Cytochrome P450 BM3 Engineering for the Production of Valuable Human Drug Metabolites

A thesis submitted to The University of Manchester for the degree of Master of Philosophy in the Faculty of Science and Engineering

Master of Philosophy

Sahara Bhanot

2020

School of Natural Sciences

Table of Contents

Word Count:

List of Tables

List of Figures

List of Equations

Abstract

P450 BM3 is an enzyme which exhibits some of the highest catalytic rates any P450, owing to the unique fusion structure. P450 BM3 has been subject to intrusive mutagenesis studies to expand its catalytic repertoire. BM3 WT preferentially binds and hydroxylates fatty acids, such as arachidonic acid.

A A82F/F87V DM was proposed….

P450 BM3 is highly relevant in the metabolism of novel and existing xenobiotics. The US Food and Drug Administration (FDA)…

Further work to..

Declaration

Copyright Statement

Acknowledgements

i) List of Abbreviations

1. Introduction

1.1 Introduction to Cytochromes P450

1.1.1 An Overview of Cytochromes P450 and their Roles

Cytochromes P450 (P450s) are an enzyme superfamily present in all domains of life and are responsible for major biotransformation involving both endogenous and exogenous substances.(Nelson, Goldstone and Stegeman, 2013). Named so for their characteristic UV- visible spectroscopic 450 nm peak upon CO binding, P450s have been thoroughly investigated for a vast number of biotechnological applications in recent years, namely drug metabolism studies and points of interest for directed evolution approaches. An attractive feature for evolutionary design is the diversity in amino acid sequences, yet retaining a conserved P450

structural fold.[2] The ubiquitous nature of P450s resides in their catalytic properties with relation to heme binding and a conserved cysteine ligand acting as an fifth axial ligand to the iron heme centre. P450s carry out several different types of reactions (Figure 1.1), and typically require molecular oxygen and 2 electrons, routinely delivered via redox partners from NAD(P)H (Equation 1).

$$R\text{-}H + O_2 + NAD(P)H + H^+ \rightarrow R\text{-}OH + H_2O + NAD(P)^+ \quad (1)$$

In the years following their discovery, the plethora of diverse roles and importance of the reactions catalysed by P450s became apparent. In humans, P450s are now known to catalyse over 95% of all known organic oxidation and reduction reactions, from the biosynthesis of life- dependant molecules such as hormones, to xenobiotic, steroid and carcinogen metabolism and toxicological regulation.[3,4]

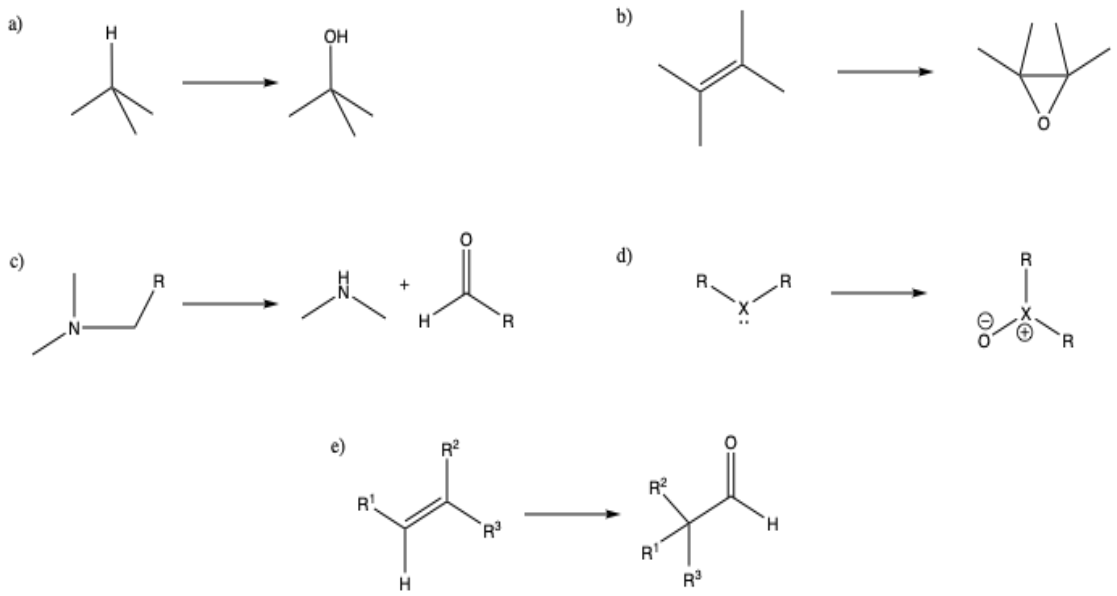


Figure 1.1 Overview outlining typical reactions carried out by P450s: a) carbon hydroxylation, b) epoxidation, c) heteroatom release, d) heteroatom oxygenation and e) group migration.

There are now over 340,000 fully identified and sequenced P450s from all major domains of life, over 41,000 P450 sequences have been assigned nomenclature and most of the remaining P450s have been sorted into clans, families and sub- families. (Nelson, 2018) (Finnigan *et al.*, 2020)

1.2 A Brief History of Cytochromes P450

1.2.1 The Discovery of Cytochromes P450

In 1949, Jim and Elizabeth (Betty) Miller reported that key enzymes in xenobiotic metabolism were present in rat liver homogenate, dubbed “the particulate fraction”. These conclusions were drawn from the metabolism of the aminoazo dye, *N*-*N*- dimethyl- 4- aminoazobenzene and its convent binding to cellular proteins upon incubation with nicotinamide adenine dinucleotide (NADPH) in its reduced form.[5] Though original investigations were based on factors in tumour formation, the Miller team described the role of NADPH in the metabolism of the aminoazo dye.[6] Subsequent investigations led to Mason hypothesising three broad enzyme classes involved in the metabolism of oxygen. Mason observed that one molecule of oxygen was consumed per molecule of substrate utilized, designating this role to a mixed- function oxidase.[7]

During the 1950s there was a surge of papers investigating xenobiotic metabolism by hepatic microsomes, with the addition of NADPH and under aerobic conditions.[8] With rapid technological developments such as the high speed centrifuge and novel analytical techniques, more in depth investigations into enzyme roles and classes revealed novel roles and functions. Klingenberg observed a characteristic optical absorption peak, λ_{max} at 450 nm upon carbon monoxide (CO) binding to a pigment in liver microsomes whilst investigating the reduction of cytochrome b_5 by nicotinamide adenine dinucleotide, in its reduced form (NADH).[9] After further investigation Klingenberg suggested that the observed excess

heme was due to something other than cytochrome b_5 alone. This CO-binding pigment was found to be present in the microsomes of other organisms which also gave the Soret peak at 450 nm.[10] These unknown enzymes were known thereafter as Cytochromes P450.

From these initial studies, the range in the catalysing ability of P450s and their dependence on redox partners and atmospheric oxygen became apparent as well as the diverse range of their substrates.

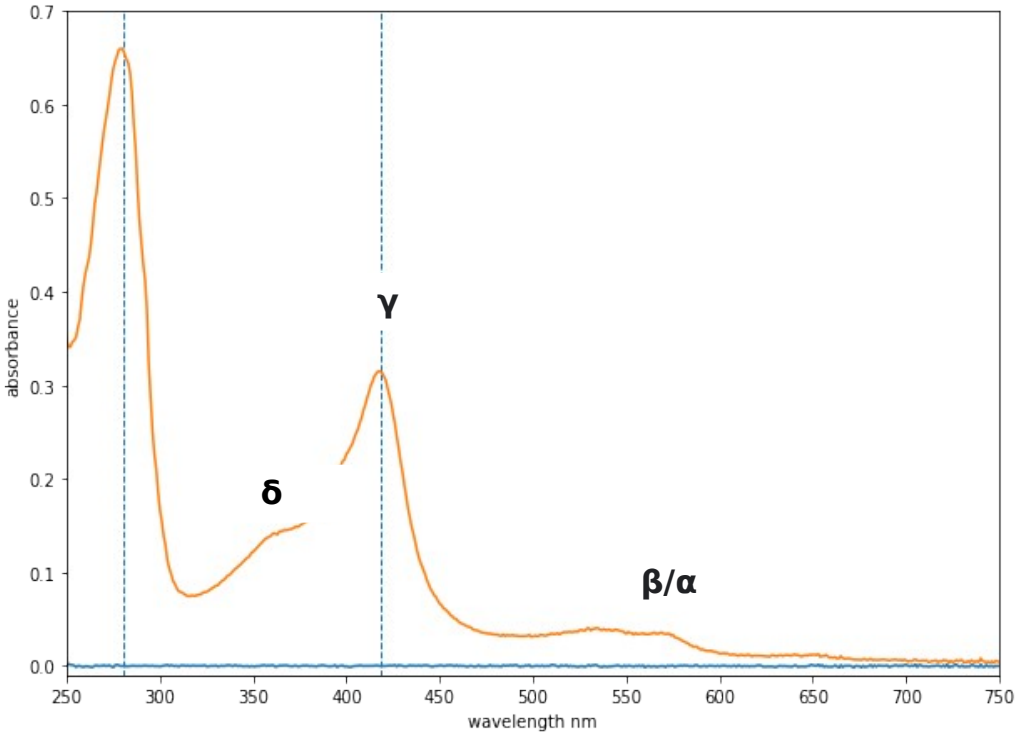


Figure 1.5 UV- Visible spectra, with P450 BM3 trace in orange and base line in blue, the dashed vertical lines are at 280 and 420 nm respectively and α/ β/ γ/ δ bands
are indicated on the trace. Figure generated in maplotlib.

The structure and role of P450s were not suggested until later, following breakthrough protein expression and purification techniques.[11] Omura and Sato carried out studies of C-21 hydroxylation of 17- hydroxy-

progesterone concluding that P450 was both a hemoprotein and a cytochrome.[12,13] Further development in the purification of P450s by Ichikawa and Yamano in 1967, found that glycerol was able to stabilize CYPs against detergents, allowing for more intrusive studies into their evolutionary past.[14]

1.1.2 Evolutionary History of the CYP superfamily

CYP enzymes are necessary for many of the ecological and physiological processes in the world and thus have been found in all kingdoms and domains of life.[1] All eukaryotes, plants, animals and fungi, possess P450s, as well as some prokaryotes. $P450_{cam}$, from the bacterium *Pseudomonas putida,* was the first successfully isolated P450 3D structure to be solved by X- ray crystallography, in both the ligand- free and ligand- bound states.[15] Though basic prokaryotic metabolism does not seem to rely on P450s, those that rely on terpenes and other organic compounds such as hydrocarbons, require forms of P450s, like the soluble, $P450_{cam}$ which is responsible for the metabolism of camphor. The structure of $P450_{cam}$ has be subject to intrusive binding and inhibition studies to give insight into key catalytic residues and conformational changes induced by various substrates.(Shiro *et al.*, 1989)[16]
Comparative studies of the amino acid sequences from P450s across all domains of life in the late 1970s showed analogous, highly conserved residues amongst all of the P450 hemoproteins, indicating divergence from a single ancestral form.[17]

It is widely thought that P450s may have originated to aid in the detoxification of oxygen in the early atmosphere, due to their affinity to bind free oxygen. As life evolved and diversified, the varying roles of P450s emerged. This has been demonstrated by phylogenetic analysis of the CYP superfamily.[18]

1.1.3 Classification of CYP superfamily

Confusion arose when naming the high volumes of P450s being sequenced. In 1987, a formal naming system was devised, based on similarity of amino acid sequences, to classify P450s as “clans”, “families” and “subfamilies”.[19] The conventional route of naming is as follows: The root symbol for cytochrome P450s is CYP, a number follows to denote the gene family, followed by a letter for the subfamily and lastly another number to characterise the particular gene. Some notable examples include CYP 101A1 and CYP 102A1 which represent $P450_{cam}$ and $P450_{BM3}$ respectively.[20]

2. Structure and Function of Cytochromes P450

2.1 Structure of CYPs

2.1.1 An Overview of P450 Redox Systems

In order for P450s to carry out monooxygenation reactions of their substrates, typically there must be a supply of molecular oxygen as well as electron carrying coenzymes NAD(P)H to shuttle two electrons.

Depending on the methods of electron shuttling to the catalytic site, all P450- containing systems can be sub-divided into four widely observed classes, I, II, III and IV. Class I P450 systems are soluble in prokaryotes (Ia) and are found within the mitochondrial membrane of eukaryotes (Ib). They require two electron donors, one from an flavin- adenine dinucleotide (FAD) containing ferredoxin reductase which requires NAD(P)H as an electron source, and the other from an soluble iron- sulfur (ferredoxin) protein.[21] These, coupled with the P450 itself, form the three-

component Class I system (Fig. 2) $P450_{cam}$ is regarded as the model bacterial cytochrome and is an example of a Class I system, made up of three soluble components, the ferredoxin protein, Putidaredoxin, the corresponding reductase, Putidaredoxin reductase and $P450_{cam}$.

Class II redox systems are found mostly in eukaryotes and are two component systems made up of a membrane associated FAD/FMN (flavin-mononucleotide), NADPH- dependent cytochrome P450 reductase (CPR), and a separate P450 which is also membrane associated (*Fig.2, ii*).[22] Class II systems are common throughout steroid hormone catabolism, as well as formation of signalling molecules. Class I and II systems both work in the metabolism of xenobiotics and detoxification pathways.

Class III systems were first reported in 2002, sharing the typical three-component bacterial (and mitochondrial) P450 systems. Class III and IV systems were named as self- sufficient families upon the discovery of CYP102A (Class III) and CYP116B (Class IV). (Roberts *et al.*, 2002). In Class III systems, the electrons are transferred as in Class I systems, from the electron donor NAD(P)H, via a FAD domain and a further redox protein, in order to reach the P450 domain. However, unlike Class I systems, in which the second redox protein is soluble ferredoxin, Class III systems rely on a FMN domain, known as a flavodoxin. $P450_{cin}$ (CYP176A1) purified from *Citrobacter braakii* uses cineole 1 as its only carbon and energy source. (Hawkes *et al.*, 2002) This bacterial system demonstrated a novel method of electron delivery, in which unlike Class I systems, the electrons are delivered via the FAD/FMN redox centres.

Similarities between Class III and Class II systems can be seen in the redox partners, where Class II systems include fused CPR domains, in Class III systems the redox proteins are distinct and separate.

There are examples of recombinant P450 systems based on both Classes I and III. For example, P450BioI (CYP107HI) is as Class I P450 system

isolated from *Bacillus subtilis*, which is able to reduce a typical Class III flavodoxin reductase (FdR) redox system.(Hannemann *et al.*, 2007)

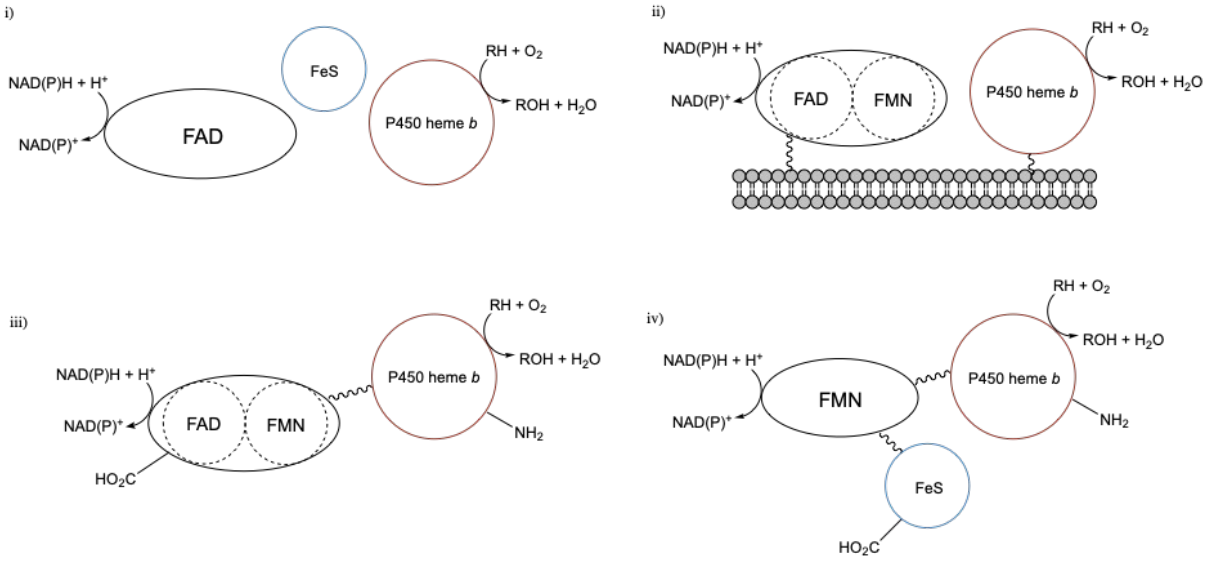


Figure 2. Simplified schematic of different P540 redox systems: i) Class I system containing a FAD domain and an iron- sulphur protein (ferredoxin) alongside the P450. Class I systems are found in the eukaryotic mitochondrial membrane. The P450 in Class II systems are paired with a diflavin reductase (FAD/FMN), in the Class III (iii) system however, the diflavin reductase is fused to the P450. Class IV systems are a single polypeptide chain, continuously linked, with a ferredoxin, FMN- reductase and the P450.

Class IV systems (Fig 2, iv) are not commonly found within nature. Instead they appear in extremophiles, with the first system discovered of its kind being CYP119A1 from *Sulfolobus solfataricus*. This small 368 residue system does not require NAD(P)H for electron donation and instead relies on pyruvic acid for its reducing ability, and then shuttles its electrons to the P450 domain with a thermostable ferredoxin. (Wright *et al.*, 1996) This find was the first high- temperature *in vitro* catalytic system reported, an

important step in the discovery of catalysts within industry.(Puchkaev and Ortiz De Montellano, 2005)

2.2 A Closer look at the Heme Domain

There are many variations of heme structures, alongside heme *b*, the two other major heme prosthetic groups are heme *a* and *c*. All three heme prosthetic groups share the same conserved iron protoporphyrin ring structure, with the ferric heme iron coordinated to four tetrapyrrole nitrogen atoms. Heme *a* , which typically appears green/ red in solution, is naturally occurring amongst many living organisms.(Poulos, 2014) An example of a heme *a* containing metalloprotease is cytochrome *c* oxidase, a protein which plays an important role in respiration. The isoprene derived hydroxyethylfarnesyl chain is thought to play an important role in energy conservation during the reduction of oxygen within the mitochondria.

Heme *c* covalently links to the protein backbone via the thiol groups, creating thioether linkages (Fig 3). These linkages ensure that heme *c* dissociates less readily than heme *b***,** also by displaying a range of reduction potentials, the *c*- type hemes show versatility in their functions, routinely serving as electron transfer proteins.(Bowman and Bren, 2008)

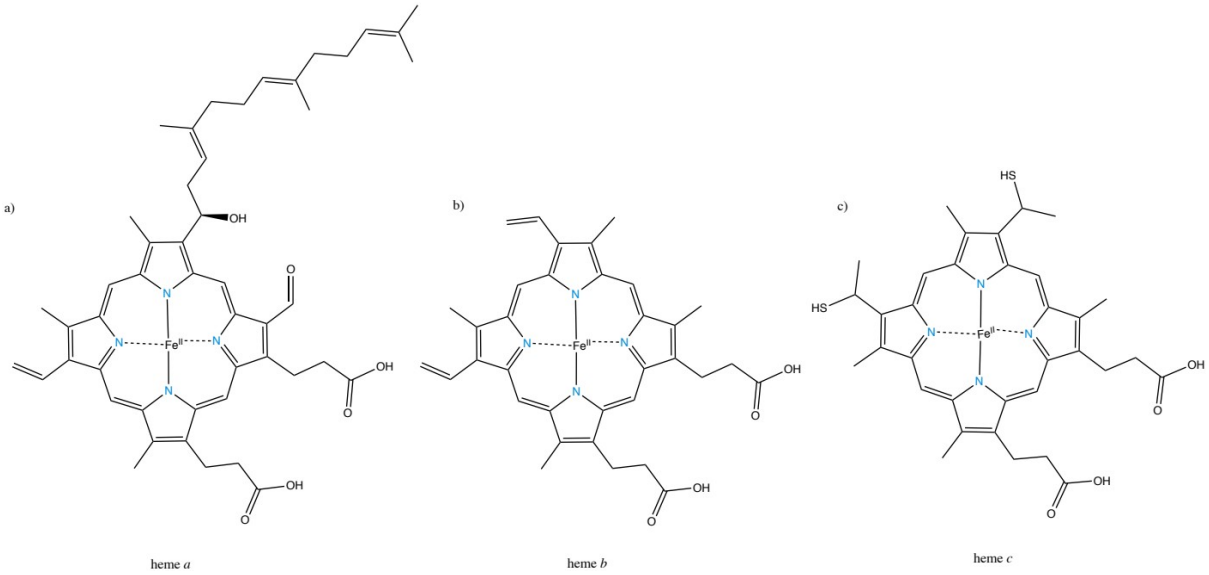


Fig 3. Comparison of the three major heme prosthetic groups, heme *a*, *b*, *c*. Each of the heme prosthetic groups contains the analogous ferriporphyrin ring structure. Heme *a* differs in structure to heme *b* due to the oxidation of the methyl group to form a formyl group and a hydroxyethylfarnesyl group, a structure derived from isoprene, creates a new side chain. Heme *c* has thioether linkages to replace the two vinyl groups, typically formed from two cysteine residues. This figure was created in ChemDraw.

The heme *b,* or ferriprotoporphyrin IX prosthetic group is found across the entire CYP P450 superfamily and is analogous to the heme in hemeoglobin and myoglobin. The heme iron is found in both the ferrous (Fe^{2+}) and ferric (Fe^{3+}) states which produces the characteristic 450 nm peak in the absorption, due to the coordination of a thiolate anion. The coordination of the cysteine residue was not confirmed until 1985, when the x- ray crystal structure of $P450_{cam}$ was solved..[23] A peak at 420 nm can be observed, known as the P420 form, arising from the reversible protonation of the cysteine thiolate anion to form a neutral thiol species.[14] (Sun *et al.*, 2013). It was previously thought that the EXXR motif, located within the K- helix of P450s was essential in CYP structure as a conserved moiety. This was disproven by the reporting of a novel CYP subfamily from *Streptommyces sp.* (CYP157C), where mutagenesis within the K- helix of variants lacking

the “essential” motif to EXXR resulted in abnormal folding (Rupasinghe *et al.*, 2006).

Upon substrate binding, the ferric heme iron spin-state shifts from low-spin (S= $\frac{1}{2}$) to high-spin (S= $\frac{5}{2}$). Electron transfer via redox partners is aided by the a corresponding increase in the heme iron reduction potential, initiating the catalytic cycle.[24]

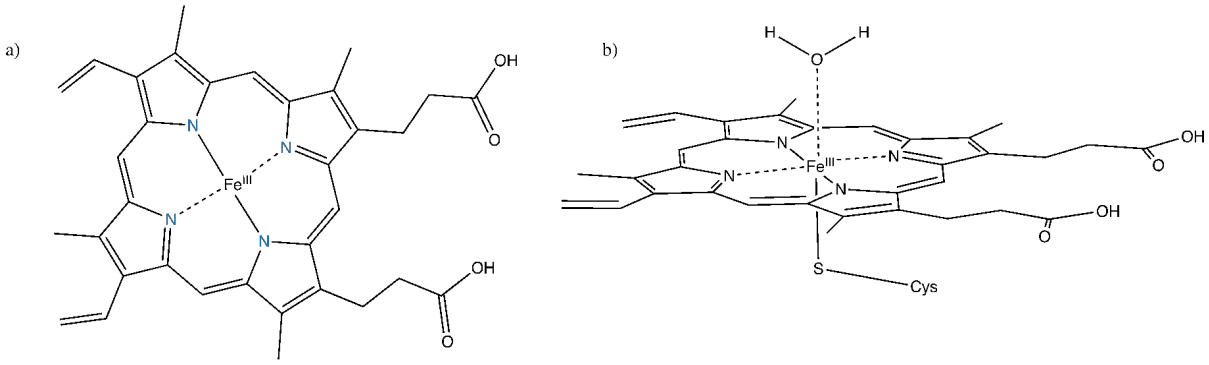


Figure 4. Heme *b* or iron protoporphyrin IX prosthetic group: a) Fe centre of *b*- type hemes are coordinated to four N atoms from the porphyrin ring. b) hexavalent Fe centre in substrate- free, “resting state”, bound to four N atoms, a cysteine residue and an exchangeable water molecule, axially bound.

Of the six ligand binding interactions with the heme iron centre, four planar nitrogen atoms coordinate to the Fe centre, originating from the tetrapyrrole ring *(Fig. 4).* The cysteine residue ligates at the C-terminus, in the proximal axial position, via a thiolate anion and the distal ligand is typically an exchangeable water molecule, in the substrate- free, “resting” state. Upon substrate or inhibitor binding, the water molecule is displaced and so the substrate is situated on the distal face of the porphyrin ring.[25]

The essential formation of the highly reactive iron- oxo intermediate via dioxygen bond cleavage is promoted by the cysteine ligand interaction, therefore making it critical for the catalytic activity of P450s.[26] The electron donating ability of the cysteine residue facilitates the reduction of

the ferric (Fe^{3+}) state to the ferrous (Fe^{2+}) state therefore weakening the O-O bond, in preparation for cleavage.[27] The subsequent reaction steps will be explored further in section 3.

3. Catalytic Activity of P450s

Throughout the catalytic cycle, shuttled electrons are responsible for the activation of molecular oxygen and the insertion of an oxygen atom into the given substrate.

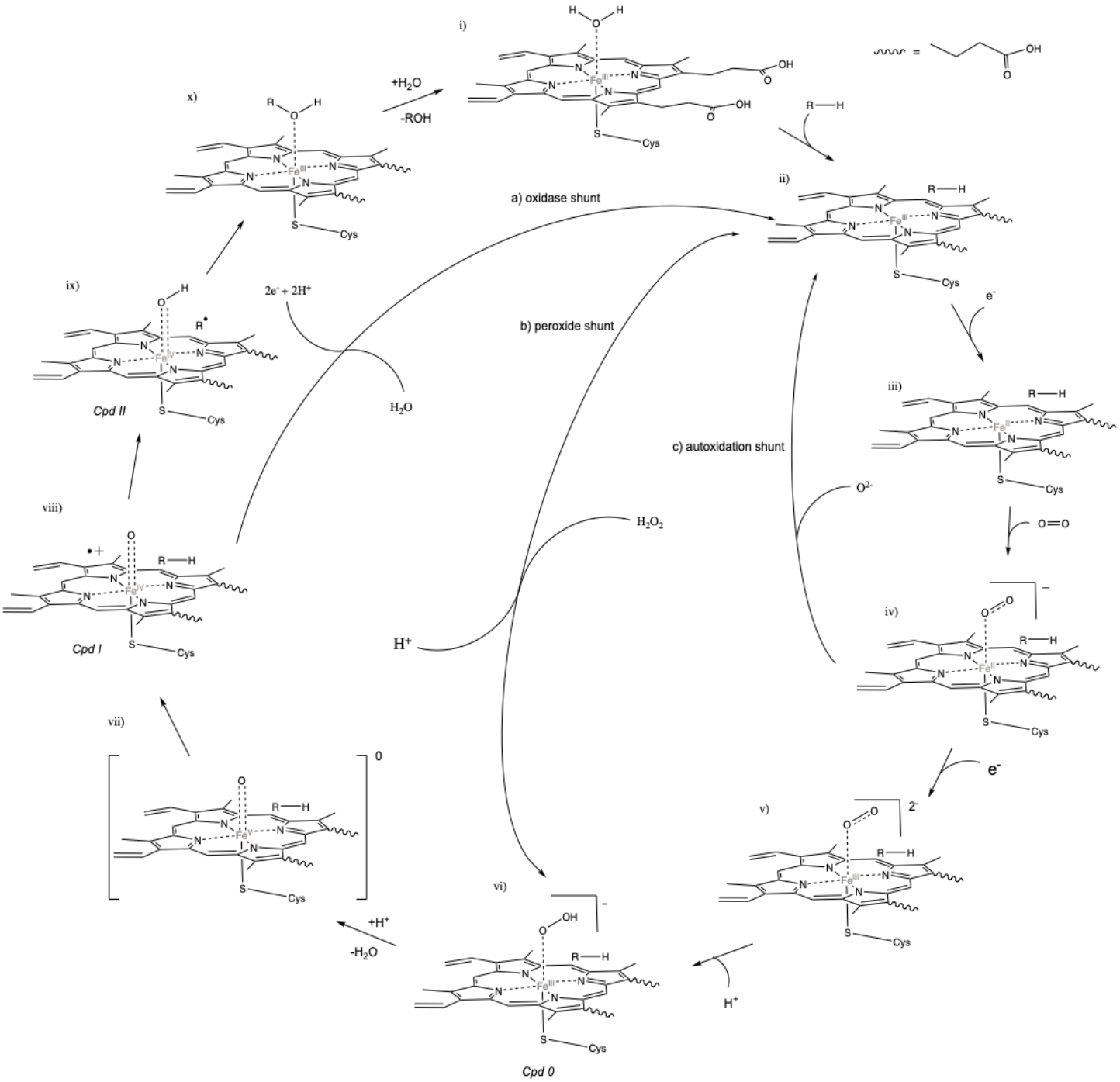


Figure 5. Catalytic cycle of P450s with shunt pathways included. i) the ferric (Fe^{3+}) resting state of the P450 heme, with an axially bound water molecule. Upon substrate binding (ii), the axial

water is displaced, inducing a low to high- spin shift in the heme iron centre ($S=\frac{1}{2}$¿$S=\frac{5}{2}$). This spin- shift facilitates the transfer of an electron, thus reducing the heme iron to a ferrous (Fe^{2+}) state. (iii). Molecular O_2 binds (iv) and the heme centre undergoes reduction once more (v) followed by a protonation, to form compound 0 (vi), the ferric- hydroperoxo species. Compound 0 is protonated and a water molecule is lost, resulting in an intermediate species (vii). The ferryl- oxo porphyrin radical cation species Compound I (viii) performs an H- abstraction from the substrate, followed by the formation of a bond between the C radical species and an oxygen, generating compound II (ix). The ROH product dissociates, leaving the heme iron centre in the ferric resting state (x to i). The shunt pathways shown: a) oxidase, b) peroxide and c) autoxidation shunt, are discussed further in the subsequent sections.

The most common reaction types catalysed by P450s are hydroxylations, which form the basis of general oxidations by P450s (see Fig 4.1). Following the formation of the reactive Cpd I species, a radical rebound mechanism results in a hydroxylation, as seen in Fig.... This is a particular reaction of interest, due to the otherwise difficult nature of oxygen insertion reactions with traditional synthetic methods. This mechanism of hydroxylation was first proposed in 1978 by Groves and McClusky, when a Fe(V)=O species was proposed as a hydroxylating agent in the CYP P450- catalysed hydroxylation of 2,3,5,6- tetradeuterated norbornane and investigating the creation endo- vs exo- hydroxy product. Due to the partial loss of stereochemistry, the radical rebound mechanism was proposed, which accounted for the large kinetic isotope effects (KIE) seen with deuterium experiments (Groves *et al.*, 1978).

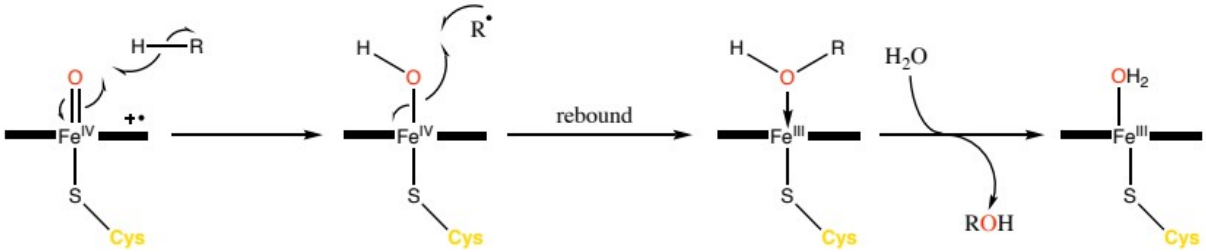


Figure 4.1 Hydroxylation mechanism of P450s via a radical rebound mechanism.

A hydrogen abstraction via Cpd I yields a highly reactive radical species, which is retained within the active site, which then goes on to react with the hydroxy moiety of the ferryl species, forming an alcohol. A water molecule coordinates to displace the -OH and the six- co- ordinate resting state is achieved, releasing the alcohol which is recycled to oxidise the subsequent molecule. (Ortiz de Montellano and De Voss, 2002)

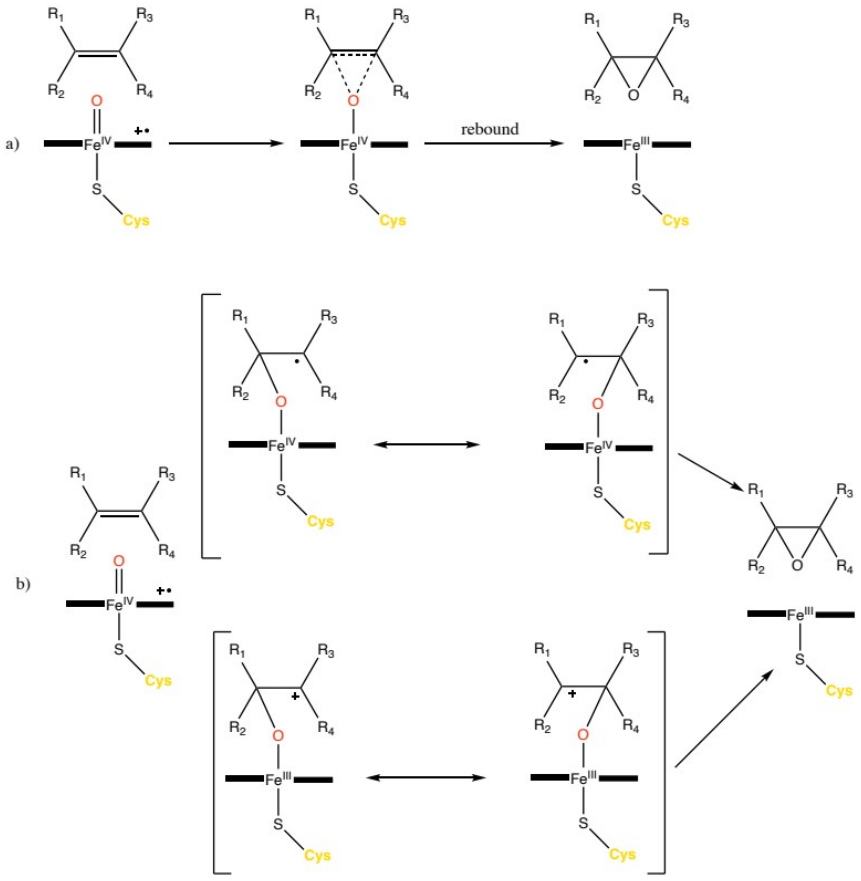


Figure 1.3 Scheme outlining two possible mechanisms of P450 mediated alkene epoxidation, one concerted (a) and non- concerted (b).

In addition to hydroxylations, P450s are able to catalyse the epoxidation of olefinic double bonds, with the retention of stereochemistry in the formed epoxide. Hanzlik and Shearer disregarded the accepted route of epoxidation via a concerted mechanism, when they performed a deuterium substitution of olefinic carbons of the two styrene derivates of styrene, p- methyl and p- phenyl (Hanzlik *et al.*, 1978).

Mechanistic investigations into the P450- mediated activation of aromatic compounds have yielded many complex pathways with intermediates and proposals of many alternative mechanistic routes. Density functional theory, in combination with molecular modelling (DFT/MM) has determined the presence of the Meisenheimer tetrahedral intermediate (I_T), and demonstrated the rate- determining π-activation step. (Vaz, McGinnity and Coon, 1998; Shaik *et al.*, 2011) These hybrid density functional studies have shown, much like the epoxidation mechanism, a hydrogen abstraction and electron transfer are not observed in the first step of the mechanism.(Jones, Mysinger and Korzekwa, 2002) There is a combination of radical and electrophilic attack where the oxo group, π system is attacked by Cpd I, to form a radical and cationic σ complex, with the major, lower energy component being the cationic intermediate. (Bathelt *et al.*, 2003) This is seen with deuterium studies. If the substrate contains deuterium at the site of hydroxylation, this should be retained within the product due to the migration toward the adjacent carbon in the mechanism of phenol formation, the NIH shift.(Jerina and Daly, 1974) This phenomenon suggests that P450- mediated aromatic oxidations are much

like aliphatic epoxidations in the isomerization of the intermediate arenes as seen in Fig 1.4.

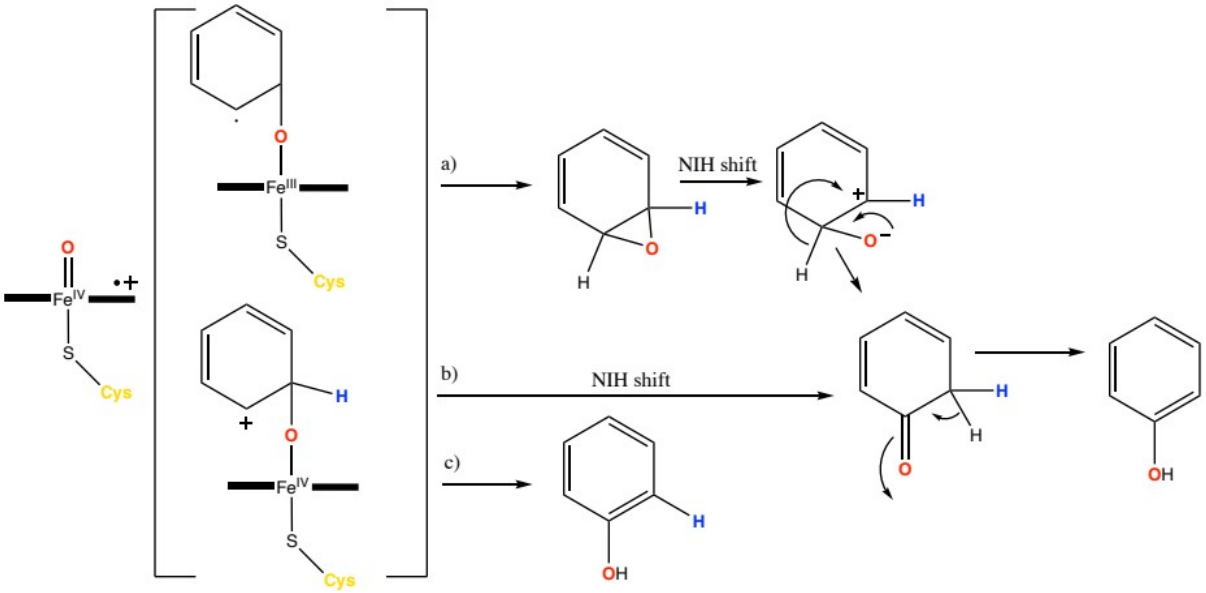


Figure 1.4 Scheme demonstrating three pathways of benzene hydroxylation. In the first pathway a) proceeds via the SIGMA- radical complex, the epoxide is formed, followed by cyclohexanone formation. NIH shift and tautomerisation gives the phenol product. b) and c) proceed via the SIGMA cation complex can form the cyclohenone via NIH shift then forms the phenol. c) the sigma cation species is protonated directly to give the phenol.

In the resting state, the ferric Fe centre is in a low- spin state. Upon substrate binding within the enzyme active site, a conformational change is induced as the distal water ligand is displaced.[28] As previously mentioned, the catalytic cycle of P450s begins with the binding of a substrate inducing a spin shift that reduces the ferric heme iron centre, Fe^{3+}, to Fe^{2+}. Following conformational changes within the active site, the thermodynamically favourable first electron transfer from the redox partner (CPR) is facilitated by the increased reduction potential of the heme Fe in the high- spin state. Molecular oxygen binds on the distal face, to the ferrous Fe centre, with the second electron transfer from CPR aiding the formation of the highly reactive “compound I” ferric- peroxo species

(see Fig 5, viii). Compound I is the intermediate state which is responsible for hydroxylation of the substrate.[29]

3.1 P450 heme iron spin- state

The redox potential of the heme iron centre and the resulting spin state underpin the catalytic activity of P450s. As seen in the P450 catalytic cycle, the ferric heme iron (Fe^{3+}) has the electron configuration: $1s^2$ $2s^2$ $2p^6$ $3s^2$ $3p^6$ $3d^5$ in its resting state. The displacement of the distal water molecule, induced by substrate binding, creates a more positive redox potential, and facilitates the transfer of an electron to the now ferrous (Fe^{2+}) iron centre. The resulting destabilisation of the hexacoordinated Fe^{2+} complex promotes a rearrangement of d- orbital electrons, prompting the formation of an energetically unfavourable pentacoordinated complex. The degeneracy of the d- orbitals comes about due to the repulsion between electrons on the N^{2-} ions and the electrons in the 3*d* orbitals of the Fe centre, causing the energy of the orbitals to increase. The *d*-orbitals in octahedral complexes are split with the e_g orbitals increasing in energy much more than the t_{2g} orbitals, as the e_g orbitals lie with overlapping symmetry with the ligand atoms, to form sigma (σ) interactions. The three t_{2g} orbitals, lie between the symmetry axes and so form pi (π) interactions.

The low- spin (LS) paired 3d electrons move to the e_g orbital, in a high-spin (HS) state (LS → HS) due to the change in octahedral crystal- field spitting energy (CFSE, Δ_{oct}). In the LS configuration, the electrons occupy the lower energy t_{2g} orbitals as the energy cost of pairing electrons is less

of that than the overall CFSE, Δ_{oct1}. In the HS state, the CFSE, Δ_{oct2}, is smaller and less of electron pairing energy and so, the electrons populate the higher energy eg orbitals ($\Delta_{oct1} > \Delta_{oct2}$). The spectroscopic properties associated with the spin- shift

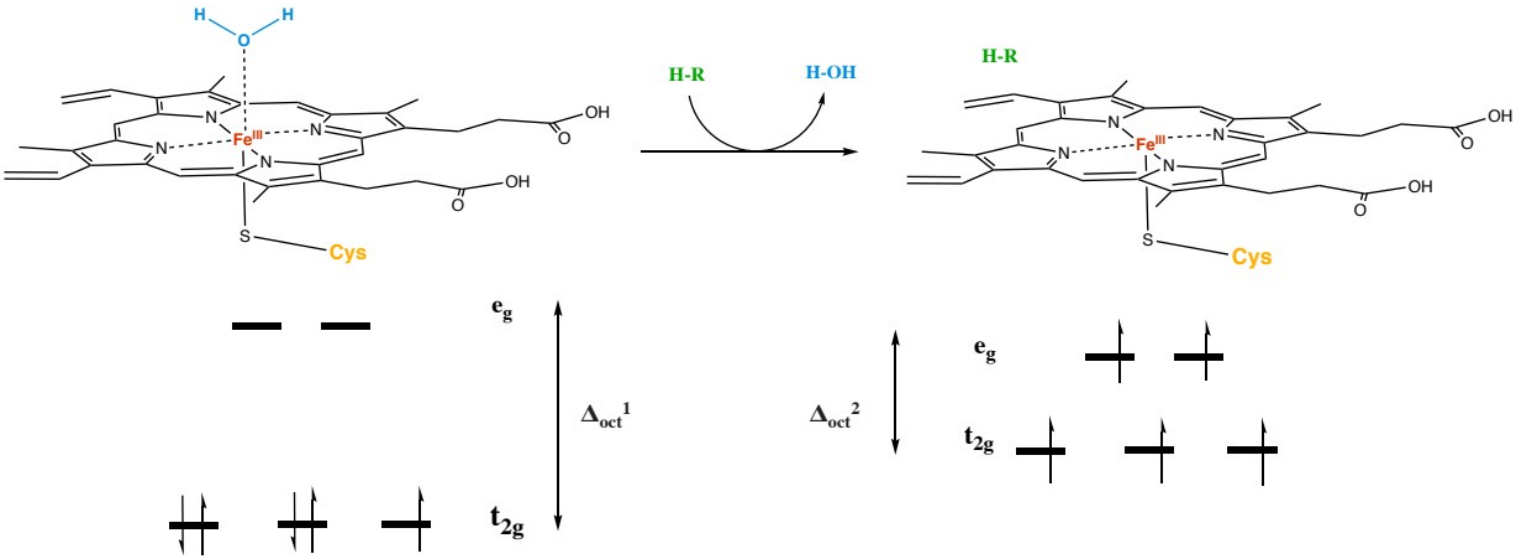


Fig 5. d- orbital splitting diagram of the ferric heme iron, showing electron rearrangement following a spin- shift induced by substrate binding. The axial water molecule is displaced up on substrate binding and a decrease in the overall crystal field splitting energy (Δ_{oct}) prompt as spin- shift from low to high- spin (LS → HS).

3.2 The Shunt pathways

Fig. 4 depicts three shunt pathways within the catalytic cycle of P450s using alternative routes of electron shuttling and result in the formation of water a reactive oxygen species, superoxide or peroxide. The typical

electron carrier, NAD(P)H, becomes uncoupled to the oxidation of the substrate.

Peroxide uncoupling (fig 4. b) is reversible and occurs when the substrate binding is unable to expel water from the catalytic site efficiently, leaving the heme iron in a non- productive mode, this facilitates the formation of Compound 0, via protonation of the Fe- bound oxygen. Hydrogen peroxide is released and the Fe centre returns to the ferric state. When concentrations of peroxide are exceedingly high however, the heme moiety begins to degrade, and so the catalytic efficiency of the enzyme will fall to an eventual inactive enzyme.

The peroxide shunt pathway has been favourably utilised in industry, allowing H_2O_2 to shuttle electrons rather than NAD(P)H. Peroxide concentrations must be high, thus forming Cpd 0 directly, and the cycle proceeds forwards. There are examples of P450s utilising this shunt pathway naturally, in order to oxidise products. OleT is an enzyme belonging to the CYP152 family, which, via oxidative decarboxylation, produced terminal olefins. (Munro *et al.*, 2018)

The oxidase shunt pathway (Fig 5. a), results when a substrate is orientated such that H- abstraction cannot occur. A water molecule is formed when the oxygen atom of the highly reactive ferryl- oxo radical species, Cpd I (Fig 5, viii), is reduced, resulting in the reformation of the pentacoordinated, substrate- bound state.(Cook *et al.*, 2016) This is also observed in the autoxidation shunt, where the autooxidation of the ferrous- oxo complex results in the release of a superoxide species. (Luthra, Denisov and Sligar, 2011) In both the oxidase and autoxidation shunt pathway, no oxidised products are formed, despite the consumption of electrons. Uncoupling reactions occur when the oxidation of a substrate is troublesome, with the substrate occupying the active site in non-favourable orientations, and when access to the heme is not controlled.

These uncoupling reactions are of particular interest to pharmacologists, biochemists and toxicologists as they are major unproductive pathways, and so the factors affecting their rates are studied in depth.

Human P450s, xenobiotic metabolism and FDA approval

Human P450s

The purification and isolation of P450s from human microsomes began in the 1970s, with P450 subfamily 3A and 2C most highly isolated due to their stability and abundance.[30] Throughout the introduction of recombinant DNA technologies made it possible for human cDNAs to be cloned and introduced primarily into mammalian and yeast systems, and by the 1990s, into bacterial systems, easing the methods of expression and purification (Gonzalez *et al.*, 1988). There are 18 human P450 families, divided into 41 subfamilies (Nelson *et al.*, 2004). Human CYPs vary in selectivity and regioselectivity, in which those variants involved in xenobiotic transformations tending to be more promiscuous in nature and therefore are able to metabolise a wide range of substrates.

CYP1A/2A

First characterized in rat liver microsomes, the CYP1A subfamily was originally thought to have induced toxic, undesirable side- effects such as activation of carcinogens.REF. However it is now known that the CYP1A subfamily contains just two functional genes, CYP1A1 and CYP1A2.

CYP1A is mainly responsible for hydroxylation and oxidation of aromatic compounds and hormones, including progesterone, where CYP1A1 substrates tend to be aromatic hydrocarbons, whereas CYP1A2 prefers

heterocyclic compounds and aromatic amines (Zanger and Schwab, 2013; Lu *et al.*, 2020). CYP1A2 is expressed within the liver at high concentrations and is the isoform that is responsible for the metabolism of 10% of clinically relevant drug molecules, metabolised by CYPs (Zhou, Liu and Chowbay, 2009). The importance of CYP1A in Phase I drug metabolism will be further reviewed in subsequent sections.

CYP2

The CYP2 gene family is the largest and most diverse amongst nature, making this enzyme family of importance to many physiological and toxicological processes (Goldstone *et al.*, 2006). Six major CYP2 isoforms are involved within hepatic metabolism pathways of over 100 structurally diverse xenobiotics, for example, CYP2C8 favours oxidation of members from the anti- diabetic, glitazone drug class (Backman *et al.*, 2016). CYP2C9 is another major isoform in xenobiotic metabolism, accounting for 15-20% of Phase I metabolism (Visser *et al.*, 2007). With many known reversible and irreversible CYP2 inhibitors, this family is a major source of CYP mediated drug- drug interactions when administered in conjunction with drugs that display even low therapeutic indexes (Holstein *et al.*, 2005).

CYP3A

Most of CYP- mediated metabolism can be accounted for by three CYP3 isoforms, majorly expressed within critical tissues found in the liver and gastrointestinal tract (Wilkinson, 1996).

CYP3A4, a very versatile enzyme, found in the highest concentrations in the liver, plays a role in the metabolism of the majority of drugs, across a variety of drug classes, including codine, statins, steroids and the macrocyclic erythromycin (Guengerich, 2015). [31] The immense range of substrates that CYP3A4 is able to oxidise is due to its conformational flexibility, making it an ideal candidate for drug metabolism studies.

.

An Introduction to Cytochrome P450 BM3 (CYP102A1)

An overview of CYP102A1

P450 BM3 was named so due to the fact it was the third isolated P450 from *Bacillus megaterium* (*B. megaterium),* following the two smaller variants BM1 and BM2. The Fulco group first isolated the CYP 102A1 from the soil bacterium, *B. megaterium,* in the 1980s, where it was first described as a medium to long- chain fatty acid hydroxylase (C12- C20) which functioned independently with only molecular oxygen and NADPH present.[43] Due to its unrivalled turnover rates, coupled with its stability, the naturally occurring fusion protein, P450 BM3, has been one of the most extensively studied P450 isoform in the past 40 years. Preferential substrates of BM3 are medium length, straight chain saturated and unsaturated fatty acids (C15/16) where hydroxylation occurs at the ω–1-ω–3 positions, with fastest turnover seen with C20 unsaturated arachidonic acid REF

BM3 is made up of a single 119000 kDa polypeptide, with Class II-mammalian redox system properties: an FAD/ FMN CPR,.[44]

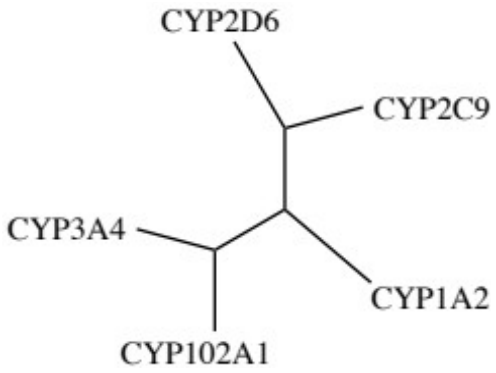


Fig.. Phylogenetic tree based on sequence of P450 BM3 (CYP102A1) in relation to other human isoforms important within drug metabolism. Figure adapted from Di Nardo et al. with original phylogenetic tree formed

from sequence alignment of substrate recognition sites (SRS) of the above CYP variants.(Di Nardo and Gilardi, 2012) REF IN TEXT

The role of P450 BM3 within nature remains somewhat ambiguous, there is however evidence to support that the role of BM3 is to aid the detoxification of polyunsaturated fatty acids, as the synthesis of BM3 is promoted by Bm3R1 repressor which is inhibited by high concentrations of polyunsaturated fatty acids. It has been hypothesised that this mechanism could be linked to an ancient metabolic homeostasis response to an abundance of these polyunsaturated fatty acids.[45]

In order to exploit the catalytic activity of BM3, it is necessary to look at the organization within the fusion structure.

Structure and Catalytic Mechanism of $P450_{BM3}$ [46]

The three discrete domains of P450 BM3 total 1048 amino acid residues: the heme domain (1-470), the FMN domain (470-664) and lastly the FAD domain from the remaining residues (664-1048). Upon further characterisation, it was found that the heme region, BMP, was fused to the reductase domain, made up of both flavin coenzymes (FMN/FAD), BMR in a stoichiometric ratio of 1:1.

Furthermore, BM3 is a model class II P450 as it is devoid of this N-terminus anchor region present in most eukaryotic P450s to tether them to the mitochondrial membrane. This membrane anchor region is necessary for electron transfer and therefore catalytic activity of eukaryotic P450s. The intact BM3 has been shown to be present as a dimeric enzyme. [47]

Resolved crystal structures of each domain in complex with a range of chemically diverse substrates have been imaged.

The Heme domain and key residues

The P450 BM3 heme domain is made up of a larger alpha and beta subdomain, with the alpha domain making up around 70% of the overall heme domain. The α- domain is comprised of 15 α- and six 3_{10}-helice and the β- domain is rich in β- sheets, with 12 strands in total. The porphrin ring structure resides between the I and L- helices, with the I- helix above. The heme iron is coordinated at the C400 sulphur residue and is also coordinated to a distal water ligand. A substrate access channel (SAC), open in the resting state, helps guide substrates directly towards the heme iron, the beginning of which is found at the residues Arg47 and Tyr51, between the F/G loop.[48] This is unlike that of $P450_{cam}$ where the SAC is too small, even in the resting state, to allow substrates to travel through. The role of the Arg47 and Tyr51 residues have been described as fatty acid anchor regions, tethering to the carboxylate functionalities. [49] In the substrate- free state, the SAC is full with solvent molecules, upon substrate binding a conformational change shifts the positioning of the F/G loop and tilts the F-/G- helices to close the SAC, therefore locking the substrate in position, distal to the heme iron centre. The I- helix structure and kink region is analogous amongst many P450s. The conformational changes upon substrate binding also induces changes in the interactions of the salt bridge with the I- helix, altering the I- helix kink and thus disrupting the hydrogen bonding network.[50] The whole porphyrin ring moves as a result of this, easing the displacement of the water molecule upon substrate binding. The displaced water molecule then goes on to interact with the Thr268 residue of the I- helix (FigureX).

Thr268 plays a role in the protonation of the iron- oxo intermediates within the catalytic cycle (Fig..) and is also a conserved residue amongst P450s.

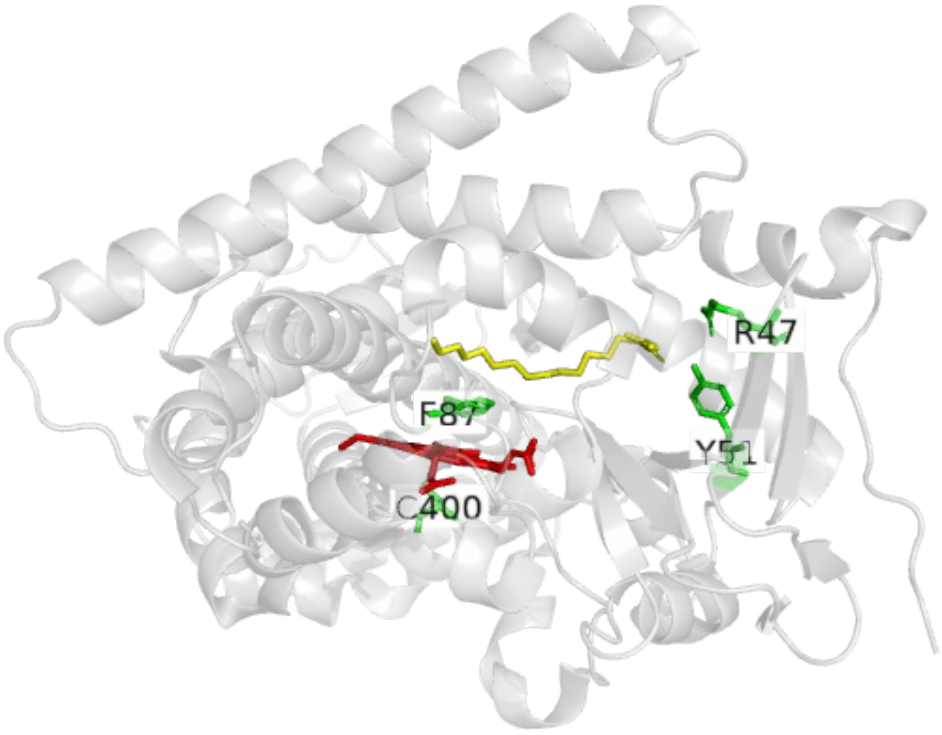


Fig.. Heme domain of BM3 in complex with N-palmityoylglycine (NPG). BM3 is shown in cartoon representation, and the heme and key residues shown as sticks. The substrate, NPG, is shown in yellow and located across the substrate access channel, protophoryrin IX ring in red, and key residues shown in green. NPG is associated to the Arg47/Tyr51

residues, found at the mouth of the substrate access channel. (PDB 1JPZ). Figure adapted from Haines et al., 2001.[51]

Add important catalytic residues separately….

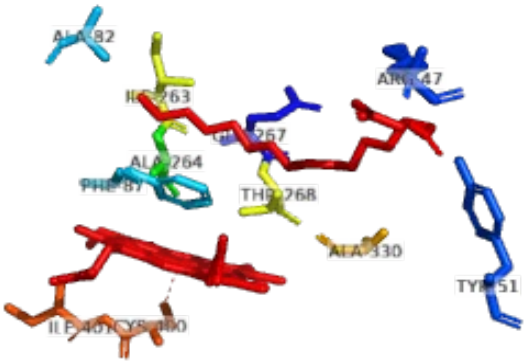


Fig.. Key residues within the active site of BM3. The key active site residues are depicted as sticks. These residues have been reported to aid…

Further key residues within the heme residue can be found when looking at the crystal structure of the BM3 heme domain complexed with N-palmityoylglycine (NPG) (Fig…) .In terms of substrate binding, Phe87, located on the distal face of the heme, has proven to be of paramount importance. As seen in Fig… Upon substrate binding, the Phe87 residue positions itself between the heme iron and the ω- methyl of the substrate, preventing monooxygenation at this terminal. The large phenyl ring on the Phe87 residue structure can be sterically hindering towards certain substrates and therefore can act as somewhat of a filter.

F87A mutagenesis studies have provided a in shift preferential hydroxylation from the preferred ω–1-3 regions. Replacement of F87 with smaller residues allows for the expansion of the active site, allowing long

chain fatty acids to fit further in, thus hydroxylating γ- and δ-positions. These initial mutations however reduced the activity of the wild- type enzyme.[52]

F87 mutagenesis in particular has allowed for intrusive regioselectivity studies….
F87A..
Another F87V.. allowed for 100% 14(S)-15(R)-arachidonic acid epoxidation.. [53]

The FMN domain

The structure of the FMN domain of BM3 was first crystallised in 1999.[54] As seen in Fig… The FMN domain consists of a central, five stranded β-sheet, surrounded by four α-helices, analogous to that of rat CPR.[55]

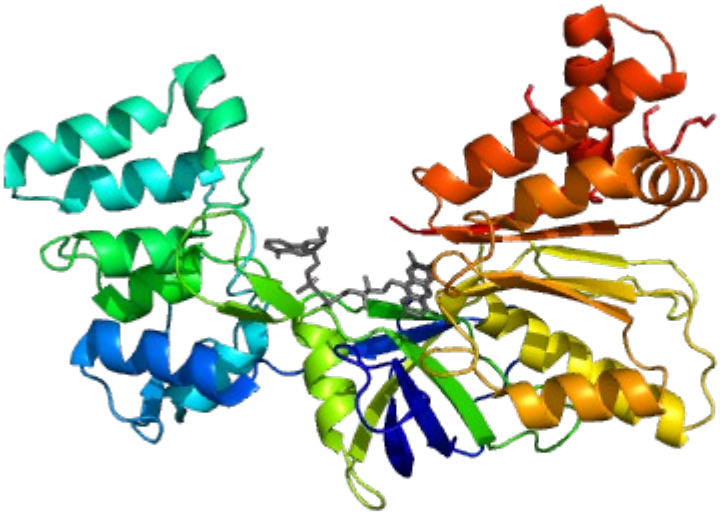
Key residues in the FMN binding region were demonstrated by mutagenesis studies. Site directed mutagenesis carried out by Fulco et al. introduced a bulkier amino acid in place of Gly570. The G570A mutant demonstrated a weakened affinity for the FMN cofactor, with the bulkier amino acid substituent protruding into the FMN binding pocked and interfering with cofactor interactions.[56]

The FMN cofactor exists between the semiquinone and fully oxidised species during the catalytic mechanism and therefore can act as a one or two electron donor. Chen et al. stabilized the neutral semiquinone state by means of insertion of an otherwise absent glycine residue. In the resting- state, the FMN cofactor is in a hydroquinone state, thermodynamically preferred and unable to shuttle electrons to the heme iron. A conserved glycine residue is absent in the BM3 FMN domain, results in a shunted, more rigid co-factor binding loop.[57] The small residue, introduced to the re- face in order to extend the FMN- binding region, allowed greater flexibility within the binding loop which was stabilizing towards the FMN semiquinone cofactor.

The FAD domain

The last domain of P450 BM3 to be structurally solved is that of the FAD domain. The stabilizing factors contributed by key residues were demonstrated by Joyce et al. in 2012 (Fig…)[58] Furthermore, the role and preference of NADPH over NADH was determined. A C773A/C999A double mutant was used to in order to remove any intra-/ intermolecular interactions leading to unwarranted disulphide bridge formation via cysteine residues and giving a homogeneous structure to be resolved.[59]

Fig.. Structure of FAD- binding domain of P450 BM3. (pdb- code



4DQK)JOYCE et al(Joyce *et al.*, 2012)

The FAD domain is approximately 42.0 kDa in a single, monomeric form and spans from amino acids residues 653- 1048 of the whole 1048 residue BM3 species.[60] The NADPH binding subdomain resides in residues 888-1048. The crystal structure of the FAD domains shows the presence of three subdomains, comprised of 5 α-helices and β-strands.

Key residues in the FAD domains have helps elucidate cofactor binding site positions at residues 660-705 and 826-887 and also the linker region connecting FAD/FMN- binding domains from residues 705-825. The most conserved motif found amongst P450 FAD- binding proteins was also found, made up of Gln756, Arg827, Tyr860 and the N- backbones of Tyr828, Tyr829, , Ile863 and Ala864. These residues are responsible for cofactor recognition and polar pyrophosphate binding interactions.[61]

It is important to note the conserved 'catalytic triad' residues: Ser830, Cys999 and Asp1044, located in proximity to the FAD isoalloxazine ring. These residues are analogous to those that were first found within rat CPR. This triad is responsible for electron shuttling via H^- ions, from NADPH to the FAD cofactor.[62]

Due to the fusion structure of P450 BM3..

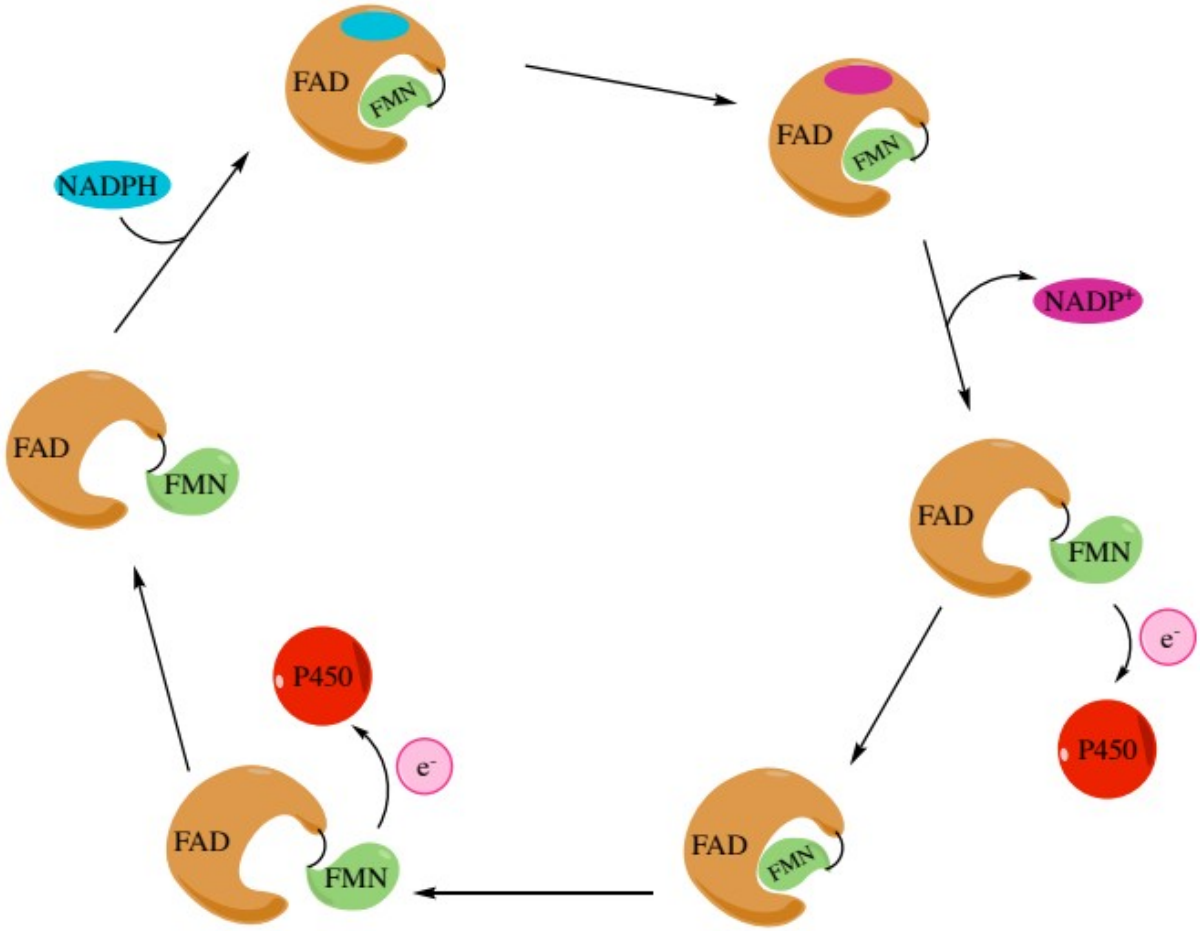
Electron transfer in P450 BM3

In the same manner as eukaryotic microsomal P450s, the route of electron transfer within P450 BM3 is as follows: NADP → FAD → FMN → heme, though the intramolecular electron transfer system[63] . The speed and efficiency of electron transfer within P450 BM3 is owed to the fusion nature of the protein domains and its existence in a dimeric form. Evidence for the dimeric enzyme form was first presented by……

Flavin reduction… (Roitel, Scrutton and Munro, 2003)

Comparative studies with nitric oxide synthase (NOS), which also exhibits loss of catalytic activity in the monomeric state, shows that this loss of activity in the monomeric species is not exclusive to BM3….

CYP 102A1 is the fastest known monoxygenase to date, when monitoring turnovers with its preferred substrate, arachidonic acid, rates of around 17000 min^{-1} have been reported (Munro *et al.*, 2002).



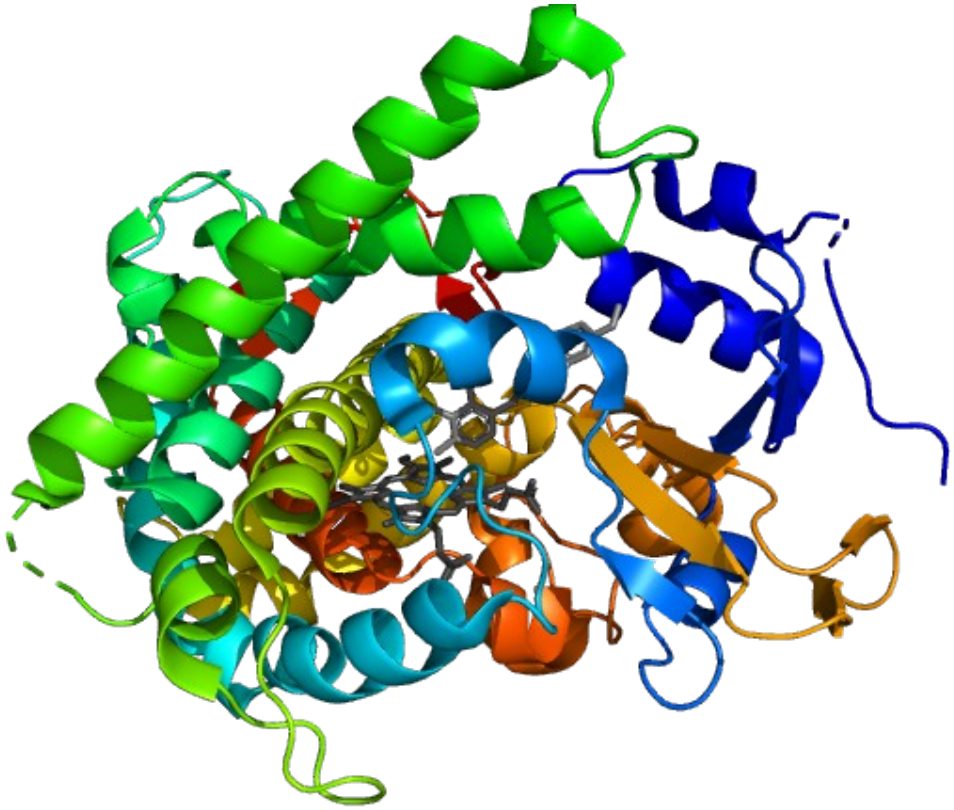
Fig… Simplified schematic of electron transfer in BM3 with CPR conformational changes.

Figure outlines conformational changed which occur in th FAD- binding domain of BM3, in order to shuttle electrons. NADPH (blue), shuttles electrons to the FAD domain (brown), then onto the FMN domain (green). $NADP^+$ (pink) is expelled. The FMN domain rotates to interact with the FAD domain. The FMN domain roates away from the FAD domain to transfer

electrons to the P450 (red). Two electrons are delivered with these conformational changes. Figure produced using ChemBioDraw.

The “Gatekeepter” mutation

Previous studies from within the Munro lab showed the outcome of the DM “gatekeeper” mutation in the P450 BM3 heme domain, and its ability to structurally alter the heme domain.



BM3 DM complexed with omeprazole. Figure adapted from (Butler *et al.*, 2013)…. EXPAND

The double “gatekeeper” mutation A82F/F87V

1.3 Drug metabolism, ADME and the Glitazone Drug class

When a foreign compound enters the body, and more specifically, the systemic circulation it is termed a xenobiotic.[32] Synthetic chemicals, pharmaceuticals, cosmetics, pesticides and additives are all examples of xenobiotics. Great care must be taken in observing and profiling the activity of xenobiotics, especially pharmaceuticals. Information on the mechanism of action of a drug molecule, the absorption, distribution, metabolism and excretion (ADME) is important in evaluating a safety and efficacy profile.[33]

1.3.1 ADME

For over 50 years, the acronym ADME has governed the laws of pharmacokinetics.[34] Administration or Absorption describes the route by which the drug is given, also factoring in how much and which times at which the drug should be administered. Bioavailability is the actual amount of the administered drug which enters the systemic circulation and is critically determined by the absorption of a drug. Distribution describes how the compound reaches the target site. Metabolism is the breakdown of the compounds ready for Excretion, predominantly via the kidneys. Excretion is an important factor to monitor in pre- clinical trials as an accumulation of a drug compound within the body may have adverse toxic effects.[35]

1.3.2 Drug Metabolism

Drug metabolism itself has important consequences on the pharmacokinetics (PK) and pharmacodynamics (PD) of a given drug molecule. Routes of drug metabolism is of paramount importance when evaluating the safety and efficacy of any novel or existing xenobiotic. The

biotransformation of a drug within the human body can be categorised by three pathways, Phase I and Phase II and occasionally a further Phase III metabolism (Zhang and Tang, 2018).

CYP- mediated metabolism is the route by which many small molecule drugs are processed for excretion. Reasons for this may be due to P450s accepting a wide variety and range of substrates in comparison with other enzymes, and their reaction products (hydroxylations, epoxide formations) may be preferred for clearance mechanisms from the body, as they tend to be unstable and reactive (Guengerich, 2003).

During Phase I metabolism, functionalisation reactions such as hydrolysis, cyclisation, oxidation and reduction are carried out by oxidases to produce more polar metabolites.[36] Following Phase I metabolism, the metabolite may be excreted if it is polar enough. C-H conversion to C-OH is an example of a common functionalisation carried out within this pathway (Wrighton and Stevens, 1992). Pharmacologically inactive substances can be activated, *in vivo*, by Phase I pathways, these compounds are known as prodrugs. This class of drugs was primarily developed to improve bioavailability by altering selectivity, which in turn decreases any unwanted side effects (Rendic, 2002).

Subsequent Phase II reactions involve conjugation of the drug metabolites with endogenous charged species to form higher molecular weight compounds (Xu, Li and Kong, 2005). A reduced form of glutathione (GSH) is a common conjugating species formed of a linear tripeptide and a free thiol group. GSH can increase the water solubility of a metabolite to allow for secretion into bile or urine. Phase III metabolism can occur following Phase II and involves further priming the metabolite in preparation for excretion by the addition of more hydrophobic functional groups (Beigneux *et al.*, 2002).

Metabolite studies are pivotal initial stages of any drug discovery and development pipeline, with the screening of new chemical entities (NCEs), providing many lead compounds which are taken onto subsequent stages as potential drug candidates. In order for these to be successful, full metabolite profiles and identification are necessary to distinguish whether the metabolites are toxic or pharmacologically active themselves (Prasad *et al.*, 2011). Metabolism studies are complex in nature, due to the varying and numerous enzymatic pathways involved in the production of many metabolites. These different metabolites are also formed in a range of concentrations, with some formed at trace concentrations whereas others formed are classed as major metabolites, which can push the limits of detection of certain analytical methods (Tsamandouras *et al.*, 2017).

1.3.3 Governing the pharmaceutical industries

Developments and research into the role of human P450s in the metabolism of xenobiotics and other exogenous substances is of paramount importance in control and regulation of chemicals. Over 85,000 chemical substances are deemed dangerous under the Toxic Substances Control Act (TSCA) implemented by the Environmental Protection Agency (EPA) in 1976.[37]

Within the pharmaceutical industry, rigorous testing and standards are upheld by regulatory authorities, such as the Food and Drug Administration in the USA. In order for a New Drug Application (NDA) to progress into further stages of development, the active pharmaceutical ingredient (API) and its metabolites, must be quantified and analysed.

The demand for P450 research in the pharmaceutical industries

1.3.4 Mimicking Drug Metabolism in Vitro

To meet the standards of FDA approval, there is a demand for large scale production of human drug metabolites. Classical synthetic routes can prove to be costly due to use of expensive starting materials and maintenance of sensitive reaction conditions.[38]

Phase I metabolism has been classically simulated in vitro with the oxidation of sp^3 hybridizes C-H bonds with non- enzyme catalysts. Chen & White's use of a Fe- containing catalyst to achieve oxidation of aliphatic C-H bonds, though highly selective, involved inefficient, multi- step synthetic routes that are difficult to scale-up for industrial production.[39] The use of these transition metal catalysts is inefficient and so, is low yielding. Traditional chemical synthesis is also non- specific resulting in the formation of complex mixtures, difficult to resolve.

Following the difficulty of traditional chemical synthesis, human microsomal liver cells were isolated and used *in vitro*. The use of human P450s allowed reaction conditions within the lab to better mimic the natural working conditions of the enzymes within the human body.[40] However, many problems are encountered when using human microsomes as differing concentrations of a cocktail of P450s produce various metabolites which require further separation and resolution steps.[41] Furthermore, the lack of stability of membrane bound P450s means that difficulties are encountered during purification.

Use of recombinant human P450 systems to produce drug metabolites has solved a significant number of difficulties mostly due to their solubility.[42] They are also able to be expressed at high yields in a prokaryotic vector such as *Eschestria coli.* As seen in the next chapter, the turnover rate of P450 BM3 vastly rivals that of any human microsomes and overcomes many of the issues faced with traditional chemical synthesis and *in vitro* human microsomal incubations.

Pharmaceutical Applications of Drug Metabolites

P450 BM3 is utilised throughout the drug discovery pipeline, which is generally a lengthy and expensive process, with most drugs taking an average of ten years to reach the commercial market.(Mohs and Greig, 2017) In the earlier stage of drug development, intrusive studies into the metabolism of novel drugs is of paramount importance, in order to determine drug- drug interactions, toxicity, which alters the drugs half-life, and various other side effects which may arise from toxic metabolites. The earlier problems with a drugs toxicity of efficacy is detected, the less time and expense is wasted downstream, where failed drugs do not progress past initial formulation studies, preclinical or clinical phase trials. (Singh, 2006) The FDA states that for all metabolites which are produced within 10% of the initial parent drug, a full metabolic profile must outline a full identification and characterization.

There has been evidence of drug metabolites having the same mode of action as the parent drug and therefore massively increase the bioactivity of the drug overall. The beta- blocker propranolol and one of its major metabolites, 4'-hydroxypropanolol exhibit this effect….
Propranolol can also be considered as a leading example, demonstrating the use of bacterial P450s to selectively hydroxylate the parent drug and therefore bypass the need for severe reaction conditions and poor yields. (Gomez De Santos *et al.*, 2018) Cut out or append to Drug met section?

1.3.5 The Glitazone Drug Class

The Glitazone drug class is a family of peroxisome proliferator-activated receptor-γ (PPARγ) agonists which are actively used in the treatment of Type- II diabetes mellitus. (Coates *et al.*, 2002). Diabetes mellitus is a disease which can present in two forms, Type I and II. People suffering with type II diabetes have an increased blood sugar (glucose) level, due to poor production of the hormone insulin, or a resistance to insulin resulting in sustained periods of hyperglycaemia.

4. Biotechnological applications of BM3

Mutant libraries from varying labs around the world, such as Arnold, hydrocarbons/ carbene transfer… and Wong (Oxford) flavour and scent compounds… complied mainly single, double, triple and quadruple mutants in several series of directed evolution studies. From these mutants, the following drugs were metabolised and found to produce active human metabolites: testosterone….

Further mutation studies have seen the number of mutations introduced to BM3 increase, For example. Arnold's 139-3 mutant has 12 mutations… These methods were aided by computational methods, Wu et al….. (Kille *et al.*, 2011)

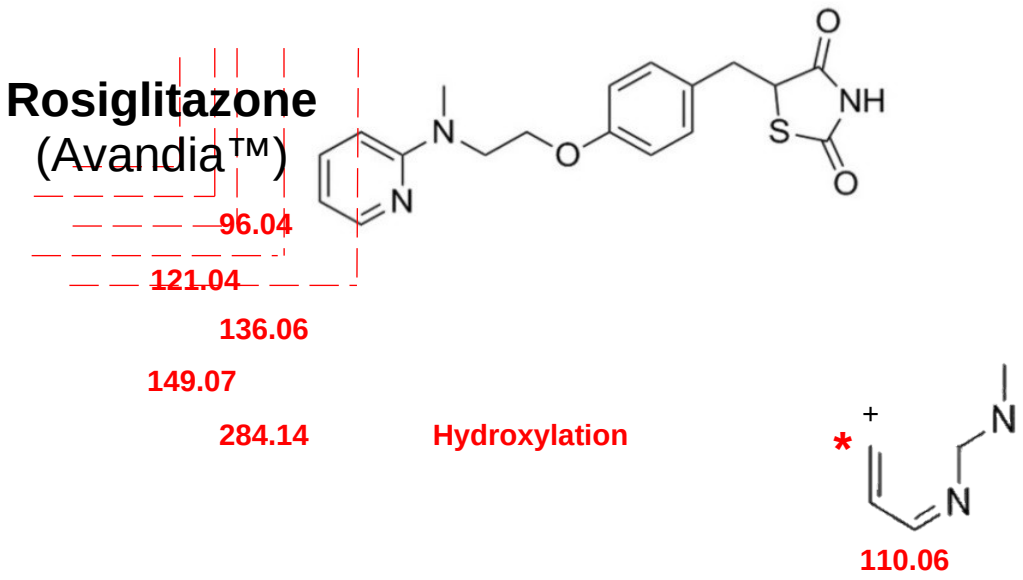
Several studies and literature has been published which demonstrates the use of P450 BM3 for the metabolism of xenobiotics. Owing to the promiscuity and flexible binding conformations belonging to the 'gatekeeper' A82F/F87V double mutant, many structurally diverse drugs are able to bind and induce Type I spectral shifts upon binding. Kd… high catalytic rates… (Jeffreys..)

Further studies by Jeffreys et al. confirmed the binding ability of BM3 DM to bind and rapidly turnover members of the glitazone drug class. The findings suggested…

Evidence from Jeffrys et al. 978 compound large FDA drug library suggested the majority of the structurally diverse compounds that were screened…. % induced a Type II inhibition spectral shift. Most of the inhibitors contained variations on the azole drug class. The N atom of the

azole ring was found to directly coordinate with the heme iron centre upon displacement of the axial water ligand in the resting state. This strong coordination????

Previous work on members of the glitazone drug class, yielded both turnover results, and metabolites. Rosiglitazone (ROSI) for example was metabolised in a similar, time- dependant fashion… (Thislethwaith, 2019)



Hypothesised ROSI fragmentations, shown in Fig.. Provided an underlying rationale for potential fragmentation and metabolism of PIOG and TROG, due to the conserved thiadolazinedione moiety (TDZ). It could be deduced that… It was important to account for the lability of the TDZ ring when considering assay parameters and sample preparation methods.

LC-MS for metabolite identification has had many recent advances such as..., making it a powerful tool, with a wide range of mass spectrometers, with varying strengths.

Pharmacological Applications of P450 BM3

Due to the inherent solubility of bacterial P450s, coupled with the ease of purification and expression in competent expression vectors, makes for a suitable start-point for a multitude of biotechnological applications. In comparison to their eukaryotic and particularly mammalian counterparts, they generally exhibit higher activity and are more stable, allowing for manipulation of their catalytic repertoire.
Furthermore, the wide range of reactions carried out by P450 enzymes allows for the formation of species with would otherwise be difficult to obtain with classical synthetic pathways, such as valuable oxidized species, particularly useful for application in the pharmaceutical industry.

FINISH COMP SECTION

Computational insights into P450 BM3

P450s are dynamic, flexi.. difficulties encountered.... How overcome? Ensembles... etc etc... More examples of docking

PyRosetta Modelling- Computational Protein modelling suite

The use of computational methods alongside both random and site-directed mutagenesis of P450 BM3 has allowed for intrusive studies of binding modes and aided the rapid progress of BM3 for the production of human drug metabolites.(Stjernschantz *et al.*, 2008) The enzyme design pipeline begins with docking of substrates into the active site of P450 BM3

in order to visualise how the substrate would orientate itself within the active site, and to gauge proximity to the heme moiety.

Designed enzymes tend to be less catalytically efficient than the wild-type variants, especially if they are designed towards binding non- natural substrates, even following an iterative design process, the catalytic ability can significantly drop in magnitude. Designed enzymes are likely to be destabilized due to the introduction of one or many mutations, which was a limiting feature in the computational design of enzymes in the past. As the need for more and more selective reactions has grown, so has the need for industrial biocatalysts and the reliance on *de novo* enzyme design has grown exponentially. The basis of computational enzyme design relies on intrusive studies into the structure- function relationships of these enzymes, and how they will tune the catalytic activity, structure, selectivity and efficiency. (Linder, 2012)

Early computational methods were pioneered by software such as Rosetta@Home, based on *in silico* methods and structure prediction models, such as side- chain rotamer optimizations. The initial Rosetta packages were the first to rely on the “theozyme” or theoretical enzyme, when evaluating the structure and redesign of enzyme active sites. Theozymes endeveour to produce a transition- state, stabilized by functional groups, which relies on a computed value for optimized geometry…
(Tantillo, Chen and Houk, 1998)

Pyrosetta is the python interface of the software package, Rosetta, which utilises scoring functions and Monte Carlo method optimizations in order to design proteins.

Pyrosetta does have its limitations as it is not able to model bond breaking or formation… This can be overcome by using a radical- based catalytical model.

Two pivotal publications in the realm of computational enzyme design culminated in successfully designed variants of… for both retro- aldol and Diels- Alder. The basis for which the enzymes were designed took into account the transition states (TS) throughout the multi- step reactions. Beginning with multiple feasible reaction mechanisms, the next step was to identify the protein scaffolds which were able to create the transition states for the desired protein ensembles which correspond to the desired reaction mechanisms, that may also account of any intermediary transition states. To simulate the dynamic active site, parameters such as the orientation and conformation of catalytic residues and internal degrees of freedom of any intermediate transition states were varied in parallel and optimized.

Biomelcular Diels- Alder HOMO LUMO…the Previous retro- aldol involved bond breaking, more to take into account with bond formation as both substrates must be in the correct conformation and orientated such that the desired stereochemical outcome is achieved.

Bot netting- ROSETTA@HOME cluster computer- distributed computing project, to model and predict protein structure.

PyMOL was written by Warren L. DeLano in the accessible python programming language, as an open- source software for protein visualization, alongside other systems. It is now widely available under academic license….

Visualizing proteins in PyMOL allows for docking runs and simulations to be evaluated by by the user. Changes may be applied to the structure, for example, changing residues… (Schrodinger, LLC, 2015.)

Rosetta has even proven to be a powerful tool in the ongoing fight against the novel SARS- CoV- 2 virus, which high- throughput screens, as well as

aiding protein modelling of the first viral protein structures, prior to any data being generated in the lab, allowing for powerful insights into the structure and function into this novel virus. The addition of Rosetta modelling has allowed for accelerated studies into the formation of vaccines…. (Naik *et al.*, 2020) and aided the search for druggable targets… MD and docking studies…

In recent years, Rosetta has evolved to a higher quality of structure prediction, with simple packages and tools to make computational enzyme design user- friendly and time efficient.
(Kaufmann et al., 2010). (Kaufmann *et al.*, 2010)

Objectives of Thesis

The aims of this project were to utilise a double mutant of P450 BM3 to form current and novel major human drug metabolites. Formation and extraction of drug metabolites were to be solely focused on members of the anti- diabetic glitazone drug class.

Following optimisations time- dependant/ bulk incubations with BM3 DM, the metabolites structures were to be determined with LC-MS/MS, coupled with a number of different 1D and 2D NMR experiments.

Due to Covid… aims changed? Here or Covid statement?

When ongoing research ceased due to Covid- 19 interruptions, the aims of the project expanded to learn methods of computational enzyme design, with the aim of designing and utilizing novel and existing packages in Python in order to design BM3 toward the selective oxidation of Pioglitazone, to form a major human metabolite.

Further aims were… method validation… and production of mutants of highest fitness generated within pipeline to be expressed *in vitro* for an end- to end enzyme design pipeline, from docking, semi- rational design, mutagenesis and comparison of mutants with the ‘gatekeeper’ BM3 variant

2 Materials and Methods

2.1 Buffers and bacterial cell strains

Buffer	Composition	pH
Buffer **A**	100 mM Kpi	7.0
Buffer **B**	20 mM imidazole in 50 mM KPi, 350 mM KCl, 10 % glycerol	8.0
Buffer **C**	10 mM imidazole, 50 mM KPi, 350 mM KCl	8.0
Buffer **D**	200 mM imidazole...	

ARE BUFFER B AND C the same and just changed during protocol?

For both WT and DM BM3 pET14b/ pET15b plasmid constructs were provided by Munro Lab.

2.2 Preparation of P450 BM3 WT and DM plasmid stocks

20 µL of NEB- 5 alpha competent *E. coli* (High efficiency) cells (New England Biolabs, UK) were thawed on ice. 1 µL of plasmid DNA (WT or DM BM3 full- length) was mixed with the cells and incubated on ice for 30 minutes. The cell mixture was heat shocked for 30 seconds at 42 °C and removed immediately to ice where they remained for a further 5 minutes. 950 µL of SOC (Super Optimal broth with Catabolite repression) medium

was added to the mixture without mixing. The mixture was incubated for 60 minutes at 37 °C in an orbital incubator (190 rpm). 150 µL of cell mixture was spread on LB agar plates containing carbenicillin (100 µg/mL final concentration). Plates were left to incubate overnight at 37 °C.

All media were autoclaved before use. A single colony was obtained from an agar plate and inoculated into 5 mL LB (Luria- Bertani) medium containing carbenicillin (50 µg/ml final concentration) The culture was incubated for 12-16 hours at 37°C in an orbital incubator shaker platform (190 rpm).

2.3 Plasmid DNA purification from *E. coli* cells

To obtain plasmid DNA for sequencing, 3 mL overnight transformant cell culture for BM3 WT or DM, was pelleted at 13000 rpm in benchtop centrifuge. Manufacturer's protocol was followed as per QIAGEN MINI-prep spin kit (QIAGEN Ltd, West Sussex, UK).

To obtain concentration of DNA extract, the absorbance was taken at 260/280 nm for a 1 µL sample on NanoDrop 2000 with a path length of 1 mm (Thermo Scientific, Wilmington, USA). The purified plasmid DNA was stored at -20°C and used as a stock for subsequent transformations.

2.4 BamHI/ NdeI analytical restriction endonuclease digest

In order to obtain the digested DNA fragments from the expression plasmids in preparation for gel electrophoresis, a BamHI restriction digests was carried out. For the WT/ DM P450 BM3 pET15b P450 full length construct, the restriction endonucleases, BamHI and NdeI, were added to a clean tube to form reaction mixture A, as shown in Table..:

Reaction Mixture **A**: made up to 20 µL
1 µL DNA (concentration 1 µg/µL) for WT and DM
15 µL H_2O (sterile)
2 µL 10X buffer
1 µL BAMHI
1 µL NdeI

Table… Reaction mixture **A** for restriction digest containing restriction endonucleases BAMHI and NdeI.

The reaction was incubated at 37°C for one hour. The digestion was stopped by the addition of 10 mM EDTA after one hour. The digested DNA fragments were then separated and visualised by agarose gel electrophoresis .

2.4.1 DNA agarose gel electrophoresis

A 0.8 % agarose gel was prepared by melting 0.4 g agarose in 25 mL 1X TAR buffer. 50 mL TAE running buffer was diluted from a 50X stock solution (242 g/L Tris- base, 57.1 mL glacial acetic acid, 100 mL 0.5 M EDTA, pH 8.0) Agarose gel was allowed to cool before the addition of SafeView (NBS Biologicals, Huntingdon, UK) for a final concentration of 0.5 µg/mL. Loading samples were prepared by mixing 10 µL sample with 2 µL of 6X TriTrack DNA Loading Dye. 10 µL DNA ladder (Thermo Scientific GeneRuler 1 kb DNA Ladder, Thermo Scientific , UK) was loaded for reference. Electrophoresis was run in 1 x TAE running buffer. The gel was run on 80 V for 45 minutes and visualised on…??

2.5 Expression of full- length P450 BM3 WT and DM

2.5.1 Transformation of Plasmids into competent E. coli cells (BL21 (DE3))

Plasmids were transformed into BL21 (DE3) competent *E. coli* cells and expressed in Terrific broth (TB) media and TB auto- induction media (Formedium, Hunstanton, UK) and 10% glycerol.

2.5.2 Day culture and auto induction of P450 BM3 WT/ DM full

BM3 WT or DM full- construct plasmids were transformed into BL21 (DE3) competent *E. coli* cells (New England Biolabs, UK) as per the manufacturers protocol.

LBA (Luria- Broth Agar) was melted and carbenicillin (100 µg/ml) was added when cool. LBA plates were poured and left to set.

Cells were mixed by inversion post incubation. 50 µL and 150 µL of culture were spread evenly. Plates were incubated in a stationary 37°C incubator overnight.

A single colony was picked from the freshly streaked plates and inoculated into 5 mL overnight culture with the relevant antibiotic, incubating at 37°C with shaking at 180 RPM.
The full 5 mL overnight culture was used to inoculate a 150 mL day culture of sterile LB media in a conical flask containing 50 µg/ mL carbenicillin. The mixture was incubated at 37°C with 180 rpm shaking for 3 hours.

2 L conical flasks were prepared with TB autoinduction media (Formedium, Hunstanston, UK). Media was made up with 500 mL distilled water, sterilised by autoclaving. 10 mL day culture was added alongside 2 mL sterile glycerol and 50 µg/ mL ampicillin. The flasks were incubated at 37 °C with 180 rpm shaking until OD_{600}= 0.6 was reached. The temperature was reduced to 25 °C and 100 µM δ-aminolevulinic acid (ΔALA) heme precursor (from a 100 mM stock prepared in ultrapure water and filter sterilised) was added.

After a further 24 hours the bacterial cells were harvested by centrifugation with a Beckman Coulter Avanti J- 26 XP centrifuge and a Beckman JLA8.1 rotor (6000 rpm, 10 minutes, 4 °C). The observed pellet colour was noted and the pellet was resuspended in Buffer B. The suspension was centrifuged again (6000 rpm, 10 minutes, 4 °C). The pellets were weighed and frozen at -20 °C until cell lysis and purification was carried out.

2.6 Lysis and Purification of WT and DM full length

The cell pellet was thawed on ice and gradually resuspended in ice- cold buffer B. Protease inhibitor cocktail (Roche Applied Science, Welwyn Garden City, UK) 1 EDTA- free cOmplete™ tablet per 100mL, 10 ug/mL DNase (Merck, Nottingham, UK). The cell suspension was then lysed by sonication on ice (35 % amplitude, 12 x 10 second pulses. 50 seconds between pulses, for 30 minutes, Bandelin Sonopuls). The cell lysate was clarified via centrifugation (18000 RPM in JA 25.50 rotor in a Avanti J Series Beckman centrifuge?) for 30 minutes at 4 °C. 30 % w/v ammonium sulphate was added to the supernatant stepwise, with gentle mixing over an hour at 4 °C. The mixture was incubated for a further 30 minutes with gentle mixing at 4 °C. Precipitate was then removed via centrifugation

(4,600 g in JA 25.50 rotor) for 30 minutes. A sample of both pellet and supernatant were taken, the supernatant was retained.

2.6.1 His- tag purification with Ni- IDA column

Both the full- length WT and DM P450 BM3 proteins contain C- terminal His- tag and so purification was via NI- IDA column. Ni- Ida resin (10 mL) was pre-equilibrated with phosphate Buffer A prior to incubation with the ammonium sulfate purified cell extract (iminodiacetic acid chelating group, Ni resin source) overnight at 4 °C with agitation.

The protein mixture bound to the Ni- IDA resin was packed into a column and washed under gravity with the following column volumes of phosphate buffer A with the addition of an imidazole concentration gradient: 20 mM (8 column volumes), 50 mM (4 column volumes), 200 mM (1.5 column volumes). The protein was eluted with the final 200 mM imidazole and phosphate buffer B elution. Each fraction was retained for analysis by SDS- PAGE. Protein concentration and purity, pre- dialysis, was determined by both SDS- PAGE and UV- Vis spectroscopy. The final elution was dialysed against phosphate Buffer A, at 4 °C.

Post dialysis, the protein was concentrated in a Vivaspin 20 centrifugal concentrator 100,000 Da molecular weight cut- off, Sartorius Vivascience, Epsom, UK). The final concentration and purity of the protein was determined with UV- Vis and SDS- PAGE. Fractions were retained throughout the lysis and purification steps and were run alongside 250 kDa molecular weight Precision Plus Protein™ Prestained protein ladder on a 10% SDS- PAGE Mini-PROTEAN TGX Stain- Free Precast gel (Biorad) in order to determine purity in 0.1 % SDS Tris Glycine running buffer, using a Biorad generator (20 minutes, 400 V).

UV- Visible spectra were recorded for both BM3 WT and DM to determine Rz values for the concentration of active, purified protein, its

concentration and spin- state. Both BM3 WT and DM were diluted 100- fold in a UV- Vis cuvette to a final volume of 1 mL. A Cary 50 UV- Vis spectrophotometer (Agilent Technologies Ltd) using Scan application 3.00, a path length of 1 cm and a scan range set for 800- 200 nm. In order to determine Rz value, the following calculation was carried out using the absorbances at both 418 and 280 nm. A_{418}/ A_{280}= Rz ratio was used as a measure of protein. Concentrations of the active heme of both WT and DM were determined by using A418 and calculating concentration based on the Beer- Lambert equation with the extinction coefficient of ε418 = 105 mM^{-1} $cm^{-1.}$

Both WT and DM proteins were flash frozen in liquid Nitrogen in 150 µL aliquots and stored at -80 °C until the week of assay.

Prior to use protein was passed through a 10 mL Lipidex 1000 column (Perkin Elmer, UK) when in order to remove any fatty acids bound in the active site. The Lipidex column was prepared by washing with hot methanol, water and equilibrated using buffer D. 1 mL of protein was loaded to the column and incubated at 4°C for one hour. The protein was eluted, ready for use in subsequent assays.

2.7 Preliminary compound tests for reaction monitoring assays

2.7.1 Compound solubility determination

Solubility limits were tested for all compounds and a suitable DMSO percentage was determined (<5 %) with serial dilutions in Buffer A. Samples were warmed if necessary to aid compound into solution.

Solid- phase extraction tests were also carried out. Compound solution was made up to concentration and percentage DMSO as determined by

previous solubility checks. The 1mL StrataX- SPE (Phenomenex, USA) column or 96- deep well plate was prepared as per the manufacturer's protocol. The column/ wells were primed with 10mL MeOH then 5 mL water. The compound sample was run through the column, washed in D_2O and dried completely under nitrogen before elution with DMSO-$_{d6}$. The sample was analysed by ^1H NMR to determine whether a sufficient amount of compound was collected.

2.7.2 Qualitative Initial rate assay

An initial assay was run in order to determine the activity and turnover of compound with BM3 DM and WT. All measurements were performed in triplicate in order to determine turnover number (mM/min)or rate? and to carry out statistical analysis to determine significance between ligand incubations and controls.

All measurements were made on a Cary 50- UV- Visible spectrophotometer (Agilent, Stockport, UK) set to the Kinetics measurement with a path length of 1 cm and regulated temperature at 30 °C with a single peltier cuvette holder. A solution of NADPH was made up in Buffer A to a concentration of 10 mM. A 200 mM stock solution of compound in DMSO was diluted to 10 mM in Buffer A. NADPH was added immediately prior to beginning measurement mixed. The spectrophotometer was blanked with buffer and the parameters were set to measure absorbance at 340 nm for 1 minute. All measurements were taken in triplicate in a quartz cuvette. Rates were calculated by determination of initial gradient with straight line fitting, and calculations were carried out using the Beer- Lambert Law, with NADPH ($\varepsilon = 6.22\ mM^{-1}\ cm^{-1}$).

10 µL NADPH (10 mM) was added to 990 µL buffer (100 µM final concentration) in a cuvette, mixed rapidly and measured. The following controls were also measured. 10 µL of chosen compound (10 mM) was

added to the corresponding amount of purified BM3 DM to make up a final concentration of 1 µM, the mixture was made up to 1 mL with buffer and reaction was initiated with 10 µL NADPH (10mM) and measured in triplicate. Which one??

2.7.3 LC- MS/MS standard test

LC- MS/MS standard was run with parent compound to determine optimal gradient and where compound elutes, on a Q Exactive Plus instrument, fitted with a heated electrospray ionisation source and a U3000 UHPLC (Thermo Fisher Scientific) was used for full scan UPLC- MS/MS analyses of each drug (scan range 50- 1000 m/z). A Hypersil Gold column (100 x 2.1 mm, 3 µm particle size, Thermo Fisher Scientific) was used.

2.8 Elucidation of BM3 DM metabolites by LCMS/MS and ^{1}H/ 2D NMR

2.8.1 LCMS/MS and ^{1}HNMR method for structural elucidation of BM3 DM metabolites

Previous methods postulated a sequential metabolite formation by BM3 DM (Jeffreys, 2018). A fully optimised method with separation and full structural elucidation via LCMS/ MS and ^{1}H NMR was used to derive metabolites of the clinically relevant drugs: Pioglitazone (PIOG) and Troglitazone (TROG) (Thistlethwaite, 2019).

Stock solutions of the components in the electron regeneration system (detailed above) were prepared and diluted to required concentrations in Buffer A. BM3 DM (5 µM), alongside the electron regeneration system (7.76 mM glucose- 6- phosphate, 0.75 U/mL glucose- 6- phosphate

dehydrogenase and 0.6 mM $NADP^+$)were made up in buffer B in a 250 mL conical flask, to give a final volume of 60 mL. The control incubations were individually prepared, with component missing from each to give a 5 mL final volume. The assay incubation and the control incubations were pre-warmed for 5 minutes in an HT Infors incubator (60 rpm, 37 °C). Initiation of the reaction monitoring assay was initiated by 75 µM and 100 µM of Pioglitazone and Troglitazone respectively, made up according to initial solubility limit checks. DMSO was added to the control incubations with 0 µM ligand.

At the following timepoints: 0, 1, 2, 3, 4, 5, 10, 15, 20, 30, 40, 60 and 180 minutes, 5 mL aliquots were taken and reactions stopped by the addition of 200 µL HCl (37 %). Control incubations were also halted with the addition of 200 µL HCl (37%) after 180 minutes. The quenched samples where neutralised with saturated NaOH solution to pH 7. The neutralised samples where then centrifuged in an Eppendorf 5810R centrifuge (4000 rpm, 4, 20 minutes). The supernatant was retained for sample extraction for analysis via LCMS/MS and NMR.

2.8.2 Preparation for UPLC- MS/MS analysis

5 mL of each sample was passed through a protein precipitation plate, (StrataX (Phenomenex, USA) and eluted in 200 µL ACN. Samples were frozen in insert HPLC vials for later analysis via LCMS/MS.

Samples were analysed on a Q Exactive Plus instrument, fitted with a heated electrospray ionisation source and a U3000 UHPLC (Thermo Fisher Scientific) was used for full scan UPLC- MS/MS analyses of each time point and control sample (scan range 50- 1000 m/z). A Hypersil Gold column (100 x 2.1 mm, 3 µm particle size, Thermo Fisher Scientific) was used for PIOG and TROG. For each sample, 5 µL of sample was injected per run at a flow rate of 0.5 mL/ min with a collision voltage of 75 C at 40 °C. Full gradient separation of each metabolite was optimised initial assay

preparations. Piog and Trog were separated in a positive ESI mode using linear ESI gradient and run time: SolvA H2O + 0.1% FA SolvB ACN + 0.1% FA over 15 with a hold of 2 minutes at 100 % SolvB, then returning to 10 % SolvA over 10 minutes. All data were processed by Thermo Scientific Xcalibur™ 3.1.66.10 (Thermo Fisher Scientific, UK). Any significant peak areas were analysed, extracted and graphically represented in OriginLab Pro 9.0 (OriginLab Corporation).

2.8.3 Preparation of samples for analysis via NMR

All samples to be analysed via NMR were prepared by solid phase extraction (SPE) with Strata-X 33 µm, 10 mg/mL Polymeric Reversed Phase columns (Phenomenex, Macclesfield, UK). All columns were primed as per the manufacturer's protocol, 10 mL HPLC grade MeOH and 5 mL water (Chromasolv). 5 mL of sample was applied to column and washed with 2 mL deuterium oxide before thoroughly drying under nitrogen then eluting DMSO-$_{d6}$. All samples were eluted in 0.7 mL DMSO-d6, into HPLC vials and later transferred to NMR tubes. TSP (20 µM final concentration) was added to each sample before extraction as an internal standard to determine extraction efficiency All NMR spectra ($^{1}H^{13}C$-HSQC and 1H13C-HSQC-TOCSY, COSY, ROESY) were recorded at 298 K on a Bruker AVIII 500 MHz spectrophotometer with Z- gradients provided by a $^{1}H/^{13}C$ QCI cryoprobe. All 1D NMR spectra were recorded using a 1D ^{1}H NMR method with preset presaturation profiling. All 2D NMR spectra were recorded using a HSQC multiplicity method with the addition of gradient selection and adiabatic carbon pulses. Solvent and water suppression were applied. All data were analysed via MestReNova 11.0.5 (Mestrelab Research).

2.9 Computational enzyme design pipeline ‘enz’

PyRosetta simulations

Enz- link github, front end pyrosetta/ vina use enz for ala sca, using results we rational design… what mutants… script in supplementary info… paste in? or .py files attached… link to git repo…

NW align, clean..
Mutants generated from enz and expressed and purified- turnover?? quantitative??

Preparation of PDB files

Files for the structures of WT and DM BM3 were downloaded in protein data bank (pdb) format from RCSB.org (Research Collaboratory for Structural Bioinformatics). PDB files were in complex with molecules such as omeprazole (Butler et al???), contained water molecules and were in the dimer conformer, in order to use the pdb files for docking it was necessary to clean the structures. BioPandas… Any file conversions made were readable by PyMOL…

In order to…. an experimental python package was developed… to… KEY WORDS… which simplifies the interface between user and wraps PyRosetta and VINA… for intrusive docking studies…

Protein object from seq… (clean, NW alignment, define AS resi, VINA energy minimised…refold… What does energy minimise do…?

Accuracy of docking… RMSD comparison to known structures… future work…. For validation of protocols…

Side chain repacking function exists within enz.. based on…

Docking in *enz*

custom score function…. based on distance between target C of substrate and heme Fe (selective -OH), dateframe…

Weighted by Vina Docking score…. (kcal/mol?????? CHECK) Soft Max??? function to.. scale E 0-1, eqn… binding energy of each pose, tighter poses weithed more heavily

Rosetta paper.. looop remodelling (Future work point!!) alpha fold

Pyrosetta design

PyRosetta version… Only compatible for one active site region. The dimeric BM3 structure was converted into its monomeric form in the structure cleaning process to comply with subsequent PyRosetta design steps.

Computational Alanine scan

Methods of the computational alanine scans were based on protocols designed by Kortemme et al., within Baker lab. (Kortemme, Kim and Baker, 2004)

An input of the cleaned pdb file of BM3 WT (3BEN?)(P) and the canonical SMILES string for Piog (L) was used. The interface between P and L were defined as…. The alanine scan computed the changes in the binding free energies ($\Delta\Delta G_{binding}$) in comparison with the WT.

"reductive mutations- intro….."

CSF- custom score function…. Based on….

Semi- rational enzyme design based on Arnold BM3 mutant library

Arnold mutant library… picked mutations semi- rationally…. (Testosterone paper…) dock in enz, evaluated any significant changes in score function/ poses…

2.6 Site- directed mutagenesis for the introduction of novel mutations

Polymerase Chain Reaction (PCR) is a commonly used method of introducing mutations and amplifying DNA sequences. Two techniques for PCR were carried out but the base components remained the same. Template plasmid, the necessary primers, deoxynucleoside triphosphates (dNTPs) and a DNA polymerase were used. INTRO

Primers were designed with different tools for each kit

Q5

QuikChange Mini

CloneAMP HIFI containing the appropropriate components….

Step Temperature (°C) Time (s) Number of cycles Initial denaturation 95 180 1 Denaturation 95 30 Annealing 57 30 25 - 35 Extension 72 60 Final extension 72 480 1 Pause 18 Pause 1

The…gene was mutated using the standard Q5 (NEB) or QuikChange (Agilent..) mutagenisis protocol for the introduction of two single point mutations. The P450 BM3 DM constructs were provided by… The individual primers were synthesised by Eurofins (…) and designed either by Benchling (…) for the Q5 kit, and with the OuikChange primer design tool (Agilent Technologies…) for the QuikChange Mini kit.

Mutation Introduced	Oligonucleotide Primer Sequence (5’ to 3’)
…188S in A82F/F87V Construct (TM1) FL	
…75W in A82F/F87V Construct (TM2)FL	
…188S in A82F/F87V Construct (TM1H) Heme	
…75W in A82F/F87V Construct (TM2H) Heme	
…188S/…75W in FL A82F/F87V Construct QMFL	
…188S/…75W in Heme A82F/F87V Construct QMH	

The PCR product from the Q5 kit required a KLD reaction mix in the provided buffer.

The PCR product from the QuikChange kit was transformed straight into BL21 Gold (DE3) competent cells.

Point mutations were validated by high-throughput Sanger sequencing performed at Eurofins..... (.....) using VecPR, SeqPR1, SeqPR2, SeqPR3 and T7F sequence primers.

5'-tctggatttgctcgctgcgacttgttcattgcttcatcca-3'
5'-gtctccgaaaaaatcacgtacaaatttccacgcttgacttaagtttttatcaaagcg-3'
5'-tggatgaagcaatgaacaagtcgcagcgagcaaatccaga-3'
5'-cgctttgataaaaacttaagtcaagcgtggaaatttgtacgtgatttttttcggagac-3'

Oligonuclotide Name	Oligo Primer Sequence (5’ to 3’)
VecPR	tacccacgccgaaacaag
SeqPR1	tgagccgcttgatgacgagaac
SeqPR2	tcgcaacgcttgattcac
SeqPR3??????	did i use?
T7 (forward)	taatacgactcactataggg

3 Results and Discussion

3.1 Protein Expression and Purification of Full Length WT and DM BM3

3.1.1 Preparation of Plasmid DNA Stocks (WT and DM)

Plasmid constructs for both BM3 full- length WT and DM were provided by the Munro Lab. Transformation attempts for the preparation of personal stocks were successful and yielded colonies.

Following the preparation of plasmid DNA stocks from existing WT and DM constructs, sequencing data and gel restriction digest data confirmed the correct DNA sequences were present within the pET14b vector for both WT and DM, and that for the DM, the point mutations existed: A82F/F87V, in preparation for expression and purification of full length WT and DM BM3.

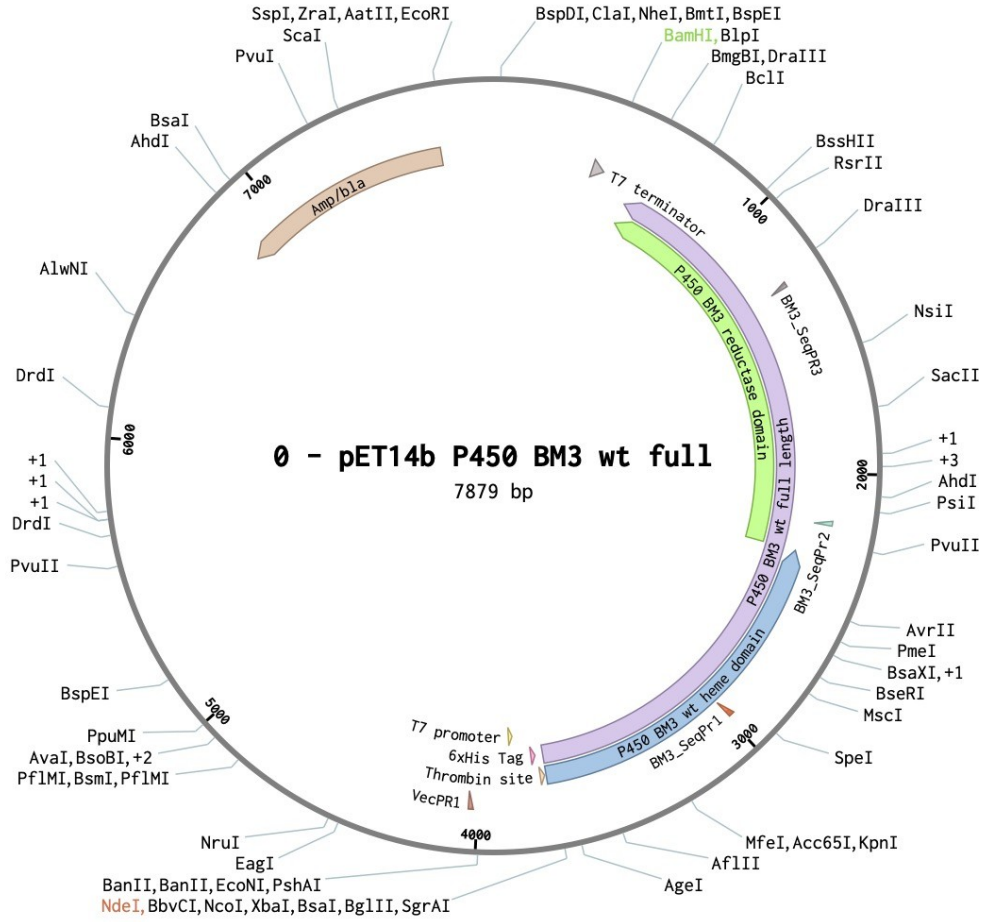


Fig... Plasmid map of P450 BM3 WT full length with labelled heme and reductase domains, and primer binding sites. Cut sites for BamH1/ NdeI digests are highlighted in green and orange respectively. Figure generated in Benchling.

3.1.2 BamH1/ NdeI restriction digest

BamH1 is a restriction endonuclease capable of recognising short DNA sequences (6 bp), which allows for specific cleavage at restriction sites. This particular restriction enzyme binds at the 5'-GGATCC-3' recognition sequence. One BamHI recignition sequence is found.

Add fig

3.1.3 BamH1 restriction digest

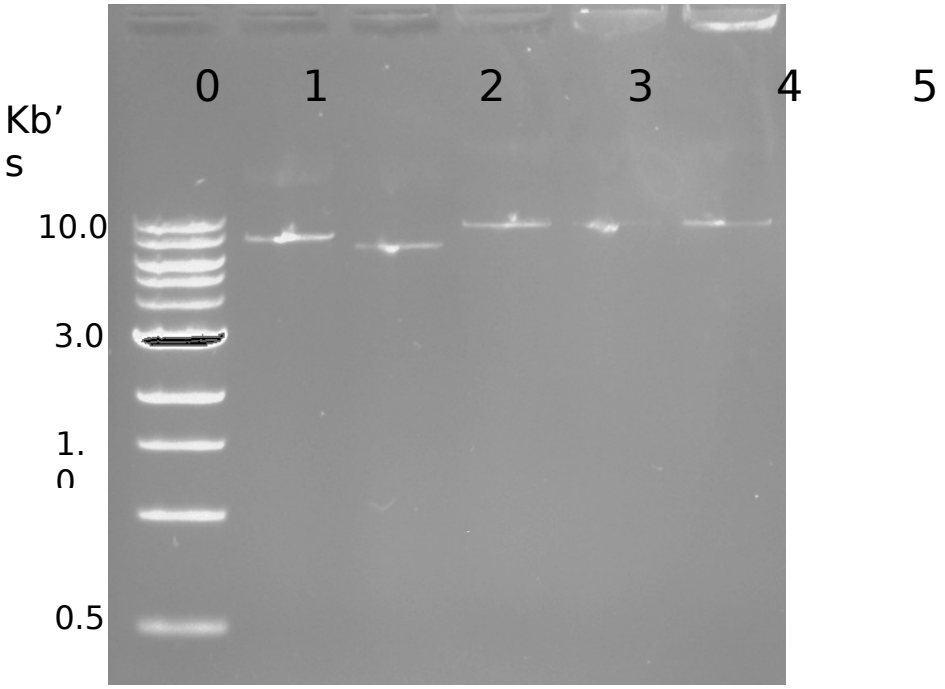


Fig..? DNA agarose gel of the BamH1 full restriction digest with 1 kb DNA ladder. The samples were loaded as follows... pET15/14 b NEED TO FIND IMAGE...

Figure 3.2: Restriction digest of the full length pET14b construct. Uncut plasmid (1) migrated more slowly than the NdeI/BamHI digested plasmid DNA (2), The samples were run in 0.8 % agarose with a GeneRuler 1 kb DNA Ladder (0), ladder bands indicated in kbp. The samples are as follows: (0)DNA ladder, (1) BM3 WT heme pET14b, (2) BM3 DM heme pET14b, (3) WT full length pET14b, (4) DM full pET14b, 5 WT heme pET20b.

3.1.3 Improving Protein Expression

Following cell harvesting (as described in Section…) cells were lysed and centrifuged. This process allowed for separation of insoluble cellular debris, from the necessary P450 protein. The use of the 35 % w/v ammonium sulphate cut

During expression initial expression and purification, low Rz values were observed for both
WT and DM full length domains, which is indicative of low protein purity. This was visualised when analysing the UV-Vis absorption spectra of WT and DM BM3, throughout each stage of protein purification. Samples were also retained at each stage of purification for visualisation on a protein gel. Initial expression attempts resulted in aggregation of apoproteins and unincorporated heme and reductase domain. This is clearly seen in Fig.. where in which a dark band appears at… The factors effecting BM3 heme incorporation has been researched thoroughly, with suggestions that temperature control and induction with delta ala must occur within narrow time margins.

Further to more precise grow up attempts, an expression trial was carried out with varying concentrations of d-ALA from the literature value of 100 mM and concentration above and below. The highest concentration of d-ALA, 200 mM, produced a pellet which was visually more red than that of the lower concentrations, however no significant increase in Rz was seen (Fig…) this may be accounted for by increased expression of other *E. coli* hemoproteins rather than increased incorporation of BM3 heme so would have no overall effect on protein activity.

Following protein expression, His- tag purification with Ni- IDA resin (Ni-iminodiacetic acid) was carried out to purify both BM3 WT and DM.

Due to the promiscuous nature of BM3 DM, it was often necessary to carry out two incubations with Lipidex resin to insure that most, if not all, of the protein resided in the unbound, low- spin state. UV- Vis spectra of WT and DM are shown in Fig… where the concentration of active heme and Rz values were determined.

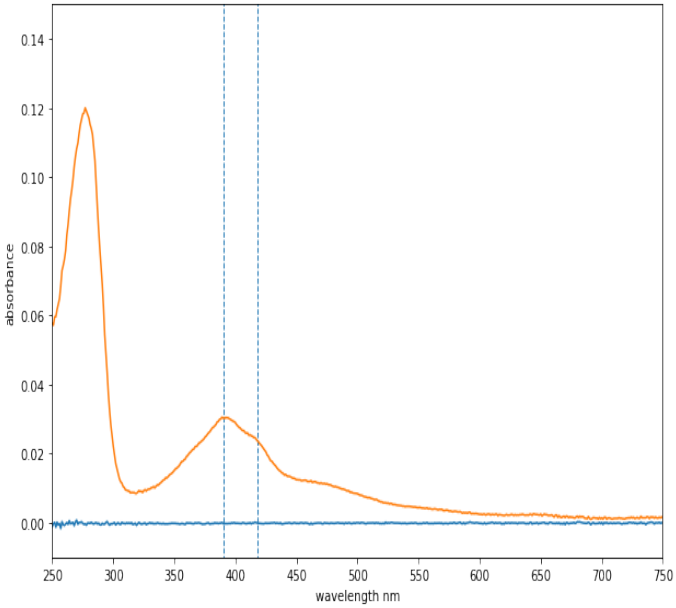


Fig... Example UV- Visible spectrum of full- length BM3 DM, post incubation with Ni- IDA The sample was diluted 100- fold in Kpi. The spectrum indicates the mixed- spin state of BM3 DM with vertical, dashed lines at 390 and 418 nm respectively. Figure generated in matplotlib.

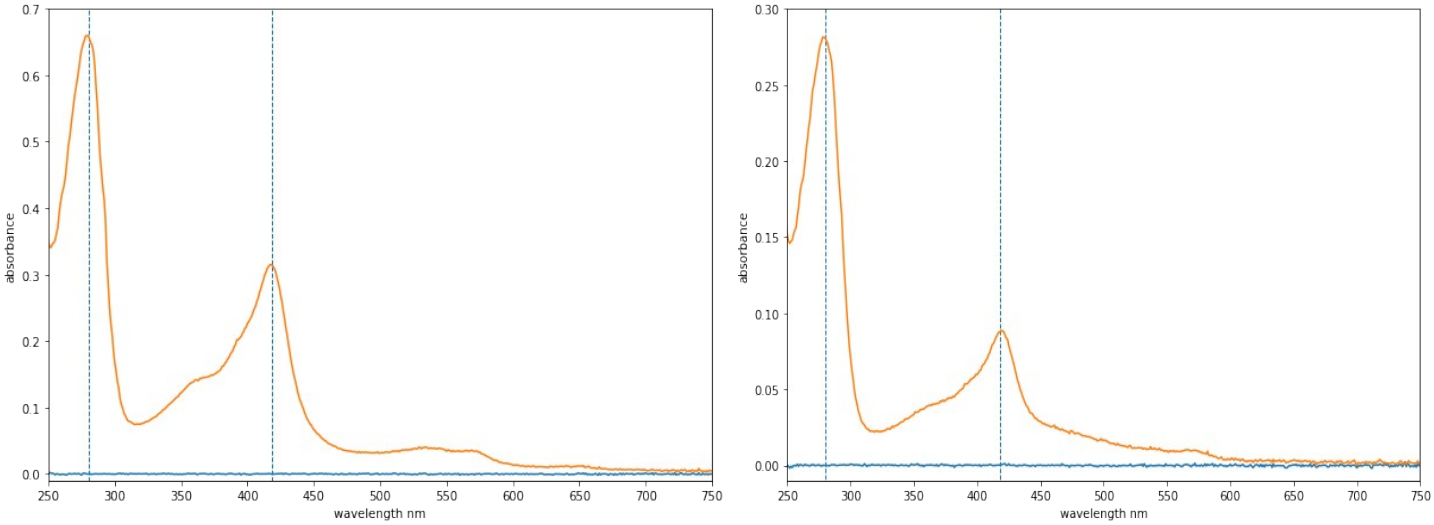


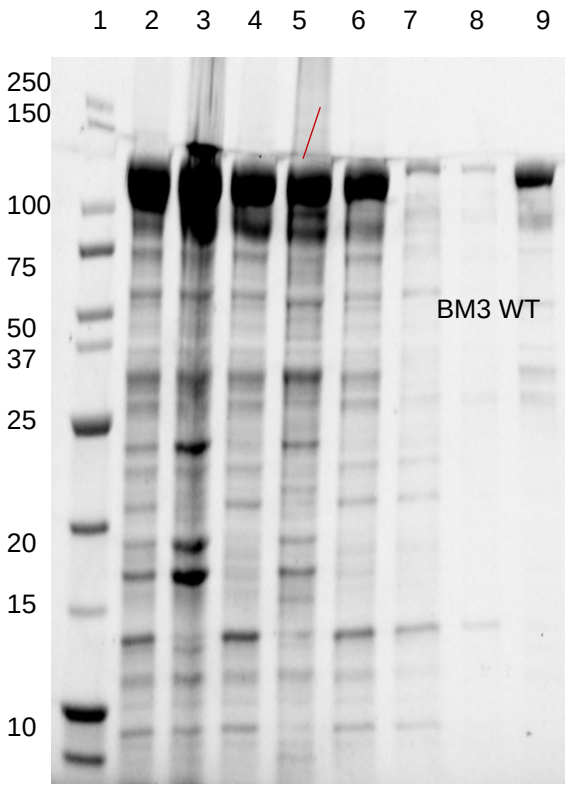
Fig... Examples of UV- Visible spectra of full- length BM3 WT and DM respectively, following purification steps with Ni- IDA, dialysis and Lipidex-1000. Each sample was taken post concentration and diluted 100- fold in Kpi. A_{418}/ A_{280}= Rz was used as a measure of protein purity. In both the BM3 WT and DM spectra, the Soret peak is present at 418 nm, indicative

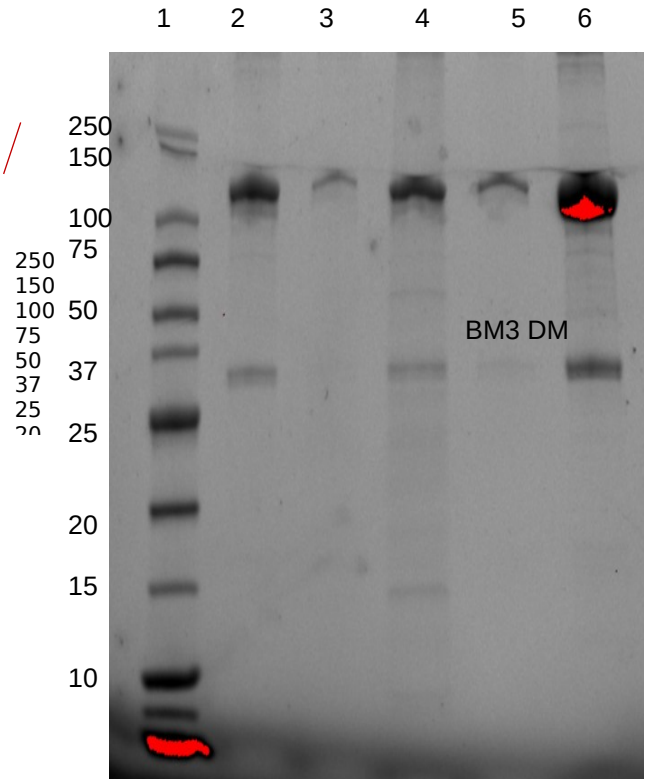
of BM3 in its low- spin state. The dashed vertical lines indicate 280 and 418 nm respectively. Concentrations were calculated using the Beer-Lambert equation. For WT, A_{280}: 0.664, A_{418}: 0.328, Rz: 0.49, Concentration: 0.467 mM. For DM, A_{280}: 0.257, A_{418}: 0.0943, Rz: 0.367, Concentration: mM. Figures generated in matplotlib.

Mixed spin...

Rz values were calculated by A_{418}/ A_{280}, as a determination of protein purity

Lipiex-1000 resin was used to remove any fatty acids bound to the active site of, once again due to the promiscuous binding ability of the DM.





Fig… Post lysis and purification SDS- PAGE gels of WT BM3 full length and DM respectively. The wells for BM3 WT contain as follows: (1) 250 kDa Protein ladder; (2) lysate supernatant; (3) lysate pellet; (4) NH_4OH supernatant; (5) lysate pellet; (6- 9) imidazole elution fractions, 10, 20, 50, 200 mM. BM3 DM full: (1) 250 kDa Protein ladder; (2, 4, 5, 6) imidazole elution fractions, 10, 20, 50, 200 mM. Well (3) was missed out. The resolved bands at ~119 kDa are labelled for WT and DM BM3. Other bands may *E. coli* proteins with affinity for nickel and cleaved BM3 domains.

~cleaved heme…… 50…

3.1.4 Expression Conclusions

Several successful rounds of expression and purifications of BM3 full-length WT and DM were carried out

It can be concluded from the SDS- PAGE gels of both the WT and DM (full-length domains) were successfully expressed, with resolved bands at ~119.5 kDa in the 200 mM imidazole elution, following Ni- IDA purification. Upon inspection of the UV- Vis spectra following each purification step, a suitable purity of protein was reached in order to carry out turn over reactions. Care was taken in subsequent expressions and growths to ensure a higher Rz was reached. This was achieved via a combination of induction with D- ALA at the critical OD = 0.6 and ensuring that the temperature was decreased following induction.

From Fig.. it can be seen that bands appear ~…. Which is indicative of unincorporated heme and/or reductase?... due to????

3.2 Preliminary assay investigations

3.2.1 Compound solubility limits

Serial dilutions of each compound (PIOG and TROG) were carried out in order to determine solubility limits in DMSO. PIOG solubility reported as approximately 2.5 mg/mL and TROG has been reported as 44.15 mg/mL.

National Center for Biotechnology Information. PubChem Compound Summary for CID 5591, Troglitazone. https://pubchem.ncbi.nlm.nih.gov/compound/Troglitazone. Accessed Jan. 4, 2021.

National Center for Biotechnology Information. PubChem Compound Summary for CID 4829, Pioglitazone. https://pubchem.ncbi.nlm.nih.gov/compound/Pioglitazone. Accessed Jan. 4, 2021.

3.2.2 SPE extraction tests

SPE extraction tests were carried out for each drug to ensure columns were not blocked, samples were adequately retained on column and to log any impurities that may be carried through in SPE extraction procedure.

3.2.3 Qualitive initial rate assay

In order to determine whether BM3 DM was able to turnover PIOG, a qualitative initial rate assay was carried out and run in triplicate for each incubation and control.

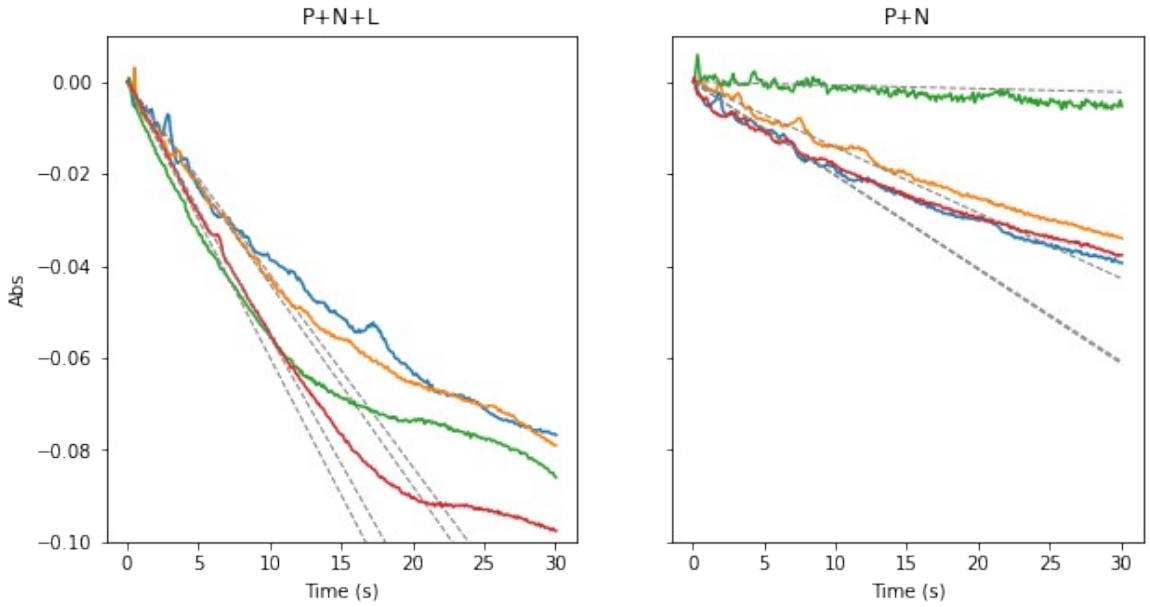


Fig.. UV- Visible traces of Absorbance vs Time for initial turnover investigations. P+N+L is the assay initiated with the ligand (PIOG) in order determine if BM3 DM was able to turnover PIOG in comparison with control, P+N, which was an incubation of BM3 DM with NADP(H) only. Gradients, displayed by dashed grey lines, were used in order to carry out statistical T- testing, in which p- values were determined to show significance between the ligand- initiated incubation and the control.

Quantification of NADP(H) consumption in steady state kinetics relies on the assumption that... and any related calculations involved absorbance at 340 nm?

Rate comparison between P+N+L and P+N reveal... a higher a rate of consumption with the ligand as seen.. and therefore can be can be

determined that BM3 DM was turning over Piog (p=….) …. As… Statistical analysis was conducted using…. And values displayed…..

The graphs in Fig… show the… The displayed gradients were used in order to work out corresponding p- values to determine significance and to calculate turnover number with the following equation:

Which t- test… students?
Significance
Talk about p- test…. 0.05…

Add notebook to supplementary…?????

3.2.4 LC- MS/MS standard tests

Standard stock solutions were made up to 200 mM for both PIOG and TROG, which were diluted as needed for use within assays.

It was seen from the initial standard LC-MS/MS chromatogram that there were issues with the sample of Trog. This is clear due to the lack of parent ion peak within the standard. Literature reports of photolysis may have been a cause of breakdown. Furthermore, the presence of unidentified ion peaks upon analysis of the fragmentation can be indicative of a contamination within the standard or internal phenacetin standard.

3.2 Drug Metabolite elucidation with LC- MS/MS and 1D and 2D NMR techniques

Previous studies by Jeffreys et al. showed a range of antidiabetic compounds binding tightly to BM3 DM (A82F/F87V). Binding studies demonstrated substantial spin- shifts in the heme iron spectra, exhibiting substrate- like binding. These studies led to formation of oxidized metabolites using full- length BM3 DM, with crystal structures alongside other analytical techniques, allowing for metabolite identification (FIG???)..

An investigation into the binding ability of the two anti- diabetic drugs, Pioglitazone and Troglitazone, were carried out using different spectroscopic methods, alongside monitoring the metabolites formed with P450 BM3 DM, with full structural elucidation provided with liquid chromatography tandem- mass spectroscopy (LC-MS/MS) and a combination of NMR (nuclear magnetic resonance) experiments such as HSQC (Heteronuclear Single Quantum Correlation), ROESY (Rotating frame Overhause Effect SpectroscopY), COSY (homonuclear COrrelation SpectroscopY) and TOCSY (TOtal Correlation SpectroscopY).

3.3.1 LC-MS/MS for the elucidation of metabolites

To ensure the full- length construct of BM3 DM turned over and produced metabolites of both pioglitazone (PIOG) and troglitazone (TROG), it was necessary to analyse assay products to deduce what oxidation products (hydroxylation or otherwise) were formed and where on the molecule the changes had occurred. LC-MS/MS was employed as an analytical tool to determine what metabolites were formed. Following analysis of the LC chromatogram of the drug standard, against the control incubations, it was possible to determine if the drug was metabolised or broken down due to other reasons, such as light- sensitivity, or reaction with one of the

cofactors...Timepoints, and/ or bulk incubation products with intact BM3 DM, were analysed via LC-MS/MS following *in vitro* turnover assays, with PIOG metabolites separated by time, and TROG metabolites formed in a bulk incubation.

3.3.2 ^{1}HNMR and 2D NMR for the elucidation of metabolites

NMR was employed as another analytical technique alongside LC-MS/MS, as though the MS data would indicate the presence of the intact standard drug, metabolite peaks, as well as allow comparison of the controls, NMR indicated where modifications, such as hydroxylations, were occurring. NMR analysis requires a larger quantity and concentration for analysis, this was therefore taken into consideration when optimising assay parameters, and so a bulk reaction volume of 60 mL was used.

3.3 Pioglitazone

Pioglitazone was the first marketed member of the thiazolidinedione (TZD) drug class, for use as an oral antidiabetic agent. The TZD drugs are also known as “insulin sensitizers”, and can be used in combination with other diabetic drugs, or alone, for the treatment of diabetes mellitus type II. As with other members of the TDZ drug class, pioglitazone is able to stimulate nuclear receptor peroxisome proliferator- activated receptor gamma (PPAR-γ) selectively (Sakamoto *et al.*, 2000).

Pioglitazone is known as an “insulin sensitizer” is due its action on improving sensitivity towards insulin in muscles and adipose tissues, allowing for improved glycaemic control for those with Type 2 diabetes. This inhibits routes of hepatic gluconeogenesis, a metabolic pathway which generates glucose from non- carbohydrate carbon sources.(Zhang

et al., 2019) Gluconeogenesis is a critical for glucose homeostasis and is regulated by many features, for example, hormone secretion.

In terms of the pharmacokinetics of pioglitazone, it is extensively metabolised by hepatic routes via hydroxylation and oxidation, with four reported primary metabolites (M-I, M-II, M-IV, M-V) and two secondary metabolites (M-III and M-VI). The major metabolites found in human serum are M-III and M-IV (Fig…).

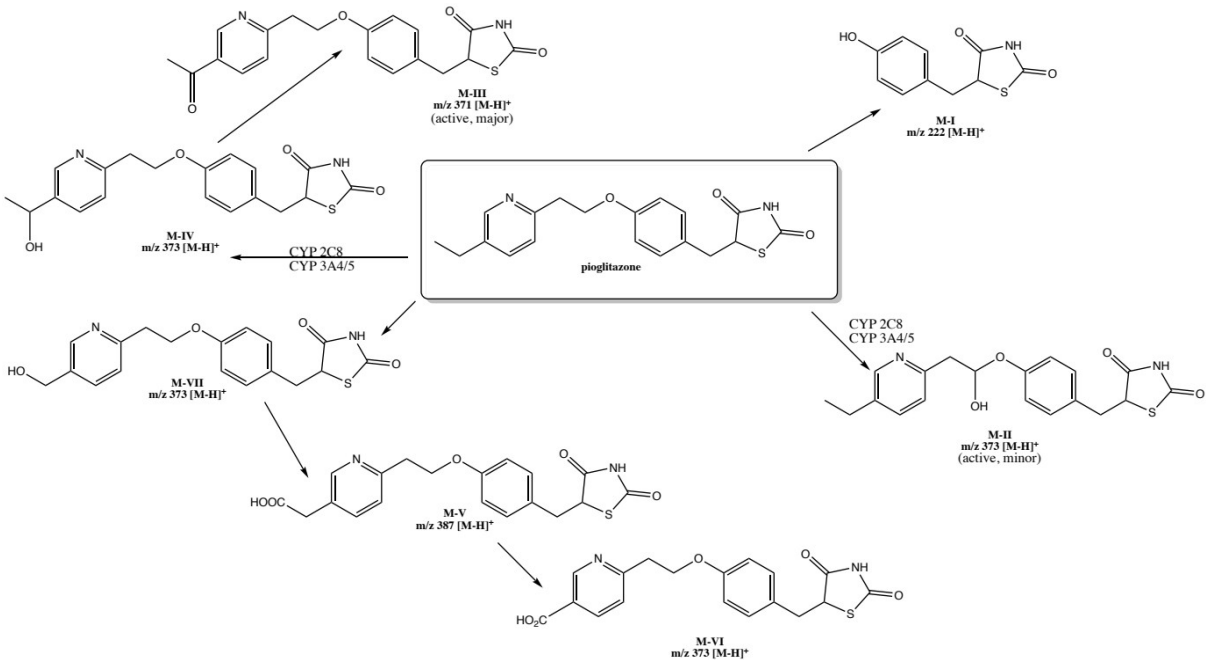
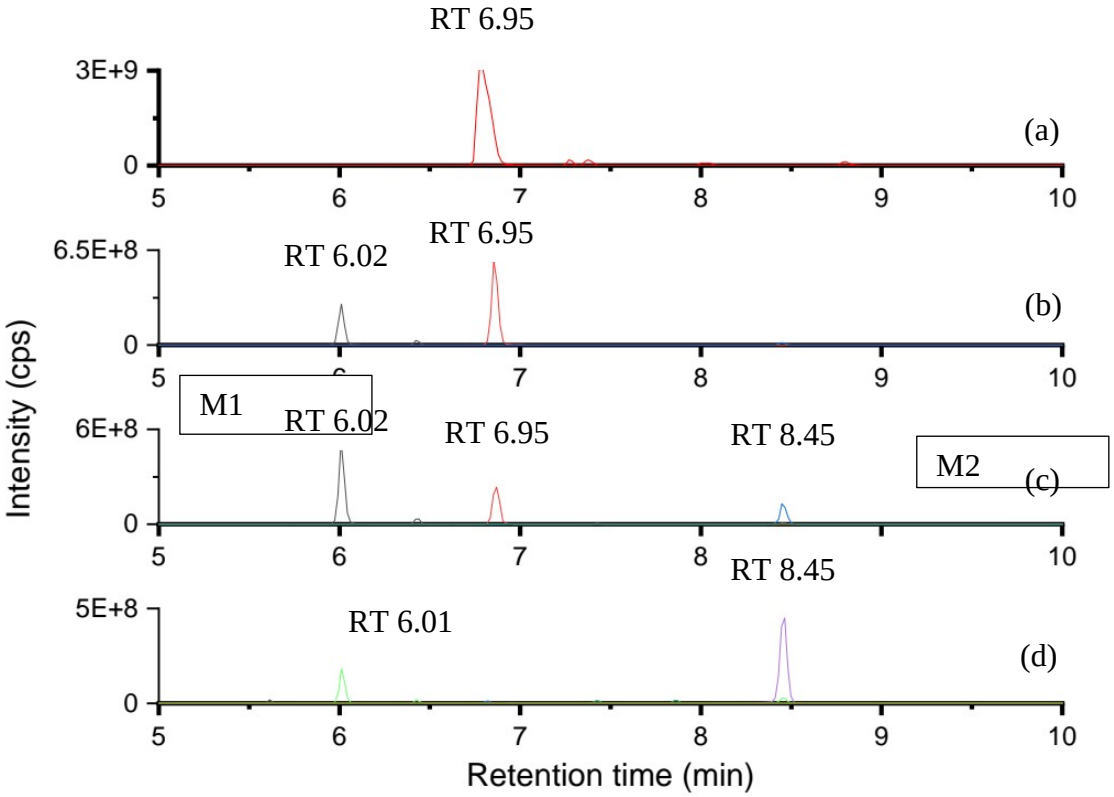


Fig.. Structures of literature reported pioglitazone metabolites, these metabolites have been most extensively reported, though novel metabolites have been identified in dog kidney and liver microsomes. M-VII….. Figure created in ChemDraw. Figure adapted from (Kawaguchi-Suzuki and Frye, 2013)

3.4.1 Pioglitazone metabolite elucidation via LC-MS/MS



Fig… LC- MS/MS plot of Relative intensity (cps) vs Retention Time (min) for the PIOG only standard (a) and the 0, 5 and 20 minute following initiation of turnover (b, c, d). Clear parent ion depletion can be seen, as well as the formation of metabolites, M1, and M2 marked. The red peak at RT 6.95 shows the depletion of the parent ion peak over time. The peaks at RT 6.01 show M1, with the peak in green showing evidence of secondary metabolism as the M1 peak begins to deplete again. The peaks at RT 8.45 are marked M2.

Owing to the unrivalled turnover rates for BM3 DM, full turnover of parent compound is seen by the 20th minute time- point, as no more parent compound peak, with s retention time of 6.95, can be seen. The 0th minute time point already shows a significant metabolism, as the parent ion peak depletes and the formation of the new species, M1, has already begun. The formation of two metabolites M1, RT 6.01, and M2, 8.45, are also marked. Evidence of secondary metabolism is shown in green as the M1 peak intensity begins to fall between the 5th and 20th timepoint.

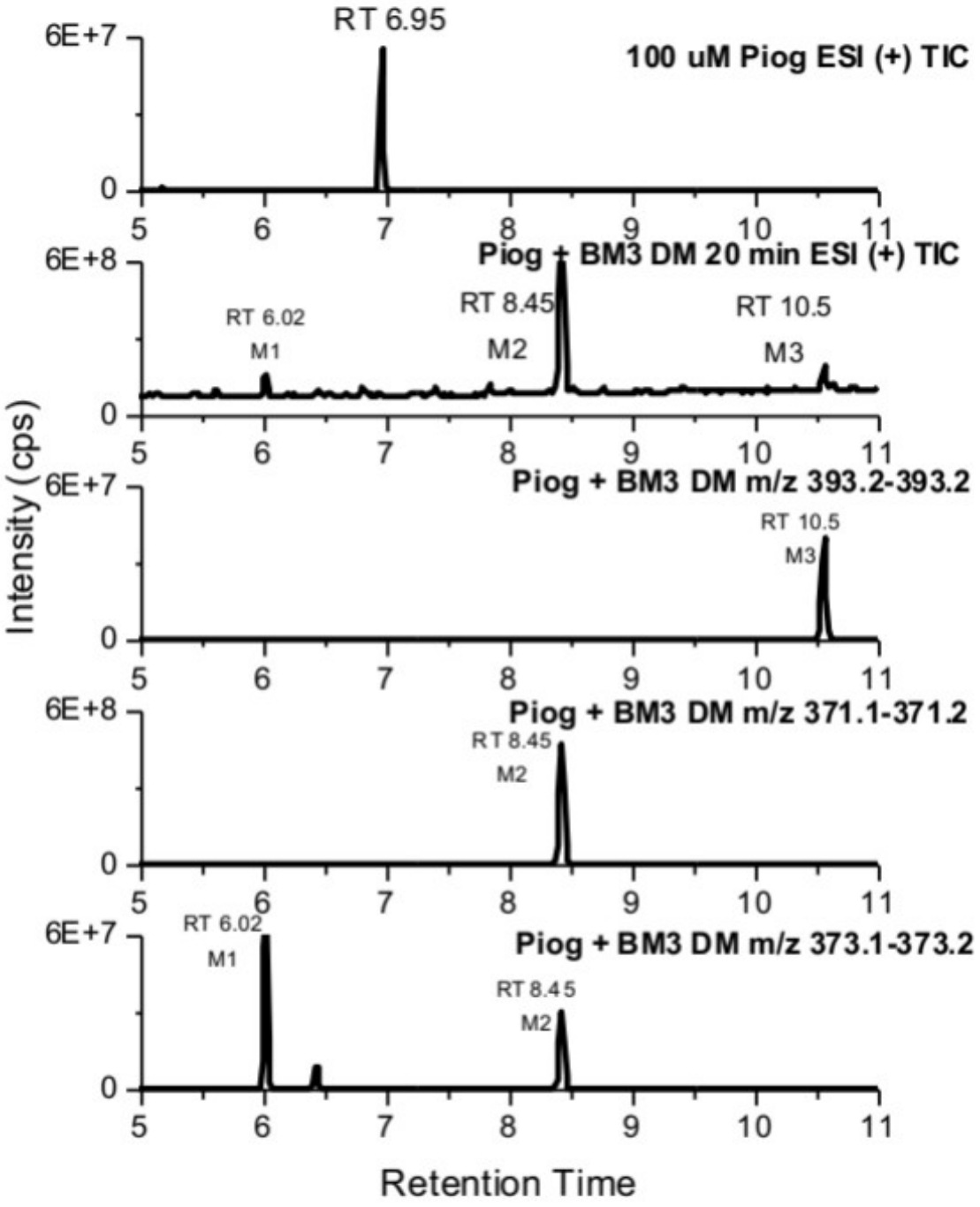


Fig...Piog standard stacked with... Add axes labels.. ADD new fig....

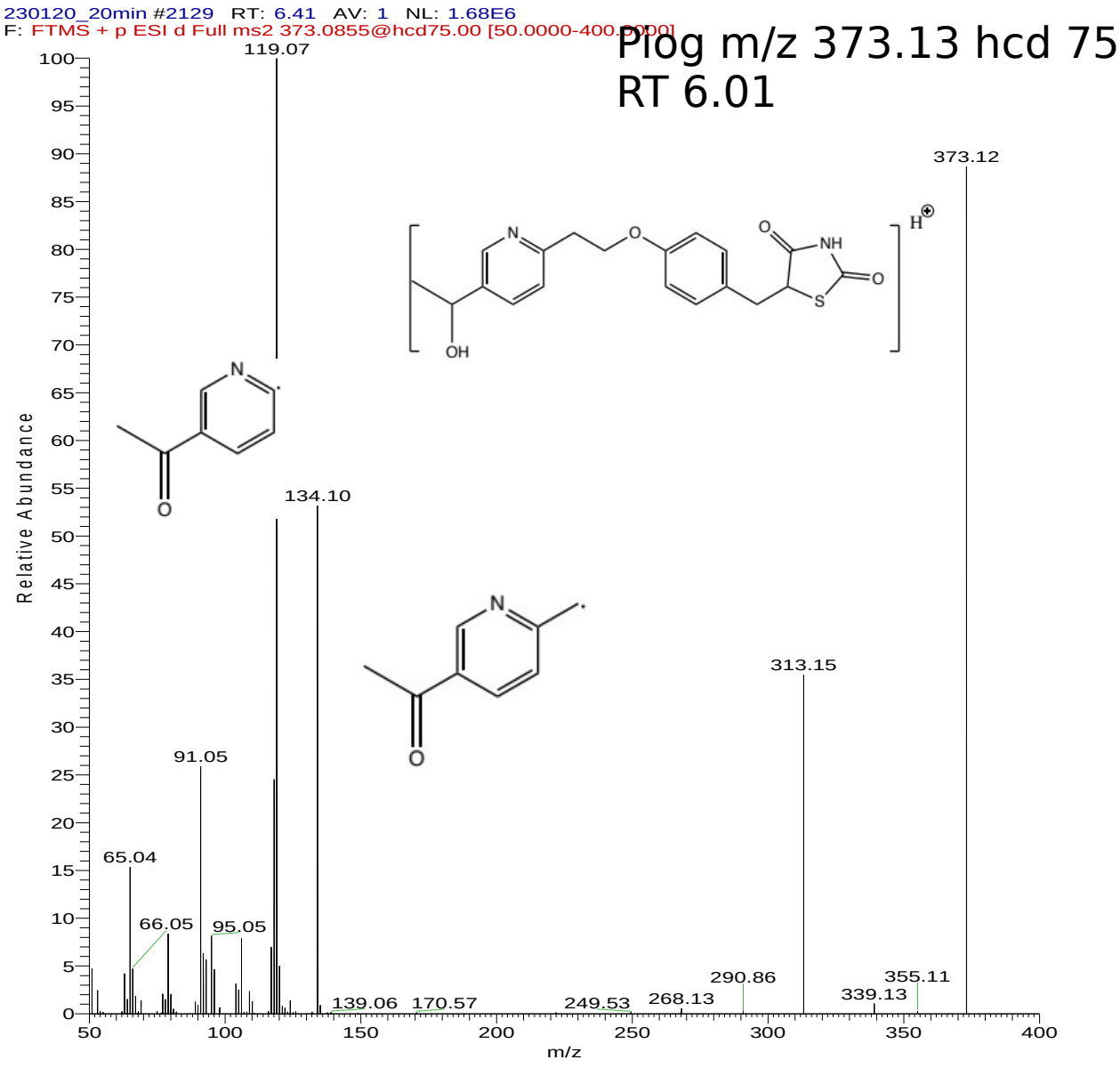
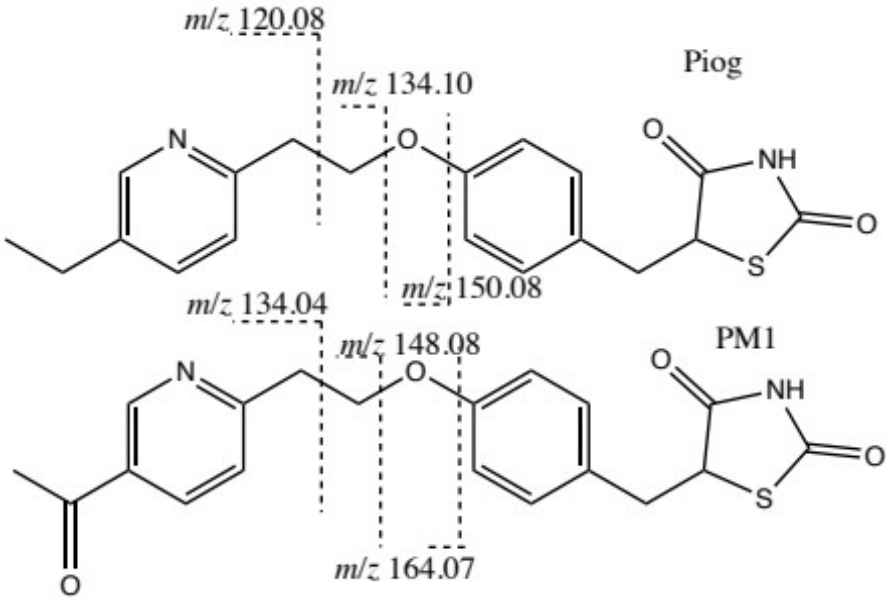


Fig... Fragmentation TIC of Piog 20 minute timepoint. Fragments of interest are labelled and indicated with an arrow. A +16 molecular ion peak can be seen, which is indicative of a hydroxylation (+OH). The other two labelled fragments of note are helpful in determining which side of the molecule the hydroxylation has occurred.

313.15? C_2? S?



Fig… Proposed fragmentation of PIOG and hypothesized metabolite PM1.

The site of hydroxylation was thought to have been on the aromatic ring due to previous work on pioglitazone and YASARA docking simulations. The LC-MS/MS data presented in Fig.. confirms the presence of a hydroxylation as a parent ion peak with a mass of +16 can be seen within the 20 minute timepoint. This addition of mass is indicative of a hydroxylation. Two other fragments are outlined in Fig.. which may be of interest in determining the site of the hydroxylation. The two marked fragments at m/z 120.04 and 134.10 show a mass increase by +16 when compared to the standard. Indicating the modification occurred within this region of the molecule.

Further elucidation by 1D and 2D NMR was necessary in order to fully evaluate changes in multiplicity in both the alkyl chain regions and aromatic regions in order to confirm which of these two regions werr hydroxylated.

3.4.2 NMR for Pioglitazone metabolite elucidation

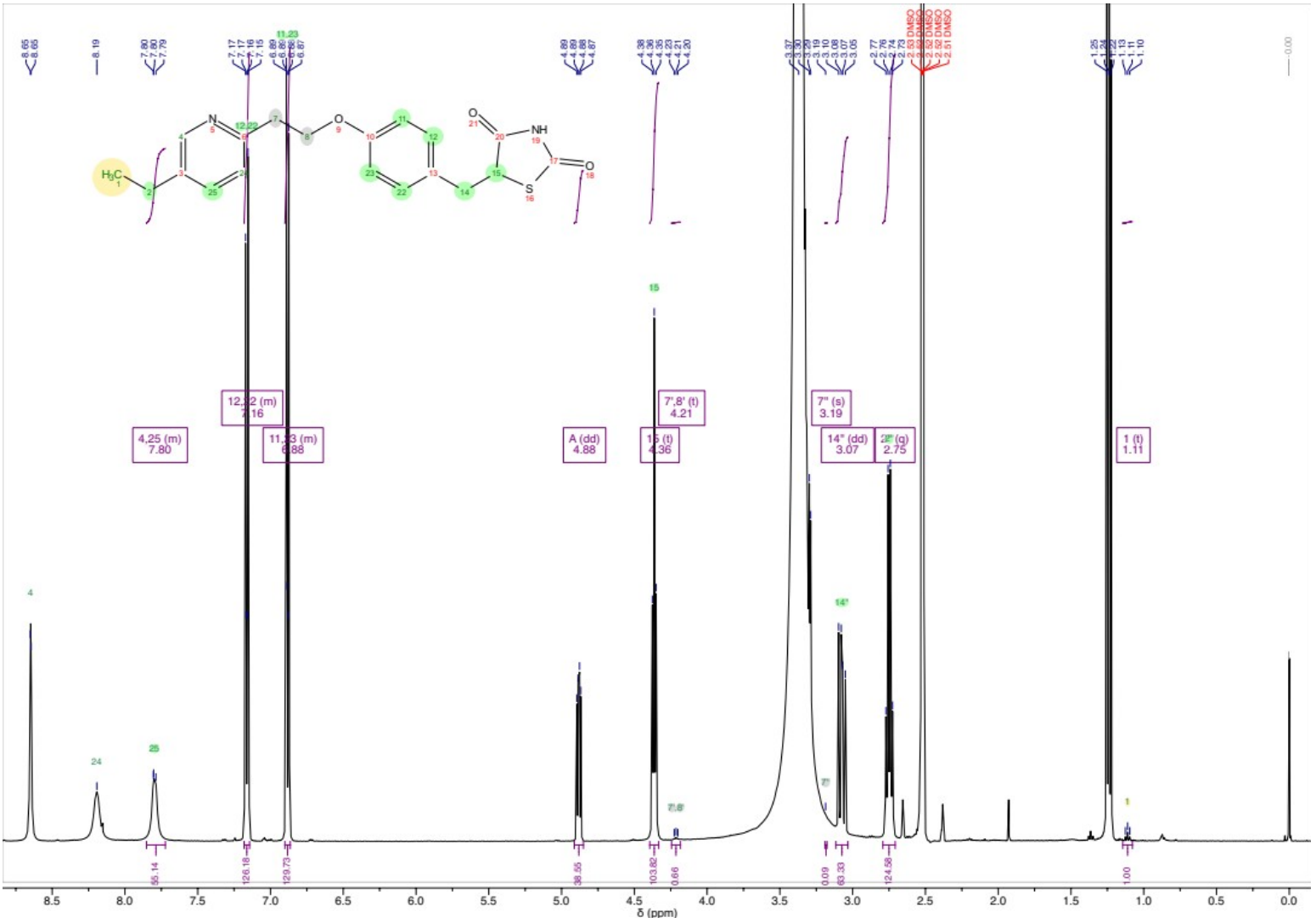


Fig.. PIOG standard, ^{1}H NMR assignment. 10 mM PIOG in DMSO-$_{d6}$. PEAK at 4.88? Data were collected on a Bruker 500 MHz NMR in DMSO-d_6.

PIOG STANDARD TOCSY

PIOG BULK ASSIGNED

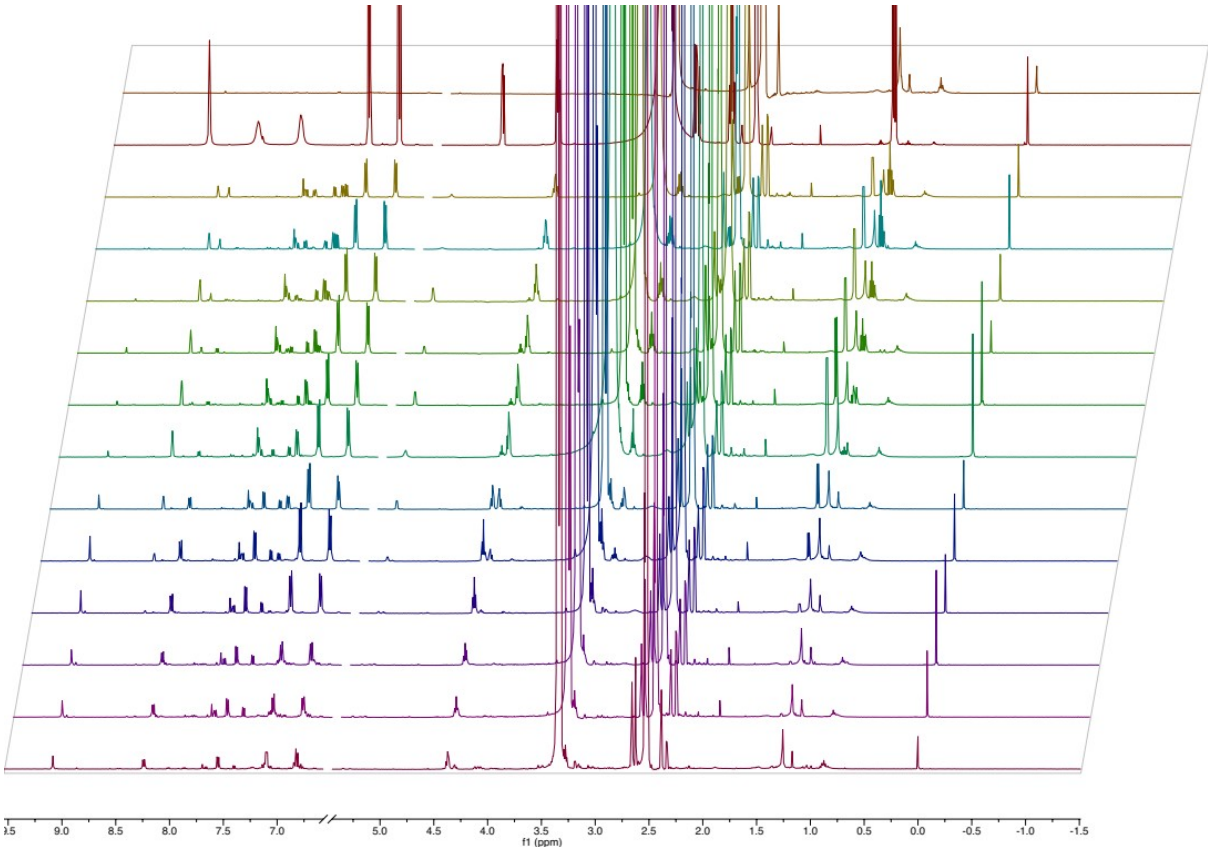


Fig... Stacked ^{1}HNMR spectra for PIOG, and corresponding time points in the following order.... Data were collected on a Bruker 500 MHz NMR in DMSO-d_6.

LABEL CONTAM AND UNKNOWNS, STACK PROPERLY WITH TIMEPOINTS CUT OUT WATER PEAK..

It can be seen from superimposing the stacked ^{1}HNMR spectra for PIOG, with the time points against the standard, where peaks shift and multiplicity changes over time with sequential metabolite formation as Piog is turned over by BM3 DM.

It can also be seen that the variation in water peak ppm, demonstrating the varying discrepancies when each column was dried under nitrogen. The drying stage is of paramount importance and should be left until the column appears cracked, to ensure as much water as possible is removed. This in turn would improve the resolution of some of the broader peaks. Furthermore, the water signal suppression… such that the signals of peaks are hidden by the bleaching of the water peaks.

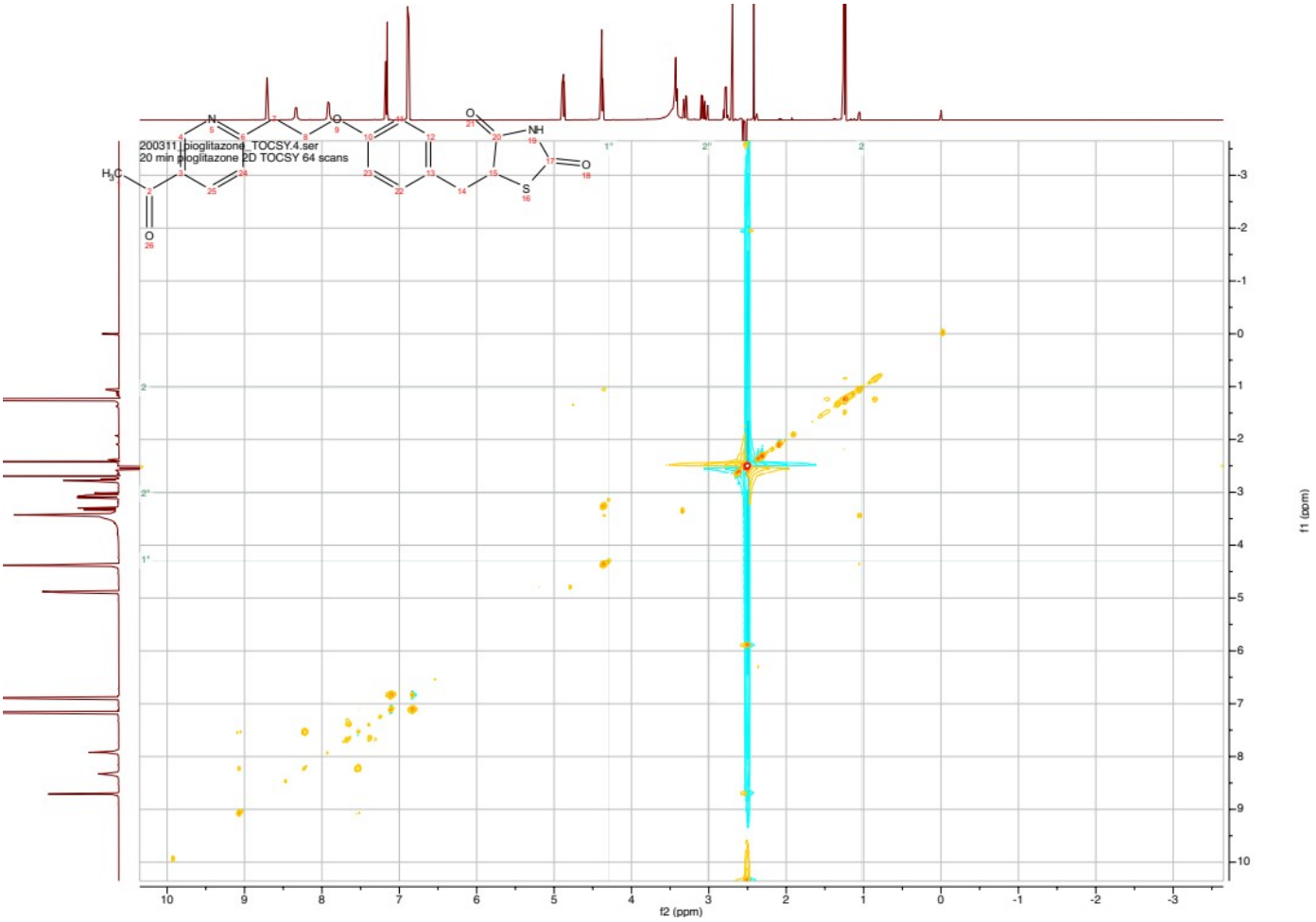


Fig.. 2D TOSCY spectra for PIOG 20 minute timepoint.... Confused with the assignment for this. Data were collected on a Bruker 500 MHz NMR in DMSO-d_6.

From analysis of the stacked spectra it is easier to deduce changing multiplicity and disappearance of peaks. The triplet at a shift of 1.07 was originally assigned to the terminal CH_3, which is not present in the 20 minute time point. This is due to a change in splitting with what were previously 2H's on C2 (see Fig…). The proposed metabolite PM1 no longer

contains a proton environment on C2, resulting in the observed change in splitting. An acetone contaminant can also be seen at around 2.11 ppm, which was present within the standard run with the timepoint but not in the standard run alongside the controls.

2D ROESY SPEC/ COSY for complete elucidation- mention coupled spin- sytem at...

3.4.3 Pioglitazone Conclusions

A combination of the LC-MS/MS and NMR techniques allowed for a full elucidation of the proposed metabolite (M1) for Pioglitazone. Initial LC data indicated fragments with +16, indicative of a hydroxylation. It was originally hypothesized that the hydroxylation would occur in the aromatic region due to previous studies and docking within *enz*, however, with LC data it can be seen which side of the molecule the oxidation occurs. In order to fully determine the position of the oxidation, ^{1}H NMR alone did not suffice. It was through the allocation of the ROSEY coupling that it could be suggested that the oxidation occurred at the C2 position.

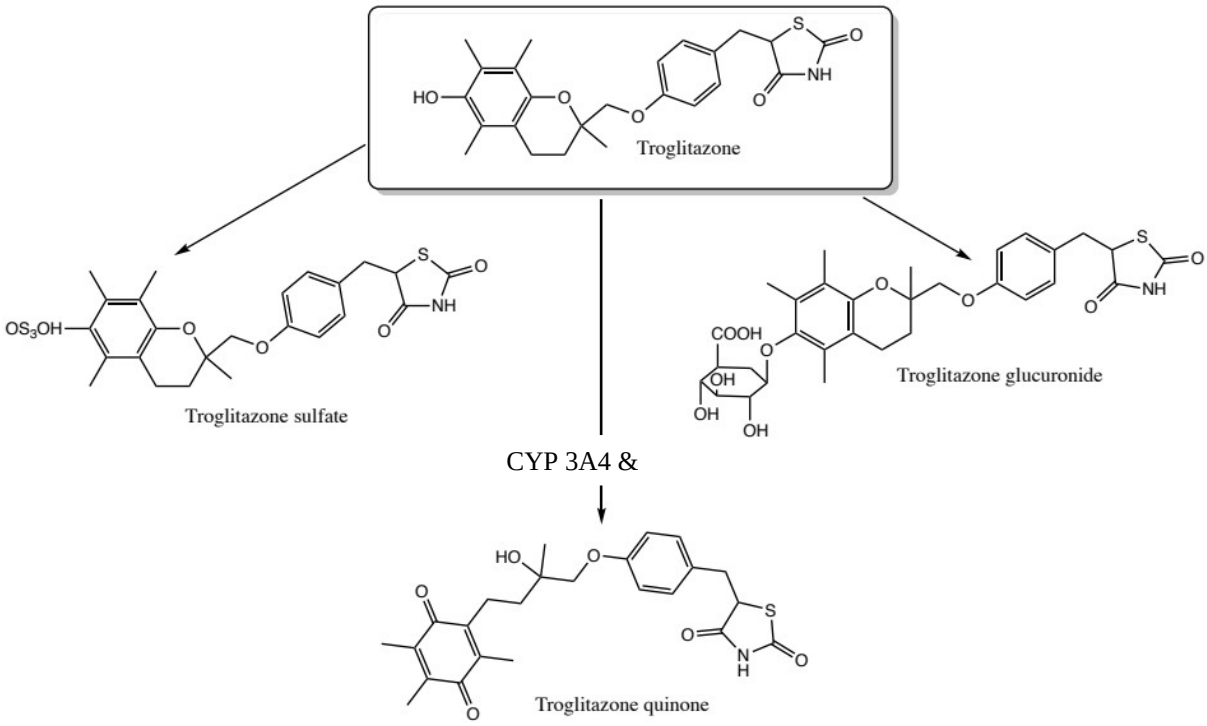
The optimized method for separation of metabolites over time made for clearer spectra and ease of determining the major CYP3A4/5 human metabolite formed.

The elucidated metabolite PM1, is a major human metabolite of Piog. The oxidation occurs at the C2 position. Using BM3 DM in order to produce this metabolite allows for ease of selective oxidation when compared with traditional synthetic methods of such exclusive oxidations.

3.5 Troglitazone

Following Pioglitazone, methods were optimised for another member of the glitazone drug class, Troglitazone (Trog). Trog is also an oral anti-diabetic drug and was first synthesized and proposed for use in treatment of type II diabetes mellitus, sometimes in combination with other anti-diabetic drugs such as Metformin.

As all members of the glitazone drug class share the thiazolidinedione (TZD)
moiety, it was hypothesised that the mode of action for each of the drugs were to be analogous in acting primarily at the..



Scheme… Outline of primary metabolic pathways in troglitazone metabolism, as outlined by(Metabolism and Chemistry, 2004)

3.5.1 LC-MS/MS for the elucidation of Troglitazone metabolites

The assay parameters were not altered from the initial pioglitazone assay, and following a bulk incubation of BM3 DM with Troglitazone for 180 mins, all controls incubations were crashed with concentrated HCl. The purpose of the acid crash was to denature any protein present within the samples, thus allowing removal of precipitates via centrifugation.

However, upon analysis of the LC-MS/MS data produced from the initial Troglitazone run, it was evident that breakdown of standards and incubations had occurred.

Several reasons for the breakdown were proposed. It was initially thought that the Troglitazone had photolyzed…. This was due to the lack of parent ion within the LC-MS/MS standard… Fig.. and the breakdown of the standard… Furthermore, the apparent lack of metabolite peaks indicated that the compound had broken down prior to incubation. This may have been due to a number of reasons, firstly improper storage of the compound…

Evidence has been found within literature to propose mechanisms for photolytic degredation of troglitazone. Fig… shows a mechanistic route.

Though with fresh compound ordered, the standard ran… and so proceeded with a second bulk incubation… Precautions were taken to avoid light exposure throught the incubation, and following, during sample extractions.

In order to eliminate sample contamination and photolysis, a brand new vial of Troglitazone was ordered and foil was used during all incubation and extraction stages to avoid irradiation…

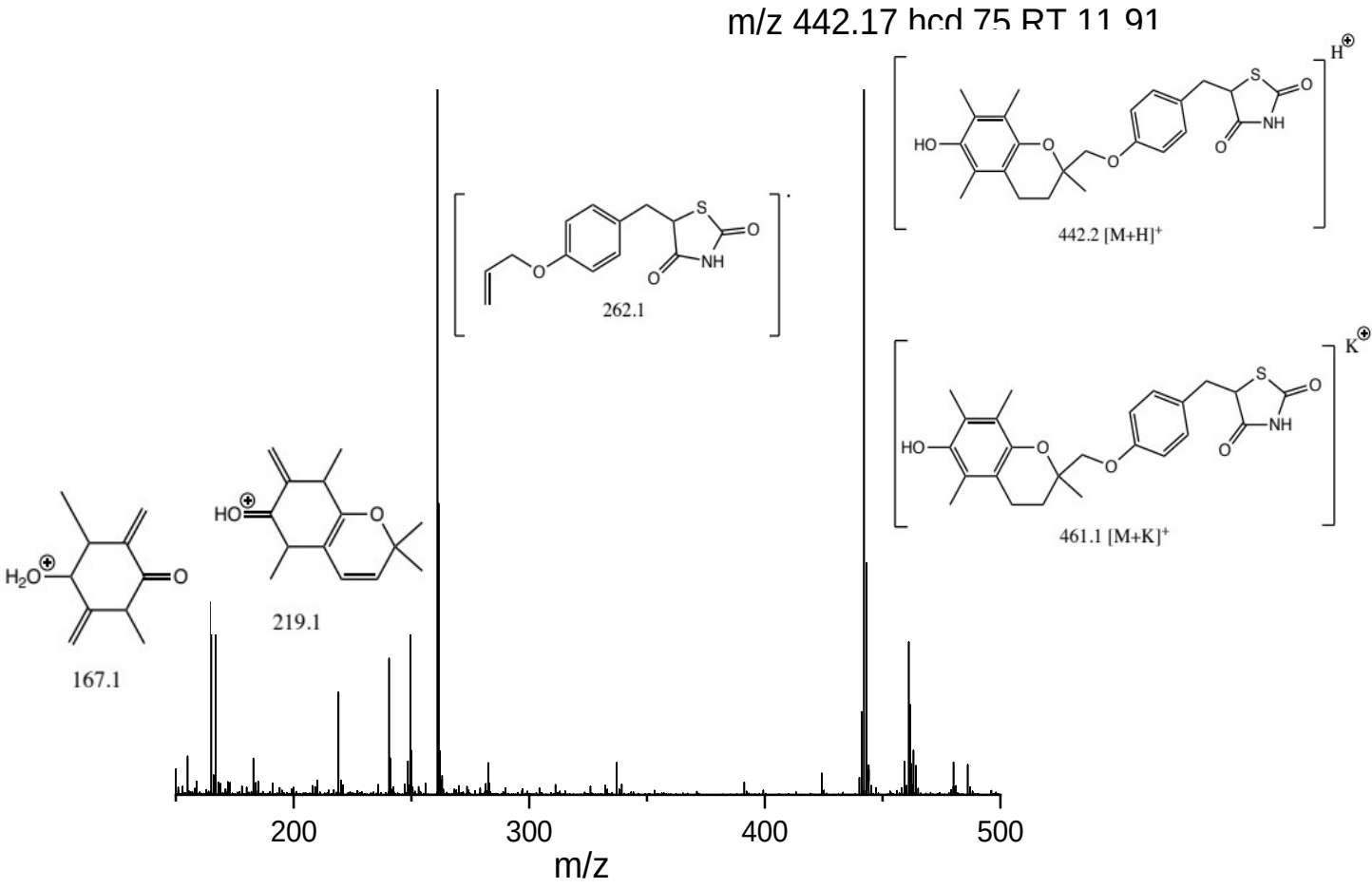
A further proposed reason for the breakdown of the formed metabolites was the cleavage of the TDZ ring moiety due to the acid crash.

In consequent method optimizations, it was ensured that the need for the acid crash and denaturation of protein was avoided.

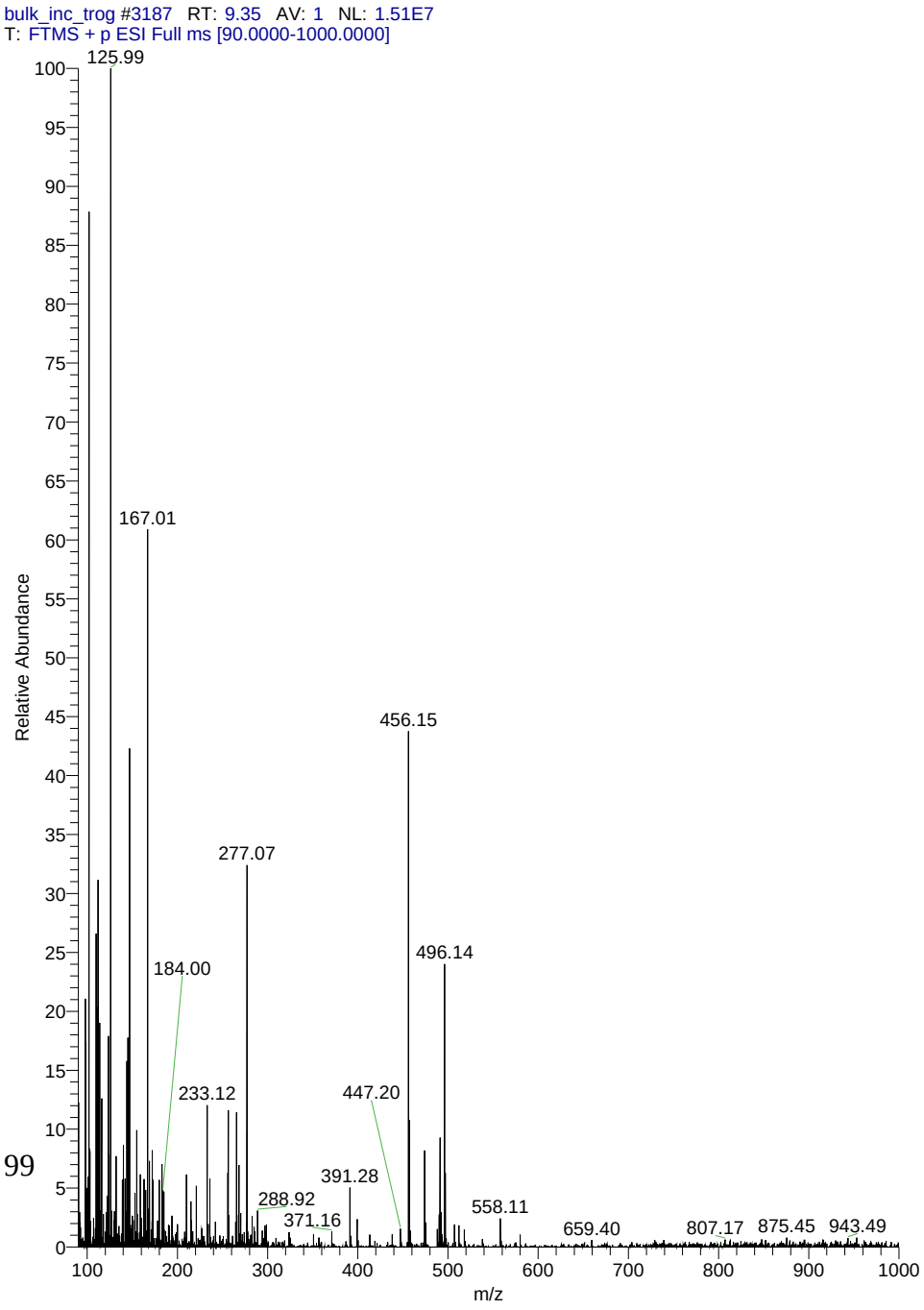
Other methods of protein removal were proposed, such as ammonium sulphate cut… but would salt out the metabolites as Trog salts exist.. Flash freezing of samples, not possible due to movement restrictions…

Finally, a protein precipitation plate was used prior to concentrating each samples on SPE columns. The plate was primed as per the manufacturers protocol and each sample was extracted under a vacuum. The use of the

protein precipitation plate prior to SPE extraction, ensured that any remaining protein would not block the SPE column. As is evident within the LC-MS/MS standard chromatogram, a clear parent ion peak can be deduced as well as the expected fragmentation…

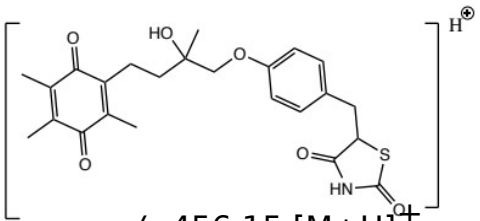


Fig… Troglitazone Standard TIC, m/z range of 150- 500…. With Parent ion $[M+H]^+$, potassium adduct $[M+K]^+$ and three significant fragments labelled. Figure produced in OriginPro 9.0 (OriginLab,…)


bulk_inc_trog #3187 RT: 9.35 AV: 1 NL: 1.51E7
T: FTMS + p ESI Full ms [90.0000-1000.0000]
125.99
167.01
456.15
277.07
496.14
184.00
233.12
447.20
391.28
288.92
371.16
558.11
659.40
807.17
875.45
943.49
Relative Abundance
m/z

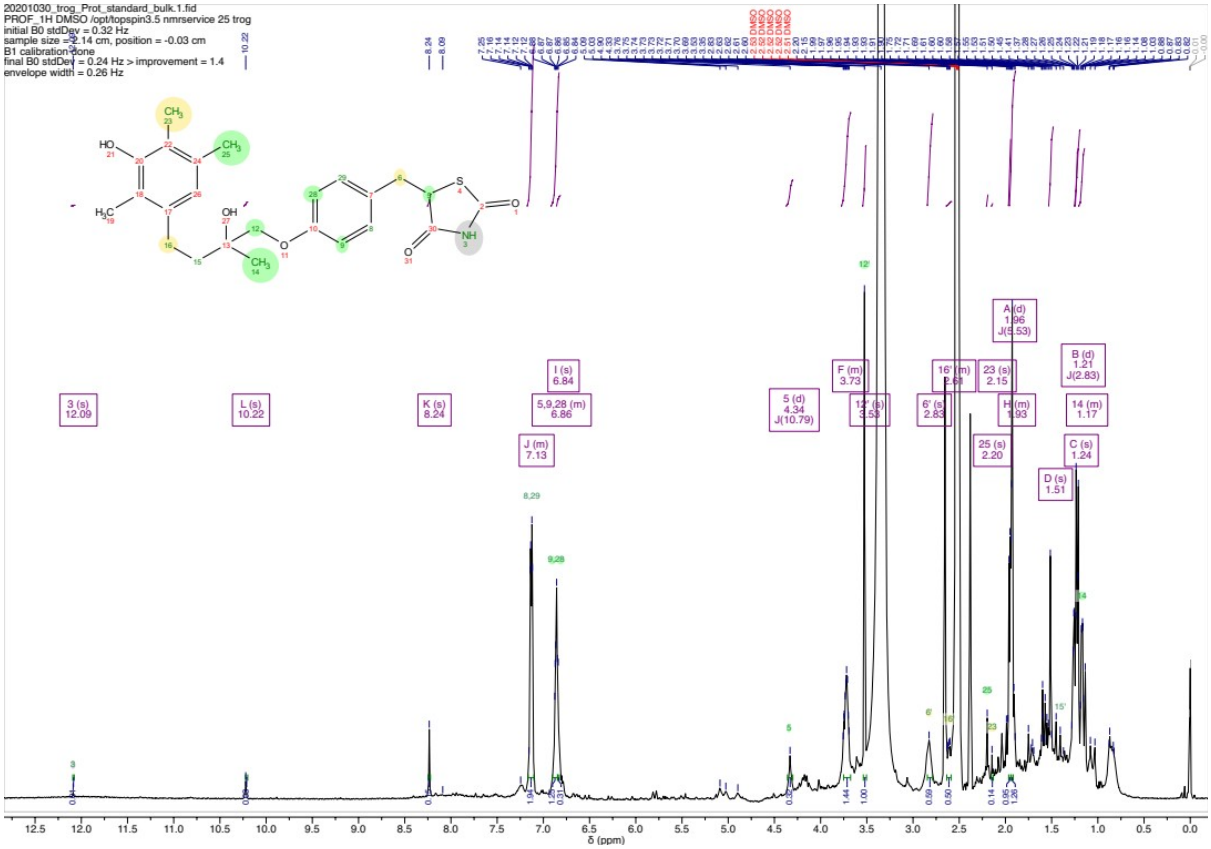
3.5.2 NMR techniques for further elucidation of Troglitazone metabolites

Fig... ^{1}H-NMR assignment of 100 uM Trog standard. Data were collected on a Bruker 500 MHz NMR in DMSO-d_6. In Appendix?



m/z 456.15 $[M+H]^+$

Fig.. Stacked 1H NMR spectra of TROG standard (a) against TROG bulk incubation after 180 mins (b). The areas of interest are the aliphatic and aromatic regions. The proposed metabolite has no changes in the aromatic region which can be seen in b) the aromatic region still shares the same peak multiplicity, indicating that any changes to TROG were not seen in either aromatic region. Data were collected on a Bruker 500 MHz NMR in DMSO-d_6.



Fig… ^{1}H- NMR Trog bulk incubation- 180 min… not crashed due to acid cleavage. Data were collected on a Bruker 500 MHz NMR in DMSO-d_6.

^{1}H NMR spectra of the new TROG standard was assigned and it was apparent that no breakdown had occurred. Upon attempts to assign the first TROG bulk standard and incubation, it quickly became apparent that there was a breakdown in the standard and subsequently the metabolites also. It can be seen from the original standard that there may have been a breakdown due to photolysis… as seen in the literature….

Insert old trog NMR/LC for standard to show breakdown…

HSQC

COSY

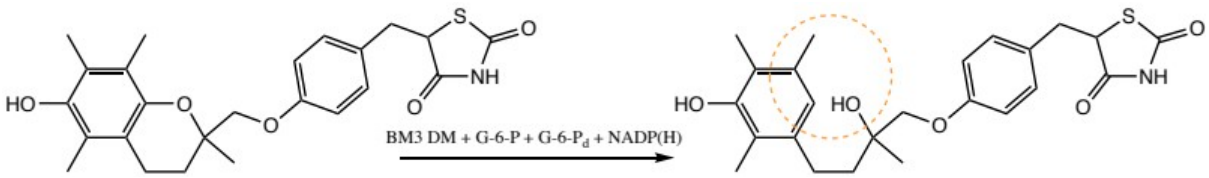
ROESY

In the ^{1}H- NMR spectra of the TROG bulk incubation, the peaks are badly resolved and the internal TSP standard peak is split abnormally, this is indicative of problems with the shimming of the internal magnet. Due to the shimming, and the resulting peak resolution, the integrations of the peaks within the spectra are less reliable, therefore it is necessary to weigh more on the LC-MS/MS data. Further elucidation would be necessary via NMR confirm the identity of the metabolite, however it…

The mixture of metabolites within the TROG bulk incubation, due to the lack of separation by time, may be futher explored and further attempts may be made in future work to separate metabolites via preparative- scale high performance liquid chromatography.

The peak at 3.53 I think is actually the singlet of the tautomer of the glitazone ring at 6H2, ive seen this for rosiglitazone too, you can justify this by saying the integration of the aromatic environments and that peak ratios are the same which means all metabolites have these environments if that makes sense? If you have a look at the troglitazone only proton through SPE this should clear up whether it does tautomerise. Ive changed it on your assignment. I think your 15 and 16 assignments look like a series of singlets from the aromatic region too so id highlight that it's too messy to distinguish but is probably the aromatic methyls. I think using the hsqc and comparing peaks to the carbon predictions for the main metabolite is going to give a bit better elucidation, cause I think the proton is too messy and shimming is too off to distinguish any peaks related to the metabolites. This is alright though as you can discuss this and say at the moment the LCMS data is the only one youre confident on.

3.5.3 Troglitazone conclusions



Scheme… Proposed major metabolite formation for TROG after 180 minute incubation with BM3 DM and electron regeneration system.

Troglitazone metabolism proved to be more difficult for a number of reasons. Firstly, initial attempts of incubations with BM3 DM and Trog failed, due to temperature dependent breakdown, or potential photooxidation or the Trog and any potential metabolites formed.

With new reagent it was possible to obtain cleaner LC-MS/MS spectra, and upon incubation with BM3, full turnover of the parent drug was achieved. Following assignment the standard fragmentation, and evaluation against a set of controls, it was possible to determine a metabolite ion peak and fragmentation (*m/z* 456.15), not seen in previous BM3 Trog

And a proposed assignment of the proton standard NMR was possible. It was however difficult to elucidate the structure of metabolite(s) formed with a combination of the two techniques. With the LC data…

The lack of resolved peaks in the ^{1}H NMMR… indicative of shimming… as well as the splitting of the TSP standard…

It would be necessary to re run samples with appropriate shimming and acquisition parameters….

Furthermore, a greater degree of separation could be achieved, either by running a tine deoendant assay or… use of separation techniques ie prep HPLC…

Optimise methods of protein removal….

However, there is enough evidence to suggest that the metabolite formed following incubation of BM3 DM with TROG is the major quinone- like human metabolite which has been reported in literature as a main source of hepatotoxicity. The formation of this metabolite eventually lead to Trog being removed from the market due to… and even reported adverse drug-drug interations (DDIs) with other common drugs such as paracetamol.

3.4.3 Production of valuable human drug metabolites with P450 conclusion

Cytochromes P450 enzymes are known to catalyse a range of structurally diverse xenobiotic substrates. The metabolites formed, are more easily excreted via renal or bilary pathways due to the increase in hydrophilicity of the molecules, resulting from hydroxylations.

In general, these drug oxidations halt any pharmacological potency…

P450 BM3 has proven to turnover a wide range of structurally diverse compounds, in this project in particular, BM3 has proven its high turnover rate…..

P450 BM3 was able to produce major human metabolites for both PIOG and TROG, needs to be further elucidated with trog… This would be of interest to…

Utilities in… scale up..???

Employing both LC-MS/MS alongside 1D and 2D NMR techniques has proven to be an effective method for metabolite elucidation of these members of the glitazone drug class. Coupled with previous work by Thisslethwaite et al. it can be seen that these techniques in tandem can be powerful in elucidating unknown metabolite structures across a range of drug classes also.

Separation of metabolites by timepoints, as with the PIOG assay, resulted in clearer, more interpretable spectra for NMR analysis. Within the TROG bulk incubation, the NMR spectra are complex in nature due to the formation of multiple metabolites. Further attempts to separate out metabolites via optimising preparation- scale HPLC methods may result in clearer spectra.

Problems encountered…

Use of BM3 DM and various other mutations for the production of human metabolites could be an effective method to scale- up. Due to the highly selective nature of BM3, DM and various mutations it is possible to introduce otherwise complex hydroxylations of a diverse range of drug classes.

HERE OR SEPARATE CONCLUSION SECTION…

3.5 Computational enzyme design python package

A computational enzyme design pipeline was constructed with the aim of docking drug molecules with P450 BM3, from a template amino acid sequence, and introducing mutations to preferentially favour the hydroxylation of the substrate, at a defined position. Pioglitazone was used as a starting point , it was first necessary to identify which major, active human metabolites would be of interest, as seen with Section.., with human drug metabolite formation with BM3 DM. For Pioglitazone specifically, the aim was to engineer BM3 from the WT sequence, to introduce novel mutations to tune specificity to hydroxylate at the C2 position. A several score criteria were involved at various steps in the design and screening of the possible BM3 mutants.

Due to time and practicality constraints, the package '*enz*' was developed with the assumptions that the protein backbone is rigid and various other assumptions in devising suitable score functions will be further discussed in subsequent sections.

The basis of this package was to merge the processes of introducing mutations and docking.

The python package *enz* wraps protein folding software, PyRosetta, with docking in AutoDock Vina. Further downstream, aims were to semi, and eventually fully automate as many processes within the computational design of BM3.

With automation in mind, API... and the introduction and rationale behind suitable score functions became more important. Generation of scores based on custom scoring fucntions (CSFs) allowed output structures to be evaluated and compared against BM3 WT and DM structures, without the need to visualise structures at each discrete step, i.e. following the

introduction of the mutation and subsequent repacking of side- chains. The minimisation of structure comparisons with visual inspection, was time saving… and… within the design process.

The introduction of mutations, side chain repacking and refolding was based functions provided by PyRosetta (Gray Lab), and the docking- end of the programme relies on… from AutoDock Vina (Trott and Olson, 2009; Chaudhury, Lyskov and Gray, 2010).
Talk about hardward…Ssh into Linode server, only 2 CPUs so weren't able to run ensembles and dynamic flexible regions. ….

assumption was made when formulating the custom score function used within the alanine scan….

Pyrosetta is a… protein modelling suite, developed in.. as Rosetta@Home.. Initially Rosetta relied on C++, a programming language that requires scientists to have a background in this language.. PyRosetta was created as the python interface, as Python was simpler coding language and much more accessible…INTRO?
The full-atom score function in Rosetta has been repeatedly improved through the introduction of new energy terms and optimization of term weighting. Although Rosetta-based simulations can generate accurate structural models, correlations between the canonical score functions and experimental data remain relatively poor (Alford *et al.*, 2015)

Rosetta Score Function

Rosetta utilises a score function reliant on an approximation of the energy of particular
molecular conformation, alongside a linear combination of weighted physical and statistical energy terms.

Ref2015 “SCORE”- semi- empirical… pdb bind dataset… based on K_d … sampling bias… drugs screen… default score function… predict new structure- refold… side- chain re- pack. Side chain conformation library… (sampling each conformer) 5A residues… :D :D :D

Vina Score Function

Generates conformations,…. Two step docking…. Genetic alg… conf of ligs… local optimisations, gradient decent… LBGBFS…

Exaustiveness….. number of runs set by this function?

Custom Score Function (CSF)

A degree of automation was introduced when generating score functions in iPython, without the need for visualisation. Aims were to produce a semi- fully automated searches, utilising a CSF with Bayesian optimisation or classical genetic algorithms was preferred.

Assumptions made when generating an appropriate fitness score were that the position of the docked substrate will be similar *in vitro*, and therefore the hydroxylations would preferentially occur at the regions defined in *enz.*

A CSF was used designed in order to score poses of ligand bases on mean weighted averaged distances to the C2 position, and its proximity to the heme Fe atom of BM3…

Fig and equation of how score works?

EQUATION for CSF

Scaling factor… distance… multiplied by energy exponent… assumption of the pose of highest fitness being proximal to Fe and therefore more likely to oxidise at the C2 position

This allowed for more intrusive docking studies with the aim of generating mutants of high fitness which would be able to selectively hydroxylate at the given C2 position. When using *enz* for docking simulations with Piog and various mutants of BM3, semi- rational design based on Arnold BM3 mutant library (REF) was used to hypothesise mutants which would be able to selectively form the major human metabolite, M1, as seen in section….

For each docking run, the default output is 8 poses…

Input equation…. Sigma… mean EUCLIDEAN distances…

More limitations… Introduction of ML models for optimisation and tuning score function (FUTURE WORK) (Shringari *et al.*, 2020)

Computational Alanine scanning

In order to determine the key residues within the active site of BM3, a computational alanine scan was run, based on methods by Baker lab? (Kortemme, Kim and Baker, 2004). Alanine scanning was a method which was developed in order to predict which amino acid residues are of energetic importance within protein- protein interactions. In the case of docking Piog to BM3, the method of alanine scanning allowed insight into which mutations had a dramatic effect on binding affinity.

Simple free- energy functions are used to calculate the effects of alanine mutated residues on the binding energies of P+L complexes. Wrapped within the function, the atomic packing interactions are defined by a Lennard- Jones potential, a hydrogen- bonding potential based on the inputted protein structure and various statistical approximations of combinations of possible rotamers.

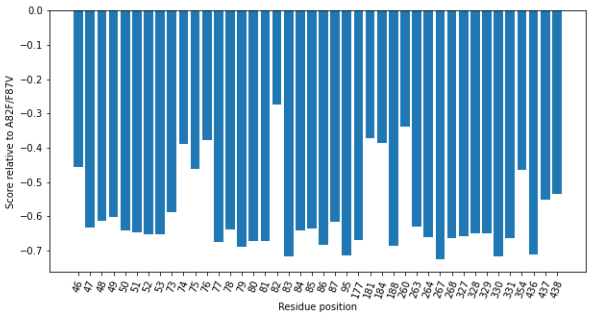
Fig… Simplified schematic of alanine scan steps within each iteration. Each output contains 8 distinct poses of Pioglitazone.

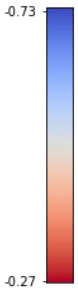
The scores used throughout each alanine scan were Vina scores…

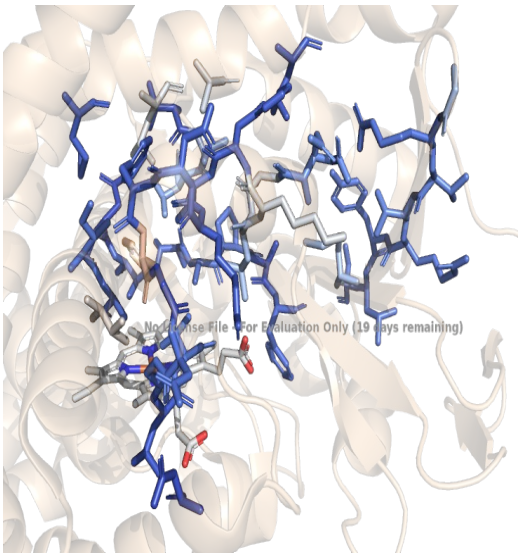
Based on following equations? (Intro?)

The alanine scan of BM3 and Piog with the key residues defined…. Revealed… Mostly, it revealed which residues were not to be mutated…

From Fig… it can be seen that…







Fig… Divergent heat map indicating the importance of selected active site residues on the binding affinities of BM3 and Piog. Residues and heme generated in stick form. … NEED TO REMAKE WITHOUT WATERMARK, need to remake alongside distance fig… Fig… Bar Chart indicating residue position against output score, relative to A82F/F87F DM following iterations within the computational alanine scan….

The scores of the alanine scan mutations relative to the DM were visualised in Fig.. it was to be expected that mutations would not score as highly as the DM because… disruptions to stability..

As seen in Fig… analysis of the various scores from the alanine scan show that the following residues were of most importance in the binding of Piog. The alanine scare results most importantly indicated which residues were not to be mutated or altered.

Semi- Rational BM3 design based on Arnold mutant library

Following the results of the alanine scan, the residues identified to be of most importance within the binding of Piog were at positions L75W and L188S.

Newer versions of *enz*… problems arose with sequence alignment early on, causing problems downstream… with mutations occurring…

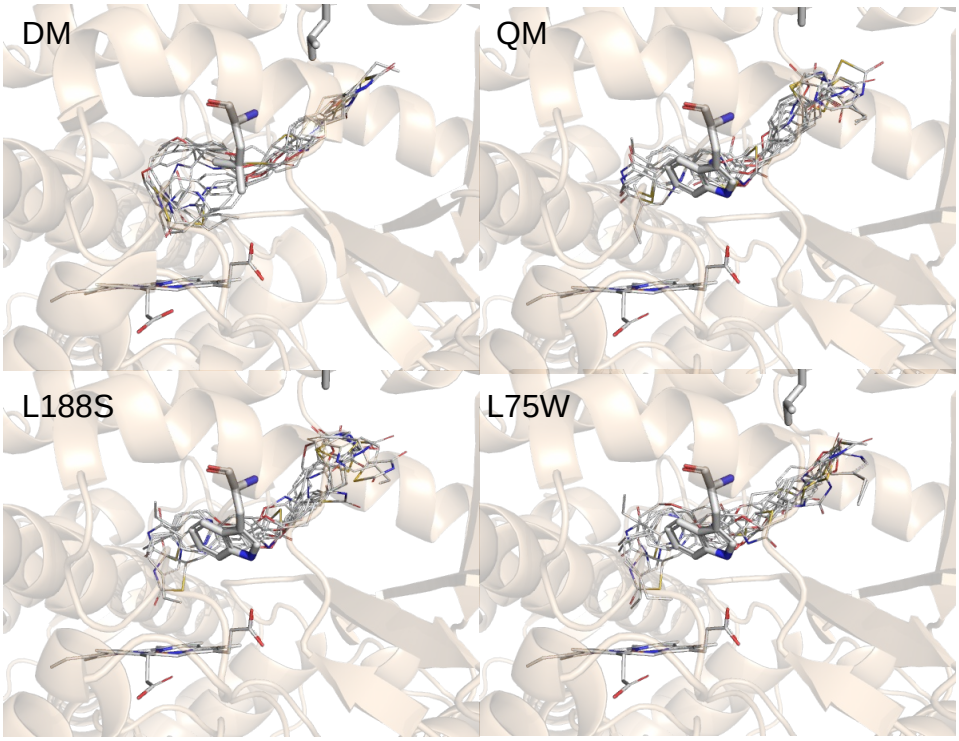
Mutations were semi- rationally picked from known mutants within the Arnold BM3 mutant library in order to improve selectivity of BM3 hydroxylating at C2 position of Piog.

These two residues coupled with the known binding capabilities of the 82F/87V DM were then applied to docking simulations to evaluate the effect on Piog poses binding to BM3 active site.

The following mutations were chosen because… as seen in Fig.. (panel fig) the DM in combination with the 75W and 188S mutations caused… and therefore a higher frequency of poses of Piog were in the correct orientation to perform C2 oxidation.

75W, constrained the position of PIOG in active site, and 188S constrained position with hydrogen bond….

Structure based accessibility…generation of binding poses to be visually inspected by the user.



Fig… Panel image comparison of docking results, DM, 188S, 75W and QM. 8 poses of pioglitazone per docking run. Constraints can be seen with poses… Heme and Piog, 188 and 75 shown as sticks…

In Fig… it can be seen by comparison of the novel mutants with the DM, that poses conformations are constrained in both the 188S and 75W TMs and QM also. The introduction of the mutations seemed to have an effect in preventing bound Piog to bend within the binding pocket….

By direct comparison of DM, the

The introduction of the 75W was based on… the bulky side chain of the W mutation aids the constraint of the pioglitazone binding…

DISTANCE FIG

Fig… Image outlining the proximity of the C2 position to the central Fe atom within the active site of BM3. IT can be seen that the following combinations of mutations… and more poses were seen in the correct orientation…

Autodock vina

The config file used… to unsure repeatability

-r tortion tree/ rigid…

De novo folding in rosetta Rosetta uses centroid and all atom folding.. we use full atom score

STATs
Gradient decent

Mean weighted distance vs soft max

$$S(y_i) = \frac{e^{y_i}}{\sum_{j=1}^{j} e^{y_j}}$$

Site directed mutagenesis with enz output

The initial aims of creating a computational enzyme design element to the project were to create and test output mutants *in vitro.* This was intended for both method validation and….

The designed mutants were to be transformed and expressed, for use in whole cell assay incubations with Pioglitazone. The selectivity would be qualitatively inspected via LC-MS/MS.

Several attempts were made in order to produce L75W and 188S BM3 TMs, using the BM3 A82F/F87V DM sequence as a template for PCR reactions. The L75W/A82F/F87V TM was used as a template in order to introduce a final 188S mutation in order to form a BM3 QM also.

Two distinct sets of primers were ordered for the mutagenesis reactions but upon analysis of sequencing data, it was seen that the mutations had not been introduced.

Furthermore, QT or DM sequence data was returned, indicative of the presence of an excess of the undigested parent plasmid, potentially due to the lack of DPNI enzyme needed to…..

Other reasons for the failed PCR attempts may be due to primer design, as it was noted that the annealing temperature of one of the sets of primers was too high.. danger of mispriming?..

Also with BM3, trouble can arise when introducing mutations due to the AT rich nature of the sequence, further leading to mispriming.

Computational design conclusions

The aim of the computational enzyme design was to produce mutants based on the results of the alanine scan and docking results.

The need for method validation…

….PCR… didn't work…
Reasons for the mutagenesis failing included…

Conclusions

Both BM3 WT and DM full length constructs were successfully purified. This is evident from the SDS- PAGE gels presented, alongside the UV-Vis spectra. The SDS- PAGE gel for both the WT and DM show resolved bands at 119 kDa indicating the successful expression of BM3 full- length.

Analysis of the UV- Vis spectra shows Rz values ranging from… to …. For WT and … for the DM. A low Rz value is indicative of low purity and therefore… the lower Rz values are to be expected for the DM as the introduction of the A82F/F87V mutations decreases the overall stability of the plasmid?

Formation of apoproteins and the lack of incorporation of the BMP heme cofactor may be further reasons of lower purity proteins, future work may include expression and purification protocol optimisations toward better heme incorporation within full- length BM3 constructs. Though lower Rz values were exhibited, there was no significant effect on the ability of the protein to turnover and metabolise the given drug molecules.

Further reasons for low the purity of both the WT and DM could be due to the lack of incorporation of the FMN domain, and heme domain…. This can be seen within the gel.. where the faint bands at.. following the ammonium sulphate cut can be indicative of aggregation of apoproteins.

Fully optimised LC-MS/MS protocols were used in elucidation of metabolite structures of Pioglitazone and further methods of sample extraction were trialled and optimised in order to process Troglitazone assay samples.

Several methods of protein removal from Troglitazone samples were hypothesised and trialled in order to retain as much sample as possible,

without break down of any formed metabolites. The use of protein precipitation plates followed by sample concentration on SPE columns was successful in retaining enough sample for analysis via both LC-MS/MS and NMR, within minimal breakdown of product and low incidence of contaminant peaks.

Optimised reaction monitoring assay development allowed for separation of drug metabolites formed, via time points, resulting in clearer, more resolved spectra for both LC-MS/MS and NMR to allow for a full elucidation of the structure of metabolite, M1, for Piog (Jeffreys et al., 2019).

Use of water suppression and appropriate shimming programmes…. Reduced any contaminant peak intensities to baseline levels, reducing effects of sample bleaching…

For both Piog and Trog, a combination of LC-MS/MS, coupled with several different NMR experiments were necessary in metabolite structure elucidation. Initial LC-MS/MS data for Piog indicated a hydroxylation and roughly where on the molecule this modification had occurred. IT was necessary to employ standard ^{1}H NMR, alongside COSY and ROESY experiments to define and identify the coupling of relevant spin systems. HSQC experiments proved useful in the case of both Piog and Trog as the spectra matched predicted structures from within the NMR database.

As the Piog metabolism assay was run in a time- dependant manner it was seen that it had been completely turned over by the 20 minute timepoint, which is an example of fast turnover by BM3 DM. This was evident from the lack of parent ion peak within the LC-MS/MS spectra of this particular time point. Though it was not seen in the same fashion with the Trog assay, the drug had been completely turned over within the 180 minutes of the assay. It is however more difficult to determine the metabolite formation ratio for Trog in the bulk incubation as sequential metabolite formation cannot be seen or monitored in the bulk incubation protocol.

Computational insights gained via docking of Piog and Trog with BM3 DM and semi- rationally designed mutants allowed for… Further crystallographic studies… to aid… method validation

Further investigations into the products of the Trog bulk assay are necessary in order to separate out metabolites. Preparative- HPLC could perhaps allow for adequate separation if enough sample is retained.

Comparative studies against human CYPs could be of interest in order to gauge whether BM3 performs metabolism of these drugs at a higher rate than human P450s. In cases where BM3 DM and other mutants outperform human CYPs, quantification and selectivity studies can determine if any BM3 variants may be subject for use in scale- up processes for metabolite production. Mass fermentation protocols have been reported… for expression of BM3 and variants…

Production of new metabolites based on *enz* to tune selectivity of BM3 towards production of specific major human metabolites could play a significant role in the ADME- Tox industry. The reduction of time necessary for *in vitro* assays and mutant determination allowed for more cost and time effective methods of generating mutants of high fitness. Once tested *in vitro,* these mutants may demonstrate ability to selectively catalyse oxidations of a range of complex drug classes…

It is evident that the expression of P450 BM3 and its use in the formation of drug metabolites produced by major human liver isoforms is progressing rapidly. The ease of expression of the ‘gate-keeper’ variant and the ability to separate out major human metabolites highlights the application within the pharmaceutical industry.

The successful production and elucidation of the major human metabolites for Piog and Trog show the scope of BM3 for drug metabolite production. Emerging xenobiotics may be… scale up….

P450 BM3, alongside other bacterial systems are rapidly coming to light and progressing greatly. With the introduction of the computational methods of docking and simulating binding with software such as PyRosetta and AutoDock Vina, the introduction of advantageous mutations is becoming ever more rapid, so the enzyme design pipeline becomes yet more efficient.

Future work

The future perspectives for any work within this project should be viewed in terms of *in vitro* metabolism assays and computational enzyme design to aid *in vitro* studies and engineer BM3 towards selectivity.

In the case of the *in vitro* drug metabolite formation assays, five members from the anti- diabetic drug class were originally planned to be carried out within this project prior to Covid and the subsequent restrictions. Covid statement?

In terms of the use for BM3 variants for the production of human metabolites, further drug classes are to be investigated to demonstrate the range of diverse molecules that BM3 can metabolise. Also further

investigations into probing selectivity of BM3 with the introduction of mutations in high- throughput enzyme design pipeline utilising *enz* and various other computational packages could expand the use of BM3 as a tool.

The *enz* pipeline requires several area of improvement with added layers of complexity and perhaps incorporation of machine- learning packages to improve efficiency and accuracy of the programme.

A high- throughput screen could be developed in order to test the fitness of mutations postulated by *enz*, *in vitro,* with LC-MS/MS as a tool to monitor selectivity of these novel mutants.

Whole- cell assays could be developed in order to test the selectivity and turnover of thesis mutants without the need to optimise methods of protein purification of each variant.

The first mutants designed semi- rationally with the use of *enz* and the Arnold mutant library should be synthesised and tested *in vitro* as a means of method validation. The 75W/188S mutations as single mutations with the 82F/87V DM to form two TMs, and the two new mutations on top of the DM in order to produce a QM should be synthesised and monitored for selectivity in appropriate assays. The data from these assays may be compared with the results of the incubations with BM3 DM in order to track any significant changes in selectivity.

Computational

Introduction of loop- remodelling… introduction of ML from alphafold..

Full automation of the search and mutation generation, heavily reliant on a score function… mutations were based off… constrained due to a known mutant library… Simulation based? Docking?

QM calc with DFT is expensive…..

Wider scope for the project….

References

https://www.nmrdb.org/13c/index.shtml?v=v2.121.0 cnmr met

‘https://www.nmrdb.org/new_predictor/index.shtml?v=v2.121.0 hnmr trog

Alford, R. F. *et al.* (2015) ‘An Integrated Framework Advancing Membrane Protein Modeling and Design’, *PLoS Computational Biology*. Public Library of Science, 11(9). doi: 10.1371/journal.pcbi.1004398.

Backman, J. T. *et al.* (2016) ‘Role of Cytochrome P450 2C8 in drug metabolism and interactions’, *Pharmacological Reviews*. American Society

for Pharmacology and Experimental Therapy, 68(1), pp. 168–241. doi: 10.1124/pr.115.011411.
Bathelt, C. M. *et al.* (2003) 'Aromatic Hydroxylation by Cytochrome P450: Model Calculations of Mechanism and Substituent Effects', *Journal of the American Chemical Society*. American Chemical Society , 125(49), pp. 15004–15005. doi: 10.1021/ja035590q.
Beigneux, A. P. *et al.* (2002) 'Reduction in cytochrome P-450 enzyme expression is associated with repression of CAR (constitutive androstane receptor) and PXR (pregnane X receptor) in mouse liver during the acute phase response', *Biochemical and Biophysical Research Communications*. Academic Press, 293(1), pp. 145–149. doi: 10.1016/S0006-291X(02)00196-1.
Bowman, S. E. J. and Bren, K. L. (2008) 'The chemistry and biochemistry of heme c: functional bases for covalent attachment'. doi: 10.1039/b717196j.
Butler, C. F. *et al.* (2013) 'Key mutations alter the cytochrome P450 BM3 conformational landscape and remove inherent substrate bias', *Journal of Biological Chemistry*, 288(35), pp. 25387–25399. doi: 10.1074/jbc.M113.479717.
Chaudhury, S., Lyskov, S. and Gray, J. J. (2010) 'PyRosetta: A script-based interface for implementing molecular modeling algorithms using Rosetta', *Bioinformatics*. Oxford University Press, pp. 689–691. doi: 10.1093/bioinformatics/btq007.
Coates, G. *et al.* (2002) 'Glitazones regulate glutamine metabolism by inducing a cellular acidosis in MDCK cells', *American Journal of Physiology-Endocrinology and Metabolism*. American Physiological Society, 283(4), pp. E729–E737. doi: 10.1152/ajpendo.00485.2001.
Cook, D. J. *et al.* (2016) 'Cytochromes P450: History, Classes, Catalytic Mechanism, and Industrial Application', in *Advances in Protein Chemistry and Structural Biology*. Academic Press Inc., pp. 105–126. doi: 10.1016/bs.apcsb.2016.07.003.
Finnigan, J. D. *et al.* (2020) 'Cytochromes P450 (P450s): A review of the class system with a focus on prokaryotic P450s', *Advances in Protein*

Chemistry and Structural Biology. Academic Press Inc., 122, pp. 289–320. doi: 10.1016/bs.apcsb.2020.06.005.

Goldstone, J. V. *et al.* (2006) 'The chemical defensome: Environmental sensing and response genes in the Strongylocentrotus purpuratus genome', *Developmental Biology*. Academic Press Inc., 300(1), pp. 366–384. doi: 10.1016/j.ydbio.2006.08.066.

Gomez De Santos, P. *et al.* (2018) 'Selective Synthesis of the Human Drug Metabolite 5′-Hydroxypropranolol by an Evolved Self-Sufficient Peroxygenase', *ACS Catalysis*. American Chemical Society, 8(6), pp. 4789–4799. doi: 10.1021/acscatal.8b01004.

Gonzalez, F. J. *et al.* (1988) 'Characterization of the common genetic defect in humans deficient in debrisoquine metabolism', *Nature*. Nature Publishing Group, 331(6155), pp. 442–446. doi: 10.1038/331442a0.

Groves, J. T. *et al.* (1978) 'Aliphatic hydroxylation by highly purified liver microsomal cytochrome P-450. Evidence for a carbon radical intermediate', *Biochemical and Biophysical Research Communications*, 81(1), pp. 154–160. doi: 10.1016/0006-291X(78)91643-1.

Guengerich, F. P. (2003) 'Cytochrome P450 oxidations in the generation of reactive electrophiles: Epoxidation and related reactions', *Archives of Biochemistry and Biophysics*. Arch Biochem Biophys, pp. 59–71. doi: 10.1016/S0003-9861(02)00415-0.

Guengerich, F. P. (2015) 'Human cytochrome P450 enzymes', in *Cytochrome P450: Structure, Mechanism, and Biochemistry, Fourth Edition*. Springer International Publishing, pp. 523–785. doi: 10.1007/978-3-319-12108-6_9.

Hannemann, F. *et al.* (2007) 'Cytochrome P450 systems-biological variations of electron transport chains', *Biochimica et Biophysica Acta - General Subjects*. Elsevier, pp. 330–344. doi: 10.1016/j.bbagen.2006.07.017.

Hanzlik, R. P. *et al.* (1978) 'Metabolism in vitro of para-substituted styrenes. Kinetic observations of substituent effects', *Biochemical Pharmacology*. Elsevier, 27(10), pp. 1435–1439. doi: 10.1016/0006-2952(78)90098-9.

Hawkes, D. B. *et al.* (2002) ‘Cytochrome P450cin (CYP176A), isolation, expression, and characterization’, *Journal of Biological Chemistry*. American Society for Biochemistry and Molecular Biology, 277(31), pp. 27725–27732. doi: 10.1074/jbc.M203382200.

Holstein, A. *et al.* (2005) ‘Association between CYP2C9 slow metabolizer genotypes and severe hypoglycaemia on medication with sulphonylurea hypoglycaemic agents’, *British Journal of Clinical Pharmacology*. John Wiley & Sons, Ltd, 60(1), pp. 103–106. doi: 10.1111/j.1365-2125.2005.02379.x.

Jerina, D. M. and Daly, J. W. (1974) ‘Arene oxides: A new aspect of drug metabolism’, *Science*. American Association for the Advancement of Science, 185(4151), pp. 573–582. doi: 10.1126/science.185.4151.573.

Jones, J. P., Mysinger, M. and Korzekwa, K. R. (2002) ‘Computational models for cytochrome P450: A predictive electronic model for aromatic oxidation and hydrogen atom abstraction’, *Drug Metabolism and Disposition*, 30(1), pp. 7–12. doi: 10.1124/dmd.30.1.7.

Joyce, M. G. *et al.* (2012) ‘The crystal structure of the FAD/NADPH-binding domain of flavocytochrome P450 BM3’, *FEBS Journal*. John Wiley & Sons, Ltd, 279(9), pp. 1694–1706. doi: 10.1111/j.1742-4658.2012.08544.x.

Kaufmann, K. W. *et al.* (2010) ‘Practically useful: What the R osetta protein modeling suite can do for you’, *Biochemistry*. American Chemical Society, pp. 2987–2998. doi: 10.1021/bi902153g.

Kawaguchi-Suzuki, M. and Frye, R. F. (2013) ‘Current clinical evidence on pioglitazone pharmacogenomics’, *Frontiers in Pharmacology*, 4 NOV. doi: 10.3389/fphar.2013.00147.

Kille, S. *et al.* (2011) ‘Regio- and stereoselectivity of P450-catalysed hydroxylation of steroids controlled by laboratory evolution’, *Nature Chemistry*. Nat Chem, 3(9), pp. 738–743. doi: 10.1038/nchem.1113.

Kortemme, T., Kim, D. E. and Baker, D. (2004) *Computational Alanine Scanning of Protein-Protein Interfaces*. Available at: www.stke.org/cgi/content/full/sigtrans;2004/219/pl2http://stke.sciencemag.org/Downloadedfromwww.stke.org/cgi/content/full/sigtrans;2004/219/pl2

(Accessed: 27 December 2020).
Linder, M. (2012) ‘Computational enzyme design: Advances, hurdles and possible ways forward’, *Computational and Structural Biotechnology Journal*. Research Network of Computational and Structural Biotechnology, 2(3), p. e201209009. doi: 10.5936/csbj.201209009.
Lu, J. *et al.* (2020) ‘New insights of CYP1A in endogenous metabolism: a focus on single nucleotide polymorphisms and diseases’, *Acta Pharmaceutica Sinica B*. Chinese Academy of Medical Sciences, pp. 91–104. doi: 10.1016/j.apsb.2019.11.016.
Luthra, A., Denisov, I. G. and Sligar, S. G. (2011) ‘Temperature Derivative Spectroscopy To Monitor the Autoxidation Decay of Cytochromes P450’, *Anal. Chem*, 83, pp. 5394–5399. doi: 10.1021/ac2009349.
Metabolism, (K H and Chemistry, T. F. W. (2004) *METABOLIC ACTIVATION OF TROGLITAZONE: IDENTIFICATION OF A REACTIVE METABOLITE AND MECHANISMS INVOLVED*. Available at: http://dmd.aspetjournals.org (Accessed: 14 October 2020).
Mohs, R. C. and Greig, N. H. (2017) ‘Drug discovery and development: Role of basic biological research’, *Alzheimer’s and Dementia: Translational Research and Clinical Interventions*. Elsevier Inc, pp. 651–657. doi: 10.1016/j.trci.2017.10.005.
Munro, A. W. *et al.* (2002) ‘P450 BM3: The very model of a modern flavocytochrome’, *Trends in Biochemical Sciences*. Elsevier Current Trends, pp. 250–257. doi: 10.1016/S0968-0004(02)02086-8.
Munro, A. W. *et al.* (2018) ‘Structure and function of the cytochrome P450 peroxygenase enzymes’, *Biochemical Society Transactions*. Portland Press Ltd, pp. 183–196. doi: 10.1042/BST20170218.
Naik, B. *et al.* (2020) ‘High throughput virtual screening reveals SARS-CoV-2 multi-target binding natural compounds to lead instant therapy for COVID-19 treatment’, *International Journal of Biological Macromolecules*. Elsevier B.V., 160, pp. 1–17. doi: 10.1016/j.ijbiomac.2020.05.184.
Di Nardo, G. and Gilardi, G. (2012) ‘Optimization of the bacterial cytochrome P450 BM3 system for the production of human drug metabolites’, *International Journal of Molecular Sciences*. MDPI AG, pp.

15901–15924. doi: 10.3390/ijms131215901.
Nelson, D. R. *et al.* (2004) ‘Comparison of cytochrome P450 (CYP) genes from the mouse and human genomes, including nomenclature recommendations for genes, pseudogenes and alternative-splice variants’, *Pharmacogenetics*. Pharmacogenetics, pp. 1–18. doi: 10.1097/00008571-200401000-00001.
Nelson, D. R. (2018) ‘Cytochrome P450 diversity in the tree of life’, *Biochimica et Biophysica Acta - Proteins and Proteomics*. Elsevier B.V., 1866(1), pp. 141–154. doi: 10.1016/j.bbapap.2017.05.003.
Nelson, D. R., Goldstone, J. V. and Stegeman, J. J. (2013) ‘The cytochrome P450 genesis locus: The origin and evolution of animal cytochrome P450s’, *Philosophical Transactions of the Royal Society B: Biological Sciences*. Royal Society, 368(1612). doi: 10.1098/rstb.2012.0474.
Ortiz de Montellano, P. R. and De Voss, J. J. (2002) ‘Oxidizing species in the mechanism of cytochrome P450’, *Natural Product Reports*. The Royal Society of Chemistry, 19(4), pp. 477–493. doi: 10.1039/b101297p.
Poulos, T. L. (2014) ‘Heme enzyme structure and function’, *Chemical Reviews*. American Chemical Society, pp. 3919–3962. doi: 10.1021/cr400415k.
Prasad, B. *et al.* (2011) ‘Metabolite identification by liquid chromatography-mass spectrometry’, *TrAC - Trends in Analytical Chemistry*. Elsevier, pp. 360–387. doi: 10.1016/j.trac.2010.10.014.
Puchkaev, A. V. and Ortiz De Montellano, P. R. (2005) ‘The Sulfolobus solfataricus electron donor partners of thermophilic CYP119: An unusual non-NAD(P)H-dependent cytochrome P450 system’, *Archives of Biochemistry and Biophysics*. Academic Press Inc., 434(1 SPEC. ISS.), pp. 169–177. doi: 10.1016/j.abb.2004.10.022.
Rendic, S. (2002) ‘Summary of information on human CYP enzymes: Human P450 metabolism data’, *Drug Metabolism Reviews*, pp. 83–448. doi: 10.1081/DMR-120001392.
Roberts, G. A. *et al.* (2002) ‘Identification of a new class of cytochrome P450 from a Rhodococcus sp.’, *Journal of Bacteriology*. American Society for Microbiology (ASM), 184(14), pp. 3898–3908. doi:

10.1128/JB.184.14.3898-3908.2002.
Roitel, O., Scrutton, N. S. and Munro, A. W. (2003) ‘Electron transfer in flavocytochrome P450 BM3: Kinetics of flavin reduction and oxidation, the role of cysteine 999, and relationships with mammalian cytochrome P450 reductase’, *Biochemistry*. Biochemistry, 42(36), pp. 10809–10821. doi: 10.1021/bi034562h.
Rupasinghe, S. *et al.* (2006) ‘The cytochrome P450 gene family CYP157 does not contain EXXR in the K-helix reducing the absolute conserved P450 residues to a single cysteine’. doi: 10.1016/j.febslet.2006.10.043.
Sakamoto, J. *et al.* (2000) ‘Activation of human peroxisome proliferator-activated receptor (PPAR) subtypes by pioglitazone’, *Biochemical and Biophysical Research Communications*. Academic Press Inc., 278(3), pp. 704–711. doi: 10.1006/bbrc.2000.3868.
Shaik, S. *et al.* (2011) ‘Trends in aromatic oxidation reactions catalyzed by cytochrome P450 enzymes: A valence bond modeling’, *Journal of Chemical Theory and Computation*. American Chemical Society, 7(2), pp. 327–339. doi: 10.1021/ct100554g.
Shiro, Y. *et al.* (1989) ‘15N NMR Study on Cyanide (C15N-) Complex of Cytochrome P-450cam. Effects of D-Camphor and Putidaredoxin on the Iron—Ligand Structure’, *Journal of the American Chemical Society*. American Chemical Society, 111(20), pp. 7707–7711. doi: 10.1021/ja00202a007.
Shringari, S. R. *et al.* (2020) ‘Rosetta custom score functions accurately predict ΔΔ: G of mutations at protein-protein interfaces using machine learning’, *Chemical Communications*. Royal Society of Chemistry, 56(50), pp. 6774–6777. doi: 10.1039/d0cc01959c.
Singh, S. (2006) ‘Preclinical Pharmacokinetics: An Approach Towards Safer and Efficacious Drugs’, *Current Drug Metabolism*. Bentham Science Publishers Ltd., 7(2), pp. 165–182. doi: 10.2174/138920006775541552.
Stjernschantz, E. *et al.* (2008) ‘Structural rationalization of novel drug metabolizing mutants of cytochrome P450 BM3’, *Proteins: Structure, Function, and Bioinformatics*. John Wiley & Sons, Ltd, 71(1), pp. 336–352. doi: 10.1002/prot.21697.

Sun, Y. *et al.* (2013) ‘Investigations of heme ligation and ligand switching in cytochromes P450 and P420’, *Biochemistry*. Biochemistry, 52(34), pp. 5941–5951. doi: 10.1021/bi400541v.

Tantillo, D. J., Chen, J. and Houk, K. N. (1998) ‘Theozymes and compuzymes: Theoretical models for biological catalysis’, *Current Opinion in Chemical Biology*. Elsevier Ltd, 2(6), pp. 743–750. doi: 10.1016/S1367-5931(98)80112-9.

Trott, O. and Olson, A. J. (2009) ‘AutoDock Vina: Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading’, *Journal of Computational Chemistry*. Wiley, 31(2), p. NA-NA. doi: 10.1002/jcc.21334.

Tsamandouras, N. *et al.* (2017) ‘Quantitative assessment of population variability in hepatic drug metabolism using a perfused three-dimensional human liver microphysiological system’, *Journal of Pharmacology and Experimental Therapeutics*. American Society for Pharmacology and Experimental Therapy, 360(1), pp. 95–105. doi: 10.1124/jpet.116.237495.

Vaz, A. D. N., McGinnity, D. F. and Coon, M. J. (1998) ‘Epoxidation of olefins by cytochrome P450: Evidence from site-specific mutagenesis for hydroperoxo-iron as an electrophilic oxidant’, *Proceedings of the National Academy of Sciences of the United States of America*. National Academy of Sciences, 95(7), pp. 3555–3560. doi: 10.1073/pnas.95.7.3555.

Visser, L. E. *et al.* (2007) ‘The risk of myocardial infarction in patients with reduced activity of cytochrome P450 2C9’, *Pharmacogenetics and Genomics*, 17(7), pp. 473–479. doi: 10.1097/01.fpc.0000236335.57046.c8.

Wilkinson, G. R. (1996) ‘Cytochrome P4503A (CYP3A) metabolism: Prediction of in vivo activity in humans’, *Journal of Pharmacokinetics and Biopharmaceutics*. J Pharmacokinet Biopharm, pp. 475–490. doi: 10.1007/BF02353475.

Wright, R. L. *et al.* (1996) ‘Cloning of a potential cytochrome p450 from the archaeon Sulfolobus solfataricus’, *FEBS Letters*. Elsevier B.V., 384(3), pp. 235–239. doi: 10.1016/0014-5793(96)00322-5.

Wrighton, S. A. and Stevens, J. C. (1992) ‘The human hepatic cytochromes p450 involved in drug metabolism’, *Critical Reviews in Toxicology*. Informa

Healthcare, 22(1), pp. 1–21. doi: 10.3109/10408449209145319.
Xu, C., Li, C. Y. T. and Kong, A. N. T. (2005) ‘Induction of phase I, II and III drug metabolism/transport by xenobiotics’, *Archives of Pharmacal Research*. Pharmaceutical Society of Korea, pp. 249–268. doi: 10.1007/BF02977789.
Zanger, U. M. and Schwab, M. (2013) ‘Cytochrome P450 enzymes in drug metabolism: Regulation of gene expression, enzyme activities, and impact of genetic variation’, *Pharmacology and Therapeutics*. Pergamon, pp. 103–141. doi: 10.1016/j.pharmthera.2012.12.007.
Zhang, X. *et al.* (2019) ‘Unraveling the regulation of hepatic gluconeogenesis’, *Frontiers in Endocrinology*. Frontiers Media S.A., p. 802. doi: 10.3389/fendo.2018.00802.
Zhang, Z. and Tang, W. (2018) ‘Drug metabolism in drug discovery and development’, *Acta Pharmaceutica Sinica B*. Chinese Academy of Medical Sciences, pp. 721–732. doi: 10.1016/j.apsb.2018.04.003.
Zhou, S.-F., Liu, J.-P. and Chowbay, B. (2009) ‘Polymorphism of human cytochrome P450 enzymes and its clinical impact’, *Drug Metabolism Reviews*. Taylor & Francis, 41(2), pp. 89–295. doi: 10.1080/03602530902843483.

Appendix

Sequences?

Turnover number calculations for initial slope…

Any computational supplementary????