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

Purification, crystallization and preliminary X-ray diffraction analysis of the RNA-dependent RNA polymerase from *Thosea asigna* virus

Diego Ferrero,[a] Mònica Buxaderas,[b] José F. Rodriguez[a] and Núria Verdaguer[b]*

[a]Centro Nacional de Biotecnología, CSIC, Calle Darwin No. 3, Madrid, 28049, Spain, and [b]Institut de Biologia Molecular de Barcelona, CSIC, Parc Científic de Barcelona, Baldiri i Reixac 10, Barcelona, 08028, Spain

Correspondence e-mail: nvmcri@ibmb.csic.es

Received 26 July 2012
Accepted 31 August 2012

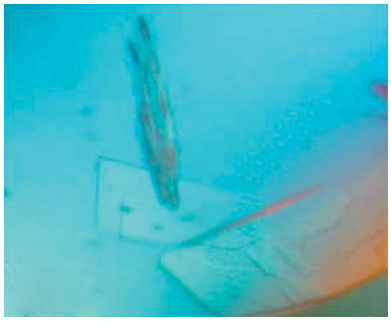
Thosea asigna virus (TaV) is a positive-sense, single-stranded RNA (ssRNA) virus that belongs to the *Permutotetravirus* genera within the recently created Permutotetraviridae family. The genome of TaV consists of an RNA segment of about 5.700 nucleotides with two open reading frames, encoding for the replicase and capsid protein. The particular TaV replicase does not contain N7-methyl transferase and helicase domains but includes a structurally unique RNA-dependent RNA polymerase (RdRp) with a sequence permutation in the domain where the active site is anchored. This architecture is also found in double-stranded RNA viruses of the Birnaviridae family. Here we report the purification and preliminary crystallographic studies TaV RdRp. The enzyme was crystallized by the sitting-drop vapour diffusion method using PEG 8K and lithium sulfate as precipitants. Two different crystal forms were obtained: native RdRp crystallized in space group $P2_12_12$ and diffracts up to 2.1 Å and the RdRp-Lu^{3+} derivative co-crystals belong to the $C222_1$ space group, diffracting to 3.0 Å resolution. The structure of TaV RdRp represents the first structure of a non-canonical RdRp from ssRNA viruses.

1. Introduction

The Tetraviridae family includes small viruses with a positive-sense single-stranded RNA (ssRNA) genome enclosed in an ∼38 nm, non-enveloped $T = 4$ icosahedral capsid. This symmetry of the capsid architecture distinguishes them from others. The existence of profound dissimilarities among members of the Tetraviridae family (*e.g.* genome organization, capsid expression strategy, replicase sequence) led to them being reorganized into three new families (*Virus Taxonomy*, 2011): Alphatetraviridae, Carmotetraviridae and Permutotetraviridae.

Thosea asigna virus (TaV), a member of the Permutotetraviridae family, infects the larvae of a moth (*Setothosea asigna*), the major defoliating pests of oil and coconut palms in Southeast Asia. TaV has a genome of ∼6.5 kb in length and a subgenomic molecule of ∼2.5 kb. The genomic molecule contains two overlapping open reading frames (ORFs): the first one encodes for a 140 kDa polypeptide, containing the RNA-dependent RNA polymerase (RdRp) and the second one codes for the capsid precursor protein, which is also present in the subgenomic molecule (Pringle *et al.*, 2001).

A characteristic of TaV that deserves special attention is the RdRp. All RdRps share a closed 'right-hand' architecture with fingers, palm and thumb subdomains encircling the active site. Sequence analyses identified five ordered sequence motifs (*A*–*B*–*C*–*D*–*E*) within the palm subdomain that are conserved in all virus replicases. However, in TaV RdRp motif *C* is located upstream of motif *A*. This sequence permutation yields a palm fold in which the canonical structural elements show a non-canonical connectivity (*C*–*A*–*B*–*D*–*E*). The permuted architecture of the palm domain is exclusive to two members of the Permutotetraviridae family having a positive-stranded RNA genome, and is present in the double-stranded RNA (dsRNA) viruses of the Birnaviridae family (Gorbalenya *et al.*, 2002). The structures of the RdRps of two different Birnaviruses have



recently been solved (Garriga *et al.*, 2007; Pan *et al.*, 2007; Graham *et al.*, 2011). This similarity suggests an evolutionary relationship between apparently unrelated viruses.

To increase our understanding of virus evolution between these distant related viruses sharing a permuted RdRp palm, we carried out a structural study of a Permutotetravirus replicase. Here we report the preliminary crystallographic studies of TaV RdRp, the first functional non-canonical RdRp from a ssRNA virus described by X-ray crystallography.

2. Materials and methods

2.1. Cloning and expression

A synthetic gene covering the whole ORF1 of TaV (1257 residues, 140 kDa; GenBank accession number AF282930.1), optimized for protein expression in *Tricoplusia ni* cells, was purchased from GenScript USA Inc., including flanking sequences which provide *Bam*HI and *Hin*dIII restriction sites in the 5′ and 3′ ends, respectively; it was cloned in a pUC57 plasmid vector, pUC57/TaV ORF1.

pUC57/TaV RdRp was digested with *Bam*HI and *Hin*dIII; the fragment containing (3776 bp) the RdRp ORF was purified from an agarose gel (QIAquick gel extraction kit, QIAGEN) and ligated to pFastBacHTb (Invitrogen) previously cleaved with the same enzymes. The resulting plasmid, pFastBacHTb TaV ORF1, encodes the viral polymerase fused to an N-terminal 6× histidine tag (hTaV ORF1), also containing the TEV protease recognition site (Fig. 1*a*). A recombinant baculovirus (rBv) expressing the hTaV ORF1 was generated according to the Bac-to-Bac protocol (Invitrogen).

2.2. Protein purification

HighFive cells (Invitrogen) were infected with rBVs at high multiplicity of infection (MOI > 1). Cells were harvested at 72 h post infection, washed twice with phosphate-buffered saline, resuspended in lysis buffer (50 m*M* bis Tris pH 6.8, 500 m*M* NaCl, 0.1% Igepal CA-630) supplemented with protease inhibitors (Complete EDTA-Free; Roche) and maintained on ice for 20 min. Thereafter, extracts were centrifuged 13 000*g* for 20 min at 277 K. Supernatants were collected and subjected to metal affinity chromatography batch purification using a Co^{2+} affinity resin (TALON, Clontech). The resin-bound hTaV ORF1 was washed four times with 5 resin volumes of washing buffer (50 m*M* bis Tris pH 6.8, 500 m*M* NaCl, 0.1% Igepal CA-63010, 5 m*M* imidazol) and eluted with 2 resin volumes of elution buffer (50 m*M* bis Tris pH 6.8, 500 m*M* NaCl, 500 m*M* imidazole). Subsequently, the elution buffer was exchanged for a lysis buffer lacking Igepal using a PD-10 desalting column (GE Healthcare). SDS–PAGE showed that the purified polypeptide was partially cleaved, resulting in a product of ∼75 kDa. The recovered polypeptide was analysed by mass spectrometry to certify the integrity of the RdRp domain, included within the first 672 amino acids of the TaV ORF 1 (Fig. 1*a*). The eluted domain was further purified by size-exclusion chromatography on a Superdex 200 HR 10/300 column exchanging the buffer for 50 m*M* MES [2-(*N*-morpholino)ethanesulfonic acid] pH 6.0, 500 m*M* NaCl, 10% glycerol and 5 m*M* DTT (dithiothreitol) (Fig. 1*b*). Finally, fractions containing the purified RdRp domain of hTaV (Fig. 1*c*) were pooled and concentrated to 10 mg ml^{-1} using VIVASPIN 10K filters (Sartorius).

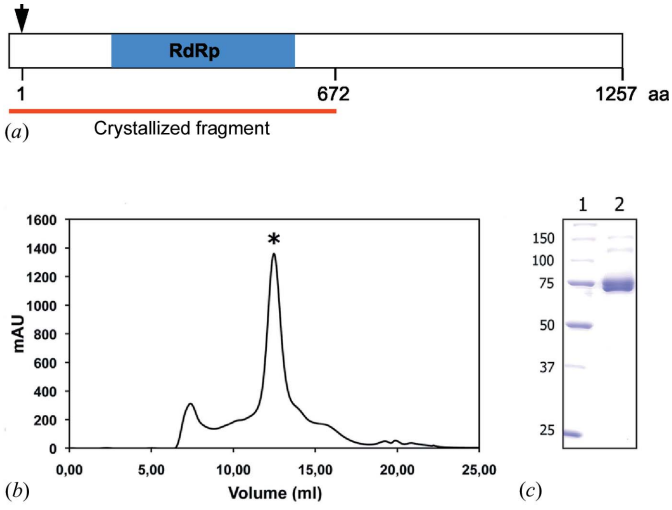


Figure 1
Production and purification of the hTaV RdRP. (*a*) Schematic representation of the recombinant hTaV ORF1 construct. The black arrow indicates the first residue of the native viral protein and the blue box shows the region containing the conserved motifs in all RdRps (Gorbalenya *et al.*, 2002). The red line indicates the crystallized fragment (672 residues of the viral polypeptide plus the fusion tag). (*b*) Size-exclusion chromatogram (Superdex 200 HR 10/300) of the purified polypeptide. According to the pre-calibration of the column, the protein elutes in a volume corresponding to a dimer of the RdRp domain (*). A small amount of aggregates were also visible. mAU on the *y* axis indicates the relative absorbance units at 280 nm. (*c*) SDS–PAGE of the purified protein after Superdex 200 chromatography. Lane 1 shows the molecular-weight marker (kDa) and lane 2 contains the sample peak, labelled with '*' in the chromatogram.

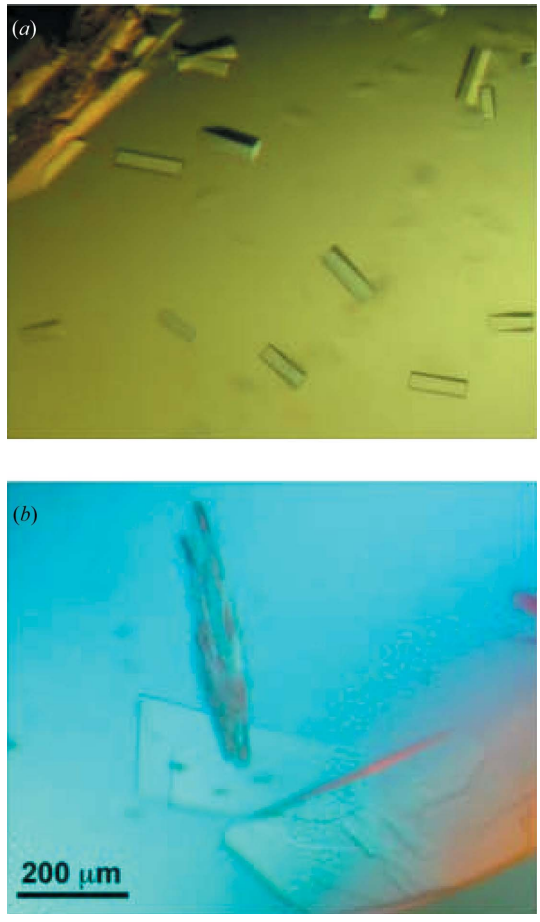


Figure 2
Image of hTaV RdRp crystals obtained from 12% PEG 8K, 750 m*M* Li_2SO_4 in native conditions (*a*) and in the presence of 2 m*M* lutetium(III) acetate (*b*).

Table 1
Data collection and processing statistics for TaV RdRP crystals.

Numbers in parentheses refer to the highest resolution shell.

Parameters	Native crystal	Derivative crystal
λ (Å)	0.98	1.3404
Space group	$P2_12_12$	$C222_1$
Unit-cell parameters (Å)	a = 134.97, b = 150.82, c = 100.61	a = 154.05, b = 224.57, c = 128.29
Resolution range (Å)	30.0–2.15 (2.27–2.15)	63.52–3.0 (3.16–3.00)
Total reflections	503448	285828
Beamline	ID14.4 (ESRF)	ID23.1 (ESRF)
Unique reflections	109020 (15398)	44236 (6247)
Multiplicity	4.6 (3.0)	6.5 (5.9)
Completeness (%)	97.5 (95.6)	98.7 (96.5)
R_{merge} (%)†	12.0 (29.4)	11.6 (37.9)
$I/\sigma(I)$	9.0 (4.1)	13.3 (4.5)
f'		−24.87
f''		34.98

† $R_{merge} = \sum_{hkl}\sum_i |I_i(hkl) - \langle I(hkl)\rangle| / \sum_{hkl}\sum_i I_i(hkl)$, where $I(hkl)$ is the observed intensity and $\langle I(hkl)\rangle$ is the weighted average intensity of multiple observations of symmetry-related reflections.

2.3. Crystallization

Initial crystallization trials of TaV RdRp were performed by the sitting-drop vapour diffusion method at both 277 and 293 K in 96-well Greiner plates using a nanolitre-drop crystallization robot (Cartesian) and the commercial screens JBScreen Classic and Basic HTS (Jena Bioscience). Protein solution (10 mg ml^{-1} in 50 m*M* MES pH 6.0, 500 m*M* NaCl, 10% glycerol, 5 m*M* DTT) droplets of 150 nl were mixed with 150 nl precipitant solution; the volume reservoir was 100 µl. Small needle-like crystals appeared within 2 d in condition G8 (JBScreen Classic), which contained 15% PEG 8K and 500 m*M* Li_2SO_4. After optimization in 24-well Greiner plates, rod-like crystals were obtained (Fig. 2*a*) using the sitting-drop technique by mixing 0.5 µl of protein with an equal volume of the reservoir solution (12% PEG 8K, 750 m*M* Li_2SO_4) equilibrated against a 750 µl reservoir.

A lutetium derivative was prepared by co-crystallization, mixing the protein solution with 2 m*M* lutetium(III) acetate. Plate-shaped crystals appeared in 2 d using the same crystallization conditions (Fig. 2*b*).

Prior to data collection, crystals were harvested in cryoloops (Hampton Research), soaked for 1 min in a reservoir solution and 20%(*v*/*v*) glycerol, and flash-cooled by immersion in liquid nitrogen.

2.4. Data collection and processing

All diffraction data were collected at 100 K from single crystals at the European Synchrotron Radiation Facility (ESRF, Grenoble, France). Native TaV RdRp data were collected up to 2.1 Å resolution on an ADSC Q4R detector on beamline ID14EH4 (λ = 0.98 Å). A total of 180 frames were collected using 0.5° oscillation and 2 s exposure. Derivative data (3.0 Å resolution) were collected on beamline ID23.1 at a wavelength corresponding to the lutetium absorption maximum (λ = 1.3404 Å) and with an oscillation angle of 2° (5 s exposure).

Diffraction images were indexed and integrated using *iMOSFLM* (Leslie, 2006; Battye *et al.*, 2011), and scaled, merged and reduced with *SCALA* (Evans *et al.*, 2006) from the *CCP*4 program suite (Winn *et al.*, 2011). The crystal parameters and diffraction data statistics are given in Table 1.

2.5. Structure determination

The structure of the polymerase was solved using single-wavelength anomalous diffraction (SAD) phasing as implemented in the *Auto-Rickshaw* pipeline (Panjikar *et al.*, 2005).

The input diffraction data were prepared and converted for use in *Auto-Rickshaw* using programs from the *CCP*4 suite (Winn *et al.*, 2011). F_A values were calculated using the program *SHELXC* (Sheldrick *et al.*, 2001). Based on an initial analysis of the data, the maximum resolution for substructure determination and initial phase calculation was set to 3.5 Å and the heavy-atom (Lu^{3+}) positions were found using the program *SHELXD* (Schneider & Sheldrick, 2002).

The correct hand for the substructure was determined using the programs *ABS* (Hao, 2004) and *SHELXE* (Sheldrick, 2008). The occupancy of all substructure atoms was refined using the program *BP3* (Pannu *et al.*, 2003). Density modification, phase extension and NCS averaging were performed using the program *RESOLVE* (Terwilliger, 2000).

A partial model was produced using the program *BUCCANEER* (Cowtan, 2006) and used for phase improvement and model completion by the MR–SAD protocol also available in *Auto-Rickshaw* (Panjikar *et al.*, 2009).

3. Results and discussion

The full-length N-terminal His6-tagged ORF1 product from TaV was overexpressed in insect cells using the baculovirus system. The protein was purified in two chromatographic steps yielding essentially pure material as judged by SDS–PAGE and size-exclusion chromatography (Fig. 1). SDS–PAGE also showed that the C-terminal moiety of the protein was cleaved probably by an unidentified protease from the insect cells, resulting in a polypeptide product of ∼75 kDa (Fig. 1*c*). Mass spectrometry analysis showed that the purified polypeptide contained the first 672 amino-acid residues of the TaV ORF1 product (Fig. 1*a*). This would correspond to the RdRp domain as predicted by computational analyses (Gorbalenya *et al.*, 2002; Zeddam *et al.*, 2010). Interestingly, the size-exclusion chromatography showed that the polymerase domain forms a dimer under the purification condition (Fig. 1*b*).

Native crystals of TaV RdRp belong to the orthorhombic space group $P2_12_12$, with unit-cell parameters a = 134.87, b = 150.82, c = 100.61 Å. The volume of the crystal asymmetric unit, 3.41 Å^3 Da^{-1}, is compatible with the presence of two RdRp molecules with a calculated solvent content of 63.96%. A data set was recorded to 2.1 Å resolution with an R_{merge} of 0.120 and completeness of 97.5% (Table 1). Phase determination by molecular replacement was initially attempted using, as starting models, the available structures of different RdRps, having either canonical (FMDV RdRp model, PDB entry 1u09; Ferrer-Orta *et al.*, 2004) or inverted palms (IBDV RdRp model, PDB entry 2pus; Garriga *et al.*, 2007). This approach was unsuccessful, probably due to the lack of sufficient homology between TaV RdRp and the models used (23 and 38%, for the FMDV and IBDV models, respectively).

A Lu^{3+} derivative was obtained by co-crystallization of the enzyme with lutetium acetate. Crystals belong to the orthorhombic $C222_1$ space group, with unit-cell parameters of a = 154.05, b = 224.57, c = 128.29 Å, with two molecules per asymmetric unit, V_M = 3.70 Å^3 Da^{-1}, corresponding to a solvent content of 66.76%. A data set was recorded to 3.0 Å resolution with an R_{merge} of 0.116 and completeness of 98.7% (Table 1). Using the Lu^{3+} data, the *Auto-Rickshaw* pipeline was adopted (Panjikar *et al.*, 2005), combining SAD and MR methods. The program *SHELXD* (Schneider &

Sheldrick, 2002) identified six heavy-atom sites. The initial phases were improved using *RESOLVE* (Terwilliger, 2000). After this procedure, the electron-density map at 3.5 Å resolution was of sufficient quality to automatically trace a partial model (629 residues), corresponding to the complete palm, most of the fingers and part of the thumb subdomains, for each of the two molecules in the asymmetric unit. This model was employed in an MR–SAD protocol of the *Auto-Rickshaw* pipeline for phase improvement. Complete model building is ongoing. This new structure will provide valuable structural information to understand the function of viral RdRps and the evolutionary relationship between apparently unrelated viruses.

We thank Ms Jenny Colom for help during protein purification, Dr César Santiago for support during the initial crystallization trials and Ms Laia Vives for help in data collection. This study was supported by the Ministerio de Economia y Competitividad (grant Nos. BIO2009-12443 and AGL2011-24758 to JFR and BIO2011-24333 to NV). DF was supported by a fellowship from the international PhD program of the 'La Caixa' foundation. Synchrotron data collection was supported by the European Synchrotron Radiation Facility and the European Union. Crystallization screenings and preliminary X-ray analysis were performed at the Automated Crystallography Platform, Barcelona.

References

Battye, T. G. G., Kontogiannis, L., Johnson, O., Powell, H. R. & Leslie, A. G. W. (2011). *Acta Cryst.* D**67**, 271–281.
Cowtan, K. (2006). *Acta Cryst.* D**62**, 1002–1011.
Evans, P. (2006). *Acta Cryst.* D**62**, 72–82.
Ferrer-Orta, C., Arias, A., Perez-Luque, R., Escarmís, C., Domingo, E. & Verdaguer, N. (2004). *J. Biol. Chem.* **279**, 47212–47221.
Garriga, D., Navarro, A., Querol-Audí, J., Abaitua, F., Rodriguez, J. F. & Verdaguer, N. (2007). *Proc. Natl Acad. Sci. USA*, **104**, 20540–20545.
Gorbalenya, A. E., Pringle, F. E., Zeddam, J. L., Luke, B. T., Cameron, C. E., Kalmakoff, J., Hanzlik, T. N., Gordon, K. H. J. & Ward, V. K. (2002). *J. Mol. Biol.* **324**, 47–62.
Graham, S. C., Sarin, L. P., Bahar, M. W., Myers, R. A., Stuart, D. I., Bamford, D. H. & Grimes, J. M. (2011). *PLoS Pathog.* **7**, e1002085.
Hao, Q. (2004). *J. Appl. Cryst.* **37**, 498–499.
Virus Taxonomy (2011). *Ninth Report of the International Committee on Taxonomy of Viruses*, edited by A. M. Q. King, M. J. Adams, E. B. Carstens & E. J. Lefkowitz. San Diego: Elsevier Academic Press.
Leslie, A. G. W. (2006). *Acta Cryst.* D**62**, 48–57.
Pan, J., Vakharia, V. N. & Tao, Y. J. (2007). *Proc. Natl Acad. Sci. USA*, **104**, 7385–7390.
Panjikar, S., Parthasarathy, V., Lamzin, V. S., Weiss, M. S. & Tucker, P. A. (2005). *Acta Cryst.* D**61**, 449–457.
Panjikar, S., Parthasarathy, V., Lamzin, V. S., Weiss, M. S. & Tucker, P. A. (2009). *Acta Cryst.* D**65**, 1089–1097.
Pannu, N. S., McCoy, A. J. & Read, R. J. (2003). *Acta Cryst.* D**59**, 1801–1808.
Pringle, F. M., Kalmakoff, J. & Ward, V. K. (2001). *J. Gen. Virol.* **82**, 259–266.
Schneider, T. R. & Sheldrick, G. M. (2002). *Acta Cryst.* D**58**, 1772–1779.
Sheldrick, G. M. (2008). *Acta Cryst.* A**64**, 112–122.
Sheldrick, G. M., Hauptman, H. A., Weeks, C. M., Miller, R. & Uson, I. (2001). *International Tables for Macromolecular Crystallography*, Vol. F, edited by M. G. Rossmann & E. Arnold, ch. 16, pp. 333–345. Dordrecht: Kluwer Academic Publishers.
Terwilliger, T. C. (2000). *Acta Cryst.* D**56**, 965–972.
Winn, M. D. *et al.* (2011). *Acta Cryst.* D**67**, 235–242.
Zeddam, J. L., Gordon, K. H., Lauber, C., Alves, C. A., Luke, B. T., Hanzlik, T. N., Ward, V. K. & Gorbalenya, A. E. (2010). *Virology*, **397**, 145–154.