Acta Cryst. (1996). D**52**, 1036–1038

Crystallization and preliminary X-ray analysis of a bifunctional enzyme: HHDD isomerase/OPET decarboxylase from *Escherichia coli*

TRACY J. JERVIS,[a] DAVID I. ROPER[b] AND IAN D. GLOVER[a]* *at* [a]*Physics Department, Keele University, Keele, Staffordshire ST5 5BG, England, and* [b]*Chemistry Department, York University, Heslington, York YOI 5DD, England. E-mail: phaφ4@oc.keele.ac.uk*

(*Received 8 November 1995; accepted 7 March 1996*)

Abstract

A bifunctional enzyme, 2-hydroxyhepta-2,4,-diene-1,7-dioate isomerase/5-oxopent-3-ene-1,2,5-tricarboxylate decarboxylase from the homoprotocatechaute (HPC) degradative pathway of *Escherichia coli* has been crystallized, using polyethylene glycol as a precipitant. The enzyme, of molecular weight 44 514 Da forms crystals belonging to the orthorhombic space group $P2_12_12_1$ with cell dimensions $a = 106$, $b = 127$ and $c = 139$ Å. The crystals diffract to at least 2.2 Å resolution using synchrotron radiation. A complete native data set has been collected to 3.3 Å resolution. The Matthews number calculated for a single molecule in the asymmetric unit is outside the normally acceptable limits and the aggregation state of the molecules in the crystal was investigated using self-rotation function studies, the results show features which are consistent with a tetramer in the asymmetric unit, giving a V_m value of 2.7 Å^3 Da^{-1}.

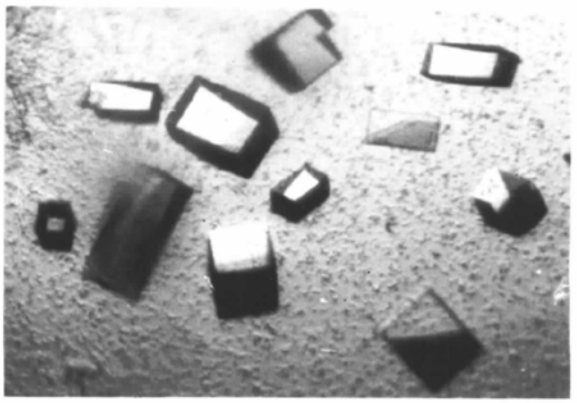
1. Introduction

The aromatic nucleus is one of the most abundant sources of carbon in the biosphere and there is considerable and growing interest in the pathways which bacteria have developed to exploit such a source of carbon as a metabolite. Bacterial catabolic pathways have evolved to metabolize inert plant substances such as alkaloids, terpenes, flavinoids and smaller aromatic compounds from the fungal degradation of lignin. The exploitation of such bacterial pathways has obvious value in the field of bioremediation as such pathways may be used in the biodegradation of man-made substances such as pesticides or petrochemicals. The catabolic pathways in *E. coli* are very similar to those in soil bacteria which would be exploited in any such application.

Homoprotocatechuic acid (HPC) is a product of the metabolism of tyrosine and phenylalanine and one of the aromatic compounds released in the degradation of lignin. The inducible *meta* cleavage of HPC is one of the longest catabolic pathways yet identified in *E. coli*. HPC degradation is achieved by a sequence of seven enzymatic reactions, producing the final products, succinic semialdehyde and pyruvic acid. The reaction sequence is initiated by extradiol (*meta*) ring cleavage with dioxygen and the ring-cleavage product is then oxidized, isomerized, decarboxylated, isomerized, hydrated and finally split by aldol fission to give the final products. The enzymes in this and other related pathways are the subject of X-ray crystallographic study and high-resolution structures for the first isomerase enzyme in the *E. coli* C HPC pathway and the isofunctional tautomerase of *Pseudomonas sp* C600 have been solved (D. I. Roper, unpublished work).

The HPC catabolic pathway bears considerable similarity to the catechol fission (TOL) pathway including the structural similarity of its intermediates especially with respect to OPET and HHDD. Although the TOL pathway contains an analogous decarboxylase enzyme activity, no activity equivalent to HHDD isomerase has been found. In *E. coli* C it has been shown that the OPET decarboxylase and HHDD isomerase activities are performed by the same magnesium-dependent protein (Roper & Cooper, 1993). The bifunctional enzyme, OPET decarboxylase/HHDD isomerase has a molecular mass of 44 514 Da from the derived amino-acid sequence and the enzyme is monomeric. Sequence comparisons show that the bifunctonal enzyme has no significant homology with other proteins, including the equivalent decarboxylase from the TOL pathway, except for two

(*a*)



(*b*)

Fig. 1. (*a*) Initial crystallization conditions yielding multiply twinned plates. (*b*) Refined crystallization conditions, as described in the text. The crystals shown have the approximate dimensions 0.2 × 0.3 × 0.2 mm.

predicted protein products with unknown functions from *A. eutrophus* and *Pseudomonas*. An analysis of the sequence of the protein itself does, however, show significant homology between the N- and C-terminal halves of the protein.

The structure determination of this molecule will, therefore, be important in a number of respects. Firstly, it will provide further information in the characterization of the aromatic degradative pathways exploited in bacteria. A detailed three-dimensional structure will also answer many intruiging evolutionary questions such as why this pathway exploits a bifunctional enzyme, and how the structure of this molecule is related to that of isofunctional enzymes in other pathways which show very limited sequence homology. The possibility of inter-active site channelling of substrates and products has been suggested but will only be confirmed with the availability of a detailed structure and elucidation of the spatial relationship of the catalytically active groups for both enzyme functions.

2. Methods

2.1. *Crystallization*

The protein was expressed in *E. coli* and purified as reported by Roper & Cooper (1993). The purified protein was concentrated to 20 mg ml^{-1}, using a Centricon YM-300 (Amicon) concentrator, in Tris buffer. Crystallization trials were carried out using the hanging-drop method at both room temperature and at 277 K using 24-well tissue-culture plates (Flow Laboratories). In all the crystallization trials 2 µl of protein solution was mixed with an equal volume of the reservoir solution prior to equilibration. Initial crystallization trials were undertaken using a variety of screening approaches including the incomplete factorial method (Carter & Carter, 1979) and sparse-matrix sampling (Jancarik & Kim, 1991). Crystals were initially grown between pH 7.0 and 8.0 in 50 m*M* HEPES buffer, using PEG 6000 and $MgCl_2$ as precipitants. Large irregular multiply twinned plates (up to 0.75 × 0.5 × 0.3 mm) grew within 1–2 d at room temperature (Fig. 1*a*). Attempts to overcome the problem of twinning have produced a number of conditions which give crystallographically identical crystals but with differing external morphologies. Twinning was eliminated by crushing the initial harvest of multiply twinned crystals and microseeding fresh hanging drops with crushed crystals and equilibrating over a reservior containing optimized concentrations of precipitant and salts.

Single crystals of the bifunctional enzyme were also grown at pH 8.0–8.5 (50 m*M* Tris buffer) using 41–45% monomethyl-ether PEG 750 and 200 m*M* $CaCl_2$, producing rhomboid plates which reached dimensions of 0.20 × 0.15 × 0.20 mm within 1 week (Fig. 1*b*). Crystals of similar external form and dimensions were also grown using similar concentrations of $SrCl_2$ in place of the calcium salt.

2.2. *X-ray crystallographic studies*

A crystal was mounted in a thin-walled glass capilliary in the normal manner. X-ray data were collected at room temperature on an 18 cm diameter MAR Research image-plate detector system on Station 9.5 (Thompson *et al.*, 1992) at the Daresbury SRS using an incident wavelength of 0.92 Å. The unit cell was determined using the autoindexing and parameter refinement program *REFIX* (Kabsch, 1988) and subsequently subjected to post-refinement in *MOSFLM*. A native data set was collected to 3.3 Å resolution using the oscillation method, employing φ ranges of 1–1.5 and exposure times of approximately 100 s per degree. The data were processed using the *MOSFLM* suite of programs (Leslie, 1992) with profile fitting of the reflections and post refinement of the unit-cell dimensions. The processed data were then scaled using the programs *ROTAVATA* and *AGROVATA* (Collaborative Computational Project, Number 4, 1994). The low-resolution data was then used in self-rotation function studies using the *CCP*4 program *POLARRFN* using data in the 10–5 Å shell.

Table 1. *Data-processing statistics*

D_{min} (Å)	$R(I)$	$\langle I \rangle$	$\langle I \rangle / \sigma$	N_{meas}	N_{ref}	% Poss	$I > 3\sigma(I)$
11.8	0.039	2912	13	961	372	88.8	96.0
8.7	0.036	3734	14	1724	668	94.5	96.8
7.2	0.039	2129	11	2072	795	93.3	96.4
6.3	0.039	1596	11	2496	942	94.4	94.5
5.7	0.046	1359	12	2922	1117	95.3	93.2
5.2	0.049	1730	10	3194	1241	95.6	92.2
4.8	0.048	2085	12	3272	1249	95.5	94.1
4.5	0.048	2569	11	3683	1408	94.2	95.3
4.2	0.051	2276	11	3950	1516	96.3	94.8
4.0	0.055	1864	10	4162	1569	94.6	91.7
3.85	0.061	1485	9	4206	1605	94.8	89.7
3.68	0.068	1218	8	4588	1713	95.0	88.0
3.54	0.077	1124	8	4814	1838	94.4	85.9
3.41	0.088	932	6	5030	1904	93.9	82.1
3.30	0.119	785	5	3435	1367	87.4	76.8
Total	0.056					93.8	89.8

3. Results

The largest crystals grown from PEG 6000 and 200 m*M* $CaCl_2$ were used in the preliminary crystallographic data collection. An examination of the recorded intensity distributions in the *h*0*l*, *hk*0 and 0*kl* zones (all possible absences to 3.4 Å included) indicated that the space group was $P2_12_12_1$. The crystallographic unit cell, derived from the 3.3 Å resolution data set discussed below, are a = 106.1, b = 127.4 and c = 139.8 Å. Data were clearly observable to at least 2.2 Å resolution.

The low-resolution native data set collected comprised of 132 618 observations and the final merged data set at 3.3 Å resolution, using only fully recorded reflections, comprised 27 163 independent reflections with an $R_{merge}(I)$ of 5.6% The data set was 94% complete, see Table 1, (completeness including partially recorded reflections is 99.3%, with a merging R of 6.8%).

The self-rotation function studies show only weak features but the most significant features are consistent with an independent tetramer in the asymmetric unit, giving a corresponding value for V_m (Matthews, 1968) of 2.7 Å^3 Da^{-1} and a solvent content of approximately 55%, both values well within acceptable values for protein crystals.

The structure determination of this interesting and important molecule is now underway. Because of the lack of significant homology with other proteins the structure solution will be carried out using MIR or MAD methods and screening for suitable heavy-atom derivatives is underway.

We are grateful to R. A. Cooper (Leicester University) for his help in establishing the collaboration between IDG and DIR. The CCLRC Daresbury Laboratory is thanked for the provision of synchrotron radiation beam time, and Keele University is

thanked for its support of TJJ through its University Studentship programme.

References

Carter, C. W. Jr & Carter, C. W. (1979). *J. Biol. Chem.* **254**, 12219–12223.

Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* D**50**, 760–763

Dagley, S. (1975). *Essays Biochem.* **11**, 81–138.

Jancarik, J. & Kim, S.-H. (1991). *J. Appl. Cryst.* **24**, 409–411.

Kabsch, W. (1988). *J. Appl. Cryst.* **21**, 67–71.

Leslie, A. G. W. (1992). In *CCP4 ESF-EACMB Newsletter for Protein Crystallography* No. 26. Warrington: Daresbury Laboratory.

Matthews, B. W. (1968). *J. Mol. Biol.* **33**, 491–497.

Roper, D. I. & Cooper, R. A. (1993). *Eur. J. Biochem.* **217**, 575–580.

Thompson, A. W., Habash, J., Harrop, S., Helliwell, J. R., Nave, C., Atkinson, P., Hasnain, S. S., Glover, I. D., Moore, P. R., Harris, N., Kinder, S. & Buffey, S., (1992). *Rev. Sci. Instrum.* **63**, 1062–1064.