The University Of Sheffield.

Investigating the industrial application potentials of cytochrome P450 BM-3 and four novel putative cytochrome P450s isolated from *Cupriavidus necator* H16

By:
Inas Al-nuaemi

A thesis submitted in partial fulfilment of the requirements for the degree of
Doctor of Philosophy

The University of Sheffield
Faculty of Engineering
Department of Chemical and Biological Engineering

Jun 2018

Dedicated to the memory of my husband, Mohammed, who always believed in my ability to succeed in the academic field. You are gone but your belief in me has made this journey possible

Acknowledgements

I would like to thank The University of Sheffield, Ministry of Higher Education in Iraq, Iraqi Cultural Attache in London, The Middle Technical University in Iraq and the Institute of Technology in Baghdad, Iraq for offering me this valuable experience and for their financial support.

I would like to express my immense gratitude to my supervisor Dr Tuck Wong for his unconditional patience and guidance throughout the entire time of my PhD. It would have been a very different story without his inspiring and motivating discussions filled with knowledge and expertise.

I would like to show my appreciation to Dr Kang Lan Tee whose constant support, positive attitude and invaluable advice has made the project to proceed and succeed!

Everyone with Tuck's group Abdulrahman, Miriam, Jose, Yomi, Pawel, Zaki and Hossam I would like to thank you for all your amazing help and support, I have been very fortunate to share lab space with you all.

Lastly, I would like to thank my family to whom I owe a great deal. I would like to thank my father Jabbar and my mother Rajaa and my sisters Wasan. Usra, Zahraa and Huda and my sister in law Masara who believed I could do this even in the hardest of times and stood by me with interest, enthusiasm and a completely unrealistic sense of pride in me. Also many thanks to my friend Marwa and her husband Mohammed for your outstanding, continues support.

And finally, the one person who has made this all possible has been my son Abdullah. He has been a constant source of support and encouragement for me to continue my PhD. Hence, great appreciation and enormous thanks are due to him, for without his understanding, I am sure this thesis would never have been completed.

Declaration

The author declares that no portion of the work referred to in this thesis has been submitted in support of an application for another degree or qualification of this or any other university or other institute of learning.

Contents page

I. List of abbreviations

AC	Affinity chromatography
AIM	Auto-induction media
ALA	Aminolevulinic acid
A_{230}, A_{260}, A_{280}	UV absorbance at a specified wavelength
BMP	Heme domain of P450 BM3
CV	Colum volume
DMSO	Dimethyl sulphoxide
DNA	Deoxyribonucleic acid
E. coli	*Escherichia coli*
EDTA	Ethylenediaminetetraacetic acid
FAD	Flavin adenine dinucleotide
FDH	Formate dehydrogenase
2Fe-2S	2 Iron-2 Sulphur cluster
FMN	Flavin mononucleotide
GSK	Glaxo Smith Kline
GST	Glutathione-S-transferase
His-tag	Histidine-tag

IEX	Ion exchange chromatography
IPTG	Isopropyl-β,D-Thiogalactopyranoside
Kan	Kanamycin
kDa	Kilodalton
LB	Lysogeny broth
mRNA	Messenger ribonucleic acid
NADP	Nicotinamide-adenine-dinucleotide phosphate
NADH	β- Nicotinamide adenine dinucleotide (reduced)
NADPH	β- Nicotinamide adenine dinucleotide-phosphate (reduced)
OD_{600}	Optical density at 600 nm
P450 BM3	Cytochrome P450 BM3 from *Bacillus megaterium*
PCR	Polymerase chain reaction
pI	Isoelectric point, the pH at which a protein has zero net surface charge
pNCA	Para-nitrophenoxycarboxylic acid
RNA	Ribonucleic acid
RT	Room temperature
rpm	Rotations per minute
SDS-PAGE	Sodium dodecyl sulphate-polyacrylamide gel electrophoresis

SEL	Trace elements solution
SEM	Standard Error Mean
SeSaM	Sequence saturation mutagenesis
SC-CO_2	Supercritical carbon dioxide
SCFs	Supercritical fluids
TEMED	Tetramethylethylendiamine
TEV	Tobacco-etch virus
THF	Tetrahydrofuran
TB	Terrific broth
TS	Super broth
UV	Ultra violet
v/v	volume per volume
WT	Wild type
α	alpha
β	beta

for amino acids abbreviation, refer to Appendix III

II. List of figures

III. List of tables

Abstract

Cytochrome P450 is a superfamily of heme-dependent monooxygenases. These enzymes play a very important role in drug metabolism and detoxification since they catalyse a broad range of reactions including hydroxylation, dehalogenation, deamination, epoxidation, peroxidation, desulphurisation and dealkylation involving various substrates, such as alkanes, phenols, steroids and fatty acids. This thesis presents results from two P450 hemoprotein investigations: I) targeting study of the most common cytochrome P450 BM-3 from *Bacillus megaterium*, and II) discovery study of four novel putative cytochrome P450s from *Cupriavidus necator* H16. In Part II of this project, P450 BM-3 WT and W5F5 variants were expressed and purified, and their activities in three green solvents, 1-butanol, 2-butanol and dimethyl carbonate, were examined to check the tolerance of the wild type and the mutant W5F5 variants towards green solvents used for the first time with this enzyme. The results revealed a new green solvent with lower denaturation effect on P450 BM-3, the wild type and the mutant, with potential application in industrial biosynthesis. Part III focuses on the characterisation of four novel cytochrome P450 proteins of unknown structure and function. The proteins were analysed structurally and the first predicted models were discussed. Optimisations of the expression and purification of these proteins as well as optical properties were analysed. Scanning the changes in light absorbance of the substrate-protein complex with eleven substrates resulting in Type I changes when lauric acid and 3-phenoxytoluene were added to the purified P450-H16-B1743, as well as the evaluation of the catalytic activity of P450-H16-B1279 showed possible catalytic activity towards indol, malonic acid, adipic acid, 3-phenoxytoluene and S-Ethyl-N,N-dipropylthiocarbama (EPTC).

1 Introduction

1.1 Objectives

In recent years, biocatalysts have been the subject of intensive investigations by pharmaceutical and industrial researchers, and that interest is continuously growing. This is not surprising given the potential advantages of enzymes: (I) the ability to display regio- chemo- and enantioselectivity; (II) the enhancement in reactions rate up to 10^{12} folds. (III) the ability to catalyse reactions in moderate conditions (pH and temperature), while these reactions require harsh conditions by using chemical catalysts (Carrea & Riva, 2000).

Cytochrome P450 (CYP) is a superfamily of multifunction monooxygenases. More than 6,000 enzymes have been classified as members of this family. They are found in nearly every organism including both prokaryotes and eukaryotes and play a crucial role in the metabolism of endogenous fatty acids and steroids as well as foreign compounds such as chemical carcinogens and drugs. Indeed, no other enzyme group can accept as many substrates (alkanes, phenols, steroids and fatty acids) or catalyse as many reaction types (hydroxylation, dehalogenation, deamination, epoxidation, peroxidation, desulphurisation and dealkylation (Bernhardt, 2006; Urlacher & Eiben, 2006; Urlacher et al., 2004). One of the main reasons that P450s attract so much attention is their ability to introduce one oxygen atom into the substrate, while the second one is reduced to water, a reaction which is difficult to catalyse by chemical catalysts (Whitehouse et al., 2012).

This thesis is dedicated to the study of cytochrome P450s from two organisms: *Bacillus megaterium* (Part II) and *Cupriavidus necator* H16 (Part III).

1.1.1 Part II: Cytochrome P450 BM-3

Despite the potential of using P450s, commercial application of the catalysed reactions of these enzymes has been hampered by the solubility of nonpolar substrates in aqueous solution. Therefore, organic co-solvents need to be added as supplements to increase substrate solubility (Wong et al., 2004).

Traditional co-solvents, further to their negative effect on enzyme activity, could cause serious problems such as environmental pollution, human health and safety issues, as well as the risk of explosion. In addition, using hazardous solvents increases the final products' cost as they require complicated separation processes, more accurate transformation and storage process as well as additional units for waste treatment. In contrast, using green solvents instead of hazardous solvents can result in potential savings of time, effort and costs by decreasing waste treatment costs, minimising safety requirements, reducing energy consumption and increasing the solvent recovery rate (Sheldon et al., 2007). For example, a comparison of ethers shows that t-amyl ethyl ether would be a highly desirable solvent for use instead of t-butyl methyl ether, which has a very low flash point and a low boiling point of 55°C, both of which lead to handling difficulties, especially on the commercial scale (Henderson et al., 2011). In addition, the removal of residual solvents from the reaction medium should be performed by either evaporation or distillation, both of which are associated with major environmental issues of global proportion and high insurance costs because of the high volatility of traditional solvents. Some of these solvents, which were classified as red solvents by GSK, such as chlorinated hydrocarbon solvents have already been banned or are likely to be in the near future. These issues of the non-green solvents have stimulated the chemical and pharmaceutical industries to seek more benign alternatives. The

intent of Part II of this project is to provide the pharmaceutical industry with the best green solvents which can be used as a replacement to the traditional organic solvents in reactions catalysed by cytochrome P450 BM-3. The behaviour of this protein will be investigated in the presence of different green solvents which are used for the first with the wild type and the W5F5 variant of P450 BM-3.

1.1.2 Part III: Putative cytochrome P450s from *Cupriavidus necator* H16

Although more than 6000 P450 cytochromes have been identified in animals, plants and microbes (Whitehouse et al., 2012), there are no general rules that can be applied to all P450s regarding their structure, substrates and reactions that can be catalysed by these superior enzymes. Due to this structural and functional diversity, identification of new P450 proteins can provide essential knowledge for industrially important biosynthetic processes. Each newly discovered cytochrome P450 represents a start point for unlimited expectations in biosynthesis applications.

Although the full sequencing of the *Cupriavidus necator* H16 genome published by Pohlmann and co-workers showed the presence of four cytochromes P450 genes, P450 H16-B2406, P450 H16-B1743, P450 H16-B1279 and P450 H16-B1009 (Pohlmann et al., 2006), the lack of structural data, spectroscopic information, relative substrates and catalytic activities of these novel P450s has motivated the investigation of the sequences of these four proteins and to align them with the other closest known P450s. In addition, it is of great interest to analyse the model and structure of the four P450s and try to express, purify, find proper substrates and discover what activities these proteins may exhibit. In this part of thesis, the main focus was to identify and characterise four novel cytochromes P450 from

Cupriavidus necator H16, the most excellent platform organism for the production of bioplastic and value-added compounds.

1.2 Organisation of this thesis

This thesis is organised in such a way that serves the diversity of its content, with a focus on the interconnections between its components to ensure that the reader understands the practical aspects of the thesis, linking it with the theoretical basis and moves from one topic to another easily. My thesis consists of three main parts, Part I, Part II and Part III, in addition to a general introduction (chapter 1) and comprehensive conclusions (chapter 12).

Part I consists of two chapters (chapters 2 and 3), providing general information about heme proteins as well as the cytochrome P450 superfamily, which is essential to understand the subsequent Part II and Part III. In chapter 2, I tried to familiarise the reader with a background about hemoproteins by including information about heme structure and heme function, as well as explaining the differences in structures and reactions between seven types of hemoproteins. Chapter 3 provides detailed information about the discovery, nomenclature, evolution, structure, optical properties and catalytic cycle of cytochromes P450 superfamily in general, as well as explaining the usual and unusual catalytic reactions and the industrial applications of different P450s.

The second part of the thesis (Part II) comprises four chapters (chapters 4-7) and is devoted to the P450 BM-3 and its tolerance towards non-conventional solvents and green solvents in particular. Chapter 4 provides detailed information about cytochrome P450 BM-3 (CYP102A1) from *Bacillus megaterium,* such as protein, domains, expression, structure and

dimerisation. In addition, it provides information about the mechanism of electron transformation, protein engineering and immobilisation approaches that have been reported and industrial applications of this enzyme. Chapter 5 highlights solvent classification and the biocatalysis in different known solvent systems, as well as the methods of stabilising enzymes in these solvents and the tolerance of P450 BM-3 against these solvents. Chapters 6 and 7 are the experimental chapters. All materials and protocols used for protein expression, purification, optical characterisation and activity evaluation in the presence of three green solvents are presented in chapter 6, while the results from these experiments are discussed in chapter 7.

The study of four novel cytochrome P450s from *Cupriavidus necator* H16 (*Ralstonia eutropha* H16) is presented in Part III. This part consists of one theoretical chapter (chapter 8) and three experimental chapters (chapters 9-11). Chapter 8 gives a brief overview of the history, genome, metabolism and industrial applications of *Cupriavidus necator* H16, as well as summarising the known P450s from this bacterium. Chapter 9 details the methodology that has been followed in the identification, sequence alignment and modelling of the four P450s, as well as the cloning, expression, purification, optical characterisation, substrate selectivity and activity investigation of these proteins. Chapter 10 presents the novel results from the sequence analysation and 3D structure modelling of P450s from *Cupriavidus necator* H16, while chapter 11 focuses on the results from expression, purification, optical characterisation, substrate binding and activity measurement of these proteins.

Part I: Heme proteins and cytochrome P450 superfamily

2 Heme proteins

2.1 Heme structure & function

Heme is an important molecule which apparently plays a pivotal role in biocatalysis and exhibits diverse biological activities, such as storing and transferring dioxygen and electron transformation (Ting et al., 2011). Although marking the active role of heme in biology as a superior biocatalyst for the organic substrate hydroxylation dates to the mid-1960s, the discovery of the enzymatic role of heme dates back earlier, to when the action of horseradish peroxidase (HRP) was indicated in 1903 (Bach & Chodat, 1903). HRP was the subject of study due to its stability and ease of purification until Mason clarified the ability to insert a single atom of oxygen from molecular dioxygen into an organic substrate enzymatically (Mason, 1957). It was not too long before P450 was distinguished as one of the most attractive oxidases in biotechnology (Garfinkel, 1958; Klingenberg, 1958; Omura & Sato, 1964).

Both the reductive and the oxidative reactions can be catalysed by heme enzymes, but this project will focus on the oxidant enzymes. Depending on the oxygen source that heme enzymes use to catalyse oxidation reactions, they are classified into two groups, oxygenases and peroxidases. The former normally use O_2, while the later use H_2O_2 to oxidise organic substrates. There are issues associated with each oxidant, by using O_2, a large kinetic barrier usually forms since O_2 is not a reactive molecule, however, H_2O_2 can lead to the creation of toxic hydroxyl radicals which negatively affect enzyme activity. In general, before heme oxygenases can bind to O_2, the iron should be reduced to ferrous (Fe^{2+}) to enable the

formation of the ferric-superoxide (Fe^{3+}-OO^{-}). After the arrival of the second electron the superoxide will be reduced to peroxide and from this point and on, both the oxidase and the peroxidase mechanism will be similar. Directly after the peroxide level, the O-O bond cleavage will take place either homolytically leaving behind two hydroxyl radicals or heterolytically to yield a water molecule and O atom with six valence electrons. In nature, the heterolytical pathway dominates the O-O bond cleavage because it is preferred to avoid the formation of hydroxyl radicals in the protein active site in most cases. The O-O cleavage leads to the most important step in the heme catalytic cycle, the formation of the active oxidant (compound I) by the oxidising equivalents rearrangement (Poulos, 2014).

2.2 Structural analysis of heme proteins

Heme as a prosthetic group can be found in many forms, but the most common forms are heme *b* which links to the host protein non-covalently and heme *c* which binds to two cysteine residues of the protein covalently by its vinyl groups (Figure 2-1). Heme proteins functional proficiency depends on several interrelated factors: I) heme protein diversity, II) protein's microenvironment variability, III) axial ligands type and V) heme accessibility to solvents. Due to the heme complexity, a good structure analysis is required for a thorough understanding of the heme binding surroundings (Bowman & Bren, 2008; Schneider et al., 2007).

A recent study on dataset of 125 heme binding proteins (heme *b* and heme *c*) identified five amino acids as axial ligands to the iron (the residue is counted as axial ligand when the distance between the heme iron and the residue's nitrogen, sulphur or oxygen is 3 Å or less).

These axial ligands are histidine (H), cysteine(C), methionine (M), lysine (K), and tyrosine(Y), with histidine domination for both heme *b* and heme *c* proteins. After histidine, cysteine represented the more utilising axial ligands in heme *b* against methionine in heme *c*. The relative frequency of heme proteins to all investigated proteins was also calculated and it was found that heme proteins have fewer cysteine (C), aspartic acid (D), isoleucine (I), lysine (K), asparagine (N), and serine residues (S) and higher number of alanine, phenylalanine (F), histidine, methionine, and tryptophan residues, while the highest relative frequencies were shown by C, F, M and Y (the axial ligands were excluded from the calculations) (Ting et al., 2011). Cysteine showed a dramatically high frequency with eight fold improvement compared to the background values in heme *c* proteins, which is expected owing to the classic binding motif in heme *c* (CXXCH) when histidine and the cysteine dominate because H works as a ligand and C binds to the vinyl groups covalently (Bowman & Bren, 2008). As well as the aromatic residues like phenylalanine and tyrosine in heme *b* and heme *c* and tryptophan in heme *b* which play a crucial role in heme interacting also showed higher relative frequency while the negatively charged residues such as glutamic acid, aspartic acid and lysine recorded fewer relative frequencies, which supports the idea that the heme binding pocket is fundamentally hydrophobic (Smith et al., 2010; Ting et al., 2011). In contrast, arginine which holds a positive charge showed a higher level of occurrences compared to other charged amino acids. A secondary structure analysis of heme *b* illustrated fewer coil types and was more helical in comparison to both heme proteins and all investigated proteins. In contrast, coil types and helices in heme *c* showed similar levels to the background. When relative heme accessibility against solvents was checked, it is found that the heme interacting residues are mostly buried residues (Ting et al., 2011). The

investigation of the sequence motifs that play a pivotal role in heme binding, marked three types of motifs, heme *c* classic motif (CXXCH), binding motif came from heme *b* (GX[HR]XC[PLAV]G) and the binding motif which represents a traditional signature of P450s (FXXGXXCXG) (Nelson, 1999; Otyepka et al., 2007; Ting et al., 2011).

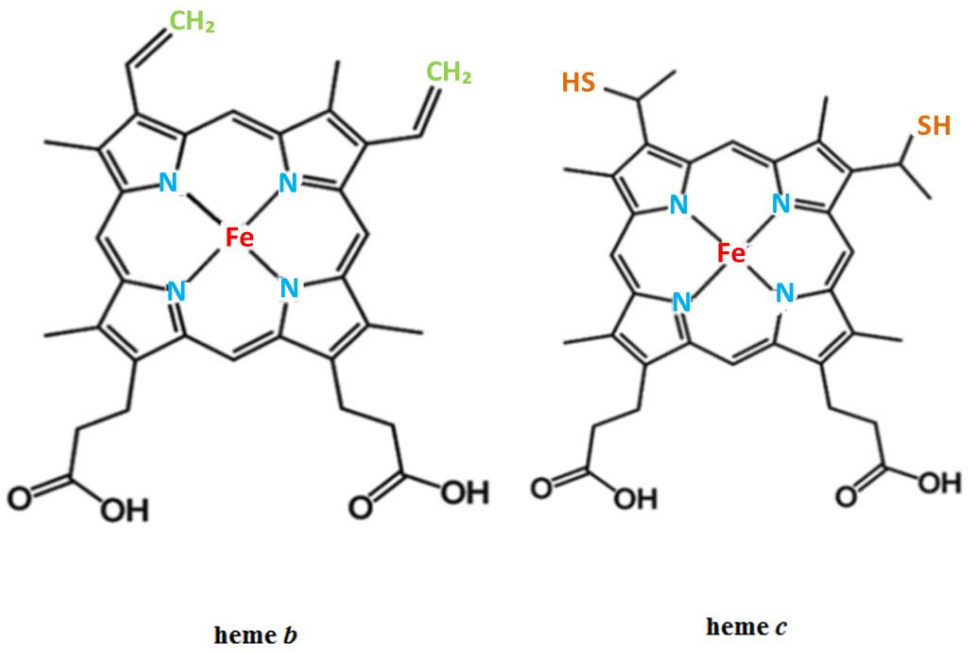


Figure 2-1: Chemical structures of of the natural occurring heme groups; **heme *b*** which links to the host protein non-covalently and **heme *c*** which binds to two cysteine residues of the protein covalently by its vinyl groups**.** The image was acquired with permission from (Ting et al., 2011)

2.3 Heme enzymes

There are many types of hemoproteins were reported previously. In this section seven types of heme enzymes will be discussed in detail. These enzymes were chosen due to their biological functions variations. Figure 2-2 shows different types of heme enzymes.

Figure 2-2: Seven types of hemoproteins. Each one of these enzymes has a biological diffrent function.

2.3.1 Cytochrome P450s

The discovery, structure, catalytic cycle, optical properties, catalysed reactions and applications of cytochrome P450s will be discussed in detail in chapter 3 of this thesis.

2.3.2 Nitric oxide synthases

The importance of nitric oxide synthases comes from their ability to produce nitric oxide (NO), an essential signalling molecule in the nervous, immune and cardiovascular systems by catalysing the oxidation of L-arginine to L-citrulline (Bredt & Snyder, 1994). However, there is no sequence similarity with P450s; the C-terminal part of the molecule is similar to P450 reductase and the N-terminal portion which binds heme gives 450 nm absorption bands when it is reduced in the presence of CO, a specific characteristic of P450, and these two halves provide a self-sufficient oxygenase (Griffith & Stuehr, 1995; McMillan et al., 1992; Stuehr & Ikeda-Saito, 1992; White & Marletta, 1992). Nitric oxide synthase is normally activated when Ca^{2+}-calmodulin binds to the linker region which connects the heme to the reductase domains (Abu-Soud & Stuehr, 1993). An unusual feature of nitric oxide synthase (NOS) is its ability to catalyse the reaction to produce NO using two different mechanisms, the first one in the presence of the cofactor tetrahydrobiopterin (BH4), while the second one is in the absence of BH4 (Poulos, 2014).

Despite the similarity in heme ligation and optical properties between NOS and P450s, these two proteins have completely different architecture. The heme in NOS is encompassed by a beta structure instead of the helices in P450. Also, both the movement of the substrates into the active site and the exit of the product (NO) are much easier than P450 because the heme is more exposed to substrates. The cystine (Cys) in both proteins, NOS and P450s, accepts an H-bond from the NH group, indole NH in NOS rather than peptide NH in P450s, and unlike other proteins, the cofactor BH4 is permanently linked to NOS, donating the one required electron for the oxygen molecule activation (Poulos, 2014). This reaction can be catalysed by NOS in two steps. The first step is similar to the traditional reaction of P450s,

except it requires just one electron and this electron is donated by BH4, forming N-hydroxy-L-arginine from L-arginine (Hurshman et al., 1999; Poulos, 2014; Wei et al., 2001). The second step is the oxidation of L-NHA to produce L-citrulline and NO. In the BH4 mechanism, the peroxy intermediate is created by the reduction of the oxy complex by BH4. After this step, the cyclic intermediate is formed, yielding L-citrulline and NO^-, while in the absence of BH4, the reactive H atom will form hydroperoxyl, followed by the formation of a cyclic intermediate which yields the products (Huang et al., 2001; Poulos, 2014).

2.3.3 Peroxidases

Peroxidases are ubiquitous enzymes owing to their ability to catalyse a variety of substrates. Except for cytochrome c peroxidase (Ccp) which uses cytochrome c as its redox partner, the most traditional substrates of these enzymes are small aromatic molecules (Hiner et al., 2002). Non-mammalian peroxidases are classified into three classes: class I are intercellular peroxidases such as the cytochrome c peroxidase, pea cytosolic ascorbate peroxidase and the gene-duplicated bacterial and fungal catalase-peroxidases; class II are the fungal extracellular peroxidases such as manganese peroxidases and class III are the plant extracellular peroxidase such as HRP. Class II and class III need extra stabilisation represented by four or five disulphide bonds and the binding sites of two Ca ions, while class I does not contain any disulphide bonds. Peroxidases are single polypeptide chains with a single heme group connected to the enzyme by a histidine residue, varying in size between 30–40 kDa (Welinder, 1992).

All three classes share the same catalytic cycle. In the first step, a porphyrin π cation radical (compound I) is formed by the removal of two electrons by the peroxide; one from the iron

and the second from the porphyrin (Dolphin et al., 1971). Compound I has a green colour and different spectral properties from other heme oxidative enzymes. In the second step, compound I will be reduced to compound II (red colour) by gaining an electron from the substrate molecule. At the final step, the Fe^{4+} is reduced to Fe^{3+} by the second substrate (Poulos, 2014). The O-O bond in peroxidase cleaves heterolytically leaving the oxygen atom with only six valence electrons, this oxygen atom is called oxenoid. The mechanism of the formation of compound I was described following the structural details of peroxide activation in cytochrome c peroxidase. This mechanism demonstrated the role of the distal His residue in the formation of compound I. This residue is essential for the acid-base catalysis of the peroxidases. The distal His52 residue transfers the peroxide proton of O1 "an oxygen atom bound to the heme iron" to the other oxygen atom O2 "an oxygen atom not bound to the heme iron" and that leads to heterolytic cleavage of the O−O bond (HAMILTON, 1974). The other residue which plays a crucial role in the peroxidases catalytic cycle is the distal Arg48. In order to stabilise the developed negative charge of the O2 atom, the anionic peroxide ligand interacts with the guanidinium side chain of Arg48 which is positively charged. So, the heterolytic cleavage of O-O bond in Ccp is achieved as a result of proton transformation from the His52 residue to the O2. The only problem with this mechanism is that the distance between O2 and the His52 is too far to form a H-bond. A functional study suggested that a water atom could play a vital role in proton transformation between O1 and O2 instead of His52 (Hiner et al., 2002; Poulos, 2014).

2.3.4 Chloroperoxidases

The main function of chloroperoxidase (CPO) is the production of chlorinate compounds by converting Cl^- into a chlorinating agent like HOCl. This step comes after the formation of

compound I. Compound I is created using the same traditional steps of the heme peroxidases, but the distal His in the heme peroxidases is replaced by the active site Glu in CPO. In addition to chlorination, CPO has the ability to behave as a peroxidase and as a P450 in oxidising different organic molecules (Morris & Hager, 1966; Piontek et al., 2010).

Chloroperoxidase has similar spectral properties to both P450s and NOS owing to the H-bond acceptance by CYS ligand from a nearby NH group. CPO has been known as a peroxide-P450 hybrid, despite the fact that the overall structure of chloroperoxidase is not similar to either P450 or peroxidases. CPO exhibits peroxidase-like distal pocket properties owing to the distal pocket polarity and acid-base catalytic group which is responsible for the heterolysis of the peroxide O-O bond, except this group is Glu instead of His in other heme peroxidases which drops the optimal pH to about 3 (Poulos, 2014).

2.3.5 Heme oxygenases

Heme oxygenases are the enzymes responsible for the degradation of heme to produce biliverdin which is converted to bilirubin. These enzymes play a crucial role in getting rid of the toxic free heme which accumulates in the mammals' bodies due to haemoglobin degradation, approximately 300 mg/day of heme is produced from the degradation of 6–8 g of haemoglobin (Kikuchi et al., 2005; Maines, 1988). Two isoforms of heme oxygenases have been identified, HO-1 and HO-2, which have completely different functions. HO-1 is responsible for heme degradation, while HO-2 is used mainly for CO formation, an important messenger molecule (Farombi & Surh, 2006; Maines, 2005). The structure of the human HO-1 and those which are structurally related to human are not similar to any known heme protein structures. The heme lies between the proximal helices and the distal helices, with the axial

heme ligand served by a histidine residue. Different from the rest of the heme enzymes, the heme itself is a substrate for HO-1 and as a result, has active site flexibility important for the functionally active enzyme. The mechanism of the α-meso heme carbon oxidation by HO-1 will be discussed in more detail because α-meso heme carbon is considered as the main substrate of most heme oxygenases. The first step of the catalytic cycle is the formation of α-meso-hydroxyheme. Initially, it was thought that Fe(IV)=O acts as oxidant in HO just like P450, but this is unlikely as His serves as an axial ligand in HO and the role that the distal ligand plays in heme proteins is different from that in P450s, also the distance between the ferryl O atom and α-meso heme carbon is too far to enable the transformation of proton to an oxygen molecule (Montellano, 1998; Poulos, 2014).

Like P450s and peroxidases, the dioxygen molecule must be cleaved into O1 and O2 by accepting a proton. From a structural point of view, it was suggested that the Asp residue plays an essential role in the O activation process until it was found recently that the most important residue in the HO catalytic cycle is the water molecule in the active site. This water molecule is the key catalytic group which transfers a proton to the iron-linked dioxygen (Matsui et al., 2010; Unno et al., 2007).

2.3.6 Dioxygenases

There are two types of dioxygenase enzymes, the human dioxygenase called indoleamine 2,3 (IDO) and the bacterial enzyme tryptophan 2,3 (TDO). Both enzymes are soluble and have a similar function, which is the conversion of L-Trp to N-formylkynurenine. These enzymes have different names since they were discovered in 1930s, they were called tryptophan pyrrolase, then tryptophan peroxidase−oxidase and eventually named as dioxygenases

(Poulos, 2014). Tryptophan apparently plays a pivotal role in building a strong immune system during the conversion into N-formylkynurenine. As a result, IDO and TDO are considered as possible therapeutic targets to fight cancer cells (Platten et al., 2012). The structures of IDO and TDO are homologous, except IDO has extra domain. The active sites of these two enzymes are also very similar, but Ser 167 in human dioxygenase is His55 in the bacterial type. Based on the crystal structure of the L-Trp binding to the TDO, it is reported that the substrate carboxylate is linked to Arg117 by H-bonds, while the α-amino group bonds to the heme propionate. There are two possible mechanisms for the TDO catalytic cycle. The first one depends on the role of the N1 substrate proton abstracts in activating the substrate. Products using this mechanism were achieved either by dioxetane intermediate or by Criegee rearrangement. The second mechanism showed that the N1 proton transformation is not essential for O_2 cleavage. Based on the later mechanism, the first step in this reaction would be the insertion of dioxygen complex to the L-Trp so the O-O bond will cleave homolytically. In the second step, two actions are combined, the oxygen atom of epoxide will receive a proton from the α-amino N atom of L-Trp and the indol ring will be opened by ferryl O atom attack (Efimov et al., 2012; Poulos, 2014).

2.3.7 Diheme proteins

Diheme proteins or MauGs are involved in tryptophan tryptophylquinone (TTQ) post translational synthesis from the bacterial methyl amine dehydrogenase (MADH) (McIntire et al., 1991). Diheme proteins have two c-type hemes similar to peroxidase, low spin and high spin heme. Unlike peroxidase, these enzymes can use either H_2O_2 or O_2 to oxidise both hemes into Fe^{4+} and form the oxyferryl species (Fu et al., 2009; Poulos, 2014). The two hemes are kept close, sharing spin and charge via the Trp93 residue which lies between the

two hemes (Geng et al., 2013). The O-O cleavage in these enzymes occurs under the control of the general local acid-base mechanism. The challenging question about how the electron is transferred between MADH and MauG has been answered by studying the mutant Trp199 in MauG. Trp199 is located between TTQ and the low spin heme in the midway. It was found that the Trp199 mutant has a negligible effect on heme properties, while it stopped TTQ production which supported the belief that Trp199 plays a pivotal role in electron transformation (Tarboush et al., 2011).

3 Cytochrome P450 superfamily

3.1 Discovery

The discovery of P450 hemoproteins dates back to 1958 when a study of heme proteins in rats and pig chromosomes highlighted a unique spectral feature of the CO-bound ferrous complex of a membrane-bound protein. This complex showed a Soret peak at 450 nm when it was pre-treated with a proper reduced agent (Garfinkel, 1958). The protein from this experiment was later identified as a member of P450 hemoprotein family, which takes its name from the distinctive Soret peak at 450 nm (Omura & Sato, 1962, 1964). The unique absorption band is a characteristic feature, not only for P450s, but also for the other three hemoprotein classes which are chloroperoxidases, protein H450 and NOSs (Danielson, 2002). Cytochrome P450s belong to a superfamily which is classified into 265 different families, with more than 2000 identified protein, genomic and DNA sequences. A member of this family can be found in almost every living organism, prokaryotic and eukaryotic. P450s are also found in archaea, which supports the concept that P450s originally date back to ancient times, approximately 3.5 billion years ago (Danielson, 2002; Whitehouse et al., 2012).

3.2 Nomenclature

In order to name and locate P450 genes into a certain family and subfamily, a standard nomenclature system was created due to the genetic diversity and sequence conservation shortage of these proteins. According to the P450 Nomenclature Committee, three factors need to be considered: “amino acid sequence identity, gene organisation and phylogenetic

association"(Nelson, 2009). Based on this system, P450's name usually starts with three italic letters (*CYP*), which represent a root for all P450s genomic and cDNA sequence names. Next, the family name is written using Arabic numerals, followed by a letter for the subfamily. Subfamily members are allocated their numbers depending on the sequence reported to the nomenclature committee. The pseudogenes in P450s nomenclature system can be distinguished by the letter (P) directly after the family Arabic numeral. A pseudogene is a defective gene responsible for the production of a non-functional protein due to either the denaturation of one copy during the gene duplication or the shortage in the "requisite regulator elements" which are required for mRNA transcription (Danielson, 2002). P450 sequences are placed in families and subfamilies depending on the similarity between their amino acids; 40% similarity qualifies P450 to be classified in the same family, while 55% similarity places it in the same subfamily and 97% similarity is considered as alleles (Nelson et al., 1993). P450 families are classified depending on the cytochrome type. For example, animal P450s are represented by *CYP* (1-49) and (301-499), *CYP* (51-69) and (501-699) represent lower eukaryotes, *CYP* (71-99) and (701-999) are for plants, with *CYP* (101-299) for bacteria (Danielson, 2002).

3.3 Evolution

Many mechanisms have been reported to be responsible for the functional diversity of the P450 superfamily, but the most recognised is the gene duplication process. It is believed that the first CYP member existed 3.5 billon years ago. At that time, due to the lack of oxygen, the P450 protein acted as nitro reductase or endo reductase. Directly after the formation of a

high amount of oxygen, the earlier CYP started to work as oxygen detoxify to protect living organisms (Danielson, 2002; Loomis, 1988; Schlezinger et al., 1999).

By looking at the evolutionary history of P450 multigene family, enormous expansion has taken place because of genome duplication. Indeed, the first expansion related to the endogenous fatty acid metabolism occurred 1.5 billon years ago. Examples of these enzymes are CYP11 and CYP4, which are important to maintain eukaryotic membrane integrity (Nebert & Gonzalez, 1985).

The second expansion is believed to have accorded the "endogenous steroid-synthesised P450 lineages" 900 million years ago. CYP19, CYP21 and CYP27 evolved dramatically to give rise to the CYP1 and CYP2, which are the focus of attention in the study of major drugs and carcinogen metabolism in mammals. The last evolutionary event was the largest to have occurred to many P450s about 400 million years ago. This expansion was motivated by two events. Firstly, the appearance of aquatic organisms and the use of toxic plants in their food. Secondly, the effect of hydrocarbon combustion products on terrestrial organisms (Gonzalez & Nebert, 1990; Nelson & Strobel, 1987).

The phylogenetic analysis of the evolution of P450s revealed that the changes in the CYP genes were very fast in comparison to other proteins. The unit evolutionary period, which is defined as the time required for a protein to show a 1% change in amino acid sequence, has been estimated to be 2–4 million years for P450s, which is a relatively short period in comparison, for example, to 400 million years for the highly conserved histone proteins (Gonzalez & Nebert, 1990; Nelson & Strobel, 1987).

As mentioned previously, the mechanism responsible for the diversity of the P450s is mainly gene duplication. This continuous process produces new genes, which may be functionally inactive in some cases, causing continuous expansion in the family's genes. As a result, the sequence diversity due to gene duplication in a single species is not only responsible for the evolutionary history of this particular species, but also the evolutionary history of the whole gene family (Gonzalez & Nebert, 1990).

3.4 Structure

Cytochrome P450 superfamily members were classified into three basic groups depending on the subcellular localisation: I) the microsomal type from the endoplasmic reticulum, II) the mitochondrial type from mitochondria and III) the cytosolic form. The majority of eukaryotic P450s are insoluble proteins found either as the microsomal type or mitochondrial type, the only exception being nitric oxide reductase (CYP55), which is thought to be acquired from bacteria and did not arise originally in eukaryotes, while the majority of prokaryotic P450s are soluble cytosolic forms (Nelson et al., 1996).

Microsomal P450 can be distinguished by the presence of 20–25 hydrophobic residues consisting of an amino-terminal signal anchor sequence with less known hydrophobic connectors. These residues link the P450 protein to the endoplasmic reticulum membrane. The other characterisation of this P450 type is the presence of the charged residues, a negatively charged residue close to the amino-terminus and polar positively charged amino acids at the carboxy-terminus. These charged residues ensure a proper protein attachment to

the endoplasmic reticulum as well as providing suitable folding to the P450s (Kusano et al., 2001).

Like the microsomal P450, the mitochondrial type can be characterised by the amino-terminus residue, but in this type, the residue is cleaved after protein synthesis and imported to the mitochondria so that the protein is connected to the mitochondria inner membrane by a string of hydrophobic amino acids, like the hydrophobic core residues in the microsomal P450. Similar to microsomal P450, the carboxy-terminus serves to provide the proper folding to the mitochondrial P450s (Kusano et al., 2001).

Nevertheless, even though the sequence identity between eukaryotic and prokaryotic is very low, about 20% or less, the secondary structure of these two types is very highly conserved. The crystal structures of many bacterial cytochromes, such as CYP175A1, CYP121A1, CYP152, CYP101A1, CYP102A1, CYP107A1, CYP108 and CYP111A1, have been determined and analysis of their three dimensional structures have shown that even with a very low sequence homology between different bacterial cytochromes or between the eukaryotic and prokaryotic P450s, the overall protein topography showed a very high similarity (Graham-Lorence et al., 1995; Hasemann et al., 1995). To date, there is a lack of understanding of eukaryotic cytochromes due to crystallisation difficulties of the membrane-bound P450s. The only crystallised eukaryotic P450 is the cytosolic CYP55A1 from the *Fusarium oxysporum,* but as was mentioned previously in this section, this protein was originally a bacterial P450, so the crystal structure of this protein is not applicable for membrane-bound P450 structures. However, two key studies apparently played a pivotal role towards a better perception of the eukaryotic P450s. The first study succeeded in solving the crystal structure of prokaryotic CYP51 from *Mycobacterium tuberculosis,* which showed a

high conservation with the eukaryotic CYP51 family. The similarity between these two proteins is not limited to their structures, but extends to their catalytic activity (Podust et al., 2001). The second important study was performed on the crystallisation of a modified mammalian cytochrome P450 (CYP2C5). In order to crystallise this protein, they created a more water-soluble protein through the replacement of the hydrophobic amino-terminal signal residue with a more water-soluble sequence. The CYP2C5 structure showed high similarity to bacterial P450s structures (Williams et al., 2000).

In general, all P450s have a triangular structural framework with two portions, the rich carboxy-terminal (alpha helix) and the rich amino-terminal (beta sheet), with both the alpha helices and beta sheets located in the same heme plane. Sequence alignments of P450s established a number of sequence motifs, one of which is located in helix I close to the protein centre and has a sequence (G/A)GX(D/E)T with terminal threonine which plays an important role in the creation of the oxygen binding pocket in the enzymes (Poulos et al., 1987). A key sequence in helix K is EXXR, the mutation of any residue in this sequence caused complete functional inactivation of the mutants. Another important residue is cysteine, the fifth ligand of the heme. The heme is buried deep inside the protein and it is surrounded by helices L and I (Danielson, 2002; Nebert et al., 1988).

The regions responsible for substrate-P450 binding gain special attention due to their importance in the catalytic activity of P450s. Many approaches have been applied for identifying the binding sites, including X-ray crystallography, which was used to identify the residues responsible for drug metabolism in eukaryotic P450s. Another predictive approach performed is site-directed mutation to determine differences in location of the amino acids which are closely related but different in their functions. The homology-based approach is

completely unbiased, providing supportive information about residues involved in substrate binding. In general, the experimental purpose of these approaches is to diagnose the amino acid changes responsible for a shift in either catalytic activity or substrate performance. CYP2A4 and CYP2A5 are the best examples of such an approach. Despite those proteins being highly identical (97.8%), they display completely different catalytic activity, but each protein starts to catalyse the reaction of the other protein by reciprocal substitutions of some amino acids. The residues that have the ability to shift the substrate specificity are called "substrate recognition sites (SRS)". The homology-based analysis of P450s proved the significance of the SRSs, not only for substrate selectivity, but also for kinetic differences between P450 isoforms that catalyse the same reaction and metabolise the same substrate. For example, CYP2C4 and CYP2C5 show more than 95% sequence identity and they catalyse the hydroxylation of progesterone 21, but the *Km* of CYP2C4 is about 10-fold greater than the K_m of CYP2C5. It was found that this K_m difference is due to one amino acid located in SRS1. Only 6 SRSs were reported in eukaryotic P450s and the positions of these sites are compatible with the substrate contact site in the prokaryotic P450s-substrate complex (Danielson, 2002; Hanioka et al., 1992).

3.5 Biochemistry

3.5.1 Optical properties

Hemoproteins are characterised by the ability of its heme group to absorb light at certain wavelengths. This chromatography phenomenon is affected by several factors, such as the surrounding environment of the heme, ligands that bind to heme and the heme oxidation

state. The spectroscopic properties of all P450s are dominated by the spin state of the heme iron. A hexacoordinated iron is represented spectroscopically by a low spin state, while a pentacoordinated iron, when the heme loses the distal ligand, is displayed as a high spin state. For a low spin state, the Soret peak is normally between 416–419 nm, while for the high spin state, the absorption maximum is shifted towards the blue region between 390–416 nm, with a trough at about 420 nm. The differences in the spectrum between the bound and unbound proteins yield a Type I difference spectrum. Many substrates, like saturated and unsaturated fatty acids and aromatic compounds, show a Type I difference spectrum when they are bound to P450s. The dissociation constant Kd can also be calculated depending on the differences between the substrate-free protein and substrate-bound protein in the spectrum. Another type (Type II) of difference spectrum is obtained when inhibitors, such as imidazole, nicotine, methylenedioxyphenyl compounds or 1,1-dialkylhydrazines, bind to a P450. The Soret peak of Type II can be seen at 426–435 nm with a trough at approximately 390–405 nm. A shift in the spin state also occurred when CO bound to the reduced P450 protein, due to the nitrogen atom of a histidine imidazole linked to the iron causing a shift in the absorption maximum to 447–452 nm (Danielson, 2002). Type I and Type II changes are shown in Figure 3-1.

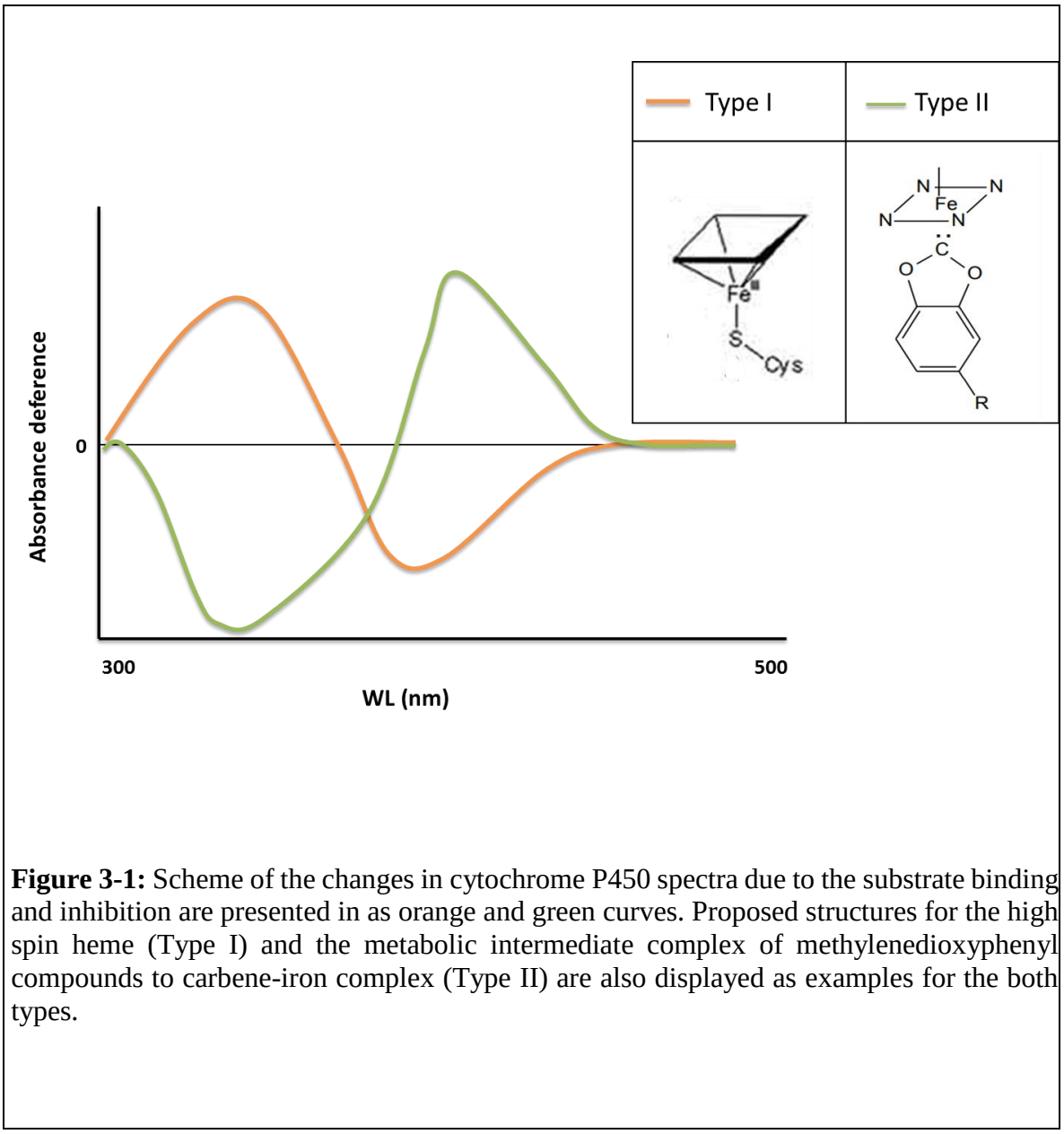


Figure 3-1: Scheme of the changes in cytochrome P450 spectra due to the substrate binding and inhibition are presented in as orange and green curves. Proposed structures for the high spin heme (Type I) and the metabolic intermediate complex of methylenedioxyphenyl compounds to carbene-iron complex (Type II) are also displayed as examples for the both types.

3.5.2 Catalytic cycle

The hydroxylation of organic compounds represents a signature reaction of all P450s. To enable this reaction, the protein needs to activate the inert molecular oxygen splitting it into two oxygen atoms, one is introduced into the H-C bond of the organic substrates like fatty acids and aromatic compounds, while the other is released as a water molecule. In addition

to atomic oxygen, this reaction requires a transfer of two electrons from a suitable donor via protein domains. NADPH and NADH are the preferred coenzymes for the chromosomal P450s, while in mitochondrial P450s, the required electrons are obtained from adrenodoxin and adrenodoxin reductase. In this protein system, a hydride ion transfers from NADPH to FAD, which is caused by the generation of the reduced form of two adrenodoxin molecules and two electrons. After that, each electron will be used to reoxide one adrenodoxin molecule. Finally, every adrenodoxin molecule will transfer a single electron to the heme iron of the hemoprotein (Danielson, 2002; Sligar & Murray, 1986).

The catalytic cycle is shown in (Figure 3-2). From this figure, the P450 heme converts from a low spin state (hexacoordinate) to the high spin state (pentacoordinate) immediately after substrate binding. Receiving one electron reduces the iron from the ferric form (Fe^{+3}) into a ferrous form (Fe^{+2}). In the next step, an oxygen molecule will bind to the ferrous iron to produce ferrous dioxygen-bound complex, then a second electron will be accepted to form the ferric peroxy complex. The oxygen molecule will then be cleaved into two oxygen atoms by receiving two protons, forming the oxyferryl intermediate with (Fe^{IV}). This intermediate represents the catalytically reactive heme. Finally, the oxygen atom of the oxyferryl will be transferred to the substrate to produce the hydroxyl form of that substrate.

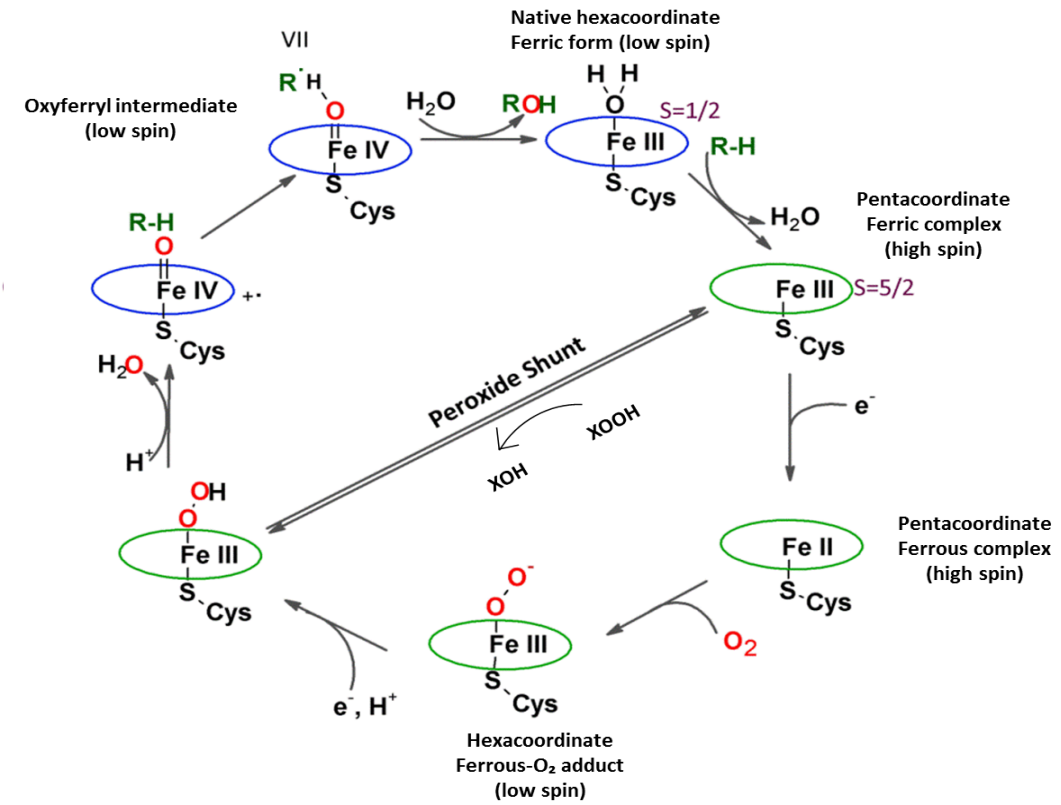


Figure 3-2: : Scheme of cytochrome P450 enzyme catalytic cycle. This figure shows the basic steps of substrate hydroxylation by P450; substrate binding, receiving the first electron, binding of molecular oxygen, accepting the second electron, oxygen cleavage, forming the oxyferryl intermediate with (Fe^{IV}) and finally produce the substrate hydroxyl. Image was acquired and modified with permission from (Erdogan, 2014)

3.6 Cytochrome P450-catalysed reactions

In general, cytochrome P450 catalysed reactions can be classified into four basic categories: I) the most common reaction is the hydroxylation of organic substrates, when a hydrogen atom is replaced by hydroxyl group; II) epoxidation reactions, in this reaction the C-C bond is cleaved when an oxygen atom is introduced to it; III) heteroatom oxidation reactions

(heteroatom such as sulphur or nitrogen); IV) reduction reactions occurring under limited oxygen conditions (Danielson, 2002; Goeptar et al., 1995).

3.6.1 Hydroxylation

The most common substrate for the hydroxylation reactions of P450s are the aliphatic and the aromatic hydrocarbons. One of the important issues in the hydroxylation is the site of the oxygen atom insertion in the C-H bond, which depends mainly on the required energy to cleave this bond. For example, the hydroxylation of hydrocarbon chains is more likely to occur at the site not at the terminal methyl group due to the differences in the C-H bond strength in these two sites. An exception to this can be seen in mammalian P450 (CYP4A), which prefers to hydroxylate fatty acids at the ω position instead of the ω-1, regardless of the fact that the later position has a weaker carbon bond than the first position. The explanation of this behaviour is still unknown, but it could be an access limitation stands behind the hydroxylation of the terminal methyl group in this case (CaJacob et al., 1988; Danielson, 2002). However, the presence of an adjacent heteroatom, such as a methyl group, on an aromatic ring gives this site the preference. Another possible hydroxylation reaction is the hydroxylation of a methyl group linked to N or O atom, but the hydroxylation of a methyl group with more than one halogen atom could cause protein acylation, which seriously affects both protein structure and function (Danielson, 2002; Mehendale et al., 1994; Raucy et al., 1993).

3.6.2 Epoxidation

The majority of P450 epoxidation studies focused on the olefins epoxidation reactions to their epoxides, such as the production of the corresponding epoxide and trichloroacetaldehyde from the epoxidation of trichloroethylene. However, the epoxidation of the terminal olefin group like octane or ethylene causes protein inactivation. Surprisingly, in limited cases (1 out of 100), the reaction diverts from its path when the olefin terminal carbon alkylates one of the heme associated nitrogen atoms. This rare reaction is known as a "suicide inactivation" because it inactivates the catalysed protein (Danielson, 2002).

3.6.3 Oxidation of aromatic rings

The discovery of ability of the P450s to hydroxylate different aromatic substrates dates back to 1950s, when acetanilide hydroxylation was studied by a number of researchers in the National Institutes of Health in Bethesda, Maryland. The objective of this study was to identify the role of microsomal hydroxylases in drug metabolism. An example of the aromatic compounds hydroxylation is the conversion of "5-nitrobenzo[b] naphtho[2,1-d]thiophene", an environmental pollutant with a high risk to human health, into DNA adducts by cytochrome P450 (Danielson, 2002).

3.6.4 Dealkylation

The ability of cytochrome P450 to catalyse heteroatom oxidation has attracted much biomedical research attention. The molecules bearing atoms with high electronegative

charge, such as nitrogen or sulphur, are more likely be oxidised by P450s than molecules with low electronegative atoms like oxygen. This reaction requires free electron pairs on the heteroatoms and two different mechanisms have been reported. The first mechanism is when ferryl oxygen (the electron deficient molecule) moves from iron to the electron pairs of sulphur or nitrogen to produce sulphoxide or N-oxide metabolites, the sulphoxide bearing electron pairs can also oxidise again to produce a sulphone. However, the oxidation of oxygen heteroatoms by P450s is unlikely to accur; the oxidation of bromide and iodide substituents by breaking the carbon-halogen bond was reported. The second mechanism involves the alkyl-substituted heteroatoms dealkylation; this reaction requires two successive steps. Firstly, the carbon atom neighbouring the heteroatom is hydroxylated, then the heteroatom atom is extruded to produce a carbonyl group. This mechanism is normally applied for sulphur and oxygen dealkylation, but it takes a different path for nitrogen dealkylation. The transformation of one electron from nitrogen to ferryl oxygen will give the nitrogen atom a positive charge. This positively charged nitrogen will cause acidification of the proton on the carbon adjacent to the heteroatom atom, and produce either N-hydroxyl derivative of the substrate or N-dealkylated product when it is transferred to the ferryl oxygen group (Danielson, 2002; Seto & Guengerich, 1993).

3.6.5 Unusual P450 reactions

Cytochrome P450s were extensively studied due to their ability to catalyse a wide range of reactions. In addition to the four traditional types of reactions that were described in sections (3.6.1–3.6.4), many P450s are also involved in different unusual reactions.

Although, reduction reactions catalysed by P450s are not common, the variety of these reactions needs to be taken in consideration. The mechanism of this reaction remains uncertain, but the reduction could happen when the $Fe^{2+}O2$ complex works as a reduced agent during the waiting of an electron. A classic example of this type of reaction is the reduction of alkyl halides in the presence of CCl_4 and halothane (Mico et al., 1983; Van Dyke & Gandolf, 1976). Another example is the reduction of nitrogen oxides, such as nitroxides, C- and N-nitroso compounds, hydroxylamines and nitro groups by P450s (Guengerich, 2001; Wislocki et al., 1980). Additional unusual reduction reactions are the reduction of glyceryl trinitrate to produce nitric oxide and the reduction of benzamidoximine to produce benzamidoxime as a final product (Clement et al., 1997; Delaforge et al., 1993). Cytochrome P450s can also reduce inorganic substrates in addition to organic compounds, for example, the reduction of SO_2 or $HSO3^-$ to $SO_2^{\cdot -}$, which is in equilibrium with dithionite, $S_2O_4^{2-}$ (Lambeth & Palmer, 1973).

A second unusual type of reaction catalysed by P450s is dehydrogenation, also known as alkane desaturation. Many organic compounds have been reported as dehydrogenation substrates, such as valproic acid, testosterone, warfarin, lovastatin, ethyl carbamate and bufuralol. It is believed that this reaction shares the same starting point with classic C-hydroxylation, that is, abstraction of a hydrogen atom. During the reaction, there is competition between oxygen rebound and a hydrogen atom abstract, but the C-hydroxylation represents a key step in this reaction. To date, there is no clear explanation why some C-hydroxylations are accompanied by dehydrogenation and others are not (Fisher et al., 1998; Guengerich, 2001; Rettie et al., 1988).

Another uncommon reaction catalysed by P450s is the oxidative cleavage of a carboxylic acid ester. This reaction was first discovered in rat P450 2C11 (Guengerich, 1987; Guengerich et al., 1988). The product of this reaction is normally carbonyl instead of alcohol. It is suggested that this reaction is a C-hydroxylation similar to the classic O-dealkylation reaction (Guengerich et al., 1988). Ring expansion, when stable five or six membered rings are formed from unstable membered rings as in the production of pyrroline from cyclobutylamine (Bondon et al., 1989), is also one of the unusual P450s reactions. P450s can catalyse, not only the ring expansion reaction, but also the formation of non-existing rings such as the oxidation of dihydropyridines which results in the production of a very stable five membered lactone ring (Guengerich et al., 1988). Cytochrome P450 catalytic reactions are not limited to the reactions described above, but also include numerous other reactions such as aldehyde scissions, dehydration, one electron oxidation, coupling reactions, isomerisation, phospholipase D activity and rearrangement of fatty acids and prostaglandin hydroperoxides. It is expected that new P450s catalytic reactions will be discovered in the near future due to the intensive research focussing on the study and improvement of reactions of this P450 superfamily (Guengerich, 2001).

3.7 Industrial applications of cytochrome P450s

CYP proteins apparently play a pivotal role in the biodegradation and modification of chemical compounds due to their substrate diversity. The adaptation of some organisms to a new diet in some circumstances leads to the further expansion of the variety of P450 substrates. Recent studies have shown that there are no similar enzymes to the cytochrome

P450 superfamily in terms of substrate acceptability and diversity of catalysed reactions, making the members of this family outstanding biocatalysts in many practical applications. In spite of the broad applicability of these enzymes, their applications have been hampered by many limitations including: I) electron transfer requirement and transformation limitation; II) the need for NAD(P)H and uncoupling between the NAD(P)H oxidation and the product formation; III) limited 3D structures for the substrate-protein complex; IV) low activity (Bernhardt, 2006; Bernhardt & Urlacher, 2014; Julsing et al., 2008; O'Reilly et al., 2011; Urlacher & Eiben, 2006).

P450 enzymes were described by many researchers as biological manufacturer enzymes of many pharmaceutical compounds and fine chemicals. CYP71AV1 is reported as a biocatalyst of the three step reaction to produce artemisinic acid, which is used as anti-malarial drug by the oxidation of amorpha-4,11-diene into intermediates, artemisinic aldehyde and artemisinic alcohol, using the engineered *S. cerevisiae* strain (Paddon et al., 2013; Ro et al., 2006). In 2010, more than 655,000 people died because of malaria and more than 200 million malaria cases were reported, therefore there is a need for a stable and relatively inexpensive source of artemisinin. The yield of artemisinic acid was very low, approximately 100 mg/l in 2008, which hindered the industrial application of this reaction until 2013, when a combination of improvements was applied, methods of synthetic biology, recombinant protein expression, downstream processing and protein engineering. These efforts successfully increased the artemisinic acid yield up to 25 g/L. Currently, CYP71AV1 is used by Sanofi (France) for the industrial manufacture of artemisinin (Bernhardt & Urlacher, 2014).

In addition to anti-malarial drugs, CYPs enzymes are involved in the production of anticancer agents, anti-inflammatory materials and antibiotics. Taxol, a chemotherapeutic agent, is

synthesised in *E*. coli at a concentration of 1 g/L from geranylgeranyl pyrophosphate (GGPP) in a series of reactions ending in taxol production. Taxadiene-5α-hydroxylase produces taxadien-5α-ol in the first step of this reaction, which is later converted to taxol (Ajikumar et al., 2010).

Another remarkable example of a multi-enzyme system application was reported in hydrocortisone production. In the first step, engineered *S. cervisiae* serves to produce ergosta-5-ene-ol and ergosta-5,22-diene-ol (CYP11A1 substrates to produce pregnenolone). The yelled pregnenolone in a multi-step reaction using 3-β-hydroxysteroid dehydrogenase, CYP17A1, CYP21A1, and CYP11B1 to produce hydrocortisone as a final industrial product (Szczebara et al., 2003).

Parvastin, which is used widely around the world to reduce the cholesterol levels, is produced in a two-step reaction. The first step is the compactin biosynthesis by *Penicillium citrinum* and the second step is the sodium compactin hydroxylation to produce parvastin using CYP105A3 (Arai, 1988; Bernhardt & Urlacher, 2014). CYPs also participate in the biosynthesis of vitamin D3 metabolites. CYP105A1 from *Streptomyces griseolus* converts vitamin D3 directly into its active form 1α,25-dihydroxyvitamin D3 through two oxidation steps (Sakaki et al., 2011; Sasaki et al., 1992).

An available industrial application of plants CYPs is their ability to biosynthesise anthocyanin, which results in the delphinidin-derived pigments in violet or blue flowers. Two CYPs are involved in anthocyanin biosynthesis, CYP75B and CYP75A. Suntory Ltd (Japan) and Florigene Pty Ltd (Australia) have successfully used these CYPs to develop genetically engineered blue roses, which is impossible to achieve by traditional plant hybridisation

methods (Bernhardt & Urlacher, 2014; Holton et al., 1993; Katsumoto et al., 2007). These P450s' products and other products are listed in Table 3-1

Table 3-1: Some of the CYP's commercial products. The table was produced depending on the data from (Bernhardt & Urlacher, 2014; Girvan & Munro, 2016)

Substrates	P450s involved	Products
Amorpha-4,11-diene	CYP71AV1	Artemisinic acid
Taxadien-5α-o	Taxadiene-5α-hydroxylase	Taxol
Ergosta-5-ene-ol and ergosta-5,22-diene-ol	CYP17A1, CYP21A1 and CYP11B1	Hydrocortisone
vitamin D3	CYP105A1	1α,25-dihydroxyvitamin D3
Roses	CYP75B and CYP75A	Blue roses
Sugar	CYP82Y2 and CYP719B1	Morphine
Tyrosine	CYP from *Petunia hybrida*	Fisetin

Part II: Tolerance of cytochrome P450 BM-3 to non-conventional solvents

4 Introduction to cytochrome P450 BM-3

4.1 Domain architecture

Cytochrome P450s are widely known heme-dependent monooxygenases found in many eukaryotes such as fungi, plants and mammals as well as prokaryotes, but the bacterial types are more preferred because of their purification and crystallisation simplicity (Munro et al., 2002; Nelson et al., 1996). A bacterial type P450 BM-3 from *Bacillus megaterium* (118 kDa) is more thermostable, highly active and easier to handle in comparison to other P450s (Bernhardt, 2006). P450 BM-3 protein has two distinct domains, the heme domain (BMP) at about 55 kDa and the reductase domain (BMR) at approximately 65 kDa. The latter consists of two prosthetic flavin groups in an equimolar ratio: flavin adenine dinucleotide (FAD) and flavin mononucleotide (FMN) (Narhi & Fulco, 1986), with both domains matching the chromosomal P450s domains well (Govindaraj & Poulos, 1995). BMR aligned well with the mammalian P450 reductases (CPR), while BMP shows similarity to the eukaryotic cytochromes such as CYP2A and CYP2A subfamilies (Porter, 1991). When BMP and BMR were expressed and purified separately, P450 BM-3 appeared to have lost its activity. This discovery showed the role of the linker region on enzyme activity. This was further supported in latter research when a number of residues were deleted from the linking region resulting in P450 BM-3 inactivity (Govindaraj & Poulos, 1995; Munro et al., 1994). BMR is less stable than BMP and undergoes proteolysis easily to the FMN binding region and an FAD/NADPH-binding region, both subdomains are inactive individually (Oster et al., 1991). P450 BM-3 BMP has remarkable homology with eukaryotic P540 sequences, making it a potential template for eukaryotic P450 sequence analysis because the crystal structure of BMP has

been fully elucidated (Amarneh et al., 1993; Loughran et al., 2000). Comparing BMP with prokaryotic heme showed some minor differences, such as CO-binding behaviour, and a major variation regarding electron transformation because in contrast to P450 BM-3, which receives two electrons vie integrated flavin domain, other bacterial P540s required an external partner to transfer electrons to the heme during catalytic reactions (Boddupalli et al., 1990; Tuck et al., 1992).

4.2 Protein expression

The expression of P450 BM-3 cannot be induced by monounsaturated acids or saturated straight-chain fatty acids in *B. megaterium* (Narhi & Fulco, 1982). It is found that polyunsaturated analogues, like arachidonic and linoleic acids, in spite of their toxic effects when they added exogenously to *B. megaterium*, bind more strongly than other straight-chain analogues to the BM-3 repressor Bm3R1 and induce the expression of BM3 (English et al., 1994; Liang et al., 1998; Ruettinger & Fulco, 1981). Other inducers of BM-3 expression in *B. megaterium* are branched-chain saturated fatty acids. These fatty acids apparently play a pivotal role in the membrane fluidity regulation and it is thought that they could be the original BM-3 substrates owing to their high ratio (about 90%) in the total fatty acid content in *B. megaterium* (English et al., 1997; Kaneda, 1991). Although, P450 BM-3 is not able to metabolise barbiturates, it is reported that barbiturate members are strong inducers for BM-3 expression (Fulco et al., 1983; Narhi & Fulco, 1982).

4.3 Protein structure

The crystal structure of P450BM-3 has been extensively studied by many researchers. The resolved structure of the substrate-free protein has a long hydrophobic substrate access channel. This channel starts from the protein surface charged residues and extends to the buried heme iron (the active site for the substrate hydroxylation) (Huang et al., 2007; Li & Poulos, 1995).

An open confirmation was distinguished in the substrate-free structure. It was found that this open conformation is too wide to serve in substrate binding. In order to enhance substrate binding, it was suggested that the protein will reshape the access channel during the binding process, resulting in more closed conformation (Whitehouse et al., 2012). A substrate-free (SF) form differs from a substrate-bound (SB) by F/G loop, F helix and G helix positions. Palmitoleiate-BMP complex was the first solved structure. From the analysis of the substrate-protein complex crystal structure, it was observed that F and G helices turn around their axes 180^{o} resulting in locking the access channel and preventing contact with the I helix and salt bridge. Fatty acid bound to P450 BM-3 in the hydroxylation position near the heme iron with (ω-1), (ω-2) or (ω-3) carbon positions, while the (ω) end is confined between Phe 81 and 87 residues due to the 90^{o} rotation of Phe 87 and rearrangement of the methyl groups of the close side-chains which are in contact with the Ala 82 residue (Huang et al., 2007). The binding of N-palmitoyl glycine (NPG) to P450 BM-3 is the most studied complex structure because NPG is more soluble than the equivalent chain fatty acids and it binds to the protein more tightly (Haines et al., 2001). The P450 substrate-free structure (PDB code 1BU7) and the structural changes in P450 BM-3 during NPG binding (PDB code 1JPZ) are shown in Figure 4-1. The changes in water molecules are the most significant phenomena in the substrate

binding process, highlighting the crucial role of water in this process. The access channel and active site of the SF contain approximately 17–21 water molecules, while the number is reduced to only three molecules in SB. One of the important water molecules that will be removed during the substrate binding is Wat202 (Figure 4-1 C). This molecule plays an important role in dioxygen binding in the SF form by stopping the formation of a hydrogen bond as a result of creating a kink in the I helix. The removal of Wat202 will allow the formation of a traditional hydrogen bond, reducing the kink angle of the I helix from 13º to 5º, the whole hydrogen bond network will be widely rearranged due to the Gly265 repositioning and His266 rotation. Wat23, the axial water molecule will be displaced to an alternative position, Wat502, due to the I helix movement during the replacement of this water molecule to stay connected to Ala264 by hydrogen bonds. As a result of Wat23 ligand replacement, the heme iron centre will change from hexacoordinate to pentacoordinate . The proteolysed FMN-BMP complex has also been determined. In this structure (PDB code 1BVY), the two asymmetric molecules of the BMP were observed together with the FMN domain situated close to the centre, interrupting the Pro382–Gln387 reign. (Whitehouse et al., 2012). In an attempt to understand the enzyme inactivity in the presence of organic solvents, a structural study on the DMSO-BMP complex was performed at 14% v/v and 28% v/v DMSO (Kuper et al., 2007). In the low solvent concentration, Wat23 was off-axial at 3.75 A˚ instead of 2.6 A˚ in the DMSO-free protein. At the high solvent concentration, DMSO was linked to the heme as a sixth ligand by the sulphur atom instead of oxygen. The DMSO-protein complex behaviour was represented well spectroscopically through the partial high spin at 14% (v/v) solvent against the typical low spin state at 28% v/v DMSO (Kuper et al., 2007).

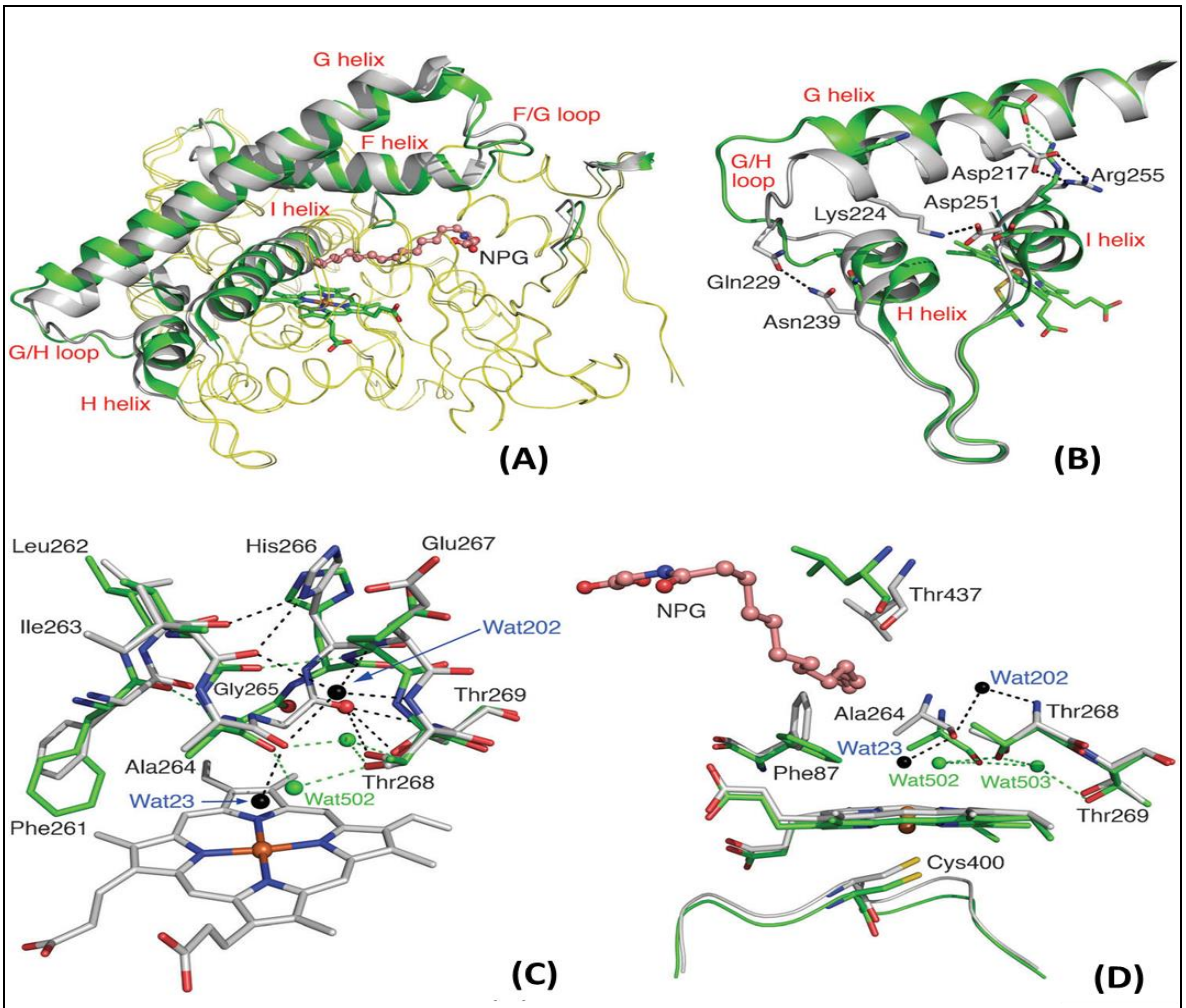


Figure 4-1: Structural changes induced in P450 BM-3 by substrate binding. The substrate-free BMP structure (PDB code 1BU7) first time crystilised by Sevrioukova et al. (Sevrioukova et al., 1999) has been shown in grey and the structural changes in P450 BM-3 during NPG binding (PDB code 1JPZ) (Haines et al., 2001) in green. The water and hydrogen bonds in black, and the haem iron in orange, and NPG in plum. **A**. Complete overlay **B**. Regions of the G-, H- and I-helices. **C**. The kink region of the I-helix. **D**. The active site. This figure acquired with permission from (Whitehouse et al., 2012).

4.4 Dimerisation

The discovery that fatty acid oxidation activity of P450 BM-3 is directly proportional to the enzyme concentration supported the belief that enzyme dilution may cause a dissociation of the function of this enzyme (Matson et al., 1977), which was further supported when it was found that electron transfer could be intermolecular. During a sedimentation velocity study, the P450 BM-3 enzyme mixture consisted of monomers, dimers, trimers and tetramers, but the overwhelming majority in the mixture was dimers. Not only P450 BM-3, but also other P450s such as eukaryotic soluble NOS and cyctochrome c reductase, lost their activity when monomeric (Black & Martin, 1994). When enzymes lose their activity due to the deletion of some residues from their linker region, it may be because this linker region becomes too short to form the proper functional dimer (Urlacher & Eiben, 2006). Aiming to prove that intermolecular electron transfer occurs in P450 BM-3, two mutants of this enzyme were mixed. The BMP of the first one was inactivated, while the BMR of the second mutant was FMN-deficient. Surprisingly, the P450 BM-3 mutant mixture showed distinguished hydroxylation activity, providing strong evidence that the electrons in P450 BM-3 transfer by intermolecularly instead of intramolecularly (Neeli et al., 2005). Despite the evidence supporting the dimerisation theory, there is experimental evidence to show that enzyme inactivity could be due to FMN depletion after a long preincubation time instead of monomerisation (Haines et al., 2001).

4.5 Electron transfer and redox partners

One of the catalytic activity requirements of P50 BM-3 enzymes is the transfer of two electrons from a cofactor such as NADH or NADPH through FAD and FMN domains to the iron heme centre. Flavin reduction and semiquinone formation play a crucial role in electron transfer and BMR function. In P450BM3, the FMN domain receives the first electron from the FADH2 and as a result of this transformation, three components are formed, anionic FMNH semiquinone (red colour), FADH semiquinone (blue colour) and heme-reducing species (Sevrioukova & Peterson, 1995; Sevrioukova et al., 1996). The transformation rate of the first FMN-to-heme electron is known as kf and is calculated by measuring the increase in the rate at 450 nm absorbance when the FeII(CO) complex is formed. Stopped-flow techniques are usually used for kf estimation, these measurements are normally taken in CO-saturated buffer after adding NADPH to the SB enzyme (Govindaraj & Poulos, 1995). In a study regarding the relationship between the catalytic reaction rate and the electron transfer speed, it was found that the electrons were transferred to the heme much slower when laurate bound to the enzyme instead of myristate. This indicated that the kf value is directly proportional to K_{cat} and both of them are dramatically decreased when some residues were deleted from the linker region (Munro et al., 1996).

All P450s require a cofactor to reduce oxygen and donate the required electrons for enzyme catalytic activity. In P450 BM-3, either NADH or NADPH is used as a cofactor. Using these cofactors is usually associated with uncoupling problems. Coupling efficiency is "the percentage of reducing equivalents from NAD(P)H that are utilised for the oxidation of substrate". The coupling efficiency varies from enzyme to enzyme and from catalytic reaction to another for the same enzyme. In the oxidation reaction of $P450_{cam}$ (the cytochrome

from *Pseudomonas putida* that catalyses the hydroxylation of camphor), the coupling efficiency was 100%, that is, all electrons that have been donated by NADH were used by $P450_{cam}$ in the oxidation forming 5-*exo*-hydroxycamphor. However, the coupling efficiency of the NADH when the same enzyme was used in the styrene epoxidation was no more 7%. There are many uncoupling mechanisms which could cause untargeted consumption of the electros such as: I) ferrous-oxy intermediate auto-oxidation; II) two electron reduction resulting in water molecule formation; III) intermediate decomposition resulting in H_2O_2 formation. Uncoupling due to peroxide formation may occur either by the promotion of H_2O_2 dissociation or by the inhibition of the cleavage of the O-O bond. The peroxide dissociation happens when the substrate is unable to remove a water molecule from the active site, so when the H_2O molecule accesses the active site, the polarity is raised and a charge is separated from iron producing a hydrogen peroxide anion (Cirino, 2004).

4.6 Protein engineering

Protein engineering using a biological approach has emerged as a superior tool to improve enzyme catalytic activity, producing tailor made enzymes able to work with new environments and substrates. This technique has already been successfully performed with many P450s, however, eukaryotic P450s catalyse a wide range of reactions and use a variety of substrates, hence their catalytic ability was hindered by solubility and isolation issues. As a result, the majority of P450s protein engineering research has focused on the soluble bacterial membranes of this family, in particular on P450BM3, a soluble self-sufficient highly active enzyme (Whitehouse et al., 2012).

Two different strategies of genetic approaches have been used to improve this enzyme: site-directed mutagenesis and the random mutagenesis/directed evolution approach (both approaches are shown in Figure 4-2). Recently, a combination of these two types were also utilised and have become increasingly important for novel enzyme expression (Urlacher et al., 2006).

Site-directed mutagenesis is the earliest engineering approach successfully used to improve protein specifications (activity, selectivity and stability). This method requires a wide knowledge of enzyme structures and the relationship between these structures and protein activity, which is a major limitation in rational approach application. The first site mutagenesis was applied at Ala328, Thr268 and Phe87 to enhance the ability of P450 BM-3 to catalyse chemical substrates, such as 1,1,2,2-tetrachloroethane and 2-phenylpropanal (Alworth et al., 1997). Other mutations, such as F87G and F87V, increased enzyme activity when used with propylbenzene and 3-chlorostyrene, while the F87A variant showed lower activity, but higher selectivity giving 54% β-hydroxylation versus 1% by the WT (Appel et al., 2001). This engineering approach was also used to produce the high-value primary-alcohols from n-alkanes. For example, the mutant A328V catalysed the oxidation of octane to produce 5% 1-octanol and 82% 2-octanol in comparison to zero and 17% in the WT respectively (Peters et al., 2003). Other mutants also achieved high production of either 1-octanol or 2-octanol from octane such as A82L, F87V/A328F, V78T and A82G (Peters et al., 2003). Site mutagenesis was also used to improve enzyme accessibility towards short chain fatty acids by applying a double mutation at L75T/L181K. By using this mutant, the enzyme was able to oxidise hexanoic acid giving K_{cat} (43 s^{-1}) instead of (3.8 s^{-1}) for WT (Ost

et al., 2000). Although a site mutagenesis approach succeeded in improving P450 BM-3 activity, selectivity and stability, good structural knowledge is required.

In recent years, directed evolution has won protein engineering researchers' attention because it does not require accurate information about protein structure. It is also able to be applied with a wide variety of design problems, in particular, obtaining new functional proteins. It is a random mutagenesis method using one or many start points, which are chosen depending on the specific goal of engineering (selectivity, stability or activity) (Arnold, 1998). In this process, the protein of interest should be encoded first, then screened before the improved protein variant can be isolated. The choice of suitable screening method is essential to the success of directed evolution experiments. Directed evolution applications have contributed to the development of protein properties knowledge and a wide range of molecule synthesis from building blocks to drug analogues, as well as the massive enhancement in activity, selectivity and stability not previously encountered in nature (Denard et al., 2015; Johannes & Zhao, 2006).

Random mutagenesis followed by site-saturating to produce a second-generation variant, A74G/F87V/L188Q, also known as GVQ variant, was used to generate indigo-producing mutants. This variant was also able to oxidise a number of chemical substrates such as chlorinated dioxins, organophosphorus pesticides and polyaromatic hydrocarbons (PAHs) (Li et al., 2000). Furthermore, when the GVQ mutation was applied to CYP102A3, GVQ_{A3} showed a great improvement in enzyme activity with capric acid against WT_{A3} (35 s^{-1} *vs.* 0.8 s^{-1}) (Lentz et al., 2004). Further investigation lead to the discovery that when $A330V_{A3}$ was used instead of Glu in $GVQ_{A3,}$ the ability of the enzyme to produce 1-octanol from octane increased to form half of the product mixture (Lentz et al., 2006). A similar role was applied

on CYP102A1 (P450 BM-3) to generate $A328V_{A1}$. Starting from F87A as a platform, eight random generations were screened and the most active mutations were combined to produce the V26T/R47F/A74G/F87A/L188K variant, which showed impressive enhancement in protein activity (Li et al., 2001). Furthermore, when the F87V mutant was used instead of F87A as a platform, even more enhancement in activity was observed (Lentz et al., 2001). One of the most interesting directed evolutions, is that identified during indigo screening in an experiment aiming to improve selectively towards small non-natural substrates. The variant from this study gave 30% propylphenol from propylbenzene in comparison to 1% for the WT. In addition, F87A and A330P were shown in the same screening to yield 78% 2-phenyl-1-propanol from propylbenzene against zero for the WT (Whitehouse et al., 2012). Directed evolution was also used by Wong in 2004 to enhance P450 BM-3 tolerance towards co-solvents (Wong et al., 2004). The engineered P450 BM-3 was significantly more resistant in the media containing DMSO and tetrahydrofuran (THF), showing high tolerance in the present of acetone, acetonitrile, dimethylformamide and ethanol (Wong et al., 2004). There have been numerous attempts to improve P450 BM-3 performance through applying a successful protein engineering approach, many of which are ongoing, giving hope for tremendous developments in this area.

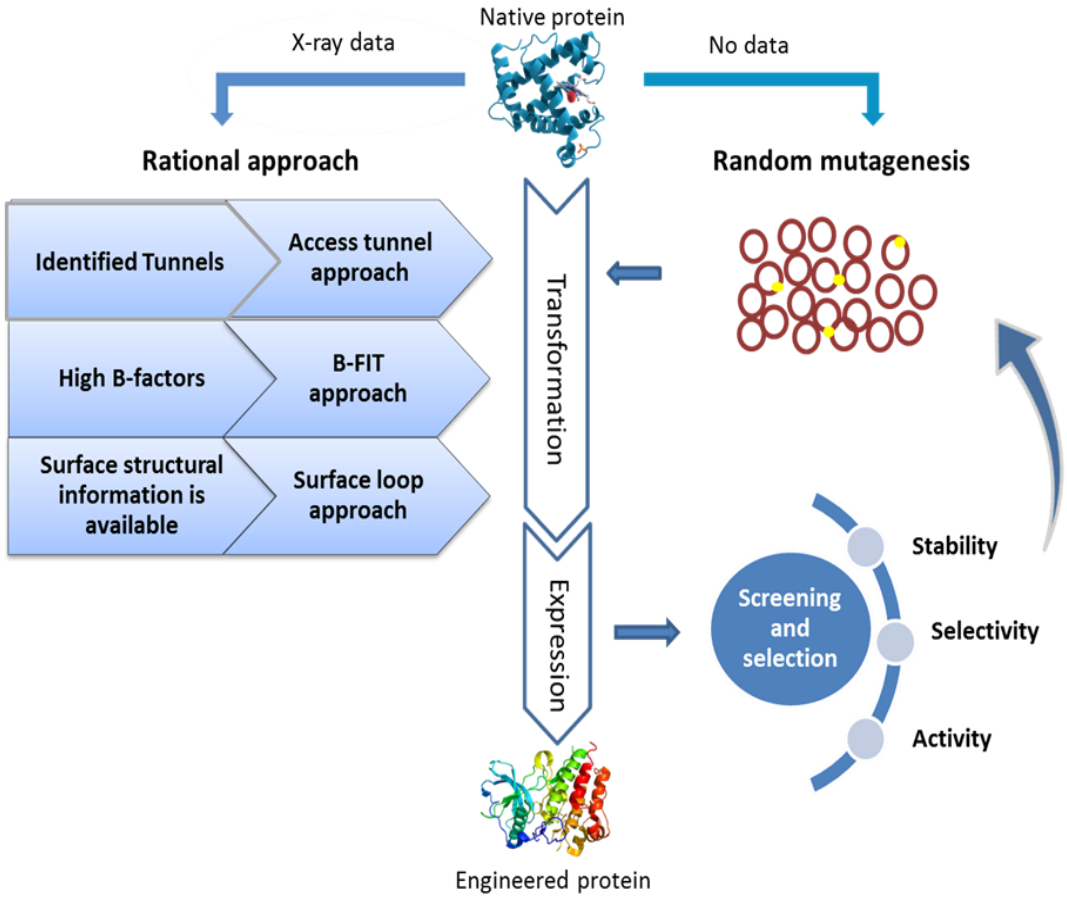


Figure 4-2: Biological approaches for enzymes engineering. Two approches are shown in the figure; random mutagenesis on the right hand when no accurate information about protein structure is required and three techniques of rational mutagenesis on the left hand, these techniques vary depending on the information available on the protein structure

4.7 Immobilisation

Enzyme immobilisation, a process which can be defined as linking the purified enzyme or the whole cell to a carrier, is required for the effective application of enzymes in industrial manufacturing by achieving a continuous operation process, recycling the enzyme and the cofactor as well as ease of product separation (Tan et al., 2016). Immobilization methods are classified into four groups: I) binding to a solid support, II) carrier-free insolubilisation by cross-linking, and III) entrapment in a polymer network and IV) encapsulation in a membrane. Types of immobilisation are presented in Figure 4-3. Although the difficulties of immobilising P450s due to their need for electron transformation could negatively affect their function as shuttle molecules, many researchers have succeeded in applying an immobilisation technique with P450 BM-3.

The first attempt to immobilise P450 BM-3 on a carrier was applied in 2003 by Maurer (Maurer et al., 2003). The enzyme was immobilised in a sol-gel matrix combined with a FDH (formate dehydrogenase), nicotinamide-adenine-dinucleotide phosphate ($NADP^+$) dependent system. This approach succeeded in achieving continuous production of hydroxyl-compounds from –ionone, naphthalene and octane using a cheap cofactor recycling system with the ability of FDH to convert $NADP^+$ into NADPH which could be used by P450 BM-3 as a source of the required electrons for hydroxylation (Maurer et al., 2003). The next step in P450 BM-3 immobilisation was achieved three years later by Waibel (Waibel et al., 2006). A triple mutant of P450 BM-3 together with *Nippostrongylus brasilliensis* AchE (NbAchE) were immobilised on a sol-gel matrix to produce an active bienzymatic biosensor used for insecticide detection in food (Waibel et al., 2006). In 2007, the continuous production of indigo from indole using whole cell immobilisation in calcium-alginate capsules was

investigated (Yan & Lehe, 2007). The immobilised *E. coli* strain showed very high thermal stability and the indigo yield was maintained at 94% of the original yield after five repeated runs Purified mutant P450 BM-3 M9 (R47F F87A M238K V281G M354S D363H W575C A595T) was entrapped on DEAE (anion matrix) combined with k-carrageenan, catalase (for immediate elimination of the hydrogen peroxide) and zinc dust as an alternative electron source. The immobilised enzyme was able to hydroxylate 3-phenoxytoluene in a plug flow reactor (Zhao et al., 2011).

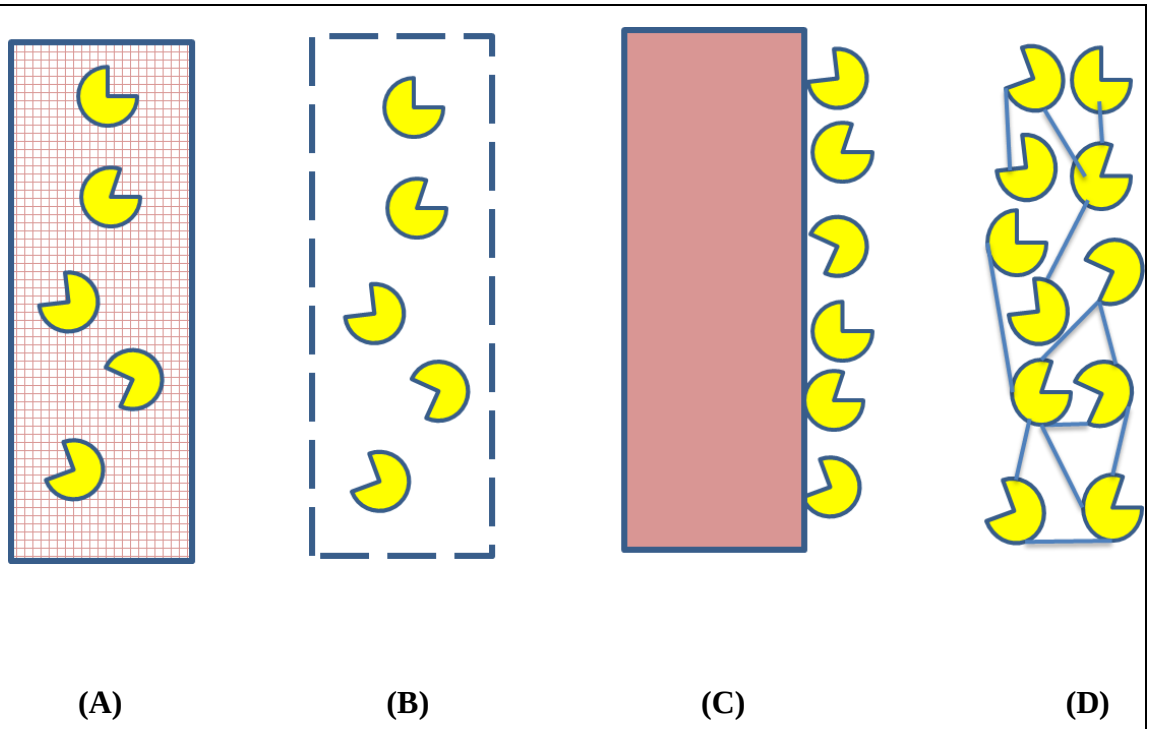


Figure 4-3: Enzyme immobilisation strategies**: (A)** entrapment in a polymer network **(B)** encapsulation in a membrane **(C)** binding to a solid support **(D)** crosslinking (carrier-free insolubilisation)**.**

4.8 Industrial applications of cytochrome P450 BM-3

By using P450 BM-3, a vast range of substrates can be utilised, including alkanes, phenols, steroids and fatty acids, and a broad range of reactions can be catalysed. Indeed, this natural soluble self-sufficient enzyme recorded the highest activity level among all other monooxygenases (Bernhardt, 2006; Urlacher & Eiben, 2006; Urlacher et al., 2004). Applying protein engineering approaches, both the rational and the random designs, made this protein more applicable, overcoming the barriers which hindered its industrial application for a long time. Examples of the applications of BM-3 variants are enormous. Multiple mutation variants of BM-3 were used to produce 4- and 5-hydroxy diclofenac products (nonsteroidal anti-inflammatory agents) used as painkillers from the conversion of diclofenac (Ren et al., 2015). Another application of BM-3 is the manufacture of (—)-perillyl alcohol (a possible anticancer drug) from terpene (—)-limonene using a triple BM-3 variant A264V/A238V/L437F (Seifert et al., 2011). The production of antibacterial and antifungal agents 9 and 10-hydroxy-β-cembrenediol (tobacco diterpene b-cembrenediol) from β-cembrenediol using two BM-3 mutants (F78A/I263L and L75A/V78A/F87G) was also reported (Venkataraman et al., 2014). In 2014, a BM-3 variant from a directed evolution approach was used to hydroxylate 1-tetralone; the hydroxylated products are of great biological importance in the synthesis of numerous biologically active compounds (Roiban et al., 2014). Another example of the application of P450 BM-3 variants is the conversion of omeprazole and esomeprazole (the potent antiulcer drugs) to their human type metabolites (CYP2C19) by a double BM3 mutant (F87V/A82F) (Butler et al., 2014). The BM3 mutant from a saturation mutagenesis of the active site was used to transfer 1-hexene into 1,2-

epoxyhexane, which was recently used to produce aluminium core-shell nanoparticles (the new energetic material in the nanoparticle field) (Kubo et al., 2006).

5 Biocatalysis in non-conventional solvents

5.1 Tolerance *vs* stability

In general, proteins are barely stable molecules and denaturation can easily occur by a number of physical elements, such as heating, agitation and freezing. Furthermore, chemical changes, such as hydrolysis, play a significant role in protein instability. Potency and safety during manufacturing, transportation, and storage processes are affected in both biochemical and pharmaceutical industries. However, a general strategy for stabilising proteins is still lacking (Klibanov, 2001).

According to Matthews (1993), protein stability is defined as the difference between folded and unfolded protein in their free energy, so protein stability depends mainly on the protein structure and could be improved either by stimulating the conversion from unfolded to folded protein or by decreasing the formation of the denatured form, or a combination of both (Matthews, 1993).

The protein in its native state (folded condition) is an ensemble of compact conformations with low energy, which is mainly characterised by low entropy due to the retention of all protein atoms in a well-known geometry. When a protein is denatured (unfolding state) by changing temperature, pH or adding denaturants, the protein conformations ensemble becomes highly heterogeneous and complex. In other words, the entropy will increase due to the loss of the native interactions and as a result, a large set of arrangements will be offered to the residues (Matthews et al., 1987).

Finding the most stable protein is a great challenge in protein science because enhancing protein stability will not provide an outstanding protein unless it is combined with improving the tolerance of this protein. Tolerance in general can be defined as the ability to bear something potentially difficult, while protein tolerance specifically is the ability of the protein to stay catalytically active in the presence of denaturants, such as organic solvents. In summary, a stable folded protein must also be kinetically accessible and the protein activity should be kept within satisfied limits in the presence of denaturants (Matthews, 1993; Pietrzykowski & Treistman, 2008).

5.2 Solvent classification

Green chemistry is the design of chemical processes that reduce or eliminate the use of hazardous substances. Solvents form an interesting part of the green chemistry philosophy because of the volume in which they are used in the chemical and pharmaceutical industries. According to the American Chemical Society Green Chemistry Institute Pharmaceutical Roundtable (ACS GCIPR), 22000 kg of solvents are used annually (Henderson et al., 2011). This quantity constitutes 80–90% of the non-aqueous mass of materials used to make an active pharmaceutical ingredient (API) (Breeden et al., 2012; Henderson et al., 2011; Sheldon et al., 2007).

There are many classifications used to sort solvents such as Pfizer results and tools, ACS GCI Solvent Selection Guide, Sanofi's Solvent Selection Guide and GSK Solvent Selection Guide (Henderson et al., 2011). The one suggested by GSK is the most common classification because it gives reliable information to researchers to help them to switch to a more

sustainable biosynthesis process. Solvents, according to GSK, are classified depending on lifecycle assessment information. They are scored from 1 (red) to 10 (green), 1–3 red, 4–7 orange and 8–10 green to give a relative ranking for every solvent in the guide in each category, where the score depends on the following categories: environmental impact (effects of solvents on the environment), health (acute and chronic effects on human health and the potential for exposure), flammability and explosion (issues affecting storage and handling of solvents), reactivity and stability (factors affecting the stability of the solvent), melting and boiling points (covering high boiling point solvents that have high energy. The GSK solvents classification and the solvents selection guide are shown in (Table 5-1 and Figure 5-1) respectively.

Table 5-1: The ranking of green solvents in each category depending on the GSK guide, 2010 (Henderson et al., 2011).

Solvent	M. P. ˚C	B. P. ˚C	Recycling	Environment	Health	Explosion	Stability	Life cycle
Water	0	100	4	10	10	10	10	10
1-Butanol	-89	118	5	7	5	8	9	5
2-Butanol	-115	100	4	6	8	7	9	6
t-Butyl acetate	-78	95	6	9	8	6	10	8
Isopropyl acetate	-73	89	5	7	7	6	9	7
Propyl acetate	-92	102	5	7	8	6	10	4
Dimethyl carbonate	-1	91	4	8	7	6	10	8

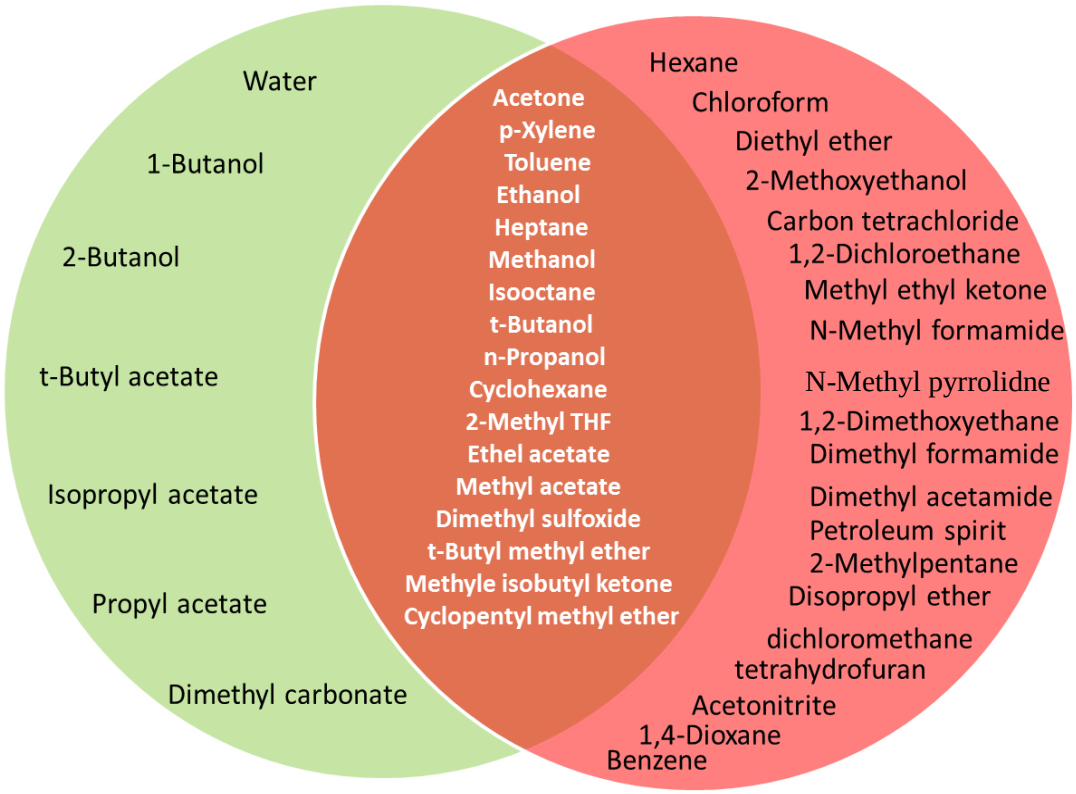


Figure 5-1: Solvents classification depending on GSK in 2010 (Henderson et al., 2011), the green region represents solvents with few issues, the orang region indicates solvents with some issues and the red area is sets solvents with major issues such as: environmental impact, health, flammability and explosion, reactivity and stability and melting and boiling points

5.3 Biocatalysis in organic media

In recent years, biocatalysts have been the subject of intensive investigations by pharmaceutical and industrial researchers, with the interest continuously growing. This is not surprising given the potential advantages of enzymes: (I) the ability to display regio-, chemo- and enantioselectivity; (II) the enhancement in reaction rate up to 10^{12} fold; (III) the ability to catalyse reactions in moderate conditions (pH and temperature), while these reactions require harsh conditions using chemical catalysts (Carrea & Riva, 2000). These superior properties of enzymes are restricted by their orientation to the aqueous reaction media (their natural media). The importance of water in enzymatic catalyst systems is an assumptive matter, since it helps to maintain proteins in their active structures as well as playing an essential role in enzyme dynamics. In spite of these advantages, using aqueous enzymatic media is faced with a variety of limitations. The insolubility of many substrates in water, the unwanted side reactions and the difficulties of products recovery are major issues indicated in biocatalytic systems using water as a solvent (Zaks & Russell, 1988).

In an effort to resolve this problem, many types of solvents were investigated to find a suitable active solvent as a replacement of water. In this chapter, different types of solvents that have been with enzymes and their impact on enzyme stability and activity will be highlighted.

5.3.1 Neat solvent

The fact that enzymatic activity and structural stability decrease in a media containing organic solvents made the later a way of study for years. In fact, the phenomenon of using

organic solvents in enzymatic media dates back to the 1930s of the last century, but this field did not command attention until a report by Zaks and Klibanov in the early 1980s (Halling & Kvittingen, 1999; Zaks & Klibanov, 1984). They indicated that the enzymatic deactivation theory is actually based on studying the enzymes in hydrous solvents not in neat solvents. Accordingly, anhydrous enzymology became a material of study and analysis by many researchers and industrial laboratories (Carrea et al., 1995; Klibanov, 1997). The hydrophobicity of organic solvents is considered valuable to provide an effective enzyme medium. It was measured and studied depending on the partition coefficient value, log P (a ratio of a compound concentration in a mixture of two immiscible phases at equilibrium). It was reported that the activity of biocatalysts in organic solvents is highly dependent on the log P value as follows: I) strongly inactivating when it is used with a solvent log $P < 2$; II) slightly denatured at log P value between 2 and 4; and III) keep its native structure in a log $P > 4$ (Laane et al., 1987). The reason behind this observed trend is that a highly hydrophobic solvent does not strip off the enzymes' water which is essential to their activity, so these solvents should not affect enzyme structures.

It was shown that protein stability and activity for numerous enzymes, such as ribonuclease, porcine pancreatic lipase (PPL) and α-chymotrypsin, can be improved by using neat organic solvents (Volkin et al., 1991; Zaks & Klibanov, 1984; Zaks & Russell, 1988). Chymotrypsin showed no appreciable effect on its crystal structure when it was introduced into neat hexane. In addition to enhancement in activity and structure stability of enzymes achieved by using neat solvents, new enzymatic reactions have come to light. Examples of these reactions, which would not have been possible with aqueous media, are the production of esters from

their constituent acids and alcohols, transesterification, aminolysis and thio-transesterification (Klibanov, 2001; Schmitke et al., 1996; Zaks & Klibanov, 1985).

5.3.2 Water-solvent mixture (co-solvent)

Water-solvent mixture or water-miscible solvents are homogeneous systems in which the enzymes are either suspended or dissolved. At high concentrations, most co-solvents cause a significant drop in enzymatic activity. These solvents form a layer around the enzyme and strip the water, which is an essential molecule for biocatalyst, from the active site of the enzyme, thereby changing the enzyme structure. As a result, most enzymes are unstable and inactive in water-solvent mixtures. However, there are some exceptions, such as polyols and glymes, when enzymes showed stability and activity very similar to their activity and stability in aqueous medium.

The dynamic and catalytic properties of enzymes greatly depend on the amount of water present in the reaction media. By using water-miscible solvents, a relationship between thermodynamic water activity and water concentration was examined using a variety of common polar solvents. This study showed that a solvent's hydrophilic nature has a significant impact on the enzyme activity. Five alcoholic solvents were investigated, methanol, 2-propanol, 1, 2-butanol, ethanol and acetonitrile (Bell et al., 1997) and the effect of water on the enzyme's secondary structure was assessed using three types of conditions, a neat organic solvent, a water-solvent mixture (co-solvent) and pure water. The secondary structures of lysozyme and subtilisin in pure solvents were very close to those of the native

enzymes, while a significant deviation was recorded in the water-solvent mixture (Griebenow & Klibanov, 1996).

5.3.3 Biphasic system

The biphasic solvent system is used to improve catalysts' recyclability, activity, stability and product separation. The prosperity of such schemes depends on the differences between the two liquids in various properties, and usually polarity. One of solvents should be polar to dissolve and retain the catalyst, and the other solvent should be nonpolar to dissolve the products (Lee et al., 2003). The first biphasic used was a mixture of water and carbon dioxide (Bhanage et al., 1999), which showed a good result when it was used with water-soluble catalysts and substrates. However, another combination needs to be applied for water-insoluble reaction media, such as ionic liquids mixed with supercritical CO_2 (Heldebrant & Jessop, 2003).

The activity, selectivity, product and catalyst recovery of two enzymes, Wilkinson's catalyst and a cationic rhodium complex were examined using two different biphasic solvents (polyethylene oxide, heptane and CH_2Cl_2 and polyethylene oxide, heptane and methanol), respectively. The results showed that using both enzymes led to high yields and selectivity (da Rosa et al., 2000). Three years later, an encouraging result was achieved using polyethylene glycol together with supercritical fluid CO_2 as a biphasic solvent in a reaction catalysed by $RhCl(PPh_3)_3$. The biphasic solvent and the catalyst were not replaced for five cycles and the catalyst activity was about 99% combined with an increase in product mass

recovery for five cycles (Heldebrant & Jessop, 2003). The effects of some solvents used with proteins are shown in Table 5-2**.**

Table 5-2: Some solvents used with native enzymes and their effects on enzymes properties

Solvents	Enzymes	Reaction	Effect on enzyme	References
99% ethanol	porcine pancreatic lipase (PPL)	transesterification	high activity at high temperature	(Zaks & Klibanov, 1984)
1-butanol & DMF	ribonuclease, chymotrypsin and lysozyme	hydrolysis of N-benzoylL-tyrosine ethyl ester (BTEE)	high thermal stability	(Volkin et al., 1991)
hexane	uncrosslinked chymotrypsin	crystallisation	stable structure	(Klibanov, 1997)
hexane, dodecane, hexadecane, ethyl ether, isopropyl, ether butyl, ether acetonitrile tetrahydrofuran, dioxane, toluene, pyridine, dimethyl sulphoxide, formamide & carbon tetrachloride	PPL, *Candida cylindracea* lipase, Mucor and Pseudomonas lipoprotein lipase	transesterification reaction between tributyrin and heptanol	enzymatic activity similar to that in water	(Zaks & Klibanov, 1985)
supercritical carbon dioxide (SC-CO_2)	amberlyst & zeolite	carbonatation	Stability improvement	(Vieville et al., 1998)
Supercritical CO_2/toluene mixture	commercial 0.5 wt.% Pd/Al_2O_3	oxidation of benzyl alcohol to benzaldehyde	high reaction rate	(Caravati et al., 2006)
Supercritical CO_2/n-hexane mixture	lipase B	esterification	high conversion rate 99%	(Knez et al., 2012)
ionic liquid 1-butyl-3-methylimidazolium hexafluorophosphate	chymotrypsin	transesterification and hydrolysis	slight enhancement in activity	(Berberich et al., 2003)
[C4mpy][Tf2N]	monellin	----------	improved thermostability	(Baker et al., 2004)
1-Butyl-3-methylimidazolium (bmim) dihydrogen phosphate (dhp), N-butyl-N-methyl pyrrolidinium dihydrogen phosphate (p1,4 dhp),choline dihydrogen phosphate	cytochrome c & P1,4dhp	----------	improved thermostability	(Fujita et al., 2005)
polyethylene oxide, heptane CH_2Cl_2	Wilkinson's catalyst	hydrogenation of hex-1-ene	high selectivity and yield	(da Rosa et al., 2000)
polyethylene oxide, heptane, methanol	cationic rhodium complex	hydrogenation of hex-1-ene	high selectivity, yield, separation and recycling efficiency	(da Rosa et al., 2000)

5.4 Biocatalysis in micellar systems

A mixture of water-in-oil (w/o) or oil-in-water (o/w) is called an emulsion. The stability of emulsions is controlled by the effect of amphiphilic compounds. Amphiphilic molecules have the ability to interact with polar as well as nonpolar molecules at the same time because of the differences in polarity between the head and the tail of these components; hydrophilic head and hydrophobic tail. The presence of these amphiphiles encourages the formation of monophasic (emulsion) inside the biphasic (microemulsion) by decreasing the interfacial water-oil tension. The ratio of water to oil is responsible for the variation in the microemulsion microstructure from tiny droplets of water-in-oil phase to the reverse microemulsions, tiny droplets of oil-in-water phase. The structure and stability of emulsions differ from those for microemulsions; emulsions are less stable and they separated into two layers once the agitation stops. In addition, the size of emulsion droplets is in micrometres range, while they are in nanometres in microemulsions. Another important difference between these two systems is the formation of spherical aggregates, these aggregates are in colloidal size and they called micelles. The micelle system is homogeneous monophasic system. Different micelles systems have different structures based on their composition, temperature and presence of amphiphilic compounds (Braga & Belo, 2015).

Microemulsions have been used successfully for enzymatic catalysis in many reactions such as hydrolysis, transesterifications, peptide and sugar acetals, and ester syntheses (Krishna et al., 2002). Microemulsions were used to overcome the heterogeneous nature of the reaction mixture when lipase was applied to hydrolyse long-chain aliphatic esters to glycerol and free fatty acids (Carvalho & Cabral, 2000). When reverse micellar system (sodium lauryl sulphate/benzene) was applied, hydrogen production by *Rhodopseudomonas sphaeroides*

increased by 25-fold. In addition, a 12-fold increase in alcohol dehydrogenases (ADH) catalysed reduction rate was recorded when Marlipal O 13/60/cyclohexane was used to reduce 2-heptanone to S-2-heptanol (Braga & Belo, 2015).

5.5 Biocatalysis in ionic liquids

Salts which have the ability to stay in a liquid state and do not crystallise at room temperature are called ionic liquids. For the names and physical properties of some ionic liquids, see Table 5-3. Imidazolium is a widely used ionic liquid and the first one discovered was $EtNH_3NO_3$. Over the past twenty years, ionic liquids have become increasingly important as green solvents for various reactions because of their remarkable results in biochemical reactions such as redox, hydrolysis and esterification. They provide an ideal solvent for engineering media for biocatalytic reactions because of the unique advantages for biocatalysis, such as their non-volatile nature which maintains a stable system pressure, ability to dissolve several compounds, as well as forming multiphase systems with many solvents (Seongsoon & Romas, 2003a).

Neat ionic liquids represent a new class of solvents, which are polar and non-aqueous at the same time depending on the cation and anion properties. Unlike organic solvents, they do not inactivate enzyme activity (Seongsoon & Romas, 2003b). Much attention has been given to improve protein stability, activity and solubility in neat ionic liquids (Weingärtner et al., 2012).

The first kinetics study to compare pure ionic liquids, aqueous ionic liquids and pure organic solvents was conducted in 2000. In this study, they highly recommended using anhydrous

ionic liquids due to the high product yield (80%) (Lau et al., 2000). Subsequently, numerous studies reported the excellence of pure ionic liquids as alternative to organic solvents (Eckstein et al., 2002; Lozano et al., 2001; Summers & Flowers, 2000).

Table 5-3: Names and physical properties of some common ionic liquids used with enzymes

IL	IL name	M. P. °C	Viscosity cP at 25 °C	Density g/ml at 25 °C	Enzyme	References
[C_2mim][BF_4]	1-Ethyl-3-midazolium tetrafluoroborate	-	43	-	Lysozyme, Catalase, Myoglobin, Trypsin Glucose isomerase, Xylanase	(Huddleston et al., 2001; Judge et al., 2009; Noritomi et al., 2011)
[C_2mim][Br]	1-Ethyl-3-methylmidazolium bromide	58–60	-	-	Trypsin	(Ge et al., 2010; Huddleston et al., 2001)
[C_4mim][BF_4]	1-Butyl-3-methylmidazolium tetrafluoroborate	−81	233	1.12	Horseradish Peroxidase, Glucose Oxidase (HRP), Esterase	(Huddleston et al., 2001; Patel et al., 2014; Persson & Bornscheuer, 2003)
[C_4mim][PF_6]	1-Butyl-3-methylmidazolium hexafluorophosphate	10	450	1.36	*Candida rugose*, Esterase, Thermolysin, Lipase Chymotrypsin	(Erbeldinger et al., 2000; Huddleston et al., 2001; Lau et al., 2004; Meng et al., 2011; Pavlidis et al., 2009; Persson & Bornscheuer, 2003)
[C_8mim][PF_6]	1-Octyl-3-methylmidazolium hexafluorophosphate	-	682	1.22	*Candida rugosa* enzyme	(Huddleston et al., 2001; Lau et al., 2004)

5.6 Biocatalysis in supercritical fluids

Supercritical fluids (SCFs) are solvents used at a combination of high pressure and temperature, usually above their critical value. From the time they were discovered in 1822

to date, their applications have witnessed considerable developments. The superior specifications by mixing gas properties (diffusivity and viscosity) with liquid properties (density and solvating ability) has opened new prospects for the use of SCFs as solvents for many applications. Furthermore, they are considered safe, cost efficient, nontoxic, easy to recover and environmentally friendly, which makes them a good alternative for traditional solvents (Knez et al., 2014). The stability of heterogeneous catalysts in a glycerol carbonatation process was investigated using supercritical solvents. The catalysts appeared to be more stable in a media containing supercritical carbon dioxide (SC-CO_2) (Vieville et al., 1998). A comparative study between neat CO_2, neat toluene and CO_2/toluene mixture was published by Caravati et al. in 2005 and it elucidated that the highest reaction rate can be achieved by adding a small amount of toluene to the SC-CO_2 (Caravati et al., 2006). The same supercritical solvent (SC-CO_2) was tested alone and as a mixture (SC-CO_2/n-hexane) in the esterification process catalysed by lipase B, showing that the SC-CO_2/n-hexane mixture was the best, with a 99% conversion rate and 11.2 (w/w) yield (Knez et al., 2012).

5.7 Stabilisation of enzymes in non-conventional solvents

Despite enzyme inactivation by organic solvents, the numerous advantages of using these solvents motivated industry and academia to find effective strategies for enzyme stabilisation in non-conventional solvents. These strategies can be classified into three categories: I) isolation of stable biocatalysts; II) enzyme structure modification, and III) solvent environment modification (Stepankova et al., 2013).

Microorganisms that have the ability to live and survive under harsh conditions, such as high and low temperature, extreme pH and organic solvents medium, offer a very effective source

of extremozymes which are highly stable enzymes. This strategy succeeded in improving the tolerance of some enzymes against organic solvents by enhancing the hydration characteristics of these enzymes, which are responsible for enzyme inactivation due to the loss of crucial water molecules (Karan et al., 2012). The tolerance of protease towards many co-solvents was investigated. This enzyme was isolated from *P. aeruginosa* PST-01 grown in soil saturated with a high concentration of different organic solvents. This enzyme showed stability in a media containing co-solvents rather than in aqueous media (Ogino et al., 1999). In addition, α-amylase and alkaline phosphatase also showed very good tolerance against organic solvents when they isolated from *Haloarcula sp.* strain S-1 and *Streptomyces clavuligerus* strain Mit-1 respectively (Fukushima et al., 2005; Thumar & Singh, 2009).

Modification of biocatalysts was also applied to improve enzyme activity using different strategies including enzyme immobilisation, ionic liquid coating, chemical modification and genetic engineering. Stabilising enzymes in solvents by using immobilisation is one of the most common strategies. All immobilisation types were reported as effective methods to enhance enzyme stability. *Candida antarctica* lipase B showed outstanding activity in neat solvents when it is adsorbed on Lewatit ion exchange resin (Hanefeld et al., 2009). In addition, when α-chymotrypsin was bound covalently to a nanoporous silica glass, it showed impressive improvement in activity and stability in the presence of organic solvents (Wang et al., 2001). Moreover, not only the enzyme, but also its cofactor NADPH were entrapped in the polymer network (polyvinyl alcohol gel beads), protected from the negative effect of pure hexane in this experiment, the entrapped dehydrogenase from *Lactobacillus kefir* successfully transformed some hydrophobic ketones to the "corresponding enantiomerically pure (R)-alcohols" (De Temiño et al., 2005). In addition to their importance as green solvents,

ionic liquids play a crucial role in enhancing enzyme tolerance towards organic solvents by acting as enzyme-coating agents. These coating agents offer a stable microenvironment for the reaction by binding to the enzymes. It was reported that when lipase B from *C. Antarctica* was coated with [BTMA][Tf2N] or [TOMA][Tf2N], it remained active at high temperatures, 95°C, in hexane (Lozano et al., 2007).

Numerous techniques and strategies have been applied to identify a suitable solution for loss of enzyme activity in a media containing solvents, but one which has been successfully used to overcome stability and activity problems, properly helping to maintain a satisfactory physical and chemical stability of enzymes in non-conventional solvents is protein engineering.

Genetic mutagenesis, both rational and directed evolution approaches were used to stabilise enzymes towards organic solvents. By applying the B-FIT method, protein rigidity was increased as a result of replacing the high B-factors residues with the lowest ones and a considerable enhancement was reported in proteins' tolerance in a high solvent media (Reetz et al., 2010). Reetz and co-workers highly recommended using the B-FIT method at sites showing high B-factors to increase thermostability and tolerance against solvents. By using this approach, lipase from *Bacillus subtilis (BSL)* showed high tolerance towards three different solvents, DMSO, ACN, and DMF (Reetz et al., 2010).

Surface loop engineering has been successfully used to overcome stability problems towards solvents and maintain protein activity through insertion, deletion and substitution of residues from the enzyme surfaces (Reich, 2014). The stability and activity of a lipase from *Bacillus subtilis* was examined in the presence of 60% DMSO by applying site saturation mutagenesis

for 91 amino acids of the lipase loops. The protein showed improvement in DMSO tolerance and a massive increase in protein activity (8-fold) (Yedavalli & Rao, 2013). A year later, two loop mutants of reductase NCR from *Zymomonas mobilis* were generated using a rational approach dependent on surface structure information, one of them showed a considerable increase in tolerance towards organic solvents (Reich et al., 2014).

In 2013, protein engineering of residues located in the access tunnels was used to enhance enzyme stability against solvents. A unique improvement in enzyme stability against 42% (v/v) DMSO for four variants generated using access tunnel mutagenesis was reported by Koudelakova and co-workers (Koudelakova et al., 2013).

Directed evolution applications contributed to the expansion of the knowledge regarding protein properties and a wide range of molecule synthesis from building blocks to drug analogues as well as the massive enhancement in activity, selectivity and stability not previously encountered in nature (Denard et al., 2015; Johannes & Zhao, 2006). An impressive example about the role of this approach in activity and stability improvement is a 500-fold enhancement in the protease subtillisin E specific activity after multiple generations of random mutagenesis were applied in the presence of 60% DMF. In addition, a 20-fold increase in p-nitrobenzyl (pNB) activity was recorded by using just four generations of random mutagenesis in approximately 15–20% DMF (Moore & Arnold, 1996). Wong et al. (2004) used directed evolution to enhance P450 BM-3 tolerance towards co-solvents. The engineered P450 BM-3 was significantly more resistant in the media containing DMSO and THF, showing a high tolerance in the presence of acetone, acetonitrile, dimethylformamide and ethanol (Wong et al., 2004). Likewise, the activity as well as tolerance of laccase in a

high concentration of acetonitrite and ethanol (20–60%) was increased significantly by the fifth generation (Zumárraga et al., 2007).

Before genetic engineering, chemical enzyme modification approach was widely applied to enhance enzyme stability and it was used successfully to improve enzyme resistance towards organic solvents. However, chemical modification has been overshadowed in recent years by protein engineering, it remained a useful technique for enzyme stabilization taking into consideration the advantages can be gained by using this technique such as the simplicity of the procedure and no structure details are required, in Figure 5-2 the advantages and disadvantages of both chemical and biological proteins engineering are summarised.

PEG was an investigated polymer for chemical modification since 1985, being used to shield and surface modify polypeptides (Veronese et al., 1985). This method was applied for the first time to improve enzyme resistance against organic solvents in 1986 by Inada and co-workers (Inada et al., 1986). When lysozyme was conjugated with methoxy polyethylene glycol, protein tolerance was enhanced against dichloromethane (Diwan & Park, 2001). The improvement in protein stability by using this approach may be very well due to a decrease in residue mobility resulting by the polymer (Nischan & Hackenberger, 2014).

The easiest method that has been used extensively for enzyme stabilisation is reaction environment modification. This technique can be applied by adding different additives to the enzyme aqueous solution or directly before enzyme lyophilisation. These agents could be sugars, inorganic salts or polyols. It was found that these additives increase protein rigidity by increasing the hydrophobicity around nonpolar amino acid residues (Stepankova et al., 2013). Three additives (glycerol, trehalose and sorbitol) were examined for thermolysin

stabilisation in the presence of different co-solvents, with trehalose improving the stability of thermolysin effectively in the presence of dimethylformamide, while both glycerol and sorbitol significantly stabilised the same enzyme in n-propanol and isopropanol. Also, a 60-fold enhancement in the fungal protease activity in neat organic solvents was reported when the enzyme was stabilised using sorbitol and PEG (Debulis & Klibanov, 1993; Pazhang et al., 2006).

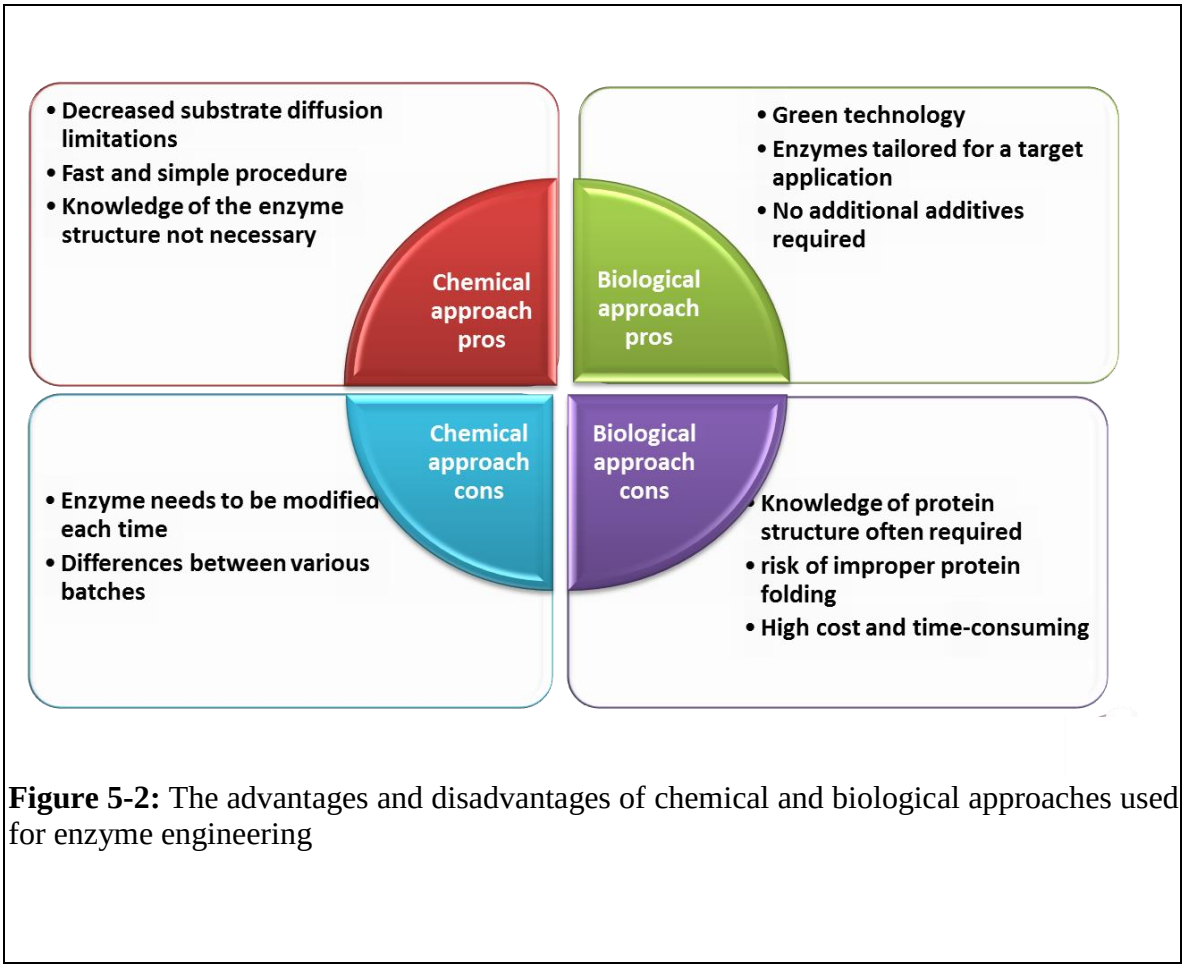


Figure 5-2: The advantages and disadvantages of chemical and biological approaches used for enzyme engineering

5.8 Tolerance of cytochrome P450 BM-3 to organic co-solvents

Using P450 BM-3 in industry as a biocatalyst is hindered by some limitations, one of which is enzyme inactivation in a medium containing organic solvents. Increased P450 BM-3

applicability requires a better understanding of the mechanism responsible for a loss in activity in the presence of co-solvents and enhancing BM-3's tolerance in such mediums. Many experimental studies were focused on investigating this phenomenon (Bailey, 1995; Kuper et al., 2007; Roccatano et al., 2005; Wong et al., 2004). The behaviour of cytochrome P450 BM-3 in water as well as changes that have been noticed in the wildtype and its mutants structures (mainly F87A variant was selected for this purpose) by adding different cosolvents and in particular dimethylsulphoxide (DMSO) were extensively analysed (Roccatano et al., 2005, 2006).

A 15 ns long molecular dynamic study (MD) simulations in pure water and in 14% DMSO/water mixture was reported by Roccata(Roccatano et al., 2005)no and co-worker (Roccatano et al., 2005). In general, no major changes were noticed in protein structure in both cases. However, the significant variations were found in helices E, F, and G helices as well as in EF and FG loops when DMSO was used instead of water. A higher mobility was recorded at the mouth of the active site around Phe42, Arg47, and Tyr51 residues in both simulations with more overall average deviations from the original crystal structure when DMSO was used (Roccatano et al., 2005). At the same study it was also observed that DMSO molecules accumulated on the protein surface increasing the solvent concentration to about 30-40 % (v/v) and allowing to only one molecule of DMSO to form a hydrogen bond with the side chain of Arg47 as can be seen from the density distribution of the DMSO molecules around the heme domain at the last 5 ns of the simulation (Figure 5-3). It was found previously that both Arg47 and Tyr51 play an essential role in facilitating the access of fatty acids substrates to the substrate access channel (SAC) since the mutation of Arg47 and Tyr51 hamper the binding of fatty acids (Denard et al., 2015; Ost et al., 2000).

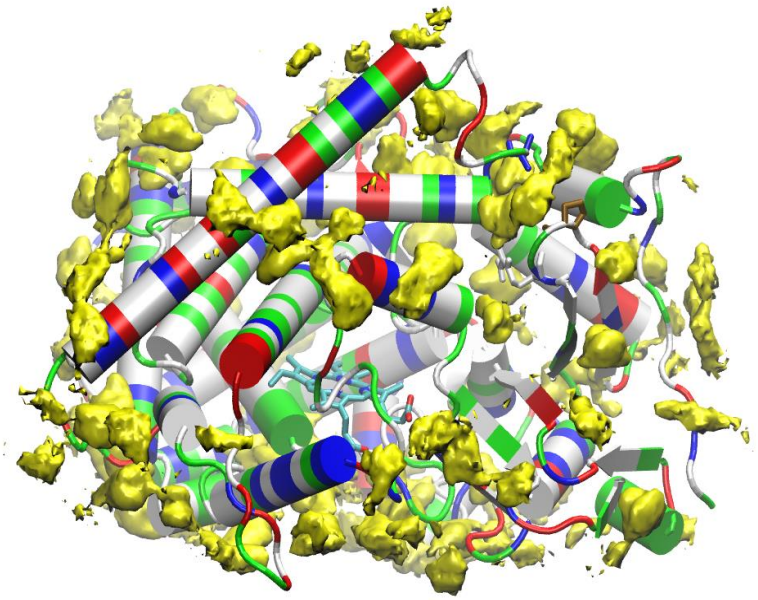


Figure 5-3: Density distribution of the DMSO molecules around the heme domain. The last 5 ns of the trajectory from (Roccatano et al., 2005)was performed to calculate the solvent distribution by using a cubic grid with 0.1 nm grid spacing in the GROMACS package. The VMD program was used to visualise the averaged volumetric density data (Humphrey et al., 1996). The protein is shown in cartoon with the amino acids coloured according to their chemical nature (red: acid, blue: basic and green: polar, gray: unpolar). Residues Pro45, Ala191 and Arg47 are shown as sticks in the figure. The DMSO high density molecules are represented in yellow. This image was acquired with permission from (Roccatano, 2015)

Generally, mutant Phe87Ala, showed lower tolerance than the WT when the catalytic activity was observed in all the considered cosolvents (Wong et al., 2004). The activity of this mutant dramatically reduced as the concentration of DMSO increased. At about 14% (v/v) DMSO, the WT retained its full activity while the activity of F87A variant decreased to only about 30% from its original activity (Wong et al., 2004). The low tolerance of this mutant against DMSO highlighted the vital role of Phe87 residues in protection the active site from the inhibitory effect of DMSO and other solvents because this residue lays in the active site cavity and likely prevents solvents from access to the heme cofactor region (Lau et al., 2000). The inhibitory effect of the cosolvent on the enzymatic activity is usually caused by the elimination of the functional water around the active site of the enzyme causing protein denaturation (Lau et al., 2000; Sørensen & Mortensen, 2005).

The changes in opening of the SAC mouth in the water and in the DMSO were also studied and analysed (Roccatano et al., 2005). The entrance opening expanding is usually checked by the distance between the Cα of the residues Pro45 and Ala191 (Humphrey et al., 1996). A study to measure the SAC opining distance of P450 MB-3 substrate-free structure (PDB ID: 1BU7) reported that this distance showed a bimodal distribution with two peaks at about 0.8 and 1.2 nm, respectively (Roccatano et al., 2005). The crystallographic unit cell of this structure contains two chains; A and B and the entrance distance were about 0.87 and 1.09 nm respectively and it was reduced further to 0.7 for chain A when it was measured in water. This closing in the SAC explain the hardness in releasing the substrates from the active site. MD Simulation on the Phe87Ala mutant in the same conditions reported that this mutant has similar distances as the crystallographic structures for both peaks of the bimodal distributions

(Roccatano et al., 2006). MD simulations of the free-substrate wildtype structure (PDB ID:1BU7) in 14% (v/v) DMSO/water mixtures concluded the tendency of the protein to retain the open SAC in the presence of solvents. On contrary, adding DMSO did not affect the shape of the bimodal distribution of the Phe87Ala mutant. However, the presence of DMSO shift both peaks to 1.6 and 1.8 nm causing entrance expanding. This widening increases the flexibility of the EF helix subdomain the presence of DMSO in comparison to the WT and gives more chance to DMSO to access the active site and reducing the catalytic activity of the enzyme (Roccatano et al., 2006).

Another way to identify the changed in protein structure and dynamic behaviour due to the solvents effect is the measure of the root mean square deviations (RMSD) and fluctuations (RMSF) (Sørensen & Mortensen, 2005). RMSD and RMSF were measured in the presence of DMSO in respect the starting crystallographic structure of the WT. A high deviations and fluctuations were highlighted in the region of E, F-G helices (Sørensen & Mortensen, 2005).

To have better understanding of the effect of DMSO on P450 BM-3 activity for both the WT and Phe87Ala mutant, both the WT and Phe87Ala heme domain crystal structures permeated with low and high concentration of DMSO were determined (Bailey, 1995; Kuper et al., 2007). The crystals structures of P450 BM-3 were solved after soaking them in 14 % (v/v) DMSO (PDB ID: 2J4S) and 28 % (v/v) DMSO/water mixture (PDB ID: 2J1M). Both crystal structures do not show large conformation change in comparison to the crystal structure in water. In 2J4S structure, the iron coordinating water molecules displaced to a distance of 0.377 nm with a nonplanar distortion in the heme region. However, 2J1M structure showed a replacement of water molecule by one of DMSO and it was confirmed that this molecule coordinates the iron with its sulfur atom. on contrast, high concentration of DMSO caused

significant changes in the active site such as increasing the I-helix kinking (Kuper et al., 2007).

On the other hand, the Phe87Ala mutant crystal structures of the hemein the presence of DMSO concentrations 14% (v/v) (PDB ID: 2X7Y and 28% (v/v) (PDB ID: 2X80) were also solved and compared with the WT structures in the same conditions (Bailey, 1995). The structure (2X7Y) showed a similar structure as in water solution while a higher RMSD was observed in the case of 2X80. In both structures, a DMSO molecule was found in the active site (they were oriented towards the heme iron but not coordinating it) with Fe-ODMSO distances of 0.325 and 0.303 nm for the 2X7Y and 2X80, respectively.

To conclude, in the WT and when the concentration of DMSO is low, the residue Phe87 is able to play its role and preventing DMSO from replacing water molecules from the coordination site of the iron atom but when the concentration of DMSO increase the chance of its molecules to access the SAC increase too. However, replacing the Phe residue by the Ala residue in the Phe87Ala variant facilitate the access of DMSO into the active site even in the low solvent's concentrations.

5.8.1 Mutants of P450 BM-3 F87A and their tolerance to organic co-solvents

The effect of DMSO and tetrahydrofuran (THF) on the activity of WT P450 BM-3 and its mutants was highlighted experimentally by Wong and co-workers (Wong et al., 2004). Random mutations were introduced into the P450 BM-3 F87A parents. In the first round, the tolerance of 6,520 clones was checked against 22.5% DMSO and 2.8% THF. From this round two mutants with multiple solvent resistances were chosen; F87AB5 (T(ACG)235A(GCG);

S(AGT) 1024R(AGA)) and F87APEC3 (R(CGC)471C(TGC)). F87AB5 showed 3.7 and 5.3-fold enhancement at 10% (v/v) DMSO and 2% (v/v) THF respectively so this mutant was selected as starting point to the next step which is saturation mutagenesis. This step produced a triple mutant F87ASB3 which has significantly higher total activity compared to F87AB5 in the presence of organic cosolvents. A random mutagenesis was applied again on F87ASB3 this time generating mutant F87A5F5 (E(GAA)494K(AAA) and R(AGA)1024E(GAG)). The specific activity of F87A5F5 is increased 5.5-fold at 10% (v/v) DMSO and 10-fold at 2% (v/v) THF in comparison to F87A.

As mentioned above cytochrome P450 BM-3 WT is significantly more tolerance to organic cosolvents than the F87A variant. To investigate the effect of F87A mutants on the WT resistance, position A87 was mutated back to phenylalanine in the evolved sequences (Wong et al., 2004). This back mutation yielded a very tolerant mutant (W5F5 from parent F87A5F5) against cosolvents especially DMSO. This final mutant showed 6-fold enhancement against 25% DMSO and 3.4-fold improvement in 2% THF as well as it is observed that W5F5 is also resistance to other cosolvents like acetone, acetonitrile, DMF, and ethanol (Wong et al., 2004). Due to its outstanding resistance against many organic solvents this back mutant (W5F5) was chosen in this project to investigate its tolerance against green solvents and compare it to the WT.

6 Materials and Methods

6.1 Chemical and biological materials

All chemicals are of analytical grade or higher quality and purchased from VWR Prolabo (Belgium), AppliChem (Germany), Sigma (UK), Merck KGaA (Germany) and Formedium LTD (UK). All biological materials used in this project are high quality and purchased from Sigma (UK), Complete Mini Roche (UK) and New England BioLab (UK).

6.2 Kits

QIAprep Spin Miniprep kit (Qiagen, Germany) was used for pETM-11 W5F5 and pETM-11 WT plasmid extraction.

6.3 Media

Auto-induction media (Terrific Broth base): To prepare 1 L of media, the following chemicals were used: 12 g tryptone, 24 g yeast extract, 3.3 g $(NH_4)_2SO_4$, 6.8 g KH_2PO_4, 7.1 g Na_2HPO_4, 0.5 g glucose, 2.2 g α-lactose monohydrate, and 0.31 g $MgSO_4.7H_20$.

Auto-induction media (2×TY): 16 g tryptone, 10 g yeast extract, 3.3 g $(NH_4)_2SO4$, 6.8 g KH2PO4, 7.1 g Na_2HPO_4, 0.5 g glucose, 2.1 g α-lactose monohydrate and 0.31 g $MgSO_4.7H_2O$.

Auto-induction media (Super Broth base): $(NH_4)_2SO_4$, 6.8 g KH_2PO_4, 7.1 g Na_2HPO_4, 0.5 g glucose, 2.2 g α-lactose monohydrate and 0.31 g $MgSO_4.7H_20$.

2×TY media: 16 g tryptone, 10 g yeast extract and 5 g NaCl were mixed with deionised water to prepare 1 L of media.

TYE agar media: 10g tryptone, 5g yeast extract, 8 g NaCl and 15 g agar. Kanamycin was added when temperature reached 50–60ºC.

6.4 Strains and vectors

Dr Tuck Seng Wong graciously provided me with pETM-11 vectors encoding the wild type P450 BM-3 WT and the modified P450 BM-3 W5F5 genes. The constructs had a kanamycin resistance gene to allow selection, polyhistidine tag and TEV-site to facilitate the purification process and lac repressor induction control. The *E. coli* strain DH5α was used for all routine cloning experiments, whereas the *E. coli* strains BL21 (DE3) and C41 (DE3) were used for recombinant protein expression.

6.5 P450 BM-3 expression

A self-replicating circular DNA (pETM11) bearing the sequences coding for the P450 BM-3 of interest was isolated using a QIAprep Spin Miniprep kit (Qiagen, Hilden, Germany) based on the alkaline lysis of bacterial cells. The plasmid DNA was quantified and evaluated using a spectrophotometer from Expedeom, UK. Both pETM11-P450 BM-3 WT and

pETM11-P450 BM-3 W5F5 plasmids were introduced into *E. coli* competent cells by transformation. In this process, 50 mM $CaCl_2$ was used to make *E. coli* cells porous, so they could take up foreign DNA. The plasmid (1 mM) was introduced to *E. coli* BL21 (DE3) or C41 (DE3) cells/$CaCl_2$ mixture and incubated on ice for 10 min before heat shock at 42°C for 1 min. Then, the mixture was immediately cooled on ice for 2 min. The cells were regenerated with 2×TY media by incubation for 1 hr at 37°C and 200 rpm. The transferred cells were then plated on TYE agar plates with 50 µM kanamycin. For validation, the plates with plasmid DNA were grown along with a plate of *E. coli* cells with no plasmid DNA.

After transformation, protein expression was induced using auto-induction media containing a low glucose concentration and a high lactose concentration. By using this technique, both the wild type and the mutant of P450 BM-3 were expressed. The expression of P450 BM-3 W5F5 was optimised using a variety of conditions such as different growth medium and various DE3 lysogen *E. coli* strains. Growth media was supplemented with 50 µg/ml kanamycin, 1 mM trace elements and 1 mM δ-aminolevulinic acid before inoculation with broth culture. Then, it was incubated for 24 hr at 30°C and 200 rpm. Expression using induction agent isopropyl β-D-1-thiogalactopyranoside (IPTG) was also investigated; the media was left to grow at 37°C until the OD 600 reached ~0.6, then expression was induced using 1 mM IPTG. The solution was incubated overnight (typically 12–16 hr) at 30°C and 200 rpm, before the cells were harvested by centrifugation at 4°C and 6000 rpm for 5 min. The pellets were stored at -20°C for later use.

6.6 Protein purification using affinity chromatography

The ÄKTA pure system from EG Healthcare, Germany was used for all protein purification during this project. *E. coli* cell pellets with P450 BM-3 variants were resuspended in equilibration buffer (buffer A1: 50 mM NaH_2PO_4, 200 mM NaCl, pH 8). The solution was supplemented with one tablet of the proteinase inhibitor and 1 mM of DNase and RNase. Bacterial lysis was performed via ultrasonication (Sonics Vibra Cell VCX 130, Sonics, USA) on ice for a total of 5 min (10 sec bursts and 20 sec cooling time) and 70% AMPL. Cell debris was eliminated by centrifugation (60 min, 6000 rpm, 4°C). The supernatant was filtered through a 0.45 µm filter to produce a clarified cell lysate. Then, protein purification was performed by affinity chromatography. Firstly, the 5 ml HisTrap prepacked column from GE Healthcare (Germany) was pre-equilibrated using buffer A1, then the filtrated sample was loaded and the desired protein was eluted using a gradient of buffer B1 (buffer A1 plus 250 mM imidazole, pH 8) to buffer A1. Finally, the fractions with the desired protein were pooled for the next step of purification. For optimisation, different elution times (10 min, 20 min and 40 min) at a constant flow rate of 2 ml/min were applied.

6.7 Ion exchange chromatography

The fractions with the desired protein from the affinity step were reloaded into the ion exchange step (IEX) to separate targeted protein as well as to concentrate the fraction volume to a volume suitable for the next step (gel filtration). Two buffers were used in this step, equilibration buffer (buffer A2: 100 mM Tris-HCl, pH 8) and elution buffer B2 (buffer A2 plus 2 M NaCl, pH 8). Fractions with the desired protein from the affinity step were diluted

8 times using buffer A2 to reduce the NaCl concentration from 300 mM to 37.5 mM and then loaded onto the ion exchange column pre-equilibrated with buffer A2. Finally, the desired protein was eluted using step elution (100% of buffer B2). Two different types of ion exchange columns were used for optimal purification, DEAE and super Q. The same steps were performed for each column, column washing, sample loading and sample elution. The fractions with desired protein were pooled for the next purification step.

6.8 Gel filtration chromatography (size exclusion chromatography)

The fractions with the desired concentrated protein from ion exchange chromatography were transferred to a 320 ml Hiload 26/600 Superdex 200 pg gel filtration column. The column was pre-equilibrated with 1.5 column volume (CV) of buffer C (50 mM Tris-HCl, 1mM EDTA, 10% (v/v) glycerol, pH 7.2). Then, the concentrated sample (up to 5 ml) was loaded onto the column and circulated buffer washed out proteins sequentially from a large to small volume. The flow rate used during this step was no more than 1.5 ml/min. Samples are eluted isocratically from a SEC column, using a single buffer system.

6.9 SDS-PAGE electrophoresis

Purified proteins were analysed using SDS-PAGE electrophoresis. A 10% acrylamide-SDS gel was used for the analysis throughout the experiments due to the high molecular weights of the proteins expressed. The gel was prepared in two layers, the lower part (resolving gel) was prepared first and left to solidify for one hour before adding the upper layer (stacking gel). After 30 min, 5 µl of PageRulerTM unstained broad range protein ladder (Thermo Fisher) and 10 µl of samples were loaded. Electrophoresis running conditions were 200 V,

400 A for 60 min. Commassie Brilliant Blue was added to stain the gel and a gel documentation system was used to capture the gel picture.

6.10 Large-scale expression and purification

After optimising the expression and purification of P450 BM-3 W5F5, the optimum strain, media and conditions were applied for the large-scale expression and purification of the wild type and mutant proteins (400 ml). For expression, the BL21 (DE3) strain was used to inoculate the SB AIM at 30°C for 24 hr. For protein purification, three chromatography steps, affinity, ion exchange and gel filtration, were used in their optimum conditions to produce a high yield of the desired purified P450s.

6.11 Spectroscopic measurement

In order to define the spectral features of P450 BM-3 WT and P450 BM-3 W5F5, the spectra of purified P450s were measured between 250–750 nm using a UV-3100PC spectrophotometer (VWR, USA). Both proteins were diluted with buffer (50 mM Tris-HCl, 1 mM EDTA and 10% v/v glycerol, pH 7.2) using a 3:1 dilution rate. The spectra obtained were used to evaluate the purity of the wild type and the mutant variant by calculating the Reinheitzahl value (RZ) (RZ is a ratio of P450 heme absorbance at 418 nm to the absorbance of all proteins in the solution at 280 ratio nm). The purest fractions by this criterion were stored for further use.

6.12 Enzyme concentration quantification

The protein concentration was calculated from the spectra using the Beer-Lambert law according to the following equation:

$$A = \varepsilon LC \quad \text{…….……..} \quad \textbf{(Equation 6.1)}$$

Where, A is the absorbance at 280 nm ($A = -\log (I_t/I_o) = -\log (T)$), where I_t is the radiant flux transmitted by that material, I_o is the radiant flux received by that material and *T* is the transmittance of that material), ε is the extinction coefficients in M^{-1} cm^{-1} at 280 nm measured in water, L is the length of path the light passes through in cm and C is the protein concentration in M.

6.13 Enzymatic assay via NADPH consumption

The activity of purified P450 BM-3 WT and P450 BM-3 W5F5 enzymes was evaluated by a NADPH consumption assay. For evaluation of the activity of the P450s, 1 ml volume of the reaction mixture contained 100 mM potassium phosphate buffer (pH 7), 800 µM lauric acid and 0.5% (v/v) P450 BM-3 W5F5 or P450 BM-3 WT. The reaction was induced by adding 250 µM NADPH and the fatty acid oxidation was measured by monitoring the absorption change at 340 nm. This assay was applied to the solvent-free medium as well as reaction media containing different concentrations of 1-butanol, 2-butanol, and dimethyl carbonate to assess the proteins' tolerance towards different green solvents by calculating the relative

activity ratio (relative activity is the ratio of specific activity in the presence of organic co-solvent to that in the absence of organic co-solvent). The P450-BM-3 wild type relative activity was compared to the mutant variants. Trace amounts of methanol used to dissolve the substrate were not taken into account.

To ensure that the activity which has been shown by P450s variants is not due to the effect of these solvents on the protein structure, the spectra of the free-solvent protein and the spectra in the presence of different solvents was compared within the wavelengths 250–750 nm to identify any change such as red or blue shifts in the Soret band as a result of adding these solvents.

7 Tolerance of cytochrome P450 BM-3 to non-conventional solvents

7.1 Optimisation of protein expression and purification

7.1.1 Expression and purification of P450 BM-3 mutant using *E. coli* C41(DE3)

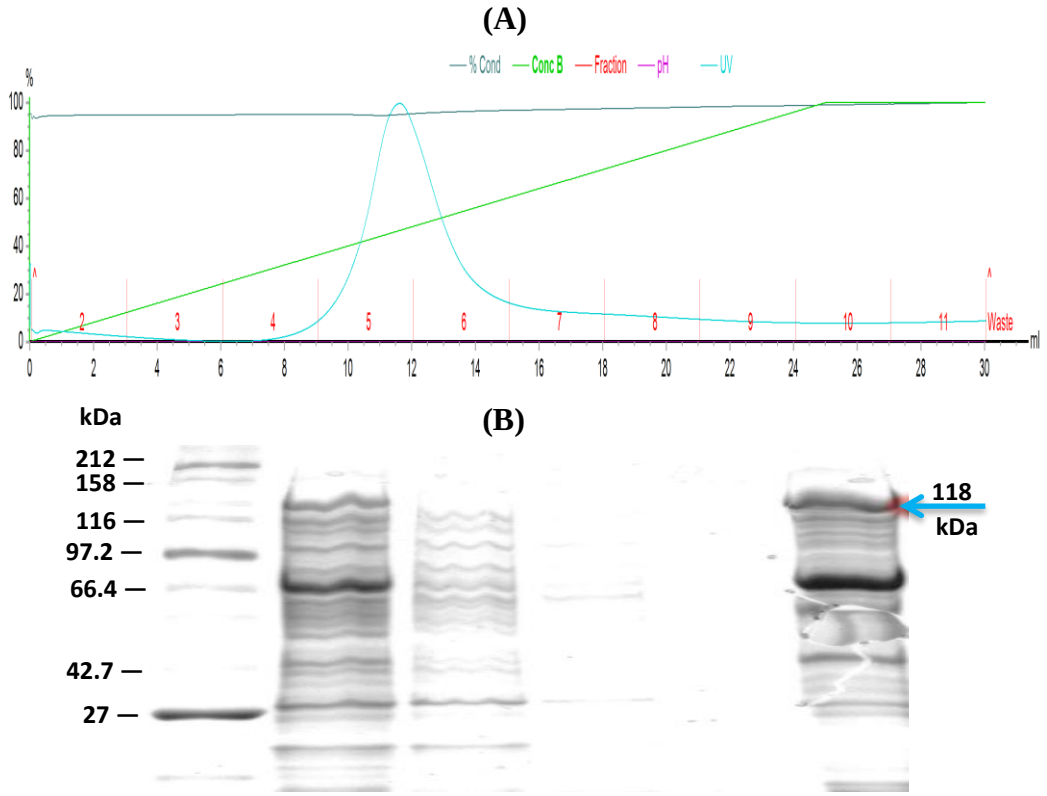
P450 BM-3 W5F5 was expressed by using TB AIM depending on metabolism changes, that is, the switch from glucose to lactose. In the absence of lactose, the *E. coli* RNA polymerase that transcribes mRNA for protein production cannot bind to the promotor sequence because the repressor lac I links to the promotor and prevents T7 RNA polymerase expression. After switching to lactose, the latter will bind to lac I and make it fall off. Therefore, T7 RNA polymerase will be transcribed and translated, and the targeted protein will be expressed (Ramos et al., 2004).

Firstly, *E. coli* C41 (DE3) strain was used for W5F5 expression. This strain was reported as superior as a host for protein overexpression (Miroux & Walker, 1996). After expression, cell pellets were centrifuged, filtrated and lysed as described in section 6.6. Then, the protein was purified by affinity chromatography. In this project, pETM-11W5F5 and pETM-11WT were designed to express a protein with a histidine tag in its C-terminus to facilitate the protein purification process. The histidine tag is a highly efficient tool for recombinant protein purification using affinity chromatography because it does not require prior knowledge about the biochemical properties of the protein. The targeted protein separates in this method depending on the tag ability to interact reversibly with a ligand linked to a chromatography medium, metal ions ($Ni2^{+}$) within an immobilised metal chelate, which has the ability to link to the histidine tag of P450 BM-3 (Janson & Rydén, 1989). From the

chromatogram in Figure 7-1 A, the desired protein can be distinguished by the increase of UV absorption at 280 nm, which can be seen as a blue peak in fraction 5 and 6 at approximately 50% (v/v) of the elution buffer.

Purified protein was analysed using 10% SDS-PAGE electrophoresis. This technique separates proteins depending on the mobility of molecules, which is a function of their length and mass-to-charge ratio. SDS was used to linearise the proteins and give them a negative charge. When an electric field was applied across the gel, the negatively charged proteins migrated towards the anode at the bottom of the electrophoresis tank, with the distance depending on their size, so small protein molecules moved easily down the gel, while the larger molecules were trapped in the gel closer to the cathode (Rath et al., 2009).

Figure 7-1 B, Lanes 3 and 4 in SDS gel picture showed no P450 BM-3 W5F5 from the column during both loading and the washing process, indicating successful binding between the column matrix and the desired protein. In line 6, P450 BM-3 W5F5 is expressed successfully at 118 kDa, but it is combined with a high rate of non-target proteins at lower kDa values. The high level of impurities prompted attempts to improve the purification efficiency as well as investigate the expression using other host cells.



1	2	3	4	5	6
Protein ladder	Crude extract	Flow-through	Wash	Peak fractions (3-4)	Fractions (5-7)

Figure 7-1: A. Chromatogram to monitor the purification of P450 BM-3 W5F5 from C41 (DE3) by affinity chromatography, 45 ml of sample was loaded onto the His Trap 5 ml column and the protein was eluted at approximately 50% elution buffer. The purification conditions were 2 ml/min flow rate and 100% elution within 10 min. **B.** SDS-PAGE protein analysis, the arrow indicates the P450 BM-3 W5F5 at 118 kDa.

7.1.2 The effect of purification duration on protein purity

It is important to have a balance between resolution and time during purification. Furthermore, the purity of the protein depends on the purpose that it is intended for. In a functional study, as in this project, a high purification level is required. Therefore, P450 BM-3 W5F5 was expressed in C41 (DE3) and purified again using affinity chromatography, but this time the elution period was increased from 10 min to 20 min using the same flow rate (2 ml/min). The protein fractions from this process were checked by SDS-PAGE. By increasing the elution period, a slight enhancement in protein quantity was observed, while the level of impurities was still high (Figure 7-2).

7.1.3 Expression and purification of P450 BM-3 mutant using *E. coli* BL21(DE3)

The expression of P450 BM-3 W5F5 protein was also checked using *E. coli* BL21 (DE3), a common bacterial strain used for P450 expression. The same expression and purification conditions for expressed protein in C41 (DE3) were also used with this strain and the results are shown in Figure 7-3. By comparing the gel picture of the protein expressed in C41 (DE3) with the same protein expressed in BL21 (DE3), it could be observed that the latter is preferred due to the high concentration of P450 BM-3 W5F5 produced; however, the protein purity still needs to be improved.

In an attempt to reach the desired protein purity, the elution period during the affinity step was increased from 20 min into 40 min. By slowing down the elution process, the peak in the chromatogram seemed to be split into two intersecting peaks (Figure 7-4); peak 1 (lanes 5) and peak 2 (lanes 6-8). The P450 protein in peak 2 showed a significant improvement in terms of purity and concentration. The eluted fractions of peak 2 showed chromatic gradients from light yellow to dark red, light red then to light yellow, again referring to the heme protein concentration in these fractions (Figure 7-5).

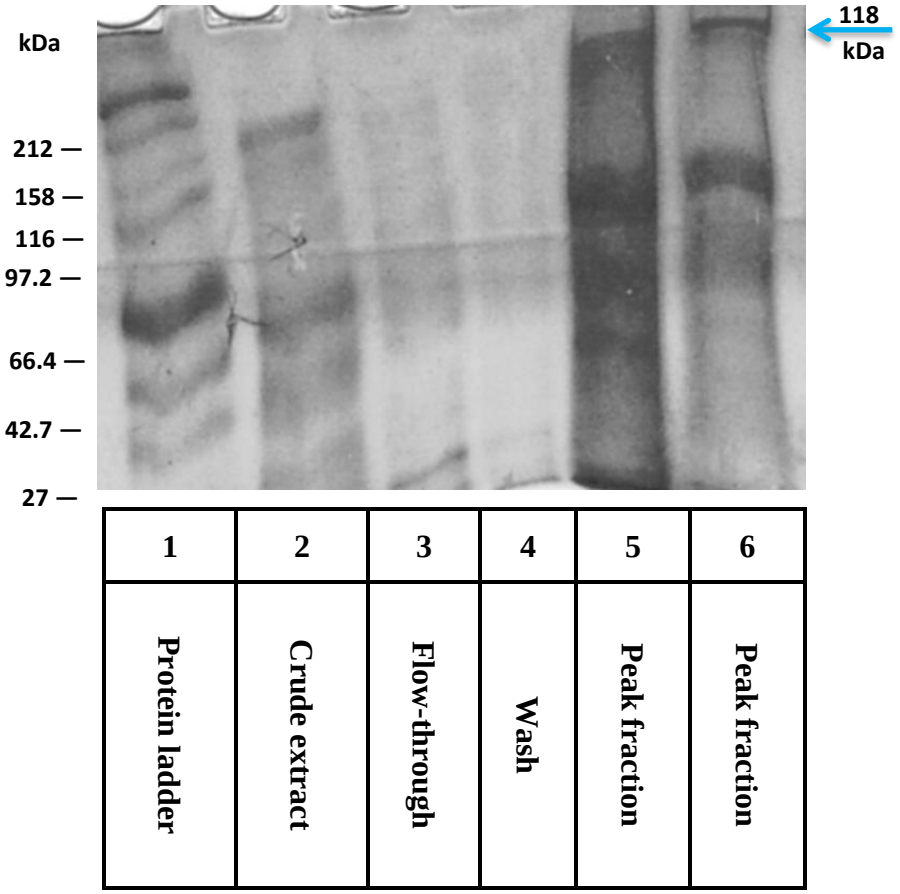
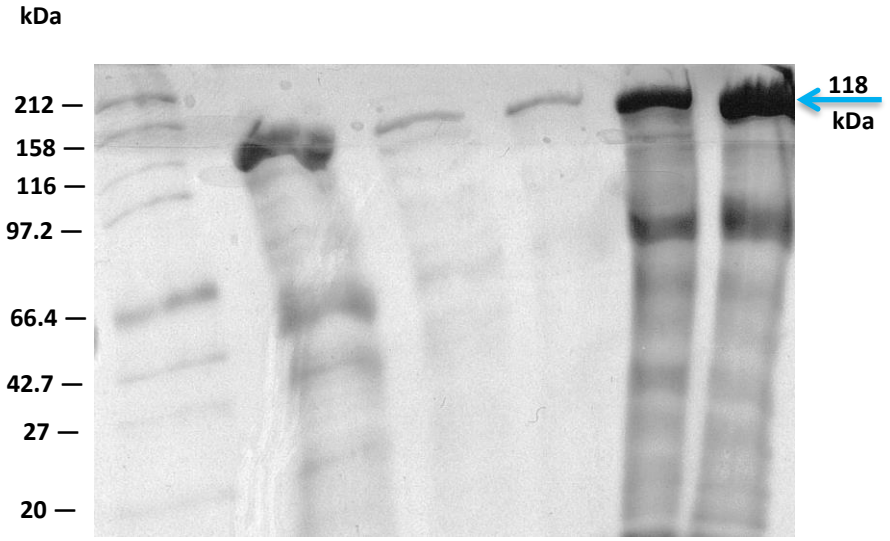
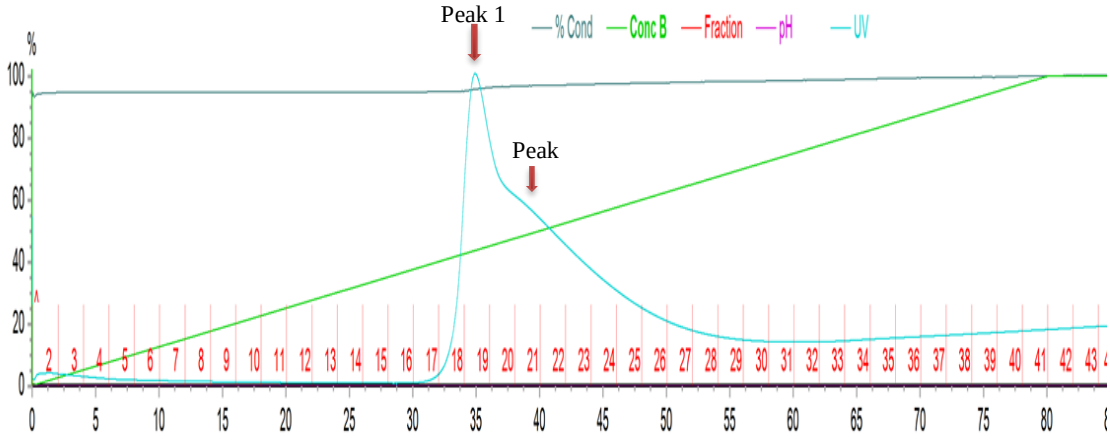


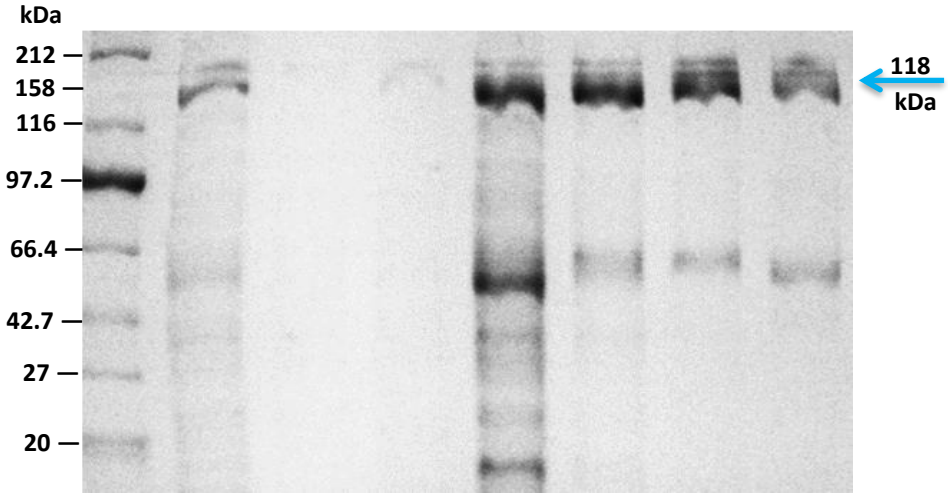
Figure 7-2: SDS-PAGE protein analysis of the eluted protein fractions from the affinity chromatography step, the arrow indicates the P450 BM-3 W5F5 at 118 kDa. Protein was expressed in C41 (DE3) and the elution flowrate was 2 ml/min for 20 min.



1	2	3	4	5	6
Protein ladder	Crude extract	Flow-through	Wash	Peak fraction	Peak fraction

Figure 7-3: SDS-PAGE protein analysis of the eluted protein fractions from the affinity chromatography step, the arrow indicates the P450 BM-3 W5F5 at 118 kDa. Protein was expressed in BL21 (DE3) and the elution flowrate was 2 ml/min for 20 min.





1	2	3	4	5	6	7	8
Protein ladder	Crude extract	Flow-through	Wash	Fraction 19 (Peak 1)	Fraction 20 (Peak 1)	Fraction 21 (Peak 2)	Fraction 22 (Peak 2)

Figure 7-4: A. Chromatogram monitored during the purification of P450 BM-3 W5F5 from BL21 (DE3) strain by affinity chromatography, 45 ml of sample was loaded to the HiTrap 5 ml column, the Peak represents eluted proteins at about 45% (v/v) elution buffer. The purification conditions are: 2 ml/min, 40 min. **B.** SDS-PAGE protein analysis, the arrow indicates the P450 BM-3 W5F5 at 118 kDa

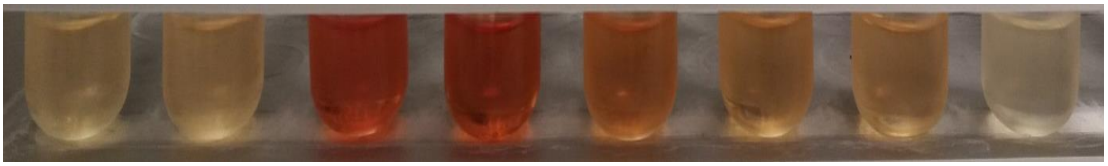


Figure 7-5: The chromatic gradients of the protein fractions from the affinity purification step from light yellow for low hemoprotein concentration to dark red for high hemoprotein concentration.

7.1.4 Using different expression inducing methods and media for W5F5 mutant expression

Before moving to another purification step, three types of auto-induction media were assessed for P450 BM-3 W5F5 expression in *E. coli* BL21 (DE3), 2×TY AIM, Terrific broth AIM (TB AIM) and Super broth AIM (SB AIM). The expression levels were visualised and quantified using SDS-PAGE gel analysis. From Figure 7-6, it can be seen that P450 BM-3 W5F5 expression level from highest to lowest could be achieved by using SB AIM, TB AIM and 2×TY AIM. The improvement in the expression gained by using SB AIM may due to the richness of this media by tryptone (35 g/l). In addition, the expression using IPTG instead of auto-induction media was investigated for P450 BM-3 production, but the protein showed a very low expression level. The harvested cell pellets of P450 BM-3 W5F5 from the expression using 50 ml 2×TY induced by 1 mM IPTG and the expression using 50 ml SB AIM are shown in Figure 7-7 A. The pellet using SB AIM shows a sign of high expression

distinguished by the redness of this pellet in comparison to the pellet from the expression using IPTG, which was less red in colour. The low expression by using IPTG was also noticed throw the SDS gel image (Figure 7-7 B).

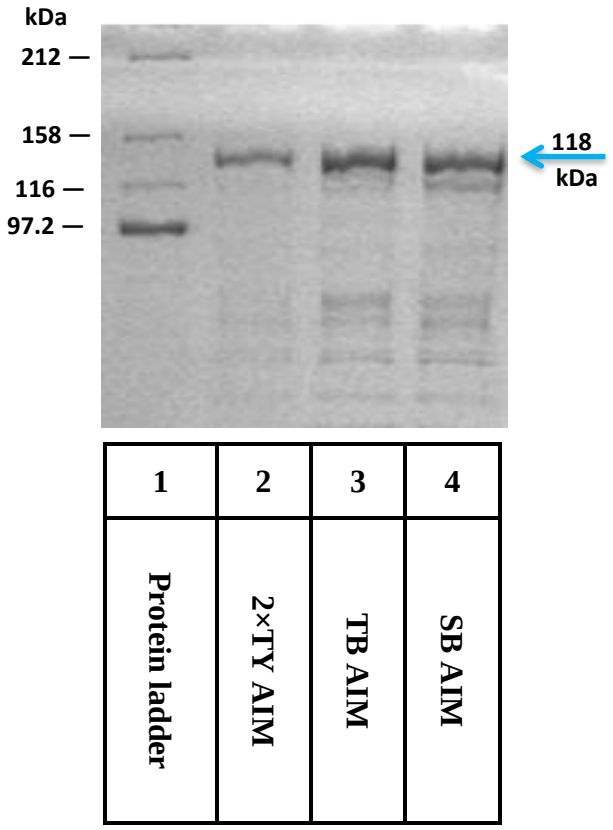


Figure 7-6: SDS-PAGE protein analysis using different ALM in BL21 (DE3); 2×TY, TB and SB. The expression was taken place at 30°C. The arrow indicates the P450 BM-3 W5F5 at 118 kDa.

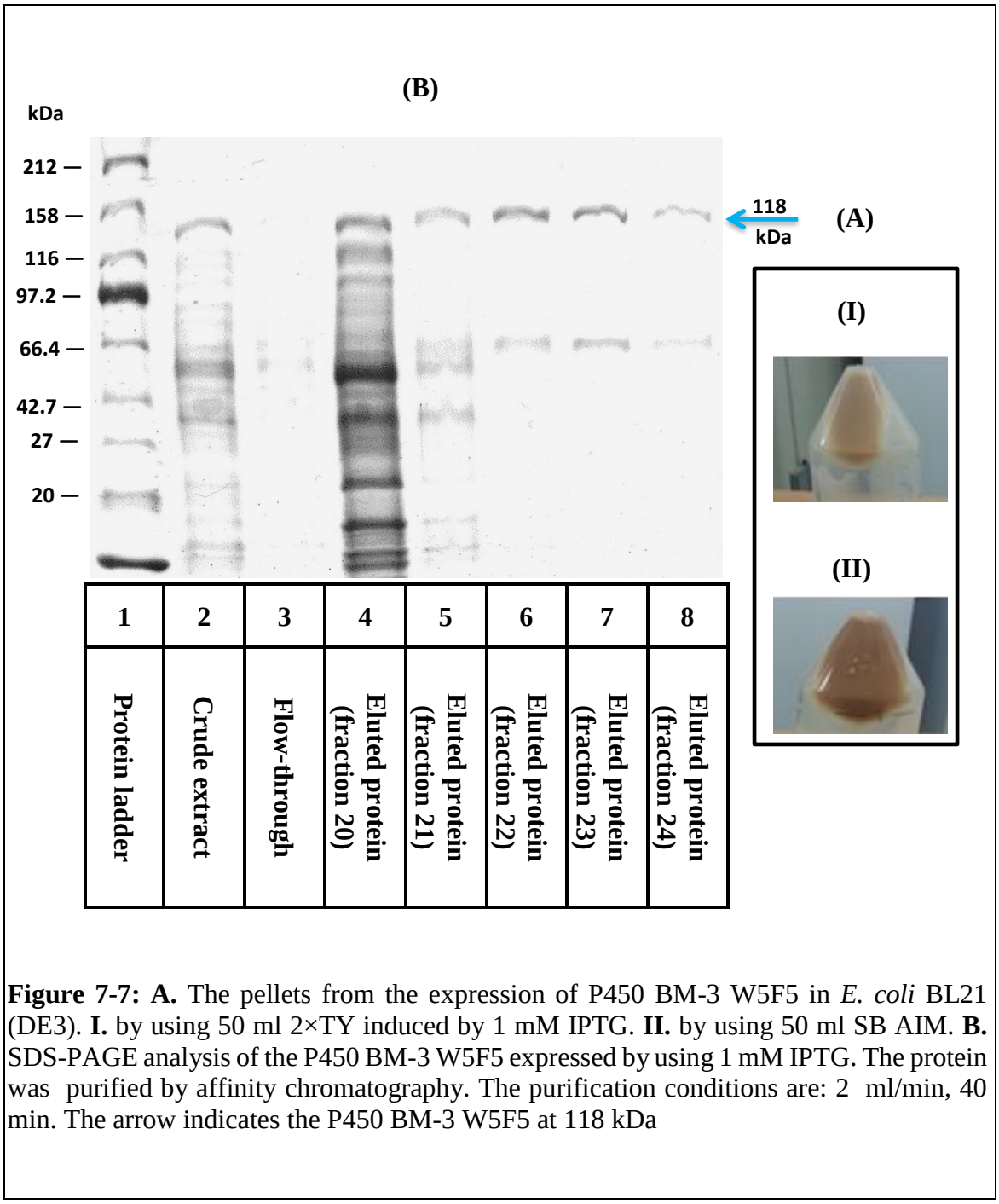
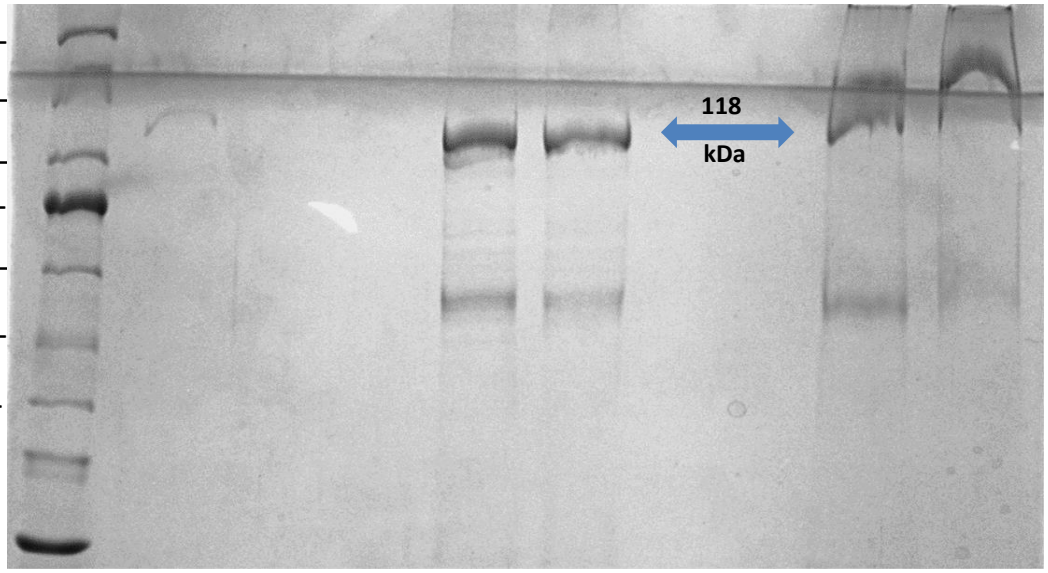


Figure 7-7: A. The pellets from the expression of P450 BM-3 W5F5 in *E. coli* BL21 (DE3). **I.** by using 50 ml 2×TY induced by 1 mM IPTG. **II.** by using 50 ml SB AIM. **B.** SDS-PAGE analysis of the P450 BM-3 W5F5 expressed by using 1 mM IPTG. The protein was purified by affinity chromatography. The purification conditions are: 2 ml/min, 40 min. The arrow indicates the P450 BM-3 W5F5 at 118 kDa

7.1.5 Ion exchange chromatography

The fractions with the desired protein from the affinity purification step were diluted and loaded onto the ion exchange column. This technique purifies proteins according to

differences in their net surface charge (Janson & Rydén, 1989). Two types of ion exchange column were compared for P450 MB-3 W5F5 purification and concentration, DEAE and super Q. From Figure 7-8, it can be concluded that the purity of protein did not improve by using any of these columns, but the protein was more concentrated using the super Q column. These results showed the importance of the addition of a third purification to achieve the desired purity.



1	2	3	4	5	6	7	8	9	10
Protein ladder	Crude extract	Flow-through (DEAE)	Wash (DEAE)	fraction 6 (DEAE)	fraction 7 (DEAE)	Flow-through (Super Q)	Wash (Super Q)	fraction 6 (Super Q)	fraction 7 (Super Q)

Figure 7-8: SDS-PAGE protein analysis during the purification of P450 BM-3 W5F5 from BL21 (DE3) strain by using two types of ion exchange chromatography columns **I)** DEAE 5 ml column **II)** Q 5 ml column. 40 ml of diluted sample (37.5 mM NaCl) was loaded to each column. 100% (v/v) of elution buffer and 2 ml/min were used. The arrow indicates the P450 BM-3 W5F5 at 118 kD

7.1.6 Gel filtration chromatography

Two proteins were eluted from the first and second purification steps, the desired protein at 118 kDa and non-target proteins at lower kDa values, approximately 50 kDa. The two proteins were different in size and could be separated using gel filtration chromatography (size exclusion). In this step, proteins are separated according to differences in their molecular size as they pass through a gel filtration medium (Janson & Rydén, 1989).

During this process, proteins are diluted extensively against buffer C (50 mM Tris-HCl, 1mM EDTA, 10% (v/v) glycerol, pH 7.2), consequently, the target protein was expressed using 100 ml of SB AIM instead of 50 ml and the sample was passed through affinity and ion exchange steps before it was ready for gel filtration. In order to check the efficiency of this technique for protein purification, 4 ml of the concentrated P450 BM-3 W5F5 solution was loaded onto a 320 ml Hiload 26/600 Superdex 200 pg gel filtration column. The chromatogram showed two separate peaks, peak 1 for the larger size protein and peak 2 for the smaller one. The fractions from both peaks were analysed using SDS-PAGE analysis (Figure 7-9). Visualisation of the gel indicated that the two proteins from the previous ion exchange column fractions were separated efficiently depending on the difference in size; large protein at 118 kDa and a smaller protein at approximately 50 kDa.

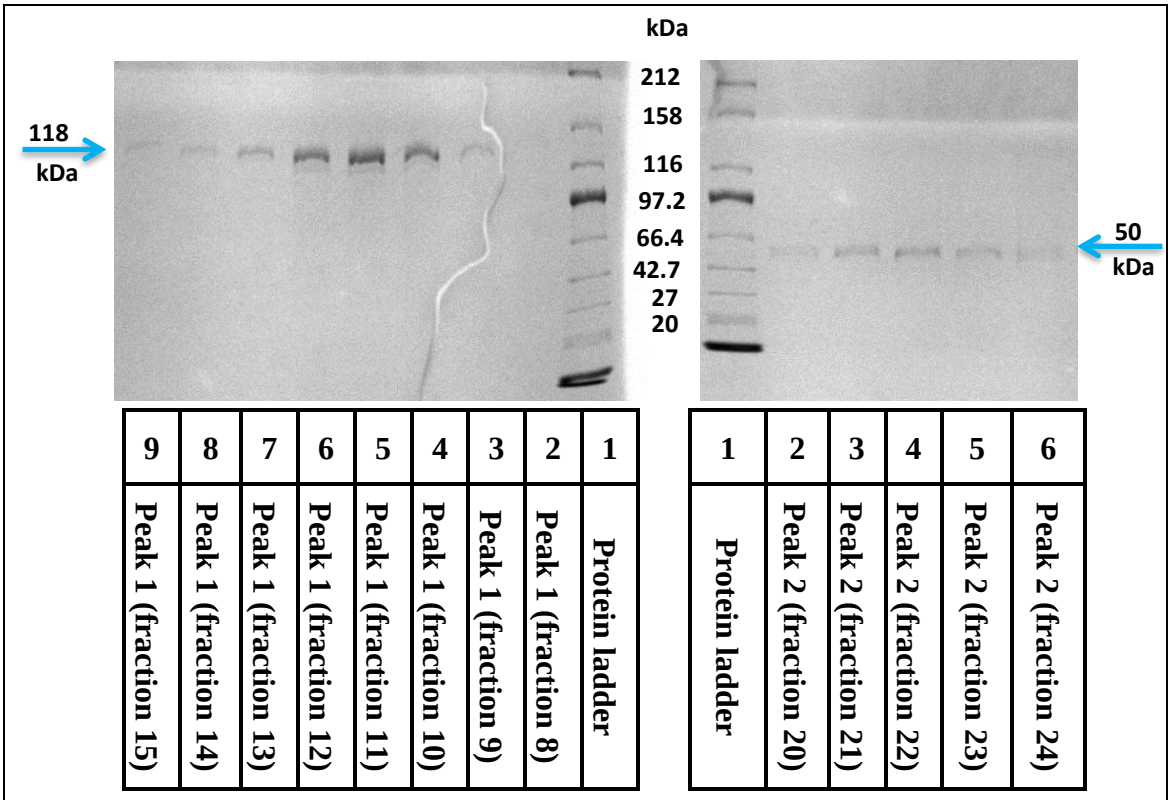


Figure 7-9: SDS PAGE analysis for the proteins from two peaks of gel filtration chromatography**. A.** SDS PAGE analysis for the first peak fractions (9 fractions from 8-16). **B.** SDS PAGE analysis for the second peak fractions (5 fractions from 20-24). 100 ml of SB AIM was used to express P450 BM-3 W5F5 at 30 C, the expression period was 24 hr. Target protein can be distinguished at 118 kDa while other unwanted protein at about 50 kDa.

7.2 Large-scale expression and purification

In order to achieve the desired protein concentration for use in the future functional investigation, 1200 ml of SB AIM was used to express P450 BM-3 W5F5. The optimum expression and purification conditions were performed. The pellets from the large-scale expression had a dark red colour, indicating a high expression level (Figure 7-10). The chromatograms of the three purification steps and SDS-PAGE gel images are shown in (Figure 7-11, Figure 7-12 and Figure 7-13). From the gel picture in Figure 7-13 B, it was evident that P450 BM-3 W5F5 was expressed and purified successfully, so could proceed with the enzyme activity assessment. The separation of two proteins in the protein solution when they passed through gel filtration column is presented in Figure 7-14

Figure 7-10: P450 BM-3 W5F5 cell pellets from 1200 ml expression culture. Each pellet comes from 400 ml expression media (SB AIM). The expression was taken place at 30°C and 200 rpm for 24 hr.

For comparison purposes, 400 ml of SB AIM was used to express the wild type protein, P450 BM-3 WT, with the same expression and purification conditions used for the mutant. The fractions from the last purification step, gel filtration chromatography, were analysed by electrophoresis and the SDS-PAGE gel showed good expression and purification levels (Figure 7-15).

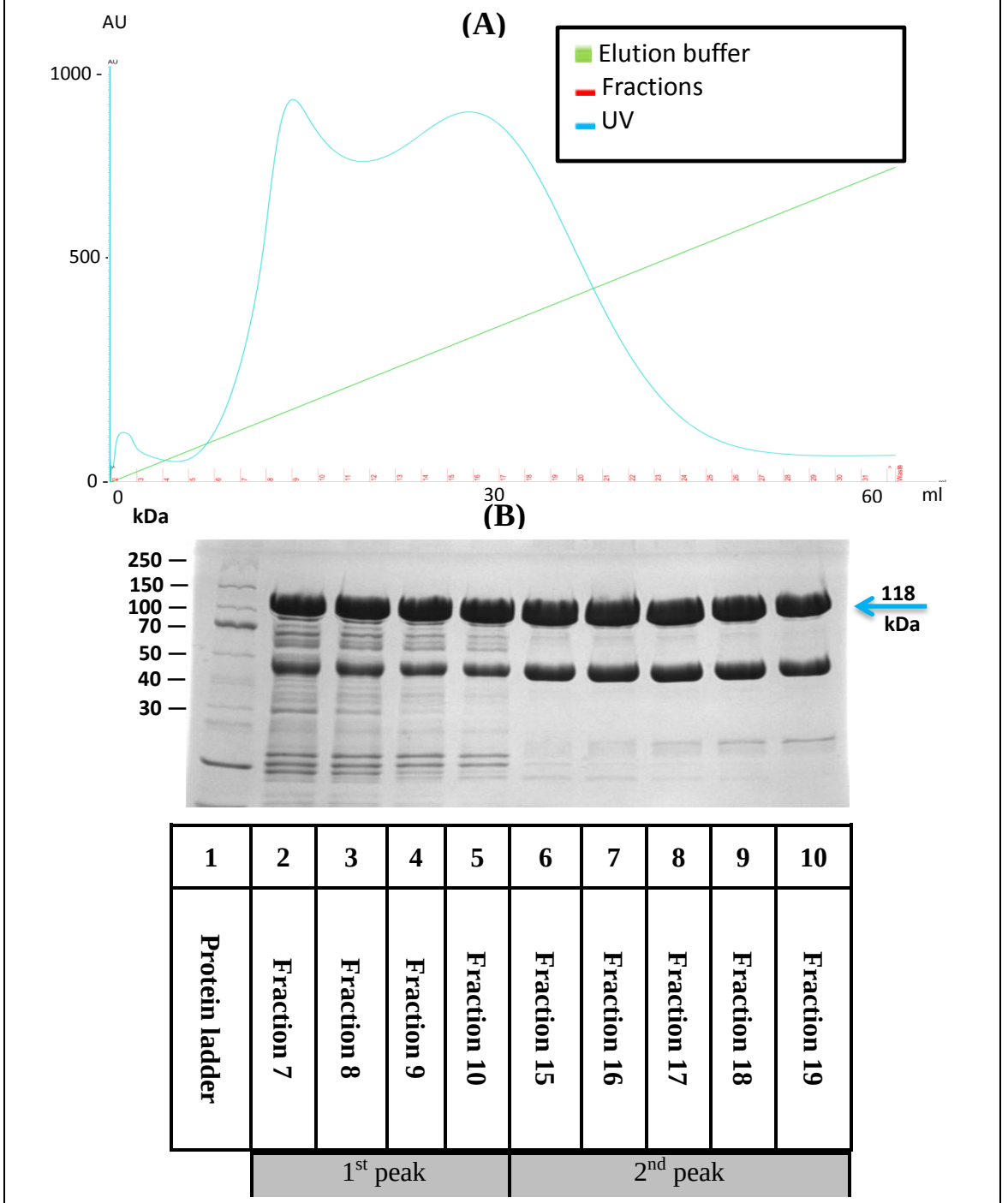
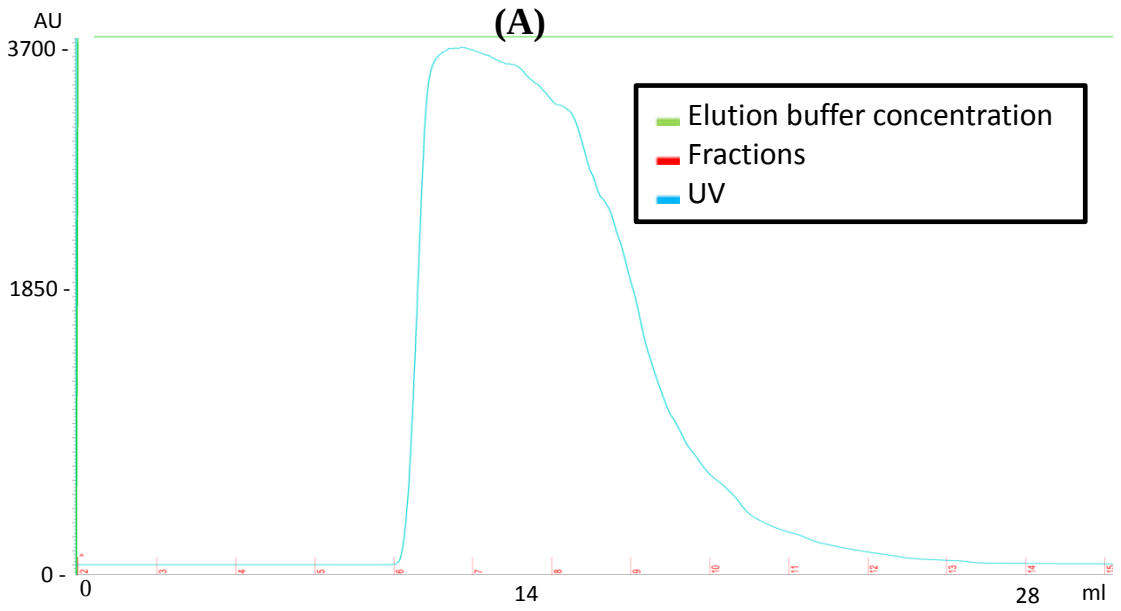
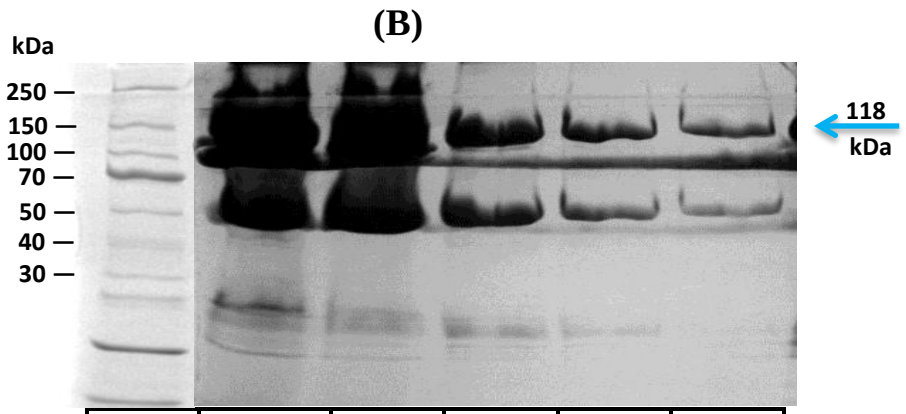


Figure 7-11: A. Chromatogram monitored during the large scale purification of P450 BM-3 W5F5 from *E. coli* BL21 (DE3) strain by affinity chromatography. The elution conditions were: 2 ml/min, 40 min. **B.** SDS-PAGE protein analysis for the first peak fractions (7-10) and second peak fractions (15-19)

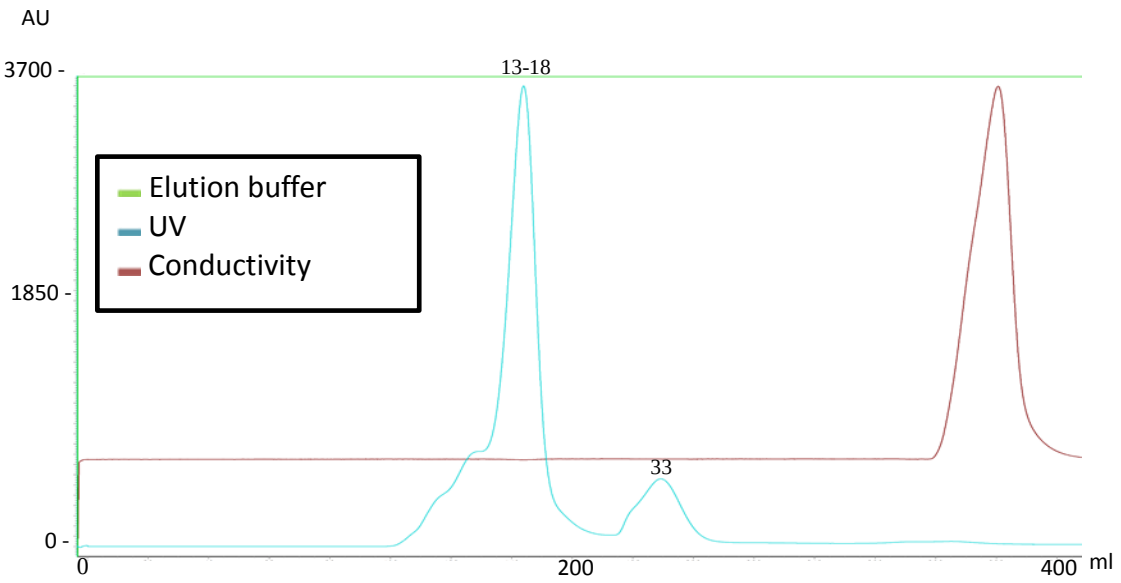




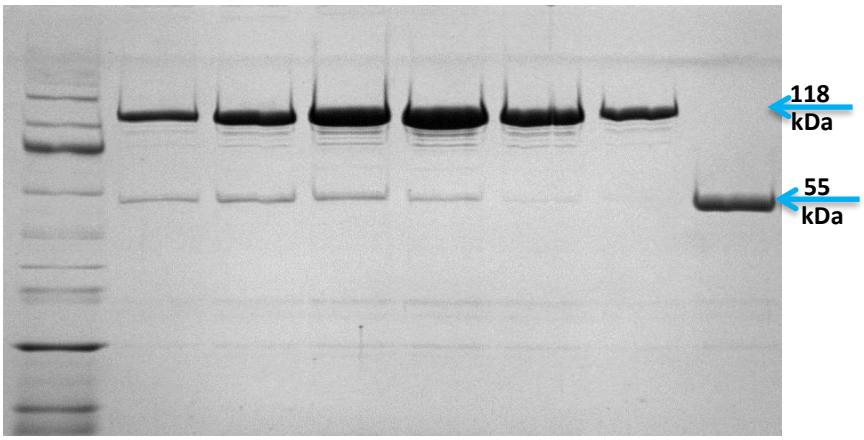
1	2	3	4	5	6
Protein ladder	Fraction 6 (The peak)	Fraction 7 (The peak)	Fraction 8 (The peak)	Fraction 9 (The peak)	Fraction 10 (The peak)

Figure 7-12: A. Chromatogram monitored during the large scale purification of P450 BM-3 W5F5 from *E. coli* BL21 (DE3) strain by ion exchange chromatography. The more pure fractions from affinity run (15-22) were loaded into super Q column. The elution conditions were: 2 ml/min, step elution. **B.** SDS-PAGE protein analysis for the peak fractions 6-10

(A)



(B)



1	2	3	4	5	6	7	8
Protein ladder	Fraction 13-peak 1	Fraction 14-peak 1	Fraction 15-peak 1	Fraction 16-peak 1	Fraction 17-peak 1	Fraction 18-peak 1	Fraction 33-peak 2

Figure 7-13: A. Chromatogram monitored during the large scale purification of P450 BM-3 W5F5 by gel filtration chromatography. **B.** SDS-PAGE protein analysis for the first peak fractions 13-18 and second peak fraction 33. The elution conditions were: 1.5 ml/min, step elution.

(A)

(B)

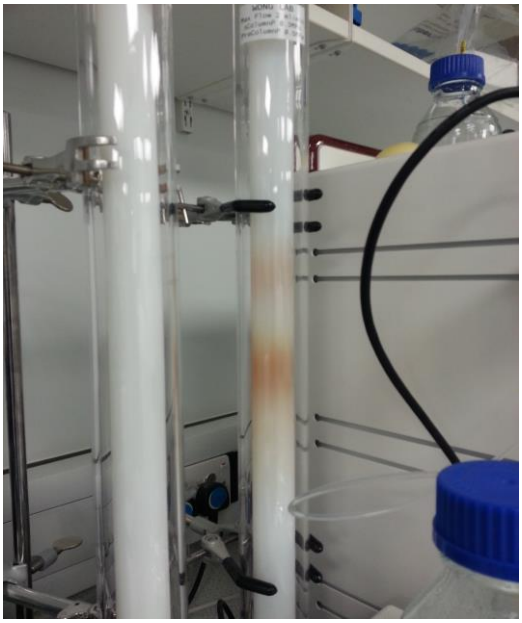


Figure 7-14: Protein purification by using gel filtration chromatography. **A.** the loaded sample before separation. **B.** two bands can be noticed due to proteins separation based on the difference in size

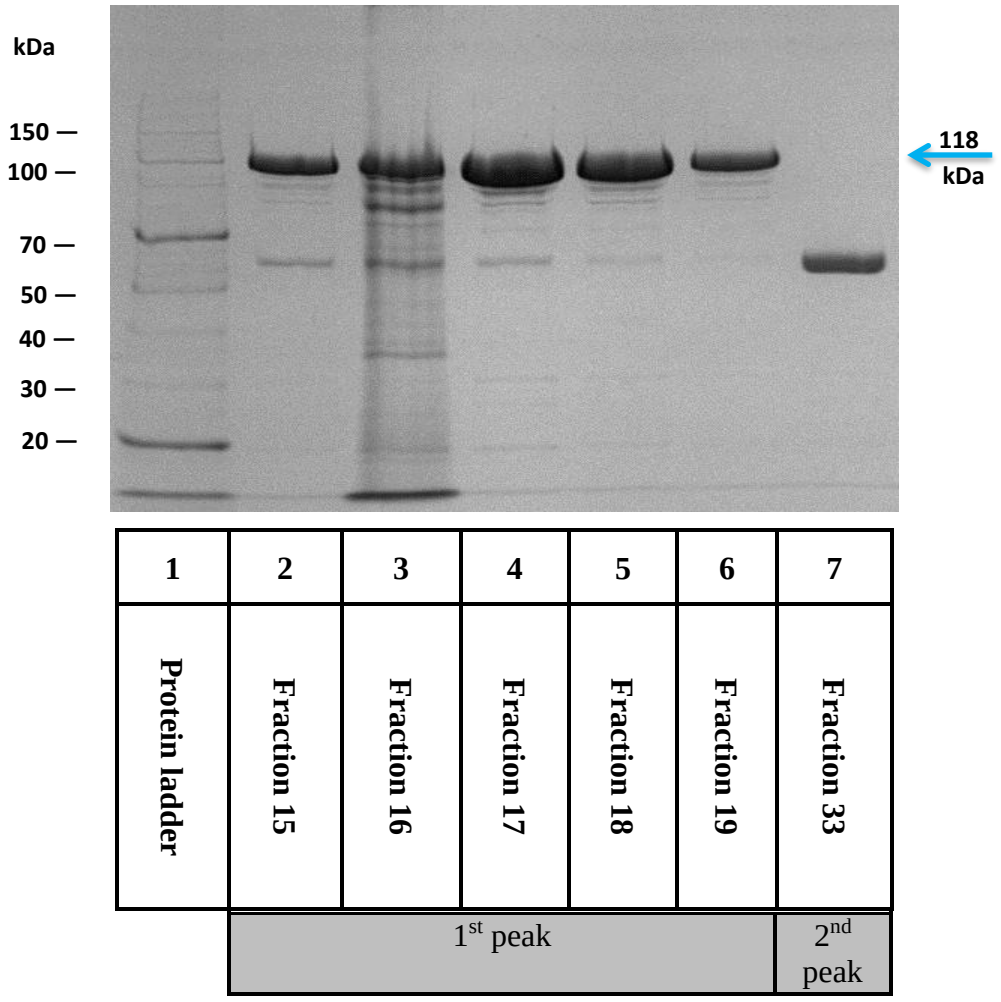


Figure 7-15: SDS-PAGE protein analysis of the large scale purification of P450 BM-3 WT by gel filtration chromatography for the first peak (fractions 15-19) and the second peak (fraction 33). The elution conditions were: 1.5 ml/min, step elution.

7.3 Water miscibility of solvents

Six solvents are considered green according to the GSK classification in 2010, 1-butanol, 2-butanol, butyl acetate, isopropyl acetate, propyl acetate and dimethyl carbonate. Before beginning the enzyme activity assay in the presence of these solvents, the solubility and miscibility of these solvents in water were checked depending on the information provided by PubChem; the online database tool of chemical molecules. According to the PubChem (https://pubchem.ncbi.nlm.nih.gov), butyl acetate, propyl acetate and isopropyl acetate are water immiscible with a very low solubility in water; 14 g/L, 18.9 g/L and 29 g/L (at 20 °C) respectively. On the other hand, the moderate solubility of 1-butanol, 2-butanol and dimethyl carbonate; 68 g/L, 181 g/L and 138 g/L respectively has increased the miscibility of these solvents. Different concentrations of these solvents 0–25% (v/v) were mixed with 100 mM potassium phosphate buffer (pH 7) and incubated for 30 min at RT. After the incubation period, the solutions were examined visually to indicate any separation in phases. Figure 7-16 shows that 1-butanol, 2-butanol and dimethyl carbonate were partially miscible in buffer, showing good miscibility at low concentrations, from 0–10% (v/v), however 1-butanol and dimethyl carbonate became immiscible at concentrations ≥ 15% (v/v) while 2-butanol stay miscible up to 20% (v/v). Depending on these results and to ensure homogenous reaction environment, P450s activity was investigated using no more than 10% (v/v) of the green solvents.

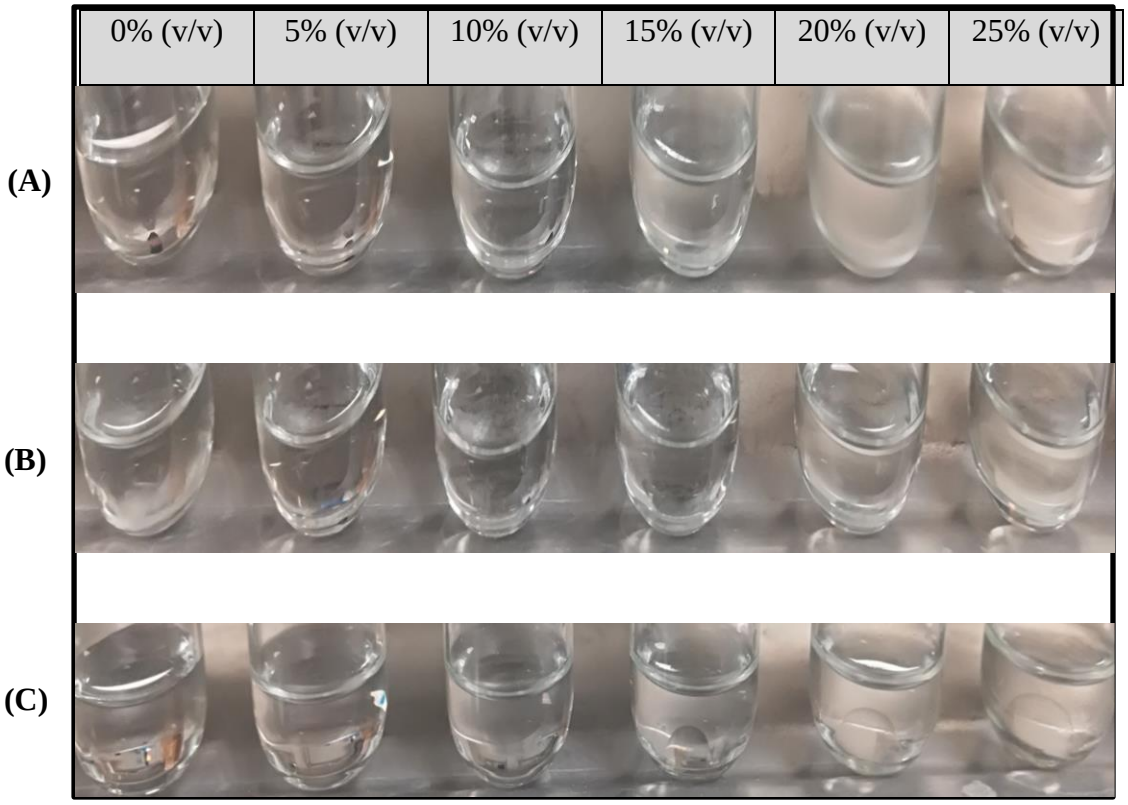


Figure 7-16: Miscibility of three green solvents**: A.** 1-butanol **B.** 2-butanol and **C.** dimethyl carbonate in 100 mM potassium phosphate buffer (pH 7). The solution started to show turbidity at concentrations ≥ 15% (v/v) of 1-butanol and ≥ 20% (v/v) of 2-butanol while the immiscibility of dimethyl carbonate became clear through the biphasic at concentrations equal or greater than15 % (v/v).

7.4 Effects of solvent on optical spectra

To check the features of the purified wildtype and its mutant W5F5, the UV-visible absorption properties of spectra at 250–750 nm were assessed. For both proteins, the spectral feature of the heme were clearly evident. The Soret band had a maximum absorption at 420 nm, with alpha/beta bands at 564 nm and 531 nm respectively, as shown in Figure 7-17. In addition, the purity of enzymes was evaluated by the RZ values (light absorbance at 420 nm for hemoprotein/light absorbance at 280 nm for total protein content in the solution). RZ values were 0.69 and 0.59 for the parents and the mutant respectively. After that, the spectra of the enzyme-solvent solutions were examined. There were no significant optical changes observed when these solvents were added to the P450 BM-3 WT and its mutant W5F5, as shown in Figure 7-18 to Figure 7-23. These results confirm that all green solvents used in this project do not affect the P450s structure or cause any blue shifts (as substrate) or red shifts (as inhibitor). Consequently, no interaction between these solvents and heme cofactor was observed through optical spectra scanning.

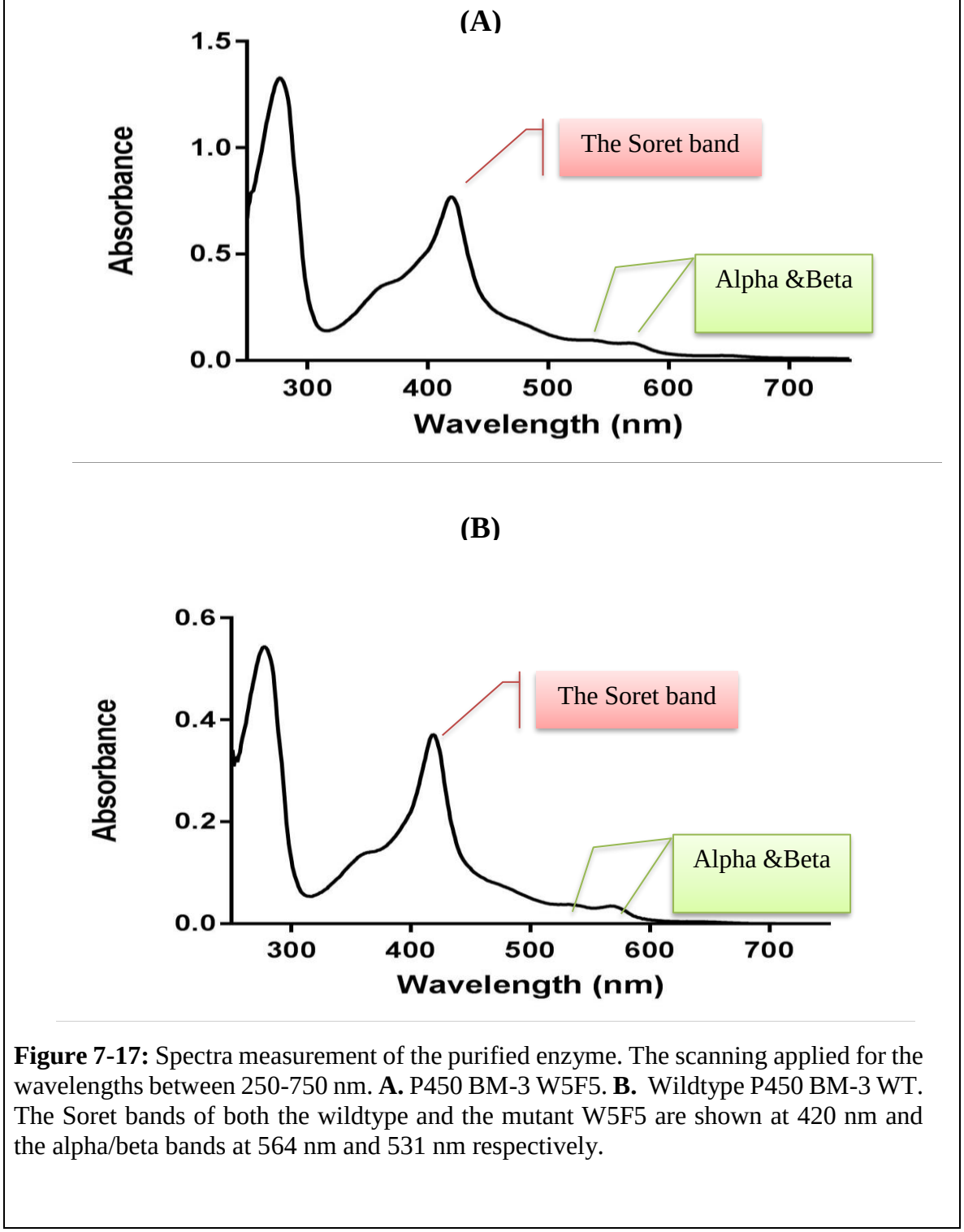


Figure 7-17: Spectra measurement of the purified enzyme. The scanning applied for the wavelengths between 250-750 nm. **A.** P450 BM-3 W5F5. **B.** Wildtype P450 BM-3 WT. The Soret bands of both the wildtype and the mutant W5F5 are shown at 420 nm and the alpha/beta bands at 564 nm and 531 nm respectively.

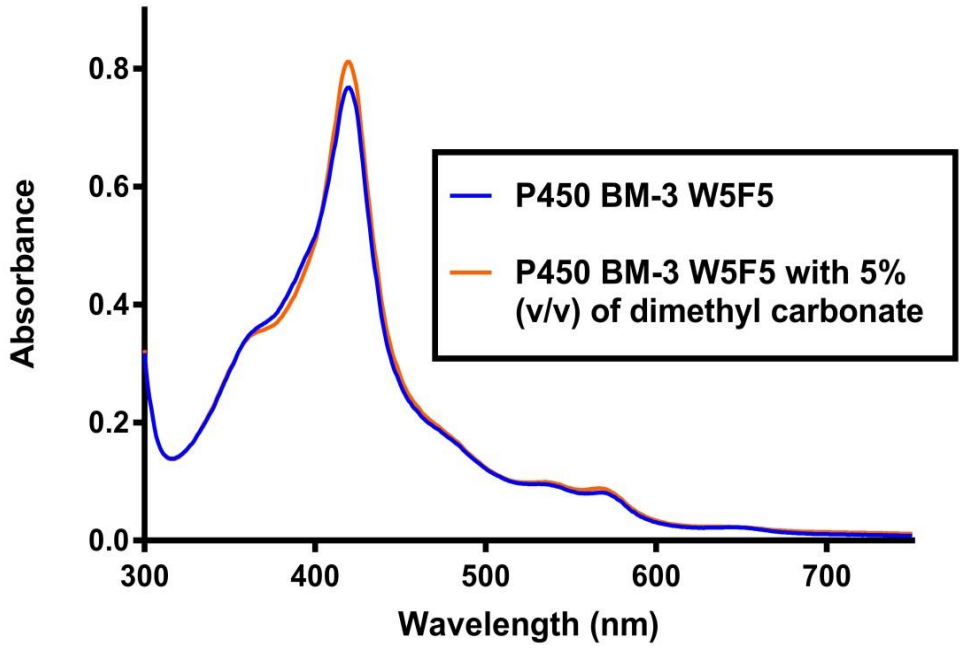


Figure 7-18: Effect of 5% (v/v) dimethyl carbonate on the spectra of P450 BM-3 W5F5. Light absorbance was checked between 300-800 nm to detect any interaction between the solvent and the heme cofactor of the protein.The concentration 5%(v/v) of dimethyl carbonate was chosen as this was the maximum concentration that used for P450 activity evaluation.

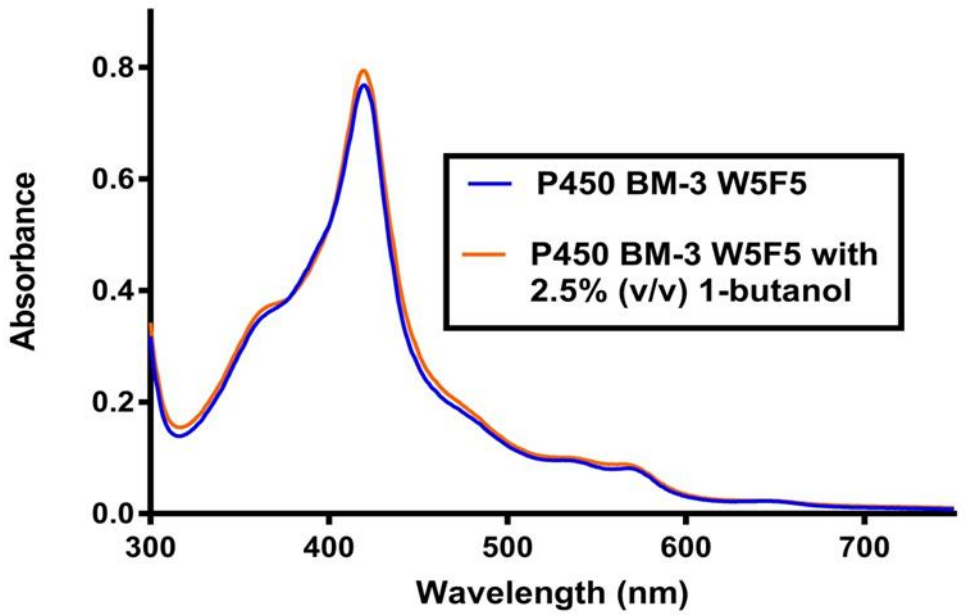


Figure 7-19: Effect of 2.5% (v/v) 1-butanol on the spectra of P450 BM-3 W5F5. Light absorbance was checked between 300-800 nm to detect any interaction between the solvent and the heme cofactor of the protein.The concentration 2.5%(v/v) of 1-butanol was chosen as this was the maximum concentration that used for P450 activity evaluation.

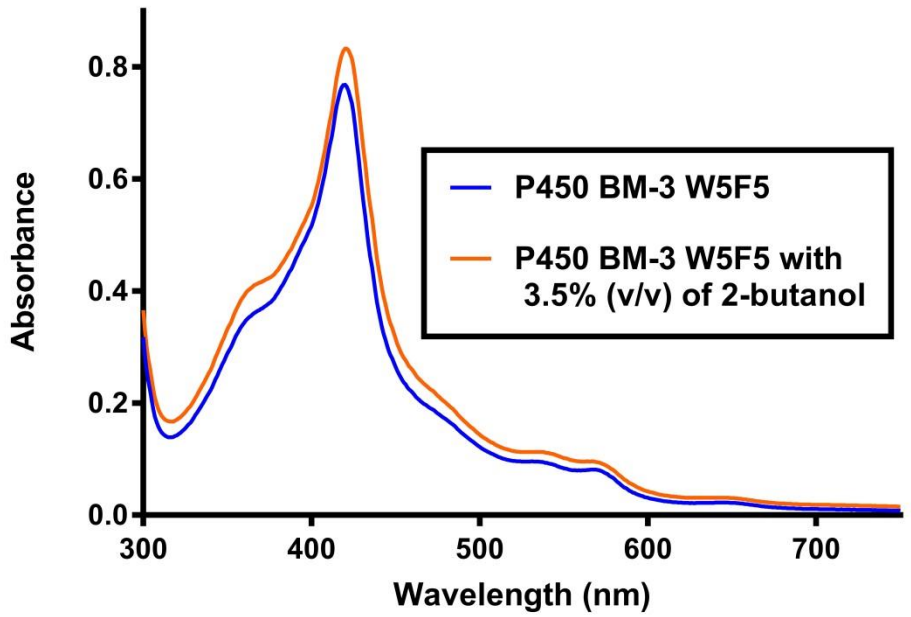


Figure 7-20: Effect of 3.5% (v/v) 2-butanol on the spectra of P450 BM-3 W5F5. Light absorbance was checked between 300-800 nm to detect any interaction between the solvent and the heme cofactor of the protein.The concentration 3.5% (v/v) of 2-butanol was chosen as this was the maximum concentration that used for P450 activity evaluation.

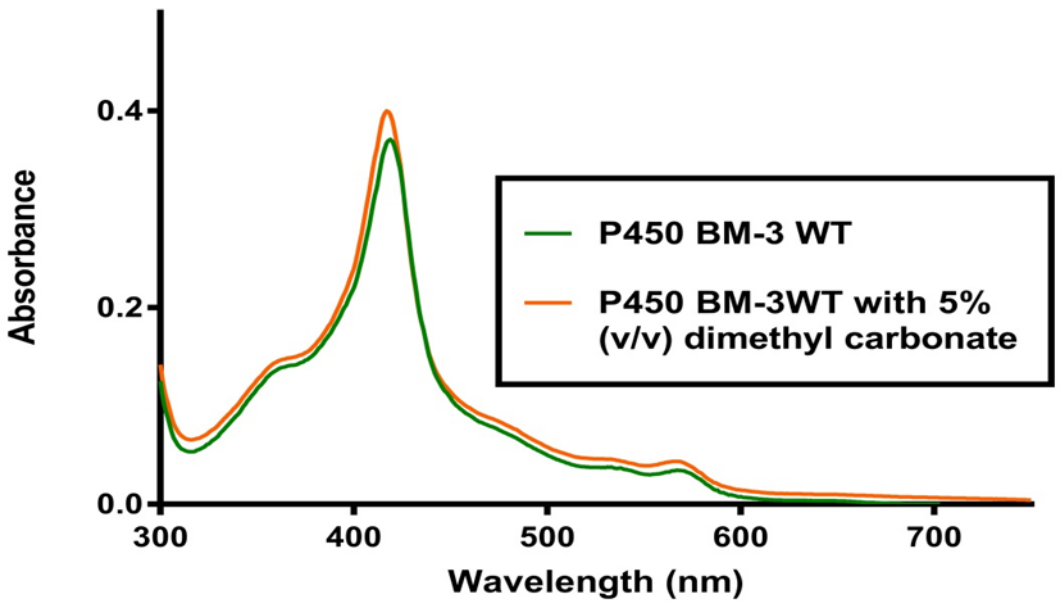


Figure 7-21: Effect of 5% (v/v) dimethyl carbonate on the spectra of P450 BM-3 WT. Light absorbance was checked between 300-800 nm to detect any interaction between the solvent and the heme cofactor of the protein.The concentration 5%(v/v) of dimethyl carbonate was chosen as this was the maximum concentration that used for P450 activity evaluation

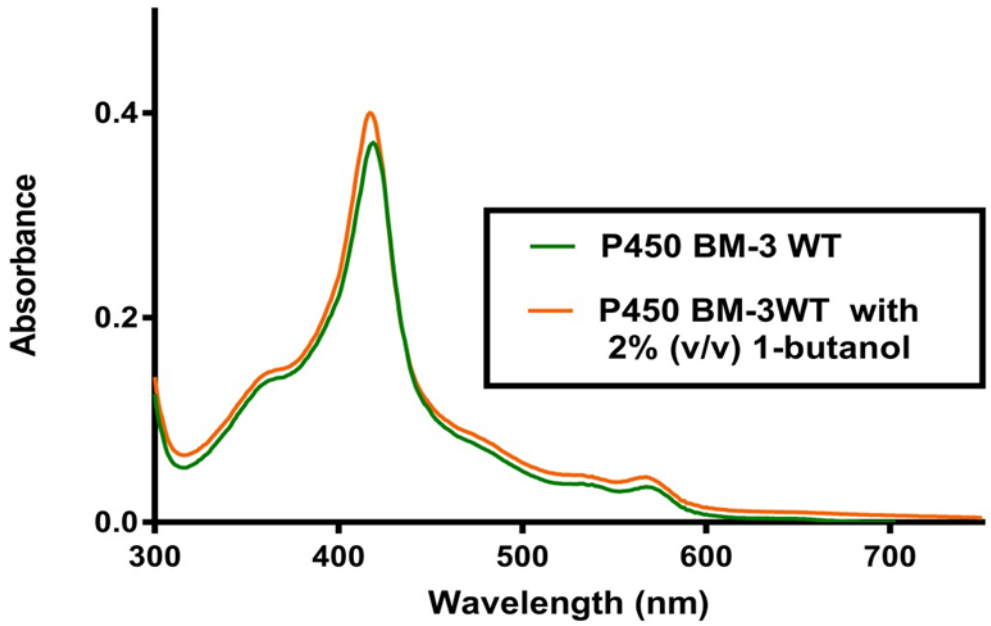


Figure 7-22: Effect of 2% (v/v) 1-butanol on the spectra of P450 BM-3 WT. Light absorbance was checked between 300-800 nm to detect any interaction between the solvent and the heme cofactor of the protein.The concentration 2%(v/v) of 1-butanol was chosen as this was the maximum concentration that used for P450 activity evaluation

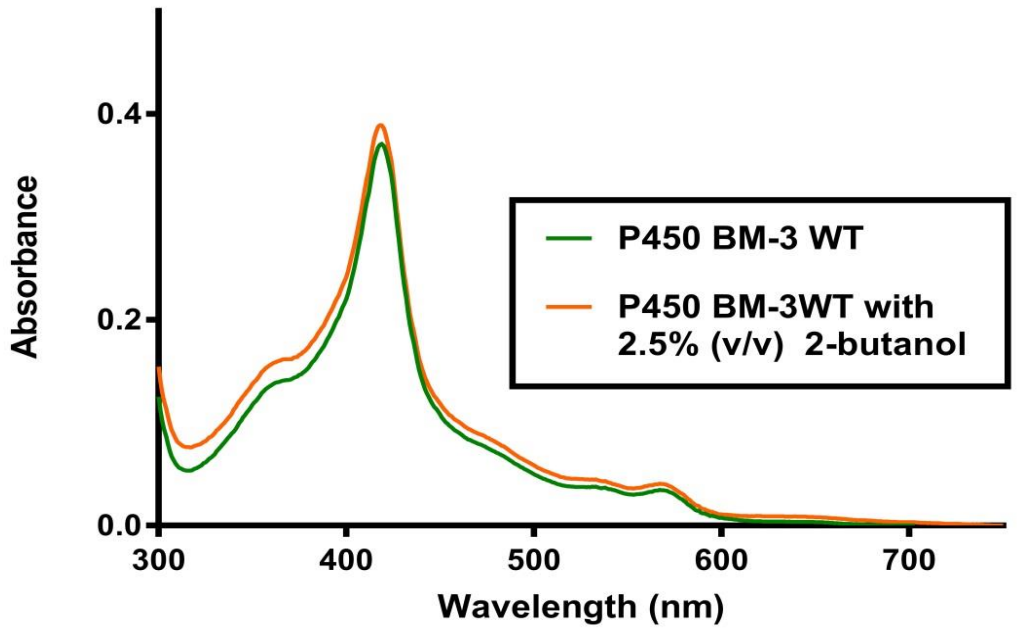


Figure 7-23: Effect of 2.5% (v/v) 2-butanol on the spectra of P450 BM-3 WT. Light absorbance was checked between 300-800 nm to detect any interaction between the solvent and the heme cofactor of the protein.The concentration 3.5% (v/v) of 2-butanol was chosen as this was the maximum concentration that used for P450 activity evaluation

7.5 Effects of protein dilution on protein stability

In this part of the project, different quantities of the enzyme were assessed to select an enzyme concentration which is able to perform in a range in which the plot of NADPH consumption versus time is linear when all other factors, including temperature, substrate concentration, reaction time and volume, are fixed. It can be seen from Figure 7-24 that 0.5% (v/v) W5F5 gave a high catalytic activity as well as a linear light absorbance within 60 sec. The concentration of 0.5% (v/v) was also used for the wildtype. The possibility of enzyme

denaturation over the experimental time was also assessed. Enzyme activity using two dilution ratios, 33.33% and 3.33%, were examined over a 6-hr experimental time period. High dilution of enzyme caused a significant decrease in enzyme activity from 100% to 69% after 6 hr, while decreasing the dilution ratio to 3.33% maintained the enzyme activity at a high rate of approximately 100%. This could be very well due to enzyme monomerisation which may be happened as a result of the excessive dilution (Whitehouse et al., 2012)(see Figure 7-25).

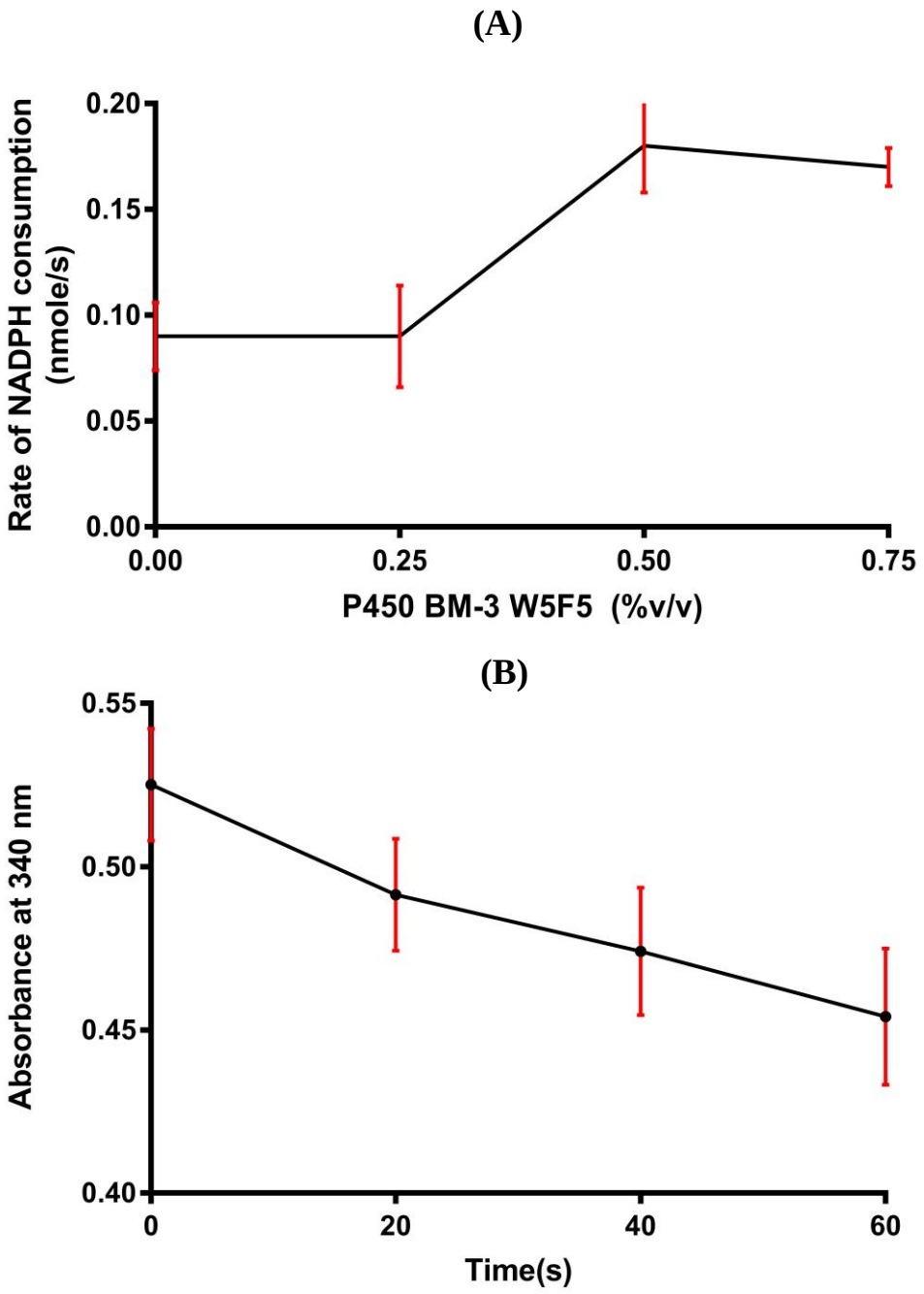


Figure 7-24: P450 BM-3 W5F5 concentration selection. **A.** NADPH consumption rate within different enzyme concentration (% v/v). **B.** Absorbance at 340 nm within 60 second using 5% (v/v) enzyme concentration. This absorbance indicates the decrease in NADPH quantity in the reaction mixture with time due to the reaction activity.

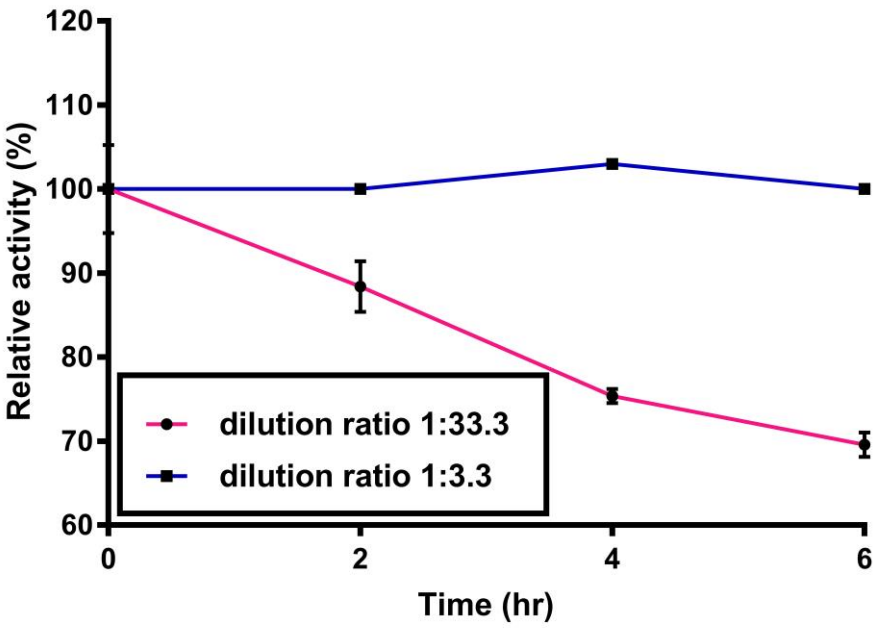


Figure 7-25: Relative activities of P450 BM-3 W5F5 variant as a function of time at two dilution ratios. Relative activity in this experiment is the ratio of specific activity at any time during the experiment to that at time zero

7.6 Tolerance of P450 BM-3 wildtype and W5F5 mutant to 1-butanol, 2-butanol and dimethyl carbonate

In order to examine the tolerance of P450 BM-3 WT and its mutant W5F5, the relative activity of these enzymes was calculated in the absence and presence of three green solvents: 1-butanol, 2-butanol and dimethyl carbonate. A continuous assay was applied to record the NADPH consumption at 340 nm per second. NADPH absorbs UV light exclusively in its reduced form and shows no absorption in oxidised forms at 340 nm. So, the decrease in NADPH concentration can be followed spectrophotometrically at 340 nm within the reaction period, and the enzyme activity is evaluated depending on this property (Noble et al., 1999) However, this assay was applied for 300 sec, but only the initial rates of the activity (60 sec) were used for the relative activity calculation. The relative activities of the purified parents and mutant in the presence of green solvents are presented in Figure 7-26, Figure 7-27 and Figure 7-28. Compared to wildtype, variant W5F5 showed a slight improvement in activity in 1-butanol and 2-butanol, but in general, the relative activity in these solvents was low, falling sharply to zero at co-solvent concentrations between 2–3.5% (v/v). Figure 7-28 indicates that variant W5F5 can tolerate significantly higher concentrations of dimethyl carbonate than the parent. The variant W5F5 showed approximately 128% relative activity in 5% (v/v) dimethyl carbonate in comparison to 30% (v/v) for the wild type. In contrast, 6% (v/v) dimethyl carbonate seems to be a critical value as the relative activity at this value is drastically reduced to approximately 80% (v/v) for the mutant variant.

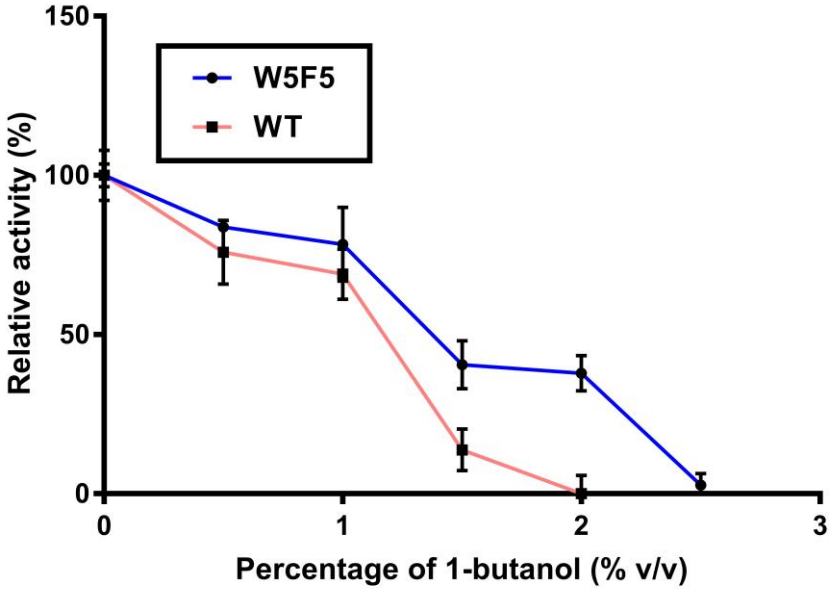


Figure 7-26: Relative activities of P450 BM-3 W5F5 variant and WT as a function of 1-butanol concentration. Relative activity in this experiment is the ratio of specific activity at any time during the experiment to that at time zero

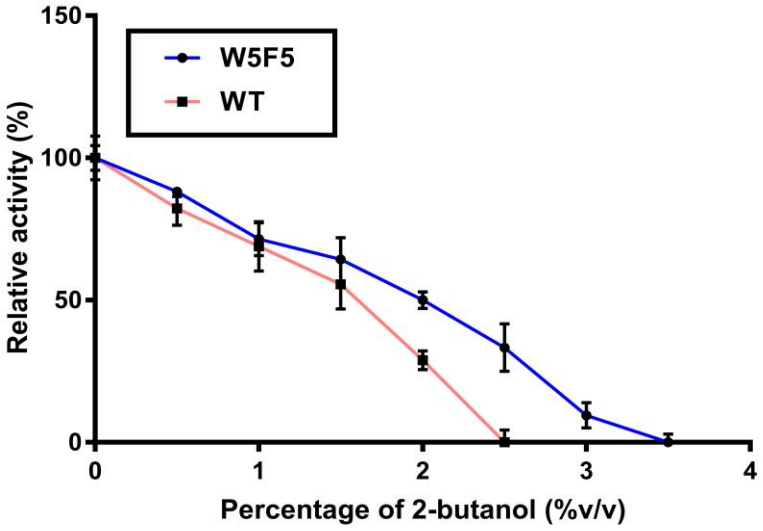


Figure 7-27: Relative activities of P450 BM-3 W5F5 variant and WT as a function of 2-butanol concentration. Relative activity in this experiment is the ratio of specific activity at any time during the experiment to that at time

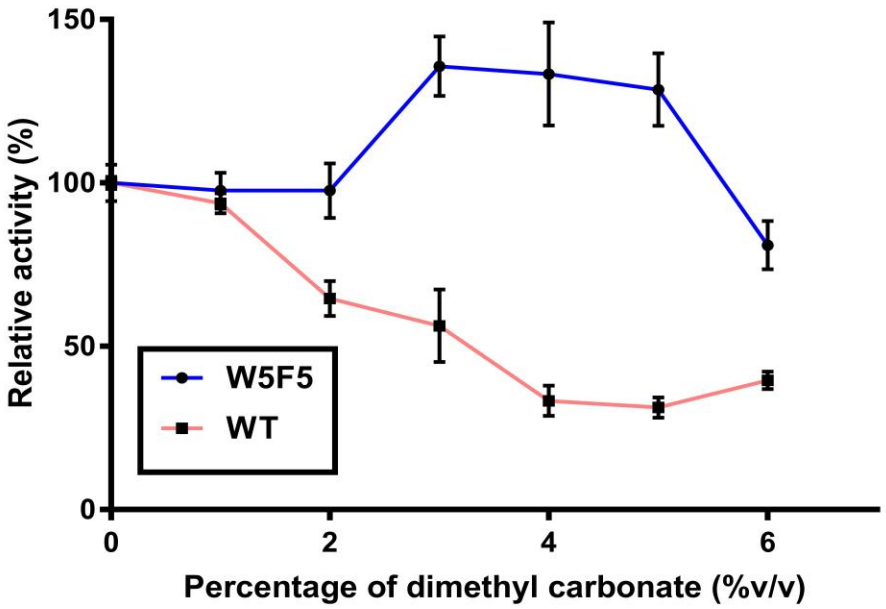


Figure 7-28: Relative activities of P450 BM-3 W5F5 variant and WT as a function of dimethyl carbonate concentration. Relative activity in this experiment is the ratio of specific activity at any time during the experiment to that at time

7.7 Conclusion

Cytochrome P450 BM-3 WT and its mutant W5F5 were successfully expressed in two *E. coli* strains, C41(DE3) and BL21(DE3). The characteristic red colour of the cell pellet reflects the ability of these strains to produce the heme cofactor. Although it is reported that C41(DE3) is superior for overexpression of seven recombinant proteins (Miroux & Walker, 1996) the yield of protein was higher in BL21(DE3). Aiming to optimise the expression of W5F5 variant, many expression factors were examined, such as induction methods and media, with the results revealing that the auto-induction media was preferable over using IPTG to induce the expression and the SB AIM was the best media for high protein expression levels.

In the preparation of microsomal monooxygenases for the functional study, the purity of protein is a vital issue. Three different steps were applied to achieve efficient separation of P450s: affinity chromatography, ion exchange chromatography and gel filtration chromatography (size exclusion). The purification efficiency of affinity chromatography step was increased by increasing the elution period from 20 min to 40 min, which slowing down the elution and increasing the efficiency of the purification process . This step was able to eliminate the majority of impurities from the crude protein extract. However, P450 BM-3 W5F5 was accompanied by a protein at lower kDa (approximately 50 kDa) as shown in Figure 7-4. The second purification step (ion exchange chromatography) in addition to improving protein purity, was utilised to concentrate the protein and reduce the protein fraction size to an acceptable level for the next step. The final step (gel filtration) aims to take advantages of the different physical properties of the proteins in the solution, in particular

different sizes. This step offered an excellent level of purity, with complete removal of the undesired protein (Figure 7-9).

Many parameters were checked prior to the investigation of the wild type and the mutant P450s' tolerance towards green solvents, such as the effect of these solvents on protein spectra and whether they interact with the heme cofactor of P450 BM-3 through scanning the protein spectra in the presence of these solvents, green solvent miscibility in water and the effect of dilution on protein activity. The activity evaluation of P450 BM-3 showed that the variant W5F5 is more tolerant than the wild type towards all green solvents tested. This is in agreement to a report by Wong and co-workers (2004), when they functionally compared this mutant with the wild type in the presence of many organic co-solvents classified as orange and red according to the GSK classification in 2010 (Wong et al., 2004). When the effect of these three green solvents on enzyme activity was investigated, it is found that the impact of 1-butanol and 2-butanol on both WT and W5F5 was high. The enzyme activity dropped to zero at solvent concentrations between 2-3.5% (v/v). In contrast, the variant W5F5 showed a significantly high tolerance against dimethyl carbonate; 128.5% relative activity at 5% (v/v) solvent concentration in comparison to 30% relative activity for the wild type at the same dimethyl carbonate concentration (Figure 7-28). The behaviour of the W5F5 variant in the presence of dimethyl carbonate is similar to its previously reported behaviour towards DMSO. Wong and co-workers recorded about 110% relative activity of the variant W5F5 at 5% (v/v) DMSO, but W5F5 remained active up to 25% (v/v) by using the later solvent, while it started to drop at 6% (v/v) dimethyl carbonate. This may be due to the immiscibility of dimethyl carbonate which was observed at 10% (v/v) and above.

Part III: Novel putative cytochrome P450s from *Cupriavidus necator* H16

8 Introduction to *Cupriavidus necator* H16 (*Ralstonia eutropha* H16)

8.1 *Cupriavidus necator* H16

Cupriavidus necator H16 is a Gram-negative bacterium found in both freshwater and soil, which multiplies by binary fission. It is a mesophilic rod-shaped bacteria, which in contrast to many other Gram-negative bacteria, is not pathogenic. This bacterium belongs to the order Burkholderiales, class Betaproteobacteria and it was isolated 60 years ago from soil near Goettingen, Germany (Alagesan et al., 2018). *Cupriavidus necator* H16 attracted the attention of bioscientists due to its ability to utilise a wide range of substrates and produce large quantities of polyhydroxyalkanoate (PHA), the biodegradable plastic with a wide range of applications (Pohlmann et al., 2006).

It was previously known as *Hydrogenomonas eutropha*, *Wautersia eutropha*, *Alcaligenes eutrophus* and *Ralstonia eutropha*, possibly due to its ability to utilise a variety of materials as nutrient sources. Recently, due to its high resistance to copper and genomic similarity with *Cupriavidus necator* strains, it was officially named as *Cupriavidus necator* H16, but many researchers contain to use *Ralstonia eutropha* in their articles (Lu et al., 2016; Vandamme & Coenye, 2004). In comparison to other family members, *C. necator* H16 represents one of the larger bacterial genomes, with a total size of 7,416,677 bp, the sizes of other members lie between 5.8 and 8.6 Mbp (Holden et al., 2004; Nierman et al., 2004; Salanoubat et al., 2002).

8.2 Genome

The *C. necator* H16 genome consists of two chromosomes and one megaplasmid: "chromosome 1 (4,052,032 bp), chromosome 2 (2,912,490 bp) and megaplasmid pHG1 (452,156 bp)". The G+C content is nearly identical in chromosome 1 and chromosome 2, with the proportion of coding regions almost the same in both chromosomes. *C. necator* H16 genome has 59 transfer RNA (tRNA) genes, the majority of which (51) are carried on chromosome 1, the reset genes are distributed as seven on chromosome 2 and one on the plasmid (pHG1). Five ribosomal RNA (rRNA) operons were reported in *C. necator* H16 genome, three on chromosome 1, two on chromosome 2. In addition, seven out of fourteen insertion elements are located on pHG1. From the gene distribution analysis of this organism, it was observed that the majority of key functions are located on chromosome 1, including DNA replication, transcription and translation, while chromosome 2 represents a refuge for many major reactions, such as aromatic compound decomposition, and utilisation of nitrogen sources as alternative nutrients. Nonetheless, all three replicons carry genes responsible for the synthesis of chemotaxis proteins and cell appendices (Pohlmann et al., 2006).

8.3 Metabolism and substrate utilisation

Over last 30 years, *C. necator* H16 was extensively studied because of its ability to produce PHA bioplastic on an industrial scale. This bacterium metabolises a wide range of organic substrates such as starch, fatty acids, tricarboxylic acid cycle intermediates, alcohols, polyols and sugars. Recently, agricultural residues were also investigated as substrates for this bacterium. In addition to the traditional organic substrate, *C. necator* H16 can also grow

autotrophically, it has the ability to use oxygen, hydrogen and carbon dioxide as alternative nutrients and energy sources (Lu et al., 2016).

8.3.1 Lithoautotrophic metabolism

C. necator can digest both carbon dioxide and formate (HCO_2^-) as carbon and energy sources via the Calvin-Benson-Bassham cycle (CBB). Lithoautotrophic/organoautotrophic metabolism can occur even under heterotrophic conditions. CO_2 is fixed by specified enzymes encoded in two *cbb* operons on chromosome 2 and pHG1, these operons are activated by a LysR-type transcriptional regulator (CbbR) located on chromosome 2. It became clear from the results of many studies that *C. necator* can induce CO_2 utilisation whenever required, but the roles of the *cbb* regulator and the positive effector (reduced metabolite) are as yet unclear and require further investigation (Bowien & Kusian, 2002). Two enzymes were reported for carbon fixation in *C. necator*: carbonic anhydrase (CA) and ribulose-1,5-bisphosphate carboxylase/oxygenase (RuBisCO). Four types of CA enzymes achieve efficient carbon fixation, encoded by the genes *can, can2, caa* and *cag*. Each gene plays a different role in CO_2 fixation. Can has important role in cell growth under atmospheric CO_2 as well as providing RuBisCO with CO_2. Caa helps in trapping the CO_2 inside the cell by converting it into more soluble bicarbonate when it passes through the cell membrane. The position of *caa* in the periplasm and being the favoured substrate for CO_2 helps this enzyme to perform its function. The functions of *cag* and *can*2 have not been fully elucidated, but it is thought that *can*2 helps in controlling the pH levels and *cag* could also work to supply CO_2 to RuBisCO. The second fixation enzyme is RuBisCO, this enzyme contributes to CO_2 fixation when two molecules of 3-phosphoglycerate are formed as a result

of the spontaneous split of an unstable intermediate (ribulose-1,5,-bisphosphate) during carbon fixation (Lu et al., 2016).

8.3.2 Heterotrophic carbon metabolism

C. necator H16 uses numerous organic components for heterotrophic growth as a carbon source. Fatty acids and lipids, when they are used as carbon sources by *C. necator* H16, gave more energy per mole carbon than any other source for carbon. Triacylglycerols from fatty acids cannot be metabolised directly by this bacterium, first they have to be digested by the extracellular lipase LipA into glycerols and free fatty acids. Chromosome 1 carries genes coding for the metabolism of organic acids, while chromosome 2 carries three genes coding for enzymes responsible for the degradation of 2-ketogluconate, glucosaminate, glucose and fructose. *C. necator* H16 has two sets of β-oxidation pathway genes (Embden-Meyerhoff-Parnas and the oxidative pentose phosphate pathways), both of which are vital for the cell growth. Moreover, the fatty acyl-CoA intermediates from these pathways contribute to the production of PHA from oils and fats (Pohlmann et al., 2006).

In contrast to many microorganisms, *C. necator* H16 is unable to digest many carbohydrates. Sugar as carbon source is restricted in the wild type of this bacterium to fructose, gluconate, xylose and N-acetyl-d-glucosamine (NAG), while the engineered variants of this organism have the ability to utilise glucose (Brandt et al., 2012; Holder et al., 2011; Raberg et al., 2012).

Cellulosic biomass is also reported as a carbon source for this bacterium. Wheat was used for PHB production as well as levulinic acid (refining cellulosic biomass) applied as a carbon

source. Recently, hydrolysed sago starch and sugarcane bagasse were used successfully to produce PHA (Koutinas et al., 2007; Wang et al., 2013; Yu & Stahl, 2008).

An interesting source of carbon for *C. necator* H16 is the aromatic compounds like phenol and benzoate. The ability of this organism to metabolise the organic compounds (known environmental pollutants) could make it a potential tool for environmental pollution treatment in the future (Johnson & Stanier, 1971; Pohlmann et al., 2006).

Engineered *C. necator* H16 can utilise some unique substrates as carbon sources. A short chain alcohol like ethanol has been converted to PHA when used by a *C. necator* variant. In addition, the by-products from various industries were used by different variants of this bacterium to produce PHA, including lactate, acetate, C2 α-hydroxy acid glyoxylate, glycolate and unprocessed crude glycerol (Lu et al., 2016). Figure 8-1 shows the difference in technique between the lithoautotrophic and the heterotrophic metabolism.

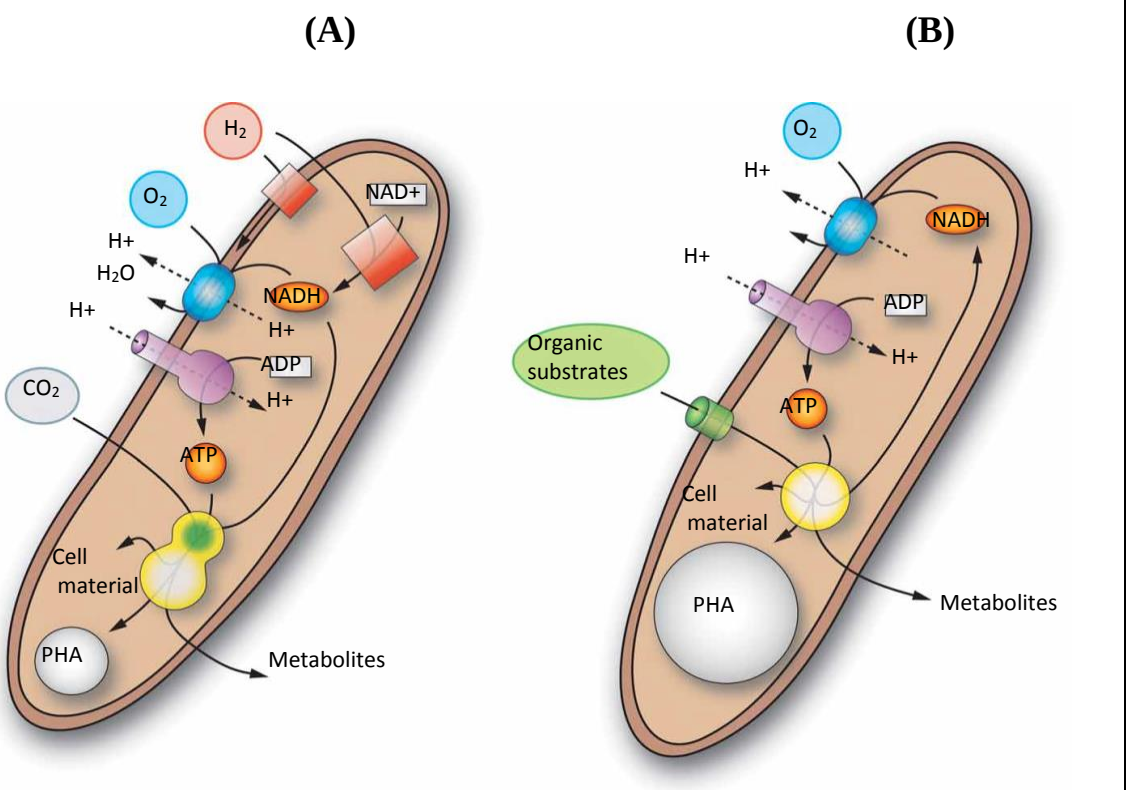


Figure 8-1: *Cupriavidus necator* H16 metabolism systems to digest different types of substrates and produce polyhydroxyalkanoate (PHA) **A.** Lithoautotrophic metabolism**,** when the bectrium uses different organic components for growth as a carbon source **B.** heterotrophic metabolism, the digestion of both carbon dioxide and formate (HCO_2^-) as carbon and energy sources via the Calvin-Benson-Bassham cycle (CBB). of This image acquired with permission from (Pohlmann et al., 2006).

8.3.3 Anaerobic metabolism

C. necator is able to grow under anaerobic conditions by using nitrogen compounds as an electron acceptor instead of oxygen. This denitrification process can be achieved by four reductase enzymes encoded on the pHG1: "Nitrate (NO_3^-) is sequentially reduced to nitrite (NO_2^-), nitric oxide (NO), and nitrous oxide (N_2O), respectively, by nitrate reductase (NAR), nitrite reductase (NIR), nitric oxide reductase (NOR), and nitrous oxide synthase (NOS)". It was suggested that sulphur also could be used as an alternative terminal electron acceptor

under anaerobic conditions through the sulphur reduction pathway. However, the genome sequence of *C. necator* H16 does not contain enzymes that form the sulphur reduction pathways such as "ArsRtype transcription regulator, sulphur dehydrogenase, sulphur-chelating/binding complex, thiosulfate-oxidising complex, and sulphate thioesterase", so reducing sulphur instead of nitrogen has not proved to date (Lu et al., 2016; Pohlmann et al., 2006).

8.4 Potential industrial applications

The ability of *Cupriavidus necator* H16 to metabolise diverse carbon sources and produce many value-added compounds means it represents an excellent microbial factory. In particular, this bacterium attracted bioscientists' interest due to its ability to produce and store many types of biopolymer, such as hydroxyhexanoate (PHH) and hydroxybutyrate (PHB), under stress conditions like nitrogen limitation or non-carbon nutrient starvation (Brigham et al., 2012). PHA represents an outstanding biodegradable replacement for petroleum-based polymer (as an example of PHA, chemical structure of PHB is presented in Figure 8-2).

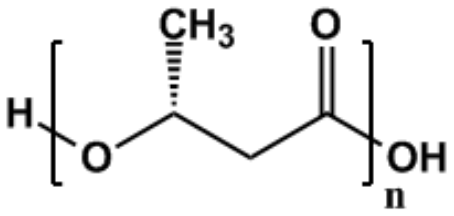


Figure 8-2: Structure of poly-(R)-3-hydroxybutyrate (PHB), a type of polyhydroxyalkanoate. The structure is generated by using ChemBioDrwa Ultra (http://www.perkinelmer.com/ chemdraw)

PHA polymers have a wide range of applications in packaging, agricultural and waste treatment, and disposable household products but the most impressive application is in the medical sector to manufacture many items such as surgical meshes, implants, sutures and scaffolds (Brigham & Sinskey, 2012; Lu et al., 2016). Many different types of monomers can be synthesised by *C. necator* H16, each with different thermal and mechanical properties: I) short chain monomer (C3-C5), (*scl*) PHA such as homopolymer (PHB); II) medium-chain monomers (C6-C12), (*mcl*) PHA and III) combination of scl and mcl monomers (*scl-co-mcl* PHA). Due to its advantageous properties and similarity to petroleum-based plastics scl-co-mcl, PHAs from engineered *C. necator* H16 became the goal of production (Lu et al., 2016).

C. necator was also used to produce non-alcohol biofuel when PHA was converted into methyl esters by an acid-catalysed hydrolysis. Free fatty acids, which can also be converted to methyl esters, were produced from engineered *C. necator* with inactivated PHB biosynthesis and β-oxidation pathways. This strain was also able to produce another fuel, methyl ketone (Chen, 2009; Müller et al., 2013; Torella et al., 2013). Another engineered strain of this bacterium, with no PHA production genes, has been shown to produce isobutanol and isopropanol by redirecting the extra energy produced as a result of PHA gene elimination (Lu et al., 2016).

Each hydroxyalkanoate monomer has a chiral centre on the carbon atom with two functional groups, a hydroxyl group and a carboxyl group (see Figure 8-2), which enable the bacterium to produce different fin chemicals like perfumes, vitamins and antibiotics (Gao et al., 2011). The monomer unit, (R)-3-hydroxybutyric in the PHB is an essential element for the production of antibiotics from *C. necator*. In starving conditions, there are seven depolymerases in this bacterium responsible for the hydrolysis of PHB into (R)-

hydroxyalkanoate acid monomers without any *in vitro* hydrolysis step. (R)-hydroxyalkanoate precursors are also produced by the elimination of the PHA synthase genes from *C. necator*, which allows the production and direct isolation of (R)-hydroxyalkanoates. The ability of *C. necator* to synthesise multi-monomer lengths of PHA increases the chance to produce different hydroxyalkanoate chiral compounds (Lee & Lee, 2003). In addition, the ability of *C. necator* to digest different types of organic compounds makes it an outstanding biological tool for bioremediation. This bacterium was used successfully to clean up soils from two pollutants, p-nitrophenol and 2,4-dichlorophenoxyacetic acid (Chen et al., 2004; Watanabe, 2001). The ability of *C. necator* H16 to produce different value-added chemicals is demonstrated in Figure 8-3

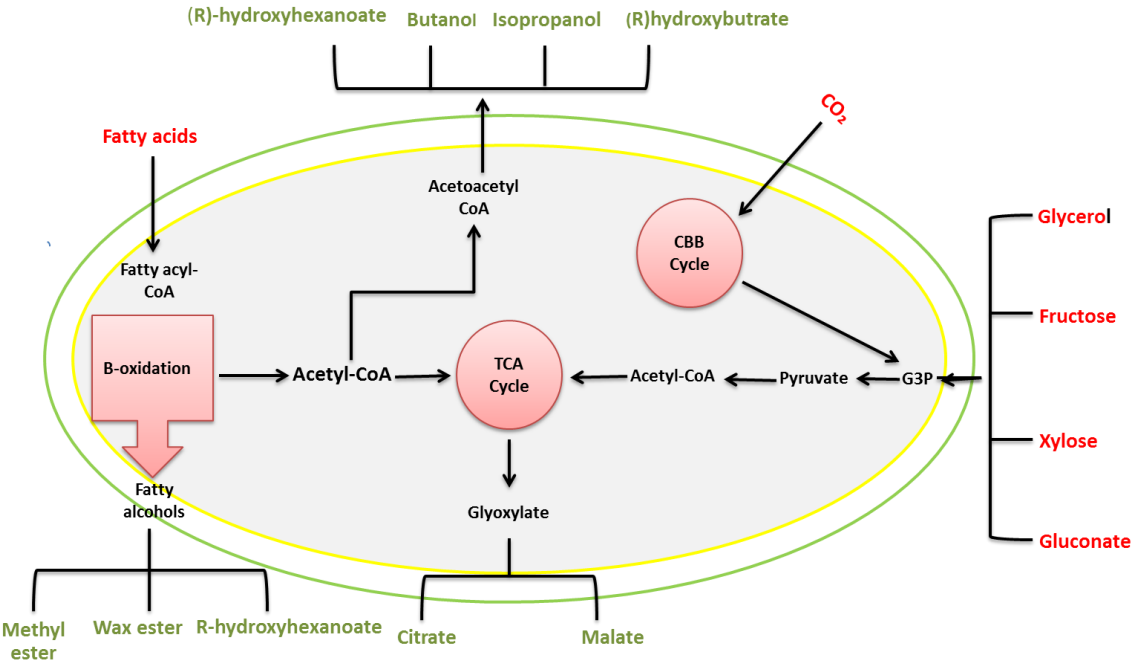


Figure 8-3: Scheme explaining the ability of ***C. necator*** H16 to metabolise and produce a variety of materials. The substrates are in red color and the products are in green. .

8.5 Summary of known cytochrome P450 enzymes from *Cupriavidus* sp.

The genome sequence study of many *Cupriavidus* sp led to the identification of numerous genes which were classified as cytochrome P450 enzymes according to the BLAST database. Cytochrome P450 enzymes were distinguished in many *Cupriavidus* sp such as *Ralstonia* sp. GA3-3, *Cupriavidus* sp. SK-4, *Cupriavidus* sp. UYPR2.512, *Cupriavidus* sp. IDO and *Cupriavidus* sp. taiwanensis (for the full list of P450s from *Cupriavidus* sp., go to BLAST (www. blast.ncbi.nlm.nih.gov)).

Although, the genome sequence analysis of many *Cupriavidus sp*. identified cytochrome P450 as a part of these sequences and gave information about the genome sequence of P450, to date only one characterisation study published in 2012 analysed the spectroscopic, biochemical and catalytic properties of a member of the *Cupriavidus* sp, CYP116B1 from *Cupriavidus metallidurans (Warman et al., 2012).*

CYP116B1 is a redox partner fusion enzyme in which the electrons are typically derived from NAD(P)H and are delivered to the P450 by the iron-sulphur centre (2Fe-2S) and FMN. A spectrophotometric scanning of this enzyme showed the feature of P450 with a Soret maximum at 418 nm and alpha/beta bands at 566 and 532 nm, respectively. The scattering of light indicated that CYP116B1 is a monomer enzyme (Warman et al., 2012). Three unsaturated fatty acids (palmitoleic, myristoleic and arachidonic acids) showed a substrate-protein binding feature (Type I) but the protein was unable to oxidise these fatty acids. However, this protein showed catalytic activity towards two herbicides, S-propyl dipropylthiocarbamate (vernolate) and S-ethyl dipropylthiocarbamate (Warman et al., 2012).

The following chapters (chapters 9-11) present the first report of the modelling, cloning, expression, purification, spectroscopic, substrate binding possibility and activity evaluation of four *C. nectar* H16 P450 enzymes (H16_B2406, H16_B1743, H16_B1279 and H16_B1009).

9 Methods and Materials

9.1 Chemicals

All chemicals are of analytical grade or higher quality and purchased from VWR Prolabo (Belgium), AppliChem (Germany), Sigma (UK), Merck KGaA (Germany) and Formedium LTD (UK). All biological materials used in this project are in high quality and purchased from Sigma (UK), Roche (UK) and New England BioLab (UK). The devises, kits and media used in this part of the thesis were mentioned in sections 6.2-6.3.

9.2 Strains and vectors

Dr Tuck Seng Wong and Dr Kang Lan Tee graciously provided four pETM-11 vectors encoding four P450 genes: H16_B2406, H16_B1743, H16_B1279 and H16_B1009. Each construct has a kanamycin resistance gene to allow selection, polyhistidine tag and TEV-site to facilitate the purification process and lac repressor induction control. The *E. coli* strain DH5α was used for all routine cloning experiments, while *E. coli* strains BL21 (DE3) and C41 (DE3) and HMS 174 (DE3) were used for recombinant protein expression.

9.3 Identification, sequence alignment and structure modelling of novel putative P450s in the *Cupriavidus necator* H16 genome

Although the protein genome sequences of four P450s from *Cupriavidus necator* H16 (CYP-H16-B2406, CYP-H16-B1743, CYP-H16-B1279 and CYP-H16-B1009) were reported, there was a lack of detailed information regarding sequence analysis, sequence alignment or

3D protein structure. Many online tools were used for P450s characterisation and structure modelling, a list of these tools and their functions and websites were presented in Table 9-1.

Expasy ProtParam online tool analysis was used to determine the detailed information about P450s such as predicted molecular weight, theoretical pI and extinction coefficient, which are important for protein expression and purification. For each protein sequence, a list of homologues that share a high level of similarity with its most closely related sequence was achieved using BLAST sequence alignment analysis and the top homologue sequences were chosen and subjected to the Clustal W online alignment tool. Prosite, another online tool was used to identify the motifs and the fingerprint sequences for the four proteins.

In the absence of structural data, a modelling study was performed to produce predicted structures of four P450s from *C. necator* H16 using two online tools; SWISS MODEL and Phyre. SWISS MODEL was used to build the heme domain of all P450 while the later was applied for modelling the reductase domains of B1279 and B1009. Same procedure was followed by using both tools. Firstly, sequences of proteins were submitted to both servers (SWISS MODEL and Phyre2). Next, models with the best hit by each tool was chosen for heme and reductase domains. Then, the quality of the 3D models was verified by using VERIFY 3D and the final models were visualised by using PyMol 2.2 program. The heme of the crystal structure of the templates were used to build the heme cofactor of the four P450s models. For further qualification of the predicted models, the 3D structure of each protein was superimposed with the crystal structure of the templet that used to build this model to determine the match between each model and its template. As well as the overall root mean deviation (RMSD) between each model and its template was calculated using PyMol 2.2. In addition, sequences alignment between models and their templates were

checked by Pairwise ClustalW alignment tool and the hydrophobic residues that bound to the heme in each heme domain were highlighted in the sequence alignment and labelled in the structure pictures. Furthermore, heme domain 3D models that produced by using SWISS MODEL were compared by models suggested by Phyre for the same heme domain in case the templates that suggested by these two servers are different. In order to extend our knowledge of these proteins, a detailed analysis of the intact enzymes was also applied using the SnapGen programme.

Table 9-1: List of software tools for *C. necator* H16 P450s modilling and their functions and links of the websites.

Software tool	Function	Link
BLAST	To compare proteins sequences and determine the closest sequences to a protein	https://blast.ncbi.nlm.nih.gov/Blast.cgi?PROGRAM=blastp&PAGE_TYPE=BlastSearch&LINK_LOC=blasthome
ClustalW	To find sequences alignment	https://embnet.vital-it.ch/software/ClustalW.html
Expasy ProtParam	To determine the protein sequence of unknown nucleotides	http://web.expasy.org/protparam/
Phyre 2	Create prediction of protein structure based on sequence alignment and existing protein crystal structure data	http://www.sbg.bio.ic.ac.uk/~phyre2/html/page.cgi?id=index
PROSITE	To identify fingerprint sequence motifs of proteins	https://prosite.expasy.org/
Protein Data Bank (PDB)	To obtain structure information of the previously crystalized protein	http://www.rcsb.org/pdb/home/home.do
PyMol	3D protein model investigation and imaging	https://www.pymol.org/
SnapGene	To identify plasmid sequence features	http://www.snapgene.com/
SWISS MODEL	Create prediction of protein structure based on sequence alignment and existing protein crystal structure data	https://swissmodel.expasy.org/

9.4 Cloning of *Cupriavidus necator* H16 P450 enzymes

Four genes from *Cupriavidus necator* H16_B2406, H16_B1743, H16_B1279 and H16_B1009, were cloned and transferred into the expression vector pETM11 by Dr Kang Lan Tee using the standard molecular biology techniques of PCR, DNA digestion, ligation, transformation and isolation. The resultant colonies were verified by DNA sequencing. A QIAprep Spin Miniprep kit was used to isolate four high purity plasmids, pETM11-H16-B2406, pETM11-H16-B1743, pETM11-H16-B1279 and pETM11-H16-B1009, from DH5α cells. The concentrations and purities of these plasmids were measured by spectrophotometry.

9.5 Protein expression and purification

Four BL21 (DE3) cultures were grown overnight and each one was transformed with one of the Cupriavidus *necator* H16 P450 vectors using the $CaCl_2$ heat shock method for pETM11-P450 BM-3 transformation (see section 6.5). For *Cupriavidus necator* H16 P450 expression, Traffic Broth AIM was supplemented with 50 μg/ml kanamycin, 1 mM trace elements and 1 mM δ-aminolevulinic acid before inoculation with 1:200 BL21 (DE3) overnight culture and incubated at 30°C for 16–18 hr. Then, the cells were harvested by centrifugation at 4°C and 6000 rpm for 5 min and cell pellets were stored at -20°C. The frozen pellets were thawed in 50 mM potassium phosphate buffer, lysed chemically using lysozyme and analysed for protein expression and solubility by SDS-PAGE electrophoresis. The molecular weights, IPs and the extension coefficients of the four proteins were calculated by the ProtParam

programme and these molecular weights were used to identify the target protein bands in the SDS-PAGE gel.

The optimal purification conditions for P450 BM-3 were applied during the purification of the four P450s from *Cupriavidus necator* H16. Pellets from 50 ml expression culture were lysed physically by sonication, centrifuged, filtrated and loaded onto the affinity column (HisTrap). The column was pre-equilibrated in buffer A1 (50 mM NaH_2PO_4, 200 mM NaCl, pH 8), then the targeted protein was eluted using a gradient of buffer A1 to buffer B1 (buffer A1 plus 250 mM imidazole, pH 8). Fractions with the desired proteins were then pooled and analysed by SDS-PAGE electrophoresis to evaluate protein expression and purity after purification and specify the possibility of moving to another purification step. Protein features and concentrations were checked after the affinity step by spectra measurement at 250–300 nm as described previously.

9.5.1 Expression optimisation of B1279 and B1009

Proteins that were lowly expressed (B1279 and B1009) were optimised under a variety of conditions (growth medium, growth temperature and DE3 lysogen *E. coli* strains BL21 (DE3) and C41 (DE3)). Each protein was expressed in twelve different expression conditions as shown in Figure 9-1. After analysis, the optimal expression conditions were applied to produce 50 ml of B1279 and B1009, and the affinity purification was repeated to compare these results with the previous results before optimisation.

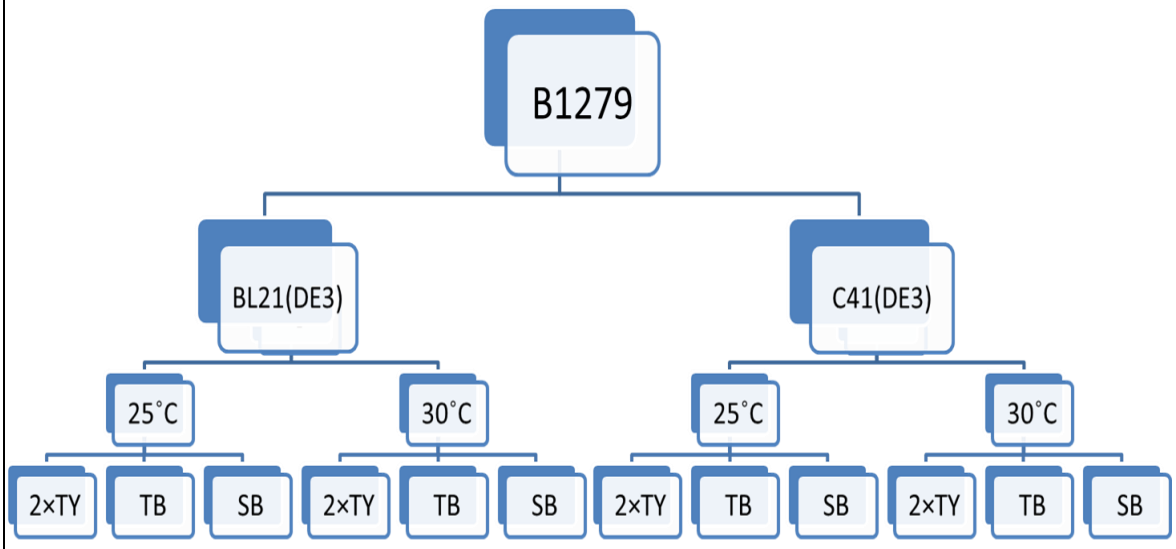


Figure 9-1: The expression optimisation of H16-B1279. Two *E. coli* strains were used; BL21(DE3) and C41(DE3), each strain was used to express P450 protein in three different AIM; 2×TY, TB and SB at two expression tempuratures; 25°C and 30°C.

9.5.1.1 B1009 expression

The expression of B1009 was investigated to obtain a sufficient quantity for functional studies. In addition to previous expression trials which were performed using two *E. coli* strains BL21 (DE3) and C41 (DE3) under a variety of conditions, B1009 expression using *E. coli* HMS174 (DE3) strain was investigated. The isolated plasmid pETM11-H16-B1009 was introduced to pre-cultured HMS174 (DE3) supplemented with 200 µg/ml rifampicin (10 mg/ml stock concentration in 67% methanol and 0.17 N NaOH) and 30 µg/ml kanamycin (30 mg/ml stock concentration) by transformation using $CaCl_2$ with a heat shock method. For protein expression, a single clone from HMS174 (DE3) pETM-11-H16-B1009 was selected and grown overnight in 5 ml of 2×TY media supplemented with 200 µg/ml rifampicin and 30 µg/ml kanamycin. The expression was examined using two different induction methods, auto-induction media (SB) and IPTG. Standard protocols for the expression using both AIM and IPTG were followed in this experiment, with only a deviation in the stock concentration of kanamycin (30 mg/ml was used instead of 50 mg/ml which had been used for the expression using both BL21 (DE3) and C41 (DE3)). The expression was investigated at two expression temperatures, 25°C and 30°C and the results were analysed for concentration, purity and solubility using SDS-PAGE electrophoresis.

9.6 Large-scale expression and purification

The expression was scaled up to 400 ml for three *Cupriavidus necator* H16 P450 enzymes (B2406, B1743 and B1279), which showed high expression in the small-scale production. The optimal expression conditions and the same protocol for the small-scale expression were

performed. Protein concentrations were checked visually depending on the pellet colour and protein bands on the SDS gel. The same affinity purification conditions that were applied in the small-scale purification were used with the large scale. Fractions from the affinity step with the desired proteins were pooled, diluted with buffer A2 (100 mM Tris-HCl, pH 8) and loaded onto the ion exchange column (Q-super) pre-equilibrated with buffer A2. P450 proteins were concentrated and eluted using step elution of buffer B2 (buffer A2 plus 2 M NaCl, pH 8). Concentrated proteins were then subjected to gel filtration to de-salt and separate proteins according to their sizes.

9.7 Spectroscopic measurement

The purified P450 proteins (B2406, B1743 and B1279) were identified optically by scanning the UV-visible spectra between 250 to 800 nm in a 1-cm path length cuvette using an UV-3100PC spectrophotometer from VWR (USA). Protein samples were diluted in buffer C (50 mM Tris-HCl, 1mM EDTA, 10% (v/v) glycerol, pH 7.2) using 3:1 (v/v) as a dilution rate. The spectra obtained were also used to indicate protein purity by calculating the RZ ratio.

9.8 Enzyme concentration estimation

The Beer-Lambert law was used as before (equation 6.1) to calculate the initial concentrations of the *Cupriavidus necator* H16 P450 proteins using the spectra at 280 nm, path length in cm and the extinction coefficient in M^{-1} cm^{-1} at 280 nm.

9.9 Purification optimisation of B1279

Different purification strategies were investigated in an attempt to enhance the concentration of the purified B1279. First, to overcome protein loss during the ion exchange process, this step was omitted so that the purest fractions of the desired protein from the affinity step were then purified by gel filtration. Secondly, gradient elution was used instead of step elution in the ion exchange step to check the purity after this step and decide if it was possible to skip the final step (gel filtration), especially as the protein was diluted extensively at this stage. Also, the concentration of imidazole was increased in the elution buffer in the first purification step (affinity) from 250 mM to 500 mM to increase the quantity of eluted protein from this step. In addition, ultrafiltration of the purified protein after the final purification step (gel filtration) was investigated. Finally, B1279 was expressed on a large scale using C41 (DE3) and purified to compare the results with that expressed in BL21 (DE3).

9.10 Substrate binding

Analysis of interactions of three P450s from *Cupriavidus necator* H16 (B2406, B1743 and B1279) with a range of saturated fatty acids, unsaturated fatty acids, aromatic component and herbicides was applied by a progressive titration in buffer C (50 mM Tris-HCl, 1mM EDTA, 10% (v/v) glycerol, pH 7.2). List of components used in this experiment are shown in Table 9-2. Spectra were scanned between 300–800 nm after the addition of each substrate. Titrations were carried out at RT and substrates were dissolved either in MilliQ water, methanol or DMSO, with the final concentration of all substrates in solution being no more than 1000 µM, except vernolate which was increased up to 2000 µM. Spectra for the titration

of the methanol-P450 and DMSO-P450 complex were assessed for any denaturation in the protein structure as a result of the addition of the solvents.

Table 9-2: Protocol and materials were used in substrate binding experiment.

Component	Negative control for B2406 & B1743 (µl)	Positive control for B2406 & B1743 (µl)	Negative control for B1279 (µl)	Positive control for B1279 (µl)
(1)* Buffer	245	245	95	95
(2)** Substrate I, II	0	0.5, 1, 1.5, 2 then 0.5 µl was added every run	0	0.5, 1, 1.5, 2 then 0.5 µl was added every run
(3)*** Enzyme	750	750	900	900
Total volume	995	995.5, 996, 996.5… 1000	995	995.5, 996, 996.5…. 1000

***Component (1)**: 50 mM Tris/HCl buffer, 1 mM EDTA, pH 7.2.

****Component (2):** I) 0.1 M of: **a)** lauric acid, MW 200.32 g/mole, 0.2 g was dissolved in 10 ml of methanol; **b)** myristic acid, MW 228.38 g/mole, 0.23 g was dissolved in 10 ml of methanol; **c)** palmitic acid, MW 256.42g/mole, 0.26 g was dissolved in 10 ml of methanol **d)** malonic, MW 148.03 g/mole, 0.15 g was dissolved in 10 ml of MilliQ water; **e)** adipic acid, MW 146.14 g/mole, 0.15 g was dissolved in 10 ml of DMSO; **f)** indole, MW 117.15 g/mole, 0.12 g was dissolved in 10 ml of DMSO; **g)** naphthalene, MW 128.17 g/mole, 0.13 g was dissolved in 10 ml of DMSO; **h)** 3-phenoxytoluene, MW 184.24 g/mole, density 1.05 g/ml, 175 µl was dissolved in 10 ml of DMSO; i**)** butyric acid, MW 88.11 g/mole, density 0.96 g/ml, 91 µl was dissolved in 10 ml of MilliQ water; j**)** EPTC, MW 189.32 g/mole, density 0.96 g/ml, 19.7 µl was dissolved in 1 ml of methanol. II) 0.4 M vernolate**,** MW 203.34 g/mole, density 0.954 g/ml, 85 µl was dissolved in 1 ml of methanol.

****Component (3)**: B2406 or B1743 proteins

9.11 Enzymatic assay via NADPH or NADH consumption

Activity of the *Cupriavidus necator* H16-B1279 was tested using the NADPH and NADH consumption assay; this assay technique was explained previously in section 6.13. The total reaction volume was 1 ml and assay was performed at RT. The reaction mixture consisted of four basic components: I) reaction buffer (50 mM Tris/HCl, 1 mM EDTA, pH 7.2); II) substrates, 100 mM stock concentration of all substrates (butyric acid, myristic acid, 3-phenoxytoluene, lauric acid, naphthalene, EPTC, malonic, adipic acid, indole, palmitic acid) was used except for vernolate, 400 mM stock concentration was used; III) B1279 purified protein, 100 nM was used as the total enzyme equivalent concentration, and V) the induction factors either 250 µM NADPH or 200 µM NADH final concentrations were used to start the reaction. These components were mixed according to the order and concentrations shown in Table 9-3.

The induction factors were added 5 seconds after the reaction start. The substrates concentrations were chosen taking into consideration the miscibility of the fatty acid solutions in the reaction buffer. Two different solvents in addition to water were used, methanol and DMSO, and trace amounts of these solvents used to dissolve the substrates were disregarded from calculations.

Table 9-3: Protocol and components used for B1279 activity evaluation

Component	Negative control (µl)	Positive control (µl)	Positive control concentrations (µM)
(1) Activity buffer	840	837.5*	41.90
		835**	41.75
		830***	41.50
(2) Substrate	0	2.5*	250
		5**	500
		10***	1000
(3) Enzyme	60	60	0.1
(4) NADPH/NADH	100	100	250/200
Total volume	1000	1000	

* Myristic acid, palmitic acid or 3- phenoxytoluene.,

**Lauric acid or naphthalene

***Adipic acid, butyric acid, indol, malonic acid, EPTC or vernolate

10 Genetic analysis and structure modelling of putative cytochrome P450s from *Cupriavidus necator* H16

10.1 Identification and sequence alignment of the genome of novel putative P450s

By using Expasy ProtParam analysis, essential information for each P450 expression analysis and purification was determined. The detailed information of P450-H16-B2406, P450-H16-B1743, P450-H16-B1279 and P450-H16-B1009 are presented in Table 10-1.

Table 10-1: Expasy ProtParam analysis for recombinant B2406, B1743, B1279 and B1009.

	B2406	**B1743**	**B1279**	**B1009**
Number of amino acids	405	429	810	1126
Molecular weight (kDa)	42.13	48.15	90	123
Total negatively charged residues (Asp + Glu)	37	61	110	126
Total positively charged residues (Arg + Lys)	30	47	85	112
Extinction coefficient (M^{-1} cm^{-1} at 280 nm measured in water)	42315	55015	95310	127365
Theoretical pI	6.08	5.90	5.75	6.15
Estimated half-life	30 hours (mammalian reticulocytes, in vitro). >20 hours (yeast, in vivo). >10 hours (*Escherichia coli*, in vivo).			

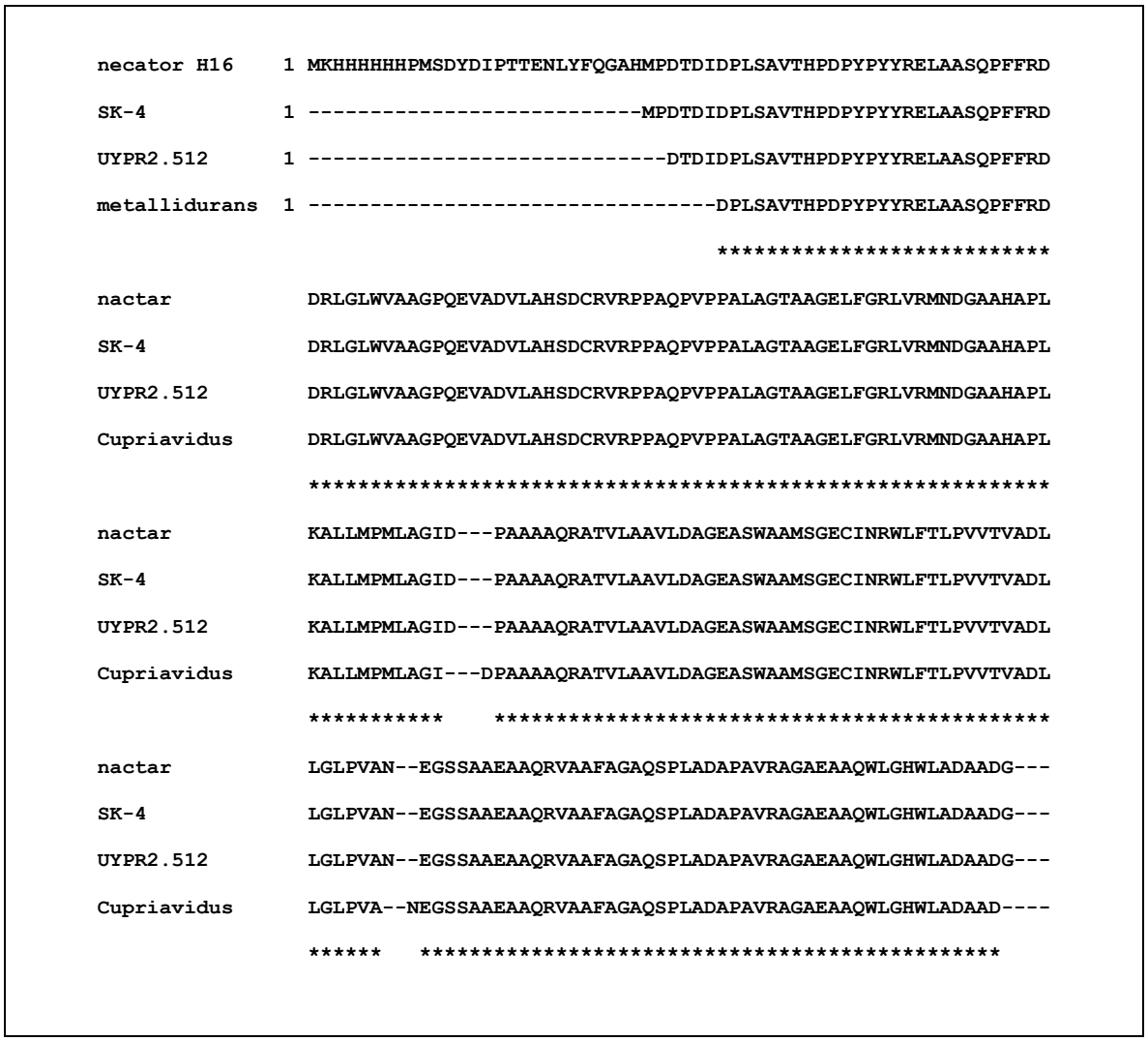
BLAST analysis gave many homologous sequences to all P450s from *Cupriavidus necator* H16. The top two closest relatives in BLAST were selected and subjected to ClustalW tool analysis for sequence alignment of B2406, B1743, B1279 and B1009 (Figure 10-1 to Figure 10-4). Furthermore, the CYP 116 B1 sequence was also compared with the four P450 sequences. This protein showed 82%, 75% and 55% identity with B1279, B1009 and B2406 respectively and is the only protein from *Cupriavidus sp.* that has been characterised (Warman et al., 2012).

The motifs and the fingerprint sequences of four P450 sequences were identified using PROSITE, the online tool, to achieve a better understanding of their structure and activity. By using this tool, (FGHGRHA**C**PG), (FGVGVHR**C**LG), (FGYGSHQ**C**MG) and (FGNGERA**C**IG) in the N-terminal of B2406, B1743, B1279 and B1009 respectively, matched the PROSITE consensus motif (FGXGXXXCXG) which represents the cytochrome P450 cysteine heme iron ligand and the cysteine residue is the heme iron fifth ligand (Minerdi et al., 2015).

The C-terminal region of P450-H16-B1279 showed sequence homology to reductase proteins as it contains three basic functional parts an FMN binding domain, a NADH binding domain and a ferredoxin-like [2Fe2S] domain. Sequence SRGGS of B1279 corresponded to the PROSITE consensus FMN binding motif [G/S]RGGS, while GIGITP corresponds to the consensus motif GXGXXP for NADH binding. Another signature sequence is represented by the [2Fe2S] ferredoxin motif CXX[G/A]XC[G/A/S/T]XC which is represented as **C**EEGL**C**GS**C** in this protein (Minerdi et al., 2015). Three of the four cysteines involved in binding the iron-sulphur cluster are located in the [2Fe2S] motif, while the fourth one is residue CYS 792. The results from the sequence analysis of B1279 showed that this protein

is a catalytically self-sufficient cytochrome P450 enzyme composed of a heme and a reductase domain.

B1009 sequence analysis via PROSITE indicated that this protein is also a self-sufficient cytochrome P450 enzyme because of the presence of both a heme domain and reductase domain. Cys463 represents the fifth cysteine heme ligand, while the reductase domain consists of a flavodoxin-like domain (FLAVODOXIN_LIKE) and ferredoxin reductase-type FAD binding domain (FAD_FR).



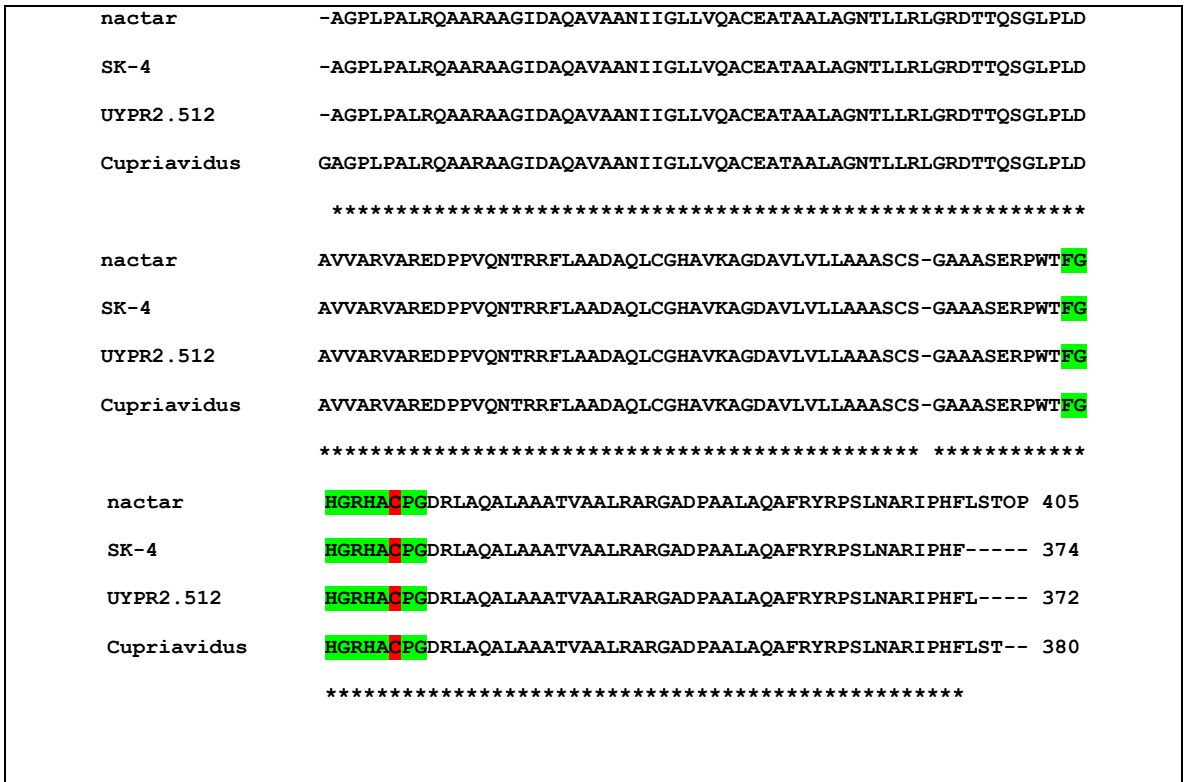
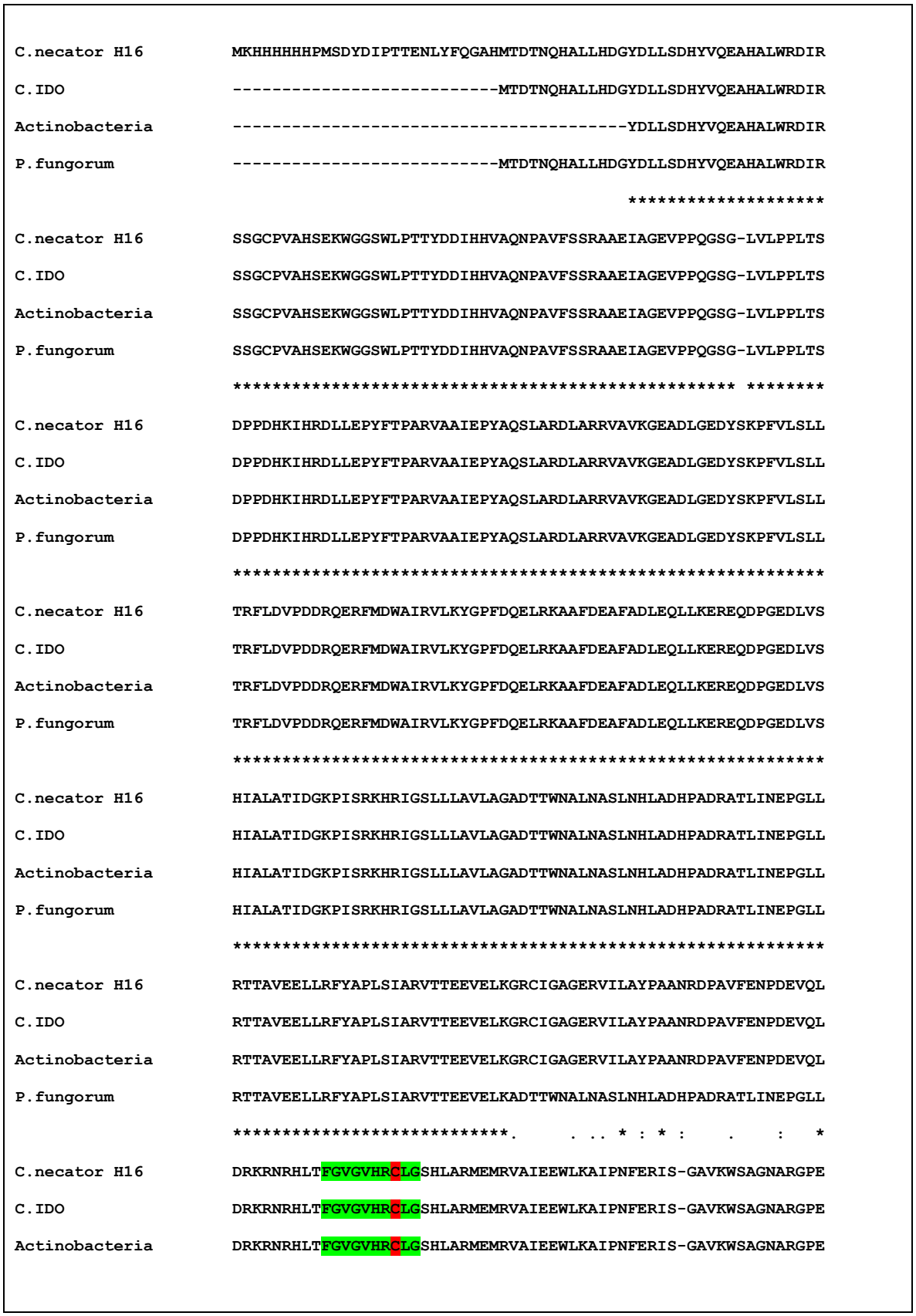


Figure 10-1: Amino acid sequence alignment of P450-H16_B2406 with other P450s from *Cupriavidus sp.* The key residues are marked as following: the conserved heme in green, the fifth ligand of heme (cysteine) in red. (*) indicating similar identity, (.) is for strongly similar residues and (:) to indicate weak similarity. The alignment was produced by using data from BLAST and ClustalW online tool.


C.necator H16 MKHHHHHHPMSDYDIPTTENLYFQGAHMTDTNQHALLHDGYDLLSDHYVQEAHALWRDIR
C.IDO ----------MTDTNQHALLHDGYDLLSDHYVQEAHALWRDIR
Actinobacteria ----------YDLLSDHYVQEAHALWRDIR
P.fungorum ----------MTDTNQHALLHDGYDLLSDHYVQEAHALWRDIR
****
C.necator H16 SSGCPVAHSEKWGGSWLPTTYDDIHHVAQNPAVFSSRAAEIAGEVPPQGSG-LVLPPLTS
C.IDO SSGCPVAHSEKWGGSWLPTTYDDIHHVAQNPAVFSSRAAEIAGEVPPQGSG-LVLPPLTS
Actinobacteria SSGCPVAHSEKWGGSWLPTTYDDIHHVAQNPAVFSSRAAEIAGEVPPQGSG-LVLPPLTS
P.fungorum SSGCPVAHSEKWGGSWLPTTYDDIHHVAQNPAVFSSRAAEIAGEVPPQGSG-LVLPPLTS
********** ********
C.necator H16 DPPDHKIHRDLLEPYFTPARVAAIEPYAQSLARDLARRVAVKGEADLGEDYSKPFVLSLL
C.IDO DPPDHKIHRDLLEPYFTPARVAAIEPYAQSLARDLARRVAVKGEADLGEDYSKPFVLSLL
Actinobacteria DPPDHKIHRDLLEPYFTPARVAAIEPYAQSLARDLARRVAVKGEADLGEDYSKPFVLSLL
P.fungorum DPPDHKIHRDLLEPYFTPARVAAIEPYAQSLARDLARRVAVKGEADLGEDYSKPFVLSLL
****
C.necator H16 TRFLDVPDDRQERFMDWAIRVLKYGPFDQELRKAAFDEAFADLEQLLKEREQDPGEDLVS
C.IDO TRFLDVPDDRQERFMDWAIRVLKYGPFDQELRKAAFDEAFADLEQLLKEREQDPGEDLVS
Actinobacteria TRFLDVPDDRQERFMDWAIRVLKYGPFDQELRKAAFDEAFADLEQLLKEREQDPGEDLVS
P.fungorum TRFLDVPDDRQERFMDWAIRVLKYGPFDQELRKAAFDEAFADLEQLLKEREQDPGEDLVS
****
C.necator H16 HIALATIDGKPISRKHRIGSLLLAVLAGADTTWNALNASLNHLADHPADRATLINEPGLL
C.IDO HIALATIDGKPISRKHRIGSLLLAVLAGADTTWNALNASLNHLADHPADRATLINEPGLL
Actinobacteria HIALATIDGKPISRKHRIGSLLLAVLAGADTTWNALNASLNHLADHPADRATLINEPGLL
P.fungorum HIALATIDGKPISRKHRIGSLLLAVLAGADTTWNALNASLNHLADHPADRATLINEPGLL
****
C.necator H16 RTTAVEELLRFYAPLSIARVTTEEVELKGRCIGAGERVILAYPAANRDPAVFENPDEVQL
C.IDO RTTAVEELLRFYAPLSIARVTTEEVELKGRCIGAGERVILAYPAANRDPAVFENPDEVQL
Actinobacteria RTTAVEELLRFYAPLSIARVTTEEVELKGRCIGAGERVILAYPAANRDPAVFENPDEVQL
P.fungorum RTTAVEELLRFYAPLSIARVTTEEVELKADTTWNALNASLNHLADHPADRATLINEPGLL
**********. . .. * : * : . : *
C.necator H16 DRKRNRHLTFGVGVHRCLGSHLARMEMRVAIEEWLKAIPNFERIS-GAVKWSAGNARGPE
C.IDO DRKRNRHLTFGVGVHRCLGSHLARMEMRVAIEEWLKAIPNFERIS-GAVKWSAGNARGPE
Actinobacteria DRKRNRHLTFGVGVHRCLGSHLARMEMRVAIEEWLKAIPNFERIS-GAVKWSAGNARGPE

```
P.fungorum              RTTAVEELLRFYAPLSIARVTTEEVELKVAIEEWLKAIPNFE-RISGAVKWSAGNARGPE
                          .  ..*    .          .:*::**************    **************
C.necator H16           NVRIRVVSTOP
C.IDO                   NVRIRVV----
Actinobacteria          NVRIR------
P.fungorum              NVRIRV-----
                        *****
```

Figure 10-2: Amino acid sequence alignment of P450-H16_B1743 with other P450s from *Cupriavidus sp.* The key residues are marked as following: the conserved heme in green, the fifth ligand of heme (cysteine) in red. (*) indicating similar identity, (.) is for strongly similar residues and (:) to indicate weak similarity. The alignment was produced by using data from BLAST and ClustalW online tool.

```
C.necator           MKHHHHHHPMSDYDIPTTENLYFQGAMANPSPSAARSGCPIDHSALTAPNGCPVSHNAAQ
F.nantongensis      ---------------------------MTISATGAARGCPIDHSTLAAPNGCPVSHNAAQ
B.UYPR2.512         --------------------------MANPSPSAARSGCPIDHSALTAPNGCPVSHNAAQ
C.metallidurans     --------------------------MPQTNAPASSGSCPIDHSALRAPNGCPVSHQAAA
                                               .  :   *****.:: *********:**

C.necator           FDPFGDGYQQDPPEYVRWSREQEPVFYSPRLGYWVVTRYEDIKAIFRDNLTFSPSIALEK
F.nantongensis      FDPFGDGYQQDPPEYVRWSREQEPVFYSPQLGYWVVTRYDDIKAIFRDNLTFSPSIALEK
B.UYPR2.512         FDPFGDGYQQDPPEYVRWSREQEPVFFSPRLGYWVVTRYQDIKAIFRDNLTFSPSIALEK
C.metallidurans     FDPFEDGYQQDPPEYVRWSRAQEPVFYSPKLGYWVVTRYDDIKAIFRDNITFSPSIALEK
                    **** *************** *****:**:*********:*********:**********

C.necator           ITPTGDEANAVLASYGYAMNRTLVNEDEPAHMPRRRALMAPFTPAELAHHEPLVRRLARE
F.nantongensis      ITPTGDEANAVLASYGYAMNRTLVNEDEPAHMPRRRALMAPFTPAELAHHEPLVRRLTRE
B.UYPR2.512         ITPTGDEANAVLASYGYAMNRTLVNEDEPAHMPRRRALMAPFTPAELAHHEPLVRRLARE
C.metallidurans     ITPTGEAANAVLASYGYAMNRTLVNEDEPAHMPRRRALMEPFTPAELAHHEPMVRKLTRE
                    *****: ******************************** ************:**:*:**

C.necator           YVDRFIDDGRADLVDQMLWEVPLTVALHFLGVPEEDMDLLRQYSIAHTVNTWGRPKPEEQ
F.nantongensis      YVDRFIDDGRADLVDQMLWEVPLTVALHFLGVPEEDMDLLRQYSIAHTVNTWGRPKPEEQ
B.UYPR2.512         YVDRFIDDGRADLVDQMLWEVPLTVALHFLGVPEEDMDLLRQYSIAHTVNTWGRPKPEEQ


C.metallidurans     YVDRFIDNGRADLVDEMLWEVPLTVALHFLGVPEEDMDLLRQYSIAHTVNTWGRPKPEEQ
                    *******:*******:*********************************************

C.necator           VAVAHAVGNFWQLAGRILDKMREDPSGPGWMQYGLRKQKELPDVVTDSYLHSMMMAGIVA
F.nantongensis      VAVAHAVGNFWQLAGRILDKMREDPSGPGWMQYGLRKQKELPEVVTDSYLHSMMMAGIVA
B.UYPR2.512         VAVAHAVGNFWQLAGRILDKMREDPSGPGWMQYGLRKQKDLPEVVTDSYLHSMMMAGIVA
C.metallidurans     VAVAHAVGNFWQLAGRILDKMREDPSGPGWMQYGLRKQRELPEVVTDSYLHSMMMAGIVA
                    **********************:**************::**:*****************

C.necator           AHETTANASANAIKLLLQHPDAWREICDDPSLIPNAVEECLRHNGSVAAWRRLVTRDAEV
F.nantongensis      AHETTANASANAIKLLLQHPDAWRELCEDPALIPNAVEECLRHNGSVAAWRRLVTRDAEV
B.UYPR2.512         AHETTANASANAVKLLLQHPDAWREICDDPSLIPNAVEECLRHNGSVAAWRRLVTRDAEV
C.metallidurans     AHETTANASANAIKLLLQHPDVWREICEDPALIPNAVEECLRHNGSVAAWRRLVTRDTEV
                    ************:********.***:*:**:**************************:**

C.necator           GGIRLPAGSKLLIVTSSANHDERHFADADLFDIRRDNASEQLTFGYGSHQCMGKNLARME
F.nantongensis      GGIRLPAGSKLLIVTSSANHDERHFADADLFDIRRDNASEQLTFGYGSHQCMGKNLARME
B.UYPR2.512         GGIRLPAGSKLLIVTSSANHDERHFADADLFDIRRDNASEQLTFGYGSHQCMGKNLARME
```

```
C.metallidurans      GGMSLAVGSKLLIVTSSANHDEHHFADADLFDIHRDNASDQLTFGYGSHQCMGKNLARME
                     **: *..***************:**********:*****:********************
E C.necator            MQVFLEELTRRLPHMRLAEQTFTYVPNTSFRGPEHLLVEWDPAQNPERRDPALLEVHQPV
F.nantongensis       MQVFLEELTRRLPHMRLAEQTFTYVPNTSFRGPEHLLVEWDPAQNPERRDPALLAVRQPV
B.UYPR2.512          MQVFLEELTRRLPHMRLAGQAFTYVPNTSFRGPEHLLVEWDPAQNPERRDPAVLDVRQPV
C.metallidurans      MQIFLEELTSRLPHMRLAAQRFTYVPNTSFRGPEHLWVEWEPARNPERTDPTVLAPRDAV
                     **:****** ******** * *************** ***:**:**** :*::*  ::.*
C.necator            RIGEPSAHTIARTVVVERVTPAADGVVRLRLAAPDGKPLPRWAPGSHIDVECGDTGLSRQ
F.nantongensis       RIGEPSAHTIARTVLVERATPAADGVLRLRLVAPDGKPLPRWAPGSHIDVECGDTGLSRQ
B.UYPR2.512          RIGEPSAHTIARTVVIERVTPAADGVVRLRLAAPDGKPLPRWAPGSHIDVECGDTGLSRQ
C.metallidurans      RIGEPTGGTTGRTLVVERVETAAEGVVRIRLVSPDGRALPRWSPGSHIDVECGHTGISRQ
                     *****:. * .*::::**. .**:**:*:**.:***:.****:**********.**:***
C.necator            YSLCGDPDDTGALEIAVLRDPDSRGGSAWVHGSVQAGDRLRIRGPRNHFRFDEGCGRAIF
F.nantongensis       YSLCGDPDDAAALEIAVLRDPASRGGSAWVHGSVRAGDRLRIRGPRNHFRFDEQCGRAIF
B.UYPR2.512          YSLCGDPDDAAALEIAVLRDPDSRGGSAWVHGSVRAGDRLRIRGPRNHFRFDESCKRAIF
C.metallidurans      YSLCGDPADTGAFEIAVLREPESRGGSAWIHASLHAGDKLKVRGPRNHFRLDETCRRAIF
                     ******* *:.*:*****::* *******:*.*:.***:*::********:** * ****
C.necator            IAGGIGITPVSAMARRARALGIDYEFHYCGRSRQAMAMLDELQALHGARLHVHASDEGQR
F.nantongensis       IAGGIGVTPVSAMARRARALGIDYTFHYCGRSRQAMAMLDELQALHGTRLQVHASDEGRR
B.UYPR2.512          IAGGIGITPVSAMARRARALGIDYEFHYCGRSRKAMAMPDELQALHGARLQVHASDEGQR
C.metallidurans      IAGGIGVTPVSAMARRAKELGVDYTFHYCGRSRASMAMIDELRALHGDRVRIHAADEGQR
                     ******:**********: **:** ******** :*** ***:**** *:::**:***:*
C.necator            ADFGKLLARPDPRTQIYACGPQRMLDALAESCAAWPEDALRVEHFVSKLGTLDASKELPF
F.nantongensis       ADFGKLLGQPDADTQIYACGPQRLLDALAECCAAWPAEALRVEHFVSRLGTLDASKEQPF
B.UYPR2.512          ADFGKLLAQPDPRTQIYACGPQRMLDALAECCAAWPEDALRVEHFVSRLGTLDASQEQPF
C.metallidurans      ADLAQVLGAPDANAQIYACGPARMVEALEALCAAWPEDSLRVEHFSSQLGTLDPSREQPF
                     **:.::*. **. :******* *:::**   ***** ::****** *:*****.*:* **
C.necator            SVELKDSGLVMEVPAGQTLLSALRGANIDVQSDCEEGLCGSCEVRVLAGQVDHRDVVLTR
F.nantongensis       TVELKDSGLVLEVPAGQTLLGALRGANIDVQSDCEEGLCGSCEVRVLAGQVDHRDVVLTR
B.UYPR2.512          TVELKDSGLVMEVPAGQTLLSALRSANIDVQSDCEEGLCGSCEVRVLAGQVDHRDVVLTR
C.metallidurans      TVELKDSGLTLEVPPDQTLLATLRAANIDVQSDCEEGLCGSCEVRVLAGEIDHRDVVLTR
                     :********.:***..****.:**.************************::*********
C.necator            SEREANQRMMACCSRACGGGRLVLELSTOP


F.nantongensis       AERDANHRMMACCSRACGGGRLVLEL----
```

```
B.UYPR2.512           SEREANQRMMACCSRACGGGRLVLEL----

C.metallidurans       GEREANNRMMACCSRAAKGGKIVLGL----

                      .**:**:*********. **::** *
```

Figure 10-3: Amino acid sequence alignment of P450-H16_B1279 with other P450s from *Cupriavidus sp.* The key residues are marked as following: the conserved heme in green, the fifth ligand of heme (cysteine) in red, the FMN binding motif in pink, the NADH binding motif in gray and the 2Fe-2S cluster in yellow. (*) indicating similar identity, (.) is for strongly similar residues and (:) to indicate weak similarity. The alignment was produced by using data from BLAST and ClustalW online tool.

```
C.necator           MKHHHHHHPMSDYDIPTTENLYFQGAHMPPPIELSSPDQASDAPHQAPARSMHAPIPEPI
E.GA3-3             ----------------------------MPPPIELSSPDQASDAPHQTPARSIHATIPEPI
D.SK-4              ----------------------------MPPPIELSSPDQASDAPHQTPARSMHAPIPEPI
C.metallidurans     -------------------------------------------------MSTATPAAALEPI
                                                                     :     .  ***
C.necator           PRDPGWPLVGNLLQITPGALGQHLLARSRHHDGIFELNFAGRRVPFVTSVALASELCDAA
E.GA3-3             PRDPGWPLVGNLLQITPGALGQHLLARSRHHDGIFELNFAGRRVPFVTSVALASELCDAA
D.SK-4              PRDPGWPLVGNLLQITPGALGQHLLARSRHHDGIFELNFAGRRVPFVTSVALASELCDAA
C.metallidurans     PRDPGWPIFGNLFQITPGEVGQHLLARSRHHDGIFELDFAGKRVPFVSSVALASELCDAT
                    * *** *:.***:**.** :*****************:***:*****:*****:*:***:
C.necator           QFRKYIGPPVSYLRGMAGDGLFTARSDEANWGKAHRILMPAFSQRAMKGYFDVMLRVANR
E.GA3-3             QFRKYIGPPLSYLRGMAGDGLFTARSDEANWGKAHRILMPAFSQRAMKGYFDVMLRVANR
D.SK-4              QFRKYIGPPLSYLRGMAGDGLFTARSDEANWGKAHRILMPAFSQRAMKGYFDVMLRVANR
C.metallidurans     RFRKIIGPPLSYLRDMAGDGLFTAHSDEPNWGCAHRILMPAFSQRAMKAYFDVMLRVANR
                    :*** ****:****.*********:***.*** ***************.***********
C.necator           LVDKWDQQGPDADIAVADDMTRLTLDTIALSGFGYDFESFASTELHPFIEAMVGALEEAM
E.GA3-3             LVDKWDQQGPDADIAVADDMTRLTLDTIALSGFGYDFESFASAELHPFIEAMVGALEEAM
D.SK-4              LVDKWDRQGPDADIAVADDMTRLTLDTIALSGFGYDFESFASAELHPFIEAMVGALEEAM
C.metallidurans     LVDKWDRQGPDADIAVADDMTRLTLDTIALAGFGYDFASFASDELDPFVVAMVGALGEAM
                    ******:************************:****** ***  :*.**: ****** ***
C.necator           SKLTRFALQDRFMHAAHQKFDQDTRFMRDLVDDVIRRRRAGDAAERPGGTANDLLGLMLE
E.GA3-3             SKLTRFALQDRFMHAAHQKFDQDTRFMRDLVDDVIRRRRAGDAAERPGGTANDLLGLMLE
D.SK-4              SKLTRFALQDRFMHAAHQKFDQDTRFMRDLVDDVIRRRRAGDAAERRGGTANDLLGLMLE
C.metallidurans     QKLTRLPIQDRFMGRAHRQAAEDIAYMRNLVDDVIRQRRVSPTSG------MDLLNLMLE
                     ****:.:*****  **::  :*  :**:*******:**.. :.        ***.****
C.necator           ARDPDTDQRLDDENIRNQVITFLIAGHETTSGLLTFALYELLRNPGVMAQAYAEVDAVLP
E.GA3-3             ARDPDTDQRLDDENIRNQVITFLIAGHETTSGLLTFALYELLRNPGVMAQAYAEVDAVLP
D.SK-4              ARDPDTDQRLDDENIRNQVITFLIAGHETTSGLLTFALYELLRNPGVMAQAYAEVDAVLP
C.metallidurans     ARDPETDRRLDDANIRNQVITFLIAGHETTSGLLTFALYELLRNPGVLAQAYAEVDTVLP
                    ****:*.:**** *********************************:********:***
C.necator           GDAAPVYADLARLPVLDRVLKETLRLWPTAPAFAVAPFEDTLLGGRYLIRKDRRLSVVLT
E.GA3-3             GDAAPVYADLARLPVLDRVLKETLRLWPTAPAFAVAPFEDTLLGGRYLIRKDRRLSVVLT
D.SK-4              GDAAPVYADLARLPVLDRVLKETLRLWPTAPAFAVAPFEDTLLGGRYLIRKDRRLSVVLT
C.metallidurans     GDAPPVYADLARMPVLDRVLKETLRLWPTAPAFAVAPFDDVVLGGRYRLRKDRRISVVLT
```


C.necator ALHRDPKVWADPERFDIDRFLPEQEAKLPRHAYMPFGNGERACIGRQFALTEAKLALALM
E.GA3-3 ALHRDPKVWADPERFDIDRFLPEQEAKLPRHAYMPFGNGERACIGRQFALTEAKLALALM
D.SK-4 ALHRDPKVWADPERFDIDRFLPEQEAKLPRHAYMPFGNGERACIGRQFALTEAKLALALM
C.metallidurans ALHRDPKVWANPERFDIDRFLPENEAKLPAHAYMPFGQGERACIGRQFALTEAKLALALM
C.necator LRNFQFTDAHDYQFRIKETLTIKPDGFTVRARRRRPHERIAAAPLC-TAQAPRAGPDVQG
E.GA3-3 LRNFQFTDAHDYQFRIKETLTIKPDGFTVRARRRRPHERIAAVPLAGTAQAPRAGPDVQG
D.SK-4 LRNFQFTDAHDYQFRIKETLTIKPDGFTVRARRRRPHERIAAAPLAGTAQAPQAGPDVQG
C.metallidurans LRNFAFQDPHDYQFRLKETLTIKPDQFVLRVRRRRPHERFVTQQAS-QAVADAAQTDVRG
C.necator RGRTLAVLCGSSLGTARELAEQVHAGALAAGFDATLRDLDDVADALPTTGLAVIIAATYN
E.GA3-3 RGRTLAVLCGSSLGTARELAEQVHAGALAAGFDATLRDLDDVADALPTTGLAVIIAATYN
D.SK-4 GGRPLAVLCGSSLGTARELAEQVHAGALAAGFDATLRDLDDVADALPTTGLAVIIAATYN
C.metallidurans HGQAMTVLCASSLGTARELAEQIHAGAIAARFDAKLADLDDAVGALPTSGLAVVVAATYN
C.necator GRAPDSARRLEALLDSGAASGYRAEALSVAVLGCGNSQWPTYQAFPRRVYEFFTNAGASP
E.GA3-3 GRAPDSARRLEALLDSGAASGYRAEALSVAVLGCGNSQWPTYQAFPRRVYEFFTNAGASP
D.SK-4 GRAPDSARRLEALLDSGAASGYRAEALSVAVLGCGNSQWPTYQAFPRRVYEFFTNAGASP
C.metallidurans GRAPDSGRKFEAMLDADDASGYRANGVRLALLGCGNSQWATYQAFPRRVFDFFITAGAVP
C.necator LLPRGEADGNGDFDQAVERWLALLWQALQAGDGVGGPAHGTPPVRVQVKDVAEIRASTLP
E.GA3-3 LLPRGEADGNGDFDQAAERWLALLWQALQAGDGVGGPAHGTPPVRVQVKDVAEIRASTLP
D.SK-4 LLPRGEADGNGDFDQAAERWLALLWQALQAGGGVGGPAHGTPPVRVQIRDVAEIRASTLP
C.metallidurans LLPRGEADGNGDFDQAAERWLAQLWQALQADGADTGGLG----VDVQVRSMAAIRAETLP
C.necator ANTEAFTVLANTELVNDPSGLWDFSHEAPRTSTRDIRLRLPDGVRYATGDHLAVYPQNQP
E.GA3-3 ANTEAFTILANAELVNDPSGLWDFSREAPRTSTRDIRLRLPDGVRYATGDHLAVYPQNQP
D.SK-4 ANTQGFTVLANTELVNDPSGLWDFSREAPRTSTRDIRLRLPEGVRYATGDHLAVYPQNQP
C.metallidurans AGTQAFTVLSNDELVGDPSGLWDFSIEAPRTSTRDIRLQLPPGITYRTGDHIAVWPQNDA
C.necator GMVAALCERIGIDPDAIVTLSASGGAARGLPLGEALSVRQLLTHFVELQDVVSRHTLRLL
E.GA3-3 GMVAALCERIGIDPDAIVTLSASGGAARGLPLGEALSVRQLLTHFVELQDVVSRHTLRLL
D.SK-4 GMVAALCERIGIDPEAIVTLSASGGAARGLPLGEALSVRQLLTHFVELQDVVSRHTLRLL
C.metallidurans QLVSDLCERLDLDPDAQATISAPHGMGRGLPIDQSLPVRQLLTHFIELQDVVSRQTLRAL

```
                      :*: ****:.:**:* .*:**. * .****:.::*..*******:********:*** *

C.necator             SQSSRCPVTRQALRQLASDDAGSGYAAQVAPRRLGLADVLDRFPAIEVDLAGLLACTVPM

E.GA3-3               SQSSRCPATRQALRQLASDDAGSGYAAQVAPRRLGLADVLDRFPAIEMDLAGLLACTVPM

D.SK-4                SQSSRCPVTRQALQQLASDDAGSGYAAQVAPRRLGLADVLDRFPAIEVDLAGLLACTVPM

C.metallidurans       AQATRCPFTKQSIEQLASDDAEHGYATKVVARRLGILDVLVEHPAIALTLQELLACTVPM

                      :* :*** *:*::.*****:*  **:::*..****: *** ..*** : *  ****:***

C.necator             RPRFYSIASSPLVSPGVATITVGTVWSPALSGRGLFRGVASTWLQGLAPGAMVAGAIRTP

E.GA3-3               RPRFYSIASSPLVSPDVATITVGTVWSPALSGRGLFRGVASTWLQGLAPGAMVAGAIRTP

D.SK-4                RPRFYSIASSPLVSPDVATITVGTVWSPALSGRGLFRGVASTWLQSLAPGAAVAGAIRTP

C.metallidurans       RPRLYSIASSPLVSPDVATLLVGTVCAPALSGRGQFRGVASTWLQHLPIGARVSASIRTP

                      ***:***********.***: ** * :.****:* ********** *. ** *:.:****

C.necator             NPTFAPAADPATPMILIGPGTGIAPFRGFLEERAAQQAAGQPVAPVQLYYGCRHPAHDWL

E.GA3-3               NPTFAPAADPATPMILIGPGTGIAPFRGFLEERAAQQAAGQPMAPVQLYYGCRHPAHDWL

D.SK-4                NPTFAPAADPATPMILIGPGTGIAPFRGFLEERAAQQAAGQPEAPVQLYYGCRHPAHDWL

C.metallidurans       NPPFAPDPDPTAPMLLIGPGTGIAPFRGFLEERALRKMAGNAVTPAQLYFGCRHPQHDWL

                      **.*** .**::**:****************** :: **:. :*.***:***** ****

C.necator             YRNNVERWQAQGVVQVHLACSVVDGEPRYVQDLLWQRRADVWARLNQGAIIYVCGDGRHM

E.GA3-3               YRNDVERWQAQGVVQVHLACSVVGGEPRYVQDLLWQRRADVWARLNQGAIIYVCGDGRHM

D.SK-4                YRDDVERWQAQGVVQVHLACSVVDGEPRYVQDLLWHRRADVWARLNQGAMLYVCGDGRHM

C.metallidurans       YREDIERWAGQGVVEVHPAYSVVPDAPRYVQDLLWQRREQVWAQVRDGATIYVCGDGRRM

                      **:::*** .***.::* * *** . *********:** :**::: :** :*******:*

C.necator             APAVRQVLIQIGAEQGGMTPEAASDWLADLVSAGRYRQDVFNSTOP

E.GA3-3               APAVRQVLIQIGAEQGGMTAEAASDWLAELVSAGRYRQDVFN----

D.SK-4                APAVRQVLIQIGAEQGGMTPEAASDWLAELVSAGRYRQDVFN----

C.metallidurans       APAVRQTLIDIGKAQGGMTDEAASDWFGGLVAQGRYRQDVFN----

                      ******.*::**  ***** ******:. **: *********
```

Figure 10-4: Amino acid sequence alignment of P450-H16_B1009 with other P450s from *Cupriavidus sp.* The key residues are marked as following: the conserved heme in green, the fifth ligand of heme (cysteine) in red, FMN in pink and FAD in yellow. (*) indicating similar identity, (.) is for strongly similar residues and (:) to indicate weak similarity. The alignment was produced by using data from BLAST and ClustalW online tool.

10.2 P450 structure modelling

The online tool SWISS-MODEL was used to predict 3D protein structures for B2406, B1743, B1279 and B1009 heme domains. This web service is designed to provide protein structure homology modelling at different levels of complexity. Building a protein homology model by SWISS-MODEL can be achieved by four basic steps; I) structural templates selection II) sequences alignment between targeted model protein and proteins of the templates III) model building depending on the alignment IV) model qualification (Schwede et al., 2003).

A list of 50 templates were suggested by SWISS-MODEL for each protein depending on sequences alignment and structures similarity. These models were visualised by using PyMol, this molecular graphics system able to graphically generate proteins structures in three dimensions and produce high-quality 3D images of biological molecules, such as proteins (Sevrioukova et al., 1999). The 3D predicted structures of four *C. necator* H16 heme domains were displayed in figures (Figure 10-5 A to Figure 10-7 A and Figure 10-9 A). Generally, all heme domain structures displayed typical P450 folds with expected helical features (α helices, β sheets and loops). Structures showed that these four proteins are probably monomers, folded into α/β domains characteristic of P450s. The heme prosthetic group of each protein appeared at the centre of the heme domain structure (shown as a stickin all heam domain structures). The heme cofactor was obtained from the crystal structure of the template of each protein.

The predicted structures of models were superimposed and compared with the crystal structures of their templates. The overlapping by using PyMol was displayed in figures (Figure 10-5 B to Figure 10-7 B and Figure 10-9 B). From the way that the secondary

structure elements of protein folding shaped, it can be concluded that all heme domains are very similar to their templets structures except minor differences that were found in the loop, N-terminal regions and C-terminal and these limited variations in the structure between P450s members are a known phenomenon because these regions are highly variable in the primary sequence (Sirim et al., 2010).

The heme cofactors of all P450s likely to be bound to highly conserved hydrophobic residues that were labelled in (Figure 10-5 C to Figure 10-7 C and Figure 10-9 C). Some of the models binding residues different from the templates once but they occupied the same position and have very similar functions to each other. The templates and the models details and PDB codes were displayed in Table 10-2. As well as the binding residues of templates and models were shown in Table 10-3. In addition, sequence alignment between P450s primary structures and their templets showed highly conserved binding residues (highlighted in yellow in Appendix I). Sequences alignments were applied by using ClustalW online tool, ClustalW is an alignment tool run with several sets of starting parameters and in each case, the alignments applied according to an objective function. ClustalW considers the individual weight of the sequence and residue-specific gaps and positions (Roccatano, 2015)

Templates that were offered by SWISS-MODEL (50 templates) covered only the heme domain of P450s, so another program was used to model the reductase domain of B1279 and B1009, Phyre the modelling online tool was chosen for reductase domain modelling because it provides an extended list of the suggested models (120 templates including both the heme and the reductase domain).The Phyre2 server predicts the three-dimensional structure of a protein sequence using the techniques of homology modelling, it is provides advanced interface and fully updated fold library (Bennett et al., 2008). The 3D structures of the

reductases domains of B1279 and B1009 were shown in Figure 10-8 A and Figure 10-11 A and the results of superimposing the models to their templets were shown in Figure 10-8 A and Figure 10-11 B. Two ligands can be recognised in the structure of reductase domain of B1279; the FMN and the iron-sulphur binding motifs (can be seen as sticks in Figure 10-8). For B1009 reductase structure three ligands were noticed in the structure, these ligands are; FAD, FMN and (nicotinamide-adenine-dinucleotide phosphate) NADP binding motifs (these ligands can be seen as sticks Figure 10-11).

The resulting models were qualified by using VERIFY 3D online tool and a final model was obtained. VERIFY 3D is an online tool measures the compatibility of a protein model (3D) with its sequence (1D). The position of each residue in the 3D structure is identified by its environment (alpha, beta, loop, polar, apolar etc). This profile is represented by the average 3D-1D score for each residue in the protein structure. According to VERIFY 3D a model will considered successful and it passes the test if 80% or more of protein's residues show average (3D-1D) $\geq$ 0.2 (Lüthy et al., 1992). In general, all models that have been shown in this chapter passed the qualification evaluation applied by VERIFY 3D test with more than 80% of residues scored average (3D-1D) $\geq$ 0.2 as can be seen in Appendix II.

More details about models and their templates for each P450 protein and their functions, ligands and binding residues were mentioned below:

10.2.1 B2406 structure analysis

Among the templets that offered by SWISS MODEL, cytochrome CYP144A1 (PDB ID: 5hdi. 1. A) was chosen to build a 3D structure of cytochrome P450 B2406 heme domain since it was the best hit with 21.51% sequence identity to the target protein (by SWISS

MODEL as well as Phyre) and 0.89 coverage. This model passed the qualification evaluation performed by VERIFY 3D with 82.02% of the protein's residues showed average (3D-1D) ≥ 0.2 (Appendix II). CYP144A1 is a monomer cytochrome isolated from *mycobacterium tuberculosis* (strain ATCC 25618 / H37Rv) and it is classified as oxidoreductase. The crystal structure of CYP144A1 was solved as substrate-free at 1.54 Å resolution (Noble et al., 1999).

A comparative structural analysis applied by superimposing the predicted model with the template showed that B2406 and P450 144 have nearly the same structural organization. For more assurance on the quality of the model, the RMSD between the resulting model and the templet was also measured and it was 1.59Å. Furthermore, the residues that bound to the heme (displayed as sticks in Figure 10-5 C) are highly conserved and this can be noticed in both the 3D structure and the sequence alignment between the templet and the model (these residues are highlighted in yellow in Appendix I). From the structure analysis of the heme cofactor, it was observed that three residues in the template structure bound to the heme; H129, R133 and R327 against three residues in the model structure; H117, K121 and R309 respectively. It could be shown that some of the secondary structure elements merge together to structure modules due to this absolute conservation. However, Arg133 in the template was replaced by Lys121 in the mode, both amino acids have positively charged side chains and very similar structures and functions.

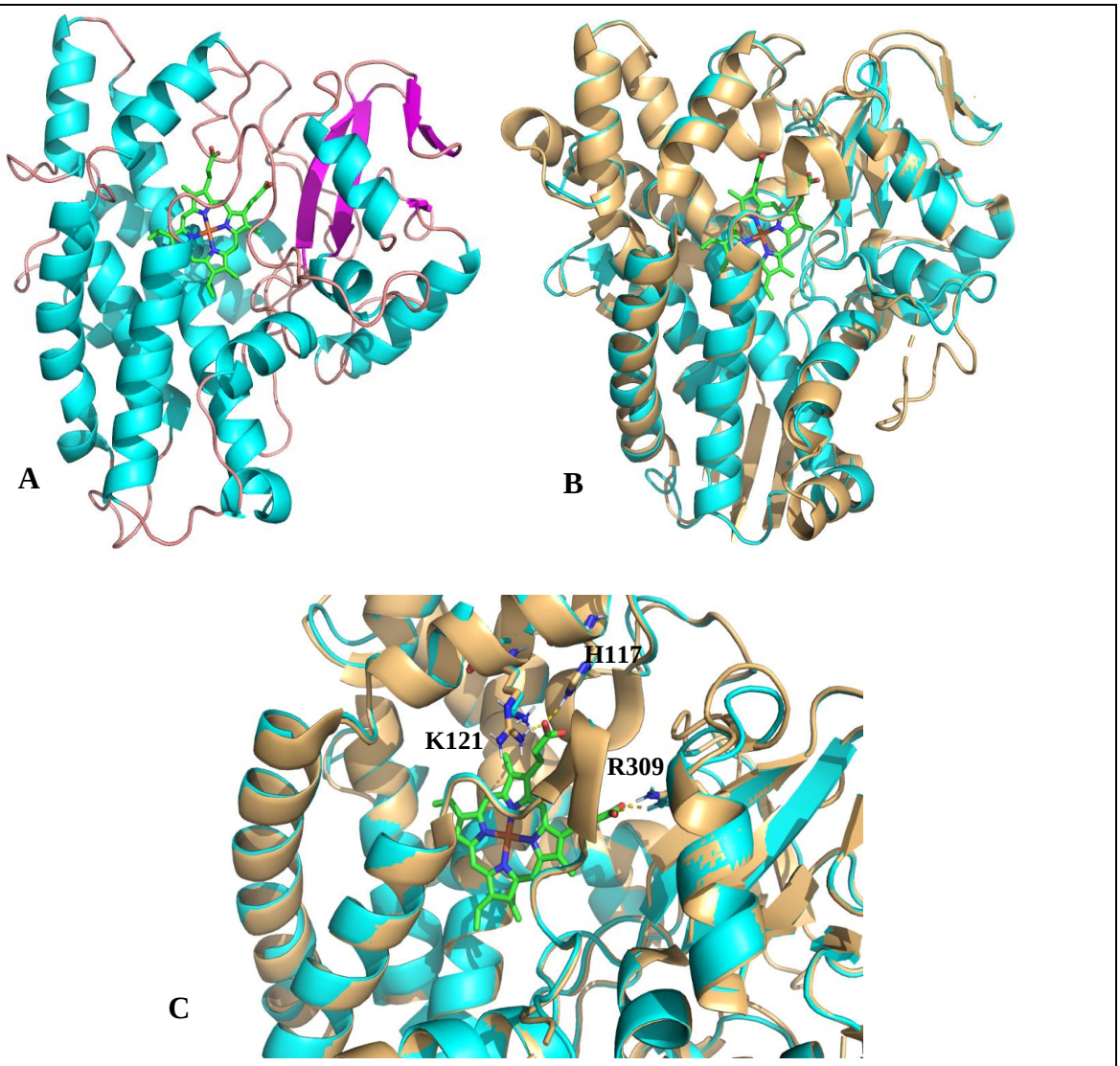


Figure 10-5: Structural analysis of P450 H16 B2406 heme domain. (**A**) 3D model of B2406 with heme cofactor. The heme cofactor is shown as stick and it is coloured by elements (carbon in green, nitrogen in blue, oxygen in red and sulphur in orange). Secondary structures are displayed in cyan (helices), magenda (sheets) and pink (coils). (**B**) Comparative structural analysis of B2406 model with structure (PDB ID: 5hdi. 1. A). Superimposed image of B2406 model (cyan) with CYP144A1 crystal structure (gold) is shown in the Figure. (**C**) Active site close view of B2406. Prosthetic group appearing at the centre of the active site is shown along with hydrogen bonds at yellow. Residues bound to the heme cofactor are shown as sticks and coloured by elements (hydrogen in gray, nitrogen in blue, oxygen in red and carbon atoms are coloured in cyan for the model and gold for the template. Amino acids in the model structure which bound to the heme cofactor are labelled in figure while the template's residues are listed in Table 10-3. This model was generated by using SWISS MODEL

10.2.2 B1743 structure analysis

C. necator H16 P450 B1743 model was built by SWISS MODEL using the template CYP130 (PDB ID: 2uvn. 1. A) isolated from *mycobacterium tuberculosis* (strain ATCC 25618 / H37Rv). CYP130 crystal structure was solved as econazole-bound CYP130 at 3 Å resolutions and it was classified as oxidoreductase (Rath et al., 2009). CYP130 was the best hit (by both SWISS MODEL and Phyre); it has 28.68 % sequence identity to P450 B1743 with 0.89 coverage. This model scored 98.70% in the VERIFY 3D quality evaluation (Appendix II).

The superimposing of the model to CYP130 structure showed good similarity with 1.41 Å RMSD over 376 residues. The structure analysis of the template showed 5 residues bound to the heme: H97, R101, R295, Y318 and S348 while the residues of the model that share the same positions with the above residues were as following: H124, R128, R318, Y341 and G370 respectively. It can be seen that 4 out of five residues are the same while the fifth one S348 in template was replaced by G370 in the model (Figure 10-6 C).

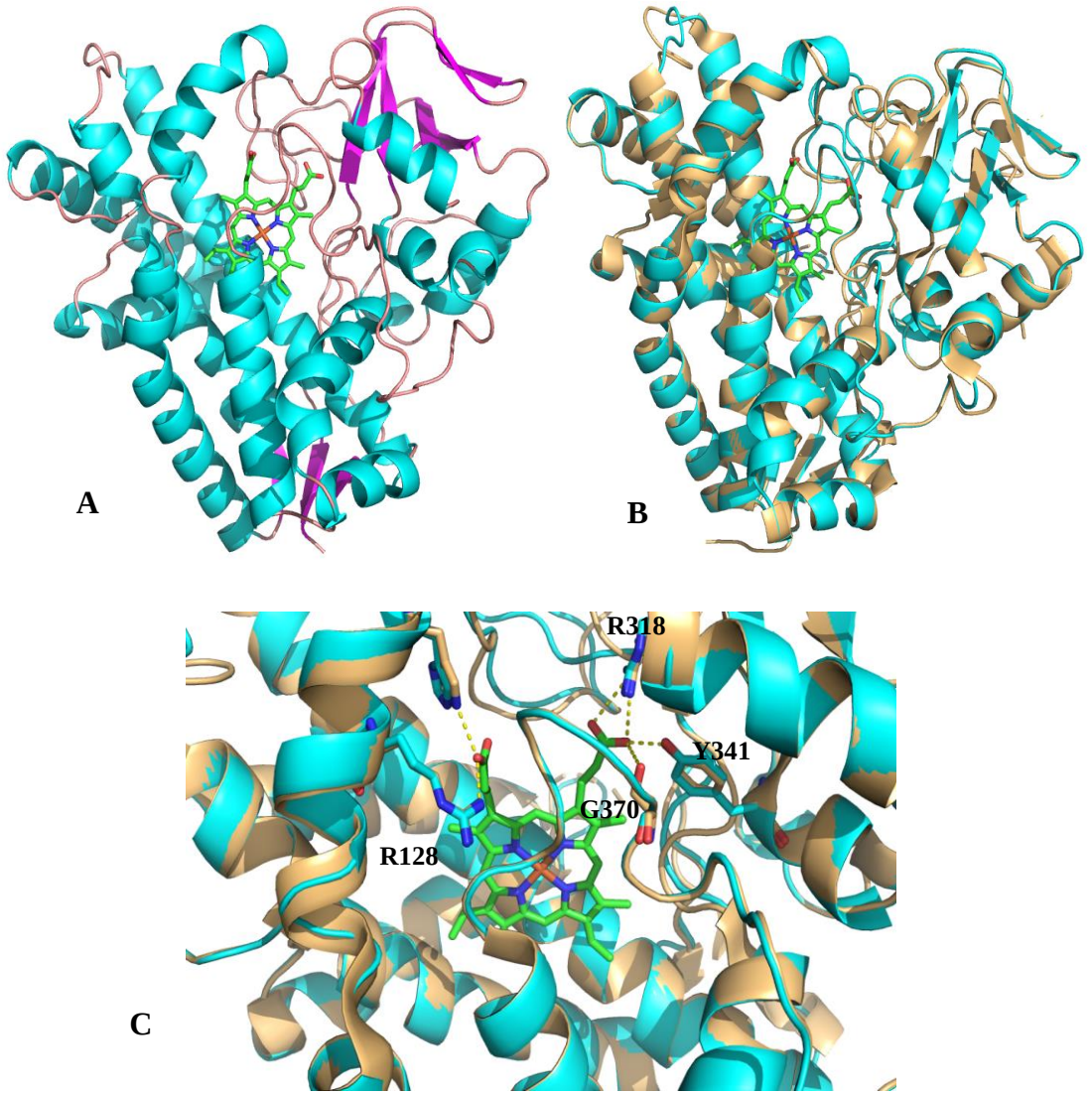


Figure 10-6: Structural analysis of P450 H16 B1743 heme domain. (A) 3D model of B1743 with heme cofactor. The heme cofactor is shown as stick and it is coloured by elements (carbon in green, nitrogen in blue, oxygen in red and sulphur in orange). Secondary structures are displayed in cyan (helices), magenda (sheets) and pink (coils). (B) Comparative structural analysis of B2406 model with structure (PDB ID: 2uvn. 1. A). Superimposed image of B2406 model (cyan) with CYP130 crystal structure (gold) is shown in the Figure. (C) Active site close view of B2406. Prosthetic group appearing at the centre of the active site is shown along with hydrogen bonds at yellow. Residues bound to the heme cofactor are shown as sticks and coloured by elements (hydrogen in grey, nitrogen in blue, oxygen in red and carbon atoms are coloured in cyan for the modwl and gold for the template. Amino acids in the model structure which bound to the heme cofactor are labelled in figure while the template's residues are listed in Table 10-3 This model was generated by using SWISS MODEL

10.2.3 B1279 structure analysis

10.2.3.1 B1279 heme domain:

The crystal structure of *tepidiphilus thermophilus* P450 heme domain (PDB ID: 6gii 1. A) was suggested by SWISS MODEL as well as Phyre as a best hit with 62.92 % sequence identity to build P450 B1279 3D model. The *tepidiphilus thermophilus* P450 heme domain is a monomer cytochrome and it is classified as oxidoreductase. The crystal structure of this template was solved as substrate-free at 1.9 Å resolution (Carugo, 2007). This model passed the qualification evaluation performed by VERIFY 3D with 94.95% of the protein's residues showed average (3D-1D) $\geq$ 0.2 (Appendix II).

The superimposing of the model to the template structure showed a very high similarity with 0.078 Å RMSD over 416 residues. The structure analysis of the template and the model showed that both structure have the same residues bound to the heme and these residues share the same positions which reflect the high accuracy of the predicted model. These residues are: H159, R163, R360 and H417and H151, R155, R352 and H409 of the template and the model respectively (Figure 10-7 A).

10.2.3.2 B1279 reductase domain:

P450 B1743 reductase domain model was built by Phyre using the template of CYP153A reductase (PDB ID: 2pia. 1. A) isolated from *burkholderia cepacia*. CYP153A crystal structure was solved at 2 Å resolutions and classified as phthalate dioxygenase reductase (Janson & Rydén, 1989). CYP153A was the best hit with 34 % sequence identity to P450 B1743. This model scored 96.21% in the VERIFY 3D quality evaluation (Appendix II).

When CYP153A structure superimposed to the model it showed good similarity with 1.03 Å RMSD. Two prosthetics groups: 2Fe-2S and FMN were noticed and represented as sticks in Figure 10-8. It is observed that four cysteine residues in both template and model bound to the 2Fe-2S ligand. Further, 9 residues in the template structure as well as model bound to the FMN and 7 out of 9 are the same while the other two are functionally similar and all binding residues occupied the same position. This structure analysis leads to the belief that the reliability of the predicted model is high and the ferredoxin reductase in the model is likely to be bound to the structure by the above binding residues. A list of the binding residues for both ligands were listed in Table 10-3.

10.2.4 B1009 structure analysis

10.2.4.1 B1009 heme domain:

C. necator H16 P450 B1009 3D heme domain model was built by SWISS MODEL using the template CYP102A1 (PDB ID: 1bu7. 1. A) isolated from *bacillus megaterium* (strain ATCC 14581 / DSM 32 / JCM 2506 / NBRC 15308 / NCIMB 9376 / NCTC 10342 / VKM B-512). P450 BM-3 crystal structure was solved in complex with FMN-binding domains of bacterial cytochrome P450BM-3 at 1.65 Å resolutions and was classified as oxidoreductase (Sevrioukova et al., 1999). This template's structure was the best hit with 46.89% sequence identity to P450 B1009 and 0.83 coverage. This model scored 90.52% in the VERIFY 3D quality evaluation (Appendix II).

A comparative structural analysis applied by superimposing the predicted model with the template showed that B1009 heme domain and P450 BM-3 are very similar and the RMSD between the resulting model and the templet was also measured and it was 0.37 Å which

indicates the high accuracy of the generated model. Furthermore, the residues that bound to the heme are highly conserved and this can be noticed from Figure 10-9 C. From overlapping the template to the model structure, it observed that three residues in the template structure bound to the heme; K69, W96 and R398 and the same residues in the model structure also bound to the heme cofactor which are; K124, W151 and R461 respectively these highly conserved binding residues increase the probability that the model predicted structure likely connects to the heme cofactor of the template.

Different model was suggested by Phyre. This model was analysed in order to compare it with the model that was offered by SWISS MODEL. Same protein was suggested by Phyre as a template which is P450 BM-3 but in deferent solved crystal structure (PDB ID: 2ij2. 1. A). This crystal structure is also for P450 BM-3 but in a novel substrate-free state when the mobile (helical) structural elements was reorganised resulting in a more open activity site cavity. This template has 46% identity with the model and 91.92% quality evaluation by VERIFY 3D (Appendix II).

When the model (PDB ID: 2ij2. 1. A) was superimposed to the template, it gave a highly similar result to what has been gained from the overlapping by using the template that was suggested by SWISS MODEL (PDB ID: 1bu7. 1. A) with the exactly same residues bound to the heme cofactor for both the model and the template (see Figure 10-10). However, the RMSD by using the structure that was suggested by Phyre2 (2ij2. 1. A) was high (1.01 Å) in comparison to the RMSD of the template given by SWISS MODEL (1bu7. 1. A), 0.37 Å and that make the first suggested template by SWISS MODEL (PDB ID: 1bu7. 1. A) more accurate (see Table 10-2.

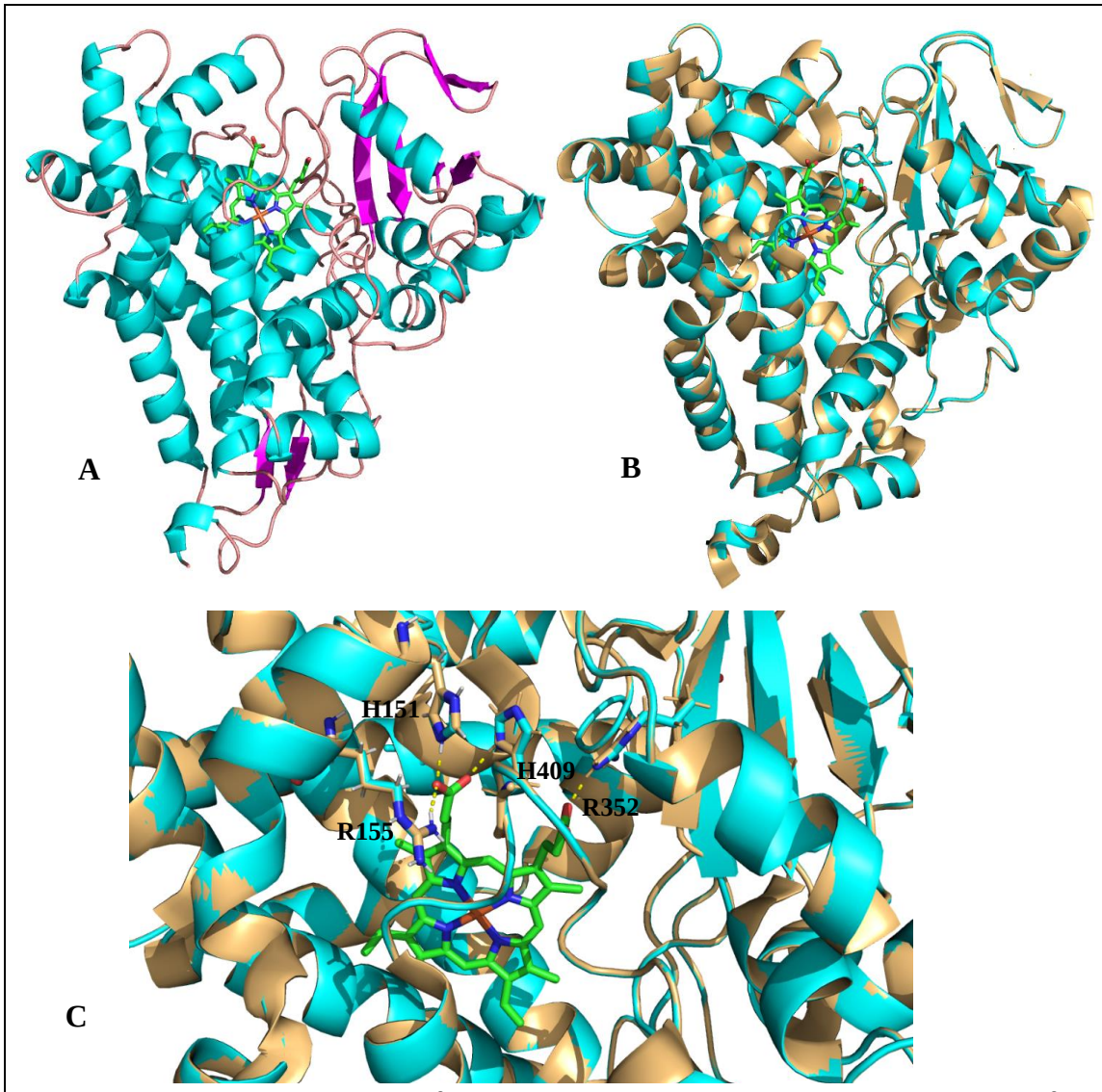


Figure 10-7: Structural analysis of P450 H16 B1279 heme domain. (A) 3D model of 1279 with heme cofactor. The heme cofactor is shown as stick and it is coloured by elements (carbon in green, nitrogen in blue, oxygen in red and sulphur in orange). Secondary structures are displayed in cyan (helices), magenda (sheets) and pink (coils). (B) Comparative structural analysis of the model with structure (PDB ID: 6gii 1. A). The model is shown in cyanand the template crystal structure in gold. (C) Active site close view of B2406. Prosthetic group appearing at the centre of the active site is shown along with hydrogen bonds at yellow. Residues bound to the heme cofactor are shown as sticks and coloured by elements (hydrogen in grey, nitrogen in blue, oxygen in red and carbon atoms are coloured in cyan for the B1279 modek and gold for the template. Amino acids in the model structure which bound to the heme cofactor are labelled in figure while the template's residues are listed in Table 10-3. This model was generated by using SWISS MODEL

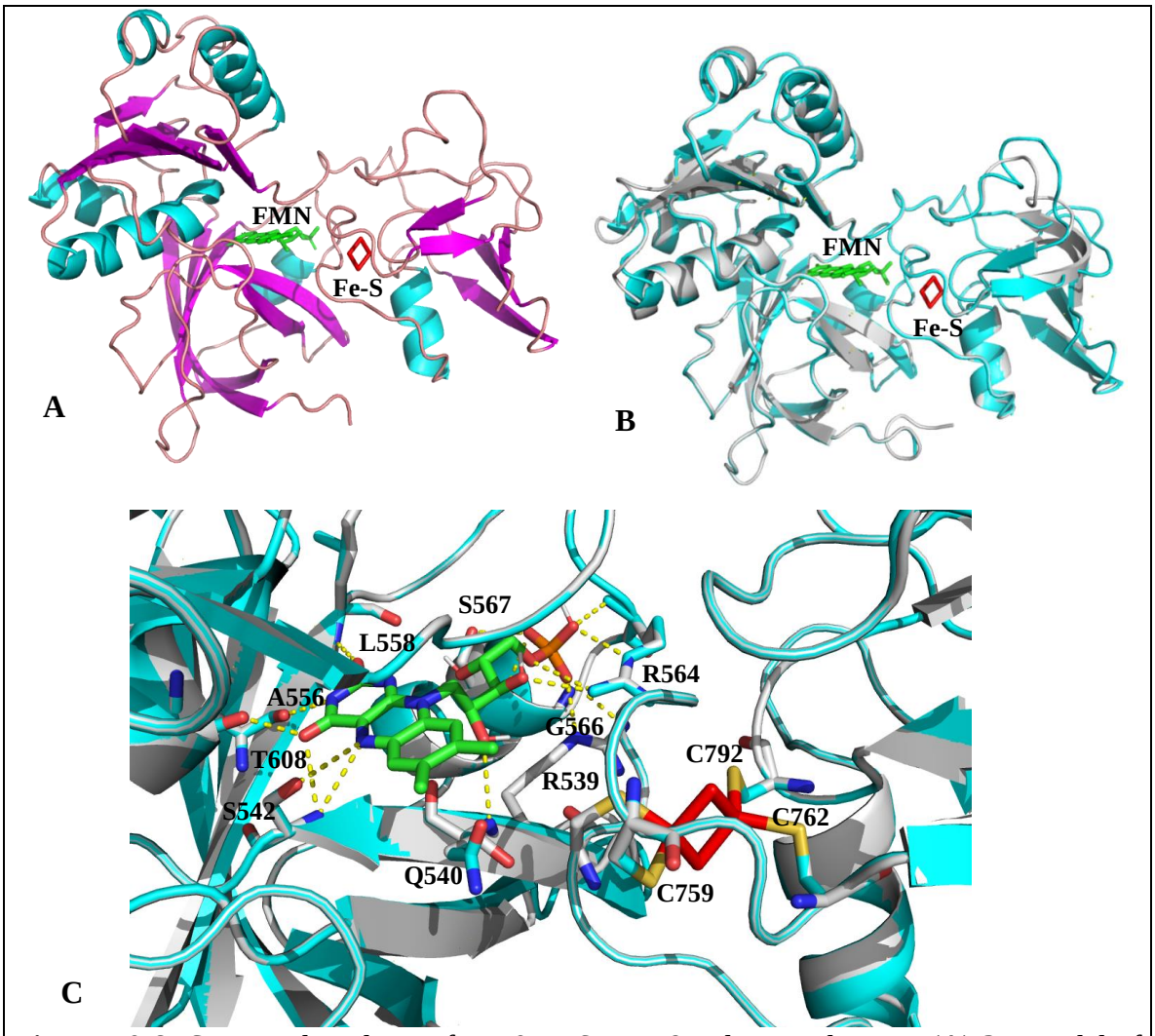


Figure 10-8: Structural analysis of P450 H16 B1279 reductase domain. (A) 3D model of B1279 reductase with two ligands. The ligands are shown as stick and it is coloured green for FMN and red for Fe-S cluster. Secondary structures are displayed in cyan (helices), magenda (sheets) and pink (coils). (B) Comparative structural analysis of B1279 model with structure (PDB ID: 2pia. 1. A). Superimposed image of the model (cyan) with CYP153 A crystal structure (grey) is shown in the Figure. (C) Reductase domain close view of B1279. Prosthetic groups appearing along with hydrogen bonds at yellow and they are coloured by elements (carbon in green, nitrogen in blue, oxygen in red and sulphur in orange). Residues bound to the heme cofactor are shown as sticks and coloured by elements (hydrogen in grey, nitrogen in blue, oxygen in red and carbon atoms are coloured in cyan for the model and grey for template). Amino acids in the model structure which bound to the heme cofactor are labelled in figure while the template's residues are listed in Table 10-3. This model was generated by using Phyre2.

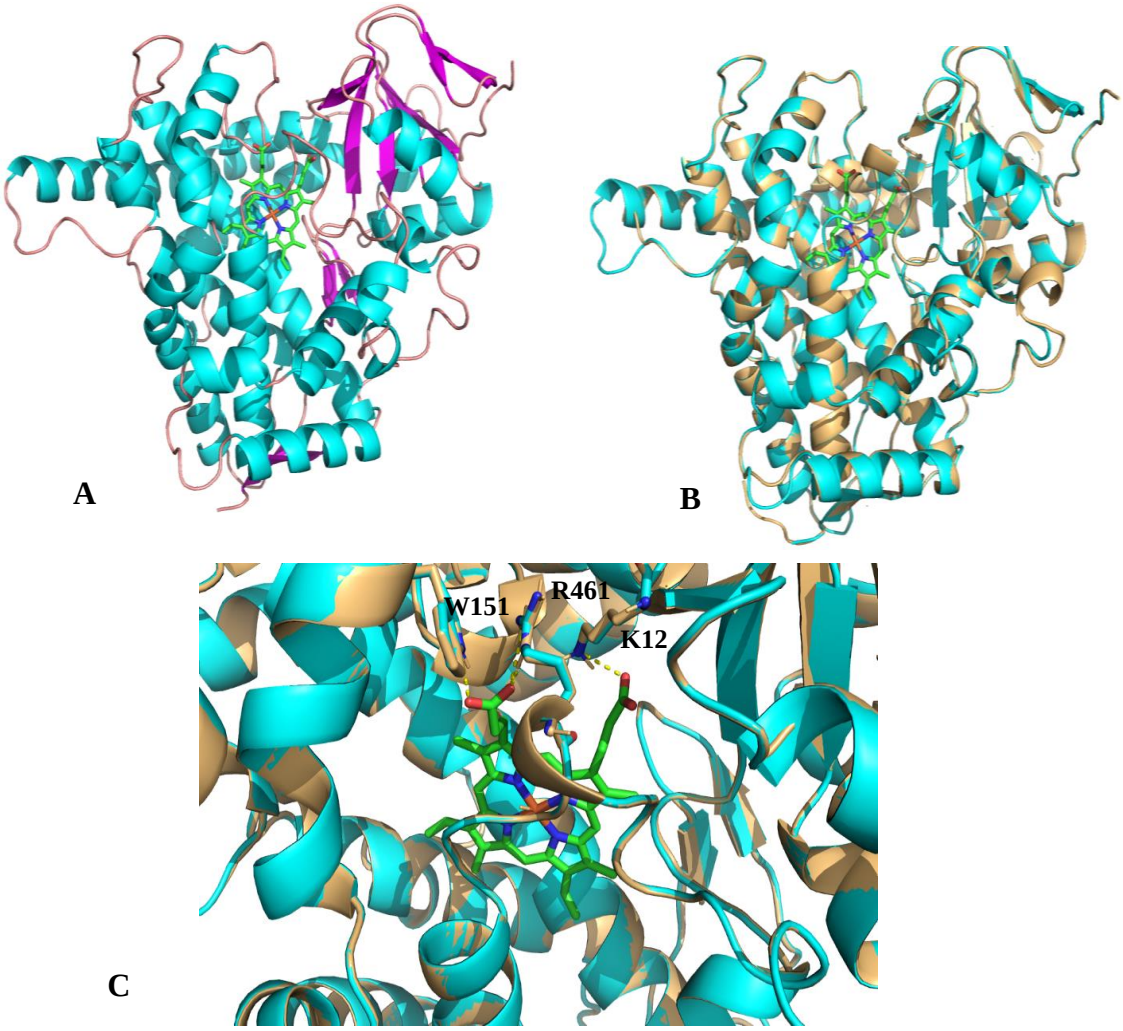


Figure 10-9: Structural analysis of P450 H16 B1009 heme domain. (A) 3D model of B1009 with heme cofactor. The heme cofactor is shown as stick and it is coloured by elements (carbon in green, nitrogen in blue, oxygen in red and sulphur in orange). Secondary structures are displayed in cyan (helices), magenda (sheets) and pink (coils). (B) Comparative structural analysis of B1009 model with structure (PDB ID: 1bu7. 1. A). Superimposed image of B2406 model (cyan) with CYP102A1 crystal structure (gold) is shown in the Figure. (C) Active site close view of B2406. Prosthetic group appearing at the centre of the active site is shown along with hydrogen bonds at yellow. Residues bound to the heme cofactor are shown as sticks and coloured by elements (hydrogen in grey, nitrogen in blue, oxygen in red and carbon atoms are coloured in cyan for the model and gold for template. Amino acids in the model structure which bound to the heme cofactor are labelled in figure while the template's residues are listed in table Table 10-3. This model was generated by using SWISS MODEL

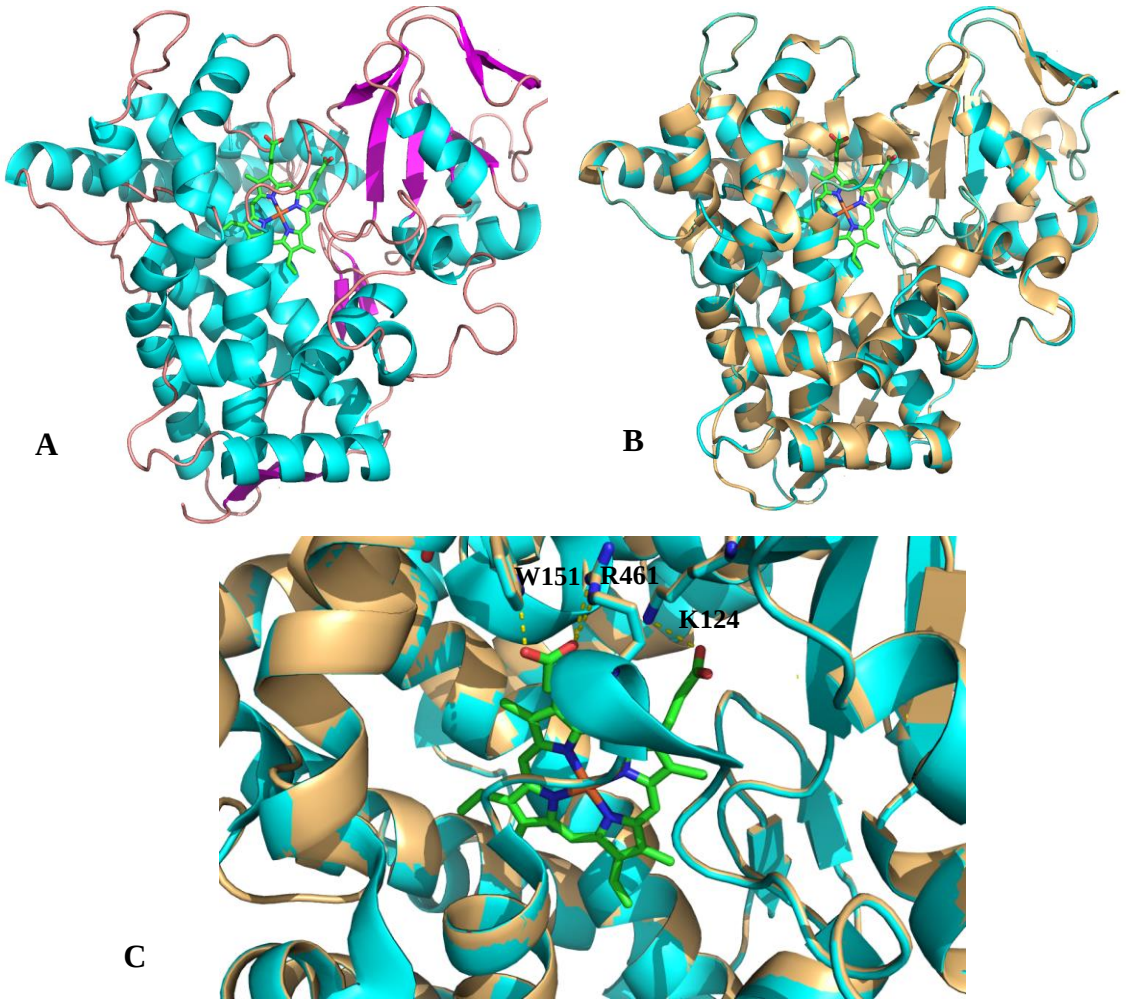


Figure 10-10: Structural analysis of P450 H16 B1009 heme domain. (A) 3D model of B2406 with heme cofactor. The heme cofactor is shown as stick and it is coloured by elements (carbon in green, nitrogen in blue, oxygen in red and sulphur in orange). Secondary structures are displayed in cyan (helices), magenda (sheets) and pink (coils). (B) Comparative structural analysis of B1009 model with structure (PDB ID: 2ij2. 1. A). Superimposed image of the model (cyan) with P450 BM-3 crystal structure (gold) is shown in the Figure. (C) Active site close view of B2406. Prosthetic group appearing at the centre of the active site is shown along with hydrogen bonds at yellow. Residues bound to the heme cofactor are shown as sticks and coloured by elements (hydrogen in grey, nitrogen in blue, oxygen in red and carbon atoms are coloured in cyan for the model and gold for the template). Amino acids in the model structure which bound to the heme cofactor are labelled in figure while the template's residues are listed in Table 10-3. This model was generated by using Phyre2.

10.2.4.2 B1009 reductase domain:

P450 B1009 reductase domain 3 D model was built by Phyre2 using the template from rat reductase complexed with nicotinamide-adenine-dinucleotide phosphate (NADP)(+) (PDB ID: 1j9z. B) isolated from *rattus norvegicus* . The template crystal structure was solved at 2.7 Å resolutions and classified as NADPH-cytochrome P450 reductase (Hubbard et al., 2001). This template was the best hit with 28% sequence identity to P450 B1009. This model scored 93.29% in the VERIFY 3D quality evaluation (Appendix II).

From superimposing the model to its template, a good similarity was observed with 1.16 Å RMSD over 576 residues. In Figure 10-11, three prosthetics groups were distinguished: a flavodoxin-like domain FMN, ferredoxin reductase-type FAD binding domain (FAD_FR) and nicotinamide-adenine-dinucleotide phosphate (NADP). These three ligands were shown as sticks in figure 10-11 and together formed a typical reductase structure.

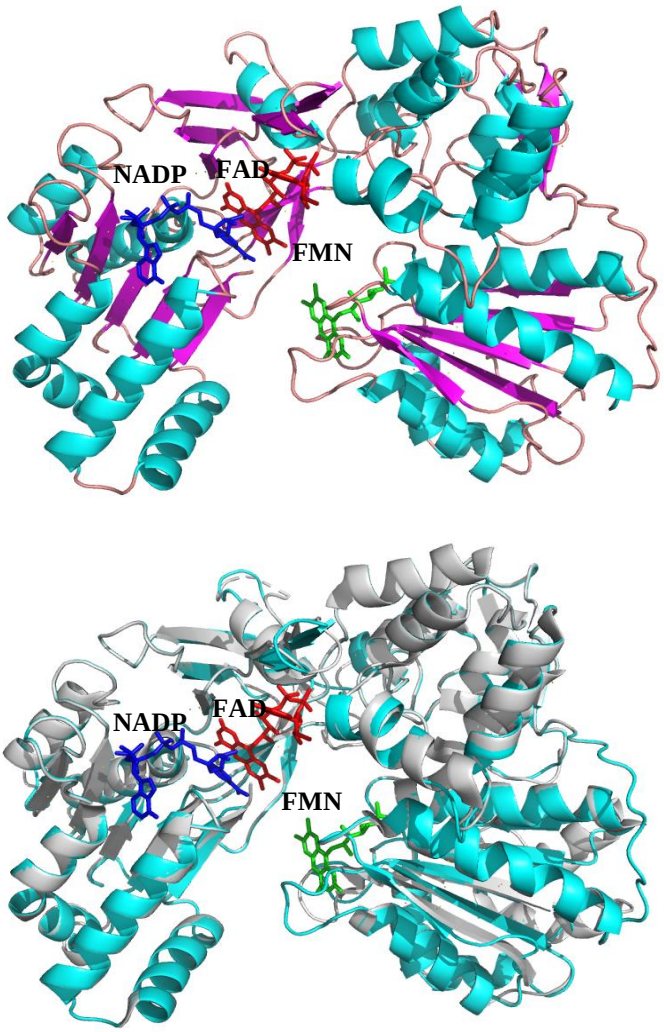


Figure 10-11: Structural analysis of P450 H16 B1009 reductase domain. (A) 3D model of B1009 reductase with three ligands. The ligands are shown as stick and it is coloured green for FMN, red for FAD and blue for NADP. Secondary structures are displayed in cyan (helices), magenda (sheets) and pink (coils). (B) Comparative structural analysis of B1009 model with structure (PDB ID: 1j9z. B). Superimposed image showed the model in cyan and the template crystal structure in grey. This model was generated by using Phyre2.

Table 10-2: List of the the results for the homology modelling of four *C. necator* H16 P450 proteins and templates used for this modilling.

Target model	Server	Template	PDB ID	Organism	Identity (%)	Overall RMSD (Å)
B2406 heme domain	SWISS MODEL	CYP144	5hdi. A	*Mycobacterium tuberculosis*	21.51	1.59
B1743 heme domain	SWISS MODEL	CYP130	2uvn . A	*Mycobacterium tuberculosis*	28.68	1.41
B1279 heme domain	SWISS MODEL	P450 heme domain	6gii. A	*Tepidiphillus thermophilus*	62.92	0.08
B1279 reductase domain	Phyre 2	CYP153A	2pia. A	*Marinobacter aquaeolei*	34	1.03
B1009 heme domain	SWISS MODEL	P450 BM-3	1bu7 A	*Bacillus megaterium*	46.89	0.37
	Phyre 2	P450 BM-3	2ij2. A	*Bacillus megaterium*	46	1.01
B1009 reductase domain	Phyre 2	NADPH-P450 reductase	1j9z. B	*Rattus norvegicus*	28	1.16

Table 10-3: This table lists the residues that bound to theprosthetic groups in models and their templates of four C. necator H16 P450 proteins

Target model	Ligands	Binding residues in the model structure	Template	Binding residues in the template structure
B2406 heme domain	Heme cofactor	H117, K121, R309	CYP144 (PDB ID: 5hdi. A)	H129, R133, R327
B1743 heme domain	Heme cofactor	H124, R128, R318, Y341, G370	CYP130 (PDB ID: 2uvn. A)	H97, R101, R295, Y318, S348
B1279 heme domain	Heme cofactor	H151, R155, R352, H409	P450 heme domain (PDB ID: 6gii. A)	H159, R163, R360, H417
B1279 reductase domain	Fe-S	C754, C759, C762, 792	CYP153 (PDB ID: 2pia. A)	C272, C277, C280, C308
	FMN	R539, Q540, S542, A556, L558, R564, G566, S567, T608		R55, T56, S58, A72, K74, R80, G82, S83, T124
B1009 heme domain by SWISS MODEL	Heme cofactor	K124, W151, R461	P450 BM-3 (PDB ID: 1bu7. A)	K69, W96, R398
B1009 heme domain by Phyre	Heme cofactor	K124, W151, R461	P450 BM-3 (PDB ID: 2ij2. A)	K69, W96, R398

10.3 Study of the basic configuration of the P450 expression vectors

The pETM-11 plasmid from Novagen was chosen in this study as an expression vector to express four P450s, B2406, B1743, B1279 and B1009, due to its ability to produce a high level of soluble and nontoxic recombinant proteins in *E. coli*, the expression vectors of these proteins are shown in Figure 10-12 to Figure 10-15. In this system, the expression of the target gene is controlled by the T7 phage RNA polymerase promoter in production hosts containing a prophage (λDE3) (Baneyx, 1999).

A commercial expression vector is designed with at least one affinity tag to allow efficient purification of large amount of protein, while some of these tags were used to enhance the protein's solubility. There are many known affinity tags such as glutathione S-transferase (GST), maltose binding protein (MBP) and polyhistidine tag (His), but the most commonly used is the polyhistidine tag consisting of 2–10 residues, which facilitate the production of pure protein in large quantities. The pETM-11 plasmid in this study was designed with a His-tag to allow easy separation of pure proteins. In spite of the advantages of using fusion proteins, in some cases they could negatively affect enzyme activity. Consequently, cleavage factors were added to the expression vector to separate the recombinant protein from affinity tags. For the pETM-11 plasmid used in this project, tobacco etch virus protease (TEV) was inserted for this purpose. The His domain was connected to the plasmid by a linker region, which enhances the flexibility and mobility of the protein to achieve an efficient purification and tag cleavage. In addition, this linker region helps to improve protein solubility. Furthermore, many restriction sites were included in the pETM-11 to cut and ligate the recombinant DNA ends into the plasmid according to changes in the requirements during the experimental work.

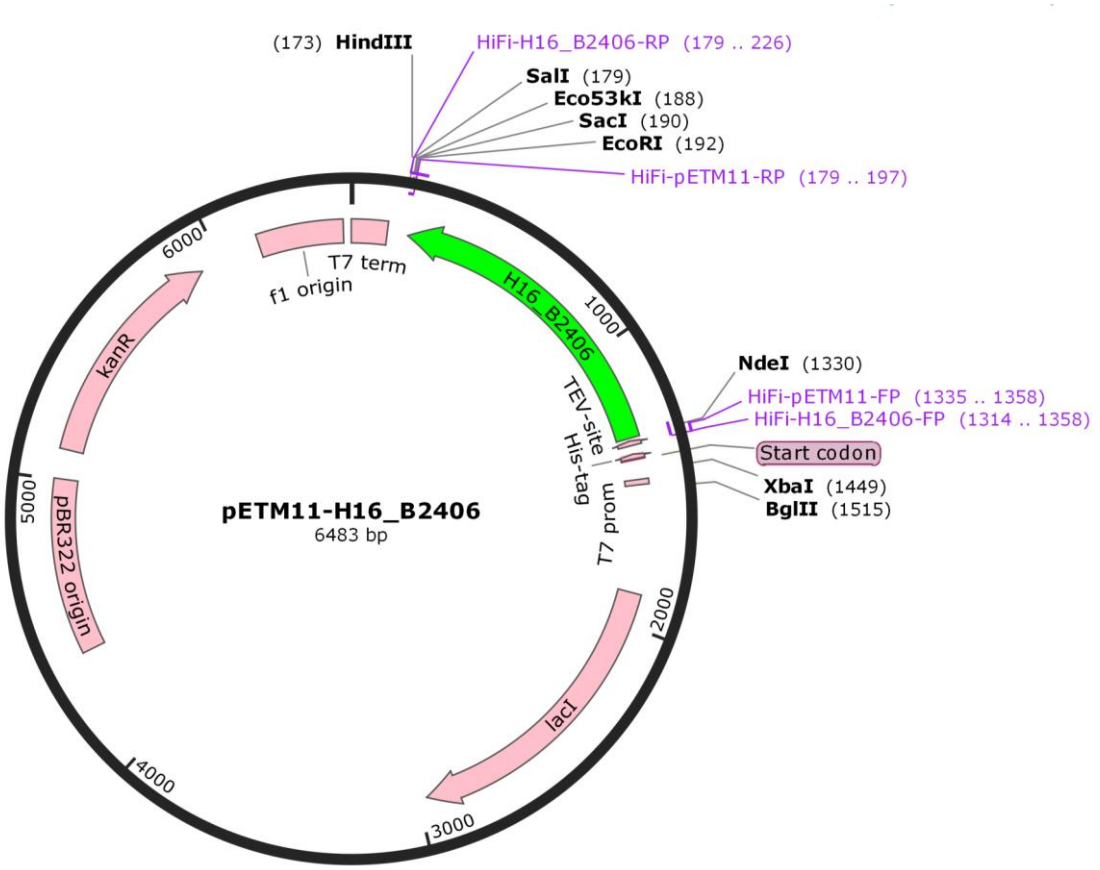


Figure 10-12: pETM11 expression plasmid (6483 bp) encoded H16_B2406 gene (green). The pETM11 vector contains the N-terminal T7 promoter and kanamycin resistance selection marker. Plasmid image was produced using SnapGene software.

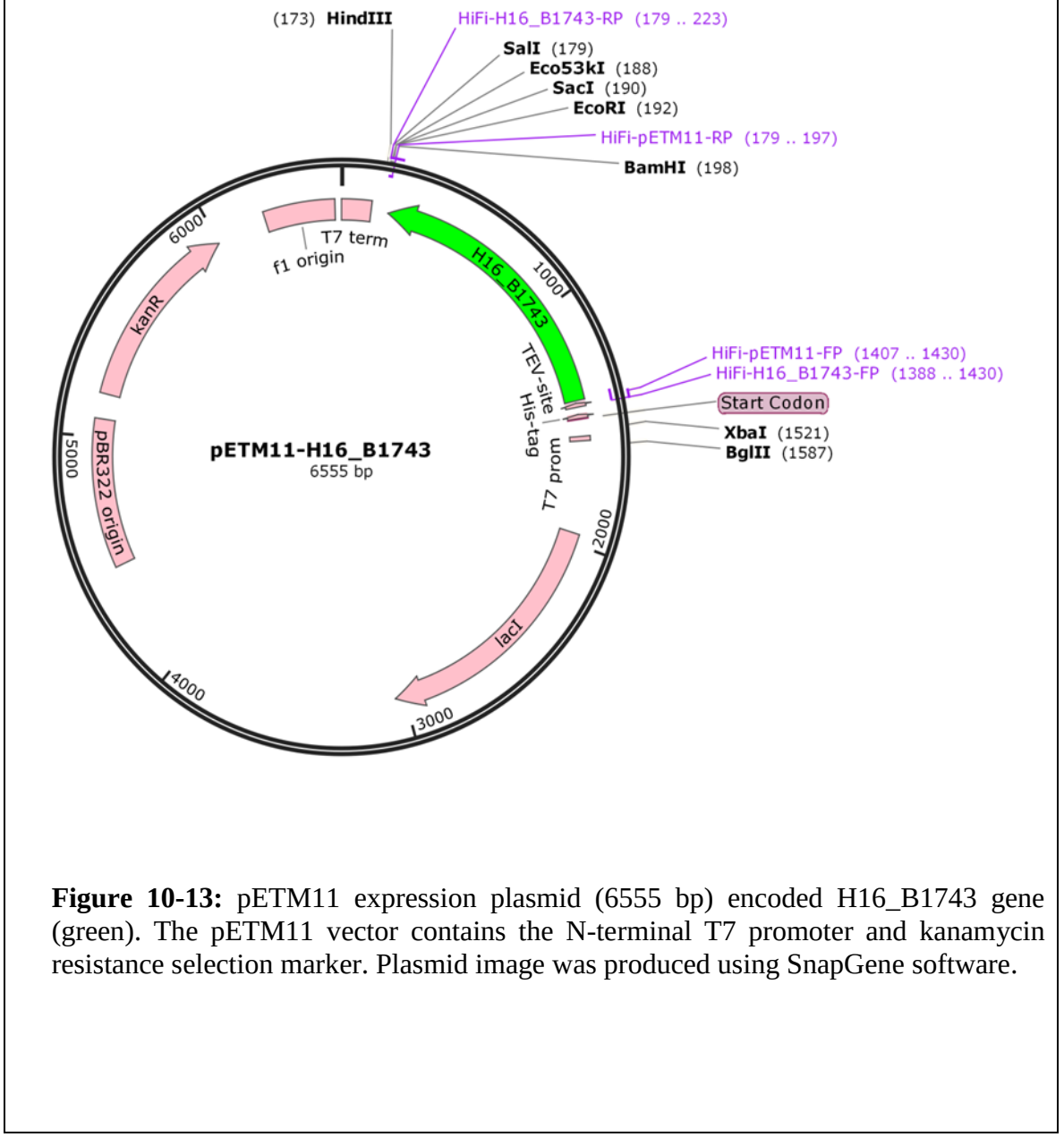


Figure 10-13: pETM11 expression plasmid (6555 bp) encoded H16_B1743 gene (green). The pETM11 vector contains the N-terminal T7 promoter and kanamycin resistance selection marker. Plasmid image was produced using SnapGene software.

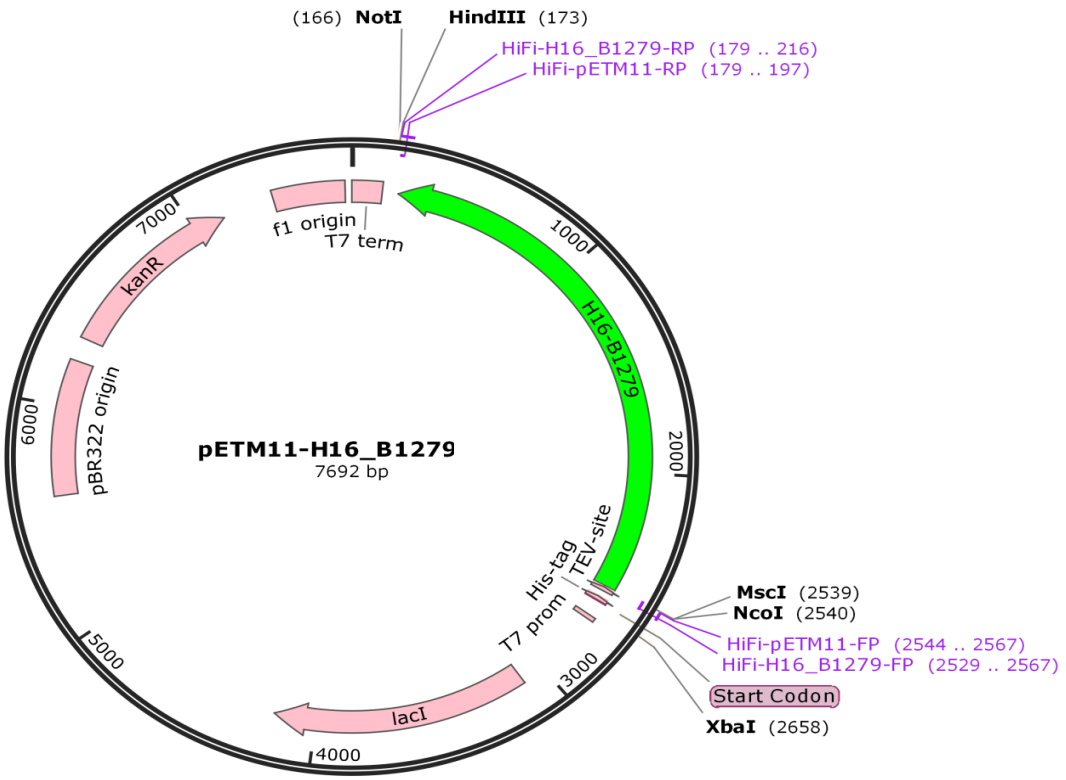


Figure 10-14: pETM11 expression plasmid (7692 bp) encoded H16_B1279 gene (green). The pETM11 vector contains the N-terminal T7 promoter and kanamycin resistance selection marker. Plasmid image was produced using SnapGene software.

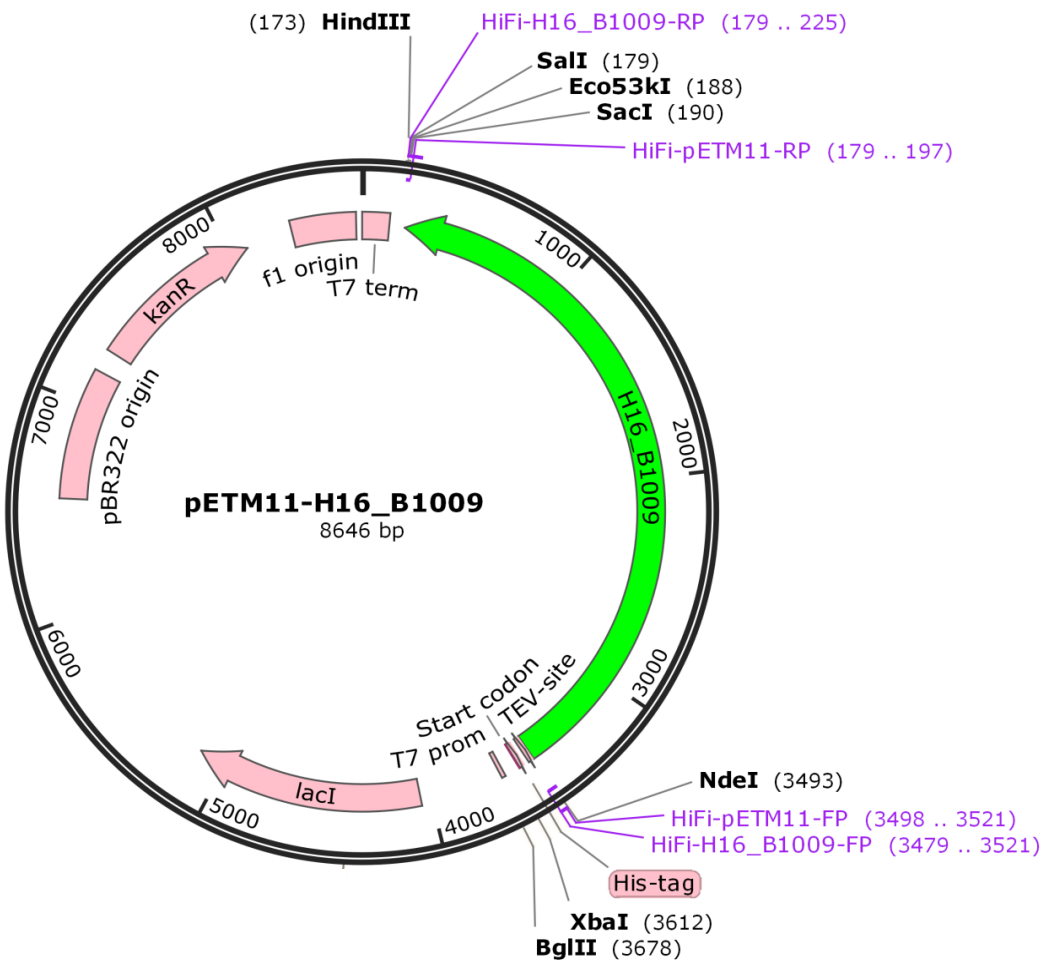


Figure 10-15: pETM11 expression plasmid (8646 bp) encoded H16_B1009 gene (green). The pETM11 vector contains the N-terminal T7 promoter and kanamycin resistance selection marker. Plasmid image was produced using SnapGene software

11 Characterisation of putative cytochrome P450s from *Cupriavidus necator* H16

11.1 Putative cytochrome P450s from *Cupriavidus necator* H16

When the concentration and the purity of the isolated plasmids; pETM11-H16-B2406, pETM11-H16-B1743, pETM11-H16-B1279 and pETM11-H16-B1009 were checked, it is found that all concentrations ranged between 70–80 nM and the A_{260}/A_{230} and A_{260}/A_{280} ratios were equal or more than 2 indicating the high purity of plasmids and their capability to be used in further investigations.

These four isolated plasmids; pETM11-H16-B2406, pETM11-H16-B1743, pETM11-H16-B1279 and pETM11-H16-B1009 were transferred into *E. coli* BL21 (DE3) cells and showed a significant growth on TYE agar media supplemented with 50 µg/ml kanamycin in comparison to no growth on the control plate (free-plasmid *E. coli* BL21 (DE3) cells) (Figure 11-1). The expression was carried out in TB auto-induction media at 30°C for 24 hr. The transformation and expression protocols were described in detail in section 9.5. Collected pellets after 24 hr were red in colour, indicating successful heme protein expression. Different expression levels can be observed depending on the pellet colour as shown in Figure 11-2.

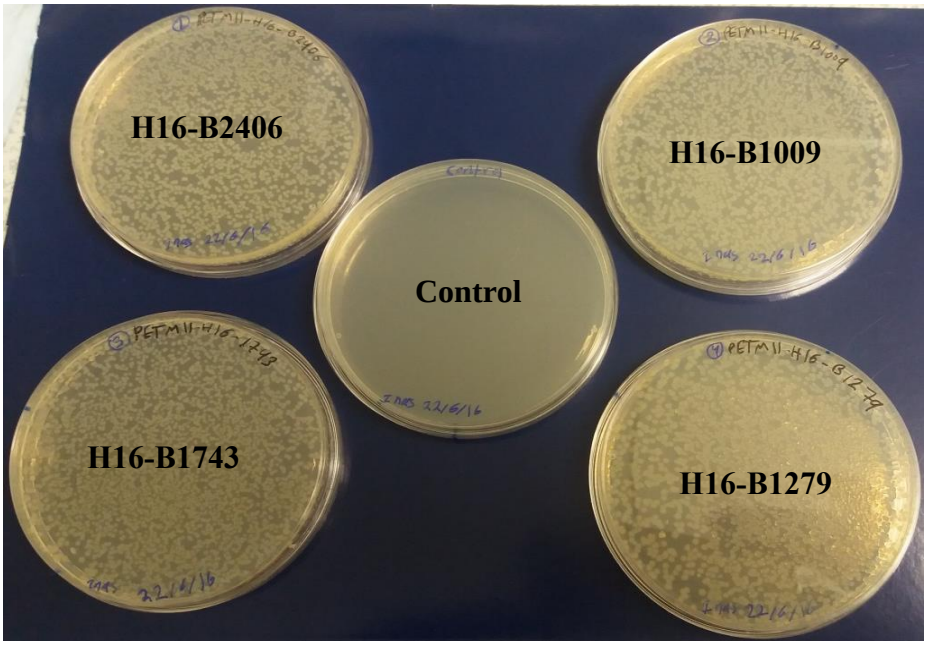


Figure 11-1: The resulting clones from the transformation of pETM11 plasmid into BL21 (BM3). Four genes were transformed. The plates from transformation were presented as: 1) H16-B2406. 2) H16-B1009. 3) H16-B1743. 4) H16-B1279 5) the control.

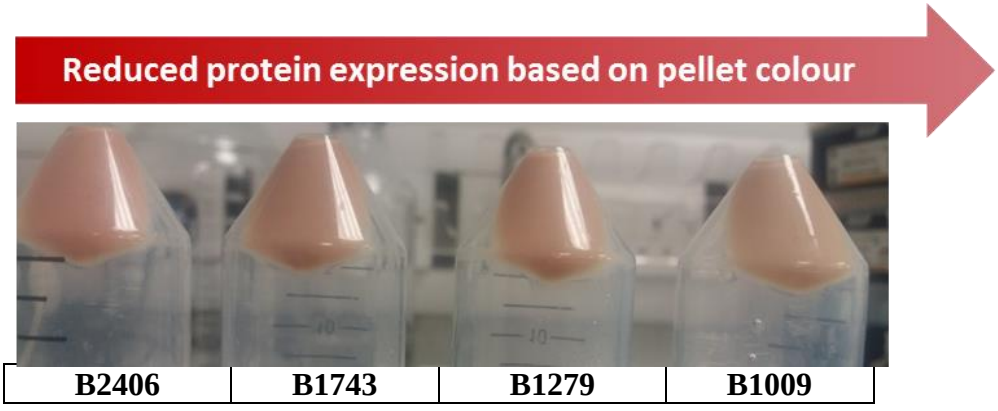
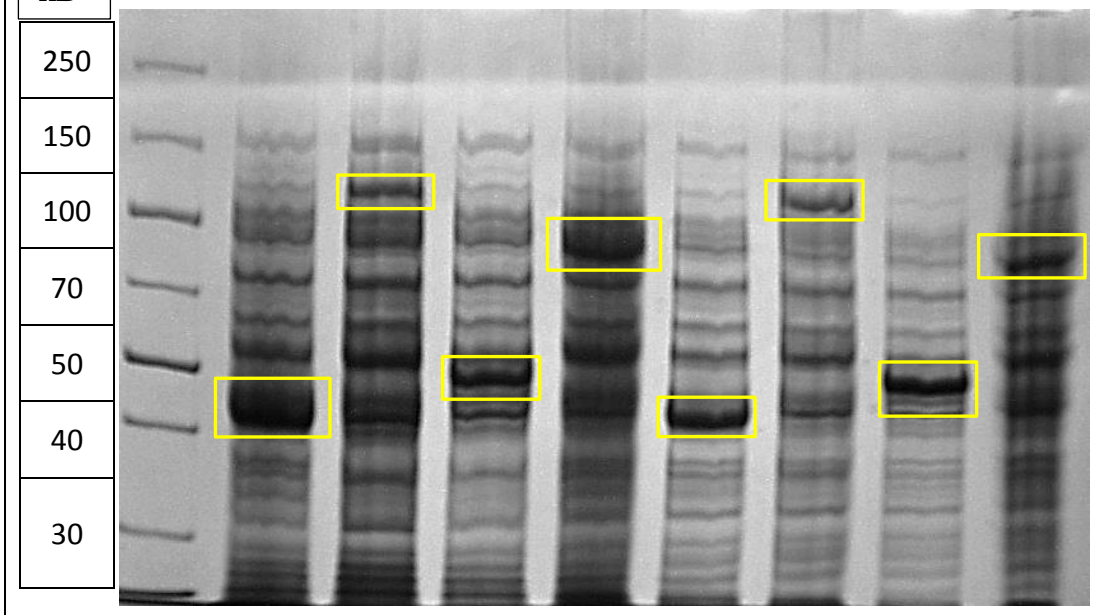


Figure 11-2: The pellet from the expression of four genes in BL21 (BM3). Those gees are: 1) H16-B2406 2) H16-B1743 3) H16-B1279 4) H16-B1009, 50 ml of TB AIM media was used for the expression of each protein at 30°C. The incubation period was 24 hr.

The solubility of *Cupriavidus necator* H16 P450s were checked by SDS-PAGE as described in section 9.5. The protein bands are surrounded by yellow rectangles to clarify the soluble/insoluble fractions (Figure 11-3). From the comparison of the soluble fractions (lane 2-5) to the insoluble fractions (lane 6-9) of the same proteins, it could be concluded that all *Cupriavidus necator* H16 P450 protein are soluble in water when they expressed in *E. coli* BL21 (DE3) as a host cells.



1	2	3	4	5	6	7	8	9
Protein ladder	H16-B2406	H16-B1009	H16-B1743	H16-B1279	H16-B2406	H16-B1009	H16-B1743	H16-B1279
	Soluble fraction				Insoluble fraction			

Figure 11-3: SDS-PAGE analysis for the soluble and insoluble fraction of four genes in pETM11 plasmid: H16-B2406, H16-B1743, H16-B1279 and H16-B1009. The yellow boxes surround the target proteins that clarified in the table underneath each lane.

11.2 Optimisation of protein expression and purification

After testing the possibility of expressing P450s from *Cupriavidus necator* H16 and checking the solubility of these proteins, it is prudent to check the yield of each protein and the possibility to purify these proteins. Proteins from 50 ml of expression culture were lysed and loaded onto a HisTrap column. The purification was performed as described in section 9.5. From Figure 11-4 and Figure 11-5, it can be observed that two proteins, B2406 and B1743, showed a very good yield and purity from the first step of purification, while the yield of B1279 was lower and associated with unknown protein at 50 kDa (Figure 11-6). In contrast, the expression and purity of B1009 was very low (Figure 11-7). These results indicate that both B2406 and B1743 can be expressed on a large scale using the optimal conditions, while the large-scale expression of B1279 and B1009 requires further improvement.

Two *E. coli* strains and a variety of conditions were applied in the expression optimisation of B1279 and B1009, which was described in section 9.5.1. The expression of B1009 remained very low with no significant deference between the expression in *E. coli* BL21 (DE3) and *E. coli* C41 (DE3), also no enhancement was observed when the incubation temperature was decreased from 30°C to 25°C and when 2×TY AIM or SB AIM were used as shown in Figure 11-8. However, using a lower expression temperature (25°C) increased B1279's expression when the other factors were fixed. Also, using SB AIM gave the highest protein production in both *E. coli* BL21 (DE3) and *E. coli* C41 (DE3). In addition, from the gel picture (Figure 11-9), C41 (DE3) strain provided higher B1279 expression compared to *E. coli* BL21 (DE3), while the later gave the purest protein. Based on these results, the optimal conditions of SB AIM, 25°C and *E. coli* BL21 (DE3) were used to express B1279 on large-scale using the protocol described in section 9.6.

The expression of B1009 was also investigated in *E coli* HMS174 (DE3) strain. The induction using auto-induction media and IPTG were examined at 25°C and 30°C incubation temperatures. The results showed an improvement in B1009 expression in *E. coli* HMS174 (DE3) using auto-induction media at 30°C but the protein was insoluble by using this strain (Figure 11-10 and Figure 11-11). From the results, it could be indicated that *E. coli* HMS174 (DE3) is favourite over *E. coli* BL21(DE3) and *E. coli* C41(DE3) for B1009 expression. However, solubility of B1009 need to be improved by using one or more of proteins' improvement solubility strategies.

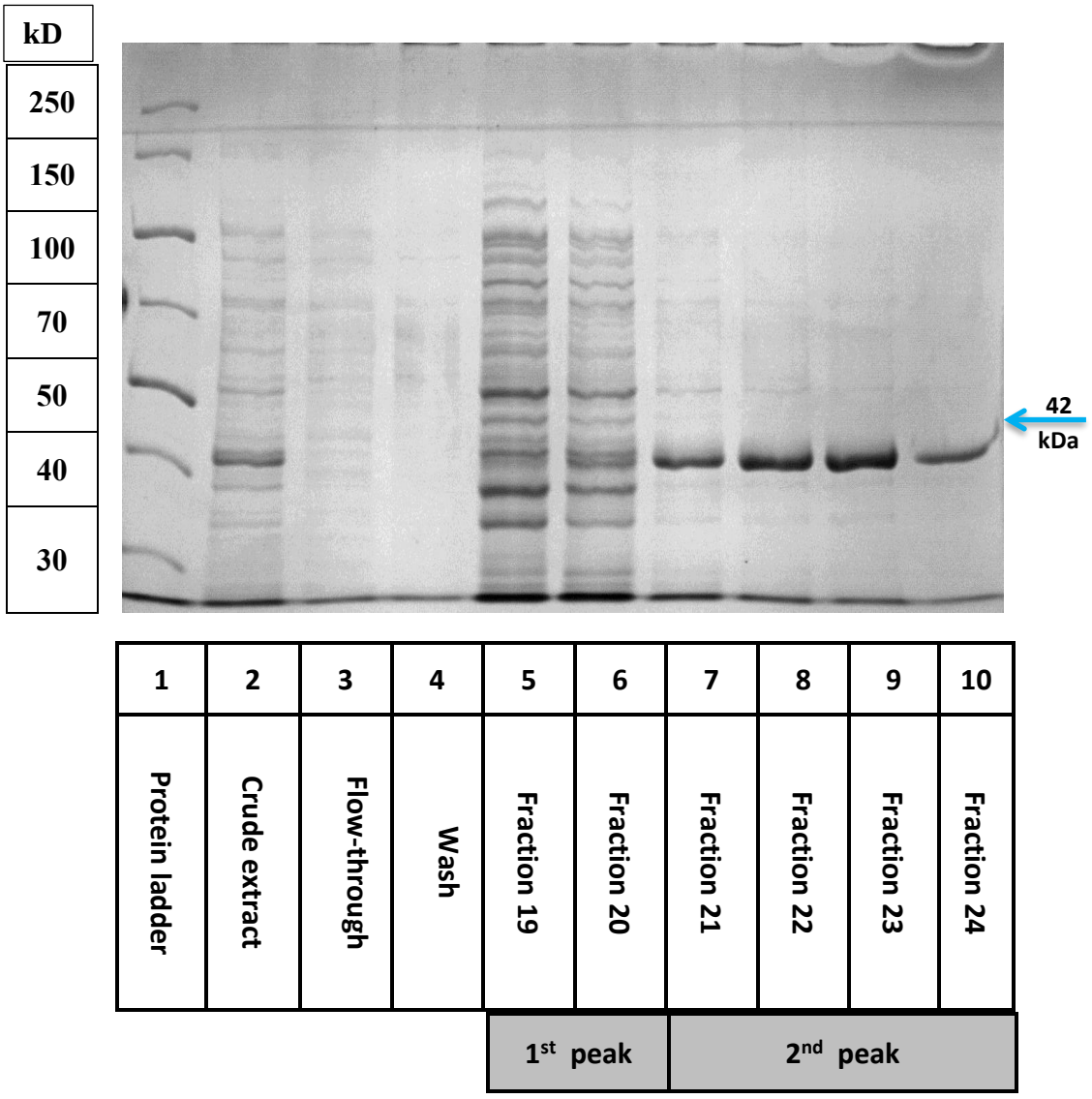


Figure 11-4: SDS-PAGE analysis of B2406 protein for the fractions eluted from affinity chromatography for the protein expressed in BL21 (DE3), 50 ml of sample was loaded to the HisTrap 5 ml column. The purification conditions were: 2 ml/min, 40 min.

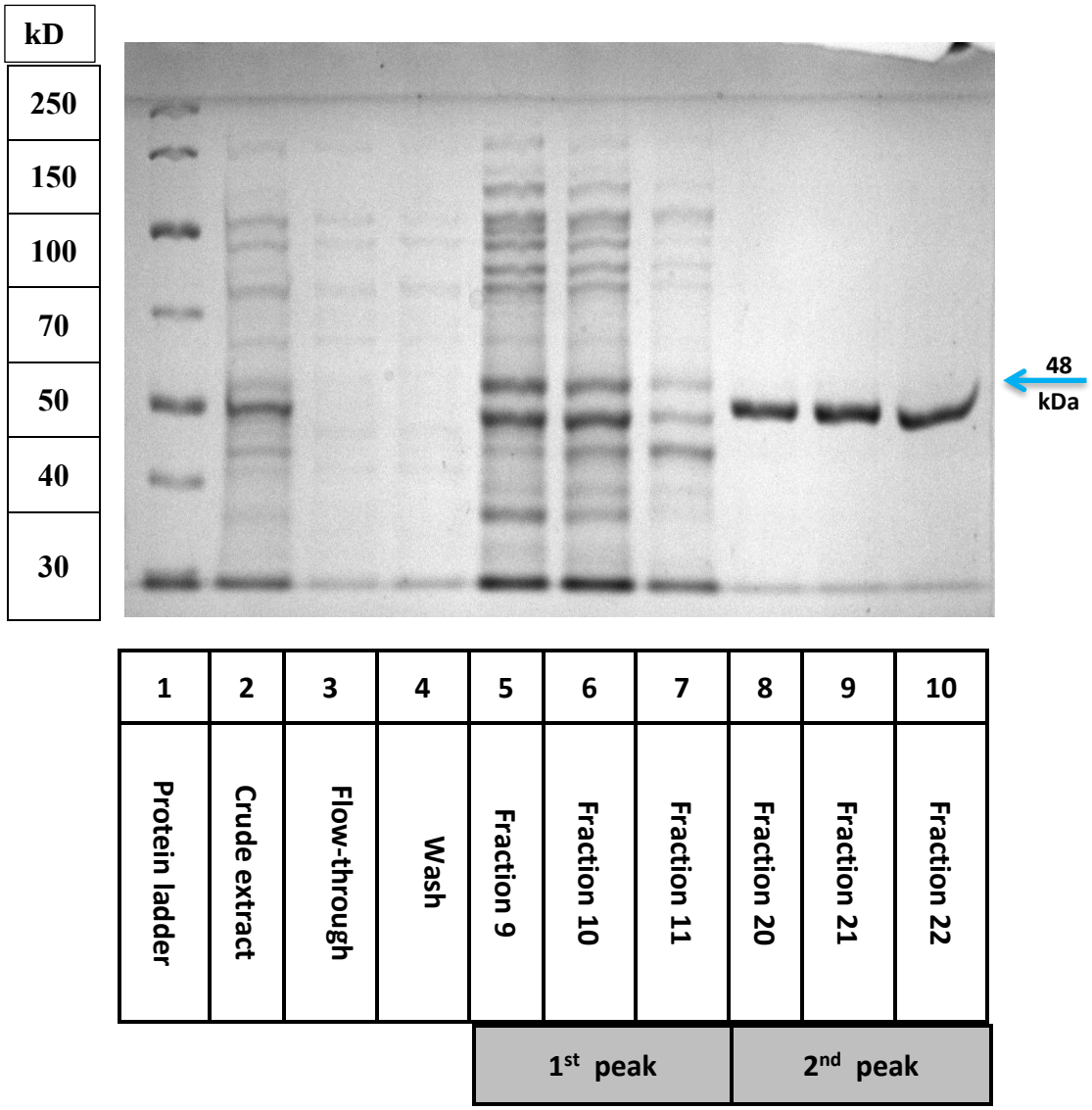


Figure 11-5: : SDS-PAGE analysis of B1743 protein for the fractions eluted from affinity chromatography for the protein expressed in BL21 (DE3), 50 ml of sample was loaded to the HisTrap 5 ml column. The purification conditions were: 2 ml/min, 40 min.

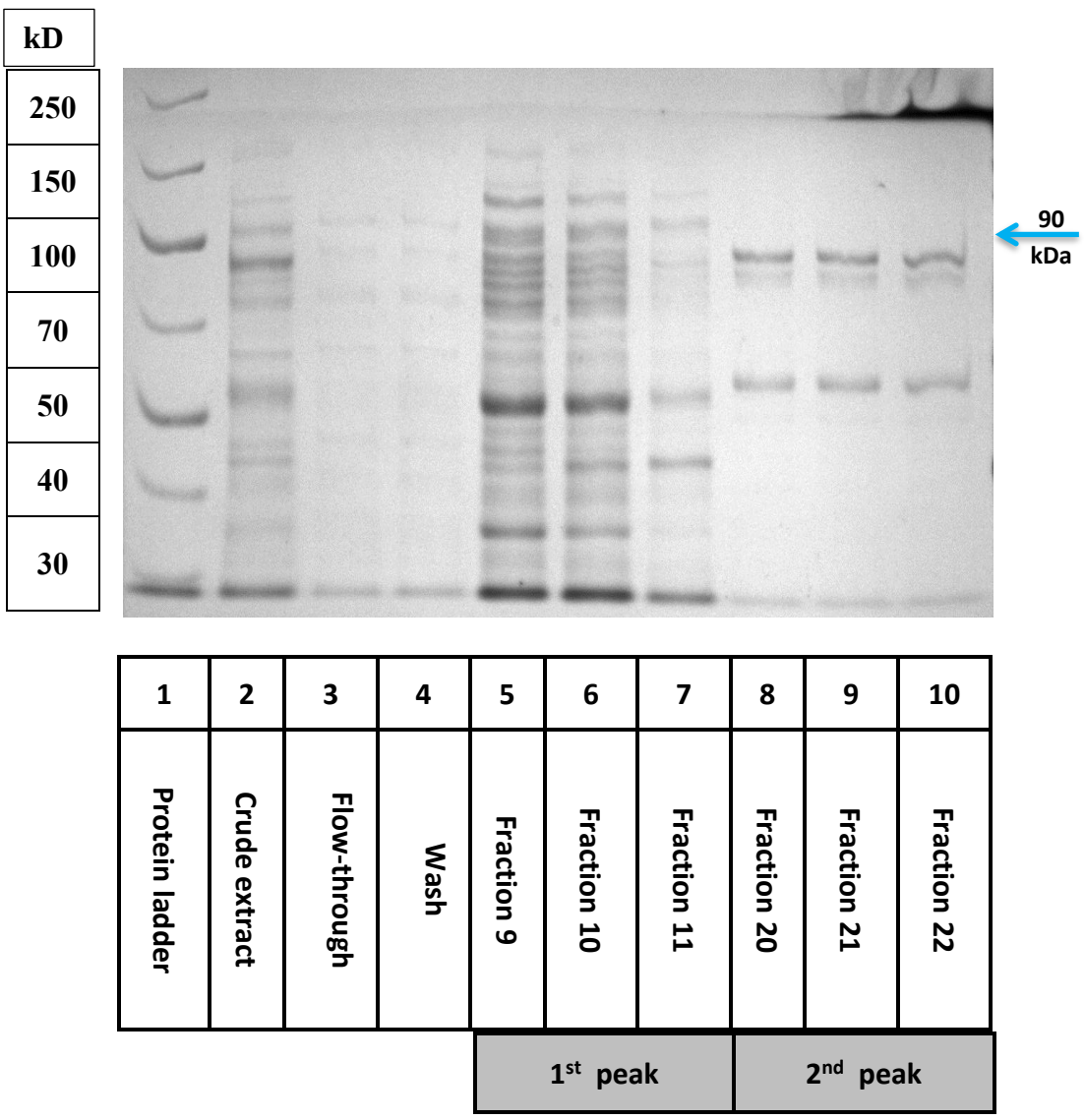


Figure 11-6: SDS-PAGE analysis of B1279 protein for the fractions eluted from affinity chromatography for the protein expressed in BL21 (DE3), 50 ml of sample was loaded to the HisTrap 5 ml column. The purification conditions were: 2 ml/min, 40 min.

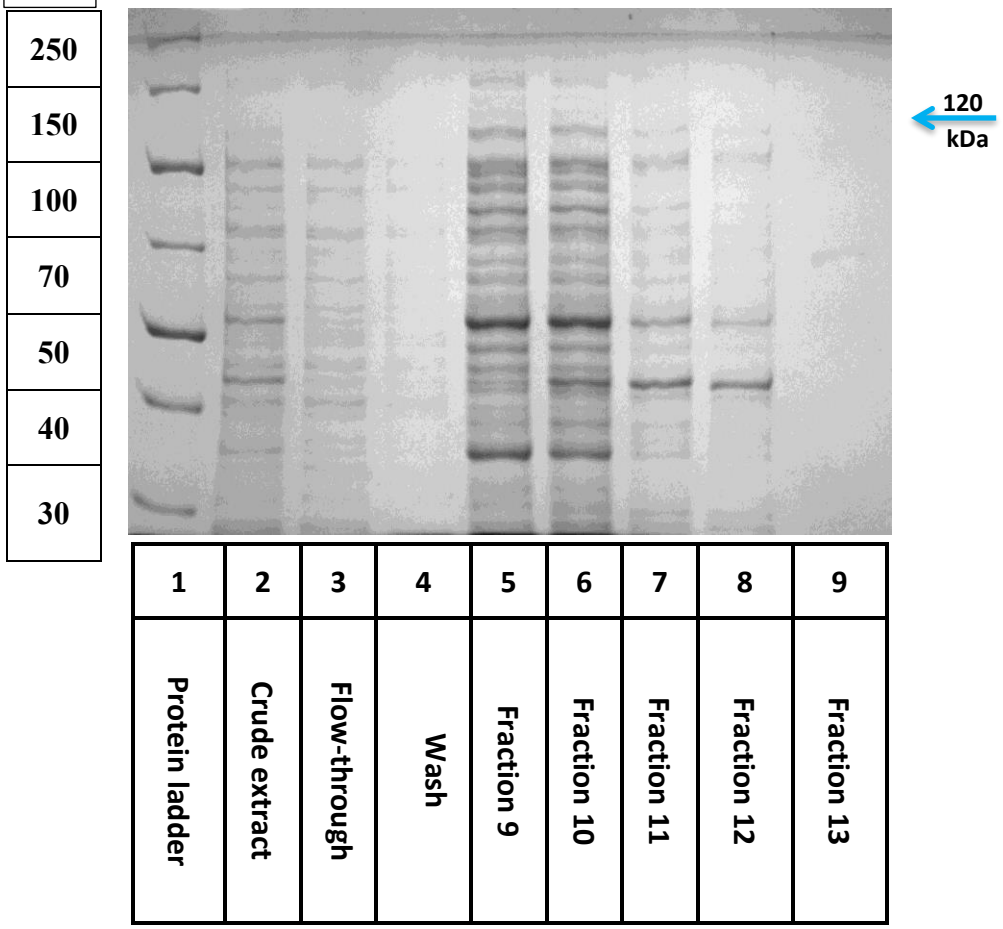


Figure 11-7: SDS-PAGE analysis of B1009 protein for the fractions eluted from affinity chromatography for the protein expressed in BL21 (DE3), 50 ml of sample was loaded to the HisTrap 5 ml column. The purification conditions were: 2 ml/min, 40 min.

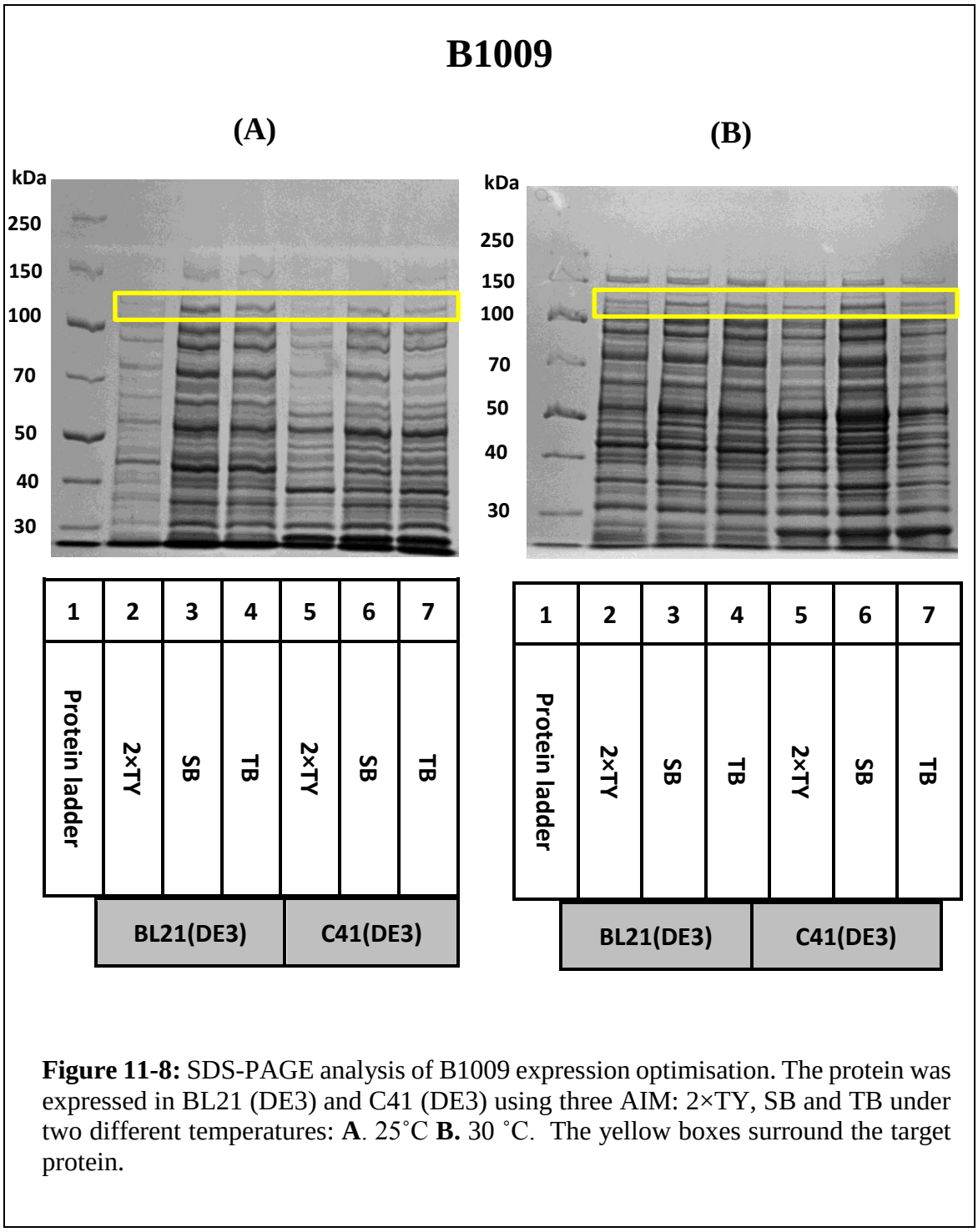


Figure 11-8: SDS-PAGE analysis of B1009 expression optimisation. The protein was expressed in BL21 (DE3) and C41 (DE3) using three AIM: 2×TY, SB and TB under two different temperatures: **A**. 25˚C **B.** 30 ˚C. The yellow boxes surround the target protein.

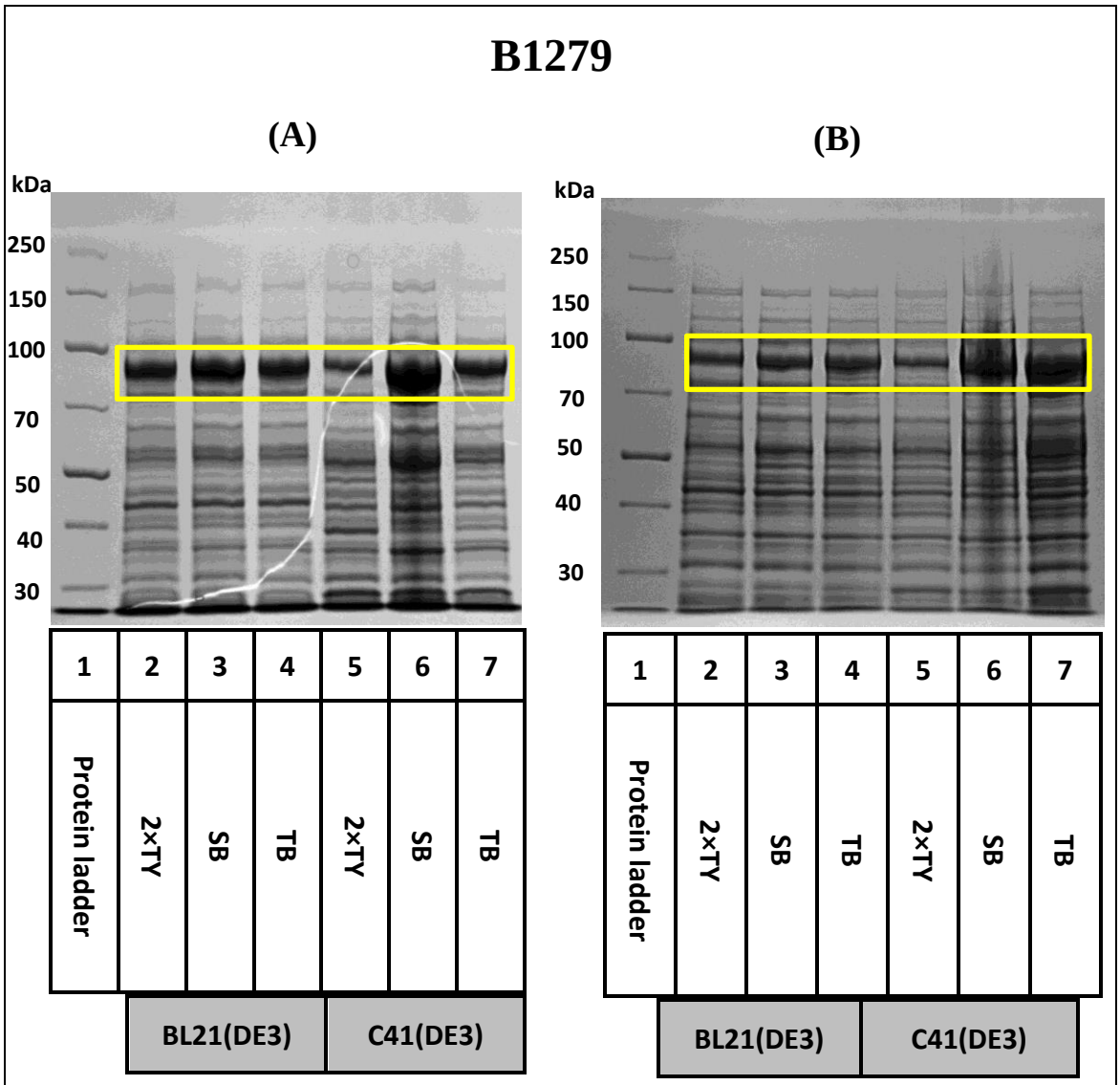


Figure 11-9: SDS-PAGE analysis of B1279 expression optimisation. The protein was expressed in BL21 (DE3) and C41 (DE3) using three AIM: 2×TY, SB and TB under two different temperatures: **A**. 25˚C **B.** 30 ˚C. The yellow boxes surround the target protein.

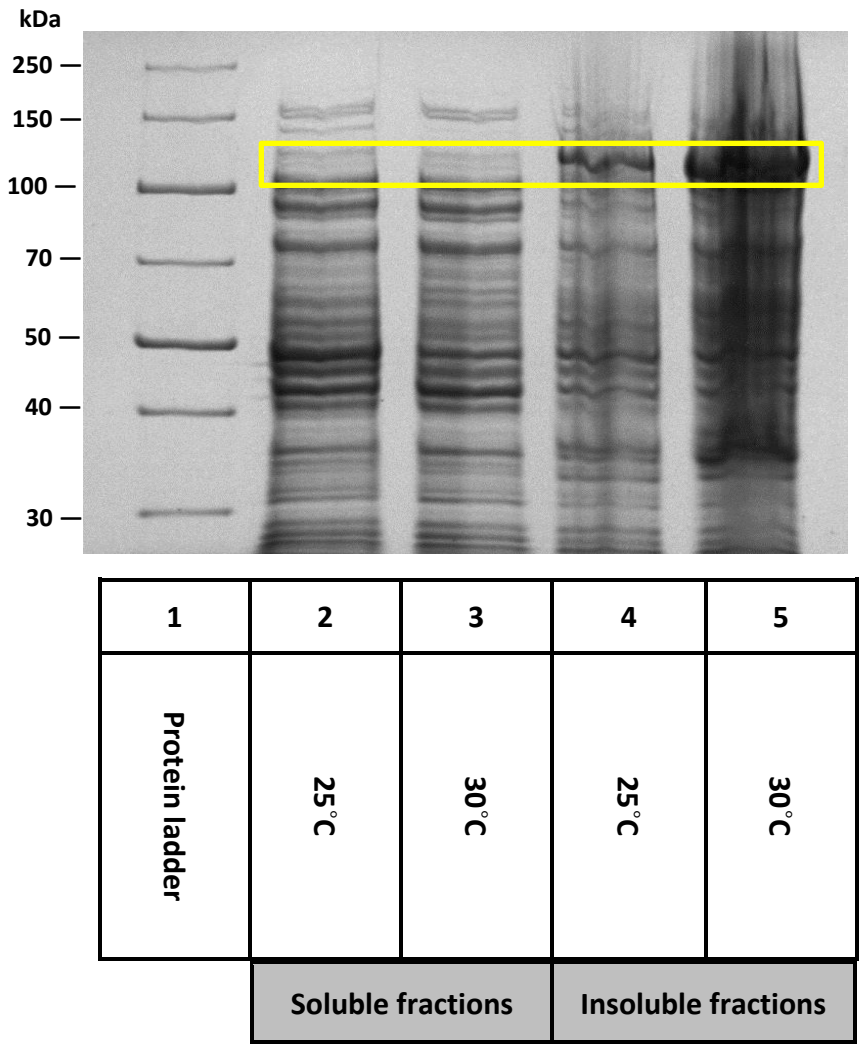


Figure 11-10: SDS-PAGE analysis of B1009 expression and solubility. The protein was expressed in *E. coli* HMS 174(DE3) using SB AIM at different incubation temperatures: 25˚C and 30˚C. The incubation was carried out for 24 hr. The yellow boxes surround the target proteins.

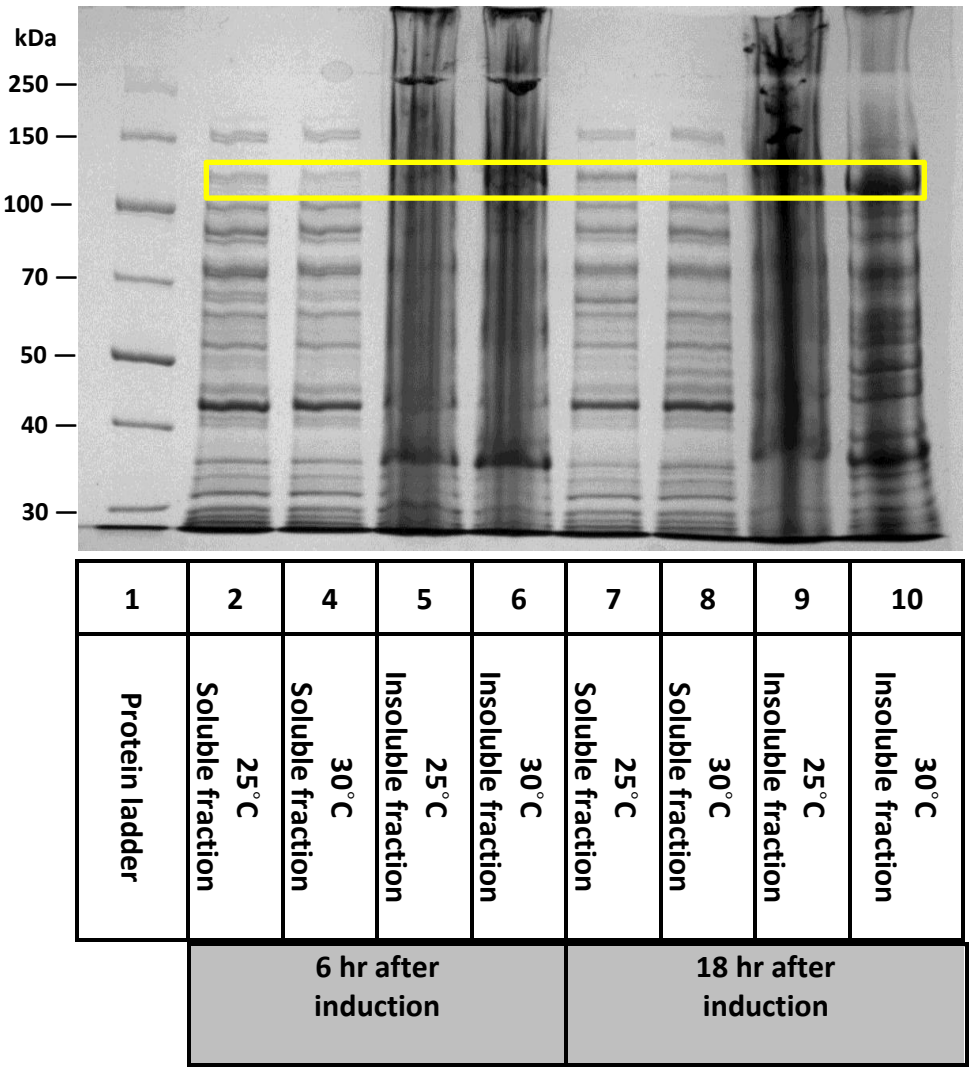


Figure 11-11: SDS-PAGE analysis of B1009 expression and solubility. The protein was expressed in HMS 174(DE3) using 100 µM IPTG at different incubation temperatures: 25˚C and 30˚C. The incubation was carried out for 6 hr and 18 hr after induction to investigate the solubility. The yellow boxes surround the target protein.

11.3 Large-scale expression and purification

The production of the three *Cupriavidus necator* H16 P450s successfully expressed in *E. coli* BL21 (DE2) was scaled up to 400 ml cultures. The expression was followed by three purification steps, affinity chromatography, ion exchange chromatography and gel filtration chromatography, as described in section 7.2.

The chromatograms and the SDS gel pictures of the purified P450s, indicated a good yield and purity of all three proteins after the affinity step as the first step of purification (Figure 11-12). In the second step (ion exchange), the three proteins were successfully concentrated (Figure 11-13) in preparation for the final step of purification, gel filtration chromatography to purify proteins depending on their size. The concentration and the purity of B2406 and B1743 after the final purification step were sufficient to proceed with the functional study (Figure 11-14 and Figure 11-15). However, while the third protein was of satisfactory purity, the concentration was still too low (1.6 nM in comparison to 17 and 20 nM for B1743 and B1279 respectively), Figure 11-16

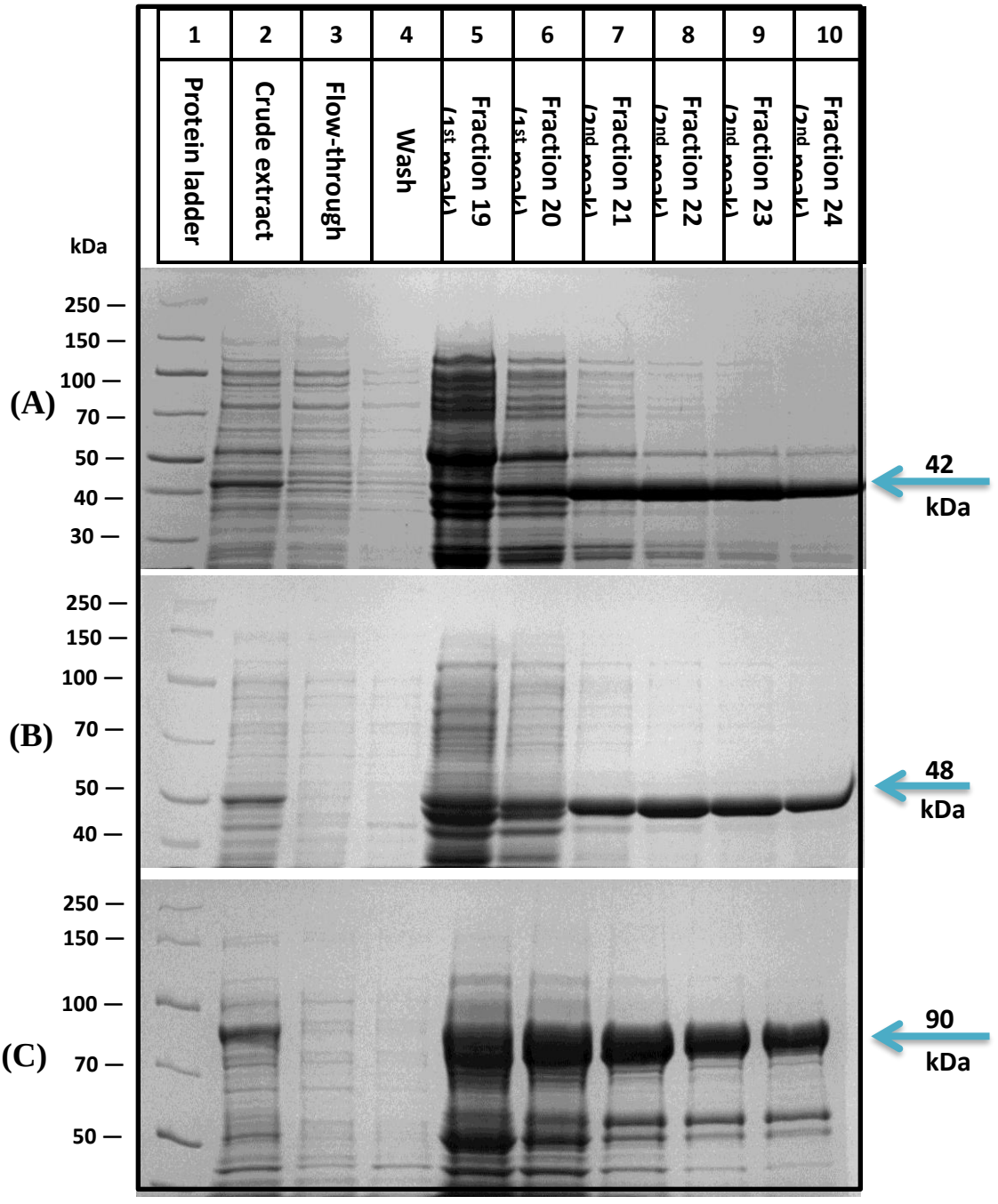


Figure 11-12: SDS-PAGE protein analysis for the first peak (19-20) and the second peak (21-24) from protein purification using affinity chromatography. The purification was carried out for 400 ml expression culture of: **A.** B2406 **B.** B1743 and **C.** B1279. Gradient elution was used, the flow rate was 2 ml/min for 40 min.

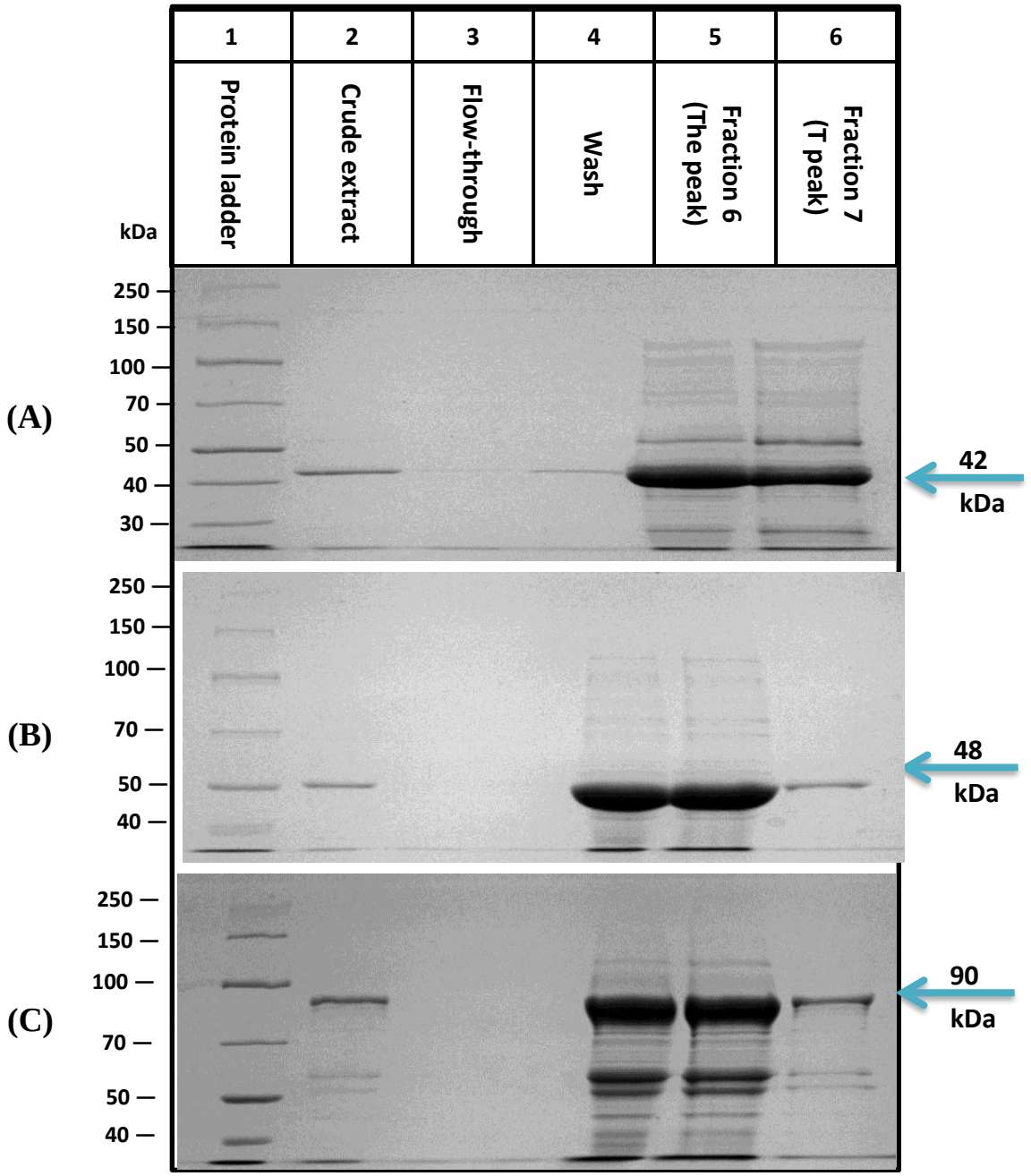


Figure 11-13: SDS-PAGE protein analysis for the peak (6 and 7) from protein purification using ion exchange chromatography of three P450s **A.** B2406 **B.** B1743 and **C.** B1279. step elution was used, the flow rate was 2 ml/.

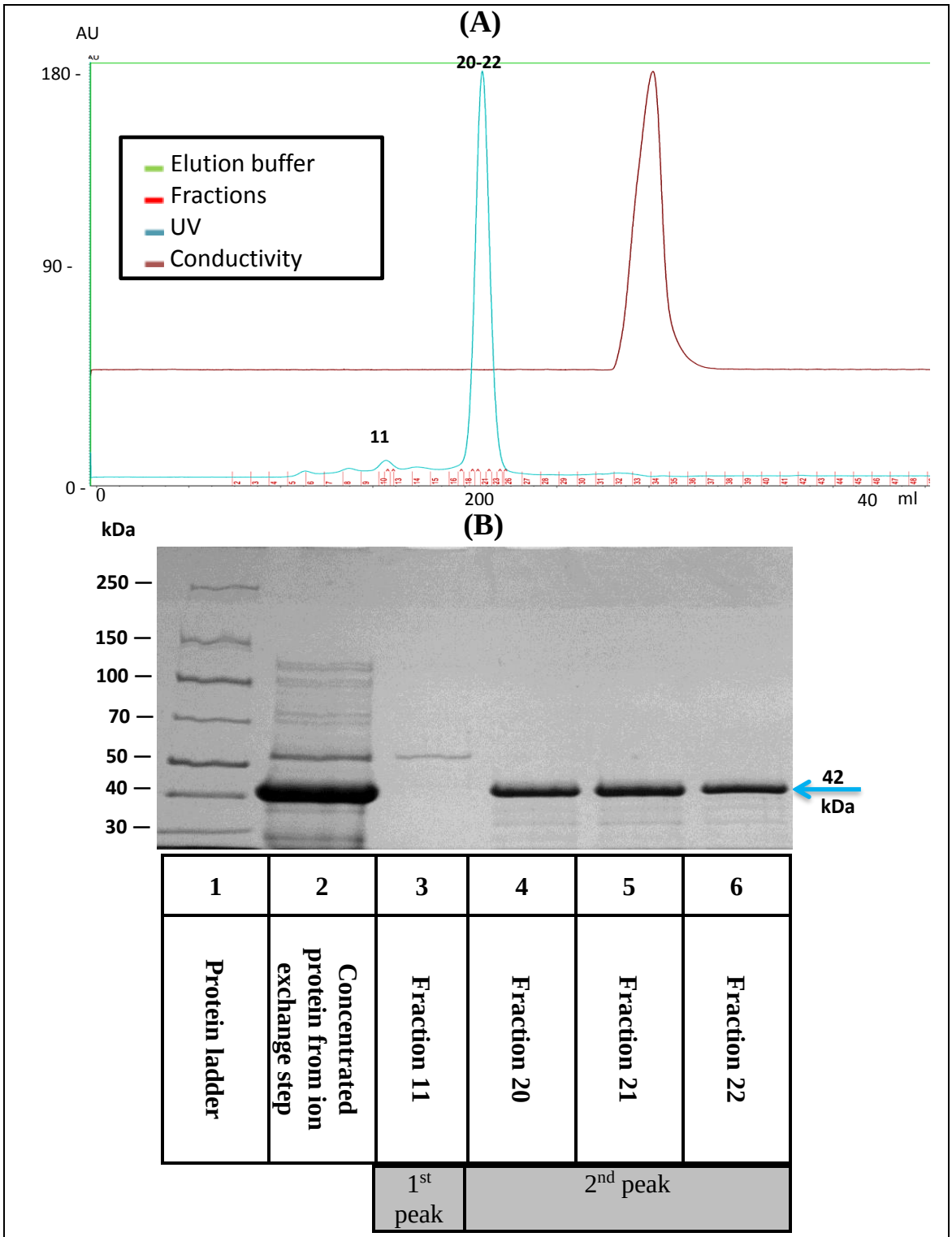


Figure 11-14: A. Chromatogram monitored during the large scale purification of H16-B2406 by gel filtration chromatography**. B.** SDS-PAGE protein analysis for the first peak fraction 11 and second peak fractions (20-22).

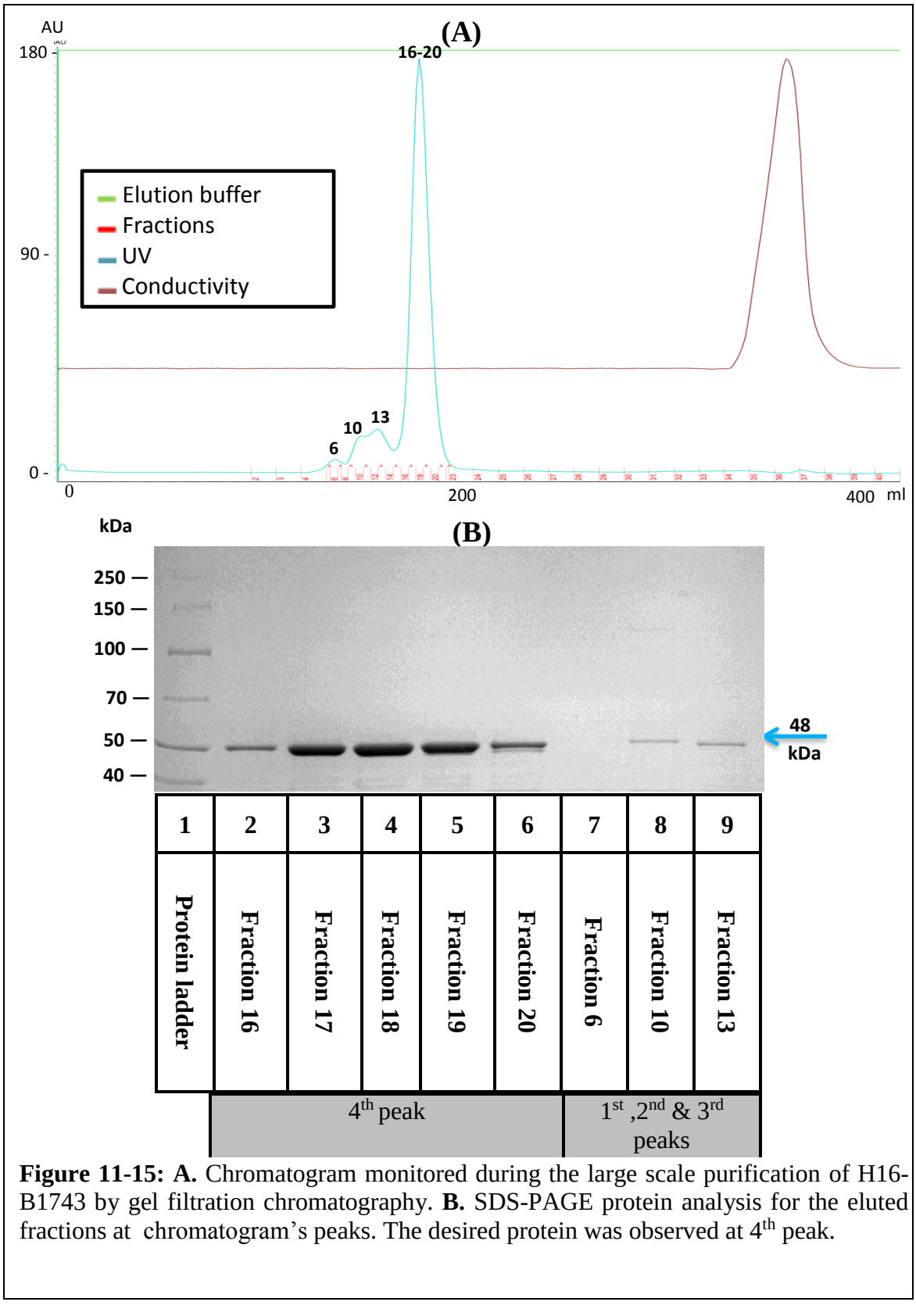
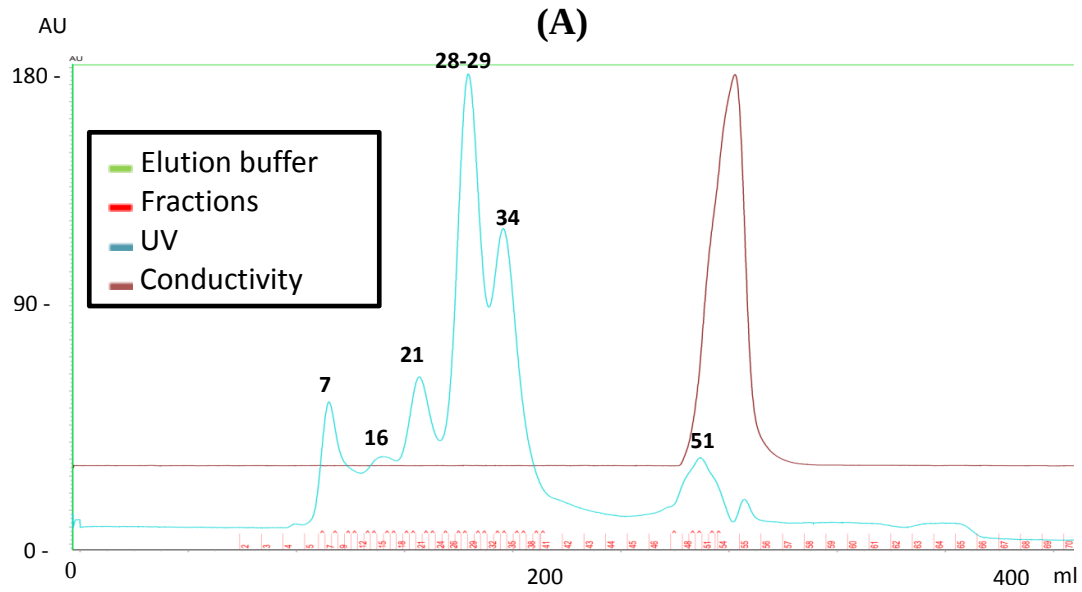


Figure 11-15: A. Chromatogram monitored during the large scale purification of H16-B1743 by gel filtration chromatography. **B.** SDS-PAGE protein analysis for the eluted fractions at chromatogram's peaks. The desired protein was observed at 4th peak.



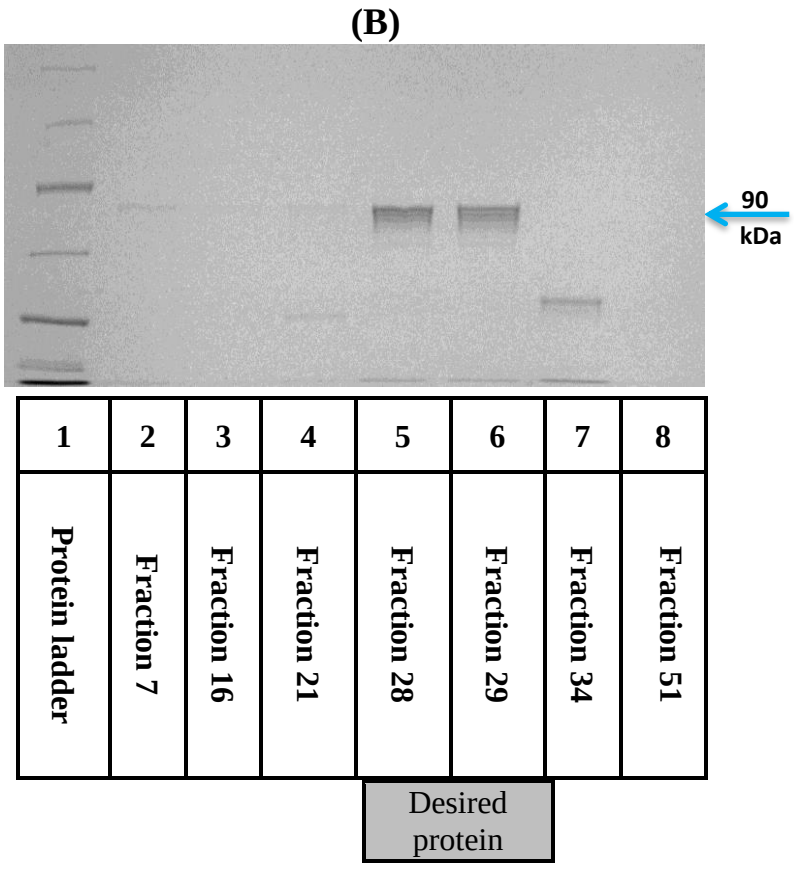


Figure 11-16: A. Chromatogram monitored during the large scale purification of H16-B1279 by gel filtration chromatography. **B.** SDS-PAGE protein analysis for the eluted fractions at chromatogram's peaks. The desired protein was observed at 4th peak.

11.4 Optical properties

The chromatogram and protein electrophoresis analysis provided an initial judgement on the protein concentrations and purities, but for more accurate visualisation, the absorbance of these proteins were scanned between 250–800 nm for more comprehensive data regarding the protein features, concentrations and purity. Figure 11-17 displaying the spectra of B2406, B1743 and B1279, all three proteins showed typical features of heme cofactors of P450s with a Soret peak at 416 nm for B2406 and 418 nm for B1743 and B1279. Alpha/beta bands were also observed at 571/535 nm, 569/537 nm and 567/533 nm for B2406, B1743 and B1279 respectively. In addition, the estimated concentrations of proteins depending on the UV absorbance and the extinction coefficients of the proteins were calculated from the spectrum values using the Beer-Lambert law (equation 6.1), which were 20, 17 and 1.6 nM for B2406, B1743 and B1279 respectively. The concentration for the later protein could be a barrier to verifying the binding of substrates and identifying potential substrates, which prompted further attempts to improve the purification of this protein.

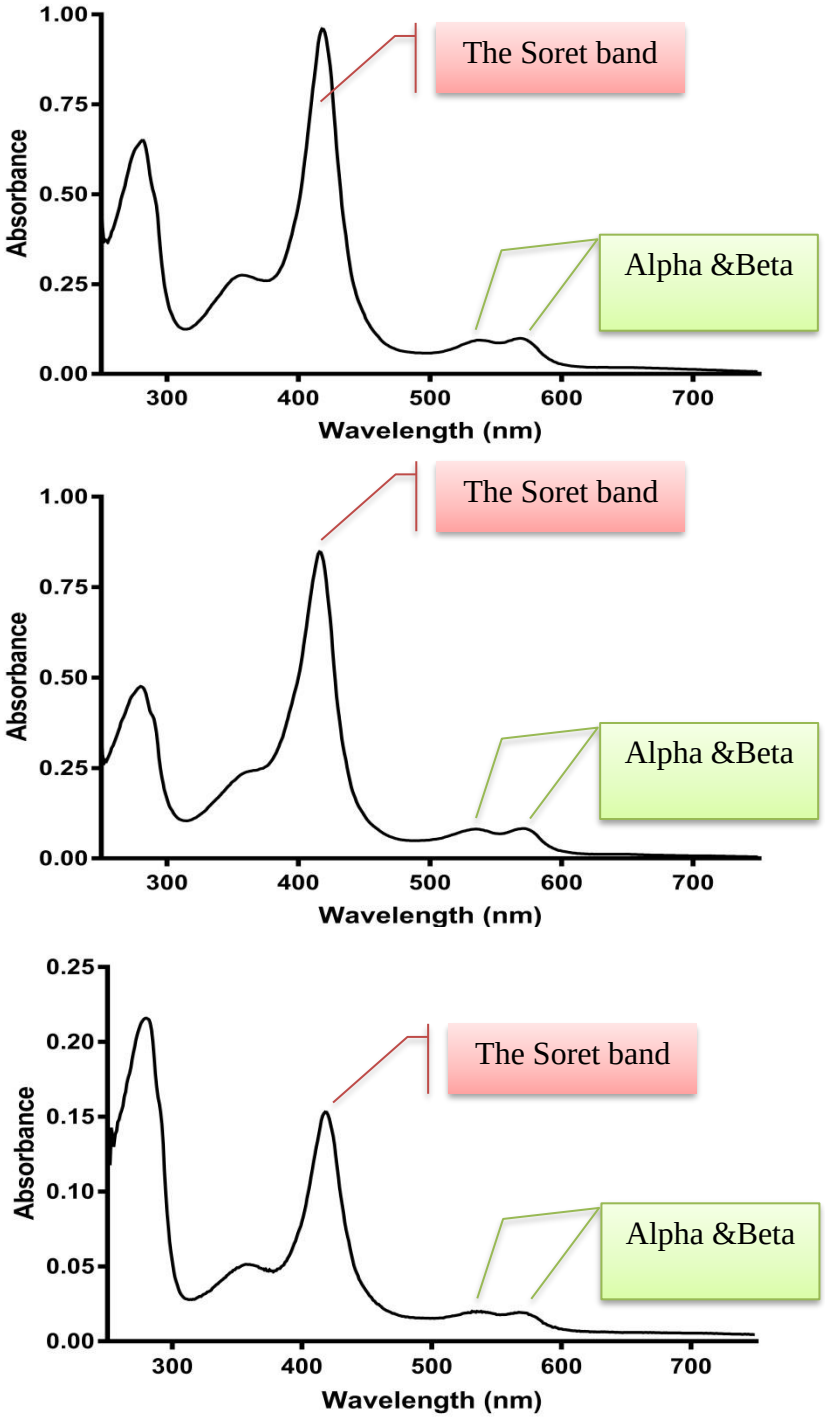


Figure 11-17: Spectra measurement for three proteins from *Cupriavidus necator* H16; **A.** H16-B2406 **B.** H16-B1743 **C.** H16-B1279. Proteins were scanned for the wavelength 250-750 nm. All three proteins showed typical features of heme cofactors of P450s with a Soret peak at 416 nm for B2406 and 418 nm for B1743 and B1279. Alpha/beta bands were also observed at 571/535 nm, 569/537 nm and 567/533 nm for B2406, B1743 and B1279 respectively

11.5 B1279 purification optimisation

With regard to the SDS gel pictures of B1279 (Figure 11-6), it was noticed that the concentration of this protein was high in the crude extract, but then decreased during the subsequent purification steps, indicating that it is the purification process that requires improvement, not the expression. Many attempts were made to overcome this problem. Firstly, the ion exchange step was omitted to reduce the amount of protein lost during the purification process, but this did not increase the protein concentration (Figure 11-18). Secondly, the elution in the ion exchange step was changed from a step elution to a gradient elution and the gel filtration was omitted as the protein is higly diluted in this final step. However, as shown in Figure 11-19, the protein after ion exchange was not pure and applying gel filtration chromatography is vital. Next, the purification of B1279 expressed in another *E. coli* strain C41 (DE3) was investigated, but this protein was also associated with a higher yield of untargeted protein at approximately 50 kDa (Figure 11-20). Other attempts to enhance B1279 concentration, such as increasing the amount of imidazole in the affinity elution buffer and applying ultrafiltration after the final purification step were not effective.

A comparison of the estimated concentrations of B1279 expressed in both BL21 (DE3) and C41 (DE3) after purification and B1743 protein is presented in Table 11-1. Both proteins have a similar concentration in the crude extract but the recovery of B1279 was nearly half that of B1743 after each purification step, leading to a decrease in concentration from 5.2 nM in the crude extract to 1.6 nM after the gel filtration step. This is probably due to the high hydrophobicity and tendency of this protein to non-specific bind to the matrix materials of the purification column and filtration membranes.

Due to the low concentration of B1279, it was decided not to proceed with the investigation of substrates binding to this protein, which requires a high protein concentration to indicate the changes in P450 features as a result of substrate binding. Instead, the activity of this protein to catalyse the reaction of different saturated, unsaturated, aromatic and herbicides substrates was assessed.

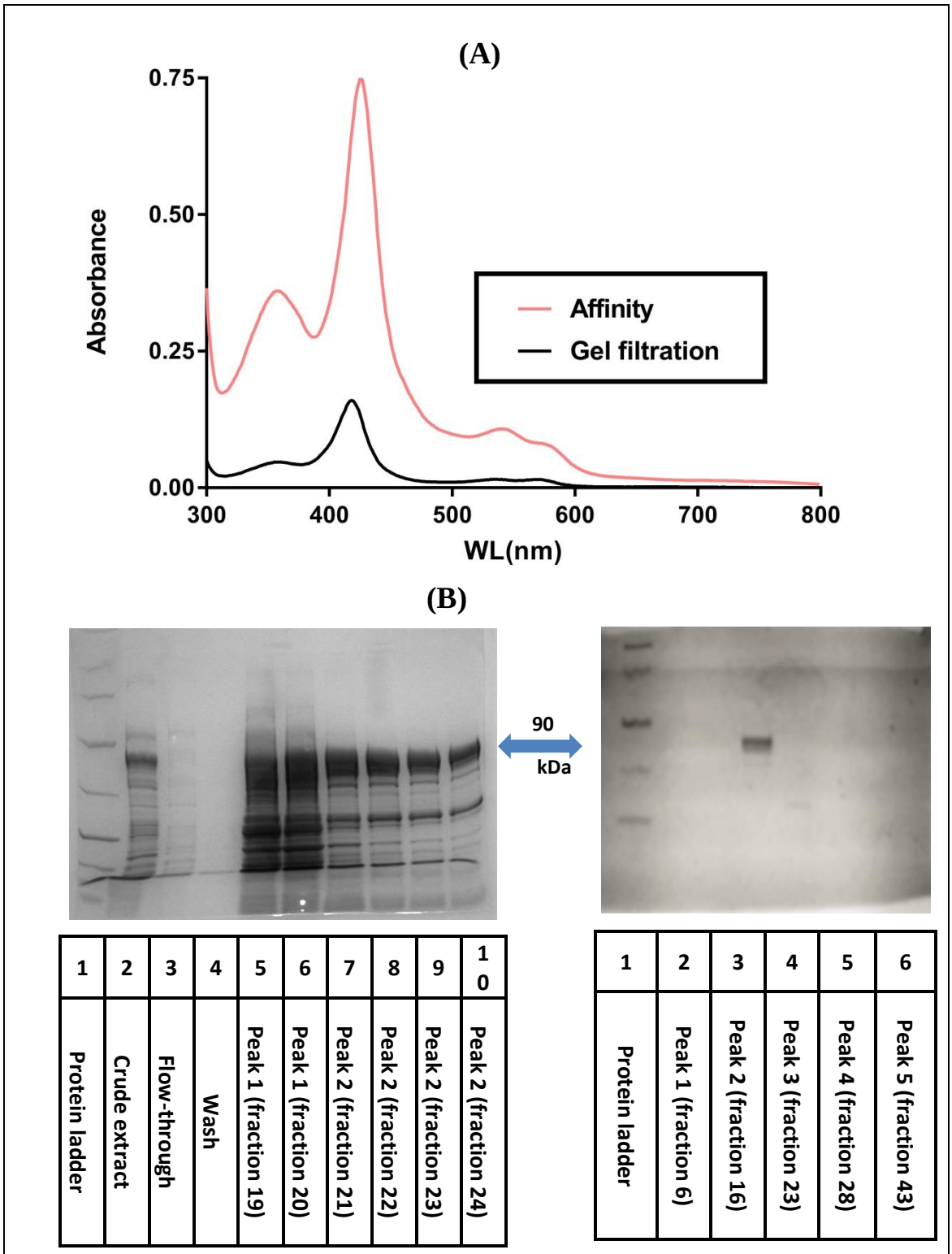


Figure 11-18: The first attempt at B1279 purification optimisation. The effect of skipping the second purification step; ion exchange on the yield of B1279 was checked by. A. spectra scanning B. SDS electrophoresis analysis.

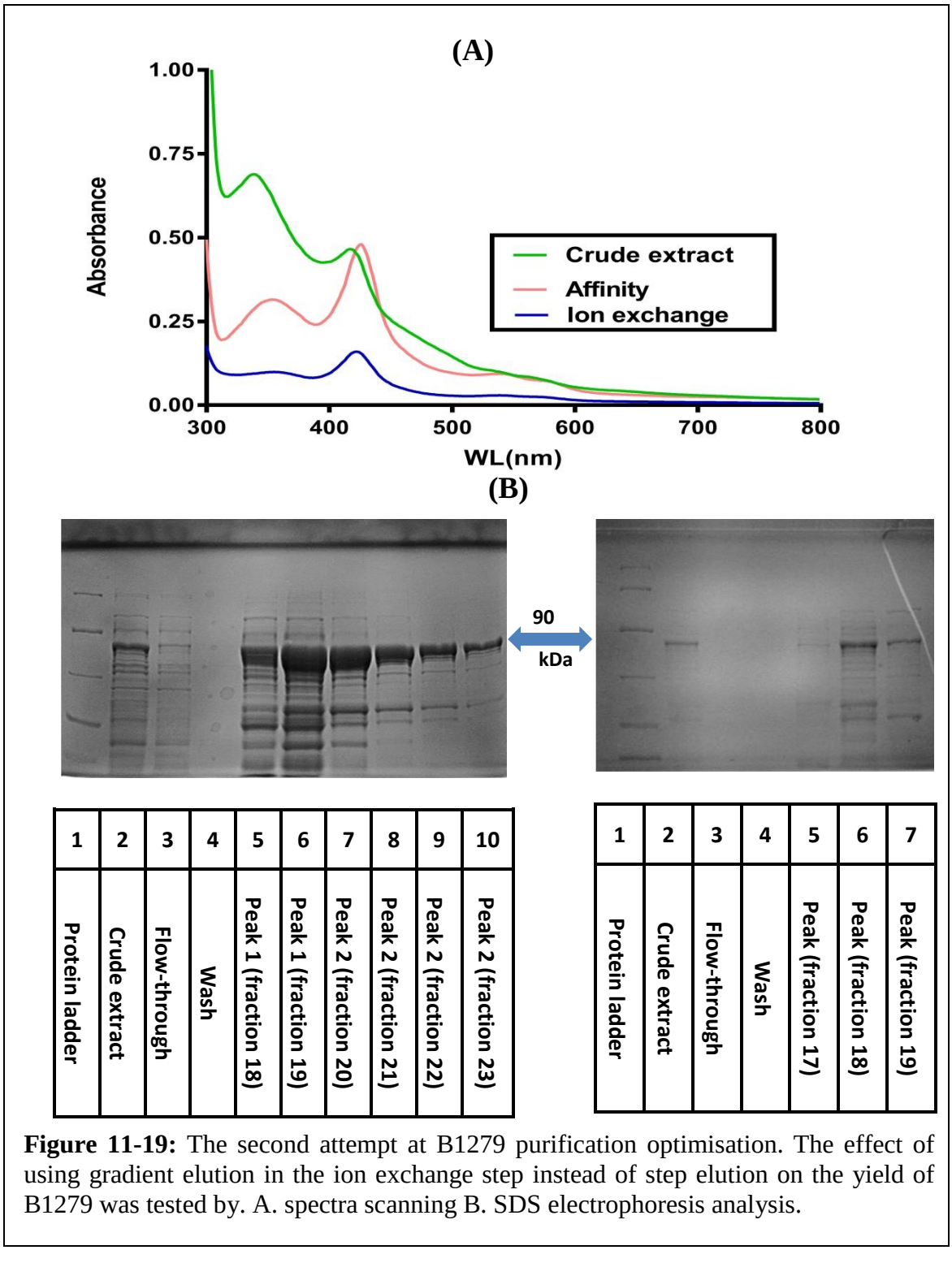


Figure 11-19: The second attempt at B1279 purification optimisation. The effect of using gradient elution in the ion exchange step instead of step elution on the yield of B1279 was tested by. A. spectra scanning B. SDS electrophoresis analysis.

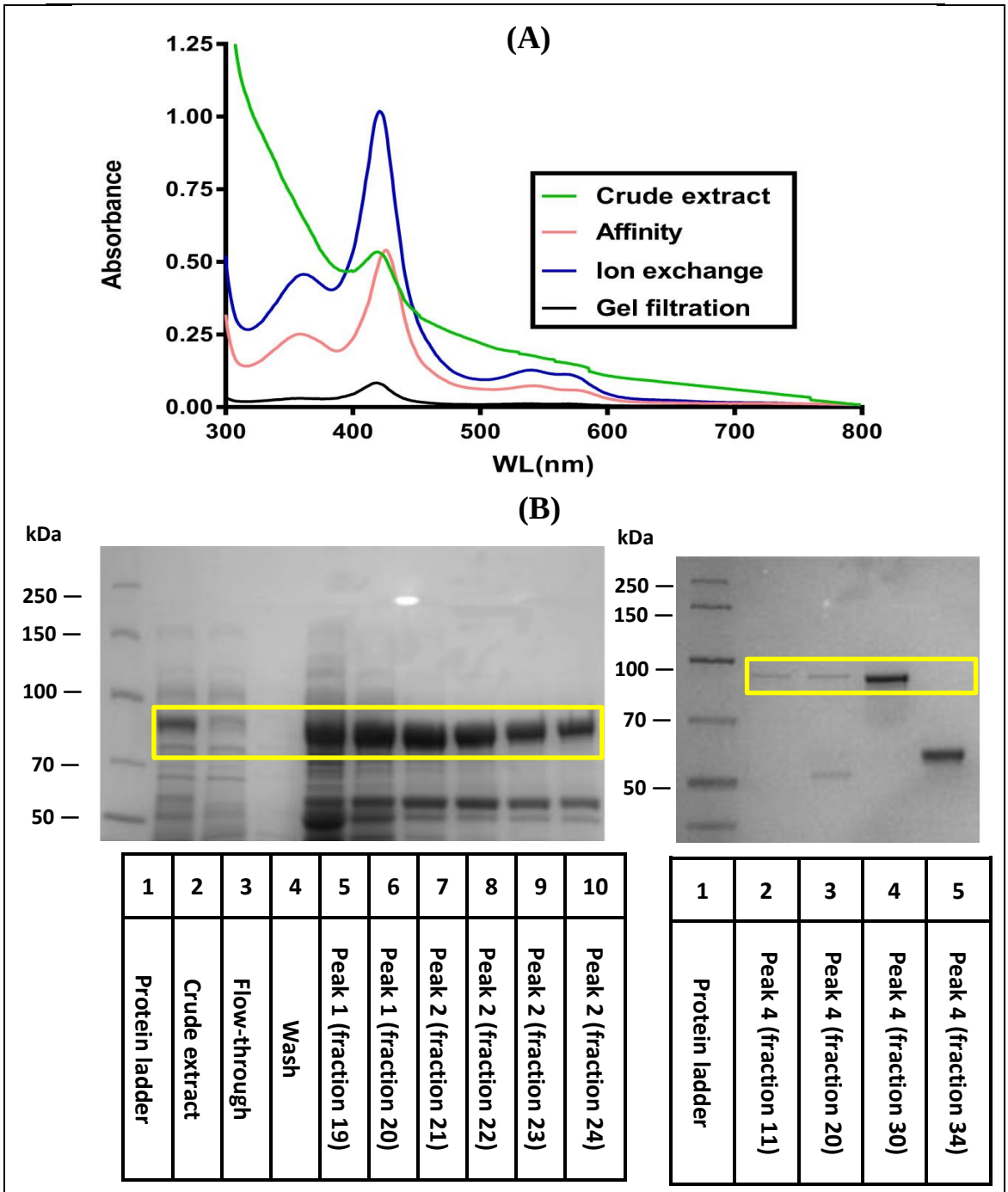


Figure 11-20: The third attempt at B1279 purification optimisation. B1279 was expressed in C41 (DE3) instead of BL21 (DE3) to evaluate the yield of B1279. The results were analysed by: A. spectra scanning B. SDS electrophoresis analysis. The yellow boxes indicate the target protein.

Table 11-1: Comparison of the purification of B1279 in BL21 (DE3) and C41 (DE3) and the purification of B1743

Purification step	B1279 (90 kDa)						B1743 (48 kDa)		
	Absorbance (soret peak)		Volume (ml)		Recovery (%)		Absorbance (soret peak)	Volume (ml)	Recovery (%)
	BL21 (DE3)	C41 (DE3)	BL21 (DE3)	C41 (DE3)	BL21 (DE3)	C41 (DE3)			
Crude extract	0.47	0.53	100	100	-	-	0.56	100	-
Affinity	0.48	0.54	14	14	14	14	1.13	14	28
Ion exchange	0.88	1	4	3	52	54	2.60	4	73
Gel filtration	0.16	0.08	9	15	48	31	0.98	9	81

11.6 Substrate binding

when the binding of 11 substrates to B2406 and B1743 proteins was assessed as described in section 9.10, no significant optical changes in the B2406 spectra were detected with any of the 11 examined substrates, while type I shifts were observed in the B1743 spectra with two substrates, lauric acid and 3-phenoxytoluene as shown in Figure 11-21.

The binding of 3-phenoxytoluene to B1743 induced a very small shift in the Soret band from 418 to 417 (blue shift), reducing the UV absorbance to 0.69 nm instead of 0.71 nm in the free substrate protein. Similar spectral changes were observed when 3-phenoxytulen was titrated to the B1743 solution; a blue shift towards 390 in the Soret maximum and the UV value decreased from 0.70 nm to 0.57 nm.

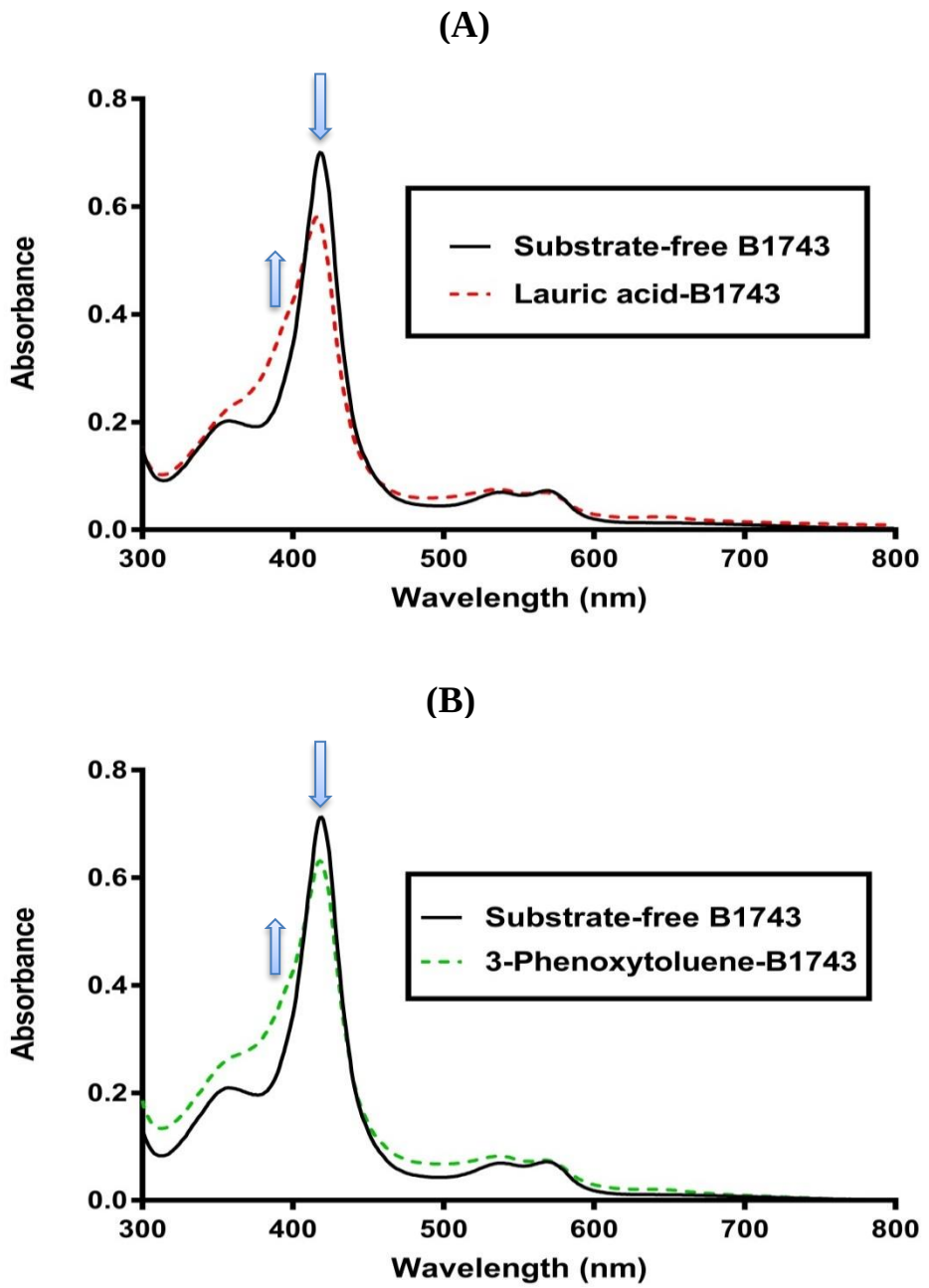


Figure 11-21: Analysis of two substrates binding to B1743. **A.** Spectral titration for B1743 with lauric acid. **B.** Spectral titration for B1743 with 3-phenoxytoluene. Arrows indicate the progressive decrease in the ferric LS Soret band (at 418 nm) and the associated increase in the ferric HS feature at about 390 nm

11.7 Activity measurement

Among the three P450s that have been successfully expressed in *E. coli* and purified during this project, B1279 was selected to be functionally analysed. In contrast to B2406 and B1743, B1279 is the only protein that contains a reductase module (FMN and 2Fe-2S) fused to its heme domain according to the genome analysis and 3D structure modelling in chapter 10, thereby increasing the probability of B1279 being catalytically active as a self-sufficient protein.

The activity evaluation of B1279 showed that the consumption rates of the redox partner, NADH, were very low, approaching its rate in the background reactions (0.048 nmole/s) with all substrates (lauric acid, palmitic acid, myristic acid, butyric acid, adipic acid, malate, indole, naphthalene, 3-phenoxytoluene, EPTC and vernolate) (Figure 11-22). The standard probability value "p-value" was more than 0.05 for all tested substrates. In biology, difference between two groups considered significant if the p-value < 0.05 (Bailey, 1995). When NADPH was used instead of NADH as electrons donor, five substrates showed a significant consumption of NADPH with p-value > 0.05. These substrates as can be seen in Figure 11-23are: indol, EPTC, malonic acid, adipic acid and 3-phenoxytolouene. As a result, B1279 is likely to be catalytically active with these five substrates.

In general, the low catalytic activity may be because the protein was extensively diluted during the purification process. Losing activity as a result of extensive dilution may be due to either monomerisation of the enzyme domains which could affect the transformation of the required electrons from the cofactor to the active site or to the FMN depletion if the enzyme was incubated for an extended period in the dilute solution (Whitehouse et al., 2012)

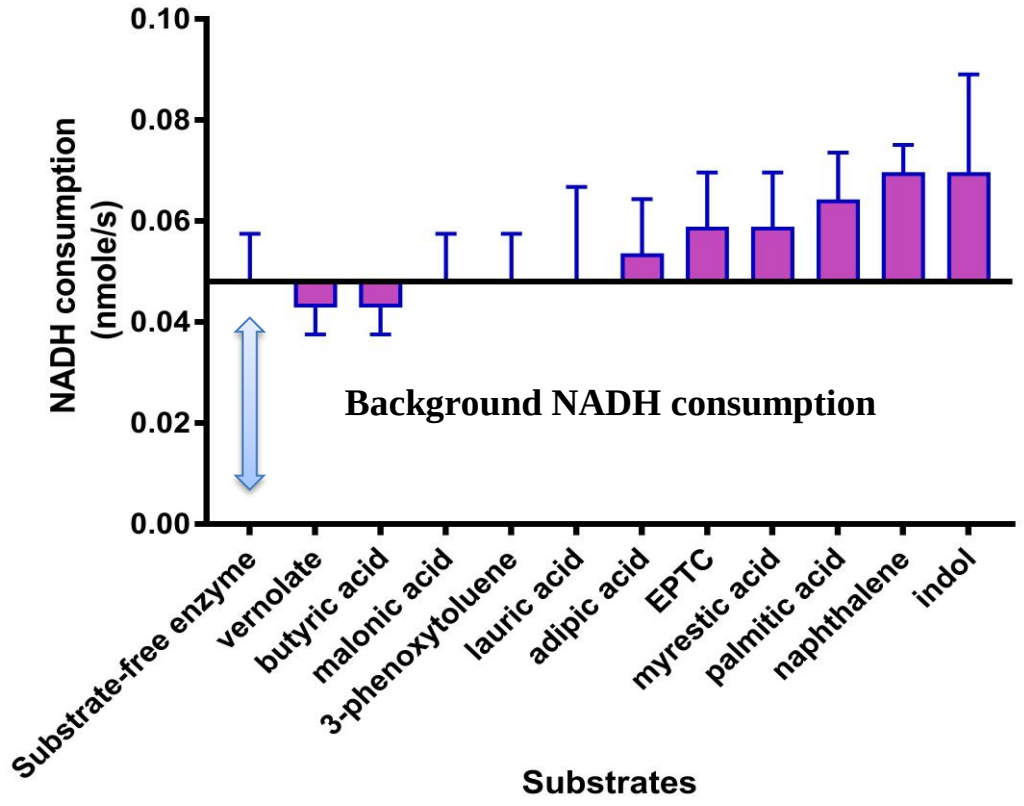


Figure 11-22: B1279 activity analysis towards eleven substrates. The bars represent the NADH consumption rate as a result to use the cofactor during the reaction. The NADH consumption indicates an interaction with the protein that is directly proportional to the reaction rate. Three independent experiments were performed with triplicates with error bar in the standard error of the mean (SEM). The calculation of the p-value showed that all the results from this experiment is not significant with p-value >0.05.

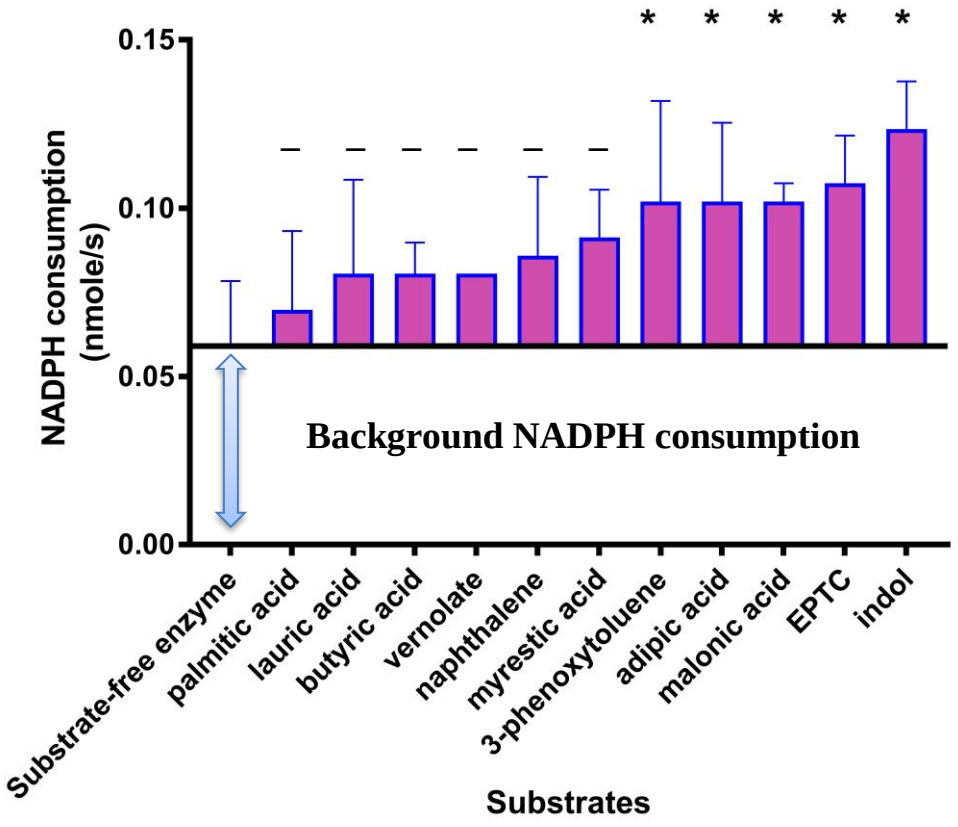


Figure 11-23: : B1279 activity analysis towards eleven substrates. The bars represent the NADPH consumption rate as a result of using the cofactor during the reaction. The NADPH consumption indicates an interaction with the protein that is directly proportional to the reaction rate. Three independent experiments were performed with triplicates with error bar in the standard error of the mean (SEM). The star above each bar (*) indicates that the result is significant with p-value <0.05 and the (-) shows that the result is not significant with p-value >0.05.

11.8 Conclusion

Although numerous bacterial cytochrome P450 enzymes have been identified and their structures and functions extensively investigated, only one cytochrome from *Cupriavidus sp.* has been studied structurally and functionally, CYP116B1from *Cupriavidus metallidurans* (Warman et al., 2012). In this chapter, the first characterisation, absorbance, relevant substrates and activity of four cytochrome P450s from *Cupriavidus necator* H16 were reported.

A PROSITE online tool and CLUSTALW multi-alignment of the amino acid sequences of these proteins with other P450 sequences from BLAST showed the conserved regions and the level of similarity. The homology search indicated similarity between three *C. necator* H16 P450s and the self-sufficient P450, CYP116B1. This protein showed 82%, 75% and 55% similarity with B1279, B1009 and B2406 respectively, but there was no similarity with B1743 in the top twenty BLAST sequences.

The generated 3D models of four C. necator H16 P450s showed a typical P450 fold with α helices, β sheets and loops. From Figure 10-5to Figure 10-7 and Figure 10-9, it is evident that the heme cofactors of all proteins are buried deeply in the heme domain structures where the heme cofactor would be expected (heme from the crystal structure of templets were used for this purpose) suggesting no aberration in the P450 heme domains for all four proteins. Generally, all heme domain structures displayed a typical P450 fold with expected helical features (α helices, β sheets and loops), structures' images showed that these four proteins are probably monomers, folded into α/β domains characteristic of P450s.

From overlapping the predicted models of each heme domain and reductase domain to its template, a high similarity was observed and this similarity was evident as the majority of the secondary structure elements merge together to structure modules due to this absolute conservation. However, some differences in N-terminal, C-terminal and loops. Differences in these regions of the P450 structure was reported previously and it could be due to the highly variable in the primary sequence of these regions (Sirim et al., 2010). The calculated RMSD between all predicted models and the crystal structures of their templates were less than 2 Å and this low RMSD indicates the high accuracy of the generated model (Carugo, 2007). As well as the quality of all models were checked by the quality verifying online tool (VERIFY 3D) and more than 80% of all models' residues showed average (3D-1D) $\geq$ 0.2.

The structure analysis of the hydrophobic residues that bound to the identified heme cofactor in the heme domains of B2406, B1743, B1279 and B1009 showed a very high conservation between the templets' and the models' binding residues. This high conservation also observed by sequences alignment of all models with the sequences of their templates. However few differences were noticed in some limited cases between binding residues of the generated models and templates but these residues still occupied same positions and have very similar functions to each other. The primary and the secondary structure analysis of four P450 from *C.necator* H16 leads to the belief that the reliability of all predicted models are high and the prosthetic groups (heme) in all heme domains are likely to be bound to the structure. Therefore, it is believed that all C. necator H16 P450 heme domains would be functionally active if they fused naturally or genetically to a proper redox partner.

In B1279 3D model reductase, a typical structure of ferredoxin reductase was clearly observed by the identification of FMN and Fe-S binding motifs with the presence of four Cys

residues in the 2Fe-2S creating a square plane to coordinate the cofactor (shown as red stick in Figure 10 7). The high conservation of the residues bound to both ligands probably facilitated the binding of both prosthetic groups FMN and Fe-S in the model to the structure, thereby catalysing the transformation of electrons from a donor (NADPH or NADH) to the active site (heme cofactor) of B1279. The existence of the active site represented by a typical P450 heme cofactor connected to the structure and the fact that this heme domain naturally fuses to the ferredoxin reductase containing two prosthetics groups (FMN and Fe-S) suggests that this protein may be catalytically active.

The 3D model of the reductase domain of B1009 showed three prosthetics groups: a flavodoxin-like domain FMN, ferredoxin reductase-type FAD binding domain (FAD_FR) and nicotinamide-adenine-dinucleotide phosphate (NADP). These three ligands were shown as sticks in Figure 10-11and together formed a typical reductase structure (detailed analysis of the prosthetics groups of the reductase domain of B1009 is absence from this study).

Both B2406 and B1743 successfully expressed using *E. coli* BL21(DE3) and SB AIM at 30°C and the high yield obtained was efficiently purified by three chromatography steps (affinity, ion exchange and gel filtration) without any further optimisation. Optimisation of B1279 expression was attempted using different *E. coli* strains, inducer agents and expression temperatures. However, during the purification process, a large amount of B1279 was lost and it was diluted extensively, with recovery percentages of this protein of 14%, 52% and 31% after affinity, ion exchange and gel filtration respectively. The recovery percentages for the protein expressed in *E. coli* BL21(DE3) were similar to those when *E. coli* C41(DE3) was used. Despite the many changes applied, unfortunately the recovery of B1279 remained low. The analysis of the results from the B1279 purification and application of a step to step

comparison with the B1743 purification (Table 11-1), indicated that the low recovery of protein was probably due to the protein nature and tendency to non-specifically bind to the purification column matrix and filtration membranes.

The expression of B1009 in *E. coli* BL21(DE3) and *E. coli* C41(DE3) was very low and the numerous unsuccessful attempts to enhance its expression in these two strains, such as different expression temperatures, different inducers and various media gave the motivation to examine another *E. coli* strain. An enhancement was observed when B1009 was expressed in *E. coli* HMS174(DE3) by using AIM SB at 30°C. However, the expressed B1009 in *E. coli* HMS174(DE3) showed low solubility.

The light absorbance scanning of B2405, B1743 and B1279 between 250-800 nm showed typical features of heme cofactors of P450, confirming the sequence analysis and 3D structure findings that these three proteins are new members of the P450 monooxygenases superfamily. With regard to potential substrates of B2406 and B1743, changes in the spectra of the lauric acid-B1743 and 3-phenoxytoluene -B1743 complexes were observed. B1279 was excluded from the binding experiments because it was unlikely to observe any change in its spectra due to the low concentration after the purification process. Instead, B1279 activity was investigated with all 11 substrates using NADPH and NADH as cofactors proteins, showed possible activity with indol, EPTC, malonic acid, adipic acid and 3-phenoxytolouene as indicated by the increase in NADPH consumption rate in substrate-enzyme solution in comparison to SF B1279 solution.

12 Conclusion and Future Prospects

12.1 Cytochrome P450 BM-3

Despite the fact that P450 BM-3 has been the subject of study for many years, only a few researchers have reported the effect of co-solvents on BM-3 activity. Previously, BM-3 tolerance was investigated in traditional solvents which were classified as hazardous materials according to many classifications such as the one published by GSK in 2010 (Henderson et al., 2011). In recent years, using alternative green solvents has earned the attention of both industrialists and academics due to their environmental advantages including waste reduction, hazard elimination and process economy improvement.

In part II of the project, I used three green solvents, 1-butanol, 2-butanol and dimethyl carbonate, to dissolve one of the most common insoluble fatty acid substrates, lauric acid, for hydroxylation catalysed by P450 BM-3. The activity of the WT and the W5F5 variant were compared in the presence of these solvents. The W5F5 variant was chosen owing to its tolerance to many co-solvents (Wong et al., 2004). To the best of my knowledge, this is the first report of the P450 BM-3 tolerance against green solvents. The results from this part of the project (Part II) showed that dimethyl carbonate is a very good green replacement of the traditional hazardous solvents, as the relative activity of the P450 BM-3, both WT and W5F5, was 31% and 128% respectively at 5% (v/v) solvent.

Both the wild type and its mutant W5F5 showed very low tolerance towards other green solvents tested, 1-butanol and 2-butanol. It is necessary to apply further steps in the future towards improving BM-3 resistance against these green solvents. This could be achieved by

applying one or more of protein stabilisation strategies such as isolation of stable biocatalysts, immobilisation, chemical modification, ionic liquid coating and genetic modification (see section 5.7 for more detail).

12.2 Putative cytochrome P450s from *Cupriavidus necator* H16

This part of the thesis presented a detailed analysis of four novel cytochrome P450s: H16_B2406, H16_B1743, H16_B1279 and H16_B1009. All these proteins are natural cytochromes from *Cupriavidus necator* H16. This bacterium can grow heterotrophically and autotrophically and is an excellent platform for the production of degradable bioplastic and value-added compounds. Three of these four proteins, B2406, B1743 and B1279, were successfully expressed in *E. coli* and purified using three chromatograph steps. The optical analysation of these three proteins confirmed typical features of heme cofactors of cytochrome P450, with a Soret peak at 416 nm for B2406 and 418 nm for B1743 and B1279, as well as alpha/beta at 571/535, 569/537 and 567/533 for B2406, B1743 and B1279 respectively.

12.2.1 P450 B2406 and P450 B1743

The sequence alignment of B2406 and B1743 with other P450s provided by the database and the predicted 3D structures of these proteins showed the heme cofactors buried deep in the heme domain structure. Also, these results confirmed that both of these proteins are missing any type of reductase modules rendering them catalytically inactive.

With regard to potential substrates of B2406 and B1743, no binding of any the 11 substrates tested was demonstrated by B2406, while lauric acid and 3-phenoxytoluene showed small blue shifts of the Soret bands towards 390 nm (Type I) when mixed with B1743. The binding

between these two substrates and B1743 suggests that this protein could be catalytically active if fused in future to an appropriate redox partner to generate a new multi-domain construct. The molecular Lego approach was used previously and proved to be successful for the enhancement of catalytic properties of P450 BM-3 heme domain. Flavodoxin from *Desulfovibrio vulgaris* (FLD) and BM-3 heme domain from *Bacillus megaterium* (BMP) were fused genetically to produce a BMP–FLD fusion protein. The catalytic activity of the BMP–FLD fused enzyme was improved by six fold when it was used to hydroxylate p-nitrophenol (Fantuzzi et al., 2006; Sadeghi et al., 2000). For future work, it is recommended to use genetic approach to produce catalytic active fused enzyme.

12.2.2 P450 B1279

The genetic analysis of the B1279 sequence, the undefined structure and substrate selectivity cytochrome, enabled the characterisation of the heme binding cofactor, 2Fe-2S centre and FMN. This analysis confirmed the hypothesis that the heme domain of this protein is naturally fused to a reductase module. Typically, in this system, electrons are driven from a cofactor redox NADPH or NADH to the heme cofactor by the iron-sulphur centre and flavoprotein (FMN).

The functional analysis of B1279 showed its potential to be catalytically active towards five substrates; indol, EPTC, malonic acid, adipic acid and 3-phenoxytolouene according to the significant increase in the consumption rate of NADPH. However, in general, the activity was very low and that may be very well because B1279 was highly diluted during the purification steps due to its tendency to form unwanted bonds with the purification column matrices. According to Whitehouse and co-workers, there are two possible scenarios for the

loss of P450 activity in a dilute solution, it may be due to monomerisation of the enzyme domains or to the FMN depletion if the enzyme was incubated for an extended period in the dilute solution (Whitehouse et al., 2012). The phenomenon of weak activity was also indicated previously with another fused protein containing the same reductase system, (FMN-[2Fe-2S]), RhF from *Rhodococcus* sp. NCIMB9784. This protein catalysed the O-dealkylation of 7-ethoxycoumarin. Although, the consumption of the NADPH and NADH were not measured in the project, Raobert and co-workers confirmed the low level of substrate turnover depended on the fluorescent properties of the product 7-hydroxycoumarin (Roberts et al., 2003).

Purification resolution depends, not only on the nature of the experimental conditions such as pH, ionic strength and elution conditions, but also on functional groups on the column matrix, so higher efficiency purification columns may be required in furture studies. Resolution in terms of efficiency can be improved by decreasing the particle size of the matrix. For example, using the Mono Q™ ion exchange column instead of Super Q and Superdex Peptide gel filtration column as an alternative to Hiload 26/600 Superdex 200 pg may increase the purification resolution to significant levels due to the small particle sizes of these matrices. Nonetheless, it would be interesting in the future to investigate the activity of B1279 with other substrates usually metabolised by *Cupriavidus necator* H16, such as acetic acid, propionic acid, glycolate, lactate, levulinic acid and benzoate (Lu et al., 2016).

Due to time limitations and technical issues, there are many analysations yet to be performed in the investigation of B1279 activity with indol, EPTC, malonic acid, adipic acid and 3-Phenoxytoluene which will be the focus of further studies of this protein. In particular, the determination of kinetic constants (K_m and K_{cat}) using Michaelis equation, the pyridine

hemochromogen assay to detect the concentration of heme and determine the extinction coefficient of the hemoprotein sample using Beer-Lambert's law, flavin cofactor identification using HPLC, stopped-flow kinetics to determine cofactor concentration dependent enzyme reduction rates and protein concentration measurement by ferrous-CO complex formation (Luciakova, 2015; Warman et al., 2012).

12.2.3 P450 B1009

H16_B1009 was studied genetically and structurally for the first time in this project. The results proved that the structure of this protein consists of two domains, heme and reductase. The reductase domain in this protein comprises two main proteins, FAD and FMN, which help in the electron transformation from the redox partner to the heme cofactor (active site).

The expression of B1009 in two *E. coli* strains; BL21(DE3) and C41(DE3) was very low. All attempts to improve B1009 expression in these strains like using different media, temperatures, inducer agents and expression time were unsuccessful. Fortunately, A significant improvement in protein expression was observed when *E. coli* HMS174(DE3) strain was used. However, B1009 showed low solubility when *E. coli* HMS174(DE3) was used. therefore it would be interesting to investigate the improvement of protein solubility in the future by using one or more of the tradition techniques to increase protein solubility in *E. coli* such as cultivation at reduced temperatures, changing media compositions and fusing to affinity tags like maltose binding protein (MBP) and N-utilizing substance A (NusA) (Sørensen & Mortensen, 2005).

In general, these findings indicate that our knowledge regarding cytochrome P450 enzymes suggests that general rules cannot be applied to all members of this superfamily or even to

all P450s from *Cupriavidus necator* H16. Every new P540 will open the door to new products, substrates and catalytic reactions as well as to unlimited biotechnological knowledge.

References

Abu-Soud & Stuehr. (1993). Nitric oxide synthases reveal a role for calmodulin in controlling electron transfer. *Proceedings of the National Academy of Sciences, 90*(22), 10769-10772.

Ajikumar, Xiao, Tyo, Wang, Simeon, Leonard, Mucha, Phon, Pfeifer & Stephanopoulos. (2010). Isoprenoid pathway optimization for Taxol precursor overproduction in Escherichia coli. *Science, 330*(6000), 70-74.

Alagesan, Minton & Malys. (2018). 13C-assisted metabolic flux analysis to investigate heterotrophic and mixotrophic metabolism in Cupriavidus necator H16. *Metabolomics, 14*(1), 9.

Alworth, Xia & Liu. (1997). Organochlorine substrates and inhibitors of P450BM-3. *Faseb Journal, 11*(9), A804-A804.

Amarneh, Corbin, Peterson, Simpson & Graham-Lorence. (1993). Functional domains of human aromatase cytochrome P450 characterized by linear alignment and site-directed mutagenesis. *Molecular Endocrinology, 7*(12), 1617-1624.

Appel, Lutz-Wahl, Fischer, Schwaneberg & Schmid. (2001). A P450 BM-3 mutant hydroxylates alkanes, cycloalkanes, arenes and heteroarenes. *Journal of Biotechnology, 88*(2), 167-171.

Arai. (1988). Pravastatin sodium (CS-154), a novel cholesterol-lowering agent which inhibits HMG-CoA reductase. *Sankyo Kenkyusyo Nempo, 40*, 1-38.

Arnold. (1998). Design by directed evolution. *Accounts of chemical research, 31*(3), 125-131.

Bach & Chodat. (1903). Über Peroxydase. *Ber. d. Deutsch, chem. Ges, 36*, 600-605.

Bailey. (1995). *Statistical methods in biology*. Aylesbury, UK: The EnglishUniversities Press LTD.

Baker, Mcceskey, Pandey & Baker. (2004). Fluorescence studies of protein thermostability in ionic liquids. *Chemical communications*(8), 940-941.

Baneyx. (1999). Recombinant protein expression in Escherichia coli. *Current opinion in biotechnology, 10*(5), 411-421.

Bell, Janssen & Halling. (1997). Water activity fails to predict critical hydration level for enzyme activity in polar organic solvents: Interconversion of water concentrations and activities. *Enzyme and Microbial Technology, 20*(6), 471-477.

Bennett, Herbert, Sternberg & Kelley. (2008). Exploring the extremes of sequence/structure space with ensemble fold recognition in the program Phyre. *Proteins: Structure, Function, and Bioinformatics, 70*(3), 611-625.

Berberich, Kaar & Russell. (2003). Use of salt hydrate pairs to control water activity for enzyme catalysis in ionic liquids. *Biotechnology progress, 19*(3), 1029-1032.

Bernhardt. (2006). Cytochromes P450 as versatile biocatalysts. *Journal of Biotechnology, 124*(1), 128-145.

Bernhardt & Urlacher. (2014). Cytochromes P450 as promising catalysts for biotechnological application: chances and limitations. *Applied microbiology and biotechnology*, 1-19.

Bhanage, Ikushima, Shirai & Arai. (1999). The selective formation of unsaturated alcohols by hydrogenation of α, β-unsaturated aldehydes in supercritical carbon dioxide using unpromoted Pt/Al 2 O 3 catalyst. *Catalysis letters, 62*(2-4), 175-177.

Black & Martin. (1994). Evidence for conformational dynamics and molecular aggregation in cytochrome P450 102 (BM-3). *Biochemistry, 33*(40), 12056-12062.

Boddupalli, Estabrook & Peterson. (1990). Fatty acid monooxygenation by cytochrome P-450BM-3. *Journal of Biological Chemistry, 265*(8), 4233-4239.

Bondon, Macdonald, Harris & Guengerich. (1989). Oxidation of cycloalkylamines by cytochrome P-450. Mechanism-based inactivation, adduct formation, ring expansion, and nitrone formation. *Journal of Biological Chemistry, 264*(4), 1988-1997.

Bowien & Kusian. (2002). Genetics and control of CO2 assimilation in the chemoautotroph Ralstoniaeutropha. *Archives of microbiology, 178*(2), 85-93.

Bowman & Bren. (2008). The chemistry and biochemistry of heme c: functional bases for covalent attachment. *Natural product reports, 25*(6), 1118-1130.

Braga & Belo. (2015). Biocatalysis in Micellar Systems *White Biotechnology for Sustainable Chemistry* (pp. 178-196).

Brandt, Raberg, Voigt, Hecker & Steinbüchel. (2012). Elevated poly (3-hydroxybutyrate) synthesis in mutants of Ralstonia eutropha H16 defective in lipopolysaccharide biosynthesis. *Applied microbiology and biotechnology, 95*(2), 471-483.

Bredt & Snyder. (1994). Nitric oxide: a physiologic messenger molecule. *Annual review of biochemistry, 63*(1), 175-195.

Breeden, Clark, Macquarrie & Sherwood. (2012). Green Solvents. *Green Techniques for Organic Synthesis and Medicinal Chemistry*, 243.

Brigham & Sinskey. (2012). Applications of polyhydroxyalkanoates in the medical industry. *International Journal of Biotechnology for Wellness Industries, 1*(1), 52-60.

Brigham, Speth, Rha & Sinskey. (2012). Whole-genome microarray and gene deletion studies reveal regulation of the polyhydroxyalkanoate production cycle by the stringent response in Ralstonia eutropha H16. *Applied and environmental microbiology, 78*(22), 8033-8044.

Butler, Peet, McLean, Baynham, Blankley, Fisher, Rigby, Leys, Voice & Munro. (2014). Human P450-like oxidation of diverse proton pump inhibitor drugs by 'gatekeeper'mutants of flavocytochrome P450 BM3. *Biochemical Journal, 460*(2), 247-259.

CaJacob, Chan, Shephard & De Montellano. (1988). The catalytic site of rat hepatic lauric acid omega-hydroxylase. Protein versus prosthetic heme alkylation in the omega-hydroxylation of acetylenic fatty acids. *Journal of Biological Chemistry, 263*(35), 18640-18649.

Caravati, Grunwaldt & Baiker. (2006). Solvent-modified supercritical CO 2: A beneficial medium for heterogeneously catalyzed oxidation reactions. *Applied Catalysis A: General, 298*, 50-56.

Carrea, Ottolina & Riva. (1995). Role of solvents in the control of enzyme selectivity in organic media. *Trends in Biotechnology, 13*(2), 63-70.

Carrea & Riva. (2000). Properties and synthetic applications of enzymes in organic solvents. *Angewandte Chemie International Edition, 39*(13), 2226-2254.

Carugo. (2007). Statistical validation of the root-mean-square-distance, a measure of protein structural proximity. *Protein Engineering, Design & Selection, 20*(1), 33-37.

Carvalho & Cabral. (2000). Reverse micelles as reaction media for lipases. *Biochimie, 82*(11), 1063-1085.

Chen. (2009). A microbial polyhydroxyalkanoates (PHA) based bio-and materials industry. *Chemical Society Reviews, 38*(8), 2434-2446.

Chen, Chang, Wu & Chang. (2004). Characterization of phenol and trichloroethene degradation by the rhizobium Ralstonia taiwanensis. *Research in Microbiology, 155*(8), 672-680.

Cirino. (2004). *Laboratory evolution of cytochrome P450 peroxygenase activity.* California Institute of Technology.

Clement, Lomb & Möller. (1997). Isolation and characterization of the protein components of the liver microsomal O2-insensitive NADH-benzamidoxime reductase. *Journal of Biological Chemistry, 272*(31), 19615-19620.

da Rosa, Martinelli, da Silva & Loh. (2000). Easy and efficient processes for catalyst recycling and product recovery in organic biphase systems tested in the hydrogenation of hex-1-ene. *Chemical Communications*(1), 33-34.

Danielson. (2002). The cytochrome P450 superfamily: biochemistry, evolution and drug metabolism in humans. *Current drug metabolism, 3*(6), 561-597.

De Temiño, Hartmeier & Ansorge-Schumacher. (2005). Entrapment of the alcohol dehydrogenase from Lactobacillus kefir in polyvinyl alcohol for the synthesis of chiral hydrophobic alcohols in organic solvents. *Enzyme and microbial technology, 36*(1), 3-9.

Debulis & Klibanov. (1993). Dramatic enhancement of enzymatic activity in organic solvents by lyoprotectants. *Biotechnology and bioengineering, 41*(5), 566-571.

Delaforge, Servent, Wirsta, Ducrocq, Mansuy & Lenfant. (1993). Particular ability of cytochrome P-450 CYP3A to reduce glyceryl trinitrate in rat liver microsomes: subsequent formation of nitric oxide. *Chemico-biological interactions, 86*(2), 103-117.

Denard, Ren & Zhao. (2015). Improving and repurposing biocatalysts via directed evolution. *Current opinion in chemical biology, 25*, 55-64.

Diwan & Park. (2001). Pegylation enhances protein stability during encapsulation in PLGA microspheres. *Journal of Controlled Release, 73*(2), 233-244.

Dolphin, Forman, Borg, Fajer & Felton. (1971). Compounds I of catalase and horse radish peroxidase: π-cation radicals. *Proceedings of the National Academy of Sciences, 68*(3), 614-618.

Eckstein, Sesing, Kragl & Adlercreutz. (2002). At low water activity α-chymotrypsin is more active in an ionic liquid than in non-ionic organic solvents. *Biotechnology Letters, 24*(11), 867-872.

Efimov, Basran, Thackray, Handa, Mowat & Raven. (2012). Heme-Containing Dioxygenases *Advances in Inorganic Chemistry* (Vol. 64, pp. 33-51): Elsevier.

English, Hughes & Wolf. (1994). Common pathways of cytochrome P450 gene regulation by peroxisome proliferators and barbiturates in Bacillus megaterium ATCC14581. *Journal of Biological Chemistry, 269*(43), 26836-26841.

English, Palmer, Alworth, Kang, Hughes & Wolf. (1997). Fatty acid signals in Bacillus megaterium are attenuated by cytochrome P-450-mediated hydroxylation. *Biochemical Journal, 327*(Pt 2), 363.

Erbeldinger, Mesiano & Russell. (2000). Enzymatic Catalysis of Formation of Z-Aspartame in Ionic Liquid− An Alternative to Enzymatic Catalysis in Organic Solvents. *Biotechnology Progress, 16*(6), 1129-1131.

Erdogan. (2014). *Probing the Catalytic Cycle of Cytochrome P450 for Reaction Intermediates.* ETH Zurich University.

Fantuzzi, Meharenna, Briscoe, Sassone, Borgia & Gilardi. (2006). Improving catalytic properties of P450 BM3 haem domain electrodes by molecular Lego. *Chemical Communications*(12), 1289-1291.

Farombi & Surh. (2006). Heme oxygenase-1 as a potential therapeutic target for hepatoprotection. *Journal of biochemistry and molecular biology, 39*(5), 479.

Fisher, Thompson, Ribeiro, Lechner & Rettie. (1998). P450-catalyzed in-chain desaturation of valproic acid: isoform selectivity and mechanism of formation of Δ3-valproic acid generated by baculovirus-expressed CYP3A1. *Archives of biochemistry and biophysics, 356*(1), 63-70.

Fu, Liu, Davidson & Liu. (2009). Heme iron nitrosyl complex of MauG reveals an efficient redox equilibrium between hemes with only one heme exclusively binding exogenous ligands. *Biochemistry, 48*(49), 11603-11605.

Fujita, MacFarlane & Forsyth. (2005). Protein solubilising and stabilising ionic liquids. *Chemical communications*(38), 4804-4806.

Fukushima, Mizuki, Echigo, Inoue & Usami. (2005). Organic solvent tolerance of halophilic α-amylase from a Haloarchaeon, Haloarcula sp. strain S-1. *Extremophiles, 9*(1), 85-89.

Fulco, Kim, Matson, Narhi & Ruettinger. (1983). Nonsubstrate induction of a soluble bacterial cytochrome P-450 monooxygenase by phenobarbital and its analogs *Enzyme Induction and Modulation* (pp. 155-161): Springer.

Gao, Chen, Wu & Chen. (2011). Polyhydroxyalkanoates as a source of chemicals, polymers, and biofuels. *Current opinion in biotechnology, 22*(6), 768-774.

Garfinkel. (1958). Studies on pig liver microsomes. I. Enzymic and pigment composition of different microsomal fractions. *Archives of biochemistry and biophysics, 77*(2), 493-509.

Ge, Wang, Tan, Tsai, Yong & Hua. (2010). A novel method of protein extraction from yeast using ionic liquid solution. *Talanta, 81*(4), 1861-1864.

Geng, Dornevil, Davidson & Liu. (2013). Tryptophan-mediated charge-resonance stabilization in the bis-Fe (IV) redox state of MauG. *Proceedings of the National Academy of Sciences, 110*(24), 9639-9644.

Girvan & Munro. (2016). Applications of microbial cytochrome P450 enzymes in biotechnology and synthetic biology. *Current opinion in chemical biology, 31*, 136-145.

Goeptar, Scheerens & Vermeulen. (1995). Oxygen and xenobiotic reductase activities of cytochrome P450. *Critical reviews in toxicology, 25*(1), 25-65.

Gonzalez & Nebert. (1990). Evolution of the P450 gene superfamily:: animal-plant 'warfare', molecular drive and human genetic differences in drug oxidation. *Trends in Genetics, 6*, 182-186.

Govindaraj & Poulos. (1995). Role of the linker region connecting the reductase and heme domains in cytochrome P450BM-3. *Biochemistry, 34*(35), 11221-11226.

Graham-Lorence, Peterson, Amarneh, Simpson & White. (1995). A three-dimensional model of aromatase cytochrome P450. *Protein Science, 4*(6), 1065-1080.

Griebenow & Klibanov. (1996). On protein denaturation in aqueous-organic mixtures but not in pure organic solvents. *Journal of the American Chemical Society, 118*(47), 11695-11700.

Griffith & Stuehr. (1995). Nitric oxide synthases: properties and catalytic mechanism. *Annual review of physiology, 57*(1), 707-734.

Guengerich. (1987). Oxidative cleavage of carboxylic esters by cytochrome P-450. *Journal of Biological Chemistry, 262*(18), 8459-8462.

Guengerich. (2001). Common and uncommon cytochrome P450 reactions related to metabolism and chemical toxicity. *Chemical research in toxicology, 14*(6), 611-650.

Guengerich, Peterson & Böcker. (1988). Cytochrome P-450-catalyzed hydroxylation and carboxylic acid ester cleavage of Hantzsch pyridine esters. *Journal of Biological Chemistry, 263*(17), 8176-8183.

Haines, Tomchick, Machius & Peterson. (2001). Pivotal role of water in the mechanism of P450BM-3. *Biochemistry, 40*(45), 13456-13465.

Halling & Kvittingen. (1999). Why did biocatalysis in organic media not take off in the 1930s? *Trends in biotechnology, 17*(9), 343-344.

HAMILTON. (1974). Chemical models and mechanisms for oxygenases *Molecular mechanisms of oxygen activation* (pp. 405-451): Elsevier.

Hanefeld, Gardossi & Magner. (2009). Understanding enzyme immobilisation. *Chemical Society Reviews, 38*(2), 453-468.

Hanioka, Gonzalez, Lindberg, Liu, Gelboin & Korzekwa. (1992). Site-directed mutagenesis of cytochrome P450s CYP2A1 and CYP2A2: influence of the distal helix on the kinetics of testosterone hydroxylation. *Biochemistry, 31*(13), 3364-3370.

Hasemann, Kurumbail, Boddupalli, Peterson & Deisenhofer. (1995). Structure and function of cytochromes P450: a comparative analysis of three crystal structures. *Structure, 3*(1), 41-62.

Heldebrant & Jessop. (2003). Liquid poly (ethylene glycol) and supercritical carbon dioxide: a benign biphasic solvent system for use and recycling of homogeneous catalysts. *Journal of the American Chemical Society, 125*(19), 5600-5601.

Henderson, Jiménez-González, Constable, Alston, Inglis, Fisher, Sherwood, Binks & Curzons. (2011). Expanding GSK's solvent selection guide–embedding sustainability into solvent selection starting at medicinal chemistry. *Green Chemistry, 13*(4), 854-862.

Hiner, Raven, Thorneley, García-Cánovas & Rodríguez-López. (2002). Mechanisms of compound I formation in heme peroxidases. *Journal of inorganic biochemistry, 91*(1), 27-34.

Holden, Titball, Peacock, Cerdeño-Tárraga, Atkins, Crossman, Pitt, Churcher, Mungall & Bentley. (2004). Genomic plasticity of the causative agent of melioidosis, Burkholderia pseudomallei. *Proceedings of the National Academy of Sciences of the United States of America, 101*(39), 14240-14245.

Holder, Ulrich, DeBono, Godfrey, Desjardins, Zucker, Zeng, Leach, Ghiviriga & Dancel. (2011). Comparative and functional genomics of Rhodococcus opacus PD630 for biofuels development. *PLoS genetics, 7*(9), e1002219.

Holton, Brugliera, Lester, Tanaka, Hyland, Menting, Lu, Farcy, Stevenson & Cornish. (1993). Cloning and expression of cytochrome P450 genes controlling flower colour. *Nature, 366*(6452), 276.

Huang, Hah & Silverman. (2001). Mechanism of Nitric Oxide Synthase. Evidence that Direct Hydrogen Atom Abstraction from the O−H Bond of NG-Hydroxyarginine Is Not Relevant to the Mechanism. *Journal of the American Chemical Society, 123*(11), 2674-2676.

Huang, Westlake, Maréchal, Joyce, Moody & Roberts. (2007). Filling a hole in cytochrome P450 BM3 improves substrate binding and catalytic efficiency. *Journal of molecular biology, 373*(3), 633-651.

Hubbard, Shen, Paschke, Kasper & Kim. (2001). NADPH-cytochrome P450 oxidoreductase structural basis for hydride and electron transfer. *Journal of Biological Chemistry, 276*(31), 29163-29170.

Huddleston, Visser, Reichert, Willauer, Broker & Rogers. (2001). Characterization and comparison of hydrophilic and hydrophobic room temperature ionic liquids incorporating the imidazolium cation. *Green chemistry, 3*(4), 156-164.

Humphrey, Dalke & Schulten. (1996). VMD: visual molecular dynamics. *Journal of molecular graphics, 14*(1), 33-38.

Hurshman, Krebs, Edmondson, Huynh & Marletta. (1999). Formation of a pterin radical in the reaction of the heme domain of inducible nitric oxide synthase with oxygen. *Biochemistry, 38*(48), 15689-15696.

Inada, Takahashi, Yoshimoto, Ajima, Matsushima & Saito. (1986). Application of polyethylene glycol-modified enzymes in biotechnological processes: organic solvent-soluble enzymes. *Trends in Biotechnology, 4*(7), 190-194.

Janson & Rydén. (1989). *Protein purification: principles, high resolution methods, and applications*. New York: VCH.

Johannes & Zhao. (2006). Directed evolution of enzymes and biosynthetic pathways. *Current opinion in microbiology, 9*(3), 261-267.

Johnson & Stanier. (1971). Dissimilation of aromatic compounds by Alcaligenes eutrophus. *Journal of bacteriology, 107*(2), 468-475.

Judge, Takahashi, Longenecker, Fry, Abad-Zapatero & Chiu. (2009). The effect of ionic liquids on protein crystallization and X-ray diffraction resolution. *Crystal Growth and Design, 9*(8), 3463-3469.

Julsing, Cornelissen, Bühler & Schmid. (2008). Heme-iron oxygenases: powerful industrial biocatalysts? *Current opinion in chemical biology, 12*(2), 177-186.

Kaneda. (1991). Iso-and anteiso-fatty acids in bacteria: biosynthesis, function, and taxonomic significance. *Microbiological reviews, 55*(2), 288-302.

Karan, Capes & DasSarma. (2012). Function and biotechnology of extremophilic enzymes in low water activity. *Aquatic Biosystems, 8*(1), 4.

Katsumoto, Fukuchi-Mizutani, Fukui, Brugliera, Holton, Karan, Nakamura, Yonekura-Sakakibara, Togami & Pigeaire. (2007). Engineering of the rose flavonoid biosynthetic pathway successfully generated blue-hued flowers accumulating delphinidin. *Plant and Cell Physiology, 48*(11), 1589-1600.

Kikuchi, Yoshida & Noguchi. (2005). Heme oxygenase and heme degradation. *Biochemical and biophysical research communications, 338*(1), 558-567.

Klibanov. (1997). Why are enzymes less active in organic solvents than in water? *Trends in Biotechnology, 15*(3), 97-101.

Klibanov. (2001). Improving enzymes by using them in organic solvents. *Nature, 409*(6817), 241-246.

Klingenberg. (1958). Pigments of rat liver microsomes. *Archives of biochemistry and biophysics, 75*(2), 376-386.

Knez, Kavčič, Gubicza, Bélafi-Bakó, Németh, Primožič & Habulin. (2012). Lipase-catalyzed esterification of lactic acid in supercritical carbon dioxide. *The Journal of Supercritical Fluids, 66*, 192-197.

Knez, Markočič, Leitgeb, Primožič, Knez Hrnčič & Škerget. (2014). Industrial applications of supercritical fluids: A review. *Energy, 77*(0), 235-243.

Koudelakova, Chaloupkova, Brezovsky, Prokop, Sebestova, Hesseler, Khabiri, Plevaka, Kulik & Kuta Smatanova. (2013). Engineering enzyme stability and resistance to an organic cosolvent by modification of residues in the access tunnel. *Angewandte Chemie International Edition, 52*(7), 1959-1963.

Koutinas, Xu, Wang & Webb. (2007). Polyhydroxybutyrate production from a novel feedstock derived from a wheat-based biorefinery. *Enzyme and Microbial Technology, 40*(5), 1035-1044.

Krishna, Srinivas, Raghavarao & Karanth. (2002). Reverse micellar extraction for downstream processing of proteins/enzymes *History and trends in bioprocessing and biotransformation* (pp. 119-183): Springer.

Kubo, Peters, Meinhold & Arnold. (2006). Enantioselective epoxidation of terminal alkenes to (R)- and (S)-epoxides by engineered cytochromes P450 BM-3. *Chemistry-a European Journal, 12*(4), 1216-1220.

Kuper, Wong, Roccatano, Wilmanns & Schwaneberg. (2007). Understanding a mechanism of organic cosolvent inactivation in heme monooxygenase P450 BM-3. *Journal of the American Chemical Society, 129*(18), 5786-5787.

Kusano, Kagawa, Sakaguchi, Omura & Waterman. (2001). Importance of a proline-rich sequence in the amino-terminal region for correct folding of mitochondrial and soluble microbial P450s. *The Journal of Biochemistry, 129*(2), 271-277.

Laane, Boeren, Vos & Veeger. (1987). Rules for optimization of biocatalysis in organic solvents. *Biotechnology and Bioengineering, 30*(1), 81-87.

Lambeth & Palmer. (1973). The kinetics and mechanism of reduction of electron transfer proteins and other compounds of biological interest by dithionite. *Journal of Biological Chemistry, 248*(17), 6095-6103.

Lau, Rantwijk, Seddon & Sheldon. (2000). Lipase-catalyzed reactions in ionic liquids. *Organic Letters, 2*(26), 4189-4191.

Lau, Sorgedrager, Carrea, van Rantwijk, Secundo & Sheldon. (2004). Dissolution of Candida antarctica lipase B in ionic liquids: effects on structure and activity. *Green Chemistry, 6*(9), 483-487.

Lee & Lee. (2003). Metabolic engineering of Escherichia coli for production of enantiomerically pure (R)-(−)-hydroxycarboxylic acids. *Applied and environmental microbiology, 69*(6), 3421-3426.

Lee, Ryoo, Smith, Arellano, Mitchell, Lagow, Webber & Johnston. (2003). Carbon dioxide-in-water microemulsions. *Journal of the American Chemical Society, 125*(10), 3181-3189.

Lentz, Feenstra, Habicher, Hauer, Schmid & Urlacher. (2006). Altering the regioselectivity of cytochrome P450 CYP102A3 of Bacillus subtilis by using a new versatile assay system. *ChemBioChem, 7*(2), 345-350.

Lentz, Li, Schwaneberg, Lutz-Wahl, Fischer & Schmid. (2001). Modification of the fatty acid specificity of cytochrome P450 BM-3 from Bacillus megaterium by directed evolution: a validated assay. *Journal of molecular catalysis B: Enzymatic, 15*(4-6), 123-133.

Lentz, Urlacher & Schmid. (2004). Substrate specificity of native and mutated cytochrome P450 (CYP102A3) from Bacillus subtilis. *Journal of biotechnology, 108*(1), 41-49.

Li & Poulos. (1995). Modeling protein–substrate interactions in the heme domain of cytochrome P450BM− 3. *Acta Crystallographica Section D: Biological Crystallography, 51*(1), 21-32.

Li, Schwaneberg, Fischer & Schmid. (2000). Directed Evolution of the Fatty-Acid Hydroxylase P450 BM-3 into an Indole-Hydroxylating Catalyst. *Chemistry-A European Journal, 6*(9), 1531-1536.

Li, Schwaneberg, Fischer, Schmitt, Pleiss, Lutz-Wahl & Schmid. (2001). Rational evolution of a medium chain-specific cytochrome P-450 BM-3 variant. *Biochimica Et Biophysica Acta (BBA)-Protein Structure and Molecular Enzymology, 1545*(1-2), 114-121.

Liang, Chen & Fulco. (1998). In vivo roles of Bm3R1 repressor in the barbiturate-mediated induction of the cytochrome P450 genes (P450BM-3 and P450BM-1) of Bacillus megaterium. *Biochimica et Biophysica Acta (BBA)-General Subjects, 1380*(2), 183-197.

Loomis. (1988). Four billion years: an essay on the evolution of genes and organisms (pp. 286). Sunderland, Massachusetts: Sinauer Associates Inc.

Loughran, Roman, Aitken, Miller & Masters. (2000). Identification of unique amino acids that modulate CYP4A7 activity. *Biochemistry, 39*(49), 15110-15120.

Lozano, de Diego, Guegan, Vaultier & Iborra. (2001). Stabilization of α-chymotrypsin by ionic liquids in transesterification reactions. *Biotechnology and bioengineering, 75*(5), 563-569.

Lozano, De Diego, Sauer, Vaultier, Gmouh & Iborra. (2007). On the importance of the supporting material for activity of immobilized Candida antarctica lipase B in ionic liquid/hexane and ionic liquid/supercritical carbon dioxide biphasic media. *The Journal of supercritical fluids, 40*(1), 93-100.

Lu, Brigham, Li & Sinskey. (2016). Ralstonia eutropha H16 as a Platform for the Production of Biofuels, Biodegradable Plastics, and Fine Chemicals from Diverse Carbon Resources. *Biotechnology for Biofuel Production and Optimization* 325-351.

Luciakova. (2015). Characterisation of Novel Cytochrome P450-fusion enzymes.

Lüthy, Bowie & Eisenberg. (1992). Assessment of protein models with three-dimensional profiles. *Nature, 356*(6364), 83.

Maines. (1988). Heme oxygenase: function, multiplicity, regulatory mechanisms, and clinical applications. *The Faseb Journal, 2*(10), 2557-2568.

Maines. (2005). The heme oxygenase system: update 2005. *Antioxidants & redox signaling, 7*(11-12), 1761-1766.

Mason. (1957). Mechanisms of oxygen metabolism. *Science, 125*(3259), 1185-1188.

Matson, Hare & Fulco. (1977). Characteristics of a cytochrome P-450-dependent fatty acid ω-2 hydroxylase from Bacillus megaterium. *Biochimica et Biophysica Acta (BBA)-Lipids and Lipid Metabolism, 487*(3), 487-494.

Matsui, Iwasaki, Sugiyama, Unno & Ikeda-Saito. (2010). Dioxygen activation for the self-degradation of heme: reaction mechanism and regulation of heme oxygenase. *Inorganic chemistry, 49*(8), 3602-3609.

Matthews. (1993). Structural and genetic analysis of protein stability. *Annual review of biochemistry, 62*(1), 139-160.

Matthews, Nicholson & Becktel. (1987). Enhanced protein thermostability from site-directed mutations that decrease the entropy of unfolding. *Proceedings of the National Academy of Sciences, 84*(19), 6663-6667.

Maurer, Schulze, Schmid & Urlacher. (2003). Immobilisation of P450 BM-3 and an NADP+ Cofactor Recycling System: Towards a Technical Application of Heme-Containing Monooxygenases in Fine Chemical Synthesis. *Advanced Synthesis & Catalysis, 345*(6-7), 802-810.

McIntire, Wemmer, Chistoserdov & Lidstrom. (1991). A new cofactor in a prokaryotic enzyme: tryptophan tryptophylquinone as the redox prosthetic group in methylamine dehydrogenase. *Science, 252*(5007), 817-824.

McMillan, Bredt, Hirsch, Snyder, Clark & Masters. (1992). Cloned, expressed rat cerebellar nitric oxide synthase contains stoichiometric amounts of heme, which binds carbon monoxide. *Proceedings of the National Academy of Sciences, 89*(23), 11141-11145.

Mehendale, Roth, Gandolfi, Klaunig, Lemasters & Curtis. (1994). Novel mechanisms in chemically induced hepatotoxicity. *The FASEB journal, 8*(15), 1285-1295.

Meng, Chen & Wang. (2011). One-pot synthesis of N, N-bis [2-methylbutyl] imidazolium hexafluorophosphate–TiO2 nanocomposites and application for protein isolation. *Journal of Materials Chemistry, 21*(38), 14857-14863.

Mico, Branchflower & Pohl. (1983). Formation of electrophilic chlorine from carbon tetrachloride—involvement of cytochrome P-450. *Biochemical pharmacology, 32*(15), 2357-2359.

Minerdi, Sadeghi, Di Nardo, Rua, Castrignanò, Allegra & Gilardi. (2015). CYP116B5: a new class VII catalytically self-sufficient cytochrome P450 from Acinetobacter radioresistens that enables growth on alkanes. *Molecular microbiology, 95*(3), 539-554.

Miroux & Walker. (1996). Over-production of proteins inEscherichia coli: mutant hosts that allow synthesis of some membrane proteins and globular proteins at high levels. *Journal of molecular biology, 260*(3), 289-298.

Montellano. (1998). Heme oxygenase mechanism: evidence for an electrophilic, ferric peroxide species. *Accounts of chemical research, 31*(9), 543-549.

Moore & Arnold. (1996). Directed evolution of a para-nitrobenzyl esterase for aqueous-organic solvents. *Nature biotechnology, 14*(4), 458-467.

Morris & Hager. (1966). Chloroperoxidase I. Isolation and properties of the crystalline glycoprotein. *Journal of Biological Chemistry, 241*(8), 1763-1768.

Müller, MacEachran, Burd, Sathitsuksanoh, Bi, Yeh, Lee, Hillson, Chhabra & Singer. (2013). Engineering of Ralstonia eutropha H16 for autotrophic and heterotrophic production of methyl ketones. *Applied and environmental microbiology, 79*(14), 4433-4439.

Munro, Daff, Coggins, Lindsay & Chapman. (1996). Probing Electron Transfer in Flavocytochrome P-450 BM3 and Its Component Domains. *The Febs Journal, 239*(2), 403-409.

Munro, Lindsay, Coggins, Kelly & Price. (1994). Structural and enzymological analysis of the interaction of isolated domains of cytochrome P-450 BM3. *Febs letters, 343*(1), 70-74.
Narhi & Fulco. (1982). Phenobarbital induction of a soluble cytochrome P-450-dependent fatty acid monooxygenase in Bacillus megaterium. *Journal of Biological Chemistry, 257*(5), 2147-2150.
Narhi & Fulco. (1986). Characterization of a catalytically self-sufficient 119,000-dalton cytochrome P-450 monooxygenase induced by barbiturates in Bacillus megaterium. *Journal of Biological Chemistry, 261*(16), 7160-7169.
Nebert & Gonzalez. (1985). Cytochrome P450 gene expression and regulation. *Trends in Pharmacological Sciences, 6*, 160-164.
Nebert, Jones, Owens & Puga. (1988). Oxidases and Related Redox Systems. *by TE King, HS Mason & M. Morrison, Alan R. Liss, New York*, 557-576.
Neeli, Girvan, Lawrence, Warren, Leys, Scrutton & Munro. (2005). The dimeric form of flavocytochrome P450 BM3 is catalytically functional as a fatty acid hydroxylase. *FEBS letters, 579*(25), 5582-5588.
Nelson. (1999). Cytochrome P450 and the individuality of species. *Archives of biochemistry and biophysics, 369*(1), 1-10.
Nelson. (2009). The cytochrome p450 homepage. *Human genomics, 4*(1), 59.
Nelson, Kamataki, Waxman, Guengerich, Estabrook, Feyereisen, Gonzalez, Coon, Gunsalus & Gotoh. (1993). The P450 superfamily: update on new sequences, gene mapping, accession numbers, early trivial names of enzymes, and nomenclature. *DNA and cell biology, 12*(1), 1-51.
Nelson, Koymans, Kamataki, Stegeman, Feyereisen, Waxman, Waterman, Gotoh, Coon & Estabrook. (1996). P450 superfamily: update on new sequences, gene mapping, accession numbers and nomenclature. *Pharmacogenetics, 6*(1), 1-42.
Nelson & Strobel. (1987). Evolution of cytochrome P-450 proteins. *Molecular biology and evolution, 4*(6), 572-593.
Nierman, DeShazer, Kim, Tettelin, Nelson, Feldblyum, Ulrich, Ronning, Brinkac & Daugherty. (2004). Structural flexibility in the Burkholderia mallei genome. *Proceedings of the National Academy of Sciences of the United States of America, 101*(39), 14246-14251.
Nischan & Hackenberger. (2014). Site-specific PEGylation of Proteins: Recent Developments. *The Journal of Organic Chemistry, 79*(22), 10727-10733.
Noble, Miles, Chapman, Lysek, MacKay, Hanzlik & Munro. (1999). Roles of key active-site residues in flavocytochrome P450 BM3. *Biochemical Journal, 339*(2), 371-379.
Noritomi, Minamisawa, Kamiya & Kato. (2011). Thermal stability of proteins in the presence of aprotic ionic liquids. *Journal of Biomedical Science and Engineering, 4*(02), 94.
O'Reilly, Köhler, Flitsch & Turner. (2011). Cytochromes P450 as useful biocatalysts: addressing the limitations. *Chemical Communications, 47*(9), 2490-2501.
Ogino, Watanabe, Yamada, Nakagawa, Hirose, Noguchi, Yasuda & Ishikawa. (1999). Purification and characterization of organic solvent-stable protease from organic solvent-tolerant Pseudomonas aeruginosa PST-01. *Journal of bioscience and bioengineering, 87*(1), 61-68.
Omura & Sato. (1962). A new cytochrome in liver microsomes. *J. biol. Chem, 237*(4), 1375-1376.
Omura & Sato. (1964). The carbon monoxide-binding pigment of liver microsomes. I. Evidence for its hemoprotein nature. *J biol Chem, 239*(7), 2370-2378.
Ost, Miles, Murdoch, Cheung, Reid, Chapman & Munro. (2000). Rational re-design of the substrate binding site of flavocytochrome P450 BM3. *Febs Letters, 486*(2), 173-177.
Oster, Boddupalli & Peterson. (1991). Expression, purification, and properties of the flavoprotein domain of cytochrome P-450BM-3. Evidence for the importance of the amino-terminal region for FMN binding. *Journal of Biological Chemistry, 266*(33), 22718-22725.

Otyepka, Skopalík, Anzenbacherová & Anzenbacher. (2007). What common structural features and variations of mammalian P450s are known to date? *Biochimica et Biophysica Acta (BBA)-General Subjects, 1770*(3), 376-389.

Paddon, Westfall, Pitera, Benjamin, Fisher, McPhee, Leavell, Tai, Main & Eng. (2013). High-level semi-synthetic production of the potent antimalarial artemisinin. *Nature, 496*(7446), 528.

Patel, Kumari & Khan. (2014). Recent advances in the applications of ionic liquids in protein stability and activity: a review. *Applied biochemistry and biotechnology, 172*(8), 3701-3720.

Pavlidis, Gournis, Papadopoulos & Stamatis. (2009). Lipases in water-in-ionic liquid microemulsions: structural and activity studies. *Journal of Molecular Catalysis B: Enzymatic, 60*(1), 50-56.

Pazhang, Khajeh, Ranjbar & Hosseinkhani. (2006). Effects of water-miscible solvents and polyhydroxy compounds on the structure and enzymatic activity of thermolysin. *Journal of biotechnology, 127*(1), 45-53.

Persson & Bornscheuer. (2003). Increased stability of an esterase from Bacillus stearothermophilus in ionic liquids as compared to organic solvents. *Journal of Molecular Catalysis B: Enzymatic, 22*(1), 21-27.

Peters, Meinhold, Glieder & Arnold. (2003). Regio-and enantioselective alkane hydroxylation with engineered cytochromes P450 BM-3. *Journal of the American Chemical Society, 125*(44), 13442-13450.

Pietrzykowski & Treistman. (2008). The molecular basis of tolerance. *Alcohol Research & Health, 31*(4), 298.

Piontek, Ullrich, Liers, Diederichs, Plattner & Hofrichter. (2010). Crystallization of a 45 kDa peroxygenase/peroxidase from the mushroom Agrocybe aegerita and structure determination by SAD utilizing only the haem iron. *Acta Crystallographica Section F: Structural Biology and Crystallization Communications, 66*(6), 693-698.

Platten, Wick & Van den Eynde. (2012). Tryptophan catabolism in cancer: beyond IDO and tryptophan depletion. *Cancer research, 72*(21), 5435-5440.

Podust, Poulos & Waterman. (2001). Crystal structure of cytochrome P450 14α-sterol demethylase (CYP51) from Mycobacterium tuberculosis in complex with azole inhibitors. *Proceedings of the National Academy of Sciences, 98*(6), 3068-3073.

Pohlmann, Fricke, Reinecke, Kusian, Liesegang, Cramm, Eitinger, Ewering, Pötter & Schwartz. (2006). Genome sequence of the bioplastic-producing "Knallgas" bacterium Ralstonia eutropha H16. *Nature biotechnology, 24*(10), 1257.

Porter. (1991). An unusual yet strongly conserved flavoprotein reductase in bacteria and mammals. *Trends in biochemical sciences, 16*, 154-158.

Poulos. (2014). Heme enzyme structure and function. *Chemical reviews, 114*(7), 3919-3962.

Poulos, Finzel & Howard. (1987). High-resolution crystal structure of cytochrome P450cam. *Journal of molecular biology, 195*(3), 687-700.

Raberg, Kaddor, Kusian, Stahlhut, Budinova, Kolev, Bowien & Steinbüchel. (2012). Impact of each individual component of the mutated PTS Nag on glucose uptake and phosphorylation in Ralstonia eutropha G+ 1. *Applied microbiology and biotechnology, 95*(3), 735-744.

Ramos, Abreu, Nascimento & Ho. (2004). A high-copy T7 Escherichia coli expression vector for the production of recombinant proteins with a minimal N-terminal His-tagged fusion peptide. *Brazilian Journal of Medical and Biological Research, 37*(8), 1103-1109.

Rath, Glibowicka, Nadeau, Chen & Deber. (2009). Detergent binding explains anomalous SDS-PAGE migration of membrane proteins. *Proceedings of the National Academy of Sciences, 106*(6), 1760-1765.

Raucy, Kraner & Lasker. (1993). Bioactivation of halogenated hydrocarbons by cytochrome P4502E1. *Critical reviews in toxicology, 23*(1), 1-20.

Reetz, Soni, Fernández, Gumulya & Carballeira. (2010). Increasing the stability of an enzyme toward hostile organic solvents by directed evolution based on iterative saturation mutagenesis using the B-FIT method. *Chemical Communications, 46*(45), 8657-8658.

Reich. (2014). *Studies on variable surface loop regions of the ene reductase NCR from Zymomonas mobilis.* Universität Stuttgart.

Reich, Kress, Nestl & Hauer. (2014). Variations in the stability of NCR ene reductase by rational enzyme loop modulation. *Journal of Structural Biology, 185*(2), 228-233.

Ren, Yorke, Taylor, Zhang, Zhou & Wong. (2015). Drug oxidation by cytochrome P450BM3: metabolite synthesis and discovering new P450 Reaction types. *Chemistry-A European Journal, 21*(42), 15039-15047.

Rettie, Boberg, Rettenmeier & Baillie. (1988). Cytochrome P-450-catalyzed desaturation of valproic acid in vitro. Species differences, induction effects, and mechanistic studies. *Journal of Biological Chemistry, 263*(27), 13733-13738.

Ro, Paradise, Ouellet, Fisher, Newman, Ndungu, Ho, Eachus, Ham & Kirby. (2006). Production of the antimalarial drug precursor artemisinic acid in engineered yeast. *Nature, 440*(7086), 940.

Roberts, Celik, Hunter, Ost, White, Chapman, Turner & Flitsch. (2003). A self-sufficient cytochrome P450 with a primary structural organization that includes a flavin domain and a [2Fe-2S] redox center. *Journal of Biological Chemistry, 278*(49), 48914-48920.

Roccatano. (2015). Structure, dynamics, and function of the monooxygenase P450 BM-3: insights from computer simulations studies. *Journal of Physics: Condensed Matter, 27*(27), 273102.

Roccatano, Wong, Schwaneberg & Zacharias. (2005). Structural and dynamic properties of cytochrome P450 BM-3 in pure water and in a dimethylsulfoxide/water mixture. *Biopolymers, 78*(5), 259-267.

Roccatano, Wong, Schwaneberg & Zacharias. (2006). Toward understanding the inactivation mechanism of monooxygenase P450 BM-3 by organic cosolvents: A molecular dynamics simulation study. *Biopolymers, 83*(5), 467-476.

Roiban, Agudo, Ilie, Lonsdale & Reetz. (2014). CH-activating oxidative hydroxylation of 1-tetralones and related compounds with high regio-and stereoselectivity. *Chemical Communications, 50*(92), 14310-14313.

Ruettinger & Fulco. (1981). Epoxidation of unsaturated fatty acids by a soluble cytochrome P-450-dependent system from Bacillus megaterium. *Journal of Biological Chemistry, 256*(11), 5728-5734.

Sadeghi, Meharenna, Fantuzzi, Valetti & Gilardi. (2000). Engineering artificial redox chains by molecular 'Lego'. *Faraday discussions, 116*, 135-153.

Sakaki, Sugimoto, Hayashi, Yasuda, Munetsuna, Kamakura, Ikushiro & Shiro. (2011). Bioconversion of vitamin D to its active form by bacterial or mammalian cytochrome P450. *Biochimica et Biophysica Acta (BBA)-Proteins and Proteomics, 1814*(1), 249-256.

Salanoubat, Genin, Artiguenave, Gouzy, Mangenot, Arlat, Billault, Brottier, Camus & Cattolico. (2002). Genome sequence of the plant pathogen Ralstonia solanacearum. *Nature, 415*(6871), 497.

Sasaki, Miyazaki, Saito, Adachi, Mizoue, Hanada & Omura. (1992). Transformation of vitamin D 3 to 1α, 25-dihydroxyvitamin D 3 via 25-hydroxyvitamin D 3 using Amycolata sp. strains. *Applied microbiology and biotechnology, 38*(2), 152-157.

Schlezinger, White & Stegeman. (1999). Oxidative inactivation of cytochrome P-450 1A (CYP1A) stimulated by 3, 3', 4, 4'-tetrachlorobiphenyl: production of reactive oxygen by vertebrate CYP1As. *Molecular pharmacology, 56*(3), 588-597.

Schmitke, Wescott & Klibanov. (1996). The mechanistic dissection of the plunge in enzymatic activity upon transition from water to anhydrous solvents. *Journal of the American Chemical Society, 118*(14), 3360-3365.

Schneider, Marles-Wright, Sharp & Paoli. (2007). Diversity and conservation of interactions for binding heme in b-type heme proteins. *Natural product reports, 24*(3), 621-630.

Schwede, Kopp, Guex & Peitsch. (2003). SWISS-MODEL: an automated protein homology-modeling server. *Nucleic acids research, 31*(13), 3381-3385.

Seifert, Antonovici, Hauer & Pleiss. (2011). An Efficient Route to Selective Bio-oxidation Catalysts: an Iterative Approach Comprising Modeling, Diversification, and Screening, Based on CYP102A1. *ChemBioChem, 12*(9), 1346-1351.

Seongsoon & Romas. (2003a). Biocatalysis in ionic liquids–advantages beyond green technology. *Current Opinion in Biotechnology, 14*(4), 432-437.

Seongsoon & Romas. (2003b). Biocatalysis in ionic liquids – advantages beyond green technology. *Current Opinion in Biotechnology, 14*(4), 432-437.

Seto & Guengerich. (1993). Partitioning between N-dealkylation and N-oxygenation in the oxidation of N, N-dialkylarylamines catalyzed by cytochrome P450 2B1. *Journal of Biological Chemistry, 268*(14), 9986-9997.

Sevrioukova, Li, Zhang, Peterson & Poulos. (1999). Structure of a cytochrome P450–redox partner electron-transfer complex. *Proceedings of the National Academy of Sciences, 96*(5), 1863-1868.

Sevrioukova & Peterson. (1995). NADPH-P-450 reductase: structural and functional comparisons of the eukaryotic and prokaryotic isoforms. *Biochimie, 77*(7-8), 562-572.

Sevrioukova, Shaffer, Ballou & Peterson. (1996). Equilibrium and transient state spectrophotometric studies of the mechanism of reduction of the flavoprotein domain of P450BM-3. *Biochemistry, 35*(22), 7058-7068.

Sheldon, Arends & Hanefeld. (2007). *Green chemistry and catalysis*: John Wiley & Sons.

Sirim, Widmann, Wagner & Pleiss. (2010). Prediction and analysis of the modular structure of cytochrome P450 monooxygenases. *BMC structural biology, 10*(1), 34.

Sligar & Murray. (1986). Cytochrome P-450: Structure, Mechanism, and Biochemistry (Ortiz de Montellano, PR, ed) pp. 429–503: Plenum Publishing Corp., New York.

Smith, Kahraman & Thornton. (2010). Heme proteins—diversity in structural characteristics, function, and folding. *Proteins: structure, function, and bioinformatics, 78*(10), 2349-2368.

Sørensen & Mortensen. (2005). Soluble expression of recombinant proteins in the cytoplasm of Escherichia coli. *Microbial cell factories, 4*(1), 1.

Stepankova, Bidmanova, Koudelakova, Prokop, Chaloupkova & Damborsky. (2013). Strategies for stabilization of enzymes in organic solvents. *Acs Catalysis, 3*(12), 2823-2836.

Stuehr & Ikeda-Saito. (1992). Spectral characterization of brain and macrophage nitric oxide synthases. Cytochrome P-450-like hemeproteins that contain a flavin semiquinone radical. *Journal of Biological Chemistry, 267*(29), 20547-20550.

Summers & Flowers. (2000). Protein renaturation by the liquid organic salt ethylammonium nitrate. *Protein Science, 9*(10), 2001-2008.

Szczebara, Chandelier, Villeret, Masurel, Bourot, Duport, Blanchard, Groisillier, Testet & Costaglioli. (2003). Total biosynthesis of hydrocortisone from a simple carbon source in yeast. *Nature biotechnology, 21*(2), 143.

Tan, Hirakawa, Suzuki, Haga, Iwata & Nagamune. (2016). Immobilization of a bacterial cytochrome P450 monooxygenase system on a solid support. *Angewandte Chemie, 128*(48), 15226-15230.

Tarboush, Jensen, Yukl, Geng, Liu, Wilmot & Davidson. (2011). Mutagenesis of tryptophan199 suggests that hopping is required for MauG-dependent tryptophan tryptophylquinone biosynthesis. *Proceedings of the National Academy of Sciences, 108*(41), 16956-16961.

Thumar & Singh. (2009). Organic solvent tolerance of an alkaline protease from salt-tolerant alkaliphilic Streptomyces clavuligerus strain Mit-1. *Journal of industrial microbiology & biotechnology, 36*(2), 211.

Ting, Bonkovsky & Guo. (2011). Structural analysis of heme proteins: implications for design and prediction. *BMC structural biology, 11*(1), 13.

Torella, Ford, Kim, Chen, Way & Silver. (2013). Tailored fatty acid synthesis via dynamic control of fatty acid elongation. *Proceedings of the National Academy of Sciences, 110*(28), 11290-11295.

Tuck, Peterson & de Montellano. (1992). Active site topologies of bacterial cytochromes P450101 (P450cam), P450108 (P450terp), and P450102 (P450BM-3). In situ rearrangement of their phenyl-iron complexes. *Journal of Biological Chemistry, 267*(8), 5614-5620.

Unno, Matsui & Ikeda-Saito. (2007). Structure and catalytic mechanism of heme oxygenase. *Natural product reports, 24*(3), 553-570.

Urlacher & Eiben. (2006). Cytochrome P450 monooxygenases: perspectives for synthetic application. *Trends in biotechnology, 24*(7), 324-330.

Urlacher, Lutz-Wahl & Schmid. (2004). Microbial P450 enzymes in biotechnology. *Applied microbiology and biotechnology, 64*(3), 317-325.

Urlacher, Makhsumkhanov & Schmid. (2006). Biotransformation of β-ionone by engineered cytochrome P450 BM-3. *Applied microbiology and biotechnology, 70*(1), 53-59.

Van Dyke & Gandolf. (1976). Anaerobic release of fluoride from halothane. Relationship to the binding of halothane metabolites to hepatic cellular constituents. *Drug Metabolism and Disposition, 4*(1), 40-44.

Vandamme & Coenye. (2004). Taxonomy of the genus Cupriavidus: a tale of lost and found. *International journal of systematic and evolutionary microbiology, 54*(6), 2285-2289.

Venkataraman, Verkade-Vreeker, Capoferri, Geerke, Vermeulen & Commandeur. (2014). Application of engineered cytochrome P450 mutants as biocatalysts for the synthesis of benzylic and aromatic metabolites of fenamic acid NSAIDs. *Bioorganic & medicinal chemistry, 22*(20), 5613-5620.

Veronese, Largajolli, Boccu, Benassi & Schiavon. (1985). Surface modification of proteins activation of monomethoxy-polyethylene glycols by phenylchloroformates and modification of ribonuclease and superoxide dismutase. *Applied biochemistry and biotechnology, 11*(2), 141-152.

Vieville, Yoo, Pelet & Mouloungui. (1998). Synthesis of glycerol carbonate by direct carbonatation of glycerol in supercritical CO2 in the presence of zeolites and ion exchange resins. *catalysis Letters, 56*(4), 245-247.

Volkin, Staubli, Langer & Klibanov. (1991). Enzyme thermoinactivation in anhydrous organic solvents. *Biotechnology and bioengineering, 37*(9), 843-853.

Waibel, Schulze, Huber & Bachmann. (2006). Screen-printed bienzymatic sensor based on sol–gel immobilized Nippostrongylus brasiliensis acetylcholinesterase and a cytochrome P450 BM-3 (CYP102-A1) mutant. *Biosensors and Bioelectronics, 21*(7), 1132-1140.

Wang, Chen, Cai, Liu, Zheng, Wang, Li & He. (2013). Biosynthesis and thermal properties of PHBV produced from levulinic acid by Ralstonia eutropha. *PLoS One, 8*(4), 60318.

Wang, Dai, Waezsada, Tsao & Davison. (2001). Enzyme stabilization by covalent binding in nanoporous sol-gel glass for nonaqueous biocatalysis. *Biotechnology and Bioengineering, 74*(3), 249-255.

Warman, Robinson, Luciakova, Lawrence, Marshall, Warren, Cheesman, Rigby, Munro & McLean. (2012). Characterization of Cupriavidus metallidurans CYP116B1–A thiocarbamate herbicide oxygenating P450–phthalate dioxygenase reductase fusion protein. *FEBS Journal, 279*(9), 1675-1693.

Watanabe. (2001). Microorganisms relevant to bioremediation. *Current opinion in biotechnology, 12*(3), 237-241.

Wei, Wang, Wang, Meade, Hemann, Hille & Stuehr. (2001). Rapid kinetic studies link tetrahydrobiopterin radical formation to heme-dioxy reduction and arginine hydroxylation in inducible nitric-oxide synthase. *Journal of biological chemistry, 276*(1), 315-319.

Weingärtner, Cabrele & Herrmann. (2012). How ionic liquids can help to stabilize native proteins. *Physical Chemistry Chemical Physics, 14*(2), 415-426.

Welinder. (1992). Superfamily of plant, fungal and bacterial peroxidases. *Current Opinion in Structural Biology, 2*(3), 388-393.

White & Marletta. (1992). Nitric oxide synthase is a cytochrome P-450 type hemoprotein. *Biochemistry, 31*(29), 6627-6631.

Whitehouse, Bell & Wong. (2012). P450BM3 (CYP102A1): connecting the dots. *Chem. Soc. Rev., 41*(3), 1218-1260.

Williams, Cosme, Sridhar, Johnson & McRee. (2000). Mammalian microsomal cytochrome P450 monooxygenase: structural adaptations for membrane binding and functional diversity. *Molecular cell, 5*(1), 121-131.

Wislocki, Miwa & Lu. (1980). Reactions catalyzed by the cytochrome P-450 system. *Enzymatic basis of detoxication, 1*, 135-182.

Wong, Arnold & Schwaneberg. (2004). Laboratory evolution of cytochrome P450 BM-3 monooxygenase for organic cosolvents. *Biotechnology and bioengineering, 85*(3), 351-358.

Yan & Lehe. (2007). Production of Indigo by Immobilization of E. coli BL21 (DE3) Cells in Calcium-Alginate Gel Capsules1. *Chinese Journal of Chemical Engineering, 15*(3), 387-390.

Yedavalli & Rao. (2013). Engineering the loops in a lipase for stability in DMSO. *Protein Engineering Design and Selection, 26*(4), 317-324

Yu & Stahl. (2008). Microbial utilization and biopolyester synthesis of bagasse hydrolysates. *Bioresource technology, 99*(17), 8042-8048.

Zaks & Klibanov. (1984). Enzymatic catalysis in organic media at 100 degrees C. *Science, 224*(4654), 1249-1251.

Zaks & Klibanov. (1985). Enzyme-catalyzed processes in organic solvents. *Proceedings of the National Academy of Sciences, 82*(10), 3192-3196.

Zaks & Russell. (1988). Enzymes in organic solvents: properties and applications. *Journal of Biotechnology, 8*(4), 259-269.

Zhao, Güven, Li & Schwaneberg. (2011). First steps towards a Zn/Co (III) sep-driven P450 BM3 reactor. *Applied microbiology and biotechnology, 91*(4), 989-999.

Zumárraga, Bulter, Shleev, Polaina, Martínez-Arias, Plou, Ballesteros & Alcalde. (2007). In Vitro Evolution of a Fungal Laccase in High Concentrations of Organic Cosolvents. *Chemistry & Biology, 14*(9), 1052-1064.

Appendices

Appendix I: Sequences alignment

B2406 heme domain:

Heme domain sequence alignment between B2406 and crystal structure (PDB ID:5HDI.A.). The model was suggested as a best hit by SWISS MODEL and Phyre online tools. Residues that bound to the heme cofactor were highlighted in yellow. (|) indicating similar identity, (.) is for strongly similar residues and (:) to indicate weak similarity. The alignment was produced by using ClustalW online tool

```
B2406           1 MKHHHHHHPMSDYDIPTTENLYFQGAHMPDTDIDPLSAVTHPDPYPYYRE   50
                             ..|....|.:|             .|.:..||||.|..
5HDI            1 ------------MTIAKDANTFF-------------GAESVQDPYPLYER   25

B2406          51 LAASQPFFRDDRLGLWVAAGPQEVADVLAHSDCRVRPP--AQPVPPALAG   98
                  :.|:....|......:...|...|.:.:.      ||.  :..:...:..
5HDI           26 MRAAGSVHRIANSDFYAVCGWDAVNEAIG------RPEDFSSNLTATMTY   69

B2406          99 TAAG--------ELFG--RLVRMNDGAAHAPLKALLMPMLAGIDPAAAAQ  138
                  ||.|         .|.|  .:::...|..|||..:.|::..|       ||:
5HDI           70 TAEGTAKPFEMDPLGGPTHVLATADDPAHAVHRKLVLRHL-------AAK  112

B2406         139 RATVLAA-VLDAGEASWA-AMSGECINRWLFT----LPVVTVADLLGLPV  182
                  |..|:.. .:.|.:..|. .|...|| .|:..    ||::.||:|:|||
5HDI          113 RIRVMEQFTVQAADRLWVDGMQDGCI-EWMGAMANRLPMMVVAELIGLP-  160

B2406         183 ANEGSSAAEAAQRVA-AFAGAQ--SPLADAPAVRAGAEAAQWLGHWLADA  229
                        ..:.||.|. .:|..|  ..|.:...:.|...|...|..::.:.
5HDI          161 ------DPDIAQLVKWGYAATQLLEGLVENDQLVAAGVALMELSGYIFEQ  204

B2406         230 ADGAGPLP------ALRQAARAAGIDAQAVAANIIGLLVQACEATAALAG  273
                  .|.|...|      .|..|..:..:|.......:.|.....|:||||.|
5HDI          205 FDRAAADPRDNLLGELATACASGELDTLTAQVMMVTLFAAGGESTAALLG  254

B2406         274 NTLLRLG-RDTTQSGLP-----LDAVVARVAREDPPVQNTRRFLAADAQL  317
                  :.:..|. |...|..:.     |.|.:....|.:||.:...|.:.....|
5HDI          255 SAVWILATRPDIQQQVRANPELLGAFIEETLRYEPPFRGHYRHVRNATTL  304

B2406         318 CGHAVKAGDAVLVLLAAASCSGAAASERPW-------------TFGHGRH  354
                  .|..:.| |:.|:||..|:....|..|.|.             :||.|.|
5HDI          305 DGTELPA-DSHLLLLWGAANRDPAQFEAPGEFRLDRAGGKGHISFGKGAH  353

B2406         355 ACPGDRLAQALAAATVAALRARGADPAALAQAFRYRPSLNARIPHFL---  401
                  .|.|..||: |.|..|..|..........|....:.||:..|....|
5HDI          354 FCVGAALAR-LEARIVLRLLLDRTSVIEAADVGGWLPSILVRRIERLELA  402
```

B1743 heme domain:

Heme domain sequence alignment between B1743 and crystal structure (PDB ID: 2UVN. A.). The model was suggested as a best hit by SWISS MODEL and Phyre online tools. Residues that bound to the heme cofactor were highlighted in yellow. (|) indicating similar identity, (.) is for strongly similar residues and (:) to indicate weak similarity. The alignment was produced by using ClustalW online tool.

```
B1743          1 MKHHHHHHPMSDYDIPTTENLYFQGAHMTDTNQHALLHDGYDLLSDHYVQ     50
                   .|||||                   |||....|..
2UVN.          1 -MHHHHH------------------HMTSVMSHEF--------------     16

B1743         51 EAHALWRDIRSSGCPVAHSEKWGGSWLP-----TTYDDIHHVA--QNPA-     92
                            .:|.:|.|...| |     ..:|.:|||.  |.|.
2UVN.         17 --------------QLATAETWPNPW-PMYRALRDHDPVHHVVPPQRPEY     51

B1743         93 --VFSSRAAEI------------------AGEVPPQGSGLVLPPLTSDP    121
                   ...||.|::                  .||:...|.....|.:..||
2UVN.         52 DYYVLSRHADVWSAARDHQTFSSAQGLTVNYGELEMIGLHDTPPMVMQDP    101

B1743        122 PDHKIHRDLLEPYFTPARVAAIEPYAQSLARDLARRVAVKGEADLGEDYS    171
                 |.|...|.|:...|||.:|..:||..:...:...::...|..|:..:..
2UVN.        102 PVHTEFRKLVSRGFTPRQVETVEPTVRKFVVERLEKLRANGGGDIVTELF    151

B1743        172 KPFVLSLLTRFLDVPDDRQERFMDWAIRVLKYGPFDQELRKA--AFDEAF    219
                 ||....::..:|.||::...:|..|...::.....|.....|  |.....
2UVN.        152 KPLPSMVVAHYLGVPEEDWTQFDGWTQAIVAANAVDGATTGALDAVGSMM    201

B1743        220 ADLEQLLKEREQDPGEDLVSHIALATI--DGKPISRKHRIGSLLLAVLAG    267
                 |....|::.|..:|.:|.:||:..|.:  ||........:.....|..|
2UVN.        202 AYFTGLIERRRTEPADDAISHLVAAGVGADGDTAGTLSILAFTFTMVTGG    251

B1743        268 ADTTWNALNASLNHLADHPADRATLINEP-GLLRTTAVEELLRFYAPL-S    315
                 .||....|..|:..|...|..|..|:::| |:  ..|||||||..:|: .
2UVN.        252 NDTVTGMLGGSMPLLHRRPDQRRLLLDDPEGI--PDAVEELLRLTSPVQG    299

B1743        316 IARVTTEEVELKGRCIGAGERVILAYPAANRDPAVFENPDEVQLDRKR--    363
                 :||.||.:|.:....|.||.||:|.|.:||||...: .||..:||..|
2UVN.        300 LARTTTRDVTIGDTTIPAGRRVLLLYGSANRDERQY-GPDAAELDVTRCP    348

B1743        364 NRHLTFGVGVHRCLGSHLARMEMRVAIEEWLKAIPNFERISGAVKWSAGN    413
                 ...|||..|.|.|||:..|||:.|||:.|.|...|:||.....:.||.|:
2UVN.        349 RNILTFSHGAHHCLGAAAAARMQCRVALTELLARCPDFEVAESRIVWSGGS    398

B1743        414 ARGPENVRIRVV    425
                    ..
2UVN.        399 YV----------    400
```

B1279 heme domain:

Heme domain sequence alignment between B1279 and crystal structure (PDB ID: 6GII.A.). The model was suggested as a best hit by SWISS MODEL and Phyre online tools. Residues that bound to the heme cofactor were highlighted in yellow. (|) indicating similar identity, (.) is for strongly similar residues and (:) to indicate weak similarity. The alignment was produced by using ClustalW online tool

```
B1279           1 --MKHHHHHH-----PMSDYDIPTTENLYF-QGAMANPSPSAARSGCPID     42
                    ..||||||     |...:....|..... :|:|.......||..||:.
6GII.           1 MGSSHHHHHHSSGLVPRGSHMASMTGGQQMGRGSMETELKETARGTCPVA     50

B1279          43 HSALTAPNGCPVSHNAAQFDPFGDGYQQDPPEYVRWSREQEPVFYSPRLG     92
                  |...::..||||...|..||||.|.|..||.::|||:|||.|:||:|:|.
6GII.          51 HGGQSSVGGCPVHRLAEDFDPFQDAYMADPAQFVRWAREQVPIFYAPKLN    100

B1279          93 YWVVTRYEDIKAIFRDNLTFSPSIALEKITPTGDEANAVLASYGYAMNRT    142
                  |||||||:.||.||||.:|||||..|:.......|...||..||||.|||
6GII.         101 YWVVTRYDTIKQIFRDPVTFSPSNVLQSFAQPSAEVRQVLERYGYAFNRT    150

B1279         143 LVNEDEPAHMPRRRALMAPFTPAELAHHEPLVRRLAREYVDRFIDDGRAD    192
                  |||||||.|:.||.||.||....||.|||:||.|.|..|:||||.||||
6GII.         151 LVNEDEPMHLERRRVLMEPFASEHLAEHEPMVRELVRRAVNRFIDTGRAD    200

B1279         193 LVDQMLWEVPLTVALHFLGVPEEDMDLLRQYSIAHTVNTWGRPKPEEQVA    242
                  |||||:||||.|||||||||.::|.:.:|:::|||||.:|||.||||:|
6GII.         201 LVDQMIWEVPFTVALHFLGVDDDDREKMRRFAIAHTVNAFGRPSPEEQLA    250

B1279         243 VAHAVGNFWQLAGRILDKMREDPSGPGWMQYGLRKQKELPDVVTDSYLHS    292
                  ||..||.|||..|.:|:|||....|||||:|.:|:||..||||||||||
6GII.         251 VAETVGQFWQFCGEVLEKMRRTADGPGWMRYSIRQQKLYPDVVTDSYLHS    300

B1279         293 MMMAGIVAAHETTANASANAIKLLLQHPDAWREICDDPSLIPNAVEECLR    342
                  ||.|.||||||||..|:.||:|.||:|...|||||.||||||.|.|||||
6GII.         301 MMQAIIVAAHETTVFATTNALKTLLEHETVWREICADPSLIPAAAEECLR    350

B1279         343 HNGSVAAWRRLVTRDAEVGGIRLPAGSKLLIVTSSANHDERHFADADLFD    392
                  :||.||.|||..||:.||.|:|||.|:.:|:|.:||||||..||.|.:.||
6GII.         351 YNGPVAGWRRRTTREVEVEGVRLPVGANILMVVASANHDSAHFDDPEFFD    400

B1279         393 IRRDNASEQLTFGYGSHQCMGKNLARMEMQVFLEELTRRLPHMRLAEQTF    442
                  |.|.||||.|.||||:|||:|:||.|||||:.:|||:||||||||||.|
6GII.         401 IGRSNASEHLNFGYGAHQCLGRNLGRMEMQIMIEELSRRLPHMRLAEQRF    450

B1279         443 TYVPNTSFRGPEHLLVEWDPAQNPERRDPALLEVHQPVRIGEPSAHTIAR    492
                  .|:.|.|||.|.||.||||||||||||||
6GII.         451 DYLHNVSFRAPRHLWVEWDPAQNPERRDP----------------------    479
```

B1009 heme domain:

Heme domain sequence alignment between B1009 and crystal structure (PDB ID: 1BU7.A.). The model was suggested by SWISS MODEL online tool. Residues that bound to the heme cofactor were highlighted in yellow. (|) indicating similar identity, (.) is for strongly similar residues and (:) to indicate weak similarity. The alignment was produced by using ClustalW online tool

```
B1009          1 MKHHHHHHPMSDYDIPTTENLYFQGAHMPPPIELSSPDQASDAPHQAPAR       50

1bu7.          1 --------------------------------------------------         0

B1009         51 SMHAPIPEPIPRDPGWPLVGNLLQITPGALGQHLLARSRHHDGIFELNFA      100
                    .|.| :|:...:..:.||..:......|.|:..:.....||:....
1bu7.          1 ----TIKE-MPQPKTFGELKNLPLLNTDKPVQALMKIADELGEIFKFEAP       45

B1009        101 GRRVPFVTSVALASELCDAAQFRKYIGPPVSYLRGMAGDGLFTARSDEAN      150
                 ||...:::|..|..|.||.::|.|.:...:.::|..|||||||:.:.|.|
1bu7.         46 GRVTRYLSSQRLIKEACDESRFDKNLSQALKFVRDFAGDGLFTSWTHEKN       95

B1009        151 WGKAHRILMPAFSQRAMKGYFDVMLRVANRLVDKWDQQGPDADIAVADDM      200
                 |.|||.||:|:|||:|||||..:|:.:|.:||.||::...|..|.|.:||
1bu7.         96 WKKAHNILLPSFSQQAMKGYHAMMVDIAVQLVQKWERLNADEHIEVPEDM      145

B1009        201 TRLTLDTIALSGFGYDFESFASTELHPFIEAMVGALEEAMSKLTRFALQD      250
                 ||||||||.|.||.|.|.||...:.||||.:||.||:|||:||.|....|
1bu7.        146 TRLTLDTIGLCGFNYRFNSFYRDQPHPFITSMVRALDEAMNKLQRANPDD      195

B1009        251 RFMHAAHQKFDQDTRFMRDLVDDVIRRRRAGDAAERPGGTANDLLGLMLE      300
                 .......::|.:|.:.|.||||.:|..|:|.      |..::|||..||.
1bu7.        196 PAYDENKRQFQEDIKVMNDLVDKIIADRKAS------GEQSDDLLTHMLN      239

B1009        301 ARDPDTDQRLDDENIRNQVITFLIAGHETTSGLLTFALYELLRNPGVMAQ      350
                 .:||:|.:.|||||||.|:|||||||||||||:||||.|::||.|:.:
1bu7.        240 GKDPETGEPLDDENIRYQIITFLIAGHETTSGLLSFALYFLVKNPHVLQK      289

B1009        351 AYAEVDAVLPGDAAPVYADLARLPVLDRVLKETLRLWPTAPAFAVAPFED      400
                 |..|...||. |..|.|..:.:|..:..||.|.||||||||||::...||
1bu7.        290 AAEEAARVLV-DPVPSYKQVKQLKYVGMVLNEALRLWPTAPAFSLYAKED      338

B1009        401 TLLGGRYLIRKDRRLSVVLTALHRDPKVWADPERFDIDRFLPEQ---EAK      447
                 |:|||.|.:.|...|.|::..||||..:|.|    |::.|.||:   .:.
1bu7.        339 TVLGGEYPLEKGDELMVLIPQLHRDKTIWGD----DVEEFRPERFENPSA      384

B1009        448 LPRHAYMPFGNGERACIGRQFALTEAKLALALMLRNFQFTDAHDYQFRIK      497
                 :|:||:.|||||:|||
1bu7.        385 IPQHAFKPFGNGQRAC---------------------------------      400
```

B1009 heme domain:

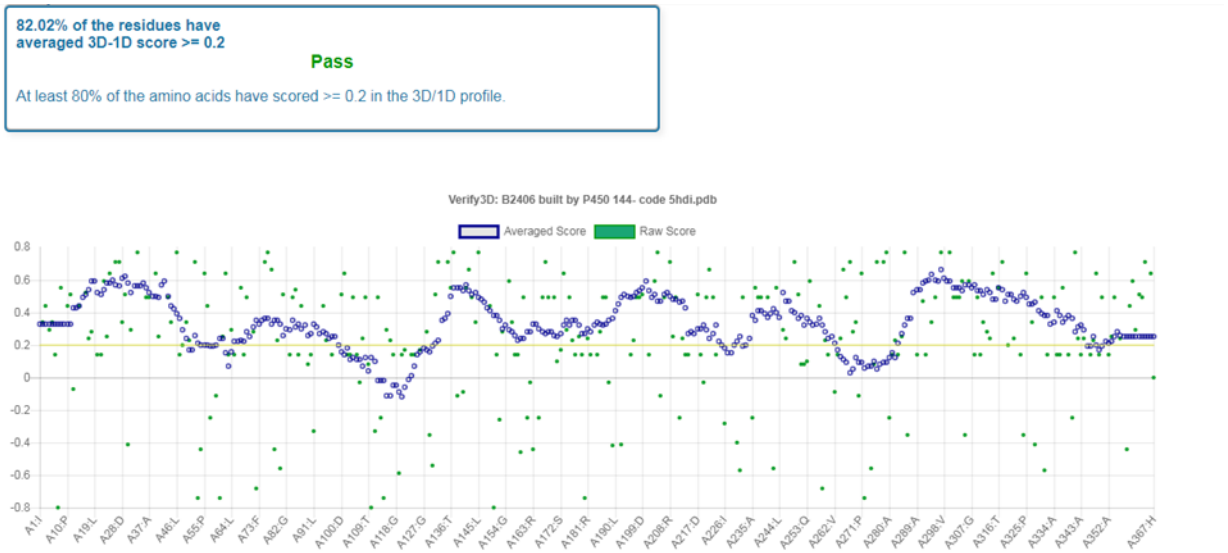
Heme domain sequence alignment between B1009 and crystal structure (PDB ID: 2ij2.A). The model was suggested as a best hit by Phyre2.2 online tool. Residues that bound to the heme cofactor were highlighted in yellow. (|) indicating similar identity, (.) is for strongly similar residues and (:) to indicate weak similarity. The alignment was produced by using ClustalW online tool

```
B1009          1 MKHHHHHHPMSDYDIPTTENLYFQGAHMPPPIELSSPDQASDAPHQAPAR      50

2IJ2.          1 --------------------------------------------------       0

B1009         51 SMHAPIPEPIPRDPGWPLVGNLLQITPGALGQHLLARSRHHDGIFELNFA    100
                     .|.| :|:...:..:.||..:......|.|:..:.....||:....
2IJ2.          1 -----TIKE-MPQPKTFGELKNLPLLNTDKPVQALMKIADELGEIFKFEAP     45

B1009        101 GRRVPFVTSVALASELCDAAQFRKYIGPPVSYLRGMAGDGLFTARSDEAN    150
                 ||...:::|..|..|.||.::|.|.:...:.::|..|||||||:.:.|.|
2IJ2.         46 GRVTRYLSSQRLIKEACDESRFDKNLSQALKFVRDFAGDGLFTSWTHEKN     95

B1009        151 WGKAHRILMPAFSQRAMKGYFDVMLRVANRLVDKWDQQGPDADIAVADDM    200
                 |.|||.||:|:|||:|||||..:|:.:|.:||.||::...|..|.|.:||
2IJ2.         96 WKKAHNILLPSFSQQAMKGYHAMMVDIAVQLVQKWERLNADEHIEVPEDM    145

B1009        201 TRLTLDTIALSGFGYDFESFASTELHPFIEAMVGALEEAMSKLTRFALQD    250
                 ||||||||.|.||.|.|.||...:.||||.:||.||:|||:||.|....|
2IJ2.        146 TRLTLDTIGLCGFNYRFNSFYRDQPHPFITSMVRALDEAMNKLQRANPDD    195

B1009        251 RFMHAAHQKFDQDTRFMRDLVDDVIRRRRAGDAAERPGGTANDLLGLMLE    300
                 .......::|.:|.:.|.||||.:|..|:|.      |..::|||..||.
2IJ2.        196 PAYDENKRQFQEDIKVMNDLVDKIIADRKAS------GEQSDDLLTHMLN    239

B1009        301 ARDPDTDQRLDDENIRNQVITFLIAGHETTSGLLTFALYELLRNPGVMAQ    350
                 .:||:|.:.|||||||.|:|||||||||||||:||||.|::||.|:.:
2IJ2.        240 GKDPETGEPLDDENIRYQIITFLIAGHETTSGLLSFALYFLVKNPHVLQK    289

B1009        351 AYAEVDAVLPGDAAPVYADLARLPVLDRVLKETLRLWPTAPAFAVAPFED    400
                 |..|...||. |..|.|..:.:|..:..||.|.||||||||||::...||
2IJ2.        290 AAEEAARVLV-DPVPSYKQVKQLKYVGMVLNEALRLWPTAPAFSLYAKED    338

B1009        401 TLLGGRYLIRKDRRLSVVLTALHRDPKVWADPERFDIDRFLPEQ---EAK    447
                 |:|||.|.:.|...|.|::..||||..:|.|    |::.|.|||:   .:.
2IJ2.        339 TVLGGEYPLEKGDELMVLIPQLHRDKTIWGD----DVEEFRPERFENPSA    384

B1009        448 LPRHAYMPFGNGERACIGRQFALTEAKLALALMLRNFQFTDAHDYQFRIK    497
                 :|:||:.|||||:|||||:||||.||.|.|.:||::|.|.|..:|:..||
2IJ2.        385 IPQHAFKPFGNGQRACIGQQFALHEATLVLGMMLKHFDFEDHTNYELDIK    434

B1009        498 ETLTIKPDGFTVRARRRRPHERIAAAPLCTAQAPRAGPDVQG-          539
                 ||||:||:||.|:|:.::       .||....:|......:.
2IJ2.        435 ETLTLKPEGFVVKAKSKK-------IPLGGIPSPSTEQSAKKV          470
```

Appendix II: Models qualification evaluation performed by VERIFY 3D. The model considers accurate if 80% or more of protein's residues showed average (3D-1D) ≥ 0.2

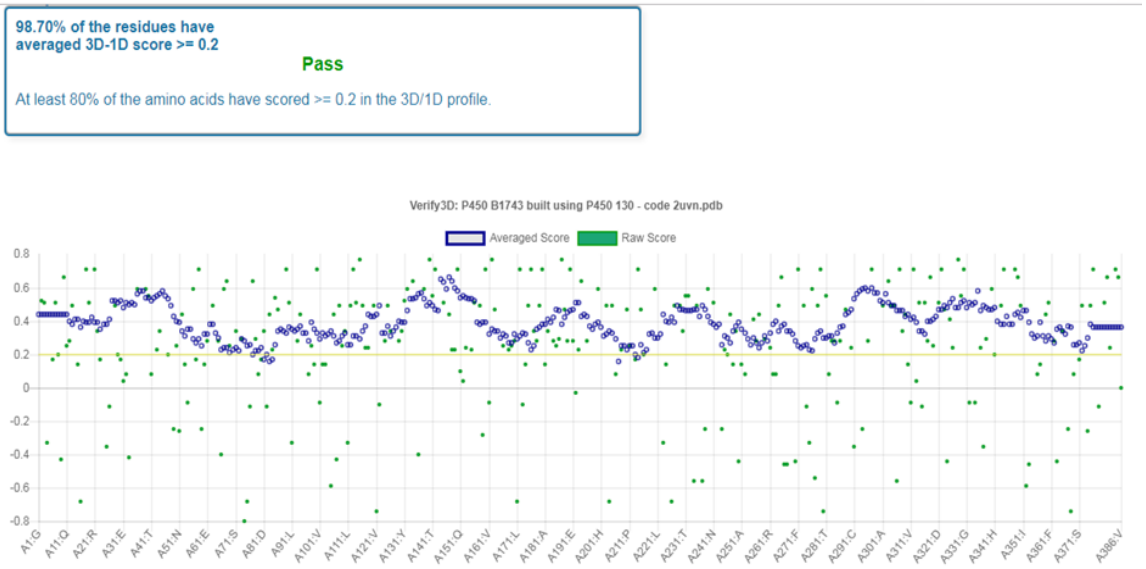
B2406 heme domain:

VERIFY 3D results of B2406 heme domain model built by SWISS MODEL using the crystal structure (PDB ID: 5hdi. 1. A.). more than 82.02% protein's residues showed average (3D-1D) ≥ 0.2



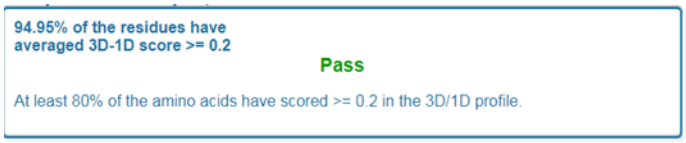
B1743 heme domain:

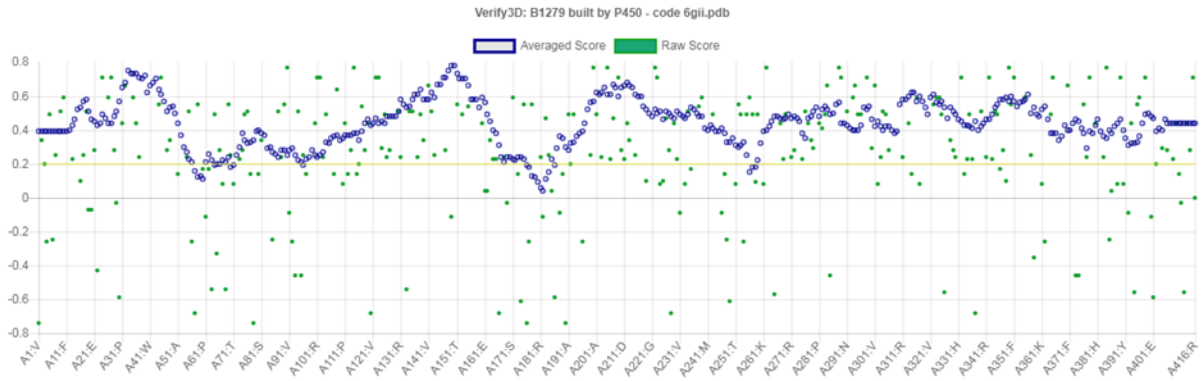
VERIFY 3D results of B1743 heme domain model built by SWISS MODEL using the crystal structure (PDB ID: 2uvn. 1. A.). more than 98.70% of protein's residues showed average (3D-1D) ≥ 0.2



B1279 heme domain:

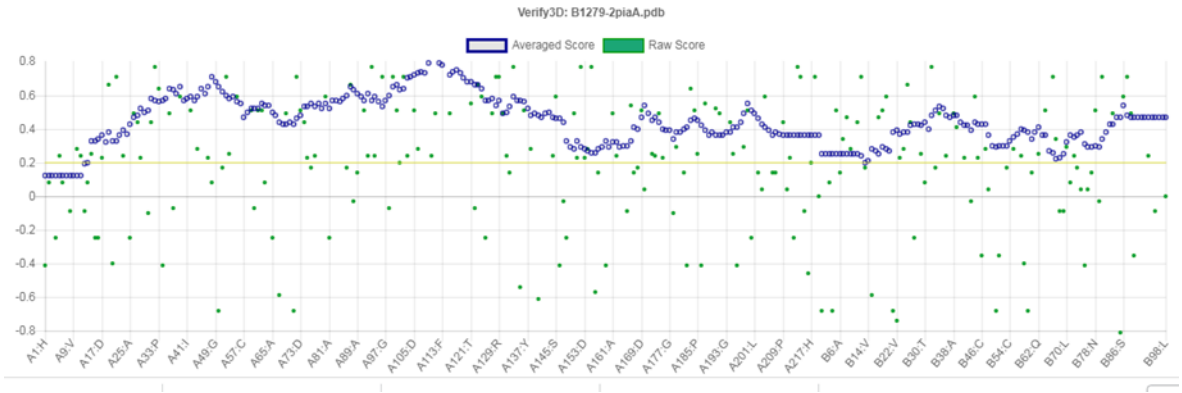
VERIFY 3D results of B1279 heme domain model built by SWISS MODEL using the crystal structure (PDB ID: 6gii. 1. A.). more than 94.95% of protein's residues showed average (3D-1D) ≥ 0.2





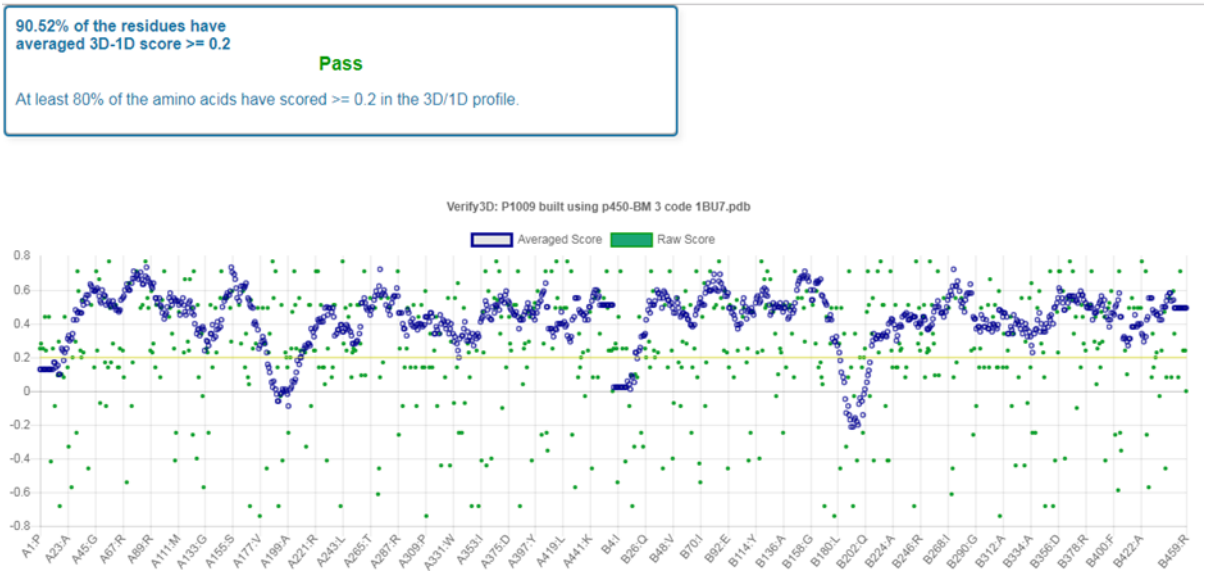
B1279 reductase domain:

VERIFY 3D results of B1279 reductase domain model built by Phyre2.2 using the crystal structure (PDB ID: 2pia. 1. A.). more than 94.21% of protein's residues showed average (3D-1D) ≥ 0.2



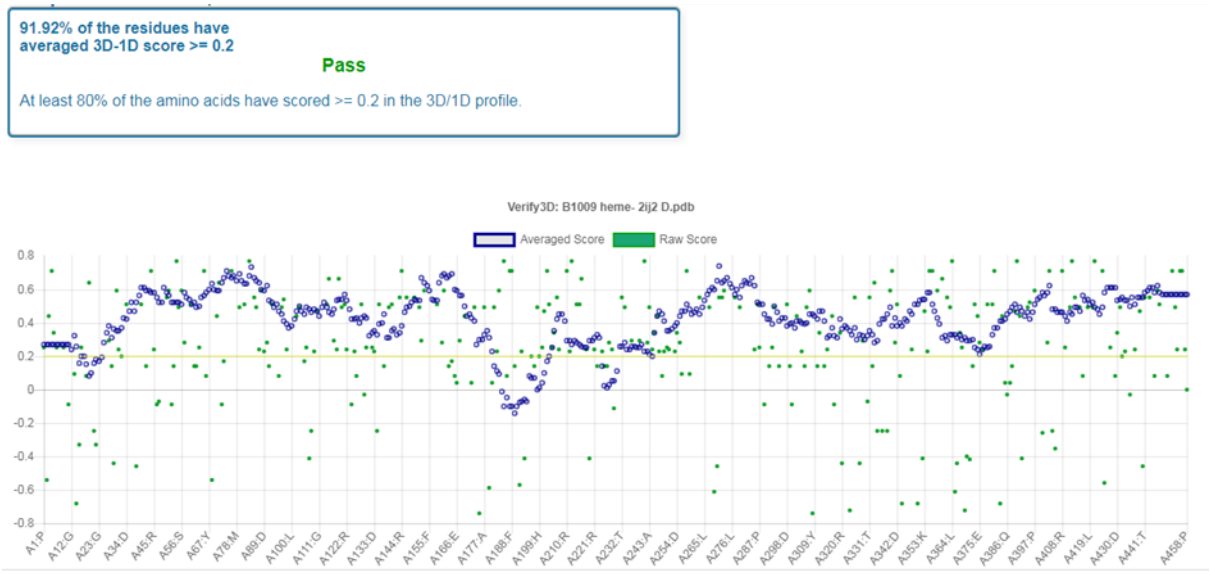
B1009 heme domain:

VERIFY 3D results of B1009 heme domain model built by SWISS MODEL using the crystal structure (PDB ID: 1bu7. 1. A.). more than 90.52% of protein's residues showed average (3D-1D) ≥ 0.2



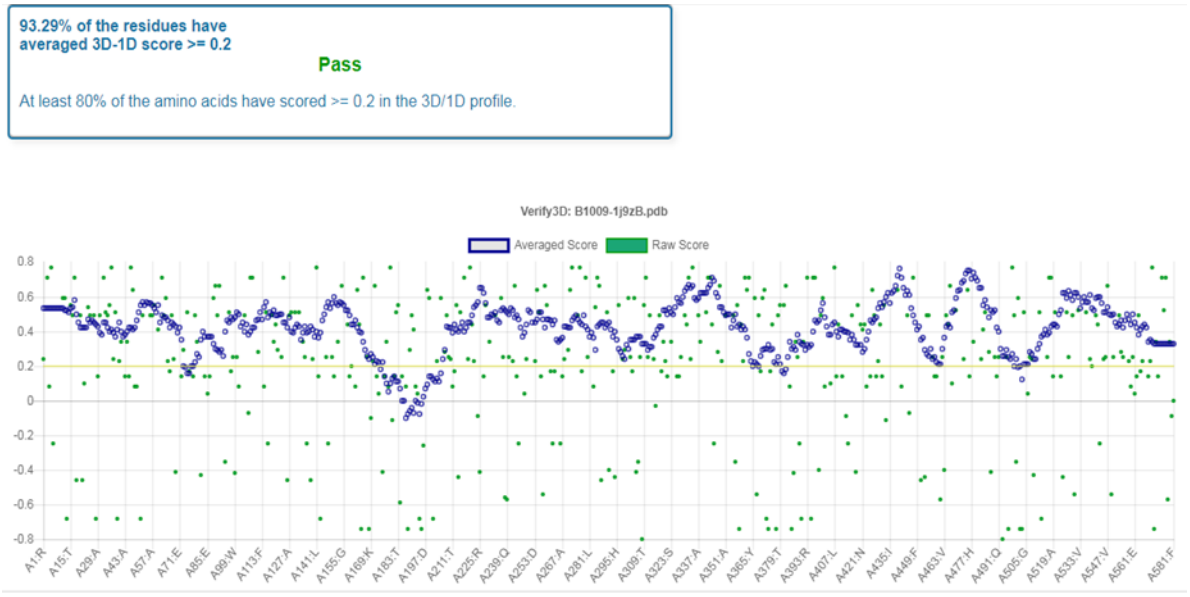
B1009 heme domain:

VERIFY 3D results of B1009 heme domain model built by Phyre2.2 using the crystal structure (PDB ID: 2ij2. 1. A.). more than 91.92% of protein's residues showed average (3D-1D) ≥ 0.2



B1009 reductase domain:

VERIFY 3D results of B1009 reductase domain model built by Phyree2.2 using the crystal structure (PDB ID: 1j9z. 1. A.). more than 93.29% of protein's residues showed average (3D-1D) ≥ 0.2



Appendix III: List of amino acids Abbreviations

Amino acid	3-letter	1-letter	Side chain class	Side chain polarity	Side chain charge
Alanine	Ala	A	aliphatic	nonpolar	neutral
Arginine	Arg	R	basic	basic polar	positive
Asparagine	Asn	N	amide	polar	neutral
Aspartic acid	Asp	D	acid	acidic polar	negative
Cysteine	Cys	C	sulfur-containing	nonpolar	neutral
Glutamic acid	Glu	E	acid	acidic polar	negative
Glutamine	Gln	Q	amide	polar	neutral
Glycine	Gly	G	aliphatic	nonpolar	neutral
Histidine	His	H	basic aromatic	basic polar	positive (10%) neutral (90%)
Isoleucine	Ile	I	aliphatic	nonpolar	neutral
Leucine	Leu	L	aliphatic	nonpolar	neutral
Lysine	Lys	K	basic	basic polar	positive
Methionine	Met	M	sulfur-containing	nonpolar	neutral
Phenylalanine	Phe	F	aromatic	nonpolar	neutral
Proline	Pro	P	cyclic	nonpolar	neutral
Serine	Ser	S	hydroxyl-containing	polar	neutral
Threonine	Thr	T	hydroxyl-containing	polar	neutral
Tryptophan	Trp	W	aromatic	nonpolar	neutral
Tyrosine	Tyr	Y	aromatic	polar	neutral
Valine	Val	V	aliphatic	nonpolar	neutral

Appendix IV: List of Publications

Publications in preparation

"Characterization and genetic analysis of putative cytochrome P450s from *Cupriavidus necator* H16". Article.

"Protein engineering for tolerance and stability improvement in non-conventional solvents". Review.

Book chapter

Omar Ali, H., Alessa, A. H. A., Al-nuaemi, I. J., Wong, T. S. "Protein engineering for lignocellulose degradation." Environmental Science and Engineering Volume 8 (2017). 1st ed. USA: Studium Press LLC.

Oral Presentation

Activity Evaluation of Cytochrome P450s, Postgraduate Research Conference, Jun 2017, Sheffield, UK.

Poster Presentations

Cytochrome P450 BM-3 tolerances in non-conventional green solvents, The First Iraqi Conference, October 2016, Sheffield, UK

Cytochrome P450 BM-3 W5F5 expression and purification optimisation, Postgraduate Research Conference, January 2016, Sheffield, UK