limiting quantity of protein. 300, 200 and 100 nl drops of protein solution (5.5–8.2 mg ml^{-1} for the native protein and 2.9–4.35 mg ml^{-1} for the SeMet protein) were mixed with 100 nl mother liquor. Several conditions from the Stura Footprint Screen for the native and SeMet forms and one from the SM1 Screen for the native protein yielded crystalline hits. After optimization, one condition led to hexagonal crystals of SeMet protein [18–26%(*w/v*) PEG 4000, 0.5 *M* MgCl$_2$, 33 m*M* HEPES, 0.2 *M* imidazole malate pH 5.2–5.6] and two conditions led to hexagonal crystals of native protein [first condition, 5–20%(*w/v*) PEG 4000, 0.2 *M* imidazole malate pH 6.8–7.0; second condition, 7–22%(*w/v*) PEG 1000, 0.2 *M* MgCl$_2$, 0.1 *M* sodium cacodylate pH 5.8/6.0/6.2]. The crystals grew in 1 d to final maximum dimensions of 0.5 × 0.2 × 0.2 mm.

5. Preliminary crystallographic analysis

The crystals of SeMet-labelled protein (Fig. 1) were cryoprotected with the same buffer as used for crystallization [25%(*w/v*) PEG 4000, 0.5 *M* MgCl$_2$, 33 m*M* HEPES, 0.2 *M* imidazole malate pH 5.2] for 5 s and flash-frozen in liquid nitrogen at 110 K, in contrast to crystals of native protein, which were flash-frozen without cryoprotectant. Diffraction data for the native and SeMet protein were collected from single crystals at the European Synchrotron Radiation Facility (ESRF, Grenoble, France) beamline ID14-1 using an ADSC Q4R CCD detector and at beamline BM-16 using a MAR CCD 165 detector, respectively. The data were processed using *MOSFLM/ SCALA* (Table 1; Collaborative Computational Project, Number 4, 1994). SIFV0014 crystals diffracted X-rays to a resolution limit of 2.95 Å. Crystals presented an hexagonal symmetry and belonged to space group *P*6$_2$22 or *P*6$_4$22, with unit-cell parameters *a* = *b* = 68.1, *c* = 132.4 Å. The asymmetric unit contains one molecule, as indicated by the V_M value (3.3 Å^3 Da^{-1}, 62% solvent content; Matthews, 1968) and the absence of extra peaks in the self-rotation function. Structure determination using the anomalous signal from the selenomethionine-substituted protein for SAD phasing has been initiated.

References

Arnold, H. P., Zillig, W., Ziese, U., Holz, I., Crosby, M., Utterback, T., Weidmann, J. F., Kristjanson, J. K., Klenk, H. P., Nelson, K. E. & Fraser, C. M. (2000). *Virology*, **267**, 252–266.
Bohm, G., Muhr, R. & Jaenicke, R. (1992). *Protein Eng.* **5**, 191–195.
Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* D**50**, 760–763.
Doublié, S. (1997). *Methods Enzymol.* **276**, 523–530.
Fauquet, C. M., Mayo, M. A., Maniloff, J., Desselberger, U. & Ball, L. A. (2005). Editors. *Virus Taxonomy: Classification and Nomenclature of Viruses.* Amsterdam: Elsevier.
Jones, D. T. (1999). *J. Mol. Biol.* **292**, 195–202.
Khayat, R., Tang, L., Larson, E. T., Lawrence, C. M., Young, M. & Johnson, J. E. (2005). *Proc. Natl Acad. Sci. USA*, **102**, 18944–18949.
Matthews, B. W. (1968). *J. Mol. Biol.* **33**, 491–497.
Prangishvili, D., Garrett, R. A. & Koonin, E. V. (2006). *Virus Res.* **117**, 52–67.
Shevchenko, A., Jensen, O. N., Podtelejnikov, A. V., Sagliocco, F., Wilm, M., Vorm, O., Mortensen, P., Boucherie, H. & Mann, M. (1996). *Proc. Natl Acad. Sci. USA*, **93**, 14440–14445.
Sulzenbacher, G. *et al.* (2002). *Acta Cryst.* D**58**, 2109–2115.
Vincentelli, R., Canaan, S., Offant, J., Cambillau, C. & Bignon, C. (2005). *Anal. Biochem.* **346**, 77–84.
Walhout, A. J., Temple, G. F., Brasch, M. A., Hartley, J. L., Lorson, M. A., van den Heuvel, S. & Vidal, M. (2000). *Methods Enzymol.* **328**, 575–592.