CHARACTERIZING A PROPOSED NOVEL ENZYME FAMILY OF REVERSIBLE HYDROXYARYLIC ACID DECARBOXYLASES

by

DELINA Y. LYON

(Under the direction of Dr. Juergen Wiegel)

ABSTRACT

In this research, two adjacent genes, *shdC* and *shdD*, were found necessary to encode an active 4-hydroxybenzoate decarboxylase (ShdCD) from *Sedimentibacter hydroxybenzoicus* JW/Z-1 (formerly *Clostridium hydroxybenzoicum* JW/Z-1). Homology searches of ShdCD revealed three homologues with the same genetic arrangement - the homologues to *shdC* and *shdD* adjacent to each other. These homologues are from *Streptomyces* sp.D7, *Bacillus subtilis,* and *Escherichia coli* O157:H7. The homologue from *Streptomyces* sp.D7 had already been identified as a vanillate decarboxylase in the laboratory of Dr. Julian Davies. The genes encoding the three homologues were cloned into *E. coli* JM109, expressed, and assayed for decarboxylase activity. The homologues from *Streptomyces* sp.D7 and *B. subtilis* are vanillate decarboxylases, and the one from *E. coli* O157:H7 is a 4-hydroxybenzoate decarboxylase. Based on sequences and enzyme activities, we propose that the decarboxylases from *S. hydroxybenzoicus*, *B. subtilis*, *Streptomyces* sp.D7, and *E. coli* O157:H7 represent a novel enzyme family.

INDEX WORDS: Aromatic compound degradation, decarboxylase, anaerobic, *Sedimentibacter hydroxybenzoicus*, hydroxyarylic acid

CHARACTERIZING A PROPOSED NOVEL ENZYME FAMILY OF REVERSIBLE HYDROXYARYLIC ACID DECARBOXYLASES

by

DELINA Y. LYON

B.A., St. Mary's College of Maryland, 1998

A Thesis Submitted to the Graduate Faculty of The University of Georgia in Partial Fulfillment of the Requirements for the Degree

MASTER OF SCIENCE

ATHENS, GEORGIA

2002

CHARACTERIZING A PROPOSED NOVEL ENZYME FAMILY OF HYDROXYARYLIC ACID DECARBOXYLASES

by

DELINA Y. LYON

Approved:

Major Professor: Juergen Wiegel

Committee: Ellen Neidle
Timothy Hoover

Electronic Version Approved:

Gordhan L. Patel
Dean of the Graduate School
The University of Georgia
August 2002

DEDICATION

I would like to dedicate my Master's thesis to my parents, Walter and Gennevive Lyon, for their unfaltering support and encouragement.

ACKNOWLEDGEMENTS

I would especially like to thank my major professor, Dr. Juergen Wiegel, for his support. I wish to express my sincere gratitude to my committee members Dr. Ellen Neidle and Dr. Tim Hoover for their advice and guidance. My thanks to everyone in the University of Georgia Microbiology Department - I'm sure I've come to every lab for help. To my friends who've helped me in and out of lab, y'all have been great. Leena Palekar, the sister I never had, thanks for listening to me whine. George Hinkal, where would I be without your love and support?

TABLE OF CONTENTS

Page

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION AND LITERATURE REVIEW

This research discusses a proposed novel enzyme family of reversible hydroxyarylic acid decarboxylases. To understand the role played by these decarboxylases, it is helpful to understand the pathways by which aromatic compounds can be degraded. This introduction will include a review of known degradation pathways of both halogenated and non-halogenated aromatic compounds. Hydroxybenzoate compounds will be defined, and known aromatic compound decarboxylases will be discussed. This review will also introduce the hydroxyarylic acid decarboxylases which are the subject of this research.

Aromatic compounds are ubiquitous in the environment, coming from both natural and anthropogenic sources (Londry and Fedorak, 1992). The largest natural source of aromatic compounds are plants, in which aromatic compounds are components in lignin and in several plant products. The most abundant man-made aromatic compounds come from petroleum refineries, wood preservatives, pulp and paper mill bleaching, and herbicides and pesticides. As for halogenated aromatic compounds, while most concern is focused on anthropogenic sources, the majority in the environment are biological in origin (Fetzner, 1998).

Aromatic compounds have varying levels of toxicity, depending on the chemical and physical properties of a given compound. One factor influencing the toxicity levels in mammals is the lipophilicity of the compound, which affects the propensity of the compound to accumulate in the adipose tissues (Philip, 2001). Aromatic compounds also accumulate in organisms as one moves higher up in the food chain, a phenomenon known as biomagnification. At low levels, most of these aromatic compounds are not very toxic, but biomagnification leads to higher concentrations and toxicities. Acute effects vary

depending on the particular chemical (Harbison, 1998). Chronic effects also vary, often including cancer, organ damage, and a series of neurological effects. Halogenated aromatic compounds are partially toxic due to both the aromatic qualities of the compound and the stability and added toxicity imparted by the halogen.

Pathways of Aromatic Compound Degradation

Many pathways are known for aerobic/anaerobic degradation of aromatic compounds, most of which are listed in the University of Minnesota Biocatalysis/ Biodegradation Database (UM-BBD, http://www.labmed.umn.edu/umbbd/index.html). Some examples of aromatic degradation are discussed below.

Aerobic Aromatic Compound Degradation

The general scheme for aerobic aromatic degradation begins with the various substrates being altered by peripheral enzymes into the central metabolites catechol, protocatechuate, or gentisate (Mishra, *et al.*, 2001). Mono- or di- oxygenases catalyze either the intradiol or extradiol cleavage of the aromatic ring, routes that are called *ortho*- or *meta*- cleavage pathways, respectively. This cleavage results in the formation of *cis,cis*-muconate or 2-hydroxy-muconic semialdehyde, and these compounds are then metabolized by the organism. These conversions are illustrated in Figure 1.1.

Aerobic degradation of aromatic compounds has been studied and reviewed extensively, and a few examples follow. An aerobic phthalate degradation pathway has been characterized in *Arthrobacter keyseri* 12B. The genes for phthalate and protocatechuate catabolism are plasmid-borne. The pathway involves

3,4-dihydroxyphthalate decarboxylase, which will be discussed in a later section (Eaton, 2001). A similar phthalate degradation pathway was found in *Burkholderia cepacia* DBO1. It uses an aerobic 4,5-dihydroxyphthalate decarboxylase, the mechanism of which is unknown (Chang and Zylstra, 1998; Pujar and Ribbons, 1985). A similar 4,5-dihydroxyphthalate decarboxylase found in *Pseudomonas testosteroni* does not seem to use cofactors, but its mechanism has not been described either(Nakazawa and Hayashi, 1978).

Enterobacter cloacae strain EM hydrolyzes paraben (4-hydroxybenzoic acid esters) to phenol aerobically (Valkova, *et al.*, 2001). Parabens are widely used as antimicrobials with low levels of toxicity. *E. cloacae* first uses esterases to form 4-hydroxybenzoate and then a decarboxylase to form phenol but does not degrade phenol further. This pathway differs from the beta-ketoadipate pathway, which is the most common pathway for aerobic aromatic degradation.

Bacteria are not the only efficient aromatic compound degraders in the environment. Fungi are among the most capable aromatic degraders. A number of different pathways are known (Wright, 1993). One example is the aerobic transformation of *p*-coumaric acid to *p*-hydroxybenzaldehyde in *Pycnoporus cinnarbarinus*, with two possible pathways to account for the observed products (Alvarado, *et al.*, 2001). There are many different correlations between fungal and bacterial degradation, with the comparison of the two yielding interesting insights (Wright, 1993). For example, fungi, like *Rhodotorula mucilaginosa*, have equivalents of the *ortho* ring cleavage pathway that is found in some bacteria. No fungi have been identified with equivalents to the *meta* ring cleavage pathway. Some research has been done on the degradation of aromatic

compounds by microalgae which would play a large role in the degradation of aromatic compounds in the oceans (Semple, *et al.*, 1999).

Anaerobic Aromatic Compound Degradation

While aerobic degradation is important, anaerobic processes are significant due to the prevalence of anoxic environments. A number of anaerobic degradation pathways have been elucidated. These pathways are of particular interest due to their applicability in bioremediation. Soils are often aerobic only for the first few centimeters, below which aromatic compounds have to be degraded anaerobically.

Several anaerobic pathways for phenolic compound degradation have been found (Schink, *et al.*, 2000). Depending on the type of organism, anaerobic degradation can take on many forms (Elder and Kelly, 1994). In photosynthetic organisms, photometabolism can be linked to organic substrates in which aromatic compounds are used as electron donors. Methanogens do not perform aromatic compound degradation but rather consume the products of the degradations, like acetate. Methanogens use only simple carbon compounds, and by utilizing waste products of aromatic compound degradation, they allow bacterial degradation, which might have been affected by waste build up, to continue. Dissimilatory nitrate reduction is another process by which aromatic compounds can be oxidized completely to CO_2. Some dissimilatory sulfate reducers use organic compounds, including aromatic compounds, as the electron donors. Some bacteria, like *Sporotomaculum hydroxybenzoicum*, are capable of fermenting aromatic compounds, using them as electron donors and acceptors.

Some common anaerobic degradation pathways include the benzoyl-CoA pathway, the resorcinol pathway, the phloroglucinol pathway, and the hydroxyhydroquinone pathway (Londry and Fedorak, 1992). These pathways are named after their central compounds and are discussed in further detail below. Another pathway exists in which 4-hydroxybenzoate is decarboxylated to phenol, and then hydroxyl group is replaced by a carboxyl group to yield benzoate (Knoll and Winter, 1989).

Aromatic-CoA compounds are common intermediates in anaerobic degradation pathways. Benzoyl-CoA, the central intermediate in the benzoyl-CoA pathway, is formed when aromatic compounds are altered at the CoA-ester level or when HSCoA attaches to benzoate and releases AMP in the ligation (Londry and Fedorak, 1992). Benzoyl-CoA is more easily reduced than benzoate, releasing cyclohexadiene carboxyl-CoA. This is an endergonic process, the energy for which may be derived from downstream degradation of the cyclohexadiene carboxyl-CoA. The benzoyl-CoA pathway has been characterized best in anaerobes such as *Desulfitobacterium* sp. strain Cat 2, *Rhodopseudomonas palustris*, and *Thauera aromatica* (Egland and Harwood, 2000; Gorny and Schink, 1994). *T. aromatica* has been well studied. Most of the genes involved in phenol metabolism have been identified (Biegert, *et al.*, 1996; Boll, *et al.*, 1997; Brackmann and Fuchs, 1993; Breese and Fuchs, 1998; Breinig, *et al.*, 2000; Harwood, *et al.*, 1999; Heider, *et al.*, 1998; Lack and Fuchs, 1992; Lack and Fuchs, 1994; Laempe, *et al.*, 2001; Schneider, *et al.*, 1997; Schneider and Fuchs, 1998). *Sporotomaculum hydroxybenzoicum* uses the benzoyl-CoA pathway to ferment 3-hydroxybenzoate, with benzoate as a transient intermediate (Muller and Schink, 2000). The end products are acetate, butyrate, and carbon dioxide. *R. palustris* also uses the

benzoyl-CoA pathway to degrade 4-hydroxybenzoate (Egland and Harwood, 2000; Elder and Kelly, 1994). A 4-hydroxybenzoate sensor regulates the expression of the genes involved in degradation of this aromatic compound. Fermentative benzoate degradation also occurs via the benzoyl-CoA pathway in *Syntrophus gentianae* (Schocke and Schink, 1999). There is also an example of aerobic degradation involving an aromatic-CoA compound. Ferulic acid is metabolized via vanillin in *Pseudomonas fluorescens* when a ferulate-CoA ligase links the CoA group to ferulic acid (Narbad and Gasson, 1998). A cleavage of the side chain leaves vanillin which is oxidized to vanillate and then demethylated to protocatechuate, which can then enter the beta-ketoadipate pathway.

Other pathways besides the benzoyl-CoA pathway have been described for the anaerobic degradation of aromatic compounds in bacteria. The anaerobic degradation of gallate (3,4,5-trihydroxybenzoate) by *Eubacterium oxidoreducens* involves the phloroglucinol pathway (Haddock and Ferry, 1993). Gallate is first decarboxylated to pyrogallol, which undergoes isomerization to phloroglucinol and is then reduced to dihydrophloroglucinol. Once the aromaticity has been destroyed degradation occurs readily to yield acetate and butyrate. The catabolism of trihydroxybenzenes in *Pelobacter acidigallici* also uses the phloroglucinol pathway (Samain, *et al.*, 1986).

Moorella thermoacetica (formerly *Clostridium thermoaceticum*) uses a variety of carboxylated aromatic compounds to feed the acetogenic pathway (Hsu, *et al.*, 1990). It seems that the decarboxylation of these aromatics yields carbon dioxide equivalents which can sustain growth of the organism under CO_2-limited conditions.

Polycyclic aromatic hydrocarbon (PAH) degradation has also been studied in bacteria (Kanaly and Harayama, 2000). These pathways often involve the separation of

the multiple aromatic rings and followed by the degradation of the individual rings using the same pathways as those used for monocyclic aromatics. In the anaerobic metabolism of naphthalene and phenanthrene, carboxylation may be the intial reaction by a sulfidogenic consortium (Zhang and Young, 1997).

An important factor in the degradation of halogenated aromatic compounds is the degree of halogenation of the benzoate ring. Halogenation can seriously alter the characteristics of the compound, especially the hydrophobicity and stability. Three basic strategies have been found for enzyme-catalyzed dehalogenations (El Fantroussi, *et al.*, 1998). (1) Dehalogenations can occur due to co-metabolism, during which the dehalogenation occurs coincidentally with metabolism. (2) In halorespiration, the organism uses halogenated compounds as electron acceptors. (3) The organism can also use the halogenated compound as a sole carbon source.

Microbes can dehalogenate aromatic compounds oxidatively using mono- or dioxygenases in co-metabolic or metabolic processes. Alternatively, dehalogenation can occur reductively, as when a dehydrohalogenase eliminates the halide forming a double bond in the organic compound (Fetzner, 1998). Substitutive dehalogenation replaces the halogen with a different side group. Most anaerobic pathways, however, go through reductive dehalogenation in which the halide is replaced by hydrogen. This can occur co-metabolically, as part of carbon metabolism, or as a method of energy conservation giving a reduced compound as the end product.

Significance of Aromatic Compound Degradation

The knowledge of degradation pathways plays a significant role for industrial applications as well as in efforts to develope waste bioremediation processes. There has been considerable discussion and research on the engineering of bacteria to optimize their degradative capacities, especially with respect to toxic waste compounds (Timmis and Pieper, 1999). These modifications involve either altering the pathways or the enzymes involved in a pathway. Researchers try to identify new pathways of aromatic compound degradation or expand the substrate ranges of existing pathways to enhance bioremediation. As researchers modify pathways, an important consideration is that no toxic byproducts or inhibitors of the pathway are made. Pathways for anaerobic aromatic compound degradation often involve a consortium of bacteria. Researchers can improve on the designs of these consortia, especially with respects to using both aerobic and anaerobic organisms. Often these pathways are supplemented with a specialist organism that is optimized to perform a difficult transformation. In terms of modifying enzymes, approaches involve increasing either stability or substrate range of an enzyme. Other factors for bioremediation include increasing pollutant bioavailability and optimizing survival of the bacteria in the environment.

Enzymes are often used for industrial scale production of aromatic compounds. Producing 4-hydroxybenzoate from phenol and carbon dioxide would greatly reduce the costs of labor intensive, fuel intensive, synthesis reactions (Aresta, *et al.*, 1998). 4-Hydroxybenzoate can be used as a base for manufacture of other industrially important aromatic compounds, such as salicylic acid, gentisic acid, gallic acid, and resorcylic acids.

Chlorophenol Degradation in Methanogenic Freshwater Sediment

One halogenated compound of environmental and health concern is 2,4-dichlorophenol. 2,4-Dichlorophenol, found in pesticides and wood treatments, is a recalcitrant aromatic pollutant that is carcinogenic and causes a variety of health problems. One pathway for the degradation of 2,4-dichlorophenol, as proposed by Zhang and Wiegel, involved a consortium of bacteria from methanogenic pond sediment from Sandy Creek Nature Park in Athens, Georgia (Zhang and Wiegel, 1992; Zhang and Wiegel, 1990). The pathway, illustrated in Figure1.2, begins with two successive reductive dechlorinations, followed by a carboxylation to form 4-hydroxybenzoate decarboxylase (Zhang, *et al.*, 1990), which is then dehydroxylated. The resulting benzoate is mineralized to carbon dioxide and hydrogen sulfide (sulfate reducers) or to methane (methanogens), depending on the species present. The rate-limiting step in this pathway appears to be the carboxylation of phenol to 4-hydroxybenzoate.

Zhang and Wiegel were able to isolate the organism responsible for this reaction, using enzyme assays to track the organism as they narrowed their isolation (Zhang and Wiegel, 1990; Zhang and Wiegel, 1994). The organism, *Sedimentibacter hydroxybenzoicus* JW/Z-1 (previously *Clostridium hydroxybenzoicum* JW/Z-1), is a Gram-type positive, spore-forming rod (Breitenstein, *et al.*, 2001; Zhang, *et al.*, 1994). The reaction of interest in this pathway is the reversible decarboxylation of 4-hydroxybenzoate to phenol, with the decarboxylation highly favored over the carboxylation (He and Wiegel, 1995; Zhang and Wiegel, 1994). The characteristics of this enzyme will be described in an upcoming section of this chapter.

Hydroxybenzoate Compounds

Since the focus of this research is the decarboxylation of hydroxyarylic acids, the following section describes the structures and natures of such compounds. Hydroxybenzoate and its derivatives are based on the following general structure, in which there is a carboxyl group at the 1 position of the benzene ring, and at least one of the other side groups is a hydroxyl group. There are three monohydroxybenzoates (2-, 3-, 4-hydroxybenzoate), six dihydroxybenzoates (2,3-, 2,4-, 2,5-, 2,6-, 3,4-, and 3,5-dihydroxybenzoate), six trihydroxybenzoates (2,3,4-, 2,3,5-, 2,3,6-, 2,4,5-, 2,4,6-, 3,4,5-trihydroxybenzoate), three tetrahydroxybenzoates (2,3,4,5-, 2,3,4,6-, 2,3,5,6-tetrahydroxybenzoate), and one pentahydroxybenzoate (2,3,4,5,6-pentahydroxybenzoate). The pK_a values of the carboxylic groups of all the hydroxybenzoates listed above are below 4.5 (Beyer and Walter, 1996). Thus, in most biological systems, in which the pH is above 5.5, the chemicals exist mainly in the carboxylate form. The π electron system in benzoic acid can function as either an electron donor or acceptor, depending on the associated side chains.

All of the listed hydroxybenzoates are found in plants, specifically in lignin (Chapman and Hall, 1996). Several of them are industrially important. 2,3-Dihydroxybenzoate, more commonly known as salicylic acid, is used as an antiseptic, perfume, and preservative. Acetylsalicylic acid, or aspirin, is an antipyretic and antineuraglic drug. It is made commercially by the Kolbe-Schmitt synthesis reaction in which dry sodium phenoxide is reacted with carbon dioxide at 120-140°C. Gentisic acid, or 2,5-dihydroxybenzoate, is an analgesic, antirheumatic, and antiarthritic drug. Gallate, or 3,4,5-trihydroxybenzoate, is found in tannins and has been used as a photodeveloper

and tanning agent. It is also an important phytohormone. Some hydroxybenzoates, such as the resorcylic acids (3,5- or 2,6-dihydroxybenzoate) are byproducts of petrochemical processes and can be used in pharmaceutical and dye industries.

Decarboxylases/Carboxylases

Carboxylases are categorized as ligases, and decarboxylases are characterized as lyases (Horton, *et al.*, 1996). Although decarboxylase reactions are often physiologically irreversible, one enzyme can be capable of performing both reactions. The physiological function of some of these enzymes depends on downstream reactions. The downstream reactions can enable thermodynamically unfavorable (de)carboxylations by keeping the concentrations of the reaction products low inside the cell. This allows the (de)carboxylation to be "pulled" in one direction or another.

Enzyme Mechanisms of (De)Carboxylases

Decarboxylases/carboxylases vary in their substrate specificities and enzyme mechanisms (O'Leary, 1992). Research into the mechanisms of aliphatic carboxylases/decarboxylase is extensive. Crystal structures are known for the pyruvate decarboxylase from *Zymomonas mobilis* and biotin carboxylase (which is one of the multi-enzyme complex of acetyl-CoA carboxylase) from *Escherichia coli*. (De)carboxylases often rely on cofactors which affect the substrate's equilibrium constant such that it favors the relevant reaction. A number of cofactors have been found associated with (de)carboxylases. These cofactors can be metals, which effectively stabilize intermediates by metal ion complexation. Other mechanisms, which may or

may not involve metals, include Schiff-base dependent decarboxylations, thiamin pyrophosphate-dependent decarboxylations, and biotin-ATP-dependent mechanisms. A survey of some of the cofactors used in enzymes is summarized in Table 1.1. A number of enzymes employ unconventional cofactors as well. Some decarboxylases, to be discussed later, do not seem to use cofactors at all, but may actually use the properties of the substrate to help stabilize (de)carboxylation. A few examples of decarboxylases are given below.

Metal-dependent decarboxylases Some decarboxylases use only metals cofactors (O'Leary, 1992). For example, oxaloacetate decarboxylase uses divalent metal ions to stablilize reaction intermediates. Ribulose-1,5-bisphosphate carboxylase is the central carbon dioxide fixing enzyme in plants, and thus the most abundant carboxylase on earth. It has been crystallized from plants (spinach) and bacteria (*Rhodospirillum rubrum*). The mechanism involves magnesium binding to the carbon dioxide-lysine adduct and stabilizing it for transfer to ribulose-1,3-bisphosphate (Horton, *et al.*, 1996).

Schiff-base dependent decarboxylases Schiff base-dependent decarboxylases stabilize the negative charge following the decarboxylation step using nitrogen rather than oxygen (O'Leary, 1992). These reactions are essentially irreversible. Pyridoxal 5'-phosphate (PLP), found in pyridoxine or vitamin B_6, has a pyridoxine ring with a phosphate group attached and a reactive aldehyde. It is often used in the decarboxylation of amino acids, during which the amino acid reacts with the PLP to form Schiff bases. PLP is first bound to the enzyme by a Schiff base linkage to a lysine amino group, and it is also held by

other noncovalent bonds. The enzyme is displaced by the substrate amino acid to a form a Schiff base with the PLP. It is in this conformation that the carboxyl group is lost from the amino acid, after which a Schiff base interchange occurs to release the substrate and recapture the enzyme. The pyridinium ring acts as an electron sink to stabilize electrons released during decarboxylation. Pyruvate-dependent decarboxylases also use a Schiff base mechanism, but the electrons are not delocalized to the extent they are in PLP.

Thiamine pyrophosphate dependent decarboxylases Thiamin pyrophosphate (TPP), or vitamin B_1, has a pyrimidine ring and a thiazolium ring with a positive charge (O'Leary, 1992). The thiazolium ring acts as an electron sink. The best studied example of this enzyme is a non-oxidative pyruvate decarboxylase, which yields carbon dioxide and a two carbon fragment. TPP is also used to oxidatively decarboxylate a number of α-ketoacids.

Biotin-ATP dependent decarboxylases Biotin is bound by enzymes via an amide bond to the ε-amino group of a lysine residue, forming biocytin (Horton, *et al.*, 1996). The enzyme has two separate sites with the covalently bound biotin swinging from one site to the other following carboxylation. ATP is first hydrolyzed with bicarbonate to form a carboxyphosphate which reacts with biotin to form *N*-carboxy-biotin. The *N*-carboxy-biotin moves to the second site where the carboxyls group is transferred to the substrate. Biotin carboxylase, a component of acetyl-CoA carboxylase, uses ATP and biotin as cofactors. Pyruvate carboxylase also uses biotin and ATP. The crystal structure of a biotin carboxylase from *Escherichia coli* has been determined, and recent studies have

focused on using site-directed mutagenesis to identify amino acid residues critical for enzyme function (Blanchard, *et al.*, 1999).

Alternative cofactors for decarboxylases Several (de)carboxylases use alternative cofactors or no cofactors at all (O'Leary, 1992). For example, prephenate dehydrogenase couples the irreversible oxidative decarboxylation of prephenic acid to NAD^+ reduction. More examples are listed in the following sections.

Most of the decarboxylases which have been studied are oxidative. The cofactors described above are primarily involved in oxidative decarboxylations. Interest in non-oxidative decarboxylases is growing as the need for enzymes that can function in anaerobic environments increases. Enzymes that are oxygen-sensitive also offer a range of mechanisms for (de)carboxylation which have not been previously considered. Alternative mechanisms allow for alternative approaches to industrial and bioremediation applications.

Examples of Hydroxyarylic Acid Decarboxylases

Aromatic decarboxylases play an important role in the degradation of aromatic compounds, but few have been characterized. The properties of these decarboxylases are summarized in Table 1.1. Decarboxylases can use conventional cofactors, such as TPP or PLP. Benzoylformate decarboxylase uses TPP while aminobenzoate decarboxylase and aromatic-L-amino-acid decarboxylases use PLP. However, the majority of aromatic decarboxylases do not utilize conventional cofactors, and, for most of them, the mechanisms are unknown.

Three decarboxylases studied use metals as possible cofactors, although the exact mechanisms have not yet been clarified. *Arthrobacter keyseri* has a 3,4-dihydroxyphthalate decarboxylase, which uses zinc to stabilize the intermediary beta-keto acid that is formed by the tautomerization of the 3-hydroxyl group. This intermediate can form an enolate, allowing elimination of the beta-carboxy substituent as CO_2 (Eaton, 2001). Phenol carboxylases and 4-hydroxybenzoate decarboxylases, which were found in an anaerobic consortium, need potassium to function (Gallert and Winter, 1992). The oxygen sensitivities of the enzymes were not established. A 3-octaprenyl-4-hydroxybenzoate decarboxylase from *E. coli* K-12 involved in ubiquinone synthesis needs Mn^{2+} and a less than 10 kDa cofactor or protein for activity (Zhang and Javor, 2000).

Another type of non-oxidative decarboxylase for which there is a plausible mechanism is represented by the 2,3-dihydroxybenzoate decarboxylase from *Aspergillus niger*. This enzyme, which is similar to one from yeast, is a homotetramer with 28 kDa subunits and is specific for 2,3-dihydroxybenzoate (Kamath, *et al.*, 1987). It has no cofactors but needs the carboxyl group and a hydroxyl group at the *ortho* position for activity (Kamath and Vaidyanathan, 1990). The suggested mechanism involves an essential histidine and cysteine both of which are in the active site. The cysteine residue was identified by differential labelling with [C14] *N*-methylmaleimide (Santha, *et al.*, 1995). Researchers looked for protection by salicylate from labelling with the sulfhydryl agent which modifies cysteine residues so that the enzyme is inactivated. A similar 2,3-dihydroxybenzoate decarboxylase was isolated from *Aspergillus oryzae* (Santha, *et al.*, 1996). This enzyme also has essential histidine and cysteine residues. It does not use

any cofactors and is a homotetramer of 38 kDa subunits. A 2,3-dihydroxybenzoate decarboxylase isolated from *Trichosoporon cutaneum* is part of the catabolism of tryptophan and anthranilate in the beta-ketoadipate pathway (Anderson and Dagley, 1981). It also does not appear to have cofactors. The native enzyme is 66.1 kDa, composed of two identical 36.5 kDa subunits. The substrate must be a benzoic acid with C-2 and C-3 hydroxyl groups.

Arylmalonate decarboxylase is a non-oxidative enzyme from *Alcaligenes bronchisepticus* KU1201. It does not need cofactors, and the proposed mechanism proceeds via a thiol ester intermediate. It seems that the electron withdrawing effects of the phenyl ring substituents and the thiol ester lower the potential energy of the negatively charged transition state (Kawasaki, *et al.*, 1996; Kawasaki, *et al.*, 1997).

The majority of aromatic decarboxylases have no defined mechanism. One of the first hydroxybenzoate decarboxylases discovered is from *Klebsiella pneumoniae* (baseonym *Aerobacter aerogenes*) (Grant and Patel, 1969). More than one hydroxybenzoate decarboxylase may be involved in the decarboxylation of 4-hydroxybenzoate, gentisate, protocatechuate, and gallate. The protein(s) have not been purified but have no apparent cofactors. A couple of hydroxybenzoate decarboxylases have been isolated from *Lactobacillus plantarum*, an important bacterium in the food industry responsible for flavoring in wine and other food products. Other aromatic decarboxylases include *p*-coumarate decarboxylase and a ferulic acid decarboxylase. *p*-Coumarate is a lignin constituent with an acrylic acid in the *para* position from the hydroxyl group on the benzene ring. Ferulic acid is defined as a substituted cinnamic acid and is widely found in plants, specifically in lignin. There is evidence of three

separate pathways for *p*-coumarate degradation (Barthelmebs, *et al.*, 2000; Cavin, *et al.*, 1997). Neither cofactors nor metals are needed for the *p*-coumarate decarboxylase, which is a homotetramer of a 23.5 kDa subunit. It is not oxygen sensitive, and no mechanism has been proposed. A homologous phenolic acid decarboxylase from *Pediococcus pentosaceus* was found to be inducible and in an autoregulated operon (Barthelmebs, *et al.*, 2000). A phenolic acid decarboxylase was purified from *Bacillus subtilis*. This enzyme decarboxylated ferulic, *p*-coumaric, and caffeic acids (Cavin, *et al.*, 1998) A ferulic acid decarboxylase from *Bacillus pumilus* was cloned and sequenced but no mechanism was proposed (Zago, *et al.*, 1995).

Vanillate decarboxylase cloned from *Streptomyces* sp.D7 needs three gene products for activity. The genes *vdcBCD* are transcribed as a polycistronic message and encode products of 201 (VdcB), 475 (VdcC), and 80 (VdcD) amino acid residues. The genes were identified using the N-terminal amino acid sequences of proteins that displayed increased expression in the presence of vanillate. The amino acid sequence was used to create nucleic acid probes to screen a library and identify the three ORFs (Chow, *et al.*, 1999). The genes were cloned and transformed into *Streptomyces lividans*1326, and the recombinant enzyme catalyzed the decarboxylation of vanillate to guaiacol (2-methoxyphenol) in a highly specific manner.

Oxygen- Sensitive Hydroxyarylic Acid Decarboxylases

A number of the decarboxylases are oxygen sensitive, although the reason for this oxygen sensitivity is not known since the reaction mechanisms have not been elucidated. Several oxygen-sensitive hydroxybenzoate decarboxylases are described below.

A mechanism using an organic acid cofactor was proposed for a reversible pyrrole-2-carboxylate decarboxylase that fixes carbon dioxide in *Bacillus megaterium* PYR2910 (Wieser, *et al.*, 1998). The proposed carboxylase mechanism involves an organic acid attack at the nitrogen of the pyrrole and an electrophilic substitution on the adjacent carbon. The best carboxyl- source is HCO_3^-, which is the limiting factor in the reverse reaction. Enzyme activity requires anaerobic or microaerobic conditions. The proposed mechanism has not been confirmed.

Some of the oxygen-sensitive decarboxylases have a metal cofactor, which could contribute to their oxygen sensitivity. For example, a number of enzymes containing iron-sulfur clusters are known to be oxygen-sensitive (Flint and Allen, 1996). Often this occurs when the iron is not completely sequestered by sulfur groups, leaving it susceptible to attack by oxygen. An irreversible gallic acid decarboxylase that requires iron was purified from *Pantoea agglomerans* T71 (Zeida, *et al.*, 1998). It is a homohexamer with a native molecular weight of 320 kDa, and it is very specific for gallic acid. There is also a gallic acid decarboxylase from *Pelobacter acidigallici* that needs magnesium (Samain, *et al.*, 1986). A *p*-cresol-yielding *p*-hydroxyphenylacetate decarboxylase from the human pathogen *Clostridium dificile* requires no typical cofactors, although it seems to need a low molecular weight fraction that may contain a small iron-sulfur protein. This enzyme is believed to be a glycyl-radical enzyme which is post-translationally activated by an iron-sulfur protein. Catalysis occurs when a glycine residue, which is located in a fingerprint sequence, transfers its radical to a cysteinyl residue in the active site. A special apo-protein, believed to be the iron-sulfur protein encoded by a gene that is close to that of the enzyme, is needed to introduce the

glycyl radical. The decarboxylation proceeds when the thiyl radical and the hydroxyl group in the *para* position of the phenyl ring allow a radical formation by homolytic cleavage of the hydroxyl bond. The phenolate is not oxidized. The hydrogen abstraction by glycyl radical enzymes has been discussed for benzylsuccinate synthase and ribonucleotide reductase (Selmer and Andrei, 2001)

Many nonoxidative aromatic acid decarboxylases have no cofactor requirement, and none of them have been crystallized. The mechanisms for these enzymes is unknown, although several possibilities have been suggested. One such decarboxylase is a 4-hydroxybenzoate decarboxylase that was purified from an anaerobic coculture. It is a homotetramer composed of 119 kDa subunits (Li, *et al.*, 2000). Another example is the catechol carboxylase from *Desulfobacterium* sp. strain Cat2 which allows the anaerobic degradation of catechol (Gorny and Schink, 1994). A fermenting culture of two marine bacteria derived from one strain, Pe23DHB, was able to grow by the decarboxylation of 2,3-dihydroxybenzoate to catechol, which was not degraded futher (Ostermann, *et al.*, 1997).

An enzyme identified in *Moorella thermoacetica* (formerly *Clostridium thermoaceticum*) decarboxylates a variety of carboxylated aromatic compounds (Hsu, *et al.*, 1990). It has a pH optimum around 6.0 at 55°C and only decarboxylates benzoates with hydroxyl groups in the *para* position and substituents at the *meta* positions. *M. thermoacetica* uses this enzyme to produce CO_2 from 4-hydroxybenzoate under CO_2-limiting conditions.

Two enzymes were isolated from *Sedimentibacter hydroxybenzoicus* JW/Z-1 (formerly *Clostridium hydroxybenzoicum* JW/Z-1) as a part of the degradation pathway

of 2,4-dichlorophenol (He and Wiegel, 1995; He and Wiegel, 1996; Zhang and Wiegel, 1990). They are both oxygen sensitive, reversible, and do not appear to use any cofactors. The 4-hydroxybenzoate decarboxylase (ShdC, named for *Sedimentibacter hydroxybenzoicus* decarboxylase) was believed to be a 350 kDa homohexamer of 57 kDa subunits (see next chapter for more information). It is able to decarboxylate a number of compounds as long as there is a hydroxyl group in the *para* position of the benzoate ring. It was designated a 4-hydroxybenzoate decarboxylase as this was the compound for which activity was first determined, although ShdC also decarboxylates vanillate. *S. hydroxybenzoicus* also contains a 3,4-dihydroxybenzoate decarboxylase (Shd34), which is a separate enzyme from ShdC. It is induced by and specific for 3,4-dihydroxybenzoate (He and Wiegel, 1996; Zhang, *et al.*, 1994). Shd34 is a 270 kDa pentamer of 57 kDa subunits.

The induced activities of ShdC and Shd34 are additive, which is evidence of there being two separate enzymes with the same function (Zhang and Wiegel, 1994). The decarboxylases were also both purified from *S. hydroxybenzoicus*. Both decarboxylases are reversible, with their equilibria favoring the decarboxylation activity. Carbon dioxide, not bicarbonate, is used as the carbon source for the reverse carboxylating activity for these enzymes. ShdC operates optimally at a temperature of 50°C and pH 5.6-6.2 with a K_{cat} at 3.3 x 10^3 per minutes at 25°C and pH 6.0 and a Km of 0.4 mM for the decarboxylation of 4-hydroxybenzoate. The purified enzyme, ShdC, is oxygen-sensitive, losing 50% of its activity after 2 hours of exposure to air at 5°C. Purified Shd34 is also oxygen-sensitive, losing 50% of its activity after 10 minutes exposure to air

at 25°C. Its activity optima are at pH 7.0 and 50°C, and it only decarboxylates a benzoate with hydroxyl groups at the 3 and 4 positions.

Description of the Project

The 4-hydroxybenzoate decarboxylase is different from previously described decarboxylases in its properties. The enzyme was originally believed to be a homohexamer based on the purified protein (He and Wiegel, 1995). In subsequent research the gene for *S. hydroxybenzoicus*' decarboxylase was sequenced. The gene was identified in a *Hin*dIII fragment of genomic DNA which had been cloned into pUC18 and expressed in *E. coli* DH5α (Huang, *et al.*, 1999). This plasmid was termed pJBH-1. An ORF in the *Hin*dIII fragment, termed *ohb*1, was designated as the coding sequence, and the protein product of pJBH-1 as expressed in *E. coli* DH5α showed decarboxylase activity. The 4-hydroxybenzoate decarboxylase from *S. hydroxybenzoicus* JW/Z-1 is purported to be a member of a novel enzyme family. The goal of this project is to identify other members of this proposed novel enzyme family of hydroxyarylic acid decarboxylases.

In Chapter 2 of this thesis, attempts to overexpress the protein after subcloning *ohb*1 yielded inactive protein. Further scrutiny of the *Hin*dIII fragment containing *ohb*1 revealed a partial ORF upstream and a complete 204 base pair ORF downstream of the putative gene. The original gene, *ohb*1, was renamed *shdC* in this study, and the ORF downstream was named *shdD* (Huang, *et al.*, 1999). Both *shd*C and *shd*D were needed for expression of active enzyme.

Continued in Chapter 2, sequence homology searches of ShdC revealed a number of hypothetical proteins but only one which represented part of an identified decarboxylase. This homologue was the vanillic acid decarboxylase (VdcBCD) from *Streptomyces* sp. D7 described earlier in the introduction (Chow, *et al.*, 1999). Homologues from three different bacteria with the same gene arrangement (with the homologues to *shdC* and *shdD* adjacent) were identified. They are from *Streptomyces* sp.D7 (VdcCD), *Bacillus subtilis* (BsdCD), and *Escherichia coli* O157:H7 (EcdCD). The genes encoding the three homologues were cloned, expressed, and assayed for decarboxylase activity. The homologues from *Streptomyces* sp.D7 and *B. subtilis* are vanillate decarboxylases, and the one from *Escherichia coli* O157:H7 is a 4-hydroxybenzoate decarboxylase.

Chapter 3 provides a description of the partial purification of ShdCD and BsdCD. An extraneous protein in the purification raises the question of a third subunit needed for activity. Based on this research, it is proposed that VdcCD, BsdCD, EcdCD, and ShdCD represent a novel enzyme family of hydroxyarylic acid decarboxylases. The members of this enzyme family share high amino acid sequence identity, catalyse the decarboxylation of hydroxyarylic acids, appear to use no common decarboxylase cofactors, and are mainly oxygen-sensitive.

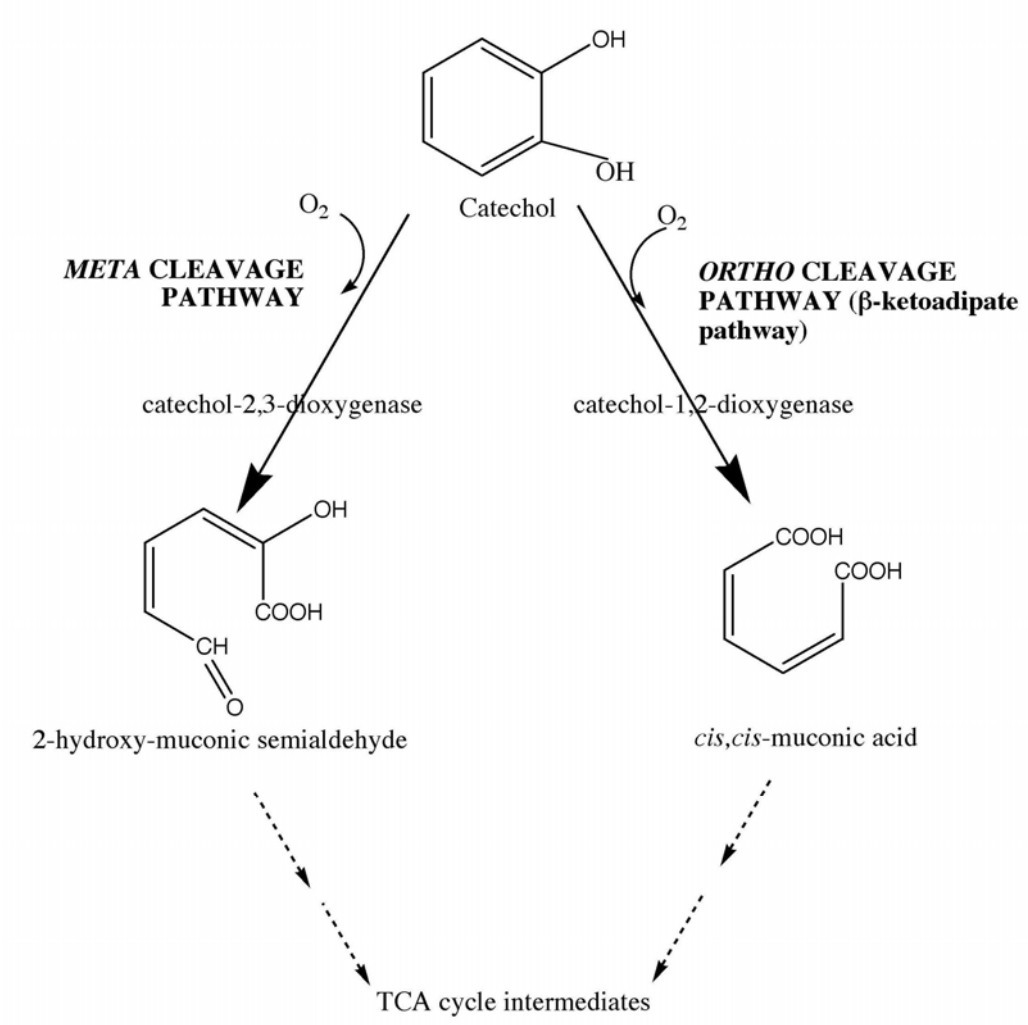
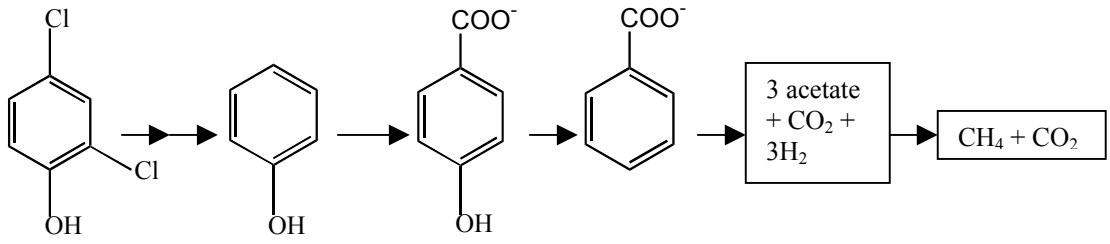


Figure 1.1- Catechol degradation pathways



2,4-Dichlorophenol Phenol 4-Hydroxybenzoate Benzoate

Figure 1.2- Proposed degradation pathway of 2,4-dichlorophenol in methanogenic freshwater sediments
The first two steps are reductive dechlorinations performed by *Desulfitobacterium dehalogenans* and other unknown microbes. The phenol is carboxylated by *S. hydroxybenzoicus*, and the 4-hydroxybenzoate is dehydroxylated by an unknown organism. The ring is cleaved open by *Methanosarcina barkeri*, a *Methanospirillum hungatii*-like organism to maintain low hydrogen concentrations, and a syntrophic anaerobe. The final methanogenesis is performed by *M. hungatii*- like and *M. barkeri*-like organisms (Zhang and Wiegel, 1990).

Table 1.1 - Hydroxyarylic acid decarboxylases with known cofactors

Cofactors	Enzyme Name	Organism of Origin	MW (kDa)	Characteristics
Thiamine Pyrophosphate	Benzylformate decarboxylase			
Pyridoxal phosphate	aminobenzoate decarboxylase			
	aromatic L-amino acid decarboxylase			
Organic acid	pyrrole-2-carboxylate decarboxylase (Wieser, *et al.*, 1998)	*Bacillus megaterium* PYR2910	2 x 52 kDa subunits	oxygen sensitive; mechanism unknown
Metal	4-hydroxybenzoate decarboxylase (Gallert and Winter, 1992)	anaerobic consortium	unknown	at least 2 separate decarboxylases; requires potassium and a divalent cation
	3,4-dihydroxyphthalate decarboxylase (Eaton, 2001)	*Arthrobacter keyseri* 12B	unknown	similar to aldolases which have histidine residues to coordinate catalytic zinc
	p-hydroxy-phenylacetate decarboxylase (Selmer and Andrei, 2001)	*Clostridium dificile*	110 + 105 = 200 kDa	oxygen-sensitive, has small iron-sulfur protein, believed to be a glycyl radical enzyme
	3-octaprenyl-4-hydroxybenzoate decarboxylase (Zhang and Javor, 2000)	*Escherichia coli* K-12	55.6 subunits = 340 kDa	needs small molecular weight cofactor and Mn^{2+}

Table 1.2- Hydroxyarylic acid decarboxylases with no known cofactors/mechanisms

Enzyme	Organism of Origin	MW (kDa)	Characteristics
4-hydroxybenzoate decarboxylase	*Aerobacter aerogenes* (Grant and Patel, 1969)	unknown	decarboxylates gentisate, protocatechuate, and gallate
	Sedimentibacter hydroxybenzoicus (He and Wiegel, 1995)	57	oxygen sensitive, decarboxylates 3,4-dihydroxybenzoate and vanillic acid
	anaerobic coculture (Li, *et al.*, 2000)	4 x 119 = 420	oxygen sensitive, reversible, also decarboxylates 3,4-dihydroxybenzoate
2,3-dihydroxybenzoate decarboxylase	*Aspergillus niger* (Santha, *et al.*, 1995)	4 x 28	mechanism involves essential histidine and cysteine
	Aspergillus oryzae (Santha, *et al.*, 1996)	4 x 38	mechanism involves histidine and cysteine
	Trichosporon cutaneum (Anderson and Dagley, 1981)	2 x 36.5	decarboxylates benzoates with *meta* and *ortho* hydroxyl groups
3,4-dihydroxybenzoate decarboxylase	*Sedimentibacter hydroxybenzoicus* (He and Wiegel, 1996)	52	oxygen sensitive
vanillic acid decarboxylase	*Streptomyces* sp.D7 (Chow, *et al.*, 1999)	36, 52, 9	specific to vanillic acid, uncertain subunit configuration
4,5-dihydroxyphthalate decarboxylase	*Burkholderia cepacia* (Chang and Zylstra, 1998; Pujar and Ribbons, 1985)	6 x 66 = 420	involved in phthalate degradation
	Pseudomonas testosteroni (Nakazawa and Hayashi, 1978)	4 x 38 = 150	involved in phthalate degradation, also decarboxylates 4-hydroxyphthalate
arylmalonate decarboxylase	*Alcaligenes bronchisepticus* KU1201 (Kawasaki, *et al.*, 1996; Kawasaki, *et al.*, 1997)	unknown	mechanism proceeds via thiol ester intermediate
p-coumarate decarboxylase	*Lactobacillus plantarum* (Cavin, *et al.*, 1997)	4 x 23.5 = 93	2 phenolic acid decarboxylases in this organism, decarboxylates caffeic acids also
ferulic acid decarboxylase	*Lactobacillus plantarum* (Barthelmebs, *et al.*, 2000)	unknown	2 phenolic acid decarboxylases in this organism, also decarboxylates *p*-coumarate
	Bacillus pumilus (Zago, *et al.*, 1995)	2 x 21.5= 42	also decarboxylates *p*-coumarate
phenolic acid decarboxylase	*Pediococcus pentosaceus* (Barthelmebs, *et al.*, 2000)	25	involved in phenolic acid metabolism
	Bacillus subtilis (Cavin, *et al.*, 1998)	2 x 22 = 45	decarboxylates ferulic, *p*-coumaric, and caffeic acids
carboxylated aromatic decarboxylase	*Moorella thermoacetica* (Hsu, *et al.*, 1990)	unknown	oxygen-sensitive, decarboxylates benzoates with a *para* hydroxyl group and any *meta* substituent

Literature cited

1. **Alvarado, I. E., A. Lomascolo, D. Navarro, M. Delattre, M. Asther, and L. Lesage-Meessen.** 2001. Evidence of a new biotransformation pathway of *p*-coumaric acid into *p*-hydroxybenzaldehyde in *Pycnoporus cinnabarinus*. Appl. Microbiol. Biotechnol. **57** (5/6):725-730.
2. **Anderson, J. J., and S. Dagley.** 1981. Catabolism of tryptophan, anthranilate, and 2,3-dihydroxybenzoate in *Trichosporon cutaneum*. J. Bacteriol. **146**(1)**:**291-7.
3. **Aresta, M., E. Quaranta, R. Liberio, C. Dileo, and I. Tommasi.** 1998. Enzymatic synthesis of 4-OH-benzoic acid from phenol and CO_2: the first example of a biotechnological application of a carboxylase enzyme. Tetrahedron **54:**8841-8846.
4. **Barthelmebs, L., C. Divies, and J. F. Cavin.** 2000. Knockout of the *p*-coumarate decarboxylase gene from *Lactobacillus plantarum* reveals the existence of two other inducible enzymatic activities involved in phenolic acid metabolism. Appl. Env. Microbiol. **66**(8)**:**3368-75.
5. **Barthelmebs, L., B. Lecomte, C. Divies, and J. F. Cavin.** 2000. Inducible metabolism of phenolic acids in *Pediococcus pentosaceus* is encoded by an autoregulated operon which involves a new class of negative transcriptional regulator. J. Bacteriol. **182**(23)**:**6724-31.
6. **Beyer, H., and W. Walter.** 1996. Handbook of Organic Chemistry. Prentice Hall, New York.

7. **Biegert, T., G. Fuchs, and J. Heider.** 1996. Evidence that anaerobic oxidation of toluene in the denitrifying bacterium *Thauera aromatica* is initiated by formation of benzylsuccinate from toluene and fumarate. Eur. J. Biochem. **238**(3)**:**661-8.

8. **Blanchard, C. Z., Y. M. Lee, P. A. Frantom, and G. L. Waldrop.** 1999. Mutations at four active site residues of biotin carboxylase abolish substrate-induced synergism by biotin. Biochemistry **38**(11)**:**3393-400.

9. **Boll, M., S. S. Albracht, and G. Fuchs.** 1997. Benzoyl-CoA reductase (dearomatizing), a key enzyme of anaerobic aromatic metabolism. A study of adenosinetriphosphatase activity, ATP stoichiometry of the reaction and EPR properties of the enzyme. Eur. J. Biochem. **244**(3)**:**840-51.

10. **Brackmann, R., and G. Fuchs.** 1993. Enzymes of anaerobic metabolism of phenolic compounds. 4-Hydroxybenzoyl- CoA reductase (dehydroxylating) from a denitrifying *Pseudomonas* species. Eur. J. Biochem. **213**(1)**:**563-71.

11. **Breese, K., and G. Fuchs.** 1998. 4-Hydroxybenzoyl-CoA reductase (dehydroxylating) from the denitrifying bacterium *Thauera aromatica*--prosthetic groups, electron donor, and genes of a member of the molybdenum-flavin-iron-sulfur proteins. Eur. J. Biochem. **251**(3)**:**916-23.

12. **Breinig, S., E. Schiltz, and G. Fuchs.** 2000. Genes involved in anaerobic metabolism of phenol in the bacterium *Thauera aromatica*. J. Bacteriol. **182**(20)**:**5849-63.

13. **Breitenstein, A., J. Wiegel, C. Haertig, N. Weiss, J. R. Andreesen, and U. Lechner.** 2002. Reclassification of *Clostridium hydroxybenzoicum* as

Sedimentibacter hydroxybenzoicus gen. nov., comb. nov. and description of *Sedimentibacter saalensis* sp. nov. Int. J. Syst. Evol. Microbiol. **52**:801-807.

14. **Cavin, J. F., L. Barthelmebs, J. Guzzo, J. Van Beeumen, B. Samyn, J. F. Travers, and C. Divies.** 1997. Purification and characterization of an inducible *p*-coumaric acid decarboxylase from *Lactobacillus plantarum*. FEMS Microbiol. Lett. **147**(2)**:**291-295.

15. **Cavin, J. F., V. Dartois, and C. Divies.** 1998. Gene cloning, transcriptional analysis, purification, and characterization of phenolic acid decarboxylase from *Bacillus subtilis*. Appl. Environ. Microbiol. **64**(4)**:**1466-71.

16. **Chang, H. K., and G. J. Zylstra.** 1998. Novel organization of the genes for phthalate degradation from *Burkholderia cepacia* DBO1. J. Bacteriol. **180**(24)**:**6529-37.

17. **Chapman, and Hall.** 1996. Dictionary of Organic Compounds, sixth ed. Chapman and Hall, Cambridge.

18. **Chow, K. T., M. K. Pope, and J. Davies.** 1999. Characterization of a vanillic acid non-oxidative decarboxylation gene cluster from *Streptomyces* sp. D7. Microbiology **145**(Pt 9)**:**2393-403.

19. **Eaton, R. W.** 2001. Plasmid-encoded phthalate catabolic pathway in *Arthrobacter keyseri* 12B. J. Bacteriol. **183**(12)**:**3689-703.

20. **Egland, P. G., and C. S. Harwood.** 2000. HbaR, a 4-hydroxybenzoate sensor and FNR-CRP superfamily member, regulates anaerobic 4-hydroxybenzoate degradation by *Rhodopseudomonas palustris*. J. Bacteriol. **182**(1)**:**100-6.

21. **El Fantroussi, S., H. Naveau, and S. N. Agathos.** 1998. Anaerobic dechlorinating bacteria. Biotechnol. Prog. **14**(2)**:**167-88.

22. **Elder, D. J., and D. J. Kelly.** 1994. The bacterial degradation of benzoic acid and benzenoid compounds under anaerobic conditions: unifying trends and new perspectives. FEMS Microbiol. Rev. **13**(4)**:**441-68.

23. **Fetzner, S.** 1998. Bacterial dehalogenation. Appl. Microbiol. Biotechnol. **50**(6)**:**633-57.

24. **Flint, D. H., and R. M. Allen.** 1996. Iron-sulfur proteins with nonredox functions. Chem. Rev. **96**(7)**:**2315-2334.

25. **Gallert, C., and J. Winter.** 1992. Comparison of 4-hydroxybenzoate decarboxylase and phenol carboxylase activities in cell-free extracts of a defined, 4-hydroxybenzoate and phenol-degrading anaerobic consortium. Appl. Microbiol. Biotechnol. **37:**119-124.

26. **Gorny, N., and B. Schink.** 1994. Anaerobic degradation of catechol by *Desulfobacterium* sp. strain Cat2 proceeds via carboxylation to protocatechuate. Appl. Environ. Microbiol. **60**(9)**:**3396-400.

27. **Grant, D. J., and J. C. Patel.** 1969. The non-oxidative decarboxylation of *p*-hydroxybenzoic acid, gentisic acid, protocatechuic acid and gallic acid by *Klebsiella aerogenes* (*Aerobacter aerogenes*). Antonie Leeuwenhoek. **35**(3)**:**325-43.

28. **Haddock, J. D., and J. G. Ferry.** 1993. Initial steps in the anaerobic degradation of 3,4,5-trihydroxybenzoate by *Eubacterium oxidoreducens*: characterization of mutants and role of 1,2,3,5-tetrahydroxybenzene. J. Bacteriol. **175**(3)**:**669-73.

29. **Harbison, R. D.** 1998. Hamilton & Hardy's Industrial Toxicology, Fifth ed. Mosby, New York.

30. **Harwood, C. S., G. Burchhardt, H. Herrmann, and G. Fuchs.** 1999. Anaerobic metabolism of aromatic compounds via the benzoyl-CoA pathway. FEMS Microbiol. Rev. **22:**439-458.

31. **He, Z., and J. Wiegel.** 1995. Purification and characterization of an oxygen-sensitive reversible 4- hydroxybenzoate decarboxylase from *Clostridium hydroxybenzoicum*. Eur. J. Biochem. **229**(1)**:**77-82.

32. **He, Z., and J. Wiegel.** 1996. Purification and characterization of an oxygen-sensitive, reversible 3,4-dihydroxybenzoate decarboxylase from *Clostridium hydroxybenzoicum*. J. Bacteriol. **178**(12)**:**3539-43.

33. **Heider, J., M. Boll, K. Breese, S. Breinig, C. Ebenau-Jehle, U. Feil, N. Gad'on, D. Laempe, B. Leuthner, M. E. Mohamed, S. Schneider, G. Burchhardt, and G. Fuchs.** 1998. Differential induction of enzymes involved in anaerobic metabolism of aromatic compounds in the denitrifying bacterium *Thauera aromatica*. Arch. Microbiol. **170**(2)**:**120-31.

34. **Horton, H. R., L. A. Moran, R. S. Ochs, J. D. Rawn, and K. G. Scrimgeour.** 1996. Principles of Biochemistry, 2nd ed. Prentice Hall, Upper Saddle River, NJ.

35. **Hsu, T., S. L. Daniel, M. F. Lux, and H. L. Drake.** 1990. Biotransformations of carboxylated aromatic compounds by the acetogen *Clostridium thermoaceticum*: generation of growth-supportive CO_2 equivalents under CO_2-limited conditions. J. Bacteriol. **172**(1)**:**212-7.

36. **Hsu, T. D., M. F. Lux, and H. L. Drake.** 1990. Expression of an aromatic-dependent decarboxylase which provides growth- essential CO_2 equivalents for the acetogenic (Wood) pathway of *Clostridium thermoaceticum*. J. Bacteriol. **172**(10)**:**5901-7.

37. **Huang, J., Z. He, and J. Wiegel.** 1999. Cloning, characterization, and expression of a novel gene encoding a reversible 4-hydroxybenzoate decarboxylase from *Clostridium hydroxybenzoicum*. J. Bacteriol. **181**(16)**:**5119-22.

38. **Kamath, A. V., D. Dasgupta, and C. S. Vaidyanathan.** 1987. Enzyme-catalysed non-oxidative decarboxylation of aromatic acids: I. Purification and spectroscopic properties of 2,3 dihydroxybenzoic acid decarboxylase from *Aspergillus niger*. Biochem. Biophys. Res. Commun. **145**(1)**:**586-95.

39. **Kamath, A. V., and C. S. Vaidyanathan.** 1990. New pathway for the biodegradation of indole in *Aspergillus niger*. Appl. Environ. Microbiol. **56**(1)**:**275-80.

40. **Kanaly, R. A., and S. Harayama.** 2000. Biodegradation of high-molecular-weight polycyclic aromatic hydrocarbons by bacteria. J. Bacteriol. **182**(8)**:**2059-67.

41. **Kawasaki, T., E. Horimai, and H. Ohta.** 1996. On the conformation of the substrate binding to the active site during the course of enzymatic decarboxylation. Bull. Chem. Soc. Jpn. **69:**3591-3594.

42. **Kawasaki, T., K. Saito, and H. Ohta.** 1997. The mode of substrate-recognition mechanism of arylmalonate decarboxylase. Chem. Lett. **:**351-352.

43. **Knoll, G., and J. Winter.** 1989. Degradation of phenol via carboxylation to benzoate by a defined, obligate syntrophic consortium of anaerobic bacteria. Appl. Microbiol. Biotechnol. **20:**318-324.

44. **Lack, A., and G. Fuchs.** 1992. Carboxylation of phenylphosphate by phenol carboxylase, an enzyme system of anaerobic phenol metabolism. J. Bacteriol. **174**(11)**:**3629-36.

45. **Lack, A., and G. Fuchs.** 1994. Evidence that phenol phosphorylation to phenylphosphate is the first step in anaerobic phenol metabolism in a denitrifying *Pseudomonas* sp. Arch. Microbiol. **161**(2)**:**132-9.

46. **Laempe, D., M. Jahn, K. Breese, H. Schagger, and G. Fuchs.** 2001. Anaerobic metabolism of 3-hydroxybenzoate by the denitrifying bacterium *Thauera aromatica*. J. Bacteriol. **183**(3)**:**968-79.

47. **Li, T., P. Juteau, R. Beaudet, F. Lepine, R. Villemur, and J. G. Bisaillon.** 2000. Purification and characterization of a 4-hydroxybenzoate decarboxylase from an anaerobic coculture. Can. J. Microbiol. **46**(9)**:**856-9.

48. **Londry, K. L., and P. M. Fedorak.** 1992. Benzoic acid intermediates in the anaerobic biodegradation of phenols. Can. J. Microbiol. **38**(1)**:**1-11.

49. **Mishra, V., R. Lal, and Srinivasan.** 2001. Enzymes and operons mediating xenobiotic degradation in bacteria. Crit. Rev. Microbiol. **27**(2)**:**133-66.

50. **Muller, J. A., and B. Schink.** 2000. Initial steps in the fermentation of 3-hydroxybenzoate by *Sporotomaculum hydroxybenzoicum*. Arch. Microbiol. **173**(4)**:**288-95.

51. **Nakazawa, T., and E. Hayashi.** 1978. Phthalate and 4-hydroxyphthalate metabolism in *Pseudomonas testosteroni*: purification and properties of 4,5-dihydroxyphthalate decarboxylase. Appl. Environ. Microbiol. **36**(2)**:**264-9.

52. **Narbad, A., and M. J. Gasson.** 1998. Metabolism of ferulic acid via vanillin using a novel CoA-dependent pathway in a newly-isolated strain of *Pseudomonas fluorescens*. Microbiology **144**(Pt 5)**:**1397-405.

53. **O'Leary, M. H.** 1992. Catalytic strategies in enzymatic carboxylation and decarboxylation, p. 235-269, The Enzymes XX, vol. XX. Academic Press.

54. **Ostermann, A., C. Gallus, and B. Schink.** 1997. Decarboxylation of 2,3-dihydroxybenzoate to catechol supports growth of fermenting bacteria. Curr. Microbiol. **35:**270-273.

55. **Philip, R. B.** 2001. Ecosystems and Human Health Toxicology and Environmental Hazards, second ed. Lewis Publishers, New York.

56. **Pujar, B. G., and D. W. Ribbons.** 1985. Phthalate metabolism in *Pseudomonas fluorescens* PHK: purification and properties of 4,5-dihydroxyphthalate decarboxylase. Appl. Environ. Microbiol. **49**(2)**:**374-6.

57. **Samain, E., G. Albagnac, and H. C. Dubourguier.** 1986. Initial steps of catabolism of trihydroxybenzenes in *Pelobacter acidigallici*. Arch. Microbiol. **144:**242-244.

58. **Santha, R., N. A. Rao, and C. S. Vaidyanathan.** 1996. Identification of the active-site peptide of 2,3-dihydroxybenzoic acid decarboxylase from *Aspergillus oryzae*. Biochim. Biophys. Acta. **1293**(2)**:**191-200.

59. **Santha, R., H. S. Savithri, N. A. Rao, and C. S. Vaidyanathan.** 1995. 2,3-Dihydroxybenzoic acid decarboxylase from *Aspergillus niger*. A novel decarboxylase. Eur. J. Biochem. **230**(1)**:**104-10.

60. **Schink, B., B. Philipp, and J. Muller.** 2000. Anaerobic degradation of phenolic compounds. Naturwissenschaften. **87**(1)**:**12-23. 0.htm.

61. **Schneider, S., M. El-Said Mohamed, and G. Fuchs.** 1997. Anaerobic metabolism of L-phenylalanine via benzoyl-CoA in the denitrifying bacterium *Thauera aromatica*. Arch. Microbiol. **168:**310-320.

62. **Schneider, S., and G. Fuchs.** 1998. Phenylacetyl-CoA:acceptor oxidoreductase, a new alpha-oxidizing enzyme that produces phenylglyoxylate. Assay, membrane localization, and differential production in *Thauera aromatica*. Arch. Microbiol. **169**(6)**:**509-16.

63. **Schocke, L., and B. Schink.** 1999. Energetics and biochemistry of fermentative benzoate degradation by *Syntrophus gentianae*. Arch. Microbiol. **171:**331-337.

64. **Selmer, T., and P. I. Andrei.** 2001. *p*-Hydroxyphenylacetate decarboxylase from *Clostridium difficile*. A novel glycyl radical enzyme catalysing the formation of *p*-cresol. Eur. J. Biochem. **268**(5)**:**1363-1372.

65. **Semple, K. T., R. B. Cain, and S. Schmidt.** 1999. Biodegradation of aromatic compounds by microalgae. FEMS Microbiol. Lett. **170:**291-300.

66. **Timmis, K. N., and D. H. Pieper.** 1999. Bacteria designed for bioremediation. Trends Biotechnol. **17**(5)**:**200-204.

67. **Valkova, N., F. Lepine, L. Valeanu, M. Dupont, L. Labrie, J. G. Bisaillon, R. Beaudet, F. Shareck, and R. Villemur.** 2001. Hydrolysis of 4-hydroxybenzoic

acid esters (parabens) and their aerobic transformation into phenol by the resistant *Enterobacter cloacae* strain EM. Appl. Environ. Microbiol. **67**(6)**:**2404-9.

68. **Wieser, M., N. Fujii, T. Yoshida, and T. Nagasawa.** 1998. Carbon dioxide fixation by reversible pyrrole-2-carboxylate decarboxylase from *Bacillus megaterium* PYR2910. Eur. J. Biochem. **257**(2)**:**495-9.

69. **Wright, J. D.** 1993. Fungal degradation of benzoic acid and related compounds. World. J. Microbiol. Biotechnol. **9:**9-16.

70. **Zago, A., G. Degrassi, and C. V. Bruschi.** 1995. Cloning, sequencing, and expression in *Escherichia coli* of the *Bacillus pumilus* gene for ferulic acid decarboxylase. Appl. Environ. Microbiol. **61**(12)**:**4484-6.

71. **Zeida, M., M. Wieser, T. Yoshida, T. Sugio, and T. Nagasawa.** 1998. Purification and characterization of gallic acid decarboxylase from *Pantoea agglomerans* T71. Appl. Environ. Microbiol. **64**(12)**:**4743-7.

72. **Zhang, H., and G. T. Javor.** 2000. Identification of the *ubiD* gene on the *Escherichia coli* chromosome. J. Bacteriol. **182**(21)**:**6243-6.

73. **Zhang, X., L. Mandelco, and J. Wiegel.** 1994. *Clostridium hydroxybenzoicum*sp.nov., an amino acid-utilizing, hydroxybenzoate-decarboxylating bacterium isolated from methanogenic freshwater pond sediment. Int. J. Syst. Bacteriol. **44**(2)**:**213-222.

74. **Zhang, X., T. V. Morgan, and J. Wiegel.** 1990. Conversion of ^{13}C-1 phenol to ^{13}C-4 benzoate, an intermediate step in the anaerobic degradation of chlorophenols. FEMS Microbiol. Lett. **67:**63-66.

75. **Zhang, X., and J. Wiegel.** 1992. The anaerobic degradation of

3-chloro-4-hydroxybenzoate in freshwater sediment proceeds via either chlorophenol or hydroxybenzoate to phenol and subsequently to benzoate. Appl. Env. Microbiol. **58**(11)**:**3580-3585.

76. **Zhang, X., and J. Wiegel.** 1990. Isolation and partial characterization of a *Clostridium* species transforming *para*-hydroxybenzoate and 3,4-dihydroxybenzoate and producing phenols as the final transformation products. Microbiol. Ecol. **20:**103-121.

77. **Zhang, X., and J. Wiegel.** 1994. Reversible conversion of 4-hydroxybenzoate and phenol by *Clostridium hydroxybenzoicum*. Appl. Environ. Microbiol. **60**(11)**:**4182-4185.

78. **Zhang, X., and J. Wiegel.** 1990. Sequential anaerobic degradation of 2,4-dichlorophenol in freshwater sediments. Appl. Environ. Microbiol. **56**(4)**:**1119-27.

79. **Zhang, X., and L. Y. Young.** 1997. Carboxylation as an initial reaction in the anaerobic metabolism of naphthalene and phenanthrene by sulfidogenic consortia. Appl. Environ. Microbiol. **63**(12)**:**4759-64.

CHAPTER 2

CHARACTERIZING A PROPOSED NOVEL ENZYME FAMILY OF REVERSIBLE HYDROXYARYLIC ACID DECARBOXYLASES

Introduction

Several genes encoding members of a proposed hydroxyarylic acid decarboxylase family were cloned, expressed, and assayed for decarboxylation activity. The proposed members are a 4-hydroxybenzoate decarboxylase (ShdCD) from *Sedimentibacter hydroxybenzoicus* JW/Z-1 (formerly *Clostridium hydroxybenzoicum* JW/Z-1), a vanillate decarboxylase (VdcCD) from *Streptomyces* sp.D7, a vanillate decarboxylase (BsdCD) from *Bacillus subtilis*, and a 4-hydroxybenzoate decarboxylase (EcdCD) from *Escherichia coli* O157:H7. The designation of the individual enzymes depends on the relevant substrate decarboxylated. For example, ShdCD was identified based on its ability to decarboxylate 4-hydroxybenzoate and therefore designated as a 4-hydroxybenzoate decarboxylase. ShdCD and VdcCD were known decarboxylases, and their encoding genes are characterized in this study. The BsdCD and EcdCD, previously designated as hypothetical proteins in the genome sequence, are both characterized as decarboxylases for the first time in this research, and their substrate ranges are established.

4-hydroxybenzoate decarboxylase (ShdCD) from *Sedimentibacter hydroxybenzoicus* JW/Z-1

The bacterium *S. hydroxybenzoicus* JW/Z-1 is a Gram-type positive, anaerobic heterotroph (Zhang, *et al.*, 1994). A 4-hydroxybenzoate decarboxylase from *S. hydroxybenzoicus* JW/Z-1 was previously purified and characterized (He and Wiegel, 1995). The most interesting facets of ShdCD are its oxygen sensitivity and the apparent lack of cofactors needed for enzyme activity. A sodium dodecyl sulfate polyacrylamide

gel electrophoresis (SDS-PAGE) analysis of the purified native protein revealed one 57 kDa band, suggesting the protein was a homohexamer since a native PAGE of the protein revealed a molecular weight of 350 kDa (He and Wiegel, 1995). N-terminal and internal amino acid sequences were obtained from this protein, and these were used to design PCR primers to amplify a segment of the encoding gene. The PCR products were used as a probe against a *Hin*dIII-digested library of *S. hydroxybenzoicus* genomic DNA. The probe identified a 2.3 kb fragment, which upon cloning and expression in *Escherichia coli* DH5α, yielded an active 4-hydroxybenzoate decarboxylase (Huang, *et al.*, 1999). The fragment contained a 1,440 bp open reading frame the product of which is a 57 kDa protein (ShdC). Database sequence homology searches of this gene reveal a large number of homologous hypothetical proteins, with only a few previously characterized proteins. This list of homologous proteins includes the proteins investigated in this research.

Further research described in this thesis has revealed that at least one gene besides *shdC* is required for expression of a functional 4-hydroxybenzoate decarboxylase. An additional gene downstream of *shdC*, termed *shdD*, is required to encode an active recombinant decarboxylase in *E. coli* JM109. Investigations of this additional gene may clarify both the mechanism of the enzyme and the reason for the enzyme's oxygen sensitivity.

Vanillate decarboxylase (VdcCD) from *Streptomyces* sp.D7

The organism *Streptomyces* sp.D7, a soil streptomycete, was isolated in the laboratory of Dr. Julian Davies on the basis of its ability to decarboxylate vanillate. The

vanillate decarboxylase, VdcBCD, from *Streptomyces* sp.D7 was identified by Chow *et al.* (Chow, *et al.*, 1999). The expression of vanillate decarboxylase activity in *Streptomyces lividans* 1326 requires expression of three genes encoding VdcBCD.

A subunit of this enzyme, VdcC, was identified during the sequence homology search of ShdC, and VdcCD bears high amino acid sequence identity to ShdCD. Besides sequence similarity, this enzyme resembles ShdCD in that it is a hydroxybenzoate decarboxylase, its expression is inducible by the substrate, and its subunits are of similar molecular weight. The enzyme mechanism for VdcCD is also unknown. Unlike ShdCD, however, VdcCD is not oxygen sensitive. In this research, we describe the cloning and expression of VdcCD in *E. coli* JM109. This in contrast with expresson in *S. lividans* 1326 in which all three subunits, VdcBCD, were needed for an active decarboxylase. Expression of an active protein in *E. coli* JM109 required only *vdcCD*.

Vanillate decarboxylase(BsdCD) from *Bacillus subtilis*

The hypothetical protein retrieved in the homology search of ShdC with the highest sequence identity is YclC from *B. subtilis*. *B. subtilis* is a Gram-type positive, facultatively anaerobic bacterium, and thus more closely related to *S. hydroxybenzoicus* than the species from which the other homologues came. Genes encoding homologues to ShdCD were discovered in *B. subtilis* in the same arrangement as in *S. hydroxybenzoicus* JW/Z-1; the gene encoding the C homologue was directly followed by the gene for the D homologue. The genes, previously annotated as *yclCD*, were renamed *bsdCD* (*Bacillus subtilis* decarboxylase) to maintain unified terminology in this research. This thesis

describes the characterization of the protein product of *bsdCD*, revealing that *bsdCD* encodes an oxygen-sensitive vanillate decarboxylase (BsdCD).

4-Hydroxybenzoate decarboxylase from *Escherichia coli* O157:H7

E. coli O157:H7, a Gram-type negative facultative anaerobe, is one of the pathogenic strains of *E. coli*. Homologues to ShdCD were found in *E. coli* O157:H7 but not other strains of *E. coli* during the sequence homology search. The genes encoding these homologues are in a similar arrangement as those in *Streptomyces* sp. D7, in that there are three adjacent genes which could encode a hydroxybenzoate decarboxylase. This cluster of genes is located in the *mutS-rpoS* genomic region which varies between pathogenic and non-pathogenic *E. coli* strains (Herbelin, *et al.*, 2000). This set of genes was renamed in this study. The *pad1* was renamed *ecdB* and *yclC* was renamed *ecdC*. The *shdD* homologue, which was not annotated in the literature, is named *ecdD*. The genes *ecdCD* were cloned and expressed in this study, and shown to encode an oxygen-sensitive 4-hydroxybenzoate decarboxylase.

Methods

Organisms and culture conditions

S. hydroxybenzoicus JW/Z-1 was grown in the medium described by Zhang *et al.* using the Hungate technique (Zhang, *et al.*, 1994). The *E. coli* JM109 and *E. coli* THU used for cloning was grown in Luria-Bertani broth (LB), both aerobically and anaerobically. The medium used for blue-white selection was LB + 1.5% agar +

100 μg/mL ampicillin + 40 ng/mL X-Gal. To grow cultures anaerobically, LB medium was made anaerobic by boiling while sparging with N_2. The media was immediately cooled on ice, after which 0.04 g/L each of NaS-nonahydrate and cysteine-HCl were added. The media was supplemented with 0.2% KNO_3 and 0.3% glucose. *E. coli* THU cultures were grown anaerobically in LB that had been boiled while sparging with N_2 and then after autoclaving, supplemented with 0.2% KNO_3 and 0.3% glucose.

The clones containing pNR were grown in M9 minimal media with casamino acids. For proper induction of the pNR plasmid with nitrate, this medium was sparged with argon instead of nitrogen. For other purposes in this research, nitrogen gas was used to sparge the media. The M9 minimal medium consisted of: 5.0 g casamino acids, 6.0 g Na_2HPO_4, 3.0 g KH_2PO_4, 0.5 g NaCl, and 1.0 g NH_4Cl, per 1 Liter. One hundred μL of resazurin (0.2%) was added as an indicator of anaerobicity. To make the medium anaerobic, 0.4 g NaS nonahydrate and 0.4 g cysteine-HCl were added after the medium was boiled and then cooled while sparging with the appropriate gas. The medium was kept at pH 7.0. Ten mL were anaerobically aliquoted into Balch tubes. 0.1 mL of filter sterilized solutions of 100 mM $MgSO_4$ heptahydrate, 20% glucose, 100 μg/mL thiamine-HCl, and 10 mM $CaCl_2$ were added to each tube.

Plasmids were transformed into *E. coli* JM109 and *E. coli* THU , which was made electrocompetent by the method described by Ausubel *et al.* (Ausubel, *et al.*, 1995). The strain *E. coli* THU was provided by Dr. George Javor (Loma Linda University, Loma Linda, CA). This strain has a point mutation in the *ubiX* gene.

The original clone containing the *vdcBCD* gene cluster from *Streptomyces* sp.D7 in pUC18 in *E. coli*, named pSub1, was provided by Julian Davies. The clone was grown

according to the method described by Chow *et al.* (Chow, *et al.*, 1999). The *B. subtilis*^{T} (ATCC 6051), obtained from Lee Kurtz (University of Georgia), was grown aerobically in LB medium at 37°C.

Cloning the genes

All the PCR primers used to amplify DNA in this research are listed in Table 2.1. PCR primers were designed using OLIGO 6.21 (Molecular Biology Insights Inc., Cascade, CO), and the primers were obtained from Integrated DNA Technologies Inc. (Coralville, IA). The PCR mixtures contained: 5 μL 10x buffer (Stratagene, La Jolla, CA), 2 μL 10 mM dNTP, 1 μL 25 mM forward primer, 1 μL 25 mM reverse primer, 1 μL genomic/plasmid DNA (giving a total of 3 μg of DNA), 39 μL sterile water, and 1 μL *Pfu* polymerase (Stratagene, La Jolla, CA). The PCR was performed in a MiniCycler (MJ Research, Waltham, MA), and the cycle parameters were optimized for each reaction. Genomic DNA was prepared using the High Pure PCR Template Preparation kit (Boehringer-Mannheim, Indianapolis, IA). The plasmid containing *vdcBCD* was isolated using the QiaSpin Miniprep kit (Qiagen, Valencia, CA). Blunt-end ligations were performed using T4 DNA ligase (New England Biolabs, Beverly, MA) and sticky end ligations were performed using T4 DNA ligase from Promega (Madison, WI). All restriction enzymes are from Promega or New England Biolabs. All DNA sequencing was performed by the Molecular Genetics Instrumentations Facility at the University of Georgia (Athens, GA).

***Cloning the 4-hydroxybenzoate decarboxylase from* S. hydroxybenzoicus** The PCR product of 4ohfor and 4ohrev, *shd*C, was ligated into pCR-SCRIPT (Stratagene) which had been cut with *Eco*RV. The sequences of the primers 4ohfor and 4ohrev are given in Table 2.1. The resulting plasmid, pCR-SCRIPT*shd*C, was transformed into *E. coli* JM109, and successful clones were identified by blue-white screening after plating on the selection media containing X-gal. White colonies were inoculated into 10 mL LB and grown overnight in a 37°C shaking incubator. Using the QiaSpin MiniPrep kit (Qiagen), the plasmid pCR-SCRIPT*shdC* was isolated from the grown cells. The plasmid was cut with *Not*I and *Sac*I (the *Hin*dIII site designed into the forward primer was too close to the *Not*I site on pCR-SCRIPT to be used) to release the *shdC* gene from the plasmid. The products were run on a 0.8% agarose gel, and *shdC* was excised and purified out of the gel using a QIAEXII gel extraction kit (Qiagen). The gene *shdC* was directionally ligated into pNR, which is an anaerobic overexpression vector designed by Dr. Amy Grunden and Dr. Mike Adams (Department of Biochemistry, University of Georgia, Athens, GA). This vector has a nitrate reductase promoter to control expression of the cloned gene. The gene will be expressed only under anaerobic conditions with nitrate. To allow directional cloning, pNR was cut with the same enzymes as those used to excise *shdC* out of pCR-SCRIPT. The plasmid, pNR*shdC*, was then transformed by electroporation into *E. coli* JM109. The resulting plasmid was isolated with QiaSpin (Qiagen), and a restriction digest confirmed the construction of the plasmid. The identity of the insert was confirmed by DNA sequencing. This construct was assayed for expression and activity of ShdCD by the methods described below. Due to the differing codon usage between *S. hydroxybenzoicus* and *E. coli*, the plasmid pRIL (Stratagene) was

also transformed into the *E. coli* containing pNR*shdC*. This pRIL carried the genes for several rare tRNAs in *E. coli*.

The primers 4ohCfor and 4ohDrev were used to amplify *shdCD*, and the appropriate restriction enzymes (Table 2.1) were used to digest the PCR product and pUC18. The product *shdCD* was ligated into pUC18, and the resulting plasmid, pUC18*shdCD*, was transformed into *E. coli* JM109. Transformants were screened by blue-white screening on plates containing X-gal. The pUC18*shdCD* construct was verified by restriction enzyme analysis and DNA sequencing. A similar method was used to clone *shdD*, using the primers 4ohDfor and 4ohDrev. These clones, pUC18*shdCD* and pUC18*shdD*, were assayed for expression and decarboxylase activity as described below.

Cloning the vanillate decarboxylase from **Streptomyces *sp.D7*** The primers, as listed in Table 2.1, were designed using the sequence of *vdcBCD* available from Genbank. The gene, *vdcC*, was PCR amplified with vdcCfor and vdcCrev; *vdcD* was amplified with vdcDfor and vdcDrev; and *vdcCD* was amplified with vdcCfor and vdcDrev. The PCR products were cut with the appropriate restriction enzymes (Table 2.1), and the products were directionally ligated into pUC18. The constructs were transformed into *E. coli* JM109, and positive clones were verified as before by restriction enzyme digest and sequencing.

Cloning the vanillate decarboxylase from **Bacillus subtilis** The gene *bsd*B was amplified by PCR using the primers bsdBfor and bsdBrev; *bsdC* was amplified by bsdCfor and bsdCrev; *bsdD* was amplified by bsdDfor and bsdDrev2; *bsdCD* was

amplified by bsdCfor and bsdDrev; and *bsdBCD* was amplified by bsdBfor and bsdDrev (Table 2.1). The PCR products bearing the genes *bsdBCD* and *bsdC* were first ligated into the pCR-SCRIPT vector (Stratagene). The pCR-SCRIPT*bsdBCD*, pCR-SCRIPT*bsdC*, and the target plasmid pUC18 were all cut with the appropriate restriction enzymes (Table 2.1). The restriction digest products were run on a 0.8% agarose gel, and the insert containing either *bsdBCD* or *bsdC* was isolated out of the gel using the QIAEXII Gel Extraction Kit (Qiagen). The other genes, *bsdB*, *bsdD*, and *bsdCD* were amplified by PCR, and together with pUC18 were cut with the appropriate restriction enzymes. Each of the five gene combinations were then ligated into the pUC18, and these constructs were electroporated into *E. coli* JM109. The clones were verified by restriction digest and sequencing.

Cloning the 4-hydroxybenzoate decarboxylase from* Escherichia coli *O157:H7 The genomic DNA of *E. coli* O157:H7 was obtained from the American Type Culture Collection (Manassas, VA). As listed in Table 2.1, to PCR amplify *ecdC*, ecdCfor and ecdCrev were used; ecdDfor and ecdDrev were used to amplify *ecdD*; and ecdCfor and ecdDrev were used to amplify *ecdCD*. The PCR products and pUC18 were each cut with the appropriate restriction enzymes, and then the PCR products were ligated into pUC18. After transformation into *E. coli* JM109, the constructs were verified by restriction digest and sequencing.

Verifying protein expression and enzyme activity

The plasmids listed and described in Table 2.2 were analyzed for expression and activity. The plasmids containing one or more of the genes homologous to *shdCD* and *vdcBCD* were transformed into *E. coli* JM109. *E. coli* strains containing the plasmids were first grown aerobically and then switched to anaerobic growth conditions to protect the enzymes from oxygen. To test for expression, the constructs were induced either by anaerobic growth with 20 mM KNO_3 for pNR constructs or with 1 mM IPTG for the pUC18 constructs. In each case, the optimal induction time was determined by monitoring protein expression levels as visualized on SDS-PAGE gels. Otherwise, the cultures were induce for 1 hour. The gels were poured using pre-mixed reagents (BioRad, Hercules, CA), and the gels were run according to manufacturer's directions. The gels were cast and run using the MiniProtean II system (BioRad). The protein samples were prepared by mixing 1 mL of cells at OD_{600}=1 (which could be harvested from the anaerobically grown cells by centrifuging an appropriate amount of cells and resuspending in 1 mL of LB broth) with 100 µL of 2x SDS loading buffer (Ausubel, *et al.*, 1995). This mixture was boiled for 3-5 minutes, and then centrifuged at 13,000 rpm for 10 minutes. A variable amount of supernatant was loaded onto the SDS-PAGE gel. The gel was run at 100 mV, 25 mAmps until the dye front ran off the gel. The gel was stained either with Coomassie blue G-250 or silver stained using the BioRad Silver Staining Kit. To visualize the small D subunit, pre-poured 16.5% Tris-tricine polyacrylamide gels (BioRad) were run.

To determine native molecular weight of the proteins, crude cell extracts were first obtained by lysing the cells with a French press. The cell lysates were analyzed on

gradient native PAGE gels. The cells lysates containing VdcCD, BsdCD, and EcdCD were compared to lysates from cells containing pUC18 and pUC18*shdCD*. The native molecular weights for the recombinant proteins were estimated by comparing the bands for VdcCD, BsdCD, and EcdCD to that of ShdCD, whose native molecular weight had been ascertained by gel filtration (He and Wiegel, 1995).

The enzyme assay, which was performed anaerobically using the Hungate technique and an anaerobic Coy chamber, was adapted from that described by He *et al.* (He and Wiegel, 1995). The cells, which had been grown anaerobically, were harvested by anaerobic centrifugation for 15 minutes at 5,000 x g at 4°C. These cells were either permeablized using toluene or were lysed with a French press. Pelleted cells to receive the toluene treatment were resuspended in Buffer II (10 mM $MgSO_4$, 1 mM $CaCl_2$, 50 mM MOPS, pH 7.0) at 10% the original volume. Toluene was added at a concentration of 100 μL /1 mL resuspended cells. The cells were shaken for 2 minutes at room temperature and then chilled on ice for 10 minutes, allowing the aqueous phase to separate. The toluene-treated cells were extracted from below the toluene layer using a syringe and needle, and this extract was used for the enzyme assay. Cells to be lysed with a French press were resuspended in a French press buffer (50 mM sodium phosphate pH 7.0, 300 mM NaCl, 0.5 mM PMSF). The cells were lysed in a SLM Aminco French Pressure Cell Press (SLM Instruments Inc, Lexington, MA) using a 1" piston at the high ratio setting yielding 9,000 psi. To maintain anoxic conditions, the French press cell was loaded into the anaerobic chamber and filled with cells in buffer. The output spout of the French pressure cell was equipped with a rubber tube and a 23 gauge needle inserted a

Hungate tube which had been flushed with nitrogen gas. This allowed the lysed cells to remain anaerobic. Whether the cells were toluene treated or lysed with a French press, 500 µL of lysed or toluized cells were mixed with 500 µL assay buffer (100 mM MES, 1 mM $MnCl_2$, 1 mM dithiothreitol, 0.5-10 mM hydroxybenzoate, pH 6.0). The substrate, the appropriate hydroxybenzoate, was added immediately before the enzyme assay. To ascertain the carboxylase activity of the enzyme, the substrate, usually phenol or guaiacol, was increased in concentration to 5-10 mM, and 100 mM $NaHCO_3$ was added as a carboxylate source. The assay mixture was incubated for 30-60 minutes, and then the assay was terminated with 2.5% trichloroacetic acid. The assay mixtures were refrigerated at -20°C for an hour to allow precipitation of protein, centrifuged at 13,000 rpm for 10 minutes, and the supernatant was analyzed by HPLC. A C-18 reverse phase Microsorb Short-one column (Rainin Instruments Company Inc., Woburn, MA) was attached to a Beckman 114M Solvent Delivery Module (Beckman, Berkeley, CA). An Alcott Model 738 Universal HPLC Autosampler (Alcott, Deerfield, IL) was used in conjunction with a DataJet SP4600 Integrator (Spectraphysics, San Jose, CA) . The UV detector, a Dynamax UV-C UV/Vis Absorbance Detector (Rainin), was set at $\lambda = 280$nm.

Oxygen sensitivity test

Oxygen sensitivity of the recombinant decarboxylases was tested by either growing the *E. coli* clones aerobically prior to the enzyme assay, performing the enzyme assay aerobically, or both. The activity of the oxygen-exposed cells was compared to that of anaerobic cells.

Substrate Range

***Vanillate decarboxylase from* Bacillus subtilis** The substrate range for the *B. subtilis* clones was determined by two methods. First, the clone pUC18*bsdCD* in *E. coli* JM109 was grown anaerobically in M9 minimal medium supplemented with casamino acids. Each tube contained one of a series of hydroxybenzoate compounds (Table 2.3). *E. coli* JM109 containing pUC18 was used as a negative control. The spent, centrifuged medium was analyzed by HPLC to determine whether any reaction had occurred. If there was a difference between the control and the cloned BsdCD, then the enzyme assay using toluene-treated cells was performed using the appropriate substrate in place of 4-hydroxybenzoate. For final verification, the assays were repeated with the partially purified enzyme described in Chapter 3.

4-hydroxybenzoate decarboxylase from* Escherichia coli *O157:H7 Substrate range was established by first growing *E. coli* JM109 cells containing pUC18*ecdCD* or pUC18 in anaerobic LB + 0.2% KNO_3 + 0.3% glucose overnight. The enzyme assay was then performed on these cells using the various substrated listed in Table 2.4 in lieu of 4-hydroxybenzoate. The cells were lysed with a French press for the enzyme assay.

Ascertaining the role of the B subunit

The following plasmids were transformed into *E. coli* THU: pSub1(provided by Dr. Julian Davies; pUC18 with *Bam*HI fragment of *Streptomyces* sp. D7 genome containing *vdcBCD*), pUC18*vdcCD*, pUC18*bsdBCD*, pUC18*bsdCD*, pUC18*ecdCD*, pUC18*shdCD*, pUC18. The clones were grown aerobically or anaerobically as described.

Expression of the proteins was ascertained by SDS-PAGE. The enzyme assay detailed above was used to measure decarboxylase activity. The activity of the enzymes was calculated by measuring the nanomoles of substrate produced per milligram of whole cell protein per minute.

Inclusion Body Assay

An inclusion body assay was performed to verify that the inactivity of ShdC was not due to the protein being trapped in an inclusion body. The inclusion body assay was performed on *E. coli* JM109 cells containing pNR*shd*C. *E. coli* JM109 containing pNR was used as a control. The cells were inoculated into 200 mL of anaerobic LB, and grown to OD_{600}=0.5. One mL of whole cells was harvested and frozen. The culture was split in half, and one half (100 mL) was induced appropriately by the methods previously mentioned. One mL of cells was harvested and frozen. Both sets of cells, induced and uninduced, were centrifuged at 6,000 x g for 15 minutes at 4°C. The cells were washed in 20 mL of the following buffer: 25 mM HEPES, 50 mM KCl, 10% glycerol, 1 mM DTT, and 1 mM PMSF (pH 7.5). The cells were spun again at 6,000 x g for 15 minutes at 4°C. The cells were resuspended in 5 mL of the same buffer, and each lysed by three passages through the French press. One hundred µL of each set of cells were spun at 15,000 x g for 15 minutes at 4°C (inclusion bodies should be in the pellet). A more thorough test involved spinning the cells at 30,000 x g for 30 minutes at 4°C.

The amount of protein in the pellet and in the supernatant was assayed by a micro-protein assay, according to a method provided by M. Thomm & J.W. Brown. Briefly, 5 mg/mL BSA was diluted eight times in a series of two-fold dilutions. One µL

of each dilution was spotted onto a nitrocellulose membrane. One μL dots of the pellet (resuspended in French Press buffer) and of the supernatant were also dotted onto the membrane. The spots were allowed to dry. The same samples were re-spotted on top of the previous spots and allowed to dry. The membrane was stained in amido black (0.25% amido black, 45% methanol, 45% distilled water, 10% glacial acetic acid) for 10 minutes. The membrane was then destained in the amido black solution minus the amido black until the background was nearly white. The intensities of the black spots were compared and used to estimate the protein concentration of the samples. The samples were analyzed by SDS-PAGE gel as described above.

Sequence Analysis

The GCG (Genetics Computer Group, University of Madison, Wisconsin) software package was used to examine sequences for restriction sites, matching sequences to each other, performing sequence homology searches (BLAST), finding open reading frames, and manipulating genomic data. Any sequences of the clones that were constructed were analyzed using Sequencher (Gene Codes Corp., Ann Arbor, MI). PCR primers were designed using OLIGO 6.2 (Molecular Biology Insights, Inc.). Estimations of molecular weight based on amino acid sequence were made using DNA Strider (Christian Marck, Cedex, France). Homology searches were performed using BLAST, and these homologues were compared to each other using PILEUP and BOXSHADE, all from GCG. The following programs were used to detect the possible nature of the secondary structure of the protein: Profil fed neural network systems from HeiDelberg

(PHD, B. Rost & C. Sander, EMBL, Heildelberg, Germany). Searches for crystallized homologues were performed in Swissprot.

Results

4-hydroxybenzoate decarboxylase from *Sedimentibacter hydroxybenzoicus* JW/Z-1

Expression and activity of* shdC *cloned in pNR After *ohb1* (*shdC*) had been identified as the gene encoding the 4-hydroxybenzoate decarboxylase, it was cloned in an expression vector so that large quantities of the protein could be produced for biochemical analysis. The clone pNR*shdC* was successfully constructed and transformed into *E. coli* JM109. The sequence of the insert was identical to that published except for one difference at position 1285 where there was a G instead of the A in the published sequence (Figure 2.1). This difference was in the wobble position for that codon, so the encoded amino acid sequence was not changed. ShdC was successfully overexpressed using the anaerobic expression vector pNR, as shown in Figure 2.2. The plasmid pRIL was introduced into strain to aid in the translation of the *shdC* mRNA, as the *S. hydroxybenzoicus* nucleotide sequence had a number of codons that are rare in *E. coli*. The recombinant protein had a molecular weight around 57 kDa as expected. However, the enzyme assay revealed no 4-hydroxybenzoate decarboxylase activity. The recombinant protein did not appear to be trapped in inclusion bodies since the majority of the protein remained in the supernatant after spinning at 30,000 x g for 30 minutes (Figure 2.3).

The original clone, pJBH-1, was re-examined to verify that the enzyme assay was functioning and that the *E. coli* strain JM109 was capable of expressing the

decarboxylase. The clone was able to express the protein (Figure 2.2), and the protein was an active decarboxylase. The plasmid pJBH-1 is pUC18 containing a *Hin*dIII fragment of the *S. hydroxybenzoicus* JW/Z-1 genome (Huang, *et al.*, 1999). The fragment contains sequences upstream and downstream of the putative 4-hydroxybenzoate decarboxylase gene (*shdC*). An examination of the upstream sequence revealed a partial ORF, and the downstream sequence held a small 204 base pair ORF, termed *shdD*, whose projected protein molecular weight was around 7 kDa.

***Examination of pUC18*shdCD** Since the protein expressed from pUC18*shdC* in *E. coli* JM109 was not an active enzyme, the other ORF's that had been identified in the original pJBH-1 plasmid were examined to clarify their potential requirement for 4-hydroxybenzoate decarboxylase activity. The plasmids pUC18*shdD* and pUC18*shdCD* were successfully constructed and transformed into *E. coli* JM109. Once again, the sequence of the insert had the same silent difference as the insert of pNR*shdC*. *E. coli* JM109 containing the plasmid pUC18*shdD* or pUC18*shdCD* was induced with 1 mM IPTG to express the proteins ShdD or ShdCD. ShdC was successfully expressed (Figure 2.2), and the protein was of the appropriate molecular weight. ShdD was not visible on either Coomassie-stained glycine or tricine SDS-PAGE gels. An enzyme assay of the cells containing ShdCD revealed that it is an active 4-hydroxybenzoate decarboxylase but ShdD was not (Table 2.2). The ShdCD was oxygen sensitive and reversible like the original enzyme purified from *S. hydroxybenzoicus* (He and Wiegel, 1995).

Homology search of ShdCD A homology search of the amino acid sequence for ShdC retrieved a number of sequences (Table 2.5). These homologues, mostly hypothetical proteins, bear high amino acid sequence similarities and identities to ShdC. A separate homology search of ShdD revealed only a few proteins (Figure 2.4). These were almost all hypothetical proteins, with the exception of VdcD. The only homologues which had a C homologue followed by a D homologue were from *B. subtilis*, *Streptomyces* sp.D7, *E. coli* O157:H7, and *S. enterica* subspecies *enterica* serovar Typhi.

Vanillate decarboxylase from *Streptomyces* sp.D7

The activity of a recombinant vanillate decarboxylase from *Streptomyces* sp.D7 was described previously (Chow, *et al.*, 1999). In that research, it was found that *vdcBCD* were needed to encode an active decarboxylase in *S. lividans* 1326. In the research presented here, *vdcC* and *vdcD* were expressed individually and together in *E. coli* JM109 to ascertain if both the C and D subunits were needed for activity, as is the case with ShdCD. The plasmids pUC18*vdcC*, pUC18*vdcD*, and pUC18*vdcCD* were each successfully transformed into *E. coli* JM109. The sequences of the inserts were identical to those published by Chow *et al* (Chow, *et al.*, 1999). All three constructs were induced with 1 mM IPTG, and the cell lysates were examined by SDS-PAGE. The C subunit was successfully expressed and visualized by SDS-PAGE (Figure 2.5) and is of the expected molecular weight of 52 kDa. VdcD could not be visualized in the cell lysate by Coomassie or silver staining, even when tricine gels were used to aid the separation of low molecular weight proteins. The enzyme assays of all three plasmids as expressed in *E. coli* JM109 revealed that both VdcC and VdcD were needed for an active reversible

vanillate decarboxylase. VdcC and VdcD alone do not have decarboxylase activity (Table 2.2). VdcCD is not oxygen sensitive, which is in accordance with oxygen sensitivity tests performed with VdcBCD expressed in *S. lividans* 1326 (Chow, *et al.*, 1999).

Vanillate decarboxylase from *Bacillus subtilis*

The hypothetical proteins homologous to ShdCD from *B. subtilis* were cloned and their expression and activity were examined. These homologues represent another member of the proposed novel hydroxyarylic acid decarboxylase family. The plasmids pUC18*bsdBCD*, pUC18*bsdCD*, pUC18*bsdB*, pUC18*bsdC*, and pUC18*bsdD* were all successfully constructed, transformed into *E. coli* JM109, and induced for expression (Figure 2.6). The induced proteins were assayed for decarboxylase activity with 4-hydroxybenzoate, 3,4-dihydroxybenzoate, and vanillate. The plasmid pUC18*bsdB* did not encode an active decarboxylase. Reversible vanillate decarboxylase activity was detected only in strains that expressed both BsdC and BsdD (Table 2.2). As with ShdD and VdcD, BsdD was not visible on SDS-PAGE gels although BsdC was easily detected.

The nucleotide sequences of *bsdB* and *bsdC* corresponded to those published in Genbank. However, there was a discrepancy in the nucleotide sequence of *bsdD*. There is a deletion in what would be position 217 of the nucleotide sequence (Figure 2.7). This difference results in a shortening of the ORF for *bsdD*, the ramifications of which will be discussed later. Due to this change, the primer sequence to amplify *bsdD* was redesigned, as detailed in Table 2.1.

BsdBCD and BsdCD are oxygen sensitive, as loss of activity resulted if either of the recombinant proteins were produced in aerobically grown cells or if the lysed cell extract of anaerobically grown cells was exposed to air for at least 30 minutes. The substrate range for BsdCD was assessed (Table 2.3). Of the twenty substrates tested, the enzyme only decarboxylated vanillate.

4-hydroxybenzoate decarboxylase from *Escherichia coli* O157:H7

E. coli O157:H7 genomic sequence contained homologues to *shd*CD. To test whether these homologues encoded active decarboxylases that could be members of the proposed novel enzyme family of hydroxyarylic acid decarboxylases, the genes were cloned and expressed, and the gene products were assayed for activity. The plasmids created, pUC18*ecdC*, pUC18*ecdD*, and pUC18*ecdCD*, were transformed into *E. coli* JM109, and the sequences of the inserts were the same as those published. As is the case with the other D homologues, SDS-PAGE gels of the proteins anaerobically expressed by pUC18*ecdC*, pUC18*ecdD*, and pUC18*ecdCD* reveal EcdC but not EcdD (Figure 2.8).

Cell lysate containing EcdC or EcdD showed no decarboxylating activity. Only EcdCD displayed decarboxylase activity towards 4-hydroxybenzoate and vanillate under anaerobic conditions. The enzyme was reversible. All activity was lost if the cells were grown in the presence of oxygen, or if the cells were exposed to oxygen at room temperature for the two hours during which the cells were lysed and assayed (Table 2.2). In contrast to the other recombinant enzymes, EcdCD had a wide substrate range, decarboxylating all but 2,4,6-trihydroxybenzoate and 3-hydroxybenzoate of the substrates tested (Table 2.4).

Role of the B subunit

The strain *E. coli* THU has a point mutation in *ubiX*, which is the *E. coli* K-12 homologue to VdcB. The activities of the clones containing the B, C, and D subunits versus those containing just the C and D subunits are compared in Table 2.6. The enzyme activity is measured in whole cells. Regardless of whether the B subunit was cloned into *E. coli* THU along with the C and D subunits, the enzyme activity was the same. Activity of the ShdCD homologues in *E. coli* JM109 compared to that in *E. coli* THU was 0-10 times higher.

Discussion

Two genes are needed to encode an active recombinant 4-hydroxybenzoate decarboxylase from *S. hydroxybenzoicus*

A recombinant 4-hydroxybenzoate decarboxylase expressed in *E. coli* DH5α resembled the native enzyme in its oxygen sensitivity, activity, and apparent molecular weight (Huang, *et al.*, 1999). No other subunits other than ShdC were visible by SDS-PAGE, and the native protein had a molecular weight which could be accounted for by a homohexameric configuration of ShdC. For this reason, it was originally concluded that the 4-hydroxybenzoate decarboxylase from *S. hydroxybenzoicus* JW/Z-1 was encoded by a single gene.

In an effort to further characterize the enzyme, we attempted to express a large amount of the 4-hydroxybenzoate decarboxylase. The plasmid pNR*shdC* transformed into *E. coli* JM109 was an effort to insert the gene into a vector that would allow for

anaerobic expression of the protein. After induction, pNR*shdC* was not able to express an active

4-hydroxybenzoate decarboxylase, even though ShdC was expressed (Figure 2.2). The inactivity of the protein was not due to the protein being trapped in inclusion bodies. The enzyme assay and growth methods were not the problem either, as *E. coli* JM109 containing pJBH-1 was able to produce a functional decarboxylase. The sequence of *shdC* was functionally identical to that of the published sequence. There was one difference when comparing the sequence of *shdC* to the original published sequence, but it did not affect the amino acid sequence. The fact that multiple sequencings of this gene in this research yielded the same result suggests that there was a sequencing error in the original publication.

Since *shdC* alone did not encode a functional decarboxylase, we re-examined the pJBH-1 construct. The plasmid, pJBH-1, consists of a *Hin*dIII fragment of the *S. hydroxybenzoicus* JW/Z-1 genome inserted into pUC18 (Huang, *et al.*, 1999). This *Hin*dIII fragment contained the putative gene sequence (termed *ohb1* but renamed *shdC* in this research), a partial ORF upstream, and a small ORF downstream (termed *shdD*). Attempts were made to express *shdD* along with *shdC*. SDS-PAGE analysis showed only ShdC but not ShdD. The inability to visualize ShdD in whole cell extracts may be attributed to a combination of the small size of the protein and the difficulty of distinguishing it from the other cellular proteins. The plasmid pUC18*shdCD* was transformed into *E. coli* JM109, and it was able to express an active 4-hydroxybenzoate decarboxylase. This lead to the conclusion that *shdD* is essential for activity, but it does not encode an active protein on its own. ShdD was not originally detected when the

4-hydroxybenzoate decarboxylase was purified from *S. hydroxybenzoicus* JW/Z-1, probably due to its small size (7 kDa). At this time, the ratio of D subunits to C subunits is unknown. The homologue to ShdD in the established vanillate decarboxylase from *Streptomyces* sp.D7, termed VdcD, was also necessary for activity of that enzyme.

Search for sequences homologous to ShdCD

The homology search of ShdC revealed a large number of hypothetical proteins, a comparison of which is shown in Figure 2.8. Other than the decarboxylases discussed in this research, there is one other characterized enzyme from the ShdC homologues. It is the UbiD, or 3-octaprenyl-4-hydroxybenzoate decarboxylase, from *E. coli* K-12. This enzyme will be discussed in detail later. *Streptomyces* sp.D7, *E. coli* O157:H7, *S. enterica*, and *B. subtilis* all have genes encoding homologues to ShdD that are immediately downstream of the genes encoding ShdC homologues. These homologues, except for the ones from *S. enterica*, were characterized in this research.

The majority of the ShdC homologues from other bacteria did not have an ShdD homologue evident in their genomes. Only an amino acid sequence was used to probe these genomes, so it is possible that the ShdD sequence exists in these genomes but the encoding nucleic acid sequence was not annotated as an ORF due to its short length. The ShdD homologues may have an iron-sulfur cluster, judging from the conserved CXXC-(21 bp)-CXXC motif. This subunit could be the reason for the oxygen sensitivity of the enzyme, as iron sulfur clusters can be inactivated by oxygen. The cluster may play a crucial role in the enzyme mechanism. Analysis of the ShdD subunit was outside the scope of this thesis but will be performed in the future.

The large number of homologues and their phylogenetic span are intriguing: do they all have the same function or are they artifacts of evolution? In this research, no attempt was made to characterize the other homologues, and they remain hypothetically annotated.

Examining the ShdCD homologues

VdcCD is a vanillate decarboxylase The vanillate decarboxylase from *Streptomyces* sp.D7 was first described as a non-oxidative decarboxylase with a narrow substrate range (Chow, *et al.*, 1999). Three polycistronic genes, *vdcBCD*, were necessary for expression of an active decarboxylase in *S. lividans*1326. In this research, the homologues to ShdCD were cloned into *E. coli* JM109. These clones, which contained pUC18*vdcC*, pUC18*vdcD*, and pUC18*vdcCD*, did not contain the gene *vdcB*, since a similar gene was not necessary for ShdCD activity as expressed in *E. coli* JM109. Each of the clones was assayed for enzyme activity. Only the clone containing both VdcC and VdcD was able to decarboxylate vanillate. However, while *vdcBCD* was necessary for an active decarboxylase in *S. lividans* 1326, only *vdcCD* were necessary for activity in *E. coli* JM109. This discrepancy was puzzling, and it prompted further investigation of the nature of the B subunit which will be discussed later in this chapter. As reported originally, the *Streptomyces* sp.D7 vanillate decarboxylase is not oxygen sensitive. This holds true even when the protein is expressed in *E. coli* JM109 and not in the original *S. lividans* 1326 system.

BsdCD is a vanillate decarboxylase Several clones were constructed to elucidate whether the ShdCD homologues from *B. subtilis* encoded a hydroxybenzoate decarboxylase and which of the genes were needed for an active enzyme. The constructs evaluated include pUC18*bsdBCD*, pUC18*bsdCD*, pUC18*bsdC*, and pUC18*bsdD*. The *bsdBCD* genes were previously named *yclBCD* and annotated as a vanillate decarboxylase. As discovered in this research, the expression of both the *bsdCD* in *E. coli* JM109 are necessary for an active vanillate decarboxylase. This enzyme was oxygen sensitive. An attempt was made to express the two subunits in two separate clones, then mix the two cell lysates together, and assay for activity. This experiment was performed to ascertain whether the two subunits could interact spontaneously. No activity was detected when BsdC-containing cell extract was mixed with BsdD-containing cell extract. Attempts to obtain an active enzyme by mixing ShdC with BsdD and BsdC with ShdD also failed. BsdCD, like the vanillate decarboxylase from *Streptomyces* sp.D7, was a specific vanillate decarboxylase (Table 2.3).

EcdCD is a 4-hydroxybenzoate decarboxylase The two genes homologous to *shdCD* from *E. coli* O157:H7, *ecd*C and *ecd*D, were cloned either separately or together into pUC18 and expressed in *E. coli* JM109. EcdCD, the result of expression of pUC18*ecdCD*, was able to decarboxylate 4-hydroxybenzoate and vanillate. As with the other enzymes discussed, EcdD, the homologue to ShdD, was needed in addition to EcdC for an active enzyme. Like ShdCD, EcdCD is oxygen sensitive. The substrate range for EcdCD, summarized in Table 2.5, is extensive compared to the other decarboxylases. While the pUC18-only control had no activity, the clone containing pUC18*ecdCD* was

able to act on numerous compounds. ShdCD needed a hydroxyl group on the *para* position of the benzoate ring but no substituent in the *ortho* position (He and Wiegel, 1995). There does not appear to be a pattern to distinguish acceptable substrates for EcdCD. A closer examination of the substrate range with purified enzyme is necessary.

There are a number of papers discussing the differences between pathogenic and non-pathogenic strains of the same species (Hayashi, *et al.*, 2001; Herbelin, *et al.*, 2000; Ohnishi, *et al.*, 2000). The gene cluster of *ecdBCD* is another difference between the pathogenic *E. coli* O157:H7 and the non-pathogenic *E. coli* K-12. This set of genes, including another gene designated *slyA,* constitute a variable genomic section between the genes for *mut*S and *rpoS* (Herbelin, *et al.*, 2000). The gene *mutS* encodes one of the four proteins required for mismatch repair, and *rpoS* encodes σ^{38} needed to regulate stationary phase and stress response genes. The genes *ecdBCD* are only conserved in the pathogenic strains related to *E. coli* O157:H7, but are not in *E. coli* K-12 or the members of ECOR group A. The cloning strain of *E. coli* JM109, which is derived from *E. coli* K-12, is not able to decarboxylate any of the hydroxybenzoate compounds without the decarboxylase genes inserted on a plasmid. EcdCD may have nothing to do with the pathogenicity of the strain, but it is another difference by which evolutionary distance or phylogenetic relationships can be mapped.

There has been another 4-hydroxybenzoate and vanillate decarboxylating enzyme identified in an *E. coli* strain, designated strain C2, which was isolated from a shea cake digester inoculated with sludge from a slaughter house (Chamkha, *et al.*, 2002). *E. coli* C2 was tolerant of tannic acid, and it decarboxylated 4-hydroxybenzoate and vanillic acid

both aerobically and anaerobically. This is similar to the VdcCD from *Streptomyces* sp.D7 which could also function either aerobically or anaerobically.

Monitoring expression of recombinant proteins For all of the recombinant proteins expressed in *E. coli* JM109 which include VdcCD, EcdCD, and BsdCD, the ShdC but not the ShdD homologues could be visualized on SDS-PAGE gels. As was the case with ShdCD, the difficulty in visualization of the ShdD homologue could be due to the low molecular weight of the fragment combined with the difficulty in distinguishing the protein from other proteins in the cell.

Discrepancy in the *bsdD* sequence

There is a discrepancy between the published sequences and the sequences reported here. In the published sequence of *bsdD*, there is a cytosine (C) in position 217 (Figure 2.7). However, in the repeated sequencings obtained from the cloning of *bsdD*, there is no C in that position. The *bsdD* was PCR amplified with *Pfu* high fidelity polymerase, which has a proof-reading function. This information along with the repeated sequencings of the *bsdD* suggest that the published sequence contains an error. This causes a shortening of the predicted protein encoded by this ORF. The predicted protein is shortened from 229 amino acids to 75 amino acids (Figure 2.4) which makes it similar in length to ShdD, EcdD, and VdcD.

Role of the B subunit

The vanillate decarboxylase from *Streptomyces* sp. D7 required VdcB in addition to VdcCD for an active decarboxylase as expressed in *S. lividans* 1326. Expression of an active enzyme in *E. coli* JM109, however, only required the C and D subunits. The question was raised whether the B subunit was necessary for an active enzyme.

There is a VdcB homologue upstream of the C subunit in all the members of the proposed family except *S. hydroxybenzoicus* (Table 2.6, Figure 2.10). The VdcB is homologous to a phenylacrylic acid decarboxylase (Pad1) from *S. cerevisiae* and to UbiX, also known as DedF, from *E. coli* K-12 (Figure 2.9) (Clausen, *et al.*, 1994). UbiX is annotated as a 3-octaprenyl-4-hydroxybenzoate carboxy-lyase, which is a protein involved in ubiquinone synthesis (Zeng, *et al.*, 1998). Its homologue in *Salmonella typhimurium* has been verified as a 3-octaprenyl-4-hydroxybenzoate decarboxylase. *E. coli* has another established 3-octaprenyl-4-hydroxybenzoate decarboxylase, encoded by *ubiD.* UbiD bears high homology to the genes encoding the C subunit of the proposed members of the novel enzyme family, but it does not bear high homology to *ubi*X, the latter also encoding a significantly smaller protein. There is the possibility that the VdcB homologue is needed for an active decarboxylase along with the ShdCD homologues, but in its absence, it may be supplemented by UbiX in *E. coli* JM109.

To determine whether the UbiX from the *E. coli* JM109 was substituting for the potentially missing B subunit, attempts were made to construct a *ubiX* knock out mutant in *E. coli* JM101, *E. coli* TB-1, *E. coli* TG-1, and *E. coli* MC13. These attempts were unsuccessful. However, the strain *E. coli* THU has a point mutation which renders the UbiX enzyme inactive (Zeng, *et al.*, 1998). Comparisons were made between the

decarboxylase activity of *E. coli* THU expressing the pUC18*bsdBCD* and pUC18*bsdCD* (Table 2.6). The enzyme activities of the two strains is the same. The same results was obtained when comparing the activities of *E. coli* THU expressing pSUB1 (pUC18*vdcBCD*) and pUC18*vdcCD*. These results led to the conclusion that the B subunit was probably not necessary for active decarboxylase expression. While the *E. coli* THU was only a point mutation and thus might have had residual protein activity, the clones that contained the B subunit did not exhibit a higher level of activity, as might be expected if the B subunit were necessary. This conclusion would be further supported if a knock out mutant of *ubiX* could be constructed, and the activities of the clones with and without the B subunit were equivalent. On average, the level of activity in *E. coli* JM109 with the same plasmids was higher than in *E. coli* THU. This may be attributed to differences in the abilities of the strains to express recombinant proteins. This difference in activity merits further investigation.

What is the physiological role of these decarboxylases?

The ShdC homologues are widely distributed among phylogenetically diverse bacteria. The only homologue to ShdC which has an established physiological function is the UbiD from *E. coli* K-12. UbiD, which is 29% identical to ShdC, is a 3-octaprenyl-4-hydroxybenzoate decarboxylase involved in ubiquinone synthesis (Zhang and Javor, 2000). It is not homologous to UbiX, which is another 3-octaprenyl-4-hydroxybenzoate found in *E. coli* K-12. UbiD has similar characteristics to those of the decarboxylases described in this research. It is a 340 kDa native protein with 55.6 kDa subunits. UbiD relies on Mn^{2+} and a small molecular weight cofactor (less than 10 kDa) for activity

(Leppik, *et al.*, 1976). Even the substrate, which is a 4-hydroxybenzoate with an octaprenyl group in the *meta* position, is similar to the substrates for ShdCD, EcdCD, BsdCD, and VdcCD. It is interesting to note that UbiD is homologous to the C subunit and the B subunit is homologous to UbiX. It could be that the enzymes in the proposed enzyme family are involved in synthesis of electron transport proteins. It is also possible that the enzymes in the proposed family simply share a common mechanism with UbiD.

Definition of the proposed reversible hydroxyarylic acid decarboxylase family

The enzymes studied in this research come from *B. subtilis* (BsdCD), *Streptomyces* sp.D7 (VdcCD), *E. coli* O157:H7 (EcdCD), and *S. hydroxybenzoicus* (ShdCD). These proteins have been characterized as vanillate decarboxylases (VdcCD and BsdCD) or 4-hydroxybenzoate decarboxylases (EcdCD and ShdCD). Based on these decarboxylases, the parameters describing the proposed reversible hydroxyarylic acid decarboxylase family are defined below.

More than one gene is needed to encode for an active decarboxylase. The genes encoding the ShdD homologue immediately follows that encoding the ShdC homologue. The high amino acid sequence identities, shown in Figures 2.4 and 2.11, supports their forming a novel enzyme family. The molecular weights of the different subunits correspond roughly to each other, and the native PAGE reveals that the native enzymes are approximately the same size (Table 2.7). All the enzymes except VdcCD are oxygen sensitive.

There are several differences between the studied enzymes. They have different substrate ranges - the vanillic acid decarboxylases seem to be very substrate specific

while the 4-hydroxybenzoate decarboxylases have a wider range of substrates. Also, of the enzymes investigated, the homologue from *Streptomyces* sp.D7 is the only enzyme that is not oxygen-sensitive. Further enzyme characterization, such as protein structure and enzyme mechanism, will reveal whether these proteins are truly related.

No homologous protein has been crystallized, thus the three dimensional structure of the decarboxylases cannot be inferred. An analysis of the primary amino acid sequence of ShdC reveals some possible features of its secondary structure. It appears that the protein is 26.9% helix, 26.7% strands, and 46.5% loops, thus the protein is of a mixed class. The accuracy of this information is less than 72% since there is no pre-characterized homologue.

Future research

Important objectives for future research are to clarify the role of the B subunit and determine whether it is needed for a functioning decarboxylase. Enzyme assays of the decarboxylases expressed in an *E. coli ubiX* knockout mutant would clarify if the B subunit is necessary. Attempts to make a knockout mutant have so far been unsuccessful.

Once the role of the B subunit is understood, enzyme characterization is needed to understand the enzyme mechanism. A partial purification of recombinant ShdCD and BsdCD has already been accomplished by tagging the C subunits with a hexahistidine and using immobilized metal affinity chromatography (IMAC) to purify the proteins. Further column chromatography could purify the enzymes to homogeneity. Once pure enzymes are available, metal analysis could elucidate both enzyme mechanism and a possible basis for the oxygen-sensitivity of the enzyme activity.

Table 2.1- List of PCR primers used for cloning ShdCD homologues

Organism	Primer Name	Primer sequence	Inserted restriction site
Sedimentibacter hydroxybenzoicus JW/Z-1	4ohfor	GGA AGC TTA GAA GGA GGT ATA TTA TGG CTA AAG TAT ACA AAG AT	*Not*I
	4ohrev	CAC ATC TAT GAC ATT TCA TAA GCG GCC GCC AAT TTA TCT G	*Hin*dIII
	4ohCfor	ACG GAG AGC TCG GAG GTA TAT TAT GGC TAA AGT ATA	*Sac*I
	4ohDfor	CGC GAC TCT AGA CTA TTT TTT CAA GGG TGG TAT	*Xba*I
	4ohDrev	GCG GAG AAG CTT CTA TTT TTT CAA GGG TGG	*Hin*dIII
Streptomyces sp.D7	vdcCfor	CGG CAG GAT CCA CAG GAG ATT CAC CAT GGC CTA TGA C	*Bam*HI
	vdcCrev	GCA CTG AAG CTT TGA GGA TCG GTG GGT GTC AGA CG	*Hin*dIII
	vdcDfor	GAT TGG ATC CAG GAG AAC AGG GCC CGT GAA CCA CC	*Bam*HI
	vdcDrev	CGA GGA AGC TTC GTG CCA GGA ACG TCA CTT GAG CAG	*Hin*dIII
Bacillus subtilis	bsdBfor	GAC CGC GGT ACC CAG GAG TAT GAT TGA AAT GAA AGC	*Kpn*I
	bsdBrev	CCT CTC TAG ACA TGA TCA AGC TCC T	*Xba*I
	bsdCfor	GCG CCC CGG GAC CAA TTC GGC ATT CGG CTT C	*Xma*I
	bsdCrev	GGC CAA GCT TAT AAA CTT CCC ATG CGC CTT C	*Hin*dIII
	bsdDfor	GAC GGT ACC AGG AGG GTT CGA AAT GCA TAC	*Kpn*I
	bsdDrev	GCA GCC TCT AGA GAT ATT ACC GTT TTA AAT CTT CCA GG	*Xba*I
	bsdDrev2*	GCA TTC TAG AGA TCA AGC CTT TCG TTC C	*Xba*I
Escherichia coli O157:H7	ecoCfor	CTG CGA ATT CGG ACG TTT GTA ATG GCA TTT GAT GAT	*Eco*RI
	ecoCrev	GCA GTC TAG ACA TCT TGT TCT CCT TAT TTA	*Xba*I
	ecoDfor	GCT GGA ATT CGG AGG TCT ATT AAT GAT TTG TCC ACG TT	*Eco*RI
	ecoDrev	GGC CTC TAG ATT ATT AGC GCT TAC CTT C	*Xba*I

* The primer bsdDrev2 was designed after a discrepancy in the published sequence of *bsdD* was discovered. This change, as shown in Figure 2.7 is the absence of a second C at position 217.

Table 2.2- ShdCD homologues expressed and tested for decarboxylase activity

Organism	Construct	Expresssion?	Decarboxylase activity?	Reversible?
Sedimentibacter hydroxybenzoicus JW/Z-1	pNR*shdC*	yes	no	
	pUC18*shdCD*	yes	yes	yes
	pUC18*shdD*	ND	no	
	pJBH-1*	yes	yes	yes
Streptomyces sp.D7	pUC18*vdcC*	yes	no	
	pUC18*vdcD*	ND	no	
	pUC18*vdcCD*	yes	yes	yes
Bacillus subtilis	pUC18*bsdB*	yes	no	
	pUC18*bsdC*	yes	no	
	pUC18*bsdD*	ND	no	
	pUC18*bsdBCD*	yes	yes	yes
	pUC18*bsdCD*	yes	yes	yes
Escherichia coli O157:H7	pUC18*ecdC*	yes	no	
	pUC18*ecdD*	ND	no	
	pUC18*ecdCD*	yes	yes	yes

* pJBH-1, which is pUC18 with a *Hin*dIII fragment of *S. hydroxybenzoicus* genomic DNA containing *shdCD* and a short upstream region called *shdB* (Huang, *et al.*, 1999).
ND = not detected

Table 2.3- Substrate range of BsdCD as expressed in *E. coli* JM109

Substrate Name	Activity?
3-hydroxybenzoate	no
4-hydroxybenzoate	no
2,3-dihydroxybenzoate	no
2,4-dihydroxybenzoate	no
2,5-dihydroxybenzoate	no
2,6-dihydroxybenzoate	no
3,4-dihydroxybenzoate	no
2,3,4-trihydroxybenzoate	no
2,4,6-trihydroxybenzoate	no
3,4,5-trihydroxybenzoate	no
3,5-dihydroxy-2-naphthalene dicarboxylate	no
2,8-dihydroxynapthoate	no
5-hydroxyisophthalate	no
1-hydroxy-2-naphthoate	no
2-hydroxy-1-naphthoate	no
3-hydroxy-2-naphthoate	no
6-hydroxynicotinate	no
2-hydroxypyridine-3-carboxylate	no
3-hydroxypyridine-2-carboxylate	no
vanillate	yes

Table 2.4- Substrate range of EcdCD as expressed in *E. coli* JM109

Substrate Name	Activity?
2-hydroxybenzoate	no
3-hydroxybenzoate	yes
4-hydroxybenzoate	yes
2,3-dihydroxybenzoate	yes
2,4-dihydroxybenzoate	yes
2,5-dihydroxybenzoate	yes
2,6-dihydroxybenzoate	yes
3,4-dihydroxybenzoate	yes
2,3,4-trihydroxybenzoate	yes
2,4,6-trihydroxybenzoate	no
3,4,5-trihydroxybenzoate	yes
vanillate	yes

Table 2.5 - List of homologues to ShdC (May 2002)

Microorganism [known protein]	Accession Number*	Percent Similarity	Percent Identity
BACTERIA			
Aerobe, Facultative Anaerobe			
Bacillus subtilis [BsdC]	CAB12158	74	58
Salmonella enterica subsp. enterica serovar Typhi	NP_457773	69	52
Salmonella typhimurium LT2	AAL21802.1	69	52
Streptomyces sp.D7 [VdcC]	AAD28782	69	53
Novosphingobium aromaticivorans	AAD03845	62	45
Sphingomonas aromaticivorans plasmid pNL1	T31294	62	45
Synechocystis sp. PCC6803	S74726	56	32
Aquifex aeolicus	D70439	54	33
Rhodospirillum rubrum	T51313	54	31
Escherichia coli O157:H7 EDL933 [EcdC]	AAG57845	53	30
Haemophilus ducreyi	AAF00114	51	32
Neisseria meningitidis (group A, strain Z2491)	F81823	50	32
Pasteurella multocida	NP_245155	50	30
Neisseria meningitidis (group B, strain MD58)	H81051	49	31
Thauera aromatica	CAC12691.1	49	30
Deinococcus radiodurans (R1)	A75533	48	30
Escherichia coli K12	AAC76846	48	29
Helicobacter pylori (strain J99)	D71864	48	25
Pseudomonas aeruginosa	PA0254	48	31
Rickettsia conorii	NP_360908	48	32
Rickettsia prowazekii	F71643	48	32
Caulobacter crescentus CB15	AAK22300	47	29
Nostoc sp. PCC 7120	AP003584	47	30
Ralstonia solanacearum	NP_518829	47	31
Sinorhizobium meliloti megaplasmid pSymA	D95411	47	27
Vibrio cholerae (group 01, strain NI6961)	G82338	47	30
Aeropyrum pernix	A72513	46	26
Chlamydia psittaci	O34023	46	23
Campylobacter jejuni (strain NCTC 11168)	B81401	45	24
Streptomyces coelicolor A3(2)	CAB92111	45	30
Agrobacterium tumefaciens	AAK89045	44	29
Brucella melitensis	NP_541935	43	26
Bacillus halodurans	BH3930	41	25
Chlamydia muridarum (strain Nigg)	F81711	40	23
Chlamydophila pneumoniae	A72092	40	23
Chlamydia trachomatis (serotype D, strain UW3/Cx)	E71557	39	23
Yersinia pestis	NC_407218	39	30
Obligate anaerobe			
Sedimentibacter hydroxybenzoicus JW/Z-1 [ShdC]	S69350	100	100

ARCHAEA			
Hyperthermophiles			
Archaeoglobus fulgidus	B69455	53	32
Methanococcus jannaschii	D64441	53	28
Pyrococcus horikoshii	F71087	53	32
Archaeoglobus fulgidus	A69276	52	25
Sulfolobus solfataricus	AAK41165	51	30
Sulfolobus tokodaii	BAB65007	49	30
Pyrococcus abyssi (strain Orsay)	E75130	45	29
Pyrobaculum aerophilum	D836164	42	27
Pyrococcus furiosus	NA	39	27
Thermophiles			
Methanothermobacter thermoautotrophicus	E69052	53	28
Mesophiles			
Methanobrevibacter smithii	S28657	51	26
Thermoplasma acidophilum	CAC12324	47	29
Thermoplasma volcanium	NP_110927	46	29
EUKARYOTES			
Saccharomyces cerevisiae	S62018	50	25

* Accession numbers refer to proteins from SWISS-PROT or GENBANK.

Table 2.6 - Comparison of the activities of clones expressed in *E. coli* THU

Strain	nmoles substrate produced/mg whole cell protein per minute
pUC18*vdcBCD* in *E. coli* THU	0.003
pUC18*vdcCD* in *E. coli* THU	0.003
pUC18*vdcCD* in *E. coli* JM109	0.030
pUC18*bsdBCD* in *E. coli* THU	0.003
pUC18*bsdCD* in *E. coli* THU	0.003
pUC18*bsdCD* in *E. coli* JM109	0.003
pUC18*ecdCD* in *E. coli* THU	0.003
pUC18*ecdCD* in *E. coli* JM109	0.033
pUC18*shdCD* in *E. coli* THU	0.009
pUC18*shdCD* in *E. coli* JM109	0.019

Table 2.7- Molecular weights and accession numbers of the proposed members of the novel hydroxyarylic acid decarboxylase family

Organism	B Subunit MW*	B Subunit Accession #	C Subunit MW*	C Subun it Accession #	D Subunit MW*	D Subunit Accession #	Native MW
Sedimentibacter hydroxybenzoicus JW/Z-1	unknown	none	57 kDa	AAD50377	7.7 kDa	none	350
Streptomyces sp.D7	36.0 kDa	AAD28781	52 kDa	AAD28782	9.0 kDa	AAD28783	350**
Bacillus subtilis	22.5 kDa	NP_388245	53 kDa	NP_388246	8.6 kDa	NP_388247	350**
Escherichia coli O157:H7	21.4 kDa	NP_311620	52 kDa	NP_311619	9.2 kDa	NP_311618	350**

* The molecular weights were calculated according to the amino acid sequence by DNA Strider.

** The native molecular weight of ShdCD was determined by gel filtration (He and Wiegel, 1995)

```
new shdC sequence  201 TCATGGTTCCTGGCAAAATCATGCTCTT 228
                       |||||||||  ||||||||||||||||||
published sequence 201 TCATGGTTCTTGGCAAAATCATGCTCTT 228
```

Figure 2.1 - Difference in the nucleotide sequence of *shdC*
The difference results in no change of the amino acid sequence of the protein.

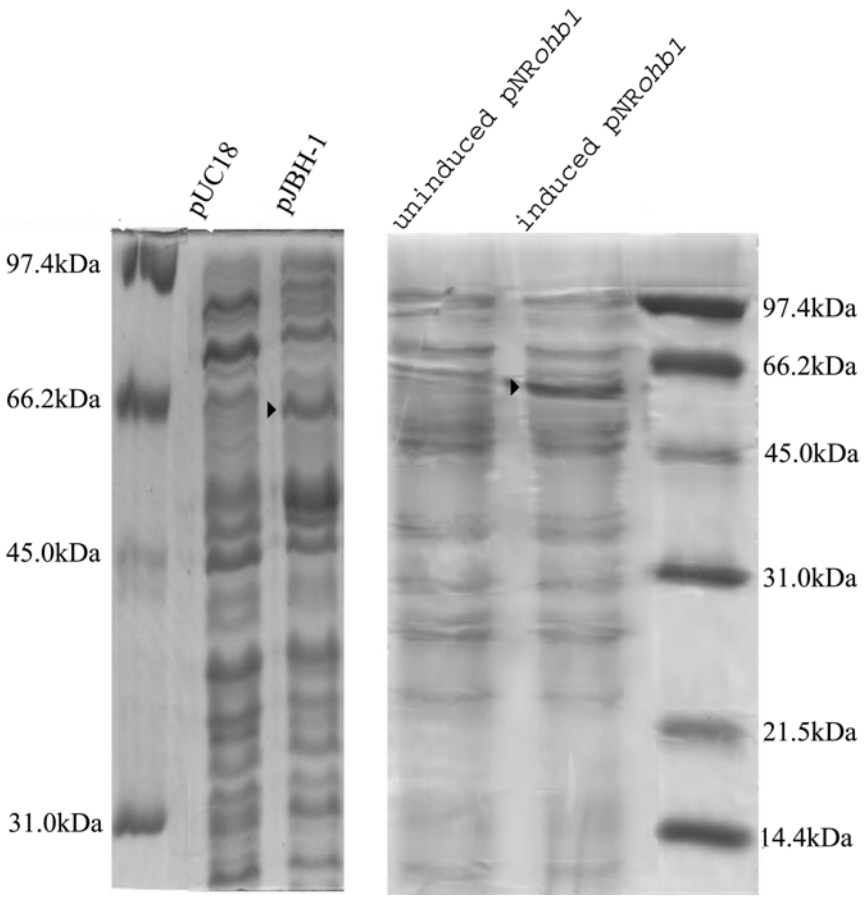


Figure 2.2- SDS-PAGE gel of ShdC expressed by pJBH-1 and pNR*shdC*pRIL in *E. coli* JM109
The SDS-PAGE gels were stained with Coomassie blue G-250. The arrows indicate the presence ShdC.

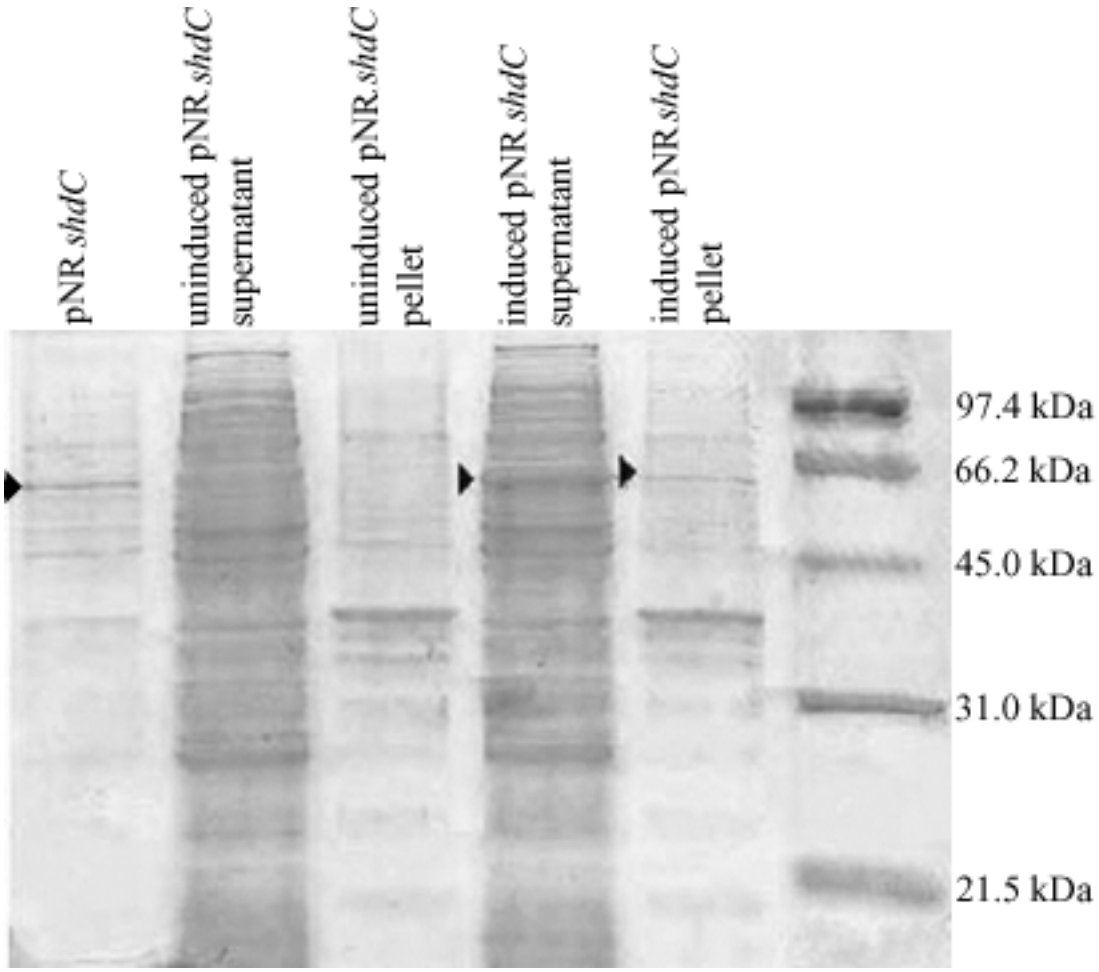


Figure 2.3- SDS-PAGE of the inclusion body test for pNR*shdC*pRIL
The arrows indicate the presence of ShdC, which is mainly in the supernatant and not in the pellet. This demonstrates that ShdC was not trapped in inclusion bodies. The SDS-PAGE gel was stained with Coomassie blue G-250.

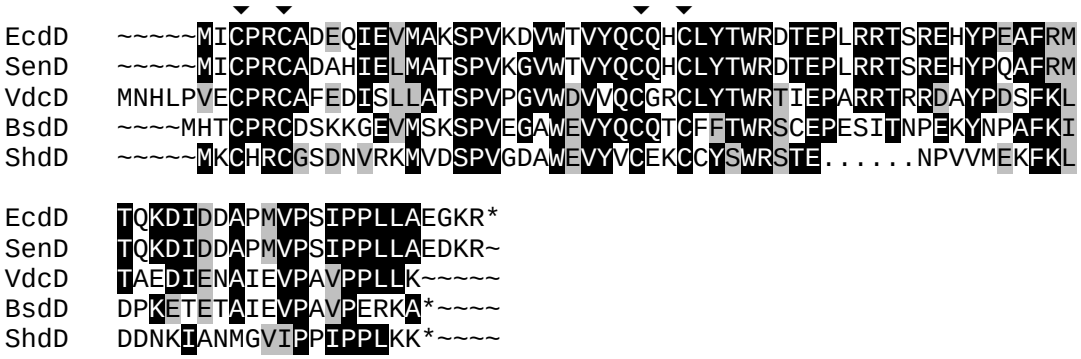


Figure 2.4- Comparing the D subunits from *E. coli* O157:H7, *S. enterica*, *Streptomyces* sp.D7, *B. subtilis*, and *S. hydroxybenzoicus*
EcdD is from *Escherichia coli* O157:H7, SenD is from *Salmonella enterica* subspecies enterica serovar Typhi, VdcD is from *Streptomyces* sp.D7, BsdD is from *Bacillus subtilis*, and ShdD is from *Sedimentibacter hydroxybenzoicus* JW/Z-1. The arrows indicate the four conserved cysteines. The alignment was created using the Boxshade program from GCG.

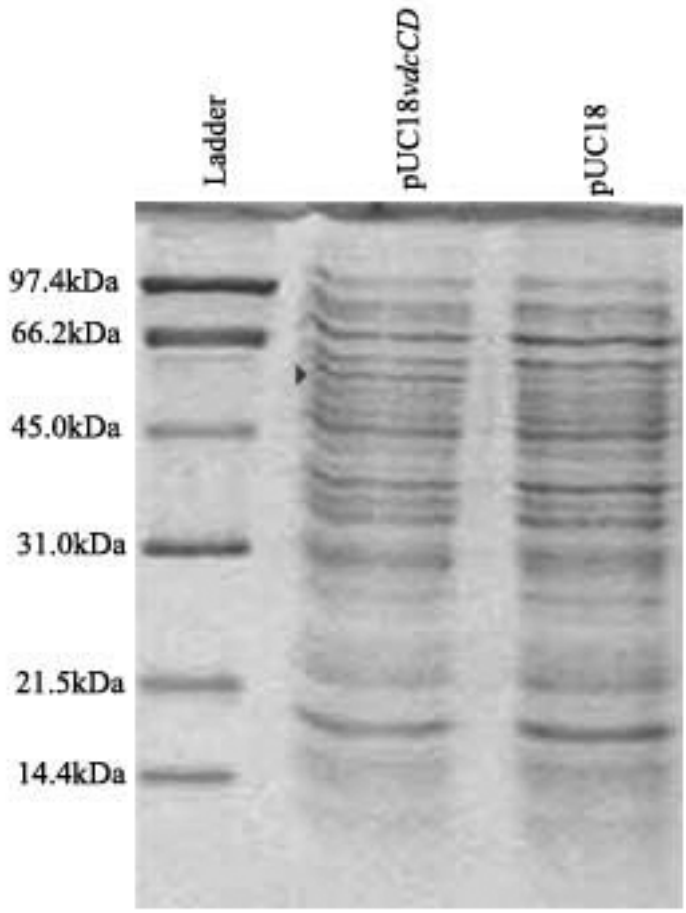


Figure 2.5- SDS-PAGE of VdcCD expressed by pUC18*vdcCD* in *E. coli* JM109
The arrow indicates the presence of VdcC. The gel was silver stained.

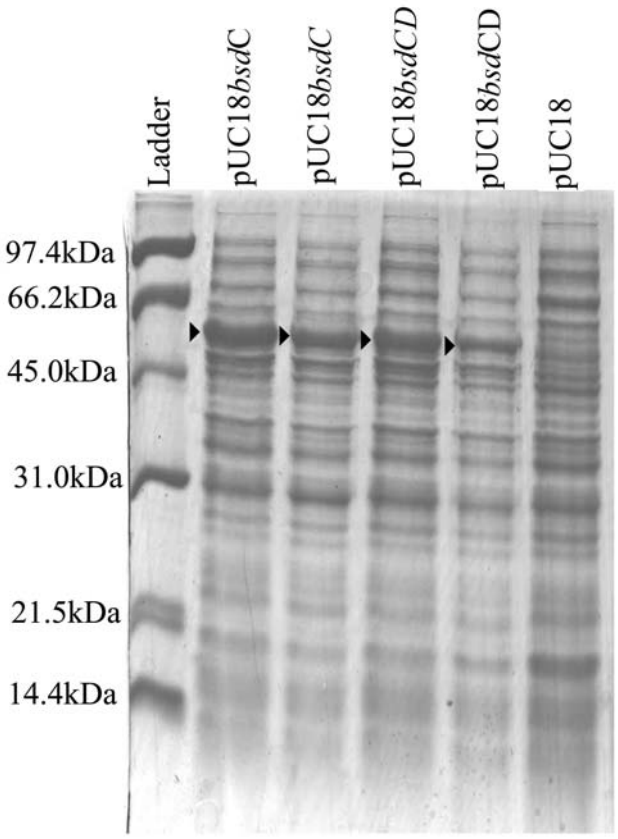


Figure 2.6- SDS-PAGE of BsdCD expressed by pUC18*bsdC* and pUC18*bsdCD* in *E. coli* JM109
The arrows indicate the position of BsdC. The gel was stained with Coomassie blue G-250.

```
published yclD  201 tccggcggtgccggaaCcgaaaggcttga 229
                    ||||||||||||||| ||||||||||||
sequenced bsdD  201 tccggcggtgccggaa.cgaaaggcttga 228
```

Figure 2.7 - Missense change in the sequence of *bsdD*
The mismatch results in a shortening of the BsdD amino acid sequence. The resulting BsdD is closer in projected molecular weight to that of the other D subunits.

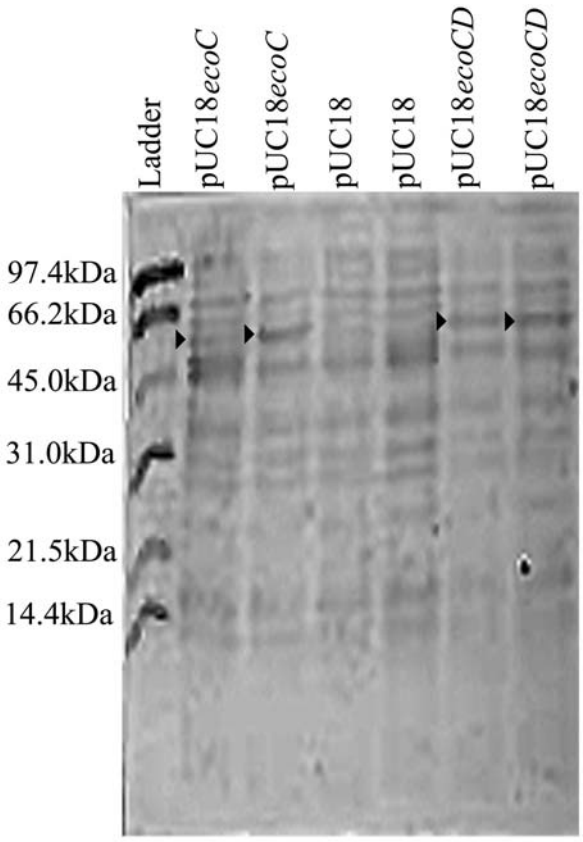


Figure 2.8- SDS-PAGE of EcdC and EcdCD expressed by pUC18*ecdC* and pUC18*ecdCD* in *E. coli* JM109
The arrows indicate the presence of EcdC. The gel was stained with Coomassie blue G-250.

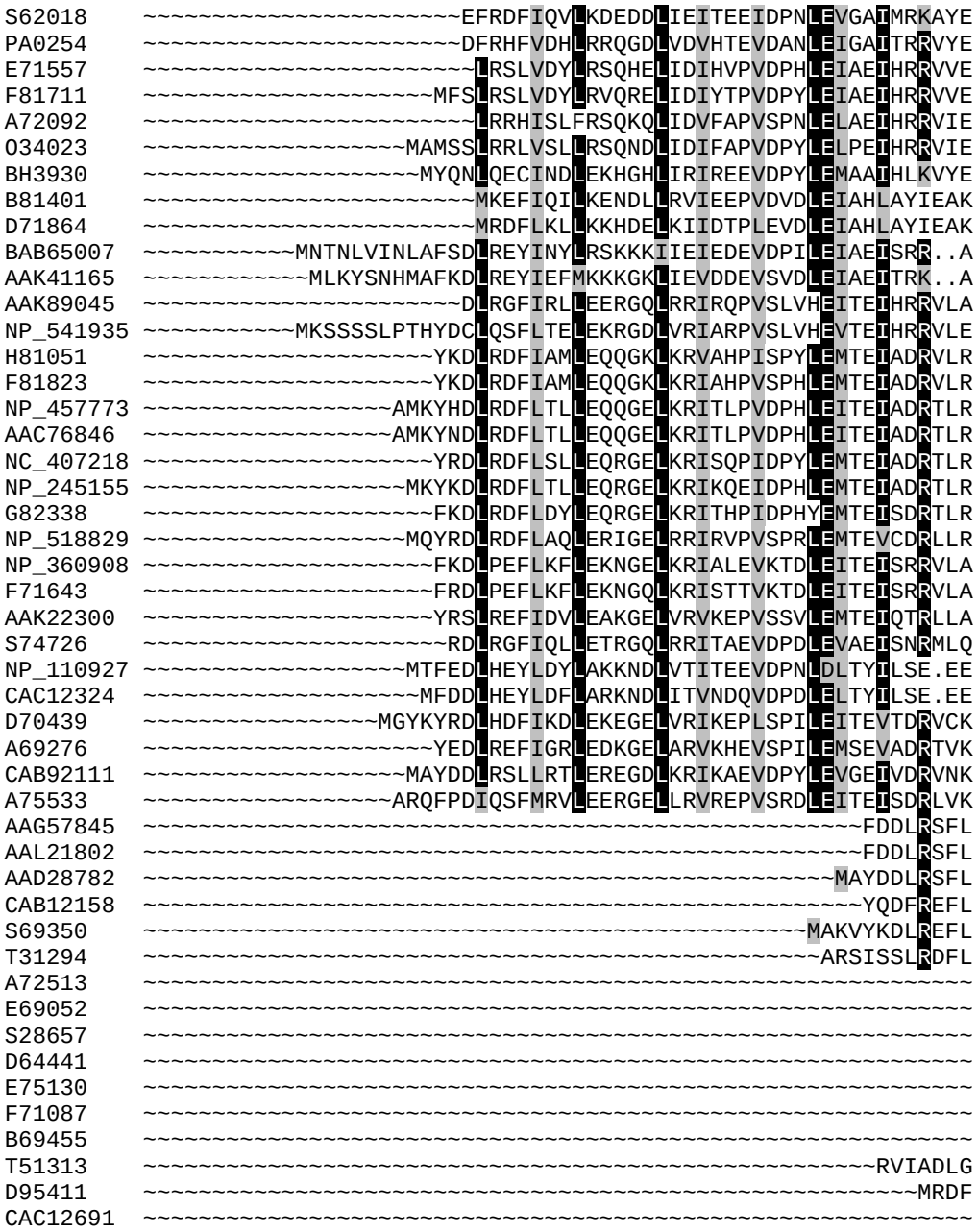
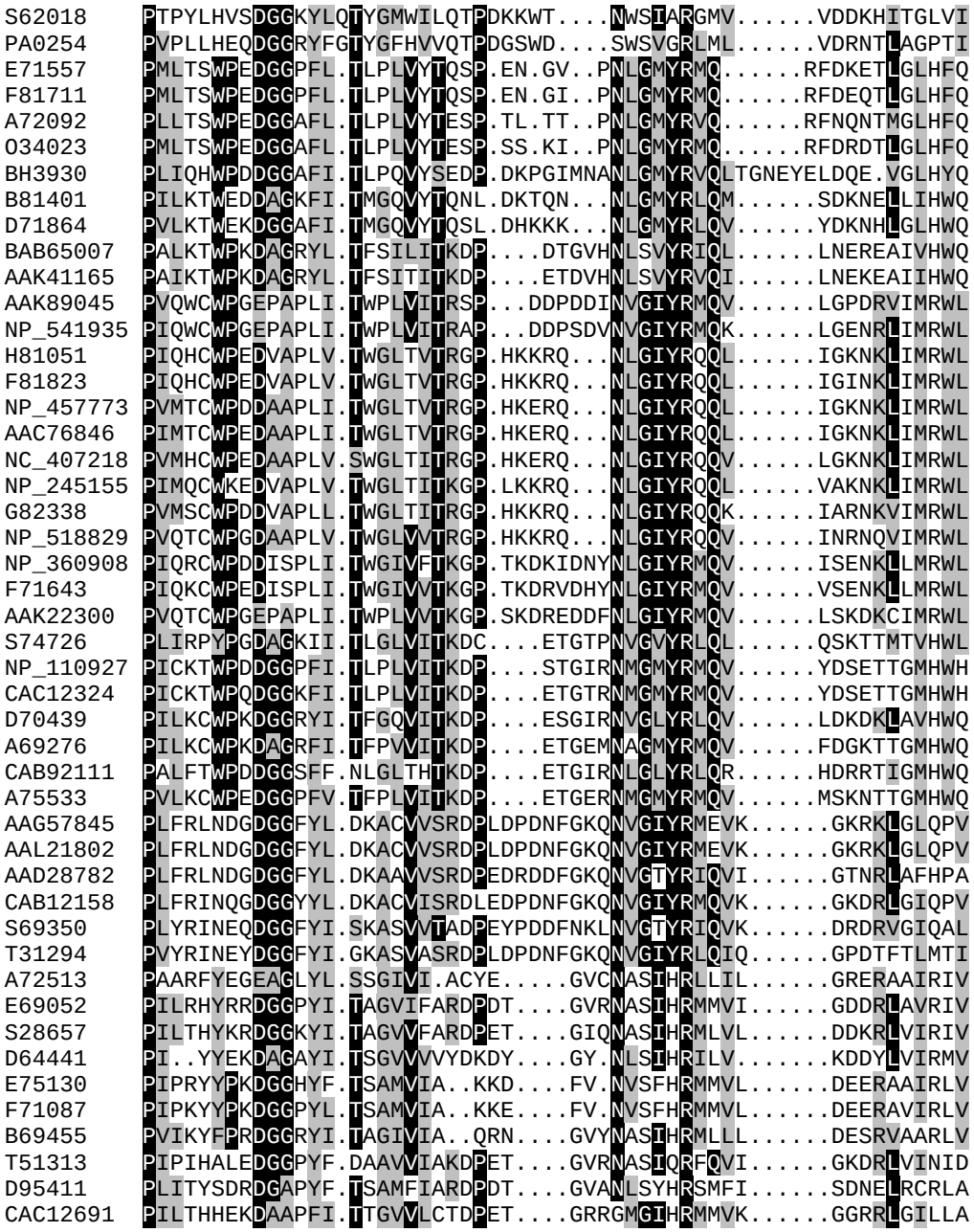
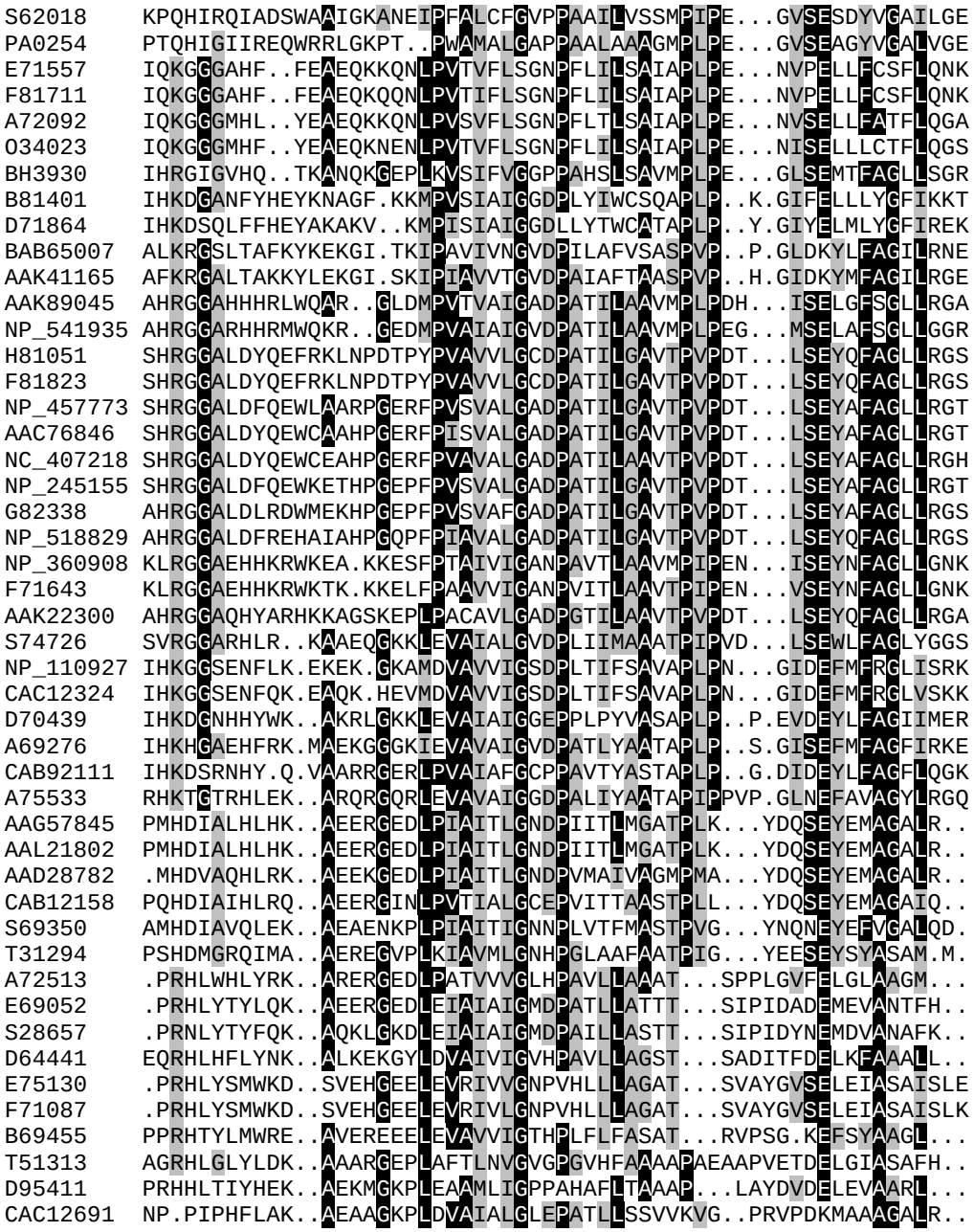


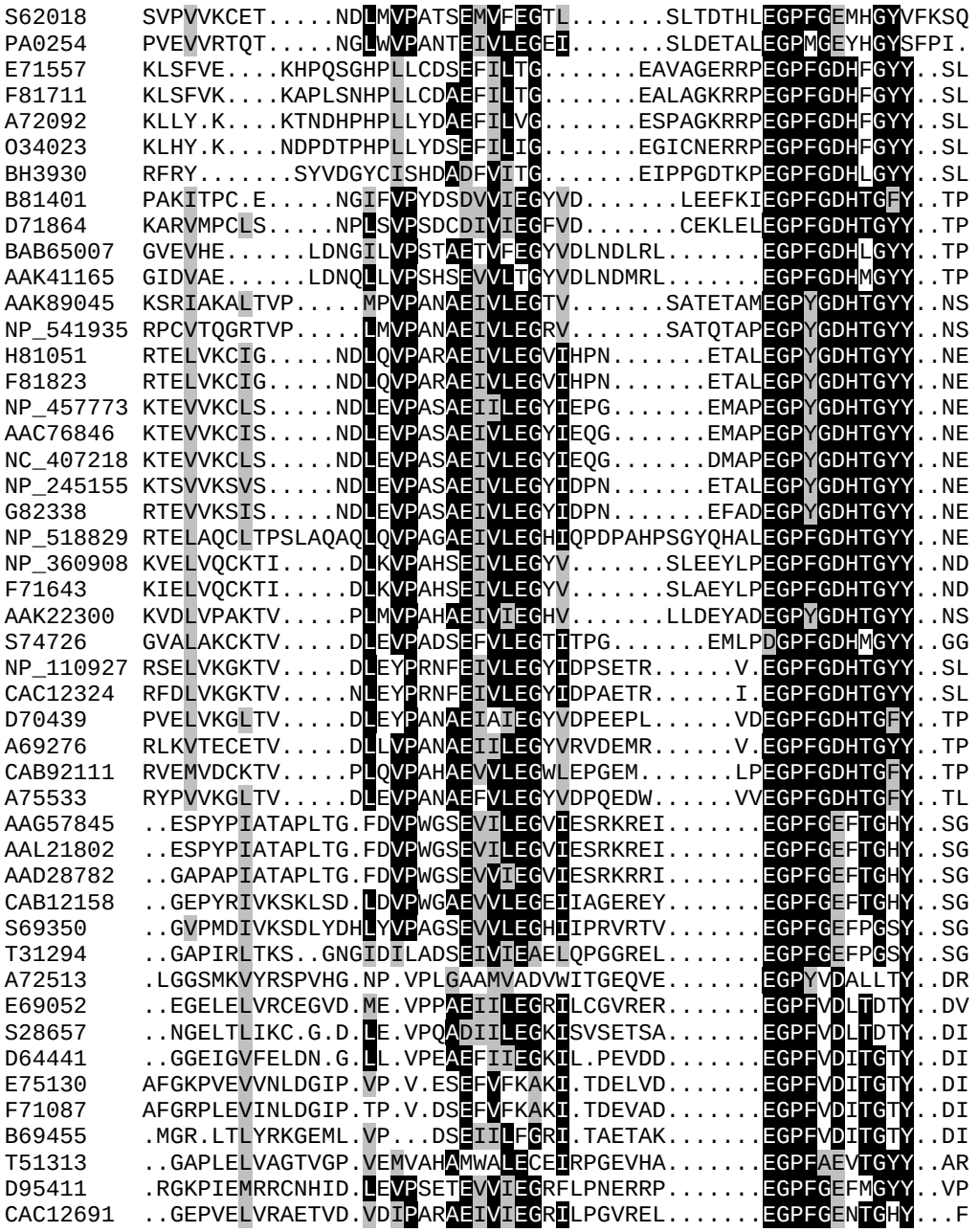
Figure 2.9 - Comparing homologues to ShdC (in groups of 60 amino acids)
The legend for this figure can be found at the end of this Boxshade comparison.

```
S62018    SHL..PAPLFKNLKGASKDLFSILGCPAGLR.....................SKEKGDH
PA0254    RRA..PAPLFHNIRDSLPGA.RVLGAPAGLR.....................ADRARAH
E71557    R..EGPALLFHQV........KGSPFPVLTNLFGTRRRVDLLFPDLSSD.........LF
F81711    N..EGPALLFHHV........KGSPFPVLTNLFGTQKRVDLLFPDLSSG.........IF
A72092    D..QGPALLFHNV........IGSSFPVLTNLFGTKHRVDQLFSQAPDN.........LI
O34023    N..QGPALLFHNV........QGASFPVLTNLFGTQKRVDQIFSKVPKG.........LI
BH3930    A..GGPALLFENV........KGSNYQAVSNLFGTMERSKFIFRQTWQS.........A.
B81401    KGEKGKALLFKNPIDKKLN..KQYKFPVLMNTFCNKKALNLAFG.........RDYKEVA
D71864    KPNGGKALLFTQPIRKEHNQIKTFGMPVLMNAFGSFKRLDLLLK.........TPIEDLQ
BAB65007  TYSKLPPLLFKNIKGYP.......NWSVITNIYYSIEAF........YDLFNTNKLEEIT
AAK41165  TYAHLPPLLFKRVKNYE.......NWKIVSNIFYSIESL........YEIFGTNKLESIS
AAK89045  ..DGGPALLFEQPVDHE...GKVRKMPLLANLFGTRQRIEWGLGLETGG......LPALG
NP_541935 ..ADGPALLFENPVDAE...GRTQTIPLLANLFGSERRIAWGLGRLPEE......LPLLA
H81051    ..AEGPALLFENPIKPD...GTRYGYPVLANLFGTPERVAMGM.....GADSVSKLREIG
F81823    ..AEGPALLFEHPVKPD...GTRYDYPVLANLFGTPERVAMGM.....GADSVSKLREIG
NP_457773 ..AGGPALLFENP........KGYAMPVLCNLFGTPKRVAMGM.....GQDDVSALREVG
AAC76846  ..AGGPALLFENP........KGYSMPVLCNLFGTPKRVAMGM.....GQEDVSALREVG
NC_407218 ..AGGPALLFENP........KGYSMPVLCNLFGTAKRVAMGM.....GQEDVSALRDVG
NP_245155 ..AGGPALLFENP........KGFDIPVLCNLFGTPKRVAMGM.....GQEQVSALRDVG
G82338    ..AKGPALLFENP........LGYDFPVLTNLFGTPERVAMGM.....GRQQVQELRDVG
NP_518829 ..AEGPAVVFERPADGA...QT.YDMPVLANLFGTPRRVALGM.....GAESLDELRDVG
NP_360908 ..QGGPALLFENVIKVD...GIKSDIPVLTNLYASINRICMGLKL.....KSPKELRELG
F71643    ..QGGPALLFENVIKAD...GSKSTIPVVTNLYASIKRICIGLNL.....KSPAELRELG
AAK22300  ..TGGPAVLFEHVLLPD...GSRSEMPALANLFGTVKRVAMGVTLGGEPRETAGELREVG
S74726    ..AGGPGLLFENV........KGSPFPVAVNLMGTVERICWAMNM.....DHPLELEDLG
NP_110927 RMGRGRTILFKNV........KGSQVPAVGNLFSTNEKLKAVLGDDP........YS.IG
CAC12324  RIGRGRTIQFNRV........KGSEVPAVGNLFSTYEKMKTVLGDDP........YQ.IG
D70439    MPGGGKALLFENP........KGYRIPVLTNLYGSEKRIKKALGYEN........LEDIG
A69276    ..AGGKALLFERP........KGYDIPVFMNAFGTERRMKLALEVER........LEEIG
CAB92111  ..AGGPALLFENV........KGSDLPLAMNVFGTDRRLLKALGLKS........YSDIS
A75533    K..GGPAVLFENV........VGSDYPVVMGLMGTRERMALAVGVND........LDELA
AAG57845  QALDDHGQLLKISEEVNAEPDLAAAANATGRIGDGAPALWFDNIRGFTDARVAMNTIGSW
AAL21802  HALDQQGQLLKISEEVNAEPDLAAAANATGRIGDGAPALWFDNIRGFTDARVAMNTIGSW
AAD28782  DTLEKEGQLLRITDEVLPEPDLAAAANATGRIGENAPALHFDNVKGFTDARIAMNVHGSW
CAB12158  AALEKEGQLLTVNEEVKPEPDLGASARAASNLGDKSPALLFNNIYGYHNARIAMNVIGSW
S69350    EVLEQEGQLIRVKEEVNPEPDIAAAGRAAANLGKNQPAVFFEKIKGYKYS.VVTNVHGSW
T31294    ELLEDAGQAITWSDAVMPEPGVRNIAVAASRDANGAPAIVFDNITGYPGKRLAVGVHGSW
A72513    ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
E69052    ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
S28657    ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~VIIKNVKG.YDLPIISGICNTR
D64441    ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
E75130    ~~~~~~~EILERFDDLVVIEKPVSKEIEITKFLLKYKDKPILFKDV.EGWEVAGNVWSTR
F71087    ~VMKMLREIVESFEDLVVIDKPVKKELELTKFLLKYKDKPVLFKDV.EGWEVAGNLWSSR
B69455    ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
T51313    RIIDRLEALGRLVRVRSEVDPRHDLAGIAARFEGGPQAVLFEKVAG.HAYPVFVGLYWSR
D95411    VRKLQERGDLLVVEREI.DPAHELAAVTHLAQKKWAKPVMFTNVKG.TRFPVVTNVYSTR
CAC12691  ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
```

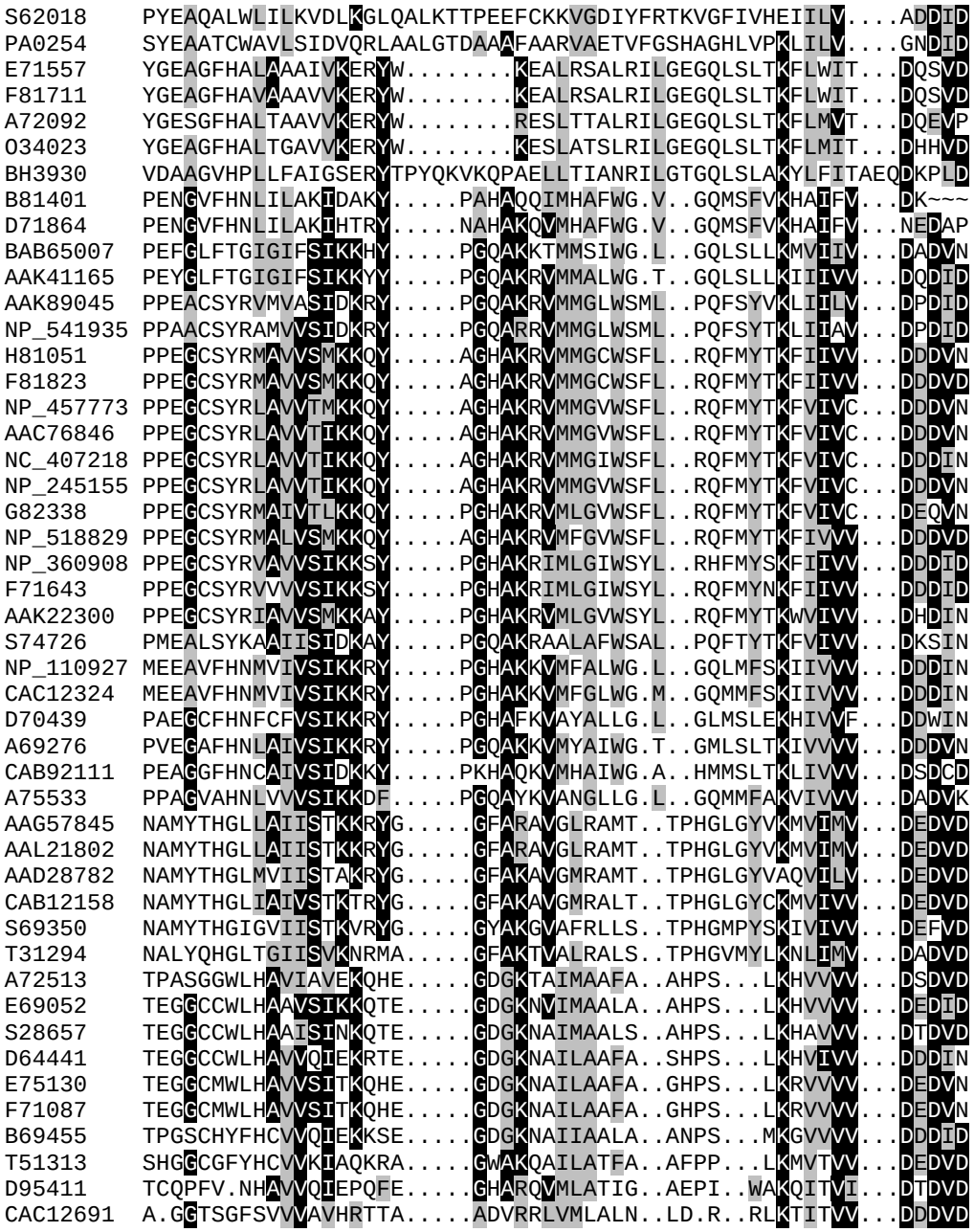
```
S62018     GRIAHHLGLDP.KTTIKEIIDYLLECKEKEPLPPITVPVSSAPCKTHILSEEKIHL.QSL
PA0254     SRLALHFGLPE.HSGPRDIVAMLRAAMRAEPIAPRR..LERGPVQENVWLGEQVDL.TRF
E71557     EQIIHLLSSP..PS.FSSLWKHRSLFKRGISALGMRKRHLR...PSPFLYQDAPNLSQ.L
F81711     DQIAILLSSP..PS.FSSLWQHRSLLKRGLSSLGLRKQRFR...PSPFLHQDAPNLLQ.L
A72092     ARVAHLISST..PK.LSSLWKSRDLLKRISS.LGLKKARFR...RFPFVSMSSVNLDH.L
O34023     PQVIHLLSSP..PK.LSQLWKHRNLLLRGLS.LGLRKARFL...KFPHKKMASVDLHQ.L
BH3930     ENVVALRNDP..MSALKHPFAHARTALAASKALPLKKSRL....PAGF...EEITISD.L
B81401     EEISKLIKL.HIPTSFKAKMDFFMNLLSFKNIPPKRLKKNKAL...YDYEILN.SLEE.L
D71864     QRMQAFLHF.DAPKNFTESLKILKDLWDLRHVFPKKTARPKDLITKQDKEV...NLWD.L
BAB65007   ESFLDKMSG.SLPVSFSNKIRNLFDILKLGKILPKSKKPA.......FKEDNNLNFNK.I
AAK41165   EGFLSNLS..NMPITFFDKIKSLREILGLGKVMPKAKSPS.......FKEEKNLDLTK.I
AAK89045   .QKLTELREPRPPKSMAEAWSKLPLLRAALSMRQRNVS..RAPVQEKVLTRDAVDLAR.L
NP_541935  .EMLAELRAPKPPRSAGEIWEKLPMAKAALNMRPRQVS..RAPVHGLVMEGASVNLDT.L
H81051     .QTLAYLKEPEPPKGIKDAFSKLPLLKDIWSMAPNVVK..NAPCQEIVWEGEDVDLYQ.L
F81823     .QTLAYLKEPEPPKGIKDAFSKLPLLKDIWSMAPNVVK..NAPCQEIVWEGEDVDLYQ.L
NP_457773  .KLLAFLKEPEPPKGFRDLFDKLPQFKQVLNMPTKRLR..GAPCQQKIASGDDVDLTR.L
AAC76846   .KLLAFLKEPEPPKGFRDLFDKLPQFKQVLNMPTKRLR..GAPCQQKIVSGDDVDLNR.I
NC_407218  .KLLAFLKEPDPPKGFRDLFDKLPKFKQVLNMPTKRLN..SAPCQEQVWQGEDVDLSR.I
NP_245155  .KLLAFLKEPEPPKGFKDFLSSIPQFKQVLNMPTKVLG..KADCQQVVLKDEDVDLYK.L
G82338     .QWLAYLKEPEPPRGLKELIEKLPVFKQVLNMPVKRLR..RAPCQEIVWQGDAVDLDK.I
NP_518829  .RLLSALKEPEPPRGLREAGKLWTMAKAVWDMAPRKVS..SPACQEIVLEGDDVDLSR.I
NP_360908  .VLLAFLKQPQPPASFKETLSMLPLAKRIFAMSPKTVS..KAACHEIVIDKPNINI...L
F71643     .ALLAFLKQPQIPESFKETLSILPLAKRIFSMSPKTIA..KGACHEVVIVKPNINI...L
AAK22300   .ELLAFLRQPQPPKGLKDALDMLPLAKTVMSMRPGTVK..KAPVQEVVLTGDQIDLSK.L
S74726     .KKLALLQQPKPPKKISQAIDFGKVLFDVLKAKPGRNF..FPPCQEVVIDGENLDLNQ.I
NP_110927  VKIADIVKPPRDSESFIG..KGLEMMRELSGLRPKVSNSIPSGYSEL....EKVDLYR.Y
CAC12324   RRIVEIAQPPGDSESFIG..KGIEMMRELGGLRPKIAGSLPSNYDEL....DRVDLFR.Y
D70439     WKLYRILK.PEVPKTFLEKIKKLPELKKLNDAIPKVVKRGKVQEEVIM...GDINLED.L
A69276     ERLLSALE.FR.PSSFMDALKGVGMLKDFMSFIPK..KTGKAPCKEVV...AE.SLDK.F
CAB92111   DKIGGLLR.PELPQGFVGVREAFGKLGTMTHVPPKKVKPGSAPVQETVLTGDDVDLER.L
A75533     QKIRALID.LGGGGSRFGLLSNLPKLRDAMNLPPRRVKT..APVQEVVWRGDEVDLSK.I
AAG57845   QNHAISLGLPPN.TPVKKQIDEFIRRWDNFPIAPE..RRANPAWAQNTVDGDEINLFDIL
AAL21802   QNHAISLGLPPN.TPVKKQIDEFIRRWDNFPVAPE..RRANPGWAENTVDGDAINLFDIL
AAD28782   ANHALALGLPKN.TPVKEQVEEFARRWDAFPVAPE..RREEAPWRENTQEGEDVDLFSVL
CAB12158   PNHAMMLGMPKD.TPVKEQFFEFAKRYDQFPMPVK..REETAPFHENEIT.EDINLFDIL
S69350     QNHALMLGLDKN.TSTKDQFYELNRRWDKFPVPPNVVKREAAPCKENVID.KDINLFEIL
T31294     DNIALLLGRPKG.TTIRELFFEIAGRWGDQEAQISFVPEAQAPVHECRIE.QDINLYDVL
A72513     ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~FEG.L
E69052     ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~L
S28657     EKIAKSINCEVS.EITQKIIEASDN.....PIKVDKFTDF.....SDY.NTTEANLDK.I
D64441     ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~I
E75130     ERIAKFLSTDNK.GLLEILYRAIEN.....PKEYAVVDKA.....EFLKNKESVNLLE.L
F71087     ERIAKFLNTDNK.GLLELLYEAMEK.....PKPFSVVEKA.....EFLKNREKVNLLE.L
B69455     ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~DVNLQE.L
T51313     ELLGALFDQPET.ALPQHVAASIKSWQSAPVDPLVVADGP.....VLEVTEAEVDLST.L
D95411     ERLGEVIGIDAG.DFCRQWSRLSSLGSAEMREPLVPANQP........PGYDEVKLSD.L
CAC12691   ~~~~~~~~~~~~~~~~~~~~~~~~~~~MHRSRRGTRPRSK.....EVIHRHPDDLLSL.L
```

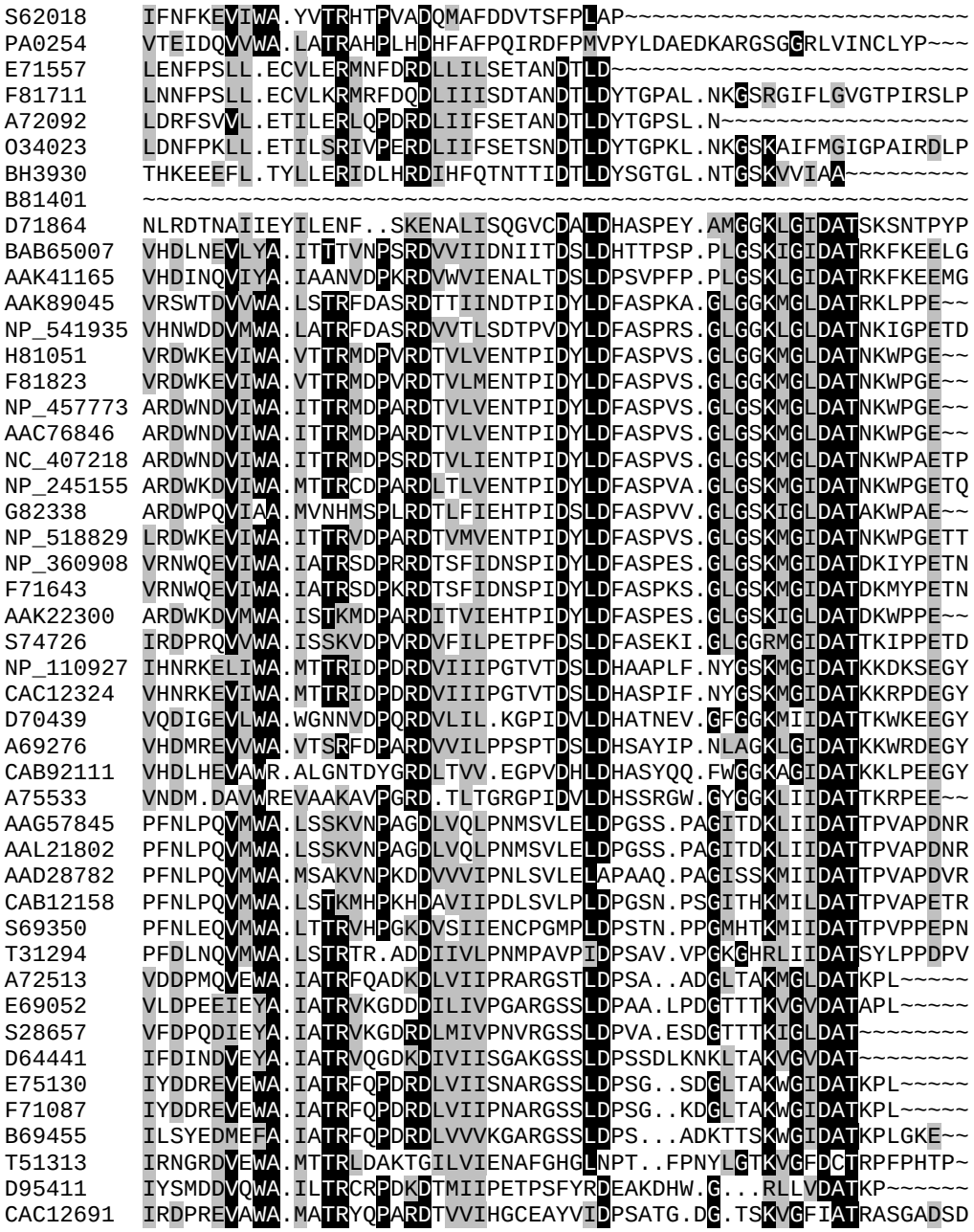

S62018 PTPYLHVSDGGKYLQTYGMWILQTPDKKWT....NWSIARGMV......VDDKHITGLVI
PA0254 PVPLLHEQDGGRYFGTYGFHVVQTPDGSWD....SWSVGRLML......VDRNTLAGPTI
E71557 PMLTSWPEDGGPFL.TLPLVYTQSP.EN.GV..PNLGMYRMQ......RFDKETLGLHFQ
F81711 PMLTSWPEDGGPFL.TLPLVYTQSP.EN.GI..PNLGMYRMQ......RFDEQTLGLHFQ
A72092 PLLTSWPEDGGAFL.TLPLVYTESP.TL.TT..PNLGMYRVQ......RFNQNTMGLHFQ
O34023 PMLTSWPEDGGAFL.TLPLVYIESP.SS.KI..PNLGMYRMQ......RFDRDTLGLHFQ
BH3930 PLIQHWPDDGGAFI.TLPQVYSEDP.DKPGIMNANLGMYRVQLTGNEYELDQE.VGLHYQ
B81401 PILKTWEDDAGKFI.TMGQVYTQNL.DKTQN...NLGMYRLQM......SDKNELLIHWQ
D71864 PVLKTWEKDGGAFI.TMGQVYTQSL.DHKKK...NLGMYRLQV......YDKNHLGLHWQ
BAB65007 PALKTWPKDAGRYL.TFSILITKDP....DTGVHNLSVYRIQL......LNEREAIVHWQ
AAK41165 PAIKTWPKDAGRYL.TFSITITKDP....ETDVHNLSVYRVQI......LNEKEAIIHWQ
AAK89045 PVQWCWPGEPAPLI.TWPLVITRSP...DDPDDINVGIYRMQV......LGPDRVIMRWL
NP_541935 PIQWCWPGEPAPLI.TWPLVITRAP...DDPSDVNVGIYRMQK......LGENRLIMRWL
H81051 PIQHCWPEDVAPLV.TWGLTVTRGP.HKKRQ...NLGIYRQQL......IGKNKLIMRWL
F81823 PIQHCWPEDVAPLV.TWGLTVTRGP.HKKRQ...NLGIYRQQL......IGINKLIMRWL
NP_457773 PVMTCWPDDAAPLI.TWGLTVTRGP.HKERQ...NLGIYRQQL......IGKNKLIMRWL
AAC76846 PIMTCWPEDAAPLI.TWGLTVTRGP.HKERQ...NLGIYRQQL......IGKNKLIMRWL
NC_407218 PVMHCWPEDAAPLV.SWGLTITRGP.HKERQ...NLGIYRQQV......LGKNKLIMRWL
NP_245155 PIMQCWKEDVAPLV.TWGLTITKGP.LKKRQ...NLGIYRQQL......VAKNKLIMRWL
G82338 PVMSCWPDDVAPLL.TWGLTITRGP.HKKRQ...NLGIYRQQK......IARNKVIMRWL
NP_518829 PVQTCWPGDAAPLV.TWGLVVTRGP.HKKRQ...NLGIYRQQV......INRNQVIMRWL
NP_360908 PIQRCWPDDISPLI.TWGIVFTKGP.TKDKIDNYNLGIYRMQV......ISENKLLMRWL
F71643 PIQKCWPEDISPLI.TWGIVVTKGP.TKDRVDHYNLGIYRMQV......VSENKLLMRWL
AAK22300 PVQTCWPGEPAPLI.TWPLVVTKGP.SKDREDDFNLGIYRMQV......LSKDKCIMRWL
S74726 PLIRPYPGDAGKII.TLGLVITKDC....ETGTPNVGVYRLQL......QSKTTMTVHWL
NP_110927 PICKTWPDDGGPFI.TLPLVITKDP....STGIRNMGMYRMQV......YDSETTGMHWH
CAC12324 PICKTWPQDGGKFI.TLPLVITKDP....ETGTRNMGMYRMQV......YDSETTGMHWH
D70439 PILKCWPKDGGRYI.TFGQVITKDP....ESGIRNVGLYRLQV......LDKDKLAVHWQ
A69276 PILKCWPKDAGRFI.TFPVVITKDP....ETGEMNAGMYRMQV......FDGKTTGMHWQ
CAB92111 PALFTWPDDGGSFF.NLGLTHTKDP....ETGIRNLGLYRLQR......HDRRTIGMHWQ
A75533 PVLKCWPEDGGPFV.TFPLVITKDP....ETGERNMGMYRMQV......MSKNTTGMHWQ
AAG57845 PLFRLNDGDGGFYL.DKACVVSRDPLDPDNFGKQNVGIYRMEVK......GKRKLGLQPV
AAL21802 PLFRLNDGDGGFYL.DKACVVSRDPLDPDNFGKQNVGIYRMEVK......GKRKLGLQPV
AAD28782 PLFRLNDGDGGFYL.DKAAVVSRDPEDRDDFGKQNVGTYRIQVI......GTNRLAFHPA
CAB12158 PLFRINQGDGGYYL.DKACVISRDLEDPDNFGKQNVGIYRMQVK......GKDRLGIQPV
S69350 PLYRINEQDGGFYI.SKASVVTADPEYPDDFNKLNVGTYRIQVK......DRDRVGIQAL
T31294 PVYRINEYDGGFYI.GKASVASRDPLDPDNFGKQNVGIYRLQIQ......GPDTFTLMTI
A72513 PAARFYEGEAGLYL.SSGIVI.ACYE.....GVCNASIHRLLIL......GRERAAIRIV
E69052 PILRHYRRDGGPYI.TAGVIFARDPDT....GVRNASIHRMMVI......GDDRLAVRIV
S28657 PILTHYKRDGGKYI.TAGVVFARDPET....GIQNASIHRMLVL......DDKRLVIRIV
D64441 PI..YYEKDAGAYI.TSGVVVVYDKDY....GY.NLSIHRILV.......KDDYLVIRMV
E75130 PIPRYYPKDGGHYF.TSAMVIA..KKD....FV.NVSFHRMMVL......DEERAAIRLV
F71087 PIPKYYPKDGGPYL.TSAMVIA..KKE....FV.NVSFHRMMVL......DEERAVIRLV
B69455 PVIKYFPRDGGRYI.TAGIVIA..QRN....GVYNASIHRMLLL......DESRVAARLV
T51313 PIPIHALEDGGPYF.DAAVVIAKDPET....GVRNASIQRFQVI......GKDRLVINID
D95411 PLITYSDRDGAPYF.TSAMFIARDPDT....GVANLSYHRSMFI......SDNELRCRLA
CAC12691 PILTHHEKDAAPFI.TTGVVLCTDPET....GRRGMGIHRMMVK......GGRRLGILLA


S62018 KPQHIRQIADSWAAIGKANEIPFALCFGVPPAAILVSSMPIPE...GVSESDYVGAILGE
PA0254 PTQHIGIIREQWRRLGKPT..PWAMALGAPPAALAAAGMPLPE...GVSEAGYVGALVGE
E71557 IQKGGGAHF..FEAEQKKQNLPVTVFLSGNPFLILSAIAPLPE...NVPELLFCSFLQNK
F81711 IQKGGGAHF..FEAEQKQQNLPVTIFLSGNPFLILSAIAPLPE...NVPELLFCSFLQNK
A72092 IQKGGGMHL..YEAEQKKQNLPVSVFLSGNPFLTLSAIAPLPE...NVSELLFATFLQGA
O34023 IQKGGGMHF..YEAEQKNENLPVTVFLSGNPFLILSAIAPLPE...NISELLLCTFLQGS
BH3930 IHRGIGVHQ..TKANQKGEPLKVSIFVGGPPAHSLSAVMPLPE...GLSEMTFAGLLSGR
B81401 IHKDGANFYHEYKNAGF.KKMPVSIAIGGDPLYIWCSQAPLP..K.GIFELLLYGFIKKT
D71864 IHKDSQLFFHEYAKAKV..KMPISIAIGGDLLYTWCATAPLP..Y.GIYELMLYGFIREK
BAB65007 ALKRGSLTAFKYKEKGI.TKIPAVIVNGVDPILAFVSASPVP..P.GLDKYLFAGILRNE
AAK41165 AFKRGALTAKKYLEKGI.SKIPIAVVTGVDPAIAFTAASPVP..H.GIDKYMFAGILRGE
AAK89045 AHRGGAHHHRLWQAR..GLDMPVTVAIGADPATILAAVMPLPDH...ISELGFSGLLRGA
NP_541935 AHRGGARHHRMWQKR..GEDMPVAIAIGVDPATILAAVMPLPEG...MSELAFSGLLGGR
H81051 SHRGGALDYQEFRKLNPDTPYPVAVVLGCDPATILGAVTPVPDT...LSEYQFAGLLRGS
F81823 SHRGGALDYQEFRKLNPDTPYPVAVVLGCDPATILGAVTPVPDT...LSEYQFAGLLRGS
NP_457773 SHRGGALDFQEWLAARPGERFPVSVALGADPATILGAVTPVPDT...LSEYAFAGLLRGT
AAC76846 SHRGGALDYQEWCAAHPGERFPISVALGADPATILGAVTPVPDT...LSEYAFAGLLRGT
NC_407218 SHRGGALDYQEWCEAHPGERFPVAVALGADPATILAAVTPVPDT...LSEYAFAGLLRGH
NP_245155 SHRGGALDFQEWKETHPGEPFPVSVALGADPATILGAVTPVPDT...LSEYAFAGLLRGT
G82338 AHRGGALDLRDWMEKHPGEPFPVSVAFGADPATILGAVTPVPDT...LSEYAFAGLLRGS
NP_518829 AHRGGALDFREHAIAHPGQPFPIAVALGADPATILGAVTPVPDT...LSEYQFAGLLRGS
NP_360908 KLRGGAEHHKRWKEA.KKESFPTAIVIGANPAVTLAAVMPIPEN...ISEYNFAGLLGNK
F71643 KLRGGAEHHKRWKTK.KKELFPAAVVIGANPVITLAAVTPIPEN...VSEYNFAGLLGNK
AAK22300 AHRGGAQHYARHKKAGSKEPLPACAVLGADPGTILAAVTPVPDT...LSEYQFAGLLRGA
S74726 SVRGGARHLR..KAAEQGKKLEVAIALGVDPLIIMAAATPIPVD...LSEWLFAGLYGGS
NP_110927 IHKGGSENFLK.EKEK.GKAMDVAVVIGSDPLTIFSAVAPLPN...GIDEFMFRGLISRK
CAC12324 IHKGGSENFQK.EAQK.HEVMDVAVVIGSDPLTIFSAVAPLPN...GIDEFMFRGLVSKK
D70439 IHKDGNHHYWK..AKRLGKKLEVAIAIGGEPPLPYVASAPLP..P.EVDEYLFAGIIMER
A69276 IHKHGAEHFRK.MAEKGGGKIEVAVAIGVDPATLYAATAPLP..S.GISEFMFAGFIRKE
CAB92111 IHKDSRNHY.Q.VAARRGERLPVAIAFGCPPAVTYASTAPLP..G.DIDEYLFAGFLQGK
A75533 RHKTGTRHLEK..ARQRGQRLEVAVAIGGDPALIYAATAPIPPVP.GLNEFAVAGYLRGQ
AAG57845 PMHDIALHLHK..AEERGEDLPIAITLGNDPIITLMGATPLK...YDQSEYEMAGALR..
AAL21802 PMHDIALHLHK..AEERGEDLPIAITLGNDPIITLMGATPLK...YDQSEYEMAGALR..
AAD28782 .MHDVAQHLRK..AEEKGEDLPIAITLGNDPVMAIVAGMPMA...YDQSEYEMAGALR..
CAB12158 PQHDIAIHLRQ..AEERGINLPVTIALGCEPVITTAASTPLL...YDQSEYEMAGAIQ..
S69350 AMHDIAVQLEK..AEAENKPLPIAITIGNNPLVTFMASTPVG...YNQNEYEFVGALQD.
T31294 PSHDMGRQIMA..AEREGVPLKIAVMLGNHPGLAAFAATPIG...YEESEYSYASAM.M.
A72513 .PRHLWHLYRK..ARERGEDLPATVVVGLHPAVLLAAAT...SPPLGVFELGLAAGM...
E69052 .PRHLYTYLQK..AEERGEDLEIAIAIGMDPATLLATTT...SIPIDADEMEVANTFH..
S28657 .PRNLYTYFQK..AQKLGKDLEIAIAIGMDPAILLASTT...SIPIDYNEMDVANAFK..
D64441 EQRHLHFLYNK..ALKEKGYLDVAIVIGVHPAVLLAGST...SADITFDELKFAAALL..
E75130 .PRHLYSMWKD..SVEHGEELEVRIVVGNPVHLLLAGAT...SVAYGVSELEIASAISLE
F71087 .PRHLYSMWKD..SVEHGEELEVRIVLGNPVHLLLAGAT...SVAYGVSELEIASAISLK
B69455 PPRHTYLMWRE..AVEREEELEVAVVIGTHPLFLFASAT...RVPSG.KEFSYAAGL...
T51313 AGRHLGLYLDK..AAARGEPLAFTLNVGVGPGVHFAAAAPAEAAPVETDELGIASAFH..
D95411 PRHHLTIYHEK..AEKMGKPLEAAMLIGPPAHAFLTAAAP...LAYDVDELEVAARL...
CAC12691 NP.PIPHFLAK..AEAAGKPLDVAIALGLEPATLLSSVVKVG..PRVPDKMAAAGALR..


S62018 SVPVVKCET.....NDLMVPATSEMVFEGTL.......SLTDTHLEGPFGEMHGYVFKSQ
PA0254 PVEVVRTQT.....NGLWVPANTEIVLEGEI.......SLDETALEGPMGEYHGYSFPI.
E71557 KLSFVE....KHPQSGHPLLCDSEFILTG.......EAVAGERRPEGPFGDHFGYY..SL
F81711 KLSFVK....KAPLSNHPLLCDAEFILTG.......EALAGKRRPEGPFGDHFGYY..SL
A72092 KLLY.K....KTNDHPHPLLYDAEFILVG.......ESPAGKRRPEGPFGDHFGYY..SL
O34023 KLHY.K....NDPDTPHPLLYDSEFILIG.......EGICNERRPEGPFGDHFGYY..SL
BH3930 RFRY.......SYVDGYCISHDADFVITG.......EIPPGDTKPEGPFGDHLGYY..SL
B81401 PAKITPC.E.....NGIFVPYDSDVVIEGYVD.......LEEFKIEGPFGDHTGFY..TP
D71864 KARVMPCLS.....NPLSVPSDCDIVIEGFVD.......CEKLELEGPFGDHTGYY..TP
BAB65007 GVEVHE......LDNGILVPSTAETVFEGYVDLNDLRL.......EGPFGDHLGYY..TP
AAK41165 GIDVAE......LDNQLLVPSHSEVVLTGYVDLNDMRL.......EGPFGDHMGYY..TP
AAK89045 KSRIAKALTVP.....MPVPANAEIVLEGTV.......SATETAMEGPYGDHTGYY..NS
NP_541935 RPCVTQGRTVP.....LMVPANAEIVLEGRV.......SATQTAPEGPYGDHTGYY..NS
H81051 RTELVKCIG.....NDLQVPARAEIVLEGVIHPN.......ETALEGPYGDHTGYY..NE
F81823 RTELVKCIG.....NDLQVPARAEIVLEGVIHPN.......ETALEGPYGDHTGYY..NE
NP_457773 KTEVVKCLS.....NDLEVPASAEIILEGYIEPG.......EMAPEGPYGDHTGYY..NE
AAC76846 KTEVVKCIS.....NDLEVPASAEIVLEGYIEQG.......EMAPEGPYGDHTGYY..NE
NC_407218 KTEVVKCLS.....NDLEVPASAEIVLEGYIEQG.......DMAPEGPYGDHTGYY..NE
NP_245155 KTSVVKSVS.....NDLEVPASAEIVLEGYIDPN.......ETALEGPYGDHTGYY..NE
G82338 RTEVVKSIS.....NDLEVPASAEIVLEGYIDPN.......EFADEGPYGDHTGYY..NE
NP_518829 RTELAQCLTPSLAQAQLQVPAGAEIVLEGHIQPDPAHPSGYQHALEGPFGDHTGYY..NE
NP_360908 KVELVQCKTI.....DLKVPAHSEIVLEGYV.......SLEEYLPEGPFGDHTGYY..ND
F71643 KIELVQCKTI.....DLKVPAHSEIVLEGYV.......SLAEYLPEGPFGDHTGYY..ND
AAK22300 KVDLVPAKTV.....PLMVPAHAEIVIEGHV.......LLDEYADEGPYGDHTGYY..NS
S74726 GVALAKCKTV.....DLEVPADSEFVLEGTITPG.......EMLPDGPFGDHMGYY..GG
NP_110927 RSELVKGKTV.....DLEYPRNFEIVLEGYIDPSETR......V.EGPFGDHTGYY..SL
CAC12324 RFDLVKGKTV.....NLEYPRNFEIVLEGYIDPAETR......I.EGPFGDHTGYY..SL
D70439 PVELVKGLTV.....DLEYPANAEIAIEGYVDPEEPL......VDEGPFGDHTGFY..TP
A69276 RLKVTECETV.....DLLVPANAEIILEGYVRVDEMR......V.EGPFGDHTGYY..TP
CAB92111 RVEMVDCKTV.....PLQVPAHAEVVLEGWLEPGEM.......LPEGPFGDHTGFY..TP
A75533 RYPVVKGLTV.....DLEVPANAEFVLEGYVDPQEDW......VVEGPFGDHTGFY..TL
AAG57845 ..ESPYPIATAPLTG.FDVPWGSEVILEGVIESRKREI.......EGPFGEFTGHY..SG
AAL21802 ..ESPYPIATAPLTG.FDVPWGSEVILEGVIESRKREI.......EGPFGEFTGHY..SG
AAD28782 ..GAPAPIATAPLTG.FDVPWGSEVVIEGVIESRKRRI.......EGPFGEFTGHY..SG
CAB12158 ..GEPYRIVKSKLSD.LDVPWGAEVVLEGEIIAGEREY.......EGPFGEFTGHY..SG
S69350 ..GVPMDIVKSDLYDHLYVPAGSEVVLEGHIIPRVRTV.......EGPFGEFPGSY..SG
T31294 ..GAPIRLTKS..GNGIDILADSEIVIEAELQPGGREL.......EGPFGEFPGSY..SG
A72513 .LGGSMKVYRSPVHG.NP.VPLGAAMVADVWITGEQVE.......EGPYVDALLTY..DR
E69052 ..EGELELVRCEGVD.ME.VPPAEIILEGRILCGVRER.......EGPFVDLTDTY..DV
S28657 ..NGELTLIKC.G.D.LE.VPQADIILEGKISVSETSA.......EGPFVDLTDTY..DI
D64441 ..GGEIGVFELDN.G.LL.VPEAEFIIEGKIL.PEVDD.......EGPFVDITGTY..DI
E75130 AFGKPVEVVNLDGIP.VP.V.ESEFVFKAKI.TDELVD.......EGPFVDITGTY..DI
F71087 AFGRPLEVINLDGIP.TP.V.DSEFVFKAKI.TDEVAD.......EGPFVDITGTY..DI
B69455 .MGR.LTLYRKGEML.VP...DSEIILFGRI.TAETAK.......EGPFVDITGTY..DI
T51313 ..GAPLELVAGTVGP.VEMVAHAMWALECEIRPGEVHA.......EGPFAEVTGYY..AR
D95411 .RGKPIEMRRCNHID.LEVPSETEVVIEGRFLPNERRP.......EGPFGEFMGYY..VP
CAC12691 ..GEPVELVRAETVD.VDIPARAEIVIEGRILPGVREL.......EGPFGENTGHY...F

```
S62018     GHPCPLYTVKAMSYRDNAILPVSNPGLCTDETHTLIGSLVATEAKELAIESGLPILDAFM
PA0254     GKPQPLFHVHALSFRDQPILPICVAGTPPEENHTIWGTMISAQLLDVAQNAGLPVDMVWC
E71557     THDFPIFKCNCLYHKKDAIYPATVVGKPFQE.DFFLGNKLQELLSPLFPLIMPGVQDLKS
F81711     THDFPIFNCQHLYHKKDAIYPATIVGKPFQE.DFFLGNKLQEILSPLFPLIMPGVQDLKS
A72092     QHDFPEFHCHKIYHRKDAIYPATVVGKPYQE.DFYIGNKLQEYLSPLFPLVMPGVRRLKS
O34023     QHDFPAFKCRKIYHRKDAIYPATIVGKPYQE.DFYLGNKLQEYLSPLFPMVMPGVRQLKS
BH3930     IHDFPVMKVHKVYAKQGAIWPFTVVGRPPQE.DTSFGALIHELTGDAVKLEIPGVKEVHA
B81401     AELFPVMKVEKIYAKKDAIYQATVVGKPPLE.DKIMGLGTERIFLPLLQTSVPDLIDYNM
D71864     IEPYPVLEVKTISYKKDSIYLATVVGKPPLE.DKYMGYLTERLFLPLLQTHAPNLIDYYM
BAB65007   QDYYPVFKLERTYSRDNPIFHATSVGKPPLE.DAWIGKAVERLFLPFIRILIPEIVDMNL
AAK41165   ADYYPVFKLERVYIREDPIFHVTSVGKPPLE.DAWIGKAVERIFLPFAKMLVPELIDMNL
AAK89045   VEAFPVMTLSAITMRRDPIYLSTYTGRPPDE.PSVLGEAMLEIFLPLVKRQFAEIVDLWM
NP_541935  VEAFPVMQVTAITMRKKPVYLSTYTERPPDE.PSRLGEVMNQLFVPVVRKQFPEIADLWL
H81051     QDYFPVFTVERITMRENPIYHSTYTGKPPDE.PAVLGVALNEVFVPLLQKQFPEITDFYL
F81823     QDHFPVFTVERITMRENPIYHSTYTGKPPDE.PAVLGVALNEVFVPLLQKQFPEITDFYL
NP_457773  VDNFPVFTVTHITQREDAIYHSTYTGRPPDE.PAVLGVALNEVFVPILQKQFPEIVDFYL
AAC76846   VDSFPVFTVTHITQREDAIYHSTYTGRPPDE.PAVLGVALNEVFVPILQKQFPEIVDFYL
NC_407218  IDNFPVFTVTHITQRQDAIYHSTYTGRPPDE.PAVMGVALNEVFVPILQKQFPEIVDFYL
NP_245155  QEYFPVFTVTHITMRKDPIYHSTYTGRPPDE.PAVLGEALNEVFIPILQKQFPEIVDFYL
G82338     VERHHVFTVTHVTMRNKPIYHSTYTGRPPDE.PAVLGVALNEVFVPILQKQFPEIADFYL
NP_518829  QDWFPVFTVERITMRRDPIYHSTYTGKPPDE.PAVLGVALNEVFVPLLQKQFPEIADFYL
NP_360908  VEEFPVFTVTAITMKKNPVYLSTYTGEPPDE.PSILGEALNEIFIPILQQQFPEIVDFWL
F71643     VEEFPIFTVTAITMKKKPIYLSTYTGKPPDE.PAILGEALNEIFIPILHQQFPEIVDFWF
AAK22300   VEKFPVFQVTAITMRKDPIYLTTFTGRPPDE.PSVLGEALNEVFIPLIRQQFPEIVDFWL
S74726     VEDSPLVRFQCLTHRKNPVYLTTFSGRPPKE.EAMMAIALNRIYTPILRQQVSEITDFFL
NP_110927  EEEFPVFHVKNIIERNDRIYPTTIVGKLWHE.DVVLGKAVERMFLPLIQMVLPEVVDINT
CAC12324   EEQFPVFHIKKIIERRDRIYPTTIVGKLWHE.DVIMGKTIERMFLPLIQMVMPEVVDINT
D70439     VDKYPQMHVTAIVMRKDPIYLTTIVGRPPQE.DKYLGWATERIFLPLIKFNLPEVVDYHL
A69276     PEPYPVFHITHITHRENPIYHATVVGKPPME.DAWLGKATERIFLPILRMMHPEIVDINL
CAB92111   QEPFPALKIDCVTMRKRPLLQSIVVGRPPTE.DGPLGRATERFFLPLLKIIVPDIVDYHL
A75533     ADLYPLFHVTCVTMRQNPVYPATIVGRPPME.DAYLIEASERLFLPAAQLIVPEIVDYHM
AAG57845   GRNMTVVRIDKVSYRTRPIFESLYLGMPWTEIDYLMGPATCVPLYQQLKAEFP..EVQAV
AAL21802   GRNMTVVRIDKVSYHSKPIFESLYLGMPWTEIDYLMGPATCVPLYQQLKAEFP..EVQAV
AAD28782   GRRMPVIRVERVSYRHEPVFESLYLGMPWNECDYLVGPNTCVPLLKQLRAEFP..EVQAV
CAB12158   GRSMPIIKIKRVYHRNNPIFEHLYLGMPWTECDYMIGINTCVPLYQQLKEAYPN.EIVAV
S69350     ARLQCEVKIDRITHRTNPIFENLYLGIPWTEIDYLMALNTSVPLYKQLKETMP..EVVAV
T31294     VRKAPIFKVTAVSHRRDPIFENIYIGRGWTEHDTLIGLHTSAPIYAQLRQSFP..EVTAV
A72513     VRRQPVVRLEAAYI.KEGEYTHTIMGGS.LEHVNLMGFPREASIWEAVRRALPRVRAVRL
E69052     VRDEPVISLERMHIRKD.AMYHAILPAG.FEHRLLQGLPQEPRIYRAVKNTVPTVRNVVL
S28657     IRDQPIINLSKMHIKKDNPHYHGILSAG.FEHKLLQGLPQEPRIFKSVKNAVPTVENVVL
D64441     VRKQPIIKIEKLY.RKEKPIFHALLPGG.IEHKTLMGMPQEPRILKGVRNTVPTVKNIVL
E75130     VRKQPVVVFEEMYH.VDDPIFHALLPGG.YEHYMLMGLPKEPQIYASVKKVVPKVHGVRL
F71087     VRKQPIVIFEEMYH.VDDPIFHALLPGG.YEHYMLMGLPKEPQIYASVKKVVPKVHGVRL
B69455     VRDEPVIVFDEMYV.KEDYIYYSITPAG.KEHQMLMGVPYEPVIYRFVSNVC.KVKNVIT
T51313     VEPRPLVRVKRIH.RRRAPIFHTLL.SG.AEVFNSVGLLGEANVLALLRVQVPGVEDVYF
D95411     VGPNAVFEVLGVTVRKDAIFHSILCGSP.EEVLTL.ELSVSANIYQRLSAALPGI..VNV
CAC12691   SNVSPVIEISAVTHR.DNFIYPGLCPWS.PEVDALLSLAAGAELLGQLQGLIDGVVDLEM
```


S62018 PYEAQALWLILKVDLKGLQALKTTPEEFCKKVGDIYFRTKVGFIVHEIILV....ADDID
PA0254 SYEAATCWAVLSIDVQRLAALGTDAAAFAARVAETVFGSHAGHLVPKLILV....GNDID
E71557 YGEAGFHALAAAIVKERYW........KEALRSALRILGEGQLSLTKFLWIT...DQSVD
F81711 YGEAGFHAVAAAVVKERYW........KEALRSALRILGEGQLSLTKFLWIT...DQSVD
A72092 YGESGFHALTAAVVKERYW........RESLTTALRILGEGQLSLTKFLMVT...DQEVP
O34023 YGEAGFHALTGAVVKERYW........KESLATSLRILGEGQLSLTKFLMIT...DHHVD
BH3930 VDAAGVHPLLFAIGSERYTPYQKVKQPAELLTIANRILGTGQLSLAKYLFITAEQDKPLD
B81401 PENGVFHNLILAKIDAKY.....PAHAQQIMHAFWG.V..GQMSFVKHAIFV....DK~~~
D71864 PENGVFHNLILAKIHTRY.....NAHAKQVMHAFWG.V..GQMSFVKHAIFV....NEDAP
BAB65007 PEFGLFTGIGIFSIKKHY.....PGQAKKTMMSIWG.L..GQLSLLKMVIIV....DADVN
AAK41165 PEYGLFTGIGIFSIKKYY.....PGQAKRVMMALWG.T..GQLSLLKIIIVV....DQDID
AAK89045 PPEACSYRVMVASIDKRY.....PGQAKRVMMGLWSML..PQFSYVKLIILV....DPDID
NP_541935 PPAACSYRAMVVSIDKRY.....PGQARRVMMGLWSML..PQFSYTKLIIAV....DPDID
H81051 PPEGCSYRMAVVSMKKQY.....AGHAKRVMMGCWSFL..RQFMYTKFIIVV....DDDVN
F81823 PPEGCSYRMAVVSMKKQY.....AGHAKRVMMGCWSFL..RQFMYTKFIIVV....DDDVD
NP_457773 PPEGCSYRLAVVTMKKQY.....AGHAKRVMMGVWSFL..RQFMYTKFVIVC....DDDVN
AAC76846 PPEGCSYRLAVVTIKKQY.....AGHAKRVMMGVWSFL..RQFMYTKFVIVC....DDDVN
NC_407218 PPEGCSYRLAVVTIKKQY.....AGHAKRVMMGIWSFL..RQFMYTKFVIVC....DDDIN
NP_245155 PPEGCSYRLAVVTIKKQY.....AGHAKRVMMGVWSFL..RQFMYTKFVIVC....DDDVN
G82338 PPEGCSYRMAIVTLKKQY.....PGHAKRVMLGVWSFL..RQFMYTKFVIVC....DEQVN
NP_518829 PPEGCSYRMALVSMKKQY.....AGHAKRVMFGVWSFL..RQFMYTKFIVVV....DDDVD
NP_360908 PPEGCSYRVAVVSIKKSY.....PGHAKRIMLGIWSYL..RHFMYSKFIIVV....DDDID
F71643 PPEGCSYRVVVVSIKKSY.....PGHAKRIMLGIWSYL..RQFMYNKFIIVV....DDDID
AAK22300 PPEGCSYRIAVVSMKKAY.....PGHAKRVMLGVWSYL..RQFMYTKWVIVV....DHDIN
S74726 PMEALSYKAAIISIDKAY.....PGQAKRAALAFWSAL..PQFTYTKFVIVV....DKSIN
NP_110927 MEEAVFHNMVIVSIKKRY.....PGHAKKVMFALWG.L..GQLMFSKIIVVV....DDDIN
CAC12324 MEEAVFHNMVIVSIKKRY.....PGHAKKVMFGLWG.M..GQMMFSKIIVVV....DDDIN
D70439 PAEGCFHNFCFVSIKKRY.....PGHAFKVAYALLG.L..GLMSLEKHIVVF....DDWIN
A69276 PVEGAFHNLAIVSIKKRY.....PGQAKKVMYAIWG.T..GMLSLTKIVVVV....DDDVN
CAB92111 PEAGGFHNCAIVSIDKKY.....PKHAQKVMHAIWG.A..HMMSLTKLIVVV....DSDCD
A75533 PPAGVAHNLVVVSIKKDF.....PGQAYKVANGLLG.L..GQMMFAKVIVVV....DADVK
AAG57845 NAMYTHGLLAIISTKKRYG.....GFARAVGLRAMT..TPHGLGYVKMVIMV....DEDVD
AAL21802 NAMYTHGLLAIISTKKRYG.....GFARAVGLRAMT..TPHGLGYVKMVIMV....DEDVD
AAD28782 NAMYTHGLMVIISTAKRYG.....GFAKAVGMRAMT..TPHGLGYVAQVILV....DEDVD
CAB12158 NAMYTHGLIAIVSTKTRYG.....GFAKAVGMRALT..TPHGLGYCKMVIVV....DEDVD
S69350 NAMYTHGIGVIISTKVRYG.....GYAKGVAFRLLS..TPHGMPYSKIVIVV....DEFVD
T31294 NALYQHGLTGIISVKNRMA.....GFAKTVALRALS..TPHGVMYLKNLIMV....DADVD
A72513 TPASGGWLHAVIAVEKQHE.....GDGKTAIMAAFA..AHPS...LKHVVVV....DSDVD
E69052 TEGGCCWLHAAVSIKKQTE.....GDGKNVIMAALA..AHPS...LKHVVVV....DEDID
S28657 TEGGCCWLHAAISINKQTE.....GDGKNAIMAALS..AHPS...LKHAVVV....DTDVD
D64441 TEGGCCWLHAVVQIEKRTE.....GDGKNAILAAFA..SHPS...LKHVIVV....DDDIN
E75130 TEGGCMWLHAVVSITKQHE.....GDGKNAILAAFA..GHPS...LKRVVVV....DEDVN
F71087 TEGGCMWLHAVVSITKQHE.....GDGKNAILAAFA..GHPS...LKRVVVV....DEDVN
B69455 TPGSCHYFHCVVQIEKKSE.....GDGKNAIIAALA..ANPS...MKGVVVV....DDDID
T51313 SHGGCGFYHCVVKIAQKRA.....GWAKQAILATFA..AFPP...LKMVTVV....DEDVD
D95411 TCQPFV.NHAVVQIEPQFE.....GHARQVMLATIG..AEPI..WAKQITVI....DTDVD
CAC12691 A.GGTSGFSVVVAVHRTTA.....ADVRRLVMLALN..LD.R..RLKTITVV....DDDVD



The labels for each line correspond to the accession number for that protein. S62018 is from *Saccharomyces cerevisiae*, PA0254 is from *Pseudomonas aeruginosa*, E71557 is from *Chlamydia trachomatis*, F81711 is from *Chlamydia muridarum*, A72092 is from *Chlamydia pneumoniae*, O34023 is from *Chlamydia psittaci*, BH3930 is from *Bacillus halodurans*, B81401 is from *Campylobacter jejuni*, D71864 is from *Helicobacter pylori*, BAB65007 is from *Sulfolobus tokodaii*, AAK41165 is from *Sulfolobus sulfataricus*, AAK89045 is from *Agrobacterium tumefaciens*, NP_541935 is from *Brucella melitensis*, H81051 is from *Neisseria meningitidis*, F81823 is from *Neisseria meningitidis*, NP_457773 is from *Salmonella enterica*, AAC76846 is UbiD from *Escherichia coli* K-12, NC_407218 is from *Yersinia pestis*, NP_245155 is from *Pasteurella multocida*,

G82338 is from *Vibrio cholerae*, NP_518829 is from *Ralstonia solanacearum*, NP_360908 is from *Rickettsia conorii*, F71643 is from *Rickettsia prowazekii*, AAK22300 is from *Caulobacter crescentus*, S74726 is from *Synechococcus* sp. PCC6803, NP_110927 is from *Thermoplasma volcanium*, CAC12324 is from *Thermoplasma acidophilum*, D70439 is from *Aquifex aeolicus*, A69276 is from *Archaeoglobus fulgidus*, CAB92111 is from *Streptomyces coelicolor*, A75533 is from *Deinococcus radiodurans*, AAG57845 is EdcC from *Escherichia coli* O157:H7, AAL21802 is from *Salmonella typhimurium*, AAD28782 is VdcC from *Streptomyces* sp.D7, CAB12158 is BsdC from *Bacillus subtilis*, S69350 is ShdC from *Sedimentibacter hydroxybenzoicus*, T31294 is from *Sphingomonas aromaticivorans* plasmid pNL1, A72513 is from *Aeropyrum pernix*, E69052 is from *Methanobacter thermoautotrophicus*, S28657 is from *Methanobrevibacter smithii*, D64441 is from *Methanococcus jannaschii*, E75130 is from *Pyrococcus abyssi*, F71087 is from *Pyrococus horikoshii*, B69455 is from *Archaeoglobus fulgidus*, T51313 is from *Rhodospirillum rubrum*, D95411 is from *Sinorhizobium meliloti* megaplasmid pSymA, adn CAC12691 is from *Thauera aromatica*. Further information on each homologue is available in Table 2.5.

```
EcdB ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~MKLIVG
SenB ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~MRLIVG
VdcB ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~MRLVVG
BsdB ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~MKAEFKRKGGGKVKLVVG
UbiX ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~MKRLIVG
Pad1 MLLFPRRTNIAFFKTTGIFANFPLLGRTITTSPSFLTHKLSKEVTRASTSPPRPKRIVVA

EcdB MTGATGAXLGVALLQALREMPNVETHLVMSKWAKTTIELETPYSARDVAALADFSHNPAD
SenB MTGATGAPLGVELLQALRAIPDVETHLVMSKWAKTTIELETPYTPAEVAALADYCHSPAD
VdcB MTGATGAPFGVRLLENLRQLPGVETHLVLSRWARTTIEMETGLSVAEVSALADVTHHPED
BsdB MTGATGAIFGVRLLQWLKA.AGVETHLVVSPWANVTIKHETGYTLQEVEQLATYTYSHKD
UbiX ISGASGAIYGVRLLQVLRDVTDIETHLVMSQAARQTLSLETDFSLREVQALADVTHDARD
Pad1 ITGATGVALGIRLLQLLKEL.SVETHLVISKWGAATMKYETDWEPHDVAALATKTYSVRD

EcdB QAATISSGSFRTDGMIVIPCSMKTLAGIRAGYADGLVGRAADVVLKEGRKLVLVPREMPL
SenB QAATISSGSFRTDGMIIIPCSMKTLAGVRAGYAEGLVGRAADVVLKEGRKLVLVPREMPL
VdcB QGATISSGSFRTDGMVIVPCSMKTLAGIRTGYAEGLVARAADVVLKERRRLVLVPRETPL
BsdB QAAAISSGSFDTDGMIVAPCSMKSLASIRTGMADNLLTRAADVMLKERKKLVLLTRETPL
UbiX IAASISSGSFQTLGMVILPCSIKTLSGIVHSYTDGLLTRAADVVLKERRPLVLCVRETPL
Pad1 VSACISSGSFQHDGMIVVPCSMKSLAAIRIGFTEDLITRAADVSIKENRKLLLVTRETPL

EcdB STIHLENMLALSRMGVAMVPPMPAFYNHPETVDDIVHHVVARVLDQFGLEHPYA..RRWQ
SenB STIHLENMLALSRMGVAIVPPMPAFYNLPQTVDDIIQHIVARVLDQFGLEHTRT..RRWQ
VdcB SEIHLQNMLELARMGVQLVPPMPAFYNNPQTVDDIVDHVVARILDQFDLPAPAA..RRWA
BsdB NQIHLENMLALTKMGTIILPPMPAFYNRPRSLEEMVDHIVFRTLDQFGIRLPEA..KRWN
UbiX HLGHLRLMTQAAEIGAVIMPPVPAFYHRPQSLDDVINQTVNRVLDQFAITLPEDLFARWQ
Pad1 SSIHLENMLSLCRAGVIIFPPVPAFYTRPKSLHDLLEQSVGRILDCFGIH..ADTFPRWE

EcdB GLPQARNFSQENE~~~
SenB GLRQTANFSQENG~~~
VdcB GMRAARAAARSFGDAA
BsdB GIEKQKGGA*~~~~~~
UbiX GA~~~~~~~~~~~~~~
Pad1 GIKSK~~~~~~~~~~~
```

Figure 2.10- Comparing proteins homologous to VdcB

EcdB is from *Escherichia coli* O157:H7, SenB is from *Salmonella enterica* subspecies enterica serovar Typhi, VdcB is from *Streptomyces* sp.D7, BsdB is from *Bacillus subtilis*, UbiX is annotated as a 3-octaprenyl-4-hydroxybenzoate decarboxylase invovled in ubiquinone synthesis in *Escherichia coli* K-12, and Pad1 is a phenolic acid decarboxylase from *Saccharomyces cerevisiae*.

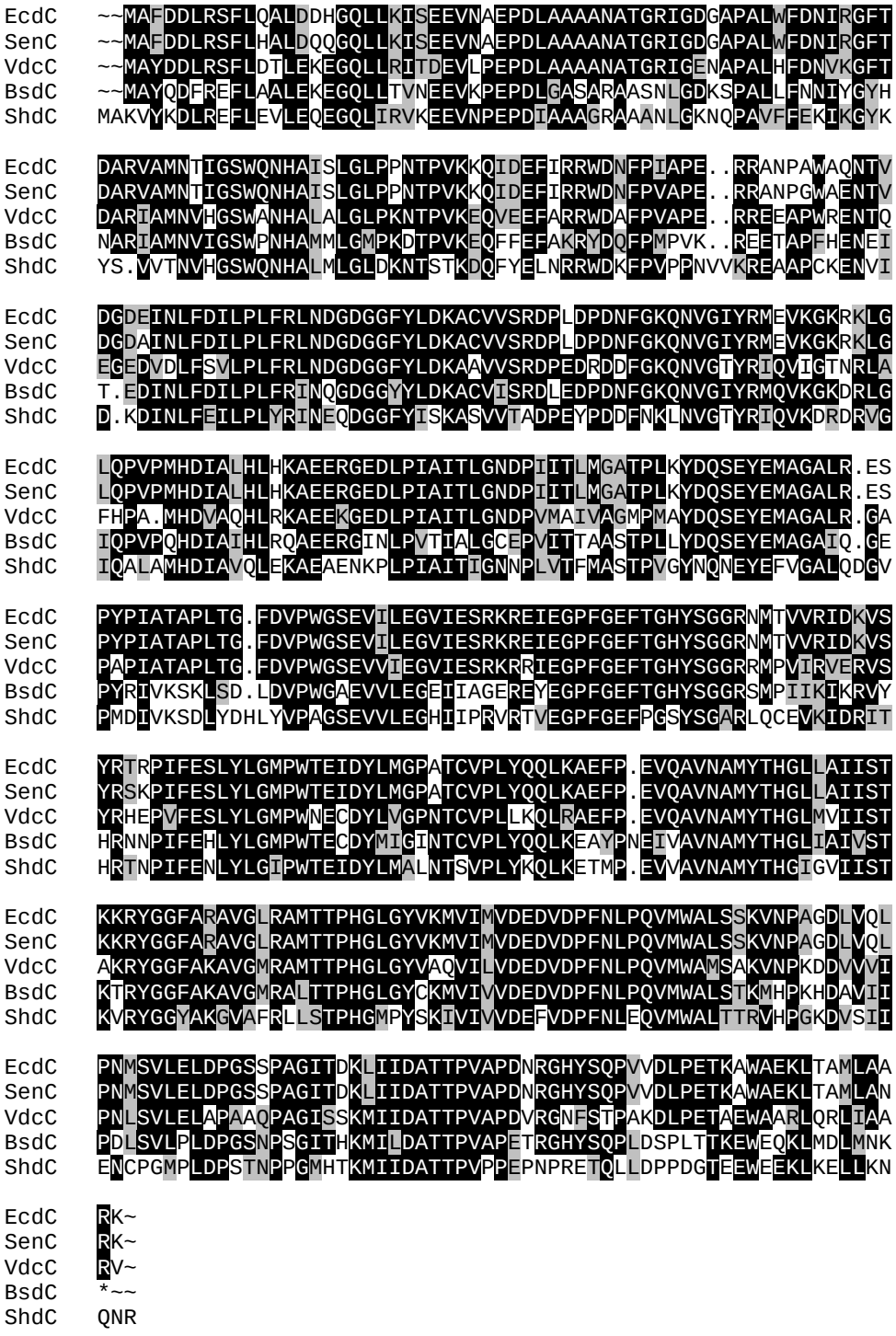


Figure 2.11- Comparing the ShdC homologues from *E. coli* O157:H7, *Salmonella enterica*, *Streptomyces* sp.D7, *B. subtilis*, and *S. hydroxybenzoicus*
EcdC is from *Escherichia coli* O157:H7, SenC is from *Salmonella enterica* subspecies enterica serovar Typhi, VdcC is from *Streptomyces* sp.D7, BsdC is from *Bacillus subtilis*, and ShdC is from *Sedimentibacter hydroxybenzoicus* JW/Z-1.

Literature cited

1. **Ausubel, F. M., R. Brent, R. E. Kingston, D. D. Moore, J. G. Seidman, J. A. Smith, and K. Struhl.** 1995. Current Protocols in Molecular Biology. John Wiley & Sons, USA.

2. **Chamkha, M., E. Record, J. L. Garcia, M. Asther, and M. Labat.** 2002. Isolation from a shea cake digester of a tannin-tolerant *Escherichia coli* strain decarboxylating *p*-hydroxybenzoic and vanillic acids. Curr. Microbiol. **44:**341-349.

3. **Chow, K. T., M. K. Pope, and J. Davies.** 1999. Characterization of a vanillic acid non-oxidative decarboxylation gene cluster from *Streptomyces* sp. D7. Microbiology **145**(Pt 9)**:**2393-403.

4. **Clausen, M., C. J. Lamb, R. Megnet, and P. W. Doerner.** 1994. PAD1 encodes phenylacrylic acid decarboxylase which confers resistance to cinnamic acid in *Saccharomyces cerevisiae*. Gene **142**(1)**:**107-12.

5. **Hayashi, T., K. Makino, M. Ohnishi, K. Kurokawa, K. Ishii, K. Yokoyama, C. G. Han, E. Ohtsubo, K. Nakayama, T. Murata, M. Tanaka, T. Tobe, T. Iida, H. Takami, T. Honda, C. Sasakawa, N. Ogasawara, T. Yasunaga, S. Kuhara, T. Shiba, M. Hattori, and H. Shinagawa.** 2001. Complete genome sequence of enterohemorrhagic *Escherichia coli* O157:H7 and genomic comparison with a laboratory strain K-12. DNA Res. **8**(1)**:**11-22.

6. **He, Z., and J. Wiegel.** 1995. Purification and characterization of an oxygen-sensitive reversible 4- hydroxybenzoate decarboxylase from *Clostridium hydroxybenzoicum*. Eur. J. Biochem. **229**(1)**:**77-82.

7. **Herbelin, C. J., S. C. Chirillo, K. A. Melnick, and T. S. Whittam.** 2000. Gene conservation and loss in the *mutS-rpoS* genomic region of pathogenic *Escherichia coli*. J. Bacteriol. **182**(19)**:**5381-90.

8. **Huang, J., Z. He, and J. Wiegel.** 1999. Cloning, characterization, and expression of a novel gene encoding a reversible 4-hydroxybenzoate decarboxylase from *Clostridium hydroxybenzoicum*. J. Bacteriol. **181**(16)**:**5119-22.

9. **Leppik, R. A., I. G. Young, and F. Gibson.** 1976. Membrane-associated reactions in ubiquinone biosynthesis in *Escherichia coli*. 3-Octaprenyl-4-hydroxybenzoate carboxy-lyase. Biochim. Biophys. Acta. **436**(4)**:**800-10.

10. **Ohnishi, M., T. Murata, K. Nakayama, S. Kuhara, M. Hattori, K. Kurokawa, T. Yasunaga, K. Yokoyama, K. Makino, H. Shinagawa, and T. Hayashi.** 2000. Comparative analysis of the whole set of rRNA operons between an enterohemorrhagic *Escherichia coli* O157:H7 Sakai strain and an *Escherichia coli* K-12 strain MG1655. Syst. Appl.Microbiol. **23**(3)**:**315-24.

11. **Zeng, H., I. Snavely, P. Zamorano, and G. T. Javor.** 1998. Low ubiquinone content in *Escherichia coli* causes thiol hypersensitivity. J. Bacteriol. **180**(14)**:**3681-3685.

12. **Zhang, H., and G. T. Javor.** 2000. Identification of the *ubiD* gene on the *Escherichia coli* chromosome. J. Bacteriol. **182**(21)**:**6243-6.

13. **Zhang, X., L. Mandelco, and J. Wiegel.** 1994. *Clostridium hydroxybenzoicum* sp.nov., an amino acid-utilizing, hydroxybenzoate-decarboxylating bacterium isolated from methanogenic freshwater pond sediment. Int. J. Syst. Bacteriol. **44**(2)**:**213-222.

CHAPTER 3

PARTIAL PURIFICATION OF A 4-HYDROXYBENZOATE DECARBOXYLASE FROM *SEDIMENTIBACTER HYDROXYBENZOICUS* JW/Z-1 AND A VANILLATE DECARBOXYLASE FROM *BACILLUS SUBTILIS*

Introduction

A new enzyme family of hydroxyarylic acid decarboxylases was proposed based on the hydroxyarylic acid decarboxylating activities of homologues of 4-hydroxybenzoate decarboxylase from *Sedimentibacter hydroxybenzoicus* JW/Z-1 (ShdCD) from *Streptomyces* sp.D7 (VdcCD), *Bacillus subtilis* (BsdCD), and *Escherichia coli* O157:H7 (EcdCD). Based on their abilities to decarboxylate both 4-hydroxybenzoate and vanillate, the enzymes from *S. hydroxybenzoicus* and *E. coli* O157:H7 were designated 4-hydroxybenzoate decarboxylases. Based on their abilities to decarboxylate vanillate but not 4-hydroxybenzoate, the enzymes from *Streptomyces* sp.D7 and *B. subtilis* were both designated vanillate decarboxylases. The parameters defining the family include high amino acid sequence homology, decarboxylation of a hydroxyarylic acid, and possible oxygen sensitivity. ShdCD appears not to require any cofactors. None of the other enzymes have been tested for cofactors. Previous attempts to purify ShdCD from *S. hydroxybenzoicus* resulted in low yields (He and Wiegel, 1995). Purification of a recombinant protein might be more successful. In an effort to further characterize these enzymes, two members of the family were chosen for purification. The ShdCD from *S. hydroxybenzoicus* and the BsdCD from *B. subtilis* were chosen since the most information was available about those two. The characteristics of ShdCD were discussed in the previous two chapters. BsdCD is an oxygen-sensitive enzyme from an aerobic organism and would provide a good comparison for ShdCD.

This chapter describes the partial purification of BsdCD and ShdCD which was accomplished by attaching hexahistidine tags to the N-terminal of the BsdC or ShdC

protein. One could then use immobilized metal affinity chromatography (IMAC) to separate the proteins from the crude cell extract.

Methods

Strains utilized and growth conditions

Escherichia coli DH5α was the strain used for cloning and protein overexpression. These cultures were grown anaerobically in Luria Bertani medium (LB) + 0.2% KNO_3 + 0.3% glucose. When necessary, 100 μg/mL of ampicillin was added.

Construction of expression plasmid

The proteins were expressed in *E. coli* DH5α using the plasmid pTrcHisC (Invitrogen, Carlsbad, CA). The primers to amplify *shdCD* are listed in Table 3.1. The PCR was performed running the following mixture on a Minicycler (MJ Research, Waltham, MA): 5 μL 10x buffer (Stratagene, La Jolla, CA), 2 μL 10 mM dNTP, 1 μL 25 mM forward primer, 1 μL 25 mM reverse primer, 1 μL genomic DNA, 39 μL sterile water, and 1 μL *Pfu* polymerase (Stratagene). The plasmid pTrcHisC was cut with the restriction enzymes corresponding to the fragment to be inserted. The PCR product was cut with its appropriate restriction enzymes (Promega, Madison, WI) (Table 3.1), and the resulting fragment and the cut plasmid were ligated with T4 DNA ligase (Promega). The construct was introduced into competent *E. coli* DH5α by electroporation, and the transformants were selected on LB + 100 μg/mL ampicillin. A number of colonies were picked, grown in LB, and the plasmids isolated using QiaSpin Miniprep kit (Qiagen, Valencia, CA). The plasmids were cut with *Nhe*I (Promega), and run on a 0.8% agarose gel. The inserts

in these plasmids were sequenced, and the sequences matched the expected sequences for *shdCD* and *bsdCD*. The resulting plasmid attaches a nucleotide sequence encoding a hexahistidine tag onto the N-terminal of the gene for BsdC or ShdC.

Protein purification

The protein purification protocol follows that of the TALON kit (Clontech, Palo Alto, CA). *E. coli* DH5α containing either pTrc*shdCD* or pTrc*bsdCD* were grown anaerobically, and the cells were harvested at mid-log phase. Induction kinetics were determined in order to choose the optimal time to harvest the cells. Two 20 mL anaerobic LB + ampicillin tubes were inoculated and incubated overnight to grow to an OD_{600}=0.3. Two mL were removed and centrifuged aerobically in a microfuge tube at 10,000 x g for 1 minute. The resulting cell pellet was frozen, and the remaining 18 mL were induced with 1 mM IPTG. Every hour for 7 hours, 2 mL of cells were harvested by centrifugation and frozen. All pellets were resuspended in 100 μL water and 100 μL 2x SDS loading buffer (Ausubel, *et al.*, 1995), and then boiled for 3 minutes. The lysed cells were centrifuged to pellet cell debris, and the resulting supernatant was analyzed by SDS-PAGE, using the method described in Chapter 2 of this thesis. Judging by protein production, the optimum induction time was determined.

For protein expression, 10 mL of aerobic LB + ampicillin were inoculated with one colony of the appropriate strain and grown overnight at 37°C. This tube was used to inoculate a 50 mL bottle with 100 μL of aerobic culture that was grown overnight at 37°C. A 1 L anaerobic bottle was inoculated with 50 mL of an anaerobic culture, and this culture was grown to mid-exponential phase. The culture was induced with 1 mM

IPTG if necessary, although induction seemed to make no difference in protein production.

The cells were collected anaerobically by centrifugation at 5,000 x g for 20 minutes at 4°C, and then resuspended in 20 mL French Press buffer (50 mM sodium phosphate, 300 mM NaCl, 10 mM imidazole, 0.5 mM PMSF, pH 7.0) Cells were lysed with a French press as described in Chapter 2. If the resulting lysate was viscous due to DNA, it was diluted 5-fold with anaerobic wash buffer (50 mM sodium phosphate, 300 mM NaCl). A plastic disposable column (Clontech) was set up in the anaerobic chamber. Two mL of a cobalt-containing resin suspension, which had been previously sparged with N_2 for 10 minutes, was transferred into the anaerobic chamber. The liquid was removed with a syringe equipped with a 23 gauge needle. The resin was equilibrated briefly with 20 mL anaerobic 1x extraction/wash buffer, and the liquid was removed by syringe. This equilibration was repeated once. Crude cell extracts were added to the resin and mixed gently for 20 minutes at room temperature in the anaerobic chamber. The liquid was removed, and the resin was then gently washed with 20 mL anaerobic 1x extraction/wash buffer for 10 minutes at room temperature. After removing the wash buffer, the resin was resuspended in 2 mL of anaerobic 1x extraction/wash buffer and transfered to a plastic column. The resin was allowed to settle in the column, and the buffer was drained until the level was just above the resin. Care was taken to ensure no air bubbles were trapped. The resin was washed once with 10 mL anaerobic 1x extraction/wash buffer, and then with 10 mL anaerobic 1x extraction/wash buffer + 25 mM imidazole. The protein was eluted with 10 mL elution buffer (50 mM sodium

phosphate, 300 mM NaCl, 250 mM imidazole). The eluate was collected in 500 µL fractions in anaerobic collection tubes which were frozen immediately.

A more stringent series of washing entailed pouring the resin directly onto the column. The column was rinsed twice with 20 mL wash buffer. The cell lysate was diluted into 100 mL of wash buffer and loaded onto the column. The column was washed with 60 mL wash buffer, followed by 40 mL of wash buffer with 25 mM imidazole. The protein was eluted with 10 mL elution buffer and collected in 2 mL fractions.

Fraction analysis and enzyme assays

The fractions were analyzed by SDS-PAGE to check for protein and identify which fractions had the most protein. A native PAGE was run to check the pertinent fractions for proper folding. Activity was ascertained by enzyme assays, which were described in Chapter 2.

Results and discussion

Protein purification

The first purification attempt, using the less stringent washes, yielded partially purified ShdCD and BsdCD. The protein eluted with the first 1 mL of elution buffer (Figure 3.1.B). Only the C subunit of each protein was tagged with the hexahistidine sequence, so it is the only subunit expected to be purified. However, SDS-PAGE analysis of the purified protein samples showed multiple bands, one of which is a 7-9 kDa band that may be the D subunit. This suggests that the two proteins copurified and

therefore interacted in their native state. A native gel showed only a single band, suggesting that the major proteins in the preparation associated with each other. The enzyme assays also show that the purified fractions containing significant amounts of protein have decarboxylating activity. This purification of ShdCD had several contaminating bands, as seen in Figure 3.1.B. A more stringent washing protocol, as detailed in the methods section, was used in an attempt to remove the contaminating proteins. Figure 3.1.A is the silver stained SDS-PAGE of the first 2 mL of that purification attempt. Only the 57 kDa protein is visible, while the 7 kDa protein is probably too faint to distinguish. The enzyme assays show only slight activity. However, the contaminating proteins remain. SDS-PAGE analysis of the negative control, being a purification of *E. coli* JM109 containing only the pTrcHis vector, did not reveal any protein bands. It may be that the contaminating protein is somehow involved in the decarboxylation complex. The molecular weight of the protein corresponds to that of a VdcB homologue which is in *E. coli* K-12-derived strains called UbiX. This protein was discussed in Chapter 2.

Further research is needed to determine whether only the two subunits C and D are copurifying or if there is an additional subunit. Once this information is obtained, the proteins can be overexpressed, purified, and studied *in vitro*. This will yield information that could help further characterize the proposed novel enzyme family. Additionally, the enzyme mechanism could be elucidated. The purified enzymes will be used for protein crystallization such that the structure of the protein can be revealed.

Table 3.1- PCR primers for cloning *shdCD* and *bsdCD* into pTRC

Primer Name	Sequence	Restriction site	Target Sequence
4ohCHis For	GAC TCG CTA GCG CTA AAG TAT ACA AAG A	*Nhe*I	*shdCD*
4ohDHis Rev	GGT CTA AGC TTT TCT ATT TTT TCA AGG G	*Hin*dIII	*shdCD*
yclCHisFor	GTC TCG CTA GCG CTT ATC AAG ATT TCA G	*Nhe*I	*bsdCD*
yclDHisRev	TAT ATA GAT CTT CAA GCC TTT CGT TCC	*Bgl*II	*bsdCD*

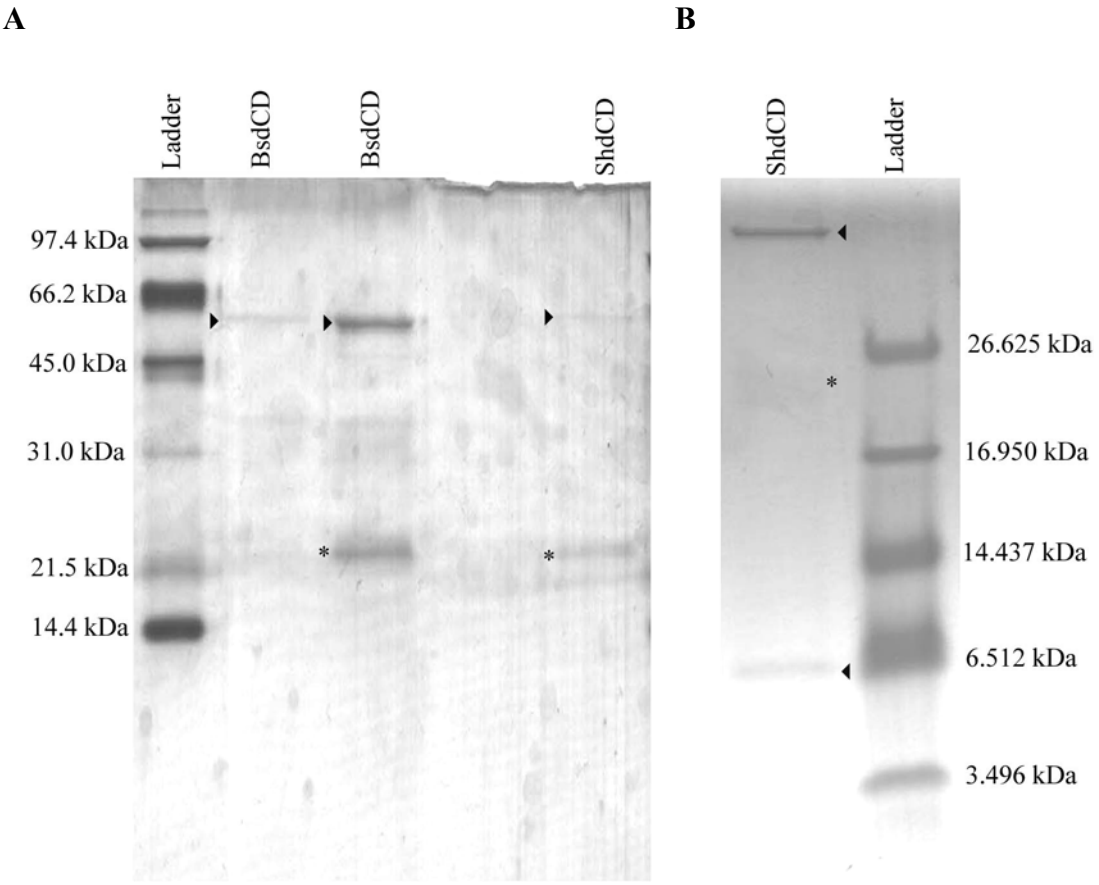


Figure 3.1 – SDS-PAGE of partially purified ShdCD and BsdCD

In Figure 3.1.A, the arrows indicate the visible C subunits. Unlike in other purifications, the D subunits are not visible in this purification, although the protein fraction is enzymatically active. Another contaminating band, indicated by the asterisk, is visible even after a stringent purification. Figure 3.1.B is a less stringent, partial purification of ShdCD run on a 16.5% Tris-Tricine gel and stained with Coomassie Blue. The C and D subunits are both visible, as indicated by the arrows. The C subunit is around 57 kDa and the D around 7 kDa. A hazy band around 20kDa is visible, corresponding to the contaminating protein found in the other purifications.

Literature cited

1. **Ausubel, F. M., R. Brent, R. E. Kingston, D. D. Moore, J. G. Seidman, J. A. Smith, and K. Struhl.** 1995. Current Protocols in Molecular Biology. John Wiley & Sons, USA.

2. **He, Z., and J. Wiegel.** 1995. Purification and characterization of an oxygen-sensitive reversible 4- hydroxybenzoate decarboxylase from *Clostridium hydroxybenzoicum*. Eur. J. Biochem. **229**(1)**:**77-82.

APPENDIX

ATTEMPTS TO OBTAIN THE NUCLEOTIDE SEQUENCE ENCODING THE 3,4-DIHYDROXYBENZOATE DECARBOXYLASE FROM *SEDIMENTIBACTER HYDROXYBENZOICUS* JW/Z-1

Introduction

Sedimentibacter hydroxybenzoicus JW/Z-1 (formerly *Clostridium hydroxybenzoicum* JW/Z-1) carries out the decarboxylation of both 4-hydroxybenzoate and 3,4-dihydroxybenzoate using a 4-hydroxybenzoate decarboxylase (ShdC) and a separate 3,4-dihydroxybenzoate decarboxylase (Shd34) (Zhang and Wiegel, 1990). While the 4-hydroxybenzoate decarboxylase can decarboxylate 3,4-dihydroxybenzoate at a lower rate, the two enzymes are distinct, based on their distinct substrate spectra, their additive activities, and that two separate proteins were isolated with two distinct N-terminal amino acid sequences (He and Wiegel, 1995, He and Wiegel, 1996). The characteristics of Shd34 include being oxygen-sensitive, the ability to catalyze the forward and reverse reactions, and the apparent lack of cofactors. Shd34 was purified through a series of chromatographic steps which were performed anaerobically at 4°C (He and Wiegel, 1996). Decarboxylase activity was assayed in the various fractions by monitoring the anaerobic conversion of 3,4-dihydroxybenzoate to catechol. The substrates and products were separated and identified using HPLC equipped with a UV detector. On an SDS-PAGE gel, the purified protein yielded only a single 57 kDa band. The native gel had a band at 270 kDa, indicating the protein might be a homopentamer or homotetramer. In retrospect, this enzyme, depending on its similarity to the 4-hydroxybenzote decarboxylase, may also have additional subunits that were not visualized on the protein gels.

The N-terminal amino acid sequence was determined for the purified ShdC cut from a SDS-PAGE gel (He and Wiegel, 1996). It is as follows: MNKVT DLRSA

IELLK TIPGQ LIETN XDV. A comparison of this N-terminal region to that of ShdC is shown in Figure A.1.

Attempts were made to clone the gene encoding Shd34 from *S. hydroxybenzoicum*. A number of methods were employed: PCR of the gene using primers designed from the N-terminal amino acid sequence with reverse primers designed from consensus sequences of ShdC with other genes, PCR of the gene using primers designed from the N-terminal and the obtained internal amino acid sequences, Southern blotting using the N-terminal sequence or the gene for ShdCD, and differential display. None were successful in obtaining a sequence.

Methods

Amplifying the gene for 3,4-dihydroxybenzoate decarboxylase by polymerase chain reaction

All primers used are listed in Table A.1. The sequence of the primers were derived from either N-terminal or internal amino acid sequences of Shd34 or consensus sequences of ShdC with similar proteins. These consensus regions are shown in Figure A.2. As the primers were designed by translating amino acid sequences into nucleic acid sequence, they were made either using the codon preference exhibited by the 4-hydroxybenzoate decarboxylase gene or were made degenerate.

PCR was carried out in a total volume of 50 µL containing 0.3 µg of template DNA, 0.5µM each of forward and reverse primers, 5 µL of a 2.5 µM dNTP mix, and 1 µL polymerase (either *Taq* or *Tfl* from Promega, Madison, WI). The reaction mixtures were sealed with mineral oil, and reactions were done in a Perkin Elmer DNA Thermal

Cycler 480 (Perkin Elmer, Foster City, CA). The cycle parameters varied according to each reaction. They mainly consisted of 40 cycles of 1 minute denaturing at 92°C, 1 minute annealing at 36-61°C, and an extension at 74°C for 3 minutes. PCR products were run on 0.8% agarose gels with λ DNA (Promega) digested with *Hin*dIII (Promega) as molecular weight markers. Bands were visualized after staining with ethidium bromide on a UV transilluminator. The genomic DNA for *S. hydroxybenzoicum* was either supplied by Jianbin Huang or isolated using a HighPure PCR Template Preparation kit (Boehringer-Mannheim, Indianapolis, IN).

Southern Blotting

Genomic DNA that was cut with various restriction enzymes, including *Hin*dIII, *Kpn*I, *Eco*RI, *Pst*I, *Sac*I, *Bam*HI, *Nde*I, and *Sap*I (Promega), was transferred onto Nitrobind nitrocellulose transfer membrane (MSI, Westboro, MA) by the method described by Ausubel *et al.* (Ausubel, *et al.*, 1995). The DNA was fixed by UV irradiation using the appropriate automatic setting on a GS Gene Linker (BioRad, Richmond, CA). The probes made for Southern blotting are listed in Table A.2. Probes for Southern blotting, being the probe made with *shdC* and the probe made with the 34MNKV, were labelled with digoxigenin-dUTP by random primed DNA labelling using the DIG DNA labelling kit (Boehringer-Mannheim). The blots were detected using colorimetric detection with nitroblue tetrazolium (NBT) and 5-bromo-4-chloro-3-indolyl phosphate (BCIP) as included in the DIG Nucleic Acids Detection Kit (Boehringer-Mannheim) or chemiluminescent methods using disodium 3-(4-methoxyspiro{1,2-dioxetane-3-2'-(5'-chloro)tricyclo[3.3.1.1^{3,2}]decan}4-yl)phenylphosphate (CSPD) as included in the DIG

Luminescent Detection Kit (Boehringer-Mannheim). Hybridizations were performed at a variety of temperatures, ranging from 55-68°C depending on the stringency.

Internal amino acid sequencing

Purified Shd34 was obtained from Zhongqui He, and the solution was concentrated using Centricon-10 filters (Amicon Bioseparations, Millipore, Bedford, MA). It was subjected to SDS-PAGE followed by staining with Coomassie R-250 (BioRad). Two fragments were present around the 50 kDa weight marker. The lower one was excised and sent to the Microchemical Facility at the Winship Cancer Center in Emory University School of Medicine for internal amino acid sequencing. The band was subjected to in-gel digestion with trypsin, and the peptides were extracted and separated by HPLC. The peptides were analyzed by MALDI-TOF mass spectrometry and sequenced.

Differential display of gene expression

The method of differential display used was described in a paper by Brzostowicz *et al.* (Brzostowicz, *et al.*, 2000). The method consists of comparing the mRNA , as amplified by RT-PCR, of induced and uninduced cultures of *S. hydroxybenzoicum* JW/Z-1. The induced cultures are exposed to the substrate that causes expression of the protein of interest. The cultures were grown as described previously (Zhang, *et al.*, 1994). 500 mL of culture were grown anaerobically to an OD_{600} of about 0.2, which took approximately 2 days. The cells were divided into two sterile anaerobic flasks in an anaerobic chamber (Coy Laboratory Products Inc.), one of which contained

3,4-dihydroxybenzoate (Sigma) such that the final concentration in 250 mL was 2.5 mM. These flasks were incubated at 37°C for one hour to allow induction of the 3,4-dihydroxybenzoate decarboxylase. Induction was verified by removal and testing of a small aliquot of cells using the enzyme assay previously described (He and Wiegel, 1996). The cells were immediately placed in an acetone/ice bath. The cells were transferred aerobically to plastic centrifuge bottles which were centrifuged for 15 minutes at 5,000 x g at 4°C. After the supernatant was discarded, the cells were placed back on ice, and the pellets resuspended in residual liquid. The cells were then aliquoted into 1.5 mL microfuge tubes, placed on dry ice, and shipped. They were kept frozen for about 1-2 weeks prior to being thawed at which time the RNA was extracted from the cells.

RNA was extracted using Trizol reagent (GibcoBRL, Life Technologies, Rockville, MD). One mL of liquified cell pellet was added to 1 mL Trizol reagent. The cells were then lysed with a bead beater at 2,400 beats/min for 2 minutes with about 0.2 mL of 0.5 mm zirconia beads (Biospec Products, Bartlesville, OK). Two mL of lysate were added to 8 mL Trizol and spun at 12,000 x g for 10 minutes at 4°C. The supernatant was saved and incubated for 5 minutes at room temperature. Two mL of chloroform were added and the mixture was shaken vigorously for 15 seconds. This mixture was incubated at room temperature for 3 minutes then centrifuged for 15 minutes at 12,000 x g at 4°C. The aqueous layer (top layer) was transferred to a fresh tube to which 5 mL isopropanol was added. This mixture was incubated at room temperature for 10 minutes then centrifuged at 12,000 x g for 10 minutes at 4°C, and the resulting supernatant liquid was discarded. The pelleted nucleic acids were suspended in 10 mL 75% ethanol and centrifuged at 7,500 x g for 5 minutes at 4°C. The supernatant was

discarded and the pellet air dried. The pellet was resuspended in 500 μL water and incubated for 10 minutes at 55-60°C. The solution was treated with DNase (Stratagene, La Jolla, CA). The RNA was analyzed by agarose gel electrophoresis. RT-PCR reactions with active or heat-killed reverse transcriptase were performed to verify the absence of contaminating DNA using the RT-PCR enzyme kit from the Superscript One-Step RT-PCR System (Gibco-BRL).

The RNA was then subjected to RT-PCR using 240 primers as designed by Walters, *et al.* (Walters, *et al.*, 2001). Five microliters of each product were run on precast polyacrylamide gels (Excel gels, Amersham-Pharmacia Biotech, Piscataway, NJ), and the gels were silver stained using the Plus-One DNA silver staining kit (Amersham-Pharmacia Biotech, Piscataway, NJ). Induced and uninduced mRNA were run side by side to allow identification of differentially expressed bands as amplified by each primer.

Bands present in the induced lane but not the uninduced were excised and extracted out of the gel using the following method. Ten mL of NaCN (1 mg/mL) in pH 11.0 water were mixed with 120 μL 2 M Tris-HCl(pH8.0), 2 mL 250 mM KCl, and 30 μL 20% NP40 (detergent). Fifty μL of the above solution was added to each excised band and heated for 20 minutes at 95°C. This extracted the DNA out of the polyacrylamide such that the solution could now be used as the DNA template for the next round of PCR.

The same primers used originally to amplify any given band were used to prime PCR reactions to obtain sufficient product for cloning. PCR reactions were performed with Taq polymerase (Perkin Elmer, Foster City ,CA) and the parameters were 94°C for

1 minutes, 60°C for 1 minutes, and 72°C for 5 minutes for 40 cycles. PCR products were analyzed on an agarose gel to verify that each reaction generated a single band.

PCR products were cloned using pCR2.1-Topo (Invitrogen, San Diego, CA). Plasmids were transformed into *E.coli* Top10 One Shot chemically competent cells (Invitrogen), and plated on LB agar supplemented with 100 μg/mL ampicillin and 40 ng/mL X-Gal. Six white colonies per plate (presumably all representing one insert) were inoculated into Terrific Broth in a 96-well format and incubated overnight at 37°C. The plasmids were miniprepped using QIAGEN Qiaprep 96 Turbo (QIAGEN, Valencia, CA). Amplified mRNAs were sequenced and the sequences were subjected to BLAST searches to identify the gene amplified.

Gene/Protein sequence analysis

All sequence analysis was performed using the GCG package (Wisconsin Package Version 10.2-UNIX), including BLAST searches run to find gene homologies. BLAST searches were alternatively run on the NIH website (www.ncbi.nlm.nih.gov). The blastp parameters were nr database, with BLOSUM 62 matrix, and gap costs were existence:11 extension:1.

Results and Discussion

Internal amino acid sequencing

Three bands were evident when the originally purified 3,4-dihydroxybenzoate decarboxylase was re-analyzed by SDS-PAGE. Two of the bands were close together and were ~50 kDa (He and Wiegel, 1996). Since the sample had been stored at -80°C for

about 4 years, the smaller band was assumed to be a degradation product. When the sample was originally run, only one band was evident at around 50 kDa. For internal amino acid sequencing, the lower molecular weight band was sequenced. The two internal amino acid sequences that were obtained are as follows:

NTSLDEFIEIGK(K)

and

LKGDSYMSIMER

A homology search (BLAST) of either of these two peptides revealed no homologous sequences. This could be due to the short length of the sequences. If the gene for ShdC were similar to the gene for Shd34, then one might expect some level of homology, but none was seen. It could be that the wrong band was sequenced (no time remained to sequence the other band), or that the two sequences bear little homology to each other.

PCR amplification of a fragment of the gene encoding 3,4-dihydroxybenzoate decarboxylase

Several attempts were made to amplify a fragment of the gene encoding Shd34 by PCR. A description of the primers used is in Table A.1. NtermA with Rev1A, NtermB with Rev1A, NtermA with Rev2A, NtermB with Rev2A, Nterm2 with Rev1A, and Nterm2 with Rev2A were all primer combinations used in an attempt to amplify a fragment of the gene encoding Shd34. These primers were derived from the N-terminal amino acid sequence of Shd34 and consensus sequences of ShdC with other homologous proteins. These consensus sequences are from a sequence comparison performed by

Jianbin Huang *et al.* (Huang, *et al.*, 1999). The only PCR product at 55°C resulted from Rev2A priming twice. This product was sequenced, and the amino acid translation revealed a protein homologous to numerous dehydrogenases, from *Clostridium dificile* and others identified by a BLAST search. Another attempt to amplify a gene fragment by PCR used a primer derived from the N-terminal amino acid sequence and an updated consensus sequence of ShdC, shown in Figure A.2. The other homologous proteins are all hypothetical, with the exception of those identified and characterized in this study. The primers used were ntermC with 34EGPFrev and ntermC with 34vivvrev. The only product was a result of EG priming twice. The lack of a PCR product from these primers may give several indications. Shd34 may not be homologous to 4-hydroxybenzoate decarboxylase, so the reverse primers designed with consensus regions may not be annealing. These reverse primers were designed using an older boxshade, so perhaps one designed with a new boxshade, with the proposed gene family well represented, could be used in the future. It could also be that the degeneracy of the primers is allowing too many PCR products such that the product of interest cannot be differentiated.

NtermA with NKDKR, NtermA with LRYER, NKDKF with LRYER, and LRYEF with NKDKR were primer combinations that were derived from both N-terminal and internal amino acid sequences. These PCR reactions were run at very low temperatures ranging from 35-45°C. There was a product from NtermA with LRYER and from NKDKF with LRYER. If the NKDKF with LRYER product was DIG-labelled during PCR, it could hybridize to itself and NtermA+LRYER in a dot blot. This made sense, seeing that NtermA+LRYER is a larger product from the N-terminal to an internal amino acid sequence and NKDKF+LRYER is a smaller product that would presumably

result from 2 internal amino acid sequences. Unfortunately, when NtermA+LRYER was sequenced, the sequence for NKDKF was not found therein. This suggests that these products are due to very promiscuous priming. The primers involved in the previous PCR were redesigned so as to be more stringent. These primers are ntermC, int1f, int1r, and int2R. The primer combinations attempted were ntermC with int1r, ntermC with int2R, and int1f with int2R. At low temperatures, around 38°C, there were too many products, but at higher temperatures around 45°C there was only one. This product was due to int1f priming twice. It is quite possible that the wrong protein fragment was sequenced, and that the internal amino acid sequences may be incorrect. Also, the degeneracy of the primers was a hindrance since too many PCR products were obtained. At higher temperatures, the other PCR products disappeared, leaving only the products obtained by the reverse primer priming twice.

A final attempt was made to amplify just the N-terminal region of the gene so as to make a Southern probe. This PCR involved ntermC with 34ntermCrev. There were too many products, as the primers were too degenerate at the low temperatures. Higher temperatures eliminated almost all of the products. After difficulty observing a 75 bp product, attempts were ceased.

Southern blotting using 4-hydroxybenzoate decarboxylase gene as a probe or the N-terminal sequence of Shd34

It was hypothesized that the sequence of Shd34 would be similar enough to ShdC that *shdC* would make a good probe for *shd34*. After hybridization with a *shdC* sequence that was DIG-labelled, there was only one band per enzyme digest. This was even the

case at low stringencies of 68°C and lower. It is possible that *shd34* does not bear enough homology to *shdCD* such that *shdCD* would make an effective hybridization probe. This seems unlikely as the major subunit of both enzymes is 57 kDa and their enzyme characteristics are quite similar. Alternatively, it is possible that colorimetric detection is not sensitive enough to detect low levels of hybridization which may have occurred if the two genes were only slightly homologous.

Attempts to use the N-terminal amino acid sequence to design a probe, namely 34MNKV from Table A.2, were made. The probe was too weakly labelled, degenerate, and did not bind to any single, small band in the genomic digest. Attempts to make a more highly labelled probe by incorporating DIG label while PCR amplifying the N-terminal sequence region using 34ntermC with 34ntermCrev were unsuccessful.

Differential display results

The RNA extracted from *S. hydroxybenzoicum* had an irregular pattern on an agarose gel, yielding bands at 900 bp and 1,100 bp instead of the expected 3 kb, 2.3 kb, 1.8 kb, and 0.9 kb RNA molecules. The expected RNA molecules are rRNA that have the same molecular weights across bacterial species. There could be degradation, although the source is unknown. The freezing of the cells was not responsible, as fresh cells yield the same banding pattern. The Trizol method is not at fault as RNA isolated by phenol preparation and the RNeasy (Qiagen) method gave similar results. The unexpected RNA pattern may offer an explanation for the failure to identify the gene of interest.

The genes that were differentially displayed and were identifiable by homology are listed in Table A.3. These sequences are the ones that were identified by at least 3 contiguous sequences and had at least 200 bp of sequence. No sequence bore homology to 4-hydroxybenzoate decarboxylase (any of the 3 genes in the *Hin*dIII fragment), *vdc*B, or the internal/N-terminal amino acid sequence of 3,4-dihydroxybenzoate decarboxylase. Of the identifiable gene products, there were none that could be construed to be involved in downstream reactions of 3,4-dihydroxybenzoate degradation. Of course, it does not seem, with either 4-hydroxybenzoate or 3,4-dihydroxybenzoate, that the organism degrades the resulting phenol/catechol further.

There were over 25 sequences that were not identifiable. They all had over 200 bp and were identifed by at least 3 contiguous sequences. If *shd34* does not bear homology to other known genes, including *shdCD*, it would be difficult to determine if the genes identified by differential display were *shd34*. No Pfam searches were performed to find possible related sequences. All sequence analysis involved pairwise comparison to identify the mRNA sequences obtained. Future experiments could involve identifying these unknown genes.

```
Shd34 MNKV.TDLRSAIELLKTIPGQLIETNXDVNTSLD
      | ||  |||   | |    ||||     ||   |
ShdC  MAKVYKDLREFLEVLEQ.EGQLIRVKEEVNPEPD
```

Figure A.1- Pairwise alignment between N-terminal amino acid sequence of Shd34 and ShdC

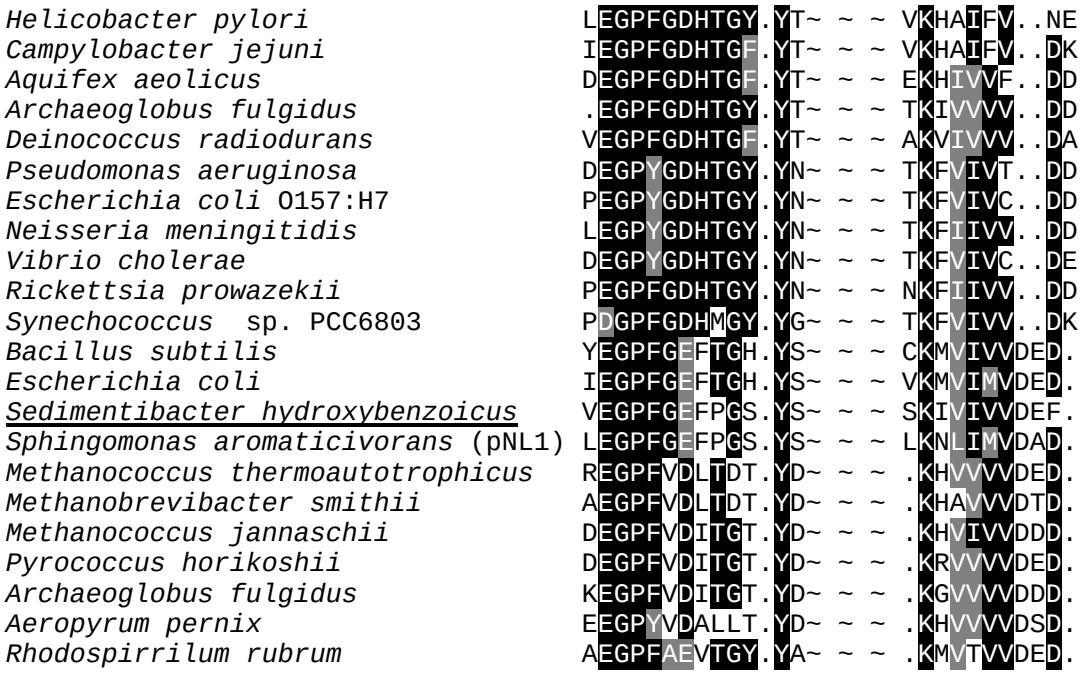


Figure A.2 - Sequence comparison illustrating the consensus regions in ShdC used to design primers to amplify *shd34*

Table A.1- Primers used to PCR amplify the gene encoding Shd34

Primer Name	Direction	Primer Sequence (5'-3')	Origin of Primer
NtermA	forward	ACA GAT TTA AGA AGT GCA	Shd34 N-term
NtermB	forward	ACA GAT TTG AGA TCA GCT	Shd34 N-term
Rev1A	reverse	AAT ATA AAA ACC TCC ATC	ShdC 143-138
Rev2A	reverse	AGG GGT TGT AGC ATC TAT	ShdC 440-445
Nterm2	forward	TTG TCC AGG TAT TGT YTT	Shd34 N-term
NKDKF	forward	GAY GAR TTY ATY GAT ATY GGN AA	Shd34 internal
NKDKR	reverse	YTT NCC RAT YTC RAT RAA YTC RTC	Shd34 internal
LRYEF	forward	TAY ATR WSN ATY ATR GA	Shd34 internal
LRYER	reverse	YTC YAT RAT NSW YAT RTA	Shd34 internal
ntermC	forward	ATG AAY AAR GTN ACN GAY YT	Shd34 N-term
int1f	forward	GAY GAR TTY ATH GAR ATH GG	Shd34 internal
int1r	reverse	CC DAT YTC DAT RAA YTC RTC	Shd34 internal
int2r	reverse	CK YTC CAT DAT NSW CAT RTA	Shd34 internal
34EGPFrev	reverse	AA TTC DCC RAA HGG DCC YTC	4OH consensus EGPFGEF
34 vivvrev	reverse	TC TTC RTC WAC WAC DAT NAC	4OH consensus VIVVDED
34ntermCrev	reverse	TTN GTY TCD ATN ARY TG	Shd34 N-term

R=A,G; Y=C,T; M=A,C; K=G,T; S=C,G; W=A,T; H=A,C,t; B=C,G,T; V=A,C,G; D=A,G,T; N=A,C,G,T

Table A.2 - Probes used for Southern hybridization to identify genomic fragment with *shd34*

Probe Name	Probe Sequence	Origin of Probe
34MNKV	ATG AAY AAR GTD ACN GAY YTD MGN WSN GCN ATW GAA YTD YTD AAR ACN ATW CCD GGH CAR YTD ATW GAA AC	translated Shd34 N-term
*shd*C	the gene sequence for *shdC*	*shdC*

Table A.3- Differentially displayed proteins as identified by homology through BLAST searches

Name of homologue	Primer number(s)	Contigs	Base Pairs
ABC Transporter	ACG63, ACG64, ACG146	13	549
ABC Transporter RbsC	ACG59, ACG163	3	1004
Amino Acid Transport Protein	ACG57	4	460
Catalase KatB	ACG182	7	829
Collagen adhesin	ACG11	4	524
Collagen-like Surface Protein	ACG158	3	750
DNA Mismatch Repair Protein	ACG20, ACG56	12	564
DNA Recombinase	ACG134, ACG53	14	717
DNA-K Type Molecular Chaperone	ACG189	3	555
Endopeptidase Clp	ACG64, ACG146	3	779
Hypothetical YqeY	ACG72, ACG75	5	402
Hypothetical YxcA - Dehydratase	ACG222,	6	1369
Hypothetical Gene for Aluminum Resistance	ACG11	3	229
Integrase	ACG19	4	472
Acetyl Coenzyme A Carboxylase	ACG53	3	387
DnaK-type Molecular Chaperone	ACG189	3	144
Exoglucanase I Precursor	ACG65	3	526
Site-Specific Recombinase	ACG59, ACG101	5	527
SRP-Cell Division Protein	ACG53	3	434
Amidohydrolase	ACG61	3	523
DNA Methyltransferase	ACG243	4	1,173
Uroconate Hydratase	ACG75	5	688
Protease Inhibitor	ACG11, ACG16, ACG25, ACG156	9	590
Pyruvate Dehydrogenase	ACG67, ACG49	14	788
Coporporphyrin Oxidase	ACG25	8	740
Acylamino Peptidase	ACG20	3	273

Literature cited

1. **Ausubel, F. M., R. Brent, R. E. Kingston, D. D. Moore, J. G. Seidman, J. A. Smith, and K. Struhl.** 1995. Current Protocols in Molecular Biology. John Wiley & Sons, USA.

2. **Brzostowicz, P. C., K. L. Gibson, S. M. Thomas, M. S. Blasko, and P. E. Rouviere.** 2000. Simultaneous identification of two cyclohexanone oxidation genes from an environmental *Brevibacterium* isolate using mRNA differential display. J. Bacteriol. **182**(15)**:**4241-4248.

3. **He, Z., and J. Wiegel.** 1995. Purification and characterization of an oxygen-sensitive reversible 4- hydroxybenzoate decarboxylase from *Clostridium hydroxybenzoicum*. Eur. J. Biochem. **229**(1)**:**77-82.

4. **He, Z., and J. Wiegel.** 1996. Purification and characterization of an oxygen-sensitive, reversible 3,4-dihydroxybenzoate decarboxylase from *Clostridium hydroxybenzoicum*. J. Bacteriol. **178**(12)**:**3539-43.

5. **Huang, J., Z. He, and J. Wiegel.** 1999. Cloning, characterization, and expression of a novel gene encoding a reversible 4-hydroxybenzoate decarboxylase from *Clostridium hydroxybenzoicum*. J. Bacteriol. **181**(16)**:**5119-22.

6. **Walters, D. M., R. Russ, H. J. Knackmuss, and P. E. Rouviere.** 2001. High-density sampling of a bacterial operon using mRNA differential display. Gene **273**(2)**:**305-15.

7. **Zhang, X., L. Mandelco, and J. Wiegel.** 1994. *Clostridium hydroxybenzoicum*sp.nov., an amino acid-utilizing, hydroxybenzoate-

decarboxylating bacterium isolated from methanogenic freshwater pond sediment. Int. J. Syst. Bacteriol. **44**(2)**:**213-222.

8. **Zhang, X., and J. Wiegel.** 1990. Isolation and partial characterization of a *Clostridium* species transforming *para*-hydroxybenzoate and 3,4-dihdyroxybenzoate and producing phenols as the final transformation products. Microbiol. Ecol. **20:**103-121.