To the Graduate Council:

I am submitting herewith a thesis written by Yang Wang entitled "Cloning and expression of a putative P450 from *Mycobacterium ulcerans.*" I have examined the final paper copy of this thesis for form and content and recommend that it be accepted in partial fulfillment of the requirements for the degree of Master of Science, with a major in Microbiology.

Pamela Small, Major Professor

We have read this thesis
and recommend its acceptance:

Accepted for the Council:

Vice Chancellor and
Dean of Graduate Studies

Thesis
2005
.W36

Cloning and expression of a putative P450 from *Mycobacterium ulcerans*

A Thesis
Presented for the
Master of Science
Degree
The University of Tennessee, Knoxville

Yang Wang
August 2005

Dedication

I dedicate this work to my family. Thank you for the love, encouragement and laughter you have given me over the years. This would not have been possible without your support.

Acknowledgements

I would like to thank Dr. Pamela Small for allowing me to pursue this degree at the microbiology department and for her patience, support and encouragement during the process. I would also like to thank the other members of my committee: Dr.Todd Renold and Dr. Barry Bruce for their suggestions and assistance during this project.

In addition, I would like to acknowledge Dr. Nitin Jain for his assistance with the CO difference spectrum. And thanks to Michael Vaughn and Brian Ranger for their in -house technical support and consulting. Special thanks to David McWilliams for mass spectrometry.

I would like to thank everyone at Dr. Small's lab and department of microbiology for the support you have given me during my time at UT. I appreciate your kindness and practical advice on more than just my research.

I would also like to thank my friends: Shanming and Chongle for keeping me positive and helping me whenever I needed anything.

And last but not least. Thank you to my family for unconditional love.

Abstract

Mycolactones are macrocyclic polyketide toxins produced by the pathogen *Mycobacterium ulcerans*, the etiologic agent of the emerging human disease known as Buruli ulcer. A giant virulence plasmid in *Mycobacterium ulcerans* encoding giant polyketide synthases is responsible for the synthesis of the lipid toxin mycolactone. *Mycobacterium ulcerans* from different geographic origins produce varieties of mycolactones including mycolactone A/B, C E. Their difference is observed by thin layer chromagraph, mass spectrometry, cytopathic assays. The presence of different mycolactone correlates with plasmid variation. Australian strains lacking the hydroxyl group at C-12' produce a mycolactone with a mass of [M + Na]+ at m/z 749, called mycolactone C. In consistency, plasmid from Australian strains has the absence of a region that includes the gene encoding a P450 hydroxylase. The product of this gene is predicted to hydroxylate the mycolactone side chain at C-12' to produce mycolactone A/B with a mass of [M + Na]+ at m/z 765. In order to know whether the P450 in the plasmid is related with the hydroxy group at C-12' in mycolactone A/B, BAC 1707 sequence was blasted by NCBI ORF and one probable P450 open reading frame was found, designated P450A. The P450A sequence was cloned into expression vector PET30a from a BAC1707. The growth parameters controlling efficient expression of heme-containing P450 holoenzyme in *E.coli* has been tested. But most of the target

protein was in the inclusion body. A new sequence, designated P450B, was identified by multi alignment which was 135 base pair shorter than the P450A. The P450B was cloned into pCWori vector and overexpressed. The P450B was purified from the soluble cytosolic fraction to electrophooretic homogeneity by affinity binding. The purified protein sequence was confirmed by mass spectrometry .The CO-reduced difference spectrum with peak at 420 nm suggests the purified protein is an inactive form of P450.

Table of Contents

Chapter 1. Introduction.....1

1.1 Buruli Ulcer.....1

1.1.1 Pathogenesis and Clinical Presentation of Buruli Ulcer Disease.....1

1.1.2 Epidemiology and Transmission of Buruli Ulcer.....3

1.1.3 Diagnosis of Buruli Ulcer and Identification of *Mycobacterium ulcerans*.....5

1.1.4 Management of Buruli Ulcer Disease.....5

1.1.5 Prevention of Buruli Ulcer Disease.....6

1.2 Mycolactone.....7

1.2.1 The Structure and Heterogeneity of Mycolactone.....7

1.2.2 Genetics of Biosynthesis of Mycolactone.....12

1.2.3 Plasmid Variation Correlates with the Variety of Mycolactone.....12

1.3 Cytochrome P450 Monooxygenases.....14

1.3.1 Gene Organization and Evolutionary History.....16

1.3.2 Structural Features.....17

1.3.3 Function in Biological Systems.....20

1.4 Objective.....21

Chapter 2. Materials and Methods.....22

2.1 Subcloning of P450A Gene into PET-30a Vector.....22

2.2 Direct Colony PCR Screening of Ligation Product.....23

2.3 Transformation of P450-PET-30a and Plating.....23

2.3.1 Induction and Growth.....24

2.3.2 SDS-PAGE and Coomassie Brilliant Blue Detaining.....24

2.3.3 Purification of P450 from Inclusion Bodies.....24

2.4 Generation of P450A-(his)$_6$-PET30a Construct.....26

2.4.1 Expression and Purification of P450-(his)$_6$-PET30a Protein.....27

2.5 Construction of P450B-(his)$_6$ into pCWori Vector.....28

2.5.1 Direct Colony PCR Screening of Ligation Product.....29

2.5.2 Sequence of the P450B-(his)$_6$-pCWori.....29

2.6 Expression and Purification of P450B-(his)$_6$-pCWori Vector.....29

Chapter 3. Results.....30

3.1 From Gene Isolation to Protein Purification.....30

3.1.1 Clonings of P450A in PET30 vector.....30

3.1.2 Cloning, Expression and Purification of P450B in pCWori vector.....33

3.2 Protein Characterization 40

3.2.1 Sequence Comparisons of the 50Kda Protein.....40

3.2.2 Spectroscopic Characterization.....44

3.2.3 Mass Spectrometric Analysis of the 50 Kda Protein.....45

Chapter 4. Discussion.....51

4.1 From Gene Isolation to Protein Purification.....51

4.2 Protein Characterization 52

Reference.....53

Vita.....57

List of Tables

Table **Page**

1. The identified proteins and their sequence coverages and spectrum counts.....50

List of Figures

Figure **Page**

1. The progression of the Buruli ulcer lesion from the preulcerative stage to the ulcerative stage.....2

2. Buruli ulcer on the right arm of a young male farmer.....4

3. Silica TLC of ASLs from *Mycobacterium ulcerans* isolates from Africa, Australia, Mexico, and Asia.....10

4. The proposed possible reactions of P450 from *M. ulcerans* of converting mycolactone C into mycolactone A/B.....11

5. Domain and module organization of the mycolactone PKS genes.....13

6. PCR analysis of 10 different *M. ulcerans* strains for the presence of eight plasmid-associated genes..... 15

7. The phylogenetic tree of the putative cytochrome P450 M.ulcerans with its top 10 homologous proteins.....18

8. A ribbon representaion of P450eryF crystal structure.....19

9. Soluble and insoluble protein comparison in cell line BL21, BLR, HMS174.....32

10. Overexpression and purification of P450A.....34

11. The sequence alignment of P450A, P450B and their homologous proteins.....35

12. The scheme of the molecular method of cloning of P450B.....37

13. PCR result and enzyme restriction analysis of PCR product.....38

14. Screening of the positive *E. coli* colonies.....39

15. Overexpression and purification of P450B at induction temperatures 30°C and 15°C. 41

16. Purification of P450 and western blot.....42

17. Alignment of P450B sequence with the well-characterized *Pseudomonas putida* P450cam and *Saccaropolyspora erythraea* P450EryF.....43

18. Carbon monoxide absorption difference spectrum of P450B-$(his)_6$.....46

19. Liquid chromatography-tandem mass spectrometry analysis.....47

20. Sequence coverage of the P450B-$(his)_6$ recombinant protein by the LC-MS^2 analysis..... 49

List of Abbreviations

ALA	aminolevulinic acid
ASL	acetone soluble lipid
BCG	bacillus Calmette-Guerin
BLAST	Basic Local Alignment Search Tool
bp	base pair
CO_2	carbon dioxide
DNA	deoxyribonucleic acid
dNTP	deoxynucleotidetriphosphate
IPTG	isopropyle- β -D-thiogalactoside
IS2404	insertion element 2404
IS2406	insertion element 2406
kb	kilobases
Kda	kilo Dalton
LB	Luria-Bertani
BMW	broad molecular weight
M	mole per liter
mg	milligram
min	minute
ml	milliter
mol	mole
m/z	mass/charge
NADPH	β -nicotinamidadenindinucleotidephosphate
Nm	nanameter
OD	optical density
ORF	open reading frame
P450-Mul	P450 from *Mycobacterium Ulcerans*
PCR	polymerase chain reaction
PKS	polyketide synthase
rpm	rounds per minute
SDS	sodium dodecyl sulfate
SDS-PAGE	sodium dodecyl sulfate-polyacrylamide gel electrophoresis
TAE	Tris/acetate/EDTA buffer
TE	Tris/EDTA buffer
TEMED	N,N,N',N'-tetramethyl-ethylendiamine
TLC	thin layer chromatography
UV	ultraviolet
VIS	visible

°C	degree Celsius
ug	microgram
ul	microliter

Chapter 1
Introduction

1.1 Buruli Ulcer

1.1.1 Pathogenesis and Clinical Presentation of Buruli Ulcer Disease

Buruli ulcer, a disease caused by *Mycobacterium ulcerans*, has emerged as a major cause of human suffering since the 1980's. The infectious organism comes from the family of Mycobacterium, which includes *Mycobacterium tuberculosis* and *Mycobacterium leprae*. It is believed that Sir Albert Cook was the first one to describe the disease in his "Mengo Hospital Notes" in 1897 [1]. However, the connection between the ulcers and *M. ulcerans* was not established until 1948. Buruli ulcer is the third most common mycobacterial infection in healthy people, after tuberculosis and leprosy, and the most poorly understood of these three diseases.

This disease is characterized by extensive, progressive necrotizing skin lesions and the lack of an acute inflammatory response（Figure 1）. Buruli ulcer can present in ways unique to different endemic regions. In Africa, the first stage of this disease is the formation of a firm, on-tender nodule or a plaque, and its second stage is the formation of a large open ulcer, or ulceration [2]. Histopathology of Buruli ulcer shows necrosis of

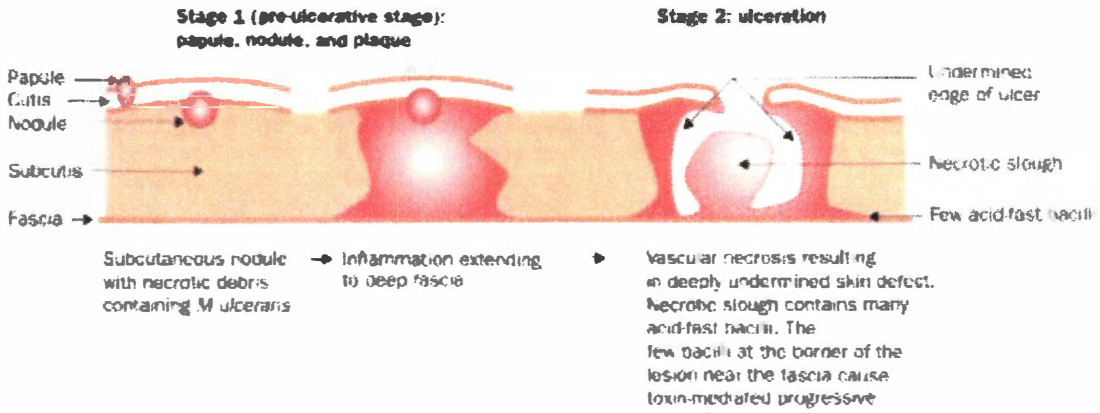


Figure 1. The progression of the Buruli ulcer lesion from the preulcerative stage to the ulcerative stage. (Adapted from van der Werff, et al.,1999.)

the subcutaneous fat with vascular damage, which results in sloughing and secondary ulceration of the overlying skin with typically undetermined edges (Figure 2). Another presentation of this disease is osteomyelitis. In Australia, the first sign of the disease is an intradermal papule or pimple. Australia and Mexico have never reported osteomyelitis, which is very common in Benin and West Africa. Also, clinical evidence suggests that *M. ulcerans* strains isolated from Asia, Mexico, and Australia may be less virulent than those isolated from Africa [3].

1.1.2 Epidemiology and Transmission of Buruli Ulcer

In the 1960's, many cases of Buruli ulcer were found in Buruli region of Uganda and hence the disease was named after this region. Currently Buruli ulcer is most common in West Africa, including all countries along the Gulf of Guinea, and an unexplained increase of its occurrence in this region has been reported recently. It is also found in marshy parts of the tropical and sub-tropical regions of Africa, Asia, Latin America and the Western Pacific. Sporadic cases have been reported or suspected in Angola, Australia, Benin, Bolivia, Burkina Faso, Cameroon, China, Congo, Côte d'Ivoire, Democratic Republic of Congo, Equatorial Guinea, French Guyana, Gabon, Ghana, Guinea, India, Indonesia, Japan, Liberia, Malaysia, Mexico, Papua New Guinea, Peru, Sierra Leone, Sri Lanka, Sudan, Suriname, Togo and Uganda. By now, more than 30 countries have reported this disease [4].



Figure 2. Buruli ulcer on the right arm of a young male farmer. (Adapted from van der Werff, et al., 1999.)

1.1.3 Diagnosis of Buruli Ulcer and Identification of *Mycobacterium ulcerans*

In endemic areas, Buruli Ulcer is diagnosed by the clinical presentation of ulcerating lesions. Its early presentation is the painless ulcer without determined edges, accompanied by a necrotic slough. Lesions are located mostly on the limbs and occasionally on the trunk or face. The patients usually do not have fever, or other inflammation. These symptoms are distinct from other bacterial infection, such as staphylococcal or streptococcal pyogenic infections [3]. Detection of acid-fast bacilli has been used in diagnosis, but neither positive nor negative results can confirm or rule out the diagnosis. Culture can not be used as a routine diagnosis because it is very expensive and time consuming and, most importantly, it often yields false-negative results [5]. PCR has been used in research settings and reference laboratories for diagnosis. Two insertion sequences (ISs), IS2404 and IS2606, are highly specific for *M. ulcerans*. IS2606 is detected in only one other species, *M. lentiflavum*. Over 40 copies of IS2606 and over 50 copies of IS2404 have been found in the genome [6]. The use of PCR probes to identify *M. ulcerans* is highly sensitive and specific. However, it has not been used as a routine diagnosis yet.

1.1.4 Management of Buruli Ulcer Disease

Surgery is the most effective treatment for this disease. Both small pre-ulcerative lesions and extensive ulcerative lesions require surgery [7]. Because it is difficult to

diagnose early infections, the surgical treatment can not be performed in very early stages, which will not need skin grafting. In order to prevent persistent infection and recurrence of this disease, the excision not only includes the necrotic tissue, but also extends to marginal healthy tissue [7].

Many antibacterial drugs such as Dapsone, streptomycin and rifampicin show very good results in vitro, and some combinations of them are effective in animal models, but none of them are effective in patients with Buruli ulcer disease [7]. The combination of antimycobacterial drugs and chemotherapy in treatment of patients with Buruli ulcer disease is not very effective. However, oral anti-microbial therapy is effective in patients with early nodular or very small ulcerations [7].

Heat treatment and hyperbaric oxygen treatment have been tried. However, considering the high cost and the expensive equipment, these methods have not been widely utilized or extensively evaluated for their effectiveness.

1.1.5 Prevention of Buruli Ulcer Disease

Bacillus Calmette–Guérin vaccine (BCG vaccine) has an incomplete protective effect against Buruli ulcers. At the present time, BCG vaccination is the only biomedical intervention that may help control Buruli ulcer in the highly affected areas [8]. Early

detection and prompt surgical treatment are the best interventions available until other means of prevention and treatment are identified.

1.2 Mycolactone

1.2.1 The Structure and Heterogeneity of Mycolactone

A variety of human diseases are caused by organisms in the genus *Mycobacterium.* A major virulence determinant of *M. tuberculosis*, *M. leprae* and *M. marinum* is their ability to survive and replicate within human macrophages. In contrast, *M. ulcerans* is an extracellular infection and there is a limited inflammatory response to the infection. Based on the extensive necrosis caused by *M. ulcerans*, as early as 1965, Connor and Lunn suggested a diffusible substance, such as an exotoxin, to be the primary cause of the ulcers [9]. In 1974, several studies have demonstrated that the sterile filtered *M. ulcerans* culture supernatant had cytopathic effects [10-12]. In animal tests using guinea pigs, the intra-dermal injection of the sterile filtrate could cause small ulcer-like lesions in their skin. In 1998, George, et al. were able to partially purify the toxin causing this effect and identify the mechanism by which it functioned[12]. They found a cytotoxic activity associated with acetone-soluble lipids present in an organic extract from *M. ulcerans* sterile filtrate.

The cytopathic effect of *M. ulcerans* on L 929 murine fibroblasts was characterized by

flow cytometric analysis which showed that *M. ulcerans* or acetone soluble lipids arrested cells in the G0/G1 stage of the cell cycle, accompanied by cytoskeletal rearrangement with [12]. In 1999, the toxin was completely purified and its structure was elucidated by mass spectral analysis and two-dimensional nuclear magnetic resonance spectral analysis and then confirmed by total synthesis [13-16].

This molecule, designated mycolactone, is a polyketide-derived macrolide from a class of molecules that are often biologically active. This was the first toxin isolated from a *mycobacteria* species, as well as the first complex polyketide isolated from a pathogenic mycobacteria [13]. It was later revealed that the compound exists as a mixture of two cis/trans isomers, mycolactone A/B, in equilibrium at an approximately 3:2 ratio. Mycolactone A/B are represented by a single peak at mass-to-charge ratio (m/z) 765.5 upon mass spectroscopic analysis [17].

Analysis of liquid chromatography/mass spectrometry (LC/MS) revealed multiple heterogeneous congeners, which primarily differ in their number of double bonds and hydroxyl groups. One of the congeners gave a prominent [M+Na]+ at m/z 750 and a weak pseudo-molecular ion at m/z 727.5, corresponding to a mycolactone congener with one less oxygen atom. Another one gave a prominent ion at m/z 747.5, which, presuming it is also a sodium adduct, would correspond to mycolactones A and B with one less

oxygen atom and two less hydrogen atoms [18]. In 2003, the congener with peaks at m/z 750 (Figure 3) was confirmed with thin layer chromatography analysis and mass spectrometry of acetone soluble lipids. [19] With the assumption that this ion has sodium adduct, this species corresponds to mycolactone A/B, lacking an oxygen atom, which is consistent with the reduction in polarity observed by TLC (Figure 3). They also showed this congener was biologically active and designated it mycolactone C [19].

Mycolactone C is the predominant mycolactone in eight Australian strains analyzed and one of the minor mycolactone in African strains. In 2004, the structure of mycolactone C was determined by total synthesis, which corresponds to C2`-deoxymycolactones A and B (Figure 4) [20]. The structure of mycolacone C supports the notion that mycolactone C is a biosynthetic precursor of mycolactones A and B.

Mycolactone A/B is the most potent of the *M. ulcerans* macrolides and mycolactone C has been shown to be 10,000 fold less cytopathic than mycolactones A and B [19, 21], since Cytopathic assays (CPA) of mycolactone A/B is 0.01ng/ml, while CPA of mycolactone C is 800 ng/ml (CPA was defined as the minimal concentration of ASLs per milliliter necessary to produce 90% cell rounding in 24 h and loss of the monolayer by 48 h)[21]. This difference in potency may partially explain the distinct and less severe pathology of Australian Buruli ulcer compared with Buruli ulcer from West African

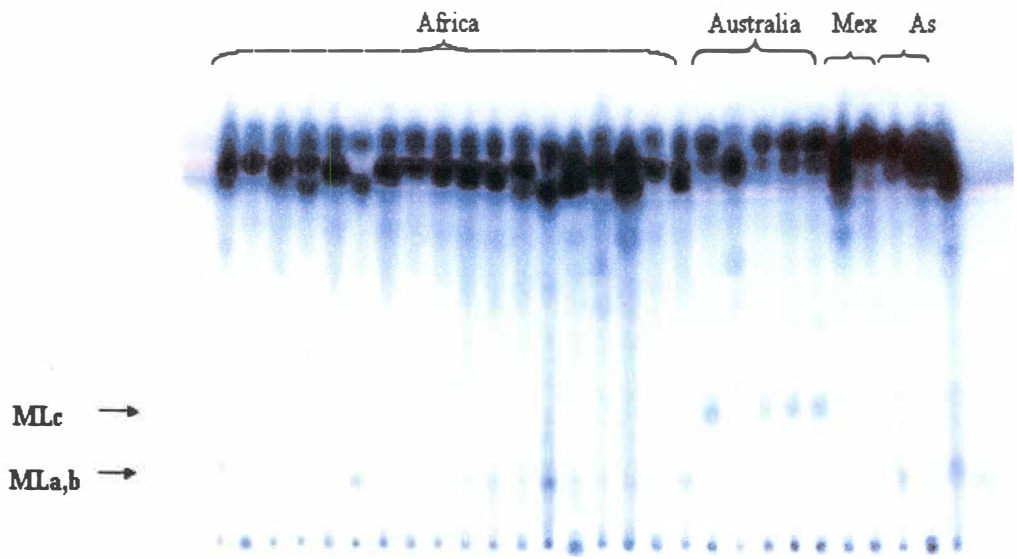


Figure 3. Silica TLC of ASLs from *Mycobacterium ulcerans* isolates from Africa, Australia, Mexico, and Asia. (Adapted from Armand Mve-Obiang, et al 2003) Silica TLC was run in chloroform-methanol-water (90:10:1, vol/vol/vol) and visualized by oxidative charring in a ceric molybdate-10% sulfuric acid stain.

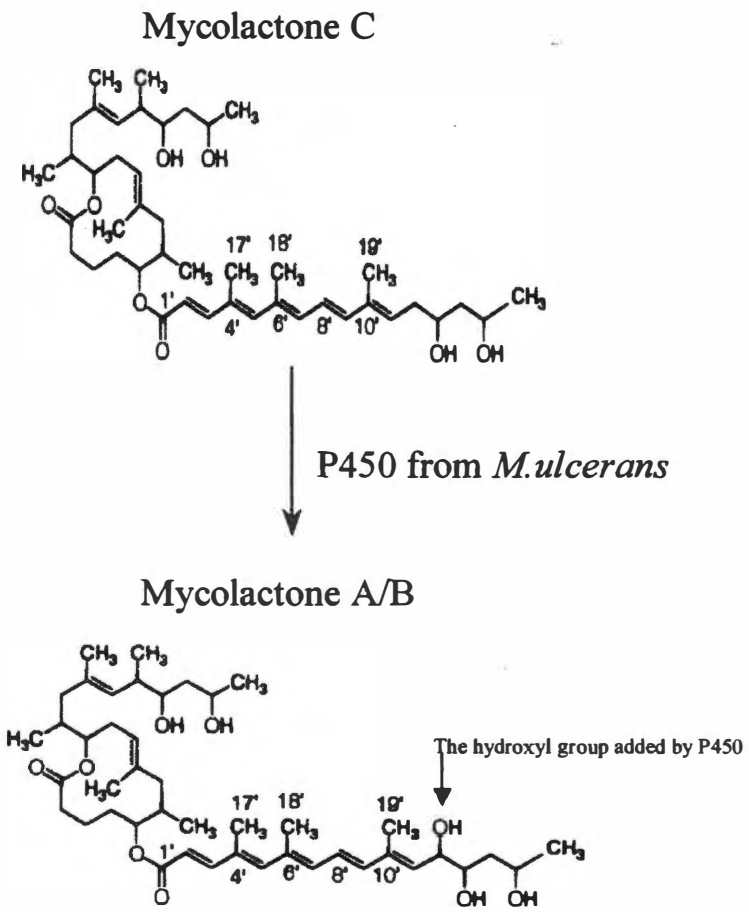


Figure 4. The proposed possible reactions of P450 from *M. ulcerans* of converting mycolactone C into mycolactone A/B. The hydroxyl group added by P450 was indicated by an arrow

strain, mycolactone profiles appear to be highly conserved within a geographic area.

1.2.2 Genetics of Biosynthesis of Mycolactone

In 2004, it was found that the 174-kb virulence plasmid pMUM001 in *Mycobacterium ulcerans* epidemic strain Agy99 contains three very large and homologous genes that encode giant polyketide synthetase (PKS), responsible for the synthesis of the lipid toxin mycolactone (Figure 5). This is the first reported example of plasmid-mediated virulence in a mycobacterium [22, 23].

The 12-membered core of mycolactone is produced by two giant, modular PKSs, mlsA1 (1.8 MDa) and mlsA2 (0.26 MDa), whereas its side chain is synthesized by mslB (1.2 MDa), a third modular PKS highly related to mslA1. There are also some genes coding for potential polyketide-modifying enzymes, including a P450 monooxygenase (*mup053*), probably responsible for hydroxylation at carbon 12 of the side chain, and an enzyme resembling FabH-like type III ketosynthases (KS) (*mup045*) [22, 23].

1.2.3 Plasmid Variation Correlates with the Variety of Mycolactone.

As early as 2003, it was found the plasmid variation with different clinical isolates with geographically diverse origins [19]. In 2005, it was also reported the plasmid variation

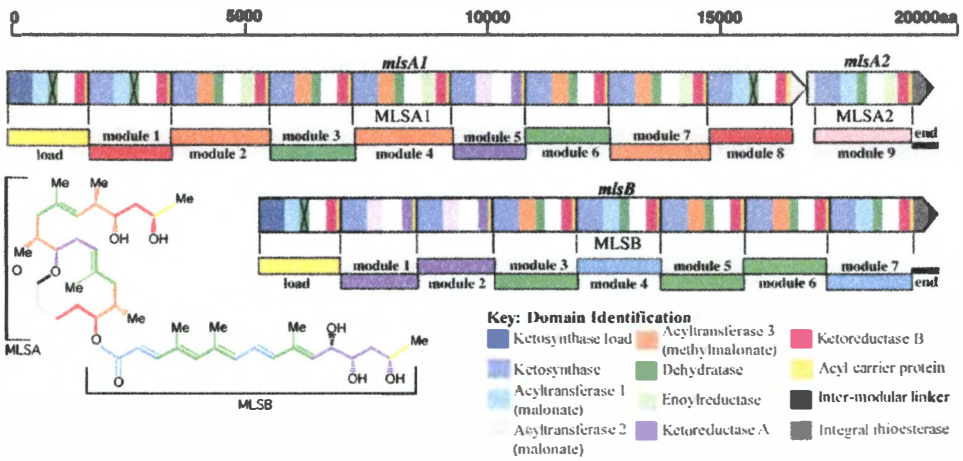


Figure 5. Domain and module organization of the mycolactone PKS genes. (Adapted from Stinear TP, et al,2004) Within each of the three genes (*mlsA1*, *mlsA2*, and *mlsB*), different domains are represented by a colored block.

For *M. ulcerans* strain Chant and ITM-941331, they lack a portion of a region that includes the gene *MUP053* (encoding a P450). The absence of *MUP053* in the Australian *M. ulcerans* strain Chant correlates well with the presence of mycolactone C and the absence of mycolactone A/B (Figure 6) [24]. This suggests that *MUP053* encodes a P450 whose substrate is the product of mlsB.

1.3 Cytochrome P450 Monooxygenases

Cytochrome P450 proteins, named for the absorption band at 450 nm of their carbon-monoxide-bound form, are one of the largest superfamilies of enzyme proteins. They are produced by a wide variety of life forms, including animals, plants, fungi and bacteria. Amino-acid sequences of the enzyme are extremely diverse. The identity between different P450s can be as low as 16% in some cases. Because of the differences in the peptide backbone, the enzymes have different substrate specificities. But all P450s contain an obligate hemecofactor. [25] This cofactor is linked to the protein peptide backbone via a thiolate ligand provided by a cysteine residue. The hemecofactor allows the enzyme to activate molecular oxygen to enable these enzymes to carry out diverse oxidative transformations. Their most conserved structural features are related to heme-binding.

M. ulcerans strain (country of origin)	pMUM001 marker[a]							
	repA	*parA*	MUP011 (STPK)	*mls* (load)	*mlsAT* (*II*)	MUP038 (TEII)	MUP045 (KSIII)	MUP053 (P450)
Agy99 (Ghana)	+	+	+	+	+	+	+	+
Kob (Ivory Coast)	+	+	+	+	+	-	+	+
1615 (Malaysia)	+	+	+	+	+	+	+	+
Chant (Australia)	+	+	+	+	+	+	+	-
IP105425 (Australia)	+	+	+	+	-	-	+	-
ITM-5114 (Mexico)	+	+	-	+	-	-	+	+
ITM-941331 (Papua New Guinea)	+	+	+	+	+	+	+	-
ITM-941328 (Malaysia)	+	+	+	+	+	+	+	-
ITM-98912 (People's Republic of China)	+	+	-	+	+	+	+	+
01G897 (French Guiana)	+	+	+	+	+	+	+	-

Figure 6. PCR analysis of 10 different *M. ulcerans* strains for the presence of eight plasmid-associated genes. (Adapted from Stinear TP,et al 2005) abbreviation as STPK, serine/threonine protein kinase; TEII, type II thioesterase; KSIII, type III ketosynthase. A panel of nine *M. ulcerans* clinical isolates with geographically diverse origins was collected and screened by PCR for the presence of eight *M. ulcerans* plasmid markers by their corresponding primers.
"+" represents positive result for PCR
"-" represents negative result for PCR

The primary physiological role of the P450 family is that of a monooxygenase, catalyzing the insertion of one of the atoms of molecular oxygen into a substrate, whereas the second oxygen atom is reduced to water [25]. The catalytic reaction can be summarized as,

$$RH + O_2 + 2H^+ + 2e^- \longrightarrow ROH + H_2O,$$

where **RH** can be one of a large range of possible substrates. The specificity of a given P450 is determined by the contact residues that define the active site of the enzyme. These can vary widely between different P450s, however the principal component of the active site of all P450s is a haem moiety. The iron ion of the haem moiety is the site of the catalytic reaction, and is also responsible for the strong 450nm absorption peak in combination with CO [26].

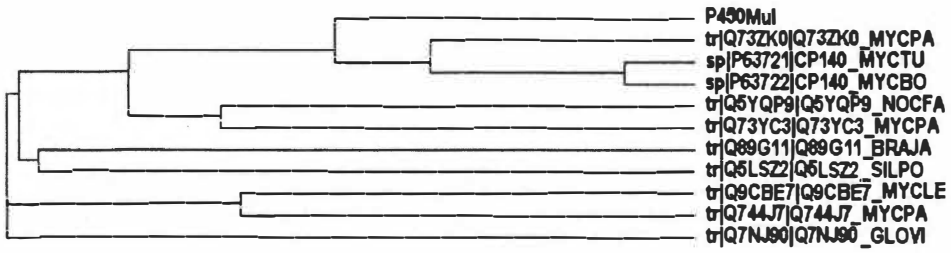
1.3.1 Gene Organization and Evolutionary History

P450 super family genes (called CYP) are classified according to the recommendation of a nomenclature committee [27], on the basis of amino acid identity, phylogenetic criteria and gene organization. The enzymes are divided into families based on amino acid sequence similarity, and each family can be further separated into subfamilies, which are designated by capital letters following the family designation (e.g., CYP3A). Individual enzymes are subsequently indicated by Arabic numerals (e.g., CYP3A4). A P450 enzyme belongs to a specific family, when the amino acid sequence reveals more than

40% sequence similarity to this family. Enzymes sharing more than 55% homology form a subfamily to this family (Figure 7) [27].

1.3.2 Structural Features

Based on the data from a number of P450 crystal structures, it is safe to state that the overall P450 fold of is quite conservative, although sequence identity among P450 proteins is often extremely low. The P450 fold is very unique since no non-P450 structure has yet been found to share the P450 fold [28]. The general structure is globular, almost triangular, with the C-terminal half being helix rich and the N-terminal half being more beta sheet rich. The C-terminal half is more conserved. The conserved core consists of a four-helix (D, E, I and L) bundle, helices J and K, two sets of ß sheets, and a coil called the 'meander' [29](figure 8). These regions includes: 1> the heme binding ligand containing the most characteristic P450 consensus sequence which is usually represented as Phe-X-X-Gly-X-Arg-X-Cys-X-Gly with the absolutely conserved cysteine that serves as fifth ligand to the heme iron. This heme-binding region is about 50 amino acids from the C-terminal of the protein. The helix rich half of the protein starts with the I-helix. 2> the consensus sequence considered as P450 signature Ala/Gly-Gly-X-Asp/Glu-Thr-Thr/Ser, which corresponds to the proton transfer groove on the distal side of the heme where the Thr residue is part of the oxygen binding site. 3> The K helix has an invariant Glu-X-X-Arg (EXXR) sequence which tolerates no



tr|Q73ZK0|Q73ZK0_MYCPA: Hypothetical protein *Mycobacterium paratuberculosis*
sp|P63721|CP140_MYCTU: Putative cytochrome P450 *Mycobacterium tuberculosis*
sp|P63722|CP140_MYCBO: Putative cytochrome P450 *Mycobacterium bovis*
tr|Q5YQP9|Q5YQP9_NOCFA: Cytochrome P450 monooxygenase *Nocardia farcinica*
tr|Q73YC3|Q73YC3_MYCPA: Hypothetical protein *Mycobacterium paratuberculosis*
tr|Q89G11|Q89G11_BRAJA: Hypothetical protein *Bradyrhizobium japonicum*
tr|Q9CBE7|Q9CBE7_MYCLE: Putative cytochrome P450 *Mycobacterium leprae*
tr|Q7NJ90|Q7NJ90_GLOVI: Putative cytochrome P450 protein *Gloeobacter violaceus*
tr|Q744J7|Q744J7_MYCPA: Hypothetical protein - *Mycobacterium paratuberculosis*
tr|Q5LSZ2|Q5LSZ2_SILPO Cytochrome P450 protein - *Silicibacter pomeroyi*

Figure 7. The phylogenetic tree of the putative cytochrome P450 *M. ulceran* with its top 10 homologous proteins. The homologous proteins of the putative cytochrome P450 *M. ulceran* were identified with the BLAST program from the SWISS-PROT database. Their phylogenetic tree was then constructed with the ClustalW program.

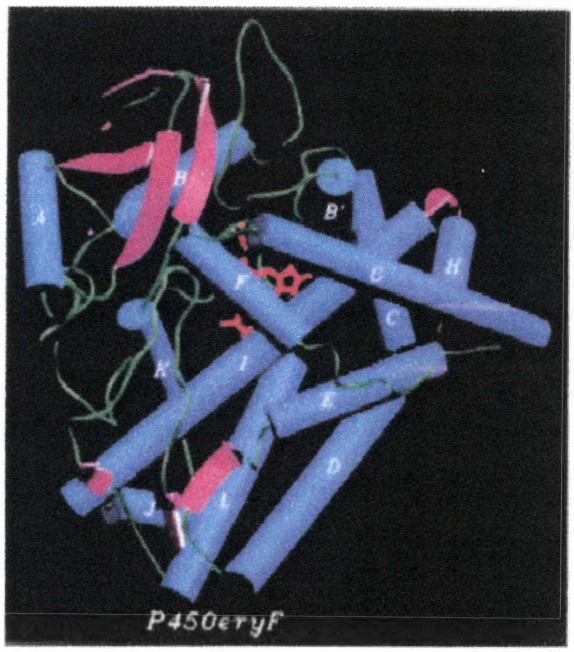


Figure 8. A ribbon representation of P450eryF crystal structure
(Adapted from Cupp-Vickery JR et al 1995) α helices are represented by blue cylinders.β sheet are represented by pink arrows. The haem prosthetic group is embedded between the I and L helices.The floor of the substrate-binding pocket is formed by the haem prosthetic group and the top by the F helix. The sides of the pocket are formed by the I helix, the β -structure consisting of residues 290-296,and the β` helix.

substitutions and is probably needed to stabilize the core structure. The E, the R and the C at the heme-binding site are the only completely conserved amino acids in P450s [30].

1.3.3 Function in Biological Systems

The ubiquity of P450 is highlighted by recent genome sequencing projects. Although *Escherichia coli* and *Salmonella typhimurium* contain no P450s, *Mycobacterium tuberculosis* has 20, *Saccharomyces cerevisiae* has 3, the fruit fly *Drosophila melanogaster* has 86, while humans have 57 P450s and most amazingly the plant *Arabidopsis thaliana* contains 249 different P450s [30]. These enzymes played very important and varied biological roles. In eukaryotes, the P450 are usually bound to the endoplasmic reticulum or inner mitochondrial membranes and are involved in the biosynthesis and catabolism of signaling molecules, steroid hormones, retinoic acid and in all the pathways of secondary metabolism. [31]. In plants, P450s are involved in the biosynthesis and catabolism of all types of hormones, in the oxygenation of fatty acids, and in all the pathways of secondary metabolism.[31]. In prokaryotes, they enable prokaryotes to catabolyze compounds used as carbon sources or to detoxify xenobiotics. Other functions described for prokaryotic P450s include fatty acid metabolism and biosynthesis of antibiotics.

1.4 Objective

The product of *M. ulcerans* putative P450 in the plasmid encoding the mycolactone is predicted to hydroxylate the mycolactone side chain of mycolactone C at Carbon-12' to produce mycolactone A/B. Strains lacking the hydroxyl group at C-12' produce mycolactone C. This is a characteristic of the Australian strains. The absence of putative P450 gene in *Mycobacterium ulcerans* strain Chant correlates well with the presence of mycolactone C and the absence of mycolactone A/B.

The goal of my research was to clone and purify *M. ulcerans* P450 in *E. coli* to study the structural and functional aspects of P450 enzymes in *M. ulcerans*, A P450 in *M. ulcerans* has not been expressed or analyzed. A high level of heterologous expression of *M. ulcerans* in *E. coli* is prerequisite to prove that the substrate of the P450 is mycolactone C and to study the enzyme activity

Chapter 2

Materials and Methods

2.1 Subcloning of P450A Gene into PET-30a Vector

A library of large fragment of *M.ulcerans* 1615 was cloned into a bacterial artificial chromosome (BAC) library. The BAC 7017 which contains the plasmid P450 gene was used as the templates for the PCR amplification. Primers corresponding to regions flanking the p450-encoding region were used to amplify the insert via standard PCR protocol that engineered *NdeI* and *HindIII* restriction sites at the 5'and 3'ends of the amplified product. The sense primer used was 5'-GAAGGAGATATACATATGTCCACAATCTACGGCTA-3' and the anti-sense primer used was 5'CGGCCGCCATATGTCATCGGATAGCCATCGATCG-3. The P450 gene amplified from the BAC 7017 were double digested with *NdeI* and *HindIII* restriction enzymes (New England Biolabs, Beverly, MA) at 37℃ for 3 hours. The digested product and PET30a plasmid were gel purified using QIAquick PCR gel purification kit (Qiagen, Valencia, CA). Ligation was performed using approximate 3:1 insert: vector molar ratio and ligase used was from Promega (Madison, WI). The ligation reaction was done at room temperature overnight. Ligation product was transformed into competent TOP 10 *E. coli* cells following normal transformation protocol.

2.2 Direct Colony PCR Screening of Ligation Product

The presence of P450 gene in PET-30a vector was first confirmed by colony screening. 25μl reactions were set up containing Takara ExTaq polymerase (Madison,WI) and the PCR reaction was performed in an Eppendorf thermal cycler (Mastercycler Gradient, Epperdorf, Hamburg, Germany). The colony PCR primers used were T7 universal primer: 5'-T AAT ACG ACT CAC TAT A_3' and reverse primer: 5'-GAG GTT GGT AAT AAG GTC ATG GGT-3'. At the same time, a patch plate was created. 6% agarose gel was run for the PCR products and positive colonies were grown up from the patch plate according to the fragment size on the gel. Plasmid was purified and sequencing confirmed as mentioned in section 2.5.

2.3 Transformation of P450-PET-30a and Plating

Eleven DE3 *E. coli* cell lines were test for their usefulness as the expression cell lines. They were: BL21, BLR, HMS 174, Turner™, Novablue, Rosseta™, Rosseta-gami, Rosseta-gami B, Rosseta Blue, Origami ™, Origami B (Novagen, CA). 2μL of construct P450-PET-30a was added into 20 μL competent cell line, the tubes were kept on ice for 5 minutes. The tubes were placed in 42℃ water-bath for 30 seconds, then immediately on ice for 2 minutes. 250 μL SOC medium was added to each tube, incubated at 37℃ for one hour and plated on the Kanamycin resistance plate (30 μg/ml) overnight.

2.3.1 Induction and Growth

Four colonies were taken out from each specific cell line's plate into 10 ml Kanamycin resistance LB medium tube. Each tube was incubated at 37, 200 Rpm until OD was between 0.4~0.6. 1 ml medium was taken out from each tube, IPTG was added to the concentration of 1 μM, incubated at 37℃, 200 Rpm for 2 hours. 100 μL sample was taken from each the original uninduced and induced medium, spinned at maximum speed for 2 minutes. 80 μL of supernatant were removed, the left 20 μL were saved for SDS-PAGE.

2.3.2 SDS-PAGE and Coomassie Brilliant Blue Detaining

Proteins were separated by SDS-PAGE (5% acrylamide-stacking and 18% acrylamide-separating) and stained in Coomassie staining solution (10% acetic acid, 50% methanol, 0.25% Coomassie Brilliant Blue R 250) for half an hour and distained in distaining buffer (10% acetic acid, 50% methanol) for 30 minutes. Pictures were taken using UVP BioImaging System (Upland, CA)

2.3.3 Purification of P450 from Inclusion Bodies

P450-PET-30a was transformed into BLR [DE3] following normal transformation protocol. Several colonies were inoculated into 10 ml LB (Kan 30 μg/ml) respectively,

and were grown at 37℃ until OD_{600} reached 0.6. Then 1 ml culture was transferred to a new tube and 1 μl 1M IPTG (isopropyl-b-D-thiogalactoside) was added for induction. An 18% polyacrylamide gel was run to determine which colony had the best expression and the remaining uninduced preculture (9 mls) was used as inoculum for 1 L LB medium. IPTG was added when OD_{600} reached 0.6 and the cells were grown for 3 more hours. The cells were then spun down at 5,000 g for 10 minutes and the supernatant was decanted. The pellet was resuspended in 30 ml 1x Buffer A (50 mM Tris-HCL, 5 mM $MgCl_2$, pH 7.6) and spun down at 40,000 x g for 10 minutes. Pellets were frozen at –80℃. Then the next day pellets were resuspended in 30 ml 1x Buffer A + 0.1% Triton X-100 and the sonicated samples were then centrifuged at 40,000 x g for 10 minutes. The supernatant was decanted and 20 μl was saved for gel analysis. Then the pellet was resuspended with 30 ml 1x Buffer A +0.1% Triton X-100 and treated with sonication and centrifugation as before. This step was repeated at least 3 more times. The pellet was then rinsed with 30 ml 1x Buffer A 3 times to remove Triton X-100. The final inclusion body pellet was incubated with 8 M urea + 50 mM DTT at room temperature overnight and then was centrifuged at 40,000 x g for 30 minutes. Supernatant was collected and protein was frozen at –80.

2.4 Generation of P450A-$(his)_6$-PET30a Construct

Multi-quick site-directed mutagenesis of P450- $(His)_6$-PET30a was performed using Stratagene QuickChange multi site-directed mutagenesis kit (Stratagene, LA, Jolla, CA). Single-strand primers were designed in order to change stop condon into an amino acid code, so the 6 histidine tag of the downstream can be expressed. The forward primer P450E: 5'-CGATGGCTATCCGAGGAAAGCTTGCGGCCG-3', the reverse primer P450F: 5'-CGGCCGCAAGCTTTCCTCGGATAGCCATCG-3'. The PCR reaction was carried out as below:

Step 1 Denaturation:	94℃	4min;
Step 2 Denaturation:	94℃	1min;
Step 3 Annealing:	55℃	1min;
Step 4 Extension:	65℃	14min
Step 5 Extension:	65℃	7 min
Step 6 Storage:	4℃	

The Taq polymerase *pfu* used was from Promega (Madison, WI). PCR product was transformed to XL-Blue competent cells after Dpn I treatment. Plasmids were extracted using Wizard vacuum mini-prep kit from Promega (Madison, WI) and confirmed by sequencing.

2.4.1 Expression and Purification of P450-(his)$_6$-PET30a Protein

The recombinant plasmid P450-(His)$_6$-PET30a was used to transform BLR(λ DE3) to kanamycin resistance. Overnight cultures (approximately 5 ml) grown in LB medium containing Kanamycin (50 μg/ml) were used to inoculate 1 liter of the same medium. This culture was grown at 15℃ until the optical density of the culture at 600 nm (OD_{420}) reached 0.5. At this time, isopropyl- β -D-thiogalactopyranoside (IPTG) was added to a final concentration of 1 mM, and the culture was grown for another 18 h. All subsequent steps were carried out at 4℃. Cells were collected by centrifugation and resuspended in lysis buffer (50mM Sodium Phosphates, 300mM NaCl, PH 7.0, 0.5% Triton X100, 3mM 2-mercaptoethanol BME). Cells were broken by passage through a French pressure cell at 20,000 psi. The cell extract was centrifuged at 100,000 X g for 1 h, and the resulting supernatant was purified with Co2+-Talon Metal Affinity Resin BD Biosciences from Clontech (Palo Alto, CA) according to the manufacturer's instructions. The Talon metal affinity resins were washed with 10 vol. of lysis buffer and then with 10 vol. of 50mM sodium phosphates, 500mM NaCl, PH 7.0, 0.5% Triton X100, 3mM BME and finally with 10 vol. of 50mM Sodium Phosphates, 500mM NaCl, PH 7.0, 0.5% Triton X100, 3mM BME, 15 mM imidazole. Elution was 50mM sodium phosphates, 500mM NaCl, PH 7.0, 0.5% Triton X100, 3mM BME, 250 mM imidazole. Successful expression and purification were verified by SDS/PAGE. Purified recombinant protein

was dialyzed against 100mM KCL, PH 7.4 , 15% glycerol at 4℃ by using Slide-A-Lyzer dialysis cassettes. (10,000 molecular weight cutoff, Pierce).

2.5 Construction of P450B-(his)$_6$ into pCWori Vector

The probable ORF of P450-Mul. sequence was aligned with other p450 gene sequence. The below sequence was used to clone into pCWori vector.

BAC1707 containing the P450 gene was used as the template for the PCR amplification. Primers corresponding to regions of conservative sequences were used to amplify the insert via standard PCR protocol that engineered *NdeI* in the forward site and 6 histidine and stop condon and Hind III restriction site in the backward site. Forward primer is CYP-A: 5'-GAAGTAGGAGGTCATATGGTGAGGCAGAGATTGAACTGG-3' Backward primer: 5'-GGTATTAAAGCTTCAATGGTGGTGATGGTGATGTCGGATA GCCATCGATCGTGC-3'. The P450 gene amplified from the BAC1707 was double digested with *Nde I* and *HindIII* restriction enzymes at 37 for 3 hours. The digested product and pCWori plasmid (nicely provided by Dr. Michael Waterman in Vanderbilt University) was gel purified using QIAquick PCR gel purification kit (Qiagen, Valencia, CA). Ligation was performed using approximate 3:1 insert: vector molar ratio and the ligase used was from Promega (Madison, WI). The ligation reaction was done at room temperature overnight. Ligation product was transformed into TOP 10 competent *E.coli*

cells following standard transformation protocol.

2.5.1 Direct Colony PCR Screening of Ligation Product

The presence of P450 gene in pCWori vector was first confirmed by colony screening. The PCR primer is the same. Plasmids were purified.

2.5.2 Sequence of the P450B-$(his)_6$-pCWori

The sequence forward primer is P450-Middle1: 5'-GTTGAGCCGCCGAATCAC-3', backward primer is P450-Middle2: 5'-GTGATTCGGCGGCTCAAC-3'

2.6 Expression and Purification of P450B-$(his)_6$-pCWori Vector

The method was the same as expression and purification of P450B-$(His)_6$-PET30a Protein except it is the P450B- $(his)_6$- -PCWori Vector was transformed into BLR(λDE3) cell lines.

Chapter 3

Results

3.1 From Gene Isolation to Protein Purification

3.1.1 Clonings of P450A in PET30 vector

A *M. ulcerans* BAC library constructed from the whole genome DNA of *M. ulcerans* 1615 was used as the sequence source for the P450 gene. Previous work showed that this gene was part of the mycolactone gene contained in the *M. ulcerans* plasmid. BAC1707 contained the whole plasmid element.

P450A sequence was identified by NCBI ORF software by scanning the full sequence of BAC1707.The full length of the putative P450A sequence was amplified from *M. ulcerans* BAC 1707 genomic DNA by PCR, using oligonucleotides designed to introduce *NdeI* and *HindIII* restriction sites. The PCR product amplified from the BAC 7017 was double digested with *NdeI* and *HindIII* restriction enzymes. The DNA of pET30a vector was double digested with the same enzymes. Ligation was performed using approximately 3:1 insert: vector molar ratio. The ligation reaction was done at room temperature overnight. After transformation, colonies were screened by PCR and restriction digestion. The insert of one positive clone was sequenced after plasmid DNA isolation in order to exclude the presence of undesired mutations. The sequence was

100% identical to the original sequence of BAC1707. Several screening tests were carried out to search for the best growth parameters for producing the most soluble protein of P450A. BLR cell line as phenotype as *E. coli* B F- dcm+ Hte ompT hsdS(rB-mB-) gal l (DE3) endA Tetr produced the most soluble P450A protein under the same condition compared with other ten commercial cell lines tested including BL21, HMS174, Tuner™, NovaBlue, Rosetta™2, Rosetta-gami, Rosetta-gami B, RosettaBlue, Origami™, and Origami B. (Figure 9) Thus BLR cell line was chosen as the expression cell line. Three different induction temperature were tested: 15℃, 30℃ and 37℃. The results proved that the most soluble protein of P450A was produced at 15℃ comparing at 30℃ and 37℃. The different induction times (3, 6, 12, 18, 24, 30 and 36 hours) and culture medium (LB medium, SV medium) were also tested, but different induction times and culture medium didn't make much difference to the quantity of soluble protein production of P450A.

The plasmid encoding P450A, designated P450A-PET30a, was specifically mutated by a multi-mutagenesis kit (Stratagene, LA, Jolla, CA), producing plasmid P450A-(his)6 -PET30a in order to facilitate purification. The expression product of plasmid P450A-(his)6 -PET30a contains the P450 with a six histidine tag on the C terminal for

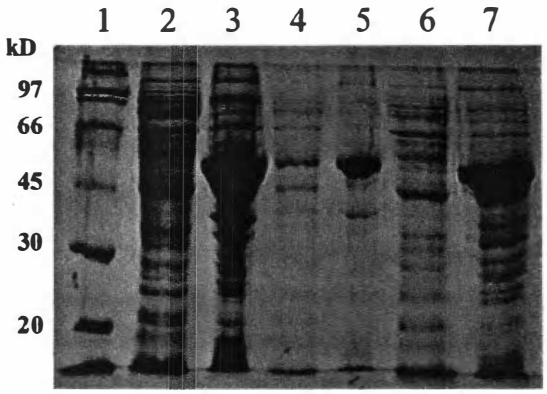


Figure 9. Soluble and insoluble protein comparison in cell line BL21, BLR, HMS174. 5 ml culture of different cell lines were induced at 15 ℃ for 18 hours in LB medium. The soluble and insoluble fraction of protein were separated by centrifugation and run on 12% acrylamide gel. Lane1: 1 : BMW marker; Lane 2 : Soluble fraction after induction in cell line BL21;lane 3 : Insoluble fraction after induction in cell line BL21;lane 4 : Soluble fraction after induction in cell line BLR; lane:5 : Insoluble fraction after induction in cell line BLR; Lane 6 : Soluble fraction after induction in cell line HMS174; lane 7:: Insoluble fraction after induction in cell line HMS174

affinity purification. The principle of purification is the 6xHis system in which six histidine residues are tagged to the recombinant protein. The histidine tag and protein can be purified using a nickel-chelating resin. The plasmid P450A-(his)6 -PET30a was transformed into BLR cell line and scaled up to 2 liters. After the OD_{600} was 0.5, IPTG was added until the final concentration was 1mM. The culture was induced for 18 hours at 15℃ and was purified by the Talon metal affinity kit. The result was analyzed by SDS-PAGE. A weak band corresponding to the target protein was seen, however the results showed that a lot of nonspecific binding occurred, (Figure 10). Results from SDS-PAGE gel analysis of purified protein showed the presence of many proteins. In summary, this method did not yield highly purified protein.

3.1.2 Cloning, Expression and Purification of P450B in pCWori vector

Multiple alignment analysis of the P450A sequence which was used for cloning into PET30a with other confirmed P450s sequence indicated that the first 135 base pairs in the P450A were not aligned with any other P450 sequences.[32] (Figure 11) This suggested that the extra 135 base pairs could be responsible for the failure to obtain soluble protein. The part of P450A sequence without P450A's first 135 base pair encoding amino acids that was the cause of insolubility, designated P450B, was aligned with other P450s using a "gtg" as the start codon. The P450B sequence was cloned into vector pCWori, a vector used previously for expressing bacterial P450's.[33]. This vector contained a TAC

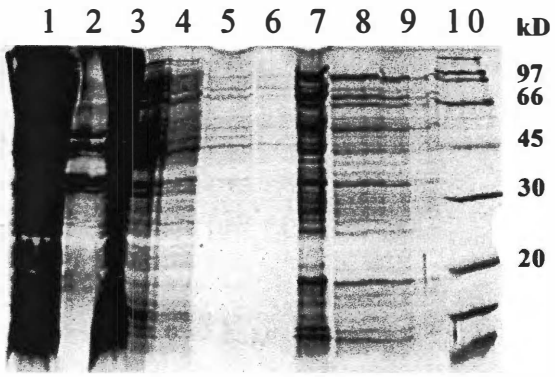
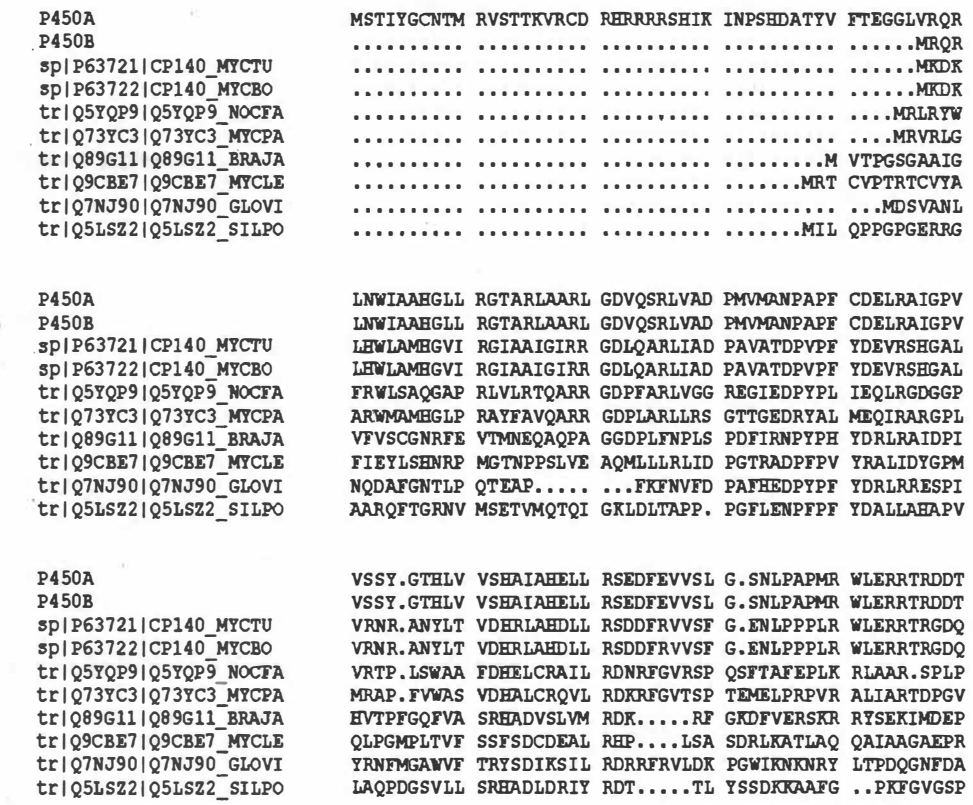


Figure 10. Overexpression and purification of P450A. P450A was cloned into PET30a vector and followed by site-specific mutation to yield a 6 histine tag on the C terminal. The overexpression of the fusion protein was carried out in *E. coli* at 15. Expression and purity of the P450A-6His analyzed by 12% SDS–PAGE. Lane 1: total cell lysate from uninduced *E. coli*. Lane 2: the supernatant protein fraction from the induced cell lysate Lane 3: the insoluble protein (pellet) fraction from the induced cell lysate separated by high speed centrifugation. Lane 4,5,6: three times washing solution; Lane 7,8,9: three times elution; Lane 10: protein standard marker



sp|P63721|CP140_MYCTU Putative cytochrome P450 *Mycobacterium tuberculosis*
sp|P63722|CP140_MYCBO Putative cytochrome P450 *Mycobacterium bovis*
tr|Q5YQP9|Q5YQP9_NOCFA Cytochrome P450 Nocardia farcinica
tr|Q73YC3|Q73YC3_MYCPA Hypothetical protein *Mycobacterium paratuberculosis*
tr|Q89G11|Q89G11_BRAJA Bll6537 protein *Bradyrhizobium japonicum*
tr|Q9CBE7|Q9CBE7_MYCLE Putative cytochrome p450 *Mycobacterium leprae*
tr|Q7NJ90|Q7NJ90_GLOVI Cytochrome P-450 like protein *Gloeobacter violaceus*
tr|Q5LSZ2|Q5LSZ2_SILPO Cytochrome P450 family protein *Silicibacter pomeroyi*

Figure 11. The sequence alignment of P450A, P450B and their homologous proteins. The homologous proteins include putative cytochrome P450 *Mycobacterium tuberculosis*; putative cytochrome P450 *Mycobacterium bovis*; cytochrome P450 *Nocardia farcinica*; hypothetical protein *Mycobacterium paratuberculosis*; putative cytochrome p450 *Mycobacterium leprae*; cytochrome P-450 like protein *Gloeobacter violaceus*; cytochrome P450 family protein *Silicibacter pomeroyi*. These sequences were retrieved by NCBI database BLAST software and aligned with CLUSTAL-W software. Only the beginning section of the multiple sequence alignment is shown.See reference 32

promoter, which is 15 times weaker than the T7 promoter in PET30a. The full length of the putative P450B sequence was amplified from *M. ulcerans* BAC 1707 genome DNA by PCR, Primers corresponding to regions of P450B sequences were used to amplify the insert via standard a PCR protocol that engineered *NdeI* in the 5' site of the gene and 6 histidines and stop condon and *Hind III* restriction site in the 3' site of the gene. The p450 gene amplified from the BAC 7017 was double digested with *NdeI* and *HindIII* restriction enzymes. The DNA of pCWori vector was double digested with the same enzymes. (Figures 12, 13) Ligation was performed using approximately 3:1 insert: vector molar ratio and was done at room temperature overnight. After transformation and the isolation of the plasmid DNA, positive clones were identified by PCR and analytical restriction analysis using *NdeI* and *HindIII* (Figures 14). The insert of positive clones was sequenced, in order to exclude the presence of undesired mutations. One clone was selected for further experiments

The soluble cytoplasmic fraction and insoluble fraction were separated by centrifugation and analyzed by SDS-PAGE. The over-expression of p450B-$(his)_6$ was detected in soluble fraction sample as a thick protein band corresponding to a protein at 50 kDa, which was consistent with the predicted molecular mass of p450B-$(his)_6$. This band was absent from the culture before induction. The culture was scaled up to 2 liters. BD TALON metal affinity resins were used to purify the histidine-tagged protein. The

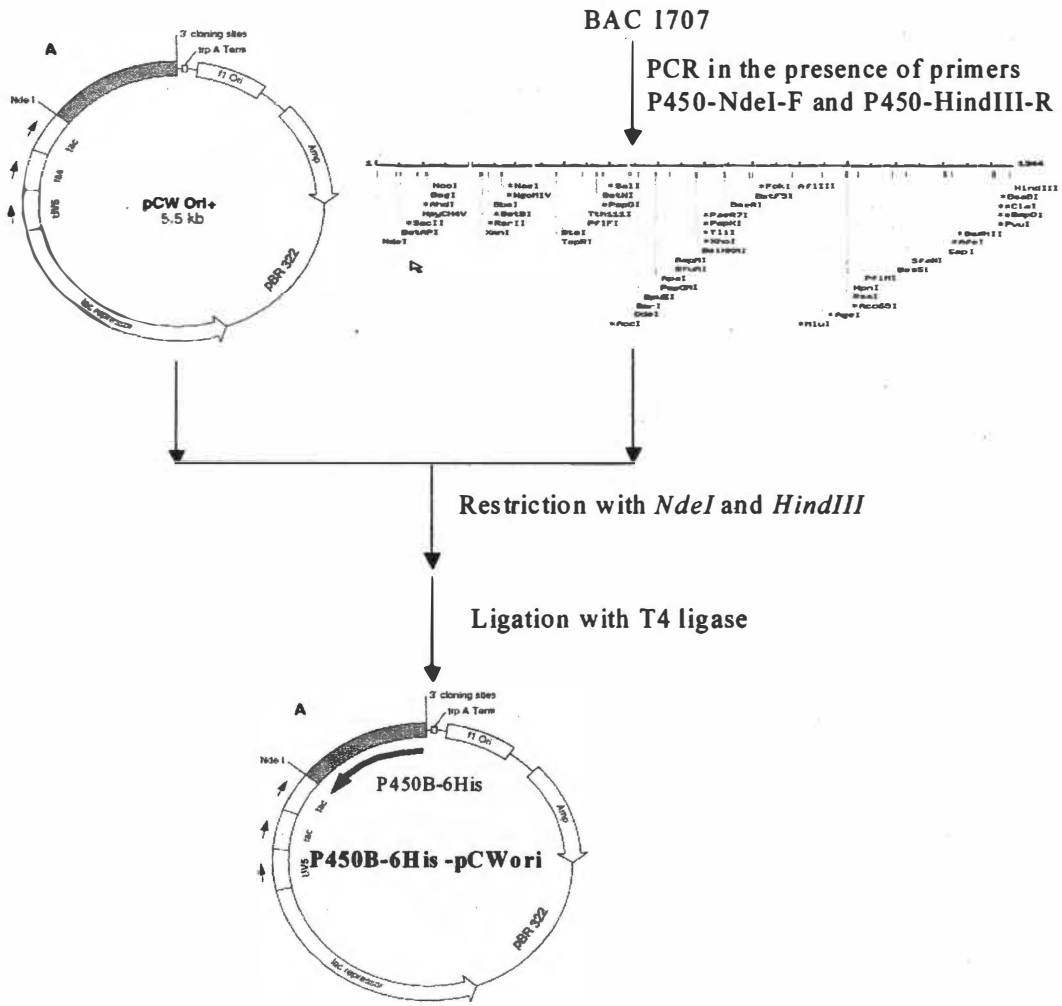


Figure 12. The scheme of the molecular method of cloning of P450B. The full length of the putative P450B sequence was amplified from *M. ulcerans* BAC 1707 genome DNA by PCR, Primers corresponding to regions of conservative sequences were used to amplify the insert via standard PCR protocol that engineered *NdeI* in the forward site and 6 histine and stop condon and *Hind III* restriction site in the backward site. The p450 gene amplified from the BAC 7017 was double digested with *NdeI* and *HindIII* restriction enzymes. The DNA of pCWori vector was double digested with the same enzymes. Ligation was performed using approximate 3:1 insert: vector molar ratio and was done at room temperature overnight.

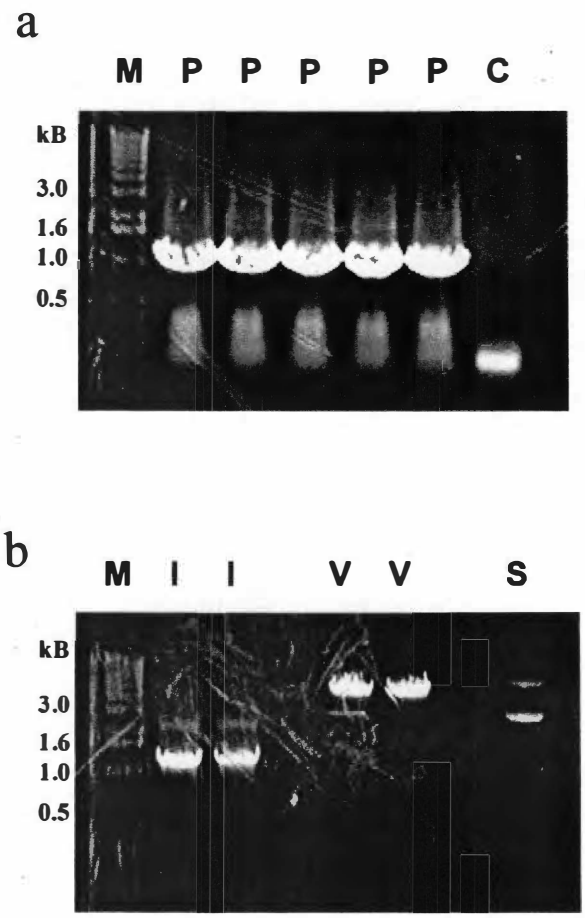


Figure 13. PCR result and enzyme restriction analysis of PCR product a> The analysis result of PCR experiment amplifying P450A gene. Lane M represents 1.0 Kb DNA marker; lane P represents the PCR product; lane C represents the negative control in PCR set up in which the water was added instead of the template DNA b> Enzyme restriction analysis of PCR product and pCWori vector. Lane M represents 1.0 Kb DNA marker; lane I represents the PCR product after double digestion of *NdeI* and *HindIII*; lane V represents the pCWori DNA after double digestion of *NdeI* and *HindIII*. Lane S represents the supercoiled pCWori DNA (undigested DNA of pCWori)

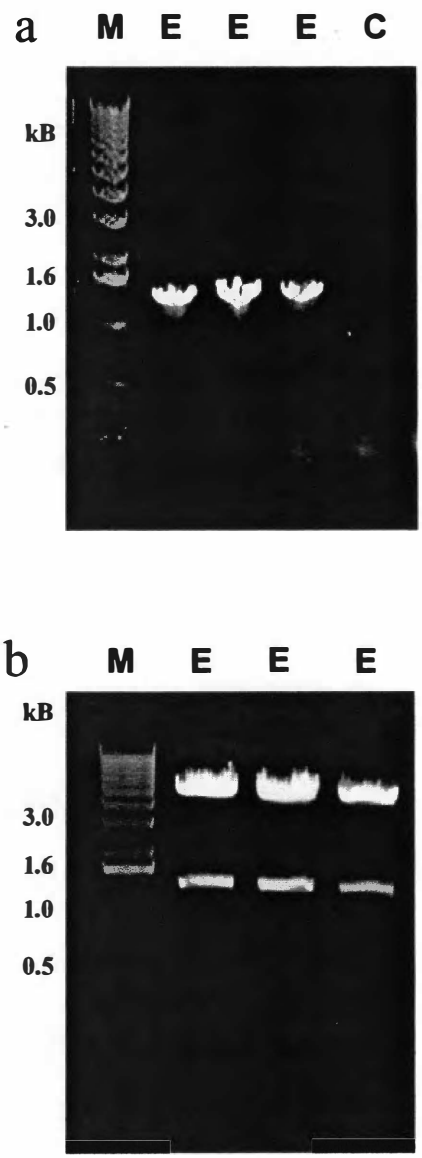


Figure 14. Screening of the positive *E. Coli* colonies. a>The PCR result of screening positive *E.coli* colonies. Lane M represented 1.0 Kb DNA marker. Lane E represented the PCR result of three *E.coli* colonies selected randomly from the plate. b> Enzyme restriction analysis of the three *E.coli* plasmid DNA. Lane M represented 1.0 Kb DNA marker; Lane E: the DNA of the plasmid after double digestion from the PCR-positive *E.coli* colony.

a reaction contained the detectable band with minor impurities (Figure 15) and reacted strongly with anti-His-tag antibodies in western blots (Figure 16).

3.2 Protein Characterization

3.2.1 Sequence Comparisons of the 50Kda Protein

The deduced amino acid sequence of P450B was very similar to those of many previously characterized P450 enzymes by protein sequence database searches. A CLUSTAL W alignment of P450B sequence with the well-characterized *Pseudomonas putida* P450cam (NCBI accession no. P00183) and *Saccaropolyspora erythraea* P450EryF is shown in Figure 17. Cytochrome P450eryF catalyzes the 6S-hydroxylation of 6-deoxyerythronolide B, the initial reaction in a multistep pathway to convert 6-deoxyerythronolide B into the antibiotic, erythromycin. Over a position alignment, the three proteins shared identity, and the extent of amino acid similarity was 12%.

A comparison with published crystallographic structures of P450cam (Protein Data Bank ID 5CP4) and P450 EryF revealed the conservation in P450B of individual residues and regions that play critical structure and functional roles for these well-characterized enzymes. The below are listed some of the most common conservations: 1. Residues R112, R299, and H355 in P450cam, which serve as H-bond donors to propionate groups of the heme prosthetic group, are conserved in P450B (R122, R320, H378, respectively).

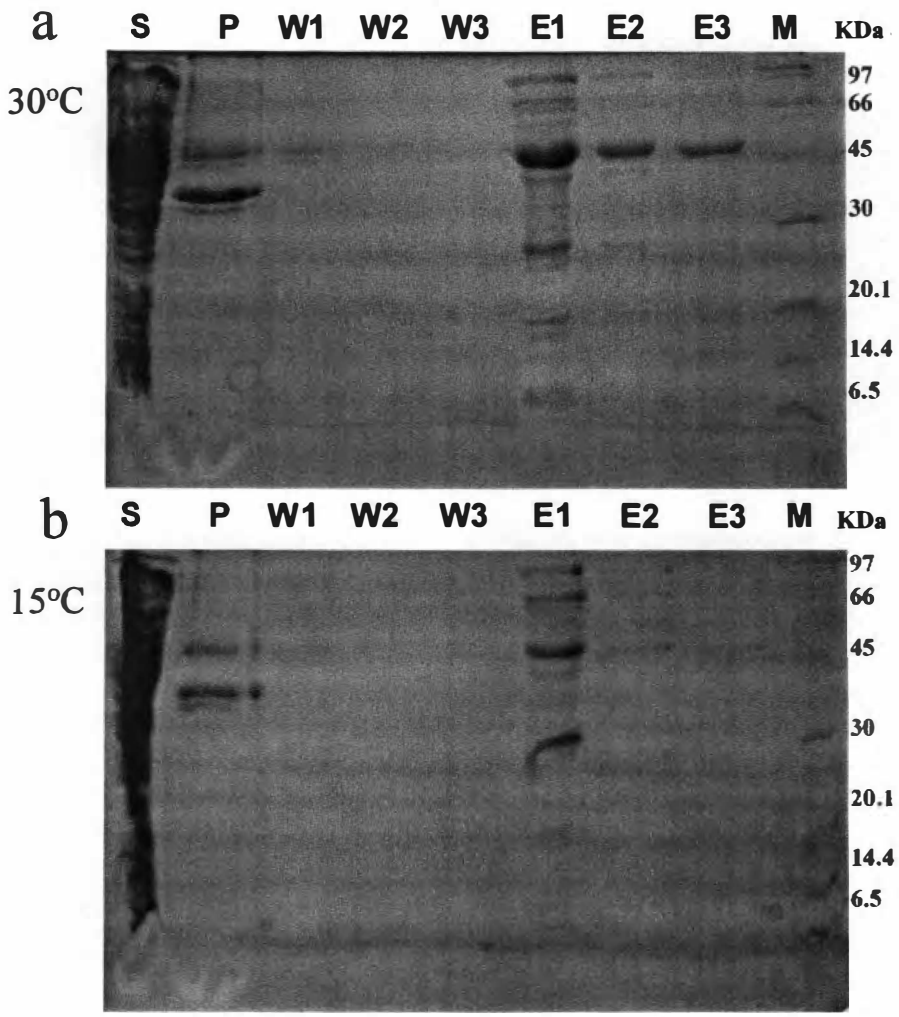


Figure 15. Overexpression and purification of P450B at induction temperatures 30°C (part a) and 15°C (part b). P450B was cloned into pCWori vector and the overexpression of the fusion protein was carried out in *E. coli* at two different temperature.Expression and purity of the P450B analyzed by 12% SDS–PAGE. Lane S: the supernatant protein fraction from the induced cell lysate separated by high speed centrifugation. Lane P: the insoluble protein (pellet) fraction from the induced cell lysate separated by high speed centrifugation. Lane W1,W2,W3: three times washing solution; Lane E1,E2,E3: three times elution; Lane M: protein standard marker.

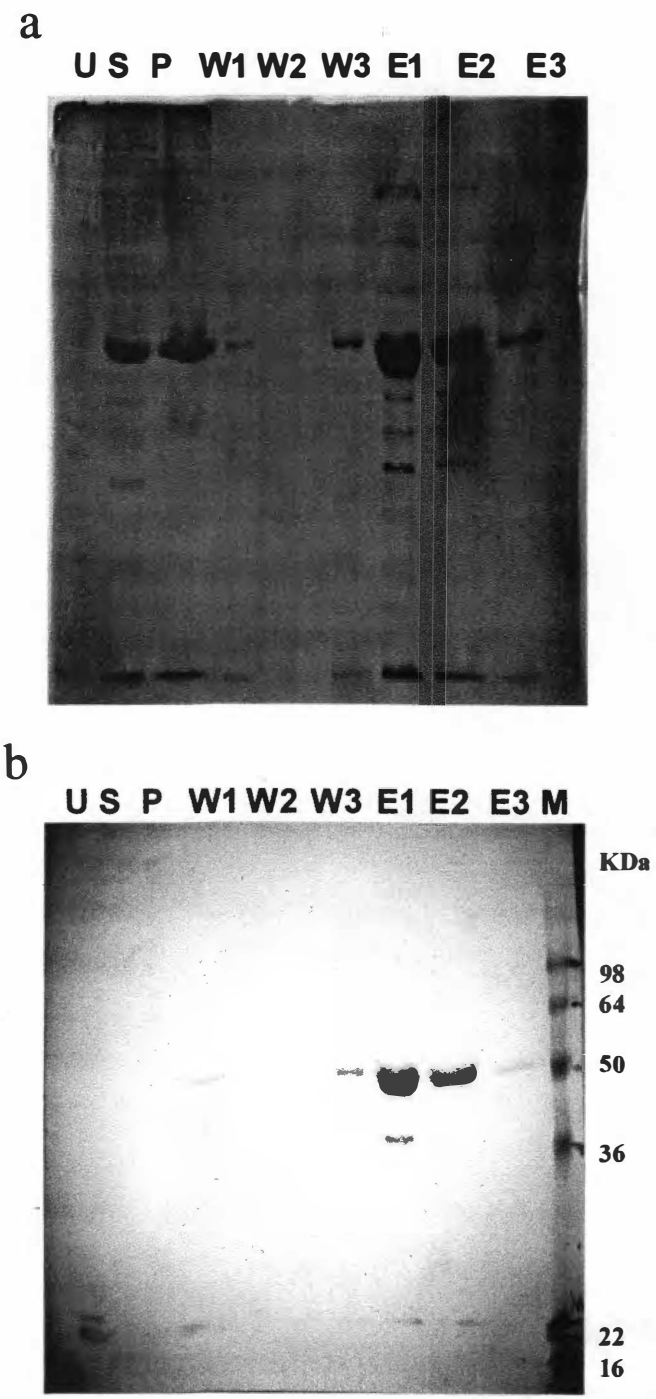


Figure 16. Purification of P450 and western blot. a>Overexpression and purification of P450B.Lane U: total cell lysate from uninduced *E. coli.* Lane S: the supernatant protein fraction from the induced cell lysate Lane P: the insoluble protein (pellet) fraction from the induced cell lysate separated by high speed centrifugation. Lane W1,W2,W3: three times washing solution ;Lane E1,E2,E3: three times elution;Lane M: protein standard marker b> Immunoblot detection of P450B-6xHis with anti-His-tag antibodies corresponding to each lane in the SDS-PAGE gel.

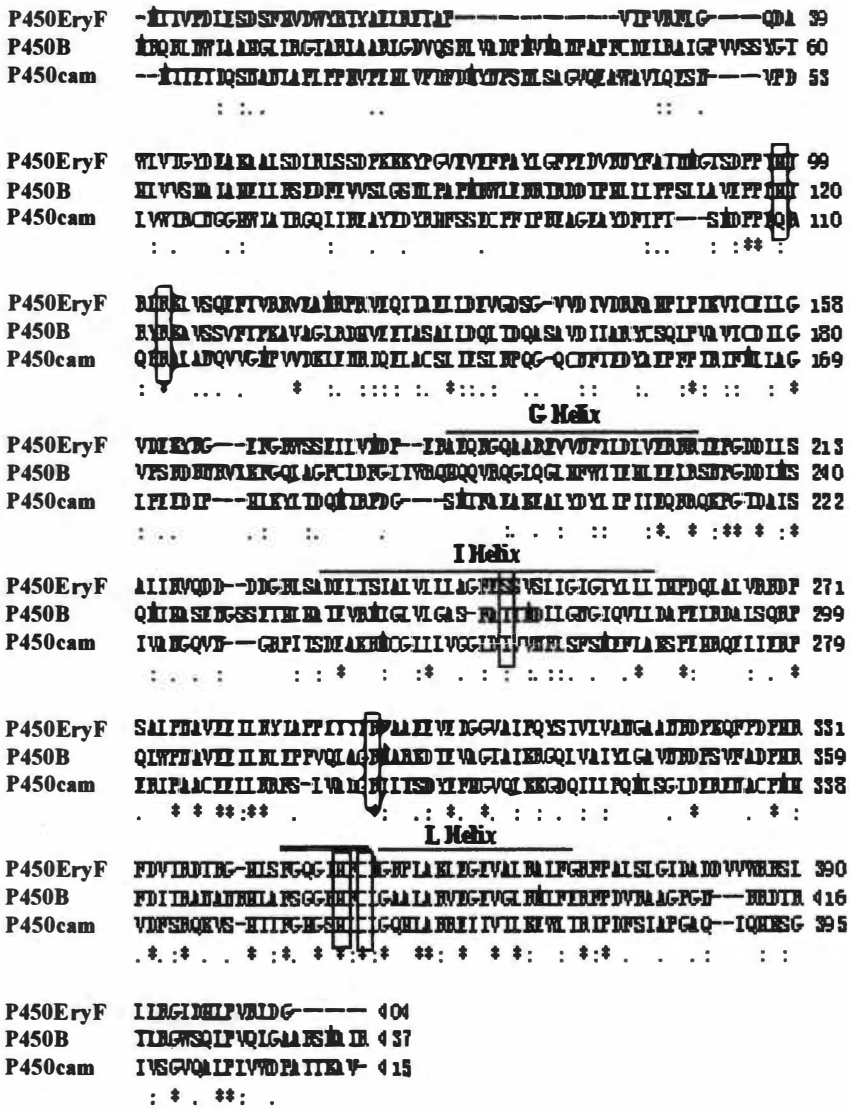


Figure 17.. Alignment of P450B sequence with the well-characterized *Pseudomonas putida* P450cam(NCBI accession no. P00183) and *Saccaropolyspora erythraea* P450EryF. Residues marked with (*) are identical, while(:)indicate strong similarity and (.) weak similarity respectively. The boxed residues are the most conserved P450 amino acids. The multiple sequence alignment was performed with the CLUSTAL W program

2. A high degree of sequence conservation was also found among the I and L helices of P450cam, and the corresponding regions of P450B. These secondary structural features as the helix I, L serve to bracket the distal and proximal faces of the heme group, respectively, in P450cam.

3. Residue T253 in helix I of P450cam, which is thought to play a role in proton translocation during the monooxygenase reaction, is also conserved in P450B (T273).

4. The region in P450B corresponding to the heme binding pocket, including the cysteinyl residue (C358 in P450cam and C380 in P450B) that forms the proximal heme iron thiolate ligand, is conserved and conforms to the consensus sequence found in all cytochrome P450 enzymes, FXXGXXXCXG.

3.2.2 Spectroscopic Characterization

P450 monooxygenase are good candidates for spectroscopic characterization. They display typical absorption spectra, which make the identification convenient. P450 shows the characteristic maximum at 450 nm in the reduced CO bound form which gets this protein the terminology of P450. The peak at 420 nm is typical for disrupted P450 enzymes, which was caused by denaturation during purification.

Spectra of the purified p450B-$(his)_6$ in storage buffer were recorded from 300 to 600 nm. The ferric heme was then reduced by adding sodium dithionite to the protein solution,

and the spectrum of the reduced enzyme was then recorded. Treatment of the sodium dithionite reduced sample with carbon monoxide was recorded too. Absorbance spectra for the ferric-oxidized, dithionite-reduced, and dithionite-reduced CO forms of p450B-$(his)_6$ were shown in figure 18. P450B exhibited a Soret absorption band at 420 nm in any form, which indicates p450B-$(his)_6$ produced in *E. coli* is a heme-containing protein with spectral properties similar to those of catalytically inactive form P450 enzymes-P420 form.

3.2.3 Mass Spectrometric Analysis of the 50 Kda Protein

The purified p450B-$(his)_6$ recombinant protein was digested with the protease trypsin to generate a mixture of peptides. The peptide mixture was then analyzed with liquid chromatogram-tandem mass spectrometry ($LC\text{-}MS^2$) (Figure 19). Briefly, the peptides were separated with reverse phase liquid chromatogram, ionized with electrospray ionization and detected with an ion trap mass spectrometer LCQ (ThermoFinnigan, San Jose, CA). During the mass spectrometric detection, the peptides were automatically selected for tandem mass spectrum measurement. The acquired tandem mass spectra were used to identify the measured peptides with the SEQUEST program [34]. The sequence database used by the SEQUEST program includes the p450B-$(his)_6$ sequence and the genome sequence of *Escherichia coli* K12. The identified peptides were then assembled to the proteins with the DTASelect program [35]. 37 peptides from the

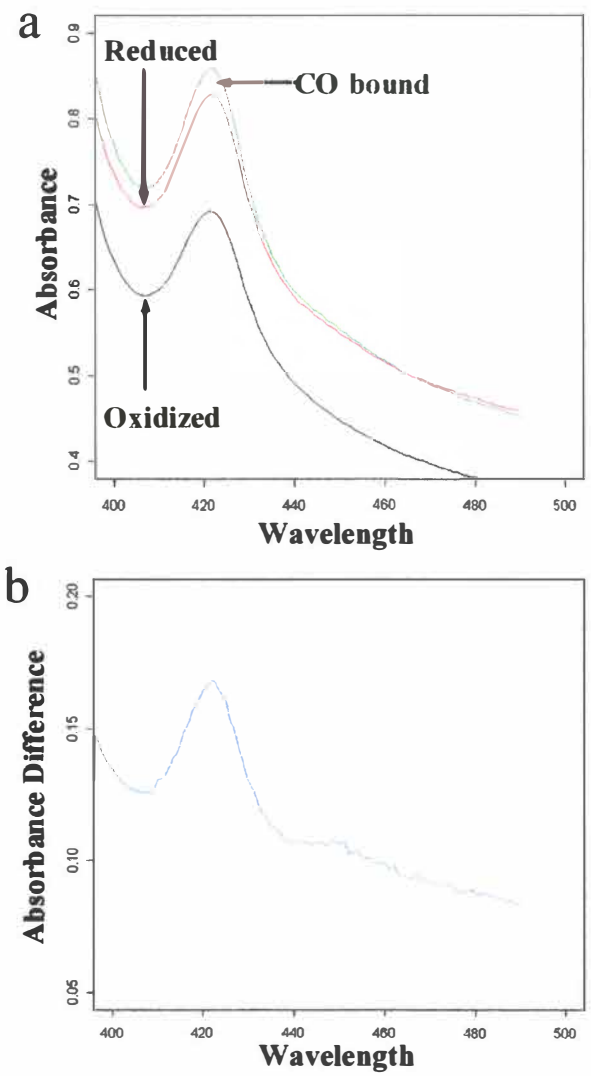


Figure 18. Carbon monoxide absorption difference spectrum of p450B-(his)$_6$. In part a, the black line represents the spectrum of oxidized p450B-(his) $_6$; the red line represents the spectrum of the reduced p450B-(his)$_6$ and the green line represents the spectrum of the reduced p450B-(his)$_6$ treated with carbon monoxide. Part b shows the absorbance difference spectrum that plots the difference in absorbance between the CO-bound form and the oxidized form of the p450B-(his)$_6$ The spectra of the purified p450B-(his)$_6$ in storage buffer were recorded from 300 to 600 nm. The ferric heme was then reduced by adding sodium dithionite to the protein solution, and the spectrum of the reduced enzyme was then recorded. The spectrum of a sample that was reduced with sodium dithionite and then treated with carbon monoxide was recorded too.

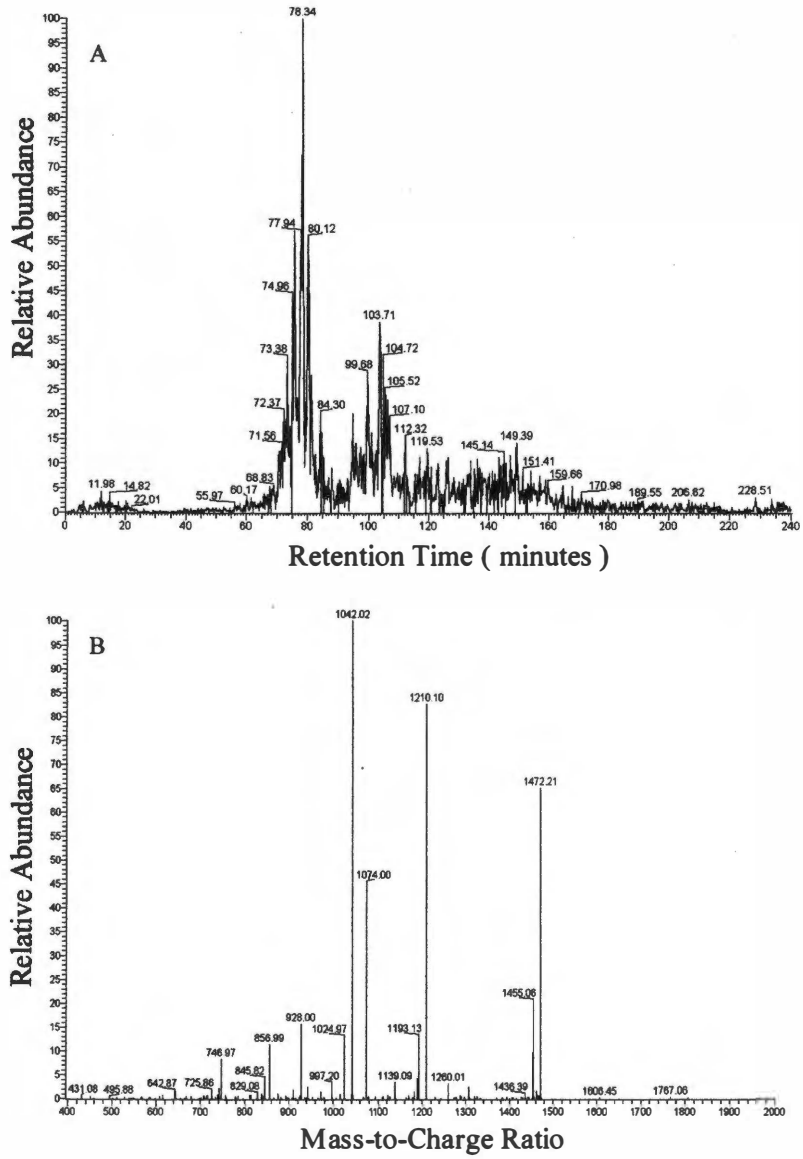


Figure 19. Liquid chromatography-tandem mass spectrometry (LC-MS^2) analysis. The purified P450B-$(his)_6$ recombinant protein was digested with trypsin and the resultant peptide mixture was examined with LC-MS^2 analysis. The base peak chromatogram is shown in part A. A selected tandem mass spectrum is shown in part B, which matches the peptide, LVADPMVMANPAPF, in the P450B-$(his)_6$ recombinant protein

p450B-$(his)_6$ recombinant protein were identified, which covers 57% of the protein sequence (Figure 20). The high sequence coverage of the p450B-$(his)_6$ protein confirms its presence in the sample. There are eighteen other proteins identified to be present in the sample (Table 1). They include ribosomal proteins and chaperon proteins, which are commonly found in the recombinant protein purification sample. The relative abundances of p450B-$(his)_6$ and the other identified proteins in the sample can be estimated with their spectrum count. The spectrum count of a protein is the number of tandem mass spectra matching to peptides from this protein, which positively correlates with the protein relative abundance. In this LC-MS^2 measurement, the spectrum count of p450B-$(his)_6$ is more than 10-fold higher than those of the other identified proteins. This suggests a good purity of p450B-$(his)_6$ in the sample.

MRQRLNWIAAHGLLRGTARLAARLGDVQSRLVADPMVMANPAPFCDE
LRAIGPVVSSYGTHLVVSHAIAHELLRSEDFEVVSLGSNLPAPMRWL
ERRTRDDTPHLLLPPSLLAVEPPNHTRYRKAVSSVFTPKAVAGLRDH
VEETASALLDQLTDQASAVDIIARYCSQLPVAVICDILGVPSRDRNR
VLKFGQLAGPCLDFGLTWRQHQQVRQGLQGLHFWITEHLEELRSNPG
DDLMSQMIHASENGSSETHLHATEVRMIGLVLGASFATTMDLLGNGI
QVLLDAPELRDALSQRPQLWPNAVEEILRLEPPVQLAGRMARKDTEV
AGTAIKRGQLVAIYLGAVNRDPSVFADPHRFDITRANANRHLAFSGG
RHFCLGAALARVEGEVGLRMLFERFPDVRAAGPGNRRDTRTLRGWSQ
LPVQLGAARSMAIRHHHHHH

Figure 20. Sequence coverage of the P450B-(his)$_6$ recombinant protein by the LC-MS^2 analysis. The LC-MS^2 analysis was used to identify the peptides from the enzymatic digestion of the purified P450B-(his)$_6$ sample . The identified peptides were then assembled computationally to the proteins. The sequence of the P450B-(his)$_6$ recombinant protein is shown with the identified peptides highlighted in red. The more the identified peptides covers the protein sequences, the more confident this protein is identified.

Table 1. The identified proteins and their sequence coverages and spectrum counts[c].

Protein Name	Sequence Coverage[a]	Spectrum Count[b]
P450B-$(his)_6$ recombinant protein	57%	457
50S ribosomal subunit protein L13	50%	10
50S ribosomal subunit protein L28	34%	7
50S ribosomal subunit protein L17	18%	2
50S ribosomal subunit protein L2	12%	2
30S ribosomal subunit protein S15	43%	6
Chaperone Hsp70	17%	18
Chaperone Hsp60	30%	30
Peptidyl prolyl cis-trans isomerase	30%	32
Cyclic AMP receptor protein	29%	7
Negative regulator protein	27%	12
Transcriptional repressor of lactose catabolism	24%	8
Conserved hypothetical protein	23%	3
Pyridine nucleotide transhydrogenase	14%	3
L-glutamine:D-fructose-6-phosphate aminotransferase	12%	8
Altronate hydrolase	9%	8
Putative formyltransferase	8%	5
2-oxoglutarate decarboxylase	8%	7

[a] the percentage of a protein's sequence covered by its identified peptides
[b] the number of tandem mass spectra matching to peptides from a protein
[c] David R. McWilliams are acknowledged for performing the LC-MS^2 measurement and data analysis

Chapter 4

Discussion

4.1 From Gene Isolation to Protein Purification

First expression experiments with the plasmid–encoding *M. ulcerans* P450 protein resulted in successful over expression though most of the protein obtained was insoluble and remained in the insoluble bodies. Several experiments were performed to determine the optimal host cell line and growth temperature for obtaining soluble protein. But the low product of soluble target protein prevents the further purification. Careful analysis of the P450A DNA sequence indicated some reasons. There were also limitations on using NCBI ORF server. Because where mycobacterium has three possible start codons (atg, gtg, or ttg), the NCBI ORF server searches the open reading frame only by start codon "atg". This suggested that the sequence of P450 cloned into PET30a was not the real open reading frame. Multiple alignment analysis of the P450A sequence which was used for cloning into PET30a with other confirmed P450s sequence indicated that the first 135 base pair in the P450A was not aligned with any other P450s sequence. P450B sequence was the part of P450A sequence without P450A's first 135 base pair and was aligned with other P450s from the very beginning as start codon "gtg". Another possible reason is the expression vector. PET expression system is one of the most successful expression system in *E.coli* expression. It utilizes the powerful T7 phage promoter Since T7 phage

promoter is one of the strongest promoters, This promoter may result in too powerful protein production in form of insoluble protein. The TAC promoter used in pCWori vector is 15 times weaker then the T7 promoter which may have facilitate production of soluble protein.

4.2 Protein Characterization

P450 monooxygenases are good candidates for spectroscopic characterization. They display typical absorption spectra, which make the identification convenient. The inability of purified P450B to exhibit the 450-nm reduced-CO difference spectrum typical of P450 enzymes, but instead displaying a 420 nm spectrum can be explained by the observation that cytochrome P450 enzymes commonly undergo, a transition to an inactive yet stable structural state referred to as P420, which was first described by Omura and Sato. It has been shown that P450-P20 conversion can be induced by temperature or pressure perturbations, as well as treatment with various chemicals or pH extrema.[36-39].

My inability to obtain recombinant p450B-$(his)_6$ that exhibited reduced CO difference spectra with a 450nm Soret absorbance maybe caused during the purification procedures or improper incorporation of the heme group into the apoenzyme in *E.coli*.

Reference

1. Ward, D.E., *Buruli ulcer.* Br Med J, 1970. **3**(718): p. 346.
2. Portaels, F., Chemlal, K., Elsen, P., Johnson, P.D., Hayman, J.A., Hibble, J., Kirkwood, R., and Meyers, W.M., *Mycobacterium ulcerans in wild animals.* Rev Sci Tech, 2001. **20**(1): p. 252-64.
3. van der Werf, T.S., van der Graaf, W.T., Tappero, J.W., and Asiedu, K., *Mycobacterium ulcerans infection.* Lancet, 1999. **354**(9183): p. 1013-8.
4. K.Asideu, R.S., &M.Raviglione,eds, *Buruli ulcer:Mycobacterium ulcerans infection.* WHO. Vol. WHO/CDS/CPE/GBUI/1 WHO. 2000. 118.
5. MacCallum, P., J.C.Tolhurst,G.Buckle,and H.A.Sissions, *A New Mycobacterial Infection in Man.* J Pathol Bacteriol, 1948. **60**: p. 4623-8.
6. Stinear, T., B. C.Ross,J.K.Davies,L.Marino,R.M.Robins-Browne,F.Oppedisano,A.Sievers,and P.D.Johnson, *Identification and characterization of IS2404 and IS2606:two distince repeated sequences for detection of Mycobacterium ulcerans by PCR.* J Clin Microbiol, 1999. **37**: p. 1018-23.
7. Thanagaraj, H.S., M.R.Evans,and M.H.Wansbrough-Jones., *Mycobacterium ulcerans disease;Buruli ulcer.* Trans R Soc Trop Med Hyg, 1999. **93**: p. 337-340.
8. Portaels, F., J.Aguiar,M.Debacker,C.Steunou,C.Zinsou,A.Guedenon,and W.M. Meyers., *Prophylactic effect of Mycobacterium bovis BCG vaccination against osteomyelitis in children with Mycobacterium ulcerans disease (Buruli Ulcer).* Clin Diagn Lab Immunol, 2002. **9**(6): p. 1389-91.
9. Connor, D.H., and H.F.Lunn., *Mycobacterium ulcerans infection(with comments on pathogenesis).* Int J lepr, 1965. **33**(Supple): p. 698-709.
10. Hockmeyer, W.T., R.E.Krieg,M.Reich,and R.D.Johnson., *Further characterization of Mycobacterium ulcerans toxin.* Infect Immun, 1978. **21**: p. 124-8.
11. Krieg, R.E., W.T.Hockmeyer,and D.H.Connor., *Toxin of Mycobacterium ulcerans:Production and effects in guinea pig skin.* Arch Dermatol, 1974. **110**: p. 783-8.
12. George, K.M., Barker, L.P., Welty, D.M., and Small, P.L., *Partial purification and characterization of biological effects of a lipid toxin produced by Mycobacterium ulcerans.* Infect Immun, 1998. **66**(2): p. 587-93.
13. George, K.M., Chatterjee, D., Gunawardana, G., Welty, D., Hayman, J., Lee, R., and Small, P.L., *Mycolactone: a polyketide toxin from Mycobacterium ulcerans required for virulence.* Science, 1999. **283**(5403): p. 854-7.
14. Song F, F.S., Benowitz AB, Kishi Y., *Total synthesis of the mycolactones.* Org Lett., 2002. **4**(4): p. 647-50.
15. Fidanze S, S.F., Szlosek-Pinaud M, Small PL, Kishi Y., *Complete structure of the mycolactones.* J Am Chem Soc., 2001. **123**(41): p. 10117-8.
16. Gunawardana, G.C., D.; George, K. M.; Brennan, P.; Whittern, D.; Small, P. L. C.;,

6(1): p. 1-42.

28. Montellano, P.R.O.d., ed. *cytochrome P450 Structure,Mechanism,and Biochemistry*. 2005, Kluwer Academic/Plenum Publishers: New York.
29. Cupp-Vickery JR, P.T., *Structure of cytochrome P450eryF involved in erythromycin biosynthesis.* Nat Struct Biol, 1995. **2**(2): p. 144-53.
30. Daniele Werck-Reichhart , R.F., *Cytochromes P450: a success story.* Genome Biology, 2000. **1**(6): p. REVIEWS3003.
31. Haqsler, J.A., *Human cytochromes P450.* Mol Aspects Med, 1999. **1999**(20): p. 1-137.
32. V.A. Lyubetsky, A.V.S., *Note on cliques and Alignments.* Mathematical models, computational methods, 2004(No3): p. 241-246.
33. Barnes, H.J., *Maximizing expression of eukaryotic cytochrome P450s in Escherichia coli.* Methods in enzymology, 1996. **272**: p. 3-14.
34. Yates, J.R., 3rd, Eng, J.K., and McCormack, A.L., *Mining genomes: correlating tandem mass spectra of modified and unmodified peptides to sequences in nucleotide databases.* Anal Chem, 1995. **67**(18): p. 3202-10.
35. Tabb, D.L., McDonald, W.H., and Yates, J.R., 3rd, *DTASelect and Contrast: tools for assembling and comparing protein identifications from shotgun proteomics.* J Proteome Res, 2002. **1**(1): p. 21-6.
36. Barnes, H.J., Arlotto, M.P., and Waterman, M.R., *Expression and enzymatic activity of recombinant cytochrome P450 17 alpha-hydroxylase in Escherichia coli.* Proc Natl Acad Sci U S A, 1991. **88**(13): p. 5597-601.
37. Hui Bon Hoa G, D.P.C., Geze M, Douzou P, Kornblatt JA, Sligar SG., *The formation of cytochrome P-450 from cytochrome P-420 is promoted by spermine.* Biochemistry., 1990. **29**(29): p. 6810-5.
38. Hui Bon Hoa G, D.P.C., Dondaine I, Sligar SG, Gunsalus IC, Douzou P., *Conformational changes of cytochromes P-450cam and P-450lin induced by high pressure.* 28, 1989. **2**: p. 651-6.
39. Healy FG, K.S., Wach M, Gibson DM, Loria R., *Involvement of a cytochrome P450 monooxygenase in thaxtomin A biosynthesis by Streptomyces acidiscabies.* J Bacteriol, 2002. **184**(7): p. 2019-29.

Vita

Yang Wang, originally from Wuhan, China, earned a Bachelor of Medicine degree from Tongji Medical College, Huazhong University of Science and Technology in May 2001. She began her graduate studies in microbiology at the University of Tennessee in January 2002, graduating with a Master of Science degree in the summer of 2005.

4533 0543, 57
11/09/05 MAB