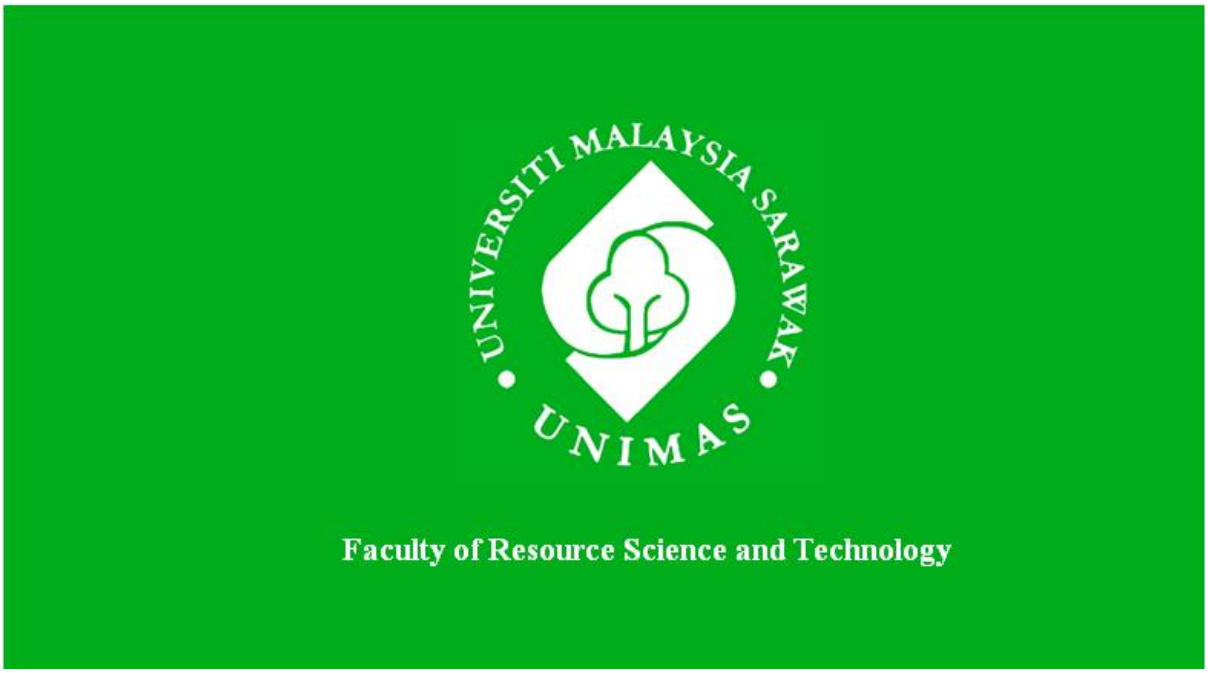

ISOLATION OF GENES RELATED TO HETEROCYCLIC COMPOUND DEGRADATION FROM MARINE BACTERIA

Ong Leng Hui (27927)

Bachelor of Science with Honours

(Resource Biotechnology)

2012

ISOLATION OF GENES RELATED TO HETEROCYCLIC COMPOUND DEGRADATION FROM MARINE BACTERIA

ONG LENG HUI (27927)

This project submitted in partial fulfillment of the requirement of the degree of Bachelor of Science with Honours (Resource Biotechnology)

SUPERVISOR: DR AZHAM ZULKHARNAIN

Resource Biotechnology Programme

Department of Molecular Biology

Faculty of Resource Science and Technology

University Malaysia Sarawak

2013

Acknowledgement

I would like to express my gratitude and appreciation to my project supervisor, Dr Azham Zulkharnain who had provided me the chance of learning and studying on the project, shared with me his valuable experiences and knowledge based on the project and given me the brilliant suggestion and advice throughout the whole project. Besides, I would like to show my gratitude to Azizah Ahmad, the research assistant in UNIMAS for giving me full support in the process of experiment. Apart from that, I would like to thank all the postgraduate students who had helped me in my laboratory work.

My gratitude also goes to my course-mates who had encouraged me and shared with me their precious finding. Also, my parents and my two siblings had given me unlimited mental support when completing the final year project, and so I would like to make use of this opportunity to show my appreciation.

Declaration

I hereby declare that this final year project have been completed on my own original work and effort and has not been submitted for any award nomination. This research is carried out under the supervision of Dr Azham Zulkharnain. Where other sources of information were used, they have been acknowledged.

NAME: ONG LENG HUI

SIGNATURE:

DATE: May 9, 2013

Contents

List of Abbreviation

A	Ampere
BPH	Biphenyl
˚C	Degree Celsius
DBF	Dibenzofuran
DNA	Deoxyribonucleic acid
dH_2O	Distilled water
EDTA	Ethylenediaminetetraacetic acid
EtBr	Ethidium bromide
g	Gram
h	Hour
MB	Marine broth
min	Minutes
ml	Mililitre
µl	Microliter
PCR	Polymerase Chain Reaction
rpm	Round per minute
s	Second
SDS	Sodium Laurieth Sulphate
V	Volt

List of tables

List of figure

ISOLATION OF GENES RELATED TO HETEROCYCLIC COMPOUND DEGRADATION FROM MARINE BACTERIA

Ong Leng Hui

Resource Biotechnology
Department of Molecular Biology
Faculty of Resource Science and Technology
Universiti Malaysia Sarawak

ABSTRACT

Dibenzofuran terminal dioxygenase gene or *dbfA1* gene and biphenyl dioxygenase gene *bphA1* gene are the genes responsible for the first step of aromatic compound metabolism found in various microorganisms to metabolize dibenzofuran and biphenyl respectively. *Thalassospira profundimaris* strain M01 and *Pseudomonas pachastrella* strain M03 are the bacteria that had been previously identified to have the capability of degrading biphenyl and dibenzofuran respectively. The purpose of this study is to detect the presence of reported heterocyclic compound degrading genes (*bphA1*, *dbfA1*) in these bacteria. Degenerative primers were designed based on the consensus regions of dibenzofuran and biphenyl degrading genes on reported strains. These primers were designed by multiple alignments of genes coding for reported dioxygenases, in order to detect the presence of these genes in bacteria *T. profundimaris* strain M01 and *P. pachastrella* strain M03 respectively via PCR. A total of 4 sets of primers were design designated bphA1F1R1, bphA1F2R2, dbfA1F1R1, and dbfA1F2R2 respectively. The primer set dbfA1F2 and dbfA1R2 produced bands on agarose gel with desired size of approximately 1,000 bp indicating the presence of *dbf*A1 gene.

Keywords: Alpha subunit terminal dioxygenase gene (*dbfA1)*, biphenyl dioxygenase gene (*bphA1)*, degenerative primers.

ABSTRAK

Gen terminal dioxygenase dibenzofuran (dbfA1) dan gen biphenyl dioxygenase (bphA1) merupakan gen-gen yang menghasilkan enzim untuk memecahkan kompoun aromatic, dibenzofuran dan biphenyl yang terdapat dalam mikroorganisasi yang berupaya untuk memecahkan kompoun-kompoun tersebut. Thalassospira profundimaris strain M01 dan Pseudomonas pachastrella strain M03 ialah bakteria yang mempunyai keupayaan untuk memecahkan dibenzofuran dan biphenyl masing-masing. Tujuan kajian ini adalah untuk mengenal pasti keadaan gen-gen yang biasa memecahkan kompoun heterocyclic dalam dua bakteria ini. Primer degenerasi telah direka dengan keputusan multiple alignment daripada gen-gen yang telah dilapor untuk mengenalpasti persamaan dalam gen-gen tersebut. Lalu setiap primer digunakan dalam PCR untuk mendapati gen yang dikaji. Empat primer telah direka, bphA1F1R1, bphA1F2R2, dbfA1F1R1,dan dbfA1F2R2, satu primer terdapat keputusan positif iaitu dbfA1F2R2 yang menghasilkan produk PCR bersaiz sebanyak 1000bp iaitu saiz produk sasaran.

Kata kunci: Gen Alpha subunit terminal dioxygenase (dbfA1), gen biphenyl dioxygenase (bphA1), primer degenerasi.

1.0 Introduction

Dibenzolfuran (DBF) is one of the members of heterocyclic compound which is generated and released to the atmosphere through the process of incineration (Happe *et al*., 1993), it is one of the components of heat transfer oil, dyeing and printing textiles and a varieties of polymers (Elvers *et al*., 1989). It has a high range of coverage throughout the commercial products of combustions as it is highly heat resistance and for these reasons, it has also becoming the most abundance and persistent pollutant in the atmosphere. Human may come into contact with DBF through inhalation as DBF will be released to the air after burning or combustion and also through ingesting contaminated food product, also through the contaminated water source.

Metabolism of DBF by microorganisms is an effective approach for the treatment of DBF contamination, and therefore it is important to understand the degrading mechanism of the microorganisms. The DBF degradation pathways of these microorganisms were studied by scientists all around the world (Becher *et al*., 2000; Fortnagel *et al.,* 1989). Besides this, the enzyme responsible for each degradation step and the genes encoding these enzymes were identified and cloned (Happe *et al.,* 1993; *Strubel et al*., 1991). Angular dioxygenation in DBF the process by which the carbon adjacent to the aromatic and the carbon bonded to the heteroatoms are both oxidized, it is also the first step and most important step in the metabolism of DBF and some other heterocyclic compound (Nojiri *et al*., 2001).

Within the DBF angular dioxygenation, gene cluster, four genes are found, *dbfA1, dbfA2, dbfA3* and *dbfA4* which code for alpha subunit of terminal dioxygenase, beta subunit of dioxygenase, ferredoxin and ferredoxin reductase (Iida *et al*., *2002*). The arrangement of these genes is investigated by Iida *et al* (2002) and the relative position of

angular dioxygenation gene cluster is shown in Figure 1. After the angular dioxygenation, DBF will be turn into an unstable form of hemiacetal-like intermediate which will spontaneously converted into 2,2',3-trihydroxydiphenyl ether (Nojiri *et al.*, 2001), Figure 2 shows the product of angular dioxygenation of DBF.

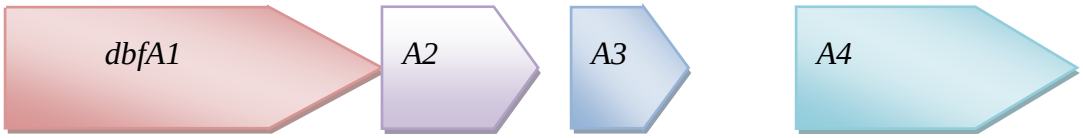

Figure 1: *dbfA* gene cluster, *dbfA1*, *A2*,and *A3* are the genes coding for alpha subunit of terminal dioxygenase, beta subunit of dioxygenase, ferredoxin and ferredoxin reductase respectively. (Iida *et al.*, 2002)

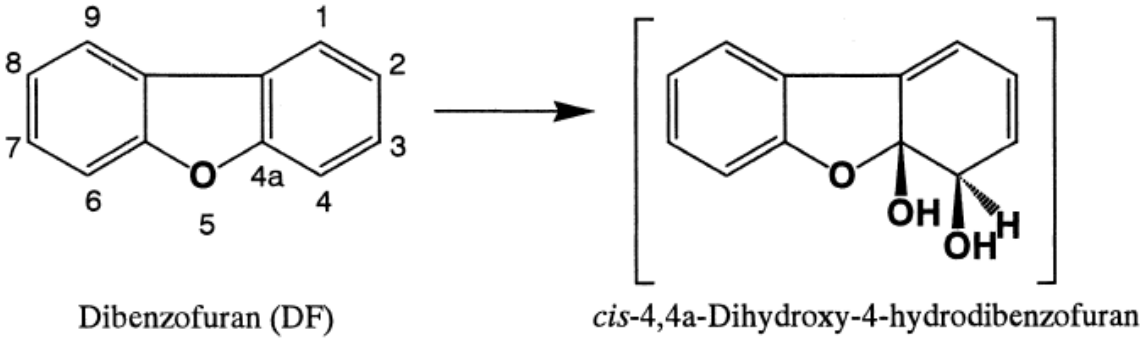

Figure 2: Product of angular dioxygenation of DBF (unstable intermediate) (Nojiri *et al.*, 2001)

Other than DBF, polychlorinated biphenyl (PCB) is also one of the most concerned environmental pollutants. The usage of PCB is wide, for example as the nonflammable dielectrics in heavy electrical transformers, pigment suspension agents on the carbonless copy paper, vehicles for pesticide application and also the dielectric in assortment of small electronic parts (Rogan *et al.*, 1986). Similar to DBF, PCB can be treated with bioremediation, by using biphenyls metabolizing microorganism (Ahmad & Focht, 1973; Furukawa & Matsumura, 1976).

Besides the studies of isolating and identifying the microorganism metabolizing biphenyls, its pathways of metabolizing, enzymes focused, and also the genes responsible for the production of those enzymes are investigated (Furukawa *et al*., 1979; Furukawa *et al*., 1978). One of the reported biphenyl metabolizing pathway are shown in Figure 3 together with the gene cluster that contains genes code for the enzymes involve in the pathway. Based on the figure, a dihydrodiol is produced by the introduction of molecular oxygen at the 2,3-position of the nonchlorinated ring by the enzyme biphenyl dioxygenase A encoded by *bphA* gene, and then it is dehydrogenated to a 2,3-dihydroxybiphenyl by dihydrodiol dehydrogenase encoded by *bphB* gene, subsequently the 2,3-dihydroxybiphenyl is cleaved at the 1,2-position by 2,3 dihydroxybiphenyl dioxygenase, product of *bphC* gene. The *meta*-cleavage compound (a chlorinated derivative of 2-hydroxy-6-oxo-6-phenylhexa-2,4-dienoate) is hydrolyzed to the corresponding chlorobenzoic acid by hydrolase (product of *bphD*) (Furukawa & Miyazaki, 1986).

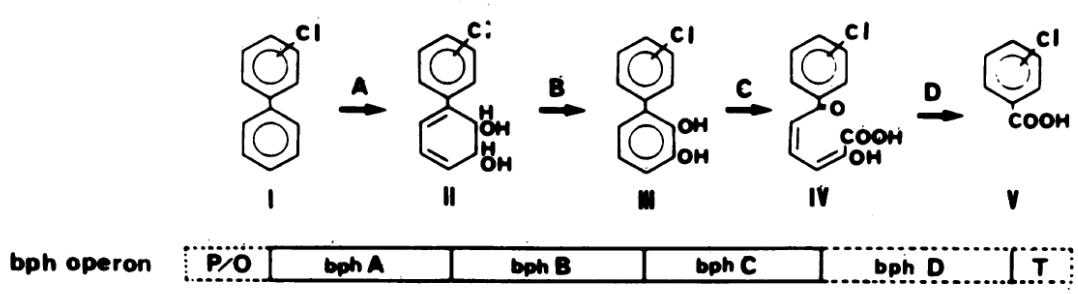

Figure 3: The biphenyls degradation pathway and its gene cluster. I: biphenyl, II: dihydrodiol, III: 2,3-dihydroxybiphenyl, IV: chlorinated derivative of 2-hydroxy-6-oxo-6-phenylhexa-2,4-dienoate, V: chlorobenzoic acid; A: biphenyl dioxygenase A, B: dihydrodiol dehydrogenase, C: 2,3-dihydroxybiphenyl dioxygenase, D: hydrolase (Furukawa & Miyazaki, 1986)

In this study, the marine bacteria *Thalassospira profundimaris* strain M01 and *Pseudomonas pachastrella* strain M03 have been previously isolated from the sea water sample from coast of Miri, Sarawak, and identified with the capability to metabolize biphenyls and DBF as sole carbon source respectively. The presence of the gene codes for large subunit of biphenyl dioxygenase, *bphA1*, and the gene codes for alpha subunit of terminal dioxygenation, *dbfA1* in the two bacterial strains metabolizing the respective compounds were detected by means of PCR with the degenerative primers designed based on the consensus regions found from the multiple alignment of previously reported genes.

1.1 Objectives

1. To design degenerative primers for the amplification of *bphA1* and *dbfA1* from *Thalassospira profundimaris* strain M01 and *Pseudomonas pachastrella* strain M03 respectively.
2. To detect the genes using designed degenerative primers by Polymerase Chain Reaction.
3. To analyze the result of the Polymerase Chain Reaction.

2.0 Literature Review

2.1 Degenerative primers

A degenerative primer is the PCR primer having several possible bases in some of its positions (Linhart & Shamir, 2002). It is often designed, either by manual approach or using bioinformatics tools in order to isolate the gene from a species that have not been sequenced. The genes sequences from its related species are retrieved and their consensus region is located through comparison with multiple alignments, these regions are the targeted priming sites and the primers are designed complementary to the nucleotides in these sites.

Degenerative primer have been widely utilized and one of the examples is the experiment of Rojas *et al* (1993) who uses degenerative primers to detect whitefly-transmitted geminiviruses, DNA of symptomatic plant sample is extracted and as the geminiviruses replicate in DNA form its DNA should also be extracted as well, PCR using degenerative primers is done to amplify and detect the consensus region of geminivirus to determine the infected samples. Besides that, according to the paper of Iserte *et al* (2013), degenerative primers can not only be designed based on DNA sequences but also protein sequences. The protein sequences will be back translated into its DNA sequence and the following downstream processing is the same with the DNA sequences primer design algorithm. Designing primers based on the protein sequence can bring to the discovery of a novel gene function when the protein is coded by an unknown gene.

There are a number of primer design programmes, for example, OLIGO (Rychlik & Rhoads, 1898), OSP (Hiller & Green, 1991; Li *et al*., 1997), Primer Master (Proutski & Holmes, 1996), PRIDE (Haas *et al*., 1998), and many others. Some of the programmes

designing specific primers where the complete nucleotide sequence of the template is needed to be known and it binds to a unique sequence while some designing degenerative primers where it contains a collection of unique sequences, it covers all the possible combinations coding for a targeted protein (Iserte *et al.*, 2013).

Besides the common concern in designing primers for example the GC content, annealing temperature, and the length of the primers, here are two major problem needed to be considered by the primer designer, the maximum coverage (number of matched sequences), and the degeneracy of the designed degenerative primers, the degeneracy of a degenerative primer refers to how many unique combination of primers it have (Linhart & Shamir, 2002). The degeneracy of the designed primers should be large enough maximum number of match sequences, however, a large degeneracy also lead to the amplification of undesired sequences which cause the downstream processes tie and cost inefficient. Therefore the main objective of degenerate primer design is to maintain the balance between maximum coverage and optimum degeneracy (Linhart & Shamir, 2002).

2.2 Nucleic acids notation- IUPAC Code

The IUPAC code or International Union of Pure and Applied Chemistry code is a list of nomenclature of incompletely specified nucleic acids, though it is widely known as IUPAC code, the actual origin of this code is the International Union of Biochemistry and Molecular Biology (IUBMB) (Johnson, 2010). In this study, the IUPAC code serve the purpose of creating degenerative primers where the ambiguous regions of nucleotides was replaced with it matching IUPAC code so that sequence of many primers can be written as one. The IUPAC codes are listed in Table 1.

Due to the lack of indicator for nucleotides proportion in the conventional IUPAC code, an extended version of IUPAC code was proposed by Johnson (2010) to amend this limitation, capital letters and small letters were used to distinguish the proportion differences though the nucleotides contained are the same. The extended version of IUPAC codes are listed in Table 2 (degenerate by 2 nucleotides) and Figure 4 (Degenerate by 3 nucleotides). However, the extended IUPAC only shows the relative abundance among the nucleotides where the exact proportion of these nucleotides is still no means to be cited.

Table 1: IUPAC codes (Johnson, 2010)

	Nucleotides	Symbol
Adenine	A	A
Cytosine	C	C
Guanine	G	G
Thymine	T	T
Uracil	U	U
Purine	A,G	R
Pyrimidine	C,T(U)	Y
Amino group	A,C	M
Keto group	G,T(U)	K
Strong interaction	C,G	S
Weak interaction	A,T(U)	W
Not G	A,C,T(U)	H
Not A	C,G,T(U)	B
Not T(U)	A,C,G	V
Not C	A,G,T(U)	D
Any	A,C,G,T(U)	N

Table 2: Extended IUPAC codes for combination of two nucleotides (Johnson, 2010)

Code	Nucleotide represented
R	A> G
r	G> A
R	A= G
Y	C>T(U)
y	T/U > C
Y	C = T/U
S	C > G
s	G > C
S	C = G
W	A > T/U
w	T/U > A
W	A = T/U
K	G > T/U
k	T/U > G
K	G = T/U
M	A > C
m	C > A
M	A= C

	Translation					
	Code	1°		2°		3°
C–G–T/U combinations (Not A)	B	C	>	G	>	T/U
	B	C	>	G	=	T/U
	B	C	=	G	=	T/U
	b	T/U	>	G	>	C
	b	T/U	=	G	>	C
	L	G	>	T/U	>	C
	L	G	>	T/U	=	C
	l	C	>	T/U	>	G
	l	C	=	T/U	>	G
	O	T/U	>	C	>	G
	O	T/U	>	C	=	G
	o	G	>	C	>	T/U
	o	G	=	C	>	T/U
A–G–T/U combinations (Not C)	D	A	>	G	>	T/U
	D	A	>	G	=	T/U
	D	A	=	G	=	T/U
	d	T/U	>	G	>	A
	d	T/U	=	G	>	A
	E	G	>	T/U	>	A
	E	G	>	T/U	=	A
	e	A	>	T/U	>	G
	e	A	=	T/U	>	G
	F	T/U	>	A	>	G
	F	T/U	>	A	=	G
	f	G	>	A	>	T/U
	f	G	=	A	>	T/U
A–C–T/U combinations (Not G)	H	A	>	C	>	T/U
	H	A	>	C	=	T/U
	H	A	=	C	=	T/U
	h	T/U	>	C	>	A
	h	T/U	=	C	>	A
	I	C	>	T/U	>	A
	I	C	>	T/U	=	A
	i	A	>	T/U	>	C
	i	A	=	T/U	>	C
	J	T/U	>	A	>	C
	J	T/U	>	A	=	C
	j	C	>	A	>	T/U
	j	C	=	A	>	T/U
A–C–G combinations (Not T/U)	V	A	>	C	>	G
	V	A	>	C	=	G
	V	A	=	C	=	G
	v	G	>	C	>	A
	v	G	=	C	>	A
	X	C	>	G	>	A
	X	C	>	G	=	A
	x	A	>	G	>	C
	x	A	=	G	>	C
	Z	G	>	A	>	C
	Z	G	>	A	=	C
	z	C	>	A	>	G
	z	C	=	A	>	G

Figure 4: Extended IUPAC code for combination of three nucleotides (Johnson, 2010)

2.3 Heterocyclic compound

Heterocyclic compounds are compound with oxygen, sulphur, phosphorous and nitrogen in the aromatic ring. Together with polycyclic compound, heterocyclic compounds form aromatic compound while 2/3 of the known aromatic compounds are heterocyclic (Kaiser, 1996). It is comprised by most geological sources for instance the petroleum, coal and oil shale. In the case of oil spill on the sea, these substances including the heterocyclic and polycyclic compound will be released to the ocean.

The water solubility of heterocyclic compounds are relatively high while its sorption to soil and aquifer materials are weak, so these compound have the tendency to be transported to the subsurface environment (Hale & Aneiro, 1997). When released to the environment, aromatic compound become the most prevalent and persistent pollutant (Seo, 2009). As it take a very long time for the natural microbial degradation process to accomplish as the concentration and the strain of the microbe is not optimum.

Bioremediation methods like biodegradation is progressively being looked into by the researcher in order to solve the environmental problem brought about by the heterocyclic compound, microbes like bacteria is isolated and multiplied to metabolize these compounds into another harmless and environmental product.

2.4 Dibenzofuran

Dibenzofuran (DBF) is one of the members of dioxins. It is a very toxic and mutagenic compound. It is a white crystal-like solid which becomes hazardous pollutant when released to the environment (Lida *et al*., 2002). DBF are formed during herbicides production (in the process of producing halogen-contain aromatics), during the combustion of dust and also during the bleaching process of pulp at paper mills (Monna *et al.*, 1993).

DBF can be found in coke dust, grate ash, fly ash and flame soot, and also it is used as an insecticide (Agency for Substances and Disease, n. d).

Human body takes in DBF through inhaling contaminated air, as the coal tar is the source of DBF, and it is one of the main constituent of smoking cigarettes, smoking is one of the main cause of DBF accumulation in the body, also DBF will be absorbed into the body through skin contact (Agency for Substances and Disease, n. d). Several symptoms like skin, eye, noses and throat irritation have been reported to be caused by short-term exposure to DBF (Agency For Substances and Disease, n. d). Due to the strong toxicity and wide range of sources, DBF is one of the major contaminant needed to be dealt with for every country.

Bacteria strains that degrade DBF have been isolate and the gene that encodes for the enzyme involving in the pathway are cloned and studied in order to understand the major pathway of degrading DBF (Bayly *et al*., 1966; Engesser *et al*., 1990; Klecka, & Gibson, 1979; Klecka & Gibson, 1980). Strubel *et al* (1991) studied *Brevibacterium* sp strain DPO 1361 and proposed a DBF degrading pathway, it is shown in Figure 5, the end product of this pathway is 2-hydroxy-4-pentenoate before entering tricarboxylic acid cycle. On the other hand, Monna *et al* (1992) studied *Staphylococcus auriculans* DBF63 which is a Gram-positive bacteria that can grow on either fluorene or DBF as sole carbon source, in order to identify the enzyme products and understand the degradation pathway (Monna *et al.*, 1993).

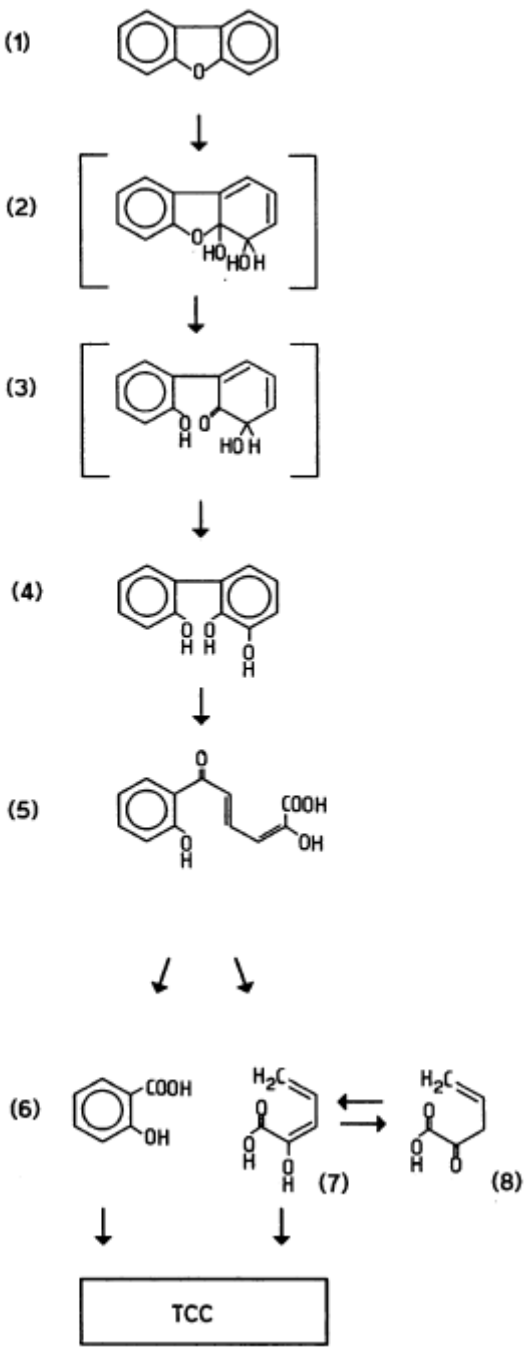

Figure 5: DBF degenerative pathway (Strubel *et al.*, 1991).

Based on Figure 5, the DBF degeneration pathway of bacteria strain Brevibacterium sp. Strain DPO 1361 proposed by Strubel *et al* (1991), where the DBF is broken down into 2,4,4*a*-dihydro-4,4*a*-dihydroxydibenzofuran before it turn into a keto tautomer HPC, followed by the convertion to 2'-OH-HOPDA, then it is converted into salicylate, followed

by 2-oxo-4-pentenoate, and before entering the tricarboxylic acid cycle it is in the form of 2-hydroxy-4-pentenoate.

2.5 Polychlorinated biphenyls

Polychlorinated biphenyls or PCBs is synthesized by the process of chlorination of biphenyls catalysed by ferric or iron (Safe, 1992), it is produced as a component of lubricant, heat-exchange fluid, insulators, plasticisers in paints, synthetic resin and plastic (Alvers *et al*., 1973). It is a very useful industrial chemical credited to its variable physical properties, chemical stability, dieletric properties, inflammability and miscibility with organic solvent (Safe, 1992).

However, it is also a well-known environmental pollutant as it is toxic, bio-concentrated, carcinogenic and also persistent once enter the living body system (Furukawa & Miyazaki, 1986). Very often the PCBs residue is found in the tissue of fishes and wildlife (Alvares *et al*., 1973) as the result of improper disposal of its final products, and when these animals are consumed as food by human being, it accumulates in our body. Human milk and adipose tissue have been reported for containing PCBs (Alvares *et al*., 1973).

The effect of PCB on living organism have been widely studied and one of the finding is that PCB is responsible for vitamin A and thyroid hormone deficiency in seal (*phoca vitulia*) based on the study of Brouwer *et al* (1989), these deficiency in turn increase the susceptibility of microbial infection, reproductive disorder and pathological alterations, Brouwer *et al* suggested this as the cause of reproductive disorder and lethal viral infection in seals and other marine mammal populations in Baltic, North Sea and Wadden Sea that was reported during that time.

In the year of 1996, a study on the intellectual impairment in children exposed to PCB in utero had been carried out by (Rogan *et al*., 1986). This is a research based on 212 children who were having higher prenatal exposure to PCBs compared to normal populations, an IQ and achievement test was carried out when they are in the age of 11, to compare their intellectual ability with the normal population. The result proved that In utero exposure to higher concentration PCB will have long term effect of the individual intellectual function.

Whilst PCBs bringing great harms to human beings and all other living things, a number of biphenyl-utilizing bacteria are capable to metabolize various components of PCB (Ahmed & Focht, 1973;Catelani *et al*., 1971;Furukawa & Matsumura, 1976; Furukawa *et al*., 1978). Four enzymes are taking part in the major biphenyl degrading pathway, biphenyl dioxygenase, dihydrodiol dehydrogenase, 2,3-dihydroxybiphenyl dioxygenase, and meta cleavage compound hydrolase, they are the products of gene *bph*A, *bph*B, *bph*C, and *bph*D respectively. The final compound of this pathway is the chlorobenzoic acid (Furukawa & Miyazaki, 1986). One of the examples of studied PCB breaking down bacteria is the *Pseudomonas pseudoalcaligenes* reported by Furukawa and Miyazaki (1986) where the gene cluster coding for the degradation of biphenyl and chlorobiphenyl was isolated and cloned.

2.6 Polymerase Chain Reaction (PCR)

Polymerase chain reaction is a reaction able to amplify specific segment of DNA (Hendry, 1991). This is an *In vitro* method of the enzymatic synthesis of DNA sequences based on the template DNA, using two oligonucleotide primers that hybridize to opposite strands and flank the region of interest in the targeted DNA (Hendry, 1991).

The PCR consists of three main processes, the denaturation, annealing and extension. The double stranded DNA is denatured into single stranded. Then, a set of primer is annealed to the template DNA which the nucleotide sequence is compatible to those primers. The extension is followed by the *Taq* polymerase which elongates the primer strand and eventually become another template strand for the next cycle of PCR. As the number of the DNA product doubles in each cycle, the quantity of product will reach more than a million.

In general, the temperature for the three processes is set at 94˚C, 50˚C, and 70˚C respectively. However this is subjected to change according to the type of PCR conducted, and also the primer used in the reaction. There are various types of PCR used in the molecular biology work, like the RT-PCR, hot-start PCR, touch-down PCR and so on.