Biotransformations of Proline by 2-oxoglutarate-dependent hydroxylases

Muhiadin Omar

PhD

University of York

Chemistry

September 2017

Abstract

Hydroxylases introduce hydroxyl groups with excellent regio- and enantioselectivity making them of significant interest for use in the production of pharmaceutical intermediates and drug metabolites. 2-oxoglutarate dependent oxygenases (2OGDOs) are non-haem dependent Fe(II) containing enzymes that catalyse various oxidation reactions, including the hydroxylation of free amino acids. Unlike the more studied cytochromes P450, these enzymes only require molecular oxygen, Fe(II) and 2-oxoglutarate for catalysis, circumventing the need for a costly cofactor regeneration system.

The targets of this work were three proline hydroxylases: a *trans*-4-proline hydroxylase from *Dactylsporangium* sp. RH1 (DOGDH), a *cis*-3-proline-hydoxylase from *Streptomyces* sp. (StP3H) and a *cis*-4-proline-hydroxylase from *Mesorhizobium loti* (MlC4H). Genes encoding all three were cloned into the pET-YSBLIC3C (and pET22b for DOGDH) expression vectors, expressed in *Escherichia coli*, and produced and purified by chromatography for use in crystallisation studies and enzymatic transformations. Extensive crystallisation trials were attempted for DOGDH including enzymatic, chemical and mutagenic modification with little success. A homology model was therefore constructed in order to identify catalytic residues within the active site that could be manipulated for enhancing the function of DOGDH.

A precolumn derivatisation assay using FMOC-Cl was developed for the analysis of proline and its hydroxylated equivalents by HPLC and LC-MS. Biotransformations were performed with L-proline using the three hydroxylases with whole cell reaction conditions deemed optimal due to the multi-component nature of the enzymes, with the cell providing machinery for the recycling of cofactors. Reactions were scaled from shake flasks to stirred tank vessels with the flow of air into the vessel and stirring rate deemed key parameters for optimal function. Finally, a high-throughput substrate screening method using a BioLecter micro-bioreactor was successfully developed and trialled with the three hydroxylases with a panel of substrates providing a platform for future investigations.

Contents

List of Figures

List of Tables

Acknowledgements

First and foremost, I would like to thank my supervisor Professor Gideon Grogan for giving me the chance to work for him on this PhD project and for the most excellent supervision he provided throughout. There is no way that I would have made it here without his support and guidance. I also want to thank Professor Neil Bruce for serving as my IPM and for the great advice he provided in TAP meetings. I also thank Dr Michael Lloyd for his role as my industrial supervisor at Dr Reddys, I couldn't have asked for a better person to have my back in the cut-throat world of industry.

I thank Simon Grist, Sally Lewis, Juliet Borgia, Louise Haigh and David Neale for their continued technical support in the YSBL and M2. I also thank David Ashford, Karl Heaton, Liz Rylott and George Evans for training and helping me with HPLC and LC-MS work, I would likely have lost my mind without them. I want to give a shout-out to the fermentation boys at Dr Reddys, Neil Triggs and Matthew Bycroft, for their help with the large scale fermentations and BioLecter work. I also want to thank Ian Taylor and Toni Fleming at Dr Reddys for their help in the microbiology labs. I am also grateful for the support of Richard Lloyd and Karen Holt-Tiffin at Dr Reddys.

Now for the Grogan group: my comrades in arms. Thank you to Tamara Mielke, Amina Frese, Anibal Cuetos-Fernandez, Mahima Sharma, Mark Petchey, Henry Man, Hamid-Reza Danesh-Azari, Elizabeth Wells and Claudia Spandolf for being the best group ever. I also want to thank Imogen Breen and Pasky Miranda for their support and friendship throughout the entirety of my time at the University of York.

Finally, I would like to thank my parents Maye and Maimuna and aunt Jimiko for always supporting and believing in me. I also want to thank my little brother Ahmed for all the financial bail outs.

Author's Declaration

This work has not previously been presented for an award at this, or any other, University. Except where stated, all the work presented in this thesis is my own. Any contribution from others is acknowledged as appropriate below or within the text.

Figures of protein structures were generated and provided by Gideon Grogan. DOGDH and StP3H were cloned into the pET-YBSLIC3C vector by Laila Roper in a previous PhD project. The BioLecter work was performed with the assistance of Neil Triggs. LC-MS analysis was conducted with the help of Karl Heaton and George Evans. Mass Spectrometry analysis of purified protein was performed by Adam Dowle at the Technology Facility in the Department of Biology at the University of York.

Chapter 1

Introduction

1.1 Enzymes in Chemistry

1.1.1 Biocatalysis

Biological catalysis has been employed by humans for thousands of years in the fermentative production and preservation of foodstuffs such as bread, cheese, wine and vinegar.[1] It was not, however, until the 19th century that scientists finally addressed the question of whether the entity responsible for fermentation was a chemical substance or a living species.[2-3] Eduard Büchner proved the latter to be true in 1897 by publishing the observation that cell-free extracts of yeast were able to ferment sugar to alcohol and carbon dioxide, proposing that a species known as 'zymase' found in yeast cells was responsible for this transformation. The biochemical pathway for fermentation was identified by the work of Embden and Meyerhoff, which was a significant milestone in mechanistic enzymology as this was the first biochemical pathway to ever be elucidated.[3]

In 1858, Louis Pasteur reached a milestone in biocatalysis after discovering that treating an aqueous solution of racemic tartaric acid ammonium salt with a culture of *Penicillium glaucum* mould resulted in the consumption of (+)-tartaric acid along with the simultaneous enrichment of the (-)-isomer; a process considered to be a forerunner of enzyme-catalysed kinetic resolution and widely recognised in both industry and academia today.[4] Pasteur's seminal work was followed in 1894 by Emil Fischer who demonstrated that the enzyme responsible for hydrolysing sucrose, which he termed 'invertin', only reacts with α-D-glucosides, while a second enzyme 'emulsin' only reacts with β-D-glucosides. This was a demonstration of the exquisite selectivity of enzyme catalysis. Fischer used this finding to suggest that the invertin and emulsin consisted of "asymmetrically built molecules", and that the enzyme and substrate fit together like a lock and key – this resulted in the 'lock and key' hypothesis of stereoselective enzyme catalysis which would remain a considerable allegory for enzyme action in the years that followed.[5]

The discovery that enzymes facilitate biological catalysis had a significant impact on the growth of biochemistry. It was however, only in the last 30 years or so that research into the use of enzymes as biocatalysts in organic chemistry gained prominence.

1.1.2 Biocatalysts in Organic Chemistry

The use of biocatalysts in chemical synthesis has increased over the last few decades.[6-9] Biocatalysts display excellent regio-, enantio- and stereoselective properties, low toxicity, high biodegradability and efficiency under mild reaction conditions, making these enzymes conform to the principle of green catalysis.[10]

Enzymes have the capacity to act as chiral catalysts due to the chiral environment of their active sites. The chiral nature of an active site ensures that an enzyme is capable of naturally binding one specific enantiomer of a chiral substrate, granting the ability to catalyse chiral functionalisation and kinetic resolutions.[11] Enzymes are becoming increasingly significant in drug discovery due to their ability to catalyse the resolution of racemic chemical mixtures to optically active compounds; a process that is extremely challenging by chemical means. In addition to the aforementioned stereoselectivity, enzymes are also capable of performing regioselective reactions by functionalising one specific chemically equivalent site in a compound. An example of a regio- and stereoselective biocatalytic reaction is the oxidative hydroxylation of the steroid progesterone to the 11α-product by the fungi *Aspergillus niger* or *Rhizopus arrhizus* (Figure 1.1), a compound that can be chemically modified to form cortisone and hydrocortisone.[12] It was later discovered that hydroxylations of this nature occurred as a result of catalysis by cytochrome P450 enzymes.[13] This biocatalytic process for the production of hydrocortisone provided a significant and viable alternative to the 40-step chemical synthesis first proposed by Woodward and co-workers.[6-7]
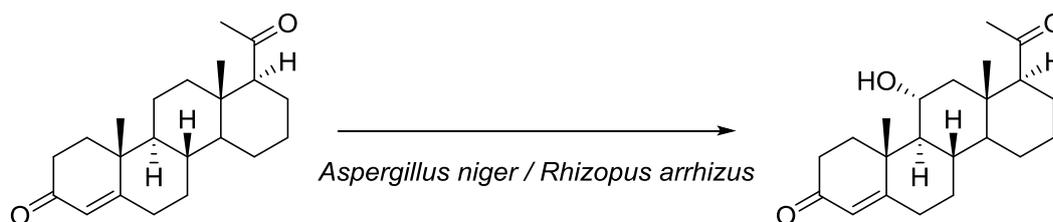
Figure 1.1: Regio- and stereoselective hydroxylation of progesterone to the 11α product by *Aspergillus niger* and *Rhizopus arrhizus*.

Natural products have long had a significant role in modern medicine and the isolation of such compounds has contributed to the discovery of a large number of biologically active compounds ranging from penicillin, morphine and quinine to anti-cancer drugs such as taxol isolated from yew bark (Figure 1.2).

| Penicllin G | Morphine | Quinine |
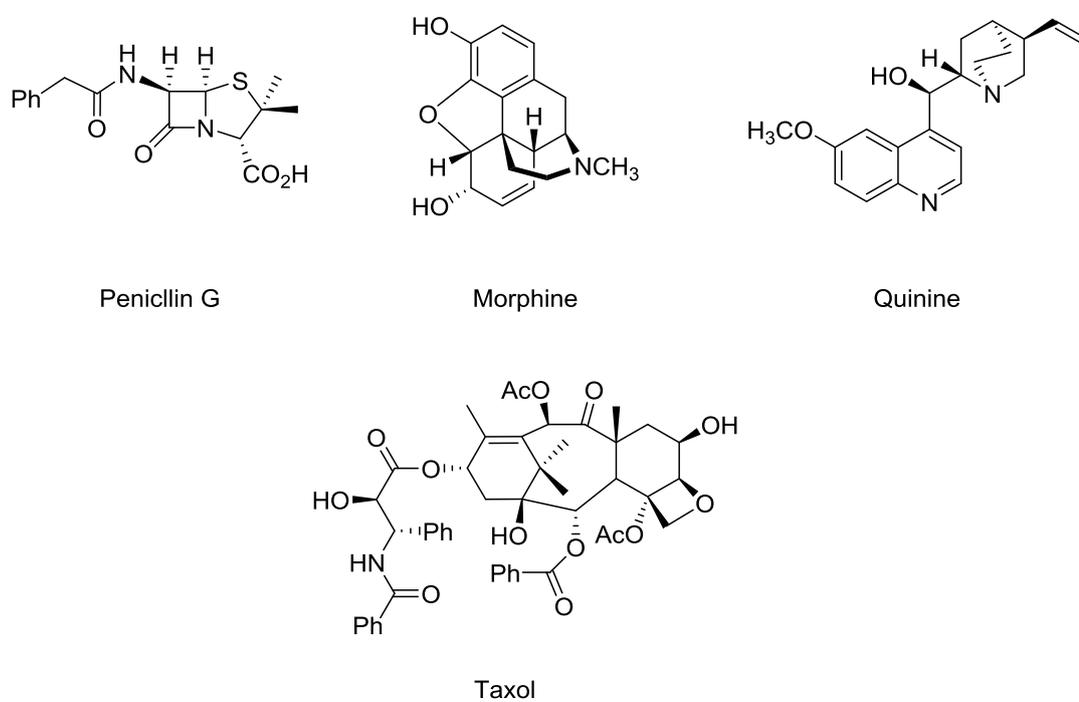
Taxol

Figure 1.2: Structures of the natural products of penicillin, morphine, quinine and taxol.

The aforementioned compounds are structurally complex making their chemical synthesis unfeasible due to the large number of synthetic steps required which would incur great economic costs. Fortunately, many of these compounds are naturally synthesised in enzyme-catalysed biosynthetic pathways, suggesting that isolating these responsible enzymes would provide *in vitro* routes for the production of natural products; this is of significant interest in the biotechnology industry. An example of such a process is the industrial production of semi-synthetic penicllins using the naturally occurring enzyme penicillin acylase (Figure 1.3).[2]
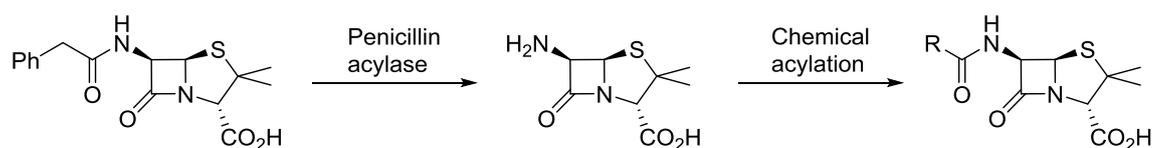
Figure 1.3: Industrial production of a semi-synthetic penicillin using a penicillin ayclase enzyme.

Biocatalysis is increasingly becoming the method of choice for the production of chiral molecules in the chemical and pharmaceutical industry. This is due to the fact that enantiomerically pure amino acids, amines, alcohols and epoxides are important classes of intermediates for the synthesis of many agrochemical and pharmaceutical products. Such intermediates are difficult to produce by chemical means, making enzymatic routes highly attractive alternatives.[14]

1.2 Enzymatic Oxygenation

The enzymatic aerobic oxidation of inactivated hydrocarbons is a highly desirable reaction in both academia and industry.[15] The main challenges faced when employing molecular oxygen for selective functionalisation to produce organic building blocks are: a) the activation of the thermodynamically strong and kinetically inert hydrocarbon bond, b) the activation of the O_2 molecule, c) the control of chemo-, stereo- and regioselectivity d) under environmentally safe and sustainable conditions.[16-18]

1.2.1 The Activation of Molecular Oxygen

The activation of molecular oxygen for oxidation reactions is extremely challenging due to inactivated O_2 being in a ground triplet state which is kinetically hindered to undergo formation of highly reactive oxygen radicals, hydroxyl radicals, hydroperoxides, or peroxides.[15]

In the ground triplet state, molecular oxygen has two unpaired electrons with parallel spins occupying two degenerate orbitals, making the reaction between O_2 and carbon in organic molecules spin forbidden. For oxidations to occur, this spin restriction must first be overcome by activating molecular oxygen in one of two ways. The first method involves the excitation of the ground triplet state ($^3O_2$) to a singlet state ($^1O_2$) in which the two electrons have opposite spins, with this diamagnetic form of oxygen being more reactive with organic molecules due to the increased presence of paired electrons. The second mode of activation is the step-wise reduction of molecular oxygen to form superoxide (O_2^-) followed by hydrogen peroxide (H_2O_2) which is reduced to a hydroxyl radical (OH) and ultimately reduced to water.[19]

1.2.2 Oxidation Chemistry

The challenges associated with employing molecular oxygen as an oxidant have resulted in the use of oxidising agents based on toxic metal ions such as osmium and chromium in organic oxygenation reactions. Many chemical oxygenation reactions also result in the generation of organo-halogens as intermediates, hence producing undesirable halogenated waste. Such reactions also often lack chemoselectivity resulting in the occurrence of undesirable side reactions.[20] Additionally, it is exceptionally difficult to achieve the regio- and stereoselective oxygenation of organic molecules by standard organic synthesis approaches, with these methods also having the caveat of often not being amenable to large scale synthesis.[13, 21-23] Examples of chemical approaches include the Sharpless asymmetric dihydroxylation and Jacobsen epoxidation (Figure 1.4).
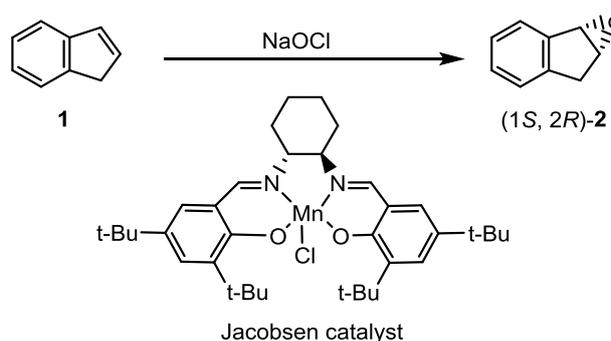
Figure 1.4: Examples of asymmetric oxidation reactions in organic chemistry. A: Sharpless asymmetric dihydroxylation for the formation of chiral diols. B: Jacobsen epoxidation of indene 1 to indene oxide 2.

1.2.3 Biocatalytic Oxidation

Enzymes are able to easily overcome the caveats associated with organic synthesis methods, possessing the capacity to activate oxygen and function with high activity and selectivity. These enzymes are generally referred to as 'oxygenases' and catalyse the

introduction of oxygen atoms from molecular oxygen or hydrogen peroxide into non-activated C-H- and C=C bonds and to heteroatoms.[24] This unique combination of activity and selectivity is thought to stem from the incorporation of reactive oxygen transfer reagents such as organic peroxides or highly oxidised iron-oxo complexes in the enzyme active site. This well-defined and chiral environment plays the multi-faceted role of positioning the substrate in the vicinity of the oxygen atom that is transferred (accounting for selectivity) while also stabilising reaction transition states thereby often leading to dramatic rate accelerations.[24]

Oxygenases can be divided into two groups. Monooxygenases catalyse the insertion of a single oxygen atom from molecular oxygen while the other is reduced to form water at the expense of a donor. Dioxygenases on the other hand, introduce both atoms into the substrate.[25]

1.3 Flavin Dependent Oxygenases

Flavin dependent oxygenases are monooxygenase enzymes involved in a wide range of biochemical processes, with some playing key roles in the catabolism of natural and pollutant compounds, while others assist in the biosynthesis of vitamins, hormones and antibiotics, contributing to host defence strategies.[26] These enzymes are renowned for their ability to catalyse a plethora of oxygenation reactions with high regio- and stereoselectivity (Figure 1.5).[27]
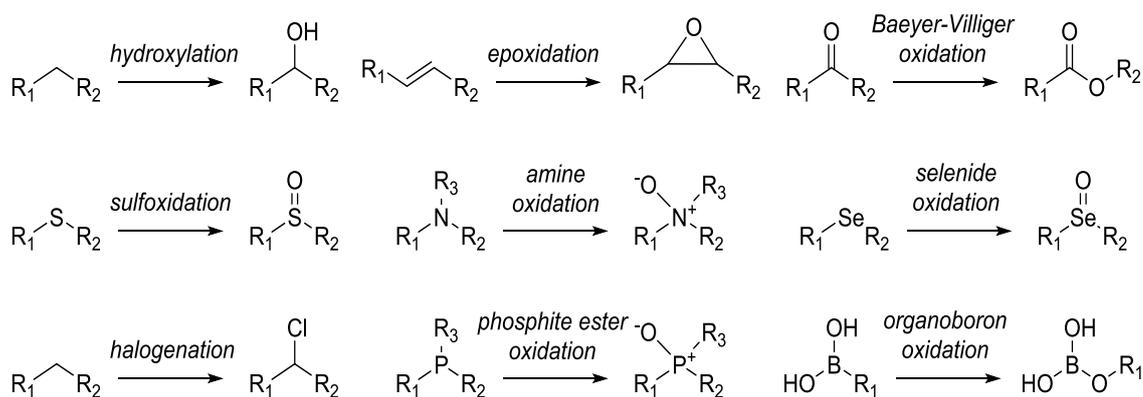
Figure 1.5: Reactions catalysed by flavin dependent monooxygenases.

As their name implies, flavin dependent monooxygenases require a purely organic species, either flavin adenine dinucleotide (FAD) or flavin mononucleotide (FMN) as the cofacot, with this species determined by the class of enzyme, reaction catalysed and mechanism of action.[28] Flavin dependent monooxygenases constitute the largest family of flavoenzymes, with at least 130 described to date.[29] van Berkel and colleagues used the structural and functional properties of flavin dependent oxygenases to propose that the enzymes be divided into six groups.[26-27] Group A and B monooxygenases (EC 1.14.13) consist of single-component enzymes that use FMN or FAD as a cofactor and are dependent on nicotinamide adenine dinucleotide (NAD) or nicotinamide adenine dinucleotide phosphate (NADPH) as external electron. Single-component monooxygenases are typified by Baeyer-Villiger monooxygenases. Group C-F monooxygenases, on the other hand, are two-component systems consisting of two enzymes. A reductase partner enzyme supplies the monooxygenase with reduced FAD or FMN, which is used directly as a cosubstrate.[30] Example of these enzymes include styrene monooxygenases and aromatic hydroxylases. A

third sub-class known as internal flavoprotein monooxygenases, which reduce the flavin cofactor through substrate oxidation, are currently rising in prominence and being placed into two newly defined groups, G and H, which has resulted from an increasing amount of information regarding their structure and function.[26]

1.3.1 Baeyer-Villiger Monooxygenases

The Baeyer-Villiger reaction describes the peracid-mediated oxidation of a carbonyl compound to the corresponding lactone or ester and was first reported by Adolf von Baeyer and Victor Villiger in 1899.[31] This reaction has wide-ranging applications in organic chemistry from the synthesis of antibiotics, pheromones and steroids to the synthesis of monomers for polymerisations.[32]
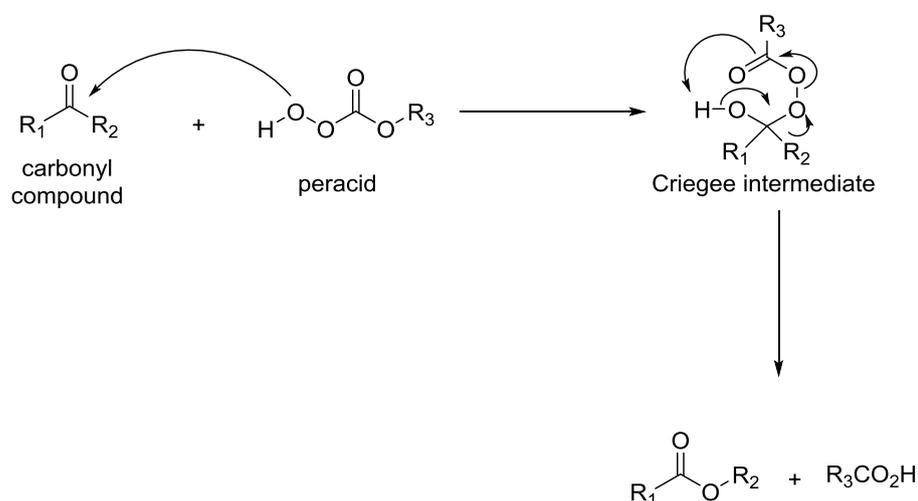
Figure 1.6: Proposed mechanism for the Baeyer-Villiger oxidation via a Criegee intermediate.

Figure 1.6 summarises the proposed two-step mechanism of the chemical Baeyer-Villiger oxidation reaction. The first step involves the nucleophilic attack or addition of the peracid to the carbonyl compound resulting in the formation of a tetrahedral Criegee intermediate. This is followed by the concerted migration of one of the adjacent carbons to oxygen resulting in the release of the carboxylate anion or carboxylic acid. The second step of this mechanism is often rate-limiting, with the leaving group playing a large role. If the migrating carbon is chiral, the stereoselectivity of the compound is retained, making the reaction highly advantageous. This chemical reaction does, however, have disadvantages which include poor selectivity and the requirement of harsh reaction conditions for catalysis making biocatalytic alternatives highly desirable.[33]

Baeyer-Villiger monooxygenases (BVMOs) catalyse the aforementioned reaction, doing so with the high selectivity typical of biocatalysts.[34] There is a striking distribution of BVMOs among organisms, with bacteria having, on average, one BVMO per genome.[35] In general, BVMOs are external monooxygenases containing a non-covalently bound FAD or FMN and require the reducing power of NADPH or NADH for the oxidoreduction of flavin.[36] Using the required cofactors as a general basis, BVMOs are commonly divided into two groups. Type I BVMOs are FAD and NADPH dependent while type II BVMOs use FMN and NADH. Of the two, type I are the more commonly occurring and well-studied with over 50 protein sequences currently available for recombinant expression.[37]

The most studied BVMO is cyclohexanone monooxygenase from *Acinetobacter* sp. (CHMO$_{Acineto}$), which now serves as a model for these enzymes.[37-38] CHMO is a type I BVMO originally found to selectively oxidise cyclohexanone but now known to have both a very broad substrate and reaction scope. The general reaction mechanism of CHMO was proposed using structural, kinetic and spectroscopic data (Figure 1.7).[39-42]
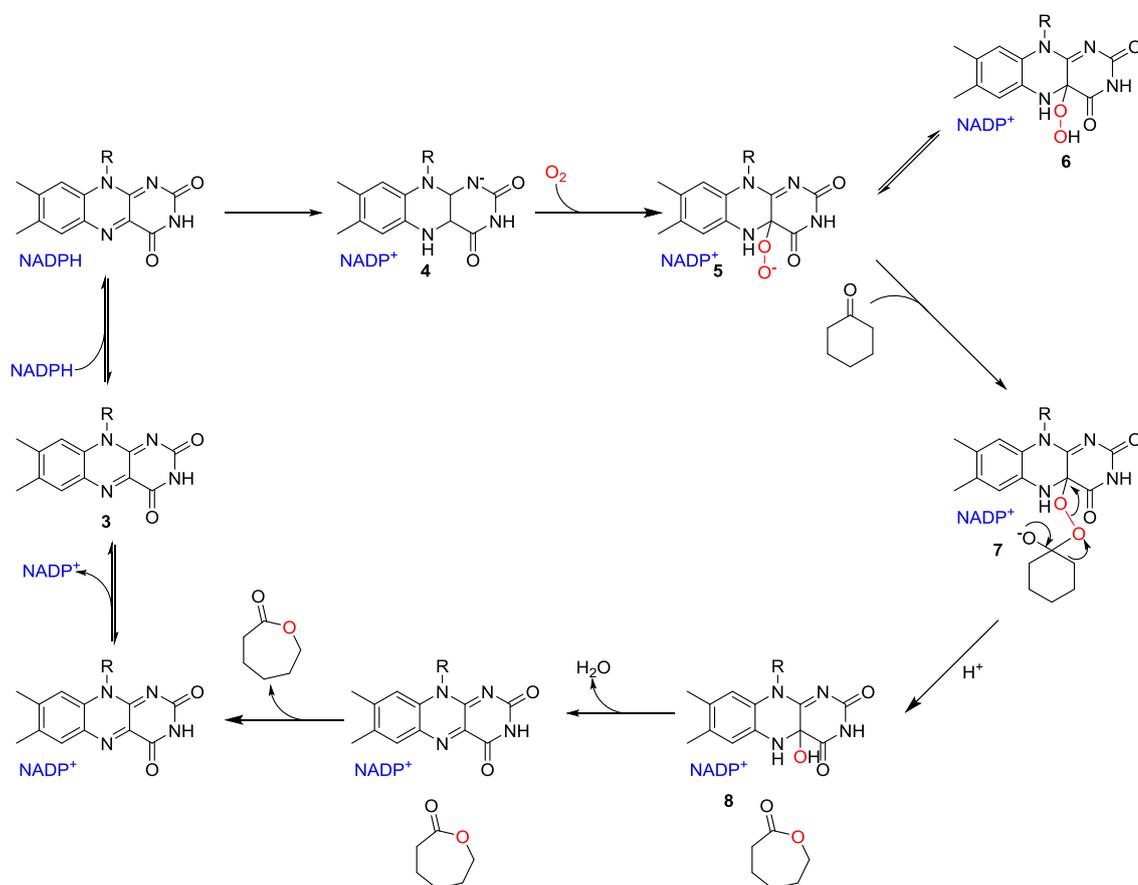
Figure 1.7: Proposed mechanism for the oxidation of cyclohexanone with CHMO from *Acinetobacter*. Figure adapted from Leisch *et al.* (2011).

9

The biocatalytic mechanism is analogous to that of the chemical Baeyer-Villiger oxidation. In BVMOs, the covalently bound flavin cofactor **3** is reduced by NADPH. The reduced flavin **4** reacts with molecular oxygen to form a C4a-peroxyflavin intermediate **5**. This intermediate plays the same role as the peracid in the chemical reaction, behaving as a nucleophile in the nucleophilic attack of the carbonyl carbon in the ketone substrate. This nucleophilic attack results in the formation of a Criegee intermediate **7** that next rearranges to form the C4a-hydroxyflavin **8** and ε-caprolactone. A molecule of water is spontaneously eliminated from the hydroxyflavin to regenerate the oxidised flavin. NADPH, the first species to bind the active site in the catalytic cycle, remains bound in the $NADP^+$ form until the final step of the reaction.[31]

The last two decades have seen increased applications of BVMOs. These include the use of BVMOs in single enzyme transformations to produce chiral lactones and in cascade reactions for the production of highly complex synthetic materials.[43-45] This is due to the high selectivity and green chemistry potential of BVMOs. There are, however, significant limitations of BVMOs that need to be overcome in order to enable industrial applications. The requirement of expensive cofactors, poor enzyme stability, substrate and product inhibition together with low product concentrations, inefficient oxygen transfer and poor solvent tolerance are among the most significant obstacles.[37]

The requirement of expensive cofactors and poor stability can be overcome by using whole cell systems for biotransformations. This exploits the host machinery to regenerate the cofactor while also avoiding the need to purify the enzyme. This approach does still suffer from substrate and product inhibition as well as poor oxygen transfer rates. An alternative approach is to use inexpensive coenzymes such as formate dehydrogenase (FDH) and glucose-6-phosphate dehydrogenase (G6PDH).[28] Difficulties associated with whole cells and solvent tolerance can be tackled by immobilising BVMOs on a solid support, with this hypothetically aiding in maintaining the protein fold and ensuring increased stability in organic solvents.[46] It must however, be noted that in the case of some BVMOs, a loss of activity has been observed despite immobilisation.[47-48] Substrate and product inhibition could be tackled by using two-phase system strategies with water immiscible organic solvents in which the organic phase acts as substrate reservoir and as an extraction solvent for product isolation.[37] Oxygen transfer is a difficult parameter to control as the rate of

transfer in whole cell reactions as it can be influenced by the geometry of reaction flasks; oxygen levels also have to be carefully maintained as levels at both extremes can have negative effects on enzyme performance and cell health.[37, 49-50]

BVMOs are highly versatile enzymes with great promise in industrial and biotechnological promise, however, a significant number of limitations need to be conquered before their potential can be fully realised.

1.3.2 Two-Component Oxygenases

Two-component flavin dependent oxygenases catalyse the oxidation of polycyclic and aromatic compounds for use as carbon sources, the biosynthesis of antibiotics and long-chain alkanes, bioluminescence, and the desulfurization of sulfonated compounds.[51] Unlike the previously discussed one-component oxygenases these enzymes require a separate flavin reductase and oxygenase for activity. This is due to two-component oxygenases using FAD or FMN directly as a cosubstrate and thus need a separate reductase enzyme to supply reduced flavin.[30] Examples of enzymes of interest in the two-component oxygenases include styrene monooxygenases and aromatic hydroxylases.

Styrene Monooxygenases

Styrene is an environmentally harmful hydrocarbon which is readily biodegradable by many classes of microorganisms, with the most common pathway initiated by a styrene monooxygenase-catalysed epoxidation of the vinyl side chain.[52] The styrene monooxygenases involved in this reaction have received widespread research interest due to their environmental and biotechnological applications. Most styrene monooxygenases have been isolated from the genus *Pseudomonas* and all catalyse the conversion of styrene in a highly enantioselective manner to (*S*)-styrene oxide (Figure 1.8), which is a useful precursor for several chiral synthons and pharmaceuticals.[53-56] Furthermore, the relaxed substrate specificity of styrene monooxygenases allows the enantioselective conversion of substituted styrene derivatives and structurally analogous compounds such as indene, dihydronaphthalene and phenylalkylsulfides, thus increasing their potential biocatalytic applications.[57-59]
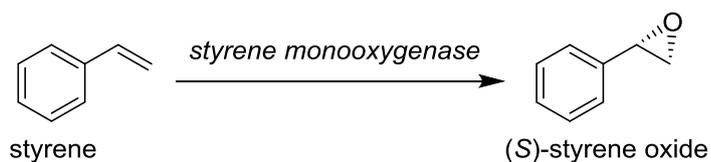
Figure 1.8: Enantioselective epoxidation of styrene to (S)-styrene oxide by styrene monooxygenase.

The standard styrene monooxygenases from *Pseudomonas* consist of two enzymatically active protein components that are encoded by genes, *styA and styB*, that are often located next to each other in the chromosome. In the proposed reaction mechanism, the flavin reductase subunit, StyB, reduces FAD at the cost of NADH. The reduced flavin, $FADH_2$, is then utilised by the monooxygenase subunit, StyA, to activate molecular oxygen to H_2O_2 for attack of styrene (Figure 1.9).
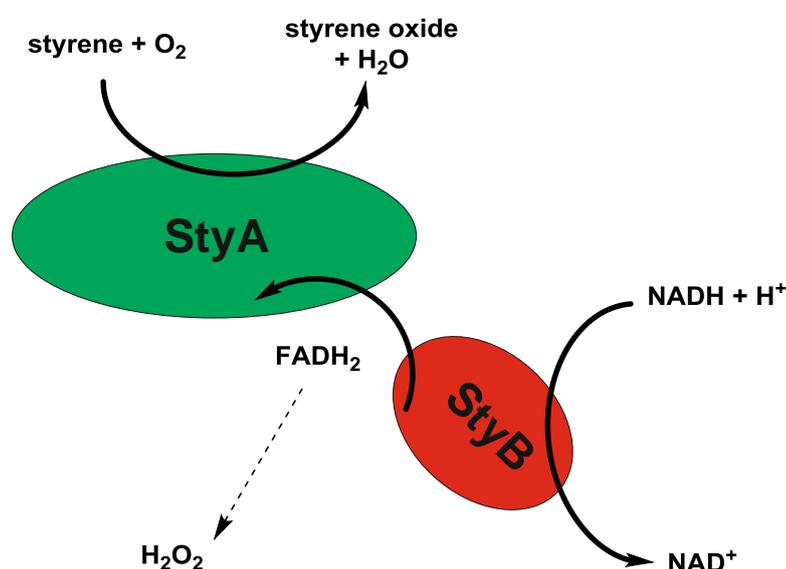
Figure 1.9: The two enzyme components, StyA and StyB, of the standard styrene monooxygenase and their proposed mechanism of action. Figure adapted from Tischler *et al.* (2010).

A limitation of the use of styrene monooxygenases in industry is ensuring the efficient inter-protein transfer of $FADH_2$ as failing to do so will result in $FADH_2$ auto-oxidation and oxidative stress on the system.[60-61] Preventing such stress is highly significant for the biotechnological applicability of multi-component oxygenases since doing so can substantially improve the long-term stability of both whole-cell and cell-free transformation processes.[52] An evolutionary strategy to overcome this limitation involves the generation of self-sufficient fusion proteins.[62-64] Cytochrome P450 BM3 from *Bacillus megaterium* has served as a well-known prototype for such fusion due to its high oxygenation capability.[63] Tischler and colleagues reported the first self-sufficient styrene monooxygenase from

Rhodococcus opacus (StyA2B), containing a monooxygenases and reductase in the same polypeptide chain.[58] StyA2B was found to have similar substrate specificity and enantioselectivity levels compared to other two-component equivalent enzymes, but displayed a low specific activity in its oxygenase domain.

Aromatic Hydroxylases

Aromatic hydroxylases are two-component oxygenases that catalyse the hydroxylation of aromatic compounds. Examples of these enzymes include *p*-hydroxyphenylacetate-3-hydroxylase from *Pseudomonas putida* and pyrrole-2-carboxylate monooxygenase from *Rhodococcus* sp. (Figure 1.10).[65-66] It was found that both components of the enzyme were required for hydroxylations in the two enzymes.
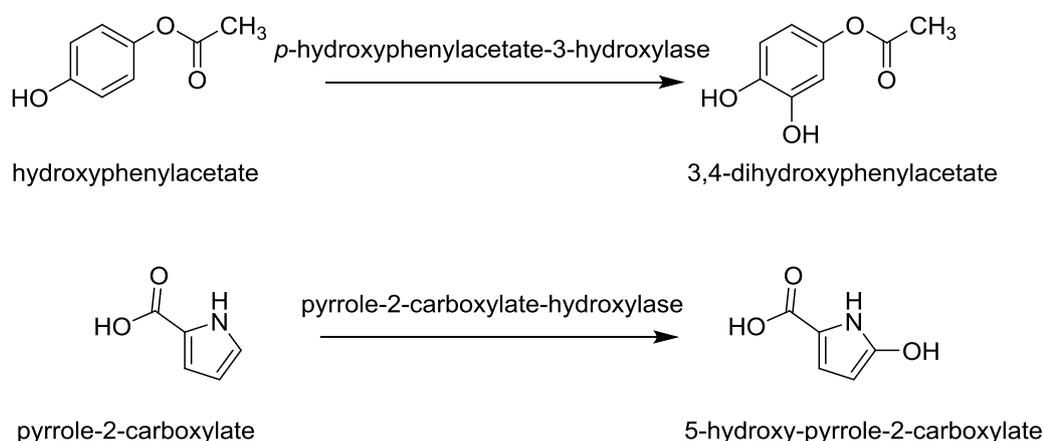
Figure 1.10: Examples of aromatic hydroxylations; Top: hydroxylation of hydroxyphenylacetate by *p*-hydroxyphenylacetate-3-hydroxylase. Bottom: hydroxylation of pyrrole-2-carboxylate by pyrrole-2-carboxylate-hydroxylase.

Aromatic hydroxylases, as with the previously discussed styrene monooxygenases, have applications in medicinal chemistry and pollutant degradation. Two-component monooxygenases have been particularly studied in the biosynthesis of antitumor agents C-1027 and rebeccamycin.[30, 67] In the area of pollutant biodegradation, 2-naphthoate monooxygenase has potential applications in the degradation of carcinogenic, mutagenic and toxic polycyclic aromatic hydrocarbons.[68] A two-component monooxygenase from *Bacillus sphaericus* JS905 was likewise found to catalyse the initial two steps in the biodegradation of *p*-nitrophenol (PNP) via a hydroxylation reaction followed by elimination of the nitro group (Figure 1.11).[69]
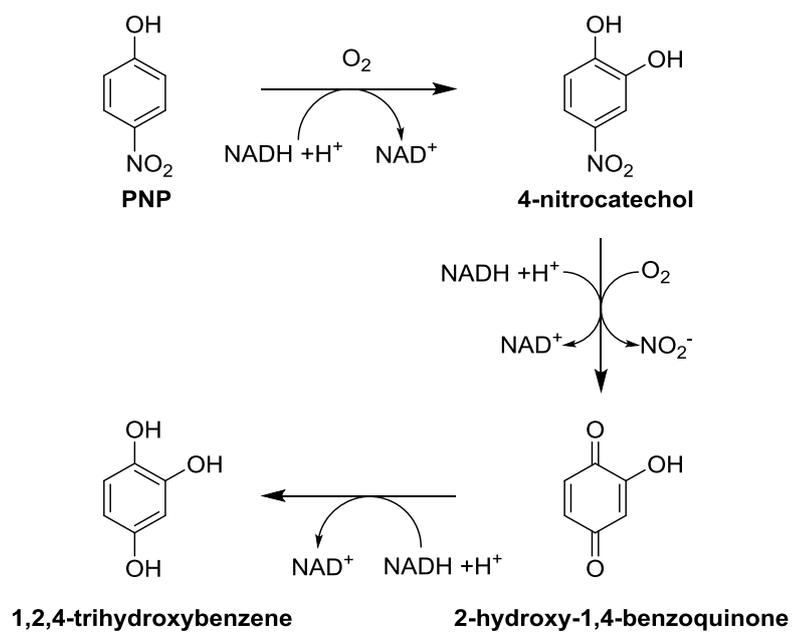
Figure 1.11: Two step biodegradation of PNP facilitated by monooxygenase from *Bacillus sphaericus* JS905.

1.4 Hydroxylation Reactions

Hydroxylations are among the most widespread enzyme reactions in Nature, playing a key role in the oxidative metabolism of both beneficial pharmaceuticals and destructive environmental pollutants. The reaction has been utilised in industry for many years in the production of fine chemicals and for bioremediation.[70] Research into biocatalytic hydroxylation reactions have gained traction over the last few years with numerous different enzyme types currently being studied.

1.4.1 Cytochrome P450 Monooxygenases

Cytochromes P450 (CYPs or P450s) are haem-containing monooxygenases present in all domains of life that catalyse the oxyfunctionalisation of a diverse range of chemical compounds including alkanes, antibiotics, xenobiotics, steroids, terpenes, fatty acids and others in the presence of molecular oxygen.[71-72] P450s constitute a distinct family of haemoproteins (collectively known as the CYP family) and are capable of catalysing the insertion of a single oxygen atom from molecular oxygen while reducing the other to water.[73] These enzymes are of significant interest due to their ability to activate molecular oxygen and react with high regio-, stereo- and enantioselectivity.[23]

P450s are named as such due to their character as haemoproteins and the unique spectral property of their haem-binding electron transfer protein (cytochrome). When the cytochrome is complexed with carbon monoxide, a characteristic shift in absorption maximum to 450 nm is observed (hence pigment 450); this ability is frequently used to estimate P450 content.[74-75] Such a feature is the result of a phylogenetically conserved cysteine thiolate which is bonded directly to the catalytic iron (Figure 1.12). This bond possesses an electron-donating character which is essential for the P450's catalytic activity. The ligand (usually water) located *trans* to the axial cysteine-thiolate is readily displaced by molecular oxygen at the start of the catalytic cycle.[23]
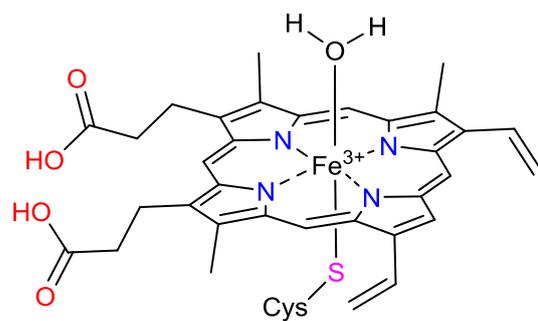
Figure 1.12: The catalytic haem present in cytochrome P450s. The catalytic iron, Fe³⁺, is haem bound and possesses both an axial cysteine thiolate and water molecule. This water molecule is displaced by molecular oxygen at the start of the catalytic cycle.
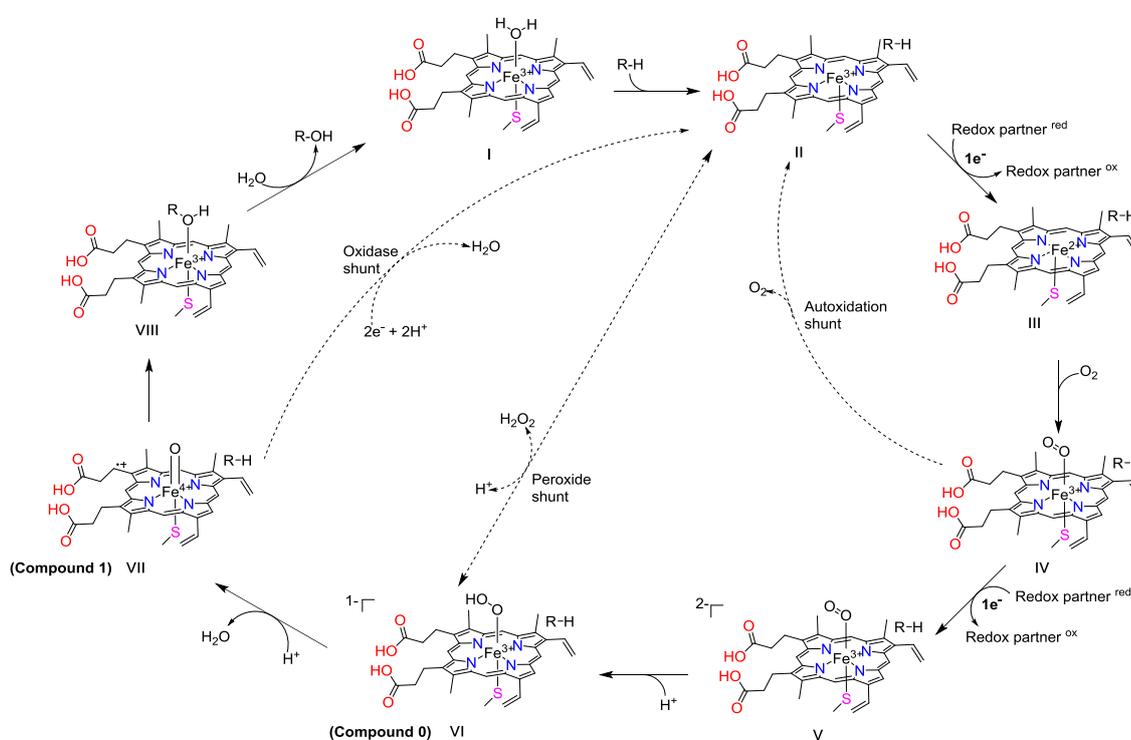
Figure 1.13: General catalytic cycle of cytochromes P450. Three "shunt" pathways responsible for the non-productive breakdown of iron-oxygen species throughout the cycle are labelled. Figure adapted from Roper and Grogan (2016).[76]

Catalytic Cycle

In order for P450 catalysis to proceed, the vast majority of these enzymes require the delivery of two electrons from pyridine nucleotide coenzymes (NADPH or NADH), with these electrons transferred to the haem *via* one or more redox partner proteins. Figure 1.13 summarises the intermediate steps of the general P450 catalytic cycle. In the first step, substrate binding results in the displacement of a water ligand, shifting the spin-state

equilibrium of the ferric haem from low-spin ($S = 1/2$) to high-spin ($S = 5/2$). This increases the positive potential of the iron and allows electron transfer from the redox partner to reduce the haem iron to the ferrous (Fe^{2+}) state. Ferrous haem then binds molecular oxygen to form the ferrous-oxy intermediate **IV** (isoelectric with the ferric-hydroperoxo form). A second electron is delivered from the redox partner to reduce the haem iron to the ferric peroxy state **V**. This species is protonated to the hydroxperoxo form (Compound 0), which is followed by a further protonation resulting in scission of the bound dioxygen, the production of water and formation of a ferryl-oxo haem species (Compound 1) with a porphyrin π radical. Compound 1 is likely the catalytically relevant oxidant substrate in most P450 reactions, attacking the nearby the substrate to result in its hydroxylation. Finally, the hydroxylated product dissociates, allowing water to rebind the ferric iron and complete the cycle.[23]

Figure 1.13 also highlights three possible routes for the collapse of the oxo intermediates. The ferrous-oxy intermediate **IV** can decay to reform the ferric P450 *via* the autoxidation shunt with production of superoxide as a side-product. Compound 0 can undergo collapse via the peroxide shunt in which peroxide is also produced. The peroxide shunt can be productively driven in the reverse direction by mixing hydrogen peroxide (or other organic peroxides) with ferric, substrate-bound P450; however, this is in most cases a very inefficient process. Compound 1 undergoes collapse via an oxidase shunt pathway coupled with the production of water.[76]

Redox Systems

As mentioned above, P450s require the delivery of two electrons to the haem for catalysis, with these electrons transferred from a reduced cofactor by redox partner proteins. This is performed using a diverse range of strategies in nature, most of which rely on electron transfer proteins that function in conjunction with cofactors. There is substantial biological diversity in P450 systems and redox part with at least 10 documented classes in nature, the four most commonly encountered in biocatalysis research are shown in Figure 1.14.[76]

(a) Class I (e.g P450$_{cam}$)

Fdx

FdR

P450

NADH + H$^+$ → NAD$^+$

RH + O$_2$ → R-OH + H$_2$O

(b) Class II (e.g CYP3A4)

CPR

P450

ER Membrane

NADPH + H$^+$ → NADP$^+$

RH + O$_2$ → R-OH + H$_2$O

(c) Class VII (e.g P450$_{Rhf}$)

FeS

FMN

P450

NADPH + H$^+$ → NADP$^+$

RH + O$_2$ → R-OH + H$_2$O

(d) Class VIII (e.g P450$_{BM3}$)

CPR

P450

NADPH + H$^+$ → NADP$^+$

RH + O$_2$ → R-OH + H$_2$O
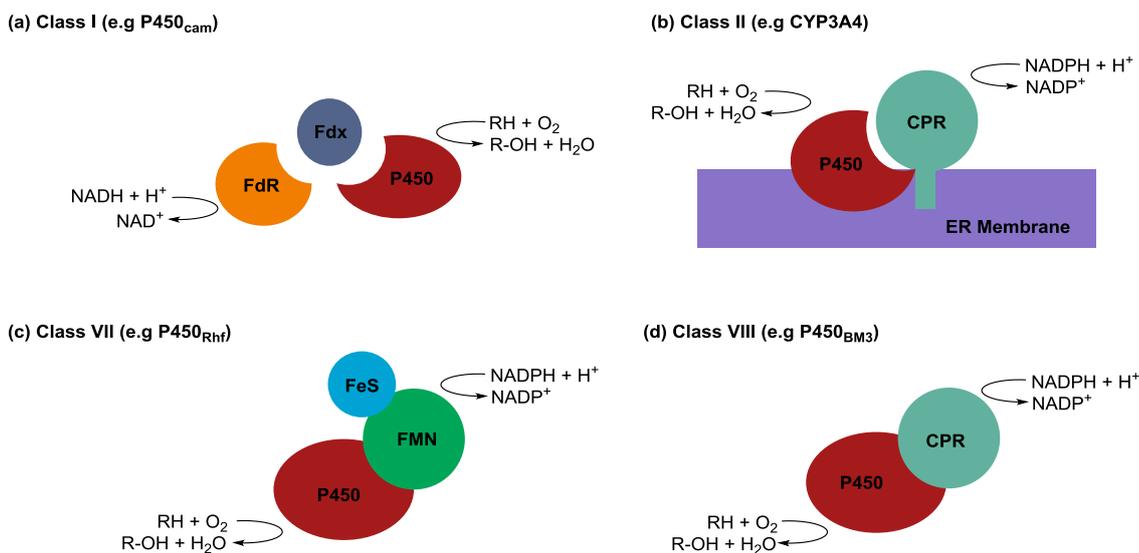
Figure 1.14: Four classes of P450 redox systems. (a) Class I bacterial system; (b) Class II observed in plant and human enzymes; (c) Class VII typified by P450$_{Rhf}$; (d) Class VIII system observed in P450$_{BM3}$. FdR = Flavin and NADPH-dependent ferrodoxin reductase; Fdx = ferredoxin; CPR = FMN- plus FAD-dependent reductase; FeS = C-terminal iron-sulfur protein; FMN = Flavin mononucleotide domain

Most bacterial P450 systems are grouped into Class I, in which electrons are abstracted from NAD(P)H by a flavin and NADPH-dependent ferrodoxin reductase (FdR). These electrons are transferred to an FAD coenzyme which reduces the next partner in the chain, ferredoxin (Fdx): a Fe-S cluster containing enzyme. The reduced Fdx then sequentially transfers two electrons to the P450 haem domain. This class of P450s is typified by CYP101A1 (P450$_{cam}$) which will be discussed in the following section.[74, 76]

Class II systems are found in eukaryotic species such as plants, humans and fungi, and are composed of a two-protein chain in which the first enzyme is an FMN- plus FAD-dependent reductase (CPR) which accepts electrons from NADPH and transfers them to the P450 haem (the second enzyme in the chain). In eukaryotes, both the reductase and P450 are anchored into the membrane of the endoplasmic reticulum, which has made these enzymes extremely difficult to study and apply in synthesis. Examples of class II enzymes include human P450s such as CYP3A4 and all of the currently known plant P450s.[76]

The requirement of additional proteins presents extra challenges for the cloning, expression and purification of P450s for use in biocatalysis. The multi-enzyme systems also present many of the limitations discussed for BVMOs previously (these limitations will be discussed in more detail later). There exist in nature P450 fusion systems in which the

electron transfer proteins and haem domain are encoded by a single gene.[76] Such P450s are typified by those in Classes VII and VIII (Figure 1.14c and d). Class VII P450s such as the prokaryotic P450$_{Rhf}$ from *Rhodococcus* sp. possess a fusion of three enzyme components for optimised function that generally consist of a C-terminal iron sulphur protein (FeS) and an FMN domain fused to an N-terminal haem domain (Figure 1.14c).[76] Class VIII P450s are composed of a CPR domain fused to the P450 haem itself (Figure 1.14d), with this arrangement having a pronounced effect on the activity of these enzymes. P450$_{BM3}$ from *Bacillus megaterium* is the most studied of the class VIII enzymes and will be discussed in its own section.

P450$_{cam}$ and P450$_{BM3}$

The field of P450 enzymology is enormous and highly complex, with research focus into multiple isoforms inevitably leading to the discovery of substantial variations in structural, catalytic and mechanistic features between members of this family.[23] As such, the search for common features binding P450s together has made the detailed characterisation of model P450 systems key to the understanding of basic principles relating P450 structure and function. The greatest amount of information pertaining to general P450 characteristics has been taken from bacterial models due to the experimentally amenable nature of soluble prokaryotic P450s. The two most significant model P450s are the *Pseudomonas putida* camphor hydroxylase P450$_{cam}$ (CYP101A1) and the *Bacillus megaterium* fatty acid hydroxylase P450$_{BM3}$ (CYP102sA1).[77-78]

Cytochrome P450$_{cam}$ is a landmark P450 serving as the first bacterial P450 system to be studied in detail from the 1960s onwards and the first P450 to have its atomic structure resolved.[79] P450$_{cam}$ is representative of the class I redox system and catalyses the *5-exo* hydroxylation of D-camphor; the initial step in the catabolism of this compound (Figure 1.15).[23]

(1*R*)-(+)-camphor → 5-*exo*-hydroxy camphor

Figure 1.15: The *5-exo*–hydroxylation of (1*R*)-(+)-camphor to 5-*exo*-hydroxy camphor by P450$_{cam}$.

P450$_{BM3}$ is a class VII P450 that catalyses the oxygenation of long chain fatty saturated and unsaturated fatty acids (with lengths of approximately C_{12} to C_{20}), performing this hydroxylation near the ω terminus (Figure 1.16).[80] The physiological role of P450$_{BM3}$ is uncertain though it has been hypothesised to play a role in the detoxification of polyunsaturated fatty acids such as linoleic acid.[81] P450$_{BM3}$ is a natural fusion protein, with its haem and reductase bound together, and was the first prokaryotic class II system to be characterised.[82] The haem domain of P450$_{BM3}$ was the second P450 enzyme to be structurally resolved by X-ray crystallography.[83] P450$_{BM3}$ possesses a eukaryotic-like redox system and a convenient fusion arrangement, which, together with the fact that it has a haem-domain strongly resembling those in eukaryotic fatty acid oxygenases, has made this system a relevant model for the mammalian class II systems.[23]
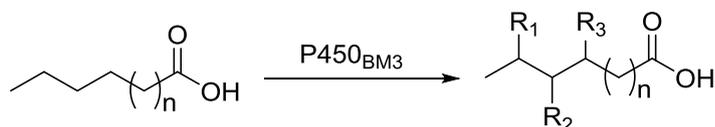
Figure 1.16: General hydroxylation of saturated, linear fatty acids at the ω-1 (R_1), ω-2 (R_2) and ω-3 (R_3) positions by P450$_{BM3}$.

P450 Structural Insights

The structures of P450cam and P450BM3 possess structural features common to P450s and can be used as general models. These P450s have a general shape akin to a triangular prism and two major structural domains, termed α and β, which surround the haem domain (Figure 1.17A).[77, 83] The structural fold of these two model P450s is generally conserved across the P450 superfamily.[84]
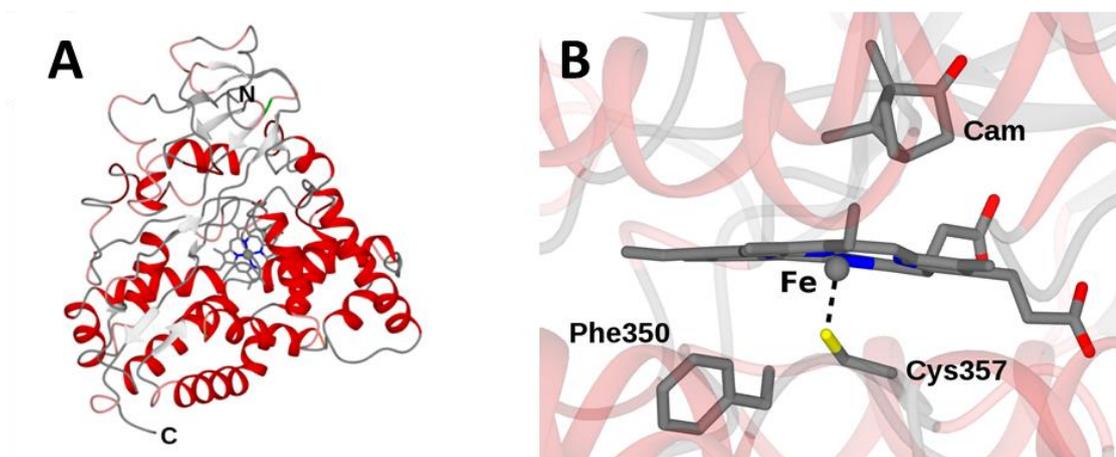
Figure 1.17: (A) The overall tertiary structure of P450$_{cam}$, such structural features are generally conserved among members of the P450 family. (B) The active site of P450$_{cam}$ with an emphasis on the conserved Cys357 and Phe350 residues integral for function. Phe350 plays a role in catalysis by interacting with Fe-Cys357 bond (dashed line) and tuning the reducing potential of the haem iron (Fe). The substrate camphor (Cam) is presented above the haem iron.

Though the α helical and β sheet components and their three-dimensional arrangement are conserved among P450s, there are significant differences in their organisation, resulting in differences between P450 active sites and other key regions.[84] Members within the P450 superfamily generally have ≥ 40% amino acid sequence identity, however, only a few amino acids are completely conserved across these enzymes. An absolutely conserved residue is the cysteine ligand bound to the haem iron as it is critical to function (Figure 1.17B).

A highly conserved phenylalanine residue located seven amino acids (Phe393 in P450$_{BM3}$) prior to the previously mentioned cysteine, interacts with the Fe-Cysteine bond and plays a key role in tuning the reductive potential of the haem iron and its reactivity with molecular oxygen.[85] The importance of this residue in P450$_{BM3}$ was determined by investigating Phe393A/H mutants, which showed a more positive haem iron potential and simultaneous acceleration of electron transfer from the redox partner, resulting in the stabilisation of the ferrous-oxy intermediate and a significant decrease in the rate of fatty acid hydroxylation.[86] As a result, the conserved phenylalanine likely plays a significant role in controlling redox potential by facilitating both efficient electron transfer to the haem iron from the redox partner, and the P450-facilitated reduction of molecular oxygen.[85-86]

Research into other conserved residues and their role in P450 enzymes is ongoing. Identifying key residues could aid in directing evolution of these enzymes for improved and novel functions which will be discussed later.

Applications of Cytochromes P450

P450s possess the ability to catalyse the regio- and stereospecific oxidation of a range of substrates (Figure 1.18) which is reflective of their key biological functions.[73] These enzymes are also highly versatile with the capacity to selectively catalyse a range of transformations, including hydroxylations, epoxidations and dehydrogenations, making them interesting candidates for biocatalysis.[74, 87]
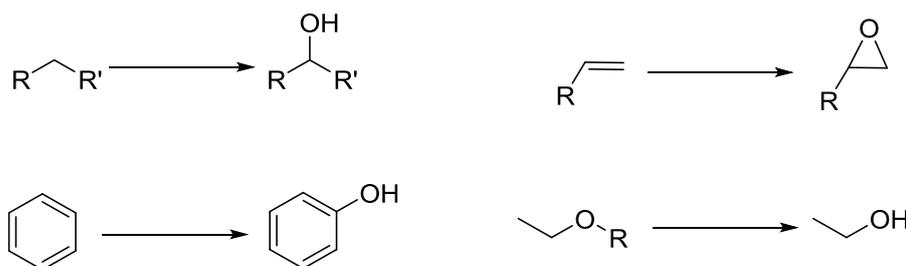
Figure 1.18: Selected examples of P450 oxidation reactions.

P450s play a large role in the biosynthesis of various steroids, vitamins, lipids and various other natural products.[88-92] In eukaryotic systems, P450s are mainly responsible for biodegradative reactions.[73] 57 putative genes in the human genome have been recognised as P450 enzymes and roughly a quarter of these have been seen to play key roles in drug metabolism.[93-97] The large P450 superfamily catalyses the oxidations of xenobiotics making them of significant interest within the pharmaceutical and fine chemical industry.[73]

Human P450s are often involved in the initial oxidation steps of metabolism, producing more soluble metabolites aiding their excretion and modulating the toxicity of these compounds.[73] The characterisation of these processes is indispensable in drug metabolism and pharmacokinetics (DMPK), allowing for the prediction of the methods in which a drug will be metabolised in the human body. DMPK is often one of the primary reasons for the failure of drug candidates.[98] The production of drug metabolites is a key priority in industry in order to assess toxicity risks, drug-drug interactions and to study metabolic pathways.[99] P450s can be used the generation of metabolites, making them essential tools in pharmacokinetics, with panels of P450s active towards common pharmaceuticals (e.g. diclofenac) showing an increased market presence.[73]

P450s can themselves be used in the synthetic routes for the production of pharmaceutical. An example of this is the process developed by Sankyo Pharma and Bristol-Myers Squibb

company produce pravastatin, a cholesterol reducing drug, *via* the microbial oxidation of compactin using a P450 from *Streptomyces* sp. (Figure 1.19).[100-101]
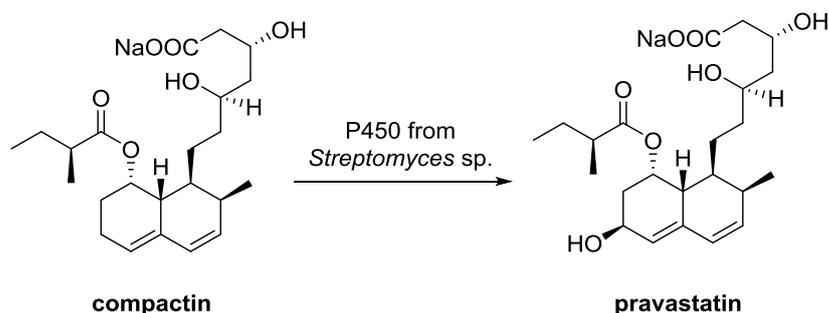
Figure 1.19: Biocatalytic synthesis of pravastatin from compactin using a P450 from *Streptomyces* sp.

The economic feasibility, high specificity and large substrate and reaction scope of P450s has made the use of these enzymes in numerous pharmaceutical processes widespread.[72] Figure 1.20 shows a select few of the many commercial and industrially relevant products that are can be produced using processes involving P450s.[72]
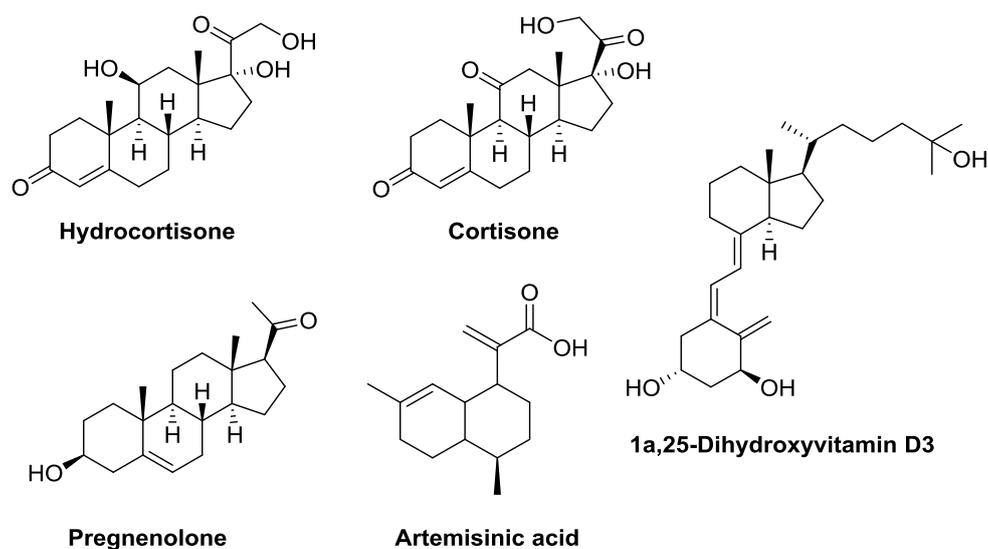
Figure 1.20: Selected industrially and commercial relevant products produced with the assistance of P450s.

In addition to the above pharmaceutical applications, P450s are also being increasingly used in bioremediation processes to remove major industrial pollutants present in the environment, including polycyclic aromatic hydrocarbons (PArHs), polychlorinated dibenzo-*p*-dioxins (PCDDs) and polychlorinated biphenyls (PCBs).[73] The use of P450s is increasing due to their capacity to perform hydroxylations, epoxidations and dealkylations on these pollutant compounds, which are key reactions in bioremediation chemistry.[102] An

example of a P450-mediated bioremediation process is the metabolism of the carcinogen benzo[*a*]pyrene using human P450 1B1.[103]

Limitations of P450s

P450s have a range of potential applications in the pharmaceutical and biotechnology sectors; however, their widespread use is hampered by a number of universally recognised technical bottlenecks.[73-74, 104-105] Firstly, P450 catalysts display productivities that are often at an insufficient level for industry scale synthesis with most displaying k_{cat} levels in the range of 1-300 min^{-1}.[104] Like many biocatalysts, P450s have the potential to be highly evolvable for improved or novel functions; however, many P450 structures do not present the substrates bound in the active site in a position favouring hydroxylation, presenting difficulties in using structures alone to rationalise mutagenesis attempts.[72, 106]

A third limitation of P450s is their need for electron transfer proteins, with specific types having different effects on the efficiency of electron flow which can greatly affect P450 performance. Another limitation is uncoupling between NAD(P)H oxidation and product formation, with it being observed that leakage of electrons may occur in the cycle which could lead to the formation of reactive oxygen species in the reaction media. In addition to this, upon consumption of four electrons, there is a risk that two water molecules can be produced without any substrate hydroxylation occurring, with the highly expensive NADPH cofactor still consumed in these cases. The requirement of an expensive NADPH cofactor is also a severe limitation and strategies to either replace its need or develop regeneration strategies are being actively pursued.[72]

The use of P450s in whole-cell reaction systems presents a further set of limitations. One problem can be low substrate solubility in water (may also be a problem during *in vitro* reactions) as most P450 substrates are hydrophobic, however, this can be overcome relatively easily by using solvent-resistant strains such as *B. subtilis and P. putida*, or by employing two-phase systems as with BVMOs.[107-109] A more challenging drawback is the uptake of substrates into the recombinant whole cells and product efflux. This can, in some cases, overcome by changing the expression host; however, engineering strategies may be required in more complex cases.[72] Finally, substrate and product toxicity can also have a serious effect on the biotechnological applications of P450s, with the risk of host-cell death

occurring when concentrations of each exceed a certain threshold value, though this is highly dependent on factors such as the recombinant system and the nature of products and reactants.[72, 109]

Engineering P450s

A substantial amount of research effort has been directed toward engineering P450s to improve function and overcome the limitations discussed above.[72]

The lack of substrate-bound P450 structures has not hindered efforts in site-directed mutagenesis and directed evolution. $P450_{BM3}$ has been successfully engineered to have improved activity towards a variety of substrates; an example of this is the nearly 9000-fold improvement of the catalytic efficiency for propane hydroxylation by a combination of site-directed and saturation mutagenesis.[110-111] $P450_{BM3}$ mutants capable of catalysing the hydroxylation of cyclododecane with an activity of up to 18 min^{-1} have also been constructed; this is a hydroxylation reaction that cannot be catalysed in the wild-type enzyme.[112]

An approach to optimise electron flow and overcome the issues associated with multi-component P450 systems has involved the engineering of artificial fusion proteins. This has been aided by the discovery of increasing number of P450-redox partner fusion proteins identified over the last decade. [63, 113-116] A model for this work was $P450_{BM3}$, a natural fusion enzyme, which displays the highest turnover numbers (for its natural substrate) of any P450 to date at 3000 min^{-1}.[74] A great amount of effort has been expended in generating artificial fusion proteins consisting of a P450 haem protein bound with all manner of redox partner proteins with a seemingly endless possibility of combinations possible.[72] Fusion proteins have even been constructed in cases in which bacterial P450s were already cloned but their autologous redox partners were still unknown; this is primarily due to the presence of many different candidates for electron transfer in the microbial cell.[117-118] In addition to this, there have been cases in which substrate conversion was observed in (previously poorly performing) bacterial P450s fused to the reductase protein of $P450_{BM3}$.[119-120]

The reliance on expensive nicotinamide cofactors, particularly NADPH, is a major barrier for the industrial use of P450s. Some engineering studies have focused on shifting P450 cofactor preference toward to more economical NADH.[121] Alternatively, the peroxide shunt has been utilised to eliminate the need for cofactors altogether, with the hypothesis that the addition of hydrogen peroxide or organic peroxide will shift the pathway in the reverse direction to generate compound 1 (Figure 1.13). Directed evolution has also been attempted to improve peroxide-driven catalysis, however, catalytic rates are often low and peroxide exposure often leads to the destruction of the active site haem (haem bleaching) and oxidative damage to the enzyme.[23, 122]

Cytochromes P450 are a widely studied enzyme superfamily with substantial industrial potential. There are, however, many caveats to their implementation (discussed above) and research into overcoming these is ongoing.

1.4.2 Peroxygenases

An alternative class of enzymes that circumvents the above limitations are peroxygenases.[123] This class of enzymes, like the P450s, contain a haem prosthetic group coordinated by a cysteine-ligand, which, in principle, grants access to the same vast range of reactions and products. Peroxygenases also require oxyferryl-haem (compound I) as the oxidising species, but, unlike P450s, generate this by directly utilising hydrogen peroxide *via* the peroxide shunt pathway rather than undergoing a series of electron transfer steps (Figure 1.13).[124]

Peroxygenases are members of a peroxidase superfamily known as fungal haem-thiolate peroxidases (HTPs) and, as their name implies, originate from various fungal sources.[125] Though structurally similar to P450s, peroxidases are generally more stable.[126] Chloroperoxidase (CPO) from the ascomycetous fungus *Leptoxyphhium (Caldariomyces) fumago* was the first HTP to be described in the 1960s and has been a focus of research by synthetic chemists since.[127] This enzyme generally catalyses oxidation of halides to hypohalous acids while also possessing limited peroxygenase activity; CPO can transfer oxygen to activated substrates such as indole or styrene but is unable to react with non-activated carbons in aromatic or alicyclic rings, or in *n*-alkanes.[126] The limitations of CPO can be overcome by utilising unspecific peroxygenase (UPO), an HTP-type enzyme first described in the fungus *Agrocybe aegerita (Aae*UPO) in 2004.[128] *Aae*UPO together with animal myeloperoxidase, plant seed peroxygenase and bacterial fatty acid peroxygenase form the second sub-class of peroxidases. Following *Aae*UPO, two similar enzymes from the mushrooms *Coprinellus radians* (*Cra*UPO) and *Marasmius rotula* (*Mro*UPO) were discovered.[129-130] These three known UPOs catalyse a range of peroxide-dependent oxygenations with catalytic efficiencies far greater than those observed in CPO and P450s.[126]

UPO Structural Features and Properties

The three previously mentioned UPOs possess a haem group linked *via* the iron to an axial cysteine ligand. Like with P450s, this feature causes a characteristic UV-vis spectrum of the reduced complex at 445-450 nm.[125, 131] UPOs are highly glycosylated proteins with up to

seven high mannose *N*-glycosylation sites and have molecular masses and isoelectric points that range from 32 to 46 kDa and 3.8 to 6.1 respectively.[132-134]

The crystal structure of AaeUPO was recently solved and revealed ten α-helices, five short β-sheets, a magnesium close to the haem propionate and a disulphide bridge in the C-terminal region.[133] The haem channel is conically shaped like a with an outer diameter and inner diameter of roughly 10 Å and 7 Å respectively, containing ten aromatic residues (9 Phe and 1 Tyr) responsible for directing the substrate to the active site.[126] Key active site amino acids include a PCP-motif (P35, C36, P37) bound axially to the haem group *via* the central cysteine, a magnesium binding EGD-motif (E122, G123, D124) and a charge stabilising arginine (R189) located close to a glutamate (E196) with the pair serving as acid-base catalysts in the peroxide cleavage reaction (Figure 1.21).[132-133]
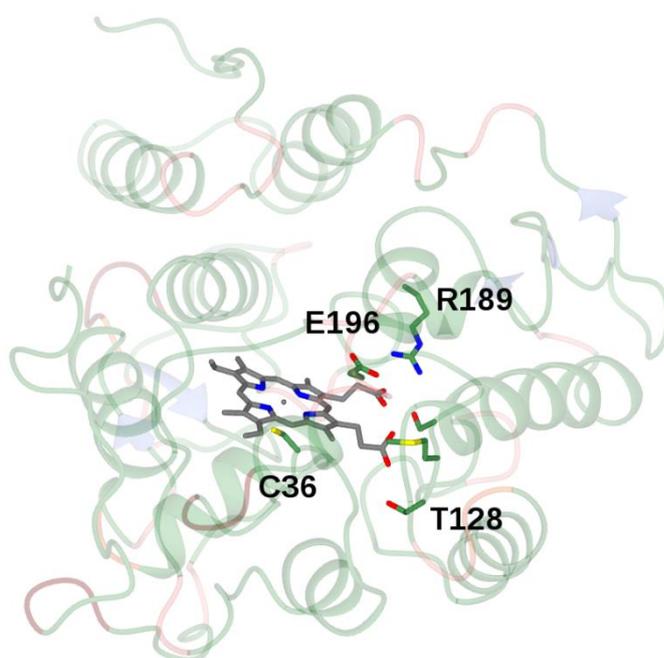
Figure 1.21: Overall structure of *Aae*UPO with an emphasis on the active site. Proposed amino acid residues that are integral to catalysis have been highlighted.

UPOs constitute a unique haem-peroxidase superfamily with BLAST searches in GenBank and other sequence databases identified over 1000 nucleotide sequences encoding for supposed UPO proteins. Most of these sequences were found in the Ascomycota and Basidiomycota phyla but are some are also present in other fungal families including the Mucoromycotina, the Glomeromycota, the Chytridiomycota and the Oomycota, suggesting that these enzymes are an evolutionary ancient superfamily specific to fungi.[126] In these

searches, there was no evidence for the presence of UPO-like genes in plants, animals or prokaryotes.[134-135]

Hydroxylations by peroxygenases

As with P450s, peroxygenases are capable of catalysing dozens of reactions including the oxidation of aromatic and heterocyclic compounds, epoxidation of alkenes, sulfoxidation, alkane and aromatic hydroxylation among many others.[123-124, 126]

UPOs have the capacity to hydroxylate varied linear, branched and cyclic alkanes together with alkyl species such as fatty acids (Figure 1.22).[126] The product distribution and selectivity seen in these reactions can vary significantly depending on the biocatalyst utilised and substrate features.[124] In the case of AaeUPO for example, fatty acids are transformed into a mixture of ω, ω-1 and ω-2 hydroxylation products (Figure 1.22(v)), reactions with linear alkanes result in similar phenomena, with a mixture of 2- and 3-alkanols usually formed (Figure 1.22(ii)).[136-137] Though the regioselectivity of alkane hydroxylation using peroxygenases is generally poor, the enantioselectivity can be very high. For example, a vastly different regioselectivity is observed when transforming fatty acids with peroxygenases from *Bacillus subtilis* (P450$_{BSβ}$) or *Clostridium acetobutylicum* (P450$_{Clα}$), with the former performing α-hydroxylation and the latter β-hydroxylation.[138-139]
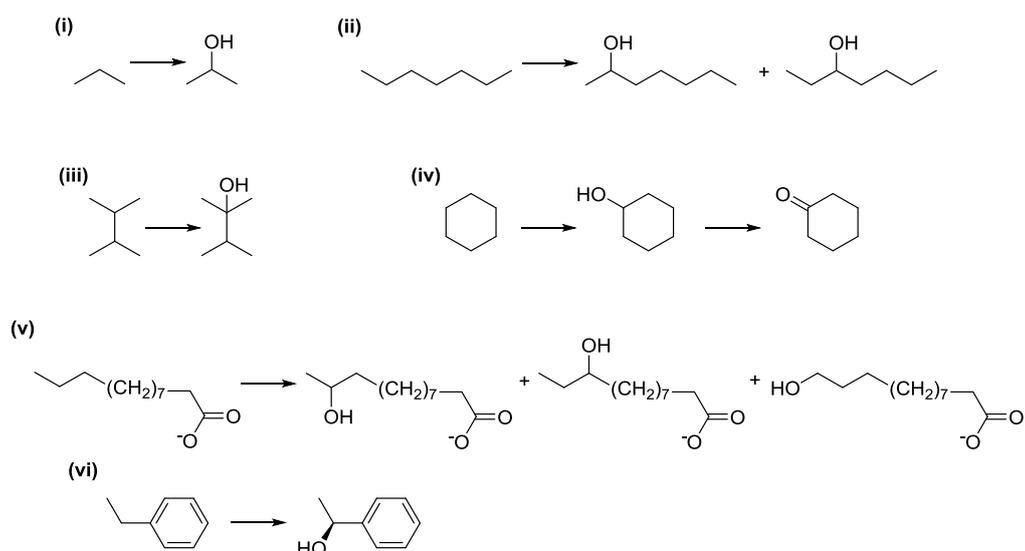
Figure 1.22: Examples of UPO-catalysed hydroxylation of alkanes and alkynes: (i) propane to propanol, (ii) *n*-heptane to 2-heptanol and 3-heptanol, (iii) 2,3-dimethylbutane to 2,3-dimethyl-butan-2-ol, (iv) cyclohexane to cyclohexanone *via* cyclohexanol, (v) lauric acid to ω-1-hydroxylauric acid and ω-2-hydroxylauric acid as major products with trace ω-hydroxylauric acid, (vi) ethylbenzene to (*R*)-1-phenylethanol. Figure adapted from Hofrichter and Ulrich (2014).[126]

Gutiérrez and colleagues recently contributed a rationale for the differences in selectivity between peroxygenases. The authors showed that the hydroxylation of cholecalciferol (Vitamin D) is predominantly unselective when catalysed by AaeUPO, while a single product is observed when the reaction is catalysed by a peroxygenase from *Coprinopsis cinerea* (*Cci*UPO) (Figure 1.23). This data was used to suggest that such a difference in selectivity was the result of variations in the size of substrate access channels and degrees of substrate translational freedom.[140]
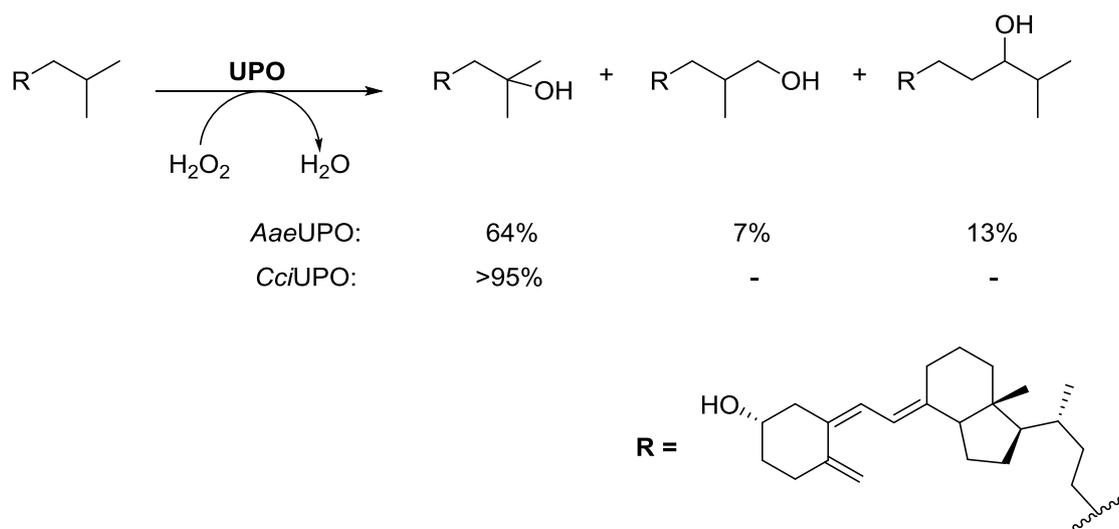
Figure 1.23: Reaction of cholecalciferol with *Aae*UPO and *Cci*UPO. The relative quantities of each product have been highlighted in the case of each enzyme. Figure adapted from Wang *et al.* (2017).[124]

A frequently observed issue in the peroxygenase-mediated hydroxylation of alkanes is the subsequent conversion of the produced alcohol to the corresponding aldehyde and ketone (Figure 1.22(iv)), with such an 'overoxidation' being extremely detrimental if the product of interest is a chiral alcohol.[141] The extent of such overoxidation is dependent on the peroxygenase utilised: *Aae*UPO and *Cci*UPO show substantially lower overoxidation of cyclohexanol to cyclohexanone than the peroxygenase from *Mariasimus rotula* (*Mro*UPO); a feature that could be manipulated for the selective accumulation of either product.[124, 141]

The selectivity of the peroxygenase-mediated hydroxylation of alkanes is highly dependent on the nature of substrate itself. Substrates with activated C-H bonds are usually transformed with greater selectivity. Ethers and secondary amines, for example, are often attacked at the α-position relative to the heteroatom resulting in dealkylation reactions.[142] Taking this further, *Aae*UPO is capable of hydroxylating benzylic C-H bonds with high

regio- and enantioselectivity, however, this latter feature was found to be strongly affected by the increasing steric demand of the alkyl chain and overoxidation remains a problem.[143]

In addition to the aforementioned alkane hydroxylations, a number of aromatic hydroxylations using peroxygenases have also been reported, among which are the regioselective hydroxylations of 2-(4-hydroxyphenoxy)propionic acid and diclofenac.[144-145] Aromatic hydroxylation (Figure 1.24) proceeds *via* formation of epoxide intermediates that spontaneously rearrange to the corresponding phenol product, with these epoxide intermediates released into the reaction mixture which suggests that the peroxygenase itself is not involved in the process of epoxide rearrangement.[142, 146] The peroxygenase-mediated hydroxylation of aromatic compounds is often faced with a challenge resulting from the fact that the phenol products are likely to experience peroxygenase-mediated H-atom abstraction to produce phenoxy radicals which often polymerise.[147-148] This drawback can generally be addressed by adding radical scavengers such as ascorbic acid to the reaction mixture or by protein engineering.[124, 147]
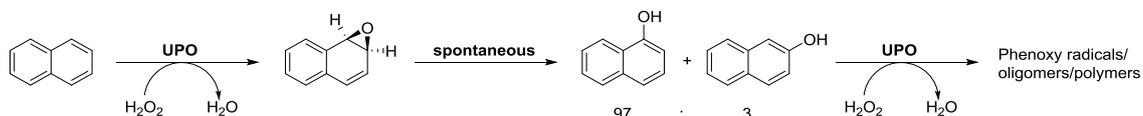
Figure 1.24: Aromatic hydroxylation mechanism. The reactions proceeds via the UPO mediated formation of an epoxide intermediate which undergoes spontaneous rearrangement to give the phenol products. The phenol products can then undergo peroxygenase mediated H-atom abstraction to produce radicals.

Challenges associated with peroxygenases

As of 2017, over 1000 putative peroxygenase sequences have been identified in genome databases, however, only a fraction of these have been characterised with respect to their catalytic potential in chemical synthesis.[126] This is primarily due to a number of challenges associated with peroxygenases.[124]

Most peroxygenases are extracellular fungal enzymes which makes protein production and isolation a complex process.[123] Fungal peroxygenases are usually produced by fed-batch cultivation of the natural organism in fermenters or shake flasks, with the enzymes secreted into the culture medium during the fermentation process.[123] The quantity of enzyme produced can vary from a few milligrams to several grams per litre depending on the

expression system employed and the enzyme itself. For example, the glycoprotein nature of *Aae*UPO makes its functional expression in *Escherichia coli* an extremely difficult process resulting in current research efforts being focussed on developing efficient fungal expression systems.[123-124] Alcalde and colleagues successfully expressed *Aae*UPO in *Saccharomyces cerevisiae*, however, enzyme yields were initially very poor (0.007 mg L^{-1}).[149] Further work focussed on evolving an attached signal peptide, which resulted in a substantially improved enzyme yield of 217 mg L^{-1}.[150] Such engineering approaches are currently being applied towards improving the expression of other peroxygenase targets.[124]

Peroxygenases, as with most enzymes, have been optimised to function in their native host systems, presenting a number of practical difficulties when used 'as obtained' for chemical reactions. Practical problems associated with peroxygenases generally fall into two categories: (1) the often limited enzyme stability and activity under chemical process conditions and (2) low reactant concentrations. Protein and reaction engineering approaches have been utilised in an attempt to overcome these issues.[123]

Stability issues in peroxygenases often result from the fact that these enzymes require hydrogen peroxide for function, yet elevated amounts of peroxide can destabilise them due to the fact peroxygenases, like other haem-dependent hydroxylases, are sensitive to it.[124] Protein engineering approaches have focused on both rational mutagenesis of easily oxidisable amino acids for enhanced oxidative stability and random mutagenesis for further stability enhancement when rational design is limited.[151] Further protein engineering has focussed on improving stability at elevated temperatures and in the presence of organic solvents.[123] Reaction engineering approaches aimed at tacking the peroxide stability issues have focused on maintaining the hydrogen peroxide at a concentration sufficient for catalysis but low enough to maintain enzyme stability. The most prevalent approach for controlling the concentration of hydrogen peroxide is currently the use of glucose oxidase to produce it from stoichiometric amounts of β-D-glucose and molecular oxygen.[152] The use of this system is popular due to its low cost and simplicity, however, it suffers from the major drawback of low atom-efficiency and the stoichiometric accumulation of gluconic acid (or gluconolactone) which could be extremely disadvantageous at an industrial scale.[124] Other reaction engineering approaches for controlling peroxide levels include the use of electrochemical methods involving the cathodic reduction of molecular oxygen; however,

these suffer from the setback of requiring electrochemical apparatus with the same flaws applicable to photochemical approaches.[153-154] Hollmann and colleagues recently proposed an enzyme cascade in which methanol is completely oxidised to carbon monoxide to produce an ideal amount of hydrogen peroxide that is fully utilised to promote efficient peroxygenase-mediated hydroxylations whilst avoiding the aforementioned stability issues.[155]

The second practical issue associated with peroxygenases is low substrate loading, with 1-10 mM of substrate usually converted by these enzymes, hindering their attractiveness as industrial biocatalysts.[156] Numerous reaction engineering approaches have been proposed to tackle this problem including a two-liquid-phase approach (previously described for BVMOs) in which a hydrophobic organic phase (the substrate) functions as both a substrate reservoir and product sink.[157]

Peroxygenases are extremely promising enzymes, holding the potential to catalyse the dream reactions performed by P450s while avoiding the need for redox proteins. These enzymes currently suffer from a number of setbacks including production difficulty and poor stability and substrate loading hindering their industrial use. However, engineering approaches are actively being pursued granting these enzymes significant promise.

1.4.3 Flavin Dependent Hydroxylases

Flavin dependent hydroxylases constitute a subgroup of the flavin dependent monooxygenases discussed in section 1.3. As previously mentioned, flavin dependent monooxygenases are classified into six groups, A-F, based on their structural and functional properties.[27] Most flavin dependent hydroxylases fall under group A, which consists of enzymes encoded by a single gene, that use NAD(P) as the electron donor and possess a glutathione reductase (GR-2) type Rossmann fold for FAD binding.[158]

The majority of Class A flavin dependent hydroxylases catalyse the regioselective hydroxylation of phenolic compounds at the *ortho-* and *para-* positions. These hydroxylases usually act *via* an electrophilic aromatic substitution mechanism in which the C4a-hydroperoxide serves as an electrophile while substrates containing hydroxyl or amino activating groups function as nucleophiles. In the catalytic mechanism (Figure 1.25), binding of the substrate greatly enhances the rate of flavin reduction by NAD(P)H and the resulting NAD(P)$^+$ is immediately liberated from the reaction site.[26] The release of NAD(P)$^+$ sets these enzymes apart from BVMOs which are Class B flavin dependent monooxygenases and keep their pyridine cofactor bound to the enzyme during catalysis.[26]
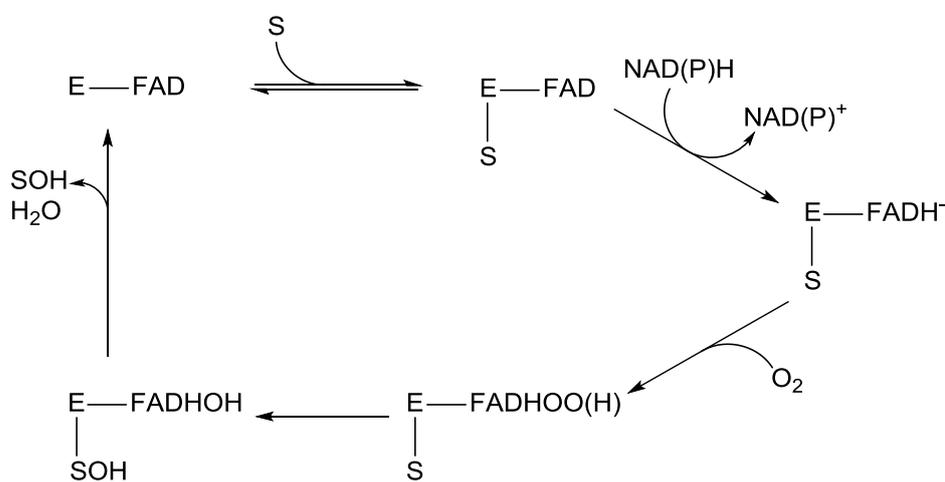
Figure 1.25: Catalytic cycle of Class A flavin dependent hydroxylases. E = enzyme; S = substrate. Adapted from Huijbers *et al.* (2014).[26]

Para-hydroxybenzoate hydroxylase (PHBH)

A great amount of the work focused on the structure and mechanism of Class A flavin dependent hydroxylases has resulted from studies of *para*-hydroxybenzoate hydroxylase (PHBH) from *Pseudomonas fluorescens*, which has served as a model for this class of enzymes

for decades having had its crystal structure elucidated in 1979.[159] PHBH is a versatile enzyme that not only catalyses the hydroxylation of 4-hydroxybenzoate but also 2,4-dihydroxybenzoate, 4-mercaptobenzoate and the chlorinated and fluorinated derivatives of its natural substrate (Figure 1.26).[76]
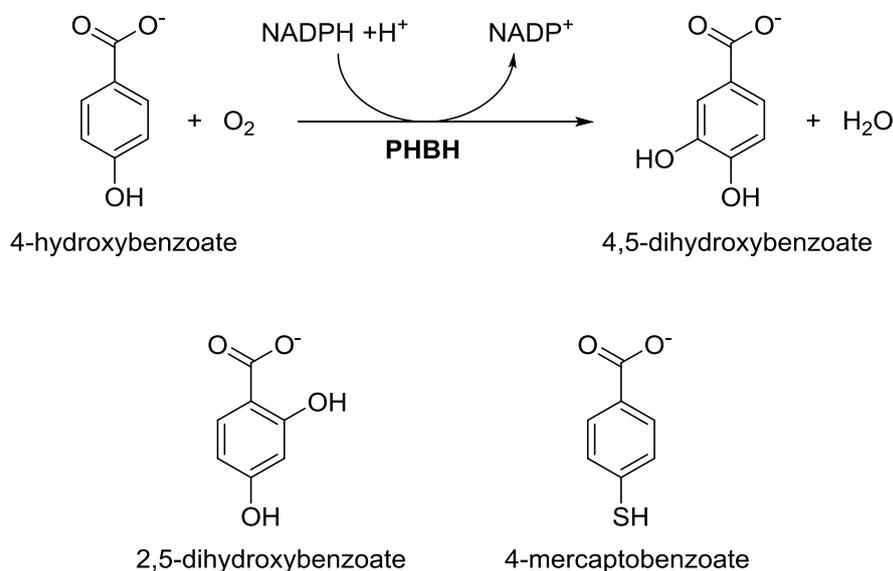
Figure 1.26: Aromatic hydroxylation reactions performed by *p*-hydroxybenzoate hydroxylase (PHBH). Adapted from Roper *et al.* (2016).[76]

PHBH is an enzyme composed of 394 amino acids, possessing three domain structure (Figure 1.27) consisting of an N-terminal Rossmann fold domain responsible for binding the ADP of the flavin coenzyme, a second domain that binds the 4-hydroxybenzoate substrate and a third predominantly helical domain which plays a role in dimerization with a second subunit.[76] The tricyclic isoalloxazine ring of FAD is bound at the interface between the first two domains and a series of mutagenesis studies revealed that this transient flavin cofactor uses different sites for substrate hydroxylation and flavin reduction with the enzyme coordinating catalysis by undergoing a series of conformational changes.[160] This flavin mobility is believed to be a general requirement of Class A enzymes, however, such a phenomenon has only been observed in the crystal structures of a select few members of this family.[161] Further investigation of the structural features of PHBH revealed that a hydrogen bond network the connects the substrate to the protein surface also acts as a sensor for flavin movement and reduction while also deprotonating the substrate to enhance its nucleophilic ability to react with the C4a-hydroperoxide.[162]
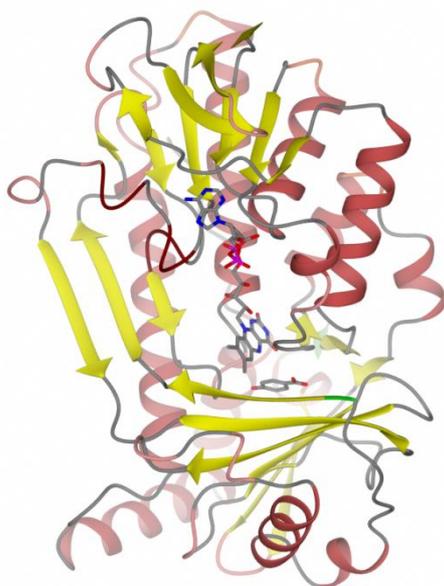
Figure 1.27: Structure of PHBH with FAD in the enzyme active site together with the substrate _p_-hydroxybenzoate directly below the FAD. Structure provided courtesy of Roper and Grogan (2016).[76]

3-Hydroxybenzoate-6-hydroxylase (3HB6H)

A second member of interest in the Class A flavin dependent hydroxylases is 3-hydroxybenzoate-6-hydroxylase (3HB6H). 3HB6H is dependent on the relatively inexpensive NADH for function and catalyses the _para_-hydroxylation of 3-hydroxybenzoate to 2,5-dihydroxybenzoate (Figure 1.28).[26]

3-hydroxybenzoate   2,5-dihydroxybenzoate
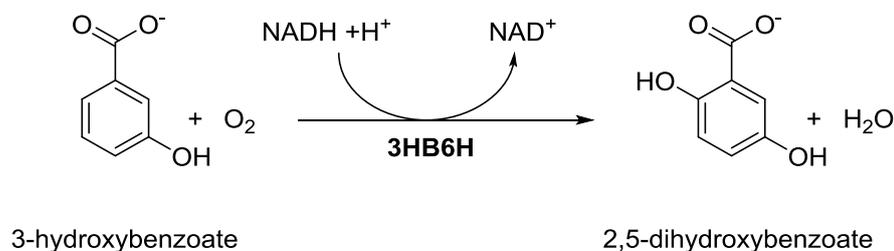
Figure 1.28: Aromatic hydroxylation of 3-hydroxybenzoate to 2,5-dihydroyxbenzoate by 3-hydroxybenzoate-6-hydroxylase (3HB6H).

The crystal structure of 3HB6H was solved by van Berkel and colleagues and revealed the enzyme to have a similar fold to that of PHBH.[163] This structure revealed that each subunit of the 3HB6H homodimer contains a hydrophobic channel filled with phospholipids that connects the protein surface to the substrate-binding site.[163] Following an overlay of this structure with other solved flavoprotein hydroxylases, van Berkel and colleagues mutated active site residues, Tyr105, His213 and Gln303, suspected of playing roles in substrate binding, particularly by interacting with carboxylate and hydroxyl groups. Crystal structures

of the three mutants were obtained and that of the His213Ser mutant with bound 3-hydroxybenzoate suggested that this residue is key for substrate hydroxylation, while a Gln-Tyr pair (N301, Y217) next to the phospholipid are important for substrate binding and coordination (Figure 1.29).[163] The active site geometry of this enzyme differs greatly from that of PHBH and the mutation work directed at it revealed important information regarding the structural determinants of regioselectivity in flavin dependent hydroxylases which was a limitation of the results obtained from the studies focused on PHBH.[26]
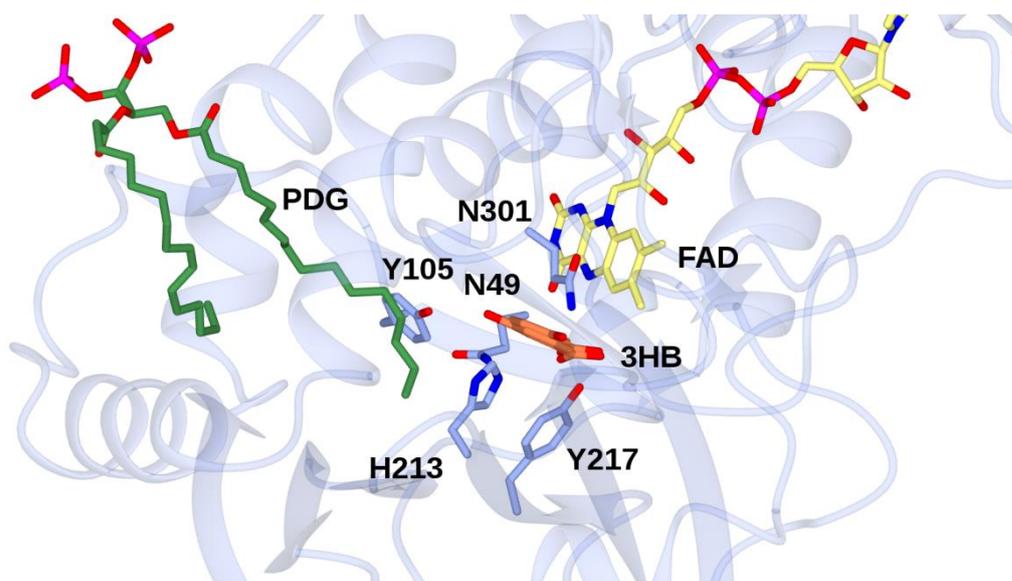
Figure 1.29: Active site of 3HB6H from *Rhodococcus jostii* RHA1 (PDB: 4BJZ). Amino acid residues of significant have been labelled. The phospholipid, phosphatidylyglycerol (PDG), is presented in green. The substate, 3-hydroxybenzoate (3HB) has also been modelled into the active site using a H213S mutant structure (PDB: 4BK1).

2-Hydroxybiphenyl 3-Monooxygenase (HbpA)

2-hydroxybiphenyl monooxygenase (HbpA) is a Class A flavoprotein monooxygenase originally isolated from *Pseudomonas azelaica* by Kohler and colleagues after it was distinguished for its ability to grow in hydroxybiphenyl as a sole carbon source.[164-165] HbpA generally catalyses the hydroxylation of 2-hydroxybiphenyl to 2,3-dihydroxybiphenyl (Figure 1.30) but is also capable of performing the *ortho*-hydroxylation of a broad range of substrates including several 3-aryl, 3-alkyl and 3-halo catechol substrates.[166]

2-hydroxybiphenyl   2,3-dihydroxybiphenyl
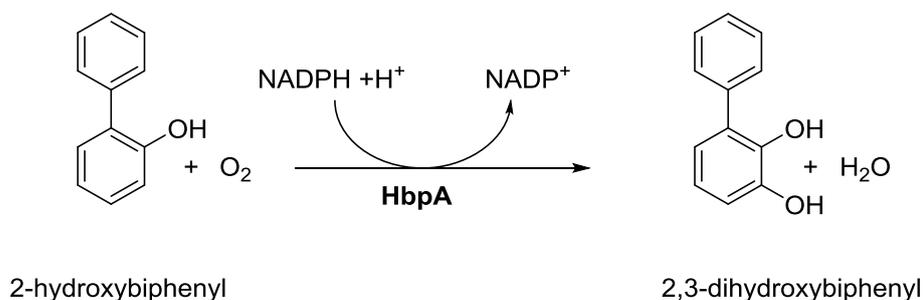
Figure 1.30: Hydroxylation of 2-hydroxybiphenyl to 2,3-dihydroxybiphenyl by HbpA.

The ability of HbpA to hydroxylate a range of 3-substituted catechol substrates is of significant interest because the 3-substituted catechol functionality is present in a large number of naturally occurring physiologically active compounds that possess pharmaceutical potential (Figure 1.31).[167] These are highly challenging to synthesise by chemical means, traditionally requiring the use of the metal halides $AlCl_3$, $FeCl_3$ and $ZnCl_2$ to act as Lewis acids in the electrophilic substitution reaction, presenting a range of safety and waste disposal issues.[168] Chemical synthesis routes are also faced with difficulties in finely controlling which regioisomer of the catechol is produced. These industrial challenges have resulted in the biocatalytic route being an extremely attractive alternative option.[167-168]

Miltrone   Urushiol   Barbatusol   Taxodione
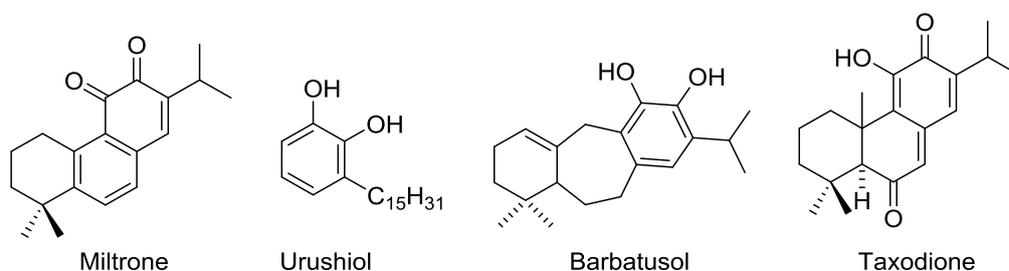
Figure 1.31: Some physiologically active compounds containing a 3-substituted catechol or *ortho*-quinone moiety. Adapted from Held *et al.* (1998).[167]

HbpA is single component enzyme that utilises NADH as cofactor, making it a model system for the investigation of methods for NADH regeneration including enzymatic and electrochemical methods.[165] The capacity of this enzyme to hydroxylate a range of substrates have also made it a target for evolution studies in order to optimise its abilities; for both these features, structural information is highly desirable in order to investigate mechanism and further rationalise mutation targets.

Grogan and colleagues solved the crystal structure of HbpA in both the apo and FAD-bound forms.[169] The structural information revealed HbpA to be a tetramer (a property unique to HbpA), with the monomers broadly resembling other members of the Class A flavin dependent monooxygenases (Figure 1.32). The monomer consists of three domains: a Rossman-type FAD-nucleotide binding domain (D1), a central substrate binding domain (D2) and a thioredoxin domain (D3).[169] A series of loops connecting the β-sheets in D2 show poor density suggesting a higher level of mobility in this region, supporting previous data claiming that Class A flavin dependent monooxygenases undergo a series of rearrangements following substrate binding.[169]
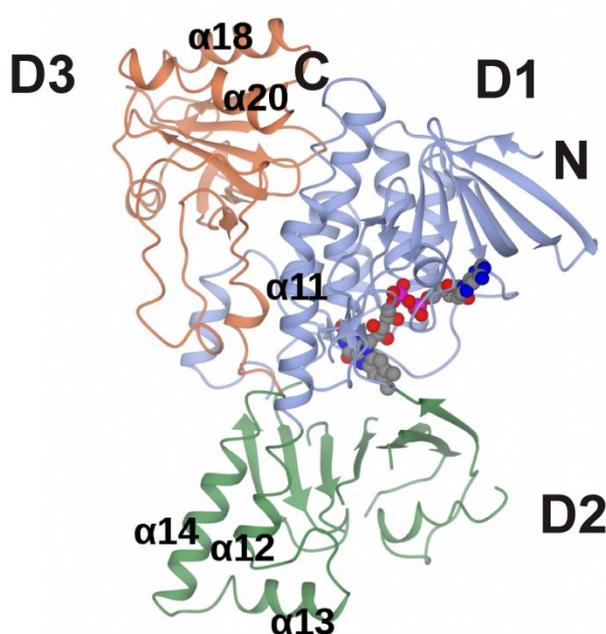
Figure 1.32: Monomer of HbpA in ribbon format in complex with FAD (shown in sphere format, carbons in grey)). Selected α-helices have been labelled. The isoalloxazine ring of FAD is bound in the interface of domains 1 and 2 (D1 and D2). N and C represent the respective termini of the monomer. Image provided by Grogan et al. (2015).[169]

Comparisons of the apo and FAD-bound structures showed a pronounced shift in the relative orientation of D2 away from D1 as a consequence of FAD-binding, resulting in a more open active site. The active site pocket within D2 contains a number of hydrophobic amino acids including Trp97, Met223 and Trp225, and interestingly, also contains two hydrophilic residues Arg242 and His48 (Figure 1.33). The His48 was found to be located 5.5 Å away from the C4a atom of FAD, the location at which the hydroperoxides forms, and 4.0 Å from a water molecule in the vicinity of the FAD that served as a well-placed mimic for the substrate, suggesting that this His48 may play a key role in catalysis.
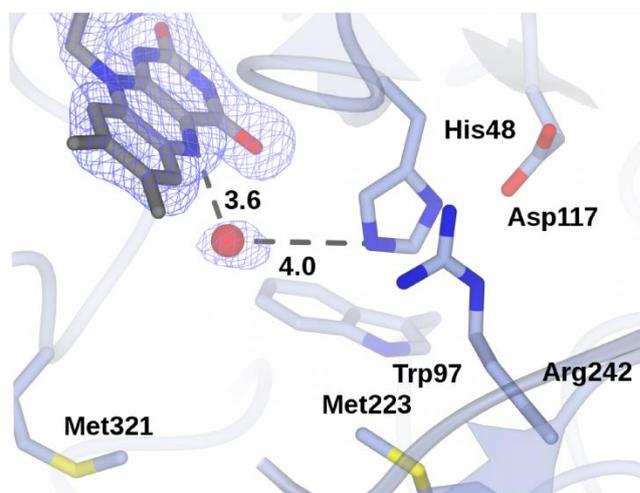
Figure 1.33: Detailed view of HbpA active site with a particular focus on the FAD binding domain. The water molecule located the closest to C4a atom of the flavin (top right structure with surrounding density) is presented in red. Amino acids of interest have been highlighted.

Mutational studies further suggested that Asp117 and Arg242 could also play a large part in catalysis.[169] This hypothesis was corroborated by Fishman and colleagues by presenting a substrate bound HbpA structure from which it was confirmed that His48 plays a role in substrate deprotonation due to the orientation of the substrate relative to it and an observed hydrogen bond network between the two. Arg242 was proposed to be central for facilitating FAD movement and reduction, while Trp225 assists in correctly positioning the entering substrate within the binding the site.[165] Further biochemical analysis, suggested that a distant Gly255 is likely located at the substrate entrance site due to the fact that mutation of this residue to a bulky phenylalanine hindered substrate entry to the active site and affected NADH binding or oxidation.[165]

In summary, flavin dependent hydroxylases are appealing biocatalysts due to ease of use as they are single component enzymes. However, these enzymes have a somewhat limited substrate scope due to being predominantly able to hydroxylate phenolic compounds.[76] Fortunately, high resolution crystal structures of members of this family are continuing to aid efforts in understanding the mechanism of the enzymes and assisting in rationalising targets for mutations in an attempt to optimise their scope and increase their potential applications.[165]

1.4.5 2-Oxoglutarate-Dependent Oxygenases

The previously discussed hydroxylases each exhibit a number of strengths; however, they all suffer from the fact that they are dependent on often costly nicotinamide and/or flavin cofactors for catalysis. 2-Oxoglutarate-dependent oxygenases (2OGDOs) are non-haem iron dependent enzymes capable of performing a diverse range of oxidation reactions, requiring ferrous iron, the inexpensive 2-oxoglutarate, ascorbic acid and molecular oxygen for catalysis.[170] These enzymes are of significant interest due to their ability to functionalise inert C-H bonds in complex molecules with high regio- and stereoselectivity without the need for a costly cofactor regeneration system as in P450s.[171]

Oxygenases that do not utilise a cofactor other than iron now constitute a distinct superfamily of redox enzymes.[172] This family contains both di-iron dependent enzymes, for example methane monooxygenases, and mono-iron dependent enzymes, which require ferric iron (lipoxygenases and intradiol cleaving cathecol dioxygenase) or ferrous iron as is the case for 2OGDOs.[172] 2OGDOs are likely the largest known family of non-haem dependent oxygenases and have been found to occur ubiquitously in numerous organisms including bacteria, fungi, plants and vertebrates, with some also observed in viruses.[172-174]

2OGDOs tend to have an absolute requirement for Fe(II) and catalyse various two-electron oxidations including hydroxylation, desaturation and oxidative ring closure reactions, with hydroxylations being the most the common.[173-174] Figure 1.34 presents a simplified hydroxylation reaction of an aliphatic C-H bond by a 2OGDO, in the reaction, the oxidation of the primary substrate is coupled to the decarboxylation of the 2-oxoglutarate into succinate and carbon dioxide with one of the molecules of dioxygen incorporated into succinate and the other into the substrate.[172]
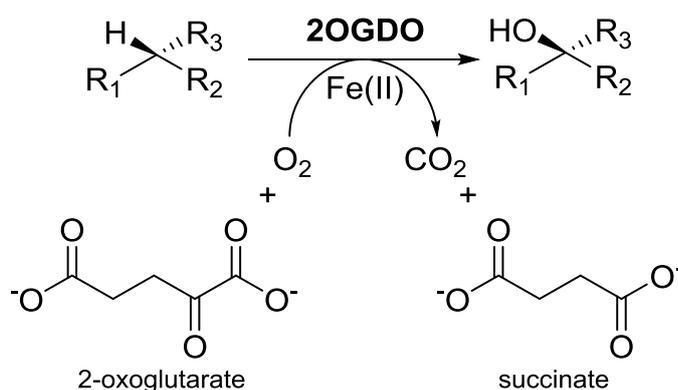
Figure 1.34: General hydroxylation of an aliphatic C-H bond catalysed by a generic 2OGDO. The oxidation of the primary substrate is coupled with decarboxylation of 2-oxoglutarate to succinate and CO_2.

Natural Roles

2OGDOs are present in almost all phyla of life. Most of these enzymes catalyse hydroxylation reactions and play key roles in a number of diverse metabolic processes, which are discussed below.[175-176]

(i) Protein Modification

At least four 2OGDOs that respectively hydroxylate the proline, lysine, aspartic acid and asparagine side chains of proteins have been identified (Figure 1.35).[170] Prolyl-4-hydroxylase (P4H) was the first 2OGDO to be identified and catalyses the *trans*-4-hydroxylation of peptidyl proline residues in mammals.[177] This transformation is essential for the formation of collagen, elastins and other proteins due to the hydroxylated residues forming intra- and interchain hydrogen bonds that stabilise the triple-helical structure of these proteins.[178] Prolyl-3-hydroxylase (P3H) is less characterised than P4H, but is known to hydroxylate peptidyl proline to 3-hydroxyproline, with this product observed at a level of up to 10% of the amount of 4-hydroxyproline observed in mature collagen.[178]
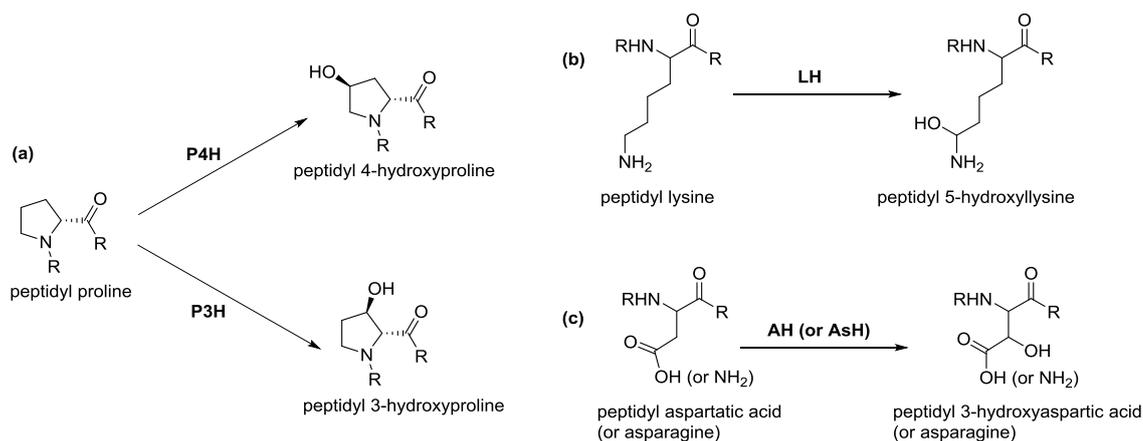
Figure 1.35: Peptidyl amino acid hydroxylation reactions catalysed by distinct 2OG-dependent hydroxylases in mammals. (a) Hydroxylation of peptidyl proline by proly-4-hydroxylase (P4H) and proly-3-hydroxylase (P3H). (b) Hydroxylation of peptidyl lysine by lysyl-hydroxylase (LH). (c) Hydroxylation of aspartic acid or asparagine by aspartyl (asparaginyl) β-hydroxylase (AH or AsH). Cofactors and co-substrates have not been depicted in this figure. R = chains in protein.

Lysyl hydroxylase (LH) is medically significant enzyme that catalyses the hydroxylation of peptidyl lysine to 5-hydroxy-lysine (Figure 1.35b), with this modification occurring in approximately 0.5 to 7% of X-Lys-Gly triplets in procollagen to act as attachment sites for galactose or glucosylgalactose units.[178] Deficiencies in LH activity are associated with the type-VI variant of Ehlers-Danlos syndrome which leads to various conncective tissue disorders.[170] Aspartyl (or asparagine) β-hydroxylase (AH or AsH) transforms the β-carbon of certain aspartic acid or asparagine residues (Figure 1.35c) in the epidermal growth factor-like domains of members of the vitamin-K-dependent proteins, complement factors and coagulation factors, with the modified residue forming a calcium-binding that likely plays a role in protein-protein interactions.[170] Human AH or AsH has been found, in numerous cases, to be expressed at elevated levels in tumour cells suggesting that these enzymes could be used as prognostic biomarkers to predict the progression of certain cancers.[179]

Protein modifications resulting from 2OGDOs also play a large role in the mammalian hypoxic response; this is primarily due to modifications of the hypoxia-inducible factor (HIF).[180] HIF is a αβ heterodimeric transcription factor that mediates the response to low oxygen tension (hypoxia) by regulating over 40 genes and control of its expression levels could be used for treatments of cancer, anaemia and ischemic disease.[181-183] In order for transcription of HIF to occur, the α-subunit needs to bind to a transcription coactivator (such as p300) and dimerise with the β-subunit.[180] Hydroxylation of the α-subunit at specific amino acid residues can render HIF inactive by either preventing it from

43

interacting with p300 or by targeting it for proteasomal destruction.[180] These hydroxylations are performed by two separate 2OGDO systems: factor inhibiting HIF (FIH) and prolyl hydroxylase domain containing isozymes (PHDs) (Figure 1.36).[180]
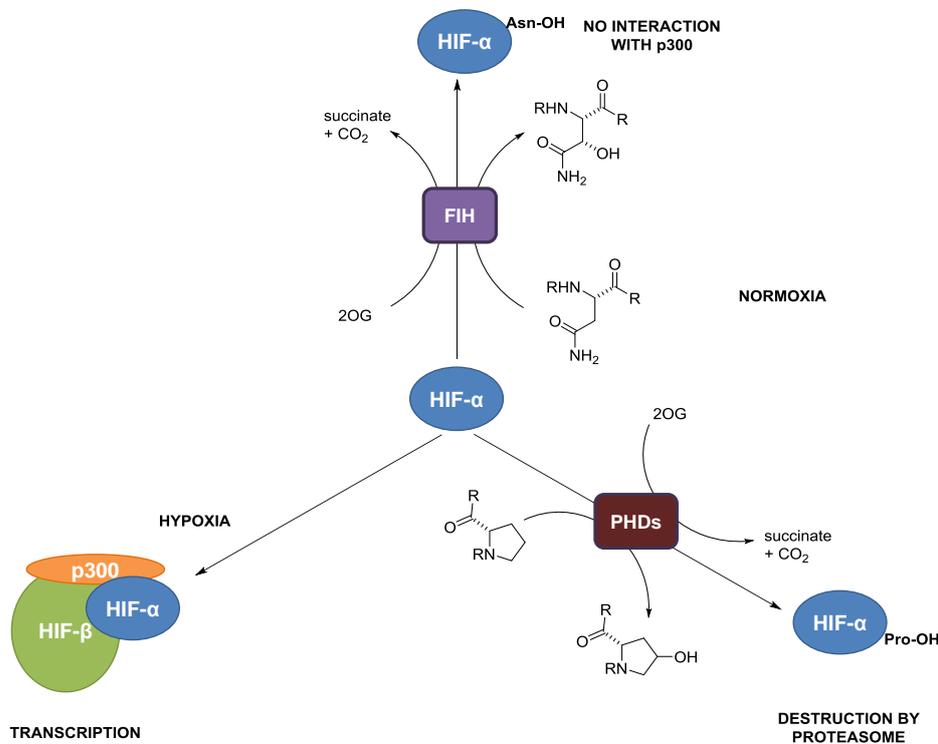
Figure 1.36: The HIF pathway. Under hypoxic conditions, HIF-α and HIF-β dimerise and recruit the transcription activator p300, resulting in transcription. Under normoxia, HIF-α is, either, hydroxylated at Asn803 by FIH and unable to interact with p300, or hydroxylated at either Pro403 or Pro564 by PHDs leading to its degradation by the ubiquitin-proteasome pathway. Figure adapted from Hewitson *et al.* (2005).[180]

FIH can hydroxylate HIF-α at the *pro-S* position of the β-carbon of Asn803 which abolishes the interaction between HIF-α and the p300 transcription coactivator, preventing the transcription of HIF. HIF-α can also be hydroxylated at the 4-position of two prolyl residues, Pro402 and Pro564 in human HIF-α, by PHDs which enables the destruction of the α-subunit by the ubiquitin-proteasome degradation pathway.[180]

(ii) DNA Repair

Some members of 2OGDO family have been found to play a role in the repair of alkylation-damaged DNA or RNA, which if untreated could lead to further toxic and mutagenic effects on the host.[170, 176] One such hydroxylase, AlkB, facilitates the repair of repair of such alkylated DNA or RNA by *via* an oxidative dealkylation mechanism in which a transient hydroxylated intermediate is produced (Figure 1.37).[184]
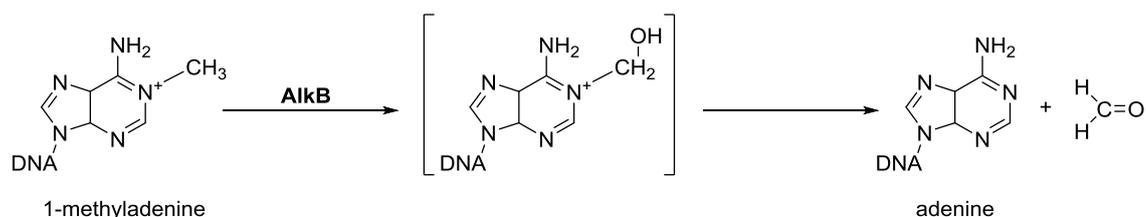
Figure 1.37: The demethylation of 1-methyladenine to adenine by AlkB.

(iii) Antibiotic Biosynthesis

2OGDOs are capable of catalysing numerous reactions in addition to various reactions such as desaturations, ring expansion and formation among others, allowing these enzymes to play key roles in the biosynthesis of diverse antibiotics.[170] These roles can vary from the production of precursors to the modification of more complete and complex intermediates.[170]

Clavaminic acid synthase (CAS) is an example of a highly versatile 2-oxoglutarate dependent oxygenase that appears to have evolved multifunctional activities, catalysing three distinct oxidative steps in the synthesis of the β-lactamase inhibitor clavulanic acid.[185] The biosynthesis of clavulanic acid from deoxyguanidoisoproclavaminic acid employs a two-enzyme cascade involving CAS and proclavaminate hydroxylase (PAH) (Figure 1.38). CAS catalyses the initial hydroxylation of deoxyguanidoisoproclavaminic acid to guanidinoproclavamic acid in which the guanidio group is then hydrolysed by PAH to yield proclavaminic acid, this acid then undergoes sequential cyclisation and desaturation reactions catalysed by CAS.[170]
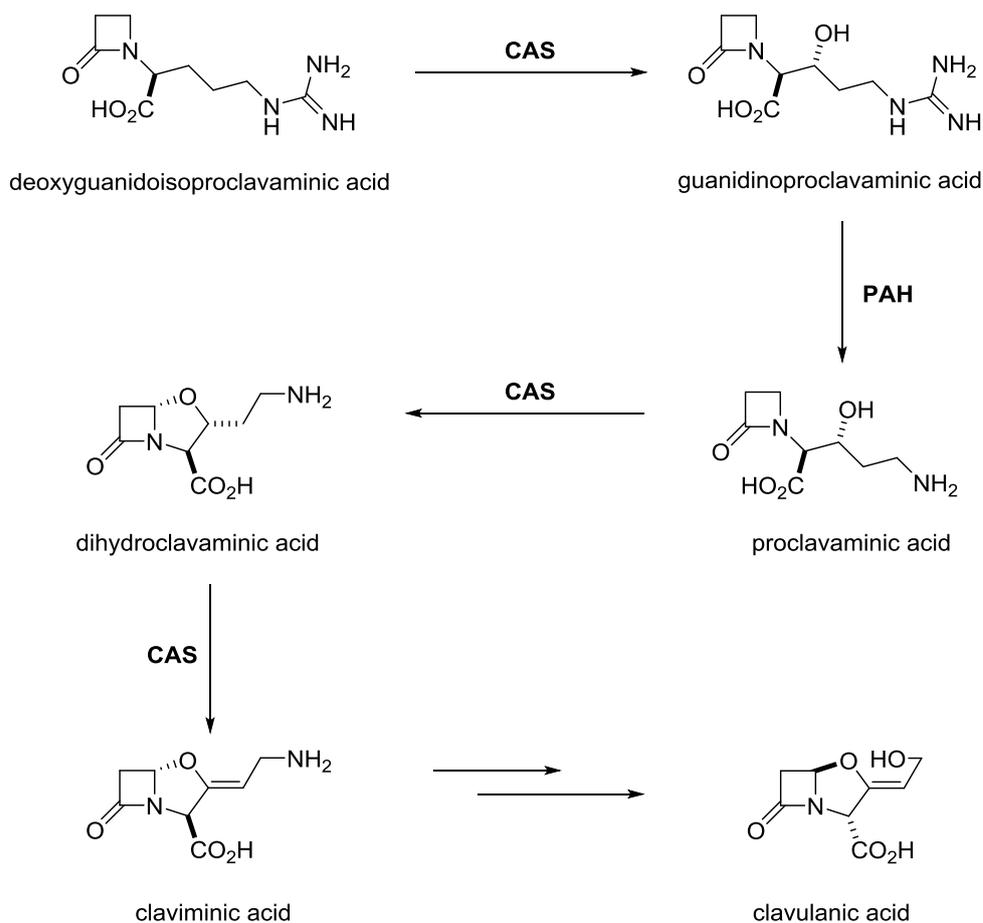
Figure 1.38: The biosynthesis of clavulanic acid from deoxyguanidoisoproclavaminic acid using a two-component enzyme cascade consisting of CAS and PAH.

(iv) Natural Product Biosynthesis in Plants

Plants utilise 2OGDOs for the biosynthesis of a variety of natural compounds including flavonoids, gibberellins and alkaloids. Flavonoids play a number of key roles in plants, including, protecting them from UV-light damage, functioning as signal and defence molecules, while also acting as antioxidants, antimalarial therapies and potential anti-cancer agents for humans.[170]

The flavonoid biosynthesis pathway, among others, highlights a particularly interesting feature of 2OGDOs, showcasing that separate enzymes are capable of catalysing closely-related or sequential steps in a pathway. In the flavonoid pathway, four 2OGDOs: flavone synthase I (FSI), flavone-3β-hydroxylase (F3H), flavanol synthase (FLS) and anthocyanidin synthase (ANS), catalyse unique oxidations on the C-ring of the substrates and intermediates present during the biosynthesis process.[186] Figure 1.39 presents a snapshot in the flavin biosynthesis pathway starting with naringenin as substrate. Naringenin is either

46

desaturated to apigenin by FSI or hydroxylated to dihydrokaempferol by F3H. Dihydroykaempferol is next either desaturated to kaempferol by FLS or reduced by dihydroflavolol reductase (DFR) to leucoperlargondin which is finally hydroxylated to perlargonidin by ANS.[186]
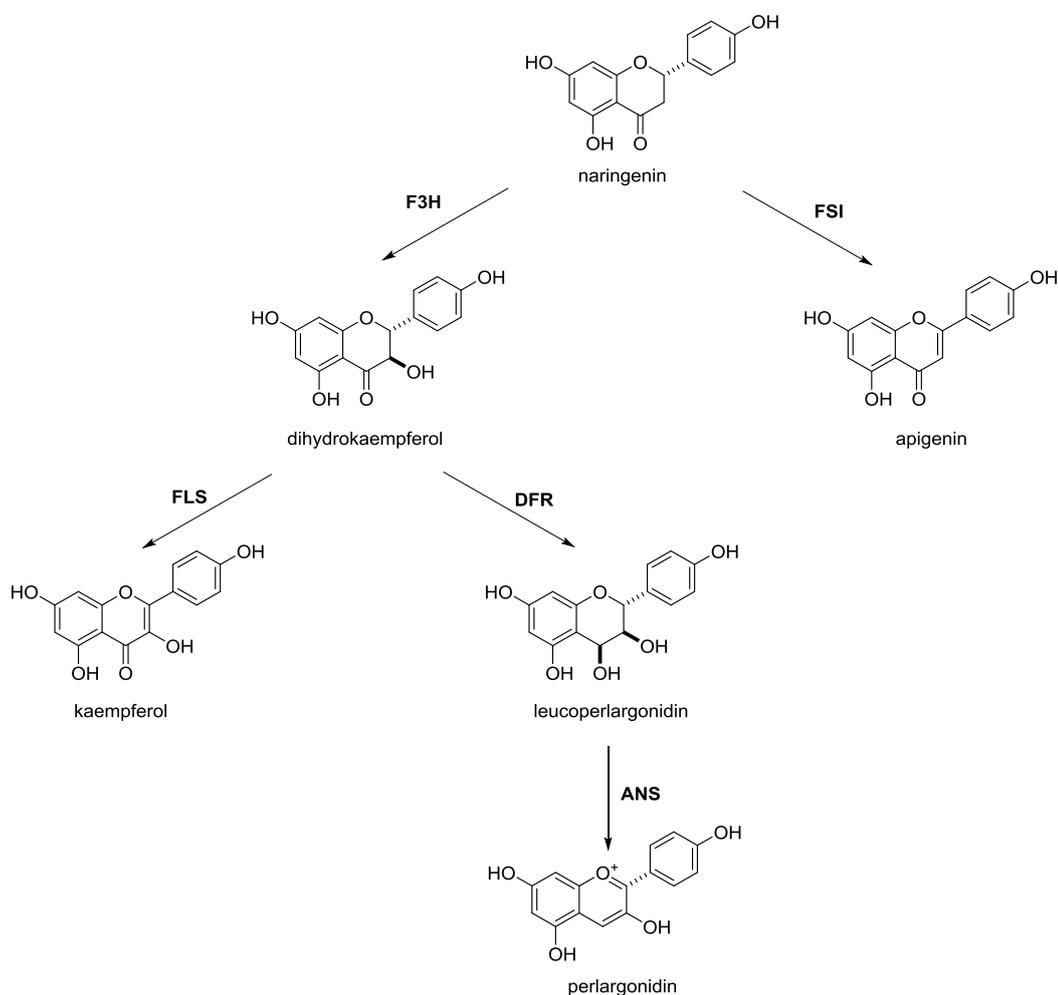
Figure 1.39: The biosynthesis of flavonoids starting with naringenin. The 2OGDO responsible for each different oxidation has been stated. Each enzyme can oxidise their respective substrate at different positions on the C-ring.

(v) Other Hydroxylations

2OGDOs perform numerous other key hydroxylation reactions in nature. Some members of this enzyme superfamily play integral roles in lipid metabolism, assisting in the synthesis of carnitine which is responsible for the transport of activated fatty acids across the inner membrane to promote their degradation by the β-oxidation pathway in the cellular matrix.[187] Another member of this family, phytanoyl-CoA hydroxylase, is required for the degradation of the plant lipid phytanic acid, with human deficiencies of this enzyme linked to disorders such as Refsum disease among others.[188]

Another interesting example is Taurine dioxygenase (TauD), a hydroxylase that is naturally expressed in *Escherichia coli*. Expression of TauD is upregulated during periods of sulphur starvation, enabling the hydroxylation of taurine to produce an unstable intermediate that readily decomposes to aminoacetaldehyde and sulphite, the latter of which is subsequently utilised by the organism as source of sulphur (Figure 1.40).[189]
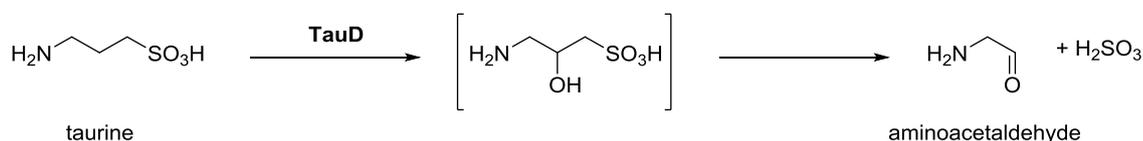
Figure 1.40: The release of sulphur from taurine as a result of hydroxylation by TauD.

Properties of 2-Oxoglutarate-Dependent Oxygenases

General Structural Features

2-Oxoglutarate-dependent hydroxylases are ubiquitous enzymes that catalyse diverse reactions; however, many structural features of members of this superfamily are highly conserved.[176] The structures of these enzymes share a common core jellyroll structural fold consisting of eight β-sheets that form eight stranded sides (Figure 1.41A).[170, 175] The enzyme active site contains a conserved His-X-Asp/Glu-X_n-His motif, ferrous iron, Fe(II), binding occurs at the facial triad of these amino acid residues (Figure 1.41B).[175]
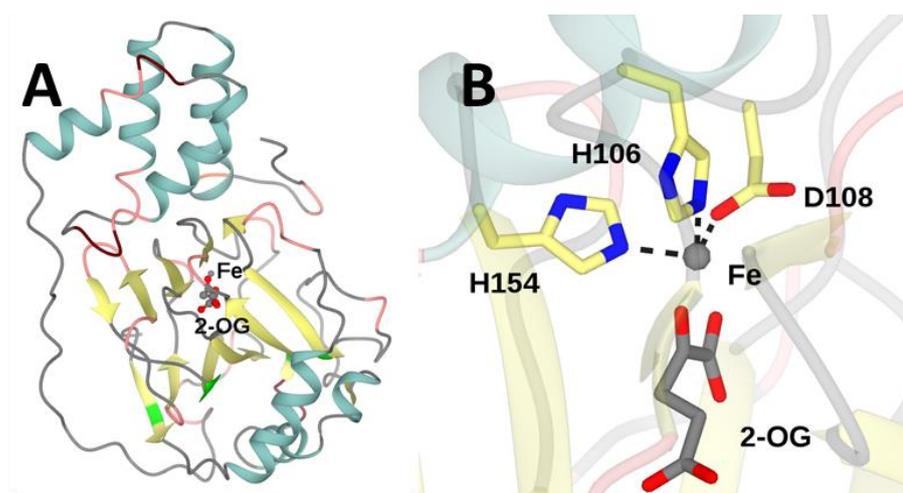
Figure 1.41: (A) The general structure of *cis*-4-proline hydroxylase from *Mesorhizobium loti* (MlC4H, PDB: 4PZW) with the active site highlighted in the centre. This highlights the general 2OGDO structural features. The central Fe is located within a jelly roll fold formed by 8 β-sheets (yellow). (B) Active site of MlC4H with the facial triad of H106, D108, H154 highlighted.

Catalytic Mechanism

Numerous structural and mechanistic studies were used to elucidate a hypothetical catalytic mechanism for oxygenation by 2-oxoglutarate-dependent oxygenases.[190-191] The mechanism (Figure 1.42) somewhat resembles that of P450s, however, the substrate free active sites generally contain ferrous iron rather than ferric iron as in P450s, with no reductase component required as the 2-oxoglutarate co-substrate functions as a two electron reductant to deactivate the molecule which itself does not react with the primary substrate.[176]

Figure 1.42 outlines the general mechanism employed by 2-oxoglutarate-dependent oxygenases for hydroxylation reactions. In the active site, an Fe(II)-centre is complexed with three water molecules and the amino acids of the facial triad (2 His, Glu) in an octahedral coordination geometry, **1**. In the first step of the reaction, 2-OG (green) displaces two of the water molecules, binding to the ferrous iron in a bidentate manner, **2**. Upon binding of the substrate, R-H (red) **3**, molecular oxygen replaces the remaining water ligand and is activated to a ferric superoxo species **4**. Following the formation of bridged alkyl-peroxo complex (brown), cleavage of the O-O bond and decarboxylation of the 2-OG (blue), an Fe(IV)-oxo complex is formed (**5**), which acts as the oxidant in the reaction. Next, hydrogen is abstracted (**6**) from the substrate R-H, which is followed by the formation of a C-O in the substrate via reaction of the substrate radical R. with the ferric hydroxyl group (rebound), **7**. This results in a recovery of the original 2+ oxidation state of the iron, together with hydroxylation of the substrate and the release of 2-OG as succinate (orange).
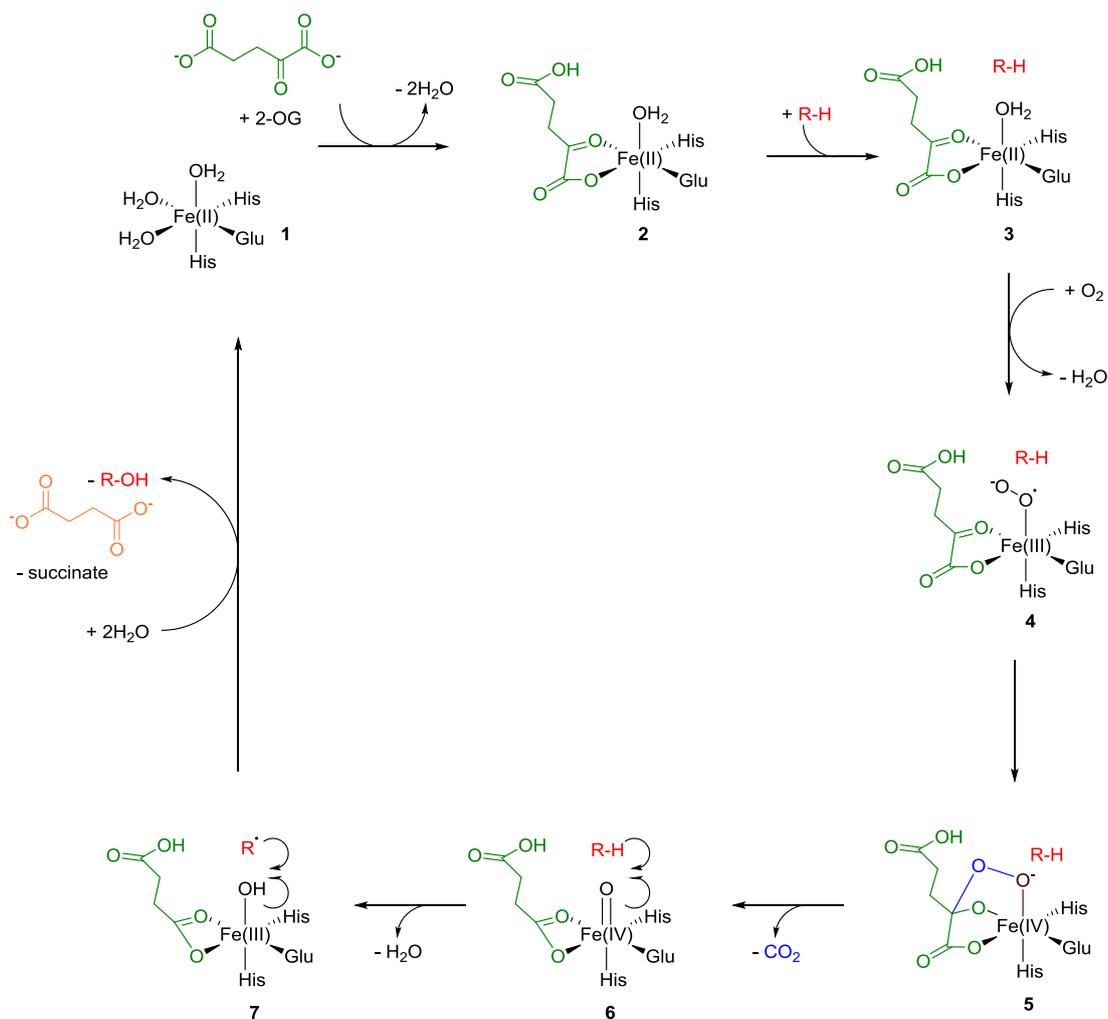
Figure 1.42: Common mechanism for hydroxylation reactions with 2-OG dependent non-haem Fe2+ hydroxylases. Fe(II), initially complexed octahedral to 3 H₂O molecules and a facial triad of His, Glu and His (1), is believed to undergo a series of oxidation/reduction steps to form a reactive Fe(IV)-oxo species (6), which reacts with the substrate R-OH to ultimately hydroxylate it. Figure adapted from Huttel *et al.* (2013).[176]

Applications of 2-oxoglutarate Dependent Oxygenases

The capacity of 2OGDOs to catalyse a broad range of reactions has resulted in the suggestion that these enzymes may be among the most versatile oxidising biocatalysts identified to date.[175] Biotechnological processes utilising these enzymes for the production of small molecules are still limited despite shown, however, there is still much promise as the catalytic scope of 2OGDOs is likely far from fully determined with new reactions continually being discovered as a result of advancing genomic analyses.[176, 192]

The most advanced biotechnological methods developed for utilising 2OGDOs have focussed on proline hydroxylation, hydroxy-isoleucine production and the transformation

of penicillin G to cephalosporin G.[176] The hydroxylation of proline and its derivatives is the focal point of the work described in this thesis.

1.4.6 Proline Hydroxylases

Amino acids are often referred to as the building blocks of life, possessing widespread functions including roles in the nutrition and health maintenance of living organisms.[193] As a result, it is not surprising that amino acids and their derivatives serve as valuable building blocks in industry. Proline, in particular, is especially useful for stereoselective synthesis due to possessing a rigid conformation and chirality-inducing properties.[194] Derivatives of proline, particularly those with substitutions in the carbon backbone, hold significant potential as useful and versatile chiral starting blocks that can be used to expand the chiral pool of compounds that can be synthesised.[195] For example, hydroxyprolines such as *trans*-4-hydroxy-L-proline (t4HyP) are extremely useful and versatile chiral synthetic units for the asymmetric synthesis of a wide range of pharmaceuticals (Figure 1.43).[196] Such drugs include carbanepem antibiotics, antispastic agents, galactosidase inhibitors and angiotensin-converting enzyme inhibitors.[194, 197]

carbanepem antibiotic backbone

acromelic acid A

galactosidase inhibitor

meroperem antibiotic

benzodiazepin derivative
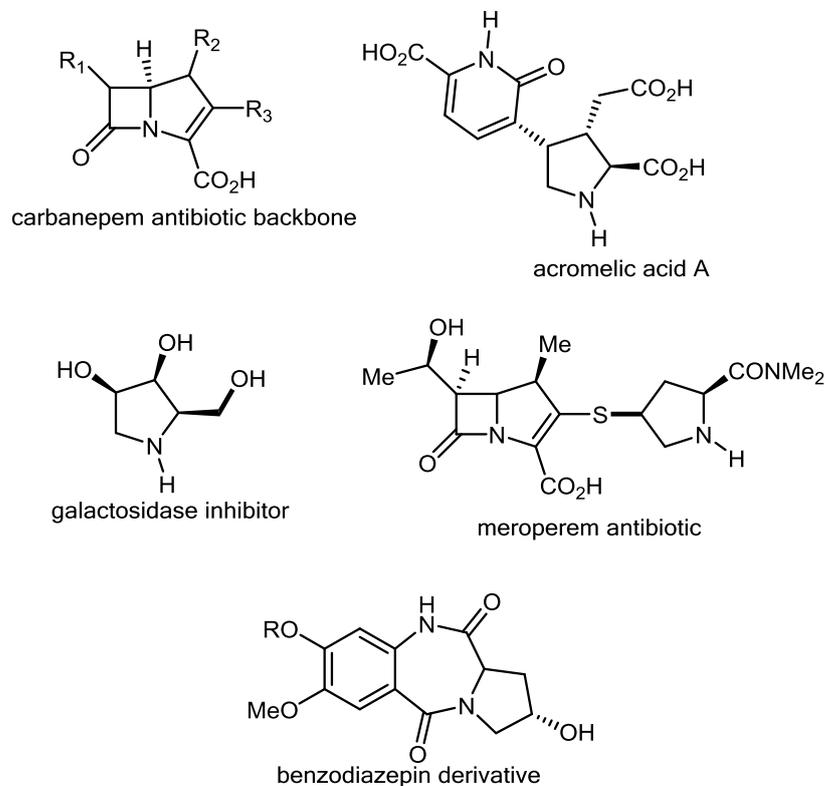
Figure 1.43: Examples of products formed from L-proline and its derivatives.

Trans-4-hydroxy-L-proline (t4HyP) is a non-essential amino acid present in both plants and animals, and is associated with a number of key pathways.[195] t4HyP possesses chiral centres at the C-2 and C-4 positions of the pyrrolidine ring with substitutions at each position of this ring possible, making this derivative a significant starting material for the access to multifunctionalised pyrrolidine rings.[195]

Proline Hydroxylation

A number of processes exist for the chemical synthesis of t4HyP, all of which share the feature of being multi-step, requiring highly volatile and toxic solvents to ultimately produce low yields of product.[195] An example of such a reaction is the Takano synthetic process for the production of L-hydroxyproline from (*S*)-*O*-benzylglycidol (Figure 1.44).[198] This is a complex process requiring the use of extremely hazardous reagents and reaction conditions for a final product yield of just 28 percent.[195] Due to such common flaws, there are now two widely used alternative approaches for the synthesis of hydroxyprolines: collagen hydrolysis and microbial biotransformations.[176]
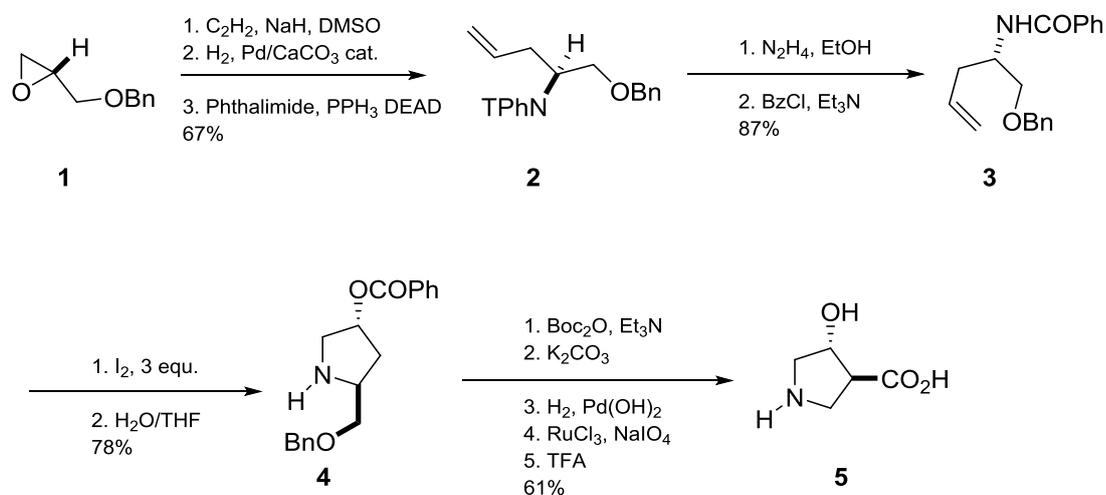
Figure 1.44: The 10-step Takano synthetic method for the production of trans-4-hydroxy-L-proline (5) using (*S*)-*O*-benzylglycidol (1) as a starting material. The percentages at the bottom of each arrow represent the product yield for those steps.

Collagen hydrolysis is a non-enzymatic process working on the principle of conducting acidic (or alkaline) hydrolysis of collagen to produce *trans*-4-hydroxy-L-proline and *trans*-3-hydroxy-L-proline (and *cis*-3- and *cis*-4-hydroxy-D-proline in trace amounts). These products are isolated from the complex hydrolytic broth following a series of ion exchange chromatography steps. This process exploits the fact that in mammalian systems, the

biosynthesis of collagen requires the hydroxylation of proline residues in procollagen with the assistance of 2-oxoglutarate–dependent *trans*-4 and *trans*-3 prolyl hydroxylases, mimicking this.[199] Despite its high specificity, this process cannot be applied for use with proline in isolation, with low product yields due to the requirement of many reaction and purification steps, the production of waste materials and the need for toxic components when extracting products.[176]

An attractive alternative to collagen hydrolysis and chemical synthesis is the use of bacterial enzymes as biocatalysts for the production of hydroxyproline.[50]

Microbial Proline Hydroxylases

Microbial screenings of various microorganisms identified 2-oxoglutarate-dependent hydroxylases capable of catalysing the conversion of free L-proline to all four possible regio- and stereoisomers of hydroxyl-L-proline (Figure 1.45).[194, 200] Of these, three have been successfully cloned and expressed in *Escherichia coli*: *cis*-3-, *cis*-4 and *trans*-4-proline hydroxylase (*cis*-P3H, *cis*-P4H and *trans*-P4H).[201] Unlike their mammalian counterparts which play a role in collagen biosynthesis, bacterial hydroxylases were found to be associated with the synthesis of host peptide antibiotics such as etamycin and telomycin.[196]
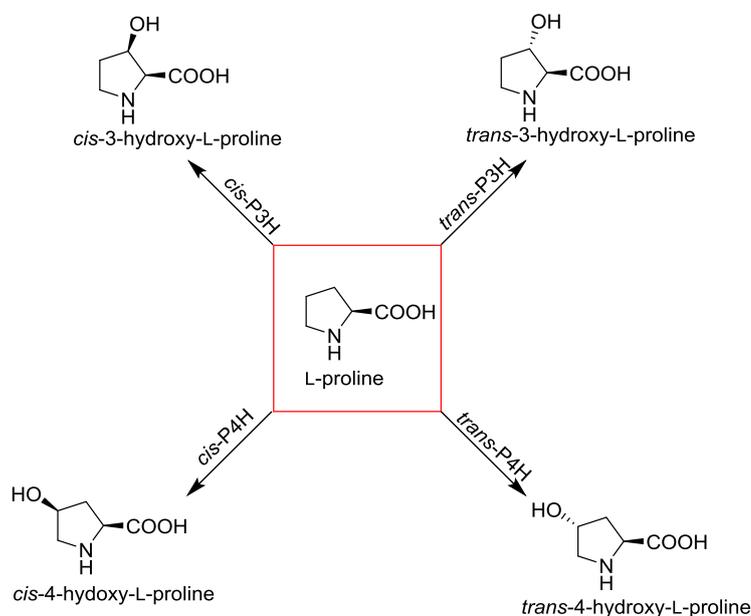
Figure 1.45: A scheme of the different proline hydroxylases found in nature. *cis*-P3H = *cis*-3-proline hydroxylase, *cis*-P4H = *cis*-4-proline hydroxylase, *trans*-P3H = *trans*-3-proline hydroxylase, *trans*-P4H = *trans*-4-proline hydroxylase. The source organisms of each enzyme were: *Steptomyces* sp. TH1 (*cis*-P3H), *Sinorhizobium meliloti* (*cis*-P4H), *Glarea lozoyensis* (*trans*-P3H) and *Dactylsporangium* sp. RH1 (*trans*-P4H). Figure adapted from Huttel *et al.* (2013).[176]

L-proline can be produced cost-efficiently by fermentation making its enzymatic conversion to hydroxy-L-proline an economical process.[202] Unlike prolyl hydroxylases, microbial proline hydroxylases hold the attractive feature of being able to catalyse the conversion of free L-proline, with these enzymes the same mechanism as other members in the 2OGDO superfamily (Figure 1.46).[202] Shibasaki and colleagues recombinantly overexpressed a *trans*-P4H gene cloned from *Dactylsporangium* sp. RH1 (DOGDH) in *E. coli* and monitored the enzymatic production of t4HyP. They observed that the enzymatic conversion was 1600-fold higher than in the native organism and that 2-OG and ferrous iron were essential for catalysis.[202] Similar findings were also observed by Mori *et al.* with a *cis*-P3H from *Streptomyces* sp. TH1 (StP3H) and Hara *et al.* for *cis*-P4Hs from *Mesorhizobium loti* and *Sinorhizobium meliloti* (MlC4H, SmC4H).[196, 201, 203]
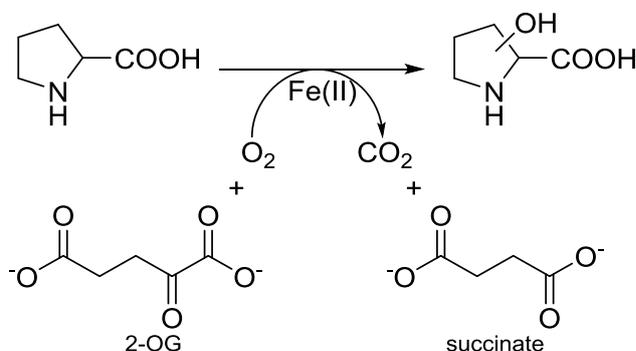
Figure 1.46: General hydroxylation of proline to hydroxyproline by a proline hydroxylase. A molecule is hydroxylated *via* a Fe^{2+} dependent mechanism requiring molecular oxygen, with 2-OG reduced to succinate.

Structural Studies of Microbial Proline Hydroxylases

The majority of structural studies of 2OGDOs have focussed on those originating from plants and animals with very few microbial structures available.[175] This is primarily due to a number of difficulties associated with isolating and crystallising microbial hydroxylases (these will be further discussed in Chapter 5).

Clifton and colleagues solved one of the first 2-OG-dependent proline hydroxylase structures – that of StP3H.[204] StP3H is a dimer in which the main chain folds into two domains: an N-terminal domain containing 10 β strands and a C-terminal helical domain. Figure 1.47 presents one of the monomers in the StP3H structure. The N-terminal domain contains a distorted jelly roll β-sheet core made up of 8 β strands. A proline-rich linker region connects the N and C-terminal domains, with the latter flanking one face of the jelly

roll. Another face of the jelly roll is flanked by an N-terminal region which forms β sheet interactions with the roll to extend the β sheet core.
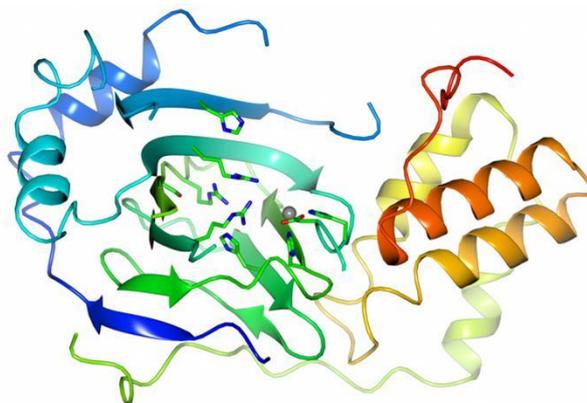
Figure 1.47: Render of a monomer of StP3H (PDB accession no. 1E5S). Dark blue = N-terminal region, Red = C-terminal and Light green = active site.

The active site of StP3H (Figure 1.48) was modelled following the co-crystallisation of monomer A of the enzyme dimer with iron (II) sulphate.[204] In the active site, the ferrous iron (Fe(II)) is ligated by a facial triad of His107, Asp109 and His158, while Arg168, Ser170 and His135 are likely involved in binding 2OG.[204] The remaining residues: Arg95, Arg97, Arg122 and His43, likely bind the carboxylate group of the proline substrate. In red in the structure (Figure 1.48) is an acidic disordered loop, which contains six sequential acidic residues that may bind to the imino group of proline.[204] Clifton *et al.* also hypothesised that Fe(II) is coordinated in a hexadentate manner to the facial triad and three water molecules.[204]
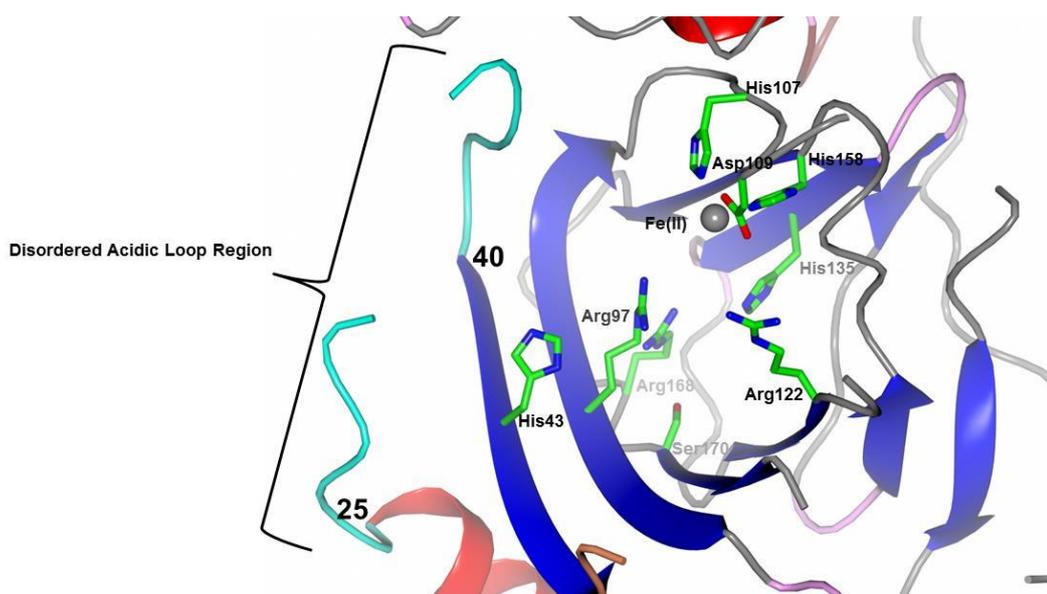
Figure 1.48: View of the StP3H active site. Dark blue = β barrel core, grey = iron(II, cyan = acidic disordered loop. PDB accession no: 1E5S.

Koketsu and coworkers recently solved the structure of the *cis*-4-proline hydroxylase from *Mesorhizobium loti* (MlC4H), presenting both an apo- and cobalt-bound structure.[205] The overall structure (Figure 1.49A) was highly similar to that of StP3H discussed above; however, this work was the first instance in which a structure of proline hydroxylase in complex with cobalt (II) (as iron mimetic), 2-OG cofactor and the L-proline within the active site was presented (the StP3H structure did not contain the latter two).[205]
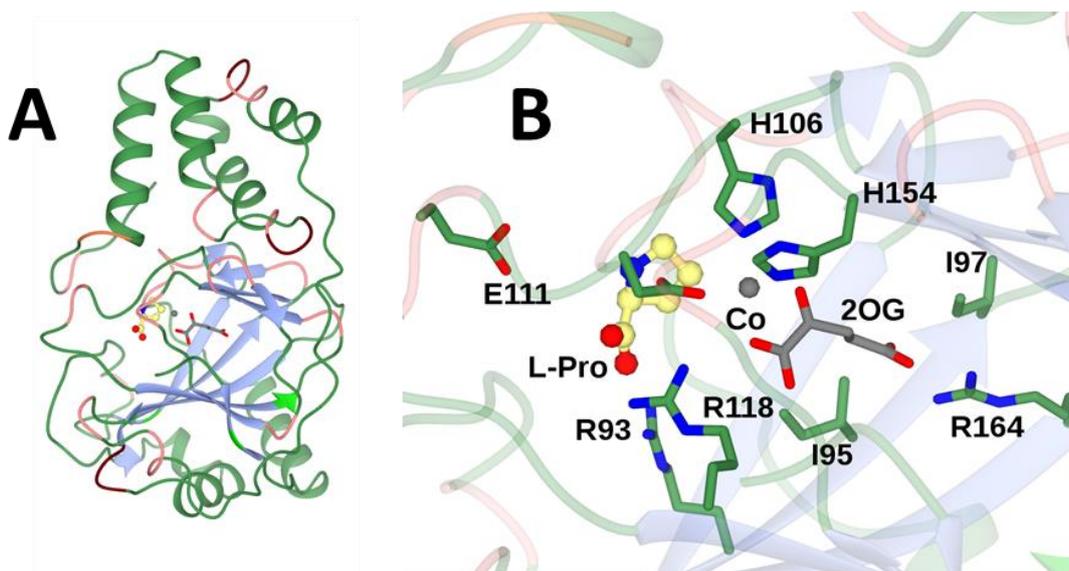
Figure 1.49: A: MlC4H monomer. The active site is located in the centre of the structure and encased by 8 β-sheets (blue) which are themselves surrounded by α-helices (green). B: Close up of MlC4H active site. Co is represented by a grey sphere and is neighboured by 2OG and L-Pro to its right and left respectively. Amino acids of interest within the active site have been labelled.

The active site (Figure 1.49B) consisted of the previously observed distorted jelly roll β-sheet motif sandwiched between the N-terminal and C-terminal α-helical domains. The active site Co^{2+} was coordinated by a triad of His106, Asp108 and His154 comprising the conserved HXD/H motif; 2-OG forms a salt bridge with Arg164 while the keto acid-moiety of the cofactor likely coordinating with the active site Co^{2+}. The L-proline was located at the centre of jelly roll β-sheet core in close vicinity to the Co^{2+}, with the amino moiety appearing to form a salt bridge with Asp108 and a water-mediated hydrogen bond to Glu111. The carboxylate group of L-proline meanwhile forms salt-bridges with Arg93 and Arg118. The correct orientation of the *cis*-face of L-Pro relative to Co^{2+} in the active site rationalised the formation of the *cis*-4-hydroxy-L-proline product.[205]

Applications of Proline Hydroxylases

The applications of proline hydroxylases in biocatalytic processes is a significant area of research interest in both industry and academia. As mentioned above, the use of these enzymes grants access to an unprecedented number of valuable chiral building blocks for the organic synthesis of pharmaceuticals.[203]

2OG-dependent L-proline hydroxylases have been found to hydroxylate substrates other than L-proline.[194] Klein *et al.* observed a recombinantly-expressed *trans*-P4H that was capable of hydroxylating L-pipecolic acid to *trans*-5-hydroxy-L-pipecolic acid in 61 % yield.[194] Taking this idea further, Hara and colleagues very recently observed a *cis*-proline hydroxylase from *Streptosporangium roseum* capable of catalysing the simultaneous formation of *cis*-3-hydroxy-L-proline and *cis*-4-hydroxy-L-proline, forming the latter in greater quantities.[197] When L-pipecolic acid was the substrate, this enzyme produced both *cis*-3-hydroxy-L-pipecolinic acid and *cis*-5-hydroxy-L-pipecolinic acid.[197] Hara and colleagues next evaluated the substrate specificities of a number of *cis*-selective hydroxylases, finding that they were also capable of oxygenating 3,4-dehydro-L-proline, L-azetidine-2-carboxylic acid, *cis*-3-hydroxy-L-proline, and L-thioproline.[206]
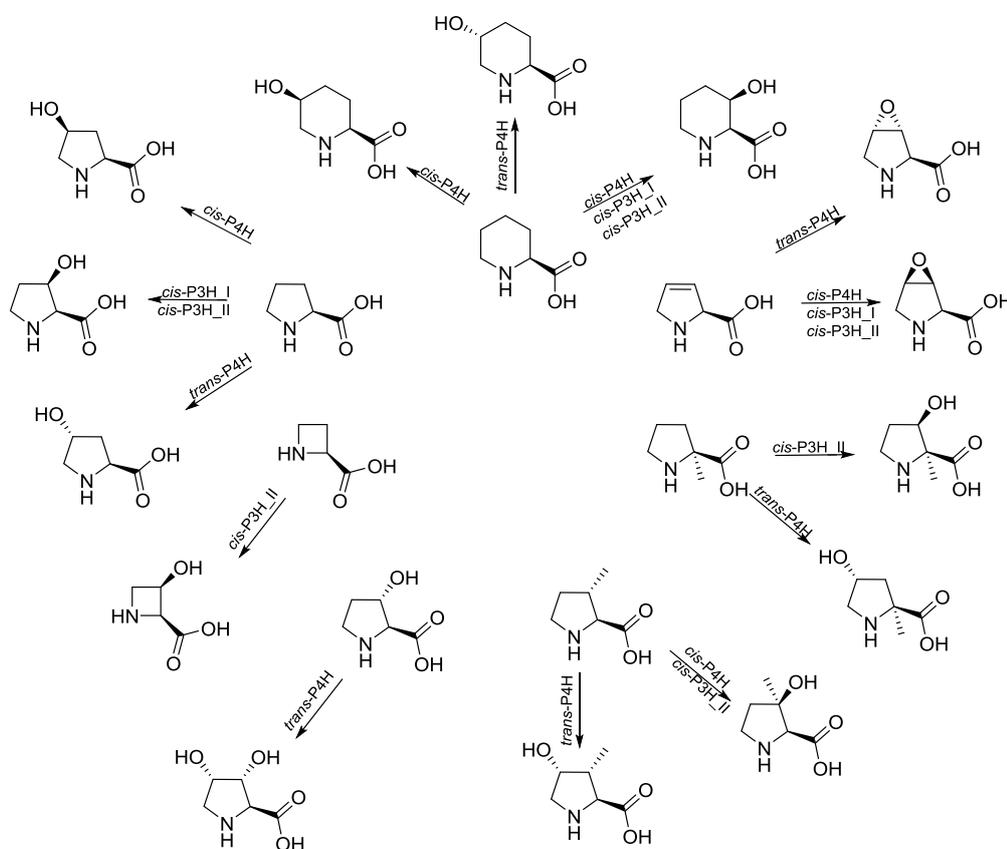
Figure 1.50: Scheme of hydroxylation of L-proline and commercially available congeners with 2OG-dependent proline hydroxylases. Figure adapted from Huttel (2013).[176]

These are only a few examples; many 2OG-dependent proline hydroxylases have very relaxed substrate specificity, catalysing a diverse range of reactions (Figure 1.50). Following investigations into the substrate spectrum of proline hydroxylases, it was found that, in a substrate, only the secondary amino acid moiety is essential for biocatalytic activity and must not be changed while changes in the hydrocarbon backbone can be tolerated.[176]

BMS-564929

BMS-564929 is a selective androgen receptor modulator (SARM) in development at Bristol-Myers Squibb (Figure 1.51).[207] SARMs hold pharmaceutical interest as alternatives to traditional anabolic drugs such as anabolic steroids, maintaining the same activity while avoiding side effects, such as liver damage, associated with the latter.[208] Various members of this drug class have also entered clinical trials as potential treatments for a range of disorders including muscle degradation due to HIV, sexual dysfunction and osteoporosis.[209]
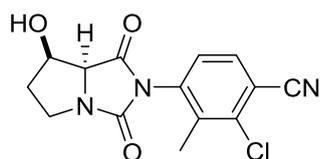
BMS-564929

Figure 1.51: BMS-564929, a potential selective androgen receptor modulator (SARM).

Johnston and colleagues presented a process in which a proline-3-hydroxylase from *Streptomyces* TH1 (P3H) can be used for the synthesis of a key intermediate for the production of BMS-549929 (Figure 1.52).[209]
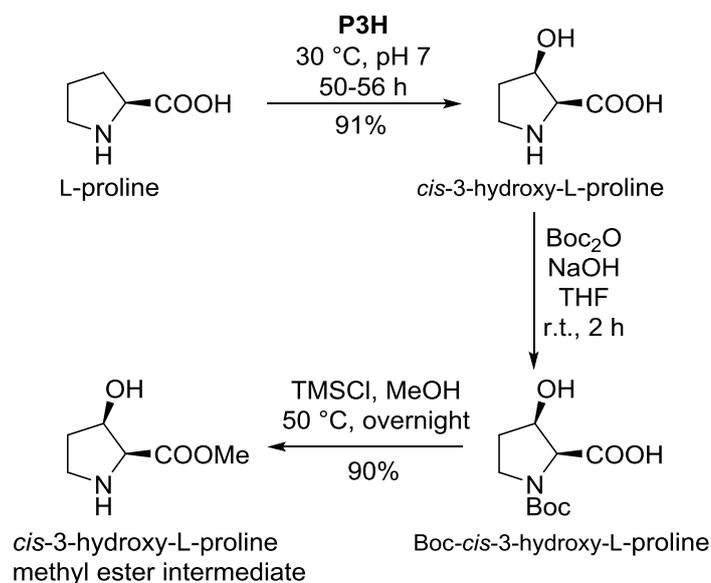
Figure 1.52: Proposed biosynthetic reaction for the production of a key methyl ester intermediate for use in synthesis of BMS-564929.

In the proposed reaction, P3H was expressed in *Escherichia coli*, produced on a fermenter-scale and used in a whole cell reaction process at a 10 L scale. L-proline was converted to *cis*-3-hydroxy-L-proline, with this product Boc-protected and subsequently deprotected back to an amine to result in the formation of methyl ester intermediate.[209]

Engineering of Proline Hydroxylases

It has only been in the last few years that mutagenesis and engineering work on microbial 2-oxoglutarate dependent oxygenases have started to come into fruition. This is likely due to a combination of increasing structural information, improved homology modelling, and enhanced strain and reaction media engineering methodologies among many other factors. An engineering example is described below.

Engineering of MlC4H for Refined Hydroxylation of L-pipecolic Acid

As mentioned above, Koketsu and colleagues managed to recently obtain the structure of *cis*-4-hydroxylase from *Mesorhizobium loti* (MlC4H) with bound substrate and cofactor.[205] This structure was used as a model to rationalise the mutagenesis of another, very closely related but difficult to crystallise *cis*-4-proline hydroxylase: a hydroxylase from *Sinorhizobium meliloti* (SmP4H). Previous work showed that SmP4H is also capable of hydroxylating L-pipecolic acid, producing *cis*-5-hydroxypipecolic acid (*cis*-5-HyPip) and *cis*-3-hydroxypipecolic acid (*cis*-3-HyPip) at an approximately 1:1 ratio.[194] The goal of this work

was to refine the regio- and stereoselectivity of the SmP4H such that only *cis*-5-HyPip was produced by it.[205]
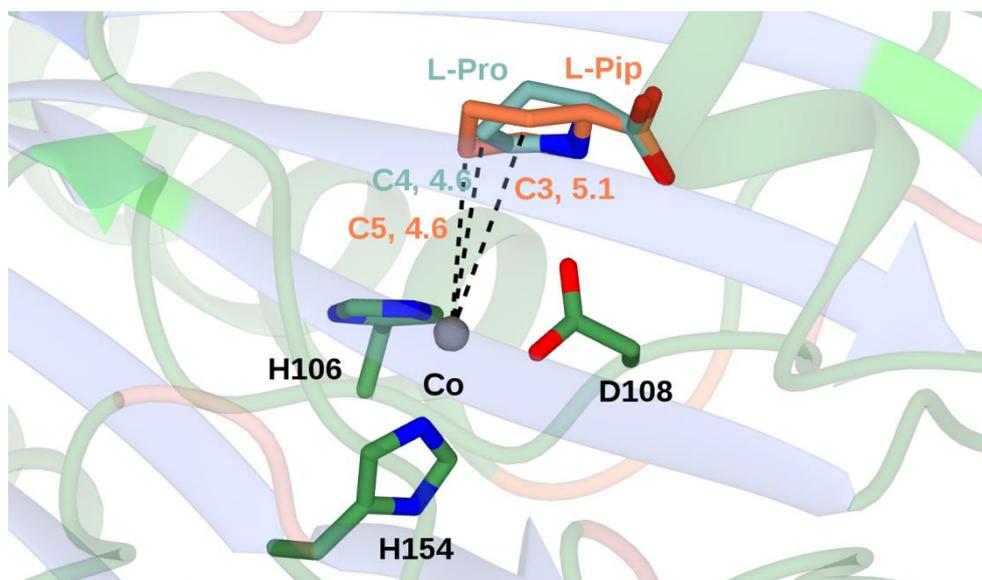
Figure 1.53: An overlay of MlC4H bound with L-Pro (blue) and L-Pip (orange) respectively. Co is represented by a grey sphere and the facial triad of H106, D108 and H154 have been labelled. The C4 carbon of L-pro is position 4.6 Å away from Co, while the C3 and C5 carbons of L-Pip are located 4.6 Å and 5.1 Å away respectively.

MlC4H was co-crystallised with L-proline and L-pipecolic acid respectively and the two structures directly compared (Figure 1.53). The complex showed that L-pipecolic acid was orientated with its C3 and C5 carbons in close vicinity to the active site-bound Co^{2+}, at distances of 4.6 and 5.1 Å respectively, showing that hydroxylation could readily occur at both positions. It was thus hypothesised that the hydroxylation ratio could be manipulated by changing the orientation and proximity of L-pipecolic acid within the active site by performing effective amino acid mutations. A triple Val95Trp/Val97Phe/Glu114Gly mutant of SmP4H ultimately resulted in an engineered enzyme that could selectivity hydroxylate L-pipecolic to exclusively produce *cis*-5-HyPip.[205]

1.5 Project Aims

2-Oxoglutarate-dependent oxygenases are highly versatile, ubiquitous enzymes that catalyse a diverse range of reactions with hydroxylations at the forefront. Proline hydroxylases are especially significant due to their ability to react with amino acids and derivatives in isolation with high regio-, stereo- and enantioselectivity, granting access to a wide pool of chiral synthons for organic synthesis.

The primary target of this project is a *trans*-4-proline hydroxylase for *Dactylsporangium* sp. (DOGDH). Work was also performed on two sub-targets: a *cis*-3-proline hydroxylase (StP3H) and *cis*-4-proline hydroxylase (MlC4H).

The aims of this work included the characterisation and optimisation of biotransformations of L-proline and congeners with the three hydroxylase targets. In order to do this, an analytical assay needed to be developed in order to assess and quantify these reactions. Prior to this work, DOGDH remained one of the few proline hydroxylases without a solved structure, it was therefore also an aim to pursue a structure in order to inform mutagenesis studies with the goal of engineering this enzyme for enhanced function and substrate scope.

In summary, the aims of this project were to:

1. Develop an analytical assay for monitoring biotransformations.
2. Clone, express and produce DOGDH, StP3H and MlC4H.
3. Pursue a structure of DOGDH using protein crystallography and related means.
4. Perform biotransformations of L-proline with hydroxylase targets at various scales and identify key parameters.
5. Optimise enzyme performance with L-proline as substrate by manipulating identified reaction conditions.
6. Screen enzymes with alternative substrates.
7. Perform mutagenesis of DOGDH to enhance performance and substrate scope.

Chapter 2

Materials and Methods

2.1 Materials

The synthetic genes for DOGDH, StP3H and MlC4H were ordered from and synthesised by Thermo Fischer Scientific Life Technologies. Cells used for cloning and expression were provided by Invitrogen. Unless stated otherwise, chemicals were ordered from Merck (formerly Sigma-Aldrich).

2.1.1 Bacterial Growth Media

Table 2.1: Lysogeny broth (LB) media composition

Component	Quantity per L Milli-Q H_2O
Tryptone	10 g
Yeast Extract	5 g
Sodium Chloride	10 g

Table 2.2: Terrific broth (TB) media composition

Component	Quantity per L Milli-Q H_2O
Tryptone	12 g
Yeast Extract	24 g
Glycerol	5 g (5% w/v)
Phosphate Buffer	100 mL (10% v/v)

Table 2.3: Composition of TB Phosphate buffer. This was prepared in Milli-Q H_2O.

Component	Quantity per L Milli-Q H_2O
KH_2PO_4	0.17 M
K_2HPO_4	0.72 M

2.1.2 Buffers for Protein Purification

These buffers were prepared all in Milli-Q H_2O and their pH was adjusted using 10 M HCl or 10 M NaCl.

Table 2.4: Compostion of Buffers utilised in the protein purification process.

Buffer	Compostion
Resuspension Buffer (A)	50 mM Tris-HCl, 500 mM NaCl, 10% v/v glycerol. pH 8.0
B	50 mM Tris-HCl, 500 mM NaCl. pH 8.0
C	50 mM Tris-HCl, 500 mM NaCl, 500 mM Imidazole, pH 8.0
Reductive Methylation (D)	50 mM Tris pH 8.0, 500 mM NaCl, 15% v/v glycerol, 10 mM β-mercaptoethanol. pH 8.0

2.1.3 Enzymatic Biotransformations

All reaction components were resuspended in 50 mM potassium phosphate (KP) buffer. A 1 M stock of this was prepared as in Table 2.5, with pH adjusted to 7.5 by adjusting the ratio of the phosphate components. This stock diluted in milli-Q H_2O to a final concentration of 50 mM prior to use.

Table 2.5: Preparation of 1 M potassium phosphate buffer pH 7.5

Component	Quantity / L^{-1} Milli-Q H_2O
1 M KH_2PO_4	13.6 g
1 M K_2HPO_4	17.4 g

80.2 mL 1 M KH2PO4 was added to 19.8 mL 1 M K2HPO4 to give a final 1 M buffer with pH 7.5

Table 2.6: Biotransformation Reaction Mix Version 1

Component	Quantity / L^{-1} 50 mM KP pH 7.5
L-proline (20 mM)	2.30 g
2-oxoglutaric acid (20 mM)	2.90 g
L-ascorbic acid (10 mM)	1.76 g
$Fe(II)SO_4$ (2 mM)	0.56 g

Components were resuspended in 50 mM KP and the pH readjusted to 7.5

Table 2.7: Biotransformation Reaction Mix II

Component	Quantity / L^{-1} 50 mM KP pH 7.5
L-proline (20 mM)	2.30 g
2-oxoglutaric acid (30 mM)	4.35 g
L-ascorbic acid (3 mM)	0.53 g
Fe(II)SO$_4$ (0.5 mM)	0.14 g

Components were resuspended in 50 mM KP and the pH readjusted to 7.5

2.1.4 High Performance Liquid Chromatography Buffers

Four High Performance Liquid Chromatography (HPLC) methods, each with their own unique columns and buffer systems, were utilised in this work.

Table 2.8: Columns utilised in HPLC and LC-MS analytical programs. The column stationary phases are highlighted in bold

Analysis System	Column
HPLC Programs 1 and 2	Phenomenex Aqua Aqua® 3 μm **C18** 125 Å, 150 x 2 mm
HPLC Program 3	Phenomenex Aqua Aqua® 3 μm **C18** 125 Å, 150 x 2 mm
HPLC Program 4	Waters XBridge **BEH C18**, 130Å, 3.5 μm, 4.6 mm x 150 mm
LC-MS Programs	Waters Acquity **C18 BEH** 130Å, 150 x 2.1 1.7 μm

Table 2.9: HPLC Program Buffer Systems

Analysis System	Buffer System Composition
HPLC Programs 1 and 2	Buffer A: Milli-Q water Buffer B: 100% Acetonitrile
HPLC Program 3	Buffer A: 20% acetonitrile/80% 50 mM sodium acetate pH 5 Buffer B: 80% acetonitrile/20% 50 mM sodium acetate pH 5
HPLC Program 4	Buffer A: 95:5:0.1 H$_2$O:Acetonitrile:Formic Acid Buffer B: Acetonitrile, 0.1% Formic Acid
LC-MS Programs	Buffer A: 95:5:0.1 H$_2$O:Acetonitrile:Formic Acid Buffer B: Acetonitrile, 0.1% Formic Acid

2.1.5 Electrophoresis and Western Blot Components

Agarose Gel Electrophoresis

Table 2.10: Components required to make Tris Acetate EDTA (TAE) Buffer (50x).

Component	Concentration
Tris	40 mM
Glacial Acetic Acid	20 mM
EDTA	1 mM

In order to prepare a 1% agarose gel, 0.6 g agarose was dissolved in 60 mL TAE Buffer (1x), and resuspended by heating in a microwave for 90 s.

Sodium Dodecyl Sulphate Polyacrylamide Gel Electrophoresis (SDS-PAGE)

Table 2.11: Recipe for 4x Running Buffer.

Component	Concentration
Tris-HCl	120 mM
Glycerol	20% w/v
SDS	4% w/v
β-mercaptoethanol	20% w/v
Bromophenol blue	0.04% w/v
Mixture was adjusted to pH 8.0 with concentrated HCl	

Table 2.12: SDS Running Buffer Recipe

Component	Concentration
Tris	25 mM
Glycine	200 mM
SDS	0.1% w/v

Table 2.13: Recipe for two gel quantities of the 12% resolving gel layer.

Component	Quantity
Deionised H_2O	3.2 mL
Resolving Gel Buffer (1.5 M Tris-HCl, 0.1% w/v SDS, pH 8.8)	2.5 mL
Acrylamide (30% w/v stock)	4.2 mL
APS (10% w/v stock)	50 µL
TEMED	8 µL

Table 2.14: Recipe for two gel quantities of the 12% stacking gel layer.

Component	Quantity
Deionised H_2O	3.2 mL
Stacking Gel Buffer (0.5 M Tris-HCl, 0.4% w/v SDS, pH 6.8)	1.3 mL
Acrylamide (30% w/v stock)	0.5 mL
APS (10% w/v stock)	35 µL
TEMED	8 µL

Table 2.15: Transfer Buffer Recipe (Western Blot)

Transfer Buffer pH 8.3	
Component	**Quantity**
Tris (25 mM)	3.03 g
Glycine (192 mM)	14.4 g
Methanol (20%)	200 mL
Milli-Q H_2O	To 1 L

Table 2.16: 10x TBST-Cl pH 7.5 Recipe

10x TBST-Cl pH 7.5	
Component	**Quantity**
Tris (100 mM, pH 7.5)	6.06 g
NaCl (1M)	29.2 g
Tween 20 (1%)	5 mL
Milli-Q H_2O	To 0.5 L

2.1.6 Primers

Table 2.17: Primers used in Cloning Experiments

Gene	Primer Name	Primer (5'-3')
DOGDH-pET22b	DOGDH_INF_FOR	GGAGATATACATATGATGCTGACCCCGACCGAAC
	DOGDH_INF_REV	GGTGGTGGTGCTCGAGCACCGGCTGCGCCAGC
DOGDH-SER	DOGDH-SER_FOR	CCAGGGACCAGCAATGCTGACCCCGACCGAACTGAAACAG
	DOGDH-SER_REV	GAGGAGAAGGCGCGTTAAACCGGCTGTGCCAGTGCAAAACCTG
MlC4H-LIC3C	MlC4H-LIC3C_FOR	CCAGGGACCAGCAATGACCACCCGTATTCTGGGTGTTGTTCAG
	MlC4H-LIC3C_REV	GAGGAGAAGGCGCGTTATTAATAGGTCATAACTTCACCGGCTGCACGATC
MP4H-LIC3C	MP4H-LIC3C_FOR	CCAGGGACCAGCAATGCTGAGCTATGAAAGCATTGATCTGTATCG
	MP4H-LIC3C_REV	GAGGAGAAGGCGCGTTATGCCAGCAGATCAGCACCCGGATC
NP4H-LIC3C	NP4H-LIC3C_FOR	CCAGGGACCAGCAATGCTGAGTGATGATCAGCTGGATACCTATTATGAAG
	NP4H-LIC3C_REV	GAGGAGAAGGCGCGTTAAAATTCATCGATCAGCGGTTCCAGTGC

2.2 Methods

2.2.1 Molecular Biology

2.2.1.1 Gene Synthesis

Genes of interest were designed using the GeneArt^{TM} tool available on Thermo Fisher Scientific Life Technology website. For each gene, it was strictly specified that codons were optimised for expression in *Escherichia coli*. Synthesis was performed by Thermo Fisher Scientific.

2.2.1.2 Primer Design

Oligonucleotide primers were designed using one of three tools depending on the purpose of these primers. For standard amplification or ligation independent cloning (LIC), primers were designed using HiTel Primer Design (http://bioltfws1.york.ac.uk/cgi-bin/primers.cgi). For mutagenesis, primers were designed using PrimerX (http://www.bioinformatics.org/primerx/). In the case of InFusion® cloning, primer design using the online Takara Clontech Infusion Design tool (http://www.clontech.com/US/Products/Cloning_and_Competent_Cells/Cloning_Resources/Online_In-Fusion_Tools).

2.2.1.3 Polymerase Chain Reaction (PCR)

Polymerase Chain Reaction (PCR) amplifications of the synthetic genes of interest were performed by first mixing the reaction components listed in Table 2.18 in a thin walled PCR tube.

Table 2.18: General PCR Reaction Mixture Components for Generic Polymerase.

Component (Stock)	Volume / μL
DNA Template (50 ng μL^{-1})	1
Forward Primer (20 pmol μL^{-1})	1
Reverse Primer (20 pmol μL^{-1})	1
dNTPs (2 mM)	5
MgSO$_4$ (25 mM)	3
Polymerase Buffer (10x)	5
DNA Polymerase	1
Milli-Q H$_2$O	33
Final Reaction Volume / μL	50

For each gene, the PCR amplification reaction was conducted in a BioEr LifeECO™ Thermal Cycler (alpha laboratories) using the programme in Table 2.19.

Table 2.19: Basic PCR Amplification Programme. Programme adapted from that recommend for KOD HotStart Polymerase.

Temperature / °C	Time	Step	Cycles
94	3 min	Initial Denature	1
94	30 s	Denature	
55	30 s	Anneal	35
72	1 min	Extend	
72	1 min	Final Extend	1
10	Hold	Cool	

2.2.1.4 Touchdown PCR

The PCR reaction components were prepared as described in Table 2.18. The Touchdown
PCR amplification was conducted in a thermocycler using the programme in Table 2.20.

**Table 2.20: Touchdown PCR Reaction Programme. This was adapted from Korbie et al.[210] using
KOD HotStart Polymerase Parameters.**

Phase 1	Step	Temperature / °C	Time
1	Denature	95	3 min
2	Denature	95	30 s
3	Anneal	72	45 s
4	Elongate	72	6 min

Repeat steps 2-4 (15 times), with the annealing temperature reduced by 1 °C each cycle.

Phase 2	Step	Temperature / °C	Time
5	Denature	95	30 s
6	Anneal	57	45 s
7	Elongate	72	6 min

Repeat steps 5-7 (20 times)

Termination	Step	Temperature / °C	Time
8	Elongate	72	5 min
9	Halt Reaction	4	15 min
10	Hold	10	-

2.2.1.5 Fragment Analysis by Agarose Gel Electrophoresis

PCR amplified samples were analysed by agarose gel electrophoresis with samples loaded
onto a 1% agarose gel (0.6 agarose in 60 mL 1 x TAE buffer with 1 μL added SYBRSafe),
and the gel run at 110 V for 50 min. The band corresponding to the amplified product was
excised with a scalpel and the DNA extracted using a GenElute™ Gel Extraction Kit
(Sigma).

2.2.1.6 Ligation Independent Cloning

For the purpose of purification, the amplified genes were cloned into the pET-YBSLIC3C (LIC3C) vector, which adds a cleavable N-terminal hexahistidine tag to the protein, by ligation independent cloning methods.

The LIC3C vector was linearised by reacting 50 μg vector DNA, 50 μL BseRI, 100 μL 10 x buffer (NEB 2) and Milli-Q water (to give a final volume of 1 mL) in a 1.5 mL Eppendorf tube at 37 °C for 110 minutes. The linearised product of this reaction was isolated by running it on a 1 % agarose gel, excising the fragment and extracting the DNA using a GenElute™ Gel Extraction Kit (Sigma).

For the purpose of the LIC annealing reaction, the linearised vector and amplified PCR product each underwent a T4 polymerase digest reaction in order to introduce complementary sticky ends to each.

Table 2.21: T4-treatment recipe for DNA Insert.

Component	Quantity
Extracted PCR product	0.2 pmol
T4 Buffer (10x)	2 μL
dATP (25 mM)	2 μL
DTT (100 mM)	0.5 μL
T4 DNA Polymerase	0.5 μL
Milli-Q H_2O	To 20 μL

Table 2.22: T4-treatment recipe for linearised vector.

Component	Quantity
Linearised pET-YSBLIC3C Vector	4.0 pmol
T4 Buffer (10x)	40 μL
dTTP (25 mM)	40 μL
DTT (100 mM)	20 μL
T4 DNA Polymerase	8.0 μL
Milli-Q H_2O	To 400 μL

The reactions were incubated at 20 °C for 30 min followed by 75 °C for 20 min.

Upon formation of the complementary sticky ends, the insert and pET-YSBLIC3C vector were annealed by reacting 4 μL insert with 2 μL vector for 10 min at room temperature. Following this, the reaction was quenched by addition of 1.5 μL EDTA (25 mM) and left to stand for a further 10 min.

The annealed product (2 μL) was transformed into 25 μL NovaBlue Singles™ Competent Cells with 30 μg mL^{-1} kanamycin as the antibiotic. A starter culture was prepared from one colony and grown overnight at 37 °C with shaking. The YSBLIC3C inserted gene product was extracted from this culture using a GenElute™ Plasmid Miniprep Kit.

A double digest of the annealing product was performed in order to determine the success of the process. This was done by mixing 7 μL of product with 1 μL Cutsmart Buffer, 1 μL NcoI and 1 μL NdeI in an Eppendorf tube, which was left to react at 37 °C for 2.5 h. The digested product was analysed by agarose gel electrophoresis.

2.2.1.7 In-Fusion® Cloning into pET22b Vector

Genes of interest were amplified by PCR and isolated by gel extraction. The circular pET22b vector was linearised by mixing the components in Table 2.23 and the reaction was incubated at 37 °C for 3 h. Following this, the linearised vector was isolated by running the sample on a 1% agarose gel, excising the fragment and extracting the product using a GenElute™ Gel Extraction Kit (Sigma).

Table 2.23: Reaction mixture for linearisation of circular pET22b vector.

Component	Quantity
Circular pET22b	2 μg
NdeI	2 μL
XhoI	2 μL
Cutsmart Buffer (10x)	5 μL
Milli-Q H$_2$O	As much as required to make up mix to 50 μL

The linearised pET22b and amplified gene insert were annealed by mixing the components in Table 2.24.

Table 2.24: In-Fusion® Annealing Reaction Mixture

Component	Quantity
Linearised pET22b	50 ng
Amplified Insert	50 ng
Infusion Mix	1 μL
Milli-Q H$_2$O	To 5 μL

This mixture was left to react at 50 °C for 15 min. All 5 μL of the annealing product was transformed into 25 μL NovaBlue Singles™ Competent Cells with 100 μg/mL ampicillin as the antibiotic. Starter cultures were prepared using a single colony for each and the process as described in the LIC annealing reaction. DOGDH-pET22b was extracted using a GenElute™ Plasmid Miniprep Kit (Sigma).

A double digest of the annealing product was performed in order to determine the success of the process. This was done by mixing 7 μL of product with 1 μL Cutsmart Buffer, 1 μL XhoI and 1 μL NdeI in an Eppendorf tube, which was left to react at 37 °C for 2.5 h. The digested product was analysed by agarose gel electrophoresis.

2.2.1.8 Sequencing

DNA sequencing was conducted in order to confirm if the gene encoding the protein of interest had been successfully inserted into the desired plasmid vector. Sequencing was performed externally by GATC Biotech.

2.2.2 General Procedures

2.2.2.1 Bacterial Transformation:

Cloning strain: NovaBlue Singles™ Competent Cells (Merck)

Expression strains: BL21(DE3) Competent cells (NEB), B834(DE3) Competent cells (Merck)

Antibiotic stocks: Kanamycin (30 mg mL^{-1}), ampicillin (100 mg mL^{-1})

General Protocol:

1 μL of plasmid was added to an Eppendorf tube containing 25 μL of cells, and left on ice for 30 min. This mixture was heat-shocked at 42 °C for 45 s and placed back on ice for 5 min. The tube was topped up with 1 mL of autoclaved LB and incubated at 37 °C for 1 h with shaking at 180 RPM. The sample was next centrifuged at 2800 g for 5 min to form a pellet of cells. 800 μL of supernatant was discarded and the pellet resuspended. 100 μL of cells were transferred to an agar plate containing respective antibiotic and gently spread with a spreader under aseptic condition. Plates were incubated overnight at 37 °C.

2.2.2.2 Preparation of Plates

Solid LB agar was microwaved until melted. Once cool to the touch, antibiotic was added in a 1:1000 dilution relative to the total volume of agar. This mixture was swirled and distributed between plates under aseptic conditions. Plates were left at room temperature in order for the agar to solidify.

2.2.2.3 Small-Scale Cultures

Under aseptic conditions, LB was transferred to a 50 mL falcon tube and antibiotic added in a 1:1000 dilution. A bacterial colony was selected from the prepared plates using a pipette tip and ejected into the LB-antibiotic solution. The tube was sealed and the mixture grown overnight at 37 °C in an orbital shaker with shaking set to 180 RPM.

2.2.3 Protein Production

General: Proteins cloned into the YSBL-LIC3C had kanamycin resistance, while those cloned into pET22b had ampicillin resistance. As a result, the antibiotic used during protein production varied depending on the vector.

2.2.3.1 Expression Tests

E. coli expression strains were transformed with recombinant plasmids as in **2.2.2.1** and two small-scale cultures prepared as in **2.2.2.2**. These starter cultures (750 μL) were used to inoculate fresh LB (7.5 mL) containing antibiotic (1:1000 dilution). 5 cultures were prepared in this manner and grown at 37 °C in an orbital shaker set to 180 RPM. The optical density at 600 nm (OD_{600}) of the cultures was monitored every 20 min until it had reached 0.8, after which three of the cultures were induced with 1 mM Isopropyl β-D-1-thiogalactopyranoside (IPTG). The induced cultures were grown overnight with shaking (180 RPM) at 16 °C, 30 °C and 37 °C respectively. The two un-induced cultured served as controls and were grown at 16 °C and 37 °C.

Cells were harvested by centrifugation at 2300 g for 1 min in a Progen Genfuge 24D Digital Microcentrifuge. This was repeated until all the cultures had been harvested. Cell pellets were formed and each resuspended 500 μL ddH$_2$O. These resuspensions were subjected to cell disruption using an ultrasonicator set to 13.0 microns on ice for 2 x 30 s, with 45 s intervals. Following this, the sonicated resuspensions were centrifuged at 16300 g for 5 min.

The supernatant believed to contain soluble protein was separated from the pellets. The pellets were resuspended in 500 μL ddH$_2$O and represented the insoluble samples. 10 μL of sample was mixed with 10 μL 2x SDS loading dye (100 mM Tris-Cl (pH 6.8); 4% (w/v) SDS (sodium dodecyl sulfate; electrophoresis grade); 0.2% (w/v) bromophenol blue; 20% (v/v) glycerol; 200 mM DTT). Samples were analysed by SDS-polyacrylamide gel electrophoresis (SDS-PAGE) on a 12 % gel run at 200 V for 50 min. Gels were developed using Coomasie Blue.

2.2.3.2 Large-Scale Protein Production

Bacterial transformations and starter cultures were performed as above. On a larger scale, each 7.5 mL (5 mL in case of TB media preparation) starter culture was used to inoculate 750 mL LB (or 500 mL TB) (with antibiotic) and the cultures grown to an optical density of 0.8 (1.2 in the case of TB growth). Following induction with 1 mM IPTG, cultures were grown overnight at 16 °C. Cells were harvested the following day by centrifugation at 5000 g for 20 min in a Sorval RC5B Plus centrifuge using a GS3 rotor. Following this, each resultant pellet was resuspended in 15 mL Buffer A (50 mM Tris-HCl, 500 mM NaCl, 10% (v/v) glycerol, pH 8.0) and the resuspensions pooled. Cells were lysed by sonication as described above but for 12 cycles of 30 s on/off intervals. Soluble fractions were collected by centrifuging the lysed cell mixture at 26892 g in a Sorval SS34 rotor for 60 min.

2.2.4 Protein Purification

2.2.4.1 Ni^{2+}-Affinity Purification

Supernatants from *2.2.3.2* were filtered using 0.45 μm filter unit and loaded onto a His-Trap^{TM} FF Crude column equilibrated with 5 column volumes 0.1 M NiSO$_4$ and 5 column volumes of Buffer B (50 mM Tris-HCl, 500 mM NaCl, pH 8.0).

The column was attached to an AKTA system (GE Healthcare Life Sciences) and Ni-affinity purification conducted by applying a gradient of Buffer B (50 mM Tris-HCl, 500 mM NaCl, pH 8.0) to Buffer C (50 mM Tris-HCl, 500 mM NaCl, 500 mM Imidazole, pH 8.0) over 100 mL (Flow Rate: 2.5 mL min^{-1}) (Table 2.25).

Table 2.25: Programme for Ni^{2+}-affinity Chromatography Purification

Step	Description	Buffer	Duration / CV
1	Initial Wash	100% Buffer B	5
2	50 mM Imidazole Wash	90% Buffer B: 10% Buffer C	7
3	Elution and Fractionation	Increase to 100% Buffer C	20
4	500 mM Imidazole Wash	100% Buffer C	5

CV = column volumes; 1 CV = 5.027 mL; Flow Rate = 2.5 mL min^{-1}

Samples of interest were analysed by SDS-PAGE and pooled together.

2.2.4.2 Buffer Exchange

Pooled fractions were subjected to buffer exchange by concentrating down to 2 mL in a VIVASPIN 20 column with 30 KDa cut-off, topped up to the starting volume with Buffer B and concentrated once more to 2 mL.

2.2.4.3 Size Exclusion Chromatography

The buffer exchanged protein was loaded onto a Hiload Superdex 75 16/600 GL column (GE Healthcare Life Sciences) equilibrated with Buffer B. Size exclusion chromatography (SEC) was performed using an AKTA System (GE Healthcare Life Sciences) using the programme described in Table 2.26.

Table 2.26: Programme for Size Exclusion Chromatography Purification

Step	Description	Buffer	Duration / CV
1	Isocratic Elution and Fractionation	100% Buffer B	1
CV = column volumes; 1 CV = 120.637 mL; Flow Rate = 1.0 mL min^{-1}			

Samples were analysed using SDS-PAGE and pooled.

2.2.5 His-tag Cleavage

Proteins fractions purified by Ni-affinity chromatography and SEC were pooled and diluted with Buffer B to a concentration of 1 mg mL^{-1}. 1 mg HRV 3C protease per 100 mg of protein was added to the diluted protein. The reaction was incubated at 4 °C for 16 h. The cleaved protein was purified using nickel-affinity chromatography using the programme in Table 2.27. The lack of His-tag on the protein should allow its isolation in Sample Wash step (2) prior to the application of an imidazole gradient in the Step 3.

Table 2.27: Ni-affinity purification for isolation His-Tag cleaved protein.

Step	Description	Buffer	Duration / CV
1	Sample Application	-	-
2	Sample Wash	100% Buffer B	5
3	Column Wash	Gradient Increase to 100% Buffer C	5
		Hold at 100% Buffer C	5
CV = column volumes; 1 CV = 5.027 mL; Flow Rate = 2.5 mL min^{-1}			

2.2.6 Reductive Methylation

The protein of interest was purified by Ni-affinity purification followed by SEC (with buffer containing 15% v/v glycerol, 10 mM β-mercaptoethanol, 50 mM Tris pH 8.0, 500 mM NaCl). The purified enzyme was concentrated to 10 mg mL^{-1} at a final volume of 1 mL. The reductive methylation reaction was commenced by adding 20 µL 1 M dimethylamine-borane complex (ABC) and 40 µL 1 M formaldehyde to the protein. The reaction solution was gently mixed at 4 °C for 2 h, after which ABC and formaldehyde were again added and solution mixed once more for 2 h. This was followed by a final addition of 10 µL ABC, after which the reaction was incubated overnight at 4 °C. The mixture was centrifuged in order to remove any resultant precipitate and the reaction quenched by addition of 5 mg glycine and 5 mg dithiothreitol (DTT) to the mixture. Finally, the protein was buffer exchanged into 50 mM Tris pH 8.0 and 500 mM NaCl by SEC.

2.2.7 Crystallisation

2.2.7.1 96-Well Sitting Drop Screening

Purified proteins obtained following SEC were concentrated to concentrations of between 10-50 mg mL^{-1} depending on the protein of interest. Concentrated proteins underwent crystallisation screening in 96 well MRC plates using three screens: PACT (Molecular Dimensions), INDEX (Hampton Research) and Clear Strategy Screen I and II (Molecular Dimensions). The screens were dispensed into the 96 well MRC plates using a HYDRA II 96-channel micro-dispenser liquid handling robot (Thermo Scientific). A mosquito liquid handling robot (TTP Labtech) was used to transfer 150 nL of protein stock and 150 nL of reservoir solution onto each well of the plate and the mixture left to incubate in a sitting well format. Plates were sealed and stored in the dark at 18 °C.

2.2.7.2 Optimisations in 24-well Plates

Screening drops deemed to hold potential were optimised on a larger scale in 24 well MRC plates. Conditions of interest were identified and optimised via addition of detergents, modification of pH and/or salt concentrations, or addition of additives. Components of the optimised mother liquor solution were mixed to give a final volume of 1 mL in their

respective wells. Varying quantities of protein (usually between 0.5-2 μL) were mixed with similarly varied amounts of the well condition on siliconized coverslips which were next placed over vacuum greased wells to create a seal for diffusion. Drops were left to incubate in a hanging drop format. Plates were stored in the dark at 18 °C.

2.2.8 Enzymatic Biotransformations

2.2.8.1 Generation of Cells

In order to perform whole cell biotransformations, it was necessary to first generate cells expressing hydroxylase targets. Cells were generated in 1 L fermenters using a Fed-Batch Fermentation protocol developed by N. Triggs and M. Bycroft at Dr Reddy's Technologies called 'Glucose Fed 5.0.'

In brief, this protocol involved three stages.

Stage 1: Preculture (50 mL)

2 x 50 mL starter cultures were prepared for each enzyme (containing: 50 mL LB; 30 μg mL^{-1} kanamycin; 1 colony from transformation plate) and grown at 37 °C with shaking at 250 RPM until $OD_{600} = 5\text{-}8$.

Stage 2: Seed culture (200 mL)

In a 1 L shake flask, an appropriate amount of preculture was added to 200 mL Vegetable peptone broth medium with 10 g L^{-1} glucose, such that starting OD_{600} was roughly 0.05. Two of these were prepared for each enzyme. These cultures were grown at 37 °C with shaking at 250 RPM until OD_{600} was between 6 and 8.

Stage 3: Fermentation (1 L)

To 500 mL starting volume of fermentation medium **GF4.0** (confidential, Dr Reddy's Laboratories Ltd) was added seed culture such that the starting OD_{600} was approximately 0.3. The temperature was set to 37 °C, dissolved oxygen maintained at 20% by manipulation of stirring rate (automatic) and pH kept at 7.2 by addition of 2 M phosphoric acid and 28% ammonium hydroxide (also automated). These conditions were maintained until OD_{600} was greater than or equal to 70, after which cells were induced with addition of

0.5 M IPTG and the growth temperature reduced to 25 °C. Cultures were then grown for 48 hours, after which they were harvested.

Two fermentations were performed for each enzyme.

Following fermentation, growth media was centrifuged at 5000 g for 30 min, the supernatant discarded and remaining cell paste was stored at -80 °C until required for biotransformations.

2.2.8.2 Whole Cell Biotransformations in Shake Flasks

Whole cell biotransformations of L-proline and L-pipecolic acid were performed using the cell paste expressing the respective hydroxylase targets produced by fermentations. Reactions were performed in both baffled and non-baffled shake flasks at 50 mL and 200 mL scales.

The reaction mixtures consisted of 10 % w/v cells and 200 mL (or 50 mL) reaction mixture I or II (see *2.1.3*), and left to react in a 2 L (or 250 mL) volumetric flask at 30 °C with shaking set to 180 RPM. 200 μL aliquots were taken from each reaction at specific time points and snap frozen in liquid nitrogen for analysis at a later point.

2.2.8.3 Whole Cell Biotransformations in MultiMax Reactors

Reactions were performed in MultiMax Reactors (Mettler Toledo) at 30 mL and 120 mL scales. Reactions mixtures were prepared as in *2.2.8.2*. Stirring rates were varied between 200 and 1000 RPM, while the reaction temperature was maintained at 30 °C. At the 120 mL scale, one of the vessels had a tube attached which served as an oxygen inlet, with oxygen flown into the system from a gas tap. 200 μL aliquots were taken from each reaction at specific time points and snap frozen in liquid nitrogen for analysis at a later point.

2.2.8.4 Reaction of L-proline with Purified Enzymes

The three hydroxylase targets were purified as standard and their stability confirmed by SDS-PAGE. A reaction mix stock containing 20 mM L-proline, 30 mM 2-oxoglutaric acid, 3 mM L-ascorbic acid and 2 mM Fe(II)SO$_4$ was prepared in 50 mM potassium phosphate buffer pH 7.5. Reactions were performed on a 4 mL scale in 50 mL falcon tubes, with this

volume consisting of 0.6 mg mL^{-1} enzyme and reaction mix, at 30 °C with shaking set to 180 RPM. Samples were taken at regular intervals for 24 h and snap frozen in liquid nitrogen.

2.2.9 Product Recovery

This protocol involved the protection of proline and hydroxyproline using Di-tert-butyl dicarbonate (Boc) followed by isolation via a series of extraction steps. Figure 2.1 presents a scheme of this reaction.
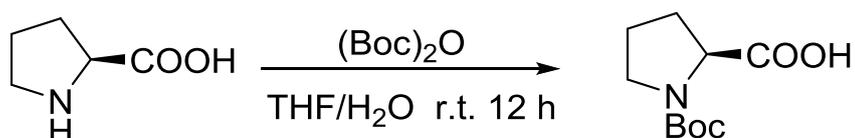
Figure 2.1:: Boc protection scheme of L-proline

Protocol:

Following reaction, cellular matter was centrifuged for 60 min at 10000 g and the supernatant retained. The supernatant was deactivated by heating at 90 °C for 10 min, after which, the precipitate was removed by filtration with celite. The filtered celite cake was washed with water and the wash combined with the filtrate. The pH of this solution was adjusted to 2.0 and the mixture left overnight at 4 °C. The resulting precipitate was filtered once more using celite, following by a wash and pooled. This solution was further cleaned up by performing an extraction with ethyl acetate in which the aqueous phase was retained. The pH of the aqueous phase was adjusted to 12.5 and 2 molar equivalents of Boc anhydride (dissolved in THF) relative to the starting concentration L-proline added to it. The reaction was left overnight with stirring. The reaction was quenched by adjusting the pH to 2.5 and the acidified solution extracted with ethyl acetate, with the organic phase retained.

2.2.10 NMR Analysis

1 mg of isolated sample was resuspended in 0.5 mL CDCl$_3$ and transferred into an NMR tube. ^{1}H and ^{13}C NMR were recorded on a Bruker Avance 400 spectrometer. Proton spectra were recorded at 400 MHz and carbon spectra were recorded at 100 MHz.

2.2.11 HPLC and LC-MS

2.2.11.1 HPLC Analysis

Samples were analysed by Reverse Phase-High Performance Liquid Chromatography (RP-HPLC) on an UltiMate™ Fully Integrated Micro-, Capillary and Nano HPLC System (LC Packings, Thermo Scientific) equipped with a Phenomenex Aqua 3u C18 125A column. The column temperature was set to 40 °C and solvent flow rate 200 μL min^{-1}. Analysis was performed at 254 nm.

The gradients and solvent systems employed were as follows:

Table 2.28: HPLC Program 1

HPLC Program 1		
Step	Time (min)	Buffer Gradient
1	0 - 6	10% - 16% Buffer B
2	6	16% - 100% Buffer B
3	6 – 10	100% Buffer B
4	10	100% - 10% Buffer B
5	10 - 15	10% Buffer B

Solvent System: Buffer A: Milli-Q H$_2$O; Buffer B: 100% Acetonitrile

Table 2.29: HPLC Program 2

HPLC Program 2		
Step	Time (min)	Buffer Gradient
1	0 – 5	10% - 40% Buffer B
2	5	40% - 100% Buffer B
3	5 – 15	100% Buffer B
4	15	100% - 10% Buffer B
5	15 - 25	10% Buffer B

Solvent System: Buffer A: Milli-Q H$_2$O; Buffer B: 100% Acetonitrile

Table 2.30: HPLC Program 3

	HPLC Program 3	
Step	Time (min)	Buffer Gradient
1	0 – 2	0% Buffer B
2	2 - 17	0% - 9% Buffer B
3	17 – 25	9% - 30% Buffer B
4	25 – 27	30% - 100% Buffer B
5	27 - 32	100% Buffer B
6	32 - 34	100% - 0% Buffer B
7	34 - 40	0% Buffer B

Solvent System: Buffer A: 20% acetonitrile/80 % 50 mM sodium acetate pH 5;
Buffer B: 80% acetonitrile/20 % 50 mM sodium acetate pH 5.

Analysis using HPLC Program 4 (Table 2.31) was performed on a Waters® 2695 HPLC Separation Module with a Waters XBridge C18 125 A Column. Flow rate was 600 μL min^{-1}, with samples analysed at 254 nm.

Table 2.31: HPLC Program 4

	HPLC Program 4	
Step	Time (min)	Buffer Gradient
1	0 – 1.5	0% Buffer B
2	1.5 - 14	0% - 95% Buffer B
3	14 -18	95 % Buffer B
4	18	100% Buffer B
5	18 - 21	100% Buffer B

Solvent System: Buffer A: 95:5:0.1 H_2O:Acetonitrile:Formic Acid
Buffer B: Acetonitrile, 0.1% Formic Acid

2.2.11.2 LC-MS Analysis

LC-MS analysis was performed using a Waters Acquity H-Class Bio UPLC-MS equipped with an Acquity C18 BEH 150 x 2.1 1.7 μm column. The column temperature was set to 40 °C.

The programs were as follows:

Table 2.32: LC-MS Long Acidic Program

LC-MS Long Acidic Program, Flow Rate = 300 μL min^{-1}		
Step	**Time (min)**	**Buffer Gradient**
1	0 – 1.5	0% Buffer B
2	1.5 - 7	0% - 95% Buffer B
3	7 -9	95 % Buffer B
4	9	0% Buffer B
5	9 – 10.5	0% Buffer B

Solvent System: Buffer A: 95:5:0.1 H_2O:Acetonitrile:Formic Acid
Buffer B: Acetonitrile, 0.1% Formic Acid

Table 2.33: LC-MS Short Acidic Program

LC-MS Short Acidic Program, Flow Rate = 600 μL min^{-1}		
Step	**Time (min)**	**Buffer Gradient**
1	0 – 0.3	0% Buffer B
2	0.3 – 1.5	0% - 95% Buffer B
3	1.5 – 2.4	95 % Buffer B
4	2.4	0% Buffer B
5	2.4 – 2.5	0% Buffer B

Solvent System: Buffer A: 95:5:0.1 H_2O:Acetonitrile:Formic Acid
Buffer B: Acetonitrile, 0.1% Formic Acid

2.2.12 Electrophoresis and Western Blotting

2.2.12.1 Agarose Gel Electrophoresis

Agarose gel electrophoresis was performed in order to analyse DNA fragments.1% agarose gels were prepared as described in *2.1.5*. DNA samples were prepared for analysis by mixing the reaction sample with 6x DNA Loading Dye (NEB) in a 1:6 ratio. A 1kb DNA ladder stock was also prepared in each case by mixing 1kb DNA ladder (NEB) with 4 x loading buffer in a 1:4 ratio.

Agarose gels were usually run for between 50 min and 1 h 30 min depending at 110 V depending on the purpose of the analysis. Follow the run, gels were visualised under UV light or with a Safe Imager (Invitrogen).

2.2.12.2 SDS-Polyacrylamide Gel Electrophoresis

SDS-PAGE was performed to analyse protein samples. The resolving and stacking gel layers for 12% SDS gels were prepared as described in Table 2.13 and Table 2.14. A 12% SDS gel was assembled by first pipetting the resolving layer into the assembly module and allowing this to solidify, after which the stacking layer was pipetted onto this and a well strip added to this.

Protein samples were prepared mixing samples with 2x SDS-loading dye in a 1:1 ratio, after which samples were denatured by heating at 95 °C for 5 min. Low molecular weight marker (Bio-Rad) was also prepared in a similar manner.

Gels were transferred into a running tank, running buffer (Table 2.12) was added to this, and the gel run for 70 min at 200 V. Following the run, gels were stained with Coomasie blue to visualise bands.

2.2.12.3 Western Blot

Protein samples were first run on a 12% SDS-PAGE gel as described above, however, the gel was not stained following completion of the run. The western blot was first thoroughly rinsed with water. One sheet of nitrocellulose was prepared by soaking the membrane with methanol, leaving for 2 mins, followed by equilibration in transfer buffer for 10 min. 8 sheets of Whatman™ 3MM paper was prepared by soaking in transfer buffer. The SDS gel was rinsed in water and transfer buffer. The apparatus was assembled in the following order: anode, 4 sheets of paper, membrane, gel, and 4 sheets of paper. The transfer process proceeded for 50 min at 25 V.

The blotting solutions (per membrane) were prepared by resuspending 2.5 g low fat milk powder (5%) in 50 mL TBST. 5 mL of this solution was aliquoted into a fresh tube to which was added 3 μL anti-His antibody and the tube left on ice until needed.

Following the transfer process, the apparatus was disassembled, the membrane rinsed with TBST and the gel stained with Ponceau Red. Following this stain, the ladder marker was visualised and marked with pencil, the blotting components were then incubated in milk for a few minutes and the solution discarded. The blot was blocked following addition of blocking solution (TBST + 5% Milk) for 1 hour at room temperature, rinsed twice with TBST, and incubated with the primary antibody for 1 hour at room temperature. Upon completion of incubation, the blot was washed with TBST three times for 10 minutes at a time, incubated with the secondary antibody for 30 min at room temperature and washed with TBST for 10 minute intervals three more times. The blot was finally visualised using a SigmaFast tablet.

Chapter 3
Assay Development

3.1 Introduction

The quantification of proline and derivatives in a mixture can be achieved by several methods, with each introducing a detectable element into the amino acid of interest. Chromatography methods are among the most widely used techniques for such analysis.[211]

Traditionally, the method of choice for the quantitative determination of amino acids in a mixture was ion exchange chromatography (IEC).[212] Moore and Stein dedicated a great deal of effort into developing such an approach, and in 1958, together with Spackman, described an amino acid analyser for the qualitative and quantification of amino acids in a protein mixture by performing IEC following postcolumn derivatisation with ninhydrin.[213-214] This seminal work awarded the pair with a Nobel Prize in 1972, and the amino acid analyser approach using IEC following derivatisation with ninhydrine or related compounds was widely used until the early 1990s.[212, 215] The described approach is frequently mentioned for its quantitative accuracy, reproducibility and large sample capacity, however, the requirement to perform more rapid analysis with smaller sample volumes resulted in the development of alternative analytical methods.[215]

Precolumn derivatisation of amino acids followed by the resolution of subsequent derivatives by reverse-phase high-performance liquid chromatography (RP-HPLC) is increasingly becoming the method of choice for quantitative amino acid analysis.[216] These derivatisation reactions introduce a covalently bound chromophore or fluorophore to the amino acid analytes of interest which serve the dual function of interacting with the apolar stationary phase in the pursuit of enhanced resolution as well as aiding in the photometric and fluorometric detection of samples.[216] These analytical procedures generally allow for rapid analysis and possess high sensitivity with detection limits reaching the femtomole scale.[215]

Table 3.1 summarises some of the commonly used derivatising agents. It has been suggested that the ideal derivatising agent should possess the capacity to rapidly and quantitatively react, under mild reaction conditions, with amino acid analytes without facing interference from matrix components such as salts, to ultimately yield a single, stable derivative for each respective amino acid of interest. The formed derivatives must be detectable with high sensitivity and the derivatising agent or its degradation products should not interfere with the chromatographic separation process.[217]

Table 3.1: Examples of compounds used for the derivatisation of amino acids.

Derivatisation Agent	Derivatisation Details
Marfey's Reagent	Marfey's Reagent (N_α-(2,4-Dinitro-5-fluorophenyl)-L-alanine amide) RP-HPLC with UV detection at 340 nm (Petersen *et al.*, 2003)[218]
NBD-Cl	NBD-Cl (4-Fluoro-7-nitrobenzofurazan) RP-HPLC with UV detection at 495 nm (Lindblad and Diegelmann, 1984)[219]
Dansyl chloride	Dansyl chloride (5-(Dimethylamino)naphthalene-1-sulphonlyl chloride) RP-HPLC with UV detection at 248 nm (Johnston *et al.*, 2009)[209]
FMOC-Cl	FMOC-Cl (9-Fluorenylmethyl chloroformate) RP-HPLC with UV detection at 254 nm (Koketsu *et al.*, 2014)[205]
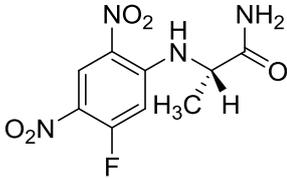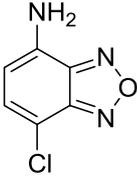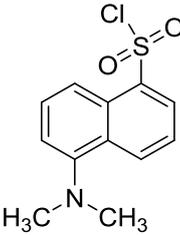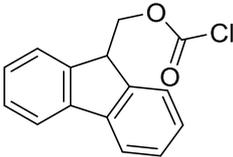
3.2 Aims

A precolumn derivatisation assay must be developed in order to analyse the activity hydroxylase targets with proline, derivatives and their hydroxylated equivalents. This assay should be suitable for application with RP-HPLC and LC-MS while also being safe, swift, reproducible and quantitative.

3.3 Marfey's Reagent

3.3.1 Overview

N_α-(2,4-Dinitro-5-fluorophenyl)-L-alanine amide (L-FDAA) (Figure 3.1) commonly known as Marfey's reagent, was first synthesised in high yield by Marfey *via* the reaction of 1,5-difluoro-2,4-dinitrobenzene (DFDNB) with L-alanine-NH_2.[220-221] L-FDAA, a pure chiral agent, is frequently used as a derivatising agent due to its ability to resolve and quantify specific enantiomers in racemic amino acids mixtures by forming both the D- and L-diastereoisomers, which can be separated under an achiral environment by chromatography due to differences in product retention factors.[211]
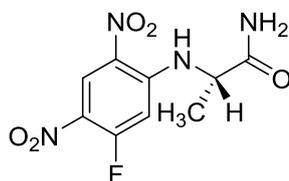
Marfey's Reagent

Figure 3.1: Structure of Marfey's Reagent (L-FDAA).

In the general derivatisation reaction (Figure 3.2), Marfey's reagent reacts with samples under alkaline conditions following the addition of base ($NaHCO_3$ in the illustrated example), with its reactive aromatic fluorine undergoing nucleophilic substitution with the free imino/amino group of the sample of interest.[222] This results in the formation of derivatised complex that can be detected by RP-HPLC at a wavelength of 254 nm (in the case of proline).
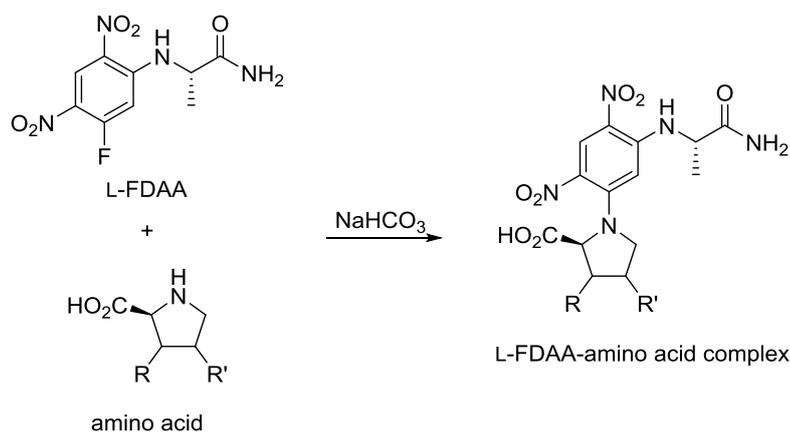
Figure 3.2: Reaction scheme of Marfey's Reagent derivatisation. In proline R and R' = H, 3-hydroxyproline R= OH and R' = H, 4-hydroxyproline R= H, R' = OH

Marfey's reagent was chosen as the initial derivatising agent in this work due to its ability to derivatise 18 of the naturally occurring amino acids, including proline and its hydroxylated equivalents.[223] Further advantages of Marfey's reagent over many other derivatising agents include the possibility to carry out chromatography on any HPLC system without column heating, its general simplicity and stable amino acid derivatives.[216]

3.3.2 Derivatisation Protocol

Standards of L-proline and *trans*-4-hydroxy-L-proline of varying concentrations were prepared by dissolving samples in 20% acetonitrile before derivatisation. 200 μL of amino acid standards were transferred into a fresh vial and to this were added 20 μL of 1 M NaHCO$_3$ and 80 μL of 36.7 mM L-FDAA (in acetone). The reaction vials were incubated in the dark at 40 °C. Derivatisation reactions were quenched by addition of 20 μL of 1 M HCl followed by 200 μL of a 50:50 v/v mixture of acetonitrile/H$_2$O. Prior to injection onto the HPLC apparatus, samples were briefly spun down to remove any particulate matter and filtered through a 0.45 μm membrane filter.

3.3.3 RP-HPLC Analysis

Standards derivatised as described above were analysed using by RP-HPLC using HPLC Program 1 detailed in Table 2.28 in section *2.2.11.1*. To summarise, this program employed a H$_2$O/acetonitrile gradient to elute detectable analytes based on polarity, with the more polar compounds eluted first.

For the purpose of assessing the effectiveness of this derivatisation assay and determining analyte retention times, standards of Marfey's Reagent (36.7 mM stock in Acetone), derivatised L-Proline (20 mM stock in 20% acetonitrile), derivatised *trans*-4-hydroxy-L-proline (20 mM stock in 20% acetonitrile) and a derivatised 1:1 mixture of L-proline and t4HyP were analysed using this method. Figure 3.3 presents the output chromatograms for these standards. These chromatograms provided qualitative evidence suggesting that the amino acid standards were being derivatised as peaks were observed for L-proline (5.4 min) and t4HyP (3.5 min) in their respective samples at a wavelength of 254 nm. Figure 3.3 summarises the retention times of these standards.
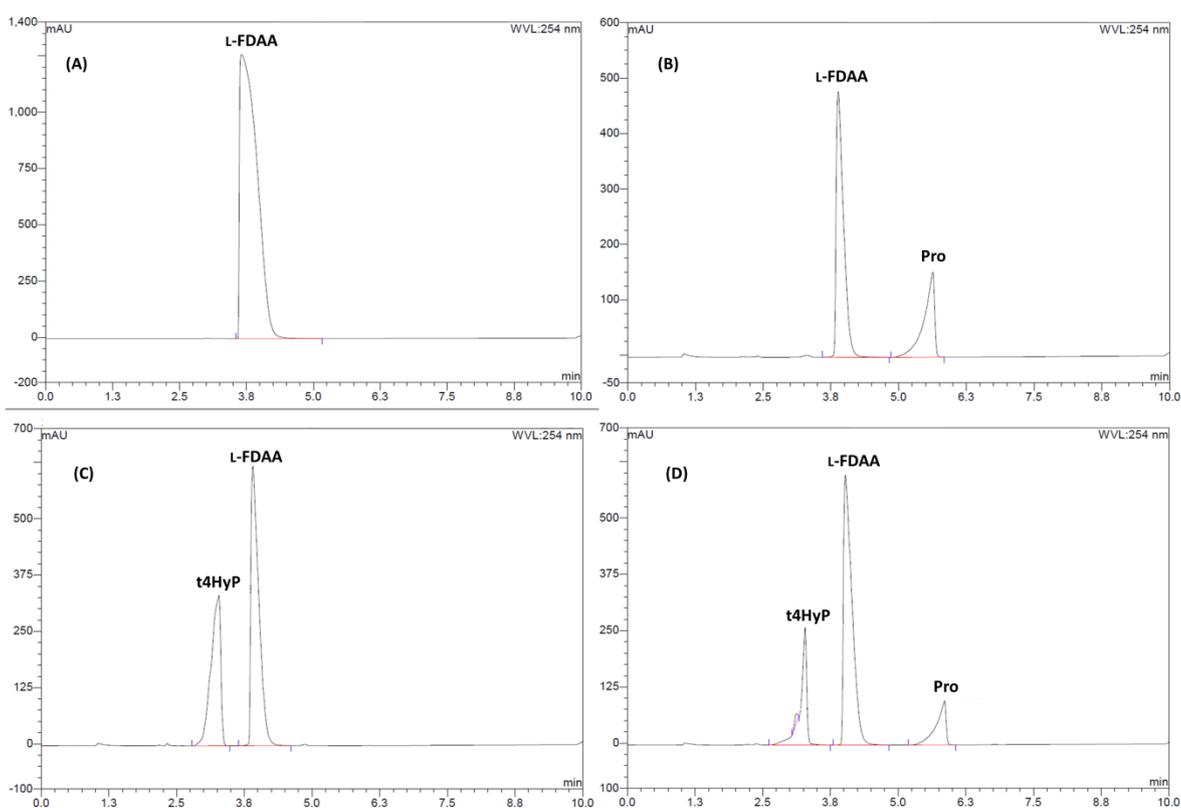
Figure 3.3: RP-HPLC chromatograms of four standards: (A) L-FDAA; (B) derivatised L-proline; (C) derivatised *trans*-4-hydroxy-L-proline (t4HyP); (D) derivatised 1:1 mixture of L-proline and t4HyP.

Table 3.2: Retention times for samples derivatised with Marfey's Reagent at UV = 254 nm

Sample	Approximate retention time (min)
Marfey's Reagent (L-FDAA)	4.1
L-proline	5.4
trans-4-hydroxy-L-proline	3.5

These retention times were consistently observed in repeated derivatisations of standards, providing confidence to proceed with Marfey's reagent as the derivatising agent. The robustness of this derivatisation reaction at varying concentrations of proline and t4HyP was next investigated to see if it was amenable for the analysis of biotransformation samples. L-proline and t4HyP standards at stock concentrations ranging from 5 mM to 100 mM were derivatised with Marfey's reagent and samples analysed as before.

Figure 3.4 presents the chromatograms obtained following derivatisation and analysis of varying stock concentrations of t4HyP (similar finding were observed for proline). The amplitude of t4HyP peak relative to the Marfey's reagent was found to increase with increasing stock concentration. This suggested that this derivatisation assay was both reproducible and robust enough to be used to analyse biotransformations at different reactant scales.
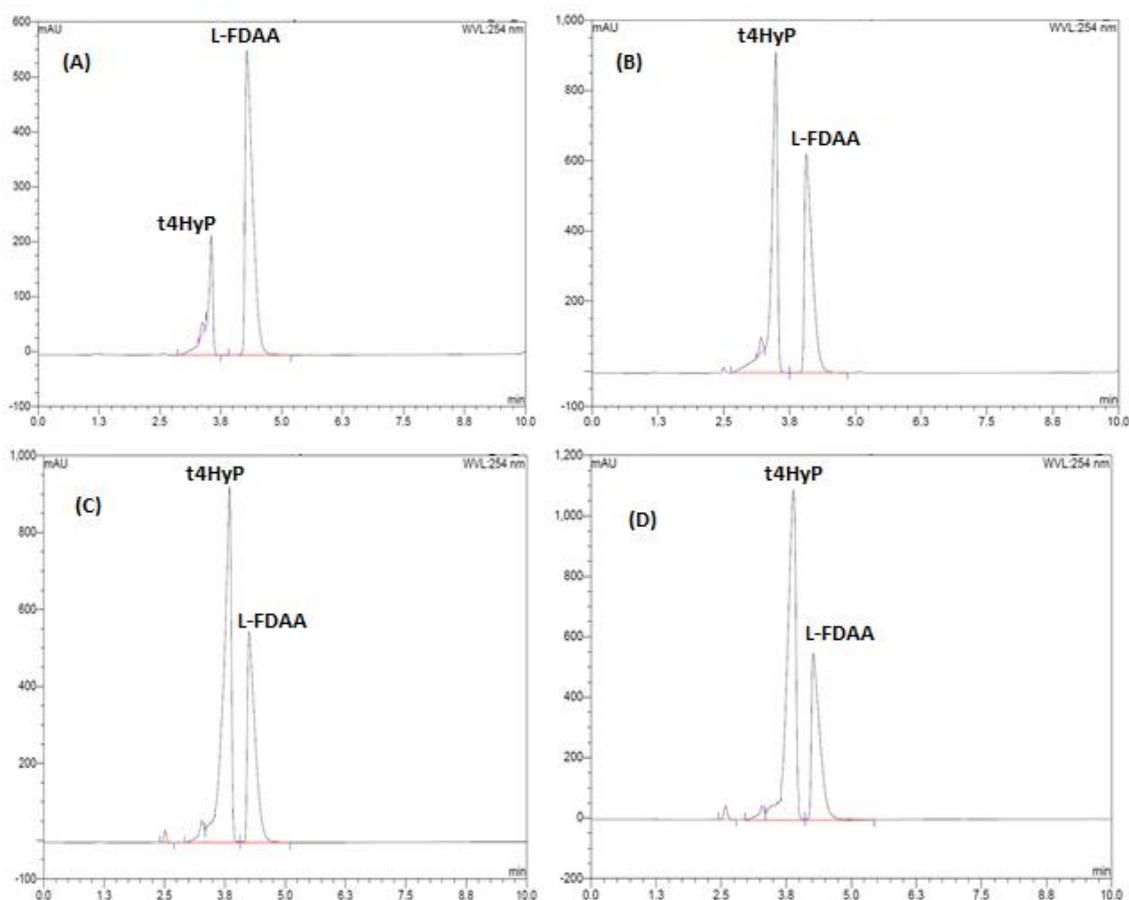
Figure 3.4: RP-HPLC chromatograms of t4HyP standards at varied stock concentrations derivatised with Marfey's reagent. (A) 5 mM t4HyP; (B) 30 mM t4HyP; (C) 60 mM t4HyP; (D) 100 mM t4HyP. The assumed peak identities have been stated. It can be seen that the amplitude of the t4HyP peak relative to L-FDAA increases with increasing stock concentration.

3.3.4 Assay Difficulties and Troubleshooting

Shortly after the generation of the above data, there was a sudden loss of reproducibility and consistency in this assay. The standards were no longer observed at the retention times reported above, and there was a substantial decrease in the peak intensities of any peaks observed.

Figure 3.5 presents two examples of the abnormal output chromatograms observed following attempted derivatisation of amino acid standards. It should first be noted that there was a substantial decrease in the intensities of the peaks observed as evident from the scale of the absorbance axes (max out at 60 mAU). Secondly, in the best cases the only familiar peak to be observed was that corresponding to Marfey's reagents at 4.1 min, suggesting the possibility that derivatisation was not occurring. Thirdly, unfamiliar peaks were observed in the void volume (less than 3 minutes in the run), this was hypothesised to have been the result of the acetonitrile concentration in the gradient being at too high of level, causing analytes to be swiftly eluted.
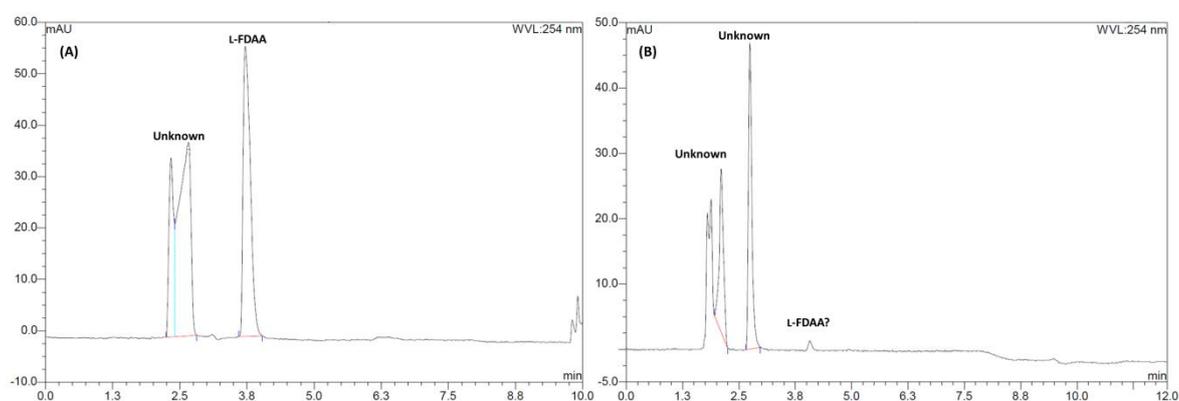
Figure 3.5: Examples of the abnormal chromatograms observed during assay troubleshooting efforts. (A) Proline sample derivatised with Marfey's reagent: there was only evidence of a Marfey's peak and a split unknown peak in the void. (B) A proline/t4HyP mixture derivatised with Marfey's reagent; a possible Marfey's peak is present together with three unknown peaks in the void volume. The poor peak intensities can also be noted based on the absorbance scale.

The hypothesised issue of sample loss in the void volume was confronted by designing a new HPLC analysis program (HPLC program 2 in Table 2.29 in section *2.2.11.1*) in which a milder acetonitrile gradient was steadily applied over a greater period of time; it was hoped that this would improve the separation of samples in the void. Unfortunately, this did not occur and the similar abnormal results continued to be observed. Following this, an

isocratic gradient of 40% acetonitrile followed by a 100% acetonitrile wash was applied with little success.

It was next hypothesised that the loss of consistency was a result of the HPLC apparatus. As a result, a significant overhaul (change in column, guard column, tubing and UV detector) was performed, yet this issue remained. LC-MS analysis of derivatised samples was next performed in order to assess the derivatisation itself, with results suggesting that the reaction was not proceeding (data not shown). This initially suggested that the chemicals employed were compromised, so fresh supplies were ordered from a different supplier. The issue unfortunately remained and as a result, it was concluded that the derivatisation was no longer occurring.

Troubleshooting of the reaction itself was performed. A preliminary hypothesis was that the issue may have resulted from loss of the derivatised complexes in either the pellet of particulate matter formed after spinning or loss during the filtration step. To assess this, the reactions were conducted under three conditions: i) without spinning and filtration, ii) without filtration and iii) without spinning. All three conditions yielded poor results.

It was finally hypothesised that the reaction may have not been proceeding due to reaction conditions not being sufficiently alkaline.[222] As a result, the derivatisation protocol was modified to incorporate boric acid buffer and sodium hydroxide, with the desire to keep conditions sufficiently basic. The modified derivatisation was as follows: 100 μL standard, 50 μL 0.5 M boric acid (pH 9.0), 9 μL 5 M NaOH and 50 μL 15 mM L-FDAA were added to an Eppendorf tube, and reacted for 1 h at 40 °C with shaking in the dark. The reaction was quenched with addition of 50 μL 1 M HCl, and the sample filtered into a GC vial prior to analysis. This too did not yield satisfactory results, with no evidence of derivatised products observed in the chromatograms.

The use of Marfey's reagent as a derivatisation agent was initially very promising. The precolumn derivatisation reaction was swift, sample resolution was high and the analysis runtime was short. However, the assay proved to be unreliable and unrepeatable, resulting in this derivatising agent not adhering to the ideal features discussed in the introduction of this chapter. It was thus deemed necessary to pursue alternative derivatisation methods.

3.4 NBD Chloride

3.4.1 Overview

4-Chloro-7-nitrobenzofurazan (NBD-Cl) was selected as a derivatisation reagent due to its frequent use as precolumn derivatisation agent for chromatography analysis in the literature, particularly by the Ozaki group.[196, 200-202, 224] NBD-Cl reacts with proline and its hydroxylated products to form analytes that are detectable at a wavelength of 495 nm.[219]
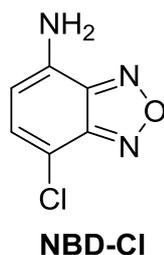
NBD-Cl

Figure 3.6: Structure of NBD-Cl

The general mechanism of the derivatisation reaction is outlined in Figure 3.7. NBD-Cl reacts with methanol and potassium tetraborate buffer in the reaction mixture to form NBD-OCH$_3$. Both NBC-Cl and NBD-OCH$_3$ can undergo nucleophilic substitution with the amino acid analyte of interest, with the latter reacting twice as fast.
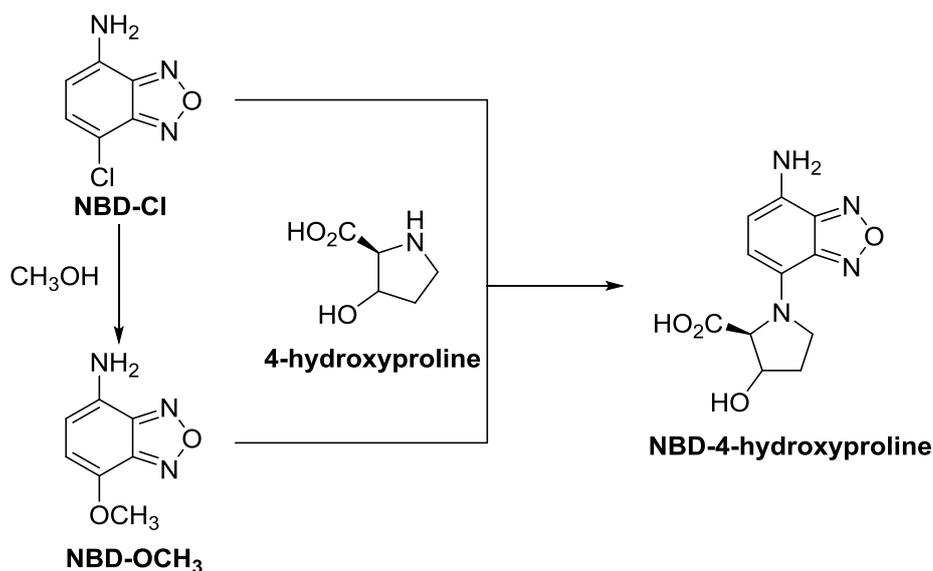
Figure 3.7: Derivatisation of hydroxyproline with NBD-Cl. Hydroxyproline can react with either NBC-Cl or NBD-OCH3, however, reaction with the latter proceeds twice as fast.

3.4.2 Derivatisation Protocol

1 mg of standard was dissolved in 500 μL Milli-Q water in an Eppendorf tube. To each tube was added 100 μL 0.4 mM potassium tetraborate buffer (pH 9.5) and 100 μL 200 mM NBD-Cl. The mixture was incubated in the dark at 37 °C for 20 min. The reaction was quenched with addition of 50 μL 1 M HCl and 150 μL 50 % methanol. Samples were spun down briefly prior to being filtered into GC vials for analysis.

3.4.3 RP-HPLC Analysis

Standards derivatised with NBD-Cl were analysed using HPLC programs 1 and 2 as well as the isocratic program described in *3.3.4*. The UV wavelength for analysis was set to 495 nm.
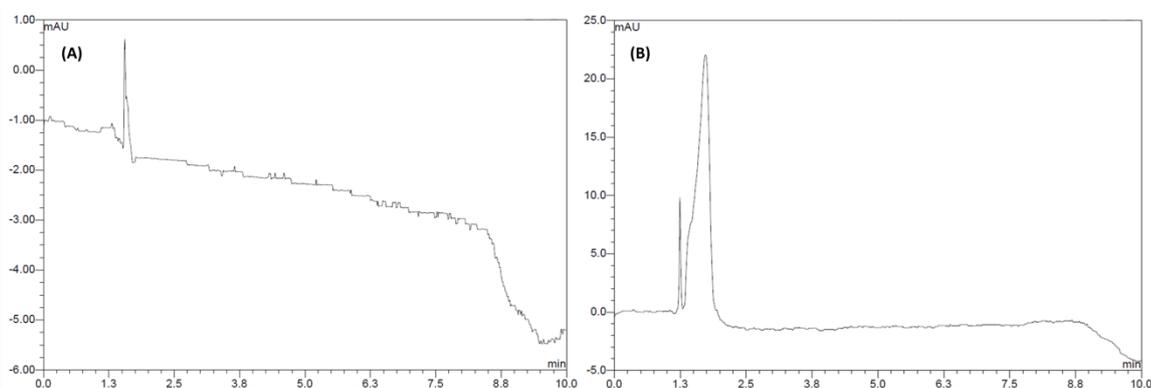
Figure 3.8: RP-HPLC chromatograms of: (A) NBD-Cl standard; (B) proline/t4HyP mixture reacted with NBD-Cl.

Figure 3.8 presents two chromatograms obtained during these trials. The first feature to note is the poor absorbance measurements in both as evident from the absorbance axes. The chromatogram for an NBD-Cl standard (Figure 3.8A) shows the absence of a peak with poor baseline, while that for a derivatised mixture of proline and t4HyP (Figure 3.8B) presents a single peak in the void volume. Such inconsistent and poor results were observed throughout the entirety of the work performed with NBD-Cl.

This lack of success despite considerable efforts with NBD-Cl as a derivatisation agent suggested that an alternative agent be pursued.

3.5 FMOC Chloride

3.5.1 Overview

9-Fluorenylmethyl chloroformate (FMOC-Cl) is argued to be a derivatisation reagent that meets most of the criteria discussed in the introduction of this chapter, having historically been observed to be detectable at high sensitivity while giving rise to a single, stable derivative per amino acid.[217] FMOC-Cl (Figure 3.9) itself serves as fluorescent label when bound to an analyte and was first used for the labelling of amino acids by Einarsson and coworkers.[225-226] The reagent rapidly reacts with both primary and secondary amino acids under mild conditions whilst itself being relatively insensitive to salts in the reaction mixture, forming highly fluorescent carbamate complexes (detection can occur at the picomole level) that are stable at room temperature and demonstrate excellent chromatographic behaviour on reverse-phase columns.[217]
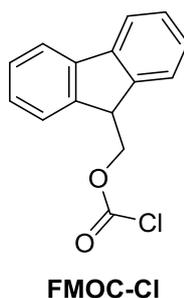
FMOC-Cl

Figure 3.9: Structure of FMOC-Cl

The general FMOC-Cl derivatisation reaction proceeds under alkaline conditions at 60 °C with vigorous shaking of samples and is followed by an acidic quench. The reaction suffers from the disadvantage of FMOC-Cl being reactive towards water, undergoing subsequent hydroxylation and decarboxylation to form a fluorescent alcohol, FMOC-OH, which elutes in the middle of the chromatogram. At high FMOC concentrations, this peak poses the risk of overlapping with other amino acids peaks, making their quantification cumbersome. As a result, it is very important the concentration of FMOC-Cl utilised in the reaction is carefully controlled.

The use of FMOC-Cl for the analysis of proline and derivatives in reaction mixtures in recent publications also contributed towards its selection as a potential derivatisation agent.[194, 205]

3.5.2 Derivatisation Method 1

The first derivatisation method to be tested was based on one utilised by Klein and coworkers.[194] Figure 3.10 outlines the general mechanism of this reaction. In brief, FMOC-Cl reacts with the amino acid in the presence of borate buffer. The reaction is quenched by the addition of 1-adamantadine (ADAM), which forms complexes with any excess FMOC-OH produced as a by-product of the reaction, theoretically eliminating them from the mixture.
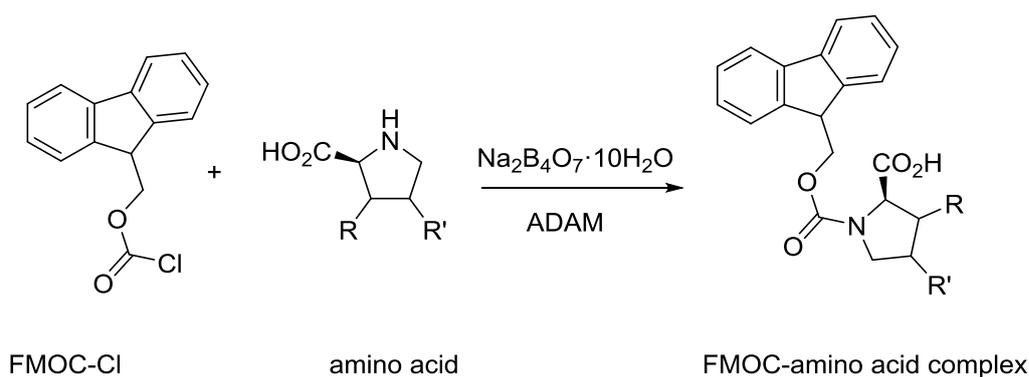
Figure 3.10: Derivatisation of an amino acid with FMOC-Cl using the Klein Method.

Protocol

The protocol involved adding 40 μL 10 mM amino acid standard, 40 μL 0.5 M sodium borate buffer, pH 7.7, and 80 μL 1.5 mM FMOC-Cl to an Eppendorf tube. The mixture was vortexed vigorously for 60 s, followed by the addition of 100 μL 40 mM ADAM and a final vortex for 45 s.

This derivatisation protocol required a fraction of the amount of time for derivatisation with L-FDAA allowing for the rapid analysis of a number of samples daily.

RP-HPLC Analysis

Following derivatisation, standards were analysed using HPLC-Program 3 (Table 2.26 in *2.2.11.1*). In order to determine the retention times of all the possible reaction components, standards of FMOC-Cl mixed with ADAM, derivatised proline, derivatised t4HyP and a derivatised pro/t4HyP mixture were analysed at 254 nm. The output chromatograms are shown in Figure 3.11.
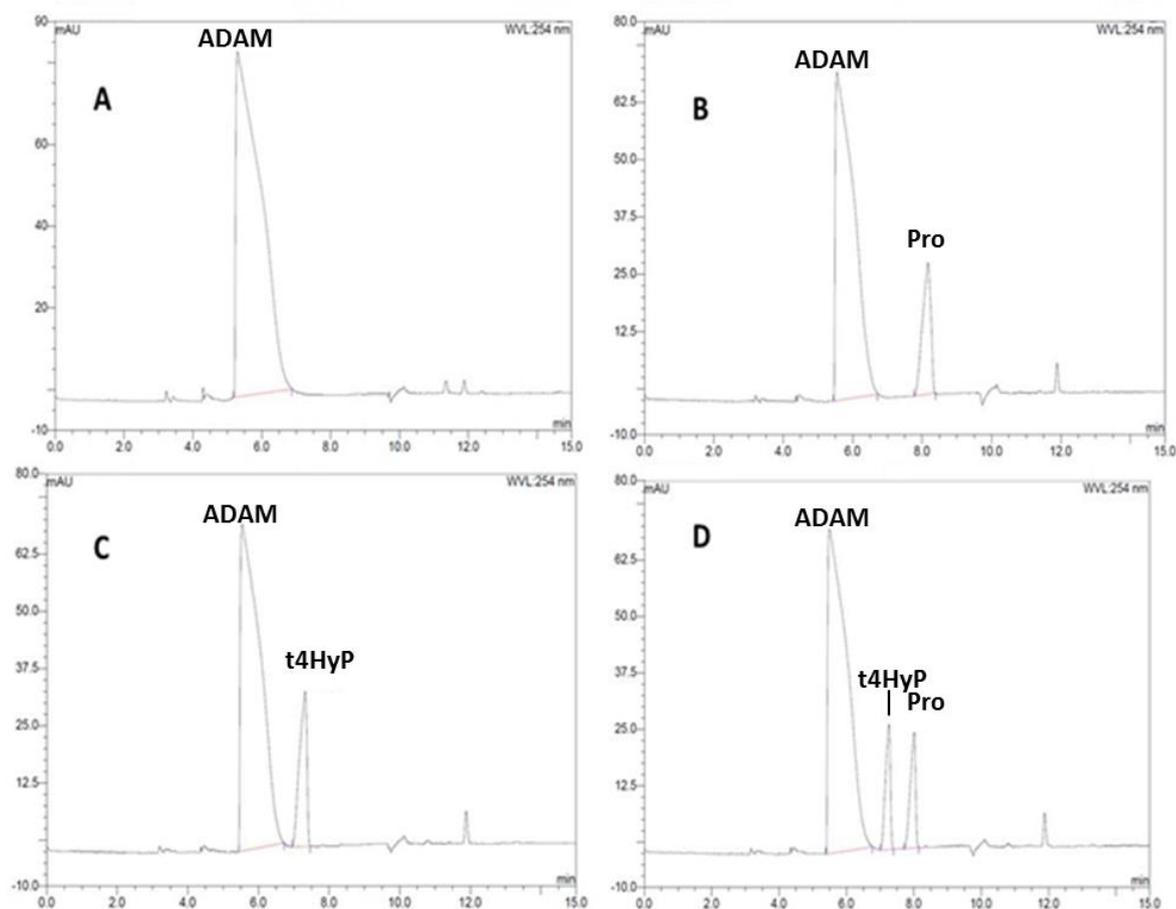
Figure 3.11: Chromatograms obtained following analysis of A: FMOC-ADAM, B: 10 mM L-proline, C: 10 mM trans-4-hydroxy-L-proline and D: 1:1 mixture of B and C. Well defined peaks can be seen in each chromatogram. Approximate retention times: FMOC-ADAM, 5.5 min; L-proline, 8 min; t4HyP, 7.3 min.

The output chromatograms suggested that the derivatisation was occurring, with FMOC-ADAM observed at 5.5 min, L-proline-FMOC at 8 min and t4HyP at 7.3 min. This data showed significant promise, however, it was important to note that proline and t4HyP eluted very close to each other which may have posed issues in future.

Though the derivatisation protocol was outwardly straightforward, a number of difficulties were encountered. Firstly, the reaction required 0.5 M sodium borate buffer at pH 7.7. Such a high concentration resulted in the buffer being very saturated causing the borate to precipitating out of solution as crystals. This required the buffer to be heated every time this phenomenon was observed and cooled to room temperature prior to adjusting the pH due to its temperature sensitivity. Second, ADAM was found to require heat in order to be resuspended and would very easily precipitate out of solution. Such sensitivity made it very difficult to keep reaction conditions consistent making reproducibility a challenge.

3.5.3 Derivatisation Method 2

Koketsu and colleagues recently published an alternative FMOC-Cl derivatisation protocol (Figure 3.12) which does not require ADAM or borate buffer at such a high concentration.[205] This method involved the use of two borate buffers; 100 mM sodium borate at pH 9 and 250 mM sodium borate at pH 5.5.
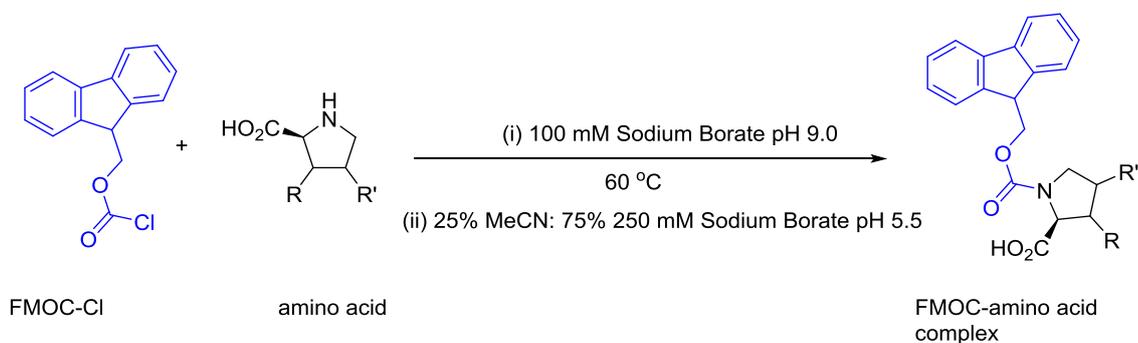
Figure 3.12: Koketsu _et al._ method of FMOC-Cl derivatisation.

Protocol

15 μL amino acid standard wass diluted with 285 μL 100 mM borate buffer pH 9.0 and 300 μL 5.6 mM FMOC-Cl added to this. The mixture was mixed vigorously and left to react for 40 min. Finally, the reaction was quenched by addition of 600 μL 25% acetonitrile: 250 mM borate buffer (pH 5.5).

RP-HPLC Analysis

Standards of FMOC-Cl, derivatised L-proline, derivatised t4HyP and a derivatised mixture of proline/t4HyP were analysed using HPLC program 4 as before with wavelength set to 254 nm.

Table 3.3 summarises the observed retention times for these standards.

Table 3.3: Retention times of FMOC-derivatised samples analysed by HPLC Program 4

Sample	Approximate retention time (min)
FMOC species	5.2 and 23.6
L-proline	16.5
trans-4-hydroxy-L-proline	13.1
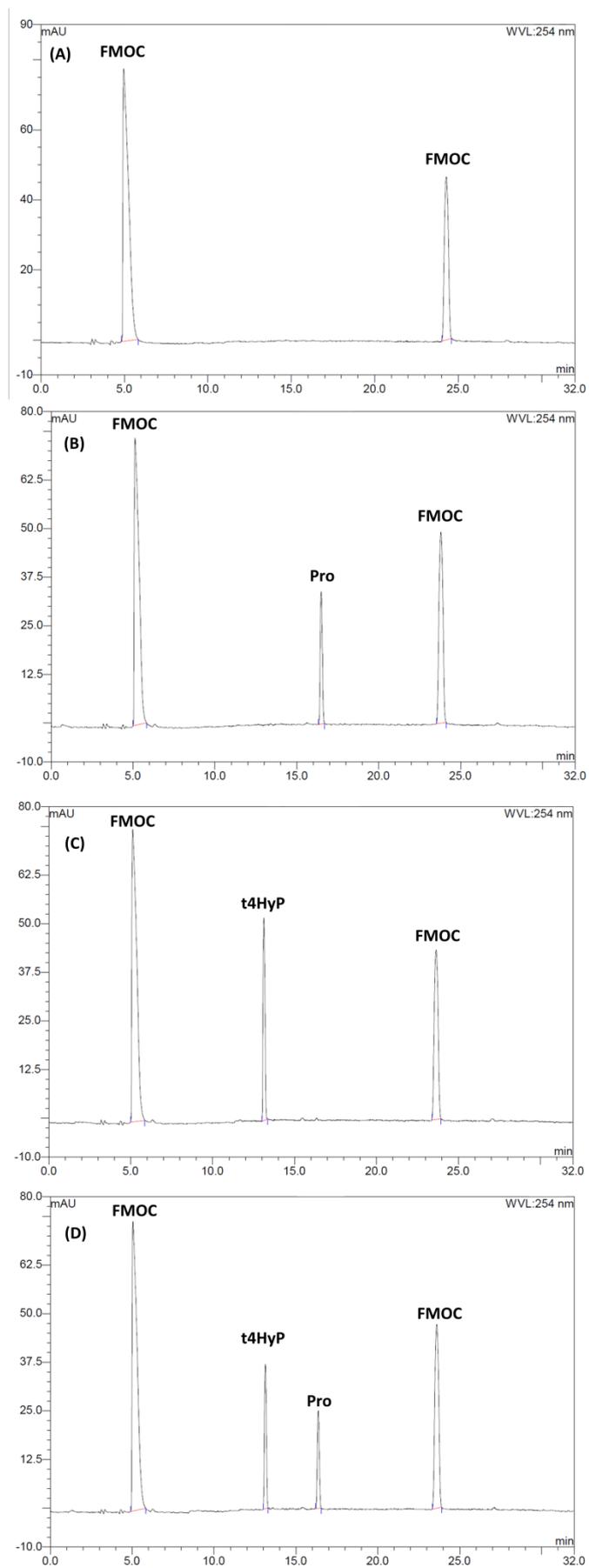
Figure 3.13: Chromatograms of samples derivatised with the Koketsu method. (A) FMOC standard; (B) FMOC-proline; (C) FMOC-t4HyP; (D) FMOC-pro/t4HyP mix.

The chromatograms obtained following analysis of standards are shown in Figure 3.13. In general, it could be seen that shifting to the Koketsu method resulted in the observation of peaks with greater resolution and separation. Peak identities were assigned by comparing chromatograms of the standards. Though such a method of assignment resulted in indications that this derivatisation protocol was effective, these finding were merely assumptions drawn from qualitative information.

Optimisation and LC-MS Analysis

Derivatisations of standards continuously performed whilst optimising the reaction conditions. The first modification was that FMOC-Cl was resuspended in acetonitrile rather than acetone for the purpose of keeping the reaction solvent system consistent with analysis solvents (the LC-MS programs employed acetonitrile gradients). Reactions were performed in glass vials following an investigation of the literature into the effect of performing derivatisations in Eppendorf tubes or glass vials, suggested that the latter is preferable due to the possibility of peptides binding to the walls of the plastic which could have an effect on sample detection at low concentrations.[227] Thirdly, the derivatisation reactions occurred with shaking at 60 °C for 40 min.

In order to confirm that the derivatisation reactions were occurring, UPLC-MS analysis of derivatised samples was performed using the programs in Table 2.32 and Table 2.33 in *2.2.11.2*.
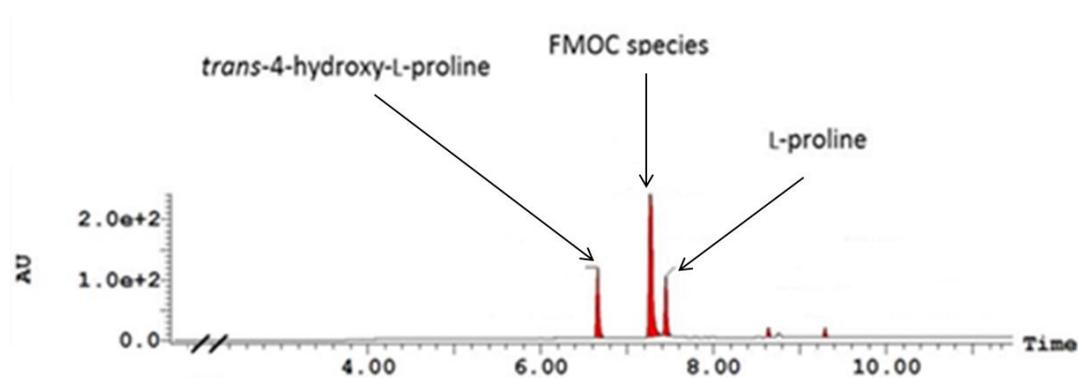
Figure 3.14: UPLC-MS chromatogram of 1:1 mixture of L-proline and trans-4-hydroxy-L-proline (20 mM each) derivatised with FMOC-Cl resuspended in acetonitrile. Clear, distinct peaks corresponding to L-proline-FMOC (7.45 min), FMOC species (7.30 min) and t4HyP-FMOC (6.70 min) were observed. Reaction was analysed using the Long Acidic Method (Table 2.28).

Figure 3.14 presents a UPLC-MS chromatogram of a derivatised 1:1 mixture of proline and t4HyP. Analysis at 254 nm presented three peaks: a 6.70 min peak corresponding to derivatised t4HyP, a 7.30 min corresponding to an FMOC species and a 7.45 min peak representing derivatised L-proline. The identities of these peaks were confirmed by MS analysis which showed that they corresponded to protonated mass of the derivatised products (Table 3.4).

Table 3.4: Retention times of standards as well as MS weight

Analyte	MH$^+$	Retention Time (min)
L-proline-FMOC	338.2	7.45
trans-4-hydroxy-L-proline-FMOC	354.2	6.70
FMOC species	179.1	7.30

Following these findings, derivatisations of standards using this modified protocol were repeated in order to test its replicability, robustness and success rate. Samples were always analysed by LC-MS in order to confirm peak identities by MS. The assay was found to fulfil all of the above criteria and it was decided that it would be employed for the analysis of biotransformations.

Final Derivatisation Method

To 15 μL of sample was added 285 μL of 100 mM sodium borate pH 9.0 and 300 μL of FMOC-Cl (5.8 mM stock for 20 mM sample). The derivatisation mixture was left to react at 60 °C for 40 min. The reaction was quenched with addition of 600 μL 25% acetonitrile: 75% 250 mM sodium borate pH 5.5. Reactions were performed in glass vials and samples were filtered prior to transfer into glass vials for LC-MS analysis.

3.6 Discussion and Conclusions

For the purposes of observing and quantifying the activity of hydroxylase targets towards proline and its derivatives, it was imperative that a derivatisation assay be developed for the analysis of said analytes by chromatography. The goal of the derivatisation reaction was to attach a detectable component to proline and its derivatives to allow for detection under UV light by chromatography. It was important that the developed assay was sensitive, robust, reliable and replicable.

Marfey's reagent was the first derivatisation agent to be utilised. Initial derivatisations of amino acid standards were very promising with both derivatised proline and t4HyP observed as clear peaks in the output chromatograms. The derivatisation reaction was also robust enough to be functional at varied concentrations of proline and t4HyP. Unfortunately, the assay inexplicably proved difficult to control shortly after confidence had been gained. Peaks were not observed where expected and in most cases, it appeared as if the derivatisation reaction itself was not occurring as only a Marfey's peak was observed. A great deal of effort was spent in troubleshooting the reaction and analytical methods as well as in overhauling the LC-MS apparatus. It was ultimately concluded that it would not be beneficial to continue pursuing Marfey's reagent derivatisation, as in this case, it was neither reliable nor reproducible.

Following the failure of Marfey's reagents, the next reagent to be utilised was NBD-Cl. This was selected due to the fact that it had been frequently used in the literature for the derivatisation of amino acids. The reaction itself was also relatively simple which worked in favour of this method. However, there was no success in derivatising the standards these resulting in this method being abandoned due to lack of promise.

The final derivatisation agent to be utilised was FMOC-Cl, an agent that had seen a great deal of success in recent work. A number of methods were adapted and modified in this work. The first method was based on one used by Klein and coworkers.[194] This reaction method was very rapid, requiring less than 20 min in total. It also appeared successful as evidenced by the chromatograms. However, the reaction required ADAM for the purpose of eliminating any produced FMOC-OH as well as 0.5 M sodium borate buffer which

could be deemed to be somewhat high. It was challenging to completely resuspend these components, making this a very difficult reaction to keep consistent hence affecting its reproducibility. As a result, a second derivatisation method was pursued, this time based on one utilised by Koketsu and coworkers in which ADAM was not utilised and the concentration of borate buffer was reduced. RP-HPLC analysis suggested that this was a functional assay. In order to confirm this, LC-MS analysis was performed which identified peaks of interest as corresponding the derivatised proline and t4HyP.

Having observed that Koketsu protocol for FMOC-Cl derivatisation was functional, the reaction protocol was further optimised. In order to limit contaminants from plastic Eppendorf tubes reactions were conducted in glass vials, the reaction was also performed at 60 °C rather than room temperature in order to promote the reaction and resuspension solvents were kept consistent with mobile phase solvents. All of these modifications resulted in a robust, reproducible and reliable assay.

It must be noted that the removal of ADAM from the reaction mixture likely resulted in the occurrence of large FMOC peak in the centre of the chromatogram. Though somewhat inconvenient and likely to add complications to product recovery steps, the adapted Koketsu protocol was retained. In a sense, the central FMOC peak could serve the positive function of acting as an internal standard of sorts with products generally appearing before it and substrates following it.

Conclusion

An assay was successfully developed for the analysis of proline and derivatives by RP-HPLC and LC-MS. This assay involved the precolumn derivatisation of analytes using FMOC-Cl, with the final assay based on one employed by Koketsu and coworkers. This assay appeared robust, reliable, sensitive and replicable. As a result, it was employed for analysis in later work.

Chapter 4

Cloning and Expression of Targets

4.1 Introduction

Proline hydroxylases are of significant industrial interest due to their ability to perform the biocatalytic hydroxylation of proline and itsderivatives. The production of enantiopure hydroxylated proline, and derivatives, is synthetically attractive as these compounds provide access to a wide-range of chiral synthons that can be particularly useful for the synthesis of a host of pharmaceuticals. Microbial and fungal screenings identified individual proline hydroxylases capable of forming each respective enantiomer of hydroxyproline. These are summarised in Table 4.1.

Table 4.1: Microbial and fungal proline hydroxylases

Enzyme	Product	Organism	Abbreviation
Trans-proline-4-hydroxylase	*Trans*-4-hydroxyproline	*Dactylsporangium* sp.	DOGDH
Trans-proline-3-hydroxylase	*Trans*-3-hydroxyproline	*Glarea lozoyensis*	GlP3H
Cis-proline-4-hydroxylase	*Cis*-4-hydroxyproline	*Mesorhizobium loti*	MlC4H
Cis-proline-3-hydroxlase	*Cis*-3-hydroxyproline	*Streptomyces* sp. TH1	StP3H

Trans-proline-4-hydroxylase from *Dactylsporangium* sp. (DOGDH) was first discovered by Shibasaki and colleagues during microbial screening.[200, 202, 224] The group successfully cloned the gene encoding DOGDH and expressed it in *E. coli*. Comparisons of the DOGDH sequence with those of other 2-oxoglutatate dependent hydroxylases revealed little sequence similarity, with the only significant shared feature being the conserved H-X/D-H motif.[200, 202, 224] DOGDH was chosen as the primary hydroxylase of this work due to the significance of its *trans*-4 hydroxylation activity, industrial potential and the fact that its structure had yet to be solved at the start of this project.

The second target, *cis*-proline-3-hydroxylase from *Streptomyces* sp. TH1 (StP3H) was discovered by Mori and coworkers in 1996.[196] This group also successfully cloned and expressed this enzyme in *E. coli*.[201] StP3H was the first proline hydroxylase to have its structure determined with this performed by Clifton and colleagues.[204] Like DOGDH, this enzyme holds significant industrial potential.

The third target of this project was a *cis*-proline-4-hydroxylase from *Mesorhizobium loti* (MlC4H). This hydroxylase was first reported by Hara and Kino.[203] Koketsu and colleagues recently solved the structure of MlC4H and used it as a model for rationalising mutagenesis in a closely related *cis*-proline-4-hydroxlase from *Sinorhizobium meliloti* (SmP4H).[205] This work was significant because following a series of mutations this group was able to evolve the regioselectivity of SmP4H towards L-pipecolic acid, showcasing the potential of proline hydroxylases.[205]

4.2 Aims

For the purposes of performing biotransformations and crystallisation trials, the hydroxylase target genes must be in a plasmid system that can be heterologously expressed in *E. coli* cells. As a result, DOGDH, StP3H and MlC4H will be cloned into pET-expression vectors.

All three hydroxylase were subjected to expression testing in *E. coli* to determine the optimal growth conditions to obtain soluble recombinant protein for further work.

4.3 Hydroxylase Target Genes

DOGDH

The DOGDH gene from *Dactylsporangium* sp. Strain RH1 was codon optimised for expression in *E. coli* and ordered from and synthesised by Thermo Fischer Scientific Life Technologies. The synthesised gene sequence is presented in Figure 4.1.

```
DOGDH 819 BP;
ATGCTGACCCCGACGGAGCTCAAGCAGTACCGCGAGGCGGGCTATCTGCTCATCGAGGACGGCC
TCGGCCCGCGGGAGGTCGACTGCCTGCGCCGGGCGGCGGCGGCCCTCTACGCGCAGGACTCGCC
GGACCGCACGCTGGAGAAGGACGGCCGCACCGTGCGCGCGGTCCACGGCTGCCACCGGCGCGAC
CCGGTCTGCCGCGACCTGGTCCGCCACCCGCGCCTGCTGGGCCCGGCGATGCAGATCCTGTCCG
GCGACGTGTACGTCCACCAGTTCAAGATCAACGCGAAGGCCCCGATGACCGGCGATGTCTGGCC
GTGGCACCAGGACTACATCTTCTGGGCCCGAGAGGACGGCATGGACCGTCCGCACGTGGTCAAC
GTCGCGGTCCTGCTCGACGAGGCCACCCACCTCAACGGGCCGCTGTTGTTCGTGCCGGGCACCC
ACGAGCTGGGCCTCATCGACGTGGAGCGCCGCGCGCCGGCCGGCGACGGCGACGCGCAGTGGCT
GCCGCAGCTCAGCGCCGACCTCGACTACGCCATCGACGCCGACCTGCTGGCCCGGCTGACGGCC
GGGCGGGGCATCGAGTCGGCCACCGGCCCGGCGGGCTCGATCCTGCTGTTCGACTCCCGGATCG
TGCACGGCTCGGGCACGAACATGTCGCCGCACCCGCGCGGCGTCGTCCTGGTCACCTACAACCG
CACCGACAACGCCCTGCCGGCGCAGGCCGCTCCGCGCCCGGAGTTCCTGGCCGCCCGCGACGCC
ACCCCGCTGGTGCCGCTGCCCGCGGGCTTCGCGCTGGCCCAGCCCGTCTAG
```

Figure 4.1: The codon optimised gene sequence of DOGDH. UniProt ID: O06499

StP3H

The StP3H gene from *Streptomyces* sp. TH1 was also codon optimised for *E. coli* and synthesised by Life Technologies. This sequence is shown in Figure 4.2.

```
StP3H 870 BP;
ATGCGTAGCCATATTCTGGGTCGTATTGAACTGGATCAAGAACGTCTGGGTCGCGATCTGGAAT
ATCTGGCAACCGTTCCGACCGTTGAAGAGGAATATGATGAATTTAGCAACGGCTTCTGGAAAAA
CATTCCGCTGTATAATGCAAGCGGTGGTAGCGAAGATCGTCTGTATCGTGACCTGGAAGGTAGT
CCGGCACAGCCGACCAAACATGCAGAACAGGTTCCGTATCTGAATGAAATCATTACCACCGTGT
ATAATGGTGAACGTCTGCAGATGGCACGTACCCGTAATCTGAAAAATGCAGTTGTTATTCCGCA
CCGCGATTTTGTTGAGCTGGATCGTGAACTGGACCAGTATTTTCGTACCATCTGATGCTGGAA
GATTCACCGCTGGCATTTCATAGTGATGATGATACCGTTATTCATATGCGTGCCGGTGAAATTT
GGTTTCTGGATGCAGCAGCAGTTCATAGCGCAGTTAATTTTGCAGAATTTAGCCGTCAGAGCCT
GTGTGTTGATCTGGCATTTGATGGTGCCTTTGATGAAAAGAAGCATTTGCAGACGCAACCGTT
TATGCACCGAATCTGAGTCCGGATGTTCGTGAACGTAAACCGTTTACCAAAGAACGTGAAGCAG
GTATTCTGGCACTGAGCGGTGTTATTGGTCGTGAAAACTTTCGTGATATTCTGTTCCTGCTGAG
CAAAGTGCACTATACCTATGATGTTCATCCGGGTGAAACCTTTGAATGGCTGGTTAGCGTTAGC
AAAGGTGCGGGTGATGATAAATGGTTGAAAAAGCAGAACGCATCCGTGATTTTGCAATTGGTG
CACGTGCCCTGGGTGAACGTTTTAGCCTGACCACCTGG
```

Figure 4.2: The codon optimised gene sequence for StP3H. UniProt ID: P96010

MlC4H

The MlC4H gene from *Mesorhizobium loti* was identified following a BLAST search, codon optimised for expression in *E. coli* and synthesised by Life Technologies. The resulting gene sequence is presented in Figure 4.3.

```
MlC4H 935 BP;
CGAATTGGCGGAAGGCCGTCAAGGCCACGTGTCTTGTCCAGAGCTCATGACCACCCGTATTCTG
GGTGTTGTTCAGCTGGATCAGCGTCGTCTGACCGATGATCTGGCAGTTCTGGCAAAAAGCAATT
TTAGCAGCGAATATAGCGATTTTGCATGTGGTCGTTGGGAATTTTGTATGCTGCGTAATCAGAG
CGGTAAACAAGAAGAACAGCGCGTTGTTGTTCATGAAACACCGGCACTGGCAACACCGCTGGGT
CAGAGCCTGCCGTATCTGAATGAACTGCTGGATAATCATTTTGATCGTGATAGCATTCGCTATG
CCCGTATTATTCGTATTAGCGAAAATGCCTGCATTATCCCGCATCGTGATTATCTGGAACTGGA
AGGCAAATTTATCCGTGTTCATCTGGTTCTGGATACCAATGAAAAATGCAGCAATACCGAAGAG
AACAACATCTTTCATATGGGTCGTGGCGAAATTTGGTTTCTGGATGCAAGCCTGCCGCATAGCG
CAGGTTGTTTTAGCCCGACACCGCGTCTGCATCTGGTTGTTGATATTGAAGGCACCCGTAGCCT
GGAAGAAGTTGCAATTAACGTTGAACAGCCGAGCGCACGTAATGCAACCGTTGATACCCGTAAA
GAATGGACCGATGAAACCCTGGAAAGCGTTCTGGGTTTTAGCGAAATCATTAGCGAAGCCAATT
ATCGTGAAATTGTTGCCATTCTGGCCAAACTGCACTTTTTTCATAAAGTGCATTGCGTGGATAT
GTACGGTTGGCTGAAAGAAATTTGTCGTCGTCGTGGTGAACCTGCACTGATTGAAAAAGCAAAT
AGTCTGGAACGCTTCTATCTGATTGATCGTGCAGCCGGTGAAGTTATGACCTATTAAGGTACCT
GGAGCACAAGACTGGCCTCATGGGCCTTCCGCTCACTGC
```

Figure 4.3: The codon optimised gene sequence of MlC4H. UniProt ID: Q989T9

4.4 Cloning of DOGDH and StP3H into the LIC3C Vector

DOGDH and StP3H were cloned into the YSBLIC3C vector by Laila Roper using the ligation independent cloning protocol described in **2.2.1.6** during a previous PhD project. At the start of this project, the identity of these cloned plasmids was confirmed by DNA sequencing performed by Eurofins Genomics.

4.5 LIC Cloning of MlC4H into the LIC3C Vector

As mentioned above, the *E. coli* optimised synthetic gene of MlC4H was obtained from Thermo Fisher Scientific Life Technologies. The synthetic gene was delivered in a pMA-T vector (Figure 4.4), a GeneArt® cloning vector utilised by Thermo Fisher Scientific.
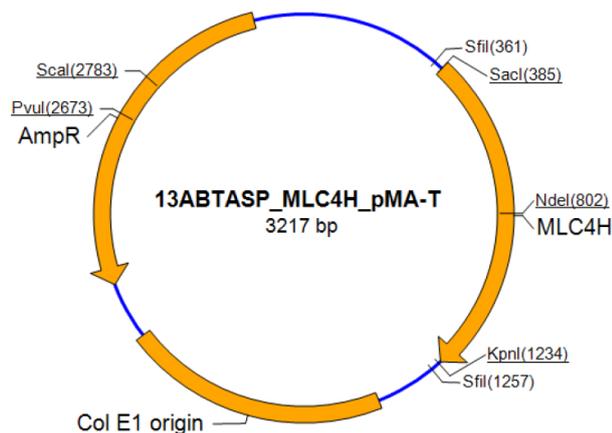
Figure 4.4: Plasmid map of the synthesised MlC4H gene in the pMA-T vector. The hexahistidine tag was inserted at the SacI restriction site so that it was positioned at the N-terminus of MlC4H

Amplification of MlC4H by PCR

Primers specific to MlC4H were ordered from and synthesised by Eurofins Genomics. The required primers were named MlC4H-LIC3C_FOR and MlC4H-LIC3C_REV respectively and their sequences are described in Table 2.17 in Section *2.1.6*. PCR amplification of the synthetic gene was performed by mixing the components in Table 2.14 in a thin-walled PCR with KOD HotStart DNA Polymerase used as the polymerase component in the mixture. The PCR amplification reaction itself was performed in Thermal Cycler using the program described in Table 2.19.

Following the PCR amplification process, the success of the reaction was determined by analysis by agarose gel electrophoresis on a 1% agarose gel. The band corresponding to the amplified MlC4H gene was excised with a scalpel and extracted using a GenElute gel extraction kit (Sigma) with 100 μL of PCR product at a concentration of 51.4 ng μL^{-1} isolated.

Linearisation of pET-YSBLIC3C Vector

Circularised pET-YBSLIC3C was obtained from a laboratory stock. A starter culture of this stock was prepared using the protocol described in *2.2.2.3*, with kanamycin (30 mg mL^{-1} stock) utilised as the antibiotic. This culture was mini-prepped using a GenElute™ Plasmid Miniprep Kit to yield 200 μL of 45.1 ng μL^{-1} product. The purified circular vector was linearised at the site of gene insertion using the restriction endonuclease, BseRI as

described in Table 2.23 and this product purified by agarose gel electrophoresis. Following gel extraction, 100 µL of linearised vector at a concentration of 36.4 ng µL^{-1} was obtained.

T4 Polymerase Reactions

The linearised vector and gene insert were treated with T4 polymerase and dTTP or dATP respectively to introduce complementary sticky ends in each to aid annealing.

LIC Annealing Reaction

The T4 polymerase treated vector and MlC4H gene insert were annealed together by simply adding 4 µL of the former to 2 µL of the latter, with this reaction mixture incubated at room temperature for 10 min. Following incubation, 2 µL EDTA (25 mM stock) was added to the mixture to quench the reaction and left at room temperature for a further 10 min. This mixture was transformed into NovaBlue cells with kanamycin as antibiotic and plates left overnight at 37 °C. The following day, colonies were observed in the plates. Three colonies were selected and used to prepare starter cultures which were minprepped to recover the DNA.

Restriction Digest

A sample of each of the three recovered samples was digested with two restriction enzymes (NdeI and NcoI) to confirm that MlC4H had been successfully inserted into the vector. The results were visualised by agarose gel electrophoresis (Figure 4.5).
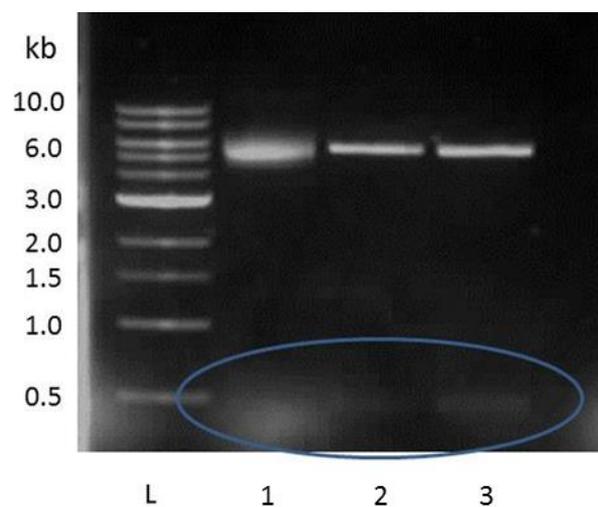
Figure 4.5: Agarose gel of MlC4H following double digest of MlC4H-LIC3C with NdeI and NcoI. All three sample lanes correspond to identical samples. Circled are the bands that likely correspond to the MlC4H gene. L = ladder; 1-3 = digested samples 1-3

The restriction digest experiment showed that in each case the MlC4H insert appeared to have been successfully inserted as evident from the circled bands at approximately 0.8 kb. The bright bands are approximately 6.0 kb likely corresponded to digested pET-YSLIC3C vector. It is important to note that such analysis only provided qualitative evidence suggesting that the insertion had occurred, as a result, further confirmation was required.

DNA Sequencing

As discussed above, the agarose gel electrophoresis analysis of the restriction digested products suggested that insertion of MlC4H had occurred. In order to concretely confirm that cloning was successful, the three DNA samples were sent for sequencing by GATC. This sequencing confirmed that all three products possessed 100% identity to MlC4H inserted into the pET-YSBLIC3C vector. As a result, it could be confidently concluded that MlC4H had been cloned into the pET-YSBLIC3C vector by ligation independent cloning.

4.6 Infusion® Cloning of DOGDH into the pET22b Vector

DOGDH was cloned into the pET22b vector in order to produce a variant containing a C-terminal hexahistidine tag. InFusion® cloning was selected as the method of choice due to its speed and efficiency as it allows for the direct insertion of one DNA fragment into another (DOGDH into the pET22b vector). In this method, two products are generated with complementary 15 bp overhangs by PCR and directly annealed to one another using the InFusion enzyme.

Linearisation of pET22b Vector and Amplification of DOGDH

Circularised pET22b vector was obtained from a laboratory stock and it was necessary to first linearise this. The pET22b vector was linearised by mixing the components in Table 2.23 in *2.2.1.7* and leaving the reaction to proceed for 3 h at 37 °C.

In parallel to this, the DOGDH gene was PCR amplified using newly synthesised primers (DOGDH-pET22b_FOR and DOGDH-pET22b_REV in Table 2.14) using the standard PCR program with KOD HotStart DNA Polymerase.

The linearised pET22b vector and newly amplified DOGDH were visualised by agarose gel electrophoresis (Figure 4.6).
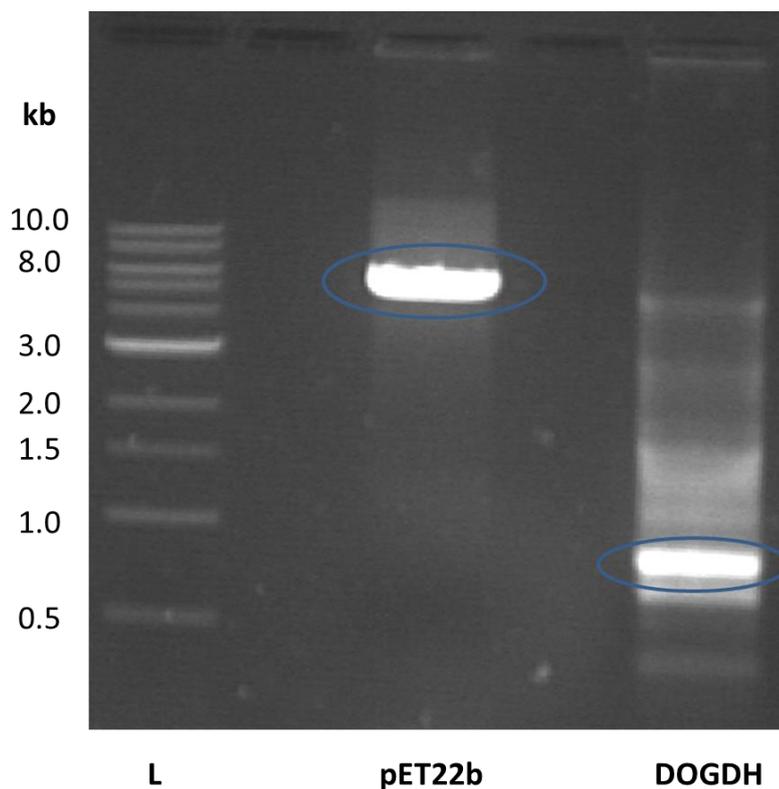
Figure 4.6: 1% agarose gel of linearised pET22 vector and PCR amplified DOGDH. The ladder (L) has been highlighted together with the size of the key marker bands. A clear linearised pET22b vector can be seen at roughly 8.0 kb (circled). PCR amplified DOGDH has also been circled and can be seen at approximately 0.8 kb.

The agarose gel presented bands that likely corresponded to linearised pET22b at 8.0 kb and PCR amplified DOGDH at 0.8 kb. In the case of the latter, other bands could also be seen in the sample suggesting non-specific primer binding; however, it was deemed okay to proceed due to how abundant the DOGDH band appeared. The circled pET22b and DOGDH bands were excised with a scalpel and gel extracted to give 100 μL of linear pET22b vector at a concentration of 45.8 ng μL^{-1} and 100 μL of DOGDH at a concentration of 120 ng μL^{-1}.

Infusion® Reaction

The linearised pET22b was mixed with amplified DOGDH and milli-Q water as described in Table 2.20. The Infusion® reaction was left to proceed at room temperature for 15 min. Following this, the entire reaction mixture was transformed into NovaBlue cells with ampicillin (100 mg mL^{-1} stock) as antibiotic. The following day, two colonies were selected

and used to prepare starter cultures with ampicillin as antibiotic. The DOGDH-pET22b plasmid was extracted from these cultures using a Miniprep kit.

Restriction Digest and Sequencing

The result of Infusion reaction was investigated by performing restriction double digests of the isolated products using XhoI and NdeI. The double digest reactions were left to proceed for 3 h at 37 °C. Following the reaction, samples were visualised by agarose gel electrophoresis (Figure 4.7).
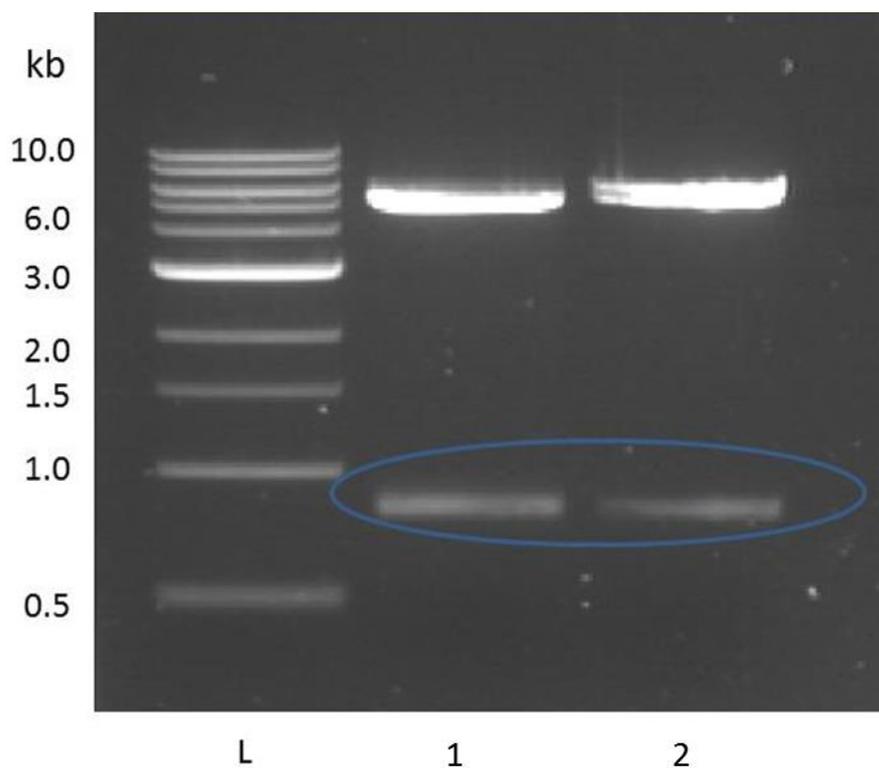
Figure 4.7: 1% agarose gel of the double digest products of DOGDH-pET22b. Circled are the bands likely corresponding to the DOGDH insert (0.9 kb), while the bright band at approximately 6 kb likely corresponds to the truncated pET22b vector. L = ladder; 1-2 = identical DOGDH-pET22b samples treated with XhoI and NdeI.

The above agarose gel suggested that the insertion of DOGDH into the pET22b vector had been successful as, for each sample, two bands were observed: a band at approximately 6 kb likely corresponding to truncated pET22b and a band at roughly 0.9 kb likely to be the DOGDH insert. These plasmids were sent for sequencing by GATC and the results confirmed that DOGDH had been inserted into the pET22b vector to give a C-terminal hexahistidine tagged variant of this enzyme.

4.7 Expression Testing

Expression testing in different strains of *E. coli* was performed in order to determine the optimal conditions for the expression of soluble protein. The detailed protocol of this process is described in Section *2.2.3.1*.

Expression Testing of DOGDH in pET-YSBLIC3C

Expression of DOGDH-LIC3C was investigated in two strains of *E. coli*; BL21(DE3) and B834(DE3), with these strains selected due to being ideal for routine T7 expression and generally suitable for the recombinant production of enzymes cloned into pET vectors. DOGDH-LIC3C had a molecular weight of 34 kDa.

BL21(DE3)

The 12% SDS-PAGE gel of this expression test is presented in Figure 4.8.
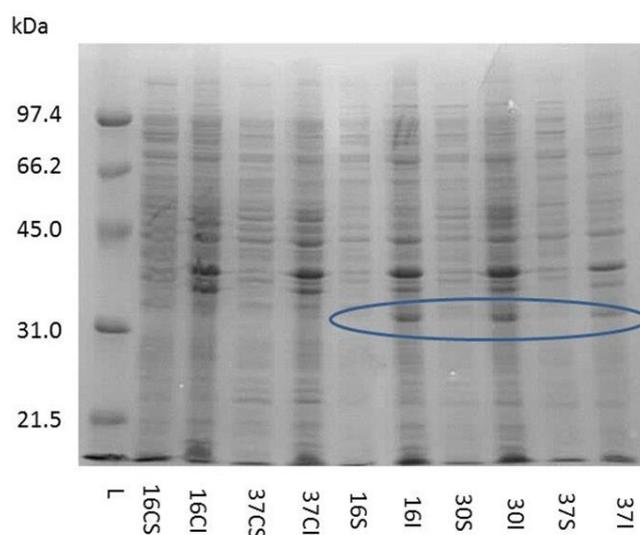
Figure 4.8: Expression test of DOGDH in *E. coli* BL21(DE3) cells. L= ladder; C=control (not induced with 1 M IPTG); S=soluble fraction; I=insoluble fraction. Number=incubation temperature. DOGDH-LIC3C is approximately 34 kDa in size. Where DOGDH should appear is circled, the faintness of bands in the soluble fractions indicate poor soluble expression

This expression test indicated poor soluble expression of DOGDH-LIC3C in BL21(DE3) cells at all three growth temperatures, as evident from the very faint bands in the circled regions.

B834(DE3)

The SDS-PAGE gel for this expression test is presented in Figure 4.9.
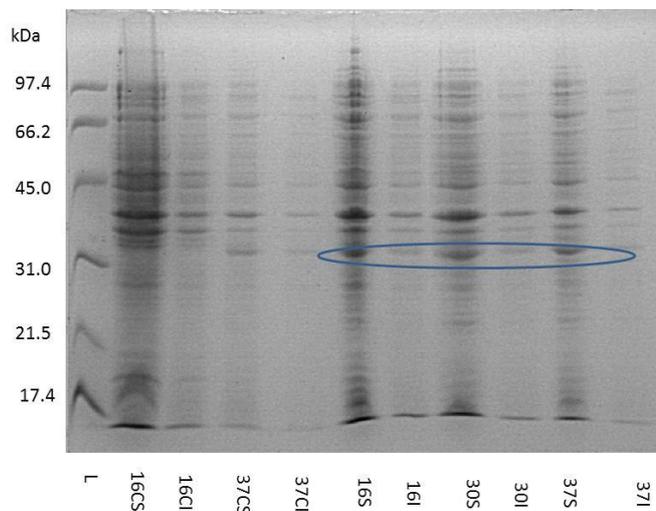
Figure 4.9: 12 % SDS-PAGE gel obtained following the expression test of DOGDH-LIC3C in B834(DE3) cells. The protein of interest is approximately 31 kDa in size. Soluble expression appeared greatest at 16 °C, as evident from the circled band. L= ladder; C=control (not induced with 1 M IPTG); S=soluble fraction; I=insoluble fraction.

There appeared to be a marked improved in the soluble expression of DOGDH-LIC3C in B834(DE3) cells compared to BL21(DE3) cells. The greatest amount of soluble expression appeared to occur when media was grown at 16 °C post-induction. This result suggested that B834(DE3) cells should be utilised for scaled-up expression.

Expression Testing of StP3H in pET-YSBLIC3C

BL21(DE3)

The expression of StP3H-LIC3C was tested in *E. coli* BL21(DE3) cells. StP3H-LIC3C has a molecular weight of approximately 35 kDa. The results of this expression test are shown in Figure 4.10.
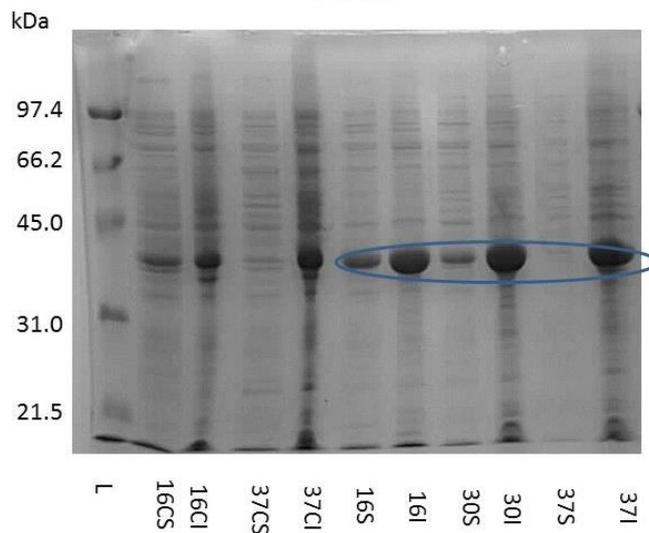
Figure 4.10: Expression tests of StP3H in E. coli BL21(DE3) cells. C=control (not induced with 1 M IPTG), S=soluble fraction, I=insoluble fraction. Number=incubation temperature. StP3H-LIC3C is approximately 35 kDa in size, there is very good soluble expression at 16 °C and 30 °C. L= ladder; C=control (not induced with 1 M IPTG); S=soluble fraction; I=insoluble fraction.

StP3H exhibited excellent soluble expression at both 16 °C and 30 °C, though it must be noted that like in DOGDH the majority of the protein appeared in the insoluble fraction. This promising result suggested that it was not necessary to screen expression in different *E. coli* strains. It was decided that scale-ups of StP3H-LIC3C would be performed using BL21(DE3) cells with post-induction growth performed at 16 °C.

Expression Testing of MlC4H in pET-YBSLIC3C

The expression of MlC4H-LIC3C was tested in *E. coli* BL21(DE3) and B834(DE3) cells. MlC4H-LIC3C is approximately 35 kDa in size.

BL21(DE3)

The SDS-PAGE gel obtained following expression testing of MlC4H-LIC3C in BL21(DE3) cells is presented in Figure 4.11.
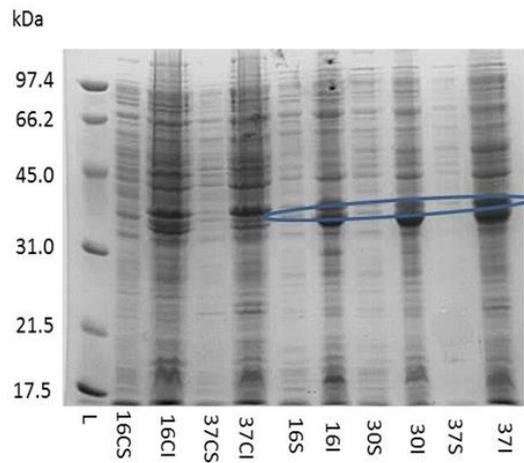
Figure 4.11: Expression test of MlC4H in BL21(DE3). The region in which the bands corresponding to it are expected have been labelled. Soluble expression is generally poor with greater amounts of insoluble expression observed. L= ladder; C=control (not induced with 1 M IPTG); S=soluble fraction; I=insoluble fraction.

Soluble expression of MlC4H-LIC3C at all three growth temperatures appeared to be poor as evident from the faint bands in the circled regions.

B834(DE3)

The SDS-PAGE gel summarising the results of this expression test is presented in Figure 4.12.
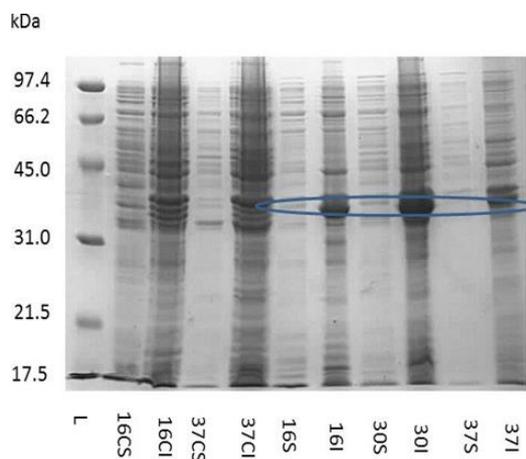
Figure 4.12: Expression test of MlC4H in B834(DE3). The region in which the bands corresponding to it are expected have been labelled. Soluble expression is poor with greater amounts of insoluble expression observed . L= ladder; C=control (not induced with 1 M IPTG); S=soluble fraction; I=insoluble fraction.

Soluble expression of MlC4H-LIC3C in B834(DE3) also appeared to be poor as only faint bands were observed at all three growth temperatures. In general, MlC4H-LIC3C seemed to be a predominantly insoluble protein in both tested *E. coli* strains.

Expression Testing of DOGDH in pET22b

Expression tests of DOGDH-pET22b were conducted in both *E. coli* BL21(DE3) and B834(DE3) strains. DOGDH-PET22b is approximately 34 kDa in size.

BL21(DE3)

Figure 4.13 presents the SDS-PAGE gel obtained following expression testing of DOGDH-pET22b in *E. coli* BL21(DE3) cells.
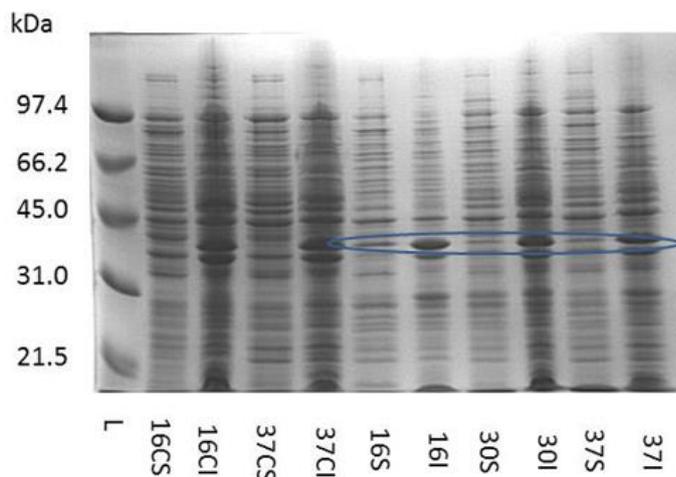
Figure 4.13: 12% SDS-PAGE gel of DOGDH-pET22b expression test in BL21(DE3) cells. Circled is the region in which the protein bands would be expected. There appears to be faint soluble expression at all three growth temperatures with the greatest seen at 16 °C. Once again, the protein appears to be mostly insoluble. L= ladder; C=control (not induced with 1 M IPTG); S=soluble fraction; I=insoluble fraction.

DOGDH-pET22b appeared to exhibit some soluble expression at all three growth temperatures as evident from the faint bands observed in the circled region of interest. The greatest soluble expression occurred at 16 °C. It must again be noted that the protein was predominantly expressed in the insoluble fractions.

B834(DE3)

The results of expression testing in B834(DE3) cells are summarised in Figure 4.14.
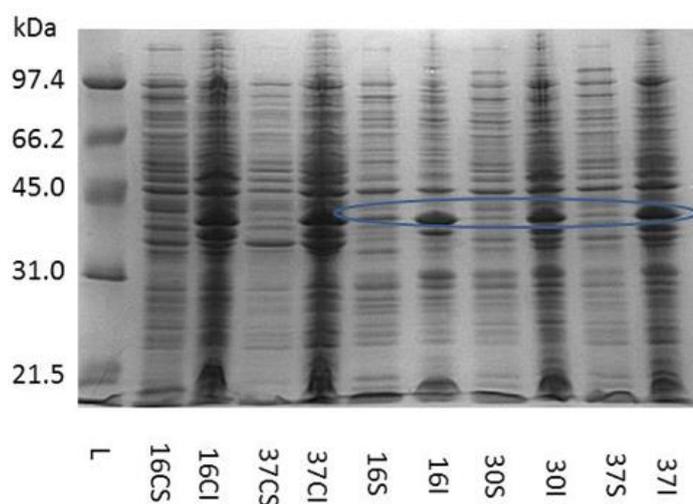
Figure 4.14: 12% SDS-PAGE gel of DOGDH-pET22b expression test in B834(DE3) cells. Circled is the region in which the protein bands would be expected. There appears to be faint soluble expression at all three growth temperatures with the greatest seen at 16 °C. Once again, the protein appears to be mostly insoluble. L= ladder; C=control (not induced with 1 M IPTG); S=soluble fraction; I=insoluble fraction.

DOGDH-pET22b exhibited moderate soluble expression at all three growth temperatures in B834(DE3) cells. The greatest soluble expression was observed at 16 °C, however, the majority of protein was again observed in the insoluble fractions.

4.8 Discussion and Conclusions

DOGDH and StP3H had been cloned into the pET-YSBLIC3C prior to the commencement of this work and the success of this was confirmed by sequencing. As a result, it was not necessary to clone either of these into this vector. MlC4H was successfully cloned into the pET-YSBLIC3C vector by ligation independent cloning.

DOGDH was cloned into the pET22b vector to produce a C-terminal hexahistidine tagged variant of the enzyme for purification and crystallisation purposes by Infusion® cloning. This method of cloning was selected because it generally requires fewer steps than ligation independent cloning, with the method based on the principle of joining two sequences (insert and vector) with 15 bp complementary overhangs.[228-229] DOGDH was successfully cloned into the pET22b vector as determined by sequencing.

Following the successful cloning of targets into their respective vectors, expression testing in various *E. coli* strains was performed in order to determine optimal growth condition for

the attainment of soluble protein. DOGDH-LIC3C showed very poor soluble expression in BL21(DE3) with the best (though still moderate) expression seen in B834(DE3). StP3H-LIC had very highly soluble expression in BL21(DE3) cells while MlC4H-LIC3C was poorly expressed in both strains. DOGDH-pET22b displayed moderate soluble expression in both BL21(DE3) and B834(DE3) cells. All of the targets displayed the greatest soluble expression when grown at 16 °C following resulting in this being applied to scale-ups. In general, all targets showed more insoluble expression which aligns with the literature.

In conclusion, hydroxylase targets were successfully cloned into their target vectors and expression tests were performed to try and ascertain optimal expression conditions and strains.

Chapter 5

Towards a structure of the *trans*-proline-4-hydroxylase DOGDH

5.1 Introduction

DOGDH is a *trans*-4-proline-hydroxylase originally discovered in *Dactylsporagium* sp. following a series of microbial screenings. This enzyme was first recombinantly overexpressed in *Escherichia coli* by Shibasaki and coworkers in 1999.[200] DOGDH is a significant 2-oxoglutarate dependent hydroxylase due to its ability to react with free L-proline in isolation to produce *trans*-4-hydroxy-L-proline (t4HyP) with high selectivity and at elevated rates.[176] T4HyP is a key intermediate for the synthesis of various pharmaceuticals, making it of significant industrial interest. This enzyme also has the capacity to react with substrates other than L-proline granting it great potential for the synthesis of an assortment of chiral synthons.[176]

Of the proline hydroxylases in the 2-oxoglutarate dependent oxygenase family, only StP3H and MlC4H have had their protein structures solved by X-ray crystallography.[204-205] This structural data was used to elucidate the mechanism of these enzymes, identify key amino acid residues for catalysis and suggested targets for mutagenesis in order to evolve the function of these enzymes. In the case of MlC4H, the structure was used to inform mutagenesis of a closely related hydroxylase, SmC4H, identifying residues for mutagenesis in order to shift regioselectivity toward L-pipecolic acid.[205]

There is currently no structure for DOGDH making its acquisition highly desirable. A DOGDH structure could be used to further characterise its mechanism of action and identify residues of interest for structure-guided mutagenesis.

5.2 Aims

DOGDH must be successfully purified by chromatography in order to perform crystallisation trials. Purified protein will be used to perform crystallisation trials using commercial screens. If necessary, modifications will be performed on DOGDH with the hope of aiding its crystallisation. If crystallisation of DOGDH itself is unsuccessful, proteins with sequence homology to it will be pursued as targets.

5.3 Purification of DOGDH

5.3.1 DOGDH-LIC3C

DOGDH had previously been cloned into the pET-YSBLIC3C vector by L. Roper. Expression tests were performed in both *E. coli* BL21(DE3) and B834(DE3) strains with poor soluble expression observed in both, though the latter showed slightly improved expression when grown at 16 °C following induction. It was therefore decided that initial purification attempts would be performed using DOGDH-LIC3C expressed in B834(DE3) cells.

DOGDH-LIC3C was expressed in B834(DE3) and protein production performed at a large scale as described in *2.2.3.2*. Ni-affinity purifications were performed using the following buffer system: Buffer 1: 50 mM Tris-HCl, 20 mM AEBSF, 20 mM Imidazole pH 7.0; Buffer 2: 50 mM Tris-HCl, 20 mM NaCl, 20 mM AEBSF, 500 mM Imidazole pH 7.0. This system was identical to that of L. Roper who had previously observed DOGDH stability issues during purification and consequently added the protease inhibitor AEBSF at a final concentration of 20 mM to purification buffers. Following sonication, cells were harvested, the supernatant containing soluble protein isolated and both Ni^{2+}-affinity and Size Exclusion Chromatography (SEC) conducted sequentially as described in *2.2.4*.

Figure 5.1 presents the SEC chromatogram obtained following purification of DOGDH-LIC3C expressed in B834(DE3) cells. A single sharp peak was observed between following flow of 55 and 70 mL of SEC buffer.
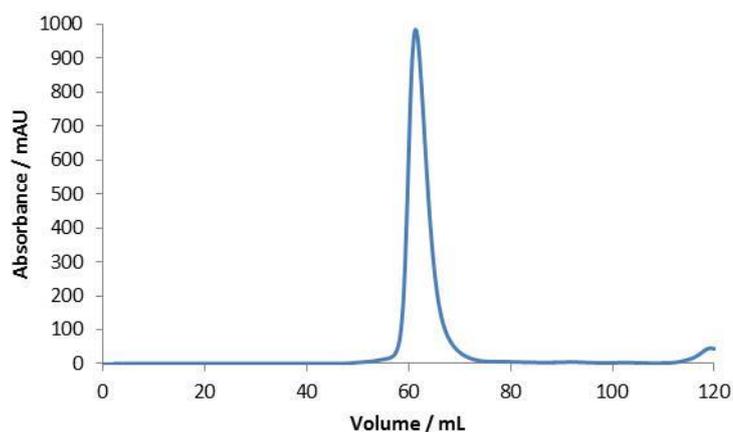
Figure 5.1: Chromatogram obtained following SEC of DOGDH-LIC3C expressed in B834(DE3) cells, a distinct peak can be between 55 and 70 mL, reaching its amplitude at roughly 63 mL. Absorbance was measured at 280 nm.

Ni^{2+}-affinity and SEC purification fractions were pooled respectively and analysed by 12% SDS-PAGE (Figure 5.2).
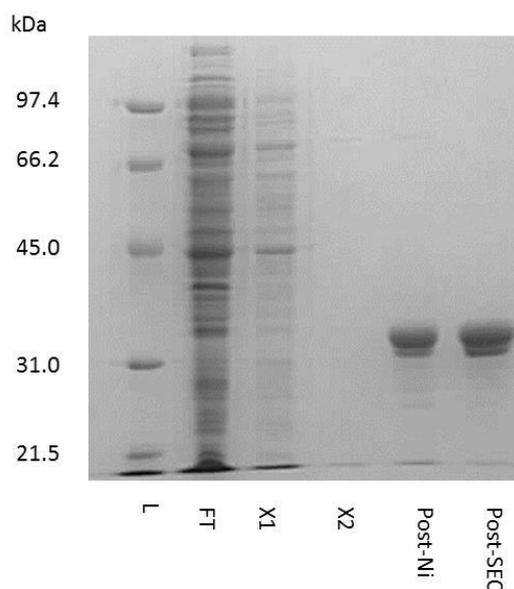
Figure 5.2: 12% SDS-PAGE gel of samples from DOGDH-LIC3C purification following expression in B834. L = protein ladder, FT=flowthrough, X1 and X2= wash fractions, Post-Ni=pooled peak fractions obtained following Ni-affinity chromatography, Post-SEC=pooled fractions obtained following SEC. SEC appeared to have further purified the sample, however, there were multiple bands suggesting protein breakdown.

The gel in Figure 5.2 suggested that DOGDH-LIC3C was successfully expressed and isolated as evident from the thick bands at 34 kDa in both the post-Ni^{2+} affinity and post-SEC samples, however, the presence of multiple bands at this position suggested that the isolated protein was undergoing degradation. This degradation occurred despite the presence of protease inhibitor. The same result was also observed for DOGDH-LIC3C

127

expressed in BL21(DE3) (Figure 5.3). The protein bands of interest were analysed using MALDI-TOF Mass Spectrometry which suggested that degradation was occurring towards the C-terminus of DOGDH-LIC3C. This conclusion was drawn due to the identification of peptide fragments that aligned with positions 237–254, 255–274 and 275–294 of the DOGDH amino acid sequence (with position 294 corresponding the C-terminus of the enzyme) (Figure 5.4).
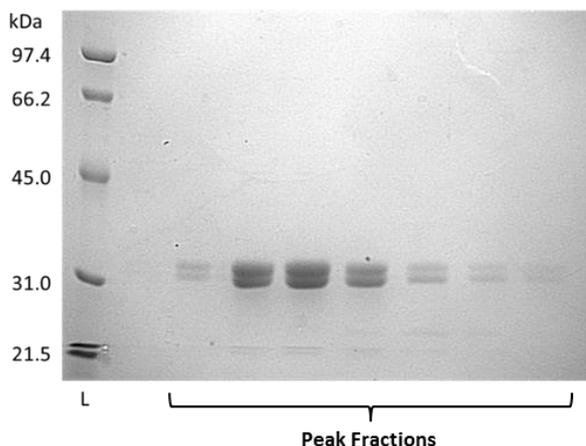
Figure 5.3: 12% SDS-PAGE gel of post SEC fractions of DOGDH-LIC3C expressed in BL21(DE3). Once again, multiple bands can be seen where DOGDH-LIC3C would be expected.

```
  1 MGSSHHHHHH SSGLEVLFQG PAMLTPTELK QYREAGYLLI EDGLGPREVD
 51 CLRRAAAALY AQDSPDRTLE KDGRTVRAVH GCHRRDPVCR DLVRHPRLLG
101 PAMQILSGDV YVHQFKINAK APMTGDVWPW HQDYIFWARE DGMDRPHVVN
151 VAVLLDEATH LNGPLLFVPG THELGLIDVE RRAPAGDGDA QWLPQLSADL
201 DYAIDADLLA RLTAGRGIES ATGPAGSILL FDSRIVHGSG TNMSPHPRGV
251 VLVTYNRTDN ALPAQAAPRP EFLAARDATP LVPLPAGFAL AQPV
```

Figure 5.4: Amino acid of DOGDH-LIC3C with the peptide fragments detected by MALDI-TOF MS highlighted in red. The peptide fragment sequences aligned with amino acids towards the C-terminus of the DOGDH suggesting that degradation was occurring in this region. The N-terminal hexahistidine tag has been highlighted in a blue rectangle.

The point at which degradation of DOGDH-LIC3C commenced was investigated by performing a time course stability test. In this, cells expressing DOGDH-LIC3C were lysed by sonication, the lysate was kept on ice and samples were taken hourly for analysis by SDS-PAGE. The results of this experiment are summarised in Figure 5.5.
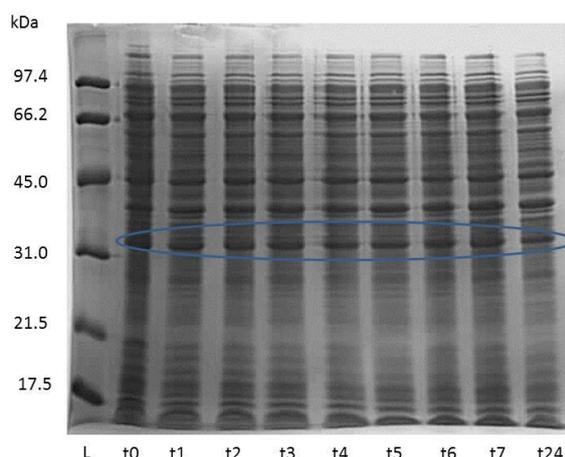
Figure 5.5: 12% SDS-PAGE gel of samples taken at different time points following cell lysis. t0= lysis point. A solid, stable 34 kDa can be seen in all the samples, this was likely DOGDH-LIC3C. t1 – t24 correspond to the hourly time point samples while L is the protein ladder.

This stability test suggested that DOGDH-LIC3C remained stable following sonication and that protein degradation was most likely occurring following Ni^{2+}-affinity purification. This degradation may have been the result of abstraction of the active site Fe(II) in DOGDH during the Ni^{2+}-affinity purification process as a result of metal ion transfer leading to the protein losing its stability.[230] This possibility would be difficult to avoid as affinity purification was deemed an essential step in protein isolation.

The concentration of salt in the purification buffers was also deemed a potential factor contributing to the instability of DOGDH. NaCl plays a number of roles in the purification buffer including stabilising the protein of interest by increasing its ionic strength which helps prevent non-specific interactions between the protein and the column. Many protocols recommend that at least 50 mM of salt is present in purification buffers to prevent this. The buffer utilised in these purification trials only contained 20 mM NaCl which could perhaps be deemed too low for its stabilising effect to occur. Other members of the research group had experienced purification successes with NaCl concentrations as high as 500 mM, as a result, the concentration of NaCl in each purification buffer was adjusted to match this.

As little difference was observed between the large scale expression of DOGDH-LIC3C expressed in BL21(DE3) and B834(DE3), expression from this point onwards was performed in the latter. Following expression and harvesting, purifications were again conducted but with modified purification buffers.
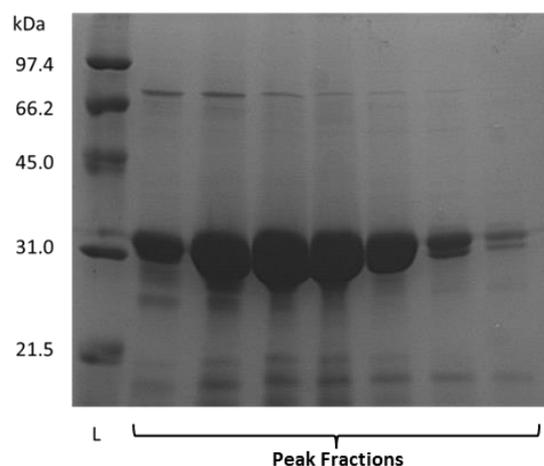
Figure 5.6: 12% SDS-PAGE gel of DOGDH-LIC3C peak fractions fractions obtained following Ni-affinity purification with modified purification buffers

Figure 5.6 above presents the SDS-PAGE gel of peak fractions obtained following Ni^{2+}-affinity purification of DOGDH-LIC3C with the modified buffer system. It was immediately evident that shifting to this system resulted in greatly improved protein yield as shown by the thicker bands. There are however multiple bands in the area of interest as seen in the lanes to the far right of the gel, which suggested that protein degradation was still occurring regardless of the increased salt concentration. This hinted toward an increased likelihood of protein instability being the result of abstraction of the active site Fe(II) metal cofactor during this purification process.

5.3.2 DOGDH-pET22b

MS data confirmed DOGDH-LIC3C underwent degradation toward the enzyme C-terminal end during and following the protein purification process, this suggested that a different construct DOGDH should be synthesised for purification and crystallisation purposes. It was thus decided that DOGDH be cloned into the pET22b vector in order to introduce a C-terminal hexahistidine tag into the protein.

DOGDH-pET22b was expressed in BL21(DE3) cells, grown in LB and purified by Ni^{2+}-affinity and SEC purification using the buffers in Table 2.4. 10% v/v glycerol was added to the resuspension buffer to aid stabilisation and prevent protein aggregation.[231] Figure 5.7 shows the SDS-PAGE gel obtained following Ni-affinity purification.

130
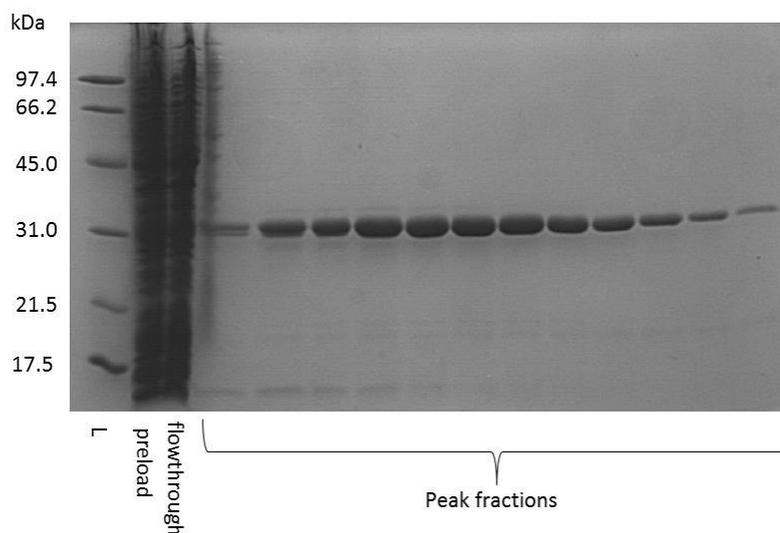
Figure 5.7: 12% SDS-PAGE gel of the peak fractions obtained following Ni-affinity purification of DOGDH-pET22b. A very clean, clear band can be seen at approximately 30 kDa in most fractions. This likely corresponds to isolated DOGDH. The fractions are also very clean, particularly following just one purification step.

A very clear, solid band likely corresponding to stable DOGDH-pET22b was observed in each fraction. This suggested that a combination of the new construct and modified buffer conditions had resulted in the prevention of proteolysis during Ni-affinity purification. These fractions were pooled and concentrated down to 2 mL prior to loading onto an S75 column. SEC was performed, yielding the chromatogram in Figure 5.8.
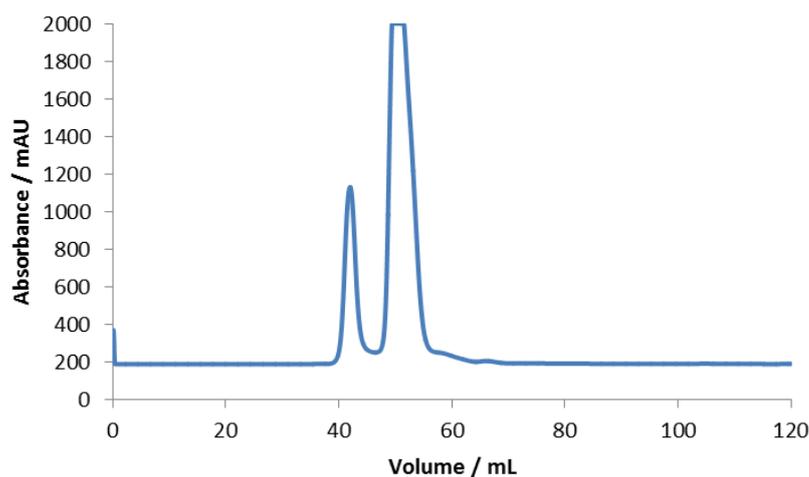
Figure 5.8: SEC Chromatogram of DOGDH-pET22b expressed in BL21(DE3). Two peaks were observed. The presence was detected by absorbance measurements at 280 nm.

Two peaks were observed and fractions corresponding to them were analysed by SDS-PAGE (Figure 5.9).
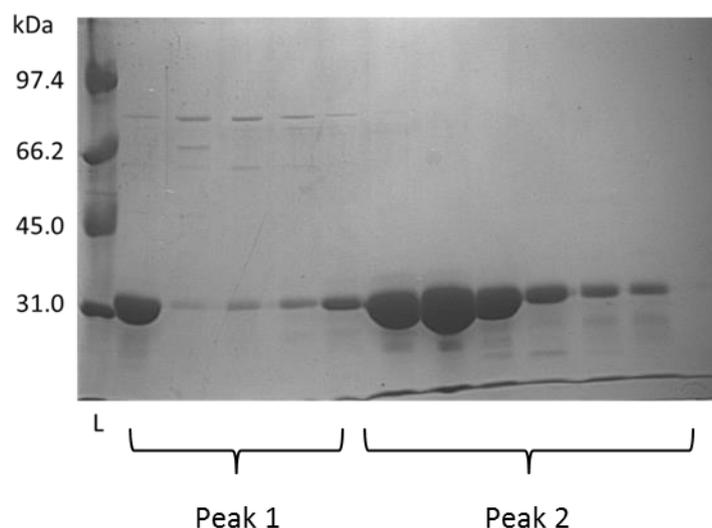
Figure 5.9: 12% SDS-PAGE gel of fractions isolated during SEC of DOGDH-pET22b. A 32 kDa band corresponding to DOGDH-pET22b can be seen in every fraction. The thickest bands correspond to fractions obtained at the height of the peak in the chromatogram

The above gel shows that the majority of DOGDH-pET22b was isolated in the second, more predominant peak on the chromatogram. Thick bands at around 32 kDa corresponding to DOGDH-pET22b can be seen in the Peak 2 fractions. These bands were also whole suggesting that stable, soluble protein had been isolated. As a result, it could be concluded that cloning of DOGDH into the pET22b vector resulted in a construct that could be expressed and purified to yield soluble protein for crystallisation trials. In general 0.8 mg of DOGDH-pET22b was isolated per gram of harvested cells.

5.4 Crystallisation Trials of DOGDH

5.4.1 DOGDH-pET22b

DOGDH-pET22b was a very promising candidate for crystallisation trials due to its heightened stability compared to the LIC3C variant following purification. This stability and ease of purification made this target amenable for broader trials.

In initial trials, purified DOGDH-pET22b was concentrated to 10 mg mL^{-1} and trialled with three commercial screens: PACT, INDEX and CSSI/II in a 96 well format as described in *2.2.7.1*. The majority of drops at low protein concentration were empty, while increasing concentration resulted in a greater incidence of debris and protein precipitate in the drops.

132

It was thought that these findings resulted from the fact that DOGDH is a multi-component enzyme that requires substrate, 2-oxoglutarate, and Fe(II) for function, with it being highly probable that Fe(II) was abstracted from the enzyme during Ni-purification. It was therefore hypothesised that adding L-proline, 2-oxoglutarate and Fe(II)SO$_4$ to the mother liquor solution in the wells of the commercial crystallisation screens may aid protein crystallisation. This was done by preparing a ligand stock containing 1 M L-proline, 1 M 2-oxoglutaric acid and 100 mM Fe(II)SO$_4$ which was added to the mother liquor in a 50-fold dilution (final concentrations of 20 mM, 20 mM and 2 mM respectively). Plates were left for 30 min prior to addition of protein in order to allow the contents of the wells to mix.

A drop of interest was observed for DOGDH-pET22b in CSS B7 (Figure 5.10). This drop consisted of what appeared to be showers of microcrystals. Though not optimal, the fact that crystalline matter in itself had been observed was promising. As a result, 24-well hanging drop optimisation trials were attempted (see *2.2.7.2*).
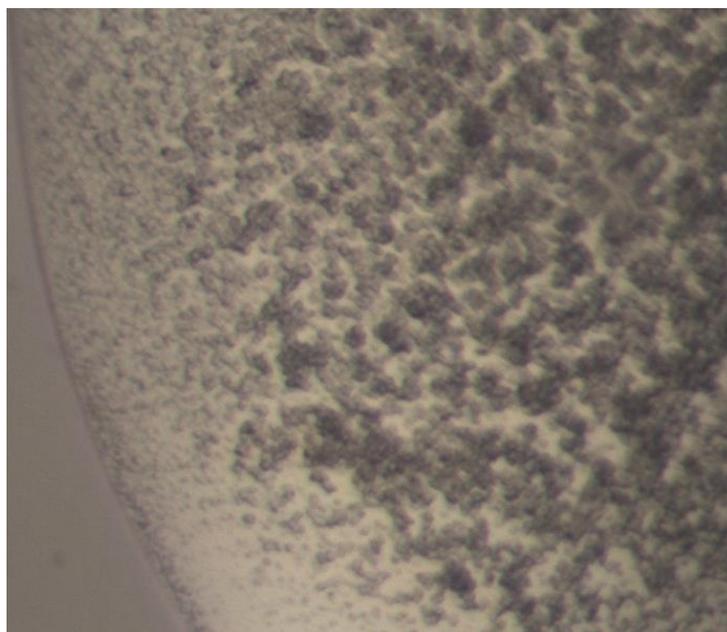
Figure 5.10: DOGDH-pET22b CSS B7 drop that yielded crystalline matter. Microcrystals were observed in the centre of the drop (dark region). The composition of the drop was 0.3 M Sodium Acetate, 15% PEG 4000, 0.1 M Bis Tris Propane pH 6.5, 20 mM L-proline, 10 mM 2-oxoglutaric acid, 2 mM Fe(II)SO$_4$. The concentration of DOGDH-pET22b was 20 mg mL^{-1}.

The crystals observed in Figure 5.10 were found in well B7 of the CSS screen. The composition of this drop was 0.3 M Sodium Acetate, 15% PEG 4000, 0.1 M Bis Tris Propane pH 6.5, 20 mM L-proline, 10 mM 2-oxoglutaric acid and 2 mM Fe(II)SO$_4$. This

condition was optimised by manipulating the pH range of Bis Tris Propane (from 5.5 to 9.0, in either 0.25 or 0.5 increments depending on the plate) and the concentration of PEG 4000 (5% to 30%, in 5% increments with the lowest concentration tested on a plate being either 5% or 15% depending on the upper limit). This, however, yielded little success with failure to even replicate the initial hit at a larger scale.

A 96 well additive screen (Hampton Research) of this condition was also attempted in order to assess whether the addition of additives could aid the crystallisation of DOGDH-pET22b. This too, however, was unsuccessful with either empty drops or precipitate observed.

Final crystallisation attempts focussed on directly supplementing the purified DOGDH-pET22b with $Fe(II)SO_4$ resuspended in Buffer B prior to dispensing the protein onto the crystallisation drops. This was unsuccessful, with $Fe(II)SO_4$ displaying poor solubility at physiological pH, swiftly precipitating out of the mixture. Koketsu *et al.* directly supplemented MlC4H with Co(II) as a Fe(II) mimetic for this reason.[205] Using this work as inspiration, $Co(II)Cl_2$ dissolved in Buffer B was directly added to purified DOGDH-pET22b, however, DOGDH was found to precipitate out of solution after addition of even the smallest possible quantity of Co(II). This suggested that Co(II) was toxic to DOGDH and could not be used a Fe(II) mimetic in this work.

5.4.2 Modification of DOGDH

Little success was achieved with crystallisation attempts of DOGDH-LIC3C and DOGDH-pET22b in their native states; this resulted in the focus of subsequent work shifting towards modifying DOGDH enzymatically, chemically or by mutagenesis with the hope of enhancing its crystallisability.

5.4.2.1 Cleavage of N-terminal 6His Tag

The N-terminal hexahistidine tag of DOGDH-LIC3C was cleaved due to the fact that the presence or absence of a tag can have a strong influence on the crystallisation behaviour of a protein. This work was performed as described in *2.2.5*. Uncleaved DOGDH-LIC3C and HRV3C protease, which both still possessed hexahistidine tags, were separated from

cleaved DOGDH-LIC3C by Ni-affinity chromatography using the program in Table 2.23. The chromatogram obtained for this purification is shown in Figure 5.11.
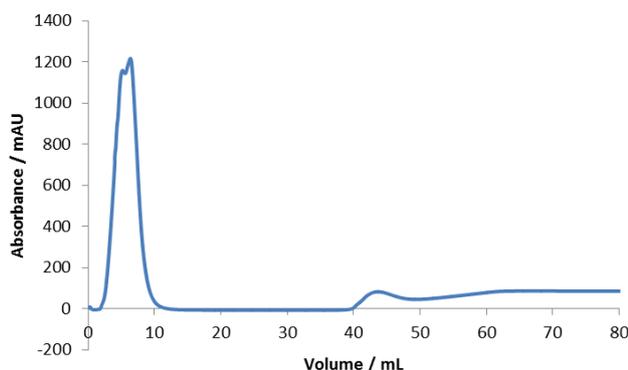
Figure 5.11: Ni-affinity purification chromatogram of DOGDH-LIC3C following overnight reaction with HRV3C protease. The cleavage appeared successful as evident from the large peak in the first 10 mL wash (no imidazole), this likely corresponded to unbound and cleaved DOGDH-LIC3C. A second peak was observed from 40 mL likely corresponding to HRV3C and uncleaved DOGDH-LIC3C. Absorbance was measured at 280 nm.

The above chromatogram suggested that the cleavage was successful as large peak was observed and isolated in the wash step of the purification. This peak likely corresponded to cleaved DOGDH-LIC3C that had not bound to the column due to the lack of a hexahistidine tag. A second, smaller peak was observed from 40 mL which likely corresponded to HRV3C protease and any remaining uncleaved DOGDH-LIC3C. The tag cleavage process was also monitored by SDS-PAGE and the resulting gel presented in Figure 5.12.
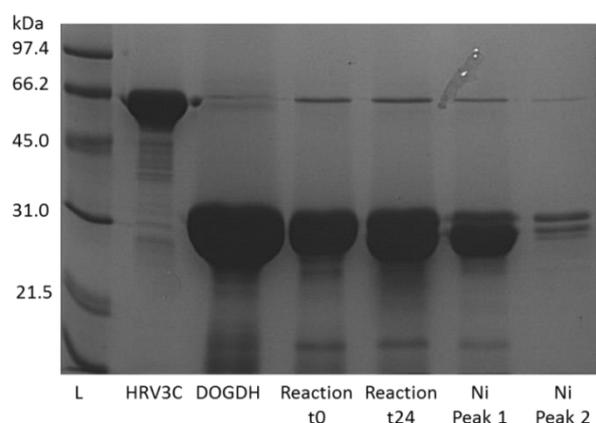
Figure 5.12: 12% SDS-PAGE gel of the His-tag cleavage process. Following the reaction at 24 h it can be seen that a new band approximately 2 kDa smaller than tagged DOGDH-LIC3C is in abundance. This is also the dominant product in the first (larger) Ni-purification peak. There was also evidence of a faint amount of HRV3C protease present in every sample at 66 kDa.

The SDS-PAGE gel provided further evidence that the cleavage reaction had likely been successful as evident from the fact that a band at roughly 2 kDa smaller than tagged DOGDH-LIC3C was the most abundant species in the sample analysed 24 hours following the reaction and in the first Ni purification peak.

The fractions in the large peak were pooled and further purified by SEC (Figure 5.13) and peak fractions analysed by SDS-PAGE (Figure 5.14).
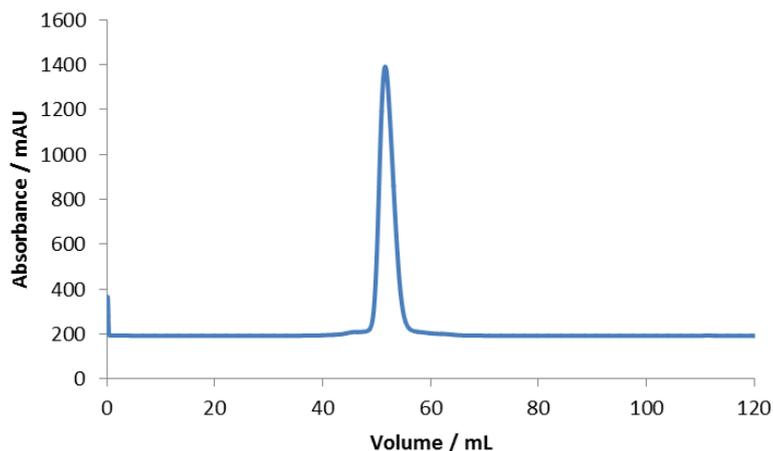
Figure 5.13: SEC chromatogram of DOGDH-LIC3C following His-tag cleavage. A single sharp band was observed between 50-60 mL. Absorbance was measured at 280 nm.
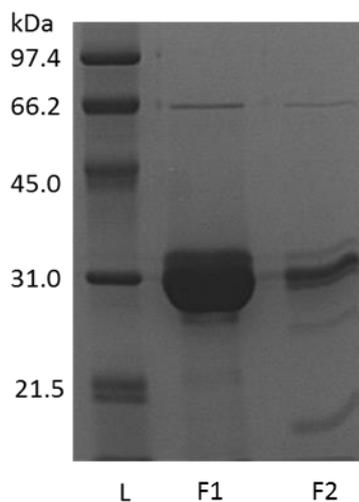
Figure 5.14: 12% SDS-PAGE gel of the two fractions (F1 and F2) obtained following SEC of His-tag cleaved DOGDH-LIC3C. A band approximately 2 kDa smaller than tagged DOGDH-LIC3C was the most abundant product. As this was DOGDH-LIC3C, multiple bands were observed where the protein was expected likely due to stability issues. Some remnants of HRV3C protease were also present at 66 kDa.

Cleaved DOGDH-LIC3C eluted in a single peak between $50 - 60$ mL of buffer. SDS-PAGE analysis supported the claim that the cleavage had been successful as seen by the fact that the abundant product band was approximately 2 kDa smaller than that for uncleaved DOGDH. Multiple bands were observed together with the cleaved protein suggesting that degradation was occurring which was to be expected as this had previously been the case for DOGDH-LIC3C.

The hexahistidine-tag free DOGDH-LIC3C was divided in two fraction, with $Fe(II)SO_4$ (0.5 mM final) added to one fraction. Crystallisation trials with the three commercial screens were performed with both fractions at protein concentrations ranging from 10 mg mL^{-1} to 50 mg mL^{-1}. Results were, however, identical to previous crystallisation attempts with empty drops, debris or precipitate observed. This suggested that cleavage of the N-terminal hexahistidine tag made little difference to the crystallisability of DOGDH.

5.4.2.2 Reductive Methylation of DOGDH-pET22b

Reductive methylation of surface exposed lysine residues works on the principle of chemically modifying these residues by methylation of their solvent exposed ε-amino group in order to reduce the surface entropy and theoretically aid crystal lattice formation without altering their protein structure or biochemical function.[232] This covalent modification was attempted on DOGDH-pET22b as described in *2.2.6*. Following this process, the DOGDH product was concentrated to 50 mg mL^{-1} and 96 well crystal trials performed using CSS, PACT and INDEX as standard.

Crystallisation trials were not successful following this attempt, with most drops appearing empty. This may have been because the protein concentration was not sufficient for crystallisation but in this instance this could not be remedied as the protein yield was very low (0.1 mg per g of cells) which may have resulted from losses during the multiple steps involved in this process.

Figure 5.15 shows a 12% SDS-PAGE gel of protein fractions obtained during the reaction process. In general, it is difficult to ascertain the success of this reaction from the bands on the gel and the yield of protein isolated was very poor with all of it used for crystal trials.
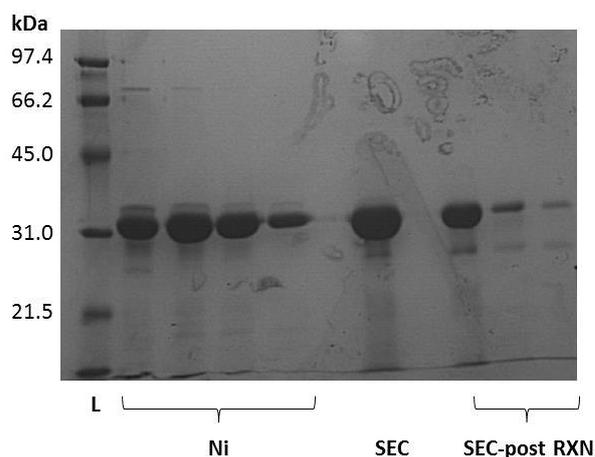
Figure 5.15: 12% SDS-PAGE gel of protein fractions obtained during the reductive methylation reaction process. Native DOGDH-pET22b is approximately 32 kDa in size and can be seen in the Ni and SEC fractions. The thick band isolated following the reaction process and buffer exchange by SEC appears larger than the native DOGDH-pET22b as evident by the fact it appears higher up on the gel. However, this qualitative measure is not enough to ascertain reaction success.

Reductive methylation proved to be inconclusive in this instance. Crystals were not obtained and the yield of protein was insufficient for downstream work. As a result, this process was set aside and an alternative pursued.

5.4.2.3 Surface Entropy Reduction Mutagenesis

As chemical modification of DOGDH-pET22b by reductive methylation did not appear to aid in producing diffraction quality crystals in the first instance, covalent modification by mutation was next attempted. Surface entropy reduction (SER) mutagenesis works under the principle of mutating surface exposed large hydrophilic amino acid residues, that may impede a protein's ability to form ordered crystal lattices by disrupting intermolecular contacts, to alanine.[233] Mutation to a more compact amino acid should in theory reduce the surface entropy and aid crystal lattice formation.[233]

The gene for a SER variant of DOGDH was designed by using an online server provided by UCLA and designed by Goldschmidt and colleagues, available at http://services.mbi.ucla.edu/SER/.[234] The original DOGDH sequence was input into the server, which yielded a job result file with a number of amino acid clusters that could be mutated to alanine, each assigned a score from 0-6 (the higher the number, the more optimal the cluster). The highest ranking result (score of 5.18) was a cluster containing glutamic acid and lysine at positions of 48 and 49. A synthetic gene containing E48A and

K49A mutations (DOGDH-SER) was ordered from Thermo Fisher Scientific (Figure 5.16).

```
ATG CTG ACC CCG ACC GAA CTG AAA CAG TAT   < 30
M   L   T   P   T   E   L   K   Q   Y

CGT GAA GCA GGT TAT CTG CTG ATT GAA GAT   < 60
R   E   A   G   Y   L   L   I   E   D

GGT CTG GGT CCG CGT GAA GTT GAT TGT CTG   < 90
G   L   G   P   R   E   V   D   C   L

CGT CGT GCA GCA GCA GCA CTG TAT GCA CAG   < 120
R   R   A   A   A   A   L   Y   A   Q

GAT AGT CCG GAT CGT ACC CTG GCA GCA GAT   < 150
D   S   P   D   R   T   L   A   A   D
```

Figure 5.16: The first 100 amino acids of the DOGDH-SER Synthetic Sequence. The E48A and K49A mutation sites have been highlighted in red.

DOGDH-SER was cloned into the pET-YBSLIC3C vector using the primers, DOGDH-SER_FOR and DOGDH-SER_REV, in Table 2.14. Double digests reactions with NdeI and NcoI were performed as previously described and samples analysed by agarose gel electrophoresis (Figure 5.17). Six samples were analysed by agarose gel electrophoresis, four of these indicated successful sub-cloning as evident from the presence of a faint band at 0.8 kb corresponding to DOGDH-SER and a bright 6.0 kb YSBLIC3C band. These four samples were analysed by sequencing which confirmed that the sub cloning had been successful.
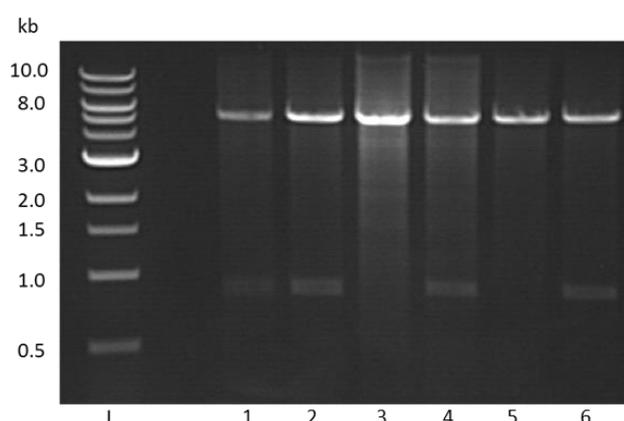
Figure 5.17: Double digest of DOGDH-SER cloned into the YSBLIC3C vector. A band at approximately 0.8 kb can be seen in 4 of the 6 samples (samples 1, 2, 4 and 6), this was likely DOGDH-SER, while a brighter band at approximately 6.0 kb likely corresponded to the LIC3C vector. 1-6 = samples treated with the restriction enzymes NdeI and NcoI.

Expression testing was performed in *E. coli* BL21(DE3) cells and samples analysed by SDS-PAGE (Figure 5.18).
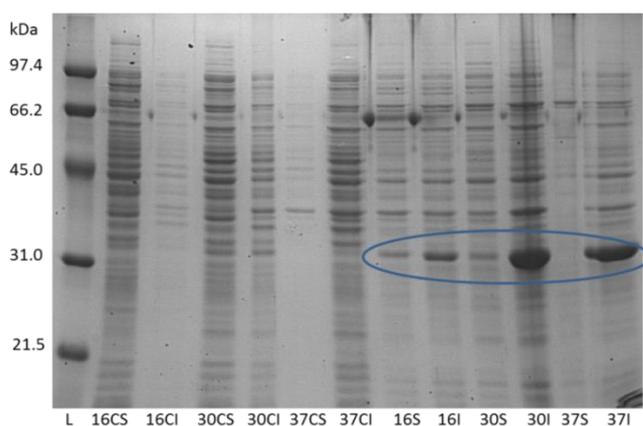
Figure 5.18: 12% SDS-PAGE gel of expression test of DOGDH-SER in BL21(DE3) cells. The growth temperature has been highlighted by the number. C = uninduced control. S = soluble fraction. I = insoluble fraction. DOGDH-SER appears soluble at all three growth temperatures as evident from the 32 kDa bands that have been circled. The greatest soluble expression was observed at 16 °C.

DOGDH-SER displayed greater soluble expression than previously studied DOGDH variants, with the highest level observed at a growth temperature of 16 °C following induction with IPTG. This provided confidence to proceed with large-scale purifications.

DOGDH-SER was expressed in *E. coli* BL21(DE3) cells and grown in 4.5 L LB. Following harvest, Ni-affinity and SEC purifications were performed as standard. The output SEC chromatogram is presented in Figure 5.19. A large and abundant peak was observed at a wavelength of 280 nm between 70 – 85 mL. The fractions corresponding to this peak were analysed by SDS-PAGE (Figure 5.20).
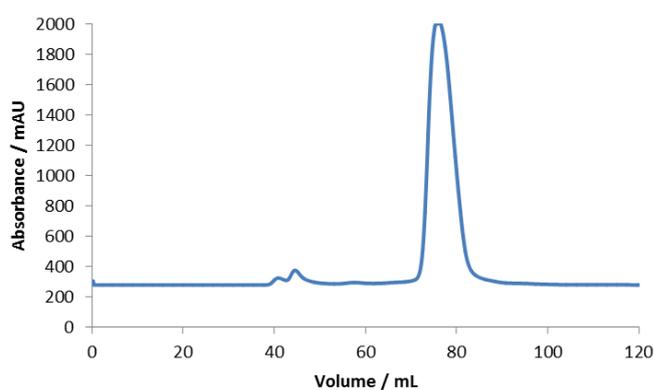
Figure 5.19: SEC chromatogram of DOGDH-SER purification. A large and abundant peak was observed between 70 – 85 mL. The peak appeared at a later volume than previously purified DOGDH equivalent due to the use of a Superdex 200 column rather than Superdex 75. Absorbance was measured at 280 nm.
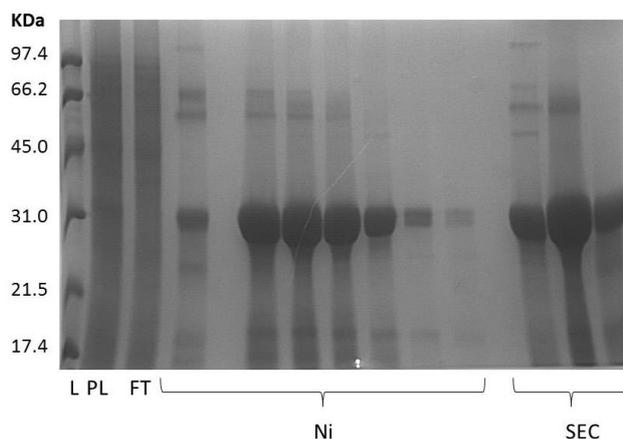
Figure 5.20: 12% SDS-PAGE gel of Ni and SEC fractions obtained during purification of DOGDH-SER. It is evident that samples were cleaned between purifications. An abundant 34 kDa peak can be seen in the key Ni and SEC fractions, this likely corresponded to DOGDH-SER-LIC3C. Purification appeared to have been successful.

SDS-PAGE analysis suggested that purification had been successful as evident from the thick and abundant 34 kDa bands in the key fractions. The protein yield markedly improved with 3 mg of protein per gram of cells isolated.

SEC fractions were pooled and the protein concentrated to 26 mg mL^{-1}. Crystal screens were performed as standard using CSS, PACT and INDEX with two drops, one with protein at 13 mg mL^{-1} and the other at 26 mg mL^{-1}. Crystalline matter was observed in the top drops (13 mg mL^{-1} of DOGDH-SER) of INDEX A10, A11 and A12 (Figure 5.21).

INDEX A10
0.1 M BisTris pH 6.5
3.0 M NaCl

INDEX A11
0.1 M HEPES pH 7.5
3.0 M NaCl

INDEX A12
0.1 M Tris pH 8.5
3.0 M NaCl

Figure 5.21: Three drops of interest observed following crystallisation trials of DOGDH-SER. The well composition has been stated below each drop. Showers of what appear nanocrystalline in natute can be seen in the outer regions of the drops.

The three drops, like with DOGDH-pET22b, consisted of nanocrystal showers in the outer regions of the drop. Though not optimal, these drops showed promise and were

optimised in 24-well hanging drop formats with drop pH and salt concentration manipulated. Figure 5.22 presents a successfully optimised drop.
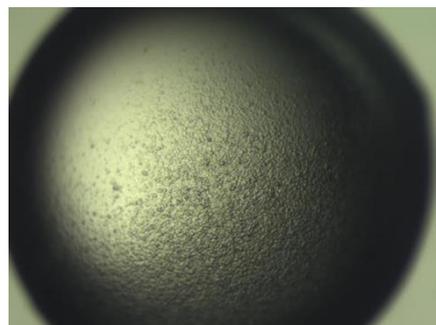
INDEX A12 opt B3:
2.5 M NaCl
0.1 M Tris pH 7.5

Figure 5.22: Drop in well B3 of the INDEX A12 optimisation. Nanocrystal showers can be seen throughout the drop. The composition of solutions in the well has been stated. This was a significant finding due to being the first time that crystalline material was observed in the hanging drop format.

This successful optimisation was significant, despite the drop being identical to what was observed in the 96 well formats, due to the fact that this was the first time for a DOGDH variant that a 96 well sitting drop was optimised and replicated in a 24 well hanging drop format. The INDEX A12 condition was further optimised by adding additives (Proline+2-OG, Ethylene glycol, DMSO, polypropylene glycol P400, glycerol, MPD, Butan-1,4-diol) at 2% and 4% v/v respectively to the drops. Figure 5.23 presents one of the drops observed following these optimisations. The crystals were yet again nanocrystal-like in nature.
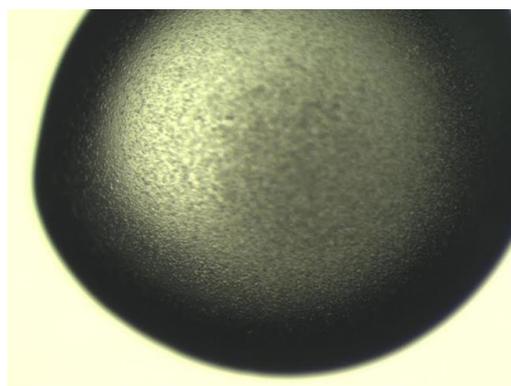
INDEX A12 rd 2.5 opt B2 top:
2.5 M NaCl
0.1 M Tris pH 6.5
4 % v/v ethylene glycol

Figure 5.23: Optimised drop in well B2 of the INDEX A12 optimisation with ethylene glycol utilised as an additive. Nanocrystal showers (some globular in shape) can be seen throughout the drop. The well composition has been stated in the figure.

142

The extensive attempts at optimising the hits observed in the INDEX screen, though promising, failed to yield diffraction quality crystals. As a result, an alternative approach was undertaken. This approach involved the addition of ligands to the mother liquor solutions of the crystallisation screens as described for DOGDH-pET22b but replaced $Fe(II)SO_4$ with $Co(II)Cl_2$, as the former was prone to oxidising in solution and the latter reported to be a good mimic.[205] The ligand stock concentration was as before but the dilution factor increased to 1000 fold. Crystal screens using 96-well commercial screens were prepared as standard.

Figure 5.24 presents two drops in which a globular crystal morphology was observed in the bottom region of the drops. These hits were observed in PACT G9 and G10, and the conditions optimised by manipulating the PEG concentration and buffer pH. Unfortunately, little success was found in doing this, with the best optimisations yielding bubble-like quasi crystals (Figure 5.25).
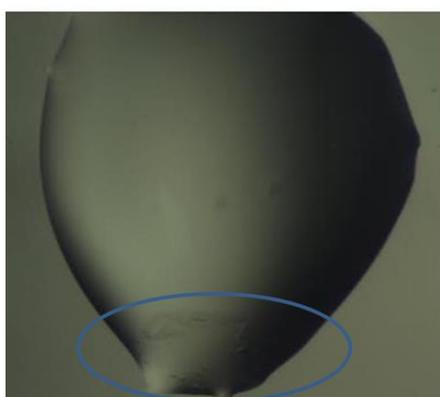
PACT G9
0.2 M Potassium Sodium Tartrate
0.1 M BisTris Propane pH 7.5
20 % PEG 3350
0.01 M Ligand Stock
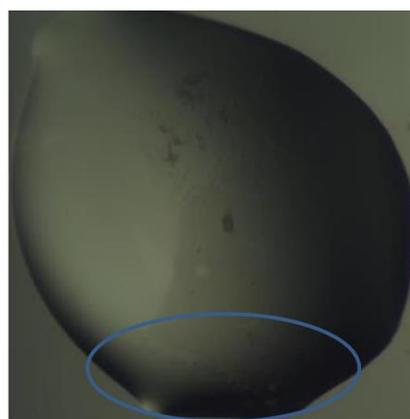
PACT G10
0.02 M sodium/potassium phosphate
0.1 M BisTris Propane pH 7.5
20 % PEG 3350
0.01 M Ligand Stock

Figure 5.24: Two drops observed in PACT G9 and G10 following replacement of $Fe(II)SO_4$ with $Co(II)Cl_2$ in the ligand stock. The well compositions have been stated below the drops. Globular material was observed at the bottom of the drops (circled).
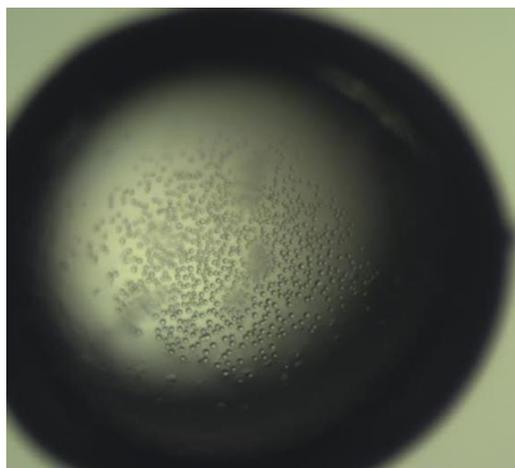
PACT G10 opt A5 top drop
0.02 M sodium/potassium phosphate
0.1 M B BisTris Propane pH 8.5
15 % PEG 3350
0.01 M Ligand stock

Figure 5.25: The top drop observed in A5 of the PACT G10 optimisation with well composition stated beneath. Bubble-like quasi crystals can be seen throughout but are particularly abundant in the centre of the drop.

The PACT G9 and G10 conditions were further optimised by assessing the effect of adding a series of detergents, which may aid in crystallising by improving solubility via the prevention of aggregation.[235] The following detergents were used: NDSB-195 (0.5 M stock), NDSB-201 (0.5 M stock), NDSB-211 (0.5 M Stock) and Triethanolamine (100% v/v stock). Detergents were added to the reservoir solutions at volumes between 0-50 µL in increments of 10 µL.

The optimised plates were left to incubate for an extended period of time but failed to yield any diffraction worthy crystals, with most droplets appearing either empty or containing bubble-like quasi crystals.

5.4.3 DOGDH Homologs

Extensive effort was employed in attempts to obtain diffraction quality crystals of DOGDH, with each attempt proving unsuccessful. Koketsu and colleagues determined the structure of SmC4H, which itself proved difficult to crystallise, by solving the structure of MlC4H which possessed a closely related sequence to it, and used this structure to model that of SmC4H.[205] A similar approach was attempted for DOGDH.

Following a BLAST search, two homologs with 50-60% sequence identity to DOGDH were selected (Figure 5.26):

(i) Hypothetical *trans*-4-hydroxylase from *Micromonospora* sp. CNB394 (MP4H)

(ii) Hypothetical *trans*-4-hydroxylase from *Nocardia* sp. BMG111209 (NP4H)

```
              10        20        30        40        50        60
               |         |         |         |         |         |
DOGDHx0   MLTPTELKQYREAGYLLIEDGLGPREVDCLRRAAAALYAQDSPDRTLEKDGRTVRAVHGC
MP4Hxx1   MLSYESIDLYREDGFLFADP-LTEKETDLLREQADREFLRDSPGRMLEKDGFTVRGVHGS
NP4Hxx2   MLSDDQLDTYYEDGFVVVPDVLGPAQRIRLRAEVERLYTVDHPGRVLEKDGRTVRAVHGC
          **:   .:.  * * *::.      *    :   **     :  * *.* ***** ***.***.
Prim.cons. MLS333LD3YREDGFL333D2LGP3E3D3LR33A3RLY33DSPGR3LEKDGRTVRAVHGC

              70        80        90       100       110       120
               |         |         |         |         |         |
DOGDHx0   HRRDPVCRDLVRHPRLLGPAMQILSGDVYVHQFKINAKAPMTGDVWPWHQDYIFWAREDG
MP4Hxx1   HVVNSTFARLVRHPKIVTPATQLLGGPVYVHQFKINAKKALTGDVWPWHQDYIFWNRGDG
NP4Hxx2   HRTSPLFSRLVRLPFLVAAAERILDGKVYVHQSKVNAKRALHGDIWPWHQDYVFWEREDG
          *  ..   ***  *  :: .* ::*.* ***** *:.*** .:  **.:******;:** * **
Prim.cons. HR33P3F3RLVRHP3LV3PA3QIL3G3VYVHQFKINAK3ALTGDVWPWHQDYIFW3REDG

             130       140       150       160       170       180
               |         |         |         |         |         |
DOGDHx0   MDRPHVVNVAVLLDEATHLNGPLLFVPGTHELGLIDVERRAPAGDGDAQWLPQLSADLDY
MP4Hxx1   MRRPDVVNVAVLLDEATDLNGPLLVLPGSHKCGSLEVARRTTVC-GDGAWRSDISADLDY
NP4Hxx2   MPAPRATNIAVFLDEATAHNGPLLFVPGSHRLGTVASGRRGGDG-----WKSNLAADLDY
          *  *   .:*:**:.***** **** .:**:*. * :  **    * .::::**** .**
Prim.cons. M3RP3VVNVAVLLDEAT3LNGPLLFVPGSH3LG333V3RR333GDGD22W3S3LSADLDY

             190       200       210       220       230       240
               |         |         |         |         |         |
DOGDHx0   AIDADLLARLTAGRGIESATGPAGSILLFDSRIVHGSGTNMSPHPRGVVLVTYNRTDNAL
MP4Hxx1   AIDAPLLSKLTETTKVTAIKGPPGSILLFDPLLVHASGVNMAPYDRRMILVTYNRVDNPL
NP4Hxx2   ALTAEQLTPLIARLGIRAATGPEGSALLFDPRLIHGSGANMSPVDRQLVIITYNSVDNVP
          *:* * *:.*    : :  .** ** **  .:*.**.**:*  *  :::.*** .**
Prim.cons. AIDA3LL33LTA33GI3AATGP3GSILLFDPRLVHGSG3NMSP3DR33VLVTYNRVDN3L

             250       260       270
               |         |         |
DOGDHx0   PAQAAPRPEFLAARDATPLVPLPAGFALAQPV
MP4Hxx1   GEVPSPRPDFLAARDNTPVRPLDPGADLLA--
NP4Hxx2   VEVPNRRPDFLCEPDSTALEPLIDEF------
          .  **:**.  * *.: **
Prim.cons. 3EVP3PRPDFLAARD3TPL3PL33GF2L22PV
```

Figure 5.26: Amino acid sequence alignment of DOGDH, MP4H and NP4H

The synthetic genes of MP4H and NP4H optimised for expression in *E. coli* were ordered from and synthesised by Life Technologies. For the purpose of cloning into the homologs into the pET-YBSLIC3C vector, the genes were PCR amplified using the primers MP4H-LIC3C_FOR, MP4H-LIC3C_REV, NP4H-LIC3C_FOR and NP4H-LIC3C_REV (described in Table 2.13). Following this, the amplified genes were extracted, purified, T4 treated and annealed to the linearised vector. Sequencing was performed by GATC which confirmed that the DOGDH homologs had been successfully cloned into the pET-YSLIC3C vector.

Expression tests were performed for both enzymes in *E. coli* BL21(DE3) cells. Figure 5.27 shows the 12% SDS-PAGE gels obtained following these tests. In general, there was more

insoluble than soluble protein expression for both proteins. The two enzymes also showed evidence of leaky expression as evident by the 32 kDa bands present in the non-induced control samples. For MP4H, the greatest soluble expression appeared to occur when cells were grown at 30 °C post-induction. In the case of NP4H, soluble expression was equally good at all three growth temperatures.
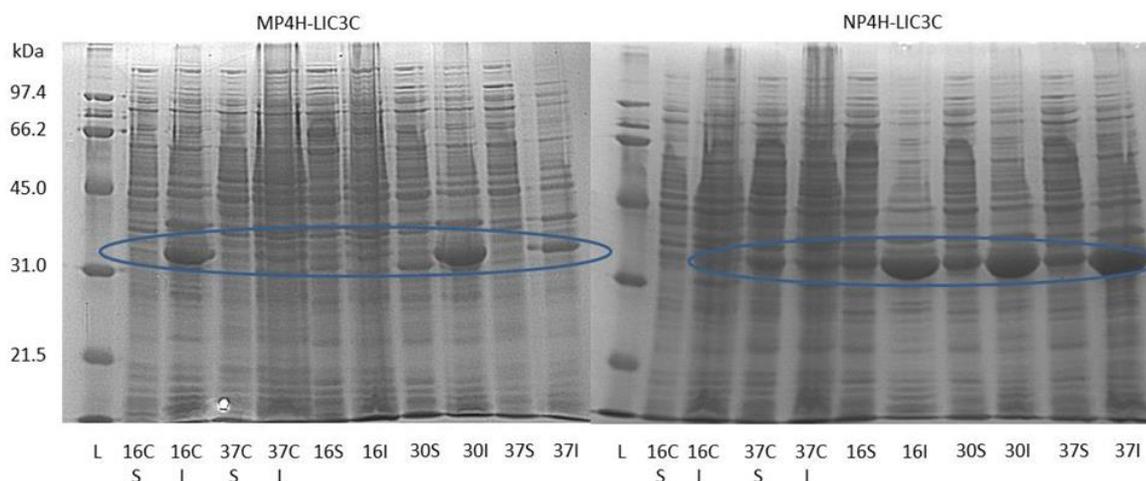
Figure 5.27: 12% SDS-PAGE gels of expression tests of MP4H-LIC3C (left) and NP4H-LIC3C (right) in *E. coli* BL21(DE3) cells. Both MP4H and NP4H are approximately 32 kDa in size. Circled are the bands likely corresponding to the proteins of interest. There appears to have been some leaky expression at 16 °C for MP4H and at 37 °C for NP4H. In general, the greatest soluble expression for MP4H appears to occur at 16 °C, while for NP4H there little difference at all three growth temperatures. S= soluble fraction; I = insoluble fraction.

Protein production was scaled up, with recombinant cells grown in 1 L Terrific Broth (12 g L^{-1} tryptone, 24 g L^{-1} yeast extract, 4 mL L^{-1} glycerol, 10 % v/v 0.17 M KH_2PO_4, 0.72 M K_2HPO_4). Protein expression was induced with addition of 1 M IPTG when the Optical Density of cells reached 2.0, and cells grown overnight at 16 °C. Cells were harvested as standard. Terrific Broth was selected for the purpose of producing a greater amount of cells from a reduced growth volume.

For both proteins, the purification process involved consecutive Ni-affinity purification and Size Exclusion Chromatography (SEC) performed in one work day. Figure 5.28 presents the chromatograms obtained following SEC of MP4H-LIC3C and NP4H-LIC3C respectively. In both cases, a large distinct peak was observed once 50-60 mL of buffer was passed through the column.

146
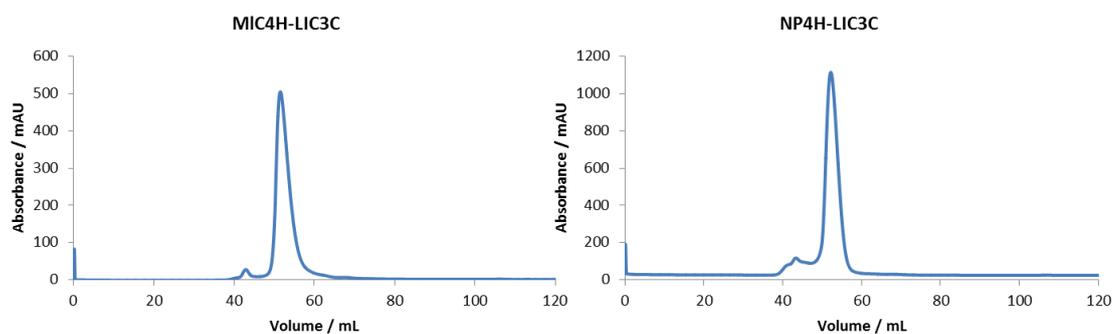
Figure 5.28: SEC chromatograms of MP4H and NP4H purifications. A clear, distinct peak can be seen for both between 50 and 60 mL.

The purity of MP4H-LIC3C and NP4H-LIC3C following each purification step was assessed by 12% SDS-PAGE analysis (Figure 5.29).
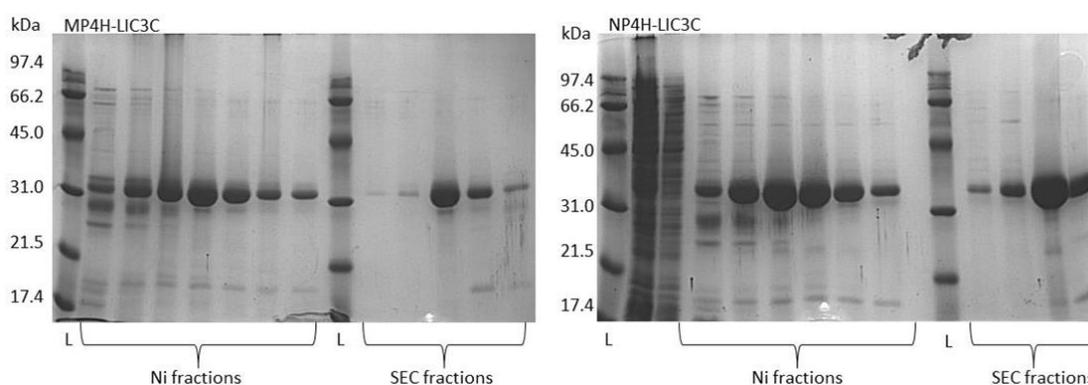
Figure 5.29: 12% SDS-PAGE gels on samples obtained following purification of MP4H (left) and NP4H (right). Both proteins are approximately 32 kDa in size. Purification appears to have been successful for both.

Both DOGDH homologs were successfully purified and found to be stable in isolation. 1 mg per g of cells of MP4H-LIC3C and 3 mg per g of cells of NP4H-LIC3C were isolated respectively.

Crystallisation trials were performed for both at concentrations of 10 mg mL^{-1} using commercial CSS (0.1 M Bis Tris Propane pH 5.5 and 6.5), PACT and INDEX screens. Initial results for NP4H-LIC3C were not very promising as most drops contained precipitate suggesting that the crystallisation conditions were not suitable for this protein. MP4H-LIC3C on the other hand showed more promise with crystalline material observed in two drops (Figure 5.30).
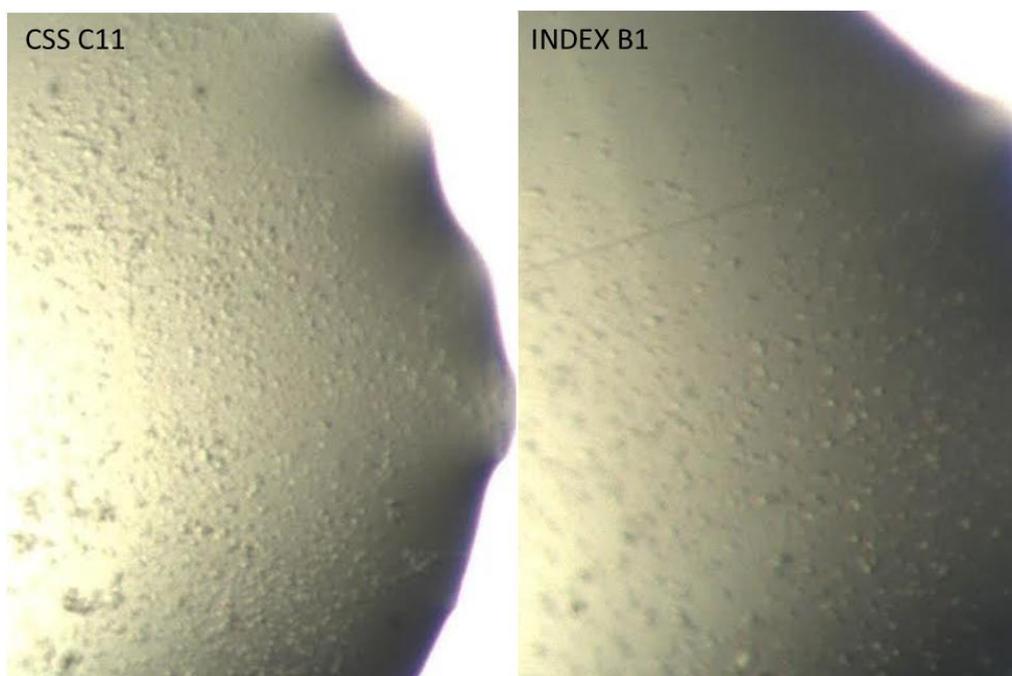
Figure 5.30: MP4H-LIC3C drops that yielded crystalline matter. Microcrystals were observed in CSS C11 (left) and INDEX B1 (right). The crystalline material was highly dispersed throughout the drops.

The conditions that yielded hits were as follows:

(i) CSS C11: 0.2 M KSCN, 10% PEG 8000, 10% PEG 1000, 0.1M Bis Tris Propane pH 6.5

(ii) INDEX B1: 0.1M Bis Tris pH 5.5, 0.5 M Magnesium Formate dihydrate

Though the crystals were not optimal, 24 well hanging drop optimisations of both these conditions were attempted.

The CSS condition was optimised by manipulating the percentage of PEG and the concentration of KSCN, while for the INDEX condition the pH of Bis Tris was changed. Rows 2-4 of each plate differed in the cofactor that they contained (Fe(II), 2-OG and L-ascorbic acid). Unfortunately, the initial hits could not be replicated in the scale ups.

Ligand addition screens, with both added Fe(II) and Co(II) respectively, were also performed for MP4H-LIC3C and NP4H-LIC3C as described for DOGDH-pET22b in *5.3.2*, using varying protein concentrations of both homologs (10 mg mL^{-1} to 50 mg mL^{-1}). However, despite these efforts, it was not possible to obtain diffraction quality crystals with

the best results being identical to what was previously observed for MP4H-LIC3C in the absence of ligands.

5.5 Discussion and Conclusions

DOGDH was initially cloned into the pET-YSBLIC3C vector to introduce an N-terminal hexahistidine tag and purified by Ni-affinity chromatography and SEC to yield unstable, degrading protein, regardless of attempts to prevent this. The continued isolation of unstable DOGDH-LIC3C was not optimal due to the fact that stability is a key prerequisite for crystallisation.[236] This is primarily the result of biomolecular crystallisation requiring the ordered self-assembly of molecules, with components within the asymmetric unit of a crystal needing to be of the same kind and shape.[236] DOGDH-LIC3C was therefore abandoned as a crystallisation target due to its stability issues.

DOGDH was next cloned into the pET22b vector to generate a C-terminal hexahistidine tagged construct. This construct was successfully purified at moderately high yield and 96-well crystal screening was performed using commercial screens with varying concentrations of protein. The resulting drops were either empty or contained precipitate or cell debris. This hinted towards continued DOGDH stability issues in the crystal drops themselves which likely hindered the protein's ability to crystallise and caused it to precipitate out of solution.

It was hypothesised that stability issues and the inability of DOGDH-pET22b to crystallise may have arisen from the absence of key enzyme components such as Fe(II), 2-oxoglutarate and/or proline in the pure enzyme and crystallisation solutions. The addition of ligands is thought to aid crystallisation by enhancing the conformational stability of the protein of interest.[237] To test this, ligand addition crystallisation trials were performed with Fe(II), 2-oxoglutarate and L-proline added to the mother liquor of the crystallisation solutions. This proved successful with nanocrystal showers observed in a single 96 well sitting drop, however, this drop could not be replicated in a 24-well hanging drop format. Attempts were also made to directly add Fe(II) and its mimetic Co(II) directly to purified DOGDH-pET22b prior to concentrating for crystallisation trials, however, this proved

unsuccessful with Fe(II)SO$_4$ precipitating out of solution and the addition of Co(II) resulting in the protein's almost instantaneous precipitation.

Little success was observed with crystallisation trials of DOGDH-LIC3C and DOGH-pET22b in their native state. As a result, the work that followed focussed on modifying the protein surface of DOGDH, a well-established strategy for protein crystallisation, by both chemical and mutagenesis methods. Reductive methylation of surface exposed residues was the chemical method attempted and proved to be inconclusive. Surface entropy reduction mutagenesis of surface exposed amino acids was the second attempted modification. This proved somewhat successful with hits observed in the 96-well sitting drop format with one of these hits successfully replicated in a 24-well hanging drop format. Attempts were made to further optimise this to little success. Ligand addition trials were performed on DOGDH-SER which yielded an alternative globular crystal morphology which also was optimised with little success even following the addition of stabilising detergents.

The final work in this chapter shifted towards homologs of DOGDH, with the hope of using any potential structural data to model DOGDH itself. Two homologs, MP4H and NP4H, were selected following a BLAST search and cloned into the pET-YSBLIC3C vector. Purifications of both homologs yielded stable proteins at high yields. Crystallisation trials were performed and two drops of interest observed for MP4H. Unfortunately, scale-ups were not successful.

In conclusion, extensive attempts were made to obtain diffraction quality crystals of DOGDH. However, the protein proved to be mostly unstable, at best generating microcrystals that could not be optimised on a larger scale.

Chapter 6

Reactions with Proline Hydroxylases

6.1 Introduction

Proline hydroxylases are a subfamily of the broader 2-oxoglutarate dependent oxygenase superfamily of enzymes. Microbial proline hydroxylases have the advantage over their mammalian equivalents of being able to catalyse the hydroxylation of proline in isolation, doing so with high regio-, enantio- and stereoselectivity.[176] In this project, hydroxylases capable of producing *trans*-4-hydroxy-L-proline (DOGDH), *cis*-3-hydroxy-L-proline (StP3H) and *cis*-4-hydroxy-L-proline (MlC4H) were successfully cloned and expressed in *Escherichia coli* providing a platform for their biocatalytic application.

The described microbial proline hydroxylases have the interesting feature of being able to hydroxylate substrates other than proline with only the secondary amino acid moiety deemed essential for activity.[176] Such a feature opens up a wealth of synthetic potential by providing access to an unprecedented number of hydroxylated chiral synthons. L-pipecolic acid (Figure 6.1) is one such alternative substrate that can be hydroxylated by proline hydroxylases to hydroxypipecolic acid, with enantiomer formed depending on the hydroxylase utilised.[238] Hydroxypipecolic acids are widely occurring amino acids in nature that are also building blocks of a number of peptide antibiotics, terpenoids and alkaloids, though this is dependent on the enantiomer of the hydroxypipecolic acid constituent.[238-241]
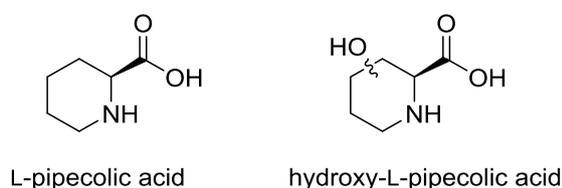
Figure 6.1: L-pipecolic acid and hydroxy-L-pipecolic acid.

The high selectivity and activity of proline hydroxylases together with their ability to hydroxylate substrates other than proline and their evolutionary potential for enhanced or modified activity towards these substrates grants these enzymes significant industrial

potential.[205] In order for such potential to be achieved, it is important to characterise the reactions they catalyse, identify key reaction parameters and investigate their structure to try and elucidate the determinants for selectivity.

6.2 Aims

Biotransformations will be performed with L-proline as substrate using hydroxylase targets as purified proteins and in whole cells in order to compare and contrast their performance. For the purposes of whole cell reactions, cells expressing the hydroxylase targets must be generated by fermentation and the presence of the targets confirmed. Biotransformations must be performed at a range of scales and key reaction parameters identified for the purposes of optimisation. A protocol must be developed for recovering and quantifying the hydroxylated products following reaction. Finally, a platform must be identified and utilised for the purpose of screening a panel of alternative substrates with the hydroxylases.

6.3 Cell-free Biotransformations

As a proof of principle test, biotransformations of L-proline with purified DOGDH-LIC3C, StP3H-LIC3C and MlC4H-LIC3C were attempted.

In order to generate the hydroxylase targets, all three were expressed in *E. coli* BL21(DE3) cells, harvested and purified by Ni-affinity chromatography and SEC as standard. The SEC chromatograms for these purifications are presented in Figure 6.2.
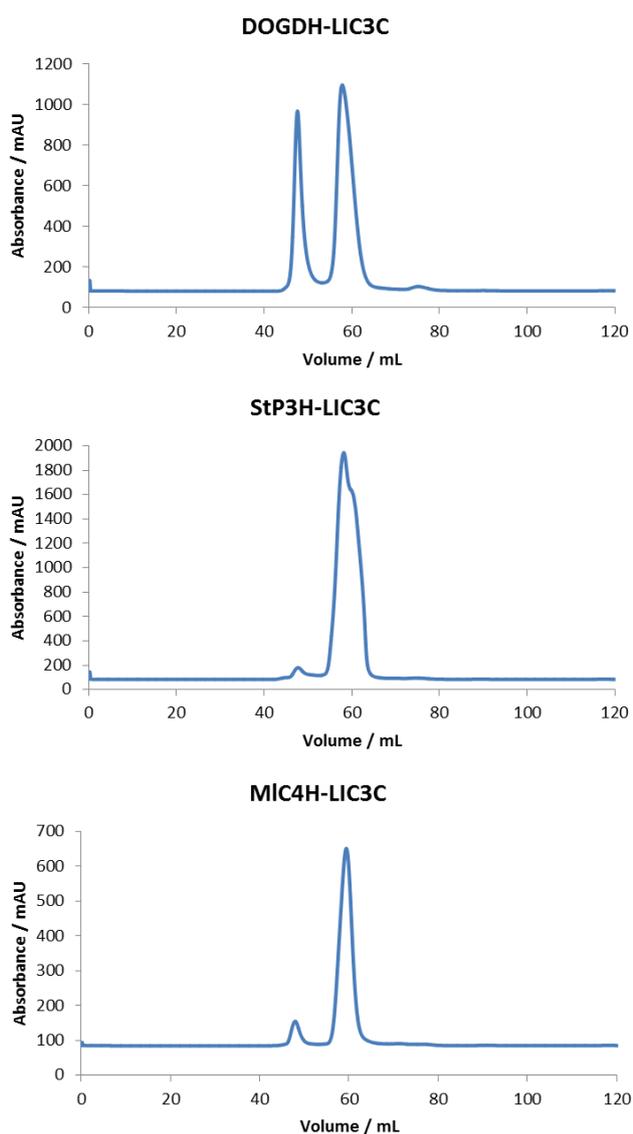
Figure 6.2: SEC chromatograms of DOGDH-LIC3C, StP3H-LIC3C and MlC4H-LIC3C purifications respectively. A single, abundant peak between 55-65 mL was observed in the StP3H and MlC4H purifications. Two peaks were observed DOGDH-LIC3C suggesting multiple products were isolated. Please note the differences in absorbance scales. Absorbance was measured at 280 nm.

The purifications appeared successful but, interestingly, two peaks were observed in the DOGDH chromatogram. Ni and SEC fractions for each hydroxylase were analysed by SDS-PAGE (Figure 6.3).
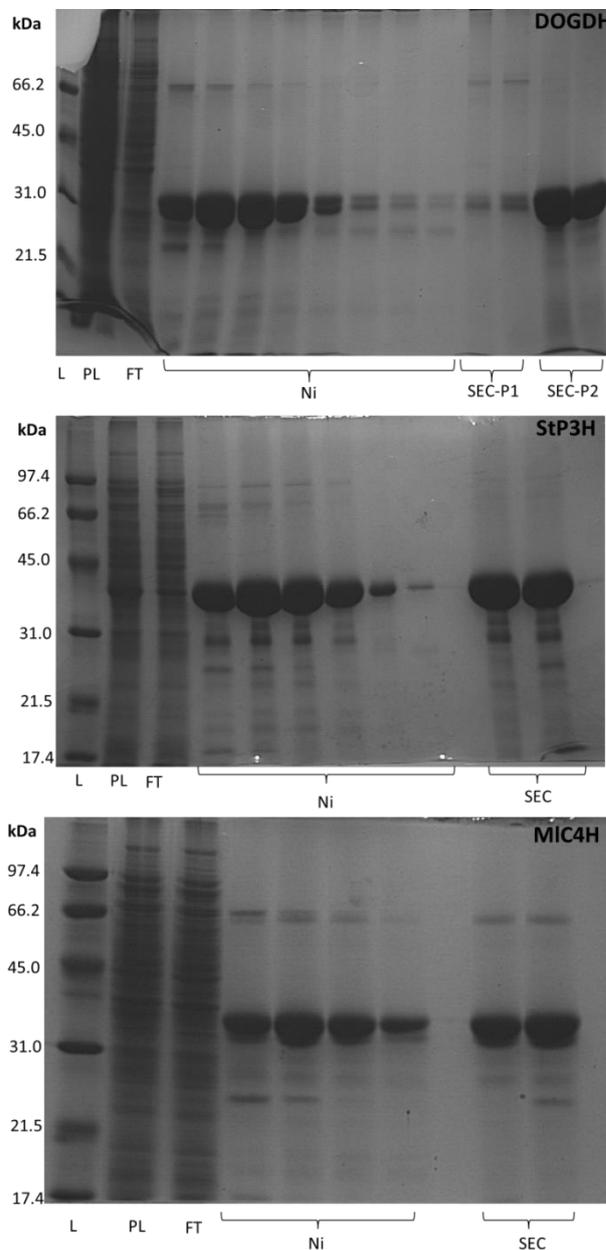
Figure 6.3: 12% SDS-PAGE gels of hydroxylase purification fractions. Top: DOGDH-LIC3C; a thick band at approximately 34 kDa likely corresponding to DOGDH-LIC3C can be seen in each purification fraction, most of the protein appears in the second SEC peak. Middle: StP3H-LIC3C; a thick band at approximately 37 kDa likely corresponding can be seen in both the Ni and SEC fractions. Bottom: MlC4H-LIC3C two bands at approximately 35 kDa are present in the Ni and SEC fractions suggesting that MlC4H was degrading.

The above gels suggested that, in general, all three hydroxylases had been successfully purified. However, DOGDH-LIC3C and MlC4H-LIC3C appeared to be undergoing degradation as evident from the multiple bands present in the fainter DOGDH purification

154

fractions and in main purification fractions of MlC4H. This was to be expected for DOGDH-LIC3C as stability issues were previously observed. StP3H-LIC3C appeared to be the most stable and well-expressed of the three targets as evident from the thick bands and the fact that its SEC chromatogram peak maxed out the UV absorbance measurement limits. Regardless of the stability issues, biotransformation work proceeded with these purified samples as this was a proof of concept test.

Whole cell biotransformations of L-proline using the SEC fractions of the three hydroxylases were performed at a 4 mL scale in 50 mL falcon tubes as described in *2.2.8.2* with the transformation mixture consisting of 0.6 mg mL^{-1} enzyme and Reaction Mix 1 (see *2.1.3*). Samples were taken at regular intervals for 24 h and analysed by HPLC. Prior to derivatisation, reaction samples were centrifuged at 16300 *g*, the supernatant retained, heat deactivated by heating in a block at 90 °C for 10 min and centrifuged once more. This heat deactivation step was performed in order to inactivate any protein and biological samples in the mixtures. Samples were derivatised using the FMOC-Cl assay developed in Chapter 3 and analysed using HPLC Program 4 described in Table 2.27.

The percentage conversation of proline to hydroxyproline was calculated by using the following equation:

$$\text{Percentage conversion} = \frac{\text{peak area of hydroxyproline}}{\text{peak area of proline} + \text{peak area of hydroxyproline}} \times 100\%$$

Results and Discussion

The percentage conversions of proline to hydroxyproline observed following 24 h of reaction time for each enzyme are summarised in Table 6.1.

Table 6.1: Maximum conversion observed for DOGDH, StP3H and MlC4H with L-proline

Enzyme	Conversion (%)
DOGDH-LIC3C	-
StP3H-LIC3C	23
MlC4H-LIC3C	-

Of the three hydroxylase targets, only StP3H showed evidence of activity in the pure enzyme state with approximately 23% conversion to *cis*-3-hydroxyproline observed after 24 h. This may have been due to the fact that this enzyme appeared the most stable of the three following purification. DOGDH-LIC3C and MlC4H-LIC3C on the other hand showed no evidence of activity as no product peaks were observed in any of the analysed samples. A likely reason for the absence of activity in DOGDH and MlC4H was that they presented greater stability issues following purification. It must be noted that both also rapidly precipitated in solution within minutes of the reactions commencing.

In summary, only StP3H appeared active in the purified state with DOGDH and MlC4H showing severe stability issues. In a sense, this finding was not surprising as these proline hydroxylases are multi-component enzymes that require Fe(II), 2-oxoglutarate (2-OG), ascorbic acid, a mean of cofactor (Fe(II)/2-OG) regeneration and enzyme stabilisation in order to successfully function.[176] Achieving such a state of equilibrium is very difficult when working in a cell free environment as additional components such as oxidoreductase enzymes for cofactor regeneration as is the case for P450s and the enzyme itself needs must also be stabilised using further complex means such as by immobilisation.[26, 47]

For the purposes of redox biocatalysis, particularly reactions in which molecular oxygen is required, whole cells continue to remain the ideal system for performing such reactions.[50, 242] This is due to the cell itself possessing the capacity to provide cofactors, regeneration strategies and an environment for the stabilisation of enzymes.[242] As a result, whole cell biotransformations were the next to be investigated.

6.4 Generation of Cells

In order to perform biotransformations using whole cells expressing DOGDH, StP3H and MlC4H, it was necessary to first produce cells. Cells were generated in 1 L fermenters (two fermentations for each hydroxylase target) using a Fed-Batch Fermentation protocol developed by N. Triggs and M. Bycroft at Dr Reddys Laboratories called 'Glucose Fed 5.0' (see *2.2.8.1*). The output data for a representative vessel is presented in Figure 6.4.
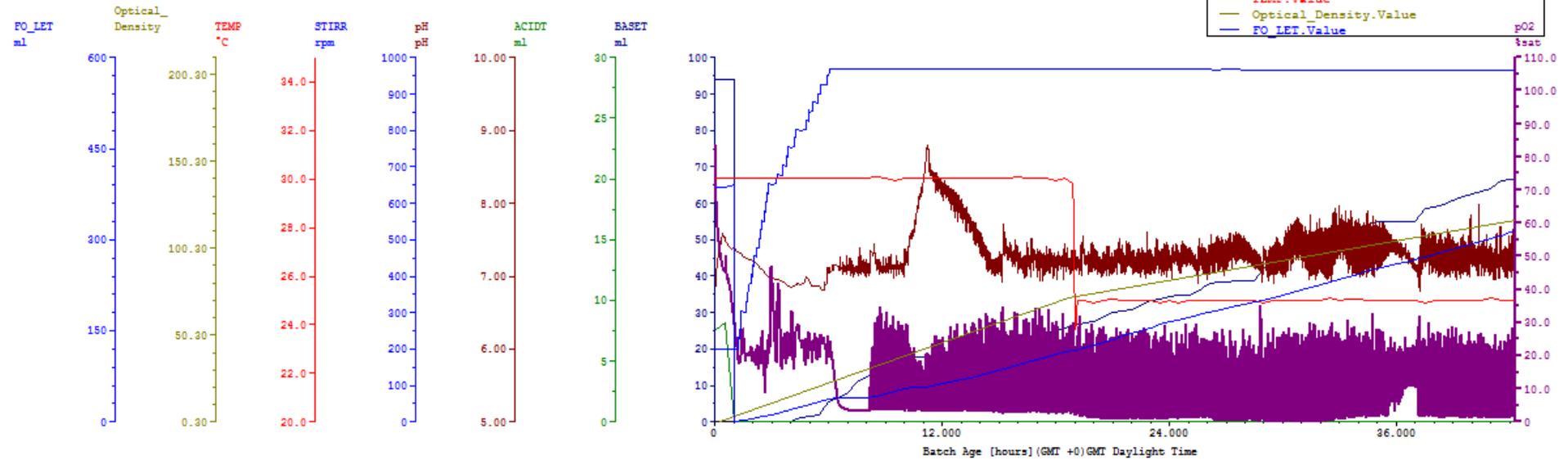
Figure 6.4: Fermentation output for DOGDH reaction 1. Variables that were monitored included: dissolved oxygen content (purple), Optical Density (yellow), vessel temperature (red), stirring rate (RPM) (light blue), vessel pH (maroon), amount of acid (green) and base (violet) added respectively. The fermentation was performed for 45 h. The stiring rate (light blue) gradually increased to 1000 RPM over the first 6 h after which it stayed at this value throughout the process. During this period, the Optical Density of the material (yellow) gradually increased to 100.0. The growth temperature (red line) was 30.0 °C for the first 18 h after which it was reduced to 25.0 °C. In order to maintain the pH at 7.2, very little acid was added following 2 h of growth time with only base supplemented into the mixture (violet). The dissolved oxygen value (purple) dropped from 60% to 20% saturation over the first 2 h, remained fairly constant between 2-8 h after which it dropped and stabilised to about 10% saturation.

The fermentation trace in Figure 6.4 is the output for one of the DOGDH growth vessels, the other five fermentations looked identical to this, so have not been presented. The optical density of the growth mixture gradually increased to approximately 100.0 during the 45 h growth period. The growth temperature was 30 °C for the first 18 h and reduced to 25 °C for the remaining time. The pH of the growth mixture was maintained at 7.2 by continued addition of 28% ammonium hydroxide. The dissolved oxygen content was initially at 60% saturation, dropping to 20% in the first two hours and staying as this level for the next 6 h. After this point, the dissolved oxygen dropped to 10% saturation and remained at this level until the end of the growth period.

Following fermentation, the final OD_{600} was approximately 100 for each growth vessel. Fermentation mixtures were centrifuged at 5000 g for 60 min and the cell pellets retained, weighed and stored at -20 °C until required. Table 6.2 summarises the final yields of cells expressing each enzyme of interest following fermentation.

Table 6.2: Quantity of cell paste produced by fermentation. The total fermentation volume for each is the sum of the two reactions.

Construct	Cell paste (g)	Total fermentation volume (mL)
DOGDH-LIC3C	330.5	1650
StP3H-LIC3C	303.7	1700
MlC4H-LIC3C	274.9	1600

The presence of the required proteins in the produced cells was assessed using Western Blot analysis. Pellets were first resuspended in 50 mM potassium phosphate buffer pH 7.5 such that the final concentration was 10% w/v. Resuspensions were sonicated at 13.0 Microns for 15 cycles of 30 s on and 30 s off, and lysates centrifuged at 10733 g for 50 min. The supernatant of each was collected and treated as the soluble protein fraction, while the remaining pellets (insoluble fractions) were resuspended in 10 mL phosphate buffer pH 7.5. Fractions were diluted 1:10 in phosphate buffer pH 7.5 and analysed on a 12 % SDS-PAGE gel. Following electrophoresis, the gel underwent Western Blot analysis using an Anti-His antibody.

The output Western Blot of the fermentation cell pellets is presented in Figure 6.5. In general, it appeared as though the fermentations had been very successful as evident from the presence of bands corresponding to each hydroxylase target in their respective cells. There also appeared to be an equal amount of soluble and insoluble expression.
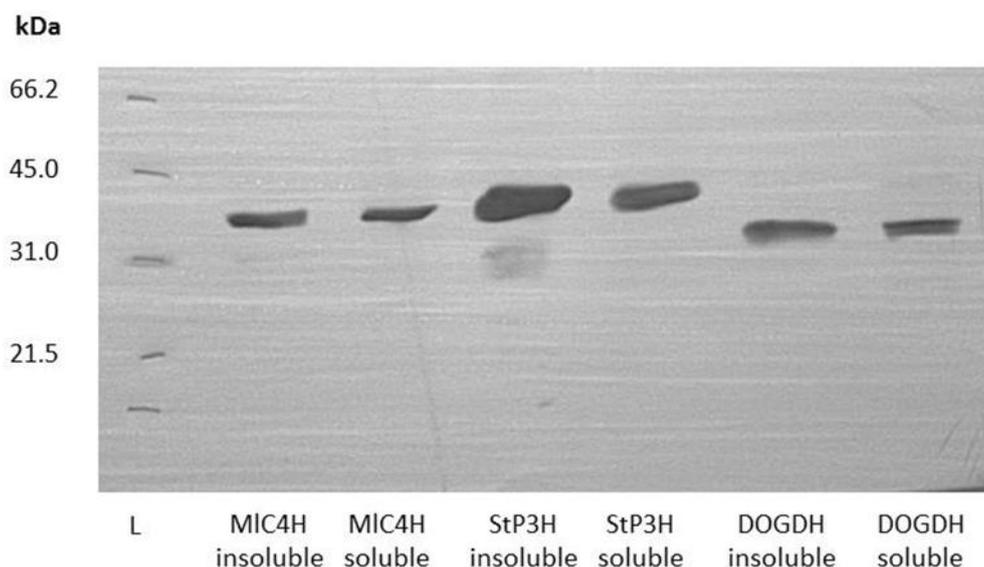
Figure 6.5: Western Blot of pellets produced following fermentation. A single band can be seen in each of the lanes at the desired size. This suggests that DOGDH-LIC3C, StP3H-LIC3C and MlC4H-LIC3C were successfully produced by fermentation.

In conclusion, DOGDH-LIC3C, StP3H-LIC3C and MlC4H-LIC3C were successfully expressed in *E. coli*. Cells expressing these targets were prepared by fermentation at a satisfactory yield and the presence of targets confirmed by Western Blot analysis. This validated and demonstrated the scalability of this process.

6.5 Whole Cell Biotransformations with L-proline

6.5.1 Shake Flask Reactions

6.5.1.1 Preliminary Reactions

Having successfully produced cells expressing DOGDH-LIC3C, StP3H-LIC3C and MlC4H-LIC3C respectively by fermentation, the next goal was to attempt whole cell biotransformations with L-proline as substrate in shake flasks. For the purposes of these biotransformations, 2 L glass Erlenmeyer baffled flasks were selected as the reaction vessels. Baffled flasks were chosen because the baffles at the base of the flasks are thought

to aid oxygen transfer during shaking which can be crucial for cell health and hydroxylase activity as these enzymes are believed to be highly dependent on molecular oxygen for catalysis.[176, 243]

Protocol

Reactions of L-proline were performed with DOGDH, StP3H and MlC4H respectively at a 150 mL scale. The reaction mixture consisting of 10% w/v frozen cell paste and 150 mL of biotransformation Reaction Mixture I (see Table 2.6) prepared in 50 mM potassium phosphate buffer pH 7.5. This ratio of reaction components were adopted from previous work performed by L. Roper. The flasks were placed in an orbital incubator set to 30 °C with shaking set to 180 RPM. 20 µL aliquots were taken from the reaction vessels at regular interval and snap frozen in liquid nitrogen for storage at -20 °C until required for analysis.

Prior to precolumn derivatisation, reaction samples were centrifuged at 16300 *g* for 10 min, the supernatant retained and deactivated by heating at 90 °C for 10 min. Samples were derivatised using the FMOC-Cl assay developed in Chapter 3 (see *3.5.3*). Derivatisation reactions were performed in glass vials and samples were filtered prior to transfer into glass vials for LC-MS analysis.

Derivatised samples were analysed by LC-MS using the LC-MS long acidic program described in Table 2.29.

Results and Discussion

In the case of LC-MS analysis, three peaks were expected in samples: an FMOC-proline peak at approximately 7.4 min, an FMOC species peak at 7.2 min and an FMOC-hydroxyproline peak between 6.6 min and 6.8 min.

LC-MS chromatograms of illustrative time points in DOGDH-LIC3C and StP3H-LIC3C biotransformations (using the long acidic program) are shown in Figure 6.6 and Figure 6.7. The peak identities have been stated and were confirmed by supporting MS data.
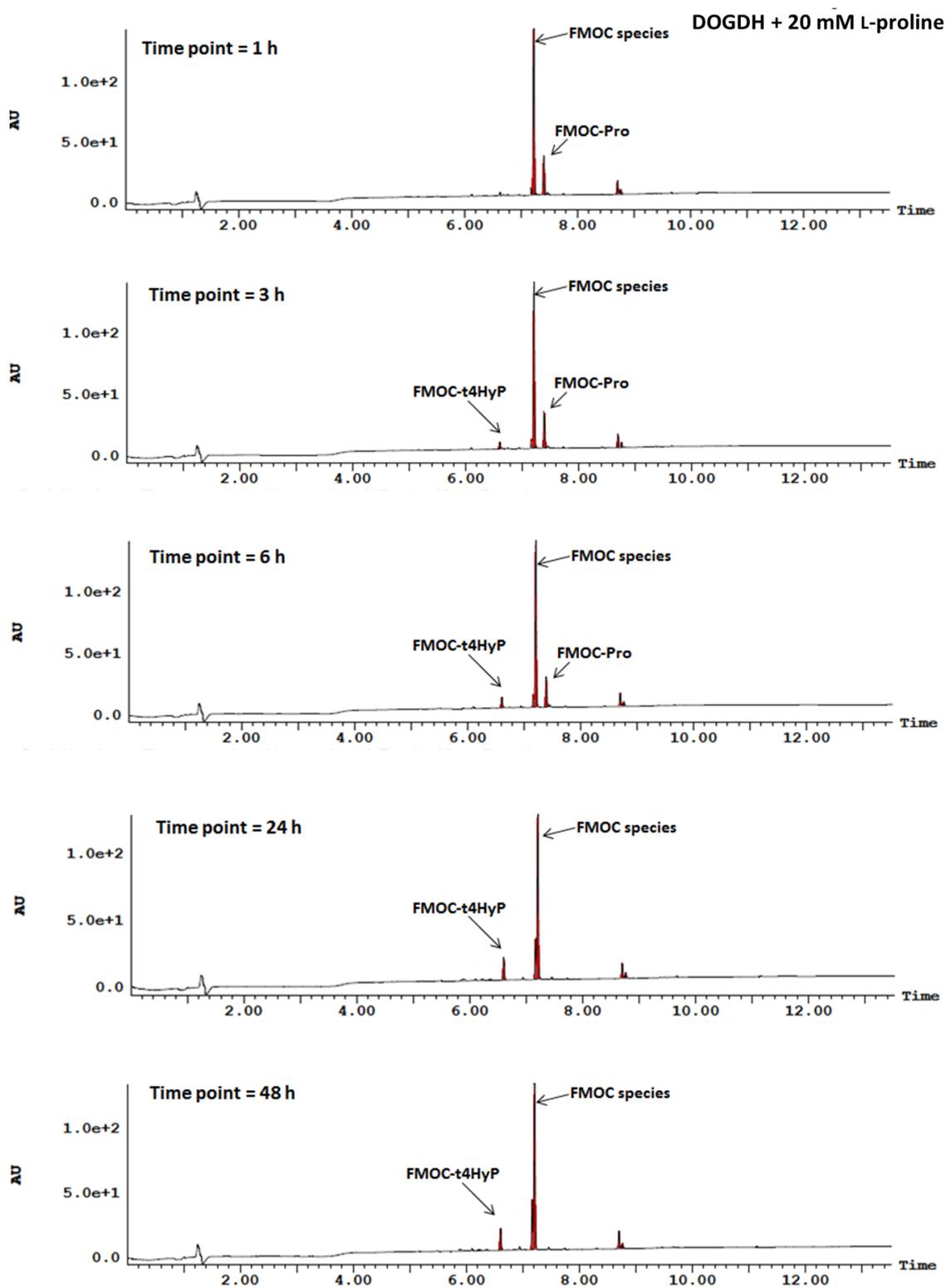
Figure 6.6: LC-MS chromatograms of DOGDH-LIC3C reactions with 20 mM L-proline. The peak identities have been labelled and were confirmed by supporting MS data. FMOC-proline was observed at around 7.4 min, an FMOC species was observed at 7.2 min while the FMOC-*trans*-4-hydroxyproline species was observed at 6.6 min. The chromatograms show that the t4HyP peak increases in amplitude over time, peaking at 24 h while the Pro peak decreases and is completely absent after 24 h.
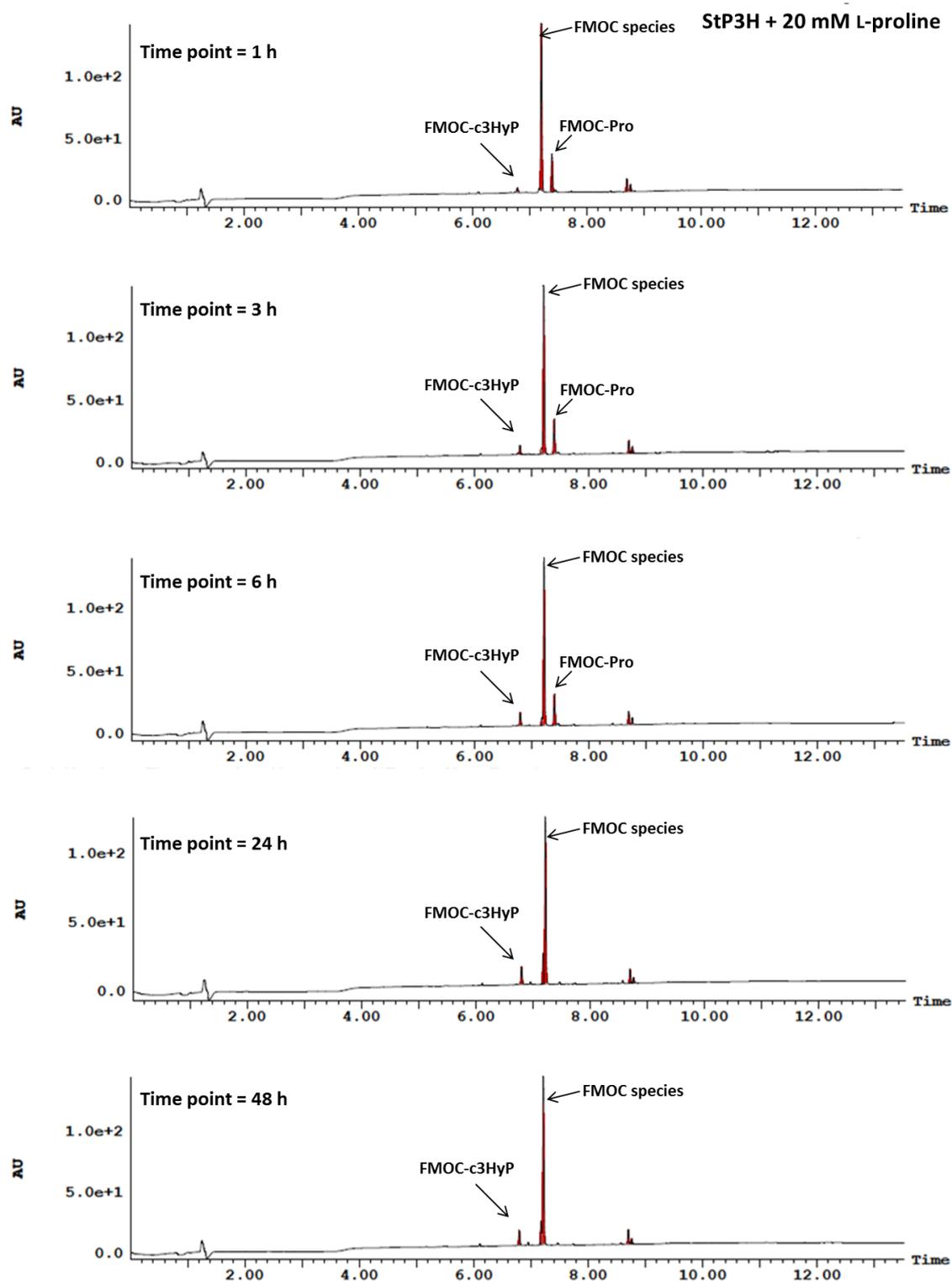
Figure 6.7: LC-MS chromatograms of StP3H-LIC3C reactions with 20 mM L-proline. The peak identities have been labelled and were confirmed by supporting MS data. FMOC-proline was observed at around 7.4 min, an FMOC species was observed at 7.2 min while the FMOC-*cis*-3-hydroxyproline species was observed at 6.8 min. The chromatograms show that the c3HyP peak increases in amplitude over time, peaking at 24 h while the Pro peak decreases and is completely absent after 24 h.

The chromatogram data for the DOGDH, StP3H and MlC4H biotransformations suggested that the biotransformations were successful as evident from the fact that the peaks corresponding to the hydroxyproline products increased in amplitude over the reaction time course, reaching a maximum following 24 h of reaction. Interestingly, proline was absent in both the 24 h and 48 h samples of these three reactions. It was, however, difficult to conclude that the reactions had gone to completion due to the fact that final peak amplitudes of the hydroxyproline products present in isolation were only a fraction of the height of the starting proline samples in the 1 h samples, especially for the MlC4H in which the final peak area of the product was roughly 20% that of the starting proline. This suggested that the proline may have been scavenged by the cells during the reaction and will be discussed later in this chapter. The peak area data for these biotransformations have been visually summarised in Figure 6.8..
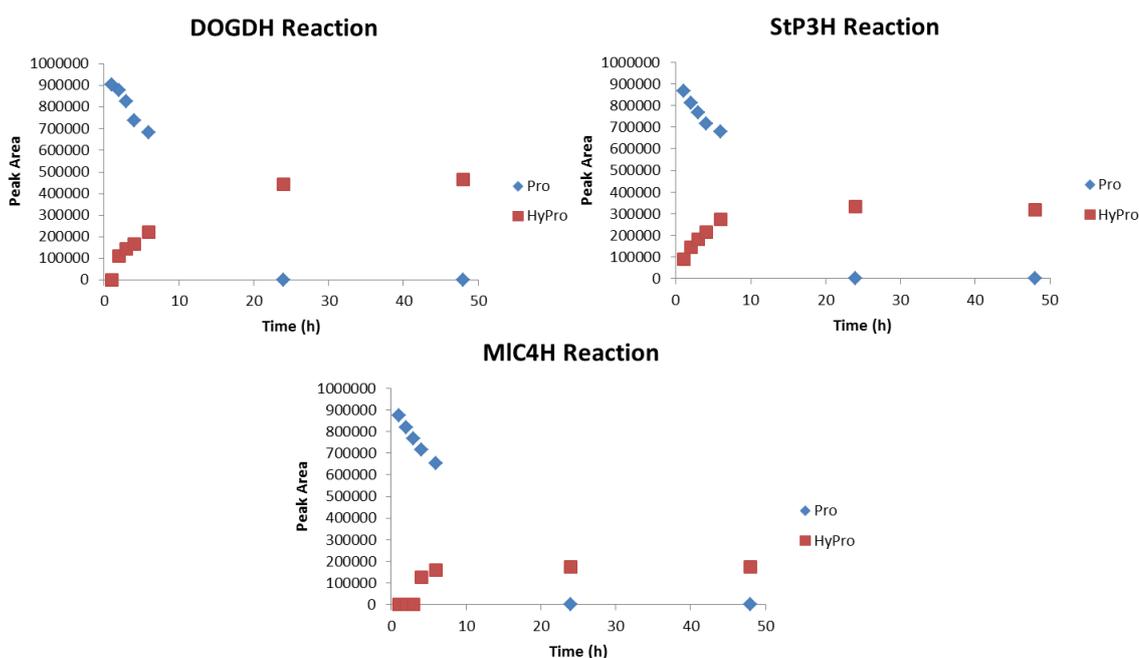
Figure 6.8: Charts summarising the LC-MS data for biotransformations of L-proline with DOGDH, StP3H and MlC4H respectively. The peak areas of proline (blue) and hydroxyproline (red) in each time point sample have been measured and visualised. DOGDH and StP3H appeared to perform similarly reaching an excess of hydroxyproline after 24 h with close final product peak areas. MlC4H also reached an excess of product after 24 but the peak area of product was approximately half the size of what was observed in DOGDH and StP3H. Interestingly, proline appeared absent in all three reactions after 24 h regardless of performance.

In general, DOGDH and StP3H appeared to perform in an almost identical manner with the peak area of the hydroxylated product gradually increasing over the first 6 h of reaction,

reaching its maximum after 24 h. A similar pattern was also seen for MlC4H, however, the final product peak area was approximately half of that observed in the DOGDH and StP3H reactions. In all three reactions, the amount of proline observed in the reaction samples gradually decreased from a maximum at the beginning to a complete absence after 24 h. The final peak area of the products in the three reactions was only a fraction of the size of the initial proline peak area in the first time point suggesting that not all of the proline had been utilised in the reactions despite its absence in the final time points, supporting the theory that the substrate was being scavenged by the cells by some means.

6.5.1.2 Optimisation of Reaction Parameters

Initial biotransformations in shake flasks were very promising in the case of DOGDH and StP3H, with the former showing improved performance compared to its purified equivalent. However, as the maximum conversion was observed after 24 h had elapsed in the reaction, it was decided that the next goal would be to optimise the reaction in order to enhance activity.

Consultation of the literature suggested that the ratio of components in the biotransformation Reaction Mix (20 mM L-proline, 20 mM 2-oxoglutaric acid, 10 mM L-ascorbic acid and 2 mM $Fe(II)SO_4$, pH 7.5) be modified. Klein and Hüttel investigated the effect that varying concentrations of $Fe(II)SO_4$ and L-ascorbic acid had on the activity of proline hydroxylases.[194] The two found that the concentration in the mixture of ascorbic acid should not exceed 3 mM, as ascorbic acid can have an inhibitory effect on the hydroxylase at concentrations exceeding this.[194] $Fe(II)SO_4$ meanwhile was found to optimally aid function at concentrations ranging from 0.1 mM to 0.5 mM depending on the hydroxylase, with a *trans*-proline-4-hydroxylase showing the greatest activity with 0.5 mM $Fe(II)SO_4$.[194]

It was also hypothesised that the concentration of 2-oxoglutaric acid in the mixture should be increased relative to substrate as it may have been limiting during the reaction time course perhaps due to being utilised in alternative processes by the cells.

As a result of these considerations, the concentrations of components in the Reaction Mix were modified such that it contained 20 mM L-proline, 30 mM 2-oxoglutaric acid, 3 mM L-ascorbic acid and 0.5 mM Fe(II)SO$_4$ (Reaction Mix II in Table 2.7).

Reactions of L-proline with DOGDH were performed in baffled flasks as before but with Reaction Mix II. Samples were taken at regular intervals, frozen until required and derivatised with FMOC-Cl for analysis by LC-MS.

Figure 6.9 presents selected illustrative LC-MS chromatograms (analysed with the long-acidic program) of time point samples in the reaction of DOGDH-LIC3C with 20 mM L-proline using the modified reaction mix. Peak identities were assumed based on the supporting MS data.
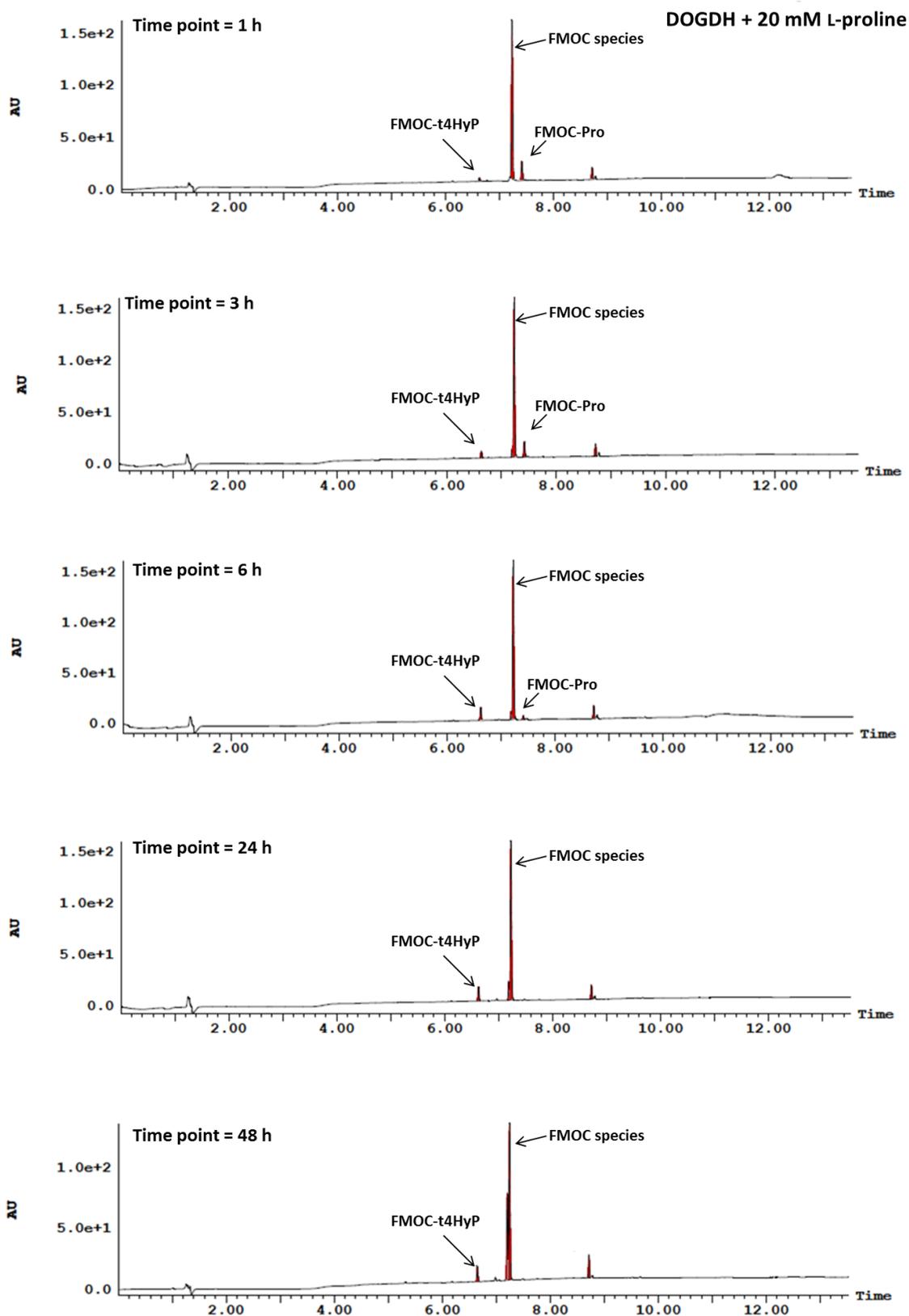
Figure 6.9: LC-MS chromatograms of selected time points in the reaction between DOGDH-LIC3C and 20 mM L-proline using the modified reaction mix. FMOC-proline = 7.4 min; FMOC species = 7.2 min; FMOC-t4HyP = 6.6 min. These chromatograms suggested that the reaction was occurring more rapidly with the modified reaction mix as evident from the increase in amplitude of the t4HyP relative to the Pro peak.

At a glance, the chromatograms suggested the reaction was proceeding more swiftly with the modified reaction as evident from the rapid increase in the amplitude of the t4HyP peak relative to the proline peak. The relative peak areas (in the LC-MS chromatograms) of proline and t4HyP in the reaction samples of the 48 h time course are visually summarised in the chart in Figure 6.10.
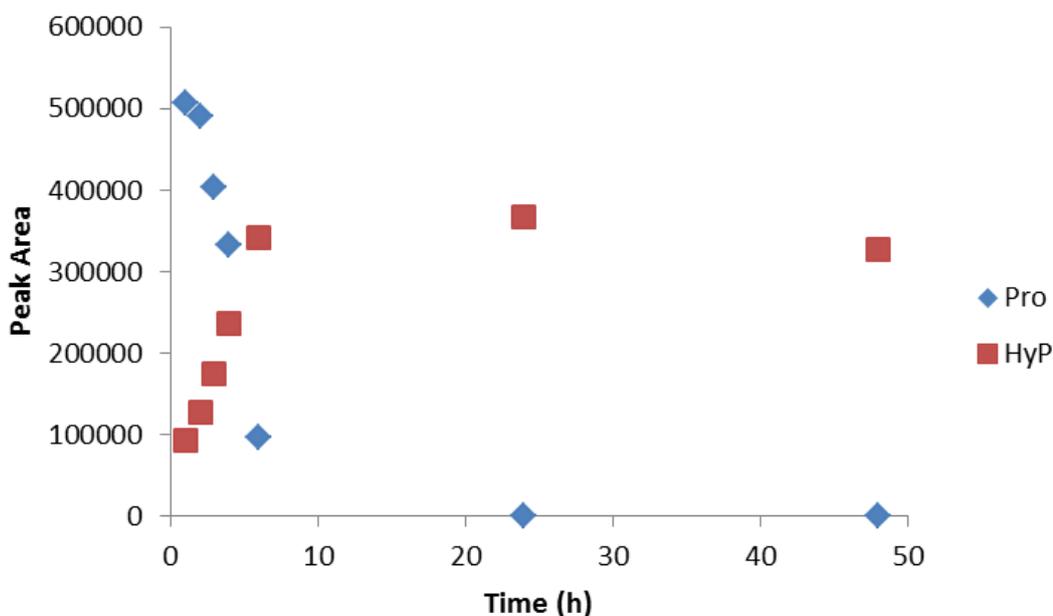
Figure 6.10: Time-course chart of the relative peak areas of proline (blue) and t4HyP (red) in the samples obtained from the reaction between DOGDH and L-proline with the modified reaction mix following LC-MS analysis. In general it was assumed that the reaction appeared to proceed more swiftly as the amount of t4HyP in the mixture rapidly increased in the first 8 h of the reaction.

Figure 6.10 presents the relative peaks area of proline and t4HyP in the LC-MS chromatograms of the reaction time point samples. The figure suggests that the reaction appeared to proceed more rapidly as evident from the sharp increase of t4HyP within the first 8 h of the reaction. From this data, it was concluded that the modified reaction mix likely optimised the reaction speed; as a result, this mix was employed in the biotransformations that followed.

It must be again be noted that following 24 h of reaction only t4HyP was observed in the samples with a complete absence of proline. However, as discussed above, it could not be concluded that the reaction proceeded to completion as the final peak amplitude of t4HyP was only a fraction the size of the starting peak amplitude of proline. This suggested that all of the starting proline was not available for reaction as evident from its gradual decrease

and complete absence at the end of the reaction course. This will be discussed further later in this chapter.

6.5.1.3 Product Recovery

It was essential that a protocol be developed for isolating products from reaction mixtures as this would allow us to quantify reaction yields and confirm the products obtained. The protocol used was adapted from one employed by Johnston and colleagues.[209] This protocol involved the protection of proline and hydroxyproline using di-tert-butyl dicarbonate (Boc) followed by isolation via a series of extraction steps. Figure 6.11 presents a scheme of this reaction.
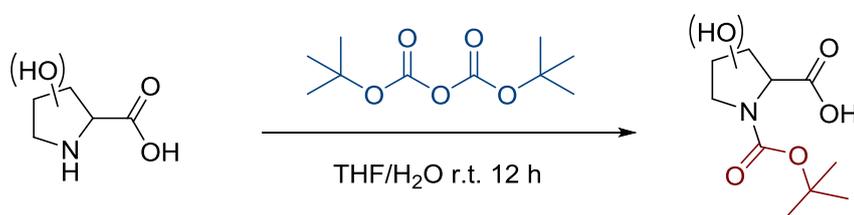
Figure 6.11: Boc protection reaction of proline (or hydroxyproline). Proline (or hydroxyproline) is reacted with di-tert-butyl dicarbonate (blue) to give a protected product (protection group in maroon) that can be isolated following a series of extraction steps.

The protection protocol was as follows: following reaction, cellular matter was centrifuged for 60 min at 10000 g and the supernatant retained. The supernatant was deactivated by heating at 90 °C for 10 min, after which, the precipitate was removed by filtration with celite. The filtered celite cake was washed with water and the wash combined with the filtrate. The pH of this solution was adjusted to 2.0 and the mixture left overnight at 4 °C. The resulting precipitate was filtered once more using celite, following by a wash and pooled. This solution was further cleaned up by performing an extraction with ethyl acetate in which the aqueous phase was retained. The pH of the aqueous phase was adjusted to 12.5 and 2 molar equivalents of Boc anhydride (dissolved in THF) relative to the starting concentration L-proline added to it. The reaction was left overnight with stirring. The reaction was quenched by adjusting the pH to 2.5 and the acidified solution extracted with ethyl acetate in which the organic phase was retained.

The presence of proline, hydroxyproline and their protected equivalents was assessed by thin layer chromatography (TLC). Figure 6.12 presents a reproduction of a TLC conducted

with Boc protected proline and hydroxyproline standards. A clear distinction could be made between bands corresponding to protected and unprotected species respectively, creating a means by which the protection reactions could be monitored.
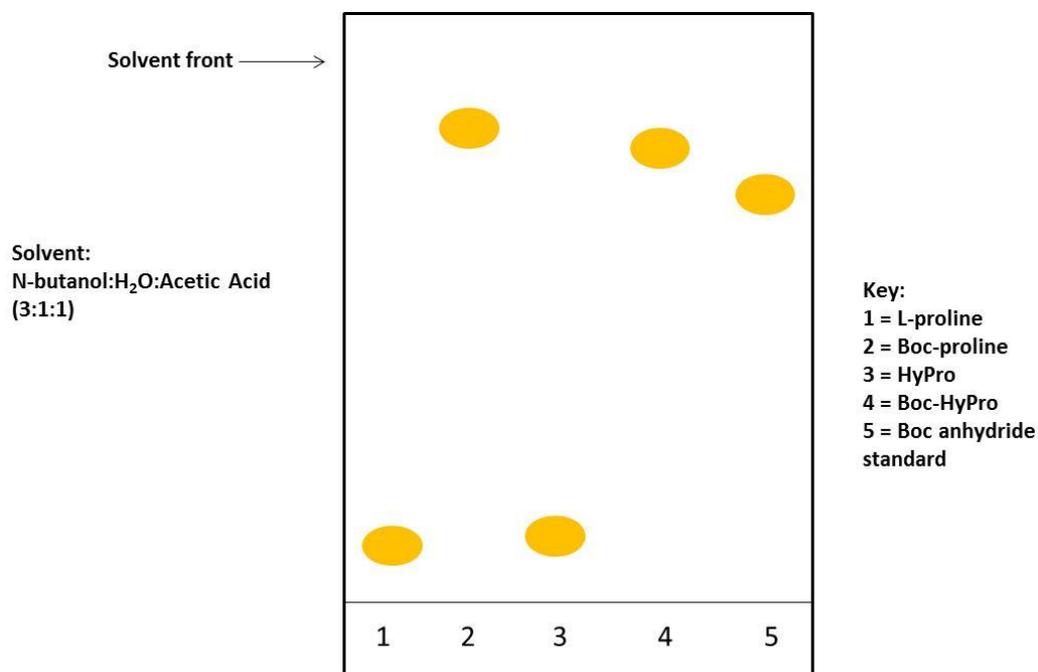
Figure 6.12: TLC reproduction of standards. The lanes can be identified in the key. There is a clear distinction between protected and unprotected species with the former migrating further on the plate. The mobile phase solvent was N-butanol:H2O:acetic acid (3:1:1).

TLC analysis of protected supernatants obtained following reactions of DOGDH with proline in **6.5.1.2** and StP3H and MlC4H with proline in **6.5.1.1** showed an absence of Boc-proline in the mixtures. This further supported the observations in the above biotransformation samples, suggesting that proline had been removed from the reaction during the reaction period.

Product Yield Calculations and NMR Analysis

Boc protection reactions were performed on final reaction supernatants obtained from the DOGDH reaction in **6.5.1.2** and the StP3H and MlC4H reactions in **6.5.1.1**. Once the Boc protection reaction had been deemed to have reached completion (based on TLC analysis), the final retained organic phases were concentrated in a rotary evaporator. Following complete evaporation of the organic solvent, the mass of the product was recorded. Prior to NMR analysis, the isolated products were further cleaned up by resuspending the concentrated products in MTBE solvent and concentrating by evaporation once more. The

mass of the products was recorded once more and this was used to calculate the final yield of the reactions.

Following clean-up, 1 mg of isolated products were resuspended in 0.5 mL CDCl$_3$, filtered and transferred into NMR tubes. NMR analysis was performed in order to observe what was present in the isolated products. For each sample, proton (^{1}H) and carbon (^{13}C) NMR were performed and recorded on a Bruker Avance 400 spectrometer. Proton spectra were recorded at 400MHz while carbon spectra were recorded at 100MHz.

The ^{1}H-NMR spectra of the isolated supernatants following reaction with DOGDH, StP3H and MlC4H have been presented in Figure 6.13, Figure 6.14 and Figure 6.15.
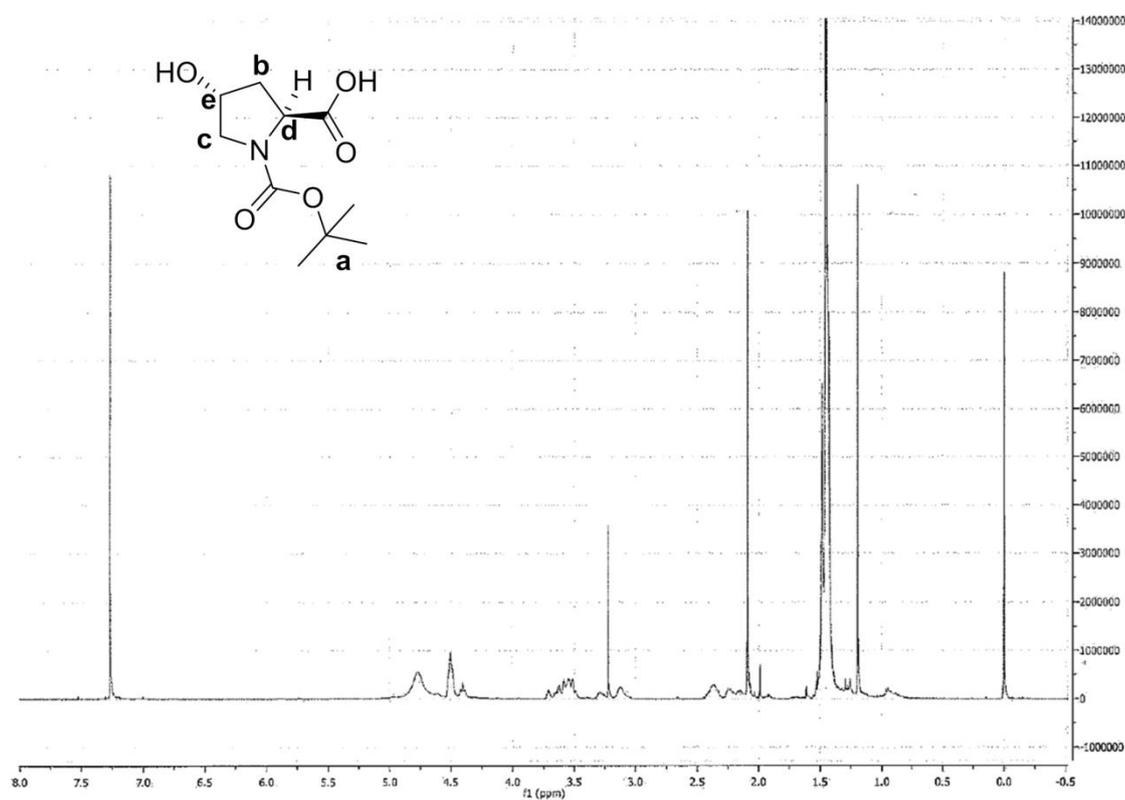
Figure 6.13: ^{1}H-NMR (CDCl$_3$) spectrum of the Boc-protected supernatant obtained from the DOGDH reaction following the final MTBE extraction. The data was recorded at 400 mHz. *Trans*-4-hydroxy-L-proline has presented with positions of interest with H-signals labelled with a letter. Peaks in the spectrum were identified as follows (the letter following the number of protons corresponds to a labelled region in the structure): δ 1.38, 1.44 (9H, A); δ 2.2, 2.4 (2H, b); δ 3.55, 3.7 (2H, c); δ 4.4, 4.55 (1H, d); δ 4.6 (1H, e). δ 1.1, 3.2 (MTBE solvent); δ 2.1 (acetic acid); δ 0.05, 7.2 (CDCl$_3$).
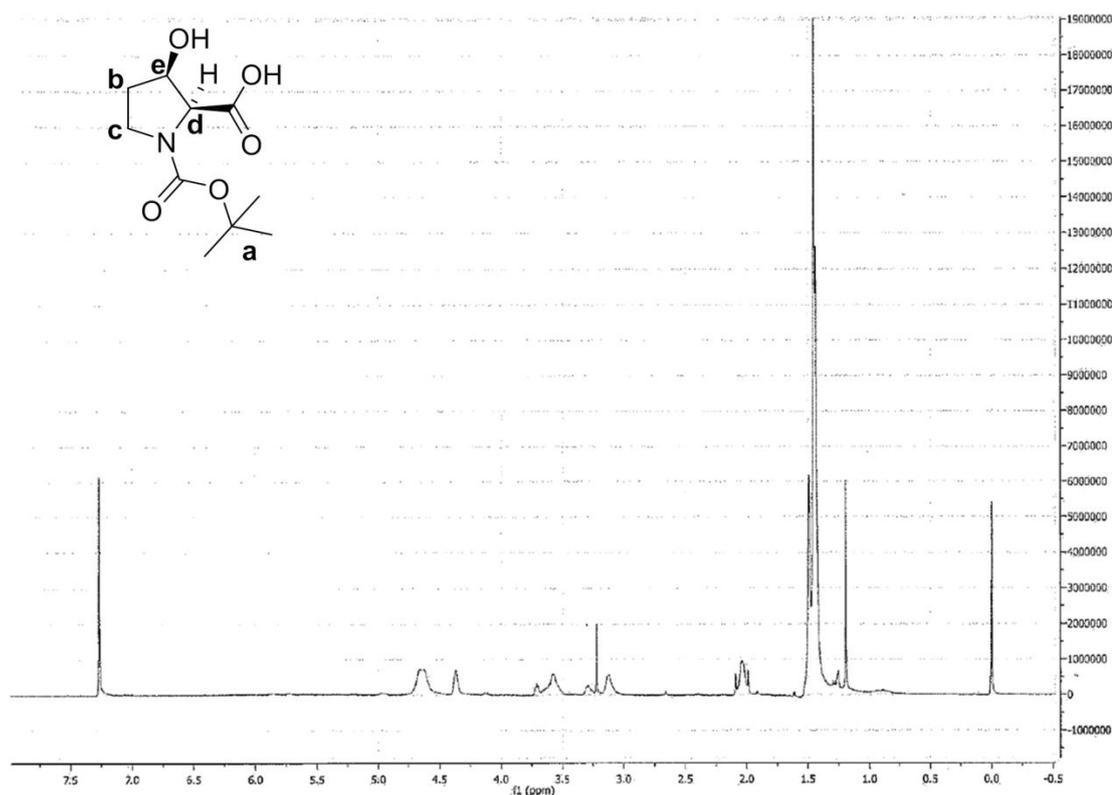
Figure 6.14: ¹H-NMR (CDCl₃) spectrum of the Boc-protected supernatant obtained from the StP3H reaction. *cis*-3-hydroxy-L-proline has presented with positions of interest with H-signals labelled with a letter. δ 1.38, 1.44 (9H, a); δ 2.1 (2H, b); δ 3.3, 3.6 (2H, c); d 4.3 (1H, d); d 4.6 (1H, e). δ 1.1, 3.2 (MTBE solvent); δ 0.05, 7.3 (CDCl₃ solvent).
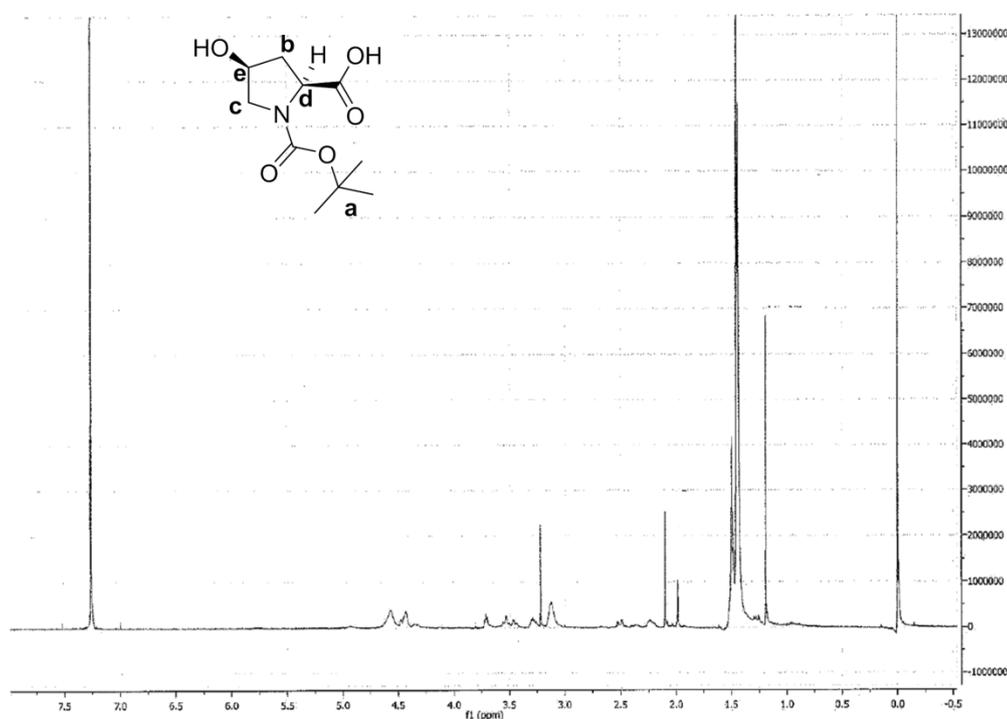
Figure 6.15: H-NMR (CDCl3) spectrum of the Boc-protected supernatant obtained from the MlC4H reaction. *Cis*-4-hydroxy-L-proline has presented with positions of interest with H-signals labelled with a letter. δ 1.38, 1.44 (9H, a); δ 2.3-2.4 (2H, b); δ 3.55, 3.6 (2H, c); δ 4.4 (1H, d); δ 4.6 (1H, e). δ 1.2. 3.25 (MTBE solvent); δ 0.05, 7.3 (CDCl3 solvent)' δ 1.95, 2.1 (acetic acid).

171

The peaks in the ^{1}H NMR spectra were assigned using information in the literature.[244-246] Boc-protected hydroxyproline products were observed in all three samples. The purest samples were isolated in the DOGDH and StP3H reaction supernatants, with only Boc-protected products and solvent peaks observed in these samples. The MlC4H sample appeared to retain contaminants following the extraction and clean-up steps as evidenced by the extra peaks in this spectrum compared to those of DOGDH and StP3H. It must be noted that all three samples predominantly contained solvent as seen by the greater amplitude of the solvent peaks relative to the product peaks suggesting that further purification and extraction steps were necessary.

Having confirmed that products had been protected and isolated in the case of the all three reactions, rough final yields were calculated using the final masses recorded following the MTBE-cleanup step for the DOGDH and StP3H reactons. These are summarised in Table 6.3. Calculations were not performed for the MlC4H reaction because the extracted sample contained a number of contaminants which would have likely skewed the rough yield calculation.

Table 6.3: Amounts of product isolated following Boc protection of reaction supernatants and MTBE clean-up. The MlC4H yield is listed as N/A because the reaction mixture contained contaminants. The final product yield (%) of Boc-hydroxyrpoline was calculated by dividing the amount of sample (in mg) isolated post MTBE cleanup by the theoretical maximum yield of Boc-hydroxyproline.

Reaction	Amount Isolated Post Boc Protection (mg)	Amount Isolated Post MTBE Clean-up (mg)	Final Product Yield (%)
DOGDH-LIC3C	530	300	43.1
StP3H-LIC3C	300	260	37.5
MlC4H-LIC3C	400	300	N/A

Note: Theoretical maximum yield of Boc-hydroxyproline is 693.3 mg based on 100% conversion in 1:1 reaction

The final yields of hydroxyproline in the DOGDH and StP3H whole cell reactions with proline were 43.1% and 37.5% respectively. These values are only estimates because sample losses were inevitable during the numerous extraction steps in the protection process. However, the quantities do align with stunted final peak amplitudes of hydroxyproline in the LC-MS chromatograms of the 48 h time point samples. This also supported the

hypothesis that the reaction was maxing out prematurely due to the unavailability of proline in the reaction mixture, perhaps due to its scavenging by the *E. coli* cellular machinery.

6.5.2 Biotransformations in Stirred Tank Reactors

Whole cell biotransformations of proline with DOGDH and StP3H in shake flasks were generally successful with approximately 43% and 37% of product isolated in each respective case. In order to better simulate industrial biotransformations which are often performed in large-scale fully instrumented bioreactors/fermenters, that utilise an impeller based agitation system rather than the rotation used in shake flasks, whole cell biotransformations of proline were next conducted in MultiMax chemical synthesis reactors (Mettler Toledo).[247] These reactions were performed for DOGDH-LIC3C and StP3H-LIC3C with the optimised reaction mix described in *6.5.1.2* as these enzymes showed the greatest base performance.

The MultiMax reactor system was composed of cylindrical glass reaction vessels placed in a remotely controlled heating jacket. The lid component of the reaction vessel contained circular inlets into which the magnetic impeller stirrer (controlled by a mechanised component linked to the overall control system), a pH probe, a sampler and air inlet could be inserted. Reaction parameters such as jacket temperature, stirring rate and system pH could be controlled and programmed using a computer desktop control panel.

For the purposes of these reactions, the reaction vessel temperature was always maintained at 30 °C and the system pH was left undisturbed. The investigated reaction parameters were stirring rate and aeration and the effect of each will be described and discussed below.

6.5.2.1 Stirring Rate

Reactions in the MultiMax vessels employed an impeller based stirring system. Stirring rates can have a significant effect on mass and oxygen transfer in the reaction solution, reaction rate and cell health.[247] As a result, it is important to carefully control the stirring rate at an optimal level for both the cells expressing the enzymes of interest and the reaction itself.

The effect of stirring rate was investigated by performing whole cell biotransformation of L-proline with DOGDH and StP3H with stirring rates set to 180 RPM and 1000 RPM respectively, with these values selected to serve as two extremes. Reactions were performed

at a 150 mL volume with 10% w/v cells resuspended in the modified reaction mix (20 mM L-proline, 30 mM 2-oxoglutarate, 3 mM ascorbic acid, 0.5 mM $Fe(II)SO_4$, pH 7.5).

Following 24 h of reaction, samples were taken from each reaction and derivatised for LC-MS analysis. Output chromatograms are presented in Figure 6.16.
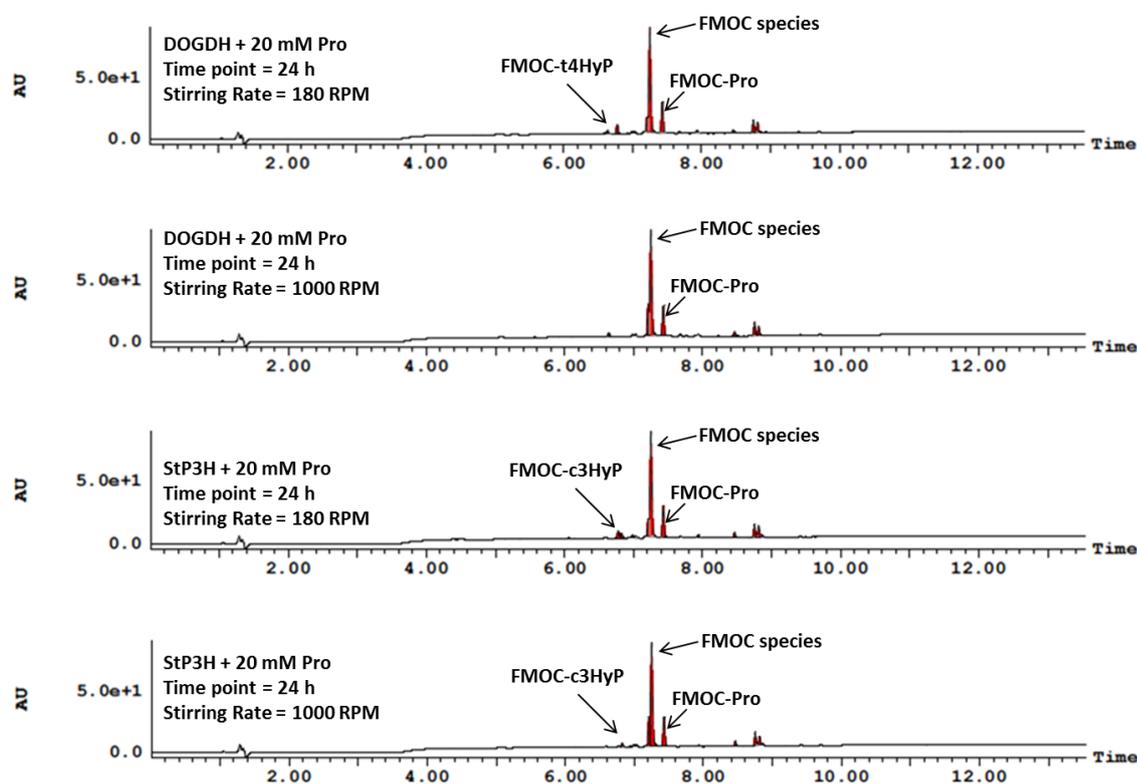
Figure 6.16: LC-MS chromatograms of 24 hour time point samples from the whole cell biotransformations of L-proline with DOGDH and StP3H in stirred vessels with stirring set to 180 RPM and 1000 RPM respectively. The identity of the sample has been stated on the chromatograms and peak identities have been labelled. FMOC-Pro = 7.4 min, FMOC species = 7.2 min, FMOC-t4HyP = 6.6 min, FMOC-c3HyP = 6.8 min.

At a glance, the chromatograms suggested that enzyme performance was generally poor in all of the reactions, regardless of stirring rate, as evident from the poor product peak signals. Despite this, the reactions performed at 180 RPM appeared moderately more successful as suggested from the presence of slight product peaks. It must also be noted that when the stirring rate was set to 1000 RPM, the DOGDH reaction appeared to yield no product at all, while the StP3H showed a slight product peak (as confirmed by MS).

The estimated conversions of the reactions have been summarised in Table 6.4.

Table 6.4: Conversions of L-proline to hydroxyproline at two stirring rates

Enzyme	Conversion at 180 RPM	Conversion at 1000 RPM
DOGDH	8.1 %	0 %
StP3H	11.4 %	Trace

Both DOGDH and StP3H performed relatively poorly in these initial trials utilising stirred tank vessels. Reactions were moderately improved at lower stirring rate as evident from the fact that product was observed in these samples. High stirring rates such as 1000 RPM appeared detrimental to activity as seen by the lack of products and to the cells themselves, as following reaction, the reaction mixture was off-colour with cell debris gathering at the bottom of the vessels. This suggested that the rapid stirring resulted in the shearing of cells.

This pilot test aimed to see whether whole cell biotransformations could be performed using DOGDH and StP3H when conducted in stirred tank vessels. The stirring rate was found to be one of the crucial parameters in this process. Though not optimal, stirring the reaction at 180 RPM appeared to yield product. As this was simply a proof of concept test, this stirring rate was utilised in the investigations that follow.

6.5.2.2 Aeration

The availability of molecular oxygen is crucial for catalysis by 2-oxoglutarate dependent hydroxylases. In the baffled flask reactions described above, oxygen transfer was assisted by a combination of factor including the baffles and surface aeration. This was more difficult to control in the stirred tank reactors due to differences in vessel size, surface area, the absence of baffles and different mode of agitation.

Biotransformations of L-proline with DOGDH were performed in MultiMax bioreactors at 30 °C with stirring set to 180 RPM. Two 150 mL reactions were conducted in parallel with one having air sparged into the system using an inlet valve and the other serving as a non-sparged control. 1 µL of antifoam was added to the sparged reaction mixture to prevent foaming due to the air flow. The reactions were performed for 24 h, after which the reaction was quenched and samples analysed by LC-MS. The output chromatograms for this analysis are presented in Figure 6.17.
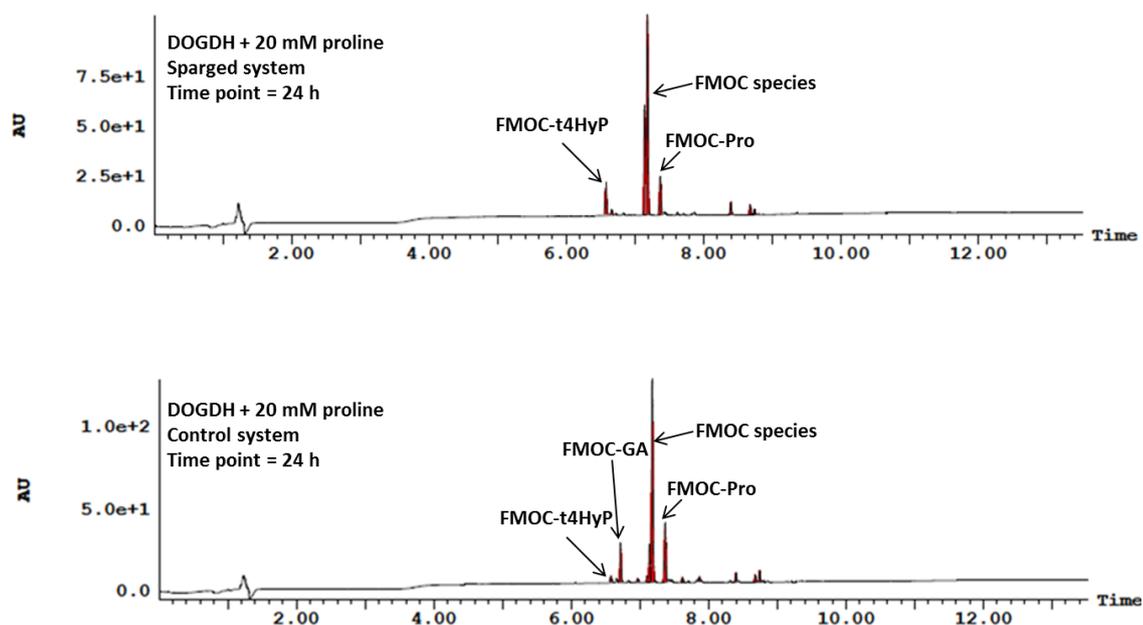
Figure 6.17: LC-MS chromatograms of samples analysed following reaction of DOGDH with 20 mM L-proline in sparged (top) and non-sparged (bottom) stirred tank vessels. Peak identities have been stated and were confirmed with the MS aspect of the analysis. FMOC-proline = 7.4 min, FMOC species = 7.2 min, FMOC-t4HyP = 6.6 min. The FMOC-GA peak represents derivatised glutamic acid contaminants in the reaction media.

The chromatograms suggested that sparging had a beneficial effect on the reaction as evident by the abundant product peak compared to the non-sparged control reaction. The conversions from these reactions are summarised in Table 6.5.

Table 6.5: Presumed conversions of L-proline into t4HyP by DOGDH in either a sparged or non-sparged stirred tank reaction vessel.

Enzyme	System	Conversion after 24 h
DOGDH	Sparged	49%
	Non-sparged	12%

There appeared to be a marked improvement in the activity of DOGDH towards proline when air was directly sparged into the stirred tank vessel, especially when compared to the non-sparged control. This finding was in line with expectations regarding the oxygen requirement of these enzymes and was significant due to displaying their potential capacity to function in stirred tank vessels in a manner akin to shake flask vessels. It was, however, found that it was imperative that the air flow into the system be tightly controlled as in once instance too much flow resulted in the expulsion of a significant amount of sample volume from the reaction vessel.

6.6 Whole Cell Biotransformations with Alternative Substrates

Whole cell biotransformations of L-proline with DOGDH, StP3H and MlC4H were performed to varying levels of success. The work that followed focussed on investigating the substrate scope of these enzymes.

6.6.1 L-pipecolic acid

As mentioned in the introduction of this chapter, L-pipecolic acid is a commonly used alternative substrate in whole cell biotransformations with proline hydroxylases.[194] As a proof of concept, initial work with alternative substrates focussed on assessing whether the expressed and produced DOGDH, StP3H and MlC4H enzyme were active towards L-pipecolic acid.

Protocol

Whole cell biotransformations were performed in 2 L baffled flasks at a 150 mL scale. The reaction solution consisted of 10% w/v cells and 150 mL reaction mix containing 20 mM L-pipecolic acid, 30 mM 2-oxoglutaric acid, 3 mM L-ascorbic acid and 0.5 mM Fe(II)SO$_4$ (reaction mix components were all initially resuspended in 50 mM potassium phosphate pH 7.5 and the pH readjusted to 7.5 prior to reaction addition of cells). Reactions were performed in an orbital shaker for 24 h at 30 °C with shaking set to 180 RPM and samples taken at regular intervals for LC-MS analysis.

Results

StP3H and MlC4H did not appear to show any activity towards L-pipecolic acid as no product peaks were observed in chromatograms, with only a 7.88 min FMOC-L-pipecolic acid peak present in each time point samples. DOGDH, on the other hand, appeared activity towards pipecolic acid with peaks corresponding to FMOC-L-pipecolic acid at 7.88 min and FMOC-hydroxypipecolic acid at 6.80 min observed in the chromatograms. FMOC-pipecolic acid was confirmed to be present at 7.88 min based on its m/z being 374.3 which was to be expected for this species. The 6.80 min was assumed to correspond to FMOC-hydroxypipecolic acid due to possessing an m/z value of 390.4 which corresponded to the mass addition of 16 (likely an OH group).

Two illustrative LC-MS chromatograms in the DOGDH and pipecolic acid biotransformation are presented in Figure 6.18.
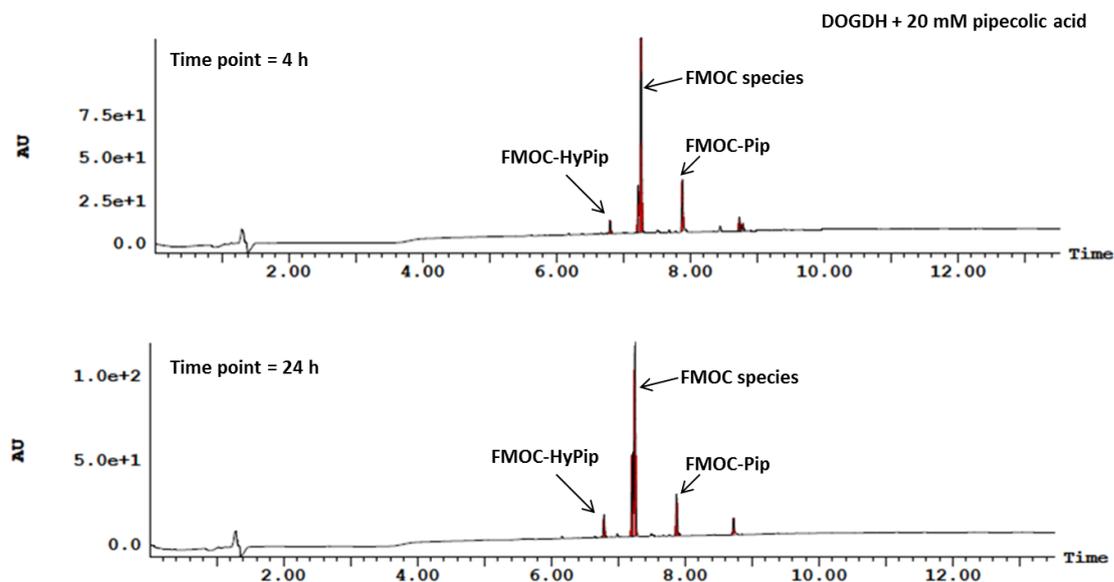
Figure 6.18: LC-MS chromatograms of two time point samples in the DOGDH biotransformation with 20 mM L-pipecolic acid. FMOC-pipecolic acid = 7.88 min, FMOC-species = 7.2 min, FMOC-hydroxypipecolic acid = 6.80 min. Peak identities were confirmed by MS in which m/z values were used as identifiers. In general, it could be seen that the peak amplitude of FMOC-HyPip increased between 4 h and 24 h.

The LC-MS chromatograms of the reaction time course samples suggested that DOGDH was active towards pipecolic acid as evident from the increasing amplitude to hydroxypipecolic acid up to 24 h and the reciprocal decreasing amplitude of the pipecolic acid substrate peak. A rough calculation based on relative peak areas suggested that the conversion into hydroxypipecolic acid after 24 h was approximately 32.3%.

Product recovery was attempted, however, this proved difficult due the presence of both substrate and product in the post reaction supernatant making their respective isolation and quantification difficult. This was due to both being Boc protected and isolated following the extraction, resulting in the requirement for further purification steps which due to time constraints was not possible. As a result, it was not possible to firmly quantify the final reaction yield.

Regardless, the fact that DOGDH was active toward L-pipecolic acid was highly promising opening up a wealth of potential if this activity was studied further.

6.6.2 High Throughput Screening of Substrate Panel

Preliminary biotransformations suggested that DOGDH was active towards L-pipecolic acid. A panel of substrates (Figure 6.19) was assembled for the purpose of performing high throughput biotransformation screenings with DOGDH, StP3H and MlC4H in order to work towards elucidating their substrate scope.
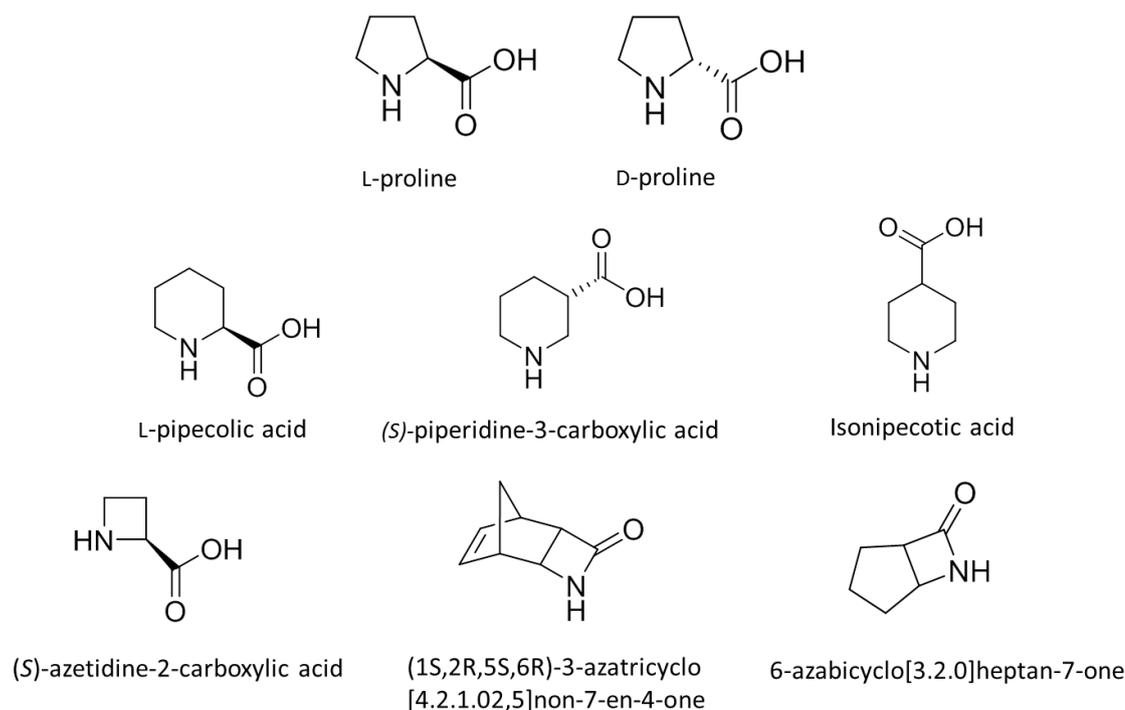
Figure 6.19: Panel of substrates selected for high throughput screening of DOGDH, StP3H and MlC4H activity scope.

Rationale for Panel Selection

L-proline was selected to serve as a control substrate in order to confirm that both the enzymes and screening assay were functional. D-proline was selected in order to assess the enantioselectivity of the hydroxylase targets towards their natural substrate. L-pipecolic acid was selected to serve as a second control substrate due to the observed activity of DOGDH towards it, while (*S*)-piperidine-3-carboxylic acid and isonipecotic acid were chosen due to possessing an identical chemical formula to pipecolic acid but differing in the positioning of the carboxylic acid group relative to the amino group, allowing the evaluation of recognition towards shifted groups. (*S*)-Azetidine-2-carboxylic acid was selected in order to test if the hydroxylases recognised the azetidine ring. (1*S*,2*R*,5*S*,6*R*)-3-azatricyclo[4.2.1.02,5]non-7-en-4-one and 6-azabicyclo[3.2.0]heptan-7-one were chosen to investigate if the hydroxylases could recognise substrates containing β-lactam rings.

Preliminary Analysis

Prior to performing the screening reactions, it was first necessary to ascertain whether the substrates could be analysed by LC-MS. This was not required for L-proline and L-pipecolic acid as they had previously been observed at 7.4 min and 7.8 min respectively following FMOC-Cl derivatiation. (*S*)-piperidine-3-carboxylic acid, isonipecotic acid and (*S*)-azetidine-2-carboxylic acid were derivatised with FMOC-Cl prior to analysis, while the β-lactams were analysed without derivatisation.

The LC-MS chromatograms of these standards are shown in Figure 6.20, Figure 6.21 and Figure 6.22.
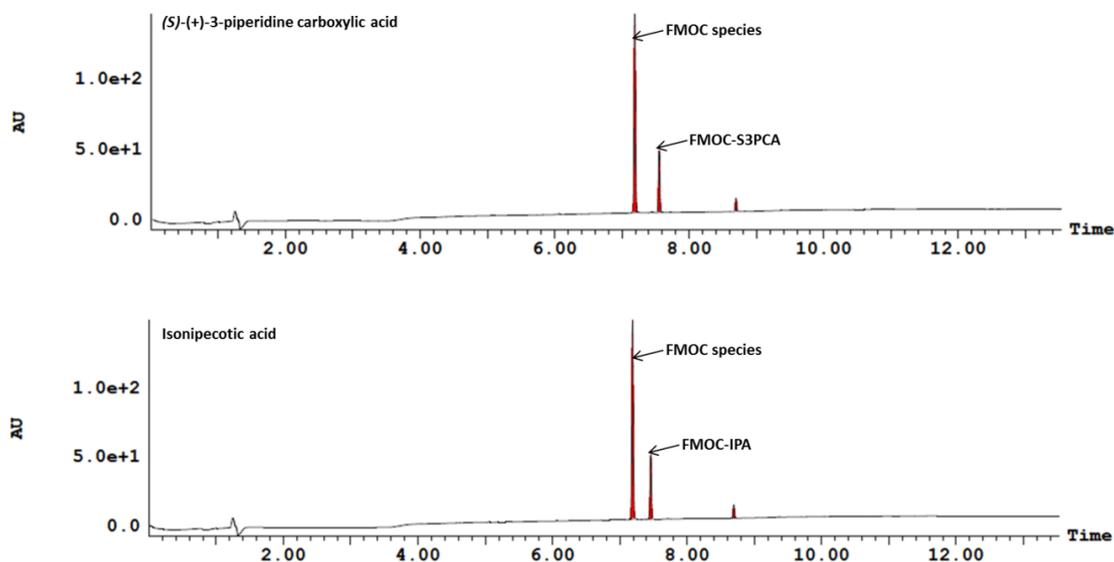
Figure 6.20: LC-MS chromatogram of (*S*)-(+)-3-piperidine carboxylic acid (S3PCA) and isonipecotic acid (IPA) standards derivatised with FMOC-Cl. FMOC-S3PCA was observed at 7.55 min while FMOC-IPA was seen at 7.48 min.
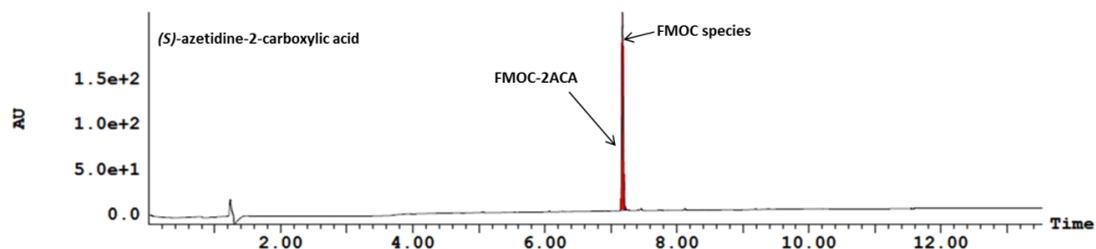
Figure 6.21: LC-MS chromatogram of (*S*)-azetidine-2-carboxylic acid (2ACA) derivatised with FMOC-Cl. Peaks corresponding to FMOC and FMOC-2ACA eluted together and appeared in a fused peak. MS was able to distinguish between them but this was not ideal.
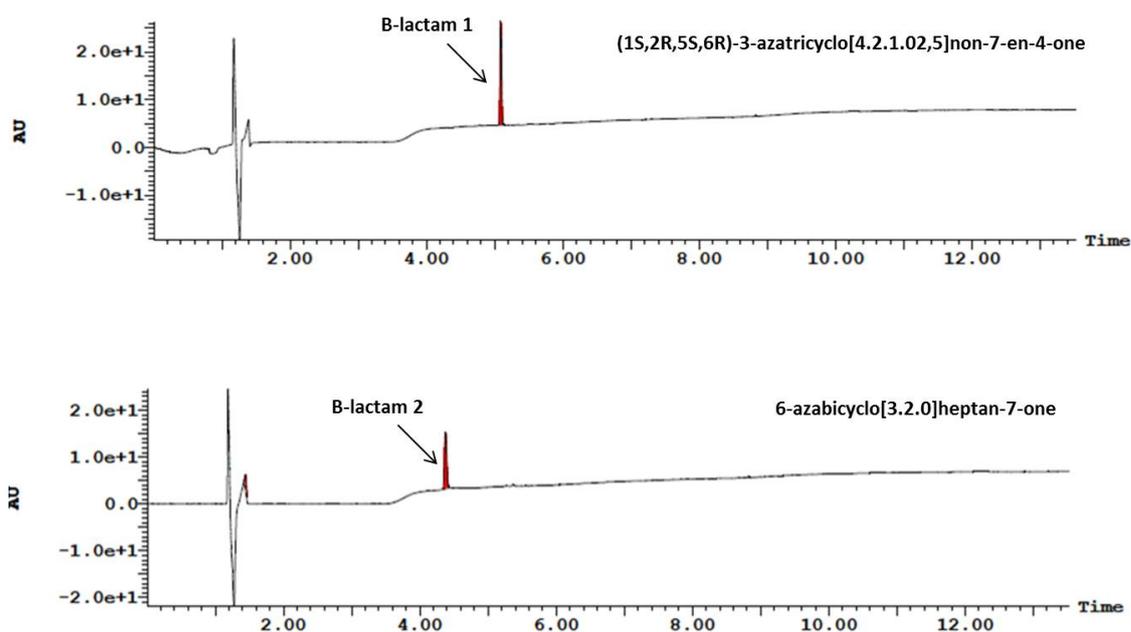
Figure 6.22: LC-MS chromatograms of the β-lactam standards. These did not require derivatisation. (1S,2R,5S,6R)-3-azatricyclo[4.2.1.02,5]non-7-en-4-one was observed at 5.09 min while 6-azabicyclo[3.2.0]heptan-7-one was observed at 4.38 min.

The features of this panel of substrates has been summarised in Table 6.6.

Table 6.6: Summary of screening substrates.

Substrate	Molecular Weight (g mol⁻¹)	MH⁺	Retention Time (min)
L-proline	115.1	360.3 (+FMOC)	7.40
D-proline	115.1	360.3 (+FMOC)	7.40
L-pipecolic acid	129.2	374.3 (+FMOC)	7.80
(S)-(+)-3-piperidine carboxylic acid	129.2	374.3 (+FMOC)	7.55
Isonipecotic acid	129.2	374.3 (+FMOC)	7.48
(S)-azetidine-2-carboxylic acid	101.1	346.3 (+FMOC)	7.18
(1S,2R,5S,6R)-3-azatricyclo[4.2.1.02,5]non-7-en-4-one	135.2	136.2	5.09
6-azabicyclo[3.2.0]heptan-7-one	111.1	112.1	4.38

Screening in the BioLecter ®

The high throughput screening of the substrate panel with DOGDH, StP3H and MlC4H was performed using a BioLecter bioreactor system designed by m2p labs. The BioLecter

181

allowed for the performance of 96 biotransformations in parallel. Reactions were performed in 96 well baffled plates that allowed reaction volumes of up to 3 mL.

Screenings were performed in duplicate for each enzyme with each substrate. Reactions were performed on a 1.5 mL scale at 30 °C with shaking set to 1000 RPM as this was deemed the optimal speed for the greatest aeration at that scale. Each reaction well contained the equivalent of 10% w/v cells resuspended in Reaction Mix containing 20 mM substrate, 30 mM 2-oxoglutaric acid, 3 mM L-ascorbic acid and 0.5 mM Fe(II)SO$_4$ (resuspended in 50 mM phosphate buffer pH 7.5 with pH readjusted to 7.5 prior to addition of cells). Following 24 h of reaction time, the reactions were stopped and samples analysed by LC-MS using the developed derivatisation protocol.

The findings of this screening study have been summarised in Table 6.7 below.

Table 6.7: Summary of screening reactions. Whether the enzyme of interest was active against the substrate or not has been stated with a yes or no.

Substrate	Enzyme		
	DOGDH	**StP3H**	**MlC4H**
L-proline	Yes	Yes	Yes
D-proline	Yes	Yes	Yes
L-pipecolic acid	Yes	No	No
Isonipecotic acid	No	No	No
(S)-(+)-3-Piperidinecarboxylic acid	No	No	No
(S)-azetidine-2-carboxylic acid	No	No	No
(1S,2R,5S,6R)-3-azatricyclo[4.2.1.02,5]non-7-en-4-one	No	No	No
6-azabicyclo[3.2.0]heptan-7-one	No	No	No

The success of the reactions was determined by the presence or absence of a product peak with an MH$^+$ value of 16 greater than the substrate in the LC-MS chromatograms. In general, all three hydroxylases were active towards L-proline as expected. Slight activity towards D-proline was also observed with small FMOC-hydroxyproline peaks observed in these samples. Figure 6.23 presents the LC-MS chromatogram of the 24 h time point

sample of the reaction between DOGDH and D-proline; similar chromatograms were observed for StP3H and MlC4H.
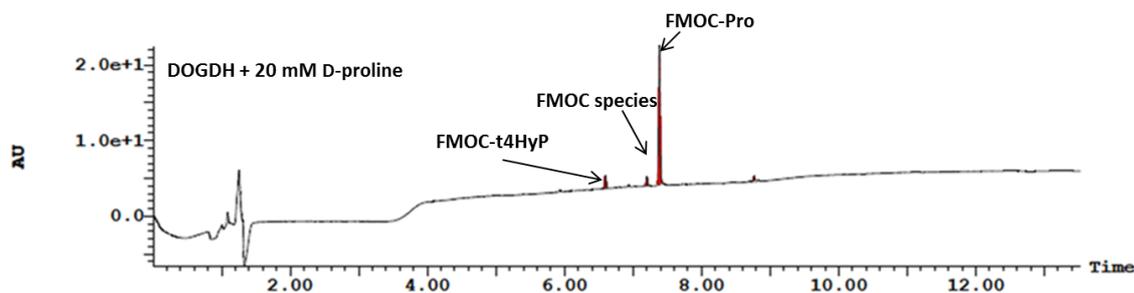
Figure 6.23: LC-MS chromatogram of 24 hour time point sample in the DOGDH reaction with 20 mM D-proline. An abundant FMOC-proline peak was observed 7.4 min, while a markedly reduced FMOC species was present at 7.2 min and a slight FMOC-hydroxyproline peak at 6.6 min. Peak identities were confirmed by MS analysis. An interesting point to note is the reduced height of FMOC peak which is usually the most abundant.

Only DOGDH was active towards L-pipecolic acid, which aligned with previous observations in this chapter. Interestingly, DOGDH was not active towards the structural isomers of L-pipecolic acid, *(S)*-(+)-3-piperidinecarboxylic acid and isonipecotic acid, suggesting that the positioning of the carboxylic acid group relative the amino group plays an important role in substrate recognition. Finally, neither DOGDH, StP3H and MlC4H appeared to be active towards *(S)*-azetidine-2-carboxylic acid and the β-lactam containing compounds.

Though the hydroxylases appeared to not be active towards most of the selected substrates, the fact that the BioLecter was successfully used in this high-throughput screening showed that the system holds the potential for use in future screening reactions if desired.

6.7 Additional Biotransformations

6.7.1 Whole Cell Biotransformations with DOGDH Homologs

As discussed in the previous chapter, two sequence homologs of DOGDH: MP4H and NP4H, were cloned, expressed and purified for the purposes of pursuing a structure of the uncrystallisable enzyme. Cells expressing MP4H and NP4H were produced in Terrific Broth for the purposes of conducting whole cell biotransformations with L-proline in order to determine if the homologs had proline hydroxylase activity.

Whole cell biotransformations of L-proline with the homologs were conducted on a 150 mL scale in 2 L baffled shake flasks as described for DOGDH in *6.5.1.2*. Once 24 hours had elapsed, the reactions were stopped and samples deactivated and derivatised with FMOC-Cl for analysis by LC-MS (Figure 6.24).
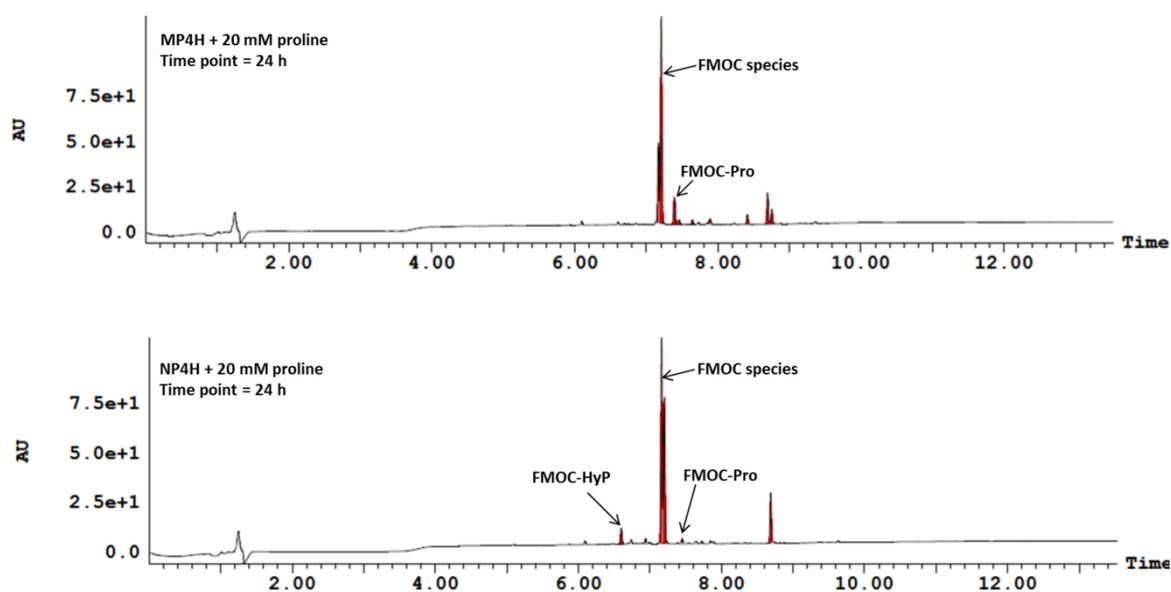
Figure 6.24: LC-MS chromatograms of whole cell biotransformation of 20 mM L-proline with DOGDH homologs. Top: MP4H. Bottom: NP4H. Peak identities have been stated and were confirmed by MS.

The LC-MS analysis suggested that MP4H was not active towards L-proline due to an absence of detectable hydroxyproline. NP4H, on the other hand, appeared to be a functional proline hydroxylase as a hydroxyproline peak was observed at 6.60 min. The NP4H product was potentially *trans*-4-hydroxyproline due to it possessing an identical retention time to the t4HyP standard.

6.7.2 Whole Cell Biotransformations with DOGDH-SER

A surface entropy reduced mutant of DOGDH was synthesised for the purposes of structural studies. This mutant differed from the wild-type DOGDH in that it possessed E48A and K49A dual mutations. As a point of interest, biotransformations of L-proline were attempted using this construct in order to assess if these mutants had any effect on the native activity of the enzyme.

Whole cells expressing DOGDH-SER were produced in Terrific Broth and reactions performed on a 150 mL scale in 2 L baffled shake flasks as described for the wild-type enzyme in *6.5.1.2*. Samples were taken from the reaction mixture at regularly intervals and snap frozen in liquid nitrogen until required. Once 24 h had elapsed, the reaction was stopped and samples process and analysed by LC-MS. Two representative LC-MS chromatograms are presented in Figure 6.25.
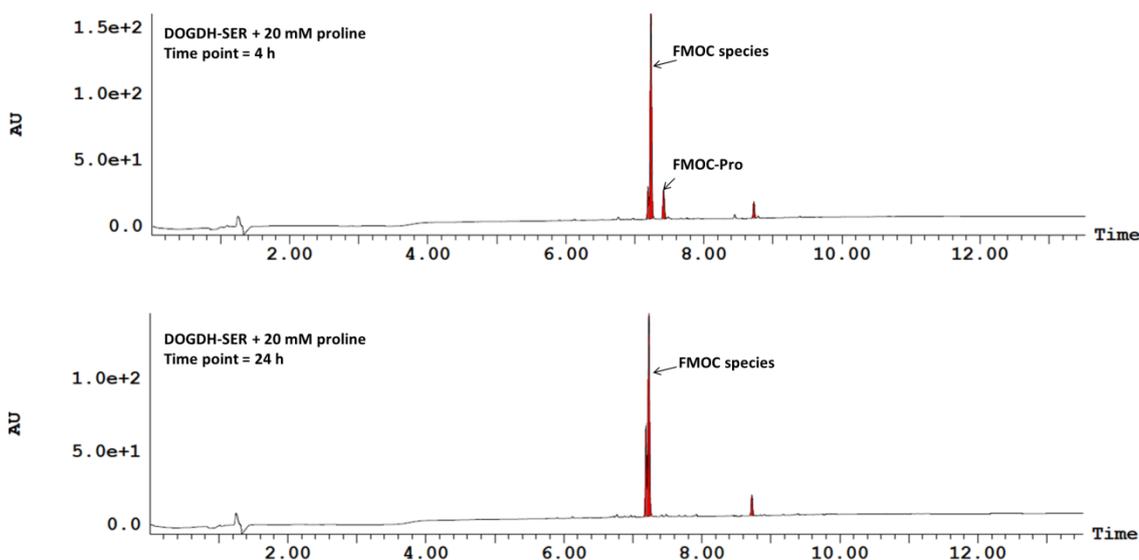
Figure 6.25: LC-MS chromatograms of 4 h and 24 h time point samples from the whole cell reaction between DOGDH-SER and 20 mM L-proline. Peak identities have been stated and were confirmed by MS.

LC-MS analysis of this reaction suggested that DOGDH-SER was not active towards L-proline as evident from the absence of hydroxyproline in both shown time points. Interestingly, the FMOC-proline is also absent in the 24 h time point sample suggesting that even without activity the proline had been removed from the reaction mixture. This is discussed further in the next section.

185

6.8 Discussion and Conclusion

Biotransformations of L-proline with purified DOGDH-LIC3C, StP3H-LIC3C and MlC4H-LIC3C were performed in order to assess the stability and activity of these enzymes in a cell-free environment. Of the three enzymes, only StP3H appeared active in the purified state with a conversion of approximately 25%. This may have been the result of StP3H-LIC3C appearing the most stable out of the three hydroxylases following purification, remaining so during the reaction course. No activity was observed for DOGDH-LIC3C and MlC4H-LIC3C with the two being generally unstable following purification, producing fragmenting bands when analysed by SDS-PAGE, with the protein rapidly precipitating out of solution during the reaction course. Another factor to consider was the enzymes' requirement for molecular oxygen, 2-oxoglutarate and Fe(II)SO$_4$, which though provided in the reaction mixture did not have a means for recycling them. As a result, either additional oxidoreductase enzymes would be required or a reaction system in which cofactors and additional reaction components could be continuously accessible. This resulted in the use of whole cell environments, as these would theoretically provide these components.[50]

Whole cell biotransformations were initially performed in 2 L baffled shake flasks and improved conversions were observed for all three hydroxylases compared to their purified equivalents. LC-MS analysis of the endpoint samples in the all three reactions showed an absence of proline in the mixtures, which almost led to the false conclusion that these reactions reached 100% conversion. The absence of proline was confirmed by TLC analysis and NMR of the extracted samples. The final yields for the DOGDH and StP3H reactions were calculated to be 41% and 37% respectively. These moderate conversions were potentially the result of proline unavailability over the reaction course. Schmid and colleagues recently discussed that *Escherichia coli* cells use proline as a carbon and nitrogen source via a proline utilisation (*put*) operon composed of two genes, *putP* encoding a proline transporter and *putA* encoding an enzyme with proline dehydrogenase activity, with this operon being transcribed and activated in the presence of proline.[242] The group found that knocking out the transporter resulted in proline availability when working with proline hydroxylases.[242] It was thus highly likely that the *put* operon was playing a key role in the

whole cell biotransformations in shake flasks described in this chapter and it would be worth attempting work with a modified *E. coli* with the transporter knocked out to see the effect on the conversions observed.

Having successfully performed whole cell biotransformations in shake flasks, pilot tests were conducted in MultiMax bioreactor vessels. These performed in order to assess whether transformations of L-proline could be conducted in a system akin to industrial bioreactors that utilise an impeller based mixing system rather than shaking. Initial transformations yielded poor results and it was assumed that this may have been the result of inefficient aeration. To test this, reactions were performed with DOGDH in which air was sparged directly into the system. It was found that sparing air into the system improved performance with greater conversions than the non-sparged equivalent observed. This work was promising because it showed that these hydroxylases could be taken from shake flasks to stirred tank vessels and function almost as well. It must however be noted that this was simply a pilot test. In order to perform the reactions in the most efficient conditions further effort would need to be conducted to determine optimal stirring rates, sample volumes and aeration levels. This is due to it not being possible to simply extrapolate findings and values from shake flasks to stirred tank due to the shifts in surface area, stirring method and mass transfer.[243]

Biotransformations were successfully performed using L-proline as substrate. The work that followed focussed on preliminary investigations of the substrate scope of DOGDH, StP3H and MlC4H. A panel of substrates was selected and screenings were performed in a BioLecter system. The findings of these preliminary studies showed that the hydroxylases were active towards both L- and D-proline while DOGDH was also active towards L-pipecolic acid. DOGDH was not active towards the structural isomers of L-pipecolic acid, (*S*)-(+)-3-piperidinecarboxylic acid and isonipecotic acid, suggesting that the positioning of the carboxylic acid group relative the amino group plays an important role in substrate recognition. Finally, none of the hydroxylases appeared to recognised substrates containing an azetidine ring or β-lactam ring. Though the majority of the reactions in the panel screen appeared to yield negative results, the fact that the BioLecter was successfully utilised to perform 96 biotransformations in parallel was very promising, as this provides a method for future screenings.

Finally, the DOGDH homologs and surface entropy reduced mutant were reacted with proline in shake flasks to ascertain whether the former possessed proline hydroxylase function and the latter was still functional following mutagenesis. Of the homologs, NP4H appeared to be a proline hydroxylase as suggested by the presence of a product peak in the LC-MS analysis. DOGDH-SER appeared to have lost function towards proline as there was no evidence of hydroxyproline in the analysed samples.

In conclusion, a series of biotransformations were performed in order to assess the factors key to performance of the hydroxylase targets. It was found that these enzymes generally appeared to function the best in a whole cell environment in which cofactors can be scavenged and regenerated, however, it is important to consider the chance of substrate itself being scavenged by cellular machinery, limiting its availability to the enzymes and hence affecting the final yield. The ratio of reaction components was also found to be important for reaction and an optimal ratio suggested. The hydroxylases were also found to be highly dependent on the availability of molecular oxygen and aeration as seen in the sparging experiment in stirred tank vessels. Finally, a preliminary screening experiment was also successfully performed in a BioLecter system and presented a means for future screenings. From this, it can be concluded that proline hydroxylases are highly sensitive and complex enzymes that hold great potential if utilised successfully.

Chapter 7

Modelling the Structure of DOGDH

7.1 Introduction

An enzyme structure can be used to obtain a great deal of information relating to a protein including (and not limited to) its mechanism of action and can also be used guide mutagenesis studies for the purposes of evolution.[248] In order to do this, the acquisition of a structure is dependent on obtaining diffraction quality crystals of the protein for X-ray crystallographic studies.[249] Obtaining crystals can be very difficult and is dependent on a number of factors, including protein purification and stability prior to, and during the crystallisation screening process.[248] Difficulties in obtaining crystals are among the most reasons for the number of known protein sequences far exceeding the number of deposited protein structures (a sequence-structure gap).[249]

In the absence of experimental 3D structures, computational methods are often used in order to predict the three-dimensional structure of protein in order to examine structure-function relationships. There are various software and servers for computational structure prediction with the output structures serving as starting points. However, human intervention is often required in further building and processing these models in order to remove and control errors that may have resulted during the computational modelling process.[250-251]

One such methodology is homology modelling, which is used to predict protein structures based on the assumption that proteins with similar sequences possess related structures. Homology modelling works on the principle of using an experimentally established protein structure with greater than 30% sequence similarity or structural homology to the target protein as a template for model prediction.[248] The general procedure can be broken down into 4 steps: (1) the identification of a known 3D structure with sequence similarity to the target to serve as a template; (2) the sequence alignment of the target protein and template protein; (3) the construction of a model structure for the target using the template structure

and sequence alignment; (4) the refinement, validation and evaluation of the generated models. The steps in this process are often repeated until a satisfactory model is generated.[252]

It must be stated that, as their name implies, models are extrapolations and may contain errors as a result.[248] It is generally accepted that models built using templates with over 50% identity to the target possess adequate accuracy for use in drug discovery applications, while structures with 25-50% identities can be used to inform mutagenesis studies. Structures resulting from 10-25% identity to templates are assumed to be speculative in the best case.[252]

Homology modelling has been used successfully in a number of cases ranging from use in drug discovery applications to mutagenesis. A relevant example for this project was previously discussed in the preceding chapters and relates to the work performed by Koketsu and colleagues. Koketsu *et al.* solved the structure of MlC4H and used this as a template for building a model structure for the sequence related proline hydroxylase SmC4H, for which crystals could not be obtained.[205] This presents a valid argument for the use of homology modelling to build a structure for DOGDH, which as described in Chapter 5, proved difficult to adequately crystallise despite extensive efforts.

7.2 Aims

The focus of Chapter 5 was the pursuit of a DOGDH structure by crystallography, which remained one of the only known microbial proline hydroxylases to not have its structure known. It was not possible to obtain diffraction quality crystals of DOGDH despite extensive efforts, resulting in the requirement of pursuing computational modelling strategies.

The aims of this chapter were to model a 3D structure of DOGDH using the Robetta server and to dock substrates and/or products into the enzyme active site in order to identify amino acid residues with potential roles in catalysis. The structure will be used to identify targets for mutagenesis and the generation of mutants will be attempted.

7.3 Homology Modelling of DOGDH

The Robetta server was selected as the automated tool for predicting a model structure of DOGDH. Robetta (http://robetta.bakerlab.org) was developed by the Baker group and is an internet service that provides automated structure prediction and analysis tools that can be used to deduce protein structures from their nucleotide sequences.[253] The server utilises a fully automated structure prediction protocol (the first of its kind) using a protein sequence as input to build a complete model in the presence or absence of sequence homology to known protein structures. The general building process involves dividing the input sequence into domains, building models for domains with sequence homology to known structures using homology modelling. Robetta also possesses the capacity to build models for the domains that lack homology using the built-in Rosetta *de novo* structure prediction method.[253] The output results are given as domain predictions and molecular coordinates of models that span the full length of the input sequence.[253]

Building a DOGDH Model

The amino acid sequence of DOGDH (Figure 4.1) was input into the Robetta server (http://robetta.bakerlab.org) and the building job left to proceed. The output DOGDH model was primarily built using ectoine hydroxylase (EctD, PDB: 3EMR) as a template.[254] An alignment of the protein sequences of DOGDH and EctD using EMBOSS Needle is shown in Figure 7.1. The alignment suggested that the two protein sequences shared 27.7% sequence identity and 40.9% similarity. As discussed above, for the purposes of homology modelling, a 25-50% sequence identity is often deemed sufficient for informing mutagenesis studies.

EctD is a member of the 2-oxoglutarate dependent oxygenases that catalyses the hydroxylation of ectoine.[254] The structure of EctD is presented in

Figure 7.2.

```
DOGDH          1 ------------------------------------MLTPTELKQYR     11
                                                    .::..:...:.
EctD           1 HHHHHHSEMQDLYPSRQRADAEMRPRLDPVVHSEWTNDAPISARQAAAFD    50

DOGDH         12 EAGYLLIEDGLGPREVDCLRRAAAALYAQ----DSPDRTLEKDGRTVRAV    57
                 ..||::||.....||..|::||..|.|.    |:.....|.....:|::
EctD          51 RDGYIVLEDIFSADEVAFLQKAAGNLLADPAALDADTIVTEPQSNEIRSI   100

DOGDH         58 HGCHRRDPVCRDLVRHPRLLGPAMQILSGDVYVHQFKINAKAPMTGDVWP   107
                 ...|.:.||...|....||...|..:|..:||:||.::|.|.....|..:.
EctD         101 FEIHAQSPVMARLAADARLADVARFLLGDEVYIHQSRLNYKPGFKGREFY   150

DOGDH        108 WHQDYIFWAREDGMDRPHVVNVAVLLDEATHLNGPLLFVPGTHELGLIDV   157
                 ||.|:..|.|||||.|...:::||.|.|..||||:.:||:|...|..|
EctD         151 WHSDFETWHVEDGMPRMRALSMSVLLAENTPHNGPLMVIPGSHRTYLTCV   200

DOGDH        158 ERRAPAGDGDAQWLPQLSADLDYAI-DADLLARLTAGRGIESATGPAGSI   206
                 ....     |..:|..|... :|.:  |.:.||.|....:.||...|.|.
EctD         201 GETP-----DDHYLSSLKKQ-EYGVPDEESLAELAHRHGIVAPTGKPGTV   244

DOGDH        207 LLFDSRIVHGSGTNMSPHPRGVVLVTYNRTDNALP---AQAAPRPEFLAA   253
                 :|||..::|||..|::|.||....:.||...|.|.    ....|||.|||.
EctD         245 ILFDCNLMHGSNGNITPFPRANAFLVYNAVSNRLEKPFGVEKPRPWFLAR   294

DOGDH        254 R--------DATPLV-PLPAGFALAQPV        272
                 |           :..|||  .:||
EctD         295 RGEPAALRVERGPLVETVPA--------        314
```

Figure 7.1: Protein sequence alignment of DOGDH and EctD obtained using EMBOSS Needle. The alignment suggested that the two proteins shared 27.7% identity and 40.9% similarity.
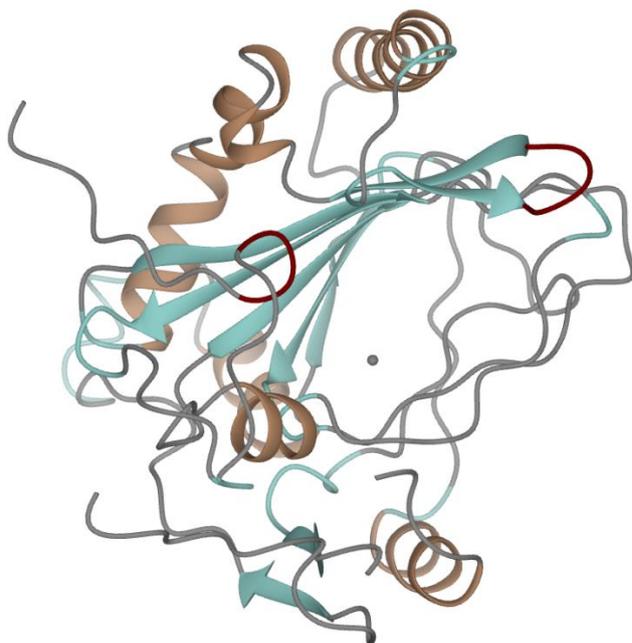
Figure 7.2: Structure of EctD monomer (PDB: 3EMR). Fe(II) is present in a central cavity and is represented as a grey sphere. The central iron is surrounded by two β-sheets (light blue) while the outer region of the monomer contains α-helices (brown).

EctD possesses a structure common to other members of the 2-oxoglutarate dependent oxygenases, containing a central core composed of β-sheets core that is flanked by a number of α-helices. Fe(II) was located at the centre of a cavity formed by the surrounding β-sheets.[254]

The DOGDH Model

The three-dimensional model of DOGDH that was the output of the Robetta server, and primarily based on EctD is presented in Figure 7.3. The DOGDH model in Figure 7.3 was constructed as a dimer and the component monomers have been labelled in blue and gold.
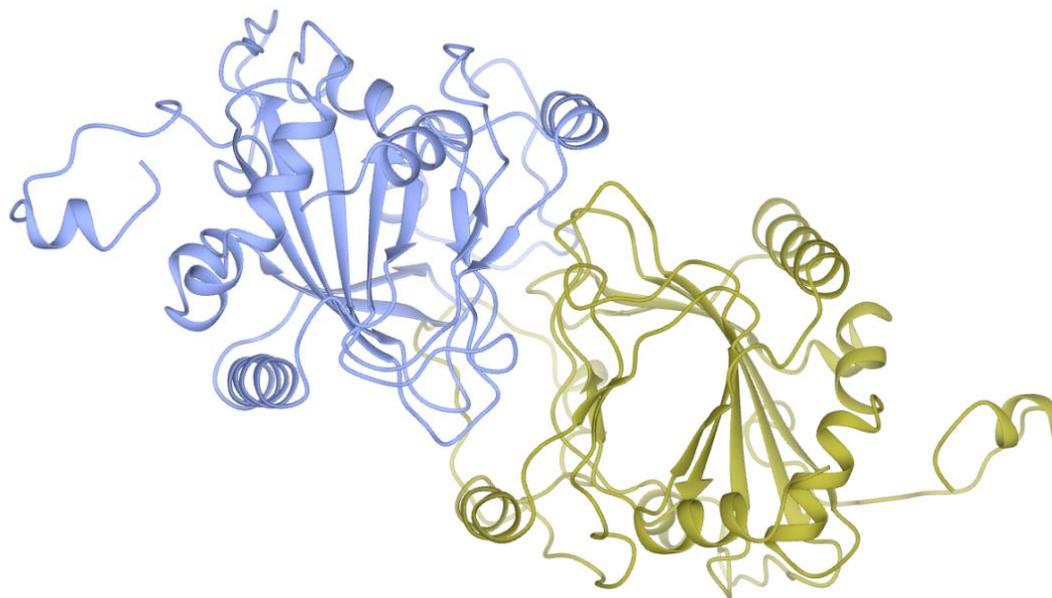
Figure 7.3: DOGDH Model. The model was constructed to be a dimer with the component monomers presented in blue and gold. The overall shape of each monomer appears to closely resemble the EctD monomer presented in Figure 7.2, possessing a central β-sheet core surrounded by α-helices.

An overlay of the DOGDH model with the apo EctD structure suggested that two shared secondary and tertiary structural homology, which together with the close resemblance of the monomer forms of the two further suggested that the latter had been used as the template for this model.

The DOGDH monomer model with docked Fe(II) in the active site channel is presented in Figure 7.4.
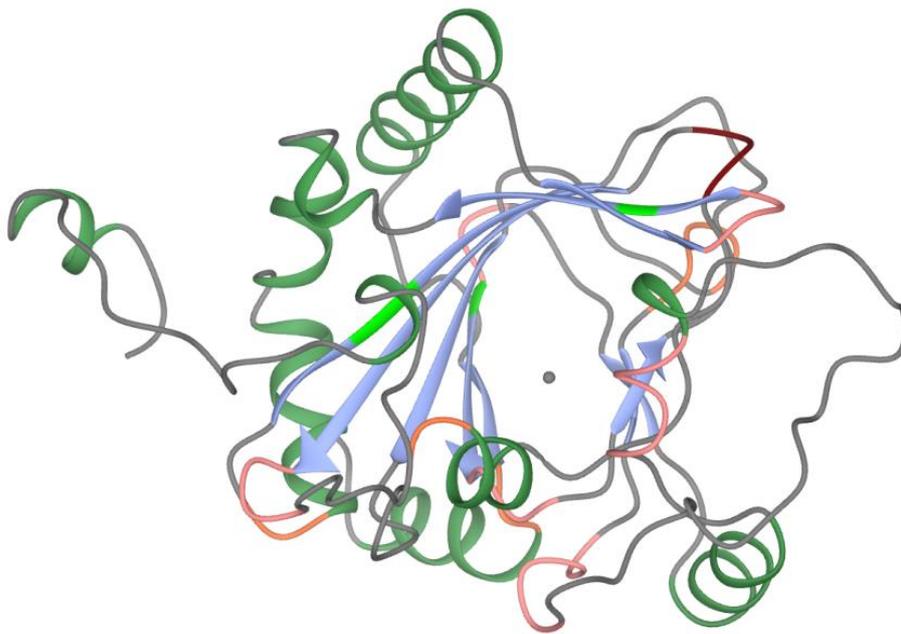
Figure 7.4: Monomer of DOGDH model with the active site cavity highlighted in the centre of the snap shot. Fe(II) is represented as a grey sphere in the centre of the active site. The central iron is located in a cavity formed by surrounding β-sheets (blue) which are in turn flanked by a series of α-helices (green).

The monomer of DOGDH shows that Fe(II) is present in a central cavity formed by 8 β-sheets (blue) that form a distorted jelly roll structural fold common to 2-oxoglutarate dependent oxygenases. The β-sheets are surrounded by α-helical domains (green).

In order to investigate the mechanism of catalysis within the DOGDH active site, the DOGDH model was superimposed with an updated EctD structure (PDB: 4Q5O) with bound product, 5-hydroxyectoine, in the active site.[255] Following superimposition *trans*-4-hydroxyproline (T4HYP) was modelled into the DOGDH active site where 5-hydroxyectoine was found in EctD. T4HYP was rotated within the active site so that NH group in the pyrrole ring and the carboxyl group were aligned with 5-hydroxyectoine.

A representation of the active site of DOGDH containing T4HYP, 2-oxoglutarate and Fe(II) is shown in Figure 7.5.
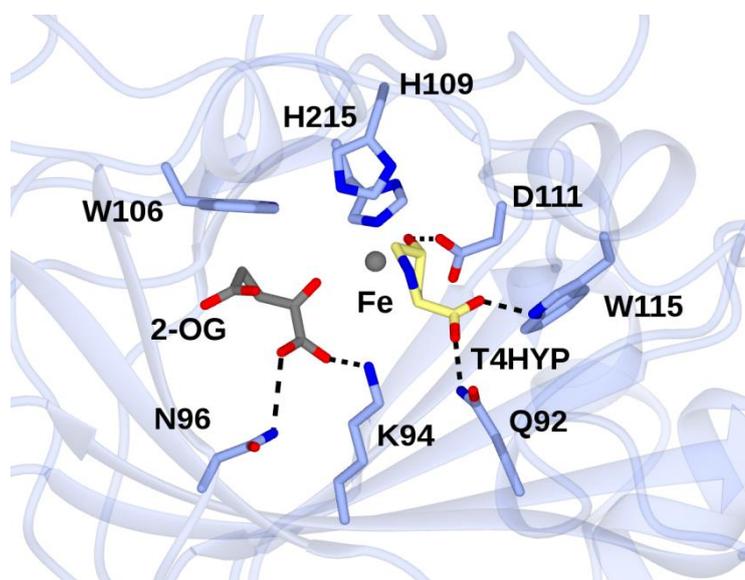
Figure 7.5: Active site of DOGDH model. A triad of H108, D111 and H215 coordinate to the central Fe ion (grey sphere). T4HYP was docked into the site and is present to the right of Fe, while 2-OG is to the left. Amino acids of interest have been highlighted. N96 and K94 potentially form interactions with 2-OG. Q92 and W1115 serve the possible function of coordinating to the carboxylate group of T4HYP. D111 possibly plays a dual role of coordinating both Fe and the C4 OH of T4HYP.

Residues of interest in the DOGDH model active site have been presented in Figure 7.5 above. A trio of H108, D111 and H215 were observed in the vicinity of the central iron metal. It was proposed that these likely serve as the facial H-X/D-H triad in DOGDH, coordinating Fe to the enzyme active site for the purposes of catalysis and enzyme stability. D111 was proposed to play the dual role of also coordinating to the C4 hydroxyl group of T4HYP by forming interactions to it from one of its carboxyl oxygens. It was hypothesised that this may play a role in orientating this C4 carbon towards the active Fe so that hydroxylation occurs here, which may partly account for the regioselectivity of the DOGDH hydroxylation reaction. W115 and Q92 were predicted to form interactions with the carboxyl group of T4HYP which may further assist in orientating it towards the Fe. Finally, K94 and D96 were hypothesised to form interactions with 2-oxoglutarate potentially playing a role in positioning or stabilising it.

The DOGDH model appeared to successfully identify residues of interest that may play a role in catalysis. These could serve as a starting point for mutagenesis studies for further investigation. It must, however, again be noted that this was simply a model hence the information drawn from it may not actually relate to the actual DOGDH structure.

Selection of Preliminary Mutagenesis Targets

The ultimate goal of mutagenesis studies of DOGDH would be to generate a variant with enhanced function, altered selectivity, broader substrate range and capacity to catalyse alternative reactions. Potentially interesting residues in and around the active site of DOGDH were identified and discussed in the previous section. The number of initial mutation targets needed to be narrowed down for the purpose of performing preliminary mutagenesis studies to test the generated DOGDH model and due to time limitations. Figure 7.6 presents a simplified representation of the active site of the DOGDH model with initial amino acid targets for mutagenesis.
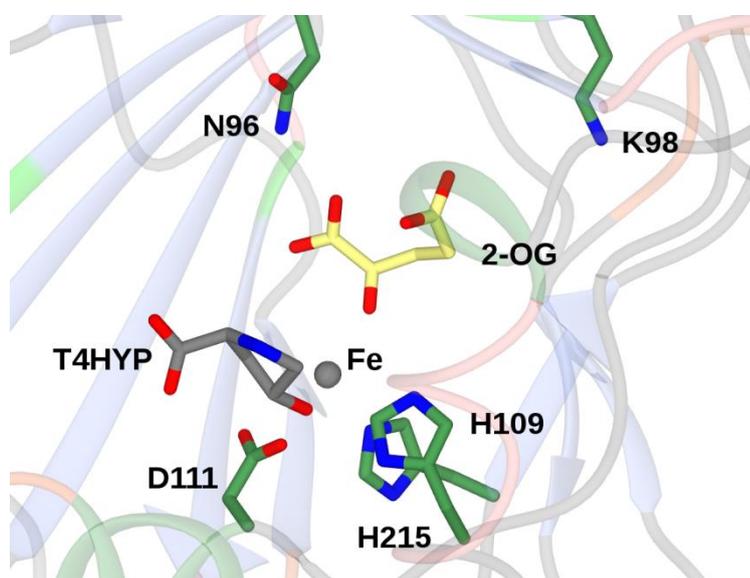
Figure 7.6: Active site of DOGDH model with amino acids of interest in preliminary mutagenesis studies. H109, D111 and H215 likely constitute the facial H-X/D-H triad that coordinate to Fe(II) and keep it bound to the enzyme active site. N96 and K98 likely correspond to key amino acid residues in the MlC4H structure (PDB: 4P7W) at positions 95 and 97.

Mutating N96 and K98 in DOGDH to Investigate Regioselectivity

Interesting mutagenesis work was performed by Koketsu and coworkers using the MlC4H structure (PDB: 4P7W).[205] An overlay of the DOGDH model and MlC4H structures suggested that the two may share similarities in overall fold (Figure 7.7), while sharing approximately 28% sequence identity.

Figure 7.7: Overlay of DOGDH model (blue) with MlC4H structure (gold; PDB: 4P7W). The two structures appear to have a shared jelly-roll core region but differ in their outer region α-helices.

Koketsu *et al.* used the MlC4H structure to model the structure of a closely related hydroxylase, SmC4H, finding that mutations of V95 and V97 in the latter had pronounced effects on the enzyme's activity, particularly on the regioselectivity of the hydroxylation reaction with L-pipecolic acid.[205] SmC4H shared 66% sequence identity with MlC4H with V95 and V97 in SmC4H corresponding to I95 and I97 in MlC4H.[205]

An overlay of the MlC4H and DOGDH structures suggested that the I95 and I97 in MlC4H corresponded to N96 and K98 in DOGDH, with the loops in which these residues were located having a strong secondary structure overlap (Figure 7.8).
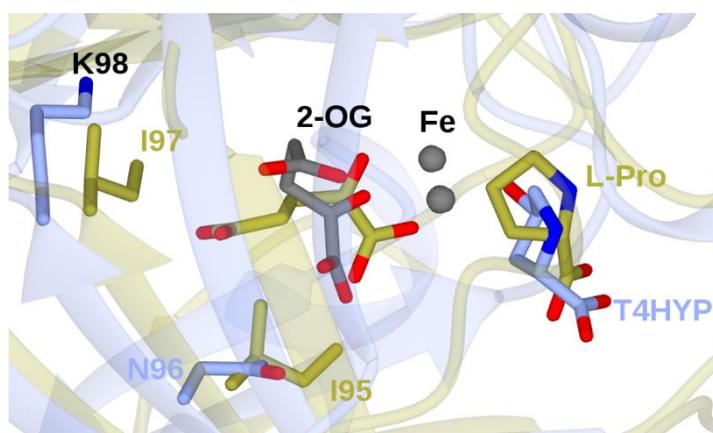
Figure 7.8: Overlay of active site loop regions of the DOGDH model (blue) and MlC4H structure (gold). Compounds and residues of interest have been labelled in gold for MlC4H and blue for DOGDH. The β-sheets and loops generally align but it must be noted that the conformations of 2-OG ad L-Pro/T4HYP differed between the enzymes. The overlay suggested that K98 and N96 in DOGDH were in the corresponding positions to I97 and I95 respectively in MlC4H.

It was decided that mutations of N96 and K98 would be performed in DOGDH for the purpose of assessing whether such mutations altered the activity of DOGDH towards L-proline and L-pipecolic acid. Mutations to alanine were selected for both in order to investigate the contribution (if any) these residues may have towards DOGDH stability and/or function by essentially knocking out the residues. Mutations of both residues to phenylalanine were also selected to assess the effect of the addition of a bulky phenyl group, while also mimicking the studies performed Koketsu and coworkers.

Mutating D111 in DOGDH to Investigate the Effects on Activity

The DOGDH active site model suggested that H109, D111 and H215 constituted the facial H-X/D-H triad that characteristically coordinate to Fe(II) anchoring it to the active site of 2-oxoglutarate dependent oxygenases.[176] A Codexis patent suggested that mutation of the Fe sphere resulting in tighter metal binding could result in enhanced activity and stability.[256] Previous mutation studies of 2-oxoglutarate dependent hydroxylases found that, for enzyme integrity and function, the first and third residue in the triad must not be changed while the middle residue presents greater flexibility for experimentation.[175] In the case of DOGDH, D111 was a promising candidate for manipulating Fe(II) binding as well as potentially altering the regioselectivity of the hydroxylation reaction (discussed above). Two mutations were initially targeted for this work; D111H to investigate the effect of having a triad of histidine residues coordinating to Fe(II) and D111N to investigate the effect of shifting from a charged side chain to a polar side chain.

7.4 Mutagenesis of DOGDH

The target mutants have been listed in Table 7.1 together with the utilised primer pairs.

Table 7.1: Primers utilised in mutagenesis

Mutant	Primer Name	Sequence (5' -> 3')
N96A	N96A_FOR_1	GTTCAAGATCGCGGCGAAGGCC
	N96A_REV_1	CCCGGAAGCGGCGCTAGAACTT
	N96A_FOR_2	GTCCACCAGTTCAAGATCGCGGCGAAGGCCCCGATGACC
	N96A_REV_2	GGTCATCGGGGCCTTCGCCGCGATCTTGAACTGGTGGAC
N96F	N96F_FOR_1	AGTTCAAGATCTTTGCGAAGGCCC
	N96F_REV_1	CCCCGGAAGCGTTTCTAGAACTTG
	N96F_FOR_2	GTCCACCAGTTCAAGATCTTTGCGAAGGCCCCGATGACC
	N96F_REV_2	GGTCATCGGGGCCTTCGCAAAGATCTTGAACTGGTGGAC
K98A	K98A_FOR_1	AGTAGCCCCGGCGGCGCAACTAGAA
	K98A_REV_1	AGTAGCCCCGGCGGCGCAACTAGAA
	K98A_FOR_2	GTTCAAGATCAACGCGGCGGCCCCGATCAGGCCGC
	K98A_REV_2	CGCCGGTCATCGGGGCCGCCGCGTTGATCTTGAAC
K98F	K98F_FOR_1	AAGATCAACGCGTTTGCCCCGATGAC
	K98F_REV_1	CCAGTAGCCCCGTTTGCGCAACTAGA
	K98F_FOR_2	CAGTTCAAGATCAACGCGTTTCGGGGCATCACCGGCGATG
	K98F_REV_2	CATCGCCGGTCATCGGGGCAAACGCGTTGATCTTGAACTG
D111H	D111H_FOR_1	CCGTGGCACCAGCATTACATCTTCTG
	D111H_REV_1	GGTCTTCTACATTACGACCACGGTGC
	D111H_FOR_2	GCCGTGGCACCAGCATTACATCTTCTGGGC
	D111H_REV_2	GGGTCTTCTACATTACGACCACGGTGCCGG
D111N	D111N_FOR_1	CCGTGGCACCAGAATTACATCTTCTG
	D111N_REV_1	GGTCTTCTACATTAAGACCACGGTGC
	D111N_FOR_2	GGCCGTGGCACCAGAATTACATCTTCTGGGCC
	D111N_REV_2	ACCGGCACCGTGGTCTTAATGTAGAAGACCCG
	D111N_FOR_3	CACCAGAATTACATCTTCTGGGCCCGAGAG
	D11N_REV_3	GATGTAATTCTGGTGCCACGGCCAGACATC

7.4.1 Site Directed Mutagenesis

Primers designed to introduce mutations at the required sites on the DOGDH gene were designed in PrimerX and synthesised by Eurofins genomics. The first sets of primers of each mutation have been labelled with "_1" in their primer name and were generally between 22 and 25 nucleotides long. Side directed mutagenesis (SDM) was attempted using the entire DOGDH-LIC3C as template using the QuickChange® protocol and the process was conducted by PCR.

The reaction components were mixed as described in Table 2.14 using KOD HotStart polymerase and the PCR reaction conducted using the program in Table 2.15. Following PCR amplification, the amplified samples were treated with 0.5 μL of Dpn1 for 4.5 h at 37 °C in order to digest any remaining template DNA, theoretically leaving just the mutant DNA. The Dpn1 treated samples (5 μL) were transformed into NovaBlue Singles™ Competent Cells (Merck) and transformed as described in *2.2.2.1* with kanamycin as antibiotic.

No colonies were observed for any of the amplified sample transformations despite repeated efforts and the use of Touchdown PCR (see *2.2.1.4*). Touchdown PCR was used in order to perform the annealing reaction at a range of temperatures. The lack of success suggested that the designed and utilised primers were not effective for this process. This may have been due to the primers being too short for QuickChange SDM with most suppliers recommending that primers be 25-45 nucleotides long in order to increase the likelihood and area over which the primers can anneal to the template. As a result of this, new and longer primers were designed in PrimerX and synthesised by Eurofins genomics. This second set of primers has been labelled with "_2" in their name in Table 7.1.

The QuickChange SDM process was repeated using standard and Touchdown PCR with this second batch of primers. Following the transformation process, colonies were observed on each plate. Starter cultures were prepared using these colonies with kanamycin as antibiotic. The DNA of interest was extracted from these overnights using a GenElute™ Plasmid Miniprep Kit. Following isolation, the concentration of the DNA products was measured using a UV spectrophotometer and found to be 30-50 ng μL^{-1}. The

isolated DNA was then sequenced by GATC in order to ascertain whether the mutations had been successful.

The sequencing results suggested that only the D111H mutation had been successfully performed (Figure 7.9) whilst the other samples containing DOGDH template DNA.

```
DOGDH          1  MLTPTELKQYREAGYLLIEDGLGPREVDCLRRAAAALYAQDSPDRTLEKD      50
                  ||||||||||||||||||||||||||||||||||||||||||||||||||
D111H          1  MLTPTELKQYREAGYLLIEDGLGPREVDCLRRAAAALYAQDSPDRTLEKD      50

DOGDH         51  GRTVRAVHGCHRRDPVCRDLVRHPRLLGPAMQILSGDVYVHQFKINAKAP     100
                  ||||||||||||||||||||||||||||||||||||||||||||||||||
D111H         51  GRTVRAVHGCHRRDPVCRDLVRHPRLLGPAMQILSGDVYVHQFKINAKAP     100

DOGDH        101  MTGDVWPWHQDYIFWAR     117
                  |||||||||||.||||||
D111H        101  MTGDVWPWHQHYIFWAR     117
```

Figure 7.9: Alignment of amino sequences of DOGDH and the isolated D111N sample. The output nucleotide sequences were translated using ExPASy tool (http://web.expasy.org/translate). For the purposes of representative alignment, the His-tag sequence was removed from the D111H translated sequence and only the first 117 amino acids of both are shown. The mutation site has been highlighted in bold and yellow. The D111H mutant appeared to have a His where DOGDH normally has Asp suggesting that the mutation was successful.

7.4.2 InFusion® Mutagenesis

The inconsistent nature of QuickChange SDM resulted in this method being deemed neither cost nor time efficient, resulting in the requirement of an alternative approach towards obtaining the desired mutants. One such alternative option was InFusion® mutagenesis which takes advantage of the streamlined InFusion® cloning system. Mutation with the InFusion® system requires the design of PCR primers possessing 15 bp overlaps with each other at the 5' end as well as introducing the desired point mutation.

As a pilot test, InFusion® mutagenesis was attempted for the D111N mutant. Primers were designed using PrimerX and have been labelled as D111N_FOR_3 and D111N_REV_3 in Table 7.1. The PCR amplification was performed by mixing the components in Table 2.14 using KOD HotStart Polymerase as the DNA polymerase. Touchdown PCR was performed (see *2.2.1.4*) and the amplified product treated with Dpn1. The treated sample (5 µL) was transformed into Stellar™ Competent Cells (Clontech) and colonies obtained. Small-scale cultures were prepared from the colonies and the DNA extracted from these using a GenElute™ Plasmid Miniprep Kit. The output DNA concentrations were between 40-50 ng µL^{-1} across the extracted samples.

The extracted DNA was sent for sequencing which confirmed that the D111N mutant had been successfully produced (Figure 7.10). This suggested that InFusion® mutagenesis was a promising method for generating the mutants of interests. Unfortunately, time constraints prevented pursuing this method for the other targets.

```
DOGDH          1 MLTPTELKQYREAGYLLIEDGLGPREVDCLRRAAAALYAQDSPDRTLEKD      50
                 |||||||||||||||||||||||||||||||||||||||||||||||||
D111N          1 MLTPTELKQYREAGYLLIEDGLGPREVDCLRRAAAALYAQDSPDRTLEKD      50

DOGDH         51 GRTVRAVHGCHRRDPVCRDLVRHPRLLGPAMQILSGDVYVHQFKINAKAP     100
                 |||||||||||||||||||||||||||||||||||||||||||||||||
D111N         51 GRTVRAVHGCHRRDPVCRDLVRHPRLLGPAMQILSGDVYVHQFKINAKAP     100

DOGDH        101 MTGDVWPWHQDYIFWAR    117
                 ||||||||||:||||||
D111N        101 MTGDVWPWHQNYIFWAR    117
```

Figure 7.10: Alignment of amino sequences of DOGDH and the isolated D111N sample. The output nucleotide sequences were translated using ExPASy tool (http://web.expasy.org/translate). For the purposes of representative alignment, the His-tag sequence was removed from the D111H translated sequence and only the first 117 amino acids of both are shown. The mutation site has been highlighted in bold and yellow. The isolated and sequenced sample appeared to have an Asn where DOGDH normally has Asp suggesting that the mutation was successful.

7.5 Discussion and Conclusions

Of the known microbial proline hydroxylases, DOGDH remains one of the few to not have a determined crystal structure. Extensive attempts were made to try and obtain diffracting crystals of DOGDH (described in Chapter 5) with little to no success. A DOGDH homology model was constructed in order to perform mutagenesis studies a homology model of DOGDH was pursued. This was performed using the Baker Robetta server using the DOGDH amino acid sequence as the input and a model constructed using EctD, a fellow 2-oxoglutarate dependent oxygenase, as a model.

The output DOGDH model was proposed to be a dimer and is presented in Figure 7.3. It is important to state that this is simply a model and it cannot be concluded with certainty that DOGDH is in fact a dimer; however, as a proof of principle, work proceeded with this model. The monomer of the DOGDH model is presented in Figure 7.4, and in general aligned with structural feature common to other 2-oxoglutarate dependent oxygenases by possessing a central cavity formed by 8 β-sheets that were surrounded by a series of α-helices. A central iron metal was modelled into the centre of the central active site cavity.

An overlay of the DOGDH model with an updated product bound EctD structure was used to dock T4HYP into the DOGDH active site. This active site model (Figure 7.5) was used to identify amino acid residues that may play roles in DOGDH activity. A potential H-X/D-H triad of H109, D111 and H215 was predicted to from the facial triad that coordinate to three sites of Fe allowing it to be anchored to the enzyme active site. W215, Q92 and D111 were hypothesised to play roles in positioning the L-proline in the active such that hydroxylation occurs at the C4-carbon of the pyrrole ring to from T4HYP, with D111 observed to from interactions with this specific carbon. K94 and N96 were proposed to potentially have a role in positioning 2-oxoglutarate in the active site perhaps so that it can bind to Fe in a bidentate manner on the opposite face of the H-X/D-H facial triad.

In order to narrow down the initial scope of mutagenesis trials it was decided that preliminary trials would focus on two areas. The first was the H-X/D-H triad of H109, D111 and H215 as a recent Codexis suggested that manipulating the Fe sphere to result in tighter metal binding could result in increased protein stability and activity.[256] The potential role of D111 in orientating proline in a manner to promote C4-hydroxylation also made it a candidate of interest for altering the selectivity of the DOGDH hydroxylation reaction. N96 and K98 were selected as the second group of mutagenesis candidates due to corresponding to amino acids in position 95 and 97 of MlC4H and SmC4H which in the case of the latter were found to play a role in the regioselectivity of the hydroxylation reaction of L-pipecolic acid together with having an effect on the enzyme activity.[205]

Eight initial mutations were selected and have been summarised in Table 7.1. QuickChange site directed mutagenesis of DOGDH-LIC3C was attempted and, despite extensive efforts to optimise the process, only the D111H mutant was produced. The seemingly inconsistent and time- and cost-inefficient nature of the QuickChange method in instances of this work suggested that an alternative protocol be pursued. InFusion® mutagenesis was selected as an alternative method and its effectiveness tested with just the D111N mutant. This alternative method was highly efficient and the desired mutant was swiftly obtained.

Unfortunately, time constraints prevented the generation of the remaining mutants by InFusion® mutagenesis as well as the assaying of the D111 mutants. If this work were to be continued, the availability of the D111 mutants would allow for experimentation in

order to determine whether the mutants express as soluble protein, are active and if the mutations have resulted in a change in activity. The InFusion® protocol could also be applied for the generation of the remaining mutants and their testing.

Chapter 8

Conclusions and Outlook

8.1 Discussion and Conclusions

The primary targets of this work were three microbial proline hydroxylases: DOGDH, StP3H and MlC4H. These three hydroxylases each possessed the natural ability to hydroxylate free L-proline to produce either *trans*-4-hydroxyproline (DOGDH), *cis*-3-hydroxyproline (StP3H) or *cis*-4-hydroxyproline (MlC4H). An interesting feature of proline hydroxylase members is their capacity to hydroxylate substrates other proline, with it being suggested that only the secondary amino moiety of the substrate be conserved, granting them significant synthetic potential for the production of intermediate and provision of access to unique chiral synthons.

Cloning and Expression of Targets

The hydroxylase targets were cloned into the pET-YSBLIC3C vector for the purposes of recombinantly expressing them in *E. coli* strains, with the added N-terminal hexahistidine allowing for their isolation by immobilised metal affinity chromatography. Small-scale expression testing suggested that all three hydroxylases generally exhibited poor soluble expression, however, this did not deter from performing larger scale expression in lysogeny broth (LB), as improved expression can be observed in some instances when growth scale is increased due to improved oxygenation of the culture in larger flasks.[257] Following immobilised Ni-affinity chromatography (IMAC) and SEC purification, DOGDH and MlC4H yielded unstable protein that appeared to be degrading as suggested by the multiple bands in the SDS-PAGE analysis gels and MS in the case of DOGDH. StP3H appeared to be the more stable of the hydroxylases yielding what appeared to be stable protein in higher yield.

Extensive attempts were made to stabilise DOGDH-LIC3C during and following purification, including the addition of protease inhibitors and glycerol to the purification buffers. It was assumed that Fe(II) may have been lost from the enzyme active site during

IMAC which may have resulted in the loss of stability. In order to have access to an alternative construct, DOGDH was cloned into the pET22b vector by InFusion® to produce a C-terminally hexahistidine tagged enzyme variant. DOGDH-pET22b showed improved soluble expression over the DOGDH-LIC3C construct resulting in large scale expression being pursued with greater confidence. DOGDH-pET22b was purified by IMAC and SEC to yield pure soluble, moderately stable protein at respectable yield.

Pursuit of a DOGDH Structure

The isolation of pure stable DOGDH-pET22b permitted the pursuit of diffraction quality crystals. This was attempted by performing 96 well crystallisation trials using commercial screens in the sitting drop format. Initial trials at a varied protein concentration resulted in either empty drops or substantial protein precipitate and debris in the drops. It was hypothesised that this may have resulted from stability issues with DOGDH as this protein is dependent on Fe(II) and cofactors for catalysis. As a result, ligand addition trials were performed in which Fe(II)SO$_4$, L-proline and 2-oxoglutarate were added to the mother liquor solution of the commercial screen wells. This yielded microcrystalline material in some drops that unfortunately proved impossible to optimise in a 24-well hanging drop format. An additive screen was also attempted with little success. Direct addition of Fe(II)SO$_4$ to the purified protein prior to screening was also attempted, however, the Fe(II)SO$_4$ proved insoluble in solution, rapidly crashing out following addition. Co(II) was added as an alternative metal to serve as an Fe(II) mimetic following the studies of Koketsu and coworkers, but this caused the protein to precipitate upon addition of the slightest amounts.

Modifications of DOGDH itself prior to crystallisation trials were next attempted. His-tag cleavage of DOGDH-LIC3C resulted in protein that displayed little difference from crystallisation attempts of the tagged variant. Reductive methylation of DOGDH-pET22b was attempted but proved inconclusive due to substantial protein loss during the process, resulting in poor output yields and mostly empty drops in the crystallisation plates. A surface entropy reduced mutant of DODGH, DOGDH-SER, was also produced. The resulting protein expressed very well with nanocrystals observed in 96 well plates which

were successfully replicated in a 24 well hanging drop format. These drops could not be optimised further despite the addition of various additives and detergents.

Having had little success crystallising DOGDH despite extensive efforts, two sequence related homologs were selected following a BLAST search. These homologs were a hypothetical *trans*-4-proline hydroxylases from *Micromonospora* (MP4H) and *Nocardia* (NP4H) respectively. Both homologs were successfully cloned into the pET-YSBLIC3C vector, expressed in *E. coli* and purified. Crystallisation trials were performed and nanocrystals observed for MP4H. Optimisations were attempted with no success.

DOGDH proved difficult to crystallise despite extensive attempts and the pursuit of sequence related homologous proteins as a last resort. This resulted in the pursuit of a homology model using the Baker Robetta server. A DOGDH model was built based on a sequence related 2-oxogluatarate dependent oxygenase (EctD). The DOGDH model possessed structural features common to its enzyme family including a central core formed by eight β-sheets surrounded by α-helices. *Trans*-4-hydroxyproline, 2-oxoglutarate and Fe were modelled into the active site and used to identify residues with potential roles in catalysis. Although only a model, the information can be used as a solid starting point for further investigation of function and to inform mutagenesis trials.

Assay Development

In order to be able to rapidly analyse samples obtained following biotransformations with the enzymes of interest, it was necessary to develop an analytic assay. L-proline and its hydroxylated derivatives were not detectable by gas chromatography leaving liquid chromatography as the primary method of choice for analysis. As proline and its derivatives do not possess UV-detectable chromophores, it was necessary to derivatise with them with a detectable component. Following extensive effort and the use of Marfey's reagent and NBD-Cl as the detectable agent a reliable and robust precolumn derivatisation assay using FMOC-Cl as the derivatising agent. The FMOC-Cl derivatisation assay permitted the high-throughput and rapid analysis of reaction samples by HPLC and LC-MS.

Biotransformations using Proline Hydroxylases

Biotransformations of L-proline were performed at a small scale using DOGDH-LIC3C, StP3H-LIC3C and MlC4H-LIC3C obtained following IMAC and SEC. These transformations suggested that DOGDH and MlC4H were inactive while StP3H showed slight activity with very mild product conversions observed. The lack of activity in DOGDH and MlC4H may have resulted from the enzymes exhibiting very poor stability following purification with the protein appearing to quickly precipitate out of solution following reaction commencement. In addition to this, there was the possibility that poor activity was the result of inefficient cofactor recycling and poor oxygen supply to the reaction, as these enzymes require the oxidoreduction of Fe(II), 2-oxoglurarate and molecular oxygen for catalysis. This resulted in the reactions that followed being performed in whole cell systems as the cell can theoretically provide machinery that can regenerate and recycle the necessary cofactors, making it an ideal reaction system if successfully utilised.

Cells expressing DOGDH-LIC3C, StP3H-LIC3C and MlC4H-LIC3C were produced by fermentation and the presence of enzymes of interest confirmed by western blot analysis. Initial whole cell biotransformations were performed at 150 mL reaction scales in 2 L baffled shake flasks. All three hydroxylases showed improved activity, with the greatest activity observed for DOGDH and StP3H. The ratio of reaction components was modified following consultation of the literature which resulted in markedly improved performance for DOGDH (the test system) with the conversions reaching what appeared to be their maximum after 8 hours of reaction time.

An interesting observation made in the baffled flask reactions of DOGDH, StP3H and MlC4H was that only hydroxylated products were observed in the LC-MS with no evidence of substrate. TLC analysis and product recovery followed by NMR analysis supported this observation. The final peak areas in the LC-MS analysis did not correspond to what would be assumed for a 100% conversion and the isolated yields following product recovery suggested conversions of 41% and 37% for DOGDH and StP3H respectively. This suggested that all of the L-proline added to the reaction was not available for reaction with it potentially being scavenged during the reaction time course by some means. Falcioni and colleagues suggested that *E. coli* possesses a proline uptake transporter that likely plays a

part in proline uptake by the cells for metabolic purposes with the transporter activating in the presence of proline.[242] It was highly probable that this transporter may have been responsible for the loss of proline from the reaction mixture during the reaction time course.

Having successfully performed whole cell biotransformation in shake flasks, reactions of L-proline were next performed MultiMax stirred tank reaction vessels in order to observe the performance of DOGDH (and StP3H) in a system mimicking industrial scale reactors. The effect of stirring rate was first investigated with it being observed that high stirring was harmful to the cells causing shearing. Following this, the effect of sparging air directly into the system was investigated. It was found that the direct sparging of air into the reaction mixture appeared to result in significantly improved conversion into product after 24 h compared to the non-sparged equivalent (49% vs 12%). It was however noted that the flow of air be carefully controlled as there were incidences of over bubbling and the expulsion of reaction mixture from the system when the air flow was set at too elevated a level. Regardless, these preliminary reactions in stirred tank vessels suggested that the tested hydroxylases were amenable to stirred tank reaction vessels and could hold industrial applicability.

The work that followed focussed on preliminary investigations of the substrate scope of DOGDH, StP3H and MlC4H. A substrate panel was assembled and screenings were performed in a BioLecter system. These preliminary studies suggested that the hydroxylases were active towards both L- and D-proline while DOGDH was also active towards L-pipecolic acid. DOGDH was not active towards the structural isomers of L-pipecolic acid, suggesting that the positioning of the carboxylic acid group relative the amino group plays an important role in substrate recognition. Finally, none of the hydroxylases appeared to recognised substrates containing an azetidine ring or β-lactam ring. Though the majority of the reactions in the panel screen appeared to yield negative results, the fact that the BioLecter was successfully utilised to perform 96 biotransformations in parallel was very promising, as this provides a method for future screenings.

8.2 Future Directions

The Crystallisation of DOGDH

Though a DOGDH model was constructed and successfully used to speculate about the role of potentially key amino acid residues within the enzyme active in catalysis. This did not change the fact that this information is resulting from a model and may not relate to the actual DOGDH structure. As a result, an X-ray structure of DOGDH remains highly desirable; however, in order to obtain such information, diffraction quality crystals of DOGDH must be obtained.

Stability issues resulting following the purification of DOGDH may have been one of the key factors influencing the enzyme's inability to crystallise. It was possible that Fe(II) was lost from the enzyme active site during IMAC with attempts to supplement the enzyme with it in the form of $Fe(II)SO_4$ proving unsuccessful due the insoluble nature of the compound. A means by which the loss of Fe from the DOGDH active site could be prevented is to avoid the use of IMAC as a purification step all together, as this could in theory keep the metal in the active site undisturbed.

Koketsu and colleagues were able to successfully crystallise MlC4H doing so by introducing a strep-tag II peptide at the N-terminus of the enzyme and purifying it using a Strep-Tactin resin followed by anion exchange chromatography and SEC.[205]   The strep-tag II introduces a small addition of 8 amino acids (Trp-Ser-His-Pro-Gln-Phe-Gly-Lys) to the N-terminal end (or C-terminal depending on preference) that serve as a synthetic peptide that can form affinity interactions with the specifically engineering streptavidin based Strep-Tactin resin. The Step-tag affinity purification process generally works on the principle of binding the post-harvest and centrifugation supernatant containing tagged protein onto a column packed with the Strep-Tactin resin, following binding, the resin is washed with phosphate-buffer saline and the target protein eluted by applying dethiobitin to the column which competes with the tagged enzyme for the biotin binding pocket within the streptavidin-based resin allowing for the gentle elution of the protein of interest. This method of purification allows for the isolation of protein without the need for IMAC allowing for the isolation of highly active and stable proteins under physiological conditions, making it a highly appealing purification strategy for DOGDH.

If this work were to be continued, it would be worth constructing an N-terminal Strep-tagged variant of DOGDH and applying the same purification strategy as employed by Koketsu and colleagues. This could potentially result in the isolation highly stable DOGDH that could be successfully crystallised. The strep-tag is available in the pET52b vector and DOGDH could perhaps be cloned into it using the InFusion® method which would allow for the rapid generation of the strep-tagged variant.

Biotransformation with Proline Hydroxylases

Reactions of L-proline in Shake Flasks

The shake-flask reactions for DOGDH, StP3H and MlC4H with L-proline initially appeared to proceed to completion based on a qualitative view of the out LC-MS sample data which suggested that only product was present in the later time point samples. TLC and NMR data supported this, however, the LC-MS peak areas of the respective products did not correspond to what would be expected for 100% hydroxylated product and the isolated yields following product recovery suggested 41% and 37% conversion for DOGDH and StP3H respectively. This suggested L-proline was being scavenged by the cells with it being suggested that a proline uptake transporter was responsible for this.[242]

It would be worthwhile to repeat these transformations using strains on *E. coli* that have been engineered to knock out this proline uptake transporter as it could lead to vastly different conversions and allow for the assessment of enzyme activity when proline is readily available. This could also allow for a more accurate quantification of the conversions of proline to hydroxyprolines with these hydroxylases.

Stirred Tank Reactions

Preliminary reactions in stirred tank vessels suggested that sparging air into the system and stirring rate were key parameters for enzyme performance. If these hydroxylases were to be applied in industry, it is highly likely that large scale reactions would be performed in stirred tank vessels. As a result, it would be highly advantageous to further optimise and characterise the parameters that are integral for the optimal performance of proline hydroxylases in stirred tank reaction vessels. This could be tackled by further tuning the

sparging method, manipulating the stirring rate and shifting the reaction scale. It is likely that detailed calculations will be necessary for this.

Alternative Substrates

The BioLecter was shown to have significant potential for the high-throughput screening of substrates, allowing for its use in future screenings and the rapid investigation of the substrate scope of proline hydroxylases. This assay could also be deployed for the rapid screening of DOGDH mutants.

Mutagenesis

Time constraints permitted the isolation of the preliminary mutagenesis targets. InFusion® mutagenesis was found to be a more effective strategy suggesting that it could be employed for the generation of the mutant targets. Upon their generation, the expression of the mutants in *E. coli*c could be assessed and the soluble mutants grown on a larger scale for both purification and whole cell biotransformation purposes. Initial transformations could be performed with L-proline as substrate to investigate if the mutations (particularly D111) have had any effect on enzyme activity or selectivity. Following initial screenings, further mutants could be selected depending on the results.

8.3 Final Conclusion

Microbial proline hydroxylases are members of the 2-oxoglutarate dependent oxygenase superfamily that possess the ideal feature of being able to hydroxylate proline in isolation with high regio- and enantioselectivity. This work focussed on three hydroxylases: DOGDH, StP3H and MlC4H, which catalysed the *trans*-4, *cis*-3 and *cis*-4 hydroxylation L-proline respectively. The three targets were successfully cloned, expressed and purified, though DOGDH and MlC4H showed poor stability following isolation.

An assay was successfully developed for the monitoring of biotransformations and reactions of L-proline performed with purified enzymes and enzymes in whole cell environments. As the 2-oxoglutarate dependent oxygenases are enzymes that require iron (II) and 2-oxoglutarate for catalysis, the reactions performed in the whole cell environment showed greater conversions than the *in vitro* equivalent, which was likely the result of the

cells' ability and machinery for cofactor regeneration and provision being utilised by the enzymes. It did however appear that proline was scavenged by the cells during the reaction period by a proline uptake transporter. It may be necessary that this transporter be inhibited by strain engineering for a more thorough analysis and investigation of the applicability of these enzymes in a whole cell system. Whole cell biotransformations were also performed in the stirred tank vessels to better simulate industrial processes and it was determined that stirring rate and the direct provision of oxygen by sparging may be integral for enzyme function.

A high-throughput assay was developed for the rapid screening of substrates with the hydroxylases using a BioLecter system and tested with a panel of substrates which in this instance showed a somewhat limited selectivity in the enzymes. This assay is however very promising due to its potential use in future work for the screening of further substrates as well as enzyme mutants.

Extensive attempts were made to obtain diffracting crystals of DOGDH as it remained one of the few known microbial proline hydroxylases to not have a determined atomic structure prior. Disappointingly, the best crystals to be obtained were too small to diffract and proved difficult to optimise further. Regardless, a homology model of DOGDH was built using the Baker Robetta server and proved useful for identifying the enzyme active site and amino acid residues that may play are role in Fe(II) coordination and enzyme regioselectivity. This model could serve as a good starting point to inform future mutagenesis studies to further characterise DOGDH, optimise its activity and enhance its industrial potential.

List of Abbreviations

2ACA   (*S*)-azetidine-2-carboxylic acid

2-OG   2-oxoglutarate

2OGDOs   2-oxoglutarate-dependent oxygenases

3HB   3-hydroxybenzoate

3HB6H   3-hydroxybenzoate-6-hydroxylase

*Aae*UPO   Unspecific peroxygenase from *Agrocybe aegerita*

ABC   Dimethylamine-borane complex

ADAM   1-adamantamine

AEBSF   4-(2-aminoethyl)benzenesulfonyl fluoride hydrochloride

AH   Aspartyl-β-hydroxylase

ANS   Anthocyanidin synthase

APS   Ammonium persulphate

AsH   Asparaginyl-β-hydroxylase

β-lactam 1   (1S,2R,5S,6R)-3-azatricyclo[4.2.1.02,5]non-7-en-4-on

β-lactam 2   6-azabicyclo[3.2.0]heptan-7-one

BLAST	Basic Local Alignment Search Tool
Boc	*tert*-butyloxycarbonyl
BVMO	Baeyer-Villiger Monooxygenase
CAS	Clavaminic acid synthase
C3HyP	*Cis*-3-hydroxyproline
C4HyP	*Cis*-4-hydroxyproline
CDCl$_3$	Deuterochloroform
*Cci*UPO	Unspecific peroxygenase from *Coprinopsis cinerea*
CHMO	Cyclohexanone Mononoxygenase
*Cra*UPO	Unspecific peroxygenase from *Coprinellus radians*
Cis-3-HyPip	*Cis*-3-hydroxypipecolic acid
Cis-5-HyPip	*Cis*-5-hydroxypipecolic acid
Cis-P3H	*Cis*-proline-3-hydroxylase
Cis-P4H	*Cis*-proline-4-hydroxylase
CPO	Chloroperoxidase
CYPs/P450s	Cytochromes P450
dATP	Deoxyadenosine triphosphate

Dansyl chloride	5-(DimethylAmino)Naphthalene-1-SulfonyL chloride
DFNDB	1,5-difluoro-2,4-dinitrobenzene
DFR	Dihydroflavolol reductase
DMPK	Drug Metabolism and Pharmacokinetics
DNA	Deoxyribonucleic acid
DNTPs	Deoxyribonucleotide triphosphate
DOGDH	*Trans*-4-proline hydroxylase from *Dactylsporangium* sp. RH1
DTT	Dithiothreitol
dTTP	Deoxythymidine triphosphate
EctD	Ectoine hydroxylase
EDTA	Ethylenediaminetetraacetic acid
F3H	Flavone-3β-hydroxylase
FAD	Flavin Adenine Dinucleotide
$FADH_2$	Flavin Adenine Dinucleotide Hydroquinone Form
FDH	Formate Dehydrogenase
$Fe(II)SO_4$	Iron (II) Sulphate
FIH	Factor inhibiting HIF

FLS	Flavanol synthase
FMN	Flavin Mononucleotide
FMOC-Cl	9-Fluorenylmethyl chloroformate
FMOC-GA	FMOC derivatised glutamic acid
FOR	Forward Primer
FSI	Flavone synthase I
G6PDH	Glucose-6-phosphate dehydrogenase
GlP3H	*Trans*-proline-3-hydroxylase from *Glarea lozoyensis*
HbpA	2-hydroxybiphenyl-3-monooxygenase from *Pseudomonas azelaica*
HIF	Hypoxia-inducible factor
HPLC	High Performance Liquid Chromatography
HTP	Fungal haem-thiolate peroxidase
H-X/D-H	His-X/Asp-His Motif
HyPip	Hydroxypipecolic acid
HyPro/HyP	Hydroxyproline
IEC	Ion Exchange Chromatography
IPA	Isonipecotic acid

IPTG	Isopropyl β-D-1-thiogalactopyranoside
IMAC	Immobilised Metal Affinity Chromatography
KP	Potassium phosphate
LB	Lysogeny Broth
LC-MS	Liquid chromatography–mass spectrometry
L-FDAA	Nα-(2,4-Dinitro-5-fluorophenyl)-L-alaninamide
LH	Lysyl hydroxylase
LIC	Ligation Independent Cloning
L-Pro	L-proline
MeCN	Acetonitrile
MlC4H	*Cis*-proline-4-hydroxylase from *Mesorhizobium loti*
*Mro*UPO	Unspecific peroxygenase from *Marasmius rotula*
MTBE	Methyl tert-butyl ether
MP4H	Hypothetical *trans*-4-hydroxylase from *Micromonospora* sp. CNB394
NAD	Nicotinamide Adenine Dinucleotide
NADPH	Nicotinamide Dinucleotide Phosphate
NBD-Cl	4-fluoro-7-nitrobenzofurazan

NDSB	Non-Detergents Sulfobetaines
NMR	Nuclear magnetic resonance
NP4H	Hypothetical trans-4-hydroxylase from *Nocardia* sp. BMG111209
OD_{600}	Optical Density at 600 nm
P3H	Prolyl-3-hydroxylase
$P450_{BM3}$	Fatty acid hydroxylase cytochrome P450 from *Bacillus megaterium*
$P450_{BS\beta}$	Fatty acid peroxygenase from *Bacillus subtilis*
$P450_{cam}$	Camphor hydroxylase cytochrome P450 from *Pseudomonas putida*
$P450_{Cla}$	Fatty acid peroxygenase from *Clostridium acetobutylicum*
P4H	Prolyl-4-hydroxylase
PAH	Proclavaminate hydroxylase
PArHs	Polycyclic Aromatic Hydrocarbons
PCBs	Polychlorinated Biphenyls
PCR	Polymerase Chain Reaction
PDB	Protein Data Bank
PEG	Polyethylene glycol
PHBH	*Para*-hydroxybenzoate hydroxylase

PHDs   Prolyl hydroxylase domain containing isozymes

Pip   Pipecolic acid

PNP   *p*-nitrophenol

REV   Reverse Primer

RP-HPLC   Reverse phase high performance liquid chromatography

S3PCA   (*S*)-(+)-3-piperidine carboxylic acid

SARM   Selective Androgen Receptor Modulator

SDM   Site Directed Mutagenesis

SDS   Sodium dodecyl sulfate

SDS-PAGE   Sodium dodecyl sulfate polyacrylamide gel electrophoresis

SEC   Size Exclusion Chromatography

SER   Surface Entropy Reduction

SmC4H   *Cis*-proline-4-hydroxylase from *Sinorhizobium meliloti*

StP3H   *Cis*-proline-3-hydroxylase from *Streptomyces* sp. TH1

StyA2B   Styrene Mononoxygenase from *Rhodococcus opacus*

T4HYP/t4HyP   *Trans*-4-hydroxy-L-proline

TAE   Tris Acetate EDTA

TauD	Taurine Dioxygenase
TBST	Tris-buffered saline + Tween 20
TEMED	Tetramethylethylenediamine
THF	Tetrahydrofuran
Trans-P3H	*Trans*-proline-3-hydroxylase
Trans-P4H	*Trans*-proline-4-hydroxylase
Tris	tris(hydroxymethyl)aminomethane
UPO	Unspecific Peroxygenase
UV-vis	Ultraviolet-visible

References

1.   Reetz, M. T., Biocatalysis in Organic Chemistry and Biotechnology: Past, Present, and Future. *J Am Chem Soc* **2013,** *135* (34), 12480-12496.

2.   Bugg, T. D. H., Introduction to Enzyme and Coenzyme Chemistry. *Wiley, Chichester, UK* **2002**.

3.   Bugg, T. D. H., The development of mechanistic enzymology in the 20th century. *Nat Prod Rep* **2001,** *18* (5), 465-493.

4.   Gal, J., The discovery of biological enantioselectivity: Louis Pasteur and the fermentation of tartaric acid, 1857 - A review and analysis 150 yr later. *Chirality* **2008,** *20* (1), 5-19.

5.   Koshland, D. E., The Key-Lock Theory and the Induced Fit Theory. *Angew Chem Int Edit* **1994,** *33* (23-24), 2375-2378.

6.   Nestl, B. M.; Nebel, B. A.; Hauer, B., Recent progress in industrial biocatalysis. *Curr Opin Chem Biol* **2011,** *15* (2), 187-193.

7.   Nestl, B. M.; Hammer, S. C.; Nebel, B. A.; Hauer, B., New Generation of Biocatalysts for Organic Synthesis. *Angew Chem Int Edit* **2014,** *53* (12), 3070-3095.

8.   Tao, J. H.; Zhao, L. S.; Ran, N. Q., Recent advances in developing chemoenzymatic processes for active pharmaceutical ingredients. *Org Process Res Dev* **2007,** *11* (2), 259-267.

9.   Straathof, A. J. J.; Panke, S.; Schmid, A., The production of fine chemicals by biotransformations. *Curr Opin Biotech* **2002,** *13* (6), 548-556.

10.   Potdar, M. K.; Kelso, G. F.; Schwarz, L.; Zhang, C. F.; Hearn, M. T. W., Recent Developments in Chemical Synthesis with Biocatalysts in Ionic Liquids. *Molecules* **2015,** *20* (9), 16788-16816.

11.   Grogan, G., Practical Biotransformations: A Beginner's Guide. *Wiley, Chichester, UK* **2009**.

12.   Hogg, J. A., Steroids, the Steroid Community, and Upjohn in Perspective - a Profile of Innovation. *Steroids* **1992,** *57* (12), 593-616.

13.   Guengerich, F. P.; Munro, A. W., Unusual Cytochrome P450 Enzymes and Reactions. *J Biol Chem* **2013,** *288* (24), 17065-17073.

14.   Breuer, M.; Ditrich, K.; Habicher, T.; Hauer, B.; Kesseler, M.; Sturmer, R.; Zelinski, T., Industrial methods for the production of optically active intermediates. *Angew Chem Int Edit* **2004,** *43* (7), 788-824.

15.   Roduner, E.; Kaim, W.; Sarkar, B.; Urlacher, V. B.; Pleiss, J.; Glaser, R.; Einicke, W. D.; Sprenger, G. A.; Beifuss, U.; Klemm, E.; Liebner, C.; Hieronymus, H.; Hsu, S. F.; Plietker, B.; Laschat, S., Selective Catalytic Oxidation of C-H Bonds with Molecular Oxygen. *ChemCatChem* **2013,** *5* (1), 82-112.

16.   Haber, J.; Mlodnicka, T., Oxidation of Hydrocarbons with Dioxygen. *J Mol Catal* **1992,** *74* (1-3), 131-141.

17.   Balcells, D.; Clot, E.; Eisenstein, O., C-H Bond Activation in Transition Metal Species from a Computational Perspective. *Chem Rev* **2010,** *110* (2), 749-823.

18. Gunay, A.; Theopold, K. H., C-H Bond Activations by Metal Oxo Compounds. *Chem Rev* **2010,** *110* (2), 1060-1081.

19. McKersie, B. D., Leshem, Y. Y., Stress and stress coping in cultivated plants. *Kluwer Academic Publishers* **1994**.

20. Faber, K., Biotransformations in organic chemistry: a textbook 5th edition. *Springer-Verlag* **2004**.

21. Jacobsen, E. N.; Marko, I.; Mungall, W. S.; Schroder, G.; Sharpless, K. B., Asymmetric Dihydroxylation Via Ligand-Accelerated Catalysis. *J Am Chem Soc* **1988,** *110* (6), 1968-1970.

22. Schaus, S. E.; Brandes, B. D.; Larrow, J. F.; Tokunaga, M.; Hansen, K. B.; Gould, A. E.; Furrow, M. E.; Jacobsen, E. N., Highly selective hydrolytic kinetic resolution of terminal epoxides catalyzed by chiral (salen)Co-III complexes. Practical synthesis of enantioenriched terminal epoxides and 1,2-diols. *J Am Chem Soc* **2002,** *124* (7), 1307-1315.

23. Munro, A. W.; Girvan, H. M.; McLean, K. J., Variations on a (t)heme - novel mechanisms, redox partners and catalytic functions in the cytochrome P450 superfamily. *Nat Prod Rep* **2007,** *24* (3), 585-609.

24. Holtmann, D.; Fraaije, M. W.; Arends, I. W. C. E.; Opperman, D. J.; Hollmann, F., The taming of oxygen: biocatalytic oxyfunctionalisations. *ChemComm* **2014,** *50* (87), 13180-13200.

25. Mihovilovic, M. D.; Muller, B.; Stanetty, P., Monooxygenase-mediated Baeyer-Villiger oxidations. *Eur J Org Chem* **2002,** (22), 3711-3730.

26. Huijbers, M. M. E.; Montersino, S.; Westphal, A. H.; Tischler, D.; van Berkel, W. J. H., Flavin dependent monooxygenases. *Arch Biochem Biophys* **2014,** *544*, 2-17.

27. van Berkel, W. J. H.; Kamerbeek, N. M.; Fraaije, M. W., Flavoprotein monooxygenases, a diverse class of oxidative biocatalysts. *J Biotechnol* **2006,** *124* (4), 670-689.

28. Pazmino, D. E. T.; Winkler, M.; Glieder, A.; Fraaije, M. W., Monooxygenases as biocatalysts: Classification, mechanistic aspects and biotechnological applications. *J Biotechnol* **2010,** *146* (1-2), 9-24.

29. Macheroux, P.; Kappes, B.; Ealick, S. E., Flavogenomics - a genomic and structural view of flavin-dependent proteins. *FEBS J* **2011,** *278* (15), 2625-2634.

30. Lin, S. J.; Van Lanen, S. G.; Shen, B., Characterization of the two-component, FAD-dependent monooxygenase SgcC that requires carrier protein-tethered substrates for the biosynthesis of the enediyne antitumor antibiotic C-1027. *J Am Chem Soc* **2008,** *130* (20), 6616-6623.

31. Leisch, H.; Morley, K.; Lau, P. C. K., Baeyer-Villiger Monooxygenases: More Than Just Green Chemistry. *Chem Rev* **2011,** *111* (7), 4165-4222.

32. Krow, G. R., The Baeyer–VilligerOxidation of Ketones and Aldehydes. In *Organic Reactions*, John Wiley & Sons, Inc.: 2004.

33. Pazmino, D. E. T.; Dudek, H. M.; Fraaije, M. W., Baeyer-Villiger monooxygenases: recent advances and future challenges. *Curr Opin Chem Biol* **2010,** *14* (2), 138-144.

34. Kirschner, A.; Bornscheuer, U. T., Kinetic resolution of 4-hydroxy-2-ketones catalyzed by a Baeyer-Villiger monooxygenase. *Angew Chem Int Edit* **2006,** *45* (42), 7004-7006.

35. Pazmino, D. E. T., Fraaije, M. W., Future Directions in Biocatalysis. *Elsevier, Amsterdam, Netherlands* **2007** 107 - 128.

36.   Balke, K.; Kadow, M.; Mallin, H.; Sass, S.; Bornscheuer, U. T., Discovery, application and protein engineering of Baeyer-Villiger monooxygenases for organic synthesis. *Org Biomol Chem* **2012,** *10* (31), 6249-6265.

37.   Bucko, M.; Gemeiner, P.; Schenkmayerova, A.; Krajcovic, T.; Rudroff, F.; Mihovilovic, M. D., Baeyer-Villiger oxidations: biotechnological approach. *Appl Microbiol Biotech* **2016,** *100* (15), 6585-6599.

38.   Donoghue, N. A.; Norris, D. B.; Trudgill, P. W., Purification and Properties of Cyclohexanone Oxygenase from Nocardia-Globerula Cl1 and Acinetobacter Ncib-9871. *Eur J Biochem* **1976,** *63* (1), 175-192.

39.   Schwab, J. M.; Li, W. B.; Thomas, L. P., Cyclohexanone Oxygenase - Stereochemistry, Enantioselectivity, and Regioselectivity of an Enzyme-Catalyzed Baeyer-Villiger Reaction. *J Am Chem Soc* **1983,** *105* (14), 4800-4808.

40.   Ryerson, C. C.; Ballou, D. P.; Walsh, C., Mechanistic Studies on Cyclohexanone Oxygenase. *Biochemistry* **1982,** *21* (11), 2644-2655.

41.   Sheng, D. W.; Ballou, D. P.; Massey, V., Mechanistic studies of cyclohexanone monooxygenase: Chemical properties of intermediates involved in catalysis. *Biochemistry* **2001,** *40* (37), 11156-11167.

42.   Mirza, I. A.; Yachnin, B. J.; Wang, S. Z.; Grosse, S.; Bergeron, H.; Imura, A.; Iwaki, H.; Hasegawa, Y.; Lau, P. C. K.; Berghuis, A. M., Crystal Structures of Cyclohexanone Monooxygenase Reveal Complex Domain Movements and a Sliding Cofactor. *J Am Chem Soc* **2009,** *131* (25), 8848-8854.

43.   Bayer, T.; Milker, S.; Wiesinger, T.; Rudroff, F.; Mihovilovic, M. D., Designer Microorganisms for Optimized Redox Cascade Reactions - Challenges and Future Perspectives. *Adv Synth Catal* **2015,** *357* (8), 1587-1618.

44.   Mihovilovic, M. D.; Grotzl, B.; Kandioller, W.; Muskotal, A.; Snajdrova, R.; Rudroff, F.; Spreitzer, H., Recombinant whole-cell mediated Baeyer-Villiger oxidation of perhydropyran-type ketones. *Chem Biodivers* **2008,** *5* (3), 490-498.

45.   Muschiol, J.; Peters, C.; Oberleitner, N.; Mihovilovic, M. D.; Bornscheuer, U. T.; Rudroff, F., Cascade catalysis - strategies and challenges en route to preparative synthetic biology. *ChemComm* **2015,** *51* (27), 5798-5811.

46.   Lima-Ramos, J.; Neto, W.; Woodley, J. M., Engineering of Biocatalysts and Biocatalytic Processes. *Top Catal* **2014,** *57* (5), 301-320.

47.   Cuetos, A.; Rioz-Martinez, A.; Valenzuela, M. L.; Lavandera, I.; de Gonzalo, G.; Carriedo, G. A.; Gotor, V., Immobilized redox enzymatic catalysts: Baeyer-Villiger monooxygenases supported on polyphosphazenes. *J Mol Catal B-Enzym* **2012,** *74* (3-4), 178-183.

48.   Mallin, H.; Wulf, H.; Bornscheuer, U. T., A self-sufficient Baeyer-Villiger biocatalysis system for the synthesis of epsilon-caprolactone from cyclohexanol. *Enzyme Microb Technol* **2013,** *53* (4), 283-287.

49.   Rebros, M.; Liptak, L.; Rosenberg, M.; Bucko, M.; Gemeiner, P., Biocatalysis with Escherichia coli-overexpressing cyclopentanone monooxygenase immobilized in polyvinyl alcohol gel. *Lett Appl Microbiol* **2014,** *58* (6), 556-563.

50.   Schrewe, M.; Julsing, M. K.; Buhler, B.; Schmid, A., Whole-cell biocatalysis for selective and productive C-O functional group introduction and modification. *Chem Soc Rev* **2013,** *42* (15), 6346-6377.

51. Ellis, H. R., The FMN-dependent two-component monooxygenase systems. *Arch Biochem Biophys* **2010,** *497* (1-2), 1-12.

52. Tischler, D.; Kermer, R.; Groning, J. A. D.; Kaschabek, S. R.; van Berkel, W. J. H.; Schlomann, M., StyA1 and StyA2B from Rhodococcus opacus 1CP: a Multifunctional Styrene Monooxygenase System. *J Bacteriol* **2010,** *192* (19), 5220-5227.

53. Badone, D.; Guzzi, U., Synthesis of the Potent and Selective Atypical Beta-Adrenergic Agonist Sr 59062a. *Bioorg Med Chem Lett* **1994,** *4* (16), 1921-1924.

54. Besse, P.; Veschambre, H., Chemical and Biological Synthesis of Chiral Epoxides. *Tetrahedron* **1994,** *50* (30), 8885-8927.

55. Rao, A. V. R.; Gurjar, M. K.; Kaiwar, V., Enantioselective Catalytic Reductions of Ketones with New 4 Membered Oxazaborolidines - Application to (S)-Tetramisole. *Tetrahedron-Asymmetr* **1992,** *3* (7), 859-862.

56. O'Leary, N. D.; O'Connor, K. E.; Dobson, A. D. W., Biochemistry, genetics and physiology of microbial styrene degradation. *Fems Microbiol Rev* **2002,** *26* (4), 403-417.

57. Hollmann, F.; Lin, P. C.; Witholt, B.; Schmid, A., Stereospecific biocatalytic epoxidation: The first example of direct regeneration of a FAD-dependent monooxygenase for catalysis. *J Am Chem Soc* **2003,** *125* (27), 8209-8217.

58. Tischler, D.; Eulberg, D.; Lakner, S.; Kaschabek, S. R.; van Berkel, W. J. H.; Schlomann, M., Identification of a Novel Self-Sufficient Styrene Monooxygenase from Rhodococcus opacus 1CP. *J Bacteriol* **2009,** *191* (15), 4996-5009.

59. van Hellemond, E. W.; Janssen, D. B.; Fraaije, M. W., Discovery of a novel styrene monooxygenase originating from the metagenome. *Appl Environ Microbiol* **2007,** *73* (18), 5832-5839.

60. Lee, K., Benzene-induced uncoupling of naphthalene dioxygenase activity and enzyme inactivation by production of hydrogen peroxide. *J Bacteriol* **1999,** *181* (9), 2719-2725.

61. Massey, V., Activation of Molecular-Oxygen by Flavins and Flavoproteins. *J Biol Chem* **1994,** *269* (36), 22459-22462.

62. Munro, A. W.; Girvan, H. M.; McLean, K. J., Cytochrome P450 - redox partner fusion enzymes. *Bba-Gen Subjects* **2007,** *1770* (3), 345-359.

63. Roberts, G. A.; Grogan, G.; Greter, A.; Flitsch, S. L.; Turner, N. J., Identification of a new class of cytochrome P450 from a Rhodococcus sp. *J Bacteriol* **2002,** *184* (14), 3898-3908.

64. Srivastava, D.; Schuermann, J. P.; White, T. A.; Krishnan, N.; Sanyal, N.; Hura, G. L.; Tan, A. M.; Henzl, M. T.; Becker, D. F.; Tanner, J. J., Crystal structure of the bifunctional proline utilization A flavoenzyme from Bradyrhizobium japonicum. *P Natl Acad Sci USA* **2010,** *107* (7), 2878-2883.

65. Arunachalam, U.; Massey, V.; Vaidyanathan, C. S., P-Hydroxyphenylacetate-3-Hydroxylase - a 2-Protein Component Enzyme. *J Biol Chem* **1992,** *267* (36), 25848-25855.

66. Becker, D.; Schrader, T.; Andreesen, J. R., Two-component flavin-dependent pyrrole-2-carboxylate monooxygenase from Rhodococcus sp. *Eur J Biochem* **1997,** *249* (3), 739-747.

67. Yeh, E.; Garneau, S.; Walsh, C. T., Robust in vitro activity of RebF and RebH, a two-component reductase/halogenase, generating 7-chlorotryptophan during rebeccamycin biosynthesis. *P Natl Acad Sci USA* **2005,** *102* (11), 3960-3965.

68.   Deng, D. Y.; Li, X. B.; Fang, X. P.; Sun, G. P., Characterization of two components of the 2-naphthoate monooxygenase system from Burkholderia sp strain JT1500. *Fems Microbiol Lett* **2007,** *273* (1), 22-27.

69.   Kadiyala, V.; Spain, J. C., A two-component monooxygenase catalyzes both the hydroxylation of p-nitrophenol and the oxidative release of nitrite from 4-nitrocatechol in Bacillus sphaericus JS905. *Appl Environ Microbiol* **1998,** *64* (7), 2479-2484.

70.   Holland, H. L.; Weber, H. K., Enzymatic hydroxylation reactions. *Curr Opin Biotech* **2000,** *11* (6), 547-553.

71.   Urlacher, V. B.; Girhard, M., Cytochrome P450 monooxygenases: an update on perspectives for synthetic application. *Trends Biotechnol* **2012,** *30* (1), 26-36.

72.   Bernhardt, R.; Urlacher, V. B., Cytochromes P450 as promising catalysts for biotechnological application: chances and limitations. *Appl Microbiol Biotechnol* **2014,** *98* (14), 6185-6203.

73.   O'Reilly, E.; Kohler, V.; Flitsch, S. L.; Turner, N. J., Cytochromes P450 as useful biocatalysts: addressing the limitations. *ChemComm* **2011,** *47* (9), 2490-2501.

74.   Bernhardt, R., Cytochromes P450 as versatile biocatalysts. *J Biotechnol* **2006,** *124* (1), 128-145.

75.   Omura, T.; Sato, R., Carbon Monoxide-Binding Pigment of Liver Microsomes .I. Evidence for Its Hemoprotein Nature. *J Biol Chem* **1964,** *239* (7), 2370-&.

76.   Roper, L.; Grogan, G., Chapter 8 - Biocatalysis for Organic Chemists: Hydroxylations A2 - Goswami, Animesh. In *Organic Synthesis Using Biocatalysis*, Stewart, J. D., Ed. Academic Press: 2016; pp 213-241.

77.   Poulos, T. L.; Finzel, B. C.; Howard, A. J., High-Resolution Crystal-Structure of Cytochrome-P450cam. *J Mol Biol* **1987,** *195* (3), 687-700.

78.   Munro, A. W.; Leys, D. G.; McLean, K. J.; Marshall, K. R.; Ost, T. W. B.; Daff, S.; Miles, C. S.; Chapman, S. K.; Lysek, D. A.; Moser, C. C.; Page, C. C.; Dutton, P. L., P450BM3: the very model of a modern flavocytochrome. *Trends Biochem Sci* **2002,** *27* (5), 250-257.

79.   Poulos, T. L.; Finzel, B. C.; Howard, A. J., Crystal-Structure of Substrate-Free Pseudomonas-Putida Cytochrome-P-450. *Biochemistry* **1986,** *25* (18), 5314-5322.

80.   Girvan, H. M.; Waltham, T. N.; Neeli, R.; Collins, H. F.; McLean, K. J.; Scrutton, N. S.; Leys, D.; Munro, A. W., Flavocytochrome P450BM3 and the origin of CYP102 fusion species. *Biochem Soc T* **2006,** *34*, 1173-1177.

81.   Palmer, C. N. A.; Gustafsson, M. C. U.; Dobson, H.; von Wachenfeldt, C.; Wolf, C. R., Adaptive responses to fatty acids are mediated by the regulated expression of cytochromes P450. *Biochem Soc T* **1999,** *27* (4), 374-378.

82.   Fulco, A. J., P450bm-3 and Other Inducible Bacterial P450 Cytochromes - Biochemistry and Regulation. *Annu Rev Pharmacol* **1991,** *31*, 177-203.

83.   Ravichandran, K. G.; Boddupalli, S. S.; Hasemann, C. A.; Peterson, J. A.; Deisenhofer, J., Crystal-Structure of Hemoprotein Domain of P450bm-3, a Prototype for Microsomal P450s. *Science* **1993,** *261* (5122), 731-736.

84.   Poulos, T. L.; Johnson, E. F., Structures of Cytochrome P450 Enzymes. In *Cytochrome P450: Structure, Mechanism, and Biochemistry*, Ortiz de Montellano, P. R., Ed. Springer US: Boston, MA, 2005; pp 87-114.

85.   Ost, T. W. B.; Munro, A. W.; Mowat, C. G.; Taylor, P. R.; Pesseguiero, A.; Fulco, A. J.; Cho, A. K.; Cheesman, M. A.; Walkinshaw, M. D.; Chapman, S. K., Structural and spectroscopic analysis of the F393H mutant of flavocytochrome p450 BM3. *Biochemistry* **2001,** *40* (45), 13430-13438.

86.   Ost, T. W. B.; Miles, C. S.; Munro, A. W.; Murdoch, J.; Reid, G. A.; Chapman, S. K., Phenylalanine 393 exerts thermodynamic control over the heme of flavocytochrome P450BM3. *Biochemistry* **2001,** *40* (45), 13421-13429.

87.   Sono, M.; Roach, M. P.; Coulter, E. D.; Dawson, J. H., Heme-containing oxygenases. *Chem Rev* **1996,** *96* (7), 2841-2887.

88.   Bishop, G.; Nomura, T.; Yokota, T.; Montoya, T.; Castle, J.; Harrison, K.; Kushiro, T.; Kamiya, Y.; Yamaguchi, S.; Bancos, S.; Szatmari, A. M.; Szekeres, M., Dwarfism and cytochrome P450-mediated C-6 oxidation of plant steroid hormones. *Biochem Soc T* **2006,** *34*, 1199-1201.

89.   Ohnishi, T.; Yokota, T.; Mizutani, M., Insights into the function and evolution of P450s in plant steroid metabolism. *Phytochemistry* **2009,** *70* (17-18), 1918-1929.

90.   Takeyama, K.; Kitanaka, S.; Sato, T.; Kobori, M.; Yanagisawa, J.; Kato, S., 25-hydroxyvitamin D-3 1 alpha-hydroxylase and vitamin D synthesis. *Science* **1997,** *277* (5333), 1827-1830.

91.   Omdahl, J. L.; Bobrovnikova, E. A.; Choe, S.; Dwivedi, P. P.; May, B. K., Overview of regulatory cytochrome P450 enzymes of the vitamin D pathway. *Steroids* **2001,** *66* (3-5), 381-389.

92.   Gibbons, G. F., The role of cytochrome P450 in the regulation of cholesterol biosynthesis. *Lipids* **2002,** *37* (12), 1163-1170.

93.   Anzai, Y.; Li, S. Y.; Chaulagain, M. R.; Kinoshita, K.; Kato, F.; Montgomery, J.; Sherman, D. H., Functional Analysis of MycCI and MycG, Cytochrome P450 Enzymes Involved in Biosynthesis of Mycinamicin Macrolide Antibiotics. *Chem Biol* **2008,** *15* (9), 950-959.

94.   Guengerich, F. P.; Wu, Z. L.; Bartleson, C. J., Function of human cytochrome P450s: Characterization of the orphans. *Biochem Biophys Res Commun* **2005,** *338* (1), 465-469.

95.   Nelson, D. R., Introductory remarks on human CYPs. *Drug Metab Rev* **2002,** *34* (1-2), 1-5.

96.   Lewis, D. F. V., 57 varieties: the human cytochromes P450. *Pharmacogenomics* **2004,** *5* (3), 305-318.

97.   Guengerich, F. P., Cytochrome P450s and other enzymes in drug metabolism and toxicity. *Aaps J* **2006,** *8* (1), E101-E111.

98.   Prentis, R. A.; Lis, Y.; Walker, S. R., Pharmaceutical Innovation by the 7 Uk-Owned Pharmaceutical Companies (1964-1985). *Brit J Clin Pharmaco* **1988,** *25* (3), 387-396.

99.   Schroer, K.; Kittelmann, M.; Lutz, S., Recombinant Human Cytochrome P450 Monooxygenases for Drug Metabolite Synthesis. *Biotechnol Bioeng* **2010,** *106* (5), 699-706.

100.   Watanabe, Y.; Ito, T.; Shiomi, M.; Tsujita, Y.; Kuroda, M.; Arai, M.; Fukami, M.; Tamura, A., Preventive Effect of Pravastatin Sodium, a Potent Inhibitor of 3-Hydroxy-3-Methylglutaryl Coenzyme-a Reductase, on Coronary Atherosclerosis and Xanthoma in Whhl Rabbits. *Biochim Biophys Acta* **1988,** *960* (3), 294-302.

101.   Park, J. W.; Lee, J. K.; Kwon, T. J.; Yi, D. H.; Kim, Y. J.; Moon, S. H.; Suh, H. H.; Kang, S. M.; Park, Y. I., Bioconversion of compactin into pravastatin by Streptomyces sp. *Biotechnol Lett* **2003,** *25* (21), 1827-1831.

102.   Kumar, S., Engineering cytochrome P450 biocatalysts for biotechnology, medicine and bioremediation. *Expert Opin Drug Met* **2010,** *6* (2), 115-131.

103.   Kim, J. H.; Stansbury, K. H.; Walker, N. J.; Trush, M. A.; Strickland, P. T.; Sutter, T. R., Metabolism of benzo[a]pyrene and benzo[a]pyrene-7,8-diol by human cytochrome P450 1B1. *Carcinogenesis* **1998,** *19* (10), 1847-1853.

104.   Julsing, M. K.; Cornelissen, S.; Buhler, B.; Schmid, A., Heme-iron oxygenases: powerful industrial biocatalysts? *Curr Opin Chem Biol* **2008,** *12* (2), 177-186.

105.   Urlacher, V. B.; Eiben, S., Cytochrome P450 monooxygenases: perspectives for synthetic application. *Trends Biotechnol* **2006,** *24* (7), 324-330.

106.   Jung, S. T.; Lauchli, R.; Arnold, F. H., Cytochrome P450: taming a wild type enzyme. *Curr Opin Biotech* **2011,** *22* (6), 809-817.

107.   Siriphongphaew, A.; Pisnupong, P.; Wongkongkatep, J.; Inprakhon, P.; Vangnai, A. S.; Honda, K.; Ohtake, H.; Kato, J.; Ogawa, J.; Shimizu, S.; Urlacher, V. B.; Schmid, R. D.; Pongtharangkul, T., Development of a whole-cell biocatalyst co-expressing P450 monooxygenase and glucose dehydrogenase for synthesis of epoxyhexane. *Appl Microbiol Biotechnol* **2012,** *95* (2), 357-367.

108.   Ruijssenaars, H. J.; Sperling, E. M. G. M.; Wiegerinck, P. H. G.; Brands, F. T. L.; Wery, J.; de Bont, J. A. M., Testosterone 15 beta-hydroxylation by solvent tolerant Pseudomonas putida S12. *J Biotechnol* **2007,** *131* (2), 205-208.

109.   Girhard, M.; Machida, K.; Itoh, M.; Schmid, R. D.; Arisawa, A.; Urlacher, V. B., Regioselective biooxidation of (+)-valencene by recombinant E. coli expressing CYP109B1 from Bacillus subtilis in a two-liquid-phase system. *Microb Cell Fact* **2009,** *8.*

110.   Whitehouse, C. J. C.; Bell, S. G.; Wong, L. L., P450(BM3) (CYP102A1): connecting the dots. *Chem Soc Rev* **2012,** *41* (3), 1218-1260.

111.   Fasan, R.; Meharenna, Y. T.; Snow, C. D.; Poulos, T. L.; Arnold, F. H., Evolutionary History of a Specialized P450 Propane Monooxygenase. *J Mol Biol* **2008,** *383* (5), 1069-1080.

112.   Weber, E.; Seifert, A.; Antonovici, M.; Geinitz, C.; Pleiss, J.; Urlacher, V. B., Screening of a minimal enriched P450 BM3 mutant library for hydroxylation of cyclic and acyclic alkanes. *ChemComm* **2011,** *47* (3), 944-946.

113.   De Mot, R.; Parret, A. H. A., A novel class of self-sufficient cytochrome P450 monooxygenases in prokaryotes. *Trends Microbiol* **2002,** *10* (11), 502-508.

114.   Rylott, E. L.; Jackson, R. G.; Edwards, J.; Womack, G. L.; Seth-Smith, H. M. B.; Rathbone, D. A.; Strand, S. E.; Bruce, N. C., An explosive-degrading cytochrome P450 activity and its targeted application for the phytoremediation of RDX. *Nat Biotechnol* **2006,** *24* (2), 216-219.

115.   McLean, K. J.; Girvan, H. M.; Munro, A. W., Cytochrome P450/redox partner fusion enzymes: biotechnological and toxicological prospects. *Expert Opin Drug Met* **2007,** *3* (6), 847-863.

116.   Jackson, C. J.; Lamb, D. C.; Marezylo, T. H.; Warrilow, A. G. S.; Manning, N. J.; Lowe, D. J.; Kelly, D. E.; Kelly, S. L., A novel sterol 14 alpha-demethylase/ferredoxin fusion protein (MCCYP51FX) from Methylococcus capsulatus represents a new class of the cytochrome P450 superfamily. *J Biol Chem* **2002,** *277* (49), 46959-46965.

117.   Chun, Y. J.; Shimada, T.; Waterman, M. R.; Guengerich, F. P., Understanding electron transport systems of Streptomyces cytochrome P450. *Biochem Soc T* **2006,** *34*, 1183-1185.

118. Khatri, Y.; Hannemann, F.; Ewen, K. M.; Pistorius, D.; Perlova, O.; Kagawa, N.; Brachmann, A. O.; Muller, R.; Bernhardt, R., The CYPome of Sorangium cellulosum So ce56 and Identification of CYP109D1 as a New Fatty Acid Hydroxylase. *Chem Biol* **2010,** *17* (12), 1295-1305.

119. Rua, F.; Sadeghi, S. J.; Castrignano, S.; Di Nardo, G.; Gilardi, G., Engineering Macaca fascicularis cytochrome P450 2C20 to reduce animal testing for new drugs. *J Inorg Biochem* **2012,** *117*, 277-284.

120. Scheps, D.; Malca, S. H.; Hoffmann, H.; Nestl, B. M.; Hauer, B., Regioselective omega-hydroxylation of medium-chain n-alkanes and primary alcohols by CYP153 enzymes from Mycobacterium marinum and Polaromonas sp strain JS666. *Org Biomol Chem* **2011,** *9* (19), 6727-6733.

121. Neeli, R.; Roitel, O.; Scrutton, N. S.; Munro, A. W., Switching pyridine nucleotide specificity in P450BM3 - Mechanistic analysis of the W1046H AND W1046A enzymes. *J Biol Chem* **2005,** *280* (18), 17634-17644.

122. Joo, H.; Lin, Z. L.; Arnold, F. H., Laboratory evolution of peroxide-mediated cytochrome P450 hydroxylation. *Nature* **1999,** *399* (6737), 670-673.

123. Bormann, S.; Baraibar, A. G.; Ni, Y.; Holtmann, D.; Hollmann, F., Specific oxyfunctionalisations catalysed by peroxygenases: opportunities, challenges and solutions. *Catal Sci Technol* **2015,** *5* (4), 2038-2052.

124. Wang, Y. H.; Lan, D. M.; Durrani, R.; Hollmann, F., Peroxygenases en route to becoming dream catalysts. What are the opportunities and challenges? *Curr Opin Chem Biol* **2017,** *37*, 1-9.

125. Omura, T., Heme-thiolate proteins. *Biochem Biophys Res Commun* **2005,** *338* (1), 404-409.

126. Hofrichter, M.; Ullrich, R., Oxidations catalyzed by fungal peroxygenases. *Curr Opin Chem Biol* **2014,** *19*, 116-125.

127. Morris, D. R.; Hager, L. P., Chloroperoxidase .I. Isolation and Properties of Crystalline Glycoprotein. *J Biol Chem* **1966,** *241* (8), 1763-&.

128. Ulrich, R.; Nuske, J.; Scheibner, K.; Spantzel, J.; Hofrichter, M., Novel haloperoxidase from the agaric basidiomycete Agrocybe aegerita oxidizes aryl alcohols and aldehydes. *Appl Environ Microbiol* **2004,** *70* (8), 4575-4581.

129. Anh, D. H.; Ullrich, R.; Benndorf, D.; Svatos, A.; Muck, A.; Hofrichter, M., The coprophilous mushroom Coprinus radians secretes a haloperoxidase that catalyzes aromatic peroxygenation. *Appl Environ Microbiol* **2007,** *73* (17), 5477-5485.

130. Grobe, G.; Ullrich, R.; Pecyna, M. J.; Kapturska, D.; Friedrich, S.; Hofrichter, M.; Scheibner, K., High-yield production of aromatic peroxygenase by the agaric fungus Marasmius rotula. *Amb Express* **2011,** *1*.

131. Luthra, A.; Denisov, I. G.; Sligar, S. G., Spectroscopic features of cytochrome P450 reaction intermediates. *Arch Biochem Biophys* **2011,** *507* (1), 26-35.

132. Pecyna, M. J.; Ullrich, R.; Bittner, B.; Clemens, A.; Scheibner, K.; Schubert, R.; Hofrichter, M., Molecular characterization of aromatic peroxygenase from Agrocybe aegerita. *Appl Microbiol Biotechnol* **2009,** *84* (5), 885-897.

133. Piontek, K.; Strittmatter, E.; Ullrich, R.; Grobe, G.; Pecyna, M. J.; Kluge, M.; Scheibner, K.; Hofrichter, M.; Plattner, D. A., Structural Basis of Substrate Conversion in a New Aromatic Peroxygenase CYTOCHROME P450 FUNCTIONALITY WITH BENEFITS. *J Biol Chem* **2013,** *288* (48), 34767-34776.

134.   Hofrichter, M.; Ullrich, R.; Pecyna, M. J.; Liers, C.; Lundell, T., New and classic families of secreted fungal heme peroxidases. *Appl Microbiol Biotechnol* **2010,** *87* (3), 871-897.

135.   Hofrichter, M.; Ullrich, R., New Trends in Fungal Biooxidation. In *Industrial Applications*, Hofrichter, M., Ed. Springer Berlin Heidelberg: Berlin, Heidelberg, 2011; pp 425-449.

136.   Gutierrez, A.; Babot, E. D.; Ullrich, R.; Hofrichter, M.; Martinez, A. T.; del Rio, J. C., Regioselective oxygenation of fatty acids, fatty alcohols and other aliphatic compounds by a basidiomycete heme-thiolate peroxidase. *Arch Biochem Biophys* **2011,** *514* (1-2), 33-43.

137.   Peter, S.; Kinne, M.; Wang, X. S.; Ullrich, R.; Kayser, G.; Groves, J. T.; Hofrichter, M., Selective hydroxylation of alkanes by an extracellular fungal peroxygenase. *FEBS J* **2011,** *278* (19), 3667-3675.

138.   Girhard, M.; Kunigk, E.; Tihovsky, S.; Shumyantseva, V. V.; Urlacher, V. B., Light-driven biocatalysis with cytochrome P450 peroxygenases. *Biotechnol Appl Bioc* **2013,** *60* (1), 111-118.

139.   Paul, C. E.; Churakova, E.; Maurits, E.; Girhard, M.; Urlacher, V. B.; Hollmann, F., In situ formation of H2O2 for P450 peroxygenases. *Bioorgan Med Chem* **2014,** *22* (20), 5692-5696.

140.   Lucas, F.; Babot, E. D.; Canellas, M.; del Rio, J. C.; Kalum, L.; Ullrich, R.; Hofrichter, M.; Guallar, V.; Martinez, A. T.; Gutierrez, A., Molecular determinants for selective C-25-hydroxylation of vitamins D-2 and D-3 by fungal peroxygenases. *Catal Sci Technol* **2016,** *6* (1), 288-295.

141.   Peter, S.; Karich, A.; Ullrich, R.; Grobe, G.; Scheibner, K.; Hofrichter, M., Enzymatic one-pot conversion of cyclohexane into cyclohexanone: Comparison of four fungal peroxygenases. *J Mol Catal B-Enzym* **2014,** *103*, 47-51.

142.   Molina-Espeja, P.; Canellas, M.; Plou, F. J.; Hofrichter, M.; Lucas, F.; Guallar, V.; Alcalde, M., Synthesis of 1-Naphthol by a Natural Peroxygenase Engineered by Directed Evolution. *ChemBioChem* **2016,** *17* (4), 341-349.

143.   Kluge, M.; Ullrich, R.; Scheibner, K.; Hofrichter, M., Stereoselective benzylic hydroxylation of alkylbenzenes and epoxidation of styrene derivatives catalyzed by the peroxygenase of Agrocybe aegerita. *Green Chem* **2012,** *14* (2), 440-446.

144.   Kinne, M.; Poraj-Kobielska, M.; Aranda, E.; Ullrich, R.; Hammel, K. E.; Scheibner, K.; Hofrichter, M., Regioselective preparation of 5-hydroxypropranolol and 4 '-hydroxydiclofenac with a fungal peroxygenase. *Bioorg Med Chem Lett* **2009,** *19* (11), 3085-3087.

145.   Kinne, M.; Ullrich, R.; Hammel, K. E.; Scheibner, K.; Hofrichter, M., Regioselective preparation of (R)-2-(4-hydroxyphenoxy)propionic acid with a fungal peroxygenase. *Tetrahedron Lett* **2008,** *49* (41), 5950-5953.

146.   Kluge, M.; Ullrich, R.; Dolge, C.; Scheibner, K.; Hofrichter, M., Hydroxylation of naphthalene by aromatic peroxygenase from Agrocybe aegerita proceeds via oxygen transfer from H2O2 and intermediary epoxidation. *Applied Microbiology and Biotechnology* **2009,** *81* (6), 1071-1076.

147.   Ullrich, R.; Hofrichter, M., Enzymatic hydroxylation of aromatic compounds. *Cell Mol Life Sci* **2007,** *64* (3), 271-293.

148.   Hollmann, F.; Arends, I. W. C. E., Enzyme Initiated Radical Polymerizations. *Polymers-Basel* **2012,** *4* (1), 759-793.

149.   Molina-Espeja, P.; Garcia-Ruiz, E.; Gonzalez-Perez, D.; Ullrich, R.; Hofrichter, M.; Alcalde, M., Directed Evolution of Unspecific Peroxygenase from Agrocybe aegerita. *Appl Environ Microbiol* **2014,** *80* (11), 3496-3507.

150.	Molina-Espeja, P.; Ma, S.; Mate, D. M.; Ludwig, R.; Alcalde, M., Tandem-yeast expression system for engineering and producing unspecific peroxygenase. *Enzyme Microb Technol* **2015,** *73-74*, 29-33.

151.	Garcia-Ruiz, E.; Gonzalez-Perez, D.; Ruiz-Duenas, F. J.; Martinez, A. T.; Alcalde, M., Directed evolution of a temperature-, peroxide- and alkaline pH-tolerant versatile peroxidase. *Biochem J* **2012,** *441*, 487-498.

152.	Bankar, S. B.; Bule, M. V.; Singhal, R. S.; Ananthanarayan, L., Glucose oxidase — An overview. *Biotechnol Adv* **2009,** *27* (4), 489-501.

153.	Lutz, S.; Steckhan, E.; Liese, A., First asymmetric electroenzymatic oxidation catalyzed by a peroxidase. *Electrochem Commun* **2004,** *6* (6), 583-587.

154.	Perez, D. I.; Grau, M. M.; Arends, I. W. C. E.; Hollmann, F., Visible light-driven and chloroperoxidase-catalyzed oxygenation reactions. *ChemComm* **2009,** (44), 6848-6850.

155.	Ni, Y.; Fernández-Fueyo, E.; Baraibar, A. G.; Ullrich, R.; Hofrichter, M.; Yanase, H.; Alcalde, M.; van Berkel, W. J. H.; Hollmann, F., Peroxygenase-Catalyzed Oxyfunctionalization Reactions Promoted by the Complete Oxidation of Methanol. *Angew Chem Int Ed* **2016,** *55* (2), 798-801.

156.	Ni, Y.; Holtmann, D.; Hollmann, F., How Green is Biocatalysis? To Calculate is To Know. *ChemCatChem* **2014,** *6* (4), 930-943.

157.	Churakova, E.; Arends, I. W. C. E.; Hollmann, F., Increasing the Productivity of Peroxidase-Catalyzed Oxyfunctionalization: A Case Study on the Potential of Two-Liquid-Phase Systems. *ChemCatChem* **2013,** *5* (2), 565-568.

158.	Eppink, M. H. M.; Berkel, W. J. H. V.; Schreuder, H. A., Identification of a novel conserved sequence motif in flavoprotein hydroxylases with a putative dual function in FAD/NAD(P)H binding. *Protein Sci* **1997,** *6* (11), 2454-2458.

159.	Wierenga, R. K.; Dejong, R. J.; Kalk, K. H.; Hol, W. G. J.; Drenth, J., Crystal-Structure of Para-Hydroxybenzoate Hydroxylase. *J Mol Biol* **1979,** *131* (1), 55-73.

160.	Entsch, B.; Vanberkel, W. J. H., Flavoprotein Structure and Mechanism .1. Structure and Mechanism of Para-Hydroxybenzoate Hydroxylase. *Faseb J* **1995,** *9* (7), 476-483.

161.	Montersino, S.; Tischler, D.; Gassner, G. T.; van Berkel, W. J. H., Catalytic and Structural Features of Flavoprotein Hydroxylases and Epoxidases. *Adv Synth Catal* **2011,** *353* (13), 2301-2319.

162.	Palfey, B. A.; McDonald, C. A., Control of catalysis in flavin-dependent monooxygenases. *Arch Biochem Biophys* **2010,** *493* (1), 26-36.

163.	Montersino, S.; Orru, R.; Barendregt, A.; Westphal, A. H.; van Duijn, E.; Mattevi, A.; van Berkel, W. J. H., Crystal Structure of 3-Hydroxybenzoate 6-Hydroxylase Uncovers Lipid-assisted Flavoprotein Strategy for Regioselective Aromatic Hydroxylation. *J Biol Chem* **2013,** *288* (36), 26235-26245.

164.	Kohler, H. P. E.; Kohlerstaub, D.; Focht, D. D., Degradation of 2-Hydroxybiphenyl and 2,2'-Dihydroxybiphenyl by Pseudomonas Sp Strain-Hbp1. *Appl Environ Microbiol* **1988,** *54* (11), 2683-2688.

165.	Kanteev, M.; Bregman-Cohen, A.; Deri, B.; Shahar, A.; Adir, N.; Fishman, A., A crystal structure of 2-hydroxybiphenyl 3-monooxygenase with bound substrate provides insights into the enzymatic mechanism. *Bba-Proteins Proteom* **2015,** *1854* (12), 1906-1913.

166. Schmid, A.; Kohler, H. P. E.; Engesser, K. H., E-coli JM109 pHBP461, a recombinant biocatalyst for the regioselective monohydroxylation of ortho-substituted phenols to their corresponding 3-substituted catechols. *J Mol Catal B-Enzym* **1998,** *5* (1-4), 311-316.

167. Held, M.; Suske, W.; Schmid, A.; Engesser, K. H.; Kohler, H. P. E.; Witholt, B.; Wubbolts, M. G., Preparative scale production of 3-substituted catechols using a novel monooxygenase from Pseudomonas azelaica HBP 1. *J Mol Catal B-Enzym* **1998,** *5* (1-4), 87-93.

168. Meyer, A.; Held, M.; Schmid, A.; Kohler, H. P. E.; Witholt, B., Synthesis of 3-tert-butylcatechol by an engineered monooxygenase. *Biotechnol Bioeng* **2003,** *81* (5), 518-524.

169. Jensen, C. N.; Mielke, T.; Farrugia, J. E.; Frank, A.; Man, H.; Hart, S.; Turkenburg, J. P.; Grogan, G., Structures of the Apo and FAD-Bound Forms of 2-Hydroxybiphenyl 3-monooxygenase (HbpA) Locate Activity Hotspots Identified by Using Directed Evolution. *ChemBioChem* **2015,** *16* (6), 968-976.

170. Hausinger, R. P., Fe(II)/alpha-ketoglutarate-dependent hydroxylases and related enzymes. *Crit Rev Biochem Mol Biol* **2004,** *39* (1), 21-68.

171. Wu, L. F.; Meng, S.; Tang, G. L., Ferrous iron and alpha-ketoglutarate-dependent dioxygenases in the biosynthesis of microbial natural products. *Bba-Proteins Proteom* **2016,** *1864* (5), 453-470.

172. Schofield, C. J.; Zhang, Z. H., Structural and mechanistic studies on 2-oxoglutarate-dependent oxygenases and related enzymes. *Curr Opin Struc Biol* **1999,** *9* (6), 722-731.

173. Prescott, A. G., A Dilemma of Dioxygenases (or Where Biochemistry and Molecular-Biology Fail to Meet). *J Exp Bot* **1993,** *44* (262), 849-861.

174. Prescott, A. G.; John, P., Dioxygenases: Molecular structure and role in plant metabolism. *Annu Rev Plant Phys* **1996,** *47*, 245-271.

175. Clifton, I. J.; McDonough, M. A.; Ehrismann, D.; Kershaw, N. J.; Granatino, N.; Schofield, C. J., Structural studies on 2-oxoglutarate oxygenases and related double-stranded beta-helix fold proteins. *J Inorg Biochem* **2006,** *100* (4), 644-669.

176. Httel, W., Biocatalytic Production of Chemical Building Blocks in Technical Scale with -Ketoglutarate-Dependent Dioxygenases. *Chem Ing Tech* **2013,** *85* (6), 809-817.

177. Hutton, J. J.; Tappel, A. L.; Udenfriend, S., Requirements for Alpha-Ketoglutarate Ferrous Ion and Ascorbate by Collagen Proline Hydroxylase. *Biochem Biophys Res Commun* **1966,** *24* (2), 179-+.

178. Kivirikko, K. I.; Pihlajaniemi, T., Collagen hydroxylases and the protein disulfide isomerase subunit of prolyl 4-hydroxylases. *Adv Enzymol Relat Subj Biochem* **1998,** *72*, 325-+.

179. Yang, H.; Song, K.; Xue, T. A.; Xue, X. P.; Huyan, T.; Wang, W.; Wang, H., The distribution and expression profiles of human Aspartyl/Asparaginyl beta-hydroxylase in tumor cell lines and human tissues. *Oncol Rep* **2010,** *24* (5), 1257-1264.

180. Hewitson, K. S.; Granatino, N.; Welford, R. W. D.; McDonough, M. A.; Schofield, C. J., Oxidation by 2-oxoglutarate oxygenases: non-haem iron systems in catalysis and signalling. *Philos T Royal Soc A* **2005,** *363* (1829), 807-828.

181. Semenza, G. L., HIF-1 and human disease: one highly involved factor. *Gene Dev* **2000,** *14* (16), 1983-1991.

182.   Hewitson, K. S.; McNeill, L. A.; Schofield, C. J., Modulating the hypoxia-inducible factor signaling pathway: Applications from cardiovascular disease to cancer. *Curr Pharm Design* **2004,** *10* (8), 821-833.

183.   Schofield, C. J.; Ratcliffe, P. J., Oxygen sensing by HIF hydroxylases. *Nat Rev Mol Cell Bio* **2004,** *5* (5), 343-354.

184.   Fedeles, B. I.; Singh, V.; Delaney, J. C.; Li, D. Y.; Essigmann, J. M., The AlkB Family of Fe(II)/alpha-Ketoglutarate-dependent Dioxygenases: Repairing Nucleic Acid Alkylation Damage and Beyond. *J Biol Chem* **2015,** *290* (34), 20734-20742.

185.   Lloyd, M. D.; Merritt, K. D.; Lee, V.; Sewell, T. J.; Wha-Son, B.; Baldwin, J. E.; Schofield, C. J.; Elson, S. W.; Baggaley, K. H.; Nicholson, N. H., Product-substrate engineering by bacteria: Studies on clavaminate synthase, a trifunctional dioxygenase. *Tetrahedron* **1999,** *55* (33), 10201-10220.

186.   Prescott, A. G.; Lloyd, M. D., The iron(II) and 2-oxoacid-dependent dioxygenases and their role in metabolism. *Nat Prod Rep* **2000,** *17* (4), 367-383.

187.   Vaz, F. M.; Wanders, R. J. A., Carnitine biosynthesis in mammals. *Biochem J* **2002,** *361*, 417-429.

188.   Gibbons, H. S.; Lin, S.; Cotter, R. J.; Raetz, C. R. H., Oxygen requirement for the biosynthesis of the S-2-hydroxymyristate moiety in Salmonella typhimurium lipid A - Function of LpxO, a new Fe2+/alpha-ketoglutarate-dependent dioxygenase homologue. *J Biol Chem* **2000,** *275* (42), 32940-32949.

189.   Eichhorn, E.; vanderPloeg, J. R.; Kertesz, M. A.; Leisinger, T., Characterization of alpha-ketoglutarate-dependent taurine dioxygenase from Escherichia coli. *J Biol Chem* **1997,** *272* (37), 23031-23036.

190.   McDonough, M. A.; Loenarz, C.; Chowdhury, R.; Clifton, I. J.; Schofield, C. J., Structural studies on human 2-oxoglutarate dependent oxygenases. *Curr Opin Struc Biol* **2010,** *20* (6), 659-672.

191.   Purpero, V.; Moran, G. R., The diverse and pervasive chemistries of the alpha-keto acid dependent enzymes. *J Biol Inorg Chem* **2007,** *12* (5), 587-601.

192.   Jia, B.; Jia, X.; Kim, K. H.; Jeon, C. O., Integrative view of 2-oxoglutarate/Fe(II)-dependent oxygenase diversity and functions in bacteria. *BBA - General Subjects* **2017,** *1861* (2), 323-334.

193.   Ivanov, K.; Stoimenova, A.; Obreshkova, D.; Saso, L., Biotechnology in the Production of Pharmaceutical Industry Ingredients: Amino Acids. *Biotechnol Biotechnol Equip* **2013,** *27* (2), 3620-3626.

194.   Klein, C.; Httel, W., A Simple Procedure for Selective Hydroxylation of L-Proline and L-Pipecolic Acid with Recombinantly Expressed Proline Hydroxylases. *Adv Synth Catal* **2011,** *353* (8), 1375-1383.

195.   Remuzon, P., Trans-4-hydroxy-L-proline, a useful and versatile chiral starting block. *Tetrahedron* **1996,** *52* (44), 13803-13835.

196.   Mori, H.; Shibasaki, T.; Uozaki, Y.; Ochiai, K.; Ozaki, A., Detection of novel proline 3-hydroxylase activities in Streptomyces and Bacillus spp by regio- and stereospecific hydroxylation of L-proline. *Appl Environ Microbiol* **1996,** *62* (6), 1903-1907.

197.   Hara, R.; Uchiumi, N.; Kino, K., Identification and characterization of 2-oxoglutarate-dependent dioxygenases catalyzing selective cis-hydroxylation of proline and pipecolinic acid from actinomycetes. *J Biotechnol* **2014,** *172*, 55-58.

198.   Takano, S.; Iwabuchi, Y.; Ogasawara, K., Concise Stereoselective Synthesis of (2s,4r)-4-Hydroxyproline from (S)-O-Benzylglycidol by a Novel Cyclization. *J Chem Soc Chem Comm* **1988,** (23), 1527-1528.

199.   Gorres, K. L.; Raines, R. T., Prolyl 4-hydroxylase. *Criti Rev Biochem Mol Biol* **2010,** *45* (2), 106-124.

200.   Shibasaki, T.; Mori, H.; Chiba, S.; Ozaki, A., Microbial proline 4-hydroxylase screening and feme cloning. *Appl Environ Microbiol* **1999,** *65* (9), 4028-4031.

201.   Mori, H.; Shibasaki, T.; Yano, K.; Ozaki, A., Purification and cloning of a proline 3-hydroxylase, a novel enzyme which hydroxylates free L-proline to cis-3-hydroxy-L-proline. *J Bacteriol* **1997,** *179* (18), 5677-5683.

202.   Shibasaki, T.; Mori, H.; Ozaki, A., Enzymatic production of trans-4-hydroxy-L-proline by regio- and stereospecific hydroxylation of L-proline. *Biosci Biotechnol Biochem* **2000,** *64* (4), 746-750.

203.   Hara, R.; Kino, K., Characterization of novel 2-oxoglutarate dependent dioxygenases converting L-proline to cis-4-hydroxy-L-proline. *Biochem Biophys Res Commun* **2009,** *379* (4), 882-886.

204.   Clifton, I. J.; Hsueh, L. C.; Baldwin, J. E.; Harlos, K.; Schofield, C. J., Structure of proline 3-hydroxylase - Evolution of the family of 2-oxoglutarate dependent oxygenases. *Eur J Biochem* **2001,** *268* (24), 6625-6636.

205.   Koketsu, K.; Shomura, Y.; Moriwaki, K.; Hayashi, M.; Mitsuhashi, S.; Hara, R.; Kino, K.; Higuchi, Y., Refined Regio- and Stereoselective Hydroxylation of l-Pipecolic Acid by Protein Engineering of l-Proline cis-4-Hydroxylase Based on the X-ray Crystal Structure. *ACS Synth Biol* **2014**.

206.   Hara, R.; Uchiumi, N.; Okamoto, N.; Kino, K., Regio- and stereoselective oxygenation of proline derivatives by using microbial 2-oxoglutarate-dependent dioxygenases. *Biosci Biotechnol Biochem* **2014,** *78* (8), 1384-8.

207.   Ostrowski, J.; Kuhns, J. E.; Lupisella, J. A.; Manfredi, M. C.; Beehler, B. C.; Krystek, S. R.; Bi, Y. Z.; Sun, C. Q.; Seethala, R.; Golla, R.; Sleph, P. G.; Fura, A.; An, Y. M.; Kish, K. F.; Sack, J. S.; Mookhtiar, K. A.; Grover, G. J.; Hamann, L. G., Pharmacological and x-ray structural characterization of a novel selective androgen receptor modulator: Potent hyperanabolic stimulation of skeletal muscle with hypostimulation of prostate in rats. *Endocrinology* **2007,** *148* (1), 4-12.

208.   Wilson, E. M., Muscle-bound? A tissue-selective nonsteroidal androgen receptor modulator. *Endocrinology* **2007,** *148* (1), 1-3.

209.   Johnston, R. M.; Chu, L. N.; Liu, M.; Goldberg, S. L.; Goswami, A.; Patel, R. N., Hydroxylation of l-proline to cis-3-hydroxy-l-proline by recombinant Escherichia coli expressing a synthetic l-proline-3-hydroxylase gene. *Enzyme Microb Technol* **2009,** *45, Numb 6-7*, 484-490.

210.   Korbie, D. J.; Mattick, J. S., Touchdown PCR for increased specificity and sensitivity in PCR amplification. *Nat Protoc* **2008,** *3* (9), 1452-1456.

211.   Bhushan, R.; Bruckner, H., Marfey's reagent for chiral amino acid analysis: A review. *Amino Acids* **2004,** *27* (3-4), 231-247.

212.   Csapó, J.; Albert, C.; Lóki, K.; Csapó-Kiss, Z., Separation and determination of the amino acids by ion exchange column chromatography applying postcolumn derivatization. *Acta Universit Sapient - Aliment* **2008,** *1*, 5-29.

213.   Moore, S.; Stein, W. H., Photometric Ninhydrin Method for Use in the Chromatography of Amino Acids. *J Biol Chem* **1948,** *176* (1), 367-388.

214.   Spackman, D. H.; Stein, W. H.; Moore, S., Automatic Recording Apparatus for Use in the Chromatography of Amino Acids. *Anal Chem* **1958,** *30* (7), 1190-1206.

215.   Cohen, S. A.; Michaud, D. P., Synthesis of a Fluorescent Derivatizing Reagent, 6-Aminoquinolyl-N-Hydroxysuccinimidyl Carbamate, and Its Application for the Analysis of Hydrolysate Amino-Acids Via High-Performance Liquid-Chromatography. *Anal Biochem* **1993,** *211* (2), 279-287.

216.   Kochhar, S.; Mouratou, B.; Christen, P., Amino Acid Analysis by Precolumn Derivatization with 1-Fluoro-2,4-Dinitrophenyl-5-L-Alanine Amide (Marfey's Reagent). In *The Protein Protocols Handbook*, Walker, J. M., Ed. Humana Press: Totowa, NJ, 2002; pp 567-572.

217.   Bank, R. A.; Jansen, E. J.; Beekman, B.; Koppele, J. M. T., Amino acid analysis by reverse-phase high-performance liquid chromatography: Improved derivatization and detection conditions with 9-fluorenylmethyl chloroformate. *Anal Biochem* **1996,** *240* (2), 167-176.

218.   Petersen, L.; Olewinski, R.; Salmon, P.; Connors, N., Novel proline hydroxylase activities in the pneumocandin-producing fungus Glarea lozoyensis responsible for the formation of trans 3- and trans 4-hydroxyproline. *Appl Microbiol Biotechnol* **2003,** *62* (2-3), 263-267.

219.   Lindblad, W. J.; Diegelmann, R. F., Quantitation of Hydroxyproline Isomers in Acid Hydrolysates by High-Performance Liquid-Chromatography. *Anal Biochem* **1984,** *138* (2), 390-395.

220.   Marfey, P.; Ottesen, M., Determination of D-Amino Acids .1. Hydrolysis of Dnp-L-Amino Acid Methyl-Esters with Carboxypeptidase-Y. *Carlsberg Res Commun* **1984,** *49* (6), 585-590.

221.   Marfey, P., Determination of D-Amino Acids .2. Use of a Bifunctional Reagent, 1,5-Difluoro-2,4-Dinitrobenzene. *Carlsberg Res Commun* **1984,** *49* (6), 591-596.

222.   Bhushan, R.; Bruckner, H., Use of Marfey's reagent and analogs for chiral amino acid analysis : Assessment and applications to natural products and biological systems. *J Chromatogr B* **2011,** *879* (29), 3148-3161.

223.   Kochhar, S.; Christen, P., Amino-Acid Analysis by High-Performance Liquid-Chromatography after Derivatization with 1-Fluoro-2,4-Dinitrophenyl-5-L-Alanine Amide. *Anal Biochem* **1989,** *178* (1), 17-21.

224.   Ozaki, A.; Shibasaki, T.; Mori, H., Specific Proline and Hydroxyproline Detection Method by Postcolumn Derivatization for High-Performance Liquid-Chromatography. *Biosci Biotechnol Biochem* **1995,** *59* (9), 1764-1765.

225.   Einarsson, S.; Folestad, S.; Josefsson, B.; Lagerkvist, S., High-resolution reversed-phase liquid chromatography system for the analysis of complex solutions of primary and secondary amino acids. *Anal Chem* **1986,** *58* (8), 1638-1643.

226.   Einarsson, S.; Josefsson, B.; Lagerkvist, S., Determination of amino acids with 9-fluorenylmethyl chloroformate and reversed-phase high-performance liquid chromatography. *J Chromatogr* **1983,** *282*, 609-618.

227.   Dai, Z. L.; Wu, Z. L.; Wang, J. J.; Wang, X. Q.; Jia, S. C.; Bazer, F. W.; Wu, G. Y., Analysis of polyamines in biological samples by HPLC involving pre-column derivatization with o-phthalaldehyde and N-acetyl-L-cysteine. *Amino Acids* **2014,** *46* (6), 1557-1564.

228.   Sleight, S. C.; Bartley, B. A.; Lieviant, J. A.; Sauro, H. M., In-Fusion BioBrick assembly and re-engineering. *Nucleic Acids Res* **2010,** *38* (8), 2624-2636.

229.   Zhu, B. G.; Cai, G. F.; Hall, E. O.; Freeman, G. J., In-Fusion (TM) assembly: seamless engineering of multidomain fusion proteins, modular vectors, and mutations. *Biotechniques* **2007,** *43* (3), 356-359.

230.   Cheung, R. C. F.; Wong, J. H.; Ng, T. B., Immobilized metal ion affinity chromatography: a review on its applications. *Appl Microbiol Biotechnol* **2012,** *96* (6), 1411-1420.

231.   Vagenende, V.; Yap, M. G. S.; Trout, B. L., Mechanisms of Protein Stabilization and Prevention of Protein Aggregation by Glycerol. *Biochemistry* **2009,** *48* (46), 11084-11096.

232.   Walter, T. S.; Meier, C.; Assenberg, R.; Au, K. F.; Ren, J. S.; Verma, A.; Nettleship, J. E.; Owens, R. J.; Stuart, D. I.; Grimes, J. M., Lysine methylation as a routine rescue strategy for protein crystallization. *Structure* **2006,** *14* (11), 1617-1622.

233.   Derewenda, Z. S., Rational protein crystallization by mutational surface engineering. *Structure* **2004,** *12* (4), 529-535.

234.   Goldschmidt, L.; Cooper, D. R.; Derewenda, Z. S.; Eisenberg, D., Toward rational protein crystallization: A Web server for the design of crystallizable protein variants. *Protein Sci* **2007,** *16* (8), 1569-1576.

235.   Mcpherson, A.; Koszelak, S.; Axelrod, H.; Day, J.; Robinson, L.; Mcgrath, M.; Williams, R.; Cascio, D., The Effects of Neutral Detergents on the Crystallization of Soluble-Proteins. *J Cryst Growth* **1986,** *76* (3), 547-553.

236.   Deller, M. C.; Kong, L.; Rupp, B., Protein stability: a crystallographer's perspective. *Acta Crystallograph Sect F Struct Biol Commun* **2016,** *72* (Pt 2), 72-95.

237.   Hassell, A. M.; An, G.; Bledsoe, R. K.; Bynum, J. M.; Carter, H. L., III; Deng, S.-J. J.; Gampe, R. T.; Grisard, T. E.; Madauss, K. P.; Nolte, R. T.; Rocque, W. J.; Wang, L.; Weaver, K. L.; Williams, S. P.; Wisely, G. B.; Xu, R.; Shewchuk, L. M., Crystallization of protein-ligand complexes. *Acta Crystallograph Sect D* **2007,** *63* (1), 72-79.

238.   Hibi, M.; Mori, R.; Miyake, R.; Kawabata, H.; Kozono, S.; Takahashi, S.; Ogawa, J., Novel Enzyme Family Found in Filamentous Fungi Catalyzing trans-4-Hydroxylation of L-Pipecolic Acid. *Appl Environ Microbiol* **2016,** *82* (7), 2070-2077.

239.   Sato, T.; Hirayama, F.; Saito, T.; Kaniwa, H., A New Alkaloid Antibiotic Tetrazomine Structure Determination. *J Antibiot* **1991,** *44* (12), 1367-1370.

240.   Son, S.; Ko, S.-K.; Jang, M.; Lee, J. K.; Ryoo, I.-J.; Lee, J.-S.; Lee, K. H.; Soung, N.-K.; Oh, H.; Hong, Y.-S.; Kim, B. Y.; Jang, J.-H.; Ahn, J. S., Ulleungamides A and B, Modified α,β-Dehydropipecolic Acid Containing Cyclic Depsipeptides from Streptomyces sp. KCB13F003. *Org Lett* **2015,** *17* (16), 4046-4049.

241.   Aiello, A.; Fattorusso, E.; Giordano, A.; Menna, M.; Müller, W. E. G.; Perović-Ottstadt, S.; Schröder, H. C., Damipipecolin and damituricin, novel bioactive bromopyrrole alkaloids from the Mediterranean sponge Axinella damicornis. *Bioorgan Med Chem* **2007,** *15* (17), 5877-5887.

242.   Falcioni, F.; Blank, L. M.; Frick, O.; Karau, A.; Buhler, B.; Schmid, A., Proline Availability Regulates Proline-4-Hydroxylase Synthesis and Substrate Uptake in Proline-Hydroxylating Recombinant Escherichia coli. *Appl Environ Microbiol* **2013,** *79* (9), 3091-3100.

243.   Klockner, W.; Buchs, J., Advances in shaking technologies. *Trends Biotechnol* **2012,** *30* (6), 307-314.

244.   Thomas, C. A.; Talaty, E. R.; Bann, J. G., 3S-Fluoroproline as a probe to monitor proline isomerization during protein folding by 19F-NMR. *ChemComm* **2009,** (23), 3366-3368.

245.   Mollica, A.; Paradisi, M. P.; Varani, K.; Spisani, S.; Lucente, G., Chemotactic peptides: fMLF-OMe analogues incorporating proline–methionine chimeras as N-terminal residue. *Bioorgan Med Chem* **2006,** *14* (7), 2253-2265.

246.   Biel, M.; Deck, P.; Giannis, A.; Waldmann, H., Synthesis and Evaluation of Acyl Protein Thioesterase 1 (APT1) Inhibitors. *Chem Eur J* **2006,** *12* (15), 4121-4143.

247.   Betts, J. I.; Baganz, F., Miniature bioreactors: current practices and future opportunities. *Microb Cell Fact* **2006,** *5* (1), 21.

248.   Cavasotto, C. N.; Phatak, S. S., Homology modeling in drug discovery: current trends and applications. *Drug Discov Today* **2009,** *14* (13-14), 676-683.

249.   Schmidt, M. F.; Korb, O.; Howard, N. I.; Dias, M. V. B.; Blundell, T. L.; Abell, C., Discovery of Schaeffer's Acid Analogues as Lead Structures of Mycobacterium tuberculosis Type II Dehydroquinase Using a Rational Drug Design Approach. *Chemmedchem* **2013,** *8* (1), 54-58.

250.   Fischer, D., Servers for protein structure prediction. *Curr Opin Struc Biol* **2006,** *16* (2), 178-182.

251.   Dalton, J. A. R.; Jackson, R. M., An evaluation of automated homology modelling methods at low target–template sequence similarity. *Bioinformatics* **2007,** *23* (15), 1901-1908.

252.   Martí-Renom, M. A.; Stuart, A. C.; Fiser, A.; Sánchez, R.; and, F. M.; Šali, A., Comparative Protein Structure Modeling of Genes and Genomes. *Annu Rev Biophys Biomol Struct* **2000,** *29* (1), 291-325.

253.   Kim, D. E.; Chivian, D.; Baker, D., Protein structure prediction and analysis using the Robetta server. *Nucleic Acids Res* **2004,** *32*, W526-W531.

254.   Reuter, K.; Pittelkow, M.; Bursy, J.; Heine, A.; Craan, T.; Bremer, E., Synthesis of 5-Hydroxyectoine from Ectoine: Crystal Structure of the Non-Heme Iron(II) and 2-Oxoglutarate-Dependent Dioxygenase EctD. *PloS One* **2010,** *5* (5).

255.   Höppner, A.; Widderich, N.; Lenders, M.; Bremer, E.; Smits, S. H. J., Crystal Structure of the Ectoine Hydroxylase, a Snapshot of the Active Site. *J Biol Chem* **2014,** *289* (43), 29570-29583.

256.   Chen, H.; Bong, Y. K.; Cabirol, F. L.; Prafulchandra, A. G.; Li, T.; Moore, J. C.; Quintanar-Audelo, M.; Hong, Y.; Collier, S. J.; Smith, D. Biocatalysts and methods for hydroxylation of chemical compounds. U.S. Patent 0118719 A1, Apr. 30, **2015**

257.   Rosano, G. L.; Ceccarelli, E. A., Recombinant protein expression in Escherichia coli: advances and challenges. *Front Microbiol* **2014,** *5*, 172.