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

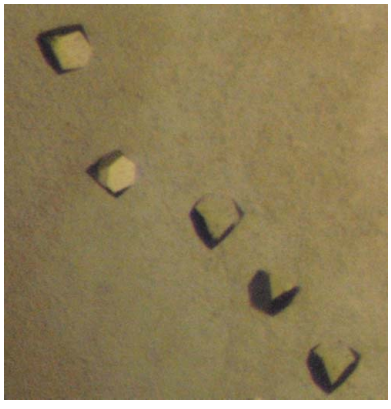
Purification, crystallization and preliminary X-ray analysis of *Mycobacterium tuberculosis* folylpolyglutamate synthase (MtbFPGS)

P. G. Young,[a]* C. A. Smith,[b] X. Sun,[c] E. N. Baker[a] and P. Metcalf[a]

[a]Laboratory of Structural Biology, School of Biological Sciences, Thomas Building, 3A Symonds Street, Auckland, New Zealand, [b]Stanford Synchrotron Radiation Laboratory, MS 992575 Sand Hill Road, Menlo Park, CA 94025, USA, and [c]HortResearch Palmerston North, Tennent Drive, Private Bag 11 030, Palmerston North, New Zealand

Correspondence e-mail: p.young@auckland.ac.nz

Received 11 April 2006
Accepted 9 May 2006



The gene encoding *Mycobacterium tuberculosis* FPGS (MtbFPGS; Rv2447c) has been cloned and the protein (51 kDa) expressed in *Escherichia coli*. The purified protein was crystallized either by the batch method in the presence of adenosine diphosphate (ADP) and $CoCl_2$ or by vapour diffusion in the presence of ADP, dihydrofolate and $CaCl_2$. X-ray diffraction data to approximately 2.0 and 2.6 Å resolution were collected at the Stanford Synchrotron Radiation Laboratory (SSRL) for crystals grown under the respective conditions. Both crystals belong to the cubic space group $P2_13$, with a unit-cell parameter of 112.6 and 111.8 Å, respectively. Structure determination is proceeding.

1. Introduction

Folates are essential vitamins that are required for normal cell growth, primarily through their central role in a variety of reactions involved in one-carbon metabolism. In these reactions, one-carbon units are utilized in the biosynthesis of various cellular components, including methionine, thymidylate and purine nucleotides (Stanger, 2002). Folates containing a single glutamate moiety are inefficient as cofactors and are not readily retained in cells. The active forms of folate are folylpolyglutamates, which contain a glutamate polypeptide of up to eight residues. The polyglutamation of folates is essential for their retention in the cell and increases their affinity for some folate-dependent enzymes (Shane, 1989; Sun *et al.*, 2001).

During the biosynthesis of folates, dihydrofolate synthetase (DHFS) adds L-glutamate to dihydropteroate (DHP) to form dihydrofolate (DHF). Dihydrofolate is converted to tetrahydrofolate (THF) by dihydrofolate reductase. Folylpolyglutamate synthetase (FPGS) adds additional glutamate residues *via* an amide linkage to the γ-carboxylate group of the folate or folate derivative to form polyglutamates (Fig. 1; Shane, 1989). FPGS activity seems to be ubiquitous in both prokaryotes and eukaryotes, whereas DHFS activity is limited to yeast, plants and most bacteria. In the majority of bacteria, both DHFS and FPGS activities are catalyzed by the same enzyme. In contrast, vertebrates have no DHFS activity associated with FPGS and cannot synthesize folates *de novo* (Suh *et al.*, 2001). The DHFS activity of bacterial FPGS could provide a target for antimicrobial chemotherapy, since folate biosynthesis is essential for bacterial survival (Pyne & Bognar, 1992). It has recently been reported that the binding site for DHP in *Escherichia coli* FPGS is situated in a distinct cavity which differs significantly from that for THF binding in *Lactobacillus casei* FPGS (Mathieu *et al.*, 2005). Because of the clear distinction between DHP and THF binding by bacterial FPGS, it may be possible to design inhibitors of bacterial DHFS activity that do not inhibit FPGS activity in humans, thereby selectively inhibiting folate metabolism in bacteria.

MtbFPGS, like *E. coli* FPGS, is thought to exhibit both DHFS and FPGS activities and has been shown to be an essential gene for the growth of *Mycobacterium tuberculosis* (Sassetti *et al.*, 2003). This raises the possibility that the development of inhibitors of MtbFPGS could also provide potential leads for chemotherapy against tuberculosis (TB). Here, we report the expression, purification and

crystallization of the 51 kDa MtbFPGS protein encoded by the *M. tuberculosis* gene Rv2447c.

2. Materials and methods

2.1. Cloning, expression and purification of MtbFPGS

A region of DNA encompassing the MtbFPGS (Rv2447c) gene was PCR-amplified from *M. tuberculosis* genomic DNA with Pfx polymerase (Invitrogen) using two synthesized oligonucleotide primers: TBFPGS-Sense2 (5′-GTTGGCTGCGCTGCAATGAA-TTCGACGA-3′) and TBFPGS-Antisense2 (5′-GGTATTGCCAG-CAGCACCACGATCGCCT-3′). The product was subsequently used for nested PCR-amplification with the oligonucleotide primers TBFPGS-sense (5′-CTGCGCTG**CCATGG**ATTCGACGAAT-3′) and TBFPGS-antisense (5′ TCGA**GGATCC**CGCGTCGCCGC-3′), which amplified the MtbFPGS gene including 5′ *Nco*I and 3′ *Bam*HI recognition sites, respectively (shown in bold). The amplified fragment was digested with *Nco*I and *Bam*HI and cloned into the respective sites in the multiple cloning site of a modified pET42a plasmid (Novagen), pET42a-rTEV, which had had the S-tag, thrombin and factor Xa cleavage sites removed and replaced with a recombinant tobacco etch virus protease (rTEV) cleavage motif. The resulting plasmid (GST-His_6-MtbFPGS) was transformed into *E. coli* (DH5α) and the nucleotide sequence was verified by sequencing in both directions.

For native MtbFPGS protein expression, GST-His_6-MtbFPGS was transformed into BL21 (λDE3) pGROELS chaperone strain cells, which were grown in Luria–Bertani (LB) medium supplemented with the required antibiotics at 310 K until OD_{600} reached 1.5. The cultures were then 'cold shocked' by cooling in ice water for 60 min. An equal volume of fresh pre-chilled media was added and expression was induced with 0.1 m*M* IPTG at 293 K for 16 h.

Selenomethionine-labelled MtbFPGS (SeMet-MtbFPGS) was produced using a modified protocol based on the inhibition of methionine biosynthesis (Doublié & Carter, 1992). Briefly, LB media was substituted with M9 minimal media and the cells were grown as in the above protocol described for the expression of native MtbFPGS protein. Once OD_{600} reached 1.5, 100 mg l^{-1} each of lysine, phenylalanine and threonine and 50 mg l^{-1} each of isoleucine, leucine and valine were added to the cultures. An abundance of L-selenomethionine (60 mg l^{-1}) was then added and the cells were grown for an additional 15 min at 310 K prior to cooling and induction with 0.1 m*M* IPTG at 293 K for 16 h.

Bacteria expressing MtbFPGS were pelleted and the cells disrupted using a cell disruptor (Constant Cell Disruption Systems) in lysis buffer [50 m*M* Tris–HCl pH 7.5, 200 m*M* NaCl, 1 m*M* $MgCl_2$, 2 m*M* β-mercaptoethanol, 10%(*v*/*v*) glycerol, 10 m*M* imidazole] containing 100 µg ml^{-1} lysozyme, 2 µg ml^{-1} DNase I, 10 µg ml^{-1} RNase A and Complete Mini protease inhibitors (Invitrogen). Insoluble material was removed by centrifugation (30 000*g*) and soluble recombinant protein was purified by nickel-chelating chromatography using Ni-Sepharose 6 Fastflow resin (Amersham Biosciences). The clarified soluble fraction from a 4 l culture was batch-bound to resin (5 ml) at 277 K for 60 min and loaded onto a column. The column was washed with lysis buffer containing 30 m*M* imidazole and recombinant protein was eluted with lysis buffer containing 200 m*M* imidazole. Fractions containing the bulk of the recombinant protein were pooled and dialyzed with a 1:20 ratio of rTEV-His_6 against buffer *B* (25 m*M* Tris–HCl pH 7.5, 150 m*M* NaCl, 0.05 m*M* EDTA) at 277 K for 16 h. MtbFPGS was separated from the rTEV-His_6 protease, the cleaved GST-His_6 tag and uncleaved protein by passing through a nickel-chelating column. The unbound protein containing MtbFPGS was concentrated using a 30 kDa molecular-weight cutoff protein concentrator (VivaScience) and loaded onto a Superdex-200 size-exclusion chromatography column (Amersham

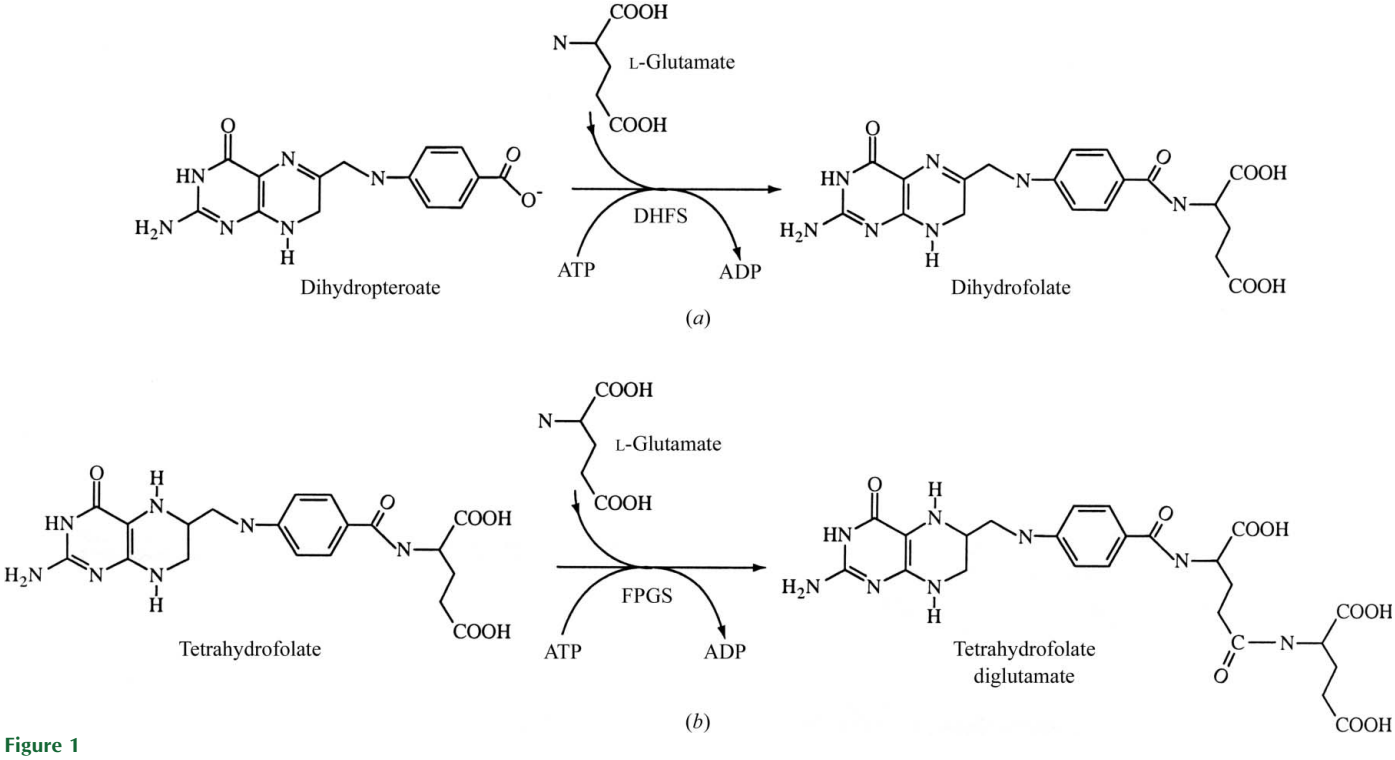


Figure 1
Schematic diagram of the conversion of (*a*) dihydropteroate to dihydrofolate and (*b*) tetrahydrofolate to tetrahydrofolate diglutamate. These reactions are performed by the same enzyme in the majority of bacteria.

Table 1
Composition of crystallization buffers used in the crystallization of MtbFPGS.

CB1 (robot screen)	CB2 (vapour diffusion)	CB3 (batch method)
15%(*w*/*v*) PEG 8000, 30%(*v*/*v*) MPD, 100 m*M* $CaCl_2$, 100 m*M* sodium acetate pH 5.5	15%(*w*/*v*) PEG 8000, 30%(*v*/*v*) MPD, 10 m*M* $CaCl_2$, 50 m*M* sodium acetate pH 5.5	14%(*w*/*v*) PEG 8000, 30%(*v*/*v*) MPD, 10 m*M* $CoCl_2$, 50 m*M* sodium acetate pH 5.5

Biosciences) pre-equilibrated with buffer *B*. Fractions containing MtbFPGS were kept separate and dynamic light-scattering (DLS) experiments were performed to measure the polydispersity of each fraction (DynaPro, Protein Solutions). SeMet-MtbFPGS was purified using the same procedures as those described for the native MtbFPGS.

2.2. Crystallization

MtbFPGS was concentrated to 10 mg ml^{-1} in buffer *B* and incubated with 1.5 m*M* ADP, 2 m*M* $MgCl_2$ and 2 m*M* DHF for 16 h at 277 K. Crystallization trials were by vapour diffusion using 100 + 100 nl drops dispensed by a Cartesian liquid-dispensing robot. Screening for crystallization utilized an in-house 480-condition set of sparse-matrix screens that included Hampton Crystal Screens I and II, Top 67, MPD, PEG Ion and Precipitant Synergy screens (Moreland *et al.*, 2005). A condition from the Precipitant Synergy reagent formulation screen produced small crystals overnight (CB1; Table 1). Crystallization conditions were optimized in 2 + 2 µl hanging drops (CB2; Table 1) by screening various buffers and pH ranges and divalent cations as additives. The mother liquor of the CB2 condition proved to be a good cryoprotectant and crystals were directly frozen in liquid nitrogen.

Diffracting crystals of MtbFPGS pre-incubated with 1.5 m*M* ADP and 2 m*M* $MgCl_2$ (Fig. 2) were also grown by the batch method under paraffin oil in 2 + 2 µl drops of protein and crystallization buffer (CB3, Table 1). The crystals appeared after 16 h and grew to a maximum dimension of 100 µm after 96 h. Before flash-freezing in liquid nitrogen, the crystals were soaked for 60 min in a 60:40 mix of cryoprotectant [CB3 crystallization buffer + 30%(*v*/*v*) glycerol] and buffer *B*. Crystals of SeMet-MtbFPGS pre-incubated with 1.5 m*M* ADP and 2 m*M* $MgCl_2$ were also grown by the batch method with crystallization buffer CB3 (Table 1) as described for the native protein.

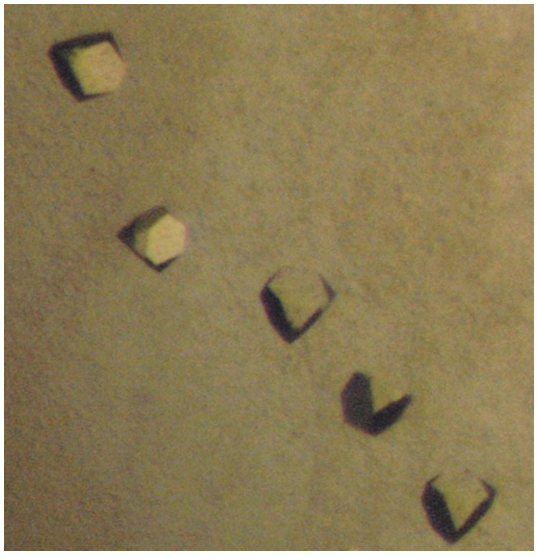


Figure 2
Crystals of *M. tuberculosis* FPGS.

Table 2
Statistics of data collection and processing.

	Native 1	Native 2
	Mg–ADP–DHF (vapour diffusion)	Mg–ADP
SSRL beamline	BL9.1	BL9.1
Wavelength (Å)	0.9798	0.9798
Space group	$P2_13$	$P2_13$
Unit-cell parameter (Å)	111.80	112.64
Resolution range (Å)	50–2.6 (2.7–2.6)	56–2.0 (2.1–2.0)
Observations	293597	428727
Unique reflections	14590	36736
Multiplicity	20	11
Completeness (%)	100	100
Mosaicity (°)	0.35	0.3
Solvent content (%)	45.0	46.3
Molecules per ASU	1	1
Mean $I/\sigma(I)$	30.9 (6.3)	20.8 (8.2)
R_{sym}† (%)	0.067 (0.62)	0.09 (0.38)

† $R_{sym} = \sum_h \sum_i [|I_i(h) - \langle I(h)\rangle|] / \sum_h \sum_i I_i(h)]$, where I_i is the ith measurement and $\langle I(h)\rangle$ is the weighted mean of all measurements of $I(h)$.

2.3. Data collection and preliminary X-ray analysis

Diffraction data from a single crystal of MtbFPGS–ADP/Mg^{2+}–DHF were collected on beamline BL9-1 at the Stanford Synchrotron Radiation Laboratory (SSRL), Stanford, CA, USA. The crystal was maintained at 100 K and diffraction data were collected on an ADSC Quantum 315 CCD. A total of 360 images were collected with an oscillation range of 0.5° per image, 45 s exposure per image and a crystal-to-detector distance of 350 mm. Diffraction data from a single crystal of MtbFPGS–ADP/Mg^{2+} grown by the batch method under paraffin oil were also collected on BL9-1 at the SSRL. A total of 200 images were collected with an oscillation range of 1° per image, a 20 s exposure per image and a crystal-to-detector distance of 240 mm. All data were indexed and integrated with *MOSFLM* (Leslie, 2006) and reduced with *SCALA* (Evans, 2006) from the *CCP*4*i* suite (Potterton *et al.*, 2003; Collaborative Computational Project, Number 4, 1994).

3. Results and discussion

Expression of MtbFPGS in *E. coli* (BL21 λDE3) at 293 K resulted in poor expression, with only a small yield of soluble highly aggregated protein as determined by DLS. Co-expression with GROELS chaperone proteins increased the yield of soluble protein, which was less severely aggregated, to ∼0.25 mg per litre of culture. However, growth of the bacteria to high density (OD_{600} = 1.5, as opposed to the customary OD_{600} = 0.5) in combination with a 'cold shock' resulted in a substantial improvement in both solubility and homogeneity (Cp/Rh = 10.5% from DLS), with an increased yield of ∼5 mg purified protein per litre of culture. The native molecular weight as estimated by DLS (51.3 kDa) closely matched that of the molecular weight predicted from the sequence (50.9 kDa) and indicates that MtbFPGS exists as a monomer.

Initial crystallization and fine-screening experiments were performed using vapour diffusion and resulted in MtbFPGS–ADP/Mg^{2+}–DHF crystals. It was observed that crystals grown under these conditions developed cracks over time. This did not occur with the crystals grown by the batch method. Growth of MtbFPGS crystals was dependent on divalent cations, with $CoCl_2$ resulting in better diffracting crystals than $CaCl_2$. No other divalent cations screened

resulted in crystal growth. The data-collection and processing statistics are presented in Table 2.

Efforts to determine the atomic structure of MtbFPGS by molecular replacement using FPGS structures from *L. casei* (Sun *et al.*, 2001; PDB code 1jbw), *E. coli* (Mathieu *et al.*, 2005; PDB code 1w78) and *Thermotoga maritima* (PDB code 1o5z), including ensembles of the whole molecule and separate domains, have not been successful. This may be the result of a combination of low sequence identity (<35%) between species and the flexible nature of these kinase-like proteins, which display large movements between their domains (Sun *et al.*, 2001).

Given these difficulties, we have prepared selenomethionine-substituted crystals in an attempt to obtain experimental phases. Three data sets at the peak and inflection point on the selenium absorption edge, along with a high-energy remote, have recently been collected at the SSRL. It is anticipated that these data will lead to the solution of the structure of MtbFPGS.

This work was supported by grants from the Health Research Council of New Zealand. The Stanford Synchrotron Radiation Laboratory (SSRL) is funded by the Department of Energy (BES, BER) and the National Institutes of Health (NCRR, NIGMS). We also thank Tom Caradoc-Davies, Christopher Squire and Richard Bunker for help with data collection.

References

Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* D**50**, 760–763.
Doublié, S. & Carter, C. W. Jr (1992). *Crystallization of Nucleic Acids and Proteins: A Practical Approach*, edited by A. Ducruix & R. Giegé, pp. 311–317. Oxford: IRL Press.
Evans, P. (2006). *Acta Cryst.* D**62**, 72–82.
Leslie, A. G. W. (2006). *Acta Cryst.* D**62**, 48–57.
Mathieu, M., Debousker, G., Vincent, S., Viviani, F., Bamas-Jacques, N. & Mikol, V. (2005). *J. Biol. Chem.* **280**, 18916–18922.
Moreland, N., Ashton, R., Baker, H. M., Ivanovic, I., Patterson, S., Arcus, V. L., Baker, E. N. & Lott, J. S. (2005). *Acta Cryst.* D**61**, 1378–1385.
Potterton, E., Briggs, P., Turkenburg, M. & Dodson, E. (2003). *Acta Cryst.* D**59**, 1131–1137.
Pyne, C. & Bognar, A. L. (1992). *J. Bacteriol.* **174**, 1750–1759.
Sassetti, C. M., Boyd, D. H. & Rubin, E. J. (2003). *Mol. Microbiol.* **48**, 77–84.
Shane, B. (1989). *Vitam. Horm.* **45**, 263–335.
Stanger, O. (2002). *Curr. Drug Metabol.* **3**, 211–223.
Suh, J. R., Herbig, A. K. & Stover, P. J. (2001). *Annu. Rev. Nutr.* **21**, 255–282.
Sun, X., Cross, J. A., Bognar, A. L., Baker, E. N. & Smith, C. A. (2001). *J. Mol. Biol.* **310**, 1067–1078.