Structural and functional characterization of an intradiol ring-cleavage dioxygenase from the polyphagous spider mite herbivore *Tetranychus urticae* Koch

Caleb R. Schlachter[1], Leily Daneshian[1], Jose Amaya[1], Vincent Klapper[1], Nicky Wybouw[2,3], Tomasz Borowski[4], Thomas Van Leeuwen[2,3], Vojislava Grbic[5,6], Miodrag Grbic[5,6], Thomas M. Makris[1], Maksymilian Chruszcz[1]*

[1] Department of Chemistry and Biochemistry, University of South Carolina, Columbia, SC 29208, USA
[2] Institute for Biodiversity and Ecosystem Dynamics, University of Amsterdam, Science Park 904, 1098 XH, Amsterdam, the Netherlands
[3] Department of Plants and Crops, Ghent University, Ghent, B-9000, Belgium
[4] Jerzy Haber Institute of Catalysis and Surface Chemistry, Polish Academy of Sciences, 30–239 Krakow, Poland
[5] Department of Biology, Western University, London, Ontario N6A 5B7, Canada
[6] University of La Rioja, Logrono, Spain

*To whom correspondence should be addressed: Maksymilian Chruszcz, Department of Chemistry and Biochemistry, University of South Carolina, Columbia, SC 29208 USA;
chruszcz@mailbox.sc.edu; Tel: (803) 777-7399, FAX: (803) 777-9521.

Keywords: two-spotted spider mite, intradiol ring-cleavage dioxygenase, horizontal gene transfer, plant-herbivore interactions, tetur07g02040

Abstract

Genome analyses of the polyphagous spider mite herbivore *Tetranychus urticae* (two-spotted spider mite) revealed the presence of a set of 17 genes that code for secreted proteins belonging to the "intradiol dioxygenase-like" subgroup. Phylogenetic analyses indicate that this novel enzyme family has been acquired by horizontal gene transfer. In order to better understand the role of these proteins in *T. urticae*, we have structurally and functionally characterized one paralog (*tetur07g02040*). It was demonstrated that this protein is indeed an intradiol ring-cleavage dioxygenase, as the enzyme is able to

cleave catechol between two hydroxyl-groups using atmospheric dioxygen. The enzyme was characterized functionally and structurally. The active site of the *T. urticae* enzyme contains an Fe^{3+} cofactor that is coordinated by two histidine and two tyrosine residues, an arrangement that is similar to those observed in bacterial homologs. However, the active site is significantly more solvent exposed than in bacterial proteins. Moreover, the mite enzyme is monomeric, while almost all structurally characterized bacterial homologs form oligomeric assemblies. *Tetur07g02040* is not only the first spider mite dioxygenase that has been characterized at the molecular level, but is also the first structurally characterized intradiol ring-cleavage dioxygenase originating from a eukaryote.

1. Introduction

Aromatic compounds are abundantly present within the environment and can have a natural (such as the plant biopolymer lignin) or man-made origin (such as phthalates in papers and several insecticides and cosmetics) (Brown et al., 2004; Guzik et al., 2013d; Sainsbury et al., 2015; Widhalm and Dudareva, 2015). These compounds are extremely stable due to their high resonance energy and are recalcitrant to biodegradation (Brown et al., 2004; Guzik et al., 2013d). The low level of biodegradation leads to an accumulation of aromatic compounds in various animal and plant tissues (Brown et al., 2004). Aromatic compounds serve various biological functions. For instance, many of the plant secondary metabolites that serve in anti-herbivore defense pathways possess an aromatic structure (Widhalm and Dudareva, 2015).

To metabolize aromatic compounds, organisms can be equipped with metalloenzymes termed ring-cleavage dioxygenases that are capable of cleaving aromatic compounds by utilizing a non-heme iron cofactor in the active site (Broderick, 1999; Dermauw et al., 2013b). Ring-cleavage dioxygenases belong to two unrelated enzyme families: extradiol and intradiol-ring cleavage dioxygenases (ID-RCDs), where the latter cleave the aromatic ring between two adjacent hydroxyl-groups (Bugg, 2003; Vaillancourt et al., 2006). In contrast to extradiol ring-cleavage dioxygenases that are phylogenetically ubiquitous, intradiol ring-cleavage dioxygenases are largely restricted to bacterial and fungal species. Genome analyses of the polyphagous spider mite herbivore *Tetranychus urticae* (or the two-spotted spider mite) revealed the presence of a set of 17 genes that code for secreted proteins that belong to the "intradiol dioxygenase-like" subgroup (Dermauw et al., 2013b; Grbic et al., 2011). Phylogenetic analyses indicate that this novel enzyme family has been acquired by horizontal gene transfer, a biological process whereby genetic information is asexually transferred between two reproductively isolated and unrelated species. This horizontal acquisition and subsequent proliferation of ID-RCDs may contribute to the polyphagous lifestyle of *T. urticae* by detoxifying allelochemicals produced by its plant hosts (Dermauw et al., 2013b). Catechol, or 1,2-dihydroxy benzene, is a common plant metabolite and a substrate for many ID-RCDs (Weng, 2014; Widhalm and Dudareva, 2015). Typically, ID-RCDs detoxify catechols (and catechol derivatives, i.e.

substituted with halides, or other functional groups) by cleaving the aromatic ring between C1 and C2, resulting in the formation of *cis, cis*-muconic acid (Fig. 1) that can be further metabolized into succinate and acetyl-CoA (Broderick, 1999; Guzik et al., 2013d).

In order to better understand the role of ID-RCDs in *T. urticae*, we have structurally and functionally characterized *Tetur07g02040*, referred to hereafter as TuIDRCD. TuIDRCD was selected for these studies, as it is the only ID-RCD in *T. urticae* that contains an intron (Dermauw et al., 2013b) and is least similar in terms of the protein sequence to other two-spotted spider mite intradiol dioxygenase-like proteins (Supplementary Materials Fig. S1). Here, we have determined the spectroscopic, structural and metabolic properties of TuIDRCD, the first arthropod ID-RCD to be characterized to date. TuIDRCD shows several deviations from the previously crystallized bacterial ID-RCDs, hereby shedding more light on the biological role of ID-RCDs in spider mites and other arthropods.

2. Materials and methods

2.1 Expression and Purification of TuIDRCD Constructs

The gene coding for *Tetur07g02040* was ordered from ATUM (formerly DNA2.0; Newark, CA), codon optimized for *E. coli* and inserted into pJExpress411 with a T7 promoter, terminator and kanamycin resistance. Initially, the TuIDRCD gene was synthesized with the first 47 amino acids truncated for two reasons: 1) residues 1-22 were predicted to be a signal peptide (Uniprot reference number: T1K8P1) (UniProt, 2015), 2) residues 28-49 were predicted to be disordered (Slabinski et al., 2007). Furthermore, the gene was synthesized to contain an N-terminal cleavable 6xHis-tag for ease of purification with the Tobacco Etch Virus (TEV) protease cut site MHHHHHHSSGVDLGTENLYFQ/SGSG where the cut site is shown with a slash. The protein coded by the initially designed gene is referred as (t48)TuIDRCD in this manuscript (Supplementary Materials, Fig. S2).

Plasmid was transformed into BL-21 (DE3) cells by heat shock and grown in 1.0 L cultures of Luria-Broth (LB) with 50 µg/mL kanamycin or 100 µg/mL ampicillin (depending on construct) shaking at 37˚C until an O.D. of 0.8 was reached. Cultures were cooled to 16˚C and protein expression was induced with 0.4 mM isopropyl β-D-1-thiogalactopyranoside for 16 hours while shaking. Cells were pelleted and frozen at -80˚C until needed further.

For purification of (t48)TuIDRCD, cell pellets were resuspended in lysis buffer (50 mM Tris, 500 mM NaCl, 10 mM imidazole, 20 mM β-mercaptoethanol (β-ME), 2% glycerol). After resuspension, cells were lysed by sonication with a Branson 45 Sonifier (ThermoFisher, Grand Island, NY). Crude extract was separated by spinning the lysate in a Beckman Coulter centrifuge (Indianapolis, IN). The clear, yellowish

supernatant was poured into a 12 x 1.5 cm Bio-Rad column (Hercules, CA) filled with 5.0 mL of NiNTA resin (ThermoFisher, Grand Island, NY) previously equilibrated in wash buffer (50 mM Tris pH 7.4, 150 mM NaCl, 30 mM imidazole, 2% glycerol, 20 mM β-ME) and then washed with wash buffer. Protein was eluted using elution buffer (50 mM Tris pH 7.4, 150 mM NaCl, 250 mM imidazole, 2% glycerol, 20 mM β-ME) and immediately put into dialysis buffer (50 mM Tris pH 7.4, 150 mM NaCl, 30 mM imidazole) in ThermoFisher SnakeSkin Dialysis Tubing (Grand Island, NY) with a 10,000 MW cutoff. Protein was dialyzed in dialysis buffer for 12 hours at 4˚C.

After dialysis, protein was concentrated with Amicon Ultra concentrators (EMD Millipore, Billerica, MA) with a 10,000 MW cutoff. Concentrated protein was put on a Superdex 200 column attached to an ÄKTA Pure FPLC system (GE Healthcare, Marlborough, MA) equilibrated in FPLC buffer (50 mM Tris pH 7.4, 150 mM NaCl, 30 mM imidazole). Peaks corresponding to (t48)TuIDRCD were pooled and the protein concentration was determined using A_{280} with the MW 27,068 Da and molar extinction coefficient 25,900 $M^{-1}cm^{-1}$ as determined by using ExPASy ProtParam tool (Gasteiger et al., 2003).

Purified (t48)TuIDRCD was red in color, indicating the presence of non-heme ferric iron, but the protein was very unstable and frequent red precipitate was observed in purified protein samples. Due to instability, constructs with the fusion partner Maltose-Binding Protein (MBP) were pursued where all purification buffers had the addition of 5.0 mM maltose. In addition, a construct including residues 23-259, TuIDRCD, was also expressed and purified which showed greater stability than (t48)TuIDRCD.

The (t48)TuIDRCD-MBP yielded decent amounts of protein (~10 mg/L culture); however, the protein did not have the red hue observed from (t48)TuIDRCD purification that was indicative of non-heme iron (and most likely properly folded protein) so this construct was not pursued further. MBP-(t48)TuIDRCD and mTuIDRCD yielded decent quantities of protein (~8 mg/L culture) and had the desired red hue after purification. Protein yields were increased by adding 3.0 mg of ferrous sulfate heptahydrate to the 1.0 L LB culture pre-induction and to the lysis buffer. Protein concentration was determined using A_{280} with the MW 67,027 Da and 27,063 Da, and molar extinction coefficients 92,375 $M^{-1}cm^{-1}$ and 24,410 $M^{-1}cm^{-1}$, for MBP-(t48)TuIDRCD, and mTuIDRCD, respectively. Attempts were not made to remove the His-tag from MBP-dioxygenase due to the presence of a potential TEV cut site in the MBP sequence. The His-tag was cleaved from mTuIDRCD with TEV protease after a 16 hour incubation at 4°C in dialysis buffer; mTuIDRCD was collected by flowing the solution over an NiNTA column equilibrated in dialysis buffer.

2.2 Molecular Cloning of TuIDRCD into pMCSG29

The dioxygenase gene coding for residues 48-259 was cloned into pMCSG29 (Eschenfeldt et al., 2013) using ligation independent cloning (LIC). This plasmid is designed to generate a fusion protein as follows: protein of interest, TEV protease cut site, 6xHis-tag, Tobacco Vein Mottling Virus (TVMV) protease cut site, MBP). The protein coded by this construct is referred to as (t48)TuIDRCD-MBP (Fig. S3). For details of the cloning protocols see Supplementary Materials).

Clones were miniprepped with a GeneJET Plasmid Miniprep Kit (ThermoFisher, Grand Island, NY) and insertion of ID-RCD was first confirmed by restriction digest with XhoI following the manufacturer's instructions (NEB, Ipswich, MA). After restriction digest, clones with correctly sized fragments were sent to EtonBioscience (Research Triangle Park, NC) for sequencing with T7 forward and T7 reverse primers.

2.3 Cloning TuIDRCD into pMBPcs1

The pMBPcs1 plasmid was generated and amplified as previously described (Schlachter et al., 2017). The (t48)TuIDRCD gene was cloned into pMBPcs1 using LIC, and this construct is referred to as MBP-(t48)TuIDRCD (Fig. S4).

Clones were miniprepped with a GeneJET Plasmid Miniprep Kit (ThermoFisher, Grand Island, NY) and insertion of dioxygenase was first confirmed by restriction digest with HindIII and XbaI following the manufacturer's instructions (NEB, Ipswich, MA). After restriction digest, clones with correctly sized fragments were sent to EtonBioscience (Research Triangle Park, NC) for sequencing with T7 forward and T7 reverse primers.

2.4 Generation of Mature TuIDRCD (mTuIDRCD)

Originally, the first construct for TuIDRCD was ordered from ATUM (Newark, CA) as a truncated protein with the first 47 amino acids missing. To observe if the predicted, unstructured residues have an impact on protein stability, residues 23-47 (25 residues total) were cloned into the (t48)TuIDRCD construct to produce mature TuIDRCD (residues 23-259, mTuIDRCD, Fig. S5)).

Primers Dioxy25-F and Dioxy25-R in Table S2 were used to produce a fragment containing the desired 25 amino acids. The fragment was generated and amplified by using PCR with Q5 polymerase (NEB, Ipswich, MA) following the manufacturer’s protocol. To insert the fragment into the TuIDRCD construct, primers p411-Dioxy25-F and p411-Dioxy25-R were used to generate megaprimers of the fragment.

Clones were selected, miniprepped, and insertion of the fragment was confirmed by sending plasmids for sequencing to EtonBioscience (Research Triangle Park, NC) using T7 forward and T7 reverse primers.

2.5 Differential Scanning Fluorimetry

Differential Scanning Fluorimetry was performed as described previously (Booth et al., 2018). Briefly, fluorescence data was collected by a Bio-Rad CFX96 RT-PCR instrument (Hercules, CA). SYPRO® Orange Dye (ThermoFisher, Waltham, MA) was diluted 1:1000 in 1.0 mL of 1 mg/mL protein (mTuIDRCD with cleaved his-tag). The solution of the dye and protein was mixed with each pH and salt screen condition 1:1 for a 20 µL total reaction volume. Bio-Rad Hardshell 96-well RT-PCR plate was used and sealed with Bio-Rad Microseal PCR Plate Sealing Film (Hercules, CA). Emission was measured at 590 nm (excitation at 488 nm) at temperatures increasing in the range of 30-90°C with a 1°C/min in 2°C increments. All screen conditions had working concentrations of 50 mM buffer with a pH range of 4-9.5 in 0.5 pH unit increments and a sodium chloride range from 0-1.0 M (No salt, 0.05 M, 0.10 M, 0.15 M, 0.20 M, 0.25 M, 0.50 M and 1.0 M). Buffers used to maintain the pH ranges were acetate (pH 4.0-5.0), Bis-Tris (pH 5.5-6.5), Tris (pH 7.0-8.0) and CHES (pH 8.5-9.5). Each experiment was performed in triplicate.

2.6 Activity Assay

All reagents for enzymatic assays were purchased from Sigma-Aldrich (St. Louis, MO). Assays were performed using catechol (1,2-dihydroxybenzene) and 4-methylcatechol as substrate following the protocol described previously (Guzik et al., 2013d). Cleavage of substrate was monitored by measuring the A_{260} nm and A_{255} nm of the product *cis*,*cis*-muconic acid and 3-methylmuconic acid with a molar extinction coefficient of 16,800 $M^{-1}cm^{-1}$ and 14,300 $M^{-1}cm^{-1}$ respectively. The measurements were performed with Synergy H1 Hybrid Multi-Mode Microplate Reader (Biotek, Winooski, VT). The production of *cis*,*cis*-muconic acid was confirmed using liquid chromatography mass spectrometry and NMR (data not shown). Reactions were performed using 200 µL total with working concentrations of 100 mM sodium phosphate (pH 6.0), Bis-Tris (pH 6.5), Tris-HCl (pH 7.0, 7.5, and 8.0), AMPSO or AMPSO/CHES (pH 8.5, 9.0, and 9.5), 5 µM (for catechol) or 0.5 µM (for 4-methylcatechol) protein (mTuIDRCD with cleaved his-tag), and 1-200 µM substrate. Activity was measured for 20 minutes. All measurements were performed in triplicate and data were fit by Origin software (Matera et al., 2010) and Microsoft Excel using the Solver add-on. Experimental results are summarized in Table 1.

2.7 Ferrozine Assay

Ferrozine assay was used to check whether the purified recombinant protein contains iron (Tchesnokov et al., 2012). All reagents for the ferrozine assay were purchased from Sigma-Aldrich (St. Louis, MO). Initially, 30 µM protein was incubated and hydrolyzed by 0.5 M HCl overnight at room temperature. Next, reducing reagent was added to the mixture and incubated for 30 minutes at room temperature for a final concentration of 0.2 M hydroxylamine (1.4 M hydroxylamine stock was prepared using 2 M HCl). Lastly, working concentrations of 0.8 mM ferrozine (10 mM ferrozine stock was prepared using 100 mM ammonium acetate) and 0.6 M ammonium acetate buffer (5 M buffer stock was prepared using ammonium hydroxide (pH 9.5)) were added and the absorbance was measured at 562 nm with the molar extinction coefficient of 27,900 $M^{-1}cm^{-1}$ on a Hewlett-Packard 8453 spectrophotometer (Palo Alto, CA). The standard curve was prepared using iron (III) chloride in the range of 5 µM to 60 µM following the same procedure.

2.8 Inductively Coupled Plasma Mass Spectrometry (ICP-MS)

For further investigation of the iron content ICP-MS was performed (Becker et al., 2010). Briefly, 200 µg of protein was prepared in 2% nitric acid and filtered by 0.2-micron (Whatman Nylon Puradisc) syringe filter. Finnigan ELEMENT XR double focusing magnetic sector field inductively coupled plasma-mass spectrometer (SF-ICP-MS) was used for the analysis with Ir and/or Rh as internal standards. 0.2 mL/min Micromist U-series nebulizer (GE, Australia), quartz torch and injector (Thermo Fisher Scientific, USA) were used for sample introduction. Sample gas flow was at 1.08 mL/min.

2.9 UV-VIS Spectroscopy

To assess whether the available iron is bound to the protein, the whole spectrum was measured. Protein concentration was measured at 280 nm on Cary 50 UV-Vis Spectrophotometer (Agilent, Santa Clara, CA) with molar extinction coefficients determined by ExPASy ProtParam. Next, the absorbance at 430 nm was determined and the molar extinction coefficient for the tyrosine to iron charge transfer transmission was calculated.

2.10 Crystallization of MBP-(t48)TuIDRCD and mTuIDRCD

All chemicals were purchased from Hampton Research (Aliso Viejo, CA), ThermoFisher (Grand Island, NY) or Sigma-Aldrich (St. Louis, MO). Crystallization experiments were performed at room temperature using the sitting-drop vapor diffusion method and MRC 2-drop 96-well crystallization plates (Hampton Research, Aliso Viejo, CA). Initially, MBP-(t48)TuIDRCD crystals were grown in 0.1 M sodium cacodylate trihydrate pH 6.5, 0.2 M magnesium acetate tetrahydrate, 20% w/v PEG8000 (3:1 of protein:mother liquor where protein ~13 mg/mL). These red colored crystals were microseeded due to their

small size. For microseeding, crystals were crushed and resuspended in a 1.5 mL tube filled with 150 µL of the cacodylate crystallization solution and a glass bead then vortexed. An equal volume of MBP-(t48)TuIDRCD (~13 mg/mL) was added to the seed stock and drops were set with 1:1 ratio of protein to mother liquor as written above. Using the seed stock, red crystals grew after about one month in 0.1 M Tris pH 7.5, 15% w/v PEG6000. The mTuIDRCD crystals were obtained using vapor diffusion and sitting drop setup. Drops were formed by mixing protein (1.1 mg/mL) and mother liquor 1:1. The mother liquor contained 0.2 M Li_2SO_4, 0.1 M HEPES pH 7.5, 10% w/v PEG3350 and 10 mM catechol. All crystallization experiments were performed at 25°C.

2.11 Data Collection, Structure Determination and Refinement

Table 2 shows the data collections statistics for crystal structures of mbp-(t48)TuIDRCD (residues 48-259; truncation product of MBP-(t48)TuIDRCD) and mTuIDRCD (residues 23-259). The crystal structures for mbp-(t48)TuIDRCD and mTuIDRCD were deposited in the Protein Data Bank (PDB) (Berman et al., 2000) with the accession codes 5VG2 and 6BDJ. Crystals were cryo-cooled in liquid nitrogen and data was collected using the Southeast Regional Collaborative Access Team (SER-CAT) 22ID beamline at the Advanced Photon Source (APS), Argonne National Lab (Argonne, IL). Data were processed with the HKL-2000 software package (Otwinowski, 1997). Molecular replacement for mbp-(t48)TuIDRCD was performed using MOLREP (Vagin and Teplyakov, 1997) integrated with HKL-3000 (Minor et al., 2006) and the PDB entry 4ILT as a starting model. BUCCANEER (Cowtan, 2006) and HKL-3000 were used to rebuild the initial model. Refinement was performed using REFMAC (Murshudov et al., 2011) and HKL-3000. Non-crystallographic symmetry was used during the whole process of refinement. TLS refinement was used during the last stages of refinement and the TLS Motion Determination server was used for partitioning protein chains into the rigid bodies undergoing vibrational motions (Painter and Merritt, 2006). Model was updated and validated with COOT (Emsley and Cowtan, 2004). MOLPROBITY was used in the final steps of the model validation (Davis et al., 2007). A similar approach was used for determination, refinement, and validation of mTuIDRCD structure; however, in this case the structure of mbp-(t48)TuIDRCD (PDB code: 5VG2) was used as the starting model for molecular replacement.

2.12 EPR Spectra

EPR spectra of the as-isolated enzyme were recorded using an X-band Bruker EMXplus spectrometer (Bruker Bio Spin, Billerica, MA) equipped with an Oxford Instruments ESR900 (Oxfordshire, UK) liquid helium continuous flow cryostat. Spectra were recorded at a temperature of 12 K and a 1 mT modulation amplitude. MBP-(t48)TuIDRCD and mTuIDRCD protein concentrations were 120 and 200 µM, respectively. The nitrosyl complexes were prepared in a Coy anaerobic chamber after degassing the

protein with N_2. The enzyme was treated with 3 molar equivalents of methyl viologen, 9 molar equivalents of sodium dithionite and 1 molar equivalent of MAHMA NONOate (2 molar equivalents of NO per protein) prior to flash freezing in liquid nitrogen.

2.13 Various Computational Calculations

XtalPred was used to design the original construct for TuIDRCD (Slabinski et al., 2007). The programs ProFunc and PDBePISA were used to analyze the crystal structures (Krissinel and Henrick, 2007; Laskowski, 2017). COOT, PyMOL (DeLano W., 2002) and UCSF-Chimera (Pettersen et al., 2004) were used to visualize and analyze the structures, as well as to generate figures. APBS (Jurrus et al., 2018) as implemented in PyMOL was used for electrostatics calculations. DALI (Holm and Rosenstrom, 2010) and PDBeFold (Krissinel and Henrick, 2004) were used to search for similar structures. The pI values were calculated using ExPASy ProtParam tool.

2.14 Phylogenetic Analysis

Based on previous studies, the closest homologs to *tetur07g02040* within Tetranychidae were included in the phylogenetic analysis (Bajda et al., 2015; Dermauw et al., 2013b). In addition, arthropod and tardigrade genome assemblies were mined for the presence of ID-RCD genes via local tBLASTn-searches using *tetur07g02040* as a query. Contamination of genome assemblies was examined by identifying neighboring eukaryotic genes on scaffolds/contigs and by verifying a continuous Illumina- and PacBio-read coverage across the coding region in case of the collembolan genome assemblies (*Folsomia candida* and *Orchesella cincta*) (Amsterdam, 2018). Alignments and Conserved Domain-searches showed that an ID-RCD gene appeared to be incorrectly annotated within the *B. tabaci* genome portal (Chen et al., 2016) and we included the NCBI annotation in our phylogenetic dataset (Supplementary Table S3). The final set of 67 protein sequences were aligned using MAFFT (G-INS-I settings). The protein model of WAG+I+G+F was selected on the Akaike Information Criterion using ProtTest 3.4.4 (Darriba et al., 2011). A phylogenetic tree was constructed based on Bayesian inference using Mr Bayes v3.2.6 (Ronquist et al., 2012). With seed and nswaps set at 21343 and 3, respectively, 7 heated chains were implemented in the MCMC analysis and were sampled every 100 generations. A total of 1,100,000 generations were run, which resulted in a final average standard deviation of split frequencies of 0.0128.

3. Results

3.1 Phylogenetic Analysis

Arthropod genome mining revealed the presence of an ID-RCD gene embedded within the *Bemisia tabaci* genome (Hexapoda: Hemiptera), an apparent horizontal gene transfer event that was not previously

detected (Chen et al., 2016). Mites within the Sarcoptiformes lineage (Chelicerata: Acariformes) also possessed ID-RCD genes within their genomes (Fig. 2). Finally, our genome screens also confirmed the presence of ID-RCD genes in collembolan genomes (Hexapoda: Entomobryomorpha) and in genomes of species that belong to the Tardigrada phylum (Faddeeva-Vakhrusheva et al., 2016; Yoshida et al., 2017). The Bayesian inference-based phylogenetic reconstruction indicated that the *B. tabaci* and tardigrade dioxygenase genes have a bacterial and fungal origin, respectively, and reflect independent horizontal gene transfer events (Fig. 2). In contrast, the precise evolutionary origin and history of the dioxygenases of the collembolan and sarcoptiform mite lineages remain unclear, but our phylogenetic reconstruction tentatively suggested a common fungal origin. The monophyletic clade that holds TuIDRCD and close spider mite homologs supported the previously postulated evolutionary scenario wherein a single horizontal gene transfer event from a fungal donor species occurred prior to speciation within the tetranychid spider mite family (Fig. 2) (Bajda et al., 2015; Dermauw et al., 2013a).

3.2 TuIDRCD Constructs – Protein Production

In total, four TuIDRCD constructs were produced (see Materials and methods, and Supplementary Figures S2-S5) due to initial solubility issues of the protein. Two of these constructs ((t48)TuIDRCD and mTuIDRCD) contained a hexahistidine tag and residues corresponding to the spider mite protein, while two additional constructs ((t48)TuIDRCD-MBP and MBP-(t48)TuIDRCD) were fused to Maltose Binding Protein (MBP). The (t48)TuIDRCD construct, where (t48) indicates truncation of the first 47 residues, was generated because this region was initially predicted to be disordered based on results from XtalPred (Slabinski et al., 2007). However, the protein was found to be unstable. Likewise, (t48)TuIDRCD-MBP, where MBP is fused to the C-terminal end, was colorless which is consistent with the absence of Fe^{3+} in the active site. Two additional constructs, MBP-(t48)TuIDRCD (MBP fused to N-terminal) and mTuIDRCD, were found to contain a pink hue and to contain iron, and were thus selected for the structural and spectroscopic studies described below. The latter mTuIDRCD construct, is comprised of residues 23-259 that correspond to a mature form of the enzyme and was used for activity studies.

3.3 UV-Visible and Electron Paramagnetic Resonance Spectroscopy

As the presence of the Fe^{3+} cofactor is critical for ID-RCD activity, the iron content of mTuIDRCD was first confirmed using metal-counting procedures that included the ferrozine assay (Tchesnokov et al., 2012) and inductively coupled plasma mass spectrometry (ICP-MS). Together, the results indicated that the purified protein contained 0.98 nmol Fe per nmol protein. ICP also revealed the presence of a minor manganese contaminant (0.11 nmol per nmol protein, Table S1) and no other metal was present in

significant abundance. The presence of the iron cofactor was further probed by optical spectroscopy. The optical spectroscopic features were similar for both MBP-(t48)TuIDRCD and mTuIDRCD. The optical spectra of mTuIDRCD and MBP-(t48)TuIDRCD are shown in Figure 3A and Figure S6A respectively. Both constructs exhibited absorption features at 325 and 460 nm. The latter feature, with an extinction coefficient ε_{460} ~2,500 $M^{-1}cm^{-1}$, is attributed to a mixture of several ligand to metal charge transfer (LMCT) transitions from the axial and equatorial tyrosinate ligands to Fe^{3+}, giving rise to its characteristic burgundy color (Davis et al., 2002). EPR spectroscopy was used to further probe the oxidation state and ligand environment of the iron cofactor. The EPR spectra of both constructs were similar and shown in Fig. 3B (mTuIDRCD) and Fig. S6B (MBP-(t48)TuIDRCD) respectively. Both constructs exhibit two main resonances at g = 9.25 and 4.29, which is diagnostic for a high-spin ferric center and nearly identical to those observed for other intradiol dioxygenases such as the well-characterized protocatechuate 3,4-dioxygenase from *Brevibacterium fuscum* (Orville and Lipscomb, 1993). No evidence for Mn^{2+} was observed. In order to further verify the presence of bound Fe^{3+}, the nitrosyl complex of MBP-(t48)TuIDRCD was prepared through reduction of the enzyme with dithionite followed by exposure to NO (Fig. S7). The resulting EPR spectrum exhibits is readily differentiated from the resting state of the enzyme, and is consistent with assignment as an S = 3/2 complex that arises from antiferromagnetic coupling between the oxidized iron and bound diatomic ligand.

3.4 mTuIDRC Stability and Enzymatic Activity

The mTuIDRC was found to be relatively stable across a broad pH range and salt concentrations (Fig. 4A) as determined with Differential Scanning Fluorimetry (DSF). The highest thermal stability of the mTuIDRCD was observed in basic solutions.

The kinetic parameters of mTuIDRCD have been determined for catechol at different pH values (Table 1). The k_{cat} increases slightly at basic pH values. The k_{cat} and catalytic efficiency are approximately 10-fold and 4-fold higher, respectively, at pH 8-9 compared to pH 6 (Fig. 4B). The K_M for catechol is similar to those observed for previously characterized enzymes; however, the k_{cat} and catalytic efficiency are significantly lower (as much as 4 orders of magnitude) than those reported for bacterial homologs (Caglio et al., 2009; Ferraroni et al., 2013; Ferraroni et al., 2006; Guzik et al., 2013a; Knoot et al., 2015; Matera et al., 2010; Travkin et al., 1997). Based on thermal stability studies with DSF (Fig. 4A), mTuIDRCD is significantly more stable at basic pH which correlates with enzymatic activity. Increased enzymatic activity at basic pH has been observed for other ID-RCDs (Caglio et al., 2013; Guzik et al., 2013b; Tsai and Li, 2007). Activity of mTuIDRCD was also tested on 3,4-dihydroxybenzoate (protocatechuate), but no activity was observed. On the other hand, 4-methylcatechol was readily cleaved by mTuIDRD with significantly better efficiency (> 10 fold higher k_{cat}) than catechol (Table 1).

3.5 Crystal Structures of mbp-(t48)TuIDRCD and mTuIDRCD

Similar to our previous work with the crystallization of MBP-cyanase (Schlachter et al., 2017), the MBP fragment was not present in the crystal structure of MBP-(t48)TuIDRCD. Therefore, the nomenclature mbp-(t48)TuIDRCD will be used to describe the fragment of the fusion construct MBP-(t48)TuIDRCD that was in the crystal structure. The mTuIDRCD construct was also crystallized. mbp-(t48)TuIDRCD and mTuIDRCD crystallized in a monoclinic system with four and two protein chains present in the asymmetric unit, respectively. In the mbp-(t48)TuIDRCD structure, each chain consists of residues 56-259 (with respect to the complete sequence reported in Uniprot (T1K8P1)) where the first eight amino acids (residues 48-55) of the truncated recombinant dioxygenase studied here are not visible in the electron density, nor is the TEV cut site and 6xHis-tag on the C-terminus. β-mercaptoethanol (β-ME) was used in purification buffers, but each chain starts at C56 which forms a disulfide bridge with C99 on the same chain. The structure of the mTuIDRCD is almost identical to the structure of mbp-(t48)TuIDRCD. In mTuIDRCD, the N-terminal residues 23-54 are disordered which is consistent with XtalPred predictions (Slabinski et al., 2007), and is the main reason these residues were excluded in the design of (t48)TuIDRCD, (t48)TuIDRCD-MBP, and MBP-(t48)TuIDRCD. The protein chains from both crystal structures superpose with RMSD values of ~0.6Å over 198 aligned C_α atoms.

TuIDRCD contains eight β-strands that form a β-sandwich and two α-helices (Figs. 5 and 6). The β-sandwich core is conserved with most intradiol-ring cleavage dioxygenases (Bianchetti et al., 2013). The overall fold of dioxygenase is most similar to that of SACTE_2871 (PDB code: 4ILT; Fig. 5C) from *Streptomyces* sp. SirexAA-E which was used as the starting model for molecular replacement (Bianchetti et al., 2013). However, there is only a 17% sequence identity and 34% sequence similarity between TuIDRCD and SACTE_2871. Dali (Holm and Rosenstrom, 2010) was used to search for proteins similar to TuIDRCD and several dioxygenases were identified: catechol 1,2-dioxygenase from *Burkholderia multivorans* (PDB code 5UMH, 167/307 residues aligned, RMSD 2.4 Å, 26% sequence identity); 1,2-dioxygenase from *Burkholderia ambifaria* (PDB code: 5VXT, 166/311 residues aligned, RMSD 2.5Å, 24% sequence identity); 3-chlorocatechol 1,2-dioxygenase from *Rhodococcus opacus 1Cp* (PDB code: 2BOY, 156/248 residues aligned, RMSD 2.0 Å, 24% sequence identity); hydroxyquinol 1,2-dioxygenase from *Pseudomonas putida* DLL-E4 (PDB code: 3N9T, 158/286 residues aligned, RMSD 2.4 Å, 23% sequence identity) and many other dioxygenases that do not have more than 25% of sequence identity to TuIDRCD over the aligned region.

TuIDRCD does not have an extensive N-terminal dimerization domain, an attribute common to ID-RCDs (Bianchetti et al., 2013). In the crystal structures, both mbp-(t48)TuIDRCD and mTuIDRCD appear to be monomeric (Figs. 5 and 6) which is consistent with size exclusion chromatography results (data not

shown). The interface area between neighboring monomers is around 600 Å^2 or 360 Å^2, which is below the cutoff value (856 Å^2) proposed for discrimination between homodimeric and monomeric proteins (Ponstingl et al., 2000). The quaternary structure of TuIDRCD is similar to that of SACTE_2871 which is currently the only other ID-RCD that is monomeric and has a crystal structure determined. However, in the case of SACTE_2871, the dioxygenase domain is fused with a lignin binding domain (Bianchetti et al., 2013).

Fe^{3+} binding sites in both mbp-(t48)TuIDRCD and mTuIDRCD structures were validated by the CheckMyMetal server (Zheng et al., 2014; Zheng et al., 2008). All Fe^{3+} cations were refined with full occupancy. The non-heme Fe^{3+} is coordinated by two tyrosines and two histidines in each active site: Y118, Y163, H169, H171 (Fig. 6B). The ferric center is coordinated in a trigonal bipyramidal manner with the four active site residues and a water molecule/hydroxyl group which is common to ID-RCDs (Guzik et al., 2013d). The mTuIDRCD was incubated with catechol prior to crystallization; despite the catechol not being present in the crystal structure, the side chain of Y163 is observed in two different orientations that correspond to iron-bound and "displaced" conformations. In the second conformation, the hydroxyl oxygen atom is displaced by 6.0 Å from the position that is observed in the iron-bound form. The presence of the second conformation permits iron environment comparisons in *apo* and *holo*-forms of the enzyme (Fig. 6). Comparison of mTuIDRCD with the catechol bound structure of catechol 1,2-dioxygenase from *R. opacus* 1CP (PDB code: 3HHY) shows changes in the coordination of Fe^{3+} upon substrate binding (Fig. 6) (Matera et al., 2010).

Comparison of chains from TuIDRCD crystal structures reveals significant conformational flexibility of the D123-K131 region that is close to the metal binding site (Fig. 7). This region includes several basic residues that, in the crystal structure, have mostly disordered side chains. However, in two chains of the structure of mbp-(t48)TuIDRCD, the side chains of K128 point toward the metal binding site.

Analysis of TuIDRCD, SACTE_2871, catechol 1,2-dioxygenase from *Rhodococcus opacus* and other ID-RCDs revealed presence of two structurally conserved water molecules. One of them (Wat 1, Fig. 6B) forms H-bonds with the peptide backbone at E64 and Y67, and side chain of W110. In addition, in the *apo*-form of the enzyme, Water 1 forms a H-bond with the hydroxyl/water molecule coordinating iron. The second conserved water molecule (Wat 2; Fig. 6B) forms H-bonds with the peptide backbone of G63 and G184, as well as side chain Nδ1 of H171. Water 2 molecule is completely buried by residues.

4. Discussion

4.1 Origin of TuIDRCD

T. urticae is able to feed on more than 1,100 plant species and therefore is considered to be one of the most polyphagous arthropod herbivores (Jeppson et al., 1975). This organism is not only able to detoxify various natural compounds originating from plants, but it also has an exceptional ability for developing pesticide resistance (Grbic et al., 2007; Van Leeuwen and Dermauw, 2016; Van Leeuwen et al., 2015). Some of the proteins, like ID-RCDs, used by *T. urticae* for xenobiotic detoxification were acquired by a lateral gene transfer. Phylogenetic analysis suggests that TuIDRCD was acquired by the spider mite from a fungal source (Fig. 2). The gene coding for TuIDRCD is the only in *T. urticae* that contains an intron (Dermauw et al., 2013b). The TuIDRCD reported here (*Tetur07g02040*) is not only the first spider mite dioxygenase that has been characterized on the molecular level, but is also the first structurally characterized ID-RCD originating from a eukaryote. All previously structurally characterized ID-RCDs are of bacterial origin (Fig. 2) and the closest bacterial homolog (SACTE_2871; (Bianchetti et al., 2013)) with the structure determined has 17% sequence identity and 34% sequence similarity to TuIDRCD. Furthermore, TuIDRCD shares only 23-30% identity and 33-53% similarity with residues of other ID-RCDs identified in *T. urticae*.

4.2 Structure of TuIDRCD

TuIDRCD is not only very distinct from its bacterial homologs in terms of the source of origin and primary structure, but it also has quite unique structural features. The structures that we have determined (mbp-(t48)TuIDRCD and mTuIDRCD) clearly indicate that the N-terminal fragment of the protein is flexible which is in agreement with predictions made based on the protein sequence. The ordered and well-structured part of the protein starts from C56 which forms a disulfide bridge with C99. Comparison of sequences suggests that this disulfide bridge is most likely present in 14 out of the 17 ID-RCDs identified in the *T. urticae* genome; the exceptions are *Tetur01g00490, Tetur04g00150*, and *Tetur07g06560*. TuIDRCD has two additional cysteine residues in the flexible N-terminal part of the protein (C25 and C40), but these amino acids are not conserved among the *T. urticae* proteins. Interestingly, superposition of TuIDRCD and SACTE_2871 (PDB code: 4ILT) structures reveals that C99 from the spider mite protein corresponds to C119 from the *Streptomyces* protein. C119 of SACTE_2871 forms an interchain disulfide bridge with a symmetry related molecule. However, most likely formation of this intermolecular bridge is an effect of crystallization of a truncated version of the protein (residues 77-230). This is consistent with the fact that in another crystal structure of SACTE_2871 (PDB code: 4ILV) such an intermolecular disulfide linkage is not present (Bianchetti et al., 2013). Therefore, it is very likely that the TuIDRCD disulfide bridge formed by C56 and C99 is homologous to C72 and C119 in SACTE_2871.

Another characteristic structural feature of TuIDRCD is its quaternary structure. This protein, similar to SACTE_2871, is monomeric and lacks the N-terminal oligomerization domain, while other proteins from this family are homodimeric or heterooligomeric. Furthermore, SACTE_2871 has a C-terminal domain that is involved in lignin binding (Bianchetti et al., 2013). The function of the N-terminal (residues 23-55) of TuIDRCD is not known, but despite being disordered, it clearly stabilizes the protein compared to (t48)TuIDRCD, which was very unstable during expression and purification (data not shown). It is possible that this fragment of enzyme is involved in interactions with other proteins or molecules similarly to the function of C-terminal domain of SACTE_2871 (Bianchetti et al., 2013).

The TuIDRCD iron binding site is composed of the same amino acids as in bacterial homologs, also reflected in very similar optical and EPR spectroscopic properties. There are two histidine (H169 and H171) and two tyrosine residues (Y118 and Y163) that coordinate Fe^{3+}. Similarly, as in other ID-RCDs, Y163 is coordinating the metal in the resting state of the enzyme (Figs. 1 and 6) but is very likely displaced upon substrate binding; Y163 also "swings" back and binds to the metal in the later steps of the reaction (Borowski and Siegbahn, 2006b; Knoot et al., 2015; Valley et al., 2005). Structural comparison of TuIDRCD with other catechol 1,2-dioxygenases shows that residues P66, W110, R166 and Q185 are completely conserved. R166 and Q185 sidechains form H-bonds, and R166 may be responsible for stabilization of reaction intermediates (Borowski and Siegbahn, 2006b). Analysis of TuIDRCD crystal structures revealed two water molecules in the vicinity of the Fe^{3+} binding site that most likely have a structural function (Fig. 6).

4.3 Enzymatic Activity

The mTuIDRCD is able to cleave catechol, which results in the formation of *cis, cis*-muconic acid (Fig. 1); however, under the same conditions the enzyme is not able to use 3,4-dihydroxybenzoate as a substrate (data not shown). Based on kinetic data, the enzyme is more efficient in solutions with basic pH. K_M values observed for catechol are in a low-micromolar range and are similar to those observed for various bacterial homologs. However, k_{cat} and k_{cat}/K_M values indicate that TuIDRCD is more than 2-fold less efficient in cleaving the substrate in comparison with SACTER_2871 and 2-4 fold less efficient in comparison with some bacterial 1,2-dioxygenases (Ferraroni et al., 2013; Ferraroni et al., 2006; Matera et al., 2010; Micalella et al., 2011). The lack of the efficiency may be related to the fact that catechol is not amongst the normal panel of catecholic substrates for the enzyme, or that TuIDRCD requires an activator/regulatory molecule (or molecules) in order to have full enzymatic potential. Evidence may be provided by the failure of the 4-methylcatechol substrate to elicit perturbations to the optical spectrum of TuIDRCD that occur upon ligation of the substrate. It is also notable that protocatechuate 3,4-dioxygenases form heterooligomers (Fig. 5D). The heterooligomers are composed from α and β units, and both units

participate in formation of the active site, although only the β unit provides Fe^{3+} binding residues (Frazee et al., 1993; Guzik et al., 2013d). Interestingly, we did not observe cleavage of protocatechuate by mTuIDRCD, while the enzyme preferred 4-methylcatechol over catechol as the substrate. It clearly shows that the type of the catechol substituent is important in gating binding and catalysis by TuIDRCD. The reason why TuIDRCD is catalytically notably less efficient than bacterial ID-RCDs is not clear, yet the observed variance of its activity for substrates with various 4-substituents allows us to speculate that at least one of the elementary steps, whose reaction energy affects the overall barrier, is the culprit. It might be either the O_2 binding step, or the following step whereby the peroxo intermediate changes its conformation (Fig. 1). In the former case the substituent effect would be electronic in nature, i.e. electron withdrawing groups deplete π electron density of the ring and hence deactivate it, whereas in the latter the effect would be steric in nature, i.e. non-bonded interactions between the substituent and protein residues lining the active site pocket would tune the energy difference between the two conformers of the peroxo intermediate. It is also worth mentioning that the active site in TuIDRCD is significantly more open in comparison with bacterial homologs that have their structures determined.

The existence of the proliferated family of ID-RCD enzymes in *T. urticae* can be linked to its ability to feed on an extremely wide range of host plants. Currently, the natural substrates for *tetur07g02040* (TuIDRCD) are not known. However, the presence of a highly homologous enzyme in *T. evansi* (Uniprot id: K9UU26), an oligophagous mite species that feeds on solanaceous plants, suggests that either *tetur07g02040* processes substrates conserved between mite species, or that *tetur07g02040* evolved the ability to process a greater spectrum of substrates. Similarly, proteins homologous to *tetur07g02040* were also identified in *Panonychus citri* (citrus red mite) and *Panonychus ulmi* (European red mite) (Bajda et al., 2015). Therefore, comparative analysis of TuIDRCD, TeIDRCD as well as *P. citri* and *P. ulmi* dioxygenases activities against a panel of potential substrates may allow discrimination between these possibilities. Furthermore, if TuIDRCD is involved in detoxification, then it is expected that it will be secreted and active in the digestive tract, a compartment where plant and mite compartments intercept. At present, it is not known where within the gut digestion and detoxification occur (Bensoussan et al., 2018). However, the basic range of optimal pH for the TuIDRCD activity suggests that its activity may not coincide with the activities of digestive proteinases, as their optimal range is within the acidic pH (Santamaria et al., 2015).

Accession Numbers

Coordinates and structure factors have been deposited to the PDB with accession codes 5VG2 (mbp-(t48)TuIDRCD) and 6BDJ (mTuIDRCD).

Acknowledgments

We would like to thank Dr. Lijiang He (Mass Spectrometry Manager, University of South Carolina) for performing ICP-MS experiments. We also would like to thank Dr. Caryn Outten for access to the microplate reader.

Structural results shown in this report are derived from data collected at Southeast Regional Collaborative Access Team (SER-CAT; 22 ID) beamline at the Advanced Photon Source, Argonne National Laboratory. Supporting institutions may be found at www.ser-cat.org/members.html. Use of the Advanced Photon Source was supported by the U.S. Department of Energy, Office of Science, Office of Basic Energy Sciences, under Contract Nos. DE-AC02-06CH11357 and W-31-109-Eng-38. This work was partially supported by an ASPIRE II and ASPIRE III grants from the Office of the Vice President of Research at the University of South Carolina. C.R.S. and M.C. were partially supported by R01AI077653 grant from National Institute of Allergy and Infectious Diseases. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. M.G. and V.G. acknowledge funding by the Government of Canada through Genome Canada and the Ontario Genomics Institute (OGI-046) and Ontario Research Fund−Research Excellence Round 8 (RE08-067). N.W. was supported by a Marie Sklodowska-Curie Action (MSCA) Individual fellowship (658795-DOGMITE) of Horizon 2020 and a Flanders Research Foundation (FWO) fellowship (12T9818N). T.V.L. was supported by the Fund for Scientific Research Flanders (FWO) (Grant Nos. G009312N and G053815N). T.V.L., V.G. and M.G. were supported by the European Commission (EC contract 618105) via FACCE ERA-NET Plus and FACCE-JP (Genomite, project ID 137). TVL has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement n° 772026).

Conflict of interest

The authors declare that they have no conflicts of interest with the contents of this article.

Author contributions

M.C., M.G., V.G. and T.V.L. initiated the studies. C.R.S performed cloning. C.R.S., L.D. and V.K. produced, and crystallized proteins. C.R.S and L.D. performed structural analysis. C.R.S and L.D. performed enzymatic studies and DSF experiments. J.A. and T.M.M. performed and analyzed EPR experiments. N.W. and T.V.L. conducted phylogenetic analysis. T.B. performed analysis of the enzyme

activity. All authors participated in the analysis of data and contributed to writing the manuscript with a final review by M.C.

References

The PyMOL Molecular Graphics System, Version 2.0 Schrödinger, LLC.

Amsterdam, V.U., 2018. Collembolomics.

Bajda, S., Dermauw, W., Greenhalgh, R., Nauen, R., Tirry, L., Clark, R.M., Van Leeuwen, T., 2015. Transcriptome profiling of a spirodiclofen susceptible and resistant strain of the European red mite Panonychus ulmi using strand-specific RNA-seq. BMC genomics 16, 974.

Becker, J.S., Zoriy, M., Matusch, A., Wu, B., Salber, D., Palm, C., Becker, J.S., 2010. Bioimaging of metals by laser ablation inductively coupled plasma mass spectrometry (LA-ICP-MS). Mass spectrometry reviews 29, 156-175.

Bensoussan, N., Zhurov, V., Yamakawa, S., O'Neil, C.H., Suzuki, T., Grbic, M., Grbic, V., 2018. The digestive system of the two-spotted spider mite, Tetranychus urticae (Koch), in the context of the mite-plant interaction. Front. Plant Sci.

Berman, H.M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T.N., Weissig, H., Shindyalov, I.N., Bourne, P.E., 2000. The Protein Data Bank. Nucleic acids research 28, 235-242.

Bianchetti, C.M., Harmann, C.H., Takasuka, T.E., Hura, G.L., Dyer, K., Fox, B.G., 2013. Fusion of dioxygenase and lignin-binding domains in a novel secreted enzyme from cellulolytic Streptomyces sp. SirexAA-E. J Biol Chem 288, 18574-18587.

Booth, W.T., Schlachter, C.R., Pote, S., Ussin, N., Mank, N.J., Klapper, V., Offermann, L.R., Tang, C., Hurlburt, B.K., Chruszcz, M., 2018. Impact of an N-terminal Polyhistidine Tag on Protein Thermal Stability. ACS Omega 3, 760-768.

Borowski, T., Siegbahn, P.E., 2006a. Mechanism for catechol ring cleavage by non-heme iron intradiol dioxygenases: a hybrid DFT study. Journal of the American Chemical Society 128, 12941-12953.

Borowski, T., Siegbahn, P.E.M., 2006b. Mechanism for catechol ring cleavage by non-heme iron intradiol dioxygenases: A hybrid DFT study. Journal of the American Chemical Society 128, 12941-12953.

Broderick, J.B., 1999. Catechol dioxygenases. Essays Biochem 34, 173-189.

Brown, C.K., Vetting, M.W., Earhart, C.A., Ohlendorf, D.H., 2004. Biophysical analyses of designed and selected mutants of protocatechuate 3,4-dioxygenase1. Annu Rev Microbiol 58, 555-585.

Bugg, T.D.H., 2003. Dioxygenase enzymes: catalytic mechanisms and chemical models. Tetrahedron 59, 7075-7101.

Caglio, R., Pessione, E., Valetti, F., Giunta, C., Ghibaudi, E., 2013. An EPR, thermostability and pH-dependence study of wild-type and mutant forms of catechol 1,2-dioxygenase from Acinetobacter radioresistens S13. Biometals 26, 75-84.

Caglio, R., Valetti, F., Caposio, P., Gribaudo, G., Pessione, E., Giunta, C., 2009. Fine-tuning of catalytic properties of catechol 1,2-dioxygenase by active site tailoring. Chembiochem 10, 1015-1024.

Chen, W., Hasegawa, D.K., Kaur, N., Kliot, A., Pinheiro, P.V., Luan, J., Stensmyr, M.C., Zheng, Y., Liu, W., Sun, H., 2016. The draft genome of whitefly Bemisia tabaci MEAM1, a global crop pest, provides novel insights into virus transmission, host adaptation, and insecticide resistance. BMC biology 14, 110.

Cowtan, K., 2006. The Buccaneer software for automated model building. 1. Tracing protein chains. Acta crystallographica 62, 1002-1011.

Darriba, D., Taboada, G.L., Doallo, R., Posada, D., 2011. ProtTest 3: fast selection of best-fit models of protein evolution. Bioinformatics (Oxford, England) 27, 1164-1165.

Davis, I.W., Leaver-Fay, A., Chen, V.B., Block, J.N., Kapral, G.J., Wang, X., Murray, L.W., Arendall, W.B., 3rd, Snoeyink, J., Richardson, J.S., Richardson, D.C., 2007. MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. Nucleic acids research 35, W375-383.

Davis, M.I., Orville, A.M., Neese, F., Zaleski, J.M., Lipscomb, J.D., Solomon, E.I., 2002. Spectroscopic and Electronic Structure Studies of Protocatechuate 3, 4-Dioxygenase: Nature of Tyrosinate− Fe (III) Bonds and Their Contribution to Reactivity. Journal of the American Chemical Society 124, 602-614.

DeLano W., S., 2002. The PyMOL Molecular Graphics System.

Dermauw, W., Osborne, E.J., Clark, R.M., Grbic, M., Tirry, L., Van Leeuwen, T., 2013a. A burst of ABC genes in the genome of the polyphagous spider mite Tetranychus urticae. BMC genomics 14, 317.

Dermauw, W., Wybouw, N., Rombauts, S., Menten, B., Vontas, J., Grbic, M., Clark, R.M., Feyereisen, R., Van Leeuwen, T., 2013b. A link between host plant adaptation and pesticide resistance in the polyphagous spider mite Tetranychus urticae. Proceedings of the National Academy of Sciences of the United States of America.

Emsley, P., Cowtan, K., 2004. Coot: model-building tools for molecular graphics. Acta crystallographica 60, 2126-2132.

Eschenfeldt, W.H., Makowska-Grzyska, M., Stols, L., Donnelly, M.I., Jedrzejczak, R., Joachimiak, A., 2013. New LIC vectors for production of proteins from genes containing rare codons. J Struct Funct Genomics 14, 135-144.

Faddeeva-Vakhrusheva, A., Derks, M.F., Anvar, S.Y., Agamennone, V., Suring, W., Smit, S., van Straalen, N.M., Roelofs, D., 2016. Gene family evolution reflects adaptation to soil environmental stressors in the genome of the collembolan Orchesella cincta. Genome biology and evolution 8, 2106-2117.

Ferraroni, M., Kolomytseva, M., Scozzafava, A., Golovleva, L., Briganti, F., 2013. X-ray structures of 4-chlorocatechol 1,2-dioxygenase adducts with substituted catechols: new perspectives in the molecular basis of intradiol ring cleaving dioxygenases specificity. J Struct Biol 181, 274-282.

Ferraroni, M., Kolomytseva, M.P., Solyanikova, I.P., Scozzafava, A., Golovleva, L.A., Briganti, F., 2006. Crystal structure of 3-chlorocatechol 1,2-dioxygenase key enzyme of a new modified ortho-pathway from the Gram-positive Rhodococcus opacus 1CP grown on 2-chlorophenol. J Mol Biol 360, 788-799.

Frazee, R.W., Livingston, D.M., LaPorte, D.C., Lipscomb, J.D., 1993. Cloning, sequencing, and expression of the Pseudomonas putida protocatechuate 3,4-dioxygenase genes. J Bacteriol 175, 6194-6202.

Gasteiger, E., Gattiker, A., Hoogland, C., Ivanyi, I., Appel, R.D., Bairoch, A., 2003. ExPASy: The proteomics server for in-depth protein knowledge and analysis. Nucleic acids research 31, 3784-3788.

Grbic, M., Khila, A., Lee, K.Z., Bjelica, A., Grbic, V., Whistlecraft, J., Verdon, L., Navajas, M., Nagy, L., 2007. Mity model: Tetranychus urticae, a candidate for chelicerate model organism. Bioessays 29, 489-496.

Grbic, M., Van Leeuwen, T., Clark, R.M., Rombauts, S., Rouze, P., Grbic, V., Osborne, E.J., Dermauw, W., Ngoc, P.C., Ortego, F., Hernandez-Crespo, P., Diaz, I., Martinez, M., Navajas, M., Sucena, E., Magalhaes, S., Nagy, L., Pace, R.M., Djuranovic, S., Smagghe, G., Iga, M., Christiaens, O., Veenstra, J.A., Ewer, J., Villalobos, R.M., Hutter, J.L., Hudson, S.D., Velez, M., Yi, S.V., Zeng, J., Pires-daSilva, A., Roch, F., Cazaux, M., Navarro, M., Zhurov, V., Acevedo, G., Bjelica, A., Fawcett, J.A., Bonnet, E., Martens, C., Baele, G., Wissler, L., Sanchez-Rodriguez, A., Tirry, L., Blais, C., Demeestere, K., Henz, S.R., Gregory, T.R., Mathieu, J., Verdon, L., Farinelli, L., Schmutz, J., Lindquist, E., Feyereisen, R., Van de Peer, Y., 2011. The genome of Tetranychus urticae reveals herbivorous pest adaptations. Nature 479, 487-492.

Guzik, U., Hupert-Kocurek, K., Salek, K., Wojcieszynska, D., 2013a. Influence of metal ions on bioremediation activity of protocatechuate 3,4-dioxygenase from Stenotrophomonas maltophilia KB2. World J Microbiol Biotechnol 29, 267-273.

Guzik, U., Hupert-Kocurek, K., Sitnik, M., Wojcieszynska, D., 2013b. High activity catechol 1,2-dioxygenase from Stenotrophomonas maltophilia strain KB2 as a useful tool in cis,cis-muconic acid production. Antonie Van Leeuwenhoek 103, 1297-1307.

Guzik, U., Hupert-Kocurek, K., Sitnik, M., Wojcieszyńska, D., 2013c. High activity catechol 1, 2-dioxygenase from Stenotrophomonas maltophilia strain KB2 as a useful tool in cis, cis-muconic acid production. Antonie van Leeuwenhoek 103, 1297-1307.

Guzik, U., Hupert-Kocurek, K., Wojcieszyńska, D., 2013d. Intradiol dioxygenases—the key enzymes in xenobiotics degradation, Biodegradation of hazardous and special products. InTech.

Holm, L., Rosenstrom, P., 2010. Dali server: conservation mapping in 3D. Nucleic acids research 38, W545-W549.

Jeppson, L.R., Keifer, H.H., Baker, E.W., 1975. Mites injurious to economic plants. Univ of California Press.

638 Jurrus, E., Engel, D., Star, K., Monson, K., Brandi, J., Felberg, L.E., Brookes, D.H., Wilson, L., Chen, J.,
639 Liles, K., Chun, M., Li, P., Gohara, D.W., Dolinsky, T., Konecny, R., Koes, D.R., Nielsen, J.E., Head-
640 Gordon, T., Geng, W., Krasny, R., Wei, G.W., Holst, M.J., McCammon, J.A., Baker, N.A., 2018.
641 Improvements to the APBS biomolecular solvation software suite. Protein Sci 27, 112-128.
642 Knoot, C.J., Purpero, V.M., Lipscomb, J.D., 2015. Crystal structures of alkylperoxo and anhydride
643 intermediates in an intradiol ring-cleaving dioxygenase. Proceedings of the National Academy of Sciences
644 of the United States of America 112, 388-393.
645 Krissinel, E., Henrick, K., 2004. Secondary-structure matching (SSM), a new tool for fast protein structure
646 alignment in three dimensions. Acta crystallographica 60, 2256-2268.
647 Krissinel, E., Henrick, K., 2007. Inference of macromolecular assemblies from crystalline state. J Mol Biol
648 372, 774-797.
649 Laskowski, R.A., 2017. The ProFunc Function Prediction Server. Methods Mol Biol 1611, 75-95.
650 Matera, I., Ferraroni, M., Kolomytseva, M., Golovleva, L., Scozzafava, A., Briganti, F., 2010. Catechol
651 1,2-dioxygenase from the Gram-positive Rhodococcus opacus 1CP: quantitative structure/activity
652 relationship and the crystal structures of native enzyme and catechols adducts. J Struct Biol 170, 548-564.
653 Micalella, C., Martignon, S., Bruno, S., Pioselli, B., Caglio, R., Valetti, F., Pessione, E., Giunta, C., Rizzi,
654 M., 2011. X-ray crystallography, mass spectrometry and single crystal microspectrophotometry: a
655 multidisciplinary characterization of catechol 1,2 dioxygenase. Biochimica et biophysica acta 1814, 817-
656 823.
657 Minor, W., Cymborowski, M., Otwinowski, Z., Chruszcz, M., 2006. HKL-3000: the integration of data
658 reduction and structure solution--from diffraction images to an initial model in minutes. Acta
659 crystallographica 62, 859-866.
660 Murshudov, G.N., Skubak, P., Lebedev, A.A., Pannu, N.S., Steiner, R.A., Nicholls, R.A., Winn, M.D.,
661 Long, F., Vagin, A.A., 2011. REFMAC5 for the refinement of macromolecular crystal structures. Acta
662 Crystallographica Section D-Biological Crystallography 67, 355-367.
663 Orville, A.M., Lipscomb, J.D., 1993. Simultaneous binding of nitric oxide and isotopically labeled
664 substrates or inhibitors by reduced protocatechuate 3,4-dioxygenase. J Biol Chem 268, 8596-8607.
665 Otwinowski, Z., Minor W., 1997. Processing of X-ray diffraction data collected in oscillation mode.,
666 Methods in enzymology: Macromolecular crystallography , part A. Academic Press, New York, pp. 307-
667 326.
668 Painter, J., Merritt, E.A., 2006. TLSMD web server for the generation of multi-group TLS models. Journal
669 of Applied Crystallography 39, 109-111.

Pettersen, E.F., Goddard, T.D., Huang, C.C., Couch, G.S., Greenblatt, D.M., Meng, E.C., Ferrin, T.E., 2004. UCSF Chimera--a visualization system for exploratory research and analysis. J Comput Chem 25, 1605-1612.

Ponstingl, H., Henrick, K., Thornton, J.M., 2000. Discriminating between homodimeric and monomeric proteins in the crystalline state. Proteins 41, 47-57.

Ronquist, F., Teslenko, M., Van Der Mark, P., Ayres, D.L., Darling, A., Höhna, S., Larget, B., Liu, L., Suchard, M.A., Huelsenbeck, J.P., 2012. MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. Systematic biology 61, 539-542.

Sainsbury, P.D., Mineyeva, Y., Mycroft, Z., Bugg, T.D., 2015. Chemical intervention in bacterial lignin degradation pathways: Development of selective inhibitors for intradiol and extradiol catechol dioxygenases. Bioorg Chem 60, 102-109.

Santamaria, M.E., Gonzalez-Cabrera, J., Martinez, M., Grbic, V., Castanera, P., Diaz, L., Ortego, F., 2015. Digestive proteases in bodies and faeces of the two-spotted spider mite, Tetranychus urticae. Journal of insect physiology 78, 69-77.

Schlachter, C.R., Klapper, V., Wybouw, N., Radford, T., Van Leeuwen, T., Grbic, M., Chruszcz, M., 2017. Structural Characterization of a Eukaryotic Cyanase from Tetranychus urticae. J Agric Food Chem 65, 5453-5462.

Slabinski, L., Jaroszewski, L., Rychlewski, L., Wilson, I.A., Lesley, S.A., Godzik, A., 2007. XtalPred: a web server for prediction of protein crystallizability. Bioinformatics (Oxford, England) 23, 3403-3405.

Tchesnokov, E.P., Wilbanks, S.M., Jameson, G.N., 2012. A strongly bound high-spin iron(II) coordinates cysteine and homocysteine in cysteine dioxygenase. Biochemistry 51, 257-264.

Travkin, V.M., Jadan, A.P., Briganti, F., Scozzafava, A., Golovleva, L.A., 1997. Characterization of an intradiol dioxygenase involved in the biodegradation of the chlorophenoxy herbicides 2,4-D and 2,4,5-T. FEBS Lett 407, 69-72.

Tsai, S.C., Li, Y.K., 2007. Purification and characterization of a catechol 1,2-dioxygenase from a phenol degrading Candida albicans TL3. Arch Microbiol 187, 199-206.

UniProt, C., 2015. UniProt: a hub for protein information. Nucleic acids research 43, D204-212.

Vagin, A., Teplyakov, A., 1997. MOLREP: an automated program for molecular replacement. Journal of Applied Crystallography 30, 1022-1025.

Vaillancourt, F.H., Bolin, J.T., Eltis, L.D., 2006. The ins and outs of ring-cleaving dioxygenases. Crit Rev Biochem Mol Biol 41, 241-267.

Valley, M.P., Brown, C.K., Burk, D.L., Vetting, M.W., Ohlendorf, D.H., Lipscomb, J.D., 2005. Roles of the equatorial tyrosyl iron ligand of protocatechuate 3,4-dioxygenase in catalysis. Biochemistry 44, 11024-11039.

Van Leeuwen, T., Dermauw, W., 2016. The Molecular Evolution of Xenobiotic Metabolism and Resistance in Chelicerate Mites. Annual review of entomology 61, 475-498.

Van Leeuwen, T., Tirry, L., Yamamoto, A., Nauen, R., Dermauw, W., 2015. The economic importance of acaricides in the control of phytophagous mites and an update on recent acaricide mode of action research. Pestic Biochem Physiol 121, 12-21.

Weng, J.K., 2014. The evolutionary paths towards complexity: a metabolic perspective. New Phytol 201, 1141-1149.

Widhalm, J.R., Dudareva, N., 2015. A familiar ring to it: biosynthesis of plant benzoic acids. Mol Plant 8, 83-97.

Wojcik, A., Borowski, T., Broclawik, E., 2011. The mechanism of the reaction of intradiol dioxygenase with hydroperoxy probe A DFT study. Catal Today 169, 207-216.

Yoshida, Y., Koutsovoulos, G., Laetsch, D.R., Stevens, L., Kumar, S., Horikawa, D.D., Ishino, K., Komine, S., Kunieda, T., Tomita, M., 2017. Comparative genomics of the tardigrades Hypsibius dujardini and Ramazzottius varieornatus. PLoS biology 15, e2002266.

Zheng, H., Chordia, M.D., Cooper, D.R., Chruszcz, M., Muller, P., Sheldrick, G.M., Minor, W., 2014. Validation of metal-binding sites in macromolecular structures with the CheckMyMetal web server. Nat Protoc 9, 156-170.

Zheng, H., Chruszcz, M., Lasota, P., Lebioda, L., Minor, W., 2008. Data mining of metal ion environments present in protein structures. J Inorg Biochem 102, 1765-1776.

Footnotes

The abbreviations used are: ID-RCD, intradiol-ring cleavage dioxygenase; MBP, Maltose Binding Protein; ICP-MS, inductively coupled plasma mass spectrometry; LMCT, ligand to metal charge transfer; DSF, Differential Scanning Fluorimetry; β-ME, mercaptoethanol; RMSD, root mean square deviation; PDB, Protein Data Bank.

730 **Table 1.** Summary of kinetic data (25°C) for cleavage of catechol and 4-methylcatechol by the His-tag
731 cleaved version of mTuIDRCD.

732

Buffers	**pH**	**K_M (µM)**	**V_{max}(µMmin^{-1})**	**k_{cat} (min^{-1})**	**k_{cat}/K_M (µM^{-1}min^{-1})**
Catechol					
Sodium phosphate	6.0	9.90 ± 0.97	0.128 ± 0.002	0.025 ± 0.001	0.002 ± 0.001
Sodium phosphate	6.5	11.8 ± 1.3	0.231 ± 0.005	0.046 ± 0.001	0.003 ± 0.001
Sodium phosphate	7.0	14.5 ± 1.9	0.32 ± 0.01	0.064 ± 0.002	0.004 ± 0.001
Sodium phosphate	7.5	16.9 ± 1.7	0.52 ± 0.01	0.104 ± 0.002	0.006 ± 0.001
HEPES	8.0	23.9 ± 2.7	1.02 ± 0.03	0.204 ± 0.006	0.008 ± 0.002
AMPSO	8.5	26.2 ± 2.6	1.14 ± 0.03	0.228 ± 0.006	0.008 ± 0.002
AMPSO	9.0	30.2 ± 3.7	1.25 ± 0.04	0.250 ± 0.008	0.008 ± 0.002
AMPSO	9.5	40.8 ± 6.3	1.37 ± 0.07	0.274 ± 0.014	0.006 ± 0.002
4-Methylcatechol					
Sodium phosphate	6.0	13.3 ± 1.2	0.201 ± 0.004	0.402 ± 0.008	0.030 ± 0.001
Sodium phosphate	6.5	21.1 ± 3.5	0.22 ± 0.01	0.44 ± 0.02	0.020 ± 0.001
Sodium phosphate	7.0	25.7 ± 5.2	0.53 ± 0.02	1.06 ± 0.04	0.040 ± 0.001
Sodium phosphate	7.5	29.3 ± 3.5	0.59 ± 0.02	1.18 ± 0.04	0.04 ± 0.01
HEPES	8.0	48.3 ± 5.5	1.35 ± 0.05	2.70 ± 0.10	0.05 ± 0.01
AMPSO	8.5	56.0 ± 7.7	1.41 ± 0.07	2.82 ± 0.14	0.05 ± 0.01
CHES	9.0	99.8 ± 13.2	1.90 ± 0.10	3.80 ± 0.20	0.03 ± 0.01
CHES	9.5	123.6 ± 10.5	2.23 ± 0.08	4.46 ± 0.16	0.03 ± 0.01

733

734

735

736

737 **Table 2.** Data collection and refinement statistics for TuIDRCD. Values in parentheses are for the highest
738 resolution shell.

Protein	mbp-(t48)TuIDRCD	mTuIDRCD
PDB accession code	5VG2	6BDJ
Data Collection		
Diffraction source	APS, 22ID	APS, 22ID
Wavelength (Å)	1.000	1.000
Space group	P2	P2
a, *b*, *c*, *β* (Å, °)	60.5, 43.1, 165.7, 95.2	61.1, 45.4, 83.0, 94.3
Resolution range (Å)	50.0-2.45 (2.49-2.45)	40.0-2.15 (2.19-2.15)
No. of unique reflections	30081 (1539)	24133 (929)
Completeness (%)	95.5 (97.7)	95.3 (72.3)
Redundancy	3.1 (3.0)	4.0 (3.1)
$\langle I/\sigma(I)\rangle$	14.5 (2.0)	32.0 (3.4)
$R_{r.i.m.}$	0.116 (0.665)	0.064 (0.300)
$R_{p.i.m.}$	0.064 (0.371)	0.032 (0.160)
Overall B factor from Wilson plot (Å^2)	37.3	49.8
Refinement		
Resolution range (Å)	50.0-2.45 (2.52-2.45)	40.0-2.15 (2.21-2.15)
Completeness (%)	95.0 (90.2)	95.3 (76.2)
No. of reflections, working set	28666 (2007)	22883 (1321)
No. of reflections, test set	1414 (103)	1095 (62)
Final R_{cryst}	0.191 (0.271)	0.214 (0.266)
Final R_{free}	0.231 (0.284)	0.248 (0.286)
Rmsd Bonds (Å)	0.015	0.011
Rmsd Angles (°)	1.7	1.6
Ramachandran Plot		
Most favored (%)	97	97
Allowed (%)	100	100

739

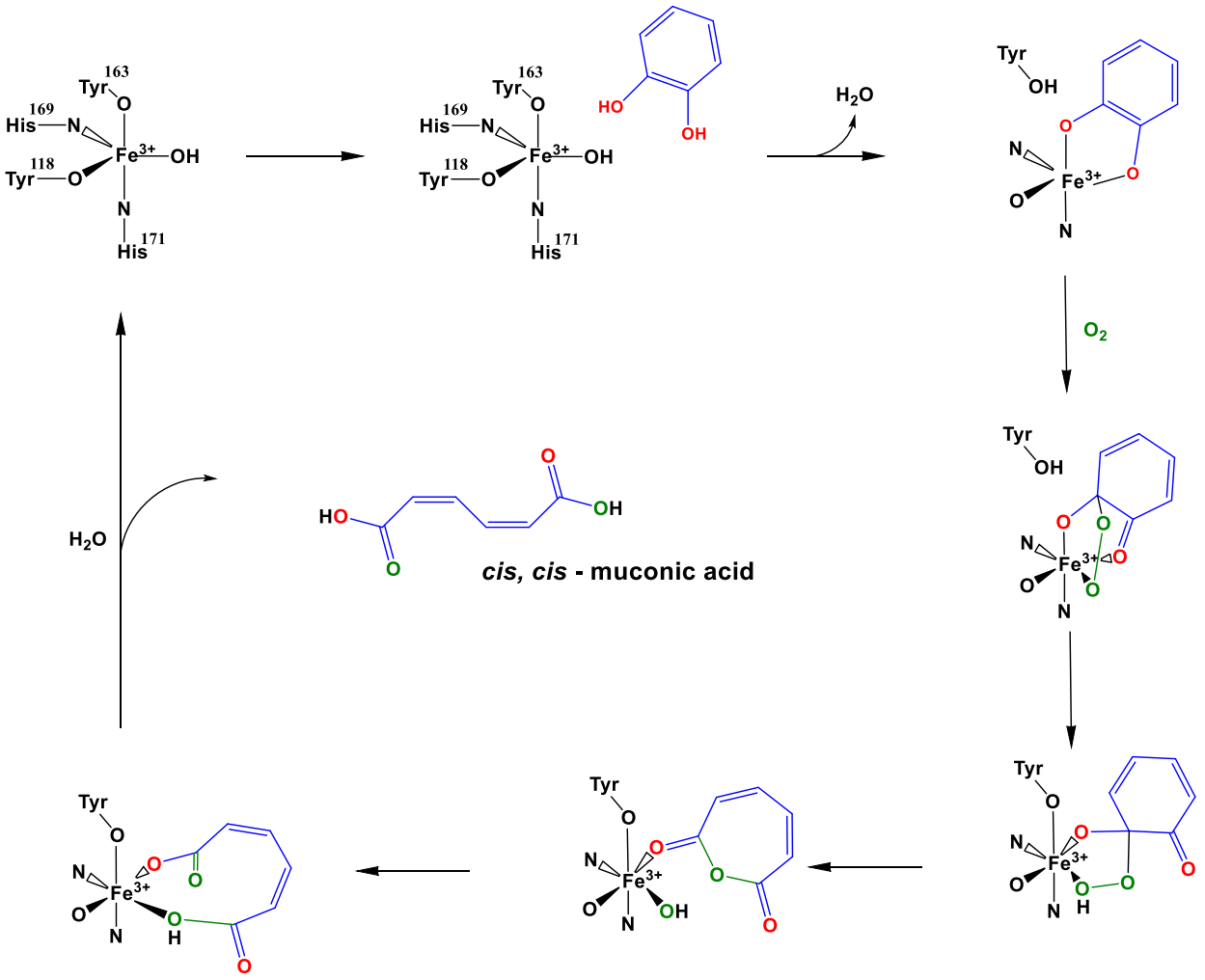


Figure 1. Schematic diagram of catechol cleavage by an intradiol ring-cleaving dioxygenase (Borowski and Siegbahn, 2006a; Knoot et al., 2015; Wojcik et al., 2011). Residue numbers correspond to those of TuIDRCD.

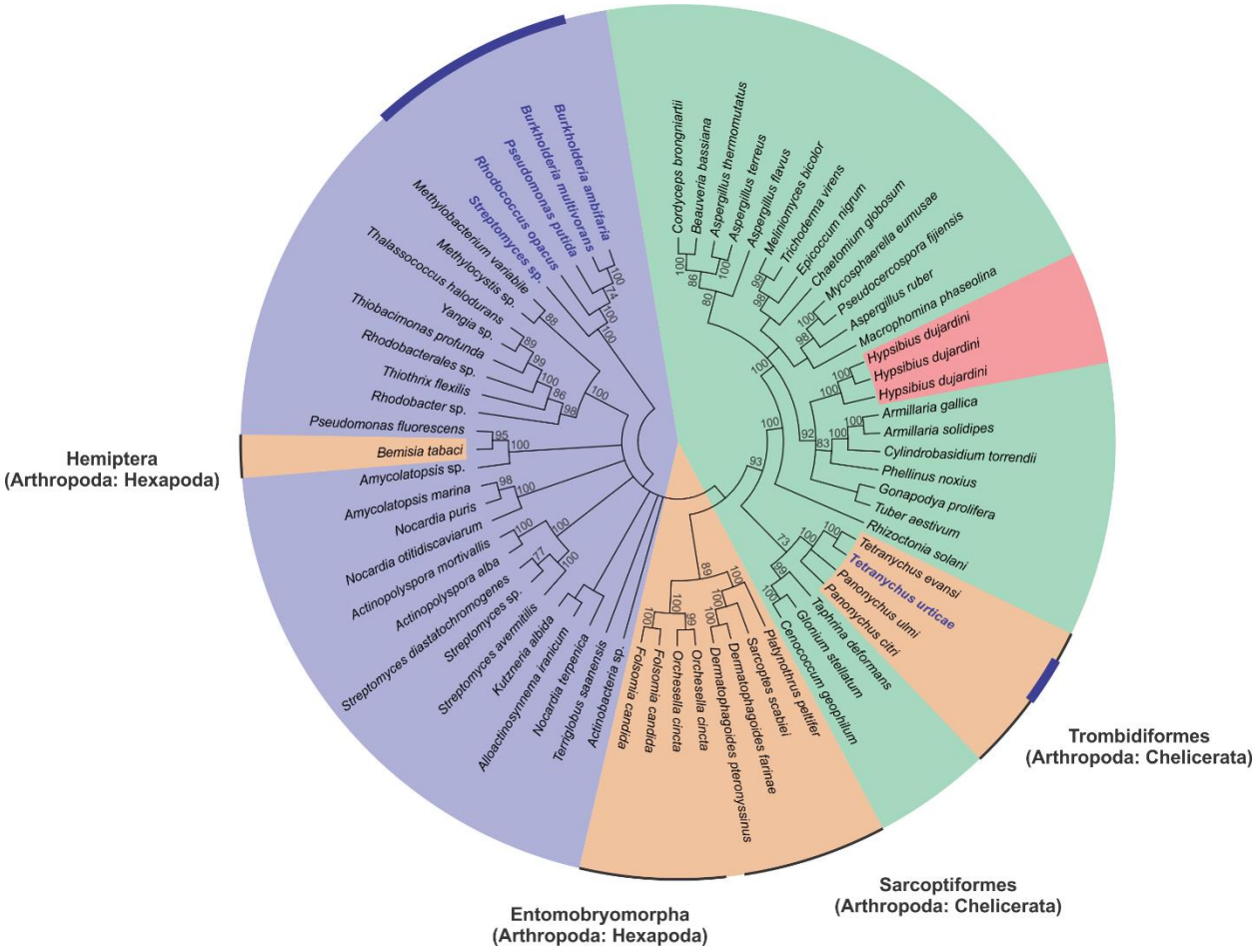
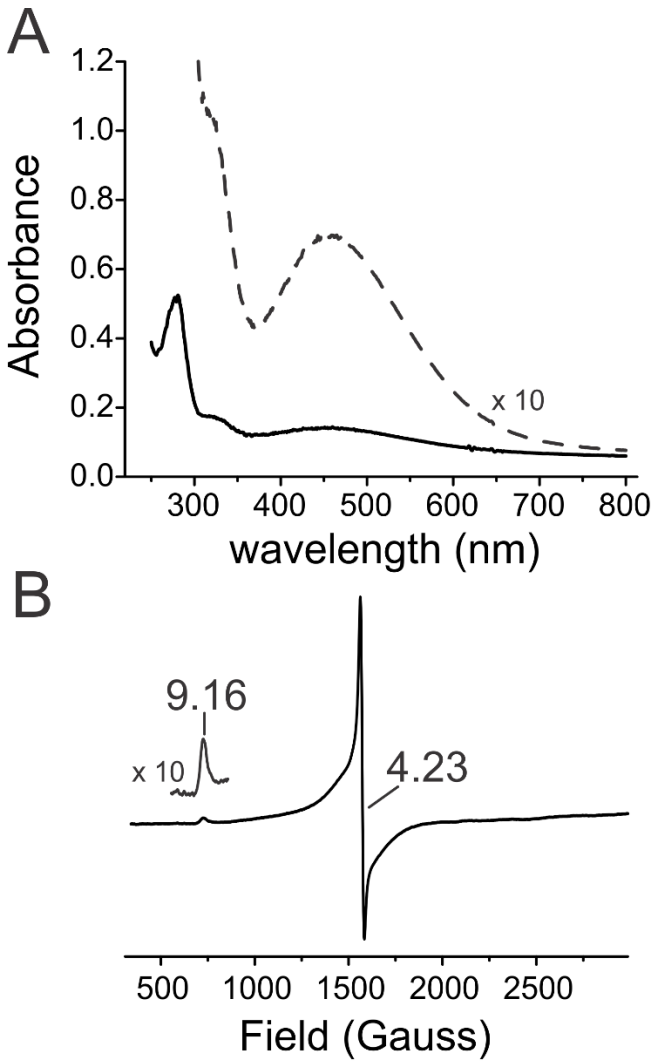


Figure 2. Phylogenetic reconstruction of the evolutionary history and origin of tardigrade and arthropod intradiol ring-cleavage dioxygenases using a Bayesian method. Only Bayesian posterior probabilities higher than 75 are depicted. Enzymes of which the crystal structure has been determined are indicated by a blue bold font and a blue circumferential line. The colored background reflects the position of the species within the tree of life with bacteria: blue, fungi: green, arthropods: orange and tardigrades: red. The distinct taxa within the Arthropoda phylum are delineated with a black circumferential line, detailing the respective orders and subphyla.

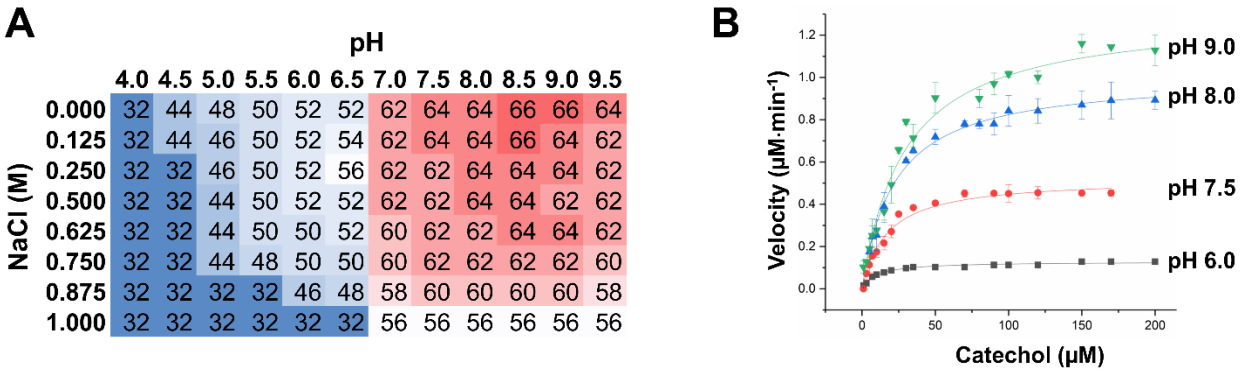


760

761

762 **Figure 3.** UV-visible (A) and electron paramagnetic resonance (B) spectra of as-purified mTuIDRCD.

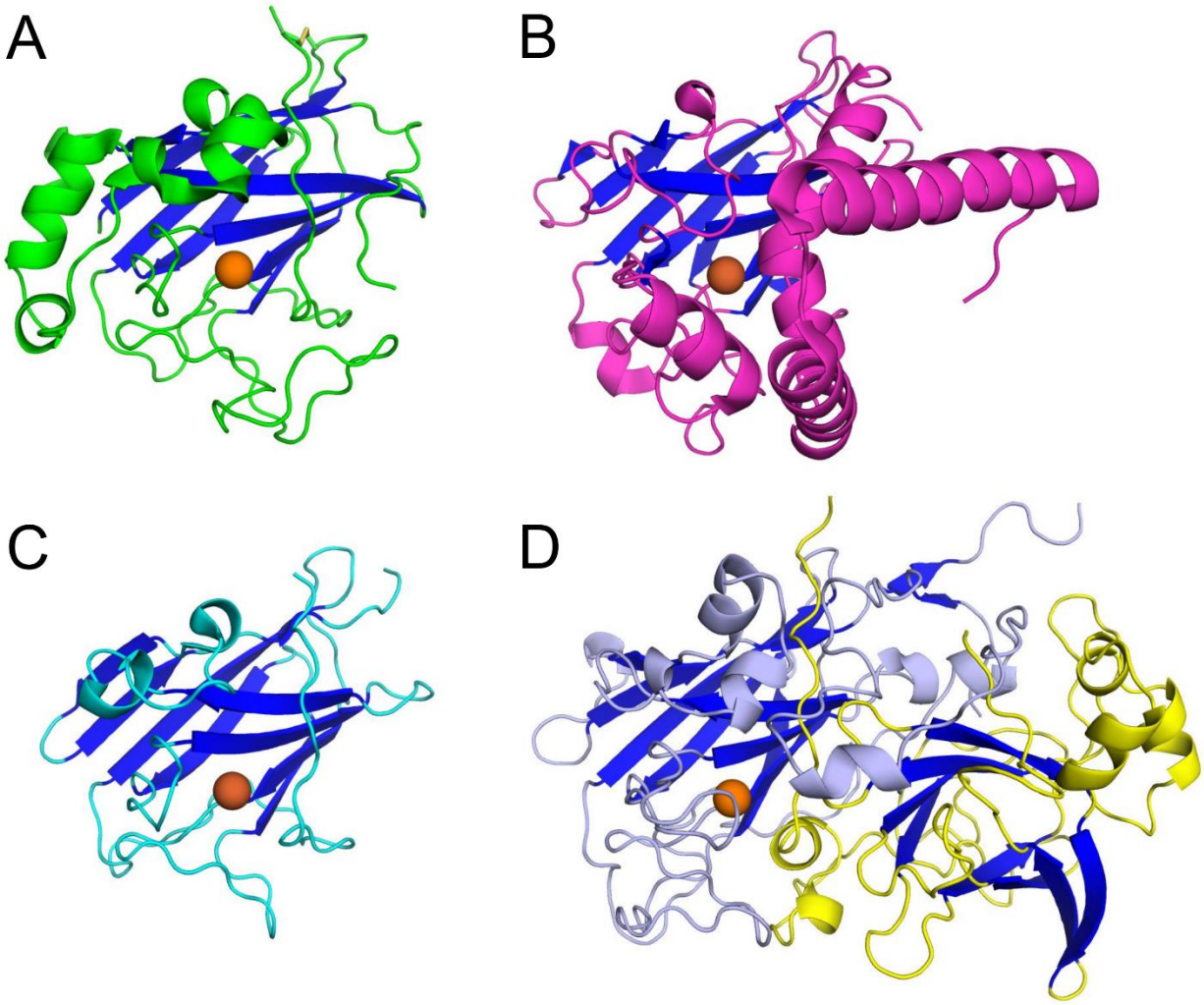
763



764

765 **Figure 4.** *A*, Effect of pH and salt concentration on stability of mTuIDRCD. Blue displays low melting
766 temperature, white is average, and red is high melting temperature. All temperatures are in °C. The standard
767 deviation was typically less than 1°C for all experiments. *B*, Michaelis-Menten graphs representing the
768 initial velocity (µM min^{-1}) vs catechol concentration (µM) in pH 6.0. 7.5, 8.0 and 9.0. The absorbance of
769 *cis*,*cis*-muconic acid was measured at 260 nm with the extinction coefficient of 16,800 M^{-1} cm^{-1} (Guzik et
770 al., 2013c).

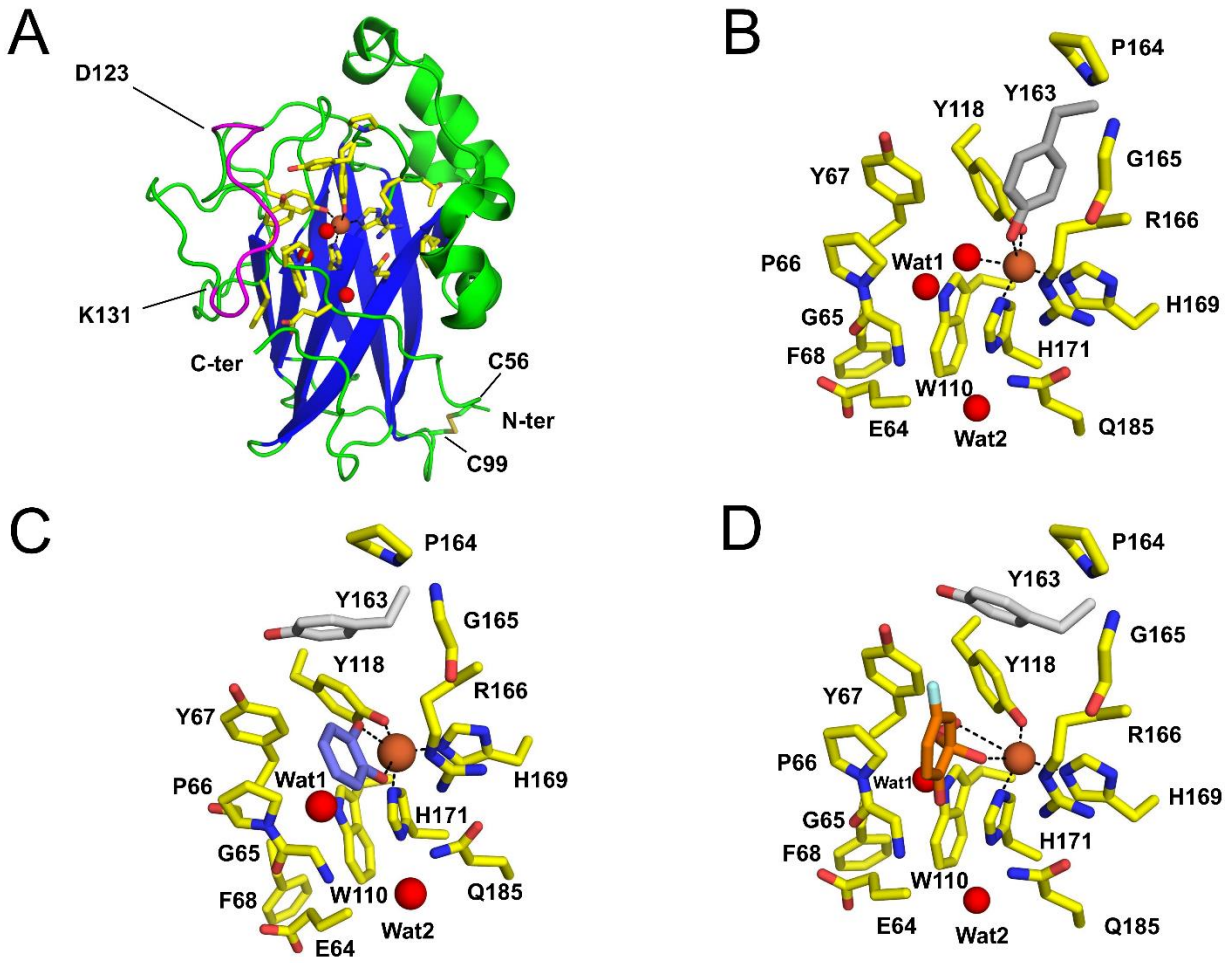
771



772

773 **Figure 5.** *A*, Cartoon representation of mTuIDRCD (PDB code: 6BDJ). The β-sandwich core is conserved
774 in ID-RCDs, and it was used to align in similar orientation all depicted structures. β-strands are marked in
775 blue. *B*, Structure of a single chain of catechol 1,2-dioxygenase from *R. opacus* (PDB code: 3I51). *C*,
776 SACTE_2871 - dioxygenase from *Streptomyces* sp. SirexAA-E (PDB code: 4ILV). *D*, Protocatechuate 3,4-
777 dioxygenase from *P. putida* (α and β chains; PDB code: 4WHR). Position of the iron is marked with orange
778 spheres. Only β chain binds Fe^{3+}.
779

780



781

782 **Figure 6.** *A*, An overall structure of mTuIDRCD. The core β-sandwich is marked in blue. The flexible
783 D123-K131 region is shown in magenta. Residues near the active site that are conserved in both TuIDRCD
784 and SACTE_2871 are shown as yellow sticks. Fe^{3+} is represented as an orange sphere, while water
785 molecules are shown as red spheres. *B*, Active site of *apo*-mTuIDRCD. Residues coordinating the iron
786 include H169, H171, Y118, and Y163. Y163 is shown in a conformation allowing for metal binding.
787 Structurally conserved water molecules are labeled as Wat1 and Wat2. *C*, A putative *holo*-form of
788 mTuIDRCD. Y163 is shown in a "displaced" conformation. Catechol molecule (purple sticks) is modeled
789 based on the structure of catechol 1,2-dioxygenase from *R. opacus* 1CP (PDB code: 3HHY). *D*, A putative
790 alkylperoxo reaction intermediate (orange sticks) modeled based on a structure of protocatechuate 3,4-
791 dioxygenase from *P. putida* (PDB code: 4WHQ).
792

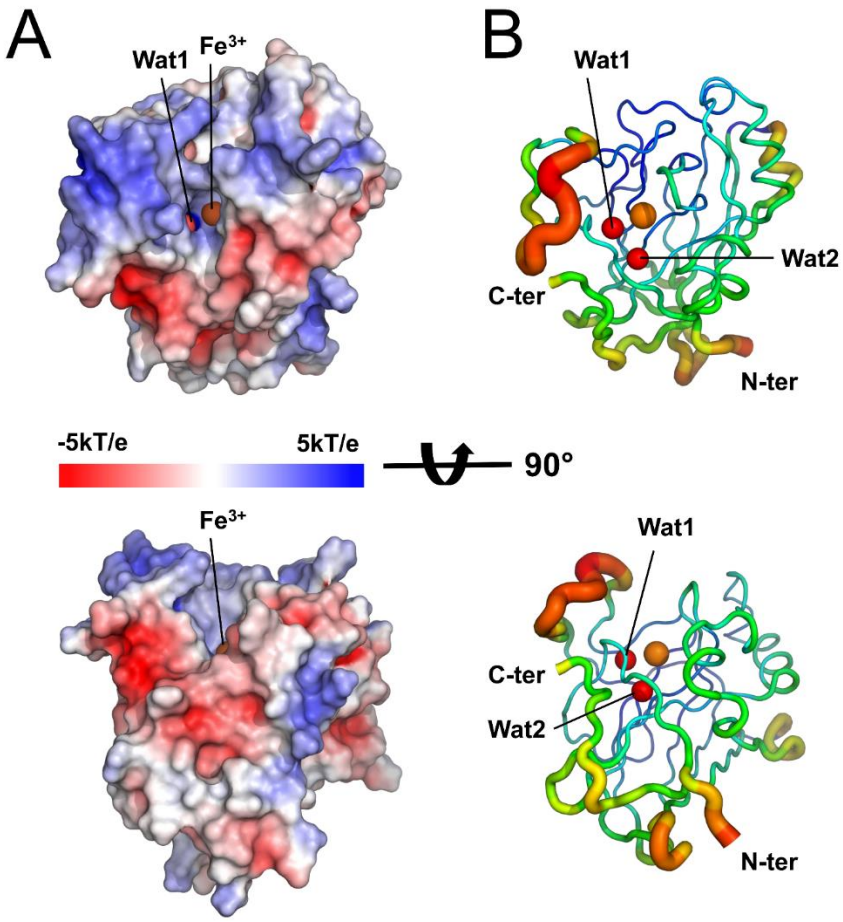


Figure 7. *A*, Distribution of charges on a surface of TuIDRCD. A protein chain with Y163 facing away from Fe^{3+} was used to generate the molecular surface. *B*, Trace of the TuIDRCD main chain showing a relative mobility of various protein fragments. A diameter of the ribbon is proportional to B-factor value. Positively charged D123-K131 loop region found near the active site has a significant mobility. Structurally conserved water molecules are labeled as Wat1 and Wat2.

Structural and functional characterization of an intradiol ring-cleavage dioxygenase from the polyphagous spider mite *Tetranychus urticae* Koch

Caleb R. Schlachter[1], Leily Daneshian[1], Jose Amaya[1], Vincent Klapper[1], Nicky Wybouw[2,3], Tomasz Borowski[4], Thomas Van Leeuwen[2,3], Vojislava Grbic[5,6], Miodrag Grbic[5,6], Thomas M. Makris[1], Maksymilian Chruszcz[1]*

[1] Department of Chemistry and Biochemistry, University of South Carolina, Columbia, SC 29208, USA
[2] Institute for Biodiversity and Ecosystem Dynamics, University of Amsterdam, Science Park 904, 1098 XH, Amsterdam, the Netherlands
[3] Department of Plants and Crops, Ghent University, Ghent, B-9000, Belgium
[4] Jerzy Haber Institute of Catalysis and Surface Chemistry, Polish Academy of Sciences, 30–239 Krakow, Poland
[5] Department of Biology, Western University, London, Ontario N6A 5B7, Canada
[6] University of La Rioja, Logrono, Spain

*To whom correspondence should be addressed: Maksymilian Chruszcz, Department of Chemistry and Biochemistry, University of South Carolina, Columbia, SC 29208 USA;
chruszcz@mailbox.sc.edu; Tel: (803) 777-7399, FAX: (803) 777-9521.

Cloning (t48)TuIDRCD-MBP

Primers used for pMCSG29 cloning are listed in Table S1. The pJExpress411 vector containing dioxygenase was used as template for the first PCR with the primers p29-Dioxy-F and p29-Dioxy-R. PCR was performed using Phusion polymerase following the manufacturer's protocol (NEB, Ipswich, MA). and thermalcycling as follows: initial denaturation 98˚C 30 seconds, 30 cycles of 98˚C for 10 seconds, 68˚C for 25 seconds and 72˚C for 25 seconds with a final extension at 72˚C for 5.0 minutes followed by an infinite hold at 12˚C. PCR product was purified using a GeneJet Gel Extraction Kit (ThermoFisher, Grand Island, NY). The pMCSG29 was amplified at the LIC site to generate blunt-ended vector with the primers pMCSG28/29-F and pMCSG28/29-R. KOD polymerase (Millipore, Billerica, MA) was used following the manufacturer's protocol with the addition of 1.0 M betaine monohydrate to amplify pMCSG29 by PCR. Thermalcycling was as follows: initial denaturation 95˚C for 2 minutes, 30 cycles of 95˚C for 20 seconds, 61˚C for 10 seconds and 70˚C for 2 minutes 30 seconds followed by an infinite hold at 12˚C. After PCR, product was purified by gel excision and cleaned up using a GeneJet Gel Extraction Kit (ThermoFisher, Grand Island, NY).

Sticky ends for LIC were made by incubating gel excised PCR products with T4 DNA polymerase (NEB, Ipswich, MA). A 40 µL reaction containing 170 fmoles of pMCSG29, 5.0 mM DTT, 2.5 mM dATP, 1X NEB2.1 buffer and 3.0 units of T4 DNA polymerase, and a 40 µL reaction containing 670 fmoles of dioxygenase insert, 5.0 mM DTT, 2.5 mM dTTP, 1X NEB2.1 buffer and 3.0 units of T4 DNA polymerase were incubated at room temperature for 30 minutes. The T4 DNA polymerase was then inactivated at 75ºC for 20 minutes and the reactions were mixed 10:10 at room temperature for 5.0 minutes. Next, 1.0 µL of 25 mM EDTA was added and the reaction was incubated another 5.0 minutes at room temperature. Transformation was performed with heat shock as mentioned previously after adding the 20 µL LIC reaction to the cells, and cells were plated on LB-ampicillin (50 µg/mL) plates at 37ºC for 16 hours.

Cloning (t48)TuIDRCD into pMBPcs1

The (t48)TuIDRCD gene was amplified using the primers pMBPcs1-Dioxy-F and pMBPcs1-Dioxy-R shown in Table S1. PCR was performed using Phusion polymerase following the manufacturer's protocol (NEB, Ipswich, MA) with thermalcycling as follows: initial denaturation at 98˚C for 30 seconds, 35 cycles of 98˚C for 10 seconds, 70˚C for 30 seconds and 72˚C for 25 seconds and a final extension of 72˚C for 5.0 minutes followed by an infinite hold at 12˚C. PCR product was purified using a GeneJet Gel Extraction Kit (ThermoFisher, Grand Island, NY). Sticky ends for LIC were made by incubating gel excised PCR products with T4 DNA polymerase (NEB, Ipswich, MA). A 40 µL reaction containing 600 fmoles of pMBPcs1, 5.0 mM DTT, 2.5 mM dATP, 1X NEB2.1 buffer and 3.0 units of T4 DNA polymerase, and a 40 µL reaction containing 1000 fmoles of dioxygenase insert, 5.0 mM DTT, 2.5 mM dTTP, 1X NEB2.1 buffer and 3.0 units of T4 DNA polymerase were incubated at room temperature for 30 minutes. The T4 DNA polymerase was then inactivated at 75ºC for 20 minutes and the reactions were mixed 10:10 at room temperature for 5.0 minutes. Next, 1.0 µL of 25 mM EDTA was added and the reaction was incubated another 5.0 minutes at room temperature. Transformation was performed with heat shock after adding the 20 µL LIC reaction to the cells, and cells were plated on LB-AMP (50 µg/mL) plates at 37ºC for 16 hours.

Generation of mature TuIDRCD

Megaprimers were made by PCR with Q5 polymerase (NEB, Ipswich, MA) using 10 ng of fragment as template and thermalcycling as follows: 98ºC for 30 seconds, 25 cycles of 98ºC for 10 seconds, 62ºC for 10 seconds, 72ºC for 10 seconds, and final extension at 72ºC for 20 seconds followed by an infinite hold at 12ºC. For insertion of the fragment, PCR with Q5 polymerase (NEB, Ipswich, MA), 400 ng megaprimer, and 60 ng of template (TuIDRCD) were used with thermalcycling as follows: 98ºC for 30 seconds, 25 cycles of 98ºC for 10 seconds, 60ºC for 30 seconds, 72ºC for 6 minutes, and final extension at 72ºC for 6 minutes followed by an infinite hold at 12ºC. Twenty units of DpnI was added to the reaction (NEB, Ipswich, MA) and incubated at 37ºC for 2 hours, then DpnI was inactivated at 80ºC for 20 minutes. The reaction was transformed into DH5-α *E. coli* cells via heat shock, plated on LB-kanamycin (50 µg/mL), and grown at 37ºC.

Table S1. Metal content of a mTuIDRCD sample as determined by ICP-MS.

Metal	**Fe**	**Mn**	**Ni**	**Zn**	**Cu**	**Co**
nM of metal/one nM of protein	0.98	0.11	0.02	0.002	0.002	0.004

Table S2. Primers used for TuIDRCD cloning. For primers Dioxy25-F and Dioxy25-R, regions in bold show complementarity to the other.

Primer	Sequence
p29-Dioxy-F	5' GTCTCTCCCATGTCGTTTGTTACCCGTTTCACCGAGT
p29-Dioxy-R	5' GGTTCTCCCCAGCGGCCCACATCAGATTGCT
pMCSG28/29-F	5' GGGGAGAACCTGTACTTCCAATCCGC
pMCSG28/29-R	5' GGGAGAGACTCCTTCTTAAAGTTAAACAAA
pMBPcs1-Dioxy-F	5' TTCGGCTGCTAGTTCGTTTGTTACCCGTTTCACCGAG
pMBPcs1-Dioxy-R	5' GGTTCTCCCCAGCGGCCCACATCAGATTGCTGC
Dioxy25-F	5' AACTCTTGTGCTTCTAAGGAAGAAGTA**GTTCCATCACCTGAAGAA**
Dioxy25-R	5' TTGTTTAACTGATTCTTGACTACACTGACC**TTCTTCAGGTGATGGAAC**
p411-Dioxy25-F	5' TGGGCACGGAAAACTTGTATTTTCAAAGCAACTCTTGTGCTTCTAAGGA
p411-Dioxy25-R	5' CACTCGGTGAAACGGGTAACAAACGATTGTTTAACTGATTCTTGACTAC

Table S3. Dataset used for phylogenetic analyses and generation of Figure 7.

Species	Accession Number	Database
Tetranychus evansi	AFY99040.1	NCBI
Tetranychus urticae	tetur07g02040	ORCAE (http://bioinformatics.psb.ugent.be/orcae/)
Panonychus ulmi	Pu_IDRCD7	(Bajda et al., 2015)
Panonychus citri	Pc_IDRCD14	(Bajda et al., 2015)
Dermatophagoides farinae	KN266226.1	NCBI
Orchesella cincta	ODM96188.1 / Ocin01_10487-PA	NCBI / http://www.collembolomics.nl/collembolomics/index.php
Orchesella cincta	ODN05189.1 / Ocin01_01489-PA	NCBI / http://www.collembolomics.nl/collembolomics/index.php
Folsomia candida	XP_021949837.1 / LOC110847243	NCBI / http://www.collembolomics.nl/collembolomics/index.php
Folsomia candida	XP_021962088.1 / LOC110857783	NCBI / http://www.collembolomics.nl/collembolomics/index.php
Nocardia terpenica	WP_098697000.1	NCBI
Nocardia puris	WP_067502736.1	NCBI
Nocardia otitidiscaviarum	WP_051037357.1	NCBI
Amycolatopsis sp.	WP_094006348.1	NCBI
Streptomyces sp.	WP_095878695.1	NCBI
Rhodobacterales sp.	OHC59208.1	NCBI
Thiobacimonas profunda	WP_083697745.1	NCBI
Thiothrix flexilis	WP_020558911.1	NCBI
Methylocystis sp.	WP_014889962.1	NCBI
Tuber aestivum	CUS10805.1	NCBI
Cylindrobasidium torrendii	KIY66130.1	NCBI
Armillaria solidipes	PBK66679.1	NCBI
Phellinus noxius	PAV17635.1	NCBI
Aspergillus ruber	EYE95593.1	NCBI
Macrophomina phaseolina	EKG10663.1	NCBI
Trichoderma virens	XP_013960844.1	NCBI
Gonapodya prolifera	KXS09040.1	NCBI
Beauveria bassiana	XP_008599008.1	NCBI
Cordyceps brongniartii	OAA34475.1	NCBI
Aspergillus terreus	XP_001215259.1	NCBI
Aspergillus flavus	KOC09022.1	NCBI
Bemisia tabaci	XP_018895614.1	NCBI
Streptomyces sp.	SirexAA-E_4ILT:A	PDB (Protein Data Bank)
Burkholderia multivorans	5UMH:A	PDB (Protein Data Bank)
Burkholderia ambifaria	5VXT:A	PDB (Protein Data Bank)
Rhodococcus opacus	1CP_2BOY:A	PDB (Protein Data Bank)
Pseudomonas putida	DLL-E4_3N9T:A	PDB (Protein Data Bank)
Sarcoptes scabiei	KPM05611.1	NCBI
Alloactinosynnema iranicum	WP_091451389.1	NCBI
Streptomyces avermitilis	OOV21026.1	NCBI
Streptomyces diastatochromogenes	WP_094220118.1	NCBI
Kutzneria albida	WP_081789622.1	NCBI
Taphrina deformans	CCG84187.1	NCBI

Glonium stellatum	OCL09936.1	NCBI
Cenococcum geophilum	OCK93203.1	NCBI
Chaetomium globosum	XP_001223092.1	NCBI
Platynothrus peltifer	LBFO01048654	NCBI
Dermatophagoides pteronyssinus	MQNO02000046.1	NCBI
Rhodobacter sp.	WP_068305520.1	NCBI
Amycolatopsis marina	WP_091675050.1	NCBI
Actinopolyspora alba	WP_092930100.1	NCBI
Actinobacteria bacterium	WP_082417602.1	NCBI
Actinopolyspora mortivallis	WP_019854124.1	NCBI
Pseudomonas fluorescens	WP_017340424.1	NCBI
Thalassococcus halodurans	SEF70789.1	NCBI
Methylobacterium variabile	WP_053080553.1	NCBI
Rhizoctonia solani	CUA69960.1	NCBI
Aspergillus thermomutatus	OXS11685.1	NCBI
Meliniomyces bicolor	PMD60147.1	NCBI
Pseudocercospora fijiensis	XP_007930445.1	NCBI
Mycosphaerella eumusae	KXT01611.1	NCBI
Terriglobus saanensis	ADV81530.1	NCBI
Armillaria gallica	PBK91973.1	NCBI
Epicoccum nigrum	OSS51998.1	NCBI
Hypsibius dujardini	BV898_00367.p01	(Yoshida et al., 2017)
Hypsibius dujardini	BV898_00365.p01	(Yoshida et al., 2017)
Hypsibius dujardini	BV898_00364.p01	(Yoshida et al., 2017)
Yangia sp.	WP_095883800.1	NCBI

Bajda, S., Dermauw, W., Greenhalgh, R., Nauen, R., Tirry, L., Clark, R.M., Van Leeuwen, T., 2015. Transcriptome profiling of a spirodiclofen susceptible and resistant strain of the European red mite Panonychus ulmi using strand-specific RNA-seq. BMC genomics 16, 974.

Yoshida, Y., Koutsovoulos, G., Laetsch, D.R., Stevens, L., Kumar, S., Horikawa, D.D., Ishino, K., Komine, S., Kunieda, T., Tomita, M., Blaxter, M., Arakawa, K., 2017. Comparative genomics of the tardigrades Hypsibius dujardini and Ramazzottius varieornatus. PLoS Biol 15, e2002266.

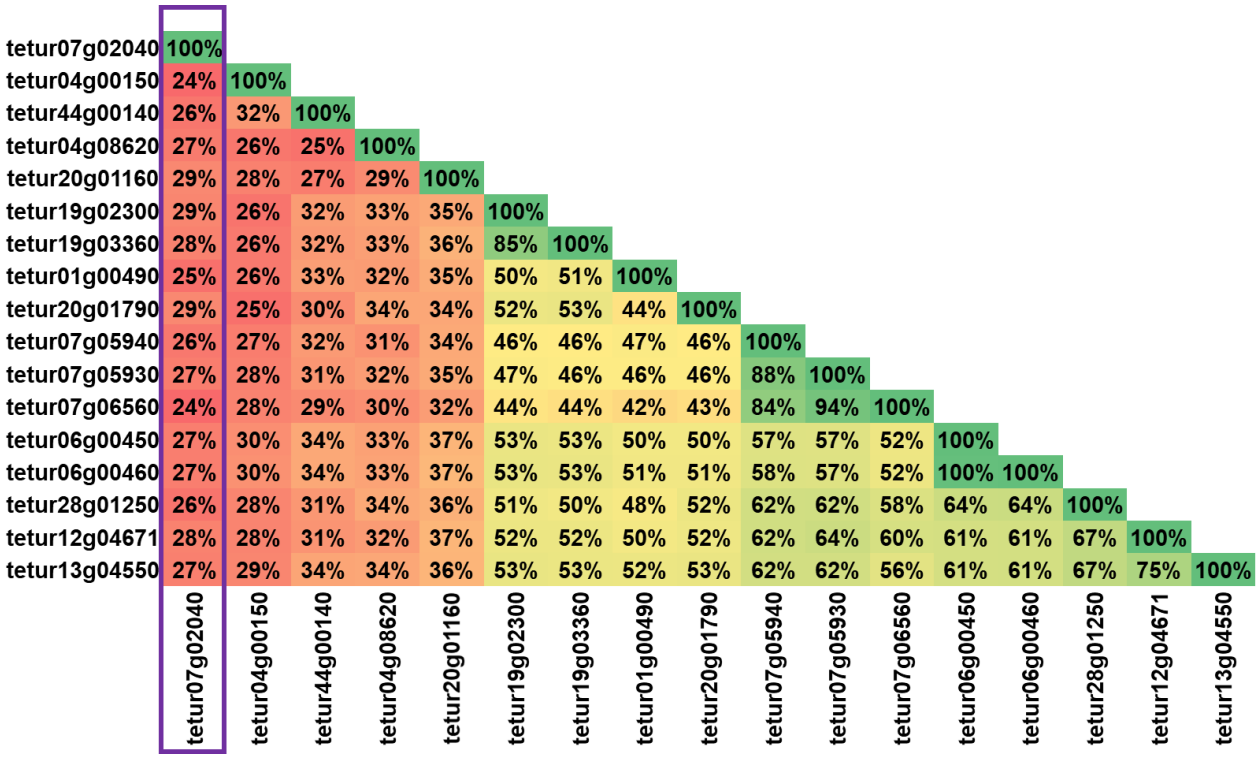


Figure S1. Comparison of sequences for the 17 TSSM intradiol dioxygenase-like proteins. Percentage corresponds to sequence identity. To highlight the differences between proteins the fields with sequence identities are color-coded. Red corresponds to the lowest values of sequence identity, green to the highest values, and yellow to intermediate values.

(t48)TuIDRCD

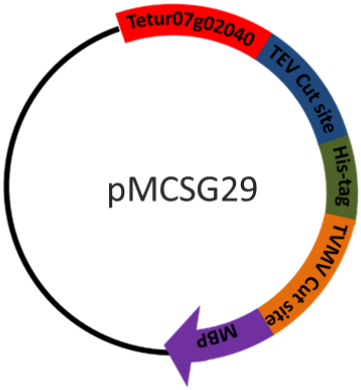
Residues 48-259 of Tetur07g02040

MHHHHHHSSGVDLGTENLYFQ/SGSGSFVTRFTECSLSPEVGEGPY
FIEEDIIRSNIVEDRIGIRLNVTLNLVDFNTCKPIKGAKVYIWQPDYSGI
YSGFMDKPRVKREKMYPKDPRRFLRGTQVTNENGTVTFETLFPGH
YPGRTPHIHYRIHANGNVAHIGQIFFDESTSQVIQSKSPYNQVHSRR
MKNEEDGEFTYFNGKKSIINIDPQSLSGDSLEGILNLAINPLHRSNL
MWA

Figure S2. Sequence of the (t48)TuIDRCD construct. Initiator methionine is shown in black, the His-tag in green, and the linker between the His-tag and the residues of *Tetur07g02040* (red) is marked in blue. The TEV-cleavage site is marked with a slash.

(t48)TuIDRCD-MBP

Residues 48-259 of Tetur07g02040

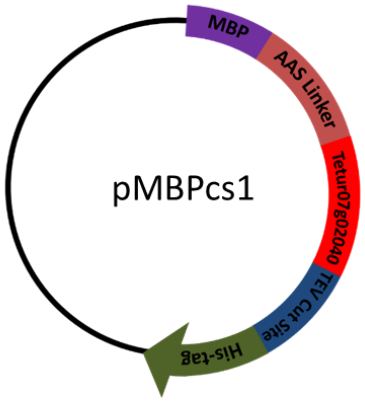


MSFVTRFTECSLSPEVGEGPYFIEEDIIRSNIVEDRIGIRLNVTLNLVD
FNTCKPIKGAKVYIWQPDYSGIYSGFMDKPRVKREKMYPKDPRRF
LRGTQVTNENGTVTFETLFPGHYPGRTPHIHYRIHANGNVAHIGQIF
FDESTSQVIQSKSPYNQVHSRRMKNEEDGEFTYFNGKKSIINIDPQS
LSGDSLEGILNLAINPLHRSNLMWAAGENLYFQ/SAGHHHHHHHHH
HETVRFQ/SKIEEGKLVIWINGDKGYNGLAEVGKKFEKDTGIKVTVE
HPDKLEEKFPQVAATGDGPDIIFWAHDRFGGYAQSGLLAEITPDKA
FQDKLYPFTWDAVRYNGKLIAYPIAVEALSLIYNKDLLPNPPKTWEEI
PALDKELKAKGKSALMFNLQEPYFTWPLIAADGGYAFKYENGKYD
IKDVGVDNAGAKAGLTFLVDLIKNKHMNADTDYSIAEAAFNKGETA
MTINGPWAWSNIDTSKVNYGVTVLPTFKGQPSKPFVGVLSAGINAA
SPNKELAKEFLENYLLTDEGLEAVNKDKPLGAVALKSYEEELAKDP
RIAATMENAQKGEIMPNIPQMSAFWYAVRTAVINAASGRQTVDEAL
KDAQTNS

Figure S3. Sequence of the (t48)TuIDRCD-MBP construct. Initiator methionine is shown in black, the His-tag in green, and the linker between the residues of *Tetur07g02040* (red) and His-tag are marked in blue. The TEV-cleavage site is marked with a slash. Residues marked in orange contain TVMV-cleavage site. The MBP sequence is marked in purple.

MBP-(t48)TuIDRCD

Residues 48-259 of Tetur07g02040



MKIEEGKLVIWINGDKGYNGLAEVGKKFEKDTGIKVTVEHPDKLEE
KFPQVAATGDGPDIIFWAHDRFGGYAQSGLLAEITPDKAFQDKLYP
FTWDAVRYNGKLIAYPIAVEALSLIYNKDLLPNPPKTWEEIPALDKE
LKAKGKSALMFNLQEPYFTWPLIAADGGYAFKYENGKYDIKDVGV
DNAGAKAGLTFLVDLIKNKHMNADTDYSIAEAAFNKGETAMTINGP
WAWSNIDTSKVNYGVTVLPTFKGQPSKPFVGVLSAGINAASPNKE
LAKEFLENYLLTDEGLEAVNKDKPLGAVALKSYEEELAKDPRIAAT
MENAQKGEIMPNIPQMSAFWYAVRTAVINAASGRQTVDEALKDAQ
TNSAASSFVTRFTECSLSPEVGEGPYFIEEDIIRSNIVEDRIGIRLNVT
LNLVDFNTCKPIKGAKVYIWQPDYSGIYSGFMDKPRVKREKMYPK
DPRRFLRGTQVTNENGTVTFETLFPGHYPGRTPHIHYRIHANGNVA
HIGQIFFDESTSQVIQSKSPYNQVHSRRMKNEEDGEFTYFNGKKSII
NIDPQSLSGDSLEGILNLAINPLHRSNLMWAAGENLYFQ/SAGHHH
HHH

Figure S4. Sequence of the MBP-(t48)TuIDRCD construct. Initiator methionine is shown in black. MBP sequence is marked in purple. The short linker between MBP and *Tetur07g02040* (red) is marked in dark orange. The His-tag is shown in green, and the linker between *Tetur07g02040* and the residues of and His-tag are marked in blue. The TEV-cleavage site is marked with a slash.

mTuIDRCD

Residues 23-259 of Tetur07g02040

MHHHHHHSSGVDLGTENLYFQ/SNSCASKEEVVPSPEEGQCSQES
VKQSFVTRFTECSLSPEVGEGPYFIEEDIIRSNIVEDRIGIRLNVTLNL
VDFNTCKPIKGAKVYIWQPDYSGIYSGFMDKPRVKREKMYPKDPR
RFLRGTQVTNENGTVTFETLFPGHYPGRTPHIHYRIHANGNVAHIGQ
IFFDESTSQVIQSKSPYNQVHSRRMKNEEDGEFTYFNGKKSIINIDP
QSLSGDSLEGILNLAINPLHRSNLMWA

Figure S5. Sequence of mTuIDRCD construct. Initiator methionine is shown in black. The His-tag is shown in green, and the linker between *Tetur07g02040* (red) and the residues of and His-tag are marked in blue. The TEV-cleavage site is marked with a slash.

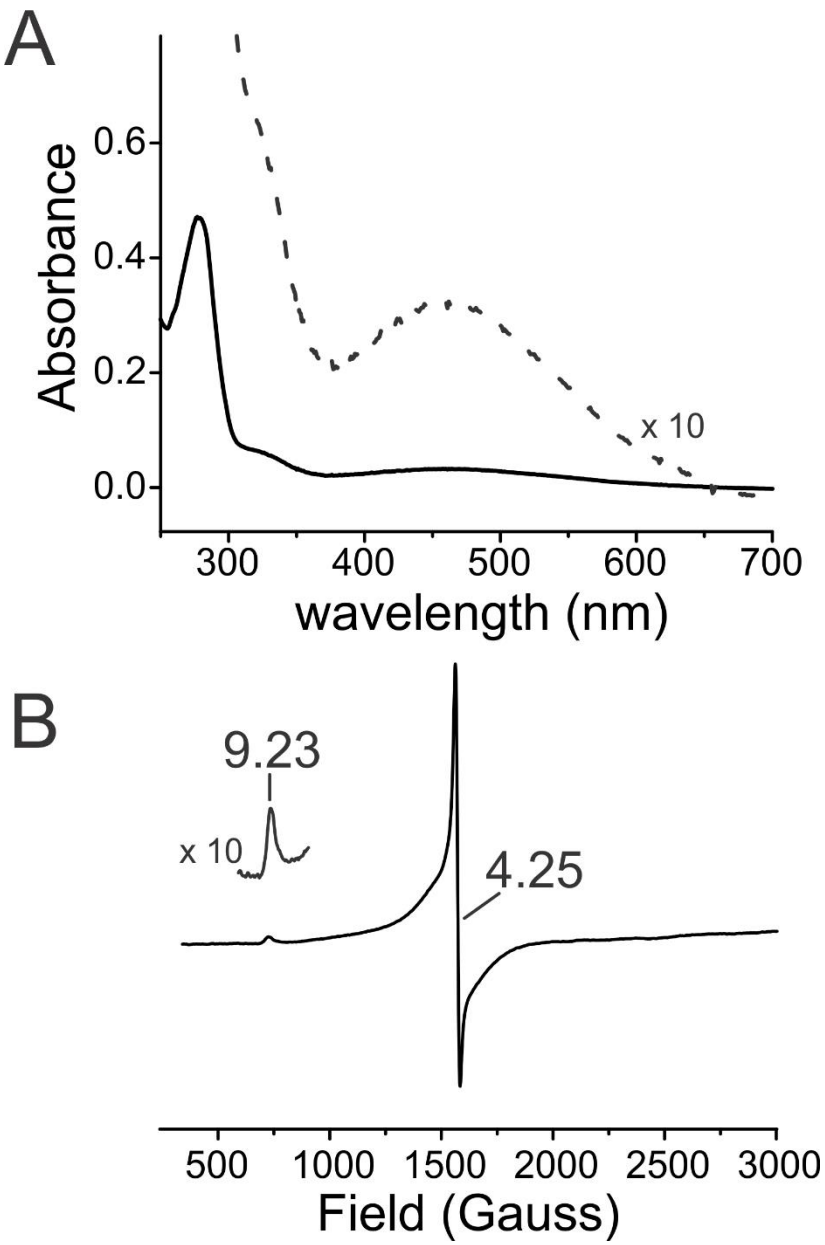


Figure S6. UV-visible (A) and electron paramagnetic resonance (B) spectra (B) of as purified MBP-(t48)TuIDRCD.

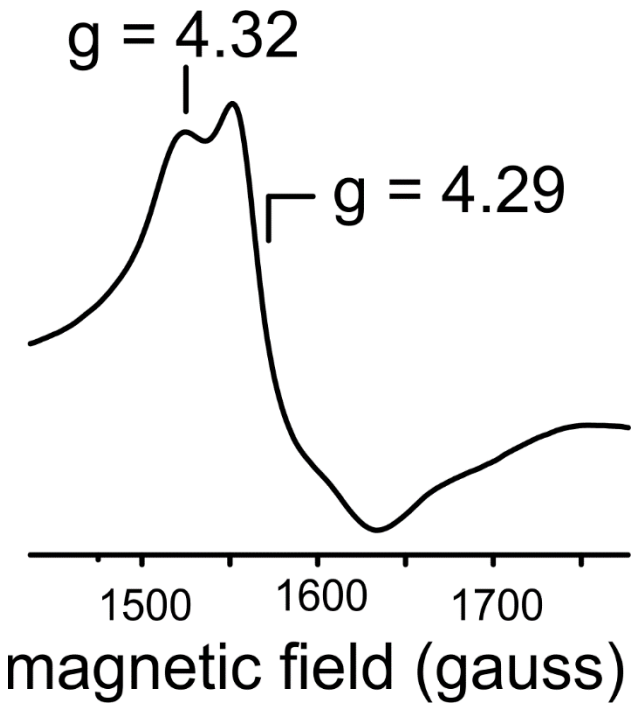


Figure S7. Electron paramagnetic resonance spectrum of MBP-(t48)TuIDRCD following reduction with dithionite and exposure to NO.