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

Crystallization of a novel metal-containing cupin from *Acidobacterium* sp. and preliminary diffraction data analysis

Andrzej Łyskowski,[a]*‡ Kerstin Steiner,[a]‡ Ivan Hajnal,[a] Georg Steinkellner,[a] Helmut Schwab[a,b] and Karl Gruber[a,c]

[a]ACIB – Austrian Centre of Industrial Biotechnology GmbH, Petersgasse 14, 8010 Graz, Austria, [b]Institute of Molecular Biotechnology, TU Graz, Petersgasse 14, 8010 Graz, Austria, and [c]Institute of Molecular Biosciences, University of Graz, Humboldtstrasse 50, 8010 Graz, Austria

‡ These authors contributed equally to this work.

Correspondence e-mail: andrzej.lyskowski@acib.at

Received 20 December 2011
Accepted 14 February 2012

Recombinant AciX9_0562 from *Acidobacterium* sp. MP5ACTX9 (UniProt ID E8WYN5) containing sequence motifs characteristic of the RmlC-type cupins superfamily and containing Pfam motif PF07883 has been successfully cloned, expressed and purified. AciX9_0562 crystallized in a number of conditions from the Morpheus protein crystallization screen. The best crystal diffracted to 2.7 Å resolution (space group $C222_1$; unit-cell parameters a = 125.29, b = 254.63, c = 82.99 Å). Structure solution was facilitated by the automated molecular-replacement pipeline *BALBES*. The initial solution was automatically rebuilt using the *PHENIX AutoBuild* wizard, with final R and R_{free} values of 0.23 and 0.26, respectively. The structure is currently undergoing manual refinement.

1. Introduction

The AciX9_0562 gene from *Acidobacterium* sp. MP5ACTX9 (UniProt ID E8WYN5) encodes a 131-amino-acid protein of uncharacterized function with a molecular weight of 14.25 kDa and a calculated isoelectric point of 5.74. A *BLAST* search showed that the protein is highly conserved among many bacteria; however, all sequences with high identity are annotated as 'hypothetical proteins'. An InterPro scan (Hunter *et al.*, 2009) for sequence motifs revealed that AciX9_0562 belongs to the RmlC-type cupins superfamily (Wilson *et al.*, 2009) and also contains Pfam motif PF07883 (Finn *et al.*, 2010), a conserved cupin 2 motif (Dunwell, 1998). Currently, the cupin superfamily comprises 50 families consisting of members with diverse functions ranging from enzymatic activities such as dioxygenases, decarboxylases, hydrolases, isomerases and epimerases to non-enzymatic functions such as auxin-binding proteins, nuclear transcription factors and seed storage proteins (Dunwell *et al.*, 2000; Agarwal *et al.*, 2009). However, no functional annotation has yet been made for AciX9_0562.

'Cupa' is the Latin term for a small barrel. All proteins that belong to the cupins adopt a barrel-like structure (Dunwell *et al.*, 2001, 2004). They can occur as monocupins or bicupins and form different oligomerization states (Dunwell *et al.*, 2004). Each cupin domain typically consists of two conserved primary-sequence motifs. Each motif corresponds to two β-strands. The motifs are separated by a less conserved region which consists of a minimum of 11 amino acids (Dunwell *et al.*, 2001).

Here, we report the successful crystallization and structure solution of AciX9_0562 from *Acidobacterium* sp. MP5ACTX9.

2. Material and methods

2.1. Plasmid construction

The synthetic gene (gene ID 322434201; see Supplementary Material[1]) encoding a putative cupin 2 conserved barrel domain protein (YP_004216413) from *Acidobacterium* sp. MP5ACTX9 was ordered codon-optimized for expression in *Escherichia coli* from GeneArt (Life Technologies). The gene was recloned into pEHisTEV

[1] Supplementary material has been deposited in the IUCr electronic archive (Reference: NJ5112).

vector *via* *Nco*I and *Hin*dIII restriction sites to introduce an N-terminal TEV-cleavable His tag (Liu & Naismith, 2009). The pEHisTEV construct was confirmed by sequencing (LGC Genomics, Berlin, Germany) and transformed into the expression host *E. coli* BL21 (DE3) Gold.

2.2. Expression and purification

E. coli BL21 (DE3) Gold cells harbouring pEHisTEV-cupin were grown in LB (lysogeny broth) medium supplemented with kanamycin (50 µg ml^{-1}) at 310 K. The expression of recombinant protein was initiated by the addition of 0.1 m*M* IPTG to the cultures on reaching an OD_{600} of ∼0.8 and cultivation was continued at 293 K for 20 h. Protein expression and localization in cell-extract fractions was monitored by SDS–PAGE. The cells were harvested, resuspended in cold buffer *A* (20 m*M* sodium phosphate buffer pH 7.4 containing 0.5 *M* NaCl and 10 m*M* imidazole) and disrupted by sonication (Branson Sonifier S-250; 6 min, 80% duty cycle, output 7). The cell lysate was centrifuged for 1 h at 50 000*g* to remove unbroken cells and insoluble material. The cell-free lysate was filtered through a 0.45 µm syringe filter and incubated with Ni Sepharose 6 Fast Flow (GE Healthcare, Uppsala, Sweden) for 20 min. The Ni Sepharose was then transferred into an empty PD-10 column. Impurities were removed by washing with 30 m*M* imidazole followed by elution of the bound protein with 300 m*M* imidazole in buffer *A*. Fractions were analysed by SDS–PAGE, pooled and concentrated using Vivaspin 20 Centrifugal Filter Units (10 kDa molecular-weight cutoff; Sartorius) and desalted on PD-10 desalting columns (GE Healthcare) into 20 m*M* Tris–HCl, 200 m*M* NaCl pH 7. The N-terminal His tag was cleaved using 100 µg TEV protease (expressed and purified as His-tagged protein in-house) per milligram of protein in 20 m*M* Tris–HCl, 200 m*M* NaCl pH 7, 1 m*M* DTT, 0.5 m*M* EDTA at 293 K for 20 h. His tag and TEV protease were removed by reloading the protein solution onto Ni Sepharose and collecting the cleaved protein in the flowthrough. The completeness of the cleavage was analysed by SDS–PAGE. The cleaved samples were concentrated and loaded onto either a Superdex 75 10/30 column for qualitative analysis of the protein size and oligomerization state or a Superdex 200 16/60 column for preparative purification. The purified protein was again concentrated using Vivaspin 20 Centrifugal Filter Units (10 kDa molecular-weight cutoff; Sartorius). For both columns 20 m*M* Tris–HCl, 200 m*M* NaCl pH 7 was used as the running buffer. Prior to crystallization, purified and concentrated samples were stored in a cold room at 279 K and were transported on ice when necessary.

2.3. Crystallization

Purified protein samples with and without His tag were subjected to screening for crystallization conditions using the Morpheus crystallization screen (Molecular Dimensions; Gorrec, 2009) and the sitting-drop method in SWISSCI 3-Well crystallization plates. Screening drops of total volume 600 nl (1:1 screening solution to protein solution at about 10 mg ml^{-1}) were dispensed using an Oryx8 Protein Crystallization Robot (Douglas Instruments Ltd) at room temperature (293 K) and the plate was stored at 290 K between inspections. Protein variants with and without His tag crystallized rapidly in several conditions. Initial crystals were observed directly after screen setup. Two conditions that produced crystals of both

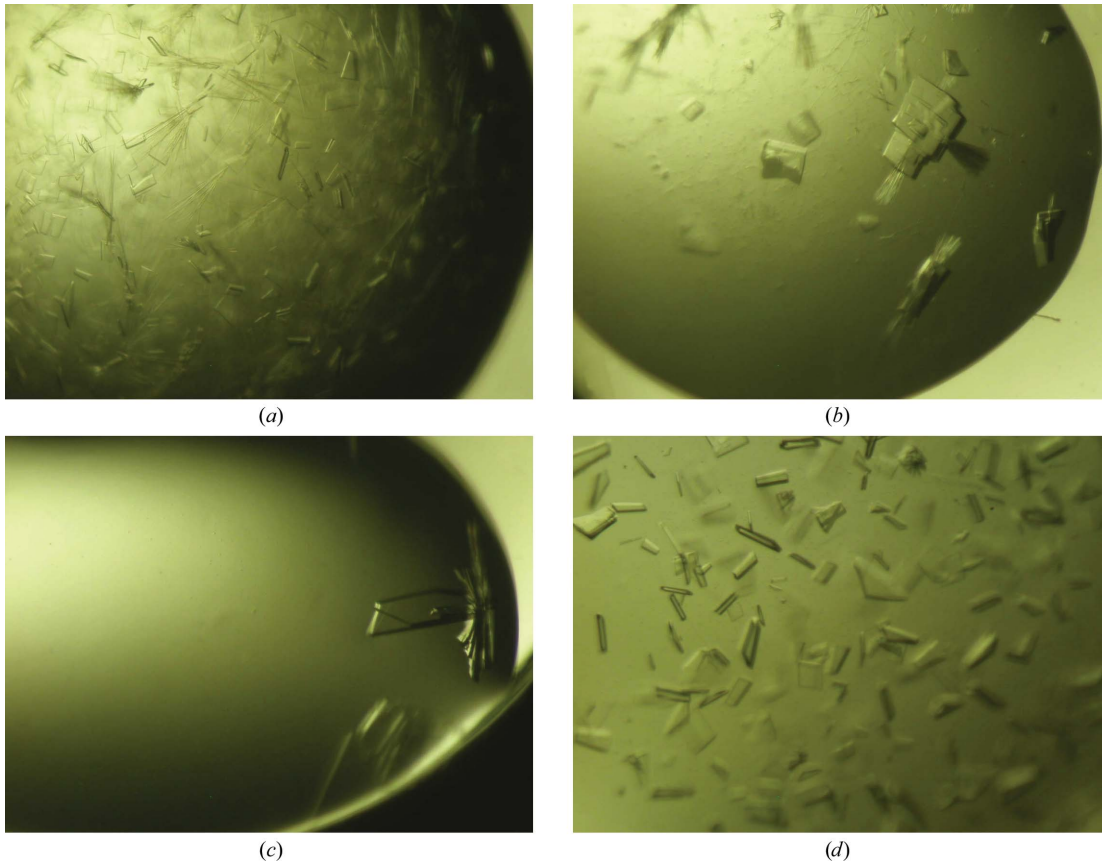

Figure 1
Crystals of AciX9_0562 from *Acidobacterium* sp. MP5ACTX9 before (*a*–*c*) and after (*d*) optimization (optimized Morpheus screen condition H02, crystal radius of between 0.05 and 0.10 mm).

Table 1
Data-processing and initial model statistics (Winn *et al.*, 2011; Evans, 2006).

Values in parentheses are for the highest resolution shell.

Data processing	
Wavelength (Å)	0.87260
Beam dimensions (µm)	8 × 5
Space group	$C222_1$
Unit-cell parameters (Å, °)	a = 125.29, b = 254.63, c = 82.99, $\alpha = \beta = \gamma$ = 90.0
Resolution limits (Å)	56.21–2.68 (2.83–2.68)
R_{merge}	0.196 (0.567)
R_{meas} (all I^+ and I^-)	0.221 (0.644)
$R_{p.i.m.}$ (all I^+ and I^-)	0.082 (0.240)
Total No. of observations	270727 (37874)
No. of unique observations	37627 (5399)
$\langle I/\sigma(I)\rangle$	8.4 (4.1)
Completeness (%)	100.0 (100.0)
Multiplicity	7.2 (7.0)
Wilson B factor (Å^2)	40.947
Model refinement	
Matthews coefficient† (Å^3 Da^{-1})	4.96
Solvent content† (%)	75.21
Resolution range (Å)	56.2–2.7
Minimum $[F_{obs}/\sigma(F_{obs})]$	1.34
Completeness for range (%)	99.97
No. of reflections	37602
Fit to data used in refinement	
R (working + test set)	0.2283
R (working set)	0.2264
R_{free}	0.2632
Free R test-set size (%)	5.03
Free R test-set count	2000

† For eight monomers in the asymmetric unit.

sample variants and that contained the visually most promising crystals were selected for further optimization: H02 [10%(*w*/*v*) PEG 8000, 20%(*v*/*v*) ethylene glycol, 0.02 *M* each of the amino acids DL-Ala, DL-Lys, DL-Ser, L-Glu and glycine, 0.1 *M* MES–imidazole buffer pH 6.5] and E05 [10%(*w*/*v*) PEG 20 000, 20%(*v*/*v*) PEG MME 550, 0.03 *M* each of the ethylene glycols diethylene glycol, pentaethylene glycol, tetraethylene glycol and triethylene glycol, 0.1 *M* MOPS/HEPES Na buffer pH 7.5]. In both conditions initial crystals were observed as early as the day after setup and reached maximum size within a one-week period.

In order to improve the initial crystals, an optimization grid was prepared. In this grid, the original drop components were diluted with Milli-Q water: the protein sample along the long plate axis and the respective precipitant solution along the short plate axis. The resulting serial dilution grid had the following compositions in the plate corners: A01, 30% protein and 50% precipitant; A12, 50% protein and 50% precipitant (original hit condition); H01, 30% protein and 30% precipitant; H12, 50% protein and 30% precipitant. Exact drop compositions were calculated using the *XStep* program of the Oryx8 Protein Crystallization Robot. Two SWISSCI 2-Well crystallization plates, one for each of the optimized conditions, were dispensed with 1000 nl total drop volumes. Each well of the plate was occupied by the protein sample variants and was equilibrated against 1 *M* NaCl solution (Newman, 2005). The best diffracting crystals for the sample without His tag were obtained from modified condition H02 with the following composition: 33.5% protein solution and 35.5% precipitant solution (Fig. 1). The crystals were vitrified without cryoprotection and the best data-set parameters are reported.

The sample with the His tag did not produce crystals of sufficient quality for structure determination.

2.4. Structure determination

The best native data were collected on beamline ID23-2 at the European Synchrotron Radiation Facility (ESRF), Grenoble, France at 100 K using a MAR 225 detector (Fig. 2). 180 images collected with a 1.0° oscillation range were processed using the program *MOSFLM*, imported into the *CCP*4 program suite using the program *POINTLESS* (Winn *et al.*, 2011; Evans, 2006) and scaled using the program *SCALA* (Evans, 2006).

The data set was submitted to the *BALBES* server for automated molecular replacement at York Structural Biology Laboratory (Long *et al.*, 2008). An initial solution with a *BALBES*-modified template based on a cupin-like protein (tm1010) from *Thermotoga maritima* (PDB entry 2f4p; Joint Center for Structural Genomics, unpublished work) contained ten copies in the asymmetric unit. It was manually inspected and the number of copies was reduced to eight owing to missing electron density for two of the ten chains. The resulting structure was then rebuilt using the *PHENIX AutoBuild* wizard (v.1.7.2_869; Adams *et al.*, 2010).

Data-processing and current model statistics are presented in Table 1.

3. Results and discussion

AciX9_0562 was successfully expressed from a pET-based vector with an N-terminal TEV-cleavable His tag with high yield in the cytoplasm of *E. coli* BL21 (DE3) Gold strain (>50% of the total protein content as judged by SDS–PAGE). The His tag was cleaved from half of the sample. Both portions were purified by gel filtration and eluted as one symmetric peak from both Superdex 75 and Superdex 200 columns. Protein preparations with and without His tag were used to set up crystallization trials. Only protein without His tag produced crystals with sufficient quality for high-resolution structure determination. X-ray diffraction data from a single crystal were collected to 2.7 Å resolution on the ID23-2 beamline at the ESRF, opening the way for structural studies and analysis.

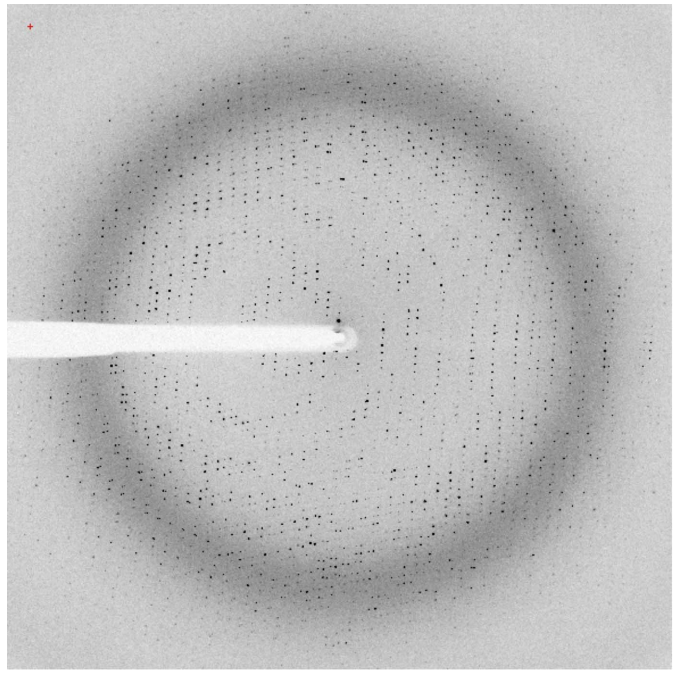

Figure 2
Diffraction image of a crystal of AciX9_0562 from *Acidobacterium* sp. MP5ACTX9 collected on the ID23-2 beamline at the ESRF, Grenoble [the image was generated with the program *ADXV* (http://www.scripps.edu/~arvai/adxv.html); the resolution at the detector edge is 2.68 Å].

Most cupins are metal-binding proteins with Mn, Fe, Ni, Cu or Zn bound in the active site. The characteristic cupin domain comprises two conserved motifs, both of which are involved in metal-ion binding. The conserved motifs comprise the following amino-acid sequences: G-X_5-H-X-H-$X_{3,4}$-E-X_6-G (motif 1) and G-X_5-P-X-G-X_2-H-X_3-N (motif 2) (Dunwell *et al.*, 2004). Two His residues and a Glu/Gln residue in motif 1 as well as the His residue in motif 2 act as ligands for binding of the active-site metal ion. At the current stage of refinement, a large unidentified electron density suitable for placement of the the hypothetical metal ion can be observed.

The biochemical and biophysical characterization as well as final and complete structure determination and refinement are currently in progress; the detailed results will be reported elsewhere.

This work was supported by the Federal Ministry of Economy, Family and Youth (BMWFJ), the Federal Ministry of Traffic, Innovation and Technology (bmvit), the Styrian Business Promotion Agency SFG, the Standortagentur Tirol and the ZIT Technology Agency of the City of Vienna through the COMET Funding Program managed by the Austrian Research Promotion Agency FFG. We acknowledge the European Synchrotron Radiation Facility for the provision of synchrotron-radiation facilities and we would like to thank the ESRF scientific and technical personnel for assistance in using beamline ID23-2.

References

Adams, P. D. *et al.* (2010). *Acta Cryst.* D**66**, 213–221.
Agarwal, G., Rajavel, M., Gopal, B. & Srinivasan, N. (2009). *PLoS One*, **4**, e5736.
Dunwell, J. M. (1998). *Biotechnol. Genet. Eng. Rev.* **15**, 1–32.
Dunwell, J. M., Culham, A., Carter, C. E., Sosa-Aguirre, C. R. & Goodenough, P. W. (2001). *Trends Biochem. Sci.* **26**, 740–746.
Dunwell, J. M., Khuri, S. & Gane, P. J. (2000). *Microbiol. Mol. Biol. Rev.* **64**, 153–179.
Dunwell, J. M., Purvis, A. & Khuri, S. (2004). *Phytochemistry*, **65**, 7–17.
Evans, P. (2006). *Acta Cryst.* D**62**, 72–82.
Finn, R. D., Mistry, J., Tate, J., Coggill, P., Heger, A., Pollington, J. E., Gavin, O. L., Gunasekaran, P., Ceric, G., Forslund, K., Holm, L., Sonnhammer, E. L., Eddy, S. R. & Bateman, A. (2010). *Nucleic Acids Res.* **38**, D211–D222.
Gorrec, F. (2009). *J. Appl. Cryst.* **42**, 1035–1042.
Hunter, S. *et al.* (2009). *Nucleic Acids Res.* **37**, D211–D215.
Liu, H. & Naismith, J. H. (2009). *Protein Expr. Purif.* **63**, 102–111.
Long, F., Vagin, A. A., Young, P. & Murshudov, G. N. (2008). *Acta Cryst.* D**64**, 125–132.
Newman, J. (2005). *Acta Cryst.* D**61**, 490–493.
Wilson, D., Pethica, R., Zhou, Y., Talbot, C., Vogel, C., Madera, M., Chothia, C. & Gough, J. (2009). *Nucleic Acids Res.* **37**, D380–D386.
Winn, M. D. (2011). *Acta Cryst.* D**67**, 235–242.