Structure and reactivity in the non-mevalonate pathway of isoprenoid biosynthesis

W.N. Hunter, C.S. Bond, M. Gabrielsen and L.E. Kemp
Division of Biological Chemistry and Molecular Microbiology, School of Life Sciences, University of Dundee, Dundee DD1 5EH, U.K.

Abstract
The function, structure and mechanism of two Escherichia coli enzymes involved in the non-mevalonate route of isoprenoid biosynthesis, 2C-methyl-D-erythritol 4-phosphate cytidylyltransferase and 2C-methyl-D-erythritol 2,4-cyclodiphosphate synthase, are reviewed. Comparisons of each with enzymes from microbial pathogens highlight important conservation of sequence suggestive of similarities in secondary structure, subunit folds, quaternary structure and active sites. Since both enzymes are validated drug targets, the models provide templates for structure-based design of anti-microbial agents targeting a number of serious human diseases.

Introduction
The isomers isopentenyl pyrophosphate (IPP) and dimethylallyl pyrophosphate (DMAPP) are the universal five-carbon precursors of isoprenoids, an important, diverse and large family of natural products which includes sterols, dolichols, triterpenes, ubiquinones and plastoquinone, and also components of macromolecules such as the prenyl groups of prenylated proteins and isopentenylated tRNAs [1,2]. Isoprenoids contribute to many biological functions including electron transport in respiration and photosynthesis, hormone-based signalling, the regulation of transcription and post-translational processes that control lipid biosynthesis, meiosis, apoptosis, protein cleavage and degradation. In addition, isoprenoids constitute a structural component of cell membranes.

Isoprenoid biosynthesis depends first on the synthesis of IPP, and in mammals, higher plants, fungi and some bacteria this is carried out in the mevalonate pathway [3–5]. This begins with the conversion of acetyl-CoA into 3-hydroxy-3-methylglutaryl-CoA followed by reduction, phosphorylation and decarboxylation to generate IPP, some of which is then isomerized to DMAPP. More recently a non-mevalonate-dependent route, the 1-deoxy-D-xylulose 5-phosphate (DOXP) or 2C-methyl-D-erythritol 4-phosphate (MEP) pathway (Scheme 1), was discovered in chloroplasts, algae, cyanobacteria, eubacteria and apicomplexa [2,5–8].

The pathway starts with the condensation of pyruvate and glyceraldehyde 3-phosphate to produce DOXP, which is then converted into MEP in reactions catalysed by DOXP synthase [9,10] and DOXP reductoisomerase [11–13] respectively. MEP is linked with CTP to produce 4-diphosphocytidyl-2C-methyl-D-erythritol (CDP-ME) and pyrophosphate in a reaction catalysed by MEP cytidylyltransferase [14,15]. An ATP-dependent 4-(cytidine 5’-diphospho)-2C-methyl-erythritol kinase [16] phosphorylates CDP-ME, producing 4-diphosphocytidyl-2C-methyl-erythritol 2-phosphate (CDP-ME2P). In the fifth stage CDP-ME2P is converted to 2C-methyl-D-erythritol-2,4-cyclodiphosphate (MECP) and CMP by MECP synthase [7,17–19]. The cyclic diphosphate product MECP is reduced to 1-hydroxy-2-methyl-2-(E)-butenyl 4-diphosphate by a reductase encoded by the ispG (formally gpe) gene in Escherichia coli. Next, the ispH gene (or lytB) gene product converts the butenyl diphosphate to IPP and also DMAPP [20], so seemingly duplicating IPP isomerase activity.

The enzymes of the DOXP pathway are attractive targets for structure-based anti-microbial drug design since they occur in many of the world’s most serious pathogenic micro-organisms, including Mycobacterium tuberculosis and Plasmodium falciparum, the causal agents for tuberculosis and malaria respectively [5,7,21–23]. The enzymes are absent from humans and, in addition, the three enzymes that catalyse the production of MECP from MEP and CTP have been validated as therapeutic targets (i.e. shown to be essential) by genetic approaches [6,23]. Such genetic validation is complemented by the chemical validation of DOXP reductoisomerase using the proven anti-microbial compound fosmidomycin [13,22].

We are investigating enzymes in the DOXP pathway and now review progress in understanding the structure and reactivity of the recombinant E. coli enzymes MEP cytidylyltransferase and MECP synthase.

Results and discussion

MEP cytidylyltransferase
Richard et al. [15] published a comprehensive study of the enzyme structure and mechanism based on highly ordered monoclinic crystals solved by the heavy-atom method.
Scheme 1 | Schematic plan of the DOXP/MEP pathway of IPP/DMAPP biosynthesis

DOXPS, DOXP synthase; DOXR, DOXP reductoisomerase; MCT, MEP cytidylyltransferase; CMK, 4-(cytidine 5′-diphospho)-2C-methylerythritol kinase; MECPS, MECP synthase.
Molecular replacement was used to solve our tetragonal structure [24] and we extended the analysis to consider the sequences of the enzyme from many different bacterial pathogens.

MEP cytidylyltransferase is a homodimer with a subunit of 236 amino acids and molecular mass of 25.7 kDa [14]. The subunit is a single α/β domain, with similarities to the Rossmann fold, into which is inserted an extended ‘β-arm’ (Figure 1). Two arms associate, providing the major contribution to the dimer interface, with a lesser contribution from side-chain interactions between residues on the C-terminal helix. The active site is formed at the dimer interface by six segments of polypeptide from one subunit and one from the partner subunit.

The dominating feature of the enzyme active site is the preponderance of basic side chains involved in binding and processing substrates. In particular there are four basic residues that are major contributors to the enzyme mechanism [15] and are strictly conserved (see below). Arg-20 and Lys-27 bind CTP phosphates, assist polarization for nucleophilic attack and also stabilize the negatively charged pentaco-ordinate transition state that develops during catalysis. Arg-157 and Lys-213 serve to bind and position the attacking nucleophile, MEP.

Sequence-based comparisons

The alignment of *E. coli* MEP cytidylyltransferase with that from numerous bacterial species was carried out. 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. About 120 amino acids are highly conserved, and 22 are strictly conserved. Sequence conservation extends throughout the polypeptide and is particularly well maintained in elements of secondary structure, at the dimer interface and in the catalytic centre.

The recognition and interactions of *E. coli* MEP cytidyltransferase with substrate and/or product involves 22 residues [15], of which 18 participate in hydrogen-bonding interactions. The 10 residues that use only main-chain functional groups are Pro-13 through to Gly-18, Gly-82, Asp-83, Ala-107 and Ala-163. Although highly conserved, there is some leeway in the residue types that can be accommodated.
Residues that use side chains are more highly conserved. In addition to the basic residues already described (Arg-20, Lys-27, Arg-157 and Lys-213), Asp-106, Arg-109 and Thr-140 are strictly conserved. The remaining residue to hydrogen bond with substrate is Ser-88, which is well conserved, although occasionally being replaced by threonine.

**MECP synthase**

We determined this structure to 1.8 Å resolution [25]. Lower-resolution structures have been reported [26,27] and although there is good overall agreement it is our high-resolution structure that best defines the details in the active site and the complexes determined by Steinbacher and colleagues [26] which explain aspects of specificity towards the MEP moiety.

The MECP synthase subunit is a single α/β domain comprising 156 residues, of which approx. 36% are in α-helices and 35% in β-strands (Figure 2). The enzyme is a compact homotrimer shaped like an extended trigonal prism, of approx. 45 Å (axial) and 56 Å (equatorial) in dimensions, with a surface area equivalent to that of one subunit buried on oligomerization. The trimer is formed using a single type of subunit–subunit interface with almost all inter-subunit hydrogen bonds involving side-chain atoms.

**The active site and interactions with CDP and two metal ions**

The homotrimer carries three active sites located in a cleft lined by residues from two subunits (Figure 2). The cleft is formed by the C-terminal section of β1, the turn leading into and the N-terminal region of α1 together with the short α2 of one subunit. These segments bind Zn^{2+} in the active site, the ribose and the diphosphate of CDP, whereas the base mainly interacts with the N-terminus of α4, and the C-terminus of β5 from the partner subunit.

The cytosine is positioned in an aliphatic pocket and forms four hydrogen bonds with main-chain atoms. The base amine N-4 donates hydrogen bonds to the carbonyl groups of cis-Pro-103′ and Ala-100′, whereas N-3 and O-2 are acceptors for such interactions with the amides of Met-105′ and Leu-106′ respectively (primes identify residues from the partner subunit).

The ribose hydroxyl groups hydrogen bond with the Asp-56 carboxylate and amide of Gly-58 and, in addition,
solvent-mediated interactions with the Asp-46 side chain and the carbonyl of Ala-131 are observed. One α-phosphate phosphoryl oxygen interacts with Thr-133 by accepting two hydrogen bonds from the amide and hydroxyl groups, while the other free α-phosphate oxygen atom co-ordinates Mn²⁺. The β-phosphate provides oxygen ligands for both Zn²⁺ and Mn²⁺ and a solvent-mediated interaction with the Thr-132 hydroxyl.

The ME2P binding site is a cavity, created by the two loops that lead into α1 and α3, with α3 providing the floor. In all of the sequences available for MECP synthase, the pentapeptide loop starting at His-34 in the E. coli enzyme, at one side of the active site, is highly conserved and is key to substrate binding.

Our high-resolution structure reveals the basis of the dependence of enzyme activity on the presence of bivalent cations and loss of activity when EDTA is present [17–19]. The active site contains two metal ions 5.7 Å apart, which adopt different co-ordination geometries. Atomic absorption spectroscopy, anomalous dispersion measurements near the Zn²⁺ K-edge and refinement of complexes crystallized in the absence of Mn²⁺ confirmed that the ion with tetrahedral geometry is Zn²⁺. The co-ordination sphere involves interactions with two histidines (His-10, His-42), Asp-8 and a CDP β-phosphate oxygen. The amino acid ligands are positioned in well-defined elements of secondary structure and presumably hold the metal tightly so that it can interact with and anchor the substrate against one side of the active site. The other ion is Mn²⁺, which adopts an octahedral environment using oxygen ligands contributed from three water molecules, both CDP α- and β-phosphate groups, and Glu-135 from a partner subunit. The Mn²⁺ site corresponds to the Mg²⁺ site identified by Steinbacher et al. [26]. We used the heavier Mn²⁺ in the crystallization conditions to ensure success in metal-ion location.

Both metal ions act as Lewis acids to polarize and also to precisely align the substrate α- and β-phosphates. An associative in-line mechanism is most likely to occur with, as the first step, nucleophilic attack by the ME2P phosphate of CDP-ME2P on the β-phosphate. This would generate a pentaco-ordinate transition state, also stabilized by metal-ion co-ordination, which in the second stage then collapses to release the CMP and cyclodiphosphate products. The enzyme requires both bivalent cations for catalysis to occur, one of which is always present in the active site (i.e. Zn²⁺); the other is presumably brought in with the substrate.

**Functional homologues**

Comparison of the E. coli enzyme with 33 MECP synthase sequences showed a range of 84% (Shigella flexneri) down to 35% (catalytic domain of the P. falciparum homologue sequence identity) [18]. The amino acid conservation extends throughout the sequence, but is particularly well maintained in elements of secondary structure and in regions of the protein involved in trimer assembly, and extends to key residues shown to interact with substrate and the metal-ion cofactors [25]. The recognition and interactions of E. coli MECP synthase with CDP and metal ions involves 22 residues of which 20 are highly conserved in MECP synthase sequences. The remaining two residues are Lys-59 and Leu-106 and in both cases they use main-chain amides to donate a hydrogen bond to the active site, so the nature of the side chain is of reduced significance.

**Concluding remarks**

Our structure/sequence analyses of E. coli MEP cytidylyltransferase and MEPC synthase suggest that they are representative models for these essential enzymes from a range of bacterial and apicomplexan species. A clear picture of enzyme mechanism and specificity has been obtained, and the structures provide templates for a rational structure-based approach to aid the search for inhibitors that provide leads for the development of broad-spectrum anti-microbials.

---

**References**


©2003 Biochemical Society

Received 16 December 2002