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

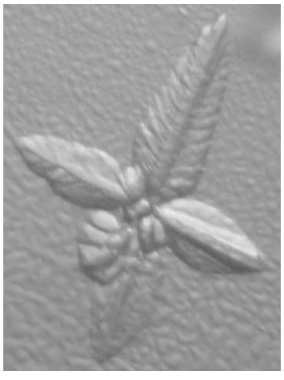
The mimivirus R355 gene product: preliminary crystallographic analysis of a putative ubiquitin-like protein-specific protease

Sandra Jeudy, Audrey Lartigue, Pascal Mansuelle, Yuki Ogata and Chantal Abergel*

Information Génomique et Structurale, CNRS UPR 2589, Institut de Microbiologie de la Méditerranée, Aix-Marseille Université, 163 Avenue de Luminy, Case 934, 13288 Marseille CEDEX 09, France

Correspondence e-mail: chantal.abergel@igs.cnrs-mrs.fr

Received 22 September 2010
Accepted 3 December 2010



The complete genome sequence of the largest known double-stranded DNA virus, mimivirus, reveals the presence of a gene (denoted R355) that potentially encodes a cysteine protease that is expressed late (after 6 h) in the infectious cycle of the virus. In order to verify a sequence-based functional prediction and understand its role during the infectious process, the R355 protein was produced to assay its proteolytic activity and solve its three-dimensional structure. Here, the preliminary crystallographic analysis of the recombinant viral protein is reported. The crystals belonged to the orthorhombic space group $P2_12_12_1$, with a monomer in the asymmetric unit. A MAD data set was used for preliminary phasing using the selenium signal from a selenomethionine-substituted protein crystal.

1. Introduction

Mimivirus, a giant virus packed into a complex particle with a diameter of 750 nm, encodes more than 1000 proteins in a 1.2 Mb genome (Raoult *et al.*, 2004; Legendre *et al.*, 2010). Mimivirus belongs to the nucleocytoplasmic large DNA virus (NCLDV) family of viruses and replicates in the amoeba *Acanthamoeba castellanii*, a protist that is ubiquitous in soil and water. The burst size of mimivirus is several hundred particles per infected cell over a 12 h infectious cycle that is terminated by the lysis of the acanthamoeba cell and the release of neo-synthesized virions into the external medium (for a review, see Claverie & Abergel, 2009). Analysis of the mimivirus transcriptome throughout its entire infectious cycle revealed a precisely timed pattern and a large dynamic range of mimivirus gene expression (Legendre *et al.*, 2010). Mimivirus genes were found to fall into three global classes of expression: early (before 3 h), intermediary (from 3 to 6 h) and late (beyond 6 h). The R355 gene encoding a putative protease belongs to the third class, which corresponds to proteins that are involved in the synthesis of structural components of the capsid and also to enzymatic proteins that will be loaded into the virus particle to play an early role in the infection process (Claverie & Abergel, 2009). The proteome analysis of purified virions confirmed the presence of the R355 gene product in the neo-synthesized virions (Renesto *et al.*, 2006; Claverie *et al.*, 2009), thus confirming its role in the early stages of the mimivirus replication cycle, such as reprogramming the host cell metabolism. Interestingly, the R355 product sequence exhibits homology to the peptidase C48 superfamily, which contain the catalytic triad Cys-His-Asn. This superfamily includes the ubiquitin-like protein-specific protease family (Ulp1). The closest homologue of the R355 protein is the pS273R cysteine protease that is present in African swine fever virus, with which it shares 22% identical residues over its entire length. The pS273R protein has been characterized as an essential component of virus maturation through the cleavage of the two polyprotein precursors pp220 and pp62 that are necessary to produce the six major structural proteins constituting the infectious particle, a typical late function (Andrés *et al.*, 2001; Alejo *et al.*, 2003; Suárez *et al.*, 2010). The R355 protein sequence also shares about 30% identity over 80 amino acids with the Ulp1 catalytic domain of the *Schizosaccharomyces pombe* deneddylating enzyme (Zhou & Watts, 2005), a homology that is more in line with a putative role in switching regulatory circuits in the host cell. Such a low level

of similarity does not permit us to discriminate between a general cysteine protease activity, an enzyme required for the maturation of structural components of the virus particle or a protease involved in metabolic signalling through the cleavage of ubiquitin-conjugated proteins. The enzymatic and structural study of the R355 protein should provide some insight into its molecular specificity and shed light on its function during the infection process, such as the regulation of host cellular processes or virion maturation. In this work, we report the cloning, expression, crystallization and preliminary phasing of the 302-amino-acid protein fused with an N-terminal histidine tag.

2. Results

2.1. Expression of the R355 gene product

The gene encoding the mimivirus R355 protein was amplified from mimivirus genomic DNA and directional cloning was performed using the Gateway system (Invitrogen) as described previously (Abergel *et al.*, 2003). A forward primer containing the 3′-specific nucleotides of the R355 gene and including an *att*B1 extension as well as a reverse primer specific for the 5′ sequence including the STOP codon of the R355 gene and an *att*B2 extension was used for amplification. The PCR product was inserted by homologous recombination in the pDEST17 expression plasmid in phase with an N-terminal His_6 tag under the control of a T7 promoter. Expression screening was performed using four *Escherichia coli* strains, three temperatures, three culture media and three induction protocols. The best condition for soluble expression of the recombinant R355 protein corresponded to *E. coli* C41 transfected cells (Avidis) grown at 290 K in SB medium (Superior Broth, AtthenaES). The induction was performed using 0.5 m*M* IPTG when the A_{600} reached 0.6–0.8. The culture was scaled up to 200 ml and the pellet was resuspended in 50 m*M* sodium phosphate, 300 m*M* NaCl buffer pH 9.0 (buffer *A*) containing 0.1% Triton X-100 and 5% glycerol and total proteins were extracted by sonication.

The selenomethionine-substituted protein was produced using an appropriate protocol to inhibit methionine synthesis in the presence of selenomethionine and M9 minimal medium (Doublié, 1997).

SYYHHHHHHLESTSLYKKAGLNICGPQRYDKENNTCFNVDQLVEMAKAYN
RYLSKTKLNPSRNYHFGDADLINIKSDKKYLLKQFKDRFGKICGSDEICL
THQAFMGELVGEMKDDILFGTFRSEGPSKSTEWLSTIDINQIMVPYENIY
PNFKFIGAVPADCDQVSVCPLYNINYDKLMDEGINYIATIFNHDRYGQPG
SHWVAMFVDINNGKLYYCDSNGKEPTKYIENSIEKFAQFYKRKTGNDIIY
KYNKNSYQKDGSECGVYSCNFIIRMLSGEPFDNIVSNSLSFQEINSCRNV
YFRNQPSKFKPHKLCDPTNSGK

Figure 1
R355 sequence as identified by N-terminal sequencing and mass spectrometry.

2.2. Purification

The cleared lysate was applied onto a 5 ml HisTrap HP Column (GE Healthcare) charged with Ni^{2+} and equilibrated with buffer *A* on an AKTÄexplorer 10S FPLC system (GE Healthcare). The column was washed with ten column volumes of buffer *A*, ten column volumes of buffer *A* containing 25 m*M* imidazole and five column volumes of buffer *A* containing 50 m*M* imidazole at a flow rate of 1 ml min^{-1}. Elution was performed with a linear gradient over seven column volumes from 50 to 500 m*M* imidazole. Microdialysis experiments were performed on the fractions corresponding to the elution of R355 with 150–200 m*M* imidazole to identify the best buffer for protein concentration. The peak was then run on a desalting column (HiPrep 26/10 Desalting, GE Healthcare) in 10 m*M* sodium cacodylate pH 5.0. The recombinant R355 corresponds to native protein in which the N-terminal methionine is replaced by an extended histidine tag inherent to the use of the Gateway system (21 residues in length: SYYHHHHHHLESTSLYKKAGL). The purified protein was characterized by mass spectroscopy and N-terminal Edman sequencing (Fig. 1). Isoelectric focusing using pH 3–10 gradient pre-cast gels (Novex) revealed a band around pI 8.2. Dynamic light scattering using a DynaPro system (Protein Solutions) indicated that the protein solution was monodisperse and monomeric as confirmed on FPLC by gel filtration of the purified protein on a Superdex S200 column (Hiload 16/60 Superdex 200, GE Healthcare).

2.3. Crystallization

The recombinant R355 protein was initially concentrated to 10 mg ml^{-1} in 10 m*M* sodium cacodylate pH 5.0 using a centrifugal filter device (Vivaspin 10K, Vivascience). The recombinant protein was tested at 293 K against 588 different conditions corresponding to

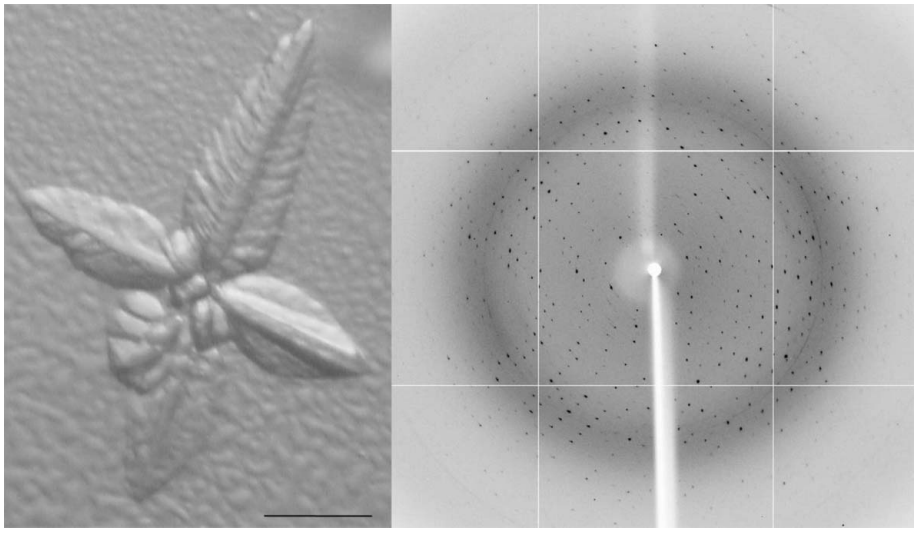


Figure 2
Picture and diffraction image of a R355 crystal. The scale bar corresponds to 0.05 mm.

Table 1
X-ray data-collection statistics (ESRF).

Values in parentheses are for the highest resolution shell.

Method	MAD				Native
Beamline	BM30A				ID29
Wavelength (Å)	0.979207	0.979423	0.968634		0.97899
Space group	$P2_12_12_1$				$P2_12_12_1$
Unit-cell parameters at 100 K (Å, °)	a = 53.93, b = 78.56, c = 79.91, α = 90, β = 90, γ = 90				a = 53.65, b = 78, c = 79.27, α = 90, β = 90, γ = 90
Resolution range (Å)	44.7–2.3 (2.39–2.3)	44.7–2.3 (2.39–2.3)	44.7–2.3 (2.39–2.3)	44.7–2.0 (2.11–2.0)	55.55–1.55 (1.63–1.55)
Observations	54428 (3782)	54508 (3818)	55734 (4480)	87010 (12546)	343773 (50469)
Unique reflections	13392 (1184)	13411 (1202)	13593 (1312)	20962 (3061)	48953 (7069)
Multiplicity	4.0 (3.2)	4.0 (3.2)	4.1 (3.5)	4.2 (4.2)	7.0 (7.1)
Completeness (%)	99.3 (99.3)	99.3 (99.3)	99.6 (99.6)	99.9 (99.9)	99.9 (99.9)
$\langle I/\sigma(I)\rangle$†	5.6 (3.6)	5.9 (3.5)	5.6 (3.6)	4.6 (2.6)	8.3 (2.2)
$R_{\rm merge}$‡	7.6 (18.3)	7.5 (19.1)	8.2 (18.4)	8.8 (26.0)	5.8 (34.7)

† $\langle I/\sigma(I)\rangle$ is the mean signal-to-noise ratio, where I is the integrated intensity of a measured reflection and σ is the estimated error in the measurement. ‡ $R_{\rm merge}$ = $\sum_{hkl}\sum_i |I_i(hkl) - \langle I(hkl)\rangle| / \sum_{hkl}\sum_i I_i(hkl)$, where $I_i(hkl)$ is the integrated intensity of the ith observation of reflection hkl and $\langle I(hkl)\rangle$ is the mean recorded intensity of reflection hkl over multiple observations.

commercially available solution sets (Crystal Screen from Hampton Research and Wizard screens from Emerald BioStructures) and conditions designed in-house using the *SAmBA* software (Audic *et al.*, 1997). Screening for crystallization conditions was performed on 3 × 96-well crystallization plates (Greiner) loaded using an eight-needle dispensing robot (Tecan WS 100/8 workstation modified for our needs), mixing 0.5 µl protein solution with 0.5 µl reservoir solution in a sitting drop (Abergel *et al.*, 2003).

Crystals appeared after few days in the 'in-house' screen in 0.1 *M* sodium cacodylate pH 6.5, 15–20%(*w*/*v*) PEG 4000, 0.1 *M* ammonium sulfate and 15–20%(*v*/*v*) MPD. We refined these conditions at 293 K by hanging-drop vapour diffusion in 24-well culture plates (Greiner). Each hanging drop was prepared by mixing 2 µl 20 mg ml^{-1} R355 solution with 0.5 µl reservoir solution. The hanging drop on the cover glass was vapour-equilibrated against 1 ml reservoir solution with an increasing PEG concentration from 4% to 12% PEG 4000, 10% MPD, 0.1 *M* ammonium sulfate, 0.1 *M* sodium cacodylate pH 6.5 in each well of the tissue-culture plate. For the selenomethionine-substituted protein, we had to use a microseeding protocol to obtain crystals that were usable for structure determination. Crystals from a previous crystal-growth experiment were ground in 50 µl reservoir solution. Serial dilution of this solution was performed to 1/16 and 0.1 µl of this diluted solution was added to each crystallization condition.

2.4. Data collection and processing

Crystals (Fig. 2) were collected in a 0.2 × 0.2 mm loop (Hampton Research), flash-cooled to 105 K in a cold nitrogen-gas stream and subjected to X-ray diffraction. In order to use the MAD method (Hendrickson *et al.*, 1990), a three-wavelength data set was collected on the BM30 beamline at the European Synchrotron Radiation Facility (ESRF; Grenoble, France) from a selenomethionine-substituted R355 crystal flash-frozen at 105 K. Data collection was performed with an oscillation angle of 1° and a crystal-to-detector distance of 360 mm and 105 images were collected at each wavelength. A fourth data set was also collected from the same crystal to 2.0 Å resolution using a 278 mm crystal-to-detector distance for the remote data set (Table 1). The crystal belonged to the orthorhombic space group $P2_12_12_1$, with unit-cell parameters a = 53.9, b = 78.6, c = 79.9 Å. The packing density for one monomer of R355 (37 293 Da) in the asymmetric unit of the crystal is 2.2 Å^3 Da^{-1}, indicating an approximate solvent content of 47.7% (Matthews, 1968). Data statistics are presented in Table 1.

A high-resolution data set for the native protein was also collected on the ID29 beamline at the ESRF using a wavelength of 0.97899 Å; 360 images were collected with an oscillation angle of 0.5° and a crystal-to-detector distance of 197.5 mm. The crystal diffracted to 1.55 Å resolution and had unit-cell parameters a = 53.6, b = 78, c = 79.3 Å. The data statistics are presented in Table 1.

MOSFLM (Leslie, 1992) and *SCALA* (Evans, 1997, 2006) from the *CCP*4 package (Collaborative Computational Project, Number 4, 1994) were used for the processing, scaling and reduction of the data sets.

Phase determination was performed with the *SOLVE* program (Terwilliger & Berendzen, 1999). Phases were calculated using the three-wavelength data in a 20–2.3 Å common resolution range and a single solution was found with seven sites and a mean figure of merit of 0.49 for all data between 20 and 2.3 Å resolution. The phases obtained were improved by solvent-flattening and histogram-matching techniques as implemented in the *DM* program (Cowtan, 1994) and the electron-density map was used to construct the main chain of the R355 protein structure.

There is no sign of a polyprotein precursor of mimivirus structural proteins in the viral genome. On the other hand, we identified a number of ubiquitin-like proteins such as ubiquitin, SUMO and Nedd8 proteins that were encoded in the *A. castellanii* genome. We thus favour the hypothesis that the R355 protein might display proteolytic activity, targeting specific *A. castellanii* proteins, *e.g.* as a ubiquitin-like protein-specific protease. The analysis of the mimivirus R355 structure as well as the characterization of its enzymatic activity and specificity should provide a better understanding of its molecular and cellular functions and help to discriminate whether the R355 is a true Ulp1-like protease that is potentially involved in the hijacking of the infected host cell or a cysteine protease simply operating at the virion level.

We thank the BM30 and ID29 teams for expert assistance on the ESRF beamlines and Professor Jean-Michel Claverie for reading the manuscript. This work was partially funded by CNRS, ANR grant No. ANR-08-BLAN-0089, IBISA and the Provence–Alpes–Côte d'Azur region.

References

Abergel, C. *et al.* (2003). *J. Struct. Funct. Genomics*, **4**, 141–157.
Alejo, A., Andrés, G. & Salas, M. L. (2003). *J. Virol.* **77**, 5571–5577.

Andrés, G., Alejo, A., Simón-Mateo, C. & Salas, M. L. (2001). *J. Biol. Chem.* **276**, 780–787.
Audic, S., Lopez, F., Claverie, J. M., Poirot, O. & Abergel, C. (1997). *Proteins*, **29**, 252–257.
Claverie, J. M. & Abergel, C. (2009). *Annu. Rev. Genet.* **43**, 49–66.
Claverie, J. M., Abergel, C. & Ogata, H. (2009). *Curr. Top. Microbiol. Immunol.* **328**, 89–121.
Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* D**50**, 760–763.
Cowtan, K. (1994). *Jnt CCP4/ESF–EACBM Newsl. Protein Crystallogr.* **31**, 34–38.
Doublié, S. (1997). *Methods Enzymol.* **276**, 523–530.
Evans, P. R. (1997). *Proceedings of the CCP4 Study Weekend. Recent Advances In Phasing*, edited by K. S. Wilson, G. Davies, A. W. Ashton & S. Bailey, pp. 97–102. Warrington: Daresbury Laboratory.
Evans, P. (2006). *Acta Cryst.* D**62**, 72–82.
Hendrickson, W. A., Horton, J. R. & LeMaster, D. M. (1990). *EMBO J.* **9**, 1665–1672.
Legendre, M., Audic, S., Poirot, O., Hingamp, P., Seltzer, V., Byrne, D., Lartigue, A., Lescot, M., Bernadac, A., Poulain, J., Abergel, C. & Claverie, J. M. (2010). *Genome Res.* **20**, 664–674.
Leslie, A. G. W. (1992). *Jnt CCP4/ESF–EACBM Newsl. Protein Crystallogr.* **26**.
Matthews, B. W. (1968). *J. Mol. Biol.* **33**, 491–497.
Raoult, D., Audic, S., Robert, C., Abergel, C., Renesto, P., Ogata, H., La Scola, B., Suzan, M. & Claverie, J. M. (2004). *Science*, **306**, 1344–1350.
Renesto, P., Abergel, C., Decloquement, P., Moinier, D., Azza, S., Ogata, H., Fourquet, P., Gorvel, J. P. & Claverie, J. M. (2006). *J. Virol.* **80**, 11678–11685.
Suárez, C., Salas, M. L. & Rodríguez, J. M. (2010). *J. Virol.* **84**, 176–187.
Terwilliger, T. C. & Berendzen, J. (1999). *Acta Cryst.* D**55**, 849–861.
Zhou, L. & Watts, F. Z. (2005). *Biochem. J.* **389**, 307–314.