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

Overexpression, purification and crystallization of the tetrameric form of SorC sorbitol operon regulator

Daniele de Sanctis,[a] Ana T. Rêgo,[a] David Marçal,[a] Colin E. McVey,[a] Maria A. Carrondo[a] and Francisco J. Enguita[b]*

[a]Instituto de Tecnologia Química e Biológica, Universidade Nova de Lisboa, Avenida da República, Apartado 127, 2781-901 Oeiras, Portugal, and [b]Instituto de Medicina Molecular, Avenida Professor Egas Moniz, 1649-028 Lisboa, Portugal

Correspondence e-mail: fenguita@fm.ul.pt

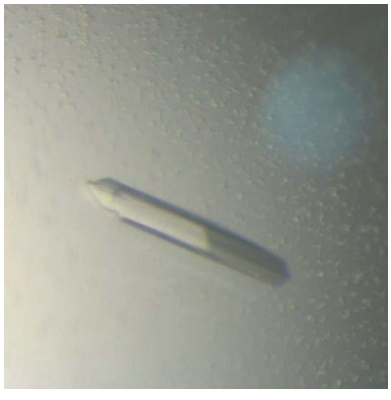
Received 1 October 2007
Accepted 19 November 2007

The sorbitol operon regulator (SorC) regulates the metabolism of L-sorbose in *Klebsiella pneumonia*. SorC was overexpressed in *Escherichia coli* and purified, and crystals were obtained of a tetrameric form. A single crystal showed X-ray diffraction to 3.20 Å. The crystal belongs to space group $P2_12_12_1$, with unit-cell parameters a = 91.6, b = 113.3, c = 184.1 Å. Analysis of the molecular-replacement solution indicates the presence of four SorC molecules in the asymmetric unit.

1. Introduction

In *Klebsiella pneumoniae* metabolism, L-sorbose is degraded *via* L-sorbose-1-phosphate and D-glucitol-6-phosphate intermediates into D-fructose-6-phosphate. The transport and phosphorylation of this ketose are catalysed by an EIIsor complex of the phosphoenol-pyruvate-dependent carbohydrate phosphotransferase system (PTS), an L-sorbose-1-P-reductase and a D-glucitol-6-P-dehydrogenase (Kelker *et al.*, 1972; Sprenger & Lengeler, 1984). The genes encoding the proteins involved in the degradation of the ketose L-sorbose are located in the *sor* operon. This operon comprises a promoter-proximal gene *sorC*, which encodes a protein of 35 kDa for which no enzymatic function has been detected. A single promoter for all *sor* genes, *sorCDFBAME*, for which SorC is the activator, appears to be located downstream of *sorC* (Wohrl *et al.*, 1990). Sequence analysis of the protein (EMBL X66059) shows that it is a member of the SorC transcriptional regulator family, which also includes DeoR (Zeng *et al.*, 2000). Members of this family are predicted to have an amino-terminal helix–turn–helix motif that binds to the DNA and a putative sugar-binding domain that is the ligand regulator-binding region. SorC from *K. pneumoniae* is regulated by L-sorbose, while other members of this family, *e.g.* DeoR, are likely to be regulated by other sugar substrates.

In the context of a detailed structural and functional characterization of a member of this class of repressors, for which no structure has been reported to date, we have produced orthorhombic crystals of SorC protein in its tetrameric form, which are described here together with their preliminary crystallographic characterization.



2. Materials and methods

2.1. Cloning and expression

The *sorC* gene (coding for sorbitol operon regulator) from *K. pneumoniae* was PCR amplified by KOD Hot Start DNA Polymerase (Novagen) using primers 5′-CACCACCACCACATGGAAAACAGTGACGATATCCGGTTGATTGTC-3′ (forward) and 5′-GAGGAGAAGGCGCGTTATTTCAGTAATAACTCAGCCGTTTCTCTGTTTG-3′ (reverse). The 1 kbp purified PCR was then inserted into pET-YSBLIC expression vector (Bonsor *et al.*, 2006) using an LIC (ligation-independent cloning) strategy. The plasmid pET-YSBLIC is derived from pET28a (Novagen) that has been adapted for LIC and adds an N-terminal hexahistidine tag to the cloned gene. The full native protein sequence is expressed fused directly to the amino-terminal sequence (MGSSHHHHHH) encoded by pET-YSBLIC.

The plasmid pET-YSBLIC-SorC was transformed into *Escherichia coli* BL21 (DE3) and grown at 310 K in Luria–Bertani medium containing kanamycin (50 µg ml^{-1}) until an optical density at 600 nm of 0.6 was reached. Cells were induced by adding 0.3 m*M* IPTG (isopropyl *β*-D-1-thiogalactopyranoside). After induction, the cells were grown at 291 K overnight and collected by centrifugation at 20 000*g*. The cell pellet was resuspended in 50 m*M* Tris–HCl pH 8.0, 500 m*M* NaCl, 1 m*M* DTT including 1 *M* NDSB (non-detergent sulfobetaine) 201 to prevent protein aggregation. Benzonase was then added to degrade nucleic acids following disruption of the cells with a French Press. The lysate was centrifuged at 20 000*g* for 20 min at 277 K. The supernatant was loaded onto a HisTrap FF nickel-affinity column (GE Healthcare). His-tagged protein was eluted with a 20–500 m*M* imidazole gradient using an ÄKTA Explorer instrument (GE Healthcare). All purification procedures were carried out at room temperature. The fractions containing SorC protein were confirmed by SDS–PAGE and were pooled and passed through a PD-10 buffer-exchange column (GE Healthcare) to remove imidazole to a final buffer containing 50 m*M* Tris pH 8.8, 0.5 *M* NaCl, 1 m*M* DTT and concentrated using a Vivapore concentrator (Sigma–Aldrich, 7500 Da molecular-weight cutoff). The protein sample was further purified by size-exclusion chromatography using a Superdex 200 (GE Healthcare) pre-equilibrated with 50 m*M* Tris pH 8.8, 0.5 *M* NaCl and 1 m*M* DTT. The purified SorC was almost homogeneous on SDS–PAGE, while gel-filtration chromatogram analysis showed that the protein was eluted in various multimeric forms.

The fractions corresponding to the tetrameric form were then concentrated to a final concentration of 4 mg ml^{-1} and stored at 253 K in gel-filtration buffer with an additional 20% glycerol. The protein concentration was determined by UV–Vis absorbance measurements with a Nanodrop ND-1000 spectrophotometer using an extinction coefficient of 41 495 M^{-1} cm^{-1}. Preliminary crystallization screens were performed by sitting-drop vapour diffusion (100 nl protein solution and 100 nl reservoir solution equilibrated against 100 µl reservoir solution) using a Cartesian Nanodrop Robot to set up 96-well plates at 294 K. Seven commercial screens were used: Classics, pH Clear I and II, Ammonium Sulfate and PEGs screens (Nextal), Index Screen (Hampton Research) and Magic Screen (BioGenova). Unique but small crystals were obtained in condition No. 93 of the Classics Screen (0.1 *M* HEPES pH 7.5, 10% PEG 6K and 5% MPD). Crystal growth was scaled up by hanging-drop vapour diffusion at 294 K in 48-well plates (1 µl protein solution and 1 µl reservoir solution equilibrated against 200 µl reservoir solution) and crystal size was optimized adding 0.1 *M* EDTA salt as an additive (Additive Screen, Hampton Research). Final crystal dimensions were 100 ×100 × 20 µm (Fig. 1).

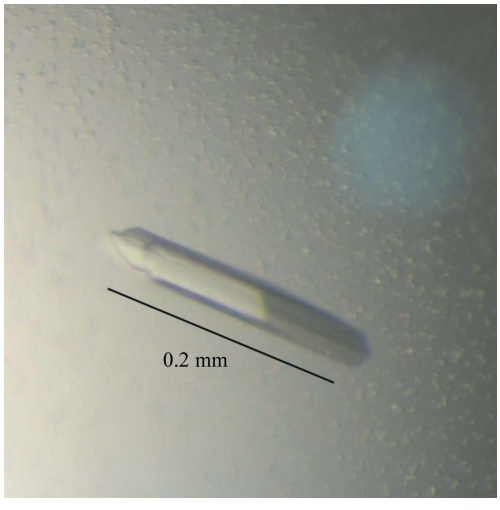


Figure 1
View of an orthorhombic SorC crystal grown at 294 K. The crystal dimensions are ∼0.1 × 0.1 × 0.02 mm.

Table 1
Data-collection statistics.

Values in parentheses are for the highest resolution shell.

Wavelength (Å)	0.87260
Space group	$P2_12_12_1$
Unit-cell parameters (Å)	a = 91.7, b = 113.3, c = 184.1
Resolution range (Å)	64.96–3.20 (3.37–3.20)
$R_{\rm merge}$†	0.156 (0.642)
$R_{\rm p.i.m.}$‡	0.096 (0.390)
Total No. of reflections	117540 (17423)
No. of unique reflections	32395 (4683)
Mean $I/\sigma(I)$	8.6 (2.1)
Completeness (%)	99.9 (100.0)
Multiplicity	3.6 (3.7)

† $R_{\rm merge} = \sum_{hkl}\sum_i |I_i(hkl) - \overline{I(hkl)}|/\sum_{hkl}\sum_i I_i(hkl)$. ‡ $R_{\rm p.i.m.} = \sum_{hkl}[1/(N-1)]^{1/2}\sum_i |I_i(hkl) - \overline{I(hkl)}|/\sum_{hkl}\sum_i I_i(hkl)$.

2.2. X-ray analysis

Crystals suitable for X-ray diffraction were transferred into cryoprotectant solution, which had the same composition as the mother liquor but with 20% glycerol added. Crystals were then flash-frozen in liquid nitrogen. Diffraction experiments were conducted at 100 K at the ID23-2 microfocus beamline at the ESRF (Grenoble, France) to best deal with the limited crystal size. An oscillation step of 0.5° over a range of 90° was used, with an exposure of 1 s per frame. Diffraction images to a resolution of 3.20 Å were indexed and integrated using *MOSFLM* (Leslie, 2006) and scaled with *SCALA* (Evans, 2006). Crystals belong to space group $P2_12_12_1$, with unit-cell parameters a = 91.6, b = 113.3, c = 184.1 Å. Data-collection statistics are reported in Table 1.

The structure was solved by molecular replacement with *Phaser* (McCoy *et al.*, 2005) using the homologous structure (25% sequence identity) of the transcriptional regulator from *Streptococcus pneumoniae* (PDB code 2gnp) as a starting model. The model used for molecular replacement was prepared using the program *CHAINSAW* (Schwarzenbacher *et al.*, 2004) based on alignment with the only homologous protein available. Subsequently, parts of the model belonging to loop regions that appear not to be conserved according to the sequence alignment were removed. The crystal-packing parameter (Matthews, 1985) indicates that six SorC molecules can be accommodated in the asymmetric unit; however, it was only possible to orient four molecules in the asymmetric unit. This arrangement of four molecules is in agreement with the tetrameric state of SorC. Different runs of *Phaser* were needed to obtain the best solution and to correctly place the four monomers. Initially, two molecules corresponding to a dimer were placed in the asymmetric unit. Subsequently, two dimers were placed in a second run, with an LLG (log-likelihood gain) of 288 and Z scores of 12.9 and 15.1 for the two dimers, using the previously obtained dimer as a search model.

The program *RESOLVE* (Terwilliger, 2000) with the prime-and-switch option was used to improve the electron-density map and remove model bias arising from the molecular-replacement solution. The quality of the map dramatically improved using NCS averaging (final FOM = 65%) between the four molecules and allowed a very good interpretation of the map despite the limited resolution. Manual

rebuilding of the molecular-replacement solution with *Coot* (Emsley & Cowtan, 2004) is in progress.

References

Bonsor, D., Butz, S. F., Solomons, J., Grant, S., Fairlamb, I. J., Fogg, M. J. & Grogan, G. (2006). *Org. Biomol. Chem.* **4**, 1252–1260.
Emsley, P. & Cowtan, K. (2004). *Acta Cryst.* D**60**, 2126–2132.
Evans, P. (2006). *Acta Cryst.* D**62**, 72–82.
Kelker, N. E., Simkins, R. A. & Anderson, R. L. (1972). *J. Biol. Chem.* **247**, 1479–1483.
Leslie, A. G. W. (2006). *Acta Cryst.* D**62**, 48–57.
McCoy, A. J., Grosse-Kunstleve, R. W., Storoni, L. C. & Read, R. J. (2005). *Acta Cryst.* D**61**, 458–464.
Matthews, B. W. (1985). *Methods Enzymol.* **114**, 176–187.
Schwarzenbacher, R., Godzik, A., Grzechnik, S. K. & Jaroszewski, L. (2004). *Acta Cryst.* D**60**, 1229–1236.
Sprenger, G. A. & Lengeler, J. W. (1984). *J. Bacteriol.* **157**, 39–45.
Terwilliger, T. C. (2000). *Acta Cryst.* D**56**, 965–972.
Wohrl, B. M., Wehmeier, U. F. & Lengeler, J. W. (1990). *Mol. Gen. Genet.* **224**, 193–200.
Zeng, X., Saxild, H. H. & Switzer, R. L. (2000). *J. Bacteriol.* **182**, 1916–1922.