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

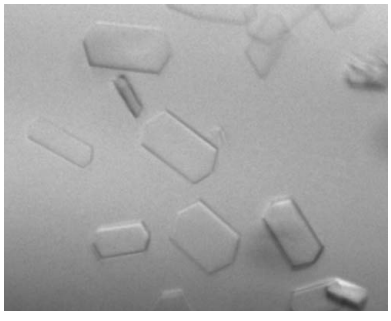
Cloning, purification and crystallization of a Walker-type *Pyrococcus abyssi* ATPase family member

Muriel Uhring, Gilbert Bey, Odile Lecompte, Jean Cavarelli, Dino Moras and Olivier Poch*

Département de Biologie et Génomiques Structurales, UMR 7104, Institut de Génétique et de Biologie Moléculaire et Cellulaire, CNRS/INSERM/ULP Strasbourg, 1 Rue Laurent Fries, 64404 Illkirch, France

Correspondence e-mail: poch@igbmc.u-strasbg.fr

Received 5 August 2005
Accepted 12 September 2005
Online 30 September 2005



Several ATPase proteins play essential roles in the initiation of chromosomal DNA replication in archaea. Walker-type ATPases are defined by their conserved Walker A and B motifs, which are associated with nucleotide binding and ATP hydrolysis. A family of 28 ATPase proteins with non-canonical Walker A sequences has been identified by a bioinformatics study of comparative genomics in *Pyrococcus* genomes. A high-throughput structural study on *P. abyssi* has been started in order to establish the structure of these proteins. 16 genes have been cloned and characterized. Six out of the seven soluble constructs were purified in *Escherichia coli* and one of them, PABY2304, has been crystallized. X-ray diffraction data were collected from selenomethionine-derivative crystals using synchrotron radiation. The crystals belong to the orthorhombic space group *C*2, with unit-cell parameters a = 79.41, b = 48.63, c = 108.77 Å, and diffract to beyond 2.6 Å resolution.

1. Introduction

In this study, we used the hyperthermophilic euryarchaeon *Pyrococcus abyssi*, the genome of which has been completely sequenced (Cohen *et al.*, 2003), as a target for high-throughput structure determination. In order to obtain new data on the puzzling role of ATPase proteins, we have started a structural study on a particular Walker-type ATPase family of *Pyrococcus abyssi.* ATPases (or NTPases) are found in all kingdoms of living organisms, where they play crucial roles in transforming chemical energy into biological events. Several ATPases play essential roles in the early steps of DNA replication, recombination and transcriptional apparatus in archaea, which are related to eukaryotes more closely than to eubacteria (for reviews, see Kelly & Brown, 2000; Lee & Bell, 2000).

The AAA+ family of proteins belongs to the Walker superfamily of A/GTPases, defined by their conserved Walker A (or P-loop) and B motifs, which mediate NTP binding and hydrolysis (Leipe *et al.*, 2002). Most AAA+ ATPases are implicated in a variety of cellular activity, including membrane fusion, proteolysis, assembly and disassembly of protein complexes, DNA replication, recombination and transcriptional regulation (Ogura & Wilkinson, 2001). The AAA (ATPases associated with diverse cellular activities) proteins form a subfamily of the AAA+ family, their distinguishing characteristic being a highly conserved motif in the ATPase domain, referred to as the SRH (for second region of homology) or AAA minimum consensus, in addition to Walker A and B motifs (Lupas & Martin, 2002). Although the proteins of the AAA+ family are all presumed to be ATPases on the basis of conserved sequence motifs, the role(s) of ATP binding and/or hydrolysis in these proteins is not always well understood and necessitates the determination of high-resolution crystal structures of AAA+ ATPase proteins.

Exhaustive bioinformatics analyses of *Pyrococcus* genomes has identified a new extended Walker-type ATPase family with a rare ATP-binding motif [G*x*R*R*x*x*GK(S,T)]. Multiple alignment analysis of these proteins leads to a division into two distinct subgroups (Lecompte *et al.*, 2001). Of 28 candidates, 16 were cloned, six were soluble in *Escherichia coli* and one of them, PABY2304, has been crystallized. The three-dimensional structure of this protein should help in understanding the molecular mechanism of these putative

archaeal ATPases. To achieve this work, we used the Structural Biology and Genomics platform (http://lbgs.u-strasbg.fr), an open inter-organisms platform ('RIO' platform) which is part of the Structural Proteomics in Europe (SPINE) network. Using a high degree of automation, the structural biology and genomics platform of Strasbourg has developed a high-throughput gene-to-structure pipeline. This pipeline is composed of three major units, cloning, purification and crystallization, in close interaction with the bio-informatics platform (http://bips.u-strasbg.fr) and *PipeAlign*, a new set of programs for high-quality protein-family analysis (Plewniak *et al.*, 2003).

Here, we described the cloning, expression, purification and crystallization of PABY2304. Crystallographic data collected from selenomethionine-derivative crystals are presented.

2. Methods and results

2.1. Construction of the expression vector

The genomic DNA of *P. abyssi* strain GE3 was used as the template for polymerase chain reaction (PCR). The PABY2304 coding gene was cloned using Gateway cloning technology (Invitrogen, The Netherlands). The primers used in cloning were 5′-GGG GAC AAG TTT GTA CAA AAA AGC AGG CTT CCT GGT GCC ACG CGG TTC TCA TAT GCT GTT TGA CCT TAG ACC CAA A-3′ containing an *Nde*I site and a thrombin-cleavage site and 5′-GGG GAC CAC TTT GTA CAA GAA AGC TGG GTC GGA TCC TCA AAT CCT TAG CAC GGT AGC AAC-3′ containing a *Bam*HI site.

The purified DNA fragment (1050 bp) was inserted into donor vector pDONR207 by BP recombination reaction (*i.e.* a recombination reaction through attB and attP recombination sites) to create entry clones. The positive clones were confirmed by colony PCR. Finally, the target gene in entry clones was transferred into the destination vector pHGWA (Busso *et al.*, 2005) *via* an LR recombination reaction (a recombination reaction through attL and attR recombination sites) to create expression clones with a histidine tag at the N-terminus of the target. This construction resulted in the mutation of the valine residue at the first position of PABY2304 to a methionine and the addition of 32 extra residues at the N-terminus (MGSSHHHHHHGTGSYITSLYKKAGFLVPRGSH) of the recombinant protein.

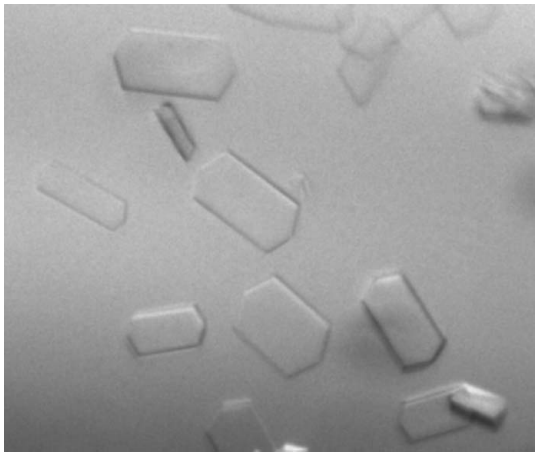


Figure 1
Crystals of selenomethionine derivative of the *PABY2304* gene product of *P. abyssi*. Typical crystal dimensions are 300 × 400 × 300 µm.

2.2. Purification of native and selenomethionyl PABY2304

The expression plasmid was transformed into *E. coli* Rosetta (DE3) strain (Novagen). This stain carries extra tRNA genes for arginine, isoleucine, proline and leucine, corresponding to codons commonly found in many archaea including the *Pyrococcus* genus but rarely used in *E. coli*. The cells were grown at 310 K in 2 l LB medium containing 100 µg ml^{-1} ampicillin and 34 µg ml^{-1} chloramphenicol. At an OD_{600} of 0.6, expression was induced with 6 m*M* isopropyl-β-D-thiogalactopyranoside (IPTG) and growth continued at 310 K for 6 h. The cells were harvested by centrifugation at 4500*g* for 15 min at 277 K and were subsequently suspended in buffer *A* (20 m*M* Tris–HCl pH 8.0 containing 250 m*M* NaCl and 2 m*M* 2-mercaptoethanol); cells were then disrupted by sonication and heated at 348 K for 15 min. The cell debris and denaturated proteins were removed by centrifugation at 18 000*g* for 30 min. The supernatant was collected and loaded onto a metal-affinity Talon (Clontech) column equilibrated with buffer *A*. The column was initially washed with 5 m*M* imidazole. The protein eluted as a clear single peak around 250 m*M* imidazole. The peak fractions were pooled and concentrated to 3 mg ml^{-1}. After this step, the sample was subjected to gel-filtration chromatography using a Superdex S75 column (Pharmacia) and a mobile phase composed of buffer *A*. The protein eluted as a single peak, followed by the excess imidazole.

Production of the selenomethionine-incorporated (SeMet) protein was carried out in the methionine auxotroph *E. coli* strain B834 (DE3) (Novagen) and protein purification proceeded as for the wild-type protein. ESI mass spectra of both native and SeMet PABY2304 were in accordance with the calculated molecular weights, indicating complete incorporation of SeMet. The purity of the SeMet protein was checked by SDS–PAGE and mass spectrometry. 12 mg pure protein was obtained from 2 l of culture as estimated by the Bradford assay (Bradford, 1976).

2.3. Crystallization and X-ray data collection

A preliminary crystallization screen was carried out by the sitting-drop vapour-diffusion method using a Tecan robot and the PEG/Ion Screen from Hampton Research. Small crystals of SeMet PABY2304 were obtained from a reservoir solution containing 100 m*M* Bis-Tris propane pH 6.5, 200 m*M* sodium sulfate and 20% PEG 3350. Larger crystals (0.3 × 0.4 × 0.3 mm; Fig. 1) were obtained after 2 d by the hanging-drop vapour-diffusion method under the same conditions. Each drop consisted of 1 µl protein solution at 3 mg ml^{-1} in 20 m*M* Tris–HCl pH 8.0, 250 m*M* NaCl, 2 m*M* 2-mercaptoethanol and 1 µl reservoir solution at 297 K.

For data collection, a single SeMet crystal was flash-cooled in a cryoprotectant solution consisting of 10%(*v*/*v*) ethylene glycol, 100 m*M* Bis-Tris propane pH 6.5, 200 m*M* sodium sulfate and 20% PEG 3350 at 100 K.

Multiple-wavelength anomalous diffraction (MAD) data were collected from this crystal to a maximum resolution of 2.6 Å using a MAR Research CCD detector on beamline BM30 at the European Synchrotron Radiation Facility (ESRF, Grenoble, France). Three wavelengths were chosen near the selenium-absorption edge based on a fluorescence absorption spectrum obtained from the frozen crystal in order to maximize the f'' component (λ_1, peak), to minimize the f' component (λ_2, inflection) and to maximize $\Delta f'$ (λ_1, remote) just before data collection. A total of 360 images of 1° rotation per image were collected for the peak and the remote wavelengths and 160 images of 1° rotation were collected for the inflection wavelength. X-ray data were indexed, processed and scaled using *HKL*2000 (Otwinowski & Minor, 1997). The crystal belonged to space group

Table 1
X-ray data-collection statistics for SeMet PABY2304.

Data were collected on beamline BM30 at ESRF, Grenoble, France. Values in parentheses are for the highest resolution shell.

Data set	Peak (λ_1)	Inflection (λ_2)	Remote (λ_3)
Wavelength (Å)	0.9796910	0.979885	0.849972
Unit-cell parameters (Å, °)	a = 79.3, b = 48.6, c = 108.6, β = 99.0	a = 79.3, b = 48.6, c = 108.6, β = 99.0	a = 79.3, b = 48.7, c = 108.8, β = 98.9
Space group	$C2$		
No. of frames (1° oscillation)	360	180	360
f'	−8.4	−11.5	−4.4
f''	7.7	3.9	3.6
Resolution	50–2.60 (2.66–2.60)	50–2.60 (2.66–2.60)	50–2.60 (2.66–2.60)
Reflections measured	86823	43495	71707
Unique reflections†	24597 (1628)	24610 (1575)	23791 (1512)
Completeness (%)	99.5 (99.9)	98.6 (99.4)	93.8 (91.7)
Redundancy	3.5 (3.4)	1.75 (1.7)	3.0 (2.7)
$I/\sigma(I)$	59.6 (15.8)	57.2 (14.3)	40.9 (8.6)
$R_{\rm merge}$‡ (%)	4.3 (12.0)	4.0 (8.6)	4.4 (14.8)

† Without merging Friedel pairs. ‡ $R_{\rm merge} = \sum |I - \langle I \rangle| / \sum I$, where I is the integrated intensity of a given reflection.

$C2$, with unit-cell parameters a = 79.41, b = 48.63, c = 108.77 Å, β = 99.02° and a corresponding unit-cell volume of 4.1×10^5 Å^3 (350 residues, 41 kDa). The asymmetric unit contains one monomer, corresponding to a $V_{\rm M}$ value of 2.55 Å^3 Da^{-1}, which is within the range observed by Matthews for protein crystals (Matthews, 1977), and a solvent content of 51.8% (assuming a partial specific volume of 0.74 cm^3 g^{-1}). Details of the data-collection statistics are presented in Table 1. The structure has been solved by classical MAD methods using *SOLVE* (Terwilliger, 2003).

We thank members of the ESRF–EMBL Joint Structural Biology Group for use of the ESRF beamline facilities and especially Dr L. Jacquamet for assistance with data collection at beamline BM30. We acknowledge the technical and scientific staff of the Structural Biology and Genomics platform and Dr H. Nierengarten for mass-spectrometry analysis. The platform was supported by funds from the Genopole Alsace–Lorraine and the European Commission as SPINE contract No. QLG2-CT-2002-00988. This work was supported by grants from the Association Française contre les Myopathies (Programme Décrypthon).

References

Bradford, M. M. (1976). *Anal. Biochem.* **72**, 248–254.
Busso, D., Delagoutte-Busso, B. & Moras, D. (2005). *Anal. Biochem.* **343**, 313–321.
Cohen, G. N., Barbe, V., Flament, D., Galperin, M., Heilig, R., Lecompte, O., Poch, O., Prieur, D., Querellou, J., Ripp, R., Thierry, J. C., Van der Oost, J., Weissenbach, J., Zivanovic, Y. & Forterre, P. (2003). *Mol. Microbiol.* **47**, 1495–1512.
Kelly, T. J. & Brown, G. W. (2000). *Annu. Rev. Biochem.* **69**, 829–880.
Lecompte, O., Ripp, R., Puzos-Barbe, V., Duprat, S., Heilig, R., Dietrich, J., Thierry, J. C. & Poch, O. (2001). *Genome Res.* **11**, 981–993.
Lee, D. G. & Bell, S. P. (2000). *Curr. Opin. Cell Biol.* **12**, 280–285.
Leipe, D. D., Wolf, Y. I., Koonin, E. V. & Aravind, L. (2002). *J. Mol. Biol.* **317**, 41–72.
Lupas, A. N. & Martin, J. (2002). *Curr. Opin. Struct. Biol.* **12**, 746–753.
Matthews, B. W. (1977). *X-ray Structure of Proteins*, edited by H. Neurath & R. L. Hill, Vol. 3, pp. 468–477. New York: Academic Press.
Ogura, T. & Wilkinson, J. (2001). *Genes Cells*, **6**, 575–597.
Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **279**, 307–326.
Plewniak, F., Bianchetti, L., Brelivet, Y., Carles, A., Chalmel, F., Lecompte, O., Mochel, T., Moulinier, L., Muller, A., Muller, J., Prigent, V., Ripp, R., Thierry, J. C., Thompson, J. D., Wicker, N. & Poch, O. (2003). *Nucleic Acids Res.* **31**, 3829–3832.
Terwilliger, T. C. (2003). *Methods Enzymol.* **374**, 22–37.