Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

Structure of 2*C*-methyl-D-erythritol-2,4-cyclodiphosphate synthase from *Shewanella oneidensis* at 1.6 Å: identification of farnesyl pyrophosphate trapped in a hydrophobic cavity

Shuisong Ni,[a] Howard Robinson,[b] Gregory C. Marsing,[c] Dirksen E. Bussiere[d] and Michael A. Kennedy[a]*

[a]Pacific Northwest National Laboratory, Biological Sciences Division, Richland, WA 99352, USA, [b]Brookhaven National Laboratory, Biology Department, Building 463, Upton, NY 11973-5000, USA, [c]Brigham Young University, A-153 ASB, Provo, UT 84602, USA, and [d]Chiron Corporation, 4560 Horton Street, Emeryville, CA 94608, USA

Correspondence e-mail: ma_kennedy@pnl.gov

Received 1 April 2004
Accepted 3 September 2004

PDB Reference: 2C-methyl-D-erythritol-2,4-cyclodiphosphate synthase, 1t0a, r1t0asf.

Isopentenyl pyrophosphate (IPP) is a universal building block for the ubiquitous isoprenoids that are essential to all organisms. The enzymes of the non-mevalonate pathway for IPP synthesis, which is unique to many pathogenic bacteria, have recently been explored as targets for antibiotic development. Several crystal structures of 2*C*-methyl-D-erythritol-2,4-cyclophosphate (MECDP) synthase, the fifth of seven enzymes involved in the non-mevalonate pathway for synthesis of IPP, have been reported; however, the composition of metal ions in the active site and the presence of a hydrophobic cavity along the non-crystallographic threefold symmetry axis has varied between the reported structures. Here, the structure of MECDP from *Shewanella oneidensis* MR1 (SO3437) was determined to 1.6 Å resolution in the absence of substrate. The presence of a zinc ion in the active-site cleft, tetrahedrally coordinated by two histidine side chains, an aspartic acid side chain and an ambiguous fourth ligand, was confirmed by zinc anomalous diffraction. Based on analysis of anomalous diffraction data and typical metal-to-ligand bond lengths, it was concluded that an octahedral sodium ion was 3.94 Å from the zinc ion. A hydrophobic cavity was observed along the threefold non-crystallographic symmetry axis, filled by a well defined non-protein electron density that could be modeled as farnesyl pyrophosphate (FPP), a downstream product of IPP, suggesting a possible feedback mechanism for enzyme regulation. The high-resolution data clarified the FPP-binding mode compared with previously reported structures. Multiple sequence alignment indicated that the residues critical to the formation of the hydrophobic cavity and for coordinating the pyrophosphate group of FPP are present in the majority of MEDCP synthase enzymes, supporting the idea of a specialized biological function related to FPP binding in a subfamily of MEDCP synthase homologs.

1. Introduction

Isoprenoids, which are essential to all organisms, are the oldest known biomolecules, with hopanoids (membrane-associated triterpenoid derivatives) having been recovered from sediments dated to be at least 2.5 billion years old (Summons *et al.*, 1999; Brocks *et al.*, 1999). Isoprenoids also represent the largest group of natural products, numbering over 30 000 (Buckingham, 1998). All isoprenoids are constructed from two five-carbon building blocks, isopentenyl pyrophosphate (IPP) and its isomer dimethylallyl pyrophosphate (DMAPP). Genomic analysis indicates that isoprenoid biosynthesis has evolved by two ancient and distinct pathways (Lange *et al.*, 2000). Bloch & Lynen originally characterized the early steps in isoprenoid biosynthesis in 1958 (Chaykin *et al.*, 1958; Lynen

et al., 1958) using cell-free homogenates obtained from rat liver and yeast cells. For nearly 40 y, it was thought that the route for IPP biosynthesis found in mammals and yeast, which starts with acetyl-CoA and proceeds through a mevalonic acid intermediate, was universal to all organisms (Spurgeon & Porter, 1981). Later, it was discovered that eubacterial hopanoids and plastid-associated isoprenoids of algae and higher plants were derived through a mevalonic acid-independent pathway, utilizing IPP derived from the condensation of pyruvate and glyceraldehyde-3-phosphate *via* a 1-deoxy-xylulose-5-phosphate intermediate (Rohmer *et al.*, 1993; Schwender *et al.*, 1996). Many human pathogens belong to the group of microorganisms that utilize the non-mevalonate pathway for isoprenoid synthesis. Consequently, enzymes in the non-mevalonate pathway for isoprenoid synthesis have gained considerable attention because of their potential as targets for antibiotic development (Jomaa *et al.*, 1999; Ridley, 1999). Indeed, this pathway has been validated as a target, as inhibition of 2*C*-methyl-D-erythritol-4-phosphate synthase by fosmidomycin has been shown to suppress the *in vitro* growth of multidrug-resistant *Plasmodium falciparum* strains and can cure mice of the rodent malaria parasite *P. vinckei* (Vial, 2000; Jomaa *et al.*, 1999). The potential efficacy of disrupting the non-mevalonate pathway for IPP synthesis in pathogenic bacteria comes from the role IPP plays in the biosynthesis of essential molecules such as dolichols, which are carbohydrate carriers in cell-wall biosynthesis, and the respiratory quinones ubiquinone and menaquinone (Sherman *et al.*, 1989; Campbell & Brown, 2002).

2*C*-Methyl-D-erythritol-2,4-cyclophosphate (MECDP) synthase is the fifth enzyme in the seven-enzyme non-mevalonate pathway for synthesis of IPP and is responsible for the cyclization of 4-diphosphocytidyl-2*C*-methyl-D-erythritol to 2*C*-methyl-D-erythritol-2,4-cyclodiphosphate. Four crystal structures of MECDP synthase have been reported at resolutions varying from 1.6 to 2.8 Å, either in the presence or the absence of substrate, from *Escherichia coli* (Richard *et al.*, 2002; Kemp *et al.*, 2002; Steinbacher *et al.*, 2002) or from *Thermus thermophilus* (Kishida *et al.*, 2003). In these structures, the protein always crystallizes as a bell-shaped homotrimer with either a crystallographic or a non-crystallographic threefold symmetry axis and with an active site formed in a cleft between adjacent monomers. While the overall shape of the proteins is highly similar in these structures, each of the structures is reported to have different combinations of metal ions in the active site, including a Zn^{2+} ion only (Steinbacher *et al.*, 2002), an Mn^{2+} ion only (Richard *et al.*, 2002), Zn^{2+} and Mn^{2+} ions (Kemp *et al.*, 2002) or two Mg^{2+} ions (Kishida *et al.*, 2003). Two of the structures contain a hydrophobic cavity along the threefold symmetry axis that is capped by a cluster of three arginine side chains (one from each monomer) at one end and a cluster of three glutamic acid side chains (one from each monomer) at the other side. In a 1.8 Å resolution structure, Kemp *et al.* (2002) reported a sulfate ion coordinated to the arginine cap and solvent trapped in the hydrophobic cavity. In a 2.8 Å resolution structure, Richard *et al.* (2002) concluded that geranyl pyrophosphate, GPP, was most likely to be trapped by the arginine cap and hydrophobic cavity (Richard *et al.*, 2002). However, the low resolution of the data together with the presence of the crystallographic symmetry axis prohibited a definitive analysis of the identity and mode of binding of the bound molecule. Kishida *et al.* (2003) reported that no cavity existed in a 1.6 Å structure of MECDP from *T. thermophilus*, presumably owing to tighter packing of the protein from the thermophilic organism. Steinbacher *et al.* (2002) did not describe a hydrophobic cavity in a lower resolution (2.5–3.2 Å) structure of the *E. coli* protein.

Here, we report a high-resolution (1.6 Å) structure of MECDP synthase from *Shewanella oneidensis* (SO3437, which refers to its gene locus) in the absence of substrate in the active site. After screening several species, MEDCP from *Shewanella* was found to produce the highest quality crystals in the absence of substrate. The *Shewanella* MEDCP synthase exists as a 51.9 kDa homotrimer of the 17.3 kDa monomer. The data confirm the presence of tetrahedral Zn^{2+} in one of the metal-binding sites and indicate an octahedral sodium ion in the second metal site in the absence of substrate. Farnesyl pyrophosphate (FPP), a downstream product of IPP, was detected in a hydrophobic cavity along the non-crystallographic threefold symmetry axis of the homotrimer, possibly indicating a feedback-regulatory role for FPP binding to MEDCP synthase. The high-resolution structure clarifies the binding mode of the pyrophosphate of FPP in the arginine cluster that caps the hydrophobic cavity.

2. Materials and methods

2.1. Cloning, expression and purification

The *SO3437* gene was amplified from *S. oneidensis* MR-1 genomic DNA (purchased from American Type Culture Collection) using a standard PCR protocol. The PCR primer sequences are 5′-ATCGATCG**CATATG**AAAATCCGAATAGGCCATGG and 5′-ATCGA**CTCGAG**CTATTGGCGGCTGAGTAATACGAC, where the sequences in bold indicate the recognition sites for restriction endonucleases *Nde*I and *Xho*I, respectively. The resulting PCR product was digested with *Nde*I and *Xho*I and cloned into the expression vector pET28b (Novagen).

The plasmid containing *SO3437* was transformed into the expression host *E. coli* Rosetta BL21(DE3) pLysS (Novagen) and the cells were grown at 310 K with vigorous shaking to an OD_{600} of ∼0.8 in 1 l of the auto-induction medium ZYP-5052 (Studier, personal communication) supplemented with 30 µg ml^{-1} kanomycin and chloramphenicol. Protein expression was spontaneously induced at 301 K overnight. Cells were then harvested and stored at 193 K. For selenomethionine (SeMet) labeling, cells of *E. coli* B834 (DE3) pLys S transformed with the pET28b plasmid containing *SO3437* were grown in 2 l of defined autoinduction medium PASM-5052 (Studier, personal communication) supplemented with 50 µg ml^{-1} SeMet and other common amino acids. All other

growth conditions for expressing the SeMet protein were similar to those for the native SO3437.

Thawed cells were resuspended in 25 ml of a buffered lysis solution (0.5 *M* NaCl, 20 m*M* Tris pH 7.8). Phenylmethylsulfonyl fluoride was added to the cell suspension to a final concentration of about 0.2 µ*M* immediately prior to cell lysis by three passes through a French Press (SLM Instruments Inc.). The cell lysate was spun at 24 000*g* for 30 min. The supernatant was loaded onto a 10 ml cobalt-affinity column (BD Biosciences) and washed stepwise with 50 ml of imidazole solutions (0.5 *M* NaCl, 20 m*M* Tris pH 7.8) of increasing concentrations (10, 50, 250 and 400 m*M*). Each fraction was analyzed on SDS–PAGE gel. SO3437 protein was eluted from the cobalt column in several fractions of about 10 ml collected between 50 and 250 m*M* imidazole. Fractions with highly purified SO3437 were combined and dialyzed at 277 K overnight against a solution containing 0.5 *M* NaCl, 10 m*M* Tris–HCl pH 8.4. The dialysed protein was digested with thrombin (2 U per milligram of protein) at 293 K for 48 h and the cleaved N-terminal His_6 tag was removed by passing the digested protein solution through another 5 ml cobalt-affinity column. The unbound fraction, which contained SO3437 without the His tag, was concentrated with an Amicon Centriprep-3 and further purified on a Pharmacia Superdex75 HiLoad size-exclusion column (SEC) equilibrated with a solution containing 0.5 *M* NaCl, 20 m*M* Tris pH 7.8. SO3437 was eluted from the SEC column in a single peak with an estimated molecular weight equivalent to a trimer. The SeMet SO3437 was prepared under the same conditions as the native protein. Standard inductively coupled plasma resonance/mass spectrometry (ICP–MS) analysis using an Agilent Technologies 4500 instrument indicated approximately stoichiometric quantities of zinc and cobalt to be present in the protein and small non-stoichiometric quantities of manganese and nickel. The presence of zinc and cobalt in protein stock solutions was confirmed by energy-resolved fluorescence data (not shown) collected at the Advanced Photon Source at Argonne National Laboratory.

2.2. Crystallization conditions

Crystallization screening of SO3437 was carried out at approximately 295 K using the hanging-drop method. Drops were set up by mixing 2 µl SO3437 solution (37.6 mg ml^{-1}) with 2 µl of each reservoir solution from the Hampton Research Crystal Screen kits. Crystal clusters appeared in a solution containing 4 *M* sodium formate after 2–3 d. Crystallization was optimized by varying the protein concentration and by adding glycerol at various concentrations. Optimized crystals were obtained at a protein concentration of 35 mg ml^{-1} in a solution containing 4 *M* sodium formate and 5% glycerol. Those crystals grew to average dimensions of 0.4 × 0.3 × 0.3 mm within a week. SeMet SO3437 was crystallized under the same conditions as the native protein.

2.3. X-ray diffraction data collection and analysis

SO3437 crystals were initially screened and characterized using an in-house Bruker AXS Proteum 6000 X-ray diffractometer equipped with a Cu rotating-anode X-ray source (FR-591). A full native data set for SO3437 was collected to 2.0 Å resolution with the Proteum 6000 at a distance of 80 mm under a steady stream of nitrogen gas (100 K) with an exposure time of 30 s and an oscillation angle of 0.3° per image. Images were indexed, integrated and scaled using the *Bruker Proteum* software (Bruker Advanced X-Ray Solutions, Madision, WI, USA). For SeMet-labeled crystals, data were collected at the National Synchrotron Light Source (NSLS) at Brookhaven National Laboratory. Anomalous data were collected at the Se peak at the NSLS/Biology beamline x12c with a Brandeis B4 CCD detector at 140 mm. The 360 images were oscillated by 0.5° during the 20 s exposures. The images were integrated and scaled with *HKL*2000 (Otwinowski & Minor, 1997). The program *SOLVE* (Terwilliger & Berendzen, 1999; http://www.solve.lanl.gov) was used to produce a SAD map at 1.6 Å from the Se anomalous peak set. Two additional anomalous data sets were collected at the Zn peak (9660 eV) and 30 eV below the Zn peak to identify the location of the Zn within the SAD map.

2.4. Phasing using molecular replacement and structure refinement

The *CCP*4 suite of programs (Collaborative Computational Project, Number 4, 1994) was used for all aspects of structure determination and refinement. Initially, the 2.0 Å resolution native data collected in-house were scaled using *SAINT Plus for Macromolecules* (Bruker Advanced X-Ray Solutions, Madision, WI, USA) and used for molecular replacement. The 1.8 Å structure of MEDCP from *E. coli* reported by Kemp *et al.* (2002) (PDB code 1gx1) was used as a search model (68% identity). All non-identical residues in the search model were replaced by alanine prior to searching for a solution using *MOLREP* (Vagin & Teplyakov, 1997; Collaborative Computational Project, Number 4, 1994). Once a solution was found, the phases and structure were refined iteratively by mutating mismatched residues and placing the side chains into the observed density using torsion-angle rotations using the *O* software package (Jones *et al.*, 1990). Structure refinement, including refinement of water positions, was conducted using *REFMAC*5 (Murshudov *et al.*, 1997; Collaborative Computational Project, Number 4, 1994). Early in the refinement, threefold non-crystallographic symmetry (NCS) averaging was applied and density modification was carried out using *DM* (Cowtan, 1994; Collaborative Computational Project, Number 4, 1994) to improve the quality of the maps. Once the 1.6 Å resolution SeMet data became available from the synchrotron, we continued refinement of the maps using the higher resolution data. Experimental structure factors were combined with the phases from the molecular-replacement solution using *SFALL* (Agarwal, 1978; Collaborative Computational Project, Number 4, 1994), followed by calculations of σ_A-weighted maps using *FFT* (Ten Eyck, 1973; Read

Table 1
Summary of diffraction data and structure-refinement statistics.

The space group was $P4_32_12$, with unit-cell parameters $a = b = 116.9$, $c = 109.1$ Å. There is one trimer per asymmetric unit. Values in parentheses are for the highest resolution shell. R_{sym} is equivalent to R_{merge} used in the *HKL*2000 software package as defined by Weiss (2001).

	Native	SeMet
Data-set statistics		
Mosaicity (°)	0.45	0.25
Resolution (Å)	65.8–2.00 (2.10–2.00)	30.0–1.60 (1.66–1.60)
Wavelength (Å)	1.54184	0.97911
No. observed reflections	754360	1445453
No. unique reflections	50872 (6615)	99854 (9809)
Signal-to-noise ratio $I/\sigma(I)$	35.2 (2.9)	55.7 (5.9)
Completeness (%)	98.4 (95.1)	100.0 (100.0)
Redundancy	16.4 (5.7)	14.5 (13.1)
R_{sym}(%)	56.1 (29.3)	8.1 (50.8)
Refinement statistics		
R value/R_{free} (%)	17.2/18.8	
Figure of merit (%)	89.5	
No. amino acids	465	
No. water molecules	611	
Average B factors (Å^2)		
Protein	16.7	
Water molecules	30.4	
R.m.s. deviations from ideal values		
Bond distances (Å)	0.010	
Bond angles (°)	1.368	
Chiral angles (°)	0.086	
Ramachandran plot: non-Gly residues in		
Most favored regions (%)	95.5	
Additionally allowed regions (%)	4.5	
Generously allowed regions (%)	0.0	
Disallowed regions (%)	0.0	

& Schierbeek, 1988; Collaborative Computational Project, Number 4, 1994). $F_o - F_c$ and $2F_o - F_c$ difference maps were generated using *REFMAC*5 (Murshudov *et al.*, 1997; Collaborative Computational Project, Number 4, 1994). All CCP4-format maps were converted to Dsn6 format using *MAPMAN* (Kleywegt & Jones, 1996).

3. Results

3.1. Data collection and analysis

The native SO3437 crystals typically diffracted to about 2.0 Å resolution on the in-house diffractometer. Diffraction data statistics are summarized in Table 1. Indexing of the diffraction data showed that the crystals were tetragonal. The space group was determined to be $P4_32_12$, with unit-cell parameters $a = b = 116.8$, $c = 108.7$ Å, $\alpha = \beta = \gamma = 90°$. The crystallographic asymmetric unit contained three molecules of SO3437, with a corresponding Matthews coefficient (Matthews, 1968) of 3.6 Å^3 Da^{-1} and a solvent content of 65.9%. SeMet crystals had the same crystal system and unit-cell parameters as the native crystals. Anomalous Fourier maps above and below the Zn edge were produced by merging the anomalous difference amplitudes with the phases from the SAD map with the program *CCP*4 (Collaborative Computational Project, Number 4, 1994). Clearly indicated in these maps were not only the locations of the three Zn atoms but also the locations of all the Met Se atoms, Cys S atoms and P atoms from the phosphate groups of the farnesyl pyrophosphate (see below). No other anomalous signals were observed above the background noise level.

3.2. Molecular replacement and structure refinement

Using *MOLREP* (Vagin & Teplyakov, 1997; Collaborative Computational Project, Number 4, 1994) and PDB file 1gx1 from Kemp *et al.* (2002), a search for three rotation peaks was conducted based on the Matthews coefficient (Matthews, 1968), which indicated three molecules in the asymmetric unit. Three peaks were found with a signal-to-noise ratio ranging from 12.5 to 13.5. The subsequent translational search for the three strongest peaks in the rotation function yielded correlation coefficients of the order of 0.45. NCS operators relating each of the molecules were determined with the r.m.s. *xyz* displacement ranging from 0.35 to 0.55 Å between the NCS-related molecules. The structure-refinement statistics are summarized in Table 1. The working R and R_{free} values were refined to 17.2 and 18.8%, respectively. About 600 water molecules with B factors less than 60 Å^2 were refined using *REFMAC*5 (Murshudov *et al.*, 1997; Collaborative Computational Project, Number 4, 1994). The initial maps generated using the molecular-replacement phases were compared with subsequent maps generated with Se SAD and Zn MAD phasing to rule out any potential bias introduced by the molecular-replacement solution.

3.3. SO3437 structure analysis

As expected, the overall structure is similar to the other reported structures for MECDP synthase from *E. coli* and *T. thermophilus*. The enzyme consists of a homotrimer with a threefold NCS axis along the long axis of the homotrimer. The ribbon structure of the monomer is shown in Fig. 1(*a*). The secondary-structure elements were determined by a consensus of *PROCHECK* (Laskowski *et al.*, 1993; Morris *et al.*, 1992) and the Kabsch & Sander (1983) algorithm implemented in *MOLMOL* (Koradi *et al.*, 1996). The monomer contains two β-sheets and eight α-helices that form a mixed α/β-sandwich arranged as indicated in Fig. 1(*b*). The four-strand β-sheet of each monomer faces inward towards the threefold NCS axis, forming a β-barrel-like structure at the core of the homotrimer (Fig. 1*c*). The interface between monomer units creates an active-site cleft where the zinc and sodium ions are located (Figs. 1*d* and 1*e*). The histidine and aspartic acid side chains on one monomer coordinate the zinc ion, while a threonine side chain on a second monomer opposite the cleft is involved in coordinating the octahedral sodium ion, indicating a proximity that is consistent with the observation that residues from two separate monomer chains are involved in the formation of the active site. The active-site cleft is illustrated in the surface-rendered structure shown in Fig. 1(*e*). Finally, a hydrophobic cavity containing additional non-protein electron density was observed. Farnesyl pyrophosphate best fitted the extra density, as will be discussed more extensively below. Fig. 1(*f*) illustrates that the farnesyl pyrophosphate and the three active sites occupy only the upper half of the homotrimer structure.

In order to confirm the presence and location of zinc in the protein crystals, anomalous electron diffraction was searched for and observed at the zinc edge. Since anomalous electron diffraction was observed at the zinc edge, a zinc MAD data set was collected and used to generate an electron-density map that was consistent with the maps obtained from the molecular-replacement solution and the SeMet SAD solution. The electron-density map at the zinc site, shown in Fig. 2(*a*), illustrates the coordination of zinc through the His12 NE2, His44 ND1 and Asp10 OD2 atoms, with zinc-to-atom distances of 2.00, 2.05 and 1.95 Å, respectively. The tetrahedral coordination geometry is completed by a disordered ligand whose identity has not been unambiguously identified (2.64 Å) but is modeled as a water molecule, as shown in Fig. 2(*b*).

The high-resolution structure reported here in the absence of substrate enables an analysis of the metal-coordination geometry prior to substrate binding. Although ICP–MS and energy-resolved fluorescence-emission spectroscopy indicated the presence of approximately stoichiometric amounts of zinc and cobalt in the protein stock solutions, the anomalous Fourier map at the Zn edge showed no defined location that could be assigned to cobalt in the SO3437 crystals. In particular, the atomic site identified as a sodium ion, 3.94 Å from the Zn site, has no detectable anomalous signal at 9660 eV. The octahedral sodium-coordination environment involves two amino-acid side chains, including OD1 from the bridging Asp10 that coordinates the Zn atom through OD2 (2.40 Å) and a threonine side-chain hydroxyl group Thr134 OG1 atom (2.50 Å). Three water molecules at 2.38, 2.40 and 2.54 Å and an unidentified fourth ligand (modeled as water) at 2.68 Å completed the octahedral sodium-coordination environment (Fig. 2*b*). Although the anomalous Fourier data at the Zn edge for the metal site near the Zn atom could be consistent with a Mg cation (Co, Ca, Cu, Zn, Mn and Fe are all not likely owing to a lack of observed anomalous signal), the ligand distances are not characteristic of Mg, which would exhibit ligand distances averaging around 2.1 Å (Harding, 1999). Similarly, potassium is unlikely at this site as its ligand distances are usually around 2.8 Å (Harding, 2002). Consequently, the most likely chemical identity for this density is sodium, where expected ligand distances are about 2.4 Å, close to the average of 2.52 Å observed here and consistent with the lack of anomalous signal at 9660 eV from this site.

Substantial non-protein electron density was observed in the $F_o - F_c$ difference map in a hydrophobic cavity along the threefold NCS axis (Fig. 2*c*). Visual inspection indicated that pyrophosphate could explain part of the

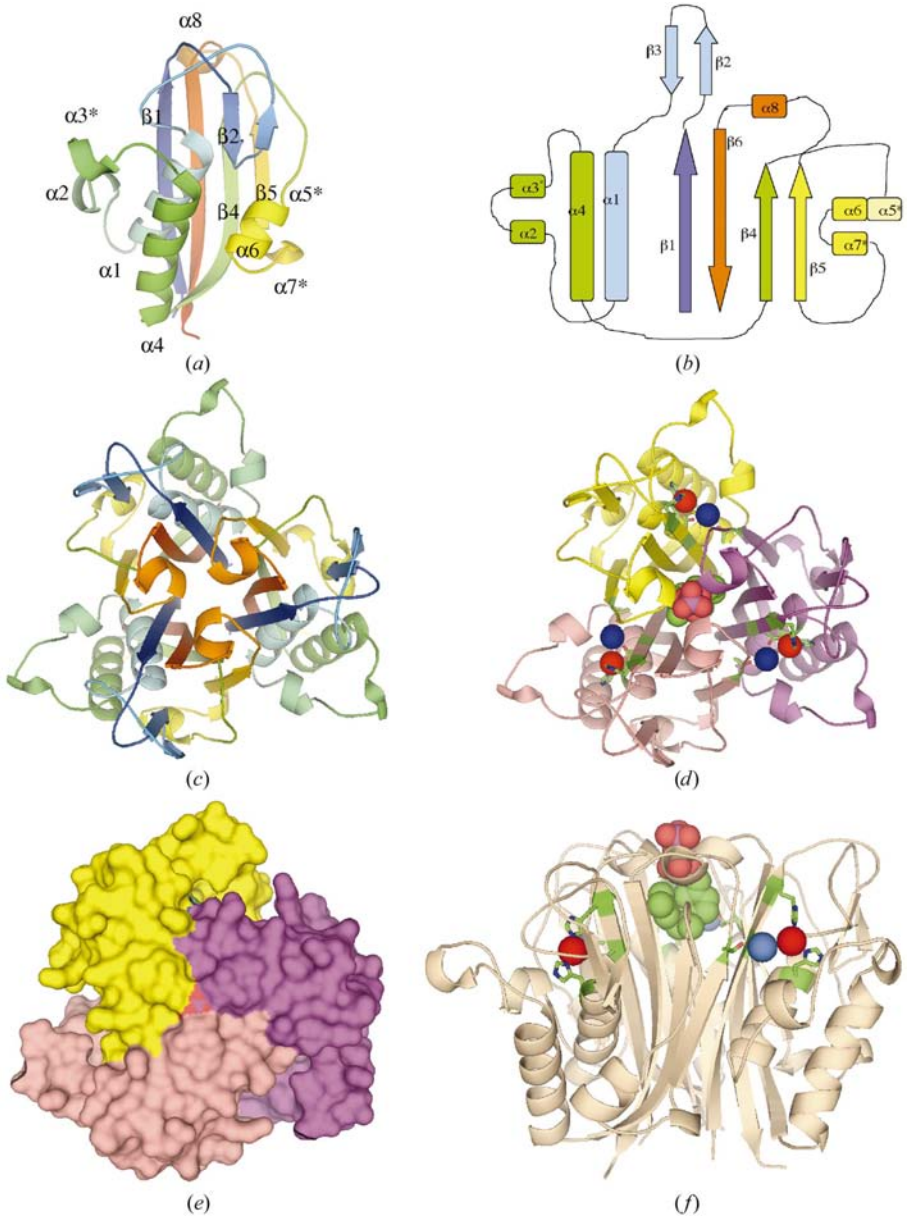

Figure 1
(*a*) *Molscript* rendering of the SO3437 monomer. (*b*) Schematic representation of the secondary-structural elements in SO3437. (*c*) *Molscript* representation of the SO3437 homotrimer structure. (*d*) *Molscript* rendering of SO3437 with the zinc (red) and sodium (blue) shown with space-filling rendering. The FPP (colored by atom type with space-filling rendering) can be seen along the threefold NCS axis. (*e*) Surface representation of the SO3437 homotrimer showing FPP at the center of the homotrimer surface. (*f*) *Molscript* rendering of the SO3437 homotrimer rotated 90° from the orientation in (*d*) with side chains involved in coordinating zinc and sodium rendered as sticks, FPP rendered with space filling and with pyrophosphate colored by atom type and the C_{15} hydrophobic chain colored green.

difference density. A survey of small molecules containing pyrophosphate groups indicated that the family of IPP derivatives best fitted the difference density. IPP, dimethylallyl pyrophosphate, geranyl pyrophosphate and FPP were all tested. In all of the molecules tested except for the FPP, the hydrophobic chain connected to the pyrophosphate group was too short to explain the difference density. FPP fits very well into the difference density, as can be seen in Fig. 2(*c*). Inspection of the structure surrounding the FPP reveals specific protein–FPP contacts. Each monomer contributes an arginine side-chain guanidinium group that forms a salt bridge with the terminal phosphate group of the FPP (Fig. 2*d*). In addition, the backbone amide of Phe141 appears to form a hydrogen bond with an O atom of the penultimate phosphate in the FPP. This bond can be observed in Fig. 3. The hydrophobic cavity occupied by FPP is formed from two phenylalanine side chains, Phe9 and Phe141, from each of the three monomers (Fig. 3). In addition, at the bottom of the hydrophobic cavity there is a cluster of three glutamic acid side chains formed from Glu151 from each of the three monomers. Each Glu151 side-chain carboxylate OE2 atom appears to form a salt bridge with a protonated His7 side-chain NE2 atom (2.89–2.95 Å) from the same monomer, balancing the high charge density in the Glu cluster. The buried salt-bridge cluster forms the floor of the hydrophobic cavity, limiting the size of the cavity and providing a structural mechanism for discriminating on the basis of size what molecules can be sequestered into the cavity. Each Glu151 carboxylate OE2 atom is also near an NE2 atom from a histidine side chain from another monomer chain, *e.g.* in the *A* chain His7 NE2 forms a salt bridge with *A* chain Glu151 OE1 at 2.94 Å and is also near the *C* chain His7 NE2 at 3.39 Å. The density at each salt bridge exhibits twofold disorder, indicating that the Glu151 carboxylate jumps between the two histidine side chains, one from the same monomer and one from the adjacent monomer.

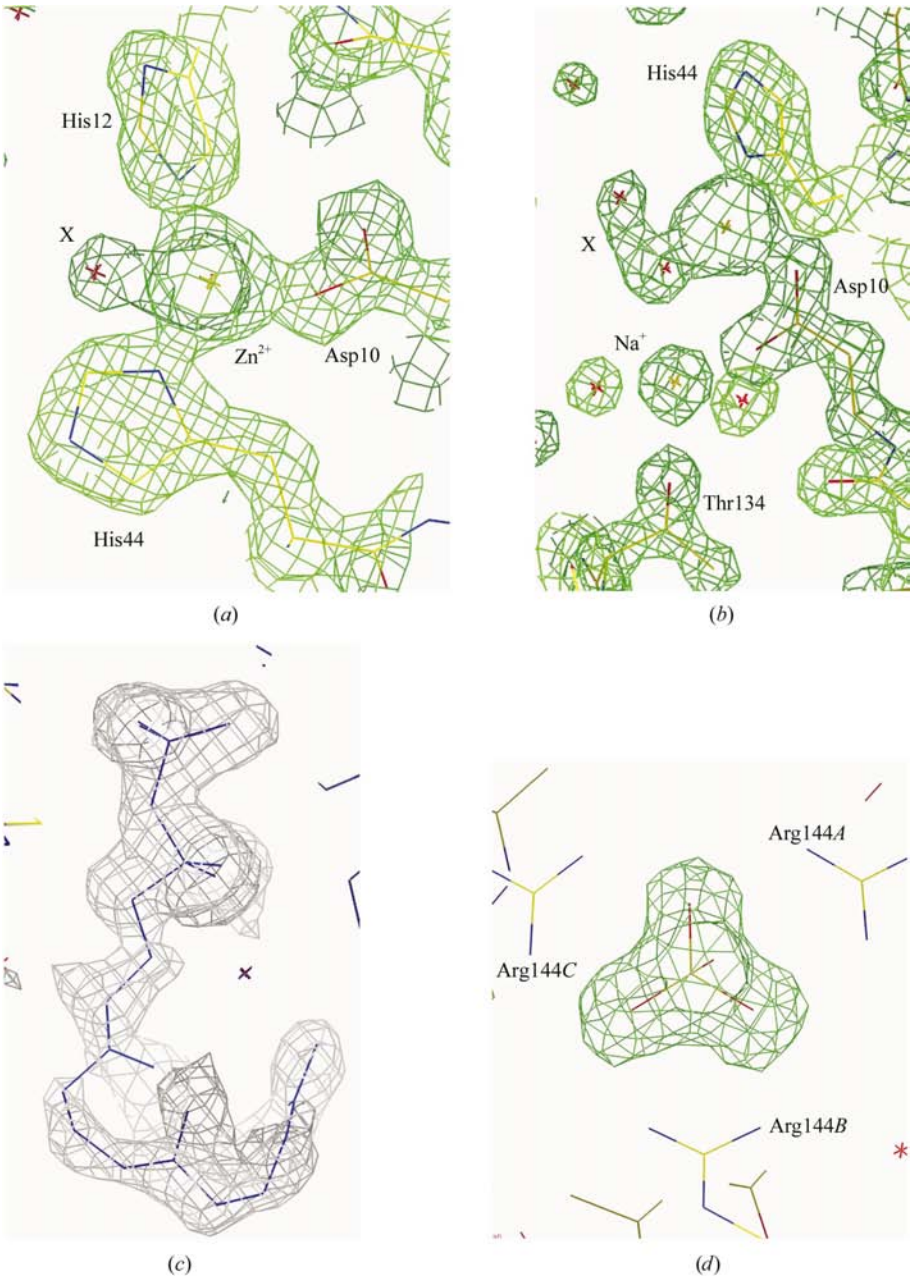

Figure 2
(*a*) Electron density in $2F_o - F_c$ map at a contour level of 2σ showing the tetrahedral zinc-coordination site. (*b*) As in (*a*) except that the orientation has been rotated to show both the zinc and the sodium sites. In this view, the unidentified electron density bridging the zinc and sodium ions is labeled X. (*c*) Electron density in the $F_o - F_c$ map at a contour level of 2σ showing the difference electron density in the hydrophobic cavity for the model in which FPP is absent. The optimized geometry of the FPP is shown fitted into the difference electron density. (*d*) The electron density in the $2F_o - F_c$ map centered on the terminal phosphate of the FPP with the radius of the electron density adjusted so that the only density displayed is for the FPP. The orientation shows how Arg144 from each monomer coordinates to each O atom of the terminal phosphate of FPP.

3.4. Conservation of the hydrophobic cavity residues in SO3437 homologs

Of the four reported SO3437 homolog crystal structures, three are of *E. coli* proteins (Kemp *et al.*, 2002; Richard *et al.*, 2002; Steinbacher *et al.*, 2002) and one is from *T. thermophilus* (Kishida *et al.*, 2003). Two of the three *E. coli* structures have been reported to contain hydrophobic cavities occupied by small molecules (Kemp *et al.*, 2002; Richard *et al.*, 2002), while a hydrophobic cavity was not discussed in the third structure (Steinbacher *et al.*,

2002). Kishida *et al.* (2003) reported that the *T. thermophilus* homolog does not contain a hydrophobic cavity. In SO3437, the residues that define the hydrophobic cavity are identified as two phenylalanine side chains, Phe9 and Phe141, contributed by each of the three monomers in the homotrimer (Fig. 3). The homotrimer architecture appears exquisitely engineered to bind the terminal and penultimate phosphate groups of the FPP through three Arg144 guanidinium side chains and three Phe141 backbone amide groups. A buried cluster of salt bridges between His7 and Glu151 from each monomer creates a floor or boundary that defines the size of the hydrophobic cavity. Given the precise identity of the residues that define the FPP-binding site, it is possible to investigate how widely the motif is conserved among SO3437 homologs. SO3437 shares sequence homology with 28 proteins identified in a cluster of orthologous group COG0245 (Natale *et al.*, 2000) (Fig. 4). Arg144 in *S. oneidensis*, which is critical for binding the terminal phosphate in FPP, is conserved in 20 of the 25 sequences compared in Fig. 4. Residues Phe9 and Phe141, which are essential to forming the hydrophobic cavity in SO3437, are conserved, similar or hydrophobic in 20 of 25 and 19 of 25 sequences, respectively. The hydrophobic residues that make up the hydrophobic cavity are always conserved when the arginine residue essential for coordinating the terminal phosphate in FPP is conserved. The cluster of salt bridges created by the pair of residues Glu151 and His7 is less conserved than the Phe9 and Phe141 residues that define the hydrophobic core and the Arg144 residues that form the pyrophosphate-binding motif, occurring in only six of the 25 sequences.

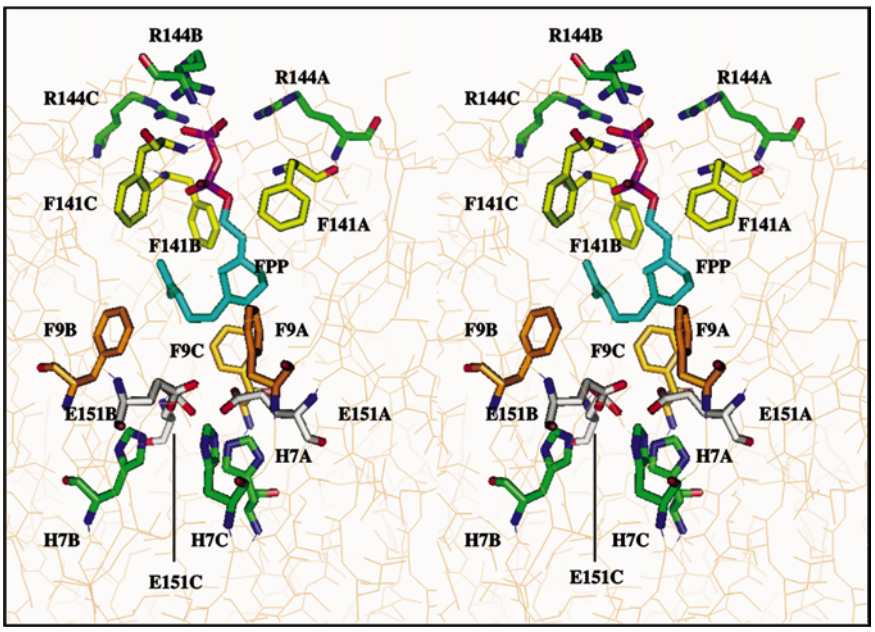

Figure 3
Stereoview of FPP (C_{15} chain colored green and phosphates colored by atom type) in the hydrophobic pocket formed by Phe141 (yellow) and Phe9 (orange) from each monomer. The Arg144 guanidinium side chain, which coordinates to the terminal phosphate of the FPP, and the Glu151 and His7 side chains, which form intra- and intermonomer salt bridges, are depicted using stick rendering.

4. Discussion

The structure of SO3437 from *S. oneidensis* reported here is the fifth experimentally determined structure of an MEDCP homolog. In the previously reported structures, four different combinations of metal ions have been observed, either in the presence or absence of substrate. MEDCP is known to be a Mg^{2+}-dependent enzyme; however, the presence of tetrahedral zinc in the active site has not been consistently observed in the other reported structures. In one case, Kishida *et al.* (2003) reported that the two metal ion sites were both Mg^{2+} based on refinement of the temperature (*B*) factors of the individual atoms. Here, multiple methods, including ICP–MS, energy-resolved fluorescence measurements and anomalous electron diffraction, were used to confirm the presence and location of tetrahedrally coordinated zinc in the SO3437 structure. ICP–MS and fluorescence measurements also indicated the presence of cobalt in the protein stock solutions used for crystallization, possibly leftover from the cobalt metal-chelate affinity column used for purification. However, careful analysis of the anomalous X-ray scattering data collected at the zinc edge indicated that cobalt was not present in the crystal. The presence of a tetrahedral zinc ion in the active site in the SO3437 structure is consistent with a naturally occurring zinc ion in native active MEDCP synthase that anchors the substrate in the active site. The second metal ion near to the tetrahedral zinc in the SO3437 structure is best explained as an octahedral sodium ion, replaced by Mg^{2+} in the active enzyme–substrate complex, that is coordinated by an Asp10 carboxylate anion as a bridging ligand that shares coordination with the zinc ion and Thr134 from the adjacent polypeptide chain. Consideration of the charge in the immediate environment of the two metal ions (Zn^{2+}, Na^{+}, COO^{-}) indicates a net charge of 2+ in the absence of an additional charged ligand. The elongated electron density bridging the sodium and zinc ions could be a disordered anion, such as carbonate, that could balance the charge between the two metal ions. It is likely that when 4-diphosphocytidyl-2*C*-methyl-D-erythritol binds in the active site the unidentified anion is displaced by the pyrophosphate of the substrate.

Two of the *E. coli* MEDCP structures were reported to have hydrophobic cavities (Richard *et al.*, 2002; Kemp *et al.*, 2002). Kemp *et al.* (2002) interpreted the extra non-protein electron density in the center of the homotrimer in a 1.8 Å resolution structure as a sulfate ion plus water trapped in the hydrophobic cavity. In a lower resolution structure (2.8 Å), Richard *et al.* (2002) suggested that the shape of the non-protein density in the hydrophobic cavity and the chemical environment of the cavity were consistent with the sequestration of a

molecule possessing a negatively charged head group and a hydrophobic tail, such as an isoprenoid pyrophosphate-like molecule. The dimensions of the cavity were such that IPP (C_5-IPP), geranyl pyrophosphate (C_{10}-GPP) or FPP (C_{15}-FPP) could be accommodated. They concluded that C_{10}-GPP best fitted the density. In the 1.6 Å SO3437 structure reported here, a well defined envelope of difference electron density could be calculated such that the length of the hydrophobic tail in IPP, GPP and FPP could be discriminately compared and it was determined that FPP best fitted the extra electron density. The architecture of the hydrophobic cavity includes an arrangement of arginine residues (Arg144) that precisely coordinate to the terminal phosphate of the FPP molecule, a hydrophobic cavity lined by the hydrophobic residues Phe9 and Phe141 from each of the three monomers and a floor made up of three intramonomer salt bridges between residues Glu151 and His7 that may serve as a boundary on the dimension of the hydrophobic cavity, perhaps enabling SO3437 to discriminate binding molecules in the hydrophobic cavity based on the size of the molecule. Sequence alignment of the MEDCP enzymes indicates that both the residues that define the hydrophobic cavity and that bind the FPP pyrophosphate are widely conserved in the protein family and always conserved together, lending further support to the suggestion that the FPP-binding motif is conserved for a biological function. The buried salt bridge that defines the floor of the hydrophobic cavity in SO3437 is less conserved than the hydrophobic cavity and FPP pyrophosphate-binding motifs, indicating a rarer and possibly specialized feature in a subgroup of the SO3437 family. The twofold disorder in the position of the Glu151 glutamate carboxylate between His7 of the same monomer and His7 from an adjacent monomer indicates a possible allosteric regulatory effect when the homotrimer binds FPP. In the *T. thermophilus* homolog that lacks a hydrophobic cavity, both the conserved hydrophobic residues Phe9 and Phe141 and the Arg144 that binds the FPP pyrophosphate in SO3437 are missing. The complex nature of the cavity, the observation that FPP, a downstream product of IPP, binds in the cavity and the wide conservation of the cavity in

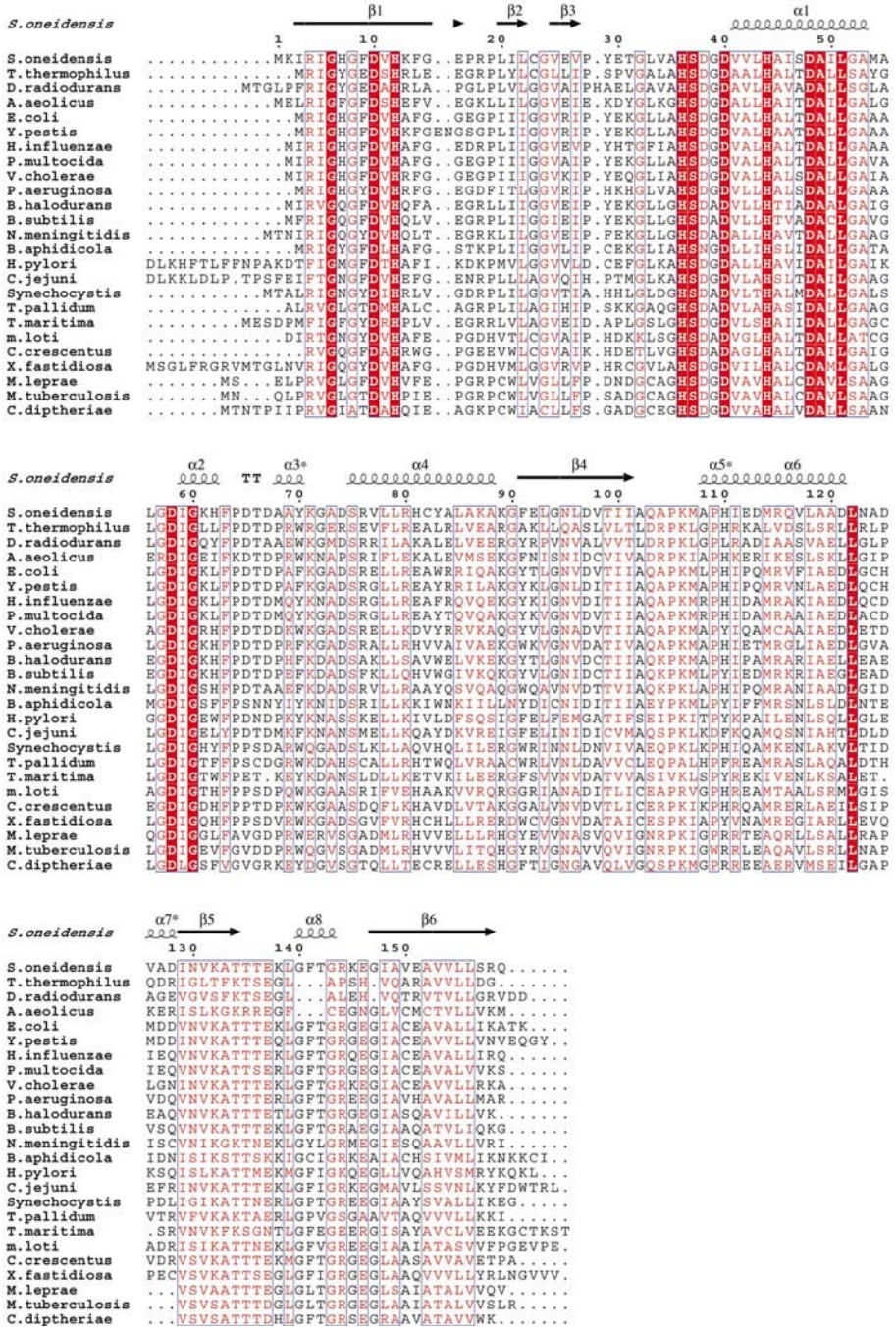

Figure 4
Sequence alignment for 25 of the 29 sequences in COG425. Full organism names and Genbank reference sequence numbers for the proteins shown in the alignment are included in parentheses: *Thermus thermophilus* (BAB86885), *Deinococcus radiodurans* (NP_293954), *Aquifex aeolicus* (NP_213652), *Escherichia coli* K12 (NP_417226), *Yersinia pestis* (AAS60600), *Haemophilus influenza* (NP_438831), *Pasteurella multocida* (NP_246548), *Vibrio cholerae* (NP_230180), *Shewanella oneidensis* (NP_718990), *Pseudomonas aeruginosa* PA01 (NP_252317), *Bacillus halodurans* (NP_240974), *B. subtilis* (NP_387972), *Neisseria meniginitidis* (NP_284427), *Buchnera aphidicola* (NP_240231), *Helicobacter pylori* (NP_207810), *Campylobacter jejuni* (NP_282735), *Synechosystis* (NP_440786), *Treponema pallidum* (NP_218953), *Thermotoga maritima* (NP_228456), *Mesorhizobium loti* (NP_102198), *Caulobacter crescentus* (NP_420546), *Xylella fastidiosa* (NP_298583), *Mycobacterium leprae* (NP_301349), *M. tuberculosis* (NP_218098), *Corynebacterium diptheria* (AF230738).

the SO3437 family indicate a possible biological function, potentially *via* a regulatory feedback mechanism, as suggested by Richard *et al.* (2002).

MEDCP has received attention as a potential drug target to disrupt the non-mevalonate pathway for isoprenoid synthesis. There are multiple approaches that can be taken in terms of drug design, including competitive binding of a drug at the active-site cleft between the two chains. As residues from two adjacent monomers bind the substrate in the active site, any means of disrupting the homotrimeric structure would be potentially effective in abolishing or altering the enzyme chemistry or efficiency. However, if a biological role, such as a feedback-regulatory mechanism, can be ascribed to a downstream product of IPP such as FPP binding in the hydrophobic cavity, drugs designed to compete for binding in the hydrophobic cavity could potentially disrupt its regulatory function. Additional enzymology will be necessary to determine which approach will be most suitable for drug discovery.

We would like to acknowledge Matt Benning of Bruker AXS for assistance in using the Bruker software, Tom Wietsma for collecting the ICP–MS at the Environmental Molecular Sciences Laboratory and Steve Heald at Argonne National Laboratory for collecting the EXAFS data at the PNC-CAT. In-house X-ray data were collected at the Environmental Molecular Sciences Laboratory (a national scientific user facility sponsored by the US Department of Energy, Office of Biological and Environmental Research) located at Pacific Northwest National Laboratory and operated by Battelle for the Department of Energy (contract KP130103). The work was supported by Battelle Corporate IR&D funding and the Department of Energy Office of Biological and Environmental Research Grant KP11-01-01:24931.

References

Agarwal, R. C. (1978). *Acta Cryst.* A**34**, 791–809.
Brocks, J. J., Logan, G. A., Buick, R. & Summons, R. E. (1999). *Science*, **285**, 1033–1036.
Buckingham, J. (1998). In *Dictionary of Natural Products on CD-ROM*, version 6.1. London: Chapman & Hall.
Campbell, T. L. & Brown, E. D. (2002). *J. Bacteriol.* **184**, 5609–5618.
Chaykin, S., Law, J., Philipis, A. H., Tchen, T. T. & Bloch, K. (1958). *Proc. Natl Acad. Sci. USA*, **44**, 998–1004.
Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* D**50**, 760–763.
Cowtan, K. (1994). *Jnt CCP4 /ESF–EACBM Newsl. Protein Crystallogr.* **31**, 34–38.
Harding, M. M. (1999). *Acta Cryst.* D**55**, 1432–1443.
Harding, M. M. (2002). *Acta Cryst.* D**58**, 872–874.
Jomaa, H., Wiesner, J., Sanderbrand, S., Altincicek, B., Weidemeyer, C., Hintz, M., Türbachova, I., Eberl, M., Zeidler, J., Lichtenthaler, H. K. & Soldati, D. (1999). *Science*, **285**, 1573–1576.
Jones, T. A., Bergdoll, M. & Kjeldgaard, M. (1990). *Crystallographic and Modeling Methods in Molecular Design*, edited by C. Bugg & S. Ealick, pp. 189–195. New York: Springer-Verlag.
Kabsch, W. & Sander, C. (1983). *Biopolymers*, **22**, 2577–2637.
Kemp, L. E., Bond, C. S. & Hunter, W. N. (2002). *Proc. Natl Acad. Sci. USA*, **99**, 6591–6596.
Kishida, H., Wada, T., Unzai, S., Kuzuyama, T., Takagi, M., Terada, T., Shirouzu, M., Yokoyama, S., Tame, J. R. & Park, S. Y. (2003). *Acta Cryst.* D**59**, 23–31.
Kleywegt, G. L. & Jones, T. A. (1996). *Acta Cryst.* D**52**, 826–828.
Koradi, R., Billeter, M. & Wüthrich, K. (1996). *J. Mol. Graph.* **14**, 29–32.
Lange, B. M., Rujan, T., Martin, W. & Croteau, R. (2000). *Proc. Natl Acad. Sci. USA*, **97**, 13172–13177.
Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. (1993). *J. Appl. Cryst.* **26**, 283–291.
Lynen, F., Eggerer, H., Henning, U. & Kessel, I. (1958). *Agnew. Chem.* **70**, 738–742.
Matthews, B. W. (1968). *J. Mol. Biol.* **33**, 491–497.
Morris, A. L., MacArthur, M. W., Hutchinson, E. G. & Thornton, J. M. (1992). *Proteins*, **12**, 345–364.
Murshudov, G. N., Vagin, A. A. & Dodson, E. J. (1997). *Acta Cryst.* D**53**, 240–255.
Natale, D. A., Shankavaram, U. T., Galperin, M. Y., Wolf, Y. I., Aravind, L. & Koonin, E. V. (2000). *Genome Biol.* **1**, Research0009.1–0009.19.
Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.
Read, R. J. & Schierbeek, A. J. (1988). *J. Appl. Cryst.* **21**, 490–495.
Richard, S. B., Ferrer, J. L., Bowman, M. E., Lillo, A. M., Tetzlaff, C. N., Cane, D. E. & Noel, J. P. (2002). *J. Biol. Chem.* **277**, 8667–8672.
Ridley, R. G. (1999). *Science*, **285**, 1503–1503.
Rohmer, M., Knani, M., Simonin, P., Sutter, B. & Sahm, H. (1993). *Biochem. J.* **295**, 517–524.
Schwender, J., Seeman, M., Lichtenthaler, H. K. & Rohmer, M. (1996). *Biochem. J.* **316**, 73–80.
Sherman, M. M., Petersen, L. A. & Poulter, C. D. (1989). *J. Bacteriol.* **171**, 3619–3628.
Spurgeon, S. L. & Porter, J. W. (1981). *Biosynthesis of Isoprenoid Compounds*, edited by J. W. Porter & S. L. Spurgeon, Vol. 1, pp. 1–46. New York: Wiley.
Steinbacher, S., Kaiser, J., Wungsintaweekul, J., Hecht, S., Eisenreich, W., Gerhardt, S., Bacher, A. & Rohdich, F. (2002). *J. Mol. Biol.* **316**, 79–88.
Summons, R. E., Jahnke, L. L., Hope, J. M. & Logan, G. A. (1999). *Nature (London)*, **400**, 554–557.
Ten Eyck, L. F. (1973). *Acta Cryst.* A**29**, 183–191.
Terwilliger, T. C. & Berendzen, J. (1999). *Acta Cryst.* D**55**, 849–861.
Vagin, A. & Teplyakov, A. (1997). *J. Appl. Cryst.* **30**, 1022–1025.
Vial, H. J. (2000). *Parasitol. Today*, **16**, 140–141.
Weiss, M. S. (2001). *J. Appl. Cryst.* **34**, 130–135.