Biochemical and structural characterization of an aromatic ring-hydroxylating dioxygenase for terephthalic acid catabolism

William M. Kincannon,[1] Michael Zahn,[2,4] Rita Clare,[1,4] Ari Romberg,[1] James Larson,[1,] Jessica Lusty-Beech,[1,4,] Brian Bothner,[1] Gregg T. Beckham,[3,4] John E. McGeehan,[2,4,*] Jennifer L. DuBois,[1,4,*]

1. Department of Biochemistry, Montana State University, Bozeman MT, United States

2. Centre for Enzyme Innovation, University of Portsmouth, Portsmouth, United Kingdom

3. Renewable Resources and Enabling Sciences Center, National Renewable Energy Laboratory, Golden CO United States

4. BOTTLE Consortium, Golden CO, United States

* Correspondence: john.mcgeehan@port.ac.uk; jennifer.dubois1@montana.edu

Classification: Biological Sciences, Applied Biological Sciences

Keywords: plastics recycling, upcycling, Rieske oxygenase, non-heme iron, aromatic catabolism; *Comamonas* sp. E6

Abstract: Several bacteria possess components of catabolic pathways for the synthetic polyester poly(ethylene terephthalate) (PET). These proceed by hydrolyzing the ester linkages of the polymer to its monomers, ethylene glycol and terephthalate (TPA), which are further converted into common metabolites. These pathways are crucial for genetically engineering microbes for PET upcycling, prompting interest in their fundamental biochemical and structural elucidation. TPA dioxygenase (TPADO) and its cognate reductase comprise a complex multi-metalloenzyme system that dihydroxylates TPA, activating it for enzymatic decarboxylation to yield protocatechuic acid (PCA). Here, we report structural, biochemical, and bioinformatic analyses of TPADO. Together, these data illustrate the remarkable adaptation of TPADO to the TPA dianion as its preferred substrate, with small, protonatable ring-2-carbon substituents being among the few permitted substrate modifications. TPADO is a Rieske [2Fe2S] and mononuclear non-heme iron-dependent oxygenase (RO) that shares low sequence similarity with most structurally characterized members of its family. Structural data show an α-helix associated histidine side chain that rotates into an Fe (II)-coordinating position following binding of the substrate into an adjacent pocket. TPA interactions with side chains in this pocket were not conserved in homologs with different substrate preferences. The binding mode of the less symmetric 2-hydroxy-TPA substrate, the observation that PCA is its oxygenation product, and the close relationship of the TPADO α subunit to that of anthranilate dioxygenase allowed us to propose a structure-based model for product formation. Future efforts to identify, evolve, or engineer TPADO variants with desirable properties will be enabled by the results described here.

Significance Statement: More than 400 million tons of plastic waste are produced each year, the overwhelming majority of which ends up in landfills. Bioconversion strategies aimed at plastics have emerged as important components of enabling a circular economy for synthetic plastics, especially those that exhibit chemically similar linkages to those found in nature, such as polyesters. The enzyme system described in this work is essential for mineralization of the xenobiotic components of PET in the biosphere. Our description of its structure and substrate preferences lay the groundwork for *in vivo* or *ex vivo* engineering of this system for PET upcycling.

Introduction

The discovery of a bacterium that assimilates poly(ethylene terephthalate) (PET) (1), the synthetic polymer used for clothing, single-use plastic bottles, and carpets has generated great excitement over the prospect of using biological catalysis for recycling this abundant waste product (2-5). The pathway for PET assimilation by *Ideonella sakaiensis* begins with a pair of secreted esterases that hydrolyze the polymer to its constituent monomers, ethylene glycol and terephthalate (TPA). Ethylene glycol is a natural product that is metabolized by multiple bacteria (6-8). TPA resembles plant-derived aromatic compounds, but it is not widely known as a substrate for bacterial growth. Because of its size and charge (-2 at pH 7), TPA must be actively transported into the cell where it is *cis*-dihydroxylated and dearomatized to yield 1,2-dihydroxy-3,5-cyclohexadiene-1,4-dicarboxylate (DCD) (9-15). The initial dihydroxylation is catalyzed by an O_2-dependent TPA dioxygenase (TPADO) working in conjunction with an NAD(P)H, flavin, and iron-sulfur-dependent reductase. A zinc-dependent dehydrogenase finally reductively decarboxylates DCD to produce protocatechuate (PCA) (**Fig. 1**) (1). Interest in using this pathway, either *in vitro* or engineered into microbes optimized for PET upcycling, is motivated by the societal, ecological, and economic benefits of plastic reclamation and recycling (5, 16, 17). Obtaining a structural and functional understanding of each of the four enzymes is an essential step toward these future applications.

Native systems for TPA import and *cis*-dihydroxylation have thus far been identified in several bacteria, including *I. sakaiensis* and several strains that do not use PET as a carbon source. The latter include *Comamonas* sp. strain E6 (10, 15), *C. testosteroni* T-2 (18), *C. testosteroni* YZW-D (11), *Delftia tsuruhatensis* T7 (19), *Rhodococcus sp.* DK17 (9), *Rhodococcus jostii* RHA1(12), *Pseudomonas umsongensis GO16* (6), and *Acinetobacter baylyi* ADP1 (TPA importer) (20). Several studies have introduced TPA catabolism into microbes for metabolic engineering applications as well (21-25).

The TPADO enzyme is a member of a large family of Rieske oxygenases (ROs). Several members of this family permit bacteria to aerobically assimilate, and thereby remediate, a wide range of environmental contaminants (26), such as naphthalene, pyrene, toluene, and chlorobenzoate. Several of these compounds resemble natural metabolites, including benzoate, cinnamate, picolinate, and salicylate, all of which are also RO substrates (27). Naphthalene dioxygenase (NDO), a paradigmatic RO (28-30), catalyzes stereo- and regio-selective dihydroxylations on a range of aromatic substrates. Other family members likewise catalyze an array of mono- and dihydroxylations, *O*- and *N*-dealkylations (27), desaturations, and sulfoxidations, where several prior studies have suggested a structural basis for reaction type or substrate preference (26, 31). ROs possess a family-defining [2Fe-2S] Rieske cluster that delivers electrons sequentially to an adjacent mononuclear non-heme iron center, where O_2 is reductively activated. The ultimate electron source is NAD(P)H, which transfers a hydride to a flavin adenine dinucleotide (FAD) cofactor in a separate reductase enzyme. One to two additional [Fe-S] clusters serve as 1e- shuttles to the active site. These clusters are found in diverse protein domain architectures that have traditionally been used for subtyping ROs in types I through V (32). Among the type II family members, which include TPADO, sequence conservation can be surprisingly low (33). Additionally, until recently, there were no structurally characterized examples of type II enzymes. Predicting TPA-directed activity among type II ROs based solely on the primary sequence of the catalytic (α) subunit is consequently challenging, limiting efforts to prospect for TPADO homologs in sequence databases.

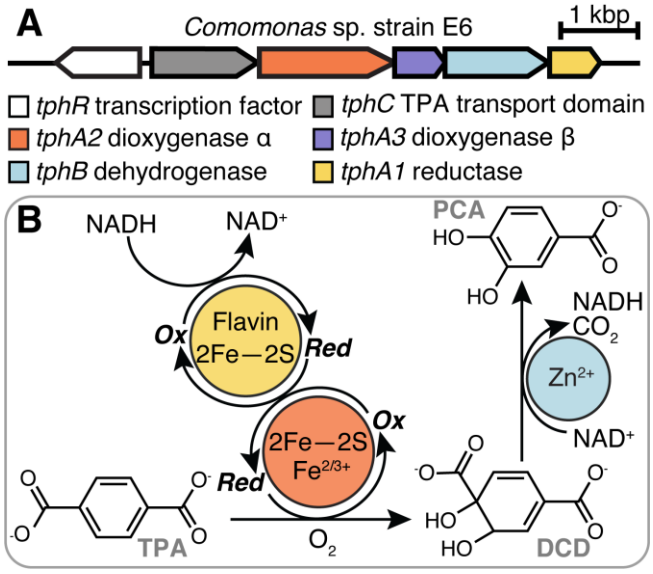


Figure 1. Enzymatic conversion of TPA to PCA. (A) Organization of the operon of genes encoding a TPA sensing transcription factor, TPA transporter domain, dioxygenase, reductase, and dehydrogenase. **(B)** TPADO (orange) uses molecular oxygen to dihydroxylate TPA, and reducing equivalents are supplied by the reductase (yellow). The product, DCD, is converted to protocatechuate via a Zn dependent dehydrogenase (blue).

In this work, we describe multiple crystal structures of the TPADO from *Comamonas* sp. strain E6 (15) in complex with both TPA and an *ortho*-substituted analog that links TPADO to salicylate 1-hydroxylase (34) and anthranilate 1,2-dioxygenase (35) in mechanism. Further, we define catalytic parameters and describe the specificity of TPADO for *para*-dicarboxylate anions that is well-explained by the binding mode observed in the structures. Finally, we show that TPADO has functionally significant sequence relationships with aryl-carboxylate directed ROs which, in conjunction with the structural data, can be used to propose a TPA-recognition sequence motif. The results establish this TPADO and its cognate reductase as a foundation for further protein and strain engineering efforts to achieve biological upcycling of PET plastic.

Results

A two-enzyme system dioxygenates TPA *in vitro*. TPADO, a 196 kDa complex of TphA2 and TphA3 subunits, and its cognate reductase (TphA1, 36.4 kDa) from *Comomonas sp.* strain E6 were heterologously expressed in *Escherichia coli* (*SI Appendix*, **Fig. S1-S2**; *Supplementary Methods*), with conditions then optimized to maximize both protein and cofactor yields as guided by studies from Ballou *et al.* (36). The α and β subunits of TPADO were co-expressed from a single IPTG-inducible pET-DUET vector, where the α subunit contained a C-terminal His_6 tag that was used for affinity purification at yields >20 mg/L culture. Native mass spectrometry identified the $\alpha_3\beta_3$ oligomer as the major component with the individual monomers as minor constituents (*SI Appendix*, **Fig. S3**). The monomeric reductase was purified via a C-terminal His_6 tag, with yields >10 mg/L culture.

UV/visible, atomic absorption, and electron paramagnetic spectroscopies confirmed the presence and type of [2Fe-2S] cluster in each protein (Rieske and plant types in TPADO and the reductase, respectively) (*SI Appendix*, **Fig. S4-S7**), based on their characteristic g-values and absorbance maxima. Combined metal and protein analyses predicted approximately stoichiometric occupancy of the expected cofactors in TPADO; however, this assumes that all of the protein (by mass) was present as intact $\alpha_3\beta_3$ oligomer, and that all iron was cofactor-associated. Consequently, the concentrations of TPADO used throughout this study, reported in terms of TPADO active sites (three per $\alpha_3\beta_3$ unit), are undoubtedly overestimated since at least some catalytically inactive α and β monomers and unpopulated cofactor binding sites were likely present.

TPA was converted by the TPADO/reductase system to an oxidized product specifically in the presence of NADH rather than NADPH. The oxygenation product could be resolved from the NAD^+ co-product by reverse phase high-performance liquid chromatography (HPLC) when subjected to a complex gradient (*SI Appendix*, *Supplementary Methods*). The distinctive UV/vis absorbance spectrum for the product (**Fig. 2A**) was consistent with absorbance maxima reported for DCD (13). However, this molecule is not commercially available, nor does extensive characterization exist in published literature. Here, we to chromatographically resolved the product from NAD^+ by HPLC, and obtained its UV/vis absorbance spectrum and mass spectral (MS) analyses. Under the ionization conditions used, DCD exhibited fragmentation consistent with the loss of CO_2 and/or H_2O (**Fig. 2B**). The expected exact mass of intact DCD is m/z = 199, while the predominant ion observed at the retention time of the product had an observed m/z = 137. TPA also displayed similar behavior, losing CO_2 during source ionization (*SI Appendix*, **Fig. S8**). Targeted fragmentation of the weak intensity, intact DCD and TPA peaks by LC-MS-MS yielded fragmentation profiles that were consistent with the observed source ionization fragmentation, supporting the definitive identification of the latter compound.

Steady-state kinetic parameters were measured as a function of variable [TPA] in ambient air (*SI Appendix*, **Fig. S9**). The data readily fit the Michaelis-Menten model, yielding k_{cat}(apparent) = 12 ± 0.3 min^{-1} and K_M[TPA](apparent) = 9.6 ± 1 μM. The parameters are *apparent* as saturating concentrations for NADH and O_2 have not been determined, and k_{cat} (maximal velocity x [enzyme]$^{-1}$) is likely underestimated. The k_{cat} and low-micromolar K_M are nonetheless similar to values measured for the TPADO homolog salicylate 5-monooxygenase NagGH from *Ralstonia* sp. strain U2 (k_{cat} = 7.81 min^{-1}, K_M = 22.4 μM) (37), suggesting that TPADO is well adapted to its substrate though slow in absolute terms. A total turnover number (TTN) of 704 (per apparent TPADO active site) was measured for the system via determination of unreacted TPA by HPLC when all other reactants were in excess. This TTN is low compared with several enzymes used in applied work, potentially due to over-counting of intact metallo-active sites here or to intrinsic instability of the enzymes. Either explanation suggests the need for engineered improvements to enzyme stability for future applications (38).

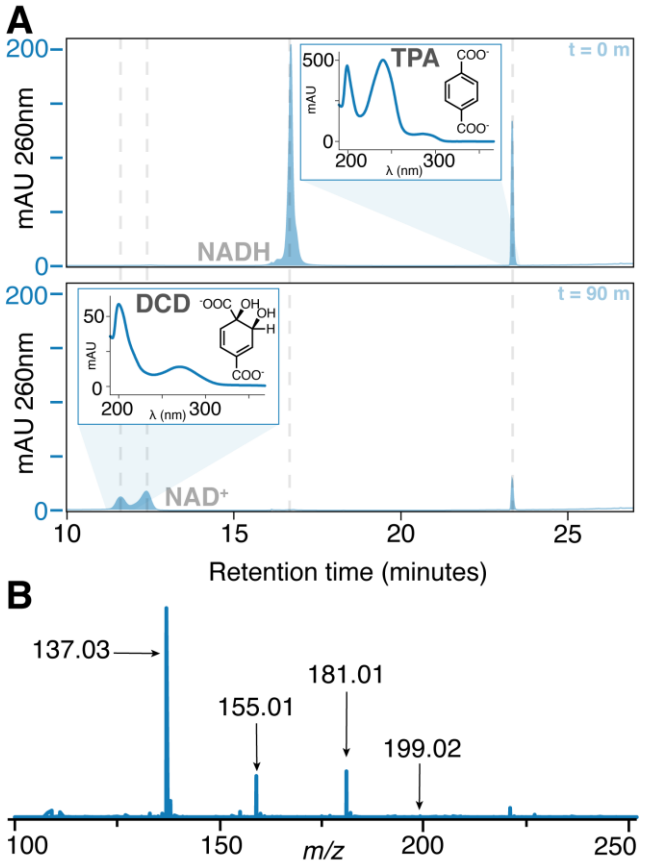


Figure 2. TPADO converts TPA to DCD. (**A**) HPLC chromatograms of a reaction mixture containing TPADO, its reductase, NADH, and TPA in air after 0 (top) and 90 min. The data illustrate disappearance of NADH with concomitant formation of new peaks at 12 min (assigned as DCD) and 13 min (NAD^+). Peak assignments were made based on identical retention times and UV/vis spectra (insets) relative to known standards (**Fig S10-S11**), except for DCD, for which no standards were available. In that case, the product was assigned based on (**B**) its mass spectrum, where the parent ion (m/z = 199.02) and fragments predicted to result from loss of water (m/z = 181.01), CO_2 (m/z = 155.01), or both (m/z = 137.03) were readily identified. (See also *SI Appendix*, **Fig. S8**.)

TPADO exhibits remarkable specificity for *para*-dicarboxylates. Many ROs are known either to accept multiple substrates (31) or to uncouple reductive O_2 activation from substrate oxygenation, expending NAD(P)H without oxygenating the organic substrate and releasing either H_2O_2 or water (*SI Appendix*, **Fig. S10**). The substrate or analog binds near to the mononuclear Fe(II), displacing water and opening a coordination position where O_2 can be reductively activated. The Fe/O_2 species can either productively oxygenate the substrate or, when an uncoupler is bound, break down to yield water or H_2O_2 (28, 39, 40).

The specificity of TPADO was assessed by analyzing substrate consumption at a fixed time point (90 min) after incubating TPADO/reductase with stoichiometrically limiting NADH and either TPA (**Fig. 2A**) or a structural analog (**Fig. 3**; *SI Appendix*, **Fig. S11-S15**). The stoichiometry of aromatic substrate as a function of NADH usage was measured by integrating their respective HPLC peaks, using a no-substrate/no-analog control sample as a baseline. Of a series of compounds screened, only TPA and closely related derivatives with hydroxyl- and amino- substituents at the ring C2 position were oxidized (**Fig. 3**). Percent coupling was calculated as: (moles substrate consumed) (moles NADH consumed above baseline)$^{-1}$ x 100%. Both TPA and 2-hydroxy TPA (2-OH-TPA) exhibited 100% coupling, while 2-amino TPA (2-NH_2-TPA) showed 22% coupling (**Fig. 3**).

Analogs with conservative substitutions to one of the carboxylates (4-nitrobenzoic acid (4-NBA), 4-carbamoylbenzoic acid (4-CBA), and 4-formylbenzoic acid (4-FBA)) yielded no detectable oxygenated products (**Fig. 3C**). 4-CBA and 4-FBA did not stimulate NADH oxidation above baseline (*SI Appendix*, **Fig. S13-S14**), while 4-NBA acted as an efficient uncoupler, promoting full consumption of NADH without reduction in the substrate analog signal (*SI Appendix*, **Fig. S15**). Taken together, these results demonstrate an extraordinarily high level of fidelity for TPA and very closely related diacids as principal substrates of the enzyme.

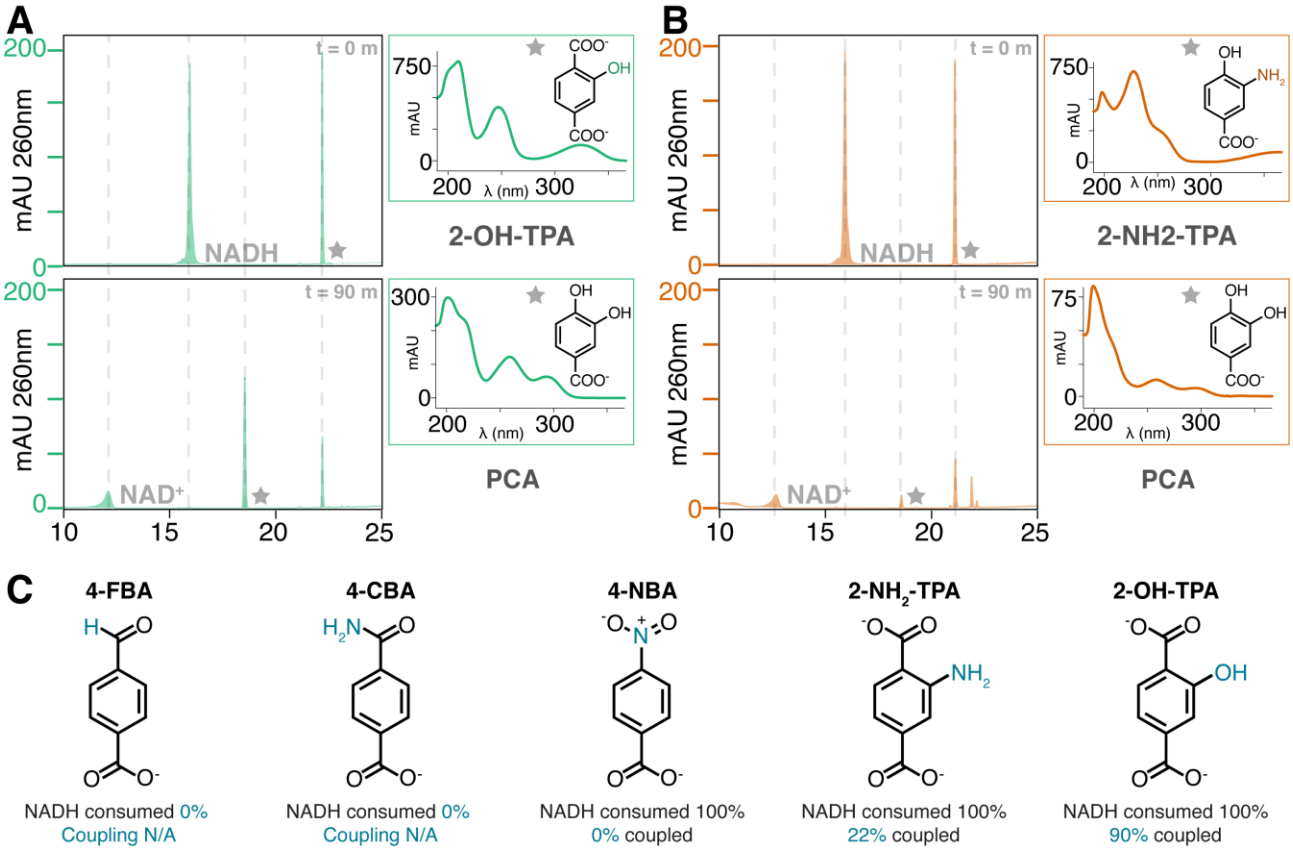


Figure 3. TPADO permits only structurally conservative substitutions on TPA analogs. Consumption of NADH and substrate analog following incubation in air with the TPADO/reductase system was measured via HPLC. (**A-B**) HPLC chromatograms measured at the outset (top) and 90-minute endpoint (bottom) of the reaction with 2-OH-TPA (**A**) and 2-NH_2-TPA (**B**). A star indicates the peak correlating to the adjacent UV/vis callout. Retention time and UV/vis spectra of the products for both the 2-OH-TPA and 2-NH_2-TPA reactions match those of PCA, though the latter achieved substantially lower coupling. The observed PCA product for both 2-substituted TPA analogs definitively indicates 1,2-dioxygenation. (**C**) Analogs 4-FBA (**1**) and 4-CBA (**2**) showed no NADH consumption, while 4-NBA (**3**) showed 100% consumption of NADH with no aromatic product formation (0% coupled). 2-NH_2-TPA (**4**) and 2-OH-TPA (**5**) showed 22% and 100% coupling, respectively.

TPADO converts 2-OH-TPA and 2-NH2-TPA to PCA. TPA analogs with -NH_2 or –OH substituents at the ring-2-position were converted to a product having a retention time and UV/vis absorbance spectrum identical to those obtained for a PCA analytical standard (CAS #99-50-3, **Fig. 3A-3B,** *SI Appendix* **Fig. S11B, Table S1**). 1,2-*Bis*-hydroxylation uniquely leads to an intermediate where decarboxylation and ring re-aromatization can proceed in conjunction with protonation and spontaneous loss of the ring-2 substituent (see Discussion). The observation of PCA as the product in both cases therefore indicates that hydroxylation must take place at the 1,2- (and not, for example, the 1,6-) ring carbons.

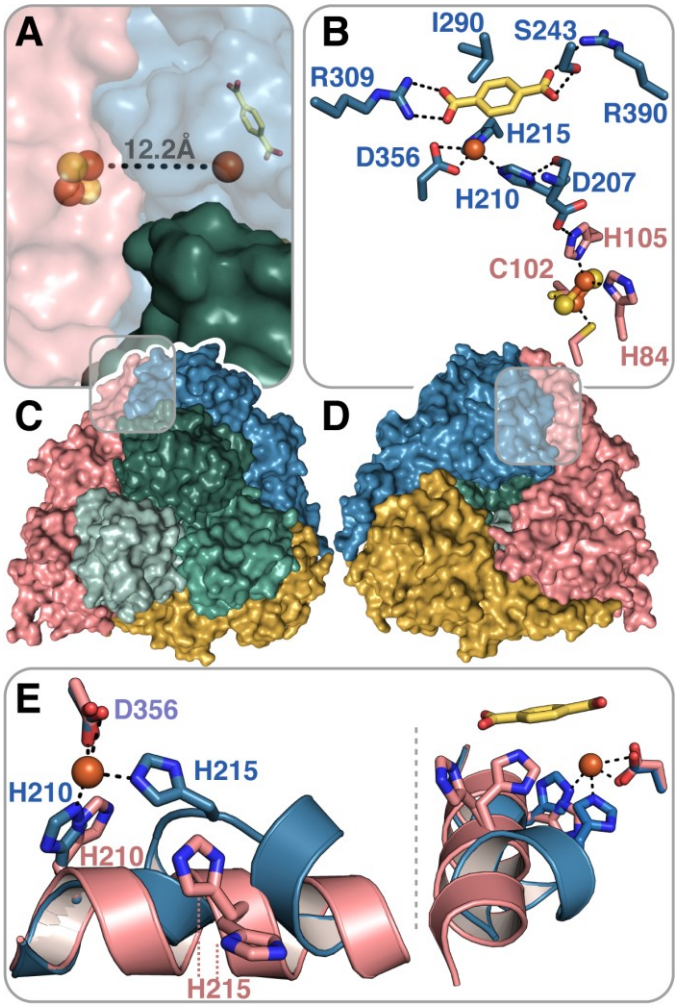
TPA dioxygenase has a canonical α3β3 structure. To understand the structural arbiters of substrate recognition, we determined the crystal structures of TPADO in the ligand-free state to 2.28 Å resolution and ligand bound structures with TPA to 2.08 Å and 2-OH-TPA to 1.95 Å resolution. During model building and refinement, clear continuous protein density was observed outside the α3β3 domains and identified as a single lysozyme molecule (*SI Appendix*, **Fig. S16**). In a rather fortuitous crystal packing, the lysozyme protein effectively acts as a crystallization chaperone that "glues" the TPADO molecules together (*SI Appendix*, **Fig. S17**), a function that, to our knowledge, has not been previously reported for lysozyme. More specifically, one lysozyme molecule forms interactions to five TPADO molecules and binds to both α- and β- subunits. TPADO forms an α3β3-heterohexamer in which three catalytic α subunits form a trimeric head-to-tail assembly atop a triad of non-catalytic β subunits (**Fig. 4A**). Like other ROs, a Rieske domain-containing a [2Fe-2S] cluster and a catalytic domain which comprises the TPA substrate and ferrous ion binding site (**Fig. 4B-C**) were located in each α-subunit (26). Residues H210, H215, and D356 coordinate the ferrous ion in the active site, which is 12.2 Å away from the [2Fe-2S] cluster of a neighboring α-subunit (**Fig. 4A**) and connected via the amino acid side chains H210, D207, and H105 (**Fig. 4B-D**). The [2Fe-2S] cluster within the same α-subunit is, by contrast, 42 Å away where it contributes to the neighboring reaction site.

The mononuclear-Fe-coordinating residues H210 and H215 are located on an α-helix spanning residues 208 to 220. In the substrate-free structure, H215 does not coordinate the ferrous ion because this helix is either disordered from residue H215 onwards in two out of three α-subunits or stabilized by crystal contacts with a symmetry related molecule; however, the helix conformation is the same within all three α-subunits (**Fig. S18**). Crystal soaking with the substrate TPA or 2-OH-TPA caused a partial unfolding of these helices except for the stabilized helix, which allows H215 to coordinate the ferrous ion (**Fig. 4E,** *SI Appendix*, **Fig. S18**) and suggests altered dynamics of the system on substrate binding.

TPA binding is supported by multiple ionic, hydrogen-bonding, and hydrophobic interactions. The substrate was located next to the mononuclear iron in an orientation confirmed by clear unbiased F_o-F_c difference electron density (*SI Appendix*, **Fig. S19**). One of the carboxylate groups from the bound TPA forms a salt bridge with R309, positioning the adjacent ring carbons 3.9 – 4.1 Å away from the reactive iron in an orientation that could promote reaction with an activated Fe/O_2 species. The side chain of I290 forms a hydrophobic-π-interaction with the aromatic ring of TPA, while the second carboxylate group forms a hydrogen bond with S243. Additionally, a potential salt bridge between R390 and this carboxylate is observed, but the distance varies among the α subunits between 2.9-4.6 Å. This apparent flexibility in the R390 side chain is reflected in its high B-factors and could have functional significance. For example, residue R390, as well as the α-helix mentioned above and subsequent residues, form the opening of a potential site for TPA entry or product release that may have a gate function (*SI Appendix*, **Fig. S20**). The TPA binding pocket is more open within the substrate-free structure because the α-helical residues from H215 onwards are disordered and not part of the model; in total, residues 215 to 227 are missing.

TPA exhibits a unique binding mode. A search in the current structural databases, utilizing the DALI protein comparison server (41), highlighted salicylate 5-monooxygenase (NagGH) (37, 42) as the closest structural homolog to TPADO with RMSD values of 1.5 Å and 1.3 Å, and sequence identities of 42% and 26%, for the α and β subunits, respectively. Furthermore, the α subunits of related toluene 2,3-dioxygenases, biphenyl dioxygenases, and

NDOs are significantly more divergent, with RMSD values ranging between 2.4–2.8 Å. A superposition of TPADO with NagGH illustrates their similar quaternary structures (*SI Appendix*, **Fig. S22**). While no substrate is present in the NagGH structure for comparison, it appears unlikely that salicylate 5-monooxygenase could bind TPA because the side chain of M257, which replaces the S243-TPA hydrogen-bonding interaction observed in TPADO, would interfere with the carboxylate group of TPA (**Fig. 5A**).



.**Figure 4. TPADO structure shows TPA binding interactions. (A, C)** Top view of the $\alpha_3\beta_3$-heterohexamer. The β subunits are colored in shades of green and the α subunits are colored in pink, yellow, and blue. Iron-sulfur cluster and mononuclear iron cofactors shown as spheres, exhibiting a head-to-tail inter-subunit pathway for electron transfer. **(B, D)** Bottom view shows the trimeric architecture of the α subunits. The bound TPA substrate is shown in yellow for the blue colored α subunit. Salt bridges with R309 and R390 and a hydrogen bonding interaction with S243 are shown. The ferrous ion, which is coordinated by the side chains of residues H210, H215, and D356, is connected to the Rieske-type [2Fe-2S] cluster of a neighboring α subunit (pink). **(E)** Superposition of the apo and ligand-bound active site structures shows the different conformations of an α-helix which contains the ferrous ion coordinating histidines H210 and H215. The pink α-helix is stabilized by crystal contacts, whereas the blue α-helix is kinked to allow the coordination of the ferrous ion by H215 when bound to TPA.

The salt bridge involving one of the TPA carboxylate groups is conserved in both TPADO (R309) and the salicylate 5-monooxygenase (R323) from *Ralstonia* sp.(**Fig. 5A**) (37), although an equivalent H-bond to R390 on the opposite end of the substrate is absent. Several additional Rieske dioxygenases possess a positively charged residue (R or K) at the same position as R309 in a pocket that is otherwise largely hydrophobic, but in those cases the homologous side chain points away from the substrate, as observed in the NDO structure (**Fig. 5B**) (43). Moreover, the residues involved in electron transfer from the Rieske cluster to the mononuclear iron superimpose well between NagGH and NDO (*SI Appendix*, **Fig. S21**).

The structure of TPADO bound to 2-OH-TPA is close to identical to the complex with TPA but surprisingly, the electron density for this ligand strongly suggests that the 2-hydroxy group is oriented towards the hydrophobic part of the active site and not towards the polar part where hydrogen bonding interactions with the carbonyl oxygen of V205 would appear possible (*SI Appendix*, **Fig. S22**). It is possible that this binding orientation of 2OH-TPA promotes catalysis by substrate destabilization, that it favors product release, or that it represents a non-productive binding mode. To investigate the latter possibility, we aligned the 2-OH-TPA bound structure with a product-bound NDO structure (PDB ID 1O7P) (29). From this alignment (**Fig. 5B-C**), the *cis*-dioxygenated carbons of the (*1R,2S*)-*cis*-1,2-dihydroxy-1,2-dihydronaphthalene product appear to align with the 1- and 2-carbons of 2-OH-TPA. These carbons are consequently implicated as the sites of hydroxylation based on the observed formation of PCA as the ultimate product, suggesting that 2-OH-TPA occupies a productive binding mode. This also suggests, by analogy, that the stereochemistry of product of the TPA reaction is predicted to be (*1S,2R*)-*cis*-1,2-dihydroxy-3,5-cyclohexadiene-1,4-dicarboxylate (DCD).

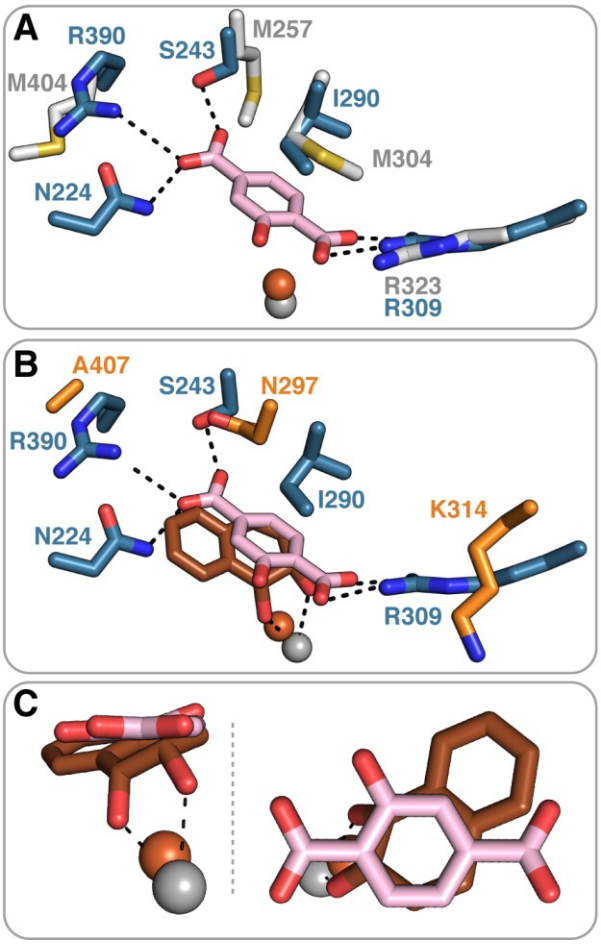


Figure 5. Active site structure of TPADO with 2-OH-TPA bound is shown superimposed on the structure of (A) NagGH (A) and (B, C) product-bound NDO. (A) Superposition of the blue colored TPADO α subunit with the α subunit of its closest structural homolog salicylate 5-monooxygenase NagGH (grey) which has no substrate bound (PDB ID 7C8Z) highlights the unique roles played by S243 and R390 in accommodating the substrate (37). (**B**) Superposition with the α subunit of NDO (orange) with its dioxygenated product bound (PDB ID 1O7P) (29). The lower label numbers belong to NDO. (**C**) Closer views of the overlay of NDO product (orange) with 2-OH-TPA substrate and their relative positions to the ferrous ion. Overlap of the 1,2-diol of portion of the NDO product with the ring-1,2 carbons of 2-OH-TPA suggests these are the sites of hydroxylation.

Sequence similarity network (SSN) analyses relate the TPADO α-subunit to a subgroup of ROs that hydroxylate hydrophilic monoaryls. Classic RO subtyping by Kweon *et al.* focused on the different domain organizations that mediate electron flow between the reductase and RO active site (32). A subsequent structure-based sequence alignment of 121 then-available catalytic α subunits was carried out by Eltis *et al.* (33). This analysis resulted in a first-

ever phylogenomic map and suggested how gene fusion events gave rise to the Rieske/mononuclear Fe α subunit.

Just over 45,000 α-subunit homologs are currently known, permitting subtyping of a much broader scope. Their sequences, members of Pfam PF00848, were submitted to an "all-by-all" comparison via the Enzyme Function Initiative's (EFI) network analysis algorithm (**Fig. 6**, *SI Appendix*, **Fig. S26**) (44). Consistent with earlier findings from Eltis *et al.*, we observed that a single subgroup of the massive and diverse family has been experimentally oversampled in prior work (cluster #2, **Fig. S26**). This node contains several aromatic/polyaromatic hydrocarbon hydroxylases of interest for bioremediation purposes, including NDO and biphenyl dioxygenase (BPO) (45).

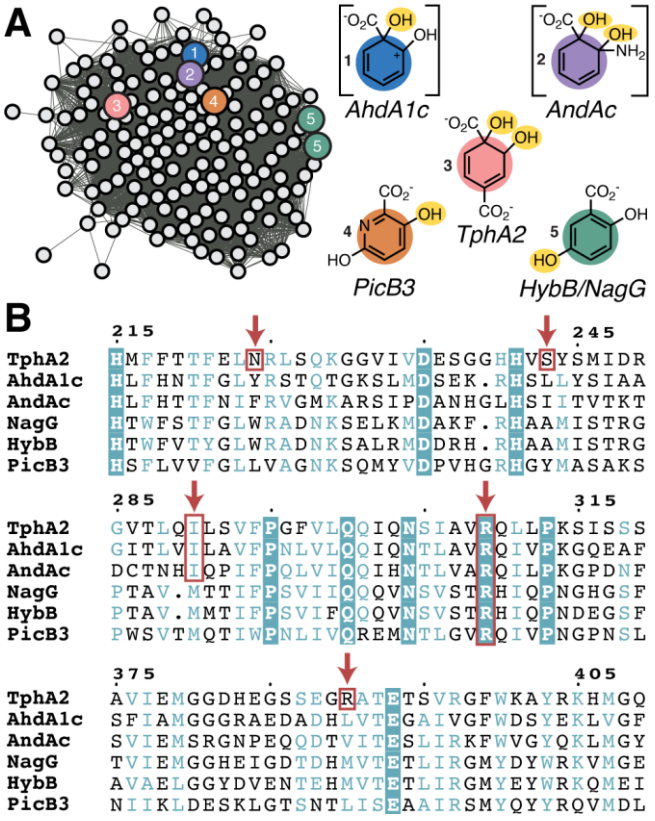


Figure 6. Sequence relationships of aryl carboxylate Rieske oxygenases. **(A)** A single cluster from a full SSN generated for Pfam00848 (*SI Appendix*, **Fig. S23**) containing 1,812 RO α subunit sequences, including TphA2 (α-TPADO) and others with characterized functions relating to aryl carboxylate oxygenation (indicated by circles with color), is shown. To the right of the cluster are structures illustrating where the enzymes are proposed to hydroxylate their aryl carboxylate substrates (O_2-derived hydroxyl groups highlighted in yellow). Intermediates are shown in brackets to clarify where hydroxyls are installed prior to their spontaneous breakdown to yield CO_2, H_2O, and catechol (AhdA1c) or CO_2, NH_3 (protonating to form NH_4^+ at neutral pH) and catechol (AndAc). **(B)** A multiple sequence alignment of the five functionally characterized homologs from (A). Orange boxes and arrows indicate residues proposed to interact with a substrate. R309 is highly conserved in these 5 sequences and across the cluster, suggesting it has a functionally significant role. N224, S243, and R390 are only found in TPADO. I290 is conserved in TphA2, AhdA1c, and AndAc.

Beyond this large subgroup, ROs with similar substrate and/or reaction types, where known, were grouped together in this analysis, suggestive of sequence-based family subdivision along functional lines (*SI Appendix*, **Fig. S23**). The cluster of sequences to which the α-subunit of TPADO (TphA2) belongs (**Fig. 6**) also contains NagG, the α-subunit of salicylate 5-hydroxylase, consistent with identification of the latter via the DALI search for TPADO structural relatives. Additionally, other functionally annotated members of the TPADO subfamily were all associated with aryl carboxylate substrates (anthranilate (35), picolinate (46), and salicylate) like TPA, although TPA is the only dicarboxylate of the group.

A sequence alignment of the α-subunits of these annotated members of the TphA2 sequence cluster showed that, of all the residues making direct contact with the substrate in the TPADO structure, only R309 is conserved. Key portions of this alignment are shown in **Fig. 6B**. As highlighted in **Fig. 5A**, the analogous arginine in NagG (R323) is proposed to form a salt bridge with the lone carboxylate of the substrate, positioning it for monooxygenation at the ring-5-carbon (yielding gentisate (47)). This active site arginine is also conserved in the α-subunits of hydroxypicolinate 3-monooxygenase (Pic3B) (46), anthranilate 1,2-dioxygenase (AndAc) (35), and salicylate 1-hydroxylase (AhdA1c; catechol product) (48, 49). A similar role in positioning the substrate via the carboxylate could be proposed for the conserved arginine in the other enzymes, though the sites of ring mono- or di-oxygenation vary (**Fig. 6A**). In TPADO, N224, S243, and R390 interact with the carboxylate at the non-reactive end of TPA but are not conserved across the subfamily. This observation suggests that they may be important for productively positioning dicarboxylate substrates.

Discussion

Plastic bioconversion using microbial enzymes depends on understanding and ultimately improving the properties of the responsible enzymes. Hydrolyzable ester linkages are ubiquitous in biology, and aromatic compounds bearing polar substituents are abundant in the metabolic pathways of diverse organisms. The presence of both ester linkages and an aromatic building block with a polar substituent in PET is consistent with this plastic constituting a readily available carbon source for bacterial consumption in the biosphere.

TPADO is an $\alpha_3\beta_3$ non-heme Rieske oxygenase which catalyzes the NADH-dependent dioxygenation of TPA, the aromatic subunit of PET. Like the better-studied cytochrome P450s, ROs catalyze a wide range of oxidations and oxygenations of diverse substrates, increasing their water solubility and activating them for further metabolism (50). However, unlike cytochromes P450, certain members of the RO family are capable, perhaps uniquely, of catalyzing the *cis*-dihydroxylation of an aryl ring. The reaction depends on the proper positioning of the substrate relative to an activated O_2 species that forms at a mononuclear iron center, where both O-atoms are poised to react on the same side of the plane defined by the substrate's aromatic ring (31). Dihydroxylation results in both the loss of aromaticity and the formation of two new chiral centers in the product, DCD. This product, in turn, is well-situated for further catabolism via an exergonic step in which CO_2 is produced, NADH regenerated, and the ring rearomatized to yield a valuable and versatile metabolite, PCA (51). This elegant metabolic arrangement, in which reductant is intrinsically recycled within the pathway, suggests that, in spite of the complexity of the TPADO/reductase system and ROs in general, this system offers a compelling starting point for engineering an efficient route for TPA bioconversion to PCA.

PET is an abundant, xenobiotic source of environmental TPA, suggesting that TPA might not be the primary but perhaps a secondary substrate of TPADO. We noted that the bacterial TPADO studied here was nonetheless highly efficient and extraordinarily specific for TPA (K_M = 9.6 μM, k_{cat}/K_M(TPA) = 2.1 $M^{-1}s^{-1}$), coupling NADH oxidation to TPA dihydroxylation with 100% fidelity and excluding a variety of structurally related compounds as potential substrates (**Fig. 3**). This observation is consistent with an early report on the *Comamonas* E6 TPADO indicating that neither the *iso*- nor the *ortho*-benzene dicarboxylate regioisomers of TPA (*para*-benzene dicarboxylate) were substrates for the enzyme (15). Two major exceptions identified here are 2-OH-TPA and, to a lesser extent, 2-NH_2-TPA, where TPADO catalyzed the conversion of each to PCA.

Each of the benzene dicarboxylate regioisomers is used in the production of plastics, whether as a repeating subunit in PET, as a synthetic precursor for monomers used in plastics, or as a noncovalently bound plasticizer (52). The *iso*- and *ortho*- benzene dicarboxylates (phthalates) serve one or both of the latter functions in several plastics. *Ortho*-phthalate has been of special concern as an endocrine disruptor with the potential to leach out of consumer products and into water (52). Bacteria that can degrade each of these compounds have been identified (53), indicating that enzymatic adaptation to these plastic-relevant compounds is not unique to one or a few strains. Recent work even suggests possible biogenic sources for and derivatives of phthalates (54), which could have helped drive RO diversification.

The *Comamonas* E6 strain has served as a paradigmatic plastic monomer degrader, capable of using each of the benzene dicarboxylate regioisomers as a sole source of carbon and energy (53), via separate operons encoding three distinct RO enzymes. While the TPADO has an α3β3 subunit structure, the *ortho*- phthalate dioxygenase from *Comamonas E6* displays α3α3 hexameric structure. It is possible that *iso*-PDO and *ortho*-PDOs from related strains have similar architecture. The strain E6 *iso*-/*ortho*- phthalate dioxygenase α subunits share 32% identity with one another, forming a separate lineage from the α subunits of TPADO and other α3β3 ROs with which they share ≤19% identity (55). Accordingly, the α subunits of TPADO and a recently reported *ortho*-phthalate dioxygenase structure superimpose poorly (*Comamonas*

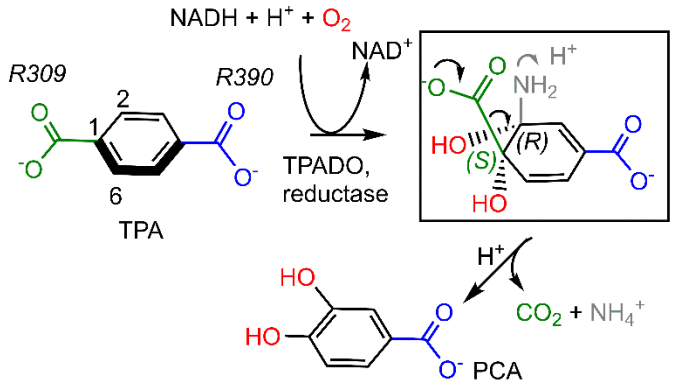
testosteroni KF1 phthalate dioxygenase (PDO), *SI Appendix*, **Fig. S23 (55)**). The active site of this PDO exhibited ionic/hydrogen-bonding interactions between the bound *ortho*-phthalate carboxylates and an Arg-Arg-Ser triad that is unconserved in the TPADO sequence or tertiary structure. Hydrophobic interactions were observed surrounding the benzylic ring. TPA was able to bind with some observed hydroxylation in the same PDO pocket, though with 80% uncoupling and a much lower k_{cat}/K_M than for *ortho*-phthalate. These results suggest that TPADO and PDO are each well-adapted to their preferred benzene dicarboxylate regioisomer.

A large-scale network analysis of >45,000 sequences of RO catalytic α-subunits was generated, aimed at understanding the origins of TPA bioconversion. Consistent with the unique homohexameric structures identified by the Eltis group, PDO α–subunits from *Comamonas* are sufficiently sequence-distant that they are not grouped with the same PFAM family (PF00848) and are not found in this SSN. Analysis of the network identified a subcluster containing the closest sequence relatives to TPADO (**Fig. 6,** *SI Appendix*, **S24**). While only a few members of the subcluster are characterized in the literature, they are all known to catalyze reactions with aryl-carboxylic acids which structurally resemble TPA. Many of the nodes in the cluster shown in **Fig. 6A** are annotated as similar to AhdA1c, the catalytic subunit of salicylate-1-monooxyenase (IPR043264). A multiple sequence alignment revealed a cluster-wide conserved arginine which forms a salt bridge to one of the carboxylate groups of TPA (R309, **Fig. 4**), and which is proposed to engage in a similar interaction with salicylate based on structural characterization of salicylate 5-monooxygenase (NagGH) (37). An additional set of residues (N224, S243, and R390) have side chains within hydrogen bonding distance of the second carboxylate in the TPA and 2-OH-TPA co-structures with TPADO (**Fig. 4-5**) described here. These are not conserved in the other functionally annotated sequences highlighted in **Fig. 6A-6B**. This conserved motif may therefore offer a means of identifying diverse TPADOs from sequence databases. Additional interesting findings resulting from the sequence similarity network (SSN) analysis cannot be discussed here at length. We are supplying the network file for readers to explore the vast sequence space. Many of the clusters in the SSN are comprised entirely of sequences with no known function, suggesting a plethora of catalytic diversity remains to be discovered.

A structure-based mechanism can be proposed to explain the observed products of the TPADO-catalyzed reactions with TPA, 2-OH-TPA, and 2-NH_2-TPA considering the data presented here and two additional observations. First, the network analysis revealed a close sequence relationship between the α subunits of TPADO and anthranilate dioxygenase from *Burkholderia cepacia* DBO1 (35). This organism can grow with anthranilic acid as a sole carbon source, where it is proposed that its RO functions to dihydroxylate anthranilate, which then spontaneously deaminates and decarboxylates to yield PCA (35). Second, extensive prior work has shown that NDO, perhaps the best-studied RO, can catalyze a variety of oxidations depending on the position of the substrate in the WT and variant enzymes relative to the site of O_2 activation. Aryl carbons that are closest to the mononuclear iron (29, 56) are generally prioritized for hydroxylation.

A mechanism (*SI Appendix*, **Fig. S10**), taking these observations and prior work with ROs into account, would proceed as follows. Substrate binding near to the mononuclear Fe in fully-reduced TPADO is expected to displace an iron-bound water molecule from TPADO (28, 39, 40). Here, we observed an unexpectedly large change in the conformation of a helix containing two Fe-ligating histidine residues in response to substrate binding (**Fig. 4C**), consistent with the role of the substrate in permitting binding and reductive activation of O_2. The initially formed ferric-η^1-superoxy intermediate was recently proposed to be the common reactive species in ROs (47), and for the sake of simplicity, we adopt this as the hypothetical intermediate here. The distal oxygen of this species would attack the nearest available substrate carbon (Fe-ring-C2 = 3.6 Å) to form the aryl radical-O-O-Fe adduct. Addition of a second active site electron from the Rieske cluster yields the ring-1,2-epoxide, adjacent to the mononuclear ferric-hydroxyl. Nucleophilic attack of the hydroxyl at ring-C1 forms the hydrogenated, dearomatized, *cis*-diol product.

Dioxygenation of the ring-1,2 carbons of 2-OH-TPA by this route would yield an unstable gem-diol at the ring-C2 position. This is expected to readily dehydrate and decarboxylate to yield PCA and CO_2 in the presence of an aqueous proton source to react with the hydroxyl leaving group. An analogous mechanism has been postulated for the close sequence relative of TPADO, anthranilate dioxygenase, in which the substrate, a 2-amino benzoate, is initially dihydroxylated and hydrogenated at the ring-1,2 carbons. Protonation of the C2-NH_2 group would catalyze the breakdown of the product to yield catechol, CO_2, and ammonia, which would rapidly acquire a proton to form ammonium cation under neutral, aqueous conditions. The dicarboxylate analog of anthranilate (2-NH_2-TPA) serves as a substrate of TPADO, with the corresponding deaminated product PCA, though the poorer amino leaving group leads to a lower level of productive turnover compared with 2-OH-TPA. The close relationship between TPADO and anthranilate dioxygenase, and the proximity of the reactive Fe(II) to the 1,2 carbons of TPA or 2-OH-TPA in the TPADO structures presented here (**Fig. 5C**) suggests an analogous route to DCD or PCA production in TPADO (**Scheme 1**).



Scheme 1. Hydroxylation of the ring-1,2 carbons of TPA yields DCD (box). The analogous intermediate generated from 2-NH_2-TPA is indicated with the amino group in grey. Arrows show the pathway for re-aromatization and loss of CO_2 + NH_4^+.

Together, these observations connect the experimentally determined binding interactions between TPADO and its substrates to a novel RO sequence subtype and a potential sequence motif specific for *para*-aryl-dicarboxylate substrates like TPA. These, in turn provide strong support for a proposed pathway for the TPADO catalyzed reaction that can now be optimized for future applied work.

Materials and Methods

TPADO and the reductase were heterologously expressed in *E. coli* and isolated in high yields via nickel affinity chromatography. Enzyme activity was monitored continuously via UV/vis monitoring of NADH disappearance and discontinuously via separation of reaction components via HPLC. Identities of starting material and products were detected with both mass-spectrometric and diode array UV-visible detectors. Quantification of reaction materials was carried out by integrating peak areas from chromatograms recorded at distinct absorption maxima. The crystal structures were solved by molecular replacement using structural homologs for the α- and β-subunits. TPADO was crystallized in the substrate-free state and soaked with substrates TPA and 2-OH-TPA overnight. Coordinates for the resulting apo and substrate-bound structures have been deposited in the Protein Data Bank (PDB IDs: 7Q04, 7Q05, 7Q06). Structural superpositions of TPADO with NagG (7C8Z) and NDO (1O7P) were generated and visualized in PyMol. An SSN of the family of proteins that the catalytic domain of TPADO belongs to (PF00848) was generated with EFI-EST web tools and visualized in Cytoscape. Detailed methods are provided in the *SI Appendix*.

Acknowledgements

We thank the Diamond Light Source (Didcot, UK) for beamtime (proposal MX-23269) and the beamline staff at I03 for support. Funding was provided by the National Science Foundation (MCB program) grant MCB1715176 (JLD, WMK, AR, JLB, RC) and U.S. Department of Energy, Office of Energy Efficiency and Renewable Energy, Advanced Manufacturing Office (AMO) and Bioenergy Technologies Office (BETO). This work was performed as part of the BioOptimized Technologies to keep Thermoplastics out of Landfills and the Environment (BOTTLE) Consortium and was supported by AMO and BETO under Contract DE-AC36-08GO28308 with the National Renewable Energy Laboratory (NREL), operated by Alliance for Sustainable Energy, LLC. The BOTTLE Consortium includes members from Montana State University and the University of Portsmouth, funded under Contract DE-AC36-08GO28308 with NREL. The views expressed in the article do not necessarily represent the views of the DOE or the U.S. Government. JEM acknowledges Research England for E3 funding. We thank Dr. Eric M. Shepard for collection of EPR spectra.

Author Contributions: **WMK**: Investigation, conceptualization, methodology, validation, formal analysis, writing, visualization; **MZ**: investigation, methodology, validation, formal analysis, writing, visualization; **AR**: investigation, methodology, validation; **RC**: validation, visualization, writing; **JL**: formal analysis, writing, visualization, methodology; **JLB**: investigation, conceptualization; **GTB**: Project administration, funding acquisition, conceptualization, writing; **JEM**: Project administration,

conceptualization, validation, methodology, writing; **JLD**: Project administration, conceptualization, methodology, funding acquisition, supervision, writing.

Competing Interest Statement: Disclose any competing interests here.

References

1. S. Yoshida, *et al.*, A bacterium that degrades and assimilates poly(ethylene terephthalate). *Science* **351**, 1196-1199 (2016).
2. R. Wei, W. Zimmermann, Biocatalysis as a green route for recycling the recalcitrant plastic polyethylene terephthalate. *Microb. Biotech.* **10**, 1302-1307 (2017).
3. L. D. Ellis, *et al.*, Chemical and biological catalysis for plastics recycling and upcycling. *Nature Cat.* **4**, 539-556 (2021).
4. A. J. Martin, C. Mondelli, S. D. Jaydev, J. Perez-Ramirez, Catalytic processing of plastic waste on the rise. *Chem* **7**, 1487-1533 (2021).
5. S. R. Nicholson, N. A. Rorrer, A. C. Carpenter, G. T. Beckham, Manufacturing energy and greenhouse gas emissions associated with plastics consumption. *Joule* **5**, 673-686 (2021).
6. T. Narancic, *et al.*, Genome analysis of the metabolically versatile *Pseudomonas umsongensis* GO16: The genetic basis for pet monomer upcycling into polyhydroxyalkanoates. *Microb. Biotech.* **14**, 2463-2480 (2021).
7. W. J. Li, *et al.*, Laboratory evolution reveals the metabolic and regulatory basis of ethylene glycol metabolism by *Pseudomonas putida* KT2440. *Environ. Microbiol.* **21**, 3669-3682 (2019).
8. M. A. Franden, *et al.*, Engineering *Pseudomonas putida* KT2440 for efficient ethylene glycol utilization. *Metab. Eng.* **48**, 197-207 (2018).
9. K. Y. Choi, *et al.*, Molecular and biochemical analysis of phthalate and terephthalate degradation by *Rhodococcus* sp strain DK17. *FEMS Microbiol. Lett.* **252**, 207-213 (2005).
10. M. Sasoh, *et al.*, Characterization of the terephthalate degradation genes of *Comamonas* sp, strain E6. *Appl. Environ. Microbiol.* **72**, 1825-1832 (2006).
11. Y. Z. Wang, Y. M. Zhou, G. J. Zystra, Molecular analysis of isophthalate and terephthalate degradation by *Comamonas testosteroni* YZW-D. *Environ. Health Persp.* **103**, 9-12 (1995).
12. H. Hara, L. D. Eltis, J. E. Davies, W. W. Mohn, Transcriptomic analysis reveals a bifurcated terephthalate degradation pathway in *Rhodococcus* sp. strain RHA1. *Journal of Bacteriology* **189**, 1641-1647 (2007).
13. H. R. Schläfli, M. A. Weiss, T. Leisinger, A. M. Cook, Terephthalate 1,2-dioxygenase system from *Comamonas testosteroni* T-2: Purification and some properties of the oxygenase component. *J. Bacteriol.* **176**, 6644-6652 (1994).
14. D. Kasai, M. Kitajima, M. Fukuda, E. Masai, Transcriptional regulation of the terephthalate catabolism operon in *Comamonas* sp. strain E6. *Applied and Environmental Microbiology* **76**, 6047-6055 (2010).
15. Y. Fukuhara, D. Kasai, Y. Katayama, M. Fukuda, E. Masai, Enzymatic properties of terephthalate 1,2-dioxygenase of *Comamonas* sp. strain E6. *Biosci., Biotech., Biochem.* **72**, 2335-2341 (2008).
16. J. Zheng, S. Suh, Strategies to reduce the global carbon footprint of plastics. *Nature Climate Change* **9**, 374-378 (2019).
17. A. Singh, *et al.*, Techno-economic, life-cycle, and socioeconomic impact analysis of enzymatic recycling of poly(ethylene terephthalate). *Joule* **5**, 2479-2503 (2021).
18. H. R. Schlafli, M. A. Weiss, T. Leisinger, A. M. Cook, Terephthalate 1,2-dioxygenase system from *Comamonas-testosteroni*-T-2 - purification and some properties of the oxygenase component. *J. Bact.* **176**, 6644-6652 (1994).
19. T. Shigematsu, K. Yumihara, Y. Ueda, S. Morimura, K. Kida, Purification and gene cloning of the oxygenase component of the terephthalate 1,2-dioxygenase system from *Delftia tsuruhatensis* strain T7. *FEMS Microbiol. Lett.* **220**, 255-260 (2003).
20. I. Pardo, *et al.*, Gene amplification, laboratory evolution, and biosensor screening reveal muck as a terephthalic acid transporter in *Acinetobacter baylyi* ADP1. *Metab. Eng.* **62**, 260-274 (2020).
21. T. Tiso, *et al.*, Towards bio-upcycling of polyethylene terephthalate. *Met. Eng.* **66**, 167-178 (2021).
22. H. T. Kim, *et al.*, Biological valorization of poly(ethylene terephthalate) monomers for upcycling waste PET. *ACS Sustainable Chem. & Eng.* **7**, 19396-19406 (2019).
23. H. T. Kim, *et al.*, Chemo-biological upcycling of poly(ethylene terephthalate) to multifunctional coating materials. *ChemSusChem* **14**, 4251-4259 (2021).
24. D. H. Kim, *et al.*, One-pot chemo-bioprocess of PET depolymerization and recycling enabled by a biocompatible catalyst, betaine. *ACS Catalysis* **11**, 3996-4008 (2021).
25. A. Z. Werner, *et al.*, Tandem chemical deconstruction and biological upcycling of poly(ethylene terephthalate) to β-ketoadipic acid by *Pseudomonas putida* KT2440. *Met. Eng.* **67**, 250-261 (2021).
26. D. J. Ferraro, L. Gakhar, S. Ramaswamy, Rieske business: Structure-function of rieske non-heme oxygenases. *Biochem. Biophys. Res. Comm.* **338**, 175-190 (2005).
27. V. Venturi, F. Zennaro, G. Degrassi, B. C. Okeke, C. V. Bruschi, Genetics of ferulic acid bioconversion to protocatechuic acid in plant-growth-promoting *Pseudomonas putida* WCS358. *Microbiology-(UK)* **144**, 965-973 (1998).
28. M. D. Wolfe, J. V. Parales, D. T. Gibson, J. D. Lipscomb, Single turnover chemistry and regulation of O-2 activation by the oxygenase component of naphthalene 1,2-dioxygenase. *J. Biol. Chem.* **276**, 1945-1953 (2001).
29. A. Karlsson, *et al.*, Crystal structure of naphthalene dioxygenase: Side-on binding of dioxygen to iron. *Science* **299**, 1039-1042 (2003).
30. S. M. Resnick, K. Lee, D. T. Gibson, Diverse reactions catalyzed by naphthalene dioxygenase from *Pseudomonas* sp. strain NCIB 9816. *J. Indust. Microbiol.* **17**, 438-457 (1996).
31. D. J. Ferraro, A. Okerlund, E. Brown, S. Ramaswamy, One enzyme, many reactions: Structural basis for the various reactions catalyzed by naphthalene 1,2-dioxygenase. *IUCRJ* **4**, 648-656 (2017).
32. O. Kweon, *et al.*, A new classification system for bacterial rieske non-heme iron aromatic ring-hydroxylating oxygenases. *BMC Biochem.* **9**, 11 (2008).
33. J. K. Capyk,L. D. Eltis, Phylogenetic analysis reveals the surprising diversity of an oxygenase class. *J. Biol. Inorg. Chem.* **17**, 425-436 (2012).
34. K. V. Ambrose, *et al.*, Functional characterization of salicylate hydroxylase from the fungal endophyte *Epichloë festucae*. *Sci. Reports* **5**, 10939 (2015).
35. H. K. Chang, P. Mohseni, G. J. Zylstra, Characterization and regulation of the genes for a novel anthranilate 1,2-dioxygenase from *Burkholderia cepacia* DBO1. *J. Bacteriol.* **185**, 5871-5881 (2003).
36. S. Jaganaman, A. Pinto, M. Tarasev, D. P. Ballou, High levels of expression of the iron-sulfur proteins phthalate dioxygenase and phthalate dioxygenase reductase in *Escherichia coli*. *Protein Express. Purif.* **52**, 273-279 (2007).
37. Y.-J. Hou, Y. Guo, D.-F. Li, N.-Y. Zhou, Structural and biochemical analysis reveals a distinct catalytic site of salicylate 5-monooxygenase naggh from Rieske dioxygenases. *Appl. Environ. Microbiol.* **87**, e01629-20 (2021).
38. T. A. Rogers,A. S. Bommarius, Utilizing simple biochemical measurements to predict lifetime output of biocatalysts in continuous isothermal processes. *Chem. Eng. Sci.* **65**, 2118-2124 (2010).
39. M. D. Wolfe, *et al.*, Benzoate 1,2-dioxygenase from *Pseudomonas putida:* Single turnover kinetics and regulation of a two-component Rieske dioxygenase. *Biochemistry* **41**, 9611-9626 (2002).
40. T. Ohta, S. Chakrabarty, J. D. Lipscomb, E. I. Solomon, Near-IR MCD of the nonheme ferrous active site in naphthalene 1,2-dioxygenase: Correlation to crystallography and structural insight into the mechanism of Rieske dioxygenases. *J. Am. Chem. Soc.* **130**, 1601-1610 (2008).
41. L. Holm, Using DALI for protein structure comparison. *Methods in Molec. Bio. (Clifton, N.J.)* **2112**, 29-42 (2020).
42. T. Fang,N. Y. Zhou, Purification and characterization of salicylate 5-hydroxylase, a three-component monooxygenase from *Ralstonia* sp. strain U2. *Applied Microbiol. Biotech.* **98**, 671-679 (2014).
43. D. J. Ferraro, A. L. Okerlund, J. C. Mowers, S. Ramaswamy, Structural basis for regioselectivity and stereoselectivity of product formation by naphthalene 1,2-dioxygenase. *J. Bact.* **188**, 6986-6994 (2006).
44. J. A. Gerlt, *et al.*, Enzyme function initiative-enzyme similarity tool (EFI-EST): A web tool for generating protein sequence similarity networks. *Biochim. Biophys. Acta* **1854**, 1019-1037 (2015).
45. C. L. Colbert, *et al.*, Structural characterization of *Pandoraea pnomenusa* B-356 biphenyl dioxygenase reveals features of potent polychlorinated biphenyl-degrading enzymes. *PloS One* **8**, e52550 (2013).
46. J. Qiu, *et al.*, Identification and characterization of a novel pic gene cluster responsible for picolinic acid degradation in *Alcaligenes faecalis* JQ135. *J. Bact.* **201**, e00077-00019 (2019).

47. M. S. Rogers, J. D. Lipscomb, Salicylate 5-hydroxylase: Intermediates in aromatic hydroxylation by a Rieske monooxygenase. *Biochemistry* **58**, 5305-5319 (2019).
48. O. Pinyakong, H. Habe, T. Yoshida, H. Nojiri, T. Omori, Identification of three novel salicylate 1-hydroxylases involved in the phenanthrene degradation of *Sphingobium* sp. strain P2. *Biochem. Biophys. Res. Commun.* **301**, 350-357 (2003).
49. Y. Jouanneau, J. Micoud, C. Meyer, Purification and characterization of a three-component salicylate 1-hydroxylase from *Sphingomonas* sp. strain CHY-1. *Appl. Environ. Microbiol.* **73**, 7515-7521 (2007).
50. F. P. Guengerich, Common and uncommon cytochrome P450 reactions related to metabolism and chemical toxicity. *Chem. Res. Toxicol.* **14**, 611-650 (2001).
51. C. S. Harwood, R. E. Parales, The beta-ketoadipate pathway and the biology of self-identity. *Ann. Rev. Microbiol.* **50**, 553-590 (1996).
52. R. Jamarani, H. C. Erythropel, J. A. Nicell, R. L. Leask, M. Marić, How green is your plasticizer? **10**, 834 (2018).
53. Y. Fukuhara, *et al.*, Characterization of the isophthalate degradation genes of *Comamonas* sp. strain E6. *Appl. Environ. Microbiol.* **76**, 519-527 (2010).
54. R. N. Roy, Bioactive natural derivatives of phthalate ester. *Crit. Rev. Biotech.* **40**, 913-929 (2020).
55. J. K. Mahto, *et al.*, Molecular insights into substrate recognition and catalysis by phthalate dioxygenase from C*omamonas testosteroni*.. *J. Biol. Chem.* (2021) ASAP.
56. R. E. Parales, *et al.*, Substrate specificity of naphthalene dioxygenase: Effect of specific amino acids at the active site of the enzyme. *J. Bacteriol.* **182**, 1641-1649 (2000).

Supplementary Appendix for

Biochemical and structural characterization of an aromatic ring-hydroxylating dioxygenase for terephthalic acid catabolism

William M. Kincannon,[1] Michael Zahn,[2,4] Rita Clare,[1,4] Ari Romberg,[1] James Larson,[1,] Jessica Lusty-Beech,[1,4,] Brian Bothner,[1] Gregg T. Beckham,[3,4] John E. McGeehan,[2,4,*] Jennifer L. DuBois,[1,4,*]

1. Department of Biochemistry, Montana State University, Bozeman MT, United States
2. Centre for Enzyme Innovation, University of Portsmouth, Portsmouth, United Kingdom
3. Renewable Resources and Enabling Sciences Center, National Renewable Energy Laboratory, Golden CO United States
4. BOTTLE Consortium, Golden CO, United States

*Correspondence: john.mcgeehan@port.ac.uk; jennifer.dubois1@montana.edu

Classification: Biological Sciences, Applied Biological Sciences

*John E. McGeehan, Jennifer L. DuBois
Email: john.mcgeehan@port.ac.uk; jennifer.dubois1@montana.edu

This PDF file includes:

Detailed Experimental Methods.

Generation of expression constructs

Genes encoding the TPA oxygenase alpha subunit (*tphA2*, BAE47077), beta subunit (*tphA3*, BAE47078) and reductase (*tphA1,* BAE47080) were synthesized in an IPTG-regulated pET expression vectors by Genscript. For TPADO, a TEV-cleavable C-terminal His_6 tag was appended to the α-subunit (TphA2), while the β-subunit (TphA3) was untagged**.** The two polypeptides were co-expressed in the same parent pET-DUET vector. For the reductase, a C-terminal His_6 tag was appended using pET45b. Vector maps and complete protein and gene sequences are given in Fig S1-S2.

Overexpression of recombinant dioxygenase

Lemo21(DE3) *E. coli* cells (New England Biolabs) harboring the pETDuet-TPADO construct were grown overnight at 37 °C in 100 mL LB starter media (5 g sodium chloride, 5 g yeast extract, 10 g tryptone per 1 L deionized water). The next morning, 500 mL TB media (1) in 2.8 L Fernbach flasks was seeded with 1/100 volume of overnight starter culture. The media was also supplemented with ampicillin (0.1 mg/mL), chloramphenicol (0.034 mg/mL), and rhamnose (2 mM). Cultures were incubated at 37 °C with agitation at 225 rpm until the OD_{600nm} reached ~4 (after ~5-6 h). Then, the flasks were cooled by complete submersion in ice for 45 minutes. After this time, the ice-chilled cultures were supplemented with 0.1 mM ferric chloride and 10 mg/L L-Cysteine hydrochloride monohydrate. Finally, protein production was induced with 1 mM IPTG and the culture flasks returned to incubation at 25 °C with agitation at 200 rpm. 18 h after the induction, cells were harvested 18 via centrifugation at 4 °C. The cell paste was frozen with liquid nitrogen, and stored at -80 °C until needed.

Purification of recombinant dioxygenase

Cell pellet was resuspended in buffer A (20 mM TrisHCl, 150 mM NaCl, 10% glycerol w/v%, pH 8.0), with approximately 20 g cell paste to 160 mL of buffer. The protease inhibitor phenylmethylsufonyl fluoride (PMSF) and chicken egg white lysozyme were added to 1 mM and 1 mg/mL, respectively. The solution was incubated at room temperature, with stirring, for 15 minutes. Then the lysate solution was sonified with a Branson digital sonifier, with stirring while incubated in an ice bath. The solution was sonified for 20 minutes, with a pulse sequence of 10 seconds on and 30 seconds off, at 50% amplitude. The sonified lysate was then clarified via centrifugation (18,000 x g, 30 minutes, 4 °C).

Clarified lysate was loaded onto 3 x 5 mL HisTrap (Cytiva) nickel affinity columns that were pre-equilibrated in buffer A. The column resin was then washed with 10 column volumes of buffer A, and then a second wash step with 60% buffer A and 40% buffer B (same as A, except with 0.25 M imidazole) for 10 column volumes. Finally, resin-bound protein was eluted via a step to 100% buffer B. Fractions (3 mL) were collected throughout the loading, washing, and eluting steps of the procedure, and analyzed for TPADO presence by SDS-PAGE. Significantly pure protein found in the final elution was concentrated to 4 mL with 30 MWCO amicon centrifugal filtration devices, and desalted into buffer A using a DG-10 column (BioRad). The protein was concentrated to 2 mL, dispensed into 0.05 mL aliquots, flash frozen with liquid nitrogen, and stored at -80 °C until needed.

Protein for crystallography was purified from 1 L scaled TB cultures, instead of 500 mL. Protein for crystallography was also buffer exchanged into Buffer A with an additional 2 mM dithiothreitol by dialysis with a 3 MWCO membrane, at the final steps of the purification. Buffer was changed two times over 24 h, and the protein was then concentrated and flash frozen as described above.

Overexpression of recombinant reductase

Lemo21(DE3) *E. coli* cells (New England Biolabs) harboring the pET-DO-Red construct (Fig S1) were grown overnight at 37 °C in 100 mL LB starter media (5 g sodium chloride, 5 g yeast extract, 10 g tryptone per 1 L deionized water). The next morning, 500 mL of TB media (1) in a 2.8 L Fernbach flask was seeded with 1/100 volume of overnight starter culture. The media was also supplemented with ampicillin (0.1 mg/mL), chloramphenicol (0.034 mg/mL), and rhamnose (2 mM). The seeded medium was then incubated at 37 °C with agitation at 225 rpm until the OD_{600nm} reached ~2.5 (or ~5 hours). Then, the culture was cooled by complete submersion in ice for 45 minutes. After this time, the ice-chilled culture was supplemented with 0.1 mM ferric chloride, 10 mg/L L-Cysteine hydrochloride monohydrate, and 0.2 mg/mL (−)-Riboflavin. Finally, protein production by the cells was induced with 0.4 mM IPTG, and cultures were returned to incubation at 20 °C with agitation at 200 rpm. The cells were harvested 18 hours after induction via centrifugation at 4 °C. The cell paste was frozen with liquid nitrogen and stored at -80 °C until needed.

Purification of recombinant reductase

All purification steps were conducted under anaerobic conditions, in a Coy anaerobic chamber containing 2.5% H_2/97.5%N_2 atmosphere. Cell pellet was resuspended in buffer C (20 mM TrisHCl, 300 mM NaCl, 15% glycerol w/v%, pH 8.0), with approximately 20 g cell paste to 160 mL of buffer. PMSF and chicken egg white lysozyme were added to 1 mM and 1 mg/mL, respectively. The solution was incubated at room temperature with stirring for 15 minutes. Then the lysate solution was sonified with a Branson digital sonifier, with stirring while incubated in an ice bath. The solution was sonified

for 20 minutes, with a pulse sequence of 10 second on and 30 seconds off, at 50% amplitude. The sonified lysate was then clarified via centrifugation (18,000 *x g*, 30 minutes, 4 °C). Clarified lysate was loaded onto 3 x 5 mL HisTrap (Cytiva) nickel affinity columns that were pre-equilibrated in buffer C. The column resin was then washed with 10 column volumes of buffer C. Protein was eluted via a step to 100% buffer D (same as buffer C, except with 100 mM imidazole). Fractions (3 mL) were collected throughout the loading, washing, and eluting steps of the procedure, and analyzed for reductase presence by SDS-PAGE. Significantly pure protein found in the final elution was concentrated to 4 mL with 30 MWCO Amicon centrifugal filtration devices, and desalted into buffer C using a DG-10 column (BioRad). The protein was concentrated to 2 mL, aliquoted into 0.05 mL volumes, flash frozen with liquid nitrogen, and stored at -80 °C until needed.

Atomic absorption spectroscopy

Dioxygenase was analyzed for iron content with three replicate preparations. Serial dilutions of each prep were conducted to obtain 10 mL of protein samples ranging from 2 to 5 μM concentration. Then, the samples were then analyzed by flame AA spectroscopy using an Agilent 240 FS instrument (Agilent Technologies Inc.) equipped with an Fe lamp utilizing an acetylene/air mixture (11:60 psi partial pressures) as a fuel source. Concentration of iron was then determined from a 0.1–2.0 ppm standard curve created from a 1000 ppm iron AA standard solution (Ricca Chemical Company).

Steady-state assays

Initial rates of reaction under steady state conditions (low [enzyme]) were measured using a Cary 60 UV-Vis spectrophotometer under aerobic conditions. Activity assays were performed by recording absorbance changes at 340 nm upon mixing of TPADO (1.9 μM), reductase (6.0 μM), and NADH (200 μM) with terephthalate (2.5–100 μM) over time. Reactions were conducted on a 2.0 mL scale, where a buffered solution of TPADO, reductase, and NADH was reacted with terephthate in a quartz cuvette (1 cm path length, concentrations after mixing). Solutions were prepared in ambient air (~250 μM O_2) and contained 20 mM TrisHCl pH 8.0, 150 mM sodium chloride, 10% glycerol (v/v), 20% DMSO (v/v), pH 8.0 buffer at 28 °C. Rates of reactions were obtained from the slope of the initial portion of the time course using the extinction coefficient of NADH at 340 nm (6.22 mM^{-1} x cm^{-1}). Rates were measured as a function of variable TPA concentration and subsequently fit to the Michaelis-Menten hyperbolic expression.

Assays for monitoring enzyme catalyzed aromatic hydrocarbon oxygenation

Assays for HPLC analysis were conducted on a 2.0 mL scale, incubated at 28 °C, and stirred at 400 RPM with a flea stir bar. Each assay contained 20 mM TrisHCl (pH 8.0), 150 mM sodium chloride, 15% DMSO, 10 % glycerol (v/v%), 0.5 mM TPA or TPA analog, 0.35 mM NADH, 1.0 μM reductase, and 2.1 μM dioxygenase. Assays were initiated with addition of dioxygenase. After dioxygenase was added, the reaction was rapidly mixed with a pipette and incubated with stirring. Reactions were quenched with ice cold methanol to a 1:1 ratio, and centrifuged to remove debris. For the measurement of total turnover number, conditions were changed to include 20% DMSO, 3 mM NADH, 1 mM TPA, 0.3 μM reductase, and 1 μM dioxygenase.

HPLC monitoring and analysis of TPADO-catalyzed reactions

Quenched reactions were analyzed for aromatic substrate/product, NADH, and NAD^+ content using a Shimadzu HPLC instrument with a diode array detector by injecting an aliquot onto a Thermo Scientific™ Hypersil GOLD™ (4.6 mm × 250 mm, 5 μm particle size) pre-equilibrated in 100% buffer A and 0% buffer B. Buffer A consisted of 0.1% (v/v) TFA (Fisher) in water. Buffer B consisted of 0.1% (v/v) TFA (Fisher) in HPLC grade acetonitrile (Fisher). The reaction components were eluted at a rate of 1 mL/min with the following program: 0% B from 0.0 to 10.0 min, 0 to 15% B from 10.0 to 18.0 min, 15 to 75% B from 18.0 to 24.0 min, 75 to 100% B from 24.0 to 24.1 min, 100% B from 24.1 to 30.1 min, 100% to 0%B from 30.1 to 30.2 minutes, and 0% B from 30.2 to 38.0 min. Authentic TPA (Sigma-Aldrich), 2-hydroxyterephthalate (Sigma-Aldrich; 10312-55-7), 2-aminoterephthalate (Alfa Aesar; 10312-55-7), 4-nitrobenzoate (ChemImpex; 62-23-7), 4-carbamoylbenzoate (AA Blocks; 6051-43-0), and 4-formylbenzoate (AA Blocks; 619-66-9), were used to develop standard curves from integrations of peaks monitored at the wavelengths and retention times reported in Table S1. The relationship of authentic substrate concentration to peak area was used to quantify substrate content in experimental reactions. Triplicate measurements were conducted for each condition, and mean concentrations were obtained with less than 10% standard error.

Partial uncoupling of NADH consumption from substrate hydroxylation was screened for in reactions involving each of the above substrates/analogs, and observed over baseline NADH consumption in the presence of 4-nitrobenzoate. To calculate %-uncoupling, the following expression was used:

$$\{\Delta[\text{substrate}] / (\Delta[\text{NADH}]_{\text{total}} - \Delta[\text{NADH}]_{\text{baseline}})\} \times 100\%$$

Here, ∆[substrate] was determined by the diminution of the area of the substrate-associated HPLC peak in a reaction mixture, and $\Delta[NADH]_{total}$ was determined by reduction of area of the NADH-associated HPLC peak in the same reaction sample following incubation for 90 minutes. In each case, data were measured for an initial time point (t_0) immediately prior to the addition of enzymes. $\Delta[NADH]_{baseline}$ was determined in a similar fashion by measuring diminution of the area of the NADH-associated HPLC peak in a no-substrate control, before and following incubation with the TPADO/reductase enzymes.

Liquid chromatography (LC) electrospray and tandem mass spectrometry (MS-MS)

To confirm the production of DCD, a reaction containing TPA/NADH prior to addition of enzymes and a quenched TPADO/reductase-catalyzed reaction using TPA and NADH as substrates (see above) were each analyzed via reverse-phase UV-UPLC-LCMS on an Agilent 1290 infinity UHPLC equipped with DAD coupled to an Agilent 6538 high-resolution Q-TOF mass analyzer. Samples were injected onto an Acquity UPLC HSS T3 column (2.1 x 100mm, 1.8 µm particle size). Molecules were eluted at a rate of 0.2 mL/min with the following program: 1% B from 0 to 2 minutes, 1 to 15% B from 2 to 10 minutes, 15 to 75% B from 10 to 16 minutes, 75 to 99% B from 16 to 16.1 minutes, 99% B from 16.1 to 18.1 minutes, 99 to 1% B from 18.1 to 18.2 minutes, and 1% B from 18.2 to 22 minutes with A consisting of water with 0.1% (v/v) formic acid and B consisting of acetonitrile with 0.1% formic acid (v/v). The combined data from UV signal, retention time, accurate mass data, and extracted ion chromatograms were used to identify TPA, NAD^+, NADH, and DCD. These data identified ions with expected m/z for TPA and DCD, as well as several putative fragmentation products generated in the source. To confirm loss of water/carbon dioxide from DCD and TPA, the samples were re-analyzed using targeted MS-MS on TPA/DCD ions at their expected elution times by collision induced fragmentation (CID) with the source voltage reduced from 3,500 V to 3,000 V. Exact mass data are summarized in Table S3 and Figure S8.

Native protein mass spectrometry

The native assembly of the dioxygenase was analyzed with native mass spectrometry (NMS). To prevent aggregation in the absence of glycerol, TPADO (initial concentration of 40 mg/mL) was first diluted to a concentration of 1 mg/mL with 170 mM ammonium acetate (Sigma-Aldrich) pH 8. The enzyme was buffer exchanged via overnight dialysis against 170 mM ammonium acetate pH 8 using a 3.5 kDa molecular weight cut-off membrane (Thermo Fisher). Sample was introduced with in-house prepared gold-coated borosilicate glass (2) capillaries and analyzed with a Synapt G2-Si electrospray time-of-flight mass spectrometer (Waters). Analysis of 10 µM TPADO in 170 mM ammonium acetate pH 8 and 2% acetonitrile (Thermo Fisher) resulted in an unsteady signal. Reducing the concentration of ammonium acetate to 50 increased signal stability considerably. Instrument parameters were: 60 °C source temperature, trap collision energy of 10 V, and capillary voltage of 2.00 V. Data were collected in positive mode at 1 Hz scan frequency over a m/z scan range of 800-12,000.

Electron paramagnetic resonance.

Concentrated dioxygenase protein was thawed in an anaerobic chamber rapidly by holding in a gloved hand. Then, 0.3 mL of protein was treated with 0.03 mL of 0.1 M sodium dithionite (dissolved in buffer A), and mixed gently with a pipettor. The mixture was incubated at room temperature in the anaerobic chamber for 30 minutes, then transferred to an EPR tube and capped with a rubber lid to maintain anaerobicity. The tube was then removed from the chamber, and immediately frozen with liquid nitrogen. Concentrated reductase protein was thawed in an anaerobic chamber rapidly by holding in a gloved hand. Then, 0.3 mL of protein was treated with 0.03 mL of 0.1 M NADH (dissolved in buffer C), and mixed gently with a pipettor. The mixture was incubated at room temperature in the anaerobic chamber for 30 minutes, then transferred to an EPR tube and capped with a rubber lid to maintain anaerobicity. The tube was then removed from the chamber, and immediately frozen with liquid nitrogen.

X-Band CW EPR spectra of dioxygenase and reductase proteins were measured with a Bruker EMX spectrometer fitted with a ColdEdge (Sumitomo Cryogenics) 10 K waveguide in-cavity cryogen-free system with an Oxford Mercury iTC controller unit and helium Stinger recirculating unit (Sumitomo Cryogenics, ColdEdge Technologies). Helium gas flow was maintained at 100 psi. Spectral parameters were 2 mW microwave power, 100 kHz modulation frequency, and 10 G modulation amplitude. Spectra were averaged over 2 scans.

Crystallography

The Ni-affinity purified TPADO was buffer exchanged into 20 mM Tris pH 8.0, 150 mM NaCl, 10% glycerol, 2 mM DTT and concentrated to 10 mg/mL. Sitting drop crystallization trials were set up with a Mosquito crystallization robot (sptlabtech) using SWISSCI 3-lens low profile crystallization plates. Crystals grew in condition C4 (12.5% MPD, 12.5% PEG 3350, 12.5% PEG 1000, 0.3 M sodium nitrate, 0.3 M sodium phosphate dibasic, 0.3 M ammonium sulfate, 0.1 M buffer Imidazole/MES pH 6.5) of the Morpheus screen (Molecular Dimensions). Apo crystals were soaked with 4 mM TPA for 24 hours before they were cryocooled with liquid nitrogen. Diffraction data were collected on beamline I03 at the Diamond Light Source (Didcot, UK) and automatically processed with STARANISO (3) on ISPyB. The structure was solved within CCP4 Cloud by molecular replacement with Molrep (4) using homology search models for the alpha and beta subunits from SWISS-MODEL (5) based

on pdb entries 3N0Q and 3EBY, respectively. Model building was performed in Coot (6) and the structures was refined with REFMAC5 (7). MolProbity (8) was used to evaluate the final model and PyMOL (Schrödinger, LLC) for protein model visualization. Data and Refinement statistics are summarized in Table S2. The atomic coordinates have been deposited in the Protein Data Bank and are available under the accession codes 7Q04, 7Q05 and 7Q06. Search for structural protein homologs and calculation of root-mean-square deviation values were performed with the DALI server (9).

Sequence Similarity Network

The network was generated by the EFI-EST (Enzyme Function Initiative – Enzyme Similarity Tool) server from PF00848, a protein family containing putative alpha subunits for aromatic ring-hydroxylating Rieske dioxygenases. A UniRef90 was applied to the 45,076 membered Pfam to reduce the total number unique sequences in the network to 17,319. The sequence alignments were conducted with E-value of 5 and alignment score of 85 to generate the full network. A repnode network was used to condense sequences with ≥40% sequence ID into single nodes. This repnode network was colorized with the EFI-EST server, and subsequently visualized in Cytoscape v3.11.1 where a y-files organic layout was applied and singletons were manually removed, resulting in the final network shown. Only clusters with at least a single node containing a sequence of a functionally verified protein are numbered. Clusters are numbered by the “Sequence Count Cluster Number” attribute bestowed by the EFI-EST server. Table S4 lists representative UniProt IDs from each cluster that are associated with proteins/genes with published activities.

Supplementary Figures

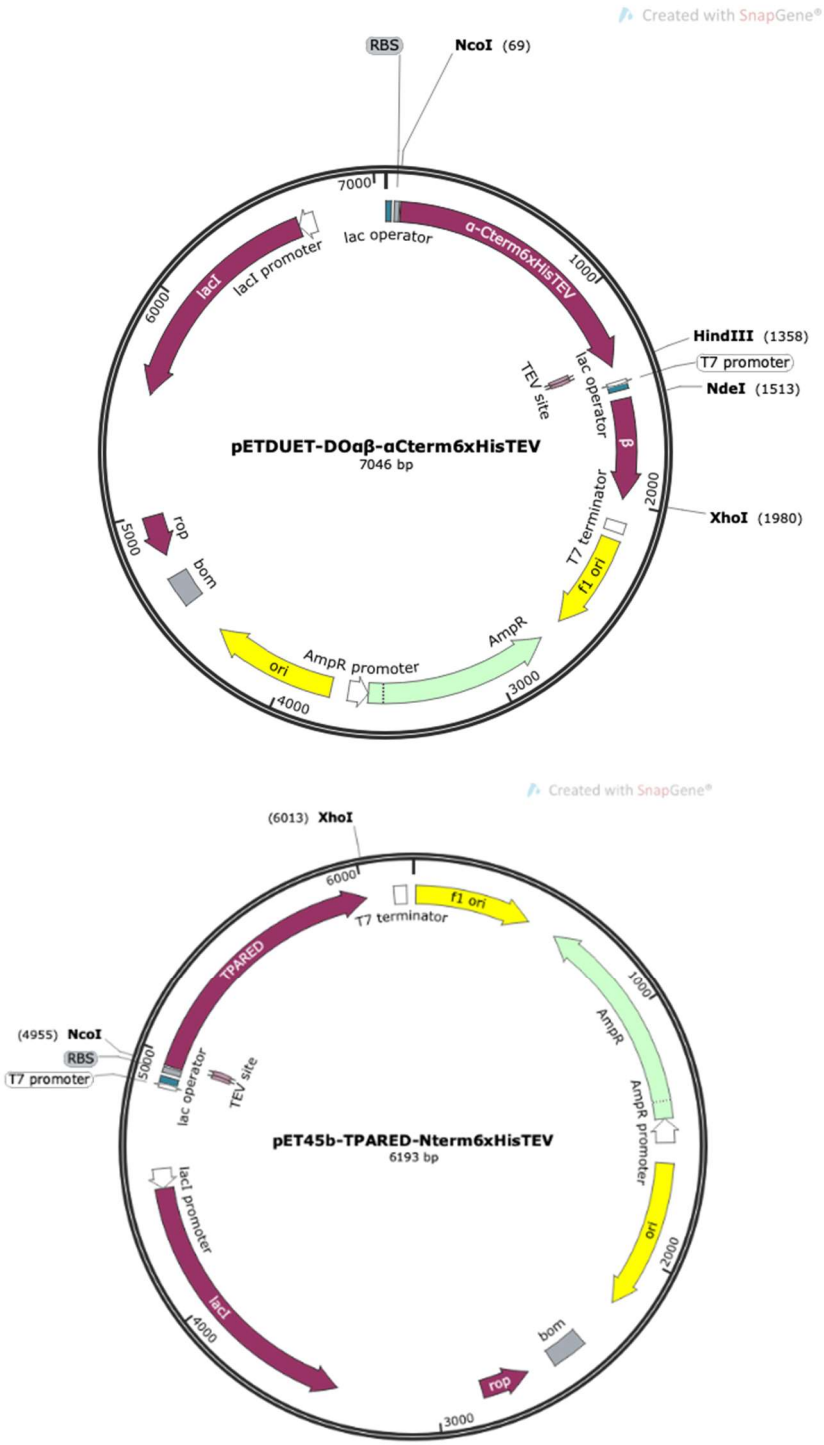


Figure S1. Plasmid maps for expressing TPADO (top) and its reductase (bottom), generated with SnapGene Viewer. A C-terminal His6 tag is appended to the α-subunit of TPADO and to the reductase.

(A) MGMQESIIQWHGATNTRVPFGIYTDTANADQEQQRIYRGEVWNYLCLESEIPGAGDFRTTFAGETPIVVVRDADQEIYAFENRCAHRGALIALEKSGRTDSFQCVYHAWSYNRQG
DLTGVAFEKGVKGQGGMPASFCKEEHGPRKLRVAVFCGLVFGSFSEDVPSIEDYLGPEICERIERVLHKPVEVIGRFTQKLPNNWKLYFENVKDSYHASLLHMFFTTFELNRLSQ
KGGVIVDESGGHHVSYSMIDRGAKDDSYKDQAIRSDNERYRLKDPSLLEGFEEFEDGVTLQILSVFPGFVLQQIQNSIAVRQLLPKSISSSELNWTYLGYADDSAEQRKVRLKQA
NLIGPAGFISMEDGAVGGFVQRGIAGAANLDAVIEMGGDHEGSSEGRATETSVRGFWKAYRKHMGQEMQA**ENLYFQG**HHHHHH*

(B) MINEIQIAAFNAAYAKTVDSDAMEQWPTFFTKDCHYRVTNVDNHAEGLAAGIVWADSQDMLTDRISALREANIYERHRYRHILGLPSIQSGDATQASASTPFMVLRIMHTGETEV
FASGEYLDKFTTIDGKLRLQERIAVCDSTVTDTLMALPL*

(C) GGGGAATTGTGAGCGGATAACAATTCCCCTCTAGAAATAATTTTGTTTAACTTTAAGAAGGAGATATACCatgggcatgcaggaaagcatcattcaatggcacggtgcgaccaac
acccgtgttccgttcggcatctataccgacaccgcgaacgcggatcaggaacagcaacgtatttaccgtggtgaggtgtggaactatctgtgcctggagagcgaaattccgggtg
cgggcgacttccgtaccacctttgcgggcgagaccccgatcgtggttgtgcgtgacgcggatcaagaaatttacgcgttcgagaaccgttgcgcgcaccgtggtgcgctgattgc
gctggaaaagagcggccgtaccgacagctttcagtgcgtttaccacgcgtggagctataaccgtcaaggtgatctgaccggcgttgcgttcgagaagggtgtgaaaggccagggt
ggcatgccggcgagcttttgcaaagaggagcacggtccgcgtaaactgcgtgttgcggtgttctgcggtctggttttcggcagctttagcgaagacgtgccgagcattgaggatt
acctgggtccggaaatctgcgaacgtattgagcgtgttctgcacaaaccggttgaagtgatcggccgtttcacccagaagctgccgaacaactggaaactgtactttgagaacgt
gaaggacagctatcacgcgagcctgctgcacatgttctttaccacctttgaactgaaccgtctgagccaaaaaggtggcgttatcgtggacgagagcggtggccaccacgttagc
tacagcatgattgatcgtggtgcgaaggacgatagctataaagaccaggcgatccgtagcgataacgaacgttaccgtctgaaggacccgagcctgctggaaggtttcgaggaat
ttgaggatggcgttaccctgcaaattctgagcgttttcccgggttttgtgctgcagcaaatccagaacagcattgcggtgcgtcaactgctgccgaaaagcatcagcagcagcga
actgaactggacctacctgggttatgcggacgatagcgcggagcagcgtaaagtgcgtctgaaacaagcgaacctgatcggtccggcgggcttcattagcatggaggatggcgcg
gttggtggctttgtgcagcgtggtattgcgggtgcggcgaacctggacgcggttattgaaatgggtggcgatcacgaaggtagcagcgagggtcgtgcgaccgagaccagcgtgc
gtggtttctggaaggcgtatcgtaaacacatgggccaggaaatgcaagcgGAGAACCTGTACTTTCAGGGTCATCACCACCACCACCACtaaAAGCTTGCGGCCGCATAATGCTT
AAGTCGAACAGAAAGTAATCGTATTGTACACGGCCGCATAATCGAAATTAATACGACTCACTATAGGGGAATTGTGAGCGGATAACAATTCCCCATCTTAGTATATTAGTTAAGT
ATAAGAAGGAGATATACATatgatcaacgagatccagattgcggcgttcaacgcggcgtacgcgaagaccgtggacagcgatgcgatggaacaatggccgacctttctttaccaaa
gactgccactatcgtgtgaccaacgttgataaccatgcggagggtctggcggcgggtattgtttgggcggacagccaggatatgctgaccgaccgtatcagcgcgctgcgtgagg
cgaacatttacgaacgtcaccgttatcgtcacatcctgggtctgccgagcattcaaagcggtgatgcgacccaagcgagcgcgagcaccccgtttatggtgctgcgtatcatgca
caccggcgagaccgaagttttcgcgagcggcgagtacctggacaagtttaccaccatcgatggtaaactgcgtctgcaagaacgtattgcggtgtgcgacagcaccgttaccgat
accctgatggcgctgccgctgtaaCTCGAGTCTGGTAAAGAAACCGCTGCTGCGAAATTTGAACGCCAGCACATGGACTCGTCTACTAGCGCAGCTTAATTAACCTAGGCTGCTG
CCACCGCTGAGCAATAACTAGCATAACCCCTTGGGGCCTCTAAACGGGTCTTGAGGGGTTTTTTGCTGAAAGGAGGAACTATATCCGGATTGGCGAATGGGACGCGCCCTGTAGC
GGCGCATTAAGCGCGGCGGGTGTGGTGGTTACGCGCAGCGTGACCGCTACACTTGCCAGCGCCCTAGCGCCCGCTCCTTTCGCTTTCTTCCCTTCCTTTCTCGCCACGTTCGCCG
GCTTTCCCCGTCAAGCTCTAAATCGGGGGCTCCCTTTAGGGTTCCGATTTAGTGCTTTACGGCACCTCGACCCCAAAAAACTTGATTAGGGTGATGGTTCACGTAGTGGGCCATC
GCCCTGATAGACGGTTTTTCGCCCTTTGACGTTGGAGTCCACGTTCTTTAATAGTGGACTCTTGTTCCAAACTGGAACAACACTCAACCCTATCTCGGTCTATTCTTTTGATTTA
TAAGGGATTTTGCCGATTTCGGCCTATTGGTTAAAAAATGAGCTGATTTAACAAAAATTTAACGCGAATTTTAACAAAATATTAACGTTTACAATTTCTGGCGGCACGATGGCAT
GAGATTATCAAAAAGGATCTTCACCTAGATCCTTTTAAATTAAAAATGAAGTTTTAAATCAATCTAAAGTATATATGAGTAAACTTGGTCTGACAGTTACCAATGCTTAATCAGT
GAGGCACCTATCTCAGCGATCTGTCTATTTCGTTCATCCATAGTTGCCTGACTCCCCGTCGTGTAGATAACTACGATACGGGAGGGCTTACCATCTGGCCCCAGTGCTGCAATGA
TACCGCGAGACCCACGCTCACCGGCTCCAGATTTATCAGCAATAAACCAGCCAGCCGGAAGGGCCGAGCGCAGAAGTGGTCCTGCAACTTTATCCGCCTCCATCCAGTCTATTAA
TTGTTGCCGGGAAGCTAGAGTAAGTAGTTCGCCAGTTAATAGTTTGCGCAACGTTGTTGCCATTGCTACAGGCATCGTGGTGTCACGCTCGTCGTTTGGTATGGCTTCATTCAGC
TCCGGTTCCCAACGATCAAGGCGAGTTACATGATCCCCCATGTTGTGCAAAAAAGCGGTTAGCTCCTTCGGTCCTCCGATCGTTGTCAGAAGTAAGTTGGCCGCAGTGTTATCAC
TCATGGTTATGGCAGCACTGCATAATTCTCTTACTGTCATGCCATCCGTAAGATGCTTTTCTGTGACTGGTGAGTACTCAACCAAGTCATTCTGAGAATAGTGTATGCGGCGACC
GAGTTGCTCTTGCCCGGCGTCAATACGGGATAATACCGCGCCACATAGCAGAACTTTAAAAGTGCTCATCATTGGAAAACGTTCTTCGGGGCGAAAACTCTCAAGGATCTTACCG
CTGTTGAGATCCAGTTCGATGTAACCCACTCGTGCACCCAACTGATCTTCAGCATCTTTTACTTTCACCAGCGTTTCTGGGTGAGCAAAAACAGGAAGGCAAAATGCCGCAAAAA
AGGGAATAAGGGCGACACGGAAATGTTGAATACTCATACTCTTCCTTTTTCAATCATGATTGAAGCATTTATCAGGGTTATTGTCTCATGAGCGGATACATATTTGAATGTATTT
AGAAAAATAAACAAATAGGTCATGACCAAAATCCCTTAACGTGAGTTTTCGTTCCACTGAGCGTCAGACCCCGTAGAAAAGATCAAAGGATCTTCTTGAGATCCTTTTTTTCTGC
GCGTAATCTGCTGCTTGCAAACAAAAAAACCACCGCTACCAGCGGTGGTTTGTTTGCCGGATCAAGAGCTACCAACTCTTTTTCCGAAGGTAACTGGCTTCAGCAGAGCGCAGAT
ACCAAATACTGTCCTTCTAGTGTAGCCGTAGTTAGGCCACCACTTCAAGAACTCTGTAGCACCGCCTACATACCTCGCTCTGCTAATCCTGTTACCAGTGGCTGCTGCCAGTGGC
GATAAGTCGTGTCTTACCGGGTTGGACTCAAGACGATAGTTACCGGATAAGGCGCAGCGGTCGGGCTGAACGGGGGGTTCGTGCACACAGCCCAGCTTGGAGCGAACGACCTACA
CCGAACTGAGATACCTACAGCGTGAGCTATGAGAAAGCGCCACGCTTCCCGAAGGGAGAAAGGCGGACAGGTATCCGGTAAGCGGCAGGGTCGGAACAGGAGAGCGCACGAGGGA
GCTTCCAGGGGGAAACGCCTGGTATCTTTATAGTCCTGTCGGGTTTCGCCACCTCTGACTTGAGCGTCGATTTTTGTGATGCTCGTCAGGGGGGCGGAGCCTATGGAAAAACGCC
AGCAACGCGGCCTTTTTACGGTTCCTGGCCTTTTGCTGGCCTTTTGCTCACATGTTCTTTCCTGCGTTATCCCCTGATTCTGTGGATAACCGTATTACCGCCTTTGAGTGAGCTG
ATACCGCTCGCCGCAGCCGAACGACCGAGCGCAGCGAGTCAGTGAGCGAGGAAGCGGAAGAGCGCCTGATGCGGTATTTTCTCCTTACGCATCTGTGCGGTATTTCACACCGCAT
ATATGGTGCACTCTCAGTACAATCTGCTCTGATGCCGCATAGTTAAGCCAGTATACACTCCGCTATCGCTACGTGACTGGGTCATGGCTGCGCCCCGACACCCGCCAACACCCGC
TGACGCGCCCTGACGGGCTTGTCTGCTCCCGGCATCCGCTTACAGACAAGCTGTGACCGTCTCCGGGAGCTGCATGTGTCAGAGGTTTTCACCGTCATCACCGAAACGCGCGAGG
CAGCTGCGGTAAAGCTCATCAGCGTGGTCGTGAAGCGATTCACAGATGTCTGCCTGTTCATCCGCGTCCAGCTCGTTGAGTTTCTCCAGAAGCGTTAATGTCTGGCTTCTGATAA
AGCGGGCCATGTTAAGGGCGGTTTTTTCCTGTTTGGTCACTGATGCCTCCGTGTAAGGGGGATTTCTGTTCATGGGGGTAATGATACCGATGAAACGAGAGAGGATGCTCACGAT
ACGGGTTACTGATGATGAACATGCCCGGTTACTGGAACGTTGTGAGGGTAAACAACTGGCGGTATGGATGCGGCGGGACCAGAGAAAAATCACTCAGGGTCAATGCCAGCGCTTC
GTTAATACAGATGTAGGTGTTCCACAGGGTAGCCAGCAGCATCCTGCGATGCAGATCCGGAACATAATGGTGCAGGGCGCTGACTTCCGCGTTTCCAGACTTTACGAAACACGGA
AACCGAAGACCATTCATGTTGTTGCTCAGGTCGCAGACGTTTTGCAGCAGCAGTCGCTTCACGTTCGCTCGCGTATCGGTGATTCATTCTGCTAACCAGTAAGGCAACCCCGCCA
GCCTAGCCGGGTCCTCAACGACAGGAGCACGATCATGCTAGTCATGCCCCGCGCCCACCGGAAGGAGCTGACTGGGTTGAAGGCTCTCAAGGGCATCGGTCGAGATCCCGGTGCC
TAATGAGTGAGCTAACTTACATTAATTGCGTTGCGCTCACTGCCCGCTTTCCAGTCGGGAAACCTGTCGTGCCAGCTGCATTAATGAATCGGCCAACGCGCGGGGAGAGGCGGTT
TGCGTATTGGGCGCCAGGGTGGTTTTTCTTTTCACCAGTGAGACGGGCAACAGCTGATTGCCCTTCACCGCCTGGCCCTGAGAGAGTTGCAGCAAGCGGTCCACGCTGGTTTGCC
CCAGCAGGCGAAAATCCTGTTTGATGGTGGTTAACGGCGGGATATAACATGAGCTGTCTTCGGTATCGTCGTATCCCACTACCGAGATGTCCGCACCAACGCGCAGCCCGGACTC
GGTAATGGCGCGCATTGCGCCCAGCGCCATCTGATCGTTGGCAACCAGCATCGCAGTGGGAACGATGCCCTCATTCAGCATTTGCATGGTTTGTTGAAAACCGGACATGGCACTC
CAGTCGCCTTCCCGTTCCGCTATCGGCTGAATTTGATTGCGAGTGAGATATTTATGCCAGCCAGCCAGACGCAGACGCGCCGAGACAGAACTTAATGGGCCCGCTAACAGCGCGA
TTTGCTGGTGACCCAATGCGACCAGATGCTCCACGCCCAGTCGCGTACCGTCTTCATGGGAGAAAATAATACTGTTGATGGGTGTCTGGTCAGAGACATCAAGAAATAACGCCGG
AACATTAGTGCAGGCAGCTTCCACAGCAATGGCATCCTGGTCATCCAGCGGATAGTTAATGATCAGCCCACTGACGCGTTGCGCGAGAAGATTGTGCACCGCCGCTTTACAGGCT
TCGACGCCGCTTCGTTCTACCATCGACACCACCACGCTGGCACCCAGTTGATCGGCGCGAGATTTAATCGCCGCGACAATTTGCGACGGCGCGTGCAGGGCCAGACTGGAGGTGG
CAACGCCAATCAGCAACGACTGTTTGCCCGCCAGTTGTTGTGCCACGCGGTTGGGAATGTAATTCAGCTCCGCCATCGCCGCTTCCACTTTTTCCCGCGTTTTCGCAGAAACGTG
GCTGGCCTGGTTCACCACGCGGGAAACGGTCTGATAAGAGACACCGGCATACTCTGCGACATCGTATAACGTTACTGGTTTCACATTCACCACCCTGAATTGACTCTCTTCCGGGC
CGCTATCATGCCATACCGCGAAAGGTTTTGCGCCATTCGATGGTGTCCGGGATCTCGACGCTCTCCCTTATGCGACTCCTGCATTAGGAAGCAGCCCAGTAGTAGGTTGAGGCCG
TTGAGCACCGCCGCCGCAAGGAATGGTGCATGCAAGGAGATGGCGCCCAACAGTCCCCCGGCCACGGGGCCTGCCACCATACCCACGCCGAAACAAGCGCTCATGAGCCCGAAGT
GGCGAGCCCGATCTTCCCCATCGGTGATGTCGGCGATATAGGCGCCAGCAACCGCACCTGTGGCGCCGGTGATGCCGGCCACGATGCGTCCGGCGTAGAGGATCGAGATCGATCT
CGATCCCGCGAAATTAATACGACTCACTATA

(D) MHHHHHH**ENLYFQG**NHQIHIHDSDIAFPCAPGQSVLDAALQAGIELPYSCRKGSCGNCASALLDGNITSFNGMAVRSELCTSEQVLLCGCTAASDIRIQPSSFRRLDPEARKRFT
AKVYSNTLAAPDVSLLRLRLPVGKRAKFEAGQYLLIHLDDGESRSYSMANPPHESDGITLHVRHVPGGRFSTIVQQLKSGDTLEIELPFGSIALKPDDTRPLICVAGGTGFAPIK
SVLDDLAKRKVQRDITLIWGARNPSGLYLPSAIDKWRKTWPQFRYIAAITDLGNVPADAHAGRVDDALRTHFGNLHDHVVHCCGSPSLVQSVRTAASDMGLLAQNFHADVFATSP
TGSH*

(E) TGGCGAATGGGACGCGCCCTGTAGCGGCGCATTAAGCGCGGCGGGTGTGGTGGTTACGCGCAGCGTGACCGCTACACTTGCCAGCGCCCTAGCGCCCGCTCCTTTCGCTTTCTTC
CCTTCCTTTCTCGCCACGTTCGCCGGCTTTCCCCGTCAAGCTCTAAATCGGGGGCTCCCTTTAGGGTTCCGATTTAGTGCTTTACGGCACCTCGACCCCAAAAAACTTGATTAGG
GTGATGGTTCACGTAGTGGGCCATCGCCCTGATAGACGGTTTTTCGCCCTTTGACGTTGGAGTCCACGTTCTTTAATAGTGGACTCTTGTTCCAAACTGGAACAACACTCAACCC
TATCTCGGTCTATTCTTTTGATTTATAAGGGATTTTGCCGATTTCGGCCTATTGGTTAAAAAATGAGCTGATTTAACAAAAATTTAACGCGAATTTTAACAAAATATTAACGTTT
ACAATTTCTGGCGGCACGATGGCATGAGATTATCAAAAAGGATCTTCACCTAGATCCTTTTAAATTAAAAATGAAGTTTTAAATCAATCTAAAGTATATATGAGTAAACTTGGTC

TGACAGTTACCAATGCTTAATCAGTGAGGCACCTATCTCAGCGATCTGTCTATTTCGTTCATCCATAGTTGCCTGACTCCCCGTCGTGTAGATAACTACGATACGGGAGGGCTTA
CCATCTGGCCCCAGTGCTGCAATGATACCGCGAGACCCACGCTCACCGGCTCCAGATTTATCAGCAATAAACCAGCCAGCCGGAAGGGCCGAGCGCAGAAGTGGTCCTGCAACTT
TATCCGCCTCCATCCAGTCTATTAATTGTTGCCGGGAAGCTAGAGTAAGTAGTTCGCCAGTTAATAGTTTGCGCAACGTTGTTGCCATTGCTACAGGCATCGTGGTGTCACGCTC
GTCGTTTGGTATGGCTTCATTCAGCTCCGGTTCCCAACGATCAAGGCGAGTTACATGATCCCCCATGTTGTGCAAAAAAGCGGTTAGCTCCTTCGGTCCTCCGATCGTTGTCAGA
AGTAAGTTGGCCGCAGTGTTATCACTCATGGTTATGGCAGCACTGCATAATTCTCTTACTGTCATGCCATCCGTAAGATGCTTTTCTGTGACTGGTGAGTACTCAACCAAGTCAT
TCTGAGAATAGTGTATGCGGCGACCGAGTTGCTCTTGCCCGGCGTCAATACGGGATAATACCGCGCCACATAGCAGAACTTTAAAAGTGCTCATCATTGGAAAACGTTCTTCGGG
GCGAAAACTCTCAAGGATCTTACCGCTGTTGAGATCCAGTTCGATGTAACCCACTCGTGCACCCAACTGATCTTCAGCATCTTTTACTTTCACCAGCGTTTCTGGGTGAGCAAAA
ACAGGAAGGCAAAATGCCGCAAAAAAGGGAATAAGGGCGACACGGAAATGTTGAATACTCATACTCTTCCTTTTTCAATCATGATTGAAGCATTTATCAGGGTTATTGTCTCATG
AGCGGATACATATTTGAATGTATTTAGAAAAATAAACAAATAGGTCATGACCAAAATCCCTTAACGTGAGTTTTCGTTCCACTGAGCGTCAGACCCCGTAGAAAAGATCAAAGGA
TCTTCTTGAGATCCTTTTTTTTCTGCGCGTAATCTGCTGCTTGCAAACAAAAAAACCACCGCTACCAGCGGTGGTTTGTTTGCCGGATCAAGAGCTACCAACTCTTTTTCCGAAGG
TAACTGGCTTCAGCAGAGCGCAGATACCAAATACTGTCCTTCTAGTGTAGCCGTAGTTAGGCCACCACTTCAAGAACTCTGTAGCACCGCCTACATACCTCGCTCTGCTAATCCT
GTTACCAGTGGCTGCTGCCAGTGGCGATAAGTCGTGTCTTACCGGGTTGGACTCAAGACGATAGTTACCGGATAAGGCGCAGCGGTCGGGCTGAACGGGGGGTTCGTGCACACAG
CCCAGCTTGGAGCGAACGACCTACACCGAACTGAGATACCTACAGCGTGAGCTATGAGAAAGCGCCACGCTTCCCGAAGGGAGAAAGGCGGACAGGTATCCGGTAAGCGGCAGGG
TCGGAACAGGAGAGCGCACGAGGGAGCTTCCAGGGGGAAACGCCTGGTATCTTTATAGTCCTGTCGGGTTTCGCCACCTCTGACTTGAGCGTCGATTTTTGTGATGCTCGTCAGG
GGGGCGGAGCCTATGGAAAAACGCCAGCAACGCGGCCTTTTTACGGTTCCTGGCCTTTTGCTGGCCTTTTGCTCACATGTTCTTTCCTGCGTTATCCCCTGATTCTGTGGATAAC
CGTATTACCGCCTTTGAGTGAGCTGATACCGCTCGCCGCAGCCGAACGACCGAGCGCAGCGAGTCAGTGAGCGAGGAAGCGGAAGAGCGCCTGATGCGGTATTTTCTCCTTACGC
ATCTGTGCGGTATTTCACACCGCATATATGGTGCACTCTCAGTACAATCTGCTCTGATGCCGCATAGTTAAGCCAGTATACACTCCGCTATCGCTACGTGACTGGGTCATGGCTG
CGCCCCGACACCCGCCAACACCCGCTGACGCGCCCTGACGGGCTTGTCTGCTCCCGGCATCCGCTTACAGACAAGCTGTGACCGTCTCCGGGAGCTGCATGTGTCAGAGGTTTTC
ACCGTCATCACCGAAACGCGCGAGGCAGCTGCGGTAAAGCTCATCAGCGTGGTCGTGAAGCGATTCACAGATGTCTGCCTGTTCATCCGCGTCCAGCTCGTTGAGTTTCTCCAGA
AGCGTTAATGTCTGGCTTCTGATAAAGCGGGCCATGTTAAGGGCGGTTTTTTCCTGTTTGGTCACTGATGCCTCCGTGTAAGGGGGATTTCTGTTCATGGGGGTAATGATACCGA
TGAAACGAGAGAGGATGCTCACGATACGGGTTACTGATGATGAACATGCCCGGTTACTGGAACGTTGTGAGGGTAAACAACTGGCGGTATGGATGCGGCGGGACCAGAGAAAAAT
CACTCAGGGTCAATGCCAGCGCTTCGTTAATACAGATGTAGGTGTTCCACAGGGTAGCCAGCAGCATCCTGCGATGCAGATCCGGAACATAATGGTGCAGGGCGCTGACTTCCGC
GTTTCCAGACTTTACGAAACACGGAAACCGAAGACCATTCATGTTGTTGCTCAGGTCGCAGACGTTTTGCAGCAGCAGTCGCTTCACGTTCGCTCGCGTATCGGTGATTCATTCT
GCTAACCAGTAAGGCAACCCCGCCAGCCTAGCCGGGTCCTCAACGACAGGAGCACGATCATGCTAGTCATGCCCCGCGCCCACCGGAAGGAGCTGACTGGGTTGAAGGCTCTCAA
GGGCATCGGTCGAGATCCCGGTGCCTAATGAGTGAGCTAACTTACATTAATTGCGTTGCGCTCACTGCCCGCTTTCCAGTCGGGAAACCTGTCGTGCCAGCTGCATTAATGAATC
GGCCAACGCGCGGGGAGAGGCGGTTTGCGTATTGGGCGCCAGGGTGGTTTTTCTTTTCACCAGTGAGACGGGCAACAGCTGATTGCCCTTCACCGCCTGGCCCTGAGAGAGTTGC
AGCAAGCGGTCCACGCTGGTTTGCCCCAGCAGGCGAAAATCCTGTTTGATGGTGGTTAACGGCGGGATATAACATGAGCTGTCTTCGGTATCGTCGTATCCCACTACCGAGATGT
CCGCACCAACGCGCAGCCCGGACTCGGTAATGGCGCGCATTGCGCCCAGCGCCATCTGATCGTTGGCAACCAGCATCGCAGTGGGAACGATGCCCTCATTCAGCATTTGCATGGT
TTGTTGAAAACCGGACATGGCACTCCAGTCGCCTTCCCGTTCCGCTATCGGCTGAATTTGATTGCGAGTGAGATATTTATGCCAGCCAGCCAGACGCAGACGCGCCGAGACAGAA
CTTAATGGGCCCGCTAACAGCGCGATTTGCTGGTGACCCAATGCGACCAGATGCTCCACGCCCAGTCGCGTACCGTCTTCATGGGAGAAAATAATACTGTTGATGGGTGTCTGGT
CAGAGACATCAAGAAATAACGCCGGAACATTAGTGCAGGCAGCTTCCACAGCAATGGCATCCTGGTCATCCAGCGGATAGTTAATGATCAGCCCACTGACGCGTTGCGCGAGAAG
ATTGTGCACCGCCGCTTTACAGGCTTCGACGCCGCTTCGTTCTACCATCGACACCACCACGCTGGCACCCAGTTGATCGGCGCGAGATTTAATCGCCGCGACAATTTGCGACGGC
GCGTGCAGGGCCAGACTGGAGGTGGCAACGCCAATCAGCAACGACTGTTTGCCCGCCAGTTGTTGTGCCACGCGGTTGGGAATGTAATTCAGCTCCGCCATCGCCGCTTCCACTT
TTTCCCGCGTTTTCGCAGAAACGTGGCTGGCCTGGTTCACCACGCGGGAAACGGTCTGATAAGAGACACCGGCATACTCTGCGACATCGTATAACGTTACTGGTTTCACATTCAC
CACCCTGAATTGACTCTCTTCCGGGCGCTATCATGCCATACCGCGAAAGGTTTTGCGCCATTCGATGGTGTCCGGGATCTCGACGCTCTCCCTTATGCGACTCCTGCATTAGGAA
GCAGCCCAGTAGTAGGTTGAGGCCGTTGAGCACCGCCGCCGCAAGGAATGGTGCATGCAAGGAGATGGCGCCCAACAGTCCCCCGGCCACGGGGCCTGCCACCATACCCACGCCG
AAACAAGCGCTCATGAGCCCGAAGTGGCGAGCCCGATCTTCCCCATCGGTGATGTCGGCGATATAGGCGCCAGCAACCGCACCTGTGGCGCCGGTGATGCCGGCCACGATGCGTC
CGGCGTAGAGGATCGAGATCGATCTCGATCCCGCGAAATTAATACGACTCACTATAGGGGAATTGTGAGCGGATAACAATTCCCCTCTAGAAATAATTTTGTTTAACTTTAAGAA
GGAGATATACCATGGGcATGCATCACCACCACCACCACGAGAACCTGTACTTCCAGGGT<u>aaccaccaaatccacattcacgatagcgacatcgcgtttccgtgcgcgccgggtca
gagcgtgctggacgcggcgctgcaagcgggcattgaactgccgtatagctgccgtaaaggtagctgcggtaactgcgcgagcgcgctgctggatggtaacatcaccagcttcaac
ggcatggcggtgcgtagcgagctgtgcaccagcgaacaagttctgctgtgcggttgcaccgcggcgagcgacatccgtattcaaccgagcagcttccgtcgtctggacccggagg
cgcgtaagcgttttaccgcgaaagtttacagcaacaccctggcggcgccggatgtgagcctgctgcgtctgcgtctgccggttggcaagcgtgcgaaatttgaagcgggccagta
cctgctgattcacctggacgatggcgagagccgtagctatagcatggcgaacccgccgcacgaaagcgacggcatcaccctgcatgtgcgtcatgttccgggtggccgtttcagc
accattgtgcagcaactgaagagcggtgtataccctggagatcgaactgccgtttggcagcattgcgctgaaaccggacgatacccgtccgctgatctgcgttgcgggtggcaccg
gttttgcgccgattaaaagcgtgctggacgatctggcgaagcgtaaagttcagcgtgacatcaccctgatttggggtgcgcgtaacccgagcggtctgtacctgccgagcgcgat
cgataagtggcgtaaaacctggccgcaattccgttatatcgcggcgattaccgacctgggtaacgtgccggcggatgcgcatgcgggtcgtgttgacgatgcgctgcgtacccac
tttggtaacctgcacgatcacgtggttcattgctgcggcagcccgagcctggtgcagagcgttcgtaccgcggcgagcgacatgggtctgctggcgcaaaacttccacgcggatg
tgtttgcgaccagcccgaccggcagccactaa</u>CTCGAGTCTGGTAAAGAAACCGCTGCTGCGAAATTTGAACGCCAGCACATGGACTCGTCTACTAGCGCAGCTTAATTAACCTA
GGCTGCTGCCACCGCTGAGCAATAACTAGCATAACCCCTTGGGGCCTCTAAACGGGTCTTGAGGGGTTTTTTGCTGAAAGGAGGAACTATATCCGGAT

Figure S2. Sequences used in this work. Amino acid sequences of the **α (A)** and **β (B)** subunits of the wild-type dioxygenase (*Comamonas* sp. E6), with TEV protease recognition site in bold, the hexahistidine tag underlined, and stop codon represented with a *. **(C)** Nucleotide sequence for the TPADO construct, with codons for the TEV protease recognition site and 6xHis tag in bold, and the codons for the insert genes underlined and in lower case. **(D)** Amino acid and **(E)** nucleotide sequences for the reductase and its vector construct, respectively.

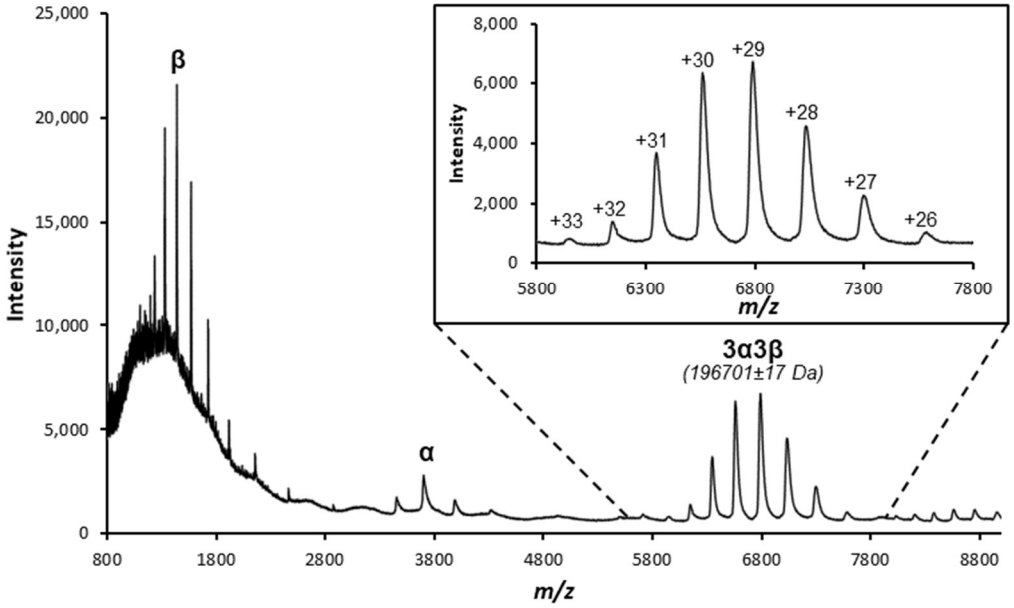


Figure S3. Native mass spectrum of TPADO. The protein has a charge envelope from +26 to +33 with an expected Gaussian distribution. The native species (spectrum shown on an amplified scale in the inset) had a deconvoluted mass of 196,915 Da corresponding to the native 3α3β complex. This is within 0.5% error of the theoretical mass of 196,012 predicted by the online ExPasy MW calculator tool (10).

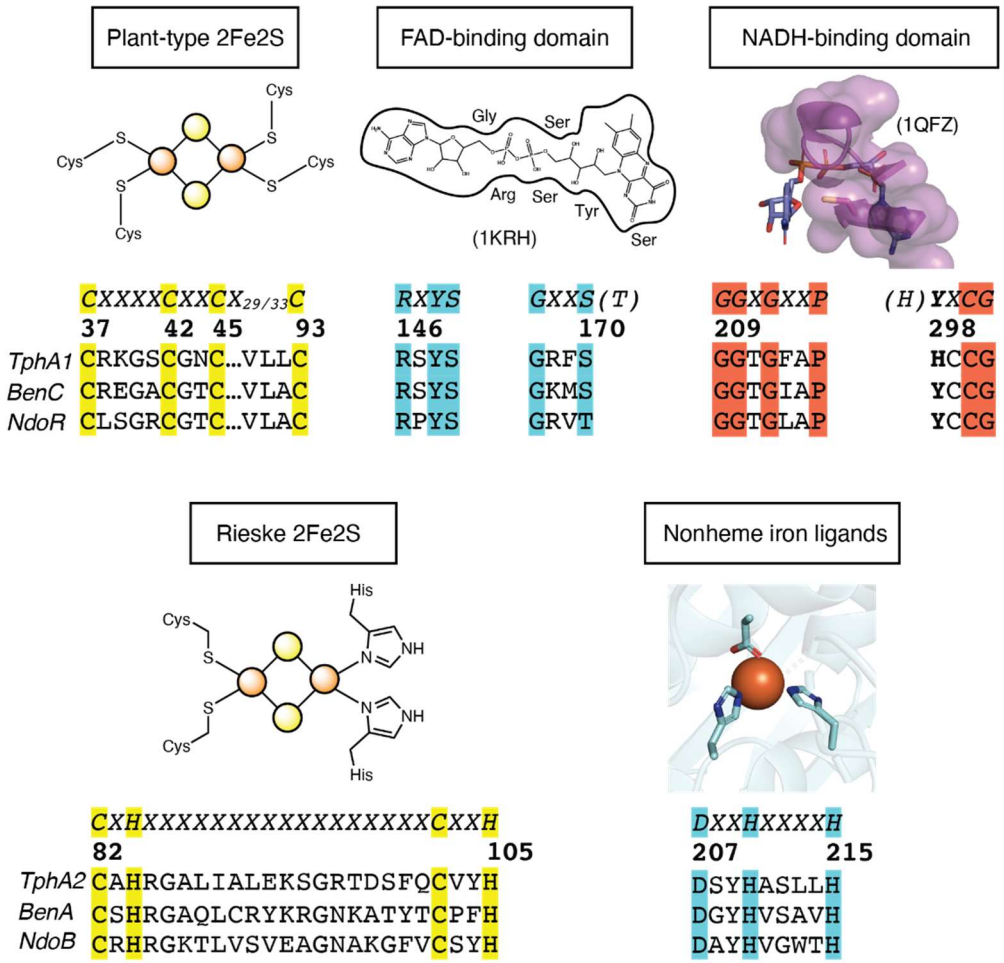


Figure S4. Conserved cofactor binding motifs in reductase and Rieske oxygenase enzymes most similar to TPADO α subunit (TphA2, bottom) and TPADO reductase (TphA1, top). Sequence alignments shown have highly conserved binding or contact residues highlighted, with a consensus sequence shown in the first line. Illustrations of how each motif is involved in cofactor binding are shown above the respective sequence alignments. BenC = benzoate dioxygenase alpha subunit; NdoB = naphthalene dioxygenase alpha subunit.

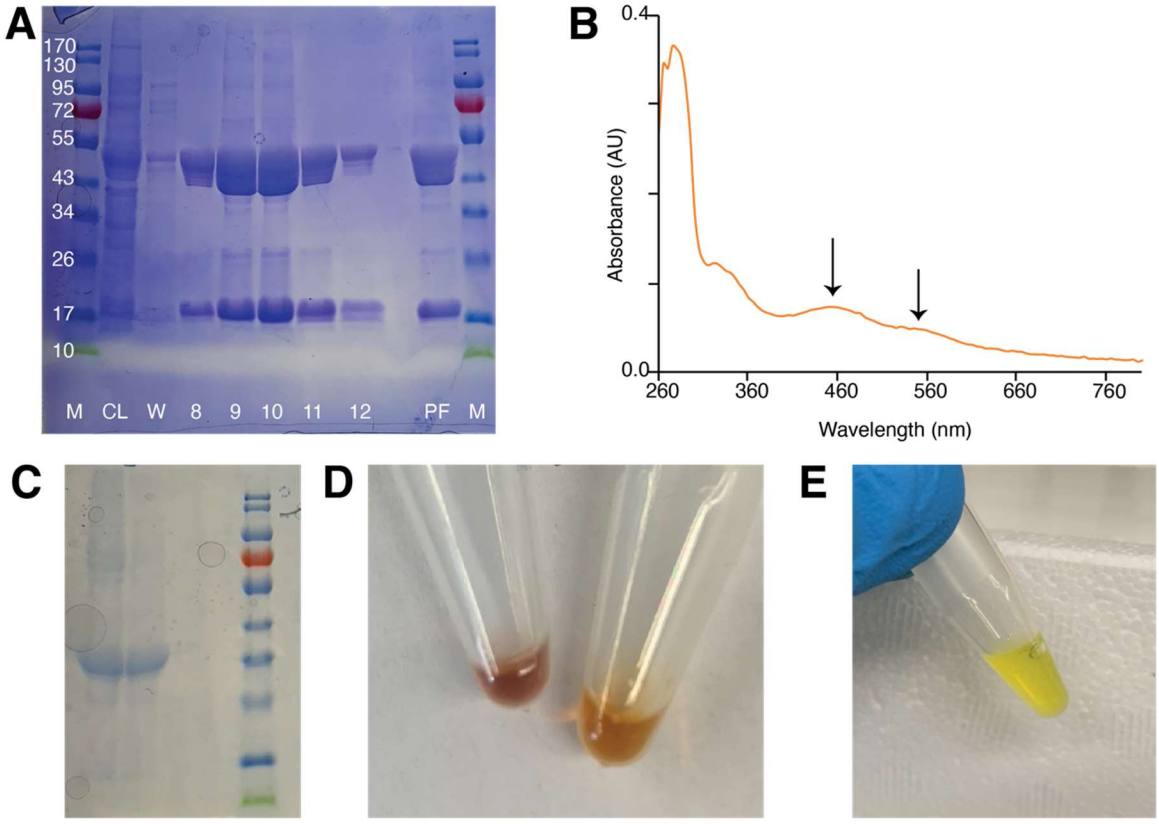


Figure S5. Isolated TPADO and reductase reflect high yield purification and the presence of cofactors. (A) SDS-PAGE and (B) UV/visible spectra of TPADO. SDS-PAGE of a molecular weight size marker (M, in kDa), clarified lysate (CL), wash (W), and Ni-NTA column eluate fraction (8-12) clearly indicates enrichment in the heavier alpha and lighter beta subunits of TPADO. The pooled fractions (PF) obtained were buffer exchanged to remove imidazole (>95% purity) and were used for the production of crystals. UV/visible solution spectrum of pure TPADO shows the expected peak maxima at 450 and 550 nm for a [2Fe-2S] cluster and mononuclear non-heme iron, respectively. The peak at ~300 nm may also be attributed to the mononuclear non-heme iron. Atomic absorption carried out on 3 independent preparations of the TPADO enzyme contained 3.5 (± 0.08), 3.1 (± 0.05), and 2.9 (± 0.06) atoms of iron per TPADO $\alpha\beta$ monomer, where the latter protein concentration was determined by Bradford assay for the complex of known molecular weight. Errors represent one standard deviation based on averaged measurements of 3 individual dilutions. (C) SDS-PAGE of reductase, showing a band near the expected mass of 36 kDa. (D) Samples of TPADO (left) and reductase (right). (E) Reductase protein was precipitated with 3% trichloroacetic acid (w/v %), and yellow flavin was extracted from the soluble portion.

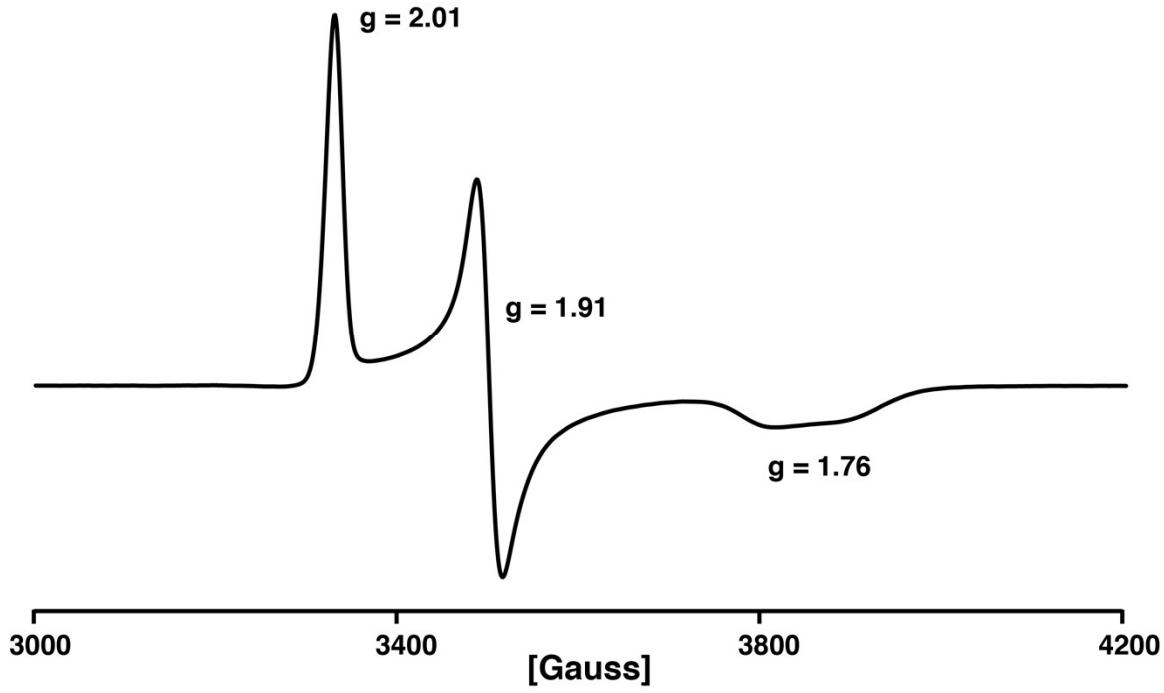


Figure S6. EPR signal from the dithionite treated (reduced) Rieske cluster of TPADO that displays the characteristic features (g = 2.01, 1.91, and 1.76) of this cofactor. Conditions: microwave power, 2.0 mW; temperature, 50 K; microwave frequency, 100 KHz (11).

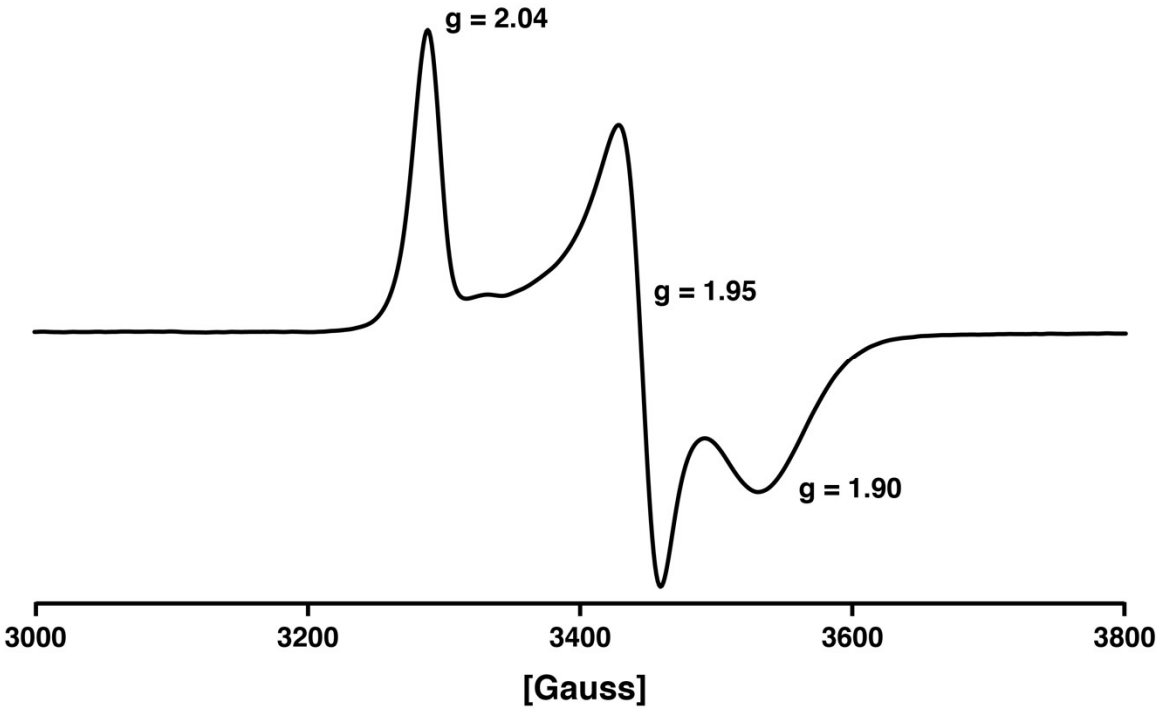


Figure S7. EPR signal from the NADH treated (reduced) Plant-type cluster of the reductase, displaying the characteristic features of this cofactor (g = 2.04, 1.95, and 1.90). Conditions: microwave power, 2.0 mW; temperature, 50 K; microwave frequency, 100 KHz (12).

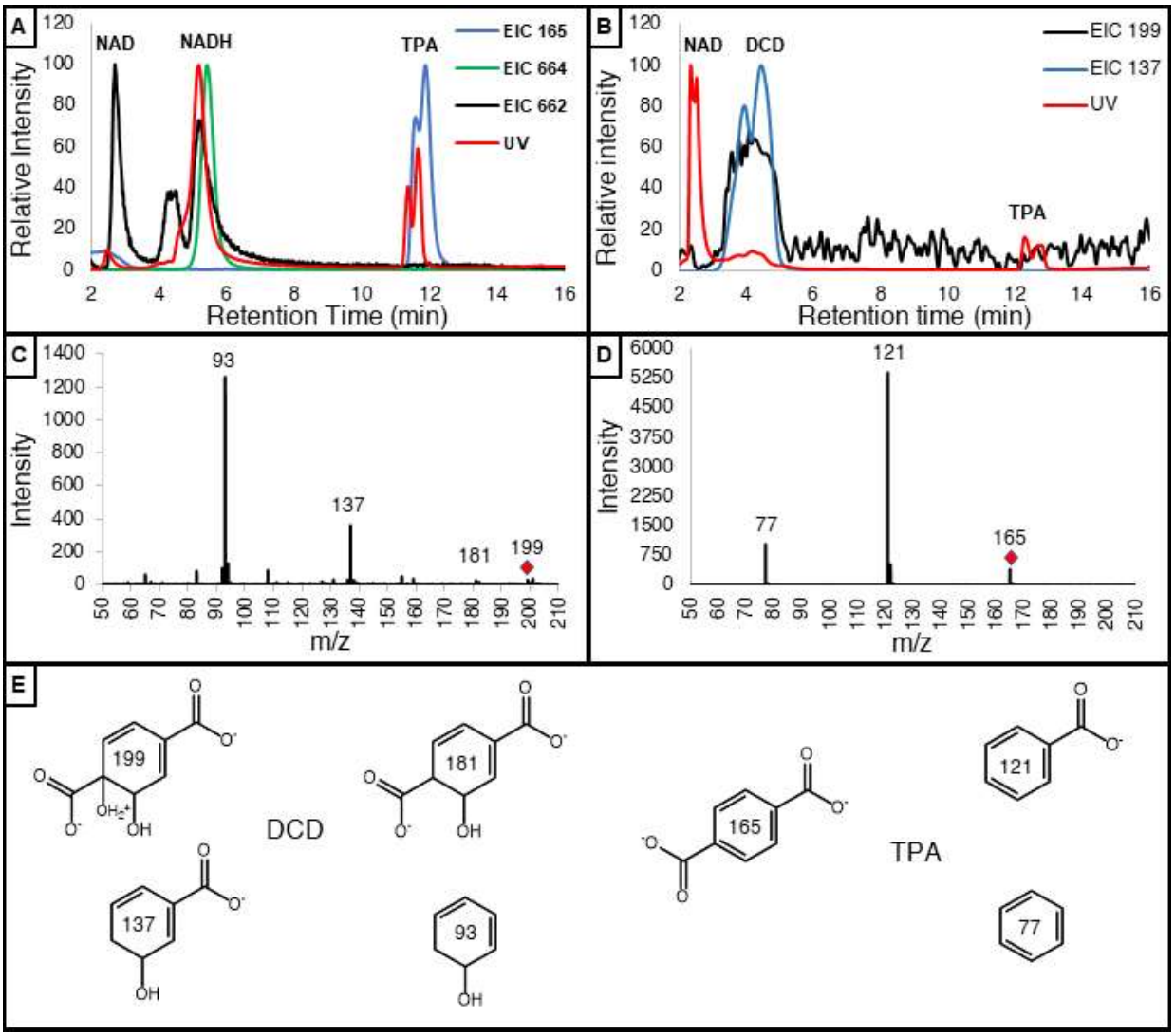


Figure S8. Extracted ion chromatograms and tandem mass spectra of TPA and DCD ions. (A) Extracted ion chromatograms (EIC) of a reaction mixture containing TPA and NADH, prior to the addition of enzymes. The expected molecular ions for NAD^+ (M-H- m/z = 662), NADH (M-H- m/z = 664) and TPA (M-H- m/z = 165) were detected. Further investigation of the spectra containing the TPA molecular ion revealed a coeluting ion of greater intensity with an m/z = 121 corresponding to loss of CO_2 (Δm=44). There is no observable peak corresponding to the molecular ion of DCD. Overlaying the UV/vis chromatogram indicates the detectability of TPA, NADH, and NAD^+. **(B)** EICs for a reaction mixture containing TPA and NADH following incubation with TPADO and reductase for 90 minutes. The same three ions, plus a species at m/z 199 which is the expected mass of the molecular ion of DCD (M-H- m/z = 199), were detected. The intensity of the ion at m/z 199 was low; however, a series of ions was observed to co-elute, the most intense ion being m/z = 137. This ion was hypothesized to come from in-source fragmentation leading to the loss of H_2O + CO_2 (Δm=62) from DCD. **(C)** A tandem mass spectrum of the parent ion of DCD (diamond), showing a fragmentation pattern consistent with the fragments observed in source ionization in the EIC above. **(D)** A tandem mass spectrum of the parent ion of TPA (diamond), showing a fragmentation pattern consistent with the fragments observed in source ionization in the EIC above. **(E)** Predicted fragment ions for DCD (left) and TPA (right). m/z values are inscribed inside each molecule. See Table S1 for exact masses of all ion species described here.

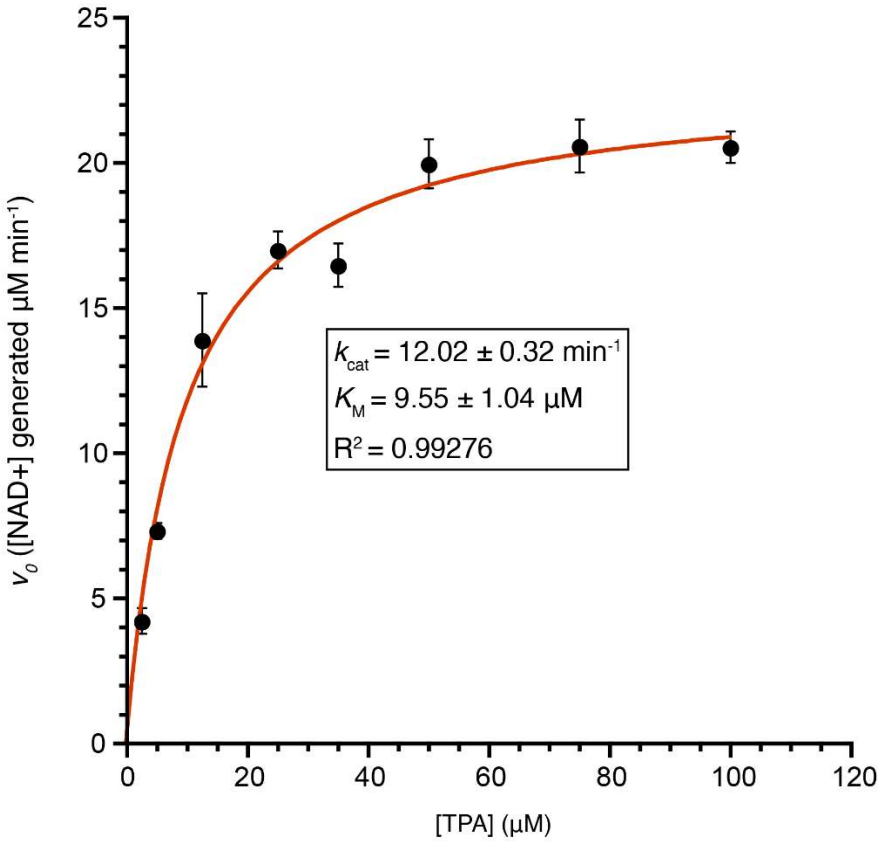


Figure S9. Apparent steady-state kinetic parameters of TPADO. Buffered solutions of TPADO (1.9 µM), reductase (0.6 µM), and NADH (200 µM) were reacted with varying concentrations of terephthalate (2.5–100 µM) in a quartz cuvette (1 cm path length, concentrations after mixing). Solutions were prepared in ambient air (~230 µM O_2) and contained 20 mM Tris, 150 mM NaCl, 20% DMSO (v/v), 10% glycerol (v/v), pH 8.0 buffer at 28 ºC. The initial rates of reactions (v_0) were obtained from the slope of the linear, first 5-10% of the time course using the extinction coefficient for NADH at 340 nm (6.22 mM^{-1} x cm^{-1}). Averages of 3 rates are plotted with error bars representing ±1 standard deviation. The red line is a fit of the data to the Michaelis-Menten hyperbolic expression.

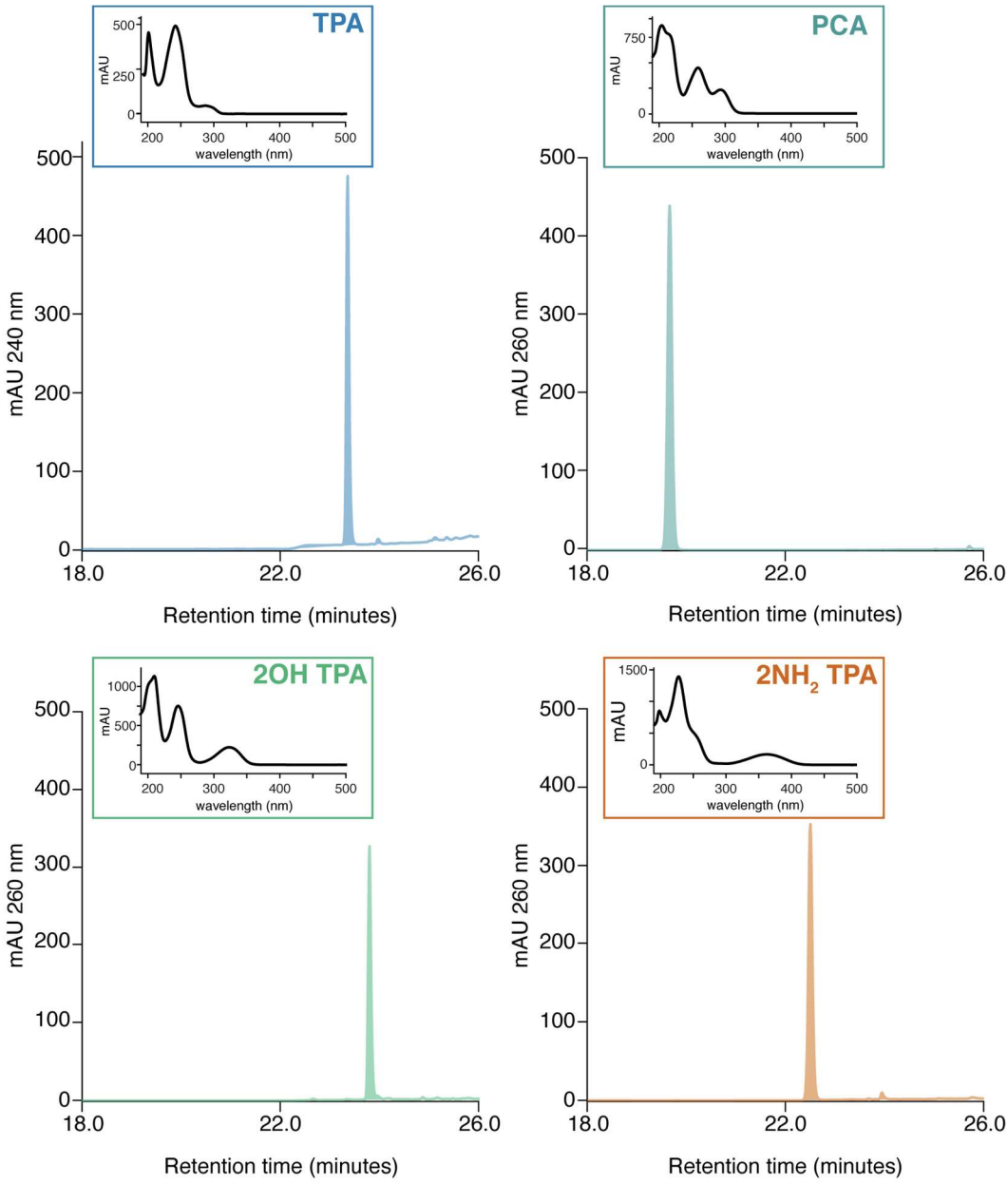


Figure S10. Standards samples of (A) TPA, (B) PCA, (C) 2-OH-TPA, and (D) $2NH_2$-TPA resolved via reverse-phase HPLC according to the program described in *Detailed experimental methods*. Retention times are given in Table S3.

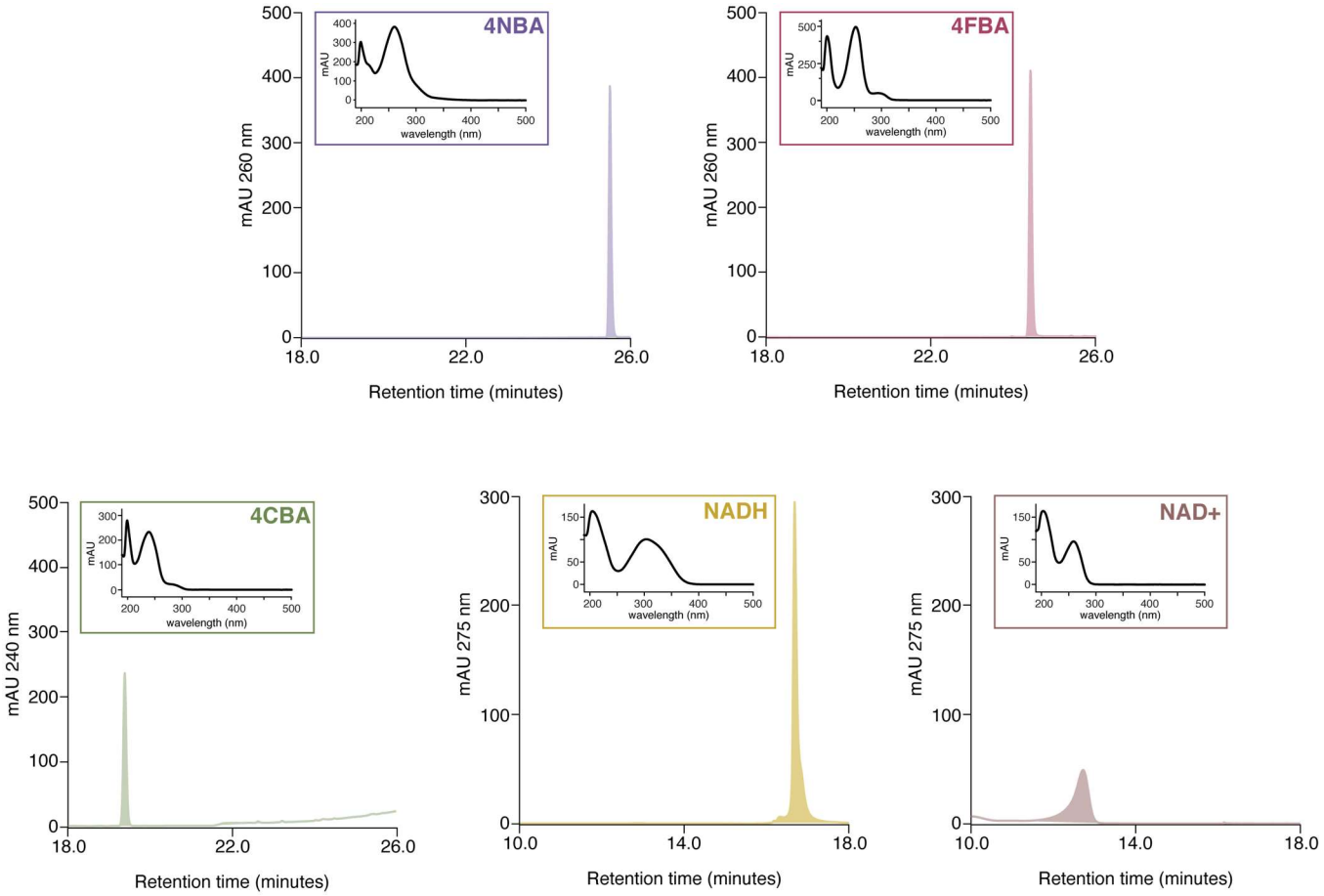


Figure S11. Standards of (A) 4-NBA, (B) 4-FBA, (C) 4-CBA, (D) NADH, and (E) NAD+ resolved via reverse-phase HPLC according to the program described in *Detailed experimental methods*. Retention times are given in Table S3.

\

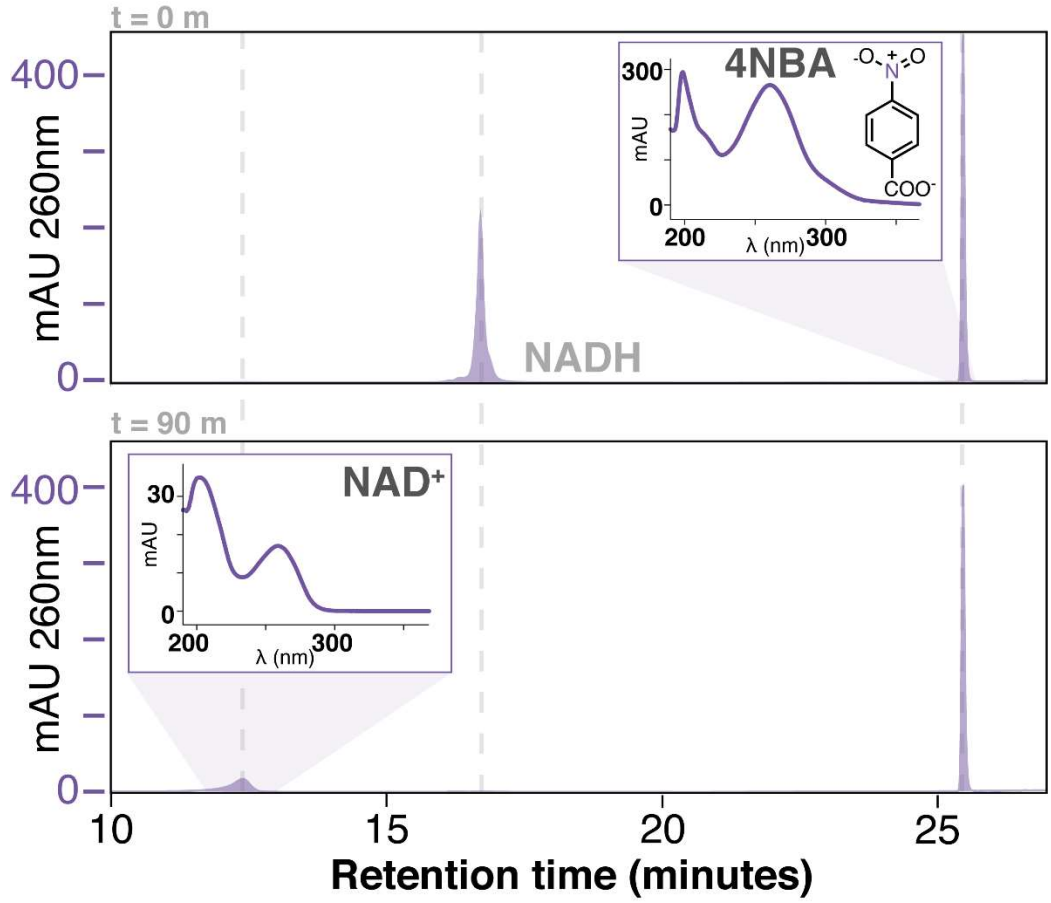


Figure S12. HPLC chromatograms monitoring 4-nitrobenzoate (4-NBA) and NADH before (top) and after incubation with TPADO/reductase. Front, t = 0 sample. Back, sample quenched at 90 minutes. These data indicate the complete conversion of NADH to NAD^+ in the presence of enzyme and 4-NBA, without the formation of any 4-NBA associated product.

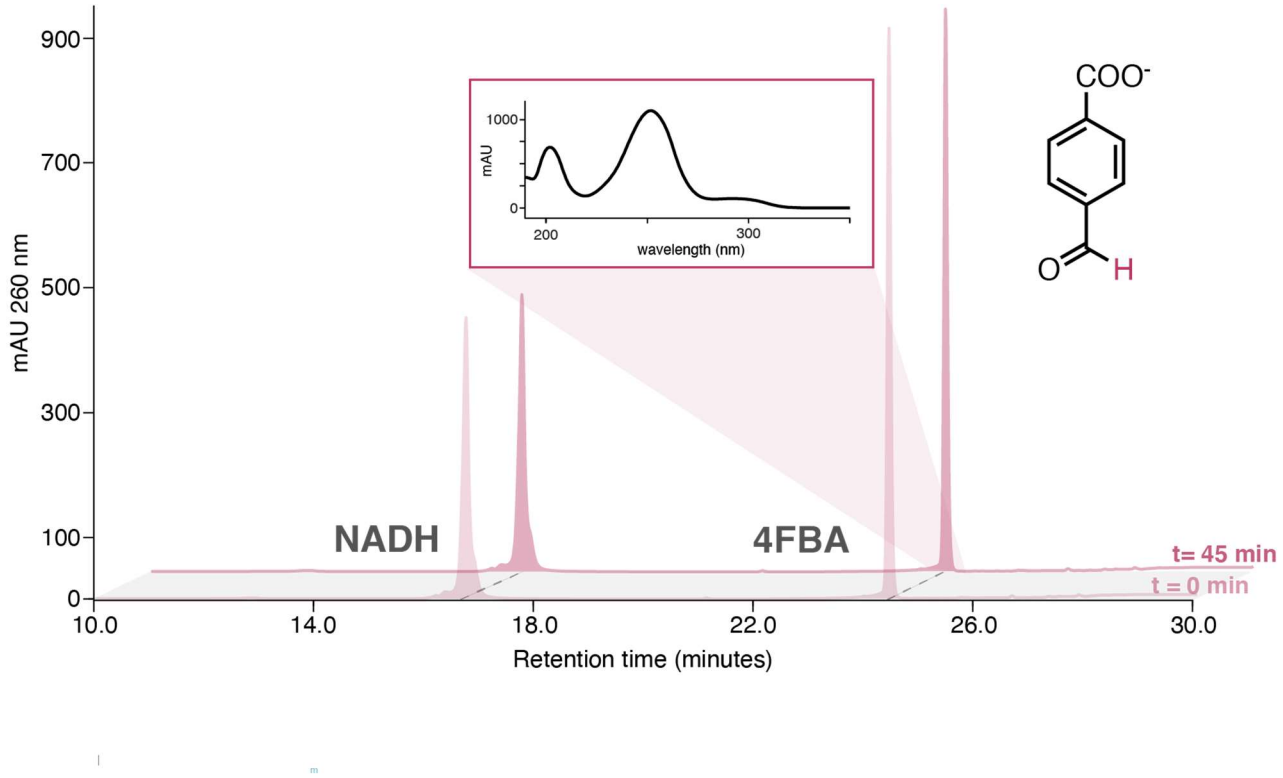


Figure S13. HPLC chromatograms monitoring 4-formylbenzoate (4FBA) and NADH before and after incubation with TPADO/reductase. Front, t = 0 sample. Back, sample quenched at 45 minutes. Inset, extracted UV/vis spectrum from the 4FBA peak. There were no changes in the NADH and 4FBA peak areas that were above background. Additionally, no quantifiable NAD+ production was observed.

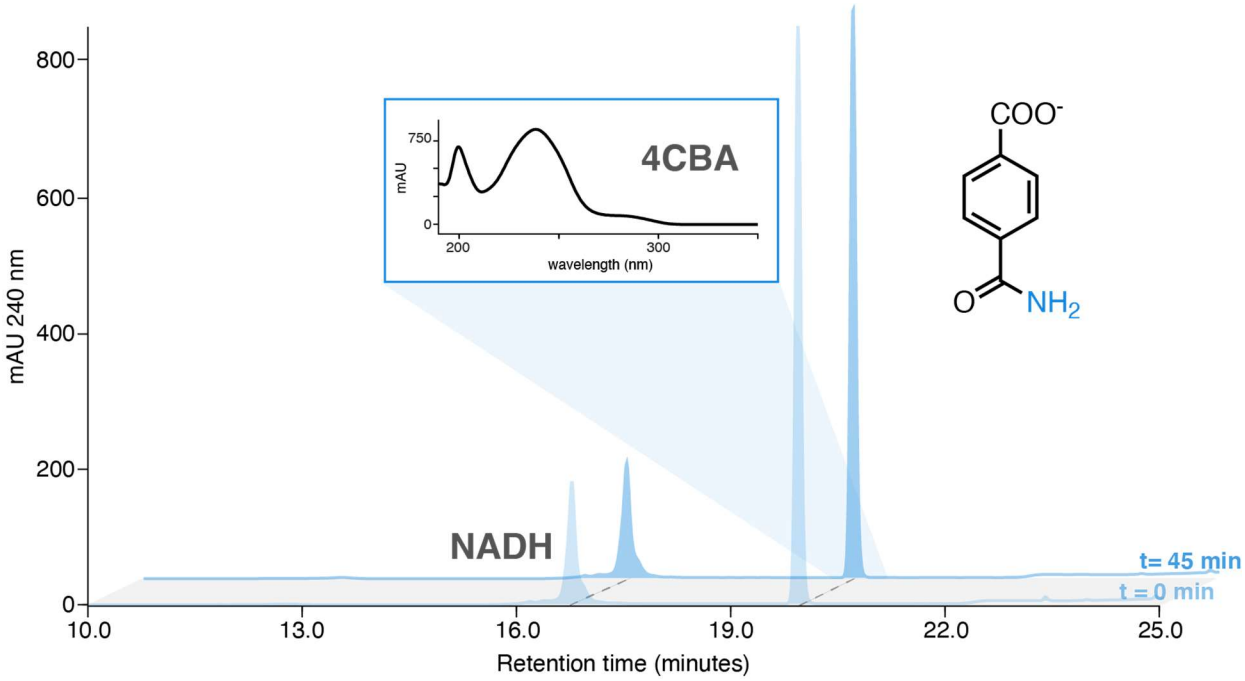


Figure S14. HPLC chromatograms monitoring the 4-carbamoylbenzoate (4CBA) and NADH before and after incubation with TPADO/reductase. Front, t = 0 sample. Back, sample quenched at 45 minutes. Inset, extracted UV/vis spectrum from the 4CBA peak. There were no changes in the NADH and 4CBA peak areas that were above background. Additionally, no quantifiable NAD+ production was observed.

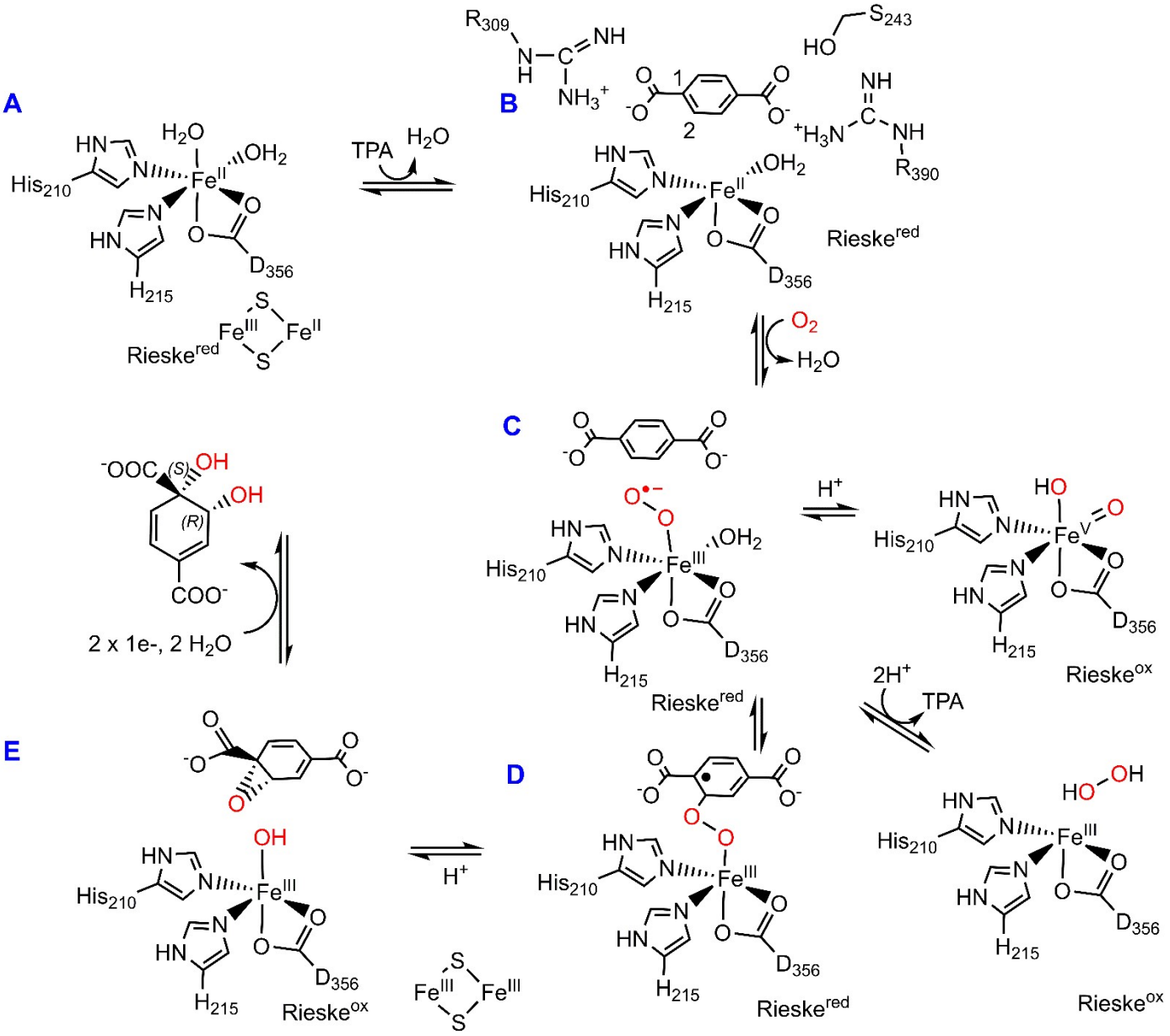


Figure S15. A canonical mechanism for ROs adapted for TPADO. A. The fully reduced, resting active site has a ferrous hexacoordinate mononuclear iron and reduced Rieske cluster ($Rieske^{red}$). B. TPA binding displaces a water molecule and primes the enzyme for reaction with O_2. The positions of residues involved in anchoring the substrate (R390, R309, S243) are indicated. The –R group illustrates the 2-position of 2-OH-TPA in the cocrystal structure with TPADO, point toward a hydrophilic pocket (not shown). C. O_2 binds in an end-on fashion to form the one-electron reduced ferric-superoxy intermediate, currently proposed by Lipscomb et al. as the catalytically reactive species. One electron is donated from the Rieske cluster to form a peroxy adduct, which can either heterolytically cleave to yield a high valent ferryl, or release H_2O_2 following di-protonation. D. Attack of the superoxide yields a radical intermediate. E. Electron transfer from the Rieske cluster supports formation of an epoxide and ferric-hydroxy intermediate. The ferric-OH attacks the epoxide, which opens and protonates to form the product. Re-reduction of the active site occurs as the product departs. The carboxylate substituent on C1 is shown maintaining a salt bridge with R390. The *cis-1S,2R*-DCD isomer is drawn as the presumptive product, based on the observed products of the reaction between NADH/O_2 and $2NH_2$-TPA. However, the absolute stereochemistry of the product has not yet been determined.

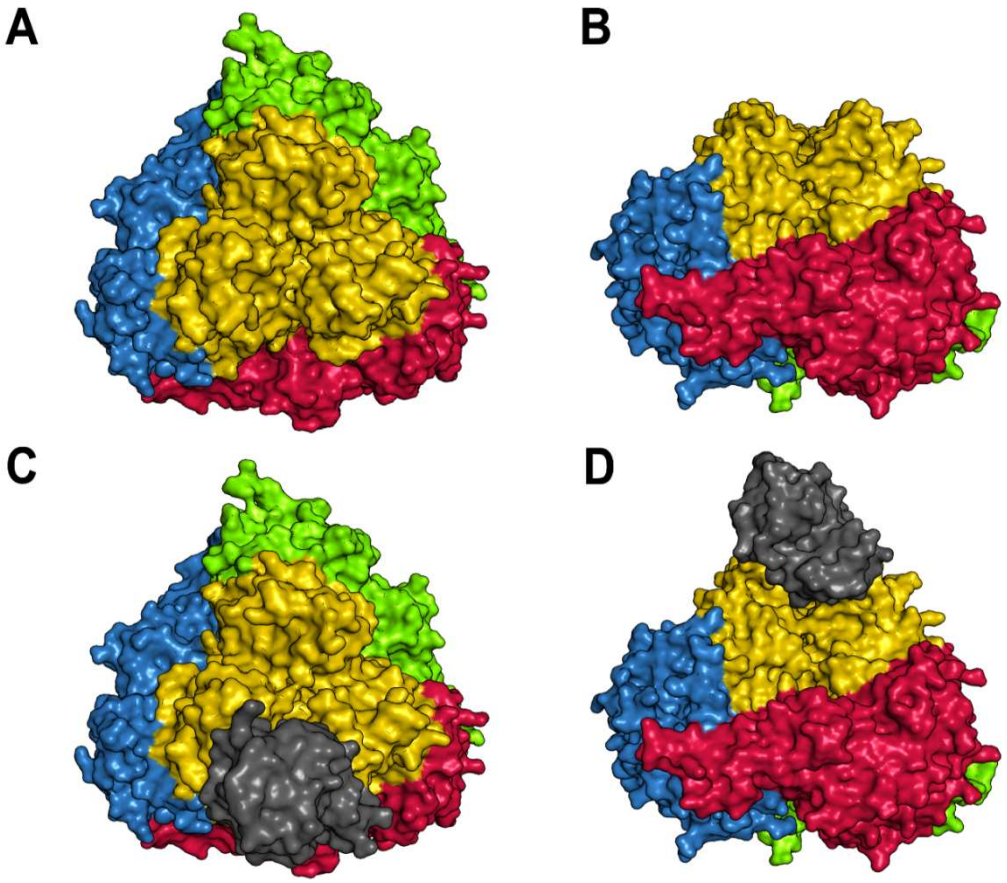


Figure S16. Lysozyme binding to the TPADO structure. A) Top and B) Side view of the TPADO $\alpha_3\beta_3$-heterohexamer. The α-subunits are colored in red, green and blue. The β-subunits are colored in yellow. C) Top and D) Side view of the TPADO $\alpha_3\beta_3$-heterohexamer including the bound lysozyme protein in gray color.

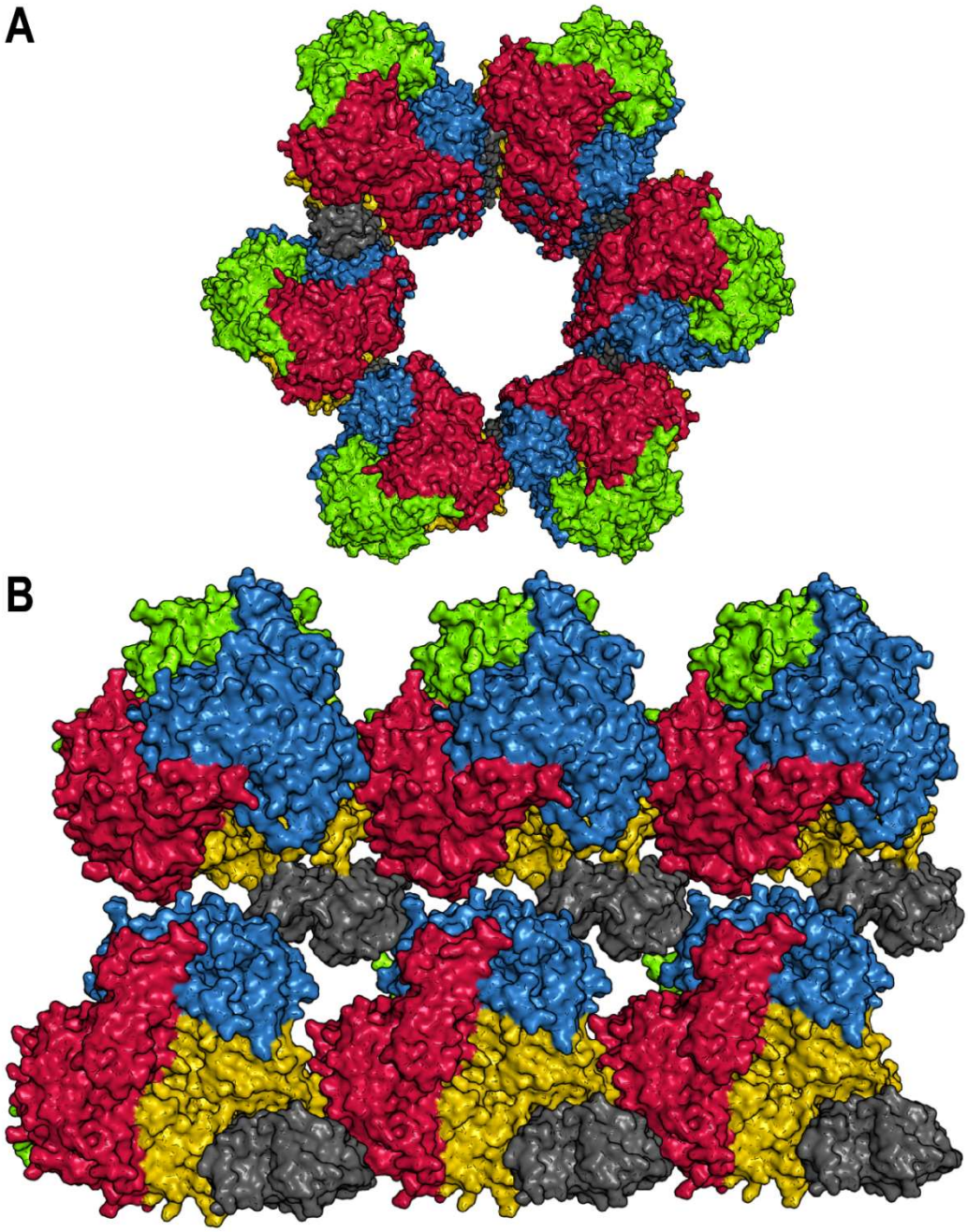


Figure S17. TPADO crystal packing interactions viewed A) along the hexagonal hole and B) inside the hexagonal hole show the influence of lysozyme (gray) as a crystallization chaperone to form stable crystal contacts.

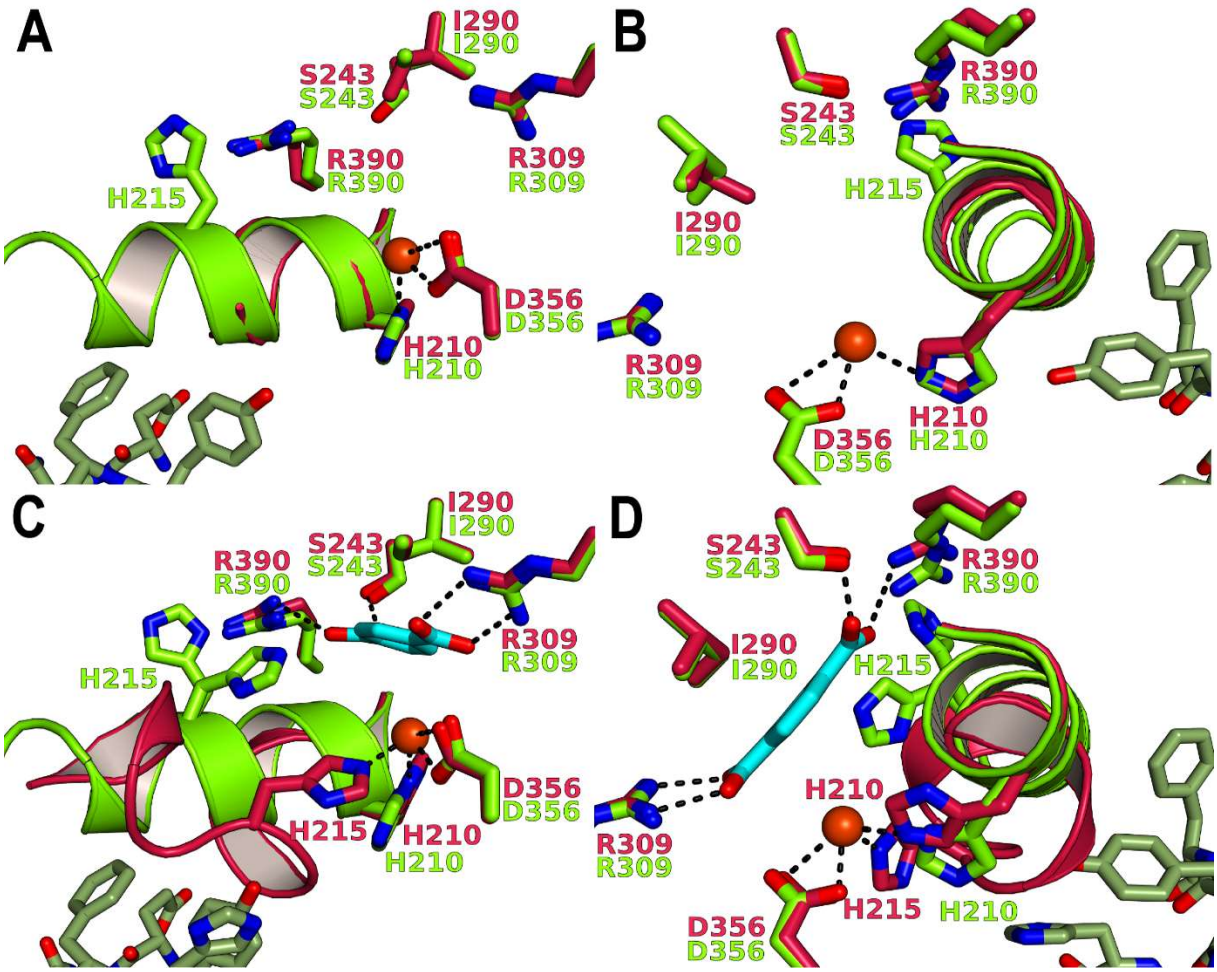


Figure S18. Superposition of the catalytic active sites between two TPADO alpha subunits in the ligand-free state (A and B) as well as in the TPA bound state crystal structure (C and D). For the light green colored alpha subunit, an alpha helix with residues 208 to 220 is stabilized by crystal packing interactions (residues of a neighboring symmetry related molecule are shown in dark green). The ferrous ion coordinating histidine residues H210 and H215 are located on this alpha helix. In the ligand-free state, the non-stabilized alpha helix within the red colored alpha subunit adopts the same alpha-helical conformation as the stabilized green helix, but is disordered from residue H215 onwards. Residue H215 does not coordinate the ferrous ion in the green colored alpha subunit. The H210 side chain residues of both the red and green colored alpha subunits align very well. In the TPA-bound state, the non-stabilized red alpha-helix unfolds upon TPA binding and enables the coordination of the H215 side chain with the ferrous ion. For residue H210, the peptide backbone still aligns very well between both alpha subunits, but the side chain conformation differs to avoid steric clashes with residues of the unfolded helix. Substrate TPA is shown in light blue. Hydrogen bonding interactions are only shown for the red colored alpha subunit for both the Apo and the TPA-bound state.

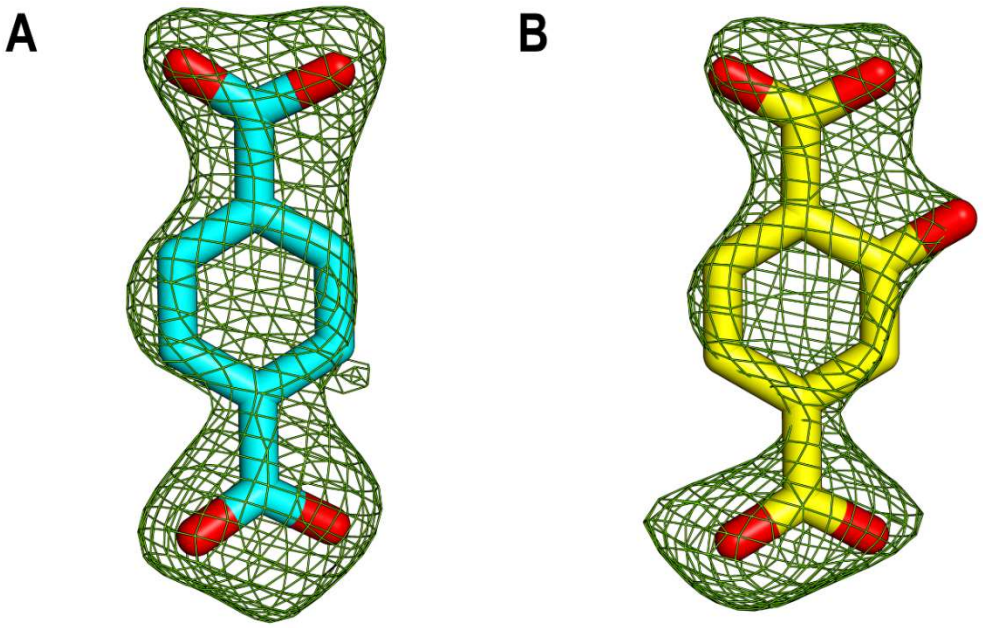


Figure S19. Omit density maps for the substrates A) TPA and B) 2OH_TPA at 2.5 σ level.

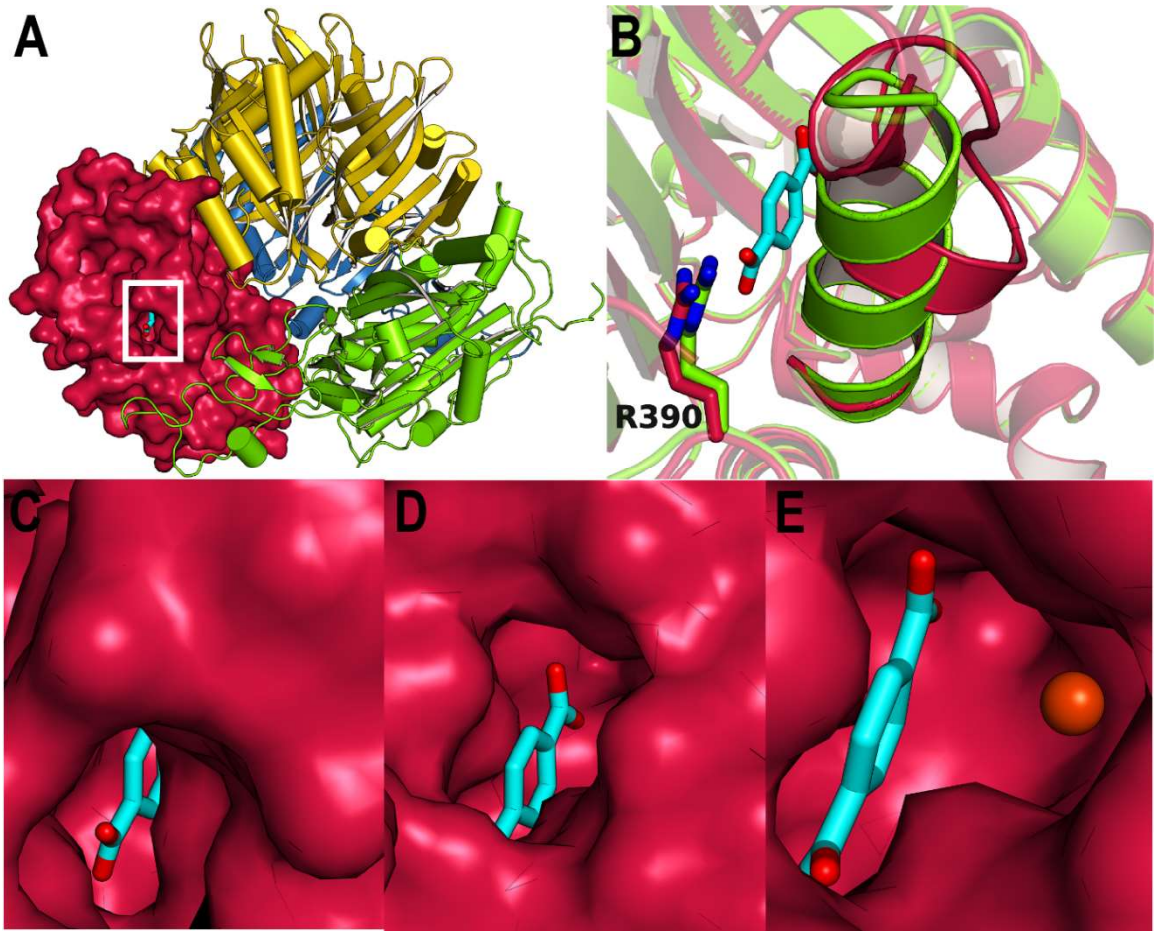


Figure S20. Visualization of the potential substrate entry/product release site of TPADO. A) Side view of the TPADO $\alpha_3\beta_3$-heterohexamer showing the red colored alpha subunit as surface presentation to locate the potential substrate entry site of TPA (white square). TPA is colored in light blue. B) Superposition of the green and red colored alpha subunits show that substrate access is gated by arginine residue R390 as well as the conformation of an α-helix which contains the ferrous ion coordinating histidine residues H210 and H215. C) A solvent accessible surface rendering of the TPADO ligand-bound structure reveals that TPA is accommodated in a deep and partially closed pocket (area shown represents the white square from A). The active site of the TPA-bound state is partially accessible due to the flexible residues 223 to 226 which were not visible within the electron density. D and E) Two alternative viewing angles are shown of the Apo structure with TPA and the ferrous ion modelled from the TPA bound state. The TPA binding pocket is more open within the substrate-free structure because the α-helical residues from H215 onwards are disordered and not part of the model (in total residues 215 to 227 are missing). The TPA binding pocket is completely closed within the green colored alpha subunit as well as for the red-colored alpha subunit of the 2-OH-TPA-bound structure in which all residues are defined within the electron density (not shown)

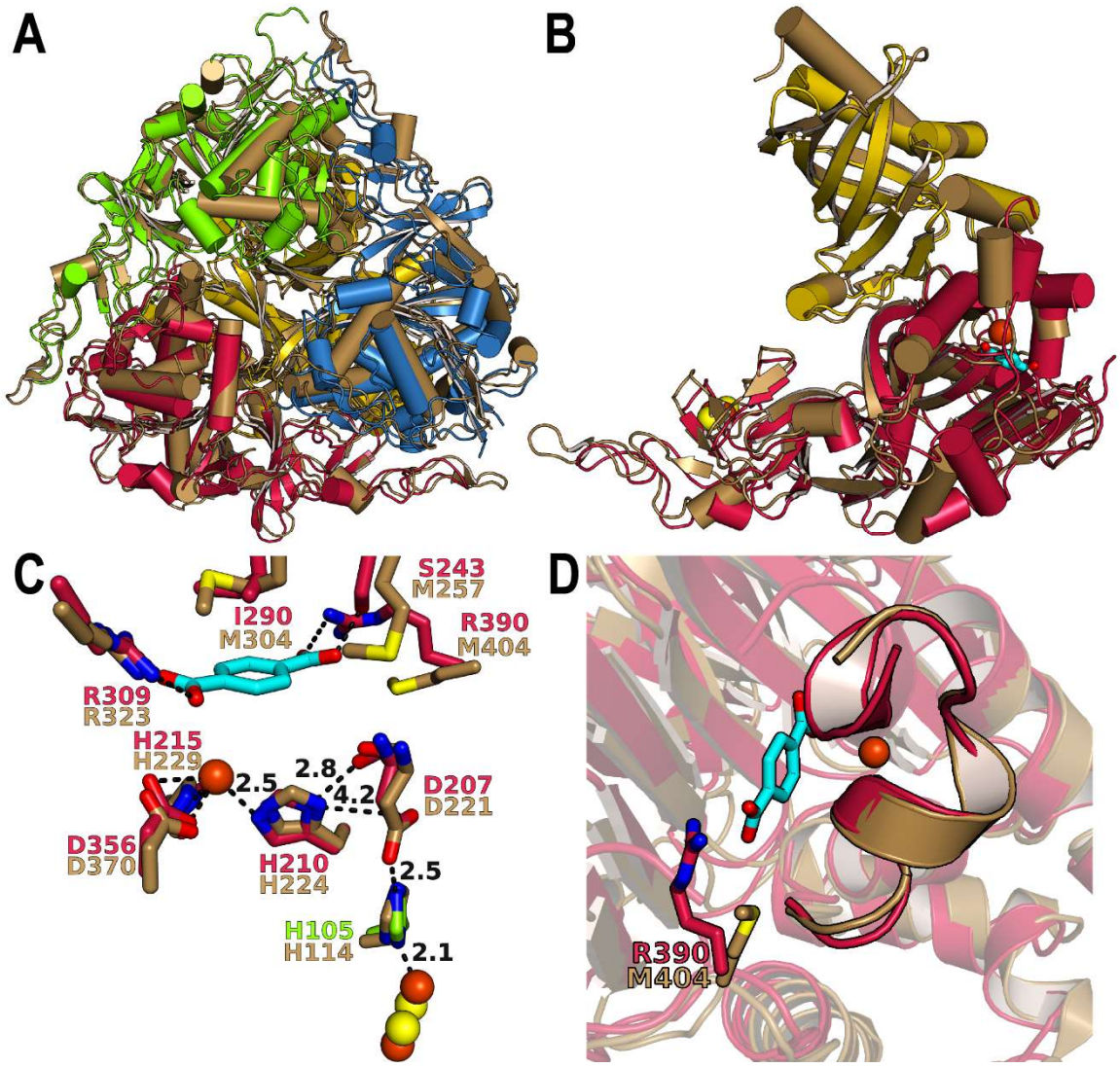


Figure S21. Structural comparison of TPADO with its closest structural homolog salicylate 5-monooxygenase (NagGH). The TPA-bound red colored TPADO alpha subunit was superimposed to an NagGH alpha subunit (pdb 7C8Z). A) Bottom view of the $\alpha_3\beta_3$-heterohexamers shows that the overall quaternary architecture between TPADO and NagGH is very similar. The TPADO alpha subunits are colored in red, green and blue. The TPADO beta subunits are colored in yellow. NagGH is colored in beige. B) Side view of the superposition shown for one αβ-subunit. C) Zoom into the catalytic active site shows that the residues coordinating the ferrous ion as well as connecting the [2Fe-2S] cluster of a neigboring alpha subunit with the ferrous ion align very well. Hydrogen bonds are only shown for TPADO for better visibility. Residue H105 belongs to a neighboring subunit and is colored in green. For residue D207, the hydrogen bonding distances of both the carbonyl oxygen and the side chain Oδ-atom are plotted to the Nδ atom of the H210 side chain. Atom distances are shown for the hydrogen bonds connecting the ferrous ion with the [2Fe-2S] cluster. D) The alpha helices which contain the ferrous ion coordinating histidine residues align very well between both structures. The substrate entry site has a methionine in NagG in contrast to TPADO which harbors an arginine residue.

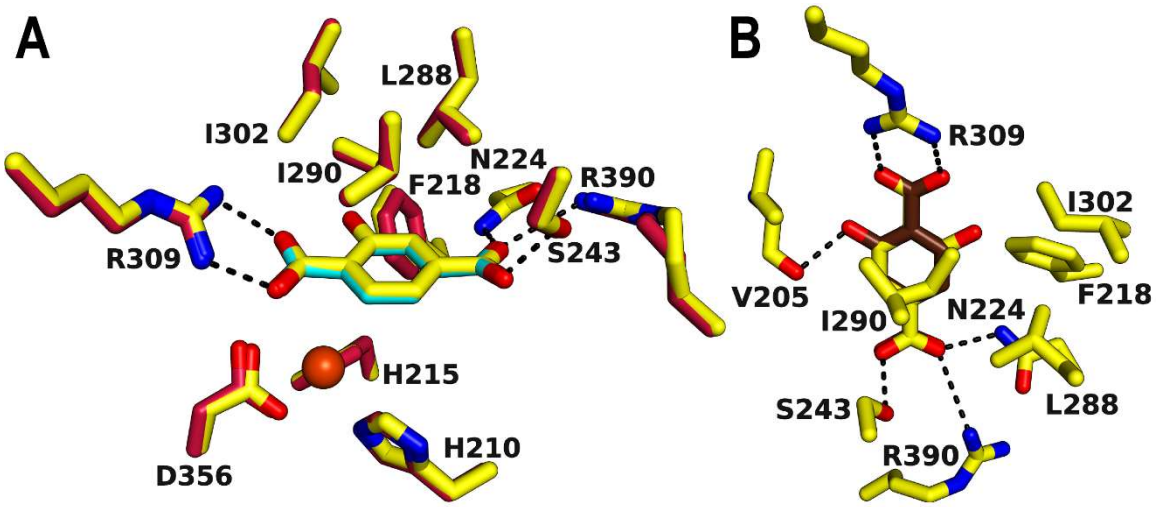


Figure S22: Comparison between the TPA and 2-OH-TPA bound structures. A) Superposition of the 2-OH-TPA-bound TPADO structure (yellow) with the TPA-bound structure (protein in red, TPA in light blue) shows that the ligands and the ligand interacting residues align very well. The 2-hydroxy-group of 2-OH-TPA is oriented towards the hydrophobic part of the catalytic site formed by residues F218, L288, I290 and I302. Residues beyond the alpha-helix ending with residues 208 to 220 are visible within on out of three alpha-subunits with N224 forming a hydrogen bond with one of 2-OH- TPA's carboxylate groups. B) Ligand 2-OH-TPA (yellow) was flipped (ligand model brown) to show that the 2-hydroxy-group would be able to form a hydrogen bond with the carbonyl oxygen of residue V205 but the electron density strongly suggests that the hydroxyl-group is oriented towards the hydrophobic side instead.

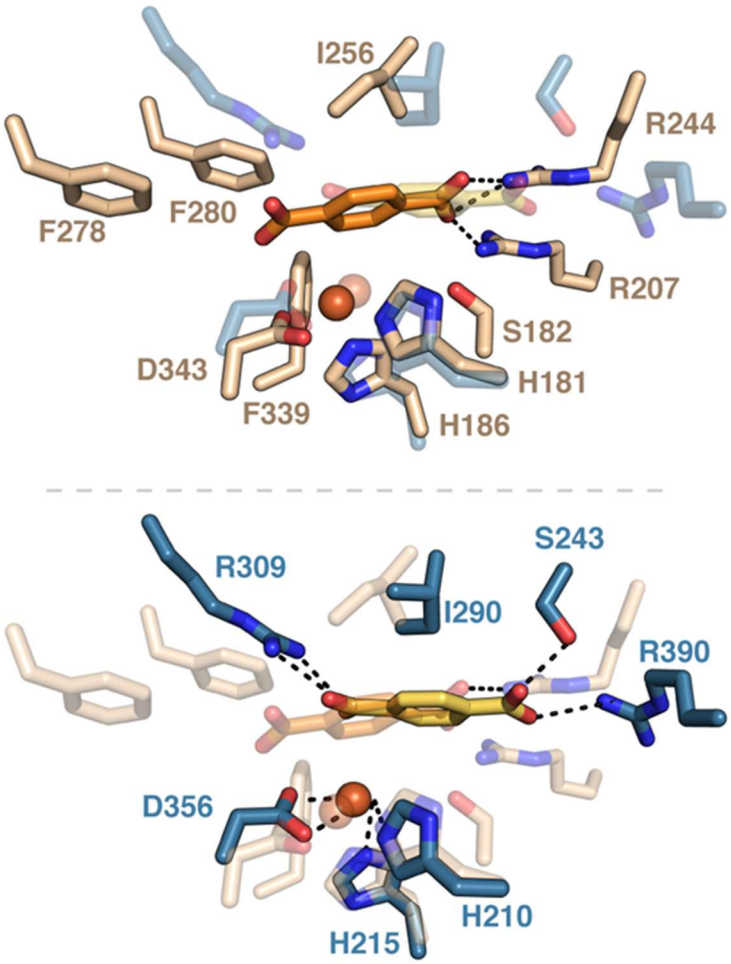


Figure S23. Superposition of the active site structures for the *ortho*-phthalate dioxygenase (PDO) from *Comamonas testosteroni* KF1 (PDB ID 7V28, carbon in steel blue, SI ref 13) and the TPADO from *Comamonas* strain E1 (PDB ID 7Q05, carbon in tan). Bound TPA is shown in the center of both structures. The top image has the PDO structure rendered at full opacity, with TPA in orange, while the bottom image has the TPAD at full opacity, with TPA in yellow. The α subunits, on the whole, overlay poorly (RMSD = 3.21Å), and their primary sequences share 19% similarity. Dashed lines indicate electrostatic and/or hydrogen bonding interactions that anchor the TPA into the respective active sites. In PDO, one of the TPA carboxylates is located in a hydrophobic pocket, while the other carboxylate forms 3 interactions with R244 and R207. In TPADO, the TPA is multiply anchored via the carboxylates at both ends of the substrate. TPA dioxygenation by PDO occurred with 20% coupling to NADH consumption, with a 25-fold reduction in k_{cat}/K_M relative to the preferred *ortho*-phthalate substrate.

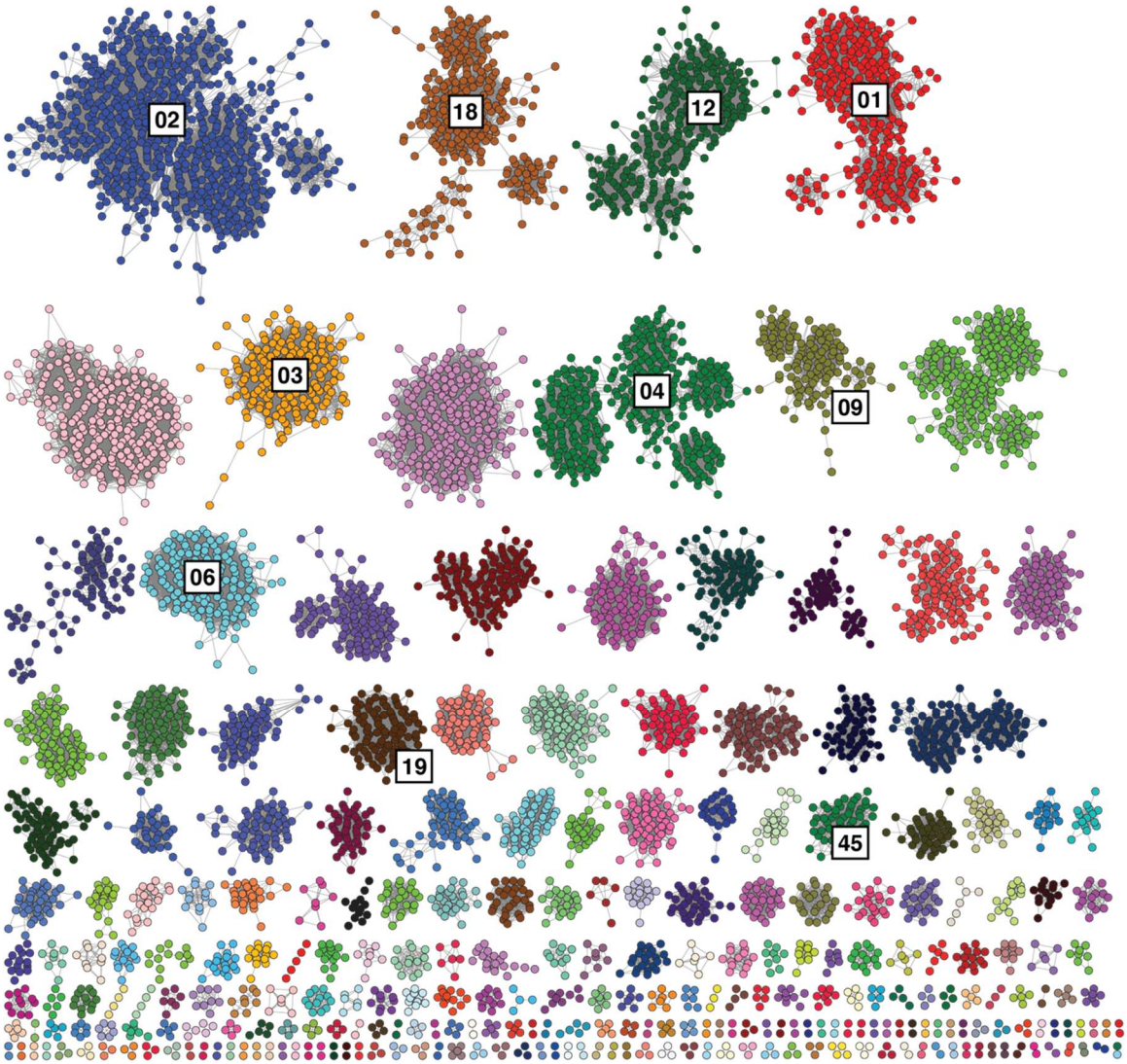


Figure S24. Color sequence similarity network of alpha subunit proteins from aromatic ring-hydroxylating Rieske oxygenases (Pfam PF00848). The network was generated by the EFI-EST server with a Uniref90 filter, an E-value of 5, and a final alignment score of 85. Clusters containing at least one sequence of a functionally verified protein are numbered. The numbers are associated with the "Sequence Count Cluster Number" attribute bestowed by the EFI-EST server. Table S4 lists representative UniProt IDs from each cluster that are associated with proteins/genes with published activities. In general cluster 1 contains glycine betaine oxidative demethylase-like proteins; 2 contains naphthalene/biphenyl/benzene/polyaromatic dioxygenase-like proteins; 3 contains benzoate and orthohalobenzoate-like proteins; 4 contains gamma glutamylanillidine dioxygenase-like proteins; 6 contains TphA2 and other monoaryl carboxylate oxygenase-like proteins; 9 contains carnitine monooxygenase-like proteins; 12 contains choline monooxygenase-like proteins; 19 contains p-cumate dioxygenase-like proteins; and 45 contains 2-aminobenzesulfonate oxygenase-like proteins.

Table S1. HPLC analysis parameters obtained from standards†

Compound	retention time (min)	λ_{maxima}
DCD	11.7	270
NAD+	12.7	259
NADH	16.7	261, 320
PCA	19.7	259, 294, 331
TPA	23.3	240, 285
2OH-TPA	23.8	247, 325
$2NH_2$-TPA	22.5	228, 362
4NBA	25.5	261
4FBA	24.4	253, 294
4CBA	19.9	239, 284

†Retention time and spectral maximum are reported for DCD generated as the product of the TPADO catalyzed reaction, as no standards for this compound are available. It is assigned on the basis of its exact mass obtained by high resolution ESI-MS. (Stereochemistry not determined.)

Table S2: Crystallographic data and refinement statistics.

	Substrate-free state	TPA bound state	2-OH-TPA bound state
Data collection			
Beamline	DLS I03	DLS I03	DLS I03
Space group	$P6_1$	$P6_1$	$P6_1$
Cell dimensions			
a, b, c (Å)	220.8 / 220.8 / 84.1	219.7 / 219.7 / 83.0	220.3 / 220.3 / 83.2
α, β, γ (°)	90.0 / 90.0 / 120.0	90.0 / 90.0 / 120.0	90.0 / 90.0 / 120.0
Resolution (Å)	110.40 – 2.28 (2.44 – 2.28) [a]	190.28 – 2.08 (2.25 – 2.08)	190.80 – 1.95 (2.06 – 1.95)
R_{merge} [%]	23.9 (276.7)	16.7 (251.4)	10.9 (237.2)
R_{pim} [%]	5.3 (60.5)	3.7 (55.7)	2.7 (56.0)
<I/σI>	11.0 (1.4)	14.0 (1.6)	17.9 (1.4)
Completeness (%)[b]	94.3 (63.3)	96.0 (65.5)	96.4 (57.4)
Redundancy	21.1 (21.8)	21.3 (21.4)	17.3 (18.5)
CC(1/2)	0.998 (0.648)	0.999 (0.667)	0.999 (0.590)
Refinement			
R_{work} / R_{free}	18.0 / 22.5	16.8 / 21.4	16.5 / 20.2
Ramachandran plot			
most favored [%]	95.2	95.8	96.5
allowed [%]	4.4	4.0	3.4
disallowed [%]	0.3	0.2	0.1
No. atoms			
protein	13645	13914	13891
water	408	953	921
ligands	13	105	83
***B*-factors**			
protein	51.5	49.5	46.1
water	39.9	47.8	46.2
ligands	52.4	79.9	66.2
R.m.s. deviations			
Bond lengths (Å)	0.0112	0.0140	0.0136
Bond angles (°)	1.70	1.82	1.81
pdb-code	7Q04	7Q05	7Q06

[a]values in parentheses are for the highest-resolution shell
[b]ellipsoidal completeness

Table S3. Accurate mass data. Theoretical and observed masses for parent and fragment ions of NAD, NADH, TPA, and DCD for both MS and MSMS acquisition.

Compound	Ion	Fragment Loss	Acquisition	Expected mass	Observed Mass	Mass error (ppm)
NAD	$[M-H]^-$		MS	662.1018	662.1045	4.1
NADH	$[M-H]^-$		MS	664.1175	664.1163	1.8
TPA	$[M-H]^-$		MS	165.0193	165.02	4.2
TPA	$[M-H]^-$	CO_2	MS	121.0295	121.0301	5.0
TPA	$[M-H]^-$	CO_2	MSMS	121.0295	121.0301	5.0
TPA	$[M-H]^-$	$CO_2 + CO_2$	MSMS	77.0397	77.0403	7.8
DCD	$[M-H]^-$		MS	199.0248	199.0253	2.5
DCD	$[M-H]^-$	H_2O	MS	181.0142	181.015	4.4
DCD	$[M-H]^-$	H_2O	MSMS	181.0142	181.0115	14.9
DCD	$[M-H]^-$	$H_2O + CO_2$	MS	137.0244	137.0248	2.9
DCD	$[M-H]^-$	$H_2O + CO_2$	MSMS	137.0244	137.0248	2.9

Table S4. Uniprot IDs and inferred substrate type for sub clusters from PF00848 SSN

Sequence Count Cluster Number	Selected Representative Uniprot IDs	Verified substrate type(s)
01	Q1QYU7	glycine betaine
02	Q9X4W9, Q93UV3, Q8RTL4, Q8GJF2, Q7DJF2, Q75WN5, Q53122, Q51743, Q45695, P0ABR5, P0A110, O86899, M9PW10, D7RED3, A5W4F2	biphenyl, benzene, toluene, cumene, nitroarene, cinnamate, napthalene, polyaromatics
03	Q51601, P07769, O85673	benzoate, orthohalobenzoate, anthranilate
04	Q76KR6	γ-glutamaylanilide
06	Q84BZ3, Q3C1E3, O87618, O52379, A0A1X9WE59, Q83VL2, Q65AS6	terephthalate, salicylate, hydroxy-picolinate, anthranilate
09	D0C9N6	carnitine
12	O04121	choline
19	Q51974	p-cumate
45	Q9RBG5	2-aminobenzene sulfonate

SI References

1. Terrific Broth. *Cold Spring Harbor Protocols* 2006(1):pdb.rec8620.
2. Luo ML*, et al.* (2016) The CRISPR RNA-guided surveillance complex in Escherichia coli accommodates extended RNA spacers. *Nucleic Acids Research* 44(15):7385-7394.
3. Vonrhein C*, et al.* (2018) Advances in automated data analysis and processing within autoPROC, combined with improved characterisation, mitigation and visualisation of the anisotropy of diffraction limits using STARANISO. *Acta Crystallogr. Sect. A Found. Adv* 74:a360-a360.
4. Vagin A & Teplyakov A (1997) MOLREP: an automated program for molecular replacement. *Journal of applied crystallography* 30(6):1022-1025.
5. Waterhouse A*, et al.* (2018) SWISS-MODEL: homology modelling of protein structures and complexes. *Nucleic acids research* 46(W1):W296-W303.
6. Emsley P, Lohkamp B, Scott WG, & Cowtan K (2010) Features and development of Coot. *Acta Crystallographica Section D: Biological Crystallography* 66(4):486-501.
7. Smart OS*, et al.* (2012) Exploiting structure similarity in refinement: automated NCS and target-structure restraints in BUSTER. *Acta Crystallographica Section D: Biological Crystallography* 68(4):368-380.
8. Williams CJ*, et al.* (2018) MolProbity: More and better reference data for improved all-atom structure validation. *Protein Science* 27(1):293-315.
9. Holm L (2020) DALI and the persistence of protein shape. *Protein Science* 29(1):128-140.
10. Gasteiger E*, et al.* (2005) Protein identification and analysis tools on the ExPASy server. *The Proteomics Protocols Handbook*:571-607.
11. Munck E, *et al.* (1984) Purification and characterization of the Rieske iron sulfur protein from *Thermus thermophilus*. *Journal of Biological Chemistry* 259(1):124-133.
12. Hantke GT, *et al.* (2011) FdC1, a novel ferredoxin capable of alternative electron partitioning, increases in conditions of acceptor limitation at photosystem. *Journal of Biological Chemistry* 286(1):50-59.
13. J. K. Mahto, et al., Molecular insights into substrate recognitionand catalysis by phthalate dioxygenase from Comamonas testosteroni. J. Biol. Chem.(2021) ASAP.