Audrey Lartigue, Sandra Jeudy,
Lionel Bertaux and Chantal
Abergel*

Information Génomique et Structurale,
IGS UMR7256, CNRS, Aix-Marseille Université,
IMM, FR3479, 163 Avenue de Luminy – case
934, 13288 Marseille CEDEX 09, France

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

Received 19 September 2012
Accepted 24 November 2012

Preliminary crystallographic analysis of a polyadenylate synthase from *Megavirus*

Megavirus chilensis, a close relative of the *Mimivirus* giant virus, is also the most complex virus sequenced to date, with a 1.26 Mb double-stranded DNA genome encoding 1120 genes. The two viruses share common regulatory elements such as a peculiar palindrome governing the termination/polyadenylation of viral transcripts. They also share a predicted polyadenylate synthase that presents a higher than average percentage of residue conservation. The *Megavirus* enzyme Mg561 was overexpressed in *Escherichia coli*, purified and crystallized. A 2.24 Å resolution MAD data set was recorded from a single crystal on the ID29 beamline at the ESRF.

1. Introduction

Megavirus chilensis was recently isolated from an environmental sample collected on the Pacific Chilean coast near the Tunquén river outfall (Arslan *et al.*, 2011). Its natural host being unknown, we successfully isolated *Megavirus* by co-cultivation with *Acanthamoeba castellanii*. *Megavirus* is an icosahedral virus of roughly 750 nm in diameter that encloses a 1.259 Mb double-stranded DNA genome encoding 1120 genes. It is clearly related to *Mimivirus*, with which it shares ~600 genes with an average of 50% sequence identity at the protein level (Arslan *et al.*, 2011; Legendre *et al.*, 2012).

As in *Mimivirus*, the replicative cycle of *Megavirus* takes place entirely in the host cytoplasm in a virion factory initiated by a single viral particle. Transcription and replication of the viral DNA take place in this transitory organelle (Claverie *et al.*, 2009; Mutsafi *et al.*, 2010).

The regulatory elements governing *Mimivirus* gene expression, such as a highly conserved early and a less stringent late promoter (Suhre *et al.*, 2005; Legendre *et al.*, 2010), are also conserved in *Megavirus*. Another example is the ‘hairpin rule’ governing *Mimivirus* transcript polyadenylation: all *Mimivirus* mRNAs are polyadenylated at a site coinciding exactly with unrelated, but strongly palindromic, genomic sequences (Byrne *et al.*, 2009; Legendre *et al.*, 2010). We verified using bioinformatic and experimental approaches that this ‘hairpin rule’ also applies to *Megavirus* (Arslan *et al.*, 2011; Legendre *et al.*, 2012). The two viruses must thus share the machinery responsible for recognition of the gene termini, transcription arrest and transcript polyadenylation.

In both genomes we identified a well conserved gene (62% protein sequence identity) predicted to encode a polyadenylate synthase which could be involved in the above processes. The first viral polyadenylate synthase was characterized in *Vaccinia virus* (Moure *et al.*, 2006). However, the sequence similarity between this enzyme and the *Megaviridae* homologue is too low (<20% sequence identity) to use the available three-dimensional structure in molecular replacement. Moreover, the ‘hairpin rule’ does not apply in poxviruses, suggesting that this high level of divergence corresponds to significant differences in their polyadenylation processes. We thus decided to conduct an *ab initio* structural and functional study of the *Megavirus* polyadenylate synthase in order to obtain insights into the polyadenylation machinery in *Megaviridae*.

© 2013 International Union of Crystallography
All rights reserved

2. Materials and methods

2.1. Expression of the Mg561 gene product

The gene encoding the *Megavirus* Mg561 protein was amplified from *Megavirus chilensis* genomic DNA and cloned by restriction/ligation.

For PCR amplification, we used a forward primer (MG561_F; CC-CGGATCCTCCGAAAATAGAAATAGAAAATTATC) containing the *Bam*HI site followed by specific nucleotides of the 5' Mg561 gene (bold sequence) and a reverse primer (Mg561_R; TTATGCGGCC-GCTTAATTAGTATTAATATCTGTAGTAGTATT) specific for the 3' sequence including the STOP codon of the Mg561 gene (bold sequence) followed by an *Not*I site.

The PCR product was inserted by restriction/ligation into an in-house-modified pETDuet expression vector (Novagen) allowing the expression of proteins fused with a His₆ tag to facilitate purification. A human rhinovirus 3C protease cleavage site allows tag removal to facilitate the recovery of highly purified proteins free from classical *Escherichia coli* contaminants.

Expression was performed in *E. coli* C41 (Avidis) transfected cells grown at 310 K in Superior Broth (SB; Athena Enzyme Systems). Induction was performed at 294 K using 0.5 mM IPTG when the *A*₆₀₀ reached ~1. The pellet from 1 l overnight culture was resuspended in 40 ml 50 mM Tris-HCl, 300 mM NaCl buffer pH 9 (buffer A) containing 0.1% Triton X-100, 5% glycerol, 0.01% (w/v) DNase, 0.01% (w/v) lysozyme and three antiprotease tablets (Roche Diagnostics), and the 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 (Doublé, 1997).

2.2. Purification

The cleared lysate was applied onto a 5 ml HisTrap FF Column (GE Healthcare) charged with Ni²⁺ and equilibrated with buffer A on an ÄKTAexplorer 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 mM imidazole and nine column volumes of buffer A containing 50 mM imidazole at a flow rate of 5 ml min⁻¹. Elution was performed with a linear gradient from 50 to 500 mM imidazole over 20 column volumes.

The in-house-produced His-tagged human rhinovirus 3C protease was added to the pooled elution fractions and dialysed overnight at 277 K against 25 mM Tris-HCl buffer pH 9, 150 mM NaCl, 1 mM DTT, 0.1 mM EDTA.

A second run of affinity purification was performed to separate the cleaved Mg561 protein from the uncleaved Mg561 protein and the His-tagged protease, which were retained on the HisTrap FF column. The 534-amino-acid Mg561 recombinant protein corresponds to the native enzyme in which the initial methionine is replaced by Gly-Pro-Gly-Ser resulting from the protease cleavage site. After recovery from the flowthrough, Mg561 was concentrated using a centrifugal filter device (Vivaspin 30K, Vivascience) and desalted on a desalting column (HiPrep 26/10 Desalting, GE Healthcare) in 10 mM Tris-HCl pH 9, 100 mM NaCl.

Gel filtration (performed on a Pharmacia Superdex S200 column) indicated that the Mg561 protein was monomeric in solution.

2.3. Crystallization

The recombinant Mg561 protein was concentrated to 5.9 mg ml⁻¹ in the desalting buffer using a centrifugal filter device (Ultracel 30K,

Millipore). The recombinant protein crystallization assays were performed at 293 K against 588 different conditions corresponding to commercially available solution sets (Crystal Screen from Hampton Research and Wizards from Emerald BioStructures) and conditions designed in-house using the *SAMBA* software (Audic *et al.*, 1997). The screening for crystallization conditions was performed on 3 × 96-well crystallization plates (Greiner) loaded using an eight-needle dispensing robot (a 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 in 0.1 M Tris-HCl buffer, 50% (v/v) PEG 200 pH 6.8. We refined these conditions at 293 K by hanging-drop vapour diffusion using 24-well culture plates (Greiner). Each hanging drop was prepared by mixing 1 µl 5.9 mg ml⁻¹ Mg561 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 10% to 50% PEG 200. The best crystals appeared after 1 d for the highest PEG concentration and reached typical dimensions of 0.2 × 0.2 × 0.1 mm (Fig. 1a). Data were collected from a crystal grown by mixing 1 µl selenomethionine-substituted Mg561 protein at 2.8 mg ml⁻¹ with

(a)

(b)

Figure 1
Photograph (a) and diffraction image (b) of an Mg561 crystal. In the inset, the edge of the diffraction circle corresponds to 2.24 Å resolution.

Table 1
X-ray data-collection statistics.

Values in parentheses are for the highest resolution shell.

Beamline	ID29, ESRF		
Method	MAD		
Space group	$P2_12_12_1$		
Unit-cell parameters at 100 K (Å)	$a = 86.7, b = 96.1, c = 153.9$		
Packing density† (Å ³ Da ⁻¹)	2.55		
Wavelength (Å)	0.9793	0.9795	0.9770
Resolution range (Å)	51.3–2.24	51.3–2.24	51.3–2.24
	(2.36–2.24)	(2.36–2.24)	(2.36–2.24)
Observations	210289 (32884)	219559 (34075)	209823 (32763)
Unique reflections	53432 (7970)	54650 (8145)	53198 (7983)
Multiplicity	3.9 (4.1)	4.0 (4.2)	3.9 (4.1)
Completeness (%)	99.0 (99.0)	99.2 (99.2)	99.0 (99.0)
$\langle I/\sigma(I) \rangle$ ‡	6.4 (1.1)	5.0 (0.9)	6.7 (1.5)
$R_{\text{merge}}^{\S}$ (%)	8.1 (62)	8.8 (69)	6.9 (46)

† For two molecules in the asymmetric unit. ‡ $\langle I/\sigma(I) \rangle$ is the mean signal-to-noise ratio, where I is the integrated intensity of a measured reflection and $\sigma(I)$ is the estimated error in the measurement. § $R_{\text{merge}} = \sum_{hkl} \sum_i |I_i(hkl) - \langle I(hkl) \rangle| / \sum_{hkl} \sum_i I_i(hkl)$, where $I_i(hkl)$ is the integrated intensity of reflection hkl having i observations and $\langle I(hkl) \rangle$ is the mean recorded intensity of reflection hkl over multiple recordings.

0.5 µl of a reservoir solution consisting of 0.1 M Tris–HCl buffer, 50% (v/v) PEG 200 pH 6.8.

2.4. Data collection and processing

This crystal was 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; the crystal diffracted to a resolution of 2.24 Å (Fig. 1b). To use the MAD method (Hendrickson *et al.*, 1990), a three-wavelength data set was collected on the ID29 beamline at the European Synchrotron Radiation Facility (ESRF; Grenoble, France). A fluorescence scan was performed to determine the peak and inflection energies; the exact values were 12.6603 and 12.6584 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 374.27 mm, and 240 images were collected for

each wavelength on a PILATUS detector. The crystal belonged to the orthorhombic space group $P2_12_12_1$, with unit-cell parameters $a = 86.7$, $b = 96.1$, $c = 153.9$ Å. Superposition of the two monomers in the asymmetric unit shows that they are related by a twofold axis located at polar angles $\omega = 111^\circ$ and $\varphi = -85^\circ$. Weak peaks are detected at these positions in the self-rotation function (Fig. 2). The packing density for two monomers of Mg561 (62743.8 Da) in the asymmetric unit of the crystal is 2.55 Å³ Da⁻¹, indicating an approximate solvent content of 51.85% (Matthews, 1968). Data statistics are presented in Table 1.

MOSFLM (Leslie, 1992) and *SCALA* (Evans, 1997, 2006) from the *CCP4* package (Winn *et al.*, 2011) were used for processing, scaling and reduction of the data sets.

A self-rotation function analysis of the data in the resolution range 9–4 Å with an integration radius of 25 Å was carried out using the *Self Rotation Function Server* (<http://services.mbi.ucla.edu/selfrot/>; Winn *et al.*, 2011).

3. Results and discussion

Phase determination was performed with *autoSHARP* (de La Fortelle & Bricogne, 1997; Bricogne *et al.*, 2003). Phases were calculated at three wavelengths in the resolution range 51.3–2.24 Å, and a single solution was found with 34 Se atoms (for 36 methionines). The mean figure of merit for this solution is 0.39. The electron-density map is currently being used to construct the main chain of the Mg561 protein structure (Fig. 3).

The transcriptome study of *A. castellanii* cells infected by *Mimivirus* through the entire viral cycle revealed that *Mimivirus* genes start to be transcribed immediately after infection (Legendre *et al.*, 2010). The expression of the *Mimivirus* and *Megavirus* polyadenylate synthase genes is controlled by the highly conserved early promoter and the *Mimivirus* enzyme exhibits continuous expression from the beginning to the end of the infection cycle. Interestingly, the *Mimivirus* polyadenylate synthase protein is also loaded into the virion capsids (Renesto *et al.*, 2006; Claverie *et al.*, 2009), suggesting that this function is required for the processing of the very first viral tran-

Figure 2
Self-rotation Patterson plot. The self-rotation function was calculated using the *Self Rotation Function Server* (<http://services.mbi.ucla.edu/selfrot/>) with default parameters. These sections have $\omega = 0$ or 180° at the centre, $\omega = 90^\circ$ around the edge and φ as marked around the periphery.

Figure 3
Projection of the solvent-flattened electron-density map produced by *autoSHARP*.

scripts, including its own. The three-dimensional structure of Mg561 will be compared with its cellular homologues (bacterial, archaeal and eukaryal polyadenylate synthases) as well as with the structure of the *Vaccinia virus* enzyme. This comparison should provide some clues to the origin of the *Megaviridae* enzymes. Molecular studies will also be performed to address the affinity of the enzyme for folded nucleic acids (RNA hairpins and DNA hairpins) and its ability to polyadenylate transcripts *in vitro*.

We thank the BM30, ID14, ID23, ID29 (ESRF) and PROXIMA (SOLEIL) beamline teams for expert assistance. We also thank Dr Alain Roussel for collecting the MAD data set and Professor Nobert Sträter for helpful suggestions. 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.
- Arslan, D., Legendre, M., Seltzer, V., Abergel, C. & Claverie, J.-M. (2011). *Proc. Natl Acad. Sci. USA*, **108**, 17486–17491.
- Audic, S., Lopez, F., Claverie, J.-M., Poirot, O. & Abergel, C. (1997). *Proteins*, **29**, 252–257.
- Bricogne, G., Vonnrhein, C., Flensburg, C., Schiltz, M. & Paciorek, W. (2003). *Acta Cryst.* **D59**, 2023–2030.
- Byrne, D., Grzela, R., Lartigue, A., Audic, S., Chenivresse, S., Encinas, S., Claverie, J.-M. & Abergel, C. (2009). *Genome Res.* **19**, 1233–1242.
- Claverie, J.-M., Abergel, C. & Ogata, H. (2009). *Curr. Top. Microbiol. Immunol.* **328**, 89–121.
- Doublé, 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.* **D62**, 72–82.
- Hendrickson, W. A., Horton, J. R. & LeMaster, D. M. (1990). *EMBO J.* **9**, 1665–1672.
- La Fortelle, E. de & Bricogne, G. (1997). *Methods Enzymol.* **276**, 472–494.
- Legendre, M., Arslan, D., Abergel, C. & Claverie, J.-M. (2012). *Commun. Integr. Biol.* **5**, 102–106.
- 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). *Int CCP4/ESF-EACBM Newsl. Protein Crystallogr.* **26**.
- Matthews, B. W. (1968). *J. Mol. Biol.* **33**, 491–497.
- Moure, C. M., Bowman, B. R., Gershon, P. D. & Quiocho, F. A. (2006). *Mol. Cell*, **22**, 339–349.
- Mutsaers, Y., Zauberman, N., Sabanay, I. & Minsky, A. (2010). *Proc. Natl Acad. Sci. USA*, **107**, 5978–5982.
- 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.
- Suhre, K., Audic, S. & Claverie, J.-M. (2005). *Proc. Natl Acad. Sci. USA*, **102**, 14689–14693.
- Winn, M. D. *et al.* (2011). *Acta Cryst.* **D67**, 235–242.