Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

Structure and catalytic mechanism of 2-*C*-methyl-D-erythritol 2,4-cyclodiphosphate (MECDP) synthase, an enzyme in the non-mevalonate pathway of isoprenoid synthesis

Hiroyuki Kishida,[a] Takashi Wada,[a,b] Satoru Unzai,[a] Tomohisa Kuzuyama,[c] Motoki Takagi,[c] Takaho Terada,[b] Mikako Shirouzu,[b] Shigeyuki Yokoyama,[b,d,e] Jeremy R. H. Tame[a] and Sam-Yong Park[a]*

[a]Protein Design Laboratory, Graduate School of Integrated Science, Yokohama City University, 1-7-29 Suehiro-cho, Tsurumi, Yokohama 230-0045, Japan, [b]Genomic Sciences Center, RIKEN Yokohama Institute, 1-7-29 Suehiro-cho, Tsurumi, Yokohama 230-0045, Japan, [c]Institute of Molecular and Cellular Biosciences, The University of Tokyo, 1-1-1 Yayoi, Bunkyo-ku, Tokyo 113-0032, Japan, [d]RIKEN Harima Institute at SPring-8, 1-1-1 Kouto, Mikazuki-cho, Sayo, Hyogo 679- 5148, Japan, and [e]Department of Biophysics and Biochemistry, Graduate School of Science, The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-0033, Japan

Correspondence e-mail: park@tsurumi.yokohama-cu.ac.jp

Received 7 June 2002
Accepted 26 September 2002

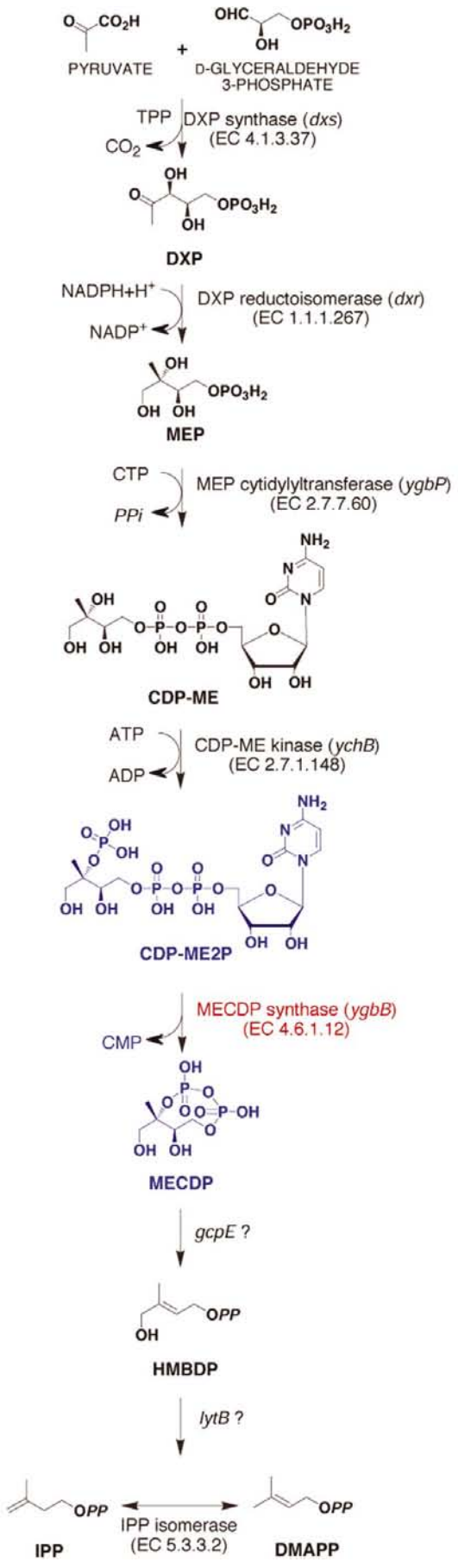
PDB References: native MECDP, 1iv1; SeMet MECDP with CDP-ME, 1iv2; SeMet MECDP with Mg, 1iv3; SeMet MECDP with CDP-ME2P, 1iv4.

Precursors for isoprenoid synthesis are essential in all organisms. These compounds are synthesized by one of two known routes: the well characterized mevalonate pathway or a recently discovered non-mevalonate route which is used in many bacteria and human pathogens. Since the second pathway is both vital and unlike any found in humans, enzymes catalysing reactions along this synthetic route are possible drug targets. The structure of one such enzyme from the thermophilic bacterium *Thermus thermophilus* has been solved to high resolution in the presence of substrate and with a substrate analogue. Enzyme co-crystallized with substrate shows only one product, cytosine monophosphate (CMP), in the active site. At the high resolution of the refinement (1.6 Å) the positions and coordination of the magnesium ions in the active site are clearly seen.

1. Introduction

Isoprenoids are a hugely diverse group of molecules found in all organisms and are constructed from just two building blocks: isopentenyl diphosphate (IPP) and its isomer dimethylallyl diphosphate (DMAPP). Yeast and higher eukaryotes synthesize IPP from acetyl CoA *via* mevalonate, using a well characterized pathway (Qureshi & Porter, 1981). Only in the last 10 y has it become apparent that many human pathogens (and plant plastids) use an altogether different synthetic route which starts with pyruvate and glyceraldehyde-3-phosphate (Rohmer *et al.*, 1993). Among the pathogenic species which use this 'non-mevalonate' pathway are *Mycobacterium tuberculosis* and *Plasmodium falciparum*. There is therefore enormous interest in uncovering the structures of the enzymes along the route for the purpose of drug design. The antibiotic fosmidomycin was discovered long before it was realised that it inhibits an early step in the non-mevalonate pathway (Kuzuyama *et al.*, 1998); this compound is able to cure mice of the rodent malaria parasite *P. vinckei* (Jomaa *et al.*, 1999).

The non-mevalonate pathway is still not completely characterized, but appears to involve seven separate enzymes (Rohdich *et al.*, 2002; Kuzuyama, 2002), several of which have already been crystallized (Richard *et al.*, 2001; Reuter *et al.*, 2002; Yajima *et al.*, 2002). Three independent groups have recently solved the structure of 4-(cytidine-5′-diphospho)-2-*C*-methyl-D-erythritol (MECDP) synthase from *Escherichia coli* (*ygbB*; Richard *et al.*, 2002; Steinbacher *et al.*, 2002; Kemp *et al.*, 2002), which catalyses the fifth step in the pathway (Herz *et al.*, 2000; Takagi *et al.*, 2000). These structures show the protein to be a trimer with three active sites shared between adjacent copies of the protein. The enzyme catalyses an intramolecular



attack by a phosphate group on a diphosphate, cytidine monophosphate (CMP) leaving the substrate to yield a cyclic diphosphate (Fig. 1). Substrate soaking into the crystals allowed the substrate and products to be positioned in the active site and some conclusions to be drawn. We have solved the structure of the equivalent protein from *Thermus thermophilus* HB8, co-crystallized with substrate and with a substrate analogue missing the attacking phosphate group. These structures, refined to 1.6 Å, show the protein in more detail and give greater insight into the catalytic mechanism, particularly with regard to the metal centres in the protein.

2. Materials and methods

2.1. Expression and purification of MECDP synthase

E. coli BL21 (DE3) and B834 (DE3) cells were transformed with the vector pET11b carrying the *ygbB* gene from *T. themophilus* HB8 and then grown in LB medium and LeMaster medium with 25 mg ml^{-1} seleno-L-methionine (SeMet), respectively (LeMaster & Richards, 1985; Hendrickson *et al.*, 1990). Cells were grown with shaking at 310 K. IPTG was added to 1 m*M* final concentration once the culture reached an optical density (at 600 nm) of 0.8. After 3 h of further growth the cells were harvested and stored at 193 K until use. The following procedure, except for the heat treatment, was carried out at 277 K. 14 g of wet cells were suspended in 160 ml of lysis buffer (20 m*M* Tris–HCl pH 8.0, 300 m*M* NaCl and 1 m*M* DTT) and then sonicated on ice. The crude extract was then incubated at 343 K for 30 min and centrifuged. Saturated ammonium sulfate solution was added to the recovered supernatant to a final concentration of 1.4 *M*. The protein was then applied to a phenyl-Toyopearl (TOSOH) column (5 ml bed volume) and eluted with a linear gradient of elution buffer (20 m*M* Tris pH 8.0, 1 m*M* DTT). The fractions containing MECDP synthase were collected and dialyzed against 20 m*M* MES pH 5.5, 50 m*M* NaCl, 1 m*M* DTT. The protein was applied to a Resource-S (Pharmacia) column (6 ml bed volume) and eluted with a linear gradient of sodium chloride (50–300 m*M* NaCl, 20 m*M* MES pH 5.5, 1 m*M* DTT). The pooled MECDP synthase was diluted with 20 m*M* MES pH 5.5, 50 m*M* NaCl, 1 m*M* DTT buffer. It was then loaded onto a Synchropak CM300 column (4 ml bed volume) and eluted with a linear gradient of sodium chloride (50–300 m*M* NaCl with 20 m*M* MES pH 5.5, 1 m*M* DTT). The fractions containing MECDP synthase were collected. 13 mg of native MECDP synthase were obtained from 6.0 g of wet cells and 53 mg of MECDP synthase containing SeMet were obtained from 14 g of wet cells.

Figure 1
The non-mevalonate synthetic pathway. MECDP synthase catalyses the cyclization of its substrate with the release of CMP. IPP, isopentenyl diphosphate; DMAPP, dimethylallyl diphosphate; DXP, 1-deoxy-D-xylulose 5-phosphate; MEP, 2-*C*-methyl-D-erythritol 4-phosphate; FMM, fosmidomycin; CDP-ME, 4-(cytidine-5′-diphospho)-2-*C*-methyl-D-erythritol; CDP-ME2P, 2-phospho-4-(cytidine-5′-diphospho)-2-*C*-methyl-D-erythritol; MECDP, 2-*C*-methyl-D-erythritol 2,4-cyclodiphosphate; HMBDP, 1-hydroxy-2-methyl-2-butenyl diphosphate.

2.2. Crystallization and data collection

Crystals were grown at 293 K in hanging drops which consisted of 1 µl protein solution and 2 µl reservoir solution. For the native protein (PDB code 1iv1), 5.1 mg ml^{-1} protein in 20 m*M* Tris–HCl pH 8.0, 150 m*M* sodium chloride, 1 m*M* DTT was crystallized with 50 m*M* Tris–HCl pH 8.5, 1.0 *M* ammonium sulfate reservoir solution. Larger crystals were grown by means of the seeding method. For the SeMet derivative with magnesium (PDB code 1iv3), 5.2 mg ml^{-1} SeMet protein in 20 m*M* MES pH 5.5, 150 m*M* sodium chloride, 1 m*M* DTT was used with 50 m*M* MES pH 6.5, 100 m*M* magnesium acetate, 10% PEG 8000 reservoir solution. Seeding was not carried out in this case. Crystals of the SeMet derivative with CDP-ME2P substrate (PDB code 1iv4) were grown in the same way except that 30 m*M* CDP-ME2P was included in the protein solution. Crystals were grown by means of seeding. Exactly the same method was used to grow crystals of the SeMet derivative with CDP-ME (PDB code 1iv2), the only difference being that 30 m*M* CDP-ME replaced the CDP-ME2P. Although the crystallization conditions were not identical for the native and SeMet protein, the two crystals are isomorphous.

Table 1
Data-collection and refinement statistics.

Values in parentheses are for the outermost resolution shell: 2.07–2.00 Å for the MAD data, 1.71–1.65 Å for the native data, 1.61–1.55 Å for the CDP-ME-Mg^{II} and CDP-ME2P-Mg^{II} data and 1.57–1.52 Å for the Mg^{II} data.

(*a*) Se MAD data set, $P4_12_12$, $a = b = 106.507$, $c = 149.212$ Å.

Wavelength (Å)	0.9781 (high remote)	0.9824 (low remote)	0.9803 (peak)	0.9808 (edge)
Resolution range (Å)	50.0–2.0	50.0–2.0	50.0–2.0	50.0–2.0
Reflections (measured/unique)	565890/57413	562930/56845	562549/57195	566988/57093
Completeness (%)	97.9 (94.5)	96.7 (82.7)	97.7 (90.3)	97.4 (87.7)
R_{merge}† (%)	4.4 (10.0)	2.8 (8.3)	5.1 (10.0)	3.2 (8.4)
Overall redundancy	9.9	9.9	9.8	9.8
Overall $\langle I/\sigma(I)\rangle$	16.6	16.3	16.3	17.9
Phasing (50.0–2.0 Å)				
R_{iso} (%)		9.9	10.3	10.2
R_{cullis} (centric/acentric)	0.62/0.70	0.65/0.72	0.80/0.85	
Phasing power (centric/acentric)	1.41/1.59	1.32/1.47	0.93/0.97	
Mean FOM after *RESOLVE* phasing (centric/acentric)	0.64/0.68			

(*b*) Refinement statistics.

Refinement statistics	Native (apo)	Mg^{II}	CDP-ME-Mg^{II}	CDP-ME2P-Mg^{II}
Refinement resolution (Å)	15.0–1.65	20.0–1.52	20.0–1.55	20.0–1.55
Reflections (measured/unique)	935741/97488	892176/128529	836026/118031	940054/118106
Completeness (%)	96.3 (73.8)	98.5 (94.6)	95.8 (78.2)	95.2 (76.0)
R_{merge}† (%)	4.8 (26.0)	4.8 (25.2)	6.8 (24.8)	7.1 (17.5)
Overall redundancy	9.6	6.9	7.1	8.0
Overall $\langle I/\sigma(I)\rangle$	7.2	7.5	7.9	10.5
σ-cutoff/reflections used	0.0/97488	0.0/128529	0.0/118031	0.0/118106
R_{cryst}/R_{free} (%)	22.6/31.2	20.3/25.7	19.7/26.3	20.2/25.1
R.m.s.d. bond lengths/bond angles (Å)	0.006/0.020	0.007/0.024	0.007/0.026	0.008/0.024
Water atoms	352	557	509	516
Average *B* factor (metal/ligand, Å^2)	None	25/None	19/17	23/15
Ramachandran plot				
Residues in most favorable regions (%)	90.0	91.7	92.0	91.5
Residues in additional allowed regions (%)	10.0	8.3	8.0	8.5

† $R_{merge} = \sum\sum_i |I(h) - I(h)_i| / \sum\sum_i I(h)$, where $I(h)$ is the mean intensity after rejections.

High-resolution diffraction data were obtained using a synchrotron-radiation source at RIKEN beamline BL44B2 station (RIKEN Structural Biology Beamline II; Adachi *et al.*, 2001), SPring-8, Harima, Japan. Intensity data were collected with a MAR CCD detector. The structure of MECDP synthase was initially solved to 2.0 Å resolution by the multiple-wavelength anomalous dispersion (MAD) method. Diffraction data were integrated and scaled with *HKL*2000 and *SCALEPACK* (Otwinowski & Minor, 1997). General handling of the scaled data was carried out with programs from the *CCP*4 suite (Collaborative Computational Project, Number 4, 1994). The positions of Se atoms were determined using the program *SOLVE* (Terwilliger & Berendzen, 1999). The anomalous differences at the peak wavelength were normalized in magnitude and density modification with *RESOLVE* (Terwilliger & Berendzen, 1999). The resulting electron-density map was sufficiently clear to build an initial model of the structure. The model was built with the program *TURBO-FRODO* (Roussel & Cambillau, 1989). Structural refinement was performed using *ARP*/*wARP* (Perrakis *et al.*, 1999), *X-PLOR* v. 3.851 (Brünger, 1996) and *SHELX* (Sheldrick & Schneider, 1997). Solvent molecules were placed at positions where spherical electron-density peaks were found above 1.5σ in the $|2F_o - F_c|$ map and above 3.0σ in the $|F_o - F_c|$ map and where stereochemically reasonable hydrogen bonds were allowed. Structural evaluations of the final models of the WT and the complex enzymes using *PROCHECK* (Laskowski *et al.*, 1993) indicated that 90–92% of the residues are in the most favourable regions of the Ramachandran plot, with no residues in 'disallowed' regions. A summary of the data-collection and refinement statistics is given in Table 1.

2.3. Assay for MECDP synthase

A radiometric assay was used for detection of MECDP synthase activity. The substrate, CDP-ME2P, was synthesized from MEP using MEP cytidylyltransferase and CDP-ME kinase. MECDP synthase was then added and the synthesis of MECDP monitored. In the first step, 50 m*M* Tris–HCl pH 7.5, 2.5 m*M* $MgCl_2$, 2.5 m*M* dithiothreitol, 2.5 m*M* CTP, 2.5 m*M* ATP, 2.5 m*M* MEP, 0.33 m*M* (2-^{14}C)-MEP (851 MBq mmol^{-1}), 1 µg MEP cytidylyltransferase and 1 µg

CDP-ME kinase were mixed in a final volume of 20 µl. CDP-ME2P synthesis was then initiated by adding MEP cytidylyltransferase and CDP-ME kinase, followed by incubation at 310 K for 60 min. The reaction was terminated by incubation at 333 K for 5 min. 1 µg of the MECDP synthase being tested was then added, followed by incubation at 310 K. After 60 min incubation, an aliquot of the reaction mixture was subjected to thin-layer chromatography on a cellulose gel (Merck). The cellulose plate, developed with a mixture of 2-butanol, acetic acid and water (3:2:2), was exposed to an imaging plate (Fujifilm), in which the intensity of the photostimulated luminescence (PSL) was proportional to the adsorbed radiation energy on the imaging plate. The amount of CDP-ME2P was determined by the estimation of the PSL of the corresponding spot with a BAS-1500 (Fujifilm).

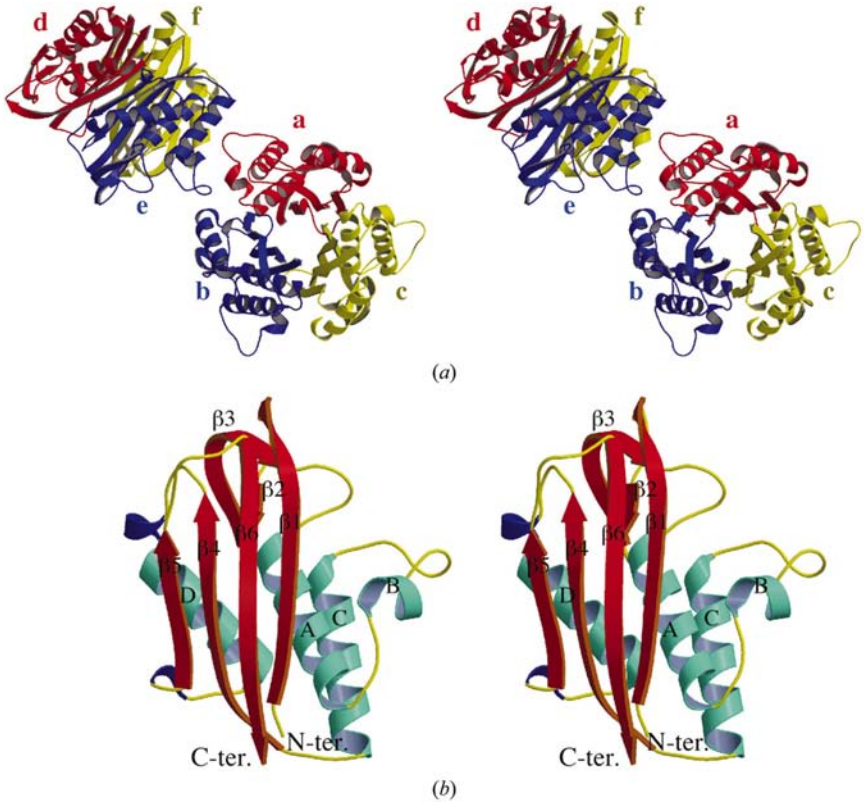


Figure 2
Stereoviews of the overall structure (*a*) and monomer structure (*b*). (*a*) shows the contents of the asymmetric unit, with six monomers arranged into two trimers. Refinement was completed without using non-crystallographic constraints or restraints. Overall, the trimers in the asymmetric unit are extremely similar with r.m.s. deviations of 0.23 Å over all main-chain atoms. The tertiary structure is shown in (*b*); the threefold axis runs up and down this view. Figures were generated using *MOLSCRIPT* (Kraulis, 1991) and rendered with *Raster3D* (Merritt & Murphy, 1994).

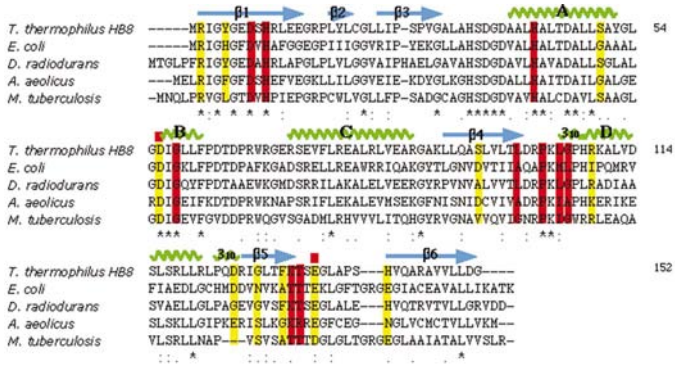


Figure 3
Sequence alignment of MECDP synthases from a variety of bacteria. Highlighted residues indicate intersubunit (yellow) and substrate (red) contacts. Although many residues at the subunit interfaces are not conserved between the proteins from *E. coli* and *T. thermophilus*, the overall structure of the enzymes is very similar. The substrate-binding site similarly shows a number of substantial mutations. The *E. coli* enzyme is unique among the sequences shown in not having a positively charged residue at position 109, close to the ribose ring binding site. Lys132 also extends into the substrate-binding site and contacts the β-phosphate. This residue is replaced by threonine in several bacterial homologues.

3. Results

3.1. Overall structure

Sedimentation-equilibrium measurements by analytical ultracentrifugation showed the 152-residue protein to have a solution molecular weight almost exactly (within 1%) that expected for a trimer, 49 560 Da (data not shown). Crystals grew in space group $P4_12_12$ from polyethylene glycol (PEG) using the hanging-drop method. Since the native protein contains a single methionine (at the N-terminus), site-directed mutations were created by the polymerase chain reaction (PCR) to change three leucine residues (41, 81, 120) to methionine. The mutant protein was expressed with selenomethionine in place of methionine and phasing was then achieved by multiple-wavelength anomalous dispersion (MAD). In contrast to the crystal form of the *E. coli* protein described by two separate groups, the *T. thermophilus* protein (mutant and native) crystallizes with six monomers in the asymmetric unit, arranged as two trimers of closely associated subunits (Fig. 2). The *E. coli* protein crystallizes in space group $I2_13$, the trimer being generated by the crystallographic threefold axis, or in space group $P2_12_12_1$ with a trimer in the asymmetric unit.

Since building our first model, three groups have reported the crystal structure of the *E. coli* enzyme (Richard *et al.*, 2002; Steinbacher *et al.*, 2002; Kemp *et al.*, 2002). Each group noted that the protein shows similar topology to the *E. coli* YjgF protein (PDB code 1qu9) and chorismate mutase (PDB code 2chs). MECDP synthase is well conserved among eubacteria, as shown in Fig. 3. The overall shape of the proteins from *E. coli* and *T. thermophilus* is therefore highly similar, each forming a trimer about 40 Å in height along the symmetry axis and about 60 Å across. Each monomer forms a single domain

including two β-sheets and four α-helices (Fig. 2). The larger β-sheet has four strands which lie roughly parallel to the threefold axis at the core of the molecule. Richard *et al.* (2002) collected only low-resolution X-ray data. They describe the protein as a 'β-barrel' and tentatively identify the electron density running along the narrow hydrophobic channel along the molecular (and crystallographic) threefold symmetry axis in their structure as an isoprenoid-like molecule. Rather than bind together through main-chain to main-chain hydrogen bonds between these β-sheets, the subunits pack through side-chain interactions. The protein is not therefore a conventional 'β-barrel'; the edge-to-face packing of the β-sheets gives a much narrower central channel than a true barrel. No such channel exists in the thermophilic protein, suggesting tighter binding of the monomers. In the *E. coli* protein, His5 and Glu149 from each subunit form salt bridges, but other residues lining the channel are apolar. In *T. thermophilus* MECDP synthase, His5 is replaced by tyrosine and the tyrosine hydroxyl groups form three hydrogen bonds with each other around the molecular symmetry axis. Glu7 and Arg145 form a cyclic salt-bridge network with their symmetry mates, the arginine residues coming close together and filling the central space. Some monomer contacts are preserved between the *E. coli* and *T. thermophilus* proteins; for example, Arg2 is a highly conserved residue on the first β-strand which forms a salt bridge with Asp125 on a neighbouring subunit. Many of the contacts are altered, however. Of eight apolar residues in the mesophilic protein involved in hydrophobic contacts between the β-sheets of different monomers, only four have apolar equivalents in the thermophilic protein. Other notable changes include Ile109 in the *E. coli* protein, which is altered to arginine in the *T. thermophilus* enzyme. The arginine side chain forms a salt bridge with the well conserved Asp56, which is involved in contacts to the substrate (see later). Since arginine is commonly found at this position, for example in the *M. tuberculosis* enzyme, it is unlikely to be related to thermostability. While different in detail, the architecture of the *E. coli* and *T. thermophilus* proteins is very similar. The latter protein relies less on apolar contacts and more on hydrogen bonds and salt bridges, however, presumably owing to the weaker hydrophobic effect at high temperatures.

The very similar quaternary structure of MECDP synthase from *E. coli* and *T. thermophilus* suggests that all members of this family form trimers. This is interesting since the *ygbB* gene is fused to the *ygbP* gene to encode a single bifunctional polypeptide in a number of bacteria (Herz *et al.*, 2000). *ygbP* encodes the enzyme MEP cytidylyltransferase, which catalyzes the reaction two steps prior to MECDP synthase. The structure of *E. coli* MEP cytidylyltransferase has been solved and is found to be a dimer, with substrate interactions with both polypeptide chains (Richard *et al.*, 2001). Perhaps the fused YgbP–YgbB polypeptide forms hexamers. No substrate channelling occurs between the two active sites in these proteins, however, as an intermediate phosphorylation is required to generate the required cyclodiphosphate.

3.2. Active sites

The previously reported structures of *E. coli* MECDP synthase include those of several crystals into which substrate or analogues have been soaked. These experiments showed the active sites to be three large symmetry-related cavities stretching across two subunits of the trimer (Fig. 4). The binding site stretches from the N-termini of helices *A* and *C* in one subunit to the C-termini of strands 4 and 5 on the next. *T. thermophilus* MECDP synthase was co-crystallized with both substrate, 2-phospho-4-(cytidine-5′-diphospho)-2-*C*-methyl-D-erythritol (CDP-ME2P), and its precursor in the pathway, 4-(cytidine-5′-diphospho)-2-*C*-methyl-D-erythritol (CDP-ME), which lacks the phosphate group involved in nucleophilic attack catalysed by the enzyme. The cytidyl group binds at the loop, linking β4 and helix *D* to residues Leu100, Pro103, Leu105 and Gly106 (Figs. 2, 3 and 5). These residues are not particularly well conserved with respect to the *E. coli* sequence, but the mode of binding is very similar, with hydrogen bonding to the cytidyl group through main-chain atoms (Fig. 3). The side chain of Ala131 in the *E. coli* protein contacts the pyrimidine base, but this residue is changed to Phe, which leads to a shift in the loop just before helix *D*. Tight substrate binding is indicated by the very well defined electron density around both the cytidyl and the ribose ring. The latter is held by Asp56′ and Gly58′ (prime indicating the neighbouring subunit) which are both extremely well conserved. Asp56 forms an Asx turn, the side-chain carboxyl H atom bonding to the main-chain N atom of Leu59. The strong conservation of the glycine suggests its flexibility is necessary for this turn to form. The replacement of Ile109 with arginine brings the guanidino group within 3.6 Å of one of

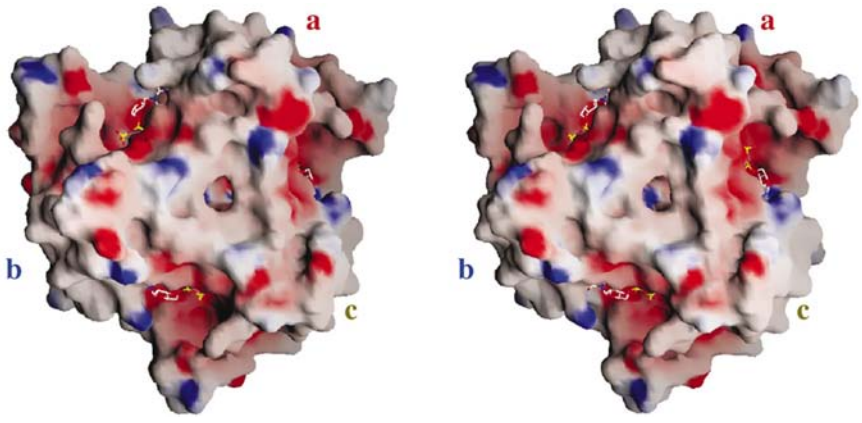


Figure 4
The van der Waals surface of the protein coloured according to electrostatic potential, red indicating negative charge and blue positive. This figure was produced using *GRASP* (Nicholls & Honig, 1991). The contour levels are −10 to +10 e Å^{-3}. The protein is viewed along the molecular threefold axis from the bottom as seen in Fig. 2(*b*). The active sites are shown by the CDP stick models fitted to the electron density from the enzyme co-crystallized with CDP-ME.

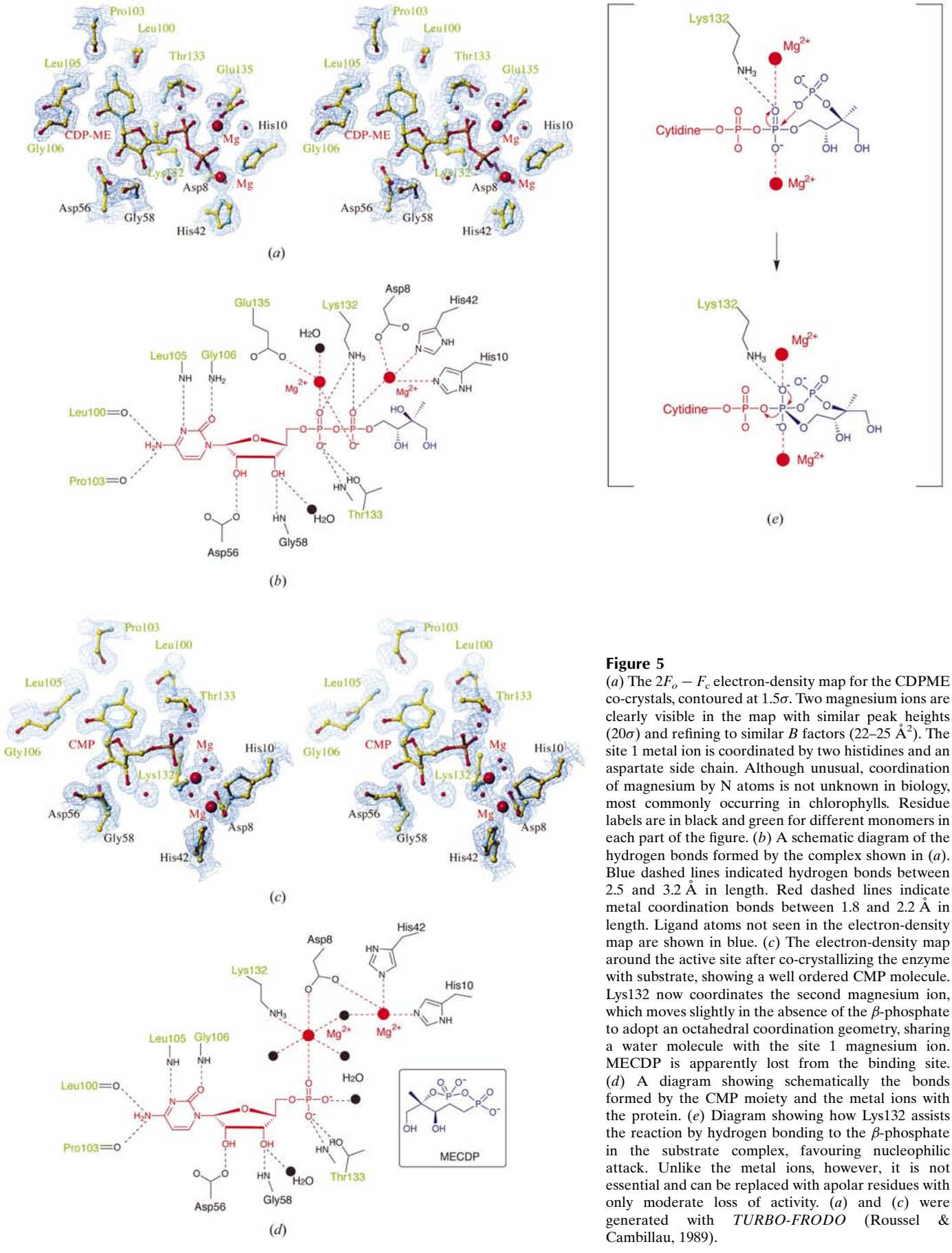


Figure 5
(*a*) The $2F_o - F_c$ electron-density map for the CDPME co-crystals, contoured at 1.5σ. Two magnesium ions are clearly visible in the map with similar peak heights (20σ) and refining to similar *B* factors (22–25 Å^2). The site 1 metal ion is coordinated by two histidines and an aspartate side chain. Although unusual, coordination of magnesium by N atoms is not unknown in biology, most commonly occurring in chlorophylls. Residue labels are in black and green for different monomers in each part of the figure. (*b*) A schematic diagram of the hydrogen bonds formed by the complex shown in (*a*). Blue dashed lines indicated hydrogen bonds between 2.5 and 3.2 Å in length. Red dashed lines indicate metal coordination bonds between 1.8 and 2.2 Å in length. Ligand atoms not seen in the electron-density map are shown in blue. (*c*) The electron-density map around the active site after co-crystallizing the enzyme with substrate, showing a well ordered CMP molecule. Lys132 now coordinates the second magnesium ion, which moves slightly in the absence of the β-phosphate to adopt an octahedral coordination geometry, sharing a water molecule with the site 1 magnesium ion. MECDP is apparently lost from the binding site. (*d*) A diagram showing schematically the bonds formed by the CMP moiety and the metal ions with the protein. (*e*) Diagram showing how Lys132 assists the reaction by hydrogen bonding to the β-phosphate in the substrate complex, favouring nucleophilic attack. Unlike the metal ions, however, it is not essential and can be replaced with apolar residues with only moderate loss of activity. (*a*) and (*c*) were generated with *TURBO-FRODO* (Roussel & Cambillau, 1989).

the ribose O atoms, but its catalytic importance is unclear. Lys132 and Thr133 contact the diphosphate group of the substrate (Figs. 5*b* and 5*e*). Unexpectedly, electron density for the protein co-crystallized with substrate (CDP-ME2P) shows strong density for CMP but no sign (even at low contour levels) of the other product, MECDP (Fig. 5*c*). The enzyme has cleaved CMP from the substrate and apparently lost MECDP. A similar result was obtained using CDP-ME. In this case the diphospho group and cytidyl are very well defined in the electron density, but the remainder of the ligand in the binding site cannot be detected. This contrasts with the previously described results obtained with the *E. coli* enzyme. Soaking crystals of *E. coli* MECDP synthase with substrate gives clear electron density for both products (in the presence of a suitable metal ion) and soaking with CDP-ME allows the entire molecule to be seen in the binding pocket. Since CDP-ME is slightly unstable, it appears hydrolysis has occurred yielding CDP (see §3.4 below).

Unlike the pockets holding the pyrimidine and ribose ring, the residues involved in binding the 2-*C*-methyl-D-erythritol moiety are rather flexible. Residues between helix *B* and helix *C*, Pro62–Ala71, form a mobile loop which is poorly ordered in the apo *E. coli* enzyme, but which becomes well ordered in the presence of the reaction products (Steinbacher *et al.*, 2002). The electron density for the *T. thermophilus* protein shows this region is very well ordered in both apo and ligand-bound forms (contoured at 1.5σ, the electron-density map shows a hole through the ring of Pro66). The apparently more rigid loop does not lead to a slower reaction rate, however, the heat-stable enzyme being significantly faster than the *E. coli* equivalent at 310 K. Asp63 is highly conserved and forms hydrogen bonds with two phosphate O atoms in the MECDP complex of the *E. coli* protein (Steinbacher *et al.*, 2002). The main-chain N atoms of His34 and Ser35 (both absolutely conserved residues) hydrogen bond to the MECDP in the active site, pulling the histidine into a position where its side chain can hydrogen bond to the carbonyl of Asp65 by lying over the substrate. In the *T. thermophilus* enzyme, both apo and liganded with CDP-ME, Asp63 occupies a position close to that of His34 in the MECDP complex of the *E. coli* protein. His34 is pushed away from the active site and points towards the solvent. Concerted movements of these residues appear alternately to open the pocket and to hold the substrate in place.

3.3. Metal-binding sites

The reaction catalysed by MECDP synthase involves intramolecular attack by the 2-phosphate group on the β-phosphate, CMP acting as the leaving group to give the cyclic diphosphate product MECDP. While conceptually simple, this requires a means of overcoming the strong electrostatic repulsion of the phosphate groups, which is the key role played by the essential metal ions found in the structure. Steinbacher *et al.* (2002) found intact substrate in the active site in the presence of the metal chelator EDTA. In the other structures they describe, two metal sites are identified, one being a magnesium bound to the diphosphate. The other metal ion, bonded to Asp8, His10 and His42, they identified as zinc, but do not describe how they did so. Richard *et al.* (2002) identified only a single metal (at the second site) as a 'putative cation', pointing out that the two histidines are unusual ligands for magnesium in proteins and suggest that the metal site represents Mn^{2+} or even Ni^{2+} instead, these metals being derived from the *E. coli* expression system. Rohdich and coworkers have in fact shown that the *E. coli* enzyme is active in the presence of Mg^{2+} or Mn^{2+} ions (Herz *et al.*, 2000). The orthorhombic crystal form of the *E. coli* protein was grown in the presence of manganese ions and CDP (Kemp *et al.*, 2002). Using atomic absorption spectroscopy and anomalous dispersion measurements with these crystals, Hunter and coworkers showed the histidine-coordinated metal ion to be zinc and the second metal ion to be manganese. In both our product-bound and CDP-ME-bound structures of the protein we find two metal ion sites. At the high resolution obtained from the tetragonal crystal form, it is possible to refine the temperature (*B*) factor of the individual atoms reliably. This indicates that both metal ions are magnesium, with very similar *B* factors and peak electron density. While it is unusual to find a magnesium ion coordinated by histidines, it is not unique (for example, the magnesium at the centre of chlorophyll has an axial histidine ligand as well as four pyrrole nitrogen ligands). We refer to the Zn-atom site of Steinbacher *et al.* (2002) as 'site 1' and the site between the α and β phosphate groups as 'site 2'. Site 2 has the usual octahedral coordination associated with magnesium, but site 1 does not. These are the same sites found in the *T. thermophilus* enzyme cocrystallized with CDP-ME; in the presence of CMP, the site 2 magnesium ion moves to a position (site 3) close to that vacated by the β-phosphate. Here it shares a coordination water molecule with the invariant site 1

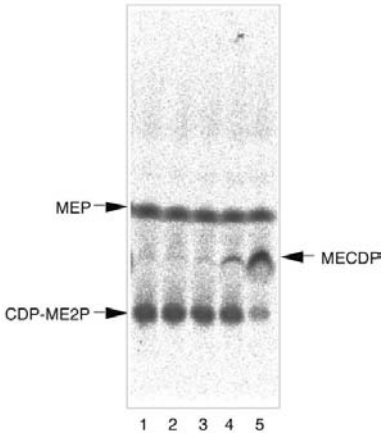


Figure 6
MECDP synthase activity of wild-type and mutant enzymes. Four mutant enzymes were tested together with native MECDP synthase from *T. thermophilus*. Lane 1, D8L; lane 2, H10F; lane 3, H42F; lane 4, K132L; lane 5, wild type. The assay was carried out at 310 K for 60 min. Replacing any of the residues coordinating the site 1 magnesium ion with phenylalanine abolished activity. Considerable activity is found for the K132L mutant, showing that the interactions of the lysine with both the substrate and the magnesium ions are not essential for catalysis.

metal ion. The β phosphate of the substrate is coordinated by both magnesium ions which hold it in a suitable position for attack by the 2-phosphate group and neutralize its charge.

In common with several other bacterial homologues, *T. thermophilus* MECDP synthase has a lysine residue in place of Thr132. This reaches towards the substrate and coordinates the site 3 Mg^{2+} ion in the presence of CMP. With CDP in the active site, the lysine hydrogen bonds to the diphosphate, possibly increasing the rate of nucleophilic attack on the β-phosphate of the substrate (Fig. 5*b*). At 310 K, the thermostable enzyme is significantly more active than the *E. coli* enzyme (data not shown). Four single-site mutants of MECDP synthase were expressed and purified for enzyme assays. Three of these changed the side chains chelating magnesium ions and all three eliminated enzyme activity (Fig. 6). In contrast, mutating Lys132 to leucine gave a reduced but substantial activity.

3.4. Catalytic mechanism

The 2-*C*-methyl-D-erythritol moiety is absent from the electron-density maps obtained by co-crystallization with both CDP-ME and CDP-ME2P. Herz *et al.* (2000) have shown that the *E. coli* enzyme can break down CDP-ME to yield a non-natural cyclophosphate product and CMP, but we see no evidence of this reaction since CDP is extremely well ordered in the binding site. It is possible that the ME is removed from the analogue by spontaneous hydrolysis, since CDP-ME appears to be slightly unstable in solution (TK, unpublished observation). Even so, the interactions with the 2-*C*-methyl-D-erythritol group are clearly rather weaker than those formed with the cytidyl moiety. In order to effect catalysis, the protein must lower the free energy of the transition state. MECDP synthase appears to lower both the entropy and enthalpy of activation by positioning the attacking phosphate appropriately and countering the negative charge around the P atom being attacked.

Many enzymes show an apparent preference for catalysing reactions in one direction; the laws of thermodynamics are not violated by this behaviour since the overall rate of catalysis depends on the binding of substrate, as described by the Haldane relationship (Jencks, 1986). For the simple reaction

$$\mathrm{E} + \mathrm{S} \underset{K_S}{\leftrightarrow} \mathrm{ES} \overset{k_f}{\underset{k_r}{\leftrightarrow}} \mathrm{EP} \underset{K_P}{\leftrightarrow} \mathrm{E} + \mathrm{P}$$

the overall equilibrium constant K_{ov} between substrate and product is given by the relationship

$$K_{\mathrm{ov}} = \frac{k_f}{k_r}\frac{K_P}{K_S}.$$

If the dissociation constants of both substrate and product (K_S and K_P) are the same, then K_{ov} is given by the ratio of the initial reaction rates in the forward and backwards directions (k_f/k_r). If the enzyme binds the product much more tightly than the substrate, it can accelerate the forward reaction more than the reverse, at the possible risk of product inhibition. The structures of *T. thermophilus* MECDP synthase show very well ordered CMP or CDP in the binding site, but no other parts of the substrate. Binding energy derived from interactions with the cytidyl moiety does not appear to be used to drive the reaction, only to lower K_S (and K_P). Catalysis is achieved through interactions with the metal ions and the mobile flap which orders on substrate binding, holding the 2-*C*-methyl-D-erythritol 2-phosphate moiety in a suitable configuration for the reaction to proceed. Our structures therefore suggests that *in vitro* the enzyme may be susceptible to product inhibition, though naturally *in vivo* CMP is constantly converted to CTP. Inhibitor design may be possible by exploiting fully the interaction energy available from the relatively rigid pockets into which the pyrimidine base and ribose ring fit. The differences found between the CMP-binding sites in the enzymes from *E. coli* and *T. thermophilus* also suggest inhibitors can be found which show some specificity for particular species.

4. Conclusions

The high-resolution crystal structures of *T. thermophilus* MECDP synthase in the presence and absence of ligands provide several important insights into the structure and mechanism of the enzyme. Although residues mediating the subunit interactions are not especially well conserved, comparison with the *E. coli* enzyme shows the overall protein fold and quaternary structure are extremely similar. Unlike Richard *et al.* (2002), who found an equilibrium of *E. coli* MECDP synthase monomer and trimer, analytical ultracentrifugation of the thermostable protein showed no detectable trace of monomer. Tighter packing is also implied by the absence of any channel through the centre of the molecule as described by Richard *et al.* (2002). The catalytic mechanism does not require any substantial movement of the protein, much of the binding site remaining rigidly fixed in the presence and absence of substrate or product. This probably partly explains the very high activity of the thermostable protein even at 310 K; many enzymes from thermophilic bacteria lose activity at moderate temperatures, becoming too rigid for catalytically necessary movements. The mobile flap of residues 62–71 which contacts the 2-*C*-methyl-D-erythritol moiety is more ordered in the thermophilic enzyme. Unlike the *E. coli* enzyme, however, binding CDP-ME to the *T. thermophilus* protein does not show ordered ligand beyond the diphosphate in the electron-density map. It is possible that the 2-phosphate group is needed to tie down the substrate in the magnesium-bound *T. thermophilus* protein. It is more likely that the CDP-ME is hydrolysed to release CDP, which appears to bind tightly to the enzyme. Tight binding of the substrate is not required for catalysis; in fact, CDP-ME2P is rather unstable and naturally cyclizes to release CDP. MECDP synthase redirects the attacking phosphate to give CMP instead. Our crystal structures demonstrate the enzyme has higher affinity for CMP than MECDP and this CMP pocket may be a suitable target for drug design.

We thank Dr S. Adachi for advice on data collection at SPring-8.

References

Adachi, S., Oguchi, T., Tanida, H., Park, S.-Y., Shimizu, H., Miyatake, H., Kamiya, N., Shiro, Y., Inoue, Y., Ueki, T. & Iizka, T. (2001). *Nucl. Instrum. Methods A*, **467–468**, 711–714.

Brünger, A. T. (1996). *X-PLOR Version 3.851 Manual*. Yale University, New Haven, CT, USA.

Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* D**50**, 760–763.

Hendrickson, W. A., Horton, J. R. & LeMaster, D. M. (1990). *EMBO J.* **9**, 1665–1672.

Herz, S., Wungsintaweekul, J., Schuhr, C. A., Hecht, S., Luttgen, H., Sagner, S., Fellermeier, M., Eisenreich, W., Zenk, M. H., Bacher, A. & Rohdich, F. (2000). *Proc. Natl Acad. Sci. USA*, **97**, 2486–2490.

Jencks, W. P. (1986). *Catalysis in Chemistry and Enzymology*. New York: Dover.

Jomaa, H., Wiesner, J., Sanderbrand, S., Altincicek, B., Weidemeyer, C., Hintz, M., Turbachova, I., Eberl, M., Zeidler, J., Lichtenthaler, H. K., Soldati, D. & Beck, E. (1999). *Science*, **285**, 1573–1576.

Kemp, L. E., Bond, C. S & Hunter, W. N. (2002). *Proc. Natl Acad. Sci. USA*, **99**, 6591–6596.

Kraulis, P. J. (1991). *J. Appl. Cryst.* **24**, 946–950.

Kuzuyama, T. (2002). *Biosci. Biotechnol. Biochem.* **66**, 1619–1627.

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

Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. (1993). *J. Appl. Cryst.* **26**, 283–291.

LeMaster, D. M. & Richards, F. M. (1985). *Biochemistry*, **24**, 7263–7268.

Merritt, E. A. & Murphy, M. E. P. (1994). *Acta Cryst.* D**50**, 869–873.

Nicholls, A. & Honig, B. (1991). *J. Comput. Chem.* **12**, 435–445.

Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.

Perrakis, A., Morris, R. & Lamzin, V. S. (1999). *Nature Struct. Biol.* **6**, 458–463.

Qureshi, N. & Porter, J. W. (1981). *Biosynthesis of Isoprenoid Compounds*, edited by J. W. Porter & S. L. Spurgeon, Vol. 1, pp. 47–94. New York: John Wiley.

Reuter, K., Sanderbrand, S., Jomaa, H., Wiesner, J., Steinbrecher, I., Beck, E., Hintz, M., Klebe, G. & Stubbs, M. T. (2002). *J. Biol. Chem.* **277**, 5378–5384.

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., Ferrer, J.-L., Bowmann, M. E., Lillo, A. M., Tetzlaff, C. N., Cane, D. E. & Noel, J. P. (2002). *J. Biol. Chem.* **277**, 8667–8672.

Rohdich, F., Hecht, S., Gärtner, K., Adam, P., Krieger, C., Amslinger, S., Arigoni, D., Bacher, A. & Eisenreich, W. (2002). *Proc. Natl Acad. Sci. USA*, **99**, 1158–1163.

Rohmer, M., Knani, M., Simonin, P., Sutter, B. & Sahm, H. (1993). *Biochem. J.* **295**, 517–524.

Roussel, A. & Cambillau, C. (1989). *Silicon Graphics Geometry Partners Directory*, pp. 77–78. Silicon Graphics, Mountain View, CA, USA.

Sheldrick, G. M. & Schneider, T. R. (1997). *Methods Enzymol.* **277**, 319–343.

Steinbacher, S., Kaiser, J., Wungsintaweekul, J., Hecht, S., Eisenreich, W., Gerhardt, S., Bacher, A. & Rohdich, F. (2002). *J. Mol. Biol.* **316**, 79–88.

Takagi, M., Kuzuyama, T., Kaneda, K., Dairi, T. & Seto, H. (2000). *Tetrahedron Lett.* **41**, 3395–3398.

Terwilliger, T. C. & Berendzen, J. (1999). *Acta Cryst.* D**55**, 849–861.

Yajima, S., Nonaka, T., Kuzuyama, T., Seto, H. & Ohasawa, K. (2002). *J. Biochem (Tokyo)*, **131**, 313–317.