doi:10.1016/j.jmb.2007.05.006 J. Mol. Biol. (2007) 370, 899–911

JMB

Available online at www.sciencedirect.com

ScienceDirect

ELSEVIER

Structure and Mechanism of HpcG, a Hydratase in the Homoprotocatechuate Degradation Pathway of *Escherichia coli*

Atsushi Izumi[1], Dean Rea[2], Tomoko Adachi[1], Satoru Unzai[1] Sam-Yong Park[1], David I. Roper[2] and Jeremy R. H. Tame[1]*

[1]*Protein Design Laboratory Yokohama City University Suehiro 1-7-29, Tsurumi Yokohama 230-0045, Japan*

[2]*Department of Biological Sciences, University of Warwick, Gibbet Hill Road Coventry, CV4 7AL England, UK*

**Corresponding author*

HpcG catalyses the hydration of a carbon–carbon double bond without the aid of any cofactor other than a simple divalent metal ion such as Mg^{2+}. Since the substrate has a nearby carbonyl group, it is believed that it first isomerises to form a pair of conjugated double bonds in the enol tautomer before Michael addition of water. Previous chemical studies of the reaction, and that of the related enzyme MhpD, have failed to provide a clear picture of the mechanism. The substrate itself is unstable, preventing co-crystallisation or soaking of crystals, but oxalate is a strong competitive inhibitor. We have solved the crystal structure of the protein in the apo form, and with magnesium and oxalate bound. Modelling substrate into the active site suggests the attacking water molecule is not part of the metal coordination shell, in contrast to a previous proposal. Our model suggests that geometrically strained *cis* isomer intermediates do not lie on the reaction pathway, and that separate groups are involved in the isomerisation and hydration steps.

Keywords: enzyme; intermediate; X-ray crystallography; aromatic; bioremediation

Introduction

Bacterial degradation of aromatic compounds is crucial to carbon recycling in the environment, and of particular interest for biodegradation and bioremediation.[1,2] A wide range of such compounds, including man-made pollutants, is broken down *via* oxygenation and subsequent meta cleavage. This is followed by a sequence of enzymatic steps, common to several pathways, leading to metabolites of the citric acid cycle. The dioxygenases at the head of these degradation pathways have been studied in detail,[3] but so far rather less attention has been paid to the downstream enzymes, which are interesting from both a biochemical and evolutionary perspective. 4-Hydroxyphenylacetic acid (4-HPA) is degraded by a long catabolic pathway involved also in the breakdown of tyrosine and phenylalanine.[4,5] 4-HPA is first converted to 3,4-dihydroxyphenylacetic acid (homoprotocatechuate, HPC), and then *via* a series of steps to pyruvate and succinic semialdehyde, which enter the central metabolic pathway.[6] Current understanding of these chemical transformations is summarised in Figure 1. The genes involved in 4-HPA degradation were first cloned from *Escherichia coli* C[6] and *E. coli* W.[7] Genome sequencing has since revealed the presence of highly related pathways in a wide range of bacteria, as well as plant and animal homologues of particular enzymes, allowing comparison of the various enzyme mechanisms and their evolution. A number of the enzymes in the 4-HPA pathway have been characterised biochemically. Crystal structures of several of the enzymes have been solved, including HpcD[8] and HpcE.[9] The enzyme activities may be useful in stereoselective chemical syntheses, and a better understanding of them may help develop biodegradable materials, as well as improve bioremediation of aromatic pollutants. The crystal structures of the enzymes before and after HpcG in

Abbreviations used: 4-HPA, 4-Hydroxyphenylacetic acid; rmsd, root-mean-square deviation; FAH, fumarylacetoacetate hydrolase; HHDD, 2-hydroxy-2,4-heptadiene-1,7-dioate.

E-mail address of the corresponding author: jtame@tsurumi.yokohama-cu.ac.jp

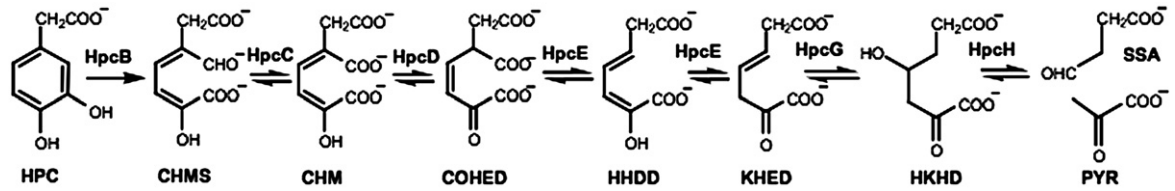

Figure 1. The homoprotocatechuate degradation pathway. The names of the enzymes, and the structures and names of the intermediates are shown. HPC, 3,4-dihydroxyphenylacetate (homoprotocatechuate); CHMS, 5-carboxymethyl-2-hydroxymuconate semialdehyde; CHM, 5-carboxymethyl-2-hydroxymuconate; 5-(carboxymethyl)-2-oxo-3-hexene-1,6 diaote, 5-carboxymethyl-2-oxo-hex-3-ene-1,6-dioate; HHDD, 2-hydroxyhepta-2,4-diene1,7-dioate; KHED, 2-ketohepta-4-ene-1,7,dioate; HKHD, 4-hydroxy-2-ketoheptane-1,7-dioate; SSA, succinic semialdehyde; PYR, pyruvate.

the pathway (HpcE and HpcH, respectively) have been solved, but so far no atomic model of HpcG has been published. The mechanism of HpcG (and the related hydratase MhpD) has been studied in some detail, but a number of important problems remain, including the question of how the attacking water molecule is activated and the nature of the intermediates. HpcG has been crystallised in the apo form and with inhibitor bound in order to understand the details of the mechanism and substrate specificity.

Results

HpcG was purified and crystallised as described.[10] The apo form of the enzyme, with no added metal or substrate, yielded crystals that diffracted to high resolution. Molecular replacement was attempted with various search models without success, but phasing was achieved by single-wavelength anomalous diffraction from selenomethionine-containing crystals. Refinement proceeded straightforwardly to the final model (Table 1). The molecular mass was measured by analytical ultracentrifugation to be 275 kDa, and sedimentation velocity data show only a single species in solution, indicating that the protein forms a stable decamer. This decamer lies across the crystallographic 2-fold, such that the asymmetric unit contains a pentameric ring (Figure 2). Comparing adjacent monomers within the pentamer using PISA shows a buried surface area of approximately 500 Å^2 per monomer.[11] The 2-fold related pentamers bury 900 Å^2 per monomer, which is consistent with a stable 10mer form. Searching the PDB with DALI[12,13] for related structures yielded the model of MhpD (PDB 1sv6), which is clearly strongly related to HpcG (32% sequence identity), but this entry has not been de-

Table 1. Summary of data collection and refinement statistics

	SeMet	Apo	Mg	Mg-oxalate
Beam line	PF, AR-NW12	SPring8, BL38B2	SPring8, BL41XU	SPring8, BL38B1
No. reflections	684,761	1,391,577	1,129,542	1,202,013
Unique reflections	80,880	223,282	195,362	188,655
Unit cell dimensions				
a = b (Å)	137.0	136.7	136.0	136.5
c (Å)	192.2	192.9	194.1	192.4
Space group		$P4_32_12$		
Wavelength (Å)	0.9789 (peak)	1.000	1.000	1.000
Resolution (Å)	2.30 (2.38 ~ 2.30)	1.60 (1.67 ~ 1.60)	1.69 (1.74 ~ 1.69)	1.70 (1.76 ~ 1.70)
R_{merge}[a]	0.070 (0.259)	0.069 (0.406)	0.086 (0.261)	0.069 (0.504)
Completeness (%)	99.3 (98.3)	93.3 (68.4)	97.1 (96.2)	95.0 (65.8)
Redundancy	8.5 (5.6)	6.2 (2.1)	5.8 (4.6)	6.4 (2.4)
$< I/\sigma\ (I) >$	28.1	17.9	22.0	10.4
R-factor (%)		20.7	19.5	19.4
R_{free} (%)		22.9	21.3	22.6
No. protein atoms		10,248	10,461	10,170
No. water molecules		1009	970	756
Average *B*-factors (Å^2)				
Protein atoms		25.4	21.5	24.3
Overall		26.3	22.2	24.7
rms deviations from ideal				
Bond lengths (Å)		0.007	0.007	0.012
Bond angles (°)		1.12	1.04	1.36
Chiral volume (Å^3)		0.076	0.075	0.091

Numbers in parentheses refer to the highest resolution shell.

[a] $R_{merge} = \sum | I_i - \langle I_i \rangle | / \sum | I_i |$, where I_i is the intensity of an observation and $\langle I_i \rangle$ is the mean value for that reflection, the summations being over all reflections.