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

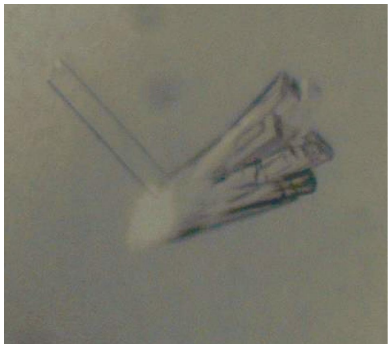
Crystallization and preliminary X-ray crystallographic analysis of MbtI, a protein essential for siderophore biosynthesis in *Mycobacterium tuberculosis*

Anthony J. Harrison,[a,b] Rochelle J. Ramsay,[b] Edward N. Baker[a,b] and J. Shaun Lott[a,b]*

[a]School of Biological Sciences, University of Auckland, Private Bag 92-019, Auckland, New Zealand, and [b]Centre for Molecular Biodiscovery, University of Auckland, Private Bag 92-019, Auckland, New Zealand

Correspondence e-mail: s.lott@auckland.ac.nz

Received 30 September 2004
Accepted 28 November 2004
Online 24 December 2004



Mycobacterium tuberculosis, the causative agent of tuberculosis, depends on the secretion of salicylate-based siderophores called mycobactins for the acquisition of extracellular iron, which is essential for the growth and virulence of the bacterium. The protein MbtI is thought to be the isochorismate synthase enzyme responsible for the conversion of chorismate to isochorismate, the first step in the salicylate production required for mycobactin biosynthesis. MbtI has been overexpressed in *Escherichia coli*, purified and crystallized. The crystals diffract to a maximum resolution of 1.8 Å. They belong to space group $P2_12_12_1$, with unit-cell parameters $a = 51.8$, $b = 163.4$, $c = 194.9$ Å, consistent with the presence of either two, three or four molecules in the asymmetric unit.

1. Introduction

Mycobacterium tuberculosis, the cause of tuberculosis (TB), is one of mankind's deadliest pathogens, being responsible for more than 2 million deaths every year (Rattan *et al.*, 1998). Recent estimates are that around a third of the world's population is infected. Although the modern multi-drug treatment for TB has a high cure rate when properly administered, current drugs are only active on replicating bacteria, necessitating long treatment times (six to nine months) for effective therapy. The ability of the bacterium to enter a non-replicating or persistent state and the emergence of multi-drug-resistant strains highlight the need for new drugs. The availability of the complete genome sequence for *M. tuberculosis* (Cole *et al.*, 1998) has provided the opportunity to identify new potential drug targets and to gain a better understanding of the biology of the pathogen.

Iron is an obligate cofactor for at least 40 different enzymes in *M. tuberculosis* and the capacity of *M. tuberculosis* to grow in animal and human hosts, especially in iron-limited extracellular spaces such as respiratory cavities and in the macrophage phagosome, is dependent on its ability to scavenge iron from its environment. Like many bacteria, *M. tuberculosis* is dependent on the synthesis and secretion of specific iron-binding compounds or siderophores for environmental iron uptake. Although other mycobacteria can produce two structurally distinct classes of siderophore, *M. tuberculosis* only synthesizes the phenyloxazolidine ring-containing class termed mycobactins (MBs; De Voss *et al.*, 1999). Mycobactins can be either soluble or cell-associated, but both types are based on a salicylate core to which modified amino-acid moieties are attached (Fig. 1). Salicylate itself is actively secreted from growing mycobacteria and is thought to play a role in extracellular iron solubilization (De Voss *et al.*, 1999). The biosynthesis of mycobactin from salicylate has been shown to be essential for mycobacterial growth and survival inside macrophages (De Voss *et al.*, 2000), making the enzymes from this biosynthetic pathway good potential targets for new anti-mycobacterial agents.

Most of the genes thought to encode the enzymes of mycobactin biosynthesis are found in a multi-gene cluster in the *M. tuberculosis* genome, spanning the open reading frames (ORFs) Rv2377c to Rv2384, annotated as *mbtH* to *mbtA*, respectively. Adjacent to this cluster is the ORF Rv2386c. This was originally annotated as encoding a putative anthranilate synthase (*trpE2*), with a presumed role in tryptophan biosynthesis. However, the biochemical function of anthranilate synthase, conversion of chorismate to anthranilate, is

very similar to that of isochorismate synthase, which converts chorismate to isochorismate in the first step of salicylate biosynthesis (Fig. 2). The enzymes are also evolutionarily related, with approximately 27% amino-acid sequence identity over a ~430-amino-acid segment. Given this similarity, its genomic context alongside other mycobactin biosynthesis genes (Quadri *et al.*, 1998) and the fact that the expression of the Rv2386c gene product is regulated by iron (Rodriguez *et al.*, 2002), the gene has been re-annotated as *mbtI* and its gene product tentatively identified as isochorismate synthase.

Here, we describe the cloning of the Rv2386c gene from *M. tuberculosis* and the expression, purification and crystallization of its gene product, the presumed isochorismate synthase MbtI, in a form suitable for high-resolution X-ray crystallographic analysis.

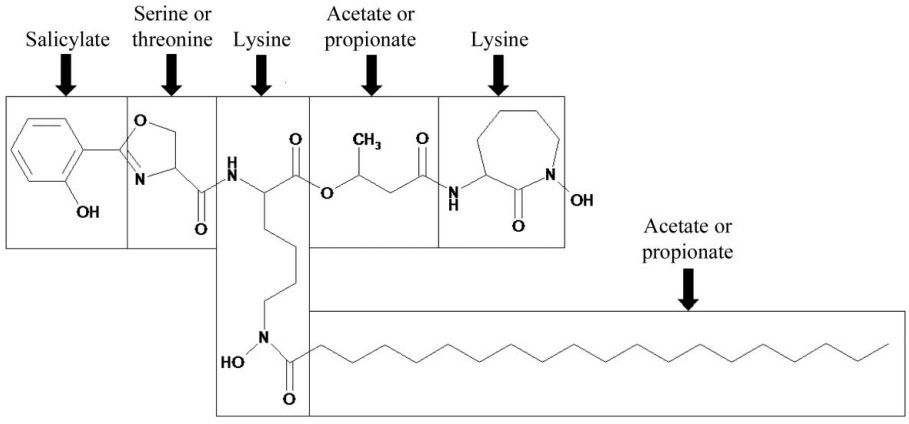


Figure 1
The chemical structure of mycobactin T, the major mycobactin siderophore produced by *M. tuberculosis*, annotated with its biosynthetic constituents. Redrawn from De Voss *et al.* (1999).

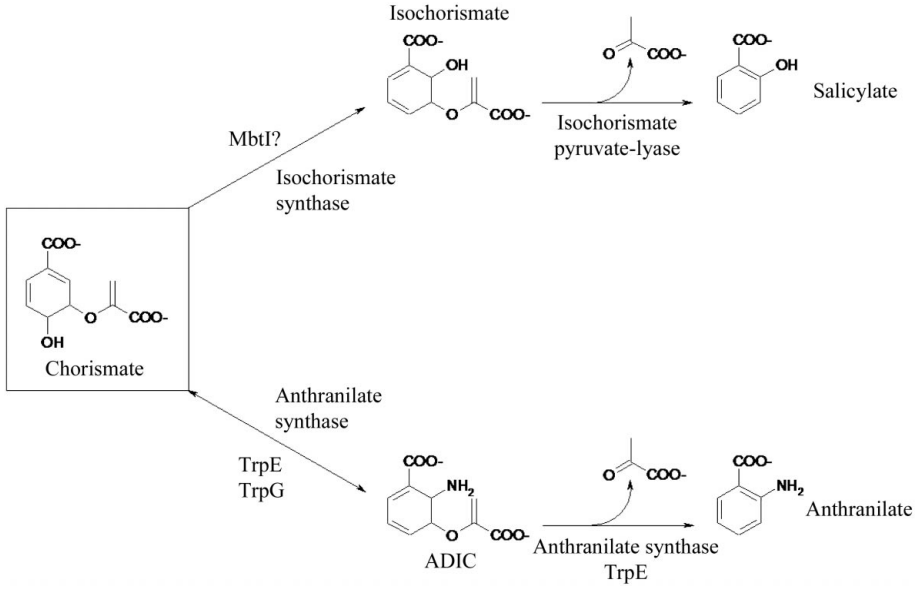


Figure 2
The biosynthesis of anthranilate and salicylate. The TrpE–TrpG complex is responsible for the conversion of chorismate to anthranilate *via* 2-amino-2-deoxyisochorismate (ADIC; the 'amino analogue' of isochorismate) with the concomitant deamidation of glutamine and release of pyruvate. MbtI is the putative isochorismate synthase, presumed to catalyze the analagous conversion of chorismate to isochorismate. The gene that encodes the isochorismate-pyruvate lyase in *M. tuberculosis* has not been identified.

2. Materials and methods

2.1. Protein expression and purification

The open reading frame Rv2386c (*mtbI*) was amplified from *M. tuberculosis* H37Rv genomic DNA using the polymerase chain reaction (PCR) with primers designed to introduce *Nco*I and *Pst*I sites at the 5′ and 3′ ends, respectively, and was cloned into a modified pET42a vector (Novagen), yielding a fusion protein in which a tobacco etch virus (TEV) protease-cleavage site (Parks *et al.*, 1994) was incorporated between MbtI and the C-terminal glutathione *S*-transferase (GST) protein. For protein expression, the plasmid was transformed into *Escherichia coli* BL21 (DE3) cells, which contained the plasmid pRI952 to supply tRNAs for decoding the 'rare' AGG, AGA and AUA codons (Del Tito *et al.*, 1995). A single transformed colony was used to seed a culture in LB medium, which was grown at 310 K for 8 h with shaking, induced with IPTG (isopropyl-β-D-thiogalactopyranoside) at a final concentration of 1 m*M* and grown at 297 K for a further 18 h before being harvested.

The cells were harvested by centrifugation (6000*g* for 15 min), resuspended in lysis buffer (20 m*M* HEPES pH 7.5, 150 m*M* NaCl, 10 m*M* imidazole, 1% glycerol; 20 ml per litre of culture) and passed through a cell disruptor (Constant Cell Disruptor Systems) at 117 MPa. After centrifugation (20 000*g* for 15 min), the supernatant was removed and filtered through a 0.22 µm filter and the cell pellet was discarded. The supernatant, containing the MbtI-GST fusion protein, was then incubated with glutathione-Sepharose 4B resin (Amersham Biosciences) pre-equilibrated in cold lysis buffer (2 ml of resin per litre of culture) at 277 K for 2 h. The resin was centrifuged, the supernatant discarded and the pellet washed three times in cold lysis buffer to remove all unbound protein. After washing, the resin was incubated overnight at 277 K with rTEV protease (0.5 mg per litre of culture) in lysis buffer (20 ml per 4 ml of resin). The resin was centrifuged (1000*g* for 5 min) and the supernatant was removed and passed down a HiTrap NTA column that had previously been loaded with Ni^{2+} ions to remove the His-tagged rTEV protease; the MbtI protein was collected in the flowthrough. The salt and imidazole in the buffer were removed by successive cycles of dilution with storage buffer (20 m*M* HEPES pH 8.0, 1% glycerol) followed by concentration using a spin-concentrator (10 kDa molecular-weight cutoff; Vivascience) and the protein was finally concentrated to 10 mg ml^{-1}.

2.2. Crystallization

Screening for crystallization conditions was performed using sitting-drop vapour diffusion in 96-well Intelliplates (Hampton Research) at 291 K by mixing 1 µl protein solution (10 mg ml^{-1} in 20 m*M* HEPES pH 8.0, 1% glycerol) with 1 µl well solution. Initial screens included the Hampton Crystal Screens I and II (Hampton Research), a systematic PEG–pH screen (Kingston *et al.*, 1994), a PEG/Ion screen (Hampton Research), the Footprint Screen No. 1

and PEG Footprint Screen (Stura *et al.*, 1992). A precipitate formed in the majority of drops within 1 h, including those that later formed crystals. Crystals appeared in 4–5 d from a reservoir solution comprising 15% PEG 4000, 0.2 *M* imidazole–malate pH 6.0 (PEG Footprint Screen condition B4).

2.3. Data collection

Crystals were soaked in a cryoprotectant solution comprising reservoir solution supplemented with 20%(*v*/*v*) glycerol and were then flash-cooled at 113 K under a stream of cold nitrogen. X-ray diffraction data were first collected using Cu $K\alpha$ radiation from a Rigaku RU300 rotating-anode generator equipped with Osmic mirror optics and a MAR345 image plate. This gave a native data set to a maximum resolution of 2.6 Å. The frozen crystal was then sent to the European Synchrotron Radiation Facility (ESRF, Grenoble, France), where a native data set was collected to a maximum resolution of 1.8 Å (beamline ID 14-4 with an ADSC Quantum 4 detector). The data were indexed and integrated using *MOSFLM* (Collaborative Computational Project, Number 4, 1994), giving a data set that was 99.5% complete with an overall $R_{\rm merge}$ of 8.6% on intensities (see Table 1 for full data-collection statistics).

3. Results and discussion

After glutathione-Sepharose affinity purification, recombinant MbtI was essentially pure, migrating with an apparent molecular weight of approximately 48 kDa on SDS–PAGE. An unidentified minor contaminating protein could be seen at approximately 50 kDa, but it was not removed and did not interfere with crystallization. The only successful crystallization conditions were those identified from the PEG Footprint screen, with precipitant conditions around 15% PEG 4000, 0.2 *M* imidazole–malate buffer pH 6.0. These conditions formed crystals that were long rods, bundled together and fused at one end. Occasionally, a single crystal formed in isolation from the others (Fig. 3) and it was from one such crystal that data were collected. The crystals were found to be orthorhombic, with systematic absences consistent with space group $P2_12_12_1$. The unit-cell parameters, $a = 51.8$, $b = 163.4$, $c = 194.9$ Å, are consistent with the presence of two, three or four molecules in the crystal asymmetric unit, corresponding to values of the Matthews coefficient $V_{\rm M}$ of 4.24, 2.82 and 2.12 Å^3 Da^{-1} and solvent contents of 71, 56 and 42%, respectively. Size-exclusion chromatography suggested that *M. tuberculosis* MbtI is monomeric in solution (data not shown), but some homologous enzymes are dimeric or tetrameric. Analysis of the self-rotation function was inconclusive as no clear peaks not related to crystal symmetry were observed, so the nature of the species in the asymmetric unit is yet to be established.

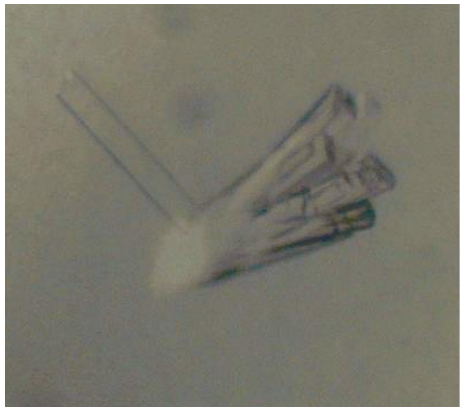


Figure 3
Crystals of *M. tuberculosis* MbtI, showing the characteristic bundled-rod morphology, with a single unbundled rod visible.

Table 1
Data-collection statistics for MbtI.

Values in parentheses are for the outermost shell.

Space group	$P2_12_12_1$
Unit-cell parameters (Å)	
a	51.82
b	163.36
c	194.93
Resolution range (Å)	46.6–1.8 (1.9–1.8)
Wavelength (Å)	0.9393
No. measured reflections	586276
No. unique reflections	153449
Multiplicity	3.8 (3.4)
Completeness (%)	99.5 (97.6)
Average $I/\sigma(I)$	4.75 (1.8)
$R_{\rm merge}$† (%)	8.6 (40.0)

† $R_{\rm merge} = \sum |I - \langle I \rangle| / \sum I$.

The high-resolution diffraction obtained from the native crystals establishes their suitability for X-ray structural analysis. Attempts were made to solve the structure by molecular replacement using *AMoRe*, *MOLREP* and *BEAST* (Collaborative Computational Project, Number 4, 1994) with a search model derived from the closest MbtI sequence homologue for which a three-dimensional structure is available. This is the anthranilate synthase (TrpE) from *Sulfolobus solfaricus*, which has about 27% amino-acid sequence identity with MbtI. The structures of other homologues (the TrpE enzymes from *Serratia marcescans*, *E. coli* and *Salmonella typhimurium*) were also tried, as was a composite model, but these experiments were unsuccessful, probably because the level of sequence identity was too low. We now plan to solve the structure of MbtI by multiwavelength anomalous diffraction (MAD) methods using selenomethionine-substituted protein, given that the MbtI monomer contains six methionines.

Thanks to Dr Didier Nurizzo for data collection at ESRF. This work is supported by the New Zealand Foundation for Research Science and Technology New Economy Research Fund. AJH is supported by a scholarship from the Health Research Council of New Zealand.

References

Cole, S. T. *et al.* (1998). *Nature (London)*, **393**, 537–544.
Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* D**50**, 760–763.
De Voss, J. J., Rutter, K., Schroeder, B. G. & Barry, C. E. (1999). *J. Bacteriol.* **181**, 4443–4451.
De Voss, J. J., Rutter, K., Schroeder, B. G., Su, H., Zhu, Y. & Barry, C. E. (2000). *Proc. Natl Acad. Sci. USA*, **97**, 1252–1257.
Del Tito, B. J. Jr, Ward, J. M., Hodgson, J., Gershater, C. J., Edwards, H., Wysocki, L. A., Watson, F. A., Sathe, G. & Kane, J. F. (1995). *J. Bacteriol.* **177**, 7086–7091.
Kingston, R. L., Baker, H. M. & Baker, E. N. (1994). *Acta Cryst.* D**50**, 429–440.
Parks, T. D., Leuther, K. K., Howard, E. D., Johnston, S. A. & Dougherty, W. G. (1994). *Anal. Biochem.* **216**, 413–417.
Quadri, L. E., Sello, J., Keating, T. A., Weinreb, P. H. & Walsh, C. T. (1998). *Chem. Biol.* **5**, 631–645.
Rattan, A., Kalia, A. & Ahmad, N. (1998). *Emerg. Infect. Dis.* **4**, 195–209.
Rodriguez, G. M., Voskuil, M. I., Gold, B., Schoolnik, G. K. & Smith, I. (2002). *Infect. Immun.* **70**, 3371–3381.
Stura, E. A., Nemerow, G. R. & Wilson, I. A. (1992). *J. Cryst. Growth*, **122**, 273–285.