

Luigi De Colibus,<sup>a</sup> Silvia Speroni,<sup>a</sup> Bruno Coutard,<sup>b</sup> Naomi L. Forrester,<sup>c</sup> Ernest Gould,<sup>c</sup> Bruno Canard<sup>b</sup> and Andrea Mattevi<sup>a\*</sup>

<sup>a</sup>Department of Genetics and Microbiology, University of Pavia, Via Ferrata 1, 27100 Pavia, Italy, <sup>b</sup>Architecture et Fonction des Macromolécules Biologiques, UMR 6098 CNRS et Université Aix-Marseille I et II, ESIL, Campus de Luminy, 13288 Marseille CEDEX 09, France, and <sup>c</sup>Centre for Ecology and Hydrology (formerly Institute of Virology), Mansfield Road, Oxford OX1 3SR, England

Correspondence e-mail: mattevi@ipvgen.unipv.it

Received 11 December 2006  
Accepted 31 January 2007

## Purification and crystallization of Kokobera virus helicase

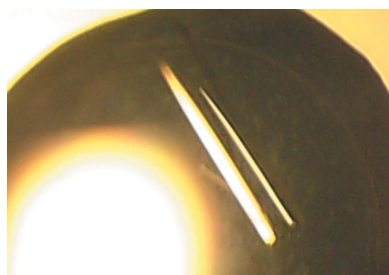
Kokobera virus is a mosquito-borne flavivirus belonging, like West Nile virus, to the Japanese encephalitis virus serocomplex. The flavivirus genus is characterized by a positive-sense single-stranded RNA genome. The unique open reading frame of the viral RNA is transcribed and translated as a single polyprotein which is post-translationally cleaved to yield three structural and seven nonstructural proteins, one of which is the NS3 gene that encodes a C-terminal helicase domain consisting of 431 amino acids. Helicase inhibitors are potential antiviral drugs as the helicase is essential to viral replication. Crystals of the Kokobera virus helicase domain were obtained by the hanging-drop vapour-diffusion method. The crystals belong to space group  $P3_121$  (or  $P3_221$ ), with unit-cell parameters  $a = 88.6$ ,  $c = 138.6$  Å, and exhibit a diffraction limit of 2.3 Å.

### 1. Introduction

The *Flavivirus* genus in the *Flaviviridae* family contains more than 70 recognized virus species. Approximately 80% of the flaviviruses are transmitted to animals and/or humans by host invertebrates such as ticks and mosquitoes (Heinz *et al.*, 2000). The flavivirus genome is composed of a single positive-sense RNA molecule that comprises several genes with replicative functions. The genome typically encodes a 370 kDa polyprotein precursor which is inserted into the membrane of the endoplasmic reticulum and processed to yield three structural proteins (C, M and E) and seven nonstructural replicative polyproteins (NS1, NS2A, NS2B, NS3, NS4A, NS4B and NS5). The mature and active forms of these proteins are generated by the combined action of the host proteases in the endoplasmic reticulum lumen and the viral proteases on the cytoplasmic side of the endoplasmic reticulum membrane. The C-terminal ~430 amino acids of the NS3 protein form a helicase domain belonging to helicase superfamily II. The precise biological functions of the NS3 helicase region are unknown, but it is thought to separate RNA daughter and template strands during replication, as well as unwinding the RNA secondary structure, particularly in the 3'-nontranslated region during initiation of RNA synthesis (Wu *et al.*, 2005).

Kokobera virus (KOKV) is a mosquito-borne flavivirus that has been isolated from mosquitoes throughout Australia and Papua New Guinea (Mackenzie *et al.*, 1994). It belongs to the Japanese encephalitis complex, along with West Nile and Murray encephalitis viruses. It was originally isolated from *Culex annulirostris* mosquitoes collected in the north of Queensland in 1960 and was subsequently named after a local Aboriginal tribe. Subsequently, KOKV has been isolated from mosquitoes collected in Western Australia, Northern Territory, New South Wales and Queensland (Doherty *et al.*, 1964). Serological evidence suggests that macropods (kangaroos and wallabies) and horses may be reservoir hosts of the virus (Doherty *et al.*, 1971; Russell, 1995). Human infections with KOKV occasionally result in an acute polyarticular disease (Mackenzie *et al.*, 1994; Doherty *et al.*, 1971; Russell, 1995; Hawkes *et al.*, 1985, 1993).

Because the helicase has been proposed as a possible therapeutic target for drug design and as structural information is necessary to guide pharmacological studies, we have undertaken structural characterization of the KOKV helicase. A recombinant protein was produced, purified and crystallized. Here, we report the purification



© 2007 International Union of Crystallography  
All rights reserved

and crystallization protocols and the data from preliminary X-ray diffraction studies.

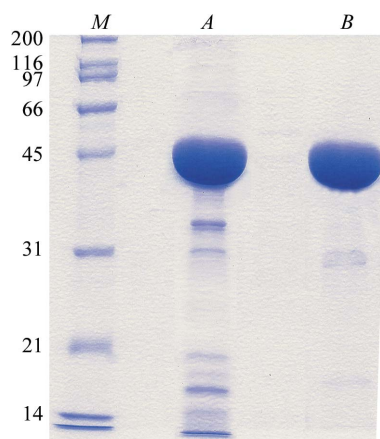
## 2. Experimental procedures and results

### 2.1. Construction of the expression vector

RT-PCR was carried out using Superscript RT II (Invitrogen) and Biotaq (Bioline) and primers 5'-GGATCGCTACGTGAGCG (forward sense) and 5'-GAGCGGCTCAAATGGCAG (reverse sense). The resulting cDNA codes for the amino-acid sequence corresponding to that between residues 189 and 620 of the predicted NS3 polypeptide. This sequence is believed to encode the functional domain of the helicase enzyme, whereas the full helicase protein ends at the NS3-NS4A cleavage site. The cDNA was re-amplified using Gateway-modified primers UNK-HEL-01HN-F (5'-GGGGACAA-GTTTGTACAAAAAGCAGGCTTCGAAGGAGATGCCACCA-TGAAACATCACCATCACCATCACAAAAGAGAGCTGACGG-TCTTGACTTG) and UNK-HEL-01HN (5'-GGGGACCACTTTGTACAAGAAAGCTGGGTCTTATTACCTTTCCCGGATGCA-AATTCCTTAAA). Nucleotides in bold correspond to viral sequences. The PCR product was subsequently cloned into pDest14 expression vector (Invitrogen), which is designed to produce an N-terminal six-His-tagged recombinant protein. The insert was checked by sequencing. Three mutations were observed (Met47Thr, Phe86Leu and Ile168Thr). Phe86 and Ile168 mutations occur in nonconserved amino acids as judged by multiple alignment using flavivirus helicases genes (not shown). Met47Thr occurs in a more conserved methionine of motif Ia, the function of which is unknown (Benarroch *et al.*, 2004). However, this methionine is not conserved in several flaviviruses (*e.g.* Kamiti river virus, cell fusing agent virus and Tamana bat virus) nor in pestiviruses, which are expected to have a very similar helicase three-dimensional structure. The mutant protein was further tested to ascertain its biochemical relevance. It was found to be enzymatically active in that it exhibited a high ATPase activity (data to be published elsewhere) and thus was further selected for crystallogensis.

### 2.2. Protein expression and purification

The construct was transformed into four *Escherichia coli* strains (BL21 pLysS, Rosetta, Origami pLysS and C41 pROS) in order to

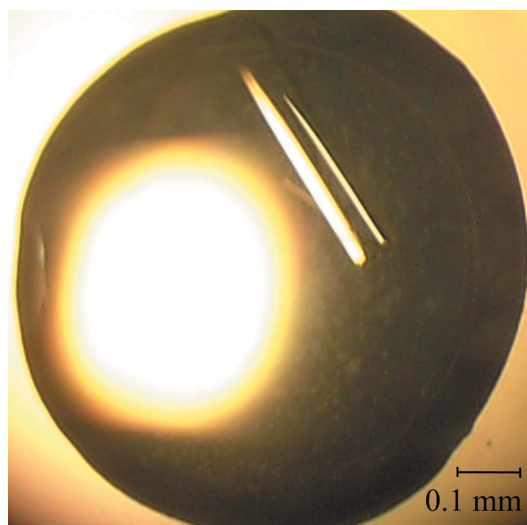


**Figure 1** Representative protein samples from each stage of the purification analysed by SDS-PAGE [10% (*w/v*) acrylamide]. Lanes *A* and *B* correspond to the protein eluted from the nickel-affinity chromatography column and the protein eluted from final gel-filtration step, respectively. Lane *M* contains molecular-weight markers (kDa).

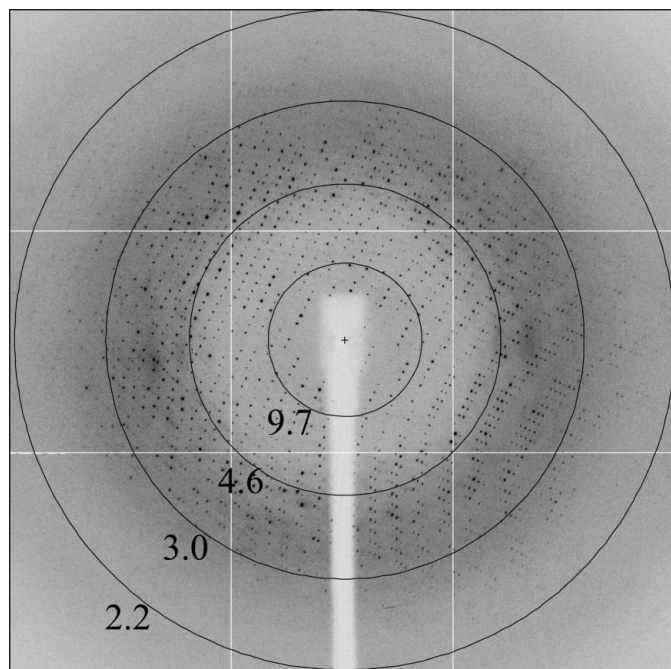
identify the optimal conditions for protein overexpression (Berrow *et al.*, 2006). Expression of soluble protein was detected by a dot-blot procedure following a previously described method (Abergel *et al.*, 2003). The best results were obtained using *E. coli* C41 pROS (a strain from Avidis carrying the pROS plasmid from Novagen) grown in 2YT medium at 293 K. These conditions were further improved, resulting in the protocol that was routinely used for large-scale protein production. Transformed cells were grown in 21 Luria-Bertani medium containing 100 µg ml<sup>-1</sup> ampicillin at 293 K until the optical density at 600 nm (OD<sub>600</sub>) reached a value of 0.6. Protein expression was induced by the addition of isopropyl β-D-thiogalactopyranoside to a final concentration of 0.5 mM. Expression of soluble protein was obtained by incubating the induced culture overnight at 293 K. Cells were collected by centrifugation and then suspended in 50 ml extraction buffer [50 mM Tris-HCl pH 8.0, 300 mM NaCl, 10 mM imidazole, 0.1% (*v/v*) Triton X-100, 5% (*v/v*) glycerol]. The cell suspension was sonicated at 400 W for 15 min using a Vibra Cell Sonicator (Bioblock Scientific), centrifuged and the supernatant was loaded onto an Ni-Sepharose High Performance column (GE Healthcare) to perform immobilized metal-affinity chromatography. The recombinant protein was eluted with a linear gradient of imidazole from 10 to 500 mM. It was then concentrated and applied onto a Hi-Load Superdex-75 column (GE Healthcare) and eluted in gel-filtration buffer (10 mM imidazole, 300 mM NaCl pH 8.0). The purity of the eluted protein was assessed by SDS-PAGE (Laemmli, 1970) (Fig. 1). The purified protein (molecular weight 49 800 Da) was concentrated to 14 mg ml<sup>-1</sup> in 10 mM imidazole, 300 mM NaCl pH 8.0. Typically, a total of about 50 mg of pure protein was obtained from 4 l bacterial culture.

### 2.3. Crystallization

Vapour-diffusion crystallization experiments were performed using an Oryx8 crystallization robot (Douglas Instruments). In a typical experiment, 0.3 µl screening solution was added to 0.3 µl protein solution on 96-well Crystal Clear P plates (Douglas Instruments); the reservoir wells contained 90 µl screening solution. Crystallization trials were performed at 277 and 293 K. Large single crystals were observed after 15 d in condition No. 21 (0.75 M trisodium citrate and 0.01 M sodium borate pH 8.5) of the Stura Footprint Screen No. 1



**Figure 2** A typical crystal of KOKV helicase domain grown in a hanging-drop vapour-diffusion experiment.



**Figure 3**  
Diffraction image of a KOKV helicase crystal collected on beamline ID23-EH1 at the ESRF. The resolution limits of the circles are in Å.

(Stura *et al.*, 1992; Molecular Dimensions Ltd). The most promising conditions were optimized using the hanging-drop vapour-diffusion method (Ducruix & Giegé, 1992). Crystallization droplets were set up manually by mixing 1  $\mu$ l protein solution and 1  $\mu$ l reservoir solution. The volume of the reservoir solution was 1 ml. The largest crystals grew at 293 K in a precipitant solution consisting of 900 mM trisodium citrate, 100 mM imidazole pH 8.0. A KOKV helicase crystal is shown in Fig. 2.

#### 2.4. X-ray diffraction data collection and preliminary analysis

For data collection, the crystals were soaked for a few seconds in a solution containing 1 M trisodium citrate, 100 mM imidazole pH 8.0 and 15% (v/v) glycerol for cryoprotection and were then flash-cooled in liquid nitrogen. X-ray diffraction data were collected at 100 K using an ADSC Q315 detector (Area Detector Systems Corporation) on beamline ID23-EH1 at the European Synchrotron Radiation Facility (ESRF; Grenoble, France). About 200 images ( $\Delta\varphi = 1.0$ ) were collected for each single crystal. The images (Fig. 3) were processed using *MOSFLM* (Leslie, 1992) and the *CCP4* suite of programs (Collaborative Computational Project, Number 4, 1994). The crystal diffracted to 2.3 Å resolution and the data-collection statistics for the best measured data set are reported in Table 1. The KOKV helicase crystals belong to the trigonal  $P3_121$  (or  $P3_221$ ) space group. The calculated value of the Matthews coefficient  $V_M$  is 3.2 Å<sup>3</sup> Da<sup>-1</sup>, assuming the presence of one monomer in the asymmetric unit (Matthews, 1968). The corresponding value of the solvent content is 62.0%. Structure determination will be carried out by molecular replacement using the structures of homologous flavivirus

**Table 1**

X-ray data-collection statistics for the KOKV helicase domain.

Values in parentheses are for the highest resolution shell.

|                            |                            |
|----------------------------|----------------------------|
| Space group                | $P3_121$ (or $P3_221$ )    |
| Unit-cell parameters (Å)   | $a = 88.57$ , $c = 138.58$ |
| Resolution (Å)             | 29.5–2.3 (2.4–2.3)         |
| Total measurements         | 587300 (17542)             |
| Unique reflections         | 28611 (994)                |
| Completeness (%)           | 99.9 (100)                 |
| Redundancy                 | 20.5 (21.1)                |
| $R_{\text{merge}}^\dagger$ | 13.1 (47.4)                |
| $I/\sigma(I)$              | 24.3 (4.2)                 |

$^\dagger R_{\text{merge}} = \sum_h \sum_i |I(h)_i - \langle I(h) \rangle| / \sum_h \sum_i I(h)_i$ , where  $I(h)_i$  is the scaled observed intensity of the  $i$ th symmetry-related observation for reflection  $h$  and  $\langle I(h) \rangle$  is the average intensity.

helicases (Wu *et al.*, 2005) as search models or by the method of single-wavelength anomalous dispersion using selenomethionine-labelled protein.

This work was supported by a Sixth Framework grant from the European Community (acronym VIZIER, contract No. 511960).

#### References

- Abergel, C., Coutard, B., Byrne, D., Chenivresse, S., Claude, J. B., Deregnacourt, C., Fricaux, T., Giancesini-Boutreux, C., Jeudy, S., Lebrun, R., Maza, C., Notredame, C., Poirot, O., Suhre, K., Varagnol, M. & Claverie, J. M. (2003). *J. Struct. Funct. Genomics*, **4**, 141–157.
- Benarroch, D., Selisko, B., Locatelli, G., Maga, G., Romette, J. L. & Canard, B. (2004). *Virology*, **328**, 208–218.
- Berrow, N. S., Bussow, K., Coutard, B., Diprose, J., Ekberg, M., Folkers, G. E., Levy, N., Lieu, V., Owens, R. J., Peleg, Y., Pinaglia, C., Quevillon-Cheruel, S., Salim, L., Scheich, C., Vincentelli, R. & Busso, D. (2006). *Acta Cryst.* **D62**, 1218–1226.
- Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* **D50**, 760–763.
- Doherty, R. L., Carley, J. G. & Gorman, B. M. (1964). *Aust. J. Exp. Biol. Med. Sci.* **42**, 149–164.
- Doherty, R. L., Standfast, H. A., Domrow, R., Wetters, E. J., Whitehead, R. H. & Carley, J. G. (1971). *Trans. R. Soc. Trop. Med. Hyg.* **65**, 504–513.
- Ducruix, A. & Giegé, R. (1992). *Crystallization of Nucleic Acids and Proteins. A Practical Approach*, edited by A. Ducruix & R. Giegé, pp. 73–97. Oxford: IRL Press.
- Hawkes, R. A., Boughton, C. R., Naim, H. M., Wild, J. & Chapman, B. (1985). *Med. J. Aust.* **143**, 555–561.
- Hawkes, R. A., Pamplin, J., Boughton, C. R. & Naim, H. M. (1993). *Med. J. Aust.* **159**, 159–162.
- Heinz, F. X., Collett, M. S., Purcell, R. H., Gould, E. A., Howard, C. R., Houghton, M., Moormann, R. J., Rice, C. M. & Thiel, H. J. (2000). *Virus Taxonomy: Classification and Nomenclature of Viruses*, edited by M. H. V. van Regenmortel, C. M. Fauquet, D. H. L. Bishop, E. B. Carstens, M. K. Estes, S. M. Lemon, J. Maniloff, M. A. Mayo, D. McGeoch, C. R. Pringle & R. B. Wickner, p. 859–878. San Diego: Academic Press.
- Laemmli, U. K. (1970). *Nature (London)*, **227**, 680–685.
- Leslie, A. G. W. (1992). *Jnt CCP4/ESF-EACBM Newsl. Protein Crystallogr.* **26**.
- Mackenzie, J. S., Lindsay, M. D., Coelen, R. J., Broom, A. K., Hall, R. A. & Smith, D. W. (1994). *Arch. Virol.* **136**, 447–467.
- Matthews, B. W. (1968). *J. Mol. Biol.* **33**, 491–497.
- Russell, R. C. (1995). *Rev. Med. Vet. Entomol.* **83**, 141–158.
- Stura, E. A., Nemerow, G. R. & Wilson, I. A. (1992). *J. Cryst. Growth*, **122**, 273–285.
- Wu, J., Bera, A. K., Kuhn, R. J. & Smith, J. L. (2005). *J. Virol.* **79**, 10268–10277.