

Vincent Cura, Nathalie Troffer-Charlier, Marie-Annick Lambert, Luc Bonnefond and Jean Cavarelli\*

Département de Biologie Structurale Intégrative, Institut de Génétique et de Biologie Moléculaire et Cellulaire (IGBMC), INSERM U964/CNRS UMR 7104/Université de Strasbourg, 1 Rue Laurent Fries, 67404 Illkirch, France

Correspondence e-mail: cava@igbmc.fr

Received 6 November 2013

Accepted 3 December 2013

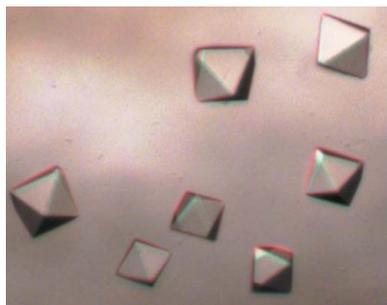
## Cloning, expression, purification and preliminary X-ray crystallographic analysis of mouse protein arginine methyltransferase 7

Protein arginine methyltransferase 7 (PRMT7) is a unique but less characterized member of the family of protein arginine methyltransferases (PRMTs) that plays a role in male germline gene imprinting. PRMT7 is the only known PRMT member that catalyzes the monomethylation but not the dimethylation of the target arginine residues and harbours two catalytic domains in tandem. PRMT7 genes from five different species were cloned and expressed in *Escherichia coli* and Sf21 insect cells. Four gave soluble proteins from Sf21 cells, of which two were homogeneous and one gave crystals. The mouse PRMT7 structure was solved by the single anomalous dispersion method using a crystal soaked with thimerosal that diffracted to beyond 2.1 Å resolution. The crystal belonged to space group  $P4_32_12$ , with unit-cell parameters  $a = b = 97.4$ ,  $c = 168.1$  Å and one PRMT7 monomer in the asymmetric unit. The structure of another crystal form belonging to space group  $I222$  was solved by molecular replacement.

### 1. Introduction

Methylation of arginine is a prevalent post-translational modification found in eukaryotes and is mediated by protein arginine methyltransferases (PRMTs). PRMTs have been implicated in a wide array of biological processes, such as regulation of transcription, RNA metabolism and DNA-damage repair (Bedford & Clarke, 2009; Yang & Bedford, 2013). Since deregulation of these processes appears to be implicated in the pathogenesis of various diseases, such as human cancers (Cha & Jho, 2012), understanding the mechanism of action of PRMTs at the atomic scale is therefore crucial both for fundamental biology and for pharmacological applications. PRMTs catalyse the transfer of the methyl group from *S*-adenosyl-L-methionine (SAM) to the side-chain N atoms of arginine residues to form methylated arginines and *S*-adenosyl-L-homocysteine (SAH). The target arginine residues, which are mainly but not exclusively located within glycine- and arginine-rich patches (GAR motifs), can be monomethylated or dimethylated. The PRMT family has been the focus of much biological research, and 11 members of the family (PRMT1–PRMT11) have been identified in mammalian cells (Wolf, 2009). PRMTs are modular proteins that vary from 320 to 970 amino acids and share a common catalytic methyltransferase domain to which additional domains are added, thus conferring structural and functional specificity to each PRMT. The PRMT catalytic domain belongs to the class I type of SAM-dependent methyltransferases (Katz *et al.*, 2003) and harbours a set of four conserved sequence motifs and a THW loop.

PRMT7 displays several unique characteristics. It is the only known PRMT member that produces only monomethylated arginine (MMA) residues (Zurita-Lopez *et al.*, 2012). PRMT7 was first identified in a genetic screen for susceptibility to chemotherapeutic cytotoxicity (Gros *et al.*, 2003). It was initially demonstrated to be capable of generating both MMA and symmetric dimethylated arginine (sDMA) depending on the substrate that was used in the assay (Miranda *et al.*, 2004; Lee *et al.*, 2005). PRMT7, together with PRMT5, is involved in the methylation of histone 3 Arg2, an activation mark of transcription (Migliori *et al.*, 2012), and histone 4 Arg3, a repressor mark of transcription for several genes implicated in DNA repair (Karkhanis *et al.*, 2012). PRMT7 methylates eukaryotic elongation factor 2 (eEF2), a function coordinated by PRMT5 and regulated by the basic fibroblast growth factor (bFGF) (Jung *et*



**Table 1**

Macromolecule-production information.

Organism	<i>Mus musculus</i>	<i>Homo sapiens</i>
DNA source	MGC:7929	Addgene 34693
Forward primer	GATCGACCATATGAAAATCTTCTGCAGTCGG	CATATGAAGGTCTTCTGTGGCCG
Reverse primer	GAGCAGATCTTCAGTCTGGGGTATCTGCATG	AGATCTTCAGCTCAAGGTGTCTG
Cloning vector	pDONR207	pCR-Blunt
Expression vector	pBacGGWH	pBacGGWH
Expression host	Sf21	Sf21
Complete amino-acid sequence of the construct produced	MSPILGYWKIKGLVQPTRLLEYLEEKYEEHLYERDEGDKW-RNKKFELGLEFPNLPYYIDGDVKLTQSMAIRYIADKHNMLGGCPKERAIEISMLEGAVLDIRYGVSRVIAYSKDFETLKVD-FLSKLPEMLKMFEDRLCHKTYLNGDHSVTHPDFMLYDAL-DVVLYMDPMCLDAFPKLVCFKKRIEAIPOIDKYLKSSKYI-AWPLQGWQATFGGGDHPKSDGTPITSLYKKAFFENLYL-FQGHMKVFCGRANPTTGSLEWLEEDHYDYHQEIARSSY-ADMLHDKDRNIKYYQGIRAAVSRVKDRGQKALVLDIGTGTGLLSMMAV TAGADFCYAIIEVFKPMAEAAVKIVERNG-FSDKIKVINKHSTEVTVGPDGDLPCRANILITELFDTELIGE-GALPSYEHAKHLVQEDCEAVPHRATVYQAQLVESRRMWSWNKLFPIHVQTSLGEQVIVPPSELERCPGAPSVCDIQLN-QVSPADFTVLSVDLPMFSDVDFSKQVSSAAACHSRRFELTSGRAQVVLSSWWDIEMDPEGKIKCTMAPFWAHSDDPEEMQWRDHWMQCVYFLPQEEPVOGSPRCLVAHDDYCVWYSLQRTSPDENDSAYQVRPVCDCQAHLLWNRPRFGEINDQDRDTHYAQALRTVLLPGSVCLCVSDGSLLSMLAHHLGAEQVFTVESSVASYRLMKRIFKVNHLEDKISVINKRPELLTAA-DLEGKKVSLLLGEPFFTTSLLPWHNLYFWYRTSVDQHLAPGAVVMPQAASLHAVIVEFRDLWRIRSPCGDCEGFDVHIMDDMIKHSDFRESREAEHPHLWEYPCRSLSKPOEILTFDFQQPIPOQPMQSKGTMELTRPGKSHGAVLWMEYQLTPDSTISTGLINPAEDKGDCCWNPCKQAVYFLSTLTLDRVPLNPRSVSYVVEFHPLTGDITMEFRLADTLS	MSPILGYWKIKGLVQPTRLLEYLEEKYEEHLYERDEGDKW-RNKKFELGLEFPNLPYYIDGDVKLTQSMAIRYIADKHNMLGGCPKERAIEISMLEGAVLDIRYGVSRVIAYSKDFETLKVD-FLSKLPEMLKMFEDRLCHKTYLNGDHSVTHPDFMLYDAL-DVVLYMDPMCLDAFPKLVCFKKRIEAIPOIDKYLKSSKYI-AWPLQGWQATFGGGDHPKSDGTPITSLYKKAFFENLYL-FQGHMKVFCGRANPTTGSLEWLEEDHYDYHQEIARSSYADMLHDKDRNIKYYQGIRAAVSRVKDRGQKALVLDIGTGTGLLSMMAV TAGADFCYAIIEVFKPMAEAAVKIVERNG-FSDKIKVINKHSTEVTVGPDGDLPCRANILITELFDTELIGE-GALPSYEHAKHLVQEDCEAVPHRATVYQAQLVESRRMWSWNKLFPIHVQTSLGEQVIVPPSELERCPGAPSVCDIQLN-QVSPADFTVLSVDLPMFSDVDFSKQVSSAAACHSRRFELTSGRAQVVLSSWWDIEMDPEGKIKCTMAPFWAHSDDPEEMQWRDHWMQCVYFLPQEEPVOGSPRCLVAHDDYCVWYSLQRTSPDENDSAYQVRPVCDCQAHLLWNRPRFGEINDQDRDTHYAQALRTVLLPGSVCLCVSDGSLLSMLAHHLGAEQVFTVESSVASYRLMKRIFKVNHLEDKISVINKRPELLTAA-DLEGKKVSLLLGEPFFTTSLLPWHNLYFWYRTSVDQHLAPGAVVMPQAASLHAVIVEFRDLWRIRSPCGDCEGFDVHIMDDMIKHSDFRESREAEHPHLWEYPCRSLSKPOEILTFDFQQPIPOQPMQSKGTMELTRPGKSHGAVLWMEYQLTPDSTISTGLINPAEDKGDCCWNPCKQAVYFLSTLTLDRVPLNPRSVSYVVEFHPLTGDITMEFRLADTLP

*al.*, 2011). PRMT7 also plays a role in male germline imprinted gene methylation through its interaction with CTCFL/BORIS (Jelinic *et al.*, 2006).

PRMT7 contains 692 amino acids (and in human) and is unusual among PRMTs in that two PRMT ‘catalytic’ domains are present in tandem, possibly as a result of gene duplication (Miranda *et al.*, 2004). The PRMTs that catalyse the formation of DMA are active as dimers, an oligomeric state that is illustrated by all previously solved crystal structures (Weiss *et al.*, 2000; Zhang & Cheng, 2003; Troffer-Charlier *et al.*, 2007; Yue *et al.*, 2007; Cheng *et al.*, 2011; Sun *et al.*, 2011; Antonysamy *et al.*, 2012). As dimerization has shown to be required for the methylation function of many PRMTs, the PRMT7 structure may reveal molecular clues to understanding the evolution of the PRMT family.

PRMT7 contains two catalytic domains in a single polypeptide chain, which raises questions about its structural organization and oligomerization. However, the sequence of the second PRMT7 catalytic domain is poorly conserved, including residues within the four conserved motifs, and it may therefore not be functional (Miranda *et al.*, 2004; Krause *et al.*, 2007). The correlation between the unusual architecture of PRMT7 and its MMA-formation activity also remains elusive. In the present study, we have solved *de novo* the structure of full-length PRMT7 from *Mus musculus* (MmPRMT7) in two different crystal forms, starting from five eukaryotic PRMT7s. It illustrates the need to screen several homologues of the target protein in parallel and different expression systems to increase the chance of obtaining crystals that diffract to high resolution.

## 2. Materials and methods

### 2.1. Cloning strategy

The *prmt7* genes from five different organisms available in various DNA banks were ordered, namely *Homo sapiens* (Addgene 34693; HsPRMT7), *Mus musculus* (PlasmID MMCD00312554; MmPRMT7), *Danio rerio* (Source Bioscience LLAM9106-008; DrPRMT7),

*Xenopus laevis* (Source Bioscience LLAM10930-L24; XIPRMT7) and *Arabidopsis thaliana* (RIKEN GSC P1997; AtPRMT7). The respective genes were amplified by PCR using primers with restriction sites at their 5’ end for subcloning in different expression vectors (Table 1 and Supporting Information<sup>1</sup>). The target vectors were pNEA-vH, pNEA-vH (Diebold *et al.*, 2011) and PGEX-NB (derived in-house from pEG-4T-1 with a different MCS) for expression in *Escherichia coli* and pDONR207 (Invitrogen) or pCR-Blunt (Life Technologies) for subcloning. The restriction sites were *NdeI* at the 5’ end of the gene and *BglII* or *BamHI* at the 3’ end, depending on the presence of internal restriction. Genes cloned in the pDONR207 vector were subcloned in pBacGGWH or pBacHGWH (Abdulrahman *et al.*, 2009) using the Gateway cloning technology (Life Technologies).

### 2.2. Cloning procedure

The PCR reactions (50 µl) consisted of Phusion HF buffer, 1.5 mM MgCl<sub>2</sub>, 200 µM of each dNTP, 0.5 µM of each primer, 1 ng template DNA and 1 U Phusion DNA polymerase (NEB). The PCR program consisted of an initial denaturation at 98°C for 30 s, followed by 30 cycles of 10 s denaturation at 98°C, 10–30 s hybridization at primer *T<sub>m</sub>* + 3°C and 30 s per kilobase extension at 72°C, and was completed by a 10 min final extension at 72°C. After analysis on 0.8% agarose gel, the amplified genes were purified using a NucleoSpin Gel and PCR Clean-up kit (Machery-Nagel). The purified DNA was then double-digested using 1 U of the corresponding restriction enzymes in the recommended NEB buffer and purified again with the same kit, except for the PCR products designed for direct insertion into the pCR-Blunt vector. The target vectors (except for pCR-Blunt) were digested using the same restriction enzymes as the PCR products, but with the addition of alkaline phosphatase (Takara) to prevent the recircularization of the vector in the absence of an insert. Usually, 3–5 µg vector was digested, providing sufficient material for several

<sup>1</sup> Supporting information has been deposited in the IUCr electronic archive (Reference: FW5438).

dozen ligation reactions. Ligation was performed for 16 h at 16°C in 10 µl with 50 ng vector and sufficient PCR fragment to exceed a 5:1 fragment:vector molar ratio, 1 U T4 DNA ligase (Thermo Scientific), 50 mM Tris-HCl pH 7.5, 10 mM MgCl<sub>2</sub>, 1 mM ATP and 1 mM DTT. Ligation products were transformed in *E. coli* DH5α CaCl<sub>2</sub> competent cells and spread on LB-agar plates supplemented with suitable antibiotics. Colonies were screened by colony PCR using *Taq* DNA polymerase (produced in-house) and positive clones were confirmed by sequencing (GATC Biotech).

Recombinant baculoviruses were obtained using the Bac-to-Bac Baculovirus Expression System (Invitrogen). DH10Bac competent cells containing the baculovirus genome were transformed with the pBacGGWH and pBacHGWH plasmids and plated onto LB agar medium containing 15 mg ml<sup>-1</sup> tetracycline, 7 mg ml<sup>-1</sup> gentamicin, 50 mg ml<sup>-1</sup> kanamycin, 25 mg ml<sup>-1</sup> X-Gal and 40 mg ml<sup>-1</sup> IPTG. Bacmid DNA purified from recombination-positive white colonies was transfected into Sf9 cells using the Lipofectin reagent (Invitrogen). Viruses were harvested 10 d after transfection.

### 2.3. Expression tests

Small-scale expression in *E. coli* was tested using the BL21(DE3) and BL21(DE3)pRARE2 strains. For each test, 5 ml LB medium supplemented with the proper antibiotic and inoculated with a clone was grown at 37°C until the OD<sub>600</sub> reached 0.5–0.6. Expression of the target protein was induced with 0.5 mM IPTG for either 4 h at 37°C or 16 h at 20°C. After harvesting, cell lysis was performed in 1.5 ml buffer A (100 mM Tris-HCl pH 7.5, 400 mM NaCl, 5% glycerol, 14 mM β-mercaptoethanol, 20 mM imidazole) and the cellular debris was sedimented by centrifugation of the lysate at 40 000g for 20 min. The supernatant was transferred into a 96 deep-well plate (Greiner) and incubated for 30 min at 4°C with 20 µl of either Ni-NTA (Qiagen) or Glutathione Sepharose (GE Healthcare) medium. After a brief centrifugation, the supernatant was discarded and the beads were washed three times with 500 µl buffer A using a Te-MO pipetting robot (Tecan). The beads were resuspended in 25 µl Laemmli buffer, boiled and analysed by SDS-PAGE.

Small-scale expression in Sf9 cells was performed in six-well plates (Falcon) at 27°C. 1.5 million cells were infected with 150 µl of the transfection supernatant. The cells were harvested 48 h post-infection. Lysis was performed by sonication in buffer B [50 mM Tris-HCl pH 8.0, 250 mM NaCl, 5% glycerol (and 10 mM DTT for GST fusion constructs)] and cellular debris was sedimented by centrifugation of the lysate at 40 000g for 30 min. The beads were resuspended in 25 µl Laemmli buffer, boiled and analysed by SDS-PAGE.

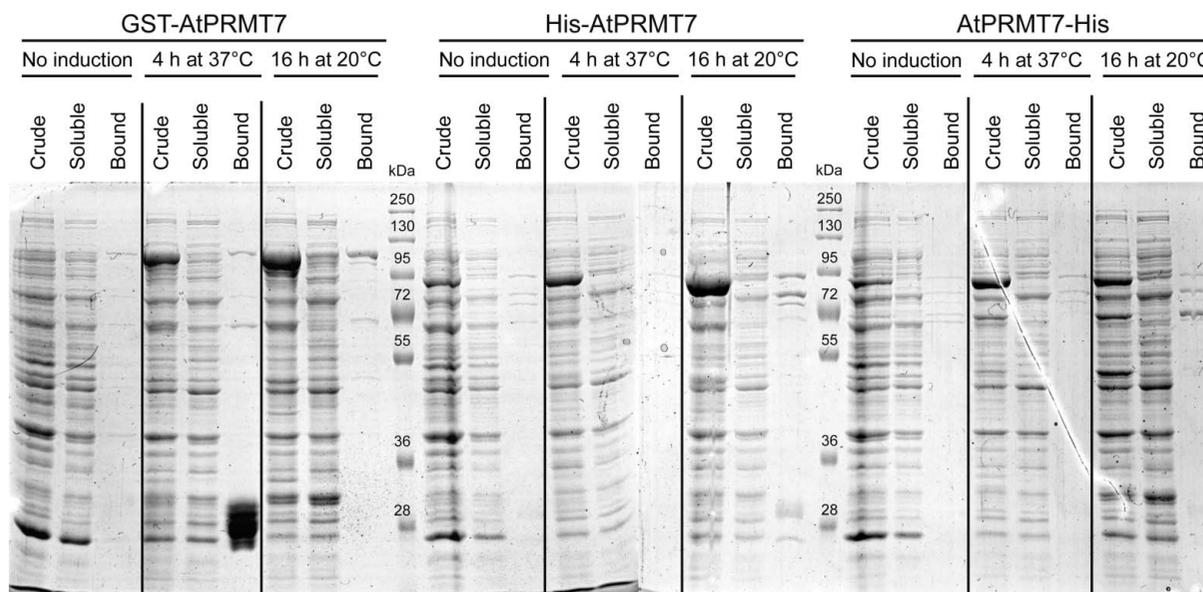
### 2.4. Expression and purification

Sf21 cells were cultured at 27°C in suspension in SF900II medium (Life Technologies) in Erlenmeyer flasks. 2 l of Sf21 cell culture (at 0.8 × 10<sup>6</sup> cells ml<sup>-1</sup>) was infected with recombinant GST-MmPRMT7 or HsPRMT7 virus with an infection multiplicity of 1. The cells were harvested 48 h post-infection. Lysis of the cells was performed in 100 ml buffer C (50 mM Tris-HCl pH 8.0, 250 mM NaCl, 5% glycerol, 10 mM DTT, 0.01% NP-40 and antiproteases) and cellular debris was sedimented by centrifugation of the lysate at 40 000g for 30 min. The supernatant was incubated at 4°C for 3 h with 3 ml Glutathione Sepharose resin (GE Healthcare). After a brief centrifugation, the supernatant was discarded and the beads were poured into an Econo-column (Bio-Rad). After two washing steps with 10 ml buffer C, 2 ml buffer C supplemented with in-house-produced TEV protease was applied onto the columns and digestion was performed overnight at 22°C with gentle mixing. The digests were concentrated with an Amicon Ultra 10K (Millipore), loaded onto a gel-filtration column (HiLoad 16/60 Superdex S200, GE Healthcare) and eluted at 1 ml min<sup>-1</sup> with buffer D (20 mM Tris-HCl pH 8.0, 150 mM NaCl, 5 mM TCEP) using an ÄKTAexplorer device (GE Healthcare). Fractions containing PRMT7 were pooled and concentrated.

### 2.5. Biophysical assays

Purified PRMT7 was analysed by dynamic light scattering in a DynaPro NanoStar device using the *Dynamics* 7.1.7 software (Wyatt Technology).

Thermal shift assays were performed with enzymes alone or in the presence of small-molecule ligands. 2 µl protein sample at 3 mg ml<sup>-1</sup>



**Figure 1**

*A. thaliana* PRMT7 expression check in *E. coli*. 10% SDS-PAGE stained with Coomassie Brilliant Blue of samples of three different AtPRMT7 constructions from small-scale expression-check assays.

was first diluted in 18  $\mu\text{l}$  buffer *E* (20 mM Tris–HCl pH 8.0, 100 mM NaCl) before the addition of 3  $\mu\text{l}$  additive (10 mM SFG, 10 mM SAH or 500 mM Arg) and 7  $\mu\text{l}$  SYPRO Orange (Invitrogen). The samples were gradually heated to thermal denaturation in a MiniOpticon Real-time PCR System (Bio-Rad) from 20 to 95°C at 0.05°C s<sup>-1</sup>. Protein unfolding was followed by fluorescence intensity (excitation wavelength 492 nm, emission wavelength 586 nm), which was plotted as a function of temperature. *Opticon Monitor* was used to estimate the melting temperature ( $T_m$ ).

## 2.6. Crystallization

Crystallization conditions for HsPRMT7 and MmPRMT7 were screened using commercially available kits by the sitting-drop vapour-diffusion method in 96-well MRC2 plates (Swissci) employing a Mosquito robot (TTP LabTech). The plates were placed in a Rock Imager (Formulatrix) at 20°C and monitored periodically. 15 screens were tested for HsPRMT7 at a concentration of 9–16 mg ml<sup>-1</sup> alone or in the presence of 1 mM SFG or 1 mM SAH, but no initial hits were obtained. An initial hit was obtained for MmPRMT7 at 16 mg ml<sup>-1</sup> in the presence of 1 mM SFG using the The Classics Suite (Qiagen) condition No. 96 [100 mM MES pH 6.5, 12% (w/v) PEG 20 000]. The condition was refined using the hanging-drop vapour-diffusion method in a 24-well XRL plate (Molecular Dimensions) using 22 mm siliconized glass cover slides (Hampton Research) and stored at 17°C. The crystals used for X-ray diffraction were obtained using an MmPRMT7 sample at 3.7 mg ml<sup>-1</sup> incubated for 20 min at room temperature with 1 mM SAH and 100 mM NDSB-256 (Sigma). Hanging drops were set up at 17°C by mixing 1  $\mu\text{l}$  protein solution and 1  $\mu\text{l}$  reservoir solution consisting of 100 mM bis-tris propane (BTP) pH 7.0, 3.2% (w/v) PEG 20 000, 12.8% (w/v) PEG 8000. Single diffracting crystals grew to 100  $\times$  100  $\times$  50  $\mu\text{m}$  within 10 d. They were cryoprotected by a short soak in 3% (w/v) PEG 20 000, 13% (w/v) PEG 8000, 100 mM BTP pH 7.0, 15% (v/v) PEG 400 and flash-cooled in liquid nitrogen for data collection. Platelet-shaped crystals of an orthorhombic form were also obtained using the method described above in 100 mM MES pH 6.5, 10–18% (w/v) PEG 20 000. They were

cryoprotected using 100 mM bis-tris pH 7.0, 15% (w/v) PEG 20 000, 15% (v/v) ethylene glycol.

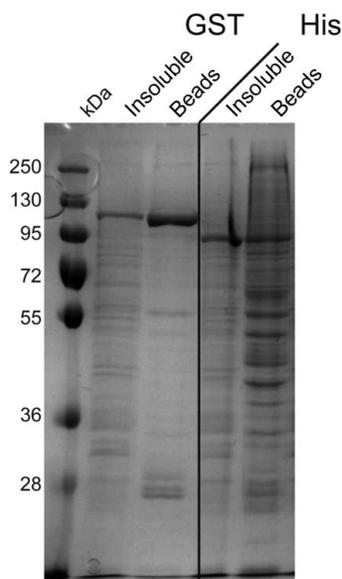
## 2.7. Data collection and processing

Data sets were collected on the PROXIMA1 beamline at SOLEIL and the ID29 beamline at ESRF using a Pilatus 6M detector (Dectris) at -173°C and were processed with *XDS* (Kabsch, 2010) and *HKL-2000* (Otwinowski & Minor, 1997). The first structure was solved using the single anomalous dispersion (SAD) method using a crystal soaked for 5 h in 5 mM thimerosal. Mercury-site determination and solvent flattening were performed using *AutoSol* in *PHENIX*. A model was built in the 2.1 Å resolution SAD map using *ARP/wARP* in the *CCP4* suite (Winn *et al.*, 2011) and *Coot* (Emsley *et al.*, 2010), and was used for phasing of the 2.65 Å resolution native data set by molecular replacement. The structure was refined with *PHENIX* (Adams *et al.*, 2010) and *BUSTER* (Bricogne *et al.*, 2011).

## 3. Results and discussion

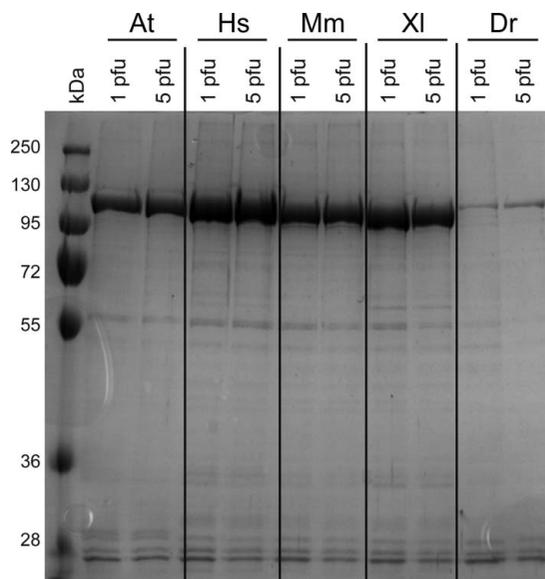
### 3.1. Poor expression in bacteria

For functional characterization, human PRMT7 (HsPRMT7) was expressed as a GST fusion in *E. coli* BL21(DE3) cells and purified (Zurita-Lopez *et al.*, 2012). However, the yield was insufficient for structural studies. We therefore decided to try several PRMT7 homologues to improve our chances of finding a protein that would be highly expressed, soluble and prone to crystallize. Five PRMT7 genes were commercially available in various gene banks and were ordered. Each gene was subcloned in three *E. coli* expression vectors to produce recombinant protein with a GST tag (PGEX-NB) or an N-terminal (pnEAvH) or a C-terminal (pnEA-vH) hexahistidine tag. An internal TEV cleavage site was also present in each construct to allow removal of the tag. All constructs could be overexpressed in *E. coli* BL21(DE3) cells, as illustrated for AtPRMT7, which is clearly visible in the crude extracts of the overnight-induced constructs (Fig. 1). However, the recombinant proteins were poorly soluble and only a minor fraction could be recovered by affinity purification. It



**Figure 2**

*A. thaliana* PRMT7 transfection test. 10% SDS-PAGE stained with Coomassie Brilliant Blue of AtPRMT7 samples from two different constructions.



**Figure 3**

PRMT7 amplification test in Sf21 cells. 10% SDS-PAGE stained with Coomassie Brilliant Blue of GST-PRMT7 samples from five different species. Pfu, plaque-forming unit.

was thus decided to switch to insect-cell expression. For this purpose, the five PRMT7 genes were subcloned in pBac vectors for expression as recombinant proteins with a hexahistidine or GST N-terminal tag and a TEV cleavage site.

### 3.2. Purification and quality control

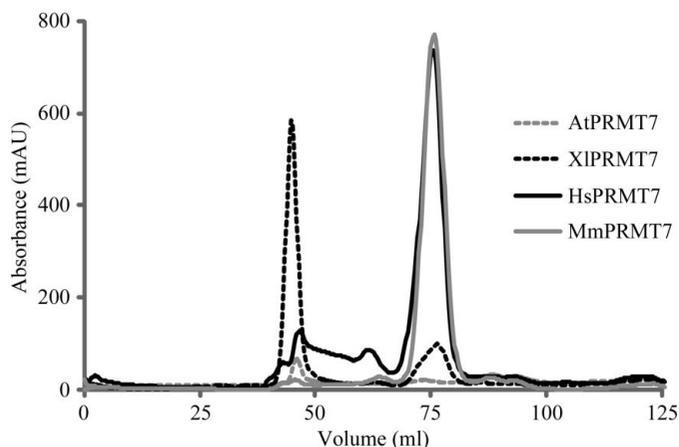
Transfection tests were performed on the ten new constructs. In all cases the GST-fusion construct gave a much larger amount of soluble protein than the hexahistidine-tagged construct, as illustrated for AtPRMT7 (Fig. 2). Therefore, amplification was only tested for the GST-tagged constructs. Among the five constructs, only PRMT7 from *D. rerio* was poorly amplified and was not further produced (Fig. 3). The four remaining recombinant PRMT7s were purified by affinity chromatography followed by removal of the GST tag and a final gel-filtration step (Fig. 4). AtPRMT7 and XI-PRMT7 were poorly produced and eluted in the void volume, suggesting aggregation of the proteins. In contrast, HsPRMT7 and MmPRMT7 were highly produced (around 3 mg per litre of medium) and eluted as a single peak at a volume compatible with a monomeric state. The homo-

**Table 2**  
Dynamic light scattering.

	Diffusion coefficient ( $\text{cm s}^{-1}$ )	Radius (nm)	Pd (nm)	%Pd	%Mass
HsPRMT7	$5.174 \times 10^{-7}$	4.657	1.346	28.9	100
MmPRMT7	$4.955 \times 10^{-7}$	4.862	1.163	23.9	99.9

**Table 3**  
Crystallization.

Method	Sitting-drop vapour diffusion	Hanging-drop vapour diffusion
Plate type	96-well MRC2	24-well XRL
Temperature ( $^{\circ}\text{C}$ )	20	17
Protein concentration ( $\text{mg ml}^{-1}$ )	16	3.7
Buffer composition of protein solution	20 mM Tris-HCl pH 8.0, 150 mM NaCl, 4 mM TCEP	20 mM Tris-HCl pH 8.0, 150 mM NaCl, 4 mM TCEP
Composition of reservoir solution	100 mM bis-tris propane pH 7.0, 3.2% (w/v) PEG 20 000, 12.8% (w/v) PEG 8000	100 mM MES pH 6.5, 10–18% (w/v) PEG 20 000
Volume and ratio of drop	200 nl (1:1 protein:reservoir)	2 $\mu\text{l}$ (1:1 protein:reservoir)
Volume of reservoir ( $\mu\text{l}$ )	50	500

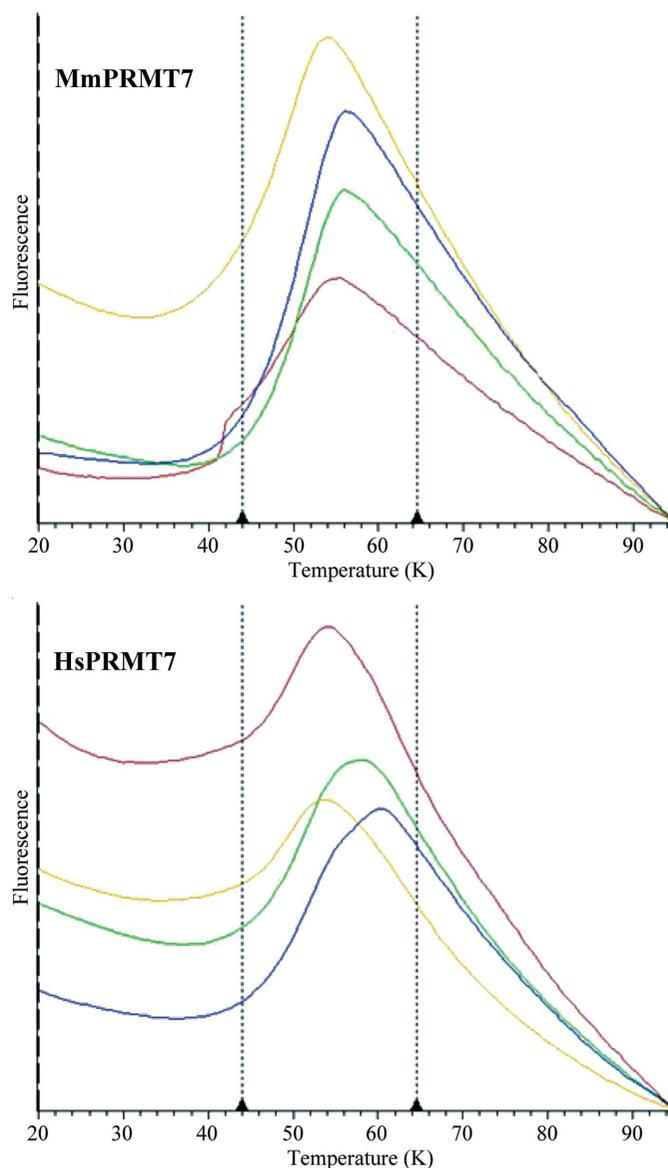


**Figure 4**  
Gel filtration and oligomeric state of PRMT7. Chromatograms of the absorbance at 280 nm during size-exclusion chromatography of the four PRMT7 proteins produced in Sf21 cells after GST-tag removal.

geneity of the latter two PRMT7s was measured by dynamic light scattering (Table 2). Both appeared to be monomeric in solution and monodisperse, with a polydispersity index below 30%, which is a good indicator of crystallizability (Ferré-D'Amaré & Burley, 1994). The thermostability of the two proteins was also measured (Fig. 5) and was  $50^{\circ}\text{C}$  for HsPRMT7 and  $49.5^{\circ}\text{C}$  for MmPRMT7, which support their potential to crystallize (Dupeux *et al.*, 2011). The presence of 1 mM SAH or SFG cofactor mimics increased the  $T_m$  by  $1.5^{\circ}\text{C}$ , but arginine, even at 50 mM, had no measurable effect.

### 3.3. Crystallization

Crystallization assays were set up for HsPRMT7 and MmPRMT7. Whereas MmPRMT7 gave several initial hits, no favourable conditions could be identified for HsPRMT7, even after extensive screening with either sparse-matrix or systematic screens. The two proteins share 589 identical residues out of 692 (85%). Once the structure has been determined, it would be interesting to map the



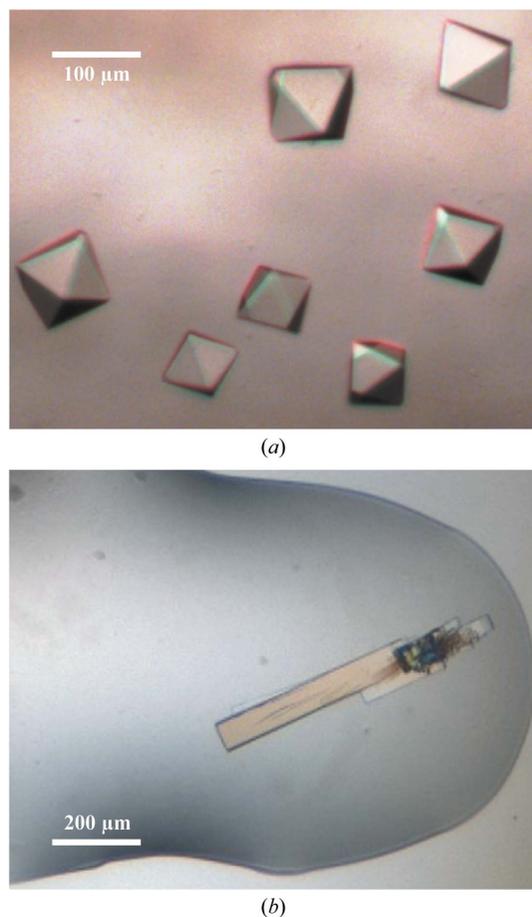
**Figure 5**  
Thermal shift assay of MmPRMT7 and HsPRMT7. The plots are coloured red for the apo form, green in the presence of SAH, blue in the presence of SFG and yellow in the presence of arginine.

differences and to attempt to rationalize the inability to crystallize HsPRMT7. Although hits were quickly identified for MmPRMT7, their reproducibility was poor. Several crystallization temperatures (4, 17 and 20°C) and cofactor mimics were tested, as was chemical lysine modification of the protein. Sulfobetaine was identified as an additive that significantly improved the reproducibility. Surprisingly, two different crystal forms appeared in the same crystallization conditions and even in the same drop, mainly at high protein concentrations (Fig. 6 and Table 3). For both crystal forms, cryoprotectant mixtures were prepared by adding minimal amounts of PEG 400 or ethylene glycol to the reservoir compositions and were tested to check the absence of ice rings on exposure to X-rays. Finally, solutions containing 15% (v/v) ethylene glycol or PEG 400 were found to be suitable to preserve crystal diffraction at  $-173^{\circ}\text{C}$ .

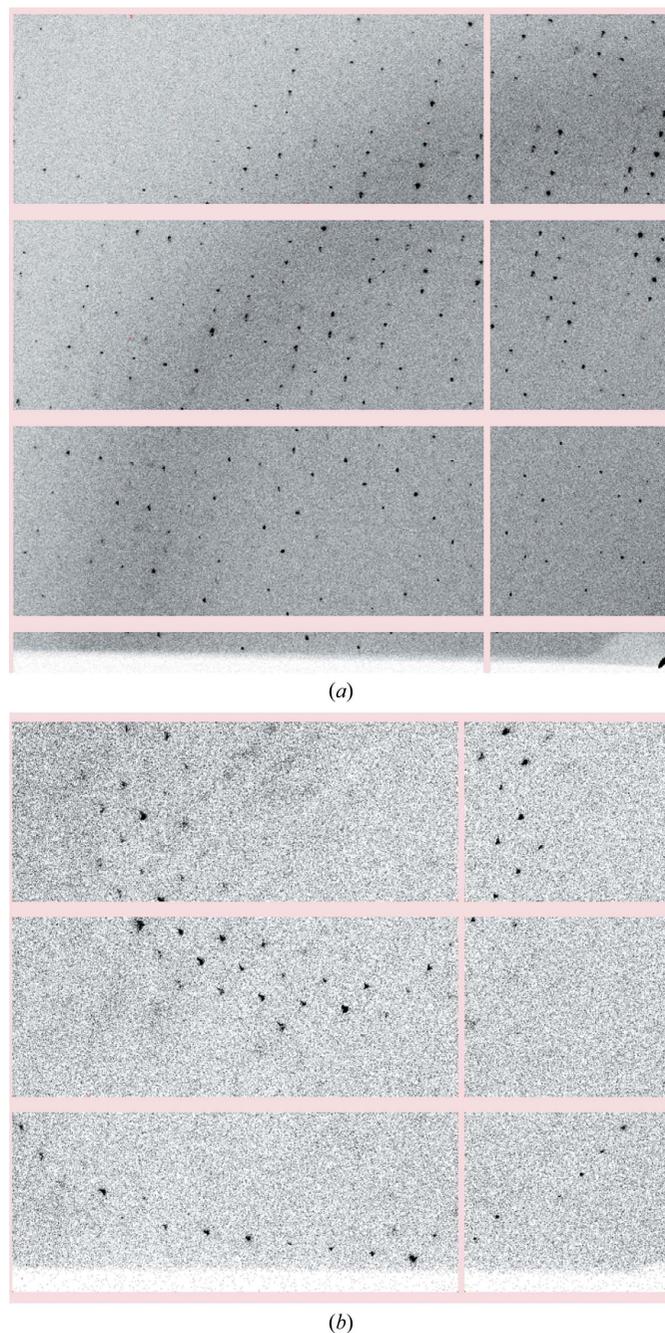
### 3.4. Structure determination

The first crystals obtained were tested using an in-house X-ray source. The diffraction limit increased with the quality of the crystals. The best diffraction patterns were collected from crystal form II at the SOLEIL synchrotron facility (Fig. 7). The crystals belonged to space group  $P4_32_12$ , with unit-cell parameters  $a = b = 97.41$ ,  $c = 168.09$  Å. Crystal form I belonged to space group  $I222$ , with unit-cell parameters  $a = 94.2$ ,  $b = 104.7$ ,  $c = 168.7$  Å. Data-collection and refinement statistics are given in Table 4. However, despite extensive trials, the native data sets collected could not be solved by molecular replacement. Several models based on known PRMT structures were

built for this purpose, with loop truncations, polyalanine mutations or removal of subdomains, but without success. A *BALBES* (Long *et al.*, 2008) run also failed to isolate a solution. Several crystals were therefore soaked with mercury derivatives (thimerosal and *para*-chloromercuribenzoic acid). A complete data set could be collected at a wavelength of 1.00637 Å from one of these crystals using an inverse-beam strategy, which allowed determination of the structure using the SAD method. A single monomer of MmPRMT7 is present in the asymmetric unit. Together with this data set, a higher resolution native data set was collected and was used to refine the MmPRMT7



**Figure 6**  
Crystals of MmPRMT7 belonging to space groups  $P4_32_12$  (a) and  $I222$  (b). Their size is shown by the white scale bar on the left.



**Figure 7**  
Excerpts of the diffraction patterns of MmPRMT7 crystals. The beam centre is in the lower right corner for both the  $P4_32_12$  (a) and  $I222$  (b) diffraction frames. The vertical and horizontal white lines correspond to the blind areas of the Pilatus detector. Owing to the small size of the spots, the full frame cannot be displayed as a reduced picture.

**Table 4**

Data collection and processing.

	Form I	Form II
Diffraction source	ID29, ESRF	PROXIMA1, SOLEIL
Wavelength (Å)	0.97625	1.00637
Temperature (°C)	−173	−173
Detector	PILATUS 6M-F	PILATUS 6M
Crystal-to-detector distance (mm)	548	332
Rotation range per image (°)	0.1	0.2
Total rotation range (°)	210	90 + 90 (inverse beam)
Exposure time per image (s)	0.04	0.2
Space group	<i>I</i> 222	<i>P</i> <sub>4</sub> <sub>3</sub> <i>2</i> <sub>1</sub> <i>2</i>
Unit-cell parameters (Å)	<i>a</i> = 94.2, <i>b</i> = 104.7, <i>c</i> = 168.7	<i>a</i> = <i>b</i> = 97.4, <i>c</i> = 168.1
Mosaicity (°)	0.55	0.43
Resolution range (Å)	50.0–2.65 (2.81–2.65)	50–2.05 (2.12–2.05)
Total No. of reflections	323775 (51629)	661421 (63140)†
No. of unique reflections	24609 (3873)	96985 (9715)†
Completeness (%)	99.6 (98.4)	99.8 (99.8)
Multiplicity	13.1 (13.3)	6.8 (6.5)
$\langle I/\sigma(I) \rangle$ ‡	17.92 (1.01)	22.6 (2.9)
<i>R</i> <sub>meas</sub> § (%)	9.7 (231.6)	11.8 (68.8)†
Overall <i>B</i> factor from Wilson plot (Å <sup>2</sup> )	84.2	32.0

† Friedel pairs have not been merged for the form II crystal (tetragonal space group). ‡  $\langle I/\sigma(I) \rangle$  falls below 2.0 in the outer shell at 2.95 Å for the form I crystal (orthorhombic space group). § *R*<sub>meas</sub> is the redundancy-independent *R* factor (Diederichs & Karplus, 1997).

structure. The details revealed by this new PRMT structure will be described elsewhere.

This work was supported by institutional support from CNRS, Université de Strasbourg, Inserm, Instruct, part of the European Strategy Forum on Research Infrastructures (ESFRI) supported by national member subscriptions as well as the French Infrastructure for Integrated Structural Biology (FRISBI; ANR-10-INSB-05-01) and grants from Agence Nationale de la Recherche (ANR-08-PCVI-0037-01), Association pour la Recherche contre le Cancer (ARC; Nos. A09/4/5005 and SFI20121205902). The authors are indebted to the baculovirus common resources of the IGBMC for recombinant virus production and thank the members of the IGBMC common services and the members of the structural biology platform of IGBMC for technical assistance. The authors thank members of the SOLEIL PROXIMA1 beamlines and the European Synchrotron Radiation Facility–European Molecular Biology Laboratory Joint Structural Biology groups for the use of beamline facilities and for help during data collection.

## References

Abdulrahman, W., Uhring, M., Kolb-Cheynel, I., Garnier, J.-M., Moras, D., Rochel, N., Busso, D. & Poterszman, A. (2009). *Anal. Biochem.* **385**, 383–385.

Adams, P. D. *et al.* (2010). *Acta Cryst.* **D66**, 213–221.  
 Antonysamy, S. *et al.* (2012). *Proc. Natl Acad. Sci. USA*, **109**, 17960–17965.  
 Bedford, M. T. & Clarke, S. G. (2009). *Mol. Cell.* **33**, 1–13.  
 Bricogne, G., Blanc, E., Brandl, M., Flensburg, C., Keller, P., Paciorek, W., Roversi, P., Sharff, A., Smart, O., Vornrhein, C. & Womack, T. (2011). *BUSTER* v2.10.0. <http://www.globalphasing.com>.  
 Cha, B. & Jho, E. H. (2012). *Expert Opin. Ther. Targets*, **16**, 651–664.  
 Cheng, Y., Frazier, M., Lu, F., Cao, X. & Redinbo, M. R. (2011). *J. Mol. Biol.* **414**, 106–122.  
 Diebold, M. L., Fribourg, S., Koch, M., Metzger, T. & Romier, C. (2011). *J. Struct. Biol.* **175**, 178–188.  
 Diederichs, K. & Karplus, P. A. (1997). *Nature Struct. Biol.* **4**, 269–275.  
 Dupeux, F., Röwer, M., Seroul, G., Blot, D. & Márquez, J. A. (2011). *Acta Cryst.* **D67**, 915–919.  
 Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. (2010). *Acta Cryst.* **D66**, 486–501.  
 Ferré-D'Amaré, A. R. & Burley, S. K. (1994). *Structure*, **2**, 357–359.  
 Gouet, P., Courcelle, E., Stuart, D. I. & Métoz, F. (1999). *Bioinformatics*, **15**, 305–308.  
 Gros, L., Delaporte, C., Frey, S., Decesse, J., de Saint-Vincent, B. R., Cavarec, L., Dubart, A., Gudkov, A. V. & Jacquemin-Sablon, A. (2003). *Cancer Res.* **63**, 164–171.  
 Jelinic, P., Stehle, J. C. & Shaw, P. (2006). *PLoS Biol.* **4**, e355.  
 Jung, G. A., Shin, B. S., Jang, Y. S., Sohn, J. B., Woo, S. R., Kim, J. E., Choi, G., Lee, K.-M., Min, B. H., Lee, K.-H. & Park, G. H. (2011). *Exp. Mol. Med.* **43**, 550–560.  
 Kabsch, W. (2010). *Acta Cryst.* **D66**, 125–132.  
 Karkhanis, V., Wang, L., Tae, S., Hu, Y.-J., Imbalzano, A. N. & Sif, S. (2012). *J. Biol. Chem.* **287**, 29801–29814.  
 Katz, J. E., Dlakić, M. & Clarke, S. (2003). *Mol. Cell. Proteomics*, **2**, 525–540.  
 Krause, C. D., Yang, Z.-H., Kim, Y.-S., Lee, J.-H., Cook, J. R. & Pestka, S. (2007). *Pharmacol. Ther.* **113**, 50–87.  
 Lee, J. H., Cook, J. R., Yang, Z.-H., Mirochnitchenko, O., Gunderson, S. I., Felix, A. M., Herth, N., Hoffmann, R. & Pestka, S. (2005). *J. Biol. Chem.* **280**, 3656–3664.  
 Long, F., Vagin, A. A., Young, P. & Murshudov, G. N. (2008). *Acta Cryst.* **D64**, 125–132.  
 Migliori, V. *et al.* (2012). *Nature Struct. Mol. Biol.* **19**, 136–144.  
 Miranda, T. B., Miranda, M., Frankel, A. & Clarke, S. (2004). *J. Biol. Chem.* **279**, 22902–22907.  
 Notredame, C., Higgins, D. G. & Heringa, J. (2000). *J. Mol. Biol.* **302**, 205–217.  
 Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.  
 Sun, L., Wang, M., Lv, Z., Yang, N., Liu, Y., Bao, S., Gong, W. & Xu, R.-M. (2011). *Proc. Natl Acad. Sci. USA*, **108**, 20538–20543.  
 Troffer-Charlier, N., Cura, V., Hassenboehler, P., Moras, D. & Cavarelli, J. (2007). *EMBO J.* **26**, 4391–4401.  
 Weiss, V. H., McBride, A. E., Soriano, M. A., Filman, D. J., Silver, P. A. & Hogle, J. M. (2000). *Nature Struct. Biol.* **7**, 1165–1171.  
 Winn, M. D. *et al.* (2011). *Acta Cryst.* **D67**, 235–242.  
 Wolf, S. S. (2009). *Cell. Mol. Life Sci.* **66**, 2109–2121.  
 Yang, Y. & Bedford, M. T. (2013). *Nature Rev. Cancer*, **13**, 37–50.  
 Yue, W. W., Hassler, M., Roe, S. M., Thompson-Vale, V. & Pearl, L. H. (2007). *EMBO J.* **26**, 4402–4412.  
 Zhang, X. & Cheng, X. (2003). *Structure*, **11**, 509–520.  
 Zurita-Lopez, C. I., Sandberg, T., Kelly, R. & Clarke, S. G. (2012). *J. Biol. Chem.* **287**, 7859–7870.