

**Design and optimisation of operon
constructs for expression and
characterisation of Class I cytochrome
P450s from actinomycetes**

Yang Lu

University College London

Thesis submitted to UCL for the degree of

Doctor of Philosophy

In

Biochemical Engineering

2020

Declaration

The genomic materials of *Bacillus licheniformis* ATCC 14580, *Streptomyces avermitilis* MA-4680 and *Saccharopolyspora erythraea* NRRL2338 were prepared by Professor John Ward and given for mining of self-sufficient cytochrome P450s in this research program. The genomic material of *Streptomyces rishiriensis* DSM 40489 was prepared by Thomas Hickman and given for mining of Class I cytochrome P450s. The dataset of drain metagenome was maintained by Dr Dragana Dobrijevic and given access for mining of Class I cytochrome P450s.

I, Yang Lu, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

Abstract

Cytochrome P450s are currently indispensable enzymes for biotechnological applications due to the variety of catalytic functions they mediate. Actinomycetes are regarded as extremely rich and valuable sources of soluble Class I cytochrome P450s. Therefore, the study of actinomycete P450s is necessary for both academia and industry, which may lead to the discovery of novel biocatalysts. The focus of this project was to establish an effective and stable expression system to produce actinomycete cytochrome P450s in *E.coli* BL21. After acquiring stable production of target P450s, further enzyme characterisation and application for biocatalysis could be carried out.

The project aimed to design the operon containing standardized genes of Class I P450, ferredoxin and ferredoxin reductase, as well as to evaluate the heterologous expression of active cytochrome P450 in *E.coli* BL21 (DE3) using the designed operon.

Initially, the operon was designed and tested (Chapter 4). The construct had three consecutive gene components assembled in the order of *cyp*, *ferredoxin* and *ferredoxin reductase*, which was controlled by a single T7 promoter at upstream. The ends of each gene component were standardized with restriction sites. The operon could lead to massively enhanced expression of some CYPs such as CYP105B1 (*Streptomyces griseolus*), CYP105AB1 (*Saccharopolyspora erythraea*) and CYP154C2 (*Streptomyces avermitilis*).

Next, the operon was successfully applied in the production of CYP105AB1 for enzyme studies (Chapter 5). The preliminary substrate screening for CYP105AB1 indicated diclofenac could be accepted. Further product characterisation confirmed the aromatic hydroxylation by CYP105AB1. Different electron transfer systems were evaluated for their cooperation with CYP105AB1 on diclofenac bioconversion. In the end, the rational engineering of CYP105AB1 was also demonstrated to create mutants.

Finally, the discovery of novel CYPs through genome mining was carried out in *Streptomyces rishiriensis* DSM 40489 (Chapter 6) and metagenome of a household drain (Chapter 7). Several novel CYPs were assembled into designed operon for stable expression and characterisation.

Impact statement

The major work devoted to this research has established an operon for the production and characterisation of Class I cytochrome P450s from actinomycetes. The utility of the operon has led to stable productions of various Class I cytochrome P450s from at least 5 different actinomycetes. The other focus of this research has been developing genome mining strategies on either single organism genome or metagenome datasets to discover novel cytochrome P450s. Several potential benefits are arising from this research that could have impacts across academia and industry.

From the academic perspective, the establishment of the operon containing *cyp* and two redox partner genes could facilitate the exploration of Class I cytochrome P450 from actinomycetes. Within the three-gene operon, it is designed to have standardized genetic components, which allows each gene to be assembled or replaced at designated positions. There is high freedom in this operon for (re-)design and fabrication of the genetic cluster of CYP and redox partners that do not already exist in nature. The operon might be a solution to difficult expression of Class I cytochrome P450s potentially existing in other research groups. The detailed performance that we have learned on the operon could also be written up for publication to facilitate wider dissemination. The genome mining work has resulted in the detailed collections of cytochrome P450s in *Streptomyces rishiriensis* DSM 40489 and the household drain microbial community. These collections expand the current superfamily of cytochrome P450s and serve as an informative database for other research groups.

Moreover, the benefit arising from this research has already impacted commercial projects in the industry. The operon is used for preliminary expression of various microbial Class I P450s potentially studied and developed for generating scalable quantities of oxidised metabolites at Hypha Discovery.

Overall, this research has generated results that may have genuine impacts across academia and industry either immediately or in the future.

Acknowledgements

First and foremost, I would like to express my sincere gratitude to my PhD supervisors Professor John Ward and Professor Helen Hailes for their support and guidance throughout my entire research project. I was lucky to be given the opportunity to experience research in the enzyme related project, and I was privileged to supervised by two Professors who never cease to enlighten me with incredible knowledge in their fields.

I am very grateful that Dr Benjamin Lichman patiently taught me through many basic molecular techniques in the first half years of the project. I am also very grateful for Dr Sebastian Schulz for his extensive understanding of microbial cytochrome P450s and for the time and efforts we spend together on designing many operon constructs used or not used in this research. I need to thank Dr Dragana Dobrijevic allowing me access to the metagenome datasets as well as explaining how genes are discovered and annotated along the metagenome. I also want to thank Dr Maria Bawn for her constant helps in troubleshooting the HPLC either instrument or results.

I am also grateful that Prof Helen Hailes allows me to use the semi-prep HPLC, LC/MS, and NMR at the Department of Chemistry. Moreover, many thanks to Jianxiong and Yu at the Chemistry Department for teaching me how to use these instruments as well as interpreting the results.

In the last four years, I have seen so many bright minds enjoying research at Ward Lab. I want to thank you all for your kind help, interesting discussions, and all the time we had in and out of the Lab. Those come before me: Benjamin, John K, Sebastian, Aisha, Dave, and Mike. Many who still in the lab: Jack, Dragana, Maria, James, Luba, Rosalia, and Tom. I must extend my gratitude to members at the Hailes group, who always pointed me directions whenever I got lost in the Chemistry Labs.

Lastly, I have my endless gratitude for my parents for their unwavering support and encouragement. For the many years of my study at UCL, the 8-hour time difference and thousands of miles distance never stopped them caring and encouraging me. The entire thesis and whatever I may achieve in the future owe much to them.

List of Figures

Figure 1. 1. The structure of heme B and the electronic configuration of the heme iron in cytochrome P450.....	24
Figure 1. 2. The landscape of bioprocess involving cytochrome P450 enzymes and some well-known examples..	25
Figure 1.3. The cytochrome P450 catalytic cycle.	30
Figure 1.4. Selected classes of P450s based in their electron transfer partners..	34
Figure 1.5. P450cam structure with all 13 helices labelled (PDB 2ZWT)(45, 69).	37
Figure 1.6. Dynamic regions of P450cam model involved in substrate binding (PDB 2ZWT)(69).....	38
Figure 1.7. Active sites of P450cam model involved in substrate binding and positioning (PDB 2ZWT)(69).....	39
Figure 1.8. Genetic colocation of P450cam and its redox partners.	41
Figure 1.9. In comparison of the processes involved in directed evolution and rational design(94).	46
Figure 2.1. Analytical HPLC method 1.	71
Figure 2.2. Analytical HPLC method 2.	72
Figure 2.3. Semi-preparative HPLC method.	73
Figure 2.4.. Liquid chromatography method.	74
Figure 2.5. Gradient of oven temperature for gas chromatography.....	74
Figure 2.6. The sample preparation for the reaction of diclofenac by CYP105AB1.	77
Figure 2.7. The sample preparation process for bioconversion of diclofenac using different CYP105AB1 lysate.	78
Figure 3.1. Genomic PCR for producing CYP102A7 gene fragments.	88
Figure 3.2. The approach of assembling expression plasmid containing <i>cyp102a7</i> gene.	89
Figure 3.3. SDS-PAGE analysis and protein purification of CYP102A7.....	90
Figure 3.4. Carbon monoxide difference spectra of CYP102A7.	91
Figure 3.5. Baseline readings for control system without adding any substrates.....	94

Figure 3.6. Potential fatty acids and terpenes catalysed by CYP102A7.....	95
Figure 4.1. Original operon for CYP, ferredoxin and ferredoxin reductase on pQR367 and pQR368.	101
Figure 4.2. An alternative operon for expression of CYP, ferredoxin and ferredoxin reductase.	102
Figure 4.3. Operon construction on plasmid pQR2239 and pQR2240.	104
Figure 4.4. The reduction of cytochrome <i>c</i> by cell lysates containing different ferredoxins and ferredoxin reductases.	105
Figure 4.5. Amplification of <i>cyp105a1</i> and <i>cyp105b1</i> from pQR367 and pQR368.	106
Figure 4.6. Expression plasmids for CYP105A1 and CYP105B1.	107
Figure 4.7. The expression level of CYP105A1 and CYP105B1.	108
Figure 4.8. The single-gene construct and three-gene operon for P450 expression.....	112
Figure 4.9. The expression level of each P450s when using <i>cyp-fd1-fdr</i> or <i>cyp-fd2-fdr</i> operons.....	113
Figure 4.10. The expression level of each P450 when using <i>cyp-fd2-fdr</i> or <i>cyp-pdx-pdr</i> operons.....	115
Figure 4.11. The expression level of P450s from <i>S.erythraea</i> when coupling with different ferredoxin genes..	119
Figure 4.12. The expression level of CYPs using <i>cyp-fd2-fdr</i> operon for different expression time.	121
Figure 4.13. The sequential differences between the old and new constructs..	123
Figure 5.1. The reconstruction strategy for inserting histidine tags to <i>cyp</i> gene within the <i>cyp-fd2-fdr</i> operon.....	128
Figure 5.2. SDS-PAGE analysis of purified CYPs, Pdx and PdR. Lane M is the marker lane with standard size protein bands.....	129
Figure 5.3. The reduction of cytochrome <i>c</i> by purified putidaredoxin and putidaredoxin reductase..	131
Figure 5.4. HPLC analysis of enzyme reaction conducted by CYP105AB1 on diclofenac.	135

Figure 5.5. Standard correlation between diclofenac concentration and peak area on HPLC chromatogram.	136
Figure 5.6. The mass spectra of NADH, diclofenac and potential products.. ..	138
Figure 5.7. The partial spectra of diclofenac and product.. ..	140
Figure 5.8. Bioconversion of diclofenac by purified CYP105AB1.	142
Figure 5.9. Concentrations of CYP105AB1 in two different cell lysates. .	143
Figure 5.10. Bioconversion of diclofenac by CYP105AB1 cell lysate..	145
Figure 5.11. Regional multiple alignment among CYP105AB1, CYP105A1 and P450 MoxA.	147
Figure 5.12. Point mutation PCR using Phusion site directed mutagenesis kit.	149
Figure 5.13. Carbon monoxide assay for CYP105AB1-wt, CYP105AB1-T108A and CYP105AB1-G287D.....	150
Figure 5.14. Proposed hydroxylation on diclofenac conducted by CYP105AB1.. ..	154
Figure 6.1. A paralogous tree of all 24 CYP sequences from <i>S.rishiriensis</i> . ..	164
Figure 6.2. Neighbour-joining tree of CYPs from 4 Streptomyces.	167
Figure 6.3. Amplification of <i>sri_cyp03</i> , <i>sri_cyp13</i> and <i>sri_cyp24</i> from <i>S.rishiriensis</i> . ..	171
Figure 6.4. Expression of Sri_CYP03 and Sri_CYP13 using <i>cyp-fd2-fdr</i> construct with different IPTG concentrations.....	173
Figure 6.5. Expression of Sri_CYP03 and Sri_CP13 using single-cyp construct.	174
Figure 6.6. The successful co-expression of active Sri_CYP03 and molecular chaperone.....	176
Figure 6.7. The successful co-expression of active Sri_CYP13 and molecular chaperone.....	177
Figure 6.8. SDS-PAGE analysis of purified Sri_CYP03 and Sri_CYP13...	178
Figure 6.9. Expression of Sri_CYP24 using <i>cyp-fd2-fdr</i> operon or single-cyp construct.	180

Figure 6.10. The unsuccessful co-expression of Sri_CYP24 and molecular chaperones.....	182
Figure 6. 11. The alignment between CYP107P2 and Sri_CYP03.	184
Figure 6.12. The alignment between CYP125A2 and Sri_CYP13.	185
Figure 6.13. The potential usage of these substrates for enzyme reaction conducted by CYP_Sri03 and CYP_Sri13.....	188
Figure 7.1. Examples of truncated CYPs found in metagenome contigs..	197
Figure 7.2. Amplification of <i>dmg_cyp08</i> and <i>dmg_cyp09</i> from drain metagenome datasets.	200
Figure 7.3. Digestion of <i>dmg_cyp08</i> with <i>NdeI</i> and <i>EcoRI</i>	201
Figure 7.4. Expression of Dmg_CYP08 and Dmg_CYP09 using <i>cyp-fd2-fdr</i> construct.	202
Figure 7.5. SDS-PAGE analysis of purified Dmg_CYP08 and Dmg_CYP09.	203
Figure 7.6. The extraction efficiency for different alkane substrates.....	205
Figure 7.7. The potential usage of alkanes as substrates for enzyme reaction conducted by Dmg_CYP08 and Dmg_CYP09.....	207
Figure B1. The calculated standard curve of 7-Ethoxycoumarin (EC).....	232
Figure B2. The calculated standard curve of Indole (IND)..	233
Figure B3. The calculated standard curve of tyrosine (TYR)..	234
Figure B4. The calculated standard curve of phenylalanine (PHA).....	235
Figure B5. The calculated standard curve of tyramine (TYA).	236
Figure B6. The calculated standard curve of phenylethylamine (PHEA).	237
Figure B7. The calculated standard curve of 1,2,3,4- Tetrahydroisoquinoline (THIQ).....	238
Figure B8. The calculated standard curve of DIC.	239
Figure B9. The negative screening results of CYP105AB1.	240
Figure B10. ¹ H -NMR spectrum for diclofenac.	241
Figure B11. ¹ H -NMR spectrum for 4'hydroxydiclofenac.	242
Figure B12. The identity matrix of CYPs in the multiple alignment.....	245
Figure C1. Multiple alignments of GXXTT, EXXR and heme-binding motif of all cytochrome P450s from <i>S.rishiriensis</i>	252

Figure C2. Coexpression of CYPs and five different chaperone systems.	255
Figure C3. Coexpression of CYPs and GroES/GroEL chaperone system..	255
Figure C4. The calculated standard curve of apigenin.....	257
Figure C5. The calculated standard curve of chrysin.....	258
Figure C6. The calculated standard curve of 7-ethoxycoumarin.....	260
Figure C7. The calculated standard curve of diclofenac.....	262
Figure C8. The calculated standard curve of THIQ.....	264
Figure C9. The screening results of CYP_Sri03 and CYP_Sri13 against five different substrates.	265
Figure D1. GC chromatogram for ethyl acetate.	269
Figure D2. GC chromatogram and mass spectrum for 0.5 mM nonane (C ₉ H ₂₀).	270
Figure D3. GC chromatogram and mass spectrum for 0.5 mM decane (C ₁₀ H ₂₂).	271
Figure D4. GC chromatogram and mass spectrum for 0.5 mM undecane (C ₁₁ H ₂₄).	272
Figure D5. GC chromatogram and mass spectrum for 0.5 mM dodecane (C ₁₂ H ₂₆).	273
Figure D6. GC chromatogram and mass spectrum for 0.5 mM tridecane (C ₁₃ H ₂₈).	274
Figure D7. GC chromatogram and mass spectrum for 0.5 mM tetradecane (C ₁₄ H ₃₀).	275
Figure D8. GC chromatogram and mass spectrum for 0.5 mM pentadecane (C ₁₄ H ₃₀).	276
Figure D9. GC chromatogram enzyme assays on nonane (C ₉ H ₂₀).	278
Figure D10. GC chromatogram enzyme assays on decane (C ₁₀ H ₂₂).....	279
Figure D11. GC chromatogram enzyme assays on undecane (C ₁₁ H ₂₄). 280	
Figure D12. GC chromatogram enzyme assays on dodecane (C ₁₂ H ₂₆).	281
Figure D13. GC chromatogram enzyme assays on tridecane (C ₁₃ H ₂₈).	282

Figure D14. GC chromatogram enzyme assays on tetradecane (C₁₄H₃₀).	
.....	283
Figure D15. GC chromatogram enzyme assays on pentadecane (C₁₅H₃₂).	
.....	284

List of Tables

Table 1. 1. Current classifications of enzymes.	22
Table 1.2. Summary of common reactions catalysed by P450cam (CYP101A1).	28
Table 1.3. Genetic colocation of P450 and redox partners in some actinomycetes.....	42
Table 2.1. Complete list of plasmids generated in each project.....	52
Table 2.2. Primers used in Eurofins Genomics sequencing service.....	58
Table 2.3. Reaction setup for routine PCR by using polymerase master mix.	59
Table 2.4. Thermocycling conditions for a routine PCR.....	59
Table 2.5. Reaction setup for routine restriction digestion.....	60
Table 2.6. Routine ligation reaction setup by using T4 DNA ligase.	60
Table 2.7. Buffers used in Ni-NTA purification system.....	66
Table 3.1. The complete protein sequences of Class VII P450s identified in <i>Bacillus licheniformis</i> ATCC 14580, <i>Streptomyces avermitilis</i> MA-4680 and <i>Saccharopolyspora erythraea</i> NRRL2338.	85
Table 3.2. The process of calculating molar amount of CYP102A7 produced per litre of cell culture.	92
Table 3.3. The main differences between previous and current expression of CYP102A7 in <i>E.coli</i> BL21 (DE3).....	93
Table 3.4. Calculation of NADPH consumption rate in first 20 min reaction period.....	96
Table 3.5. The main differences between previous and current substrate screening reactions.....	97
Table 4.1. Two pairs of primers used for amplification of <i>cyp105a1</i> and <i>cyp105b1</i>	106
Table 4.2. Enhanced expression of CYP105A1 and CYP105B1 by using different operons.	109
Table 4.3. Primers designed for PCR of five <i>cyp</i> genes.....	110
Table 4. 4. The variance of expression levels is summarized below.	116
Table 4.5. The expression levels of various P450s using different operon constructs..	123

Table 5.1. Primers used for reconstructing <i>cyp_his-fd2-fdr</i> operon.	127
Table 5.2. Concentrations of enzyme and cofactors used in enzyme assays and controls.	133
Table 5.3. The substrate panel for CYP105AB1.	133
Table 5.4. The calculated concentrations of initial diclofenac and remaining diclofenac in reaction with CYP105AB1.	136
Table 5.5. Reaction setup for bioconversion on diclofenac by cell lysates.	144
Table 5.6. Primers designed for site-directed mutations of <i>cyp105ab1</i> gene on pQR2281.	149
Table 6.1. 24 putative cytochrome P450s discovered from <i>Streptomyces rishiriensis</i> genome.	159
Table 6.2. A comparison of motif sequences in I-helix and K-helix of putative cytochrome P450s found in <i>S.rishiriensis</i>	161
Table 6.3. Closest homologues to putative cytochrome P450s found in <i>S.rishiriensis</i>	162
Table 6.4. Putative ferredoxins and ferredoxin reductases discovered from <i>S.rishiriensis</i> genome.	167
Table 6.5. Primers designed for PCR of <i>sri_cyp03</i> , <i>sri_cyp13</i> and <i>sri_cyp24</i> from <i>S.rishiriensis</i>	170
Table 6.6. Five different chaperone systems used for protein expression.	175
Table 6.7. Substrates list for preliminary screening.	186
Table 7.1. Putative cytochrome P450s discovered from drain metagenome (Dmg) datasets.	194
Table 7.2. Closest homologues to putative cytochrome P450s found in drain metagenome.	197
Table 7.3. Primers designed for PCR of <i>dmg_cyp08</i> and <i>dmg_cyp09</i> from drain metagenome datasets.	199
Table 7.4. Alkane substrates used for preliminary screening.	204
Table 7.5. Reaction setup for alkane screenings.	206
Table A1. Substrate screening list for CYP102A7.	224

Table A2. The ENA records of complete genome sequences of <i>Bacillus licheniformis</i> ATCC 14580, <i>Streptomyces avermitilis</i> MA-4680 and <i>Saccharopolyspora erythraea</i> NRRL2338.	225
Table A3. All failed PCR attempts for extracting <i>cyp102d1</i> and <i>cyp102g2</i> from their genomes.	225
Table B1. The sequential similarity between the candidate P450 at Hypha Discovery and its closest homologous P450s identified in five actinomycetes.....	229
Table B2. Substrates list for preliminary screening.	230
Table C1. Cytochrome P450s involved in secondary metabolites production by <i>Streptomyces</i>.....	252

List of Abbreviations

2Fe-2S	2 Iron-2 Sulphur cluster
Amp	Ampicillin
ATCC	American Type Culture Collection
Chl	Chloramphenicol
CPR	Cytochrome P450 Reductase
Da	Dalton
ddH₂O	Distilled, deionised water
DMSO	Dimethyl sulfoxide
dNTPs	Deoxynucleotide triphosphates
DSMZ	Deutsche Sammlung von Mikroorganismen und Zellkulturen
DTT	Dithiothreitol
EDTA	Ethylenediaminetetraacetic acid
ENA	European Nucleotide Archive
FAD	Flavin Adenine Dinucleotide
FMN	Flavin Mononucleotide
Fdx	Ferredoxin
FdR	Ferredoxin Reductase
Fldx	Flavodoxin
GC	Gas Chromatography
HCl	Hydrochloric acid
HPLC	High-Pressure Liquid Chromatography
IPTG	Isopropyl- β ,D-Thiogalactopyranoside
Kan	Kanamycin
Kb	Kilobase
kDa	Kilodalton
KPi	Potassium phosphate
LCMS	Liquid Chromatography-Mass Spectrometry
MS	Mass Spectrometry
NADH	β -Nicotinamide Adenine Dinucleotide (reduced)
NADPH	β -Nicotinamide Adenine Dinucleotide Phosphate (reduced)
NCBI	National Centre for Biotechnology Information
nm	Nanometre
NMR	Nuclear Magnetic Resonance
OD	Optical Density
PCR	Polymerase Chain Reaction
Pd	Putidaredoxin
PdR	Putidaredoxin Reductase
PDOR	Phthalate dioxygenase reductase
SDS	Sodium Dodecyl Sulfate
SRS	Substrate Recognition Site
TIC	Total Ion Current
T_m	Melting Temperature
TAE	Tris Acetic acid Buffer
UV	Ultraviolet

Table of Contents

Declaration	2
Abstract	3
Impact statement	4
Acknowledgements	5
List of Figures	6
List of Tables	12
List of Abbreviations	15
1 Chapter 1: Introduction	22
1.1 Background	22
1.1.1 Enzymes and biocatalysts.....	22
1.1.2 Metalloproteins and hemoproteins	23
1.2 Cytochrome P450 Monooxygenases.....	24
1.2.1 Physiological roles of P450	26
1.2.2 Typical reactions catalysed by P450	28
1.2.3 The P450 catalytic cycle.....	29
1.2.4 Nomenclature and classification of cytochrome P450s	32
1.2.5 Structure and function of cytochrome P450	36
1.3 Redox partners	40
1.3.1 Flavoproteins and electron transfer types	40
1.3.2 Colocation of P450 and redox partners	41
1.4 Analytical assays in cytochrome P450 study	42
1.4.1 Carbon monoxide assay.....	42
1.4.2 Cytochrome <i>c</i> assay	43
1.5 Perspectives of discovering and engineering P450s	44
1.5.1 Discovering new P450s through genome mining	44
1.5.2 Discovering new P450s through engineering	46
1.6 Research aims.....	49
2 Chapter 2: Materials and Methods	52

2.1	General materials and methods.....	52
2.1.1	Chemical compounds.....	52
2.1.2	Genes and plasmids.....	52
2.1.3	Strains and media	54
2.1.4	Transformation, inoculation, and glycerol stock.....	56
2.1.5	Sub-cloning and sequencing	56
2.1.6	DNA manipulation	58
2.1.7	Protein manipulation.....	64
2.1.8	Carbon monoxide assay.....	68
2.1.9	Cytochrome c assay	69
2.1.10	Analytical methods	70
2.2	Materials and methods for CYP105AB1 characterisation	75
2.2.1	Enzyme assay for preliminary substrate screening	75
2.2.2	Product characterisation.....	75
2.2.3	Cell-free bioconversion.....	76
2.2.4	Site-directed mutagenesis.....	79
2.3	Materials and methods for <i>Streptomyces rishiriensis</i> genome mining .	80
2.3.1	Custom tBLASTn.....	80
2.3.2	Co-expression of CYPs and molecular chaperones in <i>E.coli</i>	80
2.3.3	Preliminary substrate screening	81
2.4	Materials and methods for Drain metagenome mining	82
2.4.1	Custom tBLASTn.....	82
2.4.2	Preliminary substrate screening	82
3	Chapter 3: Building genetic constructs for heterologous expression of self-sufficient cytochrome P450 from <i>Bacillus licheniformis</i>	84
3.1	Background	84
3.2	CYP102A7 cloning and expression	87
3.3	Carbon monoxide (CO) assay	90

3.4	Preliminary substrate screening	93
3.5	Summary	97
4	Chapter 4: Design and optimisation of operon constructs for heterologous expression of Class I cytochrome P450.....	100
4.1	Expression plasmids constructed for Class I cytochrome P450 in <i>E.coli</i> BL21(DE3).....	100
4.1.1	Introduction and aims	100
4.1.2	Original operon constructed in pQR367 and pQR368	101
4.1.3	An alternative operon design.....	102
4.1.4	Construction of pQR2239 and pQR2240.....	103
4.1.5	Cloning and expression of CYP105A1 and CYP105B1.....	105
4.2	The effects of ferredoxin and ferredoxin reductase on the expression of Class I cytochrome P450s.....	110
4.2.1	<i>cyp-fd1-fdr</i> operon and <i>cyp-fd2-fdr</i> operon.....	110
4.2.2	<i>cyp-fd2-fdr</i> operon and <i>cyp-pdx-pdr</i> operon	114
4.2.3	Analysis of variance (ANOVA) test on expression level using different operon constructs.....	116
4.3	Optimised expression of CYP105AB1 from <i>Saccharopolyspora erythraea</i> NRRL2338.....	117
4.3.1	Ferredoxin 2 from <i>Streptomyces griseolus</i> and Ferredoxins from <i>Saccharopolyspora erythraea</i> NRRL2338.....	117
4.3.2	The effects of expression time when using <i>cyp-fd2-fdr</i> operon ...	120
4.4	Summary	122
5	Chapter 5: Characterisation of CYP105AB1 from <i>Saccharopolyspora erythraea</i> NRRL2338.....	125
5.1	Introduction.....	125
5.2	Expression and purification of CYP105AB1	127
5.3	Substrate screening for CYP105AB1	130
5.3.1	The use of Pdx-PdR system and cofactor NADH.....	130

5.3.2	Preliminary substrate screening	131
5.3.3	Product characterisation.....	137
5.4	The performance of cell-free biotransformation against diclofenac ...	141
5.4.1	Bioconversion by purified CYP105AB1 and Pdx-PdR system.....	141
5.4.2	Bioconversion by cell lysate	142
5.5	Construction of CYP105AB1 mutants and potential challenges	146
5.6	Summary	152
6	Chapter 6: Identification and functional analysis of cytochrome P450s discovered in <i>Streptomyces rishiriensis</i> DSM 40489	157
6.1	Introduction.....	157
6.1.1	Background and project aims	157
6.1.2	<i>Streptomyces rishiriensis</i> cultivation and genome sequencing....	157
6.2	Bioinformatic analysis of <i>S.rishiriensis</i> genome.....	158
6.2.1	Discovery of cytochrome P450s	158
6.2.2	Features of cytochrome P450s from <i>S.rishiriensis</i>	160
6.2.3	Multiple alignments and phylogenetic analysis.....	163
6.2.4	CYPs clustered with ferredoxin or ferredoxin reductases.....	167
6.2.5	CYPs involved in the production of secondary metabolites	168
6.2.6	Predictions of functions of some CYPs in <i>S.rishiriensis</i>	169
6.3	Cloning and expression of <i>sri_cyp03</i> , <i>sri_cyp13</i> and <i>sri_cyp24</i>	170
6.3.1	Cloning and plasmid construction.....	170
6.3.2	Expression of three <i>S.rishiriensis</i> CYPs in <i>E.coli</i>	171
6.4	Preliminary substrate screening for Sri_CYP03 and Sri_CYP13	183
6.4.1	Preliminary substrate pool.....	183
6.4.2	Screening results.....	187
6.5	Summary	188
7	Chapter 7: Identification and functional analysis of cytochrome P450 from a drain metagenome	193

7.1	Introduction.....	193
7.2	Bioinformatic analysis of drain metagenome	194
7.3	Cloning of <i>dmg_cyp08</i> and <i>dmg_cyp09</i>	199
7.3.1	Cloning and plasmid construction.....	199
7.3.2	Expression and purification of Dmg_CYP08 and Dmg_CYP09...	201
7.4	Preliminary substrate screening for Dmg_CYP08 and Dmg_CYP09	203
7.4.1	Alkane standards.....	203
7.4.2	Screening results of Dmg_CYP08 and Dmg_CYP09	206
7.5	Summary	207
8	Chapter 8: Conclusion and future work.....	212
8.1	Self-sufficient cytochrome P450	212
8.2	Design and optimisation of operon constructs for Class I CYP expression	213
8.3	Discovery of cytochrome P450s from <i>Streptomyces rishiriensis</i>	216
8.4	Discovery of cytochrome P450s from drain metagenome	220
9	Appendix A.....	224
10	Appendix B.....	227
10.1	The gene and protein sequences	227
10.1.1	<i>cyp105a1</i> , <i>cyp105b1</i> , <i>fd1</i> and <i>fd2</i> from <i>Streptomyces griseolus</i>	227
10.1.2	<i>cyp105ab1</i> from <i>S.erythraea</i> NRRL2338.	228
10.1.3	<i>fd_sery</i> from <i>S.erythraea</i> NRRL2338.	229
10.2	Discovery of CYP105AB1.....	229
10.3	Preliminary substrate screening for CYP105AB1	230
10.3.1	Substrates and their standard curves.....	230
10.3.2	The screening results for CYPs.....	239
10.4	NMR spectra of diclofenac and its product	241
10.5	Multiple alignment among CYP105 subfamily	243

10.6	Multiple alignment among CYP105AB1, CYP105A1 and P450_MoxA	245
11	Appendix C	246
11.1	Protein sequences of 24 CYPs identified in <i>S.rishiriensis</i>	246
11.2	Conserved regions of all 24 P450s identified in <i>S.rishiriensis</i>	252
11.3	Secondary metabolites	252
11.4	Expression of Sri_CYP03 and Sri_CYP13 with every chaperone system.....	253
11.5	The calculated standard curves for substrates used in preliminary screening.....	256
11.6	Screening results.....	264
11.7	Original gel image of Figure 6.8.....	266
12	Appendix D	267
12.1	Complete gene sequences for <i>dmg_cyp08</i> and <i>dmg_cyp09</i>	267
12.2	GC chromatograms and mass spectra of 0.5 mM alkanes.....	268
12.3	Preliminary screening of 7 alkane substrates by Dmg_CYP08 and Dmg_CYP09.....	277
13	Bibliography	285

1 Chapter 1: Introduction

1.1 Background

1.1.1 Enzymes and biocatalysts

The biomolecules that could catalyse physiologically relevant reactions are enzymes, and the development of enzymes for harnessing their catalytic ability has increased over recent years and become important industrially⁽¹⁾. Based on different chemical reactions that enzymes could conduct, there are 6 major classes of enzymes being studied in Table 1.1⁽²⁾. The selectivity of an enzyme means the type of substrates it can accept and the type of reaction it performs, which is tightly related to the tertiary structure at the catalytic site. In the current study of biocatalysts, the gaps of the desired selectivity for reactions of interest are gradually filled by the rapid discovery of new enzymes and the growing capacity of enzyme engineering.

Table 1. 1. Current classifications of enzymes⁽²⁾. There are six major classes of enzymes being recognised so far. The essential features of these reactions and the representative enzymes are shown below. The cytochrome P450s are classified as oxidoreductases. The detailed nomenclature of cytochrome P450s was described in section 1.2.4.

Class	Reaction	Example enzymes
Oxidoreductases	Transfer of electrons	Cytochrome P450s, Alcohol dehydrogenase
Transferases	Transfer of functional groups	Transaminases, glycosyltransferases
Hydrolases	Hydrolysis of a chemical bond	Lipases, proteases, GTPases
Lyases	Breaking of bonds by means other than hydrolysis	Carbonic anhydrase, ammonia lyases
Isomerases	Isomer conversion	Cyclases, Epimerases
ligases	Linkage of two macromolecules	t-RNA ligases, polyketide synthases

In the past years, enzyme catalysts have been studied and applied in chemical synthesis, which provides an alternative perspective of building synthetic schemes for complex chemicals^(3, 4). Not only chemical reagents can catalyse the organic synthesis reactions, but the biological macromolecules like enzymes can be also involved in various synthetic reactions in industry⁽⁵⁾. Over the past century, there has been a tremendous development of the classical chemical catalysts⁽⁶⁾. The catalytic steps, catalyst efficiency, and waste recycling have all been developed and improved⁽⁷⁾. Whereas, the biocatalysts have only been in use

since the 1960s, and there is still a lot to discover and learn. Compared to chemical catalysts, biocatalysts have at least four widely recognized advantages⁽⁶⁾. First is the potential of high regio-selectivity and stereo-selectivity⁽⁸⁾, which directly helps to avoid high-cost separation steps and undesired isomers. Second is that enzymes normally perform reactions at moderate conditions (normal temperature and pressure) in aqueous media, which allows the reactions to be easily set up and carried out at an acceptable rate at moderate conditions. Moreover, enzymes are biomolecules that are naturally degradable, which leaves fewer residual contaminations after the reaction. The last advantage is the potential of reducing the cost offered by a mild enzymatic reaction in comparison with conventional chemical synthesis. To be specific on cost-saving, the use of the enzyme reduces energy consumption during frequent production stops and reduces solvent use, thereby saving energy and chemicals⁽⁹⁻¹¹⁾. Therefore, the biocatalyst is thought to be a green and economically viable option for future organic synthesis^(6, 12). However, there are still many limitations to biocatalysts. For example, the high regio- and stereo-selectivity normally means that substrates are specifically positioned at the active site of the enzyme, which also may indicate a narrow substrate range. In addition, the biocatalysts are biologically active macromolecules, so the activity of enzymes may be sensitive to various factors such as temperature, pH and concentration of the substrates or products.

1.1.2 Metalloproteins and hemoproteins

Metalloprotein is a term for a protein that contains a metal ion cofactor, which accounts for nearly half of all proteins in nature⁽¹³⁾. A large class of the metalloproteins is the hemoproteins that contain a heme prosthetic group. In the heme enzyme field, cytochrome P450 was the most widely studied, which led to the discovery of around 21000 different P450s and growing with no end⁽¹⁴⁾. Since the discovery of P450 in the 1960s, scientists had learned that this pigment protein belonged to hemoprotein⁽¹⁵⁾. The heme group is essential to functions of cytochrome P450s, and the heme group is constituted of an iron (III) protoporphyrin-IX that is covalently linked to the P450 by the sulfur atom of a

proximal cysteine ligand (Figure 1.1)⁽¹⁶⁾. In the resting ferric state, it has been discovered that the heme cofactor holds an equilibration between a five-coordinate high spin state and the six-coordinate low spin state (Figure 1.1)⁽¹⁷⁾. When the catalytic cycle of heme is initiated in cytochrome P450, the H₂O at the axial level is displaced by O₂ following the reduction of the ferric iron. The details of the catalytic cycle of heme B will be discussed in section 1.2.3. The thiolate coordination to iron constructs the typical active cytochrome P450, and the transition from thiolate to thiol leads to the spectral change from P450 to biologically inactive P420⁽¹⁸⁻²⁰⁾.

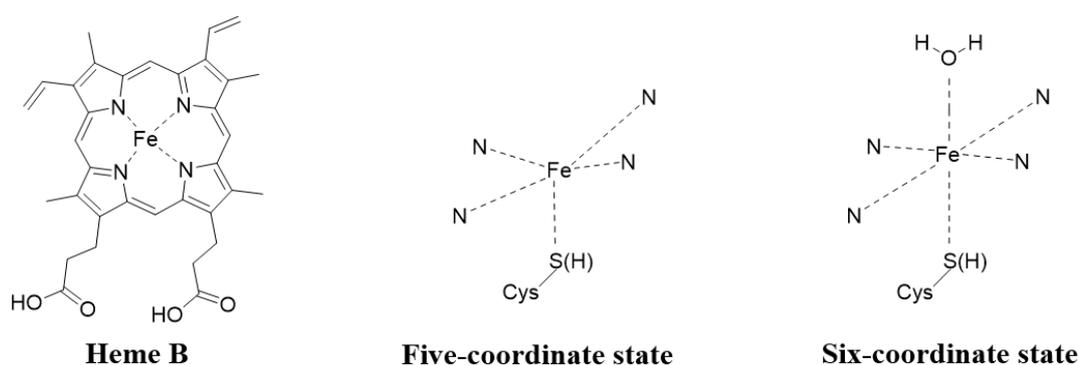
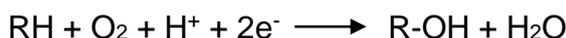


Figure 1. 1. The structure of heme B and the electronic configuration of the heme iron in cytochrome P450. The heme cofactor bound to cytochrome P450 is heme B, and the electronic configuration around the resting state ferric iron is shown above. At the resting state of the heme, there is the equilibration between five-coordinate (high-spin) and six-coordinate state (low-spin). The heme structure was directly imported from the ChemDraw database, and the state structures were drawn by the author.

1.2 Cytochrome P450 Monooxygenases

Cytochrome P450 enzymes (CYPs or P450s) are widely distributed monooxygenases, which were first identified almost 60 years ago⁽²¹⁾. These enzymes are known for their versatility of catalysing the oxidation reactions against a broad spectrum of chemical compounds⁽²²⁾. The oxidation reactions are normally indicated as hydroxylation, epoxidation, dealkylation, and some other transformations⁽²³⁾. The most common reaction performed by cytochrome P450s involves the reductive activation of dioxygen binding to the heme group and then insertion of an atom of molecular oxygen into substrates (Scheme 1)⁽²²⁾.

Performing oxygen insertion at selective non-activated C-H bonds might be one of the most difficult reactions in synthetic chemistry⁽²⁴⁾. Meanwhile, cytochrome P450 as an enzyme can carry out reactions at moderate conditions like room temperature and normal pressure. Therefore, the P450s are regarded as powerful catalysts and have attracted attention from different fields since their discovery⁽²⁵⁾.



Scheme 1. The overall oxidation reaction catalysed by general P450s⁽²²⁾. The reaction scheme requires two reducing equivalents usually derived from a nicotinamide cofactor.

Cytochrome P450s are known for their powerful capability of C-H bond activation, as well as supporting other types of reactions such as heteroatom release (dealkylation)⁽²⁶⁻²⁸⁾, epoxidation^(29, 30), and other complex reactions as reviewed elsewhere⁽³¹⁻³⁵⁾. This makes cytochrome P450 an attractive target to be engineered and developed for biotechnological purposes. Some of the classic examples shown in Figure 1.2 indicate the involvement of cytochrome P450 in many natural bioprocesses and biotechnological applications.

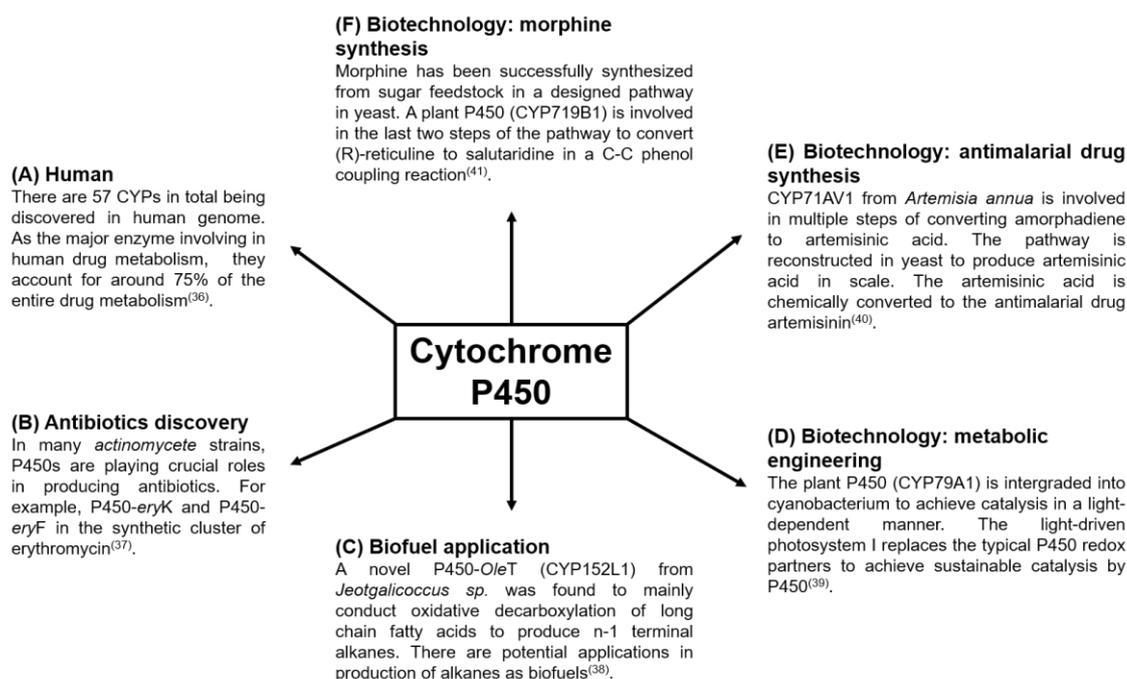


Figure 1. 2. The landscape of bioprocess involving cytochrome P450 enzymes and some well-known examples. As a widely spread class of enzymes,

cytochrome P450s play a crucial role as versatile biocatalysts in various bioprocesses. There are too many examples of the involvement of cytochrome P450 in many natural or biotechnological bioprocesses. Only six examples are given in the figure to show how cytochrome P450s are studied or re-engineered to serve various purposes. (A) 57 CYPs are found in humans, which serves as a major enzyme in drug metabolism and involves around 75% metabolism of total known drugs⁽³⁶⁾. (B) In the CYP-rich actinomycete strains, it is common to find P450s in the synthetic pathways of antibiotics. For example, the P450-*eryK* and P450-*eryF* for their natural involvement in erythromycin cluster⁽³⁷⁾. (C) A novel P450-*OleT* (CYP152L1) from *Jeotgalicoccus* sp. was found to mainly conduct oxidative decarboxylation of long-chain fatty acids to produce n-1 terminal alkanes. There are potential applications in the production of alkanes as biofuels⁽³⁸⁾. (D) Metabolic engineering of a cyanobacterium to accommodate a plant P450s to achieve electron transfer by photosystem rather than redox partners⁽³⁹⁾. (E) CYP71AV1 from *Artemisia annua* is involved in multiple steps of converting amorphaadiene to artemisinic acid. The pathway is reconstructed in yeast to produce artemisinic acid in scale. The artemisinic acid is chemically converted to the antimalarial drug artemisinin⁽⁴⁰⁾. (F) A plant P450 (CYP719B1) is involved in the last two steps of the pathway to convert (R)-reticuline to salutaridine in a C-C phenol coupling reaction⁽⁴¹⁾.

1.2.1 Physiological roles of cytochrome P450

Cytochrome P450 is widely discovered in all kingdom of life⁽²²⁾. However, the number of P450 genes varies considerably in each organism. For example, *Saccharomyces cerevisiae* (baker's yeast) has 3⁽⁴²⁾, *Drosophila melanogaster* has 83⁽⁴³⁾, *Arabidopsis thaliana* has 286⁽⁴⁴⁾, humans have 57⁽⁴⁵⁾ and also the P450-rich *Streptomyces* species as many as 18 in *S.coelicolor* and 33 in *S. avermitilis*⁽⁴⁶⁾.

The bacterial cytochrome P450s are normally soluble cytoplasmic proteins, implicated in functions like fatty acid hydroxylation, xenobiotic metabolism, antibiotic synthesis, and biotransformation⁽⁴²⁾. In practical uses, bacterial P450s are involved in the redesign of pharmaceuticals at an industrial scale, such as P450BM3 is used in the production of high-value drug metabolites⁽⁴⁷⁾, and the investigation of CYP152L1 (P450-*OleT*) in production of terminal alkanes as alternative biofuels⁽⁴⁸⁾. The research interest in this thesis lies in cytochrome P450 from actinomycetes. In actinomycetes, many species contain one or more

soluble P450s that are responsible for a variety of xenobiotic transformations (49-51).

Fungal P450s have greater divergence than animal, plant, or bacterial P450s, and have very diverse functions to meet metabolic needs in fungi⁽⁵²⁾. Fungi P450s are involved in some of the most valuable bioprocesses, e.g. in the biosynthesis of antibiotics, macrolides, and precursors of high-value pharmaceutical compounds⁽⁵²⁾. The fungal CYP51 (14- α sterol demethylase) is involved in the ergosterol biosynthesis required for membrane integrity, and the azole antibiotics are developed for inhibition of CYP51 to result in the accumulation of toxic sterol and then cell death, serving as an antifungal approach⁽⁵³⁾.

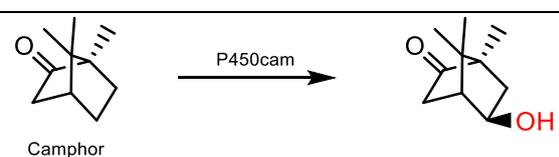
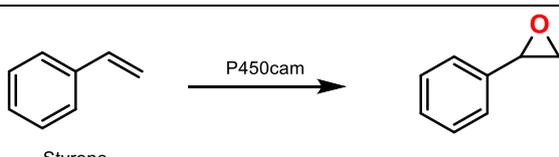
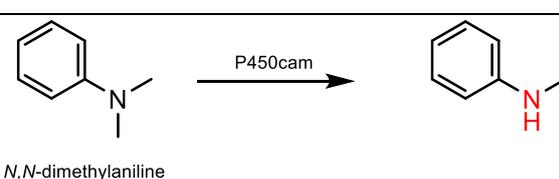
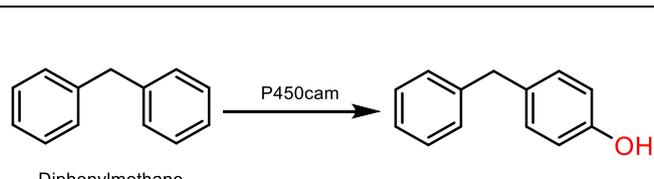
Plant P450s catalyse the oxygenation of fatty acids, catabolism of phytohormones, and synthesis of flower pigments or defence chemicals (including aromas and flavours, antioxidants and anti-cancer drugs)⁽⁵⁴⁾. The P450 CYP71AV1 from *Artemisia annua* is involved in an oxidation cascade of 4,11-amorphadiene to artemisinic acid via alcohol and aldehyde intermediates, which leads to the production of efficient antimalarial drug artemisinin⁽⁵⁵⁾. Crop P450s are targeted for genetic modifications to enhance crop yields and to improve herbicide resistance, as an increasing number of P450-dependent reactions was found in herbicide metabolism⁽⁵⁶⁾.

The human P450s are well-known as detoxifying enzymes found predominantly in liver microsomes⁽⁴⁵⁾. Both cytochrome P450 and its redox partners are membrane-bound via an N-terminal transmembrane domain, which is a feature that renders them highly difficult to isolate⁽⁵⁷⁾. Human P450s were long targeted for their ability to produce drug metabolites. Out of 18 human P450 families, the CYP1, 2, 3, and 4 families are the most abundant families and constitute the majority of the 57 CYP genes in humans, while the other 14 families have only 1 member each with very specific functions essential for life⁽⁵⁸⁾. For example, the CYP2D6 is involved in the metabolism of a quarter of all prescribed drugs in the human liver⁽⁵⁹⁾. Over 90% of human drug oxidation can be attributed to the following CYPs: 1A2 (4%), 2A6 (2%), 2C9 (10%), 2C19 (2%), 2E1 (2%), 2D6 (30%), and 3A4 (50%), which may change when more substrates for various CYPs are identified⁽⁶⁰⁾.

1.2.2 Typical reactions catalysed by P450

Cytochrome P450s constitute a large family of enzymes that are capable of diverse oxygenation catalysis in nature. In addition, enzyme engineering continues to explore the limits of CYPs to perform different reactions. As is shown in Table 1.2, the common reactions conducted by P450cam are listed as examples. The hydroxylation of camphor by P450cam is a well-known example firstly discovered in 1965⁽⁶¹⁾. The P450cam attacks the inactive C-H bond to give 5-exo-hydroxycamphor as the single product⁽⁶²⁾. The hydroxylation could also happen on C-H bonds on aromatic rings, for example, the hydroxylation of diphenylmethane by P450cam (Y96F)⁽⁶³⁾. Other typical reactions are *N*-dealkylation, including the removal of one methyl group from *N,N*-dimethylaniline by P450cam (Y96A)⁽⁶⁴⁾.

Table 1.2. Summary of common reactions catalysed by P450cam (CYP101A1). The example reactions catalysed by P450cam, and the altered groups on substrates are labelled in red.

Reactions	CYPs and example reactions
Hydroxylation ⁽⁶¹⁾	 <p>Camphor</p> <p>The reaction shows camphor (a bicyclic ketone) reacting with P450cam to form 5-exo-hydroxycamphor. The hydroxyl group (-OH) is highlighted in red in the product.</p>
Epoxidation ⁽⁶³⁾	 <p>Styrene</p> <p>The reaction shows styrene (vinylbenzene) reacting with P450cam to form styrene oxide. The epoxide oxygen atom is highlighted in red in the product.</p>
<i>N</i> -dealkylation ⁽⁶⁵⁾	 <p><i>N,N</i>-dimethylaniline</p> <p>The reaction shows <i>N,N</i>-dimethylaniline reacting with P450cam to form <i>N</i>-methylaniline. One methyl group is removed, and the remaining N-H bond is highlighted in red in the product.</p>
Aromatic hydroxylation ⁽⁶⁴⁾	 <p>Diphenylmethane</p> <p>The reaction shows diphenylmethane reacting with P450cam to form 4-hydroxydiphenylmethane. The hydroxyl group (-OH) is highlighted in red in the product.</p>

For characterisation of unknown cytochrome P450s, it is not possible to predict what reactions they could perform simply from the amino acid sequence. Therefore, the identification of products is the key to confirm what exact reactions they could perform. This principle is applied to all characterisations of P450 described in this thesis.

1.2.3 The P450 catalytic cycle

The proposed catalytic cycle in Figure 1.3 reveals the key heme intermediates appeared in the process of hydroxylation by cytochrome P450. The cycle of heme intermediates was firstly proposed in 1968 with only 4 heme intermediates being discussed^(61, 66, 67). As research went on, further understanding of the heme cycles leads to current and generally accepted form⁽⁶⁸⁾. The key step of the cycle is the activation of molecular oxygen to produce a high-valent reactive species (intermediate 7 in Figure 1.3), which was first observed and postulated by Groves in 1978^(69, 70). After that, another signature step involves the radical substrate attack intermediate 7 to form the hydroxylated product and restoring to the ferric state of heme. A total of eight heme intermediates are generally recognised in current studies of the P450 reaction mechanism.

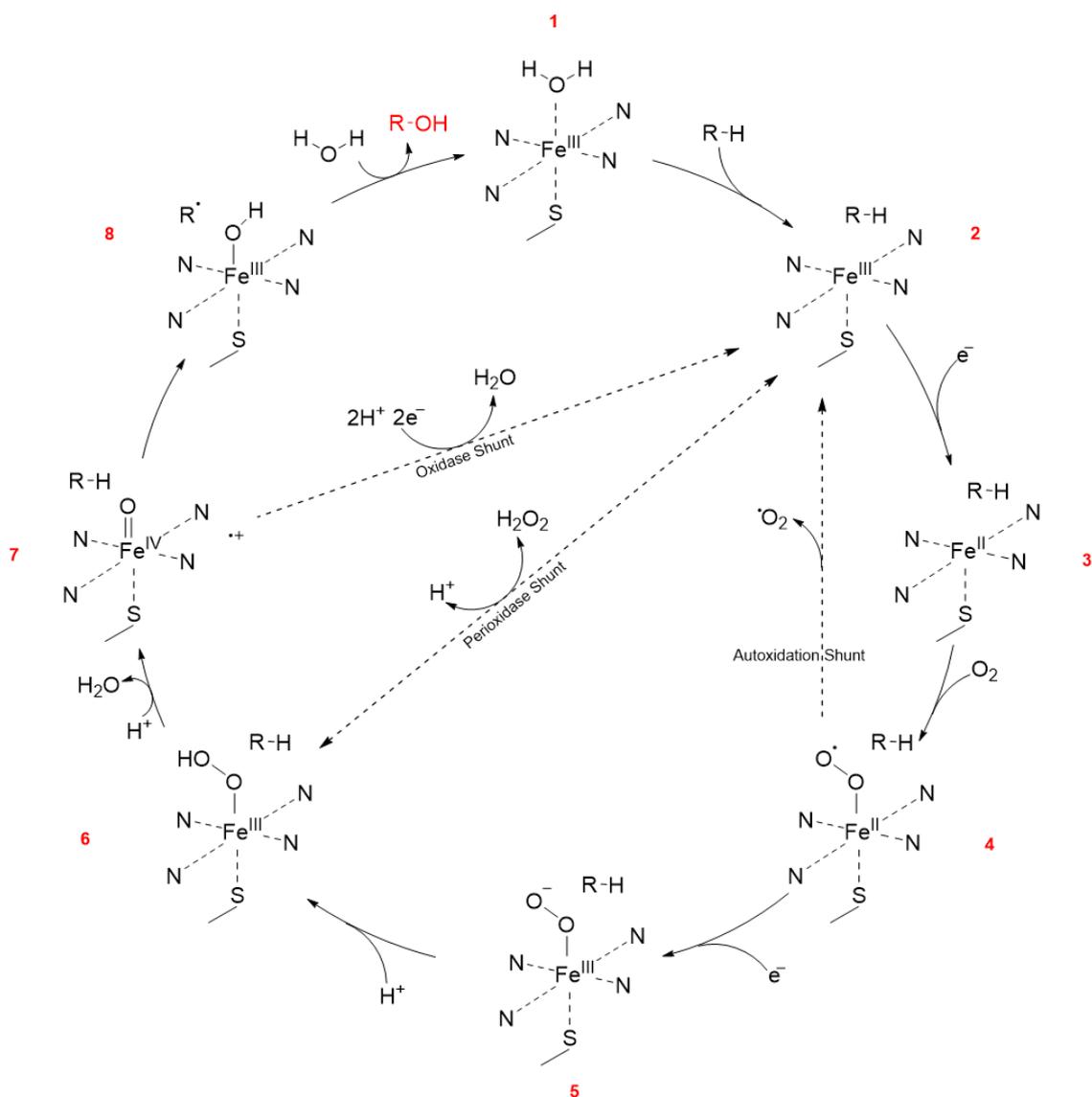


Figure 1.3. The cytochrome P450 catalytic cycle. The proposed cycle and heme intermediates are shown and explained in the text. Each intermediate compound is numbered. The axial coordination of thiolate (S) and heme (Fe) is shown, while the iron cation is also horizontally surrounded by nitrogen from four porphyrins. The substrate is shown as R-H, and the product is shown as R-OH coloured in red. The key reactive intermediate is compound 7. The figure is modified and redraw from Denisov *et al*⁽⁶⁸⁾.

Intermediates 1 and 2. The resting state of heme contains a ferric cation (Intermediate 1 in Figure 1.3), and the water molecule (6th axial ligand) is placed above the heme plane. The six-coordinate iron is defined as a low-spin state. The first step of the catalytic cycle involves the placement of the water molecule with substrate R-H on the distal side of heme, generating the intermediate 2 that contains a high-spin five-coordinate iron^(71, 72). The high-spin iron has a more

positive reduction potential⁽¹⁷⁾, which favors the next step of the reduction of ferric iron.

Intermediate 3. One electron is delivered by the external redox to reduce to ferrous state of heme^(73, 74).

Intermediate 4, and 5. The ferrous iron binds to the dioxygen molecule to form the oxy-enzyme complex (intermediate 4) that is shown as the ferric superoxide form. The oxy-enzyme complex is unstable and could collapse back to ferric heme, which is named as the autoxidation shunt. After that, the complex is reduced by a second electron to generate the peroxo intermediate 5⁽⁷⁵⁾.

Intermediate 6 and 7. The protonation of peroxo intermediate 5 generates the hydroperoxy intermediate 6. The second protonation of intermediate 6 causes the split of O-O bond and release of water, which leads to the key reactive intermediate 7. Intermediate 7 is a Fe⁴⁺ porphyrin cation radical species, which is thought to be the main oxygenating species. The intermediate 6 and 7 are all involved in the shunt reaction to short-circuit the catalytic cycle. These shunt reactions are thought to be the reason for uncoupled electron consumption from substrate oxidation. The peroxidase shunt works in both directions on, allowing the generation of reactive iron-oxo species (intermediate 6) directly from the resting ferric heme. This may offer an alternative approach to developing cytochrome P450 as catalysts. The use of hydroperoxides can bypass the steps of acquiring electrons from nicotinamide cofactors. In previous research, the enhanced activity of using hydroperoxides was confirmed on indole hydroxylation by P450BM3⁽⁷⁶⁾.

Intermediate 8. The highly reactive complex 7 attacks the C-H bond of the substrate by an abstraction of the hydrogen, which generates the intermediate 8 and a radical remaining substrate species. The radical substrate in turn attacks the hydroxyl group on intermediate 8, which restores the heme iron to its resting ferric state and releases the hydroxylated product. This mechanism is called radical rebound.

After the review of the steps of the catalytic cycle, there are several different steps such as oxygen binding, electron delivery, and protonation efficiency could

influence the efficiency of the running cycle, but it is generally accepted that the major rate-limiting steps are the first and second electron donation steps⁽⁷⁷⁾. The electron transfer to heme center is necessary for most cytochrome P450 conducted reactions. The electron transfer is performed by redox partners such as flavoproteins, ferredoxins, or flavodoxins to directly contact P450. The differences among different redox partners will be discussed in section 1.3. The placement of the heme center and redox cofactors in proximity allows the electron tunneling⁽⁷⁸⁾. Therefore, the interactions between redox partners and cytochrome P450s cause the electron transfer, which acts as the major rate-limiting step.

1.2.4 Nomenclature and classification of cytochrome P450s

The name of P450 comes from the unique spectral property of the enzyme. P stands for pigment, and 450 means the absorption peak appears at 450 nm when the reduced enzyme binds to carbon monoxide⁽¹⁵⁾. The nomenclature of P450 is defined as CYP-number-letter-number. For example, CYP1A2 is a common name of P450. CYP represents cytochrome P450 protein, and 1 means the family of the organism that produces the P450, and A indicates the sub-family, and the final 2 refers to the discovered individual P450 in the subfamily. The old family numbering system was meant to cover less than 100 P450 families in the 1980s⁽⁷⁹⁾. However, with the rapid discovering of new P450s, 100 family numbers were not enough. Therefore, the new numbering system is established and recognized to cover all members in the cytochrome P450 kingdom. In principle, the CYPs from mammals and insects belong to families 1-49 or 301-499, while families numbered between 51-69 or 501-699 are lower eukaryotes. All plant P450s are classified into families 71-99 or 701-999. All bacteria P450s are classified into families between 101-299 or 1001-2999⁽⁸⁰⁾. The subfamily is represented by single letters of the alphabet. O and I are not used to avoid confusion with zero and one. For some P450 families like CYP105, the single alphabet is not enough to cover all subfamilies, so double alphabets are also applied such as CYP105AB and CYP105AC. Normally, two P450s that have more than 40% amino acid identity are classified into the same family, and into the same subfamily when having more than 55% identity⁽⁸⁰⁾.

P450 is a heme-containing protein, which works with corresponding redox partner proteins to use electrons from cofactors like nicotinamide adenine dinucleotide hydrate (NADH) or nicotinamide adenine dinucleotide phosphate hydrate (NADPH)⁽⁸¹⁾. According to the catalytic cycle described in section 1.2.2, there are two electrons required in a stepwise manner, which are all delivered by redox partner proteins. Based on different electron supply system found for P450s, a recognised classification is shown in Figure 1.4.

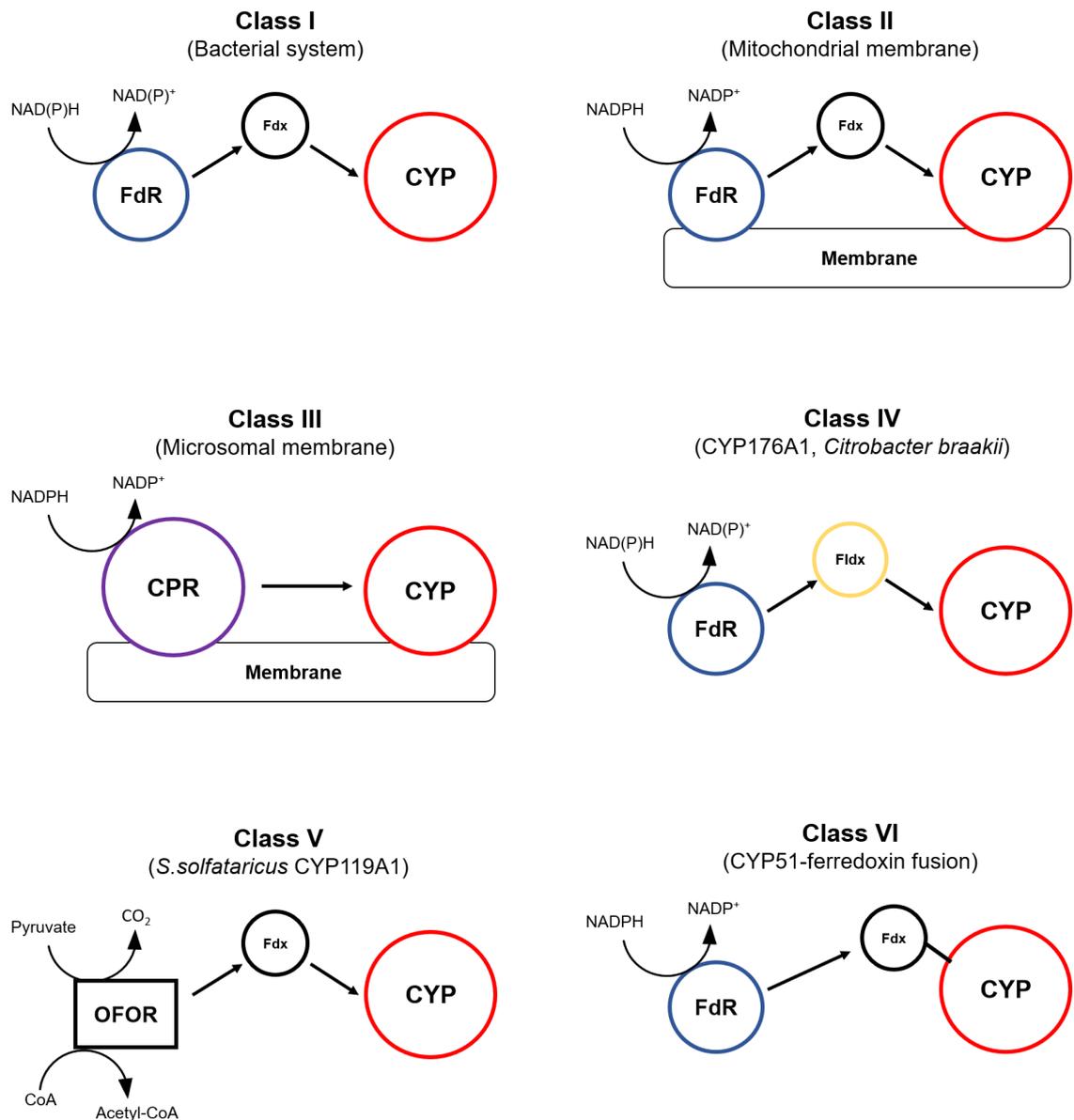




Figure 1.4. Selected classes of P450s based on their electron transfer partners. The scheme shows various P450 redox partner systems and their classifications. Fdx, ferredoxin (iron-sulfur-cluster); Fldx, flavodoxin; FdR, ferredoxin reductase; CPR, cytochrome P450 reductase; FMN, Flavin mononucleotide, a prosthetic group conjugated with enzymes to function as a redox cofactor. FAD, Flavin adenine dinucleotide, also serves as a redox centre in FdR. Class I P450 contains three separate components that are mainly found in bacteria; Class II P450 has three components, and both FdR and CYP are anchored on the membrane; Class III P450 is normally seen in eukaryotes having CPR and CYP anchored on the membrane. Class IV P450 has the similar component arrangement with Class I P450 but uses Fldx as the transfer shuttle⁽⁸²⁾; Class V P450 also has three-component system including a pyruvate-dependent 2-oxo-acid-ferredoxin-oxidoreductase (OFOR), using pyruvate rather conventional nicotinamide cofactors⁽⁸³⁾; Class VI P450 has the ferredoxin group fused to the P450 enzyme⁽⁸⁴⁾. Class VII P450; The reductase and P450 are fused as one enzyme^(85, 86); Class VIII P450. Three components are fused, and the fused part of ferredoxin and reductase is identified as the phthalate-family oxygenase reductase⁽⁸⁷⁾. All figures are modified and redrawn from a previous review of P450 classifications⁽⁸¹⁾.

Class I P450 mainly occurs in bacteria, with a three-component electron transferring system (Figure 1.4, Class I). In prokaryotes, all three components are soluble in the cytoplasm, and the electrons flow from NAD(P)H to ferredoxin reductase to ferredoxin and finally to the heme centre of P450. In Class II P450s, the electron transferring pattern is the same as that described in Class I, which in most cases relies on the soluble ferredoxin as the shuttle to accept electrons from the FAD centre of reductase and then transfer to heme centre of P450 for the oxidation reaction. Within ferredoxin protein, the major part for electron-accepting and donating is the iron-sulfur cluster, and the cluster type could be [2Fe-2S], [3Fe-4S] or [4Fe-4S]⁽⁸¹⁾. One of the representative P450s in class I is P450cam (CYP101A1) and its natural redox partners putidaredoxin and

putidaredoxin reductase from *Pseudomonas putida*⁽⁸⁸⁾. Before transferring electrons to [2Fe-2S] cluster in putidaredoxin, the putidaredoxin reductase gains electrons from NADH through its FAD redox centre⁽⁸⁸⁾.

Class III P450 is also called microsomal P450 systems, consisting of an NADPH dependent cytochrome P450 reductase and a P450 oxygenase⁽⁸⁹⁾ (Figure 1.4, Class III). The Class III P450s are normally found in the eukaryotes with two components anchored on the endoplasmic reticulum (ER)⁽⁹⁰⁾. The cytochrome reductase (CPR) contains both FAD and FMN as the redox factors, which resembles the redox centres of ferredoxin reductase and bacterial flavodoxin. In addition to the large number of eukaryotic Class III P450, only one Class III P450 from bacteria has been reported (CYP105A3 from *Streptomyces carbophilus*)⁽⁹¹⁾.

Class IV P450s are also found in prokaryotes⁽⁸²⁾. Class I and Class IV P450s are sharing a similar three-component electron transfer system, but Class IV P450 uses flavodoxin as the electron transfer mediator. The redox centre of flavodoxin is flavin mononucleotide (FMN) instead of a conventional iron-sulphur cluster in the ferredoxin. In fact, the Class IV P450 is quite similar to the Class III eukaryotic P450s as far as the redox centres are only concerned.

Class VI P450 has a unique structure compared to the standard three-component system. The C-terminal of P450 oxygenase in the class is fused to a ferredoxin or a flavodoxin domain⁽⁸⁴⁾. There are not many examples in this class, so the mechanism of electron transfer and the usage of electron cofactors remains to be determined.

Class VII P450 system contains only one component, with the reductase domain and P450 oxygenase domain fused into one peptide chain. The unique structure leads the enzyme to be self-sufficient in electron transfer, which increases the efficiency of passing electrons from cytochrome reductase to P450 oxygenase. The widely studied example in this class is CYP102A1 (P450-BM3 from *Bacillus megaterium*), which was first isolated and identified by Miura and Fulco in 1974⁽⁸⁵⁾. The P450 oxygenase domain is connected through a short chain to cytochrome reductase domain containing both FAD and FMN molecules as redox factors⁽⁸⁵⁾. The diflavin reductase domain of Class VII P450 is highly similar to the reductase partner of Class III eukaryotic P450.

Class VIII P450 is also a fused protein. Compared to Class VII, Class VIII has different redox centres attached to the P450 oxygenase. The reductase contains an FMN binding domain, which is different from the FAD-binding domain in ferredoxin reductase in other classes of P450s. The firstly identified enzyme in Class VIII is CYP116B2 (P450-RhF from *Rhodococcus sp.*)⁽⁹²⁾.

1.2.5 Structure and function of cytochrome P450

The first structural data of cytochrome P450 was resolved in 1987, presenting the structure of Class I cytochrome P450 CYP101A1 (P450cam) from *P.putida*⁽⁶²⁾. Since its breakthrough, the structure of P450cam was used as one of the model structures for Class I P450s, and there were more structures of P450s being experimentally revealed in the coming years⁽⁶⁸⁾. It seems most P450s retain a unique tertiary structure fold, which consists of 13 α -helices (labelled from A to L), and a β -sheet rich region, with several loop structures between the secondary structural elements (Figure 1.5). The P450cam structure is discussed and reviewed here as the representative structure of Class I P450, and the understanding of the structural features of P450cam facilitates the exploitation of other structurally unknown Class I P450s studied in this thesis.

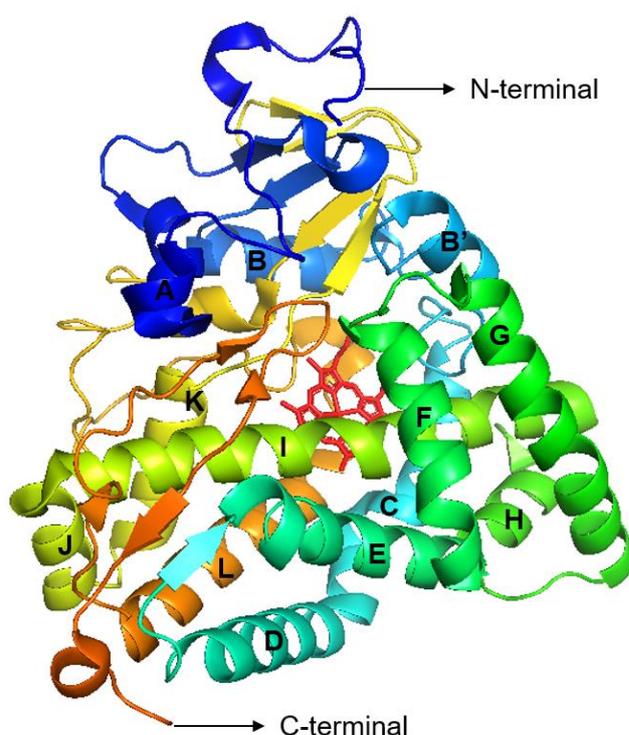


Figure 1.5. P450cam structure with all 13 helices labeled (PDB 2ZWT)^(62, 93).

The structure data was stored in PDB and was visualized using PyMOL (version 2.2.3). The heme group is coloured in red and deeply buried within the structure. The 13 helices are labeled from A to L. The cysteine group coordinating below heme group plane is located on the L-helix, and the region above the heme plane is where the oxidation reaction happens.

The structure of P450cam has several typical features that align with other structurally identified P450s. For example, the heme group is deeply buried within the structure. Three signature motifs could also be seen on the structure of P450cam, including the heme-binding motif on L-helix, the EXXR motif on K-helix, and the GXXTT on I-helix and. The heme-binding motif could be represented as FXXGXXXCXG, and the four residues F (phenylalanine), G (glycine), C (cysteine), and again G (glycine) are invariant in all P450s. The thiolate group of cysteine is responsible for axial coordination with the ferric iron of the heme ligand. The EXXR motif is located quite far away from the heme ligand. For example, the EXXR motif of P450cam is located right after the J-helix and at the N-terminal of the K-helix. The functions of the EXXR motif on the structural level include locking the heme at the heme-binding region and stabilizing the core structure. The GXXTT motif is generally located in the middle of the I-helix, which constitutes the oxygen binding site. As is known in the catalytic cycle of P450, the oxygen capturing and conversion of the ferric heme ligand to the reactive ferryl (Fe^{4+})-oxo species is the key to proceed oxidation on substrates. Therefore, it is reasonable GXXTT motif is a conserved feature for P450s. The N-terminal region of P450cam has a more β -sheet structure, while the C-terminal region is mainly α -helical. The BC and FG loops constitute most of the substrate recognition regions, which are highly involved in substrate recognition and binding. Different from those conserved motifs, these substrate recognition regions are tended to be disordered in other P450s. Therefore, the BC and FG loops are often targeted for directed mutagenesis to achieve altered substrate selectivity.

The structure of the P450 active region, including the conformation of heme ligand as well as the arrangement of the side chains surrounding the heme, is the key to understand substrate selectivity and stereospecificity. The active region is not fixed but dynamic upon substrate binding. For example, what can be seen in

P450cam (Figure 1.6), the conformational changes of several helices were observed during the substrate binding. It is accepted that the different conformations of F, G, and B' helices could lead to the structurally distinctive substrate-binding channels, which means diversity in size, shape, and specificity of the binding cavity. In previous research⁽⁹⁴⁾, the open conformation was created by the retraction of F and G helices, and the disordering of the B'-helix, which leads to a substrate channel allowing camphor substrate to access the deeply buried heme ligand.

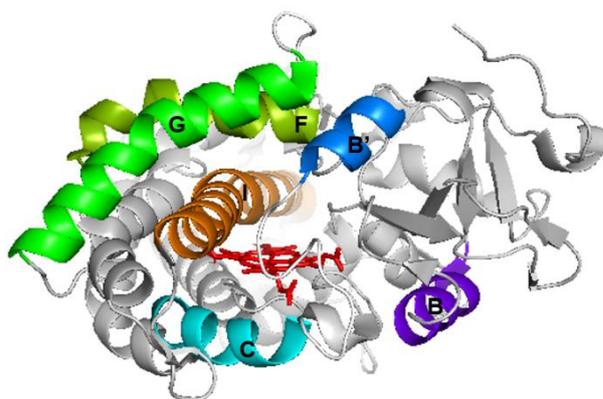


Figure 1.6. Dynamic regions of the P450cam model involved in substrate binding (PDB 2ZWT)⁽⁹³⁾. The secondary structural segments responsible for open/closed conformations that occur during substrate binding in P450cam are represented on the P450cam model (PDB 2ZWT). The I-helix (brown) is in front of the heme cofactor (red). The C-helix (cyan) and the B-helix (purple) are at the back of the heme plane. The B'-helix is located within the B-C loop, which is placed above the heme plane. The F (yellow) and G (green) helices with the FG loop trigger channel opening for substrate access to the P450cam active site⁽⁹⁴⁾.

The active site of P450cam is highly adapted to accommodate the natural substrate camphor and the correct orientation of camphor for stereospecific hydroxylation at 5'-position carbon. The carbonyl oxygen of camphor is surrounded by residues F87, L244, and most notably Y96, with which it forms a hydrogen bond (Figure 1.7)⁽⁶²⁾. The Y96F substitution removes this hydrogen bond interaction⁽⁸⁸⁾, which resulted in the decrease of binding affinity of the molecule and the production of a mixture of hydroxylation products.

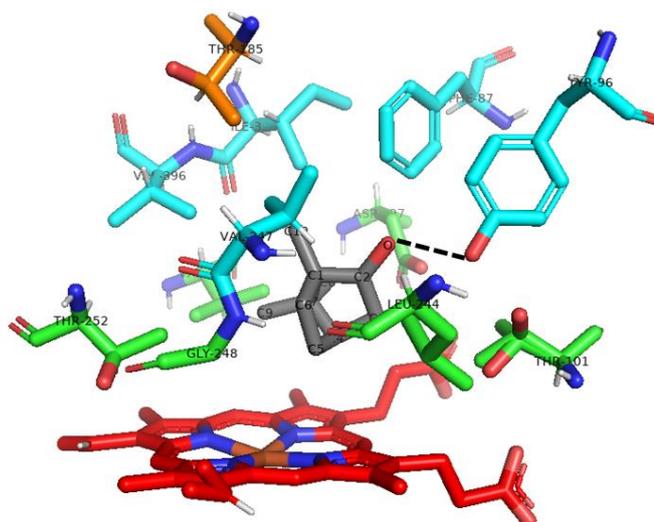


Figure 1.7. Active sites of the P450cam model involved in substrate binding and positioning (PDB 2ZWT)⁽⁹³⁾. The heme ligand is coloured in red, and the camphor molecule is in grey. The active residues are grouped into three layers. The first six residues (green), including T101, L244, G248, T252, V295, and D297, are placed over the heme plane surrounding the camphor. The second layer (cyan) contains F87, Y96, V247, I395, and V396, forming a ring above the camphor substrate. The lone residue T185 is the third layer at the top of the active site. On the first layer, the residues L244, G248, and T252 are located on the I helix. On the second layer, the F87 and Y96 are located on the B' helix, while the V247 is located on the I helix as well. The top layer of lone residue T185 is located between the F and G helix. The analysis of residues constituting the active site is consistent with the helix arrangement analysis shown in Figure 1.6. It is suspected that the residues of the three layers construct a fixed conformation to allow C5 of camphor to position towards the ferric centre.

As is summarized above, the structure features of P450cam are tightly related to the activities that P450cam can perform. The dynamic conformations of the FG loop and B-B'-C loop are the key to create a substrate channel that leads camphor to the deeply buried heme ligand. The residues at active sites play important roles in positioning camphor for stereospecific hydroxylation.

1.3 Redox partners

1.3.1 Flavoproteins and electron transfer types

Flavoproteins are found in all organisms supporting the electron dependent enzymes, which is commonly involved in various physiologically important reactions such as photosynthesis, mitochondrial respiratory chain, and fatty acid metabolism⁽⁹⁵⁾. Flavoproteins play the role of mediating the electron transfer chains between the NAD(P)H cofactor and various redox centres. The redox centers of Flavoproteins normally refer to the flavin cofactors FAD (e.g. NAD(P)⁺ reductase) or FMN (e.g. flavodoxin)⁽⁹⁶⁾. The ferredoxins are assembled with the iron-sulfur cluster as their redox centre and normally serve as the electron shuttle between the redox centre of flavoprotein and the heme centre of cytochrome P450⁽⁹⁷⁾.

The widely recognised electron flow patterns for P450 are summarized below. The direction of each electron is indicated below following the arrows.

Type I: NAD(P)H → FAD → [Fe-S]₂ → Heme

Type II: NAD(P)H → FAD → FMN → Heme

Type III: NAD(P)H → FMN → [Fe-S]₂ → Heme

Type I is commonly seen in bacteria and mitochondria, while Type II is found in both prokaryotes and eukaryotes. Type I system contains three components. P450cam is a model system in this class and consists of the NADH-dependent putidaredoxin reductase (PdR) that is reduced by 2 electrons from the dinucleotide cofactor, and the 1 electron is passed onto [Fe-S]₂ of putidaredoxin (Pdx) at a time⁽⁹⁸⁾. Type II eukaryotic system has a diflavin fusion reductase that fulfills functions of binding NAD(P)H, transferring electrons from FAD to FMN, reducing of the heme by FMN⁽⁹⁹⁾. The Type II prokaryotic system mainly refers to the self-sufficient P450s (e.g. P450BM3). Type III is only shown in the one group of fusion P450s (P450-Rhf)⁽⁸⁷⁾. Apart from the three electron transfer types, there is also one interesting exception to the electron transfer pattern in nature, which is the nitric oxide reductase (P450nor), where the NAD(P)H cofactor directly interact with the heme centre of P450⁽¹⁰⁰⁾.

There is also growing evidence that heterologous redox partners could also support the activity of cytochrome P450^(101, 102). There is one basic principle of selecting heterologous redox partners to couple CYPs for potential reactions. The complete electron flow pattern of the reconstituted system should be consistent with any of the three naturally existed types described above. Otherwise, it is very difficult to know whether the combination of P450 and redox partners is naturally favourable.

1.3.2 Colocation of P450 and redox partners

It has been noticed that some microbial Class I P450s are located adjacent to their natural redox partners on the genome and have evolved structural compatibility⁽¹⁰³⁾. For example, P450cam interacts with its physiological partner putidaredoxin (Pdx), and putidaredoxin reductase (Pdr) is encoded on a plasmid carried by the host *Pseudomonas putida* (Figure 1.8). As is shown in Figure 1.8, the genes of P450cam, Pdr, and Pdx are closely located on the genome, which is coordinately regulated and co-transcribed at the same time for physiological purposes⁽¹⁰⁴⁾.

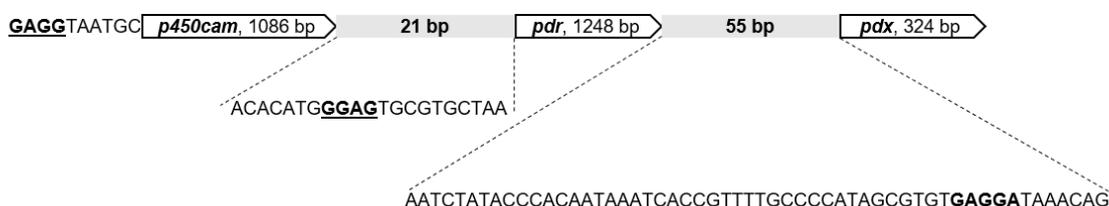


Figure 1.8. Genetic colocation of P450cam and its redox partners⁽¹⁰⁵⁾.

Genebank ID: AB771747.1. The ribosome binding sites are marked as bold and underlined. The three genes *p450cam*, *pdr*, and *pdx* are naturally arranged in order with the length of 1086 bp, 1248 bp, and 324 bp respectively. The two spacing regions among the three genes are 21 bp and 55 bp. There is no restriction site being found in the natural spacing regions.

For Class I P450, it requires two redox partners to finish the electron transfer flow. In most cases, the P450 gene and redox partner genes are independent on the genome, showing no adjacent genes to be co-regulated together. When there is a genetic colocation of P450s and redox partners, it normally means one of the redox partners is located adjacent to the P450 gene. The even rarer case is both redox partners are clustered with the P450 gene on genome, like the genetic

colocation of *p450cam*, *pdr*, and *pdx* described above. Because the P450s from actinomycetes are primarily focused and studied in this thesis, so the examples of genetic colocation of P450 and redox partners in some actinomycetes are summarized (Table 1.2).

Table 1.3. Genetic colocation of P450 and redox partners in some actinomycetes. The genetic colocation of P450 and redox partner is not commonly seen in each of the example genomes. In those cases where this happens, the P450 gene is closely located with the ferredoxin gene. The identification of P450s from *Streptomyces rishiriensis* and its colocation pattern was carried out in this thesis (Chapter 6).

Organism	Total CYPs identified	Colocation patterns discovered	Number of colocation pattern discovered so far
<i>Streptomyces griseolus</i> ⁽¹⁰⁶⁾	Unknown		2
<i>Streptomyces griseus</i> ⁽¹⁰⁷⁾	27		1
<i>Streptomyces coelicolor</i> ^(46, 108)	19		1
<i>Streptomyces avermitilis</i> ^(46, 109)	32		2
			1
			1
<i>Saccharopolyspora erytheraea</i> ⁽¹¹⁰⁾	35		1
<i>Streptomyces rishiriensis</i>	24		1 (This thesis)

1.4 Analytical assays in cytochrome P450 study

1.4.1 Carbon monoxide assay

The cytochrome P450 was first discovered as a pigment protein with the unique spectral property at 450 nm in 1958⁽²¹⁾. In cytochrome P450 studies, the spectral assay was used to determine the integrity of the enzyme as well as the concentration of total cytochrome P450s. The basic principle of the assay is that only the ferrous heme ligand could bind to carbon monoxide to form a complex that shows maximum absorption at 450 nm. The unique spectral property of the ferrous heme-CO complex is due to the axially coordinated thiolate group from

the conserved cysteine residue. An accepted concept is that the loss of peak absorption at 450 nm is generally associated with the loss of enzyme activity⁽¹¹¹⁾. Therefore, the spectral assay is necessary and useful to check the integrity of the enzyme before conducting any further studies.

The general method of conducting carbon monoxide assay involves the reduction of ferric heme to ferrous heme, incubation of ferrous heme with carbon monoxide, and the calculation of difference spectrum by subtracting reduced heme spectrum from CO complex spectrum. The extinction coefficient of the complex on the difference spectrum was determined to be $91 \text{ mM}^{-1}\text{cm}^{-1}$ ⁽¹⁵⁾, which is used for the calculation of molar concentrations of the active P450s.

1.4.2 Cytochrome *c* assay

The biological system of cytochrome P450 requires not only P450 but also its accessory reductases that are necessary for catalytic activity. It is necessary to monitor the electron transfer flow from NAD(P)H to cytochrome P450, which normally reveals the integrity of the electron transfer chain as well as how efficient the redox partners are on delivering electrons⁽¹¹¹⁾. In literature, the most convenient approach of monitoring electron transfer is to couple the CYP reductase to air-stable dyes, e.g. tetrazolium chloride⁽¹¹²⁾, or other redox components, e.g. cytochrome *c*⁽¹¹³⁾. In this thesis, equine heart cytochrome *c* is used, so the assay for monitoring electron transfer activity is also known as cytochrome *c* assay. In the assay, the cytochrome *c* receives electrons from redox partners capable of harvesting electrons from biological hydride electron donor NAD(P)H. The reduced cytochrome *c* is measured at 550 nm, and the concentration of reduced cytochrome *c* is determined with an extinction coefficient of $21 \text{ mM}^{-1}\text{cm}^{-1}$ ⁽¹¹⁴⁾. A negative control is also prepared with only cytochrome *c* and reductases, which monitors the baseline absorbance of non-reduced cytochrome *c* at 550 nm. Once an electron donor is added to the reaction system, the monitor of cytochrome *c* reduction begins. A correlation curve of cytochrome *c* absorbance and time is plotted to determine the concentration of reduced cytochrome *c* and the electron transfer efficiency.

1.5 Perspectives of discovering and engineering P450s

It is noteworthy that CYPs could conduct oxidation reactions under normal conditions. Another unique competence of CYPs is its capability to perform highly stereo/regio-selective reactions on various complex molecules. Furthermore, there is a rich pool of different CYPs identified in all kingdoms of life⁽¹¹⁵⁾. The diversification of CYPs opens novel possibilities for practical applications. However, there are still at least four commonly recognised bottlenecks that hamper the commercial applications of P450^(22, 116, 117), such as low activities, need for electron transfer partners, cofactor requirement, and uncoupling between cofactor oxidation and product formation. Both discovering new P450s and engineering P450s on improving enzyme performance could be explored to develop the capabilities for industrial applications.

1.5.1 Discovering new P450s through genome mining

Although many model P450s such as P450cam, P450BM3, and P450RhF have been identified, characterized, and even extensively engineered, the toolbox of cytochrome P450 is still growing to satisfy the potential needs. Genome mining is a common approach to discovering new enzymes⁽¹¹⁸⁾. The growing efficiency and decreasing cost of sequencing techniques have made the discovery and revealing of the full genome sequence quite simple. After gaining access to the full genome sequence, bioinformatic approaches could be applied to locate all potential P450s in the genome according to their conserved sequential regions. With the well-established classification system described in section 1.2.4, all newly identified P450s could be sorted and classified into their subfamilies for further investigation. The other common approach of discovering P450s is homologues search. It is commonly practised as using known and characterized P450s to compare against genome sequence through BLAST to find sequentially similar targets. However, the functions of homologues P450s are not known until the screening reactions are conducted and confirmed.

Homology searching is commonly used in discovering new P450s, which leads to the large expansion of two example CYP families, CYP102 and CYP152^(119, 120). CYP102 is a self-sufficient monooxygenase, consisting of a heme domain and a diflavin reductase domain, while CYP152 can efficiently use H₂O₂ as an

oxidant. The two families have high potential in biotechnological use because of their special and unique structures and functions⁽¹²¹⁾. Most of the P450s in these two families are discovered by homologues research targeting the functionally characterized CYP102A1 and CYP152A1, respectively⁽¹²¹⁾.

CYP102A1, also well known as P450-BM3 from *Bacillus megaterium*, has the unique structure of fused reductase and monooxygenase, which favours the electron transfer and exhibits extremely high turnover rate against certain substrates using NADPH as the electron source. The natural substrates for P450-BM3 are long-chain fatty acids, and hydroxylation normally happens at sub-terminal positions (ω -1, ω -2, and ω -3)⁽¹²²⁾. Furthermore, P450-BM3 prefers unsaturated long-chain fatty acids over saturated or branched ones. After homologues search of P450-BM3 using a BLAST program at NCBI, scientists found many homologs of P450-BM3 such as CYP102A2, CYP102A3, CYP102A5 and CYP102A7 in the subfamily, CYP102D1, and CYP102J1 from other bacteria genome sequences⁽¹²¹⁾. All these homologs are fused proteins and have similar extremely high turnover rates against certain substrates, but they have slightly different specificity and region-selectivity compared to P450-BM3. For example, CYP102A2 and CYP102A3 from *Bacillus subtilis* have better activity against branched fatty acids⁽¹²¹⁾, and CYP102A5 from *Bacillus cereus* shows higher region-selectivity at ω -1 and ω -2⁽¹²³⁾, and CYP102A7 from *Bacillus licheniformis* indicates more activity against saturated fatty acid and cyclic terpenes⁽¹²⁴⁾.

CYP152A1, which is also known as P450-BS β from *Bacillus subtilis*, uses H₂O₂ to conduct the oxidation reaction rather than relying on traditional NAD(P)H initiated electron transfer⁽¹²⁵⁾. The unique structure and oxidation approach make this P450 full of biotechnological interests. CYP152A1 was discovered by homologues search of a known P450-SP α from *Sphingomonas paucimobilis*, which were proved to have an approximately equal preference on α and β hydroxylation⁽¹²⁵⁾. Scientists also used the sequence of CYP152A1 to BLAST search and find another P450 from *Clostridium acetobutylicum* (CYP152A2)⁽¹²⁶⁾. CYP152A2 has a similar catalytic activity using H₂O₂, but higher preference for α hydroxylation⁽¹²⁶⁾.

1.5.2 Discovering new P450s through engineering

1.5.2.1 Enzyme engineering

The enzyme could be engineered through a rational design or evolutionary approach. Normally, the rational design can be applied when the detailed structural and functional information of the enzyme is available. The understanding of the differences between wild type enzyme structure and designed structure should be generated first, and then directed mutations at defined residues could be proposed rationally⁽²³⁾. Considering not all the P450s have well-studied structures and functions, directed evolution is another straightforward approach. Directed evolution could be conducted through firstly inserting random mutations at positions that are expected to contribute to desired property, and then screening the library of mutants for desired enzyme properties⁽²³⁾. Based on experience, most of the mutations would not make a difference or even be deleterious, but few mutations might be promising⁽¹²⁷⁾.

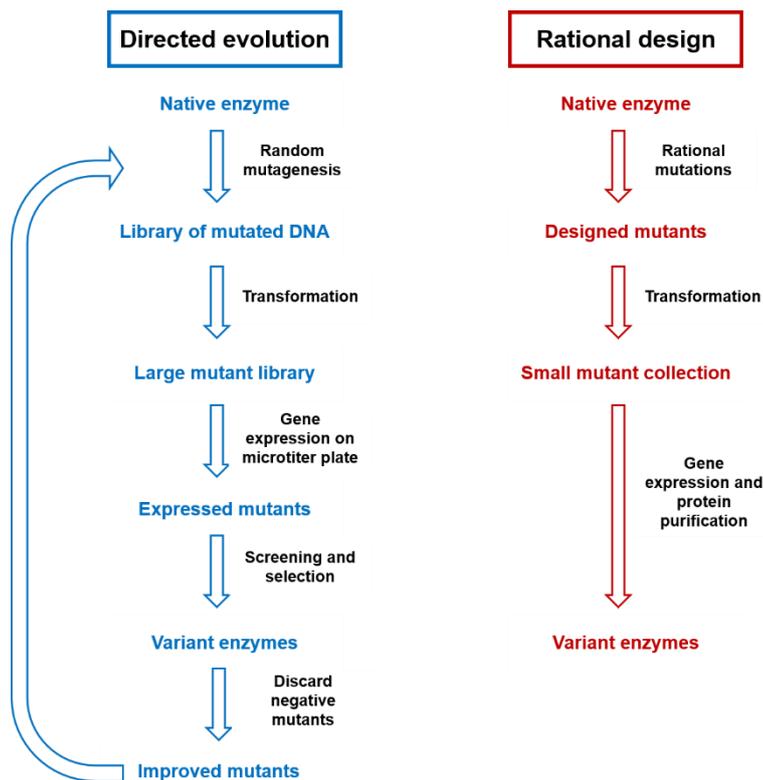


Figure 1.9. In the comparison of the processes involved in directed evolution and rational design⁽¹¹⁷⁾. The rational design is to directly plot out any expected or designed changes based on the understanding of the original structural and functional information. Direct evolution is to build a library of candidates which may carry

the expected features and then screen all the candidates for the selected property. The directed evolution may go through several iterations to find an ideal candidate with desired functions. The figure is redrawn from a previous research paper⁽¹¹⁷⁾.

What limits the utilization of wild-type P450s as biocatalysts include poor expression, being membrane-bound, and lack of known redox partner proteins. Therefore, protein engineering efforts have been put on expanding the substrate range of well-expressed bacterial P450s with known redox partners to accept desired target compounds⁽¹²⁸⁾. Both P450cam and P450BM3 are model P450s that have resolved structures and highly efficient redox partners. Much of these works have been focused on them. Protein engineering of P450cam has led to switch its substrate specificity through the introduction of point mutations to the active site, which allows P450cam to accept unnatural substrates such as pinene, butane, ethylbenzene and others⁽¹²⁹⁻¹³²⁾. This strategy has been particularly successful because the enzyme backbone remains relatively fixed during catalysis⁽⁷⁰⁾. Like P450cam, P450BM3 has been engineered to accept a variety of substrates. Most rational engineering efforts of BM3 targeted hydrophobic residues lining the substrate-binding channel as well as active site residues immediately to the heme centre^(133, 134). For example, the directed evolution of P450BM3 has created a variant F87V/L188Q/A74G with promiscuous activities against indole, alkanes, polycyclic and aromatic hydrocarbons⁽¹³⁵⁻¹³⁷⁾.

In addition to site-directed mutagenesis, another genetic engineering approach is to construct artificial P450s through shuffling the genes of different components of various P450s⁽¹³⁸⁾. Because of the unique component-based structure of P450s (redox partner and the oxygenase), it is viable to replace any of the components (normally redox partner) with heterologous components from to make artificially fused P450s. The gene fusion approach has been frequently and widely practised in the last twenty years⁽¹³⁸⁾. By using fused P450s, the limitations mentioned above like the low coupling rate and need of redox partners could be overcome.

The most widely studied natural fusion P450s are P450BM3 and P450RhF, and the self-sufficient redox partners of these two P450s have inspired various artificial fusions. By fusing novel P450s to the redox domain of the self-sufficient

P450s, it is expected that the fused protein is still self-sufficient but operates on chosen substrates. Another advantage is that the P450 domain and the redox domain are always produced in equivalent amounts. There have been various successful examples of artificial fusions using either the redox domains of P450BM3 or P450RhF. In previous research, the P450balk (CYP153A) was successfully fused to the redox domain of P450RhF for the terminal hydroxylation of octane⁽¹³⁹⁾. However, the homologues of P450balk (62% identity) were fused to RhF in the same manner but led to very poor production⁽¹⁴⁰⁾. This serves as a reminder that P450-redox fusions may not always produce a favourable interaction and are not guaranteed to deliver an efficient or even functional catalyst without further development.

1.5.2.2 Whole-cell engineering

Once the P450 is engineered with desired properties, the whole-cell system is ideal to provide sufficient redox equivalents to support the P450 based reactions. The disadvantages of using a cell-free system include an unstable environment and the continuous need for redox partners and redox factors. Therefore, the whole-cell system is a more viable option to implement the P450 involved reactions for biotechnological applications. After expressing P450 inside the host cells, it is expected that the P450 follows the designed pathways of biocatalysis and produce the right products. There are also some common problems when applying the whole-cell system like an insufficient supply of cofactors and limited product efflux, both of which will limit the production rate and may cause damage to the host cells. Some solutions, like co-expression of cofactor regeneration enzymes⁽¹⁴¹⁾ and applying effective permeabilizing agents⁽¹⁴²⁾, were proposed to overcome these limitations.

1.5.2.3 Multi-enzyme cascade design

The field of building novel P450s into an artificial cascade still needs exploration, which starts to attract attention from various disciplines. The advantages of a multi-enzyme cascade reaction are that all reactions happen in one-pot to gain the final product and several recycling systems could be built to provide recycling of cofactors. The multi-enzyme cascade could work as a closed system and

efficiently transform substrates into products⁽¹¹⁶⁾. One most extensively used method is to couple P450 with NAD(P)⁺ dependent enzymes. The extra enzyme only uses NAD(P)⁺ as the cofactor and yields NAD(P)H, while P450 relies on NAD(P)H as the redox cofactor for oxidation and produces NAD(P)⁺. The advantage of the cofactor regeneration system is providing continuous support for the P450-NAD(P)H involved reaction to oxidize substrates, but the limitation is that a large number of co-substrates is required as a driving force to push the regeneration enzyme to consume enough NAD(P)⁺ and produce NAD(P)H. The existence of some co-substrates and co-products may also inactivate the P450s and disrupt the main reaction we are expecting. Currently, there have been many regeneration enzymes being investigated, such as alcohol dehydrogenase, formate dehydrogenase, glucose dehydrogenase, glucose-6-phosphate dehydrogenase, phosphite dehydrogenase, hydrogenase, and pyridine nucleotide transhydrogenase⁽¹⁴³⁾.

1.6 Research aims

Cytochrome P450s are currently indispensable enzymes for biotechnological applications due to the variety of catalytic functions they mediate. Due to current advanced techniques in revealing genome sequence and synthesizing desired DNA sequences, the characterisation of novel enzymes is quite straightforward and time-efficient, which leads to the extensively expanding toolbox of cytochrome P450s. One of the promising directions in developing cytochrome P450 into a versatile catalyst for biotechnological use is the discovery and engineering of self-sufficient P450, which includes identification of novel natural fusion P450s from genomes and design of artificial fusions through site-directed mutagenesis or chimeric protein redesign. The chimeric protein redesign refers to the combination of redox domain and independent P450s (e.g. Class I P450 from prokaryotes and Class III P450 from eukaryotes). However, there are some limitations in the study of fusion P450s, for example, the occurrences of natural fusions are rather rare compared to the common Class I P450, which normally use specific substrates and defined reactions. It is difficult to put natural fusions directly into biotechnological uses. Besides, engineering of natural fusions could

lead to a variety of functions that are not working efficiently, e.g. preserving the advantageous traits of being self-sufficient as well as have a broad potential substrate spectrum. However, either directed mutants or chimeric redesigned fusions are not guaranteed to deliver the desired functions. Further optimisation of the artificial fusions may be required to achieve better production or quality of redesigned proteins. According to the above limitations of developing fusion P450s, other direction of developing P450 has been brought to attention, which focuses on the Class I cytochrome P450 and its naturally clustered redox partners. Just like the natural cluster of *p450cam*, *pdr*, and *pdx*, an artificial cluster could also be designed with replaceable genetic components. The three genes in one cluster should be co-activated and expressed together. Class I P450s are structurally simpler than the fusion P450, which is also a much larger group than fusion in the cytochromes P450 superfamily and potentially conducts a broader spectrum of reactions. Once the three-gene cluster is established, any Class I P450 of interest could be integrated into the cluster for production and activity test. The artificial cluster normally does not involve P450 and redox partners from the same organism. Therefore, one of the limitations is that there is no guaranteed cooperation among enzymes from different origins allowing the transfer of the electrons for catalysis.

There are three main aims of the entire research program, which are summarized below.

- The first aim of the project was to design and test the integrity of the operon containing the genes of P450 and redox partners, as well as to evaluate the effects of ferredoxin and ferredoxin reductase within the operon on enhancing the production of various actinomycete cytochrome P450s (Chapter 4). The purposes of constructing the three-gene operon include the complete production of Class I P450s and corresponding electron transfer proteins, as well as understanding how different electron transfer proteins may influence the production of P450s. There were a total of 7 P450 genes, 3 ferredoxin genes, and 2 ferredoxin reductase genes being involved in the construction of three-gene operon, and all gene fragments were prepared through PCR from genome samples stored in Ward lab. In total, 7 different

CYPs were selected, including the CYP105A1 and CYP105B1 from *Streptomyces griseolus*, CYP105D1 from *Streptomyces griseus*, CYP105D5 from *Streptomyces coelicolor*, CYP105D7 and CYP154C2 from *Streptomyces avermitilis*, CYP105AB1 from *Saccharopolyspora erythraea*. Three different combinations of redox partners were also used in construction, for example, ferredoxin 1 from *S.griseolus* and ferredoxin reductase from *S.ceolicolor*, ferredoxin 2 from *S.griseolus* and ferredoxin reductase from *S.ceolicolor*, and putidaredoxin and putidaredoxin reductase from the CAM plasmid. Each of the 7 CYPs was expressed with three different redox partners respectively, and the expression levels were measured and compared, which aimed to evaluate how redox partners may enhance the expression of CYPs. In the end, one of the combinations of redox partner was better than others on enhancing the production and quality of CYPs. The best combination of redox partners was eventually applied for expression with several other CYPs in characterisation studies.

- The second aim of the project was to apply selected operon constructs to facilitate stable and efficient production of active CYP105AB1, and to characterise the activity of CYP105AB1 against a pool of substrates, and eventually to test the potential bioconversions conducted by CYP105AB1 in a cell-free environment (Chapter 5).
- The third part of the thesis is to discover all potential CYPs within the *S.rishiriensis* genome (Chapter 6) or drain metagenome (Chapter 7) for future enzyme mining, as well as applying the established three-gene operon from Chapter 4 for potential CYP production in *E.coli*.

2 Chapter 2: Materials and Methods

2.1 General materials and methods

2.1.1 Chemical compounds

All chemicals mentioned in this thesis, if not specifically described, were purchased from Sigma Aldrich.

2.1.2 Genes and plasmids

All plasmids used in this project are listed in table 2.1.

Table 2.1. Complete list of plasmids generated in each project. Each plasmid used in this project is assigned with a unique pQR number. Other relevant characteristics of the plasmid are also described in the following tables, such as target genes for expression, antibiotic resistance, the copy number of the plasmid, and the use of the 6-histidine tag for target genes. The target genes refer to the genes assembled at the cloning site of the plasmid for expression. All antibiotics such as Kanamycin, Ampicillin, and Chloramphenicol are respectively recorded as Kan, Amp, and Chl.

Plasmids generated in Chapter 3.

pQR	Target gene*	Antibiotic resistance	Copy number	Histidine tag
2238	<i>cyp102a7</i>	Kan	High	Yes

*The target gene and its original organism are explained as below.
cyp102a7 from *Bacillus licheniformis* DSM 13.

Plasmids generated in Chapter 4 and Chapter 5.

pQR	Target genes*	Antibiotic resistance	Copy number	His tag
367	<i>cyp105a1</i> , <i>fd1</i> and <i>fdr_SCF15A</i>	Amp	High	No
368	<i>cyp105b1</i> , <i>fd2</i> and <i>fdr_SCF15A</i>	Amp	High	No
2239	<i>fd1</i> and <i>fdr_SCF15A</i>	Amp	High	No
2240	<i>fd2</i> and <i>fdr_SCF15A</i>	Amp	High	No
2241	<i>cyp105a1</i> , <i>fd1</i> and <i>fdr_SCF15A</i>	Amp	High	No
2242	<i>cyp105a1</i> , <i>fd2</i> and <i>fdr_SCF15A</i>	Amp	High	No
2243	<i>cyp105b1</i> , <i>fd1</i> and <i>fdr_SCF15A</i>	Amp	High	No
2244	<i>cyp105b1</i> , <i>fd2</i> and <i>fdr_SCF15A</i>	Amp	High	No
2245	<i>cyp105d1</i> , <i>fd1</i> and <i>fdr_SCF15A</i>	Amp	High	No
2246	<i>cyp105d1</i> , <i>fd2</i> and <i>fdr_SCF15A</i>	Amp	High	No
2247	<i>cyp105d5</i> , <i>fd1</i> and <i>fdr_SCF15A</i>	Amp	High	No
2248	<i>cyp105d5</i> , <i>fd2</i> and <i>fdr_SCF15A</i>	Amp	High	No
2249	<i>cyp105d7</i> , <i>fd1</i> and <i>fdr_SCF15A</i>	Amp	High	No
2250	<i>cyp105d7</i> , <i>fd2</i> and <i>fdr_SCF15A</i>	Amp	High	No
2251	<i>cyp105ab1</i> , <i>fd1</i> and <i>fdr_SCF15A</i>	Amp	High	No
2252	<i>cyp105ab1</i> , <i>fd2</i> and <i>fdr_SCF15A</i>	Amp	High	No
2253	<i>cyp154c2</i> , <i>fd1</i> and <i>fdr_SCF15A</i>	Amp	High	No
2254	<i>cyp154c2</i> , <i>fd2</i> and <i>fdr_SCF15A</i>	Amp	High	No
2255	<i>cyp105a1</i>	Amp	High	No

pQR	Target genes*	Antibiotic resistance	Copy number	His tag
2256	<i>cyp105b1</i>	Amp	High	No
2257	<i>cyp105d1</i>	Amp	High	No
2258	<i>cyp105d5</i>	Amp	High	No
2259	<i>cyp105d7</i>	Amp	High	No
2260	<i>cyp105ab1</i>	Amp	High	No
2261	<i>cyp154c2</i>	Amp	High	No
2262	<i>fd_sery</i> and <i>fdr_SCF15A</i>	Amp	High	No
2263	<i>cyp105ab1</i> , <i>fd_sery</i> and <i>fdr_SCF15A</i>	Amp	High	No
2264	<i>cyp105ac1</i> , <i>fd_sery</i> and <i>fdr_SCF15A</i>	Amp	High	No
2265	<i>cyp105ac1</i> , <i>fd2</i> and <i>fdr_SCF15A</i>	Amp	High	No
2266	<i>cyp296a1</i> , <i>fd_sery</i> and <i>fdr_SCF15A</i>	Amp	High	No
2267	<i>cyp296a1</i> , <i>fd2</i> and <i>fdr_SCF15A</i>	Amp	High	No
2268	<i>cyp155a2</i> , <i>fd_sery</i> and <i>fdr_SCF15A</i>	Amp	High	No
2269	<i>cyp155a2</i> , <i>fd2</i> and <i>fdr_SCF15A</i>	Amp	High	No
2270	<i>cyp295a1</i> , <i>fd_sery</i> and <i>fdr_SCF15A</i>	Amp	High	No
2271	<i>cyp295a1</i> , <i>fd2</i> and <i>fdr_SCF15A</i>	Amp	High	No
2272	<i>cyp186b1</i> , <i>fd_sery</i> and <i>fdr_SCF15A</i>	Amp	High	No
2273	<i>cyp186b1</i> , <i>fd2</i> and <i>fdr_SCF15A</i>	Amp	High	No
2274	<i>cyp105ac1</i>	Amp	High	No
2275	<i>cyp296a1</i>	Amp	High	No
2276	<i>cyp155a2</i>	Amp	High	No
2277	<i>cyp295a1</i>	Amp	High	No
2278	<i>cyp186b1</i>	Amp	High	No
2279	<i>cyp105a1-His</i> , <i>fd2</i> and <i>fdr_SCF15A</i>	Amp	High	YES
2280	<i>cyp105b1-His</i> , <i>fd2</i> and <i>fdr_SCF15A</i>	Amp	High	YES
2281	<i>cyp105ab1-His</i> , <i>fd2</i> and <i>fdr_SCF15A</i>	Amp	High	YES
2282	<i>cyp105ac1-His</i> , <i>fd2</i> and <i>fdr_SCF15A</i>	Amp	High	YES
2283	<i>cyp154c2-His</i> , <i>fd2</i> and <i>fdr_SCF15A</i>	Amp	High	YES
2284	<i>cyp296a1-His</i> , <i>fd2</i> and <i>fdr_SCF15A</i>	Amp	High	YES
2285	<i>cyp105ab1-His (R78G)</i> , <i>fd2</i> and <i>fdr_SCF15A</i>	Amp	High	YES
2286	<i>cyp105ab1-His (T108A)</i> , <i>fd2</i> and <i>fdr_SCF15A</i>	Amp	High	YES
2287	<i>cyp105ab1-His (R287D)</i> , <i>fd2</i> and <i>fdr_SCF15A</i>	Amp	High	YES
2288	<i>pdx-His</i>	Kan	High	YES
2289	<i>pdr-His</i>	Kan	High	YES
2290	<i>pdx</i> and <i>pdr</i>	Amp	High	NO
2291	<i>cyp105a1</i> , <i>pdx</i> and <i>pdr</i>	Amp	High	NO
2292	<i>cyp105b1</i> , <i>pdx</i> and <i>pdr</i>	Amp	High	NO
2293	<i>cyp15ab1</i> , <i>pdx</i> and <i>pdr</i>	Amp	High	NO
2294	<i>cyp105ac1</i> , <i>pdx</i> and <i>pdr</i>	Amp	High	NO
2295	<i>cyp154c2</i> , <i>pdx</i> and <i>pdr</i>	Amp	High	NO
2296	<i>cyp296a1</i> , <i>pdx</i> and <i>pdr</i>	Amp	High	NO
2297	<i>cyp105d1</i> , <i>pdx</i> and <i>pdr</i>	Amp	High	NO
2298	<i>cyp105d5</i> , <i>pdx</i> and <i>pdr</i>	Amp	High	NO
2299	<i>cyp105d7</i> , <i>pdx</i> and <i>pdr</i>	Amp	High	NO

* The target gene and its original organism are explained as below.

cyp105a1 and *cyp105b1* from *S.griseolus*;

cyp105ab1, *cyp105ac1*, *cyp155a2*, *cyp295a1*, *cyp296a1* and *cyp186b1* from *Saccharopolyspora erythraea*;

cyp105d1 from *S.griseus*;

cyp105d5 from *S.coelicolor*;

cyp105d7 and *cyp154c2* from *S.avermitilis*;

fd1 from *S.griseolus*;

fd2 from *S.griseolus*;

fd_sery from *S.erythraea*;

fdr_SCF15A from *S.coelicolor*;

pdx and *pdr* from *P.putida*;

Plasmids generated in Chapter 6.

pQR	Target genes*	Antibiotic resistance	Copy number	His tag
2300	<i>sri_cyp03-His, fd2</i> and <i>fdr_SCF15A</i>	Amp	High	YES
2301	<i>sri_cyp13-His, fd2</i> and <i>fdr_SCF15A</i>	Amp	High	YES
2302	<i>sri_cyp24-His, fd2</i> and <i>fdr_SCF15A</i>	Amp	High	YES
2303	<i>sri_cyp03-His</i>	Amp	High	YES
2304	<i>sri_cyp13-His</i>	Amp	High	YES
2305	<i>sri_cyp24-His</i>	Amp	High	YES
2306	<i>cyp162b1-His</i>	Amp	High	YES
2307	<i>groEL, groES, dnaK, dnaJ</i> and <i>grpE</i>	Chl	Low	NO
2308	<i>groEL</i> and <i>groES</i>	Chl	Low	NO
2309	<i>dnaK, dnaJ</i> and <i>grpE</i>	Chl	Low	NO
2310	<i>groEL, groES</i> and <i>tig</i>	Chl	Low	NO
2311	<i>tig</i>	Chl	Low	NO

*The target gene and its original organism are explained as below.

sri_cyp03, sri_cyp13 and *sri_cyp24* from *S.rishiriensis*;

cyp162b1 from *Saccharopolyspora erythraea*;

fd2 from *S.griseolus*;

fdr_SCF15A from *S.coelicolor*.

groEL, groES, dnaK, dnaJ, grpE, and *tig* are the chaperone genes from a commercial kit.

Plasmids generated in Chapter 7.

pQR	Target genes*	Antibiotic resistance	Copy number	His tag
2312	<i>dmg_cyp08-His, fd2</i> and <i>fdr_SCF15A</i>	Amp	High	YES
2313	<i>dmg_cyp09-His, fd2</i> and <i>fdr_SCF15A</i>	Amp	High	YES
2314	<i>dmg_cyp08-His, pdx</i> and <i>pdr</i>	Amp	High	YES
2315	<i>dmg_cyp09-His, pdx</i> and <i>pdr</i>	Amp	High	YES

*The target gene and its original organism are explained as below.

dmg_cyp08-His and *dmg_cyp09-His* from metagenome

fd2 from *S.griseolus*;

fdr_SCF15A from *S.coelicolor*;

pdx and *pdr* from *P.putida*.

2.1.3 Strains and media

2.1.3.1 Chemically competent *E.coli* strains

Most *E.coli* cells, if not specified, in this study were chemically competent, which were purchased from Thermo Fisher Scientific. The three main types of *E.coli* were One Shot™ TOP10, One Shot™ BL21(DE3), and One Shot™ BL21(DE3)pLysS.

Some chemically competent *E.coli* cells, that used in double transformation processes, were made in the lab. The basic principles of making chemically competent cells are harvesting cells at the exponential phase in a cold environment and complete pellet-washing with 0.1 M CaCl₂. The protocol of

making competent cells started with an inoculation of 20 mL LB with single colony and cultivation at 37°C. After OD₆₀₀ of the cell culture reached 0.6, cells were harvested by centrifugation for 3 min at 6000 rpm at 4°C. The cell pellet was gently resuspended with 10 mL cold 0.1 M CaCl₂ and then incubated on ice for 20 min. The cells were spun down at the same conditions as before, and the pellet was gently resuspended with storing buffer (75% v/v 0.1 M CaCl₂ and 15% v/v glycerol). The cell resuspension was dispensed into autoclaved microtubes (300 µL/tube) and frozen in -80°C. The CaCl₂ buffer was filtered and the glycerol was autoclaved before use.

2.1.3.2 Media

Terrific broth (TB) was the nutrient broth used in this study for cell propagation as well as protein expression. The recipe of terrific broth was adopted from the Cold Spring Harbor protocol. 12 g tryptone, 24 g yeast extract, and 4 mL glycerol were mixed with deionized water (ddH₂O) until the final volume reached 900 mL, which was sterilized by autoclaving. The terrific broth was made of the 900 mL sterile media and 100 mL filtered solution of 0.17 M KH₂PO₄ and 0.72 M K₂HPO₄. The tryptone and yeast extract were purchased from Thermo Scientific Oxoid Microbiology Products, and the glycerol was from Sigma-Aldrich.

LB-Agar was the solid media used in this study for colony selection. The LB-Agar (Miller) powder was purchased from MerckMillipore and prepared and sterilized according to the standard protocol provided.

Yeast extract-Malt extract (YEME) medium was used for the cultivation of *Streptomyces rishiriensis* DSM 40489. Every 1 litre of YEME medium was prepared with 3 g of yeast extract, 3 g of malt extract, 5 g of bacteriological peptone, 10 g of glucose, 170 g of sucrose, and distilled water⁽¹⁴⁴⁾, which was later autoclaved and dispensed in the flask for microbial cultivation.

2.1.4 Transformation, inoculation, and glycerol stock

2.1.4.1 Transformation

The transformation protocol used in this study was slightly adjusted from the standard protocol for One Shot™ competent cells provided by Thermo Fisher Scientific.

After thawing the One Shot™ competent cells, approximately 50 ng of target plasmids were added to 25 µL of cells and mixed by gentle tapping in a 2 mL microcentrifuge tube. The mixture of plasmids and competent cells was incubated on ice for 20 min before sending for heat shock treatment at 42°C for the exact 30 s. After heat shock, 250 µL of SOC medium was added to the microcentrifuge tube and the cell suspension was cultivated at 37°C for 1 h at 300rpm. In the end, 50-100 µL cell culture was evenly plated on a pre-warmed LB agar plate with appropriate antibiotics. The plates were inverted and placed at 37°C for 24 h incubation before distinctive colonies could be observed and selected for further analysis and inoculation.

2.1.4.2 Inoculation and glycerol stock

A single colony was selected for inoculation by using a sterile single-use inoculation loop, and the cells were propagated in 5 mL TB with appropriate antibiotics overnight. 800 µL of the overnight culture was mixed with 800 µL 50% v/v autoclaved glycerol to prepare glycerol stock in 2 mL microcentrifuge tubes, which was eventually stored in -80°C.

2.1.5 Sub-cloning and sequencing

2.1.5.1 Sub-cloning

For preserving DNA fragments for further usage, a sub-cloning kit called Zero Blunt® TOPO® PCR Cloning (Thermo Fisher Scientific) was used to insert any blunt-ended PCR products into a linearized plasmid. Both 3' ends of the linearized plasmid were covalently bound to DNA topoisomerase I, which were subsequently attacked by the 5' end hydroxyl group of any potential PCR fragments. Because the kit was designed specifically to insert blunt-end PCR products, so any primers used for producing these PCR products should not have added 5', phosphate group. Before conducting the insertion to the TOPO vector, the quality and quantity of PCR products were verified by using agarose gel

electrophoresis. If there was not a single, discrete band of PCR products, other operations, like optimisation of PCR conditions or gel-purification of desired products, might be conducted to acquire PCR products with higher quality and quantity.

The total volume of cloning reaction was normally 6 μL , including 1 μL of linearized TOPO plasmid, 1 μL of salt solution (1.2 M NaCl and 0.06 M MgCl_2) and 4 μL of PCR products. The reaction was conducted for 5 min at room temperature, before adding 2 μL of the reaction mixture into a vial of chemically competent cells for general transformation. The successfully transformed cells were selected for plasmid extraction, and the integrity of the inserted DNA fragment was checked through sequencing.

2.1.5.2 Sequencing

The sequencing of all plasmid studied in this project was carried out by Eurofins Genomics. Any plasmid samples were prepared to reach the concentration between 50 and 100 ng/ μL , before being shipped to Eurofins Genomics facility for sequencing. The sequencing service of Eurofins Genomics is an advanced version of traditional Sanger sequencing conducted on ABI 3730XL sequencing machines. The sequencing method is also called cycle sequencing, which involves using 4 ddNTPs fluorescently labeled with four different dyes as dideoxy terminator nucleotides. By mixing plasmid templates, primers, thermostable DNA polymerases, dNTPs, ddNTPs, as well as a buffer containing Mg^{++} and K^+ in a single reaction tube under appropriate PCR conditions, the extension reaction continues until a particular ddNTP is incorporated to generate numerous DNA fragments with different length and one labeled dideoxynucleotide at the 3' end. The reaction sample is electrokinetically injected into a capillary sequencer, which leads to all fluorescently labeled fragments migrate towards anode by size. When the DNA fragments move along in the capillary, a laser beam excites these dye molecules to produce fluorescent signals that represent the identities of the tagged ddNTPs. Finally, the colour readout is translated into a sequence using Eurofins analysis software.

The sequencing primers used in the project are summarized in Table 2.2. The sequencing results were downloaded from Eurofins Genomics and compared to

original sequence using plasmid editing software like SnapGene and ApE to check the integrity of targeted regions on DNA samples.

Table 2.2. Primers used in Eurofins Genomics sequencing service. Some of the standard primers provided by Eurofins Genomics were used, while some customized primers were synthesized before sending along with plasmid samples for sequencing.

Primer Name	Primer Sequences (5' to 3')	Description
M13 forward	TGAGTTTCGTCACCGTA	Standard primer
M13 reverse	CAGGAAACAGCTATGACC	Standard primer
T7	TAATACGACTCACTATAGGG	Standard primer
T7 term	ACGAACTCGATTGACGCGT	Synthesized
CYP reverse	CTTCCTATTAGTCGGTCACCGT	Synthesized

2.1.6 DNA manipulation

2.1.6.1 PCR profile

Every polymerase chain reaction (PCR) conducted in this project was prepared in 0.2 mL polypropylene PCR tubes manufactured by Star Lab. The thermal cycler used in this project is called Prime Thermal Cycler with a 96-well heat block, which was manufactured by TECHNE. All DNA polymerases used in this work, including Phusion® High-Fidelity DNA Polymerase and Q5® High-Fidelity DNA Polymerase, were manufactured by New England Biolabs (NEB). The polymerase master mix was prepared by NEB, which included DNA polymerase, dNTP, MgCl₂, and optimised reaction buffer. Therefore, each PCR mixture was simplified as mixing templates, primers, master mix, and DNA-free water offering robust performance and easily replicable for any other PCR reactions.

The primers synthesized by Eurofins Genomics were delivered to the lab as lyophilised samples, so the oligo powders were dissolved and diluted to 10 µM using DNA-free water. High quality and purified DNA templates were always preferred for the enhanced success of PCR reactions. The recommended DNA template concentration was between 50 and 100 ng/µL.

As is shown in Table 2.3 and Table 2.4, the routine PCR setup and thermocycling conditions were described.

Table 2.3. Reaction setup for routine PCR by using polymerase master mix.

All reaction components should be placed on ice and mixed properly before sending them to the thermocycler. For optimised PCR, DMSO might be added, and the volume was controlled to be 0.75 μL in 20 μL total reaction.

Component	20 μL Reaction	Final Concentration
10 μM Forward Primer	1 μL	0.5 μM
10 μM Reverse Primer	1 μL	0.5 μM
2X Phusion Master Mix	10 μL	1X
Template DNA	Variable	50-100 ng
Nuclease-free water	to 20 μL	

Table 2.4. Thermocycling conditions for routine PCR.

For any common PCR, the 3-step cycling reaction that included denaturation, annealing, and extension was applied. The annealing temperature was dependent on the complementary regions of both forward and reverse primers on template DNA. The T_m Calculator website developed by NEB was used to calculate the optimal annealing temperature for each primer pair (<http://tmcalculator.neb.com/#!/main>). A temperature gradient may be applied to optimize the annealing temperature for each primer pair.

Step	Temperature	Time
Initial Denaturation	98°C	30 seconds
30 Cycles	98°C	10 seconds
	Calculated Annealing Temp	30 seconds
	72°C	30 seconds per kb
Final Extension	72°C	10 minutes
Hold	4°C	

2.1.6.2 Restriction digestion

All restriction enzymes and restriction digestion buffer used in this project were purchased from NEB. CutSmart® restriction buffer from NEB was used as the only buffer for optimised single or double digestion, which provided 20 mM Tris-acetate, 10 mM Magnesium acetate, 50 mM Potassium acetate and 100 $\mu\text{g}/\text{ml}$ BSA (bovine serum albumin) at the final digestion mixture.

A routine restriction digest setup was shown in Table 2.5. The common incubation time for digestion reactions was 60 minutes, and the incubation temperature was enzyme-dependent. Although the incubation time could be reduced to 5-15 minutes by using Time-Saver™ Qualified enzyme, 60-minute incubation was always applied for sufficient digestion.

Table 2.5. Reaction setup for routine restriction digestion. For double digestion, an extra 1 μL restriction enzyme was added, but the total reaction volume was not changed.

Component	20 μL Reaction
Restriction enzyme	1 μL
DNA	Up to 1 μg
10X CutSmart buffer	2 μL
Nuclease-free Water	to 20 μL

2.1.6.3 Ligation with T4 DNA ligase

The T4 DNA ligase and reaction buffer were purchased from NEB to catalyse the formation of a phosphodiester bond between 5' phosphate and 3' hydroxyl termini in double-stranded DNA. The 10X reaction buffer was prepared by NEB, which provided 50 mM Tris-HCl, 10 mM MgCl_2 , 1 mM ATP, and 10 mM DTT in the final ligation mixture. A routine ligation reaction setup was shown below in Table 2.6. The general molar ratio of vector DNA to insert DNA was 1:3, but it was always worth to increase the portion of insert DNA to achieve higher ligation success. After preparing the ligation reaction as indicated in Table 2.6, the reaction was gently mixed by pipetting up and down and left at room temperature for 10 minutes. The reaction mixture was heat-inactivated at 65°C for 10 minutes, before being chilled on ice and transformed to chemically competent cells according to the general transformation protocol. The successfully ligated plasmids were extracted from host cells and checked for plasmid integrity through sequencing.

Table 2.6. Routine ligation reaction setup by using T4 DNA ligase. The T4 DNA ligase should be added last, and the molar ratio of 1:3 vector to insert was applied. To determine the amount of insert DNA needed in a single ligation reaction,

NEBioCalculator (<http://nebiocalculator.neb.com/#!/ligation>) was used to calculate based on the length of both vector and insert DNA.

Component	20 μL Reaction
10X T4 ligase buffer	2 μ L
Vector DNA	50 ng
Insert DNA	It depends on the length of both vector DNA and insert DNA
Nuclease-free water	To 20 μ L
T4 DNA ligase	1 μ L

2.1.6.4 Gibson assembly

Gibson assembly allows efficient assembly of multiple overlapping DNA fragments in a single tube reaction. The Gibson Assembly Master Mix was purchased from NEB, which contained exonucleases, polymerases, and ligases. The exonuclease creates the single-strand 3' overhangs that lead to annealing of fragments that share complementary region. The polymerase fills in gaps within each annealed fragment. The ligase closes the nicks in the assembled DNA.

Primers were firstly designed to amplify all fragments with appropriate overlapping regions through PCR. The recommended length of overlap between adjacent fragments is 40 bp. All PCR products were purified and quantified through gel extraction. 0.5 pmol of each fragment was mixed with 10 μ L 2X Gibson Assembly Master Mix as well as nuclease-free water to achieve a 20 μ L reaction. The reaction mixture was incubated at 50°C for 15 minutes, which was followed by subsequent transformation to chemically competent cells. It was recommended in the NEB protocol that extended incubation up to 60 minutes might help to improve assembly efficiency.

2.1.6.5 Plasmid extraction and purification

The QIAprep Spin Miniprep kit from QIAGEN was used for the isolation of high-purity plasmid DNA from *E.coli* host. The procedure is based on the alkaline lysis of bacterial cells followed by the absorption of plasmid DNA onto silica in the

presence of high salt (Vogelstein and Gillespie, 1979). In the QIAprep miniprep kit, five different buffers were being used, which included buffer P1 (50 mM Tris-Cl, pH 8.0, 10 mM EDTA and 100 µg/ml RNase A), buffer P2 (200 mM NaOH and 1% SDS w/v), buffer N3 (4.2 M guanidine hydrochloride, 0.9 M potassium acetate, pH 4.8), wash buffer PE (10 mM Tris-HCl pH 7.5, 80% ethanol) and buffer EB (10mM Tris-Cl, pH 8.5).

In the routine procedure of plasmid extraction and purification, it started from the inoculation of fresh media with a single colony or glycerol stock to prepare a 5 mL overnight culture. The cell pellet of 2 mL overnight culture was harvested and then resuspended with 250 µL buffer P1, which leads to none visible cell clumps after resuspension. 250 µL buffer P2 was then added and mixed thoroughly for lysis reaction for no more than 4 min. The reaction was stopped by mixing with 350 µL buffer N3 before being cleared by centrifugation at 13000 rpm for 10 minutes. The acquired clear supernatant was then applied to the QIAprep spin column for DNA capturing onto the silica gel. The column was washed once with 750 µL buffer PE to wash away molecules that did not bind, and the column was additionally spun down to remove residual wash buffer. 30 µL buffer EB was finally applied on the silica gel to elute any plasmid DNA bond onto the column. The eluted sample was collected in the 1.5 mL centrifuge tube and measured using Nanodrop to determine the concentration.

2.1.6.6 Agarose gel preparation

The agarose powder was mixed with TAE buffer (40mM Tris, 20mM acetic acid, and 1mM EDTA), and heated up in the microwave to prepare 0.8% (w/v) agarose gel. The boiled agarose gel was added with ethidium bromide stock to reach a final concentration of 0.6 µg/mL ethidium bromide, before being poured into a cast. A comb was placed at the front of the cast to create wells for sample loading once the gel was completely set. Both the agarose powder and ethidium bromide stock were purchased from Sigma Aldrich.

The set gel was placed into a gel tank, which was filled with TAE buffer to completely cover the surface of the gel. For any DNA samples, they were mixed thoroughly with 6X DNA loading dye (NEB) before being loaded into each well in

the gel. The accompanied DNA ladder was the 1kb HyperLadder from Bioline, which covered the size range from 200 bp to 10 kb. In this project, the recommended voltage and time for better resolution were 120 V and 45 minutes respectively. Because DNA molecules are negatively charged, so DNA samples were loaded close to the negative electrode to move along the electric field. After electrophoresis, the gel was visualised under 360 nm UV light to identify target DNA bands with specific fragment length using AlphaView software (ProteinSimple).

2.1.6.7 Gel extraction for fragment purification

The QIAquick Gel Extraction Kit from QIAGEN was used for isolation and purification of DNA fragments that separated on the agarose gel. In the kit, three buffers were being used, which included buffer QG (proprietary composition), buffer PE (10 mM Tris-HCl pH 7.5, 80% ethanol), and buffer EB (10mM Tris-Cl, pH 8.5). Gel slices were dissolved in buffer QG containing a pH indicator, allowing direct determination of the optimal pH for DNA binding. Normally, three volumes of QG buffer was used to dissolve every 0.1 g of gel slices (0.1 g ~ 100 μ L). If the colour of the mixture was orange or violet rather than yellow, 10 μ L of 3 M sodium acetate, pH 5.0, was added to optimise the pH below 7.5 for efficient membrane binding. Once the gel slice was completely dissolved in a 50°C incubator, one volume of isopropanol was added to the mixture. The mixture was then applied to the QIAquick spin column for nucleic acid adsorption to the silica membrane in the high-salt conditions provided by buffer QG. Impurities were washed away by buffer PE and pure DNA was eluted with 30 μ L low-salt buffer EB.

2.1.6.8 Measurement of DNA concentration

The NanoDrop 2000c microvolume spectrophotometer from Thermo Fisher Scientific was used for DNA quantification with a small volume. Generally, 1.5 μ L was used as the default volume for sample loading, and the blank buffer was buffer EB. Within the NanoDrop software, a modified Beer-Lambert law was used

in the background to calculate sample concentrations once the absorbance of the DNA sample was acquired.

$$c = A/(\varepsilon * b)$$

c is the calculated concentration of DNA, A is the absorbance in AU, and b is the pathlength in cm. ε is the wavelength-dependent extinction coefficient in $\text{ng} \cdot \text{cm} / \mu\text{L}$, which is generally accepted as $50 \text{ ng} \cdot \text{cm} / \mu\text{L}$ for double-stranded DNA, $33 \text{ ng} \cdot \text{cm} / \mu\text{L}$ for single-stranded DNA and $40 \text{ ng} \cdot \text{cm} / \mu\text{L}$ for RNA at 260 nm. Therefore, the final concentration displayed by Nanodrop software was reported in terms of $\text{ng} / \mu\text{L}$.

The purity ratio was also calculated by the software, which helped to roughly define if the DNA sample was pure enough for subsequent applications. A 260/280 ratio of around 1.8 was pure for DNA sample, while a 260/230 ratio was commonly higher than 260/280 value and in the range of 1.8-2.2. Apart from the purity ratio, the spectral curve was also plotted to evaluate sample quality, which should indicate a defined peak at 260 nm and a defined trough at 230 nm for high-quality DNA samples.

2.1.7 Protein manipulation

2.1.7.1 Heterologous protein expression

Plasmids containing a cytochrome P450 gene were transformed into *E.coli* BL21(DE3) cells by a standard heat-shock protocol that was described in section 2.1.4.1. The transformed cells were stored as glycerol stocks at -80°C . An aliquot from a frozen glycerol stock of plasmid containing *E. coli* BL21(DE3) was inoculated into 5 mL of TB medium (containing $100 \mu\text{g}/\text{mL}$ ampicillin). Starter cultures were incubated 18-20 h at 37°C , shaking at 250 revolutions per minute (rpm). 1 mL of this overnight culture was transferred into 100 mL fresh TB media (containing $100 \mu\text{g}/\text{mL}$ ampicillin) in 500 mL baffled conical shake flask. These expression cultures were normally incubated for 2-2.5 hours at 37°C at 250 rpm until OD_{600} of the cultures reached between 0.6 and 0.8. The culture was then followed by the addition of FeCl_3 (final concentration $500 \mu\text{M}$) and 5-aminolevulinic acid (final concentration $1000 \mu\text{M}$) to be incubated at 25°C , whilst

rotating at 180 rpm for 20 minutes. Expression was then induced by the addition of IPTG. The final concentration of IPTG, if not specifically described, was normally 500 μ M. Cultures were incubated for a further 24 hours at 25°C before harvesting, whilst shaking at 180 rpm. Cells were harvested by centrifugation at 10,000 rpm for 10 mins at 4°C. The supernatant was removed and the cells stored at -20°C until purification or carbon monoxide assay.

Plasmids containing an alcohol dehydrogenase gene were transformed into *E.coli* BL21(DE3) cells by a standard heat-shock protocol that was described in section 2.1.4.1. The transformed cells were stored as glycerol stocks at -80°C. An aliquot from a frozen glycerol stock of plasmid containing *E. coli* BL21(DE3) was inoculated into 5 mL of TB medium (containing 50 μ g/mL kanamycin). Starter cultures were incubated 18-20 h at 37°C, shaking at 250 revolutions per minute (rpm). 1 mL of this overnight culture was transferred into 100 mL fresh TB media (containing 50 μ g/mL kanamycin) in a 500 mL baffled conical shake flask. These expression cultures were normally incubated for 1 hour at 37°C at 250 rpm until OD₆₀₀ of the cultures reached between 0.6 and 0.8. Expression was then induced by the addition of IPTG (final concentration of 500 μ M). Cultures were incubated for a further 4 hours at 37°C before harvesting, whilst shaking at 250 rpm. Cells were harvested by centrifugation at 10,000 rpm for 10 mins at 4°C. The supernatant was removed and the cells stored at -20°C until purification.

2.1.7.2 Sonication and BugBuster® for cell lysis

The cell lysis was achieved either physically or chemically in this project. The physical approach of sonication was commonly applied to break *E.coli* cells. The cell pellet that expressed target proteins was resuspended in lysis buffer (10% culture volume, 50 mM potassium phosphate, pH 7.5) and homogenised by sonication (10-second on and 10-second off, 10 cycles, 8 μ m). The insoluble portion of the lysate was pelleted by centrifugation (10,000 rpm, 30 minutes). The clarified supernatant was preserved for further experiments.

The 10X BugBuster protein extraction reagent (Merck Millipore, Darmstadt, Germany) was also purchased. The concentrated BugBuster reagent was diluted with lysis buffer (50 mM potassium phosphate, pH 7.5) to get 1X BugBuster lysis

buffer that could be used directly for cell lysis. The wet cell pellet was resuspended with 1X BugBuster lysis buffer at room temperature, using 5 mL of the buffer per gram of wet cell paste. To reduce the viscosity of the lysate, 1 μ L of Benzonase Nuclease was added to every 1 mL of 1X BugBuster lysis buffer. The lysate was incubated on a rotating platform at low-speed settings for 15 minutes at room temperature, before removing insoluble debris by centrifugation at 10,000 rpm for 20 minutes at 4°C. The clarified supernatant was preserved for further experiments such as SDS-PAGE analysis or column purification.

2.1.7.3 Purification through Ni-NTA Agarose column

The Ni-NTA agarose resin from Thermo Fisher Scientific was used for purification of recombinant proteins expressed in host cells from any 6XHis-tagged vector. Ni-NTA Agarose uses nitrilotriacetic acid (NTA), a tetradentate chelating ligand, in a highly cross-linked 6% agarose matrix. NTA binds Ni²⁺ ions by four coordination sites. The nickel NTA was designed to capture histidine-tagged protein due to the high affinity and selectivity for the imidazole ring on the side chain of histidine. Proteins bound to the resin were eluted with low pH buffer or by competition with a high concentration of imidazole or histidine. As is shown in Table 2.7., three different buffers were prepared before packing the column with Ni-NTA agarose resin.

Table 2.7. Buffers used in the Ni-NTA purification system.

Buffer	Composition and pH
Binding buffer	50 mM potassium phosphate, pH 7.5
Wash buffer	50 mM potassium phosphate, 20 mM imidazole, pH 8.0
Elution buffer	50 mM potassium phosphate, 250 mM imidazole, pH 8.0

The Ni-NTA column was firstly prepared by using a clean and empty PD-10 column (GE healthcare) and the resuspended Ni-NTA Agarose. 5 mL of the agarose resin was poured into the column, allowing to settle completely by gravity (5-10 minutes) or pellet it by low-speed centrifugation (1 minute at 800 × g). 6 mL of sterile, distilled water was then added to resuspend the resin. For purification

under native conditions, 5 mL of native binding buffer was added to resuspend the resin and equilibrate the column. After the resin was completely settled, 5 mL of clarified cell lysate was gently loaded onto the surface of the resin, allowing to flow through the column by gravity. The column was then washed with 5 mL of wash buffer twice, before loading 5 mL of elution buffer to elute any target proteins. The resin in the column was finally washed with binding buffer and preserved in 20% ethanol.

The eluted protein samples were in a high concentration of imidazole, which may not be directly used in further protein characterisation. Therefore, a Sephadex G-25 in PD-10 desalting columns from GE healthcare was used to exchange the high imidazole of eluted protein samples with 50 mM potassium phosphate buffer, pH 7.5. The protein samples acquired after buffer exchange were normally concentrated using the vivaspin protein concentrator spin column (GE healthcare) with a molecular weight cut-off of 30 kDa.

2.1.7.4 SDS-PAGE

Protein analysis was achieved by using sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE), typically with 12% polyacrylamide gels unless noted otherwise. The common voltage and running time were 150 V and 45 minutes for 10-well Mini-PROTEAN® TGX™ precast gels from Bio-Rad. The running buffer for SDS polyacrylamide gel was generally the composition of 25 mM Tris, 192 mM glycine, and 0.1% (w/v) SDS. Gels were stained using Imperial Protein Stain (ThermoScientific), which was made of coomassie dye R-250. Protein purity and running mass on SDS-PAGE were measured with AlphaView software (ProteinSimple).

Protein loading samples were firstly prepared by mixing 30 µL of purified protein or clarified lysate, 15 µL of 4X Laemmli sample buffer (Bio-Rad), and 15 µL of 1M DTT. The sample mixture was then heated at 90–95°C for 5 min, before loading 20 µL of the mixture into wells of the precast Mini-PROTEAN® gel. The protein ladder used in the whole project was PageRuler™ prestained protein ladder, 10 to 180 kDa (Thermo Fisher Scientific).

2.1.8 Carbon monoxide assay

2.1.8.1 Sample preparation

The amount of active P450 was quantified by CO assay, which was developed by Omura and Sato in 1964⁽¹⁵⁾. All buffers and reagents used for carbon monoxide assay were prepared fresh on the day of assay. For quantification of cytochrome P450, 20 mL of culture was harvested and centrifuged for 10 minutes at 10000 rpm at 4°C. The cell pellet was then resuspended in 3 mL of CO assay lysis buffer (80% v/v 100 mM potassium phosphate buffer pH 7.5, 20% v/v filtered glycerol, 2 mM EDTA, 1.5 mM DTT and 0.4% v/v Triton X-100). The cell resuspension was left on ice for 20 minutes, before sonication (10-second on and 10-second off, 10 cycles, 8 μ m). The sonicated lysate was centrifuged for 45 minutes at 10000 rpm at 4°C, and the supernatant was used in the carbon monoxide assay.

2.1.8.2 Carbon monoxide assay

Before running the carbon monoxide assay, a stock solution of 250 mM sodium dithionite was prepared using CO assay lysis buffer, and the clarified lysate was left to reach room temperature. Two clean cuvettes (1.5 mL capacity) were equally filled with 900 μ L clarified lysate each. One of the cuvettes was used for carbon monoxide binding, while the other stayed the same. A fume cupboard was used, and the carbon monoxide was bubbled through lysate using a syringe needle at a rate of 1 bubble per second for 30 seconds. After carbon monoxide binding for one cuvette, both cuvettes were sealed with parafilm. 100 μ L of the stock sodium dithionite was added to each cuvette and inverted both cuvettes for 3 times. The reducing reaction with sodium dithionite was left for 3 minutes, before recording the absorbance between 400 to 500 nm using NanoDrop 2000c, which led to the plotting of a reduced spectrum and a CO binding spectrum. The NanoDrop 2000c was blanked using CO assay lysis buffer (80% v/v 100 mM potassium phosphate buffer pH 7.5, 20% v/v filtered glycerol, 2 mM EDTA, 1.5 mM DTT and 0.4% v/v Triton X-100). In the end, the carbon monoxide difference spectrum was calculated by subtracting the reduced spectrum from the CO

binding spectrum. An extinction coefficient of $91 \text{ mM}^{-1}\text{cm}^{-1}$ for carbon monoxide difference spectrum was used to determine the molar concentration of the active form of cytochrome P450 in the clarified cell lysate. Because the cell lysate was prepared from exact 20 mL of cell culture, so the total production of active cytochrome P450 for a specific batch of cell culture could be approximately determined if protein loss was not considered and the lysis was thought to be thorough and complete.

After acquiring the difference spectrum from CO assay, two absorbance readings at 450 nm (ΔA_{450}) and 490 nm (ΔA_{490}) were recorded for calculation of the molar concentration of active P450 in the measured samples. As is shown in the formula below, the extinction coefficient of $91,000 \text{ M}^{-1} \text{ cm}^{-1}$ at 450 nm in the difference spectrum was used to calculate the molar concentration of active P450 (nmol/mL) in that 1 mL cuvette mixture containing 900 μL cell lysis and 100 μL sodium dithionite⁽¹¹¹⁾.

$$C_{(P450 \text{ in cuvette})} = \frac{\Delta A_{450} - \Delta A_{490}}{91000} \times 10^6 \text{ nmol/mL}$$

Therefore, the molar concentration of active P450 (nmol) produced by each litre of cell culture was calculated as below using the $C_{(P450 \text{ in cuvette})}$ obtained above. As was described in 2.1.8.1, 3 mL of concentrated cell lysate was prepared, and 900 μL of which was used in the reduction reaction in the cuvette. The 3 mL cell lysate came from 20 mL cell culture, so the total P450 in 1 L cell culture could be calculated.

$$C_{(P450 \text{ in culture})} = \frac{C_{(P450 \text{ in cuvette})} \times \frac{1000}{900} \times 3}{20 \times 10^{-3}} \text{ nmol/L}$$

2.1.9 Cytochrome c assay

Before conducting cytochrome c assay, stock solutions of 1 mM cytochrome c from the equine heart and 0.5 M NADPH or NADH were prepared using autoclaved dH_2O . Cytochrome c reducing activities of purified reductases were determined in 1.5 mL capacity cuvette in total volumes of 1 mL containing 900 μL of reductases dissolved in 50 mM potassium phosphate, pH 7.5, and 100 μL

cytochrome *c* stock. The blank reaction was also prepared by adding 100 μL potassium phosphate buffer, pH 7.5. Reactions were started by the addition of 1 mM NADPH or NADH, which completed the electron transferring chain to reduce cytochrome *c*. The absorbance at 550 nm was immediately measured, and the mixture was measured at every minute since the reaction until the absorbance stayed constant. The cytochrome *c* reducing activities were then calculated using the extinction coefficient at 550 nm ($21 \text{ mM}^{-1}\text{cm}^{-1}$). Specific activities were given in U mg^{-1} of total protein, where 1 U is defined as the amount of enzyme catalysing the reduction of 1 μmol cytochrome *c* per minute⁽¹⁴⁵⁾. Cytochrome *c* reducing activities of soluble protein fractions obtained from *E. coli* cells lacking expressed heterologous reductase were subtracted from the results.

2.1.10 Analytical methods

2.1.10.1 HPLC

Analytical HPLC: Methods were performed with an HPLC system consisting of a rapid separation pump, an autosampler a column compartment, and a rapid separation diode array detector. All four HPLC modules belong to the DIONEX UltiMate 3000 system that developed by Thermo Scientific.

Method 1: Analytic reverse phase analysis method was used for achiral quantitative analysis. Generally, method 1 was the most used in this project. Separation of peaks was achieved with a HiChrom ACE C18-5 (150 x 4.6 mm) column. A 1 mL/min gradient of H_2O (0.1% TFA) and acetonitrile 90:10 stayed for the first minute and changed to 30:70 over 9 minutes at 30°C (Figure 2.1). Injection volumes were 20 μL . Substrate and product were detected via UV absorbance at 280 nm or 254 nm. Retention times and concentrations were determined by chemically verified standards.

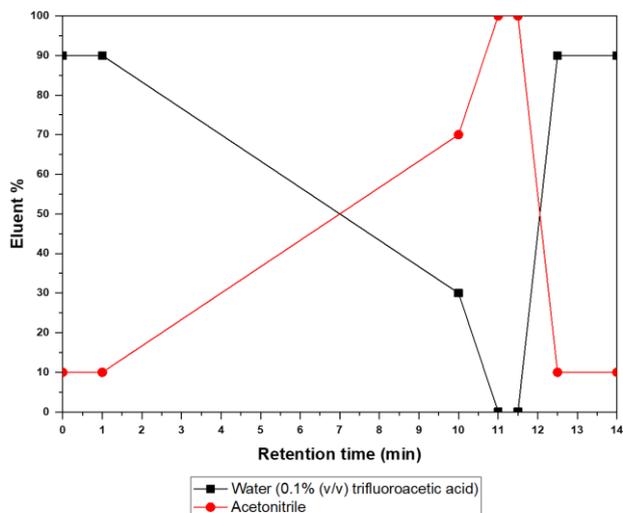


Figure 2. 1. Analytical HPLC method 1. The gradient method has a 14-min duration. The initial ratio of water to acetonitrile is 90:10. The eluent reaches to 100% acetonitrile at 11 min, which intends to elute all remaining analytes. The column oven is set at 30°C.

Method 2: The optimised analytic reverse phase analysis method was used for detecting several single aromatic compounds. The use of columns as well as any other conditions are the same as those of HPLC method 1. A 1 mL/min gradient of H₂O (0.1% TFA) and acetonitrile 70:30 stayed for the first two minutes and changed to 0:100 over 8.5 minutes at 30°C (Figure 2.1). Injection volumes were 20 µL. Substrate and product were detected via UV absorbance at 280 nm or 254 nm. Retention times and concentrations were determined by chemically verified standards.

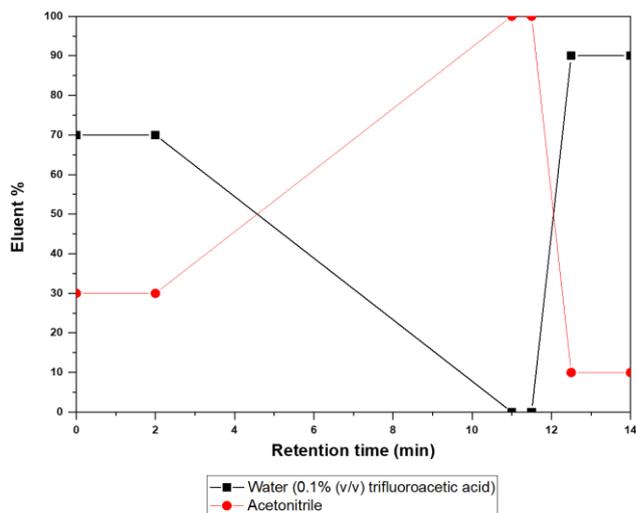


Figure 2. 2. Analytical HPLC method 2. The gradient method has a 14-min duration. The initial ratio of water to acetonitrile is 70:30. The eluent reaches to 100% acetonitrile at 11 min, which intends to elute all remaining analytes. The column oven is set at 30°C.

Method 3: Analytic reverse-phase wash method was used for the column wash. The main mobile phases used were HPLC water (50% v/v) and methanol (50% v/v), and the column was washed for 30 minutes under 0.1 ml/min at 30°C.

Preparative HPLC: Methods were performed with an HPLC system consisting of a DIONEX P580 pump, a DIONEX ASI-100 automated sample injector, a PDA-100 photodiode array detector, and an Agilent C18 reverse-phase column.

Method 4: Semi-preparative HPLC was carried out on a C18 reverse phase column, Agilent ZORBAX 300SB-C18 (250 mm × 9.4 mm, 5 μm). The flow rate was 3 mL/min with detection at 280 nm. Injection volumes varied depending on the separation achieved. The run time was 16 min, and the injection volume was set to 800 μL per run. The column temperature was set at 25°C, eluting with mobile phase A (water with 0.1% TFA) and B (acetonitrile with 0.1% TFA) in a gradient shown in Figure 2.3.

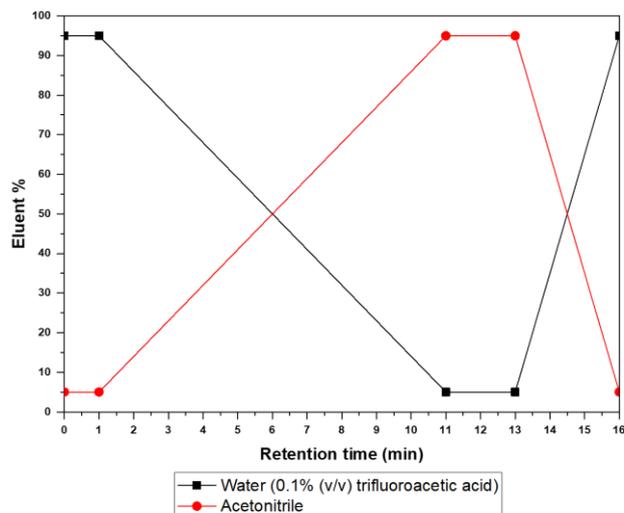


Figure 2.3. Semi-preparative HPLC method. The gradient method has a 16-min duration. The initial ratio of water to acetonitrile is 95:5. The eluent reaches 95% acetonitrile at 11 min, which intends to elute all remaining analytes. The column oven is set at 25°C.

2.1.10.2 LC-MS

Method 5: The molecular masses of new compounds were measured on an Agilent 1100 Series System with a Finnigan LTQ mass spectrometer. An ACE 5 C18 reverse-phase column (50 mm × 2.1 mm, 5 μm) was adopted with a mobile phase of eluent A (H₂O with 0.1% (v/v) formic acid) and eluent B (acetonitrile) over 5 min with a flow rate of 0.2 mL/min. The sample injection volume was 10 μL. Chemical compounds were measured in a positive ion mode, and the operating conditions of the ESI interface were set to a capillary temperature 300°C, capillary voltage 9 V, spray voltage 4 kV, sheath gas 40, auxiliary gas 10, sweep gas 0 arbitrary units. The gradient of eluents was as follows (Figure 2.4).

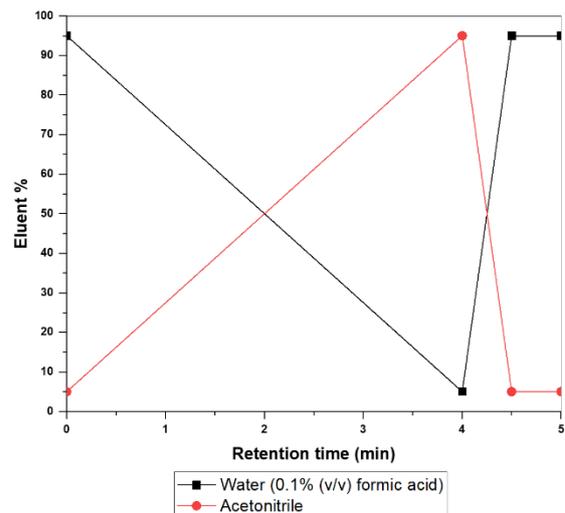


Figure 2.4. Liquid chromatography method. Two mobile phases, water (0.1% (v/v) formic acid) and acetonitrile were used in general liquid chromatography.

2.1.10.3 GC-MS

Gas chromatography-mass spectrometry was the analytical approach for identifying alkane substrates in this project. The GC-MS system consisted of a TRACE 1300 gas chromatograph and a single quadrupole mass spectrometer from Thermo Scientific. The gas chromatography started from volatilizing the samples, and the vaporized samples were propelled by the inert helium and separated through a capillary dimethylpolysiloxane column.

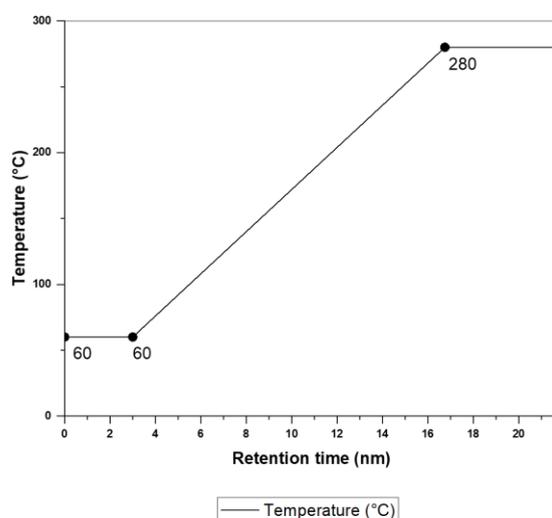


Figure 2.5. The gradient of oven temperature for gas chromatography. The initial column oven was set to be 60°C and gradually climbed to 280°C in 13 minutes.

2.1.10.4 Chemical characterisation (NMR)

NMR spectroscopy: ^1H and ^{13}C NMR spectra were performed with a Bruker Avance 600 machines at 298 K. Coupling constants were measured in Hertz and referenced to the deuterated solvent used. Infrared spectra were recorded on PerkinElmer Spectrum 100 FTIR™ spectrometer.

2.2 Materials and methods for CYP105AB1 characterisation

2.2.1 Enzyme assay for preliminary substrate screening

The enzyme assay was conducted by using purified proteins CYP105AB1, Pdx, and PdR, as well as NADH as a cofactor. All enzyme assays were conducted in 50 mM potassium phosphate buffer (pH 7.5). The purification processes for the three proteins were the same, which were described in section 2.1.7.3. The total volume of each enzyme reaction was 1000 μL , and the final concentrations of each protein components were set to be 0.2 μM of CYP105AB1, 5 μM of Pdx and 1 μM of PdR. The final concentrations of substrates and NADH were set to be 200 μM and 1 mM respectively. All substrates were firstly prepared as 20 mM stock in DMSO. The reaction was prepared in 2 mL sealable Eppendorf tubes and shaken on thermomixer at 700 rpm and 25°C for 24 hours.

The reaction was harvested by adding 100 μL DMSO to the 1000 μL original mixture and then centrifuged for 10 minutes at 13000 rpm. The supernatant was analysed through HPLC for the measurement of residue substrates. Before the analysis of reaction mixture, the substrate standards were normally prepared at 0.1 mM, 0.2 mM, 0.5 mM, or 1 mM, which were used to generate the correlation between peak areas and standard concentrations for each substrate.

2.2.2 Product characterisation

The product characterisation process of unknown chemicals involves product purification and concentration, mass determination by LC-MS, and structure characterisation by NMR. The product was analysed and purified by a C18 reverse-phase column, using semi-preparative HPLC method 5 described in section 2.1.10.2. The isolated peaks were manually collected in 15 mL falcon

tubes, and each peak sample was concentrated through lyophilisation. The final chemical powders were dissolved in 50% water and 50% acetonitrile and prepared for LC-MS. The chemical powders were also dissolved in deuterated methanol and measured for their NMR spectra.

2.2.3 Cell-free bioconversion

2.2.3.1 Bioconversion by purified CYP105AB1 and Pdx-PdR system

The CYP105AB1 was purified and used in enzyme assays along with other purified proteins to reconstruct the reaction environment for the hydroxylation of diclofenac. The reaction was set up as 1000 μL containing 0.2 μM of CYP105AB1, 5 μM of Pdx and 1 μM of PdR, and the initial amount of diclofenac and NADH was set to be 200 μM and 1 mM respectively. The reaction was conducted in 2 mL sealable Eppendorf tubes and shaken on thermomixer at 700 rpm and 25°C. The details of the reaction preparation are shown in Figure 2.6 below. Each reaction was initiated with 2 μL 0.5 M stock NADH, while the negative control was added with 2 μL water. The reaction was stopped and harvested at 2 hours, 4 hours, 20 hours, 24 hours, and 48 hours. Each sample was also prepared in duplicates, and the replicated reactions were conducted in parallel. Therefore, the final calculated DIC concentration should be the average results. The harvest stage involved the addition of 100 μL DMSO and centrifugation at 13000 rpm for 10 min. Only 20 μL of the final harvested mixture was sent for HPLC analysis.

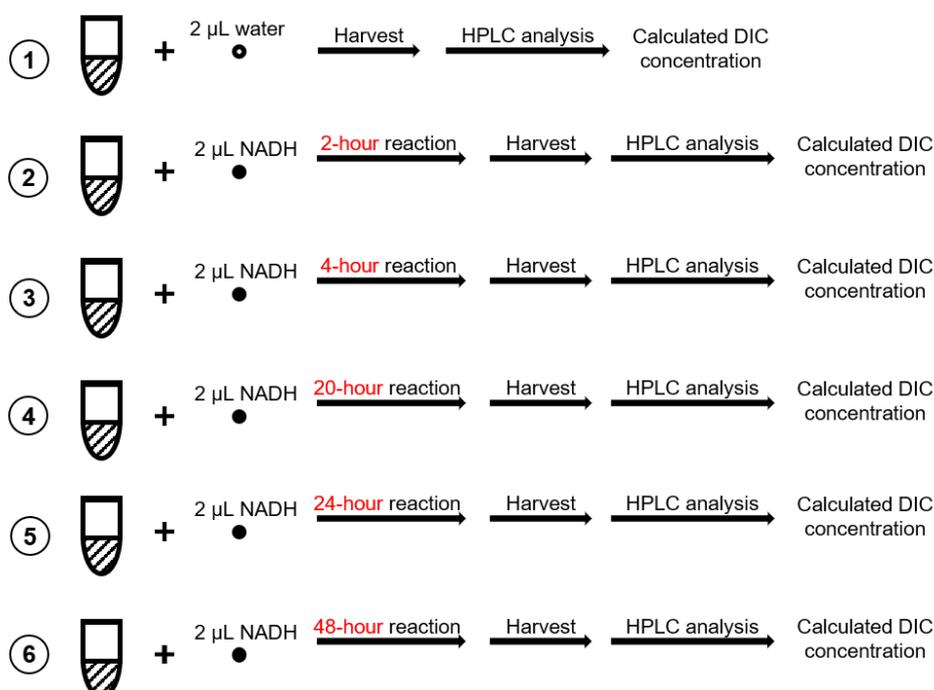


Figure 2.6. The sample preparation for the reaction of diclofenac by CYP105AB1. The total volume of the reaction was 1000 μL , and the NADH was added as the last component to start the reaction. The reaction mixture shown in sample 1 to 6 above were prepared in the same way at the same time, containing 0.2 μM CYP105AB1, 5 μM Pdx, 1 μM PdR and 200 μM diclofenac in 50 mM potassium phosphate buffer (pH 7.5). Sample 1 was added with water and immediately analysed by HPLC to measure the initial (time = 0) diclofenac concentration. Sample 2 to 6 were separately added with NADH stock, and the NADH stock was prepared as 0.5 M. The reactions were allowed for different periods such as 2 hours, 4 hours, 20 hours, 24 hours, and 48 hours before HPLC analysis. Each reaction was repeated three times under the same conditions to acquire the average DIC concentration. The HPLC method was HPLC method 1 described in section 2.1.10.1. The calculation of remaining diclofenac was based on the established mathematical correlation between diclofenac peak area and standard diclofenac concentrations.

2.2.3.2 Bioconversion by cell lysate of CYP105AB1 and partner proteins

Besides, the cell lysate containing CYP105AB1 and electron transferring proteins was also to catalyse the reaction of diclofenac. Because two different operons, including *cyp105ab1-fd2-fdr* (pQR2281) and *cyp105ab1-pdx-pdr* (pQR2293), have been used for the expression, so two different cell lysates were prepared for bioconversion of diclofenac. The sample preparation process is shown in Figure 2.7 below.

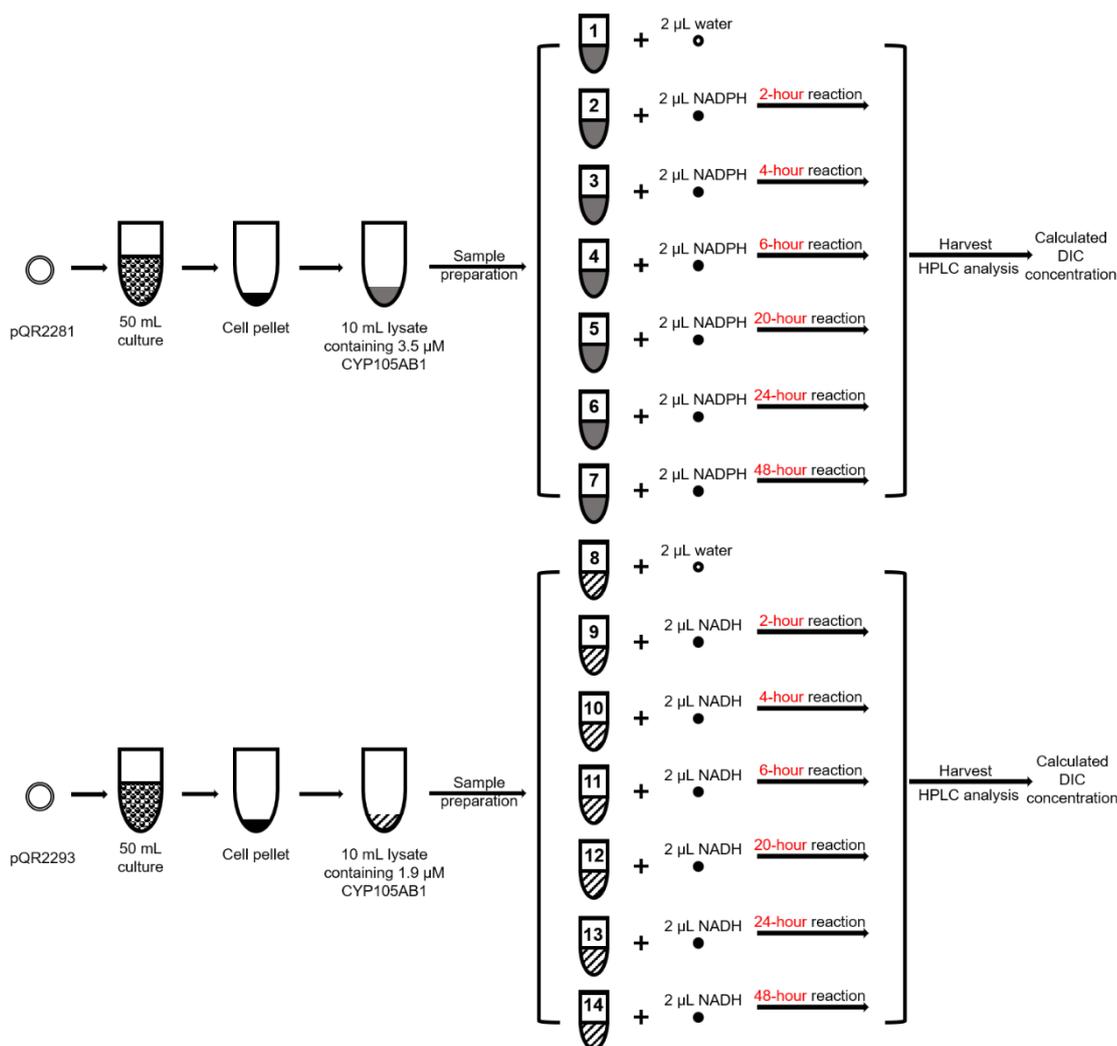


Figure 2.7. The sample preparation process for bioconversion of diclofenac using different CYP105AB1 lysate. The enzyme expression process was described in section 2.1.7.1, and the cell lysate was prepared through sonication (section 2.1.7.2). The total volume of each reaction was set to be 1000 μL. The reaction mixture shown in sample 1 to 14 above was prepared at the same time, containing 1 μM CYP105AB1 and 200 μM diclofenac in 50 mM potassium phosphate buffer (pH 7.5). Samples 1 and 8 were added with water, and immediately analysed by HPLC to measure the initial (time = 0) diclofenac concentration. Samples 2 to 7 were separately added with NADPH stock (0.5 M), while samples 9 to 14 were separately added with NADH stock (0.5 M). The reactions were allowed for different periods such as 2 hours, 4 hours, 6 hours, 20 hours, 24 hours, and 48 hours before HPLC analysis. Each reaction was repeated three times under the same conditions to acquire the average DIC concentration. The HPLC method was HPLC method 1 described in section 2.1.10.1. The calculation of remaining diclofenac was based on the established mathematical correlation between diclofenac peak area and standard diclofenac concentrations.

Every 50 mL of cell culture was centrifuged, and the pellet was lysed in 10 mL of 50 mM potassium phosphate buffer (pH 7.5) through sonication. The clarified cell lysate was filtered and measured for the concentration of CYP105AB1. The molar concentration of CYP105AB1 was determined by carbon monoxide assay. These stock cell lysates were then used for bioconversion. The bioconversion was initiated by the addition of cofactors (NADH or NADPH).

After the addition of cofactors to the reaction system, the biotransformation was monitored at 0 hour, 2 hours, 4 hours, 6 hours, 20 hours, 24 hours, and 48 hours. The 1 mL reaction was prepared in a 2 mL Eppendorf tube, which was conducted with cell lysates, 0.2 mM diclofenac and 1 mM NAD(P)H at 25°C. The final concentration of CYP105AB1 was set to be 1 µM. The harvest stage involved the addition of 100 µL DMSO and centrifugation at 13000 rpm for 10 min. Only 20 µL of the final harvested mixture was sent for HPLC analysis (method 1) to measure the remaining amount of diclofenac.

2.2.4 Site-directed mutagenesis

The Phusion Site-Directed Mutagenesis Kit (Thermo Scientific) was used to introduce point mutations on CYP105AB1. The mutagenesis process comprises four standard steps, which include PCR amplification of target plasmids using 5' phosphorylated mutation primers, digestion of template plasmids with *DpnI*, circularization of PCR products by T4 DNA ligase, and transformation of the ligation mixture with *E.coli* BL21(DE3) competent cells.

The PCR conditions used were almost the same to the ones shown in Table 2.4, except that the cycle was reduced to 25. The primers designed to have 24-30 bp and all introduced point mutations were placed on the forward primers. The template plasmid was pQR2281, which contained *cyp105ab1-fd2-fdr* operon for CYP105AB1 expression. The point mutations were introduced on the *cyp105ab1* gene through PCR.

2.3 Materials and methods for *Streptomyces rishiriensis* genome mining

2.3.1 Custom tBLASTn

Genome mining was conducted by using Geneious 11 (Biomatters Ltd). The original genome data of *S.rishiriensis* consisted of 2970 unannotated contigs, which was generated by Illumina sequencing (GATC Biotech, Eurofins Genomics). In the *S.rishiriensis* genome dataset, there were 247 contigs with the size over 900 bp, which were grouped as the DNA library for custom BLAST. The conserved protein sequence of the heme-binding motif for cytochrome P450 was used as the query sequence. Therefore, the tBLASTn approach was selected using substitution matrix BLOSUM62 as well as the empirical gap costs pattern (11, 1). Considering the library is small and finite, the e-value cut off was set at 10, which aims to find all similar sequences to the heme-binding motif.

2.3.2 Co-expression of CYPs and molecular chaperones in *E.coli*.

An effective method of constructing the co-expression system of chaperone and CYPs from *S.rishiriensis* involves two steps. The first step is preparing the chemically competent cells containing the plasmids of targeted *cyp* genes, and the second step is the transformation of these competent cells with each chaperone plasmid. The resistant markers for targeted CYP plasmid and chaperone plasmid are ampicillin and chloramphenicol respectively, so agar plates containing both antibiotics are prepared for colony selection.

To perform the co-expression, 100 mL TB medium containing 20 µg/mL chloramphenicol and 50 µg/mL ampicillin was prepared for inoculation in 500 mL baffled conical shake flask. The final concentration of IPTG was set to be 0.1 mM for induction of CYP, and the final concentrations of L-arabinose and tetracycline were set to be 2 mg/mL and 5 ng/mL for the induction of chaperone proteins. The L-arabinose or tetracycline was added into a fresh medium right before inoculation with overnight culture. After the cell culture reached 0.4 of OD600, it was added with FeCl₃ (final concentration 0.5 mM) and 5-aminolevulinic acid (final concentration 1 mM) to be incubated at 25°C, whilst rotating at 180 rpm for 20 minutes. After that, the IPTG was added at a final concentration of 0.1 mM, and the post-IPTG induction was controlled to be 24 hours at 25°C.

2.3.3 Preliminary substrate screening

The enzyme assay was conducted by using purified Sri_CYP03 or Sri_CYP13 as the catalysts, Pdx, and PdR as the electron transfer partners, as well as NADH as the cofactor. The purification process for the three proteins was achieved using a nickel column, which was described in section 2.1.7.3. Each assay was conducted in 50 mM potassium phosphate buffer (pH 7.5). The total volume of each enzyme reaction was 1000 μ L, and the final concentrations of each protein components were set to be 0.2 μ M of CYPs from *S.rishiriensis*, 5 μ M of Pdx and 1 μ M of PdR. All substrates were firstly prepared as 20 mM stock in DMSO, and the cofactor NADH was prepared as 0.5 M stock. Five different substrates were being used in the preliminary screening, which included apigenin (APG), chrysin (CHY), 7-ethoxycoumarin (EC), diclofenac (DIC) and 1,2,3,4-tetrahydroisoquinoline (THIQ). Considering the different solubility of each substrate in the 50 mM potassium phosphate reaction environment, the initial concentrations for each substrate were set differently, which led to the initial reaction concentrations of 0.05 mM APG, 0.05 mM CHY, 0.2 mM EC, 0.1 DIC, and 1 mM THIQ in each enzyme assay. The initial NADH concentration was set to be 0.5 mM, which was sufficient enough to provide electrons for the reaction system if the substrates could be accepted by cytochrome P450 in the enzyme assay. The reaction sample was prepared in 2 mL sealable Eppendorf tubes and shake on thermomixer at 700 rpm and 25°C for 24 hours. The negative control reaction was prepared without NADH, but all other components described in the reaction sample.

After 24-hour incubation, the 100 μ L DMSO was added into a 1000 μ L reaction mixture in the 2 mL reaction tube, which aims to increase the solubility of any potential hydrophobic substances. The reaction mixture was heated at 60°C for 10 minutes to stop the reactions and then centrifuged for 10 minutes at 13000 rpm. The supernatant was analysed through HPLC using method 1 (described in section 2.1.10.1). The chromatograms of reactions and negative controls were compared to establish any potential consumption of the substrates, as well as the production of any new peaks. Before conducting the reactions or negative controls, the retention time and peak area for each substrate standard were

established, which were used to calculate the standard curve of peak area to substrate concentrations.

2.4 Materials and methods for Drain metagenome mining

2.4.1 Custom tBLASTn

Genome mining was conducted by using Geneious 11 (Biomatters Ltd). The original metagenome data consisted of 69962 unannotated contigs. In the drain metagenome data, there were 53642 contigs with the size of over 800 bp, which were grouped as the DNA library database for custom BLAST. The conserved protein sequence of the heme-binding motif for cytochrome P450 was used as the query. Therefore, the tBLASTn approach was selected using substitution matrix BLOSUM62 as well as the empirical gap costs pattern (11, 1). The e-value cut-off was set at 10.

2.4.2 Preliminary substrate screening

The enzyme assay was conducted by using purified proteins Dmg_CYP08 or Dmg_CYP09, Pdx, and PdR, as well as NADH as a cofactor. The purification process for the three proteins was using a nickel column for harvesting histidine-tagged targets, which were described in section 2.1.7.3. Each enzyme assay was conducted in 50 mM potassium phosphate buffer (pH 7.5). The total volume of each enzyme reaction was 1000 μ L, and the final concentrations of each protein components were set to be 0.2 μ M of CYP105AB1, 5 μ M of Pdx and 1 μ M of PdR. The final concentrations of substrates and NADH were set to be 500 μ M and 1 mM respectively. All alkane substrates were firstly prepared as 200 mM stock in ethanol, and the cofactor NADH was prepared as 0.5 M stock. The reaction sample was prepared in 2 mL sealable Eppendorf tubes and shake on thermomixer at 700 rpm and 25°C for 24 hours. The negative control reaction was prepared without NADH, but all other components described in the reaction sample.

After 24-hour incubation, the 1000 μ L ethyl acetate was added into a 1000 μ L reaction mixture in the 2 mL reaction tube. Because the aqueous sample could not be loaded directly into the GC oven, an extraction process was conducted first. The total mixture was incubated at 300 rpm overnight for substrate extraction

into the organic layer. The mixture of reaction and ethyl acetate was centrifuged for 10 minutes. After that, 20 μL of the organic phase was used and diluted with 180 μL ethyl acetate before loading samples through GCMS analysis. The GC chromatograms of the reaction sample and the negative control sample were compared and analysed to determine the potential decrease or consumption of substrates, which eventually revealed the substrate could be potentially accepted by these P450s identified in the drain metagenome.

The retention time and peak area for each substrate standard were established before analysing the reaction and negative control. After acquiring the 200 mM substrate stock in ethanol, the 500 μM substrate were prepared in both ethyl acetate and 50 mM potassium phosphate buffer. The alkanes dissolved in ethyl acetate were diluted with ethyl acetate in 10 times and sent for GCMS analysis. For the alkanes dissolved in phosphate buffer, the same volume of ethyl acetate was used to incubate with the mixture of substrate and phosphate buffer for alkane extraction, which was placed at 300 rpm overnight. The organic layer with extracted substrates was diluted with ethyl acetate in 10 times and sent for GCMS analysis. By comparing the peak signals of alkanes dissolved in ethyl acetate and alkanes dissolved in phosphate buffer on GC chromatograms, the extraction efficiency of each substrate could be estimated.

3 Chapter 3: Building genetic constructs for heterologous expression of self-sufficient cytochrome P450 from *Bacillus licheniformis*

3.1 Background

As described in the introduction (section 1.2.4), there are various classes of cytochrome P450 in nature. The self-sufficient cytochrome P450 (Class VII cytochrome P450) catches much attention because it is unique and does not require additional reductases to assist on the delivery of electrons. Therefore, it is less complex on identification, cloning as well as characterization processes. The most widely studied self-sufficient cytochrome P450 is P450-BM3 (CYP102A1) from *Bacillus megaterium*, which is a fatty acid hydroxylase P450 fused to a mammalian-like diflavin NADPH-P450 reductase in a single polypeptide^(146, 147). Since the discovery, there have been extensive studies on identifying and characterizing homologous of P450-BM3, aiming to gain more insights into this class of P450. For example, there have been 14 close homologues of P450-BM3 being identified so far, which are all classified into the same subclass of CYP102A⁽¹⁴⁸⁾. It is considered that there are more self-sufficient P450s in various microbes, and more work on the identification of these P450s is appreciated and beneficial for the growing toolbox of biocatalysts.

As is described above, self-sufficient P450s do not need the assistance of other redox partners to conduct reactions, which makes them less complex on identification and further characterisation. Therefore, it is reasonable and necessary to conduct the research on self-sufficient P450, which provides a basic understanding of P450 biology as well as P450 characterisation techniques. Before this project, the genomic DNA of three microbes including *Bacillus licheniformis* ATCC 14580, *Streptomyces avermitilis* MA-4680 and *Saccharopolyspora erythraea* NRRL2338 had been extracted and stored at Ward Lab. At the beginning of this project, a literature review was carried out to find relevant studies on self-sufficient P450s in the three microbes mentioned above, which led to the previous studies on CYP102A7⁽¹²⁴⁾ and CYP102D1⁽¹⁴⁹⁾. CYP102A7 and CYP102D1 are Class VII P450s identified from *Bacillus licheniformis* ATCC 14580 and *Streptomyces avermitilis* MA-4680 respectively. Apart from CYP102A7 and CYP102D1, there had not been other Class VII P450s from the three microbes being identified or characterized.

The obvious questions at that time included that whether CYP102A7 and CYP102D1 were the only Class VII P450s in their genomes, and whether there was Class VII P450 in *Saccharopolyspora erythraea* NRRL2338. Fortunately, the complete genome sequences of *Bacillus licheniformis* ATCC 14580⁽¹⁵⁰⁾, *Streptomyces avermitilis* MA-4680⁽¹⁰⁹⁾, and *Saccharopolyspora erythraea* NRRL2338⁽¹¹⁰⁾ were fully determined and recorded. The complete genome sequences were recorded at ENA (European Nucleotide Archive) (Appendix A, Table A2). Therefore, a simple alignment search was conducted using the complete protein sequence of P450-BM3 as a query against the genomic sequences of three microbes, which intended to find other Class VII P450s. The alignment results showed that three P450s were being identified as homologues of P450-BM3, and details of the translated sequences are shown in Table 3.1.

Table 3.1. The complete protein sequences of Class VII P450s identified in *Bacillus licheniformis* ATCC 14580, *Streptomyces avermitilis* MA-4680, and *Saccharopolyspora erythraea* NRRL2338. The tBLASTn program was used and the protein sequence of P450-BM3 served as the query. CYP102A7 and CYP102D1 were identified and characterized in previous research. CYP102G2 was newly identified in this project.

CYP	Organism	Amino acid sequence
CYP102A7	<i>Bacillus licheniformis</i> ATCC 14580	MNKLGDGPIPKTYGPLGNLPLLDKNRVSQSLWVKIADEMGPIQFQKFADAI GVFVSSHELKVEVSESRFDKMNMGKLLKREFSGDGLFTSWTEEPNWRK AHNILLPSFSQKAMKGYHPMMQDIQVLIQKWSRLNQDESIDVDDMTRL LDTIGLGGFNFRNSFYREGQHPFIESMVRGLSEAMRQTKRFPLODKLM IQTKRFRNSDVESMFLVDRIADRKAQAESESGNDLLSLMLHAKDPETGE KLDDENIRYQIITFLIAGHETTSGLLSFAIYLLLKHPDKLKKAYEEADR LTDVPVPSYKQVQQLKYIRMLNESIRLWPTAPAFSLYAKEETVIGGKYL PKGQSVTVLIPKLHRDQSVWGEDAEAFRPERFEQMDSPAHAYKPFNGGQ RACIGMQFALHEATLVGLMILQYFDELDHANYQLKIKESLTKPDGFTIR VRPRKKEAMTAMPGAQPEENGRQEERPSAPAAENTHGTPLLVLVYGSNLGT AEEIAKELAAEEAREQGFSRTEALDQYAGAIPAEGAVIIVTASYNGNPPD CAKEFVNWLEHDQTDLLRQVYAVFGCGNRSWASTYQRIPLRIDSVLEKK GAQRLHKLGEAGDDFEGQFESWKYDLWPLLRTFEFLAEPEPNQETDR QALSVEFVNAPAAASPLAKAYQVFTAKISANRELQCEKSGRSTRHIEISL EGAAYQEGDHLGVLPQNSEVLIGRVFQRFLNGNEQILSGRNQASHLPL ERPVHVKDLFQHCVELQEPATRAQIRELAHAHTVCPHQRELEDDLLKDDVY KDQVLNKRLLTMDLLEQYPACELPFARFLALLPPLKPRYYSISSSPQLNP RQTSITVSVVSGPALSGRGHYKVASNYLAGLEPGDAISCFIREPQSGFR LPEDPETVIMVGPGTGIAPYRGLQARRIQRDAGVKLGEAHLVYFCRRP NEDFLYRDELEQAEDGIVHLHTAFSRLEGRPKTYVQDILLREDAALLIHL LNEGGRLVYCGDGRMAPAVEQALCEAYRIVQGASREESQSWLSALLEEG RYAKDWDGGVSVQHNVKADCIART
CYP102D1	<i>Streptomyces avermitilis</i> MA-4680	MTTQPETDLRPIRSPRGVPLFGHTPQIPSTNPVEYFGKLSKQFPEGLYGM EIAGIEQVFWDPDLVAEVCDETRFFKQIDKTPLAHVRDYAGALFTAHQ HEEEWGMHRVLLPVFSQRAMKGYFGQMLEIAQNLVKGWERKEGQPVNIT DDYTRLTDLTIALSGFGYRFDSEFAKEDLHPFLNALLQALVESLRRSQELP VMTKMRKADKKYRENIRLMDLVENVIKERREGKGTGEDDLLGLMLEAT DPETGKGLDDDNVRDQVVTFLIAGHETTSGLLSFATYSLMRNPHILAQY AEVDRLPGDTPDYDTIMQMDVIPRILEETLRLWAPIPMIGKSPLEDTV IGGCYGLKKGARVNILEGPLHHPKAWERPEEFDINRWLPENRVNHHPHA YKPFNGVRACIGRQFALTEARLALALVLQKFKFADTDDYKMDVKEALTR KPGGFELNVRARQEHERTVFGAADLQTDQTAQAAVSGVGMVLTVAYGSS LGSCEDLARTIADRGERSGFGTTLVGLDELGDNLPTGELLVVVASSYNGK APDNAQRFDLLAAGLPEGSLSNVRFALLGAGNTQWVATYQGFPKRIEAG LLAAGATRVIERGIADAAGDFDGMATRWMMDTLWTTLAEEYAADTSETTGP RFEVQLLTAEAVRPAIVSEQAYPLTVVANEELVSDATGLWDFSIEMPRPA AKSITIELPDGVTYDTGNHLAVFAKNEPVLVNRALARLGVDRDQVLRLDQ PGGGRTHLPVGTPTVTTGLLTFEFVELQDVATRSQIQELAHTQCPWTRPQ LQAYTADTAEEERYQKEILGKRVSVLNLLERFPAVELPLAVFLEMMGPI RPRFYSISSSPLANPRHVRLTVGLLEGPALSGDGRYRGTCSSYIAGLES DVFYGYVRVPSPTFAPPADPATPLLLIGPGTGIAPLRGFLERAHQHAHG TQVLSQVFGCRHPEHDYFYRQEMQDWEQAGIAQVHTAFSAVTGHPARF VQDAIVGAADTVWQAIQDGAYVYVCGDGRMAPAVREALAAYRKHGTGSD DEAAQQWLAQLEADERYQQDVFA
CYP102G2	<i>Saccharopolyspora erythraea</i> NRRL2338	MTQTPLHHDDVPVADVSGTGLTATPTQQAMELARRHGPVFRRTREFQSL LVSDVDLVAELSDEQRFKAVGPALENVREFAADGLFTAYNDEFNWAKAH DILMPAFALGSMRTYHPVMLRVARRLLDSWDRAAAAASAPVDVDDMTRMT LDTIGLAGFGDFGSGFRAEPHPPVGMVRCLDWSMTRLSRVPGTDHSE DEAFRADARYLASVVDVINTRAEGDTSGEDLLGLMLGARHPADGHTL AANIRNQIITFLIAGHETTSGLTSFALHYLAKNPTVLRVQREADELWGD SPDPEPSFEDIGRLTYTRQVLNETLRLWPSAPAFGRQARHDTVLGGRI RAGEAAAFLIPMLHRSPVRGDNPELDFPARFAPEAEARGPHAFKPFGTG ERACIGRQFALDEATMVLAMLAHRYRLVDHAGYRLKVKETLTKPEGLTL AVRARTAADRVTNRLALPVGLPSAAPGEPADAARRPGRVLPGTGLLLHG SNYGTCDRFAAQLALAAGELGCDTAVAPLDEYAGNLPDRPVIVVAASYN GRPTDDAVSFRWLDEAEPGAADGVDFAVLGVGDRNVAATYQHVPTRIDA RLAELGGTRILERGEADASGDLAGAVRRFSAALETALLERSGDPDAVAAA PEGDGPAYTVSEVTGGALDSLAAHGMVEMTVTEVADLTPDYPRTKRFV RLALPEGTAYRTADHLAVLPVHDAALVERAAGVLGVDLDTVDIRAKRPG RLTFDRSLTVRELLSHHVELQDPPTPDGLDALAALNCPPEERAALRGLAE EARSGTADHRTLWDLIEDHPALRDALSWSALLELLPATRPRQYSVSSSPA VDPRHVDMVSVLRAPARSGRGEFRGAGSRHLSEVRPGDVLARVQPCRE DFRVAPDEPLIMVAAGTGLAPFRGVIADRRRERVANGARQAPALCYFGCDA PDADYLHSAELRAAESAGAVAMRPAFNEAPVGGQCFVQHRIAEEAGEVWA LLESGARVLVCGDGRHMAPGVREAFRGIYRERTPGADDASAHEWLQAMIA GGCYVEDVYAG

After identifying the three Class VII P450s, the next stage involved cloning of these genes and producing enzymes in common expression cell lines. Despite many efforts devoted to the cloning of these P450 genes, only CYP102A7 was successfully cloned and expressed, and subjected to further enzyme reactions.

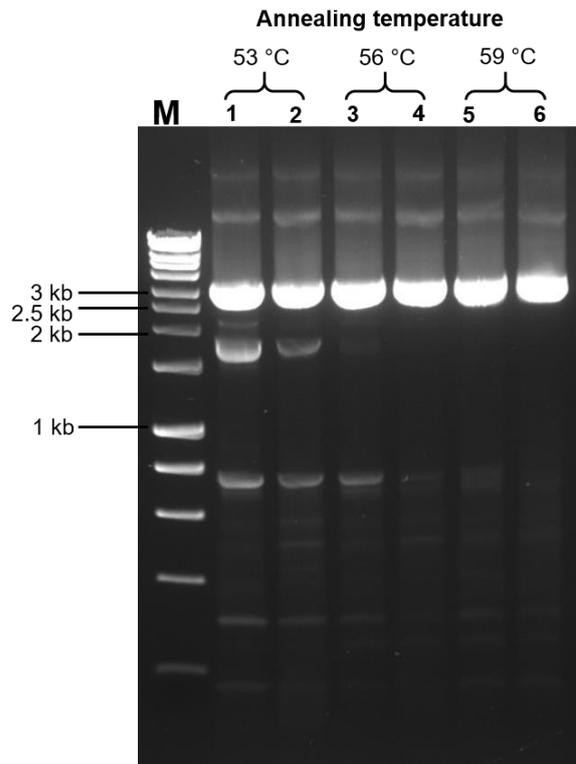


Figure 3.1. Genomic PCR for producing CYP102A7 gene fragments. The genomic DNA of *Bacillus licheniformis* was used as a template, while various annealing temperatures were tested with the designed primers. The lane M is displaying 10 μ L of 1kb Bioline HyperLadder, and the 3 kb band is labeled as above. The expected fragment length of PCR product is 3287 bp, so the brightest bands around 3kb should be the correct PCR products. As is shown in lane 6 on the gel, the PCR had less non-specific amplification at 59°C, which was used for further purification.

The purified DNA fragments containing the *cyp102a7* gene were assembled onto the pET29a backbone to construct the expression plasmid for the production of 6-histidine tagged CYP102A7. As is shown in Figure 3.2, the extended sequences on both ends of the purified PCR products were designed to be complementary to the two ends of a linearized pET29a plasmid. The linearized pET29a was generated by double digestion using restriction enzymes *NdeI* and *XhoI*. A Gibson assembly kit was used to integrate DNA fragments containing the *cyp102a7* gene into the linearized pET29a backbone at specific cloning sites (Figure 3.2) (method described in section 2.1.6.4).

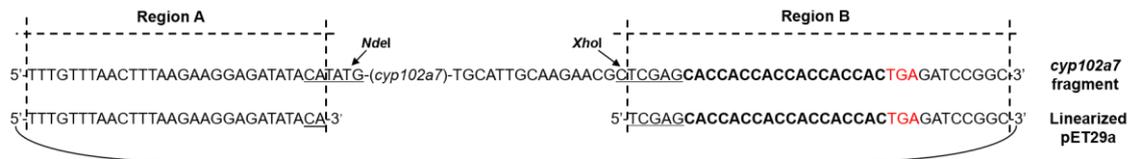


Figure 3.2. The approach of assembling expression plasmid containing the *cyp102a7* gene. The PCR fragments from Figure 3.1 were designed to contain not only the *cyp102a7* gene but also the extended region A and B showing above. The linearized pET29a was acquired by digestion with *NdeI* and *XhoI*. Gibson assembly was used in this case to integrate *cyp102a7* into pET29a. The sequences above are all aligned in a 5' to 3' direction, and the digestion sites are underlined. The bold sequences indicate the histidine tags, and the red sequences show the location of the stop codons. The ligation method was not used here due to an existing *NdeI* site within the *cyp102a7* gene, and the *NdeI* (5'-CATATG-3') site is also used as the beginning of the cloning site on pET29a. The length of complimentary region A and B were designed to be 33 bp and 35 bp respectively, which worked efficiently for the NEB Gibson assembly kit.

The assembled expression plasmids were transformed into *E.coli* BL21 (DE3). The plasmid integrity was evaluated by plasmid sequencing using T7 and T7 term primers. The *E.coli* BL21 (DE3) containing the assembled plasmid was cultivated and induced by IPTG for CYP102A7 expression, and CYP102A7 was purified by Ni-NTA column. As is shown in Figure 3.3(B) lane E, the eluted CYP102A7 was indicated in the red circle. Compared to the marker lane, the band of CYP102A7 was approximately located at 120 kDa, which matched to the calculated molecular weight of CYP102A7.

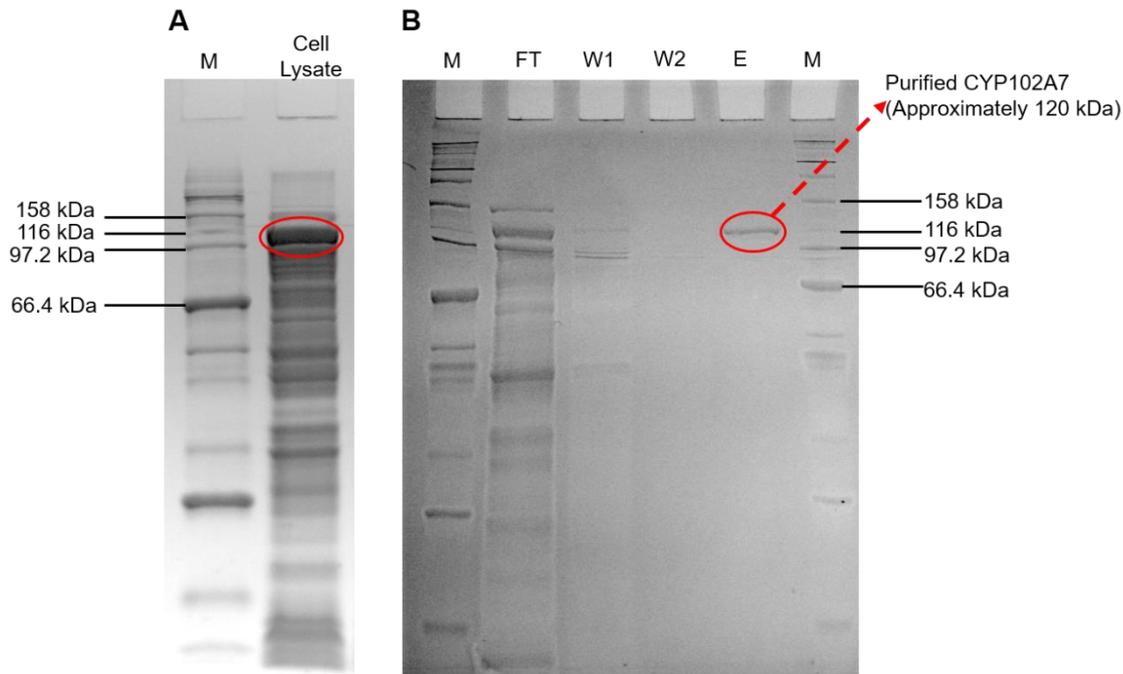


Figure 3.3. SDS-PAGE analysis and protein purification of CYP102A7. (A) The SDS-PAGE analysis of soluble fractions of the cell lysate contained CYP102A7. A relatively thick band at approximately 116 kDa was shown. (B) The cells were resuspended with lysis buffer and lysed through sonication. After applying the cell extracts through the nickel column, four fractions, including flow-through (FW), first wash (W1), second wash (W2), and elution (E), were collected. The elution lane (lane E on the gel) had a band around 116kDa as well, which indicated the purified CYP102A7.

3.3 Carbon monoxide (CO) assay

It was discovered that the presence of distinct solet peaks at 420 nm and 450 nm when reduced cytochrome P450s were bound with carbon monoxide. The discovery is now used to assay for the presence and quantification of active cytochrome P450. Therefore, a carbon monoxide assay is normally conducted to determine the production of active cytochrome P450 within cell lysate before any SDS-PAGE analysis of the purified cytochrome P450. Generally, a difference spectrum is calculated by subtracting the spectrum of reduced P450 from the spectrum of CO binding complex, which should reveal the distinctive peak at 450 nm if there are detectable amounts of active cytochrome P450 produced. In this project, a customized CO assay was performed. Firstly, 20 mL and 1 mL of *E.coli* cell culture were respectively collected and centrifuged for cell pellets. The 20-

mL cell pellet was resuspended with 3 mL of CO assay lysis buffer (section 2.1.8.1), and the clarified cell lysate was used for CO assay. The other 1 mL cell pellet was dried in an oven overnight, and the weight of the dry cell pellet (dry cell weight, dcw) was measured to 0.003497 gram.

There were generally two different approaches to conducting carbon monoxide assay being used before. The first approach is to reduce 900 μ L P450 cell lysate with 100 μ L $\text{Na}_2\text{S}_2\text{O}_4$ solution before binding with CO, while the second approach is to pre-mix CO with 900 μ L P450 cell lysate before adding 100 μ L $\text{Na}_2\text{S}_2\text{O}_4$ solution for reduction. Both approaches were applied here for the characterisation of the optical property of CYP102A7. In Figure 3.4 (A), the difference spectrum was generated by applying the first approach. An absorbance reading of -0.003 at 450 nm on the difference spectrum was recorded, which indicated a slight difference between reduced CYP and CO bound CYP. From another point of view, an absorbance reading of 0.02 at 450 nm on the difference spectrum was observed when using the second approach to conduct the CO assay (Figure 3.4 (B)). What can be concluded from the comparison between (A) and (B) in Figure 3.4 is that mixing CO into P450 cell lysate before reducing with $\text{Na}_2\text{S}_2\text{O}_4$ may lead to a better performance of CO assay. Therefore, the second approach of conducting CO assay was used throughout the entire project.

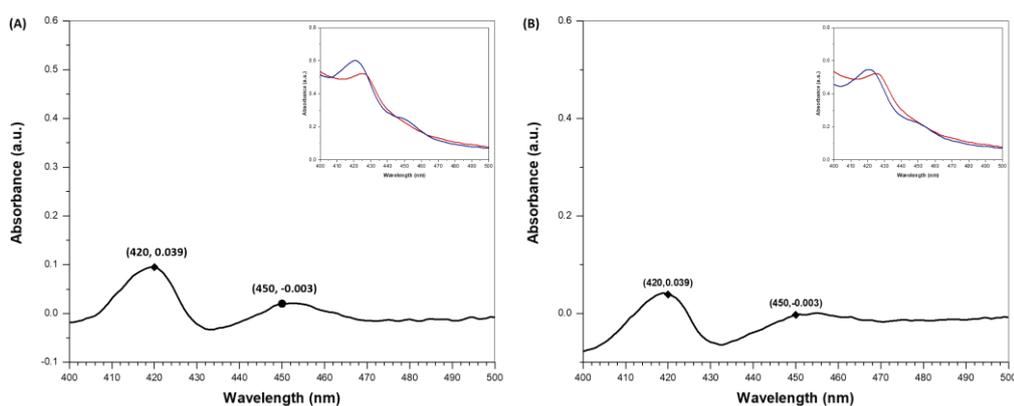


Figure 3.4. Carbon monoxide difference spectra of CYP102A7. Two approaches were used to conduct a carbon monoxide assay. (A) The difference spectrum was generated by applying the first approach that performs CO binding after $\text{Na}_2\text{S}_2\text{O}_4$ reduction. Two peaks at 420 nm and 450 nm were observed. (B) The difference

spectrum was generated by applying the second spectrum that mixes CO within P450 cell lysate before reducing with $\text{Na}_2\text{S}_2\text{O}_4$. Two peaks at 420 nm and 450 nm were recorded as well. The difference absorbance reading at 450 nm was used for further calculation of the concentration of active P450 produced by the batch of cell culture. The insert graphs are of the non-subtracted spectra.

As is described in section 2.1.8.2, the formula for calculating the molar concentration of active P450 in the sample mixture is applied. With the readings of ΔA_{450} (0.02) and ΔA_{490} (-0.015) observed in Figure 3.4 (B), the molar concentration of active CYP102A7 is 0.385 nmol/mL of the sample mixture. In Table 3.2, it shows the process of calculation that leads to the molar amount of active P450 produced per litre of cell culture, which is 64.167 nmol/L cell culture.

Table 3.2. The process of calculating the molar amount of CYP102A7 produced per litre of cell culture. The molar concentration of active cytochrome P450 per litre of cell culture was used to evaluate how much CYP102A7 was produced in *E.coli* BL21 (DE3) culture after 24-hour IPTG induction.

	Calculation	Unit
Concentration of CYP102A7 in 1 mL sample mixture for CO assay	$(0.02 - (-0.015))/0.091 = 0.385$	nmol/mL
Concentration of CYP102A7 in 3 mL cell lysis	$0.385 * 10/9 = 0.428$	nmol/mL
Molar amount of CYP102A7 in 20 mL cell culture	$0.428 * 3 = 1.284$	nmol
Molar amount of CYP102A7 in 1 L cell culture	$1.284 * 50 = 64.2$	nmol

Apart from the molar concentration of active CYP102A7, the amount of active CYP102A7 produced per gram of dry cell weight was also calculated to evaluate the expression level. The total molar amount of CYP102A7 produced by 1 L of cell culture was 64.2 nmol, and the molecular weight of 6-histidine tagged CYP102A7 is 122020.5 g/mol. Therefore, the amount of CYP102A7 produced by 1 mL of cell culture was 7.83 μg . The weight of 1 mL dry cell (dcw) was measured to be 3.5 mg, so the expression level of CYP102A7 was 2.2 mg/g of dry cell weight.

In previous research⁽¹²⁴⁾, the yield of CYP102A7 in *E.coli* BL21 (DE3) reached around 400 nmol/L cell culture after optimisation. The previous yield is much higher than the current 64 nmol/L in my study. A brief gap analysis between previous and current expression procedure indicates three main differences (Table 3.3). As is shown in Table 3.3, the expression vectors of pET28a and pET29a are essentially the same, which may have less influence on the expression level of CYP102A7. For the other two differences, IPTG concentrations and expression time might be key to affect expression level. In the current study, the focus is not to optimize expression conditions to achieve higher expression levels, but to successfully produce CYP102A7 for further substrate screening. All CYP102A7 expressed in *E.coli* will be purified and concentrated for further usage, so the current 64 mol/L of CYP102A7 expression level is acceptable.

Table 3.3. The main differences between the previous and current expression of CYP102A7 in *E.coli* BL21 (DE3).

	Previous Research in 2008	Current Study
Expression vector	pET28a	pET29a
IPTG concentration	0.25 mM	1 mM
Expression time	16 h	24 h

3.4 Preliminary substrate screening

The CYP102A7, which was purified through a Ni-NTA system, was concentrated to prepare a stock solution in 50 mM potassium phosphate buffer, pH 7.5. The CYP102A7 was then used for substrate screening against various fatty acids and terpenes (Appendix A, Table A1). A 100 μ L reaction system, consisting of 0.1 μ M CYP102A7, 200 μ M substrates, 150 μ M NADPH and 50 mM potassium buffer, pH 7.5, was constructed to define potential substrate pool for CYP102A7. If the substrate could be converted by CYP102A7, NADPH as the electron donor would be consumed. Therefore, the consumption of NADPH was determined by measuring the absorbance of the reaction mixture at 340 nm, which indicates the potential activity against substrates by CYP102A7.

As is shown in Figure 3.5, the absorbance for 150 μM NADPH and 150 μM NADP⁺ were stabilized around 0.33 and 0.17 respectively. Therefore, for any reactions performed, if the absorbance stayed below 0.33 and kept declining after mixing, the NADPH was assumed to be used by CYP102A7 and the substrate was assumed to be converted by CYP102A7 within the system. The lowest absorbance that could be reached for the mixture was around 0.17, which indicated all NADPH was used for the reaction.

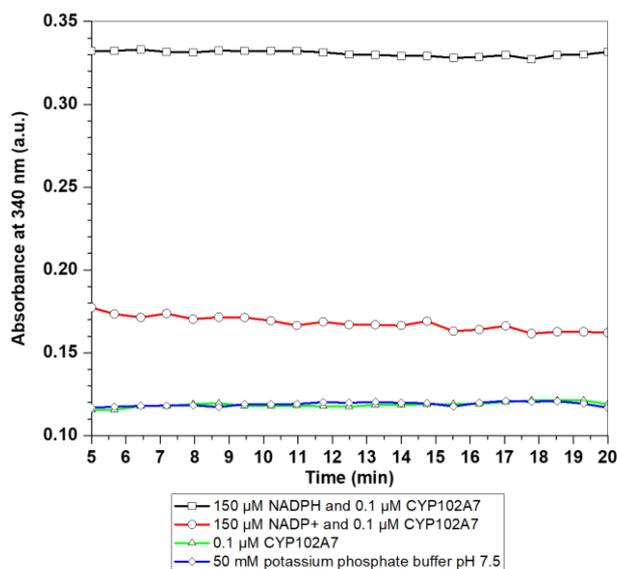


Figure 3.5. Baseline readings for the control system without adding any substrates. The control system was prepared in a 96-well plate with a total volume of 100 μL . After all components were added, the plate was sealed and shake for 5 min at room temperature for thorough mixing. After that, the plate was scanned at 340 nm every 40 seconds to measure the absorbance readings. Four control systems were prepared for baseline monitoring, which included 150 μM NADPH, 150 μM NADP⁺, 0.1 μM CYP102A7, and potassium phosphate buffer at pH 7.5.

Like all control systems, the reaction system was prepared and mixed thoroughly for 5 minutes before being measured at 340 nm. Each reaction was measured every 5 min for 80 min. As is shown in Figure 3.6, the absorbance readings for these 7 substrates (lauric acid, gamma-linolenic acid, ricinoleic acid, camphor, carvone, limonene, and tropinone) showed decreasing trend below 0.33 as incubation time increased. For other substrates in the panel (Appendix A Table A1), their reactions have the absorbance fluctuating around 150 μM NADPH baseline reading (0.33), which was considered no NADPH consumption. The

monitor of NADPH in the reaction system is a viable strategy for substrate screening, which only offers how NADPH was used after the 5-minute mixing stage.

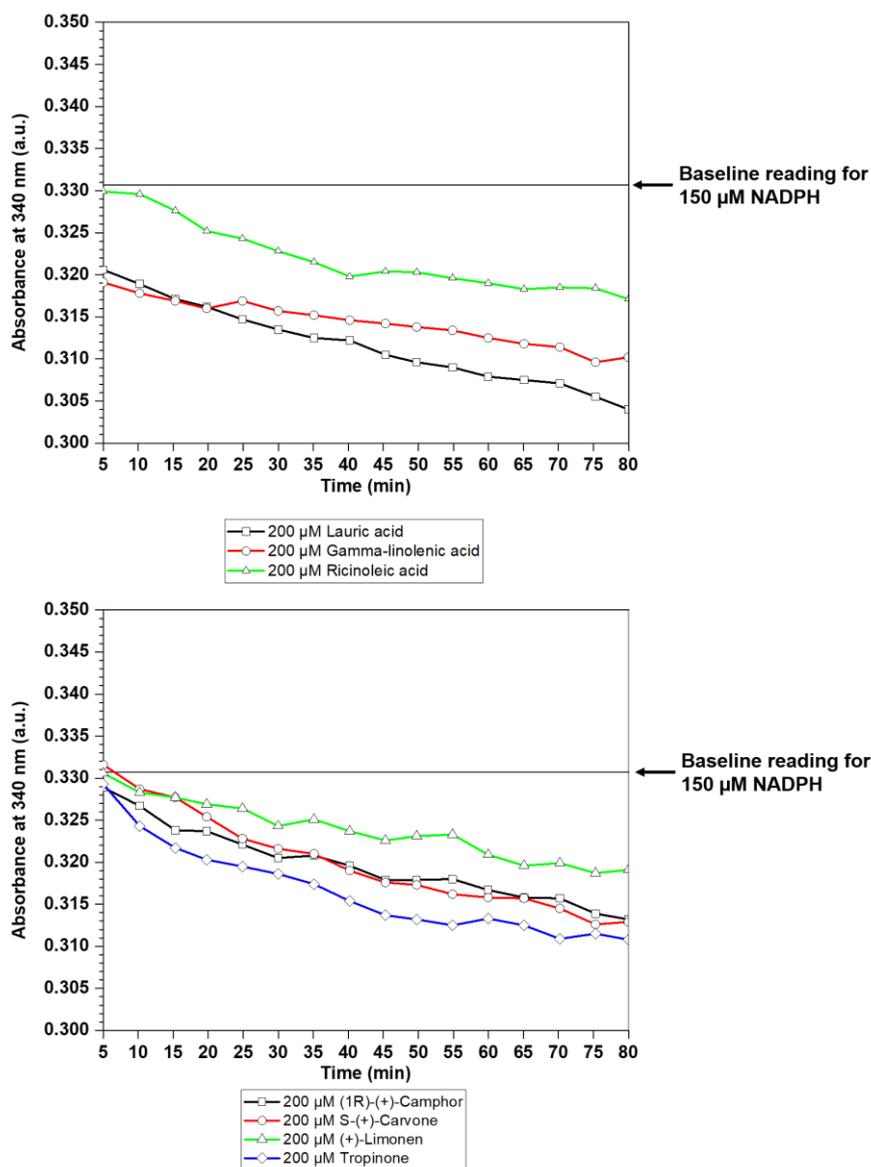


Figure 3.6. Potential fatty acids and terpenes catalysed by CYP102A7. The reaction system was prepared in a 96-well plate with a total volume of 100 μ L. After all, components were added, the plate was sealed and shaken for 5 min at room temperature for thorough mixing. After that, the plate was scanned at 340 nm every 5 minutes to measure the absorbance readings. The plate was sealed and kept on a shaker at room temperature if not being measured. The full substrate list is shown in Appendix Table A1. All substrates were prepared as 10 mM stock solution in DMSO.

Therefore, 2 μL was added to 100 μL reaction mixture to give a final concentration of 200 μM substrate as well as 2% (v/v) DMSO composition.

The NADPH turnover for the 7 reactions was calculated in Table 3.4 below. Because the baseline monitoring of NADPH at 340 nm shows a stable reading for only 20 min in Figure 3.5, so the NADPH consumption rate is calculated for the first 20 min of the reaction rather than the entire 80 min. The obvious difference in NADPH turnover for each reaction could be seen, but all consumption rates of NADPH were very low. In a previous study of CYP102A7, the NADPH turnover for lauric acid and (+)-limonene was also calculated as 1664 min^{-1} and 303 min^{-1} (124). An interesting fact is that the NADPH consumption rate on lauric acid (1664 min^{-1}) is around 5 to 6 times higher than that on (+)-limonene. The same relative rate difference is also seen in lauric acid (1.1 min^{-1}) and (+)-limonene (0.2 min^{-1}) in current reactions.

Table 3.4. Calculation of NADPH consumption rate in the first 20 min reaction period. The reaction rates were calculated using an extinction coefficient of 6.22 mM^{-1} for NADPH. The total reaction system is 100 μL containing 0.1 μM CYP102A7, which equals 0.01 nmol P450 in total.

Reactions with substrates	Abs at 20 min	Δ Abs ^a	NADPH consumed (nmol) ^b	NADPH turnover (nmol nmol P450 ⁻¹ min ⁻¹) ^c
Lauric acid	0.3162	0.0138	0.2219	1.1
γ -Linolenic acid	0.3160	0.0140	0.2251	1.1
Ricinoleic acid	0.3252	0.0048	0.0772	0.4
Camphor	0.3237	0.0063	0.1013	0.5
Carvone	0.3254	0.0046	0.0739	0.4
(+)-Limonene	0.3269	0.0031	0.0498	0.2
Tropinone	0.3203	0.0097	0.1559	0.8

^a The difference of absorbance at 340 nm is calculated as subtracting final absorbance from the baseline reading of 0.33 for NADPH

^b The amount of NADPH is calculated using the extinction coefficient mentioned above.

^c Turnover rate of NADPH is calculated using the total amount of CYP102A7 and the monitored reaction period of 20 min.

For some reason, the reactions in this study with previously confirmed substrates were much slower than their supposed rates. Therefore, a gap analysis between

the previous and current study on reaction compositions as well as reaction conditions is conducted, showing in Table 3.5, which helps to understand why the NADPH consumption was so slow and what could be improved for the current study. Both studies have prepared reactions in pH 7.5 potassium phosphate buffer at room temperature, and both reaction systems have more substrate than NADPH at the molar level. The main differences stay at the reaction volumes and CYP102A7 concentrations. The previous reaction design was set to be 1 mL total volume, which is larger than the current 100 μ L reaction on the common 96-well plate. The smaller reaction volume may lead to less accessibility to molecular oxygens that also play key roles in the oxidation reaction. The other obvious difference is the amount of P450 involved in the reaction. The molar ratio in the current reaction of the substrate to CYP102A7 is 2000:1, while it was 200:1 in the previous reaction system. It is unknown whether the much higher substrate to enzyme ratio compromises the enzyme efficiency, but a larger amount of enzyme is recommended for enzyme reactions in the future.

Table 3.5. The main differences between previous and current substrate screening reactions. All differences are highlighted as red.

	Previous Research in 2008 ⁽¹²⁴⁾	Current Study
Reaction volume	1 mL	100 μ L
Temperature	Room temperature	Room temperature
pH	7.5	7.5
Buffer system	Potassium Phosphate	Potassium Phosphate
Substrate concentration	300 μ M	200 μ M
CYP102A7 concentration	1.5 μ M	0.1 μ M
NADPH concentration	150 μ M	150 μ M
DSMO	2% (v/v)	2% (v/v)

3.5 Summary

There are three original objectives of the research activity of Chapter 3, which are summarized below.

- The *cyp102a7* gene was successfully cloned from the *Bacillus licheniformis* ATCC 14580 genome, and the gene fragment was assembled on to pET29a for expression in BL21(DE3).
- The second objective was to evaluate two approaches to conducting carbon monoxide assays. The reduction reaction of CYP should be carried out after mixing the sample with carbon monoxide.
- The last objective was to conduct preliminary substrate screening with purified CYP102A7. A group of long-chain fatty acids and terpenes was used in the preliminary screening, and positive activities were observed and recorded. There was no further product characterisation for each positive reaction.

Bioinformatic analysis was conducted to search Class VII cytochrome P450 in *Bacillus licheniformis* ATCC 14580, *Streptomyces avermitilis* MA-4680, and *Saccharopolyspora erythraea* NRRL2338, which revealed CYP102A7, CYP102D1 and CYP102G2 respectively. Genomic PCR was conducted to acquire a complete DNA sequence for further characterisation. Despite multiple attempts on genomic PCR, only the *cyp102a7* gene from *Bacillus licheniformis* ATCC 14580 was successfully copied out from the original genome and verified by sequencing. Therefore, only the *cyp102a7* gene was carried forward for cloning and expression.

CYP102A7, as a self-sufficient cytochrome P450, was successfully expressed in *E.coli* and characterised through SDS-PAGE analysis and carbon monoxide assay. The amount of active CYP102A7 produced by *E.coli* BL21 (DE3) in TB was determined to be 64.167 nmol per litre of cell culture or 2.239 mg per gram of dry cell. According to the difference spectrum of CYP102A7, there is a more distinctive sorbet peak at 420 nm, which indicates a large amount of inactive CYP102A7 produced compared to that of active CYP102A7 showing a peak at 450 nm. Therefore, the current expression approach is not optimal for producing CYP102A7 either in acceptable quantity or in good quality.

The carbon monoxide assay was carried out in two different approaches. The first approach is to conduct reduction of cytochrome P450 before carbon monoxide binding, while the second approach is to premix carbon monoxide with

cytochrome P450 before reduction reaction with $\text{Na}_2\text{S}_2\text{O}_4$. The second approach generated a more informative spectrum with distinctive and readable peaks at 420 nm and 450 nm. Therefore, it is important that the carbon monoxide is premixed with cell lysate before reduction with $\text{Na}_2\text{S}_2\text{O}_4$. After bubbling through the cell lysate with carbon monoxide, the mixture was sealed in cuvettes, which allowed immediate binding of carbon monoxide to reduced cytochrome P450 in a contained environment. For all carbon monoxide assay conducted in other projects, carbon monoxide was always firstly mixed with cell lysate or enzyme solution before the addition of $\text{Na}_2\text{S}_2\text{O}_4$.

In the preliminary substrate screening, a reaction system was set up with purified CYP102A7, substrate, NADPH, and potassium phosphate buffer at pH 7.5. Because CYP102A7 requires both electrons from NADPH for the oxidation reaction to happen, so the depletion of NADPH is the potential proof of activity on substrates. The amount of NADPH in the reaction mixture was monitored by measuring the absorbance at 340 nm. After testing all listed substrates in Table A1 (Appendix A), seven substrates including lauric acid, γ -linolenic acid, ricinoleic acid, camphor, carvone, (+)-limonene and tropinone were potentially catalysed by CYP102A7 (Figure 3.6). In previous research on CYP102A7 in 2008, a vast substrate panel was tested against CYP102A7, which revealed CYP102A7 had activities on saturated fatty acids with carbon chains not shorter than lauric acid (C12)⁽¹²⁴⁾. CYP102A7 showed the highest activity towards tetradecanoic acid (C14)⁽¹²⁴⁾, which was very similar to its homologues P450-BM3. In addition to previous findings, the current study has found CYP102A7 could react with ricinoleic acid, camphor, carvone, and tropinone. It was considered that the enzyme was too similar to the existing P450-BM3, and it was difficult to discover new substrate activities without extensive mutagenesis on the wild type. It is easily predicted that the study of CYP102A7 mutagenesis would have been all very similar to the P450-BM3 protein engineering that has been going on for more than two decades. The protein engineering approach has created P450-BM3 mutants capable of oxidizing C-H bonds with tailored selectivity and even performing new-to-nature C-H functionalization⁽¹⁵¹⁾. Therefore, the study of self-sufficient CYP102A7 was no longer the focus of my research program.

4 Chapter 4: Design and optimisation of operon constructs for heterologous expression of Class I cytochrome P450

4.1 Expression plasmids constructed for Class I cytochrome P450 in *E.coli* BL21(DE3)

4.1.1 Introduction and aims

In the last chapter, the self-sufficient cytochrome Class VII P450 CYP102A7 from *Bacillus licheniformis* was expressed in *E.coli* and purified for enzyme assays. Comparing to the common Class I cytochrome P450s, self-sufficient CYP102A7 does not require other electron transfer proteins to assist on enzymatic activities, but it is a rare P450 family in sequenced genomes. Besides, all enzymes of the BM3-type family have a limited substrate range that needs mutagenesis to become wider and more versatile. There is a large pool of Class I cytochrome P450s in nature, which indicates the diversity of enzyme functions that could be offered by these P450s. Therefore, the research focus was moved from the study of self-sufficient cytochrome P450s to Class I cytochrome P450s. The current challenge not only lies on the characterisation of Class I P450s but also involves the production of corresponding electron transfer proteins, as well as understanding how different electron transfer proteins may influence the production of P450s or the efficiency of P450's functions. Therefore, the operon design of producing CYPs and electron transfer proteins came into our attention.

In previous research, the attempts of assembling the genes of electron transfer partners and Class I cytochrome P450s within one operon were made to achieve both better expression of Class I cytochrome P450s in a microbial host and a complete chain of electron transferring^(152, 153)

In Hussain and Ward's research, the operon was designed to assemble a ferredoxin reductase gene with a DNA fragment containing naturally adjacent P450 gene and ferredoxin gene⁽¹⁵²⁾. However, it is quite rare to have a ferredoxin gene closely positioned to a class I P450 gene in the microbial genome. Therefore, it is more reasonable to have an operon design of all three genes (*cyp*, *fd*, *fdx* genes) being flexibly replaced. Based on the construct of the original three-gene operon, an alternative operon construct with defined restriction sites was designed for the expression of Class I cytochrome P450s (Figure 4.2). This

project aimed to design and test the integrity of the alternative operon, as well as to discuss the effects of ferredoxin and ferredoxin reductase within the operon on enhancing the production of various actinomycete cytochrome P450s.

4.1.2 Original operon constructed in pQR367 and pQR368

The operon, containing P450, ferredoxin, and ferredoxin reductase, was constructed onto pET21a-like plasmid in a previous study⁽¹⁵²⁾. As is shown in Figure 4.1(A), *cyp105a1* and its naturally adjacent ferredoxin *fd1* were generated together as PCR fragments from the genomic DNA of *S.griseolus* ATCC 11796, which were then assembled with ferredoxin reductase *fdr_SCF15A* gene from *S.coelicolor*. For the operon containing *cyp105b1* and its clustered ferredoxin *fd2* from *S.griseolus*, the assembly pattern was the same as that in pQR367.

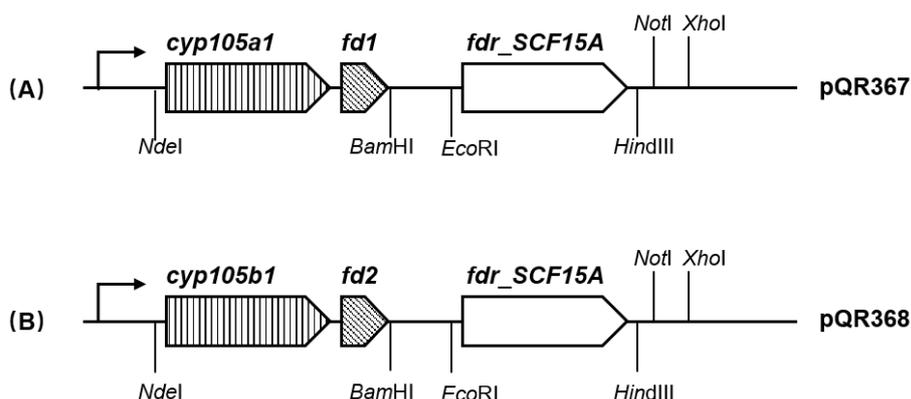


Figure 4.1. Original operon for CYP, ferredoxin and ferredoxin reductase on pQR367 and pQR368. There are two components within the original operon, containing the CYP-ferredoxin cluster and the ferredoxin reductase. There is no restriction site between the *cyp* and ferredoxin genes, and the natural distance between the *cyp* gene and the ferredoxin gene is around 20-40 bp.

According to previous research, the expression of three genes, containing CYP, ferredoxin and ferredoxin reductase, was induced by IPTG on the single T7 promoter. The three proteins together complete the electron transfer cycle that naturally serves for oxidation reactions conducted by cytochrome P450s. However, it is rare to find cytochrome P450s naturally clustered with ferredoxin, but commonly appeared as independent P450s distantly located to other ferredoxins along the genome. Therefore, the original operon is limited to

assemble those P450s clustered with ferredoxin. An alternative operon is needed, which has the potential for flexibly replacing the *cyp* component with any other Class I P450 genes.

4.1.3 An alternative operon design

As is shown in Figure 4.2, the alternative design of operon contains three gene components with standardized restriction sites. Compared to the original operon, the distance between each genetic component on the new operon is set to be minimal. For example, the sequence between the stop codon of *cyp* and the start codon of *ferredoxin* is designed to be 5'-taa**AGGA**ATTCatg-3', while the bold AGGA indicates the ribosome binding site and the underlined GAATTC is the *EcoRI* site located before the start codon of the ferredoxin gene. The sequence between ferredoxin and *fdr_SCF15A* is designed to be 5'-taa**AGGA**AAGCTTatg-3', where the AGGA is the ribosome binding site and AAGCTT is the *HindIII* site.

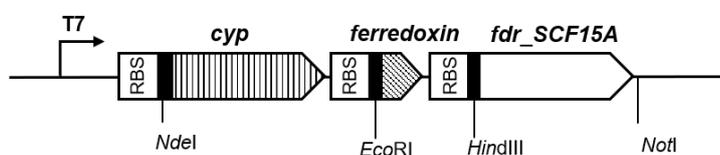


Figure 4.2. An alternative operon for expression of CYP, ferredoxin, and ferredoxin reductase. There are three components within the operon, containing *cyp*, ferredoxin and ferredoxin reductase genes. Four restriction sites, including *NdeI*, *EcoRI*, *HindIII*, and *NotI*, are designed, which allows standardized genetic components to be inserted.

The differences between the original constructs and the new construct at the sequential level were showing below. The underlined regions indicate restriction sites, and the bold regions are ribosome binding sites. The stop codon is shown in red letters, and all stop codons for the three genes in the alternative operon were unified as TAA. The alternative operon was also designed to have linking regions between genes as short and as simple as possible. All promoter regions in each operon are the T7 promoters, and not displayed below.

Original operon pQR367 in Figure 4.1:

5'-**AAGGAG**ATATACATATG-(*cyp*)-**TGAGCGGCAAGGAG**ATATA-(*ferredoxin1*)-
TGAAAGGGCA**AAGGAG**ATATAGAATTC-(*ferredoxin reductase*)-**TGAGCGGCCGC**-3'

Original operon pQR368 in Figure 4.1:

5'-**AAGGAG**ATATACATATG-(*cyp*)-**TAGGCGGCAAGGAG**ATATA-(*ferredoxin2*)-
TGAAAGGGCA**AAGGAG**ATATAGAATTC-(*ferredoxin reductase*)-**TGAGCGGCCGC**-3'

Alternative operon construct in Figure 4.2:

5'-**AGGAGGT**AAAAACATATG-(*cyp*)-**TAATAGGA**ATTC-(*ferredoxin*)-
TAATAGGAAAGCTT-(*ferredoxin reductase*)-**TAATGAGCGGCCGC**-3'

4.1.4 Construction of pQR2239 and pQR2240

In order to construct the alternative operon, a basic plasmid pQR2239 was firstly constructed, which contains only the ferredoxin gene *fd1* and ferredoxin reductase gene *fdr_SCF15A* (Figure 4.3, A). The cloning site for the *cyp* gene was designed as 5'-CATATGggcggcGAATTC-3', which involves *NdeI* and *EcoRI* as restriction sites (Figure 4.3, A). There had been several attempts on assembling *fd1* fragment, *fdr_SCF15A* fragment, and pET21a backbone, but all failed. Therefore, the initial pQR2239 was directly synthesized by DNA2.0, which was featured with Ampicillin resistant gene, pUC origin, *lac* repressor gene, as well as the designed operon, construct. The second basic plasmid pQR2240 was created by replacing the *ferredoxin 1* gene with the *ferredoxin 2* gene through digestion and ligation. The PCR products of the *ferredoxin 2* gene were shown in Figure 4.3 (B), indicating as bands with a size around 200 bp. *NdeI* and *HindIII* were the two restriction sites applied in digestion and ligation to reconstruct pQR2239 into pQR2240.

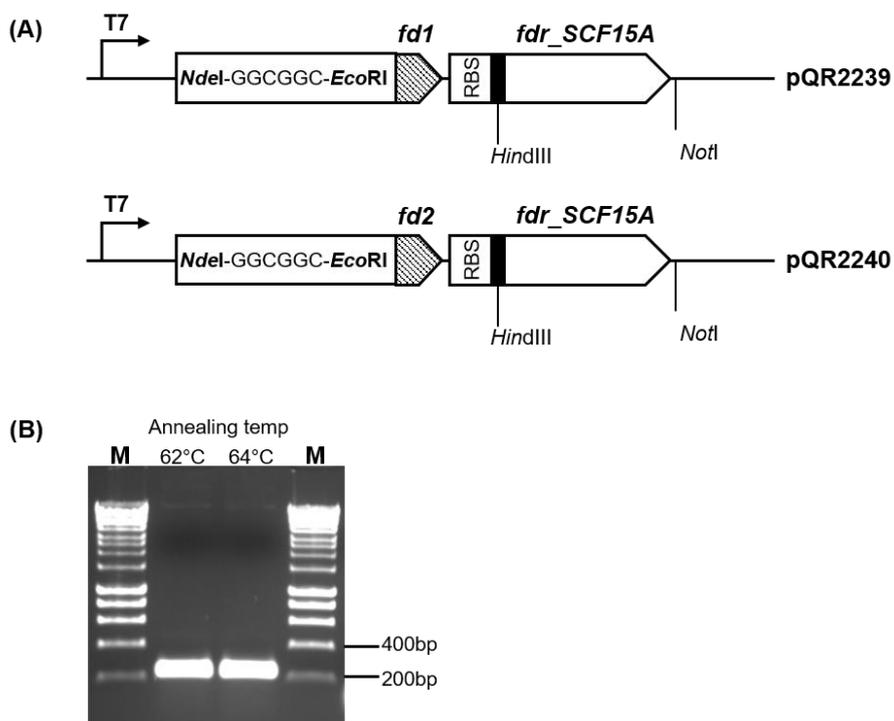


Figure 4.3. Operon construction on plasmid pQR2239 and pQR2240. (A) The two basic operon structures contain only ferredoxin and ferredoxin reductase gene. The cloning sites for P450 on both plasmids are designed to contain only *NdeI* and *EcoRI* sites for ligation. (B) The PCR products of *ferredoxin 2* gene from *S.griseolus*. The column M was displaying 10 µL of 1kb Bioline HyperLadder, and the 62°C and 64°C were the annealing temperatures used for 20 µL PCR of *ferredoxin 2*. All PCR products were loaded into the wells. The primers were designed to contain the *NdeI* site at the forward and the *HindIII* site at the reverse, which was prepared for later digestion and ligation processes to replace the *fd1* component on pQR2239.

To simply confirm if ferredoxins and ferredoxin reductases could be expressed in host cells, BL21(DE3) cells with no plasmids, with pQR2239 or with pQR2240 were incubated and induced with IPTG for 24 hours. A cytochrome c assay was conducted with the cell lysates of these three cultures, equine heart cytochrome c, and NADPH, which indicated the capability and efficiency of transferring electrons. The final cell culture was adjusted to the same optical density before being lysed for cytochrome c assay, which indicated similar cell numbers being used in cytochrome c assays. Therefore, if the ferredoxins and ferredoxin reductases were successfully expressed, the cell lysate should have higher efficiency of transferring electrons from NADPH to cytochrome c for reduction. As is shown in Figure 4.4, the cell lysates from pQR2239 and pQR2240 consume

cytochrome *c* faster than the cell lysates with no expression plasmids. There are native protein components within *E.coli* that could reduce cytochrome *c* by acquiring electrons from NADPH, and the performance of pQR2239 and pQR2240 cell lysates indicates additional electron-transferring components being present.

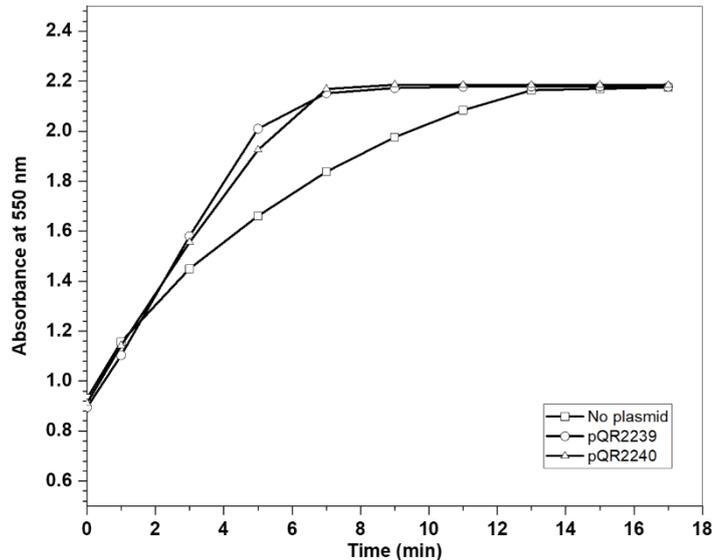


Figure 4.4. The reduction of cytochrome *c* by cell lysates containing different ferredoxins and ferredoxin reductases. The initial absorbance reading was recorded before adding any NADPH to the mixture of cytochrome *c* and cell lysate. It takes 7 minutes for pQR2239 and pQR2240 cell lysates to reduce all cytochrome *c* and more than 15 minutes for the plain cell lysates. Additional ferredoxins and ferredoxin reductases within the other two cell lysate of pQR2239 and pQR2240 increase the reduction efficiency of cytochrome *c*.

4.1.5 Cloning and expression of CYP105A1 and CYP105B1

After acquiring two basic plasmids pQR2239 and pQR2240, the genes of CYP105A1 and CYP105B1 were separately cloned into the two plasmids to test the integrity of the operon as well as to measure the expression level of P450s using the new construct. As is shown in Table 4.1, two pairs of primers were used to amplify *cyp105a1* and *cyp105b1* genes from original pQR367 and pQR368 respectively. The PCR mixture preparation was described in section 2.1.6.1.

Table 4.1. Two pairs of primers used for amplification of *cyp105a1* and *cyp105b1*.

Primer ID	Sequence 5'-3' ^a	T _m (°C) ^b
F-105a1	GCGCC <u>CATATG</u> ACCGATACCGCCAC	62
R-105a1- <i>Eco</i> RI	GCGC <u>GAATTC</u> CT TATTACCAGGTGACCGGGAGTT	64
F-105b1	GCGCC <u>CATATG</u> ACGACCGCAGAACG	64
R-105b1- <i>Eco</i> RI	GCGC <u>GAATTC</u> CT TATTACCAGGCGATCGGCA	65

^a The underlined regions indicate the restriction sites of *Nde*I and *Eco*RI, and the bold the regions indicate the ribosome binding sites.

^b The calculated annealing temperature for each primer generated by the NEB T_m calculator.

The expected band size of PCR products for both *cyp105a1* and *cyp105b1* genes are around 1200bp, and the new DNA fragment is featured with restriction sites at both ends (Figure 4.5). The bands of *cyp105a1* and *cyp105b1* were separately extracted and purified from agarose gel, and these DNA fragments were firstly sub-cloned onto TOPO vector (described in section 2.1.5) and sent for sequencing.

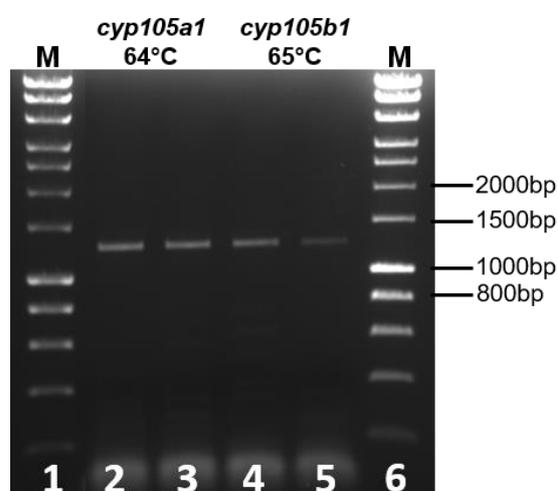


Figure 4.5. Amplification of *cyp105a1* and *cyp105b1* from pQR367 and pQR368. The annealing temperature for PCR of *cyp105a1* and *cyp105b1* were set to be 64°C and 65°C respectively. The lane 1 and 6 were the marker lands loaded with 1kb Bionline HyperLadder. The lane 2 and 3 were loaded with the PCR products of *cyp105a1*, and the annealing temperature was set to be 64°C.

Because of the designed *Nde*I and *Eco*RI sites, the sequentially correct *cyp105a1* and *cyp105b1* could be ligated onto linearized pQR2239 or pQR2240, which

leads to the formation of different plasmids that fit the design (Figure 4.6). Each of the plasmids in Figure 4.6 was transformed into *E.coli* BL21(DE3) for IPTG induced protein expression. The expression was done for 24 hours at 25°C, and the amount of active P450s was measured through carbon monoxide assay (Figure 4.7, method details described in section 2.18).

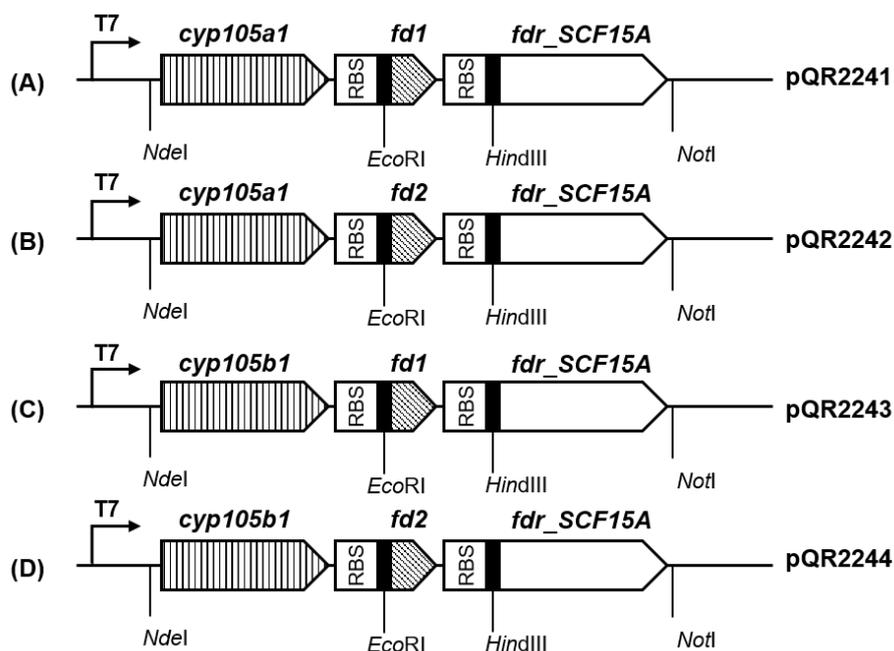


Figure 4.6. Expression plasmids for CYP105A1 and CYP105B1. The combination of a P450 gene, a ferredoxin gene, and a ferredoxin reductase gene fit the design of the three-component operon for P450 expression.

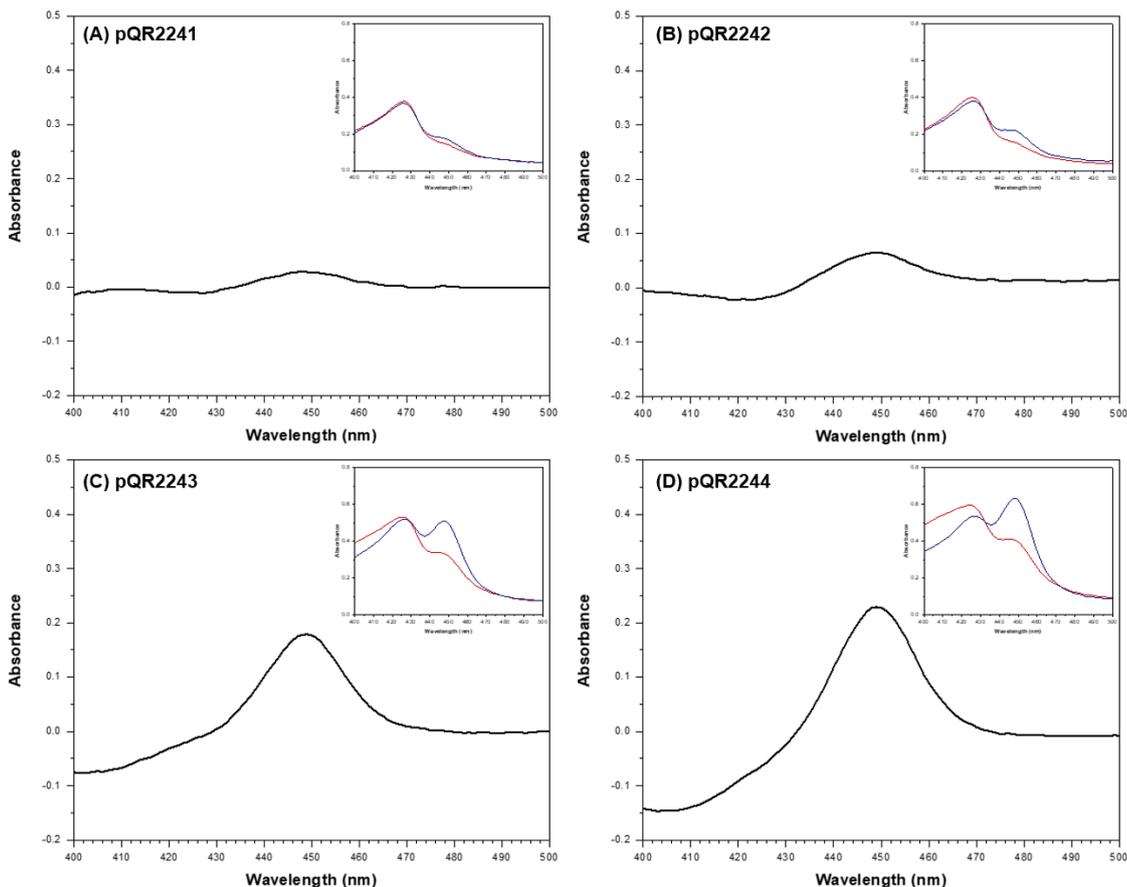


Figure 4.7. The expression level of CYP105A1 and CYP105B1. The four carbon monoxide difference spectra indicate the amount of active P450 produced by *E.coli* cells using four plasmids described in Figure 4.6. The inserted graph in each includes the spectrum of reduced cytochrome P450 (red) and the spectrum of CO-cytochrome P450 binding mixture (blue). The difference spectrum was calculated by subtracting the red spectrum from the blue spectrum. The extinction coefficient of $91,000 \text{ M}^{-1} \text{ cm}^{-1}$ at 450 nm on the difference spectrum was used to determine the molar concentration of active P450. Therefore, the higher the absorbance it is at 450 nm, the higher amount of P450s there is being produced. (A) The expression level of CYP105A1 using plasmid pQR2241. (B) The expression level of CYP105A1 using plasmid pQR2242. (C) The expression level of CYP105B1 using plasmid pQR2243. (D) The expression level of CYP105B1 using plasmid pQR2244.

As is shown in Figure 4.7 (A) and (B), the expression plasmids used for producing CYP105A1 were pQR2241 and pQR2242 respectively. The only difference between pQR2241 and pQR2242 is the different ferredoxin genes assembled in each operon. According to the peak absorbance at 450 nm in Figure 4.7 (A) and

(B), it seems the combination of *fd2* and *fdr_SCF15A* may increase the production of CYP105A1 than the use of *fd1* and *fdr_SCF15A* in the same operon configuration. In Figure 4.7 (C) and (D), CYP105B1 was expressed using pQR2243 and pQR2244 respectively. The amount of P450 produced using pQR2244 is more than that using pQR2243, which indicates the expression enhancement caused by using the *fd2* gene within the operon.

In order to compare the enhanced expression level of CYPs when assembling different ferredoxin genes within the operon, the molar concentrations of active P450s and the specific yields of active P450s were firstly calculated. As is shown in Table 4.2, the molar concentration represents the capability of the batch of cell culture on producing active P450s, and the specific yields represent how much active P450s have been produced in each host cell. When expressing ferredoxin 2 together with CYP105A1 (pQR2242), there is enhanced production of the enzyme on both molar concentration and specific yields, compared to the expression of CYP105A1 together with ferredoxin 1 (pQR2241). There is also an increased expression of CYP105B1 when using the operon containing the *fd2* gene rather than the *fd1* gene. Therefore, it is the same conclusion that the combination of *fd2* and *fdr_SCF15A* within the operon may increase the production of active P450s.

Table 4.2. Enhanced expression of CYP105A1 and CYP105B1 by using different operons. The absorbance at 450 nm on the difference spectrum as well as the dry cell weight of 1 mL cell culture was measured. The molar concentration was calculated by using the absorbance at 450 nm and the corresponding extinction coefficient value, which gives the total amount of active P450 produced per litre of cell culture. The specific yield is the amount of P450s produced per gram of dry cell weight.

	pQR2241	pQR2242	pQR2243	pQR2244
Operon construction	<i>cyp105a1</i> <i>ferredoxin 1</i> <i>fdr_SCF15A</i>	<i>cyp105a1</i> <i>ferredoxin 2</i> <i>fdr_SCF15A</i>	<i>cyp105b1</i> <i>ferredoxin 1</i> <i>fdr_SCF15A</i>	<i>cyp105b1</i> <i>ferredoxin 2</i> <i>fdr_SCF15A</i>
Molar concentration of active P450s (nmol/L)	49.45	95.24	329.67	434.07

	pQR2241	pQR2242	pQR2243	pQR2244
Specific yields of active P450s (mg per gram of dry cell weight)	0.42	1.45	2.99	5.67

4.2 The effects of ferredoxin and ferredoxin reductase on the expression of Class I cytochrome P450s

4.2.1 *cyp-fd1-fdr* operon and *cyp-fd2-fdr* operon

After investigating the effects of ferredoxin 1 and ferredoxin 2 on expressing CYP105A1 or CYP105B1 in section 4.1.4, it was concluded that placing the *fd2* gene into the operon could lead to increased production of active CYP105A1 and CYP105B1, compared to the case of using *fd1* gene in the operon. However, it is unknown whether ferredoxin 2 could also enhance the production of other CYPs as well in *E.coli* BL21(DE3) hosts while using the same operon construct. Besides, it is also not concluded that the operon containing three genes is currently the most optimal expression system for *E.coli* hosts on heterologous production of class I cytochrome P450s.

Therefore, five additional microbial P450s originating from different actinomycete strains were identified through searching for homologues of CYP105A1 and CYP105B1, which were then cloned onto pQR2239 and pQR2240 separately. These assembled plasmids for each P450 were listed and described in section 2.1.2. The cloning strategy involved amplification of *cyp* genes through PCR using genomic DNA as templates (Table 4.3), and assembly of *cyp* genes to designed plasmids pQR2239 or pQR2240. These CYPs were then expressed together with ferredoxins and ferredoxin reductases under the same conditions that used for the expression of CYP105A1 and CYP105B1.

Table 4.3. Primers designed for PCR of five *cyp* genes. The underlined sequences represent restriction sites, and the bold sequences are ribosome binding sites designed to fit in the expression operon.

<i>cyp</i> genes	Organism	Primers	Sequence 5'-3'
<i>cyp105d1</i>	<i>Streptomyces griseus</i>	Forward	GCGCCATATGACGGAATCCACGACG
		Reverse	GCGCGAATTCCTATTACCAGGCCACGGGC
<i>cyp105d5</i>	<i>Streptomyces coelicolor</i>	Forward	GCGCCATATGACGGACACCGACACGACG
		Reverse	GCGCGAATTCCTATTACCAGGCCACGGGGAG
<i>cyp105d7</i>	<i>Streptomyces avermitilis MA-4680</i>	Forward	GCGCCATATGACAGAGCCCGGTACGTCC
		Reverse	GCGCGAATTCCTATTACCAGGTCACGGGGAGTT
<i>cyp105ab1</i>	<i>Saccharopolyspora erythraea NRRL2338</i>	Forward	GCGCCATATGGCGTCTTCAGAAGCCCTGAG
		Reverse	GCGCGAATTCCTATTACCACGTACCGACATCTCG
<i>cyp154c2</i>	<i>Streptomyces avermitilis MA-4680</i>	Forward	GCGCCATATGACGACTCGTATCGCCCTGG
		Reverse	GCGCGAATTCCTATTAGCCCAGCCGCACCG

Other than the operon structure containing *cyp*, *fd*, and *fdr* genes, a plasmid contains only the *cyp* gene was also constructed (Figure 4.8), and the assembled plasmids for each P450 were listed and described in section 2.1.2. The purpose of constructing a single-gene plasmid is to estimate the basic expression level of active P450s and to compare it to the performance of using the operon constructs as the expression platform. After comparing the different expression levels of target P450s when using either the single-gene construct or three-gene operon, the advantage or disadvantage of placing different ferredoxins within an operon could be revealed. According to Figure 4.8, these operon constructs were specifically named as *cyp* construct, *cyp-fd1-fdr* operon, and *cyp-fd2-fdr* operon in this project. Each operon was induced in BL21(DE3) for expression, and the entire expression process for each P450 was repeated three times under the same conditions to acquire the average expression level of each CYP.

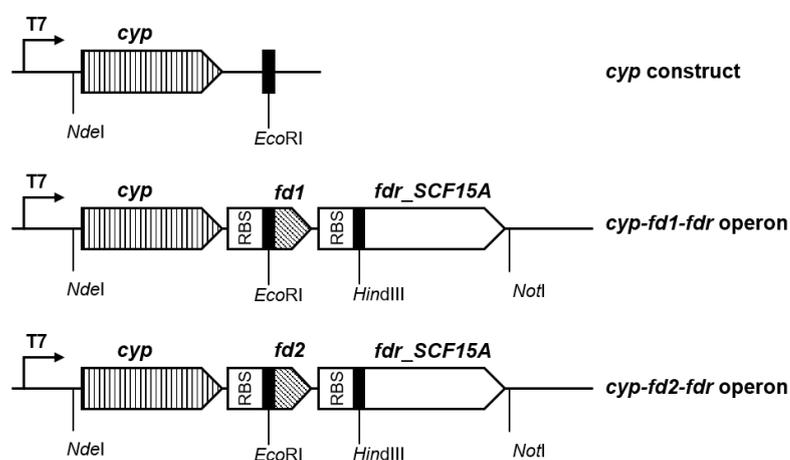


Figure 4.8. The single-gene construct and three-gene operon for P450 expression. These three constructs were assembled onto the same plasmid backbone. Because the *NdeI* and *EcoRI* sites were also present on the *cyp* construct, so the PCR products of those targeted CYPs could also be inserted. There is only the need to design one pair of primers if we want to insert a new P450s gene into all three operons.

The expression level of each P450s assembled in different operons was indicated and compared in Figure 4.9. For CYP105B1, CYP105AB1 and CYP154C2, there were enhanced expressions of CYP when using either *cyp-fd1-fdr* operon or *cyp-fd2-fdr* operon, comparing to the use of *single-cyp* construct expression in *E.coli*. Especially for CYP105AB1, the expression level of active P450s was around 20 times higher using *cyp-fd2-fdr* operon than *single-cyp* construct, and the expression level was around 4 times higher using *cyp-fd2-fdr* operon than using *cyp-fd1-fdr* operon. For CYP105B1, the expression level of active P450s was slightly higher using *cyp-fd2-fdr* operon than *cyp-fd1-fdr* operon. For CYP154C2, the clustered ferredoxin 1 and ferredoxin 2 have similar effects on enhancing the expression level of CYPs.

For CYP105A1 and CYP105D7, the expression of a single *cyp* gene in *E.coli* has the highest level, comparing to the use of other three-gene operons. It seemed that the production of these two P450s was decreased when clustering with ferredoxin and ferredoxin reductase in the expression operon. However, *cyp-fd2-fdr* operon performed better than *cyp-fd1-fdr* operon on enhancing the production of CYP105A1.

For CYP105D5, there is extremely low or no detectable expression when using the single-gene operon or the *cyp-fd2-fdr* operon. However, the *cyp-fd1-fdr* operon leads to expression in an acceptable amount. Therefore, expression of ferredoxin 1 may be vital for producing CYP105D5, and ferredoxin 2 cannot function like ferredoxin 1 on CYP105D5 production enhancement.

For CYP105D1, there are very low expression levels on both bar charts in Figure 4.9. Therefore, none of the three operon constructs lead to the successful production of CYP105D1. Further studies on either the hosts or the operon constructs are needed to achieve a detectable expression level of CYP105D1.

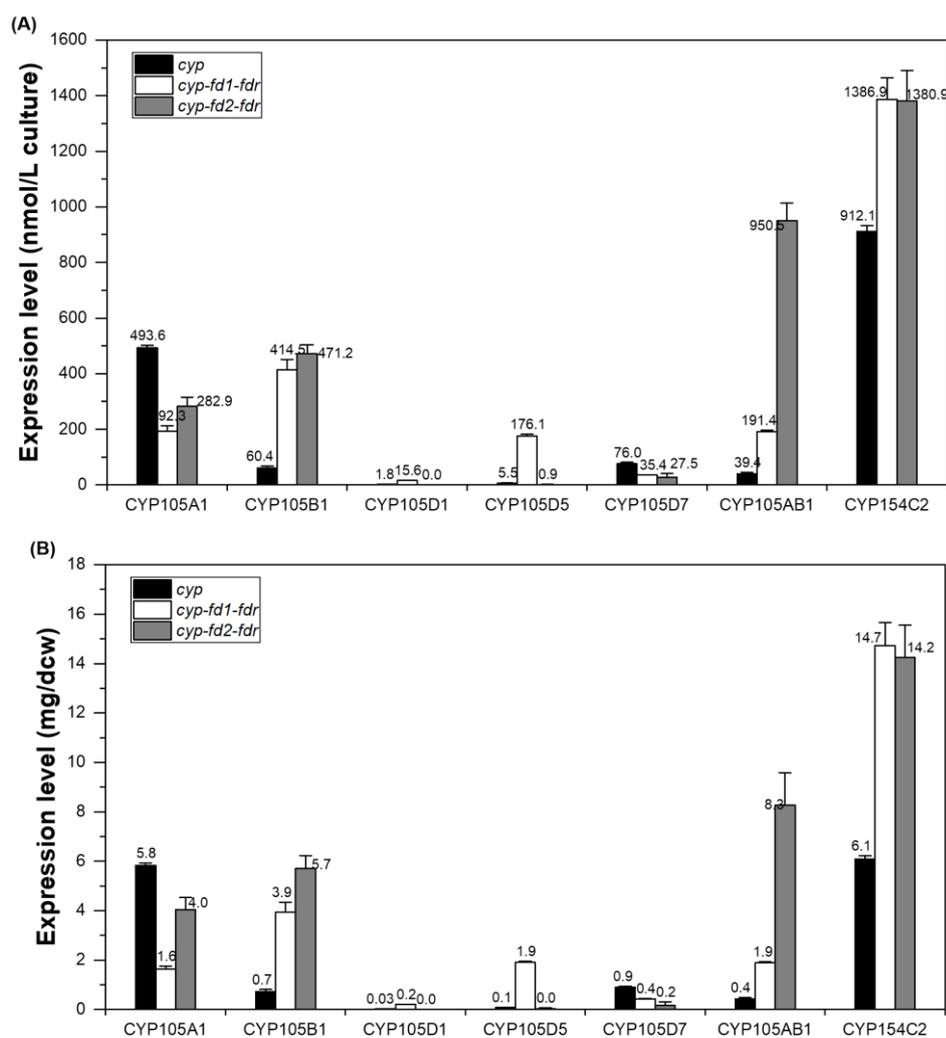


Figure 4.9. The expression level of each P450s when using *cyp-fd1-fdr* or *cyp-fd2-fdr* operons. The expression levels of each P450s were measured through carbon monoxide assay, which uses the difference absorbance at 450 nm to calculate the molar concentrations of active P450s (A) and using the molar concentration of active P450s and dry cell weight of *E.coli* to calculate the specific yields (B).

In conclusion, the expression level of class I P450s in *E.coli* could be influenced by the expression of ferredoxins and ferredoxin reductases. The expression of ferredoxin 1 or ferredoxin 2 does not always enhance the expression of P450, but they have the potentials to boost the expression of some target P450s. Besides, ferredoxin 2 (*cyp-fd2-fdr* operon) seems to perform better than ferredoxin 1 (*cyp-fd1-fdr* operon) on enhancing the expression of targeted CYPs such as CYP105A1, CYP105B1, CYP105AB1, and CYP154C2. Especially for CYP105AB1, the yield is much higher when expressing together with ferredoxin 2. Both ferredoxin 1 and ferredoxin 2 genes clustered within these operons are

from *S.griseolus*, but it is unknown why ferredoxin 2 performs better than ferredoxin 1 on significantly enhancing the production of CYP105AB1. Therefore, it is worth considering the operon containing *ferredoxin 2* gene (pQR2240) for an initial trial of heterologous expression of class I P450s from actinomycetes in the future.

4.2.2 *cyp-fd2-fdr* operon and *cyp-pdx-pdr* operon

Because class I cytochrome P450s naturally need electron transfer partners to function, so it is quite necessary and useful in our study to develop universal redox partners to reconstitute P450 activities *in vivo* or *in vitro*. In addition to the ferredoxin and ferredoxin reductases discussed above, there are also some other microbial or eukaryotic electron transfer systems being widely studied in the field. For example, the microbial Pdx-PdR system (putidaredoxin and putidaredoxin reductase) was firstly identified in *Pseudomonas putida* DSM 50198, which transfers electrons from NADH to P450cam (CYP101A1)⁽⁶⁷⁾. The eukaryotic Adx-AdR system (adrenodoxin and adrenodoxin reductase) was firstly discovered in adrenal cortex mitochondria, which transfers electrons from NADPH to target P450s⁽¹⁵⁴⁾.

The Pdx-PdR system has been widely applied in the reconstitution of P450 activity^(155, 156), which commonly involves the use of separately purified putidaredoxin, putidaredoxin reductase, and targeted P450s at defined molar concentrations. It will be more convenient to conduct bioconversion reactions in the future if all three enzymes could be expressed together at host cells. Therefore, a new operon was constructed and named as a *cyp-pdx-pdr* operon in this study (Figure 4.10A), which shares the same construct to *cyp-fd2-fdr* operon. The possible effects of *cyp-pdx-pdr* operon on the seven actinomycete P450s are also investigated and compared.

As is shown in Figure 4.10 (B) and (C), for CYP105A1, CYP105B1, CYP105AB1, and CYP154C2, the *cyp-pdx-pdr* operon did not lead to any more enhanced productions of these targeted CYPs, comparing to the use of *cyp-fd2-fdr* operon. For CYP105D1, CYP105D5, and CYP105D7, there was no detectable amount of active P450s being produced.

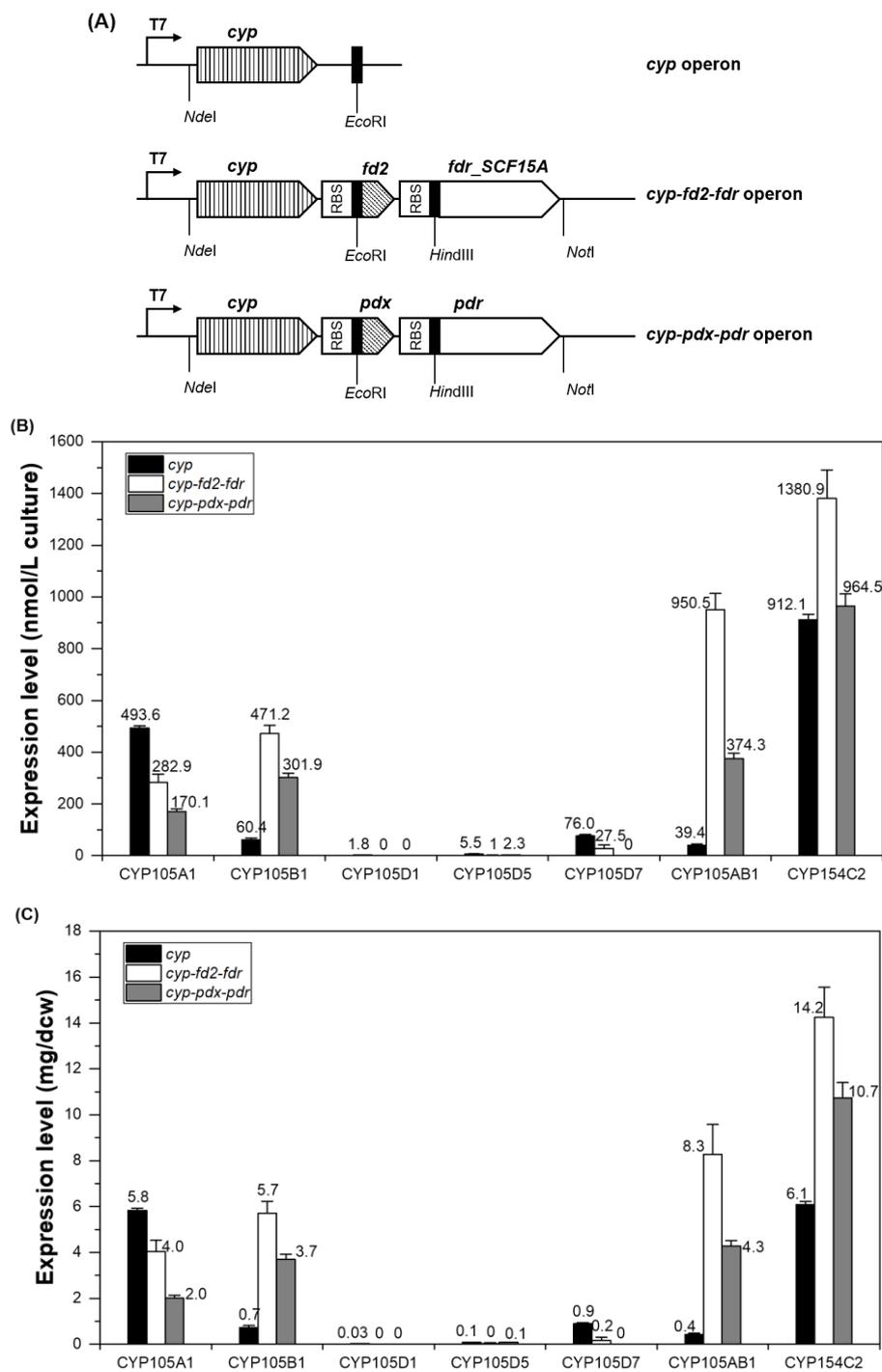


Figure 4.10. The expression level of each P450 when using *cyp-fd2-fdr* or *cyp-pdx-pdr* operons. (A) The single-gene operon and three-gene operon for P450 expression. All three operons are assembled on the same plasmid backbone. (B) Molar concentrations of active P450s. (C) The specific yields of active P450s per gram of dry cell weight.

4.2.3 Analysis of variance (ANOVA) test on expression level using different operon constructs

According to Figure 4.9 and Figure 4.10 in the last section, each candidate CYPs were assembled with four different constructs, and the expression levels were evaluated. It is important to determine if the influence of different constructs (independent variables) on expression of target CYP is statistically significant⁽¹⁵⁷⁾. The ANOVA test procedure is described in detail below.

The error bars in Figure 4.9 and Figure 4.10 represent the standard deviation, which could be used to determine the variance of the population of each CYP expression level. Only the expression levels of CYP105A1, CYP105B1, CYP105AB1, and CYP154C2 are subjected to ANOVA tests. For the expression of CYP105D1, CYP105D5 and CYP105D7, some constructs had led to no expression of target CYPs, so they are not fit for ANOVA test criteria. The four constructs are considered as four different treatments for target CYP expression, so mean square error (MSE) is calculated as the mean of variance among the four treatments. For each construct, three replicates of the expression were conducted to acquire the mean expression level of CYP. Therefore, a total of 12 expressions are carried out in the population. The other parameter of mean square between (MSB) is calculated as the variance among four mean expression levels divided by the total of expression trials. Both MSE and MSB results are shown in the table below. For each CYP expression, there are four constructs and each construct is repeated three times to acquire mean expression level, so the numerator degree of freedom is $(4-1)=3$, and the denominator degree of freedom is $(4*3-4)=8$. Therefore, we need look into the F-statistics chart of F ($\alpha=0.05$) (3,8), which is 4.0662. If the calculated F score of CYP expression is larger than 4.0662, the null hypothesis of expression level not influenced by different constructs should be rejected. Therefore, it is statistically significant that the expressions of CYP105A1 and CYP105AB1 are influenced by the change of constructs.

Table 4. 4. The variance of expression levels is summarized below. The unit of expression level was described as nmol per litre of cell culture. The mean expression level and the standard deviations were reported in Figure 4.9 and Figure 4.10. The

variance in the table was defined as the square of standard deviation. The F ($\alpha=0.05$) (3,8) is found at the F-chart as 4.0662.

Constructs	CYP105A1		CYP105B1		CYP105AB1		CYP154C2	
	Mean	Variance	Mean	Variance	Mean	Variance	Mean	Variance
<i>cyp</i>	493.6	70.9	60.4	60.4	39.4	34.0	912.1	429.4
<i>cyp-fd1-fdr</i>	192.3	377.4	414.5	1367.2	191.4	34.0	1386.9	5991.0
<i>cyp-fd2-fdr</i>	282.9	1030.1	271.2	1037.2	950.5	3900.4	1380.9	12001.7
<i>cyp-pdx-pdr</i>	170.1	114.9	301.9	280.2	374.3	434.1	964.5	2197.5

	CYP105A1	CYP105B1	CYP105AB1	CYP154C2
MSB	1814.3	1821.5	13243.5	5554.2
MSE	398.3	686.2	1100.6	5154.9
F=MSB/MSE	4.6	2.7	12.0	1.1

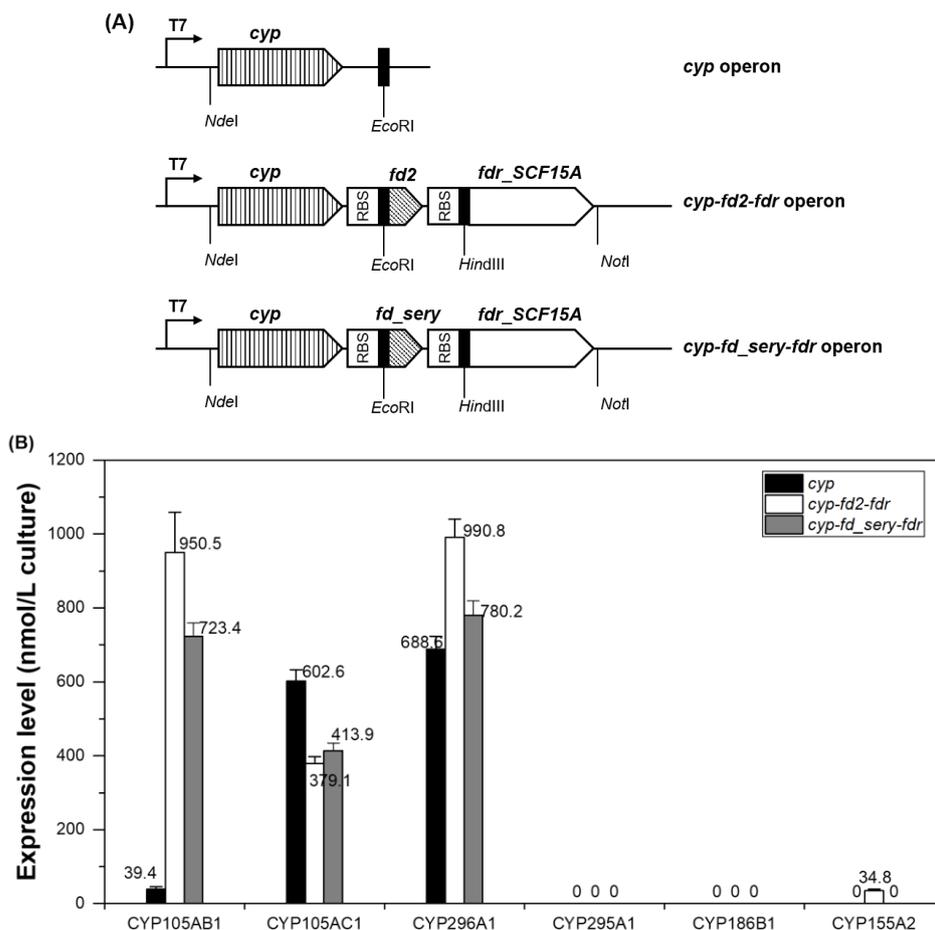
4.3 Optimised expression of CYP105AB1 from *Saccharopolyspora erythraea* NRRL2338

4.3.1 Ferredoxin 2 from *Streptomyces griseolus* and Ferredoxins from *Saccharopolyspora erythraea* NRRL2338

It has been our interest to explore the effects of redox partners on expressing cytochrome P450s especially CYP105AB1. CYP105AB1 is originated from the *Saccharopolyspora erythraea* NRRL2338. The production of CYP105AB1 reached to the highest when expressing together with ferredoxin 2 from *Streptomyces griseolus*, but it is unknown whether the expression level of CYP105AB1 could be further enhanced by ferredoxins from its genome. The *ferredoxin 2* gene from *Streptomyces griseolus* is naturally clustered with *cyp105b1* gene in its genome, so the question at the time was whether we could find a naturally clustered ferredoxin in *Saccharopolyspora erythraea* to replace ferredoxin 2 in the three-gene operon discussed in last few sections. After searching through the genome of *Saccharopolyspora erythraea*, there was only

one ferredoxin gene naturally clustered with a *cyp* gene, which was first found and named as *fd_sery* in this project. The sequence detail is shown in Appendix B, section 10.1.3. It is suspected that *fd_sery* may naturally have positive impacts on the production of CYPs originated from *S.erythraea*. Another hypothesis was that the *fd_sery* gene may have stronger enhancement than *fd2* on the expression of CYP105AB1 when using the three-gene operon in *E.coli*.

As discussed in Figure 4.10 (A), the genetic components of the *cyp-fd2-fdr* operon could be replaced with any other genes. Therefore, another operon was constructed based on the structure of *cyp-fd2-fdr* operon, and the new operon was named as a *cyp-fd_sery-fdr* operon (Figure 4.11, A). To evaluate the different effects of *fd2* and *fd_sery* on producing CYPs, the CYP105AB1 and five other CYPs, which are all from *S.erythraea*, were selected to assemble into the three different operons in Figure 4.11 (A). All operons were induced in BL21(DE3), and the amount of P450s was measured through carbon monoxide assay.



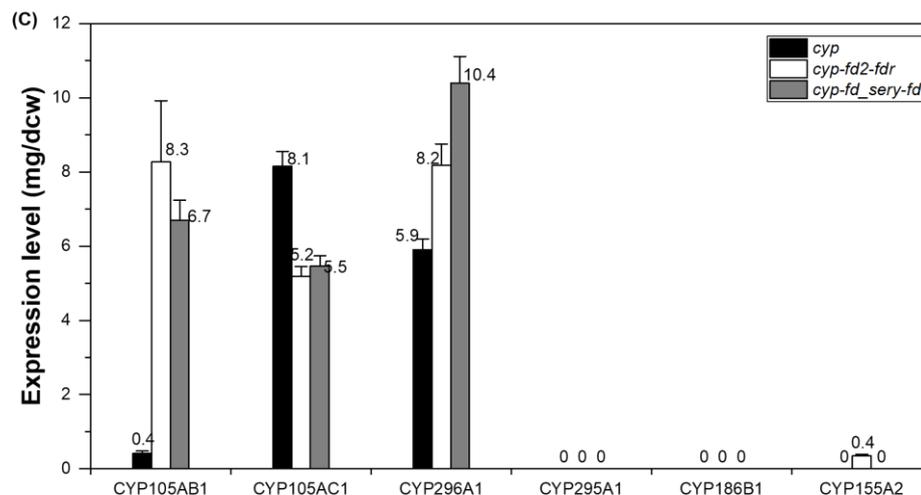


Figure 4.11. The expression level of P450s from *S.erythraea* when coupling with different ferredoxin genes. (A) Three operon constructs were used for the expression of various CYPs from *S.erythraea*. The details of each expression plasmid are listed in section 2.1.2. (B) The total molar concentration of P450s produced by *E.coli*. (C) The specific yield for each P450s per grams of dry cell weight.

According to Figure 4.11 (B) and (C), there were some CYPs from *S.erythraea* such as CYP295A1, CYP186B1, and CYP155A2 that could not be properly expressed in *E.coli* regardless of what operon constructs being used. It was expected that some CYPs could not be easily expressed using current operon design or expression strains. Further optimisation strategy on expressing foreign CYPs in *E.coli* strains will be discussed in chapter 6.

For CYP 105AB1, CYP105AC1, and CYP296A1, they were all successfully expressed using our designed operon in BL21 strains. For CYP105AB1, the operons with ferredoxin and ferredoxin reductase genes always lead to increased expression, comparing to the use of *single-cyp* construct. However, the ferredoxins (*fd_sery*) from *S.erythraea* seem not to have any better enhancement on CYP105AB1 production than the use of ferredoxins (*fd2*) from *S.griseolus*. Therefore, the *cyp105ab1-fd2-fdr* operon is shown to be the best operon that could lead to the highest production of CYP105AB1 in this project, although ferredoxin 2 is not from the same organism as CYP105AB1.

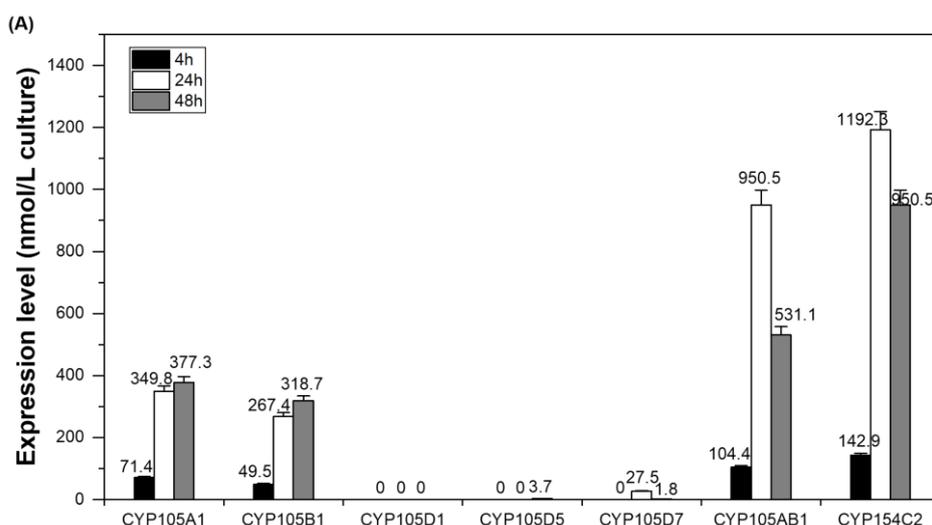
For CYP105AC1, there is a relatively similar expression level when assembling either *fd2* or *fd_sery* in the operon construct. However, the expression of single CYP105AC1 seems to be a better approach to achieve higher production of

active CYP105AC1. For CYP296A1, the *fd_sery* leads to a higher yield of CYP296A1 than *fd2* when using the three-gene operon.

Therefore, coming back to the hypotheses being made at the beginning of the section, it is not necessarily to say that the ferredoxins from *S.erythraea* may have a better enhancement than ferredoxins from *S.griseolus* on producing CYPs originated from *S.erythrae* in BL21 strains. CYP105AB1, as our target CYP that will be studied in the next chapter, had much better production when expressing with either ferredoxin (*fd_sery*) from *S.erythraea* or ferredoxins (*fd2*) from *S.griseolus*. In the future, only the *cyp-fd2-fdr* operon construct will be used in the expression of CYP105AB1, due to its better enhancement on CYP105AB1 production than using *cyp-fd_sery-fdr* operon.

4.3.2 The effects of expression time when using *cyp-fd2-fdr* operon

According to the discussion in section 4.2.1, 4.2.2, and 4.3.1, it is concluded that the *cyp-fd2-fdr* operon is currently the better construct for producing CYP105AB1 in an acceptable quantity and quality. In previous expression tests for each operon, the post-induction time was always controlled to be 24 hours. However, there are various expression time being reported in the literature, and it is unknown if the 24-hour expression is optimal to produce CYP105AB1 in this project. Therefore, different post-induction periods, including 4 hours, 24 hours, and 48 hours, were separately applied for the expression of CYPs in BL21(DE3).



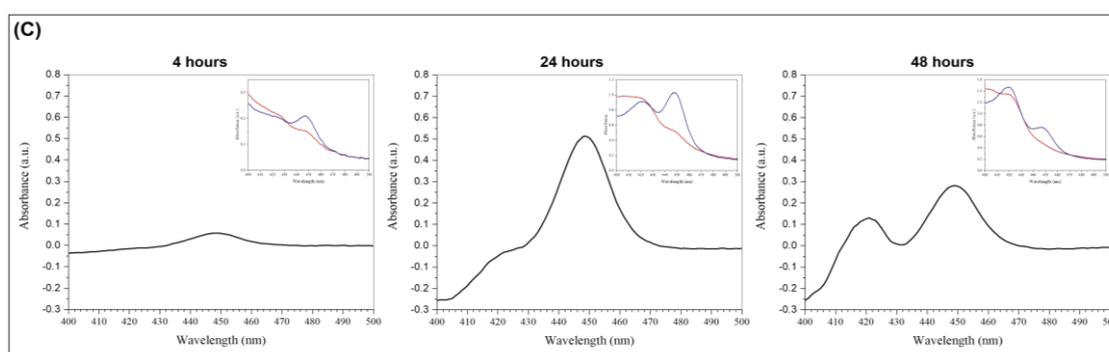
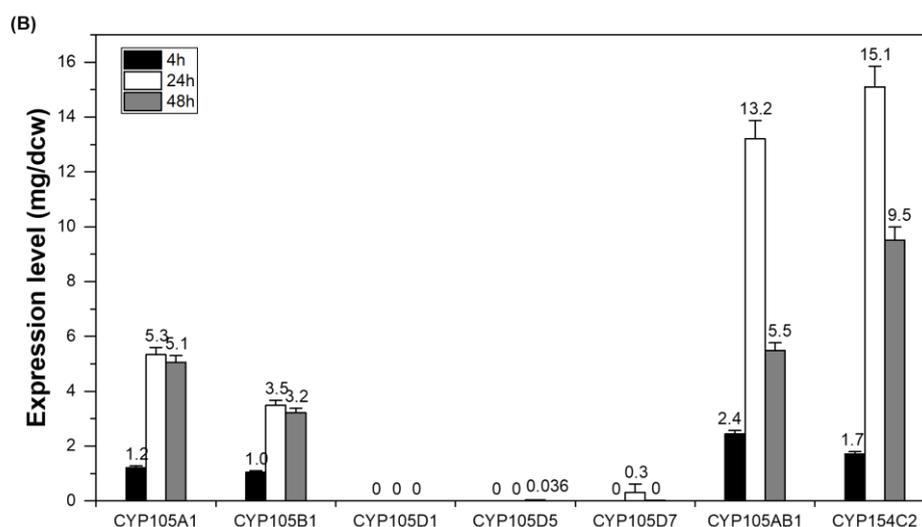


Figure 4.12. The expression level of CYPs using *cyp-fd2-fdr* operon for different expression time. Three expression periods, including 4 hours, 24 hours, and 48 hours, were tested when using *cyp-fd2-fdr* operon for producing CYPs in *E.coli*. The expression level was measured as a total molar concentration (A) or specific yield of active P450s (B). (C) is the carbon monoxide difference spectrum for CYP105AB1 that has been expressed for 4 hours, 24 hours, and 48 hours respectively. The inserted graph in each graph includes the spectrum of reduced cytochrome P450 (red) and the spectrum of CO-cytochrome P450 binding mixture (blue). The difference spectrum was calculated by subtracting the red spectrum from the blue spectrum.

As is shown in Figure 4.12 (A) and (B), the optimal expression time among the three periods is 24 hours for producing CYP105AB1. For other P450s, the specific yields for each CYP are always the highest at 24-hour expression. As is shown in Figure 4.12 (C), the carbon monoxide difference spectrum of CYP105AB1 has a much higher peak in 24-hour expression, compared to the peak in the 4-hour expression spectrum. The 48-hour spectrum shows an extra peak at 420 nm and a decreased peak at 450 nm, comparing to the spectrum of

24-hour expression. It seems the incorrectly folded CYP105AB1s has increased after the expression time lasts more than 24 hours, which means the quality of produced P450s is massively reducing at 48-hour expression.

4.4 Summary

In conclusion, there are two main objectives of the research activities carried out in Chapter 4, which are summarized below.

- The first objective was to design and test an alternative operon construct that contains a standardized P450 gene, ferredoxin gene, and ferredoxin reductase gene. Five different constructs were being successfully designed and built for the expression of CYP in *E.coli* BL21(DE3).
- The second objective was to discuss the effects of ferredoxin and ferredoxin reductase within the operon on enhancing the production of various actinomycete cytochrome P450s. The details of the expression level of CYPs using different constructs were discussed and summarized below.

Based on the construct of the original three-gene operon, an alternative operon construct was designed for the expression of Class I cytochrome P450s (Figure 4.2). The alternative construct also contains three genes that are controlled by a single T7 promoter. Comparing to the original construct, the modified features include reduction of the distances between each gene, fixed restriction sites at the beginnings and ends of each gene, and removing of redundant restriction sites along the operon. The sequential difference between the old construct and the new construct is shown in Figure 4.13. One of the main differences is the *EcoRI* site was placed between the *cyp* gene and *ferredoxin* gene in the new construct, and no restriction site in the old version. In addition, the ribosome binding sites are overlapping with *EcoRI* and *HindIII* sites in the new construct, which aims to achieve continuous and efficient usage of ribosomes for translation. The stop codon for each gene was changed to TAA, which was all followed by a thymine nucleotide. This is due to that TAAT is more efficient than the other two types of codons on terminating the translation⁽¹⁵⁸⁾. The restriction sites designed on the current construct allow modulization of each genetic component and flexible replacement. According to the design of the new construct, various

operons mentioned before, including *cyp-fd1-fdr*, *cyp-fd2-fdr*, *cyp-pdx-pdr* and *cyp-fd_sery-fdr*, were assembled. The basic frame features of these operons are the same, and what differs is just the standardized genetic component. The operon was then assembled on the plasmid backbone that being featured with ampicillin-resistant gene, pUC origin, and *lac* repressor gene. The expression plasmids were then induced in *E.coli* BL21(DE3) for CYP production.

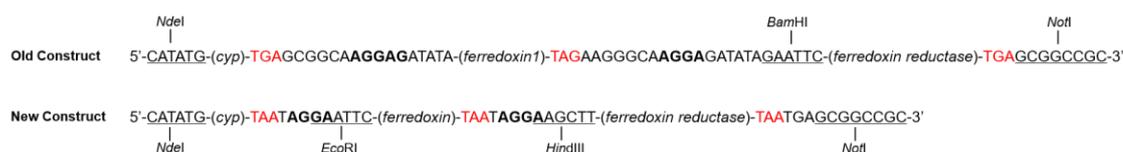


Figure 4.13. The sequential differences between the old and new constructs.

The underlined regions are restriction sites. The bold regions are ribosome binding sites. All stop codon for each gene is shown in red, and each gene is shown in italics in the brackets.

Because of the unique spectral feature of P450 when reduced and bound with carbon monoxide, the carbon monoxide assay could be used to detect and measure the amount of correctly folded cytochrome P450 being produced by host cells. The peak readings at 450 nm are used to evaluate the molar concentration of P450s, which represents the total amount of active P450 being produced at the time of harvesting. The expression levels of various P450s using different operon constructs are summarized in Table 4.4. What can be concluded from the table is that there is no guaranteed enhancement of CYP production when expressing with redox partners. It is impossible to predict whether a CYP could be produced or to predict which redox partners may have better enhancement on production. Meanwhile, it is quite convenient to put any standardized P450 gene into any of the constructs and start expression trials. The work carried out in this chapter has built a small library of constructs with consistent frame features, which provides a simple strategy on selecting constructs for CYP expression.

Table 4.5. The expression levels of various P450s using different operon constructs. There are 12 different CYPs and 5 constructs being discussed in this chapter. Each construct was assembled on the same plasmid backbone. Each

expression was conducted using BL21(DE3) strain for 24 hours in TB. The table cells that are shaded indicate the combinations of CYP and construct were not assembled or tested in this project. The expression level of active CYP is shown below as nmol per litre of culture. The highest expression achieved was highlighted in bold red, which indicated the best expression construct for each protein.

CYPs ^a		Constructs				
		<i>cyp</i> ^b	<i>cyp-fd1-fdr</i> ^c	<i>cyp-fd2-fdr</i> ^d	<i>cyp-pdx-pdr</i> ^e	<i>cyp-fd_sery-fdr</i> ^f
<i>S.griseolus</i>	CYP105A1	493.6	92.3	282.9	170.1	
	CYP105B1	60.4	414.5	471.2	301.9	
<i>S.griseus</i>	CYP105D1	1.8	15.6	0.0	0.0	
<i>S.coelicolor</i>	CYP105D5	5.5	176.1	0.9	2.3	
<i>S.avermitilis</i>	CYP105D7	76.0	35.4	27.5	0.0	
	CYP154C2	912.1	1386.9	1380.9	964.5	
<i>S.erythraea</i>	CYP105AB1	39.4	191.4	950.5	374.3	723.4
	CYP105AC1	602.6		379.1		413.9
	CYP296A1	688.6		990.8		780.2
	CYP295A1	0.0		0.0		0.0
	CYP186B1	0.0		0.0		0.0
	CYP155A2	0.0		34.8		0.0

^a All 12 different CYPs and their original organism are shown in the first two columns;

^b The single-gene construct contains only the *cyp* gene;

^c The *cyp-fd1-fdr* construct contains *cyp* gene, ferredoxin 1 gene from *S.griseolus* and ferredoxin reductase SCF15A gene from *S.coelicolor*;

^d The *cyp-fd2-fdr* construct contains *cyp* gene, ferredoxin 2 gene from *S.griseolus* and ferredoxin reductase SCF15A gene from *S.coelicolor*;

^e The *cyp-pdx-pdr* construct contains *cyp* gene, putidaredoxin gene and putidaredoxin reductase gene from *P.putida*;

^f The *cyp-fd_sery-fdr* construct contains *cyp* gene, ferredoxin gene from *S.erythraea* and ferredoxin reductase SCF15A gene from *S.coelicolor*;

CYP105AB1, as the P450 of interest, was assembled in all operon constructs for expression. The *cyp-fd2-fdr* operon was confirmed to be currently the best construct for producing CYP105AB1 in good quantity and quality (Figure 4.9 and 4.10), comparing to the use of other constructs. At the beginning of the project, it was suspected the ferredoxin (*fd_sery*) originating from the same genome of CYP105AB1 may have better performance than other ferredoxins. However, the *cyp-fd_sery-fdr* construct led to the expression of CYP105AB1 (723.4 nmol/L), which was not as high as using *cyp-fd2-fdr* construct to reach 950.5 nmol/L. The expression time for CYP105AB1 on *cyp-fd2-fdr* operon was tested to be optimal at 24 hours. For the expression of other Class I CYPs, it is not definite that *cyp-fd2-fdr* operon would be the best construct. However, it is always recommended to use the *single-cyp* construct and the *cyp-fd2-fdr* operon for expression first.

5 Chapter 5: Characterisation of CYP105AB1 from *Saccharopolyspora erythraea* NRRL2338

5.1 Introduction

Cytochrome P450s are most commonly known for their ability to use molecular oxygen to conduct hydroxylation of non-activated carbon atoms. The heme group of cytochromes P450 monooxygenases is the redox centre where the oxidation reactions happen. The reactions that happened at the redox centre require electrons that are delivered by various electron transfer components. Based on different types of electron transfer components and their structural relationship with CYPs, cytochrome P450s are divided into eight different classes⁽⁸¹⁾. The most common type is the Class I cytochrome P450, which requires the independent redox partners proteins. The electron transfer process is normally carried out by ferredoxin reductase and ferredoxin to shuttle electrons from cofactors NAD(P)H to the P450 monooxygenase. The most widely studied Class I P450 monooxygenase is the CYP101A1 (P450Cam), together with its natural electron transfer partners putidaredoxin reductase (PdR) and putidaredoxin (Pdx) from *Pseudomonas putida* DSM 50198⁽⁶¹⁾. CYP101A1 naturally catalyses camphor to produce stereospecific 5-*exo*-hydroxycamphor⁽⁶¹⁾. An interesting and rare feature of CYP101A1 is that the three genes (*cyp101a1*, *pdr*, and *pdx*) are clustered together in one operon within the *P.putida* genome. However, for most Class I cytochrome P450s like CYP105AB1 studied in this chapter, their genes are not collocated with redox partner genes in genomes. The work carried out in Chapter 4 revealed the design and optimisation of an operon construct that is similar to the natural *cyp101a1-pdr-pdx* cluster. The operon construct of *cyp105ab1-fd2-fdr* has been successfully used for stable expression of CYP105AB1 in *E.coli* BL21(DE3). Once the stable production of CYP105AB1 is achieved, further studies of CYP105AB1 could be carried out.

Since the use of Pdx and PdR has already been proven to deliver electrons from NADH to P450s other than CYP101A1^(156, 159), the CYP105AB1 activity is also reconstituted with Pdx-PdR system and cofactor NADH when screening for potential substrates in this chapter. Although the cooperation of CYP105AB1, Pdx, and PdR on transferring electrons did not exist in nature, it was assumed the Pdx-PdR could be working with CYP105AB1. For preliminary substrate

screening, a complete and working electron transferring system is enough, and the discovery and study of highly efficient electron transferring chains are not the focus in this project.

CYP105AB1, which is originated from *Saccharopolyspora erythraea* NRRL2338, was firstly annotated in a complete genome study of *S.erythraea* NRRL2338 (GenBank ID: AM420293.1)⁽¹¹⁰⁾. The complete gene sequence of CYP105AB1 is shown in section 10.1.2, Appendix B. Apart from the identification of the *cyp105ab1* gene in the *S.erythraea* genome, there has not been any characterisation on CYP105AB1 in the previous study. Because of an industrial project carried out by Dr Sebastian Schulz at Ward Lab, CYP105AB1 started to catch our interests. In the industrial project, it was aimed to find microbial CYPs that could conduct rapid oxidation of organic molecules to generate CYP derived metabolites and oxidised derivatives of drugs and agrochemicals. One of their promising candidate P450s (confidential CYP sequence) was from an actinomycete (confidential strain name). At that time, there were also other common actinomycete strains such as *S.avermitilis*, *S.coelicolor*, *S.lividan*, *S.griseolus*, and *S.erythraea* being stored at Ward Lab. Moreover, we believed there were possible homologues in these strains as well, which were as promising as the candidate CYP identified in the industrial project. Therefore, the sequence of that candidate P450 was compared against all P450s from the mentioned strains through PBLAST, and CYP105AB1 from *S.erythraea* shares the highest sequence similarity with that candidate P450. CYP105AB1 shares a 60% identity with the candidate CYP. The closest homologues in the other four strains are all sharing less than 55% protein sequence identity with the candidate CYP, details showing in Table B1 (Appendix B, section 10.2). Normally, two P450s having more than 40% amino acid identity are classified into the same family, and into the same subfamily when having more than 55% identity⁽⁸⁰⁾. The 60% sequence identity between CYP105AB1 and the candidate CYP may indicate similar protein structures or functions. Therefore, the study of CYP105AB1 was carried out in this chapter to explore more properties of this class of CYPs, which could potentially contribute to the engineering of these CYPs as well as their commercial application in rapid synthesis of CYP derived metabolites.

5.2 Expression and purification of CYP105AB1

As is discussed in Chapter 4, 6 P450s were being successfully expressed using the *cyp-fd2-fdr* operon construct, which includes CYP105A1, CYP105B1, CYP105AB1, CYP154C2, CYP105AC1, and CYP296A1. The *cyp-fd2-fdr* operon was a better construct than other designed constructs, particularly for the expression of CYP105AB1. Therefore, we intend to keep using the construct to produce CYP105AB1 and collect purified CYP105AB1 for further characterisation. However, the original *cyp-fd2-fdr* operon did not have purification tags assembled to the *cyp* gene, so a small change was first made to the operon before any expression. A 6-histidine tail was constructed to the C-terminal of *cyp* gene within the *cyp-fd2-fdr* operon, while the *fd2* and *fdr* genetic components were not changed. The reconstruction strategy involved firstly designing of reverse primers containing an extra His-tag sequence, and then PCR for amplification of the entire plasmid backbone, and finally digestion and ligation at the *EcoRI* site to close the linearized PCR products. The detailed strategy and used primers were described in Table 5.1 and Figure 5.1. Once the re-assemble was finished, the plasmids were sent for sequencing to check the correct insertion of His-tag sequence at the end of each *cyp* gene.

Table 5.1. Primers used for reconstructing *cyp_his-fd2-fdr* operon. In chapter 4, the three-gene operon was assembled to compare and evaluate the expression performance of CYPs, but it was not designed to target CYPs with purification tags. Based on the operon construct, six plasmids were assembled for CYP expression, including pQR2242 (CYP105A1), pQR2244 (CYP105B1), pQR2252 (CYP105AB1), pQR2265 (CYP105AC1), pQR2254 (CYP154C2) and pQR2267 (CYP296A1). These six plasmids were used as the DNA templates for PCR using the primers listed below. The bold TCCT sequence indicates the ribosome binding site and the underlined GAATTC is the *EcoRI* site.

Primer ID	F/R primer	Sequence 5'-3'
cyp105a1-his	Reverse	GCGCGAATTCCTATTAGTGGTGGTGGTGGTGGTGCCA GGTGACCGGGAGTTC
cyp105b1-his	Reverse	GCGCGAATTCCTATTAGTGGTGGTGGTGGTGGTGGCAGGCG ATCGGCAG

collected from the nickel column. The SDS-PAGE analysis of each purified protein is shown in Figure 5.2. The molecular weights of natural proteins were calculated as CYP105A1 (44.2 kDa), CYP105B1 (44.4 kDa), CYP105AB1 (44.5 kDa), CYP105AC1 (44.5 kDa), CYP154C2 (43.0 kDa), CYP296A1 (42.6 kDa), putidaredoxin (11.5 kDa) and putidaredoxin reductase (45.5 kDa). Considering there was an extra 6-histidine tagged to each protein, the total molecular weight of each protein was approximately corresponding to the location of each band on the gel. As is shown in Figure 5.2, each sample lane has a distinctive target protein band and other vague bands as the residue proteins from cell lysate. The distinctive bands indicated the purification through nickel column was working to deliver purified CYPs, putidaredoxin, and putidaredoxin reductase.

For all purified cytochrome P450s, the concentrations were determined through carbon monoxide assay by using the extinction coefficient of $91 \text{ mM}^{-1} \text{ cm}^{-1}$ at 450 nm in the difference spectrum⁽¹¹¹⁾. The concentration of purified putidaredoxin and putidaredoxin reductase was calculated by using an extinction coefficient of $10.4 \text{ mM}^{-1} \text{ cm}^{-1}$ at 455 nm and $10.0 \text{ mM}^{-1} \text{ cm}^{-1}$ at 454 nm respectively⁽¹⁶⁰⁾.

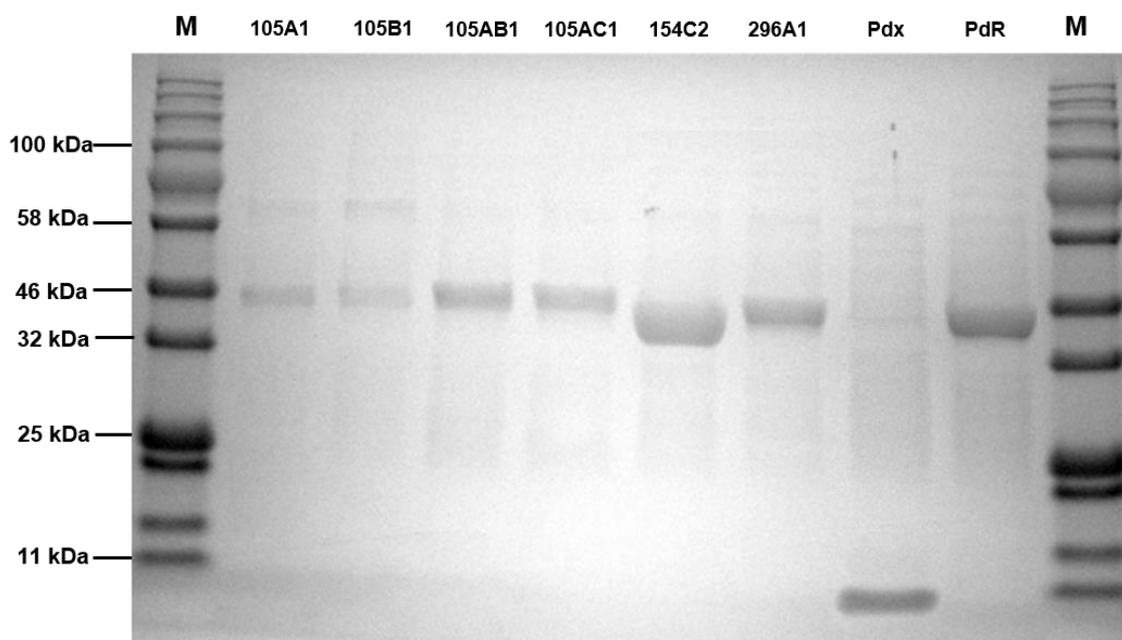


Figure 5.2. SDS-PAGE analysis of purified CYPs, Pdx, and PdR. Lane M is the marker lane with standard size protein bands. The marker is the colour pre-stained protein standards from NEB with a broad range (11-245 kDa). Six standard bands were

labeled for the lane. The sample lanes from left to right respectively represent purified CYP105A1 (44.2 kDa), CYP105B1 (44.4 kDa), CYP105AB1 (44.5 kDa), CYP105AC1 (44.5 kDa), CYP154C2 (43.0 kDa), CYP296A1 (42.6 kDa), putidaredoxin (Pdx, 11.5 kDa) and putidaredoxin reductase (PdR, 45.5 kDa).

5.3 Substrate screening for CYP105AB1

5.3.1 The use of Pdx-PdR system and cofactor NADH

The enzyme assay was designed to have 0.2 μM CYP, 5 μM Pdx and 1 μM PdR in the 50 mM potassium phosphate buffer, with a total reaction system of 1000 μL (details described in section 2.2.1). In the reaction, the combination of Pdx and PdR plays a key role in the electron transferring process, so the capability of 5 μM Pdx and 1 μM PdR on transferring electrons should be evaluated first. A cytochrome *c* assay was conducted with purified putidaredoxin and putidaredoxin reductase, which was aimed to determine how robust the Pdx-PdR system was at the current ratio (method details described in section 2.1.9). In detail, a reaction system containing 5 μM Pdx and 1 μM PdR was prepared in 50 mM potassium phosphate buffer (pH 7.5), which was used to reduce 0.1 mM cytochrome *c* from the equine heart. The reduction was started by adding NADH (1 mM at final concentration), and the absorbance at 550 nm for detecting reduced cytochrome *c* was measured. As is shown in Figure 5.3, the Pdx-PdR system with a defined molar ratio of 5:1 was proven to reduce all 0.1 mM cytochrome *c* in less than 2 minutes. Especially in the first 15 seconds after adding NADH, the absorbance at 550 nm increased rapidly, which indicates robust coordination between putidaredoxin reductase and putidaredoxin on transferring electrons. Therefore, the electron transferring system with a defined molar ratio of 5:1 (Pdx to PdR) is quite robust and suitable for enzyme assays in substrate screening.

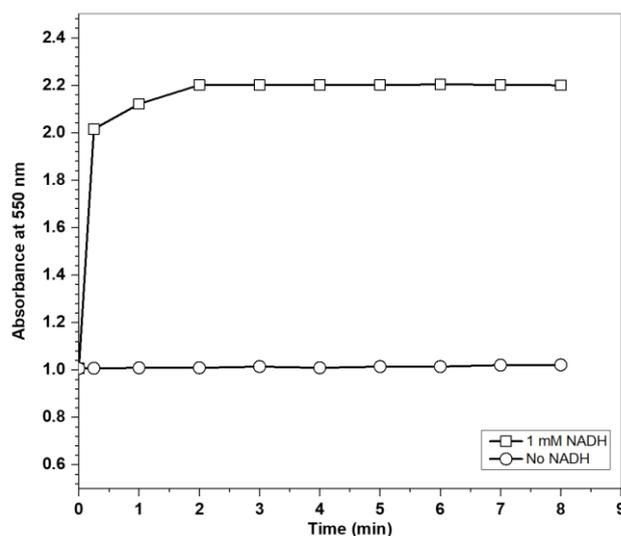


Figure 5.3. The reduction of cytochrome c by purified putidaredoxin and putidaredoxin reductase. The initial absorbance readings at time 0 min were recorded before adding any NADH to the reaction mixture. It takes approximately 2 minutes for the Pdx-PdR system to reduce all 0.1 mM cytochrome c. For reaction mixture with no addition of NADH, there is no reduction happening and the absorbance stays the same.

5.3.2 Preliminary substrate screening

In novel enzyme discovery, the straightforward approach of substrate screening is to use purified enzymes to react with a panel of substrates in a defined and stable environment⁽¹⁶¹⁾. Once the potential consumption of a substrate is observed, the reaction could be further studied on other aspects such as product characterisation, enzyme kinetics, and reaction yield. Because microbial Class I cytochrome P450 naturally relies on two partner proteins to transfer electrons from cofactors, so enzyme assay with Class I cytochrome P450 is more complex than the single enzyme reactions. In our case, the potential substrate consumption is influenced by whether the substrate could be accepted by CYP105AB1, as well as how effective the partner proteins are on delivering electrons to CYP105AB1. Therefore, the well-established Pdx-PdR system is used as the electron transferring system for CYP105AB1 at preliminary substrate screening. When CYP105AB1 was expressed in *E.coli*, the *cyp-fd2-fdr* operon was used to achieve high production of CYP105AB1. However, the Fd2-FdR

system was not considered as the electron transferring system for CYP105AB1 at substrate screening. This was due to many potential disadvantages of the Fd2-FdR system over the Pdx-PdR system. First of all, Fd2 (*S.griseolus*) and FdR (*S.coelicolor*) are originated from different organisms, unlike Pdx and PdR were coming from the same organism. It was unknown whether the Fd2-FdR system could work efficiently to obtain electrons from cofactors. Besides, it was unknown how efficiently Fd2 could deliver electrons to CYP105AB1, but there were viable applications in the reconstitution of P450 activity using the Pdx-PdR system^(155, 156), which commonly involved the use of separately purified putidaredoxin, putidaredoxin reductase and targeted P450s at defined molar concentrations. Therefore, the Pdx-PdR system was used together with purified CYP105AB1 for reaction tests with various substrates, and that is why it is necessary to have discussions in section 5.3.1 on the capability of Pdx-PdR system prepared at defined ration in this project on delivering electrons.

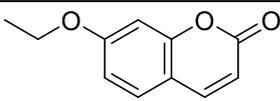
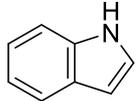
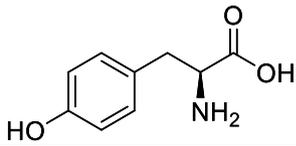
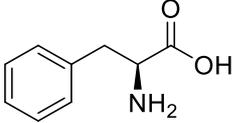
The detailed substrate screening process of using the CYP105AB1 and Pdx-PdR system is described below. As is shown in Table 5.2, a 1000 μ L reaction containing CYP105AB1, Pdx, PdR, NADH, and substrates was conducted for 24 hours. At the end of the reaction, an additional 100 μ L DMSO was added to the reaction mixture to facilitate the solubility of any potential products. Each reaction sample was filtered and finally analysed through reverse-phase HPLC. Negative control and NADH control samples were also prepared in parallel (Table 5.2), which were harvested by adding 100 μ L DMSO and then analysed through HPLC as well. The negative control sample indicated the amount of original substrate there is in the reaction system, while the NADH control sample provided the background chromatogram for the reaction system. To determine whether the reaction happens, the decrease of substrates should be detected, and new peaks should be observed on the chromatogram of reaction samples. Because the expected reaction conducted by CYP105AB1 is hydroxylation, so the products have potentially higher polarity than substrates, which should be washed out earlier than substrates through the reverse-phase C18 column. Therefore, the product peaks are expected to have earlier retention time than substrates on chromatogram.

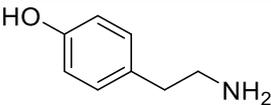
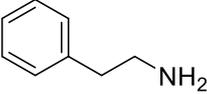
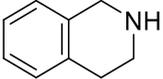
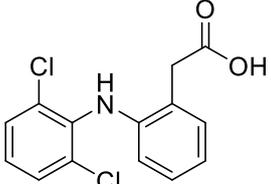
Table 5.2. Concentrations of enzyme and cofactors used in enzyme assays and controls. Both controls and reactions were prepared in 50 mM potassium phosphate (pH7.5), and the total volume of each control was set to be 1000 μ L. The concentration of each component is shown in the unit of μ M.

Sample name	CYP105AB1	Pdx	PdR	Substrate	NADH	DMSO
Reaction	0.2	5	1	200	1000	1% (v/v)
NADH control	0.2	5	1	0	1000	1% (v/v)
Negative control	0.2	5	1	200	0	1% (v/v)

As is shown in Table 5.3, a group of substrates with aromatic rings was used at the preliminary screening. The standard curve of each substrate was created before analysing any reaction mixtures through HPLC, which establish the retention time of each substrate as well as the mathematical relationship between peak area and standard substrate concentrations. The details of the standard curve for each substrate are shown in section 10.3.1, Appendix B.

Table 5.3. The substrate panel for CYP105AB1. All substrates except tyrosine and phenylalanine were dissolved in DMSO to prepare 20 mM stock solutions. For a 1000 μ L reaction system, 10 μ L of stock substrate solution was added to achieve a final concentration of 200 μ M and 1% (v/v) DMSO for the reaction system. Tyrosine and phenylalanine were prepared as 1 mM stock solutions using pH 7.5 potassium phosphate buffer. For a 1000 μ L reaction system, 200 μ L of stock solution of tyrosine or phenylalanine was used to reach a final concentration of 200 μ M.

Abbreviation	Name	Structure	MW (g/mol)	Retention time (min)
EC	7-Ethoxycoumarin		190.20	10.40
IND	Indole		117.15	10.30
TYR	Tyrosine		181.19	2.86
PHA	Phenylalanine		165.19	4.88

Abbreviation	Name	Structure	MW (g/mol)	Retention time (min)
TYA	Tyramine		137.18	3.00
PHEA	Phenylethylamine		121.18	5.48
THIQ	1,2,3,4-Tetrahydroisoquinoline		133.19	4.29
DIC	Diclofenac		318.13	10.31

All substrates were subjected to the same reaction process with CYP105AB1, and each reaction mixture was analysed through HPLC. Diclofenac is the only substrate on this panel that has positive screening results with CYP105AB1, which means additional new peaks are being observed for the diclofenac reaction. The negative screening results for other substrates are shown in section 10.3.2, Appendix B.

As is shown in Figure 5.4 (C), three peaks are being detected for the reaction sample, which are located at 1.55 min, 10.29 min, and 12.18 min respectively on the chromatogram. Comparing to the NADH control (Figure 5.4, A), they all have the NADH peak stays around 1.55 min. Comparing to the negative control (Figure 5.4, B), the peak for substrate diclofenac is shown around 12.18 min. Therefore, the new peak at 10.29 min is possible to be the products of the reaction conducted by CYP105AB1 on diclofenac. The diclofenac peak area of negative control was calculated as 27.41404 (Figure 5.4, B), while the diclofenac peak area of the reaction was calculated as 25.81758. The direct decrease of the diclofenac peak area indicates the potential consumption of substrates. According to the standard curve and standard correlation equation in Figure 5.5, the concentrations of remaining substrates in negative control and reaction mixture could be calculated, and the results are shown in Table 5.4.

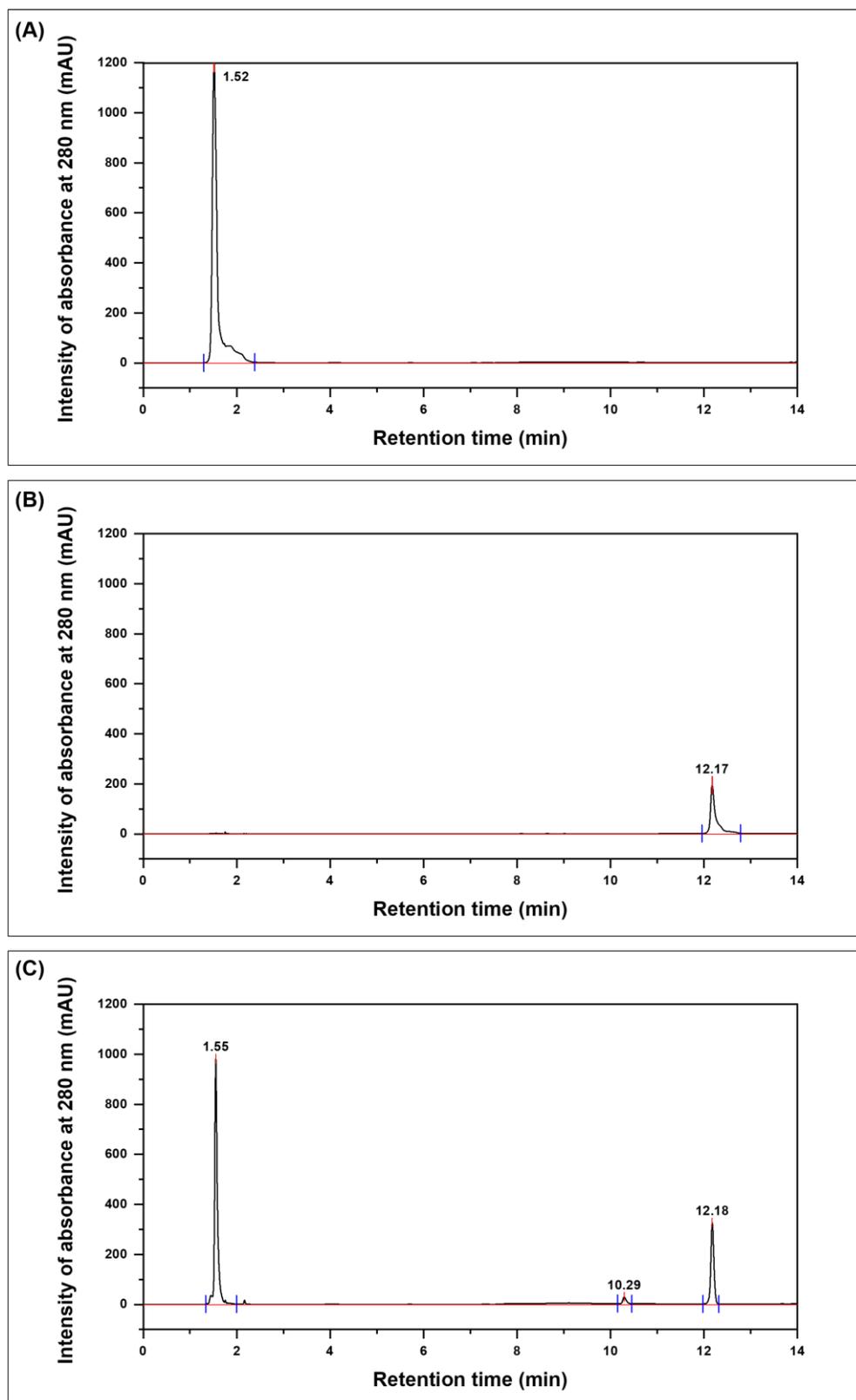


Figure 5.4. HPLC analysis of enzyme reaction conducted by CYP105AB1 on diclofenac. (A) The chromatogram of NADH control reveals only one peak at 1.52 min as NADH. (B) The chromatogram of negative control that has substrate but no NADH, which gives a peak at 12.17 min. Because there is no NADH in the negative control, so

there is no consumption of substrates. Therefore, the peak area at 12.17 represents the initial amount of 200 μM diclofenac used in the reaction system. (C) The chromatogram of reaction mixture gives three peaks at 1.55 min, 10.29 min, and 12.18 min, which represent remaining NADH, potential products, and remaining substrates respectively.

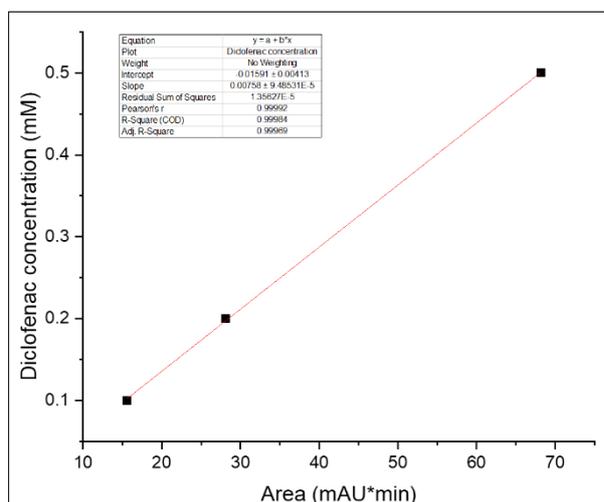


Figure 5.5. Standard correlation between diclofenac concentration and peak area on HPLC chromatogram. Three standard concentrations of diclofenac, 0.5 mM, 0.2 mM, and 0.1 mM were prepared and analysed through HPLC using method 1. The peak area has a linear relationship to standard concentrations, and the linear equation is shown as $y = 0.00758 \cdot x - 0.01591$, where y is the diclofenac concentration and x is the peak area.

Table 5.4. The calculated concentrations of initial diclofenac and remaining diclofenac in reaction with CYP105AB1. For reaction and both controls, the total volume was set to 1000 μL , and 100 μL DMSO was added to prepare samples for HPLC analysis. Therefore, the dilution factor should be considered when calculating the concentrations of diclofenac in the original 1000 μL volume.

	Negative Control	Reaction
Peak area at 12.18 min	27.41404 mAU*min	25.81758 mAU*min
Calculated concentration ($y=0.00758 \cdot x - 0.01591$)	0.1919 mM	0.1798 mM
Diclofenac concentration in 1000 μL volume	0.21109 mM	0.19778 mM
Percentage	100.0%	93.7%

Comparing to the negative control, there is a slight decrease in the calculated concentrations of remain diclofenac from 0.21109 mM to 0.19778 mM. There is only 6.3% of diclofenac loss during the reaction, which is assumed to be converted by CYP105AB1. For further study of the reaction with diclofenac, the new peak should be collected and characterised. If there is evidence indicating the new peak is the hydroxylated diclofenac, then it is confirmed that the decrease of diclofenac is due to the reaction with CYP105AB1.

5.3.3 Product characterisation

As is shown in Figure 5.5 (C), the expected product (retention time, 10.29 min) was showing a very small peak, and the peak came from 20 μ L injection using the analytical HPLC. Isolation and purification of the product peak at analytical scale were very difficult. Therefore, a semi-preparative HPLC was used to isolate the potential product peaks (details described in HPLC method 4, section 2.1.10.1). The 1000 μ L reaction containing 0.2 mM diclofenac was replicated five times, and all 5000 μ L of the reaction mixture was analysed through semi-preparative HPLC. The three peaks at 1.55 min, 10.29 min, and 12.18 min were manually collected. Each peak sample was frozen by liquid nitrogen and dried in a vacuum chamber to remove the mobile phase that was mainly water and acetonitrile. The lyophilized power for each peak was then dissolved in 50% (v/v) water and acetonitrile, and analysed by LC-MS. As is shown in Figure 5.6, the mass spectra were recorded for the NADH, diclofenac, and potential products. The calculated molecular weights for NADH and diclofenac are 664 g/mol and 296 g/mol, which correspond to the mass peak shown in Figure 5.6 (A) and (B). Because the chlorine atom has two principal stable isotopes in nature (^{35}Cl and ^{37}Cl), so there are splitting peaks at 298.1 ($M+2$) and 300.0 ($M+4$) beside the main mass peak at 295.9. The major mass peak identified in Figure 5.6 (C) has shown the mass of 312.0 and splitting peaks at 314 and 316. Comparing to the mass spectrum of diclofenac, the potential product is just 16 masses higher than the substrate, which indicates an extra oxygen atom added to diclofenac. Therefore, the proposed hydroxylation reaction happened on diclofenac catalysed by CYP105AB1.

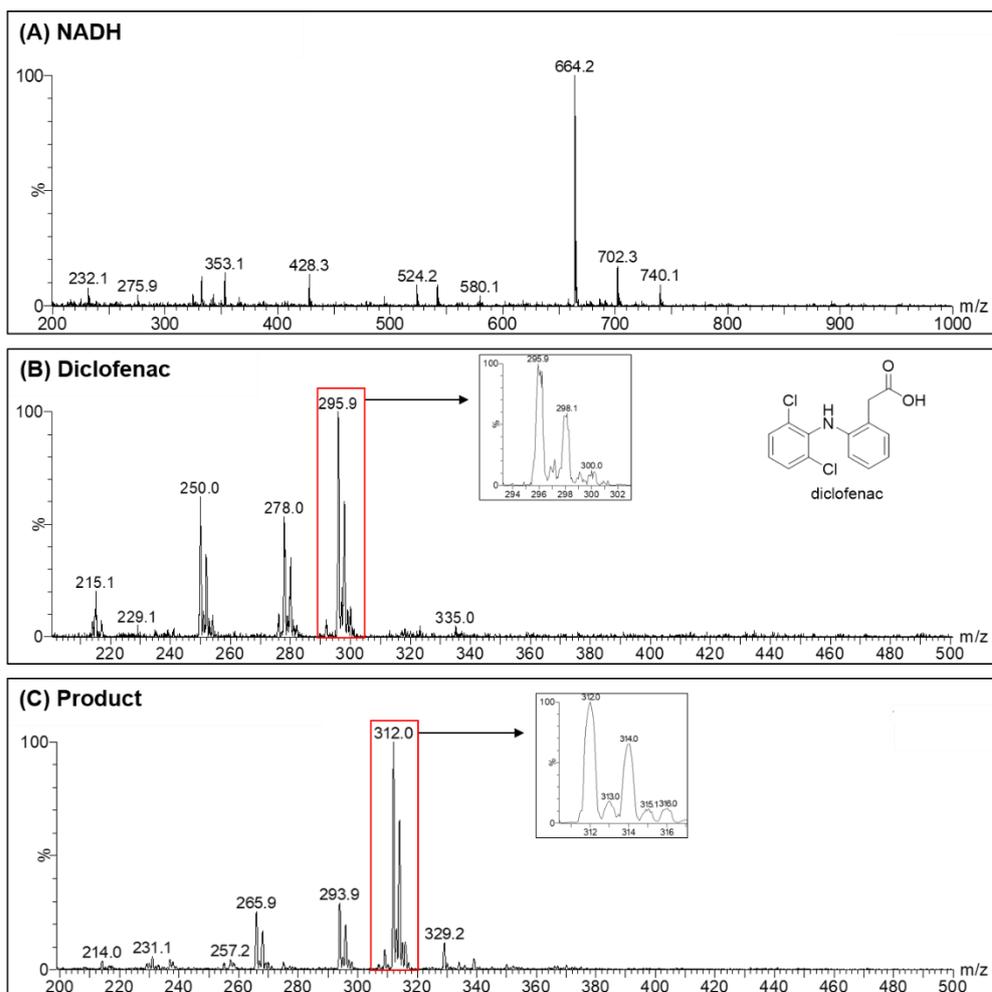


Figure 5.6. The mass spectra of NADH, diclofenac, and potential products.

The reaction with diclofenac by CYP105AB1 was analysed on HPLC, which reveals three peaks representing NADH, diclofenac as substrate, and potential products respectively. Each peak was isolated and concentrated for LC/MS. (A) The mass spectrum of NADH, which indicated a molecular weight of 664.2 g/mol. (B) The mass spectrum of diclofenac, indicating a molecular weight of 295.9 g/mol. (C) The mass spectrum of the product shows the molecular weight as 312 g/mol.

The molecular weight of the products was confirmed to be 312 g/mol by LC-MS, which indicates single hydroxylation of diclofenac. According to the diclofenac structure, there are several C-H bonds on the two aromatic rings that could potentially be activated by the insertion of a single oxygen atom. Therefore, the product was analysed through NMR for structure prediction. Firstly, the $^1\text{H-NMR}$ spectrum was recorded to determine the chemical shift patterns for all hydrogen atoms of diclofenac (Appendix B, section 10.4). The resonance frequency spectra of diclofenac and product were compared and shown in Figure 5.7. The total peak

area between δ 6 and δ 8 was around 7, which indicated there were 7 hydrogens on aromatic rings of diclofenac (Figure 5.7 (A)). In the meantime, the calculated peak area of the product was around 6, which indicated the product had 6 hydrogens directly connected to aromatic rings due to the hydroxylation reaction (Figure 5.7 (B)). As is shown in Figure 5.7 (A), the peak between δ 7.3 and δ 7.4 is the signature doublet, which corresponds to the identical hydrogens at 3' and 5' positions. On the NMR spectrum of the product in Figure 5.7 (B), the doublet peak disappeared and was replaced by singlet peak around δ 6.9. It suggested that the 4' hydrogen was replaced, which left the hydrogens at 3' and 5' to have no neighbouring hydrogens to give peak split on the spectrum. Therefore, the proposed hydroxylation should happen on 4' carbon of diclofenac, which gives product structure 4'-hydroxydiclofenac showing in Figure 5.7 (B).

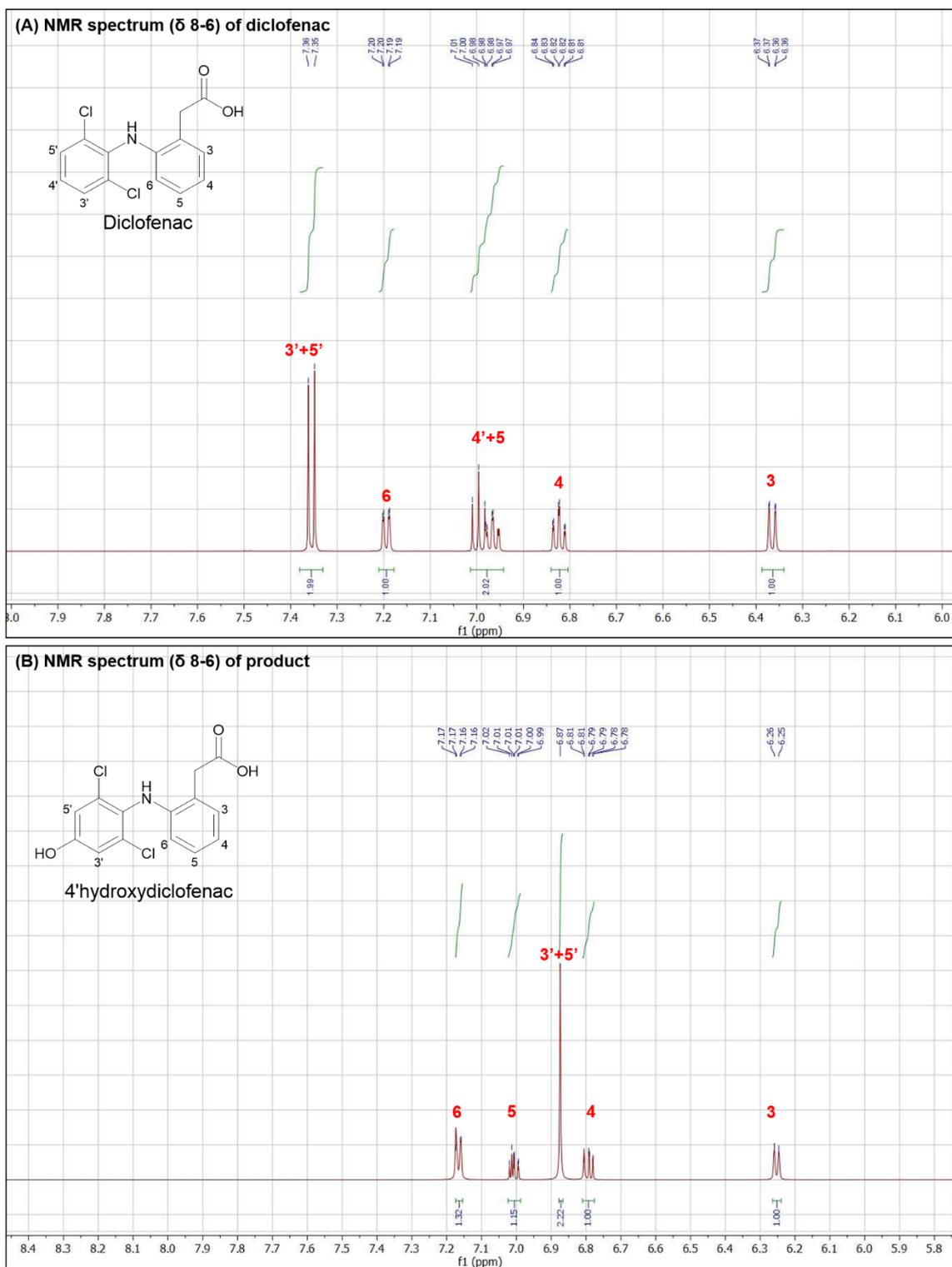


Figure 5.7. The partial spectra of diclofenac and product. (A) The peak details of diclofenac between δ 6 and δ 8. (B) the peak details of hydroxylated products between δ 6 and δ 8. The number of peak splits reveals the number of hydrogens present on neighbouring carbons. For example, the singlet indicates no hydrogen on the neighbouring carbons, and the doublet peak indicates there is one hydrogen on the

neighbouring carbons. Each peak represents a type of hydrogen within the molecule. The peak area indicates the number of hydrogens there is for the hydrogen type.

5.4 The performance of cell-free biotransformation against diclofenac

5.4.1 Bioconversion by purified CYP105AB1 and Pdx-PdR system

Since it was confirmed that CYP105AB1 could react with diclofenac, the conversion efficiency of the current reaction system was also studied. The initial diclofenac and NADH were set to be 200 μM and 1 mM and the reaction was separately harvested at 0 hour, 2 hours, 4 hours, 20 hours, 24 hours and 48 hours to monitor the consumption of diclofenac. The sample preparation was described in Figure 2.6, section 2.2.3.1. The consumption of diclofenac was evaluated and shown in Figure 5.8. The concentration of initial diclofenac was set to be 200 μM , but the actual measured concentration of initial diclofenac was around 212 μM . The main consumption of diclofenac happened in the first 4 hours, and after that, the diclofenac was slowly consumed and finally stabilized around 195 μM . Eventually, there is approximately 7% of total diclofenac being converted by 0.2 μM CYP105AB1. The decrease of the substrate in the first 4 hours was showing a linear trend, so the specific activity was calculated as below.

$$\begin{aligned}\text{Specific activity} &= \text{Consumed diclofenac} \div \text{Time} \div \text{Amount of CYP105AB1} \\ &= (212.5 - 197.0) \div (4 \times 60) \div 0.2 \\ &= 0.3 \text{ nmol/min/nmol of CYP105AB1}\end{aligned}$$

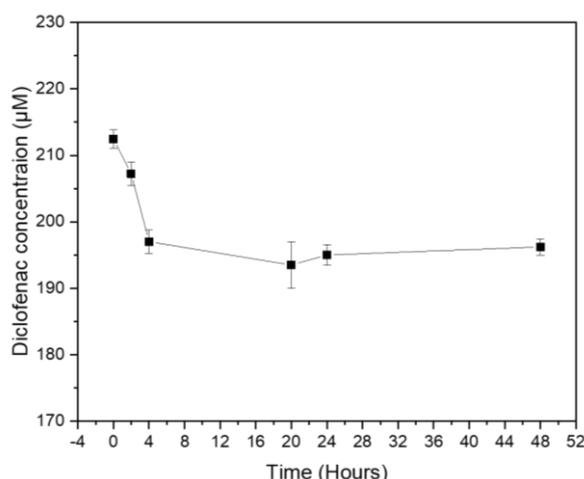


Figure 5.8. Bioconversion of diclofenac by purified CYP105AB1. The correlation between diclofenac concentration and peak area on the HPLC spectrum was established first, which was described in Figure B9. The concentration of diclofenac at each time points could be calculated by determining the peak area of remaining diclofenac on each HPLC spectrum. The concentration was calculated as the average of three repeated reactions at each time point, and the error bar represents the standard deviation among the three readings.

5.4.2 Bioconversion by cell lysate

It was clear that the purified CYP105AB1 and Pdx-PdR system could be used for diclofenac bioconversion. However, it took too much effort on the purification of each protein as well as reconstructing the reaction environment. In addition, bioconversion could also be achieved by using cell lysate, which is a more convenient approach to biotransformation involving multiple enzymes. In the case of CYP led bioconversion, there are multiple proteins involved including CYP105AB1 as well as other redox partners. Therefore, the performance of bioconversion using cell lysate that contains all protein components could be discussed here. Comparing to bioconversion using purified proteins, the molar ratio among each protein cannot be defined, and only the amount of CYP could be determined through CO assay. It is more useful to use cell lysate to quickly establish a reaction environment and produce products while using purified enzymes to establish reaction environment and study enzyme kinetics.

The aim of assembling genes of redox proteins with the *cyp105ab1* in a single operon is not only to achieve enhanced expression of CYP105AB1 within host cells but also to construct a complete electron transferring system within the cellular factory that is ready for biotransformation. There are two types of operons being constructed, including *cyp105ab1-fd2-fdr* and *cyp105ab1-pdx-pdr*, and the expression level of CYP105AB1 for each operon was discussed in chapter 4. The differences between *cyp105ab1-fd2-fdr* and *cyp105ab1-pdx-pdr* are the genes of redox proteins assembled downstream of *cyp105ab1*. Fd2-FdR system contains the ferredoxin 2 from *S.griseolus* and ferredoxin reductase SCF15A from *S.coelicolor*, while Pdx-PdR system consists of putidaredoxin and putidaredoxin reductase from *P.putida*. Neither of the two-electron transferring systems is the

natural partners for CYP105AB1, but both are confirmed for their capabilities of transferring electrons from cofactor to cytochrome c for reduction (Figure 4.4 and Figure 5.3). Therefore, the performance of the two systems assisting CYP105AB1 on bioconversion of diclofenac was investigated and discussed in this section.

The detailed sample preparation process was shown in Figure 2.7 in section 2.2.3.2. Firstly, the *cyp105ab1-fd2-fdr* operon and the *cyp105ab1-pdx-pdr* operon were separately induced and expressed in *E.coli* BL21(DE3). 20 mL of each cell culture was harvested to measure concentrations of active P450 using carbon monoxide assay. According to the CO assay methods and calculation formula described in section 2.1.8.2, the concentrations of active CYP105AB1 were calculated as 706.9 nmol (Figure 5.9, A) and 393.8 nmol (Figure 5.9, B) for every litre of cell culture.

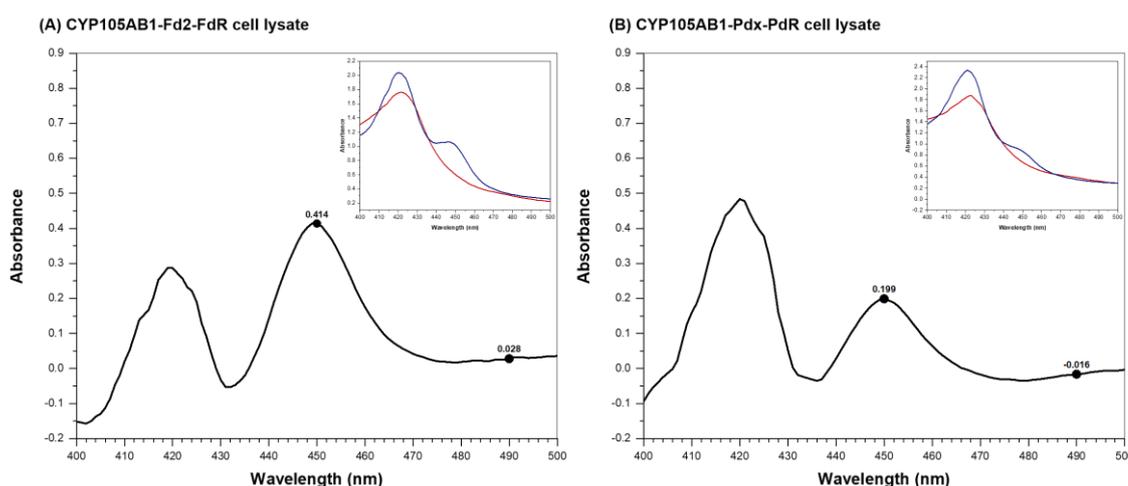


Figure 5.9. Concentrations of CYP105AB1 in two different cell lysates. The cell lysate was acquired and conducted for carbon monoxide assay (methods described in section 2.1.8). The absorbance reading was obtained for calculation of the concentration of active P450 produced by the cell culture. (A) CO spectrum of cell lysate containing CYP105AB1, Fd2, and FdR; (B) CO spectrum of cell lysate containing CYP105AB1, PdX, and PdR.

In addition, 50 mL of each cell culture was also harvested to prepare cell lysate. The cell pellet from 50 mL culture was firstly sonicated in 10 mL 50 mM potassium phosphate buffer (pH 7.5), and the filtered supernatant was used in reactions

(details of preparation described in section 2.2.3.2). The concentrations of active P450 for these two cell lysates were calculated as below:

CYP105AB1-Fd2-FdR cell lysate,

$$\frac{706.9 \text{ nmol/L} \times 50 \text{ mL}}{10 \text{ mL}} = 3.5 \mu\text{M}$$

CYP105AB1-Pdx-PdR cell lysate,

$$\frac{393.8 \text{ nmol/L} \times 50 \text{ mL}}{10 \text{ mL}} = 1.9 \mu\text{M}$$

Therefore, the composition of the reaction system with cell lysates was set below (Table 5.5). The reaction mixture was harvested after 2 hours, 4 hours, 6 hours, 20 hours, 24 hours, and 48 hours to determine the remaining diclofenac. The reaction system without cofactors was also prepared to establish the initial concentration of substrates in the reaction system. The changes in diclofenac concentration were shown in Figure 5.10.

Table 5.5. Reaction setup for bioconversion on diclofenac by cell lysates.

The concentrations of CYP105AB1 were set to be the same at 1 μM for both reactions. The natural cofactor for ferredoxin reductase (FdR) is NADPH, while the natural cofactor for putideradoxin reductase (PdR) is NADH.

Components	Bioconversion by CYP105AB1-Fd2-FdR	Bioconversion by CYP105AB1-Pdx-PdR
Cell lysate	1 μM CYP105AB1	1 μM CYP105AB1
Diclofenac	200 μM	200 μM
Cofactor	1 mM NADPH	1 mM NADH
DMSO	1% (v/v)	1% (v/v)
Buffer	50 mM potassium phosphate (pH7.5)	50 mM potassium phosphate (pH7.5)
Total volume	1000 μL	1000 μL

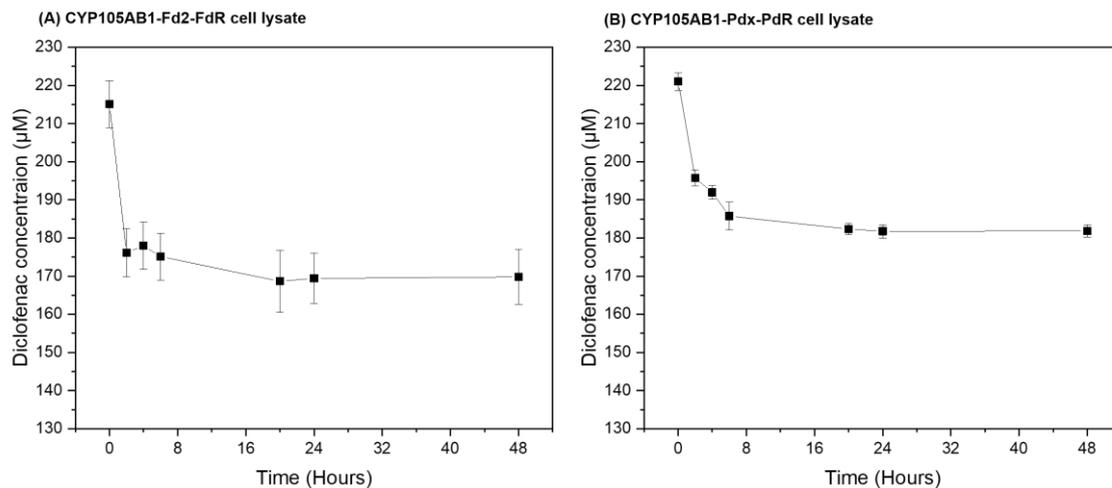


Figure 5.10. Bioconversion of diclofenac by CYP105AB1 cell lysate. The concentration of diclofenac was determined by the peak area of diclofenac on the HPLC spectrum at each time point. (A) The trend of diclofenac consumption by CYP105AB1-Fd2-FdR using NADPH. (B) The trend of diclofenac consumption by CYP105AB1-Pdx-PdR using NADH. (C) The substrate conversion rate in the first 2 hours of the reaction. The CYP105AB1-Fd2-FdR cell lysate consumed diclofenac faster than that cell lysate of CYP105AB1-Pdx-PdR.

As is described in section 5.3.2 and section 5.4.1, it is known that the commonly used Pdx-PdR system could cooperate with CYP105AB1 to achieve oxidation of diclofenac. However, it was unknown whether the Fd2-FdR system designed in this project could also work with CYP105AB1. In addition, using cell lysate to complete bioconversion is much simpler and more achievable than the bioconversion with purified enzymes. That is why the bioconversion with cell lysate was studied here. As is shown in Figure 5.10 (A), CYP105AB1 could react on diclofenac with the assist of the Fd2-FdR system. The concentration of diclofenac decreased from 215 µM and stayed at 170 µM approximately (45 µM being used), and most of the reaction happened in the first 2 hours. In Figure 5.10 (B), the Pdx-PdR system and cofactor NADH were used for the conversion of diclofenac. The concentration of diclofenac decreased from 221 µM and stayed at 181 µM approximately (40 µM being used), and most of the reaction happened in the first 6 hours. The consumption of diclofenac by CYP105AB1-Fd2-FdR system (around 45 µM) was similar to the consumption by CYP105AB1-Pdx-PdR

system (around 40 μM). However, the CYP105AB1-Fd2-FdR system used less time to complete the reaction than the CYP105AB1-Pdx-PdR system.

5.5 Construction of CYP105AB1 mutants and potential challenges

The CYP105s are commonly conserved across the actinomycetes and have been well known for their diverse involvement in biotransformation of xenobiotics and synthesis of bioactive molecules⁽¹⁶²⁾. The structure of several P450s from the CYP105 family, such as CYP105A1 from *S.griseolus*⁽¹⁶³⁾, CYP105D6 from *S.avermitilis*⁽¹⁶⁴⁾ and P450 MoxA from *Nonomuraea recticatena*⁽¹⁶⁵⁾, have been resolved, which gives more insights on catalytic characterisation. CYP105AB1 is also a P450 belongs to the CYP105 subfamily. It has been well-expressed with other electron transferring proteins in *E.coli* hosts and briefly characterised on substrate specificity in this project. As is discussed in sections above, CYP105AB1 could react on diclofenac, leading to the production of 4'-hydroxydiclofenac. In the future, enzyme engineering strategies could be introduced to explore the catalytic potential of CYP105AB1 on various substrates. The focus of this project is always the characterisation of wild type CYP105AB1, but not the structure study or rational engineering of the CYP. Therefore, only limited preliminary work on engineering CYP105AB1 was conducted and described in this section.

Before conducting any rational engineering, it would be extremely helpful if the resolved structure of CYP105AB1 has been acquired. However, there is no structure study on CYP105AB1 in this project nor will be in the foreseeable future. Therefore, the structure feature can be reasonably predicted based on the sequential comparison between CYP105AB1 and its structurally known homologues. The structure prediction of CYP105AB1 may not be complete or accurate, but it brings an insight into highly conserved regions like substrate-binding pocket.

In this section, CYP105AB1 was compared with all known P450s from the CYP105 subfamily to evaluate the sequential similarities. The calculated sequence identity of CYP105AB1 to each CYPs in the multiple alignments was displayed as a heatmap matrix (Figure B12, Appendix B, section 10.5).

involvement with the binding pocket of CYP105A1 or P450_MoxA, so the corresponding regions on CYP105AB1 were also revealed (Figure 5.11). According to the alignment, some secondary structures related to the binding pocket are well conserved, such as the helix C, helix F, helix K, β sheet 3, and the ending loop. Meanwhile, there are also some structure features like helix B' and the beginning of helix G, which varies a lot among the three CYP105s. It is suspected that these sequentially diversified regions are the reasons for different substrate specificity.

It has been reported that a library of mutants was constructed for P450_MoxA through random mutagenesis. Five single mutants were discovered due to their enhanced catalytic activity against 7-Ethoxycoumarin and diclofenac⁽¹⁶⁷⁾. The five mutants of P450_MoxA were identified as Q87W, T115A, H132L, R191W, and G294D. Three (Q87, R191, and G294) of the five mutations were involved with the residues locating at the binding pocket. The other two mutation sites (T115 and H132) were not involved with the substrate-binding pocket, and it was not clear their relationship to the enhanced activity. It is suspected the two residues at 115 and 132 may influence accelerating the electron transfer rate⁽¹⁶⁷⁾. Due to the high similarity between CYP105AB1 and P450_MoxA, the five tested mutation sites could be used as examples for the rational engineering of CYP105AB1.

After comparing the alignment between CYP105AB1 and P450_MoxA, CYP105AB1 also has conserved threonine (residue 108) and glycine (residue 287), which are corresponding sites to T115 and G294 of P450_MoxA. Therefore, two mutants for CYP105AB1 were firstly constructed as T108A and G287D. The designed primers were shown in Table 5.6, and the site-directed mutagenesis kit used as described in section 2.2.4. The two mutations were introduced through PCR with these designated primers, and the PCR products were then digested with *DpnI* to remove template plasmids, and then analysed on agarose gel (Figure 5.12). The target bands between 5-6 kb representing the linearized plasmids were extracted from the agarose gel. Circularisation of the PCR product was achieved through ligation by T4 DNA ligase. After that, the two plasmids (pQR2286 and pQR2287) containing a mutated *cyp105ab1* gene were then

prepared and sequenced to check if these single-mutations were successfully introduced.

Table 5.6. Primers designed for site-directed mutations of the *cyp105ab1* gene on pQR2281. The bold and underlined nucleotides are the ones being replaced. The changes of nucleotides led to the three mutations, T108A and G287D. All primers were designed as 5' phosphorylated primers, which were prepared for further ligation of the linearized PCR products to form the complete expression plasmids.

Mutants	Primers	Sequence 5'-3'	Codon changes
CYP105AB1-T108A	F-a322g	TCGGCGAGTTC <u>G</u> CCGTGCGCCGC	<u>ACC</u> to <u>GCC</u>
	R-a322g	GCGGCGCACGGCGAACTCGCCGA	
CYP105AB1-G287D	F-g860a	CGATCGTCGACG <u>A</u> CGCCACCTCCCG	<u>GGC</u> to <u>GAC</u>
	R-g860a	CGGGAGGTGGCGTCGTCGACGATCG	

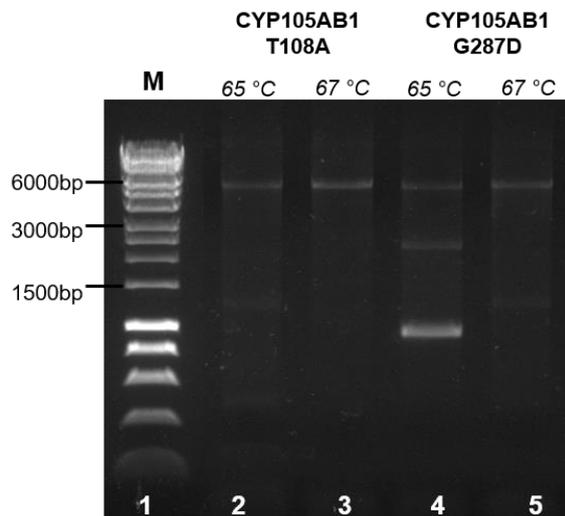


Figure 5.12. Point mutation PCR using Phusion site-directed mutagenesis kit. Two different annealing temperatures were tested. Lane 1 is displaying the 1kb DNA HyperLadder from Bioline. The other four columns from left to right are the PCR products of mutant T108A at annealing temperature 65°C (lane 2) and 67°C (lane 3), and PCR products of mutant G287D at annealing temperature 65°C (lane 4) and 67°C (lane 5). The template plasmid (pQR2281, wild type *cyp105ab1*) has a length of 5.6 kb, which contains the *cyp105ab1-fd2-fdr* operon, so the linearized PCR products should have a similar fragment length. As is shown in lane 2 to lane 5, the bands around 6kb were the linearized PCR products, which were used for further ligation to prepare plasmids pQR2286 (T108A) and pQR2287 (G287D).

After introducing the two mutations, the wild type CYP105AB1 as well as the mutants T108A and G287D were expressed in *E.coli* BL21 (DE3). All three CYP105AB1s were expressed together with Fd2 and FdR, and the general expression conditions were the same (described in section 2.1.7.1). The carbon monoxide assay was conducted to evaluate the quality of produced wild type CYP105AB1 and the other two mutants (Figure 5.13). As is shown below, there were weakly detectable peaks at 450 nm on the CO difference spectra for the two mutants (Figure 5.13, B, and C), which indicates most mutants were produced as the structurally inactive form in *E.coli* hosts. Meanwhile, the wild type CYP105AB1 had a detectable peak at 450 nm, indicating the expression of structurally active CYP105AB1 at an observable level.

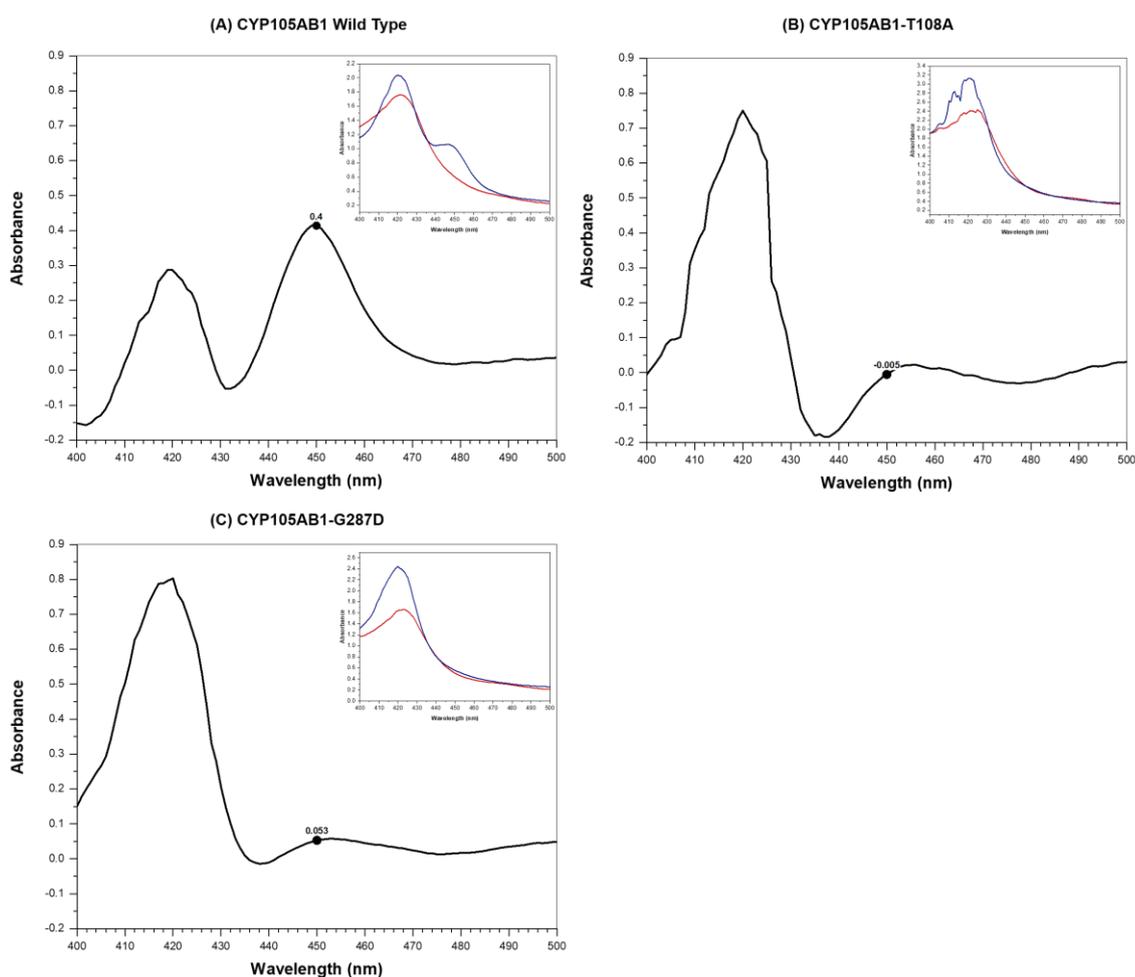


Figure 5.13. Carbon monoxide assay for CYP105AB1-wt, CYP105AB1-T108A, and CYP105AB1-G287D. All three CYP105AB1s were induced by IPTG and expressed for 24 hours. Only 20 mL of final cell culture was harvested and used for carbon monoxide assay. (A), (B) and (C) are respectively displaying the difference CO

spectra of CYP15AB1-wt, CYP105AB1-T108A, and CYP105AB1-G287D. In each graph, the main curve (black) is the difference spectrum generated by subtracting the spectrum of reduced cytochrome P450 from the spectrum of CO-P450 binding mixture. The insert graph in each include the spectrum of reduced cytochrome P450 (red) and the spectrum of CO-cytochrome P450 binding mixture (blue).

According to these difference spectra, all mutants were highly expressed as the inactive P420 form, while the wild type CYP105AB1 had both active P450 form and inactive P450 form. It is generally accepted that the spectral shift from 450 nm to 420 nm is due to the protonation of the thiolate-ferric iron coordination, resulting in a thiol-ferrous iron chelating⁽¹⁹⁾. Moreover, the generally accepted cytochrome P450 catalytic cycle has revealed that the activation of molecular oxygen starts from the reduction of ferric iron. Therefore, the two mutants (CYP105AB1-T108A and CYP105AB1-G287D) are regarded to be enzymatically inactive, comparing to the wild type CYP105AB1. Before any further characterisation of the two mutants, the expression of mutants should be optimised first to produce mutants with the correct cysteine-heme coordination.

Because there was no previous research reporting the peak shifts from 450 nm to 420 nm for the P450_MoxA mutants T115A and G294D, so it is difficult to determine how these two residues of threonine at 115 and glycine at 294 could be related to the confirmation of the heme ligand assembled in cytochrome P450. In our case, site mutations of CYP105AB1, T108A and G287D, were designed according to the mutations (T115A and G294D) constructed on P450MoxA. However, both CYP105AB1-T108A and CYP105AB1-G287D were showing complete peak shifts in their CO difference spectra, indicating the production of inactive form only of both mutants. The microbial hosts and expression conditions used were the same to produce the wild type and the mutants, and the wild-type CYP105AB1s were produced as the active form. Therefore, other approaches should be applied to the expression optimisation of these two mutants CYP105AB1-T108A and CYP105AB1-G287D, which may involve the use of different host strains, changing of expression conditions or co-expression of chaperone proteins. Considering the limited time left for this project, the mutant construction of CYP105AB1 as well as expression optimisation of mutants were left for future work.

5.6 Summary

In conclusion, there are three main objectives of the research activities carried out in Chapter 5, which are summarized below.

- The first objective was to produce CYP105AB1 using the operon construct designed in Chapter 4, and the purification of CYP105AB1 was also carried out to prepare CYP105AB1 for substrate screening.
- The second objective was to test the activities of CYP105AB1 on 8 small aromatic compounds, and the reaction mixture was analysed using reverse-phase HPLC. The substrate screening was successfully carried out, which revealed diclofenac as the only substrate could be accepted by CYP105AB1 in the substrate panel.
- The last objective was to characterize the product of reaction on diclofenac, as well as to evaluate the reaction dynamics of the cell-free reaction with defined concentrations of CYP105AB1 and redox partners.

The expression plasmid containing a three-gene operon was designed and built for microbial CYP expression, and the performance of these plasmids was studied in Chapter 4. Several three-gene operons were being constructed within the expression plasmid, including the *cyp-fd1-fdr* operon, the *cyp-fd2-fdr* operon, the *cyp-fd_sery-fdr* operon, and the *cyp-pdx-pdr* operon. The basic features of these operons were designed to be the same, including the distance between genes, the choices of restriction sites, as well as the positions of ribosome binding sites. The differences among these operons were just the ferredoxin and ferredoxin reductase genes being assembled. By comparing the performance of each operon, what we were trying to achieve is to assemble the *cyp105ab1* gene within one operon for stable expression of high-quality CYP105AB1. The operon containing the genes of ferredoxin 2 and ferredoxin reductase was proven to be the best construct for the production of CYP105AB1.

The CYP105AB1 studied in this chapter was produced by using *cyp-fd2-fdr* operon, and the purified CYP105AB1 was used for enzyme characterisation as well as substrate screening. The purification of CYP105AB1 was achieved through affinity chromatography, capturing the histidine tags attached to each

CYP105AB1. In Chapter 4, the carbon monoxide assay was conducted to reveal the spectral property at 450 nm of CYP105AB1, which indicated they belonged to the family of cytochrome P450s. In Chapter 5, the characterisation of CYP105AB1 was carried out by using SDS-PAGE for the determination of the approximate molecular weight of CYP105AB1 (Figure 5.2). The calculated molecular weight of CYP105AB1 is around 44.5 kDa, and the band on gel confirms the correct size of the purified CYP105AB1.

Other than CYP105AB1, the putidaredoxin (Pdx) and putidaredoxin reductases (PdR) were also purified using an affinity column. These two partner proteins were prepared as one of the components used in enzyme assays. The natural function of PdR was transferring electrons from NADH to the redox centre of Pdx, and Pdx serves as the shuttle between redox centres of PdR and cytochrome P450 (CYP101A1 to be specific, a camphor monooxygenase), which completed the chain of electrons used in oxidation reactions. In previous research, the combination of Pdx and PdR was commonly applied in reactions conducted by CYPs, and the molar ratio between Pdx and PdR varied a lot in different research. The molar ratio between Pdx and PdR used in this study is 5:1, which was designed to have more Pdx than PdR in the electron transfer chain to ensure the efficiency of electron delivery from one partner protein to the next. The 5:1 ration of Pdx to PdR was firstly tested for their efficiency in transferring electrons for cytochrome c reduction. A total amount of 100 μ M cytochrome c was reduced using the combination of 5 μ M Pdx and 1 μ M PdR as a shuttle for electron transferring from NADH, and all cytochrome c was used up in 15 seconds after adding NADH to protein mixture. It was indicated that Pdx and PdR were fully capable of completing the electron chains from NADH to target proteins. Therefore, it is reasonable to use Pdx, PdR, and NADH as the electron-donating system for CYP105AB1 conducted reactions.

The process of substrate screening was firstly identifying what kind of substrate could be accepted by CYP105AB1 and then characterizing the product of reaction as well as enzyme kinetics. In this project, a preliminary substrate pool was established for those CYP105AB1 conducted reactions, and the reactions were carried out in phosphate buffer (pH 7.5) with defined concentrations of

purified CYP105AB1, Pdx, PdR, and NADH. The reaction mixture was analysed through the reverse-phase column on HPLC to reveal the amount of substrate being potentially catalysed by CYP105AB1. The preliminary substrate pool in this project included 7-ethoxycoumarin, indole, tetrahydroisoquinoline, tyrosine, tyramine, phenylalanine, phenylethylamine, and diclofenac. After comparing the chromatogram between reactions and their negative controls, only diclofenac was shown to be consumed in the CYP105AB1 conducted reactions. In addition to the substrates (peak at 12.17 min), an unknown peak at 10.29 min on the chromatogram was recorded, which may be the product. To confirm the CYP105AB1's activity against diclofenac as well as investigate the chemical structure of the potential product, the potential product peak was isolated and analysed through LCMS. After comparing the mass spectra of diclofenac and potential product, the product was one oxygen atom heavier than diclofenac on the molecular weight level. Therefore, the hydroxylation on diclofenac was confirmed and the unknown new peak at 10.29 min was indeed the product peak. To identify where the hydroxyl group was inserted on diclofenac, the products were scanned and structurally characterised through NMR, which revealed the hydroxylation was happening on 4' carbon of diclofenac (Figure 5.14).

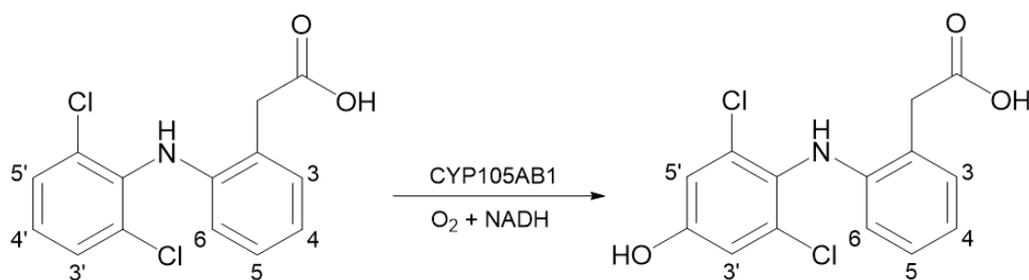


Figure 5.14. Proposed hydroxylation on diclofenac conducted by CYP105AB1. One hydroxyl group at 4' carbon of the product was confirmed by LCMS and NMR. By comparing the ¹H-NMR spectra of diclofenac and the product, the hydrogens on 3' and 5' carbon in the product had higher chemical shifts than that in diclofenac, which indicates the change of hydrogen atoms on 4' carbon.

Since diclofenac could be converted by CYP105AB1, the time-course kinetics of the reaction was also studied. The reaction containing 0.2 μM CYP105AB1, 5 μM putidaredoxin, and 1 μM putidaredoxin reductase was monitored at 2 hours, 4 hours, 8 hours, 20 hours, 24 hours and 48 hours. There was only around 7% of

diclofenac being converted, which mainly happened in the first 4 hours. The calculated specific activity for the first 4 hours was 0.3 nmol/min/nmol of CYP105AB1. In this project, there were no further kinetic studies on CYP105AB1 on the reaction with diclofenac due to the limited time left. For future kinetic studies, the strategy could include setting up various concentrations of diclofenac to react with 0.2 μ M CYP105AB1, 5 μ M putidaredoxin, and 1 μ M putidaredoxin reductase, and determination of the relationship between substrate concentrations and various reaction rates.

The original design of the three-gene operon studied in Chapter 4 was not only constructed for higher or more stable expression of CYPs in *E.coli*, but also aimed to prepare the *E.coli* cell lysate with a complete electron-donating system. Thus, for future catalytic reactions of potential chemicals, the cell lysate could be directly used rather than preparing the three purified proteins separately. The concentration of CYPs in cell lysate could be measured through carbon monoxide assay, but the concentrations of the other two-electron transferring proteins could not be directly measured. If our primary focus is to deliver enzyme conducted reactions rather than investigating the efficiency of electron transfer, the actual amount of electron transfer protein does not need to be determined. In this project, two sets of electron transferring systems, Fd2-FdR and Pdx-PdR, were expressed with CYP105AB1. The diclofenac conversion was conducted with either CYP105AB1-Fd2-FdR cell lysate or CYP105AB1-Pdx-PdR cell lysate, and the concentrations of CYP105AB1 for both reaction mixtures were adjusted to 1 μ M. As is shown in Figure 5.11 (A) and (B), the reaction rate of CYP105AB1-Fd2-FdR was higher than that of CYP105AB1-Pdx-PdR on diclofenac. Therefore, it is proven that the constructed *cyp-fd2-fdr* operon not only facilitate the enhanced expression of CYP105AB1 in *E.coli* but also provide a complete and usable electron-donating platform for potential chemical reactions.

It was believed that CYP105AB1 might have a broad substrate specificity just like its homologues P450_MoxA. However, in the preliminary screening, it was found that CYP105AB1 only react with diclofenac at a very low rate, which was expected performance as wild type cytochrome P450. What we can conclude from the current study is that the substrate pool is limited and expanding of the

substrate panel is necessary. In addition, the direction evolution of CYP105AB1 was also an approach of discovering robust enzyme candidates. In this chapter, directed engineering of CYP105AB1 was also applied to create single site mutations that were proposed to result in a higher conversion rate on diclofenac than wild type. Because there was not a known structure for CYP105AB1, so those structurally defined homologues from the CYP105 subfamily were used to determine the substrate-binding pocket of CYP105AB1. In this case, the sequences of homologues CYP105A1 and P450_MoxA were aligned with CYP105AB1, and they shared 46.3% and 57.6% identity respectively with CYP105AB1. There have been various mutation sites being studied in the pocket binding regions of CYP105A1 and P450_MoxA, and some of the corresponding sites also exist on CYP105AB1. For example, the conserved glycine (residue 294) on P450_MoxA was changed to aspartic acid, which led to enhanced activity on diclofenac. The conserved glycine was residue 287 on CYP105AB1, and the mutant G287D was created. Another mutant T108A of CYP105AB1 was created because the T115A of P450_MoxA was studied and shown enhanced activity on diclofenac as well. Two mutated genes of CYP105AB1 were assembled into *cyp-fd2-fdr* operon and checked for integrity by sequencing. The expression of CYP105AB1-T108A and CYP105AB1-G287D was conducted in the same conditions applied for the expression of wild-type CYP105AB1. However, the mutants T108A and G287D were all expressed as P420 form, which was quite different from the active P450 form for wild type CYP105AB1. The expression of T108A and G287D has been conducted for at least three times, but all results in the inactive form of two mutant proteins. It was quite interesting that the change of one residue may influence the folding structure of CYP105AB1. Since no active mutants were being produced, the reaction of diclofenac with two mutants was not conducted in this project, and there was no conclusion on whether the mutants might have enhanced activity on diclofenac. Due to the limited time for this project, further investigation of the inactive mutants as well as attempts of other expression approaches (described in detail in Chapter 6) for the production of active mutants could be included in the future.

6 Chapter 6: Identification and functional analysis of cytochrome P450s discovered in *Streptomyces rishiriensis* DSM 40489

6.1 Introduction

6.1.1 Background and project aims

In Chapter 4, the design and optimisation of a three-gene operon were conducted to increase the heterologous expression of microbial cytochrome P450s in *E.coli*, and the expression of CYP105AB1 was performed using the *cyp-fd2-fdr* operon. In Chapter 5, a group of substrates was tested with CYP105AB1 for identification and characterisation of potential activity. The work in Chapter 4 and Chapter 5 have established a working operon as well as a substrate screening strategy for the coming chapters. In this chapter, the same strategy was applied to analyse the genomic data of *Streptomyces rishiriensis* DSM 40489 for the discovery of novel CYPs. Comparing to other *Streptomyces* species such as *Streptomyces coelicolor* and *Streptomyces avermitilis*, the full genome sequence of *Streptomyces rishiriensis* was not revealed, and we did not know too much about the CYPs from *Streptomyces rishiriensis*. For all other sequentially well-characterized *Streptomyces* species like *Streptomyces avermitilis* and *Streptomyces coelicolor*, there is a rich pool of catalytic enzymes in them, which leads to the discovery of various novel CYPs and synthetic pathways. This project aimed to identify all potential CYPs within the *S.rishiriensis* genome for future enzyme mining, as well as testing the established operon construct for potential CYP production in *E.coli*. We were especially interested in one of the known CYPs from *S.rishiriensis*, which is known as Sri_CYP24 in this project. It was proposed in previous research that Sri_CYP24 was involved in the coumermycin A1 synthetic cluster for β -hydroxylation of tyrosine⁽¹⁶⁸⁾. Functionalisation of tyrosine was needed and studied in other projects within Ward Group for the development of synthetic pathways and enzyme cascades, so the study of CYPs from *S.rishiriensis* was carried out here to provide more information on CYP expression and their potential catalytic ability.

6.1.2 *Streptomyces rishiriensis* cultivation and genome sequencing

The freeze-dried *Streptomyces rishiriensis* cell samples (DSM 40489) were purchased from Leibniz Institute DSMZ-German Collection of Microorganisms and Cell Culture, which was originally isolated from soil samples in Japan. *S.*

rishiriensis was cultivated in baffled shake flask for 72 hours at 17°C and 180 rpm. The genomic DNA was extracted from the cell pellet by using GeneJET Genomic DNA Purification Kit from Thermo Fisher Scientific, which was sent to GATC biotech for Illumina sequencing. The Illumina data package was sent back in two weeks after the GATC sequencing service. Both the cultivation of *S.rishiriensis* and the extraction of genomic DNA were carried out by Mr Thomas Hickman in the Department of Biochemical Engineering at University College London. The data package was provided by GATC, containing 2970 contigs with different DNA length varying from 52 bp to 372251 bp, and the entire genomic data has a 68.4% GC content. The whole-genome data of *S.rishiriensis* was revealed using next-generation sequencing techniques for the first time and digitally preserved at Ward Group (Department of Biochemical Engineering, UCL) for mining of catalytic enzymes and pathways. All contigs provided by the sequencing company were shown as linear DNA without any annotation.

6.2 Bioinformatic analysis of *S.rishiriensis* genome

6.2.1 Discovery of cytochrome P450s

The genomic DNA of *S.rishiriensis* was used for the mining of cytochrome P450. There is a widely accepted strategy for the discovery of novel cytochrome P450, which generally refers to using functionally known P450s to find other homologs. In this chapter, the goal is to find out all cytochrome P450s that exist within the current *S.rishiriensis* genomic data package. Therefore, the approach I was taking was to use the consensus sequence of the heme-binding motif (conserved cysteine residue in hydrophobic pocket) rather than the complete cytochrome P450 protein sequence as the query sequence. The cytochrome P450 entry was found in the Pfam database, which contained 85020 different cytochrome P450s and generated a consensus of the heme-binding motif. The conserved sequence of the heme-binding motif was revealed to be **PFGAGPRNCIGERLA**, which could be simplified as **FXXGXXXCXG**. The bold phenylalanine, glycine, cysteine, and glycine are crucial and highly conserved residues for the motif sequences, which combine to represent the heme-binding region.

The consensus sequence was compared against the genomic DNA sequence through tBLASTn for protein sequence alignments, which generated multiple DNA regions along the genome that were regarded as the coding sequences for a heme-binding region. Each of the DNA regions was manually checked to determine if these heme-binding regions were located within an open reading frame. The checking process started from searching for usable ribosome binding site (5'-AAGGAGG-3') and start codon (5'-ATG-3' or 5'-GTG-3') upstream as well as a stop codon (5'-TAG-3', 5'-TGA-3' or 5'-TAA-3') downstream of the located heme-binding region. Once the coding region was manually defined from a start codon to a stop codon, the DNA sequence was translated into a protein sequence by the ExPASy Translate tool. Each of the protein sequences was sent to the Pfam database for protein definition and classification. Only the protein sequences, which were determined as a cytochrome P450 through Pfam domain analysis, were recorded. The protein sequences of all putative P450s were listed in section 11.1 Appendix C. As is shown in Table 6.1, it is the summary of all cytochrome P450s identified from the *S.rishiriensis* genome.

Table 6.1. 24 putative cytochrome P450s discovered from *Streptomyces rishiriensis* genome. The detailed positional information of each cytochrome P450 gene on genomic contigs was recorded. The protein length and molecular weight were calculated by using the ExPASy ProtParam tool. The contig number was originally assigned by the sequencing company GATC, and the numbering system was preserved at Ward Lab.

Protein ID	Size (aa)	Molecular Weight (Da)	Contig N.O. and CYP position
Sri_CYP01	413	45360.75	Contig 3, 66903-65662
Sri_CYP02	496	55119.64	Contig 6, 81356-79866
Sri_CYP03	405	45100.13	Contig 35, 83688-84905
Sri_CYP04	469	50570.74	Contig 54, 153212-154621
Sri_CYP05	410	44396.67	Contig 66, 124758-125990
Sri_CYP06	420	45939.00	Contig 66, 123499-124761
Sri_CYP07	398	43886.16	Contig 70, 24556-25752
Sri_CYP08	448	48640.89	Contig 73, 25599-26945
Sri_CYP09	453	49828.15	Contig 76, 10689-9328
Sri_CYP10	404	45032.47	Contig 76, 14197-12983
Sri_CYP11	452	49989.47	Contig 76, 9251-7893

Protein ID	Size (aa)	Molecular Weight (Da)	Contig N.O. and CYP position
Sri_CYP12	394	43031.04	Contig 76, 6712-7896
Sri_CYP13	412	45960.87	Contig 79, 19541-20779
Sri_CYP14	406	45155.62	Contig 88, 11218-12438
Sri_CYP15	399	43568.50	Contig 88, 9955-11154
Sri_CYP16	402	44090.59	Contig 100, 46182-44974
Sri_CYP17	422	46162.50	Contig 125, 9552-8284
Sri_CYP18	412	45181.54	Contig 126, 58464-59702
Sri_CYP19	395	42968.73	Contig 126, 82617-83804
Sri_CYP20	411	44909.13	Contig 127, 31718-30483
Sri_CYP21	517	56780.59	Contig 127, 57998-59551
Sri_CYP22	456	50169.63	Contig 136, 4548-5918
Sri_CYP23	431	47855.35	Contig 145, 102836-101541
Sri_CYP24	407	45059.90	Contig 176, 30890-29667

6.2.2 Features of cytochrome P450s from *S.rishiriensis*

These cytochrome P450s were initially searched based on the heme-binding motif, and other characteristic motifs like the GXXTT motif and EXXR motif were also identified to further confirm that the selected putative proteins belonged to cytochrome P450 family. As is shown in Table 6.2, both glycine (G) and cysteine (C) residues (bold and underlined in the table) are consistently conserved at the heme-binding motif for most P450s found in *S.rishiriensis*, except that Sri_CYP13 has an additional glycine. There will normally be a charged residue between the conserved glycine and cysteine at the heme-binding motif, which also applies to all P450s discovered in the *S.rishiriensis* genome. The charged residues are normally positively charged amino acids such as histidine (H), arginine (R) and lysine (K).

The threonine group of the GXXTT motif in I-helix has been implicated to play an important role in dioxygen activation and proton transfer during catalytic turnover⁽¹⁶⁹⁾. The GXXTT motif for *S.rishiriensis* P450s is also summarized in Table 6.2, where Sri_CYP04 does not have an identified GXXTT motif and Sri_CYP08 has a motif without typical threonine (T) residues conserved. Therefore, the GXXTT motif is not conserved for all cytochrome P450s.

The EXXR motif in K-helix is another commonly conserved region for cytochrome P450s. The motif feature for all *S.rishiriensis* P450s is listed in Table 6.2 as well, where Sri_CYP02, Sri_CYP04, and Sri_CYP08 all have no conserved glutamic acid (E) or arginine (R) residues.

Table 6.2. A comparison of motif sequences in I-helix and K-helix of putative cytochrome P450s found in *S.rishiriensis*. Apart from the heme-binding motif (GXXXCXG), there are two other conserved regions on cytochrome P450. The first threonine of the GXXTT motif in I-helix, glutamic acid, and arginine of EXXR motif in K-helix, as well as cysteine in the heme-binding motif are the conserved signature residues. The alignment program was ClustalW and the detailed alignment profile was shown in Appendix C Figure C1.

Protein ID	GXXTT Motif in I-helix	EXXR Motif in K-helix	Heme-binding motif
Sri_CYP01	G ²⁵¹ HHTT ²⁵⁵	E ²⁸⁹ LLR ²⁹¹	H ³⁵¹ LAFGHG <u>PHQC</u> PG ³⁶³
Sri_CYP02	A ²⁶² YENT ²⁶⁶	Unidentified (Q ³⁰⁰ LSW ³⁰³)	H ³⁶⁰ LAFGSG <u>PHEC</u> PG ³⁷²
Sri_CYP03	G ²⁴⁸ HEAT ²⁵²	E ²⁸⁶ LMR ²⁸⁹	H ³⁴⁷ ISFSAG <u>IHYC</u> IG ³⁵⁹
Sri_CYP04	Unidentified (T ²⁶⁴ SDHV ²⁶⁸)	Unidentified (R ³⁰¹ VVM ³⁰⁴)	H ³⁶² LSWGAG <u>LRQC</u> PA ³⁷⁴
Sri_CYP05	G ²⁴⁶ FETT ²⁵⁰	E ²⁸⁴ TLR ²⁸⁷	H ³⁴⁷ LAFGIG <u>VHRC</u> IG ³⁵⁹
Sri_CYP06	G ²⁴² YEPT ²⁴⁶	E ²⁸⁰ VLH ²⁸³	H ³³⁹ LSWGAG <u>AHAC</u> PA ³⁵¹
Sri_CYP07	G ²³¹ HEVSTT ²³⁷	Q ²⁷⁰ LLR ²⁷³	H ³³⁷ LAFGAG <u>PHYC</u> LG ³⁴⁹
Sri_CYP08	G ²⁵⁰ NEPL ²⁵⁴	Unidentified (D ²⁸⁴ TLW ²⁸⁷)	H ³⁴² LAWGAG <u>PHAC</u> PS ³⁵⁴
Sri_CYP09	G ²⁶⁷ SETT ²⁷¹	E ³²² ALR ³²⁵	L ³⁹² LPGFAG <u>SHKC</u> IG ⁴⁰⁴
Sri_CYP10	G ²⁴⁴ HETT ²⁴⁸	E ²⁸² LLR ²⁸⁵	P ³⁴³ LSFGGG <u>PHYC</u> IG ³⁵⁵
Sri_CYP11	G ²⁶⁶ VETT ²⁷⁰	E ³²¹ TLR ³²⁴	Y ³⁹¹ LPGFAG <u>SRKC</u> IG ⁴⁰³
Sri_CYP12	G ²³² YETA ²³⁶	E ²⁷⁰ MLR ²⁷³	H ³³² TAFGYG <u>IHHC</u> LG ³⁴⁴
Sri_CYP13	G ²⁵⁵ NETT ²⁵⁹	E ²⁹¹ IVR ²⁹⁴	H ³⁵² LGFGGG <u>GPHFC</u> LG ³⁶⁵
Sri_CYP14	S ²⁴³ LDTT ²⁴⁷	E ²⁸² LLR ²⁸⁵	H ³⁴⁴ LALGF <u>GAHQ</u> CLG ³⁵⁶
Sri_CYP15	G ²⁷⁴ HITT ²⁷⁸	E ³¹² TLR ³¹⁵	Q ³⁷³ AAFGHG <u>IHYC</u> IG ³⁸⁵
Sri_CYP16	G ²⁴⁰ HITT ²⁴⁴	E ²⁷⁸ ILR ²⁸¹	H ³³⁹ LTFGHG <u>IHF</u> CFG ³⁵¹
Sri_CYP17	G ²⁵³ GETT ²⁵⁷	E ²⁹¹ LLR ²⁹⁴	H ³⁵⁴ LGFGHG <u>RHMC</u> LG ³⁶⁶
Sri_CYP18	G ²⁴⁵ QDTT ²⁴⁹	E ²⁸³ TMR ²⁸⁶	S ³⁴⁶ LGFGHG <u>PHRC</u> MG ³⁵⁸
Sri_CYP19	S ²³² LETT ²³⁶	E ²⁷⁰ LLR ²⁷³	H ³³¹ LGFGYG <u>AHNC</u> VG ³⁴³
Sri_CYP20	G ²⁵¹ HETT ²⁵⁵	E ²⁸⁹ LMR ²⁹²	H ³⁵⁰ VAFGFG <u>IHQ</u> CLG ³⁶²
Sri_CYP21	G ²⁸⁷ HETT ²⁹¹	E ³⁴³ ALR ³⁴⁶	Y ⁴¹⁵ KPFGT <u>GARAC</u> IG ⁴²⁷
Sri_CYP22	G ²⁷⁰ SETA ²⁷⁴	E ³²⁵ TMR ³²⁸	M ³⁹⁵ TPFGVG <u>NRKC</u> PS ⁴⁰⁷
Sri_CYP23	G ²⁵⁶ FETT ²⁶⁰	E ³⁰¹ LLR ³⁰⁴	H ³⁶³ LGYGHG <u>IHYC</u> LG ³⁷⁵
Sri_CYP24	G ²⁴⁸ DETG ²⁵²	E ²⁸⁶ ALR ²⁸⁹	H ³⁴⁷ FTFAYG <u>SHYC</u> LG ³⁵⁹

According to the widely accepted taxonomy, P450s within the same family share at least 40% amino acid identity and those belonging to the same subfamily share more than 55% identity. Based on these rules, the closest homologues of each cytochrome P450s were firstly identified, and the pairwise identity of each P450s to their homologues was then calculated. Therefore, the newly discovered P450s from *S.rishiriensis* could be defined into CYP families based on the sequence similarity. As is shown in Table 6.3, Sri_CYP07 is the only P450 that could not be grouped into any of the defined P450 subfamilies or families, which means Sri_CYP07 may belong to a new group within the P450 superfamily. The detailed and accurate name for each P450s from *S.rishiriensis* needs to be defined and further archived.

Table 6.3. Closest homologues to putative cytochrome P450s found in *S.rishiriensis*. The top two homologues with high pairwise identities were selected. The CYP families as well as species of these homologues P450s were also defined.

CYP ID ^a	Size ^b (aa)	Closest homologues found in database				Predicted CYP family
		Species	CYP family ^c	Coverage ^d (%)	Identity ^e (%)	
Sri_CYP01	413	<i>Streptomyces griseolus</i>	105B1	95	49.6	105
		<i>Streptomyces tubercidicus strain R-922</i>	105B2	99	49.5	
Sri_CYP02	496	<i>Streptomyces avermitilis</i>	157C2	100	72.2	157C
		<i>Streptomyces coelicolor</i>	157C1	100	68.5	
Sri_CYP03	405	<i>Streptomyces avermitilis</i>	107P2	100	86.7	107P
		<i>Streptomyces coelicolor</i>	107P1	98	85.2	
Sri_CYP04	469	<i>Streptomyces scabies</i>	182B1	99	82.7	182B
		<i>Streptomyces avermitilis</i>	182A1	96	49.2	
Sri_CYP05	410	<i>Streptomyces fradiae</i>	154B1	99	47.2	154
		<i>Streptomyces avermitilis</i>	154C2	99	46.0	
Sri_CYP06	420	<i>Streptomyces coelicolor</i>	156A1	95	40.1	156
		<i>Streptomyces scabies</i>	156D1	93	43.4	
Sri_CYP07	398	<i>Streptomyces lavendulae</i>	107N1	99	37.8	Unknown
		<i>Saccharopolyspora erythraea NRRL23338</i>	107E2	96	37.6	
Sri_CYP08	448	<i>Streptomyces coelicolor</i>	156B1	100	72.8	156B
		<i>Streptomyces scabies</i>	156B2	98	69.3	
Sri_CYP09	453	<i>Streptomyces avermitilis</i>	183A1	95	44.0	183
		<i>Mycobacterium marinum MM0281</i>	183B1	94	43.5	
Sri_CYP10	404	<i>Roseiflexus sp. RS-1</i>	107AZ1	95	42.2	107
		<i>Bradyrhizobium japonicum</i>	107AN1	98	40.6	
Sri_CYP11	452	<i>Streptomyces avermitilis</i>	183A1	92	47.6	183

CYP ID ^a	Size ^b (aa)	Closest homologues found in database				Predicted CYP family
		Species	CYP family ^c	Coverage ^d (%)	Identity ^e (%)	
		<i>Mycobacterium marinum</i> MM0281	183B1	94	46.5	
Sri_CYP12	394	<i>Streptomyces coelicolor</i>	107T1	95	59.3	107
		<i>Streptomyces venezuelae</i>	107L1	98	57.9	
Sri_CYP13	412	<i>Streptomyces scabies</i>	125A8	99	81.5	125A
		<i>Streptomyces avermitilis</i>	125A2	100	80.7	
Sri_CYP14	406	<i>Rhodococcus sp. RHA1</i> <i>Rha04313</i>	105Y1	99	43.7	105
		<i>Streptomyces</i> <i>tubercidicus strain I-1529</i>	105AA2	99	42.9	
Sri_CYP15	399	<i>Streptomyces virginiae</i>	113C1	97	48.6	113
		<i>Streptomyces sp. Tu6071</i>	113F1	86	45.2	
Sri_CYP16	402	<i>Streptomyces sp. Tu6071</i>	113F1	94	50.4	113
		<i>Saccharopolyspora</i> <i>erythraea NRRL23338</i>	113D1	91	49.3	
Sri_CYP17	422	<i>Streptomyces avermitilis</i>	107V1	96	70.7	107V
		<i>Actinomadura hibisca</i>	107M1	98	40.9	
Sri_CYP18	412	<i>Streptomyces avermitilis</i>	154B2	98	63.5	154B
		<i>Streptomyces fradiae</i>	154B1	99	61.4	
Sri_CYP19	395	<i>Thermobifida fusca</i>	217A1	98	40.1	217
		<i>Streptomyces scabies</i>	145C1	97	39.6	
Sri_CYP20	411	<i>Streptomyces avermitilis</i>	105D7	99	71.3	105D
		<i>Streptomyces coelicolor</i>	105D5	95	76.4	
Sri_CYP21	517	<i>Streptomyces coelicolor</i>	102B1	99	78.4	102B
		<i>Streptomyces avermitilis</i>	102B2	98	76.9	
Sri_CYP22	456	<i>Streptomyces avermitilis</i>	170A2	99	74.9	170A
		<i>Streptomyces coelicolor</i>	170A1	99	73.9	
Sri_CYP23	431	<i>Streptomyces scabies</i>	107U4	99	88.6	107U
		<i>Streptomyces coelicolor</i>	107U1	96	89.0	
Sri_CYP24	407	<i>Streptomyces spheroides</i>	163A1	100	88.7	163A
		<i>Streptomyces</i> <i>roseochromogenes</i>	163A2	99	95.1	

^a The name of all putative P450s in *S.rishiriensis*.

^b Amino acid number of putative P450s.

^c Closest homologues in Bacterial cytochrome P450 collection (archived by Dr David Nelson) and family classification searched in CYPED.

^d Coverage of query sequence generated by PBLAST comparison.

^e The pairwise identity for the aligned segment.

6.2.3 Multiple alignments and phylogenetic analysis

All 24 P450s from *S.rishiriensis* are aligned using ClustalW, and the paralogous relationship among the 24 P450s is generated with the neighbour-joining method by MEGA 7.0⁽¹⁷⁰⁾. In Figure 6.1, CYP09, CYP11, CYP21, and CYP22 are relatively distant from the main group of CYPs.

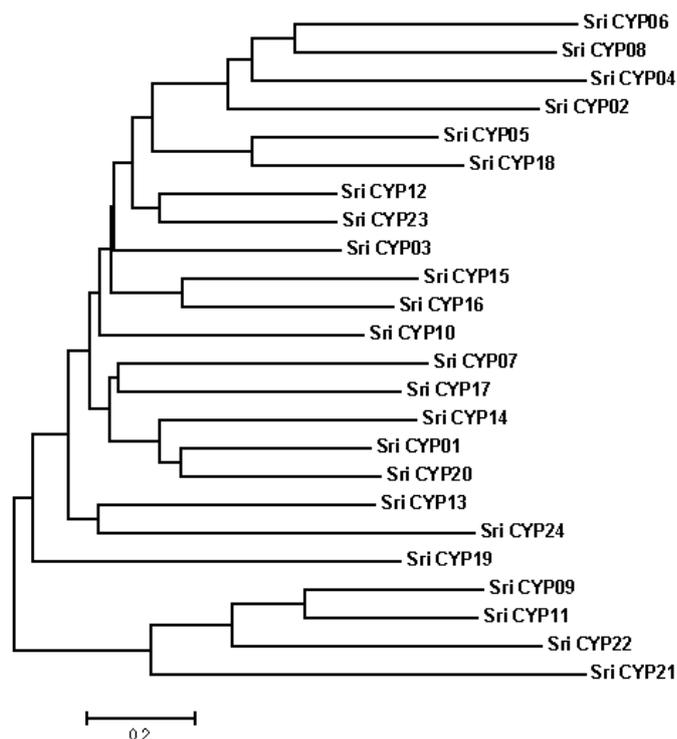


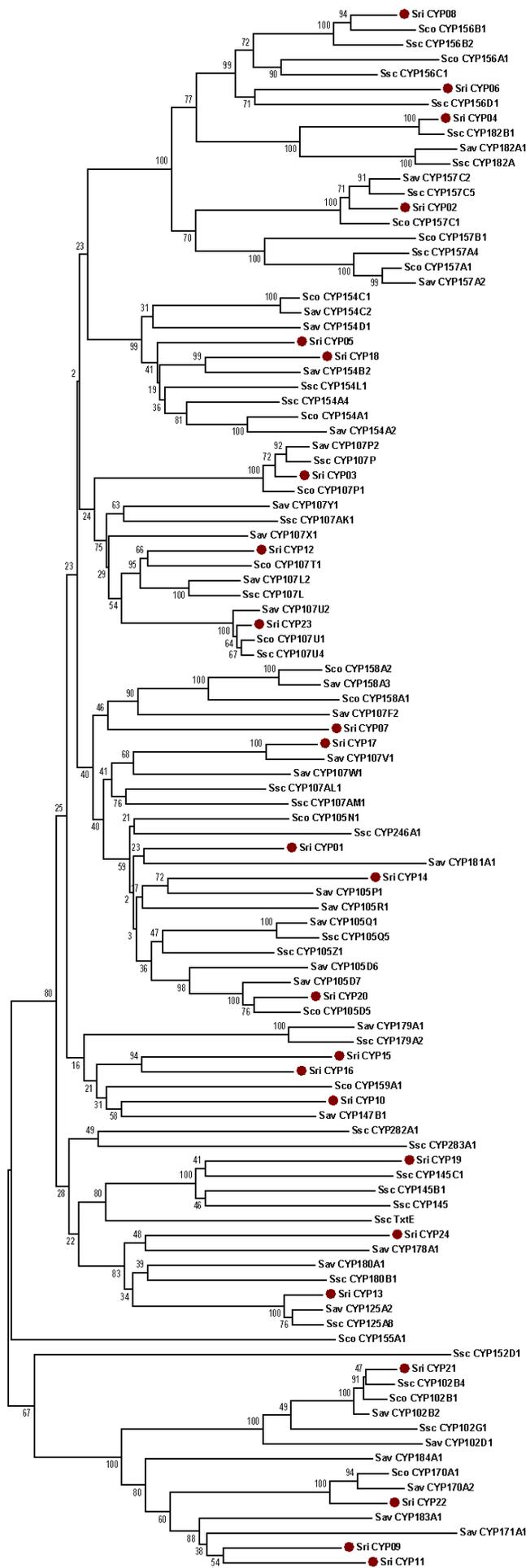
Figure 6.1. A paralogous tree of all 24 CYP sequences from *S.rishiriensis*.

The prefix Sri stands for the organism name *S.rishiriensis*. The horizontal branch lengths are proportional to the number of residue substitutions. The bar at the left corner indicates 0.2 amino acid substitutions per amino acid for the branch length.

According to Table 6.3, 11 CYPs are showing more than 70% identity to cytochrome P450s from other organisms. Among those other organisms, the most frequently appeared are *Streptomyce coelicolor*, *Streptomyces avermitilis*, and *Streptomyces scabies*. Therefore, the phylogenetic tree of combined P450s from *S.rishiriensis*, *S.coelicolor*, *S.avermitilis*, and *S.scabies* was generated in Figure 6.2, which might help to illustrate the evolutionary relationships on CYPs from *S.rishiriensis* to other *Streptomyces*.

As is shown in Figure 6.2, CYPs belonging to subfamilies such as 157C, 107P, 102B, and 107U are present in all four *Streptomyces*. CYPs from subfamilies like 105D and 170A are only present in *S.rishiriensis*, *S.coelicolor*, and *S.avermitilis*. Similarly, the CYPs from subfamily 125A only exist in *S.rishiriensis*, *S.avermitilis*, and *S.scabies*, while CYPs of subfamily 156B are only shown in *S.rishiriensis*, *S.coelicolor*, and *S.scabies*.

Apart from these generally conserved CYPs among all four *Streptomyces*, there are several CYPs such as Sri_CYP07, Sri_CYP15, Sri_CYP16, and Sri_CYP24 that are from CYP families only present in *S.rishiriensis*. Especially for Sri_CYP07, it is not evolutionarily grouped to any CYPs from the other three *Streptomyces*, and it does not have a homologue over 40% identity from any identified CYPs. It strongly indicates that Sri_CYP07 may belong to a new and undefined family of cytochrome P450.



0.2

Figure 6.2. Neighbor-joining tree of CYPs from 4 Streptomyces. The organism names are shortened as prefixes such as Sri (*S.rishiriensis*), Sco (*S.coelicolor*), Sav (*S.avermitilis*), and Ssc (*S.scabies*). All CYPs from *S.rishiriensis* were also marked with the red dot symbol to indicate where they were located on the tree. The horizontal branch lengths are proportional to the number of residue substitutions. The bar at the left corner indicates 0.2 amino acid substitutions per amino acid for the branch length.

6.2.4 CYPs clustered with ferredoxin or ferredoxin reductases

Sri_CYP20 is the only P450 clustered with ferredoxin (Sri_Fd06 in Table 6.4) in *S.rishiriensis*, and there is no ferredoxin reductase clustered to any P450s found in *S.rishiriensis*. Sri_CYP20 belongs to the P450 subfamily 105D and shares high homology to CYP105D5 from *S.coelicolor*, CYP105D1 from *S.griseus* and CYP105D4 from *S.lividan*. The three P450 homologues from 105D subfamily all happen to be clustered with a ferredoxin within their genomes. It has been suggested that the clustered ferredoxin or ferredoxin reductase could enhance the expression and activity of cytochrome P450s⁽⁴⁹⁾. The limited ferredoxins and ferredoxin reductases listed in Table 6.4 are presumably supporting the activity of all 24 P450s existed in *S.rishiriensis*.

Table 6.4. Putative ferredoxins and ferredoxin reductases discovered from the *S.rishiriensis* genome. Seven putative ferredoxins and five ferredoxin reductases were also discovered. The protein sequences of ferredoxin and ferredoxin reductases were sent through BLASTP and Pfam for domain definition. The protein length and molecular weight were calculated by the ExPASy ProtParam tool.

ID	Protein length (aa)	Molecular weight (kDa)	Contig N.O. and gene position
Putative Ferredoxins			
Sri_Fd01	101	10.66	Contig 6, 111470-111775
Sri_Fd02	107	11.86	Contig 11, 44478-44789
Sri_Fd03	61	6.53	Contig 35, 117763-117948
Sri_Fd04	90	9.39	Contig79, 31962-32234
Sri_Fd05	69	7.13	Contig 88, 9617-9408
Sri_Fd06	66	6.61	Contig 127, 30471-30271
Sri_Fd07	99	10.89	Contig 177, 11907-11608
Putative Ferredoxin reductases			

ID	Protein length (aa)	Molecular weight (kDa)	Contig N.O. and gene position
Sri_FdR01	415	43.84	Contig 18, 311154-309907
Sri_FdR02	447	46.60	Contig 38, 83217-84560
Sri_FdR03	465	48.84	Contig 92, 10998-9601
Sri_FdR04	421	44.96	Contig 169, 227-1492
Sri_FdR05	453	49.45	Contig 330, 50-1408

6.2.5 CYPs involved in the production of secondary metabolites

As is summarized in Table C1, it is very common to find various P450s are involved in the production of secondary metabolites in *Streptomyces*. For example, in *Streptomyces venezuelae*, the widely studied PikC cytochrome P450 (CYP107L1) is involved in the production of picromycin, which accepts 12-14 membered ring macrolide substrates⁽¹⁷¹⁾. In *S.rishiriensis*, one cytochrome P450, named as Sri_CYP12, shares 57.9% of identity to PikC cytochrome P450, which is a suitable candidate for further investigation on hydroxylation potentials on macrolide substrates.

Based on the anti-SMASH analysis, there is a cluster found in *S.rishiriensis* having a 100% genetic match to a known cyclooctatin biosynthetic cluster. The cluster includes core genes such as Sri_CYP11, Sri_CYP09, terpene cyclase, and type I GGDP synthase (geranylgeranyl diphosphate synthase), which is located on genomic contig 76. What is more interesting is that the cluster is closely surrounded by the other two P450s (Sri_CYP12 and Sri_CYP10) upstream and downstream respectively. It is suspected that CYP12 and CYP10 may be involved in further hydroxylation of cyclooctatin. A tetR transcriptional regulatory gene, that contains a TTA codon, is also adjacently located to the cluster. There is limited information on whether the regulatory gene is involved with the transcription of the cluster, but the regulatory gene is considered to be within the putative cluster according to anti-SMASH analysis.

In *Streptomyces spheroides* and *Streptomyces roseochromogenes*, CYP163A1 and CYP163A2 are involved in the cluster of producing Novobiocin and Clorobiocin respectively. In the meantime in *S.rishiriensis*, Sri_CYP24 shares 88.7% and 95.1% of identity to CYP163A1 and CYP163A2 respectively, and the

anti-SMASH analysis revealed that Sri_CYP24 was within the coumermycin A1 biosynthetic gene cluster. Based on a previous study on CYP163A1 and 163A2, Sri_CYP24 is proposed to conduct β -hydroxylation of tyrosine, which means introducing oxygen at β -position carbon of tyrosine⁽¹⁶⁸⁾. Further investigation on Sri_CYP24 could be carried to determine the function and efficiency of hydroxylating tyrosine as well as other potential amino acids at the β position.

6.2.6 Predictions of functions of some CYPs in *S.rishiriensis*

Among all 24 CYP in *S.rishiriensis*, a few P450s are being found to have a relatively high identity (over 80%) to those identified CYPs in the other *Streptomyces*, which includes Sri_CYP03, Sri_CYP13, Sri_CYP23 and Sri_CYP24. Because of this high similarity between these Sri_CYPs and their homologues, these newly discovered Sri_CYPs may have similar substrate spectrum as their homologues.

For Sri_CYP24, it has been discovered with the biosynthetic cluster of coumermycin A1⁽¹⁶⁸⁾. Along the synthesis pathway, it was proposed that Sri_CYP24 naturally accepts tyrosine and conducts the β -hydroxylation. Therefore, tyrosine would be the first substrate to test against when it comes to the functional analysis of Sri_CYP24. It is also reasonable to consider other amino acids like histidine and tryptophan that could be accepted by the homologues of Sri_CYP24.

For Sri_CYP23, the closest homologue is CYP107U1 from *S.coelicolor*, which shares 89% identity. CYP107U1 was found to be involved in sporulation and antibiotic production in *S.coelicolor*, and one of the natural substrates was discovered to be glycocholic acid⁽¹⁷²⁾. Therefore, CYP23 may hold similar substrates preference as well as any potential roles in the growth and development of *S.rishiriensis*.

For Sri_CYP03 and Sri_CYP13, their high identity homologues are found to be CYP107P2 (86.7%) and CYP125A2 (80.7%) respectively. Both CYP107P2 and CYP125A2 were discovered from *S.avermitilis* and had been screened regio-specific hydroxylation of flavonoids⁽¹⁷³⁾. CYP107P2 was found to have activity on

0.1 mM apigenin (4',5,7-trihydroxyflavone), which leads to the production of 3',4',5,7-tetrahydroxyflavon. CYP125A2 was tested to be active on 0.1 mM chrysin (5,7-dihydroxyflavone), which produced B-ring hydroxylated 5,7-trihydroxyflavone. Therefore, the predicted substrates for Sri_CYP03 and Sri_CYP13 are apigenin and chrysin.

6.3 Cloning and expression of *sri_cyp03*, *sri_cyp13* and *sri_cyp24*

After the complete identification and bioinformatic analysis on all CYPs as well as CYP related proteins within the *S.rishiriensis* genome, the basic understanding of several P450s has been acquired. To further conduct the functional analysis of these P450s, cloning and expression were carried out to obtain purified cytochrome P450s. Due to the length of the project, only three CYPs, Sri_CYP03, Sri_CYP13, and Sri_CYP24 were selected for study in this chapter.

6.3.1 Cloning and plasmid construction

As is discussed in Chapter 4, there are two types of constructs being applied in this project for Class I cytochrome P450 expression, which includes the single *cyp* gene expression and the three-gene *cyp-fd-fdr* operon. For both constructs, the cloning site for *cyp* gene is the same, which allows the gene fragments with standardized *NdeI* and *EcoRI* sites to be inserted and assembled. Therefore, three pairs of primers for *sri_cyp03*, *sri_cyp13*, and *sri_cyp24* were designed (Table 6.6) and used for genomic PCR (Figure 6.3) (PCR preparation details described in section 2.1.6.1). All gene fragments were firstly assembled onto the TOPO cloning vector for preservation and sequencing, before being assembled onto *cyp* construct or *cyp-fd-fdr* operon for expression.

Table 6.5. Primers designed for PCR of *sri_cyp03*, *sri_cyp13*, and *sri_cyp24* from *S.rishiriensis*. The Phusion PCR master mix was used, and the calculated annealing temperatures were strictly applied.

CYP	Primers	Sequence 5'-3' ^a	T _m (°C) ^b
Sri_CYP03	Forward	GCGCC <u>CATATGGC</u> CAGCAGTCGACGACC	70
	Reverse	GCGCGAATTCCTATTAGTGGTGGTGGTGGTGGTG GAGGGTCACGCTCAGTCCCTC	
Sri_CYP13	Forward	GCGCCATATGCCCTGTCCAGCGCTT	68
	Reverse	GCGCGAATTCCTATTAGTGGTGGTGGTGGTGGTG CCCCGCGTCGACCC	
Sri_CYP24	Forward	GCGCCATATGAGCACCCGTCCCATG	67
	Reverse	GCGCGAATTCCTATTAGTGGTGGTGGTGGTGGTG GGCCGTGATCTTGATGGG	

^a The underlined regions indicate the restriction sites of *Nde*I and *Eco*RI, and the bold regions indicate the ribosome binding sites. The gene sequence for 6His-tag was also designed in both reverse primers.

^b The calculated annealing temperature for each primer generated by the NEB T_m calculator.

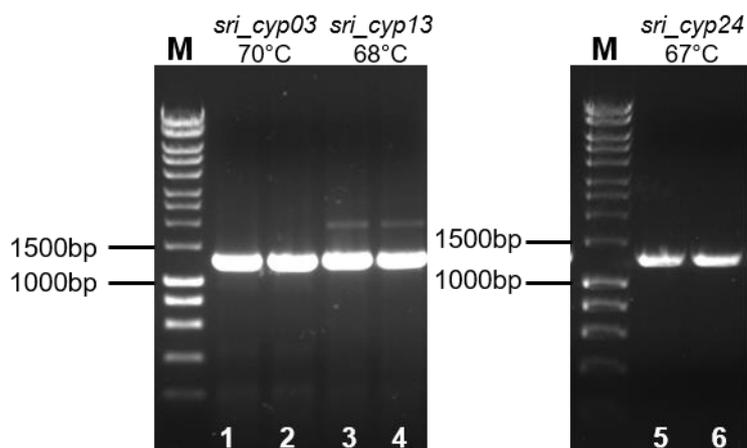


Figure 6.3. Amplification of *sri_cyp03*, *sri_cyp13* and *sri_cyp24* from *S.rishiriensis*. Genomic PCR was conducted using primers described in Table 6.6. Lane M stands for the 1kb HyperMarker purchased from Bioline. Lane 1 and 2 were the PCR products for the *sri_cyp03* gene. Lane 3 and 4 were the PCR products for the *sri_cyp13* gene. Lane 5 and 6 were the PCR products for the *sri_cyp24* gene. All bands were located between 1 kb and 1.5 kb range, which matched to the expected PCR product length of 1.2 kb approximately.

6.3.2 Expression of three *S.rishiriensis* CYPs in *E.coli*

6.3.2.1 Sri_CYP03 and Sri_CYP13

As is discussed in chapter 4, the expression of several cytochrome P450s from *Streptomyces* could be enhanced when constructing the *cyp*, *ferredoxin*, and

ferredoxin reductase within the same expression operon. Moreover, the *cyp-fd2-fdr* operon was tested to have better performance than another three-gene operon like *cyp-fd1-fdr* or *cyp-pdx-pdr*, so *cyp-fd2-fdr* is always the first operon tested when expressing any Class I cytochrome P450 in this project. Therefore, the expression of Sri_CYP03 and Sri_CYP13 started from the construction of a sole *cyp* expression and a *cyp-fd2-fdr* operon.

The three-gene operon, *sri_03-fd2-fdr*, and *sri_13-fd2-fdr*, were firstly used for expression under general conditions that being described in section 2.1.7. The concentration of IPTG was set to be either 0.5 mM or 0.1 mM, and the concentration of P450 was measured through carbon monoxide assay (Figure 6.4). As is shown in Figure 6.4, there were no detectable peaks at 450 nm in the difference spectrum of either Sri_CYP03 or Sri_CYP13. It seemed the expression of incorrectly folded Sri_CYP03 (peak at 420 nm) was slightly increased when reducing the use of IPTG from 0.5 mM to 0.1 mM (Figure 6.4 (A) and (B)). Similar changes were also observed for Sri_CYP13, showing in Figure 6.4 (C) and (D). Therefore, the decrease of IPTG could contribute to the heterologous production of inactive cytochrome P450s, but no detectable enhancement on the production of active cytochrome P450s.

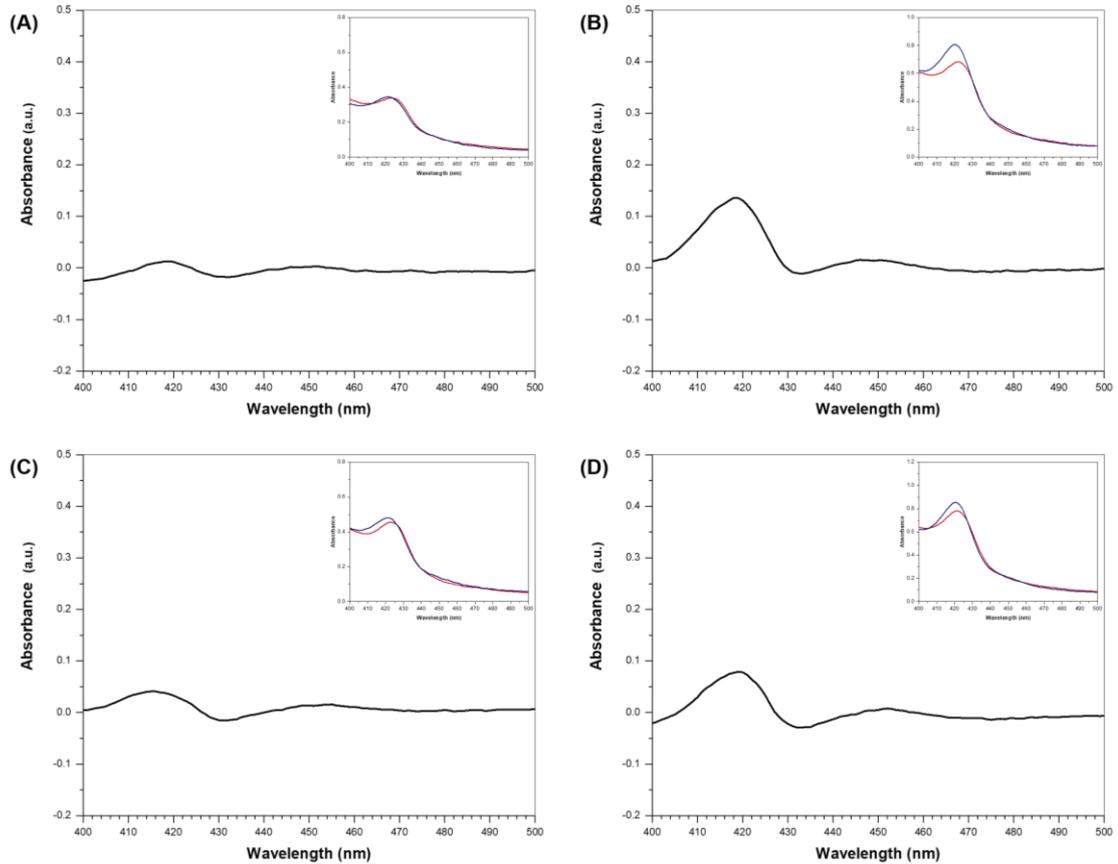


Figure 6.4. Expression of Sri_CYP03 and Sri_CYP13 using a *cyp-fd2-fdr* construct with different IPTG concentrations. All expression was conducted for 24 hours in *E.coli* BL21(DE3). The main curve in each figure was the difference spectra of carbon monoxide assay, and the inserted figure contains a red curve representing the spectrum of reduced cytochrome P450 and a blue curve representing the spectrum of CO-cytochrome P450 binding mixture. The different IPTG concentrations were indicated in (A) Sri_CYP03, 0.5 mM IPTG; (B) Sri_CYP03, 0.1 mM IPTG; (C) Sri_CYP13, 0.5 mM IPTG; (D) Sri_CYP13, 0.1 mM IPTG.

As is discussed in chapter 4, the *single-cyp* construct is always the option when the three-gene construct does not work. Therefore, the conventional *single-cyp* construct was tested for expression of Sri_CYP03 and Sri_CYP13, and the induction concentration of IPTG was set to be 0.1 mM. As is shown in Figure 6.5, there are no detectable peaks at 450 nm for either Sri_CYP03 or Sri_CYP13, but there are increased peak height at 420 nm. What it means is that the production of cytochrome P450s is boosted again, but none of these proteins are presented in the active form. Due to the simpler construct of *single-cyp* construct, the host

cells seem to have less burden on performing the heterologous expression. However, the use of either three-gene operon or single-gene construct could not influence or lead to the production of structurally active CYPs.

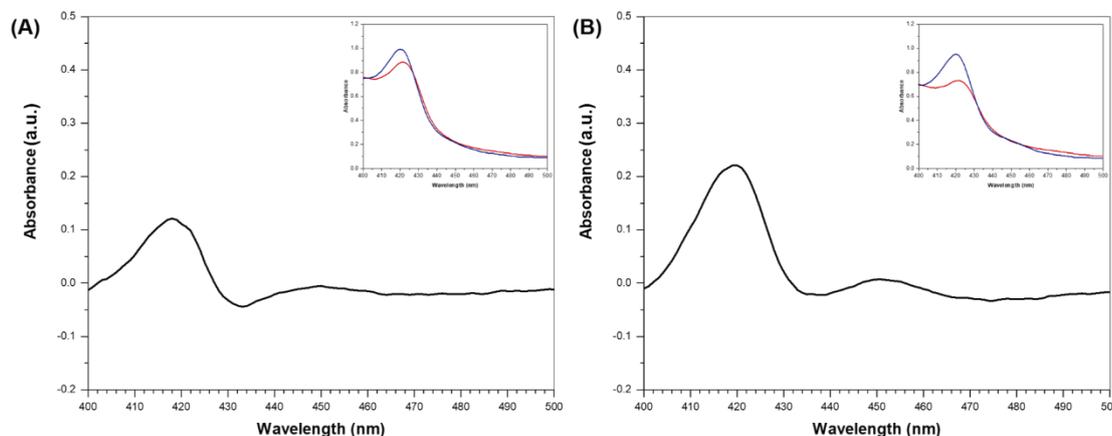


Figure 6.5. Expression of Sri_CYP03 and Sri_CP13 using a *single-cyp* construct. All expression was conducted for 24 hours by *E.coli* BL21(DE3). The main curve in each figure was the difference spectra of carbon monoxide assay, and the inserted figure contains a red curve representing the spectrum of reduced cytochrome P450 and a blue curve representing the spectrum of CO-cytochrome P450 binding mixture. (A) Sri_CYP03, 0.1 mM IPTG; (B) Sri_CYP13, 0.1 mM IPTG.

Based on the expression tests conducted above by using either *cyp-fd2-fdr* operon or *single-cyp* constructs, it is clear that the *E.coli* host is suitable to produce soluble cytochrome P450, and lower IPTG concentration could lead to increased production of heterologous protein. However, none of the produced proteins were presenting the structurally active conformation between thiolate anion of cysteine to the heme iron. It was unknown at that stage why these two *S.rishiriensis* CYPs could be expressed but not correctly assembled with heme ligands in *E.coli* host. It was suspected that the protein folding mechanism naturally existed in *E.coli* is not suitable for these two CYPs from *S.rishiriensis*.

Since the object in this chapter is still using *E.coli* as hosts to produce microbial CYPs, another approach involving the use of molecular chaperones in co-expression with the target proteins. Five different chaperone systems, that were developed by Takara Bio Inc, were tested to be co-expressed with either Sri_CYP03 or Sri_CYP13 (Table 6.7). There are previous studies on chaperone

teams including DnaK-DnaJ-GrpE, GroES-GroEL, and trigger factor (tig) within *E.coli*, which reveals their functions on facilitating the folding of recombinant proteins and preventing the aggregation of recombinant proteins^(174, 175). These chaperone teams could be used separately or together for the expression of CYP, which leads to the five combinations of chaperone systems listed in Table 6.7.

Table 6.6. Five different chaperone systems used for protein expression. All chaperone plasmids are low copy number plasmids with pACYC origin, which are compatible with Ampicillin resistant pUC origin plasmid for coexpression. Two different promoters are being situated at the upstream of the chaperone genes on the plasmids. The applicable strains for these chaperone plasmids include *E.coli* BL21 (DE3) and *E.coli* JM109.

NO.	Plasmid ID	Chaperone teams	Promoter	Inducer	Resistant marker
1	pQR2307	<i>dnaK-dnaJ-grpE</i>	<i>araB</i>	L-Arabinose	Chloramphenicol
		<i>groES-groEL</i>	<i>Pzt1</i>	Tetracyclin	
2	pQR2308	<i>groES-groEL</i>	<i>araB</i>	L-Arabinose	Chloramphenicol
3	pQR2309	<i>dnaK-dnaJ-grpE</i>	<i>araB</i>	L-Arabinose	Chloramphenicol
4	pQR2310	<i>groES-groEL-tig</i>	<i>Pzt1</i>	Tetracyclin	Chloramphenicol
5	pQR2311	<i>tig</i>	<i>araB</i>	L-Arabinose	Chloramphenicol

There are two targeted CYPs and five different chaperone systems, therefore, the co-expression started from constructing ten different strains, each of which contains one target CYP plasmid and one chaperone system at the same time. The CYP plasmids containing the *single-cyp* construct (pQR2303 and pQR2304) were firstly selected to be tested in the co-expression experiments. The co-expression method was described in section 2.3.2, which was optimised from the normal expression described in section 2.1.7.1. The expression was set to last for 24 hours after IPTG induction at 25°C. The amount of soluble P450s was measured by carbon monoxide assay, and all difference spectrum was collected in Figure C2 in Appendix C.

For Sri_CYP03, only the co-expression with groES-groEL chaperone has a detectable peak at 450 nm on the difference spectrum (Figure 6.6). At the same time, there is also a peak at 420 nm, which indicates the existence of both structurally active and inactive P450s in the cell lysate. By comparing to the spectrum displayed by expression of CYP without molecular chaperones, it is positive progress on delivering active P450 in the same *E.coli* hosts.

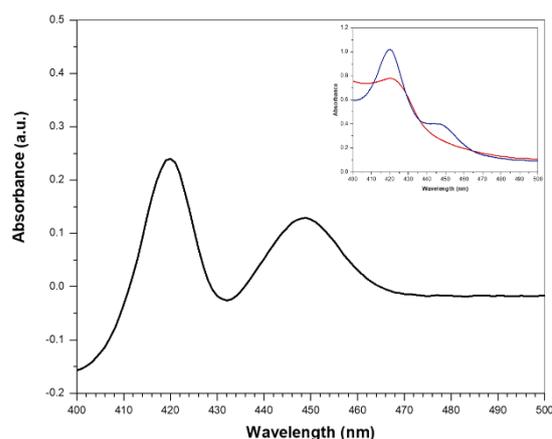


Figure 6.6. The successful co-expression of active Sri_CYP03 and molecular chaperone. The plasmid pQR2303 contains his-tagged *sri_cyp03* gene, and the plasmid pQR2308 contains the genes of molecular chaperone groES and groEL. Both plasmids were transformed into *E.coli* BL21(DE3). The expression was conducted for 24 hours with 0.1 mM IPTG and 0.2 mg/mL L-arabinose in *E.coli* BL21(DE3). The CO assay was conducted to measure the expression level of Sri_CYP03. The main curve in the figure was the difference spectrum of carbon monoxide assay, and the inserted figure contains a red curve representing the spectrum of reduced cytochrome P450 and a blue curve representing the spectrum of CO-cytochrome P450 binding mixture.

For Sri_CYP13, the co-expression with groES-groEL chaperone has a detectable peak at 450 nm on the difference spectrum (Figure 6.7, A). There are no detectable peaks around 420 nm, which indicates all expressed Sri_CYP13 are structurally active. Other than the groES-groEL team, the groES-groEL team, and trigger factor team together could also facilitate the expression of actively formed Sri_CYP13 (Figure 6.7, B). There is a detectable peak at 450 nm on the difference spectrum, which is not as high as that showing in Figure 6.7 (A). It seems that just using GroES and GroEL as chaperone proteins is more effective in enhancing the production of CYP13 with active structure, which means the trigger factor proteins do not positively influence the folding of Sri_CYP13. Therefore, it is proven that using molecular chaperones for expression of Sri_CYP13 is also a better approach, comparing to the expression of CYP13 alone in *E.coli* hosts.

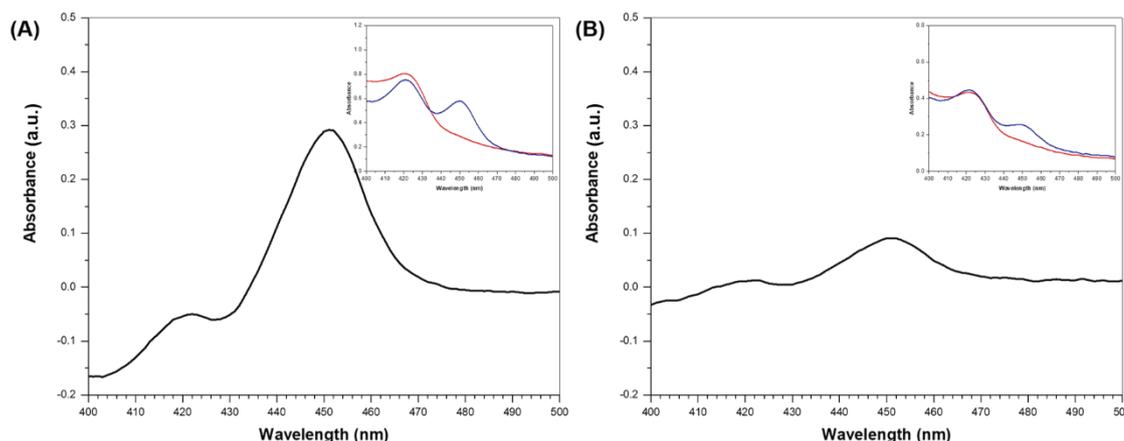


Figure 6.7. The successful co-expression of active Sri_CYP13 and molecular chaperone. All expression was conducted for 24 hours with 0.1 mM IPTG in *E.coli* BL21(DE3). The main curve in each figure was the difference spectra of carbon monoxide assay, and the inserted figure contains a red curve representing the spectrum of reduced cytochrome P450 and a blue curve representing the spectrum of CO-cytochrome P450 binding mixture. (A) Co-expression of Sri_CYP13 and groES-groEL chaperone system. (B) Co-expression of Sri_CYP13 and groES-groEL-tig chaperone system.

The expression of actively formed Sri_CYP03 and Sri_CYP13 was achieved by coupling with chaperone proteins GroES and GroEL, and all produced CYPs are tagged with 6-histidine for nickel column purification. As is shown in Figure 6.8, the purified Sri_CYP03 and Sri_CYP13 were characterised through SDS-PAGE. The targeted CYP bands were located between 40 kDa and 50 kDa, which matches the calculated molecular weights.

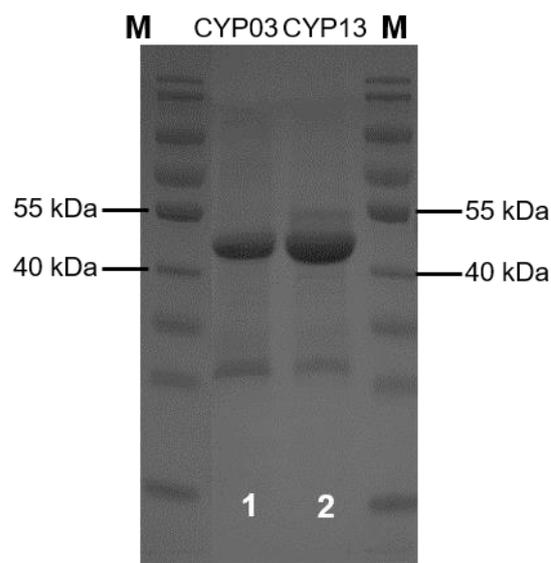


Figure 6.8. SDS-PAGE analysis of purified Sri_CYP03 and Sri_CYP13. The molecular weight of Sri_CYP03 and Sri_CYP13 are 45.1 kDa and 45.9 kDa respectively. All four lanes shown above were from the same gel. The original gel image was attached in section 11.7 (Appendix C). Lane 1, lane 2, and the right marker lane were adjacently located on the gel. The gel image was taken by AlphaView software, and it was manually annotated using PowerPoint. The marker used in both marker Lanes was the PageRuler pre-stained protein ladder (10-180kDa). Lane 1 is the purified Sri_CYP03, and Lane 2 is the purified Sri_CYP3. There were also some other impurity bands in the elution of CYP03 or CYP13 samples, but the majority of proteins are still the targeted CYPs.

After conducting all these expression trials described above, the key elements of producing Sri_CYP13 and Sri_CYP03 include coexpression with chaperone teams GroES-GroEL as well as a lower concentration of IPTG for induction. These successful expressions of Sri_CYP03 and Sri_CYP13 involves using the *single-cyp* construct to be coexpressed with the GroES-GroEL chaperone system. Other than the *single-cyp* constructs, *cyp-fd2-fdr* operons (pQR2300 and pQR2301) were also tested for the coexpression with GroES and GroEL system. Unfortunately, there was no expression of actively formed Sri_CYP03 and Sri_CYP13 when using *cyp-fd2-fdr* operons (CO spectrum shown in Figure C3). Therefore, the use of a *single-cyp* construct is also vital in the case of producing structurally active Sri_CYP03 and Sri_CYP13.

6.3.2.2 Sri_CYP24

Sri_CYP24 has been proposed in previous research to play a role in the biosynthetic cluster of coumermycin A1 that being naturally produced by *Streptomyces rishiriensis*⁽¹⁶⁸⁾. The proposed reaction conducted by Sri_CYP24 was β -hydroxylation of tyrosine. However, there have no reports on the study of Sri_CYP24 or heterologous expression of this P450 alone in *E.coli* hosts. Because there are various strategies on producing actinomycete P450s being described or developed in this project, such as using different operons for expression, change of IPTG concentrations for induction, and coexpression of chaperone proteins, so these attempts were also conducted on expressing Sri_CYP24.

Just like any CYPs being studied in this project, the P450 gene (*sri_cyp24*) was first assembled onto designated *single-cyp* construct or *cyp-fd2-fdr* operon first (Figure 6.9 E) to prepare the expression plasmids (pQR2302 and pQR2305). Both plasmids were then separately transformed into BL21 (DE3) or BL21 star (DE3) pLysS for expression. *E.coli* BL21 (DE3) is a common strain for expression of recombinant proteins, and *E.coli* BL21 star (DE3) pLysS strain is designed to have a reduced level of T7 RNA polymerase, which reduces the basal expression of recombinant genes. Because of previous experience of producing P450s from *S.rishiriensis*, it is suspected that Sri_CYP24 may also have problems with assembling the heme group into the P450. Therefore, other than the common strains, *E.coli* BL21 star (DE3) pLysS was also used for low-level expression of recombinant proteins, which may facilitate the folding process of structurally active P450s.

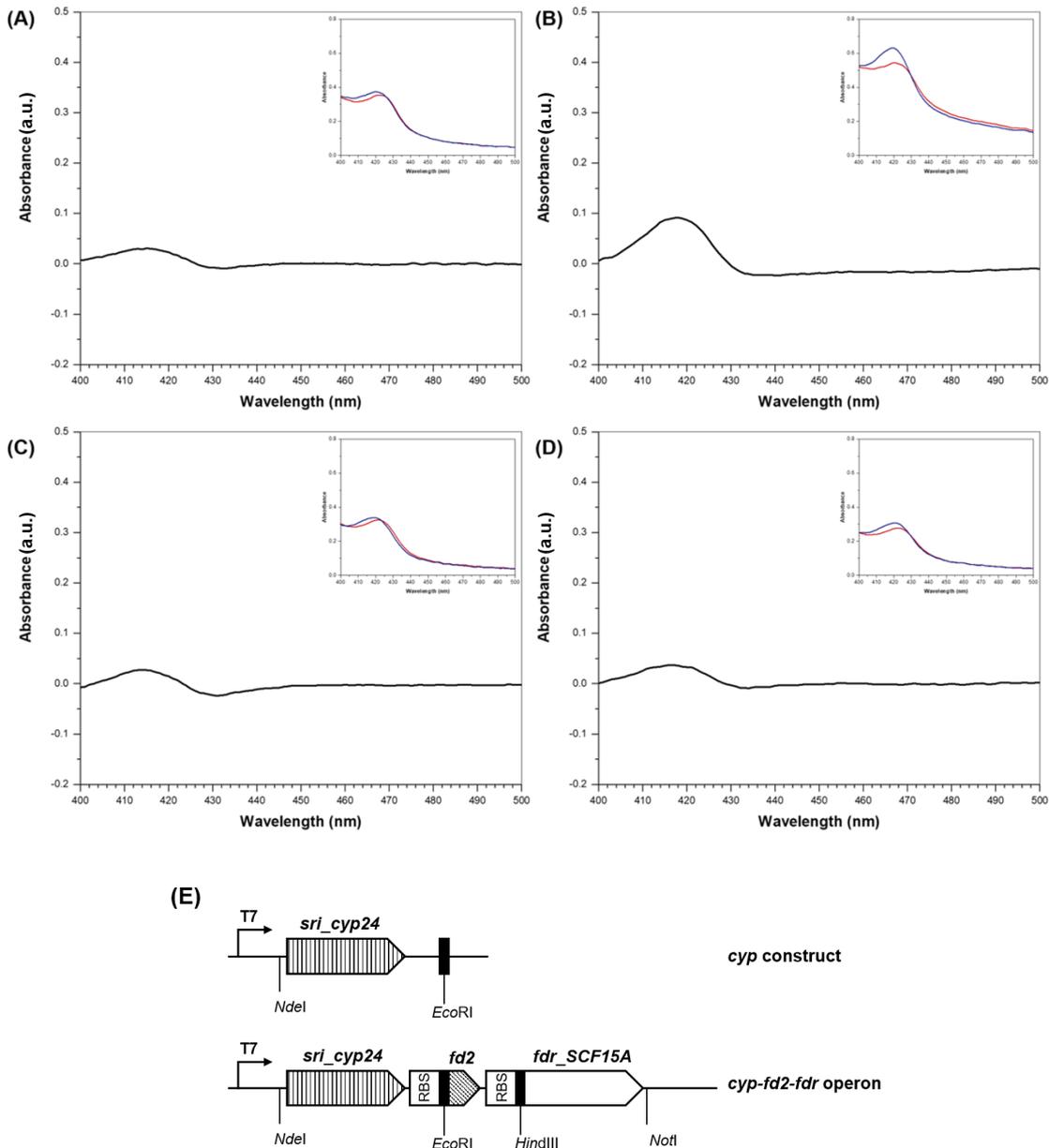


Figure 6.9. Expression of Sri_CYP24 using *cyp-fd2-fdr* operon or *single-cyp* construct. All expression was conducted for 24 hours. The main curve in each figure was the difference spectra of carbon monoxide assay, and the inserted figure contains a red curve representing the spectrum of reduced cytochrome P450 and a blue curve representing the spectrum of CO-cytochrome P450 binding mixture. The different expression strains and concentrations of IPTG were shown in (A) *sri_cyp24-fd2-fdr* operon, BL21(DE3), 0.5 mM IPTG; (B) *sri_cyp24-fd2-fdr* operon, BL21 Star (DE3) pLysS, 0.5 mM IPTG; (C) *sri_cyp24* construct, BL21(DE3), 0.1 mM IPTG; (D) *sri_cyp24* construct, BL21 Star (DE3) pLysS, 0.1 mM IPTG. (E) The difference between the *cyp* construct and *cyp-fd2-fdr* operon mentioned above.

As is shown in Figure 6.9 A to D, there are no detectable peaks at 450 nm for all four spectra, which indicates the expression of structurally active P450s was not successful for either *sri_cyp24* construct or *sri_cyp-fd2-fdr* operon. Moreover, the change of host strains from BL21 (DE3) to BL21 star (DE3) pLysS did not make a difference in the expression of Sri_CYP24 either. It seems the use of single-gene construct or three-gene operon does not have positive effects on producing Sri_CYP24, which is a similar scenario when expressing Sri_CYP03 and Sri_CYP13. Therefore, the five different chaperone systems were again co-transformed with CYP plasmids containing the *sri_cyp24* gene, which may facilitate the production of structurally active Sri_CYP24. However, the coexpression of Sri_CYP24 with any of the five chaperone systems did not lead to the production of any detectable active Sri_CYP24 (Figure 6.10 A to E).

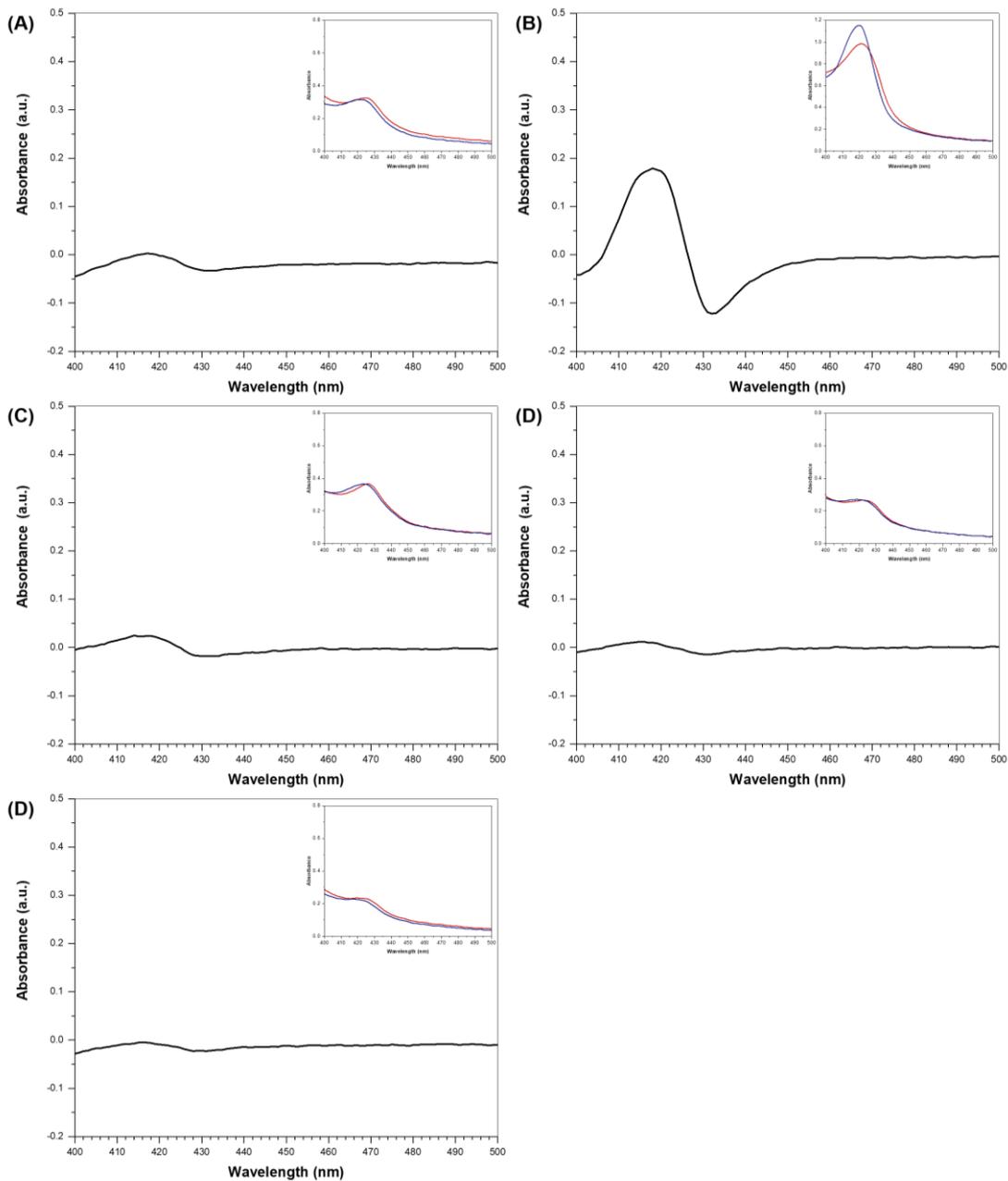


Figure 6.10. The unsuccessful co-expression of Sri_CYP24 and molecular chaperones. All expression was conducted in 0.1 mM IPTG for 24 hours at 25°C. The main curve in each figure was the difference spectra of carbon monoxide assay, and the inserted figure contains a red curve representing the spectrum of reduced cytochrome P450 and a blue curve representing the spectrum of CO-cytochrome P450 binding mixture. The CYP plasmid used for coexpression was pQR2305, containing the *single-cyp* construct. (A) Coexpression of Sri_24 and two chaperone systems DnaK-DnaJ-GrpE and GroES-GroEL. (B) Coexpression of Sri_CYP24 and GroES-GroEL system. (C) Coexpression of Sri_CYP24 and DnaK-DnaJ-GrpE system. (D) Coexpression of

Sri_CYP24 and GroES-GroEL-Trigger factor system. (E) Coexpression of Sri_CYP24 and Trigger factor.

It is unknown why the coexpression of chaperone could lead to the successful production of Sri_CYP03 and Sri_CYP13, but not contribute to the production of Sri_CYP24. Comparing to the other two P450 genes *sri_cyp03* and *sri_cyp13*, a notable difference is that *sri_cyp24* is not an independent gene within the genome but exists in the natural gene cluster of coumermycin A1 in *S.rishiriensis*⁽¹⁶⁸⁾. In previous studies, instead of a single gene from the cluster, all genes within the coumermycin A1 cluster was cloned together and expressed. It was suspected that the activation or production of Sri_CYP24 may require the presence of other proteins within the genetic cluster. Based on the expression trials in *E.coli* described above, it is also suspected that the protein production factory in *E.coli* may naturally not support the producing or folding of Sri_CYP24. The expression of Sri_CYP24 still needs further investigation, which may include the use of *Streptomyces* strains as the host or the attempts of cloning the entire cluster in *E.coli* for production.

6.4 Preliminary substrate screening for Sri_CYP03 and Sri_CYP13

6.4.1 Preliminary substrate pool

As is described in section 6.2.6, the bioinformatic analysis revealed that Sri_CYP03 shares an 86.7% sequential similarity with CYP107P2 from *S.avermitilis*, and Sri_CYP13 shares an 80.7% identity with CYP125A2 from *S.avermitilis*. The relatively high sequence identity was confirmed in the alignment shown in Figure 6.11 and Figure 6.12. As is shown, most of the aligned residues were either identical (black shaded) or strongly similar (grey shaded), which indicates they may share some common features at the structural level. In previous studies on the structural feature of cytochrome P450s, there were several substrate recognition sites (SRS) being identified and summarized in the P450s family⁽¹⁷⁶⁾. For example, the SRS1 is located on the B'-C loop at the substrate pocket and the SRS2 is located on the F-G loop being a part of the entrance to the substrate-binding pocket, as well as the SRS5 refers to the EXXR motif and five positions after. Three SRSs, including SRS1, SRS2, and SRS5,

are all involved with the substrate selectivity, which is also labeled on the alignments in Figure 6.11 and Figure 6.12. As is shown in both figures, the CYP107P2 and Sri_CYP03 are sharing highly similar SRS1 and SRS5, while CYP125A2 and Sri_CYP13 have nearly identical sequences of SRS2 and SRS5. All these comparisons on the sequential level led to the conclusion that Sri_CYP03 and Sri_CYP13 may share a similar substrate scope with their highly identical homologous CYP107P2 and CYP125A2 respectively. As was identified in previous research, the CYP107P2 has shown activities on apigenin, and CYP125A2 could react with chrysin⁽¹⁷³⁾. Therefore, the two flavonoids, apigenin (APG) and chrysin (CHY) were selected for the preliminary screening.

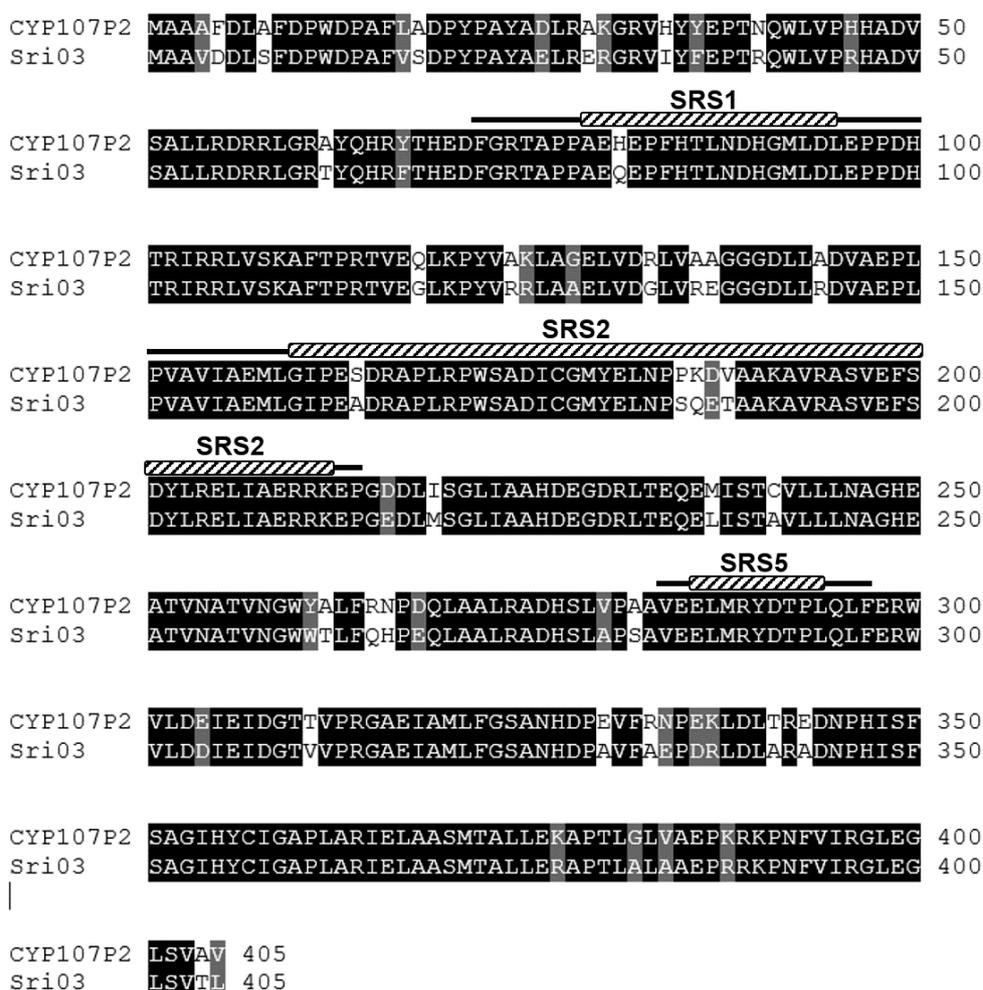


Figure 6. 11. The alignment between CYP107P2 and Sri_CYP03. The identical and strongly similar residues are shaded in black and grey respectively. The selectivity related substrate recognition sites (SRS) are labeled as above.

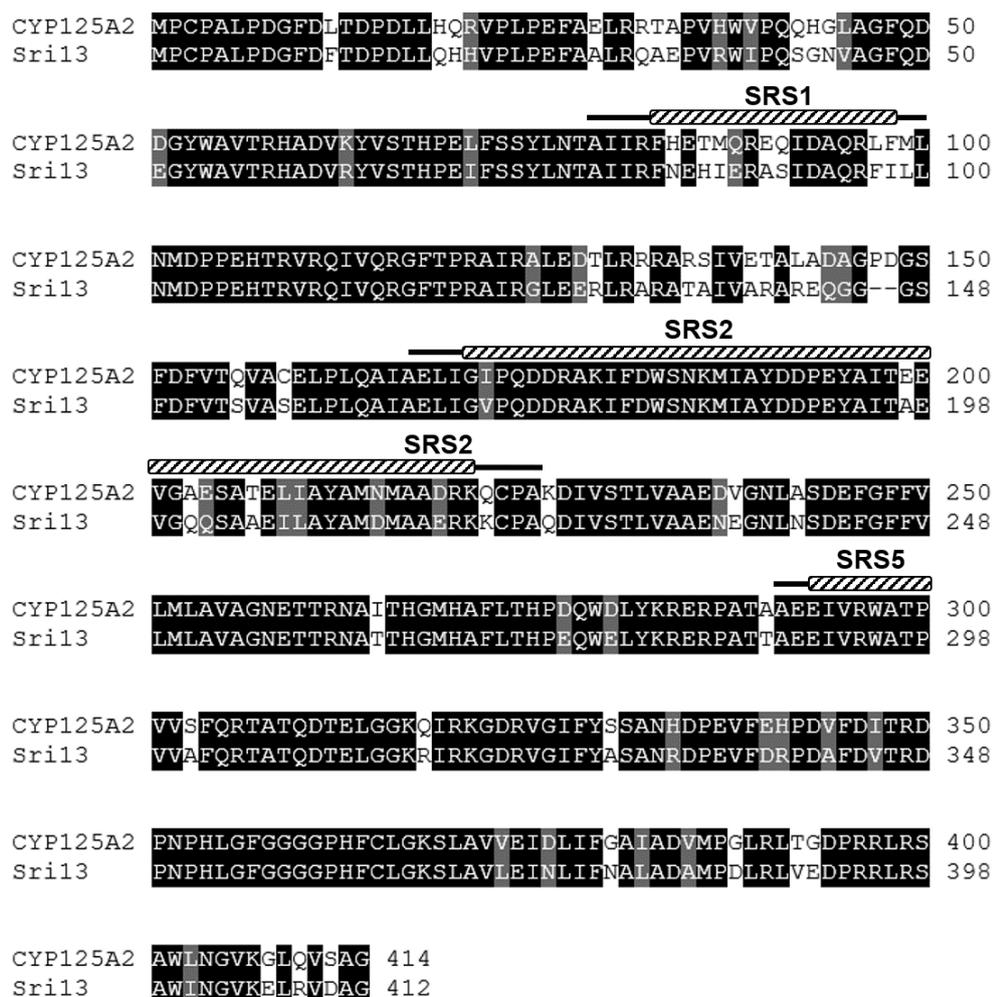
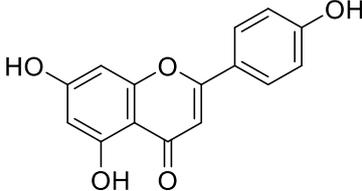
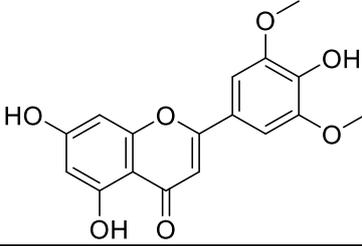
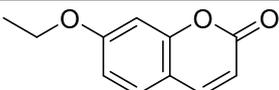
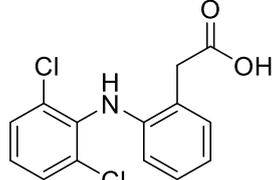
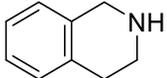


Figure 6.12. The alignment between CYP125A2 and Sri_CYP13. The identical and strongly similar residues are shaded in black and grey respectively. The three substrate recognition sites (SRS) are labeled as above.

In addition to the two flavonoids, three substrates including 7-ethoxycoumarin (EC), diclofenac (DIC), and 1,2,3,4-tetrahydroisoquinoline (THIQ) were also selected for the preliminary screening. The details of the five substrates were listed in Table 6.8. All five substrates selected for screening contain at least two ring structures and have distinctive signal peaks using the current HPLC method at 280 nm. Before conducting any enzyme assays for screening, the standard concentrations of substrates were prepared, and the correlation curves between concentrations and signal peaks were also calculated (Figure C4 to C8 in Appendix C).

Table 6.7. Substrates list for preliminary screening. All substrates were dissolved in DMSO to prepare 20 mM stock solutions. All substrates could be detected at 280 nm using HPLC method 1. The linear relationship between signal peak areas and the concentrations of substrates were calculated by Origin and recorded in Figure C4 to Figure C7 in Appendix C.

Abbreviation	Name	Structure	MW (g/mol)	Retention time (min)
APG	Apigenin		270.24	8.98
CHY	Chrysin		254.24	10.92
EC	7-Ethoxycoumarin		190.20	12.16
DIC	Diclofenac		318.13	10.31
THIQ	1,2,3,4-Tetrahydroisoquinoline		133.19	4.29

According to previous research⁽¹⁷³⁾, the flavonoids apigenin and chrysin were proven to be accepted by CYP107P2 and CYP125A2 respectively. The original enzyme assay was designed to use *E.coli* cell lysate containing CYP, putidaredoxin, and putidaredoxin reductase. The CYP genes were constructed on pET28 plasmid, while the redox partner genes were assembled on pDuet expression vector. The transformation of these two plasmids into *E.coli* led to the coexpression of CYP and redox partners *in vivo* at the same time. The estimated conversion rate of apigenin and chrysin was identified as 3% and 4% in the *E.coli* system respectively, which was indicated as being quite low⁽¹⁷³⁾. Based on these findings, it was confirmed the Pdx-PdR system could play the role of transferring

electrons to CYPs to fulfill the hydroxylation of either APG or CHY. Therefore, the Pdx-PdR system was also applied in current enzyme assays conducted by Sri_CYP03 and Sri_CYP13 (assay details are shown in section 2.3.3)

6.4.2 Screening results

These five substrates were used for preliminary screening to identify the potential activity, and they were used at different initial concentrations due to the different solubility of these substrates in 50 mM potassium phosphate solutions. The components of an assay included CYP, Pdx, PdR, NADH, and substrates, which were all dissolved in 50 mM potassium phosphate. A negative control reaction for each substrate was firstly constructed without NADH but other components and the initial substrate was quantified through HPLC. The reaction mixture was also analysed through HPLC, and the remaining substrate peaks were quantified. By comparing the initial substrates and remaining substrates, we can define any potential activity of either CYP03 or CYP13 on these substrates. As is shown in Figure 6.13, the negative controls were firstly prepared to establish the initial concentrations of each substrate used for screening assays conducted by either CYP_Sri03 and CYP_Sri13. For the determination of potential activity against the substrate, the decreasing of the substrate as well as rising of new peaks on the chromatogram should be observed. The remaining substrates for each reaction were estimated and shown in Figure 6.13. There was no observable decrease of those substrate being used in the CYP_Sri03 assays or CYP_Sri13 assays. The consumption of substrates might have been too low to be detected, so the chromatograms between negative controls and assay reactions were also compared to find any potential new peaks. Unfortunately, no new peaks was being identified, and detailed chromatogram was shown in Figure C9 (Appendix C).

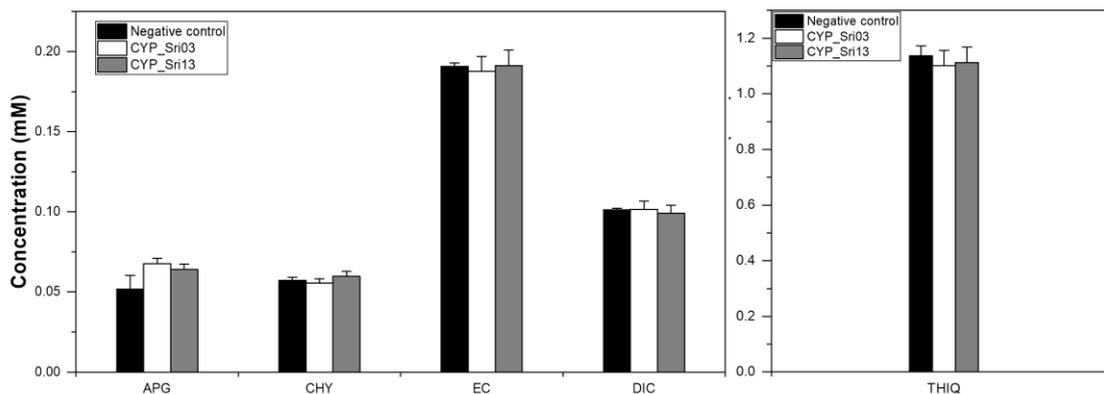


Figure 6.13. The potential usage of these substrates for enzyme reaction conducted by CYP_Sri03 and CYP_Sri13. The black bars represent the negative control reactions prepared without NADH. The white bars are the remaining substrates detected for CYP_Sri03 screening, and the grey bars represent the remaining substrates for CYP_Sri13 screening. By comparing the remaining substrates (white and grey bars) to the initial substrates (black bars), there was no observable decrease of the substrates.

6.5 Summary

In conclusion, there are two main objectives of the research activities carried out in Chapter 6, which are summarized below.

- The first objective was to identify all potential CYPs within the *S.rishiriensis* genome using a bioinformatics approach, which led to the discovery of 24 novel CYPs.
- The second objective was to test the established operon construct for potential CYP production in *E.coli*. The candidate CYPs were Sri_CYP03, Sri_CYP13, and Sri_CYP24. The expression of Sri_CYP03 and Sri_CYP13 was successfully carried out, which mainly relied on the coexpression of chaperone protein GroES and GroEL. However, the expression of Sri_CYP24 failed, and the chaperone coexpression strategy could not help either.

The well-developed next-generation sequencing techniques have led to the generation of genomic data with higher accuracy in the least amount of time. The complete genomic data contains all the information of that organism, and further bioinformatic investigations on genomic data can contribute to many other studies

such as understanding of complex phenotypes, resolve structurally complex regions on the genome, mining of functional proteins and so on⁽¹⁷⁷⁾. As is described in the background section, the genome data of *S.rishiriensis* was revealed by the Illumina sequencing platform, and it was preserved in Ward Lab for mining of novel enzymes and biosynthetic pathways. In 2017, the genome data *S.rishiriensis* were revealed, and it was the first time in the literature of acquiring the full genome data of this microbe. Since then, the genome data was involved in several projects on enzyme mining and synthetic cluster discovery. In this project, the primary aim is to identify and annotate all cytochrome P450s in *S.rishiriensis*, and the next aim is to incorporate identified cytochrome P450s from *S.rishiriensis* into the established *cyp-fd2-fdr* operon (studied in Chapter 4 and 5) for production in *E.coli*.

The genomic data of *S.rishiriensis* was stored as multiple unannotated contigs rather than a completely assembled genome sequence. Therefore, the sequence analysis software Geneious was used to group all unannotated contigs as a sequence library. The bioinformatic approach of analysing the sequence library included three steps, firstly building up a customized tBLASTn rules, and then using the conserved heme-binding motif of cytochrome P450s as the query to identify all heme-binding regions within the genome, and finally identify cytochrome P450 genes covering the heme-binding regions. According to the cytochrome P450 database recorded in Pfam (Family: P450; Accession number: PF00067), there are several conserved regions for a common cytochrome P450 including the GXXTT motif in I-helix, the EXXR motif in K-helix and the heme-binding motif. The heme-binding motif is the most conserved along with the entire P450 family, which could be displayed as **FXXGXXXCXG**. The four bold residues, phenylalanine, glycine, cysteine, and glycine, were strong representatives for constructing the hydrophobic pocket of heme ligand. Therefore, the identification of heme-binding regions leads to the identification of cytochrome P450 genes. For each identified heme-binding region, their coding sequences were also revealed on the genome. The next step is to identify if these coding sequences were within a complete gene with features of ribosome binding site, start codon, and stop codon. These identified genes were then sent to Pfam for enzyme classification to determine if they belong to the cytochrome P450 subfamilies.

Following this method, there were eventually 24 novel CYPs being discovered in the *S.rishiriensis* genome.

Only three cytochrome P450s, Sri_CYP03, Sri_CYP13, and Sri_CYP24, were subjected to heterologous expression in *E.coli*. The reason for selecting Sri_CYP03 and Sri_CYP13 for expression was their high identity to CYP107P2 (86.7%) and CYP125A2 (80.7%) from *S.avermitili*. Both CYP107P2 and CYP125A2 have been proven for their ability to hydroxylate (iso)flavonoids⁽¹⁷³⁾. It was hoping to find similar functions on Sri_CYP03 and Sri_CYP13. The reason for selecting Sri_CYP24 was its proposed β -hydroxylation on tyrosine. The functionalisation of tyrosine was one of the challenges we were interested in at the Ward group.

Several approaches were being applied for the attempts of producing these CYPs, including the use of different operon constructs, the use of the different amount of IPTG, and the attempts of chaperone coexpression. Eventually, the active form of Sri_CYP03 and Sri_CYP13 was achieved, but the production of Sri_CYP24 was still failed in this project. The detailed expression strategy is described below.

The BL21(DE3) strain and *cyp-fd2-fdr* operon were firstly used, which did not contribute to the production of active CYPs. The three-gene operon was then altered to the simpler single-gene construct, and the induction concentration of IPTG was reduced to 0.1 mM from the routine 0.5 mM. Both changes were meant to reduce the expression burden of the host cell, which may help the production. However, these optimisations did not lead to the production of active P450s but increased the production of the inactive form (P420). The last optimisation was conducted, including the use of chaperone proteins within *E.coli* host to help the heterologous expression of foreign proteins. The natural functions of GroES and GroEL are facilitating the folding of peptides and increasing the stability of protein products⁽¹⁷⁴⁾, so it was suspected these chaperone proteins may help folding of heme ligand into the heme-binding pocket as well. There were five different combinations of chaperone proteins being selected, including DnaK-DnaJ-GrpE-GroES-GroEL, DnaK-DnaJ-GrpE, GroES-GroEL, GroES-GroEL-Trigger factor, and Trigger factor. The *cyp* gene was constructed on the original Ampicillin resistant pUC origin plasmid, while the genes of chaperone proteins were

assembled on low copy number plasmids with pACYC origin. The coexpression of chaperone proteins in *E.coli* BL21 (DE3) resulted in the production of active Sri_CYP03 and Sri_CYP13, but it did not contribute to the expression of active Sri_CYP24 at all. It was also proven that the chaperon combinations of GroES-GroEL can lead to the highest production of active Sri_CYP03 and Sri_CYP13. It is unknown why the coexpression of chaperone did not contribute to the production of Sri_CYP24. Different from genes *sri_cyp03* and *sri_cyp13*, *sri_cyp24* is not an orphaned *cyp* gene within *S.rishiriensis* genome but exists in the natural gene cluster of coumermycin A1 in *S.rishiriensis*⁽¹⁶⁸⁾. In previous studies, instead of a single gene from the cluster, all genes within the coumermycin A1 cluster was cloned together and expressed. It was suspected that the activation or production of Sri_CYP24 may require the presence of other proteins within the genetic cluster. Based on the expression trials in *E.coli*, it is also suspected that the protein production factory in *E.coli* may naturally not support the producing or folding of Sri_CYP24. The expression of Sri_CYP24 still needs further investigation, which may include the use of *Streptomyces* strains as hosts or the attempts of cloning the entire cluster in *E.coli* for expression. Due to the limited time left for this project, the further studies of Sri_CYP24 expression was not carried out, and the preliminary screening was conducted for Sri_CYP03 and Sri_CYP13.

The preliminary screening pool was limited to five substrates, including 7-ethoxycoumarin (EC), diclofenac (DIC), 1,2,3,4-tetrahydroisoquinoline (THIQ), apigenin (APG) and chrysin (CHY). The enzyme reaction was conducted with 0.2 μ M CYPs, 5 μ M Pdx, 1 μ M PdR, and 0.5 mM NADH in 50 mM potassium phosphate buffer (pH 7.5). The three substrates EC, DIC, and THIQ were routinely selected as they were commonly tested in Chapter 5. The two flavonoids APG and CHY were accepted by CYP107P2 and CYP125A2 respectively. Moreover, the homologues of CYP107P2 and CYP125A2 are Sri_CYP03 (86.7% identity) and Sri_CYP13 (80.7% identity). The high sequence similarity may indicate similar enzyme structure as well as substrate selectivity. Therefore, the two flavonoids were also included in the preliminary screening pool. For each enzyme reaction, a negative control was also prepared without NADH but all other components at the same concentrations. The decrease of the substrate

was measured through HPLC for the determination of potential activity. Unfortunately, for both Sri_CYP03 and Sri_CYP13, there was no measurable decrease in the substrates or any potential new peaks of products (Figure 6.13). It was expected that none of the five substrates could be accepted by either of the CYPs, which brings to various potential reasons, including the naturally unacceptable substrates or inefficiency of the electron transfer proteins.

In summary, the current aims of the project, including the identification of CYPs from *S.rishiriensis* and establishing an expression system for CYP production, have been achieved. For future studies on Sri_CYP03 and Sri_CYP13 from *S.rishiriensis*, the substrate pool could be further expanded to other structurally similar two-ring substrates to these flavonoids, such as daidzein, genistein, curcumin, phloretin and so on. For future studies on Sri_CYP24, the expression system should be further optimised for the production of active CYPs, which includes the use of *Streptomyces* strains as the host or the attempts of cloning the entire cluster containing *sri_cyp24* in *E.coli* for production. The current approach of coexpression with chaperone systems did not result in the production of Sri_CYP24, and any optimisation on the construct or induction concentrations did not contribute to the production either.

7 Chapter 7: Identification and functional analysis of cytochrome P450 from a drain metagenome

7.1 Introduction

Metagenomics is traditionally defined as the study of the collective genomes in an environmental sample, and high-throughput sequencing technologies and computational pipelines allow us to access the largely untapped genomic resource of unculturable bacteria⁽¹⁷⁸⁾. The analysis of the collective genome data could potentially lead to the discovery of more robust enzymes, which could eventually contribute to both understanding and rational engineering of enzymes for biocatalysis. The metagenome projects conducted in Ward Group at UCL involve the mining of novel enzymes and pathways from microbiomes of various environments such as the human tongue, the parkway soil, the household drain, and so on.

The drain metagenome came from the microbial community living in a household drain for nearly two decades. The DNA sample was extracted from the drain and sequenced using the Illumina MiSeq platform, comparing to the Roche 454 Titanium platform which had been used to build other metagenome datasets in the Ward group. This generated 10 million individual sequence reads, which consists of the larger number of reads giving a much greater sequence depth, allowing better recovery of novel enzymes from the metagenome datasets. The computational approach was included in the service for metagenome assembly from the large number of short sequence reads, which generated highly fragmented thousands of contigs. The drain metagenome dataset used in this chapter contained 69962 unannotated contigs with the length varying from 500 bp to 67 kb, which was digitally saved at Ward Lab, University College London.

Research activity in this chapter aims to apply a bioinformatic approach to identify all cytochrome P450s in current drain metagenome datasets. The second aim is to incorporate novel *cyp* genes from metagenome into the established three-gene operon (discussed in Chapter 4) for cytochrome P450 expression and characterization. The performance of the *cyp-fd2-fdr* operon on producing active cytochrome P450s in this project will be compared with that of using common *cyp-pdx-pdr* operon for production in previous research. The microbial

community living in a household drain is suspected to have robust capability on degrading medium to long carbon chains, considering they live in the environment being constantly washed and incubated with domestic detergents. Therefore, it is quite interesting to find out if any identified cytochrome P450s that could potentially serve at the first step in the naturally existed carbon-chain metabolism. Other than cytochrome P450s specifically targeting carbon chains, there is also a large pool of other novel P450s in the metagenome datasets, which could potentially expand the known collection of microbial cytochrome P450s.

7.2 Bioinformatic analysis of drain metagenome

A similar strategy described in chapter 6 for P450s discovery was also used for mining cytochrome P450s from a drain metagenome. The mining approach started from identifying potential consensus region for heme-binding in every contig, and then to define the open reading frame that covers the heme-binding region, and finally to define the complete protein sequence and classify them into families. As is shown in Table 7.1, there were 44 heme-binding regions in total being found within the drain metagenome contigs, and only 16 of them were in a complete open reading frame. It means the complete gene and protein sequences for these 16 putative P450s could be properly extracted.

Table 7.1. Putative cytochrome P450s discovered from drain metagenome (Dmg) datasets. 44 heme-binding regions are being found in drain metagenome, and 16 of them are in a complete open reading frame. The positional information of each gene on contig was also recorded.

Putative CYPs with complete genes in the metagenome

ID	Size (aa)	Molecular weight (kDa)	Contig position
Dmg_CYP01	469	52.779	contig-120_98, 9776-11182
Dmg_CYP02	416	46.538	contig-120_144, 5196-6443
Dmg_CYP03	389	42.933	contig-120_259, 8843-10009
Dmg_CYP04	462	52.693	contig-120_759, 5939-4554
Dmg_CYP05	391	43.117	contig-120_2370, 1520-2692
Dmg_CYP06	404	45.675	contig-120_2781, 2518-3729
Dmg_CYP07	410	46.946	contig-120_2966, 2078-3307
Dmg_CYP08	412	46.422	contig-120_3564, 200-1435
Dmg_CYP09	412	46.768	contig-120_3564, 1512-2747

ID	Size (aa)	Molecular weight (kDa)	Contig position
Dmg_CYP10	429	47.867	contig-120_3589, 554-1840
Dmg_CYP11	429	48.371	contig-120_5735, 1344-2630
Dmg_CYP12	407	44.455	contig-120_8177, 1806-586
Dmg_CYP13	396	44.323	contig-120_9031, 513-1700
Dmg_CYP14	401	44.655	contig-120_9606, 102-1304
Dmg_CYP15	396	44.298	contig-120_11047, 1387-200
Dmg_CYP16	420	46.569	contig-120_18915, 1275-16

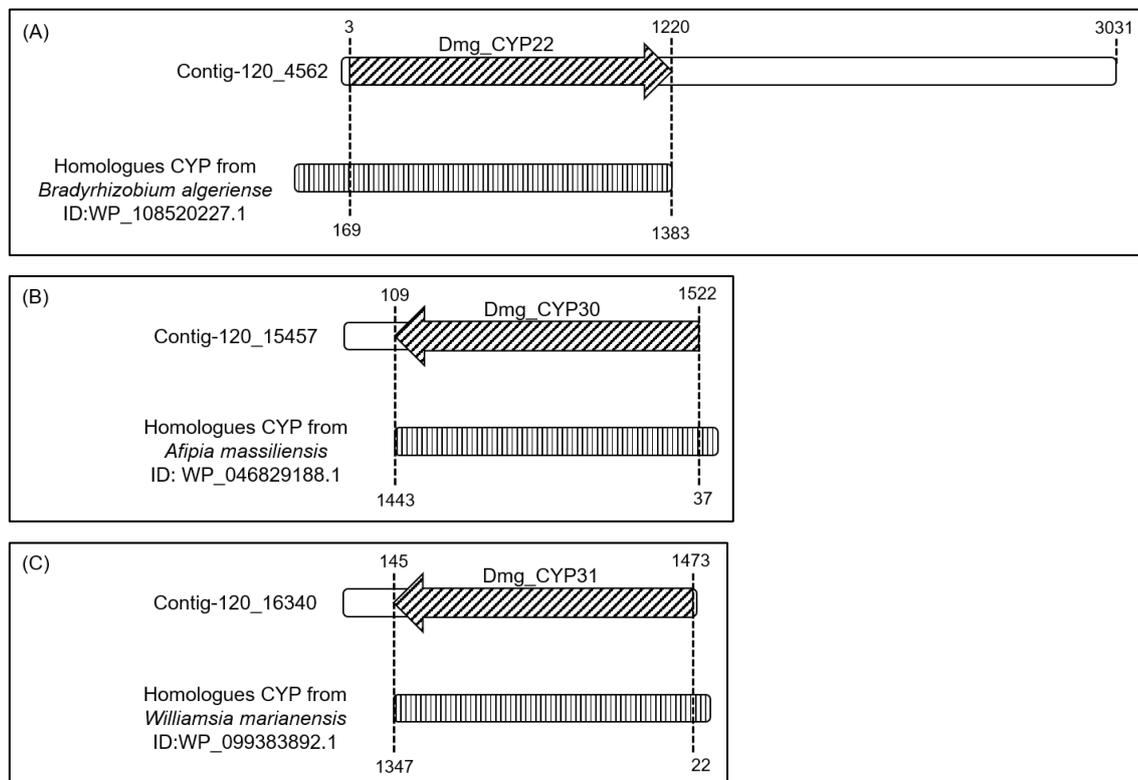
Truncated CYPs in the metagenome

ID	Partial gene length (bp)	Contig position
Dmg_CYP17	900	contig-120_1644, 4166-5065
Dmg_CYP18	537	contig-120_2028, 3-539
Dmg_CYP19	408	contig-120_2412, 3-410
Dmg_CYP20	783	contig-120_4112, 2418-3200
Dmg_CYP21	528	contig-120_4362, 2-529
Dmg_CYP22	1,218 ^a	contig-120_4562, 3-1220
Dmg_CYP23	996	contig-120_6422, 1510-2505
Dmg_CYP24	504	contig-120_7132, 3-506
Dmg_CYP25	585	contig-120_8172, 1-585
Dmg_CYP26	801	contig-120_11585, 997-1797
Dmg_CYP27	375	contig-120_11823, 1400-1774
Dmg_CYP28	783	contig-120_14419, 803-1585
Dmg_CYP29	837	contig-120_15306, 2-838
Dmg_CYP30	1,413 ^b	contig-120_15457, 109-1521
Dmg_CYP31	1,329 ^c	contig-120_16340, 145-1473
Dmg_CYP32	828	contig-120_16901, 2-829
Dmg_CYP33	675	contig-120_19112, 681-1355
Dmg_CYP34	621	contig-120_20473, 684-1304
Dmg_CYP35	810	contig-120_22277, 437-1246
Dmg_CYP36	456	contig-120_22982, 771-1226
Dmg_CYP37	1,104 ^d	contig-120_23828, 1-1104
Dmg_CYP38	1,074 ^e	contig-120_25832, 79-1152
Dmg_CYP39	381	contig-120_29907, 3-383
Dmg_CYP40	558	contig-120_33992, 448-1005
Dmg_CYP41	684	contig-120_35494, 1-684
Dmg_CYP42	537	contig-120_36877, 3-539

ID	Partial gene length (bp)	Contig position
Dmg_CYP43	504	contig-120_47470, 351-854
Dmg_CYP44	756	contig-120_50774, 68-823

a, b, c, d, e The length of the three DNA sequences are long enough for cytochrome P450 gene, but there is no coordinating start codon or ribosome binding site being detected on these contigs. When sending these truncated protein sequences through PBLAST, the protein sequences were identified as cytochrome P450, but there were missing peptides at N-terminal comparing to their potential homologous cytochrome P450s in database. Therefore, they were considered as truncated *cyp* genes.

For the last 28 sequences in Table 7.1, heme binding regions were being found on these sequences, but the sequences were in the truncated open reading frame due to limited contig length, so we can only suspect that those truncated genes are cytochrome P450s. In the 28 sequences, there were five sequences (Dmg_CYP22, Dmg_CYP30, Dmg_CYP31, Dmg_CYP37 and Dmg_CYP38) over 1 kb. The details of the five truncated sequences are shown in Figure 7.1. Each truncated CYP had partial alignment to their homologues, and the regions that were missing for these truncated CYPs were located outside the contigs. Therefore, these truncated CYPs could not be recognized as novel CYPs identified in the metagenome.



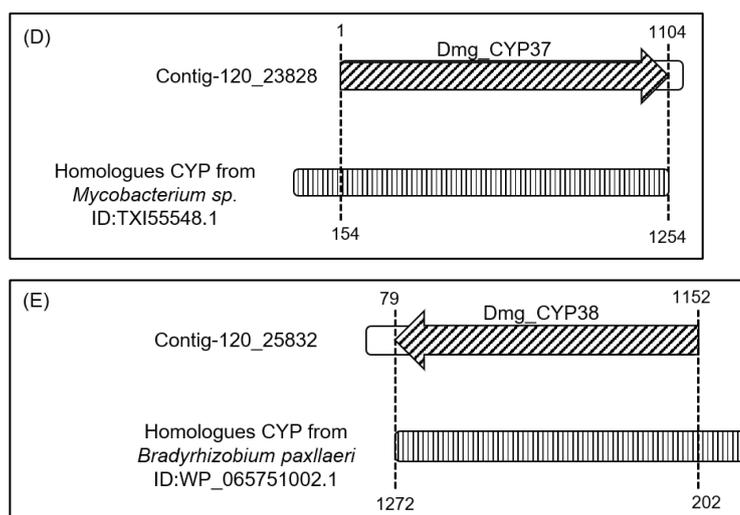


Figure 7.1. Examples of truncated CYPs found in metagenome contigs. The five sequences over 1 kb were described as examples. The arrow on contigs indicated the direction of the CYP gene. The closest homologues were displayed below each contig. The alignment regions between the homologues and truncated CYPs were labeled.

The 16 putative CYPs were compared against the bacterial P450s database to find closest homologues, which helped to define the families or subfamilies of these putative CYPs. As is shown in Table 7.2, all CYPs from the drain metagenome are novel, which has not been previously discovered. A few P450s are being considered as unknown family because they have identity lower than 40% to any known CYPs in the current library.

Table 7.2. Closest homologues to putative cytochrome P450s found in drain metagenome. The top two homologues with high pairwise identity were selected, and the CYP family as well as species of these homologues P450s were also defined.

CYP ID ^a	Closest homologues found in database ^b				Predicted CYP family
	Species	CYP family	Coverage ^c (%)	Identity ^d (%)	
Dmg_CYP01	<i>Novosphingobium aromaticivorans</i>	196A3	93	62.8	196A
	<i>Rhodopseudomonas palustris</i>	196A2	93	55.0	
Dmg_CYP02	<i>Novosphingobium aromaticivorans</i>	101D2	97	72.9	101D
	<i>Sphingomonas sp. SKA58</i>	101D3	98	71.5	
Dmg_CYP03	<i>Mycobacterium tuberculosis</i>	126A1	97	37.4	Unknown

CYP ID ^a	Closest homologues found in database ^b				Predicted CYP family
	Species	CYP family	Coverage ^c (%)	Identity ^d (%)	
	<i>Rhodococcus</i> sp. RHA1 Rha05835	125B1	95	36.9	
Dmg_CYP04	<i>Campylobacter jejuni</i> subsp. <i>jejuni</i> NCTC 11168	172A1	100	36.5	Unknown
	<i>Bradyrhizobium japonicum</i> USDA 110	201A1	98	33.4	
Dmg_CYP05	<i>Bradyrhizobium japonicum</i> USDA 110	193A1	99	69.3	193A
	<i>Agrobacterium tumefaciens</i>	104A1	94	36.3	
Dmg_CYP06	<i>Rhizobium etli</i> symbiotic plasmid p42d	127A4	99	45.4	127
	<i>Rhizobium</i> sp. BR816	127A2	98	43.8	
Dmg_CYP07	<i>Thermobifida fusca</i>	222A1	96	36.8	Unknown
	<i>Rhodococcus</i> sp. RHA1 Rha02853	254A2	97	30.8	
Dmg_CYP08	<i>Novosphingobium aromaticivorans</i>	153D1	95	56.0	153D
	<i>Sphingomonas</i> sp. HXN200	153D3	95	54.1	
Dmg_CYP09	<i>Novosphingobium aromaticivorans</i>	153D1	94	57.2	153D
	<i>Sphingomonas</i> sp. HXN200	153D3	97	54.6	
Dmg_CYP10	<i>Caulobacter crescentus</i> CB15	108G1	97	53.8	108
	<i>Novosphingobium aromaticivorans</i>	108D1	99	53.6	
Dmg_CYP11	<i>Caulobacter crescentus</i> CB15	108G1	98	61.3	108G
	<i>Ectocarpus bacterium</i>	108G2	99	56.4	
Dmg_CYP12	<i>Pseudomonas putida</i>	101A1	98	38.8	Unknown
	<i>Sphingomonas</i> sp. SKA58	101D3	97	35.4	
Dmg_CYP13	<i>Rhodopseudomonas palustris</i>	194A2	100	62.9	194A
	<i>Bradyrhizobium japonicum</i> USDA 110	194A1	98	63.4	
Dmg_CYP14	<i>Novosphingobium aromaticivorans</i>	101D1	97	40.1	101D
	<i>Novosphingobium aromaticivorans</i>	101C1	97	39.7	
Dmg_CYP15	<i>Roseiflexus</i> sp. RS-1	107AZ1	99	31.2	Unknown
	<i>Streptomyces avermitilis</i>	107L2	94	31.8	
Dmg_CYP16	<i>Sphingomonas paucimobilis</i>	152B1	94	59.1	152B
	<i>Azotobacter vinelandii</i>	152B2	96	50.7	

^a The name of all putative P450s in *S.rishiriensis*.

^b Closest homologues in Bacterial cytochrome P450 collection (archived by Dr David Nelson) and family classification searched in CYPED.

^c Coverage of query sequence generated by PBLAST comparison.

^d The pairwise identity for the aligned segment.

Reverse	GCGCGAATTCCTATTAGTGGTGGTGGTGGTGGTGGT G ACCCTTGGTGATGCGAAC
---------	--

^a The underlined regions indicate the restriction sites of *Nde*I and *Eco*RI, and the bold regions indicate the ribosome binding sites. The gene sequence for 6His-tag was also designed in both reverse primers.

^b The calculated annealing temperature for each primer generated by the NEB *T_m* calculator.

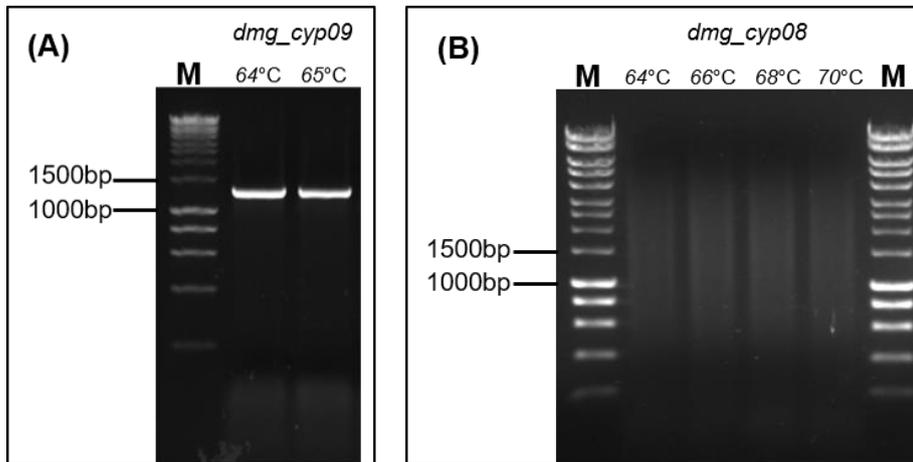


Figure 7.2. Amplification of *dmG_cyp08* and *dmG_cyp09* from drain metagenome datasets. Genomic PCR was conducted using primers described in Table 7.3. The length of expected PCR products should be around 1200 bp for both *dmG_cyp08* and *dmG_cyp09*.

As is shown in Figure 7.2, it seems the genomic PCR for the *dmG_cyp09* gene worked well at both 64°C and 65°C, but the genomic PCR for *dmG_cyp08* did not work at any of the calculated annealing temperatures. There were several potential approaches that could be applied to optimise the PCR to eventually amplified the target *dmG_cyp08* gene, which includes the use of another batch of drain metagenome sample, redesign of primers, or expanding the annealing temperatures. However, in this chapter, the gene synthesis approach rather than further optimisation PCR was selected to acquire the complete gene of *dmG_cyp08*. The synthesis was conducted at Eurofins Genomics, and the *dmG_cyp08* gene with 6-histidine tagged at C-terminal was synthesized with the *Nde*I site at the start codon, as well as *Eco*RI site after the stop codon. The gene fragments of *dmG_cyp08* were extracted through digestion with *Nde*I and *Eco*RI (Figure 7.3) for further assembly. Finally, the *dmG_cyp09* and *dmG_cyp08* gene fragments were assembled into the *cyp-fd2-fdr operon* on pQR2240. In addition,

the *dmg_cyp09* and *dmg_cyp08* gene fragments were also assembled into the *cyp-pdx-pdr operon* on pQR2290.

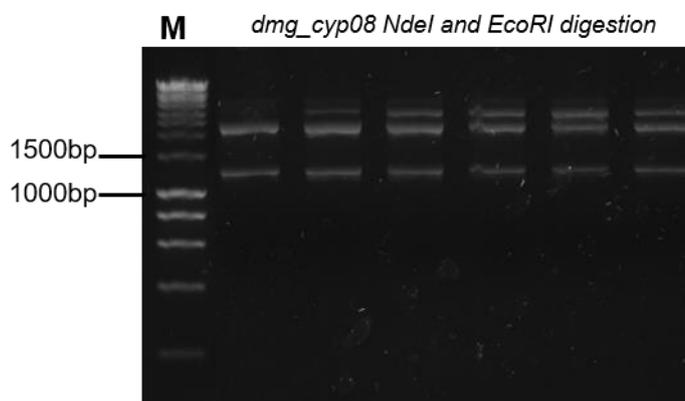


Figure 7.3. Digestion of *dmg_cyp08* with *NdeI* and *EcoRI*. Lane M stands for the 1kb marker purchased from Bioline. The rest six lanes were the same, loaded with the digestion mixture of synthetic plasmids containing the *dmg_cyp08* gene. The length of *dmg_cyp08* is 1236 bp, and all six digestions showing in the gel have bands around 1200 bp. These bands were determined as *dmg_cyp08* fragments, and then extracted and ligated with linearized pQR2240 or pQR2290.

7.3.2 Expression and purification of Dmg_CYP08 and Dmg_CYP09

As is discussed in chapter 4, the expression of several cytochrome P450s from *Streptomyces* could be enhanced when constructing the *cyp* gene, *ferredoxin* gene, and *ferredoxin reductase* gene together. Among all three-gene operons, the *cyp-fd2-fdr* operon generally performs better than others. Therefore, the *dmg_cyp08* or *dmg_cyp09* were firstly assembled with *fd2* and *fdr* to prepare the expression plasmids that were named as pQR2312 or pQR2313. In addition, *dmg_cyp08* or *dmg_cyp09* were also assembled with *pdx* and *pdr* to prepare the expression plasmids that were named as pQR2314 or pQR2315. The description of these four plasmids is mentioned in section 2.1.2.

The three-gene operons, *dmg_cyp08-fd2-fdr* and *dmg_cyp09-fd2-fdr*, were firstly used for expression under general conditions that being described in section 2.1.7. The concentration of IPTG was set to be 0.5 mM, and the concentration of P450 was measured through carbon monoxide assay (Figure 7.4). As is shown in Figure 7.4 (A) and (B), the expression of both structurally active Dmg_CYP08

and Dmg_CYP09 was achieved, according to the detectable peaks at 450 nm on both difference spectra. As is shown in Figure 7.4 (C) and (D), the expression of Dmg_CYP08 or Dmg_CYP09 was accompanied by the expression of putidaredoxin and putidaredoxin reductase (operon *dmg_cyp08-pdx-pdr* and *dmg_cyp09-pdx-pdr*). There were also structurally active P450s being produced because of the detectable peaks at 450 nm for both P450s spectra. However, peak height for Dmg_CYP08 at 450 nm in spectra (C) is slightly lower than that in spectra (A), and peak height for Dmg_CYP09 at 450 nm in spectra (D) is also lower than that in spectra (B). It means the expression of two cytochrome P450s is slightly enhanced when they were expressed with ferredoxins and ferredoxin reductases. Therefore, the *cyp-fd2-fdr* operon was used in further production of Dmg_CYP08 and Dmg_CYP09 in this chapter.

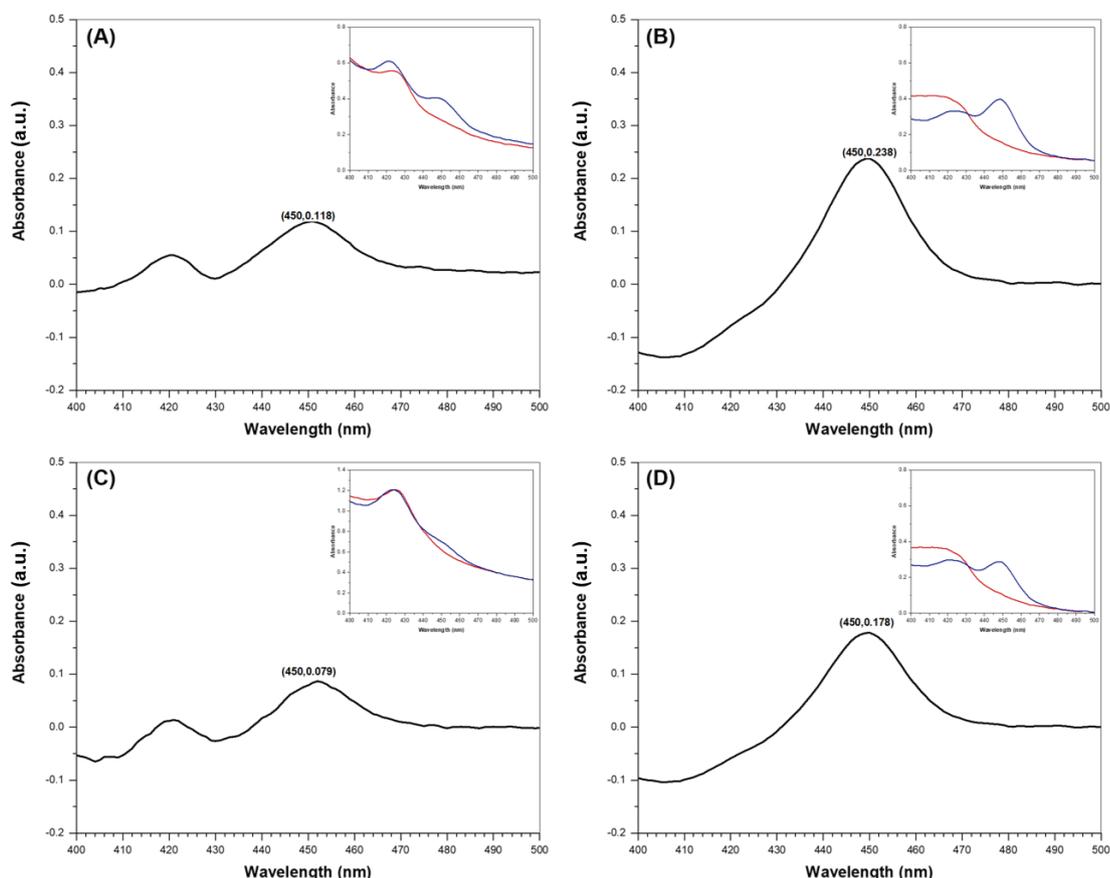


Figure 7.4. Expression of Dmg_CYP08 and Dmg_CYP09 using *cyp-fd2-fdr* construct. (A) The difference spectrum of Dmg_CYP08, and the expression plasmid was pQR2312 (*dmg_cyp08-fd2-fdr*); (B) The difference spectrum of Dmg_CYP09, and the expression plasmid was pQR2313 (*dmg_cyp09-fd2-fdr*); (C) The difference spectrum

of Dmg_CYP08, and the expression plasmid was pQR2314 (*dmg_cyp08-pdx-pdr*); (D) The difference spectrum of Dmg_CYP09, and the expression plasmid was pQR2315 (*dmg_cyp09-pdx-pdr*).

Because of the his-tag at C-terminal of both Dmg_CYP08 and Dmg_CYP09, targeted CYPs were purified through the nickel column (detailed method described in section 2.1.7). As is shown in Figure 7.5, the purified CYP08 and CYP09 were characterised on an SDS-PAGE gel, which indicates the correct protein sizes for both CYPs.

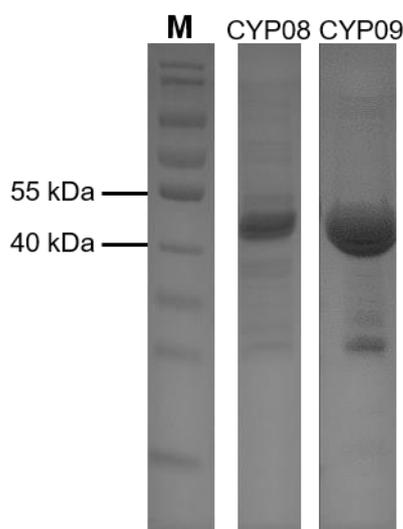


Figure 7.5. SDS-PAGE analysis of purified Dmg_CYP08 and Dmg_CYP09.

The molecular weights of Dmg_CYP08 and Dmg_CYP09 are 46.4 kDa and 46.7 kDa respectively. Lane M was the PageRuler prestained protein ladder (10-180kDa) used in SDS-PAGE. The other two lanes were loaded with purified Dmg_CYP08 and Dmg_CYP09 respectively. There are also some other bands in the purified CYP08 or CYP09 samples, but the majority of proteins are still the targeted CYPs.

.7.4 Preliminary substrate screening for Dmg_CYP08 and Dmg_CYP09

7.4.1 Alkane standards

According to Table 7.4, 7 alkanes were selected for the preliminary screening. The alkane standards were firstly prepared and analysed through GCMS (GC method is described in section 2.1.10.3), which established the retention time as well as a signal peak area for each substrate.

Table 7.4. Alkane substrates used for preliminary screening. Each substrate was dissolved in 100% ethanol to prepare the 200 mM stock solutions.

Name	Formula	Molecular weight (kDa)	Retention time (min)
Nonane	C ₉ H ₂₀	128.26	5.9
Decane	C ₁₀ H ₂₂	142.28	7.3
Undecane	C ₁₁ H ₂₄	156.31	8.4
Dodecane	C ₁₂ H ₂₆	170.33	9.5
Tridecane	C ₁₃ H ₂₈	184.36	10.5
Tetradecane	C ₁₄ H ₃₀	198.39	11.3
Pentadecane	C ₁₅ H ₃₂	212.41	12.2

The 200 mM alkane stock solutions were used to prepared 0.5 mM alkane standards in both ethyl acetate and potassium phosphate buffer (pH 7.5). For alkanes dissolved in ethyl acetate, they were analysed through GCMS directly, and their signal peaks on GC chromatogram represented the complete 0.5 mM alkanes.

Because all enzyme assays were conducted in aqueous solution and then extracted in an organic solvent for GC analysis, so it is necessary to prepare the standard concentration of alkanes in phosphate buffer as well. For alkanes dissolved in phosphate buffer, an ethyl acetate extraction was firstly performed, and the organic layer was sent for GC analysis. The extraction method was carried out as described in section 2.4.2. Considering potential alkane losses during extraction, the signal peaks on these GC chromatograms represented the actual alkanes could be detected in the extraction layer. By comparing the signal peaks of the complete alkane solutions and the extracted alkane solutions, the extraction efficiency for each alkane substrate could be estimated (Figure 7.6). The entire GC chromatograms for each alkane substrates were listed in section 12.2 Appendix D.

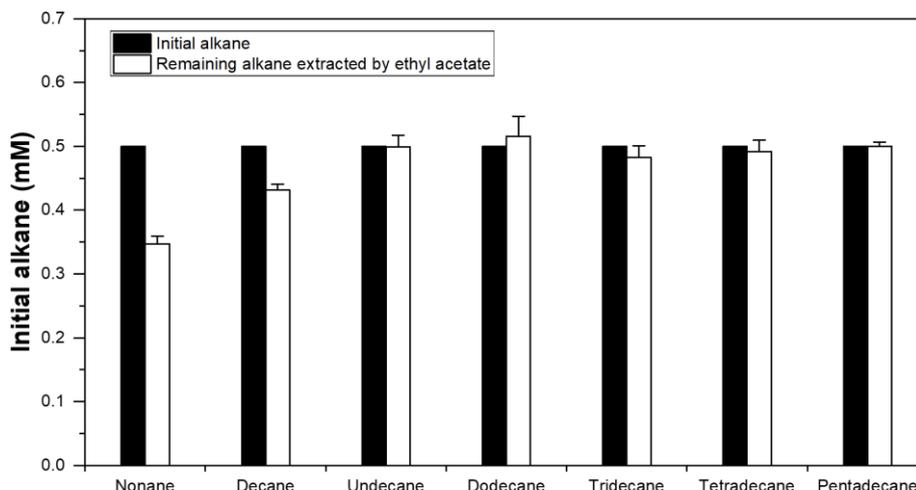


Figure 7.6. The extraction efficiency for different alkane substrates. The initial alkane was set to be 0.5 mM in both ethyl acetate and phosphate buffer. The alkane extraction by ethyl acetate could not guarantee the complete dissolving of alkanes from aqueous layer to the organic layer. The initial concentration was set to be 0.5 mM, and the remaining nonane after extraction reached 0.34 mM, which indicates a 30.5% nonane loss during the extraction process. The remaining decane after extraction was 0.43 mM, and a 13.7% decane loss during extraction was observed.

As is shown in Figure 7.6, the ethyl acetate extraction could lead to detectable substrate losses of nonane and decane, while for the other five longer chain alkanes, the substrate loss was not strongly indicated. It suggests that the extraction process does not guarantee the complete dissolving of alkanes from the aqueous phase, which can lead to inaccurate quantification of substrate consumption when it comes to enzyme kinetics studies. For example, after knowing the initial concentration of alkanes and the standard curve of alkane concentrations to signal peak areas on GC chromatogram, it is simple to determine the number of remaining alkanes after extraction as well as how much potential decrease of alkanes for the CYP initiated enzyme assay. However, the decrease of alkanes may be composed of the consumption by CYP and the substrate loss by the extraction process. Especially for nonane and decane extraction, the substrate loss reached to approximately 30.5% and 13.7%. Therefore, for preliminary screening in the next section, the exact concentration of alkanes was not calculated, but the signal peak areas were used to determine the change of alkane amount between reaction mixture and negative control.

7.4.2 Screening results of Dmg_CYP08 and Dmg_CYP09

The enzyme assay was designed with the components of three purified proteins (CYP, Pdx, and PdR), alkane as substrates, and NADH as cofactor. The reaction setup is shown in Table 7.5. All reactions were prepared in phosphate buffer, and ethyl acetate was used to extract the remaining alkanes for GC analysis. The negative control only contains the alkane in phosphate buffer, which was the initial concentration of alkanes prepared in the 1 mL reaction volume. The other two reactions Dmg_CYP08 screening and Dmg_CYP09 screening were conducted using CYP08 and CYP09 respectively, while all other reaction components were set to be the same.

Table 7.5. Reaction setup for alkane screenings. The total reaction volume was set to be 1 mL for all three reactions. 1 mL of ethyl acetate was used to extract alkanes from the three aqueous reactions.

Reaction	CYP (μM)	Pdx (μM)	PdR (μM)	Alkane (μM)	NADH (mM)
Negative control	0	0	0	500	0
Dmg_CYP08 screening	0.5	5	1	500	1
Dmg_CYP09 screening	0.5	5	1	500	1

As is shown in Figure 7.7, the initial alkanes and the remaining alkanes from screening reactions were represented by the signal peak areas on GC chromatograms. Both Dmg_CYP08 and Dmg_CYP09 seemed to react with decane ($\text{C}_{10}\text{H}_{22}$) due to the observable decrease of the peak area for the remaining decanes. Dmg_CYP09 was also considered to have reactions on the tridecane ($\text{C}_{13}\text{H}_{28}$). Dmg_CYP08 and Dmg_CYP09 also seemed to have very weak activities on tetradecane ($\text{C}_{14}\text{H}_{30}$) according to the slight decrease of the remaining alkane substrates detected. For nonane, undecane, dodecane, and pentadecane, the decreases of alkane substrates were not observed. In summary, the preliminary screening results have indicated the potential acceptance of decane and tetradecane by Dmg_CYP08, as well as the potential activity on decane, tridecane, and tetradecane by Dmg_CYP09. The GC chromatograms of the negative control reactions, Dmg_CYP08 screening, and Dmg_CYP09 screening were shown as figures in section 12.3.

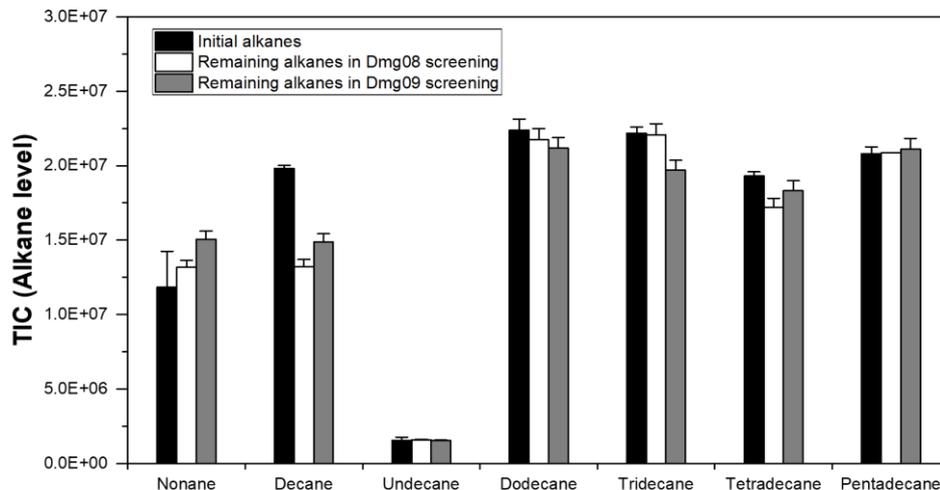


Figure 7.7. The potential usage of alkanes as substrates for enzyme reaction conducted by Dmg_CYP08 and Dmg_CYP09. The black bars represent the negative control reactions prepared without NADH, which was considered as the initial amount of substrate being detected in each system. The white bars are the remaining alkanes detected for Dmg_CYP08 screening, and the grey bars represent the remaining alkanes for Dmg_CYP09 screening.

7.5 Summary

In conclusion, there are two main objectives of the research activities carried out in Chapter 7, which are summarized below.

- The first objective was to apply a bioinformatic approach to identify all cytochrome P450s in current drain metagenome datasets. 16 novel CYPs were being discovered. Two CYPs, Dmg_CYP08, and Dmg_CYP09, were studied in this chapter due to their potential alkane hydroxylation functions.
- The second objective was to conduct the expression of target CYPs using the operon construct designed in chapter 4. The expression and purification of CYP08 and CYP09 were successfully carried out. A substrate panel containing mainly medium to long-chain alkanes was used to test the activity of target CYPs. There was no further product characterisation for the positive reactions in this chapter.

The drain metagenome dataset was analysed in this chapter for discovering any novel cytochrome P450s. The collection and processing of drain microbiomes were conducted at Ward Lab, and the process of acquiring these metagenomic

data was delivered by the Illumina NGS platform. Current high-throughput sequencing techniques have enabled the reveal of genetic information of ideally all microbes in that drain samples, but not just limited to those that are amenable to laboratory cultivation. Some potential challenges for metagenomic study include the genome assembly based on short sequence reads as well as potential complete taxonomy profiling of the microbiomes. However, for mining of particular enzymes from microbiomes, the metagenomic dataset containing fragmented contigs was usable enough. Once the genes of target enzymes were revealed, we could easily amplify these genes through PCR of metagenomic samples or synthesis of the known gene sequences.

There are many sequentially conserved regions for the cytochrome P450 family, and one of the most recognised regions is the heme-binding region. The heme-binding region was described to be **FXXGXXXCXG** in the Pfam database. The cysteine is 100% conserved in all heme-binding regions, and the other three bold amino acids are very highly conserved. Therefore, the bioinformatic approach was to first find all potential heme-binding regions on the metagenomic contigs using tBLASTn, and then to define the complete open reading frames covering these identified heme-binding regions. There were 69962 unannotated contigs with the length varying from 500 bp to 67 kb, and 44 heme-binding regions were found on these contigs. After manually identifying the potential open reading frames covering these heme-binding regions, only 16 complete cytochrome P450 genes were identified. The start codons, end codons, and RBS at proximity on the sequence were all identified for the 16 cytochrome P450 genes. The rest 28 heme-binding regions were sitting at truncated open reading frames, so it is difficult to either determine the complete gene sequences or define the family of cytochrome P450 they are belonging to. Both 16 complete *cyp* genes and 28 truncated sequences were recorded in Table 7.1. The 16 cytochrome P450s were compared against known microbial P450s using PBLAST, which aimed to determine the family or sub-family of each P450 belonging to. The classification into the same family or subfamily represents a relatively high level of sequential similarity, which may indicate potentially aligned functions on groups of substrates. The proposed family classification of 16 cytochrome P450s was summarized in Table 7.2.

Not all 16 cytochrome P450s were pulled out from the metagenome and used for expression. Because it was suspected these drain microbiomes may have robust capabilities on metabolising carbon chains due to their living environment, so we were looking for cytochrome P450 belonging to families that are known for alkane hydroxylation or fatty acid hydroxylation, for example, the CYP153 family and the CYP102 family. Among the 16 identified cytochrome P450s, two CYPs were belonging to the CYP153D subfamily, being named as Dmg_CYP08 and Dmg_CYP09. The *dmg_cyp08* and *dmg_cyp09* genes were separately assembled onto the *cyp-fd2-fdr* and *cyp-pdx-pdr* operons for expression in *E.coli* BL21(DE3). The *cyp-fd2-fdr* operon and the *cyp-pdx-pdr* operon were established and optimised in Chapter 4. The performance of using two operons on the production of Dmg_CYP08 and Dmg_CYP09 was compared. It turned out the use of *cyp-fd2-fdr* operon resulted in higher expression of Dmg_CYP08 and Dmg_CYP09, comparing to the use of *cyp-pdx-pdr* operon. The difference spectra were shown in Figure 7.3, indicating higher peaks at 450 nm for both CYPs assembled on to *cyp-fd2-fdr* operon.

The purification of histidine-tagged Dmg_CYP08 and Dmg_CYP09 was achieved through the nickel column, following the general protocol of protein manipulation described in Chapter 2. These purified CYPs were also characterised on SDS-PAGE for the identification of their molecular sizes. The preliminary substrate screening was conducted using not only the purified Dmg_CYP08 or Dmg_CYP09, but also the purified Pdx and PdR, to construct cell-free reactions. Considering the limited time for this project, a limited number of substrates were used in the preliminary screening, which included the medium-chain alkane (Table 7.4). The reactions were conducted in 50 mM phosphate buffer, and the harvesting of substrates and products was achieved through ethyl acetate extraction. All reactions were prepared with negative controls that contained all components but not NADH that served as the source of electrons, and the negative controls were used to establish the initial amount of alkanes. There was one interesting fact being revealed that the extraction efficiency of alkanes using ethyl acetate varied due to the different chain lengths. For shorter alkanes like nonane (C₉H₂₀) and decane (C₁₀H₂₂), there were around 30% and 13% losses respectively. For alkanes longer than decane, the loss of alkanes caused by

extraction into ethyl acetate was limited (Figure 7.5). In future, the extraction loss of alkanes should be taken into consideration, especially conducting the kinetic studies where various concentrations of substrates may be needed and the extraction loss should be evaluated. However, in the current stage of alkane screening, the extraction losses were not relevant to the determination of activity, because all reactions had the negative controls to be compared with. The only assumption made here was the extraction loss of the same alkane did not change, so it was sensible to have a comparison between alkanes extracted from reaction and alkanes extracted from the negative control.

The preliminary screening of seven alkanes with Dmg_CYP08 and Dmg_CYP09 indicated that decane ($C_{10}H_{22}$) and tetradecane ($C_{14}H_{30}$) were potentially accepted by both CYPs, and Dmg_CYP09 also had activities on tridecane ($C_{13}H_{28}$). In this chapter, there were no further studies being conducted on either the enzyme kinetics or the identification of potential products. The preliminary screening was aimed to identify what chain length these CYPs may prefer in that seven medium-chain alkane substrates. In future, other substrates and derivatives share the same chain length with decane, tetradecane or tridecane could be applied for screening, and the endpoint of screening is to discover any potential detergents like SDS (sodium dodecyl sulphate) or soap chemicals (mainly medium to long-chain fatty acids).

There are cytochrome P450s being involved in natural pathways of hydrocarbons degradation⁽¹⁷⁹⁾. These degradation pathways are providing a potential design of artificial pathways for cleaning hydrocarbon pollutants in the environment, and the discovering of robust cytochrome P450s offering effective oxy-functionalization of hydrocarbons could be an important part in that design. Therefore, the metagenome of drain microbes might be a viable option for mining evolutionarily advantageous CYPs on degrading hydrocarbons due to their natural detergent-rich living environment. For currently identified two CYPs (Dmg_CYP08 and Dmg_CYP09), they are all belonging to the subfamily of CYP153D that is known for alkane hydroxylation. Potential further studies are needed not only on the characterisation of enzyme activity against certain substrates but also on comparing to known CYPs coming from 153D subfamily

to evaluate the performance of CYPs from drain metagenome. There is no definite answer regarding whether Dmg08 and Dmg09 are the promising CYPs in hydrocarbon degradation, but the study of them offers a deeper understanding of the CYP153 subfamily on accepting hydrocarbon and prepare them for further biocatalysis in nature.

8 Chapter 8: Conclusion and future work

8.1 Self-sufficient cytochrome P450

The research program started with the study of self-sufficient P450s (Class VII cytochrome P450). The main reasons for working on self-sufficient P450 include that it is simpler to manipulate at the molecular level, and there is no involvement of other partner proteins when conducting reactions. The hunt for self-sufficient P450s (Class VII cytochrome P450) started from genome mining of three microbes including *Bacillus licheniformis* ATCC 14580, *Streptomyces avermitilis* MA-4680 and *Saccharopolyspora erythraea* NRRL2338. The main reasons for selecting these microbes included that there were complete genome sequences of these microbes available, and there were extracted genomic material at Ward lab ready for use. As the first project of the research program, it was straightforward and manageable for me who only had very limited experience in molecular biology. What was expected to achieve in this project included practising techniques on molecular manipulation and characterisation of macromolecules, setting up enzymatic reactions and substrate screening strategy, and finally understanding the process of studying CYP. In the end, all purposes were achieved, and every technique I had learned on studying the self-sufficient cytochrome P450 eventually contributed to the works on projects about other microbial CYPs.

The genome mining revealed that CYP102A7, CYP102D1, and CYP102G2 were the self-sufficient CYPs from *Bacillus licheniformis* ATCC 14580, *Streptomyces avermitilis* MA-4680 and *Saccharopolyspora erythraea* NRRL2338 respectively. Unfortunately, CYP102A7 was the only one being successfully cloned and expressed in this project. CYP102A7 was later characterised through SDS-PAGE and carbon monoxide assay. The amount of active CYP102A7 was measured to be 64 nmol per litre of cell culture. The yield of CYP102A7 was much lower than that confirmed in the previous CYP102A7 study (400 nmol/L cell culture)⁽¹²⁴⁾. A brief gap analysis between previous and current expression procedures indicates the main differences in concentration of IPTG and expression time. In this project, 1 mM IPTG was used for induction, and the expression lasted for 24 hours. In previous research, only 0.25 mM IPTG was used, and the expression lasted for

16 hours. For future expression of self-sufficient CYPs, the lower concentration of IPTG and shorter expression time was recommended.

8.2 Design and optimisation of operon constructs for Class I CYP expression

The main research activity of this PhD program was constructing the operon for Class I cytochrome P450 expression (Chapter 4). The key feature of the expression plasmid was a three-gene operon designed and built for microbial CYP expression. Several three-gene operons were designed, including the *cyp-fd1-fdr* operon, the *cyp-fd2-fdr* operon, the *cyp-fd_sery-fdr* operon, and the *cyp-pdx-pdr* operon. There are some common features among these operons, such as the distance between genes, the choices of restriction sites, as well as the positions of ribosome binding sites. The differences among these operons were the genes being assembled. All components within the operon were standardized, which allowed flexible combinations of P450, ferredoxin, and ferredoxin reductase genes. One of the purposes for building a three-gene operon was to explore the optimal construct for stable expression of high-quality actinomycete CYPs in *E.coli* BL21 (DE3), particularly for the expression of CYP105AB1 (*Saccharopolyspora erythraea* NRRL2338). The operon, containing the genes of ferredoxin 2 (*S.griseolus*) and ferredoxin reductase SCF15A (*S.coelicolor*), was proven to be an optimal construct for expression of CYP105AB1. The other purpose of building a three-gene operon was to set up a complete electron transfer system that could cooperate with the actinomycete CYP produced within *E.coli* host. In the future, the cell lysate or whole-cell with the operon design could be directly used for bioconversion without addition of foreign electron transferring proteins.

After expression, the characterisation of CYP105AB1 was then carried out (Chapter 5). The production of CYP105AB1 relied on using the *cyp-fd2-fdr* operon, and the purified CYP105AB1 was used as one of the components in enzyme assays for preliminary substrate screening. The calculated molecular weight of CYP105AB1 is around 44.5 kDa, which was also proven by the correct band position on SDS-PAGE gel (Figure 5.2). The *cyp105ab1-fd2-fdr* operon used in this chapter had an additional histidine tag constructed only on the C-

terminal of the *cyp105ab1* gene, which allowed purification of CYP105AB1 but no other proteins through nickel column. At the substrate screening stage, an enzyme reaction was constructed with purified proteins including CYP105AB1, putidaredoxin (Pdx) (*P.putida*), and putidaredoxin reductase (PdR) (*P.putida*). Because it was not known what electron transfer partner could naturally work with CYP105AB1, so the commonly used PdX-PdR system was selected for electron transfer.

The process of substrate screening included firstly identify potential consumption of substrate, and then to characterize the product of the reaction. The preliminary substrate pool was limited to 8 different small aromatic molecules such as 7-ethoxycoumarin, indole, tetrahydroisoquinoline, tyrosine, tyramine, phenylalanine, phenylethylamine, and diclofenac. The reaction mixture was analysed on HPLC to identify whether the amount of substrate decreased. Only diclofenac was consumed in the CYP105AB1 conducted reactions. In addition to the substrates peak (12.17 min) on the chromatogram, an unknown peak at 10.29 min was also observed, which was likely the product (Figure 5.4). In order to confirm the reaction with diclofenac as well as investigate the chemical structure of the potential product, the potential product peak and the substrate peak were isolated and analysed through LCMS. The product was one oxygen atom heavier than diclofenac on the molecular weight level. Therefore, the hydroxylation on diclofenac was confirmed and the new peak at 10.29 min was indeed the product. To identify where the hydroxyl group was inserted on diclofenac, the products were also scanned and structurally characterised through NMR, which revealed the hydroxylation was happening on 4' carbon of diclofenac (Figure 5.14).

Since diclofenac could be converted by CYP105AB1, the time-course kinetics of the reaction was also studied. The reaction containing 200 μ M diclofenac, 0.2 μ M CYP105AB1, 5 μ M putidaredoxin and 1 μ M putidaredoxin reductase was monitored for remaining substrate at 2 hours, 4 hours, 8 hours, 20 hours, 24 hours and 48 hours. There was only around 7% of total diclofenac being converted, which mainly happened in the first 4 hours. The calculated specific activity for the first 4 hours was 0.3 nmol/min/nmol of CYP105AB1. In this project, there were no further kinetic studies on CYP105AB1 on the reaction with diclofenac due to

the limited time left. For future kinetic studies, the strategy could include setting up various concentrations of diclofenac to react with 0.2 μM CYP105AB1, 5 μM putidaredoxin, and 1 μM putidaredoxin reductase, and further determination of the relationship between substrate concentrations and various reaction rates.

The original design of the three-gene operon studied in Chapter 4 was not only constructed for higher or more stable expression of CYPs in *E.coli*, but also aimed to prepare the *E.coli* cell lysate with a complete electron-donating system. Thus, for future catalytic reactions of potential chemicals, the cell lysate could be directly used rather than preparing the three purified proteins separately. The concentration of CYPs in cell lysate could be measured through carbon monoxide assay, but the concentrations of the other two-electron transferring proteins could not be directly measured. If the focus is to deliver enzyme conducted reactions rather than investigating the efficiency of electron transfer, the actual amount of electron transfer protein does not need to be determined. In this project, two sets of electron transferring systems, Fd2-FdR and Pdx-PdR, were expressed with CYP105AB1. The diclofenac conversion was conducted with either CYP105AB1-Fd2-FdR cell lysate or CYP105AB1-Pdx-PdR cell lysate, and the concentrations of CYP105AB1 for both reaction mixtures were adjusted to 1 μM . As is shown in Figure 5.11 (A) and (B), the reaction rate of CYP105AB1-Fd2-FdR was higher than that of CYP105AB1-Pdx-PdR on diclofenac. Therefore, it is proven that the constructed *cyp-fd2-fdr* operon not only facilitate the enhanced expression of CYP105AB1 in *E.coli*, but also provide a complete and usable electron-donating platform for potential chemical reactions.

In the preliminary screening, it was found that CYP105AB1 only react with diclofenac at a very low rate, which was expected performance as wild type cytochrome P450. Therefore, engineering of CYP105AB1 was carried out at the end of the project to discover robust enzyme candidates. Because there was no structural information for CYP105AB1, so those structurally defined homologues from the CYP105 subfamily were compared to reveal the common features CYP105AB1 may have. In this case, the sequences of homologues CYP105A1 and P450_MoxA were aligned with CYP105AB1, and they shared 46.3% and 57.6% identity respectively with CYP105AB1. There have been various mutation

sites being studied in the substrate-binding regions of CYP105A1⁽¹⁶⁶⁾ and P450_MoxA⁽¹⁶⁷⁾, and some of the target sites could also be found on CYP105AB1. For example, the conserved glycine (residue 294) on P450_MoxA was changed to aspartic acid, which led to enhanced activity on diclofenac. The conserved glycine on CYP105AB1 was located at residue 287, so the mutant G287D was created. Another mutant T108A of CYP105AB1 was created, because the T115A of P450_MoxA was studied and shown enhanced activity on diclofenac as well. Two mutated genes of CYP105AB1 were assembled into *cyp-fd2-fdr* operon and checked for integrity by sequencing. The expression of CYP105AB1-T108A and CYP105AB1-G287D was conducted in the same conditions applied for the expression of wild-type CYP105AB1. However, the mutants T108A and G287D were all expressed as P420 form, which was quite different from the active P450 form for wild type CYP105AB1. The expression of T108A and G287D has been conducted for at least three times, but all results in the inactive form of two mutant proteins. It was quite interesting that the change of one residue may influence the folding structure of CYP105AB1. Since there were no active mutants being produced, the reaction of diclofenac with two mutants was not conducted, and there was no conclusion on whether the mutants might have enhanced activity on diclofenac.

Further investigation on the inactive mutants as well as attempts of other expression approaches could be carried out in the future. The strategy of coexpression with chaperone proteins was proven to help CYPs from *S.rishiriensis* to fold active form P450 (results described in Chapter 6). Therefore, the coexpression of chaperone proteins and CYP105AB1 mutants is worth testing in the future.

8.3 Discovery of cytochrome P450s from *Streptomyces rishiriensis*

The well-developed next-generation sequencing techniques have led to the generation of genomic data with higher accuracy in the least amount of time. The genome data of *S.rishiriensis* was revealed by the Illumina sequencing platform, and it was preserved for genome mining of novel enzymes and biosynthetic pathways. In 2017, the genome data *S.rishiriensis* were revealed at Ward Lab, and it was the first time in the literature of acquiring the full genome data of this

microbe. Since then, the genome data was involved in several projects on enzyme mining and synthetic cluster discovery. In this project, the primary aim was to identify and annotate all cytochrome P450s in *S.rishiriensis*, and the next aim was to incorporate identified cytochrome P450s from *S.rishiriensis* into the established *cyp-fd2-fdr* operon (studied in Chapter 4 and 5) for production in *E.coli*.

The genomic data of *S.rishiriensis* was stored as multiple unannotated contigs rather than a completely assembled genome sequence. Therefore, the sequence analysis software Geneious was used to group all unannotated contigs as a sequence library. The bioinformatic approach of analysing the sequence library included three steps, firstly building up a customized tBLASTn rules, and then using the conserved heme-binding motif of cytochrome P450s as the query to identify all heme-binding regions within the genome, and finally identify cytochrome P450 genes covering the heme-binding regions. According to the cytochrome P450 database recorded in Pfam (Family: P450; Accession number: PF00067), there are several conserved regions for a common cytochrome P450 including the GXXTT motif in I-helix, the EXXR motif in K-helix and the heme-binding motif. The heme-binding motif is the most conserved along with the entire P450 family, which could be displayed as **FXXGXXXCXG**. The four bold residues, phenylalanine, glycine, cysteine, and glycine, were strong representatives for constructing the hydrophobic pocket of heme ligand. Therefore, the identification of heme-binding regions leads to the identification of cytochrome P450 genes. For each identified heme-binding region, their coding sequences were also revealed on the genome. The next step is to identify if these coding sequences were within a complete gene with features of ribosome binding site, start codon, and stop codon. These identified genes were then sent to Pfam for enzyme classification to determine if they truly belong to the cytochrome P450 subfamilies. There were 24 novel CYP being found in the *S.rishiriensis* genome.

Due to the limited project length, only three cytochrome P450s, Sri_CYP03, Sri_CYP13, and Sri_CYP24, were subjected to cloning and heterologous expression in *E.coli*. There were several approaches being applied for the attempts of producing these CYPs, including the use of different operon

constructs, the use of the different amounts of IPTG, and the attempts of chaperone coexpression. Eventually, the active form of Sri_CYP03 and Sri_CYP13 was achieved, but the production of Sri_CYP24 was still failed in this project. The detailed expression strategy is described below.

The BL21(DE3) strain and *cyp-fd2-fdr* operon were firstly used, which did not contribute to the production of active CYPs. The three-gene operon was then altered to the simpler single-gene construct, and the induction concentration of IPTG was reduced to 0.1 mM from the routine 0.5 mM. Both changes were meant to reduce the expression burden of the host cell, which may help the production. However, these optimisations did not lead to the production of active P450s but increased the production of the inactive form (P420). The last optimisation was conducted, including the use of chaperone proteins within *E.coli* host to help the heterologous expression of foreign proteins. The natural functions of GroES and GroEL are facilitating the folding of peptides and increasing the stability of protein products⁽¹⁷⁴⁾, so it was suspected these chaperone proteins may help folding of heme ligand into the heme-binding pocket as well. There were five different combinations of chaperone proteins being selected, including DnaK-DnaJ-GrpE-GroES-GroEL, DnaK-DnaJ-GrpE, GroES-GroEL, GroES-GroEL-Trigger factor, and Trigger factor. The *cyp* gene was constructed on the original Ampicillin resistant pUC origin plasmid, while the genes of chaperone proteins were assembled on low copy number plasmids with pACYC origin. The coexpression of chaperone proteins in *E.coli* BL21 (DE3) resulted in the production of active Sri_CYP03 and Sri_CYP13, but it did not contribute to the expression of active Sri_CYP24 at all. It was also proven that the chaperon combinations of GroES-GroEL can lead to the highest production of active Sri_CYP03 and Sri_CYP13. It is unknown why the coexpression of chaperone did not contribute to the production of Sri_CYP24. Different from genes *sri_cyp03* and *sri_cyp13*, *sri_cyp24* is not an orphaned *cyp* gene within the *S.rishiriensis* genome but exists in the natural gene cluster of coumermycin A1 in *S.rishiriensis*⁽¹⁶⁸⁾. In previous studies, instead of a single gene from the cluster, all genes within the coumermycin A1 cluster was cloned together and expressed. It was suspected that the activation or production of Sri_CYP24 may require the presence of other proteins within the genetic cluster. The expression of Sri_CYP24 still needs

further investigation, which may include the use of *Streptomyces* strains as hosts or the attempts of cloning the entire cluster in *E.coli* for expression. In further studies of Sri_CYP24, expression was not carried out, and the preliminary screening was conducted for Sri_CYP03 and Sri_CYP13.

The preliminary screening pool was limited to five substrates, including 7-ethoxycoumarin (EC), diclofenac (DIC), 1,2,3,4-tetrahydroisoquinoline (THIQ), apigenin (APG) and chrysin (CHY). The enzyme reaction was conducted with 0.2 μ M CYPs, 5 μ M Pdx, 1 μ M PdR, and 0.5 mM NADH in 50 mM potassium phosphate buffer (pH 7.5). The three substrates EC, DIC, and THIQ were routinely selected as they were commonly tested in Chapter 5. The two flavonoids APG and CHY were accepted by CYP107P2 and CYP125A2 respectively. Moreover, the homologues of CYP107P2 and CYP125A2 are Sri_CYP03 (86.7% identity) and Sri_CYP13 (80.7% identity). The high sequence similarity may indicate similar enzyme structure as well as substrate selectivity. Therefore, the two flavonoids were also included in the preliminary screening pool. For each enzyme reaction, a negative control was also prepared without NADH but all other components at the same concentrations. The decrease of the substrate was measured through HPLC for the determination of potential activity. Unfortunately, for both Sri_CYP03 and Sri_CYP13, there was no measurable decrease in the substrates or any potential new peaks of products (Figure 6.13). It was expected that none of the five substrates could be accepted by either of the CYPs, which brings to various potential reasons, including the naturally unacceptable substrates or inefficiency of the electron transfer proteins.

In summary, the current aims of the project, including the identification of CYPs from *S.rishiriensis* and establishing an expression system for CYP production, have been achieved. For future studies on Sri_CYP03 and Sri_CYP13 from *S.rishiriensis*, the substrate pool could be further expanded to other structurally similar two-ring substrates to these flavonoids, such as daidzein, genistein, curcumin, phloretin and so on. For future studies on Sri_CYP24, the expression system should be further optimised for the production of active CYPs, which includes the use of *Streptomyces* strains as the host or the attempts of cloning the entire cluster containing *sri_cyp24* in *E.coli* for production. The current

approach of coexpression with chaperone systems did not result in the production of Sri_CYP24, and any optimisation on the construct or induction concentrations did not help the production either. If the expression of Sri_CYP24 could be worked out in the future, the functionalization of tyrosine by Sri_CYP24 could be integrated into pathways for alkaloid synthesis that are being developed at Ward lab.

8.4 Discovery of cytochrome P450s from drain metagenome

The drain metagenome dataset was analysed for the discovery of novel cytochrome P450s. The collection and processing of drain microbiomes were conducted at Ward Lab, and these metagenomic data were delivered by the Illumina NGS platform. Current high-throughput sequencing techniques have enabled the reveal of genetic information of ideally all microbe in that drain samples, but not just limited to those that are amenable to laboratory cultivation. Some potential challenges for metagenomic study include the genome assembly based on short sequence reads as well as potential complete taxonomy profiling of the microbiomes. However, for mining of novel enzymes from microbiomes, the metagenomic dataset containing fragmented contigs was usable enough. Once the genes of target enzymes were revealed, we could easily amplify these genes through PCR of metagenomic samples or synthesis of the known gene sequences.

There are many sequentially conserved regions for the cytochrome P450 family, and one of the most recognised regions is the heme-binding region. The heme-binding region was described to be **FXXGXXXCXG** in the Pfam database. The cysteine is 100% conserved in all heme-binding regions, and the other three bold amino acids are very highly conserved. Therefore, the bioinformatic approach was to first find all potential heme-binding regions on the metagenomic contigs using tBLASTn, and then to define the complete open reading frames covering these identified heme-binding regions. There were 69962 unannotated contigs with the length varying from 500 bp to 67 kb, and 44 heme-binding regions were found on these contigs. After manually identifying the potential open reading frames covering these heme-binding regions, only 16 complete cytochrome P450 genes were identified. The start codons, end codons, and RBS at proximity on

the sequence were all identified for the 16 cytochrome P450 genes. The rest 28 heme-binding regions were sitting at truncated open reading frames, so it is difficult to either determine the complete gene sequences or define the family of cytochrome P450 they are belonging to.

Not all 16 cytochrome P450s were pulled out from the metagenome and used for expression. Because it was suspected these drain microbiomes may have robust capabilities on metabolising carbon chains due to their living environment, so we were looking for cytochrome P450 belonging to families that are known for alkane hydroxylation or fatty acid hydroxylation, for example, the CYP153 family and the CYP102 family. Among the 16 identified cytochrome P450s, two CYPs were belonging to the CYP153D subfamily, being named as Dmg_CYP08 and Dmg_CYP09. The *dmg_cyp08* and *dmg_cyp09* genes were separately assembled onto the *cyp-fd2-fdr* and *cyp-pdx-pdr* operons for expression in *E.coli* BL21(DE3). It turned out the use of *cyp-fd2-fdr* operon resulted in higher expression of Dmg_CYP08 and Dmg_CYP09, comparing to the use of *cyp-pdx-pdr* operon.

The preliminary substrate screening was conducted using not only the purified Dmg_CYP08 or Dmg_CYP09, but also the purified Pdx and PdR, to construct cell-free reactions. Considering the limited time for this project, a limited number of substrates were selected in the preliminary screening. The reactions were conducted in 50 mM phosphate buffer, and the harvesting of substrates and products was achieved through ethyl acetate extraction. All reactions were prepared with negative controls that contained all components but not NADH that served as the source of electrons, and the negative controls were used to establish the initial amount of alkanes. There was one interesting fact being revealed that the extraction efficiency of alkanes using ethyl acetate varied due to the different chain lengths. For shorter alkanes like nonane (C₉H₂₀) and decane (C₁₀H₂₂), there were around 30% and 13% losses respectively. For alkanes longer than decane, the loss of alkanes caused by extraction into ethyl acetate was limited (Figure 7.5). In future, the extraction loss of alkanes should be taken into consideration, especially conducting the kinetic studies where various concentrations of substrates may be needed and the extraction loss should be

evaluated. However, in the current stage of alkane screening, the extraction losses were not relevant to the determination of activity, because all reactions had the negative controls to be compared with. The only assumption made here was the extraction loss of the same alkane did not change, so it was sensible to have a comparison between alkanes extracted from reaction and alkanes extracted from the negative control.

The preliminary screening of seven alkanes with Dmg_CYP08 and Dmg_CYP09 indicated that decane ($C_{10}H_{22}$) and tetradecane ($C_{14}H_{30}$) were potentially accepted by both CYPs, and Dmg_CYP09 also might have activities on tridecane ($C_{13}H_{28}$). For this project, there were no further studies being conducted on either the enzyme kinetics or the identification of potential products. The preliminary screening was aimed to identify what chain length these CYPs may prefer in that seven medium-chain alkane substrates. In future, other substrates and derivatives share the same chain length with decane, tetradecane or tridecane could be applied for screening, and the endpoint of screening is to discover any potential detergents like SDS (sodium dodecyl sulphate) or soap chemicals (mainly medium to long-chain fatty acids).

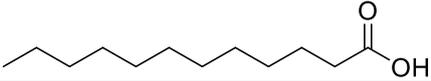
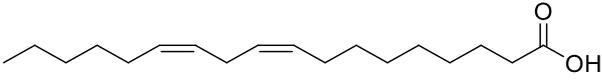
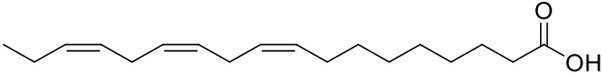
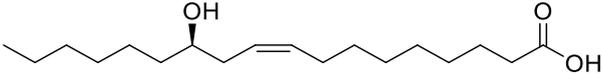
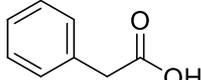
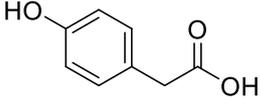
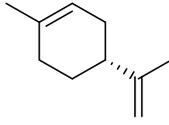
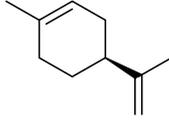
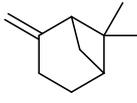
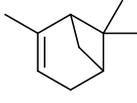
There are cytochrome P450s being involved in natural pathways of hydrocarbons degradation⁽¹⁷⁹⁾. These degradation pathways are providing a potential design of artificial pathways for cleaning hydrocarbon pollutants in the environment, and the discovering of robust cytochrome P450s offering effective oxy-functionalization of hydrocarbons could be an important part in that design. Therefore, the metagenome of drain microbes might be a viable option for mining evolutionarily advantageous CYPs on degrading hydrocarbons due to their natural detergent-rich living environment. For currently identified two CYPs (Dmg_CYP08 and Dmg_CYP09), they are all belonging to the subfamily of CYP153D that is known for alkane hydroxylation. Potential further studies are needed not only on the characterisation of enzyme activity against certain substrates, but also on comparing to known CYPs coming from 153D subfamily to evaluate the performance of CYPs from drain metagenome. There is no definite answer regarding whether Dmg08 and Dmg09 are the promising CYPs in hydrocarbon degradation, but the study of them offers a deeper understanding

of the CYP153 subfamily on accepting hydrocarbon and prepare them for further biocatalysis in nature.

9 Appendix A

This appendix contains supplementary information for Chapter 3.

Table A1. Substrate screening list for CYP102A7. All substrates were dissolved in DMSO to prepare a 10 mM stock solution. 2 μL of the stock solution was added to a 100 μL reaction system to reach a final substrate concentration of 200 μM .

Name	Molecular formula	Structure
Lauric acid	$\text{C}_{12}\text{H}_{24}\text{O}_2$	
Linoleic acid (cis,cis-9,12)	$\text{C}_{18}\text{H}_{32}\text{O}_2$	
γ -Linolenic acid (GLA) (ω -6 fatty acids)	$\text{C}_{18}\text{H}_{30}\text{O}_2$	
Ricinoleic acid (12-hydroxy-9-cis) (ω -9 fatty acids)	$\text{C}_{18}\text{H}_{34}\text{O}_3$	
Phenylacetic acid	$\text{C}_8\text{H}_8\text{O}_2$	
4-hydroxyphenylacetic acid	$\text{C}_8\text{H}_8\text{O}_3$	
R-(+)-Limonene	$\text{C}_{10}\text{H}_{16}$	
S-(-)-Limonene	$\text{C}_{10}\text{H}_{16}$	
(-)- β -Pinene	$\text{C}_{10}\text{H}_{16}$	
(+)- α -Pinene	$\text{C}_{10}\text{H}_{16}$	

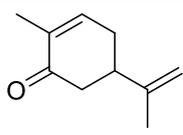
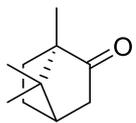
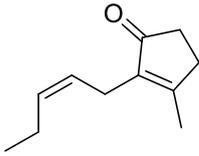
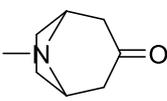
Name	Molecular formula	Structure
(+)-Carvone	C ₁₀ H ₁₄ O	
(1R)-(+)-Camphor	C ₁₀ H ₁₆ O	
cis-Jasmone	C ₁₁ H ₁₆ O	
Tropinone	C ₈ H ₁₃ NO	

Table A2. The ENA records of complete genome sequences of *Bacillus licheniformis* ATCC 14580, *Streptomyces avermitilis* MA-4680 and *Saccharopolyspora erythraea* NRRL2338. The website of ENA is <https://www.ebi.ac.uk/ena>, and the detailed genome information can be found through the corresponding accession numbers.

Strain	Genome accession number
<i>Bacillus licheniformis</i> ATCC 14580	CP034569
<i>Streptomyces avermitilis</i> MA-4680	AP017380
<i>Saccharopolyspora erythraea</i> NRRL2338	AM420293

Table A3. All failed PCR attempts for extracting *cyp102d1* and *cyp102g2* from their genomes. The primers used are shown below in the first table below. The reaction set-up and thermal cycling conditions are following the PCR profile guidelines described in section 2.1.6.1. The multiple attempts with different initial amounts of genomic materials, various annealing temperatures or various cycles are listed in the second table below.

Gene	Primers ^a		T _m (°C) ^b
<i>cyp102d1</i>	Forward	5'-GCCGCCCATATGACCACACAGCCCGAGACC-3'	66
	Reverse	5'-GCCGCCAAGCTTGGCGAAGACGTCCTGCTG-3'	63
<i>cyp102g2</i>	Forward	5'-GCCGCCCATATGACCCAGACCCCGCTCC-3'	66
	Reverse	5'-GCCGCCGCGGCCGCGCCCGCGTAGACGTCTTC-3'	62

^a The underlined regions indicate the restriction sites.

^b The calculated annealing temperature for each primer generated by NEB T_m calculator.

Attempts	Conditions
1	50 ng genomic material, $T_m = 68^\circ\text{C}$, 30 cycles
2	50 ng genomic material, $T_m = 66^\circ\text{C}$, 30 cycles
3	50 ng genomic material, $T_m = 64^\circ\text{C}$, 30 cycles
4	50 ng genomic material, $T_m = 62^\circ\text{C}$, 30 cycles
5	100 ng genomic material, $T_m = 68^\circ\text{C}$, 30 cycles
6	100 ng genomic material, $T_m = 66^\circ\text{C}$, 30 cycles
7	100 ng genomic material, $T_m = 64^\circ\text{C}$, 30 cycles
8	100 ng genomic material, $T_m = 62^\circ\text{C}$, 30 cycles
9	100 ng genomic material, $T_m = 68^\circ\text{C}$, 20 cycles
10	100 ng genomic material, $T_m = 66^\circ\text{C}$, 20 cycles
11	100 ng genomic material, $T_m = 64^\circ\text{C}$, 20 cycles
12	100 ng genomic material, $T_m = 62^\circ\text{C}$, 20 cycles

10 Appendix B

This appendix contains supplementary information for Chapter 4 and Chapter 5.

10.1 The gene and protein sequences

10.1.1 *cyp105a1*, *cyp105b1*, *fd1* and *fd2* from *Streptomyces griseolus*

> *cyp105a1*

```
ATGACCGATACCGCCACGACGCCCCAGACCACGGACGCACCCGCCTTCCCGAGCAACCG
GAGCTGTCCCTACCAGTTACCGGACGGCTACGCCAGCTCCGGGACACCCCGGCCCCC
TGCACCGGGTGACGCTCTACGACGGCCGTGAGGCGTGGGTGGTGACCAAGCACGAGGCC
GCGCGCAAACCTGCTCGGCGACCCCGGCTGTCCTCCAACCGGACGGACGACAACTTCCC
CGCCACGTACCGCGCTTCGAGGCCGTCCGGGAGAGCCCGCAGGCGTTTCATCGGCCTG
GACCCGCCCGAGCACGGCACCCGGCGGGGATGACGATCAGCGAGTTCACCGTCAAGC
GGATCAAGGGCATGCGCCCCGAGGTCGAGGAGGTGGTGCACGGCTTCCTCGACGAGATG
CTGGCCGCCGGCCCGACCGCCGACCTGGTCAGTCAGTTCGCGCTGCCGGTGCCCTCCAT
GGTGATCTGCCGACTCCTCGGCGTGCCCTACGCCGACCACGAGTTCTTCCAGGACGCGA
GCAAGCGGCTGGTGCAGTCCACGGACGCGCAGAGCGCGCTCACCGCGCGGAACGACCT
CGCGGGTTACCTGGACGGCCTCATCACCCAGTTCCAGACCGAACCGGGCGCGGGCCTGG
TGGGCGCTCTGGTGCAGGACCGAGCTGGCCAACGGCGAGATCGACCGTGAGGAACTGATC
TCCACCGCGATGCTGCTCCTCATCGCCGGCCACGAGACCACGGCCTCGATGACCTCCCTC
AGCGTGATCACCTGCTGGACCACCCGAGCAGTACGCCGCCCTGCGCGCCGACCGCAG
CCTCGTGCCCGGCGCGGTGGAGGAACTGCTCCGCTACCTCGCCATCGCCGACATCGCGG
GCGGCCGCGTCGCCACGGCGGACATCGAGGTCGAGGGGCACCTCATCCGGGCCGGCGA
GGGCGTGATCGTCGTAACCTCGATAGCCAACCGGGACGGCACGGTGTACGAGGACCCGG
ACGCCCTCGACATCCACCGCTCCGCGCGCCACCACCTCGCCTTCGGCTTCGGCGTGAC
CAGTGCCCTGGGCCAGAACCTCGCCCGGCTGGAGCTGGAGGTCATCCTAACGCCCTCAT
GGACCGCGTCCCAGCGCTGCGACTGGCCGTCCCGTTCGAGCAGTTGGTGCTGCGGCCG
GGTACGACGATCCAGGGCGTCAACGAACTCCCGGTACCTGGTGA
```

>*cyp105b1*

```
ATGACGACCGCAGAACGCACCGCTCCCCCGACGCCCTCACCGTCCCGGCCAGCCGCGC
CCCCGGCTGCCCTTCGACCCGCGCCCGACGTCACCGAGGCGGCCCGCACCGAACCG
GTCACCCGGGCCACCCTCTGGGACGGCTCCTCCTGCTGGCTGGTGACGCGCCATCAGGA
CGTCCGCGCGGTCTCGGCGACCCGCGCTTCAGCGCCGACGCCACCGCACCGGCTTC
CCCTTCTGACCGCCGGCGGCCGCGAGATCATCGGCACCAACCCGACCTTCTGCGCAT
GGACGACCCGGAGCACGCCCGACTGCGCCGGATGCTCACCGCCGACTTCATCGTCAAGA
AGGTCGAGGCGATGCGCCCCGAGGTGCAGCGCCTCGCCGACGACCTGGTGCACCGGAT
GACCACCGGACGCACCTCCGCCGACCTGGTACCGAGTTCGCGCTGCCGCTGCCGTCCC
```

TGGTGATCTGCCTGCTGCTCGGCGTCCCCTACGAGGACCACGCGTTCTTCCAGGAGCGCA
GCCGGGTCTCTGCTCACCTGCGGTCCACTCCCAGGAAGTCCGGGCCGCCAGGACGA
GTTGCTGGAGTACCTCGCCCGGCTCGCCCGGACCAAGCGGGAGCGGCCGGACGACGCC
ATCATCAGCCGCCTGGTCGCCCGCGGCGAGCTCGACGACACCCAGATCGCCACCATGGG
ACGCCTGTTGCTGGTCGCCGGCCACGAGACGACCGCCAACATGACCGCGCTCTCCACCC
TCGTGCTGCTGCGCAACCCCGACCAACTCGCCCGGCTGCGCGCCGAACCCGCGCTCGTC
AAGGGCGCCGTCGAGGAGCTGCTGCGCTACCTGACGATCGTGACAACGGCGTTCCCCG
GATCGCCACCGAGGACGTGCTCATCGGCGGCCGACCATCGCCGCCGGCGAGGGCGTC
CTGTGCATGATCAGCTCCGCCAACCGGGACGCCGAGGTGTTCCCCGGCGGGCAGCAGCT
CGACGTGGCCCGCGACGCCCGCCGACGTGGCCTTCGGCTTCGGCGTCCACCAAGTGC
CTGGGACAGCCGTTGGCCAGGGTGGAGCTCCAGATCGCCATCGAAACGCTGCTGCGCCG
CCTGCCGGACCTGCGGCTGGCCGTGCCCCACGAGGAGATCCCGTTCGCGGGCGACATG
GCGATCTACGGGGTCCACTCGCTGCCGATCGCCTGGTAG

>fd1

ATGACCATGCGGGTGAGTGCGGATCGGACGGTCTGCGTCGGTGCCGGGCTGTGTGCGCT
GACGGCGCCGGGCGTCTTCGACCAGGACGACGACGGGATCGTCACGGTGCTGACGGCC
GAACCCGCCGCCGACGACCGGCGCACCGCGCGCGAGGCCGGCCATCTCTGTCCGT
CCGGTGCGGTCCGCGTCTCGAGGACACGGAATAG

>fd2

ATGCGCATCCACGTGACCAAGACAAGTGCTGCGGCGCCGGCAGTTGCGTCCTCGCCGC
GCCCGACGTCTTCGACCAGCGGGAGGAGGACGGCATCGTGGTCCTCCTCGACACCGCGC
CGCCCGCCGCGCTGCACGACGCGGTCCGTGAGGCGGCGACCATCTGCCCGCCGCCGC
GATCACGGTGACCGACTGA

10.1.2 *cyp105ab1* from *S. erythraea* NRRL2338.

>cyp105ab1

ATGGCGTCTTCAGAAGCCCTGAGCTATCCCATGCCGCGCACCTGCCCGTACTCGCCACCG
CGGGAGTACGAGCGGTTGCGCTCGCAGGAGCCGGTCAAGCGGGTCCGCACGATCGGCC
GCGGCACCGCCTGGCTGGTGACCCGCCACGAGGACGTCCGGCGCGTCCTCTCCGACCC
GCGGATGAGCTCCGACCGGACCATGCCCGGATTTCCGAGCCTGGTCCCCGGCAGGCGG
GCCATCGTGGCCGAGAACAAGCAGGCGATGATCGGGATGGACGGCCAGGAGCACGCGG
AAGCGCGCCGCGCCGTTCATCGGCGAGTTCACCGTGCGCCGCATCAACCGGATGCGGCCG
CGCATCCAGGAGATCGTCGACGAGTGCGTCGACCGGATGCTCGCCGCCGGCGGGCCCCG

TGGACCTGGTCCGGGAGCTCTCGTTACCGGTTCCGTCGTTGGTGATCTGCGAGCTGCTCG
 GCGTGCCCTACTCCGACCACGACTTCTTCCAGAGCCGCAGCGCGCTGATGATCAGCCGCT
 CCACGCCGCCCGAGCGCAGGCGCGACGTGGTGCTCGAGCTGCGGCCGTACCTCGACGA
 GCTGGTGGCCGAGAAGGTGCGCGAGCCAGCCGACGACCTGCTCGGCCGTACGGTCCGCC
 CAGCAGTCCGAGAAGGGCGAGGTGGACCGCGAGGGCCTGGTGAGCCTGGCCTTCTGCT
 GCTGATCGCCGGGCACGAGACCACCGCAACATGATCTCGCTCGGTACCCTGGCGCTGC
 TGGACAACCCGGACCAGCTCGCGCGCATACCGAGGACCCGGCGCGGACTCCGGCCGC
 GGTGGAGGAGCTGCTGCGCTACTTCTCGATCGTCGACGGCGCCACCTCCCGGACTGCGC
 TGGCCGACATCGAGATCGGGGGAGTGCTGATCAGGGAGGGCGAAGGCCGTGGTCGCCGT
 CGGCTTGTGCGCCAACCGCGACCCCGAGGCCTTCGACTCGCCCGACGAGCTGGACCTTG
 ACCGCCAGGCGCGCAACCACGTGGCGTTCCGGGTTCCGGCGCGCACCAGTGCCTCGGCCA
 GAACCTCGCGCGGGTGGAGCTGCAGATCGTCTTCGACACGCTGTTCCGCCGCATCCCGG
 GGCTGCGCCTGGCCGACGGCCTGGACGGGATCAGGTTCAAGGACGACGCCCTGGTCTAC
 GGGGCGCACGAGATGTCGGTGACGTGGTGA

10.1.3 *fd_sery* from *S.erythraea* NRRL2338.

The *fd_sery* is the only ferredoxin gene found in the *S.erythraea* genome to be naturally clustered with a cytochrome P450 gene.

> *fd_sery*

ATGCGCGTTCGAGGTGGACCGCGAACGTTGTTGCGGGGCGGGGATGTGCGCTCTGACCGC
 CCCCATGTGTTTCGACCAGGACGGCGGGGACGGCCGGGTGCTCCTGCTGGACGCCGCG
 CCGCCCGCCGGTGTAGCACGACGCGGCCCGGGAGGCCGCGCAGCTGTGCCCCGCCGGG
 GCGATCACCGTGCTGGAGTGA

10.2 Discovery of CYP105AB1

Table B1. The sequential similarity between the candidate P450 at Hypha Discovery and its closest homologous P450s identified in five actinomycetes.

Most similar CYPs	Organism	Identity (%)	Coverage (%)
CYP105AB1	<i>Saccharopolyspora erythraea</i> NRRL2338	60	98
CYP105D7	<i>Streptomyces avermitilis</i>	51	97
CYP105D5	<i>Streptomyces coelicolor</i>	47	97
CYP105D4	<i>Streptomyces lividans</i>	47	97
CYP105A1	<i>Streptomyces griseolus</i>	47	97

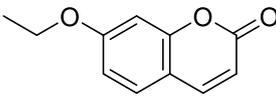
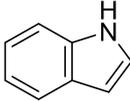
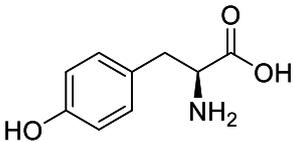
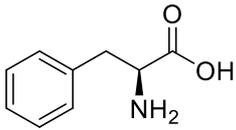
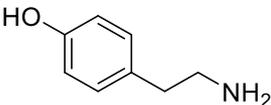
10.3 Preliminary substrate screening for CYP105AB1

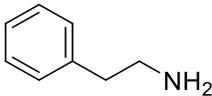
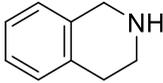
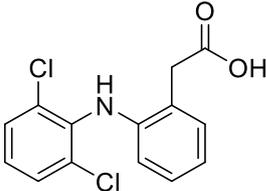
After applying the *cyp-fd2-fdr* operon for CYP production in *E.coli*, there were 6 CYPs in total being successfully expressed and purified. Each of the CYPs was subjected to a small panel of substrates for potential activity screening. The Pdx-PdR system was used for each enzyme assay for electron transfer. The assay setup and conducting methods were described in detail in section 2.2.1.

10.3.1 Substrates and their standard curves.

As is shown in Table B2, the listed substrate standards were firstly characterised on HPLC for identification of retention time and peak area.

Table B2. Substrates list for preliminary screening. All substrates except PHA and PHEA could be detected at 280 nm using HPLC method 1. The peaks for PHA and PHEA were detected at 254 nm. The linear relationship between signal peak areas and the concentrations of substrates were calculated by Origin and recorded in Figure B2 to Figure B8.

Abbreviation	Name	Structure	MW (g/mol)	Retention time (min)
EC	7-Ethoxycoumarin		190.20	10.40
IND	Indole		117.15	10.30
TYR	Tyrosine		181.19	2.86
PHA	Phenylalanine		165.19	4.88
TYA	Tyramine		137.18	3.00

Abbreviation	Name	Structure	MW (g/mol)	Retention time (min)
PHEA	Phenylethylamine		121.18	5.48
THIQ	1,2,3,4-Tetrahydroisoquinoline		133.19	4.29
DIC	Diclofenac		318.13	10.31

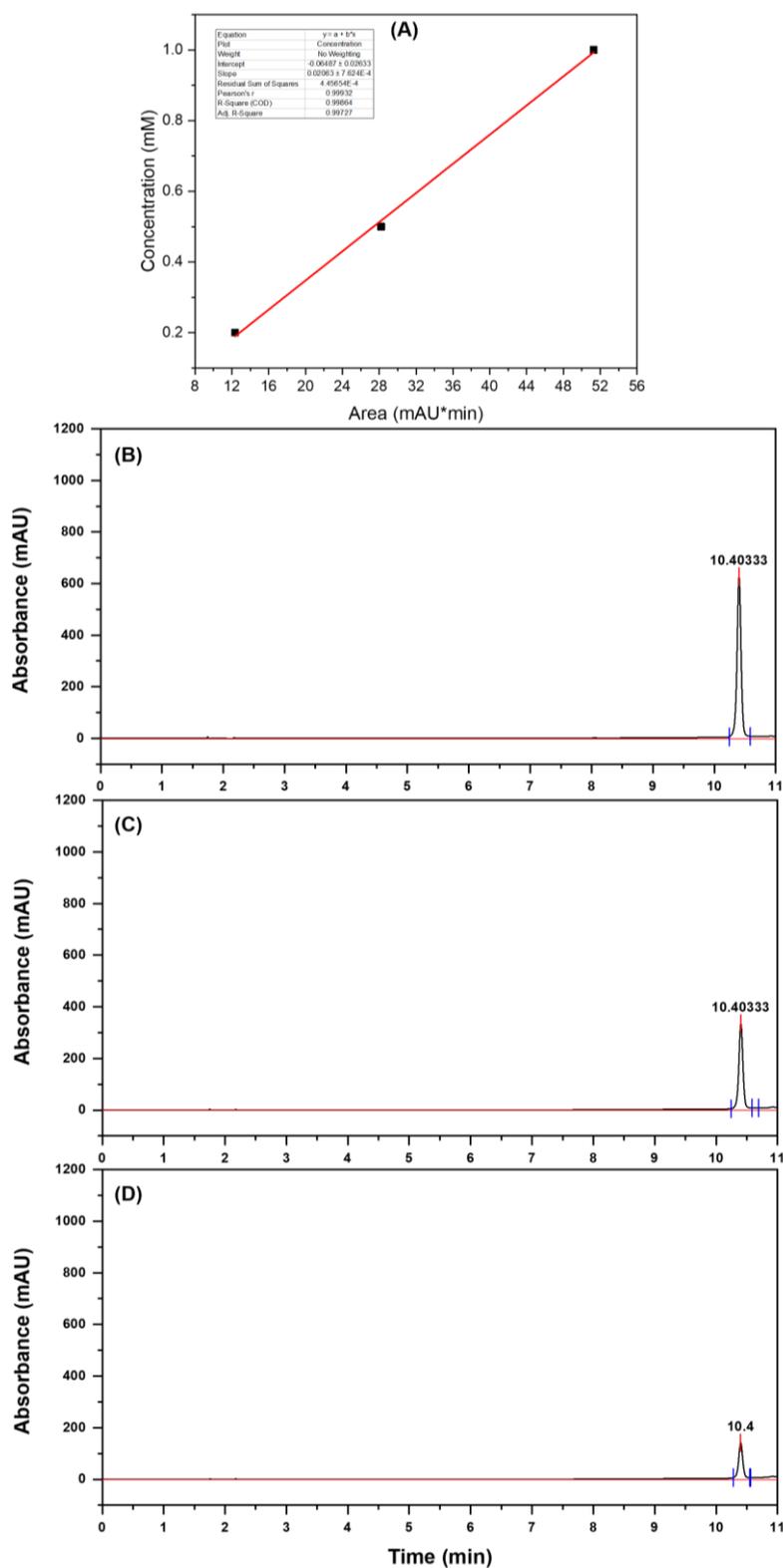


Figure B1. The calculated standard curve of 7-Ethoxycoumarin (EC). The retention time for EC was 10.4 min. (A) The calculated linear fit; (B) The HPLC chromatogram of 1 mM EC; (C) The HPLC chromatogram of 0.5 mM EC; (D) The HPLC chromatogram of 0.2 mM EC.

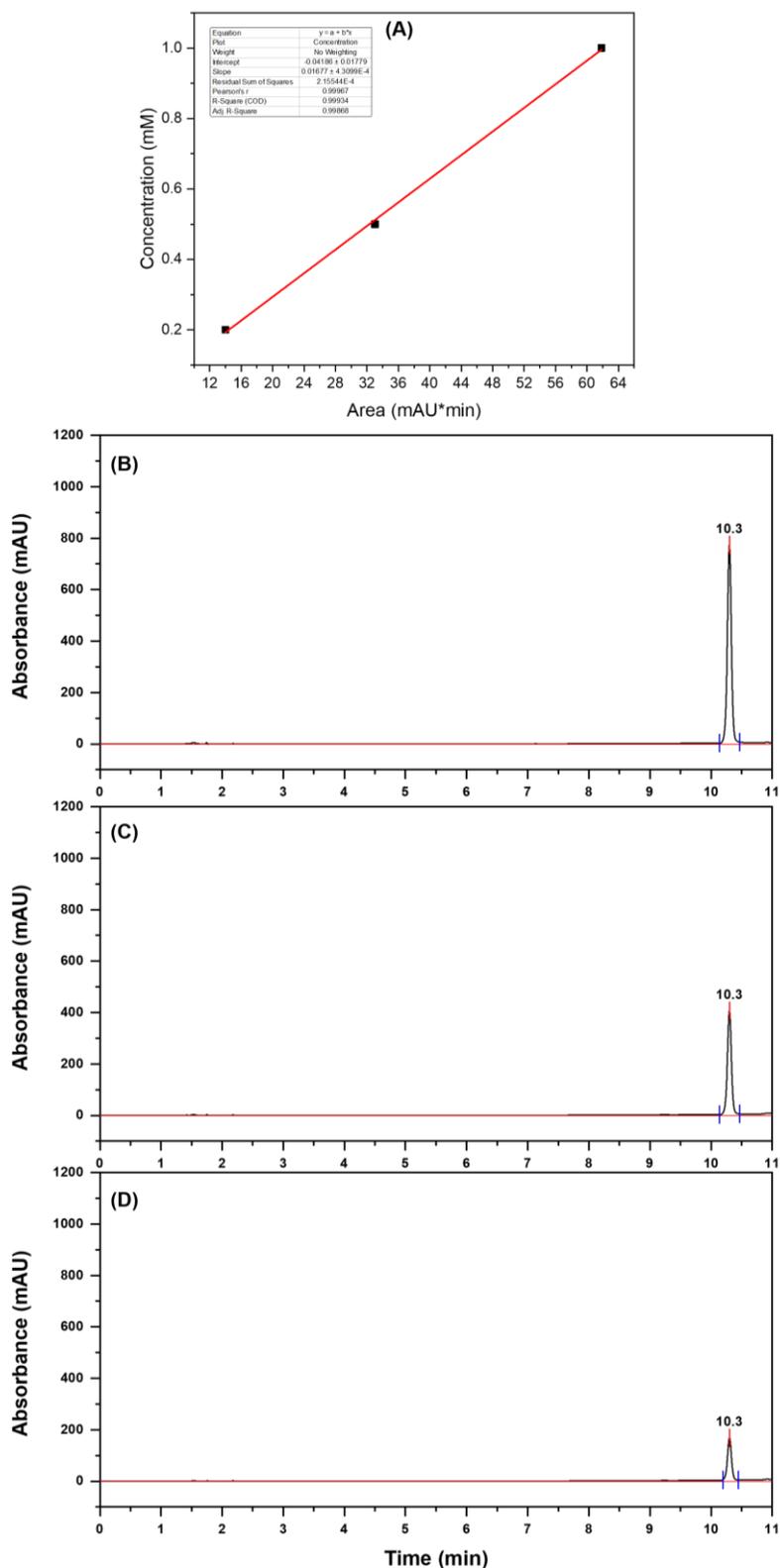


Figure B2. The calculated standard curve of Indole (IND). The retention time for IND was 10.3 min. (A) The calculated linear fit; (B) The HPLC chromatogram of 1 mM IND; (C) The HPLC chromatogram of 0.5 mM IND; (D) The HPLC chromatogram of 0.2 mM IND.

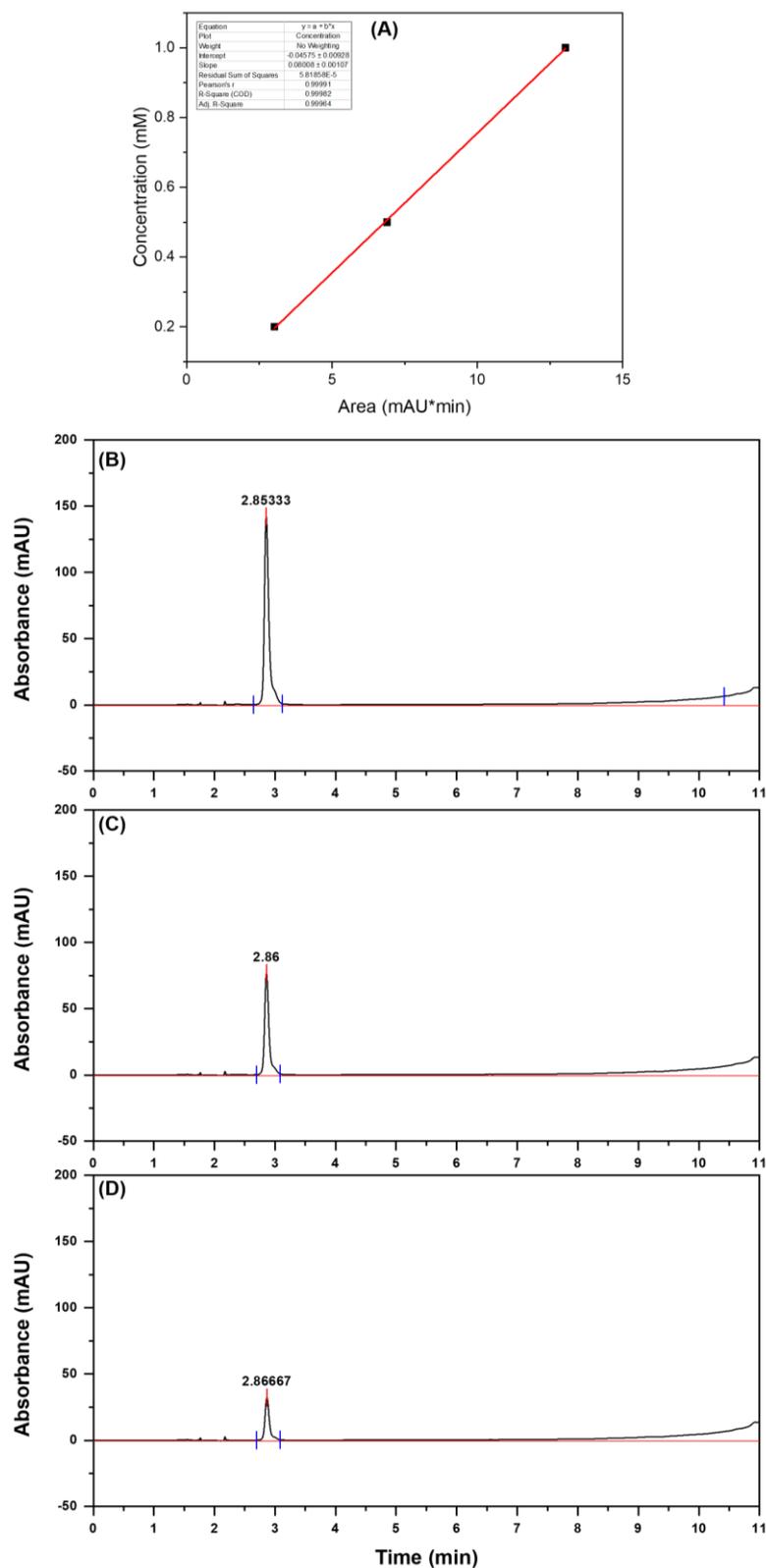


Figure B3. The calculated standard curve of tyrosine (TYR). The retention time for IND was 2.86 min. (A) The calculated linear fit; (B) The HPLC chromatogram of 1 mM TYR; (C) The HPLC chromatogram of 0.5 mM TYR; (D) The HPLC chromatogram of 0.2 mM TYR.

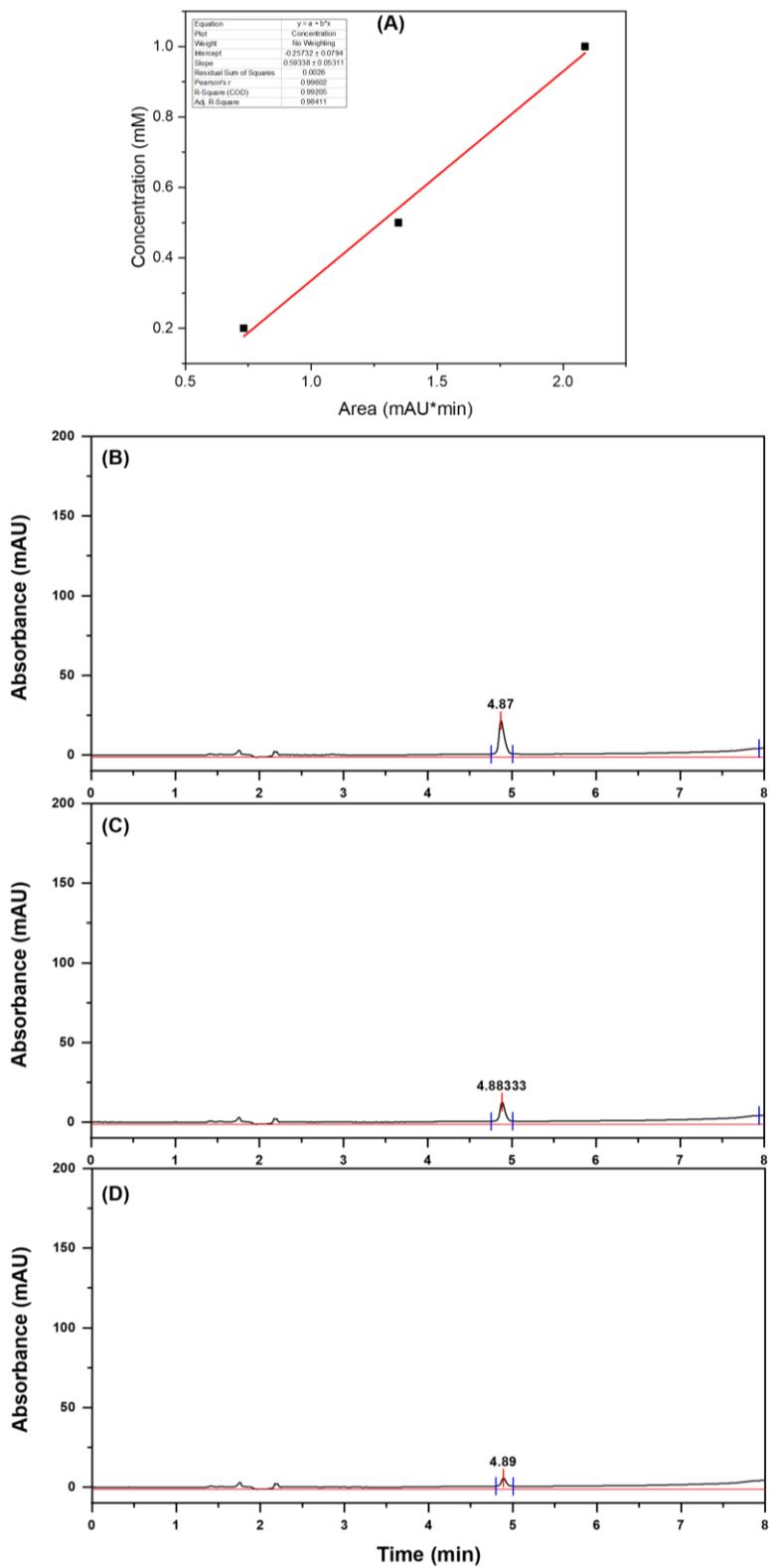


Figure B4. The calculated standard curve of phenylalanine (PHA). The retention time for PHA was 4.89 min. (A) The calculated linear fit; (B) The HPLC chromatogram of 1 mM PHA; (C) The HPLC chromatogram of 0.5 mM PHA; (D) The HPLC chromatogram of 0.2 mM PHA.

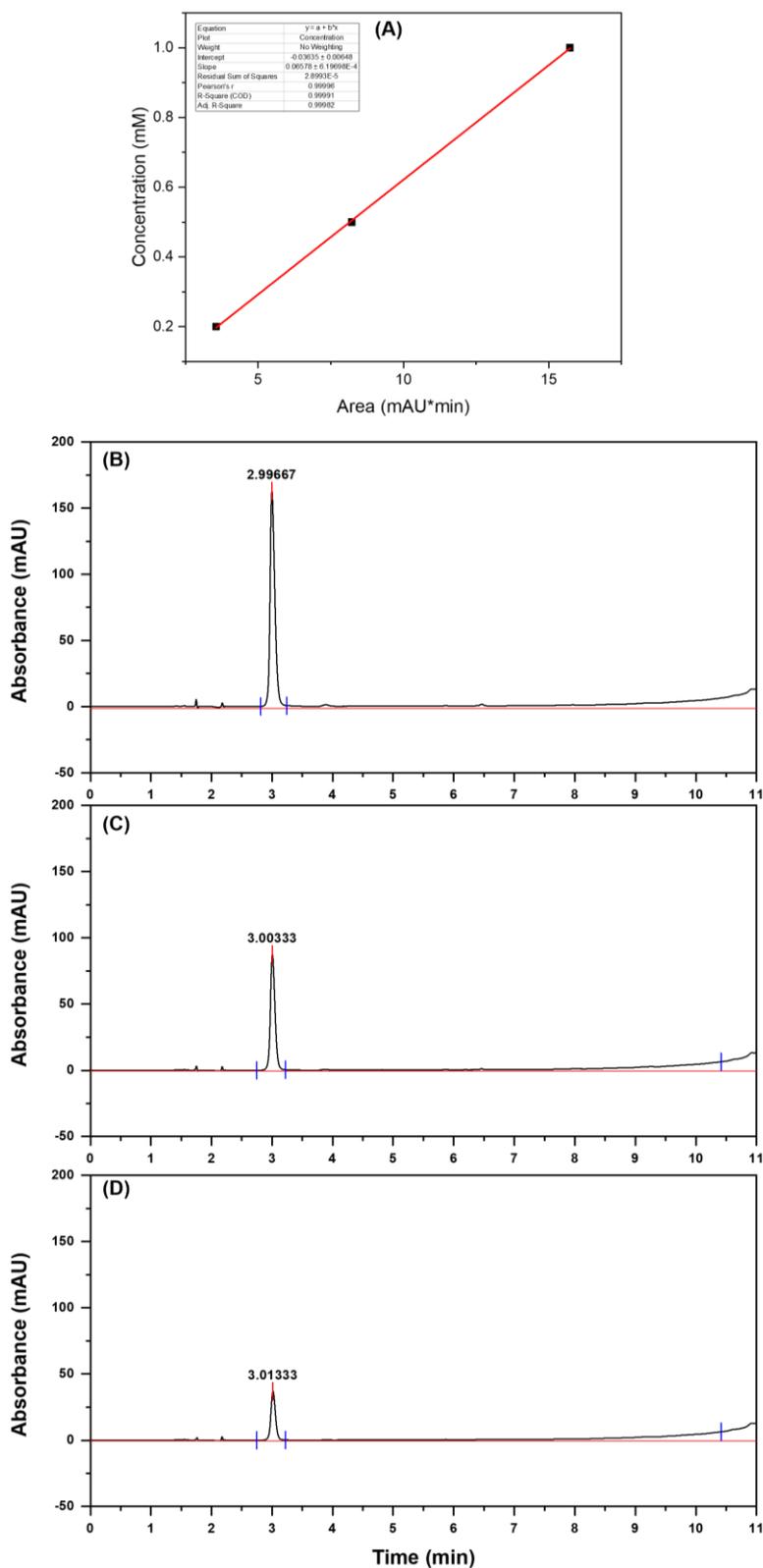


Figure B5. The calculated standard curve of tyramine (TYA). The retention time for TYA was 3.0 min. (A) The calculated linear fit; (B) The HPLC chromatogram of 1 mM TYA; (C) The HPLC chromatogram of 0.5 mM TYA; (D) The HPLC chromatogram of 0.2 mM TYA.

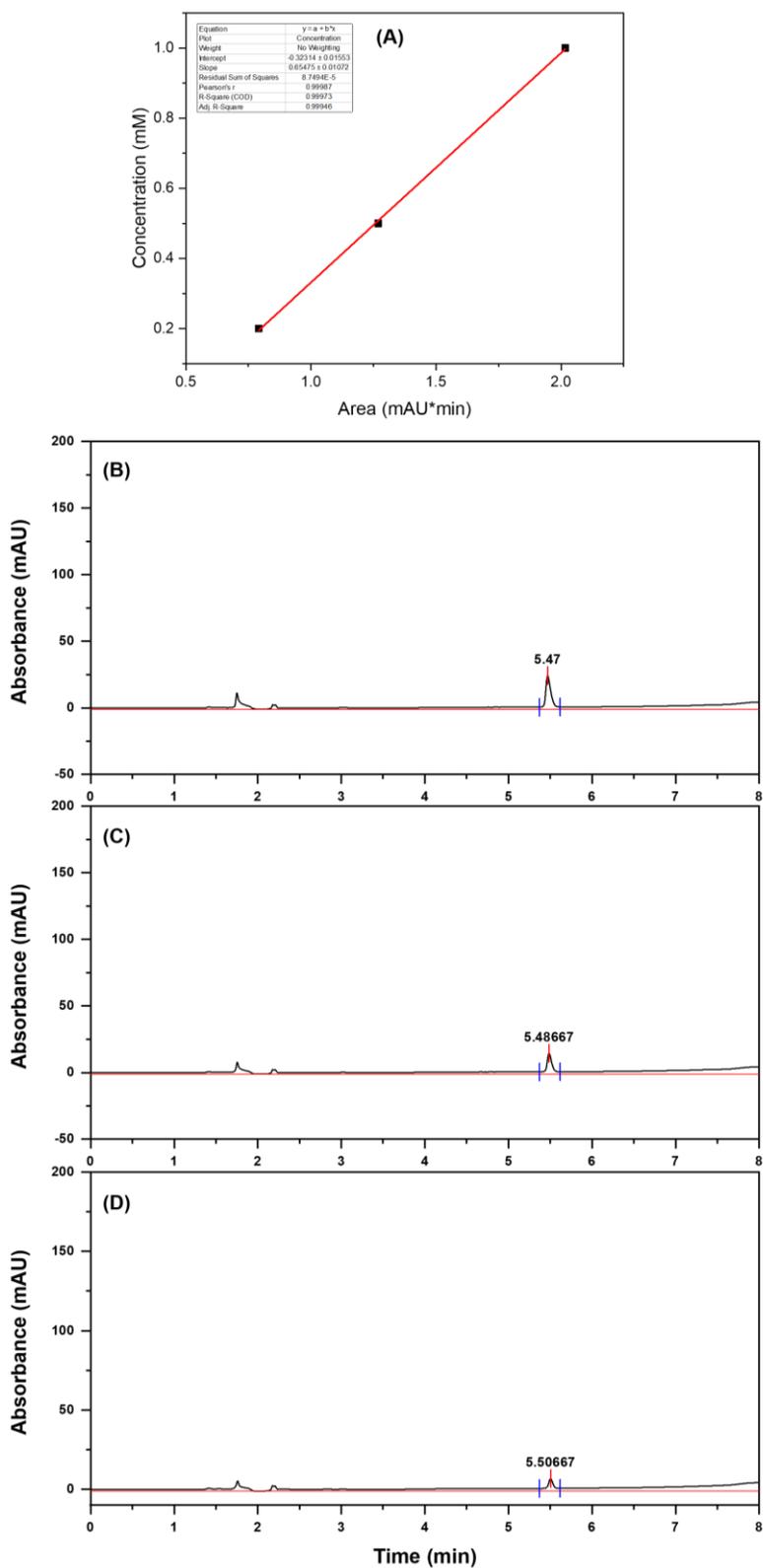


Figure B6. The calculated standard curve of phenylethylamine (PHEA). The retention time for PHEA was 5.48 min. (A) The calculated linear fit; (B) The HPLC chromatogram of 1 mM PHEA; (C) The HPLC chromatogram of 0.5 mM PHEA; (D) The HPLC chromatogram of 0.2 mM PHEA.

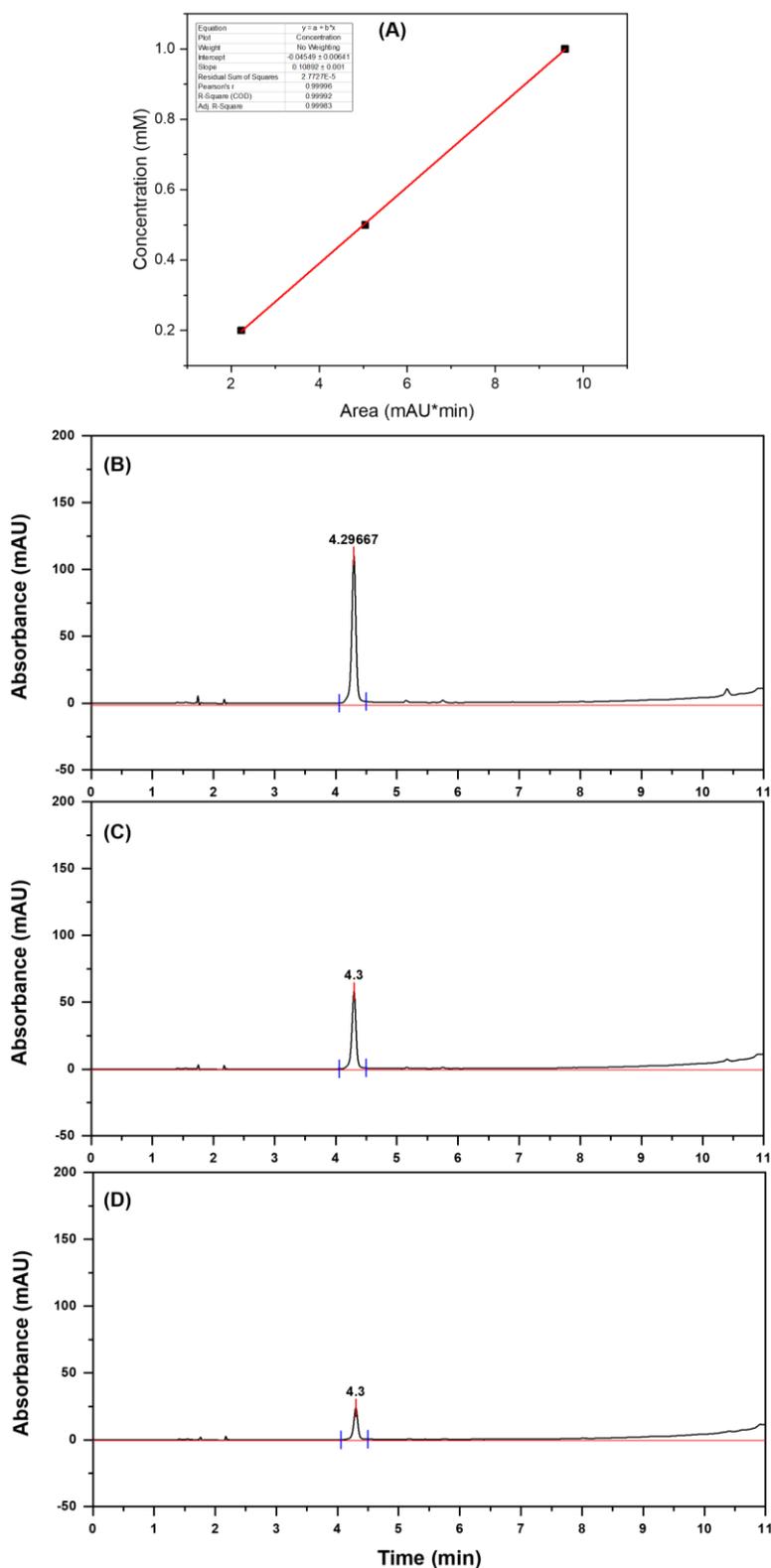


Figure B7. The calculated standard curve of 1,2,3,4- Tetrahydroisoquinoline (THIQ). The retention time for THIQ was 4.3 min. (A) The calculated linear fit; (B) The HPLC chromatogram of 1 mM THIQ; (C) The HPLC chromatogram of 0.5 mM THIQ; (D) The HPLC chromatogram of 0.2 mM THIQ.

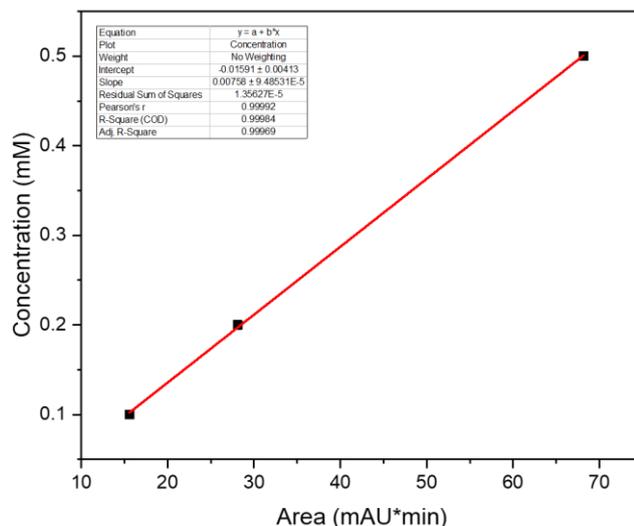
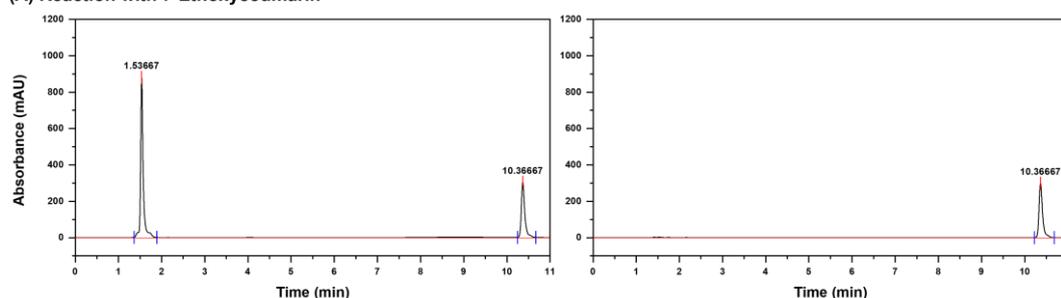


Figure B8. The calculated standard curve of DIC. Three concentrations of diclofenac including 0.5 mM, 0.2 mM and 0.1 mM were used to establish the correlation of concentration and peak area. The revealed correlation was concentration = 0.00758*Area – 0.01591, which was used for various concentration calculations.

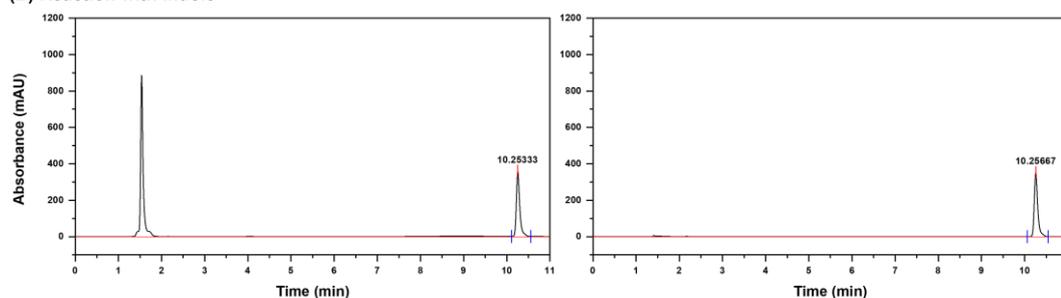
10.3.2 The screening results for CYPs

The positive screening results are shown in Chapter 5, and the negative results of the CYP105AB1 screening are shown below.

(A) Reaction with 7-Ethoxycoumarin



(B) Reaction with Indole



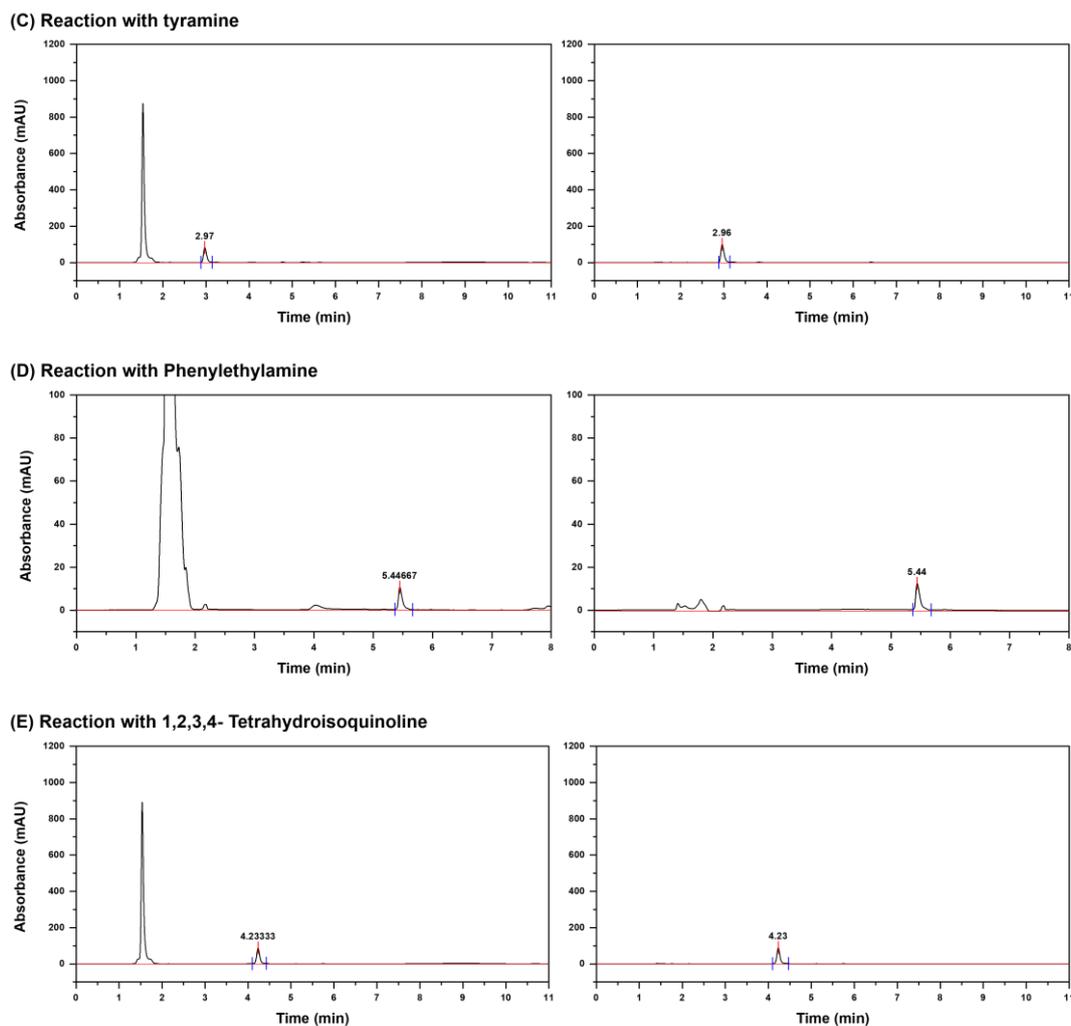


Figure B9. The negative screening results of CYP105AB1. The reaction mixture was analysed through HPLC and the substrates could be detected at 280 nm or 254 nm. (A) Chromatogram of assays with ethoxycoumarin. The retention time of ethoxycoumarin was 10.37 min, and there were no new peaks being identified in the reaction chromatogram. (B) Chromatogram of assays with indole. The retention time of indole was 10.25 min, and there were no new peaks being identified in the reaction chromatogram. (C) Chromatogram of assays with tyramine. The retention time of tyramine was 2.97 min, and there were no new peaks being identified in the reaction chromatogram. (D) Chromatogram of assays with phenylethylamine. The retention time of phenylethylamine was 5.44 min, and the peak area of the substrate did not change. (E) Chromatogram of assays with THIQ. The retention time of THIQ was 4.23 min, and there were no new peaks being identified in the reaction chromatogram.

10.4 NMR spectra of diclofenac and its product

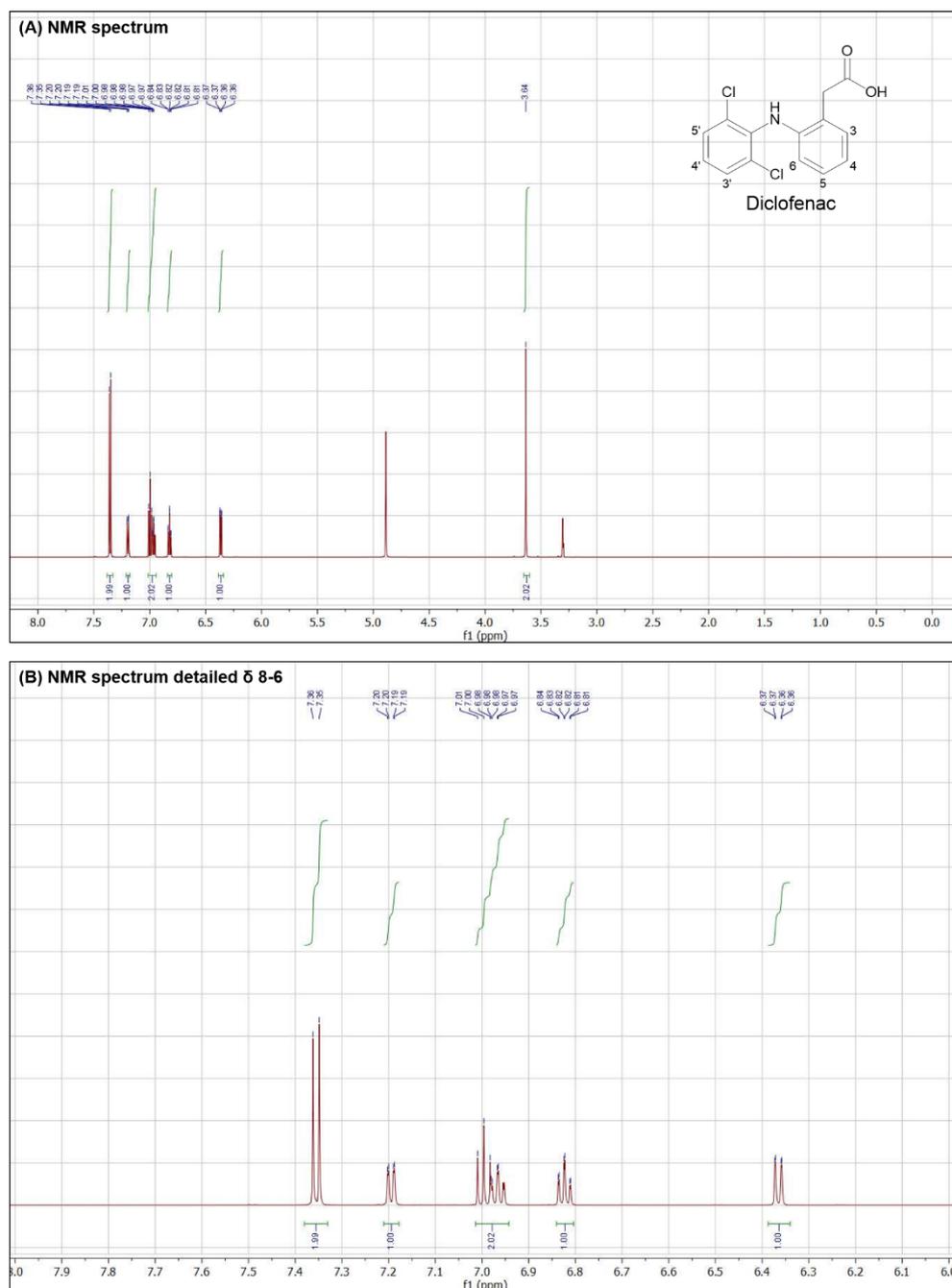


Figure B10. ^1H -NMR spectrum for diclofenac. ^1H NMR (600MHz, CD_3OD) δ 7.36-7.35 (2H, d), 7.20-7.19 (1H, d), 7.01-6.97 (2H, t), 6.84-6.81 (1H, t), 6.37-6.36 (1H, d), 3.64 (1H, s).

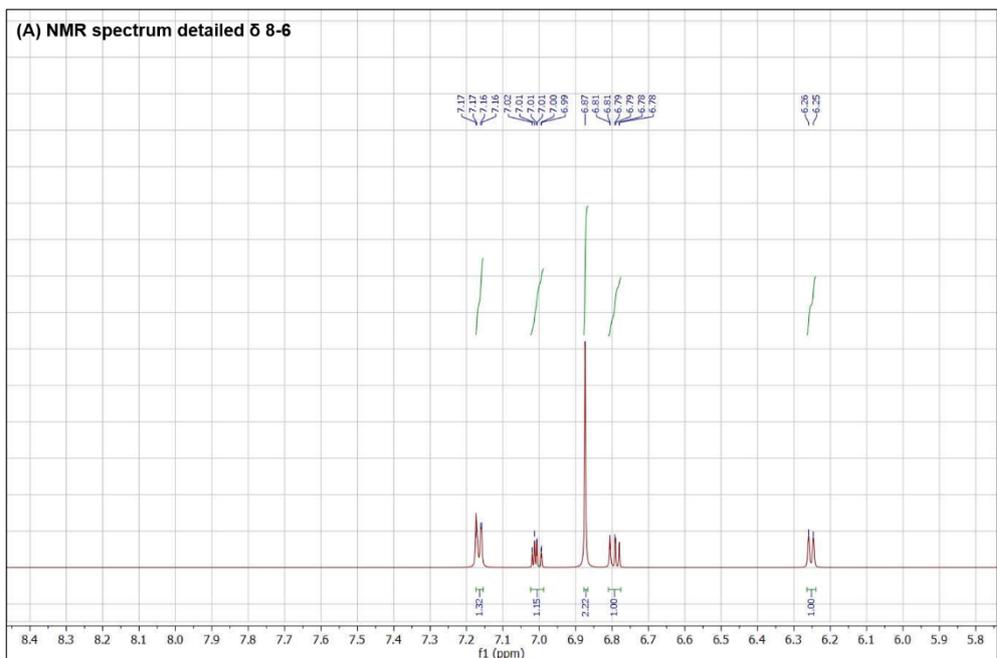
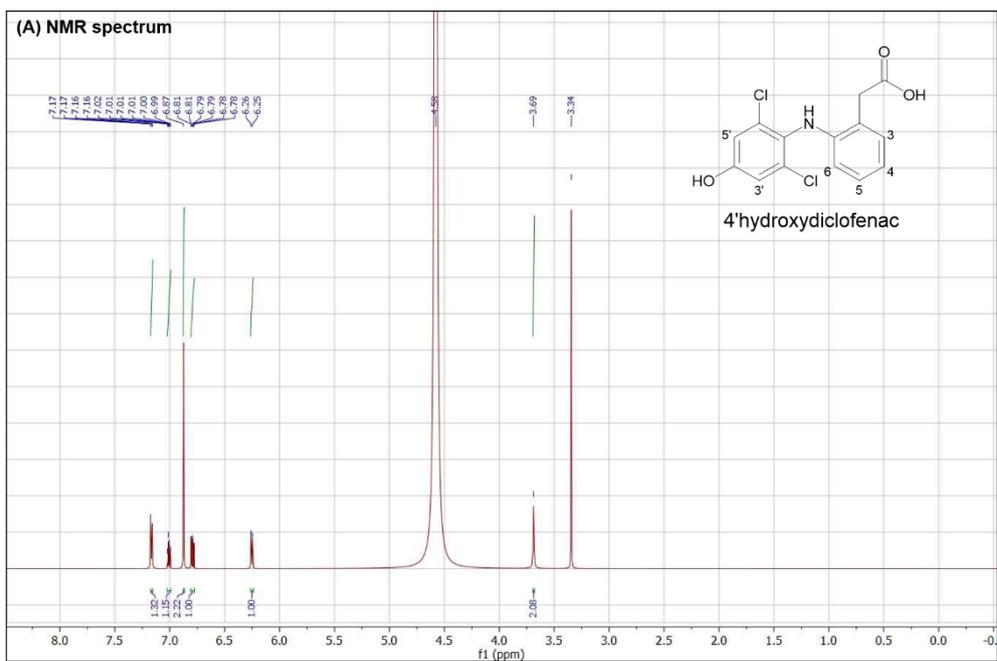


Figure B11. ^1H -NMR spectrum for 4'-hydroxydiclofenac. ^1H NMR (600MHz, CD_3OD) δ 7.17-7.16 (1H, d), 7.01-6.99 (1H, t), 6.87 (2H, s), 6.81-6.78 (1H, t), 6.26-6.25 (1H, d), 3.69 (2H, s).

10.5 Multiple alignment among CYP105 subfamily

The complete multiple alignment among 53 CYPs from the CYP105 subfamily was conducted using CLUSTALW. All CYP105s were selected from the bacterial P450 database achieved by Dr David R. Nelson. These protein sequences are publicly recorded at <http://drnelson.uthsc.edu/BLAST/allbacteria.html>. The identity percentage between CYP105AB1 and each CYP in the subfamily was calculated and displayed as heatmap in the figure below.

Figure B12. The identity matrix of CYPs in the multiple alignment. The identity is calculated as the percentage of identical residues. The darker the shade, the higher identity score there is. The identity of CYP105AB1 to every CYP in the alignment was marked in the red rectangle. P450_MoxA had the highest identity (57.6%) with CYP105AB1.

10.6 Multiple alignment among CYP105AB1, CYP105A1 and P450_MoxA

The multiple alignment among three CYPs was conducted using CLUSTALW. Identical, strongly similar, and weakly similar residues are denoted by asterisks, colons, and dots, respectively. The choices of mutation points were marked with red rectangles. Comparing to the wild type CYP105AB1, the two mutants CYP105AB1-T108A and CYP105AB1-G287D were constructed.

```

CYP105A1      MTDTATTPQTTDAPAFPSNRSCPYQLPDGYAQLRDTFGPLHRVTLYDGRQAWVVTKEEAA 60
CYP105AB1    -----MASSEALSYPMPRTCPYSPPREYERLRSQ-EPVKRVRTIGGGTAWLVTRHEDV 52
P450_MoxA    --MTKNVADELAGLELPEVERGCPFAPPAAYERLRER-APINKVRLTSGGQAWVVSHEEA 57
              . * * ** : * * :** . *:::* . * ** * : ** .

CYP105A1      RKLLGDPRLSSNRDNDNFPATSPRF--EAVRESPQAFIGLDPPEHGTRRRMTISEFTVK 117
CYP105AB1    RRVLSDPRMSSDRTMPGFPSLVPGRRAI--VAENKQAMIGMDGQEHAEARRAVIGEFTVR 110
P450_MoxA    RAVLADGRFSSDKRKDGFPLFTLDAATLQQLRSQPPLMLGMDGAEHSAARRPVIIGEFTVK 117
              * :*. * **:: . ** : . . :*: * ** . ** . * ** : ** :

CYP105A1      RIKGMRPEVEEVVHGFLDEMLAA-GPTADLVSQFALPVPSMVICRLLGVVPYADHEFFQDA 176
CYP105AB1    RINRMRPRIQEIVDECVDRLAA-GGPVDLVRELSLVPVSLVICELLGVPSYSDHDFQSR 169
P450_MoxA    RLAALRPRIQDIVDHFIDDMLATDQRPVDLVQALSPLVPVSLVICELLGVPTDHDFFQSR 177
              * : **:::*. : * ** : . ** : :*****:***.*****:***:***.

CYP105A1      SKRLVQSTDAQS-ALTARNLAGYLDGLITQFQTEPGAGLVGALVADQLANGEIDREELI 235
CYP105AB1    SALMISRSTPPERRRDVVLELRRYLDELVAEKVREPADDLLGRQVAQQSEKGEVDREGLV 229
P450_MoxA    TTMVSRSTSMEDR-RRAFaelRAYIDDLITRKESEPGDDLFSRQIARQRQEGTLDHAGLV 236
              : :. : . . : * * : * ** : . ** . * . . : * * : * : * : * :

CYP105A1      STAMLLLIAGHETTASMTSLSVITLLDHPEQYAALRADRSVPGAVEELLRYLAIAIAG 295
CYP105AB1    SLAFLLLIAGHETTANMISLGTALLDNDPQLARITEDPARTPAAVEELLRYFSIVIGAT 289
P450_MoxA    SLAFLLLTAGHETTANMISLGVVGLLSHPEQLTVVKANPGRTPMAVEELLRYFTIAGVT 296
              * *.*** *****. * **..: **.:*:* : : : . * *****:*. * .

CYP105A1      GRVATADIEVEGHLIRAGEGVIVVNSIANRDGTVYEDPDALDIHRSARHHLAFGFGVHQC 355
CYP105AB1    SRTALADIEIGGVLIREGGVVAVGLSANRDPEAFDSPDELDRQARNHVAFGFGAHQC 349
P450_MoxA    SRLATEDVEIGGVSIKAGEGVIVSMLSANWDPVAFKDPVLDVERGARHHLAFGFGPHQC 356
              . * * * : * : * : *****: ** * . . . * ** : * ** : ***** **

CYP105A1      LGQNLARLELEVILNALMDRVPTLRLAVPVEQLVLRPGTTIQGVNELPVTW 406
CYP105AB1    LGQNLARVELQIVFDTLFRRIPLRLADGLDGIRFKDDALVYGAHEMSVTW 400
P450_MoxA    LGQNLARMEQLQIVFDTLFRRIPLRLAVPMEDVVPFKGDSVIYGVHELPTW 407
              *****:***:~::~:~: *:* ***** : : : : : *.:* : **

```

11 Appendix C

This appendix contains supplementary information for Chapter 6.

11.1 Protein sequences of 24 CYPs identified in *S.rishiriensis*.

>Sri_CYP01

MTDTTETTGPLDPEAEPPVFPQEKAPGCPFDPPPAYRRLQAEDPFARVTLPSGQQARLVTR
HRDVRVAVLDDPRFSADSSHPDFPRMFPRPVPQVLQGTFFPRLDGAEHLRYRRMLARDFTGRR
AEELRPRIEQIVDARLDLMEETGGPLDLMEALAYPVPSTVCELLGVPPEDSPLFESRTRVLING
RSGPQETQRAKDDILHHLEGVVAAKEKQPGDDLISRLLDQVAPGLLDRGEVAVIAWLLLAAGH
HTTATMIGTGAMLLLENPEQLAALRADPALLPGAVDELLRHQTAMQIGMNRIATEDVAIGDETV
RAGEGVVQCQLASANRDADVFPGPDRFDVTRTARGHLAFGHGPHQCPCGQSLARVELQVTLAR
LFTRLPALRLAVPAADVFWHHLFGIHGVRELPVTW

>Sri_CYP02

VTPEHPTPTGPRDLALDPPPGCPAHTRGPGGLARLYGPGAEDLNELYERLREEHGPVAPVLIH
DDLPMWMLVGHAEENLRMVRTPSQFTKDSRIWSQLVQGRVDPDHPLMPHIAWQPICAHAEAGD
EHRRLRGAVTAAMATIDDRSVRRYINRSSQRLVNRFCQQRAELVGDFAEHLPMAVMCHVLG
MPDEYNDKMMVHAARDALKGSETAVASHEYVVGALARLTARRRAEPEEDFTSHLITHPAALTD
EVREHLRLVLFAAYENTVNLLSNVLRMVLTDPRFLARLNGGQMTVAEAVEQSLWDEPPFSTIF
GYFAKQDTELGQRIRRGDGLFFAPAPGNVDPRVDPDLSAHMQGNRSHLAFGSGPHECPGQ
DIGRAIADVGVDALLMRLPDVQLDCAEDDLRWMSIASRHLVALPVRFEPKDQQDVKHKPSHA
PIPPQLTRPVVTAPQSAPAPVHEPAAAPAPQAPAPQPTRPARRPNAWRRFLAWWRGY

>Sri_CYP03

MAAVDDLSDPWPDPFVSDPYPAYAELRERGRVIYFEPTRQWLVPRHADVSAALLDRRLGRT
YQHRFTHEDFGRTAPPAEQEPFHTLNDHGMLDLEPPDHTRIRRLVSKAFTPRTVEGLKPYVRR
LAAELVDGLVREGGDLRLDVAEPLPVAVIAEMLGIPEADRAPLRPWSADICGMYELNPSQETA
AKAVRASVEFSDYLRELIAERRKEPGEDLMSGLIAAHDEGDRLTEQELISTAVLLL NAGHEATVN
ATVNGWWTLFQHPEQLAALRADHSLAPSAVEELMRYDTPLQLFERWVLDDIEIDGTVVPRGAE
IAMLFGSANHDPVFAEPDRLLDARADNPHISFSAGIHYCIGAPLARIELAASMTALLERAPTLAL
AAEPRRKPNFVIRGLEGLSVTL

>Sri_CYP04

MESQADDGGSVDLDGATAERLETLTLEPLLTRDYETRPALVYERLRRRHGPVAPVDLLGVPA
WLVVLYREALDVLQDDAGWPKGLENWRARTDGEVPADWPLGPSLEVNHVLIQGGPGYRPLR
QAWDAALKPFQDPRHAQAKRLKAAVTAYADELISLVGQAGGTGLADLSAQFSRPLPLMVASHL
LGFPQSQGDALMDMWRVLDAGPDAEPALERLLATLAELA AVKLEKPGDDFPSYLLAAHPELS
LDELARELFMLLGMTSDHVGILISNTVVEVLSGEGGVRASLSAGMVRESMNRVVMRKPLVNF
VPRFAAKDTPLGNYTIRAGDPVWVSA AAAHADPLFAGNVCPSTTVSTRAHLSWGAGLRQCPA

RELASTVAAVGVGRLFERFSHLELALPADQLPWRSSPFMRGLRSLPVRYELAATPAQPSAPAR
TAESDATVLPDPAVRQRSSLWRYLTGLIGVGR

>Sri_CYP05

VTVTDRIALDPFGADVHAESARLRALGPIVPVELPGGIPAWAPTGYDTLKALILDPQVSKDPRRH
WRHWPELAEHPSWGWILGWVGVINMLSTYGADHARLRRLVAPSFTHRRTALRERVEAITTE
LLDALARDADGGSAVDLKAAFAHPLPMRIICELFGVPEELRPATGRLIAAIMDTSDPSTAEHAAFV
QEQIGTVLGSLIAHKSGQPDDLTTELIRVRDEEGDRLSDEELLYTLLLVIGAGFETTIVNLVGN
AVALLRRPEQLAAVRRGEIGWDAVVDETLRLYPSIATLPLRFVAVSDLKVGDTVIPAGDAITTYAA
ANADPAHYGPDAEEFDAARGAEDHLAFGIGVHRCIGAPLARMEALTALPALFERFPGLRLDTAA
GELQQVPSFIACGWREIPVRLRG

>Sri_CYP06

VNACPADRRRAFTRTAPVRLWEDGFASDPHRYAALRSQGCVGWAEAPGVVYVVTERRAA
LDVLHDTETFSHDPRWEATVPADSPVLGMMRWRPNALFADGAEHLRYRTTLIDVFDLVEPH
DLRGRVHRAVRLLAGRIGPRGEADLVADFARPVIALVNLDFGLPDSQGDRLNEGLGKMMEG
GAQAAEGEAQYQYVLELIAAKAARRGDDLPSHLLDHPAGLTREEVTWQVFLTLGAGYEPTAN
LLSNTLSRILGNPAYYSTLTSGARPVMDAVVEVLHHTPLANYGVYVRRPVSFHGVWLREAV
PVVVSYGALGRFAEREDVAGRHPADASHLSWGAGAHACPAKQHTLLLVTEAIEQLTQWLPLD
DPVLPRELRWRPFPFHRSLSALPVRFSPRSPATASAASSPSQGARP

>Sri_CYP07

MPGAPGDPLVPMHCPRVSEPLGPPFLSRLGTGTPVWVRRHADVRQMLADPRLNRAHLFAA
DAPRLTPYPNLLDNPDILNLDGIEHQRLRRTVSRTFTPRALARWRPWSSVVDLIDRLVESD
RPVDLIGGFARSLPVAVFSRLMGLDGLDNKRLAYWGDHAFAAAAFPSSHQVQEALTDFAIFGTE
LVAQRRADPGEDLVSTLVQAADDEDDEVTEQQIVSLVILLVMAGHEVSTTVFGNTLVYLLTDGRA
TWRELGSDESKAPAAVDQLLRGIPISDRDVLPGFLRAVEDVEIGGVVIPAGSVVAADTMSANL
DPEVYPEDWRVETFSPLDTPHLAFGAGPHYCLGAWLARLEMELALHRLPKRLPGLRLAVPLEA
IEWREGLLTRSPQNLPTW

>Sri_CYP08

MDRPTAPVPPPQCPAHQAGGRVPLHGPDAADPQAYYSYLRHYGPTAPVELAPGVEATLVTD
YATALQLLQDSGSFRKDARRWRDLNEGRVPPDSPVAPLLAYRPNAMFSDGADHLRLRQAITD
SMARVDSRRLSRSTAQVSDYLISQFGSRGSADLLGDYAKQLPLFVFNELFGCTADIGDRVVF
TGMFDGVNAEQAVGVMFGAIGELVALKRTRPGDDVPSWLMHQSKLTDEELVHQIALLLVAG
NEPLRNLIGHTLHRLTHDAYAHQGGLIDEAIDDTLWENPPMQNLAPHYPTSDMEFAGQKLQA
GDLVLVSFAAANTGPALTAARQAGSNRAHLAWSAGPHACPSKDPARQITVTAIENLLNKLDPV
ELAVPTESLTWRPFPFNRLTVLPAKFTPVRTAPRPAPETRAEPSNREQGPSRGRGDRGGV
WSQFLNWLTR

>Sri_CYP09

MPPTPGSPSVERTWRIGAAPGAVPLLGHVPALWRRPLEFLASLPAHGDLVEVRLGPTRAYLAA
HPELVRHVLLNPRVFDKGGVFDKARQLLGNLSVSRGEEHRFQRRLIQPAFHHTTKIAAYTTAVA
ADTRTVTDGWRAGEVRDISDAMHALLMRVAARTLFTSTGLDEDTVEEARQCLRTVSHGIYRRTV
APLGVLEKLPPTGNRAYDHANTRLRQIVADMIALRRRSADADHGDLLSTLLRAEHPETGQKLLDDA
QILDQVVTFVLVAGSETTASTLAFVFLMGLTPEVEQRVHAEVDAALAGRTPVYEDLPTLPYTRN
VITEALRLYPPSWMAMRVAEDELGGRTIPARTMILYSAYALHHNPELFPDPETFDPRWEG
ERAAHVPRGALLPFGAGSHKICIGDVLALTETALIVATVAARWRLRPAPGSRLRPEPKATLEAGP
LPMVLERR

>Sri_CYP10

MEPPGNEPLFNARDPRLRTPDYPLYKRLRDAEPVHRTPFQHWVVSQYDEVAAILRKPELSSEY
PQDPKWAHRGGQGSPIVQATRSWMLMRDGEAHRRLRGLANPVFTARAIRETAPRMTGIVH
RIMDEMGDGRDIDLIGSLAALVPVTVFCGLGGLPVEDRDLCKRKTWDALSHVIDPAIDTTRRAM
NDAAEEFGEYIAEHLARRRSDPQDDIISRLLLADFDGGKLSDDDEIIANVMMFFMGGHETTIVNLIG
NGMLALLRHPDQLLALREEPGLVDNAVDELLRYDAPVQLVARITTDGIEVGGTTIPAGAKVMVL
LGAANRDPARRYADPDRDLRRADIRPLSFGGGPHYCIGALLAKAEARLVFTELLRRHRDIRQAG
EDVVWHAKVNIRGLRELVDLIA

>Sri_CYP11

MKTQSDPVRRRWRIAKAPGRLPLAGHALHLARRPLEYLSSLPALGDLVELRLGTRLTYLPCHP
ELVQEVLLNARVYDSGGPVKEKARPILGNGLITSGWADHRRQRRMVQPAFHHTTRIATYAEVMR
EECAAHADAWRPGRPVDVGDAMQALTARVTARALFATEMAPRSVAEVQRSLPVMVRGAFRR
ALDPTGLIARLPLAANREFDAALDRLHTLIDGIIGDYRRSGEDRGDLLSALLAARDDDNGGALTD
REVHDQVMTLLLAGVETTASALTWAWHLLAAHPEVEARLHAEVDEVLGGRTPYADVPRLAHT
QRIFTETLRLYPYAWMYTRMTTQPTELAGHPLPAGTDILISPYVIHRIPEFFPRPETFDPRWLP
ERAKDVTRGSYLPFGAGSRKCIGDVFQGMTEATLALAALASRWRLRPVPGATVSPRAEMSLSA
GPLPMVPEPR

>Sri_CYP12

VTGFEDLTALGRPFVDDPYPVYERMRAQGPVHRVRTAASGEFWLVVGHQAQARSALADARLSN
DIRHSAGWQDDGGNSIGLNMVQTDPPQHTRLRGLVAKEFTSRRIEALRPKIQRIDEELLDAMP
LGRADLVESFALPLPLAVICELLGVPANDRKTFFHDWYLESTDFTRPEVAAAAAQAMTGYLAELV
AAKREPGEDLLSALAGRLGEEEDGLTPEEMLGMAFVLLVAGYETAANLLSSGTLALLRHPDQ
LAALRGDWSLLENAVEEMLRYEYEGPVETTAFRYAKEPLEIDGTVIGAGDSVAVVLAASRDPKRF
AEPDRFDIRRDTRGHTAFGYGIHHCLGAPLARMQGAIAFRALLERCPLALDTPADLRWRPS
LMLRGLHALPVTFG

>Sri_CYP13

MPCPALPDGFDFTDPDLLQHHVPLPEFAALRQAEPVRWIPQSGNVAGFQDEGYWAVTRHADV
RYVSTHPEIFSSYLNTAIRFNEHIERASIDAQRFILLNMDPPEHTRVRQIVQRGFTPRAIRGLEER

LRARATAIVARAREQGGGSFDFVTSVASELPLQAI AELIGVPQDDRAKIFDWSNKMIAYDDPEY
AITAEVGGQSSAAEILAYAMDMAAERKKCPAQDIVSTLVA AENEGNLNSDEFGFFVLM LAVAGNE
TTRNATTHGMHAF LTHPEQWELYKRERPATTAEEIVRWATPVVAFQRTATQDTELG GKRIKKG
DRVGIFYASANRDPEVFDRPDAFDVTRDPNPHLGF GGGGPHFCLGKSLAVLEINLIFNALADAM
PDLRLVEDPRRLRS AWINGVKELRVDAG

>Sri_CYP14

MKTRPDPPVIPSERGRCPFDPPAEYARLR AEQPVSPVTFEVAPKDTSGWL VTRHDLVRQILAD
DRFSHRNELLAHV VAPPFPMTEYAPQPSPPGSFAKMDAPDHTKYR RLLAGHFTLRRIQQYEPQ
LTRIVDGALAE MA AEGSPVDLVTGFAEKVSLRSVCSLMDV SPELMDGIQEHFSALMTLYTLEE
FIHHVESMDGLIR P MVRERTAEQGEDFFGRLAATGELTEDEL VNLAVLTLGGSLDTPNMLSLS
VFALLENPDQLALL RERPELYEAGAVDELLRYLTISQM GSSRCALEDVEIGDTVIKAGQTVVLSL
PAANRDPEVFTDP DRLDVTRIPRKH LALGFGAHQCLGQH LARASLRIGLRALFERFPTLR LAAP
AEEVPLRDRAVHYG VDRLLVSW S

>Sri_CYP15

VAGRSDSGADERT RPARPSYDICERGIRVAQQG TSDVLDSTPGGGFR LERAPTVEPDGGVTL
LAWAARMREEQP VWRDGS GNHVF RHADVQRILADTATFSSD TVGRLSGGEKQAPRG TLLLL
DPPLHGKMRRLV SRAFTPGLITGLEPWIAELTGQLL DAAPGDAFDLVDVLANPLPVTVIARMLG
VPTEDRDRFQGWAD QLLSTDPDDPESVRLMEETAVTISAYL QGFIEARRKEPTGDLLSTLAGA
ELDGERLEDEDIAS FATLLLLAGHITTSVLVGNALLCLDRDPELLERVRADRSLIPAVIEETLR LRP
PFTRIERVTTTATT VAGTDLGPDTLVHLWLLSANRDERVFEDADAFRPDRSNARQA AFGHGIH
YCIGAPLARLEGR IALDALLDRYARIGV DPAVRLSWHG TNVFGARHLPLRVLHA

>Sri_CYP16

VISRQLTGEPPLLEALPERWRDLHRDGPVRYDPGQSVWQVLDLETA AAVLADPVTYSSDMTG
LAPSQPDFDTFTQGNFVGM DPPEHRKLRALVSQAFTPRV VQGLAPRIEAVCARLLDSVAGHDR
FDLVDSLAYPLPII V IAE LLGIPAE EHRLFQEWASVLFGGDQLGEAPDMADLERALEIAIPTVRE
MNGYMLEHIRARRA APGDDLTSRLIAAEVDGVRLRDQEMVGFVALLLVAGHITTTALLGN AVVV
LDRHPGADSALRADPGR IPEAVEEILRWLPPFPELGRRVTRPVVLGGHEIPVD TLLMVHLGAAN
RDPARFAAPDTLDL TRHPNPHLTFGHGIHFCFGAPLARLEARIALRMVHERFPTLVVPSYEDVT
FQNP AVIIGVRHLPVEVGRA

>Sri_CYP17

MAAMTDVRDADPRPENAPAAKCPHF PFADRPGVSLDSGYLKDFRDQPLL PVTLTGGRSALLV
TRYADVRTLADPRFS REACTNGTLFARKSSALALATSDAPIHTRRRSAVQSWFTHRKAEEAR
PHIAALAERLLDR IEAAGPSVDLIAEFSALLPYQVIGDMLGLPVSDLDPLRPWMTAMMSAGRFS
AEEVTTAHESMYGYFFDQLADRRRAIAAGEPGDDL TALLTAPEDDRLSDEEIAVLGFGLLMAG
GETTASHLAMCVLQVLRTPGLAESLRQDPSAIPTVVEELLRWVWFAGTGGQPHVTVEDVELA

GTLIRAGQVVIPLTDAANRDEEVFTDPDEFRPGFRDNPHLGFHGRHMCLGAAHARVELQVGL
AAILRRLDGLELAVDETSLTWRDQMFIRGVWTLVVRWRVRGDAA

>Sri_CYP18

MDTTACPYALDVTGRDLAGEAAMLRSQGPVAVRVELPGGVVAWAVVRQQYVERLLTDSRVSK
DARQHWPDFIEGRITEDWPLYPWVANENMLFAYGASHTLRRLVAGAFTARRSEAMRPRVVQ
IAAELVDSLAESEGGQGPVLDLRTGFAELLPMRVICELFGVEQGEDTDALCTALHTVFGTTISGQ
EVEAARLEAFVRLSELVKAKRAEPGDDLTSLLIQARDQDDRLSEEELLGTLFLMIAGGQDTTAAAL
ITNAIGALLLAPEQLAHVREGRATWEDVVAETMRVHTPGAYSPMRFAVEDIDLDGVLIRKGEAV
LVSFAAGGRDPERHGPDAADRFDLLRKDRDSLGFHGHGPHRCMGAPLAAVEAESAFALFARFP
RMELACRPDELLPMPFSLMNSHRALPVLLVPRHR

>Sri_CYP19

VAVAPLTETDMFATSDRRTRYEQYRHLRELGGVVRLRDPDVYAVTRHDEVKAALADHDAFVS
GEGVGFDPGVNELMRGVTLVSDPPEHDALRSVVGAGLRPGPLRARGEDIVQKAGDLVAQVVA
KGEIDGVTELAQALPMLVIPDYLGPERNREHLYAWGQVGNLLGPASERTAHNMEMTRRLFT
FAHELAANRELAAGSPGAAMLEAADRGVIAAEKCPLLLVDFIGPSLETTAAAIGHLLVLFAERPD
QWDALRAEPGLVASTINELLRFETPVRGLTRVAARDTELGGVTPQGARIWALFASANRDERK
WERADEFDERRNPLDHLGFGYGAHNCVGGQVARLELHAVIHALLDSVERVELTGDPVIAENTIL
NTYAGIPVRFVARKAT

>Sri_CYP20

MRGEGAASPAEAAAAREQAPVAFPQDRTCPYRPPAAYDPLRDSRPLARVTLYDGATAWLVTGH
ETARALLADPRLSTDRTTRPGYPATSPRLRALPDRRVALLNVDDPEHHAQRRLVLANFTLKRAT
AMRPAIQRVVDERIEAMLAQGPPAELVSAFALPVPSTVICELGVPYADHEFFEEQSRRLLRSR
VQQEVREARDRLDGYLGALIDRKLEDPSGDGVLDALVDQQVREGTVDREEAVALATILLVAGH
ETTANMISLGVYTLQHPERLAELRADPSLLPAAVEELMRMLSADGLLRQATEDIDVGGTTIRA
GDGVVFATSVINRDESAFPAPDALDWSRSARHHVAFGFGIHQCLGQNLARAELEIALGTLARL
PGLRLAVPAEEARFKPGDTIQGMLDLPVTW

>Sri_CYP21

MARAESADALPKGFRSAERGWPELERIPHPHRIPLLDLLGASRRTPVQDSLRYAAELGPIFR
RKVFNREFVFGGAELAADMADETRFAKHVGLGIANLRPVVDALFTAYNHEPNWQLAHDVL
APGFSREAMAAYHPMMLDVAARLTGHWDAAEAAGRAVDIPGDMTKLTLETIARTGFGHDFGS
FERARPHPFVTAMVGTLTYAQRNTPVSPALLRRATRRNEADIALLNRTVDDLVRKRRSAGGG
EGDLLDRMLDTAHPSTGERLSALNVRQVVTFLVAGHETTSGALSFALHYLSLDPQLAARARA
EVDRVWGDTAVPGYEQVAKLRCVRRVLDEALRLWPTAPAFAREARADTVLGGEHPMRRGG
WALILTAMLHRDHEVWGPDAERFDPRFDAQAVRTRAPHYKPFGTGARACIGRQFALHEAV
LVLGLLLRRYELRPDPAYRLRVTERLTLMPEGLRMRLERRTRTSSDGPAGDPVSGPADDPGPE
GRGTPGSAPGCPVRGAGD

>Sri_CYP22

MTVESVKPVTPTGTEPSEPPVAGGGVPLLGHGWKMARDPLAFMARLRDHGDVVGLKLGPKT
VYAVTAPELTGALALSPDYIVSGPLWESLESLLGKEGVATANGPLHRRQRRTIQPAFRLDAIPG
YGPIMEEEAYALIERWTRGEILDVTAEAFRVSVRISARCLMRGSYMDARSERITSALATLFAGM
YQRMVVPLGPLYRVPVPANRKFNRALADLHRLVDEIVADRRASGQKPNDLLTALLEAKDENG
PIGEQEIHDQVVAILTPGSETAGSVVMSLLHVLTEHPDQVDKIREEVKNVVGDRPVAFADVRKL
QHTANVVVETMRLYPVWILTRRAVTDTELGGYRIPAGADIVYSPYAVQRDPRSYERHTEFDP
DRWLPGRSPNLPKYAMTPFGVGNRKCPSDHFSMAELTLITAVVANAFRFEPAPGSDNRTHIGI
TLRPRRLLLLRAIPG

>Sri_CYP23

VTNQPRPPADVPPAPELFTWEFATDPYPAYAWLREHAPVHRTELPSGVEAWLVTRYADAKEA
LADQRLSKNPAHHDEPAHARGKTGIPGERKAELMTHLLNIDPPDHTRLRRLVSKAFTPRRVAEF
APRVQELTDHLLDRFADRGEADLIHDFAPLPIYAICDLLGVPRDQDDFRDWAGMMIRHGGG
PRGGVARSVKKMRGYLLELIHRKRAELPARPAGEDLISGLIRASDHGEHLTENEAAAMAFILLF
AGFETTUNLIGNGAYALLTHPEQRARLQRSLTEDGGGPSLLETGIEELLYDGPVELATWRYAT
EPLRLGDQDIAPGDPVLVVLAAADRDPARFTDPDVLDSLRRDNQHLGYGHGIHYCLGAPLRL
EGRTALAALLTRLPLDLRLAADPAELRWRGGLIMRGLRTLTPVQFTPERSGT

>Sri_CYP24

MSTRPMVSADELEDIDLASPILHAHEYELDAVFRHLRANRPMYWQQPRGEQPGFWVISRYADV
NEVYKDKAHFTTENGALATLLTGGDSASGAMLAVTDGVRHHQVRNLLSKGFSPQMLELIANS
LRETVDGLLLAALDRGECDAARDISAKVPLGAICDLLEIPQTDREYLLGLTAHAWSTDYADEPPE
ESWVAKNEILLYFSKLLKERRGGDRDDMVSLLANCRIDGHPLNAAEQIANCYGLMIGGDETGR
HAITGTVLALIENPDQWRALKNGDVLKTATEEALRWTVPSLHGGRKATGDVVINGQQIKAGEV
VSVWISSANRDETVFDADEFKLARTPNKHFTFAYGSHYCLGHYLGRMEVYAVLDGLRRLVGD
LEQIGEERWIYSSILHGMSSSLPIKITA

11.2 Conserved regions of all 24 P450s identified in *S.rishiriensis*.

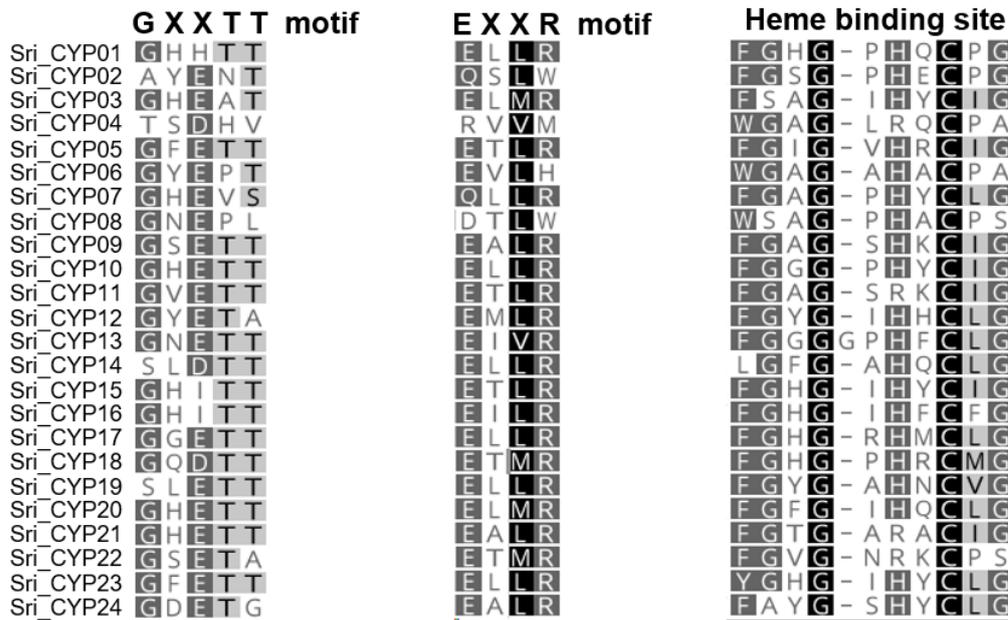


Figure C1. Multiple alignments of GXXTT, EXXR and heme-binding motif of all cytochrome P450s from *S.rishiriensis*.

11.3 Secondary metabolites

Table C1. Cytochrome P450s involved in secondary metabolites production by *Streptomyces*. *Streptomyces* produce a vast array of antibiotics, anti-parasitic agents as well as pharmacologically active metabolites. In production of those metabolites, the CYPs' involvements are summarized below⁽¹⁸⁰⁾.

<i>Streptomyces</i>	CYPs	Secondary Metabolites	Applications
<i>S. lavendulae</i>	CYP107N1 CYP160A1 CYP105F1 CYP165B5	Mitomycin C	Anti-HIV
	CYP165E1	Complestatin	
<i>S. antibioticus</i>	CYP107D1 CYP235A1	Oleandomycin	Antibacterial
	CYP163A3	Simocyclinone	Antifungal
<i>S. avermitilis</i>	CYP105P1 CYP105D6	Filipin	Antifungal
	CYP171A1	Avermectin	Antiparasitic
<i>S. fradiae</i>	CYP113B1 CYP105L1 CYP154B1	Tylosin	Antibacterial
<i>S. hygroscopicus</i>	CYP122A2 CYP107G1	Rapamycin	Antibacterial

	CYP105U1	Geldanamycin	Antitumor
<i>S. carzinostaticus</i>	CYP154J1 CYP208A2	Neocarzinostatin	Antitumor
<i>S. griseolus</i>	CYP105A1 CYP105B1	7-ethoxycoumarin	Antimicrobial
<i>S. nanchangensis</i>	CYP124B2	Nanchangmycin	Anticoccidial
	CYP171A2	Meilingmycin	Antibacterial
<i>S. natalensis</i>	CYP161A2 CYP105H3	Pimaricin	Antifungal
<i>S. nodosus</i>	CYP161A3 CYP105H4	Amphotericin	Antifungal
<i>S. noursei</i>	CYP161A1 CYP105H1	Nystatin	Antifungal
<i>S. tendae</i>	CYP162A1 CYP105K1	Nikkomycin	Insecticidal
<i>S. peucetius</i>	CYP131A1 CYP129A2	Doxorubicin	Antitumor
<i>S. sp. strain C5</i>	CYP131A2 CYP129A1	Daunorubicin	Antitumor
<i>S. ansochromogenes</i>	CYP105K2	Nikkomycin	Insecticidal
<i>S. carbophilus</i>	CYP105A3	Pravastatin	Anti-HIV
<i>S. clavuligerus</i>	CYP105M1	Clavulinic acid	Antibacterial
<i>S. caelestis</i>	CYP113B2	Niddamycin	Antibacterial
<i>S. cinnamomensis</i>	CYP124B1	Monensin	Anticoccidial
<i>S. globisporus</i>	CYP211A1	Antibiotic C-1027	Antitumor
<i>S. narbonensis</i>	CYP107L7P	Narbomycin	Antifungal
<i>S. spheroides</i>	CYP163A1	Novobiocin	Antibacterial
<i>S. thermotolerans</i>	CYP107C1	Carbomycin	Antibacterial
<i>S. roseochromogenes</i>	CYP163A2	Clorobiocin	Antibacterial
<i>S. venezuelae</i>	CYP107L1	Picromycin	Antibacterial

11.4 Expression of Sri_CYP03 and Sri_CYP13 with every chaperone system.

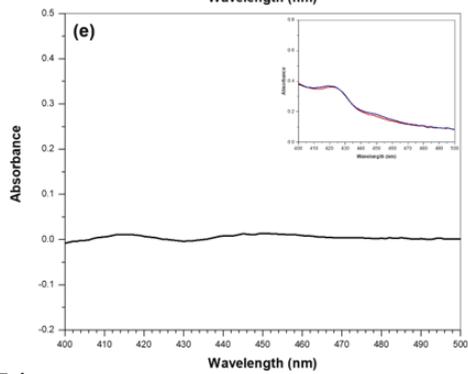
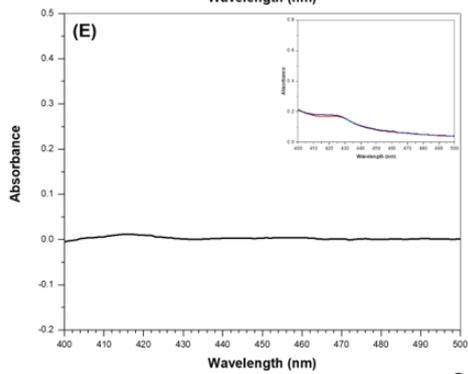
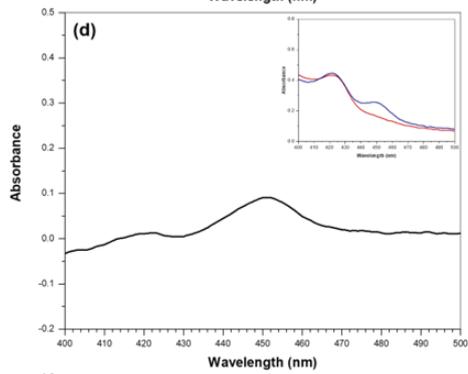
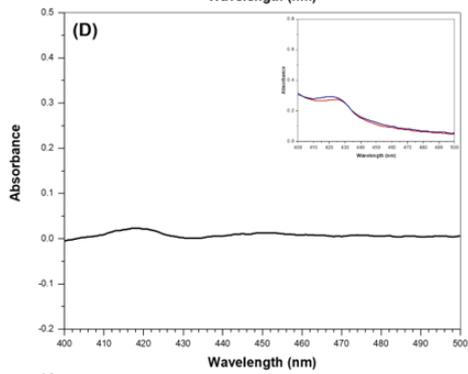
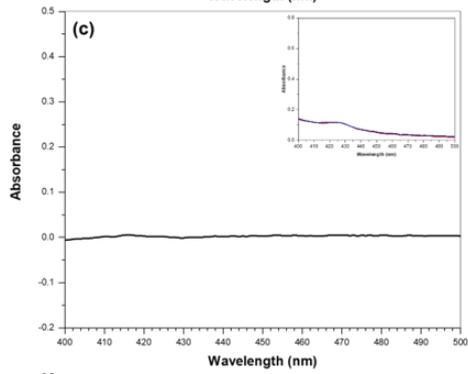
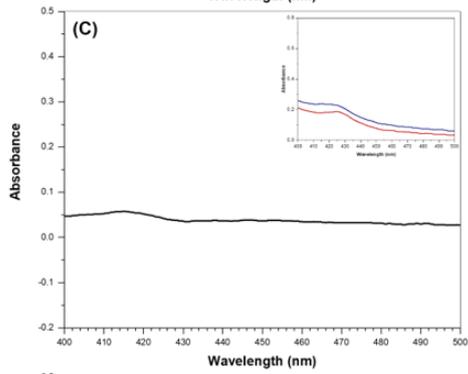
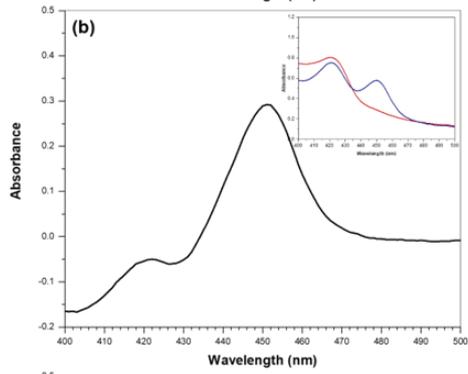
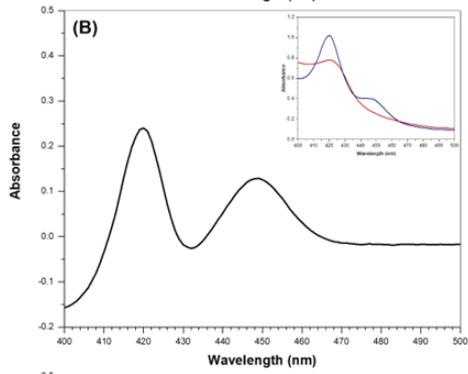
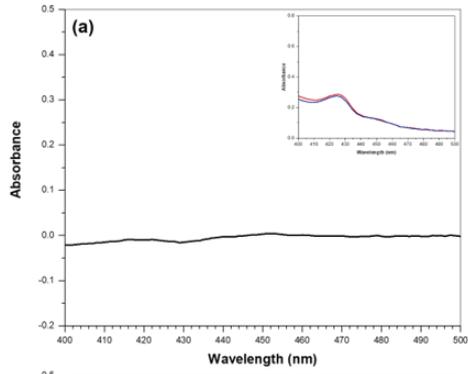
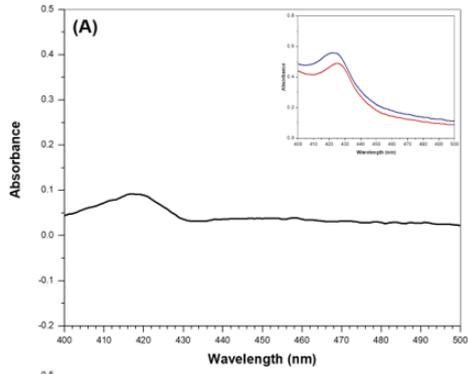


Figure C2. Coexpression of CYPs and five different chaperone systems. The two CYP plasmids used in these expression tests are pQR2303 and pQR2304, which contain only the *single-cyp* construct. (A) Coexpression of Sri_03 and two chaperone systems DnaK-DnaJ-GrpE and GroES-GroEL, 0.1 mM IPTG, 25°C, 24 hours. (B) Coexpression of Sri_CYP03 and GroES-GroEL system, 0.1 mM IPTG, 25°C, 24 hours. (C) Coexpression of Sri_CYP03 and DnaK-DnaJ-GrpE system, 0.1 mM IPTG, 25°C, 24 hours. (D) Coexpression of Sri_CYP03 and GroES-GroEL-Trigger factor system, 0.1 mM IPTG, 25°C, 24 hours. (E) Coexpression of Sri_CYP03 and Trigger factor, 0.1 mM IPTG, 25°C, 24 hours. (a) Coexpression of Sri_13 and two chaperone systems DnaK-DnaJ-GrpE and GroES-GroEL, 0.1 mM IPTG, 25°C, 24 hours. (b) Coexpression of Sri_CYP13 and GroES-GroEL system, 0.1 mM IPTG, 25°C, 24 hours. (c) Coexpression of Sri_CYP13 and DnaK-DnaJ-GrpE system, 0.1 mM IPTG, 25°C, 24 hours. (d) Coexpression of Sri_CYP13 and GroES-GroEL-Trigger factor system, 0.1 mM IPTG, 25°C, 24 hours. (e) Coexpression of Sri_CYP13 and Trigger factor, 0.1 mM IPTG, 25°C, 24 hours.

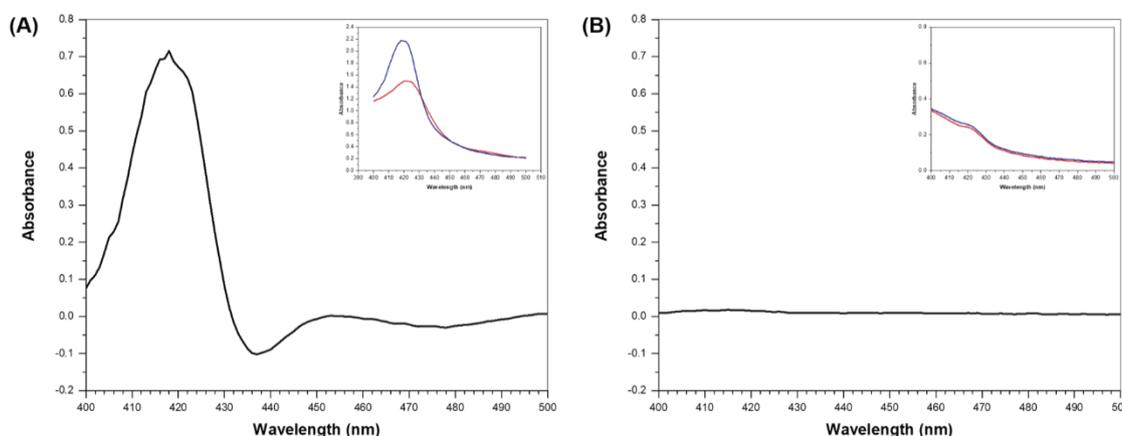


Figure C3. Coexpression of CYPs and GroES/GroEL chaperone system. The two CYP plasmids are pQR2300 and pQR2301, which contain the *cyp-fd2-fdr* operon. (A) Coexpression of Sri_CYP03 and GroES/GroEL chaperone system, 0.1 mM IPTG induction, 25°C, 24 hours. There is highly enhanced production of inactively folded P450, but no detectable active P450. (B) Coexpression of Sri13_CYP13 and GroES and GroEL chaperone system, 0.1 mM IPTG induction, 25°C, 24 hours. There is no structurally active or inactive P450 being produced in *E.coli*.

11.5 The calculated standard curves for substrates used in preliminary screening

The five substrates listed in Table 6.7 were prepared in different concentrations in DMSO and analysed through HPLC using HPLC method 1 (described in section 2.1.10.1). The retention time for each substrate was identified, and the calculated peak areas were correlated with the prepared substrate concentrations to identify the linear standard curves of each substrates (Figure C4 to C8)

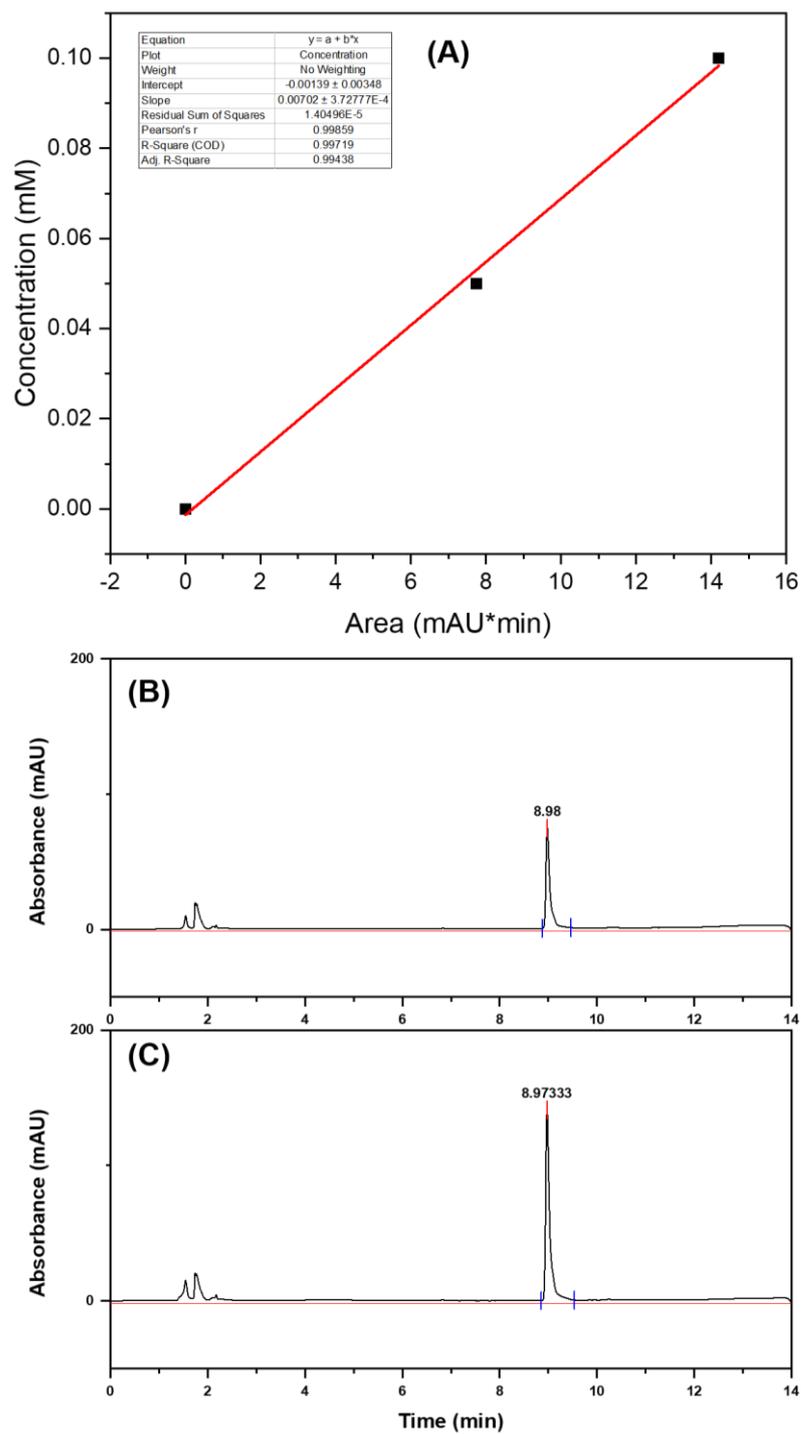


Figure C4. The calculated standard curve of apigenin. The linear correlation between peak area and standard concentration of apigenin was calculated. The retention time for apigenin was revealed to be 8.98 min. (A) The calculated linear fit; (B) The HPLC chromatogram of 0.05 mM apigenin; (C) The HPLC chromatogram of 0.1 mM apigenin.

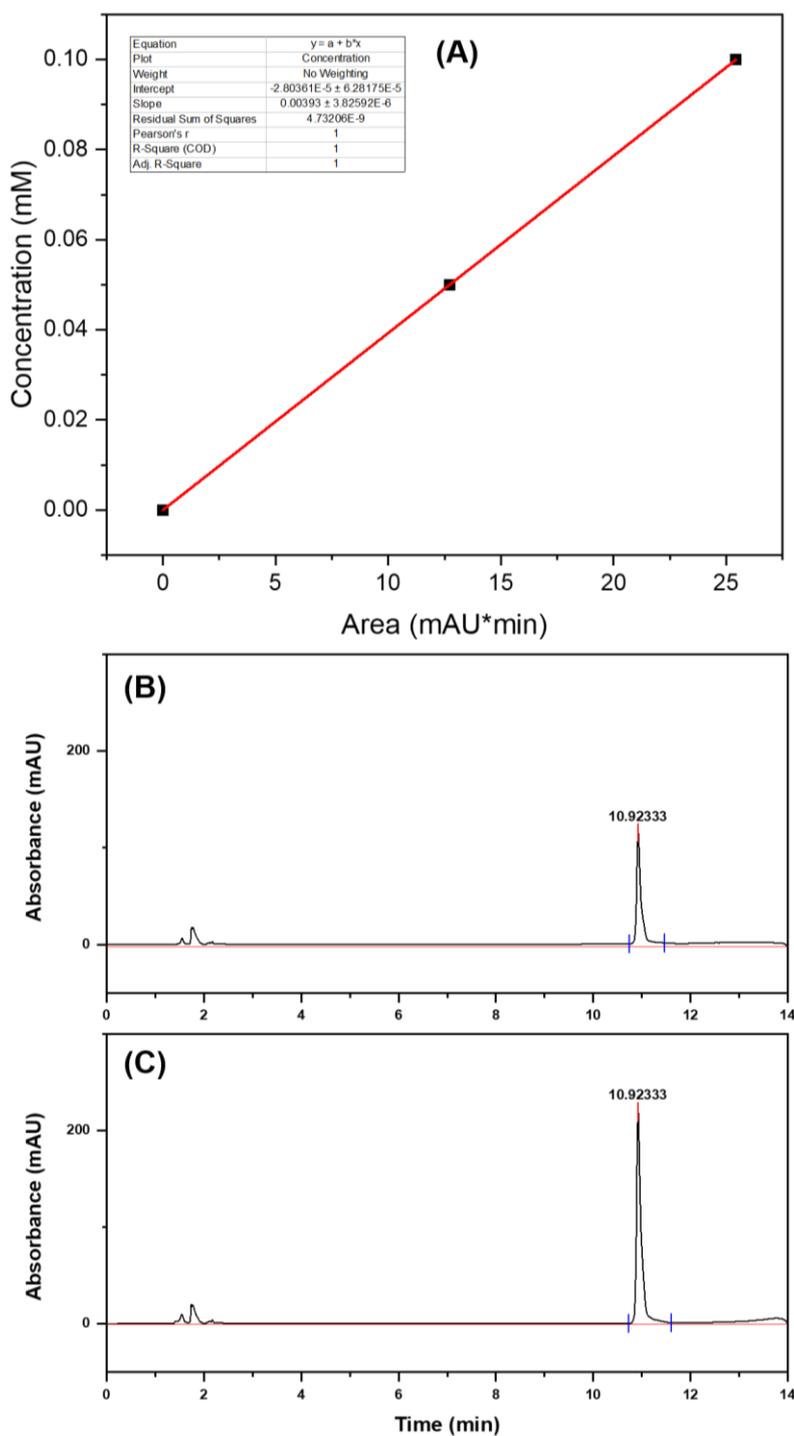


Figure C5. The calculated standard curve of chrysin. The linear correlation between peak area and standard concentration of chrysin was calculated. The retention time for chrysin was revealed to be 10.92 min. (A) The calculated linear fit; (B) The HPLC chromatogram of 0.05 mM chrysin; (C) The HPLC chromatogram of 0.1 mM chrysin.

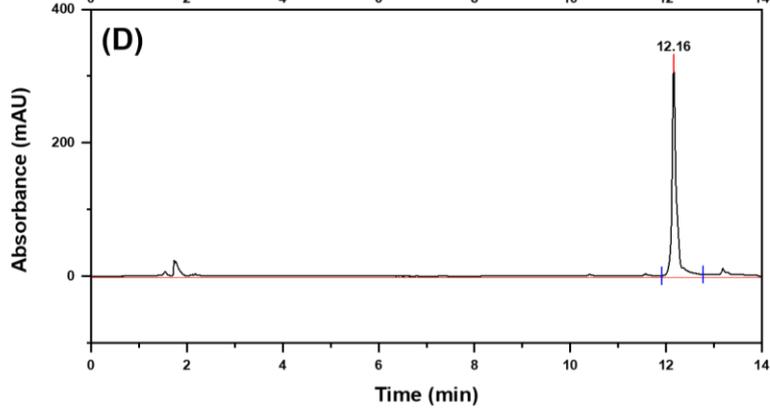
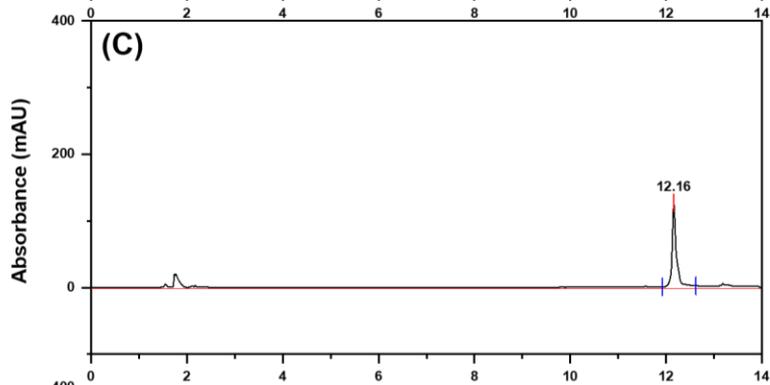
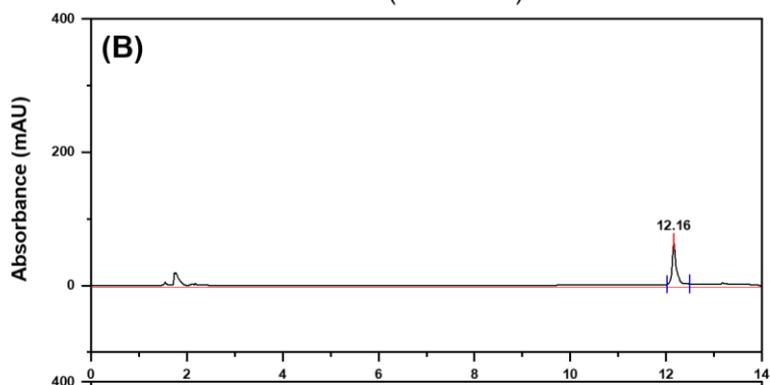
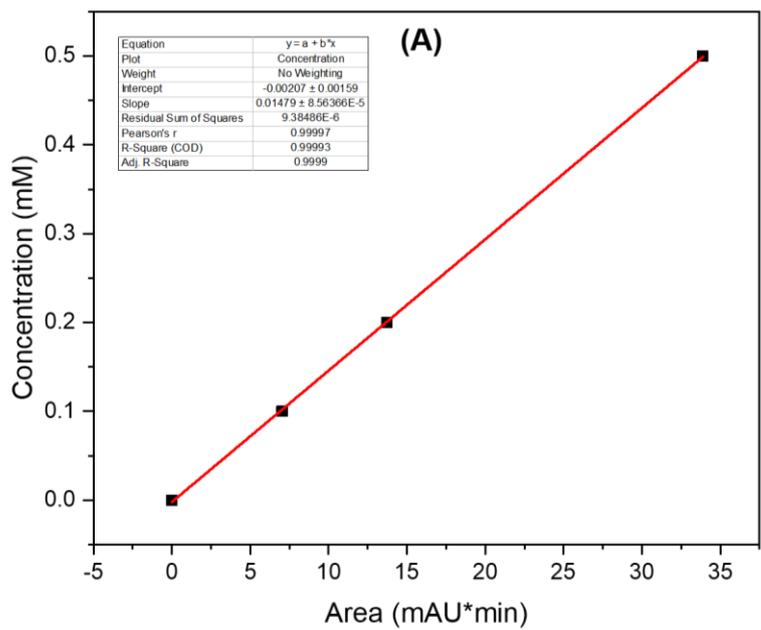


Figure C6. The calculated standard curve of 7-ethoxycoumarin. The linear correlation between peak area and standard concentration of 7-ethoxycoumarin was calculated. The retention time for 7-ethoxycoumarin was revealed to be 12.16 min. (A) The calculated linear fit; (B) The HPLC chromatogram of 0.1 mM 7-ethoxycoumarin; (C) The HPLC chromatogram of 0.2 mM 7-ethoxycoumarin. (D) The HPLC chromatogram of 0.5 mM 7-ethoxycoumarin.

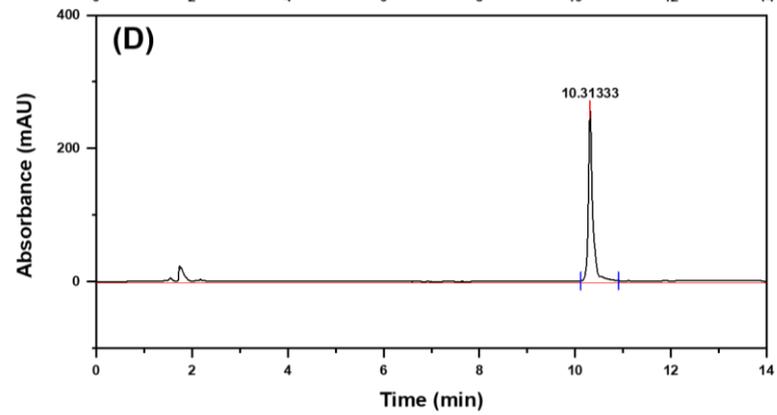
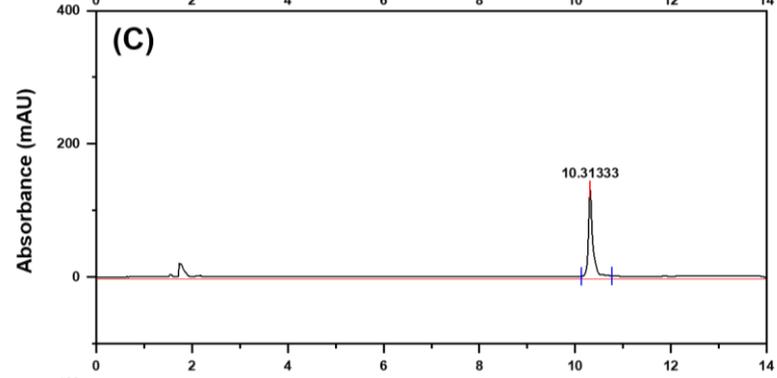
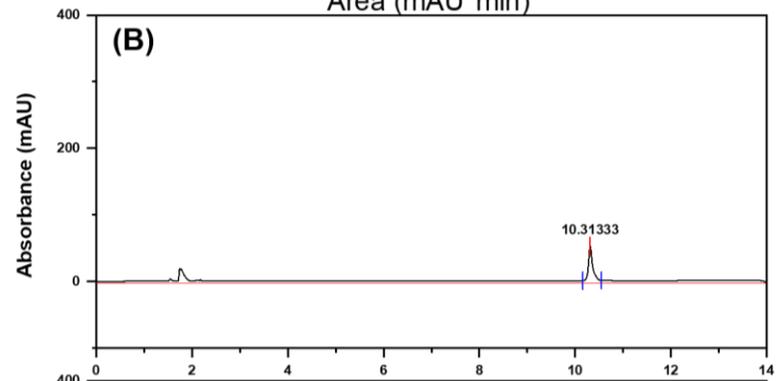
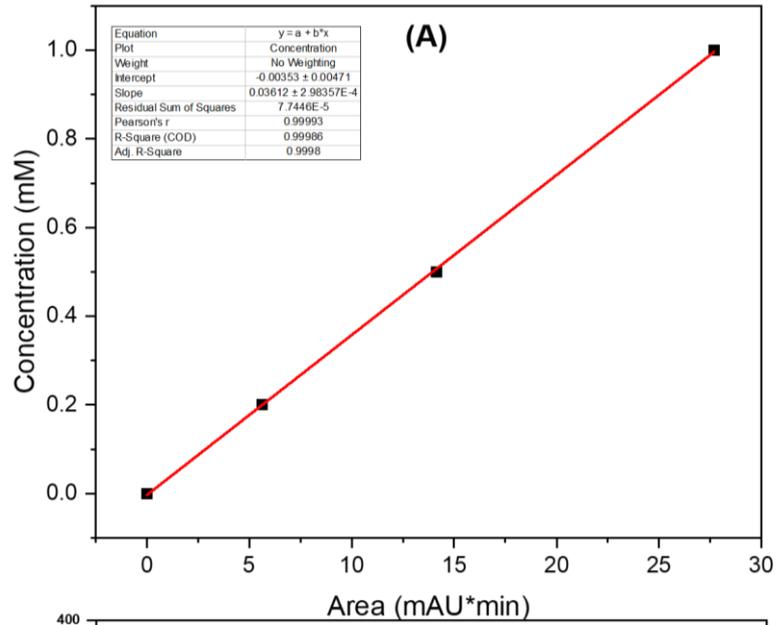


Figure C7. The calculated standard curve of diclofenac. The linear correlation between peak area and standard concentration of diclofenac was calculated. The retention time for diclofenac was revealed to be 10.31 min. (A) The calculated linear fit; (B) The HPLC chromatogram of 0.2 mM diclofenac; (C) The HPLC chromatogram of 0.5 mM diclofenac. (D) The HPLC chromatogram of 1 mM diclofenac.

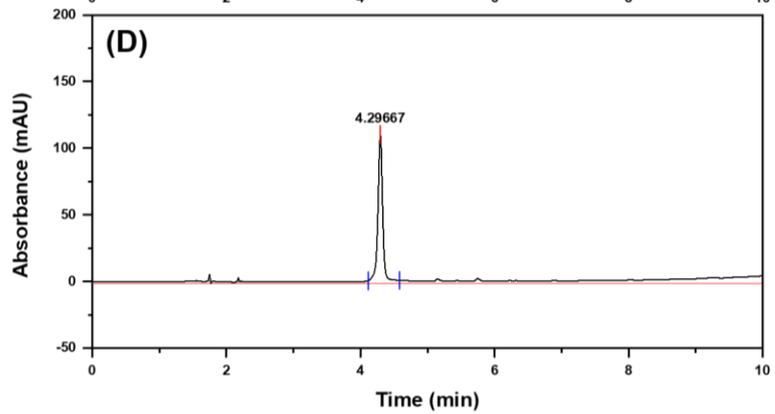
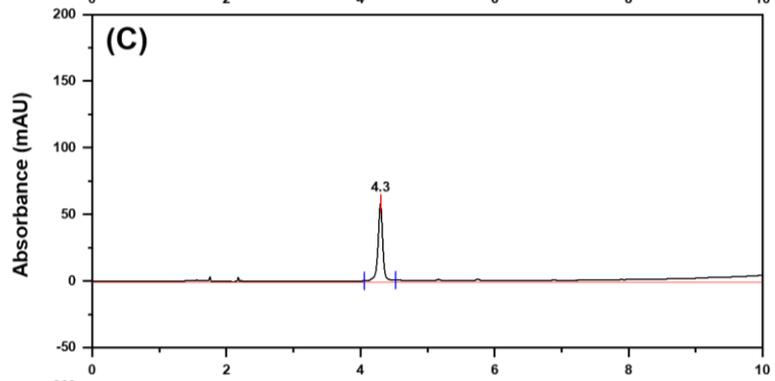
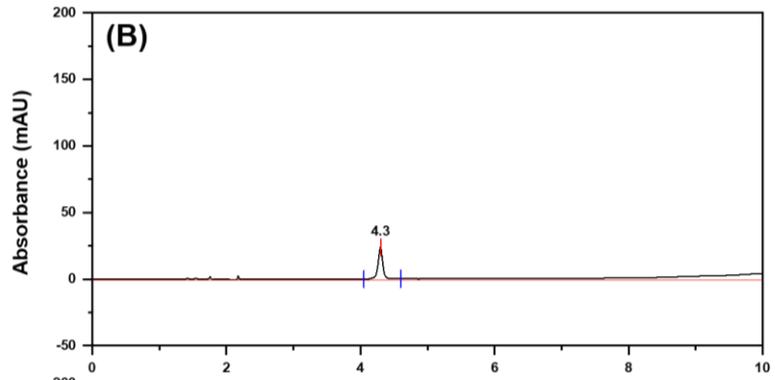
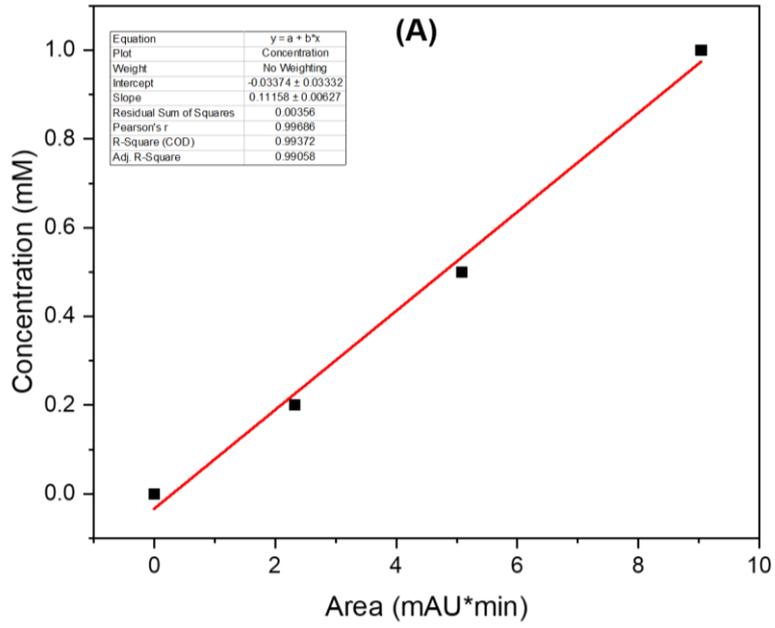
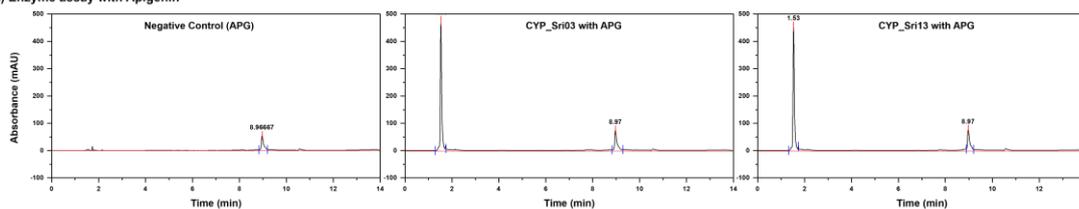


Figure C8. The calculated standard curve of THIQ. The linear correlation between peak area and standard concentration of THIQ was calculated. The retention time for THIQ was revealed to be 4.29 min. (A) The calculated linear fit; (B) The HPLC chromatogram of 0.2 mM THIQ; (C) The HPLC chromatogram of 0.5 mM THIQ. (D) The HPLC chromatogram of 1 mM THIQ.

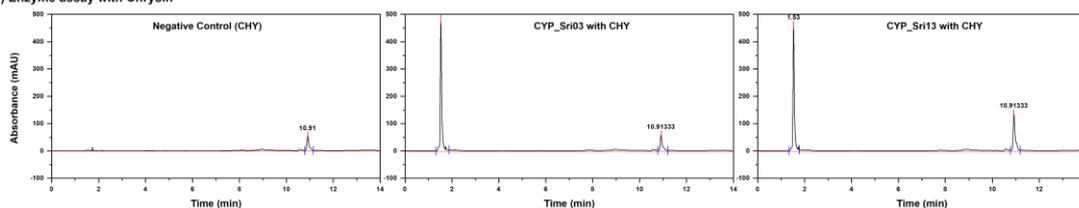
11.6 Screening results

The screening reaction was prepared with CYPs, Pdx, PdR, NADH and substrates, and analysed through HPLC. The negative control reaction revealed the initial concentration of substrates as well as the retention time of substrates. By comparing the reaction chromatogram to control chromatogram, potential new peaks were not identified. In addition, there was no obvious deduction of the NADH peak either, which indicated NADH was not consumed by the system for electron donating to potential cytochrome P450 conducted reactions.

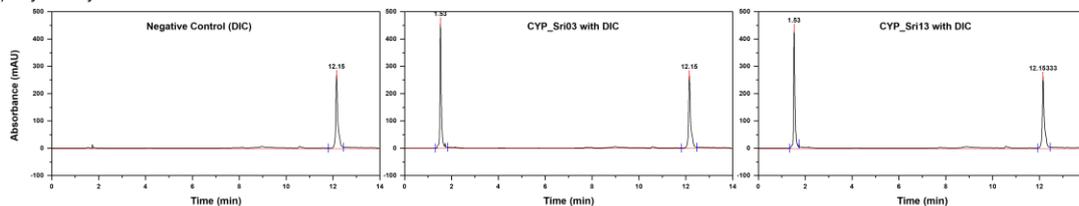
(A) Enzyme assay with Apigenin



(B) Enzyme assay with Chrysin



(C) Enzyme assay with Diclofenac



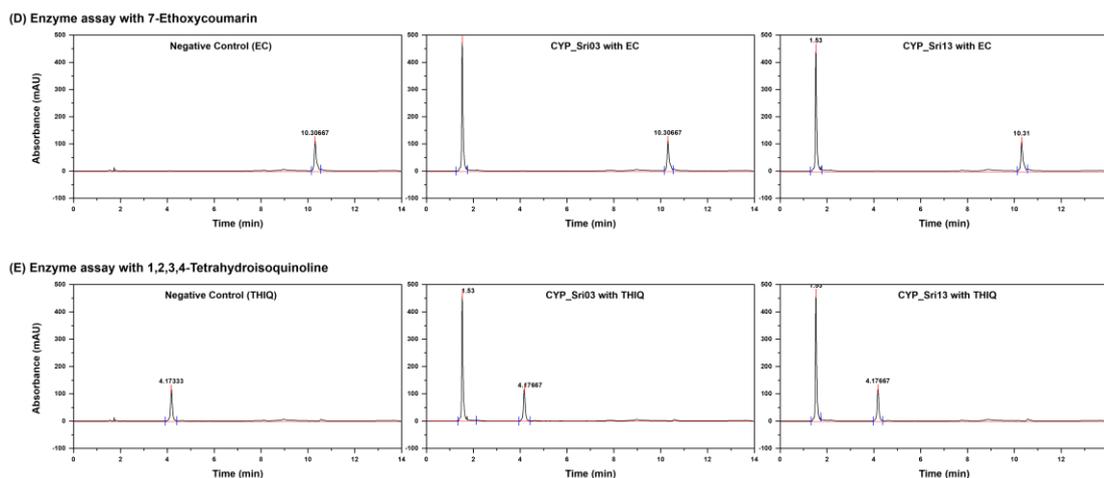


Figure C9. The screening results of CYP_Sri03 and CYP_Sri13 against five different substrates. The reaction mixture was analysed through HPLC and the substrates could be detected at 280 nm. (A) Chromatogram of assays with apigenin. The retention time of apigenin was 8.97 min, and there were no new peaks being identified in the reaction chromatogram. (B) Chromatogram of assays with chrysin. The retention time of chrysin was 10.91 min, and there were no new peaks being identified in the reaction chromatogram. (C) Chromatogram of assays with diclofenac. The retention time of diclofenac was 12.16 min, and there were no new peaks being identified in the reaction chromatogram. (D) Chromatogram of assays with ethoxycoumarin. The retention time of ethoxycoumarin was 10.31 min, and there were no new peaks being identified in the reaction chromatogram. (E) Chromatogram of assays with THIQ. The retention time of THIQ was 4.17 min, and there were no new peaks being identified in the reaction chromatogram.

11.7 Original gel image of Figure 6.8

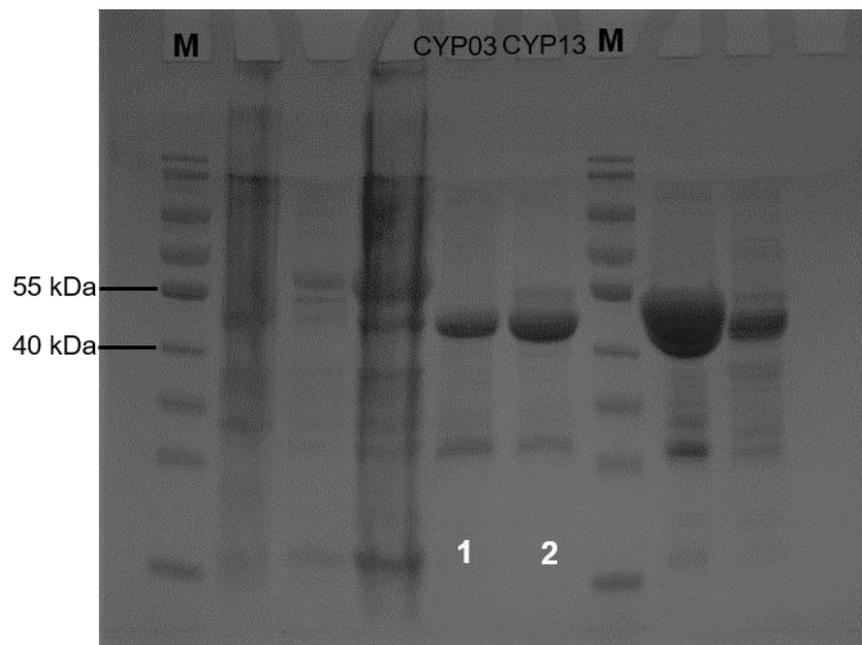


Figure C10. The gel image of Figure 6.8.

12 Appendix D

This appendix contains supplementary information for Chapter 7.

12.1 Complete gene sequences for *dmg_cyp08* and *dmg_cyp09*.

>*dmg_cyp08*

```
ATGAACGGCTTGGCCGCCGCCATAGCGCATCTTGAGTCAGGTGCGCTCG
ATGTTACGGCGAACGACTTGTACGGGGAGGGCCGTTGGCCCGACGTATT
CGCGCGCCTGCGGGCGGACGATCCCGTGCATTTCTGTCCCGCGAGCTTC
TACGGGCCTTACTGGTCCGTACGCGCCACGCCCATGTCGCGCATGTGG
AGGCACTGCCGCAGCTCTTTTCGTTCGTTCGTGGACGCACGGCGGCATCAC
CATCCTGCACAACGATCCCGACATCCAGCTGCCGATGTTTCATCGCGATGG
ACCGGCCCGAGCACACCGCGATGCGGGCGCACCGTGTCGCCTGCATTCAG
CCCGAGCGCACTCACCGACCATTCCGCCGCGATCCGCGCACGCACGGCG
GCGCTGCTCGACAGCCTTCCGGTGGGCGAAGCGTTTGATTGGGTGGACC
GGGTTTCGACCGAACTCACGACGCAGATGCTGGCGACGCTGTTTCGATTTCC
CCGTGGGAAGATCGCCGCCTGCTCGCAACATGGGCGAACTGGGCGGGTG
ATTTCCGGGCTGCGGTGAGCCCGGCGCTGCGCCGCAAGCGCCAGCGCG
TGCTGCAGCAGATGGTCGATTATTTCCGGCGGGCTGTGGGCCGAGCGGCG
GCACGCCGGGCTGGCGCCGGACCTCCTTTTCGCGGATGATCCATTCGGAC
GCGATGTCCAACCTTCGCCCGCAGGAGTTCTGCGGCAACCTGATGCTGCT
GGTGATCGGCGGCACCGATAACGACGCGCAACTCGATGACCGGGCTCGTT
GTCGCGCTCGATCAGTTTCCCGAAAACCGGGCGCTGCTGGATGCGGACC
GCGGGCTCGTCGTCAATGCCGTGTCCGAACTCATCCGCTGGCAGACGCC
GCTGTCGCACATGCGGGCGTACGGCGACGGAAGACACGGAGATCGGGCGG
CAAGATCATCCGCAAGGATGACAAGGTCGTGCTGTGGTACAACAGTGCGA
ACCGCGACGAAGCCGTGTTACCGATGCGGATCGCTTCGACGTGACGCG
CGCGAATGCGCGGGCGCCATCTCGCCTTCGGCTTCGGCATCCATCGCTGC
GTCGGCGCCCGGCTTGCCGAGCTTCAGCTCTACATCCTCCTCGAAGAGAT
GCTGAACCGCGAGATGCGGGTGACGCCGCTCCACGCGCCGATGCGGATT
CCCTCGCCCTTTGCGAACGGATAACCGCAAGCAGATGGTGACGATCACCCG
ATCCTAG
```

> *dmg_cyp09*

ATGGCGACCAAAGTGGCGGGCGGATACGAGCGATCTCGCCGCGATCGATC
CGGCGGACTACCGCTATTTACCGAGCAGACATGGCCCCGCGCTGTTTCGC
GCGGCTGCGTGCCGAAGACCCGGTGCCTACTGCACGGAGGGCATGTTT
GCGCCTTACTGGTTCGGTGACGCGCTACAAGGACATCGTGCATGTTCGAGG
CGCTGCCCCGACATTTATTCCTCGTTCGTGGGAGTACGGCGGCATCACCGTG
CTCGATGCGCTCGACCAGGAACTGAAGTTCCCCATGTTTCATCGCGATGGA
TCGGCCGGAGCACACTGCCGAGCGGGCGCGTGGTGGCGCCTGCCTTAC
GCCGAGCGAGATGCAGCGGCTCGCCGGGCGCGATCCGCGAACGGACGGC
GGTGCTGCTCGATGCGCTTCCGGTTCGGTGAGACCTTCGACTGGGTTCGAC
CGGGTTTCGATCGAACTCACACGCAGATGCTCGCCATCCTGTTTCGATTT
CCCGTGGGAGGATCGGCGGCTGCTGACCAAGTGGTTCGGACTGGGCGGG
GGACGTGCAGGCGGGCGGCCGATCCGGCGATGCGCGAGGAGCGCACGCG
GGTTCTCTACGAGATGGGCGCCTACTTCCATGGGCTGTGGCAGGCACGG
CTGAATGCGCCGCCCGCGCCGGACCTGCTTTTCGATGATGATCCATTCCGA
AGCGCTCTCGAAAATGGACGAGCAGGAGTTCATCGGGAACCTCGTGCTGT
TGATCGTGGGCGGCAACGACACGACGCGCAACTCGATGACCGGGCTCGT
GCGCGTGATGAACGAATTTCCGGCCGAATGGGACAAGCTGAAGGCGGAC
CGGGGCCTGATCCCCAATGCCGTGTCCGAACCTGATCCGCTGGCAGACGC
CGCTCGCCCACATGCGTCGCACGGCGAAAGAGGACACCGAACTCGGAGG
CAAGACCATCCACAAGGGCGACAAGGTGGTGTGATGTGGTACATTTCCGCCA
ACCGCGACACGGAGGTCTTCGAAGACCCGGACCGCGTGGATTTTGCCCG
CGAGAATGCGCGGGCGGCACCTGTTCGTTTCGGCTACGGCATCCACCGCTGC
GTGGGCGCGCGTTCGGCCGAATTGCAGCTTTTCGATCCTGATCGAGGAAAT
GCTCGCCCGTGGCATGGATGTAAAGGCTGTTGCACCGCCGGTGCCTGTG
CCATCGCCGTTTCATCAACGGCTACAAGGAGCAGATGGTTTCGCATCACCAA
GGGTAA

12.2 GC chromatograms and mass spectra of 0.5 mM alkanes

The 0.5 mM alkane standards were firstly prepared in both ethyl acetate and potassium phosphate buffer (pH 7.5). Both standards were analysed through GCMS to identify the retention time on GC chromatogram as well as the mass characterisation of each alkane substrate.

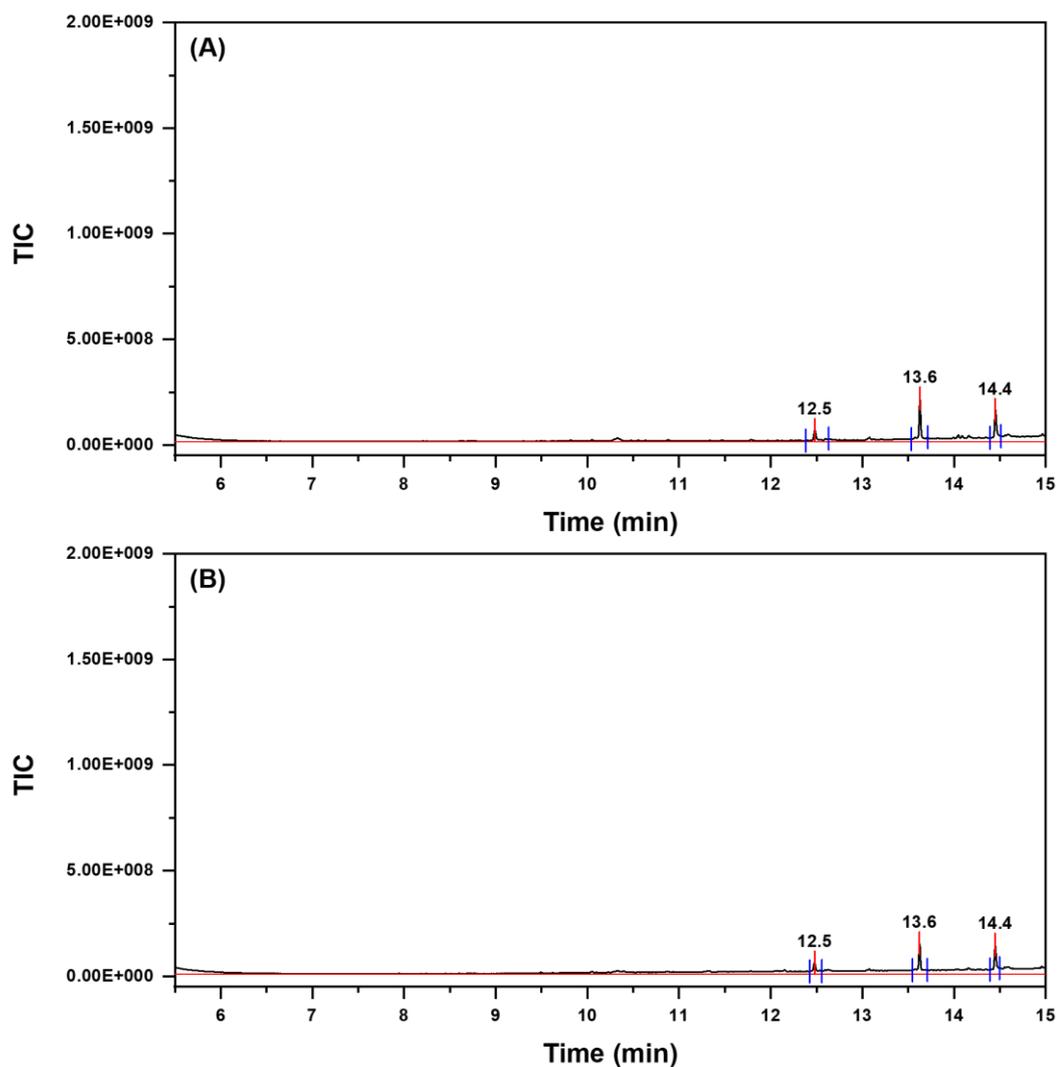


Figure D1. GC chromatogram for ethyl acetate. (A) GC chromatogram of pure ethyl acetate; (B) GC chromatogram of ethyl acetate acquired from extraction on 50 mM potassium phosphate buffer (pH 7.5).

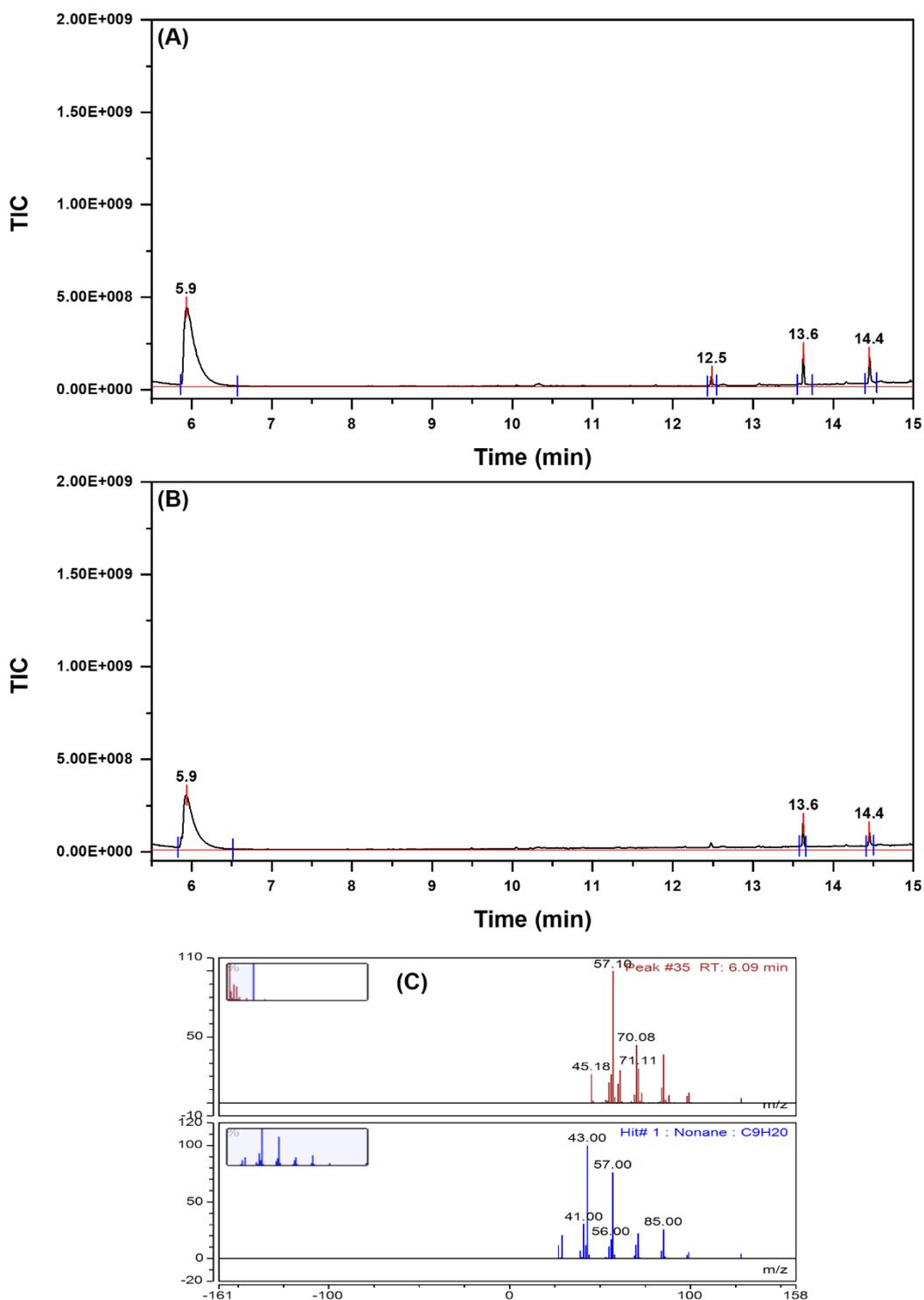


Figure D2. GC chromatogram and mass spectrum for 0.5 mM nonane (C₉H₂₀). (A) GC chromatogram of nonane dissolved in ethyl acetate; (B) GC chromatogram of nonane prepared in potassium phosphate buffer and extracted in ethyl acetate; (C) MS spectrum of nonane and comparison to standard nonane in library.

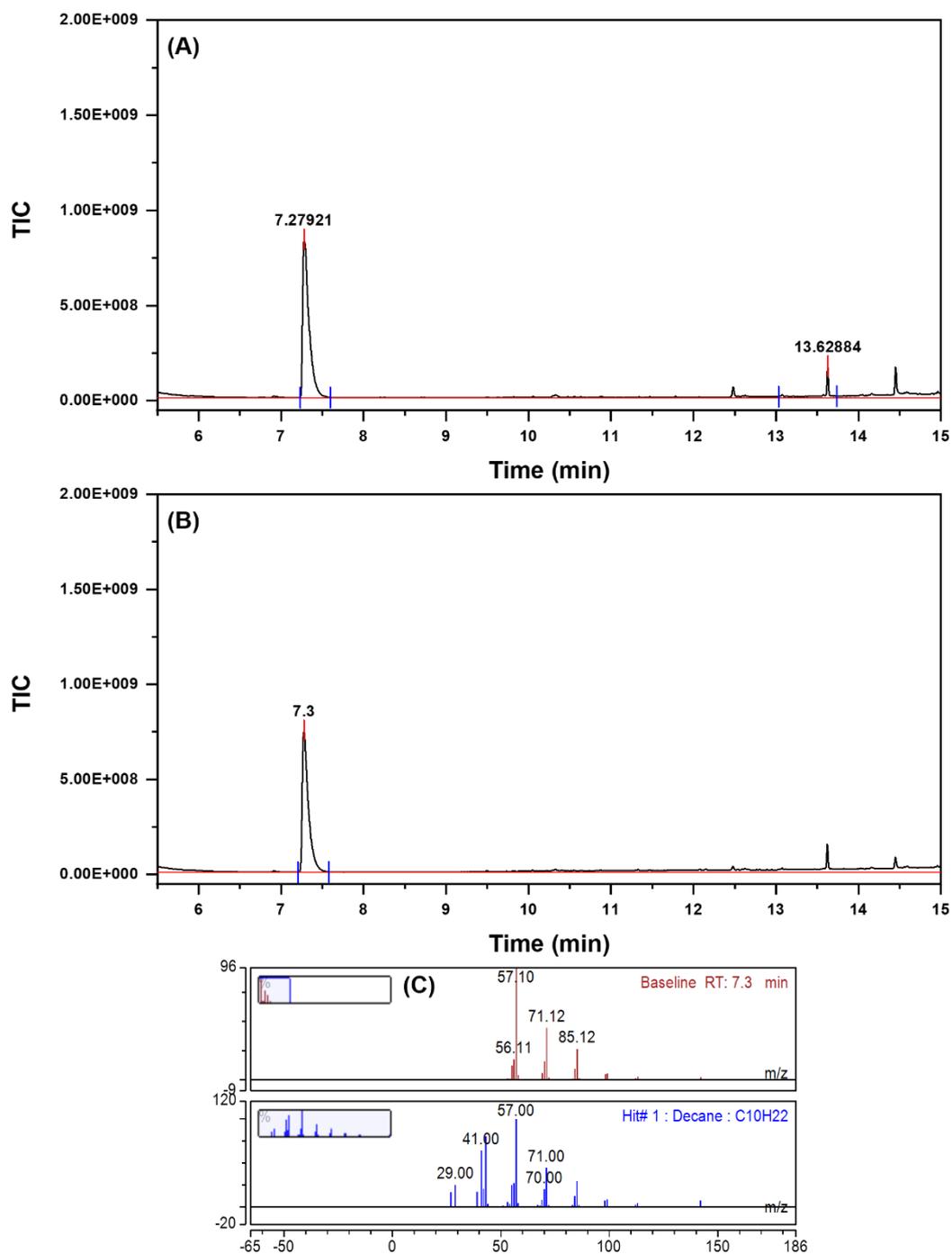


Figure D3. GC chromatogram and mass spectrum for 0.5 mM decane (C₁₀H₂₂). (A) GC chromatogram of decane dissolved in ethyl acetate; (B) GC chromatogram of decane prepared in potassium phosphate buffer and extracted in ethyl acetate; (C) MS spectrum of decane and comparison to standard decane in library.

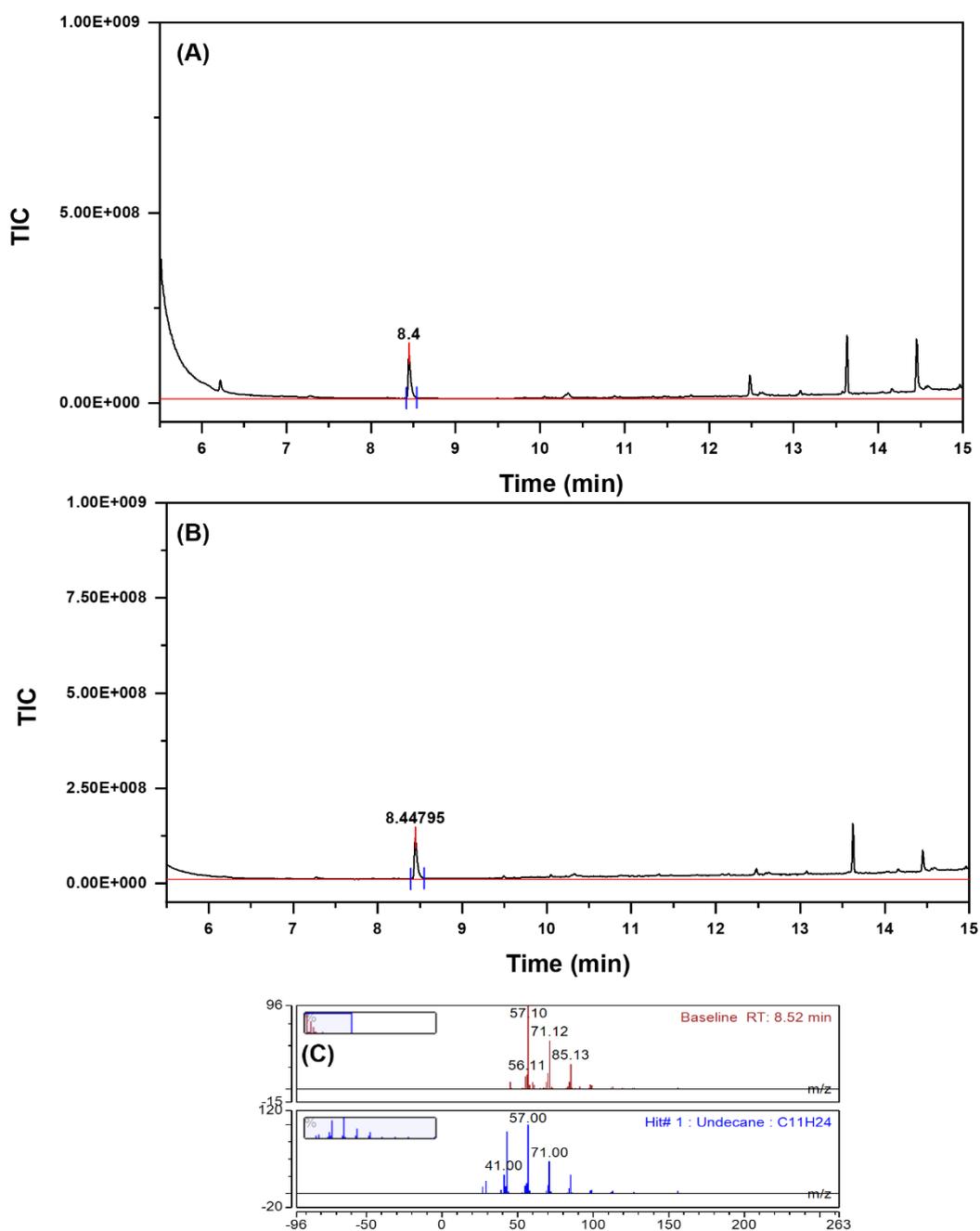


Figure D4. GC chromatogram and mass spectrum for 0.5 mM undecane (C₁₁H₂₄). (A) GC chromatogram of undecane dissolved in ethyl acetate; (B) GC chromatogram of undecane prepared in potassium phosphate buffer and extracted in ethyl acetate; (C) MS spectrum of undecane and comparison to standard undecane in the library.

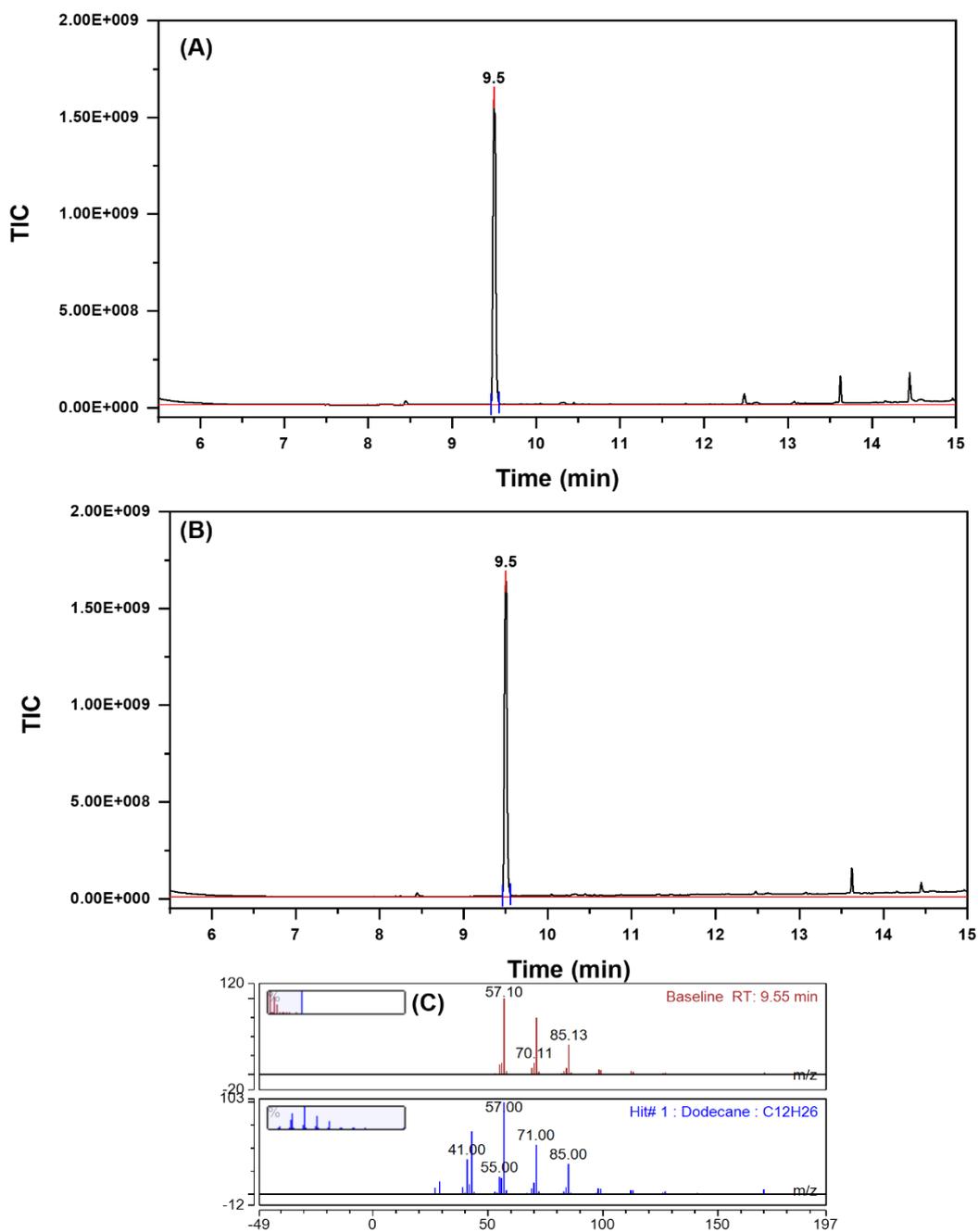


Figure D5. GC chromatogram and mass spectrum for 0.5 mM dodecane (C₁₂H₂₆). (A) GC chromatogram of dodecane dissolved in ethyl acetate; (B) GC chromatogram of dodecane prepared in potassium phosphate buffer and extracted in ethyl acetate; (C) MS spectrum of dodecane and comparison to standard dodecane in the library.

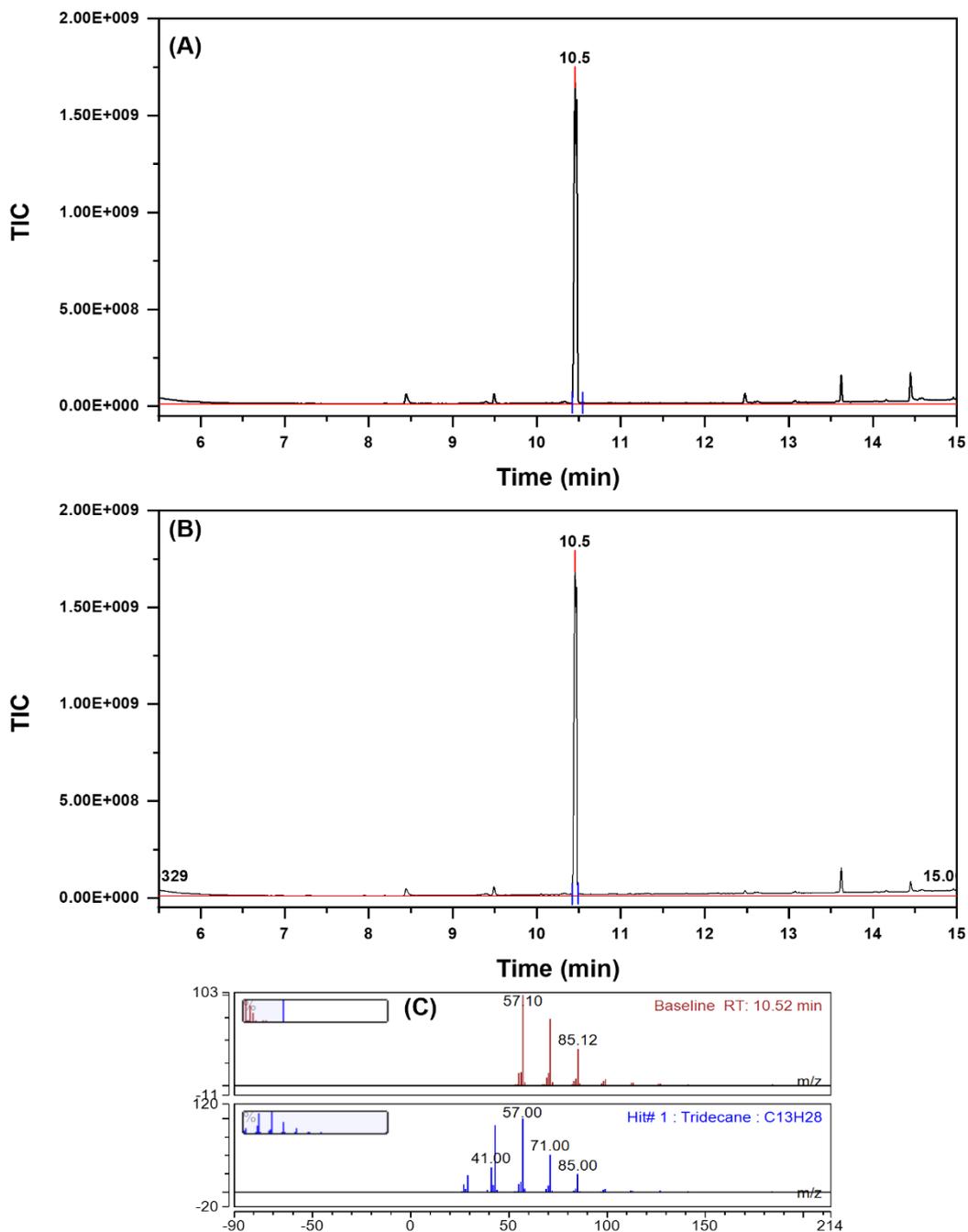


Figure D6. GC chromatogram and mass spectrum for 0.5 mM tridecane (C₁₃H₂₈). (A) GC chromatogram of tridecane dissolved in ethyl acetate; (B) GC chromatogram of tridecane prepared in potassium phosphate buffer and extracted in ethyl acetate; (C) MS spectrum of tridecane and comparison to standard tridecane in library.

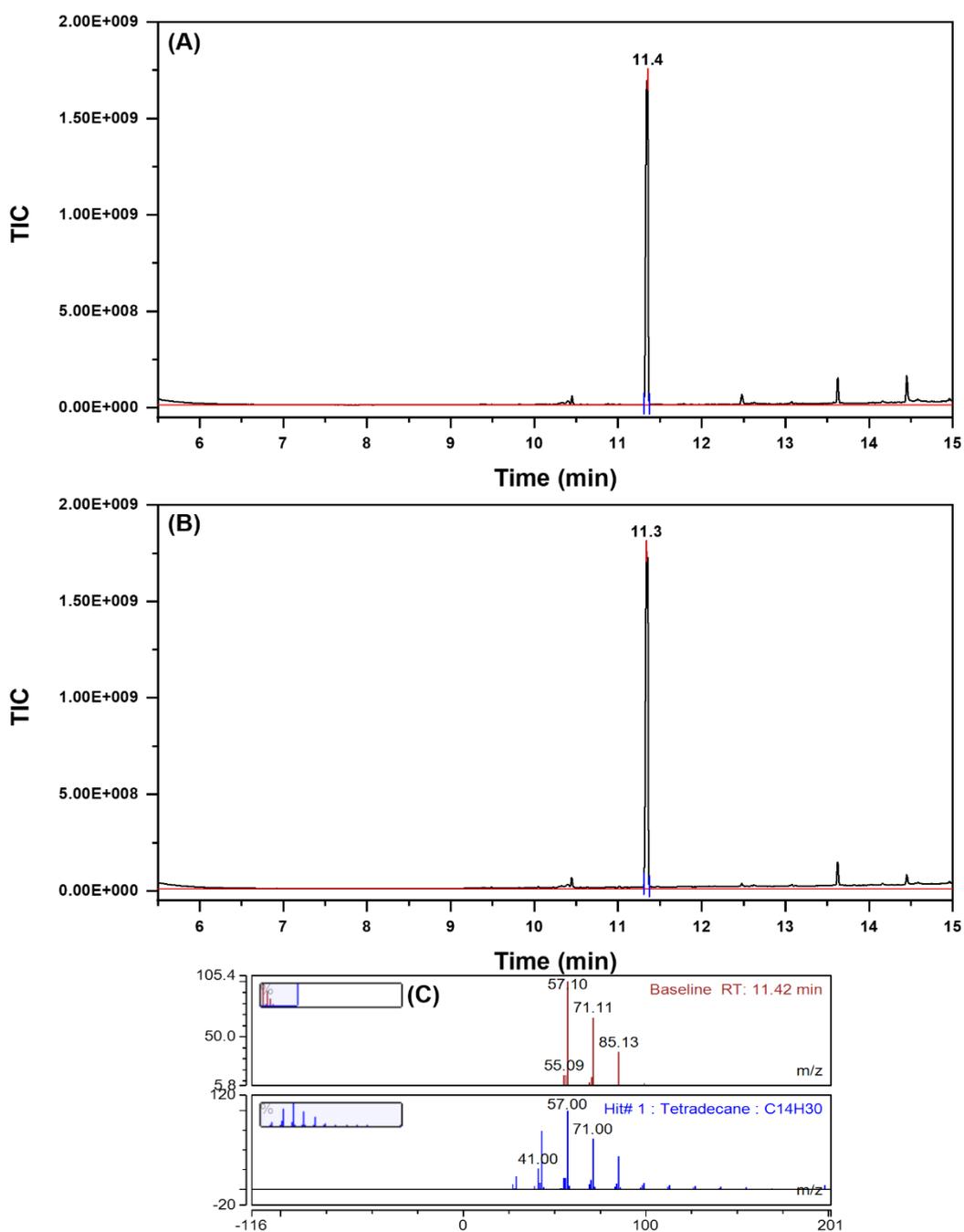


Figure D7. GC chromatogram and mass spectrum for 0.5 mM tetradecane (C₁₄H₃₀). (A) GC chromatogram of tetradecane dissolved in ethyl acetate; (B) GC chromatogram of tetradecane prepared in potassium phosphate buffer and extracted in ethyl acetate; (C) MS spectrum of tetradecane and comparison to standard tetradecane in library.

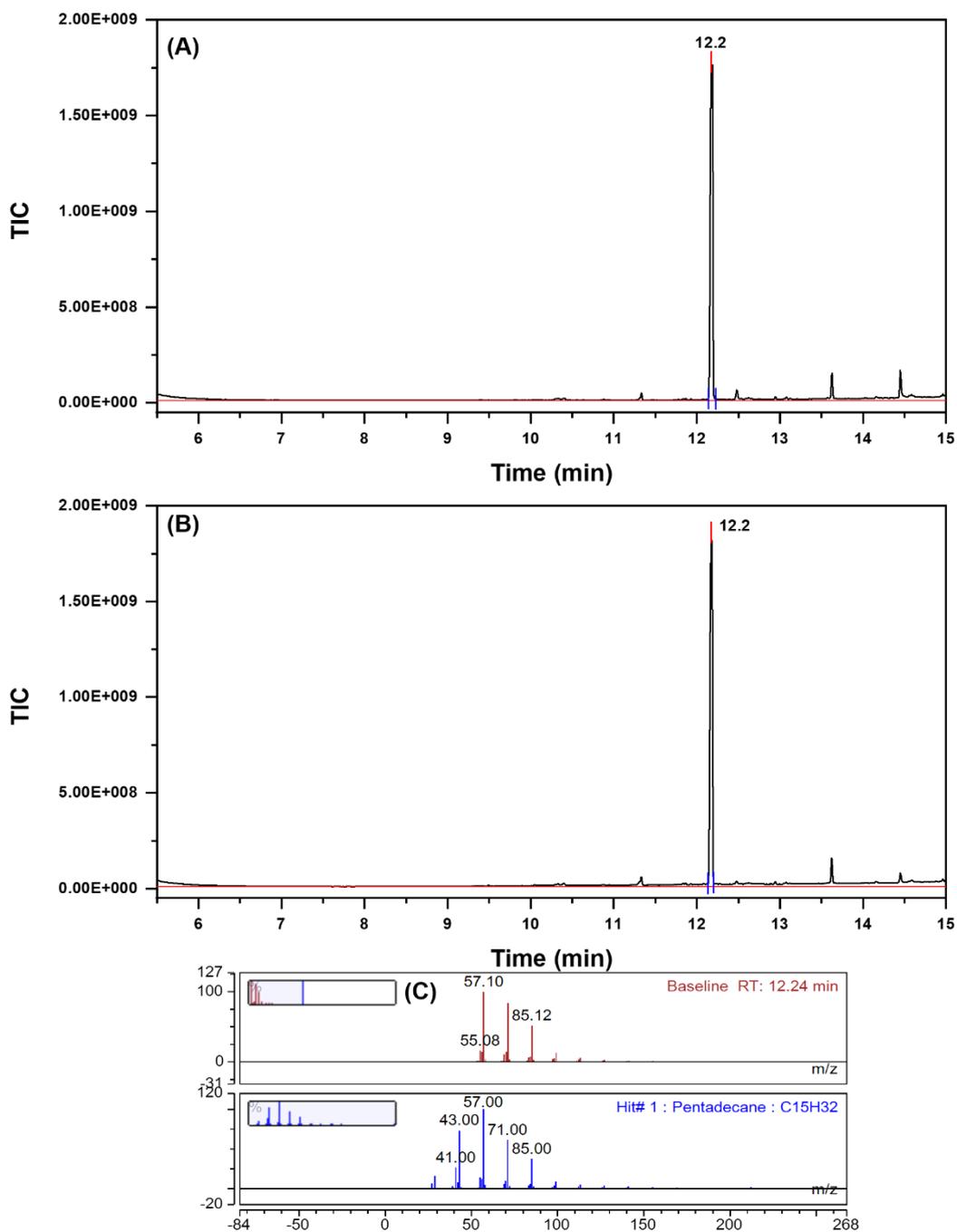


Figure D8. GC chromatogram and mass spectrum for 0.5 mM pentadecane (C₁₄H₃₀). (A) GC chromatogram of pentadecane dissolved in ethyl acetate; (B) GC chromatogram of pentadecane prepared in potassium phosphate buffer and extracted in ethyl acetate; (C) MS spectrum of pentadecane and comparison to standard pentadecane in library.

12.3 Preliminary screening of 7 alkane substrates by Dmg_CYP08 and Dmg_CYP09

The alkane screening reactions conducted by Dmg_CYP08 and Dmg_CYP09 were analysed through GC, while the negative control reaction was also analysed. According to the previously confirmed retention time for each alkane substrates, the signal peaks of alkane substrates were firstly identified for the chromatogram of negative control reactions, Dmg_CYP08 screenings and Dmg_CYP09 screenings. The peak areas were calculated by peak analyser on Origin, which reveals the remaining alkanes in either negative controls, Dmg_CYP08 screening or Dmg_CYP09 screening.

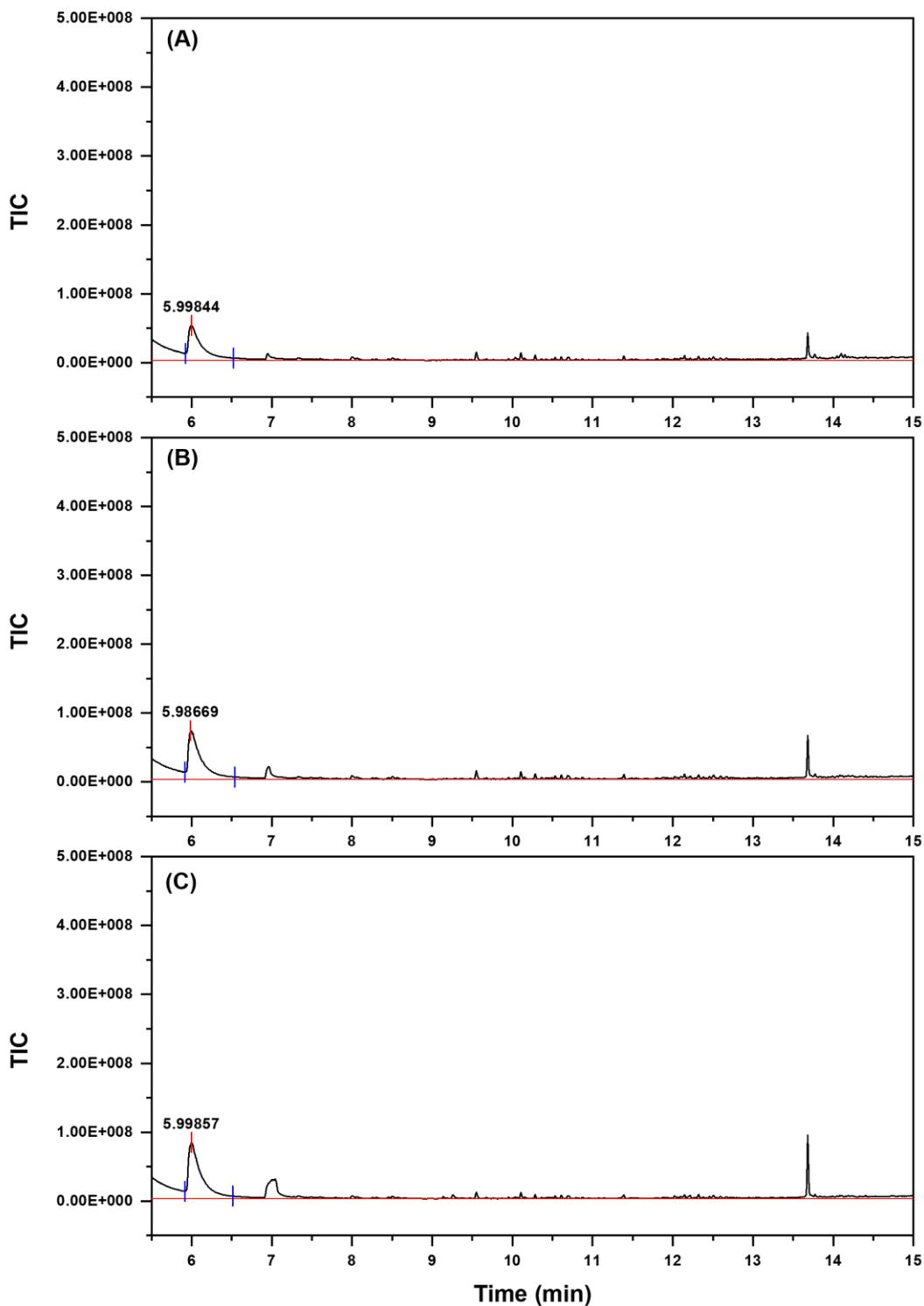


Figure D9. GC chromatogram enzyme assays on nonane (C₉H₂₀). (A) The GC analysis of negative control reaction, which represents the initial concentrations of alkanes. (B) GC analysis of Dmg_CYP08 screening; (C) GC analysis of Dmg_CYP09 screening.

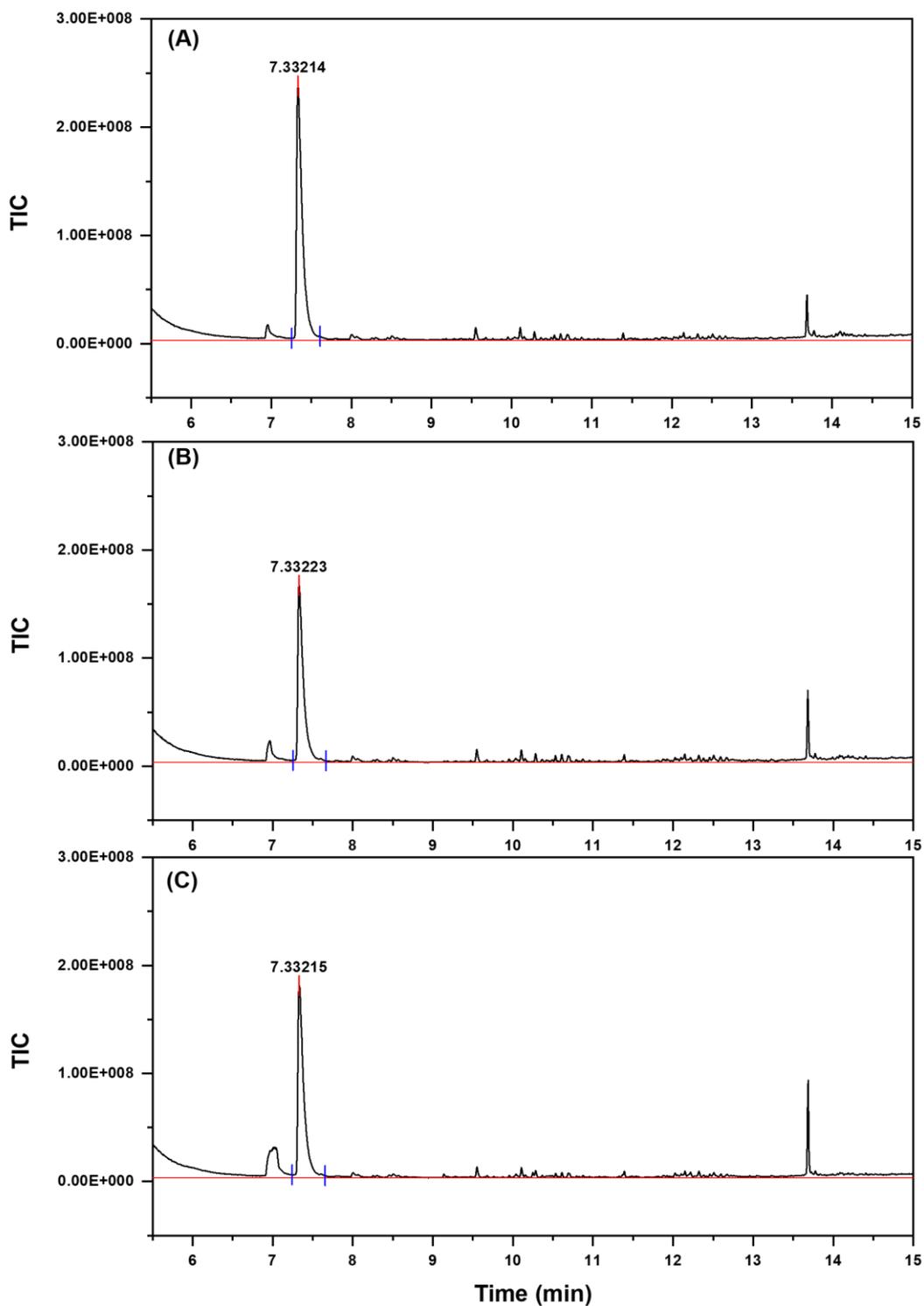


Figure D10. GC chromatogram enzyme assays on decane ($C_{10}H_{22}$). (A) The GC analysis of negative control reaction, which represents the initial concentrations of alkanes. (B) GC analysis of Dmg_CYP08 screening; (C) GC analysis of Dmg_CYP09 screening.

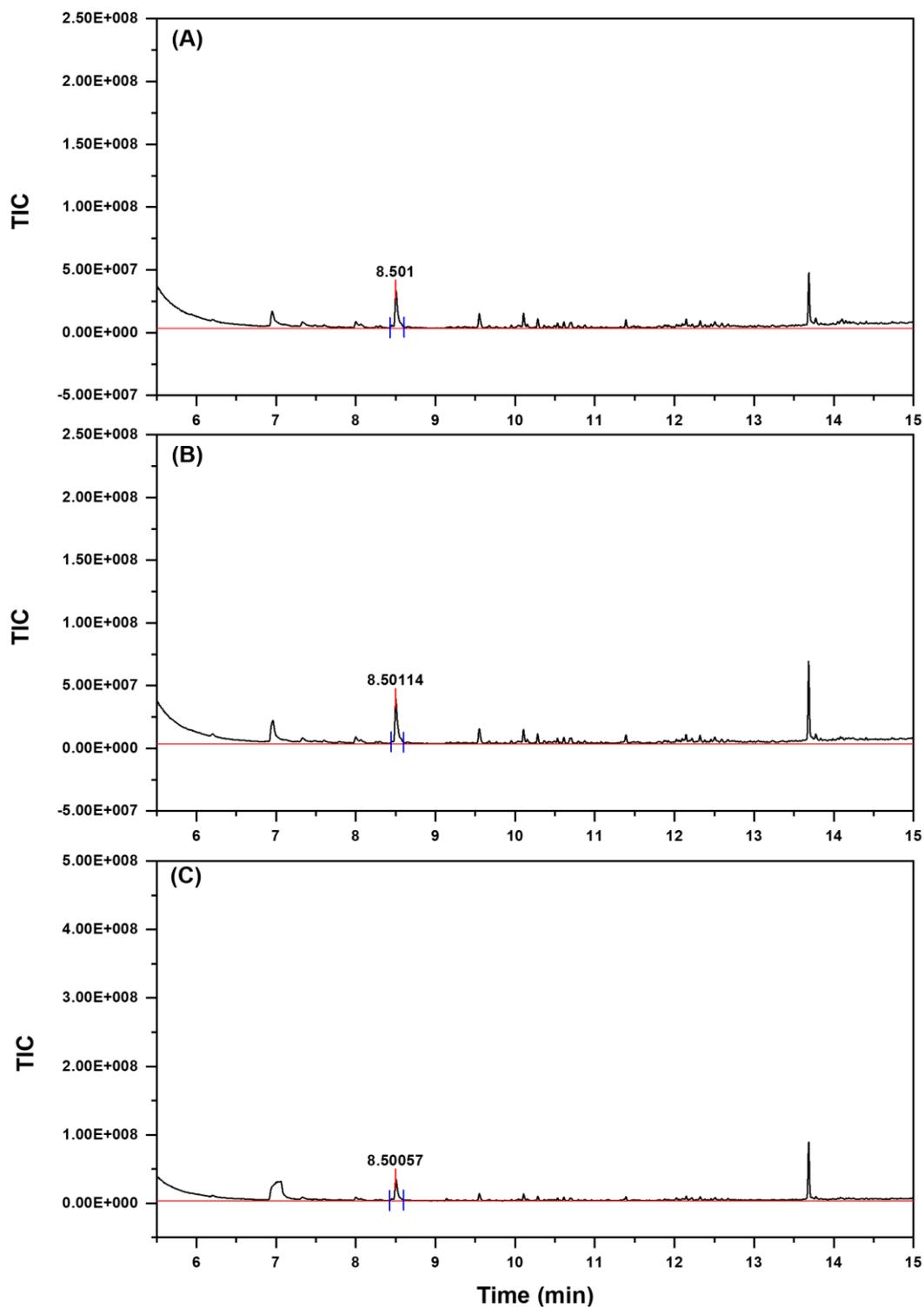


Figure D11. GC chromatogram enzyme assays on undecane ($C_{11}H_{24}$). (A) The GC analysis of negative control reaction, which represents the initial concentrations of alkanes. (B) GC analysis of Dmg_CYP08 screening; (C) GC analysis of Dmg_CYP09 screening.

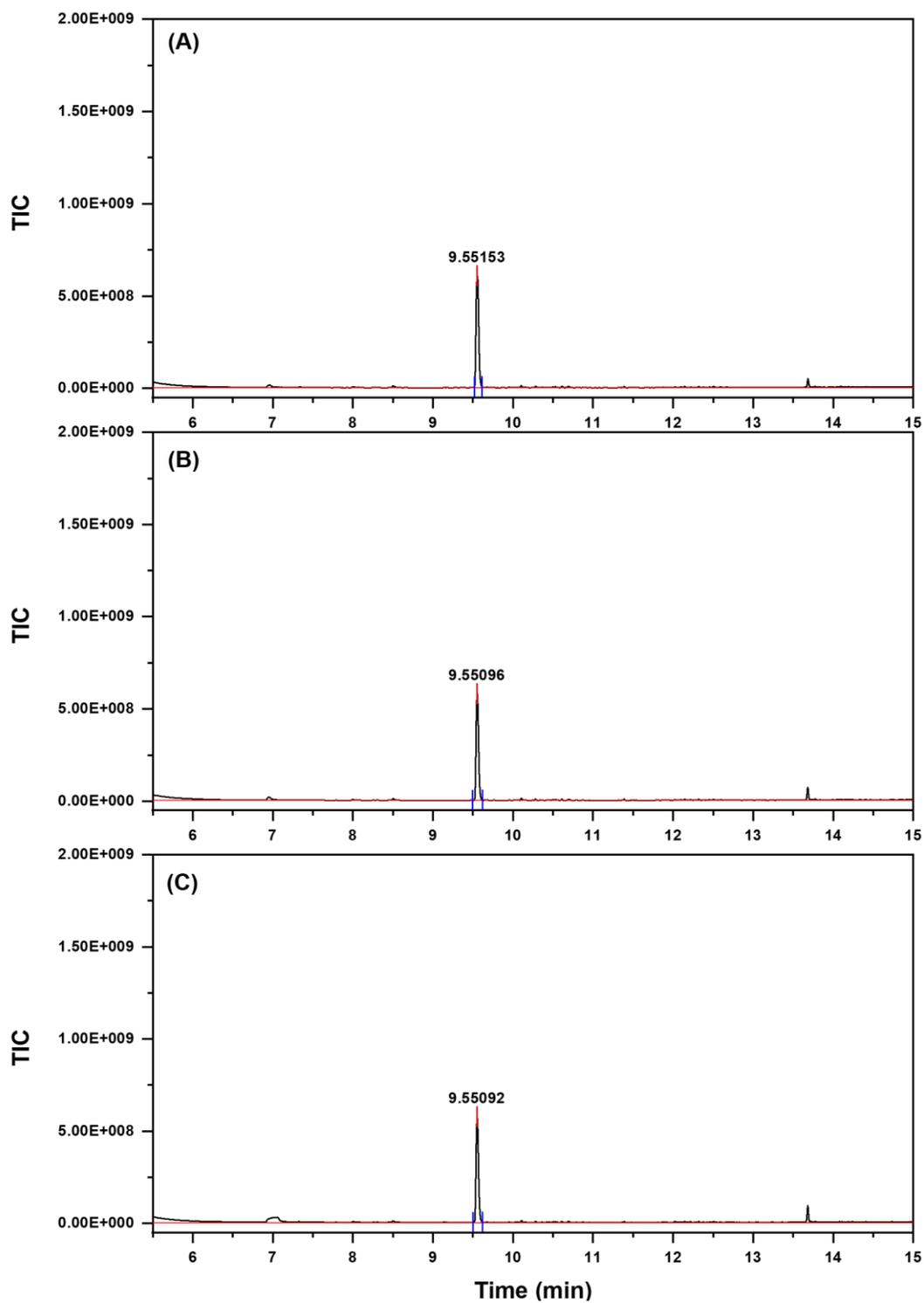


Figure D12. GC chromatogram enzyme assays on dodecane ($C_{12}H_{26}$). (A) The GC analysis of negative control reaction, which represents the initial concentrations of alkanes. (B) GC analysis of Dmg_CYP08 screening; (C) GC analysis of Dmg_CYP09 screening.

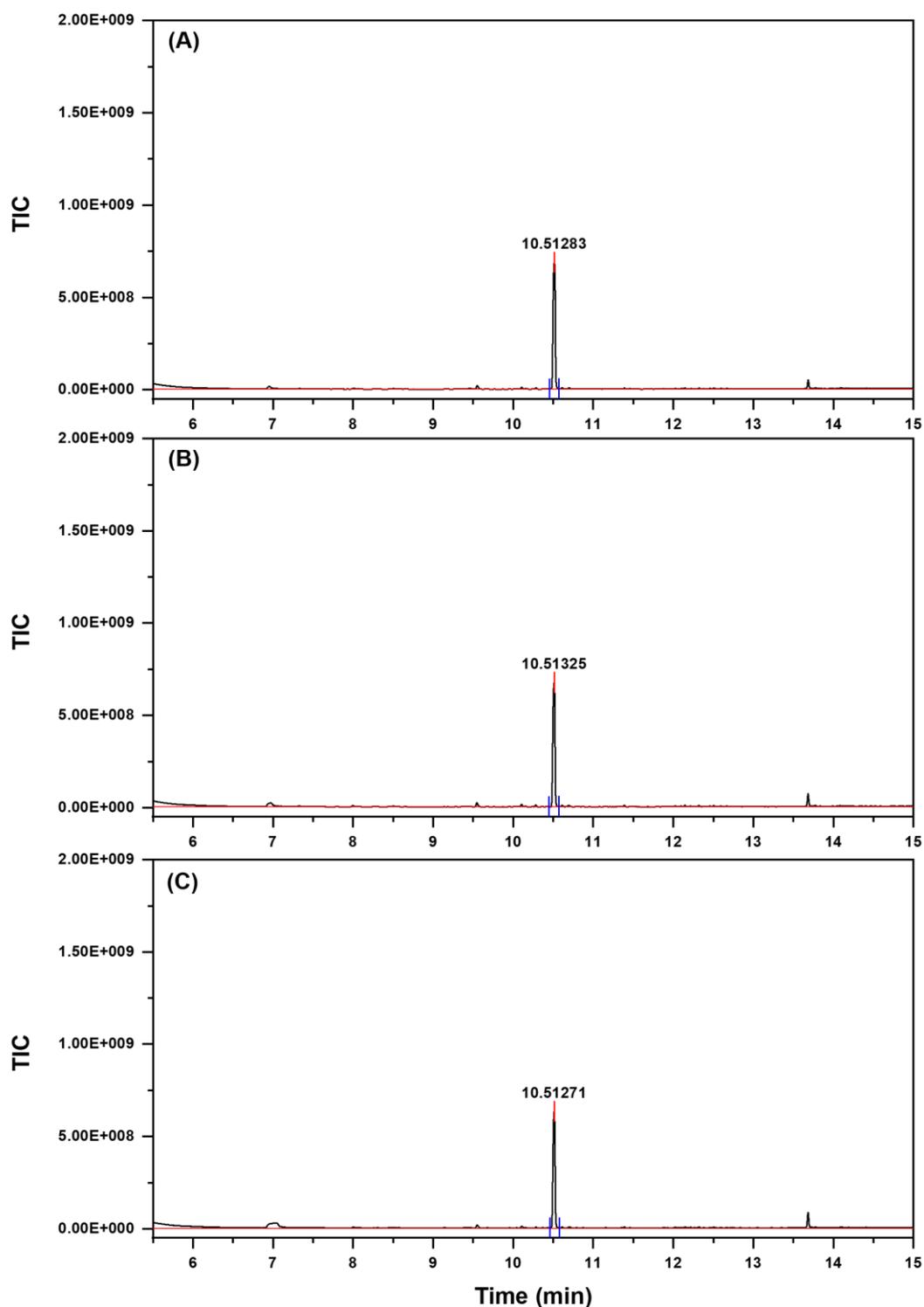


Figure D13. GC chromatogram enzyme assays on tridecane ($C_{13}H_{28}$). (A) The GC analysis of negative control reaction, which represents the initial concentrations of alkanes. (B) GC analysis of Dmg_CYP08 screening; (C) GC analysis of Dmg_CYP09 screening.

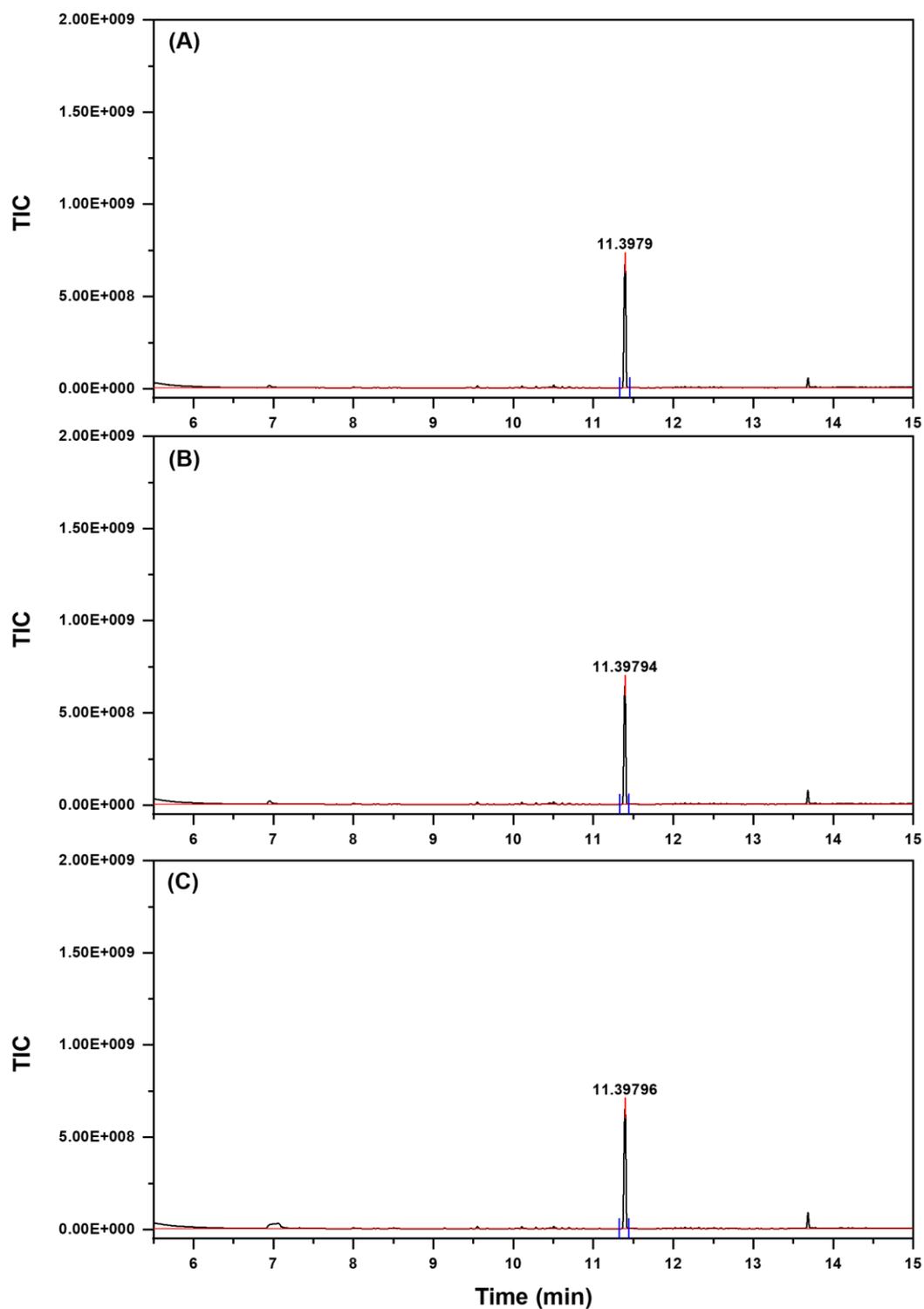


Figure D14. GC chromatogram enzyme assays on tetradecane ($C_{14}H_{30}$). (A) The GC analysis of negative control reaction, which represents the initial concentrations of alkanes. (B) GC analysis of Dmg_CYP08 screening; (C) GC analysis of Dmg_CYP09 screening.

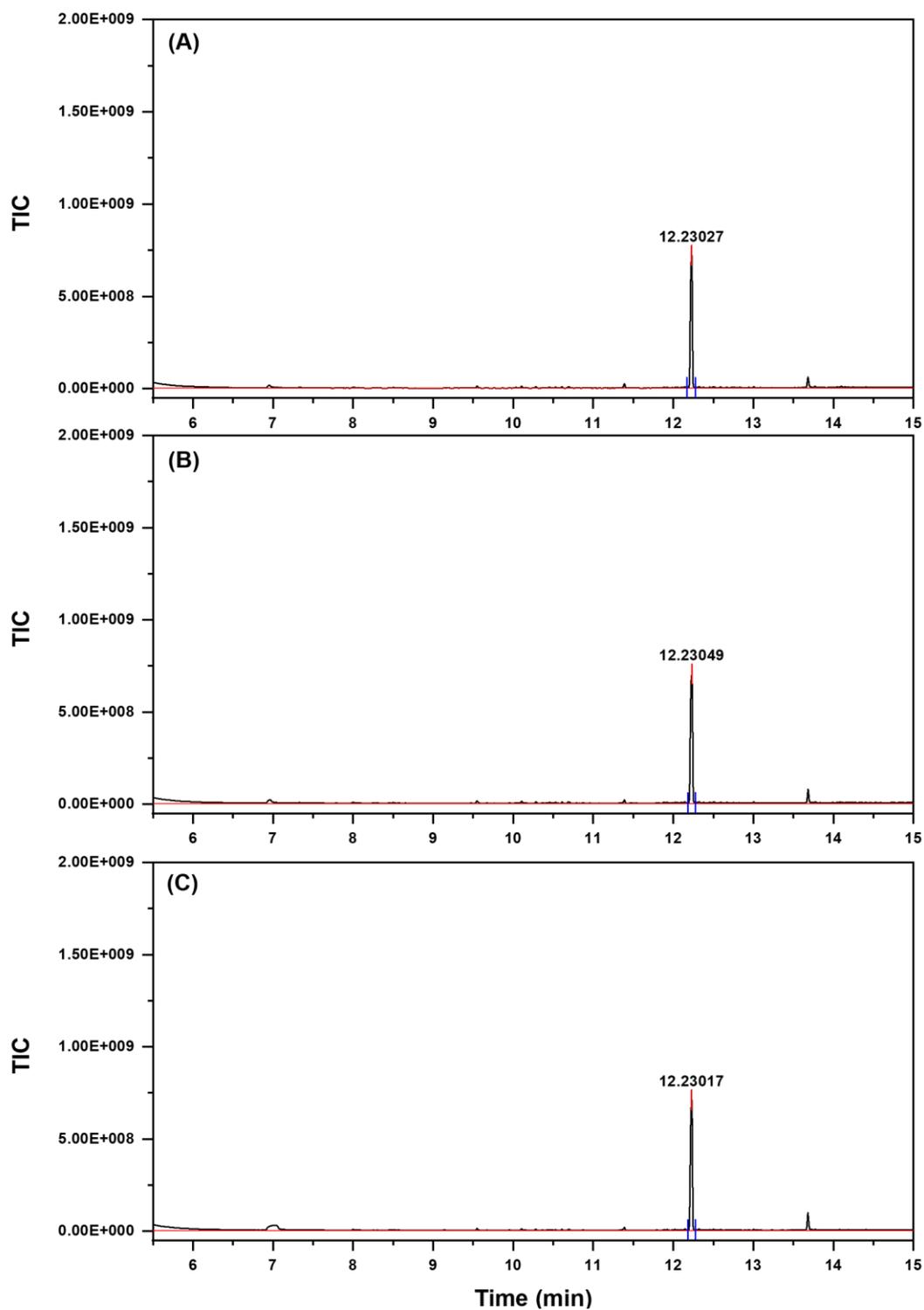


Figure D15. GC chromatogram enzyme assays on pentadecane (C₁₅H₃₂). (A) The GC analysis of negative control reaction, which represents the initial concentrations of alkanes. (B) GC analysis of Dmg_CYP08 screening; (C) GC analysis of Dmg_CYP09 screening.

13 Bibliography

1. Robinson PK. Enzymes: principles and biotechnological applications. *Essays Biochem.* 2015;59:1-41.
2. Webb EC. Nomenclature Committee of the International Union of Biochemistry and Molecular Biology on the Nomenclature and Classification of Enzymes: Academic Press; 1992.
3. Hauer B. Embracing Nature's Catalysts: A Viewpoint on the Future of Biocatalysis. *ACS Catalysis.* 2020;8418-27.
4. Thatoi H, Mohapatra PKD, Mohapatra S, Mondal KC. *Microbial Fermentation and Enzyme Technology*: CRC Press; 2020.
5. Gundersen MT, Tufvesson Pr, Rackham EJ, Lloyd RC, Woodley JM. A rapid selection procedure for simple commercial implementation of ω -transaminase reactions. *Organic Process Research & Development.* 2016;20(3):602-8.
6. Clouthier CM, Pelletier JN. Expanding the organic toolbox: a guide to integrating biocatalysis in synthesis. *Chem Soc Rev.* 2012;41(4):1585-605.
7. Sheldon RA, Brady D. The limits to biocatalysis: pushing the envelope. *Chem Commun (Camb).* 2018;54(48):6088-104.
8. Ping L, Chung DS, Bouffard J, Lee S-g. Transition metal-catalyzed site- and regio-divergent C–H bond functionalization. *Chemical Society Reviews.* 2017;46(14):4299-328.
9. Patrick K. Enzyme technology improves efficiency, cost, safety of stickies removal program. *Paper Age.* 2004;120:22-5.
10. Kuhn D, Kholiq MA, Heinzle E, Bühler B, Schmid A. Intensification and economic and ecological assessment of a biocatalytic oxyfunctionalization process. *Green Chemistry.* 2010;12(5):815-27.
11. Jegannathan KR, Nielsen PH. Environmental assessment of enzyme use in industrial production—a literature review. *Journal of cleaner production.* 2013;42:228-40.
12. Tao J, Xu JH. Biocatalysis in development of green pharmaceutical processes. *Curr Opin Chem Biol.* 2009;13(1):43-50.
13. Lu Y, Yeung N, Sieracki N, Marshall NM. Design of functional metalloproteins. *Nature.* 2009;460(7257):855-62.

14. Poulos TL. Heme enzyme structure and function. *Chem Rev.* 2014;114(7):3919-62.
15. Omura T, Sato R. The Carbon Monoxide-binding pigment of Liver Microsomes. *The Journal of Biological Chemistry.* 1964;239(7):2370-8.
16. Meunier B, de Visser SP, Shaik S. Mechanism of oxidation reactions catalyzed by cytochrome p450 enzymes. *Chem Rev.* 2004;104(9):3947-80.
17. Sligar SG. Coupling of Spin, Substrate, and Redox Equilibria in Cytochrome P450. *Biochemistry.* 1976;15(24):5399-406.
18. Sabat J, Stuehr DJ, Yeh SR, Rousseau DL. Characterization of the proximal ligand in the P420 form of inducible nitric oxide synthase. *J Am Chem Soc.* 2009;131(34):12186-92.
19. Perera R, Sono M, Sigman JA, Pfister TD, Lu Y, Dawson JH. Neutral thiol as a proximal ligand to ferrous heme iron: implications for heme proteins that lose cysteine thiolate ligation on reduction. *Proc Natl Acad Sci U S A.* 2003;100(7):3641-6.
20. Sun Y, Zeng W, Benabbas A, Ye X, Denisov I, Sligar SG, et al. Investigations of heme ligation and ligand switching in cytochromes p450 and p420. *Biochemistry.* 2013;52(34):5941-51.
21. Klingenberg M. Pigments of rat liver microsomes. *Arch Biochem Biophys.* 1958;75(2):376-86.
22. Bernhardt R. Cytochromes P450 as versatile biocatalysts. *J Biotechnol.* 2006;124(1):128-45.
23. Jung ST, Lauchli R, Arnold FH. Cytochrome P450: taming a wild type enzyme. *Curr Opin Biotechnol.* 2011;22(6):809-17.
24. Dong J, Fernández - Fueyo E, Hollmann F, Paul CE, Pesic M, Schmidt S, et al. Biocatalytic oxidation reactions: A chemist's perspective. *Angewandte Chemie International Edition.* 2018;57(30):9238-61.
25. Cook D, Finnigan J, Cook K, Black G, Charnock S. Cytochromes P450: history, classes, catalytic mechanism, and industrial Application. *Advances in protein chemistry and structural biology.* 105: Elsevier; 2016. p. 105-26.
26. McMahon RE, Culp HW, Occolowitz JC. Hepatic microsomal N-dealkylation reaction. Molecular oxygen as the source of the oxygen atom. *Journal of the American Chemical Society.* 1969;91(12):3389-90.

27. Shea JP, Valentine GL, Nelson SD. Source of oxygen in cytochrome P-450 catalyzed carbinolamine formation. *Biochemical and biophysical research communications*. 1982;109(1):231-5.
28. Kedderis GL, Dwyer LA, Rickert DE, Hollenberg PF. Source of the oxygen atom in the product of cytochrome P-450-catalyzed N-demethylation reactions. *Molecular pharmacology*. 1983;23(3):758-60.
29. WATABE T, ISOBE M, YOSHIKAWA K, TAKABATAKE E. Studies on metabolism and toxicity of styrene. I. Biotransformation of styrene to styrene glycol via styrene oxide by rat liver microsomes. *Journal of Pharmacobiodynamics*. 1978;1(2):98-104.
30. Ortiz de Montellano P, Mangold B, Wheeler C, Kunze K, Reich N. Stereochemistry of cytochrome P-450-catalyzed epoxidation and prosthetic heme alkylation. *J Biol Chem*. 1983;258(420):4213.
31. Guengerich FP. Common and uncommon cytochrome P450 reactions related to metabolism and chemical toxicity. *Chem Res Toxicol*. 2001;14(6):611-50.
32. Guengerich FP. Mechanisms of cytochrome P450 substrate oxidation: MiniReview. *J Biochem Mol Toxicol*. 2007;21(4):163-8.
33. Isin EM, Guengerich FP. Complex reactions catalyzed by cytochrome P450 enzymes. *Biochim Biophys Acta*. 2007;1770(3):314-29.
34. Guengerich FP, Munro AW. Unusual cytochrome p450 enzymes and reactions. *J Biol Chem*. 2013;288(24):17065-73.
35. Sono M, Roach MP, Coulter ED, Dawson JH. Heme-containing oxygenases. *Chemical reviews*. 1996;96(7):2841-88.
36. Guengerich FP. Cytochrome p450 and chemical toxicology. *Chem Res Toxicol*. 2008;21(1):70-83.
37. Stassi D, Donadio S, Staver MJ, Katz L. Identification of a *Saccharopolyspora erythraea* gene required for the final hydroxylation step in erythromycin biosynthesis. *J Bacteriol*. 1993;175(1):182-9.
38. Rude MA, Baron TS, Brubaker S, Alibhai M, Del Cardayre SB, Schirmer A. Terminal olefin (1-alkene) biosynthesis by a novel p450 fatty acid decarboxylase from *Jeotgalicoccus* species. *Appl Environ Microbiol*. 2011;77(5):1718-27.

39. Wlodarczyk A, Gnanasekaran T, Nielsen AZ, Zulu NN, Mellor SB, Luckner M, et al. Metabolic engineering of light-driven cytochrome P450 dependent pathways into *Synechocystis* sp PCC 6803. *Metabolic Engineering*. 2016;33:1-11.
40. Paddon CJ, Westfall PJ, Pitera DJ, Benjamin K, Fisher K, McPhee D, et al. High-level semi-synthetic production of the potent antimalarial artemisinin. *Nature*. 2013;496(7446):528-32.
41. Galanie S, Thodey K, Trenchard IJ, Filsinger Interrante M, Smolke CD. Complete biosynthesis of opioids in yeast. *Science*. 2015;349(6252):1095-100.
42. Kelly SL, Kelly DE. Microbial cytochromes P450: biodiversity and biotechnology. Where do cytochromes P450 come from, what do they do and what can they do for us? *Philos Trans R Soc Lond B Biol Sci*. 2013;368(1612):20120476.
43. Tijet N, Helvig C, Feyereisen R. The cytochrome P450 gene superfamily in *Drosophila melanogaster*: annotation, intron-exon organization and phylogeny. *Gene*. 2001;262(1-2):189-98.
44. Arabidopsis Genome I. Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature*. 2000;408(6814):796-815.
45. Nelson DR. Introductory remarks on human CYPs. *Drug Metab Rev*. 2002;34(1-2):1-5.
46. Lamb DC, Ikeda H, Nelson DR, Ishikawa J, Skaug T, Jackson C, et al. Cytochrome P450 complement (CYPome) of the avermectin-producer *Streptomyces avermitilis* and comparison to that of *Streptomyces coelicolor* A3(2). *Biochemical and Biophysical Research Communications*. 2003;307(3):610-9.
47. Sawayama AM, Chen MM, Kulanthaivel P, Kuo MS, Hemmerle H, Arnold FH. A panel of cytochrome P450 BM3 variants to produce drug metabolites and diversify lead compounds. *Chemistry*. 2009;15(43):11723-9.
48. Belcher J, McLean KJ, Matthews S, Woodward LS, Fisher K, Rigby SE, et al. Structure and biochemical properties of the alkene producing cytochrome P450 OleTJE (CYP152L1) from the *Jeotgalicoccus* sp. 8456 bacterium. *J Biol Chem*. 2014;289(10):6535-50.

49. O'Keefe DP, Harder PA. Occurrence and biological function of cytochrome P450 monooxygenases in the actinomycetes. *Mol Microbiol.* 1991;5(9):2099-105.
50. Janke D, Fritsche W. Nature and significance of microbial cometabolism of xenobiotics. *J Basic Microbiol.* 1985;25(9):603-19.
51. Bollag J-M, Liu S-Y. Biological transformation processes of pesticides. *Pesticides in the soil environment: processes, impacts and modeling.* 1990:169-211.
52. Ichinose H. Molecular and functional diversity of fungal cytochrome P450s. *Biol Pharm Bull.* 2012;35(6):833-7.
53. Alcazar-Fuoli L, Mellado E. Ergosterol biosynthesis in *Aspergillus fumigatus*: its relevance as an antifungal target and role in antifungal drug resistance. *Frontiers in Microbiology.* 2013;3:439.
54. Schuler MA. The role of cytochrome P450 monooxygenases in plant-insect interactions. *Plant Physiol.* 1996;112(4):1411-9.
55. Teoh KH, Polichuk DR, Reed DW, Nowak G, Covello PS. *Artemisia annua* L. (Asteraceae) trichome-specific cDNAs reveal CYP71AV1, a cytochrome P450 with a key role in the biosynthesis of the antimalarial sesquiterpene lactone artemisinin. *FEBS Lett.* 2006;580(5):1411-6.
56. Werck-Reichhart D, Feyereisen R. Cytochromes P450: a success story. *Genome Biol.* 2000;1(6):REVIEWS3003.
57. Baylon JL, Lenov IL, Sligar SG, Tajkhorshid E. Characterizing the membrane-bound state of cytochrome P450 3A4: structure, depth of insertion, and orientation. *J Am Chem Soc.* 2013;135(23):8542-51.
58. Nebert DW, Wikvall K, Miller WL. Human cytochromes P450 in health and disease. *Philos Trans R Soc Lond B Biol Sci.* 2013;368(1612):20120431.
59. Dean L. Codeine Therapy and CYP2D6 Genotype. In: Pratt V, McLeod H, Rubinstein W, Dean L, Kattman B, Malheiro A, editors. *Medical Genetics Summaries.* Bethesda (MD): National Center for Biotechnology Information (US); 2012.
60. Zhou SF, Liu JP, Chowbay B. Polymorphism of human cytochrome P450 enzymes and its clinical impact. *Drug Metab Rev.* 2009;41(2):89-295.

61. Hedegaard J, Gunsalus IC. Mixed function oxidation. IV. An induced methylene hydroxylase in camphor oxidation. *J Biol Chem.* 1965;240(10):4038-43.
62. Poulos TL, Finzel BC, Howard AJ. High-resolution crystal structure of cytochrome P450cam. *J Mol Biol.* 1987;195(3):687-700.
63. Mayhew MP, Reipa V, Holden MJ, Vilker VL. Improving the cytochrome P450 enzyme system for electrode-driven biocatalysis of styrene epoxidation. *Biotechnol Prog.* 2000;16(4):610-6.
64. Andrew C. Cytochrome P-450 cam monooxygenase can be redesigned to catalyse the regioselective aromatic hydroxylation of diphenylmethane. *Journal of the Chemical Society, Chemical Communications.* 1994(24):2761-2.
65. Heimbrook DC, Murray RI, Egeberg KD, Sligar SG, Nee MW, Bruice TC. Demethylation of N,N-Dimethylaniline and P-Cyano-N,N-Dimethylaniline and Their N-Oxides by Cytochromes P450Im2 and P450cam. *Journal of the American Chemical Society.* 1984;106(5):1514-5.
66. Gunsalus IC, Pederson TC, Sligar SG. Oxygenase-catalyzed biological hydroxylations. *Annu Rev Biochem.* 1975;44(1):377-407.
67. Katagiri M, Ganguli BN, Gunsalus IC. A soluble cytochrome P-450 functional in methylene hydroxylation. *J Biol Chem.* 1968;243(12):3543-6.
68. Denisov IG, Makris TM, Sligar SG, Schlichting I. Structure and chemistry of cytochrome P450. *Chem Rev.* 2005;105(6):2253-77.
69. Groves JT, McClusky GA. Aliphatic hydroxylation by highly purified liver microsomal cytochrome P-450. Evidence for a carbon radical intermediate. *Biochem Biophys Res Commun.* 1978;81(1):154-60.
70. Schlichting I, Berendzen J, Chu K, Stock AM, Maves SA, Benson DE, et al. The catalytic pathway of cytochrome p450cam at atomic resolution. *Science.* 2000;287(5458):1615-22.
71. Williams-Smith DL, Cammack R. Oxidation-reduction potentials of cytochromes P-450 and ferredoxin in the bovine adrenal: Their modification by substrates and inhibitors. *Biochimica et Biophysica Acta (BBA)-General Subjects.* 1977;499(3):432-42.

72. Fantuzzi A, Fairhead M, Gilardi G. Direct electrochemistry of immobilized human cytochrome P450 2E1. *Journal of the American Chemical Society*. 2004;126(16):5040-1.
73. Nasset MJ, Shokhirev NV, Enemark PD, Jacobson SE, Walker FA. Models of the cytochromes. Redox properties and thermodynamic stabilities of complexes of "hindered" iron (III) and iron (II) tetraphenylporphyrinates with substituted pyridines and imidazoles. *Inorganic Chemistry*. 1996;35(18):5188-200.
74. Safo MK, Nasset MJ, Walker FA, Debrunner PG, Scheidt WR. Models of the Cytochromes. Axial Ligand Orientation and Complex Stability in Iron (II) Porphyrinates: The Case of the Noninteracting $d\pi$ Orbitals. *Journal of the American Chemical Society*. 1997;119(40):9438-48.
75. Momenteau M, Reed CA. Synthetic heme-dioxygen complexes. *Chemical reviews*. 1994;94(3):659-98.
76. Li QS, Ogawa J, Schmid RD, Shimizu S. Indole hydroxylation by bacterial cytochrome P450 BM-3 and modulation of activity by cumene hydroperoxide. *Biosci Biotechnol Biochem*. 2005;69(2):293-300.
77. Eyer CS, Backes WL. Relationship between the Rate of Reductase-Cytochrome P450 Complex Formation and the Rate of First Electron Transfer. 1992.
78. Brenner S, Hay S, Munro AW, Scrutton NS. Inter-flavin electron transfer in cytochrome P450 reductase - effects of solvent and pH identify hidden complexity in mechanism. *FEBS J*. 2008;275(18):4540-57.
79. Guengerich FP. Cytochrome P450 enzymes in the generation of commercial products. *Nat Rev Drug Discov*. 2002;1(5):359-66.
80. Nelson DR. *Cytochrome P450 Nomenclature*, 2004. New Jersey: Humana Press; 2006. p. 1-10.
81. Hannemann F, Bichet A, Ewen KM, Bernhardt R. Cytochrome P450 systems--biological variations of electron transport chains. *Biochim Biophys Acta*. 2007;1770(3):330-44.
82. Hawkes DB, Adams GW, Burlingame AL, de Montellano PRO, De Voss JJ. Cytochrome P450cin (CYP176A), isolation, expression, and characterization. *Journal of Biological Chemistry*. 2002;277(31):27725-32.

83. Nishida CR, Ortiz de Montellano PR. Thermophilic cytochrome P450 enzymes. *Biochem Biophys Res Commun.* 2005;338(1):437-45.
84. Jackson CJ, Lamb DC, Marczylo TH, Warrilow AG, Manning NJ, Lowe DJ, et al. A Novel Sterol 14 α -Demethylase/Ferredoxin Fusion Protein (MCCYP51FX) from *Methylococcus capsulatus* Represents a New Class of the Cytochrome P450 Superfamily. *Journal of Biological Chemistry.* 2002;277(49):46959-65.
85. Miura Y, Fulco AJ. (ω -2) Hydroxylation of Fatty Acids by a Soluble System from *Bacillus megaterium*. *Journal of Biological Chemistry.* 1974;249(6):1880-8.
86. Fulco AJ. P450BM-3 and other inducible bacterial P450 cytochromes: biochemistry and regulation. *Annu Rev Pharmacol Toxicol.* 1991;31(1):177-203.
87. Roberts GA, Celik A, Hunter DJB, Ost TWB, White JH, Chapman SK, et al. A self-sufficient cytochrome P450 with a primary structural organization that includes a flavin domain and a [2Fe-2S] redox center. *Journal of Biological Chemistry.* 2003;278(49):48914-20.
88. Atkins WM, Sligar SG. The Roles of Active-Site Hydrogen-Bonding in Cytochrome P-450cam as Revealed by Site-Directed Mutagenesis. *Journal of Biological Chemistry.* 1988;263(35):18842-9.
89. Hanukoglu I. Electron transfer proteins of cytochrome P450 systems. *Advances in molecular and cell biology.* 14: Elsevier; 1996. p. 29-56.
90. Porter TD, Kasper CB. NADPH-cytochrome P-450 oxidoreductase: flavin mononucleotide and flavin adenine dinucleotide domains evolved from different flavoproteins. *Biochemistry.* 1986;25(7):1682-7.
91. Serizawa N, Matsuoka T. A two component-type cytochrome P-450 monooxygenase system in a prokaryote that catalyzes hydroxylation of ML-236B to pravastatin, a tissue-selective inhibitor of 3-hydroxy-3-methylglutaryl coenzyme A reductase. *Biochimica et Biophysica Acta (BBA)-Lipids and Lipid Metabolism.* 1991;1084(1):35-40.
92. Correll CC, Batie CJ, Ballou DP, Ludwig ML. Phthalate dioxygenase reductase: a modular structure for electron transfer from pyridine nucleotides to [2Fe-2S]. *Science.* 1992;258(5088):1604-10.
93. Sakurai K, Shimada H, Hayashi T, Tsukihara T. Substrate binding induces structural changes in cytochrome P450cam. *Acta Crystallogr Sect F Struct Biol Cryst Commun.* 2009;65(Pt 2):80-3.

94. Lee YT, Wilson RF, Rupniewski I, Goodin DB. P450cam visits an open conformation in the absence of substrate. *Biochemistry*. 2010;49(16):3412-9.
95. Mansoorabadi SO, Thibodeaux CJ, Liu HW. The diverse roles of flavin coenzymes-nature's most versatile thespians. *J Org Chem*. 2007;72(17):6329-42.
96. Aliverti A, Curti B, Vanoni MA. Identifying and quantitating FAD and FMN in simple and in iron-sulfur-containing flavoproteins. *Flavoprotein protocols: Springer*; 1999. p. 9-23.
97. Noodleman L, Norman Jr JG, Osborne JH, Aizman A, Case DA. Models for ferredoxins: electronic structures of iron-sulfur clusters with one, two, and four iron atoms. *Journal of the American Chemical Society*. 1985;107(12):3418-26.
98. Tripathi S, Li H, Poulos TL. Structural basis for effector control and redox partner recognition in cytochrome P450. *Science*. 2013;340(6137):1227-30.
99. Sevrioukova IF, Li H, Zhang H, Peterson JA, Poulos TL. Structure of a cytochrome P450–redox partner electron-transfer complex. *Proceedings of the National Academy of Sciences*. 1999;96(5):1863-8.
100. Daiber A, Shoun H, Ullrich V. Nitric oxide reductase (P450nor) from *Fusarium oxysporum*. *J Inorg Biochem*. 2005;99(1):185-93.
101. Jenkins CM, Waterman MR. NADPH– flavodoxin reductase and flavodoxin from *Escherichia coli*: characteristics as a soluble microsomal P450 reductase. *Biochemistry*. 1998;37(17):6106-13.
102. Hall DA, Kooi CWV, Stasik CN, Stevens SY, Zuiderweg ERP, Matthews RG. Mapping the interactions between flavodoxin and its physiological partners flavodoxin reductase and cobalamin-dependent methionine synthase. *P Natl Acad Sci USA*. 2001;98(17):9521-6.
103. Sevrioukova IF, Poulos TL. Structural biology of redox partner interactions in P450cam monooxygenase: a fresh look at an old system. *Archives of biochemistry and biophysics*. 2011;507(1):66-74.
104. Leisch H, Shi R, Grosse S, Morley K, Bergeron H, Cygler M, et al. Cloning, baeyer-villiger biooxidations, and structures of the camphor pathway 2-oxo- δ^3 -4, 5, 5-trimethylcyclopentenylicetyl-coenzyme a monooxygenase of *pseudomonas putida* atcc 17453. *Appl Environ Microbiol*. 2012;78(7):2200-12.

105. Iwaki H, Grosse S, Bergeron H, Leisch H, Morley K, Hasegawa Y, et al. Camphor pathway redux: functional recombinant expression of 2, 5- and 3, 6-diketocamphane monooxygenases of *Pseudomonas putida* ATCC 17453 with their cognate flavin reductase catalyzing Baeyer-Villiger reactions. *Applied and environmental microbiology*. 2013;79(10):3282-93.
106. Romesser JA, O'Keefe DP. Induction of cytochrome P-450-dependent sulfonylurea metabolism in *Streptomyces griseolus*. *Biochem Biophys Res Commun*. 1986;140(2):650-9.
107. Ohnishi Y, Ishikawa J, Hara H, Suzuki H, Ikenoya M, Ikeda H, et al. Genome sequence of the streptomycin-producing microorganism *Streptomyces griseus* IFO 13350. *J Bacteriol*. 2008;190(11):4050-60.
108. Lamb DC, Skaug T, Song HL, Jackson CJ, Podust LM, Waterman MR, et al. The cytochrome P450 complement (CYPome) of *Streptomyces coelicolor* A3(2). *J Biol Chem*. 2002;277(27):24000-5.
109. Ikeda H, Ishikawa J, Hanamoto A, Shinose M, Kikuchi H, Shiba T, et al. Complete genome sequence and comparative analysis of the industrial microorganism *Streptomyces avermitilis*. *Nat Biotechnol*. 2003;21(5):526-31.
110. Oliynyk M, Samborsky M, Lester JB, Mironenko T, Scott N, Dickens S, et al. Complete genome sequence of the erythromycin-producing bacterium *Saccharopolyspora erythraea* NRRL23338. *Nat Biotechnol*. 2007;25(4):447-53.
111. Guengerich FP, Martin MV, Sohl CD, Cheng Q. Measurement of cytochrome P450 and NADPH-cytochrome P450 reductase. *Nat Protoc*. 2009;4(9):1245-51.
112. Roerig DL, Mascaro L, Jr., Aust SD. Microsomal electron transport: tetrazolium reduction by rat liver microsomal NADPH-cytochrome c reductase. *Arch Biochem Biophys*. 1972;153(2):475-9.
113. Phillips AH, L RG. Hepatic Triphosphopyridine Nucleotide-Cytochrome c Reductase: Isolation, Characterization, and Kinetic Studies*. 1962.
114. Massey V. The microestimation of succinate and the extinction coefficient of cytochrome c. *Biochim Biophys Acta*. 1959;34:255-6.
115. Kelly SL, Lamb DC, Jackson CJ, Warrilow AG, Kelly DE. The biodiversity of microbial cytochromes P450. 2003.

116. Bernhardt R, Urlacher VB. Cytochromes P450 as promising catalysts for biotechnological application: chances and limitations. *Appl Microbiol Biotechnol.* 2014;98(14):6185-203.
117. O'Reilly E, Kohler V, Flitsch SL, Turner NJ. Cytochromes P450 as useful biocatalysts: addressing the limitations. *Chem Commun (Camb).* 2011;47(9):2490-501.
118. Zaparucha A, de Berardinis V, Vaxelaire-Vergne C. Genome mining for enzyme discovery. 2018.
119. Girvan H, Waltham T, Neeli R, Collins H, McLean K, Scrutton N, et al. Flavocytochrome P450 BM3 and the origin of CYP102 fusion species. Portland Press Ltd.; 2006.
120. Matsunaga I, Sumimoto T, Ueda A, Kusunose E, Ichihara K. Fatty acid - specific, regiospecific, and stereospecific hydroxylation by cytochrome P450 (CYP152B1) from *Sphingomonas paucimobilis*: Substrate structure required for α - hydroxylation. *Lipids.* 2000;35(4):365-71.
121. Furuya T, Kino K. Genome mining approach for the discovery of novel cytochrome P450 biocatalysts. *Appl Microbiol Biotechnol.* 2010;86(4):991-1002.
122. Boddupalli SS, Estabrook RW, Peterson JA. Fatty acid monooxygenation by cytochrome P-450BM-3. *J Biol Chem.* 1990;265(8):4233-9.
123. Chowdhary PK, Alemseghed M, Haines DC. Cloning, expression and characterization of a fast self-sufficient P450: CYP102A5 from *Bacillus cereus*. *Arch Biochem Biophys.* 2007;468(1):32-43.
124. Dietrich M, Eiben S, Asta C, Do TA, Pleiss J, Urlacher VB. Cloning, expression and characterisation of CYP102A7, a self-sufficient P450 monooxygenase from *Bacillus licheniformis*. *Appl Microbiol Biotechnol.* 2008;79(6):931-40.
125. Matsunaga I, Ueda A, Fujiwara N, Sumimoto T, Ichihara K. Characterization of the ybdT gene product of *Bacillus subtilis*: novel fatty acid beta-hydroxylating cytochrome P450. *Lipids.* 1999;34(8):841-6.
126. Girhard M, Schuster S, Dietrich M, Durre P, Urlacher VB. Cytochrome P450 monooxygenase from *Clostridium acetobutylicum*: a new alpha-fatty acid hydroxylase. *Biochem Biophys Res Commun.* 2007;362(1):114-9.

127. Romero PA, Arnold FH. Exploring protein fitness landscapes by directed evolution. *Nat Rev Mol Cell Biol.* 2009;10(12):866-76.
128. Rabe KS, Gandubert VJ, Spengler M, Erkelenz M, Niemeyer CM. Engineering and assaying of cytochrome P450 biocatalysts. *Anal Bioanal Chem.* 2008;392(6):1059-73.
129. Bell SG, Chen X, Sowden RJ, Xu F, Williams JN, Wong LL, et al. Molecular recognition in (+)-alpha-pinene oxidation by cytochrome P450cam. *J Am Chem Soc.* 2003;125(3):705-14.
130. Bell SG, Harford-Cross CF, Wong LL. Engineering the CYP101 system for in vivo oxidation of unnatural substrates. *Protein Eng.* 2001;14(10):797-802.
131. Bell SG, Stevenson JA, Boyd HD, Campbell S, Riddle AD, Orton EL, et al. Butane and propane oxidation by engineered cytochrome P450(cam). *Chemical Communications.* 2002(5):490-1.
132. Loida PJ, Sligar SG. Engineering cytochrome P-450cam to increase the stereospecificity and coupling of aliphatic hydroxylation. *Protein Eng.* 1993;6(2):207-12.
133. Meinhold P, Peters MW, Chen MM, Takahashi K, Arnold FH. Direct conversion of ethane to ethanol by engineered cytochrome P450 BM3. *ChemBioChem.* 2005;6(10):1765-8.
134. Coelho PS, Brustad EM, Kannan A, Arnold FH. Olefin Cyclopropanation via Carbene Transfer Catalyzed by Engineered Cytochrome P450 Enzymes. *Science.* 2013;339(6117):307-10.
135. Li QS, Schwaneberg U, Fischer P, Schmid RD. Directed evolution of the fatty - acid hydroxylase P450 BM - 3 into an indole - hydroxylating catalyst. *Chemistry - A European Journal.* 2000;6(9):1531-6.
136. Appel D, Lutz-Wahl S, Fischer P, Schwaneberg U, Schmid RD. A P450 BM-3 mutant hydroxylates alkanes, cycloalkanes, arenes and heteroarenes. *J Biotechnol.* 2001;88(2):167-71.
137. Li QS, Ogawa J, Schmid RD, Shimizu S. Engineering cytochrome P450 BM-3 for oxidation of polycyclic aromatic hydrocarbons. *Appl Environ Microbiol.* 2001;67(12):5735-9.
138. Hlavica P. Assembly of non-natural electron transfer conduits in the cytochrome P450 system: a critical assessment and update of artificial redox

constructs amenable to exploitation in biotechnological areas. *Biotechnol Adv.* 2009;27(2):103-21.

139. Nodate M, Kubota M, Misawa N. Functional expression system for cytochrome P450 genes using the reductase domain of self-sufficient P450RhF from *Rhodococcus* sp NCIMB 9784. *Applied Microbiology and Biotechnology.* 2006;71(4):455-62.

140. Fujita N, Sumisa F, Shindo K, Kabumoto H, Arisawa A, Ikenaga H, et al. Comparison of Two Vectors for Functional Expression of a Bacterial Cytochrome P450 Gene in *Escherichia coli* Using CYP153 Genes. *Biosci Biotech Bioch.* 2009;73(8):1825-30.

141. Cornelissen S, Julsing MK, Volmer J, Riechert O, Schmid A, Buhler B. Whole-cell-based CYP153A6-catalyzed (S)-limonene hydroxylation efficiency depends on host background and profits from monoterpene uptake via AlkL. *Biotechnol Bioeng.* 2013;110(5):1282-92.

142. Janocha S, Bernhardt R. Design and characterization of an efficient CYP105A1-based whole-cell biocatalyst for the conversion of resin acid diterpenoids in permeabilized *Escherichia coli*. *Appl Microbiol Biotechnol.* 2013;97(17):7639-49.

143. Pazmino DET, Winkler M, Glieder A, Fraaije MW. Monooxygenases as biocatalysts: Classification, mechanistic aspects and biotechnological applications. *Journal of Biotechnology.* 2010;146(1-2):9-24.

144. Kieser T, Bibb MJ, Buttner MJ, Chater KF, Hopwood DA. *Practical streptomyces genetics*: John Innes Foundation Norwich; 2000.

145. Schulz S, Schumacher D, Raszkowski D, Girhard M, Urlacher VB. Fusion to Hydrophobin HFBI Improves the Catalytic Performance of a Cytochrome P450 System. *Front Bioeng Biotechnol.* 2016;4:57.

146. Narhi LO, Fulco AJ. Phenobarbital induction of a soluble cytochrome P-450-dependent fatty acid monooxygenase in *Bacillus megaterium*. *J Biol Chem.* 1982;257(5):2147-50.

147. Narhi LO, Fulco AJ. Characterization of a catalytically self-sufficient 119,000-dalton cytochrome P-450 monooxygenase induced by barbiturates in *Bacillus megaterium*. *J Biol Chem.* 1986;261(16):7160-9.

148. Whitehouse CJ, Bell SG, Wong LL. P450(BM3) (CYP102A1): connecting the dots. *Chem Soc Rev.* 2012;41(3):1218-60.
149. Choi KY, Jung E, Jung DH, Pandey BP, Yun H, Park HY, et al. Cloning, expression and characterization of CYP102D1, a self-sufficient P450 monooxygenase from *Streptomyces avermitilis*. *FEBS J.* 2012;279(9):1650-62.
150. Rey MW, Ramaiya P, Nelson BA, Brody-Karpin SD, Zaretsky EJ, Tang M, et al. Complete genome sequence of the industrial bacterium *Bacillus licheniformis* and comparisons with closely related *Bacillus* species. *Genome Biol.* 2004;5(10):R77.
151. Arnold FH. Directed Evolution: Bringing New Chemistry to Life. *Angew Chem Int Ed Engl.* 2018;57(16):4143-8.
152. Hussain HA, Ward JM. Enhanced heterologous expression of two *Streptomyces griseolus* cytochrome P450s and *Streptomyces coelicolor* ferredoxin reductase as potentially efficient hydroxylation catalysts. *Applied and Environmental Microbiology.* 2003;69(1):373-82.
153. Hussain HA, Ward JA. Ferredoxin reductase enhances heterologously expressed cytochrome CYP105D1 in *Escherichia coli* and *Streptomyces lividans*. *Enzyme and Microbial Technology.* 2003;32(7):790-800.
154. Omura T, Sanders E, Estabrook RW, Cooper DY, Rosenthal O. Isolation from adrenal cortex of a nonheme iron protein and a flavoprotein functional as a reduced triphosphopyridine nucleotide-cytochrome P-450 reductase. *Archives of Biochemistry and Biophysics.* 1966;117(3):660-73.
155. Momoi K, Hofmann U, Schmid RD, Urlacher VB. Reconstitution of beta-carotene hydroxylase activity of thermostable CYP175A1 monooxygenase. *Biochem Biophys Res Commun.* 2006;339(1):331-6.
156. Schallmey A, den Besten G, Teune IG, Kembaren RF, Janssen DB. Characterization of cytochrome P450 monooxygenase CYP154H1 from the thermophilic soil bacterium *Thermobifida fusca*. *Appl Microbiol Biotechnol.* 2011;89(5):1475-85.
157. Keselman H, Huberty CJ, Lix LM, Olejnik S, Cribbie RA, Donahue B, et al. Statistical practices of educational researchers: An analysis of their ANOVA, MANOVA, and ANCOVA analyses. *Review of educational research.* 1998;68(3):350-86.

158. Poole ES, Brown CM, Tate WP. The identity of the base following the stop codon determines the efficiency of in vivo translational termination in *Escherichia coli*. *EMBO J*. 1995;14(1):151-8.
159. Furuya T, Kino K. Discovery of 2-naphthoic acid monooxygenases by genome mining and their use as biocatalysts. *ChemSusChem*. 2009;2(7):645-9.
160. Gunsalus IC, Wagner GC. Bacterial P-450cam methylene monooxygenase components: Cytochrome m, putidaredoxin, and putidaredoxin reductase. *Methods in Enzymology*. 1978;52:166-88.
161. Davids T, Schmidt M, Bottcher D, Bornscheuer UT. Strategies for the discovery and engineering of enzymes for biocatalysis. *Curr Opin Chem Biol*. 2013;17(2):215-20.
162. Moody SC, Loveridge EJ. CYP105-diverse structures, functions and roles in an intriguing family of enzymes in *Streptomyces*. *J Appl Microbiol*. 2014;117(6):1549-63.
163. Sugimoto H, Shinkyo R, Hayashi K, Yoneda S, Yamada M, Kamakura M, et al. Crystal structure of CYP105A1 (P450SU-1) in complex with 1 α ,25-dihydroxyvitamin D₃. *Biochemistry*. 2008;47(13):4017-27.
164. Xu LH, Fushinobu S, Takamatsu S, Wakagi T, Ikeda H, Shoun H. Regio- and stereospecificity of filipin hydroxylation sites revealed by crystal structures of cytochrome P450 105P1 and 105D6 from *Streptomyces avermitilis*. *J Biol Chem*. 2010;285(22):16844-53.
165. Yasutake Y, Imoto N, Fujii Y, Fujii T, Arisawa A, Tamura T. Crystal structure of cytochrome P450 MoxA from *Nonomuraea recticatena* (CYP105). *Biochem Biophys Res Commun*. 2007;361(4):876-82.
166. Hayashi K, Sugimoto H, Shinkyo R, Yamada M, Ikeda S, Ikushiro S, et al. Structure-based design of a highly active vitamin D hydroxylase from *Streptomyces griseolus* CYP105A1. *Biochemistry*. 2008;47(46):11964-72.
167. Kabumoto H, Miyazaki K, Arisawa A. Directed evolution of the actinomycete cytochrome P450moxA (CYP105) for enhanced activity. *Biosci Biotechnol Biochem*. 2009;73(9):1922-7.
168. Wang ZX, Li SM, Heide L. Identification of the coumermycin A(1) biosynthetic gene cluster of *Streptomyces rishiriensis* DSM 40489. *Antimicrob Agents Chemother*. 2000;44(11):3040-8.

169. Yoshigae Y, Kent UM, Hollenberg PF. Role of the highly conserved threonine in cytochrome P450 2E1: prevention of H₂O₂-induced inactivation during electron transfer. *Biochemistry*. 2013;52(27):4636-47.
170. Tamura K, Dudley J, Nei M, Kumar S. MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. *Mol Biol Evol*. 2007;24(8):1596-9.
171. Betlach MC, Kealey JT, Ashley GW, McDaniel R. Characterization of the macrolide P-450 hydroxylase from *Streptomyces venezuelae* which converts narbomycin to picromycin. *Biochemistry*. 1998;37(42):14937-42.
172. Tian Z, Cheng Q, Yoshimoto FK, Lei L, Lamb DC, Guengerich FP. Cytochrome P450 107U1 is required for sporulation and antibiotic production in *Streptomyces coelicolor*. *Arch Biochem Biophys*. 2013;530(2):101-7.
173. Pandey BP, Lee N, Choi KY, Jung E, Jeong DH, Kim BG. Screening of bacterial cytochrome P450s responsible for regiospecific hydroxylation of (iso)flavonoids. *Enzyme Microb Technol*. 2011;48(4-5):386-92.
174. Nishihara K, Kanemori M, Kitagawa M, Yanagi H, Yura T. Chaperone coexpression plasmids: differential and synergistic roles of DnaK-DnaJ-GrpE and GroEL-GroES in assisting folding of an allergen of Japanese cedar pollen, Cryj2, in *Escherichia coli*. *Appl Environ Microbiol*. 1998;64(5):1694-9.
175. Nishihara K, Kanemori M, Yanagi H, Yura T. Overexpression of trigger factor prevents aggregation of recombinant proteins in *Escherichia coli*. *Appl Environ Microbiol*. 2000;66(3):884-9.
176. Gricman L, Vogel C, Pleiss J. Identification of universal selectivity-determining positions in cytochrome P450 monooxygenases by systematic sequence-based literature mining. *Proteins*. 2015;83(9):1593-603.
177. Goodwin S, McPherson JD, McCombie WR. Coming of age: ten years of next-generation sequencing technologies. *Nat Rev Genet*. 2016;17(6):333-51.
178. Quince C, Walker AW, Simpson JT, Loman NJ, Segata N. Shotgun metagenomics, from sampling to analysis. *Nat Biotechnol*. 2017;35(9):833-44.
179. Nie Y, Chi CQ, Fang H, Liang JL, Lu SL, Lai GL, et al. Diverse alkane hydroxylase genes in microorganisms and environments. *Sci Rep*. 2014;4:4968.
180. Procopio RE, Silva IR, Martins MK, Azevedo JL, Araujo JM. Antibiotics produced by *Streptomyces*. *Braz J Infect Dis*. 2012;16(5):466-71.