Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

Structural and functional studies of mycobacterial IspD enzymes

Christofer Björkelid,[a] Terese Bergfors,[a] Lena M. Henriksson,[a] Ana Laura Stern,[b] Torsten Unge,[a] Sherry L. Mowbray[a,b] and T. Alwyn Jones[a]*

[a]Department of Cell and Molecular Biology, Uppsala University, Biomedical Center, Box 596, SE-751 24 Uppsala, Sweden, and [b]Department of Molecular Biology, Swedish University of Agricultural Sciences, Biomedical Center, Box 590, SE-751 24 Uppsala, Sweden

Correspondence e-mail: alwyn@xray.bmc.uu.se

Received 17 December 2010
Accepted 18 February 2011

PDB References: *Ms*IspD–CTP, 2xwl; *Ms*IspD–CMP, 2xwm; *Mt*IspD–CTP, 2xwn.

A number of pathogens, including the causative agents of tuberculosis and malaria, synthesize isopentenyl diphosphate *via* the 2-*C*-methyl-D-erythritol 4-phosphate (MEP) pathway rather than the classical mevalonate pathway found in humans. As part of a structure-based drug-discovery program against tuberculosis, IspD, the enzyme that carries out the third step in the MEP pathway, was targeted. Constructs of both the *Mycobacterium smegmatis* and the *Mycobacterium tuberculosis* enzymes that were suitable for structural and inhibitor-screening studies were engineered. Two crystal structures of the *M. smegmatis* enzyme were produced, one in complex with CTP and the other in complex with CMP. In addition, the *M. tuberculosis* enzyme was crystallized in complex with CTP. Here, the structure determination and crystallographic refinement of these crystal forms and the enzymatic characterization of the *M. tuberculosis* enzyme construct are reported. A comparison with known IspD structures allowed the definition of the structurally conserved core of the enzyme. It indicates potential flexibility in the enzyme and in particular in areas close to the active site. These well behaved constructs provide tools for future target-based screening of potential inhibitors. The conserved nature of the extended active site suggests that any new inhibitor will potentially exhibit broad-spectrum activity.

1. Introduction

Isopentenyl diphosphate (IPP) is an activated isoprene unit that is a key building block of isoprenoids, a large class of compounds that are essential to all living organisms. In the classical mevalonate pathway, activated acetate in the form of acetyl coenzyme A is converted to IPP in a four-step process that proceeds *via* the intermediate mevalonate (Beytía & Porter, 1976). The synthesis of mevalonate is the committed step in cholesterol formation and the enzyme involved is the site of action of the statin class of drugs that reduce plasma cholesterol levels. The discovery of an alternative (non-mevalonate or MEP) pathway (Rohmer, 1999) in Gram-negative and some Gram-positive bacteria, as well as in plant chloroplasts and algae, was met with some excitement, since a number of serious pathogens synthesize IPP by this pathway alone. The conversion of pyruvate and glyceraldehyde 3-phosphate to IPP in seven steps thus offers the opportunity to develop a new series of drugs and herbicides targeting enzymes that are absent in humans. This hope was strengthened when it became clear that the antimalarial drug fosmidomycin targets DXR/IspC (Kuzuyama *et al.*, 1998; Zeidler *et al.*, 1998), the enzyme that catalyses the second, committed step in the pathway in which 1-deoxy-D-xylulose 5-phosphate (DXP) is converted to 2-*C*-methyl-D-erythritol 4-

phosphate (MEP). The antibacterial activity of fosmidomycin, however, was discovered before the MEP pathway was identified (Shigi, 1989). A great deal of effort has gone into characterizing the MEP pathway to create the framework necessary for target-based drug discovery, but so far it has met with limited success in the development of new medically relevant inhibitors.

Approximately one third of the population of the world is infected with *Mycobacterium tuberculosis* (*Mtb*), the causative agent of tuberculosis (TB), and it is estimated that two million people die every year from the disease. It has been fifty years since the discovery of rifampicin (Sensi *et al.*, 1959), one of the first-line drugs that are now becoming ineffective as drug-resistant forms of the disease emerge. Ten candidate drugs are in clinical trials, but the drug-discovery pipeline needs new compounds and new targets to meet future needs for treatment (Ma *et al.*, 2010). Because isoprenoid synthesis in mycobacteria takes place *via* the MEP pathway, this pathway is an attractive source of potential targets for drug discovery against TB. The lack of success in turning DXR inhibitors into compounds with activity towards *Mtb*, despite a well defined high-resolution structure of the fosmidomycin–DXR complex (Henriksson *et al.*, 2007), has been attributed to poor uptake (Brown & Parish, 2008) and has shifted attention to other targets in the pathway. IspD (EC 2.7.7.60) is a cytidyl transferase that catalyzes the third reaction in the pathway, in which MEP is converted to 4-diphosphocytidyl-2-*C*-methyl-D-erythritol (CDP-ME; Fig. 1). The essentiality of IspD in *Mtb* has been confirmed both by transposon mutagenesis (Sassetti *et al.*, 2003) and by more focused studies (Eoh *et al.*, 2007). As a first step in a target-based approach to drug discovery, we needed to be able to produce active enzyme and determine its crystal structure. We employ a general strategy of using both *Mtb* and the closely related *M. smegmatis* (*Ms*) enzymes for structural and functional studies because of the usually better solubility of the *Ms* enzymes. In the present paper, we present our initial structural and functional studies on IspD from both *Mtb* (*Mt*IspD, 231 amino acids, 24 041.4 Da) and *Ms* (*Ms*IspD, 222 amino acids, 22 377.6 Da); the two proteins share 64% sequence identity with each other.

2. Methods

2.1. Cloning and expression

Genes encoding *Ms*IspD (MSMEG_6076) and *Mt*IspD (*Rv3582c*) were PCR-amplified from genomic DNA and the GTG start codons of each were mutated to ATG in the process. Both genes were designed with C-terminal truncations consisting of three residues for *Ms*IspD (Arg220, Gly221 and Ala222) and two residues for *Mt*IspD (Arg230 and Gly231); a C-terminal His_4 tag was appended. The resulting constructs were ligated into the pEXP5-CT/TOPO expression vector (Invitrogen) and used to transform *Escherichia coli* BL21-AI competent cells (Invitrogen). Cells expressing *Ms*IspD were grown at 310 K until an OD_{600} of 1 was reached in growth medium (10 g l^{-1} yeast extract, 10 g l^{-1} tryptone, 2 g l^{-1} NaCl, 3 g l^{-1} Na_2HPO_4, 2 g l^{-1} glycerol) with 100 µg ml^{-1} ampicillin. Protein expression was induced by adding 0.2 g l^{-1} L-arabinose and the cells were further incubated at 310 K for 3 h. Cells expressing *Mt*IspD were grown at 310 K until an OD_{600} of 0.6 was reached in Luria–Bertani medium with 100 µg ml^{-1} ampicillin. Protein expression was induced by adding 0.2 g l^{-1} L-arabinose and the cells were further incubated at 295 K for 12 h.

2.2. Protein purification

The cells were harvested by centrifugation and resuspended in lysis buffer (100 m*M* HEPES, 200 m*M* NaCl, 50 m*M* NaH_2PO_4, 50 m*M* Na_2SO_4, 10 m*M* imidazole, 0.5%(*v*/*v*) Triton X-100, final pH 7.5) with 0.05 mg ml^{-1} lysozyme, 0.01 mg ml^{-1} RNase, 0.02 mg ml^{-1} DNase I and 0.02 mg ml^{-1} phenylmethylsulfonyl fluoride. Cells were lysed in a cell disruptor (Constant Systems Ltd, UK) and cell debris was removed by centrifugation. The His-tagged proteins were purified from the supernatant on an Ni^{2+}–NTA (Qiagen) column and immediately precipitated by adding ammonium sulfate to a final concentration of 3 *M*. The sample was centrifuged and the resulting protein pellet was dissolved in 100 m*M* HEPES, 10%(*w*/*v*) glycerol, 200 m*M* Na_2SO_4, 20 m*M* $MgSO_4$ (final pH 7.0). The proteins were further purified using a Superdex 75 (GE Healthcare) size-exclusion column. Fractions containing protein of the correct size were pooled and determined to be ~99% pure by SDS–PAGE. The His_4 tag was retained on both proteins during all further studies.

2.3. Enzymatic assay

The activity of *Mt*IspD in the conversion of MEP and CTP to CDP-ME and PP_i was evaluated in a coupled assay (Bernal *et al.*, 2005). When the product PP_i is hydrolysed by inorganic pyrophosphatase, the P_i forms a complex with malachite green that can be detected at 620 nm.

A dephosphorylation method (Kötting *et al.*, 2009) was used to estimate the concentration of the substrate MEP (generously provided by AstraZeneca India Pvt. Ltd). This experiment was performed using the same setup as the assay described below. Each reaction contained 150 U ml^{-1} antarctic phosphatase (New England BioLabs), 1× antarctic phosphatase buffer (New England BioLabs) and MEP (five different concentrations from 5 to 86 µ*M*). Corresponding blanks were performed without antarctic phosphatase. The reactions were incubated at 310 K and samples were taken after 2 and 4 h to confirm that the reaction had progressed to completion. The MEP concentration was thus estimated on the basis of the amount of phosphate produced, assuming 95% conversion, as described by Kötting *et al.* (2009).

The inorganic pyrophosphatase-coupled assay was conducted at 293 K. The standard reaction mixture consisted of 50 m*M* HEPES–KOH pH 8.0, 10 m*M* $MgCl_2$, 0.2 U ml^{-1} inorganic pyrophosphatase (Roche), MEP, CTP (Fluka) and 20 n*M* *Mt*IspD (as determined by Bradford analysis using bovine serum albumin as standard). When varying MEP (nine different concentrations from 5.7 to 513 µ*M* after correction),

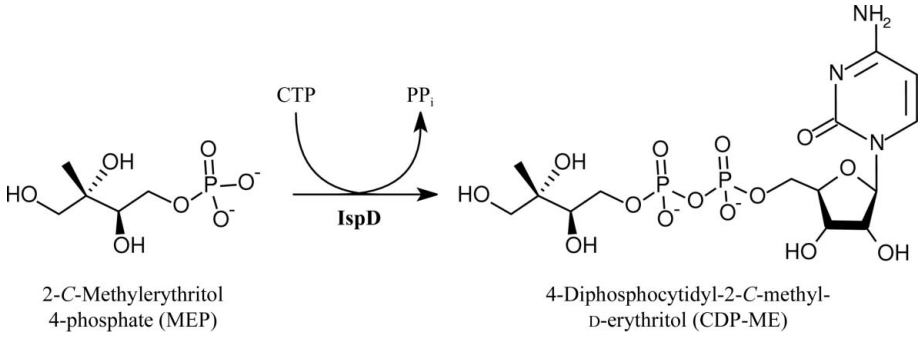

Figure 1
The third step in the MEP pathway catalysed by IspD.

the CTP concentration was kept at 250 μ*M*. When varying CTP (ten different concentrations from 30 to 1200 μ*M*), the MEP concentration was kept at 428 μ*M*.

Each reaction was performed in a total volume of 300 µl and each corresponding blank reaction, which did not contain *Mt*IspD, was performed in a total volume of 180 µl. All components except the enzyme were mixed and incubated for 10 min. The reactions were then initiated by adding *Mt*IspD (or 50 m*M* HEPES–KOH pH 8.0 in the case of the blanks).

For each separate experiment, duplicate samples of 60 µl were taken from the reactions containing enzyme and a single sample from the blank reactions at two different time points (25 and 50 min) in order to ensure that initial velocities were being measured. To stop the reaction, each 60 µl sample was mixed with 15 µl malachite green solution (P_i Colorlock Gold mix, Innova Biosciences) in a 96-well plate (SARSTEDT). After 5 min, 5 µl stabilizer (Innova Biosciences) was added to each well; after an additional 30 min, 120 µl H_2O was also added before measuring the absorbance using a spectrophotometric plate reader (Envision 2140 Multilabel Reader, PerkinElmer). The production of inorganic phosphate was quantified by comparing the absorbance with that of a calibration curve for P_i (standard solution provided with the P_i Colorlock kit).

The K_m and V_{max} values were calculated by linear regression of the data plotted in the Lineweaver–Burk plot. The kinetic parameters presented for MEP and CTP are the averages from two and three separate experiments, respectively.

2.4. Crystallization

Crystallization trials were performed using the sitting-drop vapour-diffusion method at 293 K. An Oryx 6 robot (Douglas Instruments, UK) dispensed droplets (300 nl protein solution and 300 nl screening/reservoir solution) into MRC 2 Well plates (Molecular Dimensions, UK). The droplets were equilibrated against 80 µl reservoir solution.

Despite extensive crystallization screening of the full-length *Ms*IspD and *Mt*IspD proteins, no crystals were obtained. Therefore, slightly truncated versions of the proteins (see §2.1) were constructed: residues 1–219 for *Ms*IspD and residues 1–229 for *Mt*IspD. The C-terminal His_4 tag was retained on both proteins for crystallization screening. As with the full-length proteins, we screened the apo forms of the truncated constructs and as cocrystallizations with various ligands (CMP, CDP and CTP) with and without erythritol. The cofactor $MgCl_2$ was always included in the protein buffer at a molar excess (1–10 m*M*, depending on the protein concentration).

Crystals of *Ms*IspD (at 40 mg ml^{-1} in 20 m*M* Bis-Tris propane pH 6.5, 0.1 *M* NaCl with 10 m*M* CTP and 1 m*M* $MgCl_2$) appeared in several conditions from the Morpheus screen (Molecular Dimensions, UK; Gorrec, 2009). The best diffracting crystals grew in condition E2, which consisted of 10%(*w*/*v*) PEG 8000, 20%(*v*/*v*) ethylene glycol, 0.03 *M* each of diethylene glycol, triethylene glycol, tetraethylene glycol and pentaethylene glycol and 0.1 *M* MES–imidazole buffer pH 6.5. Crystals as large as 0.3 × 0.2 × 0.2 mm grew overnight.

*Ms*IspD (in the same buffer as above) also crystallized in the presence of 10 m*M* CMP, 10 m*M* $MgCl_2$ and 10 m*M* erythritol. These crystals were of similar dimensions and morphology. They appeared overnight in Morpheus condition B5 consisting of 10%(*w*/*v*) PEG 20 000, 20%(*v*/*v*) monomethyl ether PEG 550, 0.03 *M* NaF, 0.03 *M* NaI, 0.03 *M* NaBr and 0.1 *M* MOPS–Na HEPES pH 7.5.

Crystals of *Mt*IspD (at 10 mg ml^{-1} in 20 m*M* Bis-Tris propane pH 6.5, 0.1 *M* NaCl with 10 m*M* CTP and 1 m*M* $MgCl_2$) also grew from the Morpheus screen. Crystals appeared after one week as clusters of rods with maximum dimensions 0.2 × 0.1 × 0.1 mm from condition A9 consisting of 10%(*w*/*v*) PEG 20 000, 20%(*v*/*v*) monomethyl ether PEG 550, 0.03 *M* $CaCl_2$, 0.03 *M* $MgCl_2$ and 0.1 *M* bicine–Trizma base pH 8.5.

The crystals were harvested directly from the crystallization plate and vitrified in liquid nitrogen without further cryoprotection.

2.5. Data collection and processing

Diffraction data for the *Ms*IspD–CTP complex were collected using our in-house X-ray source (Rotaflex RTP300 RC, Rigaku) to a resolution of 2.4 Å and were used to solve the initial structure by molecular replacement (data not shown). A higher resolution data set for the same complex was collected from a single fresh crystal at the European Synchrotron Radiation Facility (ESRF), as were diffraction

data for the *Ms*IspD–CMP and *Mt*IspD–CTP complexes. Diffraction data from the *Ms*IspD crystals were indexed and integrated using *MOSFLM* (Leslie, 2006) and scaled with *SCALA* (Evans, 2006), which are part of the *CCP*4 package (Winn *et al.*, 2011). Diffraction data from the *Mt*IspD crystal were indexed and integrated using *XDS* and scaled with *XSCALE*, which are part of the *XDS* program package (Kabsch, 2010). Data-collection statistics are shown in Table 1.

2.6. Structure determination and refinement

The structure of the *Ms*IspD–CTP complex was solved by molecular replacement with the program *Phaser* (McCoy *et al.*, 2007) using a search model that consisted of a truncated polyalanine model of an *E. coli* IspD (*Ec*IspD; PDB code 1vgt; Badger *et al.*, 2005) subunit, from which we omitted residues equivalent to Gly11–Ala20 and Leu129–Ala155. The structures of the *Ms*IspD–CMP complex and the *Mt*IspD–CTP complex were solved by molecular replacement with *Phaser* using the *A* chain of the refined *Ms*IspD–CTP structure as a search model. The three structures were improved by multiple alternating cycles of crystallographic refinement with *REFMAC*5 (Murshudov *et al.*, 2011) and interactive rebuilding using *O* (Jones *et al.*, 1991). Waters were included using the Water_add command in *O*, which makes use of the average carbonyl O atom three-dimensional electron-density profile. Because of the limited resolution of the *Mt*IspD–CTP complex, tight noncrystallographic (NCS) restraints were applied to the compact domain (*i.e.* excluding the β-meander 137–161 and the poorly defined active-site loop 180–191) and a suitable relative weighting of diffraction data and restraint terms was determined to limit overfitting during refinement, in which we aimed to keep $R_{\rm free}$ constant while reducing the difference between R and $R_{\rm free}$. Final refinement statistics are shown in Table 1.

2.7. Other methods

Structural figures were prepared in *O* and rendered in *Molray* (Harris & Jones, 2001). Secondary-structure assignments were made with the Yasspa command in *O* and then edited with artistic license so as to more clearly show only the longest strands and helices. Overall structural similarities were identified with *DALI* (Holm & Park, 2000). Detailed structural comparisons were made using the Lsq commands in *O* with default C^{α} matching-pair cutoffs of 3.8 Å or a close-pair cutoff of 1.0 Å (Kleywegt & Jones, 1997). Sequence alignment was performed with *PRALINE* (Simossis & Heringa, 2005) and the corresponding figure was produced with *ALINE* (Bond & Schüttelkopf, 2009). Homology models of putative *Pf*IspD domains were produced using the *Phyre* server (Kelley & Sternberg, 2009). Solvent-accessible area calculations were performed with *AREAIMOL*, which is part of the *CCP*4 package of programs (Winn *et al.*, 2011).

Table 1
Data-collection and refinement statistics.

Values in parentheses are for the outer resolution shell.

	*Ms*IspD–CTP	*Ms*IspD–CMP	*Mt*IspD–CTP
PDB code	2xwl	2xwm	2xwn
Data-collection statistics			
Beamline	ID14-4, ESRF	ID14-2, ESRF	ID14-1, ESRF
Detector	ADSC Q315r	ADSC Q4r	ADSC Q210
Wavelength (Å)	1.0045	0.9330	0.9334
Space group	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$
Unit-cell parameters (Å)	a = 45.3, b = 73.0, c = 117.2	a = 45.6, b = 71.7, c = 111.9	a = 44.1, b = 82.3, c = 133.2
$V_{\rm M}$† (Å^3 Da^{-1})	2.17	2.04	2.51
Resolution range (Å)	38.50–1.49 (1.57–1.49)	28.00–1.80 (1.90–1.80)	42.00–2.90 (2.98–2.90)
No. of reflections measured	491940	242067	79945
No. of unique reflections	64486	34794	11201
Average multiplicity	7.6 (7.5)	7.0 (7.0)	7.1 (7.3)
Completeness (%)	99.6 (99.5)	99.9 (99.9)	99.2 (98.9)
$R_{\rm merge}$‡§	0.062 (0.408)	0.086 (0.555)	0.129 (0.491)
$R_{\rm meas}$¶	0.067 (0.438)	0.094 (0.599)	0.135 (0.572)
$R_{\rm p.i.m.}$††	0.024 (0.157)	0.035 (0.225)	—
$\langle I/\sigma(I)\rangle$‡	18.6 (5.3)	16.2 (3.5)	16.2 (4.0)
Refinement statistics			
Resolution range (Å)	38.50–1.49 (1.53–1.49)	28.00–1.80 (1.85–1.80)	42.00–2.90 (2.98–2.90)
Reflections used in working set	61154 (4441)	32984 (2403)	10641 (749)
Reflections for $R_{\rm free}$ calculation	3262 (239)	1750 (115)	560 (39)
R‡ (%)	18.5 (22.9)	18.1 (23.2)	20.8 (31.5)
$R_{\rm free}$ (%)	21.0 (25.4)	21.6 (30.0)	25.3 (35.2)
No. of non-H atoms	3722	3489	3347
No. of solvent waters	469	273	0
Mean B factors (Å^2)			
Protein (chain A/B)	18.6/25.0	18.1/27.2	28.4/37.7
Solvent	31.7	31.3	—
Mg^{2+} (chain A/B)	15.1/27.6	—	15.1/30.9
CTP (chain A/B)	13.2/28.2	—	20.5/43.1
CMP (chain A/B)	—	16.9/35.5	—
Ramachandran outliers‡‡ (chain A/B) (%)	1.0/0.5	0.5/3.0	3.6/5.4
R.m.s.d. from ideal bond lengths§§ (Å)	0.008	0.009	0.009
R.m.s.d. from ideal bond angles§§ (°)	1.290	1.246	1.344

† Matthews (1968). ‡ Merging and crystallographic R factors were generated using the programs described in §2. § $R_{\rm merge} = \sum_{hkl}\sum_i |I_i(hkl) - \langle I(hkl)\rangle| / \sum_{hkl}\sum_i I_i(hkl)$. ¶ $R_{\rm meas} = \sum_{hkl}[N/(N-1)]^{1/2}\sum_i |I_i(hkl) - \langle I(hkl)\rangle| / \sum_{hkl}\sum_i I_i(hkl)$. †† $R_{\rm p.i.m.} = \sum_{hkl}[1/(N-1)]^{1/2}\sum_i |I_i(hkl) - \langle I(hkl)\rangle| / \sum_{hkl}\sum_i I_i(hkl)$. ‡‡ Calculated using a strict-boundary Ramachandran plot definition (Kleywegt & Jones, 1996). §§ Ideal values from Engh & Huber (1991).

3. Results

3.1. Enzyme activity

Activity was measured using an inorganic pyrophosphatase-coupled assay (Bernal *et al.*, 2005), with MEP concentrations validated in a separate phosphatase-based procedure. Under the conditions used, *Mt*IspD was estimated to have a $K_{\rm m}$ value of 43.5 μM and a $k_{\rm cat}$ value of

Table 2
Kinetic parameters for *Mt*IspD and *Ec*IspD.

	MEP			CTP			
	K_m (μM)	k_{cat} (min^{-1})	k_{cat}/K_m (mM^{-1} min^{-1})	K_m (μM)	k_{cat} (min^{-1})	k_{cat}/K_m (mM^{-1} min^{-1})	Reference
*Mt*IspD	43.5 ± 3.3	5.4 ± 0.8	125	45.5 ± 5.4	6.3 ± 1.4	138	This study
*Mt*IspD†	43	3.4‡	79	92	3.4‡	37	Shi *et al.* (2007)
*Mt*IspD§	58.5 ± 5.4	0.72	12.3	53.2 ± 4.5	1.0	18.8	Eoh *et al.* (2007)
*Ec*IspD	3.14	598¶	190446	131	598¶	4565	Rohdich *et al.* (1999)
*Ec*IspD	32 ± 3	1008 ± 12	31500	ND††	ND	ND	Cane *et al.* (2001)
*Ec*IspD	370 ± 60	2904 ± 660	7849	760 ± 60	3246 ± 1680	4271	Richard *et al.* (2004)
*Ec*IspD	61 ± 14	ND	ND	58 ± 6	ND	ND	Bernal *et al.* (2005)

† The standard assay system contained 2 m*M* $MgCl_2$ and the concentrations of MEP and CTP were 2 m*M* while varying the other substrate. ‡ The single reported k_{cat} value was used for calculating k_{cat}/K_m values for both MEP and CTP. § The standard assay system contained 10 m*M* $MgCl_2$ and the concentrations of MEP and CTP were 0.1 m*M* while varying the other substrate. ¶ The k_{cat} value was calculated from the published V_{max} value, assuming a molecular weight of 26 kDa for the enzyme. †† Not determined.

5.4 min^{-1} when MEP was the variable substrate and a K_m value of 45.5 μM and a k_{cat} value of 6.3 min^{-1} when CTP was the variable substrate (see Fig. 2 and Table 2).

3.2. General comments on structural work

In our structural studies of mycobacterial enzymes, we have adopted a standard approach of studying both the *Mtb* enzyme and its *Ms* homologue. The enzymes usually share high sequence homology, but the *Ms* enzymes are often easier to work with, in particular because they are generally more soluble. In the case of IspD, however, we were unable to produce suitable crystals from either the full-length *Ms*IspD or *Mt*IspD enzymes. We therefore designed, cloned and expressed constructs with short C-terminal truncations based on sequence–structure alignments (Fig. 3). However, crystallization trials with these new constructs were unsuccessful until we carried out a buffer-optimization screen (Jancarik *et al.*, 2004). On the basis of this screen, the buffers for both IspD proteins were changed to Bis-Tris propane before concentration and crystallization screening. In this new buffer, both proteins crystallized readily in the Morpheus screen. A buffer-optimization screen had also been performed on the full-length versions of the proteins. We have to date been unable to crystallize any apo form of either the truncated or full-length proteins. It therefore seems that the truncations in the constructs, the inclusion of CTP or CMP and the optimized protein buffer were all critical in obtaining crystals.

Crystal structures of *Ms*IspD have been determined in complex with CTP and with CMP. They were refined to crystallographic *R* factors of 18.5 and 18.1% at resolutions of 1.5 and 1.8 Å, respectively (Table 1). The natural sequence has been truncated by three residues at the C-terminus but was then extended with a His_4 tag. Both structures contained a dimer in the asymmetric unit and although both complexes crystallize in space group $P2_12_12_1$ they are non-isomorphous. The electron density of the CTP complex is well defined except for an active-site loop (residues 17–19) in the *B* chain, another loop nearby (residues 176–178 of both chains), the N-terminal methionine and the C-terminal histidine of the His_4 tag. Fig. 4 illustrates the quality of the electron density for this structure. In the CMP complex the loop around residue Ala176 is well defined in the *A* chain but not in the *B* chain. All four residues of the His tag have clear density in the *B* chain, while two are defined in the *A* chain; the N-terminal methionine, as expected, is missing in both chains. The temperature-factor distribution over the dimers is asymmetric in both crystal forms, such that the *A* chain has an average value that is ∼7 $Å^2$ lower. Each dimer is formed by an almost exact twofold axis and can be aligned with a root-mean-square deviation (r.m.s.d.) of 0.81 Å using all 442 C^α atoms for the CTP complex, for example. The 221 equivalent pairs of C^α atoms are related by a 178.2° rotation and can be

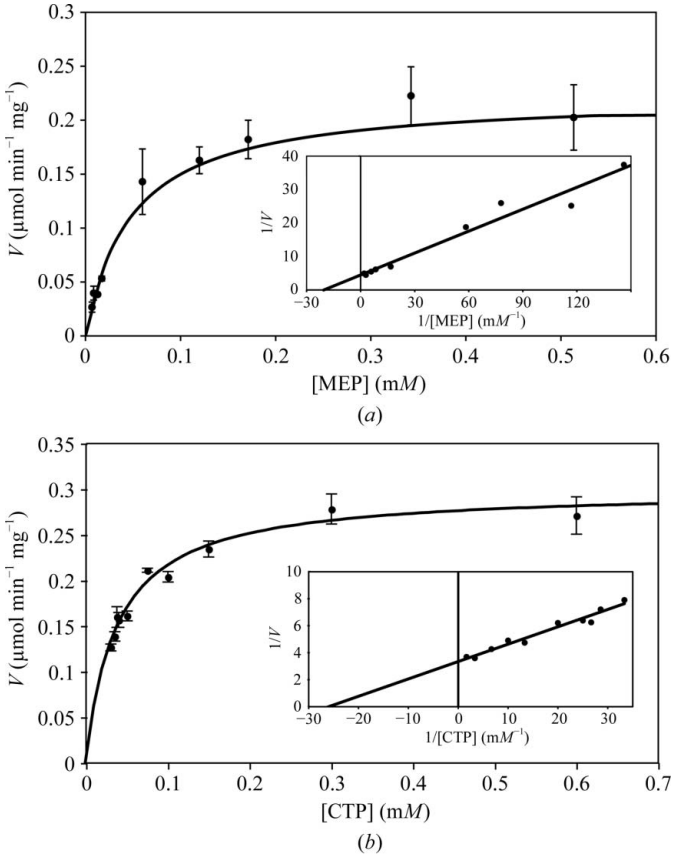

Figure 2
Enzymatic characterization of *Mt*IspD. Direct plot of velocity *versus* substrate concentration for a representative experiment. A curve was fitted according to the Michaelis–Menten equation. The inset shows data points plotted according to the Lineweaver–Burk method, where the linear fit to the data was used to calculate k_{cat} and K_m values. (*a*) MEP variation. The CTP concentration was held at 250 μM while varying the MEP concentration. (*b*) CTP variation. The MEP concentration was held at 428 μM while varying the CTP concentration.

superimposed with an r.m.s.d. of 0.68 Å. NCS breaks down in some side-chain conformations, usually owing to poor electron density (*e.g.* the side chain of Arg150 is disordered beyond the C^{γ} atom in the *A* chain of the CTP complex but is well defined in the *B* chain, where it stretches across the active site to the MEP-binding site). Pro159 has a well defined *cis* conformation in all chains. The few outliers from our strict-boundary Ramachandran definition (Kleywegt & Jones, 1996) are mostly localized to the relatively poorly defined loop around residue Ala176. However, residue Gln157 is a slight outlier in all four *Ms*IspD chains that is in well defined electron density. This conformation allows the main-chain carbonyl O atom to accept a hydrogen bond from the side-chain amide O atom of the highly conserved (Fig. 3) but buried residue Gln160.

*Mt*IspD has been crystallized in complex with CTP, again in a non-isomorphous $P2_12_12_1$ space group, and the structure was refined to an *R* factor of 20.8% at 2.9 Å resolution (Table 1). This also corresponds to a slightly truncated construct (residues 1–229) with a His_4 tag at the C-terminus and a dimer in the asymmetric unit. The electron density is well defined in both chains from residues 5 to 229 except for one of the active-site loops, in which we were unable to trace the segment 183–188; the last two residues of the His_4 tag are also missing from the model. Our final model includes a CTP and magnesium ion in each active site but, because of the limited resolution, no water molecules have been included. The complete dimer comparison, while showing an apparently exact twofold, gives an r.m.s.d. of 1.37 Å for 434 pairs of C^{α} atoms. The chain-to-chain superposition, however, shows a true rotation of 175.5° and an r.m.s.d. of 0.72 Å for the 221 equivalent C^{α} pairs. This crystal also shows temperature-factor asymmetry, particularly in that the CTP and interacting loops have lower values in the *A* chain: ∼20 $Å^2$ *versus* ∼40 $Å^2$.

3.3. *Ms*IspD structures

As expected, the structure of *Ms*IspD is similar to that of *Ec*IspD (Richard *et al.*, 2001), with a single compact domain from which a long β-meander extends (see Fig. 5). While each active site is created principally by residues from one compact domain (Fig. 5*a*), some contributions also come from the twofold-related β-meander, which is crucial to dimer formation (Fig. 5*b*). The compact domain is a variation on the Rossmann fold (Rossmann *et al.*, 1975), with a classical nucleotide-binding unit combined with an additional strand, the sixth strand, which runs antiparallel to the rest of the strands in the β-sheet. The seven-stranded sheet therefore has mixed directionality ordered 7-5-6′-4-1-2-3 (Richard *et al.*, 2001); the β-meander lies between the antiparallel strands 5 and 6 and consists of root, stem and tip regions. The meanders are critical to dimer formation because they intertwine around the local twofold axis. Each meander is involved in the formation of three well packed volume elements. Each tip and stem interacts with a twofold-related root and the final strand of the β-sheet to form a five-stranded β-barrel closed by the edge of helix α6. This volume is lined with branched-chain amino acids (Val126*A*, Val138*B*, Val144*B*, Val156*A*, Val193*A* and Ile195*A*) but also contains a cluster of polar side chains centred on Lys136*B* that interacts with the side chains of Asp181*A*, Glu187*A* and Thr147*B* (Fig. 6). Twofold-related meander stems and roots interact to form another β-barrel containing branched side chains (Ile135 and Leu153) and proline rings (Pro131 and Pro148) (Fig. 6). The fourth major set of interactions that stabilize dimer formation involves a hydrophobic patch on the C-terminal helix α7, where the side chains of Leu213 and Ala216 lie close to the

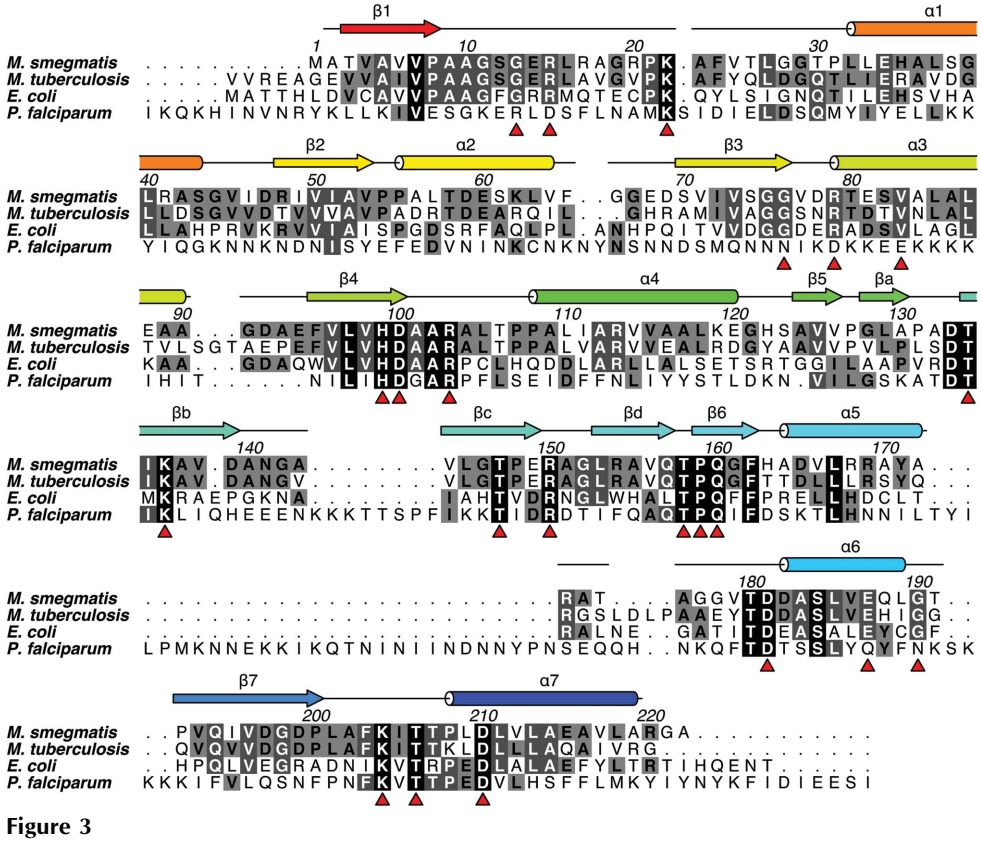

Figure 3
Sequence alignments. Sequences of *Ms*IspD, *Mt*IspD, *Ec*IspD and *Pf*IspD (UniProt identifiers A0R560, C6DMN4, C6UC96 and Q8I273, respectively) were aligned with *PRALINE* (Simossis & Heringa, 2005) and edited by hand to match the structural similarity where appropriate. We have included the secondary-structure definitions used in the main text, which were generated as described in §2. Secondary-structure elements are rainbow-coloured, as in chain *A* in the following figures. Numbering follows the *M. smegmatis* sequence. Only a portion of the *Pf*IspD sequence is shown. Red triangles indicate the 22 residues highlighted by Kemp *et al.* (2003).

twofold axis. In total, dimer formation results in a net loss in solvent-accessible surface area of ∼3200 Å^2.

Pairs of individual *Ms*IspD subunits can be superimposed with r.m.s.d.s of ∼0.8 Å for ∼216 C$^\alpha$ pairs (Fig. 7*a*). The only major difference is localized to the CMP-binding site in the *B* chain (Fig. 7*b*), which is described in more detail below. Smaller breakdowns in NCS are located in some of the connecting loops between secondary-structural elements (near residues Glu68, Ala176 and Gly190). Despite clear differences in crystallographic packing, the overall structures of the complete dimers in the CTP-bound and CMP-bound structures are very similar, with an r.m.s.d. of 0.81 Å over 434 C$^\alpha$ atoms when using a standard 3.8 Å pairwise cutoff.

CMP and CTP bind at the C-terminal edge of the main β-sheet in a broad depression created by elements from both chains in the dimer. In the CTP complex, the cytosine base slots between the flat peptide planes of residues 10–11 and 78–79 found at the end of the first and third strands, respectively. Hydrogen bonds between the base and protein are made *via* main-chain (O2 to the amide N atom of residue Ala10, N4 to the carbonyl O atom of residue Gly76) and side-chain atoms (N3 to the hydroxyl of Ser82). The ribose has a C2-*endo* conformation, with the hydroxyl groups forming hydrogen bonds to main-chain atoms (O2′ to amide nitrogen 11, O3′ to carbonyl oxygen 8) and waters. In the CTP complex each triphosphate wraps around a magnesium ion, forming an octahedral coordination that is completed by three water molecules (Figs. 4 and 7*b*). The side chains of three basic residues (Arg15, Lys22 and Lys204) form hydrogen bonds/salt links to the triphosphate, while the main-chain amide N atoms of residues 13–15 are hydrogen-bond donors. The CTP–protein interactions are essentially identical in each subunit of the dimer. In the CMP complex, however, there is a conformational change in the 12–25 loop (equivalent to the P-loop in nucleotide triphosphate hydrolases) of the *B* chain. In the *A* chain, therefore, the CMP is positioned and shares the same protein interactions as the cytosine and α-phosphate of the CTP complex (Fig. 7*b*). In the *B* chain, the CMP is shifted, almost as a rigid body, by more than 2 Å (Fig. 7*b*). Despite this, local changes in the protein structure allow all hydrogen-bonding interactions to the base to be maintained. Hydrogen-bond interactions linking the protein with the ribose and α-phosphate, however, are altered. The carbonyl O atom of residue Pro8 accepts a hydrogen bond from the O2′ hydroxyl instead of that of O3′ and while the interaction between the α-phosphate and the Lys22 side chain is maintained, new hydrogen-bonding interactions between this phosphate and the amide N atom of Gly13 and the hydroxyl of Ser12 are made. The magnesium and erythritol that had been included in the crystallization of the CMP complex could not be identified in our electron-density maps.

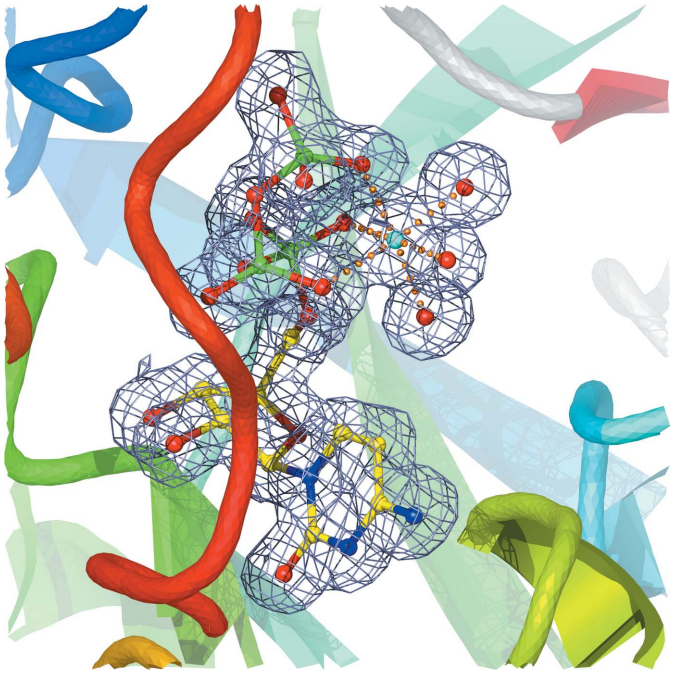

Figure 4
Representative electron density in the *Ms*IspD–CTP structure. The σ_A-weighted ($2m|F_o| - D|F_c|$) electron-density map (Read, 1986) is drawn at a level of 0.3 e Å^{-3} around the CTP- and Mg^{2+}-binding site of *Ms*IspD. The background cartoon has been created with the same red–blue rainbow colouring for one chain and a secondary-structure theme for the other (strands red, helices yellow and loops silver). The secondary-structure elements have been defined as described in §2.

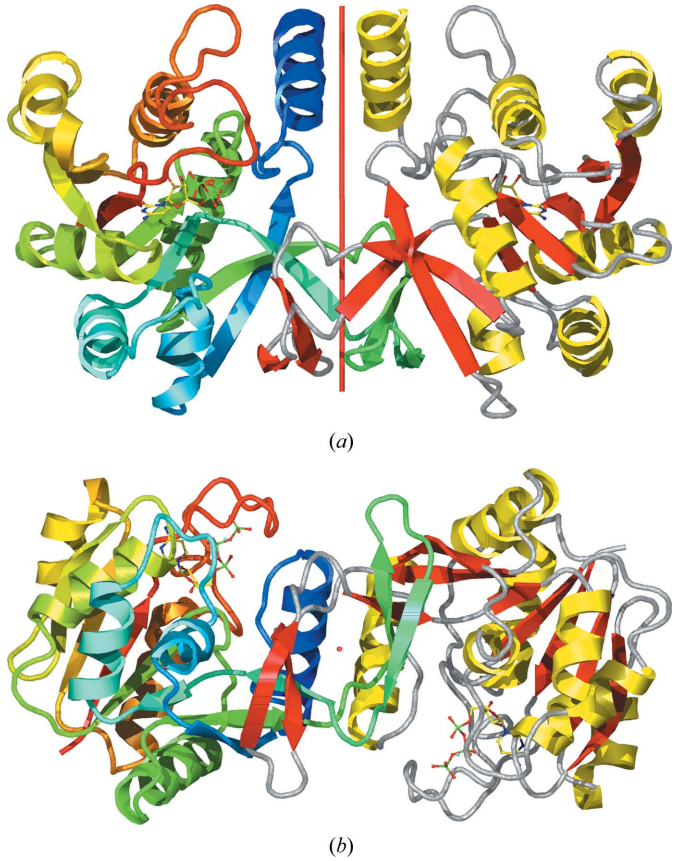

Figure 5
Cartoons of the *Ms*IspD–CTP structure. The views have been chosen (*a*) to illustrate the position of the active sites in the homodimer and (*b*) to show an overview of dimer formation. The twofold axis is drawn in each figure as a thin red cylinder. One chain has been drawn with rainbow colouring and the other with a secondary-structure theme.

Table 3
Structural comparisons with the *Ms*IspD–CTP complex *A* chain using *DALI*.

In enzymes with more than one chain, the chain with the highest *Z* has been included. The *Z* score is the degree of structural similarity in standard deviations above that expected as reported by the *DALI* server (Holm & Park, 2000). R.m.s.d. is the positional root-mean-square deviation of superimposed C^{α} atoms. lali is the total number of equivalent residues. nres is the length of the entire chain of the equivalent structure. %id is the percentage sequence identity over equivalent positions.

Z	R.m.s.d. (Å)	lali	nres	%id	PDB entry	Enzyme	Ligands	Resolution (Å)	Reference
38.7	0.5	220	220	100	2xwm	*M. smegmatis* IspD	CMP	1.8	This work
33.9	1.2	216	221	64	2wxn	*M. tuberculosis* IspD	CTP, Mg^{2+}	2.9	This work
29.7	1.4	202	203	64	3okr	*M. tuberculosis* IspD	Apo	2.4	Sacchettini *et al.* (unpublished work)
29.1	2.1	216	221	32	1vpa	*Thermatoga maritima* IspD	CTP, Mg^{2+}	2.7	Joint Center for Structural Genomics (unpublished work)
29.0	1.8	218	225	36	1i52	*E. coli* IspD	CTP, Mg^{2+}	1.5	Richard *et al.* (2001)
28.8	1.8	217	225	36	1ini	*E. coli* IspD	CDP-ME, Mg^{2+}	1.8	Richard *et al.* (2001)
26.3	2.3	210	217	36	1vgt	*E. coli* IspD	Apo	1.8	Badger *et al.* (2005)
26.3	1.9	206	213	36	1inj	*E. coli* IspD	Apo	1.6	Richard *et al.* (2001)
26.1	2.1	209	215	36	1h3m	*E. coli* IspD	Apo	2.4	Kemp *et al.* (2003)
25.0	2.2	208	214	36	1vgu	*E. coli* IspD	Apo	2.8	Badger *et al.* (2005)
27.2	2.3	200	203	44	2px7	*Thermus thermophilus* IspD	Apo	2.2	Chen *et al.* (unpublished work)
25.7	1.8	206	225	28	2vsi	*Streptococcus pneumoniae* TarI	CDP	2.8	Baur *et al.* (2009)
25.3	1.8	204	223	27	2vsh	*S. pneumoniae* TarI	Apo	2.0	Baur *et al.* (2009)
25.3	2.1	203	369	29	1w55	*Campylobacter jejuni* IspDF	CMP	2.3	Gabrielsen *et al.* (2004)
25.1	2.2	203	368	29	1w57	*C. jejuni* IspDF	CMP	3.1	Gabrielsen *et al.* (2004)
24.5	1.9	196	212	28	1w77	*Arabidopsis thaliana* IspD	CMP	2.0	Gabrielsen *et al.* (2006)
24.5	2.3	200	212	33	1vgz	*Neisseria gonorrhoeae* IspD	Apo	3.0	Badger *et al.* (2005)
23.8	2.4	201	211	34	1vgw	*N. gonorrhoeae* IspD	Apo	2.4	Badger *et al.* (2005)
24.5	2.0	210	229	25	3f1c	*Listeria monocytogenes* IspD	Apo	2.3	Patskovsky *et al.* (unpublished work)
22.9	2.5	205	445	15	1g95	*S. pneumoniae* GlmU	Apo	2.3	Kostrewa *et al.* (2001)

3.4. *Mt*IspD structure

Because of the relatively low resolution of the *Mt*IspD–CTP complex, we used NCS restraints in the refinement of this crystal structure. It became apparent, however, that there was a breakdown in the NCS that affected the tip of the dimer-forming β-meander. With a close-pair alignment in the least-squares superposition of the *A* and *B* chains, we can align 197 C^{α} pairs with an r.m.s.d. of 0.11 Å. 17 pairs in the region 140–156 are then separated by more than 1 Å and the five C^{α} atoms at the tip (residues 146–150) have differences of ∼3 Å. Despite the breakdown in NCS in *Mt*IspD, the set of interactions made by the meanders around the dimer axis are very similar to those seen in *Ms*IspD.

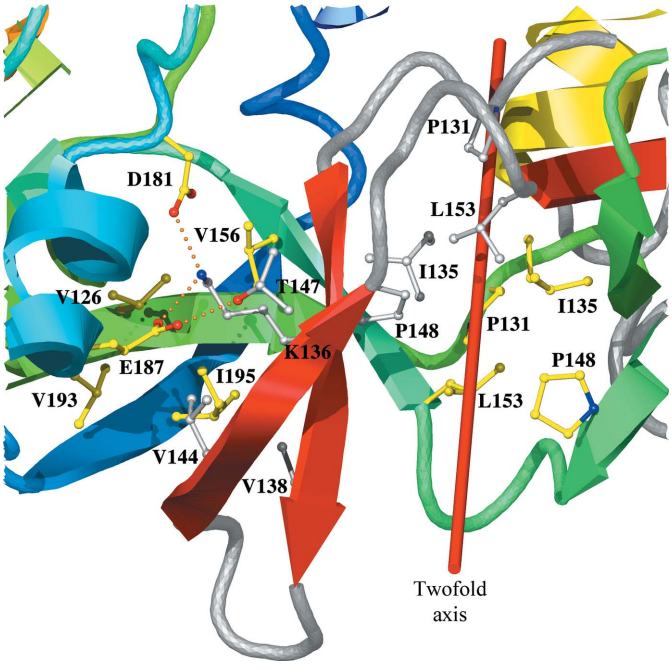

Figure 6
*Ms*IspD–CTP dimer interactions. The interacting side chains around the twofold axis (to the right of the figure) come from the β-meanders, while the interactions to the left illustrate how the β-meander stem from one chain interacts with the second chain in a small enclosed volume containing side chains of highly conserved residues. Because of symmetry, the latter set of interactions are duplicated within the dimer as a whole (not shown).

The CTP- and magnesium ion-binding sites are essentially identical in both chains, with interactions very similar to those that we observed in the *Ms*IspD structure (Fig. 7*b*). A minor difference concerns the substitution of Ser82 in *Ms*IspD for Thr86 in *Mt*IspD, where the conserved hydroxyl group makes a hydrogen bond to N2 of the base. Of the five insertions/deletions in the aligned *Ms*IspD/*Mt*IspD sequences (Fig. 3), two are at the termini, two more involve changes in links between α- and β-units and one is in a loop close to the active site (residues 176–178 and 183–190 in *Ms*IspD and *Mt*IspD, respectively). This active-site loop is longer in the *Mtb* structure, but does not extend into the CTP-binding site. Indeed, the electron density is rather weak in the *B* chain and is not possible to interpret in the *A* chain.

During the final preparation of this manuscript an as yet unpublished *Mt*IspD apoenzyme structure was deposited (PDB entry 3okr; J. C. Sacchettini, M. C. M. Reddy, J. B. Bruning & C. Thurman, unpublished work). This structure contains two dimers in the asymmetric unit in a *P*1 unit cell. The most striking differences compared with our *Mt*IspD–CTP complex involve the loops making up the cytosine base-binding site. In all four chains much of the 'P-loop' is dis-

ordered (corresponding to residues 16–25), while the loop connecting the edge strand β3 to helix α3 is shifted by up to 1.5–3 Å. The separate chains of this apoenzyme structure can be superimposed on our CTP complex with r.m.s.d.s of ∼1.0 Å for ∼198 C^{α} pairs.

3.5. Relationship to earlier structures

The most significant breakthrough in the study of IspD came in 2001 with the publication of the *E. coli* structure in three states of complexation: the apoenzyme, the CTP (*i.e.* one substrate plus Mg^{2+}) complex and the CDP-ME (one product plus Mg^{2+}) complex (Richard *et al.*, 2001). As well as providing insights into the reaction mechanism, this work demonstrated that IspD shares structural similarity to a family of cytosine transferase enzymes. IspD is now classified as a member of glycosyl transferase clan GT-A and has been assigned to Cluster of Orthologous Groups (COG) 1211 (Tatusov *et al.*, 1997) and Pfam family 01128 (Bateman *et al.*, 2004). Although *Ec*IspD remains the best-studied of the IspD enzymes, structures are also available for other IspDs, as detailed in Table 3 for the results of a structural similarity search using *DALI* (Holm & Park, 2000). One, the *Campylobacter jejuni* enzyme, is an IspD/IspF fusion protein with both enzymatic activities. Another, TarI from *Streptococcus pneumoniae*, is an IspD-like protein that uses ribitol 5-phosphate as a substrate but also has activity towards MEP. As expected, mycobacterial IspDs show most similarity to the existing IspD structures and TarI. Single subunits from these enzymes can be superimposed on *Ms*IspD with r.m.s.d.s in the range 1.8–2.3 Å for 195–218 pairs of C^{α} atoms. The variation in the number of matched pairs is a consequence in part of disordered regions in some of the crystal structures. The alignment of the *E. coli* enzyme (PDB entry 1i52 chain *A*; Richard *et al.*, 2001), for example, gives an r.m.s.d. of 1.80 Å for 218 of 222 pairs of C^{α} atoms and 36% structure-based sequence identity. The highest structure-based sequence identity is with the *Thermus thermophilus* enzyme (44%) and the lowest is with the *Listeria monocytogenes* enzyme (25%). Although all of the enzymes in this group form an active dimer as described above, there is some plasticity in the dimer interface (Gabrielsen *et al.*, 2006). If we compare *Ms*IspD with TarI, for example, an extra twist of ∼13° is required to align the second chains of the dimers around a rotation axis passing through the dimer-forming β-meander domains close to their twofold axis.

A close-pair structural alignment (described in the §2) of 11 representative IspD structures (Fig. 8*a*) clearly indicates that the β-sheet is highly conserved, as well as the root and stem sections of the β-meander. The loop following β4, the central strand of the β-sheet, is mostly similar until the middle of the following helix α4. Likewise, β7 (at one of the edges of the sheet) is similar, as is the connecting loop to the C-terminal helix α7. In the pair of loops making up the binding site for the cytosine base the main-chain conformations are less tightly clustered. In one of these loops the section following β1 shows a great deal of variation in structure (*e.g.* in the *A* and *B* chains of the *Ms*IspD–CMP complex described above), in length (*e.g.* by one residue when comparing *Ms*IspD and the *L. monocytogenes* enzyme) and crystallographic order (apoenzymes often have a disordered local structure, but the structure of *A. thaliana* CMP also has a disordered section). The second loop in this site, connecting the β3 edge strand to the next helix α3, is well conserved in length and order but has local rigid-body shifts of up to 3 Å. In the MEP portion of the substrate-binding site there is an even more striking asymmetry in the structural alignments.

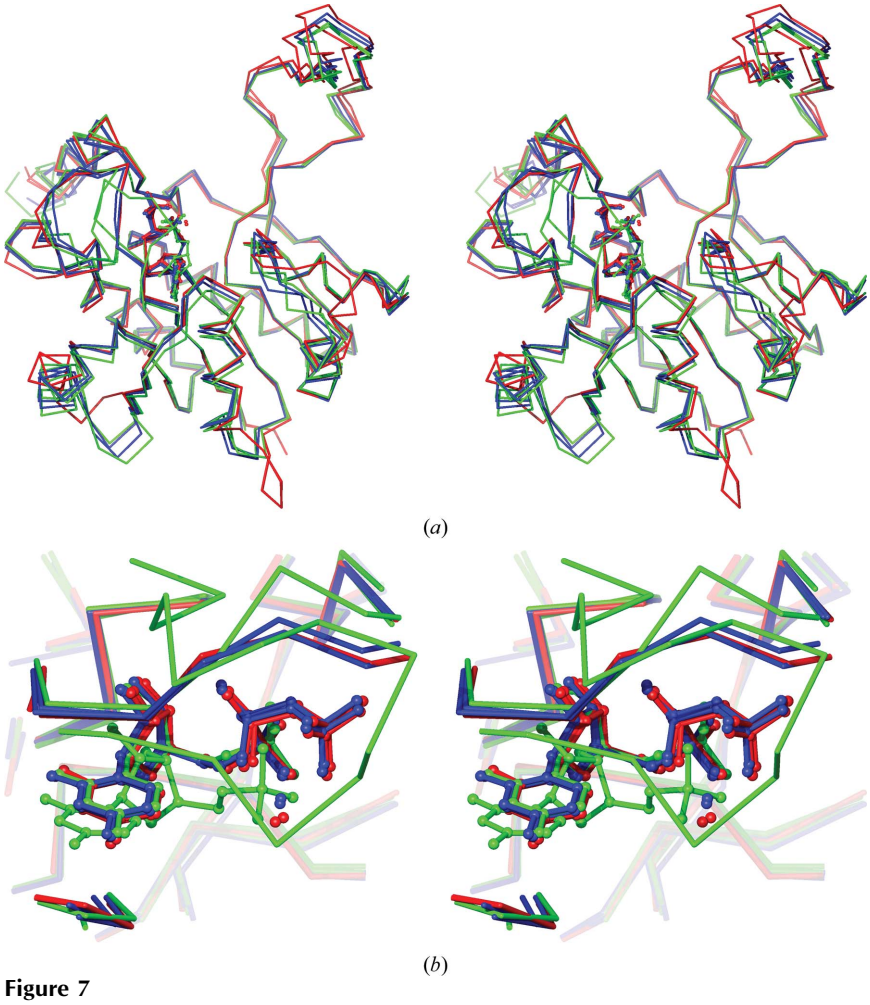

Figure 7
Superposition of the mycobacterial IspD structures. The pairs of *Ms*IspD–CTP chains and ligands are coloured blue, while *Ms*IspD–CMP structures are coloured green and *Mt*IspD–CTP structures are coloured red. (*a*) shows a stereo C^{α} trace that includes ligands and (*b*) shows a close-up of the ligands in the active site. The superpositions were made with close-pair cutoffs as described in §2.

While the three strands at the edge of the β-sheet (ordered 7-5-6) are very tightly clustered, the helices-containing loop connecting β6 and β7 shows a great deal of variation in structure and length, as do the tips of the β-meander. Despite the variation in their length, the stem sections of the twofold-related β-meanders are moderately well clustered where they help to form the MEP-binding site. However, the variation in positioning of the twofold axis in each dimer with respect to the structurally conserved regions means that there is no tight clustering of the dimer as a whole.

After the IspD group of enzymes, we find the closest member of the broader family of transferases: *N*-acetylglucosamine 1-phosphate uridyltransferase GlmU (Kostrewa *et al.*, 2001). Members of this family are larger, contain an extra β-helix domain and form trimers (Brown *et al.*, 1999). Here, portions of the transferase domain of ~205 amino acids can be superimposed with r.m.s.d.s of 2.4 Å and higher, while there is a drop in the structure-based sequence identity to 15% or less. As well as the overall fold, the location of the substrate-binding site is conserved. Although the products overlap remarkably well, the detailed interactions are different; in particular, the dimer-forming β-meander is missing. After the GlmU structures, there is a long series of more distantly related nucleotide transferases (data not shown).

4. Discussion

Our measured kinetic parameters for *Mt*IspD (Table 2) give calculated k_{cat}/K_m values of 125 mM^{-1} min^{-1} for MEP and 138 mM^{-1} min^{-1} for CTP. Our K_m values are similar to those published previously for the same enzyme (Shi *et al.*, 2007; Eoh *et al.*, 2007; see Table 2). However, a larger variation is seen among the k_{cat} values. A sixfold lower k_{cat} value for both substrates was reported by Eoh *et al.* (2007). A possible explanation may be that the earlier experiment was conducted under nonsaturating conditions, since the concentration of the fixed substrate was held at less than twice its own K_m while varying the other substrate. A smaller difference is seen when comparing with the values reported by Shi *et al.* (2007). The variation here could arise from the use of a lower concentration of magnesium chloride (2 mM instead of 10 mM), which may lower the activity (Eoh *et al.*, 2007). A comparison with the kinetic parameters of *Ec*IspD is even less straightforward because there is a large variation in the published results, as shown in Table 2. However, the k_{cat} for the *Ec*IspD enzyme is consistently found to be higher than that obtained for *Mt*IspD. In the *Ec*IspD–CTP complex, Richard *et al.* (2001) showed the γ-phosphate to be directly coordinated by the guanido group of Arg20 (Arg15 and Arg20 in *Ms*IspD and *Mt*IspD, respectively). The main-chain N atom of the preceding residue, Arg19, also formed a hydrogen bond, an interaction that is conserved in the *Ms* and *Mtb* enzymes. While not making a direct salt link, the guanido group in the side chain of Arg19 had a closest approach of 4.2 Å to the γ-phosphate and was therefore suggested to be involved in stabilizing this group in the transition state. In both *Ms*IspD and *Mt*IspD the equivalent residue is a glutamic acid whose side chain points away from the γ-phosphate of CTP in our complexes. We suggest that the loss of transition-state stabilization caused by this sequence change may account, at least in part, for the smaller k_{cat} value observed for *Mt*IspD.

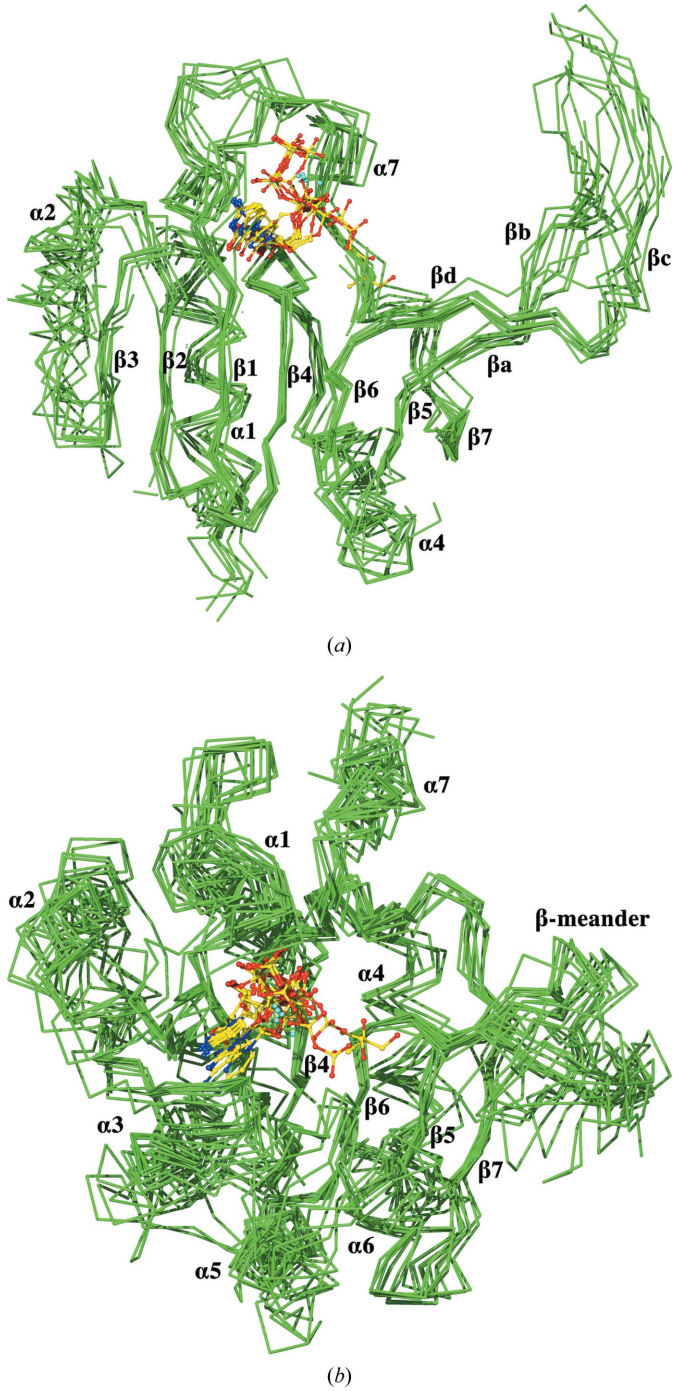

Figure 8
Superposition of representatives of 11 IspD structures. The view has been chosen in (*a*) to illustrate the well conserved core sheet and to highlight the variation in the β-meanders and some of the loops and helical segments discussed in the text. The high variability in helices α3, α5 and α6 is not apparent because we have applied a front-plane clipping action to remove them. In (*b*), the view has been chosen to highlight the structural variability. The superpositions were made with close-pair cutoffs as described in §2.

Alignment of the *Ec*IspD sequence with those from eight bacterial pathogen species gave pairwise identities in the range 30–90%; 22 residues were strictly conserved, of which 21 are in the active site (Kemp *et al.*, 2003). All of these residues are conserved in *Mt*IspD and *Ms*IspD, as are the subset of these that are thought to take part in catalysis (Richard *et al.*, 2001). Therefore, it is not surprising that *Ms*IspD and *Mt*IspD have substrate-product binding sites that are very similar to those of the *E. coli* enzyme. In Fig. 9, we have combined the structural results of the *E. coli* CDP-ME product complex with our *Ms*IspD–CTP complex structure to show the disposition of these highly conserved residues and how they are likely to interact with substrates in both *Ms*IspD and *Mt*IspD. This superposition shows the following:

(i) A highly conserved set of interactions between protein and product. Interestingly, in our *Ms*IspD–CTP complex, water molecules overlap the positions of the MEP hydroxyl groups, indicating that this part of the active site is preformed to interact with the substrate.

(ii) That four of the conserved residues are from the dimer-related meander, of which two (Thr134 and Arg150) interact directly with the product (Richard *et al.*, 2001) and presumably the MEP substrate, while the others (Lys146 and Thr147) help to create part of the dimer interface that we described earlier (Fig. 6).

(iii) That there may be space for the extension of MEP by 2–3 non-H atoms.

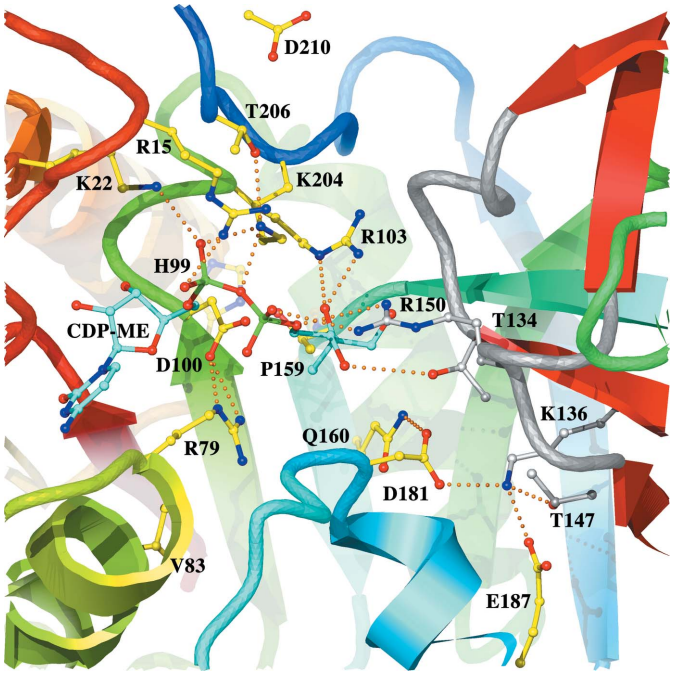

Figure 9
Highly conserved residues in the active site of a mycobacterial IspD. The *Ms*IspD–CTP active site is shown with the CDP-ME product docked from the superimposed *Ec*IspD complex of Richard *et al.* (2001). Four residues come from the β-meander of the twofold-related subunit. Dotted lines show expected hydrogen bonds. One conserved residue, Asp210, is neither in the active site nor involved directly in dimer formation; it appears at the top of the figure.

(iv) A structurally well conserved cytosine-binding site that extends away from the magnesium-binding site. However, the variation that we observe within the IspD family as a whole and in the mycobacterial IspD structures in particular indicates some degree of flexibility in two loops in the apoenzyme. These loops must be correctly formed to produce the cytosine ring-binding site and this may be relevant for inhibitor design. It suggests, for example, that a fragment-based approach may not be successful.

The *Plasmodium falciparum* enzyme (*Pf*IspD) is conspicuous by its absence from the list discussed earlier (Kemp *et al.*, 2003) despite the requirement for IspD activity in this class of parasite. The likely protein (UniProt accession No. Q8I273), which is conserved in all *Plasmodium* species that have been sequenced, is much larger than the classical IspDs described above. The putative enzyme is 734 amino acids in length, with a region of ~220 amino acids near the C-terminus that shows 19 and 16% sequence identity to the *E. coli* and *Ms*/*Mtb* IspD sequences, respectively (see Fig. 3). In some organisms, IspD and IspF exist as a covalently linked pair; in *Mtb* they are adjacent in the genome but are separated by a stop codon and a single-base frameshift. This *Plasmodium* gene, however, has no IspF-like sequence that we can identify. The conservation in the IspD-like sequence indicates that the residues making up the active site are part of a dimer-forming enzyme. This includes residues from the second subunit of the dimer that interact with the MEP substrate and the β-meander itself (*e.g.* the Lys136–Asp181 salt link in the meander–dimer interface; *Ms*IspD numbering; see Fig. 6), which appears to be extended by nine amino acids. In the β-hairpin barrel, the key branched-chain hydrophobic residues Ile135, Leu153, Pro131 and Pro148 of *Ms*IspD are matched to Ile, Ile, Ala and Ile residues in *Pf*IspD. The *cis*-proline signature of IspDs where the meander 'returns' to the sixth strand of the sheet is clearly conserved, as are the adjacent residues. Of the 22 residues highlighted by Kemp *et al.* (2003), 17 are conserved in this putative *Pf*IspD, including those from the stem of the β-meander making up the major portion of the dimer interface (see Figs. 3, 6 and 9). The largest insert (of ~25 residues) occurs in the long loop between the sixth and last β-strands (probably between α5 and α6), which is the region with largest structural variability in Fig. 8. With our alignment, this *Pf*IspD domain would start at Tyr417 and end at Tyr682 (using equivalent *Ms*IspD N- and C-terminal residues). Interestingly, this protein contains a highly polar stretch of ~60 residues close to its mid-point, which prompted us to investigate the N-terminal region in more detail. Indeed, the *Phyre* server (Kelley & Sternberg, 2009) identified another region with low sequence identity to IspD structures. Since the alignment does not extend to the critical highly conserved proline signature, we have omitted the alignment from Fig. 3. However, the possibility exists that *Pf*IspD represents fused IspD domains connected by a polar linker.

Our primary interest in the IspD enzyme and MEP pathway is in fact to develop new inhibitors for lead development. The best IspD inhibitor that has been described to date in the literature, erythritol-4-phosphate, has a rather poor IC_{50} of

1.4 m*M* (Lillo *et al.*, 2003; Wungsintaweekul, 2001). Although the activity assay is suitable for high-throughput screening, such results have not yet been published. We plan to use virtual screening to investigate binding in both the MEP- and cytosine-binding sites, as well as a fragment-based approach (although some complications arise from potential disorder in the cytosine-binding site). The high concentration of basic side chains, an essential requirement for absorbing the build-up of negatively charged phosphate groups from the CTP and MEP substrates, must be addressed as part of a program to make this target druggable.

This work was supported by funding from the Foundation for Strategic Research (SSF), the European Union Sixth Framework Program NM4TB (CT:018923), the Swedish Research Council and Uppsala University.

References

Badger, J. *et al.* (2005). *Proteins*, **60**, 787–796.

Bateman, A., Coin, L., Durbin, R., Finn, R. D., Hollich, V., Griffiths-Jones, S., Khanna, A., Marshall, M., Moxon, S., Sonnhammer, E. L., Studholme, D. J., Yeats, C. & Eddy, S. R. (2004). *Nucleic Acids Res.* **32**, D138–D141.

Baur, S., Marles-Wright, J., Buckenmaier, S., Lewis, R. J. & Vollmer, W. (2009). *J. Bacteriol.* **191**, 1200–1210.

Bernal, C., Palacin, C., Boronat, A. & Imperial, S. (2005). *Anal. Biochem.* **337**, 55–61.

Beytía, E. D. & Porter, J. W. (1976). *Annu. Rev. Biochem.* **45**, 113–142.

Bond, C. S. & Schüttelkopf, A. W. (2009). *Acta Cryst.* D**65**, 510–512.

Brown, A. C. & Parish, T. (2008). *BMC Microbiol.* **8**, 78.

Brown, K., Pompeo, F., Dixon, S., Mengin-Lecreulx, D., Cambillau, C. & Bourne, Y. (1999). *EMBO J.* **18**, 4096–4107.

Cane, D. E., Chow, C., Lillo, A. & Kang, I. (2001). *Bioorg. Med. Chem.* **9**, 1467–1477.

Engh, R. A. & Huber, R. (1991). *Acta Cryst.* A**47**, 392–400.

Eoh, H., Brown, A. C., Buetow, L., Hunter, W. N., Parish, T., Kaur, D., Brennan, P. J. & Crick, D. C. (2007). *J. Bacteriol.* **189**, 8922–8927.

Evans, P. (2006). *Acta Cryst.* D**62**, 72–82.

Gabrielsen, M., Kaiser, J., Rohdich, F., Eisenreich, W., Laupitz, R., Bacher, A., Bond, C. S. & Hunter, W. N. (2006). *FEBS J.* **273**, 1065–1073.

Gabrielsen, M., Rohdich, F., Eisenreich, W., Gräwert, T., Hecht, S., Bacher, A. & Hunter, W. N. (2004). *Eur. J. Biochem.* **271**, 3028–3035.

Gorrec, F. (2009). *J. Appl. Cryst.* **42**, 1035–1042.

Harris, M. & Jones, T. A. (2001). *Acta Cryst.* D**57**, 1201–1203.

Henriksson, L. M., Unge, T., Carlsson, J., Aqvist, J., Mowbray, S. L. & Jones, T. A. (2007). *J. Biol. Chem.* **282**, 19905–19916.

Holm, L. & Park, J. (2000). *Bioinformatics*, **16**, 566–567.

Jancarik, J., Pufan, R., Hong, C., Kim, S.-H. & Kim, R. (2004). *Acta Cryst.* D**60**, 1670–1673.

Jones, T. A., Zou, J.-Y., Cowan, S. W. & Kjeldgaard, M. (1991). *Acta Cryst.* A**47**, 110–119.

Kabsch, W. (2010). *Acta Cryst.* D**66**, 125–132.

Kelley, L. A. & Sternberg, M. J. (2009). *Nature Protoc.* **4**, 363–371.

Kemp, L. E., Bond, C. S. & Hunter, W. N. (2003). *Acta Cryst.* D**59**, 607–610.

Kleywegt, G. J. & Jones, T. A. (1996). *Structure*, **4**, 1395–1400.

Kleywegt, G. J. & Jones, T. A. (1997). *Methods Enzymol.* **277**, 525–545.

Kostrewa, D., D'Arcy, A., Takacs, B. & Kamber, M. (2001). *J. Mol. Biol.* **305**, 279–289.

Kötting, O., Santelia, D., Edner, C., Eicke, S., Marthaler, T., Gentry, M. S., Comparot-Moss, S., Chen, J., Smith, A. M., Steup, M., Ritte, G. & Zeeman, S. C. (2009). *Plant Cell*, **21**, 334–346.

Kuzuyama, T., Shimizu, T., Takahashi, S. & Seto, H. (1998). *Tetrahedron Lett.* **39**, 7913–7916.

Leslie, A. G. W. (2006). *Acta Cryst.* D**62**, 48–57.

Lillo, A. M., Tetzlaff, C. N., Sangari, F. J. & Cane, D. E. (2003). *Bioorg. Med. Chem. Lett.* **13**, 737–739.

Ma, Z., Lienhardt, C., McIlleron, H., Nunn, A. J. & Wang, X. (2010). *Lancet*, **375**, 2100–2109.

Matthews, B. W. (1968). *J. Mol. Biol.* **33**, 491–497.

McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C. & Read, R. J. (2007). *J. Appl. Cryst.* **40**, 658–674.

Murshudov, G. N., Skubák, P., Lebedev, A. A., Pannu, N. S., Steiner, R. A., Nicholls, R. A., Winn, M. D., Long, F. & Vagin, A. A. (2011). *Acta Cryst.* D**67**, 355–367.

Read, R. J. (1986). *Acta Cryst.* A**42**, 140–149.

Richard, S. B., Bowman, M. E., Kwiatkowski, W., Kang, I., Chow, C., Lillo, A. M., Cane, D. E. & Noel, J. P. (2001). *Nature Struct. Biol.* **8**, 641–648.

Richard, S. B., Lillo, A. M., Tetzlaff, C. N., Bowman, M. E., Noel, J. P. & Cane, D. E. (2004). *Biochemistry*, **43**, 12189–12197.

Rohdich, F., Wungsintaweekul, J., Fellermeier, M., Sagner, S., Herz, S., Kis, K., Eisenreich, W., Bacher, A. & Zenk, M. H. (1999). *Proc. Natl Acad. Sci. USA*, **96**, 11758–11763.

Rohmer, M. (1999). *Nat. Prod. Rep.* **16**, 565–574.

Rossmann, M. G., Liljas, A., Brandén, C. I. & Banaszak, L. J. (1975). *The Enzymes*, 3rd ed., edited by P. D. Boyer, Vol. XI, pp. 61–102. New York: Academic Press.

Sassetti, C. M., Boyd, D. H. & Rubin, E. J. (2003). *Mol. Microbiol.* **48**, 77–84.

Sensi, P., Greco, A. M. & Ballotta, R. (1959). *Antibiot. Annu.* **7**, 262–270.

Shi, W., Feng, J., Zhang, M., Lai, X., Xu, S., Zhang, X. & Wang, H. (2007). *J. Biochem. Mol. Biol.* **40**, 911–920.

Shigi, Y. (1989). *J. Antimicrob. Chemother.* **24**, 131–145.

Simossis, V. A. & Heringa, J. (2005). *Nucleic Acids Res.* **33**, W289–W294.

Tatusov, R. L., Koonin, E. V. & Lipman, D. J. (1997). *Science*, **278**, 631–637.

Winn, M. D. *et al.* (2011). *Acta Cryst.* D**67**, 235–242.

Wungsintaweekul, J. (2001). Thesis, Technical University of Munich.

Zeidler, J., Schwender, J., Muller, C., Wiesner, J., Weidemeyer, C., Beck, E., Jomaa, H. & Lichtenthaler, H. K. (1998). *Z. Naturforsch. C*, **53**, 980–986.