Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

Structure of a tetragonal crystal form of *Escherichia coli* 2-*C*-methyl-D-erythritol 4-phosphate cytidylyltransferase

Lauris E. Kemp, Charles S. Bond and William N. Hunter*

Division of Biological Chemistry and Molecular Microbiology, School of Life Sciences, University of Dundee, Dundee DD1 5EH, Scotland

Correspondence e-mail: w.n.hunter@dundee.ac.uk

Received 15 September 2002
Accepted 23 December 2002

PDB Reference: 2-C-methyl-D-erythritol 4-phosphate cytidylyltransferase, 1h3m, r1h3msf.

2-*C*-Methyl-D-erythritol 4-phosphate cytidylyltransferase is an essential enzyme in the mevalonate-independent pathway of isoprenoid biosynthesis. The structure of a tetragonal crystal form has been solved by molecular replacement and refined to 2.4 Å resolution. Structure and sequence comparisons suggest that the enzyme is a suitable target for a structure-based approach to the development of novel broad-spectrum antibiotics. However, the absence of ligands in the enzyme active site together with the moderate resolution of the structure indicates that this tetragonal crystal form is inferior to that of a previously reported highly ordered monoclinic form [Richard *et al.* (2001), *Nature Struct. Biol.* **8**, 641–647].

1. Introduction

Isoprenoids are a diverse range of natural products which participate in many biochemical reactions and also constitute structural components of membranes (Rohmer, 1999; Sacchettini & Poulter, 1997). Isoprenoid biosynthesis depends on the production of a C_5 precursor unit, isopentenyl diphosphate (IPP); in mammals, higher plants, fungi and certain bacteria, this occurs through the mevalonate (MVA) pathway (Boucher & Doolittle, 2000; Lange *et al.*, 2000). In chloroplasts, algae, cyanobacteria, many bacteria and the apicomplexa, a different metabolic pathway termed the non-mevalonate or 2-*C*-methyl-D-erythritol 4-phosphate (MEP) pathway generates IPP (Rohdich *et al.*, 2001). The non-mevalonate biosynthetic pathway is absent from mammals, but present in and the only route to IPP in many serious pathogenic microbes. Genetic approaches have validated several enzymes in this pathway by showing that they are essential for viability in *Escherichia coli* and *Bacillus subtilis* (Freiberg *et al.*, 2001). In addition, the antibacterial compound fosmidomycin has been identified as a potent inhibitor of deoxyxylulose reductoisomerase, the second enzyme in the pathway, thereby providing chemical validation of a therapeutic target (Kuzuyama, 2002). The constituent enzymes of the pathway are now attracting widespread interest as targets for structure-based antimicrobial drug design (Jomaa *et al.*, 1999; Kemp *et al.*, 2002; Ridley, 1999).

Of specific interest here is the third reaction in the MEP pathway: the formation of 4-diphosphocytidyl-2-*C*-methyl-D-erythritol from MEP and CTP, which is catalysed by 2-*C*-methyl-D-erythritol 4-phosphate cytidylyltransferase (MEPC; EC 2.7.7.60), alternatively called 4-diphosphocytidyl-2-*C*-methylerythritol synthetase (Rohdich *et al.*, 1999). This is one of the genetically validated enzymes (Freiberg *et al.*, 2001). We obtained crystals of recombinant *E. coli* MEPC (Kemp *et al.*, 2001); whilst our analysis was under way, Richard *et al.* (2001) published an excellent and comprehensive study of the enzyme structure and mechanism based on a highly ordered monoclinic crystal form. We now present the analysis of our tetragonal crystal form, brief comparisons with the already published structure and for the first time provide a structure/sequence comparison of the *E. coli* enzyme with the corresponding enzymes from a number of pathogenic bacteria.

2. Material and methods

Experimental details covering sample preparation have been described previously (Kemp *et al.*, 2001). In brief, crystals were obtained by vapour diffusion of a drop [2 µl of 9 mg ml^{-1} protein, 100 m*M* NaCl, 100 m*M* Tris–HCl pH 7.7 plus 0.6 µl of reservoir comprising 0.1 *M* sodium acetate pH 5.6, 25% PEG 4000, 0.2 *M* ammonium sulfate and, as an essential additive, 0.4 µl 30%(*w*/*v*) 1,5-diaminopentane dihydrochloride] against a reservoir solution at 293 K. Inferior crystals, which barely diffracted to 3.0 Å resolution, were grown in the presence of a number of additives including dithiothreitol. A crystal was cryopreserved by soaking in substituted mother liquor (0.2 *M* ammonium sulfate, 0.1 *M* sodium acetate, 25% PEG 4000 and 18% 2-methyl-2,4-pentanediol) for 10 s and then maintained at 100 K in a stream of nitrogen gas. Data were collected at station BM14 at the European Synchrotron Research Facility

(ESRF, Grenoble, France) and processed using *DENZO*/*SCALEPACK* (Otwinowski & Minor, 1997). The space group is $P4_12_12$, with unit-cell parameters $a = b = 73.60$, $c = 175.56$ Å. MEPC is a homodimer with a subunit of 236 amino acids and a molecular mass of 25.7 kDa (Rohdich *et al.*, 1999); the asymmetric unit contains one dimer (subunits *A* and *B*). The *CCP*4 suite of programs was used for the analysis (Collaborative Computational Project, Number 4, 1994). Each subunit was independently positioned by molecular replacement (*CNS*; Brünger *et al.*, 1998) and refined using *REFMAC* (Murshudov *et al.*, 1997) combined with graphics inspection and model/map fitting with *O* (Jones *et al.*, 1991). A conservative approach to the identification of water molecules was adopted and a bulk-solvent scattering correction was applied. Non-crystallographic symmetry restraints were imposed in the early stages of the refinement; their release for the final calculations reduced R_{work} and R_{free} by 1%.

The model consists of residues 6*A*–15*A*, 25*A*–229*A*, 5*B*–16*B*, 24*B*–229*B*, five chloride ions, 82 water molecules and one 1,5-diaminopentane, the essential additive. Several residues are poorly ordered and all or some of their atoms have been assigned zero occupancy; this applies to Gln100*A*, Glu172*A*, Glu24*B*, Asp83*B* and Pro60*B*. Dual conformations, each with an occupancy of 0.5, are included for Gln67*A* and Asn183*B*. Our sequence differs from that previously reported for *E. coli* MEPC in that residue 91 is a threonine not alanine; this results from a single base change in the gene and is either a cloning artifact or a naturally occurring sequence variation. Thr91 is at the periphery of the molecule, distant from the active site (Fig. 1), and is unlikely to influence enzyme activity.

A concern that arose during refinement was the discrepancy between R_{work} and R_{free}, which at the end of refinement was 9.4%. This was surprising given the data-processing statistics appeared to be reasonable (Table 1) and that the electron density which could be modelled was of high quality. For example, the maps clearly indicated dual conformations and identified a valine or threonine replacing alanine at position 91, subsequently confirmed to be a threonine by DNA sequencing. However, there are areas of the electron-density maps for which strong features were observed but which we could not interpret in terms of the protein model; we presume that static or conformational disorder contributes to this high level of noise. This in particular applies to amino acids 15–25 in both subunits. We conclude that the difference between R_{work} and R_{free} is most likely to contain a significant contribution from the density that we have been unable to model.

Table 1
Data-collection, refinement and model-geometry statistics.

Values in parentheses refer to the highest resolution bin.

Crystal dimensions (mm)	0.1 × 0.1 × 0.4
Resolution (Å)	2.4
Wavelength (Å)	0.97
No. of measurements	81222
No. of unique reflections	19619
Redundancy	4.1
Completeness (%)	99.5 (94.7)
$I/\sigma(I)$	37.0 (3.1)
R_{merge} (%)	3.1 (27.1)
Wilson B (Å^2)	59.2
Protein residues (total)	449
In subunit *A*	224
In subunit *B*	225
Protein atoms (total)	3316
In subunit *A*	1647
In subunit *B*	1669
R_{work} (%)	23.7
No. of reflections	18638
R_{free} (%)	33.3
No. of reflections	981
Average isotropic thermal parameters (Å^2)	
Subunit *A*	45.0
Subunit *B*	53.5
Main chain	48.6
Side chains	50.0
Solvents	49.2
R.m.s.d.s	
Bond lengths (Å)	0.016
Bond angles (°)	1.670
Ramachandran analysis	
Favoured regions (%)	91.2
Additionally allowed regions (%)	8.8

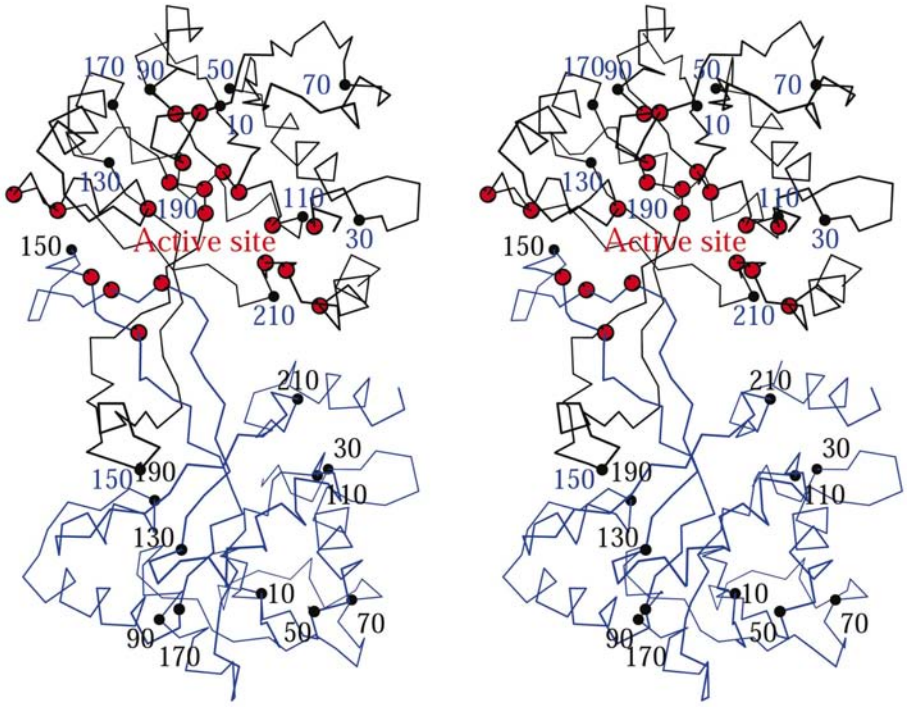

Figure 1
Stereoview C^α trace of the MEPC dimer (subunit *A*, cyan; subunit *B*, black). One active site (subunit *B*) is labelled and the C^α positions of strictly conserved residues around this active site are depicted as red spheres. Every 20th C^α position starting from residue 10, with the exception of residue 190*B*, is depicted as a black sphere and numbered in black for subunit *A* and in blue for subunit *B*. Residue Asp190*B* is strictly conserved in MEPC and is depicted with a red sphere at the active site.

Figures were prepared using the programs *MOLSCRIPT* (Kraulis, 1991) and *ESpript* (Gouet *et al.*, 1999). The amino-acid sequences for eight important bacterial pathogens were retrieved from EXPASY (http://www.expasy.ch) and aligned using *CLUSTALW* (Thomson *et al.*, 1994). Further experimental details are given in Table 1.

3. Results and discussion

3.1. Structure description and comparison with the monoclinic crystal form

The MEPC subunit is primarily a single-domain α/β structure, with similarities to the Rossmann fold, into which is inserted an extended 'β-arm' (residues 134–163; Fig. 1). Two arms associate providing the major contribution to the dimer interface, with a lesser contribution from side-chain interactions from residues 222–229 on the helix at the C-terminus. The dimer interface covers an area of approximately 1770 Å^2 and involves 26 intersubunit hydrogen bonds and four salt bridges. The diaminopentane

additive was observed to lie in a slight depression where the β-arms link, interacting with the side chains of Val155 from both subunits, hydrogen bonding to and bridging His153, again from both subunits (not shown).

The structures of the subunits are well conserved in the asymmetric unit. For example, the root-mean-square deviation (r.m.s.d.) between 215 C^{α} atoms shared by the subunits is 0.5 Å. The monoclinic form studied by Richard *et al.* (2001) has a single subunit per asymmetric unit, with the functional dimer created by the crystallographic twofold. This subunit has an r.m.s.d. of 0.9 and 1.1 Å when overlaid on subunits *A* (214 C^{α} atoms) and *B* (217 C^{α} atoms), respectively, of the tetragonal crystal form.

The enzyme active site is formed at the dimer interface by segments of the polypeptide comprising residues 15–25, 85, 104–109, 168–163, 189–190 and 213–215 from one subunit and 138–140 from the partner subunit. In the tetragonal form and in most structures of the monoclinic form (Richard *et al.*, 2001), the loop between residues 15 and 24 is disordered. An intermolecular disulfide bond is formed between Cys25*A* and a symmetry-related Cys25*A*. The formation of the disulfide bond may prevent the residues preceding Cys25 from closing over the active site, as observed in one of the ligand complexes (Richard *et al.*, 2001). We have analysed several data sets from this tetragonal crystal form that have been obtained in the presence of ligands (for example, CDP, CMP and divalent cations) and have failed to observe ordered electron density in the enzyme active site that would correspond to any of them.

3.2. Sequence-based comparison with MEPC from pathogens

The alignment of the *E. coli* protein with MEPC from eight bacterial species is presented in Fig. 2. The sequences were selected as being representative of the enzyme from serious pathogens. On the basis of pairwise alignments, the amino-acid identity ranges from 29% between *E. coli* and *Clostridium perfringens* to 91% between the *E. coli* and *Salmonella typhi* proteins. The average is approximately 51%. Approximately 120 amino acids would be considered highly conserved and 22 strictly conserved. Sequence conservation extends throughout the polypeptide but is particularly well maintained in elements of secondary structure and also in those regions of the protein that are involved in the dimer interface. Of note are the 22 amino acids strictly conserved in the nine sequences shown in Fig. 2. Of these, 21 contribute to the active site and dimerization interface. The one exception is Gly199, positioned on a loop distant from the catalytic centre.

The sequence conservation includes those residues whose side chains have been shown to interact with ligands. The recognition and interactions of *E. coli* MEPC with substrate and/or product involves 22 residues (Richard *et al.*, 2001), of which 18 participate in hydrogen-bonding interactions. The ten residues that use only main-chain functional groups are Pro13–Gly18, Gly82, Asp83, Ala107 and Ala163. Although well conserved (seven are identical in more than seven sequences; Fig. 2), there is some leeway in the residue types that can be accommodated. Residues that use side chains are more highly conserved. Arg20 and Thr140 use both main-chain and side-chain functional groups, whilst Lys27, Asp106, Arg109, Arg157 and Lys213 utilize only side-chain groups. The remaining residue that hydrogen bonds to substrates is Ser88, which is conserved in seven out of nine sequences; it is conservatively replaced by threonine in the remaining two sequences.

The dominating feature of the enzyme active site is the preponderance of basic side chains involved in binding and processing substrates. In particular, Arg20 and Lys27 bind CTP phosphate groups, assist polarization for nucleophilic attack and also stabilize the negatively charged pentacoordinate transition state. Arg157 and Lys213 serve to bind and position the attacking nucleophile, MEP phosphate. These four basic residues are major contributors to the enzyme mechanism and are strictly conserved in MEPC sequences (Fig. 2).

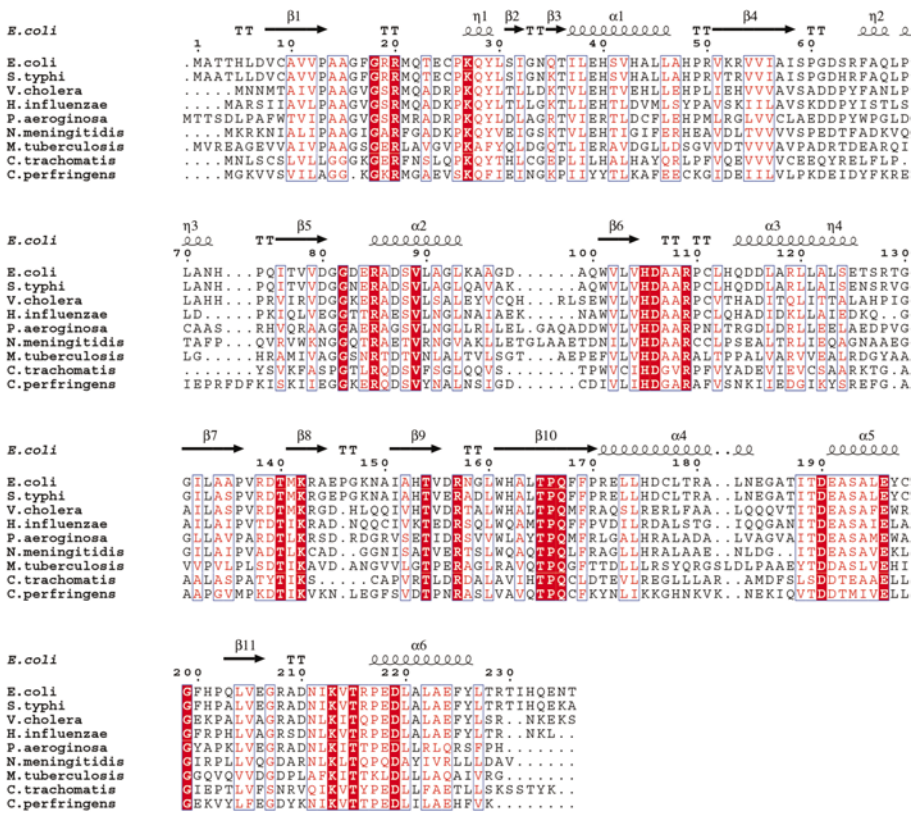

Figure 2
Sequence alignment of MEPC from *E. coli*, *Salmonella typhi*, *Vibrio cholera*, *Haemophilus influenzae*, *Pseudomonas aeruginosa, Neisseria meningitidis*, *Mycobacterium tuberculosis*, *Chlamydia trachomatis* and *Clostridium perfringens*. All of these organisms use only the non-mevalonate pathway to IPP. Elements of secondary structure assigned from the crystal structure are shown; residues whose identity is strictly conserved in all sequences are boxed in red and highly conserved residues are coloured red.

4. Conclusions

We have determined the structure of a tetragonal crystal form of *E. coli* MEPC at 2.4 Å resolution. Our structure/sequence analysis suggests that the *E. coli* enzyme is an appropriate model for MEPC from several serious human pathogens and is therefore suitable for inhibitor screening and for use in structure-based drug discovery aimed at developing broad-

spectrum antibiotics. However, despite including ligands in the crystallization solutions, the active sites are devoid of any ordered electron density. The tetragonal crystal form is therefore inappropriate for accurate studies of enzyme–ligand interactions.

This work was funded by a Wellcome Trust Senior Fellowship (WNH), an EPSRC–BBSRC studentship (LEK) and a BBSRC Sir David Phillips Research Fellowship (CSB). We thank ESRF for beam time and Dr Gordon Leonard for support.

References

Boucher, Y. & Doolittle, W. F. (2000). *Mol. Microbiol.* **37**, 703–716.

Brünger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J.-S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R., Rice, L. M., Simonson, T. & Warren, G. L. (1998). *Acta Cryst.* D**54**, 905–921.

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

Freiberg, C., Wieland, B., Spaltmann, F., Ehlert, K., Brotz, H. & Labischinski, H. (2001). *J. Mol. Microbiol. Biotechnol.* **3**, 483–489.

Gouet, P., Courcelle, E., Stuart, D. I. & Metoz, F. (1999). *Bioinformatics*, **15**, 305–308.

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**, 1573–1575.

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

Kemp, L. E., Bond, C. S. & Hunter, W. N. (2001). *Acta Cryst.* D**57**, 1189–1191.

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.

Lange, B.M., Rujan, T., Martin, W. & Croteau, R. (2000). *Proc. Natl. Acad. Sci. USA*, **97**, 13172–13177.

Murshudov, G. N., Vagin, A. A. & Dodson, E. J. (1997). *Acta Cryst.* D**53**, 240–255.

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

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**, 641–647.

Ridley, R. G. (1999). *Science*, **285**, 1502–1503.

Rohdich, F., Kis, K., Bacher, A. & Eisenreich, W. (2001). *Curr. Opin. Chem. Biol.* **5**, 535–540.

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). *Comprehensive Natural Products Chemistry*, Vol. 2, *Isoprenoids including Carotenoids and Steroids*, edited by D. Barton & K. Nakanishi, pp. 45–67. Amsterdam: Elsevier.

Sacchettini, J. C. & Poulter, C. D. (1997). *Science*, **277**, 1788–1789.

Thomson, J. D., Higgins, D. G. & Gibson, T. J. (1994). *Nucleic Acids. Res.* **22**, 4673–4680.