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

Celia J. Webby,[a] J. Shaun Lott,[b] Heather M. Baker,[b] Edward N. Baker[b] and Emily J. Parker[a]*

[a]Centre for Structural Biology, Institute of Fundamental Sciences, Massey University, Private Bag 11-222, Palmerston North, New Zealand, and [b]School of Biological Sciences and Centre for Molecular Biodiscovery, University of Auckland, Private Bag 92-019, Auckland, New Zealand

Correspondence e-mail: e.j.parker@massey.ac.nz

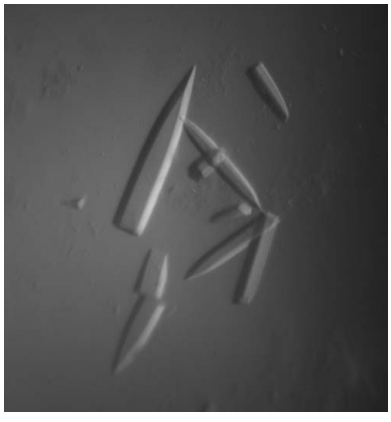
Received 10 February 2005
Accepted 14 March 2005
Online 24 March 2005

Crystallization and preliminary X-ray crystallographic analysis of 3-deoxy-D-*arabino*-heptulosonate-7-phosphate synthase from *Mycobacterium tuberculosis*

The enzymes of the shikimate pathway are attractive targets for new-generation antimicrobial agents. The first step of this pathway is catalysed by 3-deoxy-D-*arabino*-heptulosonate-7-phosphate (DAH7P) synthase and involves the condensation of phosphoenolpyruvate (PEP) and erythrose 4-phosphate (E4P) to form DAH7P. DAH7P synthases have been classified into two apparently evolutionarily unrelated types and whereas structural data have been obtained for the type I DAH7P synthases, no structural information is available for their type II counterparts. The type II DAH7P synthase from *Mycobacterium tuberculosis* was co-expressed as native and selenomethionine-substituted protein with the *Escherichia coli* chaperonins GroEL and GroES in *E. coli*, purified and crystallized. Native crystals of *M. tuberculosis* DAH7P synthase belong to space group $P3_121$ or $P3_221$ and diffract to 2.5 Å, with unit-cell parameters $a = b = 203.61$, $c = 66.39$ Å. There are either two or three molecules in the asymmetric unit. Multiwavelength anomalous diffraction (MAD) phasing using selenomethionine-substituted protein is currently under way.

1. Introduction

3-Deoxy-D-*arabino*-heptulosonate-7-phosphate (DAH7P) synthase (EC 2.5.1.54) catalyses a condensation reaction between phosphoenolpyruvate (PEP) and erythrose 4-phosphate (E4P) to form DAH7P (Fig. 1). This reaction is the first step of the shikimate pathway, a series of seven enzyme-catalysed reactions that are responsible for the biosynthesis of chorismate, a precursor for aromatic compounds including the aromatic amino acids (Bentley, 1990). The shikimate pathway is found in microorganisms and plants but is absent in higher organisms, making the enzymes of this pathway attractive as targets for the development of novel antimicrobial agents. One of the most important current challenges for drug discovery is the development of new antibiotics for *Mycobacterium tuberculosis*, which is the cause of tuberculosis (TB) and is responsible for over two million deaths annually (Rattan *et al.*, 1998). Although effective anti-TB drugs exist, the long treatment times required, the problems of latent or persistent TB (Parrish *et al.*, 1998) and the proliferation of multidrug-resistant strains of *M. tuberculosis*



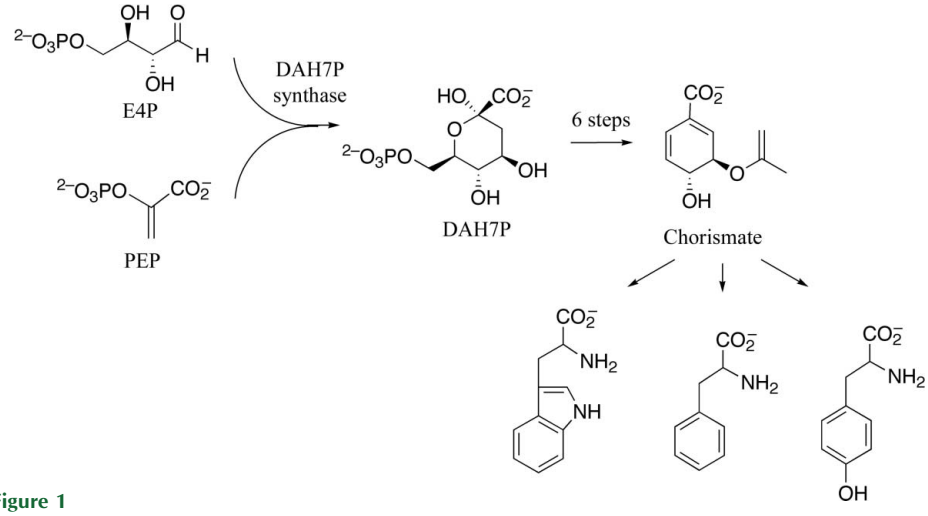


Figure 1
Biosynthesis of aromatic amino acids *via* the shikimate pathway.

(Stokstad, 2000) have all created an urgent need for the development of new antimycobacterial agents. Recent gene-disruption studies have shown that operation of the shikimate pathway is essential for the viability of *M. tuberculosis* (Parish & Stoker, 2002) and enzymes from this pathway are thus valid targets.

DAH7P synthases can be classified into two types based on sequence similarity and molecular weight (Walker *et al.*, 1996). Analysis of the genome of *M. tuberculosis* reveals the presence of a single gene encoding a DAH7P synthase belonging to the type II family. Although X-ray crystal structures have been determined for the type I enzymes from *Escherichia coli* (Shumilin *et al.*, 1999, 2003), *Saccharomyces cerevisiae* (Hartmann *et al.*, 2003; Konig *et al.*, 2004), *Thermotoga maritima* (Shumilin *et al.*, 2004) and *Pyrococcus furiosus* (Schofield *et al.*, 2005), there is limited functional and no structural information available for the type II enzymes.

Type II DAH7P synthases comprise a separate homologous group distinct from the type I enzymes (Gosset *et al.*, 2001; Jensen *et al.*, 2002). Structural information for this group is required to determine any evolutionary relationship between the two DAH7P synthase families. Type II enzymes were first identified in plants (Zhao & Herrmann, 1992; Dyer *et al.*, 1990) and subsequently a limited number of microbial examples were characterized (Walker *et al.*, 1996). However, as more microbial genomes have been sequenced the number of type II DAH7P synthases known has grown rapidly and it appears that this type of DAH7P synthase consists of a subset of plant enzymes clustered within a more divergent set of microbial enzymes (Gosset *et al.*, 2001). In some organisms, genes encoding both type I and type II DAH7P synthases have been identified, with several of the type II enzymes apparently being required for the biosynthesis of specific secondary metabolites (Silakowski *et al.*, 2000). However, the presence of only type II DAH7P synthases in the predicted proteomes of a number of species including *Streptomyces* species, *Corynebacterium diphtheriae*, *Campylobacter jejuni*, *Agrobacterium tumefaciens*, *Novosphingobium aromaticivorans*, *Helicobacter pylori* and several *Mycobacteria* species provides evidence for the role of type II DAH7P synthases in aromatic amino-acid biosynthesis. The presence of type II DAH7P synthases in a number of important pathogenic bacteria makes them prime targets for study, as exploitation of the differences between the two types may enable the development of narrow-spectrum antibiotics.

We report here the cloning, expression, *in vivo* solubilization, purification and crystallization of the type II DAH7P synthase from *M. tuberculosis*. The *in vivo* solubilization was achieved by co-expression with the *E. coli* chaperonins GroEL and GroES in *E. coli.*

2. Materials and methods

2.1. Protein expression and purification

The DNA encoding the opening reading frame encoding *M. tuberculosis* DAH7P synthase (Rv2178c) was amplified from *M. tuberculosis* H37Rv genomic DNA using the polymerase chain reaction (PCR) with primers designed to introduce *Nco*I and *Sac*I restriction sites at the 5′ and 3′ ends, respectively. The PCR product was cloned into a pProExHTa (Novagen) expression vector, yielding a protein containing an N-terminal tobacco etch virus (TEV) protease-cleavable His tag upstream of the predicted translation start site, so that cleavage with rTEV would yield a protein with the addition of two N-terminal residues (AG). For the production of soluble protein, the plasmid was transformed into *E. coli* BL21 (DE3) cells containing the pGroESL plasmid (Goloubinoff *et al.*, 1989). Transformed *E. coli* cells were grown in Luria broth (LB) medium at 310 K with shaking until mid-logarithmic phase ($OD_{600} \simeq 0.4$–0.6). The growth temperature was then lowered to 298 K and isopropyl-β-D-thiogalactopyranoside (IPTG; Applichem) was added to a final concentration of 1 m*M* to induce expression. 6 h after induction, the cells were harvested *via* centrifugation (6400*g* for 20 min) and stored at 203 K until lysis.

The cell pellet was re-suspended in lysis buffer [20 m*M* bis-Tris propane pH 7.5, 150 m*M* NaCl, 0.5 m*M* Tris(2-carboxyethyl)phosphine hydrochloride (TCEP), 0.005%(*v*/*v*) THESIT, 200 µ*M* PEP, 100 µ*M* $MnSO_4$] and lysed by passage through a cell disrupter (Constant Cell Disrupter Systems) at 97 MPa. The DNA was broken up by sonication and the cell debris was removed by centrifugation (27 000*g* for 15 min). The crude lysate was filtered through a 0.8 µm filter and then loaded onto a HiTrap NTA column (Amersham Biosciences) that had been previously charged with Ni^{2+} ions. The column was washed with five column volumes of lysis buffer before the bound proteins were eluted with a linear gradient of 0–500 m*M* imidazole in lysis buffer. The fractions exhibiting DAH7P synthase activity were pooled, exchanged into lysis buffer by dilution and concentration (10 kDa molecular-weight cutoff ultrafiltration device) and then incubated for up to 15 h at 295 K with rTEV protease (0.1 mg per litre of culture). The enzyme preparation was then concentrated to no more than 5 mg ml^{-1} using an ultrafiltration device with a 10 kDa molecular-weight cutoff (Vivascience) and further purified by size-exclusion chromatography on a Superdex S200 column (Amersham Biosciences), which also removed the rTEV protease.

DAH7P synthase activity was measured by monitoring the loss of PEP at 232 nm ($\varepsilon = 2.8 \times 10^3\ M^{-1}\ cm^{-1}$ at 303 K) in the presence of E4P using a Varian Cary 1 UV–visible spectrophotometer.

This procedure yielded approximately 4 mg homogeneous *M. tuberculosis* DAH7P synthase from a 1 l culture. The protein solution was concentrated to approximately 3 mg ml^{-1} for crystallization trials. For production of selenomethionine (SeMet) substituted protein, the plasmid bearing Rv2178c was transformed into cells of the methionine auxotroph DL41(DE3) that also contained pGroESL. The transformed cells were grown in minimal media with SeMet as the only methionine source (Hendrickson *et al.*, 1990). The SeMet-substituted protein was then purified using an identical protocol to that used for the native protein,

2.2. Crystallization

Screening for crystallization conditions was achieved using sitting-drop vapour diffusion in 96-well Intelliplates (Hampton Research) at 291 K by mixing 100 nl protein solution (3 mg ml^{-1} in lysis buffer) with 100 nl reservoir solution and equilibrating against 75 µl reservoir solution. The initial screens were set up using the Centre for Molecular Biodiscovery Crystallization facility (http://cmb1.auckland.ac.nz/), which consists of a modified Perkin–Elmer Multiprobe robot for the transfer of precipitant solutions from pre-made stock into Intelliplates and a Cartesian Honeybee robot for setting up the nanolitre crystallization experiments. Initial screens included Hampton Crystal Screens I and II (Hampton Research), a systematic PEG–pH screen (Kingston *et al.*, 1994), a PEG/Ion screen (Hampton Research) and the Footprint Screen (Stura *et al.*, 1992).

Small crystals appeared within 2–3 h in several conditions and larger needle crystals in 1–2 d in many different conditions. Further manual screening in 24-well plates (1 + 1 µl drops, 500 µl reservoir solution) was performed using these lead conditions. The precipitant condition that gave the most reproducible crystals comprised 0.1 *M* Tris–HCl pH 8.5, 1.5 *M* ammonium sulfate, 12%(*v*/*v*) glycerol

(Crystal Screen II No. 42). Refinement of crystallization conditions was achieved by altering the precipitant concentration and the pH. The best diffracting crystals were grown with 0.1 *M* Tris–HCl pH 8.0, 1.5 *M* ammonium sulfate, 12%(*v*/*v*) glycerol. The crystals grew to approximately 0.3 × 0.05 × 0.05 mm (Fig. 2). SeMet-substituted protein formed crystals using the same protein concentration that were of similar size and morphology. The best SeMet-substituted crystals were grown using the following three conditions: 0.1 *M* Na HEPES pH 7.5, 0.8 *M* sodium potassium tartrate (Crystal Screen I No. 29), 0.1 *M* Na HEPES pH 7.5, 1.5 *M* lithium sulfate (Crystal Screen I No. 16) and 0.1 *M* sodium citrate pH 5.6, 1 *M* ammonium phosphate (Crystal Screen I No. 11).

2.3. Data collection

Native crystals were soaked in a cryoprotectant solution consisting of 0.1 *M* Tris–HCl pH 8.0, 1.5 *M* ammonium sulfate supplemented with 25%(*v*/*v*) glycerol and were flash-frozen under a stream of cold nitrogen at 110 K. X-ray diffraction data were collected using a Rigaku MicroMax007 generator with Osmic blue optics and an R-AXIS IV^{++} detector. A highly redundant native data set was collected to a maximum resolution of 2.5 Å and an overall $R_{\rm merge}$ of 9.9% on intensities, using non-overlapping 0.3° oscillations collected for 15 min per frame at a crystal-to-detector distance of 180 mm (Table 1). Data were processed using *CrystalClear* v.1.3.6 (Rigaku).

3. Results and discussion

When *M. tuberculosis* DAH7P synthase was expressed in *E. coli*, no soluble recombinant protein was observed and the supernatant fraction showed no detectable enzymatic activity. Alteration of growth temperature and lysis conditions had no effect on the protein solubility. We therefore used co-expression with *E. coli* GroESL to solublize DAH7P synthase, following our previous successful solubilization of the type II DAH7P synthase from *H. pylori* (CJW and EJP, unpublished observations). Although the majority of the protein remained insoluble, approximately 30% of the total protein expressed was found in the supernatant following lysis and centrifugation and this exhibited DAH7P synthase activity. The incorporation of PEP, $MnSO_4$, TCEP and NaCl in all purification buffers was essential in order to prevent aggregation of the protein. The molecular weight of the recombinant protein after cleavage with rTEV was 50 kDa and the protein preparation was essentially pure (>95%) as estimated by SDS–PAGE. Crystallization trials were performed using protein concentrations no greater than 5 mg ml^{-1}, as protein precipitation was observed at higher concentrations. The crystals were found to be trigonal, with systematic absences consistent with the space group $P3_121$ or $P3_221$. The unit-cell parameters, $a = b = 203.61$, $c = 66.39$ Å, are consistent with the presence of two or three molecules in the crystal asymmetric unit, corresponding to values of the Matthews coefficient $V_{\rm M}$ of 3.9 and 2.6 Å^3 Da^{-1} and solvent contents of 68 and 53%, respectively. Size-exclusion chromatography indicates that *M. tuberculosis* DAH7P synthase behaves as a dimer in solution (data not shown).

Table 1
Data-collection statistics for DAH7P synthase.

Values in parentheses are for the outermost shell.

Space group	$P3_121$ or $P3_221$
Unit-cell parameters (Å)	$a = b = 203.61$, $c = 66.39$
Resolution range (Å)	39.37–2.50 (2.59–2.50)
Wavelength (Å)	1.542
No. of measured reflections	416784
No. of unique reflections	54710
Multiplicity	7.62 (7.54)
Completeness (%)	100 (100)
Average $I/\sigma(I)$	13.6 (5.4)
$R_{\rm merge}$† (%)	9.9 (34.7)

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

Molecular replacement was attempted using *E. coli* type I DAH7P synthase as a model and found to be unsuccessful. We are currently trying to solve the structure of *M. tuberculosis* DAH7P synthase by multiwavelength anomalous diffraction (MAD) methods using SeMet-substituted protein, given that the DAH7P synthase monomer contains 13 methionines.

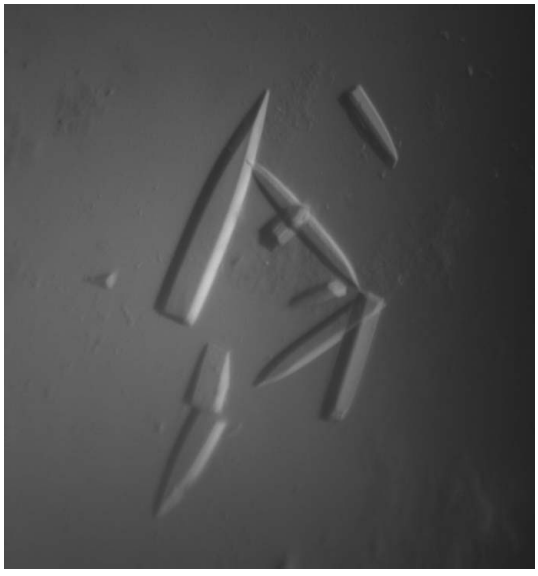


Figure 2
Crystals of *M. tuberculosis* DAH7P synthase.

We thank Mark Patchett for the pGroESL plamid and Geoffrey Jameson and Gillian Norris for assistance in obtaining the native data set. CJW is a recipient of a Massey University Vice Chancellor's scholarship. This work was funded by the RSNZ Marsden Fund (MAU008) and the Centre for Molecular Biodiscovery.

References

Bentley, R. (1990). *Crit. Rev. Biochem. Mol. Biol.* **25**, 307–384.
Dyer, W. E., Weaver, L. M., Zhao, J., Kuhn, D. N., Weller, S. C. & Herrmann, K. M. (1990). *J. Biol. Chem.* **265**, 1608–1614.
Goloubinoff, P., Gatenby, A. A. & Lorimer, G. H. (1989). *Nature (London)*, **337**, 44–47.
Gosset, G., Bonner, C. A. & Jensen, R. A. (2001). *J. Bacteriol.* **183**, 4061–4070.
Hartmann, M., Schneider, T. R., Pfeil, A., Heinrich, G., Lipscomb, W. N. & Braus, G. H. (2003). *Proc. Natl Acad. Sci. USA*, **100**, 862–867.
Hendrickson, W. A., Horton, J. R. & LeMaster, D. M. (1990). *EMBO J.* **9**, 1665–1672.
Jensen, R. A., Xie, G., Calhoun, D. H. & Bonner, C. A. (2002). *J. Mol. Evol.* **54**, 416–423.
Kingston, R. L., Baker, H. M. & Baker, E. N. (1994). *Acta Cryst.* D**50**, 429–440.
Konig, V., Pfeil, A., Braus, G. H. & Schneider, T. R. (2004). *J. Mol. Biol.* **337**, 675–690.
Parish, T. & Stoker, N. G. (2002). *Microbiology*, **148**, 3069–3077.
Parrish, N. M., Dick, J. D. & Bishai, W. R. (1998). *Trends Microbiol.* **6**, 107–112.
Rattan, A., Kalia, A. & Ahmad, N. (1998). *Emerging Infect. Dis.* **4**, 195–209.
Schofield, L. R., Anderson, B., Norris, G. E., Patchett, M. L., Jameson, G. B. & Parker, E. J. (2005). In preparation.
Shumilin, I. A., Bauerle, R. & Kretsinger, R. H. (2003). *Biochemistry*, **42**, 3766–3776.

Shumilin, I. A., Bauerle, R., Wu, J., Woodard, R. W. & Kretsinger, R. H. (2004). *J. Mol. Biol.* **341**, 455–466.
Shumilin, I. A., Kretsinger, R. H. & Bauerle, R. H. (1999). *Structure*, **7**, 865–875.
Silakowski, B., Kunze, B. & Muller, R. (2000). *Arch. Microbiol.* **173**, 403–411.
Stokstad, E. (2000). *Science*, **287**, 2391.
Stura, E. A., Nemerow, G. R. & Wilson, I. A. (1992). *J. Cryst. Growth*, **122**, 273–285.
Walker, G. E., Dunbar, B., Hunter, I. S., Nimmo, H. G. & Coggins, J. R. (1996). *Microbiology*, **142**, 1973–1982.
Zhao, J. & Herrmann, K. M. (1992). *Plant Physiol.* **100**, 1075–1076.