CHARACTERIZATION OF A VANILLATE NON-OXIDATIVE DECARBOXYLATION GENE CLUSTER FROM *STREPTOMYCES* SP. D7

by

KEVIN TOSHIO CHOW

M.Sc., The University of British Columbia, 1996

B.Sc., The University of British Columbia, 1992

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF

THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

in

THE FACULTY OF GRADUATE STUDIES

(Department of Microbiology and Immunology)

THE UNIVERSITY OF BRITISH COLUMBIA

December 1999

© Kevin Toshio Chow, 1999

National Library of Canada

Bibliothèque nationale du Canada

Acquisitions and Bibliographic Services

Acquisitions et services bibliographiques

395 Wellington Street
Ottawa ON K1A 0N4
Canada

395, rue Wellington
Ottawa ON K1A 0N4
Canada

Your file Votre référence

Our file Notre référence

The author has granted a non-exclusive licence allowing the National Library of Canada to reproduce, loan, distribute or sell copies of this thesis in microform, paper or electronic formats.

L'auteur a accordé une licence non exclusive permettant à la Bibliothèque nationale du Canada de reproduire, prêter, distribuer ou vendre des copies de cette thèse sous la forme de microfiche/film, de reproduction sur papier ou sur format électronique.

The author retains ownership of the copyright in this thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without the author's permission.

L'auteur conserve la propriété du droit d'auteur qui protège cette thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation.

0-612-48620-6

ABSTRACT

The genetics of non-oxidative decarboxylation of aromatic acids to phenolic compounds are poorly understood in both prokaryotes and eukaryotes. Although such reactions have been observed in numerous microorganisms acting on a variety of substrates, genetic analyses of these processes have not, to my knowledge, been reported in the literature. Previously, I isolated a streptomycete from soil (*Streptomyces* sp. D7), which efficiently converts 4-hydroxy-3-methoxybenzoic (vanillic) acid to 2-methoxyphenol (guaiacol). Protein two-dimensional gel electrophoresis revealed that several proteins are synthesized in response to vanillic acid, one of which was characterized by partial amino-terminal sequencing, leading to the cloning of a gene cluster from a genomic lambda phage library of *Streptomyces* sp. D7. This cluster consists of four open reading frames, *vdcA* (sequencing in progress), *vdcB* (602 bp), *vdcC* (1424 bp) and *vdcD* (239 bp). Protein sequence comparisons suggest that the product of *vdcB* (201 aa) is similar to phenylacrylate decarboxylase of yeast; the putative products of *vdcC* (475 aa) and *vdcD* (80 aa) are similar to hypothetical proteins of unknown function from various microorganisms, and are found in a similar gene cluster in *Bacillus subtilis*. *VdcA* is a putative transcriptional regulatory gene. *VdcB*, *vdcC* and *vdcD* homologues are also clustered, along with putative *p*-cresol methylhydroxylase and vanillin oxidoreductase genes, on the 184 kb catabolic plasmid pNL1 of *Sphingomonas aromaticivorans* F199. Northern blot analysis revealed the synthesis of a 2.5 kb mRNA transcript, which hybridized strongly to a *vdcC* gene probe, in vanillic acid-induced cells, suggesting that the cluster is under the control of a single inducible promoter. Expression of the entire *vdc* gene cluster in *Streptomyces lividans* 1326, as a heterologous host, resulted in that

strain acquiring the ability to decarboxylate vanillic acid to guaiacol non-oxidatively. Both *Streptomyces* strain D7 and recombinant *S. lividans* 1326 expressing the *vdc* gene cluster do not, however, decarboxylate structurally similar aromatic acids, suggesting that the system is specific for vanillic acid. By Southern blot hybridization, we detected the presence of the *vdc* gene cluster in several streptomycetes, including *Streptomyces setonii* 75Vi2, which has been previously shown to decarboxylate vanillic acid in a non-oxidative reaction. The vanillate decarboxylase catabolic system may be useful as a component for pathway engineering research focused towards the production of valuable chemicals from forestry and agricultural byproducts.

TABLE OF CONTENTS

ABSTRACT	ii
TABLE OF CONTENTS	iv
LIST OF TABLES	vii
LIST OF FIGURES	viii
LIST OF ABBREVIATIONS	xi
ACKNOWLEDGEMENTS	xii
Chapter 1: INTRODUCTION	1
1.1 Non-oxidative decarboxylation: an industrially useful but poorly characterized process	4
1.2 Non-oxidative decarboxylation in prokaryotes	7
1.3 Structure and function studies of non-oxidative decarboxylases	9
1.4 Metabolic engineering for the production of chemicals from renewable resources	10
1.5 *Streptomyces:* versatile soil microbes	11
1.6 Isolation and identification of *Streptomyces* sp. D7	15
1.7 Catabolic tests and 2D-PAGE analysis	17
1.8 Objectives	22
Chapter 2: MATERIALS AND METHODS	23
2.1 Bacterial strains and plasmids	23
2.1.1 Bacterial strains	23
2.1.2 Isolation and cloning of plasmids	23
2.1.3 16s rDNA sequence-based strain identification	28
2.2 Media and growth conditions	28
2.3 Library construction and gene cloning	31
2.3.1 Lambda phage DNA preparation	31

2.4 DNA sequencing and analysis 32

2.5 Southern blotting and hybridization 33

2.6 Chemical analyses 34

2.7 RNA isolation and analysis 34

2.8 Gene expression 38

2.8.1 pET22b(+) – *E. coli* BL21 38

2.8.2 pIJ702 – *S. lividans* 1326 38

2.8.3 pIJ680 – *S. lividans* 1326 40

2.9 Enzyme assays 42

2.10 Recombinant protein purification from *E. coli* BL21(DE3) 42

2.10.1 Lysate preparation 43

2.10.2 Inclusion body protein purification, solubilization and refolding procedure 43

2.10.3 Soluble protein purification 44

2.11 Construction of a knockout allele of the *vdcC* gene 44

2.12 *Streptomyces* sp. D7 protoplast preparation and transformation 49

2.13 *E. coli-Streptomyces* interspecies conjugation 49

2.13.1 *E. coli* S17-1/pPM801 preparation 49

2.13.2 *Streptomyces* sp. D7 preparation and mating 50

2.14 Reagents and enzymes 50

Chapter 3: RESULTS 51

3.1 16s rDNA sequence identification 51

3.1 Cloning of the VDC gene cluster 51

3.2 Detection of a putative transcriptional regulatory gene 69

3.3 Messenger RNA analyses 72

3.4 Detection of *vdc* genes in *S. setonii* 75Vi2 74

3.5 Gene knockout studies 76

3.6 Gene expression 77

3.6.1 pET22b(+) – *E. coli* BL21 77

3.6.2 pIJ702 – *S. lividans* 1326 81

3.6.3 pIJ680 – *S. lividans* 1326 81

3.7 Substrate specificity 82

Chapter 4: CONCLUSIONS 88

Chapter 5: DISCUSSION 89

5.1 Aromatic acid non-oxidative decarboxylases are multi-subunit enzymes 89

5.2 Distribution of the VDC gene cluster among streptomycetes 91

5.3 Substrate specificity 92

5.4 Primary structure motifs 93

5.5 Transcriptional activation 93

5.6 Biodegradation – a result of the microbial community gene pool 94

5.7 *Sphingomonas aromaticivorans F199* catabolic plasmid pNL1 96

5.8 Metabolic engineering applications for vanillate decarboxylase 103

5.9 Future directions 106

Chapter 6: LITERATURE CITED 109

LIST OF TABLES

Table		Page
1	Bacterial strains and plasmids	24
2	PCR primers and hybridization oligonucleotides	36
3	Variations in subunit size and configuration – characteristics of some microbial aromatic acid non-oxidative decarboxylases	90

LIST OF FIGURES

Figure		Page
1	Lignin backbone structure	2
2	Solubilization of lignin yields a mixture of low-molecular weight aromatic compounds	3
3	Reaction scheme for non-oxidative decarboxylation of vanillate to guaiacol	6
4	The *Streptomyces* growth cycle	14
5	A possible niche for *Streptomyces* in lignin degradation	15
6	Decarboxylation of vanillic acid to guaiacol by *Streptomyces* sp. D7	19
7	Experimental scheme for growth and vanillic acid induction of *Streptomyces* sp. D7 cells for 2D-PAGE analyses	20
8	Synthesis of protein 3717 in response to vanillic acid	21
9	Map of expression vector pIJ680	29
10	Map of cloning/expression vector pIJ702	30
11	pIJ702 and the 4 kb BamHI *Streptomyces* sp. D7 genomic DNA fragment inserted in both orientations in the tyrosinase gene (mel) promoter	39
12	pIJ680 and the PCR-generated gene combinations inserted downstream of the promoter for the aminoglycoside phosphotransferase (aph) gene	41
13	Strategy for a knockout of the *vdcC* gene using the thiostrepton (tsr) resistance gene	45
14	Amino terminal sequence alignment of *Streptomyces* sp. D7 protein 3717 and *Clostridium hydroxybenzoicum p*-hydroxybenzoate carboxy-lyase	54
15	Lambda DASH II phage clone carrying the putative vanillate decarboxylase gene	55
16	DNA purified from phage clone 3717C(+) and digested with *Bam*HI	56
17	pSUB1 (Lane A), pSUB3 (Lane B), pSUB4 (Lane C), pSUB5 (Lane D) and pSUB6 (Lane E) purified plasmid DNA, digested with *Bam*HI	57

18	Plasmids carrying subfragments of the *Bam*HI genomic DNA fragment from *Streptomyces* sp. D7, which was cloned in lambda DASH II phage clone 3717C(+)	58
19	Nucleotide sequence of the *vdc* gene cluster, featuring *vdcB*, *vdcC* and *vdcD*	59
20	Schematic diagram of the 4.4 kb BamHI genomic DNA fragment from *Streptomyces* sp. D7, containing the VDC gene cluster	60
21	Dendrograms depicting the sequence-based relationships between proteins and hypothetical proteins similar to the product of the *vdcB* gene and *vdcC* gene	61
22	Location of the putative divergent regulatory gene in relation to the VDC gene cluster	70
23	BLAST-X result indicating that the nucleotide region immediately upstream of *vdcB* encodes a polypeptide similar to a putative transcriptional regulator from *S. coelicolor* A3(2)	71
24	Northern blot hybridization of PCR-amplified, radiolabeled *vdcC* DNA against mRNA isolated from uninduced *Streptomyces* sp. D7 (Lane 1) and *Streptomyces* sp. D7 induced with 3.6 mM vanillic acid (Lane 2)	73
25	Southern blot hybridization of *Sal*I-digested chromosomal DNA from *Streptomyces* sp. D7 (Lane 1) and *Streptomyces setonii* 75Vi2 (Lane 2) with radiolabeled *vdcB*, *vdcC* and *vdcD* PCR-amplified DNA probes at 65°C	75
26	SDS-PAGE (12.5% acrylamide) of cell extracts showing the expression product of the *vdcC* gene, produced using pET22b(+) in *E. coli* BL21(DE3)	79
27	SDS-PAGE (12.5% acrylamide) of cell extracts showing the expression product of the *vdcC* gene, produced using pET22b(+) in *E. coli* BL21(DE3) at 25°C and 30°C with 0.1 mM IPTG for induction	80
28	Decarboxylation of vanillic acid to guaiacol by recombinant *Streptomyces lividans* 1326 strains	84
29	Results of incubating the insoluble and soluble fractions of sonicated cell extracts of *S. lividans* 1326 – pKCS8 with 1 mM vanillic acid for fifteen minutes at 25°C under both aerobic (O_2) and anaerobic (N_2) conditions	85

30	Expression of the *vdc* genes under control of the *aph* promoter in pIJ680	86
31	Chemical structures of vanillic acid and similar aromatic acids used to test the substrate specificity of vanillate decarboxylase	87
32	Amino acid sequence comparisons between the *Streptomyces* sp. D7 *vdcB* translation product (strPAD) and its *Sphingomonas* pNL1 homologue (sphPAD) **(a)**, and the *Streptomyces* sp. D7 *vdcC* translation product (strC) and its *Sphingomonas* pNL1 homologue (sphC) **(b)**	97
33	Physical map of pNL1 and the location of the VDC gene cluster homologues	102
34	Vanillate decarboxylase, Cytochrome P-450 and catechol 1,2-dioxygenase can be used in combinations to produce various industrially useful chemicals from vanillic acid as a starting material	105
35	Non-oxidative decarboxylases represent connections between the two major branches of aromatic acid catabolism, characterized by either catechol or protocatechuate central intermediates	107

LIST OF ABBREVIATIONS

Abbreviation	Definition
2D-PAGE	Protein two-dimensional polyacrylamide gel electrophoresis
aph	Aminoglycoside phosphotransferase gene
BLAST	Basic Alignment and Search Tool
DNA	Deoxyribonucleic acid
DTT	Dithiothreitol
EDTA	Ethylenediaminetetraacetate
FAD	Flavin adenine dinucleotide
FAME	Fatty acid methyl ester
HPLC	High pressure liquid chromatography
Kb	Kilobase
KDa	Kilodalton
mel	Melanin synthesis gene (tyrosinase)
MRNA	Messenger ribonucleic acid
MSMYE	Mineral salts medium with yeast extract
PAD	Phenyl acrylate decarboxylase
PCR	Polymerase chain reaction
PEG	Polyethylene glycol
pI	Isoelectric point
NMR	Nuclear magnetic resonance
RBS	Ribosome binding (Shine-Delgarno) site
RNA	Ribonucleic acid
SSC	Sodium saline citrate
SDS	Sodium dodecyl sulfate
TSB	Trypic soy broth
UV	Ultraviolet
VDC	Vanillate decarboxylase
YEME	Yeast extract malt extract

ACKNOWLEDGEMENTS

I thank all the members, past and present, of the Julian Davies Laboratory for their support, encouragement, and friendship. In particular, I thank Rumi Asano, Dr. Jeff Rogers, Sakura Iwagami-Hayward, Grace Law and Richard Kao for being great graduate school companions. In addition, I gratefully acknowledge Dr. Margaret K. Pope for providing essential guidance and technical expertise for the project. I thank Dr. Vera Webb for making it possible to prove myself in the laboratory. Vincent Martin, Professor William Mohn, Professor Douglas Kilburn, and Professor James Kronstad provided valuable technical advice for the project. I am grateful to my parents, who always believed in me, and Alisa Chan, who provided much appreciated support to get me through the most trying stages of graduate school life. Canadian Forest Products (Canfor) Research and Development Centre staff provided valuable expertise and resources. Most importantly, Professor Julian Davies (who said I would be rich and/or famous someday) provided me with the encouragement, support, finances, and worldwide network of contacts that made my efforts worthwhile. Finally, there are many others who helped me in one way or another, and to them I am sincerely thankful.

This research was funded by the GREAT Scholarship program provided by the Science Council of British Columbia (SCBC), Forest Renewal British Columbia (FRBC), and the Natural Sciences and Engineering Research Council of Canada (NSERC).

1. INTRODUCTION

Chemical manufacture of benzenoid compounds from petroleum relies on abiotic, chemical catalysts. The use of petroleum poses a number of problems, it being a non-renewable resource, a geopolitically volatile commodity, and a source of many environmentally toxic compounds. Therefore, there is growing interest in developing processes for enzymatic conversion of renewable resources such as plant biomass for the production of chemicals traditionally derived from petroleum. The abundance of phenylmethylether motifs in natural compounds such as lignin (Figure 1) and their release as a result of lignin solubilization (Figure 2) has resulted in the evolution of mechanisms for the degradation of phenolic structures by microorganisms. Microorganisms exhibiting such enzymatic biotransformation potential could be harnessed for industrial use to supplement or replace traditional chemical synthesis methods, should the need arise (Frost & Draths, 1995).

Vanillic acid is an abundant component of solubilized lignin biomass, and degradative mechanisms by which this compound is catabolized have been elucidated in several prokaryotic organisms. The genes responsible for vanillate demethylation have been cloned and sequenced from *Pseudomonas* sp. strain ATCC 19151 (Brunel & Davison, 1988), *Acinetobacter* sp. ADP1 (Segura & Ornston, 1997) and, most recently, *Sphingomonas paucimobilis* (Nishikawa *et al.*, 1998). In these microbes, vanillate is converted to protocatechuate, which is in turn degraded by enzymes of the β-ketoadipate pathway.

Figure 1: The lignin backbone structure.

Figure 2: Solubilization of lignin yields a mixture of low-molecular weight aromatic compounds, some of which are shown here.

However, in some strains of *Streptomyces* and *Bacillus* (Crawford & Olson, 1978), vanillate is catabolized via an alternative pathway involving non-oxidative decarboxylation to guaiacol (Figure 3), with further catabolism via cytochrome P-450 mediated demethylation and mineralization through the intermediate catechol. In fact, it was demonstrated that individual *Streptomyces* isolates degraded vanillate by both routes, that is, through both catechol and protocatechuate as central intermediates. There are a number of reports in the literature of aromatic acid non-oxidative decarboxylases from various microorganisms (Grant & Patel, 1969; Yoshida & Yamada, 1985; Nakajima *et al.*, 1992; Huang *et al.*, 1993; Santha *et al.*, 1995; He & Wiegel, 1995; He & Wiegel, 1996; Zeida *et al.*, 1998) but thus far there have been no molecular studies of these processes.

1.1 Non-oxidative decarboxylation: an industrially useful but poorly characterized process

Non-oxidative decarboxylation of aromatic acids involves the removal of the carboxyl moiety from the benzene nucleus via a reaction that requires neither oxygen, nor cofactors such as NAD and FAD, typical elements of the oxidative process. The non-oxidative process results in a "clean" removal of the carboxyl group, in contrast to the oxidative reaction, which substitutes a hydroxyl group at the relevant carbon atom. In nature, non-oxidative decarboxylation is not only observed for biodegradative pathways, but also for anabolic pathways, such as in the biosynthesis of naphthoquinones (Santha *et al.*, 1995). Similarly, for biotransformation and metabolic engineering applications, both oxidative and non-oxidative processes are valuable as components of hybrid pathways for the production of various industrially useful compounds. In fact, a *Klebsiella*

aerogenes protocatechuate non-oxidative decarboxylase has been engineered into a hybrid pathway to produce catechol, a useful building block for pharmaceuticals, from glucose as a renewable starting material (Frost & Draths, 1995).

COOH

CO_2

OCH_3

OH

Vanillic Acid

OCH_3

OH

Guaiacol

Figure 3: Reaction scheme for the non-oxidative decarboxylation of vanillate to guaiacol.

1.2 Non-oxidative decarboxylation in prokaryotes

Reports on microbial non-oxidative decarboxylation of aromatic acids date back as far as 1924, but the first published work that was of significance towards modern studies of these enzymes was in 1969, by Grant and Patel, then at the University of Nottingham. Their thorough physiological study of *Klebsiella aerogenes* and its abilities to non-oxidatively decarboxylate *p*-hydroxybenzoate, gentisate, protocatechuate and gallate revealed an organism that was likely responsible for decarboxylations of phenolic acids found in intestinal microflora. The study suggested that the aforementioned aromatic acids were decarboxylated by different enzymes, and that all of these enzymes were membrane associated, according to activity localization after ultrasonication and debris fractionation. The researchers noted that the non-oxidative decarboxylation of *p*-hydroxybenzoate was likely an injurious side reaction to the bacterium due to the toxic characteristics of the product, phenol.

As already mentioned, and relevant to this study, *Streptomyces* is one of the microorganisms in which non-oxidative decarboxylation has been biochemically well characterized. Crawford and Olson published data that would become the first in a series of reports detailing a streptomycete's capability to catabolize vanillate by decarboxylation to guaiacol (Crawford & Olson, 1978). *Streptomyces* strain 179 was isolated from Idaho forest soils, and exhibited what was then considered a novel catabolic reaction for the utilization of vanillic acid. Further research on another microorganism, *S. setonii* 75Vi2, revealed that, in addition to non-oxidative decarboxylation of vanillate to guaiacol, a cytochrome P-450 system is involved in demethylating guaiacol to

catechol. Catechol is then mineralized (presumably, though not demonstrated) by catechol 1,2-dioxygenase and other associated lower pathway enzymes (Sutherland, 1986). Although non-oxidative decarboxylation of vanillate by *S. setonii* 75Vi2 was well-documented biochemically, the genetic basis for the reaction was not investigated.

The purification and partial sequencing of two non-oxidative decarboxylase enzymes from the strict anaerobe *Clostridium hydroxybenzoicum* has been accomplished (He and Weigel, 1995). *C. hydroxybenzoicum*, an novel anaerobe, was isolated from freshwater sediments. The microorganism decarboxylates both 4-hydroxybenzoic acid and 3,4-dihydroxybenzoic acid under anaerobic conditions but does not metabolize the products further. The enzymes responsible for these reactions, *p*-hydroxybenzoate carboxy-lyase and 3,4-dihydroxybenzoate carboxy-lyase, showed no significant similarity to any protein sequences in the databases. The authors indicated that cloning of the genes encoding these proteins was in progress, but at the time of writing of this thesis, no such gene sequences have been reported.

From an industrial standpoint, non-oxidative decarboxylases that convert gallate to pyrogallol are particularly interesting. There are numerous reports in the literature detailing whole cell bioconversions of gallate using various microorganisms, both prokaryotic and eukaryotic. Yoshida and Yamada (1982) described the optimization of whole cell bioconversions using a *Citrobacter* sp., in which a yield of 97.4% pyrogallol was obtained from gallate as an initial substrate. A recent development with great industrial potential was a report of the purification and characterization of gallate

decarboxylase from *Pantoea agglomerans* T71 (Zeida *et al.*, 1998). *P. agglomerans* T71 expresses not only a gallate non-oxidative decarboxylase, but also a tannase, which allows the organism to produce pyrogallol from tannic acid, an abundant waste product of the forest industry. The gallate decarboxylase was highly specific for gallate. As in the other studies mentioned in this survey, the genes encoding the decarboxylase were not cloned; however, the study emphasized the industrial relevance of this seemingly simple, yet poorly characterized, reaction.

1.3 Structure and function studies of non-oxidative decarboxylases

Several reports have described the instability of aromatic acid non-oxidative decarboxylases during purification procedures. Perhaps it is this characteristic of these enzyme systems that has resulted in limited studies of the proteins and their corresponding genes. Nevertheless, several research groups have succeeded in purifying, partially sequencing, and characterizing various non-oxidative decarboxylases. Santha *et al.* (1995) reported the structure and function of 2,3-dihydroxybenzoic acid decarboxylase from *Aspergillus niger*. In this work, the active site peptide of the enzyme was determined, and a partial primary structure map was created based upon sequences derived from enzymic cleavage of the protein. The enzyme system did not require any cofactors, emphasizing the importance of multiple active site residues (yet to be characterized) in the reaction mechanism. Huang *et al.*, in a report describing the mechanism of action of ferulate decarboxylase in the yeast, *Rhodotorula rubra* (the enzyme does not decarboxylate the aromatic nucleus, but rather a side chain of the compound), also observed vanillate decarboxylation to guaiacol as a downstream

catabolic step (Huang *et al.*, 1993). The researchers included a NMR analysis of the chemical mechanism of vanillate non-oxidative decarboxylation in their studies. He *et al.*, in their study of the non-oxidative decarboxylases from *C. hydroxybenzoicum*, found that the enzyme did not require any cofactors or metal ions for activity (He *et al.*, 1995). The enzymes were observed to be reversible, depending on the reaction conditions. Finally, the aforementioned study of gallate decarboxylase in *P. agglomerans* T71 (Zeida *et al.*, 1998) details the stabilization of the enzyme and sequencing of its amino-terminus. The authors found that gallate decarboxylase from *P. agglomerans* T71 is unique among similar decarboxylases in that it requires iron as a cofactor. Information on non-oxidative decarboxylases from microorganisms is accumulating, from both fundamental and applied research, and thus far it can be generalized that these proteins form a class of enzymes which are fairly substrate specific, unstable, and for the most part function independently of cofactors.

1.4 Metabolic engineering for the production of chemicals from renewable resources

Vast amounts of aromatic carboxylic acids are available as natural products from plant and wood residues. However, exploitation of such substances as starting material for chemical syntheses or as fermentation substrates has not attracted the sustained attention of molecular biologists. Much remains to be learned about the microbiological systems that offer potential avenues for recovering the major biochemical resource that the plant-derived aromatic carboxylic acids represent. To my knowledge, there are no published studies of metabolically engineered organisms for the biotransformation of lignin residues. However, *E. coli* has been modified, in an elaborate genetic engineering

scheme, to bioconvert D-glucose to adipic acid, the building block of nylon (Draths & Frost, 1994). Briefly, *E. coli* AB2834, a mutant lacking shikimate dehydrogenase, was transformed with the plasmids pKD136 (encoding a transketolase, DAHP synthase, and DHQ synthase), pKD8.243A (encoding DHS dehydratase and protocatechuate decarboxylase), and pKD8.292 (encoding catechol 1,2-dioxygenase). The bacterial strain was kept stable due to plasmid compatibility and the use of drug resistance selections for each plasmid. *E. coli* AB2834/pKD136/pKD8.243A/pKD8.292 utilized its suite of aromatic amino acid biosynthetic genes as well as aromatic acid catabolic genes to effectively convert D-glucose, derived from many agricultural sources, to adipic acid. The process provides an effective alternative to conventional adipic acid synthesis from benzene, a process that results in the production of high amounts of nitric oxide, the "greenhouse" gas implicated in the depletion of the ozone layer of the Earth. It should be noted that almost all the genes used in the engineered *E. coli* strain have been patented, and the DNA sequences of these genes, if known, are not freely available to the public.

Vanillate decarboxylase from *Streptomyces* sp. D7 could be useful in the future as a critical link in a multi-step metabolic engineering regime similar to that described for *E. coli* AB2834. The utility of vanillate decarboxylase in value-added biomass conversions will be described in more detail in the Discussion section of this report.

1.5 *Streptomyces:* versatile soil microbes

Streptomyces is a genus of gram positive, filamentous, sporulating bacteria that are mainly native to soil, but are also found in aquatic environments. Being predominantly

soil-borne microbes (Atlas & Bartha, 1993), these organisms have developed diverse metabolic capabilities that allow for the production of a diverse array of chemical compounds. These secondary metabolites (produced during the late (stationary) phases of growth) have found many applications for humans, ranging from anti-inflammatory therapeutics to potent anti-microbial agents. The high industrial value of certain streptomycete strains for their secondary metabolites has resulted in the development of a detailed understanding of the genetics and physiology of these microorganisms (Hopwood *et al.*, 1985). Starting from spores, streptomycetes undergo germination to form substrate mycelial (hyphal) growth. Subsequently, the microorganism shifts to a secondary phase of growth, in which aerial mycelia are formed on the substrate mycelial base. It is during this phase that most of the industrially useful compounds are formed. Upon the induction of signaling responses (such as two-component serine-threonine phosphorylation cascades) due to environmental stimuli such as starvation, the organisms shift to a sporulation cycle, in which chains of uni-nucleate spores are produced at the end of the aerial mycelial tips. The streptomycete growth cycle is illustrated in Figure 4. At a genetic level, streptomycetes possess linear chromosomes, and may contain both circular and linear plasmids. As members of the Actinomycetes, they characteristically possess DNA with a high guanosine and cytosine (G+C) content, typically in the range of 70-75%. Currently, the *Streptomyces coelicolor* A3(2) genome sequencing project is well underway, and thus far, of the 8 Mb, >7000 gene (estimated) genome, just over 3 Mb have been sequenced. The data accumulated provides the following statistics (D. Hopwood, J. Davies, personal communication). The average G+C content is 71.72%. Of the open reading frames (ORFs), 3.5% have been previously sequenced, 50.5% resemble

those of known function, 19.2% resemble those of unknown function (hypothetical proteins), and 26.8% have no database match. There is an average of 1.14 kb per ORF, suggesting that the genome is tightly packed -- by comparison, the yeast *Saccharomyces cerevisiae* has 13 Mb and <6000 genes, with an average of 1.2 kb per ORF. Streptomycetes are renowned for their secondary metabolite production, but their abundance in soil, particularly in environments rich in humic matter and lignocellulose, have adapted them to degrade a wide variety of natural substances. These microorganisms provide a major contribution to the global carbon cycle by assisting fungi in the mineralization of cellulose and lignin. Surprisingly, this aspect of streptomycete biology has been little investigated; *Streptomyces viridosporus* T7A, which secretes a powerful lignin peroxidase is, thus far, the most well characterized lignin-degrading streptomycete (Thomas & Crawford, 1998). However, most streptomycetes isolated from soils do not degrade lignin, but are very efficient at the catabolism of lignin-related, low molecular weight aromatic compounds such as vanillic acid. A possible scenario (Kirk, 1987) is that in the natural consortia of microorganisms present in forest soils, fungi such as the basidiomycetes perform depolymerization of intact lignin, leading to the release of more soluble fragments which can be transformed and mineralized by other microorganisms such as streptomycetes (Figure 5). The non-oxidative decarboxylation of vanillic acid to guaiacol is one such transformation.

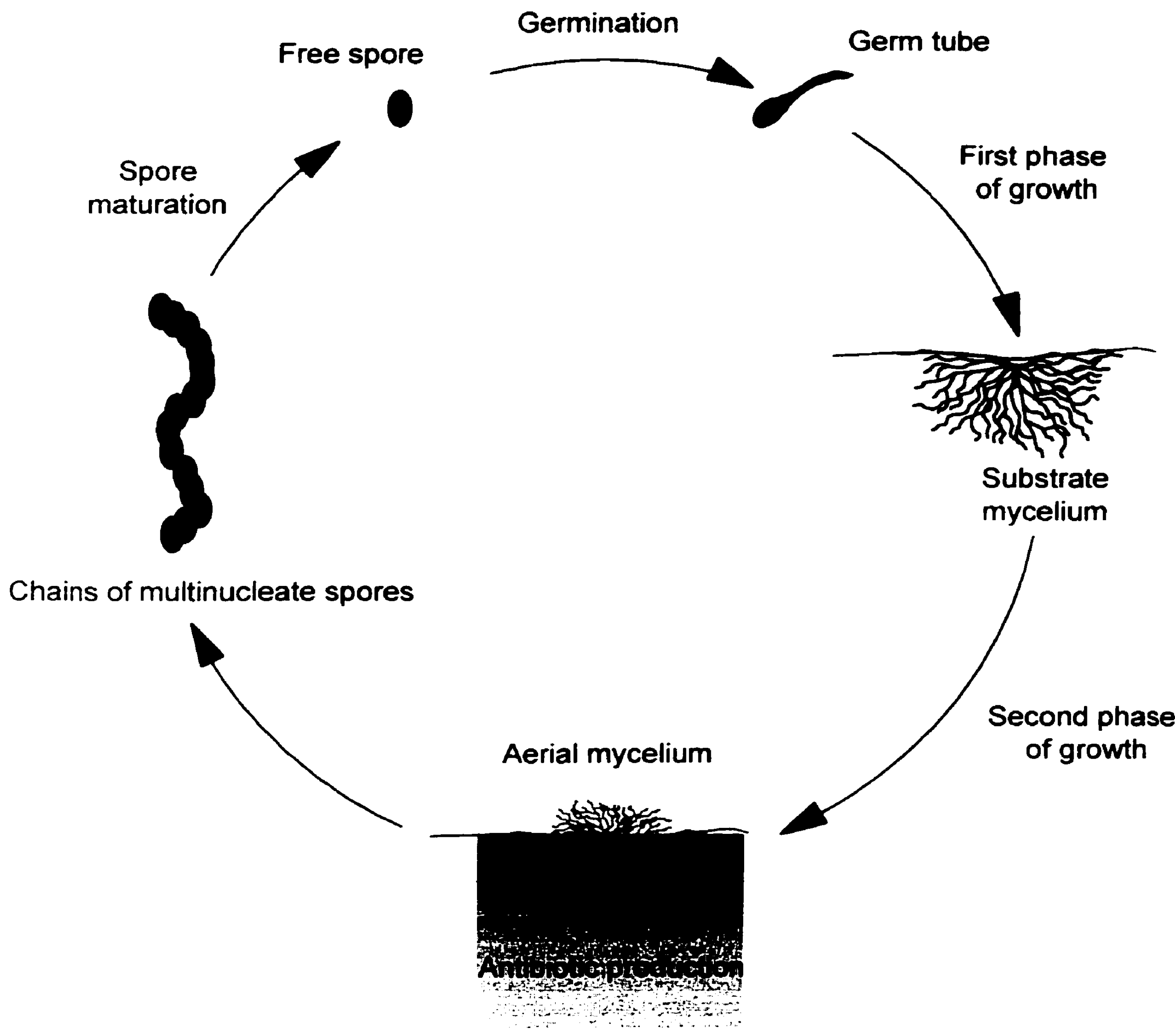


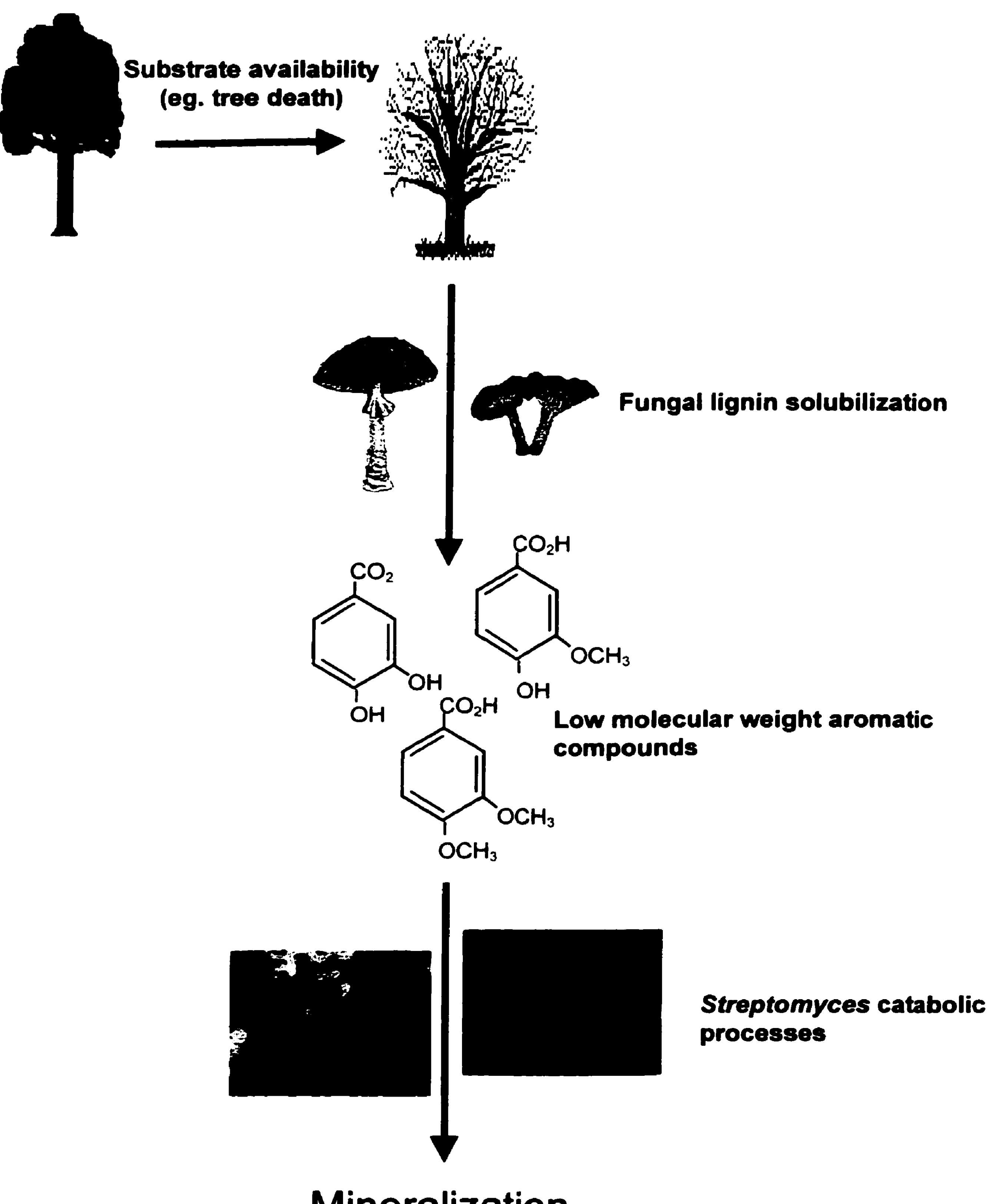
Figure 4: The *Streptomyces* growth cycle.

It is the intent of these studies to shed light on the process of non-oxidative decarboxylation of aromatic acids by utilizing a proteomics and functional genomics approach. By studying gene expression patterns during microbial catabolic processes, I have been able to partially characterize the molecular basis by which non-oxidative decarboxylation occurs in *Streptomyces*. and perhaps in other organisms with similar functions. The following information provides background details regarding the isolation of *Streptomyces* sp. D7, catabolic phenotyping, proteomic analysis of gene expression, and isolation and partial sequencing of a putative vanillate catabolic enzyme, leading to the work described in this thesis (Chow, 1996).

1.6 Isolation and identification of *Streptomyces* sp. D7

The organism used for this study, *Streptomyces* sp. D7, was isolated from a soil sample taken from forest land on the University of British Columbia campus in Vancouver, B.C., Canada (Chow, 1996). The organism produces abundant gray spores when grown on a mannitol soya agar medium, and produces a bright yellow diffusable, water-soluble pigment during growth in various solid and liquid media.

Figure 5 (following page): A possible niche for *Streptomyces* in lignin degradation.


Substrate availability
(eg. tree death)
Fungal lignin solubilization
CO2
OH
OH
CO2H
OCH3
OH
CO2H
OCH3
OCH3
Low molecular weight aromatic compounds
Streptomyces catabolic processes
Mineralization

1.7 Catabolic tests and 2D-PAGE analysis

Streptomyces sp. D7 was determined, by UV spectrophotometry and HPLC analyses of culture supernatants, to efficiently bioconvert vanillate to guaiacol (Figure 6), suggestive of the activity of a non-oxidative decarboxylase (Chow *et al.*, 1999). While *Streptomyces* sp. D7 was apparently capable of limited growth using the carboxylic acid moiety of vanillate as a sole carbon source, no further degradation of guaiacol was observed. As mentioned previously, other microorganisms and strains of *Streptomyces* have been shown to perform this enzymatic reaction (Crawford & Olson, 1978; Pometto III *et al.*, 1981; Sutherland *et al.*, 1981), but thus far, no genetic information has been published regarding these enzyme systems. In order to identify proteins synthesized during vanillate catabolism, high-resolution 2D-PAGE technology (the "Investigator System", Genomic Solutions Inc.) was used to visualize "genetic snapshots" of cellular activity when growing cells of *Streptomyces* sp. D7 were induced with non-inhibitory amounts of vanillate (Chow *et al.*, 1999; Chow, 1996). *Streptomyces* sp. D7 was grown in a mineral salts medium supplemented with 0.5% yeast extract until mycelia were in early logarithmic growth phase. At this point in the growth cycle, the culture was divided and 3.6 mM vanillic acid was added to one of the cultures to induce a response. Aliquots of both induced and uninduced cultures were pulse labeled with ^{35}S-methionine/cysteine at 1, 2, 5, 12, and 15 hours post-induction. A diagram depicting this experiment is shown in Figure 7. The labeled mycelium samples were sonicated to extract total cell protein and separated by 2D-PAGE. Compilation and analysis, by PDQUEST software (PDI, Inc.) using a SparcStation 5 workstation (Sun Microsystems), of numerous 2D-PAGE gels from several time course experiments resulted in the identification of at least six major

proteins that were synthesized in response to vanillate. The most prominent and abundant protein, of 52 kDa in molecular mass with a pI of 4.9 (Figure 8), was pooled from replicate gels, blotted to PVDF membrane and Edman-degradation sequenced, yielding sufficient and reliable amino-terminal data (AYDDLRYFLDTLEKEGQLLRIT) to allow synthesis of a degenerate oligonucleotide probe. This probe allowed me to proceed with the cloning of the gene encoding the 52 kDa, vanillate-induced protein in order to isolate and characterize the genetic elements forming the basis for vanillate decarboxylation in this organism. Such experiments form the basis for this thesis and are described herein.

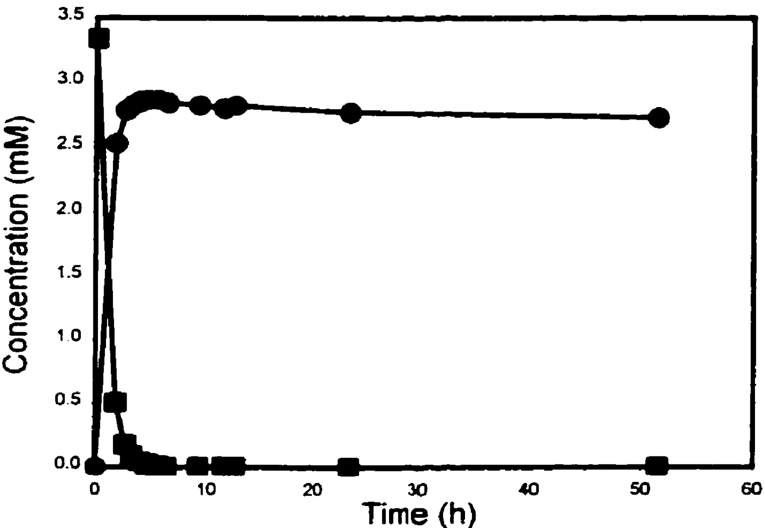


Figure 6: Decarboxylation of vanillic acid to guaiacol by *Streptomyces* sp. D7. Concentration of aromatic compounds in culture supernatants was measured using HPLC and known concentration standard solutions. Vanillic acid concentration is represented by squares (■); guaiacol concentration is represented by circles (●). Cells were grown to late log phase in YEME liquid medium, harvested, washed and resuspended in MSMYE minimal medium with approximately 3.6 mM vanillic acid. Each time point represents a concentration as measured by HPLC of culture supernatant samples.

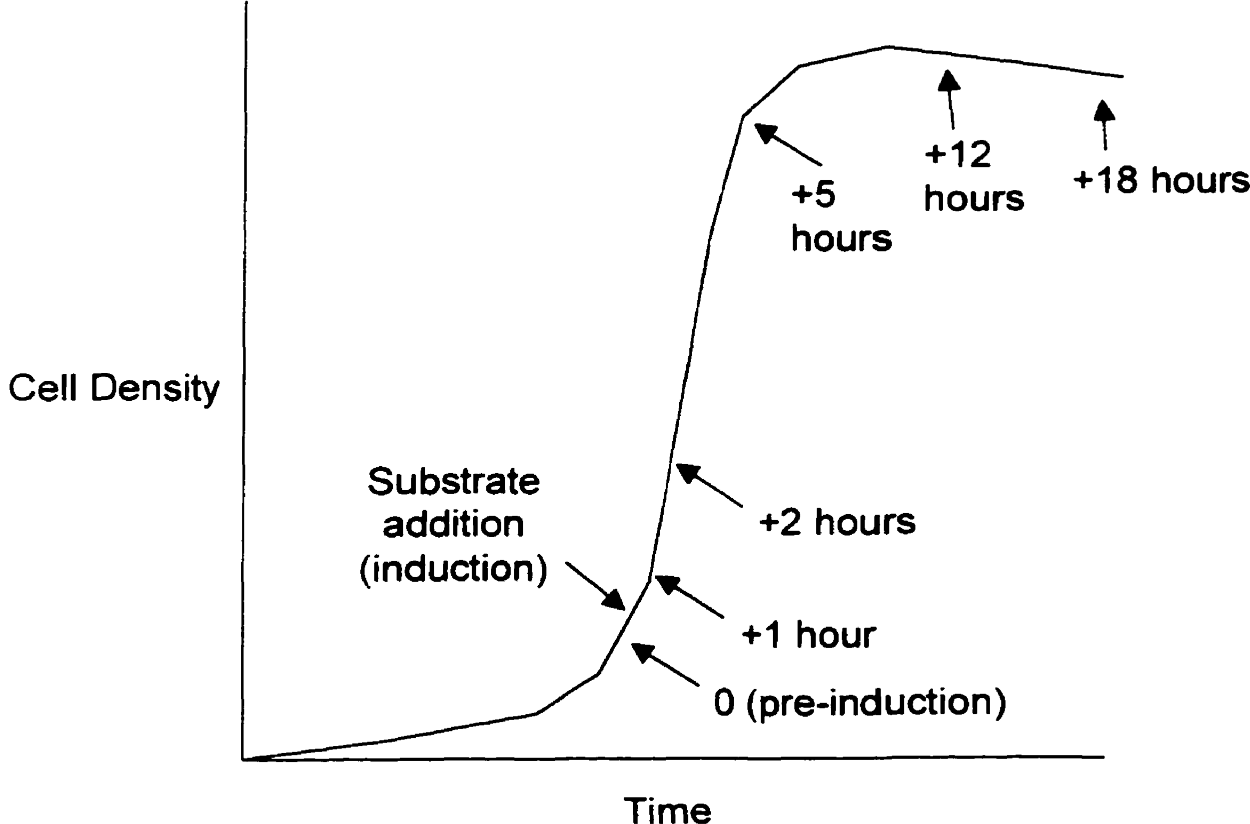


Figure 7: Experimental scheme for growth and vanillic acid induction by *Streptomyces* sp. D7 cells for 2D-PAGE analyses. 1 ml aliquots of mycelia were radioactively pulse labeled at the times indicated. The panels in Figure 6 (following page) correspond to these time points. (Chow, 1996)

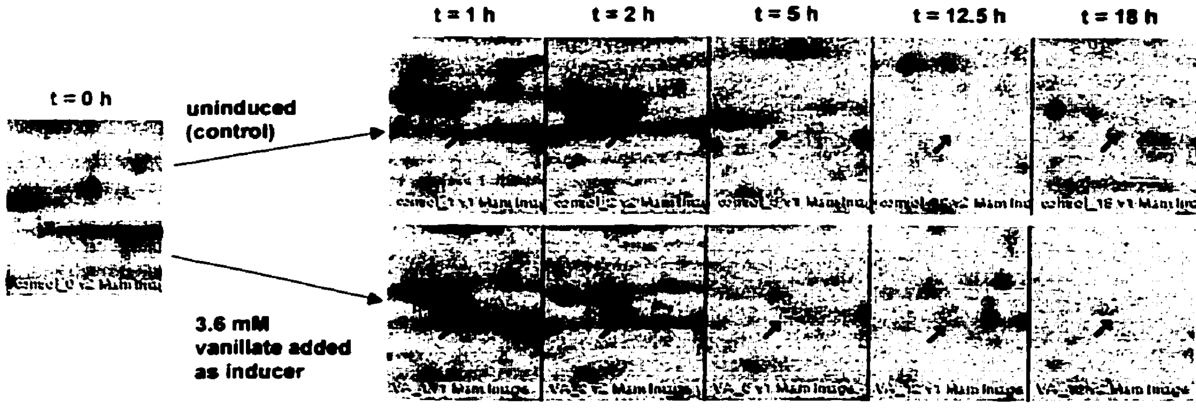


Figure 8: Protein 2D-PAGE profile of the synthesis of protein 3717 by *Streptomyces* sp. D7 in response to 3.6 mM vanillic acid. 2D-PAGE profiles of an uninduced culture are also shown for comparison. Time values are measured as hours post induction. Arrows highlight the area in which protein 3717 appears. The panels are enlargements of the 52 kDa, pI 4.9 region from eleven different 2D-PAGE gels representing eleven time point samples and two treatments. (from Chow *et al.*, 1999)

1.8 Objectives

The objectives for this study are as follows:

1. Clone and sequence the gene encoding the *Streptomyces* sp. D7 52 kDa protein induced by vanillic acid (the putative vanillate decarboxylase gene).
2. Sequence regions up- and downstream of the gene encoding the 52 kDa protein. and locate other open reading frames in the immediate vicinity.
3. Compare the putative vanillate decarboxylase gene and its translation product to similar genes and proteins from other microorganisms.
4. Determine if the putative vanillate decarboxylase gene is present in other streptomycetes.
5. Confirm the function of the putative vanillate decarboxylase gene by obtaining a gene knockout mutant of *Streptomyces* sp. D7, expressing the cloned gene in *S. lividans* 1326 as a recombinant host. or by expression of the gene in *E. coli.*
6. Characterize purified vanillate decarboxylase enzyme. if sufficient quantities can be obtained.

2. MATERIALS AND METHODS

2.1 Bacterial strains and plasmids

2.1.1 Bacterial strains

Bacterial strains used in this study are shown in Table 1. *Streptomyces* sp. D7 was isolated from forest soil on the University of British Columbia campus, as described above. *Streptomyces lividans* 1326, which was used for gene expression experiments, was obtained from the John Innes Collection in Norwich, United Kingdom. *Escherichia coli* DH5α was obtained from Gibco BRL and used for general DNA cloning and sequencing procedures. *Escherichia coli* BL21(DE3), which served as host for the T7 RNA polymerase gene expression system, was obtained from Novagen as a component of the pET22b(+) gene expression kit.

2.1.2 Plasmids – cloning and isolation

Plasmids used in this study are listed in Table 1. Subcloning of *Streptomyces* DNA (to be described in detail in Section 2.3) was performed in *Escherichia coli* DH5α with pUC19. Transformation of *E. coli* DH5α was achieved using a heat shock protocol, in which plasmid DNA was incubated with competent cells for 30 minutes on ice, then placed at 37°C for 30 seconds, then another 2 minutes on ice. One milliliter of SOC liquid medium (20 g of Bacto-Tryptone, 5 g of yeast extract, 0.5 g of NaCl, 10 ml of 250 mM KCl, 950 ml of distilled water, pH to 7.0; before use, add 5 ml of 2 M $MgCl_2$ and 20 ml/L of sterile 1 M glucose) was added to the transformation mixture followed by one hour incubation at 37°C. Aliquots were plated on LB agar supplemented with an

antibiotic appropriate for selection of the plasmid being transformed (for example, 100 μg/ml ampicillin for pUC19).

Expression studies of the VDC genes in *Streptomyces lividans* 1326 (Section 2.8) were performed using pIJ680 (Figure 9) (Hopwood *et al.*, 1985), a vector that places target genes under the control of the aminoglycoside phosphotransferase (*aph*) constitutive promoter, and pIJ702 (Figure 10) (Katz *et al.*, 1983), which provides the weaker constitutive tyrosinase (*mel*) promoter. *S. lividans* 1326 was converted to protoplasts and transformed according to published methods (Bibb *et al.*, 1978; Thompson *et al.*, 1982). Protoplasts were plated on R5 solid medium (Thompson *et al.*, 1980) and allowed to regenerate for 14 hours before transformants were selected by an overlay of soft nutrient agar containing thiostrepton to achieve a final concentration of 50 μg ml^{-1} thiostrepton per plate.

Chromosomal DNA was extracted from *Streptomyces* strains by the method of Fisher (Hopwood *et al.*, 1985). *Streptomyces* plasmids were isolated by an alkaline lysis procedure (Hopwood *et al.*, 1985) and *E. coli* plasmid DNA was routinely isolated using the Qiaprep Spin miniprep kit (Qiagen) or the NucleoSpin miniprep kit (Clontech) for sequencing and routine manipulations.

Table 1 (following 2 pages): Bacterial strains and plasmids used in this study. Construction of gene expression plasmids is described in detail in a subsequent section of Materials and Methods.

Strain or plasmid	Relevant properties	Reference/source
E. coli		
DH5α MCR	Host for pUC19 and derivatives	Gibco BRL
Streptomyces		
Streptomyces sp. D7	Wild-type vanillate decarboxylase isolate	Chow, 1996
Streptomyces lividans 1326	Wild-type *Streptomyces* heterologous expression host	John Innes Collection, Norwich
Plasmid vectors		
pUC19	2.7 kb Ap^r *E. coli* cloning vector	Gibco BRL
pKCE1	pUC19 carrying 4.4 kb *Bam*HI sub-fragment of the ~13 kb *Streptomyces* sp. D7 genomic DNA piece cloned from the phage library; contains *vdcB, vdcC, vdcD*	This study; Chow *et al.*, 1999
pKCE2	pUC19 carrying 527 bp *Sal*I sub-fragment of pKCE1 insert (see above)	This study
pKCE3	pUC19 – 3 kb *Bam*HI sub-fragment of phage clone	This study
pKCE4	pUC19 – 2.2 kb *Bam*HI sub-fragment of phage clone	This study
pKCE5	pUC19 – 1.9 kb *Bam*HI sub-fragment of phage clone	This study
pKCE6	pUC19 – 0.8 kb *Bam*HI sub-fragment of phage clone	This study

Table 1 (continued):

Strain or plasmid	**Relevant properties**	**Reference/source**
pIJ702	7.2 kb Ts^r *Streptomyces* cloning vector with *mel* promoter	Katz *et al.*, 1983/ TerraGen Discovery, Inc.
pIJ680	5.3 kb Ts^r *Streptomyces* expression vector with *aph* promoter	Hopwood *et al.*, 1985/ Dr. L. Sandercock, UBC Biotechnology Laboratory
pKCS1	pIJ702 carrying 4.4 kb BamHI insert from pKCE1 inserted in same orientation as *mel* promoter	This study; Chow *et al.*, 1999
pKCS2	pIJ702 carrying 4.4 kb BamHI insert from pKCE1 inserted in opposite orientation as *mel* promoter	This study; Chow *et al.*, 1999
pKCS4	pIJ680 carrying PCR amplified *vdcB* inserted downstream of *aph* promoter	This study
pKCS5	pIJ680 carrying PCR amplified *vdcC* inserted downstream of *aph* promoter	This study
pKCS6	pIJ680 carrying PCR amplified *vdcD* inserted downstream of *aph* promoter	This study
pKCS7	pIJ680 carrying PCR amplified *vdcBC* inserted downstream of *aph* promoter	This study

Table 1 (continued):

Strain or plasmid	**Relevant properties**	**Reference/source**
pKCS8	pIJ680 carrying PCR amplified *vdcCD* inserted downstream of *aph* promoter	This study
pKCS3	pIJ680 carrying PCR amplified *vdcBCD* inserted downstream of *aph* promoter	This study; Chow *et al.*, 1999

2.1.3 16s rDNA sequence-based strain identification

Streptomyces sp. D7 was characterized by sequencing a 505 base pair 16S ribosomal DNA fragment produced using streptomycete specific PCR primers. The primers consisted of the following sequences: forward: 5'-GAGATTTGATCCTGGCTCAG-3'; reverse: 5'-CGGACTGGTTGTTACGACTTC-3'. Thermocycling was performed as follows: 1 minute denaturation at 95°C, 2 minutes annealing at 55°C and 2 minutes extension at 72°C. The cycle was repeated 30 times, with a final extension of 10 minutes at 72°C.

2.2 Media and growth conditions

Streptomyces sp. D7 and *S. lividans* 1326 were routinely cultivated in tryptic soy broth (TSB) or on mannitol soy flour agar plates at 30°C. Catabolic tests and growth experiments were performed using mineral salts medium supplemented with 0.5% yeast extract (MSMYE: $(NH_4)_2SO_4$ 0.1 g L^{-1}, NaCl 0.1 g L^{-1}, $MgSO_4 \cdot 7H_2O$ 0.2 g L^{-1}, $CaCl_2$ 0.01 g L^{-1}, yeast extract 0.5 g L^{-1}, K_2HPO_4 1.0 g L^{-1}, KH_2PO_4 0.5 g L^{-1}, pH 7.2) and aromatic compounds of interest at concentrations of 3.6 mM to 6 mM. When appropriate, thiostrepton at 50 μg ml^{-1} was included for selection and maintenance of plasmid containing strains. For DNA extraction or protoplast preparation, strains were cultivated in YEME (liquid medium) supplemented with 0.5% glycine and 5 mM $MgCl_2$ at 30°C (Hopwood *et al.*, 1985). *E. coli* DH5α was grown in Luria-Bertani (LB) medium (supplemented with 100 μg ampicillin ml^{-1}), when maintaining pUC-based plasmids) at 37°C.

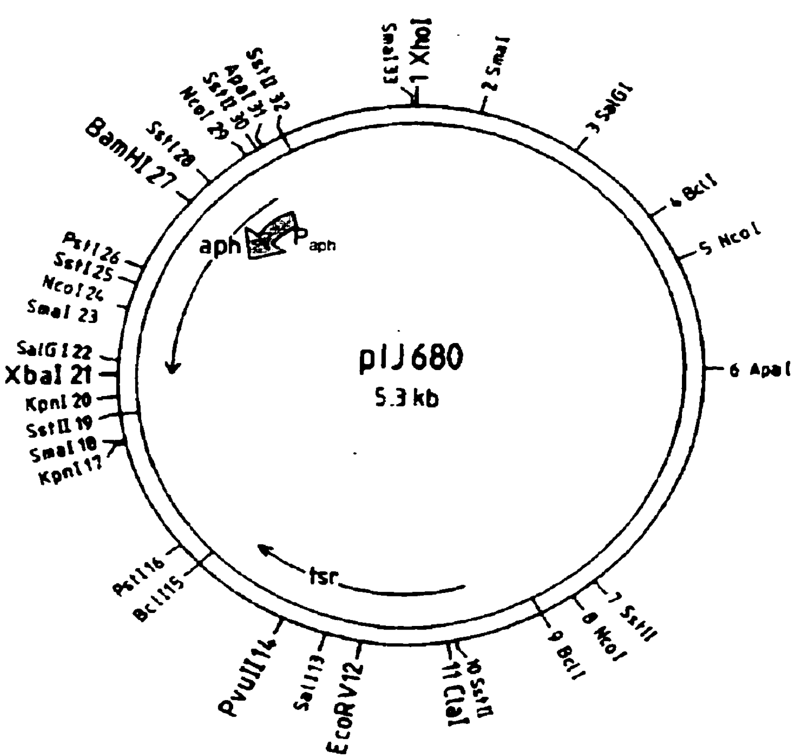


Figure 9: Map of expression vector pIJ680 (Hopwood *et al.*, 1985)

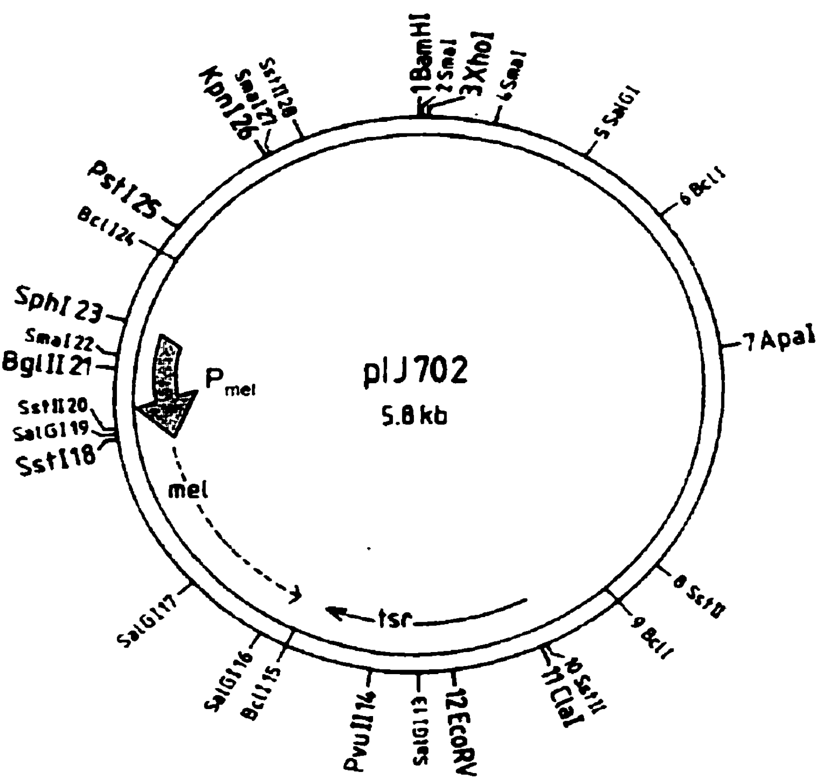


Figure 10: Map of cloning/expression vector pIJ702 (Hopwood *et al.*, 1985)

2.3 Library construction and gene cloning

A Lambda DASH II (Stratagene) genomic DNA phage library of chromosomal *Streptomyces* sp. D7 fragments was constructed by ligating 9-22 kb *Sau*3AI partially digested chromosomal DNA fragments into the *Bam*HI site of the phage arms. Phage carrying genomic DNA fragments were mixed, in soft nutrient agar, with *Escherichia coli* XL1-Blue MRA(P2) cells and plated on NZY agar plates. The plates were incubated at 37°C overnight and the resulting plaques (approximately 4000 - between 300 and 400 plaques per plate) were lifted with Hybond-N nylon membranes (Amersham). The library was screened by hybridization at 60°C with a γ^{32}P-ATP-labeled 56-mer oligonucleotide probe 3717C (5'- GC(CG) TAC GAC GAC CT(CG) CG(CG) TAC TTC CT(CG) GAC AC(CG) CT(CG) GAG AAG GAG GG(CG) CAG CT(CG) CT -3') derived from protein amino-terminal sequencing data. Of the approximately 4000 plaques in the library, twelve hybridized strongly to the probe. Lambda phage from one of these plaques were isolated and propagated in XL1-Blue MRA(P2) *E. coli* for DNA isolation (see Section 2.3.1). The *vdcB*, *vdcC* and *vdcD* genes were subcloned on a 4.4 kb *Bam*HI fragment into pUC19 in *Escherichia coli* DH5α MCR (Gibco BRL).

2.3.1 Lambda phage DNA preparation

Lambda DASH II phage DNA carrying *Streptomyces* sp. D7 chromosomal DNA *Bam*HI fragments was isolated by the following protocol. Ten milliliters of LB liquid medium was inoculated with one agar plug (from the NZY/soft agar/phage plates (described above) containing a phage plaque), 50 µL of XL1-Blue MRA(P2) *E. coli* (in 0.01 M $MgSO_4$, O.D.$_{600}$ ~0.5), 100 µL of 1 M $MgSO_4$, and shaken at 37°C overnight. After the

incubation period, 100 μL of chloroform was added and the culture was shaken for an additional 2 minutes at 37°C, then centrifuged at room temperature for 10 minutes at 5000 x g. The aqueous phase was saved, and the following were added: 100 μL 1 M $MgSO_4$, 10 ml TM buffer (50 mM Tris-HCl pH 7.4, 10 mM $MgSO_4$), 32 μL 10 mg/ml DNAse and 10 mg/ml RNAse. The solution was incubated at room temperature for 15 minutes, then 2 ml of 5 M NaCl and 2.2 g of PEG (m.w. 6000-8000) were added and allowed to dissolve completely. The mixture was the incubated for 15 minutes on ice, then centrifuged for 10 minutes at 4°C at 10000 x g. The supernatant was discarded, leaving the phage pellet, which was resuspended in a minimum of 300 μL of TM buffer. The suspension was transferred to a 1.8 ml Eppendorf tube, 300 μL of chloroform was added, mixed, then the tube was centrifuged for 5 minutes at 12000 x g. The aqueous phase was transferred to a new Eppendorf tube, and the chloroform extraction step was repeated until no interface was observed between the chloroform and aqueous phases. Fifteen microliters of 0.5 M EDTA, 30 μL of 5 M NaCl, and 350 μL of tris-buffered phenol was added, vortexed, then centrifuged for 5 minutes at 12000 x g and the aqueous phase was saved. One final chloroform extraction (350 μL) was performed, then 875 μL of 100% cold ethanol was added and incubated at -20°C overnight to precipitate the DNA. After centrifugation at 12000 x g, the ethanol was removed, the DNA pellet was air dried briefly, and resuspended in 50 to 100 μL of TE buffer.

2.4 DNA sequencing and analysis.

Automated DNA sequencing was performed using the AmpliTaq PRISM kit (Applied Biosystems) with a standard thermocycling program provided by the manufacturer, with

variations in the annealing temperature to match the melting temperature of the sequencing primer being used. Sequencing reactions were carried out by the Nucleic Acid Protein Sequencing (NAPS) Unit at the University of British Columbia and electrophoresed on an ABI Model 377 DNA sequencing apparatus (Applied Biosystems). Nucleic acid sequence was analyzed by the Wisconsin Package Version 10 (Genetics Computer Group) on a Sun Microsystems SparcStation5 (Sun Microsystems).

2.5 Southern blotting and hybridization

Genomic DNA preparations from *Streptomyces* sp. D7 and *Streptomyces setonii* 75Vi2 were digested with the restriction endonuclease *Sal*I overnight at 37°C and electrophoresed in a 0.7% agarose gel. The gel was photographed, soaked in 0.15N HCl for 15 minutes for depurination (to facilitate transfer of high molecular weight fragments) and soaked in alkaline hybridization solution (0.6M NaCl, 0.4M NaOH) for an additional 15 minutes. The treated gel was then Southern blotted to a positively charged nylon membrane (Boehringer Mannheim) using the same alkaline solution as a transfer buffer. After disassembly of the transfer apparatus, the membrane was rinsed briefly in 2X sodium saline citrate (SSC) and baked at 80°C for one hour. Hybridizations utilized a rotating incubation chamber (Hybaid) and glass hybridization bottles. The membrane was soaked in 2X SSC, rolled, placed in a glass hybridization bottle with hybridization solution (5X Denhardt's Solution, 6X SSPE, 0.1% SDS) and prehybridized at 65°C for several hours. After prehybridization, the ^{32}P-labeled probe (α-^{32}P or γ-^{32}P, depending on the labeling procedure) was denatured at 95°C for 5 minutes, cooled on ice briefly, centrifuged, then added to the hybridization bottle containing the membrane.

Hybridization was carried out overnight at 65°C. The wash regimen consisted of 2 washes of 2X SSC, 0.1% SDS for 10 minutes at room temperature, 1X SSC, 0.1% SDS for 15 minutes at 65°C once, then a brief room temperature rinse in 0.1X SSC, 0.1% SDS. The washed blot was semi-dried, then placed in plastic wrap and exposed to film (Kodak X-AR) for several hours to overnight. An identical hybridization procedure was used during the phage library screening process, and also subsequent cloning manipulations.

2.6 Chemical analyses

Culture supernatants were filtered through 0.45 μm syringe filters and processed through a C-18 hydrophobic interaction column attached to a HPLC system (Hewlett Packard, Model 1050). Conditions for separation were 30% phosphoric acid/water, 70% methanol, with a flow rate of 1.0 ml min^{-1}. Retention times for vanillic acid and guaiacol under these conditions are 5 minutes and 4 minutes, respectively. Integrated peak areas corresponding to compounds in supernatant samples were calibrated against known concentrations of vanillic acid and guaiacol standards. Additional analysis of supernatant samples was performed using a Cary 1 Bio ultraviolet/visible spectrophotometer (Varian). For UV analysis, vanillic acid characteristically displays a primary absorbance at 250 nm and a secondary absorbance at 285 nm, while guaiacol absorbs at 275 nm.

2.7 RNA isolation and analysis

Total RNA was isolated from cells grown under two different sets of conditions. Primary cultures were grown in 25 ml YEME for 48 hours before cells were pelleted by

centrifugation, washed twice with sterile water and resuspended in minimal media (MSMYE). For induced cultures, the media was supplemented with 3.6 mM vanillic acid, while no additional substrates were added to the uninduced control. These cultures were then allowed to grow an additional three hours after which cells were pelleted and washed as before. Total RNA was isolated using standard RNA isolation techniques (Hopwood *et al.*, 1985; Kirby *et al.*, 1967). For transcript detection, Northern gel electrophoresis and transfer were performed according to the manufacturer's recommendations for the NorthernMax kit (Ambion). 15 μg total RNA was loaded per lane in a polyacrylamide gel and 1 μg of an RNA standard ladder (NEB) was included for size comparison. After electrophoresis was complete, the RNA ladder lane was excised and stained with ethidium bromide to allow for visualization and to confirm RNA integrity. Transcript was detected using a probe specific for the *vdc*C gene. To generate the probe, traditional double-stranded PCR was performed on pKCE1 using oligonucleotides vdcC.F and vdcC.R (Table 2). Following amplification, excess dNTPs and oligonucleotides were removed using a QIAquick PCR Purification Kit (Qiagen). This product was then used as template for asymmetric PCR with only the vdcC.R primer. This resulted in a single-stranded PCR product that was complementary to the predicted RNA transcript. During chain elongation, ^{32}P-dCTP was provided in place of dCTP in the dNTP mix to allow for direct incorporation of radiolabel. The extension product was purified and allowed to hybridize with the immobilized RNA at 60°C for 24 hours. Excess probe was removed by washing as directed by the NorthernMax protocol, and the hybridizing transcript was visualized by exposure to autoradiography film (Kodak XAR) for exposure and development.

Table 2 (including next page): Polymerase chain reaction (PCR) primers used in this study. Engineered restriction enzyme sites are underlined.

Primer	Length (nt)	Sequence (5' → 3')	Purpose
Hybridiza-tion probes			
vdcB.F	26	ACAGGTCAGCGACAGG TTTGAGGTGG	Forward amplification of *vdcB* gene DNA.
vdcB.R	21	TACGGGGCAGGGGACT TCAGG	Reverse amplification of *vdcB* gene DNA.
vdcC.F	20	GGCGACGCCGCCTGAA GTCC	Forward amplification of *vdcC* gene DNA.
vdcC.R	20	GGGTCGGTCGGTGTCA GACG	Reverse amplification of *vdcC* gene DNA.
vdcD.F	20	CACCGATCCTCACTGA AAGG	Forward amplification of *vdcD* gene DNA.
vdcD.R	20	CATAGACCGCGTGCCG GTCG	Reverse amplification of *vdcD* gene DNA.
pIJ680 expression			
vdcBCD.FX	26	CGGATCCAGTGACAGG TTTGAGGTGG	Forward amplification of *vdcB*, *vdcBC* and *vdcBCD* genes for expression in pIJ680. Engineered *Bam*HI site.
vdcBCD.RX	28	AGTCTAGACCGGCGTC GGAGGGATGACC	Reverse amplification of *vdcD*, *vdcCD* and *vdcBCD* genes for expression in pIJ680. Engineered *Xba*I site.
vdcB.RX	21	TTCTAGACAGGGGACT TCAGG	Reverse amplification of *vdcB* gene for expression in pIJ680. Engineered *Xba*I site.
vdcC.FX	29	CGGATCCCCCGTAAAG GAATTCACCATGG	Forward amplification for *vdcC* gene and *vdcCD* genes for expression in pIJ680. Engineered *Bam*HI site.

vdcBC.RX	29	TTCTAGATCAGACGCG GGCCGCGATCAGG	Reverse amplification for *vdcB*, *vdcC* and *vdcBC* genes for expression in pIJ680. Engineered *Xba*I site.
vdcD.FX	28	CACGGATCCTCACTGA AAGGACAACTCC	Forward amplification of *vdcD* gene for expression in pIJ680. Engineered *Bam*HI site.
pET22b(+) expression			
vdcB.pETF	21	TCATATGCGGTTGGTCG TGGG	Forward amplification of *vdcB* gene for expression in pET22b(+). Engineered *Nde*I site.
vdcB.pETR	21	TTCTCGAGAGGGGACT TCAGG	Reverse amplification of *vdcB* gene for expression in pET22b(+). Engineered *Xho*I site.
vdcC.pETF	23	GGAATTCCATATGGCCT ATGACG	Forward amplification of *vdcC* gene for expression in pET22b(+). Engineered *Nde*I site.
vdcC.pETR	18	CGCGGGCTCGAGTCAG GC	Reverse amplification of *vdcC* gene for expression in pET22b(+). Engineered *Xho*I site.

2.8 Gene expression

2.8.1 Cloning and expression in pET22b(+) in *Escherichia coli* BL21

DNA encoding the *vdcB* and *vdcC* genes was PCR-amplified, using primers listed in Table 2, to incorporate *Nde*I and *Xho*I sites upstream and downstream, respectively of the target gene(s) prior to cloning into the pET22b(+) vector (Novagen). These plasmid constructs were transformed into *E. coli* BL21(DE3), which were grown to OD 0.4, then induced with IPTG to activate expression by the T7 polymerase system, characteristic of the pET22b(+) vector. Expression was performed for two hours, at which time the cells were pelleted and protein extracted. To optimize protein expression, induction of the T7 system was performed at 37°C, 30°C and 25°C, using 0.1 mM or 1 mM IPTG. Reductions in temperature and IPTG concentration are commonly used methods to increase the chances of obtaining properly folded proteins.

2.8.2 Cloning and expression in pIJ702 in *S. lividans* 1326

The 4.4 kb *Bam*HI DNA fragment containing the *vdcBCD* gene cluster was inserted into the *Streptomyces* cloning vector pIJ702 at the unique *Bgl*II site. Insertion at this site places the cluster downstream of the *mel* promoter, thereby disrupting transcription of the tyrosinase gene that serves as a color selection marker for transformants. pIJ702 carrying the insert in the same orientation as the *mel* promoter was designated pKCS1; conversely, a vector construct with the insert in the opposite orientation to the promoter was designated pKCS2 (Figure 11).

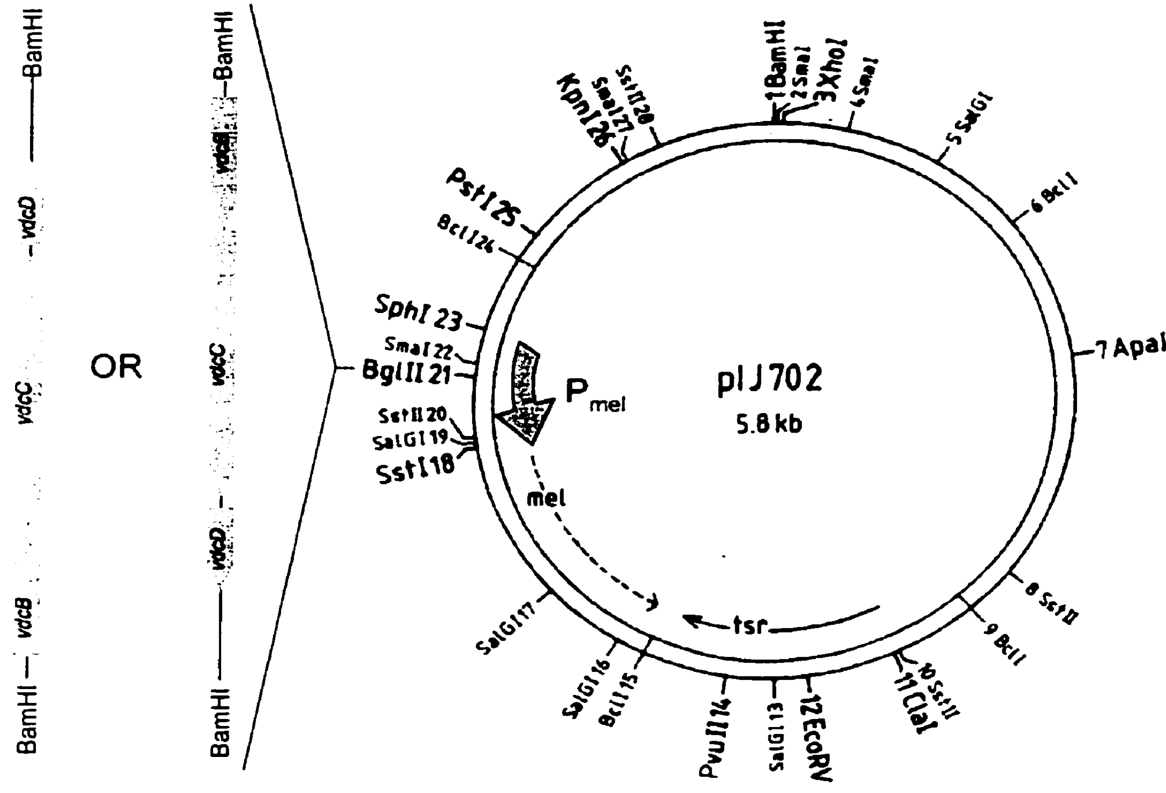


Figure 11: pIJ702 and the 4 kb BamHI *Streptomyces* sp. D7 genomic DNA fragment inserted in both orientations in the tyrosinase gene (mel) promoter.

2.8.3 Cloning and expression in pIJ680 in *S. lividans* 1326

In order to identify the gene sequences encoding the decarboxylase, DNA regions were PCR amplified using specific primers (Table 2) which included a BamHI site upstream, and a *Xba*I site downstream, of the gene(s). The PCR-generated genes were cloned downstream of the *aph* promoter in *Bam*HI-*Xba*I cut pIJ680, replacing most of the *aph* gene. The following genes were amplified using this *Bam*HI-*Xba*I PCR cloning strategy for ligation into pIJ680, creating new plasmids designated in brackets: *vdcB* (pKCS4), *vdcC* (pKCS5), *vdcD* (pKCS6), *vdcBC* (pKCS7), *vdcCD* (pKCS8), *vdcBCD* (pKCS3). pIJ680 and the PCR-generated inserts are shown in Figure 12. All plasmids were transformed individually into *S. lividans* 1326 and the resulting recombinant hosts were screened for the ability to decarboxylate vanillic acid.

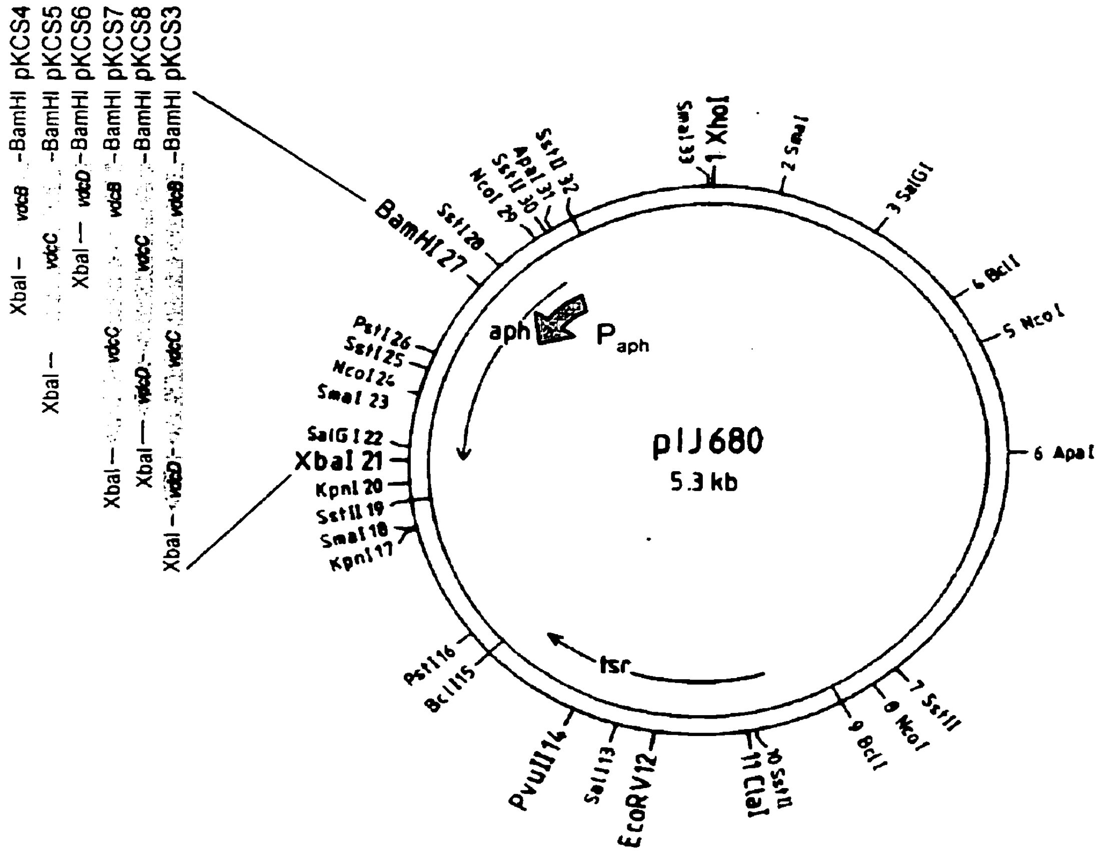


Figure 12: pIJ680 and the PCR-generated gene combinations inserted downstream of the promoter for the aminoglycoside phosphotransferase (aph) gene.

2.9 Enzyme assays

Late log phase mycelia harvested from YEME cultures were washed with phosphate buffer (10 mM Tris pH 8.0, 1 mM EDTA, 1 mM DTT) and resuspended in the same buffer, containing protease inhibitors (100 μg ml^{-1} PMSF, 1 μg ml^{-1} Pepstatin A), at a ratio of 0.1 ml buffer per 1 ml of original culture. Samples were sonicated as previously described and centrifuged at 12,000 x g for 15 minutes to remove insoluble cell debris. Soluble cell extracts were tested for decarboxylase activity by adding vanillic acid or comparative substrates to a final concentration of 1 mM. Samples were incubated for 15 minutes at 25°C, at which time they were analyzed by scanning the UV range from 300 nm to 200 nm. Soluble cell extract without substrate was used as a background in the reference cuvette. To test enzyme activity under anaerobic conditions, nitrogen gas was slowly bubbled through assay sample tubes prior to addition of substrate.

2.10 Recombinant protein purification from *E. coli* BL21(DE3)

As mentioned in Section 2.8.1, *vdcB* and *vdcC* were recombinantly expressed using the pET22b(+) plasmid vector in *E. coli* BL21(DE3). Attempts were made to purify the proteins produced by these strains. Under all expression conditions, inclusion body material was obtained, and thus efforts were mainly focused on extracting and refolding the insoluble protein aggregates. However, in the event that minute amounts of soluble, active enzyme were produced, cell lysates were passed through nickel-based affinity columns to purify the histidine-tagged recombinant proteins.

2.10.1 Lysate preparation

Ten milliliters of LB medium containing 100 µg/mL ampicillin were inoculated with material from one colony of the appropriate *E. coli* BL21(DE3) expression host and grown overnight at 37°C with shaking. Subsequently, 50 ml of prewarmed media containing 100 µg ml^{-1} ampicillin was inoculated with 2.5 ml of the overnight culture and grown at 37°C, with shaking, until the OD_{600} reached approximately 0.6 (about one hour). IPTG was added to a final concentration of either 1 mM or 0.1 mM, and cultures were grown for an additional 4-5 hours at 37°C, 30°C, or 25°C, depending on the desired expression conditions. As has already been mentioned, reduction of IPTG concentration and growth temperature is believed to reduce the amount of inclusion body formation due a slowing of the expression process. After the expression period, cells were harvested by centrifugation at 4000 x g for 20 minutes. The cell pellet was resuspended in 5 mL of lysis buffer (50 mM NaH_2PO_4, pH 8.0; 300 mM NaCl; 10 mM imidazole) and subjected to a freeze/thaw cycle using dry ice/ethanol and cold water, repeated three times. The sample was then sonicated with a microtip 6 times, on ice, at 10 seconds per burst with 10 second pauses, at 200-300 watts. The lysate was then centrifuged at 10000 x g at 4°C for 30 minutes. At this point, the supernatant contained any soluble recombinant protein, while the pellet contained the large mass of inclusion body material, visible as a white layer on top of the light yellow cell debris.

2.10.2 Inclusion body protein purification, solubilization and refolding

The pellet was resuspended in 0.1 culture volume of inclusion body wash buffer (200 mM Tris-HCl, pH 7.5, 100 mM EDTA, 10% Triton X-100) and recentrifuged. This wash

step was repeated, and the resulting purified inclusion bodies (of which the wet weight was noted) were solubilized in inclusion body solubilization buffer (500 mM CAPS, pH 11.0, 0.3% N-laurosarcosine) at 10-20 mg ml^{-1} for 15 minutes at room temperature. The (mostly) solubilized solution was centrifuged at 10000 x g for 10 minutes at room temperature, and the supernatant was removed to a clean tube. The soluble protein solution was placed in dialysis tubing, and the sample was dialyzed against 20 mM Tris-HCl, pH 8.5, 0.1 mM DTT, using at least three buffer changes of greater than 50 times the sample volume. Each dialysis step was performed for at least 3 hours at 4°C. After dialysis, the sample was concentrated using a Centricon-10, 10000 m.w. cut-off ultrafiltration device and analyzed by SDS-PAGE for purity.

2.10.3 Soluble protein purification procedure

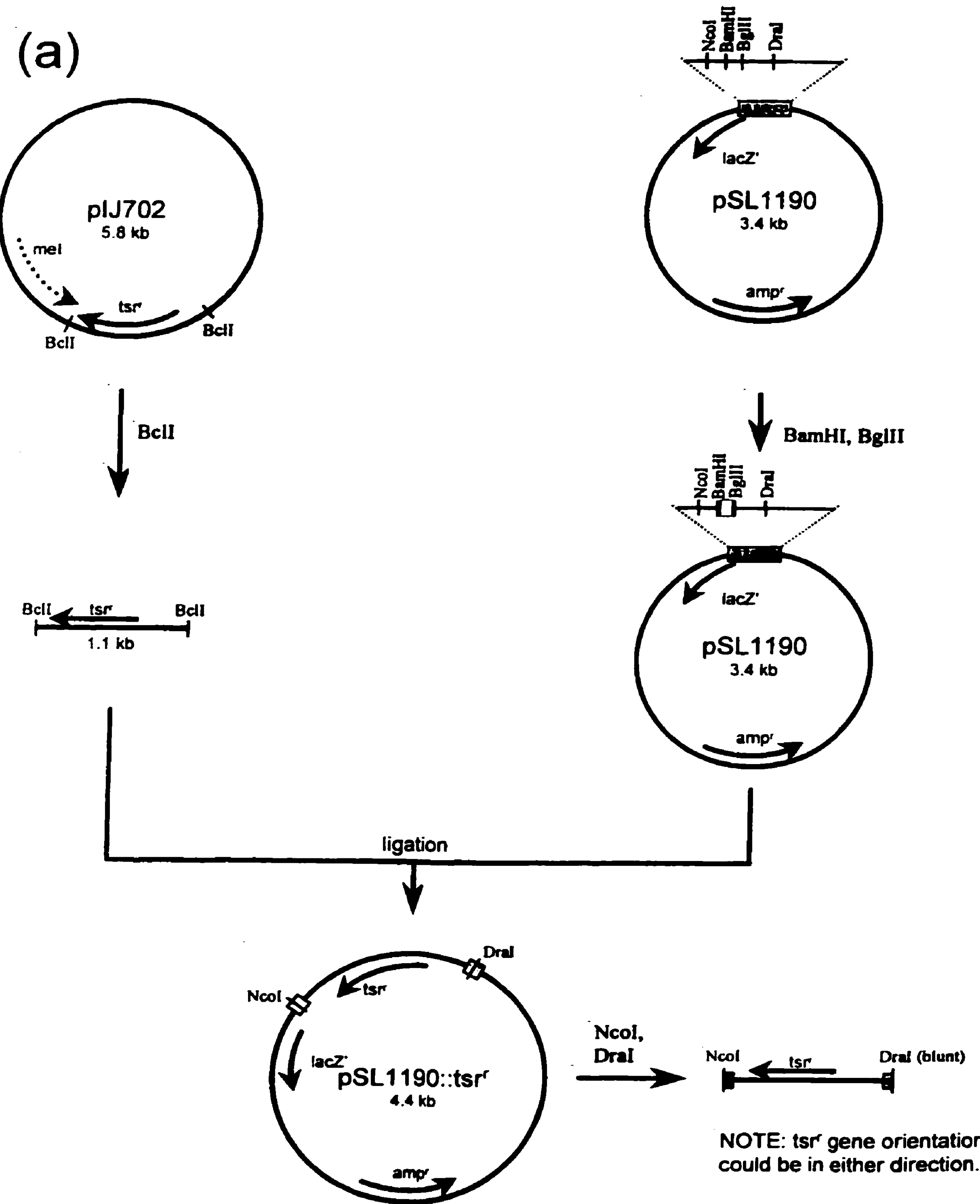
The soluble fraction of the recombinant *E. coli* BL21(DE3) cell lysates were processed through nickel-nitrilotriacetic acid (Ni-NTA) metal-affinity chromatography matrix columns in an attempt to purify any proteins that were not produced as inclusion bodies. Ni-NTA mini spin columns (Qiagen, catalogue #31313) were used for small-scale experiments, while nickel resin (ProBond, Invitrogen, catalogue #R801-01) was packed in a larger column for scaled-up procedures. Column loading, washing and elution were performed according to the manufacturers' instructions.

2.11 Construction of a knockout allele of the *vdcC* gene

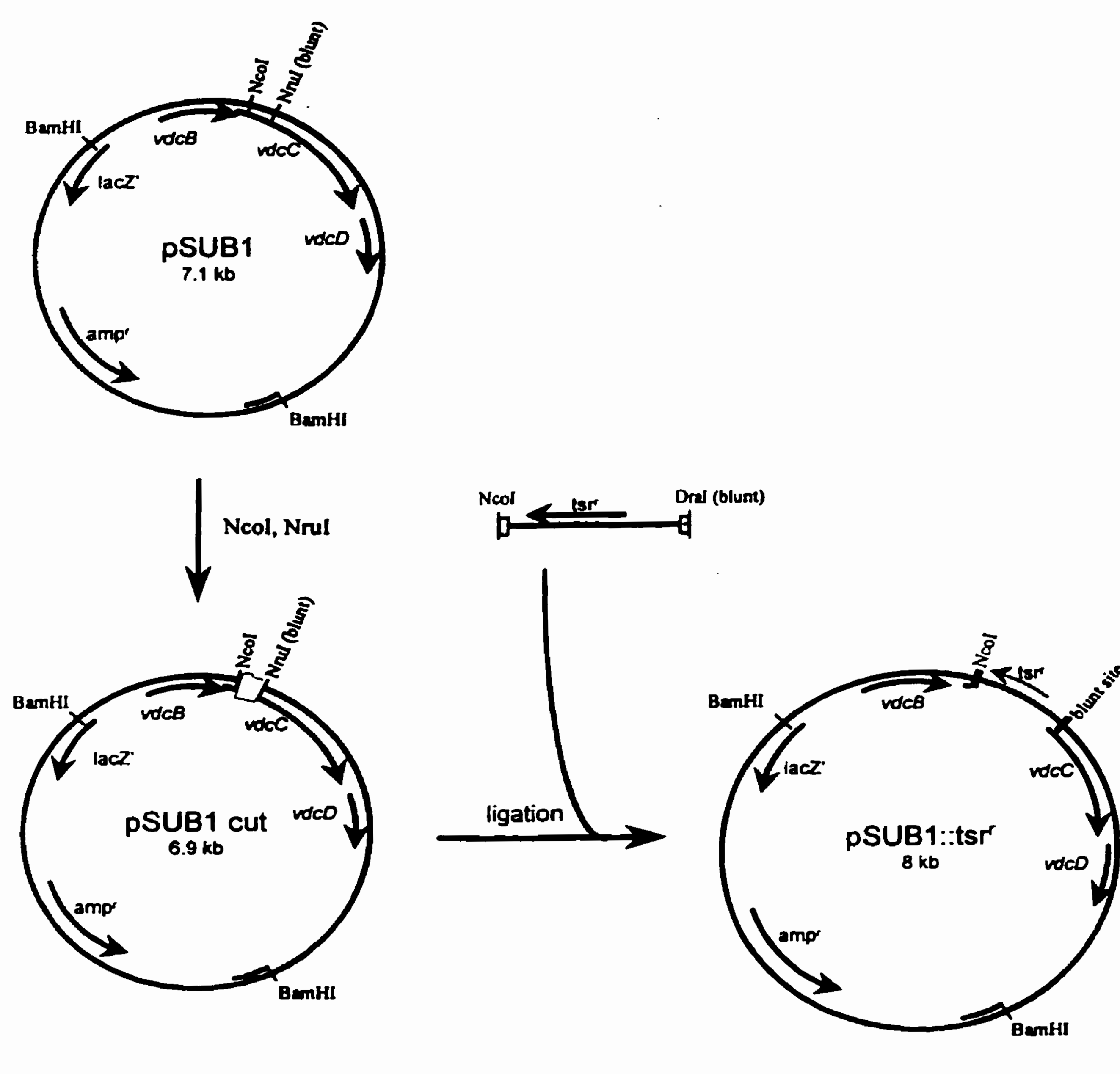
To elucidate the function of *vdcC*, the gene encoding the 52 kDa protein, a disrupted allele was created by removing a portion of the 5' region of the gene and replacing it with

the thiostrepton resistance gene (*tsr*). This strategy utilized pSL1190 (Pharmacia), which contains a superlinker with 78 restriction enzyme sites, allowing for more versatility in the construction process. The allele was designed for gene knockout experiments, which are described in subsequent sections. The allele construction strategy, using pIJ702, pSL1190 and pUC19, is shown in Figure 13.

Figure 13: Strategy for a knockout of the *vdcC* gene using the thiostrepton (tsr) resistance gene. (a) Preparation of the thiostrepton resistance marker cassette; (b) Insertion of the thiostrepton resistance cassette in *vdcC* to create the knockout allele; (c) protoplast transformation and double crossover event.


(a)
pIJ702
5.8 kb
mel
tsr
BclI
BclI
BclI
BclI
tsr
BclI
1.1 kb
NcoI
BamHI
BglII
DraI
lacZ'
pSL1190
3.4 kb
amp
BamHI, BglII
NcoI
BamHI
BglII
DraI
lacZ'
pSL1190
3.4 kb
amp
ligation
DraI
NcoI
tsr
lacZ'
pSL1190::tsr
4.4 kb
amp
NcoI,
DraI
NcoI
tsr
DraI (blunt)
NOTE: tsr gene orientation
could be in either direction.

(b)



KNOCKOUT ALLELE

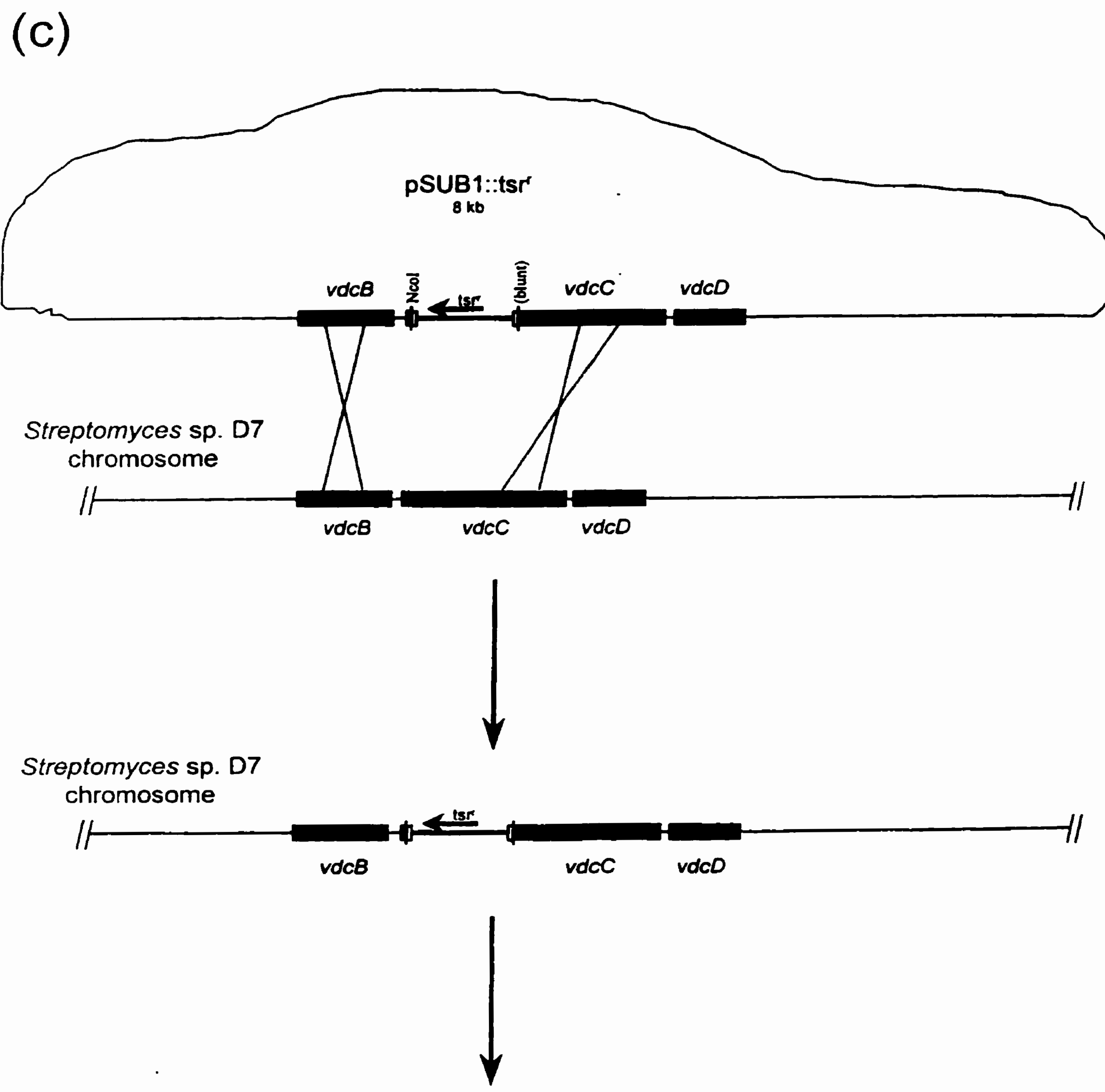


PLATE PROTOPLASTS ON REGENERATIVE MEDIA, SELECT FOR THIOSTREPTON RESISTANCE

2.12 *Streptomyces* sp. D7 protoplast preparation and transformation

Streptomyces sp. D7 protoplasts were prepared and transformed as described for *Streptomyces lividans* 1326 in Section 2.1.2, with the exception that R6 agar plates were used instead of R5. The regeneration time for *Streptomyces* sp. D7 was several days slower than that observed for *S. lividans* 1326.

The recipe for R6 solid regeneration agar medium is as follows (grams per liter): mix and autoclave 200 g sucrose, 10 g dextrose, 1 g casamino acids, 0.05 g $MgSO_4 \cdot 7H_2O$, 0.1 g K_2SO_4, 20 g agar. After autoclaving, add 11 g glutamate, 1 mL standard trace elements solution, 7 g $CaCl_2 \cdot 2H_2O$ and 100 mL MOPS buffer (0.1 M, pH 7.2).

2.13 *E. coli* – *Streptomyces* interspecies conjugation

To obtain a gene knockout mutant, attempts were made to transform *Streptomyces* sp. D7 mycelia by conjugation with *E. coli*. As an evaluation of this method, *E. coli* S17-1 carrying the vector pPM801 (aph^+, tra^+) was mated with *Streptomyces* sp. D7 mycelia to assess whether the vector, which encodes resistance to the antibiotic kanamycin, was transferred between the species.

2.13.1 *E. coli* S17-1/pPM801 preparation

One colony of *E. coli* S17-1/pPM801 was diluted in 5 mL of LB broth, supplemented with 10 μg/mL kanamycin, and incubated for 8 hours at 37°C. The grown culture was diluted 1:10 with fresh LB broth with kanamycin.

2.13.2 *Streptomyces* sp. D7 preparation and mating

Streptomyces sp. D7 spores were scraped from a lawn of bacteria on a mannitol-soy agar plate and resuspended in 5 mL of TES buffer (0.05 M, pH 7.2). The suspension was heat-shocked at 50°C for 10 minutes, then placed under cold tap water to stop the shock process. The spores were then centrifuged at 3000 x g and resuspended in 1 mL of LB broth. 150 µL of the spore suspension was plated on an AS-1 agar plate (Hopwood *et al.*, 1985) and the spores were allowed to grow for 2 to 4 hours at 30°C. Subsequently, 100 µL of the *E. coli* S17-1/pPM801 preparation (see above) was plated on top of the *Streptomyces* sp. D7 spores. After overnight growth of the *E. coli* layer, the layer was gently scraped off using a sterile glass spreader and 0.05 M TES buffer, pH 7.2. An antibiotic overlay, consisting of 2.5 mL of 50 µg/mL nalidixic acid and 25 µg/mL kanamycin, was spread over each plate. Plates were allowed to incubate for a further 6 days, at which time they were observed for any ex-conjugants.

2.14 Reagents and enzymes

All reagents used were of the highest quality, and purchased from Sigma Chemical Company unless noted otherwise. Restriction endonucleases and other modification enzymes were obtained from Gibco BRL, New England Biolabs, Boehringer Mannheim or Pharmacia.

3. RESULTS

3.1 16s rDNA sequence identification

Sequence of the 505 bp PCR product amplified from *Streptomyces* sp. D7 16s rDNA was matched against the GenBank database using the BLAST-N program (Altschul *et al.*, 1990). The result was an exact match (data not shown) with the 16s rDNA sequence from *Streptomyces* sp. strain B71277, an isolate from the BBSRC Institute for Food Research in Reading, United Kingdom (Hutson & Collins, 1997). Studies, if any, involving *Streptomyces* sp. strain B71277 remain unpublished at the time of writing this thesis, and attempts to contact the authors of the GenBank submission failed to generate a response. This identification procedure confirmed *Streptomyces* sp. D7 as a member of the streptomycetales. Previously, the microorganism had been referred to as *Streptomyces violaceusniger*, as identified by fatty acid methyl ester analysis (Chow, M.Sc. thesis, 1996). However, the results of the 16s rDNA analysis, considered a more powerful tool for taxonomy (J. Davies, personal communication), suggest otherwise.

3.1 Cloning of the VDC gene cluster

Edman degradation sequencing of protein 3717, isolated from protein 2D-PAGE gels (previous work described in Section 1.6), yielded the following N-terminal amino acid sequence: AYDDLRYFLDTLEKEGQLLRIT. This sequence matched well with the amino-terminal sequence of *p*-hydroxybenzoate carboxy-lyase from the anaerobe, *Clostridium hydroxybenzoicum* (He and Wiegel, 1995) as shown in Figure 14. The deduced degenerate oligonucleotide probe 3717C (for sequence, see Materials and

Methods) was synthesized and hybridized against a lambda DASH II phage library (Stratagene) of *Streptomyces* D7 genomic DNA. A phage clone, designated 3717C(+), hybridized strongly to the probe (Figure 15). DNA from phage clone 3717C(+) was purified and digested with the restriction enzyme *Bam*HI (Figure 16). This restriction digest revealed that the phage clone carried an approximately 13 kb *Streptomyces* sp. D7 genomic DNA insert. *Streptomyces* genomic DNA *Bam*HI digestion products of approximately 4.4 kb, 3.0 kb, 2.2 kb, 1.9 kb and <1 kb were subcloned individually into pUC19 for further manipulations and sequencing (Figure 17). The pUC19 subclones were designated pKCE1 through pKCE6 (Figure 18). pKCE1 contains the 4.4 kb *Bam*HI fragment encoding the vanillate decarboxylase gene cluster, while pKCE2 contains a 527 bp *Sal*I fragment that encodes the amino-terminal region of the *vdcC* gene.The entire sequence of the 4.4 kb *Streptomyces* sp. D7 DNA fragment (GenBank accession number AF134589) is shown in Figure 19.

Sequence analysis revealed that the gene encoding protein 3717 was contained on a 4.4 kb *Bam*HI fragment, and was determined to be the second gene in a cluster of at least three genes, designated *vdcB* (602 bp), *vdcC* (1424 bp) and *vdcD* (239 bp), as depicted in Figure 20. BLAST-X sequence analyses (Altschul *et al.*, 1990) revealed that the gene cluster, in whole or in part, is present in a variety of microorganisms. The *vdcB* translation product is highly similar to phenylacrylate decarboxylase (PAD) from *Saccharomyces cerevisiae*. The yeast PAD contains a putative trans-membrane domain close to the amino-terminus (annotated in GenBank accession number S62017), which is highly conserved among other hypothetical PAD homologues from various

microorganisms, as revealed by genome projects (this region is highlighted in the amino acid sequence alignments shown in Figure 21). The *vdcC* translation product is also highly similar to hypothetical proteins identified in various microbial genome sequencing projects, in addition to the amino-terminal similarity to 4-hydroxybenzoate carboxy-lyase as revealed from the Edman degradation sequencing of protein 3717. Dendrograms of the *vdcB* and *vdcC* translation products in comparison to other microbial homologues are shown in Figure 21. Unlike the first two genes in the cluster, the *vdcD* translation product shows similarity only to a hypothetical protein from *Bacillus subtilis*. Although these genes have homologues in a number of microbial genomes, they are not always clustered and only *Bacillus subtilis* contains all three genes in the same order as *Streptomyces* sp. D7[*]. Other microorganisms contain *vdcB* and *vdcC* homologues, but at different chromosomal locations. Interestingly, *Sphingomonas aromaticivorans* strain F199 possesses homologues to *vdcB* and *vdcC* and *vdcD* on its 184 kb catabolic plasmid pNL1 (Romine *et al.*, 1999, GenBank accession AF079317). Plasmid pNL1 contains a variety of genes encoding enzymes for the degradation of a number of toxic organic chemicals, and among these genes lie (in order; identifications by similarity only): *orf1244* (*vdcB*, phenylacrylate decarboxylase), *pchFa* (*p*-cresol methylhydroxylase), *vdh* (vanillin oxidoreductase), *orf1272* (*vdcC*), *orf1280* (*vdcD*).

[*] Interestingly, the *Bacillus subtilis* strain did not metabolize vanillic acid.

```
SD7   2   YDDLRYFLDTLEKEGQLL   19
          | ||| ||+ | + | |+
CHB   6   YRDLREFLEVLXQXGXLI   23
```

Figure 14: Amino-terminal sequence alignment of *Streptomyces* sp. D7 protein 3717 (SD7) and *Clostridium hydroxybenzoicum p*-hydroxybenzoate carboxy-lyase (CHB) (He and Wiegel, 1995). (|) = identity; (+) = similarity. Note that three amino acid residues of the *C. hydroxybenzoicum* enzyme were not identified (marked as 'X').

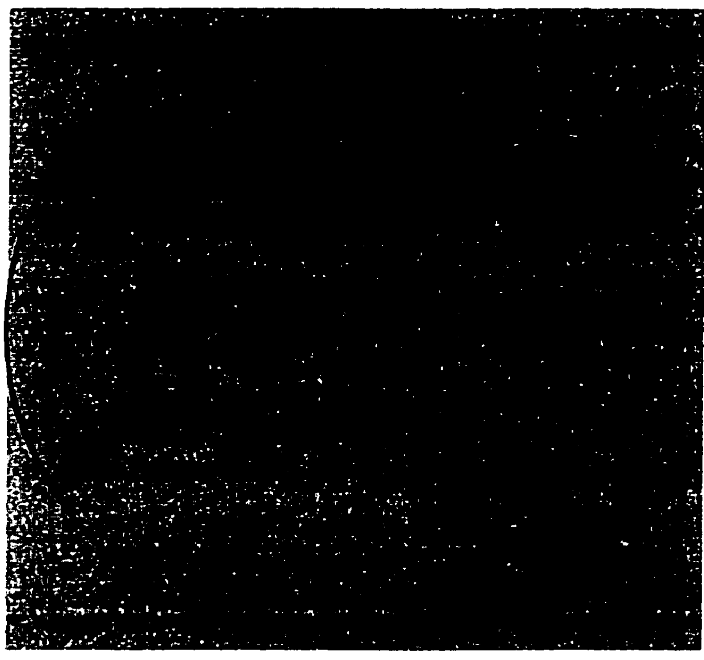


Figure 15: Lambda DASH II phage clone carrying the putative vanillate decarboxylase gene. Phage plaque lift was hybridized to ^{32}P-labeled oligonucleotide probe 3717C and exposed to X-ray film for several hours with an intensifying screen. An arrow indicates the position of the hybridizing phage clone.

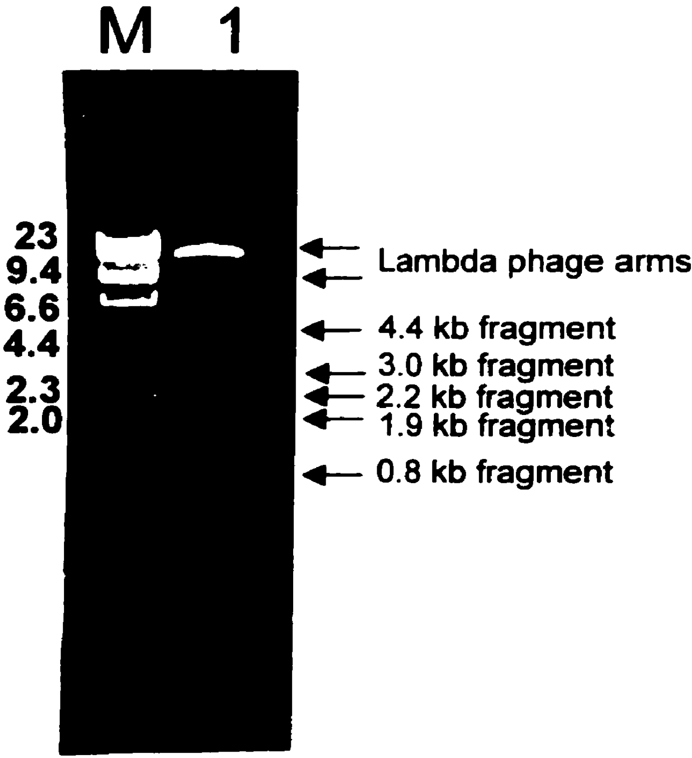


Figure 16: DNA purified from phage clone 3717C(+) and digested with *Bam*HI (Lane 1). Molecular size markers (Lane M) are denoted in kilobases. *Streptomyces* sp. D7 genomic DNA fragments were separately purified and ligated into pUC19 to form the pKCE series of plasmids. The gel consists of 0.7% agarose in TAE buffer, stained with ethidium bromide.

Figure 17: pKCE1 (Lane A), pKCE3 (Lane B), pKCE4 (Lane C), pKCE5 (Lane D) and pKCE6 (Lane E) purified plasmid DNA, digested with *Bam*HI. An arrow denotes the position of linear pUC19. M1 = λ-*Hin*dIII molecular size standards, in kb; M2 = 1 kb molecular size standards, in kb. The gel consists of 0.7% agarose in TAE buffer, stained with ethidium bromide.

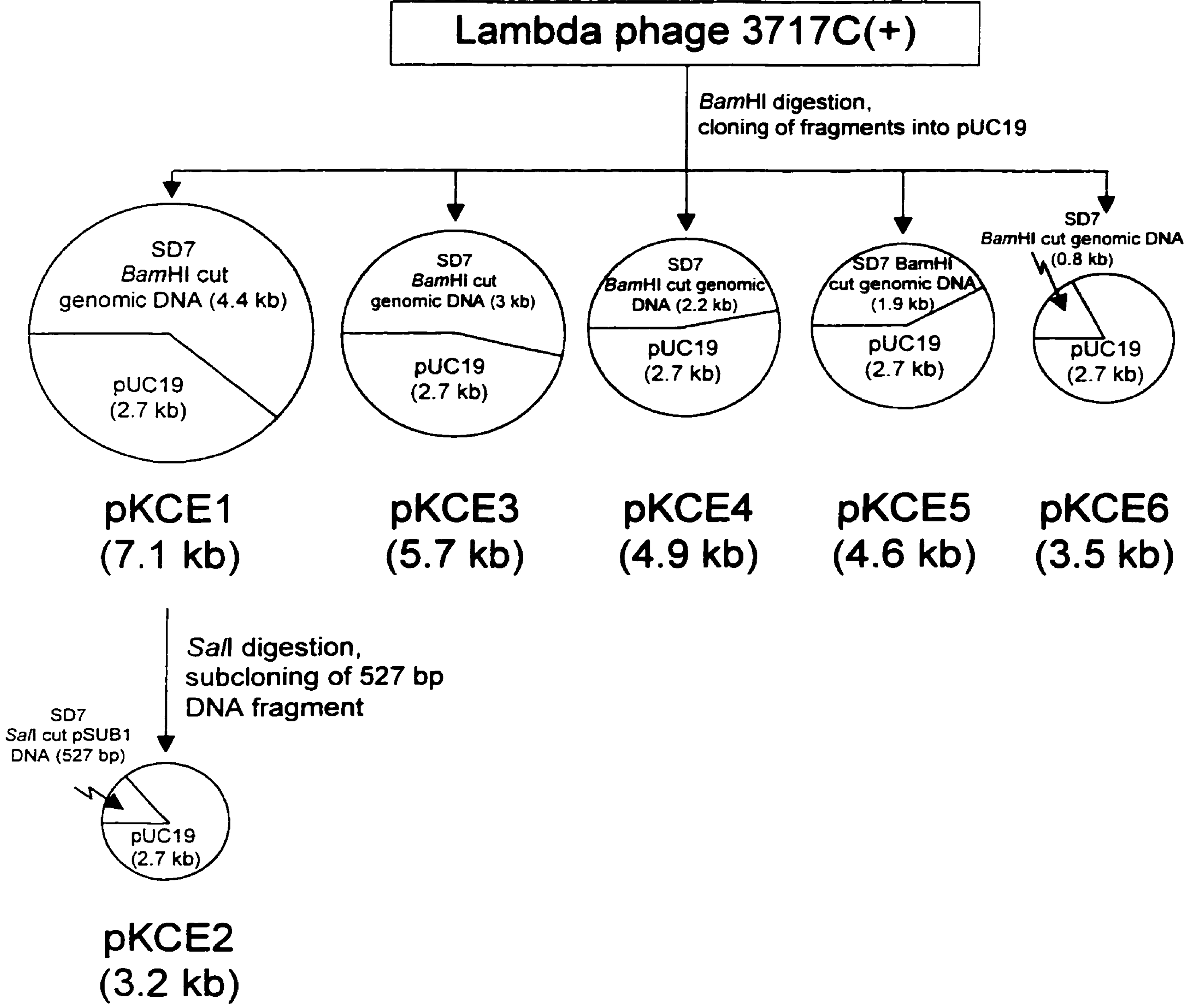


Figure 18: Plasmids carrying subfragments of the *Bam*HI genomic DNA fragment from *Streptomyces* sp. D7, which was cloned in lambda DASH II phage clone 3717C(+). Plasmids are designated pKCE1 through pKCE6, and are described in the text.

```
   1 ccgcgtccag atcgctccgt agcgtgaaag cacaggtcag cgacaggttt gaggtggtcc
     vdcB →
  61 t[illegible]cggttg gtcgtgggaa tgaccggggc gacgggtgcc ccgttcggtg tccgtcttct
 121 ggagaatctg cgccagttgc cgggcgtgga gacacatctc gtcctgtccc gctgggcgcg
 181 caccaccatc gagatggaga cgggcctgtc cgtggccgag gtgtccgccc tagcggacgt
 241 cacgcaccac cccgaggacc agggcgccac catctcctcc ggttcgttcc gcaccgacgg
 301 catggtgatc gtgccgtgct ccatgaagac cctcgccggg atcaggaccg gatacgccga
 361 agggctcgtc gcccgggccg ccgacgtggt tctcaaggag cgacgcaggc tcgtcctcgt
 421 cccgcgcgag acaccgctga gcgagatcca cctccagaac atgctcgaac tcgcccgcat
 481 gggcgtgcaa ctggtgccgc ccatgcccgc cttctacaac aacccgcaga ccgtcgacga
 541 catcgtcgat cacgtggtgg cccgcatcct cgaccagttc gacctgcccg cgcccgccgc
 601 ccggcgctgg gccgggatgc gcgccgcccg cgccgccgcc cgatccttcg gcgacgccgc
                                           vdcC →
 661 c[illegible]agtccc ctgccccgta aaggaattca cc[illegible]gccta tgacgacttg cgcagcttcc
 721 tcgacacctt ggagaaggag gggcagctgc tgcgcatcac cgacgaggtg ctgcccgagc
 781 cggatctcgc ggcggccgcc aacgcgaccg gccgcatcgg cgagaacgcc cccgccctcc
 841 acttcgacaa cgtcaagggc ttcaccgacg cccgcatcgc gatgaacgtg cacggctcct
 901 gggccaacca cgcgctcgcg ctcgggttgc cgaagaacac gccggtcaag gagcaggtgg
 961 aggagttcgc gcggcgctgg gacgccttcc ctgtcgcccc cgagcgccgc gaggaagcac
1021 cctggcgtga gaacacccag gagggcgagg acgtcgacct gttctcggtg cttcccctct
1081 tccgcctcaa cgacggagac ggtggcttct atctcgacaa ggccgccgtc gtctcccgcg
1141 acccggagga ccgggacgac ttcggcaagc agaacgtcgg cacctaccgc atccaggtca
1201 tcggcaccaa ccggctcgcc ttccaccctg ccatgcacga cgtggcccag catctgcgca
1261 aggccgagga gaagggcgag gacctgccca tcgccatcac cctcggcaac gaccccgtga
1321 tggcgatcgt ggccgggatg ccgatggcgt acgaccagag cgagtacgag atggcgggag
1381 ccctgcgcgg cgcgcccgcg cccatcgcca ccgcgccgct caccggcttc gacgtgccct
1441 gggggagcga ggtcgtcata gagggcgtca tcgagtcccg caagcgccga atagaggggc
1501 ccttcggcga gttcaccggt cattactcgg gggggcgccg catgcccgtc atccgcgtgg
1561 aacgcgtctc gtaccggcac gaaccggtct tcgagtcgct ctacctcggc atgccgtgga
1621 acgagtgcga ctacctcgtc ggacccaaca cgtgcgtgcc gctgctcaag cagctgcgcg
1681 ccgagttccc cgaggtgcag gccgtcaacg ccatgtacac gcacggcctg atggtgatca
1741 tctccacggc caagcggtac ggcggcttcg ccaaggccgt cggcatgcgc gccatgacga
1801 cgccgcacgg gctcggctac gtggcccagg tgatcctcgt cgacgaggac gtcgacccgt
1861 tcaacctgcc gcaggtcatg tgggcgatgt ccgccaaggt caacccgaag gacgacgtcg
1921 tcgtcatccc caacctgtcg gtcctggaac tcgcgcccgc cgcgcagccc gccggcatca
1981 gcagcaagat gatcatcgat gcgacgacgc cggtcgcccc ggacgtccgc ggcaacttct
2041 ccactccggc caaggacctg cccgagaccg cagagtgggc cgcccgcctg cagcgcctga
                                                          vdcD →
2101 tcgcggcccg cgtc[illegible]cac ccaccgatcc tcactgaaag gacaactccc [illegible]aaccacc
2161 tgcccgttga atgcccccgc tgcgcctttg aggacatctc cctgcttgcc acgtcccccg
2221 tcccaggcgt gtgggacgtg gtccagtgcg gccgctgtct ctacacctgg cgcacgatcg
2281 aacccgcacg ccgcacccgg cgtgacgcct acccggacag cttcaagctg acggcggagg
2341 acatcgagaa cgccatcgag gtgcccgcgg tgccgccact gctcaag[illegible] cgttcctggc
2401 acgcccggag agaacgtccg ccagaacctg aacgggcccg gcgtcgagct cccacatggg
2461 agtgagggcg ggtggtgccc aactttccct catatacacg tagttgctac tcgctcgcac
2521 tccgtctcgc gcagtgaacg tagaaaatat tcggtagtag gtagtgcaac ggaactcatt
2581 gacgcgctac ctttcgcctg ccccgtcacg gcaggcacac caacggtgtt gggccccatc
2641 cgctgaccgg ggccgaaggc gcacagttcc tgtccaaccc agagactgaa tctggatgca
2701 tgacaccgct ccgttcctcg ctgcacgaat gcgcgggcaa ccaccgttct ggccgtactg
2761 gccgcgctgg gcacctctgt cgtggctgcc gccgcgtcgc ccgccgcggc caaggacgtc
2821 tcgtatcgcg gttatcacgt acgcgtgccc gcaagttggc cagtcgtgga cctgaccgcg
2881 aaccccggca cctgcgtgcg cctcgaccgg cacgcggtct atgtcggtca tccctccgac
2941 gccggtcagg cctcgtgccc g
```

Figure 19: Nucleotide sequence of the *vdc* gene cluster, featuring *vdcB*, *vdcC* and *vdcD*. Putative Shine-Delgarno sites have been underlined, and start (ATG, GTG) and stop (TGA) codons are [illegible] for each open reading frame.

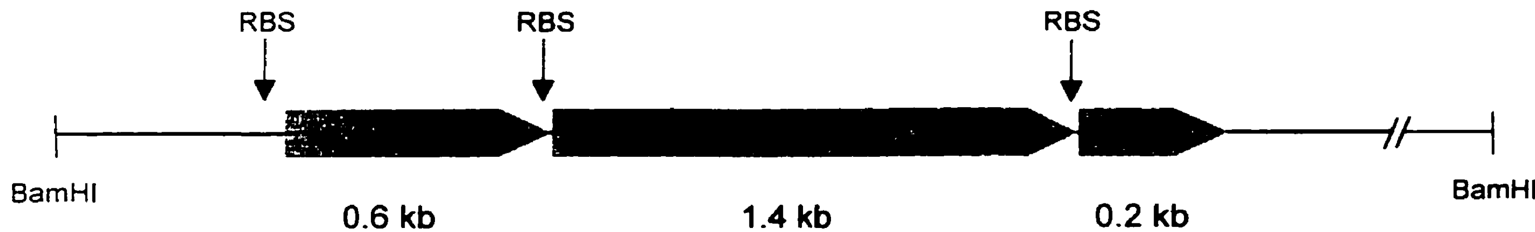


Figure 20: Schematic diagram of the 4.4 kb BamHI genomic DNA fragment from *Streptomyces* sp. D7, containing the VDC gene cluster. Putative ribosome binding site (RBS) locations have been indicated with arrows.

Figure 21 (following pages): Amino acid sequence alignments between the translation products of *vdcB* (a) and *vdcC* (b) and their respective putative homologues from various microorganisms. Dendrograms depicting these alignments are shown for the products of the *vdcB* gene (c) and the *vdcC* gene (d). These dendrograms and sequence alignments were produced using the Pileup program, which is part of the Wisconsin Package Version 10 bioinformatics software package (GCG). eco: *Escherichia coli;* sph: *Sphingomonas aromaticivorans;* bsu: *Bacillus subtilis;* str: *Streptomyces* sp. D7; scv: *Saccharomyces cerevesiae;* arc: *Archaeoglobus fulgidus;* meb: *Methanobacterium thermoautotrophicum;* mec: *Methanococcus jannaschii;* pyr: *Pyrococcus horikoshii;* aqx: *Aquifex aeolicus;* bfi: *Bacillus firmis;* chl: *Chlamydia trachomatis;* hpy: *Helicobacter pylori;* mbr: *Methanobrevibacter smithii;* rho: *Rhodospirillum rubrum;* syn: *Synechocystis* PCC6803; PAD: homologue of phenylacrylate decarboxylase; C: homologue of *vdcC*. The region of the *vdcB* homologues that is a putative membrane association domain is [illegible].

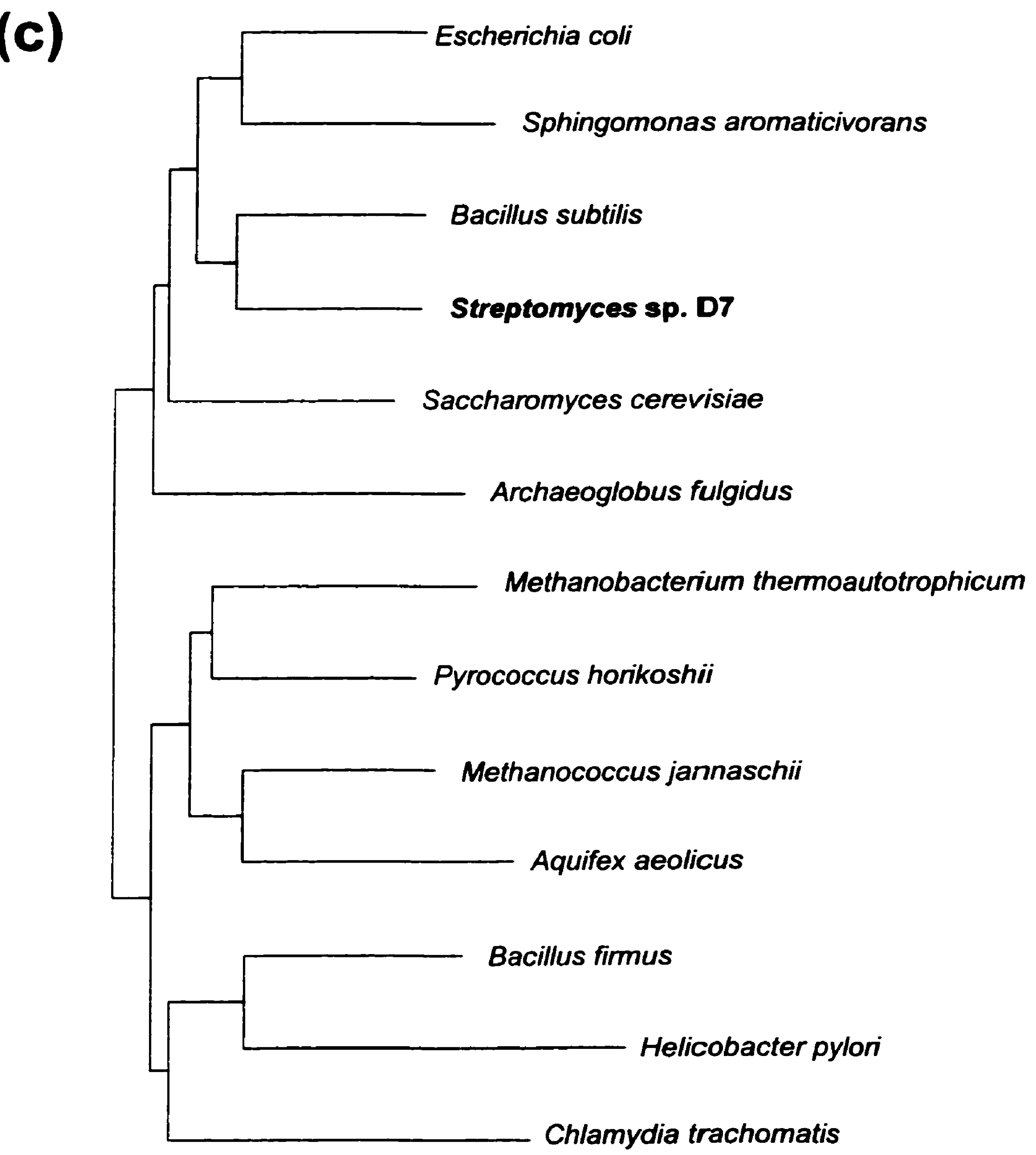
(a)

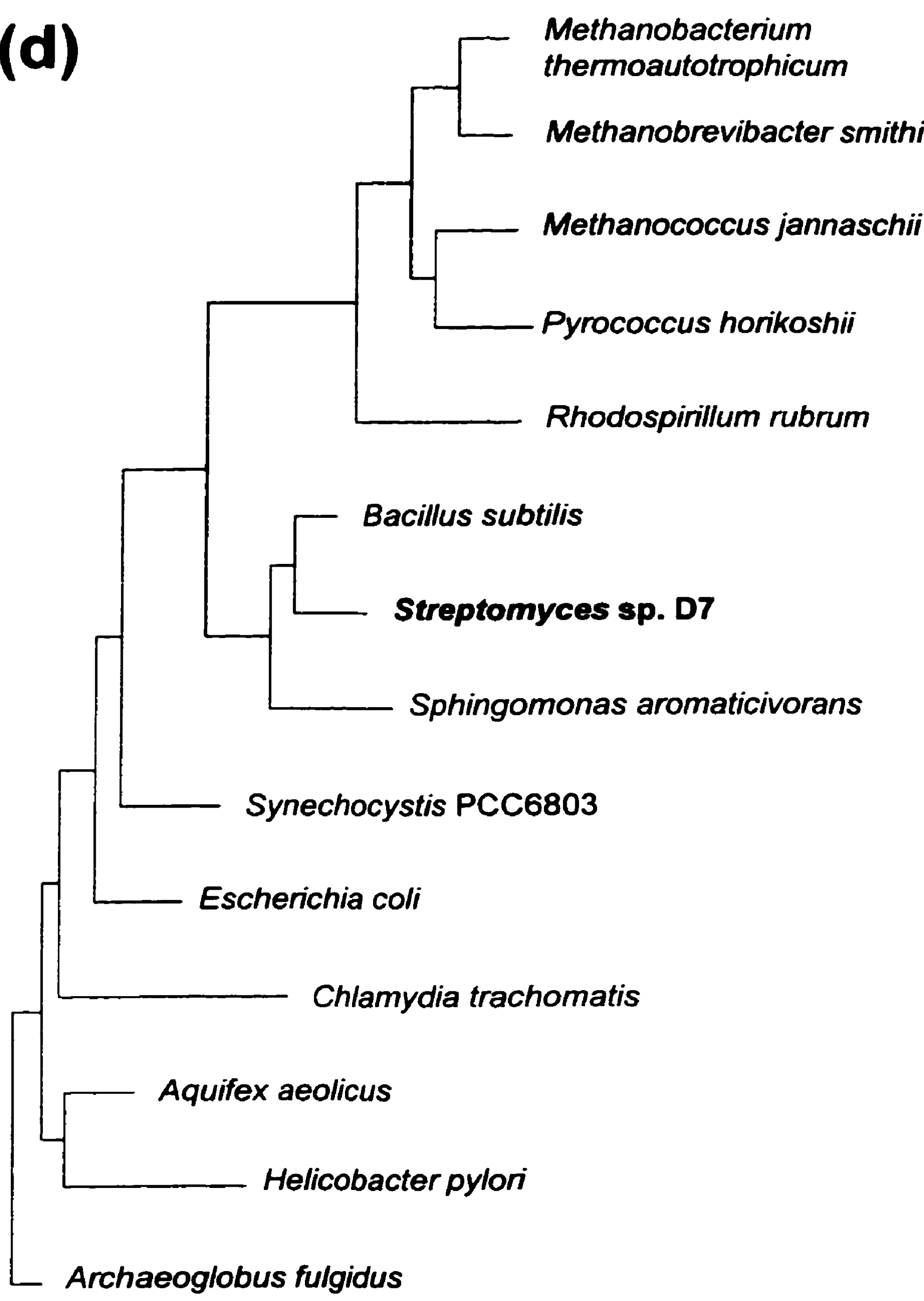
```
ecoPAD     GA~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
sphPAD     GLSKEAASLVPGAGQMEGN~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
bsuPAD     GIEKQKGGA~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
strPAD     GMRAARAAARSFGDAA~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
scvPAD     GIKSK~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
mebPAD     GKDI~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
mecPAD     TV~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
pyrPAD     MREDD~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
aqxPAD     G~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
bfiPAD     E~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
chlPAD     NPR~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
hpyPAD     MN~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
                260       270       280       290       300       310
```

(b)

```
synC    PNVGVYRLQLQSKTTMTVHWLSVRGGARHLRK..AAEQGKKLEVAIALGVDPLIIMAAATP...I
hpyC    KNLGMYRLQVYDKNHLGLHWQIHKDSQLFFHEYAKAKV..KMPVSIAIGGDLLYTWCATAP...L
chlC    PNLGMYRMQRFDKETLGLHFQIQKGGGAHFFE..AEQKKQNLPVTVFLSGNPFLILSAIAP...L
          200       210       220       230       240       250       260

mebC    PIDADEMEVANTFH....EGELELVRCEGVDMEVPPAEIILEGRILCGVRE.REGPFVDLTDTYD
mbrC    PIDYNEMDVANAFK....NGELTLIKCG..DLEVPQADIILEGKISVSETS.AEGPFVDLTDTYD
mecC    DITFDELKFAAAL....LGGEIGVFELDNGLL.VPEAEFIIEGKIL.PEVD.DEGPFVDITGTYD
pyrC    AYGVSELEIASAISLKAFGRPLEVINLDGIPT.PVDSEFVFKAKIT.DEVA.DEGPFVDITGTYD
rhoC    PVETDELGIASAFHGAPLELVAGTV...GPVEMVAHAMWALECEIRPGEVH.AEGPFAEVTGYYA
bsuC    LYDQSEYEMAGAIQGEPYR.IVKSKLSD..LDVPWGAEVVLEGEIIAGEREY.EGPFGEFTGHYS
strC    AYDQSEYEMAGALRGAPAP.IATAPLTG..FDVPWGSEVVIEGVIESRKRRI.EGPFGEFTGHYS
sphC    GYEESEYSYASAMMGAPIR.LTKSG.NG..IDILADSEIVIEAELQPGGREL.EGPFGEFPGSYS
aqxC    PPEVDEYLFAGIIMERPVE.LVKGLTVD..LEYPANAEIAIEGYVDPEEPLVDEGPFGDHTGFYT
arcC    PSGISEFMFAGFIRKERLK.VTECETVD..LLVPANAEIILEGYVRVDEMRV.EGPFGDHTGYYT
ecoC    PDTLSEYAFAGLLRGTKTE.VVKCISND..LEVPASAEIVLEGYIEQGET.APEGPYGDHTGYYN
synC    PVDLSEWLFAGLYGGSGVA.LAKCKTVD..LEVPADSEFVLEGTITPGEM.LPDGPFGDHMGYYG
hpyC    PYGIYELMLYGFMRGKKAR.VMPCLSNS..LSVPSDCDIVIEGFVDCEKLEL.EGPFGDHTGYYT
chlC    PENVPELLFCSFLQNKKLSFVEKHPQSG..HPLLCDSEFILTGEAVAGERR.PEGPFGDHFGYYS
                 270       280       290       300       310       320

mebC    VVRDEPVISLERMHIRKD.AMYHAILPAGF.EHRLLQGLPQEPRIYRAVKNTVPTVRNVVLTEGG
mbrC    IIRDQPIINLSKMHIKKDNPHYHGILSAGF.EHKLLQGLPQEPRIFKSVKNAVPTVENVVLTEGG
mecC    IVRKQPIIKIEKLY.RKEKPIFHALLPGGI.EHKTLMGMPQEPRILKGVRNTVPTVKNIVLTEGG
pyrC    IVRKQPIVIFEEMY.HVDDPIFHALLPGGY.EHYMLMGLPKEPQIYASVKKVVPKVHGVRLTEGG
rhoC    RVEPRPLVRVKRIH.RRRAPIFHTLL.SGA.EVFNSVGLLGEANVLALLRVQVPGVEDVYFSHGG
bsuC    GGRSMPIIKIKRVYHRNNPIFEHLYLGMPWTECDYMIGINTCVPLYQQLKEAYPNEIVA.VNAMY
strC    GGRRMPVIRVERVSYRHEPVFESLYLGMPWNECDYLVGPNTCVPLLKQLRAEFP.EVQA.VNAMY
sphC    GVRKAPIFKVTAVSHRRDPIFENIYIGRGWTEHDTLIGLHTSAPIYAQLRQSFP.EVTA.VNALY
aqxC    PVDKYPQMHVTAIVMRKDPIYLTTIVGRPPQE.DKYLGWATERIFLPLIKFNLPEVVDYHLPAEG
arcC    PPEPYPVFHITHITHRENPIYHATVVGKPPME.DAWLGKATERIFLPILRMMHPEIVDINLPVEG
ecoC    EVDSFPVFTVTHITQREDAIYHSTYTGRPPDE.PAVLGVALNEVFVPILQKQFPEIVDFYLPPEG
synC    GVEDSPLVRFQCLTHRKNPVYLTTFSGRPPKE.EAMMAIALNRIYTPILRQQVSEITDFFLPMEA
hpyC    PIEPYPVLEVKTISYKKDSIYLATVVGKPPLE.DKYMGYLTERLFLPLLQMNAPNLIEYYMPENG
chlC    LTHDFPIFKCNCLYHKKDAIYPATVVGKPFQE.DFFLGNKLQELLSPLFPLIMPGVQDLKSYGEA
           330       340       350       360       370       380       390

mebC    CCWLHAAVSIKKQTEGDGKNVIMAALAAHPSL...KHVVVVDEDIDVLDPEEIEYAIATRVKGDD
mbrC    CCWLHAAISINKQTEGDGKNAIMAALSAHPSL...KHAVVVDTDVDVFDPQDIEYAIATRVKGDR
mecC    CCWLHAVVQIEKRTEGDGKNAILAAFASHPSL...KHVIVVDDDINIFDINDVEYAIATRVQGDK
pyrC    CMWLHAVVSITKQHEGDGKNAILAAFAGHPSL...KRVVVVDEDVNIYDDREVEWAIATRFQPDR
rhoC    CGFYHCVVKIAQKRAGWAKQAILATFAAFPPL...KMVTVVDEDVDIRNGRDVEWAMTTRLDAKT
bsuC    THGLIAIVSTKTRYGGFAKAVGMRALTTPHGLGYCKMVIVVDEDVDPFNLPQVMWALSTKMHPKH
strC    THGLMVIISTAKRYGGFAKAVGMRAMTTPHGLGYVAQVILVDEDVDPFNLPQVMWAMSAKVNPKD
sphC    QHGLTGIISVKNRMAGFAKTVALRALSTPHGVMYLKNLIMVDADVDPFDLNQVMWALSTRTR.AD
aqxC    CFHNFCFVSIKKRYPGHAFKVAYALLG.LGLMSLEKHIVVFDDWINVQDIGEVLWAWGNNVDPQR
arcC    AFHNLAIVSIKKRYPGQAKKVMYAIWG.TGMLSLTKIVVVVDDDVNVHDMREVVWAVTSRFDPAR
ecoC    CSYRLAVVTIKKQYAGHAKRVMMGVWSFLRQFMYTKFVIVCDDDVNARDWNDVIWAITTRMDPAR
synC    LSYKAAIISIDKAYPGQAKRAALAFWSALPQFTYTKFVIVVDKSINIRDPRQVVWAISSKVDPVR
hpyC    VFHNLILAKIHTRYNAHAKQVMHAFWGVGQMSFVKHAIFVNEDAPNLRDTNAIIEYILENF..SK
chlC    GFHALAAAIVKERYWKEALRSALRILGE.GQLSLTKFLWITDQSVDLENFPSLLECVLERMNFDR
                   400       410       420       430       440       450

mebC    DILIVPGARGSSLDPAA.LPDGTTTKVGVDATAPL.ASAEKFQRVSRSE----------------
mbrC    DLMIVPNVRGSSLDPVA.ESDGTTTKIGLDATKSL.KTLDKFERVSFGE----------------
mecC    DIVIISGAKGSSLDPSSDLKNKLTAKVGVDATMSLIKGREHFERAKIPDK---------------
pyrC    DLVIIPNARGSSLDPSG..KDGLTAKWGIDATKPLDKKKE.FEKASLDF----------------
rhoC    GILVIENAFGHGLNPT..FPNYLGTKVGFDCTRPFPHT.PAFDRAKTKAMTLDGLDIVGAKR---
bsuC    DAVIIPDLSVLPLDPGSNPSGITHKMI.LDATTPVAPETRG.HYSQPLDSPLTTKEWEQKLMDLM
           460       470       480       490       500       510       520
```

```
strC     DVVVIPNLSVLELAPAAQPAGISSKMI.IDATTPVAPDVRG.NFSTPAKDLPETAEWAARLQRLI
sphC     DIIVLPNMPAVPIDPSAVVPGKGHRLI.IDATSYLPPDPVG.EAHLVTPPTGDEIDALSKRIREM
aqxC     DVLIL.KGPIDVLDHATNEVGFGGK.MIIDATTKWKEEGYTREWPEVIEMSPEVKKRIDEIWDRL
arcC     DVVILPPSPTDSLDHSAYIPNLAGK.LGIDATKKWRDEGYEREWPDVVEMDAETKRKVDAIWNEI
ecoC     DTVLVENTPIDYLDFASPVSGLGSK.MGLDATNKWPGE.TQREWGRPIKKDPDVVAHIDAIWDEL
synC     DVFILPETPFDSLDFASEKIGLGGR.MGIDATTKIPPE.TDHEWGEVLESDPAMAEQVSQRWAEY
hpyC     ENALISQGVCDALDHASPEYAMGGK.LGIDATSK.....SNTPYPTLLNDSALLALLQDKMQNIV
chlC     DLLILSETANDTLDYTGSGFNKGSKGIFLGVGAPIRSLPRRYRGPSLPGISQIGVFCRGCLVLET
           460       470       480       490       500       510       520

bsuC     NK----------------------------------------------------------
strC     AARV--------------------------------------------------------
sphC     QLGALS------------------------------------------------------
aqxC     GIE---------------------------------------------------------
arcC     RNMVL-------------------------------------------------------
ecoC     AIFNNGKSA---------------------------------------------------
synC     GLGDINLTEVNPNLFGYDV-----------------------------------------
hpyC     LLKQYYPHTRNPICVISVEKKDKSVIELAKNLLGFEEHLRIVIFVEHASNDLNNPYMLLWRIVNN
chlC     SLQQLDIPALLKEPHLADWPLVILVEDLSSALSSTKEFIWRTFTRSSPATDLHIPVSQITNHKVS
                530       540       550       560       570       580

hpyC     IDARRDILTSKHCFFIDATNKGVMDKHFREWPTETNCSMEVIENLKKKGLLKDFETLNQKFHLTH
chlC     YTPPMILNALMKPPYPKEVEADEATQNLVSSRWHSYFP----------------------
           590       600       610       620       630       640       650

hpyC     SFSTHKEDL---------------------------------------------------
                660       670       680       690       700       710
```


(c)
Escherichia coli
Sphingomonas aromaticivorans
Bacillus subtilis
Streptomyces sp. D7
Saccharomyces cerevisiae
Archaeoglobus fulgidus
Methanobacterium thermoautotrophicum
Pyrococcus horikoshii
Methanococcus jannaschii
Aquifex aeolicus
Bacillus firmus
Helicobacter pylori
Chlamydia trachomatis


(d)
Methanobacterium thermoautotrophicum
Methanobrevibacter smithii
Methanococcus jannaschii
Pyrococcus horikoshii
Rhodospirillum rubrum
Bacillus subtilis
Streptomyces sp. D7
Sphingomonas aromaticivorans
Synechocystis PCC6803
Escherichia coli
Chlamydia trachomatis
Aquifex aeolicus
Helicobacter pylori
Archaeoglobus fulgidus

3.2 Detection of a putative transcriptional regulatory gene

Preliminary DNA sequencing of the region upstream of the VDC gene cluster indicates the presence of a putative divergently transcribed regulatory gene, the position of which is illustrated in Figure 22. BLAST-X analysis of the translation product of the region matched strongly to a putative positive transcriptional activator from the *Streptomyces coelicolor* A3(2) genome (Figure 23).

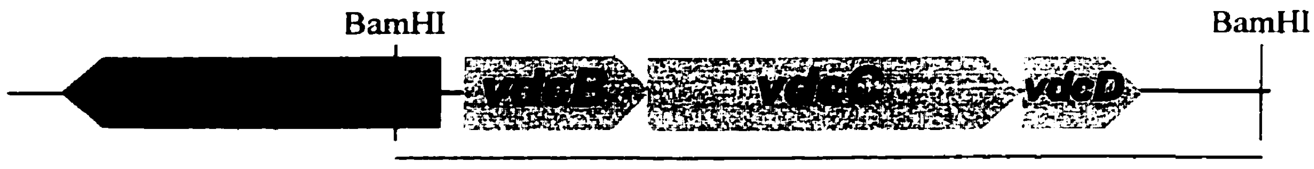


Figure 22: Location of the putative divergent regulatory gene in relation to the VDC gene cluster.

```
emb|CAA19943| (AL031107) putative transcriptional regulator [Streptomyces
            coelicolor]
            Length = 301

 Score =  119 bits (295), Expect = 5e-26
 Identities = 86/256 (33%), Positives = 111/256 (42%), Gaps = 4/256 (1%)
 Frame = -1

Query: 1180 FHQLIGVRGRPLRMWVDFVEHELRPGSWLWIRPGQVQRFGPDLAAADGVIVLFQPGFLPP 1001
            FH++    G P+R  +DF E+E   G  LWIRPGQV RF P+     G ++  QPGFLP
Sbjct: 54   FHIVMLFTGGPVRHMIDFAEYEASAGDLLWIRPGQVHRFAPE-GEYRGTVLTMQPGFLPR 112

Query: 1000 TTVSLAHMDPPYE-QRPSVLEGSDAE--GVRRALDHLVHEYGAMASLPLQAHTE-XXXXX 833
             TV    +   Y    P +L   +A   G+  AL+ L  EY    +LPL  HT
Sbjct: 113  ATVEATGL---YRYDLPPLLHPDEARLAGLTAALEQLRREYEDATTLPLSLHTAVLRHTL 169

Query: 832  XXXXXXXXXXXRGPAPRPTAAFGNVPSLSHXXXXXXXXXXGASTNTRGRSATAPHLTRRV 653
                        G A          P  S          G +TN              V
Sbjct: 170  SAFLLRLAHLAAGSARAARQGRAEAPGDSTFVLFRDAVERGFATN------------HSV 217

Query: 652  DDYAAALGYSSXXXXXXXXXXXXXXANQYVDDRVLLEAKRLLQHSGLTAREVTVRLGFTD 473
              YA ALGYS                  ++D RV+LEAKRLL H+ +    V   +GF D
Sbjct: 218  SAYADALGYSRRTLVRAVRAATGETPKGFIDKRVVLEAKRLLAHTEMPIGRVGAAVGFPD 277

Query: 472  ASDFTKFFRLRTGMTPGAFR 413
            A++F+KFF+  T  TP AFR
Sbjct: 278  AANFSKFFQQHTDQTPAAFR 297
```

Figure 23: BLAST-X result indicating that the nucleotide region immediately upstream of *vdcB* encodes a polypeptide similar to a putative transcriptional regulator from *S. coelicolor* A3(2). Query = *Streptomyces* sp. D7 translated nucleotide sequence; Subject = *S. coelicolor* A3(2) genomic DNA translation product.

3.3 Messenger RNA analyses

Messenger RNA was isolated from *Streptomyces* sp. D7 under both uninduced and vanillic acid-induced conditions, blotted to a positively charged nylon membrane, then probed with ^{32}P-labeled *vdcC* PCR amplification product. The resulting autoradiogram revealed that a transcript of approximately 2.3 kb was synthesized in vanillic acid-induced cells (Figure 24). A transcript of this size corresponds to the expected length of all three VDC genes and their associated intergenic regions. This result indicates that the gene cluster may be transcribed from a single, inducible promoter.

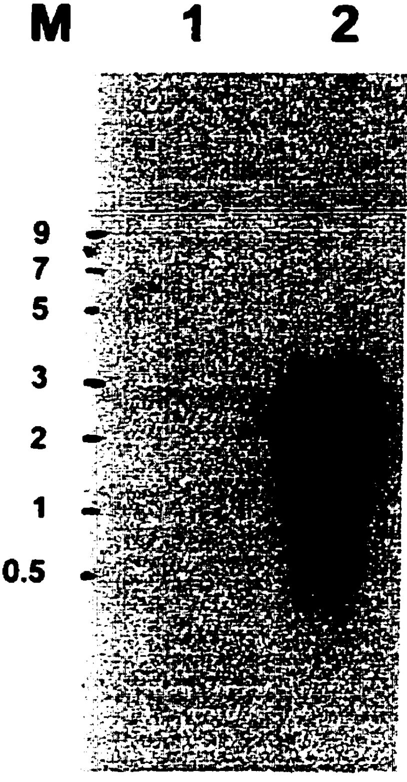


Figure 24: Northern blot hybridization of PCR-amplified, radiolabeled *vdcC* DNA against mRNA isolated from uninduced *Streptomyces* sp. D7 (Lane 1) and *Streptomyces* sp. D7 induced with 3.6 mM vanillic acid (Lane 2). RNA size standards are indicated (Lane M).

3.4 Detection of VDC cluster genes in *Streptomyces setonii* 75Vi2

Southern hybridization of radiolabeled *vdcB*, *vdcC*, and *vdcD* PCR products to *Sal*I-digested *S. setonii* 75Vi2 genomic DNA revealed that all three genes of the cluster appear to be present in the characterized vanillate decarboxylating strain (Figure 25). At the time of the writing of this report, a lambda DASH II phage clone of *S. setonii* 75Vi2 genomic DNA, which hybridizes strongly to radiolabeled *vdcC* probe, is being analyzed and sequenced (M.K. Pope, collaborative work) in order to further characterize the VDC cluster of *S. setonii* 75Vi2. Preliminary sequence information reveals that within the phage clone DNA insert, a cytochrome P-450 gene is located within several kilobases of the putative VDC gene cluster in the genome of *S. setonii* 75Vi2. In addition, the *vdcB* homologue in *S. setonii* 75Vi2 has been sequenced, with further sequencing underway to uncover the remaining genes. As mentioned earlier, cytochrome P-450 enzymes have been implicated in the demethylation of methoxyl groups on guaiacol by *S. setonii* 75Vi2 (Sutherland, 1986). No cytochrome P-450 gene was observed in the vicinity of the VDC gene cluster in *Streptomyces* sp. D7, which supports the observation that no transformation of guaiacol could be detected in the organism. These observations will be described in more detail in the Discussion section to follow.

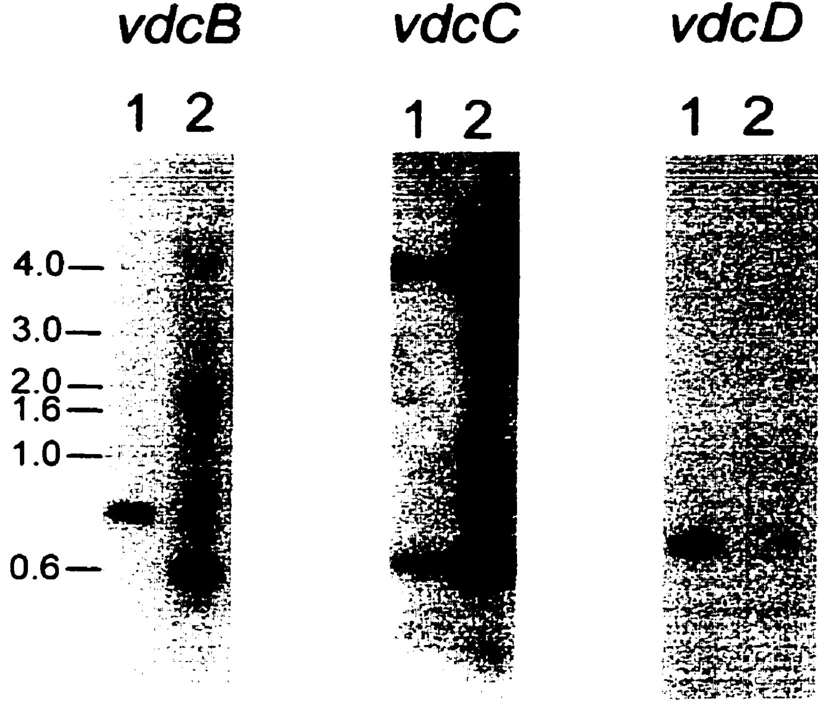


Figure 25: Southern blot hybridization of *SalI*-digested chromosomal DNA from *Streptomyces* sp. D7 (Lane 1) and *Streptomyces setonii* 75Vi2 (Lane 2) with radiolabeled *vdcB*, *vdcC* and *vdcD* (as indicated) PCR-amplified DNA probes at 65°C. Approximate molecular sizes (in kilobases) are shown on the left.

3.5 Gene knockout studies

Gene knockout experiments involve transforming cells with an inactivated form (allele) of the gene being studied, for homologous recombination with the wild type gene on the host's chromosome. Inactivation is usually achieved by replacing a section of the gene with an antibiotic resistance cassette. In streptomycete procedures, the thiostrepton resistance gene is commonly used. These experiments are commonly used for gene identification and/or functional analyses, and disruption and inactivation of the VDC gene cluster would allow determination of the functions of each of the three genes *in vivo*. In order to do this, *Streptomyces* sp. D7 had to be transformed with a disrupted allele of each gene, to obtain knockout mutants by double crossover recombination events with the chromosome. Protoplasts of *Streptomyces* sp. D7 were prepared, which regenerated efficiently on R6 agar plates designed for that purpose. However, in spite of many attempts, these protoplasts were not transformable with pIJ702 as a control plasmid using standard *Streptomyces* protoplast genetic manipulations. Likewise, attempts to introduce pPM801 into *Streptomyces* sp. D7 mycelia by interspecies *Streptomyces-E. coli* conjugation procedures failed to produce ex-conjugants expressing the kanamycin resistance marker gene of the vector. These observations suggest that *Streptomyces* sp. D7 may possess a potent restriction-modification system, which degrades foreign DNA. *Streptomyces* is particularly well known for being highly resistant to transformation procedures (Oh & Chater, 1997), which is the major reason why the vast majority of genetic research on these organisms has been conducted with either *Streptomyces lividans* or *Streptomyces coelicolor*, organisms for which transformation procedures have been well established. The resistance of *Streptomyces* sp. D7 to genetic transformation

made it necessary to devise alternative methods to characterize the functions of the VDC gene cluster. I decided to express each gene independently to examine their activity in *Escherichia coli* and/or *Streptomyces lividans* 1326 as surrogate host systems, which is described in Section 3.6.

3.6 Gene expression

3.6.1 Expression with pET22b(+) in *E. coli* BL21(DE3)

Initially, the goal of the project was to clone individual VDC cluster genes behind the T7 polymerase promoter in the expression vector pET22b(+) (Novagen), to obtain high amounts of active enzymes in the expression strain *Escherichia coli* BL21(DE3). Studies of the purified, active protein would allow the deduction of the function of each enzyme in the catabolism of vanillic acid. However, although the *vdcB* and *vdcC* genes were successfully expressed in *E. coli*, the proteins were formed as inclusion bodies, no matter how the conditions for expression were varied to achieve proper protein folding. For example, protein expression was carried out at 37°C and 25°C, with both 1 mM and 0.1 mM IPTG used for induction of the T7 RNA polymerase system. Attempts were made to solubilize and refold the inclusion body material, but with no established biochemical assay and only a putative phenotype, it was decided that streptomycete vectors and host systems would be more appropriate for the expression of the VDC genes and the determination of their function(s). A typical example of SDS-PAGE-visualized protein over expressed using pET22b(+) in *E. coli* BL21(DE3) is shown in Figure 26, in which the *vdcC* gene product was produced as inclusion bodies at 37°C with 0.1 mM IPTG used

to induce protein synthesis. An example of inclusion body formation at lower induction temperatures is shown in Figure 27.

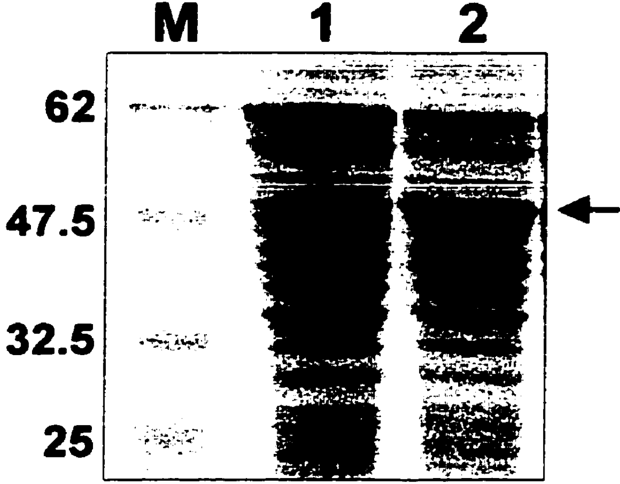


Figure 26: SDS-PAGE (12.5% acrylamide) of cell extracts showing the expression product of the *vdcC* gene, produced using pET22b(+) in *E. coli* BL21(DE3) at 37°C with 0.1 mM IPTG for induction. Lane 1: cell lysate from uninduced cells; Lane 2: cell lysate from IPTG-induced cells. The position of the *vdcC* expression product is denoted by an arrow. Molecular weight markers are shown on the left, in kilodaltons.

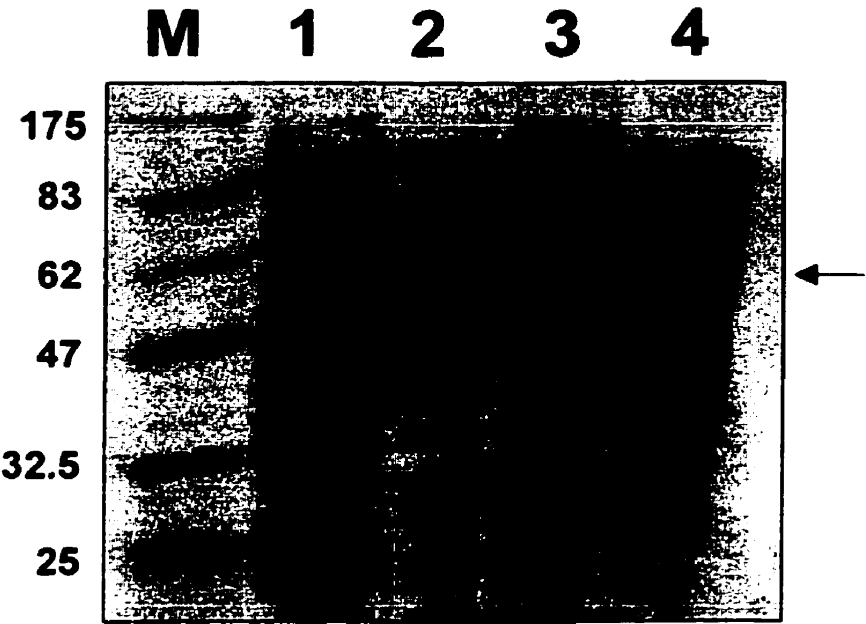


Figure 27: SDS-PAGE (12.5% acrylamide) of cell extracts showing the expression product of the *vdcC* gene, produced using pET22b(+) in *E. coli* BL21(DE3) at 25°C and 30°C with 0.1 mM IPTG for induction. Lane 1: soluble cell lysate, 25°C induction; Lane 2: insoluble protein, 25°C induction. Lane 3: soluble cell lysate, 30°C induction; Lane 4: insoluble protein, 25°C induction. The position of the *vdcC* expression product is denoted by an arrow. Molecular weight markers are shown on the left, in kilodaltons.

In hindsight, the strategy to express the *vdcB* and *vdcC* genes independently most likely would not have been successful, even if the proteins had been synthesized as soluble, active enzymes. As is detailed in Section 3.6.2 and Section 3.6.3 (below), it appears as though expression of at least the *vdcC* and *vdcD* genes, simultaneously, is required to produce vanillate decarboxylase activity.

3.6.2 Expression with pIJ702 in *S. lividans* 1326

Gene expression in *Streptomyces* is commonly performed using the high copy number plasmid, pIJ702 (Gusek & Kinsella, 1992). The vector, in combination with *S. lividans* 1326, has been used extensively to study antibiotic biosynthesis gene structure and expression. *S. lividans* 1326 carrying pKCS1 acquired the ability to efficiently decarboxylate vanillic acid to guaiacol, while *S. lividans* 1326 carrying pKCS2 produced extremely low amounts of guaiacol (Figure 28). *S. lividans* 1326 wild type did not perform any detectable vanillic acid decarboxylation. These results suggest that transcription of the genes required for vanillic acid decarboxylation by *S. lividans* 1326 carrying pKCS1 occurs from the constitutive *mel* promoter. There is possibly a low level of transcription from a natural promoter, however, as observed for *S. lividans* 1326 carrying pKCS2 that contains the gene cluster in the opposite orientation from the *mel* promoter (Figure 28 (b)).

3.6.3 Expression with pIJ680 in *S. lividans* 1326

The results of the *aph* promoter - *vdc* gene fusions are shown in Figure 30. *S. lividans* 1326 carrying pKCS3 (*vdcBCD*) or pKCS8 (*vdcCD*) converted vanillic acid to guaiacol

at approximately the same rate as the wild type strain, *Streptomyces* sp. D7. Sonicated cell-free extracts of *S. lividans* 1326 expressing the VDC system via pKCS3 exhibited decarboxylation of vanillic acid under both aerobic and anaerobic conditions (Figure 29). These results confirm the involvement of the cloned gene products in a non-oxidative system.

Transcription of *vdcC* and *vdcD*, as demonstrated by pKCS8, results in vanillate decarboxylation by *S. lividans* 1326. Therefore, it is the *vdcC* gene that possibly encodes the active enzyme, and the 239 bp *vdcD* gene may encode a protein that is essential for enzyme stability or activation. The *vdcD* gene product of approximately 9 kDa seems too small to be the decarboxylase itself, as subunits for known aromatic acid decarboxylases range from 28 kDa to 66 kDa (Table 3).

3.7 Substrate specificity

Sonicated cell extracts of *S. lividans* 1326/pKCS8 were used to test the specificity of the *Streptomyces* sp. D7 vanillate decarboxylase towards aromatic acids similar to vanillic acid. The following compounds were tested: 4-methoxy-3-hydroxybenzoate (isovanillic acid), 3,4-dimethoxybenzoate (veratrate), 3,4-dihydroxybenzoate (protocatechuate), 4-hydroxy-3,5-dimethoxybenzoate (syringate), 3,4,5-trihydroxybenzoate (gallate), 3-phenylpropenoate (*trans*-cinnamate). The structures of these compounds are shown in Figure 31. The UV spectrophotometry scans of assay mixtures at the time of substrate addition, fifteen minutes post-addition and several hours post-addition were identical, suggesting that the enzyme did not decarboxylate these compounds. By contrast,

ultraviolet spectrophotometry demonstrated vanillic acid decarboxylation to guaiacol after fifteen minutes in the soluble fractions of sonicated cell extracts, as shown in Figure 29.

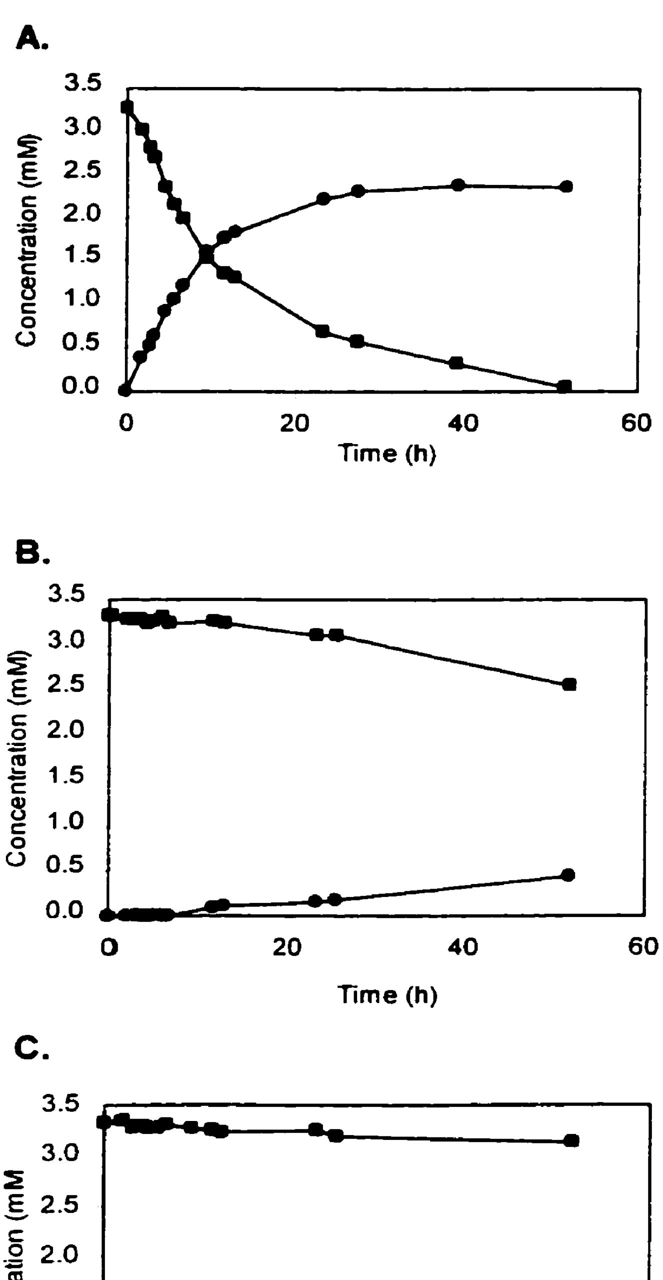


Figure 28: Decarboxylation of vanillic acid to guaiacol by recombinant *Streptomyces lividans* 1326 strains. (a) *S. lividans* 1326 carrying pKCS1; (b) *S. lividans* 1326 carrying pKCS2; (c) *S. lividans* 1326 wild type. Vanillic acid concentration is represented by squares (■); guaiacol concentration is represented by circles (●).

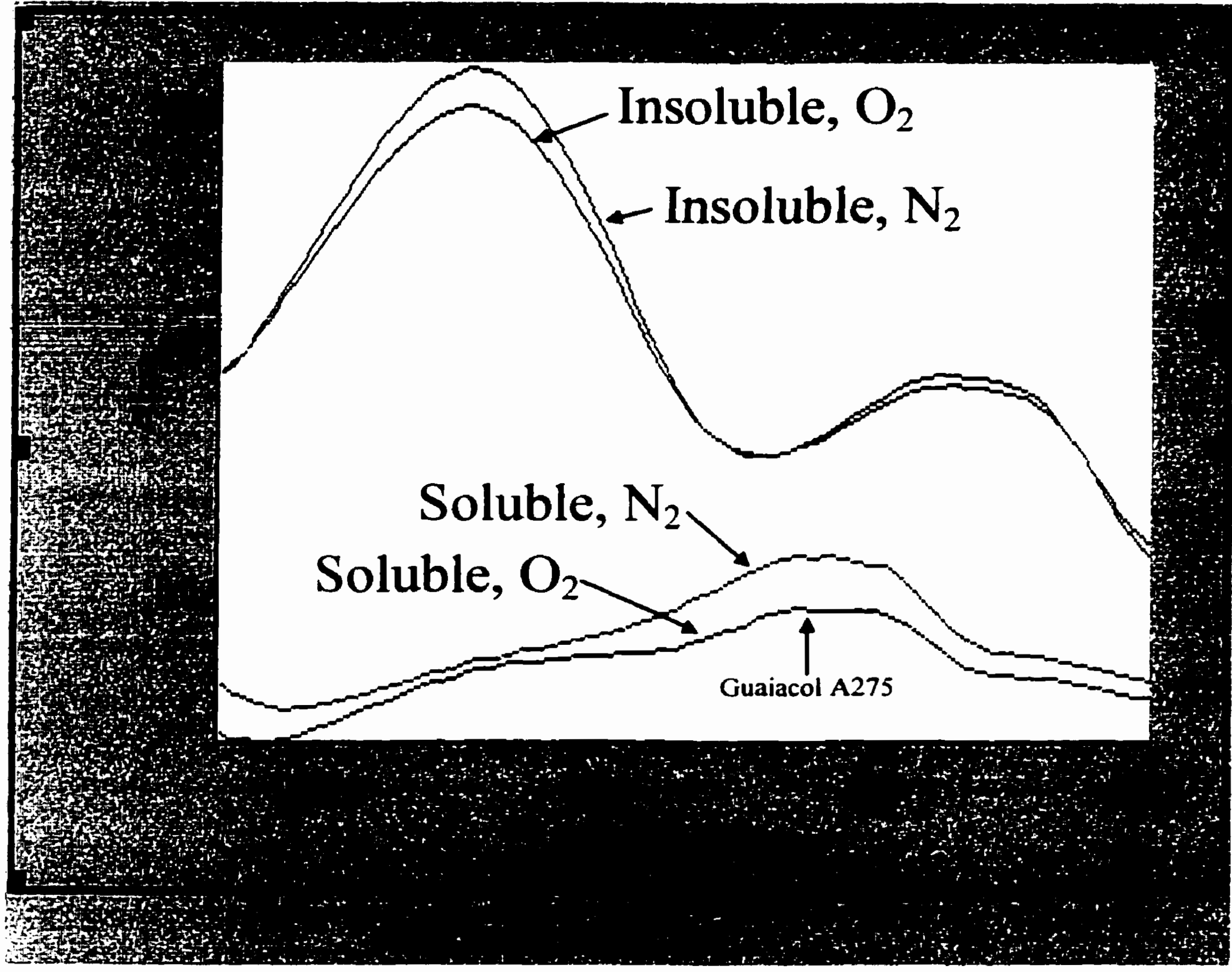


Figure 29: Results of incubating the insoluble and soluble fractions of sonicated cell extracts of *S. lividans* 1326 – pKCS8 with 1 mM vanillic acid for fifteen minutes at 25°C under both aerobic (O_2) and anaerobic (N_2) conditions. These scans indicate that the decarboxylase activity is in the soluble fraction, and that the enzyme is active under both aerobic and anaerobic conditions. Similar experiments using other aromatic acid substrates did not demonstrate any biotransformation to decarboxylated derivatives. The λ_{max} for guaiacol (275 nm) is indicated.

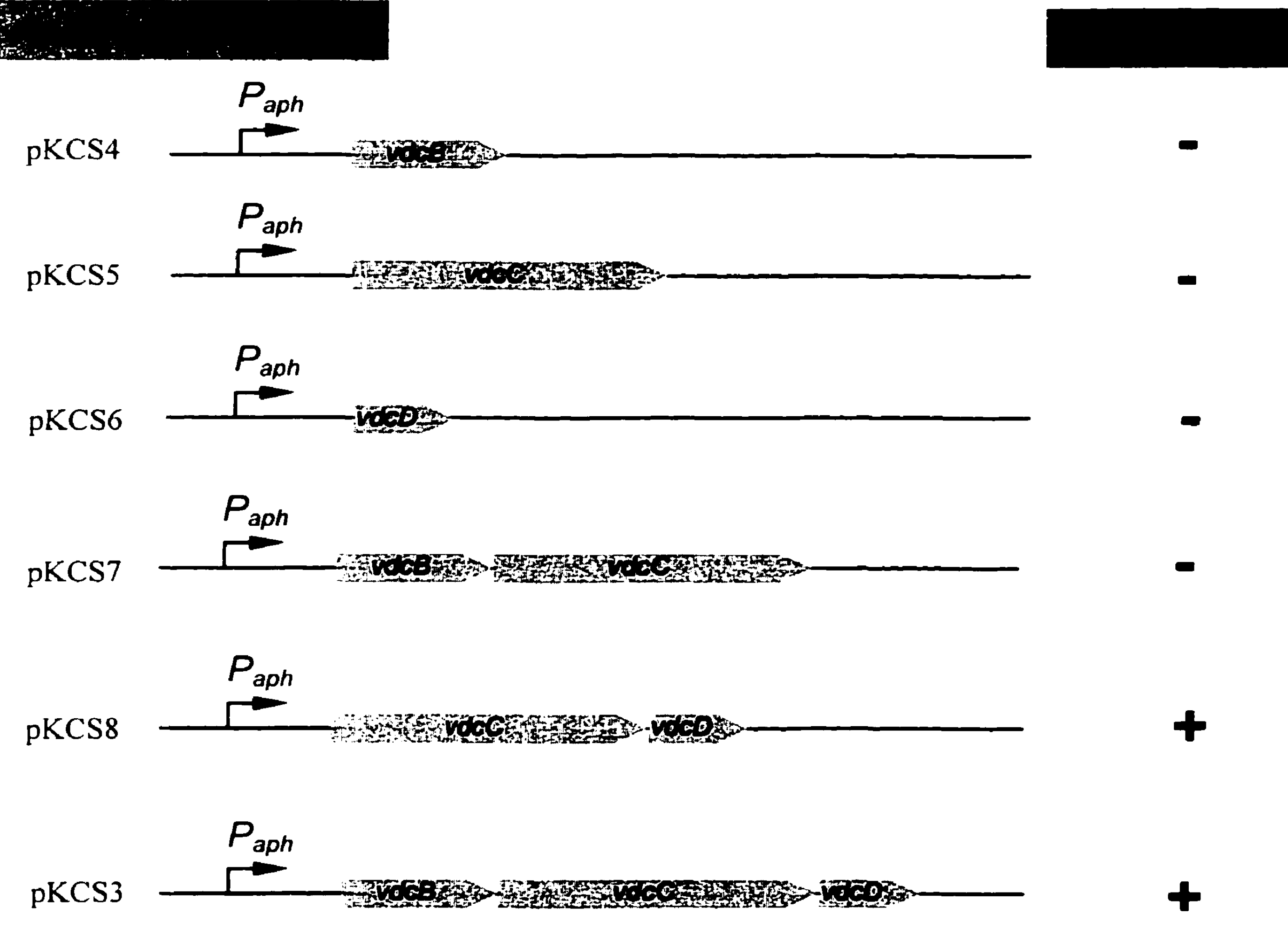


Figure 30: Results of placing the *vdc* genes under control of the aminoglycoside phosphotransferase promoter (P_{aph}) in pIJ680.

Substrate tested	*Structure*	*Decarboxylation?*
Vanillic acid	CO_2H, OCH_3, OH	YES
Isovanillic acid	CO_2H, OH, OCH_3	NO
Veratric acid	CO_2H, OCH_3, OCH_3	NO
Syringic acid	CO_2H, H_3CO, OCH_3, OH	NO
Gallic acid	CO_2H, HO, OH, OH	NO
t-cinnamic acid	CO_2H	NO

Figure 31: Chemical structures of vanillic acid and similar aromatic acids used to test the substrate specificity of vanillate decarboxylase.

4. CONCLUSIONS

The goal of this study was to determine the genetic basis of the process of non-oxidative decarboxylation of vanillic acid in *Streptomyces* sp. D7. It was demonstrated that *Streptomyces* sp. D7 performs the reaction by expressing products of a co-transcribed, three-gene cluster comprised of *vdcB*, *vdcC* and *vdcD*. The genes *vdcC* and *vdcD* encode proteins that catalyze the decarboxylation; *vdcB* is possibly involved in upstream reactions that convert more complex substrates to vanillic acid. Furthermore, the catabolic gene cluster appears to be controlled by the *vdcA* gene, which displays sequence similarity to the translated product of a putative positive transcriptional regulatory gene identified in the *Streptomyces coelicolor* A3(2) genome sequencing project. This is, to my knowledge, the first report of the genes associated with the process of non-oxidative decarboxylation of aromatic acids in a microorganism. Further studies of this system should allow these genes to be incorporated into metabolically engineered microorganisms for the production of industrially useful chemical products such as guaiacol, catechol and adipic acid.

5. DISCUSSION

5.1 Aromatic acid non-oxidative decarboxylases are multi-subunit enzymes

Although there are many reports of microbial non-oxidative decarboxylases active on aromatic acids in the literature, only recently have enzyme purifications been successful, as these proteins appear to be unstable in cell-free extracts. Of the enzymes purified thus far, all share one common feature: they are single polypeptides, which form multi-subunit enzyme complexes. However, depending on substrate and organism, the molecular mass of the polypeptide, as well as the number of subunits, is variable. Several examples of aromatic acid non-oxidative decarboxylases and their subunit configurations are listed in Table 3. I have demonstrated that *Streptomyces* sp. D7 produces a protein of approximately 52 kDa when grown in the presence of vanillic acid, and suggest that two additional functionally related proteins of 36 kDa and 9 kDa are also synthesized. The exact functions of these proteins remain unknown, although it appears that *vdcC* encodes the vanillic acid decarboxylase. Experiments in which *vdcC* and *vdcD* were expressed under the control of the *aph* promoter in pIJ680 resulted in vanillate decarboxylase activity, and it is possible that the 52 kDa VdcC is a subunit of a complex similar to non-oxidative decarboxylases described in the literature (Table 3). The amino-terminus of the VdcC protein is highly similar to the limited amino acid sequence obtained from the purified *p*-hydroxybenzoate carboxy-lyase of *Clostridium hydroxybenzoicum*. The *Clostridium* enzyme was purified and characterized, but only limited amino acid sequence was obtained. The enzyme is responsible for decarboxylation of *p*-hydroxybenzoate to phenol, a dead-end metabolite. With the exception of the amino acid sequence similarity between the vanillic acid induced protein of *Streptomyces* sp. D7 and

Enzyme	Organism	Configuration	Reference
4-hydroxybenzoate decarboxylase	*Clostridium hydroxybenzoicum*	350 kDa (6 subunits of 57 kDa)	He and Wiegel, 1995
3,4-dihydroxybenzoate decarboxylase	*Clostridium hydroxybenzoicum*	270 kDa (5 subunits of 57 kDa)	He and Wiegel, 1996
4,5-dihydroxyphthalate decarboxylase	*Pseudomonas testosteroni*	150 kDa (4 subunits of 38 kDa)	Nakazawa and Hayashi, 1978
4,5-dihydroxyphthalate decarboxylase	*Pseudomonas fluorescens*	420 kDa (6 subunits of 66 kDa)	Pujar and Gibson, 1985
2,3-dihydroxybenzoate decarboxylase	*Aspergillus niger*	120 kDa (4 subunits of 28 kDa)	Kamath *et al.*, 1987
2,3-dihydroxybenzoate decarboxylase	*Trichosporon cutaneum*	66.1 kDa (2 subunits of 36.5 kDa)	Anderson and Dagley, 1981
3,4,5-trihydroxybenzoate decarboxylase	*Pantoea agglomerans* T71	320 kDa (6 subunits of 57 kDa)	Zeida *et al.*, 1998

Table 3: Variations in subunit size and configuration – characteristics of some microbial aromatic acid non-oxidative decarboxylases.

the *C. hydroxybenzoicum* enzyme, it is difficult to assign the exact function(s) of the *vdcC* or *vdcD* gene products in the reaction. Neither polypeptide shows a relationship to any characterized enzymes in the databases, although the amino terminal sequence of the *C. hydroxybenzoicum* enzyme was noted to have weak similarity to uroporphyrinogen decarboxylase of *Synechococcus* sp (He & Wiegel, 1995). However, extending this comparison to the *Streptomyces* sp. D7 protein would be speculative, as the *Synechococcus* sp. uroporphyrinogen decarboxylase amino-terminal sequence bears almost no resemblance to the *Streptomyces* sp. D7 protein. The product of *vdcB* has primary amino acid sequence highly similar to phenylacrylate decarboxylase (PAD) from yeast. In light of this functional implication, it may be possible that the *vdcB* product is involved in transformation of acrylic phenolic compounds to substrates suitable for downstream gene products, possibly VdcC. The putative trans-membrane region noted in the yeast PAD GenBank database entry is highly conserved among the bacterial PAD homologues' amino-terminal regions (Figure 32). Therefore, it is possible that the product of the *vdcB* gene is membrane associated. Further biochemical research is needed to elucidate the true roles of the components of the VDC gene cluster.

5.2 Distribution of the VDC gene cluster among streptomycetes

The results of the Southern hybridization experiments demonstrate that the homologues of the VDC cluster genes are also apparently present in the genome of *S. setonii* 75Vi2. This observation reinforces the notion that the gene cluster described in this thesis is responsible for vanillate decarboxylation (recall that *S. setonii* 75Vi2 is the streptomycete well characterized in the literature for vanillate decarboxylation). Of thirteen

streptomycetes screened in similar Southern blots (results not shown), only three – *S. setonii* 75Vi2, *Streptomyces* sp. D7, and *Streptomyces* sp. 2065 (another environmental isolate from our collection) appeared to carry these genes. This observation suggests that the ability to decarboxylate vanillic acid to guaiacol is not a widespread trait among streptomycetes. Degradation via demethylation to protocatechuate, by the enzyme vanillate demethylase, is likely the more common route for catabolism of vanillic acid in streptomycetes, as well as other prokaryotes. In fact, a cosmid of the *S. coelicolor* A3(2) genome sequencing project contains homologues of the vanillate demethylase genes from *Pseudomonas* sp. It was previously demonstrated, by the colorimetric Rothera assay, that protocatechuate 3,4-dioxygenase is induced in *Streptomyces* sp. D7 in the presence of vanillic acid, suggesting that the initial attack on vanillic acid in this organism is performed by vanillate demethylase as well as vanillate decarboxylase (Chow, 1996). However, it is apparent from the current study that vanillate decarboxylase is very highly expressed in *Streptomyces* sp. D7, as evidenced by the almost equimolar conversion of vanillic acid to guaiacol by *Streptomyces* sp. D7 mycelia. If vanillate demethylase is synthesized in response to vanillate, the levels of that enzyme are likely very low relative to vanillate decarboxylase.

5.3 Substrate specificity

The apparent high substrate specificity of the VDC system is perhaps not surprising in light of other non-oxidative decarboxylase studies. Most decarboxylases studied thus far have been shown to be very specific for one substrate, although several are active against two related aromatic acids. For example, gallate decarboxylase from *Pantoea*

agglomerans T71 is highly substrate specific (Zeida *et al.*, 1998), while *p*-hydroxybenzoate carboxy-lyase from *Clostridium hydroxybenzoicum* is active against both *p*-hydroxybenzoate and protocatechuate (He & Wiegel, 1995). *Klebsiella aerogenes* was biochemically demonstrated to produce a number of non-oxidative decarboxylases, each enzyme specific for a different aromatic acid substrate (Grant & Patel, 1969). It will be interesting in future studies to determine the catalytic sites of these enzymes, in order to elucidate which amino acid residues affect substrate binding.

5.4 Primary structure motifs

Microbial non-oxidative decarboxylase systems reported in the literature (Grant & Patel, 1969; Yoshida & Yamada, 1985; Nakajima *et al.*, 1992; Huang *et al.*, 1993; Santha *et al.*, 1995; He & Wiegel, 1995; He & Wiegel, 1996; Zeida *et al.*, 1998) all have minimal or no requirements of cofactors for activity. The vanillate decarboxylase system of *Streptomyces* sp. D7 appears active in the absence of oxygen. Consistent with this, amino acid sequence analysis of all three VDC gene products failed to reveal the presence of any cofactor binding motifs, including NAD and FAD, which are characteristic of oxidative enzymes.

5.5 Transcriptional activation

The proteomics analyses presented previously (Chow, 1996), combined with the genetic analysis provided in this study, together demonstrate that genes encoding proteins linked to vanillic acid decarboxylation are induced, directly or indirectly, by vanillic acid itself. This observation is supported by the northern blot data, indicating that the VDC gene

cluster is transcribed in one polycistronic mRNA product upon induction by vanillic acid. These results suggest that transcription of the gene cluster is under tight regulatory control. In fact, preliminary nucleotide sequence analysis upstream of the VDC gene cluster revealed a divergent putative regulatory gene. Future work will complete the characterization of the putative regulatory gene and allow elucidation of its relationship to the VDC gene cluster. The mechanism by which vanillic acid enters streptomycete mycelia and activates transcription of the VDC gene cluster should prove to be an interesting model of environmental sensing and response. *Streptomyces* has a high proportion of two component signal transduction regulators. Recent data from the *S. coelicolor* A3(2) genome project predicts a total of 160 two-component regulators, approximately double the number found in any other genome thus far analyzed (D. Hopwood, personal communication).

5.6 Biodegradation -- a result of the microbial community gene pool

The wide distribution of homologues of the VDC genes throughout the microbial world, mostly encoded by chromosomes, but also plasmid-borne as in the case of pNL1 in *Sphingomonas aromaticivorans*, suggests that these gene products provide useful metabolic abilities for their hosts. However, non-oxidative decarboxylation of most aromatic acids yields toxic phenolic compounds and, in many cases cited in the literature, the microorganisms do not possess appropriate mechanisms to further degrade the compounds produced by these dead-end pathways. For example, in this study, *Streptomyces* sp. D7 was able to rapidly convert vanillic acid to guaiacol, but was unable to further degrade the guaiacol. In another example, *Clostridium hydroxybenzoicum* decarboxylated *p*-hydroxybenzoate to phenol, and protocatechuate to catechol without

further metabolism. Non-oxidative decarboxylation remains an enigma of microbial metabolism (Frost & Draths, 1995) and in the case of the *C. hydroxybenzoicum* enzymes, He and Weigel mention that "a direct metabolic function in the bacterium is not known for either of the two enzymes" (He & Weigel, 1996). The functions of these seemingly toxic metabolic reactions of microorganisms are not readily apparent; however, in natural ecosystems, it can be imagined that other organisms in a consortium would mineralize and remove these toxins from the environment. For example, *Streptomyces setonii* 75Vi2 (Pometto III *et al.*, 1981; Sutherland, 1986) and a *Moraxella* sp. (Sterjiades, *et al.*, 1982) were demonstrated to demethylate methoxylated aromatic compounds such as guaiacol to catechol, leading to subsequent mineralization. Evidence implicating cytochrome P-450 systems was provided in both cases. The preliminary sequence data we have obtained, indicating that a cytochrome P-450 gene is located within a few kilobases of the VDC gene cluster in *S. setonii* 75Vi2, suggests that vanillate decarboxylation to guaiacol is associated with demethylation of guaiacol to catechol in that organism. *S. setonii* 75Vi2 was also observed to degrade phenol (Antai and Crawford, 1982), adding to its reputation as a catabolically diverse organism. *Streptomyces* sp. D7 does not appear to be as catabolically diverse as *S. setonii* 75Vi2, and from additional sequencing studies of the ends of *Bam*HI subclones of the 13 kb genomic library clone, does not appear to possess a cytochrome P-450 gene nearby the VDC gene cluster. This observation is in accordance with the fact that the strain does not metabolize guaiacol – that is, guaiacol is a dead end metabolite. One could speculate that while *S. setonii* 75Vi2 possesses a complete complement of lignin-related aromatic acid catabolic genes, *Streptomyces* sp. D7 acquired (from another streptomycete or microorganism) only a partial set. Indeed,

there is a growing focus on biodegradation, not from the standpoint of an individual microorganism, but rather as a coordinated function of the entire gene pool (Wackett, 1999). Efficient biodegradation is therefore likely the result of natural consortiums of microorganisms. The observation that *Streptomyces* sp. D7 converts vanillic acid to the toxic guaiacol, but is unable to remove the guaiacol from its environment appears to support this view. However, a microorganism producing metabolites toxic to itself defies the principles of natural selection. The observation of the production of vast amounts of guaiacol from vanillic acid by *Streptomyces* sp. D7 is based on laboratory experiments, in which unnaturally high concentrations of vanillic acid were fed to the microorganism. In contrast, natural settings such as forest soils most likely do not have high amounts of vanillic acid freely available to the microorganisms. The release of such low molecular weight aromatic compounds from lignin is a rate-limited process, determined by the activity of fungi and abiotic processes. In addition, guaiacol produced by organisms such as *Streptomyces* sp. D7 would most likely be further degraded by other organisms in the environment. Therefore, *Streptomyces* sp. D7 would not be living with toxic levels of guaiacol, even if its vanillate decarboxylase system were a result of the acquisition or deletion of certain catabolic genes from its genome, leading to incomplete degradation of vanillic acid.

5.7 *Sphingomonas aromaticivorans F199* catabolic plasmid pNL1: filling in the missing links

Recent publication of the entire sequence of the 184 kb catabolic plasmid pNL1 of *Sphingomonas aromaticivorans F199* (Romine *et al.*, 1999) sheds additional light on the role of the VDC gene cluster in biodegradation processes. pNL1 is a conjugative

plasmid, which contains 186 open reading frames, 70 of which are likely associated with catabolism or transport of aromatic compounds. On this plasmid are 15 gene clusters encoding biodegradative enzymes. Among genes encoding biphenyl and *p-cresol* degradative enzymes, lie homologues to *vdcB*, *vdcC* and *vdcD*. Amino acid sequence alignments between the *Streptomyces* sp. D7 *vdcB* and *vdcC* genes and the *Sphingomonas* homologues are shown in Figure 27. Between the *vdcB* and *vdcC* homologues, there are two ORFs not found in *Streptomyces* sp. D7: *pchFa* (*p*-cresol methylhydroxylase) and *vdh* (vanillin oxidoreductase).

Figure 32 (following pages): Amino acid sequence comparisons between the *Streptomyces* sp. D7 *vdcB* translation product (strPAD) and its *Sphingomonas* pNL1 homologue (sphPAD) **(a)**, and the *Streptomyces* sp. D7 *vdcC* translation product (strC) and its *Sphingomonas* pNL1 homologue (sphC) **(b)**.

(a)

```
               1                                                  50
   sphPAD      MKRMVVGITG ATGSVYGLRL LELLRETGGW ETHLVMSPAA LLNIREELPE
   strPAD      M.RLVVGMTG ATGAPFGVRL LENLRQLPGV ETHLVLSRWA RTTIELETGL
Consensus      M-R-VVG-TG ATG---G-RL LE-LR---G- ETHLV-S--A ---I--E---

               51                                                100
   sphPAD      GKARLEALAD VVHNVRNVGA SIASGSFVCE GMAIAPCSMR TLGAVAHALS
   strPAD      SVAEVSALAD VTHHPEDQGA TISSGSFRTD GMVIVPCSMK TLAGIRTGYA
Consensus      --A---ALAD V-H-----GA -I-SGSF--- GM-I-PCSM- TL--------

               101                                               150
   sphPAD      DNLITRAADV MLKERRRLVM ITREAPLNLA HLRNMTACTE MGAVIFPPVP
   strPAD      EGLVARAADV VLKERRRLVL VPRETPLSEI HLQNMLELAR MGVQLVPPMP
Consensus      --L--RAADV -LKERRRLV- --RE-PL--- HL-NM----- MG----PP-P

               151                                               200
   sphPAD      AFYARPTSLA DVVDHTCMRV LDLFGLHAKS EKRWQGLSKE AASLVPGAGQ
   strPAD      AFYNNPQTVD DIVDHVVARI LDQFDLPAPA ARRWAGMRAA RAAARSFGDA
Consensus      AFY--P---- D-VDH---R- LD-F-L-A-- --RW-G---- -A--------

                201
   sphPAD      MEGN
   strPAD      A~~~
Consensus      ----
```

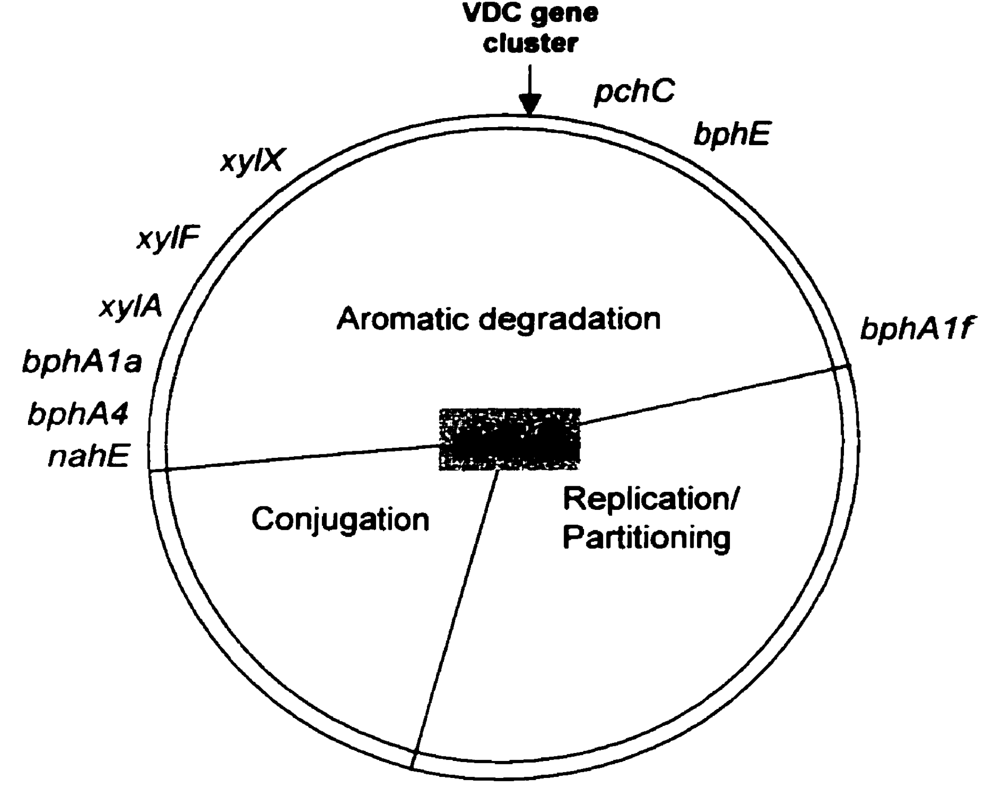
(b)

```
           1                                                 50
 strC      ~~~~~~~~~~ MAYDDLRSFL DTLEKEGQLL RITDEVLPEP DLAAAANATG
 sphC      MTMNDLPNRA RSISSLRDFL ELLEDAGQAI TWSDAVMPEP GVRNIAVAAS
 Consensus ~~~~~~~~~~ -----LR-FL --LE--GQ-- ---D-V-PEP -----A-A--

           51                                               100
 strC      RIGENAPALH FDNVKGFTDA RIAMNVHGSW ANHALALGLP KNTPVKEQVE
 sphC      RDANGAPAIV FDNITGYPGK RLAVGVHGSW DNIALLLGRP KGTTIRELFF
 Consensus R----APA-- FDN--G---- R-A--VHGSW -N-AL-LG-P K-T---E---

           101                                              150
 strC      EFARRW..DA FPVAPERREE APWRENTQEG EDVDLFSVLP LFRLNDGDGG
 sphC      EIAGRWGDQE AQISFVPEAQ APVHE.CRIE QDINLYDVLP VYRINEYDGG
 Consensus E-A-RW---- ---------- AP--E----- -D--L--VLP --R-N--DGG

           151                                              200
 strC      FYLDKAAVVS RDPEDRDDFG KQNVGTYRIQ VIGTNRLAFH PA.MHDVAQH
 sphC      FYIGKASVAS RDPLDPDNFG KQNVGIYRLQ IQGPDTFTLM TIPSHDMGRQ
 Consensus FY--KA-V-S RDP-D-D-FG KQNVG-YR-Q --G------- ----HD----

           201                                              250
 strC      LRKAEEKGED LPIAITLGND PVMAIVAGMP MAYDQSEYEM AGALRGAPAP
 sphC      IMAAEREGVP LKIAVMLGNH PGLAAFAATP IGYEESEYSY ASAMMGAPIR
 Consensus ---AE--G-- L-IA--LGN- P--A--A--P --Y--SEY-- A-A--GAP--

           251                                              300
 strC      IATAPLTGFD VPWGSEVVIE GVIESRKRRI EGPFGEFTGH YSGGRRMPVI
 sphC      LTKSG.NGID ILADSEIVIE AELQPGGREL EGPFGEFPGS YSGVRKAPIF
 Consensus -------G-D ----SE-VIE -------R-- EGPFGEF-G- YSG-R--P--

           301                                              350
 strC      RVERVSYRHE PVFESLYLGM PWNECDYLVG PNTCVPLLKQ LRAEFPEVQA
 sphC      KVTAVSHRRD PIFENIYIGR GWTEHDTLIG LHTSAPIYAQ LRQSFPEVTA
 Consensus -V--VS-R-- P-FE--Y-G- -W-E-D-L-G --T--P---Q LR--FPEV-A

           351                                              400
 strC      VNAMYTHGLM VIISTAKRYG GFAKAVGMRA MTTPHGLGYV AQVILVDEDV
 sphC      VNALYQHGLT GIISVKNRMA GFAKTVALRA LSTPHGVMYL KNLIMVDADV
 Consensus VNA-Y-HGL- -IIS---R-- GFAK-V--RA --TPHG--Y- ---I-VD-DV

           401                                              450
 strC      DPFNLPQVMW AMSAKVNPKD DVVVIPNLSV LELAPAAQPA GISSKMIIDA
 sphC      DPFDLNQVMW ALSTRTR.AD DIIVLPNMPA VPIDPSAVVP GKGHRLIIDA
 Consensus DPF-L-QVMW A-S------D D--V-PN--- ----P-A--- G-----IIDA

           451                                   489
 strC      TTPVAPDVRG NFSTPAKDLP ETAEWAARLQ RLIAARV~~
 sphC      TSYLPPDPVG EAHLVTPPTG DEIDALSKRI REMQLGALS
 Consensus T----PD--G ---------- ---------- R--------~
```

The *pchFa* gene was originally isolated from *Pseudomonas putida* NCIMB 9866, and encodes an enzyme that oxidizes the methyl group of *p*-cresol to an aldehyde. The *vdh* gene was originally isolated from *Pseudomonas fluorescens* AN103 (Walton, Genbank, 1997), and encodes an NAD^+-dependent oxidoreductase, which converts the aldehyde vanillin to vanillic acid. These genes are followed by homologues to *vdcC* and *vdcD*, for which this thesis gives evidence (pKCS7 in *S. lividans* 1326) that the gene products in *Streptomyces* sp. D7 are involved in vanillic acid non-oxidative decarboxylation. The location of the *vdc* gene homologues in relation to other genes on pNL1 is shown in Figure 33. From this one can create a catabolic scenario for these genes on plasmid pNL1, that phenylacrylate derivatives are transformed to aromatic aldehydes, which are oxidized to aromatic acids, then non-oxidatively decarboxylated to catechol derivatives for mineralization via catechol dioxygenase ring cleavage enzymes. In fact, the *meta*-cleavage enzyme catechol 2,3-dioxygenase is found on pNL1. It should be noted that the translation products of the *pchFa*, *vdh*, *vdcB* and *vdcC* homologues on pNL1 were approximately 55% identical to their *Streptomyces* sp. D7 and *Pseudomonas* spp. counterparts, and therefore may be similar in function but not substrate specificity. For example, the pNL1 genes may encode non-oxidative, decarboxylative enzymes that transform toluene derivatives instead of lignin-related aromatic compounds. *Streptomyces* sp. D7 did not contain the aldehyde oxidoreductase gene found on pNL1, and therefore would not be expected to carry through the bioconversion of phenylacrylates to aromatic acids. Indeed, *S. lividans* 1326 carrying pKCS8 effectively decarboxylated vanillic acid to guaiacol, no observable biotransformation activity was noted when the organism was exposed to ferulic acid, the acrylate derivative of vanillic

acid. The role of the VDC gene cluster for *Streptomyces* sp. D7 in nature remains to be determined, and one issue that arises from this study is whether or not the organism acquired the cluster through horizontal gene transfer events during evolution. Standard plasmid isolation procedures performed on *Streptomyces* sp. D7 did not reveal the presence of any plasmids, but this issue should be investigated further in future studies, perhaps using pulsed field gel electrophoresis techniques. Another question is whether the organism originally carried all the genes necessary for vanillic acid mineralization, but lost some downstream catabolic genes (such as those for the degradation of guaiacol) through deletion events. Finally, it is possible that the VDC gene cluster encodes enzymes for which nature did not intend to decarboxylate vanillic acid, but rather another substrate not tested in the work presented here. *Streptomyces* sp. D7 could be able to degrade a yet to be determined complex substrate with a vanillic acid-like degradation intermediate, which is processed by the genes described in this thesis work.

A.



B.

pNL1 ORF	Functional description of closest relative	% identity to *B. subtilis* ORF
orf1244	Phenylacrylic acid decarboxylase homolog	50
orf1272	conserved hypothetical protein	47
orf1280	conserved hypothetical protein	32

FIGURE 33: pNL1 of *Sphingomonas aromaticivorans*. (a) Physical map of pNL1 and the location of the VDC gene cluster homologues; (b) listing of the pNL1 open reading frames similar to ORFs from the *Bacillus subtilis* genome (From Romine, *et al.*, 1999)

5.8 Metabolic Engineering Applications for Vanillate Decarboxylase

The elaborate metabolic engineering scheme described for *E. coli* AB2834, which enables the production of adipic acid using D-glucose as a starting material, is an example of what can be achieved by mixing and matching catabolic genes in an appropriate host microorganism (Draths & Frost, 1994).

Similarly, the genes encoding the components of the vanillate decarboxylase system of *Streptomyces* sp. D7 could be used in engineered systems for the transformation of lignin-rich plant and timber waste to useful commodity or fine chemicals. Engineering such a system would probably not be as complex as the development of *E. coli* AB2834 for conversion of D-glucose to adipic acid. The by-products of wood pulp and other plant manufacturing processes, such as olive oil production (Ramos, J., personal communication), consist of large amounts of methoxylated aromatic chemicals. In theory, these aromatic chemicals could be separated from the waste stream and refined to industrially useful chemicals such as catechol and *cis,cis*-muconic acid (for drug syntheses), adipic acid (for nylon), or guaiacols (for medicines and scent compounds). To enable this process, a microorganism would have to be modified with a suite of genes encoding enzymes to produce the desired end products. For the refining of vanillate, a microorganism expressing vanillate decarboxylase would be an efficient producer of guaiacol, while a strain expressing genes for vanillate decarboxylase and guaiacol-demethylating Cytochrome P-450 (cloned and sequenced from *Streptomyces setonii* 75Vi2 by M.K. Pope and S. Baily in this laboratory) would produce catechol. If this catechol producing strain is supplemented with the gene for catechol 1,2-dioxygenase,

adipic acid could be produced. A diagram of this proposed scheme is shown in Figure 34. With the current low prices for petroleum products, commercialization of such biorefining technology is not practical as the desired products can be inexpensively produced from benzene. However, with the continually changing supply and demand of petroleum, and with increasingly stringent environmental regulations being imposed on industry, biorefining technology may become not only practical, but necessary.

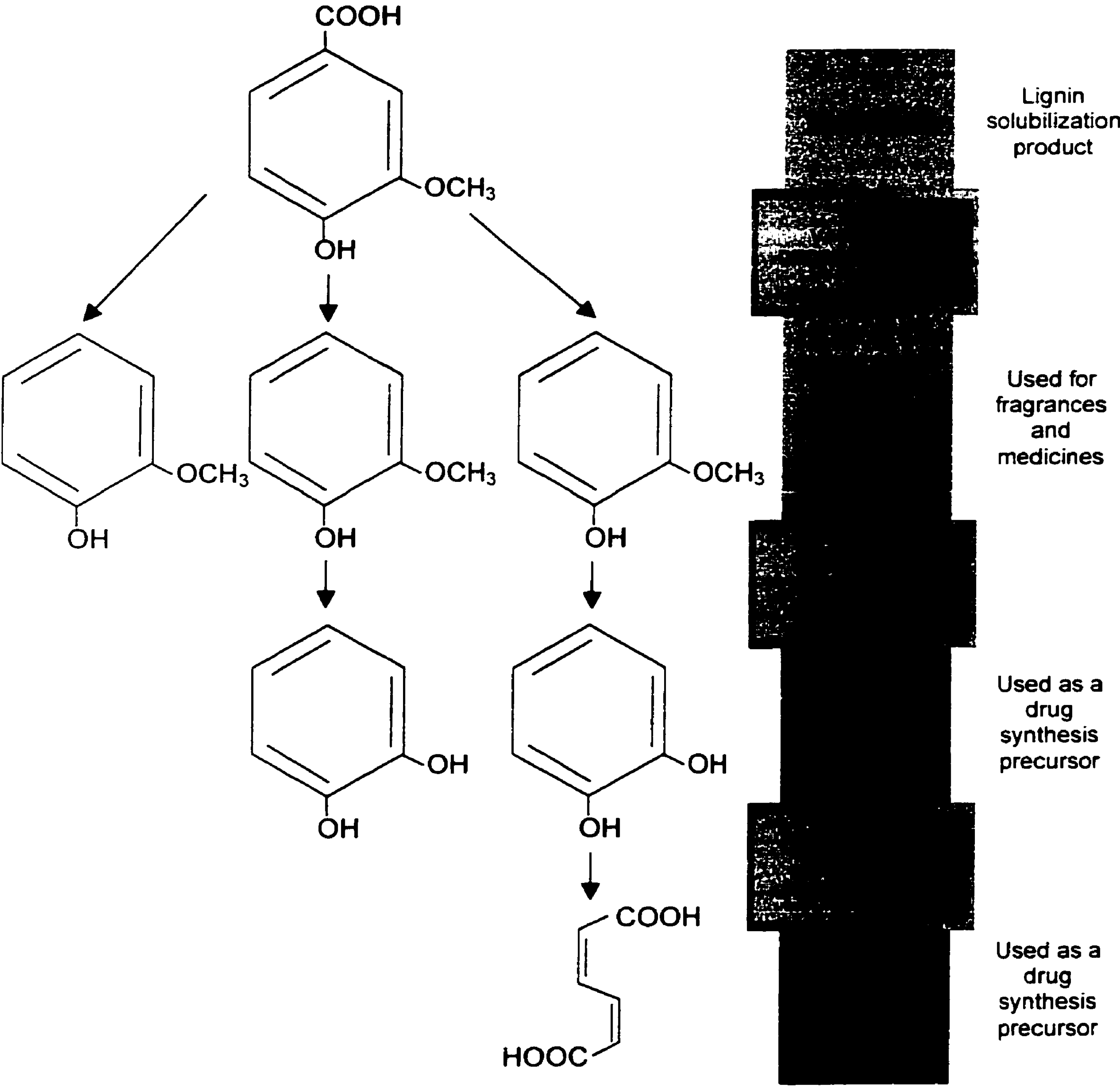


FIGURE 34: Vanillate decarboxylase, Cytochrome P-450 and catechol 1,2-dioxygenase can be used in combinations to produce various industrially useful chemicals from vanillic acid as a starting material.

5.9 Future directions

The cloning of the genes required for vanillic acid non-oxidative decarboxylation, as presented in this work, will allow us to expand our knowledge of the reaction. Purification of active recombinant protein products encoded by the gene cluster will enable detailed reaction kinetics to be determined for the decarboxylation process. Future studies will shed light on the structure and catalytic function of vanillic acid decarboxylase from *Streptomyces* sp. D7. It will be interesting to determine if the numerous sequence homologues from microbial genome databases are indeed other non-oxidative decarboxylases. The transcriptional regulation of the VDC gene cluster in response to environmental stimuli will be another interesting aspect to investigate. Accordingly, one experiment planned for the near future will be to clone the entire VDC gene cluster, including the divergent putative regulatory gene, into a suitable plasmid expression vector for expression in *S. lividans* 1326. It has already been demonstrated by 2D-PAGE and mRNA analyses that in *Streptomyces* sp. D7, the vanillate decarboxylase gene cluster is inducible in response to vanillic acid. A recombinant *S. lividans* 1326 strain carrying the VDC gene cluster and the associated regulatory gene should also demonstrate vanillic acid inducibility, unless the regulatory system requires other components not present in *S. lividans* 1326. Non-oxidative decarboxylases represent connections between the two major branches of aromatic acid catabolism, characterized by either catechol or protocatechuate central intermediates (Figure 35). By joining these two pathways, this class of enzymes supports the concept of aromatic catabolism as being a web of interconnecting biodegradative processes, and not separate, distinct pathways (Crawford and Olson, 1978). From an applied standpoint, knowledge gained from such

Lignin

CO_2H

OCH_3

OH

vanillate

VDC

OCH_3

OH

guaiacol

VanA, VanB

Cytochrome P-450

CO_2H

PDC

OH

OH

protocatechuate

OH

OH

catechol

PcaG, PcaH

CatA

mineralization

mineralization

FIGURE 35: Non-oxidative decarboxylases represent connections between the two major branches of aromatic acid catabolism, characterized by either catechol or protocatechuate central intermediates. *VDC* = vanillate decarboxylase; *PDC* = protocatechuate decarboxylase; VanA, VanB = vanillate demethylase; PcaG, PcaH = protocatechuate 3,4-dioxygenase; CatA = catechol 1,2-dioxygenase

endeavors should ultimately allow this class of enzyme to be used for a variety of metabolic engineering applications. Currently, the world's major chemical companies are refocusing their business strategies towards producing many bulk commodity chemicals via microbial fermentations (Alper, 1999). Genetic engineering of *E. coli* has already proven that aromatic acid decarboxylases are valuable tools for biomass conversion pathway development. Vanillate decarboxylase could be useful as a component in the engineered bioconversion of complex lignin molecules, through intermediates such as ferulic acid, down to industrially useful low molecular weight compounds such as catechol, adipic acid (Frost & Draths, 1995) and *cis,cis*-muconic acid (Yoshikawa *et al.*, 1990).

6. LITERATURE CITED

Alper, J. (1999). Engineering metabolism for commercial gains. Science 283, 1625-1626.

Altschul, S.F., Gish, W., Miller, W., Myers, E.W. & Lipman, D.J. (1990). Basic local alignment search tool. J Mol Biol 215, 403-410.

Anderson, J.J. & Dagley, S. (1981). Catabolism of tryptophan, anthranilate, and 2,3-dihydroxybenzoate in *Trichosporon cutaneum*. J Bacteriol 146, 291-297.

Antai, S.P. & Crawford, D.L. (1983). Degradation of phenol by *Streptomyces setonii*. Can J Microbiol 29, 142-143.

Atlas, R. & Bartha, R. (1993). Survey of microorganisms, p. 506-532. *In* Microbial Ecology: Fundamentals and Applications. Benjamin/Cummings, Redwood City.

Bibb, M.J., Ward, J.M. & Hopwood, D.A. (1978). Transformation of plasmid DNA into *Streptomyces* protoplasts at high frequency. Nature 274, 398-400.

Brunel, F. & Davison, J. (1988). Cloning and sequencing of *Pseudomonas* genes encoding vanillate demethylase. J. Bacteriol 170, 4924-4930.

Chow, K (1996). Two-dimensional polyacrylamide gel electrophoretic analysis of protein synthesis during aromatic acid catabolism by *Streptomyces violaceusniger* M.Sc. thesis.

Chow, K.T., Pope, M.K. & Davies, J. (1999). Characterization of a vanillic acid non-oxidative decarboxylation gene cluster from *Streptomyces* sp. D7. Microbiology 145, 2393-2403.

Crawford, R.L. & Olson, P.R. (1978). Microbial catabolism of vanillate: decarboxylation to guaiacol. Appl and Environ Microbiol 36, 539-543.

Draths, K.M. and Frost, J.W. (1994). Environmentally compatible synthesis of adipic acid from D-glucose. J. Am. Chem. Society 116, 399-400.

Frost, J.W. & Draths, K.M. (1995). Biocatalytic syntheses of aromatics from D-glucose: renewable microbial sources of aromatic compounds. Ann. Rev. Microbiol. 49, 557-79.

Garrels, J.I. (1979). Two dimensional gel electrophoresis and computer analysis of proteins synthesized by clonal cell lines. J Biol Chem 254, 7961-7977.

Grant, D.J.W. & Patel, J.C. (1969). The non-oxidative decarboxylation of *p*-hydroxybenzoic acid, gentisic acid, protocatechuic acid and gallic acid by *Klebsiella aerogenes* (*Aerobacter aerogenes*). Antonie Leeuwenhoek 35, 325-343.

Grund, E., Knorr, C. & Eichenlaub, R. (1990). Catabolism of benzoate and monohydroxylated benzoates by *Amycolatopsis* and *Streptomyces* spp. Appl and Environ Microbiol 56, 1459-1464.

Gusek, T.W. & Kinsella, J.E. (1992). Review of the *Streptomyces lividans*/vector pIJ702 system for gene cloning. Crit. Rev. Microbiol. 18, 247-260.

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

He, Z. & Wiegel, J. (1996). Purification and characterization of an oxygen-sensitive, reversible 3,4-dihydroxybenzoate decarboxylase from *Clostridium hydroxybenzoicum*. J Bacteriol 178, 3539-3543.

Hopwood, D.A. personal communication.

Hopwood, D.A., Bibb, M.J., Chater, K.F., Kieser, T., Bruton, C.J., Kieser, H.M., Lydiate, D.J., Smith, C.P., Ward, J.M. & Schrempf, H. (1985). Genetic manipulation of *Streptomyces:* a laboratory manual. The John Innes Foundation, Norwich.

Huang, Z., Dostal, L. & Rosazza, J.P.N. (1993). Mechanisms of ferulic acid conversions to vanillic acid and guaiacol by *Rhodotorula rubra*. J Biol Chem 268, 23954-23958.

Kamath, A.V., Dasgupta, D. & Vaidyanathan, C.S. (1987). Enzyme-catalyzed non-oxidative decarboxylation of aromatic acids: I. Purification and spectroscopic properties of 2,3-dihydroxybenzoic acid decarboxylase from *Aspergillus niger*. Biochem Biophys Res Comm 145, 586-595.

Katz, E., Thompson, C.J. & Hopwood, D.A. (1983). Cloning and expression of the tyrosinase gene from *Streptomyces antibioticus* in *Streptomyces lividans*. J Gen Microbiol 129, 2703-2714.

Kirby, K.S., Fox-Carter, E. & Guest, M. (1967). Isolation of deoxyribonucleic acid and ribosomal ribonucleic acid from bacteria. Biochem J 104, 258-262.

Kirk, T.K. (1987). Enzymatic "combustion": the microbial degradation of lignin. Ann Rev Microbiol 41, 465-505.

Nakajima, H., Otani, C. & Niimura, T. (1992). Decarboxylation of gallate by cell-free extracts of *Streptococcus faecalis* and *Klebsiella pneumoniae* isolated from rat feces. J Food Hyg Soc Japan 33, 371-376.

Nakazawa, T. & Hayashi, E. (1978). Phthalate and 4-hydroxyphthalate metabolism in *Pseudomonas testosteroni:* purification and properties of 4,5-dihydroxyphthalate decarboxylase. Appl Environ Microbiol 36, 264-269.

Nishikawa, S., Sonoki, T., Kasahara, T., Obi, T., Kubota, S., Kawai, S., Morohoshi, N. & Katayama, Y. (1998). Cloning and sequencing of the *Sphingomonas* (*Pseudomonas*) *paucimobilis* gene essential for the O demethylation of vanillate and syringate. Appl and Environ Microbiol 64, 836-842.

Oh, S.H. and Chater, K.F. (1997). Denaturation of circular or linear DNA facilitates targeted integrative transformation of *Streptomyces coelicolor* A3(2): possible relevance to other organisms. J Bacteriol 179(1):122-127.

Oritz, M.L., Calero, M., Patron, C.F., Castellanos, L & Mendez, E. (1992). Imidazole-SDS-Zn reverse staining of proteins in gels containing or not SDS and microsequence of individual unmodified electroblotted proteins. FEBS Lett 296, 300-304.

Pometto III, A.L., Sutherland, J.B. & Crawford, D.L. (1981). *Streptomyces setonii*: catabolism of vanillic acid via guaiacol and catechol. Can J Microbiol 27, 636-638.

Pujar, B.G. & Gibson, D.W. (1985). Phthalate metabolism in *Pseudomonas fluorescens* PHK: purification and properties of 4,5-dihydroxyphthalate decarboxylase. Appl Environ Microbiol 49, 374-376.

Romine, M.F., Stillwell, L.C., Wong, K.-K., Thurston, S.J., Sisk, E.C., Sensen, C.W., Gaasterland, T., Saffer, J.D., Fredrickson, J.K. & Saffer, J.D. (1999). Complete sequence of a 184 kb catabolic plasmid from *Sphingomonas aromaticivorans* strain F199. J. Bacteriol 181, 1585-1602.

Santha, R., Savithri, H.S., Rao, A. & Vaidyanathan, C.S. (1995). 2,3-dihydroxybenzoic acid decarboxylase from *Aspergillus niger*. A novel decarboxylase. Eur J Biochem 230, 104-110.

Segura A. & Ornston, N.L. (1997). pZR135: vanillate demethylase region in *Acinetobacter*. NCBI Entrez database submission (unpublished), accession AF009672.

Sterjiades, R., Sauret-Ignazi, G., Dardas, A. & Pelmont, J. (1982). Properties of a bacterial strain able to grow on guaiacol. FEMS Microbiol Lett 14, 57-60.

Sutherland, J.B., Crawford, D., & Pometto III, A.L. (1981). Catabolism of substituted benzoic acids by *Streptomyces* species. Appl and Environ Microbiol 41, 442-448.

Sutherland, J.B. (1986). Demethylation of veratrole by cytochrome P-450 in *Streptomyces setonii*. Appl and Environ Microbiol 52, 98-100.

Thomas, L & Crawford, D.L. (1998). Cloning of clustered *Streptomyces viridosporus* T7A lignocellulose catabolism genes encoding peroxidase and endoglucanase and their extracellular expression in *Pichia pastoris.* Can J. Microbiol 44, 364-372.

Thompson, C.J., Ward, J.M. & Hopwood, D.A. (1980). DNA cloning in *Streptomyces*: resistance genes from antibiotic-producing species. Nature 286, 525-527.

Thompson, C.J., Ward, J.M. & Hopwood, D.A. (1982). Cloning of antibiotic resistance and nutritional genes in streptomycetes. J Bacteriol 151, 668-677.

Wackett, L.P., Ellis, L.B.M., Speedie, S.M., Hershberger, C.D., Knackmuss, H-J., Spormann, A.M., Walsh, C.T., Forney, L.J., Punch, W.F., Kazic, T., Kanehisa, M. & Berndt, D.J. (1999). Predicting microbial biodegradation pathways. ASM News 65, 87-93.

Walton, N.J. (1998). Pseudomonas fluorescens genes encoding p-hydroxycinnamoyl CoA hydratase/lyase and vanillin: NAD+ oxidoreductase (unpublished). GenBank accession Y13067.

Yoshida, H. & Yamada, H. (1982). Microbial production of pyrogallol through decarboxylation of gallate. Agric Biol Chem 49, 659-663.

Yoshikawa, N., Mizuno, S., Ohta, K. & Suzuki, M. (1990). Microbial production of *cis,cis*-muconic acid. J Biotechnol 14, 203-210.

Zeida, M., Wieser, M., Yoshida, T., Sugio, T. & Nagasawa, T. (1998). Purification and characterization of gallic acid decarboxylase from *Pantoea agglomerans* T71. Appl and Environ Microbiol 64, 4743-4747.