This is to certify that the

dissertation entitled

CHARACTERIZATION OF THE SUBSTRATE SPECIFICITY OF 2,4-DICHLOROPHENOXYACETIC ACID/ALPHA-KETO-GLUTARATE-DEPENDENT DIOXYGENASE

presented by

JULIE CHRISTINE DUNNING HOTOPP

has been accepted towards fulfillment
of the requirements for

Ph.D. degree in MICROBIOLOGY & MOLECULAR GENETICS

Robert P. Hausinger
Major professor

Date 04/12/2002

CHARACTERIZATION OF THE SUBSTRATE SPECIFICITY OF 2,4-DICHLOROPHENOXYACETIC ACID/α-KETOGLUTARATE-DEPENDENT DIOXYGENASE

By

Julie Christine Dunning Hotopp

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

Department of Microbiology and Molecular Genetics

2002

ABSTRACT

CHARACTERIZATION OF THE SUBSTRATE SPECIFICITY OF 2,4-DICHLOROPHENOXYACETIC ACID/α-KETOGLUTARATE-DEPENDENT DIOXYGENASE

By

Julie Christine Dunning Hotopp

2,4-Dichlorophenoxyacetic acid (2,4-D)/α-ketoglutarate (αKG) dioxygenase, TfdA, couples the oxidative decarboxylation of αKG to the oxidation of the herbicide 2,4-D using a mononuclear non-heme Fe(II) active site. To better define the substrate specificity of this enzyme, a variety of non-phenoxyacetate compounds were examined as potential substrates. 2-Naphthoxyacetic acid was the best alternative substrate tested, followed by benzofuran-2-carboxylic acid, 2,4-dichlorocinnamic acid, 2-chlorocinnamic acid, 1-naphthoxyacetic acid, and 4-chlorocinnamic acid. TfdA appeared to oxidize the olefin bond of benzofuran-2-carboxylate and the cinnamic acids to form the corresponding epoxides.

To facility further studies on substrate binding, a model of TfdA was constructed and fluorescence techniques to measure binding affinities were developed. A structural model of TfdA was determined based on the coordinates of the ~30 % identical TauD. The intrinsic fluorescence of TfdA was shown to be quenched by 50-85% upon addition of Fe(II) or αKG, allowing determination of binding affinities. Trp256 was initially suspected to be the reporter since it lies within 5 Å of the metal and αKG binding sites; however, W256L and W256F variants had only negligible differences from wild-type TfdA in terms of relative fluorescence intensity and ligand-based fluorescence

quenching. Because Trp195 was predicted to be quite distant (~15 Å) from the active site, some combination of Trp113 and Trp248 is likely to serve as the reporter that senses metal and cofactor binding to TfdA.

Based on the TfdA structural model, potential 2,4-D and αKG ligands were identified and altered by site-directed mutagenesis. His214, Lys71, Arg278, and backbone amide of Ser117 were hypothesized to bind the 2,4-D carboxylate, and Lys95, and possibly Lys71, were hypothesized to interact with the 2,4-D ether atom. Additionally, Arg274 and Thr141 were suspected to be αKG ligands. Consistent with their important roles as substrate ligands, variants of Lys71, Arg274, and Arg278 had αKG K_m and K_d values and a 2,4-D K_m value that were increased 1000-, 1000-, and 100-fold, respectively. Evidence supporting an interaction between Lys95 and the 2,4-D ether atom included the K95L variant being significantly less prone to inactivation by phenylpropiolic acid and having a 4-fold lowered 2,4-dichlorocinnamic acid K_m. Thr141 plays only a minor role, if any, in interacting with αKG; the T141V had αKG K_m and K_d values and a 2,4-D K_m value similar to wild-type protein.

Studies were also conducted in an attempt to examine the substrate specificity of TfdA-like proteins in environmental and pathogenic organisms. The *tfdA*-like gene in *Bordetella pertussis* was cloned, but soluble, properly folded protein could not be obtained to analyze substrate specificity. Environmental *tfdA*-like genes proved difficult to clone due to loss of the gene in these organisms over time. During the course of these studies, the ubiquity of *tfdA*-like genes was extended to aerobic environmental isolates from soils frozen in the Siberian permafrost for upwards of 10,000 years.

ACKNOWLEDGMENTS

First, I would like to thank Bob Hausinger for being a wonderful mentor and scientific guide. I appreciate his dedication to his graduate students, his ever open door, his encouragement, his guidance, and his confidence in me as a graduate student. I also want to thank my committee members for their advice and support: John Breznak, Leslie Kuhn, Pat Oriel, and Greg Zeikus. I owe special thanks to Leslie Kuhn for her assistance in modeling the structure of TfdA. I also owe a great deal of thanks to past and present members of the lab: Scott Mulrooney and Jason Kuchar for their scientific expertise and comic relief, Deb Hogan for being my graduate student guide, Aileen Soriano for being my pipetteman buddy, Gerry Colpas for always being ready to answer my questions, Kristin Huizinga for doing many genomic preps during her rotation, Raghavakaimal Padmakumar for getting me started in the lab, and Ruth Saari, Matt Ryle, and Tim Henshaw for scientific discussions.

I am so greatly indebted to Kerry Hotopp that I am at a loss for words. He is my life guide and has been of immense support and assistance in every aspect of life. For three years, he did this all from three hundred miles away, a testament of his commitment to us. Additionally, he helped in the writing of this thesis with patent research and proofreading. I would also like to thank Sandi Clemens, Stephanie Eichorst, Jennifer Gray, Kristin Huizinga, Monica Ponder, and Kristina Stredwick for all their support and advice about science and life. I would also like to give honorable mention to EF and

BtVS. And lastly, but not least, I want to thank my parents and sister who gave me much love and support and taught me the value of education. They have always appreciated and encouraged me to be obstinate, to persevere, and to always reach for my dreams.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF SCHEMES

LIST OF EQUATIONS

LIST OF SYMBOLS AND ABBREVIATIONS

2,4,5-T	2,4,5-Trichlorophenoxyacetic acid
2,4-D	2,4-Dichlorophenoxyacetic acid
ACCO	1-Amino-1-cyclopropanecarboxylic acid oxidase
αKG	α-Ketoglutarate
AtsK	Alkylsulfatase
CarC	Carbapenem biosynthesis enzyme
CAS	Clavaminate synthase
DAOCS	Deacetoxycephalosporin C synthase
EDTA	Ethylenediaminetetraacetic acid
EpoA	Fosfomycin biosynthesis enzyme
$[E_T]$	Total concentration of enzyme subunit
HPLC	High pressure liquid chromatography
IPNS	Isopenicillin N synthase
ΔF	Change in fluorescence
ΔF_{max}	Maximal fluorescence change
[I]	Inactivator concentration
k(inact)	Inactivation rate
$k(\text{inact})_{max}$	Inactivation rate at saturating [I]
K_i	Inactivation/Inhibition constant
[L]	Total concentration of ligand
LC-ESI	Liquid chromatography electrospray ionization
LTER	Long term ecological research
MALDI	Matrix-assisted laser desorption ionization
ORF	Open reading frame
PPA	Phenylpropiolic acid
P_t	Product accumulated at time t
PvcB	Pyoverdine biosynthesis enzyme
SDS-PAGE	Sodium dodecylsulfate polyacrylamide gel electrophoresis
SyrB	Syringomycin biosynthesis enzyme 1
TauD	Taurine/αKG dioxygenase
TE	10 mM Tris (pH 7.7) and 1 mM EDTA
TFA	Trifluoroacetic acid
TfdA	2,4-Dichlorophenoxyacetic acid/αKG dioxygenase
Tris	Tris(hydroxymethyl)aminomethane
V_i	Initial velocity
YSD	Sulfonate/αKG dioxygenase

CHAPTER 1

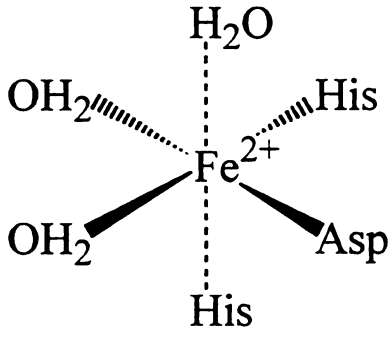
INTRODUCTION

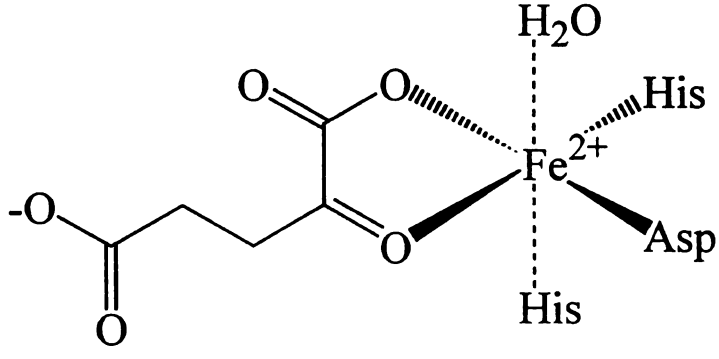
In order to provide a framework for describing my studies of various family members, this introductory chapter begins with a general description of the α-ketoglutarate (αKG) dioxygenase superfamily. Most of my efforts have focused on 2,4-dichlorophenoxyacetate/α-KG dioxygenase (TfdA), an enzyme involved in metabolism of the herbicide 2,4-dichlorophenoxyacetic acid (2,4-D). I will introduce background relevant to this herbicide, comment on its degradative pathway, and emphasize the role of TfdA. I then summarize what is known about close relatives of TfdA in environmental isolates and pathogenic bacteria. Finally, I emphasize the open questions that led to this work.

αKG DIOXYGENASE SUPERFAMILY

αKG dioxygenases are an important and understudied superfamily of enzymes. Members of the superfamily are widely distributed among eukaryotes and eubacteria and catalyze a diverse array of reactions (78). All members of the αKG dioxygenase superfamily chelate one ferrous ion per catalytic site in a 2-His-1-carboxylate facial triad (42) with three waters bound at the other adjacent coordination sites (Figure 1.1). For most family members, αKG displaces two waters to coordinate the iron via its C-1 carboxylate and C-2 keto group (Figure 1.1). In contrast, 1-amino-1-cyclopropanecarboxylic acid oxidase (ACCO) and isopenicillin N synthase (IPNS) do not require an α-ketoacid, but are placed in the superfamily based on sequence similarity (77, 80). When used as a cosubstrate, αKG is oxidatively decarboxylated in a reaction that is coupled to the oxidation of the primary substrate. The most common oxidation of the primary substrate is the hydroxylation of an unactivated carbon, but desaturations,

Fig. 1.1. 2-His-1-carboxylate facial triad. All members of the αKG dioxygenase superfamily chelate iron in a 2-His-1-carboxylate facial triad (top). For most family members, αKG binds the iron in a bidentate fashion by displacement of two water molecules (bottom).





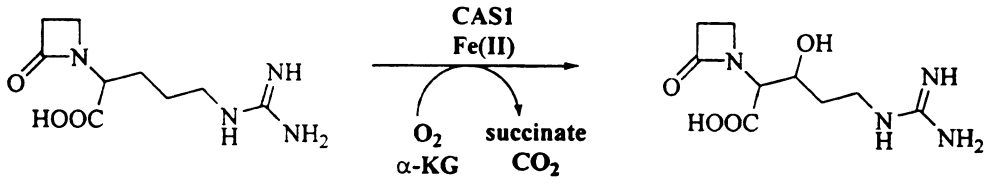
epoxidations, or ring formation and expansion reactions can also result (Scheme 1.1). The C-5 carboxylate of αKG forms a salt bridge with a basic side chain, either lysine or arginine, and in some cases is further stabilized by hydrogen bonding with a threonine or serine. In combination, the iron and αKG ligands form an $HX(D/E)X_nHX_{(6-12)}(R/K)$ conserved sequence motif in these proteins. Crystal structures of IPNS (80), clavaminate synthase 1 (CAS)1 (101), deacetoxycephalosporin C synthase (DAOCS) (109), proline hydroxylase (16), anthocyanidin synthase (106), and taurine/αKG dioxygenase (TauD) (25) reveal that the conserved residues are located on eight β-strands in a conserved jellyroll motif (Figure 1.2).

Historically, αKG dioxygenases were slow to be identified since αKG is not routinely added to assay mixtures and the enzymes show little overall sequence identity. Recently, however, there has been a rapid expansion of potential family members identified through genome sequencing projects based on homology to characterized αKG dioxygenases. For example, analysis of the genome of *Arabidopsis thaliana* reveals over sixty-four open reading frames that show strong similarity to known αKG dioxygenases; less than 20% have known functions (78). Other proteins of known function, but unknown biochemical mechanism, have been identified as αKG dioxygenases by using new search engines available to detect proteins with similar conserved motifs in sequences with little overall homology (e.g., PSI-BLAST). PSI-BLAST searches to convergence followed by alignments based on predicted secondary structure recently revealed that the extracellular matrix protein leprecan, the disease-resistance-related protein EGL-9, and the DNA-repair protein AlkB are members of the αKG dioxygenase

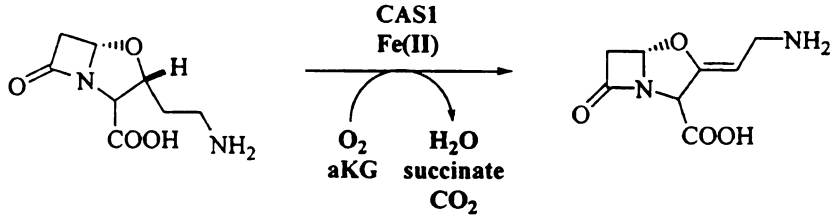
Scheme 1.1. Examples of the types of reactions carried out by αKG dioxygenases. Both hydroxylation and desaturation reactions are carried out by clavaminate synthase 1 (CAS) (9, 63). An epoxidase reaction leads to the formation of fosfomycin in a reaction carried out by EpoA (104). Ring formation and ring expansion reactions are catalyzed by IPNS (14) and deacetoxycephalosporin C synthase (DAOCS) (10, 11).

Scheme 1.1.

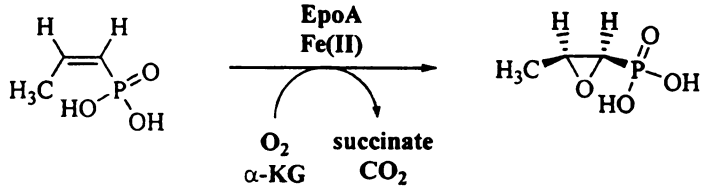
Hydroxylation



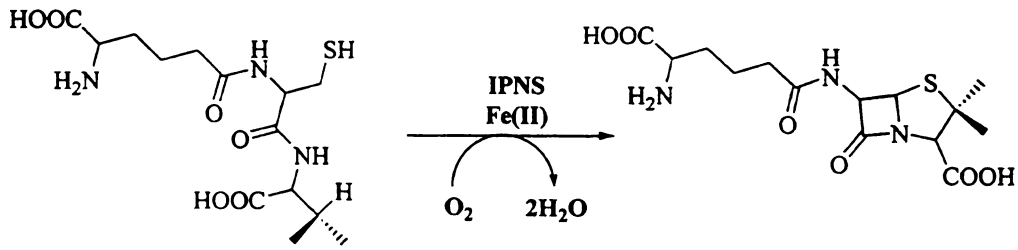
Desaturation



Epoxidation



Ring Formation



Ring Expansion

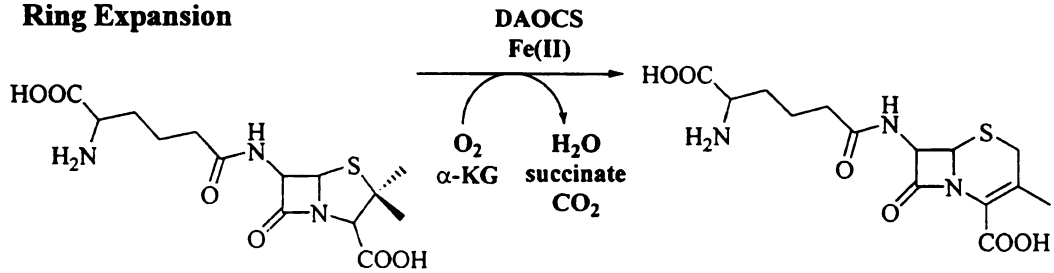
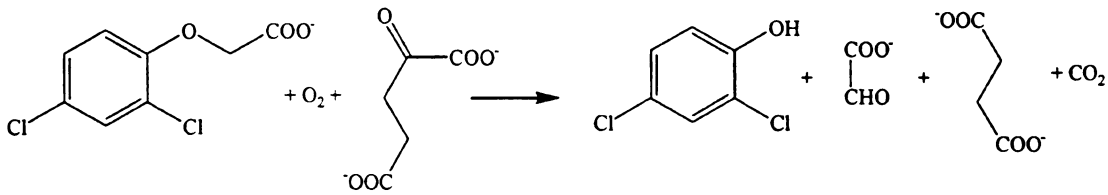


Figure 1.2. Jellyroll architecture. Conserved residues involved in binding ferrous ion and αKG are located on eight β-strands in a conserved jellyroll motif. Images in this thesis/dissertation are presented in color.

superfamily (5). These approaches set the stage for biochemical analyses to define the catalytic activity of these proteins. In a separate study, PSI-BLAST was used to divide the αKG dioxygenase superfamily into three groups each with their own more specific conserved motif. The major difference between the groups is the size of the insert between the carboxylic acid containing amino acid and the second histidine (n = 50-70, Group I; n = 138-207, Group II; n = 72-101, Group III) (47). Work in this thesis focuses on Group II members related to *Ralstonia eutropha* JMP134 (pJP4) TfdA, *Escherichia coli* TauD, *Streptomyces clavuligerus* CAS1 and CAS2, *Pseudomonas* sp. γ-butyrobetaine hydroxylase, *Caenorhabditis elegans* γ-butyrobetaine hydroxylase, *Pseudomonas putida* alkylsulfatase (AtsK), *Saccharomyces cerevisiae* sulfonate/αKG dioxygenase (YSD), *Erwinia carotovora* carbapenem biosynthesis enzyme C (CarC), *Pseudomonas aeruginosa* pyoverdine biosynthesis enzyme (PvcB), and *Pseudomonas syringae* syringomycin biosynthesis enzyme 1 (SyrB).

TfdA AND 2,4-D DEGRADATION

The *Ralstonia eutropha* JMP134 (pJP4) 2,4-D/αKG dioxygenase (TfdA) is a homodimer with a calculated monomer molecular weight of 32 KDa that hydroxylates 2,4-D while converting αKG to succinate and carbon dioxide (29).



Hydroxylation of 2,4-D at the C-2 side chain position yields an unstable hemiacetal that decomposes into glyoxylate and 2,4-dichlorophenol, which is further metabolized by the cell (29, 86). Although TfdA prefers αKG and 2,4-D, it can use a wide variety of α-

ketoacid and phenoxyacetic acid substrates (Table 1.1 and 1.2). The key residues in TfdA associated with the motif shown above are His114, Asp116, Thr141, His263, and Arg278. Site-directed mutagenesis studies confirmed His114, His263 and Asp116 as iron ligands (47). Additionally, evidence was obtained that His214 participates in substrate binding (47). Below, I discuss the history of 2,4-D, the pathway for its degradation, and homologues of this protein or its gene.

The History of 2,4-D. In 1880, Charles Darwin hypothesized that the tip of a plant influences the bending of the plant toward light, or phototropism (4). Research continued on phototropism through the early 1900's, and in 1934, Kögl and Haagen-Smit from the Netherlands discovered indoleacetic acid (auxin) to be the principle natural plant growth hormone (4). Indoleacetic acid was unstable outside of plants, so researchers began synthesizing indoleacetic acid homologues and studying their effects on plant growth activity (4). In 1940, 2,4-D was first synthesized from monochloroacetic acid and 2,4-dichlorophenol and was tested as a fungicide (4). 2,4-D was reported in 1942 to be 300 times more potent than indolebutyric acid, the major plant growth regulator at that time (4). The original 2,4-D patent was filed in February 1942 and defined plant growth hormones as compounds "containing as an essential active ingredient a monocarboxylic acid having one of its carbon atoms linked to a nuclear halogenated aromatic ring by means of a polyvalent, strongly negative, non-metallic atom, namely, oxygen, sulfur and nitrogen" (64). The patent did not mention the usage of the compounds as herbicides (64). Originally, the work was not focused on the idea of using synthetic plant growth hormones as herbicides, but instead on using them at low concentration as rooting agents (79).

Table 1.1. Phenoxyacetic acid substrate specificity (30).

Substrate	K_m (μM)	k_{cat} (min^{-1})	k_{cat}/K_m ($min^{-1}·mM^{-1}$)
Phenoxyacetate	460 ± 23	443 ± 10	960
2-Chlorophenoxyacetate	110 ± 6.5	380 ± 11	3,450
4-Chlorophenoxyacetate	117 ±6.2	595 ± 15	5,090
2,3-Dichlorophenoxyacetate	102 ± 8.5	288 ± 12	2,820
2,4-D	17.5 ± 1.0	529 ± 16	30,200
3,4-Dichlorophenoxyacetate	219 ± 7.6	307 ± 5	1,400
2,4,5-Trichlorophenoxyacetate	59.6 ± 3.4	96 ± 3	1,610
4-Chloro-2-methylphenoxyacetate	89.0 ± 8.7	233 ± 12	2,620
DL-2-Phenoxypropionate	1170 ± 120	5.1 ± 0.2	4.4
2-(2,4-Dichlorophenoxy)propionate	191 ± 6.7	61 ± 3	320
3-Phenoxypropionate	12900 ± 3200	3.2 ± 0.4	0.25

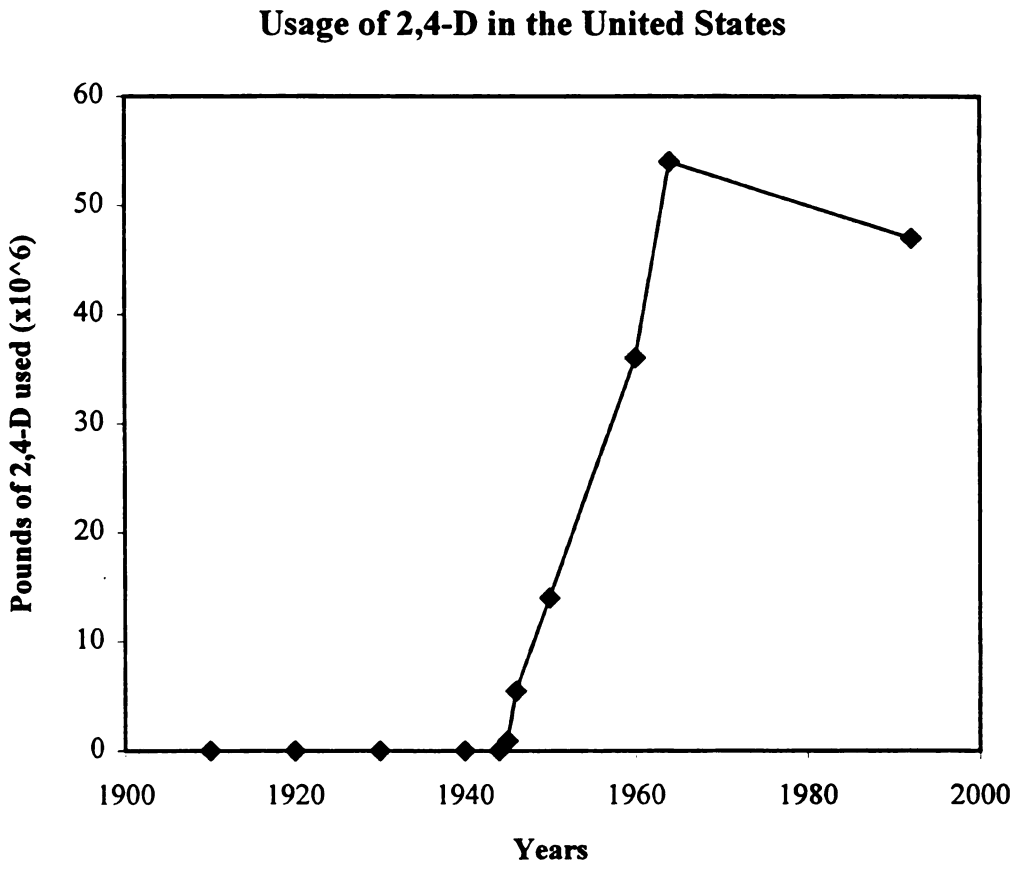
Table 1.2. α-Ketoacid substrate specificity (30).

Substrate	K_m (μM)	k_{cat} (min^{-1})	k_{cat}/K_m ($min^{-1}·mM^{-1}$)
αKG	3.20 ± 0.54	643 ± 44	200,000
α-Ketoadipate	20.6 ± 1.1	290 ± 5	14,100
Pyruvate	1020 ± 86	58 ± 2	60
α-Ketobutyrate	464 ± 61	89 ± 6	190
α-Ketovalerate	607 ± 47	404 ±14	660
α-Ketocaproate	583 ± 50	158 ± 7	270
α-Ketoisovalerate	745 ± 36	16 ± 1	20

During this time, researchers were also trying to identify selective herbicides for broadleaf weeds in grasses and grains (4). In 1941, after observing that certain plant growth hormones were phytotoxic, Kraus proposed that synthetic plant growth hormones might work as herbicides (4). In 1942, after Kraus and other scientists convinced the Secretary of War, H.L. Stimson, of the potential dangers of these selective herbicides when used as weapons, the United States Army set up Camp Detrick for the purpose of studying chemicals for biological warfare (4). The scientists at Camp Detrick pursued using herbicides to destroy crops and as forest defoliants to reduce cover of World War II Japanese defense positions (4). The research included the evaluation of over one thousand compounds, cytotoxicity effects of specific compounds including 2,4-D, demonstration of the control of weeds, demonstration of the specificity of 2,4-D to kill broad-leaf plants, and identification of the effects of applications of herbicides to soil and water. This research was published by the scientists at the end of World War II (4). Herbicides were not used in chemical or biological warfare until the introduction of Agent Orange (a mixture of 2,4,5-triclorophenoxyacetic acid (2,4,5-T) and 2,4-D) as a defoliant is the Vietnam War.

The patent for the usage of 2,4-D and other plant growth hormones as herbicides was assigned to American Chemical Paint Company in May 1945 (49). In June 1945, American Chemical Paint Company began marketing 2,4-D under the brand name Weedone, and by 1946, 2,4-D was in high demand. It continues to be used on crops today (Figure 1.3) under the brand names Hi-Dep®, Weedar® 64, Weed RHAP A-4D®. At low concentrations, 2,4-D regulates plant cell development leading to bending and swelling, proliferation and overgrowth of leaves and stems, control of root growth, and

Figure 1.3. Pounds of 2,4-D used in the twentieth century.



the initiation of new roots on stems and leaves (64). At higher concentrations, 2,4-D acts as a selective, post-emergence herbicide against broad-leaf plants (49). Because of these properties, 2,4-D is used today on major field crops, rangeland, pastureland, alfalfa for forage, turf grass, small grain production, weed control, and orchard, vineyard, soft fruit, and nut crops (4). Approved uses of 2,4-D have not been scientifically documented to cause any acute or chronic human health risks (4). In fact, a 2,4-D researcher revealed having eaten 0.5 g of 2,4-D per day for three weeks with no ill effects (4). Due to its many uses and low toxicity in humans and animals, 2,4-D is the most widely used herbicide in the world today and the third most widely used in the United States (4).

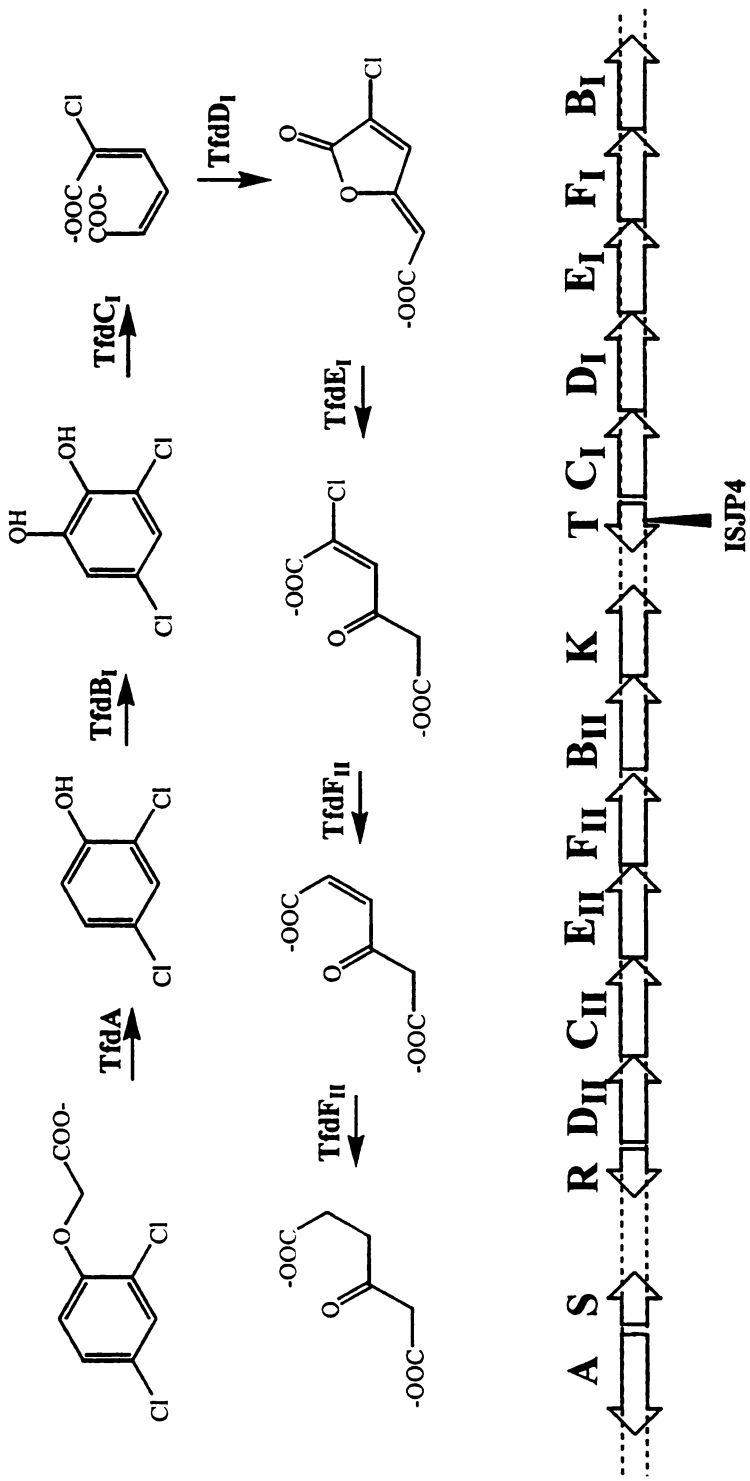
Degradation of 2,4-D. Early reports suggested that 2,4-D was rapidly decomposed in soil (7). By the 1960's organisms had been isolated that degrade 2,4-D by a pathway involving 2,4-dichlorophenol (12, 92). Since then organisms from numerous genera have been found to degrade 2,4-D including *Acinetobacter, Arthrobacter, Aspergillus, Bordetella, Bradyrhizobia, Burkholderia, Corynebacterium, Pseudomonas, Ralstonia, Rhodoferax, Stenotrophomonas, Sphingomonas, Variovorax,* and *Xanthobacter* (20, 21, 26, 27, 38, 50-52, 55, 58, 65, 67, 96, 107).

The individual genes involved in 2,4-D metabolism (the *tfd* genes) and the plasmids that typically carry them are often mobile. For example, the 2,4-D catabolic genes can be transferred in the environment between different genera of soil bacteria (31, 66, 88, 99). Transposons carrying the *tfd* genes have been identified (13). Some plasmids carrying the *tfd* genes can be transferred by conjugation (19, 72) and can be integrated and excised from the chromosome (53, 96).

Numerous pathways exist for 2,4-D degradation ranging from various oxidative schemes to completely anaerobic metabolism (40). The best-studied pathway is the one encoded by the *tfd* genes on the pJP4 plasmid from *R. eutropha* JMP134 (74). The pJP4 plasmid is a 51 MDa, Inc P1 plasmid. It carries all the genes necessary for converting 2,4-D to chloromaleylacetate (Figure 1.4). Enzymes from chromosomally encoded genes carry out the subsequent breakdown of chloromaleylacetate. The plasmid pJP4 also confers resistance to merbromin, phenylmercury acetate, and mercuric ions (21), and it confers the ability to degrade 3-chlorobenzoate (37), but not phenoxyacetate (21). Since its discovery and characterization, the pJP4-encoded catabolic system has been used as a model for the evolution and dispersal in the environment of catabolic pathways to degrade xenobiotics.

Three lines of evidence suggest that separate sets of the individual *tfd* genes were recruited onto the plasmid backbone to generate a single plasmid capable of conferring to the organism the ability to degrade 2,4-D. First, *tfdA* and the *tfdBCDEF* gene cluster are individually regulated on pJP4 (61) (Figure 1.4). Second, a comparison of 2,4-D-degradative plasmids containing homologous *tfd* genes shows that the genes are uniquely organized on unrelated plasmids (66). Finally, phylogenetic trees of the 16S rDNA, *tfdA, tfdB,* and *tfdC* produce discrepancies (71, 102). In other words, the *tfd* genes have not evolved together in one organism or on one plasmid backbone. Additionally, *tfdBCDE* operons may have predated the application of 2,4-D, since 2,4-dichlorophenol is produced in the environment from phenol by the metabolic action of *Penicillium* sp. (39). Large-scale use of 2,4-D may have increased the prevalence of these genes rather than selecting for a novel function.

Figure 1.4. Pathway for the degradation of 2,4-D and organization of the *tfd* genes in *Ralstonia eutropha* JMP134 (pJP4).



Homologues of all the genes except *tfdA* have been identified that encode enzymes with substrates similar to those in the pJP4-encoded pathway providing hints of their possible recruitment origin. The *tfdCDE* genes (encoding for dichlorocatechol degradation) are closely related to the *clcABD* genes encoding for chlorocatechol degradation (35, 36). Likewise, *tfdB* (encoding for 2,4-dichlorophenol hydroxylase) is homologous to a similar gene, *pheA*, which encodes a phenol hydroxylase (73).

Although no homologue of *tfdA* has been found to participate in decomposition of non-chlorinated aromatic substrates, some evidence suggests that one may exist. Bacterial isolates from pristine areas throughout the world have been shown to degrade 2,4-D, albeit at a much slower rate than an organism carrying the canonical pJP4 plasmid (32). One particular isolate from a pristine area in northern Saskatchewan, Canada, carries a *tfdA* gene homologous to the one on pJP4 that can insert itself onto the pJP4 backbone (55). Furthermore, bacterial isolates that possess the *tfdA* gene, but do not degrade 2,4-D, have been isolated from agricultural plots that were not treated with 2,4-D (46). Lastly, plasmid capture experiments with a recipient plasmid containing *tfdBCDEF*, but not *tfdA*, led to the identification of a plasmid containing only *tfdA* (100).

Prevalence of TfdA homologues. Apparent homologues of *tfdA* are present in approximately 30% of bacterial isolates obtained from agricultural plots. The role of such *tfdA*-like genes is unknown, as these soil microorganisms do not degrade 2,4-D (46). Similar sequences have been identified in several microorganisms including *Mycobacterium tuberculosis* and *Bordetella pertussis*. These pathogenic bacteria would

not be expected to have been exposed to 2,4-D; thus, the *tfdA*-like genes in these microbes likely function in a distinct role.

CRITICAL OPEN QUESTIONS AND SPECIFIC AIMS

The overriding question addressed by this dissertation is: what governs substrate specificity in TfdA and related enzymes? In order to answer this question, experiments were undertaken to examine the breadth of compounds that the pJP4-encoded TfdA can use as a substrate (Chapter 2), to model the structure of TfdA and develop fluorescence methods to assess metal and αKG binding (Chapter 3), and to test the model by site-directed mutagenesis and analysis of substrate-based inactivation of TfdA by phenylpropiolic acid (Chapter 4). In addition, attempts were made to compare the substrate specificity of TfdA to the substrate specificity of environmental or pathogen TfdA-like proteins mentioned above. This involved characterizing environmental isolates containing *tfdA*-like genes (Chapter 5), cloning and performing a preliminary characterization of the protein encoded by a *Bordetella pertussis tfdA*-like gene (Chapter 6), and developing methods to evolve a *tfdA*-like gene into one encoding TfdA activity (Appendix A).

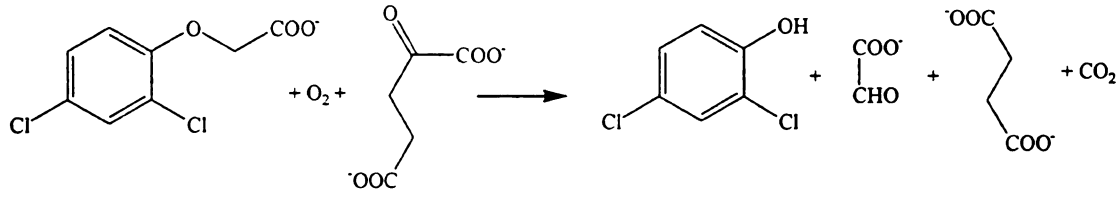
CHAPTER 2

ALTERNATIVE SUBSTRATES OF 2,4-DICHLOROPHENOXYACETIC ACID/α-KETOGLUTARATE DIOXYGENASE

This chapter is based on the published article of the same title: Dunning Hotopp, J.C. and R.P. Hausinger (2001) "Alternative substrates of 2,4-dichlorophenoxyacetate/α-ketoglutarate dioxygenase." J. Molec. Catalysis B: Enzymatic 15:155-62.

2,4-Dichlorophenoxyacetic acid (2,4-D), a broad leaf herbicide, is mineralized by a wide variety of environmental isolates (40) including the best-studied example, *Ralstonia eutropha* JMP134 (pJP4) (formerly *Alcaligenes eutrophus*). The *tfdABCDEF* genes, carried on the pJP4 plasmid in this microorganism (21), encode all of the enzymes necessary for the degradation of 2,4-D to chloromaleylacetate. Chromosomally borne genes encode enzymes that carry out the subsequent metabolism of chloromaleylacetate. Since its discovery and characterization, pJP4 has been used to study the evolution and environmental dispersal of the genes encoding a catabolic pathway for xenobiotic degradation (40).

TfdA catalyzes the first step in this 2,4-D degradation pathway. This ferrous ion and α-ketoglutarate (αKG) dependent dioxygenase hydroxylates 2,4-D while converting αKG to succinate plus CO_2 (29).



Hydroxylation of 2,4-D at the C-2 position yields an unstable hemiacetal that decomposes into 2,4-dichlorophenol and glyoxylate (29). TfdA is related in sequence (~30% amino acid identity) to a group of sulfonate degrading αKG dioxygenases (24, 45) and an alkyl sulfate ester degrading αKG dioxygenase (54). In addition to these protein homologues, nucleotide sequences closely homologous to *tfdA* are present in approximately 30% of bacterial isolates obtained from agricultural plots. The role of the *tfdA*-like genes containing these sequences is unknown since these soil microorganisms do not degrade 2,4-D (46).

The current study explores alternative substrates for TfdA in order to better define the substrate profile and to explore potential biotechnological uses for the enzyme. In particular, we demonstrate that purified enzyme or whole cells can be used to produce synthetically valuable epoxides (17, 34) of chlorinated cinnamic acids and benzofuran-2-carboxylic acid. Furthermore, since the 2,4-D-degrading enzyme may retain residual activities for the natural substrate, this work may provide clues about the role of TfdA-like proteins in nature.

EXPERIMENTAL

Purification and assay of TfdA. Escherichia coli DH5α (pUS311) carries *tfdA* on a pUC19-derived plasmid (29). These cells (4 L) were grown at 30 °C in LB medium for 16-20 hr, harvested by centrifugation, and suspended in 30 ml of TE buffer (10 mM Tris, 1 mM EDTA, pH 7.7) containing 1 mM phenylmethylsulfonyl fluoride and 10 μg/mL leupeptin. Crude cell extracts were obtained by passing the cells twice through a pre-cooled French pressure cell (16,000 psi) and clarifying the debris by centrifugation (100,000 x *g* for 45 min) at 4 °C. After splitting the sample into two pools, TfdA was enriched by chromatography at 4 °C on a column of DEAE-Sepharose (2.5 x 19 cm) using TE buffer and a 400 mL linear gradient to 200 mM NaCl. The enzyme eluted at approximately 100 mM NaCl. TfdA-containing fractions were dialyzed in TE buffer and chromatographed on a Mono Q column (HR10/10) (Pharmacia) at room temperature. The enzyme eluted at about 40 mM NaCl when using a 100 mL linear gradient to 200 mM NaCl in TE buffer. When necessary, TfdA was further purified by phenyl-Sepharose chromatography. The sample was adjusted to 1 M in ammonium sulfate, applied to a column (HR10/10) equilibrated with TE buffer containing the same concentration of this salt, and chromatographed with a 100 mL linear gradient from 1 M to 0 M ammonium sulfate. The enzyme eluted from this resin at approximately 350 mM salt. TfdA was routinely assayed with 2,4-D as the substrate by using the previously described 4-aminoantipyrene spectrophotometric assay (30). Protein concentrations were determined by using a commercial protein assay (BioRad) with bovine serum albumin as a standard.

Kinetics of non-phenoxyacetic acid substrates. The activity of TfdA toward non-phenoxyacetic acid potential substrates was determined at 30 °C in 10 mM imidazole

buffer (pH 6.8) containing 1 mM αKG, 50 μM $(NH_4)_2Fe(SO_4)_2$, and 200 μM ascorbic acid by using one or more of four different assays. For all methods, stock solutions of $(NH_4)_2Fe(SO_4)_2$ and ascorbic acid were made fresh prior to each set of experiments.

A YSI model 5300 biological oxygen monitor was used to determine rates of oxygen consumption. The probe was equilibrated using air-saturated MilliQ water, and control runs included samples lacking in ascorbic acid, iron, enzyme, αKG, and substrate.

Samples analyzed by HPLC (Hewlett Packard 1050) were resolved by using a Merck Lichrosorb RP-18 column. The peak areas or peak heights for samples absorbing at 230, 254, and/or 280 nm were determined by using the Hewlett Packard ChemStation software. Reaction mixtures containing 2,4-dichlorocinnamic acid were quenched with NaOH (0.1 M final concentration), neutralized with HCl (0.1 N final concentration), and analyzed by utilizing a mixture of 65:35:0.1 methanol:water:phosphoric acid (effluent A). A standard curve for the reaction product was prepared by complete conversion of selected concentrations of substrate. The area under the product peak was assumed to correspond to the original concentration of substrate. Reaction mixtures that contained naphthoxyacetic acid were quenched with EDTA (5 mM final concentration) and analyzed by using a 50:50:0.1 mixture of methanol:water:phosphoric acid (effluent B) as the mobile phase. Standard curves for the degradation of 1-naphthoxyacetic acid and 2-naphthoxyacetic acid were prepared by using 1-naphthol and 2-naphthol, respectively. All other potential substrates tested were quenched with NaOH (0.1 M final concentration) and neutralized with HCl (0.1 N final concentration). Standard curves for the reaction products were created by correlating the loss of substrates with gain of products at various concentrations of substrate. Effluent B was used as the mobile phase,

and the kinetics were determined in all cases by analyzing the gain of product; except, the 4-chlorocinnamic acid reaction kinetics were estimated from the loss of substrate.

Two spectrophotometric assays also were utilized. Analysis of 1-naphthoxyacetic acid was carried out by using the 4-aminoantipyrene assay with 1-naphthol as the standard. For 2,4-dichlorocinnamic acid, an alternative spectrophotometric assay made use of the 250 nm absorption associated with the conjugated system (ε_{250} = 7800 $M^{-1}cm^{-1}$). The loss of absorbance at this wavelength was monitored in studies designed to assess the ascorbic acid dependence of the reaction.

As previously documented (30, 85), TfdA activity decreased over time by both irreversible and ascorbate-reversible inactivation. To analyze the kinetics of non-phenoxyacetic acid substrates, progress curves in the presence of 200 μM ascorbate were analyzed by fitting the data to the following equation:

$$P_t = V_i (1-e^{-k(inact)t})k(inact)^{-1} \qquad \text{Eq. 2.1}$$

where P_t is the accumulated product at time t, V_i is the initial velocity, and k(inact) is the inactivation rate constant (85).

Identification of the 2,4-dichlorocinnamic acid and benzofuran-2-carboxylic acid metabolites. Samples for NMR analysis were generated at 30 °C in D_2O containing 40 mM phosphate buffer (pH 6.8), 200 μM αKG, 100 μM substrate, 50 μM $(NH_4)_2Fe(SO_4)_2$, 50 μM ascorbic acid, and 1.6 μM TfdA. The metabolites generated from 2,4-dichlorocinnamic acid and benzofuran-2-carboxylic acid were analyzed by ^{1}H-NMR using a Varian VXR 500 MHz NMR spectrometer.

The 2,4-dichlorocinnamic acid metabolite stability was assessed with a sample prepared by enzymatic conversion of 10 μmol of 2,4-dichlorocinnamic acid. The reaction

mixture was applied to a Pharmacia PEP-RPC HR10/10 column, washed with 0.1% trifluoroacetic acid, and eluted by using a 40:60:0.1 mixture of methanol:water:trifluoroacetic acid while monitoring the absorbance at 254 nm. Aliquots of the sample were incubated for varying time periods using specified conditions, and the concentrations of the remaining metabolite and the degradation products were measured by HPLC analysis with effluent A.

Assessing non-phenoxyacetic acid substrates with whole cells. R. eutropha JMP134 (pJP4) carries the pJP4 plasmid containing *tfdABCDEF* (21). *R. eutropha* JMP228 (pBH501aE) is a derivative of *R. eutropha* JMP134 (pJP4) where the *tfdA* gene has been interrupted by transposon mutagenesis (100). *R. eutropha* JMP228 is the strain lacking the plasmid. These cells were grown to late-exponential phase at 30 °C in MMO minimal medium (91) amended with the indicated carbon source. Cells were centrifuged and resuspended (A_{260} = 1.5) in fresh MMO with no carbon source. The cell suspensions were aerated on a stir plate. Carbon sources were added and samples were removed at various times, centrifuged, diluted, and transferred to HPLC vials. Isocratic HPLC with effluent A was used to analyze the samples for loss of the substrate peak and gain of the product peak.

Computing curve fits. KaleidaGraph for Windows by Abelbeck Software was used for computing all curve fits.

RESULTS

Identification of non-phenoxyacetic acid substrates. Prior studies have evaluated the ability of purified TfdA to degrade various phenoxyacetic acids, thiophenoxyacetic acids, and phenoxypropionic acids (30, 85, 86). The substrate associated with the highest catalytic efficiency was identified as the xenobiotic compound, 2,4-D. Here, we investigated a range of naturally occurring, non-phenoxyacid, aromatic compounds as substrates of the enzyme, along with some related synthetic derivatives of these compounds. In particular, cinnamic acids, auxin-like compounds, and naphthoxyacetic acids were tested by using oxygen electrode, HPLC, and spectrophotometric methods.

The oxygen electrode assay showed clearly enhanced levels of oxygen consumption over background for TfdA (1.25 μM) assay mixtures containing 2-chloro, 4-chloro, and 2,4-dichlorocinnamic acids, 1- and 2-naphthoxyacetic acids, and benzofuran-2-carboxylic acid (each at 300 μM). When each enzyme-containing sample was adjusted to 200 μM 2,4-D, oxygen consumption immediately increased to match that observed for samples containing only 2,4-D. Thus, none of these compounds exhibited significant inhibition or inactivation of 2,4-D hydroxylation activity by TfdA. For other compounds, oxygen consumption rates were close to background levels. This assay did not directly measure substrate conversion (e.g., some substrates may have uncoupled oxygen consumption from substrate oxidation), and spurious results were observed for at least one sample: 3,4-dihydroxycinnamic acid. This compound formed a blue color when mixed with ferrous ion (consistent with metal chelation by the catechol group) and consumed oxygen in the absence of enzyme when ascorbic acid was present, presumably similar to the non-enzymatic oxygen consumption observed for EDTA (not shown) (28).

Given these concerns, the potential substrates were quantitatively studied by using HPLC and spectrophotometric methods.

HPLC or (for 1-naphthoxyacetic acid) spectrophotometric methods provided similar results for conversion of the potential substrates. Under conditions equivalent to those used with the oxygen electrode, 2-chloro-, 4-chloro-, and 2,4-dichlorocinnamic acids, 1- and 2-naphthoxyacetic acids, and benzofuran-2-carboxylic acid were converted to products. In addition, small and variable amounts of 3,4-dihydroxycinnamic acid (caffeic acid), 3,5-dimethoxy-4-hydroxycinnamic acid (sinapinic acid), 4-hydroxy-3-methoxycinnamic acid (ferulic acid), 2,4-dimethoxycinnamic acid, 3,5-dimethoxycinnamic acid, unsubstituted cinnamic acid, 4-methoxycinnamic acid, and 3-methoxycinnamic acid were converted to products. No indication of substrate loss or product formation was observed for 2-hydroxycinnamic acid, 3-hydroxycinnamic acid, 4-hydroxycinnamic acid (coumaric acid), or 2-methoxycinnamic acid. Similarly, chromone-2-carboxylic acid, indole-2-carboxylic acid, indole-3-carboxylic acid, indole-3-acetic acid (auxin), indole-3-acrylic acid, hippuric acid, and phenylpropionic acid were not substrates of TfdA.

Kinetic analysis of non-phenoxyacetic acid substrates. Detailed studies of the enzyme kinetics were carried out with the test compounds that were convincingly shown to be substrates. Characterization of the kinetic parameters was complicated due to an irreversible inactivation of the ferrous containing enzyme that occurs during exposure to oxygen (85). An average k(inact) = 0.49 ± 0.26 min^{-1} was determined for all substrates, regardless of the substrate concentration. Initial rates were obtained by fitting the progress curves for substrate conversion, as described in the Materials and Methods. The

enzyme kinetics were further complicated by the dependence of the reaction on ascorbic acid. As illustrated in Figure 2.1, the concentration of this reductant had a large effect on the initial rate of 2,4-dichlorocinnamic acid utilization. Similar results were previously reported for conversion of the poor substrate, thiophenoxyacetic acid (85). To overcome this requirement, 200 μM ascorbic acid was used for all kinetics studies.

The results of the kinetic investigations are provided in Table 2.1. For comparison, data were obtained for TfdA metabolism of 2,4-D. The values of K_m, k_{cat}, and k_{cat}/K_m (the catalytic efficiency) associated with 2,4-D were similar to, and more accurate than, previous reports (17.5 ± 1.0 μM, 529 ± 16 min^{-1}, and 30,200 min^{-1} μM^{-1}, respectively) (30). Each of the non-phenoxyacetic acid substrates was associated with a higher K_m, a lower k_{cat}, and a lower catalytic efficiency than was observed for 2,4-D.

Metabolite Analysis. ^{1}H-NMR analysis of the sample mixture after complete transformation of 2,4-dichlorocinnamic acid by TfdA revealed (a) the disappearance of two doublets at δ 6.4 ppm and δ 7.6 ppm associated with protons bound to carbon atoms participating in the olefin bond and (b) the appearance of two new doublets at δ 3.5 ppm and δ 4.2 ppm (data not shown). The epoxide protons of phenylglycidic acid (the epoxide of cinnamic acid) exhibit identical resonances (18); thus, we conclude that TfdA oxidized the side chain double bond of 2,4-dichlorocinnamic acid to produce 2,4-dichlorophenylglycidic acid. In the case of benzofuran-2-carboxylic acid, NMR evidence revealed an analogous change in chemical shift of the proton on carbon-3 from δ 7.2 ppm to δ 3.6 ppm (data not shown). This result was consistent with similar formation of an epoxide during substrate conversion.

Figure 2.1. Ascorbic acid dependence of 2,4-dichlorocinnamic acid utilization by TfdA. The effect of ascorbic acid concentration on the initial rate of 2,4-dichlorocinnamic acid loss was monitored by using a continuous spectrophotometric assay. The standard assay buffer contained 391 nM TfdA dimer and 100 μM 2,4-dichlorocinnamic acid.

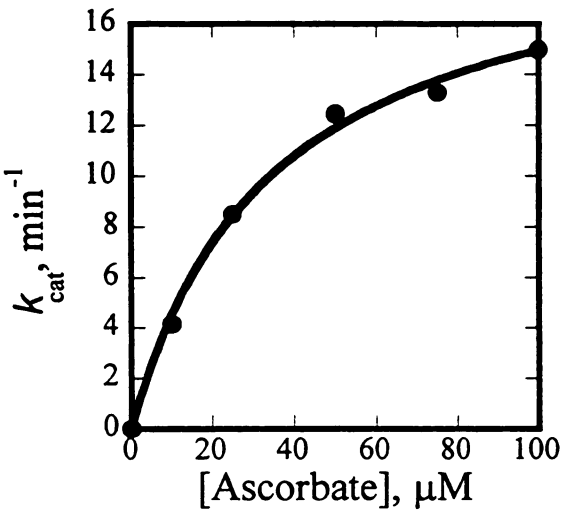


Table 2.1. Kinetics of alternative substrates of TfdA.

Substrate	K_m (μM)	k_{cat} (min^{-1})	k_{cat}/K_m ($min^{-1} \cdot mM^{-1}$)
2,4-D	20.0 ± 5.4	1020 ± 90	51,000
2-Naphthoxyacetic acid	134 ± 32	263 ± 5	1960
2,4-Dichlorocinnamic acid	190 ± 56	52.5 ± 5.1	276
Benzofuran-2-carboxylic acid	254 ± 120	105 ± 16	413
2-Chlorocinnamic acid	264 ± 52	22.1 ± 1.3	83.9
1-Naphthoxyacetic acid	622 ± 310	15.7 ± 2.9	25.2
4-Chlorocinnamic acid	>900	>38	

To further test whether the product of the 2,4-dichlorocinnamic acid is 2,4-dichlorophenylglycidic acid, we examined its pH stability and nucleophilic reactivity. Acidic conditions (pH 3) led to the rapid decomposition of the sample as two degradation products were formed (Figure 2.2, panel A). In contrast, the reaction product was stable for at least 8 days at pH 13 (data not shown). Inclusion of the strong nucleophile hydroxylamine at low pH led to the formation of two new degradation products (Figure 2.2, panel B). The low pH reactivity and distinct product profile generated in the presence of hydroxylamine are consistent with an epoxide being the initial product of 2,4-dichlorocinnamic acid metabolism by TfdA.

Whole cell studies. The potential of using intact bacterial cells to metabolize 2,4-dichlorocinnamic acid was examined using *R. eutropha* JMP134 (pJP4), *R. eutropha* JMP228 (pBH501aE), and *R. eutropha* JMP228 (Figure 2.3). The first strain encodes the entire 2,4-D pathway, whereas the second and third cultures lack *tfdA* and the pJP4 plasmid, respectively. Degradation of 2,4-dichlorocinnamic acid was observed only in the first isolate, consistent with a requirement for *tfdA* (panel A). Notably, product accumulation mirrored substrate disappearance in agreement with the inability of this strain to grow on 2,4-dichlorocinnamic acid as a sole carbon source. In contrast to the 2,4-dichlorocinnamic acid results, all three strains decomposed unsubstituted cinnamic acid (panel B). These results demonstrate that a cinnamic acid degrading system exists in the cells and does not involve *tfdA* or other loci on the pJP4 plasmid.

Figure 2.2. pH stability and nucleophilic reactivity of the TfdA reaction product derived from 2,4-dichlorocinnamic acid. The peak areas at 230 nm were determined for the 2,4-dichlorocinnamic acid reaction product (●) and degradation products associated with retention times of 2.5 min (♦), 4 min (■), 6 min (▲), and 7 min (X) of samples incubated at pH 3 (A) and pH 4 with 2 mM hydroxylamine (B).

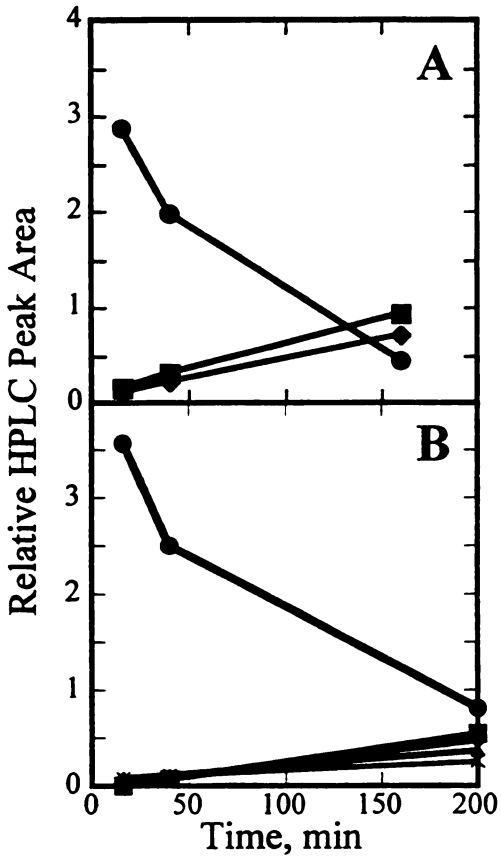
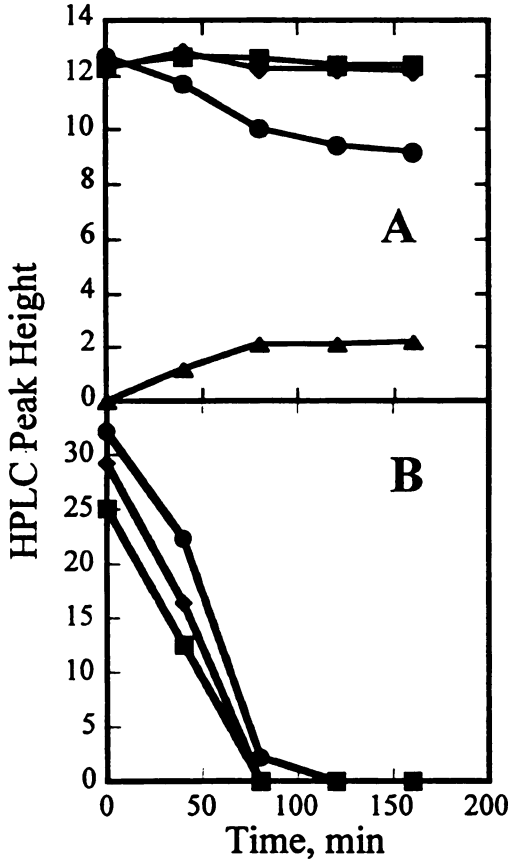


Figure 2.3. Degradation of cinnamic acids by intact cells. The time dependence was examined for decomposition of 2,4-dichlorocinnamic acid (panel A) and cinnamic acid (panel B) by *R. eutropha* JMP134 (pJP4) (●), *R. eutropha* JMP228 (■), and *R. eutropha* JMP228 (pBH501aE) (♦) by using HPLC methods and analysis of the resulting absorbances at 230 nm. In addition, the time dependence was determined for production of the metabolite derived from 2,4-dichlorocinnamic acid by *R. eutropha* JMP134 (pJP4) (▲) in panel A.



DISCUSSION

Expansion of the known substrate range of TfdA. We demonstrated that several substituted cinnamic acids are utilized (albeit poorly) as substrates by TfdA. The best of this group of substrates is 2,4-dichlorocinnamic acid, whereas the 2- or 4-monochloro derivatives are used with less efficiency and the unsubstituted cinnamic acid is metabolized even more poorly. This pattern of substrate preference parallels the situation for phenoxyacetic acids, where the 2,4-dichloro species is preferred over the monochlorinated derivatives, which are more readily metabolized than the unsubstituted species (30). Based on NMR evidence and stability/reactivity studies of the 2,4-dichlorocinnamic acid metabolite, we propose that the TfdA-generated products arising from the cinnamic acids are the side chain epoxides. Thus, TfdA is not limited to inserting oxygen into unactivated C-H bonds—it also can catalyze the easier oxidation of a C-C double bond.

Because 2,4-D chemically mimics auxin in its action as an herbicide (33), we tested whether the converse situation may occur; i.e., can auxin or auxin-like compounds chemically mimic 2,4-D and serve as a substrate for TfdA? We showed that the natural plant hormone, indole-3-acetic acid, is not a substrate of the enzyme. Furthermore, we demonstrated that indole-3-carboxylate, indole-3-acrylate, and indole-2-carboxylate also are not transformed by the enzyme. Interestingly, however, benzofuran-2-carboxylate (the analogue of indole-2-carboxylate in which the indole N is replaced by O) does serve as a substrate of TfdA. Perhaps related to this reaction, benzofuran-2-carboxylate mimics phenoxyacetic acid with the side chain linked to the aromatic ring via a methenyl carbon. NMR evidence suggests that the double bond of this substrate is converted to an epoxide

by TfdA. The ability of TfdA to selectively catalyze the transformation of benzofuran-2-carboxylate while not oxidizing indole-2-carboxylate may relate to the differences in resonance energies between the pyrrole (22 kcal/mol) and furan (16 kcal/mol) rings (23). Thus, oxygen insertion into the double bond of benzofuran-2-carboxylate is more favorable than for the indole-2-carboxylate. Another compound related in structure to both benzofuran-2-carboxylate and phenoxyacetic acid is chromone-2-carboxylate, which is not a substrate of TfdA.

1- and 2-Naphthoxyacetic acids, a third class of test compounds, were shown to be substrates of the enzyme. These compounds closely resemble phenoxyacetic acid, but contain a larger aromatic ring. The fact that TfdA is capable of transforming these compounds demonstrates that the active site is sufficiently large to allow entry of these species. Of interest, 2-naphthoxyacetic acid is a synthetic auxin predecessor to 2,4-D that was used as a rooting agent. It was never used widely as an herbicide since less 2,4-D could be applied to achieve the same benefits (79).

The ancestral role of TfdA? Soon after 2,4-D was introduced into the environment in the early 1940s, reports suggested that 2,4-D was rapidly decomposed in soil (e.g., (7)). To account for decomposition of this xenobiotic compound, we earlier proposed the existence of an ancestral gene encoding a degradative enzyme with greatest specificity toward a natural product, but capable of utilizing 2,4-D with low efficiency (40). According to this scenario, mutations arose over time to enhance the specificity of the ancestral enzyme toward 2,4-D and the new gene recombined with genes for 2,4-dichlorophenol degradation to create a 2,4-D degradation pathway. In support of this hypothesis, approximately a third of soil bacterial isolates possess a *tfdA*-like gene, but do

not measurably degrade 2,4-D (46). Thus, environmental isolates possess *tfdA*-like genes that appear to play another, yet unidentified, role besides 2,4-D catabolism. Based on the studies reported, we propose that the ancestral role of TfdA may have involved degradation of cinnamic acids. Ring-substituted cinnamic acids are widely distributed in nature and are present in sizable amounts. For example, 4-hydroxycinnamic acid is found at a concentration of 8.7 mg carbon/g carbon in the top meter of peat soil (56). Plants synthesize large amounts of cinnamic acids by phenylalanine deamination, and they catalyze a variety of ring substitution, condensation, degradation, reduction, or conjugation reactions to produce a host of cinnamic acid-related compounds including flavonoids, hydroxybenzoates, and lignin (93). Degradation of substituted cinnamic acids occurs by several pathways (for a review, see (82)), but it would not be surprising to learn that another pathway remains to be described. Future studies will test the cinnamic acid degrading abilities of the enzyme products from *tfdA*-like genes in environmental isolates.

Biocatalysis. Because of the wide range of enantiospecific chemical reactions in which epoxides can participate, epoxides serve as intermediates in a variety of medically important chemical syntheses (6). Our demonstration that purified TfdA enzyme and whole cells containing this enzyme can convert chlorinated cinnamic acids to the corresponding epoxides expands the repertoire of available synthetic building blocks. In particular, phenylglycidic acid, the epoxide of unsubstituted cinnamic acid, has been used as an intermediate in the production of taxol (17) and (2*S*,3*S*)-diltiazem (34). The bioconversion of the chlorinated cinnamic acids enables the synthesis of several drug analogues containing novel functional groups. Additionally, the use of TfdA may allow

for the production of enantiopure epoxides. TfdA has been shown previously to hydroxylate only (*S*)-dichloroprop, a phenoxypropionate herbicide (86). Additionally, a fellow member of the superfamily of α-KG dependent dioxygenases, EpoA, catalyzes the enantiomer specific production of fosfomycin ((–)-*cis*-1,2-epoxypropylphosphonic acid) from *cis*-propenylphosphonic acid in *Penicillium decumbens* (104).

ACKNOWLEDGEMENTS

I would like to thank Raghavakaimal Padmakumar for initiating several of these studies and performing NMR analysis of the benzofuran-2-carboxylate product.

CHAPTER 3

MODELING OF TfdA AND USE OF INSTRINSIC TRYPTOPHAN FLUORESCENCE AS A PROBE OF METAL AND α-KETOGLUTARATE BINDING

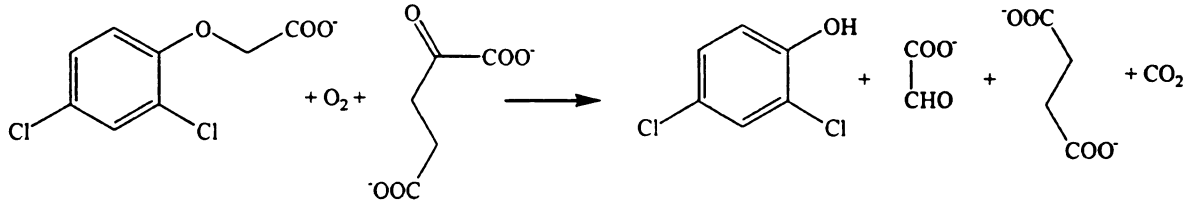
This chapter includes my contributions to two submitted papers.

The modeling of TfdA is presented in: Elkins, J. M., Ryle, M. J., Clifton, I. J., Dunning Hotopp, J. C., Lloyd, J. S., Burzlaff, N. I., Baldwin, J. E., Hausinger, R. P., and P. L. Roach (2002) "X-Ray crystal structure of *Escherichia coli* taurine/α-ketoglutarate dioxygenase complexed to ferrous iron and substrates," *Biochemistry*, in press.

The intrinsic tryptophan fluorescence is presented in: Dunning Hotopp, J. C., Auchtung, T. A., Hogan, D. A., and R. P. Hausinger (2002) "Intrinsic tryptophan fluorescence as a probe of metal and α-ketoglutarate binding to TfdA, a mononuclear non-heme iron dioxygenase," *Journal of Inorganic Biochemistry*, in press. My contributions to this paper include purification and characterization of the tryptophan mutants and measuring the dissociation constants for the various metals tested. They do not include designing or synthesizing the plasmids containing the tryptophan mutants.

α-Ketoglutarate (αKG) dioxygenases comprise an important yet understudied superfamily of enzymes. Members of the superfamily are widely distributed among eukaryotes and eubacteria where they catalyze a diverse array of reactions (78). αKG dioxygenases chelate one ferrous ion and bind one αKG per catalytic site. Substrate and oxygen then bind to the enzyme and the oxidative decarboxylation of αKG is coupled to the oxidation of the primary substrate. The most common oxidation of the primary substrate involves hydroxylation of an unactivated carbon, but desaturations, epoxidations, or ring formation and expansion reactions can also result (Scheme 1.1).

The studies described here focus on 2,4-dichlorophenoxyacetic acid (2,4-D)/αKG dioxygenase (TfdA), which hydroxylates the herbicide 2,4-D producing a hemi-acetal that spontaneously decomposes to 2,4-dichlorophenol.



A model of TfdA was developed based on the crystal structure of two homologues, taurine/αKG dioxygenase (TauD) (25) and clavaminate synthase 1 (CAS) (63). The latter proteins are 30 % and 11.5 % identical in sequence, respectively, to the target protein. TfdA is particularly well characterized by biochemical and biophysical methods (e.g., (30, 43, 47, 62, 85, 86, 105)) allowing for substantiation of the model. In addition, a simple method was developed to rapidly determine metal and αKG binding affinities so as to circumvent concerns about enzyme metallocenter oxidation. The intrinsic tryptophan fluorescence associated with TfdA was shown to be quenched by these components, providing a useful method to measure their binding affinities. Site-directed

mutagenesis methods were used to demonstrate that the tryptophan residue predicted to lie closest to the metal- and αKG-binding sites in a model generated of TfdA does not serve as the fluorescent reporter.

EXPERIMENTAL

Modeling Methods. The PHD predicted TfdA secondary structure elements (83) were aligned with the crystallographically determined secondary structures of TauD (PDB ID. 1GQW) (25) and CAS1 (PDB ID. 1DS1) (109) (Figure 3.1). The TfdA sequence was modeled onto the TauD chain A and CAS1 crystal structures ignoring iron, αKG, and primary substrates using Modeller4 (87). TfdA residues for which there were not coordinates of an equivalent residue in TauD were masked. From the initial model, the loop refinement method was used to correct an α-helix distortion near residues 34-36 and to remove six disallowed phi/psi angles. This method was also used in efforts to remodel the loop formed by residues 86-110 which contained a knot. Numerous possible models were generated for this region, including the additional criterion that residues 96-103 form an α-helix as suggested by PHD secondary structure prediction. No model was convincingly superior to the others and, since this region of TfdA contains a sixteen amino acid insertion when compared to TauD, residues 86-111 were left out of the model. Insertion of αKG and iron into the TfdA model was based on alignment of the 2-His-1-carboxylate facial triads of TfdA and TauD, with direct substitution of the αKG and iron positions. The residues interacting with the C-5 carboxylate of αKG (Thr141 and Arg274) were adjusted to optimize hydrogen bonding and ionic interactions, based on the CAS1 and TauD coordinates. 2,4-D was manually positioned into the modeled TfdA active site using InsightII (Molecular Simulations Inc.). As an initial constraint, the α carbon of the 2,4-D side chain and the carbon adjacent to the sulfur of taurine (i.e., the sites of hydroxylation) were superimposed. Since TfdA selectively oxidizes the *pro-R* hydrogen of 2,4-D (86), this hydrogen was positioned to face the iron active site. The

Figure 3.1. Secondary structure based sequence alignment of CAS1, TfdA, and TauD. An initial alignment was produced based on a 3D alignment of CAS1 and TauD. TfdA was then aligned based on a PSI-BLAST to convergence and its predicted secondary structure elements to the existing alignment. The amino acids corresponding to those in TauD secondary structure elements and the homologous residues in CAS1 and TfdA are colored red for helix, green for β-strands of the jellyroll motif and blue for other β-strands. Conserved Fe(II)-binding ligands are marked with black arrows.

```
TfdA  --MSVVANPLHPLFAAGVEDIDLREALG-------------------------------------
TauD  MSERLSITPLGPYIGAQISGADLTRPLS-------------------------------------
CAS1  --------------MTSVDC--------TAYGPELRALAARLPRTPRADLYAFLDAAHTAAASLP

TfdA  STEVREIERLMDEK---SVLVFRGQPL-----------------------SQDQQIAFARNFGPL
TauD  DNQFEQLYHAVLRH---QVVFLRDQAI-----------------------TPQQQRALAQRFGEL
CAS1  GALATALDTFNAEGSEDGHLLLRGLPVEADADLPTTPSSTPAPEDRSLLTMEAMLGLVGRRLGLH

TfdA  EGGFIKVNQRPSRF-------KYAELADISNVSLDGKVAQRDAREVVGNFANQLWHSDSSFQQPA
TauD  HIHPVYPHAEGVD--------EIIVLDTHNDNPPDND----------------NWHTDVTFIETP
CAS1  TG-----------YRELRSGTVYHDVYP---------SPGAHHLSSETSETLLEFHTEMAYHRLQ

TfdA  ARYSMLSAVVVPPS----GGDTEFCDMRAAYDALPRDLQSELEGLRAEHYALNS----RFLLGDT
TauD  PAGAILAAKELPST----GGDTLWTSGIAAYEALSVPFRQLLSGLRAEHDFRKSFPEYKYRKTEE
CAS1  PNYVMLACSRA---DHERTAATLVASVRKALPLLDERTRARLLDRRMPCCV--------------

TfdA  D---YSEAQRNAMPPVNWPLVRTHAGSG---RKFLFIGAHASH--VEGLPVAEGRMLLAELLEHA
TauD  EHQRWREAVAKNP-PLLHPVVRTHPVSG---KQALFVNEGFTTRIVDVS-EKESEALLSFLFAHI
CAS1  -DVAFRGGVDDPGAIAQVKPLYG-----DADDPFLGYDRELLA-----PEDPADKEAVAALSKAL

TfdA  TQREFVYRHRWNVGDLVMWDNRCVLHRGRRYDISA---------RRELRRATTLDDAVV------
TauD  TKPEFQVRWRWQPNDIAIWDNRVTQHYANADYLPQ---------RRIMHRATILGDKPFYRAG--
CAS1  D--EVTEAVYLEPGDLLIVDNFRTTHART------PFSPRWDGKDRWLHRVYIRTDR--NGQLSG

TfdA  --------------
TauD  --------------
CAS1  GERAGDVVAFTPRG
```

substrate carboxylate was rotated to allow favorable interactions with Arg278, Lys71, His214 and the backbone amide of Ser117.

Protein purification and kinetics. Wild-type and variant forms of TfdA were purified from *Escherichia coli* DH5α as previously described (Chapter 2). Protein concentrations were assessed using a commercial protein assay with standard curves determined with bovine serum albumin. Activity measurements were made using the 4-aminoantipyrene assay (30). Curve fits were computed using KaleidaGraph for Windows by Abelbeck Software.

Site-directed mutagenesis. The amino acid residue numbering scheme used here is based on that derived from the gene sequence rather than that of the purified protein, and differs by the addition of one residue compared to that used previously (47, 62). Trp113, Trp195, Trp248, and Trp256 TfdA variants were created using the Stratagene QuikChange Mutagenesis System, the forward mutagenic primers listed in Table 3.1, their complements, and pUS311 (30) as the starting plasmid. This pUC19 derivative contains the *Ralstonia eutropha* JMP134 (pJP4) *tfdA* gene (94).

Fluorescence Measurements. The intrinsic tryptophan fluorescence of 0.25 μM TfdA dimer in 25 mM MOPS, pH 6.75, was examined at room temperature (20 °C) by using a Hitachi F-4500 spectrofluorometer. The excitation wavelength typically was maintained at 290 nm (5 nm slit width) while monitoring emission from 300 to 360 nm (5 nm slit width). αKG was prepared in the same MOPS buffer as the protein, whereas metals were freshly dissolved in water. Cations were supplied as $Fe(NH_4)_2(SO_4)$, $ZnSO_4$, $CuSO_4$, $CoSO_4$, $MnSO_4$, $MgSO_4$, $CaCl_2$, LiCl, Li_2SO_4, NaCl, Na_2SO_4, KCl, and K_2SO_4.

Table 3.1. Forward mutagenic primer sequences.

Mutant	Forward Primer
W113F	5'-GAA CCA GCT CTT CCA CAG CGA CAG C-3'
W113L	5'-GAA CCA GCT CTT GCA CAG CGA CAG C-3'
W195F	5'-CGC CGG TCA ACT TGC CGC TGG TTC G-3'
W195L	5'-CGC CGG TCA ACT TCC CGC TGG TTC G-3'
W248F	5'-CGT GTA CCG GCA TCG CTT CAA CGT GGG AGA TCT GG-3'
W248L	5'-CCG GCA TCG CTT GAA CGT GGG AGA TC-3'
W256F	5'-GAT CTG GTG ATG TTC GAC AAC CGC TGC-3'
W256L	5'-GAT CTG GTG ATG TTG GAC AAC CGC TGC-3'

Calculation of binding constants. Protein fluorescence changes accompanying the binding of metal or αKG were analyzed by curve fitting (KaleidaGraph) according to equation 3.1.

$$\Delta F = \Delta F_{max} * \frac{K_d + [L_T] + [E_T] - \sqrt{(K_d + [L_T] + [E_T]) - 4[L_T][E_T]}}{2[E_T]} \qquad \text{Eq. 3.1}$$

In this equation, the observed change in fluorescence (ΔF) is related to the maximal fluorescence change (ΔF_{max}) extrapolated to infinite concentration of titrant. Based on the known total concentrations of enzyme subunit ($[E_T]$) and ligand ($[L_T]$), the K_d and ΔF_{max} were calculated.

RESULTS AND DISCUSSION

Modeling of the TfdA *structure.* The three-dimensional structure of TfdA was modeled and substrates were positioned at its active site, as illustrated in Figure 3.2 (PDB ID. 1GQX). Analysis of the TfdA model by Procheck gave an acceptable overall average G factor of –0.22 and no residues in disallowed regions (Figure 3.3). The αKG chelates the metal center and is predicted to form a salt bridge to Arg274 and a hydrogen bond to Thr141. Interactions with the 2,4-D carboxylic acid are proposed to involve Arg278, Lys71, His214 and the backbone amide of Ser117. Two of these interactions, Arg270 and the N-H of Val102, correspond to the taurine sulfonate-binding residues of TauD. Gly67 in TfdA replaces the third sulfonate-binding residue, His70, whereas TfdA residues His214 and Lys71 do not have similar counterparts in TauD. In addition to the potential interaction with the carboxylate, TfdA Lys71 could reasonably interact with the ether oxygen of 2,4-D. Additional interactions with the substrate are certain to involve residues 86-111 which could not be convincingly modeled. However, some potential models suggested that Lys95 could orient to hydrogen bond with the ether oxygen of 2,4-D.

TfdA is particularly well characterized by biochemical and biophysical methods (e.g., (30, 43, 47, 62, 85, 86, 105)) allowing for substantiation of the model. The TfdA model is consistent with previous observations and predictions made about TfdA. For example, Arg78 is surface exposed in the model and accounts for the reported protease sensitivity of this site (30). In addition, the model supports previous predictions based on site-directed mutagenesis and spectroscopic analyses that His114, Asp116, and His263 bind iron (47). The model vindicates a prior suggestion that Arg274 forms a salt bridge with the C-5 carboxylate of αKG (47). Thr141 is positioned to hydrogen bond this

Figure 3.2. The active site of TfdA. The crystal structures of TauD and CAS1 were used to model the homologous protein TfdA. The residues believed to be involved in substrate binding, and for which variants were made in these studies, are illustrated.

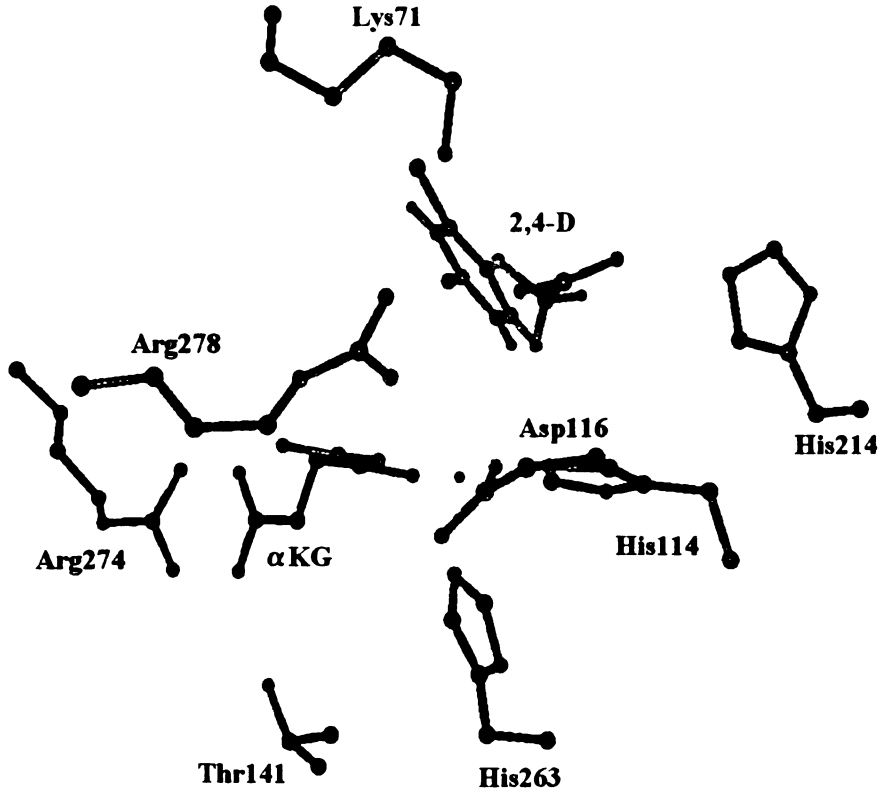
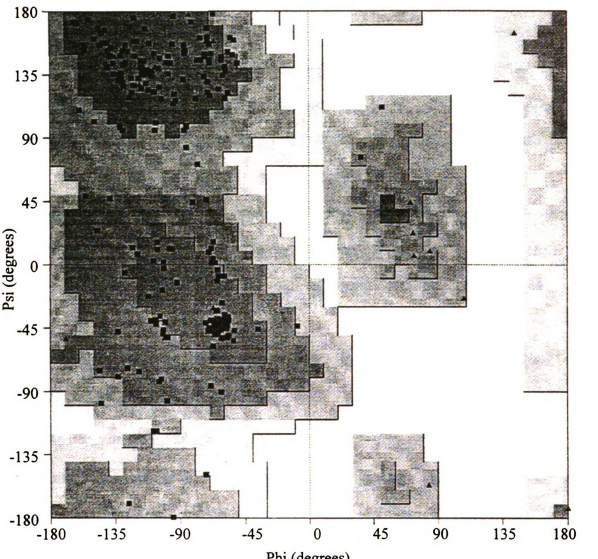


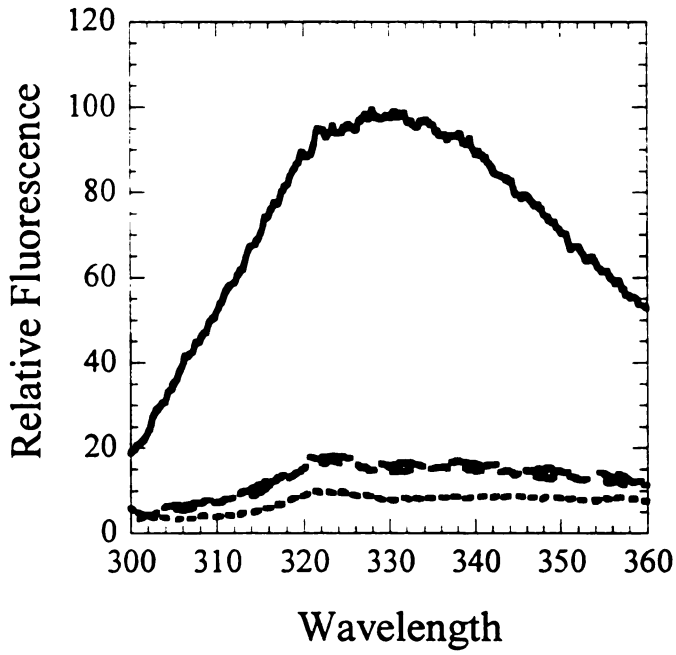
Figure 3.3. Ramachandran plot of the TfdA model. A Ramachandran plot of the TfdA model was generated using Procheck. Of the 258 residues presented in the model, no residues were found in disallowed regions and only 1.8% of residues were found in generously allowed regions. Of the remaining 98.2%, 19.2% were found in additional allowed regions and 79.0% were found in the most favored regions.



moiety of αKG, consistent with conservation of a hydroxylated amino acid at that position in Group II dioxygenases. The model is in agreement with previous results suggesting that His214 functions in 2,4-D binding or catalysis (47). Specifically, His214 interaction with a carboxylate oxygen of 2,4-D accounts for the 10-fold increase in K_m of 2,4-D in a H214A mutant. Although substitution of His217 by Ala was also shown to lead to a 2.5-fold increase in K_m of 2,4-D (47), this residue is more distant from the active site and is predicted not to directly bind substrate or participate in catalysis. The observed effects may alternatively arise from reduced substrate access to the active site or to altered positioning of a 2,4-D ligand. Finally, evidence for an interaction with the substrate ether atom, presumed to involve Lys71, but possibly also Lys95, is provided by studies involving the substrate 2,4-dichlorocinnamic acid (Chapter 2). This compound closely resembles 2,4-D in structure, but lacks the ether bond leading to a much higher K_m.

Intrinsic Tryptophan Fluorescence as a Probe of Metal and αKG Binding to TfdA. The tryptophan fluorescence spectrum of TfdA, a representative αKG-dependent dioxygenase, is quenched by addition of Fe(II), αKG, or both compounds (Figure 3.4). The presence of saturating concentrations (200 μM) of either Fe(II) or αKG leads to a 50-85% reduction in fluorescence compared to that of TfdA apoprotein. When both components are added, the fluorescence is further diminished to relative intensities that are ~10% of the initial value. Additions of 200 μM Fe(II) and/or αKG to 5 μM tryptophan in water do not cause decreases in fluorescence. The binding of the 2,4-D to TfdA was not examined by this approach because this substrate absorbs appreciably at the excitation wavelength, thus interfering with the analysis. The fluorescence changes in

Figure 3.4. Fluorescence spectra of TfdA. The fluorescence spectra of TfdA were obtained in MOPS buffer alone (solid line), or in MOPS buffer containing 200 μM Fe(II) or 200 μM αKG (dashed lines), and with both compounds (dotted line).



TfdA likely arise from changes in the environment of one or more of the four tryptophan residues in TfdA.

Fluorescence titration studies were used to directly determine the K_d for Fe(II) and αKG binding to TfdA (Table 3.2). The 3.35 μM K_d of αKG calculated by this approach closely matches the reported αKG K_m values of 3.2 μM and 4.9 μM based on kinetic studies (30, 47). Similarly, the 7.45 μM K_d of Fe(II) determined here approximates the concentration (~2.0 μM) shown to give rise to half-maximal activity (47). The presence of 200 μM αKG decreased the apparent affinity for Fe(II), perhaps due to partial chelation of the metal ion by the α-keto acid. In contrast, the presence of 200 μM Fe(II) had essentially no effect on the apparent affinity for αKG (Table 3.2). The dications of Mg, Ca, Co, Cu, Mn, and Zn bound to TfdA with K_d values similar to that observed for Fe(II) (Table 3.3), whereas monovalent cations had negligible effect. The similarities in K_d values for the dications do not correspond to their ability to inhibit TfdA; e.g., Cu exhibits competitive inhibition with K_i ~0.2 μM (43) while Mg does not inhibit.

Identification of tryptophan residues acting as reporter groups. Site-directed mutagenesis studies were carried out to create TfdA variants lacking each of the four Trp residues (Trp113, Trp195, Trp248, and Trp256) in an attempt to identify the aromatic group(s) acting as the fluorescent reporter group. Unfortunately, cultures expressing recombinant genes encoding the W248F, W248L, W195F, and W195L variants of TfdA produced insoluble protein for all cell growth conditions examined. This problem also had been encountered in prior TfdA mutagenesis studies (47). The W113F and W113L variants also could not be studied because these proteins were either not synthesized or

Table 3.2. Summary of the metal and αKG binding to wild-type TfdA

Sample	K_d of Fe(II) (μM)	K_d of αKG (μM)
TfdA	7.45 ± 0.61	3.35 ± 0.35
TfdA + 200 μM Fe(II)	--	4.78 ± 0.69
TfdA + 200 μM αKG	36.3 ± 19.1	--

Table 3.3. Summary of metal binding to wild-type TfdA.

Metal	K_d of Metal (μM)
Fe^{2+}	7.45 ± 0.61
Cu^{2+}	5.0 ± 0.4
Zn^{2+}	8.4 ± 2.9
Co^{2+}	3.0 ± 0.7
Mn^{2+}	6.8 ± 0.7
Mg^{2+}	6.2 ± 1.3
Ca^{2+}	3.8 ± 0.5

were very rapidly degraded in the cell. Perhaps related to this result, TfdA is known to catalyze enzyme self-hydroxylation during extended incubation with Fe(II), αKG, and oxygen resulting in Trp113 conversion to hydroxytryptophan (62). This reaction was hypothesized to serve a protective role for the enzyme; i.e., hydroxylation of this side chain may prevent the occurrence of more damaging oxidative reactions such as those resulting in cleavage of the peptide backbone. The lack of this protective side chain in the W113F and W113L variants may lead to their rapid oxidative degradation in the cytoplasm. The TfdA fluorescence spectrum that is quenched by Fe(II) and αKG is not associated with hydroxytryptophan based on the distinct properties (e.g., excitation beyond 300 nm) of this modified aromatic group (22). Because of these complications, only W256F and W256L TfdA variants were purified and characterized. The kinetic properties of these variants closely resembled those of the wild-type enzyme (Table 3.4).

The model of the TfdA structure based on that of TauD predicts that Trp256 lies only 5 Å from the Fe(II) site and 4.5 Å from the C-1 carboxylate of αKG (Figure 3.5). Nevertheless, this tryptophan is not the reporter group studied above as shown by the fluorescence properties of the W256F and W256L TfdA variants. The relative intensity (data not shown) and the Fe(II)- and αKG-dependent changes in fluorescence intensity of these mutant proteins closely paralleled those of the wild-type enzyme (data not shown and Table 3.4, respectively), and these mutants resemble wild-type protein in their kinetic properties (Table 3.4-3.5). The TfdA structural model predicts that Trp195 lies on the surface of the protein nearly 15 Å from, and on the opposite face to, the active site. Crystallographic studies with other αKG-dependent dioxygenases reveal the absence of significant protein conformational changes upon binding Fe(II) or αKG (101, 106).

Table 3.4. 2,4-D kinetics of variant TfdA proteins.

Mutant	Range (μM)	k_{cat} (min^{-1})	K_m (μM)	k_{cat}/K_m (min^{-1} mM^{-1})
Wt[a]	5-1000	1020 ± 90	20.0 ± 5.4	51,000
W256F	10-1000	352 ± 10	17.6 ± 1.6	20,000
W256L	10-1000	403 ± 26	13.0 ± 3.7	31,000

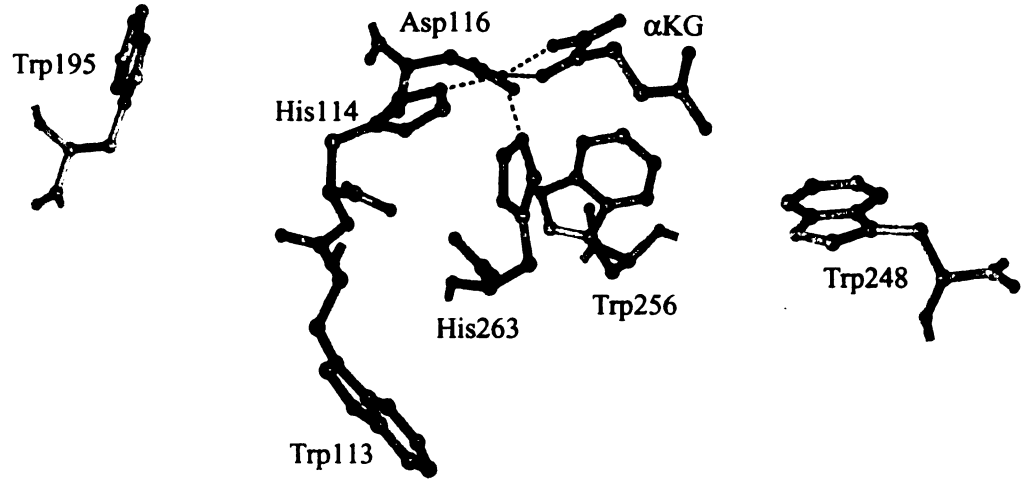
[a]Chapter 2

Table 3.5. αKG kinetics of variant proteins.

Mutant	Range (μM)	k_{cat} (min^{-1})	K_m (μM)	k_{cat}/K_m (min^{-1} mM^{-1})	K_d(μM)
Wt	1.5-10	643 ± 44[a]	3.2 ± 0.6[a]	201,000[a]	3.4 ± 0.4
W256F	5-1,000	362 ± 32	<10	>36200	2.3 ± 0.3
W256L	5-1,000	333 ± 38	<10	>33300	2.9 ± 0.3

[a]Fukumori et al. (30).

Figure 3.5. Predicted positions of tryptophan residues in TfdA. Of the four Trp residues in TfdA, three are close to the active site and could reasonably sense changes induced by Fe(II) and αKG binding: Trp113 is adjacent to a metal ligand, Trp256 lies close to both the metal and cofactor, and Trp248 is close to the αKG binding site. In contrast, Trp195 is on the surface and distant from the active site.



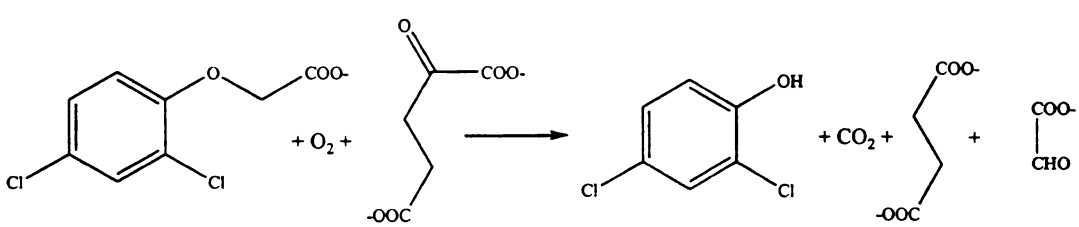
Thus, we conclude that Trp195 is unlikely to sense changes induced by Fe(II) or αKG binding. The remaining two tryptophan residues remain as reasonable options for the fluorescence reporter group(s) based on the structural model. Trp248 is only 4.1 Å from the C-5 carboxylate of αKG while also forming a hydrogen bond with Thr141 that potentially binds the α-keto acid, but this residue is more distant (9.7 Å) to the metal site. Although adjacent to a metal liganding residue, the aromatic ring of Trp113 is approximately 10 Å from both the Fe(II) and αKG binding sites. We conclude that the Fe(II)- and αKG-dependent changes in fluorescence are likely due directly to binding of molecules to the active site or to a resultant conformational change which then leads to an altered environment that influences the fluorescence of either or both Trp113 and Trp248. These residues are partially conserved among the αKG-dependent dioxygenases; thus, this technique may be applicable to estimating metal and/or αKG binding affinities of other family members.

CHAPTER 4

PROBING THE SUBSTRATE-BINDING OF 2,4-DICHLOROPHENOXYACETATE/α-KETOGLUTARATE DIOXYGENASE BY SITE-DIRECTED MUTAGENESIS AND MECHANISM-BASED INACTIVATION

α-Ketoglutarate (αKG) dioxygenases are mononuclear non-heme Fe(II) enzymes that couple the oxidative decarboxylation of αKG to oxidation of their primary substrates. Depending on the specific enzyme, the primary substrate may be a protein side chain, a cellular metabolite, or a compound taken up from the environment (78). Most commonly, substrate oxidation involves hydroxylation at an unactivated carbon atom, but desaturations, epoxidations, or ring formation or expansion reactions can also result. Based on crystal structures and site-directed mutagenesis studies, the metal is bound by two histidines and a carboxylic acid-containing side chain (16, 25, 47, 80, 101, 106, 109). The iron is also coordinate by the αKG C-1 carboxylate and C-2 keto group, while the C-5 carboxylate forms a salt bridge with an arginine or lysine side chain. A conserved jellyroll architecture serves as the platform for these residues (16, 25, 80, 101, 106, 109), which form a conserved $HX(D/E)X_nHX_{(\sim 10)}R$ motif (47).

2,4-Dichlorophenoxyacetic acid (2,4-D)/αKG dioxygenase (TfdA) is a representative family member that hydroxylates the herbicide 2,4-D producing a hemi-acetal that spontaneously decomposes to 2,4-dichlorophenol and glyoxylate.



Previous site-directed mutagenesis studies revealed the roles of His114, Asp116, and His263 in binding iron and suggested that His214 functions in 2,4-D binding or catalysis. Recently, a model of the three-dimensional structure of TfdA (Fig. 4.1) was developed based on the crystal structure of two homologues, taurine/α-ketoglutarate dioxygenase (TauD) (25) and clavaminate synthase 1 (CAS) (63). The TfdA model was consistent

with prior mutagenesis studies and identified a set of residues likely to be involved in 2,4-D binding and αKG binding.

The studies described here utilize site-directed mutagenesis approaches to test the importance of the proposed substrate binding residues of TfdA by characterizing selected mutant proteins. Additional TfdA mutants were characterized in order to assess the effects of removing a protease-sensitive site and to explore the importance of tyrosine residues located near the metal center. Furthermore, results demonstrating that phenylpropiolic acid (PPA) is a mechanism-based, irreversible inactivator of TfdA are presented and related to what is known about the structure of TfdA.

EXPERIMENTAL

Site-directed mutagenesis. The amino acid residue numbering scheme used here is based on that derived from the gene sequence rather than that of the purified protein, and differs by the addition of one residue compared to that used previously (47, 62). TfdA variants were created using the Stratagene QuikChange Mutagenesis System using the primers listed in Table 4.1 and pUS311 (30) as the starting plasmid. This pUC19 derivative contains the *Ralstonia eutropha* JMP134 (pJP4) *tfdA* gene (94).

Protein purification methods. Wild-type and all variant forms of TfdA except Y126F protein were purified from *Escherichia coli* DH5α as previously described (Chapter 2). Protein concentrations were assessed by using the BioRad assay and bovine serum albumin as the standard. Y126F TfdA was temperature sensitive so the protein was purified at 4 °C with alterations to the last two chromatographic steps. Specifically, the active fractions eluting off the DEAE-Sepharose column were pooled, adjusted to 1 M ammonium sulfate, loaded onto a phenyl-Sepharose column (2.5 cm x 19 cm) equilibrated with TE buffer (10 mM Tris, 1 mM EDTA, pH 7.7) containing the same concentration of this salt, and chromatographed with a 400 mL linear gradient from 1 to 0 M ammonium sulfate. The active fractions were dialyzed to remove the ammonium sulfate and then chromatographed on a Q-Sepharose column (2.5 cm x 19 cm) using TE buffer and a 400 mL linear gradient to 200 mM NaCl in TE.

Kinetic analyses. Typical activity measurements used the 4-aminoantipyrene assay (30) for detecting 2,4-dichlorophenol. Oxygen consumption or 2,4-dichlorocinnamic acid oxidation were monitored using previously described oxygen electrode or HPLC methods, respectively (Chapter 2). As previously documented (85),

Table 4.1. Forward mutagenic primer sequences

Mutant	Forward Primer
K71Q	5'-GGC GGT TTC ATC CAG GTC AAT CAA AG-3'
K71L	5'-GGC GGT TTC ATC TTG GTC AAT CAA AG-3'
R78Q	5'-GTC AAT CAA AGA CCT TCG CAA TTC AAG TAC GCG GAG TTG-3'
Y81F	5'-CGA GAT TCA AGT TCG CGG AGT TGG-3'
K95L	5'-CAG TCT CGA CGG CCT GGT CGC GCA AC-3'
K95Q	5'-CAG TCT CGA CGG CCA GGT CGC GCA AC-3'
T141V	5'-CGG GCG GCG ACG TCG AGT TCT GCG AC-3'
Y169F	5'-GCC GAG CAC TTC GCA CTG AAC TCC C-3'
Y244F	5'-GAA TTC GTG TTC CGG CAT CGC TG-3'
R274Q	5'-CAT CTC GGC CAG GCA AGA GCT GCG CCG G-3'
R274L	5'-CTC GGC CAG GCT TGA GCT GCG CC-3'
R278Q	5'-GTG AGC TGC GCC AGG CGA CCA CCC-3'
R278L	5'-GTG AGC TGC GCC TCG CGA CCA CCC TG-3'

TfdA activity decreases over time by a combination of irreversible inactivation and ascorbate-reversible inactivation. To analyze the kinetics of designated variant proteins whose activity decreased over the assay time course, progress curves were analyzed by fitting the data to the Eq. 4.1 where P_t is the accumulated product at time t, V_i is the initial velocity, and k(inact) is the inactivation rate constant (85).

$$P_t = V_i\,(1\text{-}e^{-k(\text{inact})t})k(\text{inact})^{-1} \qquad \text{Eq. 4.1}$$

In cases where high concentrations of substrate decreased the k_{cat}, the data were analyzed by fitting to Eq. 4.2

$$k = \frac{k_{cat}[S]}{K + [S] + \frac{[S]^2}{K_{si}}} \qquad \text{Eq. 4.2}$$

The K_d values associated with binding of Fe(II) and αKG were measured by using intrinsic tryptophan fluorescence and fitting to Eq. 4.3 as previously described (Chapter 3).

$$\Delta F = \Delta F_{\max} * \frac{K_d + [L_T] + [E_T] - \sqrt{(K_d + [L_T] + [E_T]) - 4[L_T][E_T]}}{2[E_T]} \qquad \text{Eq. 4.3}$$

In this equation, the observed changes in fluorescence (ΔF) are related to the maximal fluorescence change (ΔF_{max}) extrapolated to infinite concentration of titrant. Based on the known total concentrations of enzyme subunit ($[E_T]$) and ligand ($[L_T]$), the K_d and ΔF_{max} were calculated. Curve fits were computed using KaleidaGraph for Windows by Abelbeck Software.

Inactivation kinetics. The inactivation of TfdA by PPA was studied in 10 mM imidazole, 1 mM αKG, 5 μM $(NH_4)_2Fe(SO_4)_2$, and 20 μM ascorbic acid containing varying concentrations of the inactivator at 30 °C. Stock solutions of all reagents were

made fresh prior to each set of experiments. Aliquots were withdrawn at various time intervals, diluted 100-fold into fresh assay buffer containing 1 mM 2,4-D, and residual activities were determined as in the standard 4-aminoantipyrene assay procedure. These data were plotted according to log(residual activity) versus time in order to obtain the inactivation rates, k(inact), and further analyzed by fitting to the following equation:

$$k(inact) = \frac{k(inact)_{max} * [I]}{K_i + [I]} \qquad \text{Eq. 4.4}$$

where [I] is the inactivator concentration, $k(\text{inact})_{max}$ is the inactivation rate at saturating [I], and K_i is the [I] that gives rise to half maximal rates of inactivation.

Large-scale inactivation of TfdA by PPA followed by alkylation and protease digestion. TfdA (1-5 mg) was inactivated by PPA in 10 mM imidazole buffer (pH 6.8) containing 1 mM αKG, 1 mM PPA, 5 μM $(NH_4)_2Fe(SO_4)_2$, and 20 μM ascorbic acid at 30 °C . Inactivation was complete by 2 min as confirmed by using the 4-aminoantipyrene assay. PPA-treated protein was dialyzed against water, lyophilized, and redissolved in 6 M guanidine hydrochloride, 100 mM Tris (pH 8), and 10 mM EDTA . Using darkened vials, the sample was degassed, and dithiothreitol was added at ten-fold excess over the cysteines. After 30 min at 50 °C, iodoacetate was added in two-fold excess to dithiothreitol. The alkylation reaction was quenched with β-mercaptoethanol and the sample was dialyzed into 50 mM ammonium bicarbonate (pH 7.8). Trypsin (1 % wt/wt) was added twice with incubation at 37 °C for 1 h after each addition. Digestion was confirmed by SDS-PAGE and matrix-assisted laser desorption ionization (MALDI) mass spectrometry (Michigan State University Mass Spectrometry Facility). The MALDI mass spectra were measured at an accelerating potential of +25kV using a Voyager STR time

of flight mass spectrometer. Spectra represent the accumulation of 50-100 laser shots fired at the sample in the matrix α-cyano-4-hydroxycinnamic acid.

Separation and characterization of TfdA peptides. The tryptic digests of control and modified proteins were chromatographed on a pep-RPC (HR10/10) (Pharmacia) column by using a linear 200 mL gradient from 0.1 % trifluoroacetic acid (TFA) in water to methanol containing 0.1% TFA while monitoring the absorption at 280 nm. All peptides containing an aromatic moiety were characterized by MALDI mass spectrometry (Michigan State University Mass Spectrometry Facility).

RESULTS AND DISCUSSION

Site-directed mutagenesis of proposed 2,4-D ligands. Based on TfdA structural modeling studies, interactions with the 2,4-D carboxylic acid were suggested to involve Arg278, Lys71, His214 and the backbone amide of Ser117 (Figure 4.1) (Chapter 3). The model also indicated that a loop comprising residues 86-111 made up one face of the 2,4-D binding pocket, however a consensus structure for this region could not be obtained and these twenty-five residues were omitted from the final model. Notably, several possible structures for this loop placed Lys95 where it could assist in binding the 2,4-D via the carboxylate or the ether moiety. Site-directed mutagenic evidence for the participation of His214 in binding 2,4-D was described previously (47) and supports at least that portion of the model. In an extension of this approach, variants of TfdA with substitutions of Arg278, Lys71, and Lys95 were constructed and the proteins characterized

The R278Q, K71L, K71Q, K95L, and K95Q TfdA variants were purified and characterized to assess the roles of the residues in catalysis and 2,4-D binding. Iron binding to these proteins was not affected based on K_d determinations using fluorescence techniques (data not shown). Kinetic analyses revealed only modest changes in k_{cat}, with an approximately two-fold increase in the rate for the K71Q mutant and four- to five-fold decreases for the K71L and R278Q variants (Table 4.2). Turnover-dependent inactivation rates were unaffected (data not shown). The most striking changes observed for these proteins involved K_m, ranging from a 2-fold increase in the case of K95Q to a 200-fold increase for R278Q (Table 4.2). These results are consistent with a role for Arg278, Lys71, and perhaps Lys95 in substrate binding. To further examine the roles of these

Fig. 4.1. The active site residues in TfdA. The crystal structures of TauD and CAS1 were used to model the homologous protein TfdA. The residues believed to be involved in substrate binding, and for which variants were made in these studies, are illustrated.

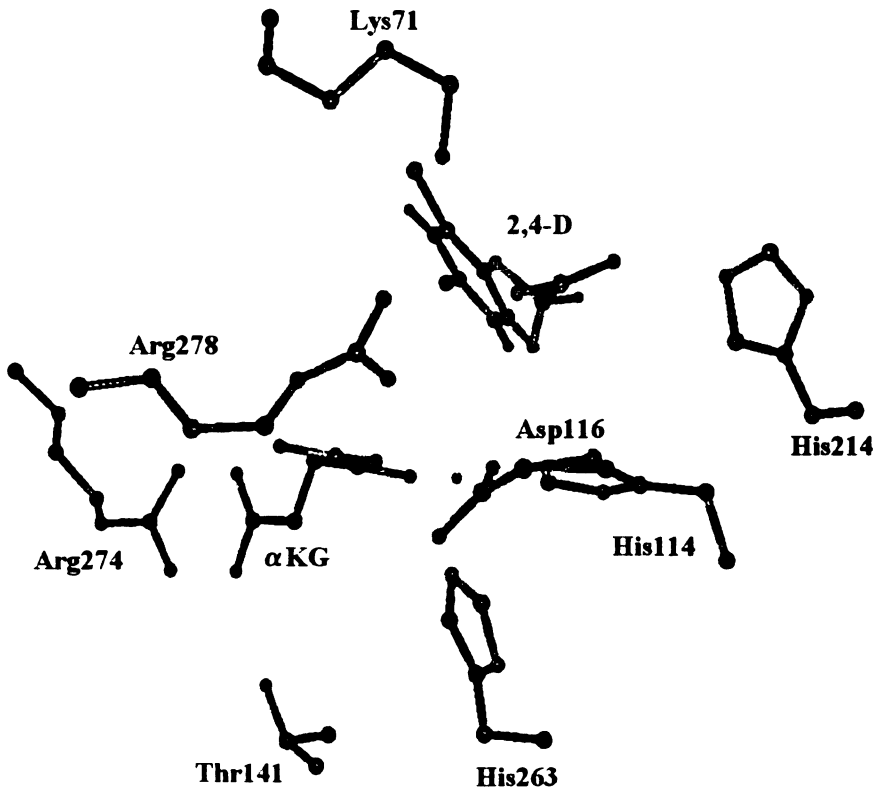


Table 4.2. 2,4-D kinetics of variant proteins.[a]

Mutant	2,4-D Range	k_{cat} (min^{-1})	K_m (μM)	k_{cat}/K_m (min^{-1} mM^{-1})	K_{si}[b] (mM)
Wt[c]	5-1000	1020 ± 90	20.0 ± 5.4	51,000	--
K71L[d,e]	100-10,000	256 ± 48	370 ± 170	692	16 ± 11
K71Q[d,e]	100-10,000	2093 ± 150	1,100 ± 300	1,900	--
R78Q	125-5,000	452 ± 21	1,500 ± 200	301	--
Y81F	5-1000	346 ± 12	33.2 ± 3.3	10,400	--
K95L[d,e]	25-1000	893 ± 41	78.5 ± 12	11,400	--
K95Q	10-1000	733 ± 22	39.5 ± 5.2	18,600	--
Y126F	5-1000	576 ± 26	31.9 ± 5.1	18,100	--
T141V	5-1000	707 ± 19	21.6 ± 2.2	32,700	--
Y169F	5-1000	444 ± 10	42.9 ± 5.1	10,300	--
Y244F	5-1000	512 ± 32	30.8 ± 4.0	16,600	--
R274Q[e,f]	250-10,000	208 ± 16	1600 ± 30	131	39 ± 18
R274L	250- 10,000	296 ± 176	2500 ± 1300	118	9.9 ± 7.4
R278Q[e,f]	250-10,000	256 ± 17	4200 ± 200	160	11 ± 8

[a]The kinetics were linear and determined at 1 mM αKG except where noted.

[b]Calculated inhibition constant for substrate inhibition when fit to Eq. 4.2.

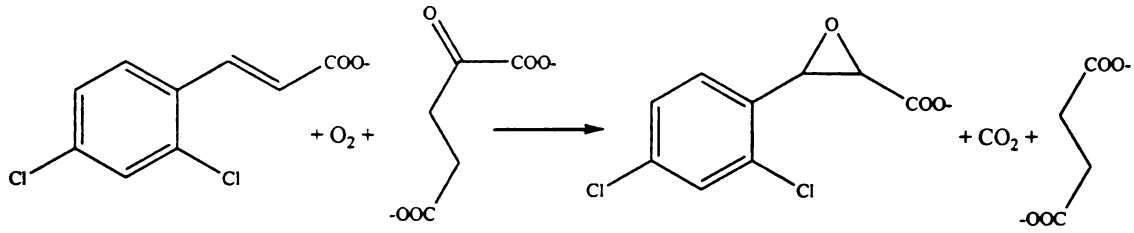
[c]Chapter 2.

[d]Determined at 10 mM αKG.

[e]Reaction progress over time was non-linear and analyzed using Eq. 4.1.

[f]Determined at 2 mM αKG.

residues, the variant proteins were incubated with the alternative substrate 2,4-dichlorocinnamic acid. This compound resembles 2,4-D, but possesses a carbon-carbon double bond in the side chain rather than an ether oxygen.



Wild-type TfdA converts 2,4-dichlorocinnamic acid to the epoxide, 2,4-dichlorophenylglycidic acid, with a K_m of 190 ± 56 μM and a k_{cat} of 52.5 ± 5.1 min^{-1} (Chapter 2). No detectable epoxide synthase activity was observed for the K71L, K71Q, and R278Q variants of TfdA in the presence of 1 mM 2,4-dichlorocinnamic acid. These results are interpreted to be due to a large increase in the K_m for this substrate similar to the situation for 2,4-D. Higher concentrations of 2,4-dichlorocinnamic acid could not be tested due to its limited solubility in water. 2,4-Dichlorocinnamic acid was oxidized by the K95Q variant with an estimated K_m > 500 μM. Furthermore, K95L TfdA had an approximate four-fold decrease in the K_m and similar k_{cat} for 2,4-dichlorocinnamic acid (K_m = 56.5 ± 14.0 and k_{cat} = 41.3 ± 3.2) when compared to wild-type enzyme. These results imply that changing the basic lysine to the more hydrophobic leucine enhances the substrate preference for the more aliphatic side chain of 2,4-dichlorocinnamic acid. Such an interpretation is consistent with Lys95 having a role in binding the ether of 2,4-D.

Lysines and arginines are expected to have multiple hydrogen-bonding partners because of their large positively charged side chains. Substitution of these residues by leucine or glutamine could reasonably have longer-range effects. One such effect observed here is the significantly increased αKG K_m and K_d of these mutants (Table 4.3),

Table 4.3. αKG kinetics of variant proteins.[a]

Mutant	Range (μM)	k_{cat} (min^{-1})	K_m (μM)	k_{cat}/K_m (min^{-1} mM^{-1})	K_d (μM)	K_{si}[b] (mM)
Wt	Look up	643 ± 44[c]	3.2 ± 0.6[c]	201,000[c]	3.4 ± 0.4	--
K71L[d,e]	2500-10^6	112 ± 10	11,700 ± 3300	9.6	2100 ± 400	--
K71Q[d,e]	2500-10^6	>1900	>20,000	--	4800 ± 300	--
R78Q	5-1,000	403 ± 45	<40	>10,100	9.9 ± 0.7	--
Y81F	5-1,000	285 ± 21	<10	>28,500	8.9 ± 0.7	--
K95L[d]	5-1,000	461 ± 51	<35	>13,200	10.0 ± 1.1	--
K95Q	5-1,000	301 ± 5	<10	>20,100	14.9 ± 2.6	--
Y126F	5-1,000	435 ± 3	<20	>21,800	14.8 ± 1.0	--
T141V	5-1,000	560 ± 48	<20	>28,000	10.9 ± 1.2	--
Y169F	5-1,000	454 ± 14	<10	>45,000	11.1 ± 2.2	--
Y244F	5,1000	512 ± 32	<10	>51,200	16.6 ± 3.7	--
R274Q[e,f]	500-10^6	263 ± 72	8200 ± 3900	32	2300 ± 200	44 ± 24
R274L	250-10^6	271 ± 68	4800 ± 2700	56	9000 ± 600	116 ± 92
R278Q[e,f]	500-10^6	224 ± 12	4100 ± 500	53	1350 ± 50	133 ± 24

[a]The kinetics were linear and determined at 1 mM 2,4-D except where noted.

[b]Calculated inhibition constant for substrate inhibition when fit to Eq. 4.2.

[c]Fukumori et al. (30).

[d]Reaction progress over time was non-linear and analyzed by using Eq. 4.1.

[e]Determined at 10 mM 2,4-D.

[f]Determined at 2 mM 2,4-D.

possibly resulting from repositioning of various active site residues, shifting of the protein backbone, and/or disruption of secondary structure elements.

Site-directed mutagenesis of proposed αKG ligands. Many of the αKG-dependent dioxygenases have a conserved arginine and threonine/serine motif, the residues of which are hypothesized to interact with the C-5 carboxylate of αKG (16, 25, 47, 80, 101, 106, 109). In TfdA, this motif is comprised of Arg274 and Thr141 (47) and is in agreement with the computer-generated structural model (Chapter 3) (Figure 4.1). To test whether these residues are critical to αKG binding, additional mutagenesis studies were carried out. The T141V, R274L, and R274Q variants of TfdA were purified and characterized (Table 4.1 and 4.2). None of the variants exhibited significant differences in iron K_d (not shown) and the substitutions had only modest effects on k_{cat}. T141V also had no significant change in its αKG K_m, αKG K_d, or 2,4-D K_m, suggesting that it has at most a minor role in binding to αKG and may merely be conserved for structural reasons. In contrast, R274L and R274Q were found to have an ~1000-fold increase in the αKG K_m and K_d (Table 4.3), confirming the critical role for Arg274 at the active site. Complimentary to the situation with variants of 2,4-D binding residues described above, the mutations involving the αKG binding residue also affected the binding of 2,4-D with a 100-fold increase in the 2,4-D K_m (Table 4.2). This is somewhat surprising since conversion of the corresponding arginine in deacetoxycephalosporin C synthase (DAOCS, and αKG dioxygenase) to a glutamine was not reported to affect the K_m of the primary substrate (41). The loss of activity arising from conversion of the corresponding arginine to glutamine in DAOCS and phytanoyl-CoA hydroxylase (another family member) can be reversed by using longer chain length α-ketoacids, a phenomenon

termed co-substrate rescue (41, 69). In particular, α-ketocaproate and α-ketoisovalerate restored activity to greater than 100% of the αKG-dependent wild-type activity in the R258Q mutant of DAOCS. Co-substrate rescue was not observed with the R274Q variant of TfdA, which had K_m values of 1.7 ± 0.5 mM and 0.70 ± 0.09 mM and k_{cat} values of 365 ± 31 min^{-1} and 22.4 ± 2.8 min^{-1} for α-ketocaproate and α-ketoisovalerate, respectively.

Arg78. In addition to mutating codons encoding the residues hypothesized from the model to be involved in substrate binding, Arg78 was converted to a glutamine in order to eliminate this protease-sensitive site in TfdA (30). No significant proteolysis product was observed for purified R78Q. Removal of this protease-sensitive site might be beneficial in future studies, such as crystallography. Although the TfdA model suggests that Arg78 is not located near the active site, the R78Q variant exhibited nearly a 100-fold increase in the 2,4-D K_m. This variant was unaffected in the iron K_d and, unlike variants of the active site 2,4-D-binding residues, did not have an increase in the αKG K_d or αKG K_m. This lack of effect on αKG binding is consistent with the residue not being in the active site. A hypothetical role of Arg78 may be to assist in 2,4-D entry into the active site. This role has also been hypothesized for His217, since the alanine variant (H217A) of this residue has an increased 2,4-D K_m but is not in the active site (47). Possibly related to these findings, the crystal structure of the αKG dioxygenase anthocyanidin synthase reveals a second binding site for the primary substrate that is proposed to help bring substrate to the active site (106).

Site-directed mutagenesis of tyrosines near the active site. Tyrosine mutants of TfdA were of interest because in TauD (a 30% identical protein) a tyrosine radical was

detected and mentioned as a possible catalytic intermediate (unpublished observations by M. J. Ryle, A. Liu, R. B. Muthukumaran, B. S. Phinney, R. Y. N. Ho, J. McCracken, L. Que, Jr., and R. P. Hausinger). In order to assess the importance of the TfdA tyrosines, the four tyrosines closest to the active site (Tyr81, Tyr126, Tyr169 and Y244) were altered to phenylalanines and characterized. All four variants were only slightly affected when compared to wild-type enzyme in terms of iron K_d, αKG K_d, αKG K_m, and 2,4-D K_m suggesting that these tryosines are not involved in catalysis or substrate binding. The Y81F, Y168F, and Y244F proteins were purified as normal, but Y126F TfdA had the unique characteristic of being thermally instable. This lability was not evident during the typical time course of the assay, but was evident in the initial purification steps at room temperature. This lack of temperature sensitivity during catalysis suggests that the presence of one or more of the substrates stabilizes the protein during activity measurements.

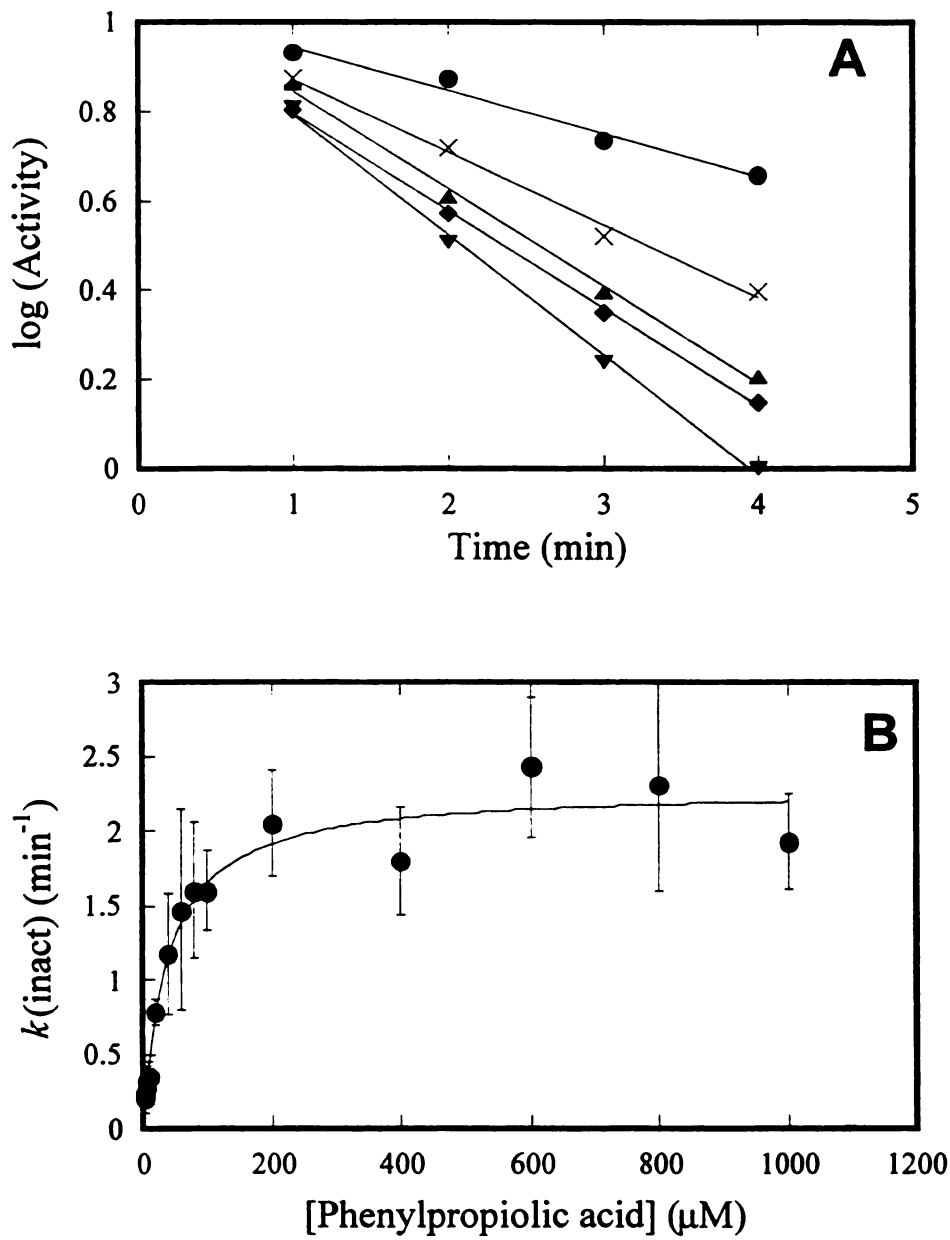
Inactivation of TfdA by PPA. In addition to examining substrate binding by site-directed mutagenesis, inactivation of TfdA by PPA, an acetylenic analogue of 2,4-D, was examined. PPA appeared to strongly inhibit or inactivate the enzyme based on the absence of oxygen consumption after adding 2,4-D to protein previously incubated with iron, αKG, ascorbate, and this compound. Kinetic analysis of enzyme inactivation revealed a first-order loss of activity (Figure 4.2, Panel A) that was dependent on the concentration of PPA and required αKG, and ferrous ions. Consistent with binding of the inactivator to the active site, high concentrations of 2,4-D protected the enzyme against inactivation. Saturation kinetics were observed for the loss of activity, allowing the calculation of $K_i = 38.1 \pm 6.0$ μM and $k(\text{inact})_{max} = 2.3 \pm 0.1$ (min^{-1}) (Figure 4.2, Panel

B). Inactivation of TfdA by PPA was not reversible by dialysis against 10 mM imidazole (pH 6.8) for 72 h, consistent with it being an irreversible, mechanism-based inactivator.

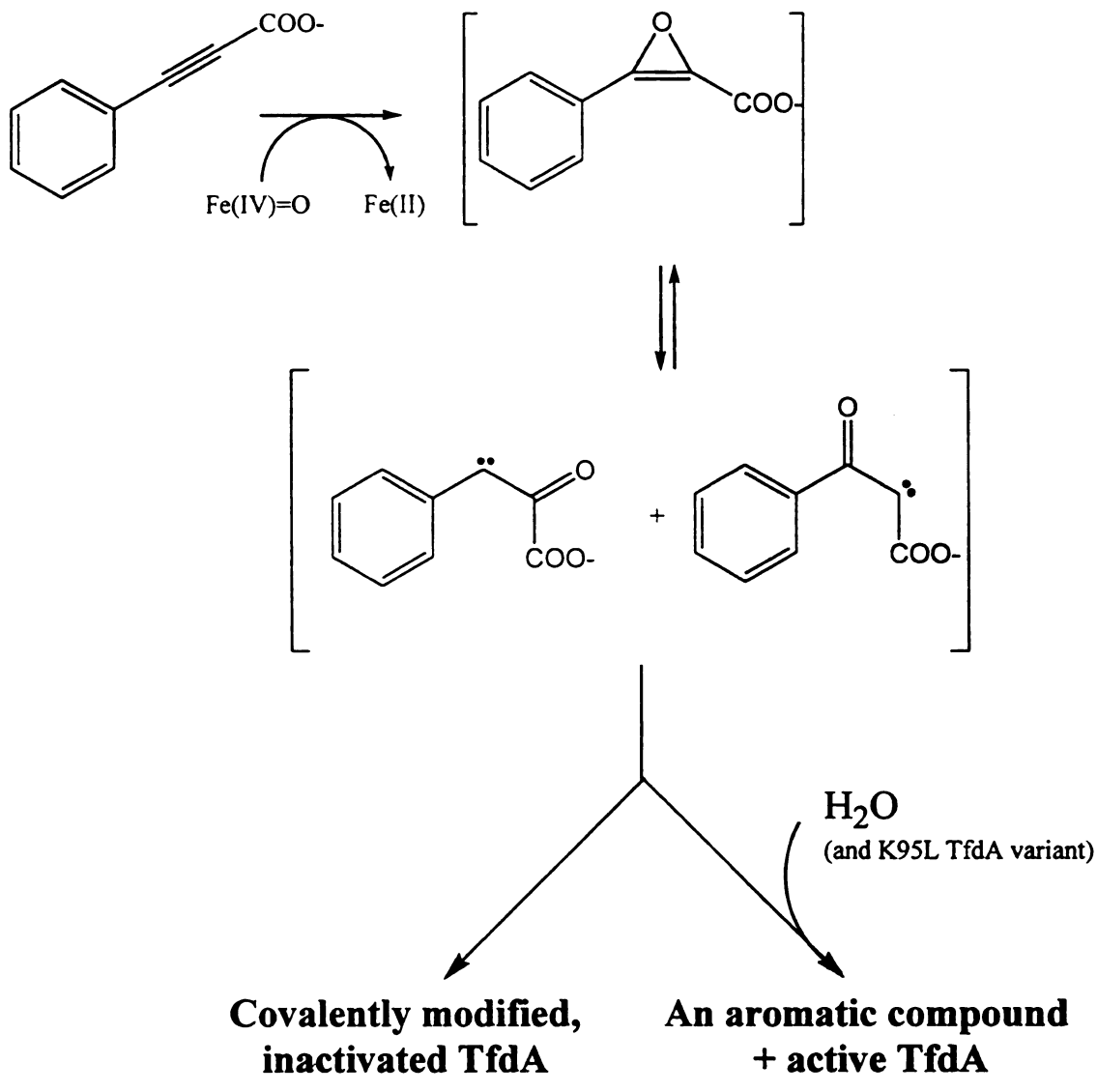
A potential mechanism for TfdA inactivation by PPA is illustrated in Scheme 4.1. This mechanism is based on that reported for thymine hydroxylase (an αKG dioxygenase) when treated with an acetylenic analogue of thymine (60, 97, 98). Thus, initial oxidation of PPA by TfdA likely produces a highly reactive oxirane intermediate that rearranges to form either of two carbene intermediates. Carbenes are known to be able to react with amino acid side chains. Thus, a covalent linkage may form between the inactivator and an active site residue. Alternatively, a carbene may react with solvent to spare the enzyme and release an aromatic product. To explore this model further, two approaches were used. The PPA-dependent inactivation kinetics of the K95L variant were analyzed and peptide studies were carried out on inactivated TfdA.

Modeling and site-directed mutagenesis studies suggest that Lys95 might be positioned near the 2,4-D ether oxygen and thus may lie close to the suspected oxirane intermediate formed from PPA acid. In order to examine the possible role of Lys95 in TfdA inactivation by this compound, the kinetics of inactivation by PPA were explored with the K95L variant. At 1 mM PPA, no inactivation of the K95L variant was observed for up to four min, while wild-type TfdA under comparable conditions was completely inactivated within one minute. The K71L, K71Q, and R278Q variants were not examined by this approach since they have significantly increased 2,4-D K_m values, and would likely bind the inactivator poorly. As a control, the Y81F variant was examined and shown to be inactivated at the same rate as wild-type enzyme. The lack of inactivation in the K95L mutant was not due to a lack of binding since PPA is a competitive inhibitor

Fig. 4.2. TfdA inactivation kinetics in the presence of PPA. Panel A shows the first-order loss of activity over time for 2 μM (●), 4 μM (X), 6 μM (♦), 8 μM (▲), and 10 μM (▼) PPA. Panel B is the plot of the individual inactivation rates versus the concentration of PPA fit to Eq. 4.4.



Scheme 4.1. Theoretical mechanism of TfdA inactivation by PPA.

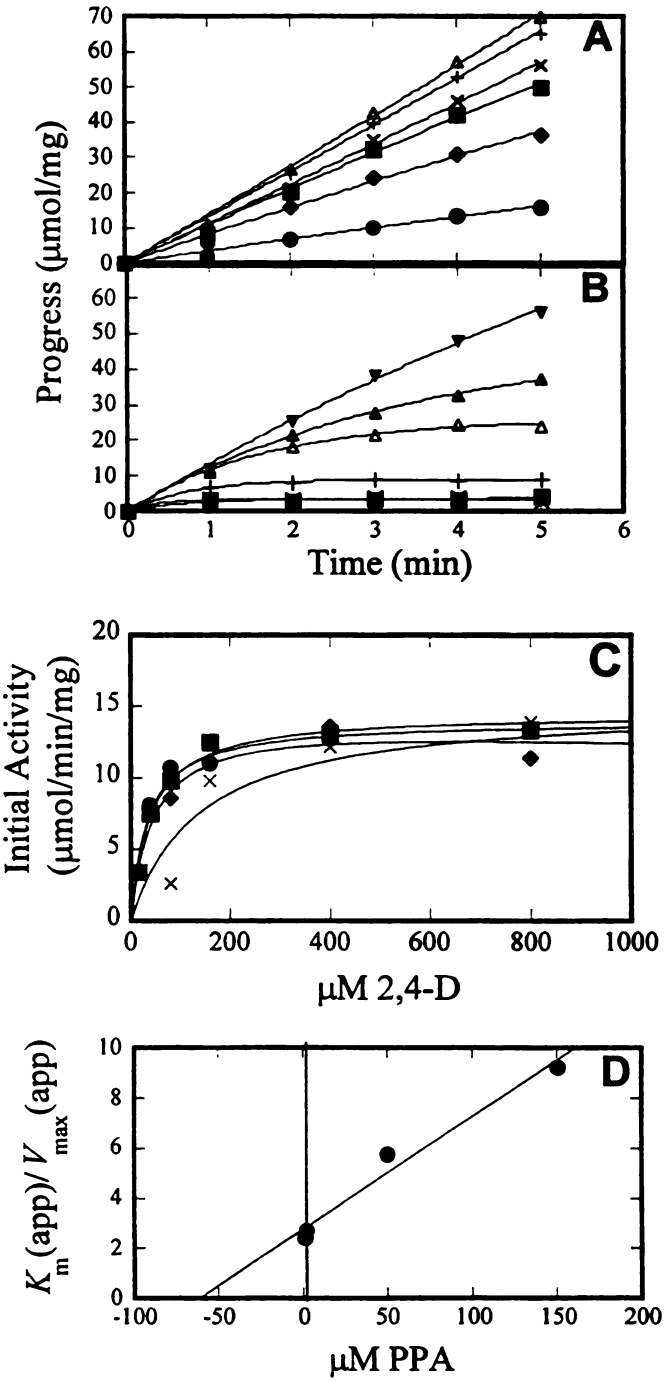


with respect to 2,4-D with a $K_i \approx 60$ μM (Figure 4.3, Panels C and D). A reasonable explanation for the observed lack of inactivation of TfdA by PPA is that any carbene intermediate that is formed reacts with water. For example, replacement of Lys95 by a leucine may result in great solvent access to the active site. If PPA is turned over by the enzyme, the reaction rate was sufficiently low to be indiscernible from background based on oxygen consumption.

In order to further characterize how wild-type TfdA is inactivated by PPA, treated and untreated samples were digested with trypsin and compared by using MALDI mass spectrometry and reverse-phase chromatography. No differences were observed between the direct analyses of digestions of modified and unmodified proteins, even though 92 % of the peptides with molecular weights between 1100 and 2600 Da were able to be identified. The peptide containing Lys95 was the one peptide in this size range that could not be identified in either digest presumably due to mass spectrometric suppression. In an extension of these studies, tryptic digests of modified and unmodified protein were chromatographed on a reverse phase column while monitoring the absorbance at 280 nm (allowing detection of peptides containing Trp, Tyr, and possibly the PPA aromatic ring). When the chromatography profiles were compared, one novel peak was observed in the digest of modified protein without a loss of intensity in any other peak (Figure 4.4, Panel A top). Each of the peaks was collected from the chromatography of the modified protein and identified by MALDI mass spectrometry (Figure 4.4, Panel B). All four Trp and all seven Tyr in TfdA were accounted for in these peptides. Notably, the peptide containing Lys95 gave a very weak MALDI signal, consistent with the suppression observed with the whole digest. Since this peptide was clearly present in peak five, Lys95 can be

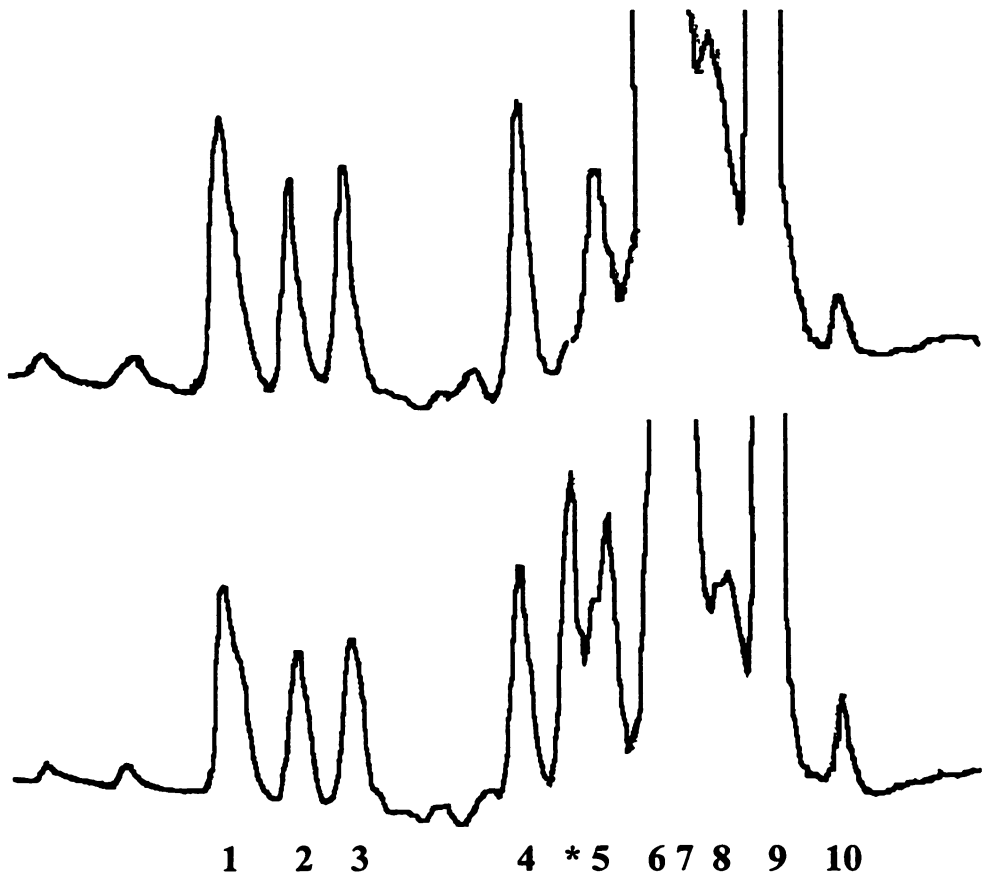
Fig. 4.3. PPA competitive inhibition kinetics for K95L TfdA. Panel A and Panel B depict progress curves of variant enzyme incubated with 2.5 μM and 150 μM PPA, respectively, with varying 2,4-D concentrations [16 μM (●), 40 μM (♦), 80 μM (■), 160 μM (X), 400 μM (+), 800 μM (Δ), 1600 μM (▲), and 4000 μM (▼)]. The initial reaction velocities of these and additional progress curve [1.25 μM (●), 2.5 μM (■), 50 μM (♦), and 150 μM (X) PPA] were plotted versus the concentration of 2,4-D (Panel C). The K_m wad deduced by replotting the apparent K_m/V_{max} as a function of PPA concentration (Panel D).

Fig. 4.3. PPA competitive inhibition kinetics for K95L TfdA.



excluded as the site of modification. Indeed, no tyrosine or tryptophan containing peptide (those absorbing at 280 nm) appears to be modified since there were no differences in the intensity or retention time of these peaks in the chromatograms. The novel A_{280}-absorbing peak that likely represents a covalent adduct of PPA and a TfdA peptide could not be identified by MALDI mass spectrometry. This result may arise from suppression effects or because the species has a mass (< 700 Da) that is obscured by the matrix features. In order to identify this peak, amino acid analysis and fast atom bombardment mass spectrometry should be performed to determine the amino acid composition and mass of this peptide, respectively. Nevertheless, the above results are consistent with PPA being a mechanism-based inactivator of TfdA that most likely becomes covalently bound to an active site peptide.

Fig. 4.4. Peptide purification and identification. Tryptic digests of PPA treated (lower trace) and untreated (upper trace) TfdA were chromatographed on a reverse phase column while monitoring at 280 nm (Panel A). Each of the peaks was collected from the chromatography of the modified protein and identified by MALDI mass spectrometry (Panel B). An asterisk indicates a peak observed only in the treated sample.



B

Peak Number	Molecular Weight	Sequence
1	1061.3	AEHYALNSR
	881.2	RYDISAR and/or YDISARR
2	714.0	EFVYR
3	877.2	AAYDALPR
4	1515.1	FLLGDTDYSEAQR
5	1595.1	YAELADISNVSLDGK
6	1394.4	NAMPPVNWPLVR
7	2589.8	EVVGNFANQLWHSDSSFQQPAAR
8	2420.2	YSMLSAVVVPPSGGDTEFCDMR
9	1505.2	WNVGDLVMWDNR
10	2234.0	Unidentified
	1525.1	Unidentified

ACKNOWLEDGEMENTS

I thank Rhonda Husain and Beverly Chamberlain for discussion about interpreting mass spectrometric data and Jon Elkins for the preparation of Figure 1.

CHAPTER 5

PHYSIOLOGICAL AND GENETIC CHARACTERIZATION OF ENVIRONMENTAL ISOLATES CONTAINING *tfdA*-like SEQUENCES

In prior studies from this laboratory, bacteria isolated from agricultural soils (specifically a culture collection derived from the MSU-Kellogg Biological Research Station long-term ecological research (LTER) plots) were examined for the presence of *tfdA*-like genes and their ability to degrade 2,4-D (46). PCR amplification of an internal fragment of *tfdA* revealed that approximately 37% of these isolates contain a *tfdA*-like gene, although none of the isolates degraded or incorporated 2,4-D (46). In this chapter, I report three sets of studies to better define the role of *tfdA*-like genes in environmental isolates.

I initiated a series of experiments to differentiate between two hypotheses that are compatible with the widespread prevalence of *tfdA*-like genes in 2,4-D non-degrading isolates. In the first hypothesis, *tfdA*-like genes encoded enzymes that degraded 2,4-D (or more generally, phenoxyacetic acid), but in the absence of 2,4-D exposure, the regulatory or downstream genes were lost and *tfdA* became a cryptic gene. Such a sequence of events would have required very rapid dissemination of the gene followed by a rapid loss of function since 2,4-D was patented and first applied to the environment in the 1940s. An alternative, and more tenable, hypothesis is that the *tfdA*-like genes present in these bacteria function in a distinct and highly ubiquitous role in soil having nothing to do with phenoxyacetate decomposition. To address these possibilities, I examined thirty-three Siberian permafrost cultures using the same procedure of PCR amplification of genomic DNA with *tfdA*-specific primers. These isolates had been frozen in the Siberian permafrost from present time to over a thousand years. In those cells believed to have been frozen over a thousand years, 2,4-D exposure is highly unlikely and detection of a *tfdA*-like gene would be indicative of an alternative function for *tfdA*.

In a second set of studies, I examined whether the *tfdA*-like gene of environmental isolates produced enzymes that might function to degrade cinnamic acid or ferulic acid. Cinnamic acid and its derivatives are abundant in nature, structurally resemble 2,4-D, and were found to be substrates of TfdA (see Chapter 2). Therefore, the degradation of cinnamic acids may explain the abundance of the *tfdA*-like gene in the environmental isolates.

In an additional study, I undertook efforts to clone a *tfdA*-like gene from an LTER isolate in order to characterize the substrate utilization profile of the recombinant protein.

EXPERIMENTAL

Media and cultures. Escherichia coli DH5α (pUS311) carrying *tfdA* on a pUC19-derived plasmid (29) was grown in LB medium. *R. eutropha* JMP134 carries the pJP4 plasmid containing *tfdABCDEF* (21). *R. eutropha* JMP228 (pBH501aE) is a derivative of *R. eutropha* JMP134 (pJP4) where the *tfdA* gene has been interrupted by transposon mutagenesis (100). These cells were grown in MMO minimal medium (91) amended with 2,4-D or succinate. The permafrost isolates were supplied by Dr. James Tiedje and were routinely cultured on 20% tryptic soy broth. The LTER isolates are a collection of environmental isolates that have been phylogenetically characterized and examined for the presence of the *tfdA* gene (46). They were cultured on MMO amended with 2,4-D or succinate, 20% TSB, or R2 (0.5 g/L yeast extract, 0.5 g proteose peptone 3, 0.5 g/L casamino acids, 0.5 g/L dextrose, 0.5 g/L starch, 0.3 g/L sodium pyruvate, 0.3 g/L potassium phosphate, 0.05 g/L magnesium sulfate).

Amplification of the tfdA-like gene. Genomic DNA was isolated from overnight cultures (8). A 360-bp internal *tfdA* fragment was amplified from 10-100 ng of genomic DNA using previously described primers (46).

Cinnamic acid and ferulic acid utilization studies. The isolates were grown by shaking at 30 °C in tubes of R2 amended with 0.5 mg/mL *trans*-cinnamic acid or ferulic acid. After 2 days, the tubes were centrifuged and the culture supernatants were analyzed by HPLC (Hewlett Packard 1050). Usage of cinnamic acid or ferulic acid was determined by analyzing the peak height at 280 nm after isocratic chromatography on a Merck Lichrosorb RP-18 column with 50:50:0.1 mixture of methanol:water:phosphoric acid.

RESULTS AND DISCUSSION

Prevalence of tfdA-like genes in Siberian permafrost samples. It has been hypothesized that the 2,4-D biodegradation pathway arose rapidly upon application of 2,4-D in the 1940s and continues to arise rapidly in pristine soils treated with 2,4-D (32). This situation requires that a form of each of the *tfd* genes be present in soils prior to herbicide application. Additionally, the encoded proteins must either use or rapidly evolve to use the xenobiotic or corresponding pathway intermediate. For the genes downstream of *tfdA* in the pJP4 plasmid, homologues have been identified that work on more "natural" substrates (35, 36, 73). In contrast, a homologue of *tfdA* functional toward a naturally occurring chemical has remained elusive. To further examine the evolution and dissemination of *tfdA*, I carried out a set of studies with Siberian soil isolates. I examined the presence of *tfdA*-like sequences in Siberian permafrost cultures using the previously described method of PCR amplification of genomic DNA. Of thirty-three permafrost culture stocks from which genomic DNA was purified, eight genomic preps (or 24%) tested positive for the presence of a *tfdA*-like gene (Table 5.1). Since some of these isolates have been frozen in the Siberian permafrost for over a thousand years prior to culturing, 2,4-D exposure with these isolates is highly unlikely. Theoretically, the permafrost samples have been isolated away from even incidental 2,4-D exposure; yet, the *tfdA* gene was prevalent in isolates from these soils at levels similar to those found in current agricultural soils. Detection of *tfdA*-like genes in isolates frozen away more than a hundred years ago is indicative of an alternative function for *tfdA*.

Table 5.1. Characteristics of the Permafrost Isolates Tested.

Strain	Estimated Age	Putative Identification	*tfdA*-like gene[1]
23-9		*Flavobacterium* sp. (76)	+ and -
26-2			-
33-1	1000K (76)	*Arthrobacter* sp. (76)	-
33-9			+
215-1	20-30K (76)		-
215-2	20-30K (76)		-
215-4	20-30K (76)		+
215-5	20-30K (76)		+
215-14	20-30K (76)		-
215-15	20-30K (76)		-
215-30			-
309-5	20-30K (76)		+ and -
309-16	20-30K (76)		-
312-12			-
342-5			-
392-1	Modern (76)		-
392-7	Modern (76)		-
1892			-
2411			-
5138-1		*Exiguobacterium* sp. (57)	+
6411			-
ED 23		*Acinetobacter* sp.	-
ED 28			+ and -
ED 37			-
EDM 6-4			+ and -

[1] A positive sign indicates the presence of an amplification product derived from a *tfdA*-like gene. Duplicate freezer stocks existed when stocks were made at different times or when a spontaneous mutation yields a new colony phenotype (a color change, a change in antibiotic susceptibility, etc). In the case of ED28, 23-9, 309-5, and EDM6-4, one culture tested positive while the other tested negative. This phenomenon may be similar to the results seen with the LTER isolates where the gene seems to disappear from the freezer stocks over time. In other duplicates tested (ED 37 and EDM 33-1), neither genomic prep contained the *tfdA*-like gene.

Assessment of cinnamic acid utilization by environmental isolates. Substituted cinnamic acids are prevalent in nature, resemble 2,4-D in structure and were recently found to be substrates of TfdA (Chapter 2). Thus, it is plausible that a *tfdA*-like gene exists in nature where it participates in degradation of a non-chlorinated compound like cinnamic acid. Incorporation of a few mutations may have converted the gene product to one that degrades the chlorinated xenobiotic 2,4-D. For example, an *E. coli* homologue of an atrazine-degrading enzyme specific for the natural compound melamine differs from the herbicide-degrading protein by only 9 of 475 amino acids (89). Similarly, directed evolution experiments using DNA shuffling or random mutagenesis have provided other examples of how a few changes in the amino acid sequence can result in dramatic changes in substrate specificity (e.g. (108)).

I grew the LTER isolates in R2 amended with cinnamic acid or ferulic acid in order to assess their ability to degrade these natural products. After analyzing media supernatants by HPLC to examine the loss of cinnamic or ferulic acid (Tables 5.2-5.4), I concluded that the presence of a *tfdA*-like gene in an LTER isolate was not predictive of its ability to degrade cinnamic acid or ferulic acid.

Cloning of a tfdA-like gene from an environmental isolate. In an effort to determine the functions of the *tfdA*-like genes in the environmental isolates, I isolated genomic DNA from four isolates (LTER 3, 8, 28, and 40) so as to clone the regions flanking the *tfdA*-like gene. Unfortunately, in the three years that have elapsed since the LTER isolates were initially examined, the freezer stocks seem to have lost the gene. Using genomic DNA preparations prepared identically to those described by Hogan et al. (46), I obtained DNA from which the 16S rDNA could be amplified, but

Table 5.2. Growth of LTER isolates on media amended with cinnamic acid.

	Isolates containing a *tfdA*-like gene	Isolates lacking a *tfdA*-like gene	Total
No cinnamic acid degradation	10	24	34
Cinnamic acid degraded	17	23	40
Total	27	47	74

χ^2=1.35; p > 0.20 and the null hypothesis cannot be rejected

Table 5.3. Growth of LTER isolates on media amended with ferulic acid.

	Isolates containing a *tfdA*-like gene	Isolates lacking a *tfdA*-like gene	Total
No ferulic acid degradation	16	37	53
Ferulic acid degraded	11	10	21
Total	27	46	74

χ^2=3.25; p > 0.20 and the null hypothesis cannot be rejected

Table 5.4. Cinnamic acid and ferulic acid degradation by LTER isolate.

LTER Isolate	Degrades cinnamic acid	Degrades ferulic acid
1	+	+
2	-	+
3	+	+
4	-	-
5	+	-
6	-	-
8	+	+
9	-	-
10	+	+
11	+	-
12	+	-
13	-	-
14	+	+
15	+	+
17	-	-
18	-	-
20	+	+
21	+	+
22	-	-
24	-	-
25	+	+
26	-	-
27	+	-
28	+	+
29	+	+
30	-	-
31	+	-
32	+	-
33	+	+
34	+	-
36	-	-
37	+	-
38	-	-
39	-	-
40	+	-
41	+	+
42	+	+
43	+	-

LTER Isolate	Degrades cinnamic acid	Degrades ferulic acid
44	+	+
47	-	-
48	+	-
52	-	-
54	-	+
57	-	-
59	+	-
61	+	-
62	+	-
63	-	-
64	-	-
65	-	+
66	-	-
67	+	+
68	-	+
69	+	-
70	+	-
71	+	-
72	-	-
73	-	+
74	-	-
75	-	-
77	+	-
78	-	-
79	-	-
81	-	-
82	ND	ND
86	-	-
88	+	-
89	+	-
91	-	-
92	+	-
94	-	-
95	-	-
97	+	-
98	+	-
100	+	-

Isolates numbered in bold contain a *tfdA*-like gene (46).

the *tfdA*-like gene could not. I varied the growth condition (e.g., using a variety of media such as TSB, MMO, and R2) in an attempt to enhance the persistence of *tfdA*, but I obtained no amplification products with the *tfdA* specific primers. In addition, I examined LTER 40 genomic DNA that was still available from 1997 and quantified the abundance of *tfdA* by a PCR titration. I observed a visible band on an ethidium bromide stained gel after PCR amplification of pUS311 (containing *tfdA*) DNA at 20 fg/μL but not at 10 fg/μL. This corresponded to 4 x 10^4 plasmids/μL. Similarly, I found that the lowest level of amplification of the LTER 40 genomic DNA required 1 ng/μL or 1 x 10^6 Mb/μL genomic DNA. Assuming a 10 Mb genome, this corresponds to 1 x 10^5 genomes/μL. Therefore, I conclude there is approximately 1 *tfdA*-like gene per 25 genome equivalents in this previously prepared DNA, and less in freshly prepared samples.

What happened to the tfdA-like genes in the LTER freezer stocks? The internal fragment of *tfdA* is no longer amplifiable from cultures grown from the LTER freezer stocks. This lack of reproducibility may be due to contamination of the original genomic preparations, to loss of the *tfdA*-like gene from the freezer stocks over time, or to loss of the cells containing the *tfdA*-like gene more rapidly in the freezer than those lacking the *tfdA*-like gene. Several lines of evidence suggest that the original amplification of the *tfdA* fragment was not due to contamination. PCR amplification was confirmed by at least one additional amplification reaction (46). If contamination occurred, one might expect contamination in batches of genomic preparations or that the contaminants would be all the same. Neither the amplification nor the hybridization profile of the *tfdA* fragments could be correlated to certain groups of genomic preparations and amplification reactions (44). Furthermore, the sequences of the fragments amplified were different. It was shown

that the PCR fragments exhibited differential hybridization to the canonical *R. eutropha* JMP134 (pJP4) *tfdA* gene. Approximately half of the genes hybridized to the *R. eutropha* JMP134 (pJP4) *tfdA* gene with high stringency. It was confirmed that these were highly homologous to the *R. eutropha* JMP134 (pJP4) *tfdA* gene by sequencing the PCR products of a select few (46). LTER 40 which hydribidized at high stringency to pJP4 was shown by sequencing to be identical to the pJP4 *tfdA* (46). Those that did not hybridize at high stringency to the pJP4 *tfdA* were either novel sequences (LTER 5 and 75), or were homologous to the *tfdA* encoded on the chromosome of *Burkholderia cepacia* RASC (LTER 1, 8, and 11) (46). Additionally, genomic DNA from several isolates was routinely obtained and amplified up to a year later (44). The results of this current study suggest that the *tfdA*-like gene may have been present in only 4% of the genomes in this DNA.

There are several examples of unstable DNA elements that might account for this loss. *Acinetobacter* strain Adp1 loses a 100-kb fragment of its chromosome that contains genes for the degradation of ferulic acid when the isolate is not grown on this carbon source (90). More commonly, plasmids and transposable elements can be lost upon culturing when selection is lacking. The genes required for 2,4-D degradation are often associated with plasmids and transposable elements (13, 75); however, such genes are not expected to be lost in freezer stocks. Perhaps related to the observed instability, loss of cell viability on antibiotic containing agar plates has been noted when freezer stocks of cells containing pET vectors are made with a concentration of glycerol greater than 10% (95). The phenomenon has not been studied in detail, but perhaps higher concentrations of glycerol inhibit the ability of the cells to grow under selection due to loss of the

plasmid between freezing and plating. The current LTER freezer stocks contain 15-20% glycerol (44); thus, a similar uncharacterized phenomenon may have occurred. The presence of the *tfdA*-like gene at low levels in the 1997 genomic preparations may be due to the freezer stocks having been subcultured three times (44). The repeated culturing of *Ralstonia* sp. strain TFD41, a 2,4-D degrading organism, is known to cause alterations in the genomic DNA (70).

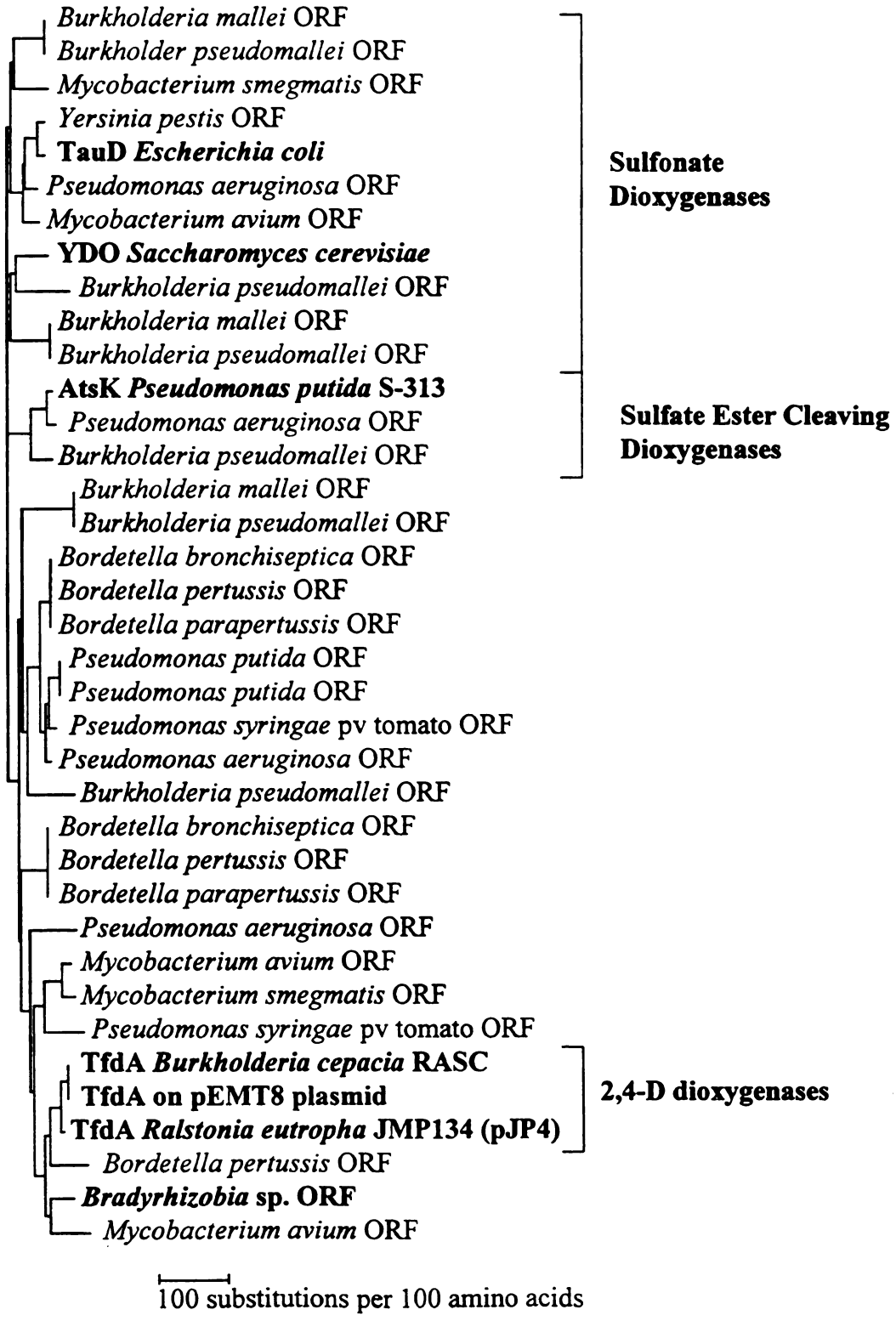
The future for the LTER isolates. The lack of stability of the *tfdA* gene confounds the interpretation of results with the isolates. I suggest that any LTER studies subsequent to the 1997 paper are suspect. The *tfdA*-like gene could be cloned from the older genomic preps using a BAC library approach, but the loss of the *tfdA*-like gene in the freezer stocks makes it impossible to go back to the isolate and do physiological studies to confirm the function of the gene. For these reasons, further studies with these environmental isolates were abandoned.

CHAPTER 6

CLONING OF A *tfdA*-like SEQUENCE FROM *Bordetella pertussis* AND PRELIMINARY CHARACTERIZATION OF THE CORRESPONDING PROTEIN

B. pertussis is a gram negative, obligate aerobic, coccobacillus, and the causative agent of pertussis (whooping cough). The genomes of *Bordetella pertussis*, *Bordetella parapertussis,* and *Bordetella bronchiseptica* have recently been sequenced and shown to contain two, two, and three open reading frames, respectively, related to *tfdA* (Figure 6.1) (1-3). The *tfdA*-like sequence at ~1.75 Mb on the *B. pertussis* chromosome encodes the most closely related protein to TfdA of any microorganism not known to metabolize 2,4-D. Although free-living *B. pertussis* is not thought to be present in nature (its only known reservoir is humans), other *Bordetella* strains have been found in the environment. Furthermore, a *Bordetella* strain isolated from enrichment cultures inoculated with activated sewage sludge has been shown to degrade 2,4-D (38). Like many characterized 2,4-D-degrading strains (i.e., *Ralstonia eutropha* JMP134 and *Burkholderia cepacia* RASC), *B. pertussis* is a β-proteobacterium. I hypothesize that the *B. pertussis tfdA*-like gene is a close relative of *tfdA*'s ancestor. Characterization of the *B. pertussis* protein may elucidate the function of proteins encoded by *tfdA*-like sequences in the environment. In order to determine the function of the *B. pertussis tfdA*-like gene, I cloned the open reading frame and studied the corresponding recombinant protein and a recombinant maltose binding protein (MBP) fusion of the protein.

Figure 6.1. Phylogenetic tree of characterized and hypothetical proteins homologous to TfdA. Cloned genes with corresponding characterized recombinant proteins are in bold.



EXPERIMENTAL

Phylogenetic tree of TfdA relatives. Protein sequences more than 30% homologous to the entire open reading frame of TfdA were identified in GenBank and the genome sequencing projects using tblastn. A rooted phylogram of these protein sequences was developed using the GrowTree program in the GCG Wisconsin Package with a blosum62 sequence comparision matrix, the Jukes-Cantor distance correction method, and the neighbor joining tree construction method.

Cloning. B. pertussis genomic DNA was graciously supplied by Trevor Stenson of Alison Weiss's lab at the University of Cincinnati. The BP1665 open reading frame (corresponding to nucleotides 1751989-1752825 on the *B. pertussis* genome) was PCR amplified from the genomic DNA using the forward primer 5'-GCTCTAGAATGACCATCACCATTACC-3' that contains an *Xba*I cleavage site 5'-extension and the reverse primer 5'-CATCAAGCTTCAGACCGGTTCCAGC-3' that contains a *Hin*dIII cleavage site 3'-extension and cloned into pMAL-c2 for expression in *Escherichia coli* DH5α. PCR was carried out in 20 mM Tris-HCl (pH 8.4), 50 mM KCl, 2.5 mM $MgCl_2$, 1.25 mM each dNTP (Gibco), 1 μM each primer, and 0.05 units/μL Gibco *Taq* DNA polymerase. Initial denaturation occurred at 94 °C for 5 min, followed by elongation for 35 cycles of denaturation for 1 min at 95 °C, annealing at 55 °C for 45 sec, and elongation at 72 °C for 45 sec, with a final 10 min elongation step at 72 °C. The resulting PCR product was cleaved with *Hin*dIII and *Xba*I and inserted into the pMAL-c2 cut with the same enzymes to create pMAL-BP1665. The resulting plasmid was sequenced with the New England BioLabs M13/pUC and *malE* sequencing primers to ensure the presence of the appropriate insert.

The *B. pertussis tfdA*-like gene was subcloned into pET24a(+) (for production of the native protein under regulation of the T7 promoter) and into pET28a(+) (to produce a His-tagged protein under regulation of the T7 promoter) and expressed in *Escherichia coli* C41 (68). The *B. pertussis tfdA*-like gene was PCR amplified from 2 pg/μL pMAL-BP1665 plasmid using as the forward primer 5'-AGGATATACATATGACCATCACCATTACC-3' that contains a 5'-extension and mutation creating a *Nde*I cleavage site and New England BioLabs M13/pUC as the reverse primer. The PCR product was cleaved with *Nde*I and *Hin*dIII and inserted into pET28a(+) and pET24a(+) cut with the same enzymes to produce pET28-BP1665 and pET24-BP1665 respectively.

Production and purification of the B. pertussis *TfdA-like protein. Escherichia coli* DH5α (pMAL-BP1665) was grown from a 1% inoculation of an overnight starter culture at 30 °C in LB medium containing 100 μg/mL ampicillin. At an A_{600} of ~0.4, the cultures were induced for two h with a final concentration of 0.4 mM IPTG, harvested by centrifugation, and suspended in 30 ml of column buffer (20 mM Tris, 1 mM EDTA, 200 mM NaCl, pH 7.8) containing 1 mM phenylmethylsulfonyl fluoride and 10 μg/mL leupeptin. Crude cell extracts were obtained by passing the cells twice through a pre-cooled French pressure cell (16,000 psi) and clarifying the debris by centrifugation (100,000 x *g* for 45 min) at 4 °C. After diluting the cell extracts to 150 mL with column buffer, the cell extracts were applied at 4 °C to a column containing New England Biolabs amylose resin (2.5 x 6 cm). The column was washed with an additional 150 mL of column buffer and the protein was removed from the column with column buffer amended with 10 mM maltose. The fractions were examined by 12% SDS-PAGE gel and

pooled accordingly. The resulting MBP-BP1665 was dialyzed against 10 mM imidazole to remove the EDTA. The protein concentration was determined by using a commercial protein assay (BioRad) with bovine serum albumin as a standard.

E. coli C41 (DE3) (pET28-BP1665) and *E. coli* C41 (DE3) (pET24-BP1665) were grown from 1% inocula of overnight starter cultures at 30 °C in LB medium containing 100 μg/mL ampicillin. At A_{600} of ~2, the cultures were induced for 2 h with a final concentration of 0.4 mM IPTG, harvested by centrifugation, and suspended in 30 ml TE buffer (20 mM Tris, 1 mM EDTA pH 7.8) containing 1 mM phenylmethylsulfonyl fluoride and 10 μg/mL leupeptin. Crude cell extracts were obtained in the same manner as already described. Native length BP1665 protein was enriched by chromatography at 4 °C on a column of DEAE-Sepharose (2.5 x 19 cm) using TE buffer and a 400 mL linear gradient to 200 mM NaCl. The enzyme eluted at approximately 100 mM NaCl. TfdA-containing fractions were dialyzed in TE buffer and chromatographed on a Mono Q column (HR10/10) (Pharmacia) at room temperature. The enzyme eluted at about 20 mM NaCl when using a 100 mL linear gradient to 100 mM NaCl in TE buffer.

Assessment of substrate specificity. The activity of the MBP-BP1665 and native length BP1665 toward potential substrates was determined at 30 °C in 10 mM imidazole buffer (pH 6.8) containing 1 mM α-ketoglutarate (αKG), 50 μM $(NH_4)_2Fe(SO_4)_2$, and 200 μM ascorbic acid by using one of two different assays. For all methods, stock solutions of $(NH_4)_2Fe(SO_4)_2$ and ascorbic acid were made fresh prior to each set of experiments. The capability to degrade phenoxyacetate and 2,4-D was tested by using the previously described 4-aminoantipyrene spectrophotometric assay (30). The release of sulfite from the sulfonates taurine, isethionate, cysteate, and taurocholate was examined

by using the previously described Ellman's reagent assay (24). The activity toward other potential substrates was determined by monitoring the rates of oxygen consumption at 30 °C with a YSI model 5300 biological oxygen monitor as previously described (Chapter 2).

Molecular weight determination. The molecular weights of the MBP-BP1665 and native length BP1665 were estimated by gel filtration on a Pharmacia HR 10/30 Superose 6 column in the above column buffer. Calibration was carried out by using gel filtration standards (BioRad) containing thyroglobulin (670 kDa), bovine γ-globulin (158 kDa), chicken ovalbumin (45 kDa), equine myoglobin (17 kDa) and vitamin B-12 (1.35 kDa).

UV/vis spectroscopy. Detection of a charge-transfer transition characteristic of αKG dioxygenases was determined by the method of Ryle et al. (84). The UV/vis spectrum was taken of a degassed solution of 190-250 μM protein and 1 mM αKG under nitrogen. Degassed ferrous ammonium sulfate under nitrogen was added to a concentration equal to the protein concentration.

RESULTS AND DISCUSSION

Genome sequencing has revealed a preponderance of uncharacterized αKG dioxygenases. For example, the genome of *Arabidopsis thaliana* contains over sixty-four open reading frames that show strong similarity to known αKG dioxygenases and less than 20% have known functions (78). To roughly assess the prevalence of αKG dioxygenases related to TfdA, BLAST was used to identify proteins in the genome sequencing projects with >30% identity to the entirety of TfdA. Although the genome sequencing projects searched covered bacteria from most of the phylogenetic branches of the eubacteria, homologues of TfdA were only found in the *Mycobacteria* in the Gram positives, the *Bordetella* and *Burkholderia* in the β-proteobacteria, and the *Pseudomonas* and *Enterobacteriaceae* in the γ-proteobacteria.

Upon examination of a phylogenetic tree made from these protein sequences (Figure 6.1), only eleven of the thirty uncharacterized open reading frames (ORFs) could be grouped with characterized proteins. All three known 2,4-D/αKG dioxygenases grouped together. Two uncharacterized ORFs formed a clade with AtsK, a *Pseudomonas putida* S-313 sulfate-ester cleaving/αKG dioxygenase (103). Nine uncharacterized ORFs grouped with the sulfonate/αKG dioxygenases of *Escherichia coli* (24) and *Saccharomyces cerevisiae* (45). Interestingly, *Burkholderia pseudomallei* and *Burkholderia mallei* have three and two ORFs respectively clustering with the sulfonate/αKG dioxygenases suggesting that these organisms widely utilize sulfonates for growth.

Numerous ORFs seem to be strictly conserved among phylogenetically related organisms. Three ORFs are similar when compared between the species in *B. mallei* and

B. pseudomallei. Likewise, *B. pertussis, B. parapertussis, and B. bronchiseptica* contain two ORFs that are very similar in these genera. Intriguingly, *B. pertussis* has an open reading frame (BP1665) that is distinct from the open reading frames of the other *Bordetella* and that is most closely related to the 2,4-D/αKG dioxygenases. In order to determine if this open reading frame produces an enzyme that could use 2,4-D and to explore the ancestral substrate specificity of TfdA, this open reading frame was cloned for the production of recombinant protein.

MBP-BP1665 was produced from the pMAL-BP1665 plasmid and purified by using the amylose resin. No activity was detected for this protein when assayed with phenoxyacetates, sulfonates, benzofuran-2-carboxylic acid, chromone-2-carboxylic acid, cinnamate, hippurate, indole-3-acetic acid, indole-3-acrylic acid, indole-2-carboxylic acid, indole-3-carboxylic acid, indole-3-pyruvate, 2-naphthoxyacetic acid, PPA, butyrate, sodium dodecylsulfate, 2-hydroxycinnamate, 3-hydroxycinnamate, 4-hydroxycinnamate, 2-methoxycinnamate, 3-methoxycinnamate, 4-methoxycinnamate, ferulate, 3,5-dimethoxycinnamate, 4-aminoethylphosphonic acid, thymine, isethionate, DL-alanine, L-cysteine, L-aspartate, L-glutamate, L-phenylalanine, L-glycine, L-histidine, L-isoleucine, L-lysine. L-leucine, L-methionine, L-asparagine, L-proline, L-glutamine, L-serine, L-threonine, L-valine, L-tryptophan, L-tyrosine, or phosphonoformate.

Gel filtration chromatography followed by electrophoresis of selected fractions on an SDS-PAGE gel indicated that MBP-BP1665 was mainly present as high molecular weight multimers (>4,000 kDa) that eluted in the column void volume. TauD and TfdA, other closely related αKG dioxygenases, are dimers (24, 30), and it is presumed that the BP1665 should be a dimer as well. I conclude that the high molecular weight multimers

are non-native aggregates, perhaps analogous to a previously described state of two inactive mutants of TfdA (47). In those cases, the non-fusion H245A and H262A mutants of TfdA formed inclusion bodies while the corresponding maltose-binding protein fusions were soluble but formed high molecular weight multimers (47).

Additionally, MBP-BP1665 did not develop a charge-transfer transition characteristic of αKG dioxygenases. In the absence of oxygen and the presence of αKG and ferrous ions, TfdA and TauD have a characteristic metal-to-ligand charge-transfer transition at 500-600 nm arising from the bidentate binding of αKG to iron (43, 84). Lack of this charge-transfer transition is consistent with mis-folded protein that cannot bind substrate. I conclude that overproduced MBP-BP1665 did not fold properly in *E. coli.*

The *B. pertussis tfdA*-like gene was further subcloned into pET28 and pET24. The His-tagged fusion of the BP1665 protein expressed from pET28 was insoluble and formed inclusion bodies. The non-fusion version of BP1665 expressed from pET24 was mostly insoluble, but the small proportion of soluble protein was able to be purified from the cell lysates. No activity was detected for this protein when assayed with phenoxyacetates or sulfonates. Gel filtration chromatography followed by electrophoresis of selected fractions on an SDS-PAGE gel indicated that the protein was mainly present as monomers (27 kDa). Similarly, this protein did not develop the metal-to-ligand charge-transfer transition at 500-600 nm. I conclude that the overproduced non-fusion BP1665 protein also does not fold properly in *E. coli.*

CHAPTER 7

CONCLUSIONS AND FUTURE DIRECTIONS

The goal of this work was to determine what governs substrate specificity in TfdA and related enzymes. 2,4-D is the preferred substrate for TfdA followed by 2-naphthoxyacetic acid, benzofuran-2-carboxylic acid, 2,4-dichlorocinnamic acid, 2-chlorocinnamic acid, 1-naphthoxyacetic acid, and 4-chlorocinnamic acid. A generic structure of a substrate for TfdA might be an aromatic ring with a three atom side chain in an extended conformation and ending in a carboxylate. Compounds with a terminal sulfonate or phosphonate were not found to be substrates.

The latter characteristic can be understood upon examination of the TfdA structural model. In the model of TfdA, the two carboxylate oxygens are coordinated to Arg278, Lys95, His214, and the backbone amide of Ser117. Kinetic analyses of variants of these residues were consistent with these roles. In contrast, the ~30% identical TauD selects for a sulfonate-containing substrate through coordination of the three sulfonate oxygens to arginine, histidine, and a backbone amide. Additional insights into the specificity of TfdA for binding 2,4-D were obtained by site-directed mutagenesis of Lys95 and inactivation of the enzyme by PPA. The results are consistent with Lys95 being hydrogen bonded to the ether of 2,4-D, partly explaining the preference of TfdA for 2,4-D over 2,4-dichlorocinnamic acid. Residues involved in interacting with the ring of 2,4-D or with the ring chlorines were not identified and could be the focus of further research. In the TfdA structural model, no amino acid side chains interact with the ring or these chlorines, but residues participating in such interactions may be present in a protein loop that was omitted from the structure. In order to further explore the role of these omitted residues, alanine-scanning mutagenesis could be carried out and/or a crystal structure obtained. TfdA samples were provided to the crystallographer who deduced the

TauD crystal structure, however a crystal structure of TfdA is unavailable because thus far the crystals have been amorphous. At least some of the crystallography problems are likely due to a proteolysis site in TfdA that now has been eliminated in the R78Q variant, so further efforts with this variant should be continued. Although analysis of substrate interactions may be complicated in this variant because of its high 2,4-D K_m, it would be worthwhile especially since Arg78 is not expected to be an actual 2,4-D ligand.

Based on the TfdA structural model, Arg274 and Thr141 were identified as possible αKG ligands. Site-directed mutagenesis data are consistent with Thr141 having only a minor role, if any, and Arg274 having a major role in binding αKG. Importantly, the activity of the R274Q variant did not return to wild-type levels when using other α-keto acids (termed co-substrate rescue), highlighting the differences among the members of the αKG dioxygenase superfamily. These differences between TfdA and other αKG dioxygenase family members may be important when trying to understand the trade-offs made when a protein evolves a new substrate specificity.

The intrinsic tryptophan fluorescence of the protein was used to develop a binding assay for αKG and Fe(II). In addition, each of the four tryptophan residues were altered in an effort to identify the fluorescent reporter in TfdA. Even though Trp256 was the most likely candidate due to its close proximity to the active site, it was excluded as the reporter group. The reporter is expected to be some combination of Trp113 and Trp248. Further work should be carried out to determine the role of Trp113 in the protein. This tryptophan has been identified as becoming hydroxylated when protein containing αKG, but not substrate, is incubated in the presence of oxygen for thirty minutes. Additionally, in the studies described here, variants of Trp113 could not be expressed suggesting an

important role for this tryptophan. The hydroxylation of Trp113 has been hypothesized to serve a protective role for the enzyme by preventing the occurrence of more damaging oxidative reactions such as those resulting in cleavage of the peptide backbone. The lack of this protective side chain in the W113F and W113L variants may lead to their rapid oxidative degradation in the cytoplasm. Alternatively, *in vivo* hydroxylation of Trp113 may signal a lack of substrate and lead to degradation of the protein. Further work on Trp113 could involve fluorescence techniques since the characteristics of hydroxytryptophan differ from that of tryptophan. Additionally, examination of the W113F and W113L variants in a different expression system may allow for expression of the variants.

In order to examine the evolution of substrate specificity of a *tfdA* gene product, I sought to clone a *tfdA*-like gene from environmental organisms isolated from modern agricultural soil. Unfortunately, *tfdA*-like genes proved to be unstable in the available soil isolates and this approach was abandoned. With projects like the meta-genome project (81), where DNA from soil is being cloned and sequenced, the sequence of an environmental *tfdA*-like gene is likely to be determined. This information will facilitate the cloning of the gene and characterization of the encoded protein. Although my preliminary efforts to evolve TfdA-like activity were unsuccessful, I did develop plate assay methods to detect formation of a protein that preferably hydroxylates 2,4-D. As an alternative direction to the evolution work, I showed that *tfdA*-like genes are present in permafrost soil samples with an estimated age greater than 10,000 years. These results suggest that *tfdA*-like genes have not appeared in response to 2,4-D exposure; rather, they are prevalent in the environment and of ancient ancestry.

In an effort to further evaluate the evolution of substrate specificity of a *tfdA* gene product, the *Bordetella pertussis tfdA*-like gene was cloned and overexpressed, and the protein characterized. Unfortunately, enzyme activity was not detected for any potential substrate examined. The *B. pertussis tfdA*-like gene may be of further interest to study since it is not found in the initial shotgun sequencing of *Bordetella parapertussis* or *Bordetella bronchiseptica.* If these results hold up, this might suggest that the target gene encodes one of several proteins that differentiates *Bordetella pertussis* from the other *Bordetella* sp. Further information about the function of this ORF may be provided in microarray work when *B. pertussis* arrays become available. If this protein is involved in pathogenesis, its function should be further explored after expression in alternative systems such as yeast or a high GC organism.

When I began this research, a homologue to TfdA had just been identified, nothing was known about the specific role of any amino acid in TfdA, and little was known about the substrate specificity of TfdA or the residues involved in governing that substrate specificity. Four years later, numerous alternative substrates of TfdA have been identified that may be of potential biosynthetic importance (Chapter 2), a model for the structure of TfdA has been obtained based on the crystal structure coordinates of a ~30% identical homologue (Chapter 3), and this model has been used to identify potential substrate ligands which were confirmed by site-directed mutagenesis techniques (Chapter 4).

REFERENCES

1. The *Bordetella pertussis* genome sequence data was produced by the Pathogen Sequencing Unit at the Sanger Centre, and can be accessed from http://www.sanger.ac.uk/Projects/B_pertussis.

2. The *Bordetella parapertussis* genome sequence data was produced by the Pathogen Sequencing Unit at the Sanger Centre.

3. The *Bordetella bronchiseptica* genome sequence data was produced by the Pathogen Sequencing Unit at the Sanger Centre.

4. Biologic and economic benefits from the use of phenoxy herbicides in the United States. NAPIAP 1-PA-96. U. S. Department of Agriculture.

5. Aravind, L., and E. V. Koonin. 2001. The DNA-repair protein AlkB, EGL-9, and leprecan define new familes of 2-oxoglutarate- and iron-dependent dioxygenases. Genome Biol. 2(3):research0007.1-0007.8.

6. Archelas, A., and R. Furstoss. 1997. Synthesis of enantiopure epoxides through biocatalytic approaches. Annu. Rev. Microbiol. 51:491-525.

7. Audus, L. J. 1952. The decomposition of 2,4-dichlorophenoxyacetic acid and 2-methyl-4-chlorophenoxyacetic acid in the soil. J. Sci. Food Agric. 3:268.

8. Ausubel, F., R. Brent, R. Kingston, D. Moore, J. Seidman, I. Smith, and K. Struhl. 1987. Current protocols in molecular biology. John Wiley and Sons, New York.

9. Baggaley, K. H., A. G. Brown, and C. J. Schofield. 1997. Chemistry and biosynthesis of clavulanic acid and other clavams. Nat. Prod. Rep. 14(4):309-333.

10. Baldwin, J. E. 1990. The biosynthesis of penicillins and cephalosporins. J. Heterocycl. Chem 27:71-78.

11. Baldwin, J. E., and E. Abraham. 1988. The biosynthesis of penicillins and cephalosporins. Nat. Prod. Rep. 5(2):129-145.

12. Bell, G. R. 1957. Some morphological and biochemical characteristics of a soil bacterium which decomposes 2,4-dichlorophenoxyacetic acid. Can. J. Microbiol. 3:821-840.

13. Bulinski, D. A., and C. H. Nakatsu. 1998. Involvement of a transposon in the dissemination of 2,4-D catabolic genes. 8th International Symposium on Microbial Ecology, 9-14 August, Halifax, N.S.:115.

14. Burzlaff, N. I., P. J. Rutledge, I. J. Clifton, C. M. H. Hensgens, M. Pickford, R. M. Adlington, P. L. Roach, and J. E. Baldwin. 1999. The reaction cycle of isopenicillin N synthase observed by X-ray diffraction. Nature 401(6754):721-724.

15. Cadwell, R. C., and G. F. Joyce. 1992. Randomization of genes by PCR mutagenesis. PCR Methods and Applications 2:28-33.

16. Clifton, I. J., L. C. Hsueh, J. E. Baldwin, K. Harlos, and C. J. Schofield. 2001. Structure of proline 3-hydroxylase - Evolution of the family of 2-oxoglutarate dependent oxygenases. Eur J Biochem 268(24):6625-6636.

17. Denis, J.-N., A. E. Greene, A. A. Serra, and M.-J. Luche. 1986. An efficient, enantioselective synthesis of the taxol side chain. J. Org. Chem. 1986(51):46-50.

18. DiGennaro, P., A. Colmegna, E. Galli, G. Sello, F. Pelizzoni, and G. Bestetti. 1999. A new biocatalyst for production of optically pure aryl epoxides by styrene monooxygenase from *Pseudomonas fluorescens* ST. Appl. Environ. Microbiol. 65(6):2794-2797.

19. DiGiovanni, G. D., J. W. Neilson, I. L. Pepper, and N. A. Sinclair. 1996. Gene transfer of *Alcaligenes eutrophus* JMP134 plasmid pJP4 to indigenous soil recipients. Appl. Environ. Microbiol. 62(7):2521-2526.

20. Ditzelmuller, G., M. Loidl, and F. Streichsbier. 1989. Isolation and characterization of a 2,4-dichlorophenoxyacetic acid-degrading soil bacterium. Appl. Microbiol. Biotechnol. 31:93-96.

21. Don, R. H., and J. M. Pemberton. 1981. Properties of six pesticide degradation plasmids isolated from *Alcaligenes paradoxus* and *Alcaligenes eutrophus*. J. Bacteriol. 145(2):681-686.

22. Eftink, M. R. 1997. Fluorescence methods for studying equilibrium macromolecule-ligand interactions, p. 221-257. *In* L. Brand and M. L. Johnson (ed.), Methods in Enzymology: Fluorescence spectroscopy, vol. 278. Academic Press, New York.

23. Ege, S. N. 1994. Organic chemistry: Structure and reactivity, 3rd ed. D. C. Heath and Company, Toronto.

24. Eichhorn, E., J. R. van der Ploeg, M. A. Kertesz, and T. Leisinger. 1997. Characterization of α-ketoglutarate-dependent taurine dioxygenase from *Escherichia coli*. J. Biol. Chem. 272(37):23031-23036.

25. Elkins, J. M., M. J. Ryle, I. J. Clifton, J. C. Dunning Hotopp, J. S. Lloyd, N. I. Burzlaff, J. E. Baldwin, R. P. Hausinger, and P. L. Roach. 2002. X-Ray crystal structure of *Escherichia coli* taurine/α-ketoglutarate dioxygenase complexed to ferrous iron and substrates. Biochemistry 41(16):5185-5192.

26. Evans, W. C., B. S. W. Smith, H. N. Fernley, and J. I. Davies. 1971. Bacterial metabolism of 2,4-dichlorophenoxyacetate. Biochem. J. 122:543-551.

27. Faulkner, J. K., and D. Woodcock. 1964. Metabolism of 2,4-dichlorophenoxyacetic acid ('2,4-D') by *Aspergillus niger* van Tiegh. Nature(4947):865.

28. Feig, A. L., and S. J. Lippard. 1994. Reactions of non-heme iron(II) centers with dioxygen in biology and chemistry. Chem. Rev. 94:759-805.

29. Fukumori, F., and R. P. Hausinger. 1993. *Alcaligenes eutrophus* JMP134 "2,4-dichlorophenoxyacetate monooxygenase" is an α-ketoglutarate-dependent dioxygenase. J. Bacteriol. 175(7):2083-2086.

30. Fukumori, F., and R. P. Hausinger. 1993. Purification and characterization of 2,4-dichlorophenoxyacetate/α-ketoglutarate dioxygenase. J. Biol. Chem. 268(32):24311-24317.

31. Fulthorpe, R. R., C. McGowan, O. V. Maltseva, W. E. Holben, and J. M. Tiedje. 1995. 2,4-Dichlorophenoxyacetic acid-degrading bacteria contain mosaics of catabolic genes. Appl. Environ. Microbiol. 61(9):3274-3281.

32. Fulthorpe, R. R., A. N. Rhodes, and J. M. Tiedje. 1996. Pristine soils mineralize 3-chlorobenzoate and 2,4-dichlorophenoxyacetate via different microbial populations. Appl. Environ. Microbiol. 62(4):1159-1166.

33. Garraway, J. L., and R. L. Wain. 1976. The design of auxin-type herbicides. Drug Design 7:115-164.

34. Gentile, A., and C. Giordano. 1992. The enzymatic preparation of (2*R*, 3*S*) phenyl glycidic acid Esters. J. Org. Chem. 57:6635-6637.

35. Ghosal, D., and I.-S. You. 1988. Nucleotide homology and organization of chlorocatechol genes of plasmids pJP4 and pAC27. Mol. Gen. Genet. 211:113-120.

36. Ghosal, D., and I.-S. You. 1989. Operon structure and nucleotide homology of the chlorocatechol oxidation genes of plasmids pJP4 and pAC27. Gene 83:225-232.

37. Ghosal, D., I.-S. You, D. K. Chatterjee, and A. M. Chakrabarty. 1985. Genes specifying degradation of 3-chlorobenzoic acid in plasmids pAC27 and pJP4. Proc. Natl. Acad. Sci., USA 82:1638-1642.

38. Greer, L., J. Robinson, and D. Shelton. 1992. Kinetic comparison of seven strains of 2,4-dichlorophenoxyacetic acid-degrading bacteria. Appl. Environ. Microbiol. 58:1027-1030.

39. Gribble, G. W. 1992. Naturally occurring organohalogen compounds--A survey. J. Nat. Prod. 55(10):1353-1395.

40. Hausinger, R. P., F. Fukumori, D. A. Hogan, T. M. Sassanella, Y. Kamagata, H. Takami, and R. E. Saari. 1997. Biochemistry of 2,4-dichlorophenoxyacetic acid (2,4-D) degradation: Evolutionary implications, p. 35-51. *In* K. Horikoshi, M. Fukuda, and T. Kudo (ed.), Microbial Diversity and Genetics of Biodegradation. Karger, New York.

41. Hee, H.-J., M. D. Lloyd, I. J. Clifton, K. Harlos, A. Dubus, J. E. Baldwin, J.-M. Frere, and C. J. Schofield. 2001. Alteration of the co-substrate selectivity of deacetoxycephalosporin G synthase--The role of arginine 258. J. Biol. Chem. 276:18290-18295.

42. Hegg, E. L., and L. Que. 1997. The 2-his-1-carboxylate facial triad: an emerging structural motif in mononuclear non-heme iron (II) enzymes. Eur. J. Biochem. 250:625-629.

43. Hegg, E. L., A. K. Whiting, R. E. Saari, J. McCracken, R. P. Hausinger, and L. Que, Jr. 1999. Herbicide-degrading α-keto acid-dependent enzyme TfdA: metal coordination environment and mechanistic insights. Biochemistry 38:16714-16726.

44. Hogan, D. A. 2001. Personal communication.

45. Hogan, D. A., T. A. Auchtung, and R. P. Hausinger. 1999. Cloning and characterization of a sulfonate/α-ketoglutarate dioxygenase from *Saccharomyces cerevisiae*. J. Bacteriol. 181:5876-5879.

46. Hogan, D. A., D. H. Buckley, C. H. Nakatsu, T. M. Schmidt, and R. P. Hausinger. 1997. Distribution of the *tfdA* gene in soil bacteria that do not degrade 2,4-dichlorophenoxyacetic acid (2,4-D). Microb. Ecol. 34:90-96.

47. Hogan, D. A., S. Smith, E. Saari, J. McCracken, and R. P. Hausinger. 2000. Site-directed mutagenesis of 2,4-dichlorophenoxyacetic acid/α-ketoglutarate dioxygenase. J. Biol. Chem. 275(17):12400-12409.

48. Itoh, K. 2001. Personal communication.

49. Jones, F. D. May 4, 1945. U. S. patent 2,390,941.

50. Ka, J. O., W. E. Holben, and J. M. Tiedje. 1994. Analysis of competition in soil among 2,4-dichlorophenoxyacetic acid-degrading bacteria. Appl. Environ. Microbiol. 60:1121-1128.

51. Ka, J. O., W. E. Holben, and J. M. Tiedje. 1994. Genetic and phenotypic diversity of 2,4-dichlorophenoxyacetic acid (2,4-D)-degrading bacteria isolated from 2,4-D-treated field soils. Appl. Environ. Microbiol. 60:1106-1115.

52. Ka, J. O., W. E. Holben, and J. M. Tiedje. 1994. Use of gene probes to aid in recovery and identification of functionally dominant 2,4-dichlorophenoxyacetic acid-degrading populations in soil. Appl. Environ. Microbiol. 60:1116-1120.

53. Ka, J. O., and J. M. Tiedje. 1994. Integration and excision of a 2,4-dichlorophenoxyacetic acid-degradative plasmid in *Alcaligenes paradoxus* and evidence of its natural intergeneric transfer. J. Bacteriol. 176(17):5284-5289.

54. Kahnert, A., and M. A. Kertesz. 2000. Characterization of a sulfur-regulated oxygenative alkylsulfatatse from *Pseudomonas putida* S-313. J. Biol. Chem. 275(41):31661-31667.

55. Kamagata, Y., R. R. Fulthorpe, K. Tamura, H. Takami, L. J. Forney, and J. M. Tiedje. 1997. Pristine environments harbor a new group of oligotrophic 2,4-dichlrophenoxyacetic acid-degrading bacteria. Appl. Environ. Microbiol. 63(6):2266-2272.

56. Katase, T., and R. Kondo. 1989. Vertical profiles of *trans*- and *cis*-4-hydroxycinnamic acids and other phenolic acids in Horonobe peat soils, Japan. Soil Science 148(4):258-264.

57. Kathariou, S. 2001. Personal communication.

58. Kilpi, S. 1980. Degradation of some phenoxyacid herbicides by mixed cultures of bacteria isolated from soil treated with 2-(2-methyl-4-chloro)phenoxypropionic acid. Microb. Ecol 6:261-270.

59. King, R. J., K. A. Short, and R. J. Seidler. 1991. Assay for detection and enumeration of genetically engineered microorganisms which is based on the activity of a deregulated 2,4-dichlorophenoxyacetate monooxygenase. Appl. Environ. Microbiol. 27(6):1790-1792.

60. Lai, M.-t., W. Wu, and J. Stubbe. 1995. Characterization of a novel, stable norcaradiene adduct resulting from the inactivation of thymine hydroxylase y 5-ethynyluracil. J. Am. Chem. Soc. 117:5023-5030.

61. Leveau, J. H. J., A. J. B. Zehnder, and J. R. van der Meer. 1998. The *tfdK* gene product facilitates uptake of 2,4-dichlorophenoxyacetate by *Ralstonia eutropha* JMP134(pJP4). J. Bacteriol. 180(8):2237-2243.

62. Liu, A., R. Y. N. Ho, and L. Que, Jr. 2001. Alternative reactivity of an α-ketoglutarate-dependent iron(II) oxygenase: enzyme self-hydroxylation. J. Am. Chem. Soc. 123:5126-5127.

63. Lloyd, M. D., K. D. Merritt, V. Lee, T. J. Sewell, B. Wha-Son, J. E. Baldwin, C. J. Schofield, S. W. Elson, K. H. Baggaley, and N. H. Nicholson. 1999. Product-substrate engineering by bacteria: Studies on clavaminate synthase, a trifunctional dioxygenase. Tetrahedron 55(33):10201-10220.

64. Lontz, J. F. February 20, 1942. U. S. patent 2,322,761.

65. Loos, M. A., R. N. Roberts, and M. Alexander. 1967. Phenols as intermediates in the decomposition of phenoxyacetates by an *Arthrobacter* species. Can. J. Microbiol. 13:679-690.

66. Mäe, A. A., R. O. Marits, N. R. Ausmees, V. M. Kôiv, and A. L. Heinaru. 1993. Characterization of a new 2,4-dichlorophenoxyacetic acid degrading plasmid

pEST4011: Physical map and localization of catabolic genes. J. Gen. Microbiol. 139:3165-3170.

67. Mai, P., O. S. Jacobsen, and J. Aamand. 2001. Mineralization and co-metabolic degradation of phenoxyalkanoic acid herbicides by a pure bacterial culture isolated from an aquifer. Appl. Microbiol. Biotechnol. 56:486-490.

68. Miroux, B., and J. E. Walker. 1996. Over-production of proteins in *Escherichia coli*: mutant hosts that allow synthesis of some membrane proteins and globular proteins at high levels. J. Mol. Biol. 260:289-298.

69. Mukherji, M., W. Chien, N. J. Kershaw, I. J. Clifton, C. J. Schofield, A. S. Wierzbicki, and M. D. Lloyd. 2001. Structure-function analysis of phytanoyl-CoA 2-hydroxylase mutations causing Refsum's disease. Hum Mol Genet 10(18):1971-1982.

70. Nakatsu, C. H., R. Korona, R. E. Lenski, F. J. de Bruijn, T. L. Marsh, and L. J. Forney. 1998. Parallel and divergent genotypic evolution in experimental populations of *Ralstonia* sp. J. Bacteriol. 180(17):4325-4331.

71. Nakatsu, C. H., R. Korona, R. E. Lenski, F. J. deBruijn, T. L. Marsh, and L. J. Forney. 1998. Parallel and divergent genotypic evolution in experimental populations of *Ralstonia* sp. J.Bacteriol. 180(17):4325-4331.

72. Nielson, J. W., K. L. Josephson, I. L. Pepper, R. P. Arnold, G. D. DiGiovanni, and N. A. Sinclair. 1994. Frequency of horizontal gene transfer of a large catabolic plasmid (pJP4) in soil. Appl. Environ. Microbiol. 60(11):4053-4058.

73. Nurk, A., L. Kasak, and M. Kivisaar. 1991. Sequence of the gene (*pheA*) encoding phenol monooxygeanse from *Pseudomonas* sp. ESP 1001: expression in *Escherichia coli* and *Pseudomonas putida*. Gene 102:13-18.

74. Pemberton, J. M., B. Corney, and R. H. Don. 1979. Evolution and spread of pesticide degrading ability among soil micro-organisms, p. 287-299. *In* K. N. Timmis and A. Puhler (ed.), Plasmids of Medical, Environmental, and Commercial Importance. Elsevier/North-Holland Biochemical Press, Amsterdam.

75. Pemberton, J. M., and P. R. Fisher. 1977. 2,4-D plasmids and persistence. Nature 268:732-733.

76. Ponder, M. 2001. Personal communication.

77. Prescott, A. G. 1993. A dilemma of dioxygenases (or where biochemistry and molecular-biology fail to meet). J. Exp. Bot. 44:849.

78. Prescott, A. G., and M. D. Lloyd. 2000. The iron (II) and 2-oxoacid-dependent dioxygenases and their role in metabolism. Nat. Prod. Rep. 17:367-383.

79. Que Hee, S. S., and R. G. Sutherland. 1981. The phenoxyalkanoic herbicides: volume I chemistry, analysis, and environmental pollution. CRC Press, Inc., Boca Raton, Florida.

80. Roach, P. L., I. J. Clifton, V. Fulop, K. Harlos, G. J. Barton, J. Hajdu, I. Andersson, C. J. Schofield, and J. E. Baldwin. 1995. Crystal structure of isopenicillin N synthase is the first from a new structural family of enzymes. Nature 375:700-704.

81. Rondon, M. R., P. R. August, A. D. Bettermann, S. F. Brady, T. H. Grossman, M. R. Liles, K. A. Loiacono, B. A. Lynch, I. A. MacNeil, C. Minor, M. Gilman, C. L.

Tiong, M. S. Osburne, J. Clardy, J. Handelsman, and R. M. Goodman. 2000. Cloning the soil metagenome: a strategy for accessing the genetic and functional diversity of uncultured microorganisms. Appl. Environ. Microbiol. 66(6):2541-2547.

82. Rosazza, J. P., Z. Huang, L. Dostal, T. Volm, and B. Rousseau. 1995. Review: biocatalytic transformations of ferulic acid: An abundant aromatic natural product. J. Ind. Microbiol. 15(6):457-471.

83. Rost, B., and C. Sander. 1993. Prediction of protein secondary structure at better than 70-percent accuracy. Journal of Molecular Biology 232(2):584-599.

84. Ryle, M. J., R. Padmakumar, and R. P. Hausinger. 1999. Stopped-flow kinetic analysis of *Escherichia coli* taurine/α-ketoglutarate dioxygenase: interactions with α-ketoglutarate, taurine, and oxygen. Biochemistry 38:15278-15286.

85. Saari, R. E., and R. P. Hausinger. 1998. Ascorbic acid-dependent turnover and reactivation of 2,4-dichlorophenoxyacetic acid/α-ketoglutarate dioxygenase using thiophenoxyacetic acid. Biochemistry 37:3035-3042.

86. Saari, R. E., D. A. Hogan, and R. P. Hausinger. 1999. Stereospecific degradation of the phenoxypropionate herbicide dichlorprop. J. Molec. Catalysis B: Enzymatic 6:421-428.

87. Sali, A., and T. L. Blundell. 1993. Comparative protein modelling by satisfaction of spatial restraints. J. Mol. Biol. 234:779-815.

88. Sandmann, E. R. I., and M. A. Loos. 1988. Aromatic metabolism by a 2,4-D degrading *Arthrobacter* sp. Can. J. Microbiol. 34:125-130.

89. Seffernick, J. L., M. L. deSouza, M. J. Sadowsky, and L. P. Wackett. 2001. Melamine deaminase and atrazine chlorohydrolase: 98 percent identical but functionally different. J. Bacteriol. 183(8):2405-2410.

90. Segura, A., P. V. Bunz, D. A. D'Argenio, and L. N. Ornston. 1999. Genetic analysis of a chromosomal region containing *vanA* and *vanB*, genes required for conversion of either ferulate or vanillate to protocatechuate in *Acinetobacter*. J. Bacteriol. 181(11):3494-3504.

91. Stanier, R., N. Palleroni, and M. Doudoroff. 1966. The aerobic Pseudomonads: A taxonomic study. J. Gen. Microbiol. 43:159-271.

92. Steenson, T. I., and N. Walker. 1957. The pathway of breakdown of 2,4-dichloro- and 4-chloro-2-methyl-phenoxyacetic acid by bacteria. J. Gen. Microbiol. 16:146-155.

93. Strack, D. 1997. Phenolic metabolism, p. 387. *In* P. M. Dey and J. B. Harbone (ed.), Plant Biochemistry. Academic Press, San Diego.

94. Streber, W. R., K. N. Timmis, and M. H. Zenk. 1987. Analysis, cloning, and high-level expression of 2,4-dichlorophenoxyacetate monoxygenase gene *tfdA* of *Alcaligense eutrophus* JMP134. J. Bacteriol. 169:2950-2955.

95. Studier, F. W., A. H. Rosenberg, J. J. Dunn, and J. W. Dubendorff. 1990. Meth. Enzymol. 185:60-89.

96. Suwa, Y., A. D. Wright, F. Fukimori, K. A. Nummy, R. P. Hausinger, W. E. Holben, and L. J. Forney. 1996. Characterization of a chromosomally encoded 2,4-dichlorophenoxyacetic acid/α-ketoglutarate dioxygenase from *Burkholderia* sp. strain RASC. Appl. Environ. Microbiol. 62(7):2464-2469.

97. Thornburg, L. D., and J. Stubbe. 1993. Mechanism-based inactivation of thymine hydroxylase, an α-ketoglutarate-dependent dioxygenase, by 5-ethynyluracil. Biochemistry 32:14034-14042.

98. Thornburg, L. D., and J. Stubbe. 1989. Mechanism-based inhibition of thymine hydroxylase. J. Am. Chem. Soc. 111:7632-7633.

99. Top, E. M., W. E. Holben, and L. J. Forney. 1995. Characterization of diverse 2,4-dichlorophenoxyacetic acid-degradative plasmids isolated from soil by complementation. Appl. Environ. Microbiol. 61(5):11691-1698.

100. Top, E. M., O. V. Maltseva, and L. J. Forney. 1996. Capture of a catabolic plasmid that encodes only 2,4-dichlorophenoxyacetic acid: α-ketoglutaric acid dioxygenase (TfdA) by genetic complementation. Appl. Environ. Microbiol. 62(7):2470-2476.

101. Valegard, K., A. Terwisscha van Scheltinga, C., M. D. Lloyd, T. Hara, S. Ramaswamy, A. Perrakis, A. Thompson, H.-J. Lee, J. E. Baldwin, C. J. Schofield, J. Hajdu, and I. Andersson. 1998. Structure of a cephalosporin synthase. Nature 394:805-809.

102. Vallaeys, T., L. Courde, C. McGowan, A. D. Wright, and R. R. Fulthorpe. 1999. Phylogenetic analyses indicate independent recruitment of diverse gene cassettes during assemblage of the 2,4-D catabolic pathway. FEMS Microbiol. Ecol. 28(1999):373-382.

103. Vermeij, P., C. Wietek, and M. Kertesz. 2000. Characterization of a sulfur-regulated oxygenative alkylsulfatase from *Pseudomonas purtida* S-313. J. Biol. Chem. 275(41):31661-31667.

104. Watanabe, M., N. Sumida, S. Murakami, H. Anzai, C. J. Thompson, Y. Tateno, and T. Murakami. 1999. A phosphonate-induced gene which promotes *Penicillium*-mediated bioconversion of *cis*-Propenylphosphonic acid to fosfomycin. Appl. Environ. Microbiol. 65(3):1036-1044.

105. Whiting, A. K., R. E. Saari, M. A. Fredrick, J. McCracken, R. P. Hausinger, and L. Que. 1997. Metal coordination environment of a Cu(II)-substituted α-keto acid-dependent dioxygenase that degrades the herbicide 2,4-D. J. Am. Chem. Soc. 119(14):3413-3414.

106. Wilmouth, R., J. Turnbull, R. Welford, I. J. Clifton, A. G. Prescott, and C. J. Schofield. 2002. Structure and mechanism of anthocyanidin synthase from *Arabidopsis thaliana*. Structure 10(1):93-103.

107. Xia, X.-S., S. Aathithan, K. Oswiecimska, A. R. W. Smith, and I. J. Bruce. 1998. A novel plasmid pIJB1 possessing a putative 2,4-dichlorophenoxyacetate degradative transposon Tn5530 in *Burkholderia cepacia* strain 2a. Plasmid 39:154-159.

108. Yano, T., S. Oue, and H. Kagamiyama. 1998. Directed evolution of an aspartate aminotransferase with new substrate specificities. Proc. Natl. Acad. Sci. USA 95:5511-5515.

109. Zhang, Z., J. Ren, D. K. Stammers, J. E. Baldwin, K. Harlos, and C. J. Schofield. 2000. Structural origins of the selectivity of the trifunctional oxygenase clavaminic acid synthase. Nature Struct. Biol. 7(2):127-133.

APPENDIX A

DEVELOPMENT OF METHODS TO EVOLVE A *tfdA*-like GENE

The *tfd* genes appear to have been recruited from multiple metabolic pathways onto a single plasmid backbone to generate a plasmid capable of conferring to the organism the ability to degrade 2,4-D (40). Homologues of the *tfdBCDE* genes are known to encode proteins that utilize naturally occurring non-chlorinated substrates related to the 2,4-D degradative intermediates (35, 36, 73), but the substrate of the environmental homologues of TfdA have not been identified. Based on PCR amplification studies, *tfdA*-like sequences are widespread in the environment but the organisms containing these putative genes generally do not degrade 2,4-D (46). Furthermore, these *tfdA*-like sequences have been present in the environment for over a thousand years based on my analysis of Siberian permafrost soil isolates (Chapter 5). I hypothesize that *tfdA*-like genes in nature work on a substrate similar in structure to 2,4-D and that, through a small number of mutations, the gene was able to mutate into a *tfdA* gene encoding the 2,4-D degrading enzyme. I suggest that this process should be able to be reproduced artificially by directed evolution. For example, by artificially introducing small numbers of point mutations into a *tfdA*-like gene and screening the constructs for increased activity with 2,4-D it may be possible to obtain reasonable TfdA activity after several rounds.

To attempt to mimic the natural evolution of TfdA, I used error prone PCR to introduce small numbers of point mutations into *Bordetella pertussis* and *Bradyrhizobia* sp. *tfdA*-like genes. The *B. pertussis* and *Bradyrhizobia* sp. genes were chosen as candidates for directed evolution because the encoded proteins are closely related to TfdA, yet do not prefer 2,4-D as a substrate.

EXPERIMENTAL

Media and cultures. Escherichia coli DH5α (pMAL-tfdA) carrying *tfdA* on the pMAL plasmid (47) was used for experimental positive controls. *E. coli* DH5α (pMAL-brad) containing a *Bradyrhizobia* sp. *tfdA*-like gene, graciously supplied by Kazuhito Ito, and *E. coli* DH5α (pMAL-BP1665), containing a *Bordetella pertussis tfdA*-like sequence previously described (Chapter 6), were used as sources of plasmid for the directed evolution studies. The proteins encoded for by the plasmids had an N-terminal maltose binding protein fusion. All cultures were grown in LB or on LB containing 1.5% agarose.

Random mutagenesis. Plasmid DNA was isolated from overnight cultures using Promega Wizard preps. The *tfdA*-like genes were mutagenized by amplification of the corresponding pMAL plasmid by the method of Cadwell and Joyce (15) with the New England Biolabs *malE* and M13 primers in 20 mM Tris-HCl (pH 8.4), 50 mM KCl, 7 mM $MgCl_2$, 0.5 mM $MnCl_2$, 5 mM dCTP, 5 mM dTTP, 1 mM dATP and 1 mL dGTP, 1 μM each primer, and 0.05 units/μL Gibco *Taq* DNA polymerase. Non-mutagenic amplification was carried out with the same primers in 20 mM Tris-HCl (pH 8.4), 50 mM KCl, 2.5 mM $MgCl_2$, 1.25 mM each dNTP (Gibco), 1 μM each primer, and 0.05 units/μL Gibco *Taq* DNA. Initial denaturation occurred at 94 °C for 5 min, followed by elongation for 35 cycles of denaturation for 1 min at 95 °C, annealing at 55 °C for 45 sec, and elongation at 72 °C for 45 sec, with a final 10 min elongation step at 72 °C.

Screening. The resulting PCR products were cleaved with *Hin*dIII and *Xba*I, and inserted into pMAL-c2 cut with the same enzymes to create a library of mutagenized plasmids. This library of mutagenized plasmids was transformed into Gibco DH5α MAX $E_{fficiency}$ cells so as to yield approximately 1000 colony forming units on a 150 mm plate.

The colonies were transferred to sterile nitrocellulose filters, and the filters were soaked in 50 mM Tris buffer (pH 7) with the colony side up for 30 min. (It should be noted that the filters were not soaked in IPTG to induce the cells; instead, the plate assay relied on the leakiness of the *tac* promoter when cells are grown on LB not amended with glucose.) The filters were transferred to a shallow tray containing 10 mM imidazole, 1 mM αKG, 1 mM 2,4-D, 10 μM ferrous ammonium sulfate, and 10 μM ascorbate and soaked for 10 min. 4-Aminoantipyrene was added to 0.02%, pH 10 buffer to 0.1X, NaOH to adjust the pH to 9.0, and 0.08% potassium ferricyanide (59). The color was developed for 5 min. Areas identified as having more activity than non-mutagenized colonies were marked on the original plates. The original plates were re-grown and colonies from that region of the plate were streaked for isolation and re-screened by the filter assay to identify the exact colony with enhanced activity.

Activity analysis. Cell lysates were prepared as previously described (Chapter 2) for isolates identified as having enhanced activity as well as for *E. coli* (pMAL-brad) and *E. coli* (pMAL-tfdA) as controls. Cell lysates were prepared from both static cultures and well-aerated cultures. Purified proteins were obtained by passing the cell lysates over an amylose column as previously described (47). Cell lysates and purified proteins were assayed by using the 4-aminoantipyrene assay (59) and proteins concentrations were determined by using a commercial protein assay (BioRad) with bovine serum albumin as a standard.

Plate matings. Overnight cultures of *E. coli* DH5α (pACYC) in LB amended with 34 μg/mL chloramphenicol and *R. eutropha* JMP228 (pBH501aE) in LB amended with 50 μg/mL kanamycin and 100 μg/mL rifampin were diluted 1000-fold into 10% LB. LB

plates were prepared with only 1 mL of the *R. eutropha* JMP228 (pBH501aE) dilution, 1 mL of the *E. coli* DH5α (pACYC) dilution, or 1 mL of the *R. eutropha* JMP228 (pBH501aE) combined with 1 mL of the *E. coli* DH5α (pACYC) dilution. After incubating overnight at 30 °C, the plates were scraped into 5 mL sterile PBS, serially diluted and plated on LB amended with 34 μg/mL chloramphenicol, 50 μg/mL kanamycin, and 100 μg/mL rifampin. The colonies were counted after overnight incubation at 30 °C.

RESULTS AND DISCUSSION

The *B. pertussis* and *Bradyrhizobia* sp. *tfdA*-like genes were chosen as candidates for directed evolution because the encoded proteins are closely related to TfdA (Figure 6.1), yet they do not prefer 2,4-D as a substrate. The maltose binding protein fusions of the *B. pertussis* TfdA-like protein had no detectable activity with 2,4-D and the *Bradyrhizobia* sp. TfdA-like protein degraded 0.03 μmol of 2,4-D min^{-1} mg^{-1} (48). This value corresponds to 0.4% of the activity of the maltose binding protein fusion of wild-type TfdA (30). Consistent with these results with purified proteins, cells producing the *B. pertussis* protein did not turn red, those producing the *Bradyrhizobia* protein turned pink, and those producing the *R. eutropha* TfdA protein turned dark red.

Once I confirmed that the colorimetric plate assay could be used to screen for colonies with increased activity towards 2,4-D as substrate, the *B. pertussis* and *Bradyrhizobia* sp. *tfdA*-like genes were mutagenized by PCR mutagenesis. I utilized a method involving the addition of manganese and an altered dNTP ratio since this procedure was determined to maximize the mutation rate without bias for certain types of mutations (15). An initial round of mutagenesis and cloning was carried out with the *B. pertussis tfdA*-like gene, the *Bradyrhizobia tfdA*-like gene, and the *tfdA* gene. From each set, six colonies were picked and grown in LB. Fourteen of the eighteen cultures contained plasmids of the expected size for vector with insert (Table A.1). Of the other four, one plasmid was larger than expected, one was smaller than expected, and two did not yield plasmid. Of the fourteen cultures containing appropriately sized plasmids, seven produced appropriately sized proteins, five produced truncated proteins, and two

Table A.1 Characteristics of selected mutagenized clones.

Expected Insert	Clone Number	Expression (assessed on SDS-PAGE gel)	Insert (assessed on agarose gel)
pMAL-brad (mutagenized)	1	Full length	Yes
	2	Full length	Yes
	3	Full length	No plasmid
	4	No	Smaller than expected
	5	Truncated	Yes
	6	Full length	Yes
pMAL-BP1665 (mutagenized)	1	Full length	Yes
	2	Almost full length	Yes
	3	Almost full length	Yes
	4	No	No plasmid
	5	Truncated	Yes
	6	No	Yes
	7	Full length	Yes
pMAL-tfdA (mutagenized)	1	Full length	Bigger than expected
	2	Full length	Yes
	3	Almost full length	Yes
	4	No	Yes
	5	Full length	Yes
pMAL-tfdA		Full length	Yes

did not produce any TfdA-like protein. Although there seems to be a disproportionate number of truncations the mutagenesis was deemed successful and a larger screening was undertaken.

Six plates, two for the mutagenized *Bradyrhizobia* gene, two for the mutagenized *Bordetella* gene, and two for the mutagenized *tfdA*, were prepared with approximately one thousand colonies each, replica filtered, and assayed. The plates containing the mutagenized *B. pertussis* gene yielded no pink or red colonies. Approximately seven percent of the colonies containing mutagenized *tfdA* turned red. The plates containing colonies with the mutagenized *Bradyrhizobia* gene contained twenty areas that were red and numerous more areas that were pink. The twenty red areas were picked, streaked, and reanalyzed by the assay. Of these, four were deemed to be active. A colony from each of these was further analyzed (1-4-6-4, 1-4-7-4, 1-4-8-2, and 1-4-9-2).

Shaken and static cultures of 1-4-6-4, 1-4-7-4, 1-4-8-2, 1-4-9-2, *E. coli* (pMAL-brad), and *E. coli* (pMAL-tfdA) were grown, cell lysates were prepared, and the proteins were purified by amylose column chromatography. No activity was detected in the purified proteins from the mutagenized clones or the non-mutagenized *E. coli* (pMAL-brad) cultures. In contrast, the control protein was active as expected (47).

As an alternative to the phenol staining assay approach, I explored using *R. eutropha* JMP228 (pBH501aE) for screening the evolved proteins. *R. eutropha* JMP228 (pBH501aE) is a derivative of *R. eutropha* JMP134 (pJP4) that contains the *tfdBCDEF* genes, but lacks *tfdA*. I showed that pACYC184 is compatible with the plasmid already in the host and found that both plasmids could be maintained. Initially studies with *R. eutropha* JMP228 (pBH501aE) and *E. coli* (pACYC184) confirmed that pACYC184 was

mobilized into *R. eutropha* JMP228 (pBH501aE) during plate matings with a yield of about 6700 transformants per plate mating. Six attempts were carried out using four different locations for inserting the *Bradyrhizobia* sp. *tfdA*-like gene into pACYC184; none were successful. As a control, the same attempts were made with *tfdA*, but these were also unsuccessful. The final attempt included containing different antibiotics on each piece of DNA being ligated and screening with both antibiotics. No appropriately sized transformants could be obtained in any of the cases. I presume that the genes are not stable in the plasmid.

Kazuhito Ito has cloned the *Bradyrhizobia* sp. *tfdA*-like gene into a different plasmid and a construct of *R. eutropha* JMP228 (pBH501aE) containing this plasmid has been made and is currently being characterized (48). Future studies could be carried out in that construct if the plasmid and gene are stable and if protein isolated from those cells dimerizes correctly.