Expression, purification and crystallization of the SARS-CoV *macro* domain
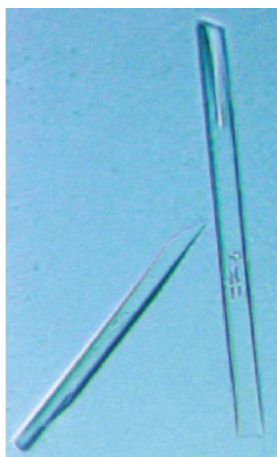
Hélène Malet,‡ Karen Dalle,‡ Nicolas Brémond,‡ Fabienne Tocque, Stéphanie Blangy, Valérie Campanacci, Bruno Coutard, Sacha Grisel, Julie Lichière, Violaine Lantez, Christian Cambillau, Bruno Canard and Marie-Pierre Egloff*

Centre National de la Recherche Scientifique and Universités d'Aix-Marseille I et II, UMR 6098, Architecture et Fonction des Macromolécules Biologiques, UMR 6098-Case 932, 163 Avenue de Luminy, 13288 Marseille CEDEX 9, France

‡ These authors made an equal contribution to this work.

Correspondence e-mail: marie-pierre.egloff@afmb.univ-mrs.fr

Received 14 December 2005
Accepted 13 March 2006

Macro domains or X domains are found as modules of multidomain proteins, but can also constitute a protein on their own. Recently, biochemical and structural studies of cellular *macro* domains have been performed, showing that they are active as ADP-ribose-1″-phosphatases. *Macro* domains are also present in a number of positive-stranded RNA viruses, but their precise function in viral replication is still unknown. The major human pathogen severe acute respiratory syndrome coronavirus (SARS-CoV) encodes 16 non-structural proteins (nsps), one of which (nsp3) encompasses a *macro* domain. The SARS-CoV nsp3 gene region corresponding to amino acids 182–355 has been cloned, expressed in *Escherichia coli*, purified and crystallized. The crystals belong to space group $P2_1$, with unit-cell parameters $a = 37.5$, $b = 55.6$, $c = 108.9$ Å, $\beta = 91.4°$, and the asymmetric unit contains either two or three molecules. Both native and selenomethionine-labelled crystals diffract to 1.8 Å.

1. Introduction

Coronaviruses have the largest genomes of all positive-stranded RNA viruses. Translation of their replicase genes results in two large polyproteins which are subsequently cleaved into a large number of non-structural proteins (16 for the severe acute respiratory syndrome coronavirus; SARS-CoV) whose function is to ensure genomic and subgenomic RNA synthesis (Lai & Holmes, 2001; Siddell, 1995; Siddell *et al.*, 2005). The particularly large size of the genome probably results from the fact that it encodes for unusual enzymatic activities within the viral world (Snijder *et al.*, 2003). The precise roles of some of these proteins have yet to be identified (Ziebuhr, 2005).

SARS-CoV nsp3 is a large (213 kDa) multidomain protein believed to carry several enzymatic activities involved in the replicative complex, which includes a *macro* domain, also known as an X domain (Gorbalenya *et al.*, 1991, 2004; Snijder *et al.*, 2003). To date, 330 proteins have been reported to possess such a domain (from the SMART database; Letunic *et al.*, 2004). The *macro* domain is found in a number of multidomain proteins, including nonstructural proteins of several ssRNA viruses (such as rubella virus, alphaviruses and some coronaviruses), some poly-ADP-ribose polymerases and the C-terminal domain of macro-H2A histone proteins, but it can also constitute a stand-alone protein by itself, as is the case in bacteria, archaebacteria and eukaryotes (F. Forouhar, I. Lee, S. M. Vorobiev, R. Xiao, T. B. Acton, G. T. Montelione, J. F. Hunt & L. Tong, unpublished work; Allen *et al.*, 2003; Shull *et al.*, 2005; Letunic *et al.*, 2004). This wide distribution suggests an essential and ubiquitous role in cellular processes.

Several studies have recently revealed that *macro* domains act as phosphatases that remove the 1″-phosphate group from ADP-ribose-1″-phosphate (Martzen *et al.*, 1999; Karras *et al.*, 2005; Shull *et al.*, 2005; Putics *et al.*, 2005). The role of this activity, particularly in the viral world, is still puzzling. We are interested in identifying the exact role of viral *macro* domains by a combination of enzymatic characterization studies and structural studies of the enzyme alone or in complex with biologically relevant molecules. Therefore, we have initiated the structural and functional characterization of Semliki Forest virus, human coronavirus 229E and SARS coronavirus *macro* domains. During the course of this study, the crystal structure of a slightly different construct of another strain of SARS-CoV (strain

Tor2) was solved as an apo enzyme and published (Saikatendu *et al.*, 2005). The Tor2 protein is dimeric, both in solution and in the crystal, and its structure reveals that the active site is occluded by the dimeric association, which explains why soaking experiments with various ligands failed for this particular crystal structure.

Here, we report the cloning, expression, purification, crystallization and preliminary X-ray characterization of a monomeric form of the SARS-CoV strain Frankfurt 1 *macro* domain (an 18.6 kDa protein consisting of amino acids 182–355 of nsp3), which may enable more successful ligand-soaking experiments and hence provide some insight into the mechanism of viral *macro* domains.

2. Cloning, expression and purification

As soon as the SARS-CoV genome sequence became available, we performed the annotation of its genome independently using the *VaZyMolO* ('viral enzyme module localization') tool that had been previously developed in our laboratory. The scope of *VaZyMolO* is to define within large viral polyproteins viral proteins and protein modules that might be expressed in a soluble and functionally active form (Ferron *et al.*, 2005). Our annotation of the proteins (non-structural, structural and accessory) encoded by the SARS-CoV genome was similar to that of Snijder *et al.* (2003). In the case of nsp3, we defined six independent domains, including a *macro* domain. Being located between two poorly characterized domains (namely a glutamic acid-rich acidic domain and a SARS-specific unique domain; SUD), the SARS-CoV *macro* domain did not have clearly defined borders. Our annotation within the *VaZyMolO* database was based on multiple criteria, including bioinformatics approaches, examination of the available three-dimensional structures of cellular *macro* domains and existing biochemical data. The latter suggests that in some viruses the N-terminal border definition may be particularly crucial for detection of ADP-ribose-1″-phosphatase activity (T. Ahola, personal communication). The coding sequence for this domain (consisting of 174 amino acids, amino acids 182–355 of nsp3, isolate Frankfurt 1, DDB/EMBL/Genbank accession No. AY291315)
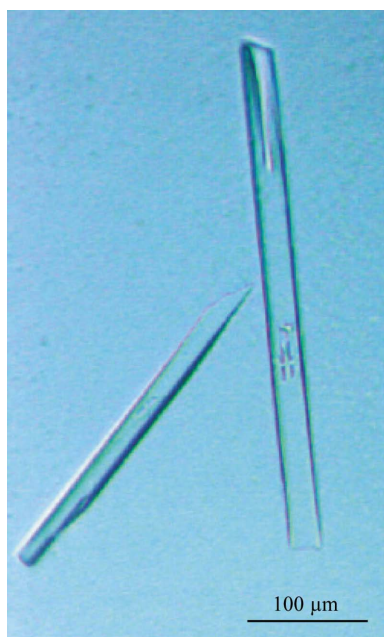
Figure 1
Crystals of SARS-CoV *macro* domain.

was amplified from the pMal-SARS-CoV-X plasmid kindly provided by Akos Putics and John Ziebuhr (Putics *et al.*, 2005). Amplification was performed by PCR using two primers containing the attB sites of the Gateway recombination system (Invitrogen). A sequence encoding a hexahistidine tag was added at the 5′ end of the gene. The cDNA was then subcloned in the pDest14 plasmid (Invitrogen). The open reading frame of the final construct was checked by sequencing (MilleGen, Toulouse, France).

Expression was performed in *Escherichia coli* strain C41pROS, a C41 strain into which we have introduced the pLys plasmid from the Rosetta strain (Novagen). Cultures were grown in TB at 310 K and $250g$ to mid-exponential phase (an OD_{600} of 0.6). Media were supplemented with 100 mg ml^{-1} ampicillin to select for *E. coli* recombinants. Protein expression was induced by the addition of a final concentration of 50 µM isopropyl 1-thio-β-D-galactopyranoside (IPTG) and the cells were incubated for a further 20 h at 298 K and 250 rev min^{-1}. Cells were collected by centrifugation at 5000g for 20 min at 277 K and the bacterial pellets were resuspended and then frozen at 253 K in 50 mM Tris–HCl, 150 mM NaCl, 10 mM imidazole pH 8.0 (10 ml per OD_{600} unit and per litre of culture). Cellular suspensions were thawed, supplemented with 1 mM PMSF, 0.2 mg ml^{-1} lysozyme, 0.1 µg ml^{-1} DNase, 20 mM MgSO$_4$ and protease-cocktail inhibitor (Sigma), lysed by sonication and centrifuged (20 000g) for 40 min at 277 K to produce cell-free extract. All purification steps were performed at 277 K. The supernatant was applied onto a 5 ml bed-volume HiTrap nickel immobilized metal-ion affinity chromatography (IMAC) column (Amersham Biosciences) connected to an FPLC system (Amersham Biosciences). The protein was eluted with 50 mM Tris–HCl, 150 mM NaCl, 250 mM imidazole pH 8.0. Fractions containing the protein were determined by SDS–PAGE, pooled and then applied onto a preparative Superdex 200 gel-filtration column pre-equilibrated in 10 mM MOPS, 150 mM NaCl pH 7.0. Protein was eluted with a retention volume corresponding to a monomeric *macro* domain. The purity of the protein was evaluated by SDS–PAGE. Protein was concentrated to 8 mg ml^{-1} using a Viva-spin 10 kDa molecular-weight cutoff centrifugal concentrator (Viva-science). The protein concentration was determined from the molar extinction coefficient of the enzyme at 280 nm (9840 M^{-1} cm^{-1}) as calculated by the *Expasy* server (http://www.expasy.org/tools/protparam.html). The purified protein was stored at 277 K. Seleno-methionine (SeMet) labelled protein was prepared following standard procedures (Doublié, 1997) and the protein was purified using the same protocol employed for the native protein except that 1 mM TCEP was added to the protein immediately after its elution from the IMAC column and to the gel-filtration buffer. Typical yields were 15–20 mg of pure protein per litre of culture for both native and SeMet-labelled protein.

3. Crystallization

After checking the protein purity, folding and monodispersity using SDS–PAGE, circular dichroism and dynamic light scattering (Zulauf & D'Arcy, 1992), crystallization of the SARS-CoV *macro* domain was screened at 292 K by the sitting-drop vapour-diffusion method in 96-well Greiner crystallization plates which had been automatically filled using a Tecan Genesis robot. As the amount of protein was not a limiting parameter, we used a wide-screen approach using the following commercial kits: Wizard Screens I and II (Emerald Bio-structures), Structure Screens I and II (Jancarik & Kim, 1991; Hampton Research), Stura Footprint Screen (Stura *et al.*, 1992; Molecular Dimension Limited) and Natrix Screen (Scott *et al.*, 1995;

Table 1
Comparison of experimental procedures and results.

	Saikatendu *et al.* (2005)	Present work
Strain	Toronto 2 isolate, gi:34555776	Frankfurt 1 isolate, gi:31581502
Construct	184–365 of nsp3	182–355 of nsp3
No. of amino acids	182	174
N-terminal tag	MGSDKIHHHHHH	MHHHHHH
Protein buffer	25 m*M* Tris pH 7.8, 150 m*M* NaCl	10 m*M* MOPS pH 7.0, 150 m*M* NaCl
Crystallization conditions	1.85 *M* sodium malonate pH 7.0	0.1 *M* imidazole pH 7.9, 1.3 *M* sodium citrate
Space group	$P2_12_12_1$	$P2_1$
Unit-cell parameters (Å, °)	$a = 76.9, b = 81.2, c = 125.7$	$a = 37.4, b = 55.4, c = 108.9, \beta = 91.6$
Diffraction limit (Å)	1.4	1.8
Solvent content (%)	51	38.9

Table 2
Data-collection statistics.

Values in parentheses are for the highest resolution shell.

	Native data	SeMet data
X-ray source	ESRF ID29	ESRF ID29
Wavelength (Å)	0.9792	0.9792
Crystal parameters		
Space group	$P2_1$	$P2_1$
Unit-cell parameters		
a (Å)	37.4	37.5
b (Å)	55.4	55.6
c (Å)	108.8	108.9
β (°)	91.6	91.4
Resolution range (Å)	40–1.8 (1.9–1.8)	28–1.8 (1.9–1.8)
Data processing		
Observed reflections	238756	432355
Unique reflections	41448	41654
R_{merge} (%)	7.7 (18.2)	10.9 (38.6)
Completeness (%)	99.6 (98.7)	99.9 (99.9)
Multiplicity	4.8 (5.6)	10.4 (8.9)
Mean $I/\sigma(I)$	16.2 (7.9)	17.7 (7.1)

† $R_{merge} = \sum_{hkl}\sum_i |I - \langle I \rangle|/\sum_{hkl}\sum_i I$.

Hampton Research). Using the 8 mg ml^{-1} purified protein sample, three drops were set up by a Cartesian robot controlled by the *AXSYS* software above each reservoir; these drops were made up of 300, 200 and 100 nl of protein solution, respectively, and 100 nl reservoir solution (Sulzenbacher *et al.*, 2002). From these 864 conditions, two hits were identified in the Stura Footprint Screen. A primary optimization of these conditions was performed using home-made solutions, which were dispensed in 64 wells of a Greiner crystallization plate in a 8 × 8 format by the Tecan Genesis robot. The best conditions were then transposed to hanging drops in 24-well tissue-culture Limbro plates. As in the Greiner plates, hundreds of plate-like multiple crystals grew from a single nucleation point overnight. Microseeding in undersaturated equilibrated drops led to single crystals that were suitable for X-ray analysis (Fig. 1). The final conditions for crystallization were as follows: 2 µl protein solution (8 mg ml^{-1}) was mixed with 1 µl reservoir solution (total volume of 0.5 ml) containing 0.1 *M* imidazole pH 7.9, 1.3 *M* sodium citrate. Multiple crystals grew overnight. They were crushed and used to microseed a pre-equilibrated drop made of 2 µl protein solution (8 mg ml^{-1}) mixed with 1 µl of a reservoir solution containing 0.1 *M* imidazole pH 7.8, 0.9 *M* sodium citrate. These conditions differ from those found by Saikatendu and coworkers (see Table 1), which is not surprising considering that the constructs are quite different (considering the N-terminal tag, the construct of Saikatendu and coworkers is three amino acids longer at the N-terminal and ten amino acids longer at the C-terminal extremities).

4. Data collection and processing

Both native and SeMet crystals were cryoprotected in a solution containing the reservoir solution to which 20% glycerol was added and harvested into a rayon-fibre loop prior to flash-freezing in liquid nitrogen. All data were collected from single crystals at 100 K at beamline ID29 of the European Synchrotron Radiation Facility, Grenoble, France using an ADSC Quantum 210 charge-coupled device detector. Data were processed with *MOSFLM* (Leslie, 1992) and were reduced and scaled using the *SCALA* program from the *CCP*4 suite (Collaborative Computational Project, Number 4, 1994). Crystals belonged to space group $P2_1$, with unit-cell parameters $a = 37.5, b = 55.6, c = 108.9$ Å, $\beta = 91.4$°. Data-collection statistics are shown in Table 2. The asymmetric unit may contain either three monomers, giving a V_M value (Matthews, 1968) of 2.0 Å^3 Da^{-1} and 38.9% solvent, or two monomers, giving a V_M value of 3.0 Å^3 Da^{-1} and 59.3% solvent. A self-rotation function did not resolve this ambiguity, as no significant peak corresponding to either twofold or

threefold non-crystallographic symmetry was observed. The native Patterson map also did not reveal any peaks. Attempts to solve the structure by molecular replacement using various models based on the coordinates of the *macro* domain from *Archaeoglobus fulgidus* (Allen *et al.*, 2003; PDB code 1hjz), which shares 29% identity with the SARS-CoV *macro* domain and using various programs [*AMoRe* (Navaza, 1994), *MOLREP* (Vagin & Teplyakov, 1997) and *Phaser* (McCoy *et al.*, 2005)] proved unsuccessful. Using a SAD data set collected from an SeMet crystal, we have determined the positions of all Se atoms and structure refinement is ongoing.

As previously mentioned, the molecular packing observed by Saikatendu and coworkers did not allow either cocrystallization or ligand-soaking experiments as the active site of both molecules was occluded by the dimer interface. The different crystal form obtained in this study may enable more successful ligand-soaking experiments.

References

Allen, M. D., Buckle, A. M., Cordell, S. C., Löwe, J. & Bycroft, M. (2003). *J. Mol. Biol.* **330**, 503–511.

Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* D**50**, 760–763.

Doublié, S. (1997). *Methods Enzymol.* **276**, 523–530.

Ferron, F., Rancurel, C., Longhi, S., Cambillau, C., Henrissat, B. & Canard, B. (2005). *J. Gen. Virol.* **86**, 743–749.

Gorbalenya, A. E., Koonin, E. V. & Lai, M. M. (1991). *FEBS Lett.* **288**, 201–205.

Gorbalenya, A. E., Snijder, E. & Spaan, W. J. M. (2004). *J. Virol.* **78**, 7863–7866.

Jancarik, J. & Kim, S.-H. (1991). *J. Appl. Cryst.* **24**, 409–411.

Karras, G. I., Kustatscher, G., Buhecha, H. R., Allen, M. D., Pugieux, C., Sait, F., Bycroft, M. & Ladurner, A. G. (2005). *EMBO J.* **24**, 1911–1920.

Lai, M. M. C. & Holmes, K. V. (2001). *Fields Virology*, 4th ed., edited by D. M. Knipe & P. M. Howley, pp. 1163–1185. Philadelphia: Lippincott, Williams & Wilkins.

Leslie, A. G. W. (1992). *Jnt CCP4/ESF–EACBM Newsl. Protein Crystallogr.* **26**.

crystallization communications

Letunic, I., Copley, R. R., Schmidt, S., Ciccarelli, F. D., Doerks, T., Schultz, J., Ponting, C. P. & Bork, P. (2004). *Nucleic Acids Res.* **32**, D142–D144.

McCoy, A. J., Grosse-Kunstleve, R. W., Storoni, L. C. & Read, R. J. (2005). *Acta Cryst.* D**61**, 458–464.

Martzen, M. R., McCraith, S. M., Spinelli, S. L., Torres, F. M., Fields, S., Grayhack, E. J. & Phizicky, E. M. (1999). *Science*, **286**, 1153–1155.

Matthews, B. W. (1968). *J. Mol. Biol.* **33**, 491–497.

Navaza, J. (1994). *Acta Cryst.* A**50**, 157–163.

Putics, A., Filipowicz, W., Hall, J., Gorbalenya, A. E. & Ziebuhr, J. (2005). *J. Virol.* **79**, 12721–12731.

Saikatendu, K. S., Joseph, J. S., Subramanian, V., Clayton, T., Griffith, M., Moy, K., Velasquez, J., Neuman, B. W., Buchmeier, M. J., Stevens, R. C. & Kuhn, P. (2005). *Structure*, **13**, 1665–1675.

Scott, W. G., Finch, J. T., Grenfell, R., Fogg, J., Smith, T., Gait, M. J. & Klug, A. (1995). *J. Mol. Biol.* **250**, 327–332.

Shull, N. P., Spinelli, S. L. & Phizicky, E. M. (2005). *Nucleic Acids Res.* **33**, 650–660.

Siddell, S. G. (1995). In *The Viruses*, edited by H. Fraenkel-Conrat & R. R. Wagner. New York: Plenum Press.

Siddell, S. G., Ziebuhr, J. & Snijder, E. J. (2005). *Topley and Wilson's Microbiology and Microbial Infections*, 10th ed., Vol. 1, edited by B. W. J. Mahy & V. ter Meulen, pp. 823–856. London: Hodder Arnold.

Snijder, E. J., Bredenbeek, P. J., Dobbe, J. C., Thiel, V., Ziebuhr, J., Poon, L. L., Guan, Y., Rozanov, M., Spaan, W. J. & Gorbalenya, A. E. (2003). *J. Mol. Biol.* **331**, 991–1004.

Stura, E. A., Nemerow, G. R. & Wilson, I. A. (1992). *J. Cryst. Growth*, **122**, 273–285.

Sulzenbacher, G. *et al.* (2002). *Acta Cryst.* D**58**, 2109–2115.

Vagin, A. & Teplyakov, A. (1997). *J. Appl. Cryst.* **30**, 1022–1025.

Ziebuhr, J. (2005). *Curr. Top. Microbiol. Immunol.* **287**, 57–94.

Zulauf, M. & D'Arcy, A. (1992). *J. Cryst. Growth*, **122**, 102–106.