**Supplementary Information**

**Methods**

Production of C23O64

The gene encoding *C23O64* was cloned into a pET-21a expression vector and transformed into *E.* *coli* BL21 (DE3) cells for protein expression (Table 1). The cell cultures were grown at 37 °C to an OD of 0.4-0.6 at 600 nm, followed by the induction with 100 µM isopropyl-β-D-1-thiogalactopyranoside (IPTG). Cells were then grown for 12 hours at 28 °C and harvested by centrifugation at 4000g for 30 min at 4 °C. The cell pellet was resuspended in buffer A (50 mM Tris-HCl pH 7.5, 50 mM NaCl, 0.5 mM DTT). Cells were lysed by ultrasonication for 15 min., the cell lysate was centrifuged at 25000 g for 30 min at 4 °C to remove cell debris, and the supernatant was collected.

All purification steps were carried out at 4 °C, and SDS-PAGE was performed to identify the protein of interest. Purification of the enzyme was achieved by ammonium sulphate precipitation (0-40 %). The precipitate was collected by centrifugation and dissolved in a minimum volume (~10 mL) of buffer A. Residual salt removal was achieved through dialysis against buffer A. Anion exchange chromatography was carried out with Q-FF (16/60) (GE Healthcare), pre-equilibrated with buffer A. The protein was eluted at 20-40% step-gradient with buffer B (50 mM Tris-HCl pH 7.5, 1 M NaCl, 0.5 mM DTT). The active fractions were pooled and concentrated using Amicon ultra-centrifugal filters (30 K MW, Merck Millipore). In the final step, size exclusion chromatography was performed on an S200 column (16/60, prep; GE Healthcare) pre-equilibrated with buffer A. The protein fractions were pooled and concentrated. These were kept as glycerol stock at -20 °C. Protein quantification at each step was determined by Bradford assay using bovine serum albumin as standard. The glycerol stock was diluted and used for crystallization.

Table S1: Catechol 2,3-dioxygenase from *Diaphorobacter* sp DS2, C23O64 production information

|  |  |
| --- | --- |
| Source organism | Diaphorobacter sp. strain DS2 |
| DNA source | **Genomic DNA** |
| Forward primer | **TCTGCAGGATCCATGGGCGTACTACGCATA** |
| Reverse primer | **TCTGCAAAGCTTTCAGGTATAGACGTCCGT** |
| Cloning vector | **TA vector** |
| Expression vector | **pET-21a** |
| Expression host | ***E. coli* BL21-DE3** |
| The complete amino acid sequence of the construct produced | **MGVLRIGHASLKVMDMDAAVRHYENVLGMKTT MKDKAGNVYLKCWDEWDKYSVILTPSDQAGMN HLAYKVEKEADLEALQQKIEAWGVKTTMLDEGT LPSTGRMLQFKLPSGHEMRLYASKEFVGTDVGNI NPDPWPDGLKGAGAHWLDHCLLVCEMNPEAGIN TVADNTRFVTECLDFFLTEQVLVGPGGSIQATTFL ARTTTPHDIAFVGGPTSGLHHIAFFLDSWHDVLKA ADVMAKNKVRIDVAPTRHGITRGETIYFFDPSGNR NETFAGLGYLAQRDRPVTTWTEDQLGSAIFYHTG**  **YLEPSFTDVYT** |

† Restrictions sites incorporated in the primers have been underlined

Crystallization

The purified C23O64 protein was concentrated to 22 mg ml-1 in buffer A for crystallization. A nanolitre-dispensing robot (Mosquito®, TTP Labtech) was used to set up hanging drop crystallization trials using commercial sparse matrix crystallization screens. The crystallization screens were from Hampton Research (Crystal Screen 1 and Crystal Screen 2), Rigaku Reagents (Wizard 1 and 2), and Qiagen (The Classic and Classic Suite II). The crystallization drops contained 350 nl protein solution and 350 nl reservoir solution with 100 µl in the reservoir. Among the few conditions that gave good diamond-shaped crystals, the crystals from the well containing 0.2 M MgCl2, 0.1 M HEPES sodium salt pH-7.5, 30% (v/v) PEG-400 were chosen and used for soaking with the natural and analogous substrates (Table 2).

****

**Fig S1**- Showing the crystals of catechol 2,3-dioxygenase from *Diaphorobacter* sp strain DS2 from the condition 0.2 M MgCl2, 0.1 M HEPES sodium salt pH-7.5, 30% (v/v) PEG-400 and the diffraction pattern of the crystal.

The crystals were soaked in a 20% cryoprotectant (PEG-400) solution followed by additional soaking in a 5 mM ligand in cryoprotectant. The crystals were flash cooled and stored in liquid nitrogen for X-ray diffraction.

Table S2: Crystallization condition for C23O64

|  |  |
| --- | --- |
| Method | Vapour diffusion, hanging drop |
| Plate type | **96-well** |
| Temperature (K) | **291** |
| Protein concentration | **22 mg ml-1** |
| Buffer composition of protein solution | **50 mM tris buffer pH-7.5, 50 mM NaCl, 0.5 mM DTT** |
| Composition of reservoir solution | **0.2 M MgCl2, 0.1 M HEPES sodium salt pH-7.5, 30% (v/v) PEG-400** |
| Volume of drop | **700 nl** |
| Volume of reservoir | **100 µL** |

Data collection and processing

X-ray diffraction data were collected from crystals of ligand-free and ligand-bound C23O64 at the ALS Beamline 5.0.2 Berkeley USA at 100 K at a wavelength of 1.01 Å. The indexing and integration of the diffraction data were performed using XDS ([Kabsch, 2010](#_ENREF_7)). The data reduction and scaling were achieved using AIMLESS ([Evans & Murshudov, 2013](#_ENREF_6)) from CCP4i suite ([Winn et al., 2011](#_ENREF_12)). Phenix.xtriage from PHENIX suite ([Adams et al., 2010](#_ENREF_1)) was then used to analyze the resulting intensity data. Matthews\_coeff from the CCP4i suite ([Matthews, 1968](#_ENREF_9)) ([Kantardjieff & Rupp, 2003](#_ENREF_8)) provided an estimate of the number of molecules in the asymmetric unit. The summary is shown in Table 3.

Table S3: Data collection and processing

Values for the outer shell are given in parentheses.

|  |  |  |  |
| --- | --- | --- | --- |
|  | Substrate free | 3-fluorocatechol bound | 4-methylcatechol bound |
| Diffraction source | **ALS beamline 5.0.1 Berkeley** | **ALS beamline 5.0.1 Berkeley** | **ALS beamline 5.0.1 Berkeley** |
| Wavelength (Å) | **1.005** | **1.005** | **1.005** |
| Temperature (K) | **100** | **100** | **100** |
| Detector | **CCD** | **CCD** | **CCD** |
| Crystal-detector distance (mm) | **425.0** | **400.0** | **400.0** |
| Rotation range per image (°) | **0.5** | **0.5** | **0.5** |
| Space group | **P 62 2 2** | **P 62 2 2** | **P 62 2 2** |
| a, b, c (Å) | **102.45 102.45 113.34** | **102.28 102.28 113.85** | **101.89 101.89 114.36** |
| α, β, γ (°) | **90 90 120** | **90 90 120** | **90 90 120** |
| Resolution range (Å) | **46.7-2.4(2.48- 2.4)** | **47.9-2.2 (2.27-2.2)** | **46.5-2.4(2.48-2.4)** |
| Total No. of reflections | **267657 (25464)** | **334861 (32625)** | **256653 (25511)** |
| No. of unique reflections | **14296 (1387)** | **18445 (1798)** | **14277 (1402)** |
| Completeness (%) | **99.88 (100.00)** | **98.89 (89.27)** | **99.81 (100.00)** |
| Redundancy | **18.7 (18.3)** | **18.2 (18.1)** | **18.0 (18.2)** |
| 〈 I/σ(I)〉 | **26.4 (3.7)** | **31.3 (5.6)** | **22.5 (4.7)** |
| Rp.i.m. | **0.09 (0.77)** | **0.07 (0.77)** | **0.09 (0.68)** |
| Overall B factor from Wilson plot (Å2) | **49.4** | **40.3** | **45.8** |

Structure solution and refinement

A monomer of Lap B from *Pseudomonas* sp KL28 (PDB entry-3hpv; ([Cho, Jung, Lee, & Rhee, 2009](#_ENREF_4)) was used as a search model for molecular replacement with *PHASER*© ([McCoy et al., 2007](#_ENREF_10)) in the *PHENIX* suite for ligand-free C23O64. *LapB* had a sequence identity of 42% with the *C23O64*. The Phenix.autobuild ([Terwilliger et al., 2008](#_ENREF_11)) was used for the initial model building. Phenix.refine ([Afonine et al., 2012](#_ENREF_2)) was used for structure refinement. Coot ([Emsley, Lohkamp, Scottc, & Cowtan, 2010](#_ENREF_5)) Was used for iterative model building in the maps and with alternate cycles of refinement and per residue analysis. Water and ligand molecules were added via Coot and then manually fit into the electron density. The refinement statistics are summarized in Table 4.

Sequence comparison based on structural alignment

T-Coffee Expresso ([Armougom et al., 2006](#_ENREF_3)) server performed the structure-based sequence alignment of several type 1 extradiol dioxygenases, and a figure for the same was generated using ESPript 3. The *C23O64* gene sequence identities with metapyrocatechase from *P*. *putida*, LapB from *Pseudomonas* sp. KL28, HPCD from *B*. *fuscum*, HPCD from A. *globiformis,* and 2,3-DHBN from *Rhodococcus* sp. were observed.

The three crystals of C23O64 were submitted to PDB and their id for the substrate free crystal, 4-methylcatechol bound, and 3-fluorocatechol bound crystals were 5zsz, 5znh, 5zsx respectively.

**Active site comparison**

The active site of the three crystal structures were super-imposed using pymol software. The comparison shown significant movement of the residues from substrate free to substrate bound structure.

**Tetrameric assembly of C23O64 protein**

****

**Fig S2-** Showing tetrameric assembly of 3-fluorocatechol bound C23O64 protein the orientation of the substrate shows their orientation with respect to each other. This figure was made using pymol.



**Fig S3**- Showing active site pocket of purple: substrate free, green: 4-methylcatechol bound and cyan: 3-fluorocatechol bound crystal structure of C23O64. The figure was made using pymol.

Adams, P. D., Afonine, P. V., Bunko´czi, G. b., Chen, V. B., Davis, I. W., Echols, N., . . . Zwarta, P. H. (2010). PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Cryst. D, D66* 213–221.

Afonine, P. V., W, R., Grosse-Kunstleve, Echols, N., Headd, J. J., Moriarty, N. W., . . . Adams, P. D. (2012). Towards automated crystallographic structure refinement with phenix.refine. *Acta Cryst. D, D68*, 352-367.

Armougom, F., Moretti, S., Poirot, O., Audic, S., Dumas, P., Schaeli, B., . . . Notredame, C. (2006). Expresso: automatic incorporation of structural information in multiple sequence alignments using 3D-Coffee. *Nucleic Acids Research, 34*, W604–W608.

Cho, J.-H., Jung, D.-K., Lee, K., & Rhee, S. (2009). Crystal Structure and Functional Analysis of the Extradiol Dioxygenase LapB from a Long-chain Alkylphenol Degradation Pathway in Pseudomonas. *J Biol Chem., 284*, 34321-34330.

Emsley, P., Lohkamp, B., Scottc, W. G., & Cowtan, K. (2010). Features and development of Coot. *Acta Cryst. D, 66*, 486–501.

Evans, P. R., & Murshudov, G. N. (2013). How good are my data and what is the resolution? *Acta Cryst. D, 69*, 1204–1214.

Kabsch, W. (2010). Integration, scaling, space-group assignment and post-refinement. *Acta Cryst. D, 66*, 133–144.

Kantardjieff, K. A., & Rupp, B. (2003). Matthews coefficient probabilities: Improved estimates for unit cell contents of proteins, DNA, and protein–nucleic acid complex crystals. *Protein Science, 12*, 1865–1871.

Matthews, B. W. (1968). Solvent Content of Protein Crystals. *J. Mol. Biol., 33*, 491-497.

McCoy, A. J., W, R., Grosse-Kunstleve, Adams, P. D., Winn, M. D., Storonia, L. C., & Read, R. J. (2007). Phaser crystallographic software. *J. Appl. Cryst., 40*, 658–674.

Terwilliger, T. C., W., R., Grosse-Kunstleve, Afonine, P. V., Moriarty, N. W., Zwart, P. H., . . . Adams, P. D. (2008). Iterative model building, structure refinement and density modification with the PHENIX AutoBuild wizard. *Acta Cryst. D, 64*, 61–69.

Winn, M. D., C.vBallard, C., Cowtan, K. D., Dodson, E. J., Emsley, P., Evans, P. R., . . . Wilson, K. S. (2011). Overview of the CCP4 suite and current developments. *Acta Cryst. D, D67*, 235–242.