The Crystal Structure of HpcE, a Bifunctional Decarboxylase/Isomerase with a Multifunctional Fold†

Jeremy R. H. Tame,*,§,‡ Keiichi Namba,§,|| Eleanor J. Dodson,⊥ and David I. Roper#

Protein Design Laboratory, Yokohama City University, Suehiro 1-7-29, Kanagawa 230-0045, Japan, Protonic NanoMachine Project, ERATO, JST, 3-4 Hikaridai, Seika, Kyoto 619-0237, Japan, Advanced Technology Research Laboratories, Matsushita Electric Industrial Co., Ltd., 3-4 Hikaridai, Seika, Kyoto 619-0237, Japan, York Structural Biology Laboratory, Department of Chemistry, University of York, York YO10 5DD, England, and Department of Biological Sciences, University of Warwick, Coventry CV4 7AL, England

Received August 23, 2001; Revised Manuscript Received November 29, 2001

ABSTRACT: The structure of the bifunctional enzyme HpcE (OPET decarboxylase/HHDD isomerase) from *Escherichia coli* shows that the protein consists of highly similar N and C terminal halves. Sequence matches suggest that this fold is widespread among different species, including man. Many of these homologues are uncharacterized but apparently connected with the metabolism of aromatic compounds. The domain shows similar topology to the C terminal domain of fumarylacetoacetate hydrolase (FAH), a functionally related enzyme, despite lacking significant overall sequence similarity. HpcE is known to catalyze two rather different reactions, and comparisons with FAH allow some tentative conclusions to be drawn about the active sites. Key mutations within the active site apparently allow enzymes with this fold to carry out a variety chemical processes.

Bacterial degradative pathways of aromatic compounds are of considerable interest for bioremediation and biodegradation of waste products. One such pathway in *Escherichia coli* C breaks down homoprotocatechuate (3,4-dihydroxyphenylacetate, HPC),[1] a product of tyrosine and phenylalanine metabolism also released by lignin catabolism. The inducible extradiol (*meta*) cleavage HPC pathway is one of the longest degradative pathways known, and shows a strong resemblance to the dehydrogenative catechol catabolic pathway encoded by the TOL plasmid (*1*, *2*). The 4-hydroxyphenylacetate pathway of *E. coli* W is highly similar, the necessary genes being carried on a 15-kb cassette (*3*). Seven sequential steps reduce HPC to succinic semialdehyde and pyruvic acid. Sequencing of the genes involved and characterization of the enzymes showed that two consecutive steps, a decarboxylation and an isomerization, were carried out by the same protein, HpcE (Figure 1). The sequence of the 45-kD monomeric protein shows clear sequence similarity between the N and C terminal halves, suggestive of two similar domains and raising the possibility that each domain corresponded to an active site, the substrate being channelled between the two. HpcE from *E. coli* C is identical to HpaG of the *E. coli* W pathway (*3*). The crystal structures of several isomerases in similar pathways have been reported. HpcE however shows little sequence similarity to these including CHMI (5-carboxymethyl-2-hydroxymuconate isomerase), which catalyzes an earlier step in the HPC pathway and acts on a similar substrate (*4*). Indeed, sequence homology searches found no apparent similarity between HpcE and any previously solved structures. To examine the differences in the two domains of the protein, and to examine the basis of its marked preference for 2-hydroxyhepta-2,4-diene-1,7-dioate (HHDD) over CHM (5-carboxymethyl-2-hydroxymuconate), we have solved the structure by multiple isomorphous replacement (MIR) and refined it to 1.7 Å.

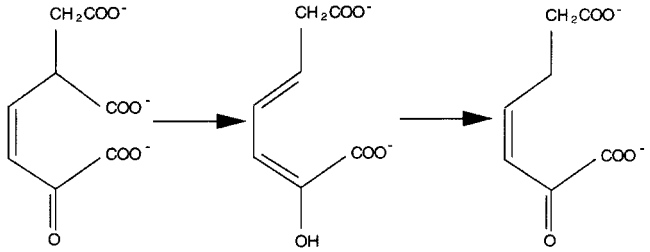

FIGURE 1: The chemical reactions catalyzed by HpcE. Carbon dioxide is lost in the first step, and the subsequent isomerization is required to lose a second carbon dioxide molecule in the next step in the degradation pathway. The isomerization is reversible, but the biological direction is shown.

RESULTS AND DISCUSSION

Overall Structure. The final model consists of four copies of the 429-residue protein in the asymmetric unit, two pairs of monomers being related by a simple translation of 1/2 along *z* and 0.03 along *x*. A diagram of the overall monomer structure is shown in Figure 2, and refinement statistics are given in Table 1. Not all residues are visible in the electron density map, a break occurring from residues 236 to 243

† Funding is from the Royal Society University Research Fellowship (JRHT), Japan Science and Technology Corporation (KN), BBSRC (EJD), and Medical Research Council (DIR).

* Corresponding author. E-mail: jtame@tsurumi.yokohama-cu.ac.jp. Tel. +81 (0)45 508 7228. Fax +81 (0)45 508 7366.

‡ Yokohama City University.

§ Protonic NanoMachine Project, ERATO.

|| Matsushita Electric Industrial Co., Ltd.

⊥ University of York.

University of Warwick.

[1] Abbreviations: HHDD, 2-hydroxyhepta-2,4-diene-1,7-dioate; OPET, 5-oxopent-3-ene-1,2,5-tricarboxylate; FAH, fumarylacetoacetate hydrolase; 4-OT, 4-oxalocrotonate tautomerase; CHMI, 5-carboxymethyl-2-hydroxymuconate isomerase; HPC, homoprotocatechuate (3,4-dihydroxyphenylacetate); ODC, orotodine 5′ monophosphate decarboxylase.

10.1021/bi015717t CCC: $22.00
Published on Web 02/08/2002

FIGURE 2: The overall structure of of HpcE. This figure was drawn using MOLSCRIPT (*29*) and RENDER (*30*). β sheets are shown as arrows, the secondary structure assignments being made automatically by MOLSCRIPT. The polypeptide chain is colored from the N terminus (blue) to the C terminus (red). This view, approximately down the pseudo 2-fold axis of the protein, shows residues 1 and 2 plugging one end of the N domain barrel. The position of the calcium ion in the C domain is shown as a gray sphere.

Table 1.[a]

data collection statistics		
resolution	91.2−1.7 Å	1.75−1.70 Å
R merge (%)	5.5	54.6
$<I/\sigma(I)>$	18.6	2.2
completeness (%)	98.5	98.3
B factor	22.0 Å^2	
multiplicity	3.4	3.0
refinement statistics		
resolution limits (Å)	91.2−1.7	
no. of used reflections	186 453	
percentage observed	98.46	
percentage free reflections	5.0550	
overall *R* factor	0.2155	
free *R* factor	0.2520	
overall correlation coefficient	0.9586	
free correlation coefficient	0.9438	
ML based su of positional parameters	0.0978	
ML based su of thermal parameters	3.134	
average *B* value for all atoms	38.2	
average *B* value for main chain atoms	35.9	
bond length (Å)	0.019 (weight 0.021)	
bond angle (°)	1.851 (weight 1.964)	

[a] R factor $= \Sigma||F_{obs}| - |F_{calc}||/\Sigma F_{obs}$. The high R factors are due to the translational NCS that gives rise to a class of systematically weak reflections. No sigma cutoff was applied at any stage of data handling. Four residues appear in "forbidden" areas of the Ramachandran plot. Two of these, Glu 58 and Arg 212, are found at flexible loops on the molecule and have relatively poor density. Asp 234 occurs just before the break in the density at His 235, which is clearly highly mobile. Lys 29 is found at a surface loop close to a cis-proline (Pro 27) but has relatively strong electron density in three copies of the molecule.

inclusive, which are omitted from each copy of the protein in the model. These residues form a loop lying away from the body of the protein and possibly only take up a fixed conformation on substrate binding. Crystal contacts are formed by cysteine 48, which makes a disulfide bond with its equivalent on neighboring subunits. Oxidation is not complete, and the sulfur has been modeled into alternate conformations.

There is 34% sequence identity between the N and C domains of HpcE, with residues 1−200 matching 221−429 (Table 2), which led to the suggestion that the two have similar folds (*2*). It therefore was not surprising to find each monomer consists of two highly similar incomplete barrel-like domains each measuring 30 × 40 × 40 Å. No significant similarity to known protein folds was found by searching PDB using DALI (*5*). The cores of the two domains may be superimposed with remarkably small deviations in the main-chain atom positions; the fit of the main-chain atoms for the 137 residues in the barrel of each domain gives an rms difference of only 1.03 Å. The structure is very different from two previously solved enzyme structures involved in similar reactions (*4*), but the significant sequence homology of the N and C terminal domains with numerous other proteins suggested the same fold occurs in mammalian as well as bacterial proteins, many of unknown function.

Domain Structure. HpcE is a mainly β protein, with long stretches of coil connecting the strands, which form two principal sheets in each domain. There are only three helices longer than four residues, two of which are found in equivalent positions of the N and C terminal domains. The extra helix occurs between the first and second strands of sheet 1 in domain N, at a position equivalent to the flexible loop in domain C which is not visible in the electron density maps. A schematic diagram of the connectivity is shown in Figure 3, and the residue numbering of the β strands is given in Table 3. The two domains form a heterodimer related by a pseudo 2-fold rotation axis. Each domain is roughly 200 residues long, and residues 206 to 220 form a linker that reaches across to the start of domain C. The domains consist of two mainly antiparallel β sheets which exchange strands, the chain passing from one sheet to the other five times. The larger sheet (I) is strongly twisted into a barrel-like structure, with the smaller sheet (II), also highly twisted, lying against it. The gap between the edge strands of sheet I is filled by residues 106 to 113 in domain N, and residues 319 to 330 in domain C, both stretches including a one-turn helix. The barrels of both N and C domains are plugged at one end, by residues 1−2 and 220−221, respectively. The other end of each barrel forms a cavity. Overall, the barrel may be likened to a cupped hand, the fingers (running across the β strands) curling over the palm and the thumb blocking one end of the tunnel formed. A short helix sits between the finger tips and base of the thumb, leaving a cavity lined by the little finger. Side-chains within the cavity in each domain possibly involved in catalysis are shown in Figure 4. There is a calcium ion bound in this cavity in the C domain only. The open end of the N domain cavity is rather more occluded by a loop from residues 98−110. The entry of a substrate into this active site would have to be through a hole lined by Phe 102, Val 36, Ala 9, Phe 38, Pro 105, and Lys 110. The short equivalent loop in the C domain (residues 297−299) lies away from the axis of the barrel and does not obstruct small molecules from entering the lumen. In Figure 4b, it is therefore possible to look down the core of the barrel behind the calcium ion to residues 220 and 221 which block the far end.

The close similarity of the two domains among all four copies of the protein in the asymmetric unit suggests a fairly rigid core structure, although movement of the loops to allow substrate access to the active sites is clearly possible. Removing noncrystallographic symmetry restraints in the later stages of refinement did not lead to large changes, although there are some differences in surface side-chains close to crystal contacts. The loop from Leu 97 to Ala 109 may in fact move as a lid covering the mouth of the N domain active site.

Examining other related structures, it became obvious that the hetero "dimer" consisting of the N and C domains of

Table 2: Sequence Alignment of the N and C Domains of HpcE[a]

```
                      EEEEE   HHHHHH                  EEEEE           EEE
N domain   1     MKGTIFAVALNHRSQLDAWQEAFQQSPIKAPPKTAVWFIKPRNTVIGCGEPIPFPQ-GEN  59
C domain 221     PHGTLFALGLNYADHASELEFK--------PPEEPLVFLKAPNTLTGDNQTSVRPNNIEY 272
                     EEEEE                          EEEEE         EEEEE
                  :**:**:.**: .: .  :          **: .: *:*. **: * .:.   *:  *

                  EEEEEEEEE                  EEEEEEE                        EEE
N domain         LLSGATVALIVGKTATKVREEDAAEYIAGYALANDVSLPE--ESFYRPAIKAKCRDGFCP 117
C domain         MHYEAELVVVIGKQARNVSEADAMDYVAGYTVCNDYAIRDYLENYYRPNLRVKSRDGLTP 332
                 EEEEEEEEEEE               EEEEEEEE                        EEE
                 :   * :.:::** * :* * ** :*:***::.** :: :  *.:*** ::.*.***: *

                 E                EEEEEE   EEEEEE        HHHHHHHHH         EEE
N domain         IGETVA----LSNVDNLTIYTEINGRPADHWNTSDLQRNAAQLLSA<b>LSEFATLNPGDAI</b>L 173
C domain         MLSTIVPKEAIPDPHNLTLRTFVNGELRQQGTTADLIFSVPFLIAY<b>LSEFMTLNPGDMI</b>A 392
                 EE               EEEEEE  EEEEEEEE        HHHHHHHH       EEEE
                 : .*:.    :.: .***: * :**.  :: .*:**  ... *:: **** ****** *

                              EEEEEE     EEEEEEE
N domain         LGTPQARVEIQPGDRVRVLAEGFPPLENPVVDEREVT 210
C domain         TGTPKGLSDVVPGDEVVVEVEGVGRLVNRIVSEETAK 429
                             EEEEEEE  EEEEEEEE
                  ***:.  :: ***.* * .**.  * * :*.*. ..
```

[a] Residues 1−210 were aligned with residues 221−429 using ClustalW. Identical residues in the alignment are indicated by asterisks (*), very similar residues are indicated with colons (:), and similar residues are indicated with a period (.). The secondary structure is indicated in the lines above (N domain) and below (C domain) the sequences (H − helix, E − β strand). The highly conserved 12-residue sequence Leu 160 to Ile 172 matches Leu 379 to Ile 391 (shown in bold). These sequences form almost identical loop structures in the two domains with an overall rms deviation of 0.57 Å over the 52 main-chain atoms compared. Phe 163 and Phe 382 are equivalent residues in the N and C domains, respectively. They lie at the C terminal end of an α helix and pack against two positively charged residues on the opposite domain, forming a key contact between the two halves of the protein.

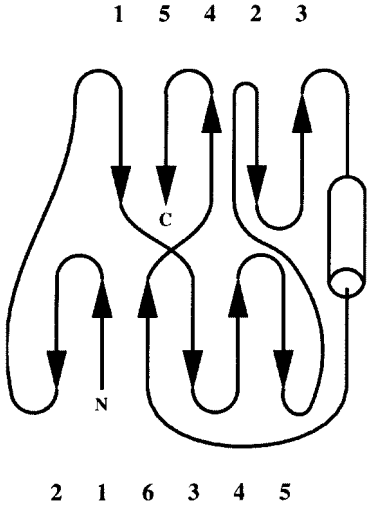

FIGURE 3: Topology diagram showing the common features of the individual domains of HpcE. Short helices (less than five residues) are omitted, and β strands are shown as arrows. This diagram shows the connectivity between the two β sheets in each domain. FAH shows identical connectivity but has longer surface loops and an extra domain that make the protein larger overall.

Table 3: Residue Numbers of Strands in Sheets I and II of the N and C Domains

N domain	sheet I	sheet II
strand 1	5−9	52−54
strand 2	36−40	131−136
strand 3	61−69	140−145
strand 4	87−93	188−193
strand 5	115−118	199−205
strand 6	171−173	
C domain	sheet I	sheet II
strand 1	225−229	262−266
strand 2	248−252	350−355
strand 3	273−283	358−365
strand 4	299−306	407−413
strand 5	330−334	416−423
strand 6	389−392	

HpcE is related to the C terminal domains of the homodimer of mouse fumarylacetoacetate hydrolase (FAH; *6*). A comparison between the principal β strands is shown in Figure 5. It is not clear why this similarity was not detected by DALI, but the differences in the structures are apparently large enough to mask the underlying topological similarity. FAH hydrolyzes a variety of diketoacids and catalyzes the last step in tyrosine and phenylalanine catabolism. In humans, mutations in the FAH gene lead to hereditary tyrosinemia type I, which is associated with liver failure and may be fatal. FAH subunits have a molecular weight of 46 kDa. The crystal structure has a calcium ion coordinated to several carboxylate side chains (*6*), and mutations at this site inactivate the enzyme (*7*). The FAH−product complex has been solved by crystal soaking with fumarate and acetoacetate, showing the acetoacetate binds directly to the calcium (*6*).

Structural Comparison with FAH. FAH consists of N and C terminal domains, the N domain being about 120 residues long and largely forming β strands. HpcE has no counterpart to this domain, but its N and C terminal halves each strongly

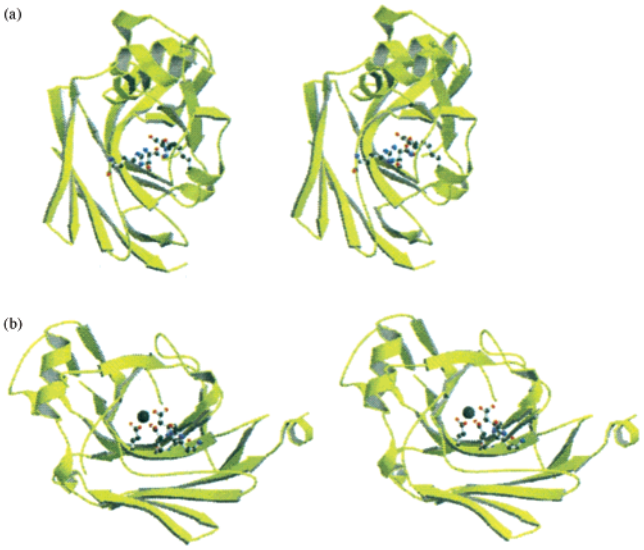

FIGURE 4: The active sites of the protein. Despite the highly conserved nature of the barrel structure in each domain, different residues are found within the pocket enclosed by β sheet I in the N and C domains. (a) In the N domain, a Lys 110 and Arg 180 each form salt-bridges with Asp 94 and Glu99. The short loop from 98 to 100 (seen to the right in this view) partially occludes the entrance to the cavity. (b) In the C domain, Glu 276, Glu 278, and Asp 307 coordinate a calcium ion. The 2Fo-Fc electron density map in the region of these side-chains shown in Figure 6.

FIGURE 5: Comparison of the main β sheet structure of the N domain of HpcE and FAH. The upper panel shows HpcE, and the lower panel shows FAH. The equivalent β sheets have been colored blue and green. FAH was the first protein structure determined in which a β sandwich and a β roll are combined in this way (*6*). The domains of HpcE represent a simpler version of the same fold. This figure was made with MOLSCRIPT (*29*) and RENDER (*30*).

resemble the C terminal domain of FAH. In solution, FAH exists as a dimer linked by contacts between this C terminal barrel domain. Analytical ultracentrifugation experiments show HpcE to be monomeric (data not shown). Overlapping the similar structural features between HpcE and FAH revealed some sequence similarity (Table 4). (The sequences of both HpcE and FAH are available from the SwissProt database, entries P37352 and P33505, respectively, but the entry for HpcE P37352 is incorrect and gives a truncated sequence of 405 amino acid residues. The latest GenBank sequence entry has accession number X75028, but the X-ray model shows the dipeptide sequence Val−Pro is inserted between Val 402 and Gly 403 of this sequence. The modified sequence is used throughout this paper.) Sequence comparison between the last 210 residues of HpcE and FAH gives a sequence identity of 15.0%, with a region of about 50 residues being highly conserved between the two. Superposing the proteins also overlaps the calcium ion in each structure. Remarkably, overlapping the main-chain atoms of the two structures on this stretch of sequence (residues 215−265 of FAH with 288−338 of HpcE) automatically aligns the N domain of HpcE with the dimer partner of FAH, showing the relative orientation of the two domains is very similar in both proteins. The rms deviation of the 204 atoms used is 3.8 Å. Using this alignment to match residues visually, 46 residues (184 main-chain atoms) were selected from the barrel core which gave an rms deviation of only 0.65 Å. The barrel structure of FAH and HpcE is therefore clearly very highly conserved between the mammalian and bacterial enzymes. Despite holding the barrels in very similar positions, the domain interfaces are not very similar. This region is very hydrophobic in HpcE and forms a typical protein-core like structure. In FAH, the 2-fold axis between the identical monomers passes between two histidine residues (His 160) which form main-chain to side-chain hydrogen bonds with each other. The buried surface area was calculated with the programs SURFACE and AREAIMOL from the CCP4 suite (*8*), treating residues 1−206 and 210−429 of HpcE as separate domains. A total surface area of 3300−3400 Å^2 is buried between the two. The interface is larger in FAH, partly due to an extra helix between residues 237 and 246, and 5500−5600 Å^2 are excluded from solvent on dimer formation. This represents about 15% of the surface area of each FAH monomer (18 200 Å^2) and suggests very strong self-association. The smaller interface found between HpcE domains presumably reflects the large entropic advantage of covalent attachment, and the domains associate without requiring a very large number of atomic contacts.

N Domain. There are two significant cavities within the protein, located at equivalent positions within the center of each domain. These cavities are at the equivalent positions of the substrate binding site and calcium ion in FAH. The N domain cavity has an unusual tetrad of charges formed by Asp 94, Glu 99, Lys 110, and Arg 180, which lie roughly in a plane (Figure 6). The opposite face of the cavity is much more apolar, although the bottom of the pocket is also relatively polar but uncharged. Leu 10, Asn 11, and His 12 all present their main-chain nitrogen atoms to the interior, and a carboxyl group can fit into this position with the charge cluster sitting next to a possible substrate. An extensive hydrogen bonding network holds the side-chains inside the pocket in place, and their temperature factors are similar to those of other internal residues.

Since FAH PDB entry 1QCO has fumarate and acetoacetate modeled into the active site, overlapping it onto HpcE as described above allows the substrate binding site of FAH to be compared with equivalent residues in HpcE. The loss of calcium binding in the N domain is readily explained by changes in the metal coordinating residues in FAH. Asp 126,

Table 4: Sequence Alignment of HpcE C Domain and FAH[a]

```
                      = =                                  =
HpcE          YMHYEAELVVVIGKQARN---VSEADAMDYVAGYTVCNDYAIRDYLENYYRPN------- 321
FAH           LLDMELEMAFFVGPGNRFGEPIPISKAHEHIFGMVLMNDWSARDIQQWEYVPLGPFLGKS 254
               :. * *:...:*   *    :. :.* ::: * .: **:: **  :  * *

HpcE          ------LRVKSRDGLTPMLSTIVPKEAIPDP---HNLTLRTFVNGELRQQG---TTADLI 369
FAH           FGTTISPWVVPMDALMPFVVPNPKQDPKPLPYLCHSQPYTFDINLSVSLKGEGMSQAATI 314
                     * . *.* *:: .   ::. * *   *. .    :* .:  :*   : *  *

HpcE          FSVPFLIAYLSEFMTLN----------PGDMIATGTPKG--------------------- 398
FAH           CRSNFKHMYWTMLQQLTHHSVNGCNLRPGDLLASGTISGSDPESFGSMLELSWKGTKAID 374
                  *   * : :  *.          ***::*:** .*

HpcE          ------LSDVVPGDEVVVE--------VEGVGRLVNRIVSEETAK 429
FAH           VEQGQTRTFLLDGDEVIITGHCQGDGYRVGFGQCAGKVLPALSPA 419
                    : :: ****::           *.*: ..::::.  :.
```

[a] Sequence alignment of the C terminal domain of HpcE with mouse FAH. The alignment starts at Tyr 272 (HpcE) and Leu 195 (FAH). A close match between the sequences starting at Pro 387 (HpcE) and Pro 342 (FAH), both shown in bold, overlaps the 12-residue sequence conserved between the N and C domains of HpcE highlighted in Table 2. The following residues form a strand of the barrel structure. The equal signs above the top line of the alignment indicate the calcium-binding carboxyl groups conserved between HpcE and FAH (Glu 276, Glu 278 and Asp 307 in HpcE). Extra loops found in FAH as compared to HpcE are apparent in the alignment.

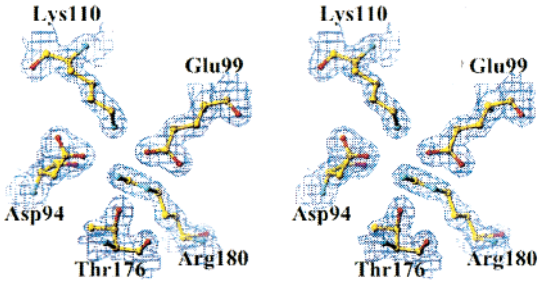

FIGURE 6: Close-up view of the active site in the N terminal domain. The electron density map is shown at 1.2 σ.

Glu 199, and Glu 201 in FAH have equivalent residues Ala 7, Gly 63, and Thr 65 in the HpcE N domain, and the side-chain of Arg 180 in HpcE occupies the same position as Glu 199 in FAH. The N domain has room to bind acetoacetate but not fumarate, whose binding site is blocked by Phe 102. The fumarate binding site in FAH includes Arg 142, from the domain having no counterpart in HpcE. The residues around acetoacetate are conserved to some extent; Phe 38, Lys 110, and Thr 176 closely mirror the FAH residues Tyr 159, Lys 253, and Thr 350. Acetoacetate hydrogen bonds to His 133 in FAH, which lies close to Gln 15 in the superposed HpcE. Since acetoacetate also binds directly to the calcium ion however, the N domain is unlikely to bind it strongly.

C Domain. Unlike the N domain, the C domain has a triad of carboxyl groups coordinated to a metal ion (Figure 7a). The three side-chains involved, Glu 276, Glu 278, and Asp 307 are also held in place by two lysine side-chains (Lys 252 and Lys 325) and the main-chain nitrogens of Thr 393 and Gly 394. Decarboxylase enzyme activity is reported to be magnesium dependent, and the protein loses this activity if purified in the absence of magnesium (*9*). (Since no magnesium was added to the protein and the crystals were grown in 200 mM calcium chloride the metal ion present in the structure is presumably calcium.) This suggests the metal-binding domain is the site of decarboxylase activity. This is

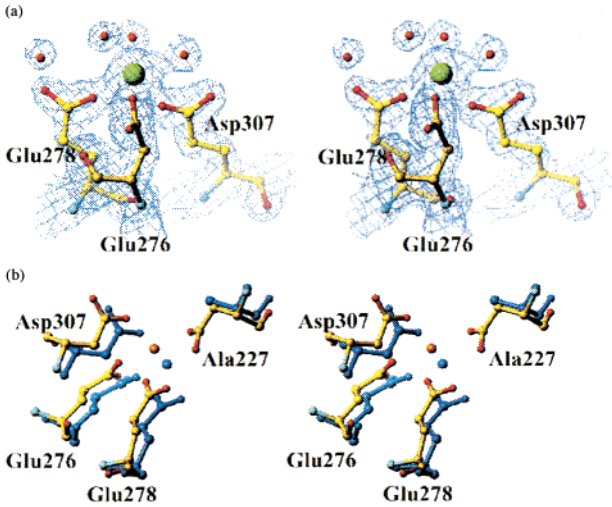

FIGURE 7: (a) Close-up view of the active site in the C terminal domain. The electron density map is shown at 1.2 σ. Water molecules are shown as red spheres, and the calcium ion is colored green. (b) The overlap of the calcium binding sites of FAH and HpcE. The two proteins were overlapped on 46 residues in the barrel core. The main-chain atoms were fitted between 123 and 126 (224−227), 196−206 (273−283), 226−234 (300−308), 325−334 (371−380), 338−349 (383−394), the HpcE residue numbers being given in brackets. The rms displacement is 0.65 Å, and the maximum displacement is 1.6 Å. HpcE protein is shown in blue and its calcium ion can be seen as an isolated sphere. FAH has its nitrogen atoms colored blue, oxygen is red, carbon is yellow, and calcium is orange. The calcium binding site is extremely similar in both proteins except that one carboxyl group in FAH is missing in HpcE, which has alanine at this position.

consistent with the presence of calcium in the active site of FAH, which also breaks a carbon−carbon bond, although the chemistry involved appears to be rather different (*6*). FAH uses a histidine and a glutamate, the latter bound to the calcium ion, to position a water molecule that attacks the substrate nucleophilically. As noted by Timm and co-workers (*6*), calcium is more commonly a structural than catalytic component of proteins, and it may well serve simply

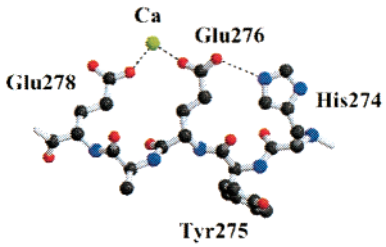

FIGURE 8: The position of His 274. The imidazole of His 274 and the glutamate side-chains of residues 276 and 278 lie roughly in a plane. All three residues lie on the third β strand of sheet I in the C domain. The histidine side-chain is close enough to form a salt-bridge with the carboxyl of Glu 276, indicated by a dashed line. Rotating the imidazole group through 180° about χ_2 would allow a hydrogen bond to form between Nϵ and the carbonyl oxygen of Tyr 275 instead. Neither of these atoms forms a hydrogen bond in the conformation shown. The calcium ion is shown as a green sphere.

to hold the active site residues in place. Both HpcE and FAH use a pair of glutamate residues on a β strand separated by one residue as ligands for the metal ion (Figure 8), but FAH uses four carboxyl groups, and HpcE uses only three. The positions of these three glutamate residues and the metal ion are extremely well conserved between the two proteins (Figure 7b). Like the N domain, the C domain conserves the FAH residue Thr 350 (as Thr 395) and replaces FAH Tyr 159 with a phenylalanine (HpcE Phe 250). These residues lie close to the acetoacetate binding site in FAH, but a three-residue insertion at residue 229 causes a bulge of residues in HpcE that partially blocks its binding site. The fumarate binding site is not as well conserved. Leu 230 and Leu 314 make a rather hydrophobic pocket in HpcE, although Arg 314 is close-by.

The imidazole side-chain of His 274 may play some role in catalysis. As modeled, the histidine Nϵ atom hydrogen bonds to Glu 276, a ligand of the calcium ion (Figure 8). Neither the Nδ atom nor the carbonyl oxygen of Tyr 275 form hydrogen bonds, sitting in a highly apolar pocket excluded from water. Rotation of the imidazole group through 180° about χ_2 would bring these atoms within 3 Å of each other, and place Nϵ the same distance from a water molecule, but break the bond with Glu 276. His 274 and Glu 276 may therefore play important roles in protonating the substrate to assist departure of carbon dioxide, but mutagenesis and substrate soaking studies will be needed to confirm this.

Although Tyr 232 and Glu 246 are clearly visible in the electron density map, the residues between cannot be seen even at very low contour levels (0.5 σ). This surface loop stretches just over 13 Å across two hydrophobic residues (Ile 369 and Phe 370), which are well-ordered and visible. The residues missing from the model are distant from the active site and presumably play no direct role in catalysis.

Comparison with Other Isomerases. Although the catechol pathway has enzymes equivalent to 5-carboxymethyl-2-hydroxymuconate isomerase (CHMI) and OPET decarboxylase of the HPC pathway, it has no equivalent to HHDD isomerase. Several enzymes are known that catalyze the interconversion of keto and enol forms. In some cases, they are required to give the keto form, but in others the enol form is the substrate of the successive enzyme. Over 230 structures of isomerases are present in PDB, 62 of these being xylose isomerase and 38 triosephosphate isomerase. Not surprisingly, given the range of substrates, the remainder are quite diverse, although structural similarity is not a guarantee of catalytic similarity. Xylose isomerase and triose phosphate isomerase are both barrel structures, but the former uses hydride transfer and the latter uses an ene-diol intermediate (*10*, *11*). A number of isomerases have an absolute requirement for metal ions, but some do not. Relatively few enzymes catalyzing a simple keto−enol transformation have been solved. Of these, two enzymes also involved in processing *meta* cleavage products of aromatic compounds are most relevant to HpcE, CHMI, and 4-oxalocrotonate tautomerase (4-OT).

CHMI is the enzyme in the HPC pathway immediately upstream from HpcE, a homodimer with 126-amino acid subunits. 4-OT is a hexameric enzyme with identical subunits of only 62 amino acid residues, the smallest enzyme subunit known. The two enzymes catalyze similar reactions with similar substrates (and will process each other's substrates slowly). The crystal structures of these enzymes unexpectedly showed that they have very similar overall structures, the 4-OT dimer adopting a very similar fold to the CHMI monomer, despite the lack of any sequence similarity between the two (*4*). Although substrate soaks were not possible with CHMI due to the high concentrations of sulfate used for crystallization, the active site could be identified by the positions of ordered sulfate ions and comparison with the 4-OT structure. Both enzymes use the N-terminal proline as a catalytic base. The very different nature of CHMI and HpcE precludes a simple comparison of the active sites. It is therefore not possible at present to determine the reasons for the different substrate specificity of the two proteins. Since the HHDD isomerase activity works on a smaller substrate than CHM, it is however highly probable that steric hindrance is involved. As in the case of other isomerases such as phosphomannose isomerase (*12*), the structure of HpcE does not allow the catalytic residues or the orientation of bound substrate to be determined. More work is needed both to test the activity of mutant proteins and to make suitable substrate analogues.

Comparison with Decarboxylases. Despite the potential industrial interest in decarboxylases due to their carboligase activity, relatively few of these enzymes have been studied structurally. PDB currently contains 12 unique decarboxylase structures. A common requirement is an electron sink to accept the electron left on the substrate by the carboxyl group which departs as carbon dioxide. This electron sink is of often a cofactor such as pyridoxal phosphate, but several decarboxylases function without any cofactor. For example, Whitby and co-workers (*13*) solved the structure of Uro-D, an eight-stranded β-barrel containing enzyme involved in porphyrin synthesis. The active site could be identified from the fairly hydrophobic pocket with positively charged residues present to bind the carboxyl groups on the substrate, but the catalytic mechanism is unclear. It has been suggested that the pyrrole ring of the protonated substrate acts as the electron sink. Orotidine 5′ monophosphate decarboxylase (ODC) also has the TIM barrel fold, and the distinction of enhancing the rate of reaction faster than any other known enzyme. The structure and its mechanism have been studied by various groups. One group emphasized Jencks' "Circe effect" in the action of the enzyme, the carboxyl group being

destabilized by close approximation to a aspartate side-chain (*14*), while another group stressed the importance of an active site lysine residue that may stabilize a carbanion transition state intermediate (*15*). A third group suggests that ODC has a unique mechanism because the carbanion generated is located in an sp^2 orbital perpendicular to a π system, and favors a scheme in which substrate protonation and decarboxylation occur simultaneously (*16*). Despite its catalytic prowess, ODC does not have any structurally obvious catalytic features except a lysine and an aspartate. OPET decarboxylation catalyzed by HpcE also requires no pyridoxal or related cofactor. As compared to the reaction carried out by ODC, the chemistry is facilitated by the carbonyl on the substrate itself, which accepts the electron left behind by the departing carbon dioxide. HpcE may therefore achieve adequate catalysis of the relatively simple transformation simply by binding the substrate and destabilizing the carboxyl group to be removed, as has been proposed for ODC (*14*). The role of the metal ions is unclear. No other decarboxylase has metal ion dependence, and it may be that the calcium ions in the structure serve only to hold the active site residues in place.

CONCLUSION

Certain protein folds are known to be capable of supporting a variety of enzymic activities. One of the best known enzyme folds is the TIM barrel, which is adopted by enzymes from four main EC classes carrying out at least 15 different functions (*17*). No other fold is associated with so many functions. The incomplete barrel-like fold of HpcE and FAH is not related to true barrel structures since the strands lie in opposite directions, but it too can clearly carry out more than one function. Gene duplication has apparently occurred at some point in evolution to create the two copies of the domain in the polypeptide chain of HpcE. Since the isomerization reaction has a relatively low activation energy, it may be that this step originally proceeded without enzymic assistance, as in the TOL pathway. Duplication subsequently allowed a novel active site to evolve which catalyzed the reorganization of HHDD into a form readily decarboxylated. The fact that the apparent active site is found in the same place in each domain suggests the fold may be a useful scaffold into which different active sites may be engineered. Although a large number of the homologues remain to be studied, it seems that most HpcE-like proteins are involved in the catabolism of aromatic compounds; the ORF21 protein from *Rhodococcus erythropolis* for example was sequenced from a plasmid associated with biphenyl degradation (*18*). The HpcE fold appears therefore to be rather less catalytically multipotent than the TIM barrel, although a number of sequence matches have been labeled "isomerase/decarboxylase" or "isomerase" simply due to the known function of HpcE. The domains of HpcE and their alignment with similar sequences suggest the minimal domain size to be about 200 amino acids. The rather larger FAH is more decorated with surface loops than HpcE, although it is not clear the extra residues have any biological function.

At present, the nature of the interactions with substrate and the catalytic mechanism are unclear. Substrate soaks are required to identify the mode of binding, and site-directed mutagenesis is needed to determine which residues play a role in catalysis. Further studies of the metal dependence of the enzyme are also necessary to discover the role of metal ions in the decarboxylase activity.

Table 5: Derivative Datasets and Phasing Statistics

	Hg	Pt
resolution limits	15.0−2.6 Å	15.0−2.5 Å
no. measured reflns.	204 635	321 675
no. unique reflns.	52 639	63 117
completeness[a] (%)	93.3 (89.8)	99.7 (95.7)
Rmerge[a] (%)	8.2 (16.4)	6.0 (16.8)
$<I/\sigma(I)>$[a]	11.9 (5.1)	12.9 (6.3)
phasing power[b]	1.03/0.79	0.76/0.59

[a] Numbers given in brackets refer to the outer shell of data. [b] Phasing power is given for acentric and centric reflections. Phasing power = rms ($|F_H|/|F_{PH}-|F_P + F_H||$).

MATERIALS AND METHODS

Crystallization and Data Collection. The purification and crystallization of HpcE have been reported previously (*19*). The crystals were originally reported to be $P2_12_12_1$ with cell dimensions $a = 106.1$ Å, $b = 127.4$ Å, $c = 139.8$ Å. Crystals grown using the reported conditions (using PEG 6000 as precipitant) often grew as clusters of very thin crystals, and a native dataset of only 3.3 Å could be collected at Daresbury SRS station PX 9.5. Replacing PEG 6000 with monomethyl ether PEG 2000 and including 200 mM calcium chloride improved the crystal quality considerably. The final conditions used were 15% MME-PEG 2000, 200 mM $CaCl_2$, and 50 mM Tris-Cl, pH 8.0. Hanging drops were set up using 15 mg/mL HpcE, mixing the protein and reservoir solution in a 1:1 ratio. Crystals grew over 2−3 weeks at 20 °C. The Matthews coefficient (*20*) is 2.54 assuming four copies of the monomer in the asymmetric unit. These crystals are orthorhombic with cell dimensions $a = 104.0$ Å, $b = 126.1$ Å, $c = 138.2$ Å. (The shrinkage in all axes as compared to the earlier crystals may be related to the improved diffraction.) Absences were observed along all three axes, *h*00, 0*k*0, and 00*l*. However, the native Patterson map showed a strong peak at (0.5, 0.03, 0.0) indicating simple translational symmetry and making it impossible to decide whether the absences along the shortest axis indicated that the space group was $P2_12_12_1$, or were due to the translation and that the true space group was $P2_12_12$. The data were therefore reindexed to give $a = 126.1$ Å, $b = 138.2$ Å, $c = 104.0$ Å. Native data were collected at Station X11 of the DESY synchrotron using a 300 mm Mar image plate detector. A total of 198 754 reflections were measured using a single crystal. A total of 229 images were collected in two sweeps; 134 high-resolution images were collected using 0.7° oscillations, and a further 95 low resolution images (to 2.6 Å) using 1° oscillations. Data collection statistics are given in Table 1.

Phasing. Two isomorphous derivatives were prepared, using mercury acetate (soaking 2.5 h in a 2 mM solution) and potassium tetrachloroplatinate (by leaving crystals of K_2PtCl_4 next to a protein crystal and waiting 2 h). Both derivative data sets were collected at Daresbury SRS station PX 9.5. All data (native and derivative) were indexed and integrated with DENZO (*21*); subsequent data manipulation was carried out using the CCP4 suite (*8*). Data collection statistics for the derivative data sets are given in Table 5.

The difference Patterson maps showed clear peaks cor-

responding to mercury and platinum sites, but also had a large peak at (0.03, 0.0, 0.5), and similar features were found both on the sections $w = 0.0$ and $w = 0.5$. The heavy atom sites could be determined assuming the spacegroup to be either $P2_12_12$ or $P2_12_12_1$, the crystal packing being nearly identical in the two cases. These differ only in having a pair of molecules related by crystallographic 2-fold symmetry with an *x* shift of 0.0, or a pair related by noncrystallographic 2-fold symmetry with an *x* shift 0.03. Phasing was attempted with MLPHARE in both space-groups, yielding similar figure of merit statistics for both. Molecules could be identified in the density, allowing a single molecule to be masked out. The masks were edited in O (*22*), and then masked density was used for molecular replacement. This gave a higher correlation coefficent for spacegroup $P2_12_12_1$, so all further calculations used this choice. Noncrystallographic symmetry averaging was carried out using DM (*23*), initially with NCS operators derived from the heavy atom positions. Improved initial phases from SHARP (*24*) together with further 4-fold NCS averaging led to excellent maps in space-group $P2_12_12_1$. Phases were improved further and main-chain tracing was carried out with ARP/wARP (*25*). The identification of individual chains among the four copies of the protein in the asymmetric unit was performed manually using Xtalview (*26*), which was also used for all subsequent modeling.

Refinement. In early cycles of refinement, NCS restraints were applied to the four copies of the molecule in the asymmetric unit. A short region of the polypeptide missing from the model (residues 4−17) could be sequenced directly from the map, and other missing residues were added as the phases improved. Water molecules were added using ARP (*27*), and by hand. All refinement was carried out with REFMAC (*28*). The use of maximum likelihood methods allows all the measured data to be used despite the Rmerge rising above 50% in the outer shell. The high R factor for the final model reflects the noncrystallographic symmetry which gives rise to a class of systematically weak reflections. Such structures do not usually refine to low R factors. Low resolution data were included in the refinement, allowing a bulk solvent correction to be applied. Data processing and refinement statistics are shown in Table 1. Protein coordinates and native X-ray data have been deposited with PDB and assigned the accession code 1GTT.

ACKNOWLEDGMENT

We thank Prof. Sam-Yong Park for help in making the figures. We are grateful to Elizabeth Duke and Daresbury SRS for the use of beam-line PX 9.5 at very short notice, and DESY for the use of beamline X11. We also thank Prof. R. Cooper for advice and comments, and would like to dedicate this paper to him in the year of his retirement.

REFERENCES

1. Roper, D. I., and Cooper, R. A. (1993) *Eur. J. Biochem. 217*, 575−580.
2. Roper, D. I., Fawcett, T., and Cooper, R. A. (1993) *Mol. Gen. Genet. 237*, 241−250.
3. Prieto, M. A., Diaz, E., and Garcia, J. L. (1996) *J. Bacteriol. 178*, 111−120.
4. Subramanya, H. S., Roper, D. I., Dauter, Z., Dodson, E. J., Davies, G. J., Wilson, K. S., and Wigley, D. B. (1996) *Biochemistry 35*, 792−802.
5. Holm, L., and Sander, C. (1993) *J. Mol. Biol. 233*, 123−138.
6. Timm, D. E., Mueller, H., Bhanumoorthy, P., Harp, J. M., and Bunick, G. J. (1999) *Structure 7*, 1023−1033.
7. Aponte, J. L., Sega, G. A., Hauser, L. J., Dhar, M. S., Withrow, C. M., Carpenter, D. A., Rinchik, E. M., Culiat, C. T., and Johnson, D. K. (2001) *Proc. Natl. Acad. Sci. U.S.A.* 98, 641−645.
8. CCP4. *Acta Crystallogr. D50*, 760−763.
9. Garrido-Pertierra, A., and Cooper, R. A. (1981) *Eur. J. Biochem. 117*, 581−584.
10. Allen, K. N., Lavie, A., Farber, G.K., Glasfeld, A., Petsko, G. A., and Ringe, D. (1994) *Biochemistry 33*, 1481−1487.
11. Albery, W. J., and Knowles, J. R. (1976) *Biochemistry 15*, 5627−5631.
12. Cleasby, A., Wonacott, A., Skarzynski, T., Hubbard, R. E., Davies, G. J., Proudfoot, A. E. I., Bernard, A. R., Payton, M. A., and Wells, T. N. C. (1996) *Nat. Struct. Biol. 3*, 470−479.
13. Whitby, F. G., Phillips, J. D., Kushner, J. P., and Hill, C. P. (1998) *EMBO J. 17*, 2463−2471.
14. Wu, N., Mo, Y., Gao, J., and Pai, E. (2000) *Proc. Natl. Acad. Sci. U.S.A. 97*, 2017−2022.
15. Miller, B. G., Hassell, A. M., Wolfenden, R., Milburn, M. V., and Short, S. A. (2000) *Proc. Natl. Acad. Sci. U.S.A. 97*, 2011−2016.
16. Appleby, T. C., Kinsland, C., Begley, T. P., and Ealick, S. E. (2000) *Proc. Natl. Acad. Sci. U.S.A. 97*, 2005−2010.
17. Hegyi, H., and Gerstein, M. (1999) *J. Mol. Biol. 228*, 147−164.
18. Kosono, S., Maeda, M., Fuji, F., Arai, H., and Kudo, T. (2000) *Appl. Environ. Microbiol. 63*, 3282−3285.
19. Jervis, T. J., Roper, D. I., and Glover, I. D. (1996) *Acta Crystallogr. D52*, 1036−1038.
20. Matthews, B. W. (1968) *J. Mol. Biol. 33*, 491−497.
21. Otwinowski, Z. (1991) *Methods Enzymol. 276*, 307−326.
22. Jones, T. A., Zou, J.-H., Cowan, S. W., and Kjeldgaard, M. (1991) *Acta Crystallogr. A47*, 110−119.
23. Cowtan, K. D. (1994) *Joint CCP4, and ESF-EACBM Newsletter on Protein Crystallography 31*, 34−38.
24. de la Fortelle, E., and Bricogne, G. (1997) *Methods Enzymol. 276*, 472−494.
25. Perrakis, A., Morris, R. J. H., and Lamzin, V. S. (1999) *Nat. Struct. Biol. 6*, 458−463.
26. McRee, D. E. (1999) *J. Struct. Biol. 125*, 156−165.
27. Lamzin, V. S., and Wilson, K. S. (1993) *Acta Crystallogr. D49*, 129−149.
28. Murshudov, G. N., Vagin, A., and Dodson, E. J. (1997) *Acta Crystallogr. D53*, 240−255.
29. Kraulis, P. J. (1991) *J. Appl. Crystallogr. 24*, 946−950.
30. Merritt, E. A., and Murphy, M. E. P. (1994) *Acta Crystallogr. D50*, 869−873.

BI015717T