FUNCTIONAL RECOMBINANT EXPRESSION OF CYTOCHROME P450S AND NITRILASES IN ESCHERICHIA COLI, AND THEIR USE IN BIOCATALYSIS

M A ZOUPALI

PhD

2022

FUNCTIONAL RECOMBINANT EXPRESSION OF CYTOCHROME P450S AND NITRILASES IN ESCHERICHIA COLI, AND THEIR USE IN BIOCATALYSIS

MIGKENA ARTEMIS ZOUPALI

A Thesis Submitted in Partial Fulfilment of The Requirements for the award of Doctor of Philosophy of the University of Northumbria at Newcastle.

Research undertaken in the School of Life Sciences, University of Northumbria

December 2022

Abstract

The use of enzymes to perform biocatalytic reactions has emerged as a key area in industrial biotechnology due to the increased demand for a sustainable and economical method of producing a variety of chemicals, while eliminating the harmful biproducts formed and the harsh conditions used.

Two different types of enzymes were used throughout this work: the cytochrome P450s and the nitrilases. Cytochrome P450s are a superfamily of haem-containing enzymes that catalyse a vast array of oxidative reactions, including the synthesis of active pharmaceutical ingredients, with a high degree of regio- and stereoselectivity. However, they exhibit low recombinant expression in *E. coli*, poor stability, and low turnover number. Moreover, eight out of ten existing classes require the help of electron transport proteins for biocatalytic processes, with only class VII and VIII being self-sufficient systems which contain all the redox partner proteins needed for biocatalysis.

On the other hand, nitrilase enzymes catalyse the hydrolysis of nitrile compounds into the appropriate carboxylic acid and ammonia consequently have numerous commercial and biotechnological uses. Nitriles are intermediates in the synthesis of pharmaceuticals, and their chemical hydrolysis requires high temperatures and pH, and formation of toxic by-products. Nitrilases can hydrolyse the nitriles in a single step in environmentally friendly conditions.

This project involved the cloning and recombinant expression of already characterised human enzymes CYP2A6, CYP2B6, CYP3A4, CYP3A5 and CYP2D6 in *E. coli* strains, and their participation in 2S, 6S; 2R,6R hydroxynorketamine production from racemic norketamine in whole cell biotransformations; a biotransformation of particular interest to Quotient Sciences. CYP2B6 and CYP3A4, indeed, hydroxylated norketamine to hydroxynorketamine in whole cell biotransformations.

Moreover, after mining the Prozomix database two novel metagenome genes, 914 and 198 were identified with sequence identity up to 24% to human CYP3A5. Both 914 and 198 were cloned and recombinantly expressed. Metagenome gene 914 revealed activity against ketamine, which N-demethylated to norketamine, and hydroxylated to hydroxyketamine, while 198 did not hydroxylate ketamine or norketamine.

Furthermore, 20 candidate nitrilases taken from the private collection of Prof. Gary Black, were expressed, and assayed against a panel of six nitriles to determine their catalytic activity. Only one nitrilase enzyme showed good activity against two

nitriles, benzyl (2-cyanoethyl) carbamate and 2-(4-cyanophenyl)-N-ethylacetamide, which are of interest to Quotient Sciences.

In summary, the development of comparative methods in different cloning strategies will allow future scientist to enhance the expression and catalytic activity of the P450s. Moreover, the development of hydroxynorketamine from CYP2B6 and CYP3A4, hydroxyketamine from metagenome genes, and the development of both nitriles from nitrilase enzyme Q2GR86 will permit their use to be applied in industrial processes.

Contents

List of tables

Table of figures

List of abbreviations

Abbreviation	Term
1KP	One Thousand Plants
5-ALA	5-aminolevulinic acid
Å	Angstrom
ACN	Acetonitrile
ADHs	Alcohol dehydrogenases
Adx	Adrenodoxin-type ferrodoxins
AIM	Autoinduction media
ALA	Aminolevulinic acid
ALAD	Aminolevulinic acid dehydratase
ALAS	Aminolevulinate synthase
AOL	Allene oxide synthases
API	Active pharmaceutical ingredient
APS	Ammonium persulfate
ATP	Adenosine triphosphate
BLAST	Basic local alignment search tool
BM3	Bacillus megaterium
bp	Base pair
BSA	Bovine serum albumin
BSA	Bovine album serum
CFE	Cell free extract
CN	Cyanide group
CO	Carbon monoxide
CoA	Coenzyme A
Cpd-I	Compound I
CPgenIII	Coproporphyrinogen III
CPO	Coproporphyrinogen oxidase
CPR	Cytochrome P450 reductase
CPR	Cytochrome P450 reductase
CYP	Cytochrome P450
DDA	Data dependent analysis
DES	Divinyl ether synthases
DMSO	Dimethyl sulfoxide

DTT	Dithiothreitol
EDTA	Ethylenediaminetetraacetic acid
EICs	Extracted Ion Chromatograms
EICs	Extracted Ion Chromatograms
Ethanol	EtOH
FA	Formic acid
FAD	Flavin adenine dinucleotide
Fds	Bacterial-type ferredoxin
Fdx	Ferredoxin
FdxR	Ferredoxin reductase
FECH	Ferrochelatese
FMN	Flavin mononucleotide
GDH	Glutamate dehydrogenase
GIDH	Glutamate dehydrogenase
HCl	Hydrochloric acid
HEPES	4-(2-Hydroxyethyl) piperazine-1-ethanesulfonic acid
HK	Hydroxyketamine
HMB	hydroxymethylbilane
HMBS	Hydroxymethylbilane synthase
HMM	Hidden Markov Model
HMM	Hidden Markov model
HNK	Hydroxynorketamine
HO	Haem oxygenase
HPL	Hydroperoxide lyases
HPLC	High performance liquid chromatography
HT	High throughput
IAA	Indoloacetic acid
IAN	Indoloacetonitrile
IDT	Integrated DNA Technologies
IMAC	Immobilised metal affinity chromatography
IPTG	Isopropyl ß-D-1-thiogalactopyranoside
IPTG	Isopropyl β- d-1-thiogalactopyranoside
kDa	kilodalton
KREDs	Ketoreductases
LB	Luria Bertani broth

LbCR	Lactobacillus brevis ketoreductase
LC-MS	Liquid chromatography mass-spectrometry
M	Mol
mg	Milligram
MgCl	Magnesium Chloride
MIBiG	Minimum Information of biosynthetic gene cluster
mM	Millimole
MSA	Multiple sequence alignment
MW	Molecular weight
NaCl	Sodium Chloride
NAD(P)H	Nicotinamide adenine dinucleotide phosphate
NCBI	National Centre for Biotechnology Information
NCE	Normalized Collison energy
NDMA	N-methyl-D aspartate
NEB	New England Biolabs
NH	Amino group
NiSO4	Nickel sulphate
nm	Nanometre
NO	Nitric oxide
OD	Optical density
OFOR	2-oxoacid ferredoxin oxidoreductase
OPA	O-phthalaldehyde
OxR	Oxireductase
P450	Cytochromes P450
P450	Cytochrome P450
PAHs	Polycyclic aromatic hydrocarbons
PBG	Porphobilinogen
PBGS	Porphobilinogen synthase
PCBs	Polychlorinated biphenyls
PCDDs	Polychlorinated dibenzo-p-dioxins
PCR	Polymerase chain reaction
PCR	Polymerase chain reaction
PDB	Protein databank
pDNA	Plasmid DNA
PDOR	Phthalate dioxygenase reductase

PdR	Putidaredoxin reductase
Pfam	Protein family
PFAM	Protein family
PGIS	Prostacyclin synthases
PPgenIX	Protoporphyrinogen IX
PPIX	Protoporphyrin IX
ppm	Parts per million
PPO	Protoporphyrinogen oxidase
RE	Restriction endonucleases
REase	Restriction endonuclease
RMSD	Root-mean-square deviation
ROS	Reactive oxygen species
SDS-PAGE	Sodium dodecyl sulphate–polyacrylamide gel electrophoresis
SEC	Size Exclusion Chromatography
SIM	Single Ion Monitoring
Std	Protein standard (marker)
TBAX	Thromboxane synthases
TCA	Trichloroacetic acid
TCP	Total cell Protein
TEMED	Tetramethylethylenediamine
TFA	Tetrafluoruacetic acid
TIC	Total Ion Chromatogram
TMH	Transmembrane helix
U	Units
UniProt	Universal Protein Resource
UPgenIII	Uroporphyrinogen III
UROD	Uroporphyrinogen decarboxylase
UV	Ultraviolet
UV	Ultraviolet
WCB	Whole cell biotransformations
δ-ALA	Aminolevulinic acid
μM	Micromole

Acknowledgements

I would like to start by expressing my gratitude to my supervisor, Prof. Gary Black, for giving me this great opportunity and for having faith in me from the beginning. His encouragement and support enabled me to complete my study, but more importantly, his supervision and counsel on scientific and personal issues kept me motivated. I'll always be appreciative to him. Additionally, I want to thank Gareth Jenkins and Richard Castledine, the project managers at my organisation, for their confidence in me and for entrusting me with their project. I am grateful.

I want to express my gratitude to my family, including my partner, Ilias, my mum and dad, my sister Anneta, and my brother Leo, for their unwavering support over the years. Additional thanks are given to my spouse for believing in me, supporting me, and most importantly, putting up with me while I worked all those late nights and weekends.

I am grateful of each and every one of the friends and coworkers I have gained over the years. I'm so grateful to my friends and lab colleagues Efi Kritikaki, Pedro Fernandes Julia, Josef O'Sullivan, Claire Morley, Mingaile Valancuscaite, Ifeanyi Enekwa, Olga Cugaj, Sophia Quigley, and Gina Abdelaal for their support over the past four years at work and for the enjoyable nights out that kept us joking at trying times. Furthermore, I'd like to thank, Theodora Mantso, Elena Eftychiou, Sotiris Kyriakou, Victoria Artzanidou, Lynn Anderson, and Joe Watson, and all the PhD office people. Thank you for your love and friendship.

I would especially like to thank Nicola Brown and Jennifer Wright for the lab instruction and priceless counsel they have given me over the past four years. I wouldn't have made it without you.

I'd also like to express my gratitude to a group of people who have supported me much throughout my PhD. Jose Munoz and Dr. Ciaran Kelly for their assistance with all of my scientific inquiries, Sam Bowerbank and Graeme Turnbull for their training and support with the HPLC, Dr. William Cheung for running all of my LC-MS samples and helping me with the analysis, Henry Gould for his assistance with the LC-MS analysis, Dr. Edward Taylor for aiding me with crystallisation, and Drs. Bob Fin and Nicola C. for their assistance.

A thank you goes to technical support team Gordon, Karen, Dawn, Gary, Zara and Anthea, and to everyone working in the Ellison's labs at Northumbria University.

Finally, thank you to the Quotient Sciences Pharmaceutical Manufacturing (Arcinova), and the ERDF funding partners.

Declaration

I declare that the work contained within this thesis has not been submitted for any other award and the work is all my own work. I also confirm that this work fully acknowledges opinions, ideas and contributions from the work of others.

I declare that the Word Count of this Thesis is 37,476 words.

Name: Migkena A Zoupali

Date: 30/12/2022

1. Introduction

1.1. What is Biocatalysis

Green chemistry and biocatalysis has become a core field in the industrial biotechnology world, due to the need for a sustainable and affordable method for producing a variety of chemicals, such as pharmaceuticals, fine chemicals, and polymers, but most importantly due to enzymes' excellent catalytic efficiency. The utilisation of enzymes as biocatalysts has improved the synthesis of chemicals, while minimising the production costs and bioproduct waste. Most chemical and drug intermediates produced through chemical synthesis achieve maximum overall yield of about 50%, while the enantiomeric purity is compromised. Moreover, their manufacturing requires the use of catalysts, reagents and solvents which employ harsh working conditions, such as high pH and temperatures (Ran et al., 2008). Enzymes have excellent chemo-, regio- and stereo- selectivity on a wide range of substrates, producing high purity enantiomers that are difficult to achieve chemically. Enzymes are effective and incredibly simple to utilise and the variety of reactions that can be impacted has expanded due to the replacement of water with organic solvent (Rey et al., 2004).

Since the beginning of civilization, natural enzymes have been employed to produce goods like leather, indigo and linen, as well as foods like cheese, sourdough, beer, wine, and vinegar. The above processes relied on either enzymes being produced by naturally growing microorganisms or added from enzymes found from other resources such as fruits and vegetables. Thanks to the advent of fermentation techniques developed in the latter half of the 20th century, enzymes can now be produced as purified, well-characterized preparations on a large scale using desired production strains. For more than 30 years, enzymes have been included in detergents due to their ability to degrade various natural accruing substances found in stains (Kirk et al., 2002). One of the most prominent examples of biocatalysis is the utilisation of a nitrile hydratase, from *Rhodococcus rhodochrous J1*, that catalyses the conversion of acrylonitrile to produce acrylamide, on industrial scale (Yamada and Kobayashi, 1996). In the food industry, one of the oldest examples involve the xylose isomerase-3-catalyzed isomerization of D-glucose to D-fructose, which results in the manufacture of high-fructose corn syrup (Jensen and Rugh, 1987). Additionally, the utilisation of enzymes for the productions of fine chemicals and polymers have been successfully applied to industrial scale. Such enzymes include lipases/ esterases, proteases, amidases and dehalogenases. One example is the resolution of phenylglycidyl ester by lipase, which produces a diltiazem precursor. Also, protease chymotrypsin has

been used for the resolution of N-acyl amino acids and amino acid esters in industrial scale (Schulze and Wubbolts, 1999, Kadokawa and Kobayashi, 2010). Furthermore, through the asymmetric reduction of prochiral ketones, ketoreductases (KREDs), a subset of alcohol dehydrogenases (ADHs), have been used to produce chiral secondary alcohols (Sheldon et al., 2020).

The recent large-scale acceleration of (meta)genome sequencing, the evolution of structural biology and metabolic engineering and protein expression, and high throughput screening of enzymes used in biocatalysis has evolved into a game-changing technology for chemical synthesis. Also, the widespread availability of synthetic genes, at low cost, enables quick and inexpensive screening of a wide range of enzyme variations, and the utilisation of enzymes in whole microbial cells, cell free extracts, purified and/or immobilised enzymes as catalysts for chemical synthesis, has supported this expanding field (Gong et al., 2012). Additionally, focussed databases and search tools have been designed to increase the availability of these biocatalysts to a wider scientific community, and a variety of custom enzymes with high efficiencies and selectivities have now been generated rapidly and on a gram to kilogramme scale (Bell et al., 2021).

Numerous applications have been made possible by these characteristics, particularly in the food and pharmaceutical industries. In the pharmaceutical industry, thousands of tonnes of different beta lactam antibiotics, including the penicillin and cephalosporins, are synthesised industrially using penicillin G amidase. The chemical synthesis of these widely used antibiotics involved several steps and non-environmentally friendly solvents and reagents, and generated considerate amount of waste, but with the development of the recombinants DNA technology and the immobilisation of the enzymes, the cleavage of penicillin G can be achieved in water at 37ºC with ammonia as the only used reagent (Sheldon and Rantwijk, 2004). In the food industry, a fine example of biocatalysis is the synthesis of L-aspartame. L-aspartame is an artificial sweetener used by diabetics, initially manufactured by chemical synthesis, but now its two building blocks, L-phenylalanine and L-aspartic acid are linked together through bacterial enzyme thermolysine (Doble and Kruthiventi, 2007).

Additionally, mutagenesis studies have contributed significantly to the advancement of biocatalysis. Mutagenesis studies of *Lactobacillus brevis* ketoreductase (LbCR) lead to improved activity and thermostability of the enzyme with a 3.2-fold increase in kcat/KM, and a half-life that is 1900 times longer at 40°C. When LbCR was co-expressed with the glutamate, produced with high efficiency the chiral alcohol precursor for atorvastatin, which the latter is used for lowering cholesterol levels (Gong et al., 2019). Site-directed mutagenesis of a metagenome nitrilase with UniProt code Q6RWG0 resulted in the formation of

(R)-4-cyano-3-hydroxybutyrate with an enantiomeric excess of 98.5% from the desymmerization of 3-hydroxyglutaronitrile. The product is used in the atorvastatin synthesis (DeSantis et al., 2003).

Another group of enzymes that their potential has accelerated due to biocatalysis and molecular engineering, are cytochrome P450s. P450s perform a wide number of transformations including but not limited to N-, O-, and S-dealkylations, C-C bond cleavage, epoxidation, N- and S-oxidation, and desaturation (Urlacher and Girhard, 2019). With more than 1000 experimentally characterised biosynthetic gene clusters, cytochrome P450 monooxygenases rank as one of the most promiscuous and frequent PFAM domains in the Minimum Information about a biosynthetic gene cluster (MIBiG) database. It has been demonstrated that cytochrome P450s can alter substances from almost every major class of natural products. Engineered P450s have significantly surpassed the capabilities of naturally occurring biocatalysts to catalyse the synthesis of carbon-silicon, carbon-boron, and olefin cyclopropane bonds (Robinson et al., 2021). The most recent additions to the hemoprotein-catalysed reaction repertoire, such as intermolecular nitrene C-H insertion, carbene Si-H and C-H insertions, Doyle-Kirmse reactions, aldehyde olefinations, and azide-to-aldehyde conversions (Brandenberg et al., 2017). The use of P450s in synthetic multi-enzymatic cascade reactions and chemo-enzymatic processes opens up fascinating new possibilities for getting access to novel molecules that cannot be manufactured naturally or chemically (Robinson et al., 2021).

In this research project we will be focusing in biocatalysis reactions realised by two types of enzymes: The P450s and the nitrilases.

1.2. The P450 Superfamily

The cytochrome P450 superfamily (CYPs or P450s) consisted of more than 300,000 P450 proteins spread across all forms of life (Nelson, 2018), and as of December 2022, there are over 1 million P450 proteins in the GenBank (https://www.ncbi.nlm.nih.gov/protein/?term=P450). P450s have been found to exist in bacteria, archaea, and eukaryotes, suggesting that all known P450 genes originated billions of years ago from a common ancestral gene. However, there are no identifiable P450s in some archaea or eubacteria, and in the genome of *E. coli* (Danielson, 2002, Nelson, 2018).

With the introduction and commercialization of DNA sequencing (Sanger et al., 1977), many putative P450 encoding genes were quickly identified. In 1982, the first entire cytochrome P450 rat sequences of CYP2B1 and CYP2B2 were published (Fujii-Kuriyama et al., 1982, Waxman and Walsh, 1982). Currently, over 62,000 bacterial CYPs and over 85,000 fungal CYPs have been described in public and private information databases, of which 11,000 genes have been assigned a name. Moreover, over 13,000 animal and over 16,000 plants P450 have been assigned nomenclature (Nelson, 2018).

Depending on their sequence similarity, the P450 superfamily is divided into families (Nebert et al., 2013), and many subfamilies. Genes with more than 40% similarity are assigned to the same family, and are distinguished by a number added after the CYP root, while genes with equal or greater than 55% similarity are assigned to the same subfamily which are distinguished by a capital letter (i.e. CYP 2B6) (Bezirtzoglou, 2012). There are also a considerable number of pseudogenes in each of the species. Pseudogenes are genes that do not translate to a functional protein, either because of gene duplication where one of the genes has generated and lost function, or because a reverse transcribed mRNA has been reincorporated into genomic DNA, resulting in the absence of transcriptional regulatory elements. Pseudogenes are assigned a “P” after the family number. Hence, genes with more than 97% sequence similarity are considered pseudogene (Danielson, 2002).

The human genome contains 57 P450 genes encoded by 18 mammalian cytochrome families. P450s remain the major metabolisers of the drugs in the human body, converting lipophilic drugs to hydrophilic forms, resulting in their excretion through the kidneys (Quehl et al., 2016). The CYP2, CYP3, and CYP4 families have significantly more genes than the other 15 families, and are responsible for metabolising almost 75-80% of the drugs in clinical use (Zanger and Schwab, 2013).

During drug oxidation the required electrons which derive from NADPH, are delivered by the P450 reductase through its cofactors FAD and FMN to the haem of the monooxygenase (Munro et al., 2007, Urlacher and Girhard, 2012, Quehl et al., 2016). P450 enzymes are mainly expressed in the cytoplasm of endoplasmic reticulum of the liver (Pelkonen et al., 2008), but their expression can occur elsewhere in the body including the lungs, kidneys and small intestine and adrenal cortex (Cooper et al., 1965, Lynch and Price, 2007). In humans these enzymes are responsible for metabolism of drugs and other chemicals, breakdown of steroids and fatty acids, as well as hormone synthesis and breakdown (Danielson, 2002).

1.2.1. Background of P450s

The P450 superfamily are haem-containing mono-oxygenase enzymes that catalyse a wide range of oxidation reactions and accept numerous substrates.

P450s are in high demand due to their many functions. P450s are used within the biocatalysis community, most notably for hydroxylation reactions as they hydrolyse the conversion of a substrate to the final product and where one oxygen molecule is introduced into the substrate and the other is reduced to water. One example is the hydroxylation of Progesterone to 11-deoxycortisone by enzyme P450 21A2 (Figure 1-1) (Pallan et al., 2015). More than 20 reactions are performed by P450s according to Sono et al. (1996), such as hydrocarbon (and aromatic) hydroxylation, alcohol and aldehyde oxidations, nitric oxide reduction, alkyne oxygenation to carboxylic acids, olefin (and arene) epoxidation, and N-/S-/O-dealkylation reactions. P450s are responsible for the metabolism of numerous endogenous substrates, as well as drugs and other chemicals which are commonly termed exogenous compounds. Exposure to some exogenous compounds will induce the production of P450s (Pan et al., 2011, Pikuleva and Waterman, 2013). For example, some P450s are activated in terms of cellular expression by certain chemicals, among them ketamine (Chan et al., 2005), while their expression is suppressed, or are inactivated by other substances (Amit et al., 2007).

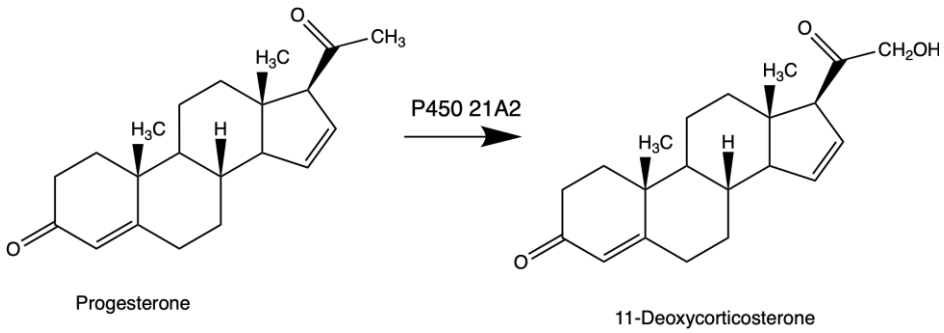


Figure 1-1 Hydroxylation of progesterone to 11-Deoxycorticosterone (Pallan et al., 2015). Figure created with ChemDraw.

P450s were first mentioned in 1955 by Brodie et al., where an enzyme in the human liver microsomes oxidised xenobiotic compounds. Then in 1958, Klingenberg and Garfinkel detected and described in human liver microsomes a carbon monoxide containing (CO) pigment which would absorb at 450 nm by difference spectroscopy (Omura and Sato, 1964). They named it Pigment-450 nm.

The protein moiety influenced the pigment's peak intensity and precise absorbance. In the oxidised state, less noticeable peaks in cytochrome P450s were seen at approximately 570, 540, 410 and 392 nm (Hall, 1993).

The identity of P450 was not totally evidenced however until Omura and Sato (1964) reported its nature to be that of a type-b haemoprotein (Figure 1-2). Cooper et al. (1965) discovered that the pigment P450 was dependent on nicotinamide adenine dinucleotide phosphate (NADPH) for the carbon 21 hydroxylation of 17-hydroxyprogesterone in bovine adrenal microsomes (Cooper David et al., 1965).

Type b-haemoproteins contain a protoporphyrin IX ring structure consisted by four nitrogen atoms bound by four carbons each to form four rings, otherwise stated as pyrroles, coordinated by an iron ion (Fe^{+}) in the centre of the ring. The iron is also coordinated by two axial ligands; on the fifth position an imidazole nitrogen of a cysteine on the proximal side of the ring, while the sixth position on the distal side of the planar ring is vacant, and is reserved for the water molecule. The axial thiol ligand connects the haem to the apoprotein (Collins and Dawson, 2013).

Electrostatic, salt bridges $-NH^{+3}$ hold the four porphyrin subunits in a roughly tetrahedral configuration and the iron molecule in place. Each nitrogen molecule faces the interior of the bigger ring they produce (Butterworth and Heinemann, (1997).

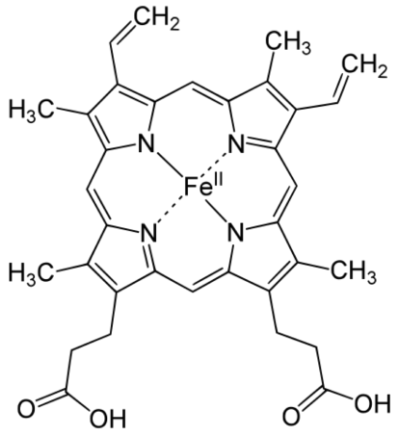


Figure 1-2 Type-b haem. Chemical structure of haem-b with the four pyrroles coordinated by the iron (Fe^{+}) ion. Figure created with ChemDraw.

When Omura and Sato (1964) reduced an isolate of rabbit liver microsomes with sodium dithionite and then bubbled it through carbon monoxide, they observed a difference spectrum with a prominent peak appearing at 450 nm, the maximum absorbance of the haem (named the Soret Peak), indicating the presence of P450 in the liver microsomes. When comparing their results to cytochrome b5, another haemoprotein, they confirmed that the P450s are a different class of haem-containing enzyme that utilise a novel haem prosthetic group (Omura and Sato, 1964).

Katagiri et al. (1968) then discovered a soluble camphor-inducible enzyme in *Pseudomonas putida*. They gave it the name P450cam in reference to its ability to activate the first stage in the breakdown of camphor, the 5-exo-hydroxylation of D-camphor. P450cam allowed camphor to be used as a sole source of carbon and energy for growth of *P. putida* (Katagiri et al., 1968). Their team made the pivotal discovery that the axial thiol residue was a cysteine residue (C^{357}), which helped explain the Soret Peak consistent with P450s enzymes. This is the most conserved residue among P450s.

P450cam was successfully investigated by X-ray crystallography as the first P450 enzyme after being purified (Figure 1-3) which allowed Poulos et al. (1987) to conduct mechanistic and spectroscopic studies. When they compared the P450cam crystal structures of camphor free and camphor bound, they discovered that in the high spin camphor-bound state the haem ring is penta-coordinate, while in the low spin camphor-free the haem iron is hexa-coordinate, with a water molecule occupying the distal axial ligand (Poulos et al., 1987).

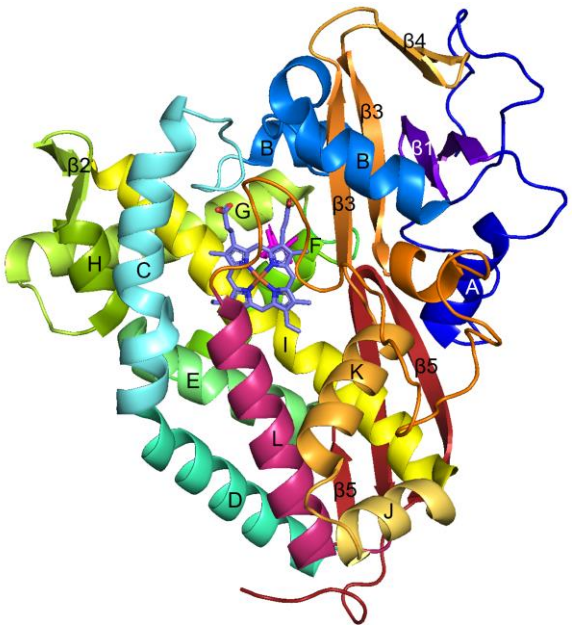


Figure 1-3 Structure of P450cam (PDB: 2CPP) edited with PyMOL according to secondary structure (Poulos et al., 1987). The P450cam is separated into helix rich domain, where all the helices are found, and the helix-poor or β-barrel domain. The A helix coloured in dark blue (aa 37-46), B helices in marine (aa 67-77 and 89- 96), C helix in sky blue (aa 106- 126), D helix in green cyan (aa 127-145), E helix in lime green (aa 149-169), F helix in green (aa 173-185), G helix in lemon (aa 192-214), H helix in dark green (aa 218-225), I helix in bright yellow (aa 234-267), J helix in yellow (aa 267-276), K helix in bright orange (aa 280-292), and L helix in warm pink (aa 359-378). β1-sheet in purple (aa 52-66), β2-sheet in dark green (aa 226-233), β3-sheets in orange (aa 295-301, 315-323), β4-sheet in bright orange (aa 305-312), and β5-sheets in fire brick (aa 382-405, 146-150). The haem moiety can be seen located in the centre in light purple.

Poulos et al. discovered from the first X-ray crystallography analysis of the crystal structure in 1987, that A-L helical segments make up 40% of the structure, and P450cam, unlike many other enzymes solved then, does not fold into an N- and C-terminal domain as other enzymes do. Instead, it starts on the left-hand side, it crosses over to the right, where most of the helices are found, crosses back over to the left, and then again to the right where Val 414, the C-terminus is located adjacent to the D helix (Figure 1-3) (Poulos et al., 1985). P450cam does not contain any cleft or hinge but is instead divided into right side being helical rich and left side being helical poor domain containing helices C-L, and helices A and B, respectively. Many residues were discovered to have an important role in substrate binding, notably that of Tyr96 which anchors the substrate via a hydrogen bond, which was formed between carboxyl and carbonyl group between Tyr96 and camphor, respectively. Others noteworthy molecules that orientate camphor in the active site with hydrophobic interactions are Phe 87; Leu 244, Val 247, Thr 252, and Val 295 (Poulos et al., 1985).

Haem is positioned between proximal and distal helices L and I, respectively, and is in direct contact. The β-bulge hosting the Cys ligand is the most notable segment, which holds it into place to be in close distance between two amino groups. Another difference is that the Cys^{357} which is responsible for the formation of haem thiolate ligand, extends from the N-terminus of the proximal helix in P450cam, whereas the histidine ligand extends from the C-terminus of the proximal helix in haemoglobins and cytochrome c peroxidases. Because the position that haem it is surrounded by apolar residues, and buried, with the closest position to the molecular surface being 8 Å , there is a necessity for an electron donor system within close proximity for catalysis (Poulos et al., 1985).

1.2.2. Cytochrome P450 Conserved Structure and Specificity Towards Substrates

P450s share the same or similar overall fold with each other but differ significantly towards their structural element position. The fold and charge confer a definite specificity to the active site towards binding to a certain substrate. All P450 structures are progressively more conserved the closest to the haem region, with the most conserved being the cysteine residue, where the haem Fe is bound to the protein through the thiolate sulphur of the cysteine residue. Site directed mutagenesis of the cysteine residue results in loss of catalytic activity of the enzyme (Lamb and Waterman, 2013).

Close to haem thiolate oxygen activation region are the I and L helices which are in contact with the haem, and the β-bulge located prior to L helix. The significant role of ß-bulge, where cysteine ligand is located, is also to protect and stabilise the cysteine ligand so that it is in close proximity to the two peptide NH groups in order to accept H-bonds. H-bonds regulate the haem iron redox potential (Poulos and Johnson, 2015).

In contrary, regions that regulate the substrate specificity, such as B-C loop, vary in orientation in different P450s, resulting in differences in orientating the substrate in the active site. Also, substrate specificity is dependent on the active site shape and size (Poulos and Johnson, 2015).

On the surface of the P450, close and on the same side as the haem, are located a stretch of arginines that bind the flavoprotein NADPH-dependent cytochrome P450 reductase (CPR), as well as the cytochrome b5, which can be an accessory protein to some P450s (Guengerich et al., 2016).

1.2.3. Haem Biosynthesis Pathway

Haem is a prosthetic group for a number of different types of proteins, including catalase, peroxidase, cytochrome c, myoglobin, and cytochrome c. Haem is critical for P450 protein synthesis, assembly, repair and disposal of the proteins (Correia et al., 2011, Swenson et al., 2020).

In eukaryotes, haem biosynthesis is initiated in the matrix of the inner membrane of mitochondria (Figure 1-4). Eight enzymes are required to synthesise haem from glycine, succinyl-coenzyme A (CoA), along with molecular oxygen (O_2) and iron (Fe). Four of the enzymes exist in the mitochondria, while the other four are in the cytosol. Since glycine is produced in the cytosol, different trafficking mechanisms take place to deliver glycine in the mitochondrial matrix to interact with succinyl CoA (Swenson et al., 2020).

The first enzyme required in haem biosynthesis is 5-aminolevulinic acid (ALA). In the first step of haem synthesis, 5- aminolevulinate synthase 1 (ALAS), an obligatory cofactor pyridoxal 5'-phosphate, form ALA from glycine and succinyl-CoA, while CO_2 and CoA are released (Hunter and Ferreira, 2009). Then ALA is transferred back to the cytosol where the next four enzymatic reactions happen (Furuyama et al., 2007). ALAS-1 and -2 contains three classical haem regulatory motifs (HRMs), of which two are located in the targeting sequence, while the third is located at the N- terminus of the mature ALAS. Haem binds to the regulatory motifs and is translocated in the mitochondria (Hamza and Dailey, 2012). Glycine is only hypothesised to be trafficked outside to the mitochondria through SLC25A38, although there is not enough research to confirm this (Yien and Perfetto, 2022).

In the second step of haem synthesis, porphobilinogen synthase (PBGS) (previously known as ALA dehydratase (ALAD)), condensates two molecules of ALA from cytosol, to form a single pyrrole, porphobilinogen (PBG). Then, hydroxymethylbilane synthase (HMBS) (previously known deaminase (PBGD)), combines four PBG molecules to form a linear tetrapyrrole hydroxymethylbilane (HMB). Uroporphyrinogen III synthase (UROS) catalyses the inversion and cyclisation of HMB pyrrole D ring to form uroporphyrinogen III (UPgenIII) (Swenson et al., 2020).

Uroporphyrinogen decarboxylase (UROD) catalyses the carboxylation of the four-pyrrole acetic acid side chains of UPgenIII to coproporphyrinogen III (CPgenIII), and four carbon monoxide molecules. At this step, CPgenIII is transported back into the inner mitochondrial membrane space by ABCB6, or 2-oxoglutarate carrier (SLC25A11) (Furuyama et al., 2007, Swenson et al., 2020). Once in the inner

mitochondrial membrane, CPgenIII is converted to protoporphyrinogen IX (PPgenIX) by coproporphyrinogen oxidase (CPO), and then to protoporphyrin IX (PPIX) by protoporphyrinogen oxidase (PPO), by the incorporation of three molecules of oxygen and production of three molecules of hydrogen peroxide (H_2O_2). Then, PPIX serves as a substrate to mitochondrial catalyst ferrochelatese (FECH) which, in the mitochondrial matrix, embeds the ferrous iron into the porphyrin ring, forming haem-b. The newly synthesised haem is transferred to the cytosol, where it will bind and form the new haemoproteins. Any unbound haem will be fragmented down into carbon monoxide, iron and biliverdin, by haem oxygenase (Correia et al., 2011) (Furuyama et al., 2007, Swenson et al., 2020).

Haem is a highly reactive and cytotoxic moiety, hence its synthesis and degradation in the liver, that subsist in a balanced manner, and is regulated by negative feedback (Correia et al., 2011). Haem synthesis is reduced by suppressing the expression of ALAS, while haem degradation is stimulated by inducing the expression of haem oxygenase (HO)-1. There exist two ALAS forms expressed in eukaryotes, ALAS1 and ALAS2. ALAS1 is ubiquitously expressed in all tissue forms, including erythroid cells, whereas ALAS2 is unambiguous to erythroid cells. Only ALAS1 synthesis is repressed by haem negative feedback in order to maintain an intracellular balance, whereas large ALAS2 expression is required for haemoglobin assembly, and therefore is not repressed by haem (Furuyama et al., 2007).

More than 50% of the haem produced in the liver, is consumed in P450 synthesis, while the rest is used for either protein binding, or exists loosely bound to intracellular proteins in a “free form”, ready to regulate cellular processes and interact with proteins in prerequisite of haem (Correia et al., 2011).

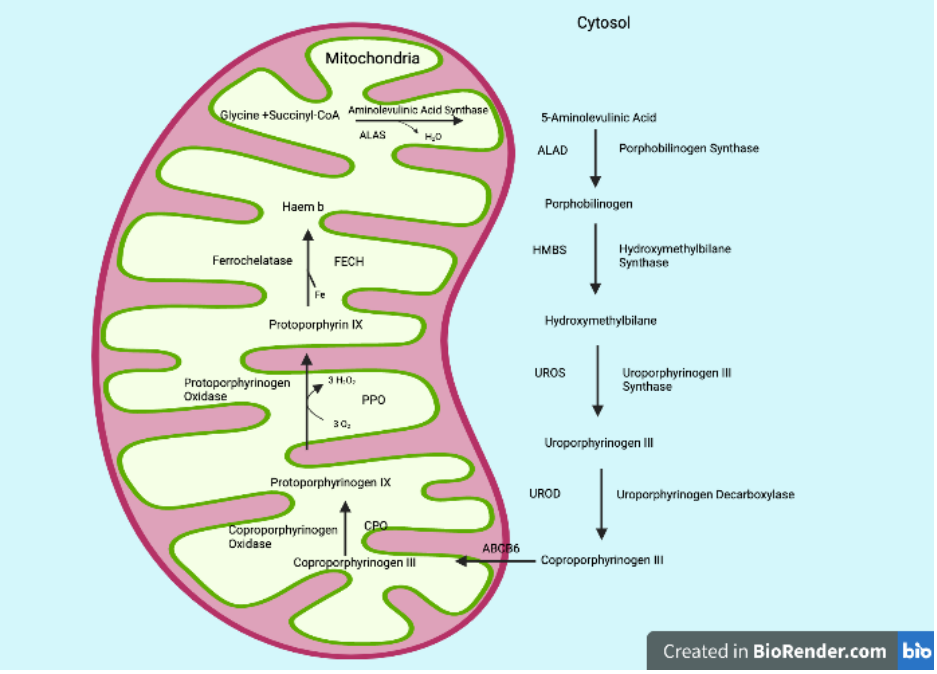


Figure 1-4 Haem biosynthesis pathway in eukaryotes as described by Swenson et al, 2020. In the figure can be observed the eight enzymes required to synthesise haem b in the final step, as well as the enzymes that catalyse the reactions. The first and the rate limiting enzyme for the haem biosynthesis, 5- aminolevulinic acid (ALA), is synthesised in the mitochondria matrix, as well as protoporphyrinogen IX (PPgenIX), protoporphyrin IX (PPIX), whereas porphobilinogen (PBG), hydroxymethylbilane (HMB), uroporphyrinogen III (UPgenIII) and coproporphyrinogen III (CPgenIII) are synthesised in the cytosol. The final enzyme in the pathway, ferrochelatese (FECH) on the mitochondrial matrix, catalyses the insertion of the iron to the porphyrin ring to form haem b. Figure created with BioRender.com

1.2.4. The P450 Catalytic Cycle

The catalytic cycle of the cytochrome P450 (Figure 1-5) is a very complex, multiple step mechanism where numerous intermediates are employed. Specifically, the molecular oxygen is enzymatically activated and cleaved into two molecules, where one of them is inserted into the substrate, so the monooxygenation of the substrate molecule can take place, and the other is reduced to water with the release of the product (Lewis and Pratt, 1998).

In stage one the iron (Fe) of the P450 haem starts in the ferric (Fe^{III}) low spin resting state, where the Fe is hexacoordinated by four nitrogens of the four pyrrole rings, a cysteine thiol ligand on the proximal site of the planar haem, and a water molecule, as an axial ligand on the distal side of the haem (Lamb and Waterman, 2013).

In stage two of the catalytic cycle the substrate binds in the active site of the haem domain, and the water molecule that covers the active site is dismissed, resulting in a high spin pentacoordinate Fe, ready for redox potential. The active site undergoes conformational changes. The crucial step in starting the catalytic cycle is binding of the substrate to the active site. The attached substrate causes the outer orbital to have one or more unpaired electrons, which elevates its energy level and causes it to transition from a low to a high spin state (Lewis and Pratt, 1998) .

In stage three, the first reduction takes place. In the presence of NAD(P)H, an electron is delivered from the redox partner flavin adenine dinucleotide (FAD) and flavin mononucleotide (FMN) for NAD(P)H dependent cytochrome P450 reductase, such as human reductase to the Fe^{III} of the haem, which generates the ferrous (Fe^{II}) pentacoordinate haem (Lamb and Waterman, 2013). The change of energy creates the "difference spectra" when the enzyme substrate complex absorbance is measured. When one such complex is formed, it has been observed a decrease in absorbance at 420 nm and an increase at 390 nm.

At stage four the dioxygen will bind in the distal axial of the high spin Fe^{II} of the haemoprotein leading to the $Fe^{II} + O_2$ (ferrous dioxygen) species being generated. The haemoprotein has to be in the ferrous form for the oxygen to bind and be activated (Munro et al., 2007). At this stage, the iron-oxygen bond can decouple, generating a ferric superoxide radical ($Fe^{III+} + O_2^-$) that is highly reactive and can destabilise the catalytic cycle (Lewis and Pratt, 1998). This step is quasi stable and autooxidation can happen, which can lead the iron to the resting state (in stage 2). The autooxidation is dependent in the redox potential, meaning the autooxidation is possible only if only is faster than the second electron transfer (Luthra et al., 2011).

At stage five the second reduction takes place, where the second electron is transferred to the haem from the redox partner, resulting in ferric peroxy low spin complex (Sono et al., 1996).

In stage six the oxyferrous P450 complex is double protonated attracting a hydrogen from a neighbouring bound water molecule and amino acid side chains. This results in the cleavage of the dioxygen bond, which binds to the hydrogen and forms a water molecule and the Fe (IV) oxo-porphyrin π-radical P450 or compound I intermediate (Cpd-I) (Lamb and Waterman, 2013). Following that, an oxygen atom is inserted into the substrate's C–H bond generating the hydrophilic product (RO-H) which is released from the active site. A water molecule is bound to the active site returning the iron to the hexacoordinate resting state, where it can be used to bind and oxygenate another substrate molecule (Waskell and Kim, 2015). Cpd-I is commonly regarded as the final intermediate stage in the P450 catalytic cycle and is considered to be the primary oxidising agent that generates the hydroxylated product (Rittle et al., 2010).

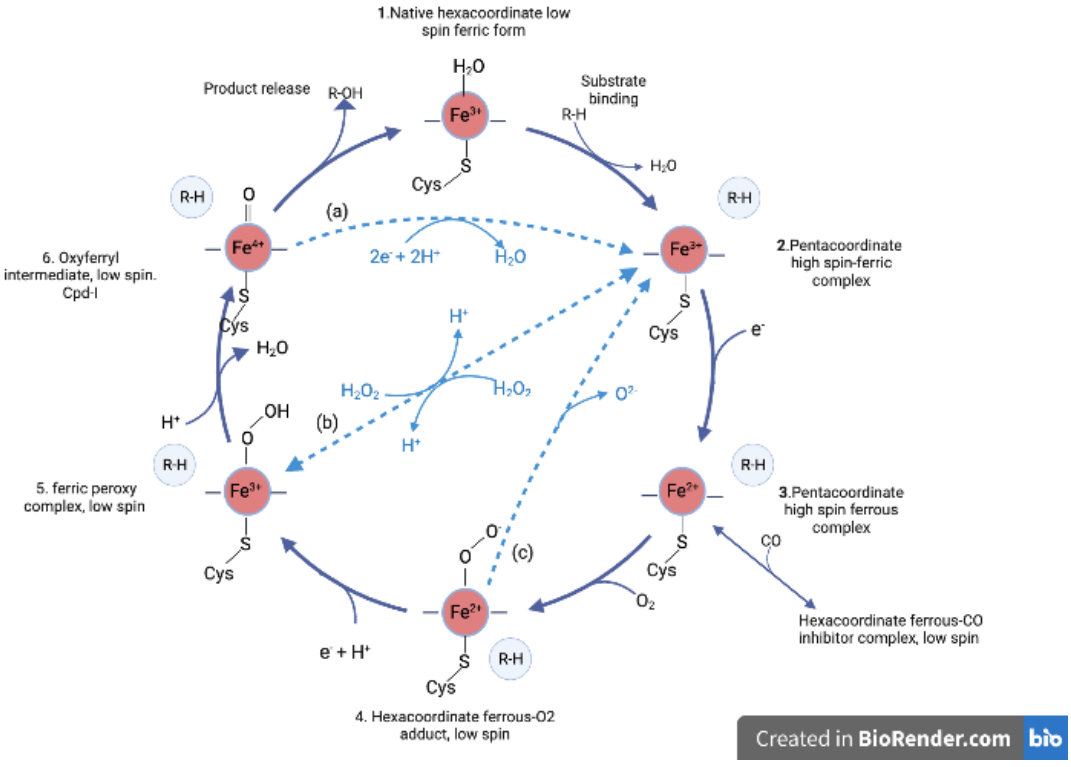


Figure 1-5 Cytochrome P450s catalytic cycle. Figure created with BioRender.com

1.2.5. P450 Uncoupling

During the catalytic cycle of the P450 enzyme, one molecular oxygen binds to the substrate while the other is reduced to water, in an ideal situation. In eukaryotic P450s though, atoms of the activated oxygen can be released from the enzyme without the oxygenation of the substrate and produce reactive oxygen species (ROS), which can destabilise the catalytic cycle (Lewis and Pratt, 1998). Those are the unstable oxygen atoms that can interact with other molecules. There are three branches that can produce ROS. First branch is the autooxidation shunt pathway (Figure 1-5 c) where degradation of ferrous-peroxo complex leads to the release of anion superoxide radical. The second branch where this can happen is the peroxidase shunt pathway (Figure 1-5 b), where the P450-hydroxy-ferric complex is protonated and forms hydrogen peroxide, and the third branch is the oxidase shunt pathway (Figure 1-5 a) where an atom of oxygen is released in the form of water molecule following the double reduction and diprotonation of the Cpd I intermediate (Zangar et al., 2004).

Coupling is the efficiency which the electrons are transferred from NAD(P)H to the redox partners, and eventually to the haem of the P450 to catalyse the oxygenation of the substrate, which is a good determinant of the efficiency of the catalytic reaction expressed in mole % (Jung et al., 2011). The three above mentioned pathways are instead the uncoupling of the oxygen. P450cam has a coupling efficiency of 100%, which means all the donated electrons are used to catalyse the oxygenation of the camphor to 5-hydroxycamphor. There are totally different degrees of coupling from enzyme to enzyme, with the prokaryotic enzymes being the most efficient and the microsomal ones the most poorly coupled (Zangar et al., 2004). Enzyme engineering efforts have focused on the efficacy of the P450s regarding coupling (Jung et al., 2011).

1.2.6. Classification of P450s

P450s can perform catalytic reactions after interacting with redox partners to source the electrons they require for their activity. There are various P450 systems, such as those that, although expressed from different genes, co-function with the redox partners and utilise them for catalytic activity. Another group is the self-sufficient P450-monoogygenase systems, where the redox partners are fused within the enzyme. The self-sufficient P450s do not require any additional protein component for catalytic activity. Such an example is nitric oxide reductase CYP55, which reduces nitric oxide (NO) to produce N_2O using as electron donor NADPH without the need of a redox partner system (Degtyarenko and Kulikova, 2001).

Depending on the nature of the electron transfer proteins they use, the domain architecture and the cellular location of the P450-monooxygenase systems, P450s are classified into 10 different classes (Figure 1-6). P450s require redox partner proteins for catalysis, such as flavodoxin reductases and flavodoxins for electron transfer from NADPH to the haem centre in the P450, depending on the class they belong (Munro et al., 2007, Urlacher and Girhard, 2012).

All P450-containing monooxygenase mechanisms have a mutual structural and functional domain architecture. These mechanisms can contain different redox partners that exist in different species in various combinations such as: cytochrome P450 reductase containing FAD and FMN; bacterial-type ferredoxin (Fds); adrenodoxin reductase (Pan et al.); putidaredoxin reductase (PdR); cytochrome b5 and adrenodoxin-type ferrodoxins (Adx).

The general electron flow strategy for P450-containing monooxygenase systems includes NAD(P)H acting as the electron reservoir that are transferred through FAD domain to the P450 domain, and finally transferred to the bimolecular oxygen (O_2), which ends up with the reductive cleavage of the oxygen to substrate and water, as shown below in the catalytic mechanism of P450s (Degtyarenko and Kulikova, 2001).

$$RH + O^2 + 2e^- + 2H^+ \rightarrow ROH + H_2O$$

Initially it was hypothesised that there existed only two P450 redox systems, corresponding to the already discovered P450s at the time; (i) the class I prokaryotic system containing the soluble protein from the previously characterised P450cam, as well as the system's two soluble redox enzymes, the FAD-containing NADH-specific putidaredoxin reductase, and the iron sulphur cluster of the type 2Fe-2S containing ferredoxin, the putidaredoxin, and (ii) class II eukaryotic system containing membrane P450s and only one additional protein, the CPR containing

FAD and FMN to perform monooxygenase activity in the presence of NADPH and oxygen (McLean et al., 2005).

More extensive studies and the discovery of the self-sufficient P450 monooxygenase from soil bacterium *Bacillus megaterium* (P450 BM3) provided the evidence that more diverse P450 systems exist, and not only the two hypothesised classes. The P450 BM3 exists as a long single polypeptide where the P450 is fused with the CPR and catalyses the ω-2 hydroxylation of long chains fatty acids, alcohols and amides, and the hydroxylation and epoxidation of monounsaturated fatty acids with the only addition of NADPH and O_2 (Narhi and Fulco, 1986).

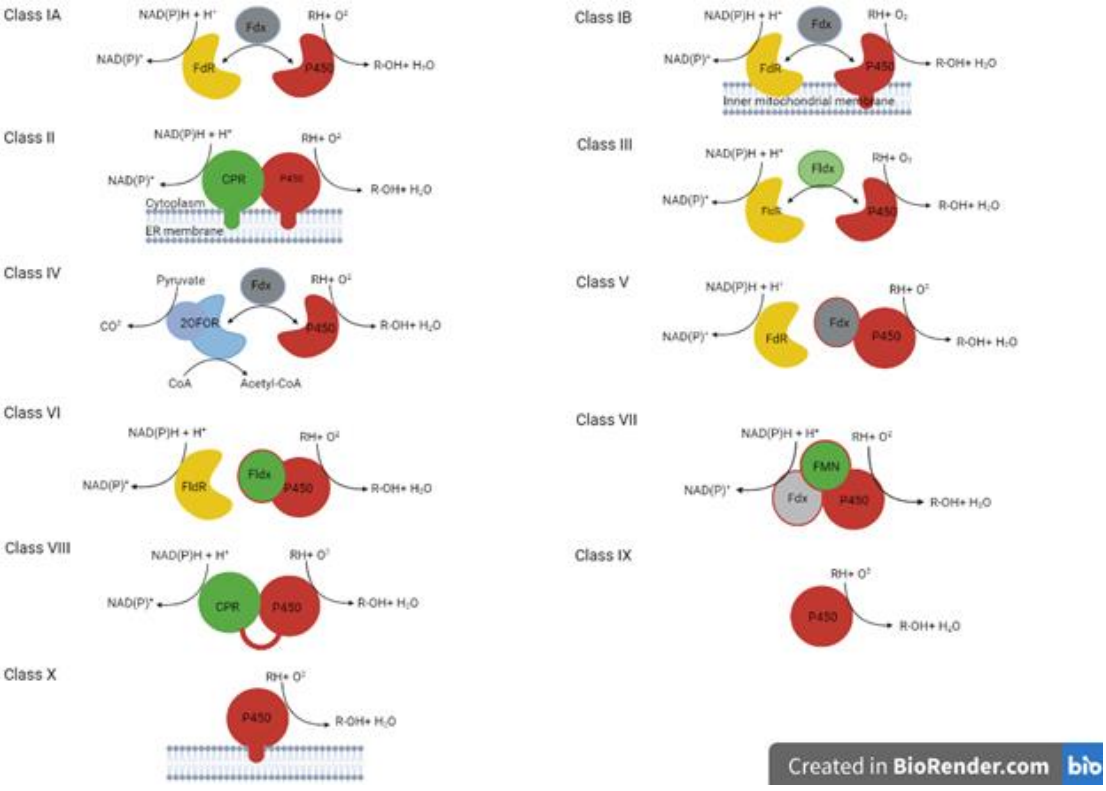


Figure 1-6 A diagram illustrating the domain architecture of the various P450 systems, created with BioRender.com, and adapted from (Hannemann et al., 2007). Class IA, bacterial system; class IB, mitochondrial system; class II, microsomal system; class III, bacterial system; class IV, thermophilic bacterial system; class V, bacterial fused Fdx-P450 system; class VI, bacterial fused Fldx-P450 system; class VII, bacterial fused PFOR_P450 system; class VIII, bacterial fused CPR-P450 system; class IX, soluble eukaryotic P450nor; class X, acyl hyperoxide utilising systems (Hannemann et al., 2007).

The Class I P450 system includes the majority of bacterial cytochrome P450 systems, the cytosolic P450s class IA, and the eukaryotic mitochondrial P450 systems, class IB. Both systems share their three-protein domain architecture of an FAD-containing ferredoxin reductase, which transfers the electrons from the NAD(P)H to an iron-sulphur cluster ferredoxin of the type: either 1Fe-0S, 2Fe-2S, 3Fe-4S, 4Fe-3S, or 4Fe-4S, and to the haem of the P450. Both these systems are not phylogenetically related, even though belong to the same class. In bacteria all three proteins of the system are soluble, while in the mitochondrial systems only the ferredoxin is soluble, with the other two enzymes being internally bound to the mitochondrial membrane (McLean et al., 2005, Hannemann et al., 2007). P450cam from Pseudomonas putida (CYP101) is a good example of prokaryotic class IA systems (Poulos et al., 1985, Pylypenko and Schlichting, 2004), whereas for class IB a good example is the adrenal steroid hydroxylase P450scc (CYP11A1) ("scc"-side chain cleavage) which uses cholesterol as natural substrate and cleaves its side chain to convert it to pregnenolone (Hanukoglu and Jefcoate, 1980).

The Class II P450 system mostly consisted of the eukaryotic microsomal systems. Microsomal class II systems employ a two-component system, the P450 domain and the NADPH dependent P450 reductase. The electrons are transferred from NADPH through FAD to the FMN of the CPR, which carries them to the haem of P450 and molecular O_2. CPR and nitric-oxide synthase (NOS) are the only two reductases found in mammalian cells that carry both FAD and FMN prosthetic groups. Both P450s and P450-reductase are located in the endoplasmic reticulum membrane facing the cytoplasmic side (Hanukoglu, 1996). X-ray crystallography of the recombinantly expressed rat CPR (Figure 1-7), at a resolution 2.6 Å proved that the CPR molecule is made up of two functional domains: the hydrophobic N-terminal domain that anchors the enzyme to the endoplasmic reticulum nuclear envelope, maintaining adequate spatial interaction for electron transfer between the reductase and cytochromes P450, and the hydrophilic C-terminus domain, which the latter can be subdivided further into two structural domains, and the connecting domain (Wang et al., 1997). The C-terminus subdomain is consisted of a fusion of bacterial flavodoxin FMN binding reductase, and two NAD(P)H dependent FAD containing enzymes, the ferredoxin reductase and cytochrome b5 reductase. Truncating the hydrophobic anchor, the CPR is not capable of transferring electrons to the P450 anymore, which means loss of activity, although it can transfer electrons to cytochrome c and other synthetic electron acceptors (Wang et al., 1997).

P450sca, isolated from the soil bacterium *Streptomyces carbophilus*, is the only prokaryotic P450 currently belonging to bacterial class II systems. P450sca was shown to catalyse the hydroxylation of mevastatin (ML-236B) to pravastatin,

which can be used in therapy of hypercholesterolemia, as it reduces significantly the cholesterol levels in the plasma (Serizawa and Matsuoka, 1991).

Class II enzyme systems are extensively used in drug metabolism studies despite being membrane bound proteins, and hence considered difficult to work with, and to achieve sufficient recombinant expression (Ingelman-Sundberg, 2004, Zanger and Schwab, 2013).

Figure 1-7 Rat's CPR crystal structure shown in cartoon diagram. The FMN-binding domain is represented in blue with the cofactor FMN represented as ball and stick in light blue. Next to it is the connecting domain in pink, and the FAD- and NADP(H)-domains in green, with the cofactors FAD in yellow and $NADP^+$ in orange. A disordered region that represents a "hinge" is between the FMN-binding and the connecting domain (Wang et al., 1997). Figure annotated with PyMOL.

The Class III P450 system was discovered by Hawkes et al. in 2002. They isolated and purified a novel P450 from a strain of *Citrobacter braakii*, P450cin, which uses cineole 1 as a sole carbon source. Based on sequence homology, they identified three open reading frames: cytochrome P450, a NADPH-dependent flavodoxin/ferredoxin reductase, and a flavodoxin. This system resembles the class I of the majority of bacterial, and some eukaryotic mitochondrial systems, in the sense that it employs the same three protein component system for catalytic activity. While the electron flow in the Class I is from the NADPH-dependent FAD containing ferredoxin reductase to the iron-sulphur cluster protein, to the P450 centre, in the class III the iron-sulphur cluster is replaced by a flavoprotein,

suggesting that uses flavodoxin as its natural redox partner. In class III the electrons are delivered from FAD to FMN redox centres, and that makes class III similar to eukaryotic microsomal system class II, only that instead of the fused CPR domain of class II, in class III the FAD and FMN are expressed as independent proteins (Hawkes et al., 2002). P450cin was recombinantly expressed in the presence of cineole 1 hypothesising that even though *E. coli* does not have a P450 catalyst, it's flavodoxin reductase might mimic the P450cin mechanism. The result showed the biodegradation of cineole 1 and the production of (S)-6β-hydroxycineole, which support the hypothesis that cineole 1 is the enzyme's natural substrate (Hawkes et al., 2002).

Class IV is made up entirely of extremophiles, of which only a few P450 systems have been discovered. The first thermophilic P450 discovered, and hence being extensively studied, is the soluble CYP119 from archaeon *Sulfolobus solfataricus* (Wright et al., 1996). CYP119 has high stability to temperatures as high as 91 °C, and pressure to 200 MPa, with pressure at 380 MPa causing half inactivation. Upon decompression from 500 MPa, the inactive P420 was able to convert to an active P450 (McLean et al., 1998, Tschirret-Guth et al., 2001). P450s in this family utilise a 2-oxoacid: ferredoxin oxidoreductase (OFOR), which is serving as an electron acceptor from coenzyme-A and pyruvic acid rather than NAD(P)H, and its co-partner ferredoxin as redox partners, both originating from the same *Sulfolobus* family. This P450 is active against lauric acid at an optimum temperature of 70ºC (Puchkaev et al., 2002, Puchkaev and Ortiz de Montellano, 2005).

Class V consists of a single P450 characterised to date, that of a sterol 14α-demethylase from *Methylococcus capsulatus,* CYP51. This CYP represents a new system, which consists of two separate protein components, an NAD(P)H dependent ferredoxin reductase, and a P450 domain fused to the C-terminus of the ferredoxin type 3Fe-4S enzyme, via an alanine rich linker which acts as a hinge between both domains to allow interaction. This protein domain arrangement suggests that this new class V is related to class I but has evolved into a separate class through evolution (Jackson et al., 2002).

Class VI system was discovered on analysis of the P450-encoding gene (xplA) from *Rhodococcus rhodochrous* strain 11Y. Class VI is similar to class V in the sense that it consists of two separate proteins; a P450 haem domain fused to the C-terminus of the flavodoxin, and an NAD(PH) dependent flavodoxin reductase. *Arabidopsis thaliana* plant was engineered to express gene *xplA* and the result showed that the expressed P450 catalysed the degradation of hexahydro-1,3,5-trinitro-1,3,5-triazine (RDX), an explosive that is widely used by the military, in the absence of oxygen. *A. thaliana* expressing xplA gene was resistant to any

phototoxicity when grown in RDX contaminated soil, and moreover removed RDX from liquid media. This suggests that expressing xplA gene can be used in the bioremediation of RDX and other classes of this explosive (Rylott et al., 2006, Seth-Smith et al., 2002).

Class VII is a completely novel class according to the domain architecture, consisting of a single fused gene, which are classed as "self-sufficient" P450s, and they don't require external redox partners. The C-terminus of the P450 is fused to a phthalate dioxygenase reductase (PDOR), which is unlike the other P450s systems discovered so far. The reductase flavin cofactor, accepting electrons from NADPH and transferring them to the reductase, in this instance is the FMN instead of the FAD as in the other classes (Correll et al., 1992). The first described P450 belonging to class VII was CYP116B2 (P450RhF), isolated from *Rhodococcus* sp. strain NCIMB 9784. The reductase partner consists of three distinct domains: the flavin mononucleotide domain, the NAD(P)H binding domain, and the ferredoxin centre of the type 2Fe-2S. Recombinant expression of the P450 into *E. coli* performed O-dealkylation of the 7- ethoxycoumarin to 7-hydroxycoumarin in the absence of any redox partners (Roberts et al., 2002). Since then, an array of different reactions were confirmed for P450RhF, such as epoxidation, aromatic hydroxylation, dealkylation and asymmetric sulphoxidation confirming P450RhF promiscuous substrate activity (O'Reilly et al., 2013). Moreover, when self-sufficient P450RhF was expressed recombinantly in a whole cell system, the hydroxylation of diclofenac, an anti-inflammatory drug, to 5-hydroxydiclofenac, and further autooxidation to the quinonimine derivative was demonstrated. The autooxidation is enzymatic free and increases overtime, at elevated temperatures and high concentration of salts (Klenk et al., 2017).

Class VIII of P450 systems consist of a single peptide chain containing a P450 protein fused to a reductase like the eukaryotic P450s CPR, containing a FAD and FMN binding domain. The most intensively studied P450 of this class has become P450 BM3 (CYP102A1) from *Bacillus megaterium*. P450 BM3 was the first P450 discovered to have its P450 haem-containing oxygenase connected with a short linker to the NAD(P)H cytochrome P450 reductase containing FMN and FAD cofactors, making it a self-sufficient system. Due to its similarity in sequence, structure and function with the mammalian drug metabolising P450 enzymes of class II, and being easily produced and solubly expressed recombinantly, P450 BM3 has been subject to extensive research into its structure and mechanism and its associated domains (Narhi and Fulco, 1986). P450 BM3 has drawn a lot of biotechnological value as well, due to its role of metabolising alkanes, alkenes, and short chain of alkanoic acids through its intensive engineering. P450 BM3 catalyses the hydroxylation of long and medium chain fatty acids at positions ω-1, ω-2 and

ω-3, in the presence of NAD(P)H and oxygen. This was the highest catalytic reaction with Kcat = 17,000 min^{-1} (McLean et al., 2005).

Class VIII systems have been found in other prokaryotic and fungal species. *Bacillus subtilis* P450s CYP102A2 and CYP102A3 are homologues to BM3, and P450foxy, from fungal *Fusarium oxysporum* (CYP505A1), are some well-known examples of class VIII systems. The bacterial CYP102A2 and CYP102A3, and fungal CYP505A1 have high sequence identity with CYP102A1, and catalyse the hydroxylation of fatty acids at ω-1- ω-3. Both bacterial are water soluble resembling CYP102A, while fungal CYP505A1 resembles that of the eukaryotic membrane bound P450s (Kitazume et al., 2000).

Although class VIII is not as ubiquitous as class I and II, it has attracted a lot of biotechnological interest due to being one single fused protein used for catalytic activity, without the need of external redox partners (Durairaj et al., 2016, Kunst et al., 1997, Whitehouse et al., 2012b).

Class IX are a totally unique class in the sense that they do not require any redox partners for electron transfer from an NADPH reservoir. Instead, they consist of nitric oxide reductases. A perfect example of this unique class is the first P450 (CYP55) discovered in this class, the P450nor from the filamentous fungi *Fusarium oxysporum*. Amino acid sequence showed that they lack the N-terminus hydrophobic signal peptide that is required for membrane localisation in microsomal P450s, and any mitochondria-targeting sequence required for mitochondrial localisation. P450nor resembled the sequences of soluble bacterial P450s, instead of the eukaryotic yeast and fungal P450s, suggesting that this is the only soluble eukaryotic P450 (Kizawa et al., 1991). CYP55 is localised in the mitochondrial and cytoplasm of the eukaryotic P450. Its physiological role in the denitrification reaction is the nitrate respiration in the fungus under anaerobic conditions. P450nor catalyses the reduction of nitric oxide to nitrous oxide (Takaya et al., 1999).

Like class IX, class X systems are unique, as they do not require redox partners for the transport of the electrons, oxygen, or NAD(P)H electron reservoirs, but they utilise a different intramolecular system to catalyse the conversion of the substrates. Although they are identified as P450s, they can be classified into different classes of enzymes such as: allene oxide synthases (AOL, CYP74A), divinyl ether synthases (DES, CYP74D), hydroperoxide lyases (HPL, CYP74 B/C), prostacyclin synthases (PGIS) and thromboxane synthases (TBAX). All these are members of the CYP74 family are detected in the membranes of the chloroplasts (Froehlich et al., 2001). They utilise oxygen donors from acyl hyperoxide of their substrates, to

form new carbon-oxygen bonds in the products, which are consequently used as signals against pest attacks in plants (Hannemann et al., 2007).

1.2.7. P450 Kingdom and Their Physiological Function in Nature

P450s are responsible for the oxidation of up to 95% of all organic chemicals, and up to 75% of drug metabolites. As a result, the range of applications for P450s has expanded to include bioremediation, agriculture, bacteriology, and drug development (Rendic and Guengerich, 2015).

Prokaryotic P450s are cytosolic, soluble enzymes with a diverse range of catalytic activities including drug and fatty acid (Lamb and Waterman, 2013). P450cam from *Pseudomonas putida* was the first isolated prokaryotic P450 with a biological role of hydroxylating camphor to 7-hydroxycamphor (Keller et al., 1972), and the first to have its structure crystallised in 1985 by Poulos et al. The P450cam's crystallisation was likely influenced by its soluble nature and simplicity of manufacturing (Poulos et al., 1985). Both P450cam and P450 MB3 enzymes have been well studied and have provided a significant amount of information relating to prokaryotic P450s.

There are 602 prokaryotic P450 families, of which 2979 bacterial and 64 archaeal CYP sequences have been named, and more exist in private and public databases. There are > 760,000 bacterial CYPs already identified in the databases (https://www.ncbi.nlm.nih.gov/protein/?term=P450).

Three out of the 18 human CYP families, CYP1, CYP2 and CYP3 metabolise most of the drugs in use, and CYP3A4 alone is responsible for the metabolism of 27% of the drugs and 13% of general chemicals (Fisher et al., 2009, Rendic and Guengerich, 2015) .

Most eukaryotic P450 are membrane bound proteins of the endoplasmic reticulum (microsomal), although there are some eukaryotic P450s that are located in the matrix side of the mitochondria (mitochondrial), and a third type which is the cytosolic soluble P450s, albeit in less abundance. The eukaryotic P450 have a length of 480 to 560 amino acids (Danielson, 2002) .

Eukaryotic P450, and especially human P450s, are among the most studied P450 enzymes due to their role in therapeutics, drug and chemical degradation from the human body, and steroid hormone biosynthesis and procarcinogen activation

Different strategies exist for successful expression and catalytically active mammalian P450s, since they are membrane bound proteins, and hence difficult to work with (Williams et al., 2000).

On the other hand, the fungal family is extremely large and most diverse. There are between 1.0 and 1.5 million fungal species that exist in all temperature zones of the earth. They have a significant and far-reaching impact on the Earth's ecosystem with their biological roles ranging from production of food to other eukaryotes and humans to breaking down various biopolymers and other biological compounds in both dead and living hosts (Choi and Kim, 2017).

There are 369,808 P450 proteins to be found in fungi. The great diversity of fungal CYPs, which primarily resulted through gene duplication, is strongly related to the way that fungi live in their habitats. When it comes to diversified metabolism and fungal adaptation to particular ecological niches, CYPs in fungi play a variety of important and crucial roles (Chen et al., 2014).

Some fungi have at least one P450, while others have more than 200 P450s. *Schizosaccharomyces* species has only two and *Saccharomyces cerevisiae* having only three CYPs. In contrary, *Dendrothele bispora* has 564 CYPs and *Sphaerolobus stellata* has 601 CYPs (Nelson, 2018). According to the One Thousand Fungal Genomes Project (1KFG) from JGI only 7,925 P450s that have been assigned a CYP name spread across 805 families. Most fungal species occupy families CYP51 and CYP61 which are involved in sterol biosynthesis (Chen et al., 2014).

Plants have the most abundant P450 discovered in any species. After sequencing more than 1200 transcriptomes from algae and angiosperms, more than 171,940 plant CYP sequences were generated, according to the One Thousand Plants (1KP) initiative, which belong to 277 families. These sequences were further grouped into 9 plant taxa with some of them being algae, liverworts, angiosperms and gymnosperms (Nelson, 2018).

Up to 1% of P450 genes are estimated to exist in each plant genome. *Arabidopsis thaliana* on its own contains 245 CYPs, the largest single gene family in *Arabidopsis*. Only 60 of these have been actively characterised, with a large proportion remaining uncharacterised. The rice plant *Oryza sativa* contains 334 CYPs, 316 exist in grapevine *Vitis vinifera*, 372 in *Sorghum bicolor*, 312 in *Populus trichocarpa* and 332 in soybean *Glycine max* (Mizutani, 2012).

Plant P450s have been shown to engage in a wide range of biochemical pathways that leads to the production of primary and secondary metabolites such as terpenoids, alkaloids, phenylpropanoids and plant hormones, which help support their chemical defence mechanism (Mizutani, 2012). Land plant P450s have been divided in four categories depending on their functions in the plant. For example, category I is divided between three families, where CYP51 and CYP710 are involved in sterol biosynthesis, and CYP97 family that is important in light harvesting, photosynthesis and light protection, whereas category II is involved in biopolymer and other chemical synthesis, which helps plants survive and proliferate in their habitats (Mizutani, 2012). All these functions suggests that plant P450s can be used in the chemical biosynthesis and environmental bioremediation, which will be analysed in detail later.

The first viral P450 discovered was after the sequencing of the viral genome of the *Acanthamoeba polyphaga mimivirus*. *Acanthamoeba polyphaga mimivirus*, a huge virus that grows in amoeba, has a genome of 911 predicted coding sequences, of which only 298 have a known function (Lamb et al., 2009). Previously, 2180 viral genomes had been sequenced with no successful identification of P450s. Two putative P450s exist in this viral genome known as MIMI_L532 (YP_142886) and MIMI_L808 (YP_143162). YP_142886 showed 23% identity to bacterial P450s of *Chloroflexus aurantiacus* and *Streptomyces peucetius*, while YP_143162 showed 25% and 24% identity with nematode *Caenorhabditis briggsae* CYP37B1 and sea squirt *Ciona intestinalis*, respectively (Lamb et al., 2009). Moreover, YP_143162 showed a strong homology to CYP51 from different species (Lamb David et al., 2019), and encodes lanosterol demethylase, as does CYP51, which is essential in sterol biosynthesis (Aoyama et al., 1984).

Since then, more viral P450s have been discovered in a new order of giant viruses, the Megavirales. P450s have been discovered in land viruses, ocean, human viruses, such as *Ranid herpesvirus 3*, and even in phages, such as *Mycobacterium phage Adler* (Lamb David et al., 2019).

1.2.8. Important P450 applications

P450s are important enzymes used in drug metabolism, drug synthesis and active pharmaceutical ingredient (API) production. They metabolise drugs and chemicals in the human body, and eventually lead to their excretion through the kidneys (Peltoniemi, 2013). They are used in the synthesis of steroids, sterols, vitamins, fatty acids, and prostaglandins. Plant P450s have been found to convert terpenes, alkanes, and aromatic compounds, which are in high demand. Also, they are used in the bioremediation of land and ground water (Bernhardt, 2006).

Roughly, two-thirds of naturally occurring antibiotics, drugs precursor and other secondary metabolites, including antifungal drugs, antitumor agents and herbicides, are produced by Streptomyces species. CYPs with specific functions have been found in 21 different strains of Streptomyces, which are clustered in the macrolide antibiotic biosynthesis genes locus (Lamb et al., 2003). One important example of API production through biocatalysis is that of *Streptomyces carbophilus* P450sca, which was shown to catalyse the hydroxylation of mevastatin (ML-236B) to pravastatin, which can be used in the therapy of hypercholesterolemia, as it significantly reduces the cholesterol levels in the plasma (Serizawa and Matsuoka, 1991, Watanabe et al., 1995).

Another example is one of the two P450 genes in *Streptomyces griseolus*, CYP105A1 which catalyses the hydroxylation Vitamin D_3 at C-25 producing the intermediate 25-hdroxyvitamin D3, which is further hydroxylated to 1α, 25-dihydroxyvitamin D_3, in a two-step reaction (Sugimoto et al., 2008).

As mentioned previously, *Rhodococcus rhodochrous* strain 11Y was isolated from military land, and its P450, encoded by gene xplA, was capable of degrading the high explosive and major contaminant RDX, which is found in military bases, to nitrite, formaldehyde, and formate, with no ammonia or other dead-end metabolites. *R. rhodochrous* strain 11Y utilises RDX as a sole nitrogen growth source. This CYP can be used as a decontaminant in the highly contaminated military areas with this class of explosives (Seth-Smith et al., 2002).

Another example is that of self-sufficient BM3 from *Bacillus megaterium*, which metabolises some of the major environmental pollutants, the polycyclic aromatic hydrocarbons (PAHs), the polychlorinated biphenyls (PCBs) and the polychlorinated dibenzo-p-dioxins (PCDDs), and makes it a very important enzyme in the bioremediation industry (Kumar, 2010).BM3s ability to biotransform alkanes, fatty acids and aromatic compound without the need of any accessory proteins has made it a role model for engineering of self-sufficient P450 systems (McLean et al., 2005).

1.2.9. Several P450 engineering techniques that enhanced activity and solubility.

It is noteworthy that P450s have an important role in clinical pharmacology. Moreover, their widespread biotechnological applications means they are in high demand within the biocatalysis community (Xu and Du, 2018). However, their limited enzyme stability, substrate specificity, poor activity and requirement for electron supply cofactors, restriction their widespread application in biocatalysis at commercial scale. To circumvent these issues, enzyme engineering has been widely employed, with model enzymes (Rentmeister et al., 2011, Urlacher and Girhard, 2012, O'Reilly et al., 2011).

P450 BM3 was engineered through site directed mutagenesis to metabolise PAHs such as naphthalene, fluorene and acenaphthene, and the activity of the mutant was 4-fold higher than that of the wild type (Li et al., 2001).

Robin et al. (2011) fused different bacterial P450s to the reductase domain of the self-sufficient P450RhF from *Rhodococcus sp.* strain NCIMB 9784 using a short linker. P450RhF is a self-sufficient enzyme which comprises a P450 domain fused to a reductase domain containing FMN- and Fe/S and catalyses the dealkylation of of 7-ethoxycoumarin to 7-hydroxycoumarin (Robin et al., 2009).

Furthermore, Dodhia et al. (2006) fused three human cytochrome P450s to the oxidoreductase domain of the P450 BM3 and proved that a chimera of two different enzymes from different origins can work together and have the same catalytic activity as the native enzymes. Moreover, the chimeras were easily purified, without the need of detergent, properly folded and active enzymes, and their solubility was improved compared to the native ones. In addition, creating the P450-reductase chimeras can facilitate the high-throughput screening of the enzymes. With this in mind, all human cytochrome P450s can potentially be fused to different reductases to create a chimera, or even better, can be fused to the human reductase system, to create one single self- sufficient peptide that can be catalytically active. Such an approach was applied to one part my study (Dodhia et al., 2006).

1.2.10. Drug Metabolism and Interaction

Drug metabolism is initiated by P450s, primarily in the intestine and liver, although it can take place in other sites of the body including lungs, kidneys, brain and plasma (Peltoniemi, 2013). Expression of CYPs is induced or inhibited by numerous factors including induction by drugs, multiple genetic polymorphisms, age, sex, inflammation, or other substances. Multiallelic hereditary polymorphisms regarding CYPs 2A6, 2B6, 2D6 and 3A5, which are enormously influenced by ethnicity, are responsible for their function, and categorise them as poor, extensive, intermediate and ultrafast metabolisers (Zanger and Schwab, 2013).

1.2.11. P450s Mediated Metabolism of Ketamine

Ketamine (CI-581) is a phencyclidine derivative which functions as an antagonist of the N-methyl-D aspartate (NDMA) receptor in the central nervous system, and has been used in for the management of pain in both central and peripheral nervous system and anaesthetic or analgesic for the past 30 years (Kohrs and Durieux, 1998) (Sabia et al., 2011).

Ketamine is metabolised extensively by microsomal P450 enzymes. Studies with human liver microsomes suggest that ketamine is metabolized through human cytochrome P450 system via CYP3A subfamily of enzymes, which is the most expressed subfamily in the small intestine (Bezirtzoglou, 2012), and CYP2 family (Portmann et al., 2010a). According to Desta et al. (2012) (Figure 1-8) (R, S) ketamine can be hydroxylated to produce hydroxyketamines (HK), (2S,6S;2R,6R)-HK (5a) and (2S,6R;2R,6S)-HK (5b), with 5b further N-demethylated to produce (2S,5S;2R,5R)-hydroxynorketamine (HNK). Additionally, ketamine can be directly N-demethylated to produce (R, S)-norketamine, which can be further hydroxylated to produce diastereomeric HNK resulting from cyclohexanone ring hydroxylation at the carbon 6 (C6), C4, and C5 positions (Desta et al., 2012).

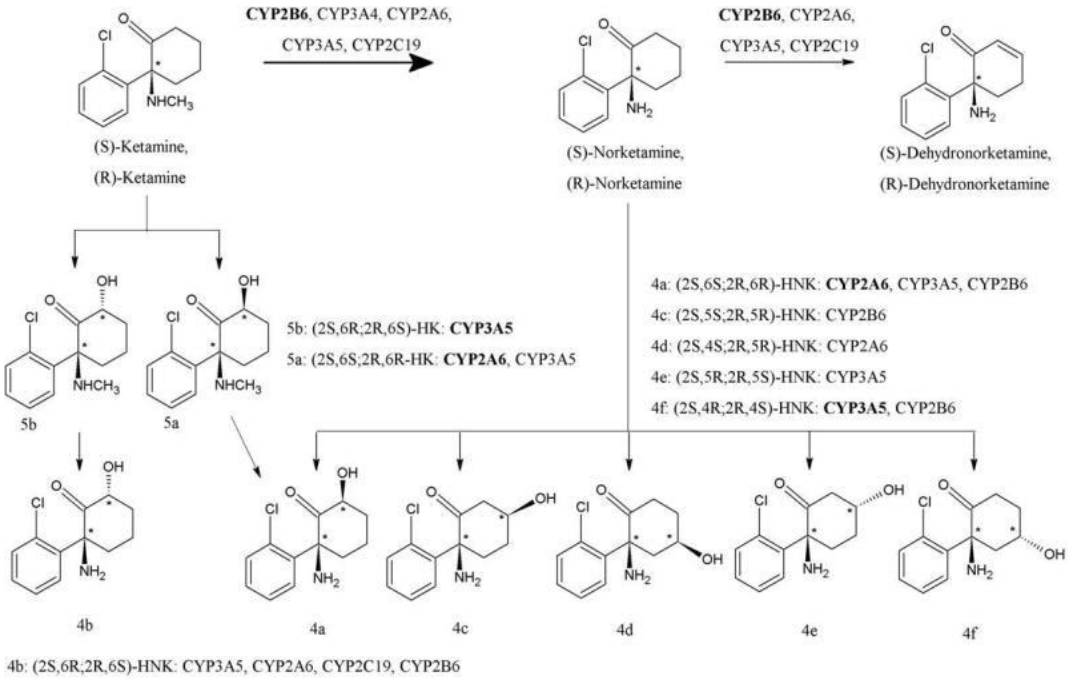


Figure 1-8 A very good summary of the metabolic pathway of (R, S)-Ketamine and the P450 enzymes involved in the metabolism of ketamine and norketamine by Desta et al, 2012. (R, S)-Ketamine was hydroxylated to 2S,6S;2R,6R- HK mainly by CYP2A6 (5a) which was further N-demethylated to 4a by CYP2B6, and to 2S,6R;2R,6S- HK by CYP3A4 and CYP3A5 enzyme (5b), and was N-demethylated to 2S, 6R;2R, 6S (4b) by CYP3A5, CYP2B6, CYP2C19 and CYP2B6. (R, S)-norketamine was hydroxylated to (2S,6S;2R,6R)-HNK by CYP2A6, CYP2B6 and CYP3A5 (4a); to (2S,5S;2R,5R)- HNK by CYP2B6 (4c); to (2S,4S;2R,5R)- HNK by CYP2A6 (4d); to (2S,5R;2R,5S)- HNK by CYP3A5 (4e); to (2S,4R;2R,4S)- HNK by, mainly CYP3A5, and CYP2B6 (4f) (Desta et al., 2012).

1.3. Biocatalytic Hydrolysis of Nitriles by Nitrilases

1.3.1. The Importance of Nitriles

Nitriles (or cyanides) are organic compounds that contain one or more cyanide groups (-CN) attached to a functional group, with the general formula being R-C≡N. The use of the cyano group as a radical acceptor in cascade reactions offers up a number of opportunities for the fast synthesis of carbocycles and heterocycles (Shen et al., 2021).

Nitrile compounds are naturally occurring in the environment, because they play a variety of roles as metabolites in numerous biological systems and industrial applications. They mainly occur as cyanoglycosides, cyanolipids, cyanohydrins. Many biological systems, including some plants, microbes, insects, and arthropods, create nitriles under biotic and abiotic stress conditions in the form of glycosides and cyanolipids, which are essential for plant-microbial interaction (Bhalla et al., 2018).

Even though most nitriles are extremely poisonous, mutagenic, and carcinogenic, because of their cyano group, hydrolysis, oxidation and chemical synthesis routes for these substances to obtain a wide range of beneficial amides, carboxylic acids, and ketones, such as pharmaceuticals, fine chemicals, agrochemicals, polyamides and polyesters, are well-known (Gong et al., 2012). Chemical processes frequently call for extreme conditions, such as the utilisation of metal catalysts, or strong acids or bases at high pressures and extreme temperatures, and require multiple steps reactions, which more than often result in poor selectivity, and the formation of side products. In contrast, biocatalytic processes for synthesis and hydrolysis of nitriles and nitrile intermediates has proved beneficial due to their efficient enantioselectivity, product purity and their environmentally friendly production methods (Wang, 2015). A well-known function is the bioremediation of extremely hazardous nitriles found in environmental waste and pollutants through nitrile-degrading enzyme, the nitrilases (Veselá et al., 2010).

1.3.2. Biological Synthesis and Biocatalytic Hydrolysis of Nitriles

In biological systems, such as bacteria and plants, nitriles are synthesised through two different pathways: either an aldoxime dehydratase, that catalyses the dehydration of the aldoxime to form a nitrile, or a hydroxynitrile lyase, that catalyses the formation of cyanohydrins from an aldehyde and a hydrogen cyanide (Figure 1-9). Cyanoglycosides and cyanolipids are immediately generated from cyanohydrins (Bhalla et al., 2018). An example is the aldoxime dehydratase from *Bacillus* sp. strain OxB-1, that catalyses the conversion of aldoxime to nitrile, which is the first step in aldoxime breakdown (Kato et al., 2000).

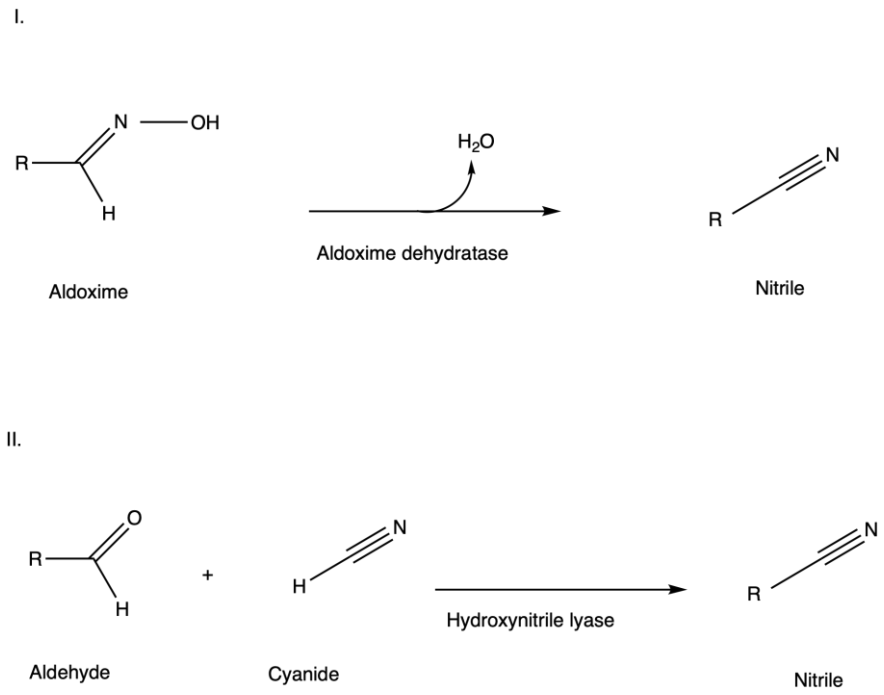


Figure 1-9 Chemical reaction portraying the two distinct pathways for the biological synthesis of nitriles. An aldoxime dehydratase catalyses the dehydration of an aldoxime, or an hydroxynitrile lyase forming a nitrile from the combination of an aldehyde and a hydrocyanide. Figure adapted from (Bhalla et al., 2018). Figure created with ChemDraw (https://perkinelmerinformatics.com/products/research/chemdraw).

There are also two unique pathways for the biocatalytic hydrolysis of nitriles (Figure 1-10). The nitrilases catalyse the direct hydrolysis of a nitrile to form the corresponding carboxylic acid with the release of ammonia in one step, or in two steps, a nitrile hydratase adds a molecular water to the carbon-nitrogen bond to create the corresponding amide, which the latter acts as a substrate for an amidase enzyme to be hydrolysed to the carboxylic acid product and the subsequent release of ammonia (Piotrowski, 2008, Black et al., 2015).

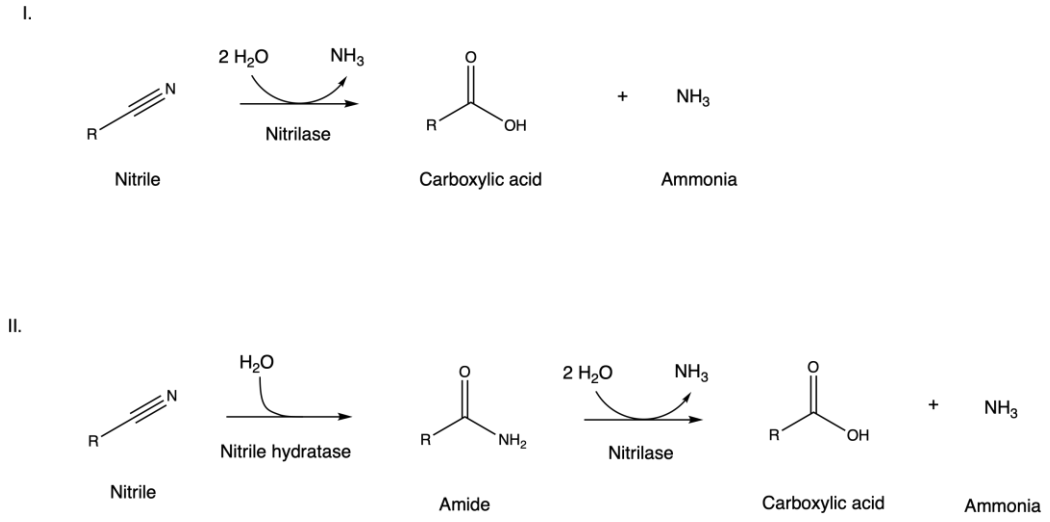


Figure 1-10 The two distinct pathways of hydrolysing a nitrile to its carboxylic acid with the release of release of ammonia using nitrilases and nitrile hydratases. Figure created with ChemDraw (https://perkinelmerinformatics.com/products/research/chemdraw).

1.3.3. Nitrilases

Nitrilases are naturally occurring enzymes in bacteria, filamentous fungi, yeasts, and plants. They are key enzymes due to their ability to transform nitriles into a variety of products and intermediates. In the decade between 2000 and 2010, the number of publications in nitrilase and nitrilase-facilitated biocatalysis accelerated 8 times higher (Gong et al., 2012). Nitrilases are from a big family known C–N hydrolases (Piotrowski, 2008). Nitrilases have regio- and stereo-selectivity and offer a more attractive route to the intermediate biosynthesis than their chemical synthesis counterparts, as they produce less harsh by-products and have less effect on the environment. Moreover, their biocatalytic hydrolysis occurs at environmental temperatures and near physiological pH, and in aqueous phase, in contrast to the conventional chemical approaches. As mentioned previously, nitrilases catalyse the conversion of nitriles directly in one, or indirectly with the help of nitrile hydratase, in two conventional steps to the corresponding carboxylic acids and ammonia (Singh et al., 2006, Gong et al., 2012).

The first nitrilase enzyme was discovered and purified from barley leaves in 1964 by Thimann and Mahadevan for its ability to convert indoloacetonitrile (IAN) to the plant growth hormone indoloacetic acid (IAA) and ammonia. Thimann and Mahadevan tested 29 plants from 21 families, and only members of the families Hordeum Vulgare (barley), Cruciferae (cabbage, cauliflower group and radish), Gramineae (grasses) and Musaceae (banana family) showed the relevant activity (Thimann and Mahadevan, 1964). Nearly two decades later, the biocatalytic conversion of a nitrile to the carboxylic acid though the synergistic function of an amidase and a nitrile hydratase was discovered. Asano et al, (1980) discovered the biocatalytic conversion of acetonitrile to acetic acid via an intermediate of acetamide was carried out by a nitrile hydratase in conjunction with an amidase in Arthobacter sp. J-1(Asano et al., 1980). Computational biology has advanced dramatically in recent years due to the exponential increase in the volume of biological data, and therefore more nitrilases from different species have been discovered with great potential in industrial chemical synthesis. According to the GenBank database, there are 6844 sequences which have been labelled either as nitrilases, or hypothetical proteins with a high degree of sequence similarity to known nitrilases. They belong to many different organisms, with the higher position occupied by eukaryotes with 4752 proteins, of which 1720 belong to fungi and 767 to plants, followed by bacteria with 1378 proteins. Archaea with 679 proteins and viruses with 34 proteins (https://www.ncbi.nlm.nih.gov/gene/?term=nitrilase).

1.3.4. Classification of Nitrilase Superfamily

The nitrilase superfamily, as proposed by Pace and Brenner in 2001, can be divided into 13 branches based on sequence similarity and the existence of extra domains. Eight branches appear to contain amidase or amide-condensation activities, and only one branch is known to have nitrilase activity, from a nitrile to the carboxylic acid and ammonia (Pace and Brenner, 2001).

The branches were classified according to the E-values generated from the alignment of the 176 identified sequences and the hallmark motifs around active-site residues. Values less than $1x\ 10^{-25}$ are classified into the same branch. Nitrile hydratases are not classified as members of the nitrilase superfamily, whereas amidases, due to the relation of the aliphatic amidases with the nitrilases, are included in the superfamily branches, although some amidases and thiol proteases that are not related to the nitrilase family (Pace and Brenner, 2001). Figure 1-11 highlights all the branches and domains of the nitrilase family. Branch 1 is occupied by nitrilases, branch 2, 3 and 4 by aliphatic amidases, N-terminal amidases and biotinidase, respectively, branches 5 and 6 are occupied by amidases carrying out the specific carbamylase reaction, whereas branches 7 and 8 are prokaryotic and eukaryotic NAD^+ synthetases, respectively. Branch 9 is occupied by apolipoprotein N- acyltransferases, which transfer a fatty acid from phospholipid to a polypeptide amino terminus, which is usually an amino-terminally diacylglyceride-modified cysteine. Branch 10 consists of Nit and NitFhit proteins which encode for diadenosine polyphosphate hydrolases, while branches 11 to 13 have unknown functions. The subcellular localization, selectivity, and cell biology of uncharacterized protein sequences can be predicted with the aid of genetic and biochemical analyses of the family member protein domains (Pace and Brenner, 2001).

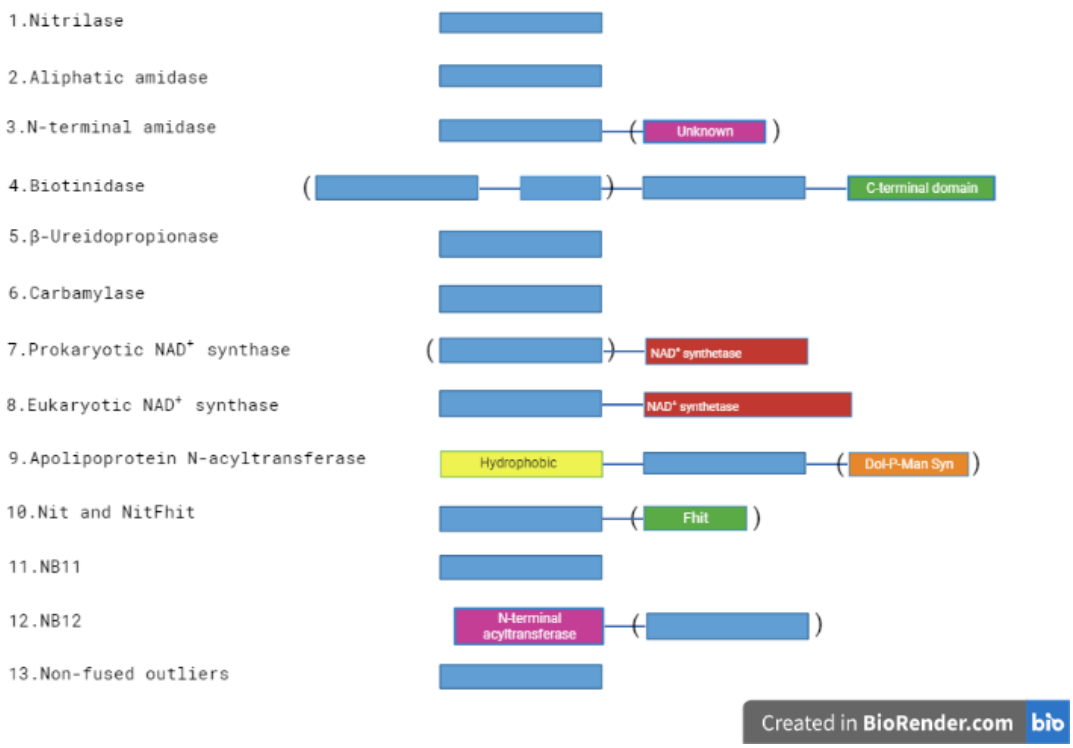


Figure 1-11 Illustration of all the branches that make up nitrilase superfamily, as well as all the individual domains on each branch. The domains in blue are nitrilase-related domains, while domains in brackets means they do not occur in each member of the branch, but only in some of them. Branch 3 has an extra, unknown domain, while branch 4 has a C-terminus domain. Branches 7 and 8 have the NAD^+ synthase domain, while branch 9 has a hydrophobic domain at the amino-terminus, and the nitrilase-related domain sometimes fused to a dolichol phosphate mannose synthetase at the C-terminus. Branch 10 consists of Nit and Fhit fused proteins, while branch 12 members are predicted to have functions as are fused to an acyltransferase at the amino terminus. Figure adapted from Pace and Brenner, 2001 and created with Biorender.com.

A recent, machine learning technique, which was developed by Sharma et al. (2018), was used to classify nitrilases as aromatic, aliphatic and aryl acetonitrile (Figure 1-12) depending in amino acid composition. The composition of aliphatic nitrilases was high in aliphatic amino acids (Ala (A), Gly (G), Leu (L), Ile (I), Val (V), Met (M) and Pro (P)), whereas aromatic nitrilases were rich in aromatic residues (Tyr (Y), Trp (W), His (H) and Phe (F)), which suggests that the activity of the nitrilases is dictated by the amino acids composition, according to the above. It is believed that this approach can help future scientists in predicting new nitrilase sequences as aliphatic or aromatic based on amino-acid composition (Sharma et al., 2018).

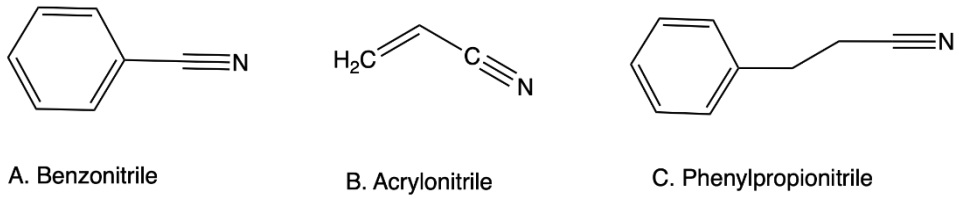


Figure 1-12 An example of aromatic nitrile (A), an aliphatic nitrile (B), and an arylacetonitrile (C). Figure drawn with Chemdraw.com

1.3.5. Nitrilase and Nitrile Hydratase Conserved Structure and Function

Structural studies on different organisms and strains have provided insight about the catalytic active site of the nitrilase enzyme superfamily and the conserved catalytic residues among them. Specifically, all bacterial nitrile hydratases are hetero-oligomeric assemblies of two types of subunits, most frequently dimers or tetramers, but they can potentially be bigger structures with up to 20 subunits (Yamada and Kobayashi, 1996). The active site of the nitrile hydratase almost always exists a metal ion in the form of non-haem Fe^{3+} and/or Co^{3+} in bacterial enzymes, while fungus *Myrrothecium verrucaria* contains Zn^{2+}(Maier-Greiner et al., 1991). This was demonstrated from studies in the three-dimensional structures of the nitrile hydratases of *Pseudomonas chlororaphis* B23 (Sugiura et al., 1987, Sakurai et al., 1988), *Rhodococcus sp*. R312 (Huang et al., 1997), *Rhodococcus sp*. N-774 (Nagashima et al., 1998) and *Pseudonocardia thermophila* JCM 3095 (Miyanaga et al., 2001). The non-haem iron centre in the active site of the enzymes was confirmed by electron spin resonance (ESR) to be low spin (Sugiura et al., 1987, Sakurai et al., 1988). From the crystal structure of nitrile hydratase of *Rhodococcus sp*. N-774 it was determined that the iron ion is coordinated by five ligands, three sulphur atoms of cysteine residues, of which cysteine residues Cys111 and Cys113 underwent posttranslational modification to become cysteine-sulfinic acid and cysteine-sulfenic acid, respectively, and two nitrogen atoms. Nitric oxide (NO), which binds reversibly to the sixth coordination site of the iron, likely the catalytic site, deactivates this enzyme, while in the absence of NO a hydroxide binds to that position, which functions as a nucleophile and attacks the nitrile carbon either directly or through an activated water molecule to form the amide. Tyr72, which is believed to be involved in substrate binding in the iron containing hydratase, is replaced by tryptophan residue in the cobalt containing enzyme (Nagashima et al., 1998) (Miyanaga et al., 2001).

In contrast, nitrilases lack a metal cofactor but are categorised as thiol enzymes as they contain a sulfhydryl group that is necessary for catalytic action. They are homo-oligomers or, less frequently, monomers with a molecular weight of about 40 kDa (O'Reilly and Turner, 2003). The covalent thioimidate complex is hydrolysed after being attacked by a sulfhydryl group of a cysteine residue through nucleophilic attack on the nitrile carbon atom (Kobayashi and Shimizu, 1994). According to Layh et al. (1998), the presence of thiol-binding substances like silver nitrate ($AgNO_3$) and copper sulphate ($CuSO_4$) can limit enzyme activity while thiol-reducing substances like dithiothreitol can increase it (Layh et al., 1998).

Nitrilases exhibit a multimeric alpha-beta-beta-alpha sandwich structure fold and create a super-sandwich scaffold α-β-β-α-α-β-β-α using a similar dimerization interface. They share a conserved catalytic trio of Glu54, Lys127, and Cys169 that is responsible for the covalent catalysis. The catalytic glutamic acid serves as the general base, catalytic lysine/glutamic acid as the general acid, and catalytic cysteine residue as the nucleophile (Martínková and Křen, 2010). The catalytic trio is shared by nearly all members of the nitrilase superfamily, and every catalytic process seems to involve a thiol acyl enzyme intermediate (Pace and Brenner, 2001). Additionally, after a comparative Bayes analysis was performed between branch 1 and 2 of the nitrilases to identify adaptive evolution and conserved residues, three conserved residues that are near the catalytic triad, Thr41, Gln157, and Tyr184, were discovered buried within the protein, and is hypothesised that these residues contribute to changes in substrate selectivity (Podar et al., 2005). Other conserved catalytic residues between the members of the superfamily are Tyr125, His129, Tyr170, Asp171, Arg173 and Phe174, as described for the active site of the Nit domain from the NitFhit enzyme of *Caenorhabditis elegans* (Kumari and Poddar, 2019). A conserved Cys-Trp-Glu motif located near the Cys169 residue of the catalytic triad helps to differentiate the nitrilase branch from other superfamily members, (Howden and Preston, 2009). The importance of the cysteine as catalytic residue was confirmed by mutation studies performed in *Alcaligenes faecalis* JM3 and *Arabidopsis thaliana*, where the mutant enzymes lost catalytic activity (Kobayashi et al., 1993, Vorwerk et al., 2001). Additionally, mutation of Lys129, in *Pseudomonas stutzeri* AK61 resulted in loss of activity as well, (O'Reilly and Turner, 2003).

Mutation studies performed by Kaushik (2012) with *Pseudomonas putida* also identified conserved catalytic triad residues, and even discovered another tetrad, Pro-Gly-Tyr-Pro, to be largely conserved in additional nitrilase superfamily members, such as cyanide hydratases and dehydratases, which suggests that they are likely key or evolutionary residues. Additionally, the catalytic triad was discovered to be conserved even for nitrilases with sequence identities as low as 27% to those of known nitrilase (Kaushik et al., 2012).

Most organisms have been shown to produce either a nitrilase or nitrile hydratase. That is not the case for *Rhodococcus rhodochrous* strains LL 100-21 and J1, which hold biotransformation routes that are both nitrilase and nitrile hydratase-catalysed (Wang, 2005). *R. rhodochrous* LL 100-21 was grown in different substrates such as acetonitrile, acetamide and other nitriles, amides and acids to utilise as a sole source of carbon and/or nitrogen, and both enzymes, the nitrile hydratase and amidase were induced. Both acetonitrilase and acetamidase activities were induced during growth on either acetonitrile or acetamide. With the exception of

propionitrile and propionamide, simple aliphatic amides may co-induce the hydratase and amidase enzymes at levels comparable to those produced by growing on the respective nitriles, suggesting that the biosynthetic pathway of these two enzymes is actually associated (Linton and Knowles, 1986, Collins and Knowles, 1983). In contrast, ε-caprolactam, isovaleronitrile and isobutyronitrile are great inducers to produce a nitrilase in *Rhodococcus rhodochrous* J1 cells when grown in the presence of the substrate (Nagasawa et al., 1988, Nagasawa et al., 1990).

Additionally, there exist enzymes that contain both nitrilase and nitrile hydratase activity. Such an example is enzymes of the type NIT4 (nitrilase-like protein 4), which have both nitrilase and nitrile hydratase activities. They have been found in many species in plants such as *Arabidopsis thaliana* and *Lupinus angustifolius* (Piotrowski *et al*., 2001).

1.3.6. Nitrilase Enzymes in the Environment

The nitrilase enzymes that have been investigated so far exhibit a wide variety of biochemical traits. Particularly, the considerable variations in substrate specificity of the enzymes. In nature, nitrilases have an important role in plant and microbe interactions such as defence mechanisms, nitrogen utilisation as carbon sources, plant hormone synthesis, and detoxification mechanisms (Howden and Preston, 2009).

1.3.7. Nitrilase Enzymes in Plants

Perhaps the most thoroughly studied nitrilases in terms of their biological roles are those found in plants. In plants, nitriles can exist in the intermediate form of cyanolipids which can be hydrolysed at the final step to form hydrogen cyanide, which plants could utilise as a protective substance against herbivores (Legras et al., 1990). Moreover, cyanogenic glycosides and glucosinolates are used as defence chemicals to shield plants from herbivore attack, or in the plant hormone biosynthesis, and their metabolism in plants results in the production of nitrile molecules (Ludwig-Müller and Cohen, 2002).

According to their affinity for certain substrates, plant nitrilases can be divided into two groups: the arylacetonitrilases with strong affinity toward arylacetonitriles, such as NIT1, NIT2 and NIT3 from *A. thaliana*, or the NIT4-type enzymes which are nitrilases that have strong catalytic activity towards b-cyano-L-alanine to produce aspartic acid, asparagine and ammonia, (Howden and Preston, 2009).

1.3.8. Nitrilase Enzymes in Bacteria

The first bacterial nitrilase to be discovered was the rinicine nitrilase from a *Pseudomonas* species, isolated from soil in the 1960s (Robinson and Hook, 1964). Since then, have been extensively studied and many bacterial nitrilases have been discovered belonging to the genera *Acinetobacter, Alcaligenes, Bacillus, Corynebacterium, Nocardia, Pseudomonas, Rhodobacter and Rhodococcus* with broad substrate specificity and applications (Gong et al., 2012) (Chhiba-Govindjee et al., 2019). It has been demonstrated that microbial nitrilases can construct homo-oligomeric spirals. For example, *Rhodococcus rhodochrous* J1 nitrilase, which is the enzyme best described for the transformation of acrylonitrile and 3-cyanopyridine into acrylic acid and nicotinic acid, respectively, exists in the form of inactive dimer, which can become active during a C-terminal autolytic cleavage, or in a substrate-dependent manner in the presence of benzonitrile, oligomerising to create active spirals. Most bacterial enzymes are inactive dimers (Thuku et al., 2009). It is common to find bacterial nitrilase genes clustered alongside other genes important to specific metabolic pathways, with the nitrilase substrate selectivity suited for the function of the cluster (Chhiba-Govindjee et al., 2019).

1.3.9. Nitrilase Enzymes in Filamentous Fungi

Numerous fungal strains from the genera of *Aspergillus, Gibberella, Fusarium* and *Penicillium* have reported nitrilase activity towards indole-3-acetonitrile which transform to indole-3-acetic acid, although fungal nitrilases were not as well studied as the bacterial ones until recent years (Gong et al., 2012). The nitrilase from *Fusarium solani* IMI 196840, was the first to be discovered in 1977. It can hydrolyse the herbicide bromoxynil, and the benzonitrile to produce benzoate. The organism grew in the presence of benzonitrile and utilised it as sole carbon and nitrogen source, which indicates its role and contribution significantly

in the biodegradation of several nitrilic herbicides in the environment (Harper, 1977). The first nitrilase from *Aspergillus niger* K10 strain was isolated in 2006, the third nitrilase from the fungi species to be purified and characterised since the *Fusarium solani* IMI 196840 nitrilase discovered in 1977 (Kaplan et al., 2006). The nitrilase showed affinity towards seven different aromatic nitriles, and was recombinantly expressed, fully purified and characterised in 2011, with best substrates being benzonitrile and 4-cyanopyridine. This was the first fungal nitrilase to be recombinantly expressed (excluding cyanide hydratases) (Kaplan et al., 2011). Since then, numerous fungal nitrilases have been recombinantly expressed and characterised, such as nitrilases from *Auricularia, Macrophomina, Nectria, Neurospora, Talaromyces, Trichoderma* and *Trichpophyton* (Martínková, 2019).

1.3.10. Nitrilase Enzymes in Yeast

More than 60 yeast strains with nitrilase and/or nitrile hydratase activities have been reported so far isolated from fermented foods, soil and cyanide contaminated bioreactors that can metabolise nitriles. Species such as *Candida, Pichia, Saccharomyces,Hanseniaspora, Debaryomyces, Geotrichum, Williopsis, Torulopsis, Exophiala, Kluyveromyces, Aureobasidium, Cryptococcus,* and *Rhodotorula* use various aromatic, cyclic, and heterocyclic nitriles as a source of nitrogen (Brewis et al., 1995) (Bhalla et al., 2009). Acidic environments are typically favourable for yeast nitrilase-mediated biotransformation which benefited the hydroxylation of hydroxynitriles and aminonitriles under acidic conditions, as these substrates spontaneously disintegrate under neutral conditions, but are more stable at an acidic pH (Gong et al., 2012). For example, *Exophiala oligosperma* R1 from black yeast was grown at pH 4.0 in the presence of phenylacetonitrile, the only source of nitrogen and carbon. The produced nitrilase hydrolysed phenacetylnitrile to phenylacetic acid and 2-hydroxyphenylacetic acid. This benefited the reaction massively, as phenylacetonitrile is not stable under physiological pH (Rustler and Stolz, 2007).

1.3.11. Application of Nitrilases in Industry

Growing interest has been shown in using nitrile converting biocatalysts for the detoxification of nitriles and their conversion to fine chemicals, pharmaceuticals, important chemical intermediates, and food additives. The first procedure that used nitrile-converting biocatalysts on a large scale was the enzymatic synthesis of acrylamide from acrylonitrile utilising entire cells of *Pseudomonas chlororaphis* B23, *Rhodococcus sp*. N-774, and *Rhodococcus rhodochrous* J1 (Mylerova and Martinkova, 2003).

Different microbial nitrilases have been used in the bioremediation and decontamination of soil and water from nitriles and cyanides. These organisms live in contaminated environments and benefit from the contaminant by utilising them as carbon or nitrogen sources. Such an example is the cyanide hydratase from *Fusarium solani* which may hydrolyse free or metal-complexed cyanide, allowing this fungus to survive doses of cyanide that would be hazardous to other organisms (Barclay et al., 1998). Cyanide can also be hydrolysed by cyanide dihydratases, such as cyanide dihydratases from *Pseudomonas stutzeri* AK61 (Watanabe et al., 1998). (Cluness et al., 1993) Both the cyanide dihydratases and cyanide hydratases belong to nitrilases group of enzymes (Sewell et al., 2003).

Another use of the biocatalytic conversion of nitriles is that of the production of food supplements and vitamins. An example is the production of vitamin B3 also known as niacin or nicotinic acid that people cannot adequately synthesise on their own. Nicotinic acid is added to animal feed but is also utilised in medicine for the treatment of different ailments such as cholesterol, diabetes, and cataract among others. The amidase and nitrile hydratase from *Microbacterium imperiale* CBS 498-74 were used to produce nicotinic acid from 3-cyanopyridine in a cascade reaction with the intermediate product being nicotinamide (Cantarella et al., 2011). In addition, nicotinic acid is employed in pollution control as a deodorant for waste gases and air as well as a biostimulator for the production of activated sludge (Mathew et al., 1988).

Nitrile converting biocatalysts have been successfully used in the formation of fine chemicals and drug intermediates. A recent example is the enzyme AaNIT from *Arabis alpina*, which is a Nit type enzyme. AaNIT exhibited high catalytic activity against isobutylsuccinonitrile, which was hydrolysed to the (S)-3-cyano-5-methylhexanoic acid, but the formation of bioproduct (S)-3-cyano-5-methyl hexanoic amide was detected as well. This was due to AaNIT exhibiting both nitrilase and nitrile hydratase activities, and hence when used in combination with an amidase from *Pantoea* sp., the biproduct was eliminated in a one-pot cascade

reaction (Zhang et al., 2019). The product, (S)-3-cyano-5-methylhexanoic acid, is a chiral intermediate of the drug pregabalin, which is used for the treatment of anxiety, social phobia, post-herpetic neuralgia, and diabetic peripheral neuropathy, among other nervous system diseases. Pregabalin had become the focus of research in the pharmaceutical industry due to its broad application with different strategies being employed for its synthesis, but always ending in low yield. The discovery and generation of N258D, a mutant of AaNIT, exhibited excellent enantioselectivity towards the substrate, and resulted in high yield product with high enantiomeric excess of 99.5% (Zhang et al., 2019). These are only a few examples that highlight of the importance of the nitrilases and other nitrile converting enzymes in biocatalytic processes in the industrial world.

1.4. Aims and Objectives

Cytochromes P450s offer a fascinating alternative to the current techniques for developing APIs, fine chemicals, and creating metabolite standards. A variety of biocatalytic applications can benefit from the superfamily's amazing diversity in substrate range and catalytic activities. The diversity demonstrated by P450s in the substrate world is still largely unexplored. It has long been assumed that P450 enzymes metabolise norketamine to hydroxynorketamine, although it hasn't been proved yet. Currently, hydroxynorketamine is synthesised chemically, and although adjacent to an existing ketone, oxidation at the 5-position requires three chemical stages with an overall yield of approximately 50% (Figure 1-13). Since 2S, 6S; 2R, 6R-hydroxynorketamine is a known major metabolite of ketamine with oxidation and demethylation catalysed by P450 CYP2A6, CYP2B6 and CYP3A5, it is therefore reasonable to assume that these P450 enzymes could also hydrolyse norketamine in a single step. Various cloning and expression approaches were employed in this study in an effort to boost the expression catalytic activity of the P450 enzymes.

Moreover, the use of metagenomics for obtaining protein sequence data and, ultimately, novel putative biocatalysts is swiftly emerging as a technology of significant economic and scholarly value within the biocatalysis field. Through mining the metagenome library available at Prozomix Ltd (an enzyme manufacturer), two metagenome genes were identified S0zJJln2Kzf8zW_C33315_O22890198 (198) and S0zJJln2Kzf8zW_C5419_O20953914 (914) that shared homology 20 % and 24 % with the human gene CP3A5, respectively. Due to their homology to CYP3A5 both metagenome genes were hypothesised to metabolise ketamine to hydroxyketamine, and/ or norketamine to hydroxynorketamine in a single step.

Additionally, nitriles, which are intermediates in the drug manufacturing, are hydrolysed by nitrilase enzymes into the equivalent carboxylic acid and ammonia. Six nitriles, whose chemical synthesis had repeatedly failed, were approached enzymatically. More than 20 nitrilases were employed in biotransformation assays against the panel of 6 nitriles to determine if their corresponding carboxylic acid products could be achieved through biological routes.

Finally, our aim was to develop biologically catalysed routes to the above compounds that were of interest to Quotient Sciences (previously Arcinova, a contract development and manufacturing organization), and more generally, to develop enzymes and methodology to provide a toolbox for remote hydroxylation.

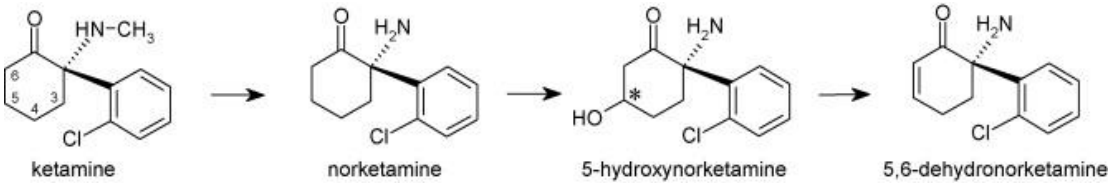


Figure 1-13 Chemical structures of ketamine, norketamine, 5-hydroxynorketamine and 5,6-dehydronorketamine. As seen in the structures norketamine is hydroxylated to hydroxynorketamine at the 5-carbon in the cyclohexanone ring (Portmann et al., 2010b) .

2. P450 Materials and methods

2.1. Materials

2.1.1. E. coli Cloning and Expression Strains

- *E. coli* TOP10 (Invitrogen™) was used as the cloning host and has the following genotype.

  F- mcrA Δ(mrr-hsdRMS-mcrBC) φ80lacZΔM15 ΔlacX74 nupG recA1 araD139 Δ (ara- leu)7697 galE15 galK16 rpsL (Str^R) endA1 λ$^-$

- E. coli C41(DE3) (Novagen) was used as the expression host and has the following genotype.

  F – ompT hsdSB (rB- mB-) gal dcm (DE3)

- E. coli C43(DE3) (Novagen) was used as the expression host and has the following genotype.

  F– ompT gal dcm hsdSB(rB- mB-)(DE3)

- E. coli BL21(DE3) (Novagen) was used as the expression host and has the following genotype.

  E. coli str. B F– ompT gal dcm lon hsdSB(rB–mB–) λ (DE3 [lacI lacUV5-T7p07 ind1 sam7 nin5]) [malB+] K-12(λS)

- E. coli Tuner cells (MERCK) was used as the expression host and has the following genotype.

  F–ompT hsdSB (rB– mB–) gal dcm lacY1(DE3)

2.1.2. Vectors

All vectors, stocks and working antibiotic concentrations, and media are outlined below in Table 2-1, Table 2-2, Table 2-3, Table 2-4.

Table 2-1 vectors used throughout the study.

Vectors	Supplier
pET-22b	Novagen®
pET-28a	Novagen®
pET-29b	Novagen®
pCWOri+	Addgene
pCDFDuet-1	Novagen®

2.1.3. Antibiotic Stocks

All antibiotic stocks were prepared with sterilised 18.2 MΩ/cm H_2O obtained from a Merck Millipore Milli-Q® water system. Ampicillin, chloramphenicol, spectinomycin and kanamycin stocks were prepared to at least 1000 x working concentration.

Table 2-2 Antibiotics used throughout the study.

Antibiotic used	Stock solution	Working concentration
ampicillin	10 mg/ mL	100 µg/ mL
kanamycin	10 mg/ mL	50 µg/ mL
chloramphenicol	34 mg	34 µg/ mL
spectinomycin	150 mg	120 µg/ mL

2.1.4. Sterilisation

Unless stated otherwise, all solutions, apparatus and consumable items were sterilised by autoclaving using a GEMINI Rodwell Scientific Instruments MP 25 Control autoclave at 126 ˚C for 20 min.

2.1.5. Bacterial Growth Media

All growth media were prepared using 18.2 MΩ/cm deionised H_2O, the pH adjusted to 7.2 with 1 M NaOH or 1 M HCl as necessary, and then sterilised by autoclaving at 126 °C for 20 minutes. Liquid growth media were stored at room temperature and solid growth media at 4°C.

IPTG was dissolved in sterile 18.2 MΩ/cm H_2O at 1000 times the working concentration (1 M) and stored at -20 ºC. A final concentration of 1 mM was employed to induce protein expression. *For Tuner cells multiple IPTG concentrations were trialled.

Table 2-3 LB (Luria-Bertani) Miller broth – (Merck)

10 g	Tryptone
5 g	Yeast Extract
10 g	NaCl

Table 2-4 2 x YT per litre (Merck)

16 g	Tryptone
10 g	Yeast Extract
5 g	NaCl

2.1.6. Reagents for Molecular Biology

DNA Standard Ladder

Hypperlader™ 1 kb (New England Biolabs®) was used throughout the study and comprised the following standard sizes; 0.2 kb, 0.4 kb, 0.6 kb, 0.8 kb ,1 kb, 1.5 kb, 2 kb, 2,5 kb, 3 kb, 4 kb, 5 kb, 6 kb, 8 kb and 10 kb.

Q5 High-Fidelity DNA Polymerase 2x Master Mix

The Q5 DNA Polymerase is supplied at a 2 x concentration. When used at the recommended 1 x final concentration the Q5 High-Fidelity Master Mix contains 2 mM Mg^{++}, 200 µM of each deoxynucleotide (dNTP) and 1 x final concentration broad-use buffer. All the buffers used for PCR and ligation are summarised below in Table 2-5, Table 2-6, Table 2-7.

Table 2-5 10 x KOD reaction buffer without Mg^{2+} (pH8.8 at 25 °C) (components given for 1 x concentration)

1.2 M Tris-HCl
100 mM KCl
60 mM $(NH_4)_2SO_4$
1% Triton X-100
0.01% BSA

Table 2-6 One Taq DNA polymerase 2 x Master Mix (pH 8.9 at 25 °C (components given for 1 x concentration)

20 mM Tris-HCl
22 mM KCl
22 mM NH_4Cl
1.8 mM $MgCl_2$
5% Glycerol
0.05% Tween® 20
0.06% IGEPAL® CA-630

0.2 mM dNTPs
25 units/ml OneTaq® DNA Polymerase

Table 2-7 10 x T4 DNA ligase buffer (pH 7.5 at 25°C) (components given for 1 x concentration)

50 mM Tris-HCl
10 mM $MgCl_2$
1 mM ATP
10 mM DTT

Ethidium bromide stain solution

Ethidium bromide was used at working concentration of 250 µL of 100 mg/mL stock solution (Sigma) / 500 mL distilled H_2O. In Table 2-8 are shown the components of 50 x stock TAE. 1% agarose was used in 1x TAE to make up agarose gels.

Table 2-8 TAE 50 x stock concentration (1 x Working Concentration diluted with 18.2 MΩ/cm H_2O) Per litre

242.0 g	Tris base (Melford)
100 mL	50 mM EDTA (pH 8.0) (Sigma)
57.1 mL	1 M Glacial acetic acid (Sigma)

2.1.7. Protein Purification Buffers

In Table 2-9, Table 2-10 and Table 2-11 are given the purification buffers used.

Table 2-9 Bio-Rad Econo-Pac gravity columns Purification Buffers

Start Buffer (10 mM)	Wash Buffer	Elution Buffer (100 mM)	Elution Buffer (200 mM)	Elution Buffer (300 mM)	Elution Buffer (400 mM)	Elution Buffer (500 mM)
20 mM Na_2HPO_4	20 mM Na_2HPO_4	20 mM Na_2HPO_4	20 mM Na_2HPO_4	20 mM Na_2HPO_4	20 mM Na_2HPO_4	20 mM Na_2HPO_4
10 mM Imidazole	20 mM Imidazole	100 mM Imidazole	200 mM Imidazole	300 mM Imidazole	400 mM Imidazole	500 mM Imidazole
500 mM NaCl	500 mM NaCl	500 mM NaCl	500 mM NaCl	500 mM NaCl	500 mM NaCl	500 mM NaCl

Table 2-10 AKTA prime plus purification buffers.

Buffers
Start Buffer
100 mM EDTA
18.2 MΩ/cm H_2O
20% Ethanol
100 mM $NiSO_4$
Elution Buffer

Table 2-11 AKTA explorer Size Exclusion Chromatography purification buffers

20 mM	NA_2HPO_4 and 150 mM NaCl, pH 7.4

2.1.8. Protein Assay buffers

Bradford reagent assay protein standards

To determine the protein concentration, a Bradford Assay was performed. Protein standards were performed ranging from 0.1 – 1.4 mg/mL (Table 2-12). Table 2-13 and Table 2-14 summarise the CO assay buffer components and Whole cell biotransformation buffer components, respectively.

Table 2-12 Bradford Reagent Assay

Tube number	BSA (mg/mL)	Stock (µL)	Tris 20 mM solution (µL)
Blank	0	0	1000
1	0.1	10	990
2	0.2	20	980
3	0.3	30	970
4	0.4	40	960
5	0.5	50	950
6	0.6	60	940
7	0.7	70	930
8	0.8	80	920
9	0.9	90	910
10	1.0	100	900
11	1.1	110	890
12	1.2	120	880
13	1.3	130	870
14	1.4	140	860

Table 2-13 CO assay Spec Buffer (pH 7.4)

20 mL	Glycerol (40%)
10 mL	Tris (1 M)
10 mL	100 mM CHAPS
1 mL	100 mM EDTA
9 mL	Sterile 18.2 MΩ/cm water

Table 2-14 Whole Cell Biotransformation (WCB) Screening Buffer

Ingredients	Volume used from stock conc.	Stock concentration
Potassium phosphate	100 µL	0.5 mM
Acetonitrile (ACN)	50 µL	Analytical gradient
Glycerol	40 µL	50%
Substrate	100 µL	100 µg/mL for MS, or 5 mg/mL for HPLC
NADPH	25 µL	10 mg/mL
Distilled sterile H_2O	685 µL	-

2.1.9. SDS-PAGE (Sodium Dodecyl Sulphate–Polyacrylamide Gel Electrophoresis)

All expressed enzymes were separated in SDS-PAGE gels. PageRuler Prestained and/or unstained protein ladder (ThermoFisher) of sizes 250, 130, 100, 70, 55, 35, 25, 15, 10 kDa were used throughout the study. 15 µL of protein and 5 µL SDS-PAGE loading buffer were added to a 1.5 mL microcentrifuge tube. The samples were mixed briefly by vortexing, pulse spun, boiled for 3 minutes, and separated on SDS-PAGE gels. Tables 2-15 to 2-26 summarise all the buffers used for SDS gels.

Table 2-15 SDS-PAGE Loading Buffer 1 M

0.6 ml	Tris-HCl pH 6.8 (60 mM)
5 mL	50% Glycerol
2 mL	10% (w/v) SDS
0.5 mL	14.4 mM β-mercaptoethanol
1 mL	1% Bromophenol Blue
0.9 mL	18.2 MΩ/cm H_2O

Table 2-16 SDS-PAGE Cracking Buffer (24%)

7.6 mL	SDS-PAGE Loading Buffer
2.4 g	Urea

Table 2-17 1 M Tris-HCl buffer pH 6.8

121.14 g	Tris Base
Adjust pH	with HCl
Up to 1 L	18.2 MΩ/cm H_2O

Table 2-18 2 M Tris-HCl buffer (pH 8.0)

242.2 g	Tris Base
Adjust pH	with HCl
Up to 1 L	18.2 MΩ/cm H_2O

Table 2-19 10 X Stock SDS-PAGE Running Buffer

30.3 g	Tris- HCl (pH 8.0)
144.0 g	Glycine
10 g	SDS
Up to 1 L	18.2 MΩ/cm H_2O

Table 2-20 Resolving Buffer (Buffer B)

325 mL	2 M Tris-HCl pH 8.8
20 mL	10% (w/v) SDS
155 mL	H_2O

Table 2-21 Stacking Buffer (Buffer C)

250 mL	1 M Tris-HCl pH 6.8
20 mL	10% (w/v) SDS
230 mL	H_2O

Table 2-22 APS 10% (w/v)

1 g	APS
10 mL	H_2O

Table 2-23 Resolving 12% Polyacrylamide Gel

3.0 mL	40% Acrylamide/Bis solution (37:5:1)
2.5 mL	Buffer B
4.5 mL	H_2O
100 µL	APS 10%
20 µL	TEMED

Table 2-24 Stacking 12% Polyacrylamide Gel

0.5 mL	40% Acrylamide/Bis solution (29:1)
1 mL	Buffer B
2.5 mL	H_2O
30 µL	APS 10%
10 µL	TEMED

Table 2-25 Coomassie Blue Stain Solution

1.0 g	Coomassie brilliant blue R-250 (Sigma)
400 mL	Methanol
100 mL	Acetic acid
Up to 1 L	18.2 MΩ/cm H_2O

Table 2-26 Destain Solution Per litre

400 mL	Methanol
100 mL	Acetic acid (glacial)
Up to 1 L	18.2 MΩ/cm H_2O

2.2. Bioinformatics

All the bioinformatic tools and databases used are listed in Table 2-27.

Table 2-27 Bioinformatic analysis tools and applications utilised in this study.

Bioinformatics Tool	URL	Application
National Center for Biotechnology Information (NCBI)	http://www.ncbi.nlm.nih.gov/	Genome/metagenome sequence data and multiple sequence alignments (BLAST).
Integrated DNA Technologies (IDT)	https://eu.idtdna.com/pages	Create genomic constructs with sequence-verified gBlocks fragments, and codon optimisation tool.
Candida Genome Database (CGD)	http://www.candidagenome.org/	Primer Design tool.
Webcutter 2.0	http://rna.lundberg.gu.se/cutter2/	Restriction mapping genes of interest.
Uniprot	http://www.uniprot.org/	Accessing protein sequences with functional information.
ClustalW	https://www.genome.jp/tools-bin/clustalw	Multiple protein sequence (≤ 500 sequences) alignments/ dendrograms.
ExPASy Proteomics Server	http://www.expasy.org/	Access to general protein bioinformatics tools i.e., Translate and ProtParam.
SignalP 5.0 Server	https://services.healthtech.dtu.dk/service.php?SignalP-5.0	Prediction of the presence of signal peptide and the cleavage sites location within amino acid sequences of interest.
Pfam	https://pfam.xfam.org/	Protein domain prediction server based on multiple sequence alignments and HMM.
Phyre2	http://www.sbg.bio.ic.ac.uk/~phyre2/html/page.cgi?id=index	Protein fold and structure prediction server.
PyMOL	https://pymol.org/2/	Protein structure visualisation system.
ProzOMIGO™ BLAST	http://prozomigo.com/	Basic Local Alignment Search Tool used for identification of protein sequences homologous to of the target human genes.

2.3. Molecular Biology

2.3.1. PCR

Two DNA polymerases were utilised throughout the study to generate PCR product, KOD hot-start DNA-polymerase (Invitrogen) and Q5® High-Fidelity 2XMaster Mix (New England Biolabs, NEB). The PCR reaction reagents and annealing temperatures of both KOD and Q5 polymerases are outlined in the Tables below (Table 2-28, Table 2-29, Table 2-30 and Table 2-31. Colony PCR was performed in colonies to confirm the recombinants, and OneTaq® polymerase enzyme was used. The reaction reagents and annealing temperatures of OneTaq® polymerase are summarised below in Table 2-32 and Table 2-33, respectively. An Eppendorf MasterCycler® thermocycler was used for all PCR reactions. All PCRs were run in a 20-degree gradient format, with the calculated annealing temperature and minus and plus 10 degrees from the annealing temperature used. Primary and secondary temperatures were also used for annealing. The primary annealing temperature used, was calculated from the Tm values of the primers without any 5' sequence that was added to the primer that does not anneal to the original template, and the secondary temperature being the one calculated from the full primer sequences. Annealing temperatures were 5 ˚C less than the lowest Tm value of the primer pair. PCR products were pooled, 6x loading dye was added to it, and separated in an agarose gel (section 2.3.2).

Table 2-28 PCR reaction using Q5 High-Fidelity DNA Polymerase.

Reagents	25 µL Reaction	50 µL Reaction	Final concentration
Q5 High-Fidelity 2x Master Mix	12.5 µL	25 µL	1x
10 µM Forward Primer	1.25 µL	2.5 µL	0.5 µM
10 µM Reverse Primer	1.25 µL	2. 5 µL	0.5 µM
Template DNA	1 µL	1 µL	<1000 ng
H_2O	9 µL	19 µL	n/a

Table 2-29 Thermocycler set up temperatures for Q5 High-Fidelity DNA Polymerase.

Stage	Cycles	Temperatures	Time
Initial Denaturation	1 x	98 °C	30 sec
Denaturation		98 °C	5-10 sec
Primary annealing temperature	5 x	*X °C	10-30 sec
Elongation		72 °C	20-30 sec/kb
Denaturation		98 °C	5-10 sec
Secondary annealing temperature	25 x	*X °C	10-30 sec
Elongation		72 °C	20-30 sec/kb
Final extension		72 °C	2 min
Hold		4-10 °C	n/a

Table 2-30 PCR reaction using KOD Hot Start DNA Polymerase

Components	Volume (Final 50 µL)	Final concentration
10x KOD reaction buffer	5 µL	1 x
dNTPs (2 mM each)	5 µL	0.2 mM each
MgSO4 (25 mM)	3 µL	1.5 mM
Forward primer (10 mM)	1 µL	2 mM
Reverse primer (10 mM)	1 µL	2 mM
Template DNA	1 µL	>500 ng
KOD Hot Start DNA polymerase (1 U / µL)	1 µL	0.02 U / µL
18.2 MΩ/cm H2O	32 µL	n/a

Table 2-31 Thermocycler set up temperatures for KOD Hot Start DNA Polymerase.

Stage	Cycles	Temperatures	Time
Initial Denaturation	1 x	95 °C	30 sec
Denaturation	5 x	95 °C	10 sec
Primary annealing temperature		*X °C	30 sec
Elongation		72 °C	20 sec/kb
Denaturation	25 x	95 °C	30 sec
Secondary annealing temperature		*X °C	30 sec
Elongation		72 °C	20 sec//kb
Final extension		72 °C	10 min
Hold		4-10 °C	n/a

Table 2-32 Thermocycler set up temperatures for OneTaq DNA Polymerase.

Reagents	25 µL Reaction	Final concentration
OneTaq 2 x Master Mix	6 µL	1 x
10 µM Forward Primer	1 µL	1 µM
10 µM Reverse Primer	1 µL	1 µM
Colony resuspended in H_2O /Template DNA	4 µL	<1000 ng

Table 2-33 Thermocycler set up temperatures for OneTaq DNA Polymerase.

Stage	Cycles	Temperatures	Time
Initial Denaturation	1 x	94 °C	30 min
Denaturation	5 x	94 °C	30 sec
Primary annealing temperature		45-68 °C	15- 60 sec
Elongation		68 °C	1min/kb
Denaturation	25 x	94 °C	30 sec
Secondary annealing temperature		45-68 °C	30 sec
Elongation		68 °C	1min/kb
Final extension		68 °C	5 min
Hold		4-10 °C	n/a

2.3.2. Agarose Gel Electrophoresis

A 1% (w/v) agarose mini gel was prepared in a 250 mL Erlenmeyer flask (section 2.1.6) and boiled in a microwave until the agarose was fully dissolved, without any visible speckles. The agarose solution was left to cool down for 5 min, and then it was poured into a Bio-Rad casting tray with the desired combes inserted. The gel was left about 20 min until completely set and was then moved from the casting tray and transferred into a Bio-Rad horizontal gel tank. The gel was covered with 1 x TAE buffer diluted from 50 x stock concentration buffer. The DNA samples were prepared as stated in (section 2.1.6), mixed in a vortex and pulse spun down. The comb(s) were then removed from the agarose gel and the DNA samples were loaded in the wells using a Gilson pipette. A Hypperlader™ 1 kb size marker from NEB was loaded on a well as size reference ((section 2.1.6).A Bio-Rad powerpack was used to run the gels set at 110 volts for 35-40 min (small gel size) to allow for good DNA separation. The gel was then soaked for 10 min in ethidium bromide solution, was rinsed with 18.2 MΩ/cm H_2O, and was visualised under UV light in a Pharos FX Plus Molecular Imager.

2.3.3. PCR Product Gel Extraction

The desired PCR product from the gradient were pooled, separated in an agarose gel as stated in section 2.3.2. Following the gel staining in ethidium bromide, the gel pieces of the desired size band were excised with a scalpel under a UV light and transferred to a 15 mL falcon tube. The before and after the addition of the gel weight of the falcon tube was recorded, and therefore the weight of the gel was determined. DNA extraction was performed using the NZYTech Gelpure kit according to manufacturer's instructions (Table 2-35). The PCR product was eluted in 50 µL elution buffer (section 2.1.6).

2.3.4. PCR Product Restriction Digestion

All restriction enzymes and buffers are in Table 2-34. PCR products were digested for 2 hours at 37 °C using 48 µL of gel-purified PCR product, 1 µL of each restriction enzyme (final 20 U/ µL), 6 µL restriction enzyme buffer (final 1x), and 4 µL 18.2 MΩ/cm H_2O to make it up to 60 µL reaction volume.

Analytical digests were performed for 2 hours at 37 °C using 7 µL mini prep DNA, 1 µL of each restriction enzyme, and 1 µL required buffer (final conc. 1 x).

Table 2-34 Molecular biology enzyme suppliers and applications

Enzyme	Supplier	Application
T4 DNA ligase	NEB	Ligation
*Avr*II, *Bam*HI, *Eco*RI, *Hin*dIII		Restriction digestion
*Nco*I, *Nde*I, *Sal*I, *Xho*I, *Not*I	NEB	
Antarctic phosphatase	NEB	Dephosphorylation of DNA
Buffer 2, Buffer 3, Cutsmart	NEB	Restriction enzyme buffers

2.3.5. Enzymatic Clean Up

Following a restriction digest, the PCR product was cleaned up using NZYTech NZYGelpure kit according to manufacturer's instructions. DNA was eluted in 30 µL of 18.2 MΩ/cm H2O (pH 7.5). If sufficient DNA concentration for the insert/plasmid ratio was not received, then the DNA was concentrated down using a SpeedVac (Thermo Scientific SpeedVac DNA 130).

2.3.6. Preparation of *E. coli* Chemically Competent Cells

An aliquot of *E. coli* TOP10, BL21 (DE3), C41(DE3), C43(DE3), Tuner or XLI-Blue comp cells was streaked out on an agar plate without any antibiotic and was incubated overnight at 37 ºC in a static incubator. A colony was picked off the agar plate and was inoculated into 5 mL LB broth without antibiotic in a 30 mL universal and was incubated overnight at 37 ºC with rotary shaking at 200 rpm (INFORST HT Multitron Standard). A straight sided 2 L conical flask containing 500 mL LB was inoculated with the 5 ml starter culture and was incubated at 37 ºC until the OD_{600} reached between 0.4- 0.6 (approx. 2 hr). The culture was chilled on ice for 10 min with occasional swirl to make sure all the cells were chilled. The chilled culture was then split equally between two prechilled 500 mL Nalgene centrifuge pots and spun down for 10 min at 4000x g in a Sigma 3-18KS pre-cooled centrifuge at 4 ºC. The supernatant was aseptically discarded, and the cells were washed by resuspending each pot with 100 mL 0.1 M sterile, ice cold $MgCl_2$ by gentle swirling. The cells were centrifuged again at 4000 x g for another 10 min, and the supernatant was aseptically discarded once again, ensuring the cells were always kept on ice. The first pellet was resuspended by gentle swirling in 15 mL of 0.1 M ice cold $CaCl_2$, and the resuspension was added to the second pellet where the cells were swirled into the suspension. The suspension was left on ice between 1.5-4 hours to mature, followed by adding pre-chilled sterile glycerol to a final concentration of 15% (v/v; using 50% (v/v). The suspension was incubated overnight on ice, and aliquots of 150 µL competent cells were distributed in 1.5 mL prechilled microcentrifuge tubes, which were stored in the -80.

2.3.7. Plasmid DNA Preparation Mini

A colony was picked from the plate and inoculated in a 5 mL sterile LB broth made up in a 30 mL glass universal. The inoculum was incubated overnight at 37 °C in a shaking incubator (New Brunswick Scientific) at 200 rpm. The plasmid DNA (pDNA) was isolated following the manufacturers' instructions either the NZYTech kit, or the Thermo Scientific kit (section 2.3.12, DNA manipulation kits). The DNA was eluted in 50 of elution buffer into a sterile 1.5 mL microcentrifuge tube and stored at -20°C.

2.3.8. Plasmid DNA Preparation Midi/Maxi

Some plasmids were purchased from Novagen, while others were used from the personal collection of Prof. Gary Black. A loopful (10 µL) of *E. coli* bacteria containing the target plasmid was streaked in an antibiotic plate (antibiotic of selection) and were incubated for 16-18 hours (overnight) at 37 °C. A colony was picked from the antibiotic plate and inoculated in 50 mL LB broth made up in a 250 mL non-baffled flask, for a Midiprep, or in 250 mL LB broth made up in a 2 L non-baffled flask, for Maxiprep, supplied with the required antibiotic. The inoculum was incubated overnight at 37 °C in a shaking incubator (New Brunswick Scientific) at 200 rpm. The plasmid DNA was purified following the manufacturers' instructions either the NZYMidiprep/Maxiprep kit, or the Thermo Scientific kit. The DNA was eluted in 500 µL for Midiprep and 1 mL of elution buffer for Maxiprep, into sterile 1.5 mL microcentrifuge tube and stored at -20°C.

2.3.9. Plasmid DNA restriction digest

For pDNA restriction digest, 10 µg of pDNA was used as starting concentration, and 5 µL of 20 U/ µL concentration of each restriction enzyme, the appropriate restriction enzyme buffer at final concentration 1x (buffer 2, buffer 3 or cutsmart), and the rest 18.2 MΩ/cm H2O, to make up to volume. The digestion reaction was placed at 37 °C for at least 4 hours. Following the digest, the pDNA was subjected to an enzymatic clean up according to section 2.3.5 and eluted in 50 µL of elution buffer into sterile 1.5 mL microcentrifuge tube. The pDNA was added 1 x loading

buffer, and was separated overnight in medium size agarose gels, set at 20 milliamp (mA). For multiple plasmid preparations large size agarose gels were used. The DNA was gel extracted as stated in PCR Product Gel Extraction and eluted in 50 µL 18.2 MΩ/cm alkaline H2O (pH 7.5). If a concentration less than 50 ng/µL was recovered, then the pDNA was concentrated down using the SpeedVac until the desired concentration was achieved.

2.3.10. Dephosphorylation of the Digested DNA

When pDNA was linearised with a single restriction enzyme, or restriction enzymes that their overhangs happen to be complementary to each other, and hence a self-ligation can occur, the DNA was subjected to a dephosphorylation step. In a 70 µL digested pDNA, 1 µL phosphatase, 8 µL Antarctic phosphatase buffer and 1 µL sterile 18.2 MΩ/cm H_2O were added in a 1.5 mL eppendorf and the reaction mixture was incubated at 37 °C for 30 min. The reaction was heat shocked at 80 °C for 2 min to inactivate the phosphatase, and the reaction was subjected to an enzymatic clean up (section 2.3.5) and eluted in alkaline H_2O.

2.3.11. Ligation

The ligation components concentration was calculated according to the downloadable NEB tools app - ligation, and the ligation reaction was performed in 0.5 mL PCR tubes. Specifically, 2 µL (50 ng/µL) REase-digested plasmid DNA was always used, x µL of insert (to a ratio between 1:3 to 1:6 (vector: insert) was employed, equation below), REase-digested PCR product, 1 µL T4 10 x ligation buffer, and 1 µL T4 ligation enzyme, and 18.2 MΩ/cm H_2O if necessary to make it up to 10 µL final volume. The reaction was briefly vortexed and spun down and was incubated in the thermocycler set at 16 °C for 48 hours, followed by hold at 10 °C. The concentration of the DNA was determined using a NANODROP ONE from Thermo Scientific.

$$\text{Insert DNA (ng)} = \frac{\textbf{Insert Ratio x Insert size (bp) x Vector mass (ng)}}{\textbf{Vector size (bp)}}$$

2.3.12. DNA Manipulation Kits

Table 2-35 displays the DNA manipulation kits used in this study. All kits were stored at ambient temperature, with some components at 4°C according to the manufacturers' instructions.

Table 2-35 DNA manipulation kits.

Kit	Manufacturer	Application
NZYMaxiprep kit	NZYTech	Large scale vector purification
NZYMidiprep kit	NZYTech	Large scale vector purification
NZYMiniprep kit	NZYTech	Recombinant DNA purification
NZYGelpure kit	NZYTech	DNA Extraction
NZYtech NZY Tissue gDNA	NZYTech	Metagenomic DNA purification
GeneJET Plasmid Maxiprep Kit	Thermo Scientific™	Recombinant DNA purification
GeneJET Plasmid Midiprep Kit	Thermo Scientific™	Recombinant pDNA purification
GeneJET Plasmid Miniprep Kit	Thermo Scientific™	Recombinant pDNA purification
GeneJET PCR Purification Kit	Thermo Scientific™	DNA Extraction

2.3.13. Transformation of Plasmid DNA in *E. coli* TOP10 Chemically Competent Cells

Aseptically, 5 µL of the ligation was added to 150 µL *E. coli* chemically competent cells in a 1.5 mL microcentrifuge tube. Transformations were gently mixed by stirring with a Gilson pipette and left on ice for 5 hours, heat shocked at 42 °C for 90 seconds, and placed again on ice for 2 minutes. 200 µL sterile LB broth was added to each of the transformations and the transformed cells were incubated at 37 °C in a water bath (Grant JB Nova) for 1 hour, followed by plating out on antibiotic agar plates (for the required selectivity) using a spreader. The plate was incubated at 37 °C overnight in the static incubator.

2.3.14. Confirmation of Transformants by Colony PCR and/or Restriction Digest

3 to 5 colonies were picked from the ligation-transformation agar plate and were resuspended in 12 µL sterile 18.2 MΩ/cm H2O in a PCR tube. 2 µL were transferred from the suspension to a new, sterile PCR tube, where 2 µL of sterile 18.2 MΩ/cm H2O, 0.5 µL of the forward and reverse primers (final concentration 10 µM each), and 5 µL of One Taq 2 x Master Mix (section 2.1.6) were added. The PCR reaction was mixed by vortexing and placed in the thermocycler. The recombinants were confirmed by agarose gel electrophoresis Agarose Gel Electrophoresis. With successful recombinants, the rest of the colony resuspension (10 µL) was inoculated in a pre-prepared and sterilised 5 mL LB in a 20 mL glass universal and incubated at 37 °C overnight. The pDNA was purified using a mini prep kit, and a diagnostic digest was performed for further confirmation of the recombinant (section 2.3.4), and sequencing of the clones using GENEWIZ®. A 25% glycerol stock was prepared from sterile 50% glycerol stock in a 1.5 mL microcentrifuge tube by adding 500 µL 50 % (v/v) glycerol and 500 µL recombinant culture. The glycerol stock was stored in the -80.

2.4. Recombinant Protein Expression

2.4.1. Transformation of the Clones in *E. coli* BL21(DE3), C41(DE3), C43(DE3) or Tuner cells

Aseptically, 2 µL of pDNA was transformed in 150 µL *E. coli* competent cells of choice, in a 1.5 mL microcentrifuge tube for expression of the recombinant enzyme. When essential co-transformation of two or more plasmids in the same *E. coli* cells, 2 µL of each plasmid containing the target gene were transformed in the cells. The transformation was gently mixed by stirring with a Gilson pipette and was incubated on ice for 2 hours, then heat shocked for 90 sec at 42 °C, followed by resting on ice for 2 minutes. Then, 200 µL of sterile LB broth were added to the transformation reaction and was placed at 37 °C for 1 hour. The transformations were plated out in antibiotic agar plates and incubated at 37 °C overnight.

2.4.2. Protein Expression

One colony was picked from the antibiotic plate, was inoculated in a 5 mL LB liquid broth, and incubated at 37 °C overnight. An 1% inoculum was inoculated in 50 mL media (different media were trialled to achieve best expression) supplemented with the required antibiotic/s, made up in 250 mL baffled flask. The culture was incubated at 37 °C until the OD_{600} reached between 0.6-1.0 absorbance (approx. 2.5 hours). The absorbance was measured using 1 mL culture in a microcuvette in a Helios α, Spectronic Unicam spectrophotometer at 600 nm, and pure LB as blank. Firstly, 1 mM δ-aminolevulinic acid (δ-ALA, or 5-ALA) was added to the cultures, which is the haem synthesis initiator. Induction of GroEL and GroES chaperones was achieved with the addition of 2 mg/mL L-arabinose, which was employed to help with the protein folding. Induction of P450s was achieved by addition of 1 mM IPTG. This was applied only to LB cultures, while different concentrations of IPTG were used with Tuner cells. The cultures were cultivated at 20 °C and 30 °C overnight for LB with rotary shaking at 100 rpm (INFORST HT Multitron Standard), and at 48 h for AIM at 180 rpm. Expression conditions with and without pGro7 were trialled for CYP expression.

When large scale protein expression was required, 500 mL LB broth was made up each in 4x 2 L baffled flasks (total 2 L expression media), and induced with 1 mM δ-ALA, 2 mg/mL L-arabinose and required concentration of IPTG.

2.4.3. Cell Harvesting and TCP and CFE Preparing

A 50 mL induced culture was harvested after approximately 24 hours for LB broth (Hausjell et al.) and 48 hours (AIM), and their OD_{600nm} was recorded to determine the cell density. Usually, LB media would be harvested between OD_{600nm} 3.5 – 5.5 A, whereas AIM between OD_{600} 6.5 – 12 A. AIM reached higher densities than the LB. The harvesting point and densities of the cultures would vary also depending on the clone.

2.4.4. Preparation of Total Cell Protein

Preparation of the TCP from E. coli were performed at 4 °C. 0.5 mL of culture was collected in a 1.5 mL microcentrifuge tube and spun down for 1 minute at 13000 x g. After discarding the supernatant, the cell pellet was resuspended in 100 µL cracking buffer (section 2.1.9). The TCP samples were prepared according to (section 2.1.9), separated according to their molecular masses with 12 % SDS-PAGE and visualised with Coomassie blue dye (section 2.1.9).

2.4.5. Preparation of Cell Free Extracts (CFE)

Preparation of the CFEs from *E. coli* were performed at 4°C. The cells were harvested by centrifugation for 10 min at 4000 x g (SIGMA 3-18KS) and were resuspended in 5 mL Start buffer pH 7.4 (section 2.1.7). The suspensions were sonicated at 13000 amplitude (Soniprep 150 Plus) for 2 min, with 10 second on /off intervals, and then centrifuged for 40 minutes at 18500 x g and 4 °C (Thermo Scientific SL 16R centrifuge). The supernatant (approx.4 mL CFE) was transferred into a 15 mL clean falcon tube, and if not purified immediately, glycerol to 15 % (v/v) final concentration was added and the cell lysates were stored in the – 20. All CFE samples were prepared according to section 2.1.9, separated according to their molecular masses with 12 % SDS-PAGE and visualised with Coomassie blue dye. The supernatants were purified using either AKTA prime plus, or Bio-Rad Econo-pac® Chromatography columns (section 2.4.6 and 2.4.7).

2.4.6. Purification through Econo-pac® Chromatography Columns

The expressed enzymes were purified using gravity-flow Econo-pac chromatography columns. First, the columns were filled with 5 mL Talon Metal Affinity Resin (stored in 20% ethanol) and were let until the resin set at the bottom of the column. Second, the column was equilibrated with 5 bed volumes of Start buffer (20 mL). Third, the 5 mL sample was applied to the column and the flow through was collected in a 5 mL bijou. Then, sequentially, 5 mL of each of the buffers from start buffer to elution buffer (section 2.1.7) were loaded to the column and the flow throughs were collected in separate labelled universal tubes. All

purified enzyme fractions were stored on ice at 4 °C. The purified proteins were prepared according to section 2.1.9 separated according to their molecular masses with 12 % SDS-PAGE and visualised with Coomassie blue stain. The purified enzymes were used in biotransformation assays, while the remaining protein was supplemented with final concentration of 15% glycerol and stored for short time in ice placed at 4 °C.

2.4.7. Protein purification through Immobilised Metal Affinity Chromatography (IMAC) using AKTA Prime Plus System (Cytiva)

Proteins were purified using medium pressure chromatography through prepacked Ni^{2+} Sepharose, a 5 mL His Trap HP columns attached to an AKTA system. The column consists of small diameter beads which are cross-linked agarose, compact in the column. A binary of buffers was used for the elution step and consisted of start buffer and elution buffer (section 2.1.7). The sample, 5 mL of CFE, was loaded into the column through the tubing line position 8 by a manual method at a flow rate 2 mL per minute, to the pre-charged Nickel column, then 20 mL of start buffer were loaded to the column to remove any nonspecific bund proteins. Then elution method 3 (Table 2-36) is initiated, where a linear gradient of start and elution buffer is used which starts at 100 % start buffers and finishes with 100% elution buffer. 5 mL fractions were collected from the delivery arm, and flow rate was 1 mL per min, pressure limit 0.3 MPa while all other method options were left to default according to manufacturer's instructions. Then 100 % elution buffer was run through the column for 20 min at 1 mL per min flow rate to wash the column, followed by re- equilibration with 5 min start buffer. The chromatograms were interpreted with PrimeView program.

Confirmation of expressed enzymes in purified fractions and quantification of enzymes was accomplished through Bradford reagent assays (section 2.5.1) according to manufacturers' instructions. All purified proteins were prepared (section 2.1.9), were separated with SDS-PAGE and visualised with Coomassie blue dye.

Table 2-36 AKTA methods used for protein purification.

Method 1: Ni Striping and Charging

30 mL	18.2 MΩ/cm H_2O
30 mL	100 mM EDTA
30 mL	18.2 MΩ/cm H_2O
30 mL	100 mM $NiSO_4$

Method 2: Ni Equilibration

100 mL	Start Buffer

Method 3: Elution Method

200 mL	Start Buffer
200 mL	Elution Buffer

Method 4: Ni EtOH Store Method

100 mL	20 % EtOH

2.4.8. SDS-PAGE Analysis of All Expressed Proteins

All expressed enzymes were separated with SDS-PAGE gels, which were freshly poured on the day according to the recipe in section 2.1.9, were placed in the appropriate vertical trays, and into a Bio-Rad mini-PROTEAN® tetra system SDS-PAGE gel tank. The tank was filled with the required volume of 1 x SDS running buffer, and the combs were carefully removed, and the wells were rinsed with more 1 x SDS running buffer with a Pasteur pipette. Then, 10-20 µL (depending on well size) of prepared protein sample (boiled TCP, CFE or purified enzyme with loading buffer) were loaded in each well, and a protein standard (ThermoFisher Scientific) as a size reference. The tank was connected to a Bio-Rad powerpack, and the gel was run at 200 V for 50 minutes. Once the process was completed the gels were carefully removed from the trays and the glass plates and were placed in square petri dishes where they were submerged in Coomassie blue stain (section 2.1.9). The SDS gels were stained for 10 min by placing them on a rocker (KS125 basic IKA LABOTECHNIK), followed by destaining overnight in destain solution (section 2.1.9) with rocking. The gels were then rinsed with distilled water and imaged with the SYNGENE G: BOX.

2.4.9. IN- GEL Trypsin Digestion

i. Preparation and destaining of Gel pieces

To confirm that target proteins were expressed, an in-gel trypsin digestion was performed. Specifically, freshly poured gels with proteins separated were stained with Coomassie stain and destained in Destain solution. Then the stained protein bands were excised from the gel with a sterile scalpel into approximately 1 x 1 mm^3 pieces and placed in LoBind microcentrifuge tubes (In-Gel Tryptic Digestion Kit, Thermo Scientific™). The Coomassie stained gel pieces were washed three times with 100 µL of 100 mM NH_4HCO_3 (ammonium bicarbonate) followed by 60 µL 100 % acetonitrile (ACN) for 15 min with agitation at room temperature, or as long needed to remove Coomassie stain from the gel pieces. Then, the gel pieces were shrunk for 5 - 10 min by adding 200 µL 100 % ACN, and ACN was pipetted off.

ii. Reduction

The gel pieces were rehydrated in 100 µL 20 mM DTT (Dithiothreitol) for 30 min at 56 °C, followed by removing the excess liquid and re-shrinking of the gel pieces with 100 % ACN as with the previous step.

iii. Alkylation

The gel pieces were rehydrated by adding 100 µL 55 mM iodoacetamide (IAA) in 100 mM NH_4HCO_3 and alkylated in the dark for 20 min at room temperature. All the excess liquid was removed, and the gel pieces were washed twice with 100 µL NH_4HCO_3 in room temperature with agitation. The gel pieces were dehydrated with 100 µL ACN for 15 min at room temperature and agitation. ACN was removed and gel pieces were dried in a vacuum centrifuge for 5 minutes.

iv. Trypsin Digestion

The gel pieces were dehydrated with 30 µL of 20 µg/mL trypsin solution and placed on ice for 20 min to allow to absorb solution. 100 µg trypsin stock was diluted with 1 mL 50 mM acetic acid, made 100 µg/mL stock solution, of which then was diluted 5 times in 50 mM NH_4HCO_3 to give 20 µg/mL. If all solution was absorbed, then more was added for gel pieces to be completely saturated in trypsin solution. After 20 min, 50 µL 50 mM NH_4HCO_3 were added and incubated 16-24 h 37 °C.

v. Extraction of peptides

After overnight incubation, 50 µL 50 % (v/v) ACN/ 5 % (v/v) formic acid (FA) was added to stop the enzymatic reaction and were incubated at room temperature with agitation. The supernatant was removed by pipette and transferred to another fresh LoBind tube the extraction was repeated by adding 50 µL 83 % (v/v) ACN / 0.2 % (v/v) FA. Once again, the supernatant was removed and added to the previous LoBind tube.

The supernatant was frozen at -80 with a hole punched on top of the LoBind tube, followed by freeze drying and the lyophilised samples stored in -80. The lyophilised samples were resuspended 20 µL 5 % (v/v) ACN/ 0.1 % (v/v) FA and analysed by LC-MS for proteomic analysis.

2.4.10. Proteomic Analysis of Peptides with Liquid Chromatography Mass Spectrometry (LC-MS)

The peptide identifications were performed using a Nanoflow Dionex™ 3000 RSLC liquid chromatography system hyphenated to a Q-Exactive Plus High-resolution mass spectrometry system with a C18 EasySpray column emitter (Thermo, Hemel Hempstead, UK, part number ES803).

The LC buffer for the peptide separation was a binary solvent system: Buffer A consist of 95 % LC/MS grade water 95%/5 % ACN with 0.1 % formic acid) v/v, Buffer B consist of 95% ACN/5% ultrapure water with 0.1 % formic acid. The loading and transport buffers were 95 % LC/MS grade water/ 5% ACN with 0.1 % Tetrafluoruacetic acid (TFA). The flow rate was set to 300 nL/min for column and 25 µL/min for the trap.

The trap column used was Acclaim™ Pep Map™ 100 C18 LC column (Thermo Scientific™) (5 µm particle size; pore size 100 Å), maintained at 45 °C. The trap column is a C18 stationary phase, and it allows the pre-concentration of low abundance peptides to be accumulated and eluted during the analysis. The easy spray column emitter was connected to the nanoflow HESI source (part number ES801) with the inlet capillary operating at 275^{o}C.

The data were acquired on a Orbitrap plus hi resolution mass analyser. The data was acquired using the data dependent analysis (DDA) methodology with the following settings: full scan was performed at 35,000 Mass resolution at 200m/z at rate of approximately 7 scans/sec with an automatic gain control (AGC) of 1e^{6} and injection time of 100 ms^{-1}. The mass acquisition range was set to 400 to 1400 m/z. The DDA MS/MS acquisition were performed also at 35,000 with an AGC of 1e^{5} with a maximum injection time of 100 ms^{-1}. The isolation window was set to 1.3 m/z, with a underfilled ratio of 0.4 %, Dynamic exclusion was set to 15 seconds, and the top 10 most abundant ions were selected for MS/MS with a normalized Collison energy (NCE) level of 20. The sample injection volume of 3uL was used. The total run time was approximately 110 min injection to injection. The Thermo Raw post-acquisition data files were processed for peak table generation using Maxquant and Perseus workflow (https://www.maxquant.org/) using the developer recommended sittings.

After peak table generation and removal of contaminants using Maxquant build in library, protein selection was performed using a minimum of two or more unique peptide sequences. Missing value imputation was performed within Perseus (https://maxquant.net/perseus/). The annotated peak table was exported for visualization and multivariant statical analysis in metaboanalyst V5 (https://www.metaboanalyst.ca). The peak table was initially processed log2 normalisation follow by unit-variance and mean centring (autoscaling) prior to unsupervised PCA analysis and visualization and PLS-DA analysis.

2.5. Enzymatic Assays

2.5.1. Determination of Protein Concentration

The concentration of the purified enzyme was determined using Bovine Serum Albumin (BSA) and Bradford reagent (Appendix 3). A 10 mg/mL BSA stock concentration was used to make a standard curve by making serial dilutions of the stock concentration in 20 mM Tris-HCl pH 7.4 in 10 different 1.5 mL microcentrifuge tube (section 2.1.8). Then, in a flat bottom 96 well plate was added 196 µL of Bradford reagent and 4 µL of each BSA concentration and were mixed with the pipette. The reactions were developed in the dark for 10 min and read in a plate reader (TECAN SPARK 10M) at 595 nm. The absorbance readings were used to generate a standard concentration graph and curve in Microsoft Excel, with the formula of the curve calculated. Then, all proteins of unknown concentration were subjected to the same method and formula to determine their concentrations. All unknown concentration protein dilutions were performed in triplicate where a mean absorbance number was generated.

2.5.2. P450 CO Reduced Spectral Assay in Whole cells, CFEs and Purified Enzymes

CO reduced spectral scanning was performed in all P450s in order to confirm activity of the P450s. 1 mL was harvested from E. coli expressing cultures was pelleted at 17000 x g at 4 °C for 2 minutes (Thermo Scientific SL 16R centrifuge). The pellet was resuspended in 1 mL Spec buffer (section 2.1.8) plus 950 µL sterilised 18.2 MΩ/cm water, and approximately 5-10 mg of sodium dithionite (Na2S2O4) added to the solution. The mixture was mixed gently and separated in two UV micro cuvettes and their absorbance was recorded at OD_{595nm} in a dual beam recording spectrophotometer (Helios α, Spectronic Unicam). One of the cuvettes was used as a reference while the other was utilised for the sample. The reference was used to set up a baseline between 400 nm and 500 nm, while the sample was flushed for 1 minute under the CO, and its CO-reduced difference spectrum measured between $OD_{400\text{-}500}$ nm. A peak is expected to be seen at 450 nm when P450 activity is detected. A negative control CO assay was performed with empty plasmid (Appendix 3, Figure 3.2).

Alternatively, 100 µL of CFE was added to 1 mL spec buffer plus 900 µL sterilised 18.2 MΩ/cm water, and the above procedure was repeated. The total concentration of the P450 in a sample preparation can be calculated by the difference of the absorbance (ΔOD) between 490 and 450 nm and divided by the extinction coefficient that Omura and Sato generated for the P450s, $\varepsilon = 91\ mM^{-1}\ cm^{-1}$ (Omura and Sato, 1964). The below formula shows the calculation of P450 per mL of assay.

$$(\Delta A_{450} - \Delta A_{490})/0{:}091 = \text{nmol of P450 per ml}$$

2.5.3. Catalytic Activity of CYPs in Whole Cell Biotransformations (WCB) and CFE Assays

P450 characterisation assays were performed on whole cell and cell free extracts. All assays were performed in triplicate, and 100 µL time point samples were taken at defined intervals.

2.5.3.1 Whole Cell Biotransformations

In 15 mL falcon tubes, 14 mL of the 50 mL LB E. coli expression cultures containing the relevant P450 were harvested at 4000 xg and resuspended in 875 µL P450 screening buffer (section 2.1.8). The enzymatic reaction was initiated with the addition of 100 µL substrate and 25 µL 10 mg/mL NADPH reducing agent. The concentration of the substrate varied depending on the further analysis performed. For reactions that were analysed using the HPLC, the substrate stock concentration was 5 mg/mL and final concentration was 0.5 mg/mL, whereas for LCMS analysis the stock concentration of the substrate was 100 µg/mL, and final concentration 10 µg/mL. The reactions were placed on an Innova 5000 environmental shaker at 8000 rpm, set at 37 ºC, with 100 µL samples taken at defined intervals (0 hr, 15min, 3hr and 24hr). All drawn samples were centrifuged at 13000 x g in a bench top centrifuge for 5 minutes, then the supernatants were transferred into fresh, sterile 1.5 mL microcentrifuge tubes, boiled for 10 min, centrifuged again 13000 x g for 10 more minutes, and stored at -80 °C, until analysed through HPLC (Ultimate 3000), or Mass spectrometer system (ThermoScientific Vanquish UHPLC connected to an IDX tribrid).

2.5.3.2 Cell Free Extract Assays

All CFE assays were undertaken in 1 mL Screening buffer (section 2.1.8) and placed at 37 ºC overnight. At selected intervals of 0, 3 and 24 hours 100 µL samples were taken. 13000 x g in a bench top centrifuge for 5 minutes, then the supernatants were transferred into fresh, sterile 1.5 mL microcentrifuge tubes, boiled for 10 min, and stored at -20 °C, until analysed through HPLC (Ultimate 3000).

2.5.4. (HPLC) Analysis of All Enzymatic Assays

Some of the WCB timepoint samples were analysed using a HPLC system and the chromatographic method and buffers as in Table 2-37. Chemical standards of ketamine, norketamine and coumarin were included in the analysis for identification within the sample. The standards and the substrate were analysed in the assay conditions, for their stability to be determined over a 24-hour timeframe, but without the addition of the cells. The same final concentration was always applied for standards and substrate.

Table 2-37 High Performance Liquid Chromatography (HPLC) buffers and conditions

Column:		Stationary Phase:	Phenomenex, Luna C18 (2), 5µm
Material/Dimensions:		Stainless Steel, 150 mm x 4.6 mm	
Mobile Phase A:		50 mM Potassium dihydrogen phosphate buffer, pH 5.0 / acetonitrile (97:3 v/v)	
Mobile Phase B:		50 mM Potassium dihydrogen phosphate buffer, pH 5.0 / acetonitrile (32:68 v/v)	
Gradient Times:	Time	A	B
	minutes	%	%
	0.0	100	0
	8.0	100	0
	55.0	0	100
	60.0	0	100
	60.1	100	0
	70.0	100	0
Column Temperature:		30°C	
Injection Volume:		24 µL	
Needle Wash:		Diluent	
Detection wavelength:		220 nm	
Total Run Time:		70.0 minutes	

2.5.5. Metabolomic analysis of WCB with LC-MS

LC-MS analysis was performed in the WCB timepoint samples with the concentration of substrate used at 10 µg/mL final. The separation was performed on a ThermoScientific liquid chromatography Vanquish system with 20 µL injection loop system. An Acclaimed 120^{TM} C18 Liquid chromatography analytical column (particle size 3 µm, internal diameter 2.1 mm length 50 mm) was used for chromatographic separation maintained at 45 ℃. A binary buffers system was used: buffer A consisted of 95/5 (v/v) LC-MS grade water and Acetonitrile and buffer B was 95/5 (v/v) Acetonitrile and LC-MS grade water respectively, both buffer systems contain 0.1 % (v/v) formic acid with a pH of 3. The flow rate was set to 400 µL/min. Normal pressure profile range from 130 bars to 50 bars.

Each chromatographic run had a length of 20 min, with the first 1 min diverted to waste. The starting condition was 95 % buffer A and 5% buffer B, which remained for 2 min. Then, a linear increase to 95 % buffer B, reached at 11 min, which remained for 4 min (cleaning period), and immediately returned to starting condition and remained for 4 min. The void volume was 0.4 mL.

The MS acquisition was performed on a ThermoScientific High resolution IDX tribrid mass analyser with a OptaMax heated electrospray ionisation (HESI) system. The spray voltage was set to 3.5 kV (positive mode) with a scan range set from 100-1,000 *m/z* and a mass resolution at 30,000. The automatic gain control (AGC) was set 25% and maximum injection time of 50ms. The data acquisition was performed in MS1 mode only. The HESI vaporiser temperature was set at 100 °C, the ion transfer tube (ITT) temperature was set at 325 °C. The sheath gas flow rate was maintained at 50 the auxiliary gas flow was set to 10, and the sweep gas set to 1.

The data analysis was done using the Thermo Xcalibur Qual browser, version 4.3. Chemical standards of ketamine, norketamine and coumarin were included in the analysis for identification within the sample, using single ion monitoring (SIM) mode methodology with a mass tolerance set to 10ppm. The difference between the substrate's diminishing and the product produced by matching the ion mass plus the mass of the retention time was used to quantify the products. To verify the identity, the mass of the products was computed while also accounting the protonated adducts and compared to the peak that had appeared. The samples prepared for LCMS analysis without pre-concentration step and were injected neat.

2.6. Nitrilases Methods

2.6.1. Transformation and Expression of Target Enzyme

Twenty nitrilases were taken from the personal collection of Prof. Black (Table 2-38), which were already cloned into pET28a plasmid with restriction enzymes NdeI-XhoI. 2 μL of each DNA provided was transformed into E. coli BL21(DE3) competent cells. All clones were recombinantly expressed in *E. coli* BL21(DE3), as previously described, and were stored as lyophilised CFEs. Expressions were attempted in LB and AIM at 20 °C and 30 °C for nitrilases 1-10, whereas nitrilase 11 to 20 expressed in LB 20 °C except for nitrilase 17, which expressed in AIM 30 °C (Table 4-1). All cells were harvested as described in the previous method section, and TCPs and CFEs were prepared and separated in 12 % SDS gels. Suspensions of the lyophilised enzymes at concentrations of 0.054 mg/ mL in 10 mM phosphate buffer, pH 7.2 were used to determine their substrate specificity and relative activity for a range of six nitriles (Table 2-39).

Table 2-38 Nitrilase enzymes used in this study.

UniProt Code	Recognised Activity	Organism
Q2J474	Nitrilase/cyanide hydratase and apolipoprotein N-acyltransferase	*Rhodopseudomonas palustris*
A5ETE9	Putative nitrilase/ CN hydrolase	*Bradyrhizobium sp*
B4AL96	Nitrilase	*Bacillus pumilus*
Q89PT3	Aromatic and aliphatic nitrilase/ CN hydrolase	*Bradyrhizobium diazoefficiens*
A5EKU8	Aliphatic nitrilase	*Bradyrhizobium sp.*
A0R378	Nitrilase 2/ CN hydrolase	*Mycolicibacterium smegmatis*
Q89GE3	Aromatic Nitrilase	*Bradyrhizobium diazoefficiens*
Q819F0	Nitrilase/ CN hydrolase	*Bacillus cereus*
Q5LLB2	Nitrilase/ CN hydrolase	*Ruegeria pomeroyi*
Q6N284	Putative nitrilase/ CN hydrolase	*Rhodopseudomonas palustris*
A5MYU1	Predicted nitrilase/ CN hydrolase	*Clostridium kluyveri*
A7TP0T	CN hydrolase domain-containing protein	*Vanderwaltozyma polyspora*

P20960	Mostly arylacetonitrilases/ CN hydrolase	*Alcaligenes faecalis*
Q2GR86	CN hydrolase domain-containing protein	*Chaetomium globosum*
A0LKP2	Nitrilase/cyanide hydratase and apolipoprotein N-acyltransferase	*Syntrophobacter fumaroxidans*
E3HN55	Nitrilase 1(Nit1)/ / CN hydrolase	*Achromobacter xylosoxidans*
F0Q9Y1	Nitrilase/ CN hydrolyse	*Acidovorax avenae*
G8CXY5	Nitrilase/ CN hydrolyse	*Alcaligenes faecalis*
D1C8L7	Nitrilase/cyanide hydratase and apolipoprotein N-acyltransferase	*Sphaerobacter thermophilus*
C5DH06	Nitrilase/ CN hydrolyse	*Kluyveromyces thermotolerans*

2.6.2. High throughput Screening of Nitrilases Activity with OPA Reagent

A high-throughput screening method was developed by Black et al. (2015) to semi-quantitatively measure the nitrilase activity in cell free extracts (CFEs) against nitriles (Black et al., 2015). O-phthalaldehyde (OPA) reacts with ammonia and under acidic conditions generates a chromophore which can be measured with UV/visible spectroscopy. Trichloroacetic acid (TCA) was used in their assay as the acid to generate a deep blue colour (Black et al., 2015).

Lyophilised forms of all nitrilases were used in activity assays to determine their activity against target nitriles. All assays were performed in 96 well flat bottom microtiter plates, in triplicate. The known activity of a nitrilase Q89GE3 against 3-phenylpropionitrile was used as a positive control in the CFE suspensions for the detection of ammonia.

2.6.3. Preparations of Stock Concentrations and Buffers

CFEs suspensions were prepared in concentrations of 0.56 mg/ mL in 10 mM phosphate buffer, pH 7.2, while nitrile suspensions were prepared in concentrations of 100 mM in DMSO or EtOH, depending on their solubility. Only nitriles 1 to 6 were of importance to Quotient Sciences, while nitriles 7 and 8 were used as controls (Table 2-39). o-Phthalaldehyde was dissolved to 200 mg/mL in methanol, before being diluted 1 in 100 (final concentration 2 mg/mL) in 15 mM sodium tetraborate buffer pH 9.5. The reagent was stored at 4 °C in the dark until use. Stock concentrations of NH_4Cl solution (2-12 mM) were prepared as well and used to create a standards concentrations graph (Appendix 3, Table 3.1 and Table 3.2, and Figures 3.3 and 3.4).

Table 2-39 Nitriles used in the assays and their given numbers for ease.

Labelled as	Nitrile	Resuspended in
1	Benzyl(2-cyanoethyl) carbamate	DMSO
2	n-Nonadecanenitrile	EtOH
3	N-(3-cyanopropyl) phthalimide	DMSO
4	4-bromo-2-nitrobenzonitrile	DMSO
5	Tert-butyl ((S)-1cyanopropan-2-yl) carbamate	DMSO
6	2-(4-cyanophenyl)-N-ethylacetamide	DMSO
7	3-phenylpropionitrile (Control)	DMSO
8	4-chlorobenzonitrile (Control)	DMSO

2.6.4. Microplate preparation

In a flat bottom microtiter plate were pipetted 207 µL of CFE suspension (final concentration 54 µg/mL) and 23 µL of substrate solution (10 mM final), and the plates were covered and incubated overnight at 37 °C for 18 hours.

2.6.5. Detection of ammonia

In a fresh microtiter plate were added 140 µL of DMSO, 100 µL of OPA reagent, 50 µL test solution which was taken from the incubated overnight plate, and 50 µL of 10 % (w/v) TCA solution. Each nitrilase suspension was loaded in triplicate into the wells with the same substrate. The resulting mixture was diluted 1:3 with DMSO in a new microtiter plate to allow the determination of concentration by UV/visible spectroscopy (100 µL of the above test solution were added to 200 µL DMSO) and incubated at room temperature for 10 min for the colour to develop. The absorbance of the plate was recorded at 675 nm using a BioTek Synergy HT multimode microplate reader. The concentrations of ammonia generated were calculated based on the standard curve of NH_4Cl graph, while deionised water was used as blank.

2.6.6. Protein purification through Bio-Rad Econo-Pac® Chromatography columns

All the nitrilases that showed activity against one or more nitriles were purified using Bio-Rad Econo-Pac chromatography columns as described previously in the methods. Briefly, CFEs of target proteins were applied to pre-equilibrated, with 5 bed volumes of Start Buffer, Talon Metal Affinity Resin columns, and the proteins were eluted with increasing concentrations of imidazole (10 mM, 20 mM, 100 mM, 200 mM, 300 mM, 400 mM, and 500 mM imidazole, (section 2.1.7). All elution fractions were collected in separate 20 mL universal tubes. Pure enzymes were confirmed from their separation on 12 % SDS-PAGE gels and visualised with Coomassie blue stain. All purified enzyme fractions were pooled together and concentrated using Amicon Ultra 4 centrifugal filter units with 10 kDa cut off (depending on protein size). Then, a buffer exchange was performed where the

proteins were washed from imidazole by adding 3 times buffer A containing 20 mM Na_2HPO_4 and 150 mM $NaCl_2$ and were concentrated down to final volume 2.5 mL.

2.6.7. Size Exclusion Chromatography

Proteins were subjected to size exclusion chromatography using an AKTA explorer system and sephacryl columns. A Hiprep 16/600 Sephacryl S-200 HR column was used, and 2 mL loop. The column was equilibrated with 1.5 column volumes of buffer A and a 2 mL loop was used. The loop was first rinse-cleaned with 1 loop volume of 20 % ethanol (EtOH), and then the 2.5 mL of the protein was injected to the column. The protein was eluted using buffer A, and the eluted protein fractions were separated in 12 % SDS-PAGE gels. Fractions that contained the pure protein were pooled and concentrated once again with the Amicon Ultra 4 centrifugal filter units. A buffer exchange was performed with 10 mM Hepes buffer and pH 2 degrees above or below the theoretical PI of the protein. A desired concentration of 10 mg/mL enzyme was aimed for before the protein was used for crystallisation.

2.6.8. Crystallisation Method

Only protein Q2GR86 was prepared for crystallisation and screened in 300 nl drops using a mosquito robot (TTP Labtech) together with Crystal screen, Hampton PEG/Ion screen (Hampton research, US), Structure Screen 1 and 2 (Molecular Dimensions LTD, UK). The vapor diffusion sitting drop method was used and the plates were incubated at 37°C for a week. The plates were checked under a microscope for crystal formation. Although the crystallisation procedure was repeated twice, no crystals were formed.

2.6.9. Catalytic Activity Assays

An assay with enzyme and substrate was performed to measure the nitrilase/ nitrile activity, and the conversion rate of substrate into product. Both nitriles, nitrile 1: Benzyl (2-cyanoethyl) carbamate, and nitrile 6: 2-(4-cyanophenyl)-N-ethylacetamide, were assayed, plus the negative control (no enzyme, only buffers and nitrile), and the positive control (4-chlorobenzonitrile). All assays were performed in 1 mL reaction volume in eppendorf, where 100 µL of nitrile solutions from 100 mM stock solution (final concentration 10 mM), 23 µL of purified nitrilase Q2GR86 solution from stock concentration 2.38 mg/ mL in 10 mM potassium phosphate buffer, pH 7.2 (final concentration 54 µg/ mL), and 877 µL potassium phosphate buffer, pH 7.2, were pipetted in. The reaction mixtures were briefly vortexed and placed in an Eppendorf Thermomixer® C set at 37 °C, overnight. Samples of 100 µL were collected from the reaction at set intervals of 0, 1, 2, 4, 6, and 24 hours, with Timepoint 0 (T0 hours) working as the negative control.

2.6.10. Ammonia Detection Assay

Ammonia Rapid Assay Kit (Megazyme) was used to determine the rate of conversion of nitrile to its carboxylic acid. The components included: 2 mL deionised H_2O, 100 µL reaction sample from the nitrilase/ nitrile assay (taken at each timepoint), 300 µL solution 1 from the kit (buffer plus 2-oxoglutarate) and 200 µL solution 2 from the kit (NADPH). All the above components were mixed in 1.6 semi-micro cuvette, and the absorbance of the reaction was recorded after 2 minutes at 230 nm. Then, 200 µL suspension 3 from the kit (glutamate dehydrogenase, or GIDH enzyme) were added. The reaction was mixed with gently pipetting in and out, and the absorbance was measured again after 5 minutes. The same was repeated for each timepoint sample drawn from the reaction solution, and the assays were repeated in triplicate. All the absorbance measurements were used to create the rate graph (see results).

3. P450s Results

After an extensive literature search, we determined that while there has been a substantial effort to tackle the human P450 expression, there was still room for improvement. We developed in this research three different cloning strategies to enhance human P450 expression, and catalytic activity. In the first strategy we co-cloned human P450s with their native reductase. In the second strategy we cloned the human P450s in a pET22b plasmid that positions the mature enzymes in the periplasm which confers them less hydrophobic interaction and aggregation, and proper folding, while the reductase was cloned in pCDF-Duet-1 compatible plasmid and both proteins were co-expressed. In the third strategy the human P450s were cloned in a pET29b plasmid without periplasm location and co-expressed with the reductase already cloned in pCDFDuet-1 from second strategy. All these strategies are explained further in the text.

3.1. Challenges of Producing Human P450s and Strategies Employed to Overcome Them

In general, microsomal P450s have been a challenge to express due to them being membrane bound but also their requirement for complex cofactor needs and the lack of redox partner, which present difficulties.

Microsomal P450s consist of a hydrophobic stretch of a about 20-25 amino acids, followed by a short, positive charged linker region of 10 amino acids, which all together form the N-terminal polypeptide chain of the P450 enzyme that facilitates membrane targeting. Only the hydrophobic stretch is inserted into the membrane, where it is retained (Poulos and Johnson, 2015). A similar sequence exists at the amino terminus of the mitochondrial P450s that targets the mitochondrial membrane, and some of their functions are sterols, steroids, and bile acid formation. In microsomal P450s the negative stretch of the leader sequence is inserted in the endoplasmic reticulum membrane during protein synthesis, whereas in mitochondrial P450s this mitochondrion targeting sequence is cleaved during import of the protein into the matrix side of the mitochondria. The linker region unites the transmembrane helix (TMH) to a conserved proline at the N- terminus of the P450 (Hanukoglu, 1996).

Expression of microsomal P450s with deleted N- terminal transmembrane helices can produce active enzyme, although in the absence of detergents they have a low solubility. Such examples are the rat liver P450c7 that has cholesterol 7-alpha hydroxylase activity (Li and Chiang, 1991). Another example is P450 2E1, which is responsible for the oxygenation of alcohol, N-nitrosodiethylamine, aniline, and p-nitrophenol, which when was truncated at the N-terminus (Δ3-29), remained membrane bound. Moreover, deletion of the N-terminal hydrophobic segment (S1) from the phenobarbital-inducible cytochrome CYP2B4 and additionally deleting the 10 amino acid hydrophilic region (L1), one third of the enzyme remained membrane bound, indicating that there is still some membrane interaction between F and G helices that display increased hydrophobicity (Werck-Reichhart and Feyereisen, 2000). On the other hand, changing the neutral amino acids 3 and 8 from the rabbit P450 2E1, 50% of the expressed enzyme was cytosolic from the previous 30%, whereas deleting 2-27 amino acids from P450 2B4, resulting in the amino terminus becoming positive charged, changing the soluble enzyme from 27% to 67% (Pernecky et al., 1993).

Eukaryotic P450s rely on NADPH-dependent cytochrome P450 reductases (CPR) for electron transport from reduced pyridine nucleotides. Although *E. coli* contains flavoproteins for coupling with heterologously expressed P450s, *in vivo* catalytic action progresses very slow, and living bacterial cells lack the endogenous electron donors therefore, there is a need for redox partners. Parikh and colleagues (1997) implemented the naturally occurring fusion proteins system of CYP102 from *Bacillus megaterium* and expressed P450 and CPR bicistronically in pCWOri^{+} plasmid, with the first cistron being the P450 and second cistron the CPR. The double tac promoters were located close to the first operon, and expression was initiated with the addition of IPTG. One mRNA was generated but two proteins were expressed. This expression system increased the redox potential due to co-localisation of both enzymes. The expression level of both operons were similar, while generated catalytic reaction levels were comparable to those obtained using human liver microsomes as well as those of reconstituted in vitro systems composed of purified proteins, lipids, and cofactors (Parikh et al., 1997). Additionally, they proved that certain P450s are not dependent on cytochrome b5 for catalytic activity, due to b5 being part of the electron chain supply, as their recombinantly expressed CYP3A4 displayed a high hydroxylation activity against testosterone without the need of cytochrome b5 (Parikh et al., 1997). Adding to this, Dodhia et al (2006) fused three N-terminally modified P450s to the soluble NADPH-depended oxidoreductase domain of CYP102A1 from *Bacillus megaterium* and co-expressed them as one single peptide in *E. coli*. The enzymes were purified without any detergents and were both active and properly folded without resulting in inclusion bodies. They exhibited enhanced solubility, and were

catalytically self-sufficient, while presented catalytic activities like the reconstituted native enzymes (Dodhia et al., 2006).

Blake and colleagues (1996) also co-expressed P450 and P450 reductase in pB216 plasmid. They used a cloning strategy where they fused the 5' prime end of the P450 reductase to the pelB bacterial leader sequence in plasmid pJR4, using the double Ptac promoters from pCWOri^{+} plasmid, and subcloned the whole cDNA into pB216 plasmid, creating head to tail arrangement with the already cloned CYP3A4 in it (Figure 1.5 in Appendix 1). They achieved a 40-fold increase in the expression level of the P450 reductase compared to the one expressed without the pelB leader sequence, while the content of the recombinant P450 and reductase enzymes in the membrane fractions were similar to those of human liver microsomes (Blake et al., 1996). On the other hand, Pritchard and colleagues used both bacterial leader sequences, pelB and OmpA, to clone and express with their human P450s (Blake et al., 1996, Pritchard et al., 1997). Pan and colleagues co-expressed the human P450s and oxireductase in two separate compatible plasmids. Specifically, they cloned the CYP in pCWOri^{+} and OxR in pACYC plasmid fused to the ompA bacterial leader sequence (Pan et al., 2011).

Moreover, soluble, self-sufficient, human P450- *Bacillus megaterium* reductase (BMR) constructs have been created by removing the hydrophobic N-terminal membrane anchor domain from the insoluble human enzyme before fusion to the BMR sequence (Talmann et al., 2017).

Another cloning and co-expression strategy that resulted in high catalytic activity, low maintenance costs, elimination of purification costs and the poor enzyme stability all together, was the utilisation of the bacterial whole cell catalyst to display the human P450 and reductase in the outer membrane of *E. coli*. Specifically, Quehl and colleagues used two autotransporters found in the outer membrane of gram-negative bacteria to transport their recombinant proteins to the outer membrane using two compatible plasmids. The recombinant proteins CPR and P450 were modified to contain the N- terminal CtxB signal peptide, had their transmembrane domains (amino acids 1–56 and 1–29, respectively) removed, and contained a C-terminal b-domain of the *E. coli* autotransporters EhaA and AIDA-I, respectively. After translation, the proteins were transported through the sec pathway to the inner membrane where the signal peptide is truncated, and the b-barrel is inserted in the outer membrane and the "passenger" proteins were transferred to the extracellular space (Quehl et al., 2016).

According to the lengthy literature review performed by Zelasko in 2013, most of the P450s expressed to date were cloned in pCWOri^{+} plasmid, followed by

pKK233, while very few P450s were cloned into the popular pET plasmid systems for hyper-expression in *E. coli* (Zelasko et al., 2013).

After generating an extensive literature review on existing methods for better expression and catalytic activity of P450, we developed three different cloning and expression strategies.

First cloning strategy involved the co-cloning of the P450s and the human CPR in the same plasmid, connected by a linker as fusion proteins, where a single peptide of both proteins will be generated during translation. This has the benefit of both P450s and human reductase expressing as a single peptide which can facilitate the catalytic activity of the P450s. This strategy reassures that both enzymes will be co-expressed and localised in the same compartment of the cell, rather than random localisation.

Second strategy involved cloning of P450s in a pET22b plasmid, and CPR in the compatible pCDFDuet-1 plasmid, for co-expression in the same *E. coli* strain. pET22b introduces a bacterial leader sequence to the N-terminus of the P450s, which redirects the expressed P450s in the periplasm of the bacteria. After translocation the pelB leader sequence is cleaved off by signal peptidases of the bacteria. Directing P450s to the periplasm has as a benefit of helping with protein 3-dimensional folding, and consequently better / or catalytically active enzyme. The creation of disulfide bonds, proper folding, and improved protein activity are provided by periplasm with a more oxidative area than the cytoplasm (Aghaei et al., 2016).

And finally, third strategy involved cloning of P450s in pET29b plasmid, and co-expressing them with the already cloned reductase in pCDFDuet-1 in the same *E. coli* strains. This strategy was devised as a comparison with co-expression of the clones in pET22b plasmids. pET29b plasmid does not contain a pelB leader sequence, and therefore the expressed enzyme is not translocated in the periplasm. In contrary, the expressed enzyme is located in the plasma membrane of the bacteria. With this comparison we wanted determine which cloning and expression strategy generates higher level of expression and catalytically active enzyme.

Expression condition of all human and metagenome genes involved using different *E. coli* strains, such as BL21(DE3), C41(DE3), C43(DE3) and Tuner cells. Different concentrations of IPTG were trialled with Tuner cells in order to find the best expression of enzymes and catalytic activity. Moreover, two different media and temperatures were employed: Autoinduction media (AIM) and Luria Bertani broth (LB) in both, 20 °C and 30 °C. Another parameter trialled was that of co-expression with the pGro7 plasmid expressing chaperones GroES (60 kDa) and GroEL (10 kDa), which help with protein folding, and hence active enzyme yield.

Expressing media were supplemented with different concentrations of 5-ALA, in order to determine best working concentration for higher protein expression, as 5-ALA facilitates with haem synthesis. Moreover, addition of ferrous and ferric chloride to the media was implemented, which can enhance protein expression through 5-Ala synthesis. CO assays were performed in all the cultures co-expressing the P450s and CPR. A culture expressing an empty plasmid was used as the negative control, for determining the suitability of the assay. Additionally, whole cell biotransformations (WCB) were performed on all CO active enzymes, and samples were collected at set timepoint of 0 h, 15 min, 3 h and 24 h. Briefly, 14 mL worth of *E. coli* cells were diluted in 1 mL 0.05 mM potassium phosphate containing 5 % acetonitrile, 4 % glycerol and 10 μM substrate final concentration. The reaction was initiated with the addition of 25 μL of 10 mg/mL (250 μg) NADPH. Ketamine was used as a positive control substrate for CP2B6, CYP2D6, CP3A5 and CP3A4, while coumarin was used as a positive control substrate for CP2A6, and norketamine was the test substrate for all expressed and CO-active enzymes. In contrary, metagenome enzymes were novel, and therefore ketamine and norketamine were both used as test substrates. The choice of the substrates was based in literature (Table 3-1). CYP2D6 has already been proved to N-demethylate ketamine to norketamine (Portmann et al., 2010b).

The initial substrate concentration was 0.5 mg/mL, and all the WCB datapoints were analysed via HPLC, but that did not result in any catalytic activity even for the positive controls (data not shown). Potentially, the substrate concentration was too high and was inhibiting the enzyme. So, the substrate concentration was changed to final concentration of 10 μM, and the results were analysed by liquid chromatography-mass-spectrometry (LC-MS). All the cloning and expression strategies are explained further down in the thesis.

Table 3-1 Control substrates used for the characterised human genes. Ketamine was used as control substrate for CYP2D6, CYP2B6, CYP3A5, CYP3A4. Coumarin was used as a control substrate for CYP2A6, and ketamine and norketamine were used as substrates for the novel metagenome genes.

Gene	Hydroxylation activity	According to research
CYP2A6	Ketamine	(Desta et al., 2012)
CYP2A6	Coumarin	(Pelkonen et al., 2000)
CYP2B6	Ketamine	(Desta et al., 2012)
CYP3A5	Ketamine	(Desta et al., 2012)
CYP3A4	Ketamine	(Desta et al., 2012)
CYP2D6	Coumarin, ketamine	(Nakamura et al., 2001) (Portmann et al., 2010b)
Meta 198	Novel (test substrate: norketamine and ketamine)	-
Meta 914	Novel (test substrate: norketamine and ketamine)	-

3.1.1. N-terminus Modification of Enzymes

Before any cloning took place, the N-terminus hydrophobic signal peptide of the human enzymes were deleted, and the negative charged transmembrane domain that anchors the protein to the endoplasmic reticulum was replaced by a previously recognised positively charged sequence, LLLAVFL (Barnes et al., 1991). The modification was introduced through PCR mutagenesis for CP2B6, CP2D6, CP3A4 and CP3A5, while CP2A6 was modified through gene synthesis. Different N-terminus modification was applied to the genes according to the literature (Table 3-2).

Both metagenome genes homologous to CYP3A5, 198 and 914, were cloned into three different plasmids with and without the N-terminal modification. The sequences of these two genes were presented as a hit for human P450 3A5 after mining the Prozomix database, and their primers were based on their sequences.

The metagenome genes were amplified from a sample called Meta-4 that contained a variety of Prozomix-provided genes. As the metagenome genes are in fact bacterial genes, we did not know if the modification would generate adequate yields of expressed enzyme, or enhance enzyme solubility, or, expressing the original non-modified version of the metagenome genes would generate more active and/ or soluble enzyme, therefore both strategies were employed (Table 3-3).

Table 3-2 N-terminus modification of the CYP forms.

CYP form	Amino acid Sequence deleted	Amino acid Sequence added in	Modification performed	According to publication:
CYP2A6	MLASGMLLVALLVCLTVMVLMSVWQQRKSK	MAKKTSSK	Through Gene Synthesis	(Smith *et al*., 2007, DeVore *et al*., 2008)
CYP2B6	LSVLLFLALLTGLLLLLVQ	MEKKTSSKGK	Through PCR mutagenesis	(Scott *et al*., 2001, Zhao and Halpert, 2007)
CYP2D6	MGLEALVPLAVIVAIFLLLVDLMHRRQRWAARY	MARQVHSSWNL	Through PCR mutagenesis	*(Gillam et al., 1995)*
CYP3A5	MDLIPNLA	MALLLAVF	Through PCR mutagenesis	(Lee et al., 2007)
CYP3A4	MALIPDLAMETWLLLAVSLVLLYLYGTHS	MALLLAVF	Through PCR mutagenesis	(Pan *et al*., 2011)

Table 3-3 N-terminus modification of the Metagenomic DNA.

Gene sequence from Metagenome 4	Score identity to CP3A5	Amino acid Sequence added	Modification performed
S0zJJln2Kzf8zW_C33315_O22890198	20%	MALLLAVF	Through PCR mutagenesis
S0zJJln2Kzf8zW_C5419_O20953914	24%	MALLLAVF	Through PCR mutagenesis

3.2. Co-Cloning and Expression of P450s and NADPH-dependent CPR using Bio-Bricks.

As our aim was to express these human enzymes, and to determine if they have any catalytic activity against Norketamine, so we employed Dodhia's method with adaptions as the first method to be used for co-expression of P450's and CPR (Appendix 1, Figure 1.5). We used *Hin*dIII as linker between the two genes, instead of *AvrII*, while human reductase was used instead of the BMR. *Nde*I and *Hind*III were the two non-destroyed restriction sites in pCWOri^{+} when we purchased it, so we opted for *Hind*III as a linker. Moreover, we decided to clone them in parallel in pET29b, and determine which plasmid confers better expression-catalytic activity. pCWOri^{+} plasmid is a low copy number cloning plasmid, whereas pET29b is high copy number expression plasmid, so we decided to co-clone in both and determine best expression system.

3.2.1. Co-Cloning of P450s and CPR

Different primers (Appendix 1) were generated for the co-cloning of the P450s and the CPR in pET29b and pCWOri^{+}, and different restriction sites existed in each plasmid. Moreover, all genes were cloned in both plasmids with and without the 6x His-tag incorporated in the C-terminus of the CPR. It has been mentioned that the His-tag can interact with the active site and substrate binding, hence the need for creating a construct without a His-tag.

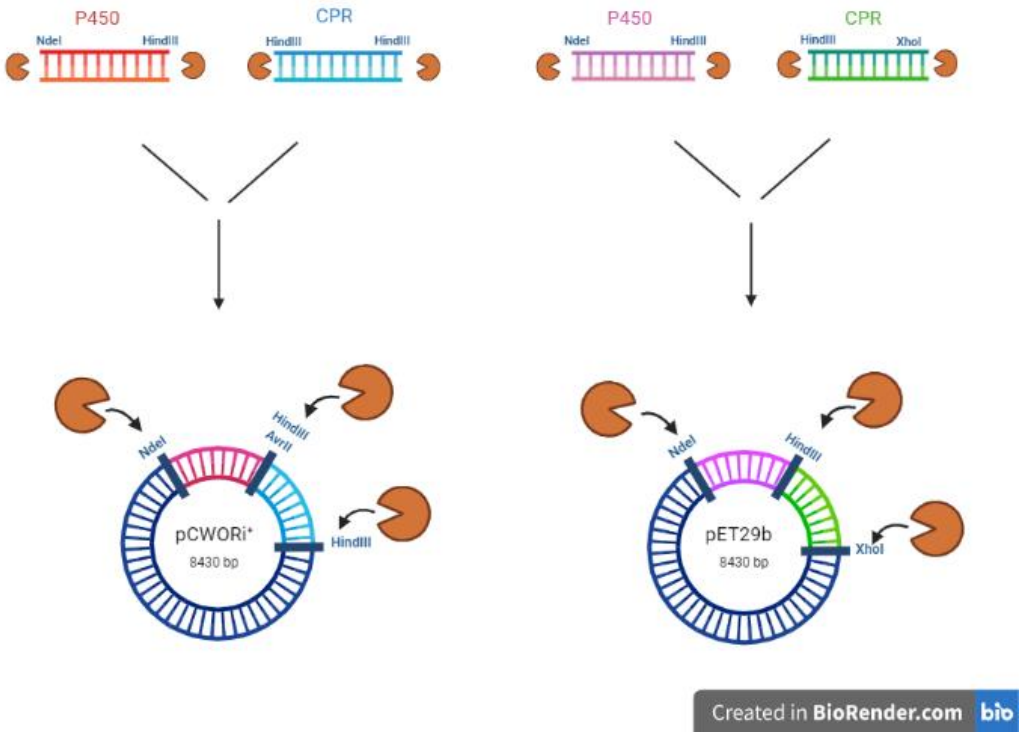


Figure 3-1 Representation of co-cloning strategy of P450s and CPR in both plasmids. All P450s were co-cloned with the CPR into pCWOri^{+} and pET29b plasmids using different combination of primers. For pCWOri^{+} all P450s were digested with NdeI-HindIII and cloned in, then pCWOri^{+} was linearised again only with HindIII and CPR digested with HindIII both sides, was cloned in (assembly on the left-hand side of the figure). For pET29b all P450s were digested with NdeI- HindIII and ligated into pET29b. Then pET29b plasmid was re-digested with HindIII -XhoI and CPR was cloned in (assembly on the right-hand side of the figure). Figure created with BioRender.com.

The cloning (Figure 3-1) involved both plasmids, pCWOri$^+$ and pET29b being digested with *Nde*I and *Hind*III restriction enzymes, and the P450 genes digested with the same restriction enzymes, were cloned in using T4 ligase. Then pCWOri$^+$ was linearised again with *Hind*III, was treated with Antarctic phosphatase, and the CPR gene digested with *Hind*III both sides was cloned into the plasmid. On the other hand, pET29b was re-digested with *Hind*III and *Xho*I and *Hin*dIII – *Xho*I cut CPR was cloned in.

The generated constructs are arranged in Table 3-4 below [(-) refers to no His-Tag used, while (+) refers to His tag incorporated at C-terminus].

Table 3-4 All constructs generated from the co-cloning strategy.

Construct	Usable Clones
CP2A6-CPR-pCWOri$^+$ (-)	(Introduced termination codon at cloning)
CP2A6-CPR-pCWOri$^+$ (+)	(Introduced termination codon at cloning)
CP2A6-CPR-pET29b (+)	√
CP2A6-CPR-pET29b (-)	√
CP2B6-CPR-pCWOri$^+$ (-)	√
CP2B6-CPR-pCWOri$^+$ (+)	√
CP2B6-CPR-pET29b (+)	√
CP2B6-CPR-pET29b (-)	√
CP3A5-CPR--pCWOri$^+$ (-)	√
CP3A5-CPR- pCWOri$^+$ (+)	√
CP3A5-CPR-pET29b (+)	√
CP3A5-CPR-pET29b (-)	√
CP3A4-CPR-pCWOri$^+$ (-)	√
CP3A4-CPR- pCWOri+ (+)	(Introduced a 5 amino acid deletion at CPR sequence)
CP3A4-CPR-pET29b (+)	√
CP3A4-CPR-pET29b (-)	√
CP2D6-CPR-pCWOri+ (-)	√
CP2D6-CPR-pCWOri+ (+)	√
CP2D6-CPR-pET29b (+)	√
CP2D6-CPR-pET29b (-)	√

3.2.2. Expression in BL21(DE3) with 0.5 mM 5-ALA with and without pGro7 in both LB and AIM at 20 °C and 30 °C.

In this experiment all expressions were performed in BL21(DE3) with and without pGro7 chaperones, in LB and AIM at 20 °C and 30 °C induced with 0.5 mM 5-ALA. Good expression was observed for CP2A6-CPR-pET29b (-) without pGro7 in LB 20°C at 130.7 kDa (Figure 3-2 **A,** lane 3 indicated by a red arrow), and active through CO (Figure 3-3), while less expression was observed for the same enzyme (lane 9) in LB 30 °C. The same expression experiment was performed with pGro7, and again CYP2A6-CPR-pET29b (-) in LB 20 °C was expressed, although less expression was observed (Figure 3-2 **B**), but the enzyme did not produce the characteristic Soret peak at 450 nm, when the CO assay was performed. There was no active enzyme for this instance when co-expressed with the chaperones.

Moreover, the expression experiment in BL21(DE3) was repeated in AIM 20°C and 30 °C with and without pGro7 (Figure 3-2 **C**). Again, CYP2A6-CPR-pET29b (-) expressed without pGro7, but no active enzyme could be determined on the CO reduced spectra. In this instance AIM, being a richer media, did not have a positive impact in the expression of the enzymes, instead better expression was observed in LB media. The MS sample number, which serves as a reference for all MS samples, and lists all the conditions under which the protein was synthesised, is presented in Table 3-8. The primary purpose of the number is to serve as a cross reference between the CO peak and the circumstances that produced it (some samples were analysed by HPLC hence no data for them).

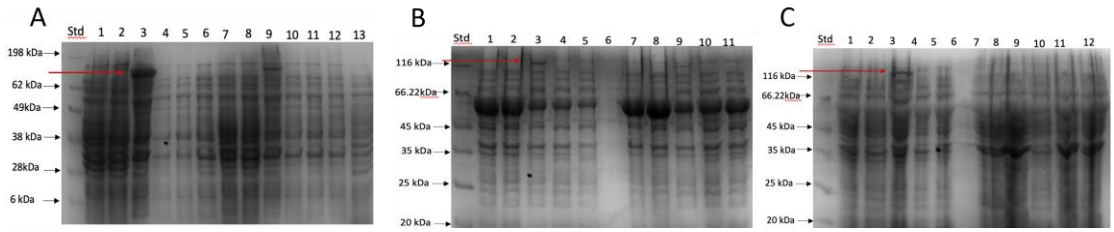


Figure 3-2 Total protein of CYP2A6-CPR fusion in pET29b expressed n BL21(DE3) and induced with 0.5 mM 5-ALA (TCP). In figure **A,** lanes1-6 are shown expression conditions without pGro7 in BL21(DE3) in LB 20°C, and in lanes 7-12 in LB 30 °C. In figure **B**, in lanes1-6 are shown expression conditions with pGro7 in BL21(DE3) in LB 20°C, and in lanes 7-12 in LB 30 °C. Figure C, lanes 1-6 show expression conditions with pGro7 in BL21(DE3) in AIM 20°C, and in lanes 7-12 in AIM 30 °C. From left; protein standard (ThermoFisher Scientific), lane 1; CP2A6-CPR- pCWOri^{+} (-) in LB 20 °C, lane 2; CP2B6-CPR-pCWOri^{+} (-) in LB 20 °C, lane 3; CP2A6- CPR- pET29b (-) in LB 20 °C, lane 4; CP3A5-CPR-pET29b (+) in LB 20 °C, lane 5; CP3A5-CPR-pET29b (-) in LB 20 °C, lane 6; empty plasmid, lane 7; CP2A6-CPR- pCWOri^{+} (-) in LB 30 °C, lane 8; CP2B6-CPR-pCWOri^{+} (-) in LB 30 °C, lane 9; CP2A6- CPR- pET29b (-) in LB 30 °C, lane 10; CP3A5-CPR-pET29b (+) in LB 30 °C, lane 11; CP3A5-CPR-pET29b (-) in LB 30 °C. CP2A6-CPR-pET29b at 130.7 kDa has been expressed only in LB 20 °C and AIM 20 °C with and without pGro7 as can be observed in lane 3 of all three gels, (indicated by a red arrow), while CP2A6- CPR- pET29b (-) in LB 30 °C expressed only without pGro7 (gel **A** lane 9).

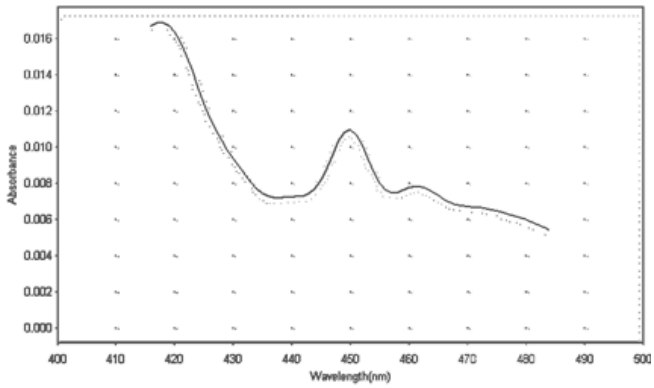


Figure 3-3 CO spectrum (400-500nm) of CYP2A6-CPR-pET29b (-) fusion expressed in LB 20 °C without pGro7, induced with 0.5 mM 5-ALA. (MS sample number 5, Table 3-8).

3.2.3. Expression in BL21(DE3) without pGro7 in LB at 20 °C and 30 °C, with 0.25 mM 5-ALA, and 0.25 mM $FeCl^{2+}$ and $FeCl^{3+}$

According to Miura and Kumondai, supplementing the media with 0.25 mM ferrous and ferric chloride, and lowering the concentration of the 5-ALA from 0.5 mM to 0.25 mM may enhance the P450 expression and activity (Miura et al., 2015, Kumondai et al., 2020). So, we implemented this method and indeed, different enzymes were expressed in these conditions. Additionally, most of the expressed enzymes were catalytically active, apart from one, CYP2B6- CPR- pET29b (-) in LB 20 °C and 30 °C, which expressed well, as seen in SDS gel (Figure 3-4) but was not active in a CO assay.

The rest, CYP3A4-CPR-pCWOri^{+} (+) in LB 30 °C (Figure 3-5 **A**), CYP3A4-CPR-pET29b (+) in LB 20 °C (Figure 3-5 **B**), CYP3A4-CPR-pET29b (-) in LB 20 °C (Figure 3-5 **C**), CYP3A5- CPR-pCWOri^{+} (+) in LB 20 °C (Figure 3-5 **D**) when expressed without pGro7, were all active against CO assay.

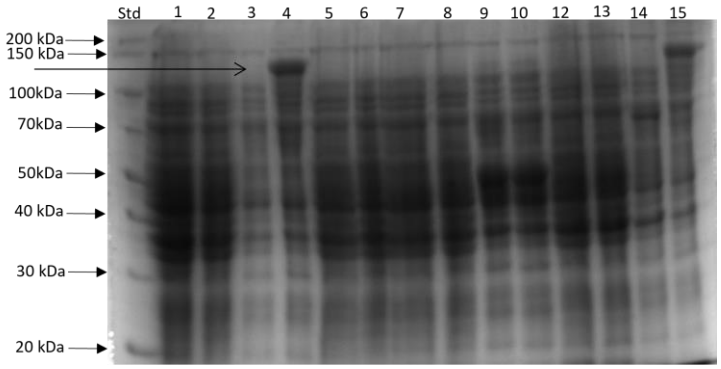


Figure 3-4 Total protein of CYP2B6-CPR-pET29b fusion expressed in BL21(DE3) without pGro7, and induced with 0.25 mM 5-ALA, and 0.25 mM Fe^{2+} and Fe^{3+} chloride. protein standard (ThermoFisher Scientific), lane 1; 2A6-CPR-pCWOri^{+} (-) in LB 20 °C, lane 2; 2A6-CPR- pCWOri^{+} (+) in LB 20 °C, lane 3; 2B6-CPR-pET29b (+) in LB 20 °C, lane 4; CYP2B6- CPR- pET29b (-) in LB 20 °C, lane 5; 3A5-CPR-pCWOri^{+} (-) in LB 20 °C, lane 6; 3A5-CPR- pCWOri^{+} (+) in LB 20 °C, lane 7; 3A4-CPR- pCWOri^{+} (-) in LB 20 °C, lane 8; 3A4-CPR- pCWOri^{+} (+) in LB 20 °C, lane 9; 3A4-CPR-pET29b (+) in LB 20 °C, lane 10; 3A4-CPR-pET29b (-) in LB 20 °C, lane 12; 2A6-CPR-pCWOri^{+} (-) in LB 30 °C, lane 13; 2A6-CPR- pCWOri^{+} (+) in LB 30 °C, lane 14; 2B6-CPR-pET29b (+) in LB 30 °C, lane 15; CYP2B6- CPR- pET29b (-) in LB 30 °C.

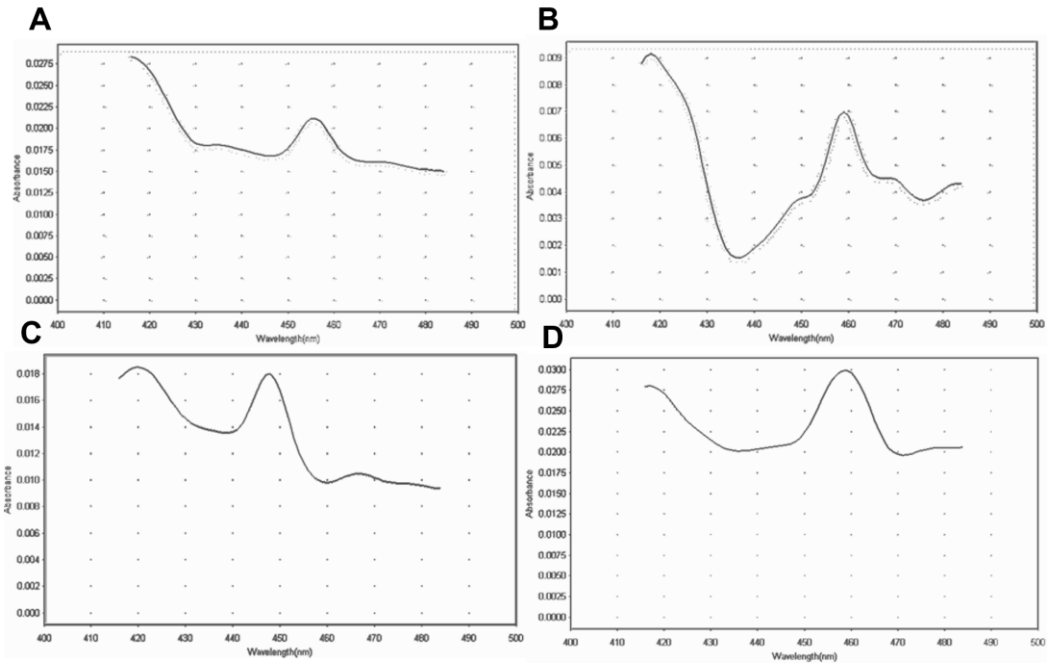


Figure 3-5 CO spectrum (400-500nm) of fused P450s expressed without pGro7 in BL21(DE3), induced with 0.25 mM 5-ALA and 0.25 mM $FeCl^{2+}$ and $FeCl^{3+}$. **A**: CYP3A4-CPR-pCWOri^{+} (+) without pGro7 in LB 30 °C, **B**: CYP3A4-pET29b (+) in LB 20 °C, **C**: CYP3A4-pET29b (-) in LB 20 °C and **D**: CYP3A5- CPR- pCWOri^{+} (+) in LB 20 °C.

3.2.4. Expression in C41(DE3) with and without pGro7, with 0.5 mM 5-ALA (no $FeCl^{2+}$ and $FeCl^{3+}$) in LB and AIM at 20 °C and 30 °C.

Initially an expression experiment was performed in C41(DE3) and CD43(DE3) with and without pGro7 in LB 20 °C and LB 30 °C to determine if high expression levels of P450 can be achieved, but expression was achieved only in C41(DE3) (Appendix 2, Figure 2.1 to 2.4, lanes 3 and 4, and 9 to 11). As mentioned before, C41(DE3) and CD43(DE3) strains are widely used to overexpress membrane proteins. Ichinose and colleagues achieved recombinant expression yields of more than 2000 nmol/L of culture for 141 fungal P450s when co-expressing with pGro7 in C41(DE3) strain, although it was not as successful for other fungal P450s (Ichinose et al., 2015). Additionally, Cheng and colleagues reported the same remarkably high expression levels of recombinant CYP2E1, achieving 900–1400 nmol/L culture (Cheng et al., 2004).

So consequently, expression was attempted in in *E. coli* C41(DE3) in LB and AIM at 20 °C and 30 °C supplemented with $FeCl^{2+}$ and $FeCl^{3+}$. All fused genes were co-expressed with and without pGro7 plasmid.

The cultures exhibited delayed growth, and the culture densities were lower when compared to expression in BL21(DE3). Active enzyme was achieved only for two enzymes (cloned with and without his tag) when expressed without the presence of the chaperones GroEL and GroES: CYP2B6-CPR-pET29b (+) in LB 20 °C (Figure 3-6 **A**), CYP2B6-CPR-pET29b (-) in LB 20 °C (Figure 3-6 **B**), CYP3A5-CPR-pET29b (+) in LB 30 °C (Figure 3-6 **C**) and CYP3A5-CPR-pET29b (-) in LB 30 °C (Figure 3-6 **D**) as seen via CO assays.

SDS-PAGE analysis of the enzymes demonstrated very low expression, very fine protein bands were visible on SDS gels (Appendix 2, Figures 2.5 to 2.9), which corresponded with CO activities. Moreover, when performing CO assays, no active enzyme was recovered from AIM, at both temperatures, which indicated that AIM did not have a positive impact on expression. This was backed up from our previous experiment when expressing in BL21(DE3), so all expression experiments in AIM were ended at this point and not continued in the next experiments.

Another experiment was performed in C41(DE3) but this time with increased concentration of 5-ALA from 0.5 mM to 1 mM to determine if higher expression and/or catalytic activity can be achieved with higher concentrations of 5-ALA. Only one enzyme, CYP2B6-CPR-pET29b (+) in LB 20 °C with pGro7 (Figure 3-7 **A**), and CYP2B6-CPR-pET29b (-) in LB 20 °C without pGro7 (Figure 3-7 **B**) was active according to the CO reduced spectra, although their activity had increased

compared to lower concentration of 5-ALA. So, 5-ALA had an impact on the expression and activity of CYP2B6 enzyme expressed in C41(DE3) without pGro7. Higher concentrations of 5-ALA had resulted in better expression of CYP2B6, so the concentration of 1 mM was determined to be the optimal concentration was employed in the rest of expression experiments.

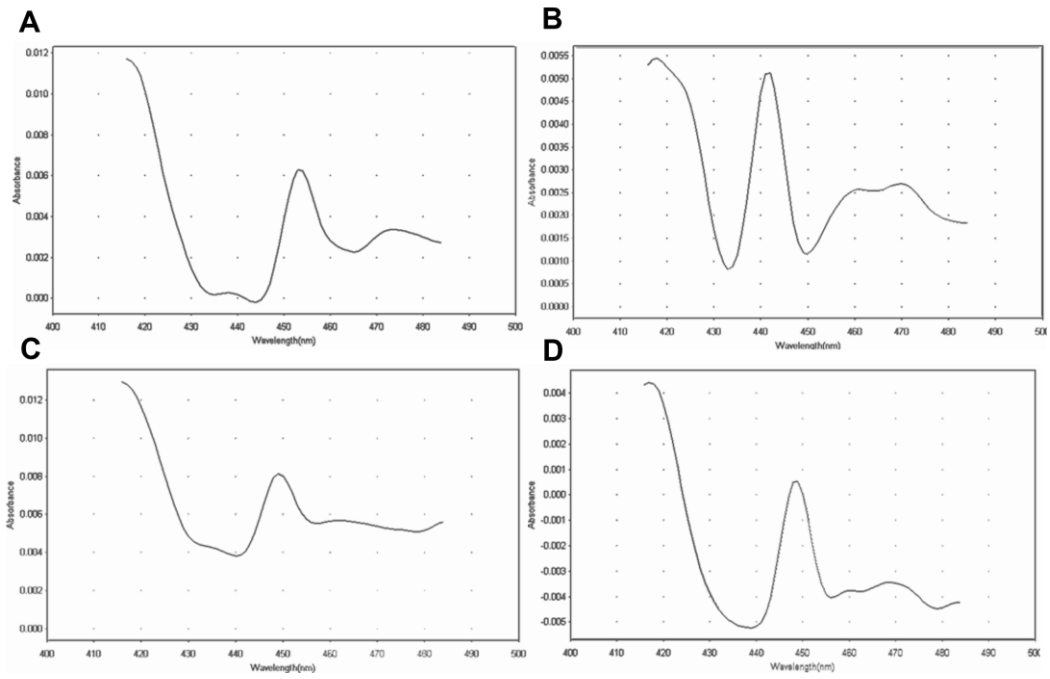


Figure 3-6 CO spectrum (400-500nm) of fused P450s expressed without pGro7 in C41(DE3), induced with 0.5 mM 5-ALA ((no $FeCl^{2+}$ and $FeCl^{3+}$). **A**: CYP2B6-CPR-pET29b (+) in LB 20 °C, **B**: CYP2B6-CPR-pET29b (-) in LB 20 °C, **C**: CYP3A5-CPR-pET29b (+) in LB 30 °C, **D**: CYP3A5-CPR-pET29b (-) in LB 30 °C.

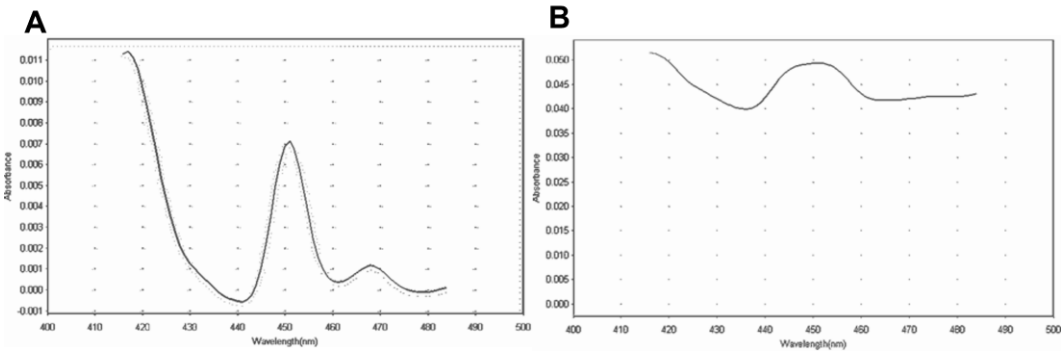


Figure 3-7 CO spectrum (400-500nm) of fused P450s expressed in C41(DE3) induced with 1 mM 5-ALA ((no $FeCl^{2+}$ and $FeCl^{3+}$). **A**: CYP2B6-CPR-pET29b (+) in LB 20 °C + pGro7, (MS sample number 6, Table 3-7), **B**: CYP2B6-CPR-pET29b (-) without pGro7 in LB 20 °C.

3.2.5. Expression in Tuner cells with 0.25 mM IPTG, with and without pGro7 in LB at 20 °C and 30 °C.

IPTG concentration can have a detrimental effect in the recombinant enzyme production. High protein expression in a high copy number plasmid can lead to metabolic imbalance to *E. coli* cells, which can result in the misfolding of the protein and inclusion bodies (Bhatwa et al., 2021). As mentioned previously, *E. coli* Tuner cells are mutants of BL21(DE3) with deleted lacZY, which enables them to adjust the levels of protein expression throughout the cells in a culture, depending on IPTG concentration. Moreover, protein over expression has proven problematic due to difficulties with folding, lack of function following overexpression, and the overexpression process results in cell toxicity and cell count loss according to James et al, 2021. The recommended IPTG concentration for BL21(DE3) is between 0. 4 and 1 mM IPTG, so we reduced the lowest recommended IPTG concentration by almost half in order to reduce the rate of expression, which increases the protein folding, but also the cell viability and protein production capacity (James et al., 2021, Studier, 2005). So, an IPTG concentration of 0.25 mM was used for this experiment. Co-expression with the chaperones GroEL and GroES resulted in active enzymes CYP2A6-CPR-pET29b + pGRo7 (-) LB 30 °C (Figure 3-9 **A**), CYP2B6-CPR-pET29b + pGRo7 (+) in LB 20 °C (Figure 3-9 **B**), CYP2B6-CPR-pET29b + pGRo7 (-) in LB 20°C (Figure 3-9 **C**), and CYP3A4-CPR-pET29b + pGRo7 (-) in LB 30 °C (Figure 3-9 **D**), while expression without the chaperones generated active enzyme CYP2A6-CPR-pET29b (+) (Figure 3-9 **E**) and CYP3A5-CPR-pET29b (+) (Figure 3-9 **F**).

Also, expressed but not active enzyme was observed for CYP2A6-CPR-pET29b (E) and CP2B6 - CPR - pET29b (-) without pGro7 (Figure 3-8). WCB assays were performed to the active enzyme cultures, and the timepoint samples were analysed by LC-MS.

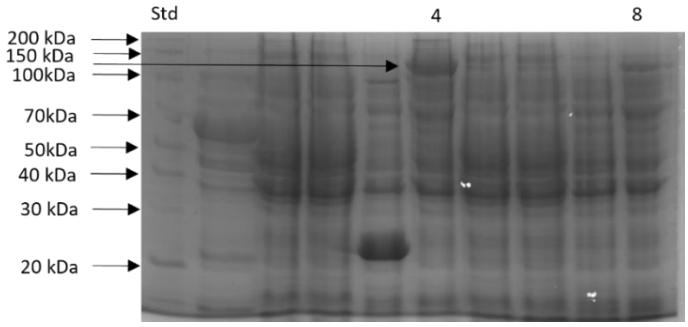


Figure 3-8 Total protein of fused P450s expressed in Tuner cells in LB 20°C and induced with 0.25 mM IPTG. From left; protein standard (ThermoFisher Scientific), lane 4; CP2A6-CPR-pET29b (-) + pGro7 at 130.7 kDa, and lane 8; CP2B6-CPR-pET29b (-) at 131.7 kDa have been expressed without pGro7, as indicated by a black arrow.

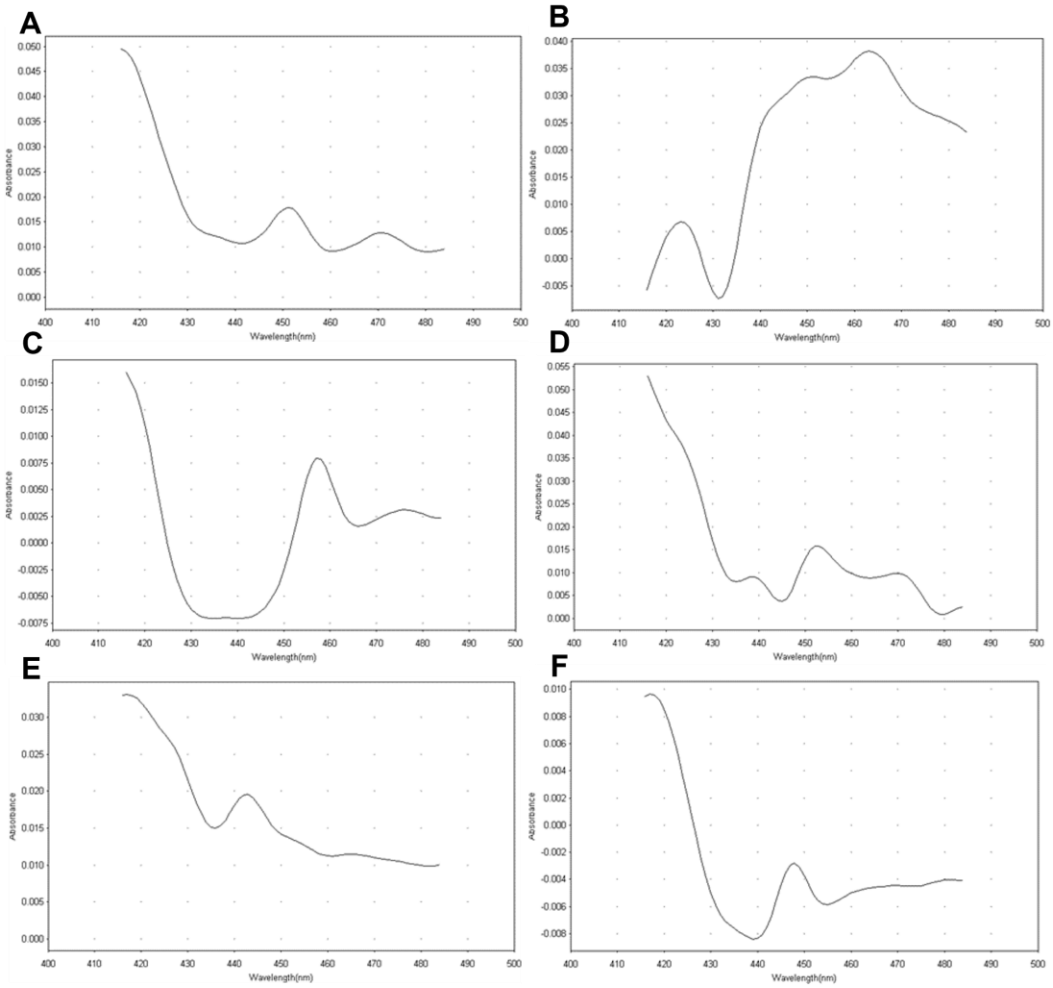


Figure 3-9 CO spectrum (400-500 nm) fused P450s expressed in Tuner and induced with 0.25 mM IPTG. **A**: CYP2A6-pET29b+ pGRo7 (-) in LB 30 °C, **B**: CYP2B6-pET29b+ pGRo7 (+) in LB 20 °C (MS sample number 7), **C**: CYP2B6-pET29b+ pGRo7 (-) in LB 20 °C (MS sample number 8), **D**: CYP3A4-pCWOri^{+} with pGRo7 (+) in LB 30 °C (MS sample number 16), **E**: CYP2A6-CPR-pET29b (+) without pGro7 in LB 20 °C (MS sample number 3), **F**: CYP3A5-CPR-pET29b (+) without pGro7 in LB 30 °C (MS sample number 11). All MS sample numbers are summarised in Table 3-8.

3.3. Co-expression of CYPs and CPR in pET22b and pCDFDuet-1 respectively in *E. coli* cells

As already stated, cloning P450s with a bacterial leader sequence, such as pelB or OmpA, can direct the CYP sequence in the periplasm, which can help with proper protein folding and hence catalytic activity. All target human P450s were cloned into pET22b plasmid, which has already incorporated in the backbone of the plasmid a sequence encoding a pelB leader sequence. Cloning in between *Nco*I and *Xho*I restriction endonucleases can introduce the pelB leader sequence to the *N*-terminus of the protein, where all will be translated as one protein, and exported to the periplasm, where the pelB sequence is cleaved by signal peptidases of the bacteria. Additionally, CPR was cloned in the compatible plasmid pCDFDuet-1. All cloning and their co-expression strategy is detailed below in Figure 3-10 (primers in Appendix 1). Both genes were co-expressed in *E. coli* BL21(DE3, C41(DE3) and Tuner strains, in different temperatures and expression conditions to determine best expression and catalytic activity. All genes were cloned between restriction endonucleases *Nco*I and *Xho*I into pET22b, apart from CYP3A5 that was cloned between *BamH*I-*Xho*I. CPR was cloned with *BamH*I-*Hind*III in pCDFDuet-1.

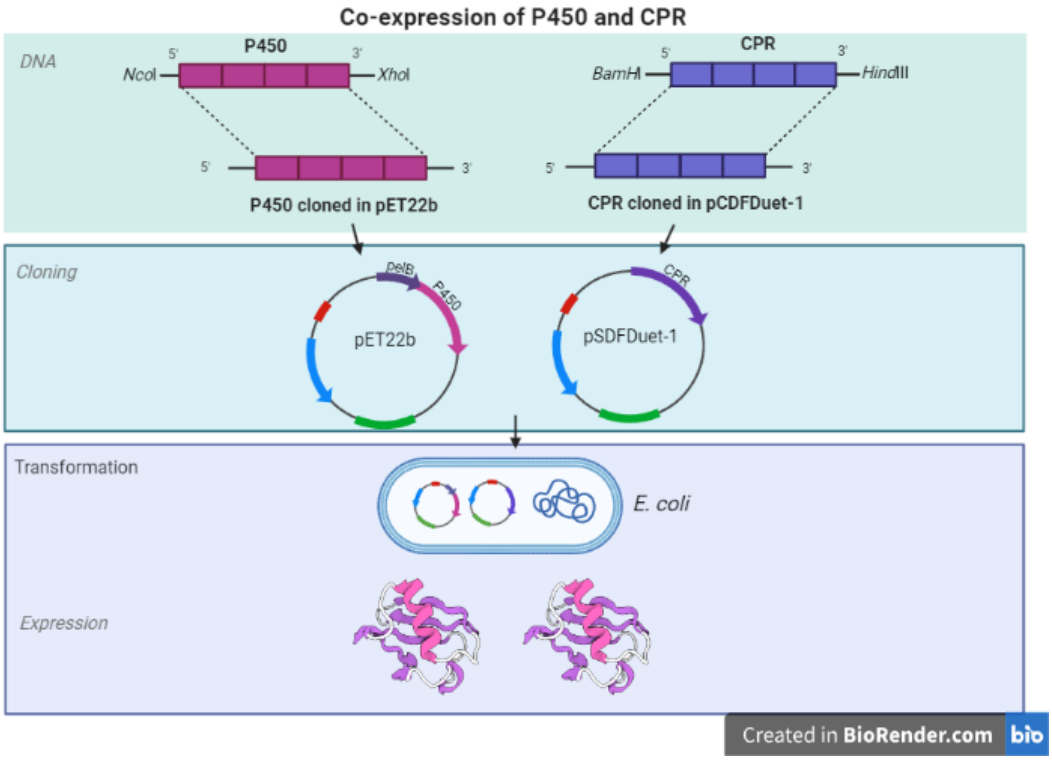


Figure 3-10 P450s and CPR cloned in pET22b and pCDFDuet-1 respectively, and their simultaneous expression in *E. coli*.

3.3.1. Co-Expression with and without pGRo7 in BL21(DE3) and C41(DE3) in LB 20 °C and 30 °C, induced with 1 mM IPTG.

Parallel expression of each clone took place in BL21(DE3) and C41(DE3) with and without pGro7, in LB 20 °C and 30 °C, to compare their expression, and were separated in SDS gels side by side. Good expression level was achieved for CPR when expressed with pGro7 in BL21(DE3), in LB 20 °C and 30 °C, as seen in the SDS gels (Figure 3-11 **A** and **B**, lanes, 5, 7, 9, 11, 13, 15, 17 and 19 at 77.5 kDa), and without pGro7 in BL21(DE3) (Figure 3-11, **C**, lanes 1- 10, at 77 kDa), whereas no expression of CPR was achieved in C41(DE3) with and without the chaperones (Figure 3-11 **A** and **B**, lanes 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, and SDS gel **D** entirely).

It's difficult to say if the P450s showed any expression when co-expressed with the chaperones, as all P450 proteins of approximately 55 kDa were masked by the chaperone expression GroES at 60 kDa. In contrary, good expression can be seen for CYP2A6 and CYP2B6 in LB 20 °C and 30 °C, CYP3A5 in LB 20 °C, and CYP2D6 in LB 20 °C and 30 °C when expressed in BL21(DE3) without pGRo7 (Figure 3-11 **C** indicated by an arrow at approximately 55 kDa), whereas no expression of any of the P450s can be seen in C41(DE3) in any condition (Figure 3-11 **D**).

In BL21(DE3), CO-active enzyme was obtained for CYP2B6-pET22b in both LB 20 °C and 30 °C (Figure 3-12 **A** and **B**), and CYP3A5-pET22b in LB 30 °C when co-expressed with pGro7 (Figure 3-12 **C**), whereas CYP2D6-pET22b in LB 20 °C when expressed without pGro7 (Figure 3-12 **D**).

In contrary, active enzyme was obtained only for CYP2A6-pET22b (Figure 3-12 **E**) and CYP3A4-pET22b (Figure 3-12 **F**) when co-expressed with pGRo7 in LB 20 °C. All active enzymes were used in WCB assays and analysed by LC-MS.

As can be observed from Figure 3-12 spectra, the P450 peak is sometimes close to 450 nm, whereas other times is close to 460 nm. Despite the fact that all P450s belong to the same enzyme family, and that the P450 peak should be the same, the isozyme can occasionally affect the peak. When a significant fraction of the total P450 is made up of individual P450 proteins with peaks that deviate by one or two nanometres from the 450-nm norm, the peak of the whole P450 may shift as a result (Feyereisen, 2019).

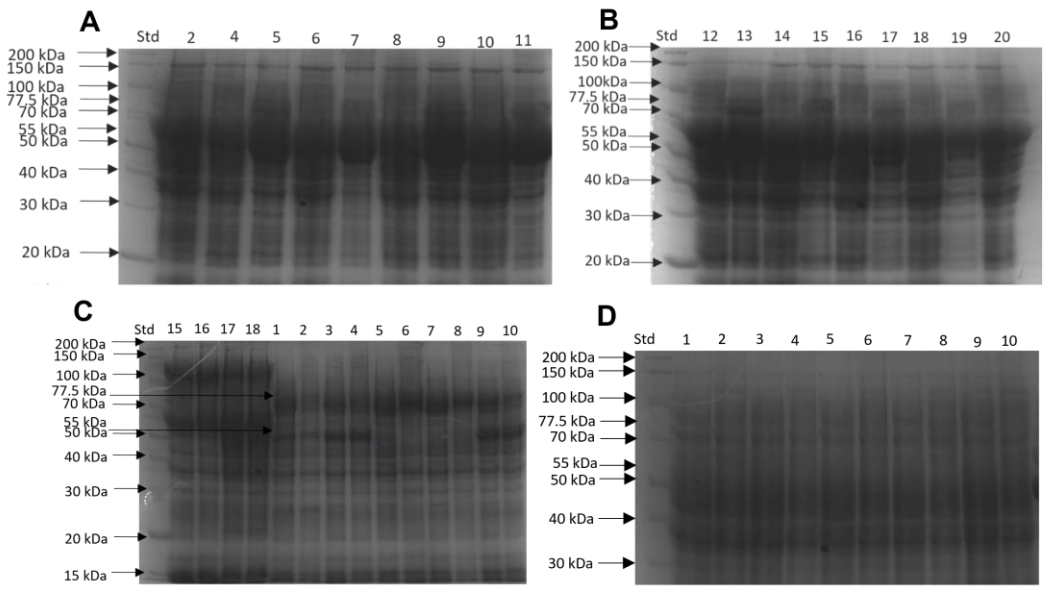


Figure 3-11 A comparison of co-expressed P450s-pET22b and CPR-pCDFDuet-1 total protein in BL21(DE3) and C41(DE3) induced with 1 mM IPTG. A: expressed with pGro7; From left; protein standard (ThermoFisher Scientific), lane 2; CYP2A6 - pET22b in C41(DE3) in LB 20 °C, lane 3; CYP2A6 - pET22b in BL21(DE3) in LB 30 °C, lane 4; CYP2A6 - pET22b in C41(DE3) in LB 30 °C, lane 5; CYP2B6 - pET22b in BL21(DE3) in LB 20 °C (55.9 kDa), lane 6; CYP2B6 - pET22b in C41(DE3) in LB 20 °C, lane 7; CYP2B6 - pET22b BL21(DE3) in LB 30 °C, lane 8; CYP2B6 - pET22b in C41(DE3) in LB 30 °C, lane 9; CYP3A5 - pET22b in BL21(DE3) in LB 20 °C (57.9 kDa), lane 10; CYP3A5 - pET22b in C41(DE3) in LB 20 °C, lane 11; CYP3A5 - pET22b in BL21(DE3)in LB 30 °C.

B: expressed with pGro7; protein standard, lane 12; CYP3A5 - pET22b in C41(DE3) in LB 30 °C, lane 13; CYP3A4 - pET22b in BL21(DE3) in LB 20 °C (55.7 kDa), lane 14; CYP3A4 - pET22b in C41(DE3) in LB 20 °C, lane 15; CYP3A4 - pET22b in BL21(DE3) in LB 30 °C, lane 16; CYP3A4 - pET22b in C41(DE3) in LB 30 °C, lane 17; CYP2D6 - pET22b in BL21(DE3) in LB 20 °C (54 kDa), lane 18; CYP2D6 - pET22b in C41(DE3) in LB 20 °C, lane 19; CYP2D6 - pET22b in BL21(DE3) in LB 30 °C, lane 20; CYP2D6 - pET22b in C41(DE3) in LB 30 °C.

C: expressed without pGro7 in BL21(DE3); protein standard, lane 1; CYP2A6 - pET22b in LB 20 °C (54.8 kDa), lane 2; CYP2A6 - pET22b in LB 30 °C, lane 3; CYP2B6 - pET22b in LB 20 °C (55.9 kDa), lane 4; CYP2B6 - pET22b in LB 30 °C, lane 5; CYP3A5 - pET22b in LB 20 °C (57.9 kDa), lane 6; CYP3A5 - pET22b in LB 30 °C, lane 7; CYP3A4 - pET22b in LB 20 °C (55.7 kDa), lane 8; CYP3A4 - pET22b in LB 30 °C, lane 9; CYP2D6 - pET22b in LB 20 °C (54 kDa), lane 10; CYP2D6 - pET22b in LB 30 °C (all indicated by a black arrow at 55k Da, second from the top). The top black arrow indicates the co-expressed CPR enzyme at 77.5 kDa.

D: expressed without pGro7 in C41(DE3); protein standard, lane 1; CYP2A6 - pET22b in LB 20 °C (54.8 kDa), lane 2; CYP2A6 - pET22b in LB 30 °C, lane 3; CYP2B6 - pET22b in LB 20 °C (55.9 kDa), lane 4; CYP2B6 - pET22b in LB 30 °C, lane 5; CYP3A5 - pET22b in LB 20 °C (57.9 kDa), lane 6; CYP3A5 - pET22b in LB 30 °C, lane 7; CYP3A4 - pET22b in LB 20 °C (55.7 kDa), lane 8; CYP3A4 - pET22b in LB 30 °C, lane 9; CYP2D6 - pET22b in LB 20 °C (54 kDa), lane 10; CYP2D6 - pET22b in LB 30 °C (all indicated by an arrow at 55k Da). The CPR enzyme as seen expressed at 77.5 kDa.

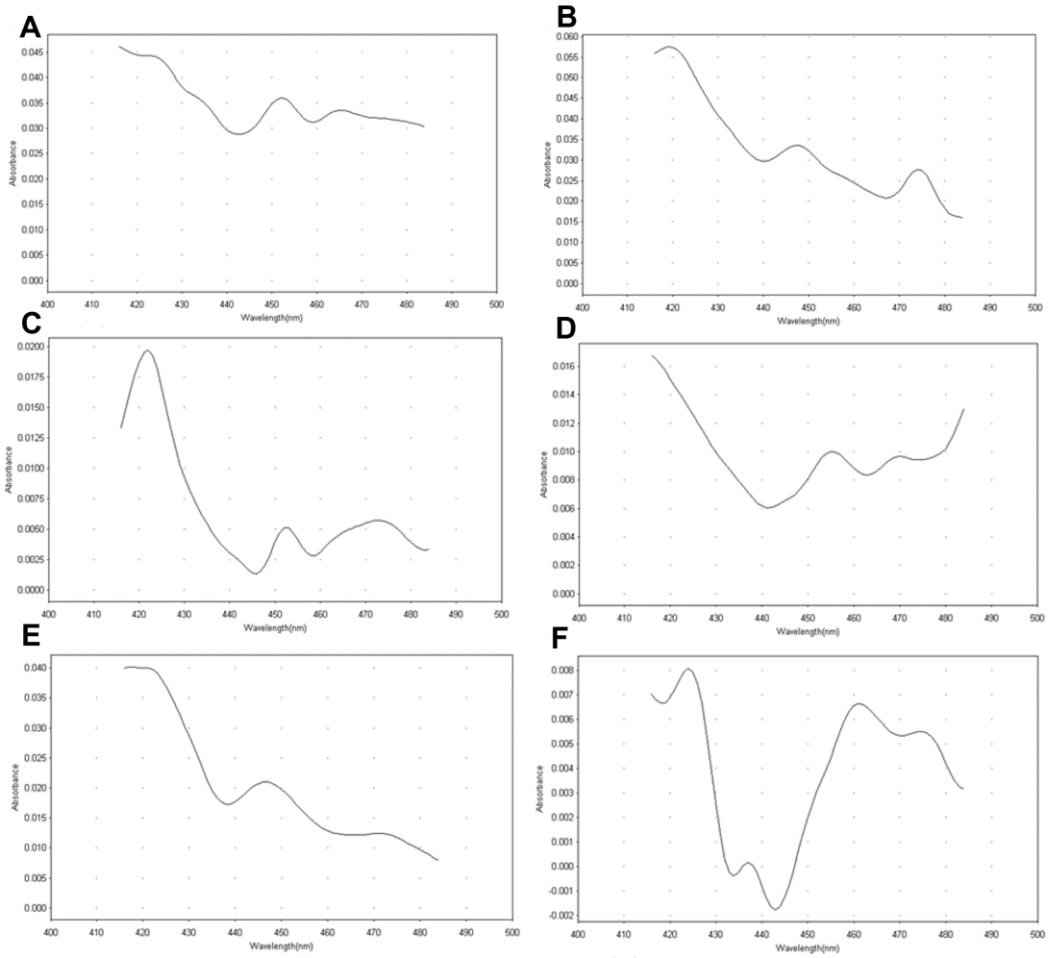


Figure 3-12 CO spectrum (400-500nm) of co-expressed P450s-pET22b and CPR-pCDFDuet-1 in BL21(DE3) and C41(DE3). **A:** CYP2B6-pET22b and CPR-pCDFDuet-1 + pGro7 in BL21(DE3) in LB 20 °C (MS sample number 5), **B:** CYP2B6-pET22b and CPR-pCDF + pGro7 in BL21(DE3) in LB 30 °C, (MS sample number 7), **C:** CYP3A5-pET22b and CPR-pCDFDuet-1 + pGro7 in BL21(DE3) in LB 30 °C (MS sample number 11), **D:** CYP2D6- pET22b and CPR- pCDFDuet-1 without pGro7 in BL21(DE3) in LB 20 °C (MS sample number 9), **E:** CYP2A6 - pET22b and CPR- pCDFDuet-1 + pGro7 in C41(DE3) LB 20 °C (MS sample number 2), and **F:** CYP3A4- pET22b and CPR-pCDFDuet-1 + pGro7 in C41(DE3) LB 20 °C (MS sample number 14) All MS sample numbers are summarised in Table 3-8.

3.3.2. Co-Expression of P450s with and without pGRo7 in Tuner Cells, Induced with 0.25 mM, 0.5 mM and 1mM IPTG, in LB 20 °C and 30 °C.

In this experiment, Tuner cells were used to assess expression in LB 20 °C and 30 °C with and without pGRo7, and cultures were induced with 0.25 mM, 0.5 mM and 1mM IPTG. Co-expression of clones with pGro7 in Tuner cells generated active enzymes for: CYP2B6-pET22b and CPR-pCDFDuet-1 with 0.5 mM IPTG in LB 30 °C, CYP3A4-pET22b and CPR-pCDFDuet-1 with 0.25 mM IPTG in LB 30 °C, CYP2D6-pET22b and CPR-pCDFDuet-1 with 0.25 mM IPTG in LB 20 °C, CYP2D6-pET22b and CPR-pCDFDuet-1 with 0.25 mM IPTG in LB 30 °C (Figure 3-13 **A**, **B**, **C** and **D**), while when expressed without pGro7 for CYP2D6-pET22b and CPR-pCDFDuet-1 with 1 mM IPTG in LB 20 °C, CYP3A5-pET22b and CPR-pCDFDuet-1 with 0.5 mM IPTG in LB 20 °C, CYP3A5-pET22b and CPR-pCDFDuet-1 with 0.5 mM IPTG in LB 30 °C, (Figure 3-14 **A**, **B** and **C**). It can be observed that active enzyme was recovered mostly in 0.25 and 0.5 mM IPTG, while only CYP2D6 was CO-active when induced with 1 mM IPTG, which corresponds to lane 18 of SDS gel (Figure 3-15). IPTG concentration can have an impact on expressed enzyme, as the lower the concentration of IPTG, the slower the rate of the expression, and the more time there is for translation of the enzyme and proper folding. Fastly and improperly folded enzyme can result in inclusion bodies, and subsequently in non-active enzyme. Also, it can be seen that different enzymes had different requirements for temperature and concentrations of IPTG, as well presence and absence of pGro7 to generate active enzyme. No expression at all was achieved for co-expressed CYP2D6 and CPR with 0.5 mM IPTG in LB 30 °C (Figure 3-15, lane 19).

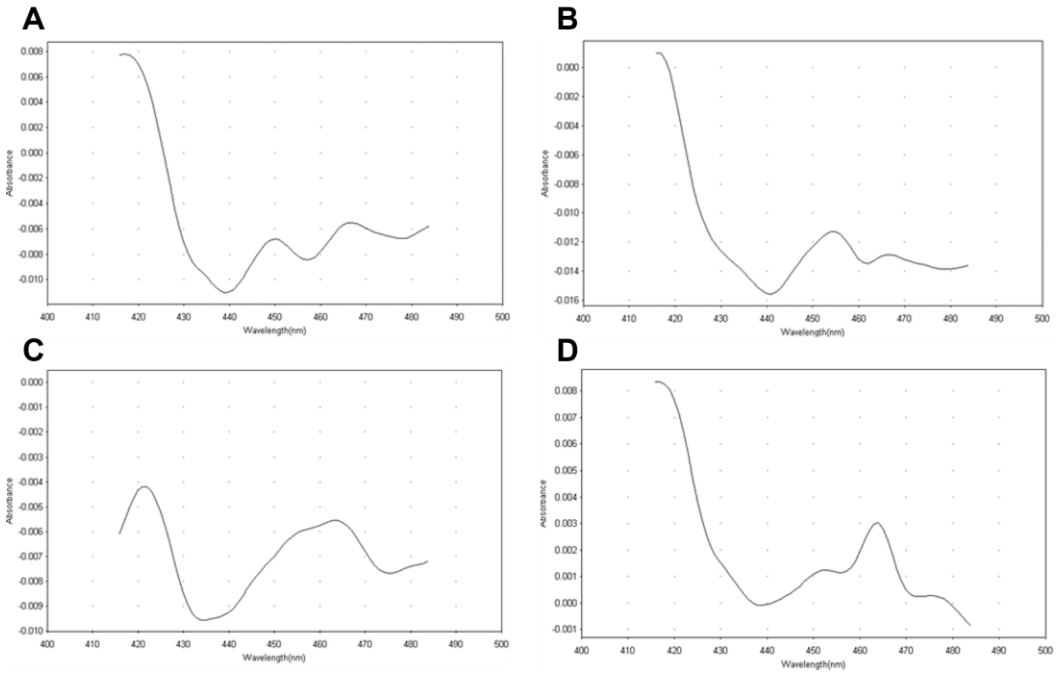


Figure 3-13 CO spectrum (400-500nm) of co-expressed P450s-pET22b and CPR-pCDFDuet-1 +pGro7 in Tuner cells. **A:** CYP2B6 - pET22b and CPR- pCDFDuet-1 + pGro7 induced with 0.5 mM IPTG in LB 30 °C (MS sample number 8), **B:** CYP3A4 - pET22b and CPR- pCDFDuet-1 + pGro7 induced with 0.25 mM IPTG in LB 30 °C (MS sample number 15), **C:** CYP2D6-pET22b and CPR-pCDFDuet-1 + pGro7 induced with 0.25 mM IPTG in LB 20 °C (MS sample number 17), and **D:** CYP2D6-pET22b and CPR-pCDFDuet-1 + pGro7 induced with 0.25 mM IPTG in LB 30 °C (MS sample number 19). All MS sample numbers are summarised in Table 3-8.

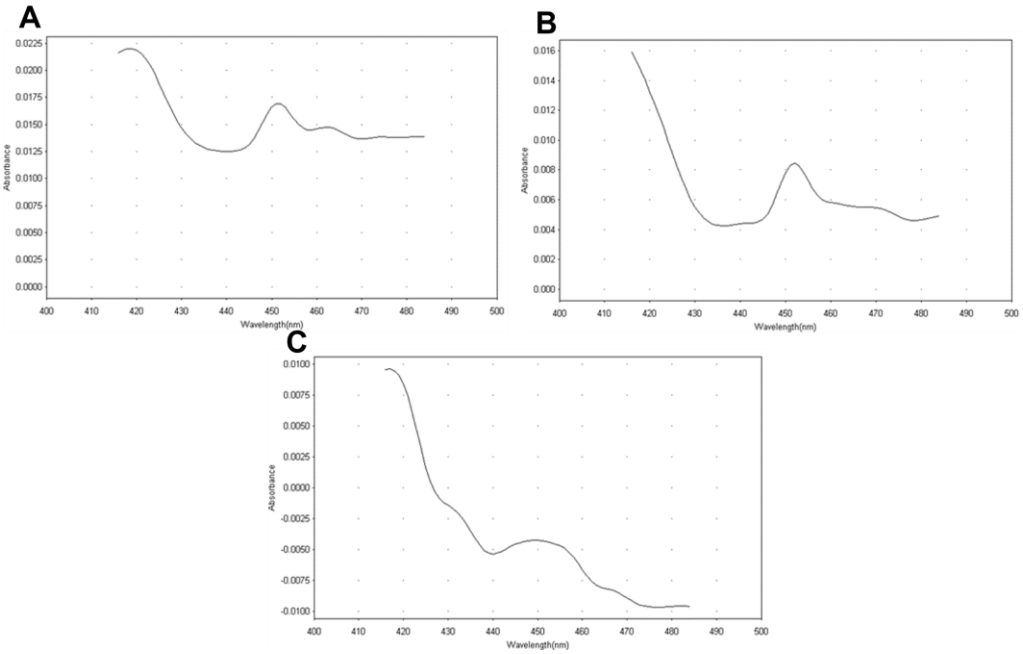


Figure 3-14 CO spectrum (400-500nm) of co-expressed P450s-pET22b and CPR-pCDFDuet-1 without pGro7 in Tuner cells. A: CYP2D6-pET22b and CPR-pCDFDuet-1 induced with 1 mM IPTG in LB 20 °C (MS sample number 18-2), **B:** CYP3A5-pET22b and CPR-pCDFDuet-1 induced with 0.5 mM IPTG in LB 20 °C (MS sample number 9-1) and **C:** CYP3A5-pET22b and CPR-pCDFDuet-1 iduced with 0.5 mM IPTG in LB 30 °C. All MS sample numbers are summarised in Table 3-8.

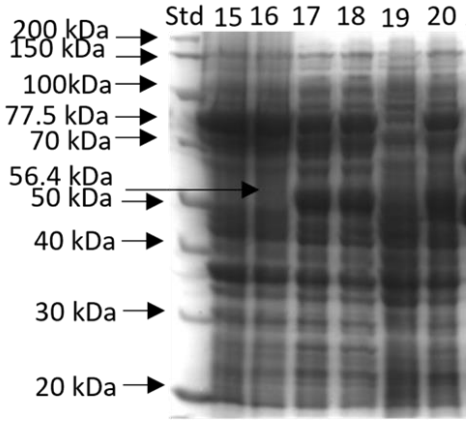


Figure 3-15 Total protein of co-expressed P450s-pET22b and CPR-pCDFDuet-1 without pGro7 in Tuner cells, induced with 0.5 and 1mM IPTG. From left; protein standard (ThermoFisher Scientific), lane 17; CYP2D6 at 56.4 kDa (indicated by a black arrow) induced with 0.5 mM IPTG in LB 20 °C, lane 18; CYP2D6 induced with 1 mM IPTG in LB 20 °C, lane 19; CYP2D6 induced with 0.5 mM IPTG in LB 30 °C, lane 20; CYP2D6 induced with 1 mM IPTG in LB 30°C, while only lanes 17, 18 and 20 the CPR can be observed at 77.5 kDa.

3.4. Co-expression of CYPs and CPR into pET29b and pCDFDuet-1, respectively

In this experiment we intended to determine if co-expression and co-localisation of both P450s and the reductase would generate more active enzyme compared to the expression and localisation of the P450s in the periplasm (as we did with expression in pET22b, Figure 3-10).

So, target P450s were cloned in pET29b, and co-expressed with the CPR cloned in pCDFDuet-1. The same cloning strategy was applied as in Figure 3-10, but with the difference the P450s were cloned in pET29b instead of pET22b. The advantage of pET29b is that the expressed P450s are not inserted into the periplasm; rather, P450 and CPR are positioned in the cell membrane, where they can interact and transfer electrons with one another. Moreover, overexpressed P450s and their need for NADPH can stress the organism and lead to metabolic imbalance. The co-overexpression with CPR is a common strategy to avoid the imbalance (Tang et al., 2010, Pan et al., 2011, Rawal et al., 2010). The primers can be found in the Appendix 1. Briefly, all target P450s were cloned in pET29b with *Nde*I and *Hind*III so that a 6 x His tag was incorporated at the C-terminus of the recombinant protein.

3.4.1. Co-Expression in Tuner cells with 0.25 mM and 0.5 mM IPTG, in LB 20 °C and 30 °C, with and without pGRo7

Expression assays were performed with and without pGRo7 in Tuner cells induced with 0.25 mM and 0.5 mM IPTG, in LB 20 °C and 30 °C. When expressed with pGRo7, active enzyme was recovered for only CYP2A6 with 0.5 mM IPTG in both LB 20 °C and 30 °C (Figure 3-16 **A** and **B**, respectively), whereas, when expressed without pGro7, active enzyme was obtained for CYP2A6 induced with 0.25 mM IPTG in LB 20 °C, and with 0.5 mM IPTG in LB 30 °C, and for CYP2B6 expressed with 0.25 mM IPTG in LB 30 °C (Figure 3-17 **A**, **B** and **C**, respectively). WCB transformations were performed for enzymes producing a positive CO spectrum and products were analysed by mass spec. Coumarin was used as positive control substrate for CYP2A6 and ketamine for CYP2B6.

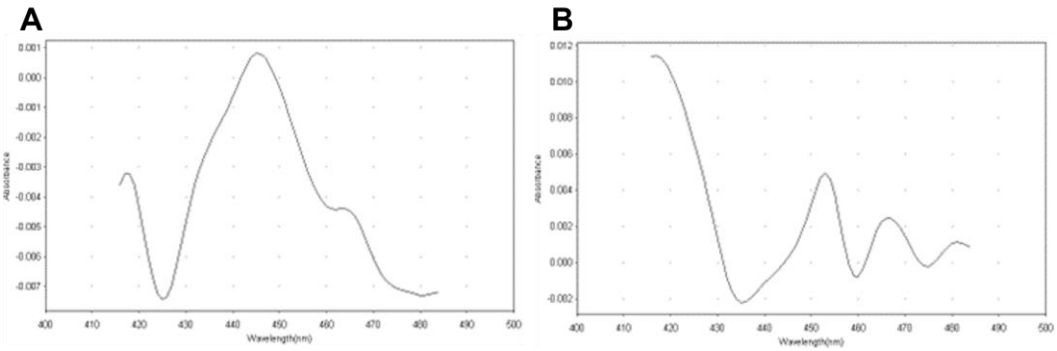


Figure 3-16 CO spectrum (400-500nm) of co-expressed P450s-pET29b and CPR-pCDFDuet-1 + pGro7 in Tuner cells. A: CYP2A6- pET29b and CPR- pCDFDuet-1 + pGro7 induced with 0.5 mM IPTG in LB 20 °C (MS sample number 3), **B:**CYP2A6-pET29b and CPR-pCDFDuet-1 + pGro7 induced with 0.5 mM IPTG in LB 30 °C (MS sample number 4). All MS sample numbers are summarised in Table 3-8.

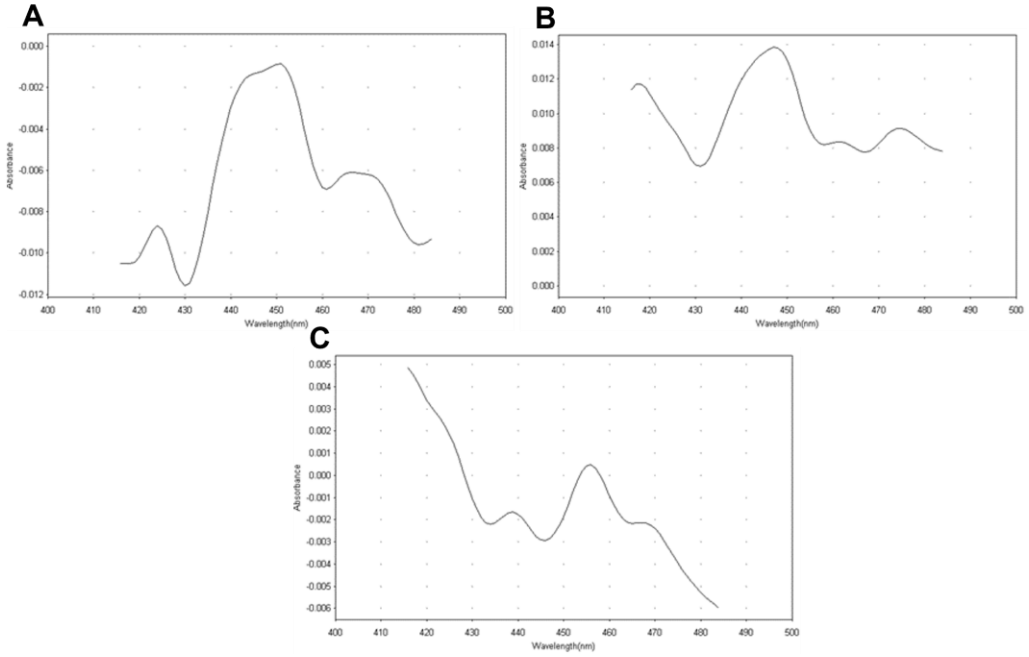


Figure 3-17 CO spectrum (400-500nm) of co-expressed P450s-pET29b and CPR-pCDFDuet-1 without pGro7 in Tuner cells. A: CYP2A6- pET29b and CPR-pCDFDuet-1 without pGro7 induced with 0.25 mM IPTG in LB 20 °C (MS sample number 1), B: CYP2A6-pET29b and CPR-pCDFDuet-1 without pGRo7 induced with 0.5 mM IPTG in LB 30 °C (MS sample number 4), and C: CYP2B6-pET29b and CPR-pCDFDuet-1 induced with 0.25 mM IPTG in LB 30 °C (MS sample number 6). All MS sample numbers are summarised in Table 3-8.

3.5. Cloning and Expression of Metagenome Genes with the N-terminus (MALLLAVFL) modification

Through mining the metagenome library available through Prozomix Ltd, two metagenome genes were identified, 198 and 914, that shared 20% and 24% homology, respectively, with the human gene CYP3A5. According to structure prediction, both encoded proteins are P450s, and therefore hypothesised to metabolise parent drug ketamine to hydroxyketamine, and potentially other metabolites such as norketamine to hydroxynorketamine. If the hypothesis is proved, that would abolish the use the need of human CYPs, which are difficult to express actively, and work with, due to external requirements for cofactors, and benefit the pharmaceutical community in the enzymatic production of hydroxyketamine and hydroxynorketamine.

Both metagenome genes were cloned into three different plasmids, according to their properties, and expressed in different *E. coli* strains to determine the best expression and catalytic activity. While pET29b and pCWOri^{+} are high and low copy number plasmids, respectively, pGEX-6P-1 allows the addition of a GST tag instead of the conventional His tag. Sometimes the His tag can confer changes to protein structures, as it can interact with the active site and the haem group directly or indirectly (Lin et al., 2011).

An *N*-terminus modification of the metagenome genes encoding proteins according to the modification applied to the human CYP3A5. Specifically, a positively charged amino acid sequence, previously known from Barnes *et al.* (1991), was incorporated. Adding a hydrophilic stretch can lead to soluble protein expression located in the cytoplasm (Barnes *et al.*, 1991). Both genes were expressed initially in BL21(DE3) in LB and AIM at 20 °C and 30 °C, with and without pGro7. Then another experiment was performed in Tuner cells.

3.5.1. Expression of Metagenome Genes with and without pGRo7, in BL21(DE3), in LB 20 °C and 30 °C.

The metagenome genes were cloned into different plasmids and expressed in different media in *E. coli* strains BL21(DE3) and Tuner cells. Each of the constructs 198-pET29b, 198-pCWOri^{+}, 198-pGEX-6P-1, and 914-pET29b, 914-pGEX-6P-1 were individually transformed into *E. coli* BL21(DE3) strain with and without pGro7 plasmid-expressing chaperones GroEL and GroES. The cultures were cultivated at 20 °C and 30 °C overnight for LB, and for 48 h for AIM, with rotary shaking at 150 rpm (Table 3-5 and Table 3-6), in order to determine best possible expression. Good expression was achieved for 198-pET29b in LB 20 °C, 198-pET29b in AIM 20 °C and 198-pET29b in AIM 30 °C, all with pGro7, and 914-pGEX-6P-1 in LB 20 °C and 30 °C. In the SDS gel below (Figure 3-18 **A**), a distinct band of protein expression can be seen running horizontally all along the gel, which corresponds to the chaperone expression GroEL at 60 kDa, whereas in lane 3 can be seen another distinct band at approximately 50 kDa, which corresponds to protein 198-pET29b, in LB 20 °C (52.3 kDa). In Figure 3-18 **B**, lane 1 can be seen at approximately 50 kDa again the expressed 198-pET29b, while in lane 2 can be observed a protein band at 144.4 kDa, which corresponds to protein 914-pGEX-6P-1 in LB 20 °C. It was determined that the best cloning plasmid and expression conditions that generated active enzyme were constructs 198-pET29b in LB 20 °C and 914-pGEX-6P-1 in LB 20 °C, so all further assays were carried out with only these two constructs. CO spectra were performed for all expressing conditions. Different version of active 914-pGEX-6P-1 enzyme in the same conditions can be observed in the chromatograms below (Figure 3-19 **A**, **B** and **C)**. WCB were performed with active enzymes and all the samples were analysed by mass spec.

Table 3-5 Metagenome gene 198 constructs generated and expression conditions in *E. coli* BL21(DE3).

		Temperature in ºC				
198 BL21(DE3)	Cloning plasmids	20		30		Media
1 mM IPTG	pET29b	ϰ	√	ϰ	ϰ	LB
		ϰ	√	ϰ	√	AIM
1 mM IPTG	pCWOri+	ϰ	ϰ	ϰ	ϰ	LB
		ϰ	ϰ	ϰ	ϰ	AIM
1 mM IPTG	pGEX-6P-1	ϰ	ϰ	ϰ	ϰ	LB
		ϰ	ϰ	ϰ	ϰ	AIM
		no pGro7	pGro7	no pGro7	pGro7	
		Presence of chaperone Plasmid				

Table 3-6 Metagenome gene 914 constructs generated and expression conditions in *E. coli* BL21(DE3).

		Temperature in ºC				
914 BL21(DE3)	Cloning plasmids	20		30		Media
1mM IPTG	pET29b	ϰ	ϰ	ϰ	ϰ	LB
		ϰ	ϰ	ϰ	ϰ	AIM
1mM IPTG	pGEX-6P-1	ϰ	√	ϰ	√	LB
		ϰ	ϰ	ϰ	ϰ	AIM
		no pGro7	pGro7	no pGro7	pGro7	
		Presence of chaperone Plasmid				

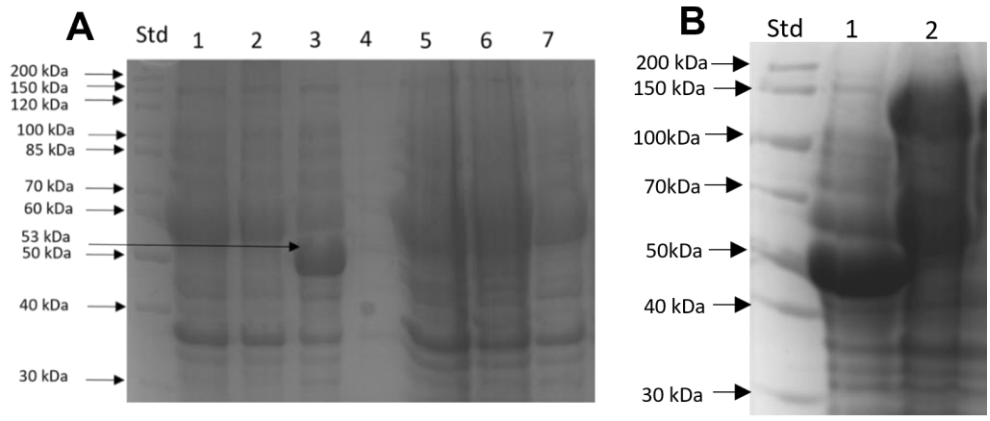


Figure 3-18 Total protein of metagenome 198 and 914 proteins co-expressed with pGro7 in BL21(DE3). A: from left; protein standard (ThermoFisher Scientific), lane 1; 198-pCWOri$^+$ in LB 20 °C, lane 2; 198- pCWOri$^+$ in LB 30 °C, lane 3; 198-pET29b in LB 20 °C, lane 4; 198-pET29b in LB 30 °C, lane 5; 198-pGEX-6P-1 in LB 20 °C, lane 6; 198-pGEX-6P-1 in LB 30 °C, lane 7; pET29b in LB 20 °C (as a control). The molecular weights of 198 constructs are as: 198-pCWOri$^+$ at 52.3 kDa, 198-pET29b at 52.3 kDa, 198- pGEX-6P-1 at 75.2 kDa, pGro7 GroEL chaperone at 60 kDa. **B:** lane 1; expressed 198-pET29b with pGro7 in LB 20°C at 52.3 kDa size band, lane 2; 914-pGEX-6P-1 with pGro7 in LB 20°C at 144.4 kDa size band.

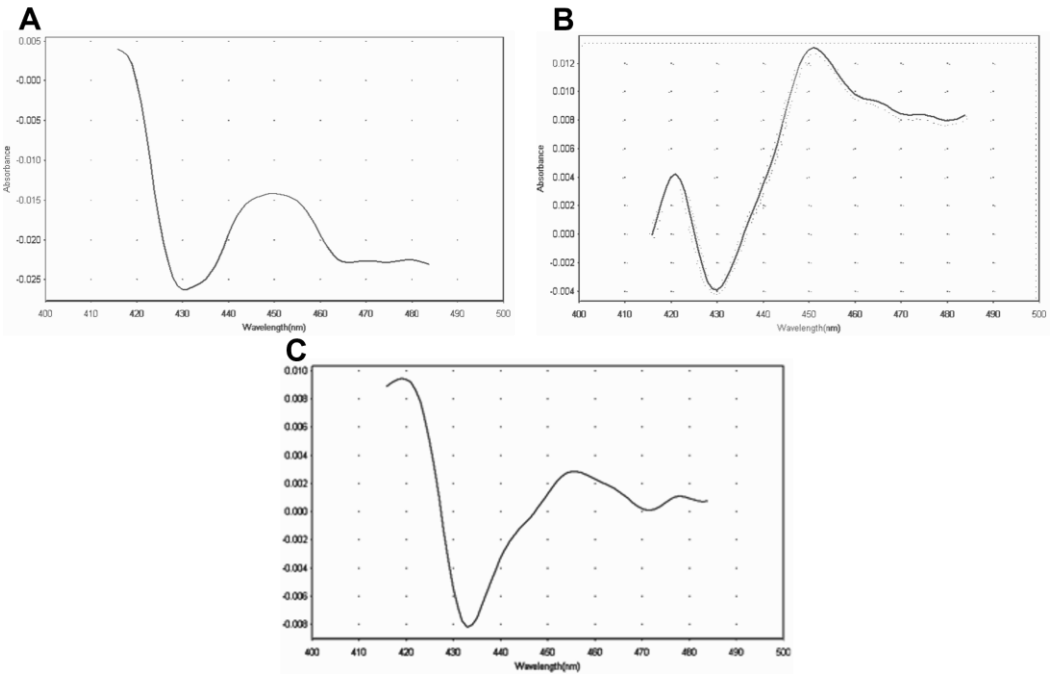


Figure 3-19 CO spectrum (400-500 nm) of metagenome 198 and 914 proteins co-expressed with pGro7 in BL21(DE3). A: 914-pGEX-6P-1 in LB 20 °C (MS sample number 4), **B:**914-pGEX-6P-1 in LB 20 °C, **C:**914-pGEX-6P-1in BL21(DE3) LB 20 °C. All MS sample numbers are summarised in Table 3-8.

3.5.2. Expression of Metagenome Proteins with and without pGro7 in Tuner cells, induced with 0.25 mM, 0.5 mM, and 1 mM IPTG, in LB 20 °C and 30 °C.

Further expression assays were performed with and without pGro7 in Tuner cells, induced with 0.25 mM, 0.5 mM, and 1 mM IPTG, in LB 20 °C and 30 °C. Good expression level was achieved for both metagenome genes encoding proteins 198 and 914 without pGro7, when induced with 0.25 mM and 0.5 mM IPTG. In Figure 3-20 **A,** at 52.8 kDa (lanes 21- 23, and Figure 3-19 **B**, lane 24) can be observed metagenome gene encoding protein 198, and in Figure 3-20 **B** at 144.8 kDa (lanes 25 to 28) can be observed metagenome gene encoding protein 914.

Similar expression level was achieved for both metagenome genes encoding proteins when co-expressed with pGro7. In Figure 3-21 (lanes 21 to 24) can be observed again at 52. 8 kDa metagenome gene 198 encoding protein (indicated with an arrow), at 60 kDa the GroES chaperone (lanes 21 to 28), and at 144.8 kDa (lanes 25 to 28) metagenome gene encoding protein 914.

CO-active enzyme was recovered from metagenome gene 198-pET29b without pGro7, when induced with 0.25 mM IPTG in LB 30 °C (Figure 3-22 **A**), and with pGro7, when induced with 0.5 mM IPTG in LB 30 °C (Figure 3-22 **B**), and when induced with 1 mM IPTG LB 20 °C (Figure 3-22 **C**). Higher catalytic activity was achieved when expressed with pGro7 in LB 30 °C induced with 0.5 mM IPTG, as seen from the CO spectra chromatograms (HPLC analysis not successful).

On the other hand, metagenome gene 914- pGEX-6P-1 expressed without pGro7 when induced with 0.5 mM, and 1 mM IPTG in LB 20 °C (Figure 3-23 **A** and **B**), and with pGro7 when induced with 0.5 mM, and 1 mM IPTG in LB 20 °C (Figure 3-23 **C** and **D**). Higher catalytic activities were achieved when co-expressed with pGro7, as seen from the CO chromatograms. All the expressed 914 enzymes, and only 198 induced with 0.25 mM IPTG in LB 30 °C without pGro7 were used in WCB and analysed using mass spec.

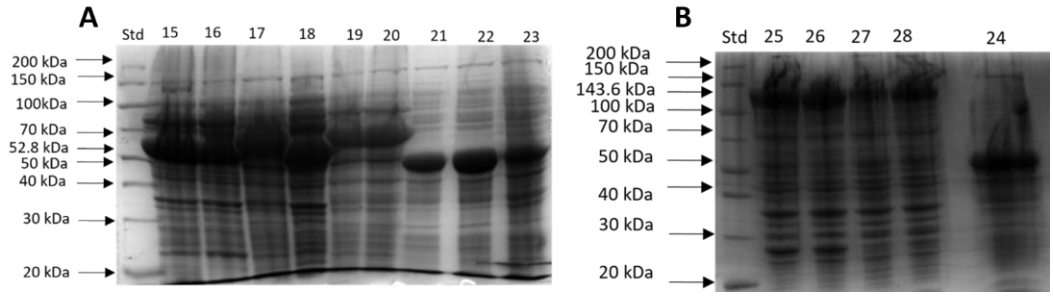


Figure 3-20 Total protein of metagenome proteins 198 (52.8 kDa) and 914 (144.8 kDa) expressed without pGRo7 in Tuner cells, induced with 0.25 mM and 0.5 mM IPTG. From left; protein standard (ThermoFisher Scientific), lane 21; 198-pET29b in LB 20 °C with 0.25 mM IPTG, lane 22; 198-pET29b, in LB 20°C with 0.5 mM IPTG, lane 23; 198-pET29b in LB 30°C with 0.25 mM IPTG, lane 24; 198-pET29b in LB 30°C with 0.5 mM IPTG, lane 25; 914-pGEX in LB 20°C with 0.25 mM IPTG, lane 26; 914-pGEX in LB 20°C with 0.5 mM IPTG, lane 27; 914-pGEX in LB 30 °C with 0.25 mM IPTG, lane 28; 914-pGEX in LB 30 °C with 0.5 mM IPTG.

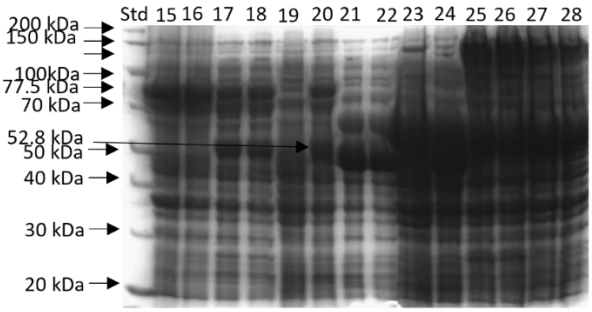


Figure 3-21 Total protein of metagenome genes encoding proteins 198 (52.8 kDa) and 914 (144.8 kDa) + pGRo7 in Tuner cells, induced with 0.5 mM and 1 mM IPTG. From left; protein standard (ThermoFisher Scientific), lane 21; 198-pET29b in LB 20°C induced with 0.5 mM IPTG, lane 22; 198-pET29b in LB 20°C induced with 1 mM IPTG, lane 23; 198-pET29b in LB 30°C induced with 0.5 mM IPTG, lane 24; 198-pET29b in LB 20°C induced with 1 mM IPTG, lane 25; 914-pGEX- pGRo7 in LB 20°C induced with 0.5 mM IPTG, lane 26; 914-pGEX in LB 20°C induced with 1 mM IPTG, lane 27; 914-pGEX in LB 30 °C induced with 0.5 mM IPTG, lane 28; 914-pGEX in LB 30 °C induced with 1 mM IPTG.

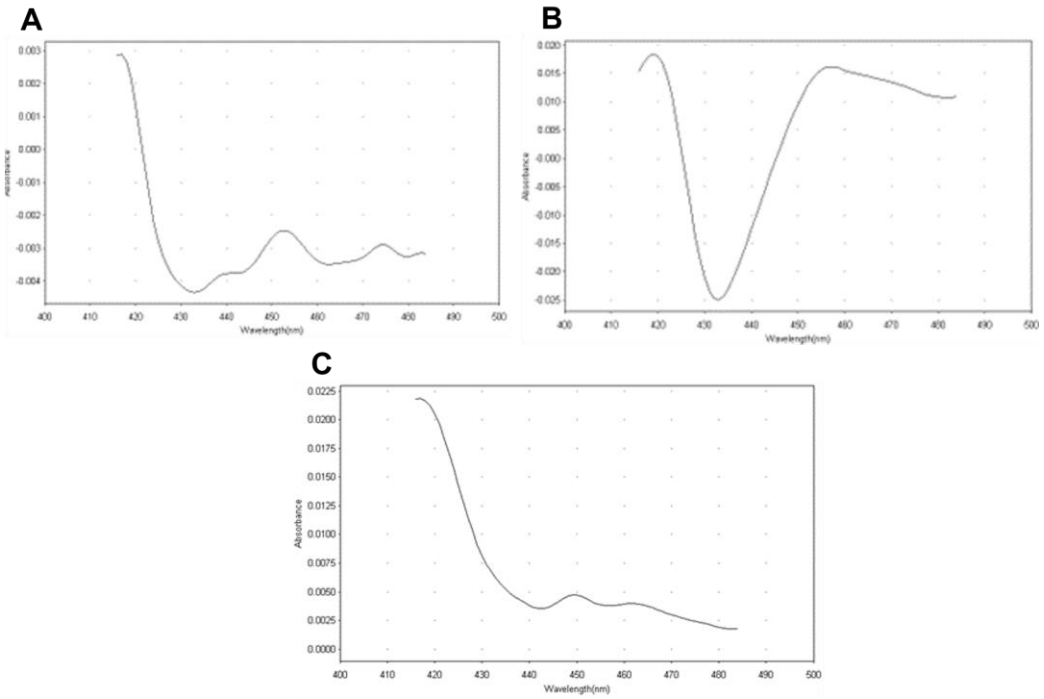


Figure 3-22 CO spectrum (400-500 nm) of recombinantly expressed metagenome protein 198 in tuner cells. A: 198-pET29b without pGro7 in LB 30 °C, induced with 0.25 mM IPTG (MS sample number 21), **B:** 198-pET29b +pGRo7 in LB 30 °C, induced with 0.5 mM IPTG, and **C:** 198-pET29b +pGro7 in LB 20 °C induced with 1 mM IPTG. All MS sample numbers are summarised in Table 3-8.

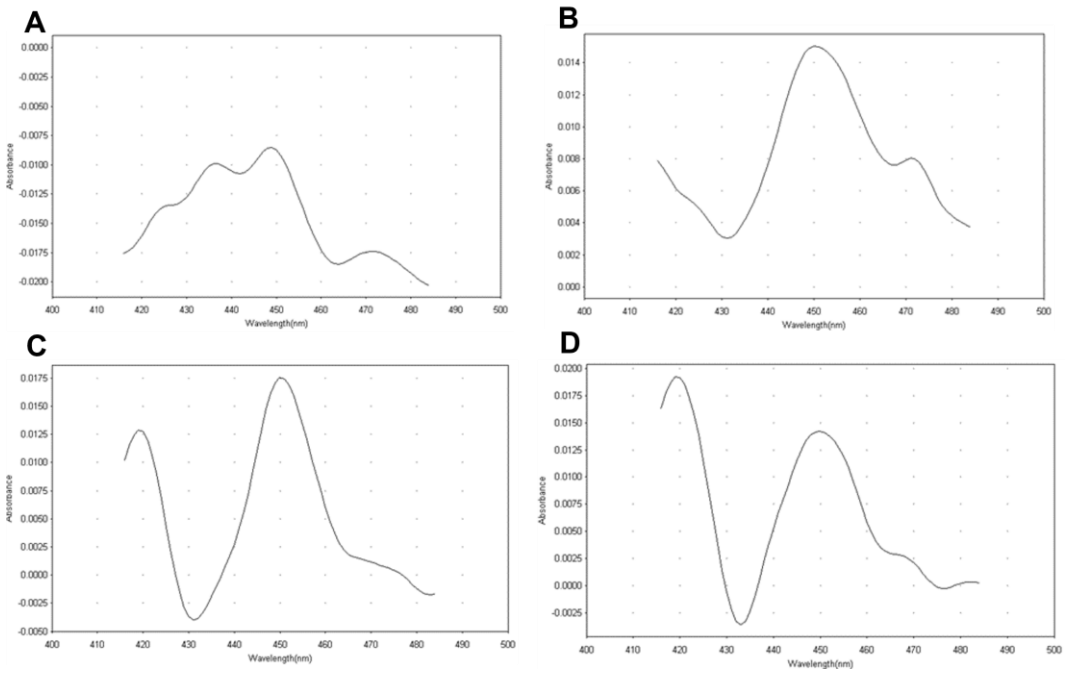


Figure 3-23 CO spectrum (400-500 nm) of recombinantly expressed metagenome protein 914 in tuner cells, in LB 20 °C. A: 914-pGEX-6P-1 without pGro7, induced with 0.5 mM IPTG (MS sample number 25), **B:** 914-pGEX-6P-1 without pGro7, induced with 1 mM IPTG (MS sample number 26), **C:** 914-pGEX-6P-1+ pGro7, induced with 0.5 mM IPTG (MS sample number 25-3), **D:** 914-pGEX-6P-1+ pGro7, induced with 1 mM IPTG (MS sample number 26-4). All MS sample numbers are summarised in Table 3-8.

3.6. Cloning and expression of Metagenome genes without the N-terminus (MALLLAVF) modification

The non-modified *N*-terminus metagenome genes were cloned; 198 in pET29b, and 914 in pGEX-6P-1. This strategy was employed to determine if the native P450 metagenome genes without the modification MALLLAVF can generate higher expression and/ or soluble enzymes. Expression was assayed only in Tuner cells with and without pGRo7, in LB 20 ℃ and 30 ℃ induced with 0.25 mM and 0.5 mM IPTG. Tuner cells have proved so far to generate comparatively the same level of expressed enzyme, but higher level of catalytic activities of the expressed proteins.

The same level of expression was achieved for both 198 (Figure 3-24, lanes 19 to 23, at 51.6 kDa) and 914 enzymes expressed with pGro7 (Figure 3-24, lanes 25 to 28, at 143.7 kDa), and without pGro7 (data not shown).

Enzyme 914 was CO-active when co-expressed with pGro7 in LB 20 ℃ and 30 ℃, and induced with 0.25 mM, and 0.5 mM IPTG (all conditions when co-expressed with pGro7, Figure 3-26 **A**, **B**, **C** and **D**), whereas, when expressed without pGro7, active enzyme was only achieved in LB 20 ℃ when induced with 0.25 mM IPTG (Figure 3-26 **E**). Highest catalytically active enzyme according to CO was recovered when co-expressed with pGro7 and induced with 0.5 mM IPTG in LB 20 ℃ and LB ℃ (Figure 3-26 **C** and **D**). Additionally, 914 was detected in the soluble fraction when co-expressed with pGro7 in LB 20 ℃ and induced with 0.25 mM and 0.5 mM IPTG (Figure 3-24, lanes 25 to 27), and without pGro7 (Figure 3-25, lanes 25 and 27, at 143.7 kDa). From the solubly expressed 914 enzyme, catalytical activity through CO exhibited only when expressed without pGro7 in LB 20 ℃ and induced with 0.25 mM IPTG (Figure 3-26 **E**). WCB transformations and mass spec analysis was performed to all active enzyme for the detection of catalytic activity.

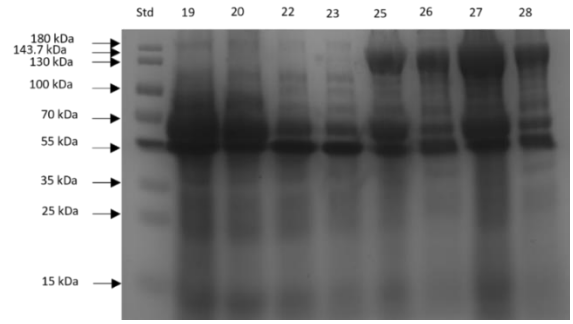


Figure 3-24 Total protein of metagenome enzymes 198-pET29b (51.6 kDa) and 914-pGEX-6P-1 (143.7 kDa) co-expressed with pGro7 in Tuner cells, induced with 0.25 mM and 0.5 mM IPTG. From left; protein standard (ThermoFisher Scientific), lane 22; 198-pET29b in LB 30 °C, induced with 0.5 mM IPTG, lane 23; 198-pET29b in LB 20°C, induced with 0.25 mM IPTG, lane 25; 914-pGEX in LB 20°C, induced with 0.25 mM IPTG, lane 26; 914-pGEX in LB 30°C, induced with 0.25 mM IPTG, lane 27; 914-pGEX in LB 20 °C, induced with 0.5 mM IPTG, lane 28; 914-pGEX in LB 30 °C, induced with 0.5 mM IPTG.

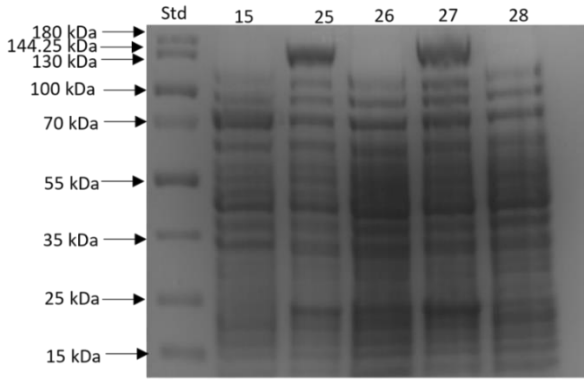


Figure 3-25 CFE of metagenome enzyme 914-pGEX-6P-1 (143.7 kDa) expressed without pGro7 in Tuner cells, induced with 0.25 mM and 0.5 mM IPTG. From left; protein standard (ThermoFisher Scientific), lane 25; 914-pGEX in LB 20°C induced with 0.25 mM IPTG, lane 26; 914-pGEX in LB 30°C, induced with 0.25 mM IPTG, lane 27; 914-pGEX in LB 20 °C, induced with 0.5 mM IPTG, lane 28; 914-pGEX in LB 30 °C, induced with 0.5 mM IPTG.

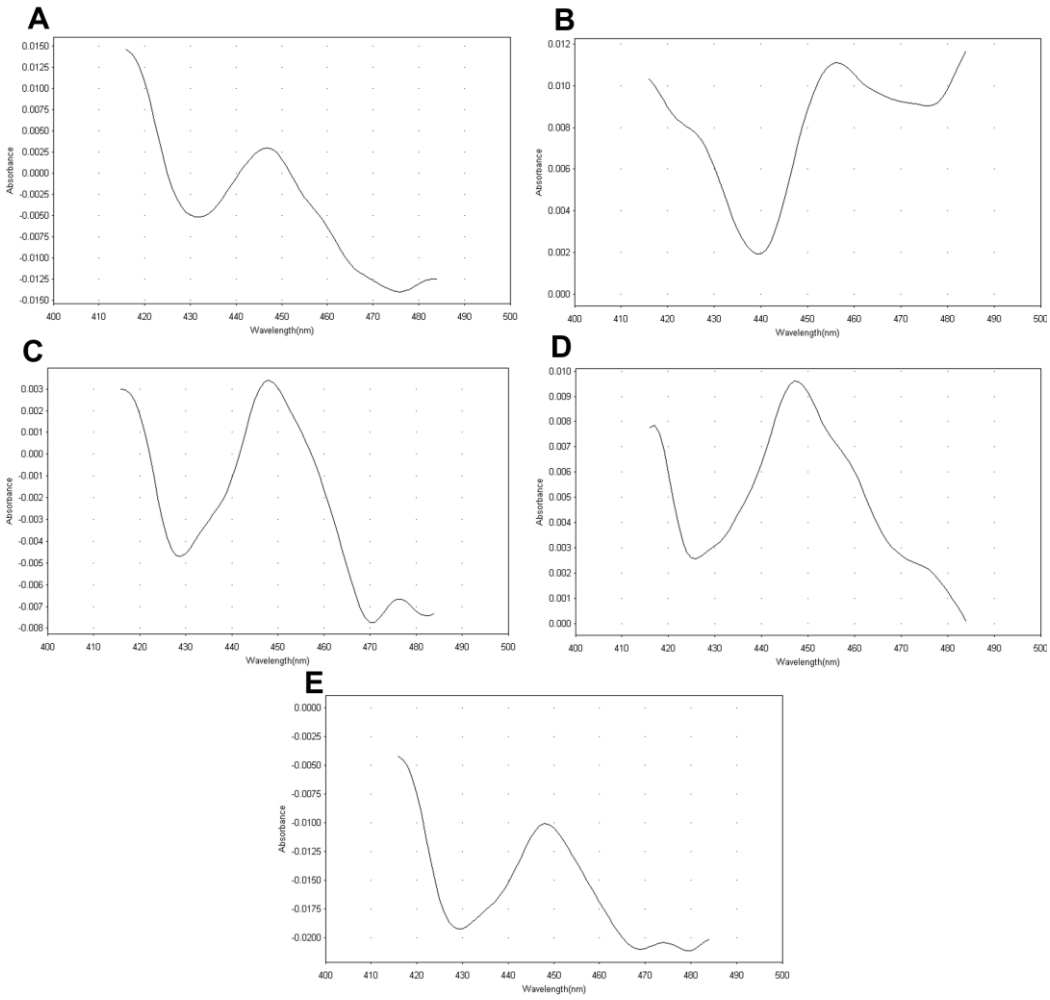


Figure 3-26 CO spectrum (400-500 nm) of metagenome enzyme 914-pGEX-6P-1 (no-MALLLAVF) expressed in Tuner cells. **A:** 914 in LB 20 °C induced with 0.25 mM IPTG (MS sample number 25), **B:** 914+pGro7 in LB 30 °C, induced with 0.25 mM IPTG (MS sample number 26), **C:** 914+pGro7 in LB 20 °C, induced with 0.5 mM IPTG (MS sample number 27), **D:** 914+pGro7 in LB 30 °C, induced with 0.5 mM IPTG (MS sample number 28), E: 914 without pGro7 in LB 20 °C, induced with 0.25 mM IPTG (MS sample 24). All MS sample numbers are summarised in Table 3-8.

3.7. Metabolomic Analysis of WCB with HPLC

All expressed enzymes were subjected to CO assays to determine the activity status. All cells expressing active enzymes were used in whole cell biotransformations with the enzyme resuspended in 1 mL 0.05 mM potassium phosphate buffer containing 5 % ACN, 4 % glycerol and 0.05 mg/mL final substrate concentration, which is equivalent of 1.823 mM for norketamine, and the reactions were initiated with the addition of 10 mg/mL reducing agent NADPH. Briefly, the reactions were proceeded for 24 hours incubated at 37 °*C*. The volume of 100 µL samples were taken at set intervals of 0 hours (negative control sample), 15 minutes, 3 hours and 24 hours. All the samples were spun down, supernatants transferred to a fresh microcentrifuge tube, then boiled for 10 minutes to stop the reactions, and once again were spun down for 10 minutes and supernatants were transferred to fresh microcentrifuge tube. The samples were kept at -80 degrees Celsius until they could be tested.

All the analysis was performed using HPLC, but no activity was detected. While the substrates sampled alone on the HPLC were detected on the chromatograms by their retention times, no product was detected. Additionally, positive control assays were performed with probe substrates and active enzymes, and while the enzymes were confirmed to be active against CO, no substrate depletion, or product conversion was observed for the probe substrates either. We made the decision to lower the substrate concentration from 0.05 mg/mL to 10 µM, because it was too high, and employ LC-MS to analyse our WCB assays.

3.8. Metabolomic Analysis of WCB with Liquid Chromatography Mass Spectrometry (LC-MS)

Based on the above facts, we changed the concentration of the substrate from 0.05 mg/mL to 10 µM final concentration, because it was too high, and repeated all WCB (section 2.5.3.1). All the timepoint samples were then analysed using LC-MS, as a greater sensitivity was required with the lower concentrations of substrate used.

High resolution mass measurement of ions of substrates in MS experiments were accomplished using a final 10 µM of each substrate in the same conditions / buffers as the assays (section 2.1.8). High-resolution Extracted Ion Chromatograms (EICs) of coumarin, ketamine and norketamine were obtained by processing the full-scan MS data set using potential hydroxyketamine, hydroxycoumarin and hydroxynorketamine metabolite ions with 10 ppm mass tolerance. The calculated masses for all the metabolites used in LC-MS analysis are presented in Table 3-7.All the expressed enzymes generated from the different cloning and expression strategies, were used in WCB and analysed by MS are organized below (Table 3-8).

Table 3-7 The calculated masses of all the metabolites used in LC-MS analysis.

Chemical	Formula	Mass (M)	Protonated adduct (M+H)
Ketamine	C13H16ClNO	237.09204	238.09932
5-hydroxyketamine	C13H16ClNO2	253.086957	254.094237
Norketamine	C12H14ClNO	223.07639	224.08367
Hydroxynorketamine	C12H14ClNO2	239.07137	240.07865
Coumarin	C9H6O3	146.03678	147.04406
4-Hydroxycoumarin	C9H6O2	162.03169	163.03897

3.9. Metabolism of Norketamine to Hydroxynorketamine by CYP2B6 and CYP3A4

As mentioned previously, all CO active enzymes were used in WCB, where samples were drawn at set intervals of 0 hours, 15 min, 3 hours and 24 hours (T0, T15, T3 and T24), and were analysed with mass spec. Only two human enzymes hydroxylated norketamine to hydroxynorketamine. CYP2B6-pET29b co-expressed with CPR-pCDFDuet-1 + pGro7 chaperones in Tuner cells in LB 30 °C, and induced with 0.25 mM IPTG, and CYP3A4-pET22b co-expressed with CPR-pCDFDuet-1 + pGro7 in Tuner cells in LB 30 °C, when induced with 0.25 mM IPTG. Both enzymes metabolise norketamine at timepoint 24 hours (T24).

Substrate standards were run on a mass spectrophotometer to find their retention times in order to confirm the applicability of the assay. In Figure 3-27 is presented norketamine standard; top panel is the Total Ion Chromatogram (TIC) and norketamine peak with retention time 5.25, middle panel is shown the SIM channel of mass (M) norketamine (red peak) with retention time 5.25 and abundance 8.17 $x10^7$ (seen on the right hand side of the panel) , and bottom panel is the high resolution mass spectrum of norketamine with exact m/z 224.08389. The calculated mass of norketamine taking into consideration the protonated adduct (M+H) is 224.08367.

Additionally, all the substrates were exploited under test conditions without the enzyme over a period of 24 hours to observe any degradations. Figure 3-28 shows top panel norketamine standard (black peak), 2^{nd} panel norketamine at T0 (red peak), 3^{rd} panel norketamine at T3 (green peak), 4^{th} panel norketamine at T15 (blue peak), and bottom panel is norketamine at T24 (light green peak).

Figure 3-29, Figure 3-30, Figure 3-31, Figure 3-32 and Figure 3-33 show all the chromatograms from the LC-MS analysis of the CYP2B6 whole cell activity against norketamine over timepoints T0, T15, T3 and T24.

At T0 (Figure 3-29) on top panel can be seen the TIC and norketamine peak with retention time 4.53 (black peak), in the second panel is the SIM channel of mass norketamine (negative control for CYP2B6-pET29b, indicated with a red arrow) with retention time 4.49 an abundance of $5.87x10^7$, and bottom panel is the high resolution mass spectrum of norketamine with an exact m/z ratio of 224.08437, while third panel is empty (expected hydroxynorketamine peak).

Similarly, at T15 and T3 (Figure 3-30 and Figure 3-31) the TIC and norketamine peak can once more be seen in the top panel with retention times of 4.52 and 4.54 (black peak), respectively. In the second panel, the SIM channel of mass norketamine can be seen with peaks at 4.49 and 4.50 (indicated with a red arrow), respectively,

and their abundances of 5.53 x 10^7 and 5.42 x 10^7, respectively. The high-resolution mass spectrum of norketamine is shown in the bottom panel with exact m/z ratio of 224.08427 and 224.08440, respectively.

In contrary, at T24 (Figure 3-32) can be seen the TIC in the top panel, SIM channel of mass norketamine with retention time 4.75 and abundance 5.08 x10^7 in the second panel, but in the third panel there is a peak with retention time 5.13 and abundance and 9.33 x10^3, which when selected (Figure 3-33, red arrow), can be seen in the bottom panel its high-resolution mass spectrum with an exact m/z of 240.07748, which is within 10 ppm of the estimated hydroxynorketamine m/z 240.07865. This indicates that at T24 CYP2B6 converted norketamine to hydroxynorketamine, although in low abundance. As can be seen, the hydroxynorketamine peak is relatively low, and the signal is weak. As stated in the methods, the samples were prepared for LC-MS analysis without the use of a pre-concentration step, and they were injected neat. Nonetheless, from T0 to T24 the abundance of norketamine decreases from 5.87 x 10^7 to 5.08 x10^7, which is 13.46 % of substrate conversion to product.

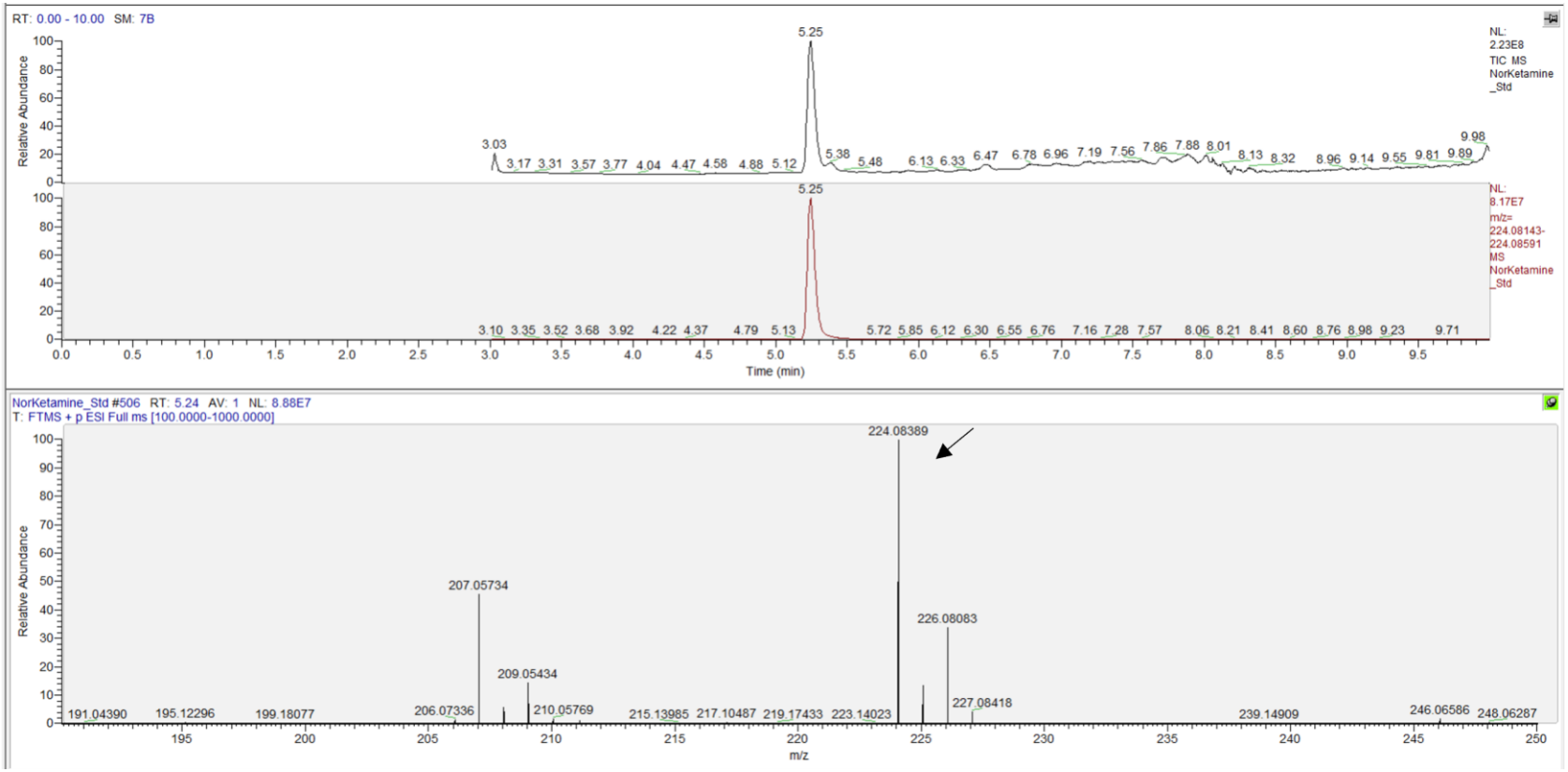


Figure 3-27 LC- MS Chromatogram of norketamine standard. Top panel is the Total Ion Chromatogram (TIC) of norketamine, middle panel is the Single Ion Monitoring (SIM) channel of mass norketamine with retention time 5.25 and abundance 8.17 x 107 (shown on right of the panel), and bottom panel is the high-resolution mass spectrum of norketamine with the exact m/z of 224.08389.

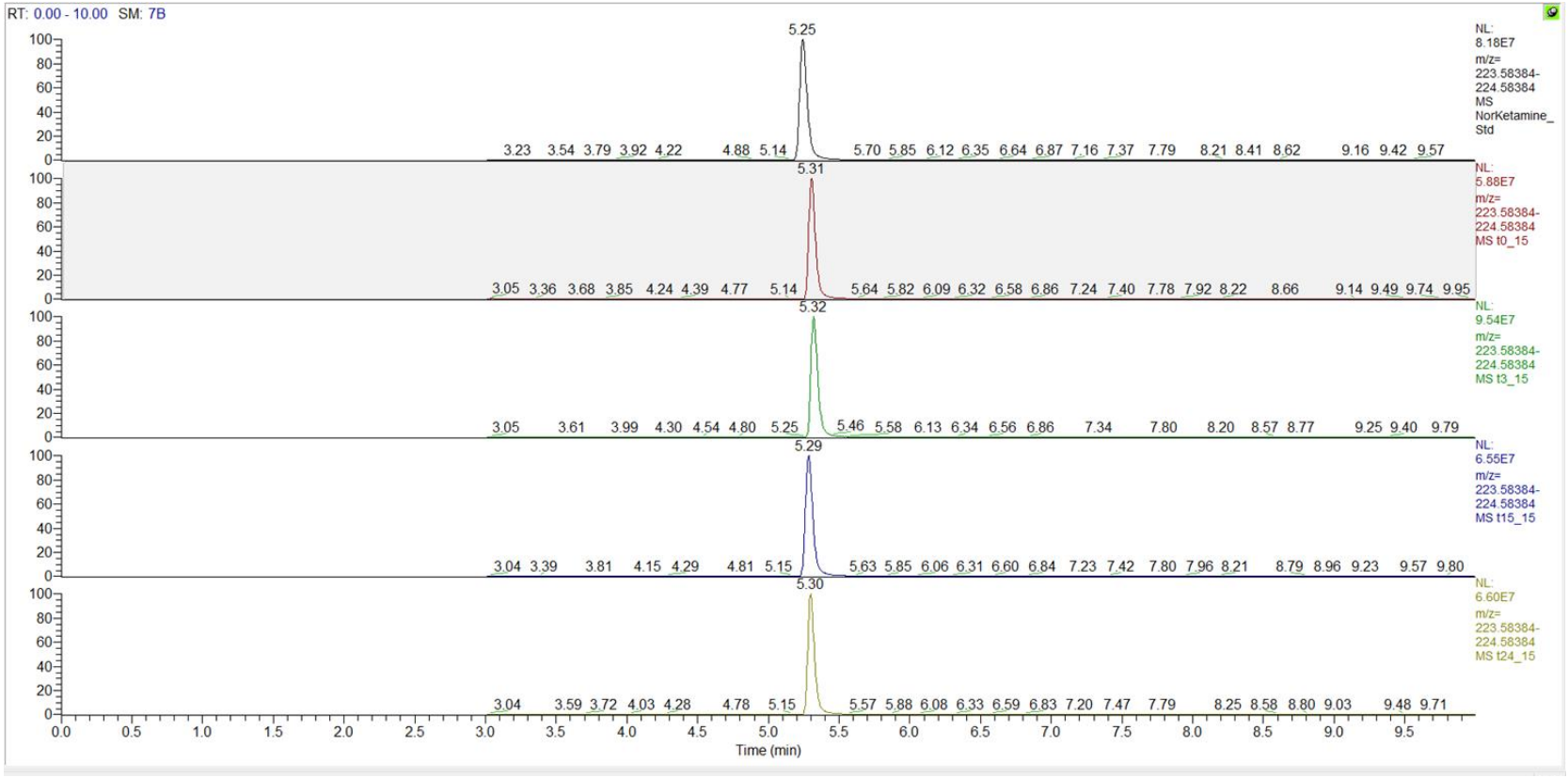


Figure 3-28 LC- MS Chromatogram of substrate norketamine standards in assay conditions for 24 hours (without enzyme). The calculated mass of norketamine (M+H) is 224.08367, while on the right of the panels is shown the mass of norketamine within 10 ppm range, as well as the abundance presented into 10^7. From top panel to bottom; SIM channel of norketamine standard, SIM channel of norketamine at T0 (red peak), SIM channel of norketamine at T3 hours (green peak), SIM channel of norketamine at T15 minutes (blue peak), and SIM channel of norketamine at T24 hours (light green peak).

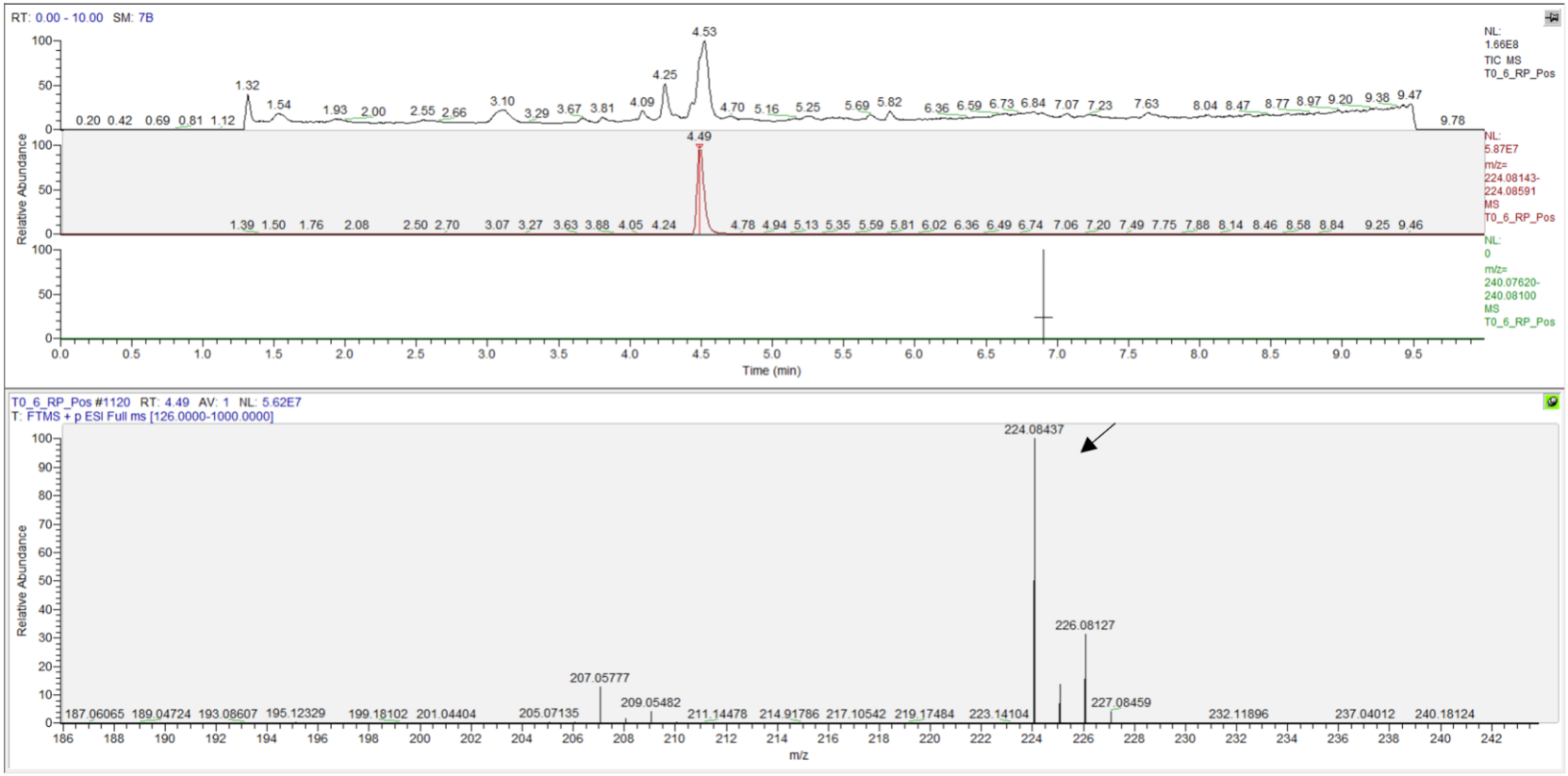


Figure 3-29 LC- MS Chromatogram of the human CYP2B6 -pET29b whole-cell activity against norketamine at T0. Top panel is shown the TIC with norketamine peak at retention time 4.53 (black peak), in the second panel is shown the SIM channel of mass norketamine with retention time 4.49 (red peak), in the third panel (blank) is the SIM channel of hydroxynorketamine peak (blank), and in the bottom panel is shown the high-resolution mass spectrum of norketamine with exact m/z 224.08437 and abundance 5.87 x10^7.

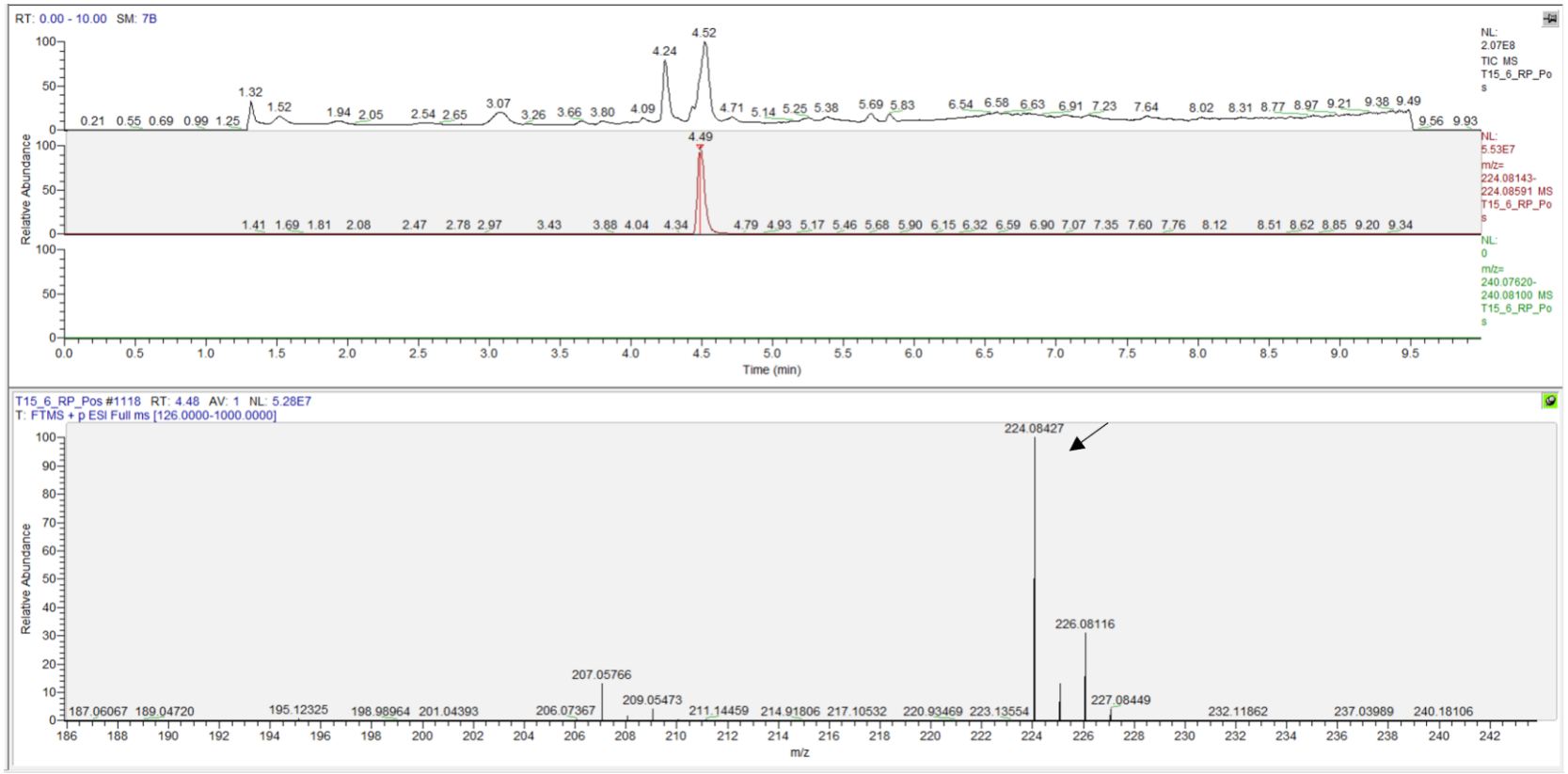


Figure 3-30 LC- MS Chromatogram of CYP2B6 -pET29b whole cell activity against norketamine at T15. Top panel is shown the TIC with norketamine peak at retention time 4.52 (black peak), in the second panel is shown the SIM channel of mass norketamine with retention time 4.49 (red peak), in the third panel is shown the SIM channel of mass hydroxynorketamine (blank) with expected m/z 240.07865, and in the bottom panel is shown the high-resolution mass spectrum of norketamine with exact m/z 224.08427 and abundance 5.53 x10^7.

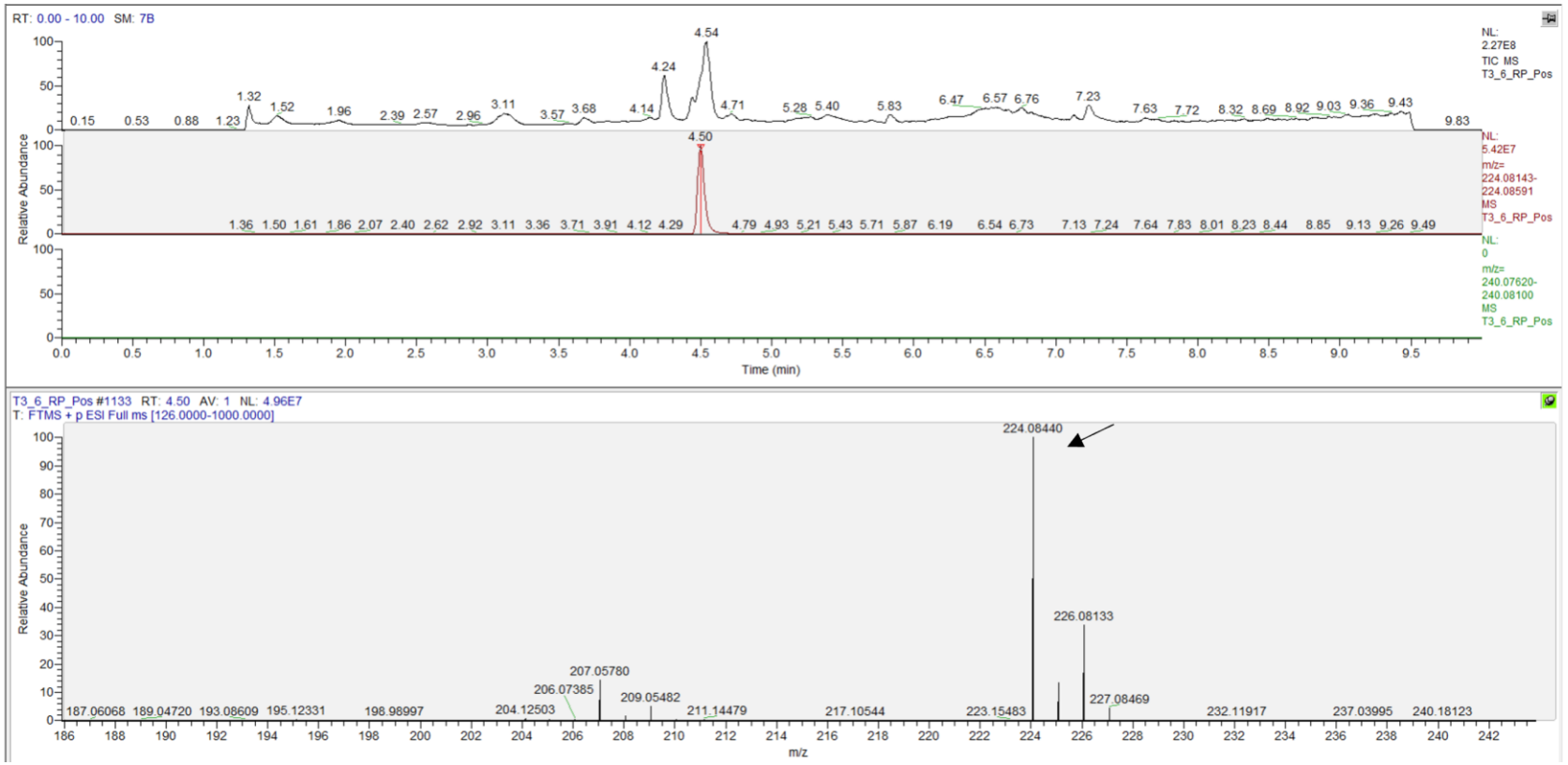


Figure 3-31 LC- MS Chromatogram of CYP2B6 -pET29b whole cell activity against norketamine at T3. Top panel is shown the TIC with norketamine peak at retention time 4.54 (black peak), in the second panel is shown the SIM channel of mass norketamine with retention time 4.50 (red peak), in the third panel is shown the SIM channel of mass hydroxynorketamine (blank) with expected m/z 240.07865, and in the bottom panel is shown the high-resolution mass spectrum of norketamine with exact m/z 224.08440 and abundance 5.42 x10^7.

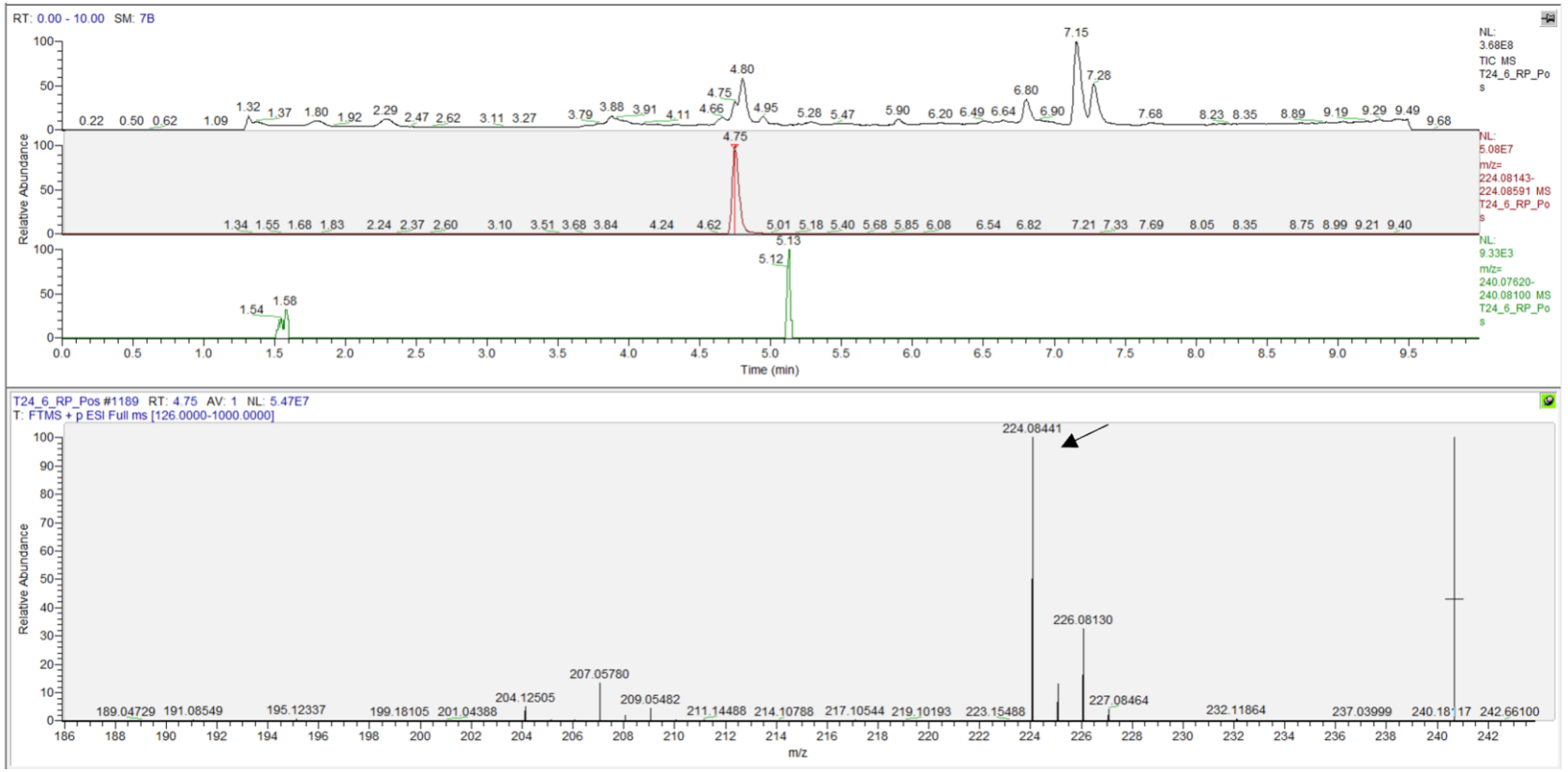


Figure 3-32 LC- MS Chromatogram of CYP2B6 -pET29b whole cell activity against norketamine at T24. Top panel is shown the TIC with norketamine peak at retention time 4.80 (black peak), in the second panel is shown the SIM channel of mass norketamine with retention time 4.75 (red peak), in the third panel is shown the SIM channel of mass hydroxynorketamine (green peak) with retention time 5.13 and abundance 9.33 $x10^3$, and in the bottom panel is shown the high-resolution mass spectrum of norketamine with exact m/z 224.08441 and abundance 5.08 $x10^7$.

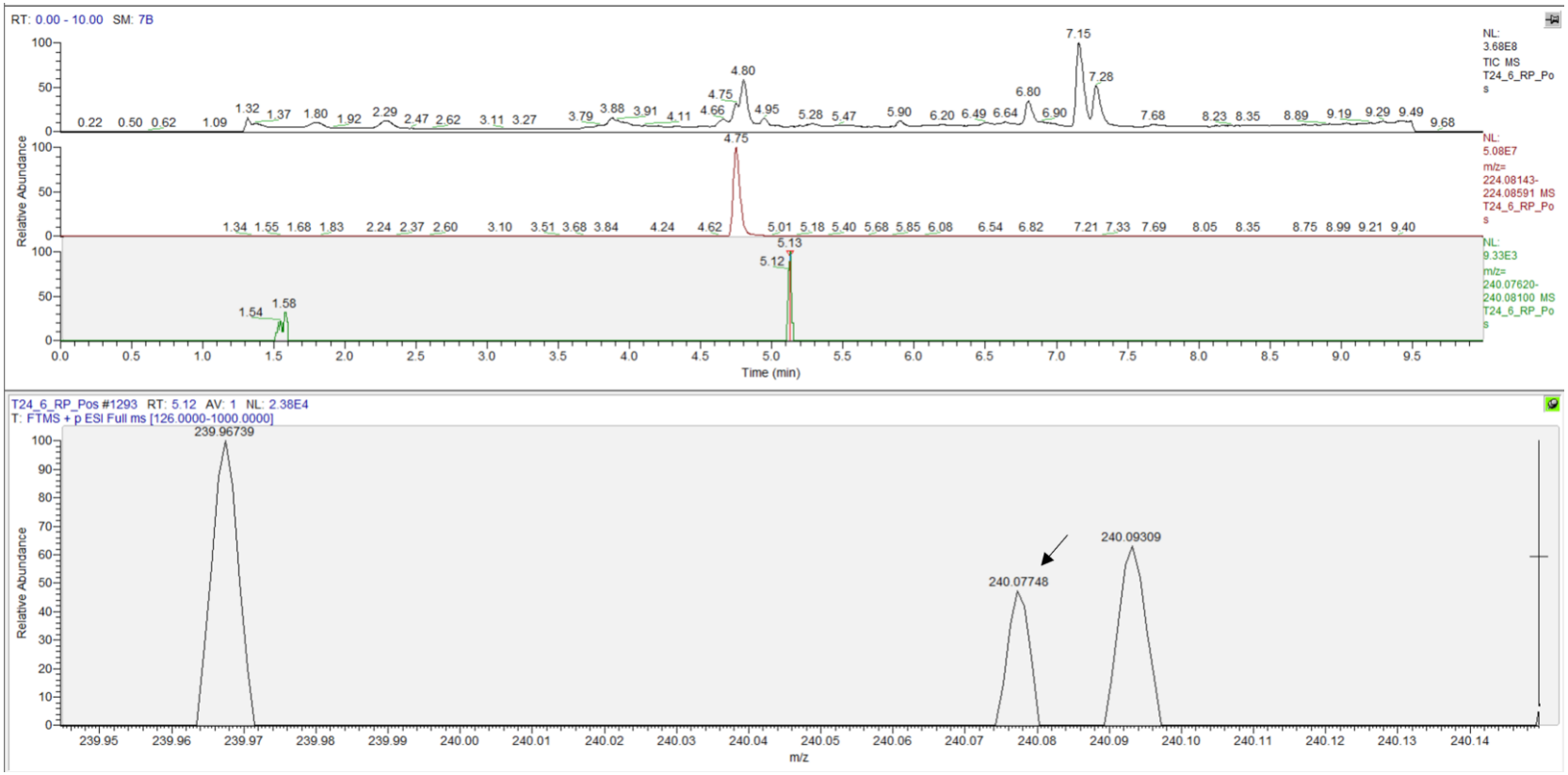


Figure 3-33 LC- MS Chromatogram of CYP2B6 -pET29b whole cell activity against norketamine at T24. Top panel is shown the TIC with norketamine peak at retention time 4.80 (black peak), in the second panel is shown the SIM channel of mass norketamine with retention time 4.75 (red peak), in the third panel is shown SIM channel of mass hydroxynorketamine (green highlighted with a red arrow) with retention time 5.13 and abundance 9.33 x10^3 with estimated m/z 240.07865, and the bottom panel is shown the high-resolution mass spectrum of hydroxynorketamine with exact m/z 240.07748 (indicated with an arrow), which is within 10 ppm from the estimated mass.

The second enzyme that converted norketamine to hydroxynorketamine, although in even lower abundance was CYP3A4-pET22b co-expressed with CPR-pCDFDuet-1 +pGro7 in Tuner cells in LB 30 °C, when induced with 0.25 mM IPTG.

The LC-MS analysis of the whole cell activities against norketamine at different timepoints (T0, T15, T3, and T24) is depicted in the chromatograms below, in Figure 3-34, Figure 3-35, Figure 3-36, Figure 3-37, and Figure 3-38.

At T0 (Figure 3-34) can be seen in the top panel the TIC with norketamine peak at retention time 5.20, in the second panel is shown the SIM channel of mass norketamine with retention time 5.31 and abundance 5.88 x 10^7, in the third panel is shown the SIM channel of mass hydroxynorketamine (blank), and in the bottom panel is shown the high-resolution mass spectrum of norketamine with exact m/z 224.08376.

Similarly, at T15 and T3 (Figure 3-35 and Figure 3-36) is shown in the top panels the TIC and norketamine peak with retention times of 5.29 and 5.32 (black peak), respectively. In the second panel is shown the SIM channel of mass norketamine with retention time 5.29 and 5.32 (indicated with a red arrow), respectively, and their abundances of 6.55 x 10^7 and 9.54 x 10^7, respectively. In the third panel is shown the SIM channel of mass hydroxyketamine (blank), and in the bottom panel is shown the high-resolution mass spectrum of norketamine with exact m/z ratio of 224.08376 and 224.08325, respectively.

In contrary, at T24 (Figure 3-37) can be seen the TIC in the top panel, in the second panel is shown the SIM channel of mass norketamine with retention time 5.30 and abundance 6.60 x10^7. In the third panel is shown the SIM channel of mass hydroxynorketamine where a peak has appeared with retention time 5.14 and abundance and 2.69 x10^3, which when selected (Figure 3-38, red arrow) can be seen in the bottom panel its high-resolution mass spectrum with an exact m/z of 240.07854, which is within 10 ppm of the estimated hydroxynorketamine m/z 240.07865. This indicates that at T24 CYP3A4 converted norketamine to hydroxynorketamine, although in very low abundance. Again, can be seen that the hydroxynorketamine peak is relatively low, and the signal is weak, which can possibly be attributed to the way the samples were prepared for injecting in the LC-MS, while no exact conversion can be calculated.

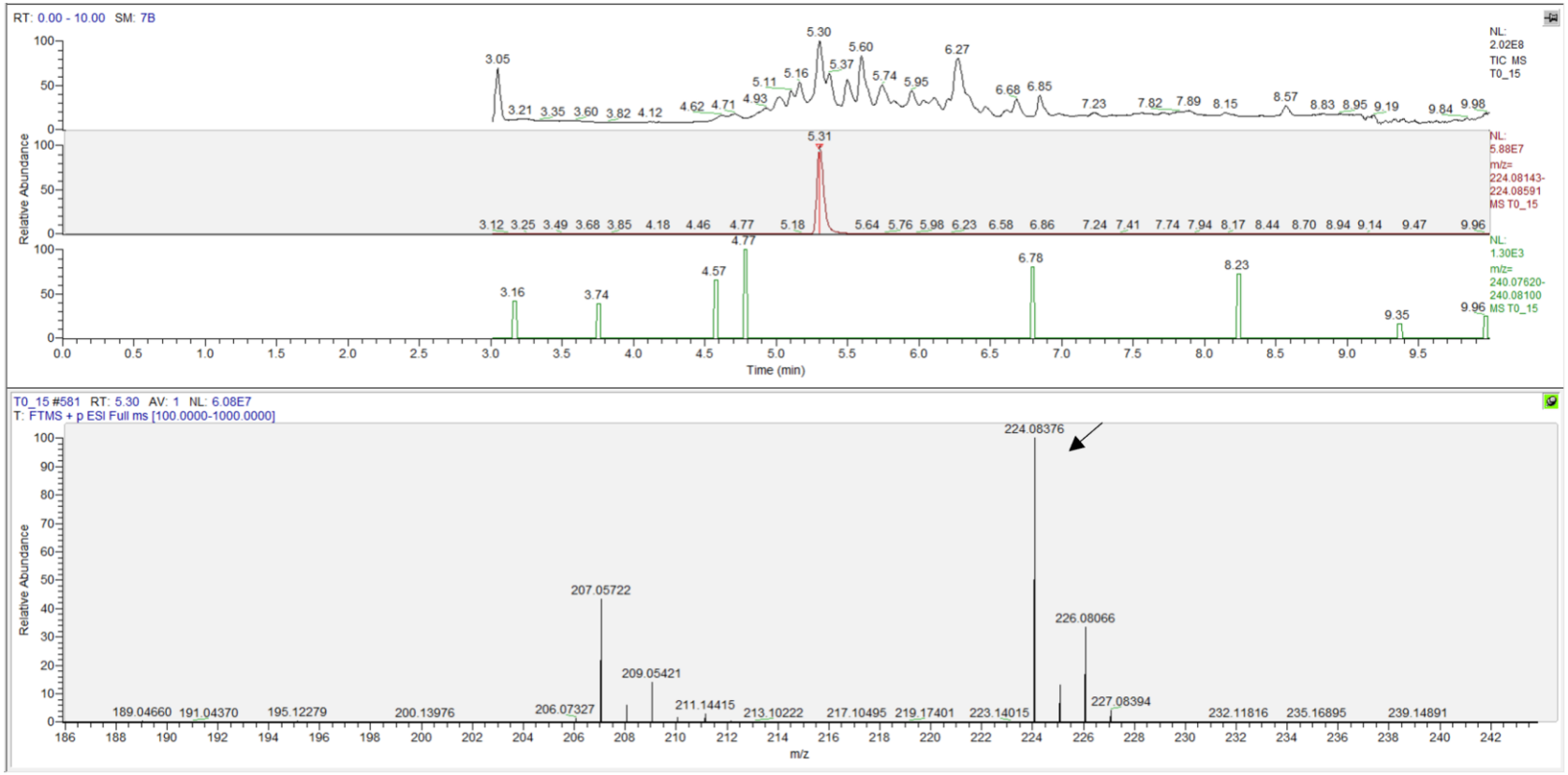


Figure 3-34 LC- MS Chromatogram of CYP3A4 -pET29b whole cell activity against norketamine at T0. Top panel is shown the TIC, in the second panel is shown the SIM channel of mass norketamine with retention time 5.31 (red peak), in the third panel is shown the SIM channel of mass hydroxynorketamine peak (blank), and in the bottom panel is shown the high-resolution mass spectrum of norketamine with exact m/z 224.08376 and abundance 5.88 x10^7.

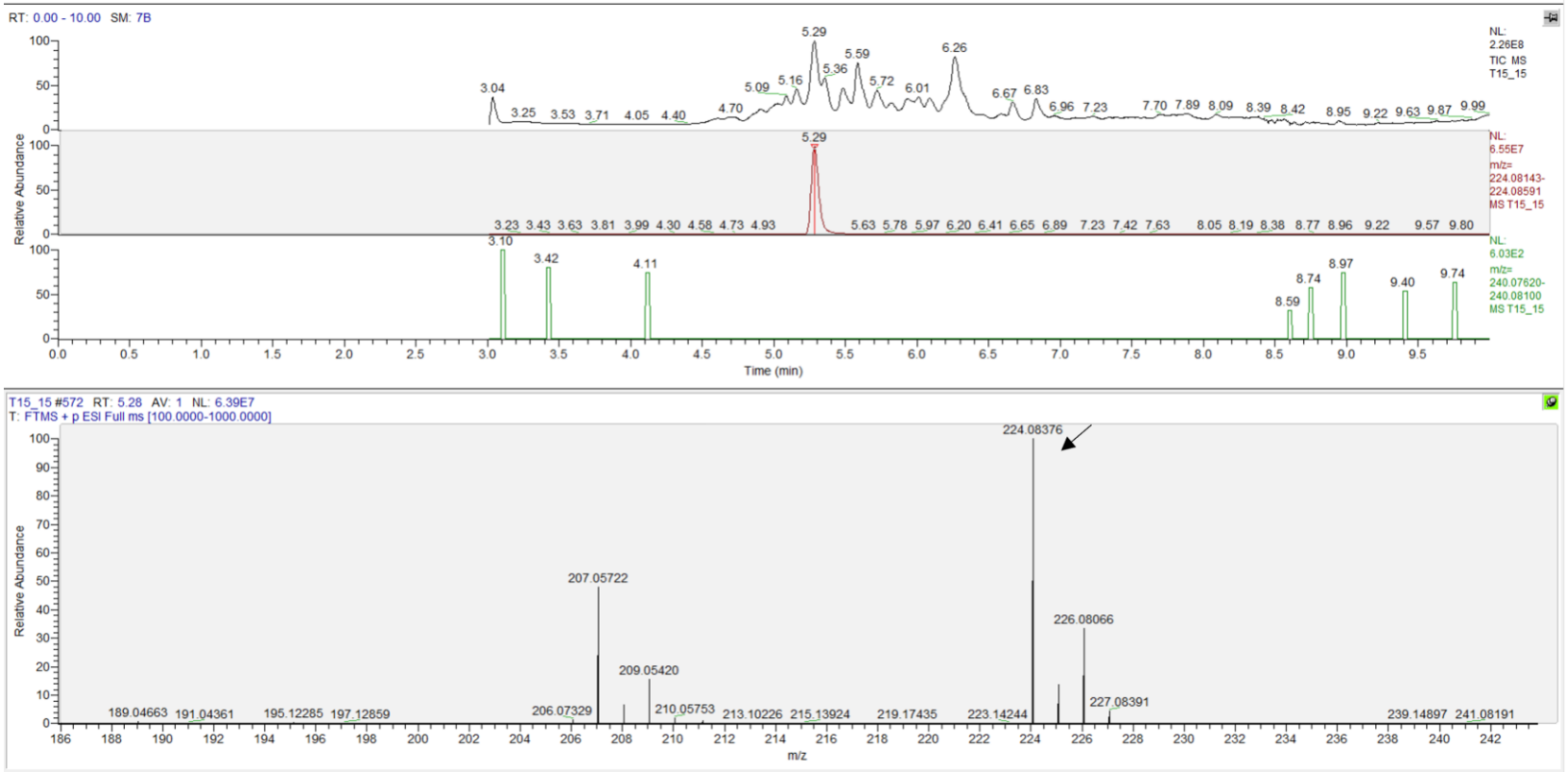


Figure 3-35 LC- MS Chromatogram of CYP3A4 -pET29b whole cell activity against norketamine at T15. Top panel is shown the TIC, in the second panel is shown the SIM channel of mass norketamine with retention time 5.29 (red peak), in the third panel is shown the SIM channel of mass hydroxynorketamine (blank), and in the bottom panel is shown the high-resolution mass spectrum of norketamine with exact m/z 224.08376 and abundance 6.55 x10^7.

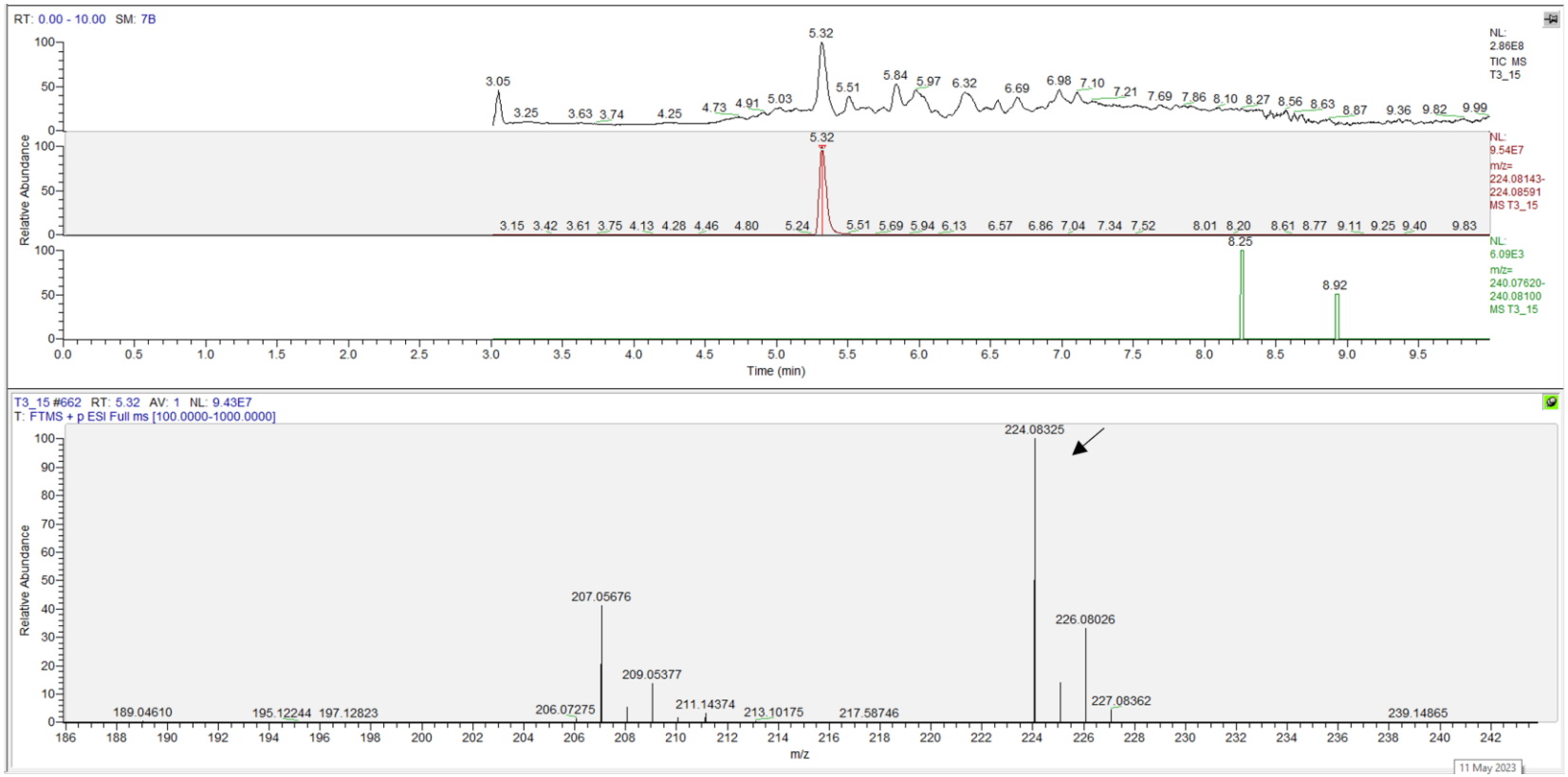


Figure 3-36 LC- MS Chromatogram of CYP3A4 -pET29b whole cell activity against substrate at T3. Top panel is shown the TIC with norketamine peak at retention time 5.32, in the second panel is shown the SIM channel of mass norketamine with retention time 5.32 (red peak), the third panel is SIM channel of mass hydroxynorketamine (blank), and the bottom panel is shown the high-resolution mass spectrum of norketamine with exact m/z 224.08325 and abundance 9.54 x10^7.

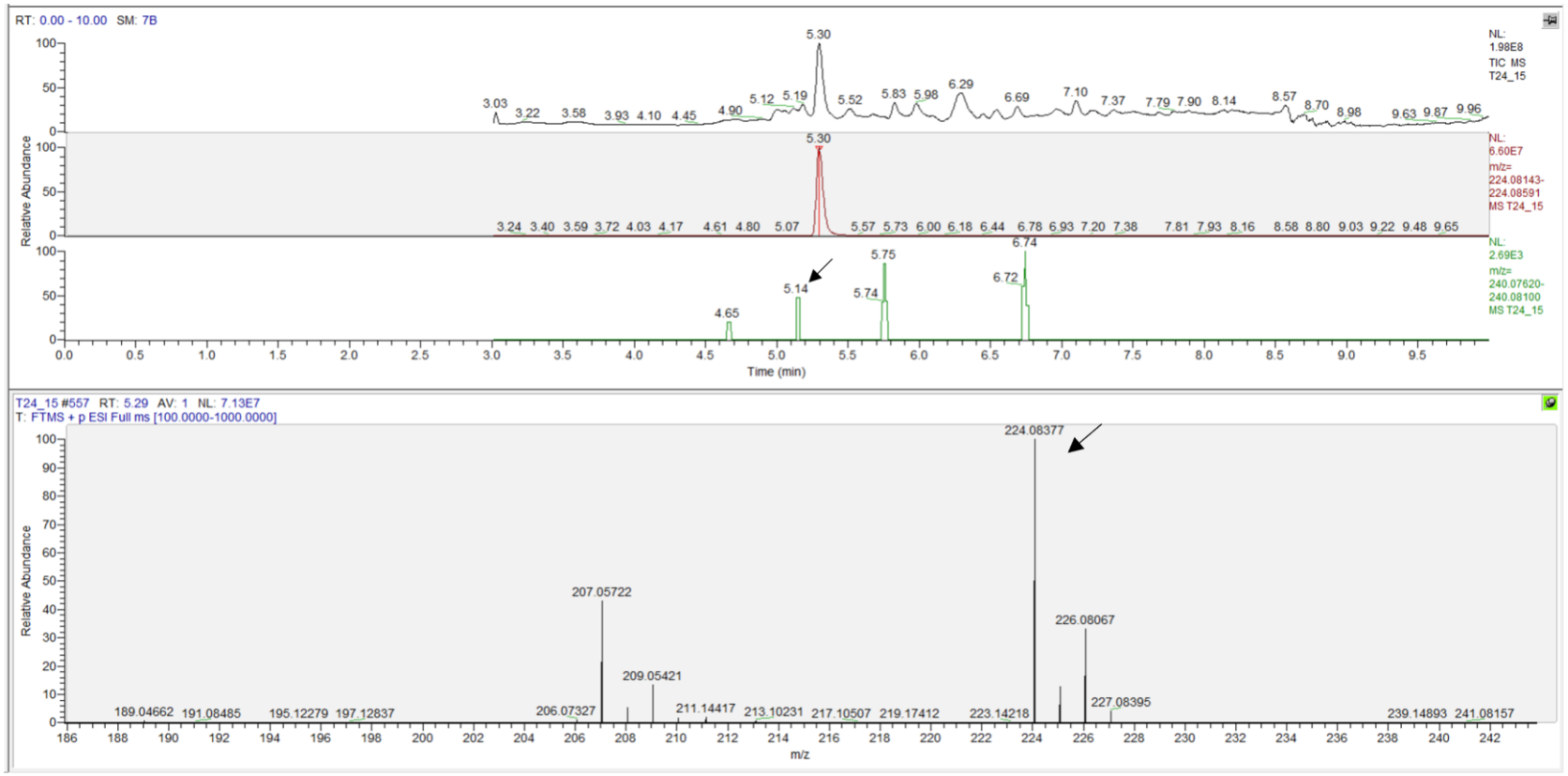


Figure 3-37 LC- MS Chromatogram of CYP3A4 -pET29b whole cell activity against norketamine at T24. Top panel is shown the TIC with norketamine peak at retention time 5.30, in the second panel is shown the SIM channel of mass norketamine with retention time 5.30 (red peak), the third panel is shown the SIM channel of mass hydroxynorketamine where a peak has appeared with retention time 5.14 (indicated by a black arrow), and the bottom panel is shown the high-resolution mass spectrum of norketamine with exact m/z 224.08377 and abundance 6.60 x10^7.

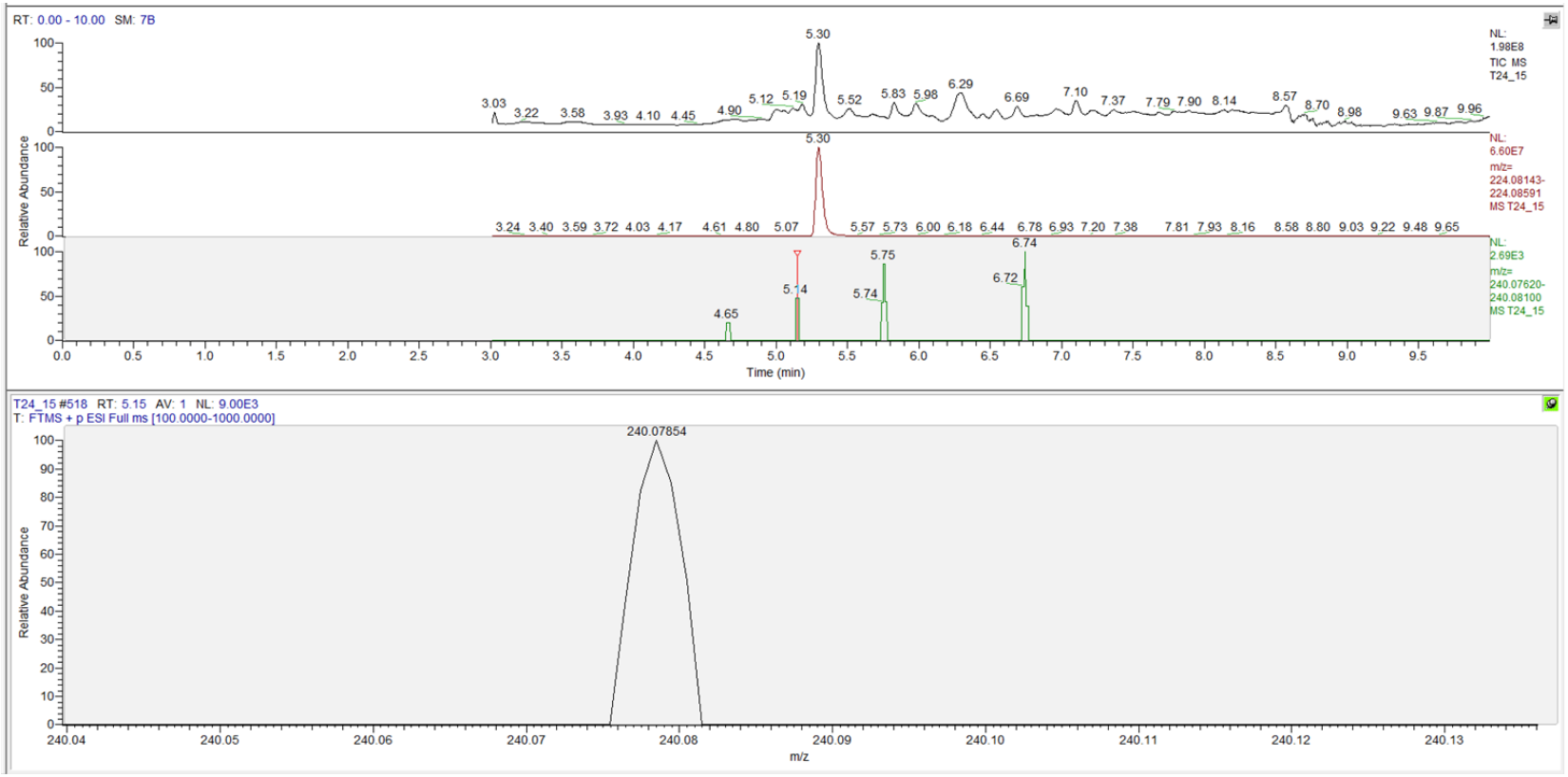


Figure 3-38 LC- MS Chromatogram of CYP3A4 -pET29b whole cell activity against norketamine at T24. Top panel is shown the TIC with norketamine peak at retention time 5.30, in the second panel is shown the SIM channel of mass norketamine with retention time 5.30 (red peak), in the third panel is shown the SIM channel of mass hydroxynorketamine retention time 5.14 (indicated by a red arrow), and in the bottom panel is shown the high-resolution mass spectrum of hydroxynorketamine peak with exact m/z 240.07854 and abundance 2.69 x10^3, which is within 10 ppm of estimated hydroxynorketamine m/z 240.07865.

3.10. Metagenome gene 914 assisted N-Demethylation of Ketamine to Norketamine, and Hydroxylation to Hydroxyketamine

After mining the Prozomix metagenome library, two metagenome genes, 198 and 914, that share 20% and 24% homology with the human CYP3A5, respectively, which were employed in ketamine hydroxylation assays to produce hydroxyketamine, and possibly its metabolite, norketamine to hydroxynorketamine. Metagenome enzyme 914 was expressed and was CO active in most al of the expression assays with and without the N-terminus modification, but only soluble without the N-terminus modification. In contrast, metagenome enzyme 198 was CO active only when the modification MALLLAVF was applied at the N-terminus, and co-expressed with pGro7 in Tuner cells, and induced with 0.5 mM IPTG. Moreover, 914-pGEX-6P-1 enzyme hydroxylated ketamine to hydroxyketamine and N-demethylated it to norketamine, rapidly, while 198-pET29b did not show any activity against ketamine or its metabolite, norketamine.

Ketamine standards were run in the LC-MS, and the results are shown in Figure 3-39. In the top panel is presented the TIC of ketamine with ketamine peak at retention time 5.35, in the second panel is presented the SIM channel of mass ketamine with retention time 5.35 and abundance 2.68 x 10^8, and in the third panel is presented the high-resolution mass spectrum of ketamine with the exact m/z of 238.09973.

In Figure 3-40, Figure 3-41, Figure 3-42, and Figure 3-43 are presented the LC-MS chromatograms for 914 -pGEX-6P-1 whole cell activity against ketamine at T0, T15, T2 and T24, respectively. As can be seen from the chromatograms enzyme 914-pGEX-6P-1 when co-expressed with pGro7 in BL21(DE3) in LB 20 °C (MS Sample 4) hydroxylated ketamine to hydroxyketamine, and N-demethylated it to norketamine since T0. Although this is labelled T0, it is T10 in real time, which is the time it takes from starting the assay with the addition of NADPH, (collect the sample, spin down, and separate into pellet and supernatant), to stopping the assay by boiling the sample. During those 10 minutes ketamine is rapidly hydroxylated. In the top panels across all four chromatograms are presented the TICs, in the second panels are presented the SIM channels of mass norketamine (except for 1st chromatogram where norketamine and ketamine SIM channels are presented in reverse), in the third panels are presented the SIM channels of mass ketamine, and in the bottom panels are presented the SIM channels of mass hydroxyketamine.

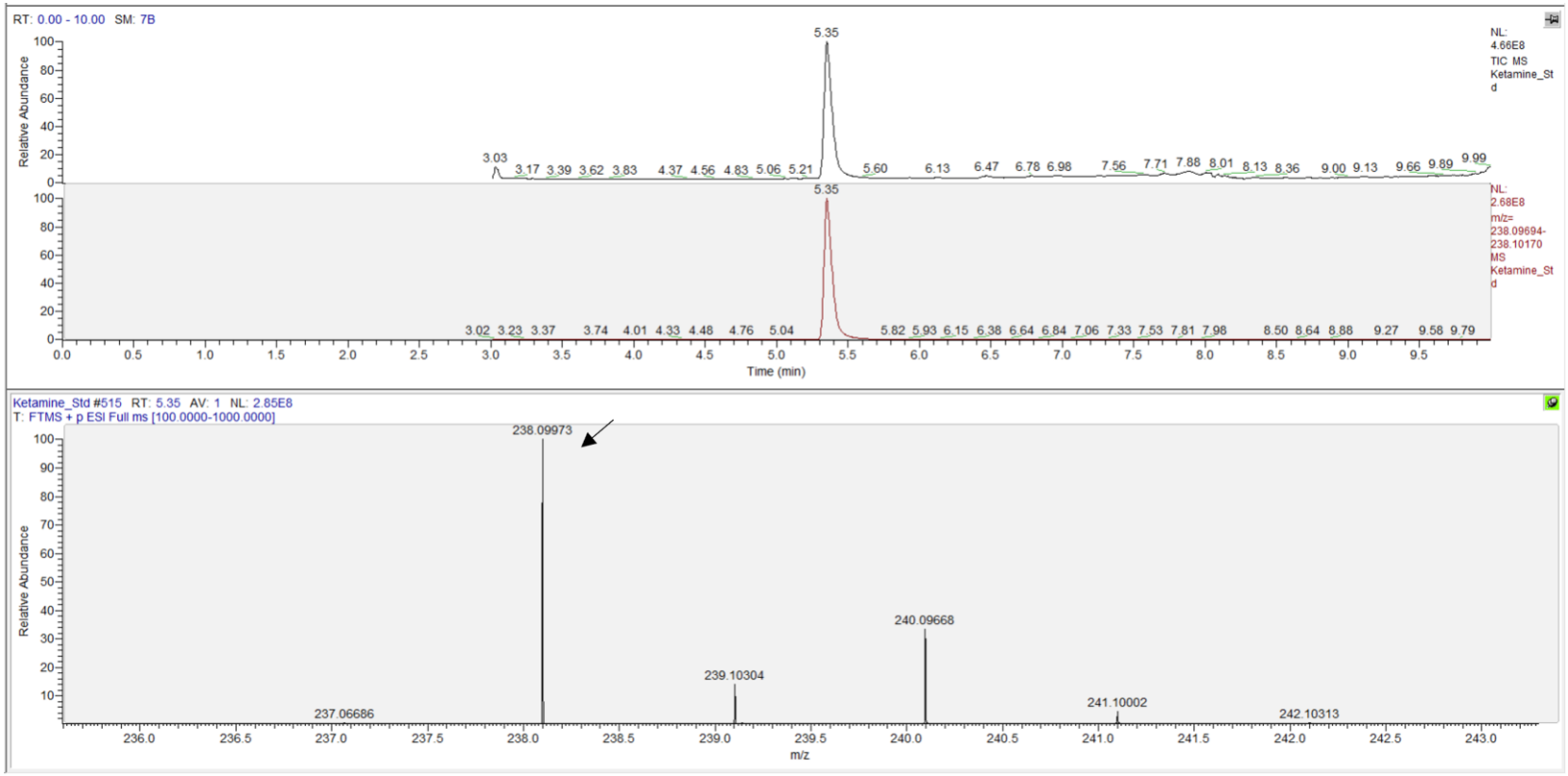


Figure 3-39 LC- MS Chromatogram of ketamine standard. Top panel is the TIC of ketamine with retention time 5.35, in the second panel is the SIM channel of mass ketamine with retention time 5.35 and abundance 2.68 x 10^8 (shown on right of the panel), in the third panel is presented the high-resolution mass spectrum of ketamine with the exact m/z of 238.09973. The calculated mass of ketamine (taking into consideration the protons) is 238.09932.

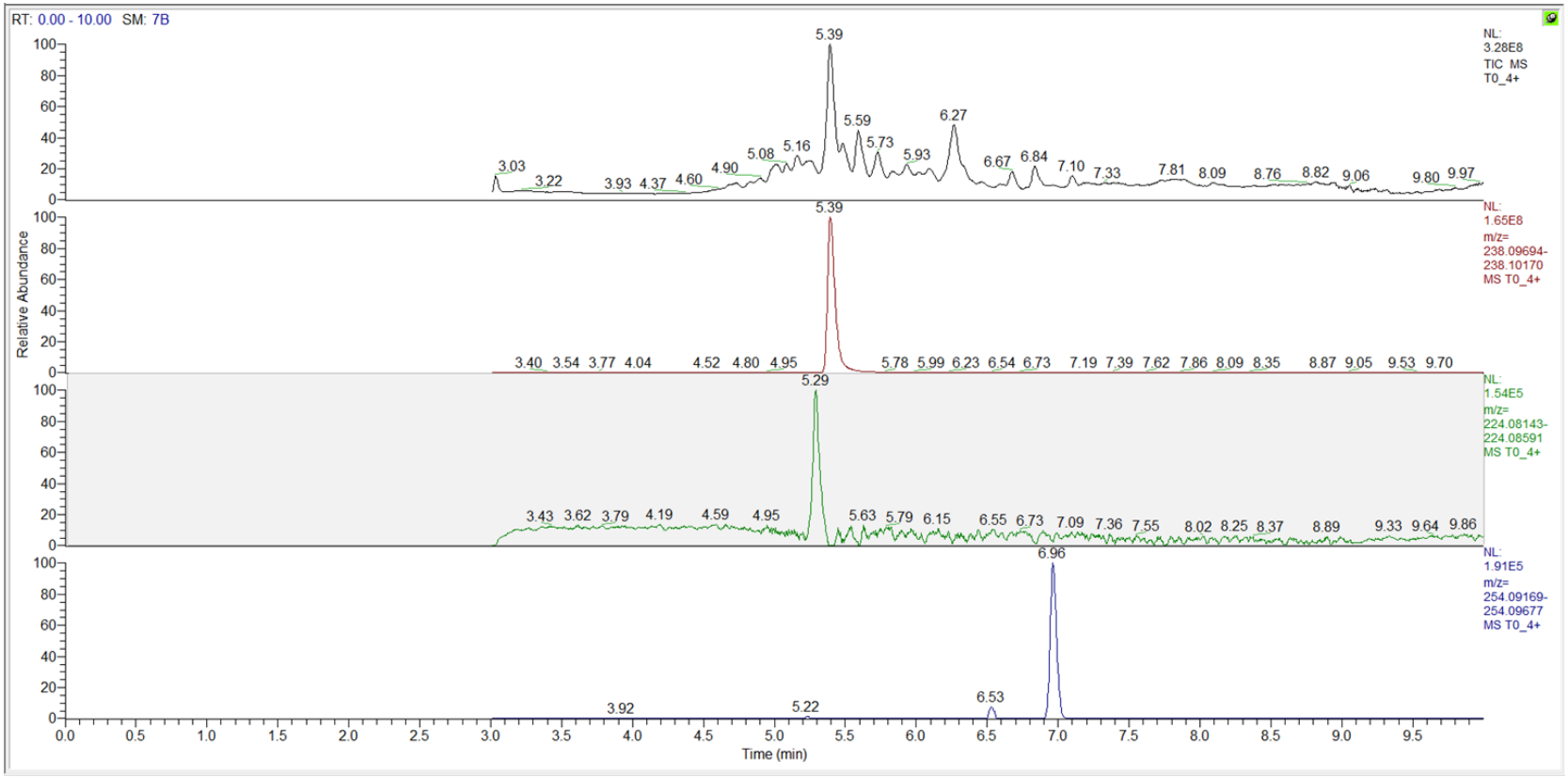


Figure 3-40 LC-MS Chromatogram of 914 -pGEX-6P-1 whole cell activity against ketamine at T0. Top panel is shown the TIC with ketamine peak at retention time 5.39 (black peak), in the second panel is shown the SIM channel of mass ketamine with retention time 5.39 (red peak) and abundance 1.65 $x10^8$, in the third panel is shown the SIM channel of mass norketamine (green peak) with retention time 5.29 and abundance 1.45 $x10^5$, and in the bottom panel is shown the SIM channel of mass hydroxyketamine with retention time 6.96 and abundance 1.91 $x10^5$.

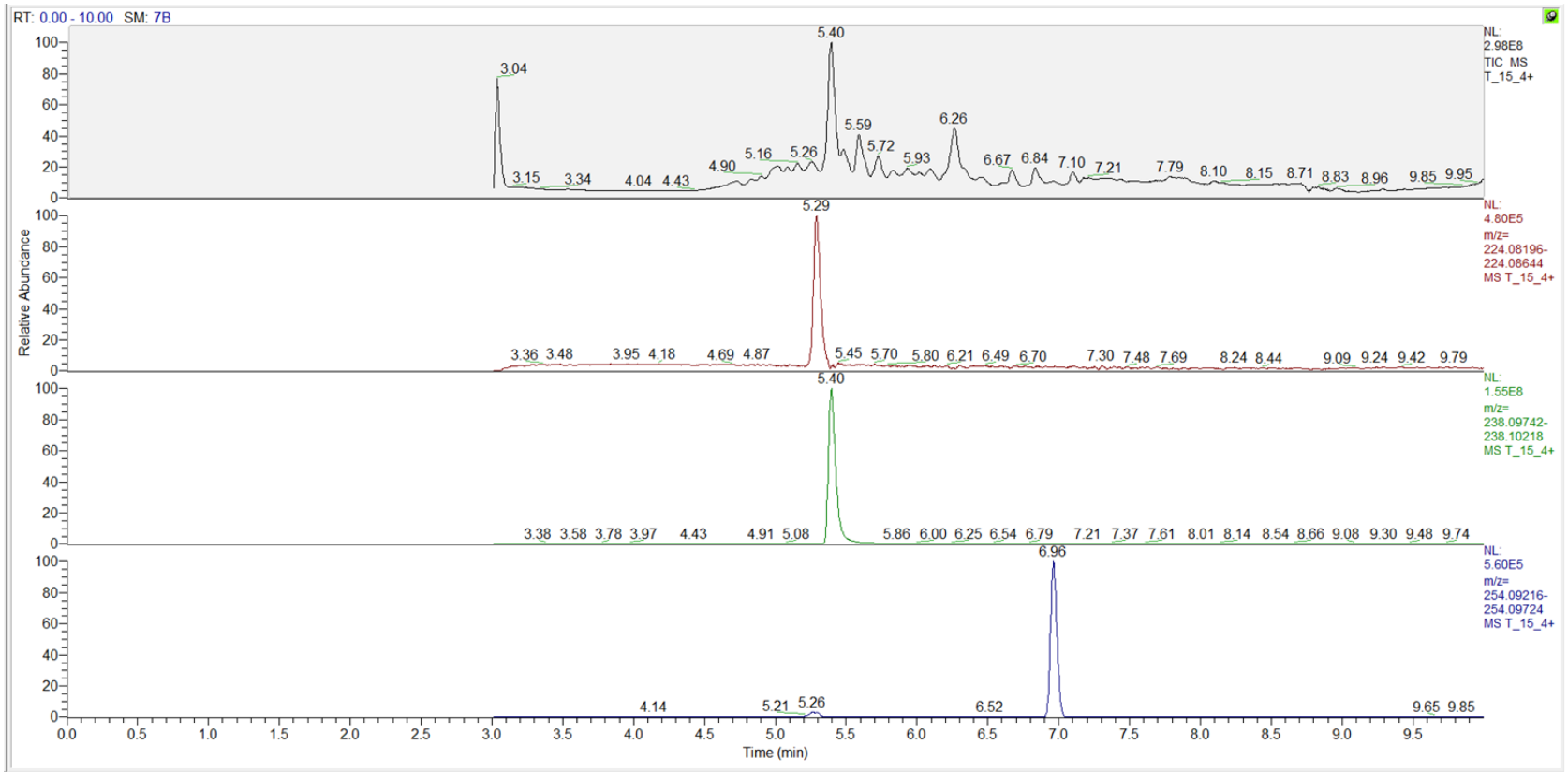


Figure 3-41 LC-MS Chromatogram of 914 -pGEX-6P-1 whole cell activity against ketamine at T15. Top panel is shown the TIC with ketamine peak at retention time 5.40 (black peak), in the second panel is shown the SIM channel of mass norketamine with retention time 5.29 (red peak) and abundance 5.80 x10^5, in the third panel is shown the SIM channel of mass ketamine (green peak) with retention time 5.40 and abundance 1.55 x10^8, and in the bottom panel is shown the SIM channel of mass hydroxyketamine with retention time 6.96 and abundance 5.60 x10^5.

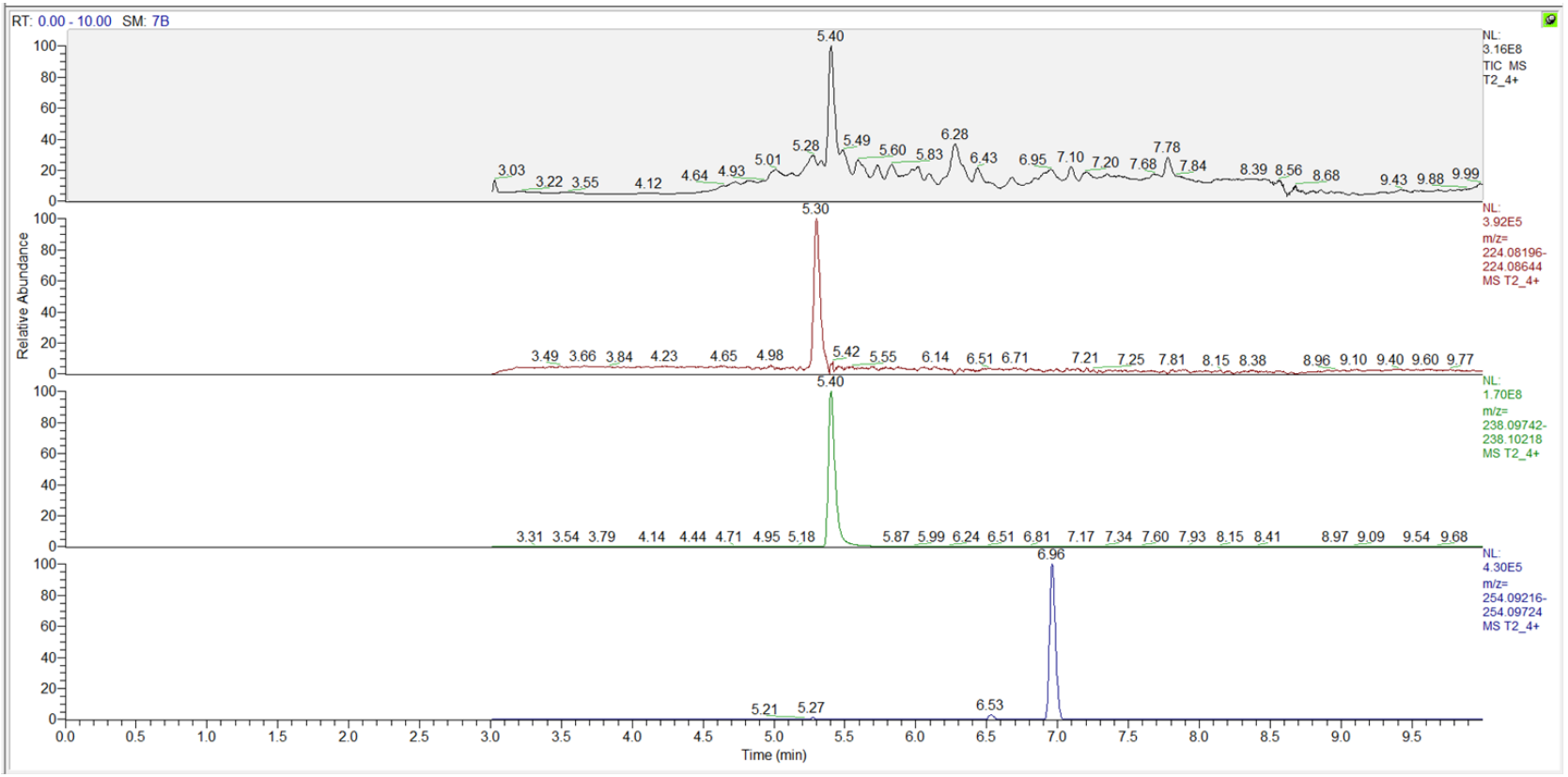


Figure 3-42 LC-MS Chromatogram of 914 -pGEX-6P-1 whole cell activity against ketamine at T2. Top panel is shown the TIC and ketamine peak at retention time 5.40 (black peak), in the second panel is shown the SIM channel of mass norketamine with retention time 5.30 (red peak) and abundance 3.92 x10^5, in the third panel is shown the SIM channel of mass ketamine (green peak) with retention time 5.40 and abundance 1.70 x10^8, and in the bottom panel is shown the SIM channel of mass hydroxyketamine with retention time 6.96 and abundance 4.30 x10^5.

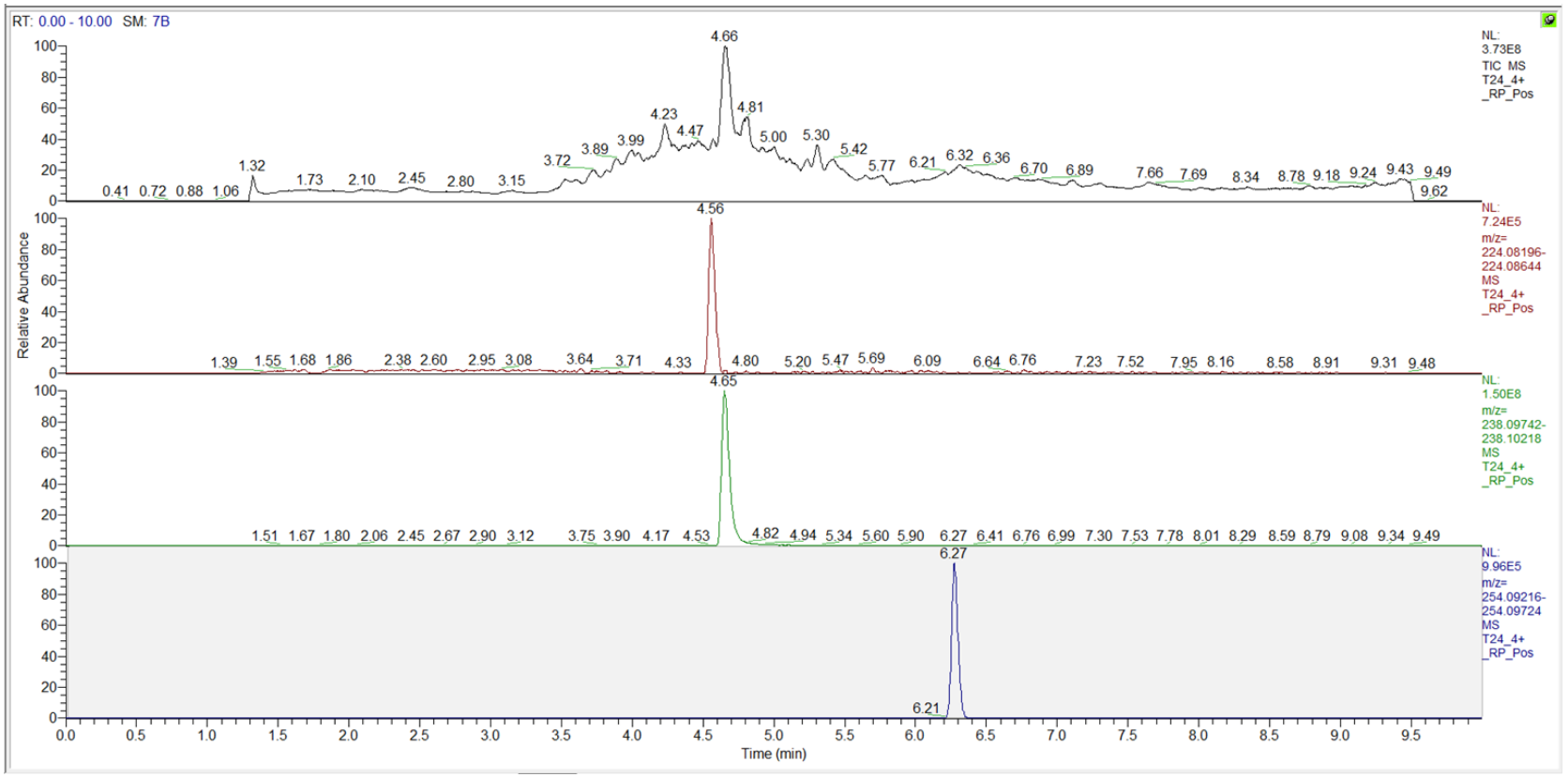


Figure 3-43 LC-MS Chromatogram of 914 -pGEX-6P-1 whole cell activity against ketamine at T24. Top panel is shown the TIC with ketamine peak at retention time 4.6 (black peak), in the second panel is shown the SIM channel of mass norketamine with retention time 4.56 (red peak) and abundance 7.24 x10^5, in the third panel is shown the SIM channel of mass ketamine (green peak) with retention time 4.65 and abundance 1.50 x10^8, and in the bottom panel is shown the SIM channel of mass hydroxyketamine with retention time 6.27 and abundance 9.96 x10^5.

Additionally, in Figure 3-44 to Figure 3-55 are presented the individual peaks of every single metabolite (ketamine, norketamine and hydroxyketamine) across all timepoints (T0, T15, T2 and T24). Specifically, in the first four LC-MS chromatograms (Figure 3-44 Figure 3-45, Figure 3-46 and Figure 3-47) are shown the TICs in the top panels, the SIM channels are shown in the middle panels, and the high-resolution mass spectra of metabolite ketamine across all four timepoints are shown in the bottom panels. In the next four chromatograms (Figure 3-48, Figure 3-49, Figure 3-50 and Figure 3-51) in the top panels are shown are shown the TICs, in the middle panels are presented the SIM channels, and in the bottom panels are presented the high-resolution mass spectra of metabolite norketamine across all four timepoints, and in the last four chromatograms (Figure 3-52,Figure 3-53 , Figure 3-54 and Figure 3-55) in the top panels are demonstrated the TICs, in the middle panels are demonstrated the SIM channels, and in the bottom panels are represented the high-resolution mass spectra of metabolite hydroxyketamine across all four timepoints. All abundances of each metabolite can be seen on the right side of each corresponding channel, while the exact m/z ratios can be seen in the high-resolution mass spectra panels, which are all within 10 ppm of the predicted masses.

The results of an investigation of the relationship between the decreasing substrate abundance and the rising product abundance over time are shown in Figure 3-56. Although there is a minor reduction of ketamine abundance (represented as a blue bar) between T0 and T2 as can be seen, ketamine abundance decreased to 30% after 24 hours. On the other hand, both norketamine (represented as orange bar) and hydroxyketamine (represented as green bar) increase to some extent from T0 to T2. Then at 24 hours, from looking at the graph, ketamine has increased to 30% of the total sample, whereas hydroxyketamine has increased to 40% of the total sample. This means that the starting material (ketamine) and products (hydroxyketamine, norkeketamine) are in a ratio of substrate to compound 3 to 4 to 3.

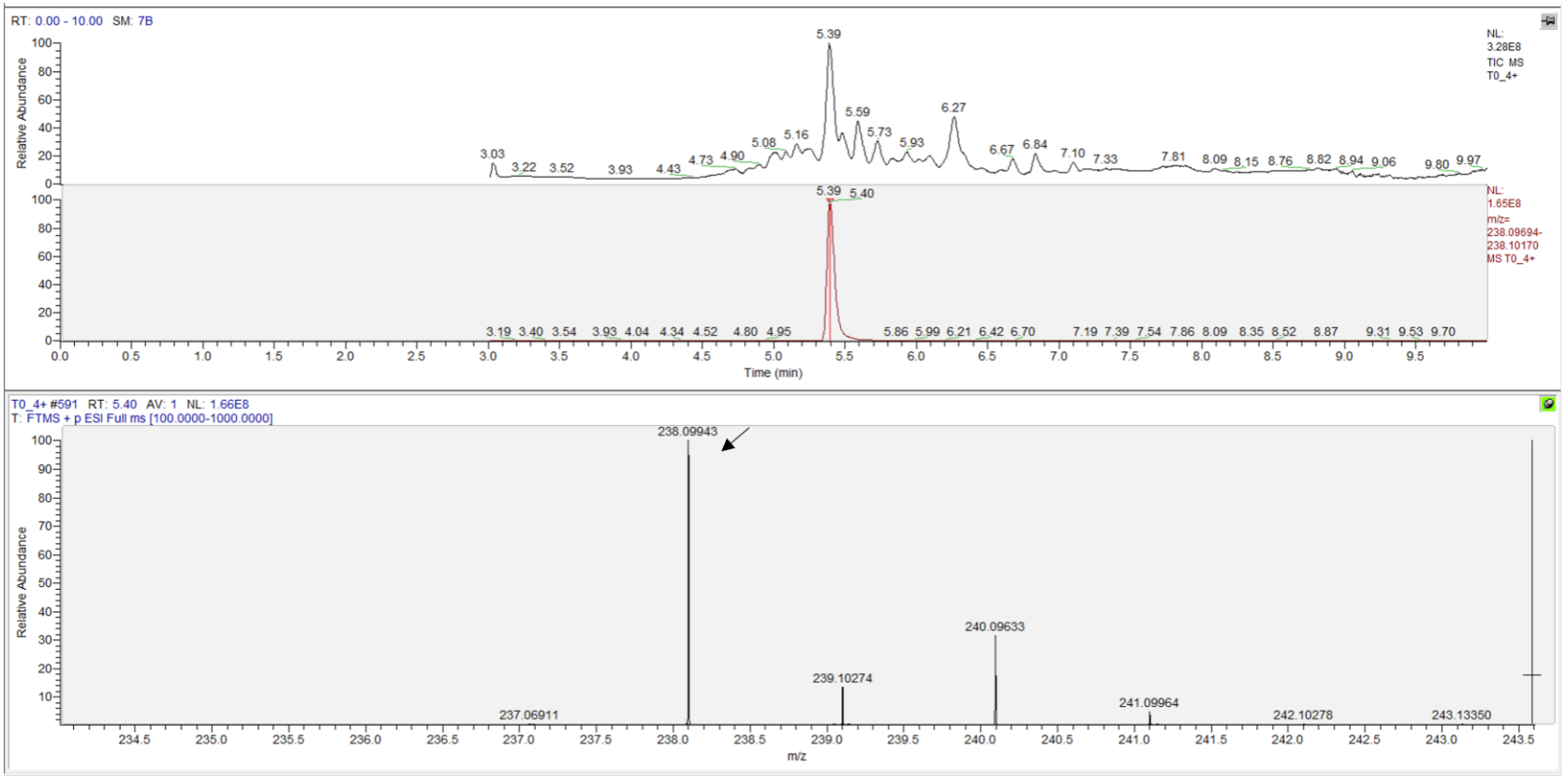


Figure 3-44 LC- MS Chromatogram of 914 -pGEX-6P-1 whole cell activity against ketamine at T0. Top panel is shown the TIC and ketamine peak with retention time 5.39 (black peak), in the second panel is shown the SIM channel of mass ketamine with retention time 5.39 (red peak) and abundance 1.65 x10^{8}, in the third panel is shown the high-resolution mass spectrum of ketamine with the exact m/z of 238.09943.

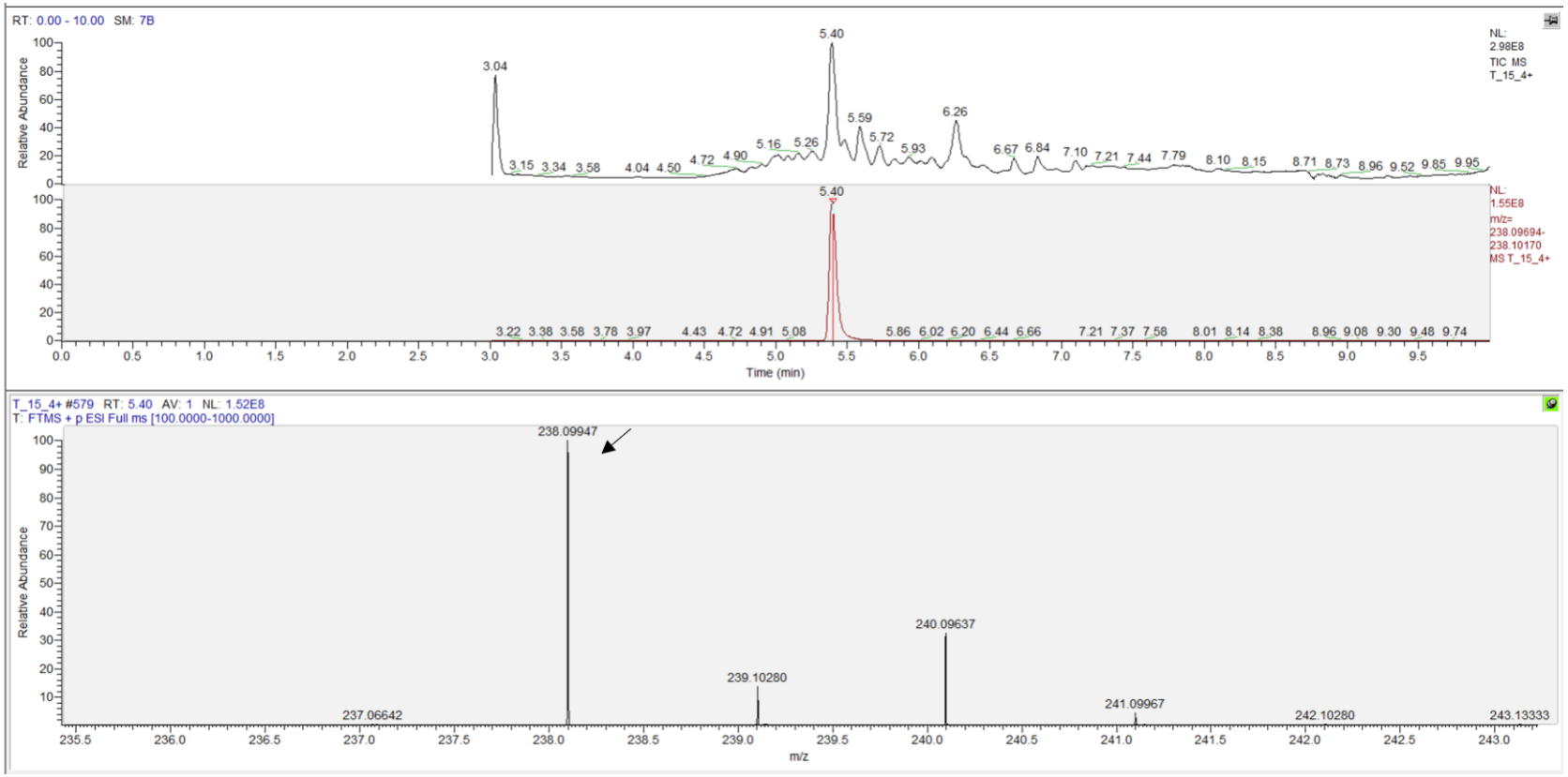


Figure 3-45 LC- MS Chromatogram of 914 -pGEX-6P-1 whole cell activity against ketamine at T15. Top panel is shown the TIC and ketamine peak with retention time 5.40 (black peak), in the second panel is shown the SIM channel of mass ketamine with retention time 5.40 (red peak) and abundance 1.55 x10^8, and in the third panel is shown the high-resolution mass spectrum of ketamine with the exact m/z of 238.09947.

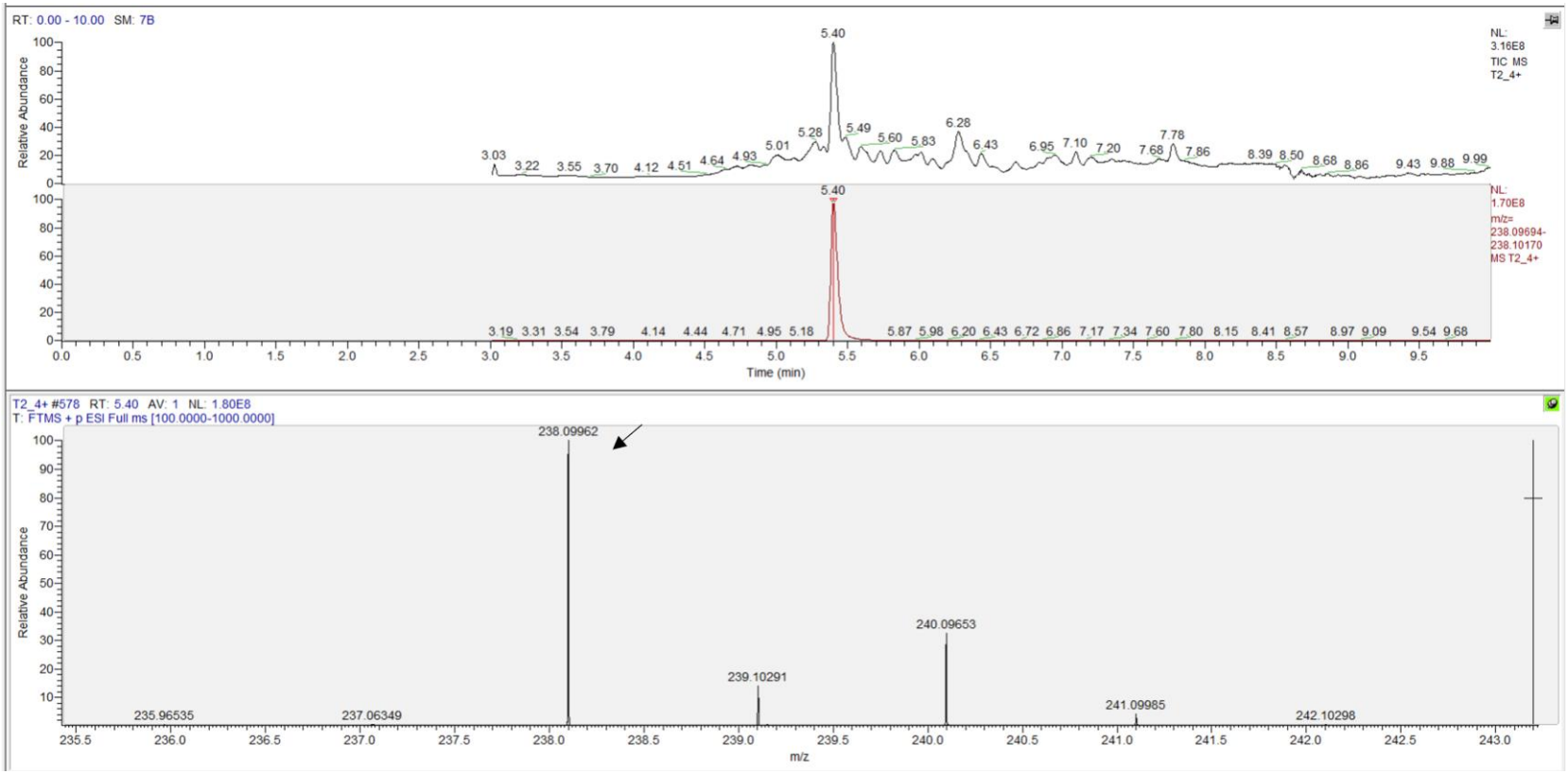


Figure 3-46 LC- MS Chromatogram of 914 -pGEX-6P-1 whole cell activity against ketamine at T2. Top panel is shown the TIC and ketamine peak with retention time 5.40 (black peak), in the second panel is shown the SIM channel of mass ketamine with retention time 5.40 (red peak) and abundance 1.70 x10^8, and in the third panel is shown the high-resolution mass spectrum of ketamine with the exact m/z of 238.09962.

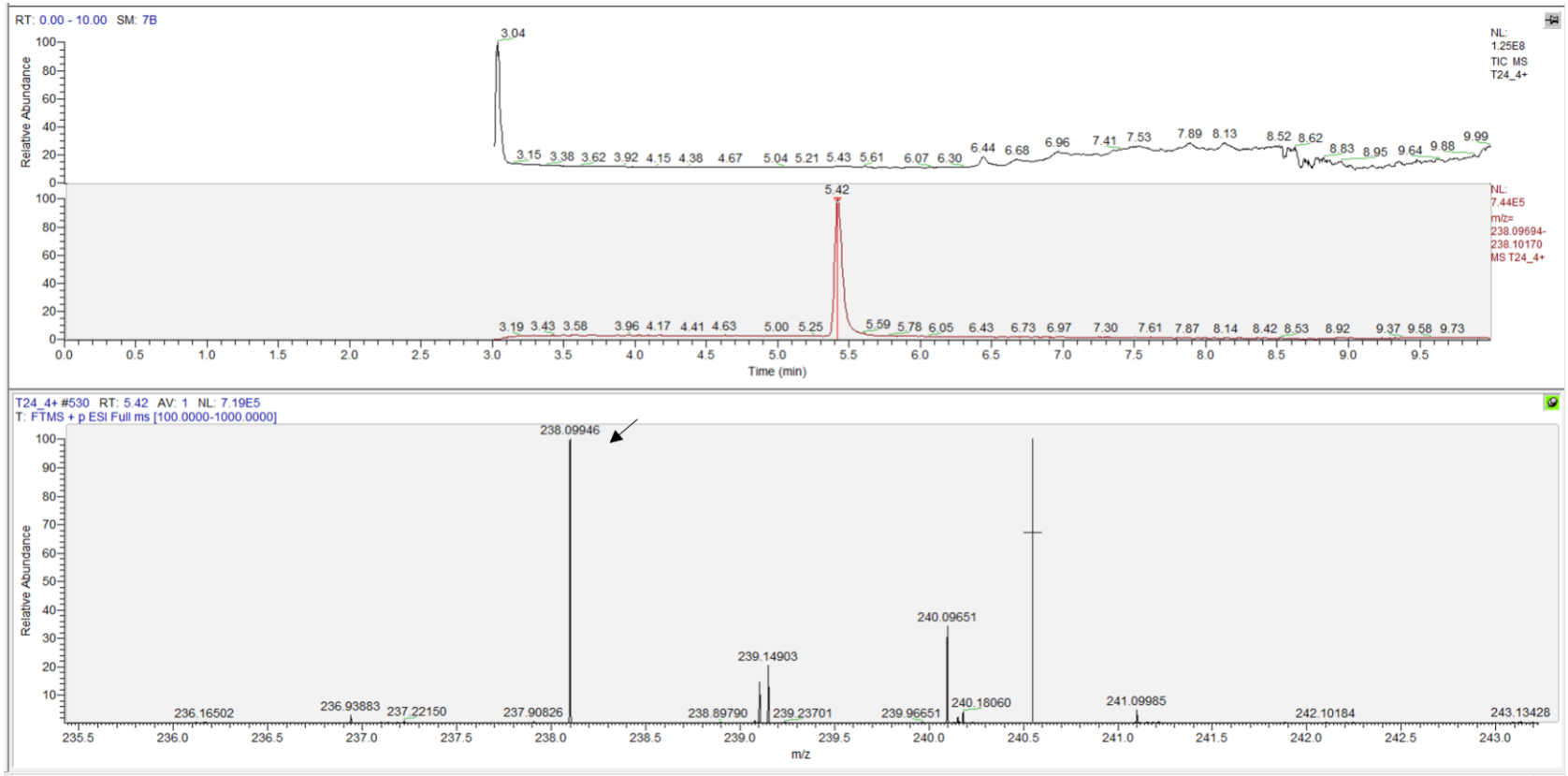


Figure 3-47 LC- MS Chromatogram of 914 -pGEX-6P-1 whole cell activity against ketamine at T24. Top panel is shown the TIC, in the second panel is shown the SIM channel of mass ketamine with retention time 5.42 (red peak) and abundance 7.44 x10^5, and in the third panel is shown the high-resolution mass spectrum of ketamine with the exact m/z of 238.09946.

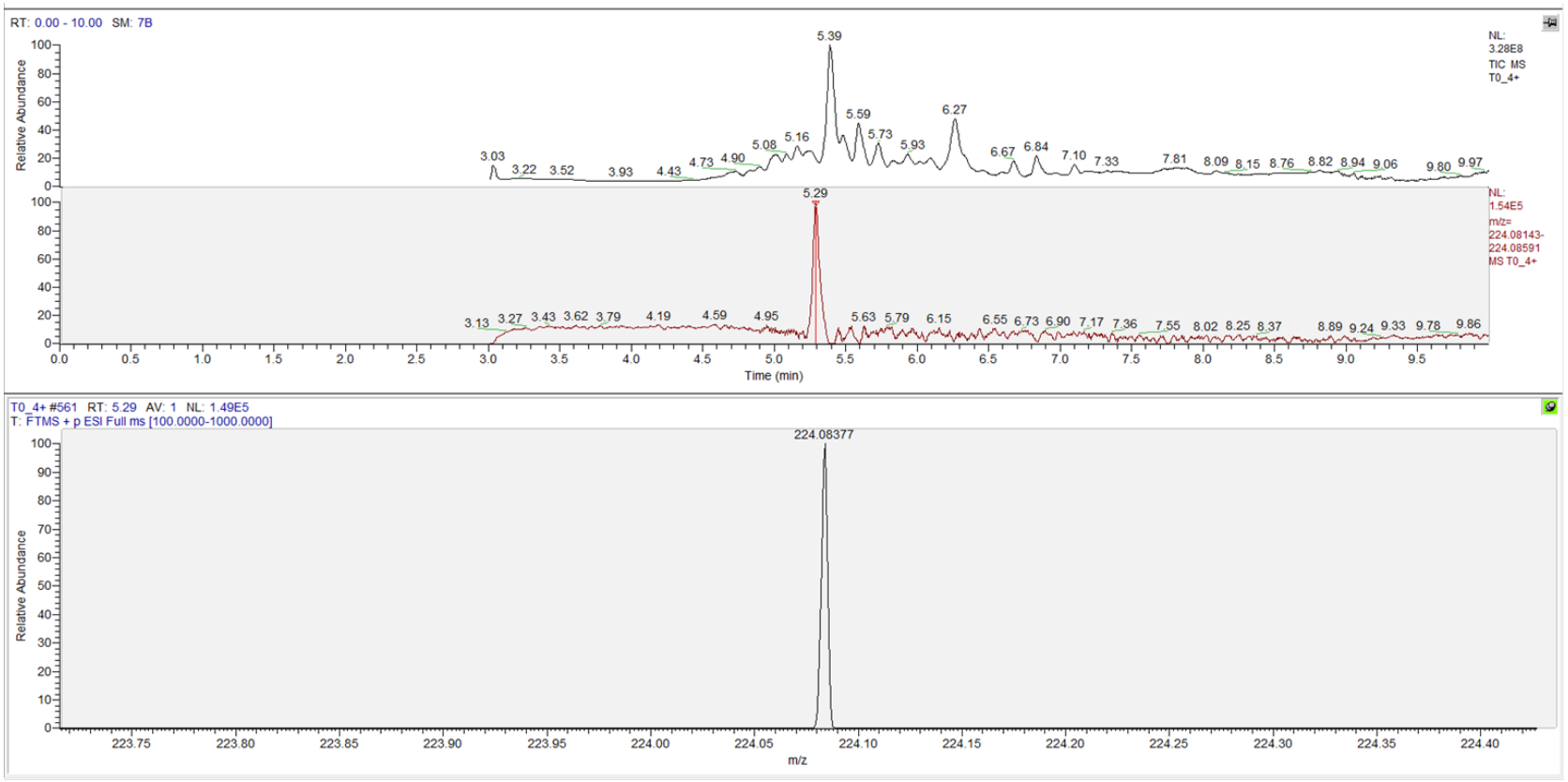


Figure 3-48 LC- MS Chromatogram of 914 -pGEX-6P-1 whole cell activity against ketamine at T0. Top panel is shown the TIC and ketamine peak with retention time 5.39 (black peak), in the second panel is shown the SIM channel of mass norketamine with retention time 5.29 (red peak) and abundance 1.54 x10^5, and in the third panel is shown the high-resolution mass spectrum of norketamine with the exact m/z of 224.08377.

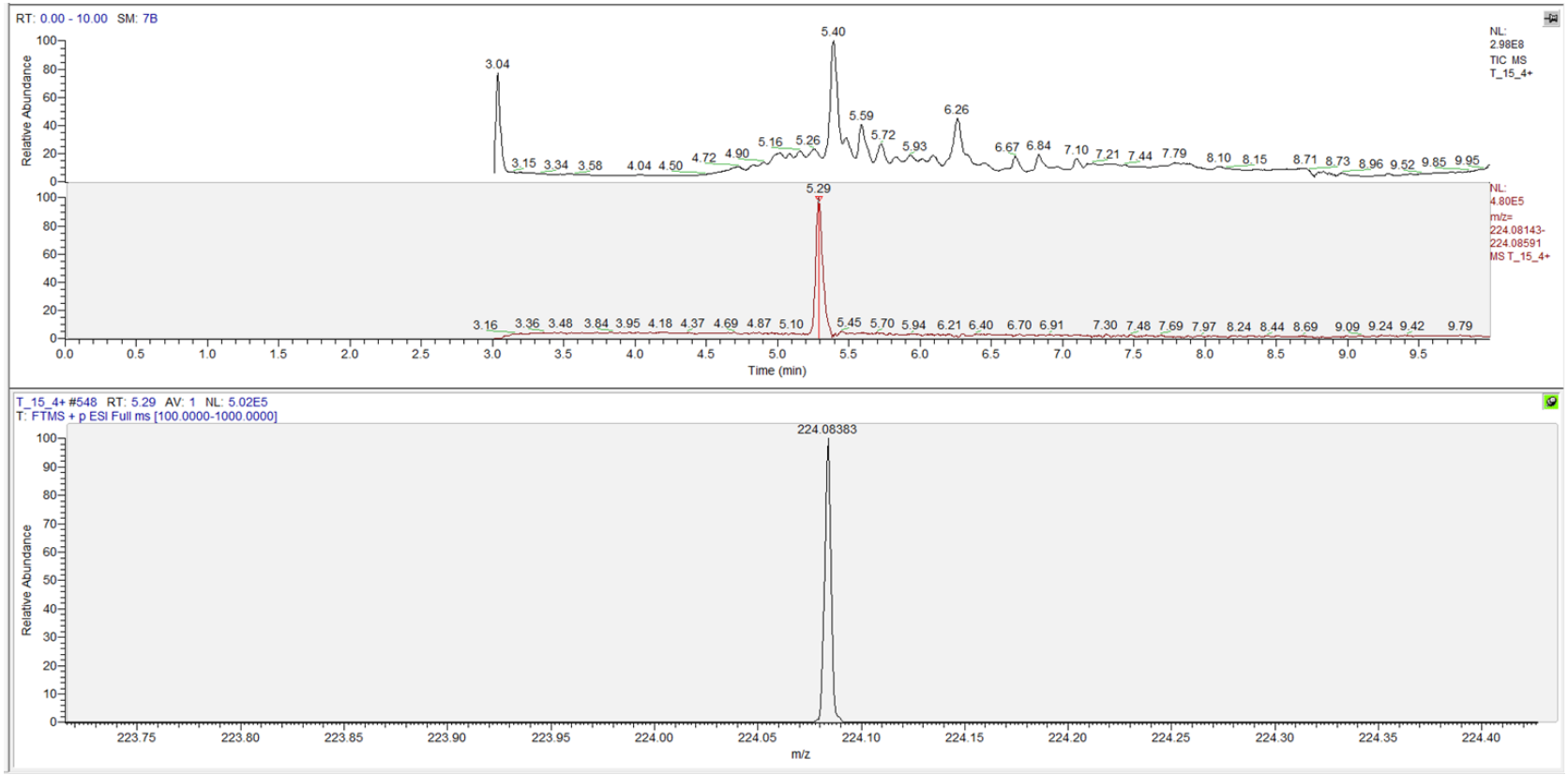


Figure 3-49 LC- MS Chromatogram of 914 -pGEX-6P-1 whole cell activity against ketamine at T15. Top panel is shown the TIC and ketamine peak with retention time 5.40 (black peak), in the second panel is shown the SIM channel of mass norketamine with retention time 5.29 (red peak) and abundance 4.80 x10^5, and in the third panel is shown the high-resolution mass spectrum of norketamine with the exact m/z of 224.08383.

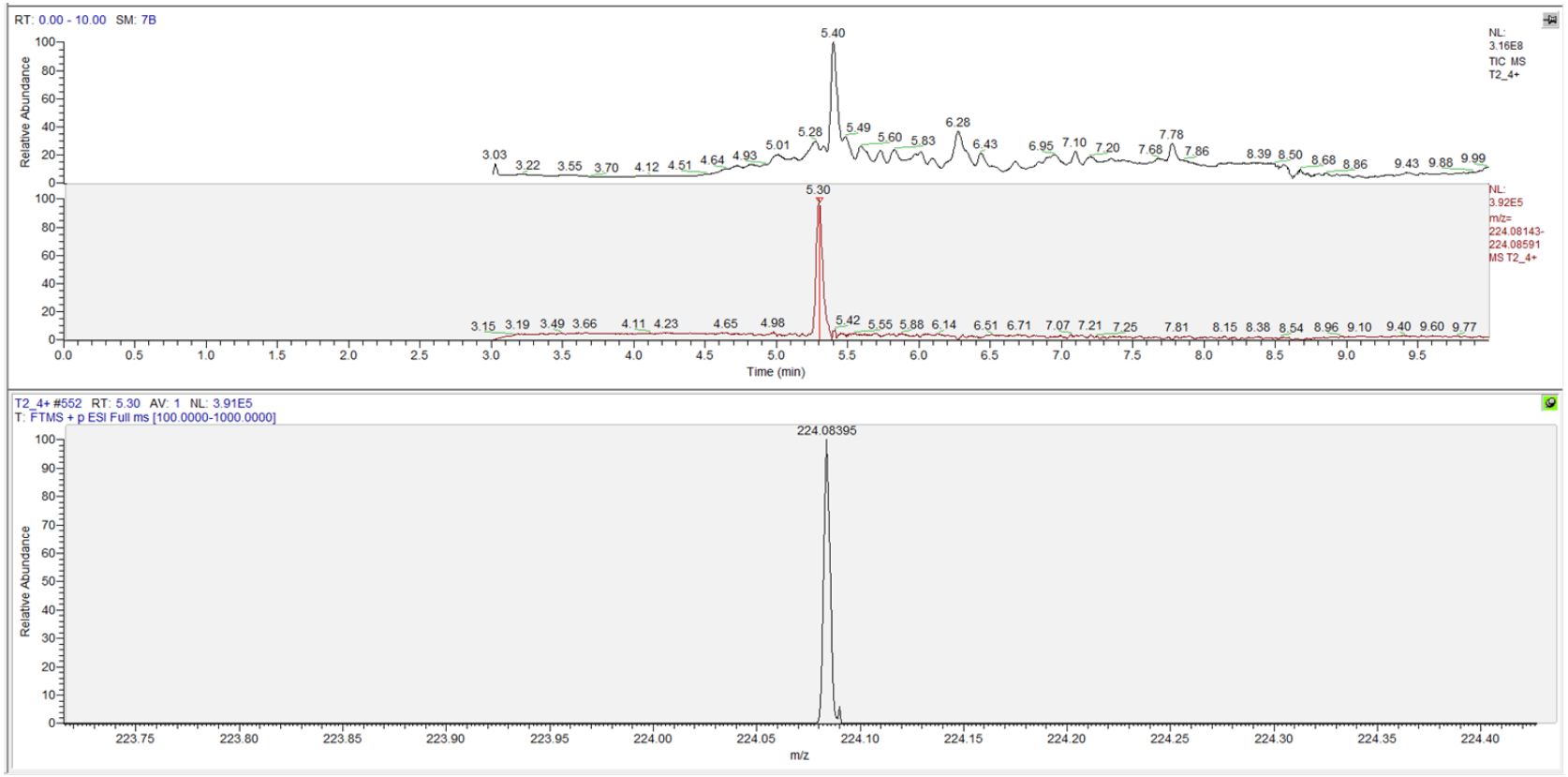


Figure 3-50 LC- MS Chromatogram of 914 -pGEX-6P-1 whole cell activity against ketamine at T2. Top panel is shown the TIC and ketamine peak with retention time 5.40 (black peak), in the second panel is shown the SIM channel of mass norketamine with retention time 5.30 (red peak) and abundance 3.92 x10^5, and in the third panel is shown the high-resolution mass spectrum of SIM of norketamine with the exact m/z of 224.08395.

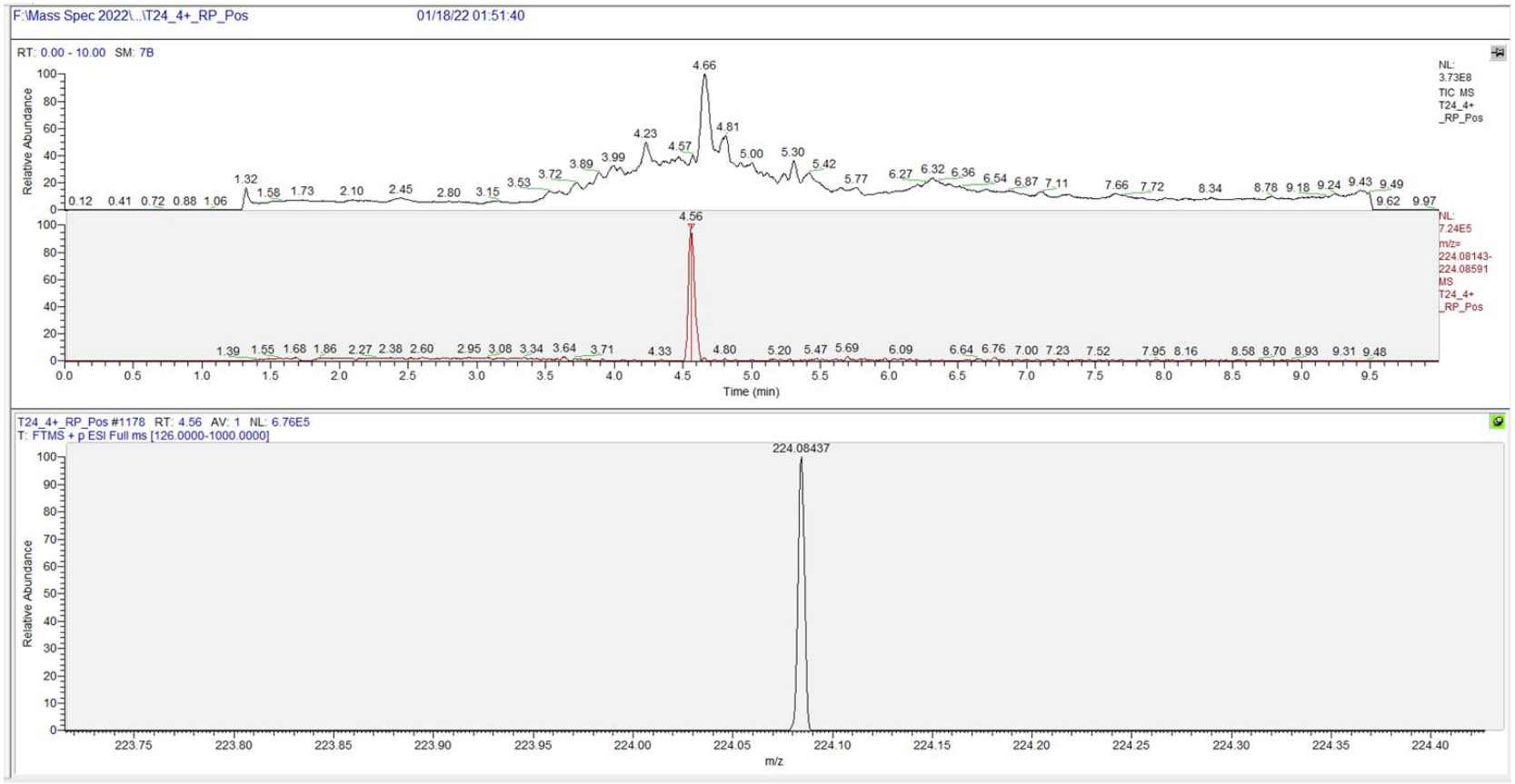


Figure 3-51 LC- MS Chromatogram of 914 -pGEX-6P-1 whole cell activity against ketamine at T24. Top panel is shown the TIC, in the second panel is shown the SIM channel of mass norketamine with retention time 4.56 (red peak) and abundance 7.24 x10^5, and in the third panel is shown the high-resolution mass spectrum of SIM of norketamine with the exact m/z of 224.08437.

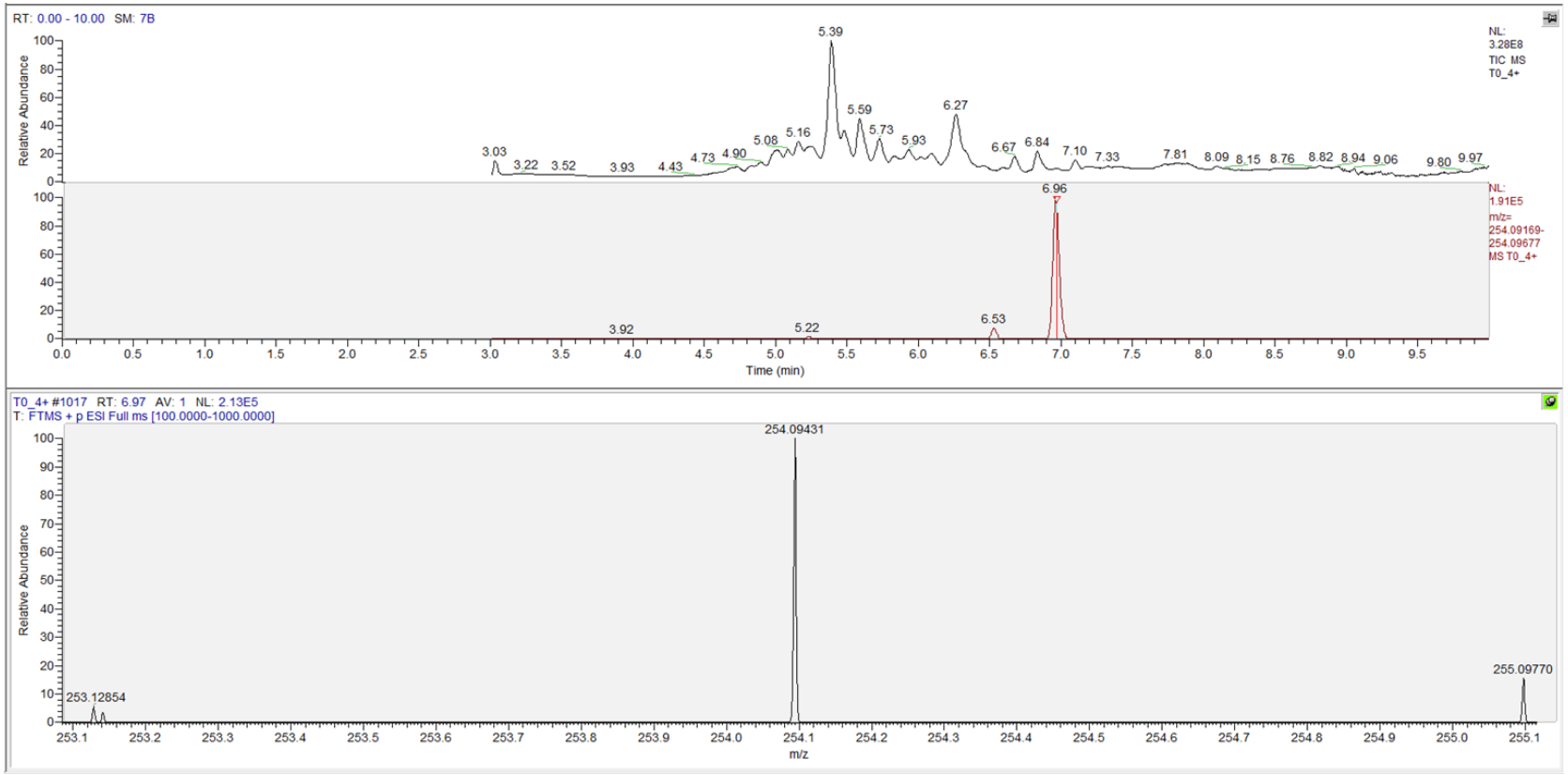


Figure 3-52 LC- MS Chromatogram of 914 -pGEX-6P-1 whole cell activity against ketamine at T0. Top panel is shown the TIC and ketamine peak with retention time 5.39 (black peak), in the second panel is shown the SIM channel of mass hydroxyketamine with retention time 6.96 (red peak) and abundance 1.91 x10^5, and in the third panel is shown the high-resolution mass spectrum of hydroxyketamine with the exact m/z of 254.09431.

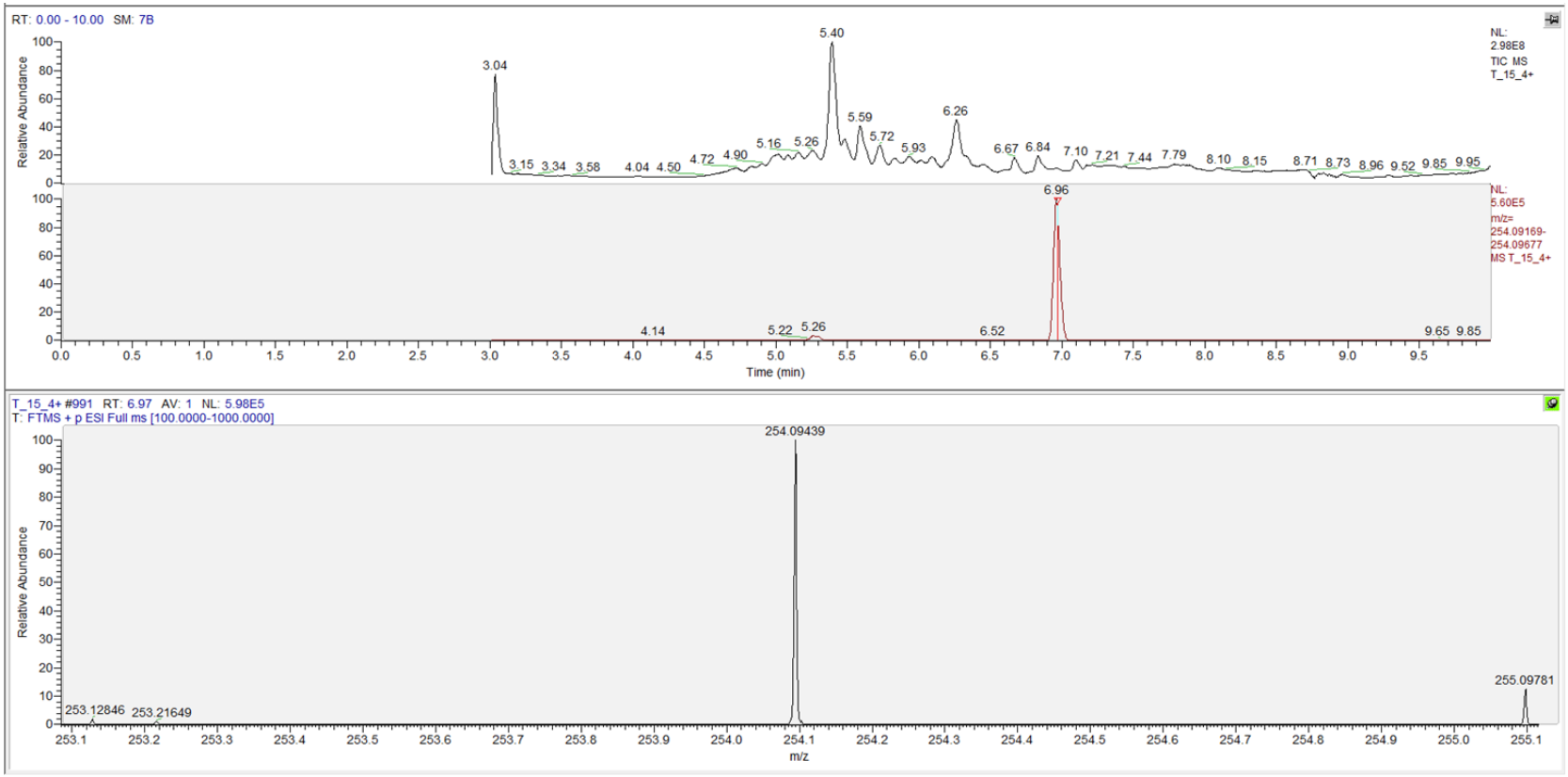


Figure 3-53 LC- MS Chromatogram of 914 -pGEX-6P-1 whole cell activity against ketamine at T15. Top panel is shown the TIC and ketamine peak with retention time 5.40 (black peak), in the second panel is shown the SIM channel of mass hydroxyketamine with retention time 6.96 (red peak) and abundance 5.60 x10^5, and in the third panel is shown the high-resolution mass spectrum of hydroxyketamine with the exact m/z of 254.09439.

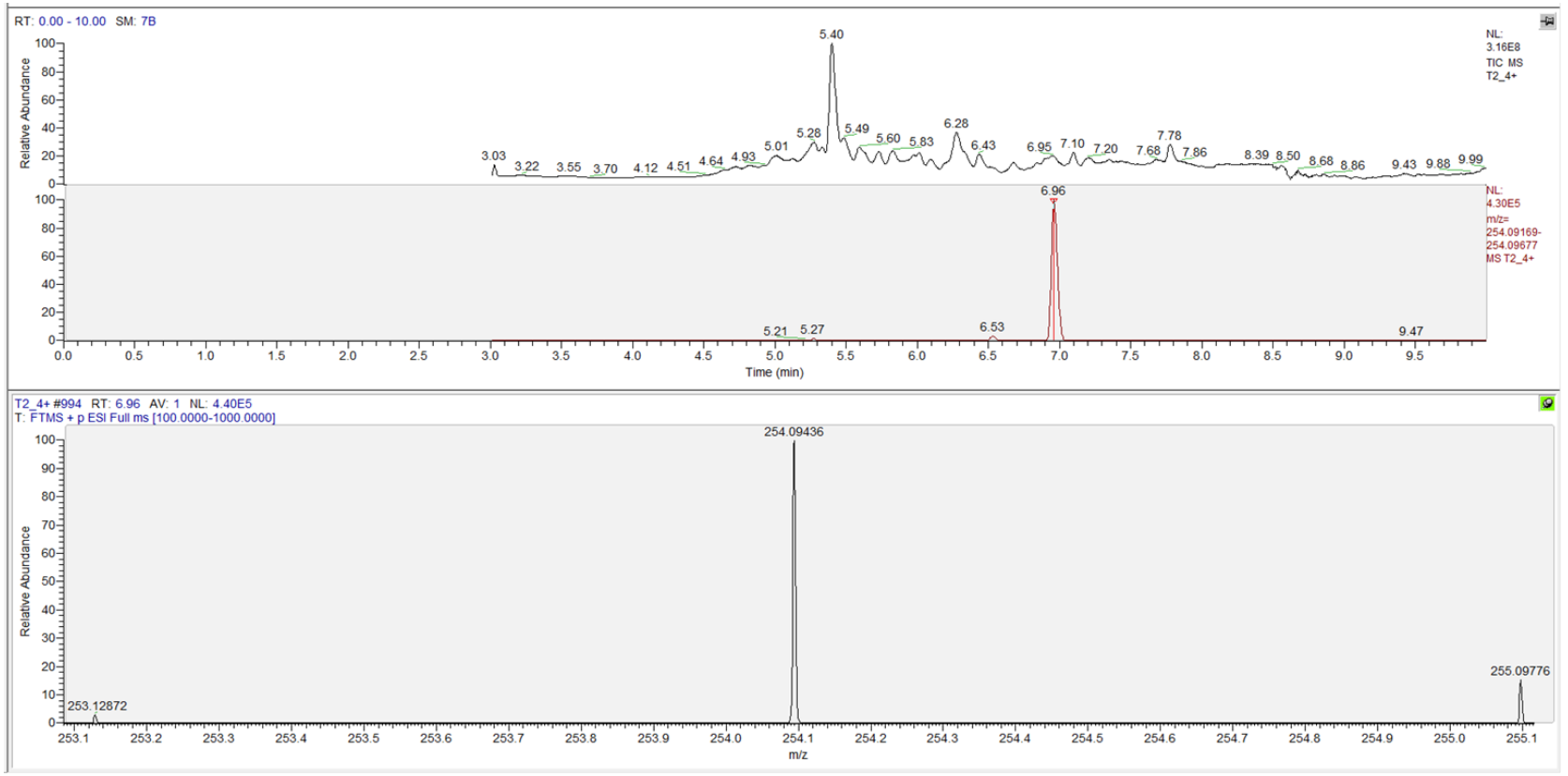


Figure 3-54 LC- MS Chromatogram of 914 -pGEX-6P-1 whole cell activity against ketamine at T2. Top panel is shown the TIC and ketamine peak with retention time 5.40 (black peak), in the second panel is shown the SIM channel of mass hydroxyketamine with retention time 6.96 (red peak) and abundance 4.30 x10^5, and in the third panel is shown the high-resolution mass spectrum of hydroxyketamine with the exact m/z of 254.09436.

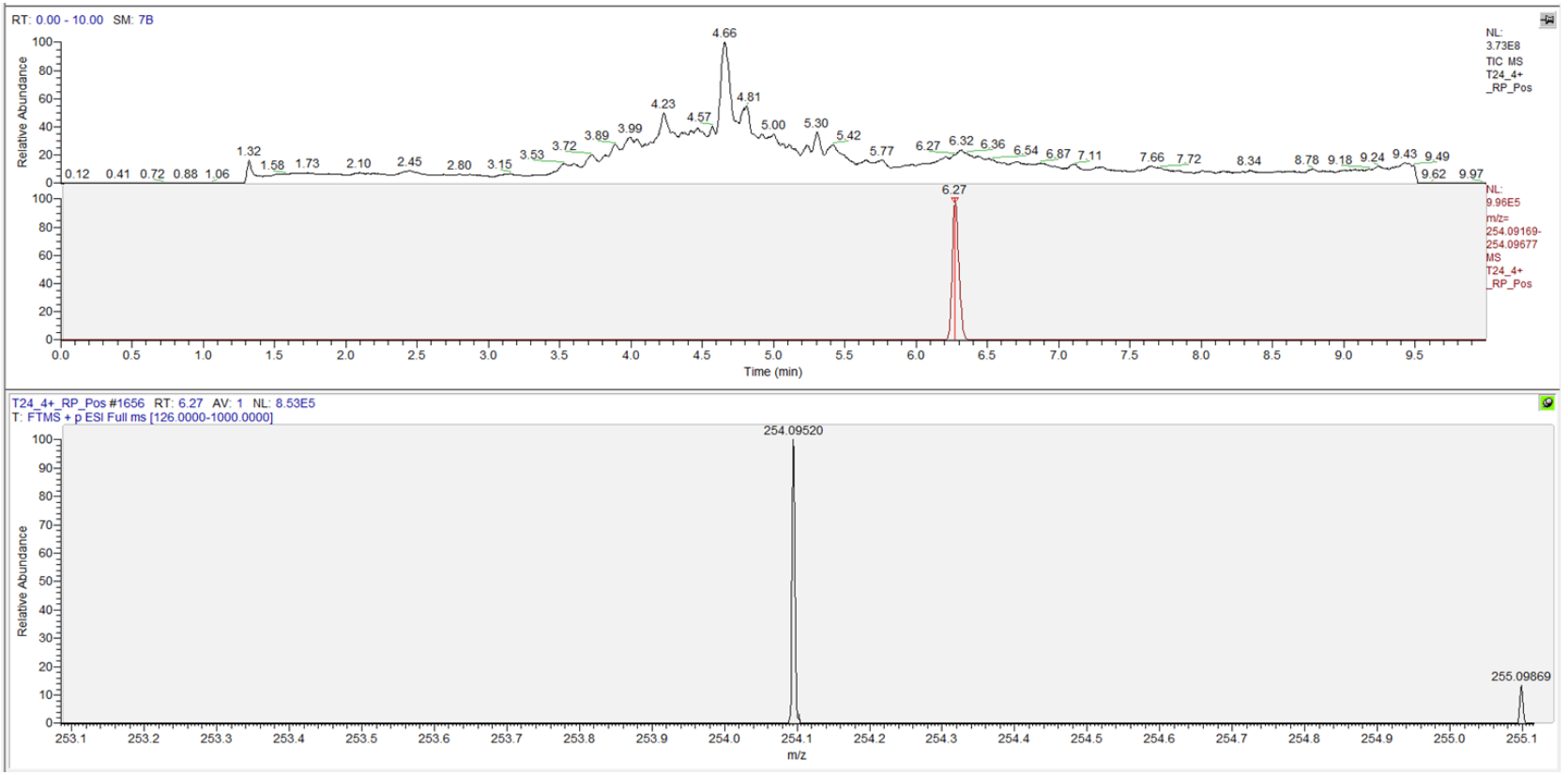


Figure 3-55 LC- MS Chromatogram of 914 -pGEX-6P-1 whole cell activity against ketamine at T24. Top panel is shown the TIC, in the second panel is shown the SIM channel of mass hydroxyketamine with retention time 6.27 (red peak) and abundance 9.96 x10^5, and in the third panel is shown the high-resolution mass spectrum of hydroxyketamine with the exact m/z of 254.09520.

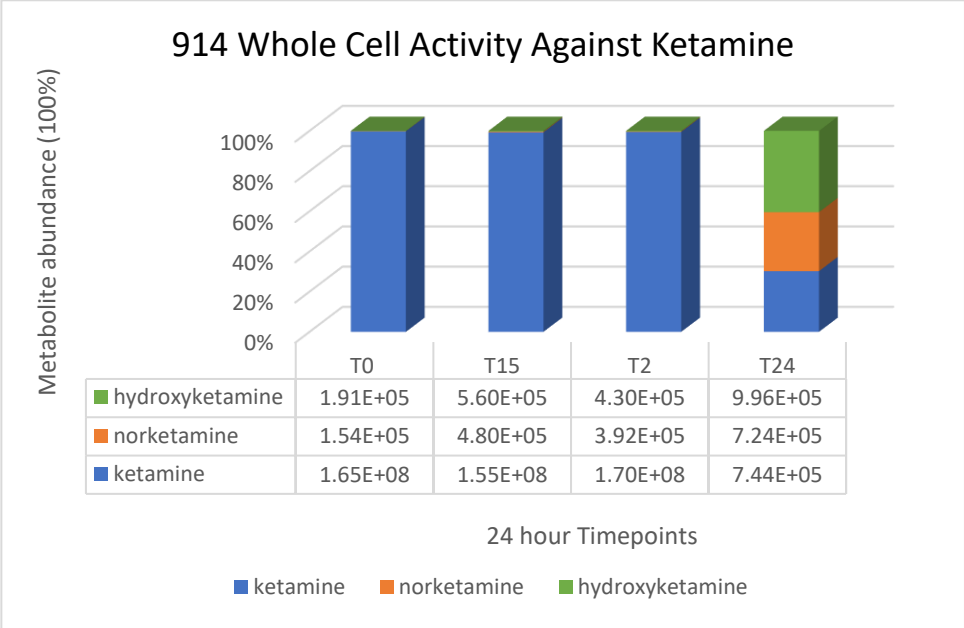


Figure 3-56 Ketamine, norketamine, and hydroxyketamine concentrations throughout a 24-hour period. Figure shows 914 whole cell activity against ketamine, and the concentration (%) of the starting material and concentration and products generated over the 24-hour assay. Ketamine is represented in blue, norketamine in orange an hydroxyketamine in green.

3.11. Analysis of Cloning and Expression Strategies, and Impact on Enzyme Activity

Table 3-8 Summary of all conditions that generated CO-active enzymes. Reference control substrates in Table 3-1

Name	MS Number	Plasmid cloned/expressed	Strain	pGro7	Media	°C	5-ALA (Baeshen et al.)	IPTG (Baeshen et al.)	Control Substrate Active	Norketamine
CYP2A6	3 *	Co-cloned in pET29b (-)	Tuner	-	LB	20	1	0.25	√	-
	5	Co-cloned in pET29b (-)	BL21(DE3)	-	LB	20	1	1	√	-
	2	pET22b, pCDF	C41(DE3)	-	LB	20	1	1	-	-
	1	pET29b, pCDF	Tuner	-	LB	20	1	0.25	√	-
	4	pET29b, pCDF	Tuner	-	LB	30	1	0.5	√	-
	3	pET29b, pCDF	Tuner	+	LB	20	1	0.5	√	-
	4	pET29b, pCDF	Tuner	+	LB	30	1	0.5	√	-
	6	Co-cloned in pET29b (+)	C41(DE3)	+	LB	20	1	1	√	-
	7**	Co-cloned in pET29b (+)	Tuner	+	LB	20	1	0.25	√	-
	8**	Co-cloned in pET29b (-)	Tuner	+	LB	20	1	0.25	√	-
	8	Co-cloned in pET29b (-)	C41(DE3)	+	LB	20	1	1	√	-

CYP2B6	5***	pET22b, pCDF	BL21(DE3)	+	LB	20	1	1	√	-
	7	pET22b, pCDF	BL21(DE3)	+	LB	30	1	1	√	-
	8	pET22b, pCDF	Tuner	+	LB	30	1	0.5	√	-
	6	pET29b, pCDF	Tuner	-	LB	30	1	0.25	√	√
CYP2D6	9	pET22b, pCDF	BL21(DE3)	-	LB	20	1	1	√	-
	18-2	pET22b, pCDF	Tuner	-	LB	20	1	1	√	-
	17	pET22b, pCDF	Tuner	+	LB	20	1	0.25	√	-
	19	pET22b, pCDF	Tuner	+	LB	30	1	0.25	√	-
CYP3A4	16**	Co-cloned in pET29b (-)	Tuner	+	LB	30	1	0.25	√	-
	14	pET22b, pCDF	C41(DE3)	+	LB	20	1	1	√	-
	15	pET22b, pCDF	Tuner	+	LB	30	1	0.25	√	√
CYP3A5	11	Co-cloned in pET29b (+)	Tuner	-	LB	30	1	0.25	√	-
	11	pET22b, pCDF	BL21(DE3)	+	LB	30	1	1	√	-
	9-1	pET22b, pCDF	Tuner	-	LB	20	1	0.5	√	-
Meta 198	21	pET29b	Tuner	-	LB	20	1	0.5	-	-
	24	pET29b	Tuner	-	LB	30	1	1	-	-
	4	pGEX-6p-1	BL21(DE3)	+	LB	20	1	1	√	-
	25	pGEX-6p-1	Tuner	-	LB	20	1	0.5	√	-
	26	pGEX-6p-1	Tuner	-	LB	20	1	1	√	-
	25-3●	pGEX-6p-1	Tuner	+	LB	20	1	0.5	√	-

Meta 914 MALLLAVF	26-4	pGEX-6p-1	Tuner	+	LB	20	1	1	-	-
	25	pGEX-6p-1	Tuner	+	LB	20	1	0.25	√	-
Meta 914	26	pGEX-6p-1	Tuner	+	LB	30	1	0.25	√	-
No MALLLAVF	27	pGEX-6p-1	Tuner	+	LB	20	1	0.5	√	-
	28	pGEX-6p-1	Tuner	+	LB	30	1	0.5	√	-
	24!	pGEX-6p-1	Tuner	-	LB	20	1	0.25	√	-

Ketamine and norketamine were used as test substrates for both metagenome enzymes, even though there was no previous known catalytic activity.

*Ketamine was used as probe substrate, which was metabolised at timepoint 24 hours to hydroxynorketamine and N-demethylated to norketamine.

** Ketamine only N-demethylated to norketamine, but not metabolised to hydroxyketamine!

*** Ketamine was metabolised to hydroxyketamine and N-demethylated to Norketamine at timepoint 0 hours, at timepoint 15 minutes only N-demethylated to Norketamine, and at timepoint 3 and 24 hours no visible metabolites were detected!

● Highest catalytically active enzyme produced from metagenome gene 914-pGEX-pGro7 induced with 0.5 mM IPTG in LB 20 °C.

! The only conditions that generated catalytically active and at the same time and soluble 914 enzyme.

In Table 3-8 are presented all the cloning plasmids, strains and conditions of all the target genes that were solubly expressed and bound CO, and the ones that were active against the test substrate. From the co-cloned assay only one enzyme expressed in BL21(DE3), CYP2A6 CPR-pET29b (-) without pGro7 in LB 20 °C. When the media were supplemented with ferric and ferrous chloride, and the concentration of the 5-ALA was reduced to 0.25 mM only two enzymes expressed again in BL21(DE3) without pGro7 in different conditions: CYP3A4-CPR-pCWOri^{+} (+) in LB 30 °C, CYP3A4-CPR-pET29b (+) in LB 20 °C, CYP3A4-CPR-pET29b (-) in LB 20 °C, CYP3A5-CPR-pCWOri^{+} (+) in LB 20 °C. Additionally, two different enzymes expressed again without pGro7 in C41(DE3), in different temperatures: CYP2B6-CPR-pET29b (+) in LB 20 °C, CYP2B6-CPR-pET29b(-) in LB 20 °C, CYP3A5-CPR-pET29b(+) in LB 30 °C and CYP3A5-CPR-pET29b(-) in LB 30 °C. In contrary, when expressed in Tuner cells and induced with 0.25 mM IPTG, all enzymes expressed in different temperature and chaperone combinations apart from CYP2D6.

From the co-expression assays of CYP-pET22b and CPR-pCDFDuet-1 good expression was obtained in BL21(DE3) for genes CYP2A6, CYP2B6, CYP2D6, and CPR with and without pGro7 in both temperatures, although CO-active enzymes were only CYP2B6, CYP3A5, and CYP2D6, but not CYP2A6. When co-expressed with pGro7 in C41(DE3) CO-active enzyme was achieved for CYP2A6 and CYP3A4, but CPR did not express, which impeded any catalytic activity. However, better /or expression was achieved in BL21(DE3) with pGro7 for P450s, rather than in C41(DE3) with or without pGro7. In contrary, when expressed in Tuner cells, again all target enzymes expressed at different IPTG concentrations, temperatures, and chaperone combinations.

From co-expression assay of CYP-pET29b and CPR-pCDF in Tuner cells, only two enzymes expressed. CYP2A6 in LB 20 °C induced with 0.25 mM IPTG, and in LB 30 °C induced with 0.5 mM IPTG, and CYP2B6 in LB 30 °C when induced with 0.25 mM IPTG.

As can be seen from the expressed enzymes in Table 3-8 different conditions produced CO-binding active enzymes. All CYPs had different requirements for generating CO-active enzyme, such as temperature, strain and IPTG concentrations, and the co-expression with or without the chaperones. All enzymes were active against the probe substrates, apart from only one (CYP2A6- pET22b when co-expressed with CPR-pCDFDuet-1 in LB 20 °C and induced with 1mM IPTG).

Metagenome enzyme 198 was active only when the modification MALLLAVF was applied at the N-terminus, and co-expressed with pGro7 in Tuner cells, in LB 20 °C and 30 °C induced with 0.5 mM and 1 mM IPTG. In contrast, metagenome enzyme 914 produced active enzyme with and without the MALLLAVF modification, with the highest catalytic activity achieved with the MALLLAVF modification when co-expressed with pGro7 in Tuner cells, in LB 20 °C and induced with 0.5 mM IPTG. In contrast, soluble enzyme was achieved for metagenome enzyme 914 only without the MALLLAVF modification, when expressed with and without pGro7 in LB 20 °C and 30 °C, and induced with 0.25 mM and 0.5 mM IPTG, of which soluble and catalytically active enzyme was produced when expressed in Tuner cells without pGro7 in LB 20 °C and induced with 0.25 mM IPTG. All the above suggests that there is no silver bullet when cloning and expressing human and metagenome P450s, as to which method is going to work best, and produce active enzyme. Moreover, Table 3-8 proves once again that all enzymes have different requirements for media, temperatures, cloning plasmids, and supplementations of the media, as well as chaperone co-expression.

Generally, complementing the expression with the chaperones and inducing with low levels of IPTG when expressing in Tuner cells can generate active enzyme according to our experiments summarised in the above table.

Metagenome enzyme 914 metabolised ketamine at timepoint 0 hours to hydroxyketamine, and N-demethylated it to norketamine. This metagenome gene is a homolog of CYP3A5 and shares 24% sequence identity. A protein structure prediction was proposed by Phyre2 based on partial fragment of already known protein structures using PSI-Blast and alignment between the predicted Hidden Markov Model (HMM) of the query and already known protein structures. Metagenome enzyme 914 aligned against the self-sufficient P450BM3 from *Bacillus megaterium* (CYP102A1, PDB: 1bu7), with 100 % confidence and 46 % identity, and with SiRFP-60 (PDB: 6efv) with 100 % confidence, and 31% identity, and 100 % confidence and 30% identity with CYPOR (PDB: 6j7a). These suggest that 914 metagenome enzyme is a cytochrome P450.

In contrast, metagenome enzyme 198 does not have any metabolic activity against ketamine or norketamine, although this metagenome enzyme shares 20 % sequence identity with CYP3A5. Instead, a structure prediction was performed from aligning the HMM of the query sequence and already known protein structures. 94 % of the structure was modelled against CYP97A3 (PDB: 6j94) from plant organism *Arabidopsis thaliana* with 100% confidence and 30% structure identity. This prediction suggests that 198 is again a cytochrome P450 with the possibility to

catalyse the hydroxylation of the β- and ε-rings of alpha-carotene and produce lutein, based on the activity of the template enzyme. Due to shortage of time, no further characterisation assays were performed for this enzyme, but we are aiming to conclude these experiments in the future.

3.12. Metagenome Genes Structure Prediction Data

A structure prediction was performed for metagenome enzymes 914 and 198. This was accomplished to find the closest protein structure to both our proteins, but also to determine if the catalytic residues from any of the known template structures were supported with the same amino acids in the predicted model. Proteins that share similar structures have often similar functions, and when they share percentage identity (i.d %) that is more than expected by chance for random proteins, most likely they originate from common ancestors (Pearson and Sierk, 2005). The only way to establish this would be by generating the crystal structure of the query protein, and site directed mutagenesis on the suspected catalytic residues.

A FASTA format of the proteins was submitted to Phyre2 server for structure prediction, and a set of templates were generated organised from the highest alignment coverage, confidence of alignment and percentage identity to the query sequence. The 3D models of templates with the highest alignment coverage, confidence and percentage identity were used to align against the predicted structure and were visualised using PyMOL.

3.12.1. 914 Structure Prediction

The 914 structure was predicted based on 9 templates (PDB: 5gxu, 3hf2, 1tll, 1j9z, 6j7a, 2ij2a, 7f3w, 6efv and 2bpo) based on heuristics to maximise confidence, percentage identity and alignment coverage, while 20 residues were modelled ab initio.

Metagenome enzyme 914 aligned with the highest homology and percentage identity at the N-terminus with 2ij2 with 100 % confidence and 46 % identity (11-462 aa of sequence aligned). 2ij2 is a mutant of the self-sufficient P450BM3 from *Bacillus megaterium* CYP102A1 (PDB: 1bu7). Predicted enzyme structure 914 was superimposed with 1bu7, and very little difference is observed with Cα root-mean-square deviation (RMSD) Ca 0.819 Å. Since 914 aligns only at the N-terminus with 1bu7, some section of 914 structure is not superimposed (Figure 3-57).

Figure 3-57 Cartoon view of superimposed structures 914 (in deep purple) and 1bu7 (in cyan). PDB: 1bu7 (CYP102A) from *Bacillus megaterium*, shown. The protoporphyrin ring is shown in magenta with the iron ion in blue (nitrogen atoms) and orange (iron ion), in the centre of the superimposed structures. Figure is created with PyMol 2.5.4 (https://pymol.org/2/).

At C-terminus 914 aligned with 100 % confidence, and 31 % identity with 6efv (PDB: SiRFP-60) (residues 491-1063, hence large sections of 914 and 6efv structures have not superimposed) (Figure 3-58), and 30 % identity with 6j7a (PDB: CYPOR) (residues 488-1064) (Figure 3-59). For robust alignment the identity value should be higher that 30-40 %, although percentage identity less than 15 % with 100 % confidence can generally mean that the structure predicted is correct (http://www.sbg.bio.ic.ac.uk/~phyre2/html/page.cgi?id=index).

Structural differences and deviations were observed when 914 was superimposed to 6efv, and 6j7a, with Ca RMSD 4.035 Å and 5.03 Å, respectively, although RMSD cannot be reliable for structure pairs of different sizes, as RMSD measures the distance (d) of the n pair of atoms in two superimposed structures (Carugo and Pongor, 2001), and there is a size difference between the 914 and 6efv structures. 6j7a is a characterised fusion protein of haem oxygenase-1 and NADPH cytochrome P450 reductase, while 6efv is an NADPH-dependent sulphite reductase flavoprotein (Tavolieri et al., 2019, Sugishima et al., 2019).

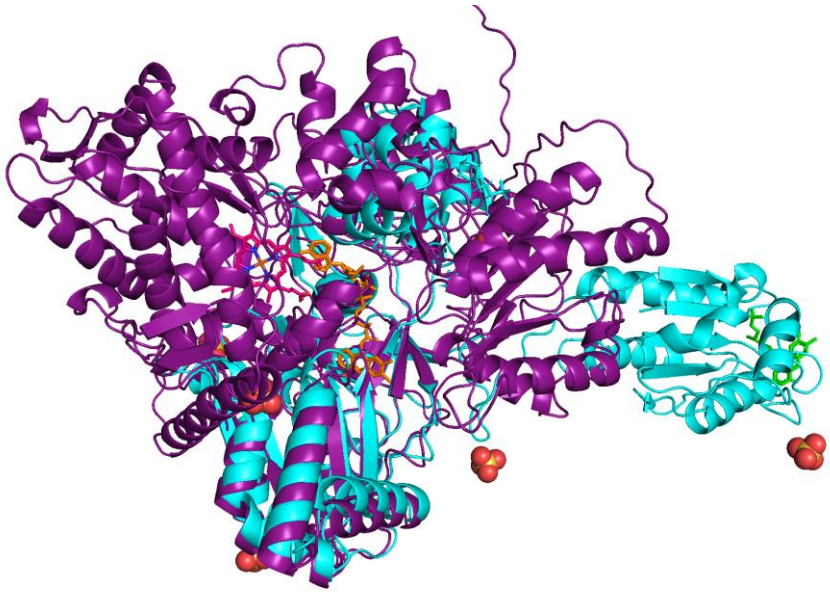


Figure 3-58 Cartoon view of superimposed structures of 914 (in deep purple) and 6efv (in cyan). PDB: 6efv -SiRFP-60 from *Escherichia coli*. The porphyrin ring is shown in magenta with the iron ion in the centre in blue (nitrogen atoms) and orange (iron ion), FAD domain is shown in orange, while FMN domain is coloured in green, on the far right. The orange and yellow spheres are SO_4 ions. Figure is created with PyMol 2.5.4 (https://pymol.org/2/).

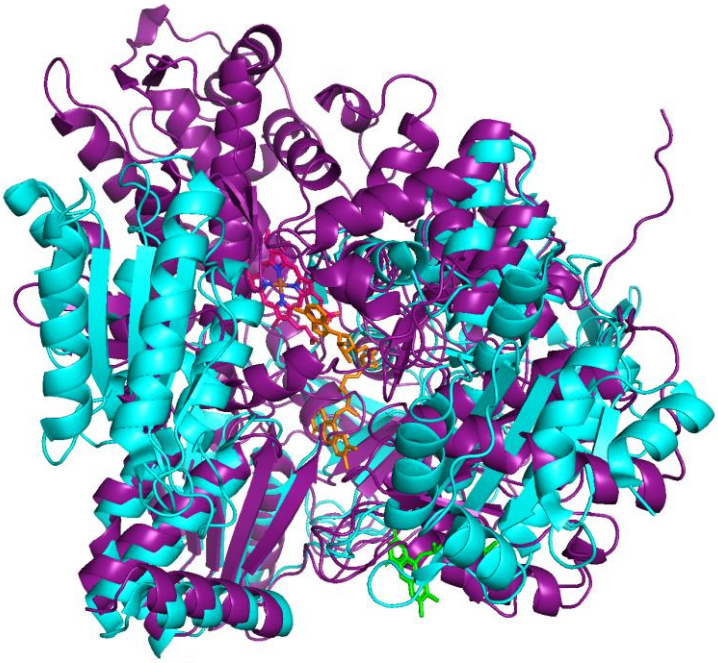


Figure 3-59 Cartoon view of superimposed structures of 914 (in deep purple) and 6j7a (in cyan). PDB: 6j7a-Fusion protein of haem oxygenase-1 and NADPH cytochrome P450 reductase (17aa). The haem is shown in magenta with the iron ion in the centre blue (nitrogen atoms) and orange (iron ion), FAD domain in coloured orange, just below haem, and FMN domain shown in green. 914 aligns with 6j7a at the C-terminus, hence a part is not aligned. Figure is created with PyMol 2.5.4 (https://pymol.org/2/).

The P450 BM3 model, a high-activity fatty acid hydroxylase in which the P450 is fused to its redox partner, is arguably the most significant P450 model. Residues Arg47 and Tyr51 are of great importance in P450BM3, as they interact with the substrate, while Phe87 is well documented to experience a significant shift upon substrate binding and to play a crucial role in regulating the regional specificity of substrate oxidation in BM3. At the opening of the P450's binding pocket is Phe42 which forms a cap over the active site. According to mutation studies, arginine at position 47 is very crucial for interacting with the substrate in an ion pair manner according to mutation studies (NOBLE et al., 1999). Further mutation studies suggested that both Arg47 and Tyr51 are responsible for enticing substrates to the access channel opening in addition to serving as a binding site, as double substitution of Arg47Ala and Tyr51Phe residues ended in reduced catalytic activity (Ost et al., 2000).

After superimposing both 914 and 1bu7 structures, we observed that Arg47 is replaced in 914 structure by another positively charged residue of the same group, Lys55, and Tyr51 is replaced in the 914 structural space by another hydrophobic residue of the same group, Phe59, while Phe87 aligns with Phe95 in 914 (Figure 3-60). In 2012, Whitehouse and colleagues (2012) reported that CYP102A was the only fatty acid hydroxylase in the subfamily to have a positively charged residue in position 47, and polar residue in position 51. Moreover, they suggested that Tyr51 is conserved in microsomal fatty acid hydroxylases, whereas Arg47 is not, while neither residue is conserved in P450foxy which has the same active site residues as BM3, suggesting that sub-family members may incorporate substrate at different locations (Whitehouse et al., 2012a).

Moreover, two conserved residues are found in BM3, Glu267 and Thr268. Mutations on these resulted in increased uncoupling levels. Threonine's side chain positions a water molecule at the active site (Cryle and De Voss, 2008, Whitehouse et al., 2012a). Glu267 align in 914's structure with Glu275 and Thr268 with Thr276.

Ala264 is another conserved residue, and in almost all the crystal structures that have been reported to date, the carbonyl oxygen of Ala264, which is a component of the I-helix, is hydrogen-bonded to the water molecule that is closest to the haem iron. Ala264 aligns in our structure with Ala272. Additionally, Cys400 in BM3 is the conserved cysteine that binds to haem in the P450 family, which aligns in 914 with Cys410.

Moreover, mammalian isoforms often have insertions at the N-terminus end and in the F/G and/or G/H loops associated to the CYP102A subfamily. In fact, two residues, Asn381 and Ala388, that are present to eukaryotic P450s and absent from other bacterial P450s, were discovered in the BM3. These residues are believed to be involved into inter-domain binding (Whitehouse et al., 2012a). Ala388 aligns well in our structure with Ala399, and Apn381 is replaced in our structure with Asn391 (Figure 3-60). The evidence strengthens the idea that 914 might be a member of the self-sufficient family.

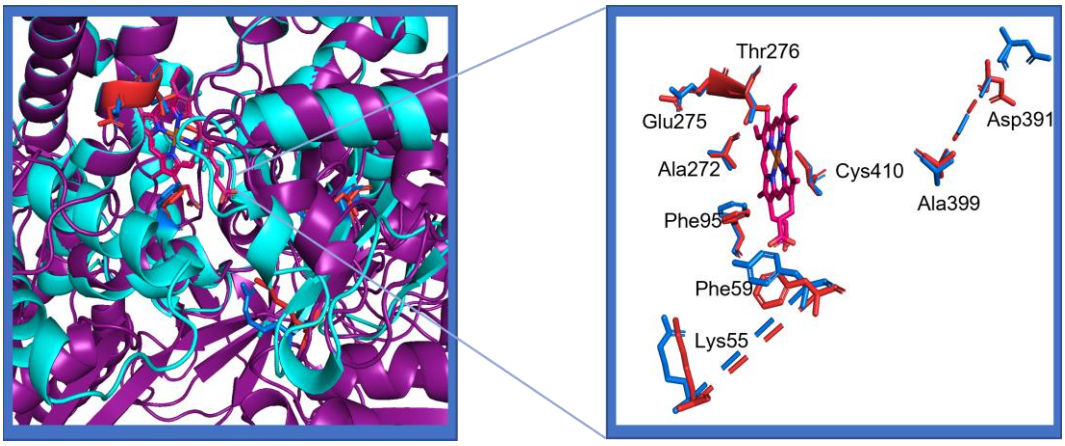


Figure 3-60 Catalytic residues of 914 aligned from P450BM3. On the left figure is presented a cartoon view of superimposed 914 structure (deep purple) and 1BU7 (PDB: CYP102A) from *Bacillus megaterium (*in cyan). The porphyrin ring is shown in magenta, P450BM3 catalytic residues are shown in blue, and predicted catalytic residues from 914 are in red. On the right figure are seen all catalytic residues from 914 and 1bu7 aligned in space, with the haem in plane, and the conserved cysteine residue found in the proximal side of the haem. Figure is created with PyMol 2.5.4 (https://pymol.org/2/).

Members of the CYP102A sub-family proteins, including CYP102A2 and A3 from *B. subtilis*, A5 from *B. cereus*, and A7 from *B. licheniformis*, have been expressed and characterised. According to an alignment of the A1 to A15 enzymes within the CYP102A subfamily, the C-helix and B'/C loop, the E-helix and most of the E/F loop, a large portion of the I-helix and the b5-sheet that comes before it, and the region on the N-terminus side together with the proximal loop are all highly conserved regions (residues 391-402) (Whitehouse et al., 2012a). This highly conserved region aligns well in our 914's structure with residues 410-412, and it is at the opening of the binding pocket.

Moreover, five out of six highly conserved regions across all P450s, named substrate recognition sites (SRSs) are highly conserved in BM3 as well. SRS4 and SRS6 (residues 253-271, and 434-441, respectively) are highly conserved, SRS1 (residues 69-92) is conserved only at the C-terminus, whereas the N-terminus displays dissimilarities, SRS2 and SRS5 (residues 181-188, and 325-335, respectively) both have intriguing peculiarities, while SRS3 (residues 200-208) is not conserved at all in BM3 (Whitehouse et al., 2012a).

After superimposition of BM3 structure to our 914, was observed that SRS4 (266-279), SRS6 (444-451), SRS1 (77-100), SRS5 (334-344) align completely, SRS3 (208-216) almost aligns as at the N-terminus two residues fold into a loop instead of a barrel, and SRS2 does not align at all as is a loop that forms a different fold in our structure, but beta barrel in BM3 (Figure 3-52).

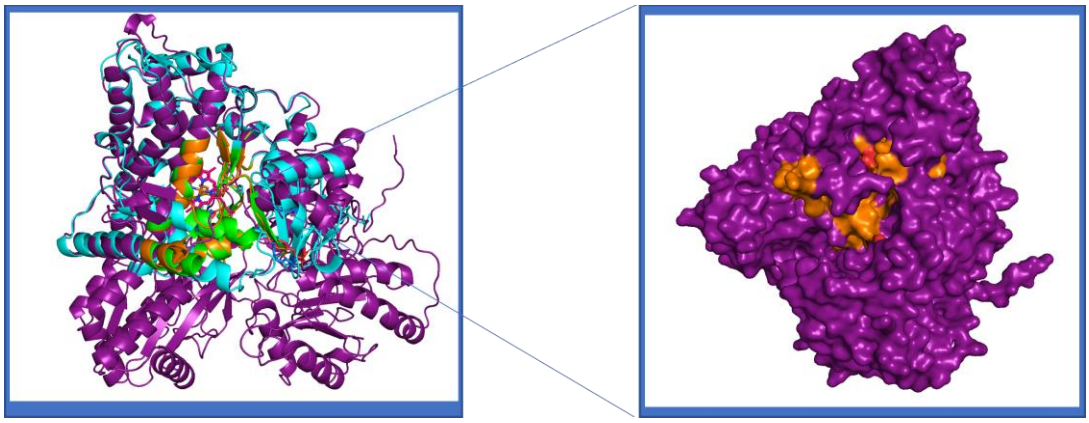


Figure 3-61 Conserved substrate recognition sites. On the left figure are seen superimposed structures of 914 (in purple) with substrate recognition sites in orange, with P450 BM3 (CYP102A) (in cyan) with substrate recognition sites in green. On the right is shown the surface of 914 with the conserved substrate recognition regions in orange, which are found in and around the catalytic pocket. Also, some of the catalytic residues can be seen highlighted in red (from 914) in both figures, while in P450BM3 (in blue) only on the left figure. Figure is created with PyMol 2.5.4 (https://pymol.org/2/)

All of the aforementioned evidence points to gene 914 possibly being a member of a self-sufficient P450 that is connected to the P450BM3. We performed a Blastp alignment (NCBI), metagenome enzyme 914 exhibited 98.9 % identity with CYP120A1 from organism *Cupriavidus metallidurans* strain CH34 and hence can be assigned to P450 family 120 and subfamily A (or CYP120A), while the number of the subfamily member remains to be assigned once fully characterised.

Two classes of self- sufficient P450s have been discovered, such as CYP102A from *Bacillus megaterium (*P450BM3), where the N- terminus domain is fused to the C-terminus NADPH- depended on cytochrome P450 reductase containing the flavin mononucleotide (FMN) binding flavodoxin and a FAD- binding NADPH domain (see Class VIII at introduction). The electron flows from the NADPH to FAD to FMN and to the haem of the P450. The other class of self-sufficient P450s CYP116 contain an N-terminus fused to the reductase domain of the type of phthalate dioxygenase which contains the NADPH-reductase domain and the ferredoxin 2Fe-

2S cluster domain at the C-terminus (see Class VII at introduction). The electron transfer in this type of self-sufficient P450 flows from NADPH to the FMN to the 2Fe-2S to the haem (Zhang et al., 2020).

Structure prediction and intramolecular transfer system of self-sufficient P450 enzyme classes is of great importance, especially since these classes are being used in biocatalysis and chemical production. Their importance relies on being self-sufficient enzymes equipped already with the required redox partners to perform the electron transfer for catalytic activity and substrate reduction.

There are different suggestions as to how the electron transfer is facilitated in the self-sufficient species. For BM3 has been suggested that it either has a CPR domain movement that can accelerate the electron flow from FMN to the haem (Barnaba et al., 2017), or alternatively that BM3 functions as homodimer where the cofactors can facilitate electron transfer from the centres of the same or neighbouring domains (Kitazume et al., 2007).

Several reductase domains of either the BM3 or that of P450 RhF from Rhodococcus species have been artificially fused to other P450s to produce a range of important chemicals, which enhances their catalytic activity, and emphasises their importance in biocatalytic applications. P450 BM3 has been very extensively studied due to the above reason as well as for its catalytic efficiency. P450 BM3 catalyses the hydroxylation of fatty acids at speeds up to 17,000 min^{-1} (or 280s^{-1}) which is twice as fast as the observe rates for the eukaryotic P450s. The fusion system affords P450 BM3 with the fastest rate of electron transfer from flavin to haem, which saves on electron waste while delivering considerable amounts of hydroxylated product (NOBLE et al., 1999). There are suggestions that more self-sufficient enzymes fused to their redox partners exist. The CYP102A1 gene has more than fifty homologues in bacteria (CYP102) and fungi (CYP505), according to microbial genome databases (Whitehouse et al., 2012a) .

Based on the above, discovering a new enzyme belonging to the same class of the self-sufficient system may be of benefit to the biocatalysis community, and potentially in the drug and chemical hydroxylation.

3.12.2. 198 Structure Prediction

The 198 structure was predicted based on 6 templates (PDB: 3el3, 3dbg, 2ve3, 5t6q, 6j94 and 2q9f) based on heuristics to maximise confidence, percentage identity and alignment coverage. The structure with the highest percentage identity was that of CYP97A3 (PDB: 6j94) which was aligned with 100 % confidence and share 30 % sequence identity, and 94 % sequence coverage (residues 25-441). The root square mean value was 1.404 Å, which suggests that the structures are very similar, and can be observed from the alignment in Figure 3-62.

The CYP monooxygenase CYP97A3 from the organism *Arabidopsis thaliana* catalyses the hydroxylation of the β- and ε- rings of α-carotene and produces lutein (Niu et al., 2020). The members of the CYP97A family resemble the mitochondrial CYPs, as they utilise the same ferredoxin-$NADP^+$ reductase mechanism (FNR-ferredoxin) for electron transport. Moreover, CYP97A3 is peripherally associated with the membrane through its hydrophobic regions and has specificity towards 3R hydroxylation. All the residues that facilitate this specificity, some of which are conserved as well in our structure, are the conserved oxygen activating Thr381 which aligns in 198 structure with Thr261, and Ala377 which align in 198 with Ala257 in the I helix (Figure 3-63). Thr555 in the β4 sheet is replaced in the 198's structure with hydrophobic residue Val429, and Ile190 is replaced in 198 by Leu80, which are located in the B/C loop. The conserved haem binding residue Cys516 in CYP97, aligns with Cys391 in 198. Additionally, the pro-3S side of the cyclohexene ring contains a cis proline which is located at position 441, which is replaced in the 198's structure with a hydrophobic residue, Ala319 (Niu et al., 2020).

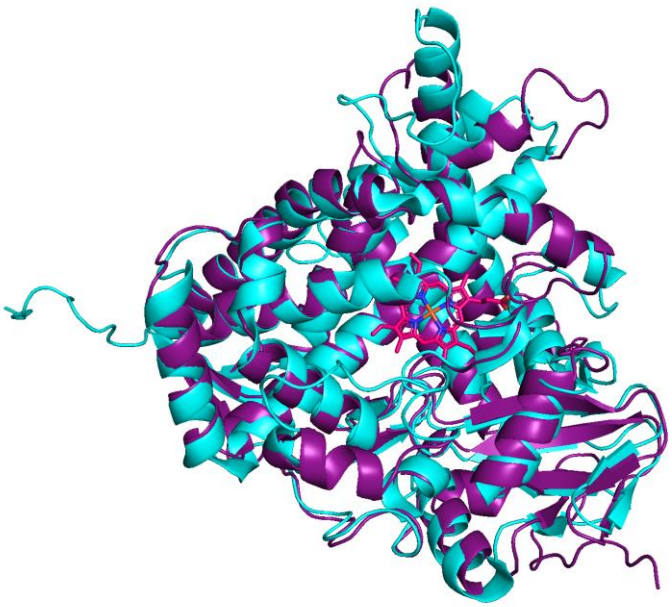


Figure 3-62 Cartoon view of superimposed structures 198 (deep purple) and 6j94 (cyan). PDB: 6j94 (CYP97A) an ε-ring hydroxylase of α-carotene to lutein from *Arabidopsis thaliana*. Haem is shown in magenta with the iron ion in the centre in blue and orange. The structures have a root square mean distance of 1.404 Å, meaning that the structures are very similar, which can be observed from the alignment as well. Figure created with PyMol 2.5.4 (https://pymol.org/2/).

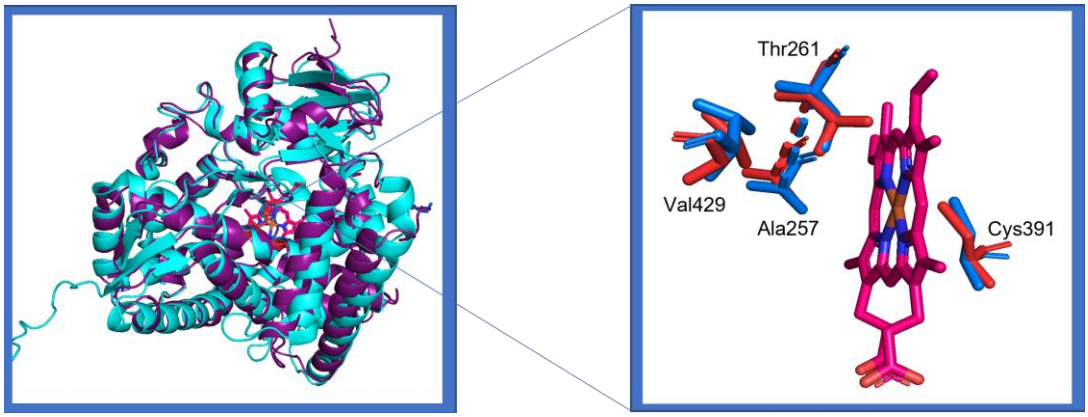


Figure 3-63 Catalytic residues of 198 as aligned from CYP97A. Cartoon view of superimposed structures of 198 (in deep purple) and 6j94 in (cyan). Haem is shown in magenta with the iron ion in the centre in blue and orange. Cys391 (Cys516) is shown in the proximal side of the haem, whereas on the distal side of the haem ring are superimposed residues Thr261 (Thr381) and Arg257 (Arg377). In parenthesis are the residues as found in CYP97A. Figure created with PyMol 2.5.4 (https://pymol.org/2/).

Additionally, the B′ helix residues Ile176, Leu177, Ile180, and Leu181, the F/G loop Ile292, the β1-sheet Val443, Ile445, and Phe466, and the β4-sheet Ala554 all contribute to the formation of the hydrophobic pocket of CY97A (Niu et al., 2020).

Ile180 aligns with Leu75, Leu181 aligns with Leu76, Ile292 aligns with Phe20, Phe466 aligns with Met345, Ala554 aligns with Ile428, whereas the rest are replaced by polar amino acids (Ile176 by Asn71, Leu177 by Ser72, Val443 by Ser321, Ile445 by Thre323) (Figure 3-64).

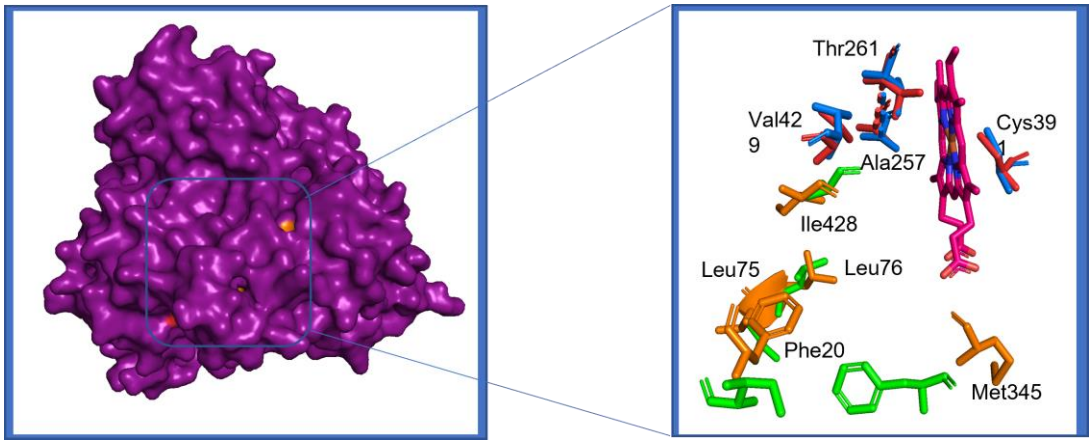


Figure 3-64 Hydrophobic residues forming binding pocket in 198 as aligned from CYP97A. On the left figure is shown a surface view of 198 with all the hydrophobic residues in orange placed in the binding pocket. On the right are shown the individual hydrophobic residues in orange as aligned from CYP97A (in green). Haem is shown in magenta with the iron ion in the centre in blue and orange, whereas all the catalytic residues are in the distal side of the haem (mentioned previously). Figure created with PyMol 2.5.4 (https://pymol.org/2/).

The above-mentioned suggests that the 198-metagenome enzyme may metabolise the β- and ε- rings of a-carotene to create lutein, while additional research is required to pinpoint the hydroxylation activities. According to a BLASTp alignment, metagenome gene 198 shares 99.77% sequence identity with the P450 family 132 of species *Dyadobacter endophyticus*, and 99 % coverage (440 of 441 residues). Metagenome gene 198 can be assigned to P450 family 132.

3.13. P450 Discussion and Future Work

Multiple cloning strategies were employed during this study to determine the best possible way to achieve P450 expression and catalytic activity. The P450s were initially cloned into pET29b and/ or pET28a as a first attempt of expression (data not shown). pET plasmids are a new derivatives of plasmids, constructed by Studier and colleagues in 1986 (Studier and Moffatt, 1986), which were developed by Novagen with new features to permit easier subcloning, detection and purification of target genes. Gene transcription is based on a strong bacteriophage T7 RNA polymerase and a strong T7 promoter that does not occur normally in *E. coli*. T7 RNA polymerase directs a high-level expression where the target protein can comprise more than 50% of the total cell protein, hours after induction.

Two different pET plasmid version exist; the transcription plasmids which are used for expression of genes derived from prokaryotic genes that contain their own prokaryotic ribosome binding site and ATG start codon; and the translation plasmids, which are used for the expression of eukaryotic derived genes and contain the ribosome binding site of the phage T7 major capsid protein and a reading frame recognition sequence relative to *BamH*I cloning site. The addition of a letter suffix to the plasmid name distinguishes translation vector names from transcription vector names. For example, pET-21a (+) indicates the reading frame in relation to the *Bam*HI cloning site recognition sequence, which indicates is a translation plasmid. In this study the translation version of these plasmid systems was used.

Then, human P450s and human reductase were co-cloned in pET29b plasmid fused together with a linker, and as a comparison in pCWOri$^+$ plasmid with the same strategy. This two-step cloning in a pET system proved to be a better cloning strategy as none of the genes mutated in these plasmids. pET plasmids can maintain the target genes transcriptionally silent till induction, which eliminates plasmid instability due to expression of proteins potentially toxic to the host. In contrast, two genes that were cloned in the pCWOri$^+$, CYP2A6- CPR- pCWOri$^+$ (-) and (+) clone, and CYP3A4-CPR-pCWOri$^+$ (+) carried mutations from PCR. Moreover, we achieved active human P450s only when cloned and expressed in pET plasmids!

Another cloning and expression strategy that we pursued, as a comparison for better expression, was human P450s that we cloned in pET22 and pET29b, and co-expressed with the human reductase, cloned in pETDuet-1 plasmid. As mentioned previously, pET22b plasmid leads the expressed protein in the periplasm, which can help with the enzyme folding. From this experiment we confirmed part of our

hypothesis once again, that human P450 CYP3A4 converted norketamine to its metabolite, hydroxynorketamine. CYP3A4 was found to be the principle enzyme used by human liver microsomes for ketamine N-demethylation, followed by a minor contribution by CYP2B6 and CYP2C9 (Chen and Chen, 2010). In our experiments, CYP3A4, cloned into pET22b and co-expressed with the human CPR-pETDuet-1, was catalytically active against norketamine. The biotransformations were performed for 24 hours. It is our aim in the future experiments to perform a pre-concentration step of all the samples before injecting them in the LC for analysis. Also, MS/MS assays are required in the future work to determine which hydroxynorketamine enantiomer was generated as product from the biotransformations, which we didn't elucidate in this study.

During the co-expression of P450s cloned into pET29b and CPR-pCDFDuet-1, we found that one of the genes in this cloning strategy metabolised norketamine to hydroxynorketamine. CYP2B6-pET29b and CPR-pCDFDuet-1, metabolised norketamine to its active metabolite hydroxynorketamine. The co-expression was of benefit, as both enzymes required for activity, i.e., the P450 and the reductase, were co-localised in the same compartment of the cell expressed along each other, and hence the electron flow from the $NADPH^{+}$ to the reductase FAD connecting domain, to FMN domain, and to the haem centre was facilitated. The actual yield of the product was low and the signal weak, as we performed the whole cell biotransformation assays without pre-concentrating the samples. Instead, once again we injected the samples in the LC neat. Something to look at in the future experiments, would be to perform docking studies and Quantum Mechanics, and predict the catalytic residues. Additionally, performing site directed mutagenesis to the catalytic residues might improve catalytic affinity towards the substrate and determine if higher product yield is achieved. Finally, performing some MS/MS experiments with the biotransformations to determine the exact hydroxynorketamine enantiomer are required in future work.

To achieve to high expression of the enzymes, an *N*-terminus modification was performed for all P450s. The *N*-terminus of a membrane protein is formed of a signal sequence consisting of several hydrophobic residues, that is recognised by a signal recognition particle and directs the protein to the endoplasmic reticulum in eukaryotes, or plasma membrane in prokaryotes (Wiedmann et al., 1987). Once the protein sequence is translated then the signal sequence is cleaved by signal peptidases in prokaryotes, but with eukaryotic P450s this sequence is not cleaved. If a eukaryotic protein is expressed in a bacterial system with its signal sequence, then it can lead to low expression as the signal sequence is not efficiently recognised by the bacterial

translocation mechanism. Additionally, replacing the eukaryotic signal sequence with that of bacterial ones led to increased P450s expression and significant improvement in recovering active P450s from the bacterial membranes (Pritchard et al., 1997). On the other hand, truncating all the *N*-terminal signal sequence, and adding a hydrophilic stretch can lead to soluble protein expression located in the cytoplasm (Li and Chiang, 1991). According to the above, the *N*-terminus hydrophobic signal peptide of our target human enzymes were deleted. Moreover, eukaryotic P450s are bound to the endoplasmic reticulum by an *N*-terminal transmembrane domain, and since bacteria do not have organelles, the eukaryotic proteins need to be directed to the plasma membrane of the bacteria to retain the membrane binding characteristics. That was done by inserting at the *N*-terminus a previously recognised sequence "LLLAVFL". Barnes et al. (1991) were the first to introduce the bovine CYP17a modification at the *N*-terminus (MALLLAVFL), and successfully expressed eukaryotic CYP17a in *E. coli* host cells with catalytic activity. The sequence is recognised easily by the ribosome for translation, and by the signal recognition particle for membrane localisation for *E. coli* (Barnes et al., 1991). The modification was introduced through PCR mutagenesis for CP2B6, CP2D6, CP3A4 and CP3A5, while CP2A6 was modified through gene synthesis (Table 3-2). On the other hand, the metagenome genes were cloned with and without the *N*-terminal modification. As the metagenome enzymes are in fact bacterial enzymes, we did not know if the modification would generate adequate yields or enhance enzyme solubility, or, simply expressing the original non-modified version of the metagenome genes would generate more soluble and active enzyme, therefore both strategies were employed.

Although both metagenome genes were active with the *N*-terminal modification, soluble enzyme was achieved from metagenome gene 914 only in the natural form, without the *N*-terminal modification. Moreover, the highest catalytically active 914 enzyme was recovered without an *N*- terminal modification, when co-expressed with pGro7 and induced with 0.5 mM IPTG in LB 20°C and LB°C.

In contrast, metagenome enzyme 198 expressed only with the MALLLAVF modification incorporated. From modelling studies we performed, 198 is associated with the membrane through transmembrane domains.

Furthermore, human P450s and metagenome genes were cloned with and without the incorporation of the 6 x histidine tags at their *C*-terminus. His tags are used for purification of the expressed enzymes using metal affinity chromatography, but sometimes the His tag can confer changes to protein structures, as it can interact with the active site and the haem group directly or indirectly (Lin et al., 2011) . Emoto et

al (2009), showed that midazolam could not access the centre of the haem of the hexahistidine tagged CYP3A5, with a high interaction energy compared to the non-tagged one (Emoto et al., 2009). Thielges and colleagues (2011) observed that the His-tagged proteins had an increased occurrence of B factors, which is an indicator of structural disorder and altered protein dynamics, compared to the non-modified proteins (Thielges et al., 2011). Moreover, results from another study indicated that the His-tag attached to the *C*-terminus of the tropinone reductase interferes with the catalytic activity. After performing protein modelling, they determined that only the *C*-terminus hexahistidine tagged enzyme was functionally impaired, and not the N-terminus, and the hexahistidine tag interfered with the active site by steric or electrostatic interactions (Freydank et al., 2008). In contrast, Lin and colleagues (2011) proved that attaching a six histidine tag at the N-terminus of cytochrome b5 and after performing structural prediction and molecular dynamics simulation, b5 had similar property and functions to that of the isolated cytochrome b5 (Lin et al., 2011).

To achieve P450 expression but also catalytically active enzyme without the probability of His tag interacting with the active site, some of the target P450s were cloned and attempted expression in glutathione S-transferase (GST) tag plasmid pGEX-6P-1, as GST fusion proteins. Metagenome enzyme 914 only expressed and was active when cloned in pGEX-6P-1 plasmid. We did not perform any purification assays of the soluble enzyme yet, but that is to take place in our future experiments, in order to fully characterise metagenome enzyme 914.

Another parameter that we incorporated in our study was plasmid pGro7 expressing GroES and GroEL chaperones. Often, P450 expressed in *E. coli* result in misfolding of the enzyme and the formation of inclusion bodies. Chaperones enhance the expression of P450s as helps in the folding of proteins (Mitsuda and Iwasaki, 2006). All cloned human and metagenome genes were co-expressed with the pGro7 plasmid to determine if the enzyme was enhanced this way. Indeed, we achieved catalytically active enzyme with and without pGro7 co-expression, although in most instances the active enzyme was achieved only when co-expressed with pGro7. This was down to the individual enzyme, and the requirement of each of the enzymes were different.

5-ALA is the first compound in the haem biosynthesis pathway. 5-ALA is formed from the condensation of succinyl-CoA and glycine by the ALA synthase enzyme in animals, lower eukaryotes and bacteria, including *E. coli* (Beale, 1990). Although 5-ALA is synthesised in *E. coli*, according to literature, addition of the 5-ALA and iron ions (ferric and ferrous) in the expressing media could increase P450 expression (Miura et al., 2015, Kumondai et al., 2020). Therefore, the media was supplemented

with final concentration of 0.25 mM of each Fe^{2+} and Fe^{3+}, whereas concentrations of 0.5 mM and of 1 mM 5-ALA were added to all the expressing cultures at the point of induction. When we increased the concentration of 5-ALA from 0.5 mM to 1mM, the enzyme activity was increased compared to the lower concentration (i.e., CYP2B6 expressed in C41(DE3) without pGro7), therefore all experiments were carried out at this concentration.

E. coli BL21(DE3) strain together with T7 promoter-based plasmids have been widely used for the expression of the recombinant proteins. Expression is tuned by T7 RNA polymerase, located in the DE3 prophage of the chromosome, and is controlled by L8-UV5 lac promoter, which is induced by IPTG. T7 RNA polymerase is more powerful than the wild type lac promoter, due to 3 mutations occurring; two base pair substitutions at the -10 promoter sequence bring the sequence closer to the one recognised by bacterial sigma factors, leading to RNA polymerase being recruited more effectively, reducing its dependence on CAP/cAMP stimulation for activation, and a mutation CAP/cAMP site decreases the affinity for CAP/cAMP (Mathieu et al., 2019). Moreover, BL21(DE3) strain is missing the Lon protease, which degrades many foreign proteins. BL21(DE3) it by far the most used strain for recombinant protein expression (Gottesman, 1996).

Finally, T7 RNA polymerase transcribes 8 times faster than the native *E. coli* RNA polymerase. This can lead to higher levels of mRNA being transcribed and more protein being synthesised, which can have a detrimental effect on protein production as can lead to improper protein folding which can result in inclusion bodies, or exceeding the capacity of membrane protein synthesis and secretion (Mathieu et al., 2019).

Attempted expression of all the clones was undertaken in *E. coli* C41(DE3) in LB at 20°C and 30 °C. C41(DE3) and C43(DE3) strains were designed by Miroux and Walker (1996), when isolating BL21(DE3) strains with the enhanced ability to express membrane protein strains, or toxic membrane proteins. C41(DE3) and C43(DE3) have also three mutations in the lacUV5 promoter, with two mutations at the -10 region reversing it back to the wild type lac promoter, and is susceptible to catabolite regulation. The transcription rate is reduced which explains why these two strains can express membrane or toxic proteins so effectively, without resulting in toxicity, and have improved membrane protein yields (Miroux and Walker, 1996).

Comparatively, *E. coli* Tuner cells are mutants of BL21(DE3) with deleted lacZY which enables them to adjust the levels of protein expression throughout the cells in a culture, by varying the IPTG concentration added. So, Tuner cells were also

employed in our study to tune recombinant protein expression. Hausjell and colleagues (2018) reported in their review that an IPTG concentration of 0.5 and 0.6 mM were found to be optimal in CYP expression optimisation (Hausjell et al.).

Although previous researchers generated exceptionally high expression levels of recombinant enzyme when using C41(DE3) (Ichinose et al., 2015, Cheng et al., 2004), we achieved greatest activity of target human P450s when expressed in Tuner cells in LB 20 ℃ and/ or 30 ℃ and induced with 0.25 mM IPTG. Tuner cells generated comparatively the same level of expression, but higher levels of catalytic activity of the expressed enzymes, which can be attributed to the IPTG concentration, as the lower the expression rate, increased the chance of human P450s folding properly, and hence not ending up in inclusion bodies.

Also, different culturing conditions and media were employed for the best possible expression of enzymes: LB and AIM at 20℃ and 30℃. Although AIM is a self-inducing and richer media, it did not have a positive impact in the expression of the enzymes. Better expression and catalytically active enzymes were retrieved from LB media when cultured at 20 ℃.

Additionally, we changed the concentration of the substrate mid-way in our WCB from 0.05 mg/mL to 10 µM, because it was too high, and employed LC-MS to analyse our WCB assays. There is fine line of what is the optimal concentration of substrate in a whole cell biotransformation, as high concentrations may inhibit the enzyme. Experiments to determine the optimal concentration of substrate must be performed first. Ketamine concentrations between 0.05 and 0.005 mM have been used in previous experiments. Moreover, CYP2A6 enzyme assays and phenacetin concentrations from 0 to 250 µM have also been used (Chen and Chen, 2010).

In conclusion, we confirmed with our study that there is no one rule that applies all. Each P450 enzyme is different, despite belonging to the same family, and / or subfamily, in the sense that they require different expression conditions such as cloning and expression plasmids, expression strains, cultivation temperatures, IPTG concentration, supplementation of the media with ferric and ferrous chloride (Fe_3 and $FeCl_2$), and 5-ALA concentration, as well as chaperone co-expression.

Conducting this comparison of all these cloning and expression methods will hopefully help future scientist achieve good expression levels and active P450s in fast and efficient manner, and we hope that we have provided that through our research.

Further characterising the two metagenome P450s, 198 and 914, is in our future plans. Metagenome enzyme 914 metabolises ketamine to its metabolite, hydroxyketamine already, but since the enzyme expressed solubly, we would firstly attempt to crystallise it. Solving the structure of the enzyme and providing mechanistic details can provide an insight as to which class of P450s according to its electron transfer system, as well as to prove our hypothesis if belongs to the self-sufficient class of enzymes and is a P450BM3 homolog. Secondly, performing kinetic assays to determine the rates of conversion, the optimal temperature and pH that the enzyme has its best catalytic activity. In addition, we only hypothesised the likely activity of the P450 198 is metabolising a-carotene to lutein, so in order to establish that activity, we need to perform biotransformation assays with enzyme and substrate.

4. Nitrilases Results

4.1. SDS-PAGE Analysis of the Expressed Enzymes

As mentioned earlier in the text, nitrilase enzymes have a wide range of industrial and biotechnological applications, including the synthesis of commercially significant carboxylic acids, which are intermediates in the API production, and the bioremediation of cyanide and toxic nitriles. These enzymes have been found in fungi, bacteria and plants, as well as in the genomes of eukaryotes.

We attempted the expression of 20 nitrilases, with the aim to be used in hydrolysis assays against a panel of six nitriles, which their chemical synthesis had failed previously. The 20 nitrilases were obtained from the personal collection of Prof. Gary Black. Additionally, 10 more nitrilases were provided as a lyophilised form ready to be assayed from Prozomix and the results are presented in Appendix 3 (Figures 3.10 and 3.11) due to the assays not being any further owing to IP restriction. Expression of nitrilases 1 to 10 were performed in LB 20 °C and LB 30 °C, while the expression conditions of nitrilase 11 to 20 were already known (from previous assays, Black et al., 2015 research paper). Nitrilases 11 to 20 expressed in LB 20 °C except for nitrilase 17, which expressed in AIM 30 °C (Table 4-1). Most of the nitrilases expressed in either or both temperatures. TCP and CFEs of expressed enzymes were separated in 12 % SDS-PAGE and visualised with Coomassie Blue dye. From Figure 4-1 can be seen that nitrilases 1, 2, 3, 4, 5, 6, 7 and 8 expressed well in LB 20 ℃, and LB 30 °C (Appendix 2, Figure 2.11, lanes 4, 5 and 6 in LB 30°C). All nitrilases expressed in LB 20 °C and LB 30 °C were found in the soluble fraction (Appendix 2, Figures 2.12 and 2.13). Nitrilases 11 to 20 expressed in LB 20 °C except for nitrilase 17, which expressed in AIM 30 °C.

Table 4-1 Summary of expressed nitrilases and their specifications.

Number given	UniProt code	Molecular Weight (kDa)	Expressing Conditions
1	Q2J474	34	LB 20 °C and LB 30 °C
2	A5ETE9	34.2	LB 20 °C and LB 30 °C
3	B4AL96	40.1	LB 20 °C and LB 30 °C
4	Q89PT3	36.6	LB 20 °C and LB 30 °C
5	A5EKU8	39.2	LB 20 °C and LB 30 °C
6	A0R378	39.1	LB 20 °C and LB 30 °C
7	Q89GE3	38.4	LB 20 °C and LB 30 °C
8	Q819F0	31.7	LB 20 °C and LB 30 °C
9	Q5LLB2	39.4	LB 20 °C and LB 30 °C
10	Q6N284	35.9	LB 20 °C and LB 30 °C
11	A5MYU1	37.9	LB 20 °C, 100 rpm
12	A7TP0T	34	LB 20 °C, 100 rpm
13	P20960	41.1	LB 20 °C, 100 rpm
14	Q2GR86	39.2	LB 20 °C, 100 rpm
15	A0LKP2	38.2	LB 20 °C, 100 rpm
16	E3HN55	36.9	LB 20 °C, 100 rpm
17	F0Q9Y1	39.6	AIM 30 °C, 180 rpm
18	G8CXY5	41.1	LB 20 °C, 100 rpm
19	D1C8L7	42.9	LB 20 °C, 100 rpm
20	C5DH06	37.4	LB 20 °C, 100 rpm

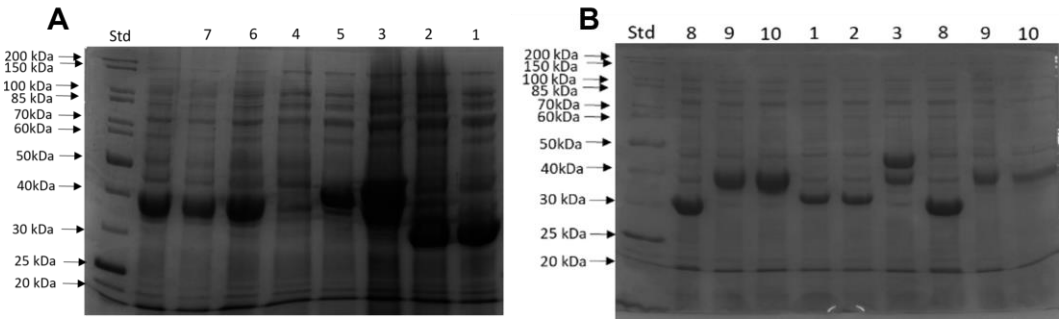


Figure 4-1 TCP of Nitrilase enzymes in LB 20 ℃ separated in SDS-PAGE gel. A: From left protein standard (ThermoFisher Scientific) lane 7; Q89GE3=38.4 kDa, lane 6; A0R378=39.1, lane 4; Q89PT3=36.6 kDa; lane 5; A5EKU8=39.2 kDa, lane 3; B4AL96=40.1 kDa lane 2; Q2J474=34 kDa; lane 1; A5ETE9=34.2 kDa, **B:** lanes 8, 9 and 10 from left are in LB 20 ℃, while lanes 1 to 10 are in LB 30 ℃. protein standard, lane 8; Q819F0=31.7 kDa, lane 9; Q5LLB2=39.4 kDa, lane 10; Q6N284=35.9 kDa, lane 1; Q2J474=34 kDa, lane 2; A5ETE9=34.2 kDa, lane 3; B4AL96=40.1 kDa, lane 8; Q819F0=31.7 kDa, lane 9; Q5LLB2=39.4 kDa, lane 10; nitrilase Q6N284=35.9 kDa.

4.2. High-Throughput 96 Well-Plate Assay to Determine the Catalytic Activity of Nitrilases Against 6 Nitriles.

Initially, microtiter plate assays were performed with all nitrilases to determine their catalytic activities against the panel of six nitriles, plus a control nitrile. The assay determines the activity according to the ammonia generated which reacts with the OPA under acidic conditions and generates a chromophore which can be measured with UV/visible spectroscopy. All nitrile structures and their corresponding carboxylic acids are summarized below (Figure 4-2). Nitriles 1, 3 and 7 belong to the arylacetonitriles, nitriles 2 and 5 are aliphatic nitriles, whereas 4, 6 and 8 are aromatic nitriles (refer to Table 2-39).

According to the high-throughput assay nitrilase Q2GR86 showed good activity against benzyl (2-cyanoethyl) carbamate (nitrile 1) and 2-(4-cyanophenyl)-N-ethylacetamide (nitrile 6), and some activity for N-(3-cyanopropyl) pthalimide (nitrile 3). The results are presented in Figure 4-3, while the 96 well plates are shown in Appendix 3, figure 3-9. The rest of the nitrilases did not show any activity against the nitriles. All assays were performed in triplicate to confirm results. The known activity of a nitrilase Q89GE3 against nitrile 3-phenylpropionitrile was used as a positive control in the 96 well plate assay (Appendix 3, Figure 3-8), whereas *E. coli* BL21(DE3) CFEs with solutions of all nitriles were used as the negative controls.

According to the quantitative data from the absorbance traces shown in the graph below (Figure 4-3), the nitrilase Q2GR86 has greater catalytic affinity towards nitrile 6 than nitrile 1. According to the results of the assay, nitrilase Q2GR86 appears to favour aromatic nitrilases.

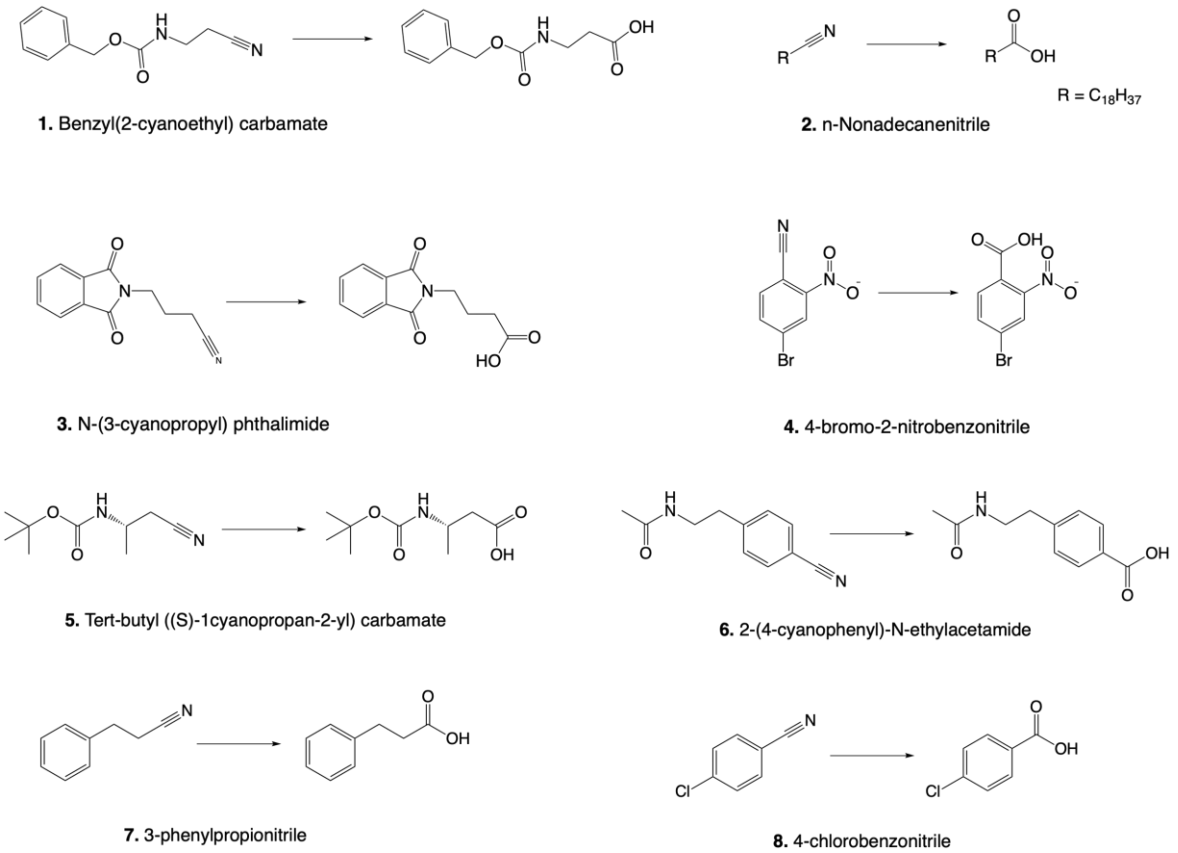


Figure 4-2 Nitriles used in the biocatalytic assays, and opposite them, indicated with an arrow, their corresponding carboxylic acids. Nitriles 1, 3 and 7 belong to the arylacetonitriles, nitriles 2 and 5 are aliphatic nitriles, whereas 4, 6 and 8 are aromatic nitriles. Nitriles 7 and 8 were used as a control. Nitrile 7 was used for the detection of ammonia in 96 well plate assays, whereas Nitrile 8 was used in whole cell biotransformation assays against Q2GR86. Figure drawn with ChemDraw.com

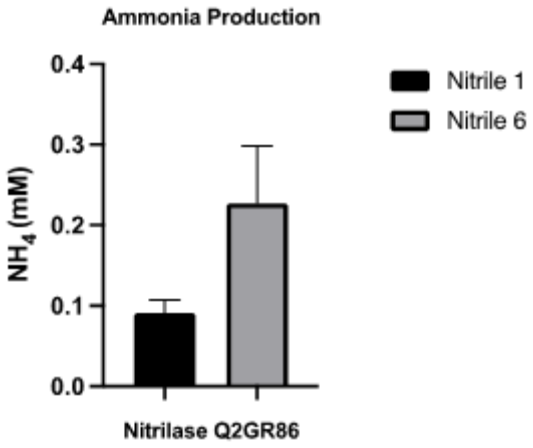


Figure 4-3 Catalytic affinity of nitrilase Q2GR86 against nitriles 1 and 6. Graph showing the Q2GR86 hydrolysis of nitriles 1 and 6, and the release of ammonia in the ammonia detection assay.

4.3. Purification of nitrilase Q2GR86

Nitrilase Q2GR86 was purified using Bio-rad Econo-pac® Chromatography Columns. In Figure 4-4 **A** are shown all the purified fractions separated via 12% SDS-PAGE. After purification, all the fractions were pooled together, concentrated using Amicon Ultra 4 centrifugal filter units (10 kDa cut off) and were purified using Size Exclusion Chromatography (SEC) via an AKTA system and Hiprep 16/600 Sephacryl S-200 HR column (Figure 4-4 **B**). The collected samples were pooled together once again and concentrated again. The concentration of the enzyme was recorded using NanoDrop ONE (Labtech). The concentrated enzyme was put through the crystallisation process twice, but no crystals were recovered. Instead, structure prediction was performed through PHYRE2, and analysed with PyMol (see section 2.2.12).

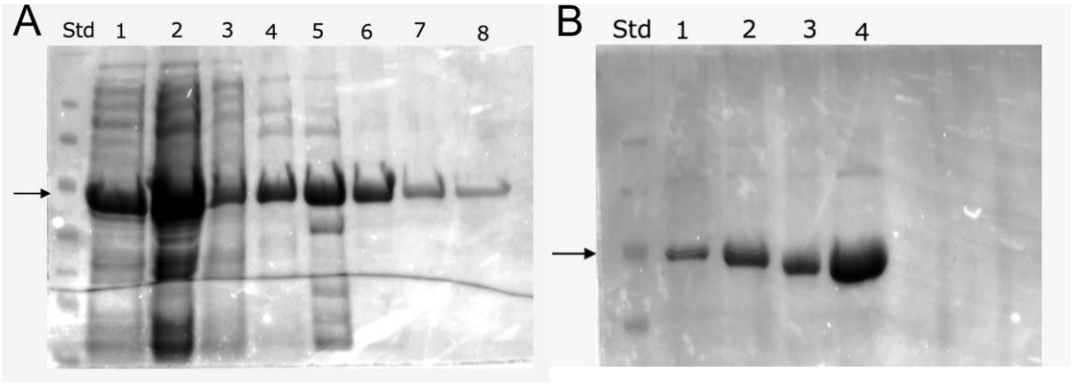


Figure 4-4 A: Purified fractions of nitrilase Q2GR86 on SDS gel. From left; protein standard (ThermoFisher Scientific), lane 1; sample flow through, lane 2; binding buffer wash, lane 3; 20 mM imidazole, lane 4; 100 mM imidazole, lane 5; 200 mM imidazole, lane 6; 300 mM imidazole, lane 7; 400 mM imidazole, lane 8; 500 mM imidazole. **B:** Size exclusion chromatography fractions of nitrilase Q2GR86 on SDS gel. Lanes 1 to 4 are the fractions collected after the purified enzyme fractions were pooled together and were further purified using Size Exclusion Chromatography (SEC) via an AKTA system via an AKTA system (GE Healthcare) and sephacryl columns.

4.4. Conversion Rates of Nitriles to Carboxylic Acid with Release of Ammonia

After determining the activity of nitrilase Q2GR86 through the high-throughput 96 well assay, another experiment was performed with purified and concentrated enzyme to determine the conversion rates. The conversion of nitrile to carboxylic acid and ammonia was measured, through an indirect method, using the Ammonia Assay Kit (Rapid) (Megazymes). Essentially, 2-oxoglutarate is converted to L-glutamic acid by the enzyme glutamate dehydrogenase (GDH) in the presence of ammonia, where NADPH is consumed and $NADP^+$ is generated (Anusree and Nampoothiri, 2015), which was quantified in this assay. As mentioned previously in the methods, 0.054 mg of enzyme Q2GR86 was used in 1 mL reaction volumes and 10 mM final concentration of each nitrile. In the first experiment, nitrile 1 was totally hydrolysed over the course of 24 hours, whereas nitrile 6 was totally hydrolysed in 4 hours (Figure 4-5). Nitrile 6 was hydrolysed by nitrilase Q2GR86 to its carboxylic acid and release of ammonia in 4 hours, where 41.703 mM of nitrile 6 were hydrolysed /mg of enzyme /h/ mL of solution, or 8.4×10^{-4} mM of substrate 1/Mol of enzyme/ hour were consumed in the reaction.

As the rate of conversion was slow for nitrile 1 this assay was repeated with increased enzyme concentration to final concentration 204 µg/ mL (stock enzyme concentration 8.89 mg/ mL), which is 4 times higher the previously used concentration, whereas the concentration of the nitrile remained the same in the repeat reaction (Figure 4-6).

At these new conditions, nitrile 1 was hydrolysed by Q2GR86 to its carboxylic acid with the release of ammonia in 6 hours, where 23.88 mM of Nitrile 1 were hydrolysed /mg of enzyme/ h /mL solution, or 4.8×10^{-4} mM of substrate /Mol of enzyme/ hour were consumed in the reaction.

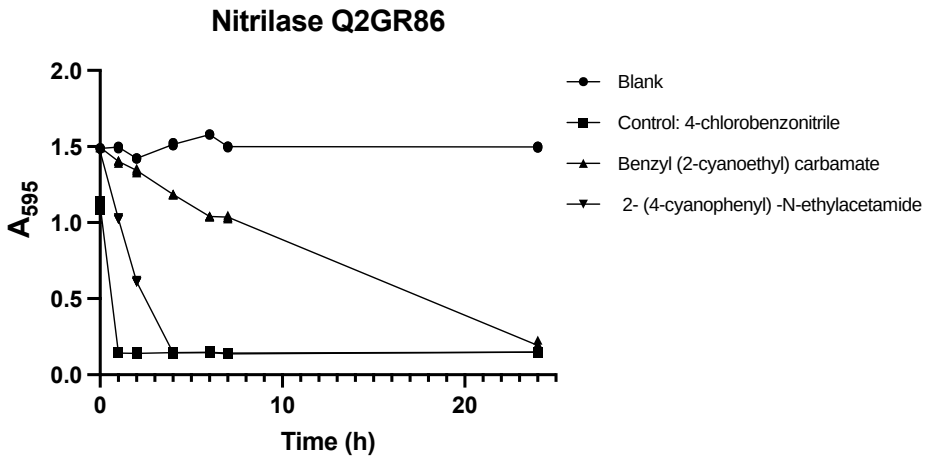


Figure 4-5 Nitrilase Q2GR86 metabolising nitriles 1; Benzyl (2-cyanoethyl) carbamate, and nitrile 6; 2-(4-cyanophenyl)-N-ethylacetamide to carboxylic acid and ammonia. In the graph can be seen the negative control (blank) remained at same absorbance throughout our assays, control nitrile; 4-chlorobenzonitrile was hydrolysed in 1 hour, nitrile 1; benzyl (2-cyanoethyl) carbamate was hydrolysed in 24 hours, and nitrile 6; 2- (4-cyanophenyl) -N-ethylacetamide was hydrolysed in 4 hours.

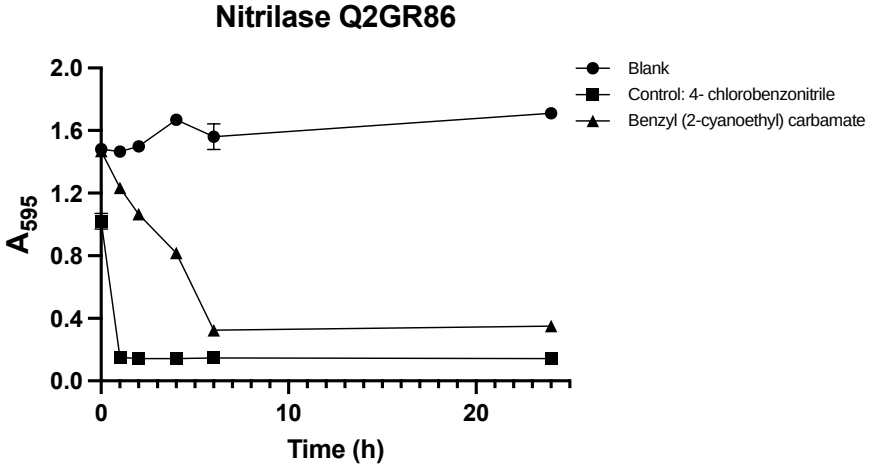


Figure 4-6 Nitrilase Q2GR86 hydrolysing nitrile 1 (best concentration of nitrilase used for the hydrolysis of nitrile 1). The same in this repeat assay, the negative and positive control confirm the suitability of the assay, while nitrile 1; benzyl (2-cyanoethyl) carbamate was hydrolysed in 6 hours to its carboxylic acid.

4.5. Structure Prediction of Nitrilase Q2GR86

The Q2GR86 structure was predicted based on 3 templates (PDB: 6zby, 3wuy and 6i00) based on heuristics to maximise confidence, percentage identity and alignment coverage, while 17 residues were modelled *ab initio*.

Enzyme Q2GR86 aligned with 6zby with 100 % confidence and 32 percentage identity and 93 % coverage (residues 5-323), with 3wuy with 100 % confidence and 34 percentage identity and 86 % coverage (residues 1-297), and 6i00 with 100 % confidence and 33 percentage identity and 92 % coverage (residues 8-324). Q2GR86 superimposed with 3wuy had an RMSD of 0.550 Å, while 6zby had RMSD value of 0.728 Å, and with 6i00 had RMSD value of 1.125 Å. All three RMSD values are quite small, which means all the templates and predicted structure are very similar (Kufareva and Abagyan, 2012).

Q2GR86 predicted structure is in the monomeric level, 3wuy forms a dimer (Zhang et al., 2014), whereas 6zby and 6i00 form homo-oligomers. All three templates have nitrilase activity, while 6i00 also has nitrile hydratase activity. 3wuy was chosen for mapping of Q2GR86 due to their RMSD value of 0.550, for their close structural similarity at the monomer level, and sequence identity of 34 % (Figure 4-7). Enzyme 3wuy (Nit6803) belongs to *Syechocystis* sp. strain PCC6803, and mononitriles, dinitriles, aromatic nitriles, and aliphatic nitriles are just a few of the wide range of substrates that Nit6803 exhibits hydrolysis activity toward (Zhang et al., 2014).

Nit6803 forms a dimer, but forms a decamer in solution, with molecular weight 397 kDa. On the monomeric level, Nit6803 consists of two helical layers. Layer one is made up of α1 and α4 helices (aa 26–44 and 87–100, respectively), and layer two is made up of α7 and α8 helices (aa 170-185 and 200–217, respectively). In contrast, the hydrophobic pocket of Nit6803 is formed by a central two-layer β-sheet made up of eleven β-strands. The two helical layers surround the central β-sheets on both sides. Moreover, at its C-terminus Nit6803 consist of a α-helix structure (aa 279 - 289), and two α-helix insertions. The first insertion consists of α3 and α2 helices (aa 61-73) connected by four amino acid loop, and second insertion by α9 and α10 helices (aa 228-248) connected by a two amino acid loop. These insertion fragments are proposed to be the substrate binding loops, and structural differences are subject to substrate profiles of the nitrilases (Zhang et al., 2014). Structural comparison of Nit6803 and Q2GR86 reveals the notable structural similarity of both structures and suggests that the above-mentioned features apply to our structure, although few structural deviations are observed (Zhang et al., 2014).

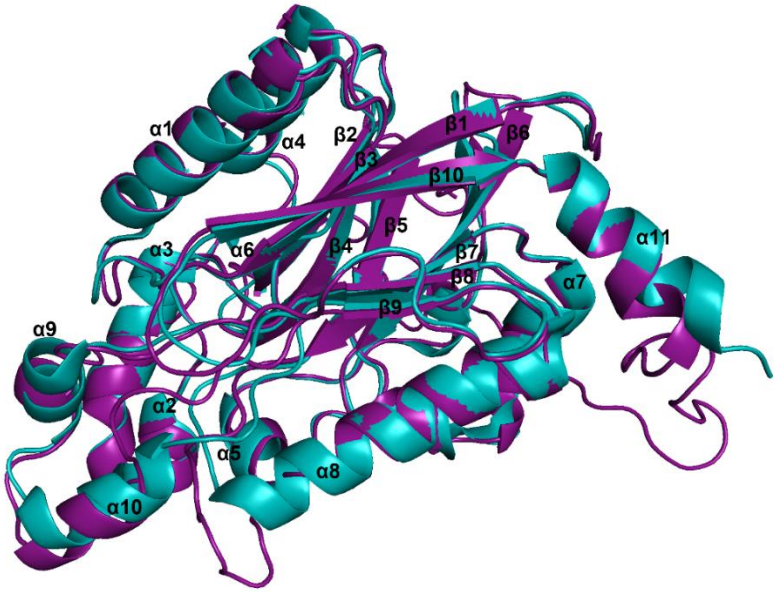


Figure 4-7 Cartoon view showing the superimposed structures of Q2GR86 (in purple) and Nit6803 (in teal). The labelling of α helices and β-sheets was done based on the structure of Nit6803. A notable structure similarity can be observed from the alignment between Q2GR86 and Nit6803. In the middle of both structures can be seen the 11 β-sheets that form the hydrophobic active site, whereas around them are seen the α-helices, all labelled according to Zhang et al, (2014). Small structural differences are observed, such as the distance between the α11-helix structure on the right side of the image. Image is produced with PyMol 2.5.4 (https://pymol.org/2/).

4.6. Nitrilase Q2GR86 Catalytic Residues and Active Site

According to mutagenesis studies, there is a crucial residue involved in substrate preference in Nit6803, namely Trp146. Structural alignments between Q2GR86 and Nit6803 suggested that this residue aligns with Trp145 in Q2GR86 (Zhang et al., 2014). The most conserved catalytic residues in the nitrilase super-family are Glu54, Lys127, and Cys169 (Pace and Brenner, 2001) which are very similar or identical to the catalytic residues in Nit6803 and Q2GR86, as the identified catalytic residues in Nit6803 are Glu53, Lys135 and Cys169, whereas Glu53 aligns with Glu48, and Lys135 with Lys134 in our structure, whereas Cys169 is conserved in Q2GR86 structure (Figure 4-8). These are identified as the catalytic residues responsible for substrate recognition and stabilisation. Another conserved residue is Glu142, which aligns with Glu141 in our structure. This residue is within hydrogen bond distance of Lys135 which confirms that Glu53, Lys135 and C169 form the catalytic trio for the catalysis. Lys135 aligns in our structure with Lys134. Mutational studies within the catalytic residues of Nit6803 resulted in loss of activity (Zhang et al., 2014).

Additionally, residues Tyr125, His129, Tyr170, Asp171, Arg173 and Phe174 are highly conserved within the family (Pace and Brenner, 2001), and of these Tyr170 aligns with Trp170, Asp171 with Glu171, and Phe174 aligns with Met174 in Q2GR86, whereas Tyr125, His129 and Arg173 are not conserved in our structure. All the conserved residues identified from Pace and Brenner 2001, are seen in the active site coloured in lime green (PDB: 1ems), and in Q2GR86 in light magenta (Figure 4-8). Furthermore, mutations of residues E180Q, D264N and Y52F resulted in increased activity as these residues are involved in the folding of enzyme, and mutations led to loss of enzyme stability (O'Reilly and Turner, 2003). Phe52 aligns with Phe46 in Q2GR86 structure.

According to Zhang et al. (2014) few aromatic and hydrophobic residues combine to generate the hydrophobic pocket of Nit6803, whereas in the brackets are presented how they align in our structure: the hydrophobic residues are Phe64 (no alignment), Tyr140 (Tyr139), His141 (no alignment), Trp170 (Trp170), Pro194 (Pro194), Met197 (Met199), Val198 (Val205), and Phe202 (Ala206). This suggests that the hydrophobic pocket of our enzyme consists of the same hydrophobic residues according to the alignment.

Moreover, at the base of the binding pocket lie the catalytic residues Glu38, Lys134 and Cys169, which can establish hydrogen bonds with Asn118 (Asp118) and Glu141 (Glu142) and stabilise further. The residues in brackets are the ones seen in Nit6803 that aligned in our structure (Zhang et al., 2014).

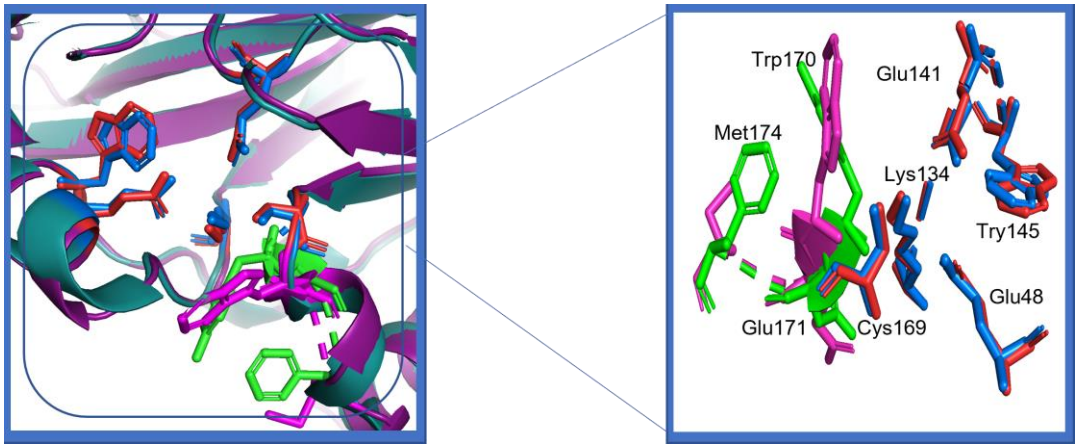


Figure 4-8 Cartoon view of superimposed structures of Q2GR86 (purple) and Nit6803 (teal), and the conserved catalytic residues of the nitrilase family. On the left figure can be seen superimposed the catalytic residues responsible for substrate recognition and stabilisation coloured in red (Q2GR8) and in blue (Nit6803). Those are Glu48 (Glu53), Cys169 (Cys169), Lys134 (Lys135), Glu141 (Glu142) and Trp145 (Trp144). In brackets are the conserved catalytic residues of Nit6803. Other aligning residues coloured in light magenta (Q2GR86) and lime green (PDb:1ems) are the other super-conserved catalytic residues according to Pace and Brenner (2001), Tyr170 which aligns with Trp170 in our structure, Asp171 aligning with Glu171, and Phe174 aligns with Met174 in Q2GR86 (Pace and Brenner, 2001). On the right image are isolated only the above-mentioned residues. Image is produced with PyMol 2.5.4 (https://pymol.org/2/).

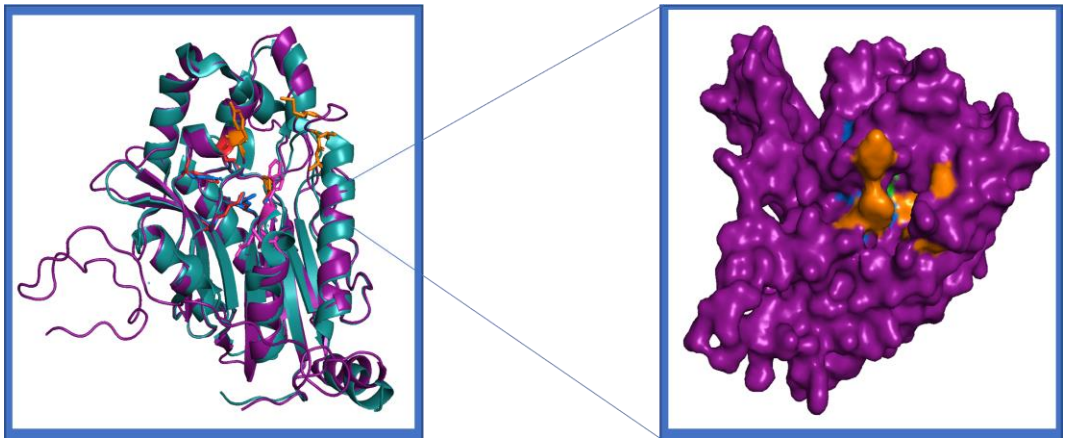


Figure 4-9 Surface display of Q2GR86 structure coloured in purple. In blue are coloured the conserved catalytic residues (Glu48, Lys134, Glu141, Cys169 and Trp145), whereas in orange are coloured the hydrophobic residues that form the catalytic pocket (Tyr139, Trp170, Pro194, Met199, Val205, and Ala206), and the Asn118 and Glu141 that can form hydrogen bonds with the catalytic residues, shown n green. Image is produced with PyMol 2.5.4 (https://pymol.org/2/).

4.7. Nitrilases Discussion and Future Work

Nitrile hydratases consist of two types of subunits assembled in dimers or tetramers, but can be bigger structures as well (Yamada and Kobayashi, 1996).

In the active site of the nitrile hydratases exists an iron atom coordinated by three sulphur atoms of cysteine residues, such as Cys111 and Cys113 in the nitrile hydratase of *Rhodococcus sp*. N-774, and two nitrogen atoms. NO coordinates the iron atom as the sixth ligand, which is believed to be the active site, and is replaced by hydroxide in the active form of enzyme and acts as the nucleophile to attack the nitrile carbon and to form the amide. The conserved Tyr72 involved in substrate recognition and binding, is one of the predominant catalytic residues in iron containing hydratases (Nagashima et al., 1998). Tyr72 aligns in Q2GR86 structure with Tyr73, whereas in Nit8603 is replaced by Leu78. Tyr72 is instead replaced by tryptophan in cobalt containing enzymes, such as in *Pseudonocardia thermophila* JCM 3095, and this replacement is believed to be involved in the aromatic rather than aliphatic nitrile hydrolysis (Miyanaga et al., 2001).

Nitrilases on the other hand, do not contain an iron atom at the catalytic centre, but contain a sulfhydryl group. The nitrilases can be a form of homo-oligomers, and less frequently a monomer with a protein of approximately 40 kDa which aggregates to form the catalytic enzyme (O'Reilly and Turner, 2003). They can be divided in 3 groups depending on their substrate affinity: to aromatic, aliphatic and arylacetonitriles. Enzyme Q2GR86 is a characterised nitrilase/cyanonitrilase according to UniProt and showed the best structural alignment and similarity to Nit8603, which is another characterised nitrilase. Nitrilase Q2GR86 forms a monomer according to its predicted structure and contains the conserved catalytic trio of Glu48 (aligning with the conserved Glu53 in the family), Lys135 and Cys169 required for catalysis. The sulfhydryl group of Cys169 acts as the nucleophile to attack the nitrile carbon atom, the catalytic glutamic acid serves as the general base and the catalytic lysine/glutamic acid as the general acid (Kobayashi and Shimizu, 1994, Martínková and Křen, 2010). In the Nit6803 structure, the distance between the sulphur of Cys169 and the oxygen of Glu53 is 3.2 Å indicating that electron transfer can happen between Cys169 and Glu53, which might activate the cysteine residue (Zhang et al., 2014). In Q2GR86, that distance is 3.9 Å, which indicates that the same interaction can happen in our structure.

Moreover, nitrilase Q2GR86 contains also other conserved catalytic residues that are shared between the members of the superfamily, which are Tyr125, His129, Tyr170, Asp171, Arg173 and Phe174, as described from the active site of the Nit domain

from the NitFhit enzyme of *Caenorhabditis elegans*, whereas Tyr125 and Arg173 are not conserved in Q2GR86 (Kumari and Poddar, 2019).

The hydrophobic pocket of nitrilases Q2GR86 is made up of several hydrophobic residues, the same as in Nit6803, and two aromatic residues, Trp170 and Phe202. The aromatic residues might offer the π-π interactions needed to stabilise the substrates for aromatic nitriles, according to Zhang et al, (2014). Q2GR86 exhibited high catalytic affinity towards 2 arylacetonitriles, benzyl(2-cyanoethyl) carbamate, and 2-(4-cyanophenyl)-N-ethylacetamide, hence can be categorised as an aromatic nitrilase enzyme from our study, supporting the findings of affinity towards aromatic nitriles. Moreover, the aromaticity of crucial residue Trp146 may be key in determining the substrate specificity according to Zhang et al. (2014).

The characterisation of nitrilases Q2GR86 was only partially done based on the sequence identity (34 %) and structural alignment with Nit6803. A full characterisation would require growing crystals and predicting the catalytic residues. A second step is to carry out mutagenesis tests to identify the catalytic residues. It would also be advantageous to do a thorough examination of the catalytic residues, specific substrate specificity, and affinities. We have already conferred some insight into the types of substrates that nitrilase Q2GR86 can accommodate, as well as which residues might be its catalytic ones depending on how Q2GR86 is superimposed with Nit6803. However, more research is necessary to fully explain these findings, so we intend to expand this work in our subsequent experiments. We wish to characterise Q2GR86 thoroughly by finishing all of the kinetics for the nitrilase Q2GR86 at various temperatures and pH levels and determining the Km and Vmax, with the goal of publishing all of the data shortly. As per the importance of the nitriles used in this study, they are subject of Quotient Sciences, and we cannot know what their corresponding carboxylic acids will be used for.

5. Final Discussion and Conclusions

Biocatalysis is a green technology used in the conversion of substrates in a more sustainable way with less bioproduct formation and less burden to the environment. Reactions are realised in aqueous environments, at physiological pH and temperature, without the harsh conditions that are required for the synthetic chemical reactions. Cytochrome P450 present an exciting group of enzymes with a vast functionality upon a wide range of substrates. P450s are the cornerstone in the drug development and discovery due to their ability to catalyse stereo- and regio- catalytic reactions and produce pure compounds compared to chemical methods, classifying them as extremely valuable to the pharmaceutical industry.

Our aim in chapter 3 was to clone and express human P450s 2A6, 2B6, 3A4, 3A5, and 2D6 and use them in whole cell assays to determine if they have any catalytic affinity towards norketamine. There was already research indicating that CYP2B6 and CYP3A4 could potentially metabolise norketamine. The other P450s used were selected based on sequence identity that they shared with these two P450s. Additionally, to increase the likelihood of discovering active P450s, we used a blast search to mine the Prozomix database and discovered two metagenome genes, 198 and 914 that shared up to 24 % sequence identity the human P450 3A5.

After performing a literature survey, we came to the conclusion that there is no clear approach for producing high expression but also biocatalytically active P450s, so we decided to investigate that area of the research and determine the most effective way to do so. The P450s' expression turned out to be challenging and time-consuming. To address this issue, various cloning techniques were used, and a variety of expression method were employed. A N-terminus modification was applied also to the P450s to help with their solubility. All the active enzymes were used in whole cell biocatalytic applications against norketamine. All the work completed in chapter 3 resulted in human P450s 2B6 and 3A4 to hydroxylate norketamine to hydroxyketamine in whole cell biotransformations, although the activity was low.

On the other hand, metagenome genes turned out to be simple to express. Furthermore, they underwent no time-consuming co-cloning with reductase due to them relying on the bacterial reductase for activity, and one of them, 914 was found in the soluble fraction when expressed without the N-terminal modification. Although both metagenome genes expressed and were active, only metagenome 914 enzyme showed good activity against ketamine in whole cell biotransformation, which was hydroxylated to create hydroxyketamine and N-demethylated to norketamine. Looking at the graph (Figure 3-56), after 24 hours the starting material

(ketamine) reduced to 30 % of the total, while norketamine increased by 30 %, whereas hydroxyketamine increased by 40 %.

Structural analyses were performed on 914 and the result showed that 914 shares highest homology and percentage identity up to 46 % with P450BM3 from *Bacillus megaterium* at the amino terminus, while at the C-terminus shares 30 % identity with CYPOR which is a cytochrome P450 reductase. Further analysis revealed that all the catalytic residues identified in P450BM3 as well as the substrate recognition sites align perfectly in 914 protein structure. All the evidence suggests that gene 914 may be a part of a self-sufficient P450 that is associated with the P450BM3. We used Blastp alignment (NCBI), and enzyme 914 from the metagenome showed 98.9 % sequence identity with CYP120A1 from the organism *Cupriavidus metallidurans* strain CH34. As a result, it can be categorised as belonging to P450 family 120 and subfamily A (or CYP120A), though the subfamily member's number will not be assigned until it has been thoroughly characterised.

On the other hand, metagenome enzyme 198 shared 30 % structural identity with CYP97A3 from the organism *Arabidopsis thaliana*, which catalyses the hydroxylation of the β- and ε- rings of α-carotene and produces lutein. Moreover, all the catalytic residues from CYP97A3 align well in 198 structure. A BLASTp alignment shows that the metagenome gene 198 and the P450 family 132 of the species *Dyadobacter endophyticus* have 99.77 % sequence identity and 99% coverage (440 of 441 residues). P450 family 132 can be used to classify metagenome gene 198.

Although all the structural analysis on both metagenome genes are only predictions, it is in our future plans to fully characterise both enzymes and assign them P450 family and subfamily name and number. Additionally, we will attempt to produce crystals for 914 enzyme and solve the structure, and further determine the catalytic residues. Furthermore, we would like to complete our catalytic assays of 198 enzyme against α-carotene and if does hydroxylate it to lutein.

Another exciting group of enzymes are nitrilases due to their ability to catalyse the conversion of a wide number of nitriles to their corresponding carboxylic acid and ammonia. The nitriles are intermediate used in the formation of pharmaceuticals and other fine chemicals, and their hydrolysis is important in the biocatalysis community for API production. The purpose of chapter 4 was to create nitrilase enzymes capable of biocatalytically hydrolysing six nitriles to their carboxylic acids. The chemical manufacture of these nitriles had failed through numerous rounds before. All nitrilases were relatively easy to express. 20 nitrilases were produced in this process, and 10 nitrilases were received in lyophilised form from Prozomix.

A high throughput was performed where identified that nitrilase Q2GR86 hydrolysed benzyl (2-cyanoethyl) carbamate (nitrile 1) and 2-(4-cyanophenyl)-N-ethylacetamide (nitrile 6) and showed some activity for N-(3-cyanopropyl) pthalimide (nitrile 3). Kinetic assays were performed with nitrilase Q2GR86 against nitrile 1 and 6. The results showed that nitrile 6 was hydrolysed to the carboxylic acid by nitrilase Q2GR86 in 4 hours, with 41.703 mM of nitrile 6 hydrolysed /mg of enzyme /h/ mL of solution, or 8.4 x 10^{-4} mM of substrate 1/Mol of enzyme/hour were consumed in the reaction, whereas nitrile 1 was hydrolysed by Q2GR86 to its carboxylic acid with the release of ammonia in 6 hours, where 23.88 mM of Nitrile 1 were hydrolysed /mg of enzyme/ h /mL solution, or 4.8 x 10^{-4} mM of substrate /Mol of enzyme/ hour were consumed in the reaction. All the kinetics were determined at 25 °C, and we intend to expand the kinetic studies across a range of temperatures and pH levels.

Furthermore, the crystallisation of nitrilase Q2GR86 was unsuccessful both times. Instead, a structure prediction generated, showed that Q2GR86 aligned with Nit6803 (PDB: 3wuy) with 100 % confidence and 34 percentage identity and 86 % coverage and superimposed with 3wuy had an RMSD value of 0.550 Å, which means the share high structural homology. Moreover, all the catalytic residues and the hydrophobic residues constructing the active site of Nit6803 align perfectly in nitrilase Q2GR86. It is possible to classify nitrilase Q2GR86 as an aromatic nitrilase, providing validity to the findings of affinity for aromatic nitriles according to our study, although this remains to be confirmed in further studies.

In conclusion, this project focused on the cloning and expression of enzymes for remote hydroxylation. Biocatalysis has emerged as an important procedure in the synthesis of APIs, as evidenced by the fact that numerous pharmaceutical companies have blended this green technology with traditional medicinal chemistry. There are numerous other families of enzymes that could offer considerable advantages in biocatalytic processes over traditional chemical routes. In this study, we investigated the use of human P450s, metagenomic P450s, and nitrilases which generated promising results and therefore hold the potential for use in upscale biocatalytic processes.

References:

1997. 25 - Iron, Ruthenium and Osmium. *In:* GREENWOOD, N. N. & EARNSHAW, A. (eds.) *Chemistry of the Elements (Second Edition).* Oxford: Butterworth-Heinemann.

AGHAEI, S., SAFFAR, B., GHAEDI, K. & MOBINI DEHKORDI, M. 2016. Functional analysis of recombinant codon-optimized bovine neutrophil β-defensin. *Journal of Advanced Research,* 7**,** 815-821.

AMIT, S. K., OBACH, R. S. & TRISTAN, S. M. 2007. Mechanism-Based Inactivation of Cytochrome P450 Enzymes: Chemical Mechanisms, Structure-Activity Relationships and Relationship to Clinical Drug-Drug Interactions and Idiosyncratic Adverse Drug Reactions. *Current Drug Metabolism,* 8**,** 407-447.

ANUSREE, M. & NAMPOOTHIRI, K. M. 2015. Chapter 12 - White Biotechnology for Amino Acids. *In:* PANDEY, A., HÖFER, R., TAHERZADEH, M., NAMPOOTHIRI, K. M. & LARROCHE, C. (eds.) *Industrial Biorefineries & White Biotechnology.* Amsterdam: Elsevier.

AOYAMA, Y., YOSHIDA, Y. & SATO, R. 1984. Yeast cytochrome P-450 catalyzing lanosterol 14 alpha-demethylation. II. Lanosterol metabolism by purified P-450(14)DM and by intact microsomes. *J Biol Chem,* 259**,** 1661-6.

ASANO, Y., TANI, Y. & YAMADA, H. 1980. A New Enzyme “Nitrile Hydratase” which Degrades Acetonitrile in Combination with Amidase. *Agricultural and Biological Chemistry,* 44**,** 2251-2252.

BAESHEN, N. A., BAESHEN, M. N., SHEIKH, A., BORA, R. S., AHMED, M. M. M., RAMADAN, H. A. I., SAINI, K. S. & REDWAN, E. M. 2014. Cell factories for insulin production. *Microbial cell factories,* 13**,** 141-141.

BARCLAY, M., TETT, V. A. & KNOWLES, C. J. 1998. Metabolism and enzymology of cyanide/metallocyanide biodegradation by Fusarium solani under neutral and acidic conditions. *Enzyme and Microbial Technology,* 23**,** 321-330.

BARNABA, C., MARTINEZ, M. J., TAYLOR, E., BARDEN, A. O. & BROZIK, J. A. 2017. Single-Protein Tracking Reveals That NADPH Mediates the Insertion of Cytochrome P450 Reductase into a Biomimetic of the Endoplasmic Reticulum. *J Am Chem Soc,* 139**,** 5420-5430.

BARNES, H. J., ARLOTTO, M. P. & WATERMAN, M. R. 1991. Expression and enzymatic activity of recombinant cytochrome P450 17 alpha-hydroxylase in Escherichia coli. *Proceedings of the National Academy of Sciences,* 88**,** 5597-5601.

BEALE, S. I. 1990. Biosynthesis of the Tetrapyrrole Pigment Precursor, delta-Aminolevulinic Acid, from Glutamate. *Plant Physiol,* 93**,** 1273-9.

BELL, E. L., FINNIGAN, W., FRANCE, S. P., GREEN, A. P., HAYES, M. A., HEPWORTH, L. J., LOVELOCK, S. L., NIIKURA, H., OSUNA, S., ROMERO, E., RYAN, K. S., TURNER, N. J. & FLITSCH, S. L. 2021. Biocatalysis. *Nature Reviews Methods Primers,* 1**,** 46.

BERNHARDT, R. 2006. Cytochromes P450 as versatile biocatalysts. *Journal of Biotechnology,* 124**,** 128-145.

BEZIRTZOGLOU, E. E. V. 2012. Intestinal cytochromes P450 regulating the intestinal microbiota and its probiotic profile. *Microbial ecology in health and disease,* 23**,** 10.3402/mehd.v23i0.18370.

BHALLA, T. C., KUMAR, V., KUMAR, V., THAKUR, N. & SAVITRI 2018. Nitrile Metabolizing Enzymes in Biocatalysis and Biotransformation. *Applied Biochemistry and Biotechnology,* 185**,** 925-946.

BHALLA, T. C., SHARMA, M. & SHARMA, N. N. 2009. Nitrile Metabolizing Yeasts. *In:* SATYANARAYANA, T. & KUNZE, G. (eds.) *Yeast Biotechnology: Diversity and Applications.* Dordrecht: Springer Netherlands.

BHATWA, A., WANG, W., HASSAN, Y. I., ABRAHAM, N., LI, X. Z. & ZHOU, T. 2021. Challenges Associated With the Formation of Recombinant Protein Inclusion Bodies in Escherichia coli and Strategies to Address Them for Industrial Applications. *Front Bioeng Biotechnol,* 9**,** 630551.

BLACK, G. W., BROWN, N. L., PERRY, J. J., RANDALL, P. D., TURNBULL, G. & ZHANG, M. 2015. A high-throughput screening method for determining the substrate scope of nitrilases. *Chem Commun (Camb),* 51**,** 2660-2.

BLAKE, J. A. R., PRITCHARD, M., DING, S., SMITH, G. C. M., BURCHELL, B., WOLF, C. R. & FRIEDBERG, T. 1996. Coexpression of a human P450 (CYP3A4) and P450 reductase generates a highly functional monooxygenase system in Escherichia coli. *FEBS Letters,* 397**,** 210-214.

BRANDENBERG, O. F., FASAN, R. & ARNOLD, F. H. 2017. Exploiting and engineering hemoproteins for abiological carbene and nitrene transfer reactions. *Curr Opin Biotechnol,* 47**,** 102-111.

BREWIS, E. A., VAN DER WALT, J. P. & PRIOR, B. A. 1995. The Utilization of Aromatic, Cyclic and Heterocyclic Nitriles by Yeasts. *Systematic and Applied Microbiology,* 18**,** 338-342.

CANTARELLA, L., GALLIFUOCO, A., MALANDRA, A., MARTÍNKOVÁ, L., SPERA, A. & CANTARELLA, M. 2011. High-yield continuous production of nicotinic acid via nitrile hydratase–amidase cascade reactions using cascade CSMRs. *Enzyme and Microbial Technology,* 48**,** 345-350.

CARUGO, O. & PONGOR, S. 2001. A normalized root-mean-square distance for comparing protein three-dimensional structures. *Protein Sci,* 10**,** 1470-3.

CHAN, W. H., SUN, W. Z. & UENG, T. H. 2005. Induction of rat hepatic cytochrome P-450 by ketamine and its toxicological implications. *J Toxicol Environ Health A,* 68**,** 1581-97.

CHEN, J.-T. & CHEN, R.-M. 2010. Mechanisms of ketamine-involved regulation of cytochrome P450 gene expression. *Expert Opinion on Drug Metabolism & Toxicology,* 6**,** 273-281.

CHEN, W., LEE, M.-K., JEFCOATE, C., KIM, S.-C., CHEN, F. & YU, J.-H. 2014. Fungal cytochrome p450 monooxygenases: their distribution, structure, functions, family expansion, and evolutionary origin. *Genome biology and evolution,* 6**,** 1620-1634.

CHENG, D., KELLEY, R. W., CAWLEY, G. F. & BACKES, W. L. 2004. High-level expression of recombinant rabbit cytochrome P450 2E1 in Escherichia coli C41 and its purification. *Protein Expression and Purification,* 33**,** 66-71.

CHHIBA-GOVINDJEE, V. P., VAN DER WESTHUYZEN, C. W., BODE, M. L. & BRADY, D. 2019. Bacterial nitrilases and their regulation. *Applied Microbiology and Biotechnology,* 103**,** 4679-4692.

CHOI, J. & KIM, S.-H. 2017. A genome Tree of Life for the Fungi kingdom. *Proceedings of the National Academy of Sciences,* 114**,** 9391-9396.

CLUNESS, M. J., TURNER, P. D., CLEMENTS, E., BROWN, D. T., APOS & REILLY, C. 1993. Purification and properties of cyanide hydratase from Fusarium lateritium and analysis of the corresponding chy1 gene. *Microbiology,* 139**,** 1807-1815.

COLLINS, D. P. & DAWSON, J. H. 2013. 3.05 - Recent History of Heme-Containing Proteins: Advances in Structure, Functions, and Reaction Intermediate Determination. *In:* REEDIJK, J. & POEPPELMEIER, K. (eds.) *Comprehensive Inorganic Chemistry II (Second Edition).* Amsterdam: Elsevier.

COLLINS, P. A. & KNOWLES, C. J. 1983. The Utilization of Nitriles and Amides by Nocardia rhodochrous. *Microbiology,* 129**,** 711-718.

COOPER DAVID, Y., LEVIN, S., NARASIMHULU, S., ROSENTHAL, O. & ESTABROOK RONALD, W. 1965. Photochemical Action Spectrum of the Terminal Oxidase of Mixed Function Oxidase Systems. *Science,* 147**,** 400-402.

COOPER, D. Y., LEVIN, S., NARASIMHULU, S. & ROSENTHAL, O. 1965. PHOTOCHEMICAL ACTION SPECTRUM OF THE TERMINAL OXIDASE OF MIXED FUNCTION OXIDASE SYSTEMS. *Science,* 147**,** 400-2.

CORREIA, M. A., SINCLAIR, P. R. & DE MATTEIS, F. 2011. Cytochrome P450 regulation: the interplay between its heme and apoprotein moieties in synthesis, assembly, repair, and disposal. *Drug metabolism reviews,* 43**,** 1-26.

CORRELL, C. C., BATIE, C. J., BALLOU, D. P. & LUDWIG, M. L. 1992. Phthalate dioxygenase reductase: a modular structure for electron transfer from pyridine nucleotides to [2Fe-2S]. *Science,* 258**,** 1604-10.

CRYLE, M. J. & DE VOSS, J. J. 2008. The Role of the Conserved Threonine in P450BM3 Oxygen Activation: Substrate-Determined Hydroxylation Activity of the Thr268Ala Mutant. *ChemBioChem,* 9**,** 261-266.

DANIELSON, P. B. 2002. The Cytochrome P450 Superfamily: Biochemistry, Evolution and Drug Metabolism in Humans. *Current Drug Metabolism,* 3**,** 561-597.

DEGTYARENKO, K. & KULIKOVA, T. A. 2001. Evolution of bioinorganic motifs in P450-containing systems. *Biochemical Society transactions,* 29**,** 139-47.

DESANTIS, G., WONG, K., FARWELL, B., CHATMAN, K., ZHU, Z., TOMLINSON, G., HUANG, H., TAN, X., BIBBS, L., CHEN, P., KRETZ, K. & BURK, M. J. 2003. Creation of a Productive, Highly Enantioselective Nitrilase through Gene Site Saturation Mutagenesis (GSSM). *Journal of the American Chemical Society,* 125**,** 11476-11477.

DESTA, Z., MOADDEL, R., OGBURN, E. T., XU, C., RAMAMOORTHY, A., VENKATA, S. L. V., SANGHVI, M., GOLDBERG, M. E., TORJMAN, M. C. & WAINER, I. W. 2012. Stereoselective and regiospecific hydroxylation of ketamine and norketamine. *Xenobiotica; the fate of foreign compounds in biological systems,* 42**,** 1076-1087.

DEVORE, N. M., SMITH, B. D., URBAN, M. J. & SCOTT, E. E. 2008. Key residues controlling phenacetin metabolism by human cytochrome P450 2A enzymes. *Drug metabolism and disposition: the biological fate of chemicals,* 36**,** 2582-2590.

DOBLE, M. & KRUTHIVENTI, A. K. 2007. CHAPTER 4 - Biocatalysis: Green Chemistry. *In:* DOBLE, M. & KRUTHIVENTI, A. K. (eds.) *Green Chemistry and Engineering.* Burlington: Academic Press.

DODHIA, V. R., FANTUZZI, A. & GILARDI, G. 2006. Engineering human cytochrome P450 enzymes into catalytically self-sufficient chimeras using molecular Lego. *J Biol Inorg Chem,* 11**,** 903-16.

DURAIRAJ, P., HUR, J.-S. & YUN, H. 2016. Versatile biocatalysis of fungal cytochrome P450 monooxygenases. *Microbial Cell Factories,* 15**,** 125.

EMOTO, C., MURAYAMA, N., WAKIYA, S. & YAMAZAKI, H. 2009. Effects of Histidine-Tag on Recombinant Human Cytochrome P450 3A5 Catalytic Activity in Reconstitution Systems. *Drug metabolism letters,* 3**,** 207-11.

FEYEREISEN, R. 2019. Insect CYP Genes and P450 Enzymes☆. *Reference Module in Life Sciences.* Elsevier.

FISHER, C. D., LICKTEIG, A. J., AUGUSTINE, L. M., RANGER-MOORE, J., JACKSON, J. P., FERGUSON, S. S. & CHERRINGTON, N. J. 2009. Hepatic cytochrome P450 enzyme alterations in humans with progressive stages of nonalcoholic fatty liver disease. *Drug Metab Dispos,* 37**,** 2087-94.

FREYDANK, A. C., BRANDT, W. & DRÄGER, B. 2008. Protein structure modeling indicates hexahistidine-tag interference with enzyme activity. *Proteins,* 72**,** 173-83.

FROEHLICH, J. E., ITOH, A. & HOWE, G. A. 2001. Tomato allene oxide synthase and fatty acid hydroperoxide lyase, two cytochrome P450s involved in oxylipin metabolism, are targeted to different membranes of chloroplast envelope. *Plant physiology,* 125**,** 306-317.

FUJII-KURIYAMA, Y., MIZUKAMI, Y., KAWAJIRI, K., SOGAWA, K. & MURAMATSU, M. 1982. Primary structure of a cytochrome P-450: coding nucleotide sequence of phenobarbital-inducible cytochrome P-450 cDNA from rat liver. *Proc Natl Acad Sci U S A,* 79**,** 2793-7.

FURUYAMA, K., KANEKO, K. & VARGAS, P. D. 2007. Heme as a magnificent molecule with multiple missions: heme determines its own fate and governs cellular homeostasis. *Tohoku J Exp Med,* 213**,** 1-16.

GILLAM, E. M. J., GUO, Z. Y., MARTIN, M. V., JENKINS, C. M. & GUENGERICH, F. P. 1995. Expression of Cytochrome P450 2D6 in Escherichia coli, Purification, and Spectral and Catalytic Characterization. *Archives of Biochemistry and Biophysics,* 319**,** 540-550.

GONG, J.-S., LU, Z.-M., LI, H., SHI, J.-S., ZHOU, Z.-M. & XU, Z.-H. 2012. Nitrilases in nitrile biocatalysis: recent progress and forthcoming research. *Microbial Cell Factories,* 11**,** 142.

GONG, X.-M., QIN, Z., LI, F.-L., ZENG, B.-B., ZHENG, G.-W. & XU, J.-H. 2019. Development of an Engineered Ketoreductase with Simultaneously Improved Thermostability and Activity for Making a Bulky Atorvastatin Precursor. *ACS Catalysis,* 9**,** 147-153.

GOTTESMAN, S. 1996. PROTEASES AND THEIR TARGETS IN ESCHERICHIA COLI. *Annual Review of Genetics,* 30**,** 465-506.

GUENGERICH, F. P., WATERMAN, M. R. & EGLI, M. 2016. Recent Structural Insights into Cytochrome P450 Function. *Trends in pharmacological sciences,* 37**,** 625-640.

HALL, P. F. 1993. 10 - Molecular Biology of Testicular Steroid Secretion. *In:* DE KRETSER, D. (ed.) *Molecular Biology of the Male Reproductive System.* San Diego: Academic Press.

HAMZA, I. & DAILEY, H. A. 2012. One ring to rule them all: Trafficking of heme and heme synthesis intermediates in the metazoans. *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research,* 1823**,** 1617-1632.

HANNEMANN, F., BICHET, A., EWEN, K. M. & BERNHARDT, R. 2007. Cytochrome P450 systems—biological variations of electron transport chains. *Biochimica et Biophysica Acta (BBA) - General Subjects,* 1770**,** 330-344.

HANUKOGLU, I. 1996. Electron Transfer Proteins of Cytochrome P450 Systems. *In:* BITTAR, E. E. (ed.) *Advances in Molecular and Cell Biology.* Elsevier.

HANUKOGLU, I. & JEFCOATE, C. R. 1980. Mitochondrial cytochrome P-450scc. Mechanism of electron transport by adrenodoxin. *J Biol Chem,* 255**,** 3057-61.

HARPER, D. B. 1977. Fungal degradation of aromatic nitriles. Enzymology of C-N cleavage by Fusarium solani. *Biochemical Journal,* 167**,** 685-692.

HAUSJELL, J., HALBWIRTH, H. & SPADIUT, O. 2018. Recombinant production of eukaryotic cytochrome P450s in microbial cell factories. *Bioscience Reports,* 38.

HAWKES, D. B., ADAMS, G. W., BURLINGAME, A. L., ORTIZ DE MONTELLANO, P. R. & DE VOSS, J. J. 2002. Cytochrome P450cin (CYP176A), Isolation, Expression, and Characterization*. *Journal of Biological Chemistry,* 277**,** 27725-27732.

HOWDEN, A. J. & PRESTON, G. M. 2009. Nitrilase enzymes and their role in plant-microbe interactions. *Microb Biotechnol,* 2**,** 441-51.

HUANG, W., JIA, J., CUMMINGS, J., NELSON, M., SCHNEIDER, G. & LINDQVIST, Y. 1997. Crystal structure of nitrile hydratase reveals a novel iron centre in a novel fold. *Structure,* 5**,** 691-699.

HUNTER, G. A. & FERREIRA, G. C. 2009. 5-aminolevulinate synthase: catalysis of the first step of heme biosynthesis. *Cellular and molecular biology (Noisy-le-Grand, France),* 55**,** 102-110.

ICHINOSE, H., HATAKEYAMA, M. & YAMAUCHI, Y. 2015. Sequence modifications and heterologous expression of eukaryotic cytochromes P450 in Escherichia coli. *Journal of Bioscience and Bioengineering,* 120**,** 268-274.

INGELMAN-SUNDBERG, M. 2004. Human drug metabolising cytochrome P450 enzymes: properties and polymorphisms. *Naunyn Schmiedebergs Arch Pharmacol,* 369**,** 89-104.
JACKSON, C. J., LAMB, D. C., MARCZYLO, T. H., WARRILOW, A. G., MANNING, N. J., LOWE, D. J., KELLY, D. E. & KELLY, S. L. 2002. A novel sterol 14alpha-demethylase/ferredoxin fusion protein (MCCYP51FX) from Methylococcus capsulatus represents a new class of the cytochrome P450 superfamily. *J Biol Chem,* 277**,** 46959-65.
JAMES, J., YARNALL, B., KORANTENG, A., GIBSON, J., RAHMAN, T. & DOYLE, D. A. 2021. Protein over-expression in Escherichia coli triggers adaptation analogous to antimicrobial resistance. *Microb Cell Fact,* 20**,** 13.
JENSEN, V. J. & RUGH, S. 1987. [33] Industrial-scale production and application of immobilized glucose isomerase. *Methods in Enzymology.* Academic Press.
JUNG, S. T., LAUCHLI, R. & ARNOLD, F. H. 2011. Cytochrome P450: taming a wild type enzyme. *Current opinion in biotechnology,* 22**,** 809-817.
KADOKAWA, J.-I. & KOBAYASHI, S. 2010. Polymer synthesis by enzymatic catalysis. *Current Opinion in Chemical Biology,* 14**,** 145-153.
KAPLAN, O., BEZOUŠKA, K., PLÍHAL, O., ETTRICH, R., KULIK, N., VANĚK, O., KAVAN, D., BENADA, O., MALANDRA, A., ŠVEDA, O., VESELÁ, A. B., RINÁGELOVÁ, A., SLÁMOVÁ, K., CANTARELLA, M., FELSBERG, J., DUŠKOVÁ, J., DOHNÁLEK, J., KOTIK, M., KŘEN, V. & MARTÍNKOVÁ, L. 2011. Heterologous expression, purification and characterization of nitrilase from Aspergillus nigerK10. *BMC Biotechnology,* 11**,** 2.
KAPLAN, O., VEJVODA, V., PLÍHAL, O., POMPACH, P., KAVAN, D., BOJAROVÁ, P., BEZOUŠKA, K., MACKOVÁ, M., CANTARELLA, M., JIRKŮ, V., KŘEN, V. & MARTÍNKOVÁ, L. 2006. Purification and characterization of a nitrilase from Aspergillus niger K10. *Applied Microbiology and Biotechnology,* 73**,** 567-575.
KATAGIRI, M., GANGULI, B. N. & GUNSALUS, I. C. 1968. A Soluble Cytochrome P-450 Functional in Methylene Hydroxylation. *Journal of Biological Chemistry,* 243**,** 3543-3546.
KATO, Y., NAKAMURA, K., SAKIYAMA, H., MAYHEW, S. G. & ASANO, Y. 2000. Novel heme-containing lyase, phenylacetaldoxime dehydratase from Bacillus sp. strain OxB-1: purification, characterization, and molecular cloning of the gene. *Biochemistry,* 39**,** 800-9.
KAUSHIK, S., MOHAN, U. & BANERJEE, U. 2012. Exploring residues crucial for nitrilase function by site directed mutagenesis to gain better insight into sequence-function relationships. *Int J Biochem Mol Biol,* 3**,** 384-91.
KELLER, R. M., WÜTHRICH, K. & DEBRUNNER, P. G. 1972. Proton magnetic resonance reveals high-spin iron (II) in ferrous cytochrome P450 cam from Pseudomonas putida. *Proceedings of the National Academy of Sciences of the United States of America,* 69**,** 2073-2075.
KIRK, O., BORCHERT, T. V. & FUGLSANG, C. C. 2002. Industrial enzyme applications. *Current Opinion in Biotechnology,* 13**,** 345-351.
KITAZUME, T., HAINES, D. C., ESTABROOK, R. W., CHEN, B. & PETERSON, J. A. 2007. Obligatory intermolecular electron-transfer from FAD to FMN in dimeric P450BM-3. *Biochemistry,* 46**,** 11892-11901.
KITAZUME, T., TAKAYA, N., NAKAYAMA, N. & SHOUN, H. 2000. Fusarium oxysporum Fatty-acid Subterminal Hydroxylase (CYP505) Is a Membrane-bound Eukaryotic Counterpart of Bacillus megaterium Cytochrome P450BM3*. *Journal of Biological Chemistry,* 275**,** 39734-39740.
KIZAWA, H., TOMURA, D., ODA, M., FUKAMIZU, A., HOSHINO, T., GOTOH, O., YASUI, T. & SHOUN, H. 1991. Nucleotide sequence of the unique nitrate/nitrite-inducible cytochrome P-450 cDNA from Fusarium oxysporum. *J Biol Chem,* 266**,** 10632-7.
KLENK, J. M., NEBEL, B. A., PORTER, J. L., KULIG, J. K., HUSSAIN, S. A., RICHTER, S. M., TAVANTI, M., TURNER, N. J., HAYES, M. A., HAUER, B. & FLITSCH, S. L. 2017. The self-sufficient P450 RhF expressed in a whole cell system selectively catalyses the 5-hydroxylation of diclofenac. *Biotechnol J,* 12.
KOBAYASHI, M., IZUI, H., NAGASAWA, T. & YAMADA, H. 1993. Nitrilase in biosynthesis of the plant hormone indole-3-acetic acid from indole-3-acetonitrile: cloning of the Alcaligenes gene and site-directed mutagenesis of cysteine residues. *Proc Natl Acad Sci U S A,* 90**,** 247-51.
KOBAYASHI, M. & SHIMIZU, S. 1994. Versatile nitrilases: Nitrile-hydrolysing enzymes. *FEMS Microbiology Letters,* 120**,** 217-223.
KOHRS, R. & DURIEUX, M. E. 1998. Ketamine: Teaching an Old Drug New Tricks. *Anesthesia & Analgesia,* 87**,** 1186-1193.
KUFAREVA, I. & ABAGYAN, R. 2012. Methods of protein structure comparison. *Methods Mol Biol,* 857**,** 231-57.
KUMAR, S. 2010. Engineering cytochrome P450 biocatalysts for biotechnology, medicine and bioremediation. *Expert Opin Drug Metab Toxicol,* 6**,** 115-31.
KUMARI, P. & PODDAR, R. 2019. A comparative multivariate analysis of nitrilase enzymes: An ensemble based computational approach. *Comput Biol Chem,* 83**,** 107095.

KUMONDAI, M., HISHINUMA, E., GUTIÉRREZ RICO, E. M., ITO, A., NAKANISHI, Y., SAIGUSA, D., HIRASAWA, N. & HIRATSUKA, M. 2020. Heterologous expression of high-activity cytochrome P450 in mammalian cells. *Scientific Reports,* 10**,** 14193.

KUNST, F., OGASAWARA, N., MOSZER, I., ALBERTINI, A. M., ALLONI, G., AZEVEDO, V., BERTERO, M. G., BESSIÈRES, P., BOLOTIN, A., BORCHERT, S., BORRISS, R., BOURSIER, L., BRANS, A., BRAUN, M., BRIGNELL, S. C., BRON, S., BROUILLET, S., BRUSCHI, C. V., CALDWELL, B., CAPUANO, V., CARTER, N. M., CHOI, S. K., CODANI, J. J., CONNERTON, I. F., CUMMINGS, N. J., DANIEL, R. A., DENIZOT, F., DEVINE, K. M., DÜSTERHÖFT, A., EHRLICH, S. D., EMMERSON, P. T., ENTIAN, K. D., ERRINGTON, J., FABRET, C., FERRARI, E., FOULGER, D., FRITZ, C., FUJITA, M., FUJITA, Y., FUMA, S., GALIZZI, A., GALLERON, N., GHIM, S. Y., GLASER, P., GOFFEAU, A., GOLIGHTLY, E. J., GRANDI, G., GUISEPPI, G., GUY, B. J., HAGA, K., HAIECH, J., HARWOOD, C. R., HÉNAUT, A., HILBERT, H., HOLSAPPEL, S., HOSONO, S., HULLO, M. F., ITAYA, M., JONES, L., JORIS, B., KARAMATA, D., KASAHARA, Y., KLAERR-BLANCHARD, M., KLEIN, C., KOBAYASHI, Y., KOETTER, P., KONINGSTEIN, G., KROGH, S., KUMANO, M., KURITA, K., LAPIDUS, A., LARDINOIS, S., LAUBER, J., LAZAREVIC, V., LEE, S. M., LEVINE, A., LIU, H., MASUDA, S., MAUËL, C., MÉDIGUE, C., MEDINA, N., MELLADO, R. P., MIZUNO, M., MOESTL, D., NAKAI, S., NOBACK, M., NOONE, D., O'REILLY, M., OGAWA, K., OGIWARA, A., OUDEGA, B., PARK, S. H., PARRO, V., POHL, T. M., PORTETELLE, D., PORWOLLIK, S., PRESCOTT, A. M., PRESECAN, E., PUJIC, P., PURNELLE, B., et al. 1997. The complete genome sequence of the Gram-positive bacterium Bacillus subtilis. *Nature,* 390**,** 249-256.

LAMB DAVID, C., FOLLMER ALEC, H., GOLDSTONE JARED, V., NELSON DAVID, R., WARRILOW ANDREW, G., PRICE CLAIRE, L., TRUE MARIE, Y., KELLY STEVEN, L., POULOS THOMAS, L. & STEGEMAN JOHN, J. 2019. On the occurrence of cytochrome P450 in viruses. *Proceedings of the National Academy of Sciences,* 116**,** 12343-12352.

LAMB, D. C., IKEDA, H., NELSON, D. R., ISHIKAWA, J., SKAUG, T., JACKSON, C., OMURA, S., WATERMAN, M. R. & KELLY, S. L. 2003. Cytochrome P450 complement (CYPome) of the avermectin-producer Streptomyces avermitilis and comparison to that of Streptomyces coelicolor A3(2). *Biochemical and Biophysical Research Communications,* 307**,** 610-619.

LAMB, D. C., LEI, L., WARRILOW, A. G., LEPESHEVA, G. I., MULLINS, J. G., WATERMAN, M. R. & KELLY, S. L. 2009. The first virally encoded cytochrome p450. *J Virol,* 83**,** 8266-9.

LAMB, D. C. & WATERMAN, M. R. 2013. Unusual properties of the cytochrome P450 superfamily. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences,* 368**,** 20120434-20120434.

LAYH, N., PARRATT, J. & WILLETTS, A. 1998. Characterization and partial purification of an enantioselective arylacetonitrilase from Pseudomonas fluorescens DSM 7155. *Journal of Molecular Catalysis B: Enzymatic,* 5**,** 467-474.

LEE, S.-J., VAN DER HEIDEN, I. P., GOLDSTEIN, J. A. & VAN SCHAIK, R. H. N. 2007. A new CYP3A5 variant, CYP3A5*11, is shown to be defective in nifedipine metabolism in a recombinant cDNA expression system. *Drug metabolism and disposition: the biological fate of chemicals,* 35**,** 67-71.

LEGRAS, J. L., CHUZEL, G., ARNAUD, A. & GALZY, P. 1990. Natural nitriles and their metabolism. *World J Microbiol Biotechnol,* 6**,** 83-108.

LEWIS, D. F. & PRATT, J. M. 1998. The P450 catalytic cycle and oxygenation mechanism. *Drug metabolism reviews,* 30**,** 739-786.

LI, Q.-S., OGAWA, J., SCHMID ROLF, D. & SHIMIZU, S. 2001. Engineering Cytochrome P450 BM-3 for Oxidation of Polycyclic Aromatic Hydrocarbons. *Applied and Environmental Microbiology,* 67**,** 5735-5739.

LI, Y. C. & CHIANG, J. Y. 1991. The expression of a catalytically active cholesterol 7 alpha-hydroxylase cytochrome P450 in Escherichia coli. *Journal of Biological Chemistry,* 266**,** 19186-19191.

LIN, Y. W., YING, T. L. & LIAO, L. F. 2011. Molecular modeling and dynamics simulation of a histidine-tagged cytochrome b_5. *J Mol Model,* 17**,** 971-8.

LINTON, E. A. & KNOWLES, C. J. 1986. Utilization of Aliphatic Amides and Nitriles by Nocardia rhodochrous LL100-21. *Microbiology,* 132**,** 1493-1501.

LUDWIG-MÜLLER, J. & COHEN, J. D. 2002. Identification and quantification of three active auxins in different tissues of Tropaeolum majus. *Physiol Plant,* 115**,** 320-329.

LUTHRA, A., DENISOV, I. G. & SLIGAR, S. G. 2011. Spectroscopic features of cytochrome P450 reaction intermediates. *Archives of Biochemistry and Biophysics,* 507**,** 26-35.

LYNCH, T. & PRICE, A. 2007. The effect of cytochrome P450 metabolism on drug response, interactions, and adverse effects. *Am Fam Physician,* 76**,** 391-6.

MAIER-GREINER, U. H., OBERMAIER-SKROBRANEK, B. M., ESTERMAIER, L. M., KAMMERLOHER, W., FREUND, C., WÜLFING, C., BURKERT, U. I., MATERN, D. H., BREUER, M., EULITZ, M. & ET AL. 1991. Isolation and

properties of a nitrile hydratase from the soil fungus Myrothecium verrucaria that is highly specific for the fertilizer cyanamide and cloning of its gene. *Proc Natl Acad Sci U S A,* 88**,** 4260-4.
MARTÍNKOVÁ, L. 2019. Nitrile metabolism in fungi: A review of its key enzymes nitrilases with focus on their biotechnological impact. *Fungal Biology Reviews,* 33**,** 149-157.
MARTÍNKOVÁ, L. & KŘEN, V. 2010. Biotransformations with nitrilases. *Current Opinion in Chemical Biology,* 14**,** 130-137.
MATHEW, C. D., NAGASAWA, T., KOBAYASHI, M. & YAMADA, H. 1988. Nitrilase-Catalyzed Production of Nicotinic Acid from 3-Cyanopyridine in Rhodococcus rhodochrous J1. *Appl Environ Microbiol,* 54**,** 1030-2.
MATHIEU, K., JAVED, W., VALLET, S., LESTERLIN, C., CANDUSSO, M.-P., DING, F., XU, X. N., EBEL, C., JAULT, J.-M. & ORELLE, C. 2019. Functionality of membrane proteins overexpressed and purified from E. coli is highly dependent upon the strain. *Scientific Reports,* 9**,** 2654.
MCLEAN, K. J., SABRI, M., MARSHALL, K. R., LAWSON, R. J., LEWIS, D. G., CLIFT, D., BALDING, P. R., DUNFORD, A. J., WARMAN, A. J., MCVEY, J. P., QUINN, A.-M., SUTCLIFFE, M. J., SCRUTTON, N. S. & MUNRO, A. W. 2005. Biodiversity of cytochrome P450 redox systems. *Biochemical Society Transactions,* 33**,** 796-801.
MCLEAN, M. A., MAVES, S. A., WEISS, K. E., KREPICH, S. & SLIGAR, S. G. 1998. Characterization of a Cytochrome P450 from the Acidothermophilic ArchaeaSulfolobus solfataricus. *Biochemical and Biophysical Research Communications,* 252**,** 166-172.
MIROUX, B. & WALKER, J. E. 1996. Over-production of proteins in Escherichia coli: mutant hosts that allow synthesis of some membrane proteins and globular proteins at high levels. *J Mol Biol,* 260**,** 289-98.
MITSUDA, M. & IWASAKI, M. 2006. Improvement in the expression of CYP2B6 by co-expression with molecular chaperones GroES/EL in Escherichia coli. *Protein Expression and Purification,* 46**,** 401-405.
MIURA, M., ITO, K., HAYASHI, M., NAKAJIMA, M., TANAKA, T. & OGURA, S.-I. 2015. The Effect of 5-Aminolevulinic Acid on Cytochrome P450-Mediated Prodrug Activation. *PLOS ONE,* 10**,** e0131793.
MIYANAGA, A., FUSHINOBU, S., ITO, K. & WAKAGI, T. 2001. Crystal Structure of Cobalt-Containing Nitrile Hydratase. *Biochemical and Biophysical Research Communications,* 288**,** 1169-1174.
MIZUTANI, M. 2012. Impacts of Diversification of Cytochrome P450 on Plant Metabolism. *Biological & pharmaceutical bulletin,* 35**,** 824-32.
MUNRO, A. W., GIRVAN, H. M. & MCLEAN, K. J. 2007. Cytochrome P450–redox partner fusion enzymes. *Biochimica et Biophysica Acta (BBA) - General Subjects,* 1770**,** 345-359.
MYLEROVA, V. & MARTINKOVA, L. 2003. Synthetic Applications of Nitrile-Converting Enzymes. *Current Organic Chemistry,* 7**,** 1279-1295.
NAGASAWA, T., KOBAYASHI, M. & YAMADA, H. 1988. Optimum culture conditions for the production of benzonitrilase by Rhodococcus rhodochrous J1. *Archives of Microbiology,* 150**,** 89-94.
NAGASAWA, T., NAKAMURA, T. & YAMADA, H. 1990. ε-caprolactam, a new powerful inducer for the formation of Rhodococcus rhodochrous J1 nitrilase. *Archives of Microbiology,* 155**,** 13-17.
NAGASHIMA, S., NAKASAKO, M., DOHMAE, N., TSUJIMURA, M., TAKIO, K., ODAKA, M., YOHDA, M., KAMIYA, N. & ENDO, I. 1998. Novel non-heme iron center of nitrile hydratase with a claw setting of oxygen atoms. *Nat Struct Biol,* 5**,** 347-51.
NAKAMURA, K., HANNA, I. H., CAI, H., NISHIMURA, Y., WILLIAMS, K. M. & GUENGERICH, F. P. 2001. Coumarin substrates for cytochrome P450 2D6 fluorescence assays. *Anal Biochem,* 292**,** 280-6.
NARHI, L. O. & FULCO, A. J. 1986. Characterization of a catalytically self-sufficient 119,000-dalton cytochrome P-450 monooxygenase induced by barbiturates in Bacillus megaterium. *J Biol Chem,* 261**,** 7160-9.
NEBERT, D. W., WIKVALL, K. & MILLER, W. L. 2013. Human cytochromes P450 in health and disease. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences,* 368**,** 20120431-20120431.
NELSON, D. R. 2018. Cytochrome P450 diversity in the tree of life. *Biochimica et biophysica acta. Proteins and proteomics,* 1866**,** 141-154.
NIU, G., GUO, Q., WANG, J., ZHAO, S., HE, Y. & LIU, L. 2020. Structural basis for plant lutein biosynthesis from α-carotene. *Proc Natl Acad Sci U S A,* 117**,** 14150-14157.
NOBLE, M. A., MILES, C. S., CHAPMAN, S. K., LYSEK, D. A., MACKAY, A. C., REID, G. A., HANZLIK, R. P. & MUNRO, A. W. 1999. Roles of key active-site residues in flavocytochrome P450 BM3. *Biochemical Journal,* 339**,** 371-379.
O'REILLY, C. & TURNER, P. D. 2003. The nitrilase family of CN hydrolysing enzymes – a comparative study. *Journal of Applied Microbiology,* 95**,** 1161-1174.
O'REILLY, E., CORBETT, M., HUSSAIN, S., KELLY, P. P., RICHARDSON, D., FLITSCH, S. L. & TURNER, N. J. 2013. Substrate promiscuity of cytochrome P450 RhF. *Catalysis Science & Technology,* 3**,** 1490-1492.
O'REILLY, E., KÖHLER, V., FLITSCH, S. L. & TURNER, N. J. 2011. Cytochromes P450 as useful biocatalysts: addressing the limitations. *Chemical Communications,* 47**,** 2490-2501.

OMURA, T. & SATO, R. 1964. The Carbon Monoxide-binding Pigment of Liver Microsomes: I. EVIDENCE FOR ITS HEMOPROTEIN NATURE. *Journal of Biological Chemistry,* 239**,** 2370-2378.

OST, T. W., MILES, C. S., MURDOCH, J., CHEUNG, Y., REID, G. A., CHAPMAN, S. K. & MUNRO, A. W. 2000. Rational re-design of the substrate binding site of flavocytochrome P450 BM3. *FEBS Lett,* 486**,** 173-7.

PACE, H. C. & BRENNER, C. 2001. The nitrilase superfamily: classification, structure and function. *Genome Biology,* 2**,** reviews0001.1.

PALLAN, P. S., WANG, C., LEI, L., YOSHIMOTO, F. K., AUCHUS, R. J., WATERMAN, M. R., GUENGERICH, F. P. & EGLI, M. 2015. Human Cytochrome P450 21A2, the Major Steroid 21-Hydroxylase: STRUCTURE OF THE ENZYME·PROGESTERONE SUBSTRATE COMPLEX AND RATE-LIMITING C-H BOND CLEAVAGE. *J Biol Chem,* 290**,** 13128-43.

PAN, Y., ABD-RASHID, B. A., ISMAIL, Z., ISMAIL, R., MAK, J. W. & ONG, C. E. 2011. Heterologous Expression of Human Cytochromes P450 2D6 and CYP3A4 in Escherichia coli and Their Functional Characterization. *The Protein Journal,* 30**,** 581-591.

PARIKH, A., GILLAM, E. M. & GUENGERICH, F. P. 1997. Drug metabolism by Escherichia coli expressing human cytochromes P450. *Nat Biotechnol,* 15**,** 784-8.

PEARSON, W. R. & SIERK, M. L. 2005. The limits of protein sequence comparison? *Curr Opin Struct Biol,* 15**,** 254-60.

PELKONEN, O., RAUTIO, A., RAUNIO, H. & PASANEN, M. 2000. CYP2A6: a human coumarin 7-hydroxylase. *Toxicology,* 144**,** 139-147.

PELKONEN, O., TURPEINEN, M., HAKKOLA, J., HONKAKOSKI, P., HUKKANEN, J. & RAUNIO, H. 2008. Inhibition and induction of human cytochrome P450 enzymes: current status. *Arch Toxicol,* 82**,** 667-715.

PELTONIEMI, M. 2013. *Effects of cytochrome P450 enzyme inhibitors and inducers on the metabolism of S-ketamine.* Doctorate Doctoral Dissertation, UNIVERSITY OF TURKU.

PERNECKY, S. J., LARSON, J. R., PHILPOT, R. M. & COON, M. J. 1993. Expression of truncated forms of liver microsomal P450 cytochromes 2B4 and 2E1 in Escherichia coli: influence of NH2-terminal region on localization in cytosol and membranes. *Proceedings of the National Academy of Sciences of the United States of America,* 90**,** 2651-2655.

PIKULEVA, I. A. & WATERMAN, M. R. 2013. Cytochromes P450: Roles in diseases. *Journal of Biological Chemistry,* 288**,** 17091-17098.

PIOTROWSKI, M. 2008. Primary or secondary? Versatile nitrilases in plant metabolism. *Phytochemistry,* 69**,** 2655-2667.

PIOTROWSKI, M., SCHÖNFELDER, S. & WEILER, E. W. 2001. The Arabidopsis thaliana isogene NIT4 and its orthologs in tobacco encode beta-cyano-L-alanine hydratase/nitrilase. *J Biol Chem,* 276**,** 2616-21.

PODAR, M., EADS, J. R. & RICHARDSON, T. H. 2005. Evolution of a microbial nitrilase gene family: a comparative and environmental genomics study. *BMC Evolutionary Biology,* 5**,** 42.

PORTMANN, S., KWAN, H. Y., THEURILLAT, R., SCHMITZ, A., MEVISSEN, M. & THORMANN, W. 2010a. Enantioselective capillary electrophoresis for identification and characterization of human cytochrome P450 enzymes which metabolize ketamine and norketamine in vitro. *J Chromatogr A,* 1217**,** 7942-8.

PORTMANN, S., KWAN, H. Y., THEURILLAT, R., SCHMITZ, A., MEVISSEN, M. & THORMANN, W. 2010b. Enantioselective capillary electrophoresis for identification and characterization of human cytochrome P450 enzymes which metabolize ketamine and norketamine in vitro. *Journal of Chromatography A,* 1217**,** 7942-7948.

POULOS, T. L., FINZEL, B. C., GUNSALUS, I. C., WAGNER, G. C. & KRAUT, J. 1985. The 2.6-A crystal structure of Pseudomonas putida cytochrome P-450. *Journal of Biological Chemistry,* 260**,** 16122-16130.

POULOS, T. L., FINZEL, B. C. & HOWARD, A. J. 1987. High-resolution crystal structure of cytochrome P450cam. *Journal of Molecular Biology,* 195**,** 687-700.

POULOS, T. L. & JOHNSON, E. F. 2015. Structures of Cytochrome P450 Enzymes. *In:* ORTIZ DE MONTELLANO, P. R. (ed.) *Cytochrome P450: Structure, Mechanism, and Biochemistry.* Cham: Springer International Publishing.

PRITCHARD, M. P., OSSETIAN, R., LI, D. N., HENDERSON, C. J., BURCHELL, B., WOLF, C. R. & FRIEDBERG, T. 1997. A general strategy for the expression of recombinant human cytochrome P450s in Escherichia coli using bacterial signal peptides: expression of CYP3A4, CYP2A6, and CYP2E1. *Arch Biochem Biophys,* 345**,** 342-54.

PUCHKAEV, A. V. & ORTIZ DE MONTELLANO, P. R. 2005. The Sulfolobus solfataricus electron donor partners of thermophilic CYP119: an unusual non-NAD(P)H-dependent cytochrome P450 system. *Arch Biochem Biophys,* 434**,** 169-77.

PUCHKAEV, A. V., WAKAGI, T. & ORTIZ DE MONTELLANO, P. R. 2002. CYP119 plus a Sulfolobus tokodaii strain 7 ferredoxin and 2-oxoacid:ferredoxin oxidoreductase constitute a high-temperature cytochrome P450 catalytic system. *J Am Chem Soc,* 124**,** 12682-3.

PYLYPENKO, O. & SCHLICHTING, I. 2004. Structural aspects of ligand binding to and electron transfer in bacterial and fungal P450s. *Annu Rev Biochem,* 73**,** 991-1018.

QUEHL, P., HOLLENDER, J., SCHÜÜRMANN, J., BROSSETTE, T., MAAS, R. & JOSE, J. 2016. Co-expression of active human cytochrome P450 1A2 and cytochrome P450 reductase on the cell surface of Escherichia coli. *Microbial cell factories,* 15**,** 26-26.

RAN, N., ZHAO, L., CHEN, Z. & TAO, J. 2008. Recent applications of biocatalysis in developing green chemistry for chemical synthesis at the industrial scale. *Green Chemistry,* 10**,** 361-372.

RAWAL, S., YIP, S. S. M. & COULOMBE, R. A. 2010. Cloning, Expression and Functional Characterization of Cytochrome P450 3A37 from Turkey Liver with High Aflatoxin B1 Epoxidation Activity. *Chemical Research in Toxicology,* 23**,** 1322-1329.

RENDIC, S. & GUENGERICH, F. P. 2015. Survey of Human Oxidoreductases and Cytochrome P450 Enzymes Involved in the Metabolism of Xenobiotic and Natural Chemicals. *Chemical research in toxicology,* 28**,** 38-42.

RENTMEISTER, A., BROWN, T. R., SNOW, C. D., CARBONE, M. N. & ARNOLD, F. H. 2011. Engineered Bacterial Mimics of Human Drug Metabolizing Enzyme CYP2C9. *ChemCatChem,* 3**,** 1065-1071.

REY, P., ROSSI, J.-C., TAILLADES, J., GROS, G. & NORE, O. 2004. Hydrolysis of Nitriles Using an Immobilized Nitrilase: Applications to the Synthesis of Methionine Hydroxy Analogue Derivatives. *Journal of Agricultural and Food Chemistry,* 52**,** 8155-8162.

RITTLE, J., YOUNKER, J. M. & GREEN, M. T. 2010. Cytochrome P450: The Active Oxidant and Its Spectrum. *Inorganic Chemistry,* 49**,** 3610-3617.

ROBERTS, G. A., GROGAN, G., GRETER, A., FLITSCH, S. L. & TURNER, N. J. 2002. Identification of a new class of cytochrome P450 from a Rhodococcus sp. *J Bacteriol,* 184**,** 3898-908.

ROBIN, A., ROBERTS, G. A., KISCH, J., SABBADIN, F., GROGAN, G., BRUCE, N., TURNER, N. J. & FLITSCH, S. L. 2009. Engineering and improvement of the efficiency of a chimeric [P450cam-RhFRed reductase domain] enzyme. *Chemical Communications***,** 2478-2480.

ROBINSON, S. L., PIEL, J. & SUNAGAWA, S. 2021. A roadmap for metagenomic enzyme discovery. *Nat Prod Rep,* 38**,** 1994-2023.

ROBINSON, W. G. & HOOK, R. H. 1964. Ricinine Nitrilase: I. REACTION PRODUCT AND SUBSTRATE SPECIFICITY. *Journal of Biological Chemistry,* 239**,** 4257-4262.

RUSTLER, S. & STOLZ, A. 2007. Isolation and characterization of a nitrile hydrolysing acidotolerant black yeast—Exophiala oligosperma R1. *Applied Microbiology and Biotechnology,* 75**,** 899-908.

RYLOTT, E. L., JACKSON, R. G., EDWARDS, J., WOMACK, G. L., SETH-SMITH, H. M. B., RATHBONE, D. A., STRAND, S. E. & BRUCE, N. C. 2006. An explosive-degrading cytochrome P450 activity and its targeted application for the phytoremediation of RDX. *Nature Biotechnology,* 24**,** 216-219.

SABIA, M., HIRSH, R. A., TORJMAN, M. C., WAINER, I. W., COOPER, N., DOMSKY, R. & GOLDBERG, M. E. 2011. Advances in translational neuropathic research: example of enantioselective pharmacokinetic-pharmacodynamic modeling of ketamine-induced pain relief in complex regional pain syndrome. *Current pain and headache reports,* 15**,** 207-214.

SAKURAI, H., TSUCHIYA, K. & MIGITA, K. 1988. An active-site model for nitrile hydratase: axially coordinate non-heme iron complexes in the low-spin ferric state. *Inorganic Chemistry,* 27**,** 3877-3879.

SANGER, F., NICKLEN, S. & COULSON, A. R. 1977. DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci U S A,* 74**,** 5463-7.

SCHULZE, B. & WUBBOLTS, M. G. 1999. Biocatalysis for industrial production of fine chemicals. *Current Opinion in Biotechnology,* 10**,** 609-615.

SCOTT, E. E., SPATZENEGGER, M. & HALPERT, J. R. 2001. A Truncation of 2B Subfamily Cytochromes P450 Yields Increased Expression Levels, Increased Solubility, and Decreased Aggregation While Retaining Function. *Archives of Biochemistry and Biophysics,* 395**,** 57-68.

SERIZAWA, N. & MATSUOKA, T. 1991. 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. *Biochim Biophys Acta,* 1084**,** 35-40.

SETH-SMITH, H. M., ROSSER, S. J., BASRAN, A., TRAVIS, E. R., DABBS, E. R., NICKLIN, S. & BRUCE, N. C. 2002. Cloning, sequencing, and characterization of the hexahydro-1,3,5-Trinitro-1,3,5-triazine degradation gene cluster from Rhodococcus rhodochrous. *Appl Environ Microbiol,* 68**,** 4764-71.

SEWELL, B. T., BERMAN, M. N., MEYERS, P. R., JANDHYALA, D. & BENEDIK, M. J. 2003. The Cyanide Degrading Nitrilase from Pseudomonas stutzeri AK61 Is a Two-Fold Symmetric, 14-Subunit Spiral. *Structure,* 11**,** 1413-1422.

SHARMA, N., VERMA, R., SAVITRI & BHALLA, T. C. 2018. Classifying nitrilases as aliphatic and aromatic using machine learning technique. *3 Biotech,* 8**,** 68.

SHELDON, R. A., BRADY, D. & BODE, M. L. 2020. The Hitchhiker's guide to biocatalysis: recent advances in the use of enzymes in organic synthesis. *Chemical science,* 11**,** 2587-2605.

SHELDON, R. A. & RANTWIJK, F. V. 2004. Biocatalysis for Sustainable Organic Synthesis. *Australian Journal of Chemistry,* 57**,** 281-289.

SHEN, J.-D., CAI, X., LIU, Z.-Q. & ZHENG, Y.-G. 2021. Nitrilase: a promising biocatalyst in industrial applications for green chemistry. *Critical Reviews in Biotechnology,* 41**,** 72-93.

SINGH, R., SHARMA, R., TEWARI, N. & RAWAT, D. S. 2006. Nitrilase and its application as a 'green' catalyst. *Chem Biodivers,* 3**,** 1279-87.

SMITH, B. D., SANDERS, J. L., PORUBSKY, P. R., LUSHINGTON, G. H., STOUT, C. D. & SCOTT, E. E. 2007. Structure of the Human Lung Cytochrome P450 2A13. *Journal of Biological Chemistry,* 282**,** 17306-17313.

SONO, M., ROACH, M. P., COULTER, E. D. & DAWSON, J. H. 1996. Heme-Containing Oxygenases. *Chemical Reviews,* 96**,** 2841-2888.

STUDIER, F. W. 2005. Protein production by auto-induction in high-density shaking cultures. *Protein Expression and Purification,* 41**,** 207-234.

STUDIER, F. W. & MOFFATT, B. A. 1986. Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes. *J Mol Biol,* 189**,** 113-30.

SUGIMOTO, H., SHINKYO, R., HAYASHI, K., YONEDA, S., YAMADA, M., KAMAKURA, M., IKUSHIRO, S.-I., SHIRO, Y. & SAKAKI, T. 2008. Crystal Structure of CYP105A1 (P450SU-1) in Complex with 1α,25-Dihydroxyvitamin D3. *Biochemistry,* 47**,** 4017-4027.

SUGISHIMA, M., SATO, H., WADA, K. & YAMAMOTO, K. 2019. Crystal structure of a NADPH-cytochrome P450 oxidoreductase (CYPOR) and heme oxygenase 1 fusion protein implies a conformational change in CYPOR upon NADPH/NADP(+) binding. *FEBS Lett,* 593**,** 868-875.

SUGIURA, Y., KUWAHARA, J., NAGASAWA, T. & YAMADA, H. 1987. Nitrile hydratase. The first non-heme iron enzyme with a typical low-spin iron(III)-active center. *Journal of the American Chemical Society,* 109**,** 5848-5850.

SWENSON, S. A., MOORE, C. M., MARCERO, J. R., MEDLOCK, A. E., REDDI, A. R. & KHALIMONCHUK, O. 2020. From Synthesis to Utilization: The Ins and Outs of Mitochondrial Heme. *Cells,* 9**,** 579.

TAKAYA, N., SUZUKI, S., KUWAZAKI, S., SHOUN, H., MARUO, F., YAMAGUCHI, M. & TAKEO, K. 1999. Cytochrome P450nor, a Novel Class of Mitochondrial Cytochrome P450 Involved in Nitrate Respiration in the Fungus Fusarium oxysporum. *Archives of Biochemistry and Biophysics,* 372**,** 340-346.

TALMANN, L., WIESNER, J. & VILCINSKAS, A. 2017. Strategies for the construction of insect P450 fusion enzymes. *Zeitschrift für Naturforschung C.*

TANG, Z., SALAMANCA-PINZÓN, S. G., WU, Z.-L., XIAO, Y. & GUENGERICH, F. P. 2010. Human cytochrome P450 4F11: Heterologous expression in bacteria, purification, and characterization of catalytic function. *Archives of Biochemistry and Biophysics,* 494**,** 86-93.

TAVOLIERI, A. M., MURRAY, D. T., ASKENASY, I., PENNINGTON, J. M., MCGARRY, L., STANLEY, C. B. & STROUPE, M. E. 2019. NADPH-dependent sulfite reductase flavoprotein adopts an extended conformation unique to this diflavin reductase. *J Struct Biol,* 205**,** 170-179.

THIELGES, M. C., CHUNG, J. K., AXUP, J. Y. & FAYER, M. D. 2011. Influence of histidine tag attachment on picosecond protein dynamics. *Biochemistry,* 50**,** 5799-805.

THIMANN, K. V. & MAHADEVAN, S. 1964. Nitrilase: I. Occurrence, preparation, and general properties of the enzyme. *Archives of Biochemistry and Biophysics,* 105**,** 133-141.

THUKU, R. N., BRADY, D., BENEDIK, M. J. & SEWELL, B. T. 2009. Microbial nitrilases: versatile, spiral forming, industrial enzymes. *Journal of Applied Microbiology,* 106**,** 703-727.

TSCHIRRET-GUTH, R. A., KOO, L. S., HUI BON HOA, G. & ORTIZ DE MONTELLANO, P. R. 2001. Reversible Pressure Deformation of A Thermophilic Cytochrome P450 Enzyme (CYP119) and Its Active-Site Mutants. *Journal of the American Chemical Society,* 123**,** 3412-3417.

URLACHER, V. B. & GIRHARD, M. 2012. Cytochrome P450 monooxygenases: an update on perspectives for synthetic application. *Trends in Biotechnology,* 30**,** 26-36.

URLACHER, V. B. & GIRHARD, M. 2019. Cytochrome P450 Monooxygenases in Biotechnology and Synthetic Biology. *Trends Biotechnol,* 37**,** 882-897.

VESELÁ, A. B., FRANC, M., PELANTOVÁ, H., KUBÁČ, D., VEJVODA, V., ŠULC, M., BHALLA, T. C., MACKOVÁ, M., LOVECKÁ, P., JANŮ, P., DEMNEROVÁ, K. & MARTÍNKOVÁ, L. 2010. Hydrolysis of benzonitrile herbicides by soil actinobacteria and metabolite toxicity. *Biodegradation,* 21**,** 761-770.

VORWERK, S., BIERNACKI, S., HILLEBRAND, H., JANZIK, I., MÜLLER, A., WEILER, E. W. & PIOTROWSKI, M. 2001. Enzymatic characterization of the recombinant Arabidopsis thaliana nitrilase subfamily encoded by the NIT2/NIT1/NIT3-gene cluster. *Planta,* 212**,** 508-16.

WANG, M.-X. 2005. Enantioselective Biotransformations of Nitriles in Organic Synthesis. *Topics in Catalysis,* 35**,** 117-130.

WANG, M.-X. 2015. Enantioselective Biotransformations of Nitriles in Organic Synthesis. *Accounts of Chemical Research,* 48**,** 602-611.

WANG, M., ROBERTS DAVID, L., PASCHKE, R., SHEA THOMAS, M., MASTERS BETTIE SUE, S. & KIM JUNG-JA, P. 1997. Three-dimensional structure of NADPH–cytochrome P450 reductase: Prototype for FMN- and FAD-containing enzymes. *Proceedings of the National Academy of Sciences,* 94**,** 8411-8416.

WASKELL, L. & KIM, J.-J. P. 2015. Electron Transfer Partners of Cytochrome P450. *In:* ORTIZ DE MONTELLANO, P. R. (ed.) *Cytochrome P450: Structure, Mechanism, and Biochemistry.* Cham: Springer International Publishing.

WATANABE, A., YANO, K., IKEBUKURO, K. & KARUBE, I. 1998. Cloning and expression of a gene encoding cyanidase from Pseudomonas stutzeri AK61. *Appl Microbiol Biotechnol,* 50**,** 93-7.

WATANABE, I., NARA, F. & SERIZAWA, N. 1995. Cloning, characterization and expression of the gene encoding cytochrome P-450sca-in2 from Streptomyces carbophilus involved in production of pravastatin, a specific HMG-CoA reductase inhibitor. *Gene,* 163**,** 81-85.

WATERMAN, M. R. 1993. HETEROLOGOUS EXPRESSION OF CYTOCHROME-P-450 IN ESCHERICHIA-COLI. *Biochemical Society Transactions,* 21**,** 1081-1085.

WAXMAN, D. J. & WALSH, C. 1982. Phenobarbital-induced rat liver cytochrome P-450. Purification and characterization of two closely related isozymic forms. *Journal of Biological Chemistry,* 257**,** 10446-10457.

WERCK-REICHHART, D. & FEYEREISEN, R. 2000. Cytochromes P450: a success story. *Genome Biology,* 1**,** reviews3003.1.

WHITEHOUSE, C., BELL, S. & WONG, L. 2012a. P450(BM3) (CYP102A1): connecting the dots. *Chemical Society reviews,* 41**,** 1218-60.

WHITEHOUSE, C. J. C., BELL, S. G. & WONG, L.-L. 2012b. P450BM3 (CYP102A1): connecting the dots. *Chemical Society Reviews,* 41**,** 1218-1260.

WIEDMANN, M., KURZCHALIA, T. V., HARTMANN, E. & RAPOPORT, T. A. 1987. A signal sequence receptor in the endoplasmic reticulum membrane. *Nature,* 328**,** 830-833.

WILLIAMS, P. A., COSME, J., SRIDHAR, V., JOHNSON, E. F. & MCREE, D. E. 2000. Mammalian Microsomal Cytochrome P450 Monooxygenase: Structural Adaptations for Membrane Binding and Functional Diversity. *Molecular Cell,* 5**,** 121-131.

WRIGHT, R. L., HARRIS, K., SOLOW, B., WHITE, R. H. & KENNELLY, P. J. 1996. Cloning of a potential cytochrome P450 from the archaeon Sulfolobus solfataricus. *FEBS Lett,* 384**,** 235-9.

XU, L.-H. & DU, Y.-L. 2018. Rational and semi-rational engineering of cytochrome P450s for biotechnological applications. *Synthetic and Systems Biotechnology,* 3**,** 283-290.

YAMADA, H. & KOBAYASHI, M. 1996. Nitrile Hydratase and Its Application to Industrial Production of Acrylamide. *Bioscience, Biotechnology, and Biochemistry,* 60**,** 1391-1400.

YIEN, Y. Y. & PERFETTO, M. 2022. Regulation of Heme Synthesis by Mitochondrial Homeostasis Proteins. *Front Cell Dev Biol,* 10**,** 895521.

ZANGAR, R. C., DAVYDOV, D. R. & VERMA, S. 2004. Mechanisms that regulate production of reactive oxygen species by cytochrome P450. *Toxicology and Applied Pharmacology,* 199**,** 316-331.

ZANGER, U. M. & SCHWAB, M. 2013. Cytochrome P450 enzymes in drug metabolism: Regulation of gene expression, enzyme activities, and impact of genetic variation. *Pharmacology & Therapeutics,* 138**,** 103-141.

ZELASKO, S., PALARIA, A. & DAS, A. 2013. Optimizations to achieve high-level expression of cytochrome P450 proteins using Escherichia coli expression systems. *Protein Expression and Purification,* 92**,** 77-87.

ZHANG, L., XIE, Z., LIU, Z., ZHOU, S., MA, L., LIU, W., HUANG, J.-W., KO, T.-P., LI, X., HU, Y., MIN, J., YU, X., GUO, R.-T. & CHEN, C.-C. 2020. Structural insight into the electron transfer pathway of a self-sufficient P450 monooxygenase. *Nature Communications,* 11**,** 2676.

ZHANG, L., YIN, B., WANG, C., JIANG, S., WANG, H., YUAN, Y. A. & WEI, D. 2014. Structural insights into enzymatic activity and substrate specificity determination by a single amino acid in nitrilase from Syechocystis sp. PCC6803. *J Struct Biol,* 188**,** 93-101.

ZHANG, Q., WU, Z.-M., HAO, C.-L., TANG, X.-L., ZHENG, R.-C. & ZHENG, Y.-G. 2019. Highly regio- and enantioselective synthesis of chiral intermediate for pregabalin using one-pot bienzymatic cascade of nitrilase and amidase. *Applied Microbiology and Biotechnology,* 103**,** 5617-5626.

ZHAO, Y. & HALPERT, J. R. 2007. Structure–function analysis of cytochromes P450 2B. *Biochimica et Biophysica Acta (BBA) - General Subjects,* 1770**,** 402-412.

Appendices

Appendix 1- Cloning plasmids and primers

Figure 1.1 pET22a-c plasmid map-Novagen

pET-22b(+) Vector

TB038 12/98

The pET-22b(+) vector (Cat. No. 69744-3) carries an N-terminal *pelB* signal sequence for potential periplasmic localization, plus optional C-terminal His•Tag® sequence. Unique sites are shown on the circle map. Note that the sequence is numbered by the pBR322 convention, so the T7 expression region is reversed on the circular map. The cloning/expression region of the coding strand transcribed by T7 RNA polymerase is shown below. The f1 origin is oriented so that infection with helper phage will produce virions containing single-stranded DNA that corresponds to the coding strand. Therefore, single-stranded sequencing should be performed using the T7 terminator primer (Cat. No. 69337-3).

pET-22b(+) sequence landmarks	
T7 promoter	361-377
T7 transcription start	360
pelB coding sequence	224-289
Multiple cloning sites	
(*Nco* I - *Xho* I)	158-225
His•Tag coding sequence	140-157
T7 terminator	26-72
lacI coding sequence	764-1843
pBR322 origin	3277
bla coding sequence	4038-4895
f1 origin	5027-5482

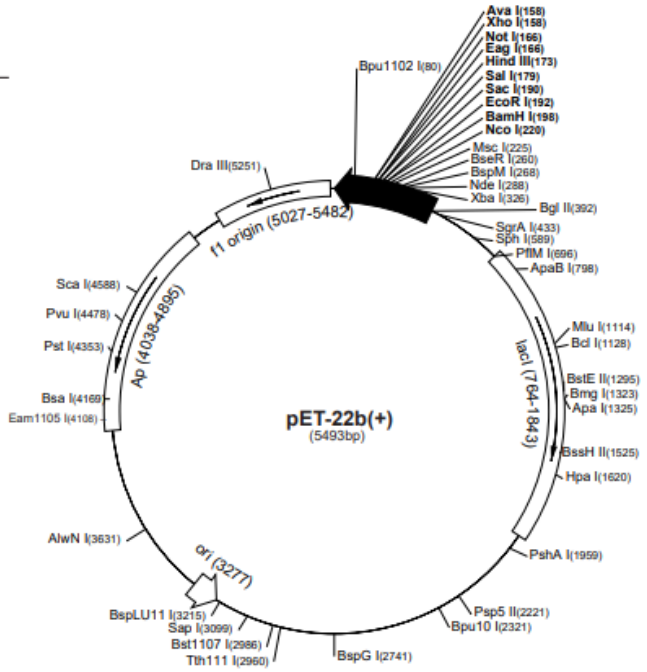


Figure 1.2 pET29a-c plasmid map – Novagen

pET-29a-c(+) Vectors

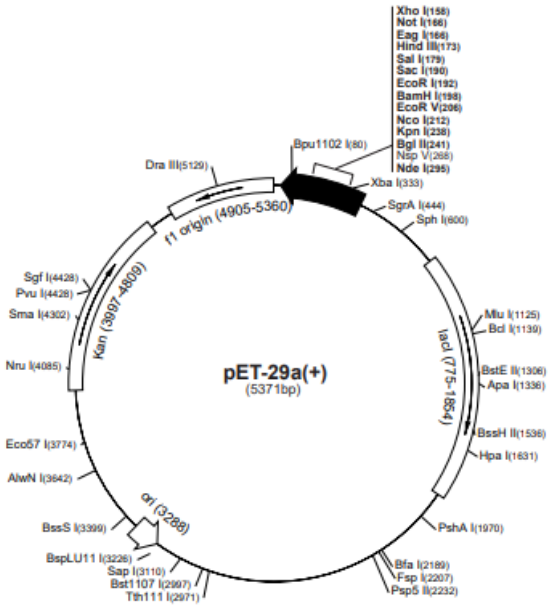
TB076 12/98

	Cat. No.
pET-29a DNA	69871-3
pET-29b DNA	69872-3
pET-29c DNA	69873-3

The pET-29a-c(+) vectors carry an N-terminal S•Tag™/thrombin configuration plus an optional C-terminal His•Tag® sequence. Unique sites are shown on the circle map. Note that the sequence is numbered by the pBR322 convention, so the T7 expression region is reversed on the circular map. The cloning/expression region of the coding strand transcribed by T7 RNA polymerase is shown below. The f1 origin is oriented so that infection with helper phage will produce virions containing single-stranded DNA that corresponds to the coding strand. Therefore, single-stranded sequencing should be performed using the T7 terminator primer (Cat. No. 69337-3).

pET-29a(+) sequence landmarks	
T7 promoter	368-384
T7 transcription start	367
S•Tag coding sequence	249-293
Multiple cloning sites (*Nco* I - *Xho* I)	158-217
His•Tag coding sequence	140-157
T7 terminator	26-72
lacI coding sequence	775-1854
pBR322 origin	3288
Kan coding sequence	3997-4809
f1 origin	4905-5360

The maps for pET-29b(+) and pET-29c(+) are the same as pET-29a(+) (shown) with the following exceptions: pET-29b(+) is a 5370bp plasmid; subtract 1bp from each site beyond *Bam*H I at 198. pET-29c(+) is a 5372bp plasmid; add 1bp to each site beyond *Bam*H I at 198.



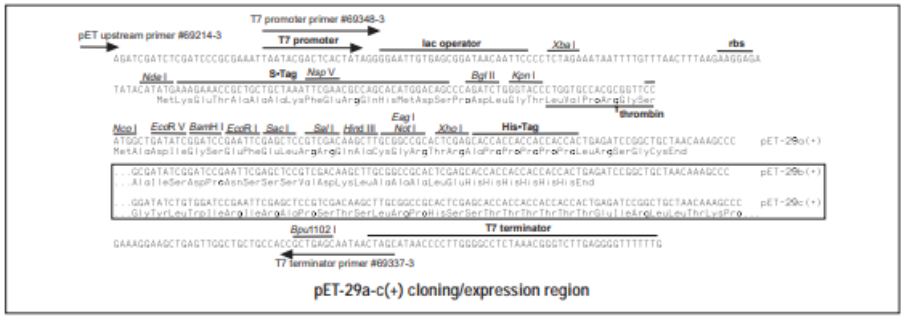


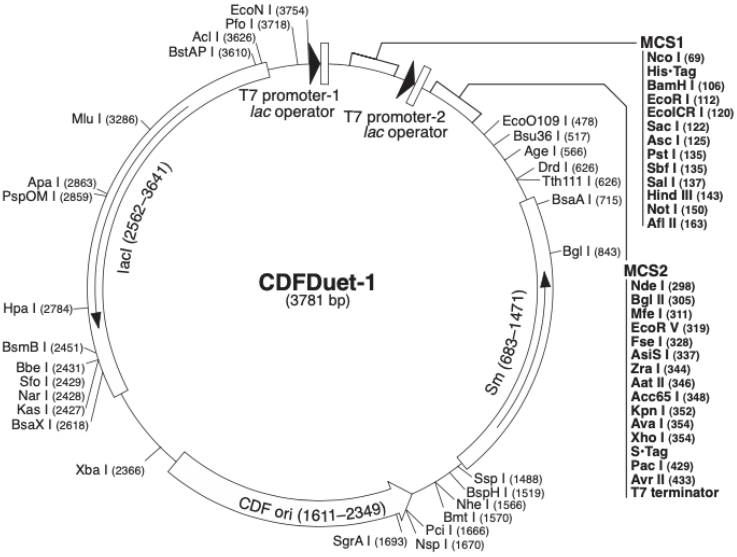
Figure 1.3 pCDFDuet-1 plasmid map-Novagen

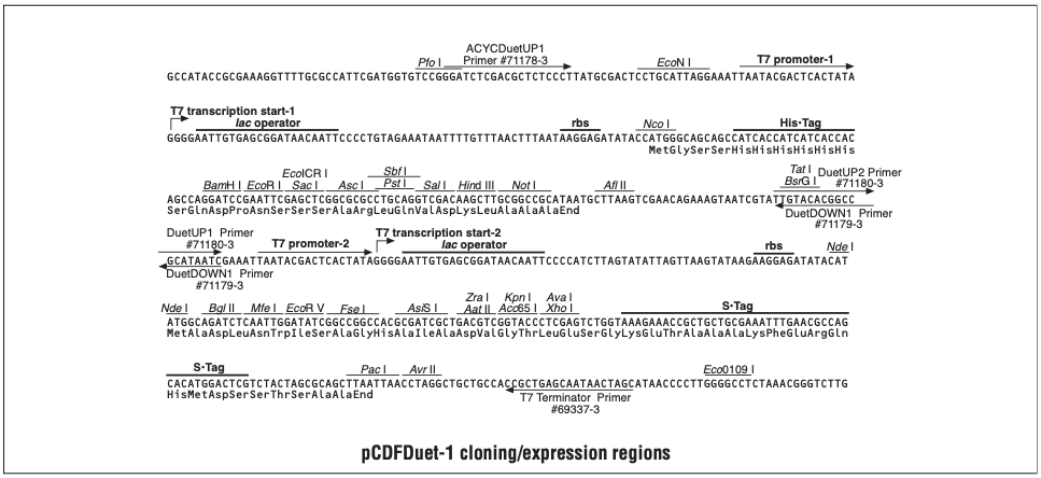
pCDFDuet-1 Vector

TB390 0903

	Cat. No.
pCDFDuet-1 DNA	71340-3
pCDFDuet-1 sequence landmarks	
T7 promoter-1	3765–3781
T7 transcription start-1	1
His•Tag® coding sequence	83–100
Multiple cloning sites-1 (*Nco* I–*Afl* II)	69–168
T7 promoter-2	214–230
T7 transcription start-2	231
Multiple cloning sites-2 (*Nde* I–*Avr* II)	297–438
S•Tag™ coding sequence	366–410
T7 terminator	462–509
aadA (SmR) coding sequence	683–1471
CDF origin	1611–2349
lacI coding sequence	2562–3641

pCDFDuet™-1 is designed for the coexpression of two target ORFs. The vector contains two multiple cloning sites (MCS), each of which is preceded by a T7*lac* promoter and ribosome binding site (rbs). The vector also carries the CloDF13-derived CDF replicon, *lacI* gene and streptomycin/spectinomycin resistance gene (Sm^R). This vector can be used in combination with pACYCDuet™-1 (Cat. No. 71147-3), pRSFDuet™-1 (Cat. No. 71341-3), and pETDuet™-1 (Cat. No. 71146-3) in an appropriate host strain for the coexpression of 4 to 8 target proteins. ORFs inserted into MCS1 can be sequenced using the ACYCDuetUP1 Primer (Cat. No. 71178-3) and DuetDOWN1 Primer (Cat. No. 71179-3). ORFs inserted into MCS2 can be sequenced using the DuetUP2 Primer (Cat. No. 71180-3) and T7 Terminator Primer (Cat. No. 69337-3). Unique sites are shown on the circle map.





pCDFDuet-1 cloning/expression regions

Figure 1.4 pCWOri^{+} plasmid map-Addgene

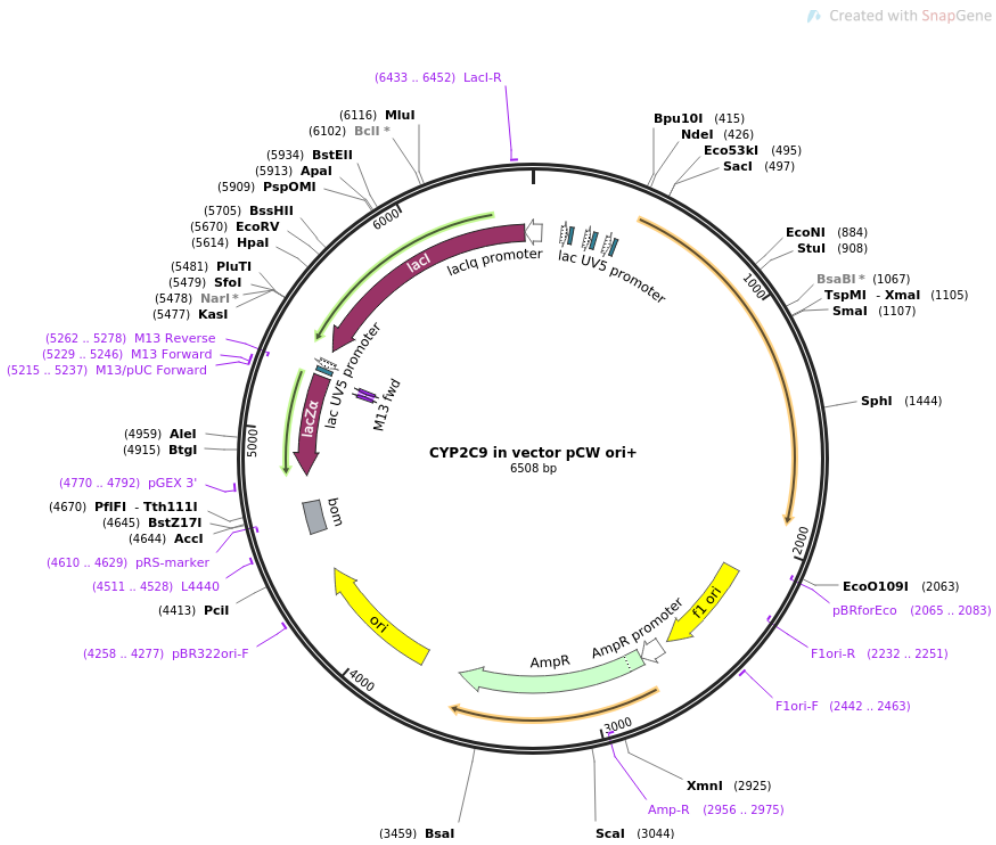


Figure 1.5 pB216 plasmid map- Blake at al. (1996)

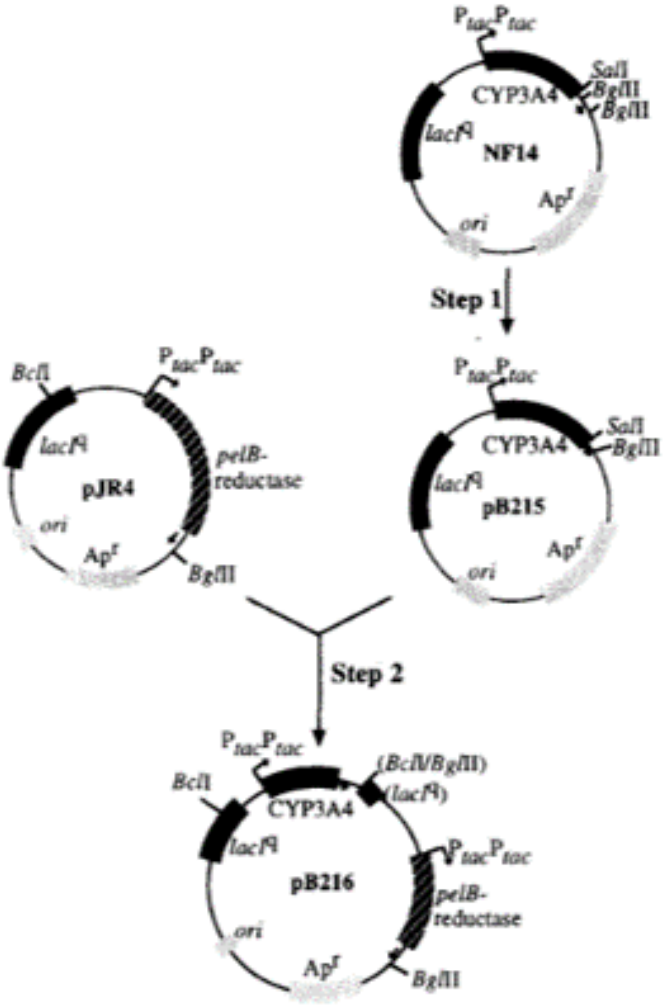


Figure 1.6 Co-cloning of P450s and CPR with the linker in pCWOri+

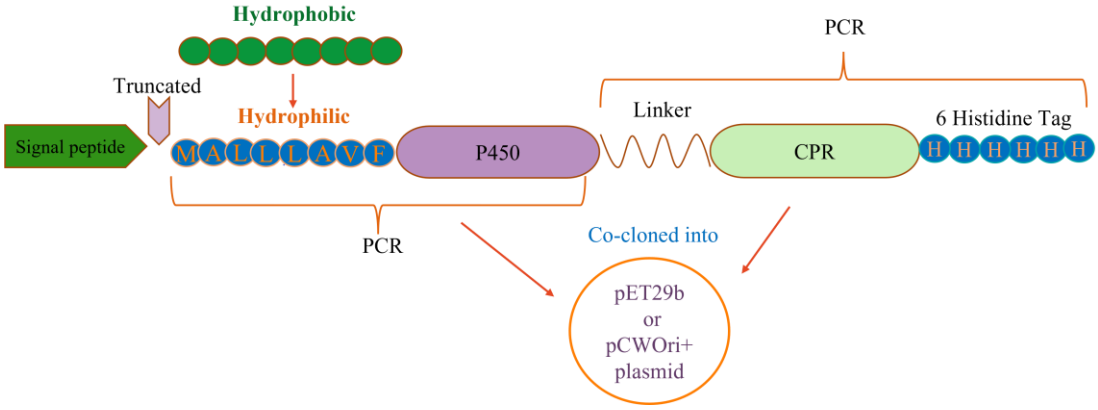


Figure 1.6 Representation of co-cloning strategy of P450s and CPR in pCWOri^{+}/ pET29b plasmids. As can be seen from the figure above, the signal peptide was truncated for all human P450s, the hydrophobic stretch was replaced with a hydrophilic one for better solubility, then the P450 was cloned into the plasmids first. The plasmid was digested again and the CPR with incorporated a His Tag (for pCWOri^{+}) was cloned in.

Table 1.1 Generated for co-cloning of P450 with the CPR

In the first cloning strategy, the human P450s were co-cloned with the human reductase (CPR) in pET29b and pCWOri^{+} plasmids and were co-expressed as fusion proteins in different *E. coli* strains.

CYP Form	Cloning Plasmid	Nucleotide sequence of primers
		The forward primer (*NdeI**)
		5′ TGCCATAG CAT **ATG GCTAAAAAAACATCCTCA** 3′
	pCWOri^{+}	The reverse primer (*AvrII**)
		5′ TGCCATAG CCTAGG **ACGAGGCAAGAAGGACATTGT** 3′
CP2A6		The reverse primer (*HindIII**)
		5′TGCCATAG AAGCTT **ACGAGGCAAGAAGGACATTG** 3′
	pET29b	The forward primer (*NdeI**)
		5′ TGCCATAG CAT **ATG GCTAAAAAAACATCCTCA** 3′
		The reverse primer (*HindIII**)
		5′TGCCATAG AAGCTT **ACGAGGCAAGAAGGACATTG** 3′
		The forward primer (*NdeI**)

		M E K K T S S K G K
		5′TGCCATAG CAT **ATG**GAG AAGAAGACATCCTCGAAAGGGAAA **CGGCATCCAAACACTCAC** 3′
	pCWOri$^+$	The reverse primer (*Avr*II*)
		5′TGCCATAG CCTAGG **GCGCGGAAGAAAGCGAATTT** 3′
CP2B6		The reverse primer (*HindIII**)
		5′ TGCCATAG AAGCTT **GCGCGGAAGAAAGCGA** 3′
		The forward primer (*NdeI**)
	pET29b	M E K K T S S K G K
		5′TGCCATAG CAT **ATG**GAG AAGAAGACATCCTCGAAAGGGAAA **CGGCATCCAAACACTCAC** 3′
		The reverse primer (*HindIII**)
		5′ TGCCATAG AAGCTT **GCGCGGAAGAAAGCGA** 3′

		The forward primer (*NdeI**)
		M A L L L A V F
		5′ TGCCATAG CAT **ATG** GCTCTTCTTTTGGCCGTGTTC **GTAGAGACCTGGCTTTTGCTG** 3′
	pCWOri$^+$	The reverse primer (*Avr*II*)
		5′ TGCCATAG CCTAGG **TTCACCGCTGAGTGTACCGT** 3′
		The reverse primer (*Hind*III*)
CP3A5		5′ TGCCATAG AAGCTT **TTCACCGCTGAGTGTACCGT** 3′

		The forward primer (*NdeI**)
		M A L L L A V F
	pET29b	5′ TGCCATAG CAT **ATG** GCTCTTCTTTTGGCCGTGTTC **GTAGAGACCTGGCTTTTGCTG** 3′
		The reverse primer (*Hind*III*)
		5′ TGCCATAG AAGCTT **TTCACCGCTGAGTGTACCGT** 3′
		The forward primer (*Nde*I**)*
		M A L L L A V F
		5′ TGCCATAG CAT **ATG** GCTCTTCTTTTGGCCGTGTTC **CATGGTTTATTTAAGAAGCTC 3′**
	pCWOri+	The reverse primer (*Avr*II*)
		5′ TGCCATAG CCTAGG **AGCGCCAGATACTGTGCC** 3′
CP3A4		The reverse primer (*Hind*III*)
		5′ TGCCATAG AAGCTT **AGCGCCAGATACTGTGCC** 3′
		The forward primer (*Nde*I**)*
	pET29b	M A L L L A V F
		5′ TGCCATAG CAT **ATG** GCTCTTCTTTTGGCCGTGTTC **CATGGTTTATTTAAGAAGCTC 3′**
		The reverse primer (*Hind*III*)
		5′ TGCCATAG AAGCTT **AGCGCCAGATACTGTGCC** 3′

		The forward primer (AvrII)*
	pCWOri+	5′TGCCATAG CCTAGG **ATGGGGGATTCACACGTC** 3′
		The forward primer (HindIII)*
		5′TGCCATAG AAGCTT **ATGGGGGATTCACACGTC** 3′
CPR		The reverse primer (*Hind*III*)stop codon (B)
		5′TGCCATAG AAGCTT TCA **AGACCACACGTCTAATGAGT** 3′
		The reverse primer (*Hind*III*)6x His Tag + stop codon (F)
		5′TGCCATAG AAGCTT TCAGTGGTGGTGGTGGTGGTG **AGACCACACGTCTAATGAGT** 3′
		The forward primer (*HindIII*)*
	pET29b	5′TGCCATAG AAGCTT **ATGGGGGATTCACACGTC** 3′
		The reverse primer (*Xho*I*)no stop codon (His tag with pET29b)(D)
		5′TGCCATAG CTCGAG **AGACCACACGTCTAATGAGT** 3′
		The reverse primer (*Xho*I*)stop codon (E)
		5′TGCCATAG CTCGAG TCA **AGACCACACGTCTAATGAGT** 3′

1. ** The restriction enzyme site is underlined in the primer sequence, while the functional primer is highlighted in bold. The N-terminus modification of the amino acid sequence is shown above the nucleotide sequence.*

Table 1.2 P450 cloning in pET22b and CPR in pCDFDuet-1

Second cloning strategy, the human P450s were cloned into pET22b, whereas human reductase was cloned in pCDFDuet-1 plasmid and co-expressed in *E. coli* strains.

CYP Form	Cloning Plasmid	Sequence of nucleotide primers
CP2A6	pET22b	The forward primer (*Nco*I*) 5′ TGCCATAG **CC ATGGCTAAAAAAACATCCTCA** 3′ The reverse primer (*Xho*I*) 5′ TGCCATAG **CTCGAG ACGAGGCAAGAAGGACATTGT 3′**
CP2B6	pET22b	The forward primer (*Nco*I*) M E K K T S S K G K 5′TGCCATAG **CC ATGG**AG AAGAAGACATCCTCGAAAGGGAAA**CGGCATCCAAACACTCACGAT** 3′ The reverse primer (*Xho*I*) 5′ TGCCATAG **CTCGAG GCGCGGAAGAAAGCGAATTT** 3′
CP3A5	pET22b	The forward primer (*BamHI*) M A L L L A V F 5′TGCCATAG GGATTC G **ATG** GCTCTTCTTTTGGCCGTGTTC **GTAGAGACCTGGCTTTTGCTG3′** The reverse primer (*Xho*I*) 5′TGCCATAG **CTCGAG TTCACCGCTGAGTGTACCGT** 3′

CP3A4	pET22b	The forward primer (*NcoI**) M A L L L A V F 5′TGCCATAG **CC ATG**GCTCTTCTTTTGGCCGTGTTC **CATGGTTTATTTAAGAAGCTC 3′** The reverse primer (*XhoI**) 5′TGCCATAG **CTCGAG AGCGCCAGATACTGTGCCA** 3′
CP2D6	pET22b	The forward primer (*NcoI**) M A R Q V H S S W N L 5′TGCCATAG **CC ATG** GCTCGTCAGGTCCACTCGTCCTGGAATCTT **CCGCCAGGGCCGTTGCCT 3′** The reverse primer (*XhoI**) 5′TGCCATAG **CTCGAG CCGCGGCACTGCGCATAA** 3′
CPR	pCDFDuet-1	The forward primer (*BamHI**) 5′TGCCATAG GGATCC G **ATGGGGGATTCACACGTC 3′** The reverse primer (*Hind*III*) 5′TGCCATAG AAGCTT **AGACCACACGTCTAATGAGT** 3′

G highlighted in red was added to put the sequence in frame with the plasmid

Table 1.3 P450 cloning in pET29b and CPR in pCDFDuet-1

In the third cloning strategy, all human P450s were cloned de novo in pET29b and co-expressed with the cloned human reductase in pCDFDuet-1 plasmid in different *E. coli* strains.

CYP Form	Cloning Plasmid	Sequence of nucleotide primers
CP2A6	pET29b	The forward primer (*Nde*I*) 5′ TGCCATAG **CAT ATGGCTAAAAAAACATCCTCA** 3′ The reverse primer (*Xho*I*) 5′ TGCCATAG **CTCGAG ACGAGGCAAGAAGGACATTGT 3′**
CP2B6	pET29b	The forward primer (*Nde*I*) M E K K T S S K G K 5′TGCCATAG **CAT ATGG**AG AAGAAGACATCCTCGAAAGGGAAA **CGGCATCCAAACACTCACGAT**3′ The reverse primer (*Xho*I*) 5′ TGCCATAG **CTCGAG GCGCGGAAGAAAGCGAATTT** 3′
CP3A5	pET29b	The forward primer (Nde*I*) M A L L L A V F 5′TGCCATAG CAT**ATG** GCTCTTCTTTTGGCCGTGTTC **GTAGAGACCTGGCTTTTGCTG3′** The reverse primer (*Xho*I*) 5′TGCCATAG **CTCGAG TTCACCGCTGAGTGTACCGT** 3′

CP3A4	pET29b	The forward primer (*NdeI**) M A L L L A V F 5′TGCCATAG **CAT ATG** GCTCTTCTTTTGGCCGTGTTC **CATGGTTTATTTAAGAAGCTC 3′** The reverse primer (*XhoI**) 5′TGCCATAG **CTCGAG AGCGCCAGATACTGTGCCA** 3′
CP2D6	pET29b	The forward primer (*NdeI**) M A R Q V H S S W N L 5′TGCCATAG **CAT ATG** GCTCGTCAGGTCCACTCGTCCTGGAATCTT **CCGCCAGGGCCGTTGCCT3′** The reverse primer (*XhoI**) 5′TGCCATAG **CTCGAG CCGCGGCACTGCGCATAA** 3′

Appendix 2- SDS-PAGE gels

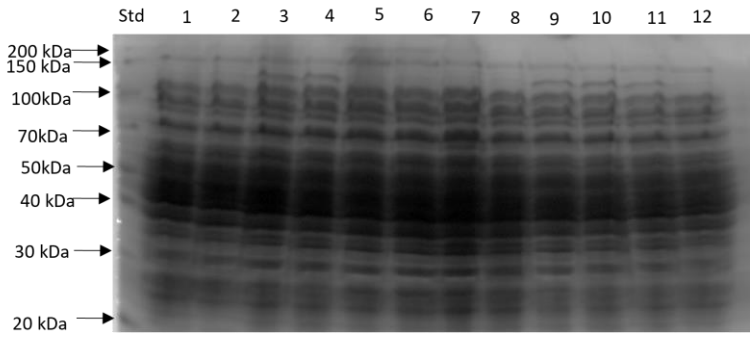


Figure 2. 1 SDS gel showing total protein of fused genes expressed in C41(DE3) without pGro7, with 0.5 mM 5-ALA in LB 20 °C. lane 1; 2A6-CPR-pCWOri^{+} B, lane 2; 2A6-CPR-pCWOri^{+} F, lane 3; 2B6-CPR-pET29b D, lane 4; 2B6-CPR-pET29b E, lane 5; 3A5-CPR-pCWOri^{+} B, lane 6; 3A5-CPR- pCWOri^{+} F, lane 7; 3A4-CPR- pCWOri^{+} B, lane 8; 3A4-CPR- pCWOri^{+} F ,lane 9; 3A5-CPR-pET29b D, lane 10; 3A5-CPR-pET29b E, lane 11; 2A6-CPR-pET29b E, lane 12; 2B6-CPR- pCWOri^{+} B.

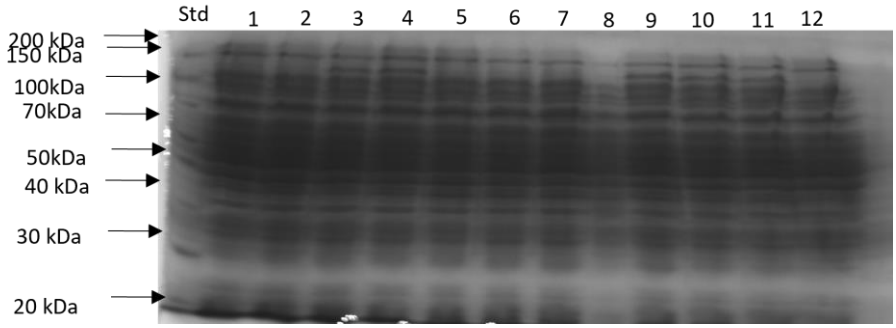


Figure 2. 2 SDS gel showing total protein of fused genes expressed in C41(DE3) without pGro7, with 0.5 mM 5-ALA in LB 30 °C. lane 1; 2A6-CPR-pCWOri^{+} B, lane 2; 2A6-CPR-pCWOri^{+} F, lane 3; 2B6-CPR-pET29b D, lane 4; 2B6-CPR-pET29b E, lane 5; 3A5-CPR-pCWOri^{+} B, lane 6; 3A5-CPR- pCWOri^{+} F, lane 7; 3A4-CPR- pCWOri^{+} B, lane 8; 3A4-CPR- pCWOri^{+} F ,lane 9; 3A5-CPR-pET29b D, lane 10; 3A5-CPR-pET29b E, lane 11; 2A6-CPR-pET29b E, lane 12; 2B6-CPR- pCWOri^{+} B.

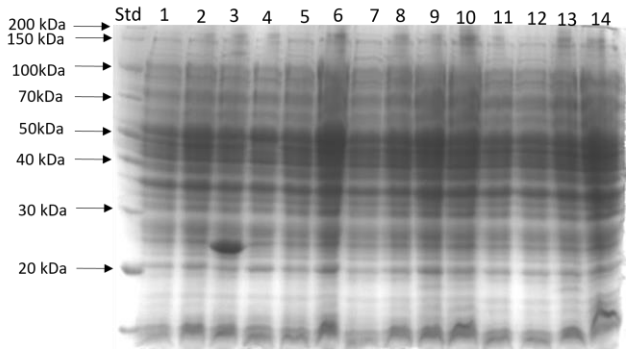


Figure 2. 3 SDS gel showing total protein of fused genes expressed in C41(DE3) with pGro7, with 0.5 mM 5-ALA in LB 20 °C. lane 1; 2A6-CPR-pCWOri^{+} B, lane 2; 2A6-CPR-pCWOri^{+} F, lane 3; 2B6-CPR-pET29b D, lane 4; 2B6-CPR-pET29b E, lane 5; 3A5-CPR-pCWOri^{+} B, lane 6; 3A5-CPR- pCWOri^{+} F, lane 7; 3A4-CPR- pCWOri^{+} B, lane 8; 3A4-CPR- pCWOri^{+} F ,lane 9; 3A5-CPR-pET29b D, lane 10; 3A5-CPR-pET29b E, lane 11; 2A6-CPR-pET29b E, lane 12; 2B6-CPR- pCWOri^{+} B, lane 13; 3A4-CPR-pET29b D, lane 14; 3A4-CPR-pET29b E.

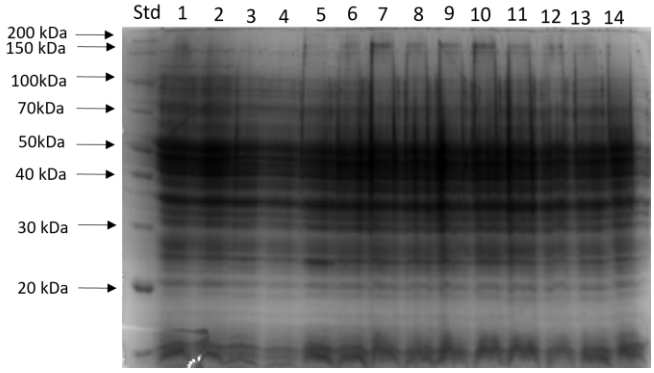


Figure 2. 4 SDS gel showing total protein of fused genes expressed in C41(DE3) with pGro7, with 0.5 mM 5-ALA in LB 20 °C. lane 1; 2A6-CPR-pCWOri^{+} B, lane 2; 2A6-CPR-pCWOri^{+} F, lane 3; 2B6-CPR-pET29b D, lane 4; 2B6-CPR-pET29b E, lane 5; 3A5-CPR-pCWOri^{+} B, lane 6; 3A5-CPR- pCWOri^{+} F, lane 7; 3A4-CPR- pCWOri^{+} B, lane 8; 3A4-CPR- pCWOri^{+} F ,lane 9; 3A5-CPR-pET29b D, lane 10; 3A5-CPR-pET29b E, lane 11; 2A6-CPR-pET29b E, lane 12; 2B6-CPR- pCWOri^{+} B, lane 13; 3A4-CPR-pET29b D, lane 14; 3A4-CPR-pET29b E.

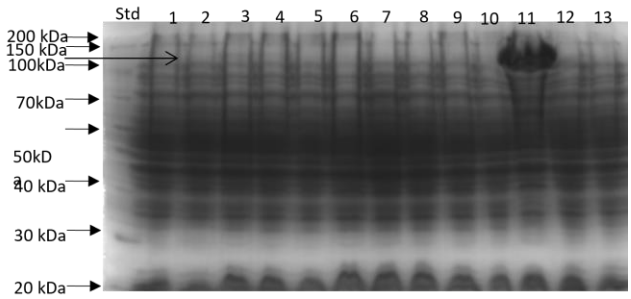


Figure 2. 5 SDS gel showing total protein of fused genes genes expressed in C41(DE3) without pGro7, with 0.5 mM 5-ALA (no ferrous and ferric chloride) in LB 20 °C. lane 1; 2A6-CPR-pCWOri^{+} B, lane 2; 2A6-CPR- pCWOri^{+} F, lane 3; 2B6-CPR-pET29b D, lane 4; 2B6-CPR-pET29b E, lane 5; 3A5-CPR- pCWOri^{+} B, lane 6; 3A5-CPR- pCWOri^{+} F, lane 7; 3A4-CPR- pCWOri^{+} B, lane 8; 3A4-CPR- pCWOri^{+} F ,lane 9; 3A4-CPR-pET29b D, lane 10; 3A4-CPR-pET29b E, lane 12; 2A6-CPR-pET29b E1, lane 13; 2B6-CPR- pCWOri^{+} B.

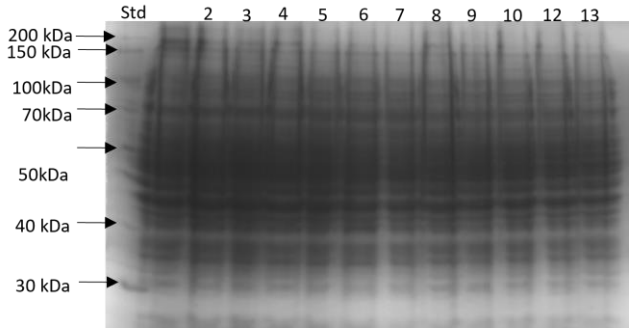


Figure 2. 6 SDS gel showing total protein of fused genes expressed in C41(DE3) without pGro7, with 0.5 mM 5-ALA (no ferrous and ferric chloride) in LB 30 °C. lane 1; 2A6-CPR- pCWOri^{+} B, lane 2; 2A6-CPR- pCWOri^{+} F, lane 3; 2B6-CPR-pET29b D, lane 4; 2B6-CPR-pET29b E, lane 5; 3A5-CPR- pCWOri^{+} B, lane 6; 3A5-CPR- pCWOri^{+} F, lane 7; 3A4-CPR- pCWOri^{+} B, lane 8; 3A4-CPR- pCWOri^{+} F ,lane 9; 3A4-CPR-pET29b D, lane 10; 3A4-CPR-pET29b E, lane 12; 2A6-CPR-pET29b E, lane 13; 2B6-CPR- pCWOri^{+} B.

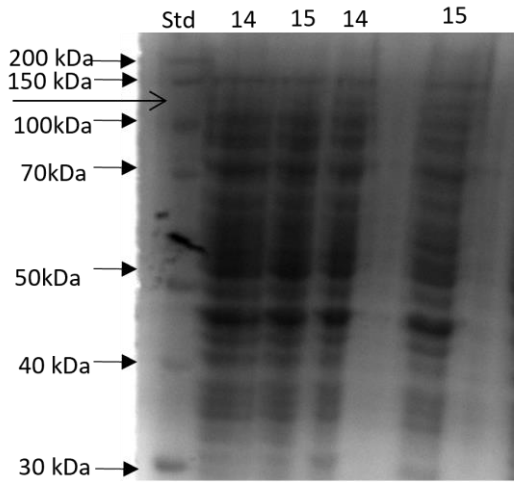


Figure 2. 7 SDS gel showing total protein of fused genes expressed in C41(DE3) without pGro7, with 0.5 mM 5-ALA (no ferrous and ferric chloride) (continuing from two previous gels). Lanes 14 and 15 on the left are in LB 20 °C, whereas lanes 14 and 15 on the right are in LB 30 °C. Lane 14; 3A5-CPR-pET29b D, lane 15; 3A5-CPR-pET29b E.

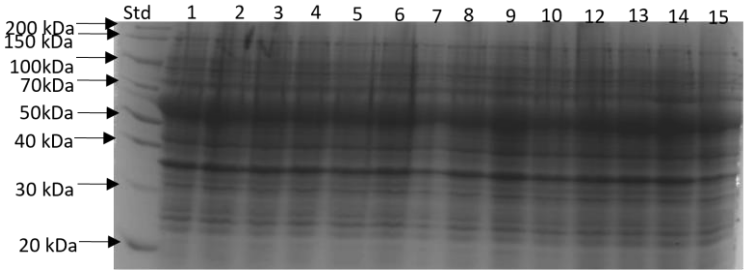


Figure 2. 8 SDS gel showing total protein of fused genes expressed in C41(DE3) with pGro7, with 0.5 mM 5-ALA (no ferrous and ferric chloride) in LB 20 °C. lane 1; 2A6-CPR- pCWOri$^+$ B, lane 2; 2A6-CPR- pCWOri$^+$ F, lane 3; 2B6-CPR-pET29b D, lane 4; 2B6-CPR-pET29b E, lane 5; 3A5-CPR- pCWOri$^+$ B, lane 6; 3A5-CPR- pCWOri$^+$ F, lane 7; 3A4-CPR- pCWOri$^+$ B, lane 8; 3A4-CPR- pCWOri$^+$ F ,lane 9; 3A4-CPR-pET29b D, lane 10; 3A4-CPR-pET29b E, lane 12; 2A6-CPR-pET29b E1, lane 13; 2B6-CPR-pCWOri$^+$ B, lane 14; 3A5-CPR-pET29b D, lane 15; 3A5-CPR-pET29b E.

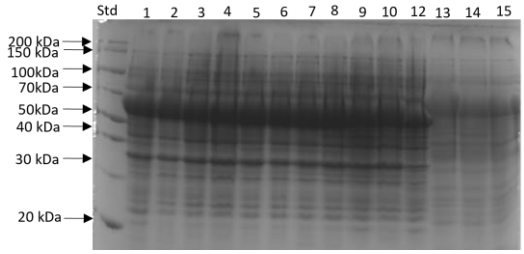


Figure 2. 9 SDS gel showing total protein of fused genes expressed in C41(DE3) with pGro7, with 0.5 mM 5-ALA (no ferrous and ferric chloride) in LB 30 °C. lane 1; 2A6-CPR- pCWOri$^+$ B, lane 2; 2A6-CPR- pCWOri$^+$ F, lane 3; 2B6-CPR-pET29b D, lane 4; 2B6-CPR-pET29b E, lane 5; 3A5-CPR- pCWOri$^+$ B, lane 6; 3A5-CPR- pCWOri$^+$ F, lane 7; 3A4-CPR- pCWOri$^+$ B, lane 8; 3A4-CPR- pCWOri$^+$ F ,lane 9; 3A4-CPR-pET29b D, lane 10; 3A4-CPR-pET29b E, lane 12; 2A6-CPR-pET29b E1, lane 13; 2B6-CPR-pCWOri$^+$ B, lane 14; 3A5-CPR-pET29b D, lane 15; 3A5-CPR-pET29b E.

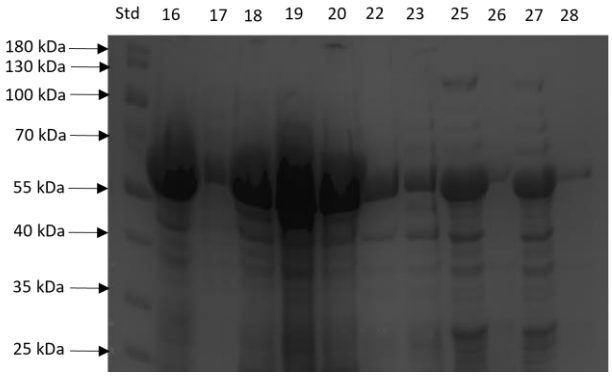


Figure 2. 10 SDS gel showing all CFE of metagenome genes 198-pET29b (51.6 kDa) and 914-pGEX-6P-1 (116.8 kDa) co-expressed with pGRo7 in Tuner cells with 0.25 mM and 0.5 mM IPTG. From left; protein standard (PageRuler, ThermoFisher Scientific), lane 22; 198-pET29b-pGRo7 in LB 30 °C with 0.5 mM IPTG, lane 23; 198-pET29b-pGRo7 in LB 20°C with 0.25 mM IPTG, lane 25; 914-pGEX- pGRo7 in LB 20°C with 0.25 mM IPTG, lane 26; 914-pGEX- pGRo7 in LB 30°C with 0.25 mM IPTG, lane 27; 914-pGEX- pGRo7 in LB 20 °C with 0.5 mM IPTG, lane 28; 914-pGEX- pGRo7 in LB 30 °C with 0.5 mM IPTG.

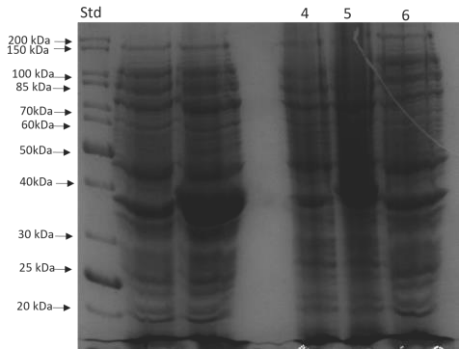


Figure 2.11 TCP of Nitrilases in LB 30 ℃. From left protein standard (PageRuler, ThermoFisher Scientific), lane 4; Q89PT3=36.6 kDa, lane 5; A5EKU8=39.2 kDa, lane 6; A0R378=39.1 kDa,

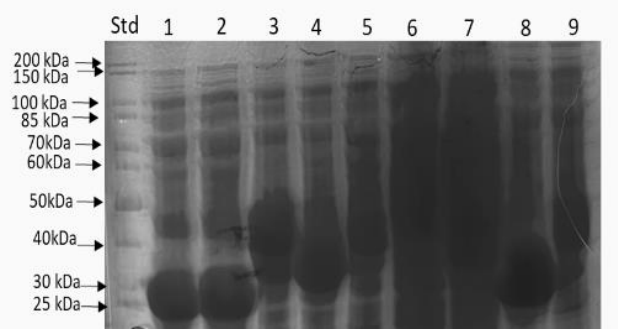


Figure 2.12 CFE of Nitrilases in LB 20 ℃. From left protein standard (PageRuler, ThermoFisher Scientific), lane 1; Q2J474=34 kDa, lane 2; A5ETE9=34.2 kDa, lane 3; B4AL96=40.1 kDa, lane 4; Q89PT3=36.6 kDa, lane 5; A5EKU8=39.2 kDa, lane 6;

A0R378=39.1 kDa, lane 7; Q89GE3=34.4 kDa, lane 8; Q819F0=31.7 kDa, lane 9; Q5LLB2=37.2 kDa.

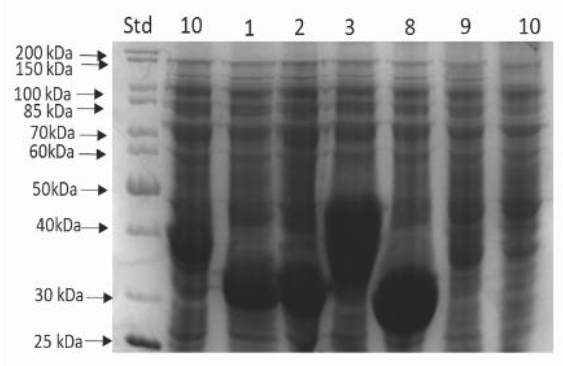


Figure 2.13 CFE of Nitrilases in LB 30 ℃. From left protein standard (PageRuler, ThermoFisher Scientific), lane 10 in LB 20 ℃; Q6N284=35.9kDa; lane 1; Q2J474=34 kDa, lane 2; A5ETE9=34.2 kDa, lane 3; B4AL96=40.1 kDa; lane 8; Q819F0=31.7 kDa, lane 9; Q5LLB2=37.2 kDa, lane10; Q6N284=35.9kDa

Appendix 3- Enzyme assays

CO assay of a positive control

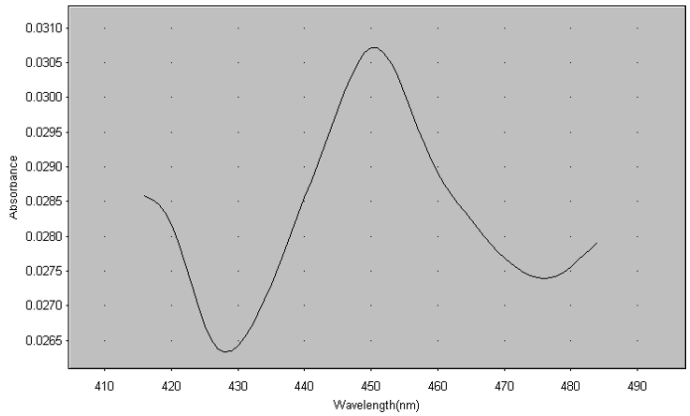


Figure 3. 1 CO reduced spectrum (400-500nm) of recombinant hexa-histidine tagged protein WP-011442524.1 from Rhodopseudomonas palustris HaA2 expressed in E. coli BL21(DE3). The peak at 450 nm indicates its activity, and was used as a positive control, to determine the suitability of the method.

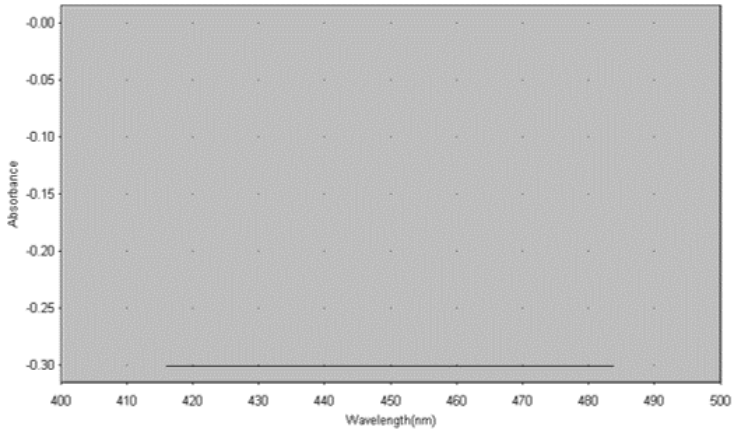


Figure 3. 2 Negative control of CO assay with an of empty plasmid in *E. coli* cells.

BSA Bradfords standard curve

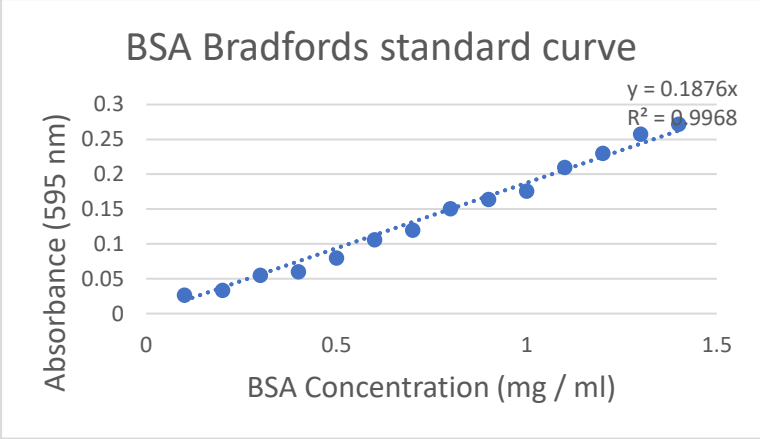


Figure 3.3 Graph plotted based on standard concentrations of Bradford reagent and the absorbances measured at 595 nm.

NH_4Cl standard curve

Table 3.4 NH4Cl standards concentration curve components.

NH4Cl concentration (mM) (Baeshen et al.)	Volume (µL) NH4Cl to add	Volume (µL) of H2O
2	20	980
3	30	970
4	40	960
5	50	950
6	60	940
7	70	930
8	80	920
9	90	910
10	100	900
11	110	890
12	120	880

Table 3.5 The concentrations of NH_4Cl used and the absorbances recorded.

Concentration (mM)	Absorbance
2	0.126939
3	0.292562
4	0.275686
5	0.281422
6	0.636119
7	0.794259
8	0.727936
9	0.929789
10	0.923029
11	0.881812
12	1.230982

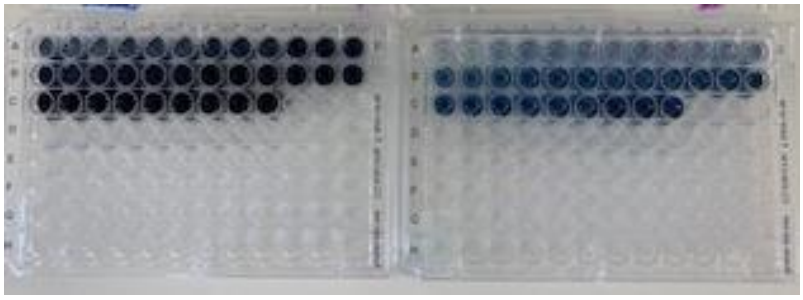


Figure 3.6 Microtiter plate of NH_4Cl standard solution. On the left side plate is the neat standard solutions of NH4Cl, while on the right plate is the diluted in DMSO to allow concentration determination by UV/visible spectroscopy.

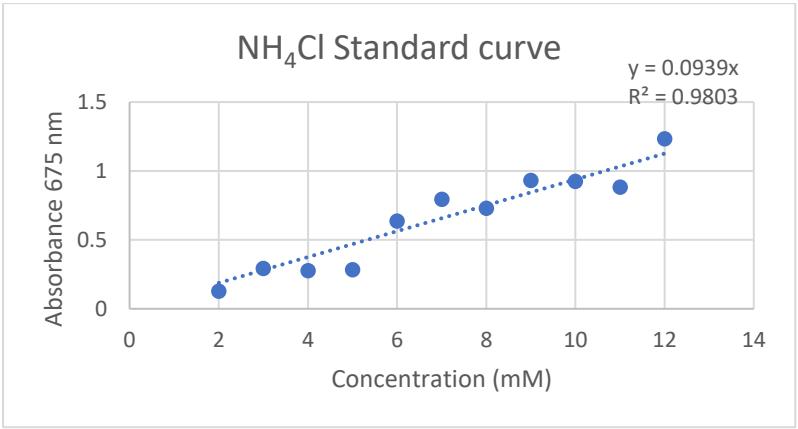


Figure 3.7 Graph plotted based on standard concentrations of NH_4Cl and the absorbances measured at 675 nm.

Figure 3.8 Microtiter plate with positive control. On the left side plate is the neat standard solutions of NH4Cl, while on the right plate is the diluted in DMSO to allow concentration determination by UV/visible spectroscopy.

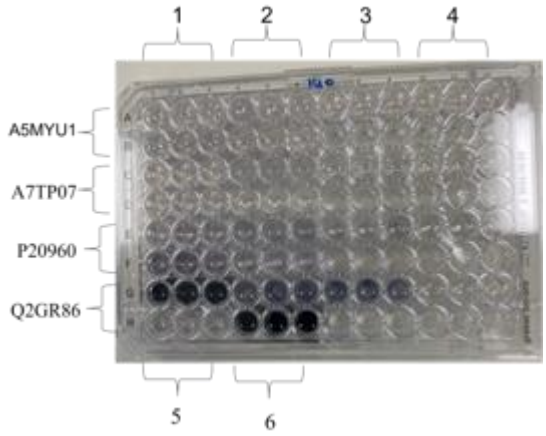


Figure 3.9 Microtiter plate assay of lyophilised (CFE) Q2GR86 enzyme hydrolysing benzyl (2-cyanoethyl) carbamate) (Nitriles 1), and 2-(4-cyanophenyl)-N-ethylacetamide (Nitrile 6) into their carboxylic acid, and the release of ammonia. A chromophore is generated after the addition of TCA in solution. The numbers on top and bottom of plate indicate the number of the nitriles (Table 2-39), while on the left of the plate are the names of the nitrilases.

Additionally, 10 nitrilases were provided as lyophilised CFE forms from Prozomix Ltd, UK, labelled by codes PRO-NITR 010, and PRO-NITR 027 to PRO-NITR 035.

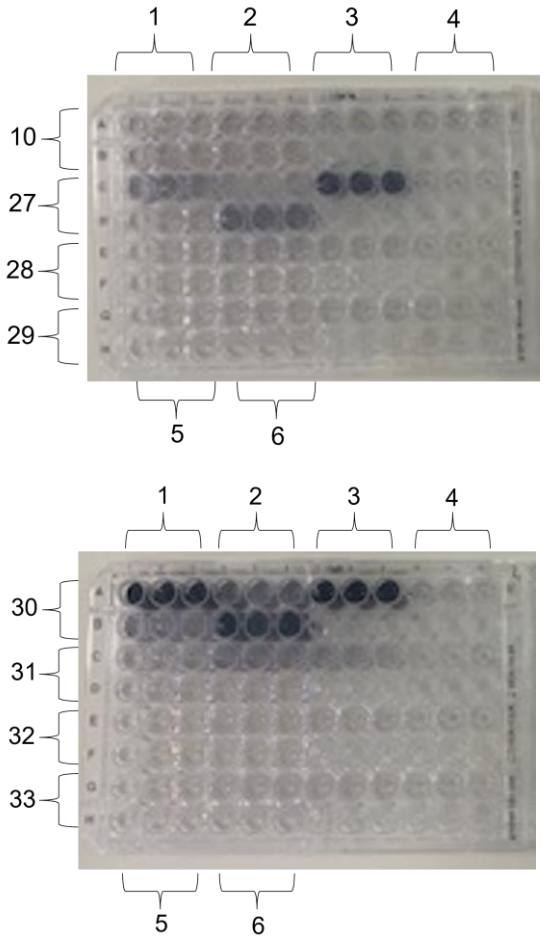


Figure 3.10 Microtiter plate assay of lyophilised (CFE) nitrilases 10, 27, 28 and 29 and 30-34 provided by Prozomix Ltd. . As can be seen, nitrilase 27 showed some catalytic activity against Benzyl(2-cyanoethyl) carbamate (nr 1), N-(3-cyanopropyl) pthalimide (nr 3) and against 2-(4-cyanophenyl)-N-ethylacetamide (nr 6), and nitrilase 30 showed good catalytic activity against Benzyl(2-cyanoethyl) carbamate; (nr 1), N-(3-cyanopropyl) pthalimide; (nr 3), and against 2-(4-cyanophenyl)-N-ethylacetamide; (nr 6).

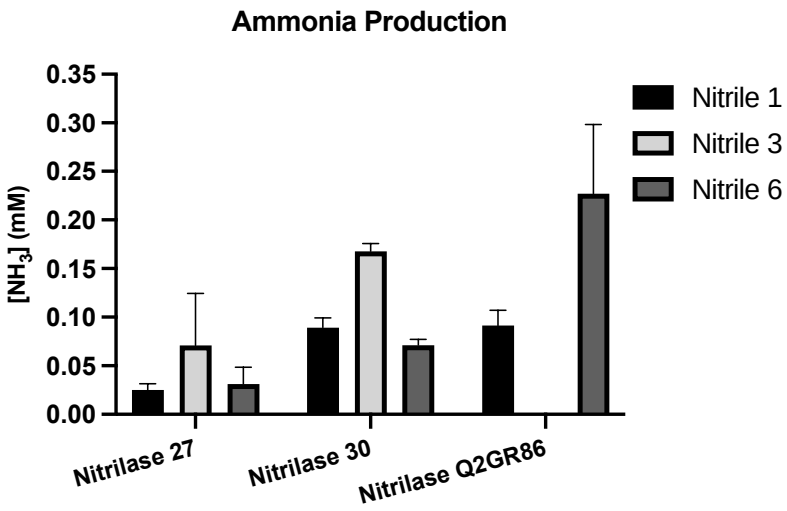


Figure 3.11 graph showing the catalytic affinity of nitrilases 27, 30 and Q2GR86 against nitriles 1, 3 and 6.