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

# Preliminary crystallographic analysis of a possible transcription factor encoded by the mimivirus L544 gene

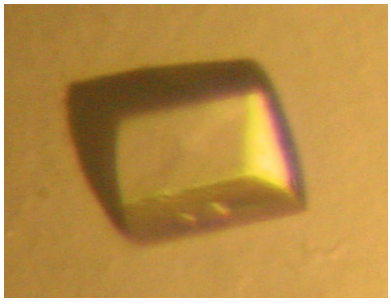
**Alexandre Ciaccafava,‡ Audrey Lartigue,‡ Pascal Mansuelle, Sandra Jeudy and Chantal Abergel***

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

‡ These authors contributed equally to this work.

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

Received 28 April 2011
Accepted 30 May 2011





Mimivirus is the prototype of a new family (the *Mimiviridae*) of nucleocytoplasmic large DNA viruses (NCLDVs), which already include the *Poxviridae*, *Iridoviridae*, *Phycodnaviridae* and *Asfarviridae*. Mimivirus specifically replicates in cells from the genus *Acanthamoeba*. Proteomic analysis of purified mimivirus particles revealed the presence of many subunits of the DNA-directed RNA polymerase II complex. A fully functional pre-transcriptional complex appears to be loaded in the virions, allowing mimivirus to initiate transcription within the host cytoplasm immediately upon infection independently of the host nuclear apparatus. To fully understand this process, a systematic study of mimivirus proteins that are predicted (by bioinformatics) or suspected (by proteomic analysis) to be involved in transcription was initiated by cloning and expressing them in *Escherichia coli* in order to determine their three-dimensional structures. Here, preliminary crystallographic analysis of the recombinant L544 protein is reported. The crystals belonged to the orthorhombic space group $C222_1$ with one monomer per asymmetric unit. A MAD data set was used for preliminary phasing using the selenium signal present in a selenomethionine-substituted protein crystal.

## 1. Introduction

Mimivirus, the largest and most complex virus characterized to date, exhibits a giant hairy particle with a diameter of 750 nm packing a 1.2 Mb genome that encodes 979 proteins, six tRNAs and 33 non-coding RNAs (Raoult *et al.*, 2004; Legendre *et al.*, 2011). Despite these exceptional figures, mimivirus nevertheless shares numerous features (and a probable common ancestor) with previously described large double-stranded DNA viruses that infect a variety of eukaryotic hosts such as vertebrates (*Poxviridae*, *Asfaviridae*), invertebrates (*Iridoviridae*), algae (*Phycodnaviridae*) and protists (*Mimiviridae*). These different viruses are collectively referred to as nucleocytoplasmic large DNA viruses (NCLDVs; Iyer *et al.*, 2006), as their replication cycle was initially thought to consist of two phases: a nuclear phase, during which DNA transcription and replication take place in the host cell nucleus, and a cytoplasmic phase, in which protein translation, capsid assembly and the final virion maturation are performed. Interestingly, and despite their probable common evolutionary origin, it is now becoming clear that *Iridoviridae*, *Asfaviridae* and *Phycodnaviridae* have a nuclear and a cytoplasmic phase, while *Poxviridae* and *Mimiviridae* exclusively replicate in their host's cytoplasm and do not go through a nuclear phase. Consequently, their genomes must encode a complete transcriptional apparatus, which in the case of mimivirus and the vaccinia virus has been shown to be loaded in the virions (Renesto *et al.*, 2006; Van Vliet *et al.*, 2009), suggesting that viral transcription can be initiated immediately upon infection (Claverie *et al.*, 2009; Mutsafi *et al.*, 2010). We studied the mimivirus replication cycle in the amoeba *Acanthamoeba castellanii*, a protist that is ubiquitous in soil and water. Analysis of the transcriptome of infected *A. castellanii* cells through the mimivirus infectious cycle highlighted three main temporal classes of gene expression (Legendre *et al.*, 2011). The early expressed genes mostly correspond to unknown functions or genes encoding proteins involved in translation, while the intermediate class correspond to genes encoding proteins involved in DNA replication and

the late class correspond to genes encoding either structural proteins of the capsid or enzymes loaded in the virion to be used immediately upon infection such as the transcription machinery. The L544 protein belongs to this third class; although it has no clear homology to any known subunits of the RNA polymerase II complex or other transcription factors, we set out to determine its three-dimensional structure in order to obtain insight into its role in the early stage of infection.

In this work, we report the cloning, expression, crystallization and preliminary phasing of the 387-amino-acid L544 protein fused to a C-terminal histidine tag.

## 2. Methods and materials

### 2.1. Expression of the L544 gene product

The gene encoding the mimivirus L544 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 (L544ct1F; GGGGACAAGTTTGTACAAAAAAGCAGGCTTCGAAGGAGATAGAACCATGAGAATGTTGATTTTCACATATAAGTTAGAG) containing the 5′ *attB*1 extension followed by specific nucleotides of the 5′ L544 gene and a reverse primer (L544ct1R; GGGGACCACTTTGTACAAGAAAGCTGGGTTATATATCAATTTTTTTTTGGTTAAAAAATTGG) specific for the 3′ sequence excluding the STOP codon of the L544 gene followed by the *attB*2 extension were used for amplification. The STOP codon is provided by the pDEST42 plasmid.

The PCR product was inserted by homologous recombination in the pDEST42 expression plasmid for expression in phase with a C-terminal $His_6$ tag under the control of a T7 promoter. Expression screening was performed using four *Escherichia coli* strains [Rosetta (DE3) and C41 (DE3) from Avidis and BL21 (DE3) and RIPL (DE3) from Stratagene], three temperatures (290, 298 and 310 K), three culture media [2×YT from Difco and Superior Broth (SB) and Turbo Broth (TB) from Athena Enzyme Systems] and three induction protocols (0.05 m*M* IPTG, 2 m*M* IPTG and autoinduction). The best-scoring condition for soluble expression of recombinant L544 protein corresponds to RIPL transfected cells grown at 290 K in TB medium. Induction was performed using 0.05 m*M* IPTG when the $A_{600}$ reached 0.75. The culture was scaled up to 1 l; the pellet was resuspended in 50 m*M* Tris, 300 m*M* NaCl buffer pH 7.5 (buffer *A*) containing 0.1% Triton X-100, 5% glycerol, 0.01%(*w*/*v*) DNAse, 0.01%(*w*/*v*) lysozyme and two antiprotease tablets (Roche Diagnostics) and total proteins were extracted by sonication.

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

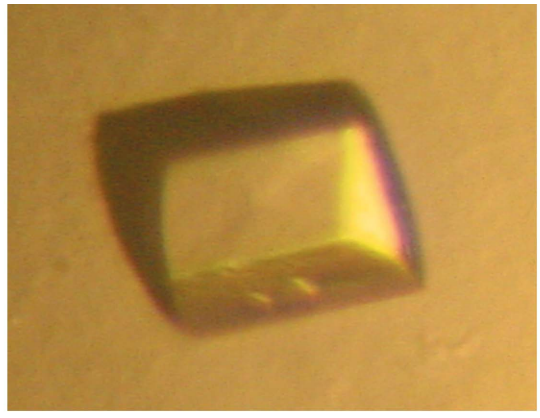
### 2.2. Purification

The cleared lysate was applied onto a 5 ml HisTrap FF Column (GE Healthcare) charged with $Ni^{2+}$ and equilibrated with buffer *A* on an AKTÄexplorer 10S FPLC system (GE Healthcare) at a flow rate of 1.5 ml $min^{-1}$. The column was washed with nine column volumes of buffer *A*, nine column volumes of buffer *A* containing 25 m*M* imidazole and nine column volumes of buffer *A* containing 50 m*M* imidazole at a flow rate of 5 ml $min^{-1}$. Elution was performed with a linear gradient over 20 column volumes from 50 to 500 m*M* imidazole.

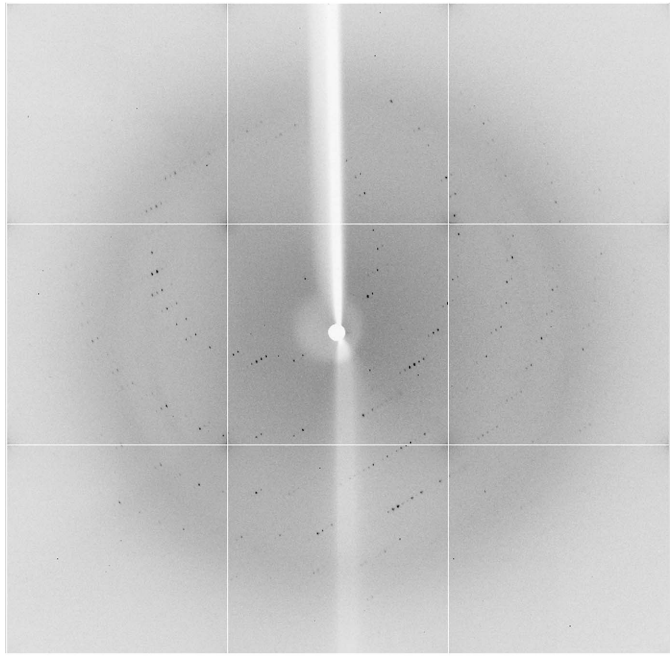
Microdialysis experiments were performed on fractions corresponding to the elution of L544 to identify the best buffer for protein concentration. The peak was then run on a desalting column (HiPrep 26/10 Desalting, GE Healthcare) in 10 m*M* CHES pH 9 for native L544 protein and 10 m*M* CAPS, 150 m*M* NaCl pH 10.5 for selenomethionine-substituted L544. The recombinant L544 corresponds to the native protein extended by a C-terminal histidine tag inherent to the use of the Gateway system (38-residue tag: NPAFLYKVVINSKLEGKPIPNPLLGLDSTRTGHHHHHH). The purified protein characterized by mass spectroscopy on a MALDI–TOF Microflex II (Bruker) had a mass of 50 140 Da (theoretical mass: 50 131 Da). Interestingly, the recombinant protein was found to copurify with *E. coli* nucleic acids. We used the Qubit system (Invitrogen) to characterize them and estimated the protein/DNA/RNA content of the purified fraction. 2 g of L544 protein contained 150 mg RNA and 50 mg DNA. After proteinase K digestion of the L544 protein, the nucleic acids appeared on an agarose gel as a smear corresponding to nucleotide lengths of 300–600 nucleotides.

### 2.3. Crystallization

The recombinant L544 protein was initially concentrated to 10 mg $ml^{-1}$ in 10 m*M* CHES buffer pH 9.0 using a centrifugal filter



(*a*)



(*b*)

**Figure 1**
(*a*) Picture and (*b*) diffraction image of an L544 crystal.

**Table 1**
X-ray data-collection statistics.

Values in parentheses are for the highest resolution shell.

| | | | |
|---|---|---|---|
| Beamline | BM30A, ESRF | | |
| Method | MAD | | |
| Wavelength (Å) | 0.9800 | 0.9801 | 0.9611 |
| Space group | $C222_1$ | | |
| Unit-cell parameters at 100 K (Å, °) | $a = 59.63$, $b = 90.78$, $c = 152.36$, $\alpha = 90$, $\beta = 90$, $\gamma = 90$ | | |
| Resolution range (Å) | 76.69–2.60 (2.74–2.60) | 76.69–2.74 (2.91–2.74) | 76.69–2.91 (3.11–2.91) |
| Observations | 38045 (5265) | 33052 (5169) | 28213 (5003) |
| Unique reflections | 11931 (1668) | 10295 (1612) | 8749 (1555) |
| Multiplicity | 3.4 (3.4) | 3.4 (3.4) | 3.4 (3.5) |
| Completeness (%) | 100 (100) | 100 (100) | 100 (100) |
| $\langle I/\sigma(I)\rangle$† | 4.1 (1.4) | 3.3 (1.1) | 1.7 (1) |
| $R_{\rm merge}$‡ (%) | 6.5 (53) | 10.1 (63) | 14 (51) |

† $\langle I/\sigma(I)\rangle$ is the mean signal-to-noise ratio, where $I$ is the integrated intensity of a measured reflection and $\sigma$ 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 intensity of the $i$th observation of reflection $hkl$ and $\langle I(hkl)\rangle$ is the mean recorded intensity of reflection $hkl$ over multiple recordings.

device (Vivaspin 10K, Vivascience). The recombinant protein was tested at 293 K against 588 different conditions corresponding to commercially available screens (Crystal Screen from Hampton Research and Wizards from Emerald BioStructures) and conditions designed in-house using the *SAmBA* software (Audic *et al.*, 1997). Screening for crystallization conditions was performed in three 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).

Crystalline precipitates appeared in screen conditions containing PEG between pH 6 and 8 in the presence of divalent cations such as calcium and magnesium (magnesium chloride or acetate and calcium acetate). We refined these conditions at 293 K by hanging-drop vapour diffusion using 24-well culture plates (Greiner) with various divalent cations. Each hanging drop was prepared by mixing 1 µl 10 mg ml$^{-1}$ L544 solution with 1 µl reservoir solution. The hanging drop on the cover glass was vapour-equilibrated against 1 ml reservoir solution in each well of the tissue-culture plate. We used increasing PEG concentrations from 4 to 12% PEG 4000, 0.1 *M* sodium cacodylate between pH 6 and 7 and between 0.05 and 0.2 *M* $MgCl_2$, $CaCl_2$ or $MnCl_2$. The best crystals appeared after one month when the salt was 0.1 *M* $MnCl_2$ or 0.2 *M* $MgCl_2$ and reached typical dimensions of 0.2 × 0.15 × 0.15 mm (Fig. 1). Data were collected from a crystal grown in 0.1 *M* sodium cacodylate pH 6, 10% PEG 4000 and 0.2 *M* $MgCl_2$.

### 2.4. Data collection and processing

A previously described evaporation protocol (Abergel, 2004) was used to improve the crystal diffraction from 8 to 2.6 Å resolution (Fig. 1). Crystals were soaked for 15–30 min in 10 µl reservoir solution containing 10% ethylene glycol as a cryoprotectant; they were then collected in a Hampton Research 0.2 × 0.2 mm loop, flash-cooled to 105 K in a cold nitrogen-gas stream and subjected to X-ray diffraction. 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 L544 crystal flash-frozen at 105 K. A fluorescence scan was performed to determine the peak and inflection energies: the exact values were 12.6522 and 12.6497 keV for the peak and inflection, respectively. Data collection was performed with an oscillation angle of 0.5° and a crystal-to-detector distance of 395.5 mm; 180 images were collected at each wavelength on an ADSC detector. The crystal belonged to the orthorhombic space group $C222_1$, with unit-cell parameters $a = 59.63$, $b = 90.78$, $c = 152.36$ Å. The packing density for one monomer of L544 (50 131 Da) in the asymmetric unit of the crystal was 2.06 Å$^3$ Da$^{-1}$, indicating an approximate solvent content of 40.17% (Matthews, 1968). Data statistics are presented in Table 1.

*MOSFLM* (Leslie, 1992) and *SCALA* (Evans, 1997, 2005) from the *CCP*4 package (Winn *et al.*, 2011) were used for the processing, scaling and reduction of the data sets.

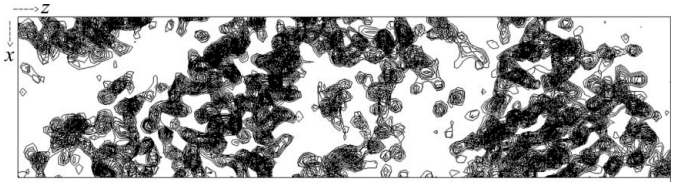



**Figure 2**
Projection of the solvent-flattened electron-density map produced by *autoSHARP*.

## 3. Results and discussion

Phase determination was performed with *autoSHARP* (de La Fortelle & Bricogne, 1997; Bricogne *et al.*, 2003). Phases were calculated using data at three wavelengths in the 25–2.6 Å resolution range and a single solution was found with two Se atoms (for three methionines). The mean figure of merit for this solution was 0.26. The electron-density map is currently being used to construct the main chain of the L544 protein structure (Fig. 2).

The transcriptome study of infected *A. castellanii* cells through the entire viral cycle revealed that mimivirus genes start to be transcribed immediately after infection (Legendre *et al.*, 2011). The mimivirus genes encoding proteins involved in transcription, such as the DNA-dependent RNA polymerase complex subunits, are expressed in the late phase of the viral cycle and proteomic study of the purified virion capsids revealed that the corresponding proteins are loaded into the virion capsids (Renesto *et al.*, 2006; Claverie *et al.*, 2009) and could be active immediately upon infection. Mimivirus anonymous genes that share an expression profile similar to the transcription genes and encode proteins which are also present in the mimivirus particles are thus good candidates for potential transcription functions. The L544 gene is one of them and, interestingly, the behaviour of the over-expressed recombinant L544 protein, which copurifies with *E. coli* RNA, also supports its possible involvement in viral transcription. Analysis of the diffraction data with the program *DIBER* (Chojnowski & Bochtler, 2010) suggested that the crystal contained only protein and no nucleic acids. Since the L544 protein does not show any significant similarity to other sequences in the available databases, its three-dimensional structure could either reveal a new fold or establish its distant relationship to a previously described structural family, thus providing some clues to its molecular function. Specificity studies are also in progress to investigate its affinity for nucleic acids and search for a specific motif that may govern the L544 protein–nucleic acid interaction, thus providing some hints to its cellular function.

We thank the BM30, ID14, ID23 and ID29 teams for expert assistance on the ESRF beamlines, Professor Jean-Michel Claverie for reading the manuscript and Dr Youri Timsit for helpful discussions. We acknowledge the use of the IMM Proteomic Platform

(MaP). 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. (2004). *Acta Cryst.* D**60**, 1413–1416.
Abergel, C. *et al.* (2003). *J. Struct. Funct. Genomics*, **4**, 141–157.
Audic, S., Lopez, F., Claverie, J.-M., Poirot, O. & Abergel, C. (1997). *Proteins*, **29**, 252–257.
Bricogne, G., Vonrhein, C., Flensburg, C., Schiltz, M. & Paciorek, W. (2003). *Acta Cryst.* D**59**, 2023–2030.
Chojnowski, G. & Bochtler, M. (2010). *Acta Cryst.* D**66**, 643–653.
Claverie, J.-M., Abergel, C. & Ogata, H. (2009). *Curr. Top. Microbiol. Immunol.* **328**, 89–121.
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.
Iyer, L. M., Balaji, S., Koonin, E. V. & Aravind, L. (2006). *Virus Res.* **117**, 156–184.
La Fortelle, E. de & Bricogne, G. (1997). *Methods Enzymol.* **276**, 472–494.
Legendre, M., Santini, S., Rico, A., Abergel, C. & Claverie, J.-M. (2011). *Virol. J.* **8**, 99.
Leslie, A. G. W. (1992). *Jnt CCP4/ESF–EACBM Newsl. Protein Crystallogr.* **26**.
Matthews, B. W. (1968). *J. Mol. Biol.* **33**, 491–497.
Mutsafi, Y., Zauberman, N., Sabanay, I. & Minsky, A. (2010). *Proc. Natl Acad. Sci. USA*, **107**, 5978–5982.
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.
Van Vliet, K., Mohamed, M. R., Zhang, L., Villa, N. Y., Werden, S. J., Liu, J. & McFadden, G. (2009). *Microbiol. Mol. Biol. Rev.* **73**, 730–749.
Winn, M. D. *et al.* (2011). *Acta Cryst.* D**67**, 235–242.