PROTEINS: Structure, Function, and Genetics 49:135–138 (2002)

STRUCTURE NOTE

Structure of 2*C*-Methyl-D-Erythrol-2,4-Cyclodiphosphate Synthase From *Haemophilus influenzae*: Activation by Conformational Transition

Christopher Lehmann,[1] Kap Lim,[1] John Toedt,[1†] Wojciech Krajewski,[1] Andrew Howard,[3,4] Edward Eisenstein,[1,2] and Osnat Herzberg[1*]

[1]*Center for Advanced Research in Biotechnology, University of Maryland Biotechnology Institute, Rockville, Maryland*
[2]*The National Institute of Standards and Technology, Gaithersburg, Maryland*
[3]*Advanced Photon Source, Argonne National Laboratory, Argonne, Illinois*
[4]*Illinois Institute of Technology, Chicago, Illinois*

Introduction. 2*C*-methyl-D-erythrol-2,4-cyclodiphosphate synthase (YgbB, or IspF) functions in the non-mevalonate biosynthesis pathway of isoprenoids[1] unique to prokaryotes and to some plants. The Mg(II)-dependent enzyme converts 4-diphosphocytidyl-2*C*-methyl-D-erythritol 2-phosphate into the cyclic compound 2*C*-methyl-D-erythritol 2,4-cyclodiphosphate, and CMP. The crystal structure of the enzyme from *Escherichia coli* was recently reported[2] as a complex with product or substrate and in the unbound state. The structures revealed a trimeric enzyme, with a bound Zn(II) in each of the three active sites. The Zn(II) is tetrahedrally coordinated. Two histidine residues and an aspartic acid provide three ligands. The fourth ligand is a solvent molecule in the unbound enzyme. In the enzyme complexes with CDP (substrate fragment) or with 2,4-cyclodiphosphate (product), the fourth ligand is a phosphate group. We report here the crystal structure of the enzyme from *Haemophilus influenzae* (HI0671). The protein was selected as a Structural Genomics target[3] (http://www.s2f.umd.edu), focusing on proteins of unknown function, before the functional annotation of the sequence.[4] The trimeric association and overall fold of the protein is the same as that of the *E. coli* enzyme, consistent with their high amino acid sequence identity (77%). However, a large polypeptide segment adopts an entirely different conformation, such that a histidine residue, invariant in 45 of 50 known sequences (conservatively replaced by asparagine in the remaining five sequences), coordinates to Zn(II). We propose that enzyme activity may be regulated by a conformational transition of this structural unit.

Materials and Methods. The gene encoding HI0671 was amplified from *H. influenzae* KW20 genomic DNA and cloned into pET17b for expression of the native polypeptide in *E. coli* strain BL21(DE3). Cells were grown at 37°C in LB medium containing 100 mg/L ampicillin. Expression was induced at mid-log phase with 1 mM IPTG and was continued for 4 h. Cells were lysed by using a French press, and the crude cell lyaste was dialyzed against 20 mM Tris-HCl, pH 8.4. Protein purification was performed on a BioCAD 700E chromatography workstation (PerSeptive Biosystems). The extract was first applied to a Poros HQ 50 anion exchange column equilibrated with 20 mM Tris-HCl at pH 8.4 and eluted with a linear gradient from 0 to 1 M NaCl. The fractions enriched for HI0671 were pooled, dialyzed against buffer containing a blend of 20 mM MES, 20 mM HEPES, and 20 mM sodium acetate, to yield a pH of 6.5, and applied to a Poros HS 20 cation exchange column equilibrated in the same buffer. Pure protein was obtained by eluting with a linear gradient from 0 to 1 M NaCl. A VoyagerDE MALDI mass spectrometer (PerSeptive Biosystems) was used to determine the molecular weight of 17,197 close to the theoretical value 17194.

Crystals belonging to two different space groups ($I4_122$ and $P6_3$) were obtained by the vapor diffusion method in hanging drops at room temperature. The reservoir solutions consisted of 1.7 M ammonium sulfate, 0.1 M MES buffer at pH 6.5, and 1 mM $CoCl_2$. The hanging drops contained equal volumes of protein (15 mg/mL) and reservoir solution. A mercury heavy atom derivative was obtained by soaking crystals overnight in mother liquor containing 1 mM ethyl mercury phosphate. For data collection, crystals were briefly soaked in 90% saturated lithium sulfate solution buffered at pH 6.5 with 0.1 M MES and flash frozen in liquid propane.

Diffraction data were collected at the Advanced Photon

Grant sponsor: National Institute of Health; Grant number: P01 GM57890.

†John Toedt's current address is Department of Physical Science, Eastern Connecticut State University.

*Correspondence to: Osnat Herzberg, Center for Advanced Research in Biotechnology, University of Maryland Biotechnology Institute, 9600 Gudelsky Drive, Rockville, MD 20850. E-mail: osnat@carb.nist.gov.

Received 10 April 2002; Accepted 11 April 2002

TABLE I. Data Processing Statistics

Space group	$I4_122$	$P6_3$ (native)	$P6_3$ (Hg)
Cell dimension (Å)	$a = b = 112.3, c = 187.6$	$a = b = 118.2, c = 174.9$	$a = b = 114.4, c = 170.5$
No. of molecules in the asymmetric unit	3	8	8
Wavelength (Å)	1.00	1.00	1.00
Resolution (Å)	29–2.90	29–3.05	29–4.0
No. of measured reflections	399,292	1,095,449	234,477
No. of unique reflections	13,670	27,674	11,193[a]
Completeness (%)[b]	100 (100)	100 (100)	98 (96.8)
R_{merge}[b,c]	0.053 (0.294)	0.073 (0.310)	0.090 (0.175)
$\langle I/\sigma(I)\rangle$[b]	40.5 (9.4)	26.0 (9.2)	14.4 (10.2)
Redundancy	18.6	15.6	11.6

[a]Friedel pairs are treated as independent reflections.
[b]The values in parentheses are for the highest resolution shell.
[c]$R_{merge} = \Sigma_{hkl}((\Sigma_j|I_j - \langle I\rangle|)/\Sigma_j|I_j|)$, for equivalent reflections (for the Hg-derivative data, Bijvoet pairs were separated).

Source (APS) IMCA-CAT beam line 17-ID (Argonne National Laboratory). All data were collected at 100 K. Native data were collected for both crystal forms, and mercury derivative data were collected at 1.0Å for the $P6_3$ crystal form. The data were processed and scaled using HKL.[5] Data processing statistics are listed in Table I. Analysis of the anomalous differences at 4Å resolution using the program SHELXD[6] led to the identification of eight mercury sites. Phase calculations were carried out at 4Å resolution using MLPHARE.[7] Density modification, eightfold noncrystallographic symmetry (NCS)-averaging and phase extension (native hexagonal data) were performed at 3.0Å resolution using DM.[8] A model of one molecule was build manually, and the remaining asymmetric unit was generated by using the NCS relations. A roughly refined model was used to determine the structure of the crystal form $I4_122$ by molecular replacement using CNS.[9] Refinement was carried out with CNS, treating the three molecules in the asymmetric unit independently. Because the crystallization solution contained 1 mM $CoCl_2$, the metal ion in the active site was refined as Co(II). Refinement statistics are provided in Table II.

TABLE II. Refinement Statistics

Space group	14, 22		
Resolution (Å)	20.0–2.90		
Unique reflections F·2σ(F)	12,725		
Completeness (%)[a]	97.8 (94.8)		
No. of protein atoms	3,597		
No. of Co(II)	3		
No. of sulfate ions	1		
R[a,b]	0.198 (0.294)		
R_{free}[a,b]	0.290 (0.387)		
RMSD from ideal geometry			
Bond length	0.018 Å		
Bond angle	2.0°		
Average B factor (Å^2)	57		
Ramachandran plot (%)			
Most favored	82.6	Allowed	16.9
Generously allowed	0.5	Disallowed	0.0

[a]The values in parentheses are for the highest resolution shell, 3.0–2.9 Å.
[b]$R = \Sigma_{hkl}||F_o| - |F_c||/\Sigma_{hkl}|F_o|$, where F_o and F_c are the observed and calculated structure factors, respectively. R_{free} is computed for 8% of the reflections that were randomly selected and omitted from the refinement.

Results and Discussion. For the following discussion, the numbering scheme of the *E. coli* ispF is adopted. The numbers are lower by one compared with the sequential numbering of HI0671.

HI0671 asymmetric unit of the form crystallized in space groups $P6_3$ contains two trimers, and two monomers, each of which forms a trimer using a crystallographic threefold symmetry axis. The asymmetric unit of the $I4_122$ crystal form contains one trimer. Thus, there are a total of 11 independent views of the monomer, all exhibiting the same structural departures from the *E. coli* IspF structure (PDB entry code 1JY8). The molecules pack in both crystal forms as tetramers of trimers. As with the enzyme from *E. coli*, each monomer folds into a four-stranded β-sheet, and the β-sheets of three molecules face one another in a triangular cross section. The perimeter of the trimer is covered by α-helices [Fig. 1(a)]. An unusual set of interactions occurs at the center of the trimer, where three noncrystallographic symmetry related glutamic acid residues (Glu149) interact closely with one another. There is no space to accommodate a cation in between, but three equivalent histidine residues (His5) form salt bridges with the glutamates. The same interactions are present in the *E. coli* ispF. Another interaction that promotes trimer formation is due to a sulfate that binds at one end of the interface, forming charge-charge interactions with three symmetry-related arginine residues (Arg142, conserved in many but not all IspF sequences). An anion is absent in the *E. coli* IspF, although the three Arg142 guanidinium groups are in close proximity. The remaining trimer contacts are mediated by hydrophobic interactions.

Superimposition of the 1JY8 and HI0671 coordinates, omitting two loop regions that exhibit very different conformations (residues 14–38 and 60–70) resulted in a root mean square deviation (RMSD) between α-carbon atom pairs of 0.70 Å, demonstrating that the core fold is the same.

A one-turn α-helix (residues 31–34) is associated with the active site, and the invariant His34 provides the fourth ligand to the metal ion. The same segment of the *E. coli*

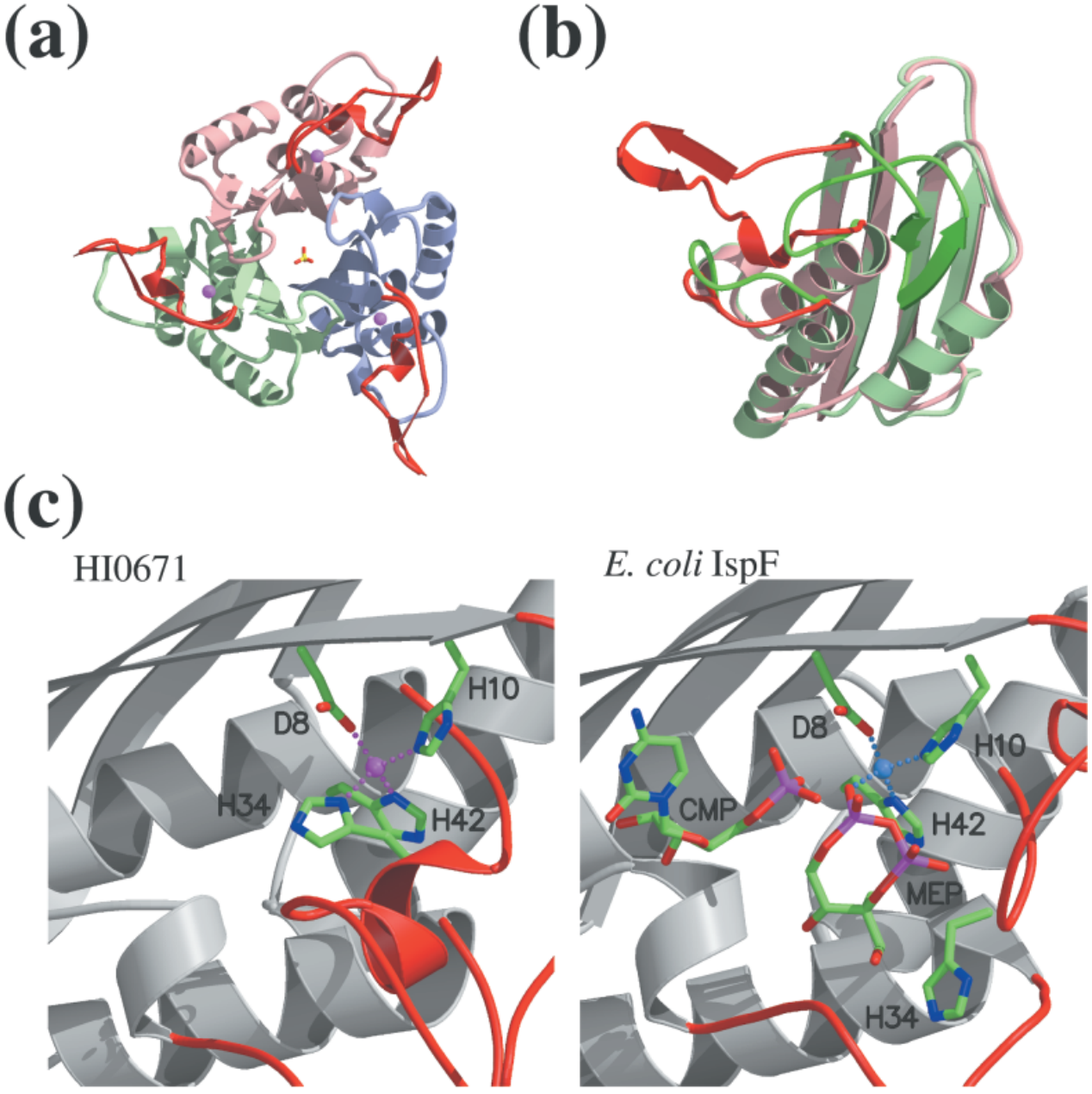

Fig. 1. Crystal structure of 2*C*-methyl-D-erythrol 2,4-cyclodiphosphate synthase (HI0671) from *H. influenzae*. **a:** Ribbon diagram with each monomer highlighted in different color. Cobalt(II) ions are shown as magenta spheres. The loop region that exhibits different conformation compared with that in the *E. coli* enzyme structure is colored in red. A sulfate ion is shown as a ball and stick model with the sulfur colored yellow and oxygen atoms colored red. **b:** Superpostion of the monomers of HI0671 and the IspF enzyme from *E. coli*. The common core structures are colored green-gray (*E. coli*) and red-gray (*H. influenzae*), and the deviating loop regions are colored green (*E. coli*) and red (*H. influenzae*). **c:** Comparison of the binding sites. The active site of HI0671 is blocked by the loop carrying His34, and the metal is coordinated to three histidine residues and one aspartic acid. The active site of the *E. coli* enzyme is open and occupied by 2*C*-methyl-D-erythritol 2-phosphate and CMP. Only two histidine residues coordinate the metal, and the fourth coordination is provided by a phosphate group of the product.

enzyme adopts a different conformation, with His34 pointing away from the active site. The superposition of the two structures [Fig. 1(b)] reveals that His34 of HI0671 occupies the position that is occupied by 2,4-cyclodiphosphate in the *E. coli* IspF. The preceding loop (residues 13–30), containing a short β-hairpin, protrudes into the solvent in HI0671. In contrast, this loop inserts between two α helices in the *E. coli* IspF, so that the β-hairpin packs against the four-stranded β-sheet, leaving the active site accessible to substrate. It is of interest that in both crystal forms of HI0671, the β-hairpin also inserts between the same two helices, but of a monomer that belongs to a different trimer. This is the packing interaction that leads to the formation of the tetramer of trimers. Finally, to prevent clashes with the 13–34 segment, HI0671 and *E. coli* IspF exhibit correlated conformational variability of

an adjacent polypeptide segment (residues 61–71), which are also shown in Figure 1(c).

The structures of IspF from two different organisms reveal that the 13–34 polypeptide segment can adopt at least two vastly different conformations, and this is associated with an invariant histidine switching between two states. In one state His34 coordinates to a catalytic metal, thus blocking access to the active site, and in the second state the histidine is displaced by a phosphoryl group of the substrate. The two conformations provide for a regulatory mechanism of enzyme activity used to prevent a wasteful side reaction, because IspF converts 4-diphosphocytidyl-2*C*-methyl-D-erythritol into 2*C*-methyl-D-erythritol 3,4-cyclophosphate, an undesirable product.[4] 4-Diphosphocytidyl-2*C*-methyl-D-erythritol is synthesized by IspD and is converted into the true IspF substrate, 4-diphosphocytidyl-2*C*-methyl-D-erythritol 2-phosphate, by the kinase IspE. In the genomes of many organisms, IspD and IspF are adjacent or even expressed as a single polypeptide chain. Inactivation of IspF before IspE action may therefore be essential. The identity of the factor that trigger enzyme activation await further investigations.

In the context of a Structural Genomics project, it is worthwhile to note that the function of HI0671, which was originally annotated "hypothetical," would not be decipherable from the structure alone. Although the fold is reminiscent of the trimeric proteins chorismate mutase,[10] and YjgF/YabJ,[11,12] (protein of unknown function), the topology is novel. Moreover, although the active site is located at the monomer-monomer interface, as seen in chorismate mutase, and is postulated for YjgF/YabJ, the architecture of each active site is vastly different. In the absence of the detailed biochemical work that elucidated the unusual reaction catalyzed by IspF, the structural analysis together with multiple-sequence alignment would only lead to a limited conclusion that the protein is an enzyme and that zinc is involved in activity.

Acknowledgments. We thank John Moult and Eugene Melamud for the development and help with their bioinformatics web site (http://www.s2f.umbi.edu). We thank the staff at the Advanced Photon Source, IMCA-CAT, for their help during data collection. The U.S. Department of Energy, Basic Energy Sciences supported use of the Advanced Photon Source under contract W-31-109-Eng-38. Protein Data Bank entry code: 1JN1.

REFERENCES

1. Rohmer M, Knani M, Simonin P, Sutter B, Sahm H. Isoprenoid biosynthesis in bacteria: a novel pathway for the early steps leading to isopentenyl diphosphate. Biochem J 1993;295:517–524.
2. Steinbacher S, Kaiser J, Wungsintaweekul J, Hecht S, Eisenreich W, Gerhardt S, Bacher A, Rohdich F. Structure of 2C-methyl-d-erythritol-2,4-cyclodiphosphate synthase involved in mevalonate-independent biosynthesis of isoprenoids. J Mol Biol 2002;316:79–88.
3. Eisenstein E, Gilliland GL, Herzberg O, Moult J, Orban J, Poljak RJ, Banerjei L, Richardson D, Howard AJ. Biological function made crystal clear—annotation of hypothetical proteins via structural genomics. Curr Opin Biotechnol 2000;11:25–30.
4. Herz S, Wungsintaweekul J, Schuhr CA, Hecht S, Luttgen H, Sagner S, Fellermeier M, Eisenreich W, Zenk MH, Bacher A, et al. Biosynthesis of terpenoids: YgbB protein converts 4-diphosphocytidyl-2C-methyl-D-erythritol 2-phosphate to 2C-methyl-D-erythritol 2,4-cyclodiphosphate. Proc Natl Acad Sci USA 2000;97:2486–2490.
5. Otwinowski Z, Minor W. Processing of x-ray diffraction data collected in oscillation mode. Methods Enzymol 1997;276:307–326.
6. Sheldrick GM. SHELX applications to macromolecules. In: Fortierr S, editor. Direct methods for solving macromolecular structures. The Netherlands: Kluwer Academic Publishers; 1998. p 401–411.
7. Otwinowski Z. Maximum likelihood refinement of heavy atom parameters. In: Wolf W, Evans PR, Leslie AGW, editors. A conf proceedings: Isomorphous replacement and anomalous scattering. Warrington, UK. Daresbury Laboratory. p 80–86.
8. Cowtan K. "DM": an automated procedure for phase improvement by density modification. Joint CCP4 and ESF-EACBM Newsletter on Protein Crystallography 1994;31:34–38.
9. Brünger AT, Adams PD, Clore GM, DeLano WL, Gros P, Grosse-Kunstleve RW, Jiang JS, Kuszewski J, Nilges M, Pannu NS, et al. Crystallography & NMR system: a new software suite for macromolecular structure determination. Acta Crystallogr D Biol Crystallogr 1998;54:905–921.
10. Chook YM, Ke H, Lipscomb WN. Crystal structures of the monofunctional chorismate mutase from Bacillus subtilis and its complex with a transition state analog. Proc Natl Acad Sci USA 1993;90:8600–8603.
11. Volz K. A test case for structure-based functional assignment: the 1.2 A crystal structure of the yjgF gene product from Escherichia coli. Protein Sci 1999;8:2428–2437.
12. Sinha S, Rappu P, Lange SC, Mantsala P, Zalkin H, Smith JL. Crystal structure of Bacillus subtilis YabJ, a purine regulatory protein and member of the highly conserved YjgF family. Proc Natl Acad Sci USA 1999;96:13074–13079.