Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

Crystallization and preliminary X-ray analysis of *N*-acetyl-1-D-myo-inosityl-2-deoxy-α-D-glucopyranoside deacetylase (MshB) from *Mycobacterium tuberculosis*

Andrew A. McCarthy,[a,b] Rainer Knijff,[a] Neil A. Peterson[a] and Edward N. Baker[a]*

[a]School of Biological Sciences, University of Auckland, Auckland, New Zealand, and [b]European Molecular Biology Laboratory, 6 Rue Jules Horowitz, BP 181, 38042 Grenoble, France

Correspondence e-mail: ted.baker@auckland.ac.nz

Received 1 August 2003
Accepted 23 September 2003

This paper is dedicated to Rainer Knijff, a talented graduate student, who died tragically before the work could be completed.

Mycobacteria synthesize mycothiol (MSH) as a low-molecular-weight thiol that protects against oxidative stress in a similar role to that of glutathione in many other species. The absence of MSH in mammals suggests that enzymes from its biosynthetic pathway in *Mycobacterium tuberculosis* could be useful targets for drug design. The gene for MshB (Rv1170), the enzyme that catalyses the second step in MSH biosynthesis in *M. tuberculosis*, has been cloned and the protein has been expressed in *Escherichia coli* both in native and SeMet-substituted forms and crystallized in two crystal forms. One of these, prepared in the presence of β-octylglucoside as a key additive, is suitable for high-resolution X-ray structural analysis. The crystals are orthorhombic, with unit-cell parameters a = 71.69, b = 83.74, c = 95.65 Å, space group $P2_12_12_1$ and two molecules in the asymmetric unit. X-ray diffraction data to 1.9 Å resolution have been collected.

1. Introduction

All cells require protection against the reactive oxygen species that are generated during normal metabolic processes. In eukaryotes and many bacteria, glutathione plays a key role in this protection. In mycobacteria, however, glutathione is absent and an alternative low-molecular-weight thiol compound known as mycothiol [1-D-myo-inosityl-2-(*N*-acetyl-L-cysteinyl)amido-2-deoxy-α-D-glucopyranoside] is used (Newton *et al.*, 1996). Mycothiol (MSH) is produced at millimolar levels in mycobacteria, in which it serves to maintain the correct redox potential in the cell, protect against oxidative stress and possibly provide an oxidation-resistant storage for cysteine (Anderberg *et al.*, 1998).

The biosynthetic pathway for MSH in *Mycobacteria* is believed to involve at least four steps (Newton & Fahey, 2002; Newton *et al.*, 2003). The first step produces a key intermediate 1-D-myo-inosityl-2-amino-2-deoxy-α-D-glucopyranoside (GlcNac-Ins), which is a conjugate of 1-D-myo-inositol and *N*-acetylglucosamine. Subsequent steps involve removal of the acetyl group by a deacetylase MshB, addition of L-cysteine by a ligase MshC and transfer of an acetyl group by an acetyltransferase MshD. Candidate genes for all of these activities have now been identified in *M. smegmatis* and their homologues found in *M. tuberculosis* (Newton & Fahey, 2002; Newton *et al.*, 2003).

It has recently been shown that a disruption of the mycothiol-biosynthesis pathway in *M. smegmatis* results in an increased sensitivity to antibiotics (Rawat *et al.*, 2002). This, together with our interest in thiol redox proteins (McCarthy *et al.*, 2000) and our participation in the TB structural genomics consortium (Goulding *et al.*, 2002), caused us to decide to undertake the structure determination of enzymes from this pathway in *M. tuberculosis*. Here, we describe the successful expression, purification and crystallization of MshB, the deacetylase that catalyses the second step in mycothiol biosynthesis in *M. tuberculosis* (Newton *et al.*, 2000).

2. Materials and methods

2.1. Cloning and expression

The DNA fragment encoding the MshB protein from *M. tuberculosis* (Rv1170) was obtained by PCR from genomic DNA. The primers 1, 5′-AAAAGACCATGGCTGAGACGCCCCGGCTGCTG-3′, and 2, 5′-AAGAATTCCTACGTGCCGGACGCGGTGAAGCCCAG-3′, were used to amplify the relevant region and incorporate the desired restriction sites (*Nco*I/*Eco*RI) for cloning into the expression vector pProEXHt (Life Technologies). This vector adds a hexa-His tag and linker peptide to the N-terminus of the expressed protein. The plasmid was transformed into the *Escherichia coli* cloning strain DH5α and then into the BL21(DE3) expression strain using ampicillin resistance of the plasmid for selection. Positive colonies were checked for a DNA fragment of the correct size for the *mshB* gene by both colony PCR and restriction digest. The plasmid was also sequenced at the Auckland University

sequencing facility to ensure no mutations had been introduced.

The cells were grown in Luria–Bertani broth containing 100 μg ml^{-1} ampicillin. Small-scale tests confirmed the presence of an expressed and soluble protein of the correct molecular weight and were used to find the optimal soluble expression conditions. Large-scale expression (2–4 l) was carried out by first growing the cells at 310 K. The cultures were then moved to 298 K when the OD_{600} = 0.6–0.8. After 1 h, expression was induced by addition of isopropyl-β-D-thiogalactopyranoside to a final concentration of 1 m*M* and incubation was continued overnight at 298 K. The cells were then harvested at 5000*g* for 15 min at 277 K.

The plasmid was later transformed into the methionine-auxotrophic *E. coli* strain BL41 (DE3) and grown in 2 l of LeMaster medium supplemented with 0.5 m*M* selenomethionine (SeMet) for preparation of SeMet protein (Hendrickson *et al.*, 1990). The same expression protocol as above was used for protein production.

2.2. Purification

The frozen bacterial cells were resuspended in ice-cold lysis buffer [100 m*M* HEPES pH 7.5 containing 200 m*M* NaCl, 2 m*M* tri(2-carboxyethyl)phosphine hydrochloride (TCEP) and 0.5 m*M* EDTA]. A protease-inhibitor tablet (Roche) was added and the cells were lysed with a French press. The cell debris was removed by centrifugation at 14 000*g* for 20 min at 277 K. The clarified supernatant was loaded onto a 5 ml Hi-Trap column (Pharmacia) and washed with five column volumes of lysis buffer containing 25 m*M* imidazole. The protein was eluted using a linear gradient of 0.025–0.5 *M* imidazole in lysis buffer. Protein-containing fractions were pooled and dialysed into 50 m*M* Tris–HCl pH 8.0 containing 150 m*M* NaCl, 2 m*M* TCEP and 0.5 m*M* EDTA overnight. The polyhistidine tag was removed by digestion with tobacco etch virus (TEV) protease for 2 h, followed by passage of the digestion mixture through the Hi-Trap column again to remove uncleaved protein and the TEV protease. The protein was then concentrated and loaded onto a Superdex75 gel-filtration column (Pharmacia). The protein eluted as a single peak and the protein-containing fractions were dialysed and concentrated to an approximate concentration of 10 mg ml^{-1} by ultrafiltration in 10 m*M* HEPES buffer, pH 7.5. The SeMet protein was purified using the same protocol, except that the polyhistidine tag was not removed from the protein that gave the crystals used for data collection.

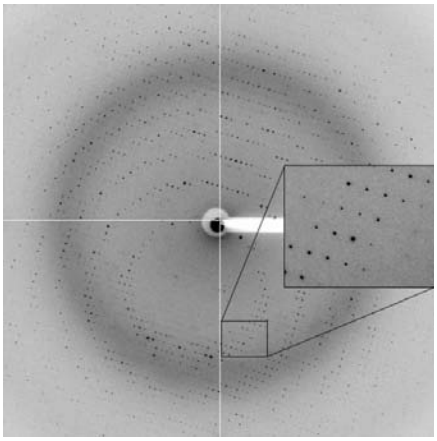

Figure 1
X-ray diffraction pattern obtained from form 2 crystals of MshB on beamline ID29 at ESRF, taken as a 1° oscillation, 1 s exposure with a crystal-to-detector distance of 170 mm. The resolution at the edge is 1.9 Å.

Table 1
Crystal data and data-collection statistics.

Values in parentheses are for the outermost shell of data collected.

	Form 1	Form 2
Unit-cell parameters		
a (Å)	161.63	71.69
b (Å)	71.44	83.74
c (Å)	150.78	95.65
β (°)	103.02	
Space group	*C*2	$P2_12_12_1$
Matthews coefficient (Å^3 Da^{-1})	3.3	2.1
Solvent content (%)	62.7	41.9
No. molecules in the AU	4	2
Resolution range (Å)	25–2.3 (2.38–2.30)	20–1.9 (2.1–1.9)
Wavelength used (Å)	1.08	0.979
Total No. unique reflections collected	72607	44945
No. observed reflections [$I/\sigma(I) > 1$]	224043	324835
Completeness (%)	95.9 (80.1)	97.6 (94.3)
Multiplicity	3.0	7.2
$R_{\rm merge}$†	0.072 (0.364)	0.071 (0.223)
Mean $I/\sigma(I)$	14.2 (2.6)	17.8 (7.7)
Percentage of reflections with $I/\sigma(I) > 3$	95.0 (76.3)	91.1 (78.2)

† $R_{\rm merge} = \sum|I_i - \langle I\rangle|/\sum\langle I\rangle$, where I_i and $\langle I\rangle$ are the measured and mean intensities, respectively, of a reflection *hkl*.

2.3. Crystallization and preliminary X-ray data

All crystallization trials were performed using the hanging-drop vapour-diffusion method. Initial crystallization conditions were found using commercially available crystallization screens (Hampton Research) and subsequent crystals were prepared by fine variation of these conditions. Native crystals were obtained at 291 K from 24–27% monomethylether (mme) PEG 2000, 0.2 *M* ammonium sulfate in 100 m*M* sodium cacodylate pH 6.1. Crystals (form 1) appeared overnight as stacked plates and grew to full size in a few days. They could be broken apart into single plates suitable for diffraction experiments, but were fragile and difficult to work with. Addition of β-octylglucoside (BOG) to the initial crystallization conditions resulted in another crystal form (form 2). Fine screening of these conditions resulted in an optimal crystallization condition comprising 15–20% PEG 3000, 200 m*M* ammonium sulfate and 0.6–0.7% BOG in 100 m*M* sodium cacodylate pH 6.3. These crystals grew as large twinned blocks and proved to be a much more robust crystal form. Again, pieces suitable for diffraction could be broken off and occasionally single crystals were obtained. These conditions were used to crystallize the SeMet protein because the crystals were easier to manipulate.

The native crystals were transferred to 40% mme PEG 2000, 0.2 *M* ammonium sulfate in 100 m*M* sodium cacodylate pH 6.1 for flash-freezing in a 100 K nitrogen-gas stream. The SeMet crystals were first transferred to 25% PEG 3000, 200 m*M* ammonium sulfate and 0.6–0.7% BOG in 100 m*M* sodium cacodylate pH 6.3 for manipulation before transferring to a similar condition with a higher PEG concentration (37–38%) for flash-freezing. Native diffraction data were measured at the Stanford Synchrotron Radiation Laboratory (SSRL) beamline 9.2 using a 3 × 3 ADSC Quantum 315 detector. SeMet diffraction data were measured on beamline ID29 at the European Synchrotron Radiation Facility (ESRF) using a 2 × 2 ADSC Quantum 210 CCD detector (Fig. 1). The data-collection strategies were optimized with the computer program *MOSFLM* (Leslie, 1992). The SSRL data

were processed using the *HKL* program package (Otwinowski & Minor, 1997) and the ESRF data were processed using the *XDS* program package (Kabsch, 2001). Data statistics are given in Table 1.

3. Results and discussion

The first crystal form obtained proved to be very difficult to work with. Although single crystals could sometimes be separated from the stacks of plates and did give a native data set to 2.3 Å resolution, their fragility made them unsuitable for soaking experiments. The discovery that the use of β-octylglucoside induced a second and much more favourable crystal form made this the approach of choice for a high-resolution crystal structure analysis; these crystals, which have only two molecules in the asymmetric unit compared with four for form 1, diffracted to higher resolution and were larger and much more robust. These properties are probably related to the lower solvent content (42% compared with 63% for the form 1 crystals). It has been noted before that the use of small concentrations of detergents can aid crystallization of soluble proteins (McPherson *et al.*, 1986). In this case, we speculate that the β-octylglucoside molecules may actively participate in the packing of the protein molecules in the form 2 crystals. Alternatively, since the natural substrate of MshB is a substituted glucose derivative, it is possible that β-octylglucoside binds in or around the active site and may induce a more favourable conformation for crystallization. Once the form 2 crystallization conditions had been discovered, it was found that these crystals could be obtained for both native and SeMet protein and both with and without the polyhistidine affinity tag. The SeMet data were collected from crystals in which the protein molecules were still His tagged and these crystals are now being used for a high-resolution X-ray structure analysis.

We thank Clyde Smith for help with synchrotron data collection and acknowledge financial support from the Marsden Fund, the Health Research Council and the New Economy Research Fund of New Zealand.

References

Anderberg, S. J., Newton, G. L. & Fahey, R. C. (1998). *J. Biol. Chem.* **273**, 30391–30397.
Goulding, C. W. *et al.* (2002). *Curr. Drug Targets Infect. Disord.* **2**, 121–141.
Hendrickson, W. A., Horton, J. R. & LeMaster, D. M. (1990). *EMBO J.* **9**, 1665–1672.
Kabsch, W. (2001). *International Tables for Crystallography*, Vol. *F*, edited by M. G. Rossmann & E. Arnold, pp. 730–734. Dordrecht: Kluwer Academic Publishers.
Leslie, A. G. W. (1992). *Jnt CCP4/ESF–EAMCB Newsl. Protein Crystallogr.* **26**.
McCarthy, A. A., Haebel, P. W., Torronen, A., Rybin, V., Baker, E. N. & Metcalf, P. (2000). *Nature Struct. Biol.* **7**, 196–199.
McPherson, A., Koszelak, S., Axelrod, H., Day, J., Robinson, L., McGrath, M., Willams, R. & Cascio, D. (1986). *J. Cryst. Growth*, **76**, 547–553.
Newton, G. L., Arnold, K., Price, M. S., Sherrill, C., Delcardayre, S. B., Aharonowitz, Y., Cohen, G., Davies, J., Fahey, R. C. & Davis, C. (1996). *J. Bacteriol.* **178**, 1990–1995.
Newton, G. L., Av-Gay, Y. & Fahey, R. C. (2000). *J. Bacteriol.* **182**, 6958–6963.
Newton, G. L. & Fahey, R. C. (2002). *Arch. Microbiol.* **178**, 388–394.
Newton, G. L., Koledin, T., Gorovitz, B., Rawat, M., Fahey, R. C. & Av-Gay, Y. (2003). *J. Bacteriol.* **185**, 3476–3479.
Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.
Rawat, M., Newton, G. L., Ko, M., Martinez, G. J., Fahey, R. C. & Av-Gay, Y. (2002). *Antimicrob. Agents Chemother.* **46**, 1–8.