Computational studies to elucidate the role of proteins in the prevention of malaria

A thesis submitted to the University of Manchester for the degree of Doctor of Philosophy in the Faculty of Engineering and Physical Sciences

2010

Jaclyn Bibby

School of Chemical Engineering and Analytical Science

Contents

Word count 72,194

Figures

12

13

19

Tables

Abstract

Jaclyn Bibby

Computational studies to elucidate the role of proteins in the prevention of malaria

This thesis was submitted to the University of Manchester for the degree of Doctor of Philosophy (PhD) in 2010.

Malaria is a disease that affects half the worlds population and is caused by a parasite spread by mosquito. The control of malaria relies on the use of insecticide treated nets to prevent transmission by mosquito bite. Nets alone act as poor barriers to insect bites and the insecticides used to treat the net have a role in killing the insect on contact. As the nets are in frequent contact with people, pyrethroids are the only class of insecticide recommended for use due to its low toxicity to mammals. Insecticide resistance to pyrethroids can occur in insects reducing the effectiveness of the nets to prevent transmission, and insects can become resistant to insecticides by the over expression of cytochrome P450 enzymes. P450 enzymes are a super family of heme containing enzymes involved in the detoxification of drugs and xenobiotics. This study uses computational modelling techniques to give an insight into the structures of the P450s involved in the detoxification of these insecticides. These computational studies complement experimental work and give an understanding of experimental results by giving an insight at the atomic level into the structures of these enzymes. The computational models give explanations for the observed results, but are also predictive and can be used to guide experimental studies. In this study, homology modelling and bioinformatics was used to build structural models of the P450s. These models were validated structurally to ensure that the proteins were correctly folded, docking studies were used to ensure that there was a good correlation between the experimental and computational results. A number of P450s are overexpressed in insecticide resistant mosquitoes these were studied to understand their ability to bind to pyrethroids and comparisons were made to P450s incapable of metabolism. Based on this, a fingerprint for metabolism was constructed that may be used to predict the capacity for metabolism in unknown P450s by identifying residues involved in metabolism. The models produced can be used to explain the profiles of metabolites observed to be produced by these enzymes. The studies on CYP6M2 investigate its ability to metabolise pyrethroids and in particular its metabolism pathway for deltamethrin. It was shown to metabolise deltamethrin at specific sites that can be explained by the models produced in this thesis. The models can also explain the specificity of the enzymes towards a number of fluorescent substrates by identifying the residues that have a steric influence. The models can be used to guide the development of novel pyrethroids that are resistant to metabolism. In addition, the models identified factors external to the active site that influence the metabolism of pyrethroids including its interaction with binding partners and the membrane as well as ligand access to the buried active site. Such factors explain the selectivity of enzymes for the logP of their substrates. These models were used to design probes specific to metabolising enzymes that can be used to identify novel P450s involved in insecticide resistance, and could be used to monitor resistance in insect populations. In the human host, toll like receptors are involved in sensing the malaria parasite and initiate an inflammatory response, an excessive inflammatory response can lead to severe forms of malaria. Mal has a central role in this pathway and the affect of malaria on the human host can be determined by the variant of this protein. Understanding the role of Mal can lead to the identification of targets for drugs that can modulate the immune response and prevent hyper inflammatory disorders.

<u>Declaration</u>

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

<u>Publications from this Thesis.</u>

L. A. Mclaughlin, U. Niazi, J. Bibby, J.-P. David, J. Vontas, J. Hemingway, H. Ranson, M. J. Sutcliffe and M. J. I. Paine, Characterization of inhibitors and substrates of Anopheles gambiae CYP6Z2. Insect molecular biology, 2008. 17(2): 125-35.

1. Introduction

1.1 Vector borne diseases

1.1.1 Malaria

According to the World Heath Organisation (WHO), half the world's population of around 3.3 billion people live in areas that have a risk of malaria transmission; 109 countries are classified as endemic or at risk (Figure 1.1). In 2006 there were an estimated 247 million malaria cases, and nearly a million deaths with 81% of these among children under 5 years. Malaria is caused by parasites of the *Plasmodium* genus and is spread from person to person through the bites of infected *Anopheles* mosquitoes. Severe malaria occurs when the infection is complicated by blood abnormalities and organ failures. Destruction of red blood cells, hemolysis, can cause anemia. Cerebral malaria is associated with abnormal behavior, seizures and coma. Severe malaria can also manifest as pulmonary oedema, acute respiratory distress syndrome, cardiovascular collapse and shock.

*, China, Indonesia, Philippines, Solomon Islands, Sudan, Vanu-atu and Yemen have localized malaria-free projects

Certified malaria-free and/or no ongoing local transmission for over a decade

Prevention of reintroduction

Elimination

Pre-elimination

Control
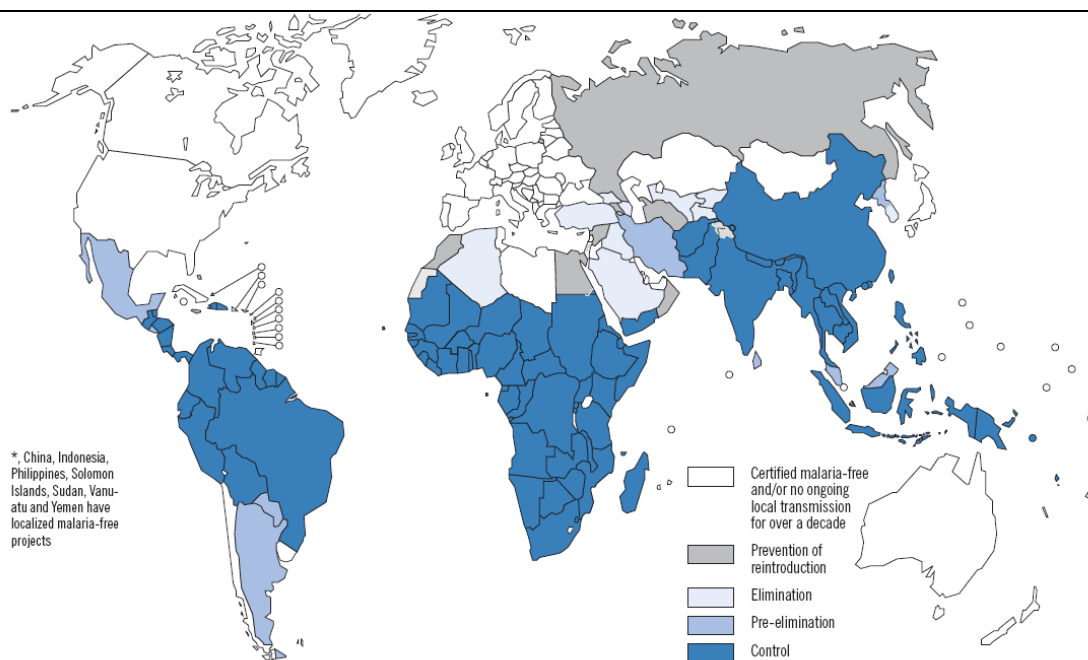
Figure 1.1 Countries where malaria is endemic (taken from World Health Organization world malaria report, 2008).

1.1.2 Dengue

According the WHO, Dengue is the most common mosquito borne viral disease. The *Aedes aegypti* mosquito is the main vector of transmission and its distribution is comparable to that of malaria, 2.5 billion people live in areas where Dengue is transmitted (Figure 1.2). The dengue virus is a Flaviviridae (single stranded non-segmented RNA virus) with four distinct serotypes. Infection with one serotype does not provide immunity to the others so that people living in dengue-endemic areas can

have up to four infections, and sequential infection increases the risk of more severe disease (Gubler, 2006).
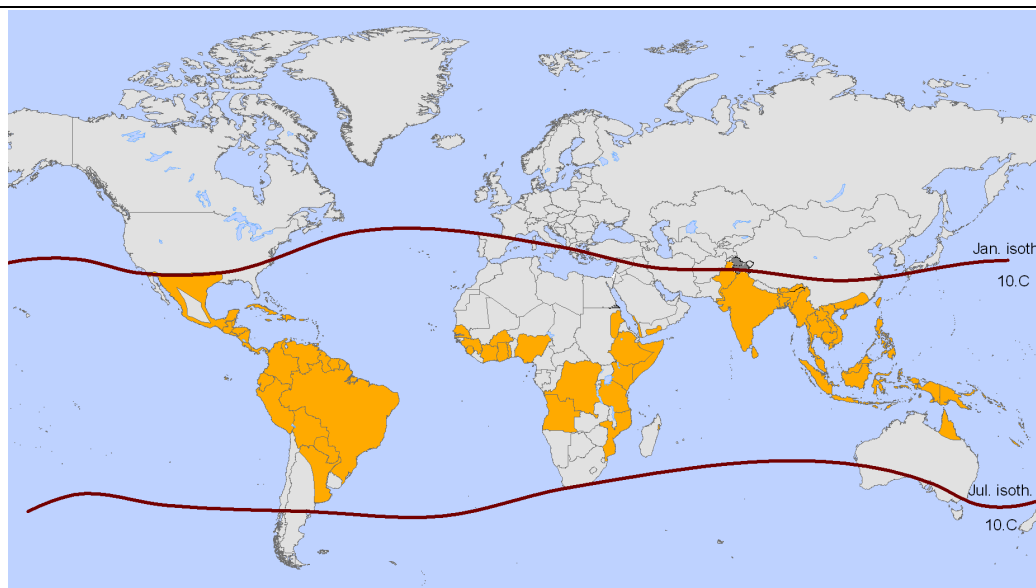
Figure 1.2 Areas at risk of dengue transmission, (taken from Nathan and Dayal-Drager, 2006). The contour lines of the January and July isotherms indicate geographical limits.

The Dengue virus causes a severe flu-like illness called Dengue Fever (DF) usually consisting of a fever and a rash in infants or high fever, headache, muscle and joint pain in older children and adults. Dengue Hemorrhagic fever (DHF) is a potentially fatal complication characterised by high fever, enlargement of the liver, circulatory failure, and hemorrhagic phenomena such as bleeding from the nose, gums and gastro intestinal tract or injection sites. Each year there are around 50-100 million cases of DF and 250-500,000 cases of DHF. The fatality rate of DHF is around 5% with most of these among children and young adults. Dengue has become a major cause of mortality in children especially in South-East Asia. The number of cases of DF and DHF is currently growing with a record number of cases occurring in the Americas in 2002 with 968,723 cases and 234 deaths. The most common causes of death from DHF include bleeding into the vital organs, massive bleeding, and shock caused by the loss of intra-vascular fluids due to increased capillary permeability (Lloyd, 2003).

1.2 Vectors

1.2.1 *Anopheles*

The *Anopheles gambiae* mosquito is a vector for *Plasmodium* parasites that cause malaria. It is highly anthropophilic with a distribution throughout sub-Saharan Africa. Sattler et al. (2005) found that the preferred breeding sites of *A. gambiae* were slopes to

riverbeds, riverbeds, borders of swamps, stagnant drains and rivers, and valleys with vegetable gardens and crops. They also found *A. gambiae* breeding in a sewage pond and one swamp extremely polluted with organic matter. It was suggested that it is likely that anopheles can breed in every type of water accumulation. The habitat preferences of *Anopheles gambiae* have been found to have changed over the 20th century as a result of ecological change. Chinery (1984) found that breeding in water filled man made containers has increased from 1.97% in 1911 to 21% of all breeding in the 1980s. This was also accompanied by an increase in breeding in polluted habitats from 5.3% to 25% of all breeding. *Anopheles* is also anthropophagic in that it prefers to bite humans or cattle; this and its habitat preferences allow it to act as a vector for disease transmission.

1.2.2. *Aedes*

The *Aedes aegypti* mosquito has a worldwide distribution between latitudes 45°N and 35°S. It is a known vector of the yellow fever and dengue viruses. *Ae aegypti* tends to occur in urban areas due to its breeding and behavioral preferences. It breeds in artificial containers, such as water storage containers, neglected cups and jugs that contain fresh water (Simard et al., 2005). There has been an increase in the occurrence of dengue since 1970, from 9 countries affected to 60 by 1999, caused by an increase in urbanisation, poor urban infrastructure and unreliable water supply that has promoted the use of water storage and the accumulation of discarded water containers (Kay and Vu, 2005). *Ae aegypti* also prefers to oviposit, rest and bite hosts indoors, and does not require vegetation for sugar feeding. *Ae aegypti* is therefore anthropophilic and tends to be closely associated with humans (Braks et al., 2003). *Aedes aegypti* is found worldwide in the tropics and is an efficient vector because it is susceptible to the dengue virus, preferentially feeds on human blood, is a daytime feeder with an imperceptible bite and can bite many people in a short period for one blood meal. Unlike malaria, dengue occurs primarily in urban areas and *Aedes* is adapted to urban habitats as it breeds in clean stagnant water in man made containers. At present the only method of controlling dengue is to control the vector mosquitoes. Most dengue control programs rely on the use of insecticides to control the larval and adult stages of *Aedes aegypti* (Wilder-Smith and Schwartz, 2005).

1.3 Insecticide treated nets

Insecticide treated nets (ITNs) are a vector control tool against malaria. Nets have long been used as a method of protection against mosquitoes. However, nets are limited as a physical barrier as they can become torn, or they can be hung in a way that allows mosquitoes to bite through them. To prevent this, nets can be treated with insecticides and treated nets offer greater protection than untreated nets by irritating, repelling or killing the mosquitoes as they contact the net.

The requirements of insecticides for the treatment of nets is strict due to the frequent contacts with people. The insecticide must be non-toxic to children and be non-irritating to the skin. According to the WHO Communicable Disease Control, Prevention and Eradication WHO Pesticide Evaluation Scheme (WHOPES, 2002), pyrethroids are the only insecticides recommended for the treatment of mosquito nets. They satisfy this criteria by having a low volatility to reduce inhalation toxicity, no teratogenic, carcinogenic, or mutagenic effects and low toxicity to mammals (Najera and Zaim, 2002). Pyrethroids affect mosquitoes with knockdown and killing effects at doses below the threshold for mammalian toxicity. In addition the pyrethroids recommended by WHOPES (Table 1.1) all have an exito-repellent effect on the vector species, and the presence of one net in a room may partially protect individuals outside the net.

Table 1.1 The insecticides recommended for the treatment of nets (taken from WHOPES, 2002).

Insecticide	Dose (mg active ingredient/m^2 of netting)
α-cypermethrin	20-40
Cyfluthrin	50
Deltamethrin	15-25
Etofenprox	200
Lamda-cyhalothrin	10-15
Permethrin	200-500

1.4 History of pyrethroid development

Pyrethroids were originally found in the flowers of the herbaceous perennials *Pyrethrum cinerariifolium* and *Chrysanthemum cinerariifolium.* Pyrethroids have been found to mostly act on the sodium channel; insect sodium channels are 100 fold more sensitive to pyrethroids than mammalian channels, and partly because of this, mammals tend to be resistant to pyrethroids (Ray and Forshaw, 2000)

Pyrethroids are the oldest class of insecticides. Pyrethrum flowers or their extracts have been used for centuries. In the 1800s they were used by the Caucasian tribes to control lice, by 1828 the flowers were being produced commercially in Armenia and by 1840 in Dalmatia. Since the 1920s the insecticidal components, pyrethrins, were extracted with kerosene to produce liquid sprays. After the Second World War, production continued in East Africa and South America and by the 1980s pyrethrum production had reached 15,000 tons of dried flowers per year. However, the dried flowers contain only 1-2% pyrethrins by dry weight producing 150-200 tons per year. This production was insufficient and provided only half of the world market demand (Casida, 1980).

The first development of synthetic pyrethroids was a result of wartime research in the US and the UK to find substitutes to minimise dependency on imported natural materials, where shipping channels may be disrupted. Allethrin was the first commercial pyrethroid produced from this, and further pyrethroids were developed until 1965 that were pyrethrum substitutes. These pyrethrins and pyrethrin substitutes were structurally similar to the natural pyrethrins and can undergo photooxydation at the isobutenyl, furan and allyl substituents, and were unstable for agricultural use (Figure 1.3a). The development of pyrethroids for crop protection involved replacement of the photolabile isobutenyl methyls with stable functions (Figure 1.3b), and the inclusion of a phenoxybenzyl alcohol group which increased activity and photostability (Figure 1.3c) (Casida, 1980; Elliott, 1980). Metabolism of the insecticide was taken into account in the design of insecticides, oxidation at the isobutenyl methyls was prevented by replacement with a dihalovinyl. As the structure of the pyrethroid was altered, its metabolism in insects was also altered.

1.5 Pyrethroid structure

Almost all pyrethroids are esters, composed of an acid and alcohol groups connected by an ester bond. The acid tends to be a substituted cyclopropane carboxylic acid. The alcohol groups consist of pyrethrolone for the pyrethrins and phenoxybenzyl groups for the pyrethroids. The pyrethroids with a cyclopropane acid group have two common chiral centres at carbons 3 and 1 (Figure 1.4), and isomers differing at either of these centres have varying toxicity.
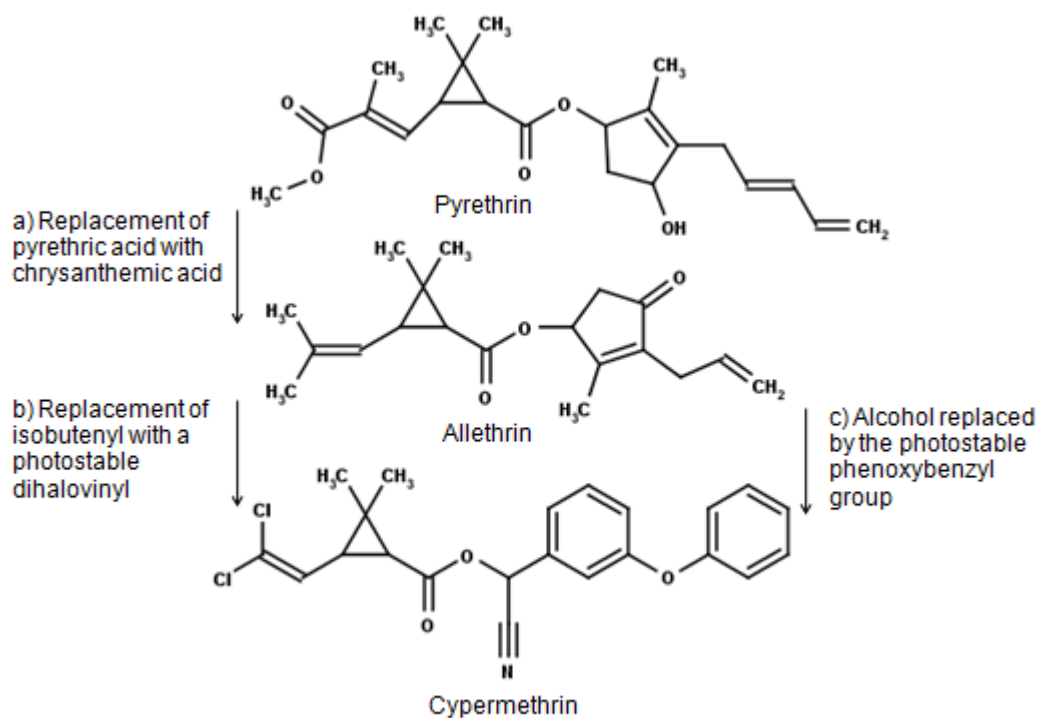
Figure 1.3 The substitution of pyrethrin acid and alcohol groups with photo stable groups.
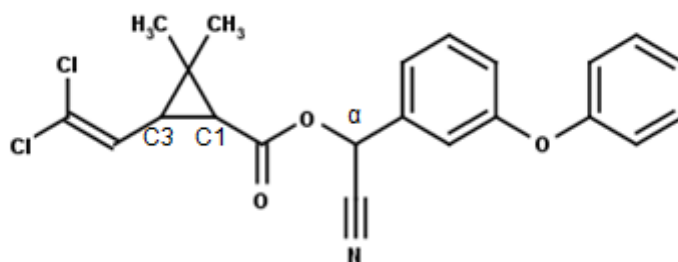
Figure 1.4 The positions of the C1, C3 and α carbons in cypermethrin.

Isomers differing at the 1 carbon are termed 1R and 1S. The toxicity of these isomers vary according to the structure of the alcohol group, the C3 substituent (a halovinyl or isobutenyl) and the species tested. Isomers differing at the 3 carbon are termed trans or cis depending on their orientation relative to the 1 carbon. The acid moieties of the naturally occurring pyrethroids are exclusively in the 1R trans configuration. Pyrethroids derived from (1S)cyclopropanecarboxylic acids have been found to be less active than those derived from (1R) acids (Anderson et al., 1985).

The trans chrysanthamates can be more active than the cis isomers (Elliott, 1971). However, where the isobutenyl group of the crysanthamic acid is replaced by a

halovinyl group, the activity of the cis and trans may vary, but generally where a halovinyl group is present, the cis forms of pyrethroids are more toxic than the trans forms. Alzogaray et al. (1998) compared isomer mixtures of cis and trans permethrin against the bug *Triatoma infestans*. They found that the cis isomer was 3 fold more effective than the trans in the first instar, 25 fold in the third instar, and 54 fold in the fifth instar larvae. Stereospecificity also occurs in the alcohol moiety. The alpha carbon can become a chiral centre by the inclusion of an α-cyano group; when this occurs only one isomer has insecticidal activity. The (S) α-cyano-3-phenoxybenzyl alcohol has a higher activity than the R isomer (Anderson et al., 1985).

1.6 Pyrethroid resistance

1.6.1 Housefly pyrethroid resistance

Insecticide resistance is defined as the ability of some individuals to tolerate doses of a toxicant that would prove lethal to the majority of individuals in a normal population of the same organism. The learn PyR (LPR) strain was collected from a dairy near Horseheads NY in 1980 and selected with permethrin until 6000-fold resistance was attained. It was also shown to have 100,000-fold resistance to deltamethrin and has been extensively studied because of its resistance. This strain has been shown to be resistant due to three mechanisms: increased detoxification by P450s, decreased cuticular penetration, and target site insensitivity (Scott and Georghiou, 1986).

Scott and Georghiou (1986) found that in LPR there is slower penetration of permethrin than in the susceptible NAIDM strain, with 60% remaining external in LPR compared to 49% in NAIDM. Slower penetration may be a mechanism of resistance in housefly.

Scott and Georghiou (1986) found that LPR was highly resistant to knockdown, a characteristic of target site insensitivity, and found LPR to have a low neuronal sensitivity to permethrin. Target site insensitivity was determined by applying permethrin and monitoring for burst discharges in the dorsal longitudinal muscle, a marker for neuronal dysfunction. The threshold dose causing burst discharges was found to be greater in LPR than that of the Super-KDR strain, homozygous for the knock down resistance (kdr) gene.

In LPR pyrethroid resistance is also associated with elevated P450 activity, demonstrated by the synergistic affect of piperonyl butoxide, and also with increases in P450, b5 and P450 reductase activity (Scott and Georghiou, 1986).

1.7 P450s

Cytochromes P450, named after their absorption at 450 nm, are one of the largest superfamilies of proteins. They are a superfamily of heme containing enzymes that are involved in the metabolism of endogenous and exogenous compounds. They are involved in the classical 'Phase I' metabolism in which a substrate is oxidised. During Phase II other enzymes use this oxygen for further metabolism. Potentially toxic drugs and xenobiotics can be detoxified by this oxidation and be made more water soluble to increase its elimination by the kidney (Spatzenegger and Jaeger, 1995).

1.7.1 P450 structure

The nomenclature system was introduced by Nebert et al. (1997) which proposes that members of a family share >40% identity and members of a sub family >55% identity. CYP is used as a prefix, followed by a number for the family, a letter for the subfamily and a number for the individual gene. Sequence identity among P450s can be as low as 20% with only three absolutely conserved amino acids. Despite high sequence variation, there is a high degree of conservation of overall fold with the highest conservation around the heme, indicating a conserved mechanism of electron transfer and oxygen activation. The conserved fold consists of a core bundle formed by four helices D,E,I and L, helices J and K, two sets of beta sheet, and a coil called the 'meander' (Figure 1.5). These regions contain conserved consensus sequences; (F-X-X-G-X-R-X-C-X-G) is part of the heme-binding loop and occurs on the L helix, the conserved Cys is the heme iron's fifth ligand, and the heme iron being bonded to the four nitrogens of the prophyrin. (E-X-X-R) occurs on the K helix and stabilises the core. (A/G-G-G-X-D/E-T-T/S) is the P450 signature and corresponds to the electron transfer groove of the I helix. P450s contain variable regions that are associated with the N-terminal transmembrane helix, and substrate binding and recognition (substrate recognition sites) (Werck-Reichhart and Feyereisen, 2000).
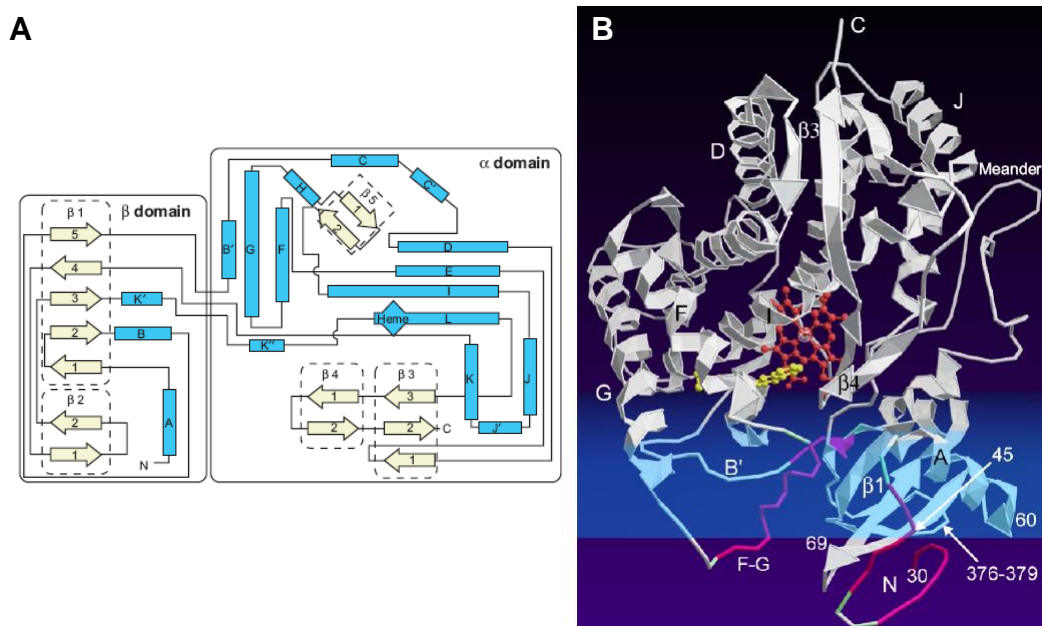
Figure 1.5 (A) Topology diagram of the secondary structure of a typical P450. A-helices: blue boxes; arrows outlined with doted lines: β-sheets. (B) Ribbon diagram of the distal face of CYP2C5 showing its association with the membrane (purple). Taken from Werck-Reichhart and Feyereisen (2000).

1.7.2 History of P450 structure determination

The first crystal structure for a P450, P450cam (CYP101) (Poulos et al., 1985) was published in 1985. It was several years before the next structure, P450BMP (CYP102) was published in 1993 (Ravichandran et al., 1993) and until 2000, the only crystal structures available were microbial P450s. Crystalisation of membrane bound mammalian P450s was difficult due to their hydrophobicity. The first eukaryotic microsomal P450 was CYP2C5 (Williams et al., 2000b), which had to be modified to promote crystallization. Crystallisation was only achieved after engineering to improve its solubility, the N-terminal 22 residues forming the membrane binding helix was removed and the FG loop between residues 201 and 210 replaced with those from P4502C3 (Cosme and Johnson, 2000). With the availability of these structures, it was shown that there is a conserved structural fold even though the sequence identity may be less than 20% and with only three absolutely conserved amino acids (Graham and Peterson, 1999).

1.7.3 Substrate Recognition Sites (SRS)

Although P450s tend to show an overlap in the substrates that they bind, they also show distinct substrate specificities. Gotoh (1992) attempted to identify the parts of a P450

that are involved in substrate recognition or binding, that determine specificity. Six substrate recognition sites (SRS) were identified that constitute 16% of the total residues of a P450. These are: SRS1, the B' area; SRS2, the C terminus of F-helix; SRS3, the N terminus of the G-helix; SRS4, the N terminal half of the I-helix; SRS5, the β3 area, and SRS6 the central region of β5 (Gotoh, 1992). Hyper-variable regions within these six SRSs can account for the variations in substrate specificities of P450s. This assignment of SRS regions has been supported by finding direct counterparts on other P450s and the creation of point mutations in these areas that significantly affect substrate specificity.

Investigations into substrate selectivity have involved a number of approaches including site directed mutagenesis, and homology modelling. Homology modelling is needed as the crystal structures of only a few P450s have been determined. In the absence of a crystal structure homology models can be used to predict the structure of the P450 and the positions of residues involved in substrate binding. Homology modelling has been used successfully to interpret the results of mutagenesis studies and to guide mutagenic studies to alter substrate specificity (de Groot, 2006).

1.7.4 Common active site binding

There are a number of examples where biochemical pathways contain P450s in different families capable of catalysing the same reaction. Rupasinghe et al. (2003) identified four highly divergent P450s that were capable of catalysing reactions in the phenylpropanoid pathway, the plant biosynthetic pathway for the formation of lignins, flavonoids and anthocyanins. These were CYP75A5, CYP84A1, CYP75B1 and CYP8A3. Although the primary sequence identities could be as low as 13% they identified a common substrate recognition mechanism among them in the orientation of the substrate, location of residues in the SRS regions that contact the substrate as well as sequence similarities in the SRS regions.

It was found that sequence comparisons among the SRS regions showed conservations, particularly in the SRS5 that contacts the substrate. In SRS5 they found a highly conserved sequence that has a sequence identity significantly higher than in comparisons of the full length proteins or other SRS regions. They predicted that the conserved residues in the SRS5 region would come within 4 Å of the substrate and contact the substrate. These residues were embedded in a Pro-rich sequence that

produced kinks in the backbone to project them into the active site (Rupasinghe et al., 2003).

The P450s that catalyse the metabolism of pyrethroids may also have some active site commonalities with conserved residues in the SRS regions. The conservations may be used to identify a pyrethroid metabolism 'fingerprint' that could be used to identify other metabolisers to guide experimental investigation.

1.7.5 The P450 cycle

Cytochrome P450 are a superfamily involved in oxidative metabolism of endogenous and exogenous chemicals. They primarily function as mono-oxygenases, activating molecular oxygen and inserting a single oxygen atom into the substrate:

$$NADPH + H + O_2 + R \rightarrow NADP^+ + H_2O + RO$$
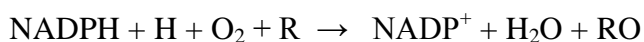
The enzyme catalyses the insertion of oxygen atoms between C-H, C-C, C-N and N-H bonds, and can mediate epoxidations, S-oxidations, H-hydroxylation and dealkylations. The input of two reducing equivalents ($2e^-$) and two proton equivalents are required. These are usually derived from HADPH or NADH, and transferred via a flavoprotein (Shaik et al., 2005).

There are different electron transfer chains and P450s can be classified accordingly (Hannemann et al., 2007):

Class I require both an FAD-containing reductase and an iron sulfur redoxin.

Class II require only an FAD/FMN-containing P450 reductase.

Class III are self sufficient and require no electron donor.

Class IV receive electrons from NADPH or NADH.

The active center for the reaction is the porphyrin ligated iron of the iron-protoporphyrin IX (heme). At the proximal face, the thiolate of a conserved cysteine coordinates the iron as the fifth ligand. At the distal face there is access for water, oxygen and a site for substrate binding and orientation relative to the iron. The electronic characteristics of the heme allow the iron to exist as either a low spin or high spin state depending on the ligand environment (Lewis and Pratt, 1998). In the heme system, 6-fold coordinated Fe is found to be low-spin and 5-fold coordinate Fe is found

to be in the high-spin state. In the resting state the iron is low spin (III) and a water molecule forms a sixth axial ligand on the distal face of the heme to stabilize the low-spin state (Segall et al., 1998).

The P450 operates by means of a catalytic cycle which involves seven steps:
1) The cycle begins in the resting state (Figure 1.6, 1) with a water bound to the ferric iron at the distal side. In this hexacoordinated Fe^{III} complex the d-orbitals of the iron contain five electrons in a low spin doublet configuration.

2) The entry of the substrate into the active site displaces this water molecule and leaves a five coordinated iron (Figure 1.6, 2). The iron moves from the plane of the heme to a position below the heme and becomes a sextet high-spin species.

3) The complex is now a better electron acceptor than the resting state and can take up an electron from the reductase leading to a high spin ferrous complex (Figure 1.6, 3).

4) Following reduction, oxygen is able to bind to the Fe^{II} high spin state to produce a ferrous dioxygen complex (Figure 1.6, 4).

5) This complex has a singlet spin state and is a good electron acceptor which triggers a second reduction to generate a ferric-peroxo anion species (Figure 1.6, 5).

6) As this is a base, it becomes protonated to a ferric-hydroperoxide species also called Cpd 0 (Figure 1.6, 6).

7) Cpd 0 is also a base and attracts a second proton to form Cpd I and water. This species transfers an oxygen atom to the substrate (Figure 1.6, 7). After the reaction, the product exits the active site and water enters to restore the resting state (Shaik et al., 2005).
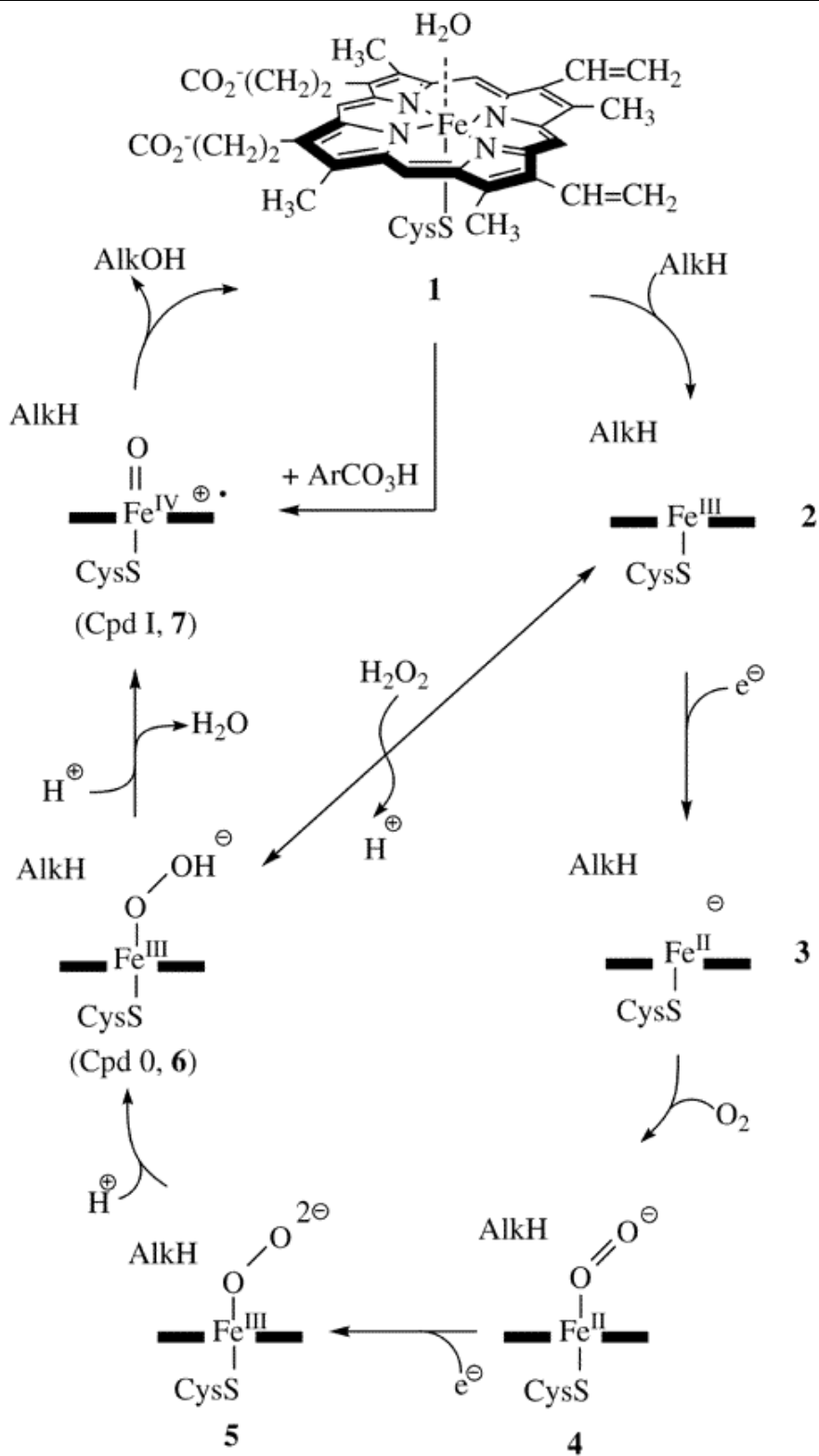
Figure 1.6 The P450 cycle taken from Shaik et al. (2005).

The redox partners for microsomal P450s are P450 reductase and cytochrome b5. P450

reductase accepts a pair of electrons from NADPH and transfers these to cytochromes

P450. Cytochrome b5 is known to augment P450 reactions by electron transfer. While P450 reductase can transfer the first and second electrons, b5 can only donate the second electron (Bridges et al., 1998).

1.7.6 P450 inhibitors

Ligands can act as inhibitors of P450s. Type I ligands displace the water ligand shifting the iron spin towards the high-spin form and accelerate P450 reduction. Ligands that contain nitrogen heterocycles and anilines can replace the water ligand to stabilize the low-spin form and are termed type II. These ligands act as inhibitors as they stabilize the low reduction potential P450 and prevent oxygen binding to the heme. The nature of the interaction can be determined by the visible spectrum of the ligand bound P450 measured with a spectrophotometer (Locuson et al., 2007). Type I is characterised by a trough at around 420mµ and peak at 380mµ, type II is characterised by a peak at 430mµ and trough at 394mµ (Gigon et al., 1968).

1.7.7 P450 regulation and induction

Each P450 has its own expression pattern which is developmentally regulated, often with expression limited to a particular developmental phase, and tissue specific. Insect xenobiotic detoxification occurs in the midgut, Malphigian tubules and fatbody, and transcripts of P450s involved in detoxification are enriched in these tissues. The expression and transcription rates of P450 genes are inducible by xenobiotics via a number of pathways. Aryl hydrocarbons such as dioxin act as inducers mediated by the Ah receptor, dioxin binds to the Ah receptor which dimerises with the protein Arnt, and the dimer induces a battery of genes including P450s (Giraudo et al.). Many insect P450s are induced by their substrates, CYP6B metabolises furanocoumarins, and its expression is 5-fold higher in individuals after dietary exposure. This insect gene contains response elements similar to those in mammalian genes that are targets of aryl hydrocarbon regulatory cascades (Brown et al., 2005). Insect P450s involved in insecticide resistance may be constitutively over-expressed, in the case of CYP6D1 the insert of a 15-bp fragment acts to disrupt repressor binding leading to high levels of expression in resistant strains (Scott et al., 1999).

1.8 Experimental determination of structure

The structure of a protein is important for its function and its structure can be experimentally determined using X-ray crystallography and NMR.

1.8.1 X-ray crystallography

For an object to be seen its size must be at least half the wavelength of the light being used to see it. To get a higher resolution, shorter wavelengths can be obtained from electrons and X-rays. X-ray scattering from a single molecule would be weak, a crystal arranges large numbers of molecules in the same orientation so that scattered waves in phase can amplify a signal. Crystallography relies on the scattering of X-rays by electrons in the sample to create a diffraction pattern. Crystals are used because they have a unit cell that is repeated throughout the crystal in a periodic way and this repetition amplifies the diffractions. These diffractions are governed by the geometry of the crystal unit cell and the wavelength of the X-ray, and the crystal structure is encoded in the diffractions. In a crystal, with the repetition of the unit cell, their diffraction patterns interfere with each other. The diffraction pattern is zero everywhere except for the spots where crystal diffraction is allowed, and at these spots the diffraction pattern equals that of the unit cell. In the Bragg model, the crystal contains families of equally spaced parallel planes if waves reflected from adjacent planes are in phase, if the path length distance for the two waves equals a whole number of wavelengths, constructive interference occurs and a strong diffraction is observed. This leads to Bragg's Law (Figure 1.7):

$$n\lambda = 2d \sin\theta$$
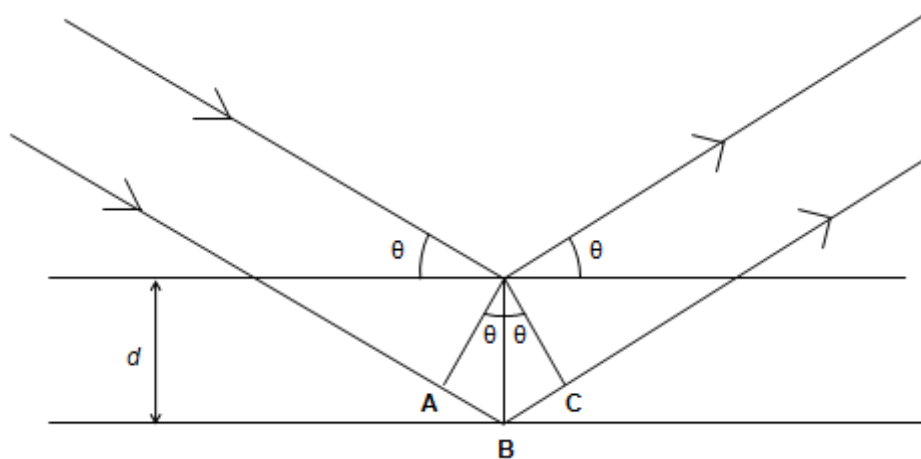
Figure 1.7 Bragg's Law (taken from Lattman, 2008). The two horizontal lines represent two planes belonging to a family of Bragg's planes with a spacing of d, incoming X-rays are reflected. The ray reflected by the lower plane travels a greater distance (ABC $= 2d \sin\theta$), when this distance equals a whole number of wavelengths λ, the reflected waves are in phase.

In most directions the scattered waves from different unit cells are out of phase and cancel to give zero. Non-zero diffraction is observed in those directions for which the diffracted waves from all the unit cells are in phase. The diffraction pattern consists of discrete spots or Bragg peaks where the diffracted waves from every unit cell add together.

To obtain the structure of the diffracting motif a Fourier transformation of the structure factors are calculated. These factors or F values represent the reflection amplitudes and phases. Each reflection is characterised by amplitude and phase but only amplitude can be measured and no direct information about phase is provided by the experiment, but the phases can be determined by a number of methods. This leads to an initial electron density distribution of the crystal which can be improved iteratively. The result of an X-ray diffraction experiment is a map of the electron density of the crystal and molecular modelling is used to interpret this as individual atoms and molecules. This model is refined to give the best match between the observed reflection amplitudes and those calculated from the model. This agreement between the observed and calculated amplitude is determined by the R-factor (Wlodawer et al., 2008).

1.8.2 NMR

Nuclear magnetic resonance (NMR) spectroscopy originated with the observation of nuclear magnetic resonance, and has been applied to determine protein structure. Many proteins do not crystallise, and for those that do some may diffract poorly. Unlike X-ray crystallography, NMR can operate in the solution state and protein dynamics and ligand binding can be measured. It is possible that X-ray structures may not be a true picture of the molecule as it exists in solution. NMR structure determination can be carried out under conditions that more closely match experimental conditions and can eliminate perturbations in geometry that may occur in the crystalline environment (MacArthur and Thornton, 1993). NMR spectroscopy is based on the magnetic properties of nuclei; nuclei that have an odd atomic mass or number behave as spinning electrical charges and have a magnetic moment. When placed in a magnetic field, its magnetic moment will align either parallel (low energy state) or antiparallel (high energy state) to the magnetic field. If energy is provided at the proper frequency, energy is absorbed and a transition from the lower energy to higher energy state can take place. As the nuclei return to the lower energy state, a signal with the same frequency as the one absorbed is emitted (Ingwall, 1982). Protein conformational data are upper distance limits derived

from nuclear Overhauser effects (NOEs). NOEs result from cross relaxation due to interactions between nearby pairs of nuclear spins in a molecule, and can give distance information (Herrmann et al., 2002).

NMR has an advantage over crystallography in that it is free of crystallisation artifacts, surface residues are often perturbed by intermolecular contacts in the crystal. This has a biological relevance as surfaces are often involved in interactions with other proteins and ligands. NMR also differs from crystallography in that each cross peak contains information on a single torsion angle or distance. In contrast each peak in a diffraction pattern contains information on the entire structure. NMR can also estimate the timescales of intra-molecular motions (Wagner et al., 1992). However, unlike crystallography, NMR is not useful for all proteins. It has a size limit of around 35kDa as large proteins have slower tumbling rates and shorter NMR signal relaxation times, the increase in size also introduces more complexity because there are more NMR-active nuclei (Yu, 1999).

1.9 Modelling

Proteins fold into specific compact structures and the information coded in the amino acid sequence is sufficient to determine the structure. The three dimensional structures of proteins can be predicted from the amino acid sequence using ab initio or comparative modelling. Ab initio modelling methods use physical principles alone to find the folded structure. It is assumed that a protein sequence folds to its native conformation that is at the global free-energy minimum. Although simulating protein folding can be carried out using molecular dynamics, this is computationally expensive. Ab initio modelling usually involves a representation of protein geometry, a force field and energy surface searching. A representation of the protein is needed but limits on computational resources mean an all atom model cannot be used. A simple geometry model is often used where a single particle represents a number of atoms of the protein. A potential function is needed to predict the native structure of the protein. Force fields such as CHARMM can be used. A conformational search technique is used to search the conformational space (Osguthorpe, 2000).

Comparative modelling uses experimentally determined structures to predict the conformation of a target protein with a similar sequence, and is possible because small changes in sequence may result in only small changes in structure. This modelling

involves finding an homologous 3D structural template that can be identified from the PDB, the sequences of the target and template are aligned and a structure is calculated. However, comparative modelling is restricted to sequences with closely related proteins with known structures (Sali and Blundell, 1993).

1.9.1 History of P450 structure prediction

Modelling and structure prediction of P450s was initially limited by the availability of crystal structures of only soluble microbial P450s. These structures were of limited use for modelling eukaryotic microsomal P450s that have adaptations for binding to the membrane and redox partners. With the availability of the first crystal structure of a eukaryotic P450 it was shown that although the elements that define the active site are the same as the microbial P450s, the arrangements of these elements differ. Of the substrate recognition sites, only SRS4 on the I helix is conserved in its location. The other SRS regions diverged significantly with RMDS of 3.3 Å for SRS5, 6.0 Å for SRS1, 6.4 Å for SRS2 and 3.9 Å for SRS3. Although different domains of CYP2C5 can be superimposed with those of P450BM3, the interdomain positioning differs and the fit is poor if both domains are used in the alignment (Williams et al., 2000b).

The differences between microbial and eukaryotic P450s affected the modeling, homology models of CYP2C5 based on CYPBM3 before its structure was determined had large errors. The positions of SRS1 differ giving 5 Å differences in position of residues, the F helix is moved significantly in CYP2C5 with a 12 Å displacement compared to the model (Williams et al., 2000b). The crysalisation of CYP2C5 was a landmark in modelling due to its similarity to other mammalian CYPs. After the CYP2C5 structure became available there was an improvement in the modelling of other mammalian P450s. The use of CYP2C5 to model CYP2D6 improved the models over those based on the microbial structures, principle component analysis showed improvements and the model successfully identified residues in the active site (Kirton et al., 2002) that were later confirmed in the crystal structure. A model of CYP2D6 based on CYP2C5 in another study (Marechal et al., 2008) also showed good agreement with the crystal structure. Residues within 5 Å of the heme overlaid well with the crystal and the Cα RMSD between the model and crystal was a reasonable 0.8 Å. The main region of error was the FG region which adopted a different conformation in the crystal. This region is known to vary in size and shape across CYPs.

1.10 Kinetics

P450 kinetics have a clinical relevance as the determination of P450 kinetics can be used to predict drug dosing and clearance in humans. There are a number of models that are used to determine P450 activity. Microsomes are vesicles of the endoplasmic reticulum and can be used as a source of P450s. The transmembrane part of the P450s insert into the microsomal membrane with the N terminus facing the luminal side and the body of the P450 located on the cytoplasmic side. The reductase and b5 are also located on the cytoplasmic side. Microsomes are commonly used as they can be easily isolated by centrifugation. However, there can be considerable differences in activity between microsomal preparations and as they contain a large number of P450s any activity may not easily be attributed to an individual isoform. To study the kinetics of an individual P450, baculovirus mediated expression can be used. A baculovirus can be used as a vector to carry a P450 gene into cultured insect cells, and is then expressed at a high level (Kramer and Tracy, 2008).

The kinetics of a P450 can be described by the Michaelis-Menten equation. K_m is defined as the substrate concentration at which half V_{max} is reached.

$$E + S \underset{k_{.1}}{\overset{k_1}{\rightleftarrows}} \quad E - S \xrightarrow{k_2} E + P$$
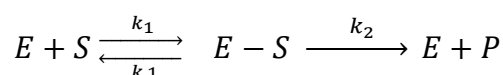
where E is the enzyme, S is the substrate and P the product.

$$v = \frac{V_{max} \cdot [S]}{K_m + [S]}$$

where V_{max} is the maximum velocity, K_m the Michaelis constant and $[S]$ the substrate concentration.

K_m is the combination of several rate constants and is equal to $k_{-1} + k_2/k_1$. K_m is the apparent dissociation constant often associated with the affinity, but the dissociation constant K_s is a more accurate description of affinity. The binding of an enzyme and its substrate can be represented as:
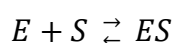
$$E + S \rightleftarrows ES$$

The equilibrium constant is the substrate dissociation constant K_s:

$$K_s = \frac{[E][S]}{[ES]}$$

The velocity of the enzyme reaction is the rate of product formation and expressed as the amount of product produced per time. V_0 is the initial velocity of product formation and proportional to the steady state concentration of ES. V_{max} is the maximal velocity of the reaction defined as the value of V_0 when the enzyme is completely saturated.

In the presence of an inhibitor there is a decrease in the rate of turnover. Competitive inhibitors bind to the same binding region as the substrate to prevent the substrate binding. This has no effect on V_{max} but increases K_m:

$$v = \frac{V_{max}\,[S]}{K_m\left(1 + \frac{[I]}{K_i}\right) + [S]}$$

where [I] is the inhibitor concentration, K_i is the inhibition constant $K_i = [E][I]/[EI]$.

Determining K_i usually involves multiple substrate and inhibitor concentrations and represents the dissociation constant of the enzyme inhibitor complex. IC50 is the concentration of a compound that results in 50% inhibition. it can also be used to describe the inhibitory potential of a compound but unlike K_i only one concentration of substrate is used in the presence of multiple inhibitor concentrations. Unlike K_i, the IC50 measurement is dependent on the concentration of both substrate and inhibitor and IC50s may not be comparable unless made at the same enzyme and substrate concentrations. Both K_i and IC50 can be used to describe the inhibitory potential of a compound, the lower the K_i and the lower the IC50, the higher the affinity. Protein-ligand binding interactions and binding energies can also be predicted using docking (Kramer and Tracy, 2008).

1.11 Structure based drug design

It is possible to rationally design drugs using the structural model of a protein. Drug design can be both computer aided or experimental. Computer aided design can involve virtual screening where databases of molecules are docked into the active site and scored to find the best binders. De novo drug design involves docking small fragments into the site and scoring them, novel drugs can then be created by linking the high scoring fragments. Visual inspection of molecules binding to the active site can lead to modifications to maximise the interactions. As proteins are flexible in solution they may occur as an ensemble of conformations. Dockings in a single rigid structure may not

reflect the results in solution. Protein flexibility can be accounted for by using NMR ensembles, or by using molecular dynamics or normal modes (Anderson, 2003).

1.12 Modelling and characterization of mosquito P450s

The next chapter, Chapter 2, covers details of the methods used in the studies in this thesis.

In Chapter 3 The modelling procedure for each P450 is discussed, each model is tested using data obtained from the collaborators or the literature to determine if the model can replicate known experimental results. This aims to demonstrate that the models can replicate the known activity of the enzymes and may serve to rationalize further experimental data and to priorities ligands for study.

The characteristics of pyrethroid metabolising and non-metabolising enzymes are discussed in Chapter 4. Insects express a large number of P450s and it is time consuming to characterize an individual enzyme to determine if it is capable of metabolism. An understanding of the requirements for binding pyrethroids can be used to select candidate P450s for further study. In this chapter, metabolisers are compared to the non metabolisers to determine these characteristics, and to suggest a 'fingerprint' to identify metabolisers by their active site characteristics.

The metabolism of other substrates is discussed in Chapter 5. This chapter discusses the activity of a number of P450s towards a range of fluorescent probes and acetylene inhibitors. As these probes are diverse in structure, the substrate preference gives an indication of the structure of the active sites of the enzymes. The residues forming the active site that produce the preference could be identified.

A number of factors external to the active site affecting metabolism are discussed in Chapter 6. As active site characteristics alone may not explain P450 substrate preference and activity, other factors were investigated. The role of b5 in enhancing metabolism and shifting product profiles, the effects of membrane binding, and the role of SNPs are discussed.

In Chapter 7, the results of the previous chapters are used to design novel pyrethroids and probes. Understanding how pyrethroids can bind to the active site for metabolism

can be used to place functional groups to mask metabolism, to produce toxic products upon metabolism or to produce poor contacts with the active sites to reduce metabolism. The design of pyrethroid-like acetylene probes can be used to label pyrethroid metabolising P450s for their identification.

Chapter 8 discusses proteins with a role in inflammation in the host response to malaria. The central role of Mal and its interactions with proteins in the inflammatory cascade are discussed.

2. Method

2.0 Preface

This chapter explains the theory behind the different methods used to carry out the work in this thesis. A range of tools are available that can be used for the characterisation of a protein. The overall strategy is illustrated in Figure 2.1.
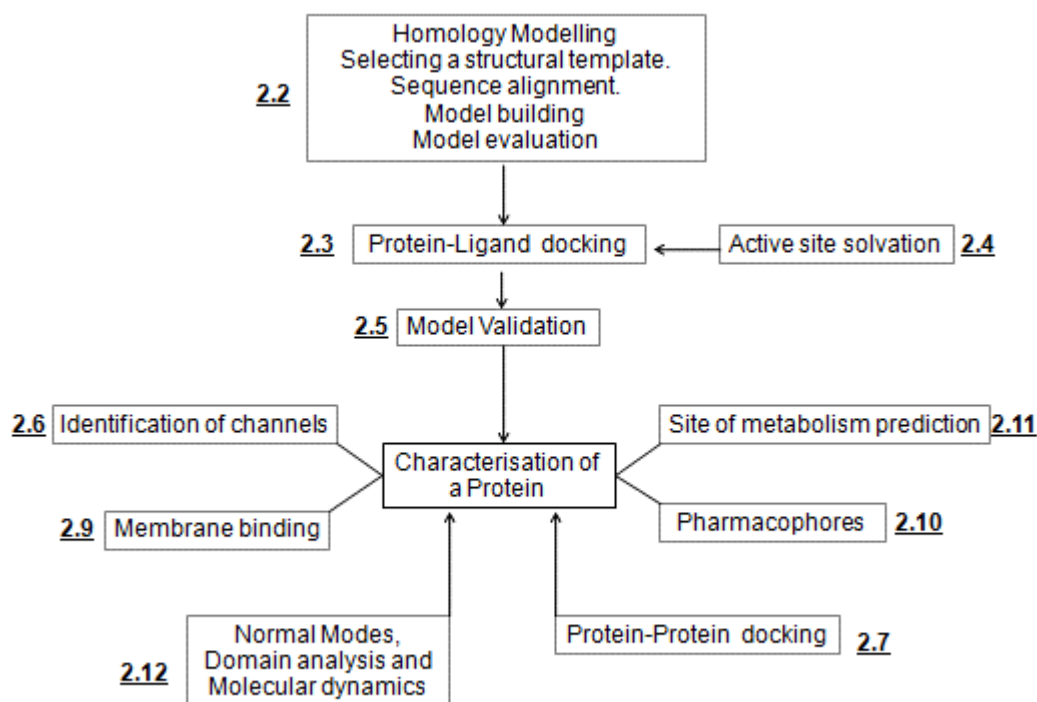
Figure 2.1 Overview of the strategy used in the characterisation of a protein. See the underlined sections for details of each step.

2.1 Introduction

With the development of large-scale genome sequencing, RNA expression and proteome scans, unknown proteins have been identified that are targets for a particular function. Although there are experimental structures for some proteins, there is a gap between the known sequences and known structures, and elucidating a protein's function relies on the analysis of structure/function relationships (Guerrucci and Belle, 1995). A number of tools are available that allow the elucidation of protein structure and function and an understanding of these tools is required.

2.2 Homology modelling

Homology modelling aims to produce a 3D model of an unknown protein based on similarity to a known structure. It is based on the observation that the structure of a

protein is determined by its amino acid sequence, and similar sequences produce similar structures. The structure of proteins from the same family are more conserved than the sequence as many combinations of side chains can produce structures that are similar (Lesk and Chothia, 1980). Comparative modelling consists of five steps: finding related protein structures, selecting one of these structures as a template, aligning the target sequence to the template, model building and evaluation. If the evaluation suggests that the model is incorrect, the process can be repeated (Sanchez and Sali, 2000).

2.2.1 Template selection

Searching the PDB can suggest a number of potential templates. To select one or more template a number of factors are taken into account (Kopp and Schwede, 2004):

Sequence similarity: The quality of the template is related to the overall sequence similarity and inversely related to the number and length of gaps.

Phylogeny: Families of proteins can be organised into subfamilies and phylogenetic trees can be used to identify the template from the closest subfamily.

Functional similarity: The similarity between the target and template can be considered. This can include similarities in ligands. A template that is bound to a similar ligand may be preferred for studying protein-ligand binding. However, in this study many of the targets were newly identified and had unknown ligand specificities.

Template quality: The quality of the crystal structure needs to be taken into account. The resolution and R factor can indicate the accuracy of the structure. In this study where appropriate, the best quality crystal structure was used.

In addition to this Baudry et al. (2006) suggested that to improve the predictive capabilities of P450 models there should be discrimination between P450 classes with structures subdivided into classes based on their known or presumed electron transfer partners. They found that using alignments that discriminate between the different classes could give models that are closer to crystal structures than when templates from different classes are combined. Therefore, templates from within the same class as the target were used for the modelling.

2.2.2 Sequence identity

The first step in homology modelling is to identify a structural template. Templates can be identified based on similarities of the amino acid sequence. While transferring structural information from a template to a target is straightforward when similarity is high, it can be difficult where similarity is low or restricted to a region. A threshold of similarity sufficient for homology modelling was proposed by Sander and Schneider (1991), where they quantified the relationship between sequence and structural similarity, and alignment length. They produced a database of homology derived secondary structure of proteins (HSSP) and aligned sequences and structures. From a scatter plot of sequence and structural similarity and length, they determined a threshold for each alignment length so that any alignment with a similarity over the threshold represents structural homology. This length dependant threshold for significant sequence identity (the HSSP-curve) is the inverse square root of the length of alignment, for alignments between 7-80 residues, and 25% identity for sequences over 80 residues (Sander and Schneider, 1991) (Figure 2.2).
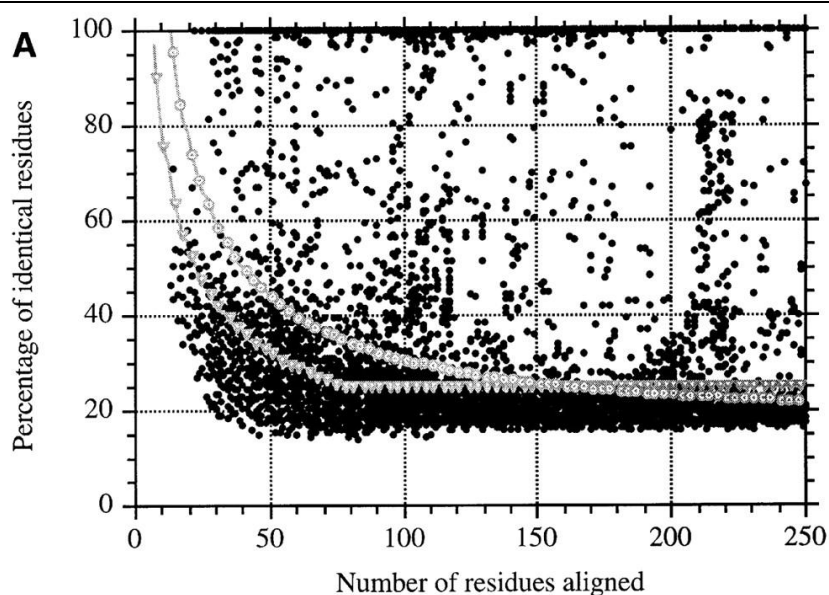
Figure 2.2 Taken from Rost (1999). The HSSP Curve (dotted circles) proposed by Sander and Schneider (1991) and the curve proposed by Rost (1999) (diamonds).

This original curve was derived for sequence identity and not for similarity and does not take into account that the physico-chemical nature of amino acids, such as hydrophobicity, is more conserved than residue type. Rost (1999) produced a threshold for sequence similarity. This curve is 100% for alignments of less than 12 residues and

10% for alignments above 500 residues. Sequence identity between proteins is used to infer evolutionary relationships and similarities in structure. As proteins differ in length, identity is corrected and expressed as percent identity but the denominator used can affect this value. The four denominators used are the length of the shorter sequence, arithmetic mean sequence length, number of aligned positions (including gaps) and number of aligned residue pairs (excludes gaps). May (2004) suggested that as the number of gaps has a large affect, the arithmetic mean should be used.

2.2.3 Sequence Alignment

Comparative modelling methods assume a structural equivalence between aligned target and template residues. A misalignment of one residue can result in a 4 Å error in the model (Fiser and Sali, 2003). The alignment is the most important step in homology modelling as if the alignment is incorrect, all other steps will not lead to reliable models. Sequence alignment methods can give an optimal alignment given a table of scores for matches and mismatches between residues (PAM and BLOSUM matrices) and penalties for insertions or deletions of different lengths (Thompson et al., 1994).

Sequence alignments between the template and target can be carried out by 3DCoffee (Poirot et al., 2004). This program uses three alignment methods: sequence-sequence, structure-structure and structure-sequence threading, and can produce alignments that are more accurate than alignments based on sequence information alone. The algorithms SAP was used for structure-structure alignment and Fugue was used for the structure-sequence alignment. Given a data set, the program identifies sequences associated with a structure and those that are not. It considers all possible pairs and applies the correct algorithm. If both sequences have structures, 3DCoffee makes a global pairwise alignment, a local pairwise alignment, and a structure based alignment with SAP. If only one has a structure Fugue is used instead of SAP (Poirot et al., 2004).

Sequence structure alignment methods such as Fugue use threading to fit a probe sequence onto the backbone of a known structure and evaluate its compatibility. This relies on structural similarity and can recognize sequence structure pairs with fold-level similarity. Residues in secondary structure tend to be more conserved than those in coil, and gaps are less likely to occur in secondary structure. Threading methods use a structure based gap penalty which calculates a penalty for each residue according to its position relative to secondary structure elements (Shi et al., 2001).

The resulting pairwise alignments are compiled into a list of pairs of residues and weighted according to the level of identity. When two or more alignments contribute the same pair their weights are added. This library of weighted residue pairs is used as a substitution matrix. During the multiple sequence alignment, the distance matrix of the sequences are used to estimate a guide tree. This tree controls the order sequences are included into the multiple sequence alignment (Poirot et al., 2004).

The alignments can be improved manually and also by using programs such as RASCAL (Thompson et al., 2003). RASCAL divides a given alignment into regions and realigns the badly aligned areas without affecting the whole. It can therefore be used to correct alignments.

2.2.4 Secondary structure prediction

Alignments can be improved by taking secondary structure into account. Gaps should be avoided in structural elements and the alignment needs to be inspected in the view of the template structure. Secondary structure prediction tools can be used to predict the occurrence of helices, sheets and coils and predicted structural similarities to a template can also be used to confirm template selection (Arnold et al., 2006). Secondary structure prediction can be obtained from a number of tools.

The GOR4 (Garnier et al., 1996) method predicts secondary structure by the use of an information function:

$$I(S; R) = \log[P(S|R)/P(S)]$$

where *S* is a conformation and *R* is an amino acid, *P*(*S*|*R*) is the probability of conformation *S* when residue *R* is present, and *P*(*S*) is the probability of *S*. An estimation of *I*(*S;R*) can be derived from a database of known sequences and corresponding secondary structure. The method relies on the frequencies of structural states observed for residues in a 17-residue sliding window. These frequencies are converted to structure propensity for the central position of the window (Heringa, 2000).

Nnpred (Kneller et al., 1990) uses neural networks to predict secondary structure. Neural networks can determine the mapping between a set of inputs and outputs such as that between sequence and secondary structure. The network acts as a pattern matcher

by learning the optimal mapping between input and output patterns, with the mapping specified by weights of the network variables. The network is composed of a set of computational units linked to other units with connections of modifiable weight trained on a structural database. Input units receive the input sequence. Three output units correspond to the secondary structure α-helix, β-strand and coil. A sliding window of 13 residues is inputted and the secondary structure of the central residue is assigned based on the output unit with the largest output.

Phyre (Bennett-Lovsey et al., 2008) uses an ensemble of fold recognition systems that detect the similarity of a protein sequence with a sequence of known 3D structure, which are then combined to give a secondary structure prediction for the sequence.

LOOPP (Meller and Elber, 2001) uses a threading approach which involves the matching of a sequence to a shape. The compatibility of a sequence to a structure is evaluated by using representations of protein structures and an energy score derived from the sum of interaction energies.

2.2.5 BLAST
Comparative modelling starts with searching known protein structures in the Protein Data Bank (PDB) (Berman et al., 2000) using the target sequence. Sequence similarity algorithms can be global or local. Global similarity methods optimise the overall alignment of two sequences while local methods align conserved subsequences. Local methods are preferred for database searching as the database may contain only partial sequences and as proteins may share regions of similarity. These methods assign scores to inserts and deletions to produce the least costly alignment to minimise the evolutionary distance and to maximise similarity. Although dynamic programming algorithms such as Needleman Wunsch can give an optimal alignment, they are impractical for searching large databases (Altschul et al., 1990).

Blast (Basic Local Alignment Search Tool) (Altschul et al., 1990) was developed for database searches due to its speed. It uses a matrix of similarity scores for all possible pairs of residues such as the BLOSUM matrix. Identities and conservative replacements are assigned positive scores while non conservative replacements are given negative scores. Blast searches for exact matches or maximal segment pairs (MSP) which are the

highest scoring pair of identical length segments, with the boundaries of the MSP chosen to maximise the score. The MSP is a measure of local similarity for any pair of sequences and the highest MSP score S can be estimated. Firstly a list of high scoring words are compiled. BLAST seeks segment pairs that contain word pairs (of fixed length w, usually 3 for proteins) with a score of at least threshold T. A database can be scanned to see if it contains a word of length w that can pair with the query, with a score greater than the threshold T. This alignment is then extended outwards. The algorithm returns the best scoring local alignment.

PSI-BLAST (Altschul et al., 1997) is preferred where there is a low sequence identity between targets and templates as this program is suited to finding distantly related members of a protein family, and could identify related templates with low sequence identities. PSI-BLAST is a position specific iteration of a BLAST search in which a profile or position specific scoring matrix is constructed automatically from a multiple alignment of an initial BLAST search. The profile is generated from the position scores with high scores for conserved regions and low scores for weakly conserved regions. This profile is used to perform a second BLAST search, and used to further refine the profile. PSI-BLAST searches therefore increase the sensitivity of a standard blast search.

Templates of known 3D structure are needed for modelling and can be identified from the Protein Data Bank (PDB) (Berman et al., 2000) database by BLASTP (Altschul et al., 1990) or PSI-BLAST (Altschul et al., 1997).

2.2.6 Phylogeny

Neighbour joining (Saitou and Nei, 1987) is widely used to construct phylogenetic trees. The algorithm identifies neighbours. Neighbours are operational taxonomic units (OTUs) that are connected through a single interior node. The algorithm starts with a starlike unrooted tree with the assumption that there is no clustering of OTUs. Some pairs of OTUs will be more closely related than other pairs, and pairs are chosen to give the smallest branch lengths or total number of substitutions. This pair of OTUs is then considered to be a single OTU and the next pair of OTUs are chosen. This process is iterated until all interior branches are found (Figure 2.3). As this method does not explore all possible tree topologies, this may not produce the minimum evolution tree but is efficient.

A   8   B
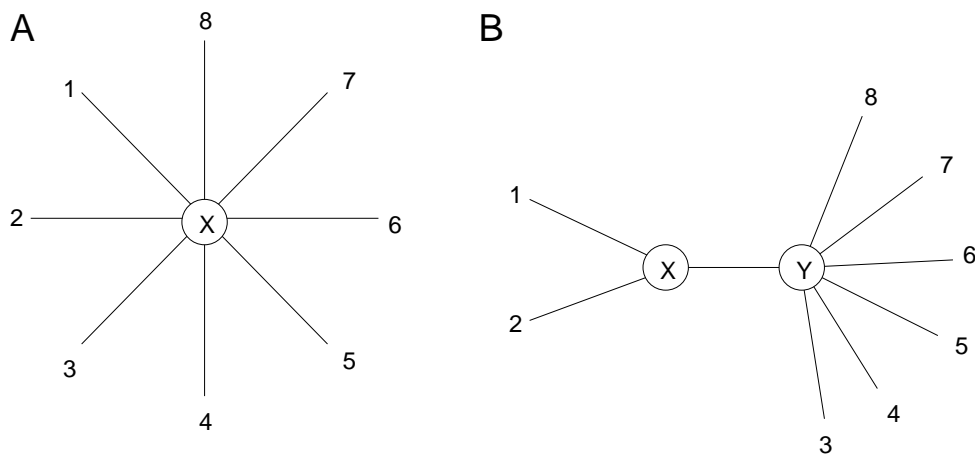
Figure 2.3 (A) A starlike tree with no hierarchical structure. (B) a tree in shich OTUs 1 and 2 are clustered into a single OTU. Taken from Saitou and Nei (1987).

2.2.7 Model Building

MODELLER (Sali and Blundell, 1993) attempts to find the most probable structure for a given alignment to a template. It involves a modelling method of three stages: alignment of the sequence to the template, extraction of spatial restraints on the sequence using the alignment, and satisfaction of the restraints to obtain a model. A restraint is defined as a probability density function p(x) for the feature x. A probability density function (pdf) is used as it gives more information than the mean of measurements and upper and lower bounds of a distance, and can be calculated using classical mechanics or empirically using a database. The spatial restraints on the target can be derived from homology, dihedrals and distances in the target are extracted from the alignment with the template. Spatial restraints on a sequence can be obtained from a statistical analysis of the relationships between features of protein structure. A database of 17 family alignments including 80 proteins was constructed to quantify relationships between Cα-Cα distances or between main-chain dihedral angles from related proteins. The values of main-chain dihedral angles are calculated from the type of residue, from main-chain conformations of an equivalent residue and from sequence similarity between the two proteins, and are described as pdfs. Stereochemical restraints such as bond length and angle are derived from the molecular mechanics force field CHARMM22, and statistical preferences for angles and distances from a representative set of proteins. Stereochemical restraints can be obtained from the amino acid sequence and include bond angles and distances, planarity of peptide groups and rings, chiralities,

60

and vdW contact distances.

Models are then produced by optimising the pdfs and limiting violations of spatial restraints. A molecular pdf is assembled from the collection of feature pdfs and this function is optimised. The optimum of the molecular pdf is found by successive optimisations of increasingly complex target functions, starting with local restraints and introducing more long range interactions, and finally the true molecular pdf incorporating all restraints. Loops are modeled using an optimisation approach that relies on conjugate gradients and molecular dynamics with simulated annealing.

The MODELLER program is able to generate high quality models automatically without user intervention. The alignment of the target and template tends to be the only input to the program and is therefore the determining factor influencing the quality of the model. MODELLER considers aligned residues to be structurally equivalent, and tries to derive a 3D model for the target that is as close to the template as possible. Inserted regions, that have no equivalent in the template, are modeled in the context of the whole molecule using their sequence alone (Sali and Blundell, 1993). ModLoop is a web server for the automated modelling of protein loops that uses the loop modelling routine of MODELLER that can be used to remodel loops.

2.2.8 Model evaluation

Homology modelling can provide good results when the sequence identity is high, but at low identities, errors can occur (Rodriguez et al., 1998). Models can be evaluated by checking the coordinates or stereochemistry of the model using programs such as PROCHECK (Laskowski et al., 1993). PROCHECK uses a Ramachandran plot to identify residues in the disallowed regions, and identifies bad contacts between non bonded atoms, and measures bond lengths and angles. This gives an assessment of the quality of the structure.

2.2.8.1 ERRAT

ERRAT (Colovos and Yeates, 1993) is a protein verification tool that identifies incorrectly folded regions of protein structures based on characteristic atomic interactions. ERRAT is useful as it gives the error score for each residue, identifying regions that need attention. ERRAT was developed to evaluate the correctness of protein structures. Carbon, nitrogen and oxygen atoms can form six distinct types of non-

covalently bonded interactions, CC, CN, CO, NN, NO and OO. Only non bonded interactions where the distance between the atoms is less than a pre set limit (3.5 Å) are considered, and for a protein of n length the fractions of interactions are calculated. For example the fraction of CC interactions represent the fraction of all pairwise interactions that are of the CC type. These were calculated from a database of reliable crystal structures.

These interactions occur with non random frequencies in proteins. CC, NO and NN interactions are more abundant than in a predicted random association, while CN and OO interactions are disfavoured and CO is random. The higher incidence of NO interactions may reflect H-bonding while the low incidence of OO interactions may be due to repulsions between charged oxygen atoms. Random distributions are expected to occur in incorrect regions and by distinguishing between correct and incorrect patterns of interactions, incorrectly folded regions can be identified (Colovos and Yeates, 1993).

2.3 Docking

Docking involves the prediction of the ligand orientation in the binding site, and can be used to predict the residues that could be involved in substrate binding. Docking is usually multistage, with docking algorithms that pose the ligand into the site, and scoring functions that evaluate the interaction, these are reviewed by Kitchen et al. (2004).

There are three types of docking method: random search method where random changes are made to the ligand position, such as that used by GOLD (Jones et al., 1997) and AutoDock (Goodsell et al., 1993); a systematic search where all degrees of freedom of the ligand are explored, such as DOCK (Ewing et al., 2001), FlexX (Rarey et al., 1996) and Glide (Friesner et al., 2004); and simulation methods where molecular dynamics is used to simulate ligand binding such as the minimisation stages of DOCK.

There are also three classes of scoring function: force-field based where molecular mechanics force-fields quantify the energies of the receptor-ligand interaction, such as DOCK and AutoDock; empirical functions that are fitted to reproduce experimental data, such as ChemScore (Baxter et al., 1998; Eldridge et al., 1997) and LUDI (Bohm, 1994); and knowledge based that are designed to reproduce experimental structures rather than binding energies such as DrugScore (Gohlke et al., 2000).

In preparation for docking, a number of modifications need to be made to the model. MODELLERs output is a 3D model for the target sequence containing the mainchain and sidechain non-hydrogen atoms, therefore both the hydrogens and the heme need to be added to the model. MOE (Molecular Operating Environment, Chemical Computing Group, Inc., Montreal, Quebec) can be used to prepare the heme and to add hydrogens to standard positions on the protein.

2.3.1 GOLD

GOLD (Genetic Optimisation for Ligand Docking) (Jones et al., 1997) uses an evolutionary strategy to perform docking. This approach can rapidly identify good but not necessarily optimal solutions. It uses a genetic algorithm that mimics evolution by the manipulation of data structures called 'chromosomes'. Each chromosome is a possible ligand orientation within the binding site and is assigned a score. Instead of one large population of chromosomes, they are distributed between subpopulations, or islands, with migration between them. Migration involves chromosomes being copied from one island to the next with migration occurring 5% of the time. Parent chromosomes are chosen and genetic operators, crossover and mutation are applied. Crossover combines chromosomes while mutations introduce random permutations. Parents are chosen randomly with a bias towards the fittest to introduce an evolutionary pressure, the selection method used is termed 'roulette-wheel-selection' as it is analogous to a roulette wheel with each member of the population occupying a slice of the wheel that is proportional to its fitness.

To prevent the GA converging to a sub-optimal solution and to maintain population diversity a low selection pressure of 1.1 is used. This is the relative probability of the best individual being selected compared with the average. To further increase diversity, when an individual is added to a population, niching is used to compare that individual to every member of the population to see if they inhabit the same niche. Chromosomes occupying the same niche have RMSDs of less than 1 Å. If the new chromosome is in the same niche as an individual in the population, then it replaces the least fit chromosome in that niche rather than the least fit in the population. This algorithm is stochastic in nature, starting from a randomly generated population, with mutation introducing random changes and parent chromosomes are selected at random with a bias towards the fittest (Jones et al., 1997). One of the problems of this algorithm is

reproducibility and many runs may be needed to elucidate the correct binding mode. In this study at least 50 runs were used for each ligand.

The chromosome is represented as four strings. Conformational information such as the angle of rotation around a bond are represented by strings for both the protein and ligand in two binary strings. Two other strings suggest mappings between the ligand and active site. One of these is a mapping between lone pairs in the ligand to hydrogen atoms in the protein, the other a mapping of hydrogen atoms in the ligand to lone pairs in the protein. By mapping lone pairs with hydrogen atoms, H-bonds are suggested and when the chromosome is decoded, a least squares method is used to form as many H-bonds as possible. However, for ligands with fewer than three hydrogen bond donors or acceptors, fitting is not possible. To account for this, for ligands with fewer than five donors or acceptors, a mapping is chosen and decoded by placing the acceptor on the donor hydrogen fitting point. Although this produces a result, it is less effective than the least squares fitting process (Jones et al., 1997).

GOLD requires the user to define the size and location of the active site by specifying an origin and a radius. GOLD uses a flood-fill algorithm to locate the solvent accessible surface within the radius of the origin, and then identifies concave surfaces to which the ligand can bind. In this study the origin was specified as the heme Fe, and the radius as 20 Å.

2.3.2 ChemScore

Ligand binding affinity represents the difference in free energy between the protein plus unbound ligand in solution, and their complex (Weber et al., 1992). ChemScore (Baxter et al., 1998; Eldridge et al., 1997) is an empirical scoring function that calculates an estimate of the free energy of binding for a given ligand receptor complex. The ChemScore function assigns atom types to all ligand atoms and to all receptor atoms in contact with the ligand:

lipophilic: Cl, Br, and I which are not ions; sulphurs which are not polar or acceptor; carbons which are not polar.

H bond donor: Nitrogens with hydrogens attached; hydrogens attached to N or O.

H-bond donor/acceptor: Oxygens attached to hydrogens; imine nitrogen.

H-bond acceptor: Oxygens not attached to hydrogens; N with no hydrogens and one or two connections; halogens that are ions; sulphurs with one connection.

Polar: N with no hydrogens and more than two connections; phosphorous; carbons and sulphurs attached to two or more polar atoms; fluorine; carbons attached to two or more polar atoms; carbons in nitriles or carbonyls; N with no hydrogens and four connections.

Metal: metal atoms.

$$\Delta G_{binding} = \Delta G_0 + \Delta G_{hbond} \sum_{il} g_1(\Delta r) g_2(\Delta \alpha)$$

$$+ \Delta G_{metal} \sum_{aM} f(r_{aM}) + \Delta G_{lipo} \sum_{IL} f(r_{IL})$$

$$+ \Delta G_{rot} H_{rot}$$

Equation 2.1 The ChemScore empirical scoring function. ΔG terms are coefficients derived from multiple linear regression analysis on a training set.

In the ChemScore equation (Equation 2.1) the H-bond term $\Sigma_{il}g_1g_2$ is calculated for all possibilities of hydrogen bonds between the receptor (I) and ligand (i) atoms with a penalty score for deviation from ideal angles and distances (Equation 2.2). The metal term is a simple contact term between acceptor/donor atoms in the receptor and metal atoms in the ligand (Equation 2.3). The lipophilic term is calculated for all lipophilic atoms in the receptor (L) and ligand (l) (Equation 2.4) (Eldridge et al., 1997).

$$\sum_{il} g_1 \, g_2$$

$$g_1(\Delta r) = \begin{cases} 1 & \text{if} & \Delta r \leq 0.25 \text{ Å} \\ 1 - (\Delta r - 0.25)/0.4 & \text{if} & 0.25\text{Å} < \Delta r \leq 0.65 \text{ Å} \\ 0 & \text{if} & \Delta r > 0.65 \text{ Å} \end{cases}$$

$$g_2(\Delta \alpha) = \begin{cases} 1 & \text{if} & \Delta \alpha \leq 30° \\ 1 - (\Delta \alpha - 30)/50 & \text{if} & 30° < \Delta \alpha \leq 80° \\ 0 & \text{if} & \Delta \alpha > 80° \end{cases}$$

Equation 2.2 The ChemScore H-bond term. Δr is the deviation of the H-bond length from 1.85 Å, and $\Delta \alpha$ the deviation from the ideal angle 180°.

$$\sum {}_{aM} f(r_{aM})$$

Equation 2.3 ChemScore metal term.

where:
a= ligand acceptor/donor atoms
M= receptor metal atoms
$f(r)$ = a contact term (Figure 2.4) where R1 is 2.2 Å and R2 is 2.6 Å
r_{aM} = interatomic distance between ligand and receptor atoms.

$$\sum {}_{IL} f(r_{IL})$$

Equation 2.4 The ChemScore lipophilic term.
where:
l= lipophilic atoms in the ligand.
L= lipophilic atoms in the receptor.
$f(r)$ = a contact term (Figure 2.4) where R1 is the vdW radius of l and L +0.5 Å, and R2 is R1 + 3.0 Å.
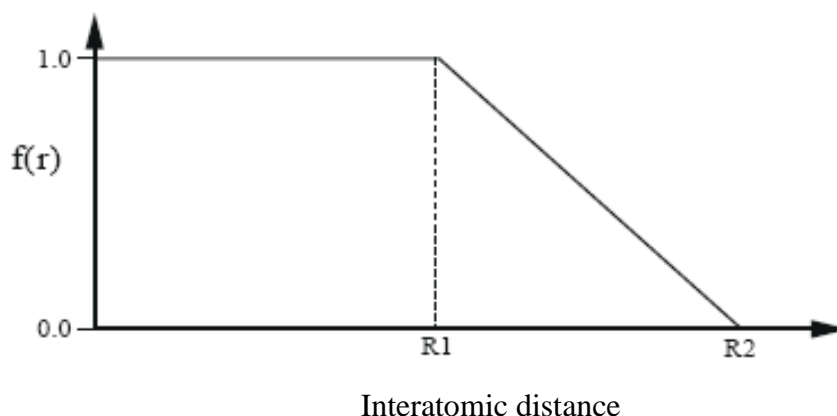
Figure 2.4 The form of the function f(r).

The torsional entropy term identifies frozen rotatable bonds. These bonds are considered frozen if both sides of the bond are in contact with the receptor, if the distance is less than the sum of the vdW radii plus 0.5 Å. If only one side of the bond is in contact, then rotation is only partly impaired and the penalty is not large. Large peptide ligands can have many frozen rotatable bonds and have an unrealistically high rotatable bond penalty. The contribution from rotatable bonds can be reduced as lipophilic peptides can fold into a low energy conformation in solution and the number of freely rotatable bonds is reduced. Lipophilic sidechains in contact with the receptor can retain entropy associated with the non-local nature of lipophilic interactions (Equation 2.5) (Eldridge et al., 1997).

$$\text{Hrot} = 1 + (1 - 1/N_{rot}) \sum r(P_{nl}(r) + P'_{nl}(r)) / 2$$

Equation 2.5 The ChemScore torsional entropy term.
Where:
Nrot = number of rotatable bonds.
Pnl(r) + P'nl(r) = percentage of non-lipophilic atoms either side of the bond.

A clash term is used to penalise close contacts between ligand and receptor heavy atoms and depends on the atom types in the clash.

For two atoms that could be involved in hydrogen bond contact the term is zero when $r > 1.60$ Å, and for $r \leq 1.60$ Å is:

$$E = \frac{20.0}{\Delta G_{hbond}} \left(\frac{1.60 - r}{1.60} \right)$$

For a metal contact the term is zero when $r > 1.38$ Å and for $r \leq 1.38$ Å the term is:

$$E = \frac{20.0}{\Delta G_{metal}} \left(\frac{1.38 - r}{1.38} \right)$$

For all other heavy atoms the term is zero when $r > r_{clash}$ and for $r \leq r_{clash}$ it takes the form:

$$E = 1.0 + 4.0 \left(\frac{r - r_{clash}}{-r_{clash}} \right)$$

r is the distance between the ligand and receptor atom. $r_{clash} = 3.35$ Å if the receptor atom is a sulphur, otherwise $r_{clash} = 3.10$ Å.

The internal energy of the ligand is a sum of torsion and internal clash terms. The internal clash takes the same forms as the ligand receptor clash terms but is only evaluated for ligand atoms. The torsion term is:

$$E = A[1 - \cos(n\Phi - \Phi_0)]$$

Where: Φ is the torsion angle.

For sp^3-sp^3 bonds $A = 0.1875$, $n = 3.0$, $\Phi_0 = \pi$. For sp^3-sp^2 bonds $A = 0.09375$, $n = 6.0$, $\Phi_0 = 0.0$. For sp^2-sp^2 bonds $A = 0.1875$, $n = 2.0$, $\Phi_0 = 0$ (Baxter et al., 1998).

The final ChemScore value is obtained by adding the clash penalty and internal torsion terms.

Constraints can be applied to the dockings such as a distance constraint. The distance between a ligand and protein atom can be constrained to within a maximum and minimum distance bounds. If a distance lies outside of these bounds, a spring energy term is used to reduce its fitness score (Equation 2.6) (Eldridge et al., 1997).

$$E = kx^2$$

Equation 2.6 ChemScore constraint.
where: x is the difference between the distance and constraint bound.
k is the user defined spring constant.

The ChemScore function was optimised against a training set of 82 protein ligand complexes taken from the PDB for which binding affinities (kJ/mol) are known. The function was able to reproduce the binding affinities with an error of 8.86 kJ/mol (Eldridge et al., 1997).

2.3.3 DOCK

The DOCK method (Ewing et al., 2001) is a rapid docking algorithm suited for docking databases. It is divided into three parts:

1) Representation of the receptor and ligand and identifying the binding site. To identify the active site a program generates a set of spheres to fill all of the grooves on the protein surface, these spheres are then collected into binding sites. The sphere centres are the putative ligand atom positions. To orient the ligand into the active site, active site spheres are paired with ligand atoms. DOCK tries to produce a representation of the key in a 'lock and key' representation of ligand binding, If a ligand provides a good match then it should fit within the receptor spheres.

2) Matching the receptor and ligand. Ligand and receptor spheres are matched by comparing the internal distances ligand-centre/ ligand-centre and receptor-centre/receptor-centre distances. Sets of ligand centres match sets of receptor centres if the internal distances match within a distance tolerance. This allows the identification of geometrically similar clusters of spheres in the receptor site.

3) Optimisation of the ligand position. Using a least squares algorithm, the ligand is rotated into the receptor spheres to reduce atom overlaps. Finally after the initial

orientation, the ligand may be minimised using a rigid body simplex to find a local minimum and minimise the energy score.

2.3.4 DOCK anchor and grow

For flexible docking, DOCK uses an anchor-and-grow algorithm. During flexible docking the ligand is represented by rigid segments separated by rotatable bonds. Using the anchor first search, a conformation is constructed and minimised one segment at a time starting from an anchor segment (Figure 2.5).

Step 1) The largest rigid substructure is identified and rigidly oriented into the active site by matching ligand atoms to the receptor spheres of the active site.

Step 2) The anchor orientations are evaluated and optimised using a scoring function and an energy minimiser. Orientations are ranked, clustered by RMSD. and pruned.

Step 3) The flexible parts of the ligand are then built or grown into the receptor. During the growth phase an internal energy score computes the Lennard-Jones and Coulombic energy between all ligand atoms to minimise internal clashes.

Step 4) When complete each conformation is locally optimised.

Step 5) If any additional portions of the ligand can act as anchors the process is repeated.

It is believed that the shape of the active site will restrict the sampling of the ligand to those that are most relevant (Ewing et al., 2001).

2.3.5 DOCK Scoring

A grid based score can be generated by DOCK. A simple energy score is based on an implementation of force field scoring. The force field scores approximate molecular mechanics interaction energies. These consist of van der Waals and Coulombic electrostatics components (Equation 2.7) (Meng et al., 1992).
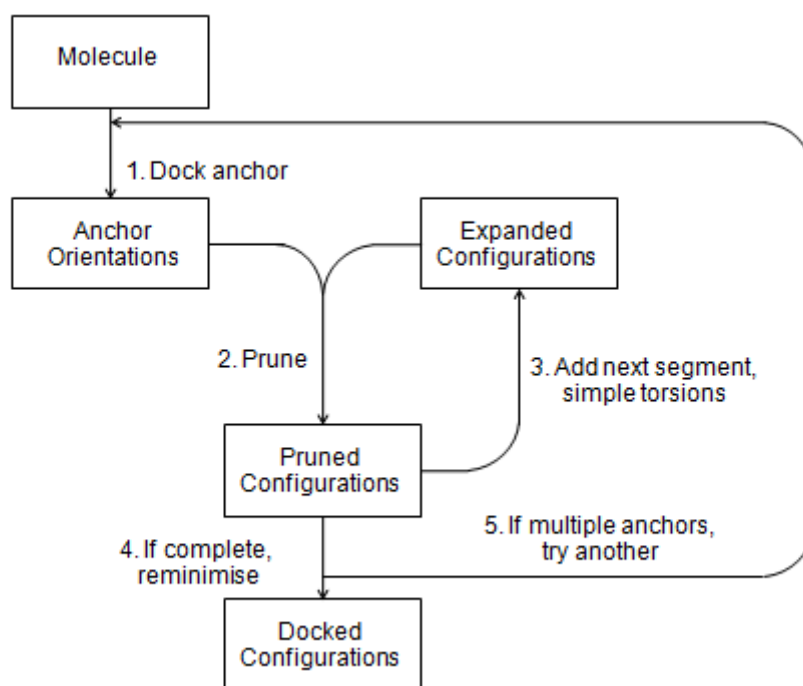
Figure 2.5 The DOCK method taken from Ewing et al. (2001).

$$E = \sum_{i=1}^{lig} \sum_{j=1}^{rec} \left(\frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^{6}} + 332 \frac{q_i q_j}{D r_{ij}} \right)$$

Equation 2.7 The DOCK score.

Where:

E	Intermolecular interaction energy.
r_{ij}	Distance between atoms i and j.
A_{ij} and B_{ij}	van der Waals repulsion and attraction parameters.
q_i and q_j	Point charges on atoms i and j.
D	Dielectric function.
332	Factor to convert electrostatic energy to kcal/mol.

DOCK uses the program DISTMAP that produces a grid for a vdW contact score, for every receptor atom within contact range, the sum at a grid point is incremented by one unless a close contact limit is violated in which case a negative number is added. The electrostatic score is based on potentials calculated by DelPhi (Meng et al., 1992). The energy function lacks an explicit H-bonding, desolvation and hydrophobicity terms and it is known that the limitations of the scoring function affect the accuracy of the binding orientations (Kitchen et al., 2004).

2.4 Active site solvation

The binding of a ligand to a protein takes place in aqueous solution and water can influence this process by mediating H-bonds, and the presence of a water in the binding interface has been shown to increase the range of ligands that can be bound by a protein (Ladbury, 1996).

2.4.1 Solvation with DOWSER

As side chain packing in proteins is not perfect, cavities exist that can accommodate waters. To stabilise a water, the cavity has to provide an energetically favourable environment comparable to that of liquid water so that the energy transfer of a molecule from liquid water to the cavity is negative. DOWSER (Zhang and Hermans, 1996) gives a quantitative measure of the hydrophilicity of cavities in proteins by calculating the energies of introducing a water molecule into the cavities. DOWSER identifies cavities on the surface of the protein, it then searches for the minimum energy positions of a water within a cavity. Polar cavities with interaction energies of below -12 kcal/mol tend to be filled with water and non-polar cavities with energies greater than this tend to be empty. This threshold energy is below the energy of liquid water (-10 kcal/mol) and near that of ice (-11.5 kcal/mol).

2.4.2 MOE water soak

The active site was also solvated using the 'water soak' option of the MOE program (Molecular Operating Environment). This solvates the selected parts with a sphere or cube with the size determined by the user. The added waters are extracted from a predefined box containing the water molecules. While the H-bond networks between the water molecules are optimised, H-bond networks between the waters and the protein may not be.

2.4.3 GOLD Ligand Docking with water

GOLD allows waters to be switched on or off and to rotate around three axes to optimise H-bonding. To predict if a specific water molecule should be bound or displaced, GOLD calculates the free-energy change (ΔG_b) associated with transferring a water from the bulk solvent to its binding site in the protein. The free energy change for a given water molecule is:

$$\Delta G_b(w) = \Delta G_p(w) + \Delta G_i(w)$$

where $\Delta G_p(w)$ is a constant penalty for each water that is switched on and represents the loss of rigid-body translational and rotational entropy upon binding. $\Delta G_i(w)$ is the binding affinity containing contributions from interactions that the water forms with the protein and ligand. For a water to be bound in the protein-ligand complex, its binding affinity must outweigh the loss of rigid-body entropy.

This method was found to correctly predict whether a water is present or displaced in 92% of the cases. The quality of the binding modes has been assessed for three categories: primary mediated complexes where a H-bond donor or acceptor H-bonds with a water but not with the protein; secondary mediated complexes where ligand donors and acceptors can H-bond with water and the protein; non-mediated complexes where the ligand displaces all waters in the active site. There was an improvement in the binding modes for primary mediated complexes, but there was no affect on the quality of the predicted binding modes for secondary and non-mediated complexes. This lack of improvement could occur because the interaction between a ligand and water represents only a small fraction of interactions a ligand forms and a reasonable binding mode can be predicted in the absence of important structural waters (Verdonk et al., 2005).

In test sets where the water should be displaced or where decoy waters were present, Verdonk et al. (2005) found a decrease in success rate. They suggested that as these waters are in good positions to interact with the protein, they are not easily displaced, and including them prevents the algorithm from finding the correct mode. Each water included increases the search space, increasing the likelihood of a false positive, and it was advised to limit the number of waters in the active site to only include those that are known to be involved in ligand binding.

2.5 Model validation

The development of a computer based representation of the active site of P450s is an iterative process that relies on substrate binding data and mutagenesis to test the hypotheses driven by the initial models. The model produced should be consistent with the known experimental observations such as site directed mutagenesis and ligand binding, and the results of the dockings can be used to validate an homology model. Where substrates are able to dock with positions of known sites of metabolism above the heme, the models can be shown to produce results that are consistent with experimental results. A comparison between the ChemScores of docked ligands and the

experimentally derived IC50 values can be used as an assessment of model quality. The experimental log IC50 scores can be plotted against the negative ChemScores of the best ranked solution for each ligand, and the correlation can be used as an assessment of quality (Kemp et al., 2004).

2.6 Identification of channels

Molecular channels are a feature of many biomolecules. In P450s, the active site is buried inside the protein and routes exist that allow ligands to enter and leave the site. Identifying these channels is important as substrate specificity may not only be determined by the active site, but also by the selectivity of access routes (Wade et al., 2004).
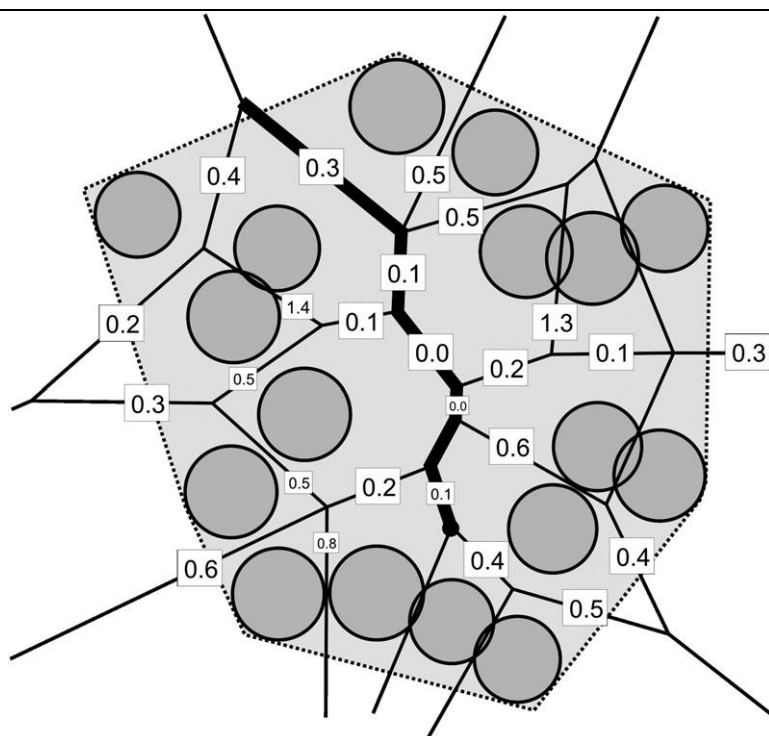
Figure 2.6 Voronoi diagram (thin lines) and outer boundary (dotted line) of a molecule represented as atoms (circles).The optimal path (thick line) from a given starting point (black circle) represents a found tunnel (taken from Petrek et al, 2007).

2.6.1 MOLE

MOLE (Petrek et al., 2007) was developed for the rapid location of channels, tunnels and pores in proteins and is able to give smooth channel profiles and identify bottlenecks. It finds channels by seeking an exit pathway from a point inside the protein cavity. With both the cavity and exterior defined, it finds the shortest, widest pathway between them. MOLE is based on Voronoi diagrams, it represents the protein by atom centred van der Waals spheres and constructs a Voronoi diagram. This diagram divides a

space according to the distances between the protein atoms and consists of cells representing the points closest to the atom in the centre of each cell, the boundary of a cell is comprised of facets that form the interface between neighbouring cells (Figure 2.6). This mesh can be used to search for molecular channels leading from a given point to the outside. The boundary between the protein and environment is approximated and the cheapest pathway from the starting point and the exterior is identified. MOLE has been shown to be capable of identifying access or egress channels and channel networks.

2.7 Protein-protein docking

Protein-protein interactions play an important role in many biochemical processes. The goal of protein-protein interaction predictors is to take two proteins and derive a model of the bound structure. Protein docking is carried out in two stages, the initial stage and the refinement stage. In the initial stage the proteins are treated as rigid and all docking sites are explored to produce a number of complexes. In the refinement stage a small number of the complexes produced are refined with conformational searches using side and main chain rotamers and scored using more detailed energy functions (Chen et al., 2003).

2.7.1 GRAMM

GRAMM (Vakser, 1995) contains a geometric algorithm. The coordinates of each protein are projected onto a three dimensional grid. Grid points are considered inside the molecule if they are within the van der Waals radius of an atom. Grid points at the surface of the protein are distinguished from the interior as a layer of finite width between the outside and inside. The surfaces of the two proteins are then matched. One molecule is moved relative to the other in each dimension and a correlation between the molecules is determined. If there is no contact between the molecules the correlation is 0; if there is contact the correlation is positive (Equation 2.8).

To prevent the penetration of one molecule into the other, ρ is given a large negative score while δ is given a small positive score so that when penetration occurs the multiplication of negative numbers by positive numbers results in a negative score. A good geometric match occurs where there are coordinates with a high positive correlation peak and the width of the peak gives a measure of displacement of the molecule before the match is lost. All possible orientations of the ligand molecule are

sampled with the positions of the ligand with the highest score retrieved (Katchalski-Katzir et al., 1992).

$$a_{l,m,n} = \begin{cases} 1 & \text{the surface of the molecule} \\ \rho & \text{inside the molecule } (-15) \\ 0 & \text{outside the molecule} \end{cases}$$

$$b_{l,m,n} = \begin{cases} 1 & \text{the surface of the molecule} \\ \delta & \text{inside the molecule } (1) \\ 0 & \text{outside the molecule} \end{cases}$$

Equation 2.8 The GRAMM geometric recognition algorithm.

A number of factors are known to affect this algorithm. As the atomic coordinates of both molecules are used excluding the hydrogens, the quality of the match is affected. Although conformational change is likely to occur, both molecules are assumed to be rigid leading to either gaps or overlaps in the contact area. While this can be tolerated by increasing the layer of grid points that form the surface, a thicker layer can lead to false matches by increasing the tolerance to the deviation away from the correct match. The grid size is an important parameter. A fine grid uses excessive computational time while with a large grid spurious correlations can occur. As no feature smaller than the grid step is considered it is the threshold for molecular data, and a large grid step can be used for molecules with no local details such as low resolution structures or models where there may be inaccuracies in the surface details. As geometric surface complementarity methods only maximises surface contact between molecules, they are known to have various problems. There may be a number of solutions for geometric fit so that the correct fit may occur along with many false solutions. As the algorithm assumes rigid bodies, it does not take into account the conformational changes that may occur and may result in failure when using the structures of the free compounds (unbound docking) (Katchalski-Katzir et al., 1992).

2.7.2 PatchDock

PatchDock (Duhovny et al., 2002) also determines surface complementarity but using a different strategy. Firstly, the surface of the protein is divided into patches based on geometric shape (concave, convex, flat). These are then filtered with hot-spots retained.

Protein-protein interaction faces tend to have conserved polar and aromatic hot-spot residues. Secondly, the patches are matched between the proteins, concave with convex and flat with any type. Thirdly, the complexes are scored with penalties for overlaps and scores for complementarity.

2.7.3 Zdock

Zdock (Chen et al., 2003; Chen and Weng, 2003) uses pairwise shape complementarity (PSC), desolvation (DE) and electrostatics (ELEC) to create a scoring function PSC+DE+ELEC.

The favourable component of PSC is defined as the number of receptor ligand atom pairs within a distance cut off. This cut off is defined as a parameter D plus the receptor atom radius to account for different atomic radii. The unfavourable component is proportional to the number of overlapping grid points between the receptor and ligand. Rpsc and Lpsc describe the geometric characteristics for the receptor and ligand in an NxNxN grid. A solvent accessible surface layer is not defined. Instead for grid points in the open space, the number of atoms within a distance cut off are recorded.

$$\text{Re}\big[R_{psc}(l,m,n)\big] = \begin{cases} \text{number of receptor atoms within} & \text{open space} \\ \quad (D + \text{receptor atom radius}) \\ \qquad\qquad 0 & \text{otherwise} \end{cases}$$

$$\text{Re}[L_{PSC}(l,m,n)] = \begin{cases} 1 & \text{if this grid is nearest grid of a ligand atom} \\ 0 & \text{otherwise} \end{cases}$$

$$\text{Im}[L_{PSC}(l,m,n) = \begin{cases} 3 & \text{solvent excluding surface of protein} \\ 9 & \text{protein core} \\ 0 & \text{open space} \end{cases}$$

Equation 2.9 The Zdock PSC term.

The penalties for different grid overlaps:

core-core   -9*9 = -81

surface-core   -3*9 = -27

surface-surface -3*3 =-9

The small penalty for surface overlaps allows a toleration of structural flexibility. The Re[] terms compute the favourable component of PSC, for each grid point in open space, Re[Rpsc] denotes the number of receptor atoms within the distance cut off. Re[Lpsc] records the nearest grid point for each ligand atom. Multiplying these two terms gives the total number of receptor-ligand atom pairs within the distance cut off.

The favourable and unfavourable components can be summed to give a single score with high scores indicating good shape complementarity.

To estimate desolvation, the atomic contact energy (ACE) is used. ACE is defined as the change in free energy when breaking two protein-water contacts and forming protein-protein and water-water contacts. The total desolvation score is the sum of the ACE scores of all receptor and ligand atoms within a distance cut off of 6 Å. ACE scores range from 1.334 (unfavourable) to -1.827 (favourable) while the PSC scores give a score of 1 for atom pairs. To make the PSC and ACE scores compatible, the signs of the PSC are flipped and the penalty score reduced to -1.334. When PSC and ACE are summed, the more negative the score, the more favourable the interaction energy. R_{de} and L_{de} are used to describe the desolvation properties and R_{psc} and L_{psc} describe the shape.

$$R_{PSC} = L_{PSC} = \begin{cases} 3 & \text{solvent excluding surface layer} \\ 3^2 & \text{protein core} \\ 0 & \text{open space} \end{cases}$$

$$\text{Re}[R_{DE}] = \text{Re}[L_{DE}] = \begin{cases} \text{sum PSC } + \text{ACE scores of all nearby atoms} & \text{open space} \\ 0 & \text{otherwise} \end{cases}$$

$$\text{Im}[R_{DE}] = \text{Im}[L_{DE}] = \begin{cases} 1 & \text{if grid point is nearest grid point of an atom} \\ 0 & \text{otherwise} \end{cases}$$

Equation 2.10 The Zdock desolvation term.

Electrostatics are computed using the Coulombic formula. This is expressed as a function of the electrical potential generated by the receptor and partial charges of ligand atoms. This is multiplied by a scaling factor and added to the PSC and DE scores (Chen et al., 2003).

2.7.4 DOT

DOT (Daughter of Turnip) (Mandell et al., 2001) incorporates both van der Waals (geometric fit) and solvent continuum electrostatics provided by the Poisson-Boltzmann equation. Unlike the Coulombic model, this gives a more accurate energy term as it captures the dielectric constants of water, protein and lipid phases of a system. The Poisson equation finds the electric potential for a charge distribution and the Poisson-Boltzmann equation is obtained when the charge distribution of counter ions in the solution is added to the charge distribution of the protein. DOT calculates the free energy landscape and can generate a hot-spot or cluster of favourable energies

surrounding a crystallographic solution. Such a cluster of favourable free energies could imply an energy funnel and productive binding.

2.7.5 ClusPro

Docking algorithms such as GRAMM produce thousands of millions of putative complexes covering the relative positions of the two molecules producing a few native structures among many false conformations. These need to be evaluated to identify the native complex. Complexes can be first filtered assuming that the native complex is at a global free energy minimum using desolvation and electrostatic energies to eliminate many false complexes It is also assumed that the native structure is the global minimum where the energy landscape will be broadest and deepest but that the surface will also have local minima narrower and shallower than the global minimum. Therefore the centre of the most populated cluster is expected to be closest to the native complex (Comeau et al., 2004a).

ClusPro (Comeau et al., 2004a; Comeau et al., 2004b) is a fully automated web-based program for protein-protein docking. ClusPro uses either DOT or ZDOCK to perform rigid body dockings to generate many docked conformations. DOT uses a 128 Å x 128 Å x 128 Å grid with a grid spacing of 1 Å and 13,000 rotations to generate 2.7×10^{10} structures. ClusPro retains 20,000 of these based on surface complementarity scores which are then subjected to filtering by free energy. ZDOCK differs from DOT in that it combines shape complementarity with desolvation and electrostatics. Only 2000 of these structures are kept to be filtered.

The complexes generated by either ZDOCK or DOT are filtered using the atomic contact potential to give a statistical measure of desolvation free energy, and a Coulombic model to give the electrostatic free energy. While the atomic contact is a smooth potential and resistant to side chain conformation, the electrostatic potential is sensitive to incorrect rotamers and a larger number of electrostatic structures pass though the filter. To identify the global minimum the 2000 most favourable structures are then clustered based on a pairwise binding site RMSD. For each of the structures where the ligand protein has an atom within 10 Å of the receptor protein, the distance between the Ca of that complex and the Ca of the remaining complexes are calculated. Clusters are formed by selecting the ligand with the most neighbours. Clustering occurs until at least 30 clusters are formed (Comeau et al., 2004a).

2.7.6 FireDock

FireDock (Fast Interaction REfinement in molecular DOCKing) (Andrusier et al., 2007) is a method of refining and re-scoring rigid body docking solutions. Such refinement tries to improve the rigid body docking by considering the flexibility of the sidechains and backbone. Side chain flexibility is restricted to clashing residues at the interface. The interface is defined as a residue with at least one atom within 6 Å of any atom of the docking partner. Side chain flexibility is modeled by rotamers using the Dunbrack backbone dependent rotamer library. A rotamer set for each residue is created by collecting high probability rotamers. This is then followed by refining the relative position of the docking partners using Monte Carlo minimisation of the docking score. The refined candidates are then ranked by the docking score. The top 25 are refined again with full interface sidechain optimisation, in this stage the atomic radii are less smoothed.

The docking score is an ACE based free energy score which is an approximation of the binding free energy. The binding free energy is the change in the free energy of the system which occurs upon complex formation:

$$\Delta G = G^C - (G^R + G^L)$$

where G^C is the free energy of the receptor ligand complex and G^R and G^L are the free energies of the unbound receptor and ligand. As only interface residues can change their conformation the complex is split into interface and non-interface regions and the complex energy is split into intramolecular and intermolecular energy:

$$\Delta G = G_{intrf} = G^C_{intrf} - (G^R_{intrf} + G^L_{intrf})$$

$$= G^C_{intrf\_inter} + \Delta G^L_{intrf\_intra} + \Delta G^R_{intrf\_intra}$$

ΔG is then described by the change in intermolecular complex energy and change in interface intramolecular energies of each molecule caused by binding. The desolvation free energy is estimated by the ACE potential, and is defined as the free energy change when a bond between two atoms is replaced by solvent bonds. The electrostatic contribution is approximated by the pairwise Coulombic energy (Andrusier et al., 2007).

2.8 Electrostatics

An important feature of the function of proteins is their specificity of binding to one and other and electrostatic interactions play an important role in this. Charge-solvent interactions are also important factors in the binding of charged substrates to proteins

and protein folding, ionised amino acids prefer to occur on the surfaces of proteins due to the favourable interactions between them and solvent (Gilson and Honig, 1988). DelPhi (Gilson et al., 1988; Honig and Nicholls, 1995; Rocchia et al., 2002) can be used to calculate solution electrostatics, it solves the Poisson–Boltzmann equation (PBE) for a protein-solvent system. PBE relates the charges on the atoms to the electrostatic potential at all points in space taking into account the effects of solvent. The electrostatic potential can be visually represented and displayed, and proteins can have unique patterns that indicate a functional role (Honig and Nicholls, 1995).

2.9 Membrane binding

P450s are membrane bound enzymes attached to the endoplasmic reticulum, and the interaction with the membrane may have a role in its function (Schleinkofer et al., 2005).

2.9.1 MAPAS

MAPAS (Sharikov et al., 2008) was used to predict the number of residues in contact with the membrane and to predict the plane of attachment. MAPAS assumes that the planar region contacting the membrane must provide the binding energy to keep the protein at the membrane surface. MAPAS identifies all of the planar surfaces and produces a number of scoring methods to evaluate them. The membranephilic residues score (MRS) is calculated by using the membrane disengagement score W from a steered MD of each residue. An MRS>4 indicates support for membrane binding.

$$MRS = \left(\sum W(Nu_i \ * \ Su_i/Su_{imax}) \right) + \sum \left(\frac{(W \ * \ Nc_i)}{(Nu_i \ + \ Nc_i)} \right)$$

Equation 2.11 The membranephilic residues score.

where:
Nu_i = number of uncharged residues in the plane
Nc_i = number of charged residues in the plane
Su_i = solvent accessible surface of the uncharged residue that is included in the plane
Su_{imax} = maximum solvent accessible surface of the uncharged residue that is included in the plane

The membranephilic area score (MAS) assumes that the membrane contact area will contain membranephilic residues. An MAS of >40% indicates support for membrane binding.

$$MAS = \left(\frac{S_{top(plane)}}{S_{all(plane)}} \right)$$

Equation 2.12 The membranephilic area score.

$S_{top(plane)}$ = solvent accessible surface of the top 5 membranephilic residues in the plane

$S_{all(plane)}$ = solvent accessible surface of all residues in the plane.

Kmpha assumes that the overall surface properties will influence the tendency of the membranephilic region to bind to the membrane and is a measure of 'membranephilic asymmetry. A Kmpa of >2 indicates that the protein is membrane contacting.

$$K_{mpha} = \frac{(MAS)}{\left(\frac{S_{top(protein)}}{S_{all(protein)}} \right)}$$

Equation 2.13 The Kmpha score.

$S_{top(protein)}$ = the solvent accessible surface of the membranephilic residues of the entire protein

$S_{all(protein)}$ = the solvent accessible surface of all residues of the protein.

2.9.2 HotPatch

HotPatch (Pettit et al., 2007) predicts the locations of functional sites by finding patches of properties on protein surfaces. It can identify specific patches identified as the fraction of residues in the region that have a property. To identify a patch it evaluates the property of interest for all atoms of the protein and clusters the atoms with a high value of the property. It scores each patch to assign a statistical score to describe how probable it is for the patch to overlap a functional site.

2.10 Pharmacophores

A pharmacophore is a set of structural features in a ligand that are related to its recognition by a receptor and its biological activity. The spacial arrangement of these features to form a 3D pharmacophore represents the set of interactions of a ligand with its receptor. A given pharmacophore is limited in representing the mode of action of ligands that bind to the same target. There are two approaches to developing a pharmacophore, either the analysis of the receptor (structure based design) or from the set of ligands that bind to a target (ligand based design) (Wolber and Langer, 2005).

2.10.1 MOE

The MOE pharmacophore analysis module (Chemical Computing Group) uses a ligand-based approach. Features on a ligand are represented by points in space that are labeled to represent the pharmacophoric features at that location. A pharmacophore consensus can be used to generate a pharmacophore that is consistent with a set of aligned conformations. This consensus pharmacophore is composed of all features of all molecules.

2.10.2 LigandScout

LigandScout (Wolber and Langer, 2005) uses a structure based approach to pharmacophores. This method is associated with docking as the first step involves aligning the ligand into the receptor. It uses a rule set to automatically detect and classify protein-ligand interactions into H-bond, charge transfers and lipophilic interactions. This set of interactions forms the pharmacophore.

2.10.3 SARvision

Structure activity trends are important in drug discovery. SARvision (Reichard, 2008) is a tool to analyse patterns in collections of compounds. From these collections an algorithm identifies and sorts scaffolds into a tree of substructures. This tree allows the sorting or filtering of compounds, by clicking on a scaffold it creates a substructure filter that allows the interrogation of the subset, and as scaffolds can be sorted according to frequency and size, the structural distribution of a compound collection can be determined. Where a compound collection contains biological data, scaffolds can be sorted by this value to identify structure-activity trends for drug discovery.

2.11 Site of metabolism prediction

MetaSite (Molecular Discovery Ltd., London, UK) (Cruciani et al., 2005) predicts the metabolic biotransformations that are carried out by P450s based on both the reactivity and accessibility to the heme of ligand atoms. Each atom in the substrate is assigned a score that is proportional to the exposure of the substrate towards the heme and its reactivity.

MetaSite calculates two sets of descriptors, one for the protein and one for the substrate, based on GRID interaction fields. A probe is used to measure the potential at each point

and many different probes are used to represent different chemical groups to give information on how the protein might interact with the ligand.

The flexible mode of GRID was used. This allows the sidechains to move in response to attractive or repulsive interactions with the probe. Flexible interaction fields are independent of the initial sidechain position and can simulate the adaptation of the enzyme to the substrate by mimicking the movements that may occur. As substrates are flexible and as the conformation is relevant to substrate recognition, the substrate is modeled based on a low energy minimum conformation and the obtained conformers are induced by the interaction field and shape of the active site. Sets of descriptors of the protein and substrate are compared.

The prediction of the site of metabolism is based on the hypothesis that the distance from the reaction centre of the protein and interaction points in the cavity should correlate with the distance between the reactive centre of the molecule and the position of the different atom types in the molecule. Each atom in the substrate is assigned a score that is proportional to the exposure of the substrate towards the heme.

MetaSite calculates the reactivity of the atoms in the substrate. The formation of a radical species is needed for the reaction and R_i is the reactivity of atom i and represents the energy required to produce the intermediate. As ab initio methods to calculate R_i are slow, they collected druglike substrates for P450s and detected the non-redundant fragments and carried out ab initio calculations on the fragments to simulate hydrogen abstraction. Atomic positions were classified and ranked from stable (0.0) to strongly reactive (1.5). The fragments are recognised in the substrate and the R_i are assigned to the atomic positions (Cruciani et al., 2005).

2.12 Dynamics

Proteins can sample a large ensemble of conformations around an average structure but crystal structures and homology models are static snapshots. To understand protein function dynamics needs to be taken into account.

2.12.1 Normal modes

Normal mode analysis (NMA) is a computational method for studying the large amplitude motions that are involved in the function of proteins. It has been shown that

important transition pathways often follow the trajectories of one or a few low frequency normal modes. Protein structure is thought to have evolved to follow one or a few normal modes so that their structural flexibility facilitates the important conformations. NMA can be used to predict the conformational changes that a protein undergoes to carry out its function and protein motions can be represented as a superposition of normal modes fluctuating around an energy minimum (Suhre and Sanejouand, 2004). NMA can be used to address the problem of the flexibility of the protein during docking as it can be used to explore the collective motions that are functionally relevant and can provide structures that can be used for docking (Floquet et al., 2006).

2.12.2 ElNémo

The input to ElNémo (Suhre and Sanejouand, 2004) is a protein model that is the reference, and it determines the interaction matrix for the elastic model and computes its normal modes. The ElNémo normal mode calculation is based on an approximation of the energy function around a minimum energy conformation. The approximation allows the solution of the equations of motion by diagonalising the Hessian matrix. The eigenvectors of this matrix are the normal modes. A single-parameter Hookean potential is used which can yield low-frequency normal modes as accurate as those obtained by empirical force fields:

$$E_p = \sum_{d_{ij}^0 < R_c} c\left(d_{ij} - d_{ij}^0\right)^2$$

where d_{ij} is the distance between two atoms i and j, d_{ij}^0 is the distance between the atoms in the three dimensional structure, c is the spring constant of the Hookean potential, and R_c is a cut-off. This implies that the reference structure is the energy minimum conformation and all atoms have the same fixed value for the kinetic energy term. Building block approximation is used to group several residues into a single block super-residue. This approximation allows for a reasonable computational time and has little influence on the low frequency modes.

The degree of collectivity is calculated that indicates the fraction of residues that are affected by a given mode. Maximal collectivity is indicated by 1 while for motions that involve few atoms it approaches 0. Low frequency modes related to function are expected to be collective. Localised motions tend to correspond to extended parts of the

system like the termini and are usually meaningless. ElNémo produces normal mode perturbed models that are the original reference structure with a perturbation that corresponds to the normal mode applied to every atom (Suhre and Sanejouand, 2004).

2.12.3 Identification of protein domains

Proteins can be crystallised in more than one conformation. They may have open and closed conformations, and these are thought to represent the conformations that are accessible under normal conditions. If a protein has more than one conformation, domains can be defined based on the groups of residues that move in a concerted fashion and the function of a protein can be can be related to the domain motions. DynDom (Hayward and Berendsen, 1998) identifies protein domains. Domains can be identified based on their differing rotational properties and can be identified from a structure and a set of displacement vectors. Such displacement vectors can be determined from a normal mode analysis.

2.12.4 Molecular dynamics

NAMD (Phillips et al., 2005) was designed to be an easy to use molecular modelling tool allowing access to dynamic information extrapolated from structures. NAMD simulates how the atoms of a molecule move according to Newtonian equations of motion.

The total potential energy is dependent on all atomic positions and couples the motions of atoms. Potential energy is represented through the MD force field and is the most crucial part of the simulation as it must represent the interaction between atoms. For an all atom MD, all atoms experience a force specified by the force field that accounts for the interaction of that atom with the rest of the system. NAMD uses a common potential energy function:

$$U_{total} = U_{bond} + U_{angle} + U_{dihedral} + U_{vdw} + U_{coulomb}$$

where U_{bond}, U_{angle} and $U_{dihedral}$ describe the stretching, bending and torsional bonded interactions:

$$U_{bond} = \sum_{bonds\ i} k_i^{bond} (r_i - r_{0i})^2$$

$$U_{\text{angle}} = \sum_{\text{angles } i} k_i^{\text{angle}} (\theta_i - \theta_{oi})^2$$

$$U_{\text{dihedral}} = \sum_{\text{dihedral } i} \begin{cases} k_i^{\text{dihe}}[1 + \cos{(n_i \Phi_i - \gamma_i)}], n_i \neq 0 \\ k_i^{\text{dihe}} (0_i - \gamma_i)^2 n = 0 \end{cases}$$

where *bonds* counts each covalent bond, *angles* are the angles between each pair of bonds sharing an atom and *dihedral* describes atom pairs separated by three covalent bonds and the central bond subject to a torsional angle Φ.

U_{vdw} and U_{coulomb} describe interactions between non bonded atoms corresponding to the van der Waals force approximated by a Lennard-Jones 6-12 potential, and electrostatic interactions:

$$U_{\text{vdW}} = \sum_i \sum_{j>i} 4\epsilon_{ij} \left[\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}}\right)^{6} \right]$$

$$U_{\text{Colomb}} = \sum_i \sum_{j>i} \frac{q_i q_j}{4\pi \epsilon_0 r_{ij}}$$

For every particle the parameters such as k_i^{bond} and r_{0i} for the interactions are stated in the force field parameter files. In this study the parameters from CHARMM are used.

To avoid surface effects at the boundary of the simulated system, periodic boundary conditions are used. The system is enclosed in a cell that is replicated by periodic translations so that when a particle leaves the cell on one side, it is replaced by a copy entering the cell on the opposite side. Each particle is subject to the potential from all other particles in the system including the images in the surrounding cells, thereby eliminating surface effects, although under the minimum-image convention each particle interacts with the closest image.

3. Modelling and characterisation of P450s

3.0 Preface

This chapter covers the homology modelling and characterisation of a number of pyrethroid metabolising and non-metabolising P450s. Many of these have been identified by studies done in collaboration with the Liverpool School of Tropical Medicine where the expression and characterisation of these proteins were carried out while others have been identified from the literature.

3.1 CYP6 family

3.1.1 Evolutionary relationships between metabolisers

The CYP6 family are related to the CYP3 and CYP5 families and are the most numerous of the insect P450 genes (Feyereisen, 2006). Sasabe et al. (2004) noted that in insects, a number of genes linked to insecticide resistance appear to be related to the CYP6 family. There have been studies demonstrating the metabolism of cyclodiene insecticides by CYP6A1 (Andersen et al., 1994), and pyrethroids by CYP6D1 and CYP6B8, and it has been suggested that there could be an evolutionary relationship between the allelochemical-inducible CYP6 genes and resistance. This suggestion has been further supported by the finding that CYP6A1, CYP6B8 and CYP321A1 can metabolise plant compounds. CYP6A1 was found to metabolise farnesyl, geranyl and neryl esters of plant origin (Sasabe et al., 2004), while CYP6B8 and CYP321A1 metabolised a wide range of plant compounds as well as insecticides (Li et al., 2004; Sasabe et al., 2004).

It is possible there could be evolutionary relationships between the pyrethroid metabolisers of humans and insects. The insect pyrethroid metabolising CYP6s are in the CYP3 clade and are related to vertebrate CYP3 and 5 families (Feyereisen, 2006), but show a high sequence identity to the CYP3A subfamily, members of which have been shown to metabolise pyrethroids. The insect CYP6 family has been considered to have only recently diverged from the CYP3 family, estimated divergence time between 470 mya and 80 mya (Lewis et al., 1998). Here the *Anopheles* CYP6 and CYP9 are also identified as pyrethroid metabolising enzymes also belonging to the CYP3 clade.

However, functional conservation between homologs does not explain all cases of pyrethroid metabolism. Not all pyrethroid metabolising enzymes occur within the CYP3 clade. Members of the mammalian CYP1 and 2 families have also been found to metabolise pyrethroids. The CYP1 and 2 families diverged from the CYP3 clade an

estimated 800 mya (Lewis et al., 1998) and the rat CYP1A2 has been shown to be the most active P450 in deltamethrin metabolism. The overlapping substrate specificity for pyrethroids may be due to commonalities in active sites between pyrethroid metaboliers.

3.1.2 Template selection

Members of the CYP6 family were individually submitted to the PSI-BLAST server for three iterations. CYP3A4 was consistently identified as among the top hits and as having the highest sequence identity. In addition, the CYP6s have been identified as being members of the CYP3 clade along with CYP3A4. This is also shown by a neighbour-joining tree (Figure 3.1). The selection of CYP3A4 as a template was further supported by secondary structure predictors as the CYP6 sequences are predicted to have similar structural features as CYP3A4 (1TQN). Alignments to this template were submitted to Modeller v8.0 using the input script in Appendix A and the lowest energy models were selected.
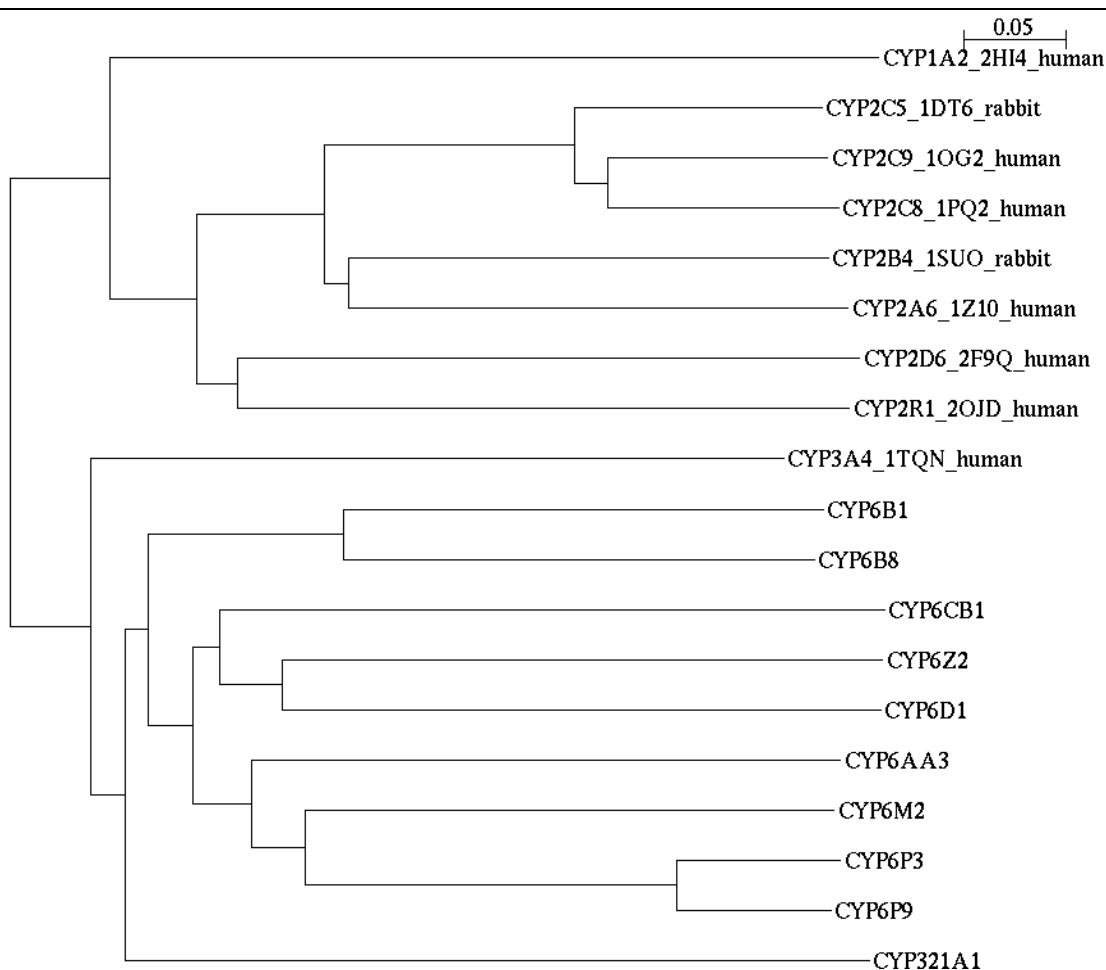
Figure 3.1 ClustalW neighbour joining tree for the insect CYP6s and the hits from PSI-BLAST labelled as PDB ID_CYP_species.

3.1.3 Anopheles gambiae CYP6M2

Muller et al. (2007) compared the pyrethroid resistant Odumasy strain with the susceptible Kisumu strain and found that a number of genes including CYP6M2 are over expressed in the resistant *Anopheles gambiae*.

3.1.3.1 Alignment

CYP3A4 was identified as the closest related to CYP6M2 and had the highest identity with a 30% identity over the aligned region or 33% over length the shortest sequence. CYP3A4 can also metabolise similar compounds such as deltamethrin and luciferin. 3DCoffee was used to produce an alignment. However, an insert occurred in the BC loop that affected the positioning of residues in the active site. The 3DCoffee output was altered by moving the position of the insert as was suggested by a multiple alignment of related CYP3 clade pyrethroid metabolisers (Figure 3.2). In the model this insert is positioned at the surface of the protein which is consistent with the insertion of polar residues.
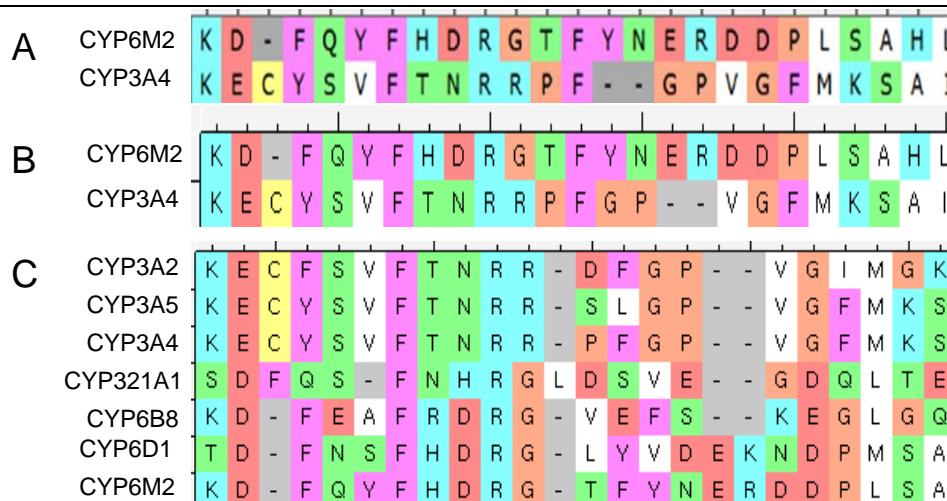
Figure 3.2 Alignments of CYP6M2 with the template CYP3A4. (A) 3DCoffee output. (B) Edited alignment. (C) Multiple alignment of CYP3 clade pyrethroid metabolisers. The alignment display colours the amino acids according to physicochemical property (polar positive: blue; polar negative: red; polar neutral: green; non-polar aliphatic: white; non-polar aromatic: purple; Pro and Gly: brown; Cys: yellow).

3.1.3.2 CYP6M2 model evaluation

From an output of 30 structures from modeller a model was selected based on ERRAT profile and MODELLER score. The model chosen showed few regions of disorder and a good ERRAT score, and most disordered regions that were present corresponded to

inserts. The disordered region between residues 50 and 60 corresponded to a partial unwinding of the A helix due to the substitution for a proline within the helix (Figure 3.3).
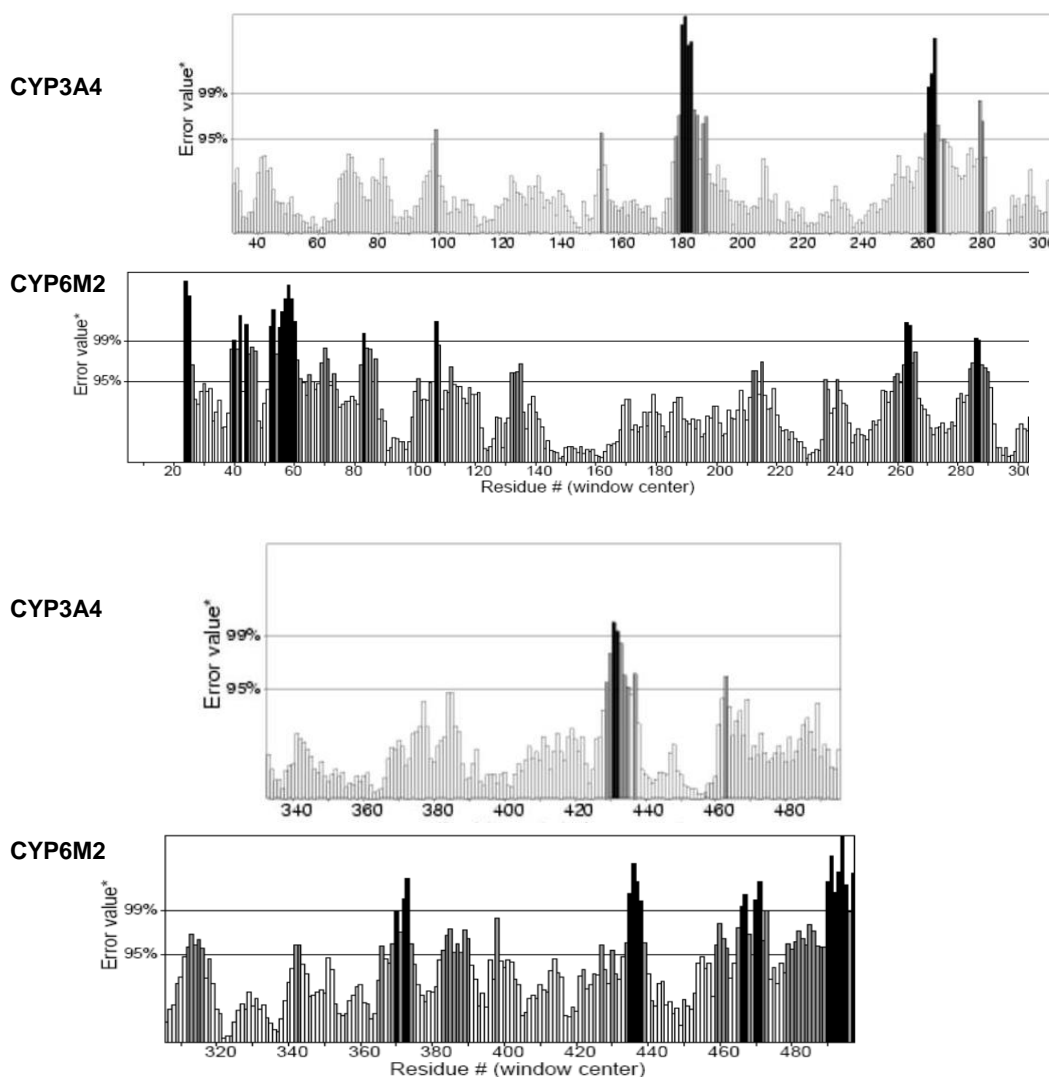
Figure 3.3 Errat profile of the CYP6M2 model and template CYP3A4. * On the error axis the two lines represent the confidence with which it is possible to reject regions that exceed that error value. The overall quality is the percentage of the protein for which the error value is below the 95% rejection limit, for CYP6M2 this is 74%.

3.1.3.3 Validation

The activity of CYP6M2 towards luciferins has been determined for range of luciferins (Figure 3.4). The scores of the dockings reflect the activity towards the luciferins. The preferred substrates (L-PPXE, L-PFBE and L-BE) show higher scores than the poor substrates (L-H, L-ME) except for L-CEE (Table 3.1). The first ranked mode of PPXE was too distant for metabolism at 5.2 Å. A constrained docking was used to place the PPXE site of metabolism close to the heme between 1.5-4.5 Å. Although this was a

91

constrained docking, it produced a higher score than the free docking and indicates that PPXE is able to bind for metabolism with a high score (Table 3.2). Genetic algorithms like GOLD may not find the optimal solution but find reasonable solutions. The greater score of the constrained docking may be due to a smaller search space allowing for a denser sampling of conformations to find a more optimal solution.
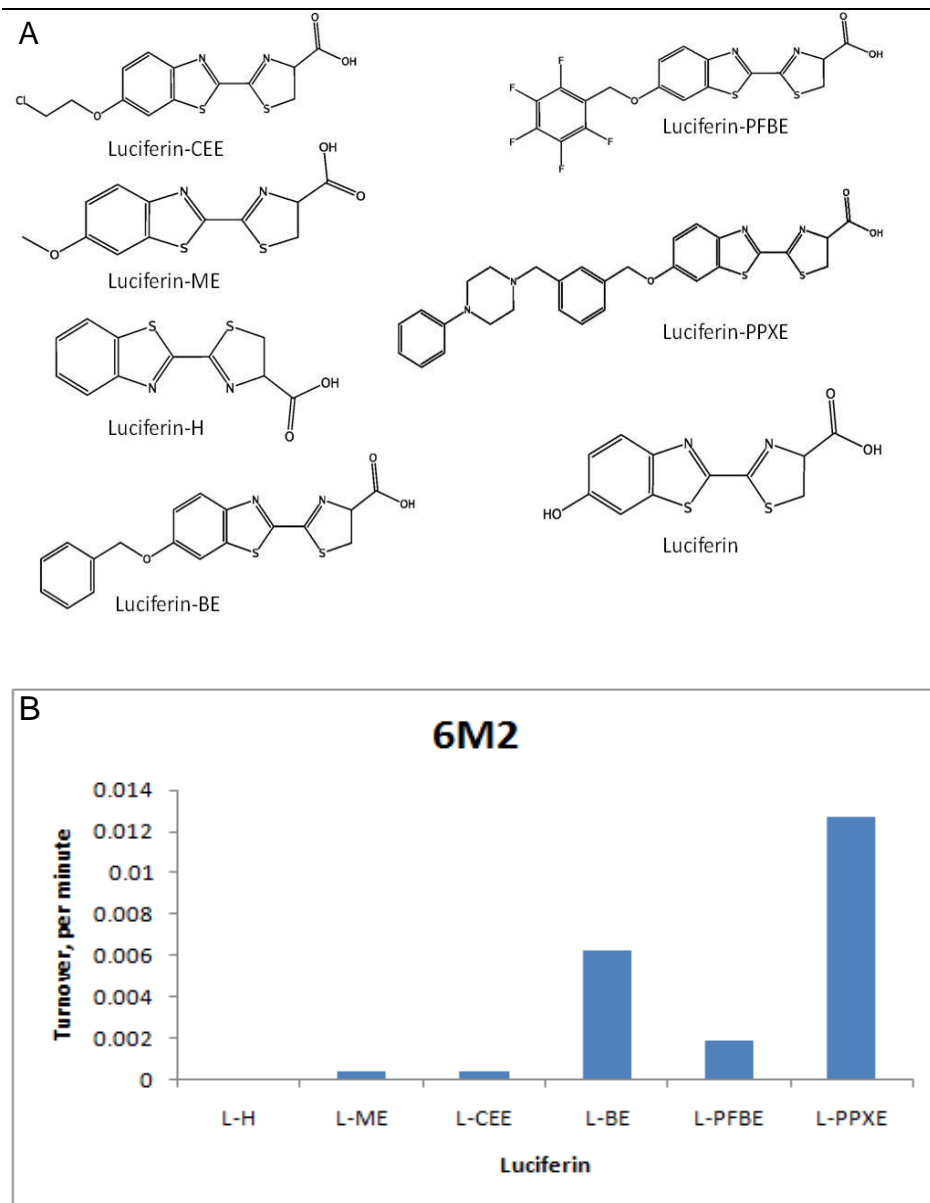
Figure 3.4 (A) The luciferin compounds tested: 6′-deoxyluciferin (-H), luciferin 6′-methyl ether (-ME), luciferin 6′-chloroethyl ether (-CEE), luciferin 6′ benzyl ether (-BE), luciferin 6′-pentafluorobenzyl ether (-PFBE) and luciferin-6' phenylpiperazinylyl ether (-PPXE). (B) Activity of CYP6M2 towards the luciferins: 6′-deoxyluciferin (L-H), luciferin 6′-methyl ether (L-ME), luciferin 6′-chloroethyl ether (L-CEE), luciferin 6′ benzyl ether (L-BE), luciferin 6′-pentafluorobenzyl ether (L-PFBE) and luciferin-6' phenylpiperazinylyl ether (L-PPXE) (McLaughlin, unpublished).

Table 3.1 ChemScores (kJ/mol) of productive modes of luciferins in CYP6M2.

	rnk	Score	ΔG	S(hbond)	S(metal)	S(lipo)	H(rot)	ΔE(clash)	ΔE(int)
PPXE	1st	52.4	-57.9	0.5	0.0	484.9	2.3	2.5	2.9
BE	1	47.1	-49.0	0.9	0.0	389.5	2.1	0.7	1.2
PFBE	1	36.2	-44.3	0.9	0.0	361.9	2.7	6.2	1.8
ME	1	33.8	-42.4	1.7	0.0	306.6	1.8	8.2	0.4
CEE	1	38.8	-45.8	1.6	0.0	344.4	2.1	6.7	0.4
H	1	33.7	-41.0	1.9	0.0	283.9	1.5	7.1	0.3

Table 3.2 ChemScore (kJ/mol) of a restrained docking of PPXE in CYP6M2.

	Score	ΔG	S(hbond)	S(metal)	S(lipo)	H(rot)	ΔE(clash)	ΔE(int)
PPXE	65.8	-70.5	1.8	0.0	555.7	2.3	3.2	1.5

To detect the fluorescent product, metabolism needs to occur on the carbon next to the oxygen to break the C-O bond and release the fluorescent luciferin. The binding modes of -ME, -CEE and the constrained -PPXE are similar and occur with the body of the luciferin perpendicular to the heme and H-bonds to residues of the G'G loop. In this position the body of the luciferin can π-stack with F123 and form hydrophobic interactions with H121 and the I helix, two H-bonds occur with Q241 and T242 of the G'G loop. The chain of PPXE can form additional aromatic interactions as it can π-stack between F110 of the BC loop and F367 of SRS5 (Figure 3.5, A and B).
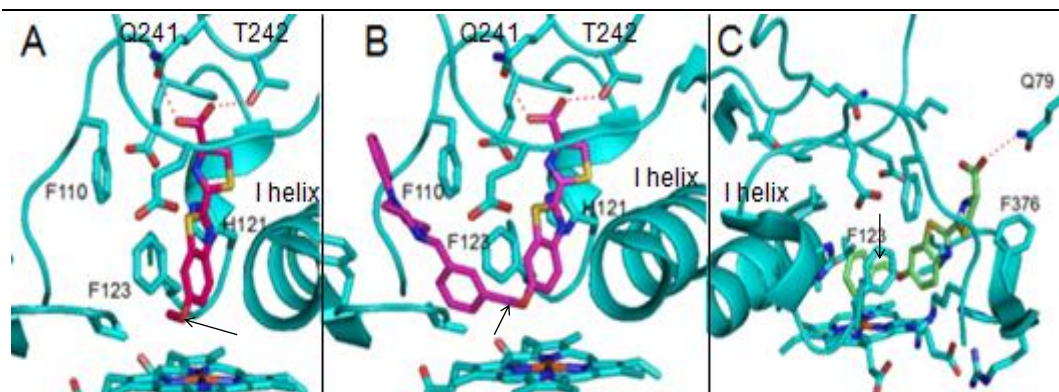
Figure 3.5 CYP6M2 binding -ME free docking (A), -PPXE constrained docking (B) and –BE free docking (C). Sires of metabolism indicated by arrows.
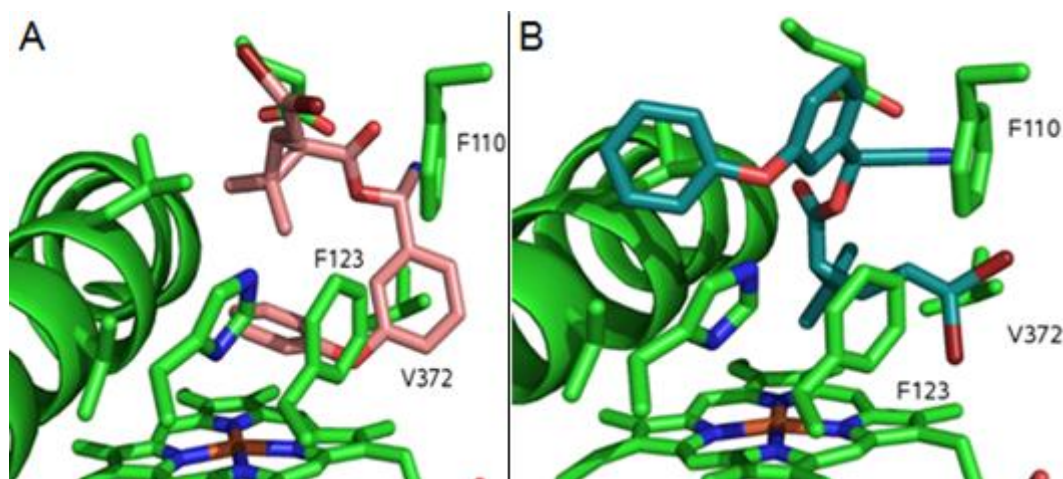
Figure 3.6 Two binding modes of deltamethrin in CYP6M2, binding for metabolism at the 4' (A) and the trans methyl (B).

The binding mode of -BE and -PFBE differ, with the body of the luciferin towards SRS5 forming hydrophobic interactions with F376, the benzyl ring and the aromatic ring of the body stacks around F123 while a H-bond occurs with Q79 of beta strand 1-1 (Figure 3.5, C). CYP6M2 has also been found to metabolise deltamethrin at the 4' and trans methyl. Deltamethrin binds according to the known regiospecificity with a mode that allows metabolism on the trans methyl and one that allows metabolism on the 4' (Figure 3.6).

3.1.4 *Anopheles gambiae* CYP6P3

CYP6P3 has been found to be upregulated in pyrethroid resistant *A. gambiae* populations (Djouaka et al., 2008). A PSI-BLAST search of the CYP6P3 sequence identified CYP3A4 as a template and a neighbour joining phylogenetic tree also confirms homology to the CYP6 family and the CYP3 clade. The CYP6P3 sequence was aligned to the template using 3DCoffee. However, this automated alignment may mis-position an insert. A structural re-alignment was carried out using secondary structure predictors to predict the structure and then realigning structural elements to the template.

In the template the A helix is followed by a short loop of a single residue, and then beta strand β1-1, loop, β1-2. Compared to CYP3A4, CYP6P3 has an insert at this position. The 3DCoffee alignment positions this insert between beta strands β1-1 and β1-2. However, the secondary structure predictions suggest that this insert could form a beta strand rather than a coil. The insert was repositioned to between helix A and strand β1-1 to align predicted secondary structure to that in the template (Figure 3.7). Models based

on the unedited 3DCoffee output show a region of disorder in the region of the insert around residue 60, but models based on the structural re-edit show less disorder in this region. In models based on both alignments there are regions of disorder around positions 260 and 280 that correspond to large inserts in the GH and HI loops (Figure 3.8).

Figure 3.7 Alignments of CYP6P3 with the template CYP3A4. (A) The 3DCoffee output. (B) The realignment based on secondary structure prediction.
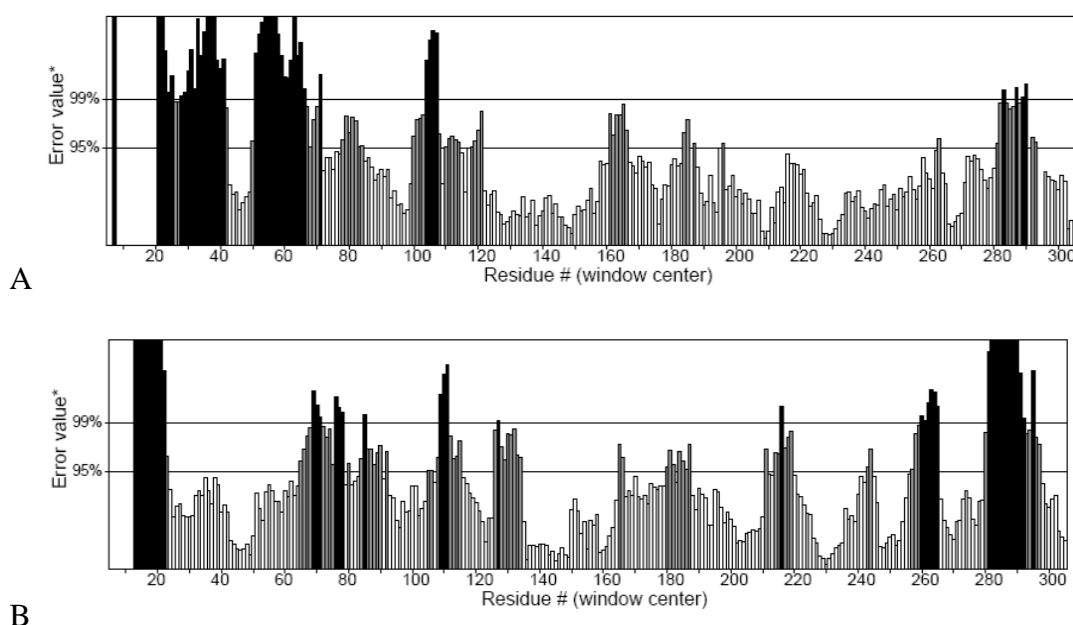
Figure 3.8 Errat profiles of the model based on the 3DCoffee output (A) and the model based on the structural realignment (B).

3.1.4.1 Validation

CYP6P3 has been found to metabolise deltamethrin with a similar HPLC metabolite profile as CYP6M2 and sites of metabolism are suggested to be on the trans methyl and 4' (Figure 3.9) . The best ranked docking mode of deltamethrin is consistent with the experimental findings as it is positioned for metabolism on the trans methyl. This mode may occur due to an aromatic or hydrophobic network formed by F309, H121, F123 and F110 that is in a position to bind the alcohol group distant from the heme and allow the acid group to approach the heme. In this mode the phenyl ring forms hydrophobic interactions with F309 of the I helix and H121, the benzyl ring can π-stack with F110. The residues immediately above the heme V380 and F123 form hydrophobic

interactions with the acid group and these vdW contacts may act to stabilise the site of metabolism. Although it was not found, a H-bond between the phenoxybenzyl O and the histidine H121 is possible (Figure 3.9).

The activity of CYP6P3 towards a number of luciferins has been determined (Figure 3.10). The docking scores reflect the activities: -PPXE and -BE are efficiently metabolised and show high docking scores, while the poor substrates show poor scores (Table 3.3). Although modes that allow metabolism of PPXE are low ranking, they are higher scoring than the highest ranked dockings of the other luciferins.
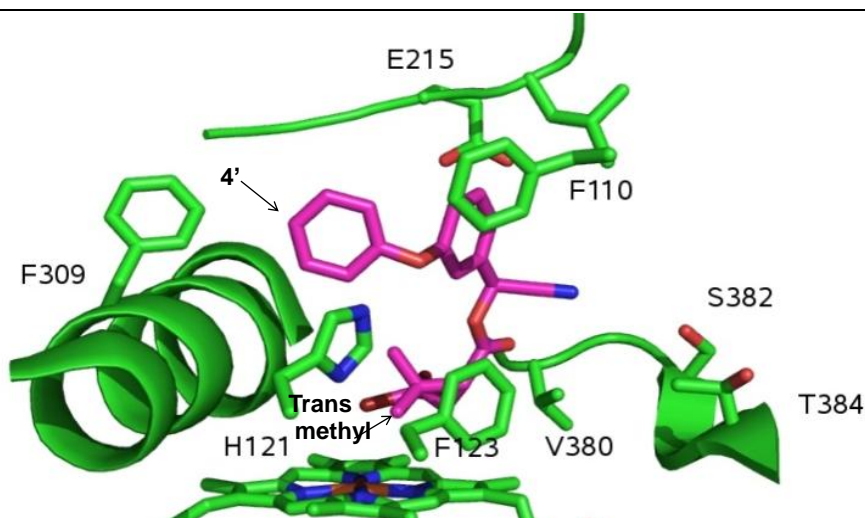
Figure 3.9 CYP6P3 binding deltamethrin for metabolism on the trans methyl.
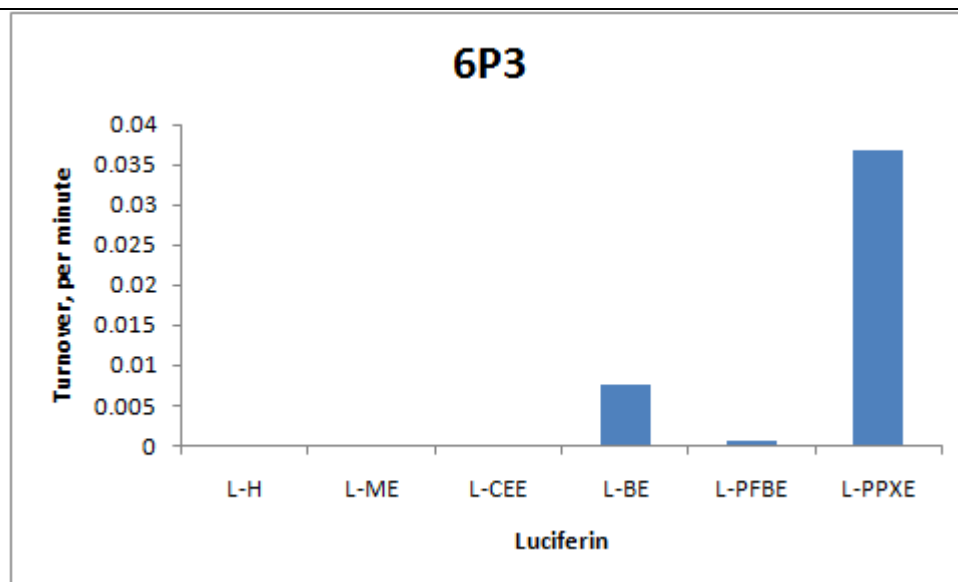
Figure 3.10 Activity of CYP6P3 towards luciferins: 6′-deoxyluciferin (L-H), luciferin 6′-methyl ether (L-ME), luciferin 6′-chloroethyl ether (L-CEE), luciferin 6′ benzyl ether (L-BE), luciferin 6′-pentafluorobenzyl ether (L-PFBE) and luciferin-6' phenylpiperazinylyl ether (L-PPXE) (McLaughlin, unpublished).

Table 3.3 The ChemScores (kJ/mol) of docking modes of luciferins that allow metabolism in CYP6P3.

	rnk	Score	ΔG	S(hbond)	S(metal)	S(lipo)	H(rot)	ΔE(clash)	ΔE(int)
PPXE	54	51.9	-60.8	0.0	0.0	523.7	2.3	6.4	2.6
BE	1	47.4	-49.9	0.9	0.0	399.5	2.1	1.5	1.2
PFBE	1	37.2	-46.4	0.7	0.0	388.1	2.7	5.4	3.8
ME	1	33.9	-35.1	0.9	0.0	264.9	1.8	0.5	0.6
CEE	1	38.2	-39.0	0.9	0.0	307.1	2.1	0.3	0.6
H	1	34.9	-35.4	0.9	0.0	260.6	1.5	0.3	0.0

Luciferin-BE and -PFBE have similar structures differing only in that the H of the phenyl ring are replaced by fluorine in -PFBE. Despite these similarities, CYP6P3 metabolises -BE well and -PFBE poorly. In the dockings the best ranked modes of both of these compounds bind for production of the fluorescent product but have different scores. The differences between the scores may be due a larger clash score in -PFBE, the -PFBE fluorines are larger than hydrogens and may provide steric clashes that are indicated in the higher clash score. However, the fluorine has a small size similar to hydrogen. Its vdW radius is 1.47 Å similar to that of hydrogen (1.2 Å) and can mimic hydrogen in ligands, and is unlikely to have a steric effect. Fluorine is known to have other effects on ligand metabolism, fluorine increases stability towards oxidation (Jeschke, 2004), and in the case of -PFBE it could stabilise the aromatic ring to oxidative attack.

The preference for metabolism may be due to the size of the substrate, smaller substrates have fewer vdW contacts and show poor lipophilic scores while large ligands such as -PPXE can have more contacts producing a higher score. The activities may also be due to stability, small substrates may be more mobile within the active site while larger ligands with optimal contacts are more stable and can be metabolised more efficiently. However, some ligands such as -PFBE can only bind in positions for metabolism with clashes due to the shape of the active site.

3.1.5 *Anopheles funestus* CYP6P9

Anopheles funestus is a major African malaria vector. As no target site resistance has been found in this species, resistance is mediated by detoxification of pyrethroids by P450s and Amenya et al. (2008) found that CYP6P9 is overexpressed in pyrethroid resistant strains. CYP6P9 is the ortholog of CYP6P3 and is similar in sequence, and so CYP3A4 was also selected as a template. The insert near helix A is also conserved between the proteins, and structural prediction suggested that this should be situated

between helix A and beta strand 1-1 and this was manually realigned (Figure 3.11). Similar to CYP6P3, this also produced an improvement in errat profile in the region, and shows similar areas of disorder in the GH and HI loops (Figure 3.12).
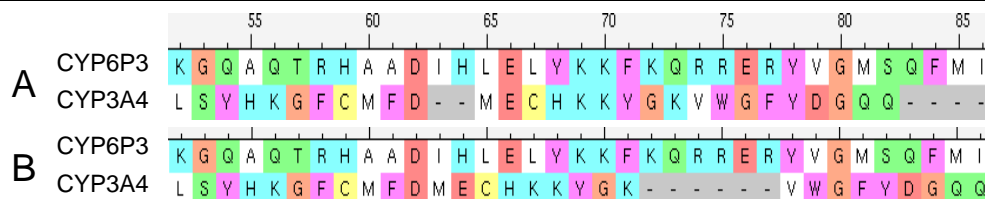
Figure 3.11 Alignment of CYP6P9 with the template CYP3A4. (A) The 3DCoffee output; (B) realignment based on structural prediction.
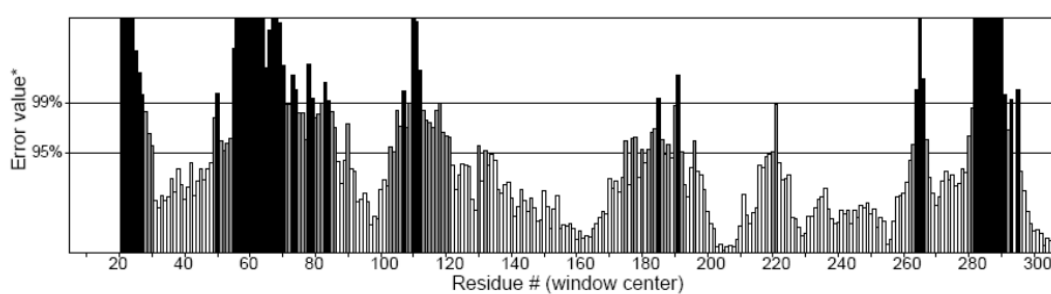
A
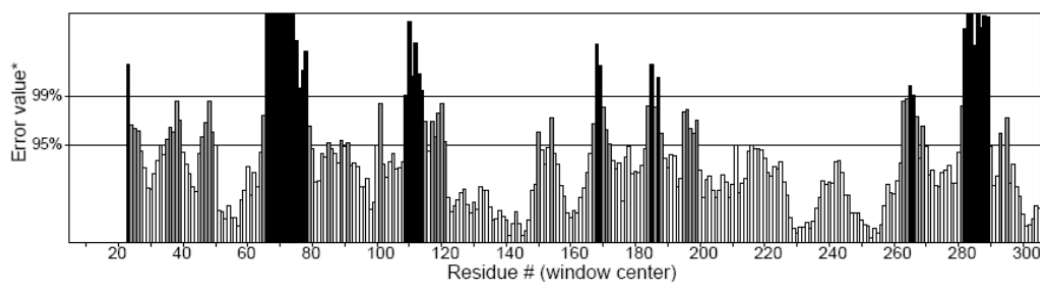
B

Figure 3.12 Errat profiles of the realigned region of CYP6P9. The 3DCoffee output (A) and manual re-alignment (B).

3.1.5.1 Validation

As with CYP6P3, the activity of CYP6P9 towards luciferins has been determined (Figure 3.13). The binding scores are consistent with the activity of the enzyme, the preferred substrates -BE and -PPXE showing higher binding scores than the poor substrates (Table 3.4).
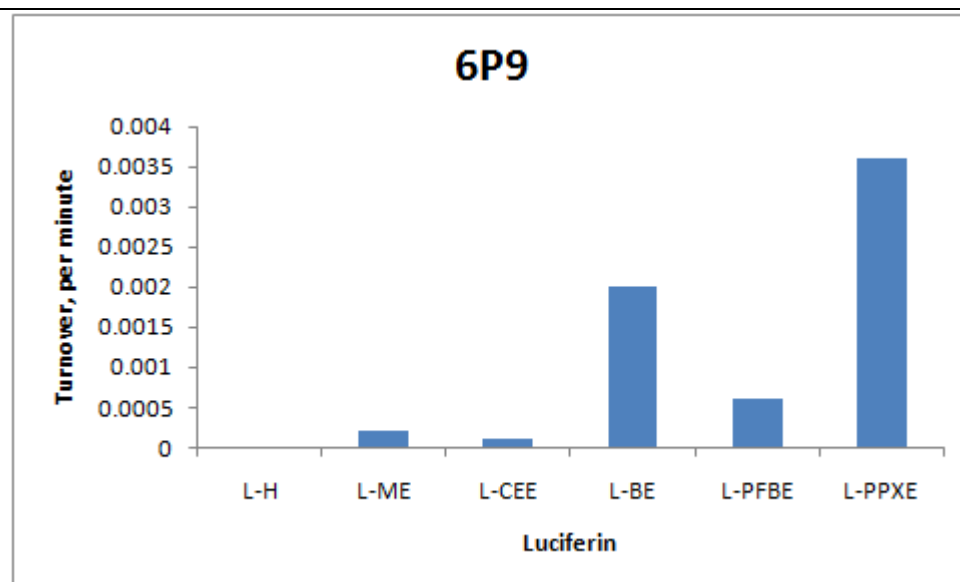
Figure 3.13 The activity of CYP6P9 towards luciferins (McLaughlin, unpublished).

Table 3.4 The ChemScores (kJ/mol) of productive modes of luciferins in CYP6P9.

	rnk	Score	ΔG	S(hbond)	S(metal)	S(lipo)	H(rot)	ΔE(clash)	ΔE(int)
PPXE	31st	54.3	-66.4	1.0	0.0	542.5	2.3	9.8	2.2
BE	1st	47.8	-49.2	2.8	0.0	339.1	2.1	1.2	0.2
PFBE	6th	34.2	-40.5	1.5	0.0	313.8	2.7	0.7	5.6
ME	1st	34.2	-35.3	2.8	0.0	215.6	1.8	0.2	0.8
CEE	5th	36.6	-38.5	2.9	0.0	246.9	2.1	0.9	0.9
H	1st	37.9	-38.5	2.7	0.0	238.2	1.5	0.5	0.2

3.1.6 *Anopheles gambiae* CYP6Z family

CYP6Z2 was identified as being over expressed in the DDT resistant strain XAN/U
compared to the resistant strain kisumu. CYP6Z1 is expressed in the permethrin
resistant Odumasy strain. Members of the CYP6Z family have been modelled
previously by Chiu et al. (2008) and Mclaughlin et al. (2008). Chiu et al. (2008)
modelled CYP6Z1 from the DDT resistant RSP strain and CYP6Z2 from the insecticide
susceptible PEST strain, and compared their activities and active sites (Table 3.5).

Table 3.5 Taken from Chiu et al. (2008) the activities of CYP6Z1 and CYP6Z2.

Activity (nmol/min/nmol P450)

Substrate	6Z1	6Z2
DDT	3.91	ND
Carbaryl	2.20	0.49
Xanthotoxin	1.51	ND

However, these results contradict those of Mclaughlin et al. (2008) where CYP6Z2 was found to metabolise xanthotoxin, and where xanthotoxin was found to be a mechanism based inhibitor. Xanthotoxin has been shown to act as a mechanism based inhibitor of P450s caused by epoxidation and opening of the furan ring. Mao et al. (2006) suggested that metabolism of xanthotoxin is likely to occur by epoxidation of the 2'-3' position because of the higher electron density of the electron rich π bond (Figure 3.14). Letteron et al. (1986) suggested that epoxidation forms an unstable radicaloid that may bind covalently to P450s to inactivate them.
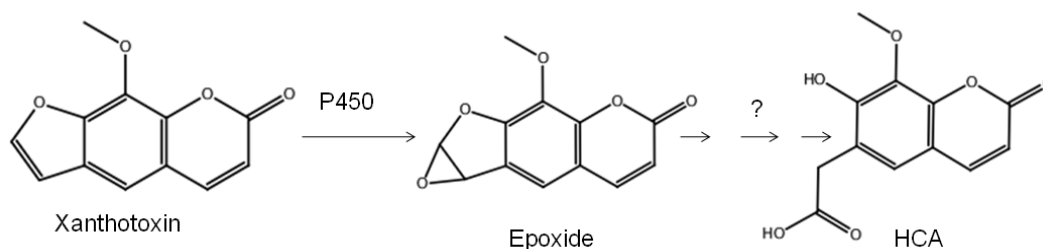
Figure 3.14 Metabolism of xanthotoxin Taken from Mao et al. (2006).

It is unclear what may produce the differences between the results of Mclaughlin et al. (2008) and Chiu et al. (2008). In both studies, the same CYP6Z2 sequence was used but in the Mclaughlin et al. (2008) study CYP6Z2 was co expressed only with Anopheles gambiae P450 reductase while in the Chiu et al. (2008) study this was co-expressed with both housefly P450 reductase and fruit fly b5.

Chiu et al. (2008) found that the activities of CYP6Z1 and CYP6Z2 differed and suggested that the differences could be due to steric restrictions in the CYP6Z2 active site. Their dockings into the active sites indicated that while DDT and xanthotoxin could dock close to the heme in productive modes in CYP6Z1, they were either distant or in unproductive modes in CYP6Z2, while carbaryl could dock productively in both CYP6Z2 and CYP6Z1. They overlaid their CYP6Z2 model with the CYP6Z1 model with DDT bound to identify clashes between DDT and the side chains of CYP6Z2. They suggested that substitution or positioning of three active site residues (F115, G298I, N208R) caused clashes. R208 in CYP6Z2 is substituted for Asn in CYP6Z1, while they did not find any steric clashes with this residue, it projects substantially farther into the active site than Asn. The side chain of I298 of CYP6Z2 is longer than G298 in CYP6Z1 and was suggested to occur at a position on the I helix that could affect substrate binding. While F115 is conserved in both proteins, it projects further

into the active site of CYP6Z2 causing clashes. The substitution of small residues in CYP6Z1 for bulkier residues in CYP6Z2 reduces the active site volume (3,471 Å^3 in CYP6Z1; 2,601 Å^3 in CYP6Z2). They suggested that the three residues constrain the active site of CYP6Z2 to give unproductive positionings of DDT and xanthotoxin.

3.1.6.1 Sequence Alignments

The CYP6Z family were also modelled in this study. An alignment with the 6Z family show that CYP6Z2 and CYP6Z3 are very similar but that CYP6Z2 differs in the FG region due to the substitutions of polar and aromatic residues for aliphatic. The aromatic residues that are suggested to prevent metabolism in CYP6Z2 (F212, F222) are conserved in all 3 family members (Figure 3.15).

In CYP6Z1 the substitution of one Phe for a Ile at position 216 occurs at the opening of the access channel 3 between the F and G loops and is in a position that may affect the access of hydrophobic substrates from the membrane (Figure 3.15). Another difference between CYP6Z1 and CYP6Z2 is in the I helix at position 302. in CYP6Z1 and CYP6Z3 this is a Glu while in CYP6Z2 this is an Asp. This position is equivalent to Asp251 of P450 101 which has been proposed to be involved in O_2 activation. It has been suggested that an Asp may be more effective than a Glu as in P450d the mutation of the wild type Glu to Asp increases activity (Ishigooka et al., 1992). In CYP1A2, the mutation of Glu318Asp increases Kcat with no change in Km. It was suggested that the Asp carbonyl may be better positioned than the Glu (Hiroya et al., 1994).

Models of CYP6Z2 have been produced previously by Chiu et al. (2008) and Mclaughlin et al. (2008). The model produced by Chiu et al. (2008) used a different alignment to that of Mclaughlin et al. (2008). In the Chiu et al. (2008) model, R208 projected into the active site rather than R210. Here a third CYP6Z2 model was produced with R208 projecting into the active site. However, the Chiu et al. (2008) alignment was discarded as it produced models with poor errat scores and misaligned a number of residues (Figure 3.16).
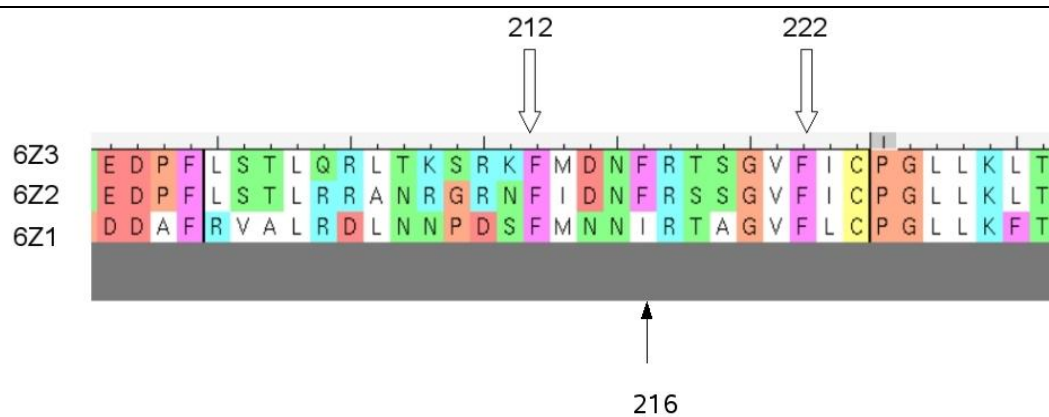
Figure 3.15 The alignemnt of the CYP6Z family. The positions of residues predicted to affect metabolism are indicated by arrows.
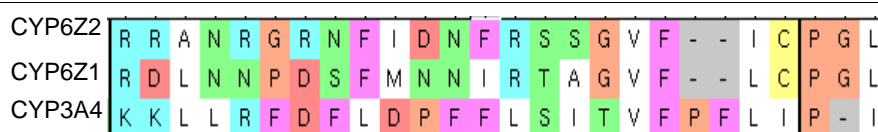
Figure 3.16 Alignment produced by Chiu et al. (2008).

The alignments used to make the modes in this study involved finding an optimum alignment to the template for each protein and to give good errat scores. Unlike Chiu et al. (2008), different alignments were used for CYP6Z1 and CYP6Z2 because of substitutions between them, although this produced different positions for inserts. As substitutions between N and D are favoured while the substitutions of R and D are less favourable, in CYP6Z2 the N is aligned with the D of the template. The CYP6Z1 alignment positions the gap in a different position but aligns the D with a D in the template (Figure 3.17, A and B). The CYP6Z3 alignment is similar to that of CYP6Z2 (Figure 3.17, C).

Figure 3.17 Alignments of the BC region of CYP6Z2 (A), CYP6Z1 (B) and CYP6Z3 (C).

CYP6Z1 overall quality factor 66%
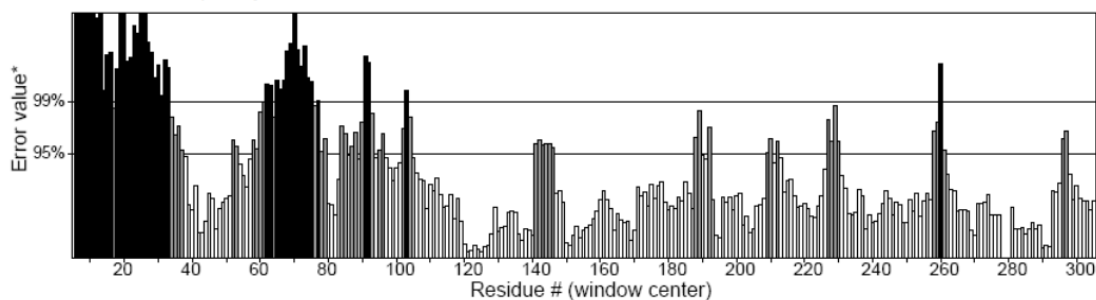
CYP6Z3 overall quality factor 66%
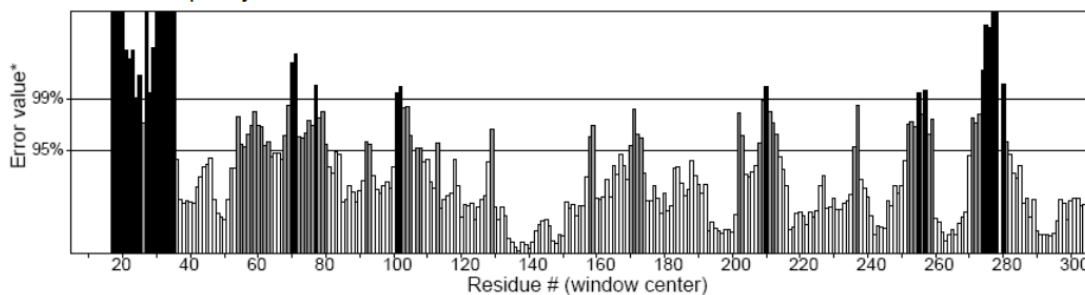
CYP6Z2 overall quality factor 66%
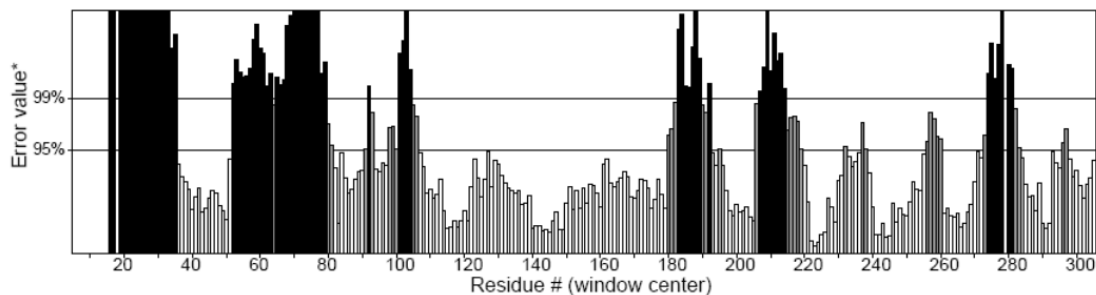
Figure 3.18 ERRAT scors for CYP6Z1/2/3 manually aligned region.

Although both 1TQN and 1W0E were used as templates, the positioning of the basic residue within the active site was similar to 1TQN. As this residue is known to rotate out of the active site as its position varies between the 1TQN and 1W0E structures, it was repositioned using the MOE rotamer explorer. Rotamers that were close to that of the crystal structure 1W0E were chosen and only higher scoring rotamers than the one present in the modeller output were used. The ERRAT quality of the three models were consistantly high with areas of high error scores primarily occuring at the C and N termini (Figure 3.18).

3.1.6.2 CYP6Z1 Verification

It has been determined experimentally that CYP6Z1 is able to metabolise xanthotoxin, DDT and carbaryl (Chiu et al., 2008). This CYP6Z1 model is able to bind xanthotoxin and carbaryl. Although the metabolites are unknown, the model binds these in positions that correspond to known sites of metabolism in other P450s. Xanthotoxin binds for

epoxidation at the 2'3' bond (Figure 3.19, A), suggested to be the site of metabolism by Mao et al. (2006). The model is able to bind carbaryl in a position that corresponds to known sites of metabolism in human P450s 1A1, 1A2, 2B6, 2C19 and 3A4 (Tang et al., 2002) (Figure 3.19, B). The best ranked mode allows for the production of 5-hydroxycarbaryl. The binding mode of carbaryl is similar to that of xanthotoxin, the aromatic rings stack against F115 while two H-bonds can occur with N208. However, the dockings of DDT are not in positions that allow metabolism (Figure 3.19, C and D).
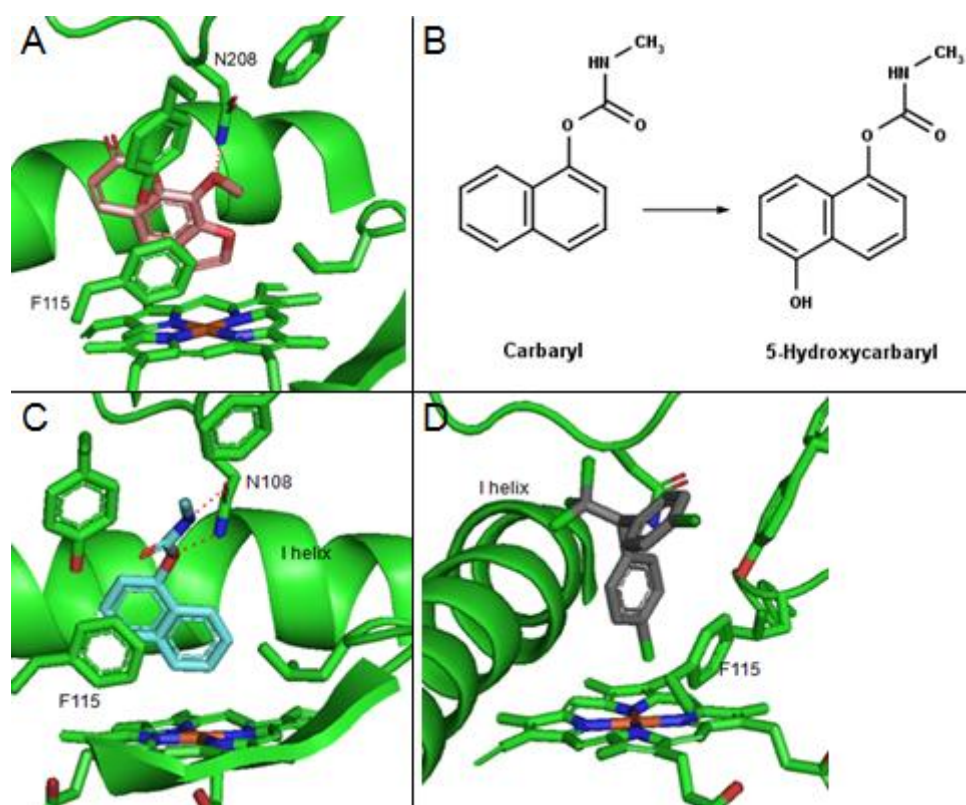
Figure 3.19 (A) CYP6Z2 binding xanthotoxin for metabolism at the 2'3' bond. (B) Metabolites of carbaryl taken from Tang et al. (2002). (C) CYP6Z1 binding carbaryl for metabolism at the 5 position. (D) CYP6Z1 binding DDT in an unproductive mode.

3.1.6.3 CYP6Z2 verification

To evaluate this model, the ligands with known binding data were docked and IC50s compared to ChemScores as suggested by Kemp et al (2004). Some classes of compounds showed a good correlation between ChemScores and IC50 while others showed a poor correlation (Table 3.6 and Figure 3.20).

Table 3.6 Correlations of log10 IC50 vs negative ChemScore.

	R	R^2
Phytoestrogens	0.47	0.2
Insecticides	0.8	0.64
Hormones/steroids	0.59	0.34
Drugs	0.16	0.02
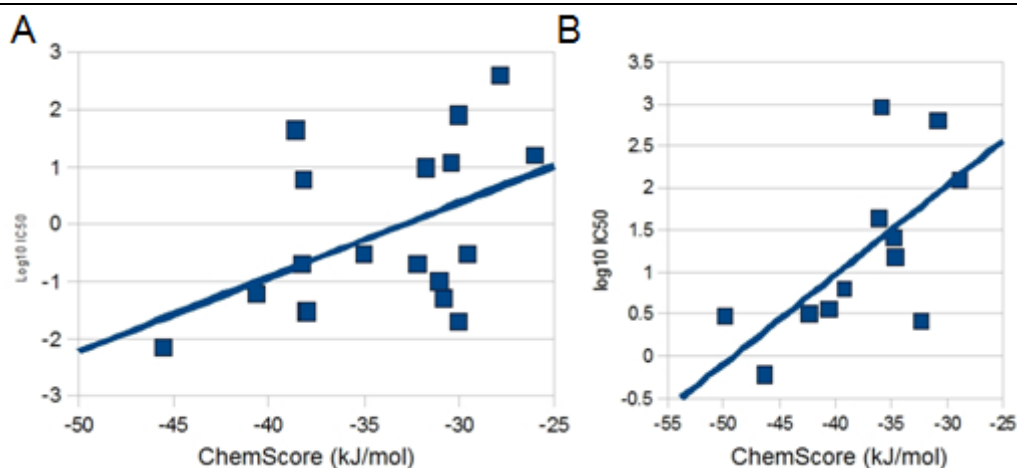
Figure 3.20 The correlation of ChemScore (kJ/mol) and IC50 for ligands docked into CYP6Z2. (A) graph of phytoestrogens, (B) graph of insecticides and hormones.

The mode of binding was also examined to determine if the model could bind compounds in positions that would allow metabolism. CYP6Z2 has been shown to be unable to metabolise carbaryl. However, in the CYP6Z2 model, carbaryl is able to bind in a position that allows metabolism. The 5th ranked mode of xanthotoxin allows metabolism at the 2'3' bond. This is consistent with the suicide inhibition by furan-containing compounds seen in CYP6Z2 (Figure 3.21). Unlike the CYP6Z1 dockings, CYP6Z2 can bind DDT close to the heme. However, this may be too distant for type II coordination that has been suggested to occur for the non-enzymatic dehalogenation, and may be consistent with a lack of metabolism. Isomers of permethrin and cypermethrin were docked. For all isomers there were both productive and unproductive modes. For (S trans) (R) cypermethrn the productive and non-productive modes were similar. The first ranked mode allowed metabolism on the methyls while the second ranked mode did not allow metabolism. Unlike the first, the second ranked mode showed a H-bond as well as π-stacking with F212 (Figure 3.22).
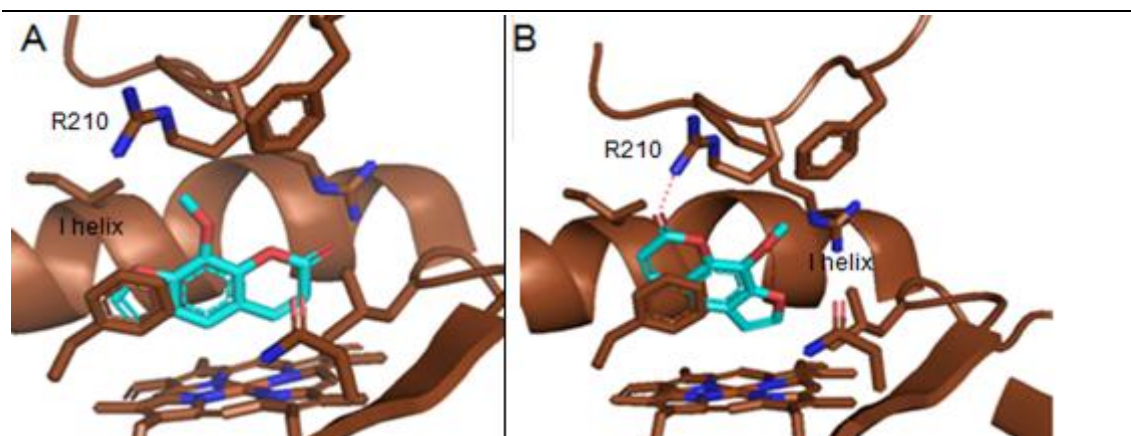
Figure 3.21 (A) First ranked, (B) 5th ranked dockings of xanthotoxin in CYP6Z2 with a H-bond (red dotted line) with R210.
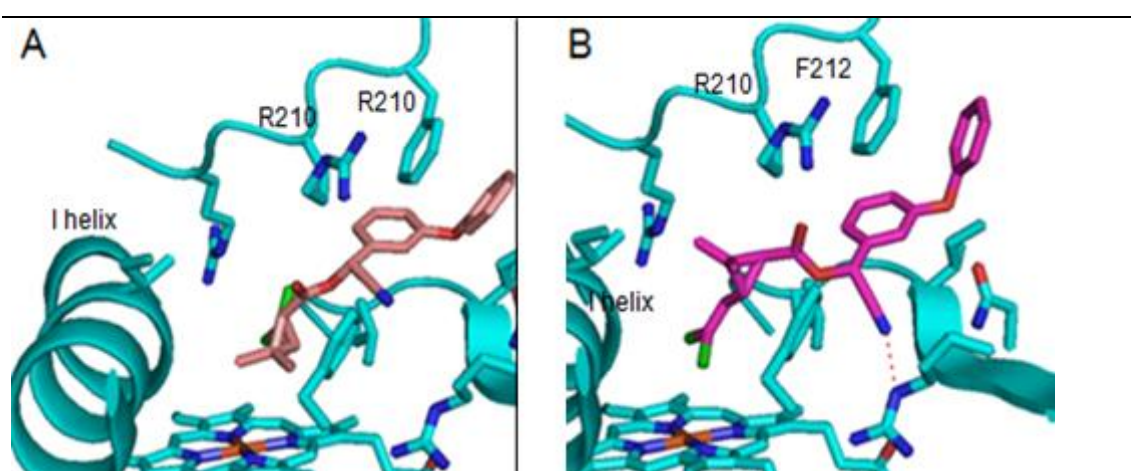
Figure 3.22 (A) first ranked mode of (S trans)(R), (B) second ranked mode.

A similar situation occurred with (Rtrans)(R), although the best ranked mode allowed metabolism but did not show a H-bond, the second ranked mode with a H-bond was an unproductive mode. The first ranked mode of (Rtrans)(S), the (Strans)(S) second ranked and (Scis)(S) 6th ranked modes docked within the access channel indicating that an intermediate binding site may occur. This may be consistant with a high affinity binding but a lack of metabolism.

3.1.6.4 Fluorescent Markers

It has been shown that CYP6Z2 can metabolise benzyl esters of resorufin efficiently, methoxy and ethoxyresorufin less efficiently, and is unable to metabolise pentoxyresorufin.

In the Mclaughlin et al. (2008) model, benzyloxyresorufin binds with the benzyl ring between the BC loop and SRS5. In this model benzyloxyresorufin binds in a different position. Instead of stacking between F115 and L365, it lies between SRS5 and the I

106

helix. This mode may occur where the Arg in the FG loop is rotated out of the active site and not occupying this space. No dockings occurred that allowed methoxyresorufin metabolism and only the 95th ranked docking permitted metabolism of ethoxyresorufin. This may be due to the presence of R210 in the roof of the active site. This residue formed a H-bond with the resorufin in an unproductive mode, alternative rotamers of this residue may prevent this. Resorufins with small chains are able to bind in unproductive modes. They bind between SRS5 and the I helix placing the site of metabolism distant from the heme (Figure 3.23). This mode may be due to a lack of steric restrictions unlike benzyloxyresorufin. With benzyloxyresorufin, the steric bulk of the benzyl ring prevents binding in this unproductive mode. This is consistent with the experimental data showing a higher activity towards benzyloxyresorufin and low activity towards methoxy- and ethoxyresorufin.
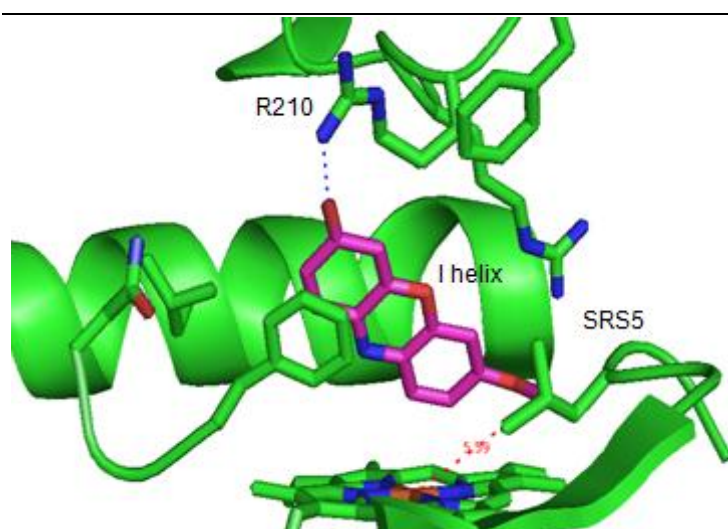
Figure 3.23 1st ranked mode of methoxyresorufin with the site of metabolism 5.59 Å from the heme.

With pentoxyresorufin a productive mode occurred (21st ranked) but had a higher clash score. The higher scoring modes placed the site of metabolism distant from the heme with the body of the resorufin inverted. Modes also occurred where the site of metabolism was sterically screened by atoms of the pentyl chain. This is consistent with pentoxyresorufin as a non substrate. The scores of the dockings indicate improved vdW interactions for benzyloxy resorufin. The lower vdW scores for the smaller resorufins may indicate an increase in mobility that may cause the lower activity (Table 3.7).

Table 3.7 ChemScores (kJ/mol) of productive modes of resorufins in the CYP6Z2 model.

		Score	ΔG	S(hbond)	S(metal)	S(lipo)	H(rot)	ΔE(clash)	ΔE(int)
Eth	95th	34.9	-35.5	0.9	0.0	253.9	1.1	0.4	0.2
Pent	21st	40.5	-41.9	0.9	0.0	313.0	1.3	0.5	0.8
Benz	9th	43.1	-44.8	0.8	0.0	339.7	1.2	1.0	0.6

CYP6Z2 has also been found to metabolise luciferins (Figure 3.24). The luciferins are different from the resorufins. Although benzyloxyresorufin (7-BR) is metabolised efficiently, luciferin-BE is metabolised less efficiently, and while methoxyresorufin is poorly metabolised, methoxyluciferin is efficiently metabolised. The dockings tend to agree with the experimental results in that although no dockings were found that would allow -BE to bind for metabolism, the best ranked dockings place -ME and -CEE for metabolism and no productive modes of -PFBE or -PPXE were found. The modes of -ME and -CEE as well as 7-BR involve an H-bond with N113 while the luciferins form a second H-bond with R210 (Figure 3.25). Although metabolism of -CEE is reduced 6 fold compared to -ME, the score for -CEE is greater than that of -ME due to a greater hydrophobic score (Table 3.8).
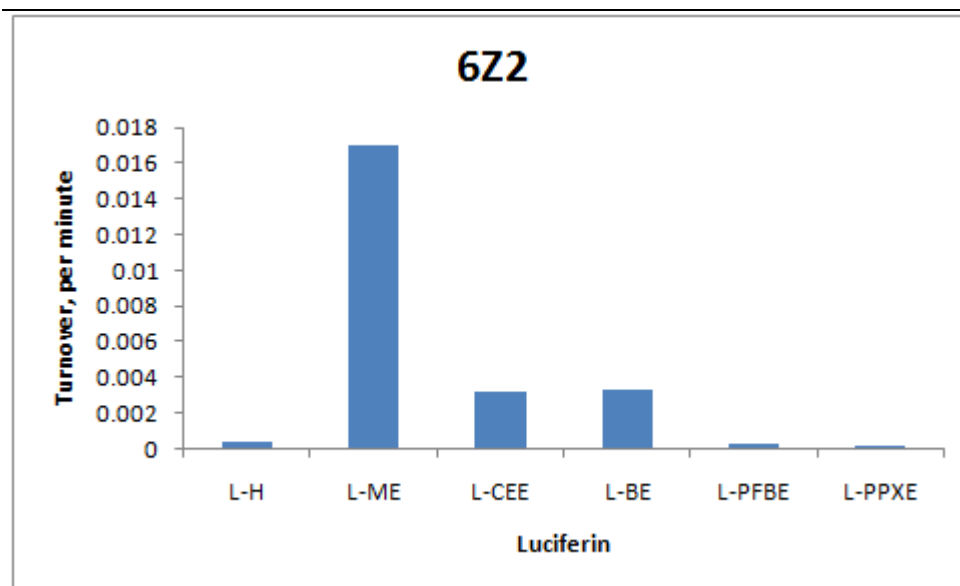
Figure 3.24 Experimental results for CYP6Z2 (McLaughlin, unpublished).
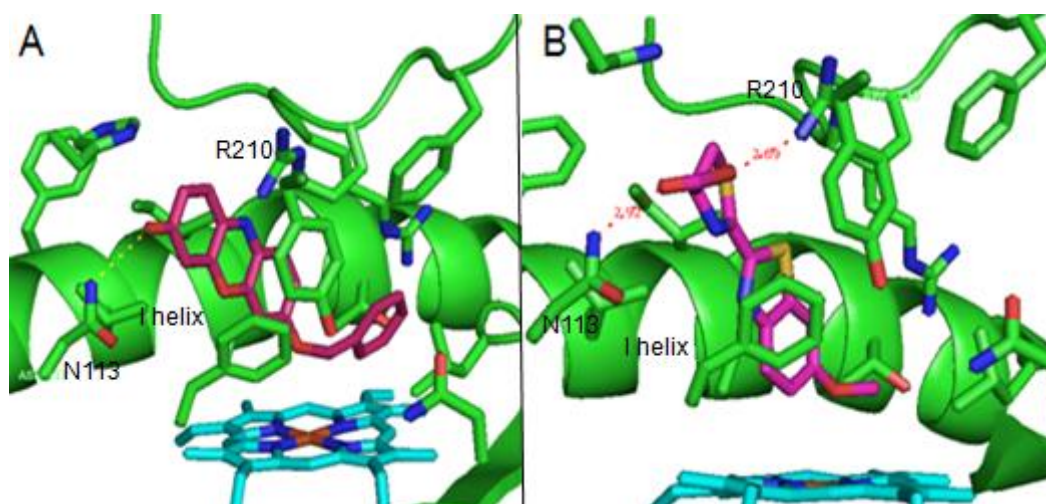
Figure 3.25 Binding modes of 7-BR (A) and -ME (B) in CYP6Z2.

Table 3.8 ChemScores (kJ/mol) of productive modes of luciferins in CYP6Z2.

		Score	ΔG	S(hbond)	S(metal)	S(lipo)	H(rot)	ΔE(clash)	ΔE(int)
ME	1st	38.3	-38.9	1.1	0.0	293.8	1.8	0.2	0.4
CEE	1st	42.5	-43.4	1.2	0.0	337.0	2.1	0.1	0.8
H	9th	35.9	-36.6	1.0	0.0	270.3	1.5	0.1	0.6

3.1.7 *Musca domestica* CYP6D1

In Housefly *Musca domestica*, CYP6D1 has been identified as a metaboliser of pyrethroids and is over expressed in resistant strains. CYP3A4 1TQN was identified as a template by PSI-BLAST and phylogenetic analysis, and aligned with 3DCoffee.

Housefly pyrethroid resistance is correlated with an increase in metabolism of other substrates. To determine this contribution, Wheelock and Scott (1992) treated housefly microsomes with anti 6D1lpr. Anti-lpr antibodies inhibited MROD activity by 83%, AHH activity by 100%, EROD activity by 78%, ECOD activity by 65% while PROD activity was not inhibited. CYP6D1 is the major cytochrome involved in metabolising B(a)P, methoxyresorufin and ethoxyresorufin. Ethoxycoumarin activity was only inhibited at the highest concentration of antiserum and may only be a poor substrate for CYP6D1 while pentoxyresorufin may not be a substrate. The rank order of substrate turnover in CYP6D1 is Benzo(a)pyrene > phenanthrene > methoxyresorufin. B(a)P may similar to the endogenous substrate of CYP6D1 as it is more efficiently metabolised (Korytko et al., 2000).

B(a)P has been identified as the preferred substrate of CYP6D1 but no metabolites have been identified. In dogs the primary metabolite is 3 or 9 phenol, with also 7,8 and 4,5 or 9,10 diol produced. In rats there is a tendancy to produce B(a)P 7,8 dihydrodiol. Dockings of B(a)P allow 2 possible docking positions: one that would allow 7,8 or 9,10 metabolism, and the other metabolism at the 2,3,4 or 5 positions. This is consistant with the known sites of metabolism in mammals. In these positions there may be aromatic interactions with F115 and Y102 (Figure 3.26, A).
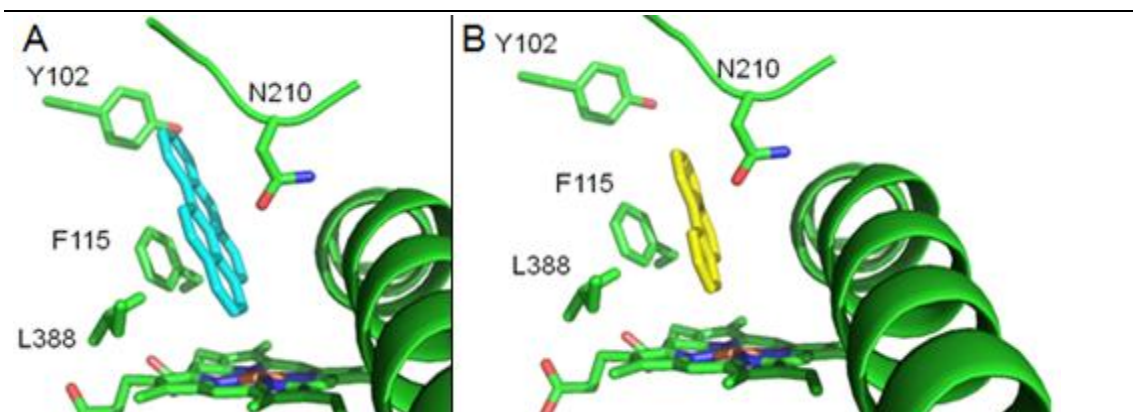
Figure 3.26 Best ranked dockings of B(a)P (A) and phenanthrene (B) in CYP6D1.

CYP6D1 was found to metabolise phenanthrene to produce a single unidentified metabolite. In the same study, dog liver microsomes were found to produce one metabolite of phenanthrene that had the same Rf value as the metabolite produced by CYP6D1, suggesting that these metabolites are identical (Korytko et al., 2000). The metabolite produced by CYP6D1 may be either 3,4; 9,10 or 1,2 dihydrodiol as these are the metabolites produced by other mammals (Nordqvist et al., 1981). In the best ranked dockings, phenathrene can bind for 1,2 metabolism that is consistant with the metabolism in mammals (Figure 3.26, B). The model was shown to bind these substrates with the positions of known sites of metabolism above the heme. The CYP6D1 model can therefore yield results consistent with most of the experimentally determined metabolite production.

3.1.7.2 IC50 ChemScore correlation

For CYP6D1, the IC50 data was produced by Scott et al. (2000). They used housefly microsomes from female houseflies homogenised and centrifuged to collect the supernatant. The supernatant pellet was re-suspended in buffer and used as the enzyme source. MROD activity was used to measure CYP6D1 activity, although in these microsomes CYP6D1 accounts for 83% of MROD activity. Also the IC50 scores were

only determined up to a cut off point of 10^{-4} (M). The log IC50 –ve ChemScore plots for CYP6D1 show a trend. The plots for all inhibitors, including competitive and non-competitive, show poor R^2 values. The plot of the competitive inhibitors alone show values closer to 0.2 (Figure 3.27).
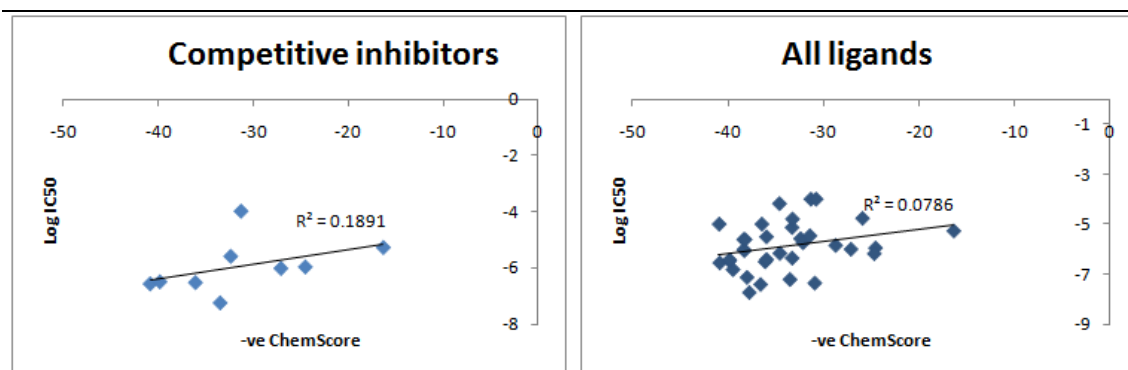
Figure 3.27. The R^2 values for all ligands and for competitive inhibitors docked in CYP6D1.

The poor correlation between ChemScore and experimental IC50 may be due to a number of factors. The model itself may be of poor quality regarding the docking of ligands. The method of scoring may have an affect. The ChemScore method has been found by Kemp et al. (2004) to underpredict the strength of binding of the tightly binding ligands, and they also found initial R^2 scores of 0.28 and 0.11 for their ligand sets, and removal of the tightest binding ligands may improve the scores.

Alternatively, the types of ligands and experimental methods used may affect the IC50 scores. There can be variation between published studies in Km and Vmax values as well as inhibition constants Ki and IC50. Masimirembwa et al. (1999) suggested that these differences could be due to individual variation from using microsomes from different individuals, different expression systems or different experimental procedures or conditions. A large number of factors have been shown to affect enzyme kinetics: type of buffer and ionic strength, amounts of P450 reductase and b5, protein content, and phospholipids, glutathione and detergents. Although there are variations in the values of kinetics, it was found that the ranking of the substrate specificity and inhibitor selectivity were the same regardless of the conditions used. The rank order of ChemScores of the dockings is generally in accordance with the rank order of IC50 for the ligands, with a small number of exceptions, and may indicate that the model is of sufficient quality. The types of ligands used by Scott et al. (2000) were predominantly potential suicide inhibitors rather than competitive inhibitors or substrates. These included aryl acetylene compounds and methylenedioxyphenyl compounds. When these

suicide inhibitors were removed from the data set, the R^2 values increased from 0.08 to 0.2.

3.1.8 CYP6B family

Helicoverpa zea (corn earworm) is one of the most devastating crop pests due to its ability to metabolise both toxic xenobiotics in host plants, and insecticides. The polyphagous *H. zea* has been found to be resistant to a wide range of insecticides such as carbamides, organophosphates and pyrethroids. CYP6B8 and CYP321A1 have been identified as involved in the metabolism of both insecticides and xenobiotics but with differing rates (Table 3.9 and 3.10). While CYP6B1 is able to metabolise xenobiotics it is unable to metabolise pyrethroids.

Table 3.9 CYP6B8 and CYP6B1substrates taken from Li et al. (2003).

Xenobiotic	Vmax µmol/µmol of P450 per min	Km, mM	Clearance ml/µmol of P450 per min
CYP6B8			
Allelochemicals			
Coumarin	NDA		
Flavone	6.95	0.158	44.1
Rutin	0.89	0.02	44.3
Xanthotoxin	3.67	0.066	55.6
Chlorogenic acid	13.3	0.235	55.7
Indole-3-carbinol	1.56	0.013	121.8
Quercetin	5.3	0.02	260.5
Insecticides			
Carbaryl	NDA		
Diflubenzuron	NDA		
α-Cypermethrin	12.72	0.085	149.3
Aldrin	22.42	0.084	267.8
Diazinon	38.92	0.088	442.9
CYP6B1			
Allelochemicals			
Coumarin	NDA		
Flavone	2.35	0.059	39.8
Rutin	NDA		
Xanthotoxin	22.24	0.012	1,859.90
Chlorogenic acid	NDA		
Indole-3-carbinol	NDA		
Quercetin	NDA		
Insecticides			
Carbaryl	NDA		
Diflubenzuron	NDA		
α-Cypermethrin	NDA		
Aldrin	NDA		
Diazinon	50.52	0.197	256.5

NDA, no detectable activity.

Table 3.10 Activity of CYP321A1 and CYP6B8. Taken from Sasabe et al. (2004).

Substrate	Metabolic activity (μmol/μmol P450 per minute)	
	321A1	6B8v1
Xanthotoxin	1.28	1.88
Angelicin	0.88	0.45
α-Naphthoflavone	0.94	0.23
Aflatoxin	2.60	ND
α-Cypermethrin	1.60	2.44
Aldrin	5.28	14.33
Diazinon	1.35	2.33

ND, not detectable.

<u>3.1.8.1 Alignments</u>

Li et al. (2003) found that CYP321A1 is closely related to the CYP6 family as it forms a clade with the CYP6B genes (Figure 3.28). However, CYP321A1 has been suggested to have evolved from a different progenitor gene due to differences. The CYP6B genes contain a single intron positioned two residues downstream of the conserved cysteine that binds the heme, while the CYP321A1 does not contain any introns (Sasabe et al, 2004). Consistent with the findings of Li et al. (2003), in this study, in a ClustalW phylogenetic tree CYP321A1 forms a cluster with the CYP6B family. This group was identified as related to the CYP3 clade with the highest identity to CYP3A4 (Figure 3.29). CYP3A4 was identified as a template and 3DCoffee was used to produce an alignment. Modelling was attempted with either 1TQN and 1W0E as templates. For CYP6B8, the 1W0E template produced the best model.
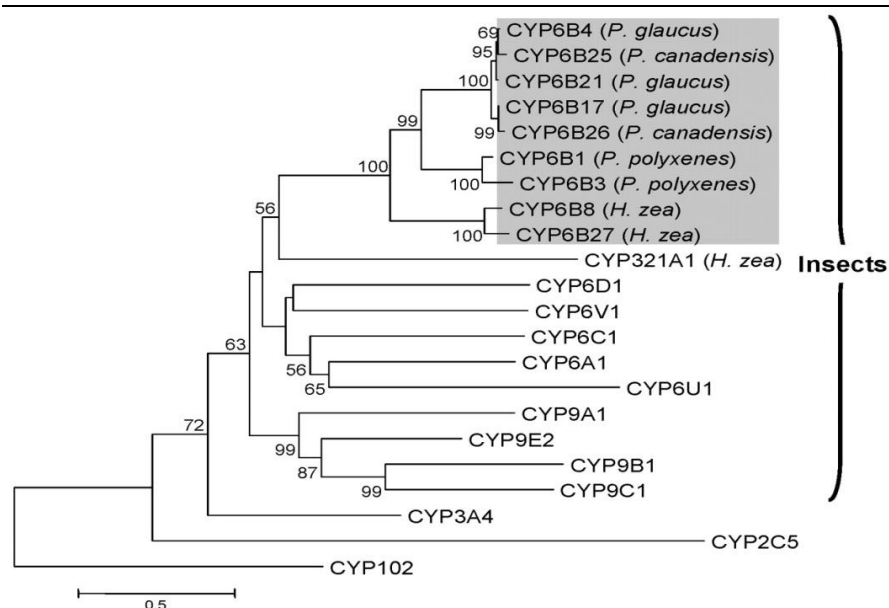
Figure 3.28 CYP321A1 identified as forming a clade with the CYP6B family (grey) Taken from Li et al. (2003), identifying CYP321A1 as closely related.

113
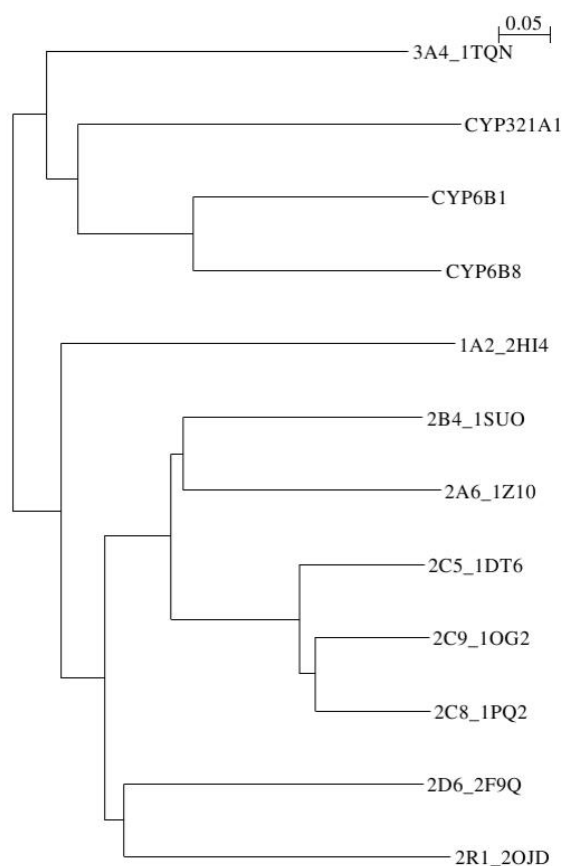
Figure 3.29 NJ tree produced from sequences aligned with clustalW2 (CYP_PDB ID).

3.1.8.2 *Helicoverpa zea* CYP6B8 verification

CYP6B8 is able to metabolise quercetin, and metabolism by P450s at the 6 or 8 position has been previously shown to occur (Halbwirth and Stich, 2006) (Figure 3.30).
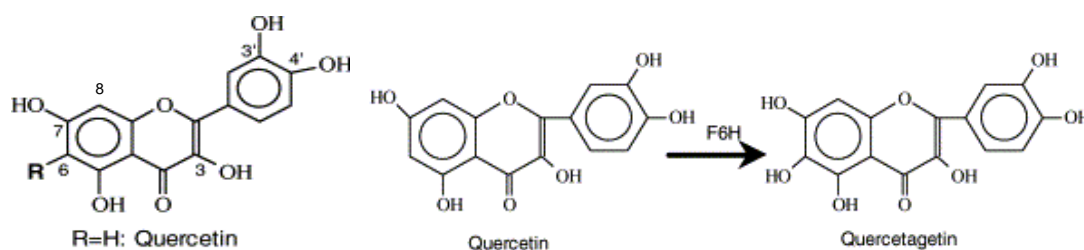
Figure 3.30 Numbering and metabolism of quercetin by P450 F6H. Adapted from Halbwirth and Stich (2004).

The dockings into the CYP6B8 model are consistent with this mode of metabolism as the first ranked docking allows metabolism on the 8 position while the second ranked mode allows metabolism on the 6 position. Quercetin makes a number of H-bonds with the sidechains of E107 and Y214 as well as the main chain of Q375, S210 and A305. The A ring also π-stacks with F118 (Figure 3.31).

114
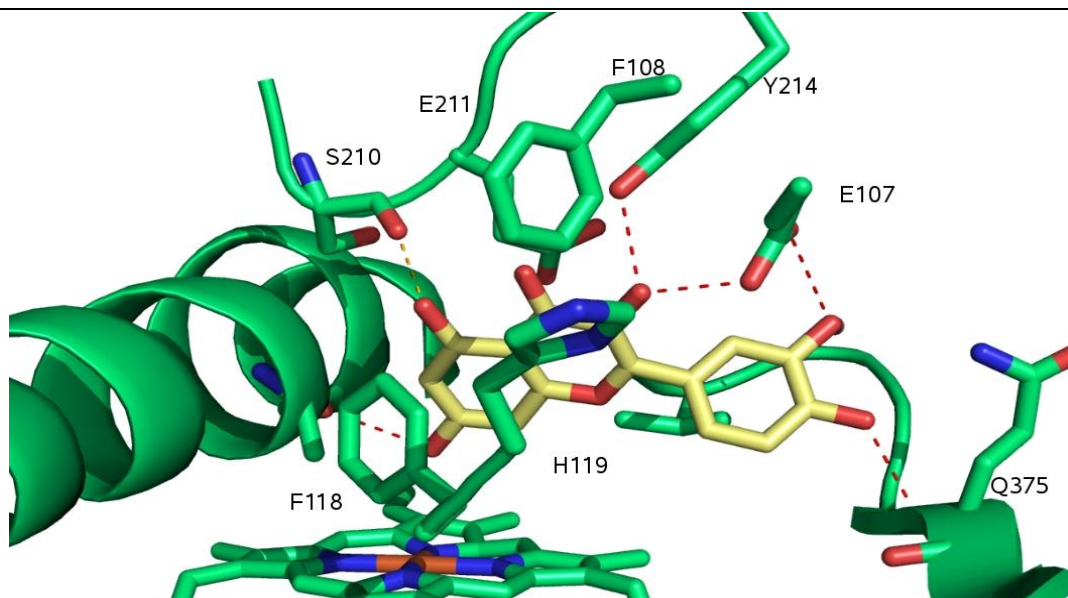
Figure 3.31 The best ranked mode of quercetin in CYP6B8 allows metabolism on the A ring at position 8.
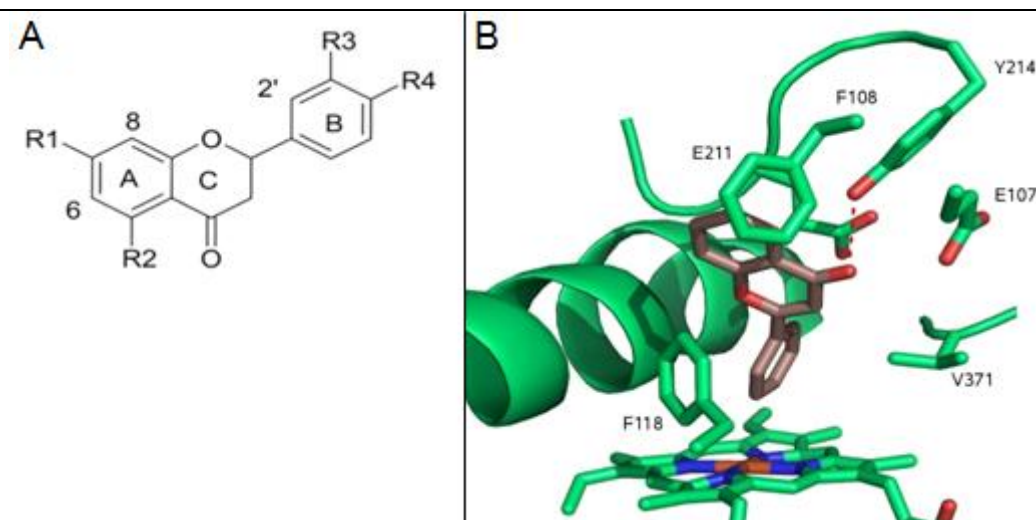
Figure 3.32 (A) Flavone structure and nomenclature (taken from Nikolic et al, 2004). (B) Flavone binding for metabolism on the B ring in CYP6B8.

Flavone can bind for metabolism on the A or B rings. Unlike quercetin, only one H-bond is found with Y214, but π-stacking is found with F118 (Figure 3.32). The CYP6B8 metabolism of flavone and quercetin differ with a higher clearance rate and lower Km for quercetin. This difference may be due to the large number of H-bonds that quercetin is able to form that may stabilise the ligand within the active site although the docking score is lower for quercetin than flavone.

A-Naphthoflavone (ANF) can be metabolised and binds in a similar mode as the flavones but is lacking the H-bond that occurs with Y214 as the H-bond acceptor is orientated away from Y214. The CYP6B8 model binds aldrin in a mode that is

115

consistent with epoxidation (Figure 3.33).
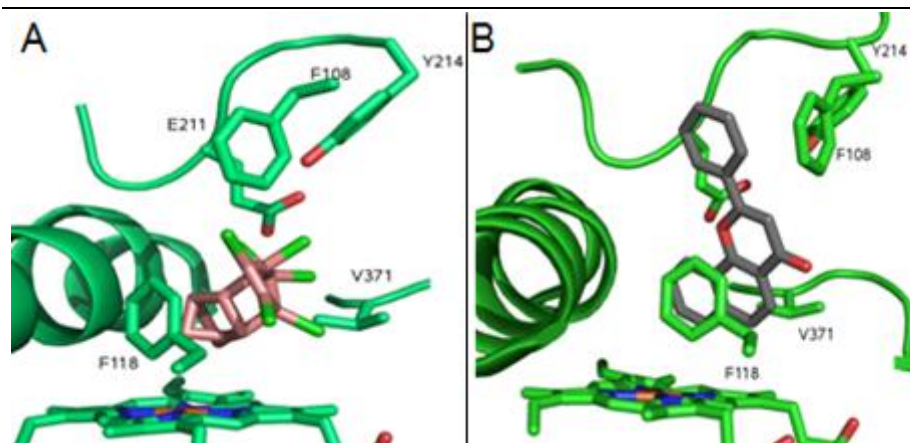
Figure 3.33 Best ranked mode of aldrin (A) and ANF (B) binding in CYP6B8.

CYP6B8 can metabolise indole-3-carbinol, although the metabolites have not been determined, the metabolism of indole is known to occur on the 6 and 3 positions, although the 6-hydroxylation product was only found to be produced at low levels by CYP2A6 (Gillam et al., 2000) (Figure 3.34).
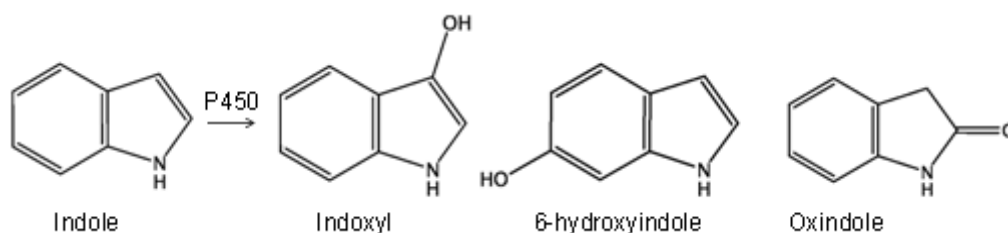
Figure 3.34 Metabolites of indole P450 metabolism. Adapted from Gillam et al. (2000).

CYP6B8 binds indole-3-carbinol in a single mode for metabolism at the 6 position. This is constant with known positions of metabolism (Gillam et al., 2000). In this mode, 3 H-bonds are possible with Y214, E107 and the backbone of S210, π-stacking also occurs with F118 (Figure 3.35).
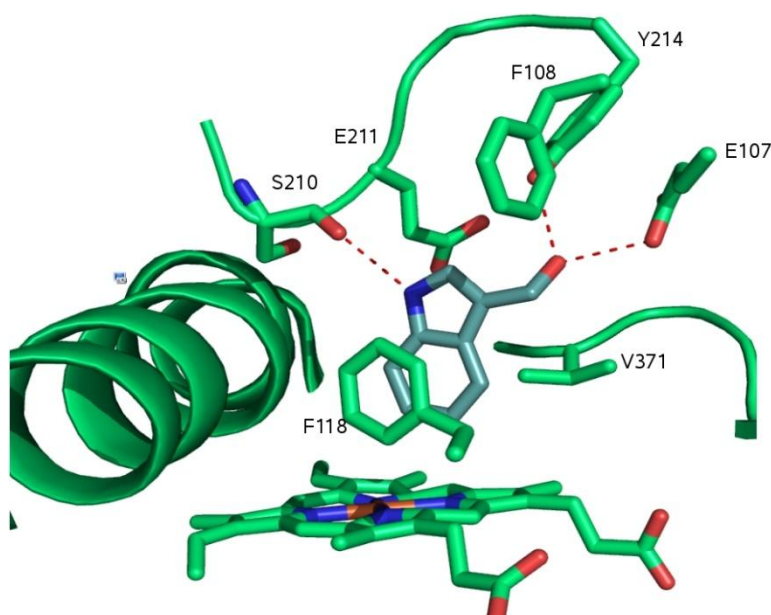
Figure 3.35 The best ranked binding mode of indole-3-carbinol in CYP6B8.

Although indole-3-carbinol is a substrate, coumarin and carbaryl that have a similar structure, are poor or non-substrates. As with indole-3-carbinol, both coumarin and carbaryl bind in modes that allow metabolism with similar modes and interactions, but with fewer H-bonds possible. While 3 H-bonds are possible with indole-3-carbinol, two are possible with carbaryl and one is possible with coumarin. Although all three of these small compounds have similar structures, only indole-3-carbinol acts as a substrate. While the binding mode of indole-3-carbinol is consistent with the experimental findings, the dockings of coumarin and carbaryl are inconsistent. The docking scores are also inconsistent with equivalent scores for coumarin and indole-3-carbinol, and higher scores for carbaryl (Table 3.11).

Table 3.11 ChemScores (kJ/mol) of ligands in CYP6B8.

	Score	ΔG	S(hbond)	S(metal)	S(lipo)	H(rot)	ΔE(clash)	ΔE(int)
Indole	31.5	-31.6	2.7	0.0	176.5	1.4	0.0	0.1
Carbaryl	34.5	-35.3	1.6	0.0	236.4	1.3	0.4	0.4
Coumarin	30.3	-32.2	0.9	0.0	202.9	0.0	1.8	0.0

Xanthotoxin is a substrate and known to be metabolised on the furan ring. However, the top ranked binding modes are inconsistent with this metabolism, although the 15th ranked mode allows metabolism on the furan ring (Figure 3.36). This mode is lower ranking than the non-productive mode because no H-bonds occur, although no H-bonds occur, there is a higher vdW score.

117
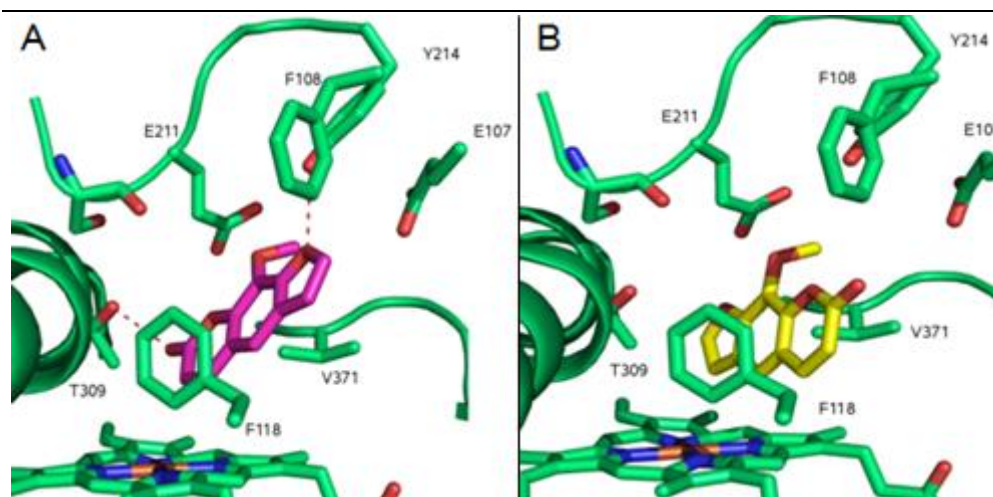
Figure 3.36 In CYP6B8, xanthotoxin binding in a mode inconsistent with metabolism of the furan ring, forming two H-bonds (A). Xanthotoxin binding in a mode consistent with metabolism of the furan ring, no H-bonds are possible (B).

Similar to xanthotoxin, angelecin is also expected to be epoxidised on the furan ring. The highest scoring mode of binding is consistent with this metabolism with the furan ring binding close to the heme. Chlorogenic acid and rutin are large molecules that are known to be oxidatively cleaved by P450s, and CYP6B8 has been found to be able to metabolise both of these. Although the model can bind both of these molecules in positions that allow hydroxylation, they do not bind in positions that would allow oxidative cleavage (Figure 3.37).

Although CYP6B8 has been shown to metabolise α-cypermethrin, no metabolites have been identified. Two modes occur for both isomers of α-cypermethrin: the highest scoring allows metabolism on the dimethyls while the lower scoring allows metabolism at the 4' position. Both isomers bind in similar modes with similar interactions. In the best ranked mode, the phenoxybenzyl group is surrounded by aromatic residues and stacking can occur with F300, F108, H119 and F118. A H-bond is also possible between the cyano group and Y214. The acid group is surrounded by F118 and V371 that provide vdW interactions. In the lower ranked mode, the molecule is inverted and metabolism is possible on the 4' position. In this mode π-stacking is also possible with F108 and F118 but no H-bonds occur (Figure 3.38).
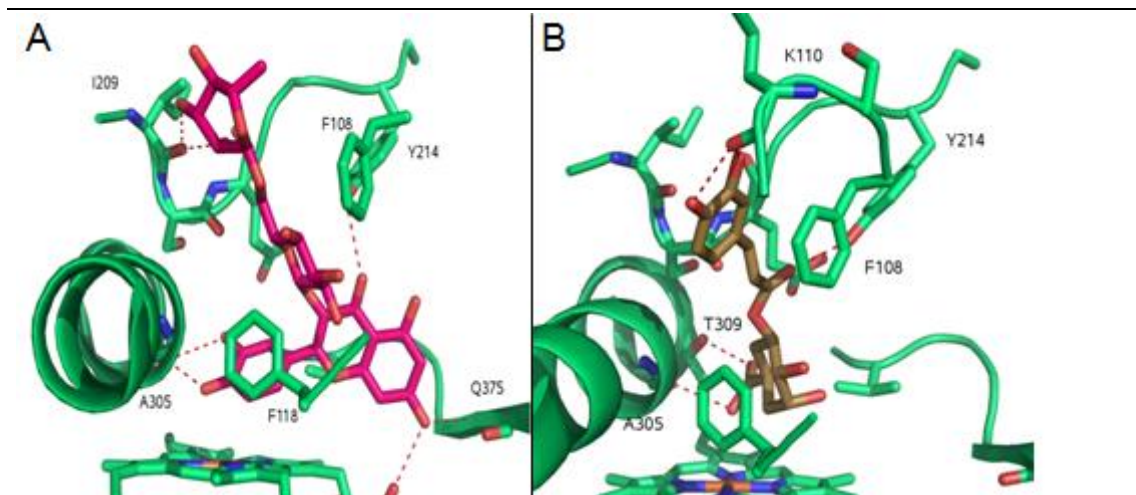
Figure 3.37 (A) The 5th ranked docking of rutin binds in a position that allows hydroxylation. (B) The first ranked mode of chlorogenic acid allows metabolism in CYP6B8.
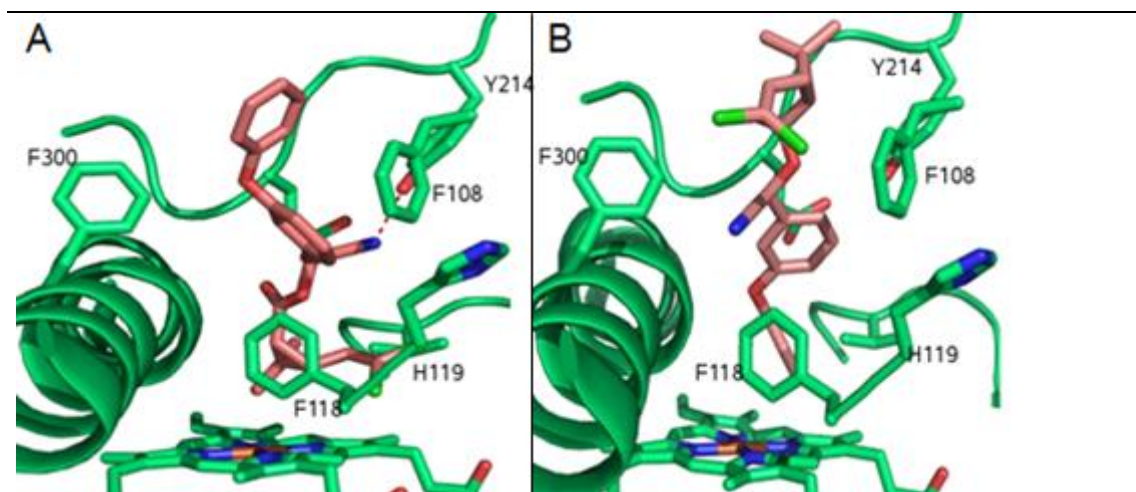
Figure 3.38 (A) Best ranked mode, (B) 45th ranked mode of cypermethrin (Rcis)(S) bound in CYP6B8.

Aflatoxin is a non-substrate of CYP6B8 but is able to bind within the active site in a mode that is consistent with metabolism. However, the binding scores for aflatoxin are poor compared to the scores for the other ligands.

Overall the CYP6B8 model tends to be able to bind substrates in positions that place known sites of metabolism above the heme and indicates a good quality model able to replicate experimental findings.

3.1.8.3 *Helicoverpa zea* CYP321A1 verification

CYP321A1 also occurs in *Helicoverpa zea*, and its subsrates have been identified

119

(Table 3.12). It metabolises similar substrates as CYP6B8 but with a reduced rate.

Table 3.12 CYP321A1 substrates taken from Sasabe et al. (2004).

Compound	nmol metabolised / min / nmol p450
Xanthotoxin	1.46
Angelicin	2.50
α-Naphthoflavone	3.27
α-Cypermethrin	3.22

CYP321A1 differs from CYP6B8 in that it is missing the aromatic residues in the BC loop. These are replaced by polar residues (D107 and T119). It does have an aromatic residue in the G helix (F236) and F299 of the I helix projecting into the site that may contact the ligand. These aromatic residues may have a similar role as F118 of CYP6B8 or F115 of CYP6D1, but may have a non optimal position to form aromatic interactions. The missing aromatic residues and the possible non-optimal location of the aromatic residues that are present may contribute to the reduced metabolism in CYP321A1.

The best ranked modes of both isomers of α-cypermethrin bind in similar positions for metabolism on the gem dimethyl groups. In this mode, the alcohol group is surrounded by aromatic residues of the G'G loop and I helix, and hydrophobic residues of the BC loop. There are a number of aromatic residues near an access channel. ScisR can bind in an alternative mode with the alcohol group binding near this channel surrounded by these residues F212, F371 and H103. These residues may hold the ligand distant from the heme (Figure 3.39). CYP321A1 can metabolise xanthotoxin and the model of CYP321A1 binds xanthotoxin in a mode that is consistent with epoxidation.
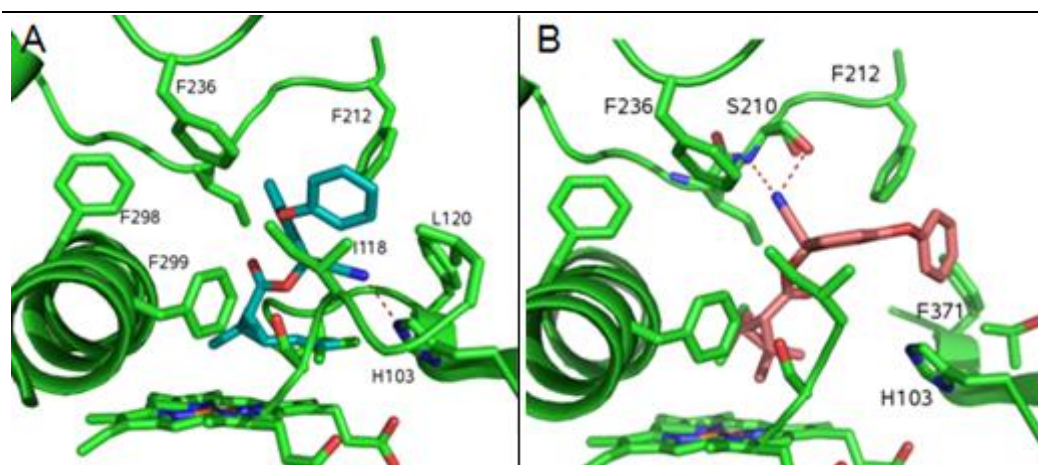
Figure 3.39 Cypermethrin binding modes in CYP321A1.
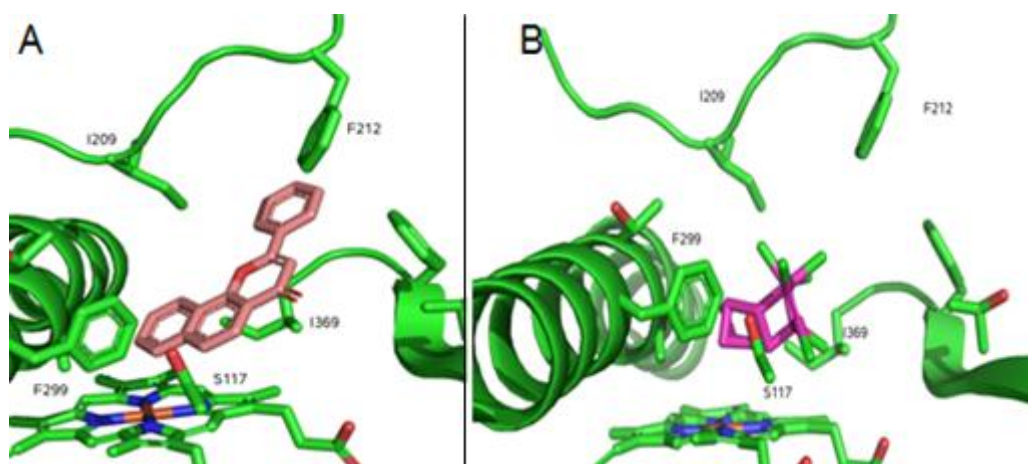
Figure 3.40 The best ranked mode of ANF binding for 7 or 8 metabolism (A), the best ranked mode of aldrin binding for epoxidation (B) in CYP321A1.

The ANF metabolites produced by CYP321A1 have not been identified but metabolism has been predicted to occur on the 7 or 8 carbons (Rupasinghe et al., 2007). The model is consistent with this prediction as it binds ANF for metabolism on the 7 or 8 positions (Figure 3.40, A). CYP321A1 has been found to metabolise aldrin. Although metabolites have not been identified, Rupasinghe et al. (2007) predicted that metabolism in CYP321A1 occurs by epoxidation to produce dieldrin as occurs in CYP6A1. The model is consistent with this prediction as it binds aldrin in a position for this epoxidation (Figure 3.40, B).

Niu et al. (2008) found that CYP321A1 metabolises aflatoxin B1 (AFB1) to two metabolites, with the primary mode of metabolism O-demethylation to AFP1. A second unidentified minor metabolite was also produced with a ratio of 1.8:1. In the model produced here, although the best ranked modes do not allow metabolism, two modes occur that allow metabolism. The best ranking of these modes allows metabolism on the 8-methoxy group consistent with the O-demethylation product as the primary metabolite. A second low ranking mode occurs that would allow epoxidation of the 8-9 bond. The docking results for aflatoxin are consistant with the known experimental data for CYP321A1, where two metabolites are detected with the O-demethylation product being the major metabolite (Table 3.13 and Figure 3.41).

Table 3.13 ChemScores (kJ/mol) of modes that allow metabolism of aflatoxin in CYP321A1.

Mode(rank)	Score	ΔG	S(hbond)	S(metal)	S(lipo)	H(rot)	ΔE(clash)	ΔE(int)
8-methoxy(7)	30.5	-31.1	1.6	0.0	195.7	1.0	0.5	0.1
8-9 bond(24)	29.9	-32.6	1.8	0.0	201.0	1.0	2.3	0.3
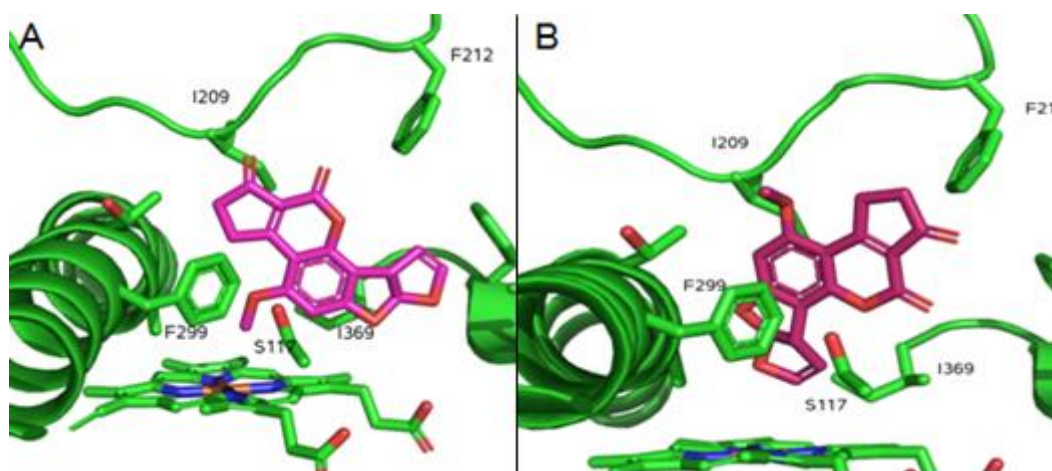
Figure 3.41 In CYP321A1 the 7th ranked mode allows for metabolism on the methyl to give the O-demethylated product (A). The 24th ranked mode allows epoxidation on the 8-9 double bond (B).

Although experimental data shows CYP321A1 incapable of metabolising coumarin, the model is able to bind coumarin for metabolism. However, the docking score is the lowest with a poor vdW score which may be consistent with poor metabolism (Table 3.14). Although the model reproduces the experimental results for the majority of ligands, there are some inconsistencies but indicates a good quality model.

Table 3.14 ChemScores (kJ/mol) of substrates of CYP321A1.

	Score	ΔG	S(hbond)	S(metal)	S(lipo)	H(rot)	ΔE(clash)	ΔE(int)
Angelicin	33.3	-33.5	0.9	0.0	211.8	0.0	0.1	0.0
ANF	45.1	-45.9	0.9	0.0	320.9	0.0	0.8	0.0
Aflatoxin	32.1	-34.5	2.6	0.0	195.0	1.0	2.1	0.3
Xanthotoxin	32.1	-33.1	0.9	0.0	232.4	1.0	1.0	0.0
Coumarin	30.4	-30.4	0.7	0.0	194.7	0.0	0.1	0.0
ScisR	45.7	-51.7	2.8	0.0	351.0	1.6	2.3	3.8
RcisS	47.6	-49.9	1.0	0.0	386.8	1.6	0.3	1.9
Diazinon	31.6	-32.8	1.9	0.0	219.4	1.9	0.4	0.9
Aldrin	40.6	-40.7	0.0	0.0	300.9	0.0	0.0	0.0

3.1.9 *Anopheles minimus* CYP6AA3

The sequence was aligned to CYP3A4 in 3DCoffee. However, the 3DCoffee output places an insert within helix A which produced a poor ERRAT profile. Secondary structure prediction suggested that the loop could be positioned between helix A and the beta strand and this was realigned. In addition to this there was a repositioning of a gap in the BC region.

3.1.9.1 Verification

It has been found by Boonsuepsakul et al. (2008) that CYP6AA3 metabolises deltamethrin to produce phenoxybenzaldehyde by oxidative cleavage of the ester bond. In a free docking deltamethrin binds in a similar position as in CYP6M2 with the methyls close to the heme. A constrained docking of deltamethrin to place the alpha carbon above the heme also placed the cyano group in a position to coordinate with the iron in the top ranked modes. However, in the 73rd ranked mode, the cyano group can H-bond with the Lys of the FG region to position the cyano group away from the heme and prevent coordination to allow ester bond cleavage. Bioallethrin has been found not to be a substrate of CYP6AA3, in the dockings. The best ranked modes bind for metabolism but other low ranked modes position the ligand distant from the heme.

3.2 *Aedes aegypti* CYP9J family

Aedes aegypti is a vector for yellow fever and dengue fever and a number of CYP9s have been tested for metabolism of pyrethroids. CYP9J24, 9J26, 9J28 and 9J32 were found to be metabolisers while CYP9J19 was a non-metaboliser.

3.2.1 Alignments

Both PSI-BLAST and a phylogenetic tree produced by ClustalW indicates that the CYP9s are related to the CYP3s (Figure 3.42) and 1TQN was used as a template. Initial alignments were produced by 3DCoffee and manually realigned. The models of CYPs 9J24, 9J26 and 9J28 resemble the CYP6 metabolisers CYP6M2 and CYP6P3/9 and may indicate necessary conservations in the active site for pyrethroid metabolism, while CYP9J19 has substitutions that may affect metabolism.

3.2.2 Model structure

3.2.2.1 BC loop

Generally in P450s, the BC loop can contain varied secondary structure such as a large helix as in 3DAX or beta sheet as in 2Q9G. In all of the CYP9 family, a large insert occurs relative to CYP3A4. The 3DCoffee output positioned the loop in a position that would disrupt the B' helix and so this was repositioned. Some secondary structure predictors indicated that secondary structure could occur between the C and B helices. This was predicted to be either helix or sheet, but with sheet more commonly predicted. A number of secondary structure predictors were used which varied in their predictions for a given sequence. The predictions also vary among the members of the family. A single beta strand was consistently predicted to occur in CYP9J24, two strands were predicted to occur in CYP9J28 but no strands were predicted to occur in CYP9J26, 9J32

and 9J19. The locations of the predicted strands in CYP9J28 suggest that a sheet could occur similar to that of CYP46A1 (Figure 3.43). The BC loop has been shown to adopt a variety of conformations between P450s and due to the existence of a large insert and predicted sheet, the conformation of the BC loop may be inaccurate.
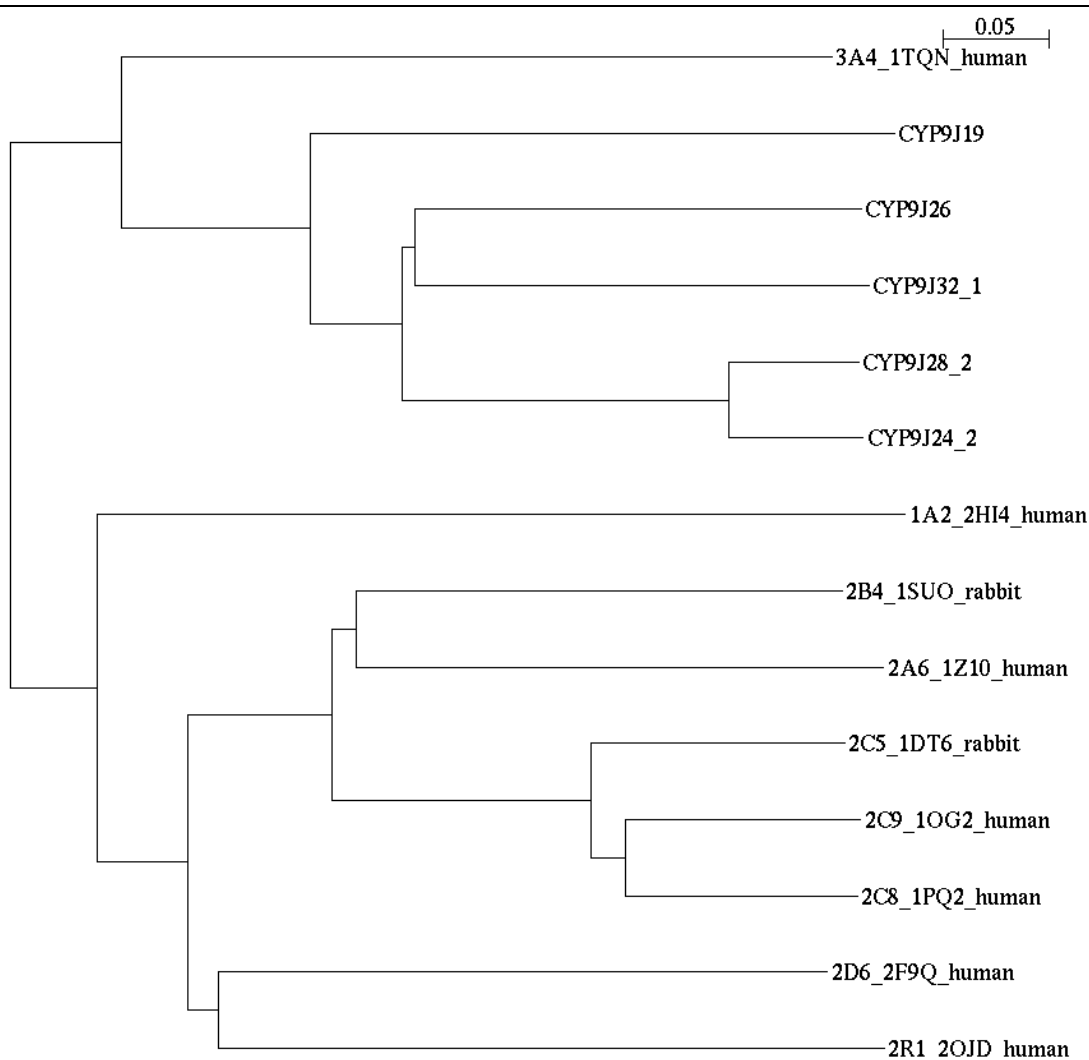
Figure 3.42 Phylogenetic tree produced by ClustalW of the CYP9 family and the hits from PSI-BLAST labelled as CYP_PDB ID _species.
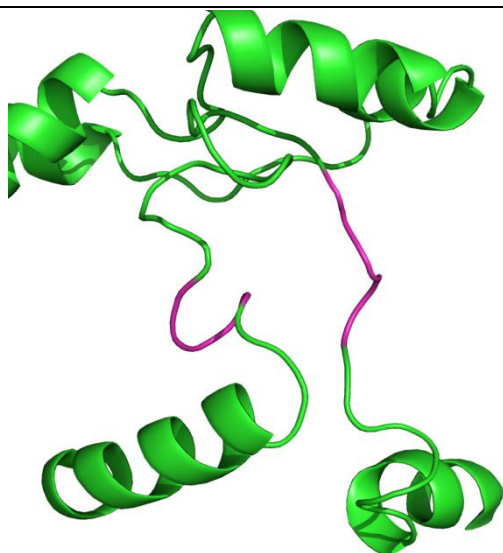
Figure 3.43 CYP9J28 with the positions of the predicted beta strands highlighted in pink.

3.2.2.2 G-I region

Compared to the template, the CYP9J family have a large insert in the region between the G and I helices. In most of the family, this insert was predicted to form a helix. A difference occurred between CYP9J32 and the other CYP9J family members in this region. In CYP9J32, only one helix was predicted to occur, and this could correspond to the H helix that occurs in the template. In the other CYP9J members, two helices were predicted to occur in this region (Figure 3.44). As no additional helix occurs in the template, the remainder of the insert occurred as a random coil and produced a poor Errat profile. The large inserts between helices G and I and within the BC loop could have large affects on the structure of the protein that cannot be represented in the models.

A large insert is predicted to occur between the H and I helices, and secondary structure predictors suggest that this may form an alpha helix. Helix-helix packing has a role in determining tertiary structure, with preferred geometries of interaction to produce an optimal packing. In the template the H-I loop is not clearly identified and this loop is known to show high mobility as it is completely exposed to the solvent. The presence of an additional helix in the 9Js could have large affects on helix packing and rearrangements could alter the tertiary structure or the shape of the active site.
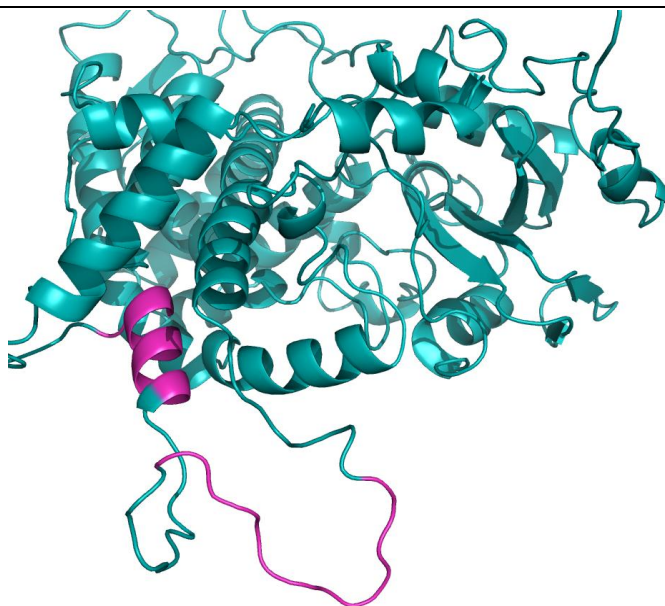
Figure 3.44 CYP9J24 with the predicted locations of helices highlighted in pink.

3.3 Conclusion

In insects there has been an expansion of the CYP6 family, and many of the insect pyrethroid metabolisers tend to occur in this family, while in mammals metabolisers have been identified across families. This study modelled a range of CYP6 and CYP9 P450s that had been tested for activity towards pyrethroids. This identification and modelling of both metabolisers and non-metabolisers is a starting point to identify commonalities between metabolisers and their differences from the non-metabolisers, to understand the requirements for the metabolism of insecticides.

In this study, homology models were produced for a range of identified metabolisers and non-metabolisers. Templates were selected for each P450 and alignments were produced to optimise the strucuture. The models produced in this study were validated in a number of ways. Firstly, the structural evaluation of the models produced, verified that the models were structurally sound and not misfolded. An automated alignment, such as that carried out by 3DCoffee, does not fully take into account secondary strucure as inserts were found to occur in structural elements. Using secondary structure predictions and multiple sequence alignments of family members, the placement of inserts can be manually corrected to improve model quality.

Each model was also extensivly validated to show that the models were consistant with the known experimental observations of each individual enzyme. Experimentally

126

derived substrate binding data and kinetics were compared with the model docking scores to show correlations that give an indication of the model quality. The poses of the ligands docked gives a prediction of the site of metabolism as ligand atoms exposed to, and within range of the heme iron could be hydroxylated. This prediction when compared to the reported sites of metabolism show that a model is consistant with the activity of the enzyme. In this study, the models produced were generally consistant in showing a correlation between known kinetics and docking score, and in binding substrates for metabolism with the correct regiospecifity.

There were some inconsistancies with the data, but a number of factors were found that may explain the inconsistancies. Generally, inconsistencies occured where the experimental studies were carried out on whole microsomes containing many P450 isoforms rather than on a single isolated isoform. Inconsistancies also occured when comparing the IC50s of suicide inhibitors to the docking score. Another inconsistancy occurred where models were able to bind non metabolites for metabolism. However, factors external to the active site may prevent metabolism, such as access from the membrane, and may not indicate a poor model.

4.Pyrethroid metabolism by P450s

4.0 Preface

This chapter identifies P450s involved in pyrethroid metabolism and their metabolites, and attempts to identify commonalities between them. This study uses modelling to identify active site residues, and docking to identify at the atomic level the mode of binding of pyrethroids. Identifying commonalities between metabolisers can be used to produce a 'fingerprint' of binding to identify uncharacterised metabolisers so that candidate P450s can be prioritised for further study.

4.1 Insect P450 CYP6 metabolisers

4.1.1 Housefly CYP6D1

4.1.1.1 Metabolism of natural pyrethrins

The metabolism of naturally occurring pyrethrins in housefly was studied by Yamamoto et al. (1969). They found that housefly homogenates could only metabolise pyrethroids in the presence of NADPH and that stereochemistry did not affect the metabolism pathway. All of the metabolites of allethrin were esters that were more polar than the original allethrin and it was determined that metabolism was carried out by P450s.

They found that the major modifications occurred on the acid, the ester linkage was not affected and the alcohol moiety was not modified. The major metabolic pathway for the metabolism of allethrin was oxidation of the trans-methyl group of the isobutenyl moiety. They also found a minor pathway involving the hydroxylation of the cis-methyl of the isobutenyl moiety (Figure 4.1). The metabolism of pyrethrin, phthalthrin and dimethrin were also presumed to be metabolised in a similar way although their products were not well characterised. It was found that these trans hydroxymethyl compounds have a low toxicity and that the increase in polarity or ionisation may prevent or slow the penetration of the nerve (Yamamoto et al, 1969).

4.1.1.2 Pyrethrin metabolism by CYP6D1

The regiospecificity of metabolism of the naturally occurring pyrethrins and pyrethrin analogues have been shown in housefly. As the primary pyrethroid metabolising P450 in housefly is CYP6D1, it is assumed that this is able to metabolise the pyrethrins as described by Yamamoto et al. (1969).
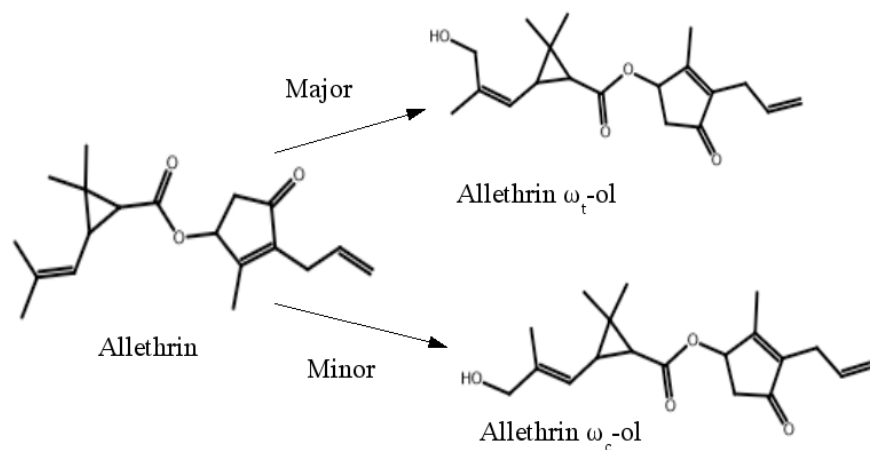
Figure 4.1 Metabolism of allethrin in housefly, adapted from Yamamoto et al. (1969).
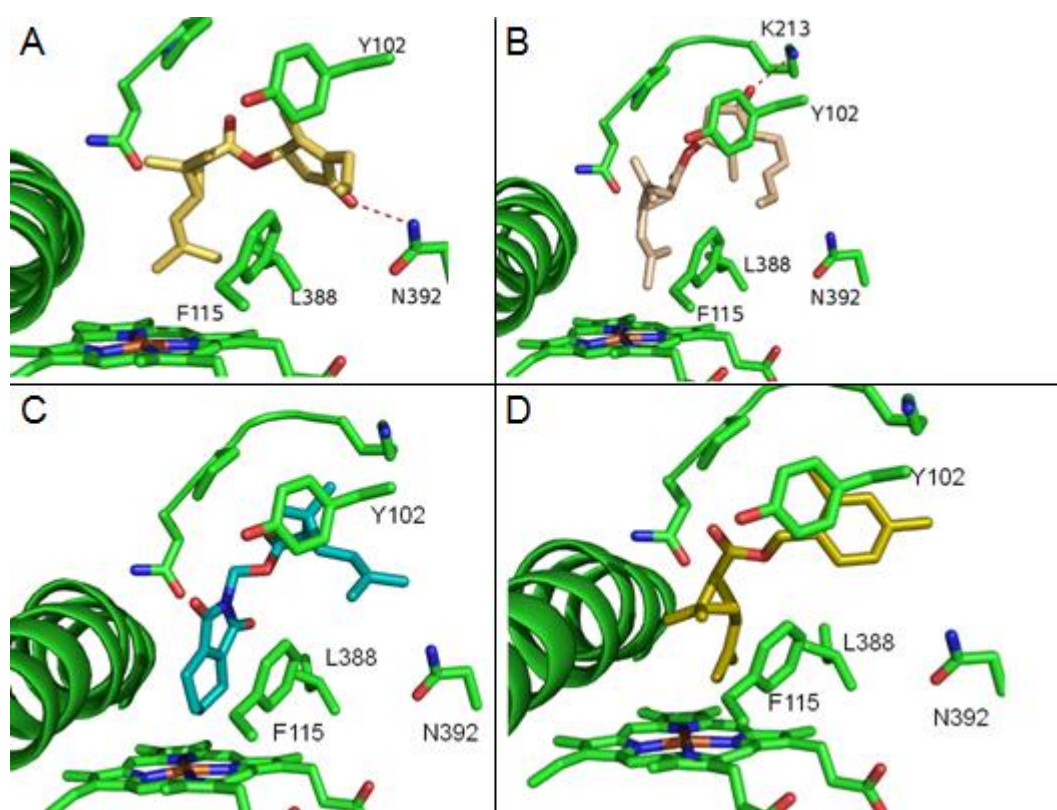
Figure 4.2 The best ranked dockings of allethrin (A) and pyrethrin I (B) both binding in a position for trans methyl metabolism. Best ranked mode of phalthrin (C) and 33rd ranked mode of dimethirn (D) docked into the CYP6D1 model.

The pyrethrins were docked and the regiospecificity of the dockings for allethrin and pyrethrin I agree with the experimental findings as the best ranked dockings allow metabolism of the trans methyl. The model may explain this regiospecificity. The shape of the active site may pre-dispose the ligand to dock in a position for metabolism of the trans methyl. The alcohol moiety of these pyrethroids contains H-bond acceptors and

130

may prefer to bind to the more polar part of the active site near the access channel 2a (Wade et al., 2004) and form H-bonds with N392 or K213 as well as hydrophobic interactions with F115 and Y102. With the alcohol group in this position, the acid group lies within the hydrophobic pocket above the heme. This orientation places the trans methyl closest to the heme (Figure 4.2, A and B).

However, the dockings for dimethrin and phalthrin do not support the suggestion that metabolism occurs on the trans methyl. The highest scoring modes of dimethrin and phalthrin do not place the acid group above the heme, but place the alcohol for metabolism (Figure 4.2, C). However, a lower scoring mode of dimethrin places the acid group above the heme that would allow metabolism of the trans methyl.(Figure 4.2, D).

4.1.1.3 Metabolism of synthetic pyrethroids by CYP6D1

As the isobutenyl group of crysanthamates is photolabile and susceptible to oxidative metabolism, this group is often substituted. The replacement with a dichlorovinyl group often increases potency. The substitution of the methyls in allethrin for halogens to produce dichloroallethrin increases the killing potency in houseflies by 3 fold. Incorporation of an α-cyano group also increases potency, with the α-S configuration having more potency than the α-R isomer (Ruzo and Casida, 1977). Unlike the pyrethrins, in the cis- and trans-dichoro pyrethroids the isobutenyl group methyls are no longer present as they are substituted by chorines, and metabolism of these compounds is shifted either onto the gem dimethyl groups, with a stereospecificity that varies between organisms or onto the 4', 2' or 6 positions on the phenoxybenzyl group (Casida, 1980). Generally the substitution of the methyls of the isobutenyl group with chlorine increases the toxicity of the pyrethroid, and may remove sites of hydroxylation and shift the metabolism to less preferred sites.

The metabolites produced by housefly microsomes have been identified. Shono et al. (1979) found that there was a stereospecificity in the metabolism of the gem dimethyl groups of permethrin in houseflies with the trans methyl of the trans permethrin, and the cis methyl of the cis permethrin the preferred sited of metabolism (Table 4.1). However, Ruzo et al. (1977) found that in the metabolism of an isomer mixture the trans methyl is preferably metabolised.

Table. 4.1 Isomer metabolism in housefly taken from Shono et al. (1979).

Isomer	rank order of metabolism
trans permethrin	trans > 4' > 6
cis permethrin	cis > 4' = trans > 6

There may be more than one P450 involved in pyrethroid resistance in housefly, Wheelock and Scott (1992) found that treatment with anti P450lpr decreased the gem dimethyl metabolism of deltamethrin while increasing metabolism of the alcohol moiety. The metabolites identified from microsomes may be the products of multiple isoforms. The isoform CYP6D1 has been shown to be involved in the metabolism of a number of pyrethroids. CYP6D1 has been shown to metabolise cypermethrin with 4'OH cypermethrin as the major metabolite, as anti-CYP6D1 prevents metabolite production (Zhang and Scott, 1996). CYP6D1 has also been shown to metabolise deltamethrin (Wheelock and Scott, 1992).

4.1.1.4 Dockings of synthetic pyrethroids in CYP6D1

The dockings of the four isomers of permethrin show that each isomer can bind for metabolism on either the 4' or trans methyl. This is inconsistent with the studies on whole microsomes but is consistent with the metabolism of pyrethroids in isolated CYP6D1. The best ranked modes of Rtrans, Scis, and Strans place the 4' in a position for metabolism while Rcis binds for metabolism on the trans methyl while a lower scoring mode places the 4' for metabolism. Similarly, deltamethrin and all isomers of cypermethrin except RtransS bind for metabolism of the 4'. RtransS binds for metabolism of the cis methyl but a lower scoring mode places the 4' for metabolism (Figure 4.3). The 4' binding mode for these synthetic pyrethroids is similar for deltamethrin cypermethrin and permethrin. The phenyl ring π-stacks with F115 while the benzyl ring interacts with both F115 and Y102. These interactions position the phenoxybenzyl group for metabolism at the 4' position. The steric bulk of L388 in SRS5 may restrict the space available above the heme and restrict metabolism on the bulky acid group while allowing metabolism on the phenyl ring. These interactions may contribute to the regiospecificity of oxidation on the 4'.
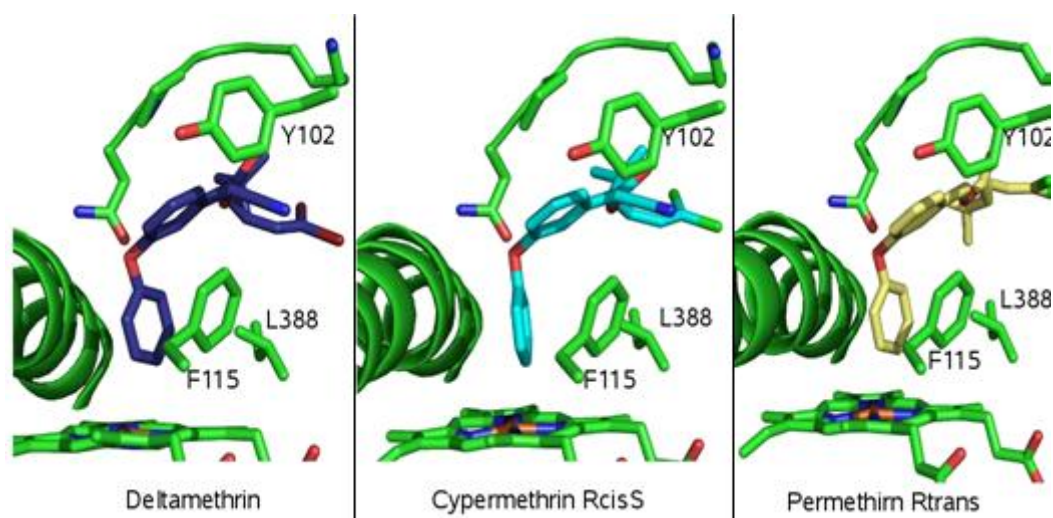
Figure 4.3 The CYP6D1 model binding synthetic pyrethroids.

4.1.2 *Anopheles gambiae* CYP6M2

CYP6M2 has been shown to metabolise deltamethrin into a range of products in a three step process. The native deltamethrin is metabolised at either the 4' or trans methyl, with the 4'OH the major metabolite. The 4'OH product is then oxidised at the 1' to cleave the ether bond to give cyano(3-hydroxyphenyl)methyl deltamethrate. This product is further metabolised at the alpha carbon for ester cleavage to give deltamethric acid (Figure 4.4).
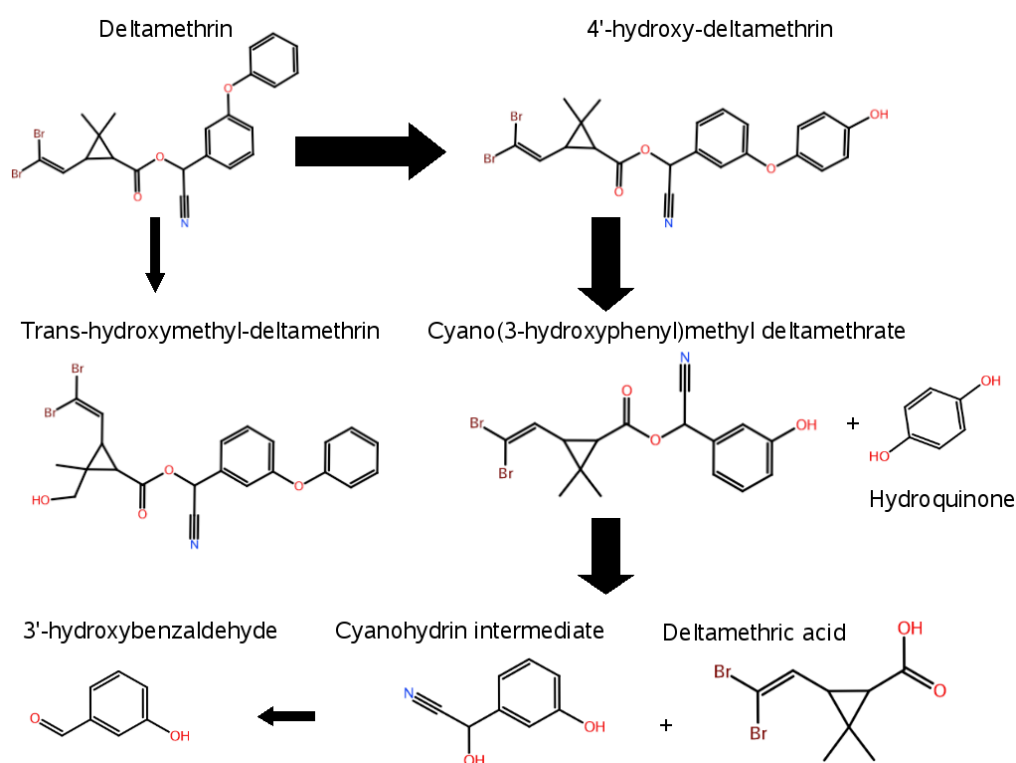
Figure 4.4 The metabolism pathway of deltamethrin in CYP6M2 (McLaughlin, unpublished). Deltamethrin is metabolised in a three step process.

4.1.2.1 Regiospecificity of the native deltamethrin in CYP6M2

When deltamethrin is docked into the CYP6M2 model, two modes occur that score highly, a mode that allows metabolism on the trans methyl and one that allows metabolism on the 4'. The best ranked mode places the trans methyl above the heme while the cis methyl is distant. This is consistent with the known regiospecificity of deltamethrin metabolism on the trans methyl. The second highest scoring mode places the phenyl ring above the heme in a position that would allow 4' metabolism. This is also consistent with the known metabolite produced. However, the scores of the dockings are inconsistent with the preference for metabolism, the higher scoring mode for trans-OH is a minor metabolite while the lower scoring 4' mode is the major metabolite. This may be explained by inadequacies in the scoring method. A constrained docking to place the oxygen bonded carbons of either the phenyl ring or the benzyl ring above the heme produces modes that are higher scoring than the free dockings. There may be inadequacies in the search space of the free dockings or in the scoring.

There may be a number of factors affecting regiospecificity. The similarity in score between the modes (Table 4.2) may indicate that there are two high scoring orientations and that metabolite preference may be determined by reactivity. This was tested using MetaSite. MetaSite indicated that the 4' carbon was the most reactive site of the ligand while the gemdimethyl group was relatively unreactive. Although CYP6M2 may be able to bind the ligand to place both the trans methyl and 4' carbons above the heme, the reactivity of the 4' carbon could lead to a greater amount of product produced. Both the binding modes of deltamethrin and the MetaSite reactivity prediction are consistent with the experimental findings.

Table 4.2 ChemScores (kJ/mol) of the native deltamethrin bound in CYP6M2.

	Score	ΔG	S(hbond)	S(metal)	S(lipo)	H(rot)	ΔE(clash)	ΔE(int)
trans(1st)	41.5	-46.5	0.0	0.0	386.9	1.6	2.7	2.4
4'(2nd)	41.1	-43.3	0.0	0.0	358.7	1.6	0.5	1.7

The dockings offer an explanation for the regiospecificity for the sites of metabolism of the native deltamethrin. The regiospecificity for the trans methyl may be explained partly by the orientation of the acid group in the active site and partly by the relatively exposed nature of the trans methyl. With the acid group docked above the heme and the alcohol group stacking with F110, the halovinyl group lies between F123 and V372. In this position the trans methyl projects towards the heme while the cis is distant. It may

be the steric restrictions produced by F123 and V372 that restrict the space available and constrain metabolism to the trans (Figure 4.5, A).
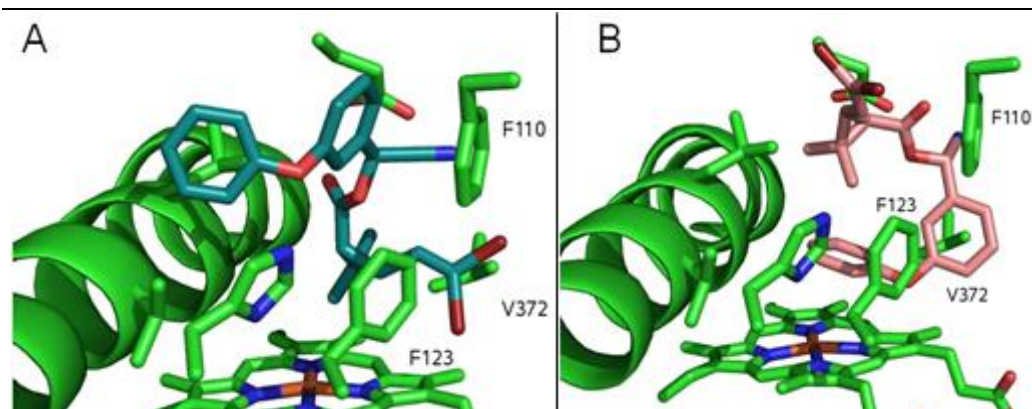
Figure 4.5 Best ranked mode of native deltamethrin in CYP6M2 placing the trans methyl 3.3 Å above the iron, and the cis methyl 5.7 Å (A). The 2nd ranked mode of deltamethrin placing the phenyl ring above the heme (B).

In the second high scoring mode, the phenyl ring can bind above the heme, although the regiospecificity for 4' metabolism is less well explained as the 4' carbon as well as the 2' and 6' are close to the heme, although the 4' carbon is closer at 4.25 Å while the ortho carbons are 4.44 Å and 4.32 Å (Figure 4.5, B).

4.1.2.2 Successive reactions and unobserved metabolites

It has been shown with a number of P450s that a single substrate can be repeatedly metabolised to a number of products. $P450_{17\alpha,lyase}$ catalyses the 17α-hydroxylation of progesterone and pregnenolone, and the C17-C20 bond cleavage of the 17α-hydroxylated steroids to form androgens such as androstenedione. Tagashira et al. (1995) showed that the 17α-hydroxy intermediate was successively oxidised without being displaced from the active site. When pregnenolone was metabolised in the presence of radio-labelled 17α-OH intermediate, the androgens produced retained the same isotope as the original pregnenolone, showing that the released 17α-OH cannot be the intermediate for further metabolism. If the intermediates did leave, the active site would be occupied by the substrate that is present in large amounts. The androgens are produced from a non-dissociating intermediate. In this successive reaction, a portion of the 17α-OH progesterone is displaced while the remainder is metabolised by a second monooxygenase reaction. The rate of the first reaction is higher than the second. In the second reaction, it is necessary for the intermediate to change its orientation in the active site and this was suggested to cause the slower rate in the successive reaction.

After the initial metabolism of the native deltamethrin, the 4'OH is then further metabolised to 3-hydroxyphenyl while the trans-OH is not. Free dockings of the metabolites show that there is a large difference in score between the metabolites that are further metabolised and those that are not. Trans-OH binds with a low score while 4'OH binds with a high score. As 4'OH deltamethrin is able to bind with a high score it may be retained in the active site to be further metabolised, while the poorer binding metabolites may be displaced (Table 4.3). However, the modes of the free dockings of the metabolites do not reflect the sites of metabolism for 4'OH or 3-hydroxyphenyl (Figure 4.6). It is possible that while the metabolite is retained, it tumbles within the site to reposition for further metabolism. In this case little movement of the ligand is required as the second site of metabolism is close to the first.

Table 4.3 ChemScores (kJ/mol) of the metabolites of deltamethrin bound in CYP6M2.

	Score	ΔG	S(hbond)	S(metal)	S(lipo)	H(rot)	ΔE(clash)	ΔE(int)
4'OH	40.3	-43.7	0.9	0.0	354.2	2.4	1.5	1.9
4=O	42.9	-45.2	1.0	0.0	353.8	1.9	1.0	1.4
transOH	39.9	-43.3	0.5	0.0	366.6	2.6	1.1	2.3
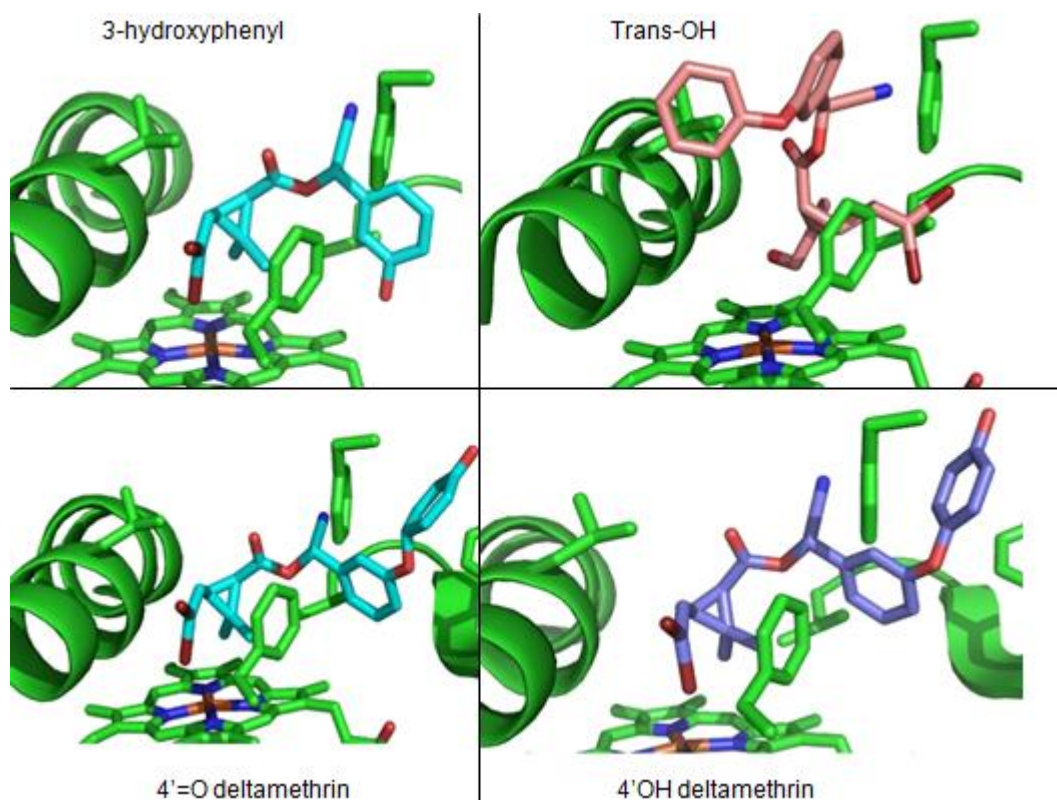
Figure 4.6 The docking modes of the free dockings of the deltamethrin metabolites in CYP6M2.

4.1.2.3 Alcohol metabolism and ether bond cleavage

From the metabolites identified, cleavage of the ether bond of the phenoxybenzyl group

136

may occur. P450s have been shown to be able to cleave the diarylether bond of *p*-(*p*-nitrophenoxy)phenol to produce *p*-nitrophenol and hydroquinone. Ohe et al. (1994) using ^{18}O showed that this was accompanied by replacement of *p*-nitrophenoxy group by the oxygen atom in an ipso-substitution. In this case the hydroxy group was found to be necessary as *p*-(*p*-nitrophenoxy)phenol analogues that lack the hydroxy group were not cleaved. Ohe et al. (1995) suggested a mechanism for the cleavage that involves abstraction of one electron and one proton or one hydrogen radical to give a phenoxy radical. This radical de-localises on the aromatic ring and redistributes to the ipso position. Hydroxylation at the ipso position gives a hemiketal which breaks down (Figure 4.7).
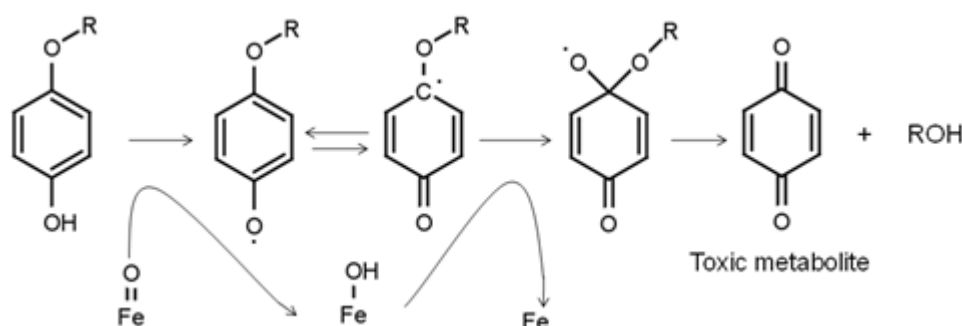
Figure 4.7 Ipso substitution reaction adapted from Ohe et al. (1994).

This reaction can only occur in arylethers after aromatic hydroxylation. For ether bond cleavage to occur with pyrethroid phenoxybenzyl groups, prior hydroxylation at the 4' may be required. As ether bond cleavage by P450s has been shown to occur by ipso substitution of an hydroxylated phenyl ring, ipso substitution of the 4' hydroxylated deltamethrin could lead to ether bond cleavage. The 4'OH deltamethrin needs to be oxidised on the 1' carbon for ipso substitution. Free dockings support this suggestion as this mode occurred but was low scoring. Although the scores of the free docking do not explain the preference for metabolism, this may be due to inadequacies of the search space as a constrained docking to place the oxygen bonded carbon of the phenyl ring above the heme produces modes that are higher scoring than the free dockings (Table 4.4).

Table 4.4 ChemScores (kJ/mol) of the best ranked modes of a constrained docking in CYP6M2. The dockings involved constraining the oxygen bonded phenyl carbon close to the heme.

	Score	ΔG	S(hbond)	S(metal)	S(lipo)	H(rot)	ΔE(clash)	ΔE(int)	ΔE(con)
Native	42.3	-43.9	0.0	0.0	364.3	1.6	0.2	1.5	0.0
4'OH	41.9	-45.7	0.9	0.0	370.7	2.4	0.8	2.9	0.0

When the oxygen bonded carbon of the phenyl ring is constrained close to the heme, the phenyl ring lies close to the heme in a similar mode as in the free docking, but with an improved score. This may indicate a deficiency of sampling in the free docking that this higher scoring mode was not found. After 4' hydroxylation the product may be retained in this position for hydrogen abstraction. MetaSite indicated that after hydrogen abstraction, the 1' carbon is activated for metabolism so that ipso substitution can occur.

4.1.2.4 Ester bond cleavage

In the free dockings of native deltamethrin no modes occur that allow metabolism of the alpha carbon. In constrained dockings to place the alpha carbon above the heme, the cyano group coordinates with the iron in a mode that would produce type II inhibition. This may be consistent with the finding that native deltamethrin is not cleaved at the ester bond but must be modified before this occurs. Deltamethrin is converted to 3-hydroxyphenyl before the ester bond is cleaved, the removal of the phenyl ring may alter the binding mode to allow metabolism on the alpha carbon.

To prevent finding coordination with the cyano, two restraints were used one to constrain the alpha carbon above the heme with a restraint of 1.5-4.5 Å and a second to prevent the cyano approaching the heme.

Table 4.5 ChemScores (kJ/mol) for dual restrained dockings.

	Score	ΔG	S(hbond)	S(metal)	S(lipo)	H(rot)	ΔE(clash)	ΔE(int)	ΔE(con)
Native	33.1	-37.2	0.0	0.0	306.9	1.6	0.5	2.6	0.9
4'OH	31.4	-47.7	1.0	0.0	386.3	2.4	10.5	5.7	0.0
transOH	33.6	-38.0	1.0	0.0	307.6	2.6	1.0	2.9	0.5
3-hyd	28.1	-33.1	0.7	0.0	263.1	2.2	1.7	1.3	1.9

Although 3-hydroxyphenyl deltamethrin is further metabolised at the alpha carbon, no free modes find this pose and the restrained docking produces poor scores (Table 4.5). However, the absence of the phenyl ring produces a lower internal torsional strain than when the phenoxybenzyl ring is intact and a lower clash score than the 4' OH. To allow

the 3-hydroxyphenyl alpha carbon to approach the heme for ester cleavage, the phenyl ring stacks between F123 and the I-helix (Figure 4.8). As there is limited space available in the area between the BC loop and I-helix, the absence of this ring in 3-hydroxyphenyl may be preferred. MetaSite also indicated the alpha carbon as a reactive site for metabolism.
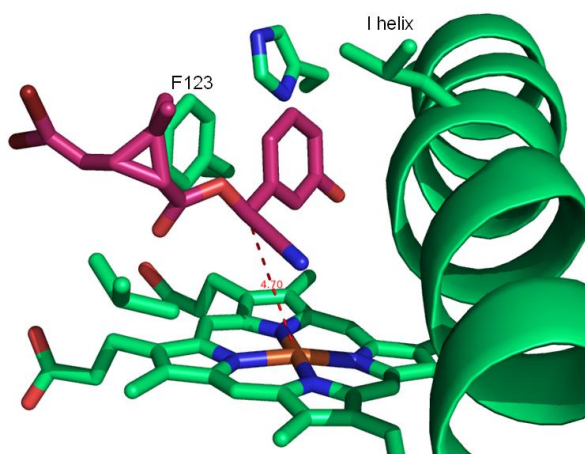
Figure 4.8 The docking modes of a dual restrained docking of 3-hydroxyphenyl in CYP6M2.

4.1.3 *Anopheles gambiae* CYP6P3

CYP6P3 and CYP6M2 produce similar metabolite peaks of deltamethrin and can bind in similar modes. The best ranked mode places the 4' in a position for metabolism while a lower ranking mode places the trans methyl for metabolism (Figure 4.9). In the mode that allows 4' metabolism, the benzyl ring stacks with F123 while the phenyl ring stacks above the heme. F110 and L216 form hydrophobic contacts with the acid group. In the mode that allows metabolism of the trans methyl, the benzyl ring stacks with F110 while the acid forms hydrophobic contacts with L380 and F123.
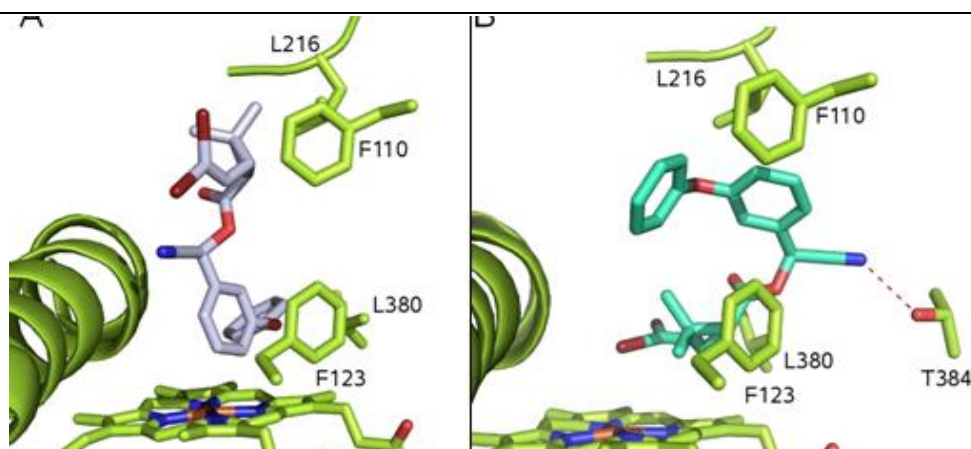
Figure 4.9 Deltamethrin binding in CYP6P3 the best ranked mode (A) and second ranked mode (B).

4.1.4 *Anopheles funestus* CYP6P9

Like CYP6M2 and CYP6P3, CYP6P9 is also predicted by the dockings to bind deltamethrin for metabolism on either the 4' or trans methyl. In the best ranked mode, deltamethrin binds for metabolism on the trans methyl. Like CYP6P3, this mode involves π-stacking of the benzyl ring with F110 and hydrophobic contacts between the acid group and F123 and V380 (Figure 4.10, A). A mode occurs that also allows metabolism at the 4' position. Although CYP6P9 is the orthologue of CYP6P3 it has substitutions within the binding site and this binding mode differs. In this mode there is a H-bond with R54 that positions the acid group distant from the heme. The benzyl ring binds between V380 and F123 and π-stacking may occur with F123. The phenyl ring may also stack with F123 (Figure 4.10, B).

4.1.5 *Anopheles minimus* CYP6AA3

CYP6AA3 has a similar active site to CYP6M2 and CYP6P3, CYP6AA3 has similar aromatic residues in the BC loop (H120, F112, Y109) and conserved aliphatic residue in SRS5 (V376), but shows a different regiospecifitiy for deltamethrin. It has been found that CYP6AA3 metabolises deltamethrin to produce phenoxybenzaldehyde by oxidative cleavage of the ester bond. In a free docking, deltamethrin binds in a similar position as in CYP6M2 with the methyls close to the heme (Figure 4.10, C). A constrained docking of deltamethrin to place the alpha carbon above the heme also placed the cyano group in a position to coordinate with the iron in the top ranked modes. However, in the 73rd ranked mode, the cyano group can H-bond with the Lys of the FG region to position the cyano group away from the heme and prevent coordination (Figure 4.10, D). It may be this substitution that shifts the preference for metabolism onto the alpha carbon for the native deltamethrin.
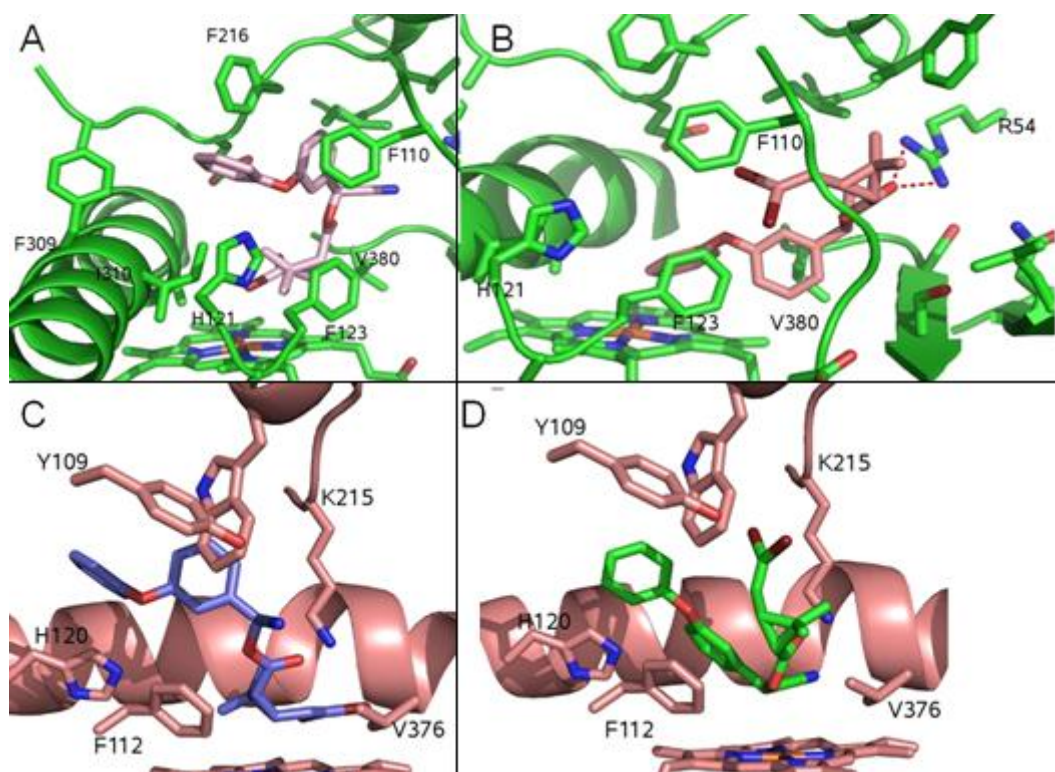
Figure 4.10 The best ranked binding mode of deltamethrin in CYP6P9 places the ligand in position for metabolism on the trans methyl (A) while the 6th ranked mode places it for metabolism on the 4' (B). The best ranked mode (C) and the 73rd ranked mode (D) of deltamethrin binding in CYP6AA3.

4.1.6 *Helicoverpa zea* CYP6B8

Although CYP6B8 has been shown to metabolise alpha-cypermethrin, no metabolites have been identified. Two modes occur for both isomers of alpha-cypermethrin: the highest scoring allows metabolism on the dimethyls while the lower scoring allows metabolism at the 4' position. Both isomers bind in similar modes with similar interactions. In the best ranked mode, the phenoxybenzyl group is surrounded by aromatic residues and stacking can occur with F300, F108, H119 and F118. A H-bond is also possible between the cyano group and Y214. The acid group is surrounded by F118 and V371 that provide vdW interactions (Figure 4.11, A). In the lower ranked mode, the molecule is inverted and metabolism is possible on the 4' position. In this mode π-stacking is also possible with F108 and F118 but no H-bonds occur (Figure 4.11, B).
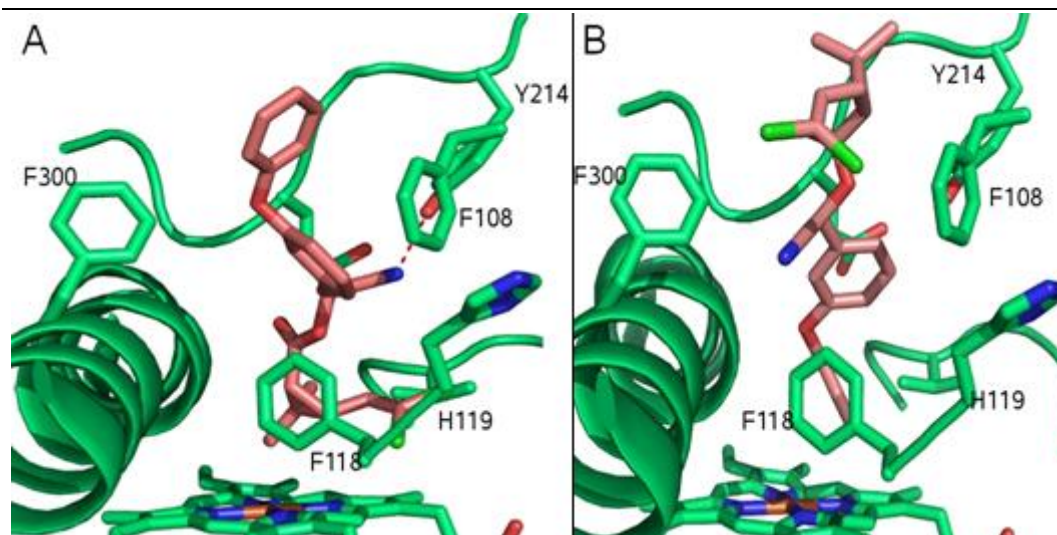
Figure 4.11 The best ranked mode (A) and 45th ranked mode (B) of cypermethrin (Rcis)(S) binding in CYP6B8.

4.1.7 Regiospecificity comparisons

The identified metabolisers metabolise deltamethrin with different regiospecificities, CYP6D1 metabolised deltamethrin on the 4' position while CYP6M2 and CYP6P3 metabolise both the alcohol and acid groups. There are a number of conservations and substitutions between these P450s that could indicate the reason for the difference in regiospecificity.

4.1.7.1 BC loop

The CYP6 metabolisers have two conserved aromatic residues in the BC loop. In these P450s, the BC loop aromatic residues are involved in binding the ligand in different binding modes. In CYP6D1 both BC loop aromatic residues are involved in stacking with the aromatic rings of deltamethrin while in CYP6M2/6P3 they may be involved in hydrophobic interactions as well as π-stacking. CYP6M2 and CYP6P3 have an additional aromatic residue H121 in the BC loop that is in a position to interact with the ligand, this is absent in CYP6D1 and may contribute to differences in regiospecificity. This additional His may provide aromatic interactions, or may form a H-bond with the phenoxybenzyl oxygen, but as the His replaces an Asn in CYP6D1, H-bonding may not the primary factor (Figure 4.12).
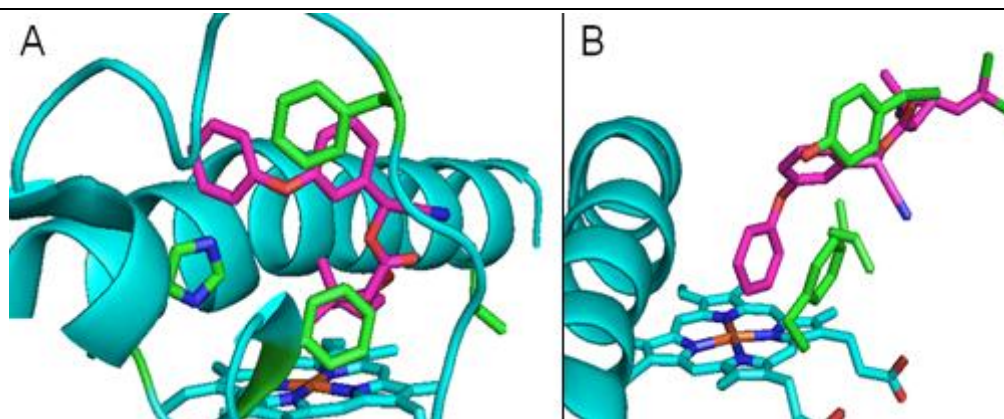
Figure 4.12 CYP6P3 binding deltamethrin for metabolism on the methyls (A) and CYP6D1 binding deltamethrin for metabolism at the 4' position (B). In CYP6P3 the additional His is in a position to provide additional aromatic interactions where deltamethrin is in a position for metabolism on the acid group.

4.1.7.2 FG loop

CYP6D1 has a H-bond donor in the FG loop that is in a position to form H-bonds with the phenoxybenzyl oxygen while deltamethrin is in a position for 4' metabolism. In both CYP6M2 and CYP6P3 this is substituted and a H-bond is not possible.

4.1.7.3 SRS5

There are differences between CYP6D1 and CYP6M2/6P3 in SRS5. In CYP6D1 this is a Leu while in CYP6M2/CYP6P9 this is a smaller Val. In CYP6D1, the size of the active site is sterically restricted as both F115 and L388 project into the site. To access the heme a ligand must bind between F115/L388 and the I helix. As the acid group is relatively bulky, metabolism on the methyls may be sterically restricted while the planar phenyl ring is able to bind. These steric restrictions may also be responsible for CYP6D1s preference for planar ligands. The substitution of the large L388 in CYP6D1 for the smaller V372 in CYP6M2/6P3 may provide space for the acid group to bind. CYP6D1 also has a H-bond donor (N392) in SRS5. This can form a H-bond with the cyano group while deltamethrin is in a position for 4' metabolism. N392 in SRS5 is substituted in CYP6M2 for F376 and it is possible that the loss of this H-bond could contribute to differences in regiospecificity (Figure 4.13).
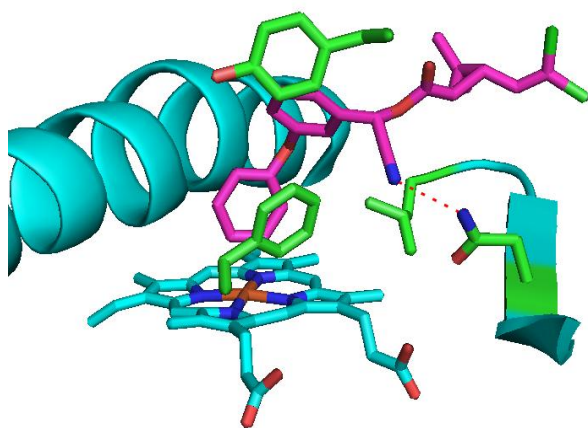
Figure 4.13 CYP6D1 binding cypermethrin with H-bonds to N392 (dotted red line).

4.2 Insect CYP6 non-metabolisers

4.2.1 *Anopheles gambiae* CYP6Z2

It has been found that CYP6Z2 is capable of binding cypermethrin and permethrin as they are capable of inhibiting BROD activity with a low IC50 (0.6 and 3.0 μM respectively), but has not been shown to metabolise them (McLaughlin et al, 2008). Although this may be due to the absence of b5, the dockings suggest that pyrethroids can adopt a non productive binding mode due to aromatic and polar residues distant from the heme. Two aromatic residues F212 and F222 appear to be a major determinant of binding with additional H-bonds formed with K48 and N369. In the non-productive binding mode the phenoxybenzyl group is held distant by aromatic interactions with F222 and F212 while the halogens approach the heme. In contrast, the metaboliser CYP6D1 has both of these substituted for non aromatic residues K213 and N222 and the ligand is able to bind productively (Figure 4.14).

4.2.2 *Aedes aegypti* CYP6CB1

CYP6CB1 has also been shown to be a non-metaboliser and the model shows a similar binding mode to CYP6Z2 with the alcohol group held distant by an aromatic network. An alignment shows that this aromatic network is conserved between CYP6Z2 and CYP6CB1; the residues F212 and F222 of CYP6Z2 align with W211 and F221 of CYP6CB1 (Figure 4.15). These aromatic residues in the FG loop were identified as involved in preventing metabolism in CYP6Z2 and may also contribute to a lack of activity in CYP6CB1. Both of these positions are substituted in the metabolisers CYP6D1, CYP6M2 and CYP6P9/3.
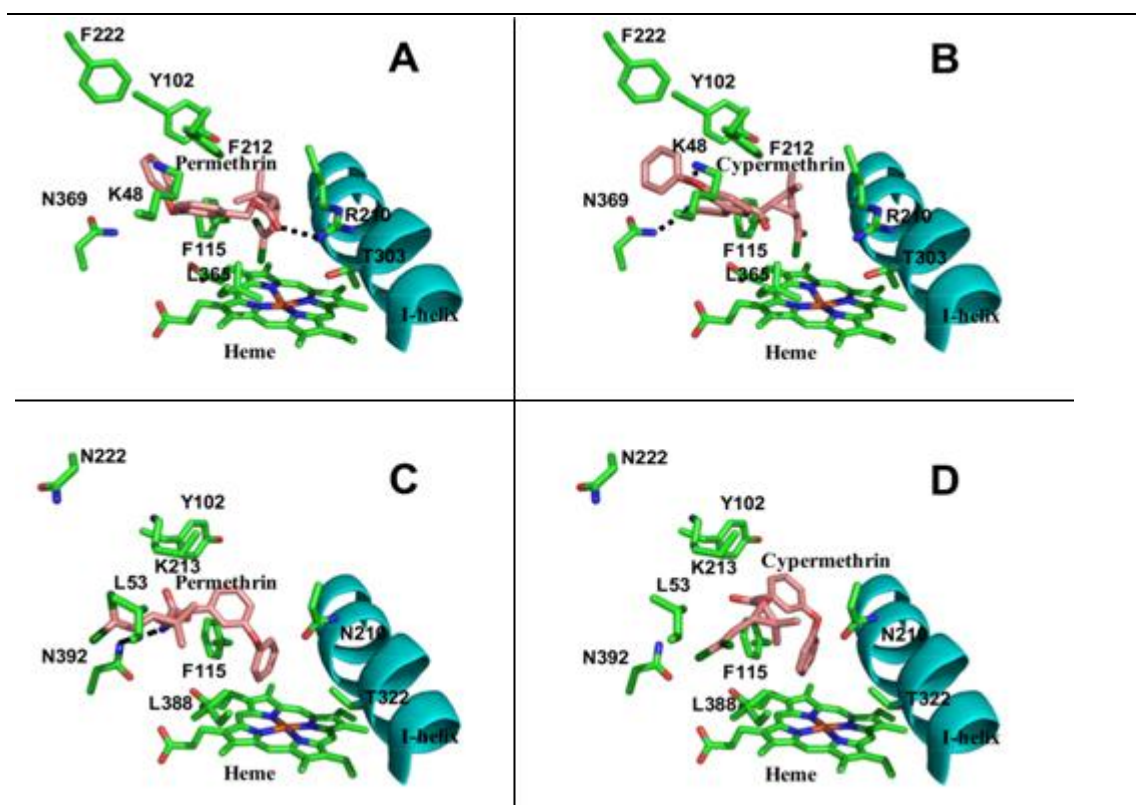
Figure 4.14 The binding modes of pyrethroids in CYP6Z2 and CYP6D1. (A) Permethrin in CYP6Z2; (B) cypermethrin in CYP6Z2; (C) permethrin in CYP6D1 and (D) cypermethrin in CYP6D1 taken from McLaughlin et al. (2008).

Figure 4.15 An alignment of the FG region of CYP6Z2 and CYP6CB1 showing conserved aromatic residues (arrows).

W211 is in a position to interact with the ligand but F221 is distant, although, as it occurs in an access channel it may have an influence on access. Unlike CYP6Z2, CYP6CB1 has additional aromatic residues contributing to this binding mode, in addition to W211, H104 of the BC loop and F378 of SRS5 are also in positions to stack with the ligand in this non productive mode (Figure 4.16).

Table 4.6 ChemScore (kJ/mol) of the best ranked mode of deltamethrin in CYP6CB1.

Score	ΔG	S(hbond)	S(metal)	S(lipo)	H(rot)	ΔE(clash)	ΔE(int)
43.2	-54.6	1.7	0.0	408.3	1.6	7.9	3.4
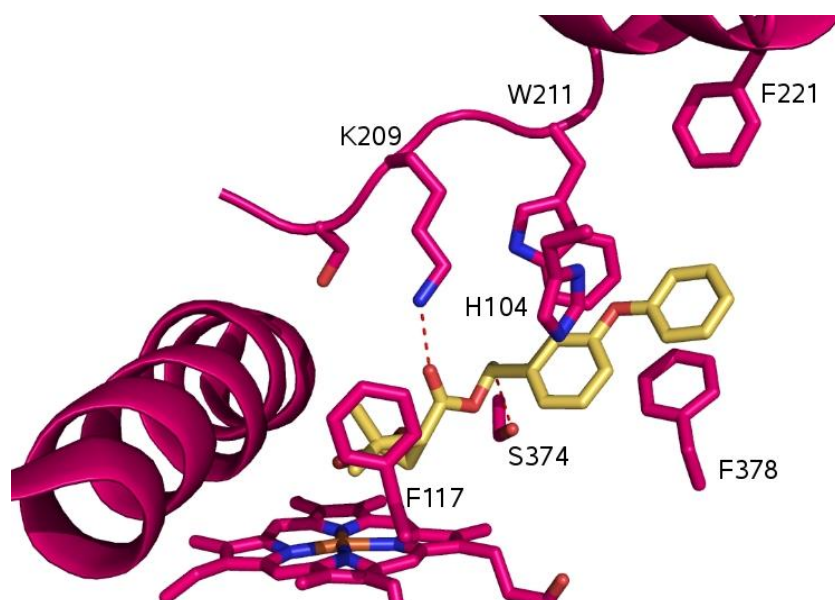
Figure 4.16 First ranked mode of deltamethrin in CYP6CB1, deltamethrin binds in a non-productive mode with two H-bonds to K209 and S374 (red dotted lines).

This non-productive mode is also high scoring (Table 4.6) due to the presence of two H-bonds with K209 and S374 (Figure 4.16). There may be other factors contributing to the lack of activity, the steric bulk of W211 may restrict access or binding within the active site. The aliphatic residue of SRS5 that is conserved in the metabolisers is substituted with a Ser, while this forms H-bonds with the ligand in the unproductive mode, it may also affect the hydrophobicity of the active site and affinity for hydrophobic ligands such as pyrethroids.

4.2.3 *Papilio polyxenes* CYP6B1

It has been previously shown that CYP6B1 is incapable of metabolising α-cypermethrin (Rcis S and Scis R mixture). Similar to CYP6Z2, in the CYP6B1 model cypermethrin binds with the halogens oriented towards the heme. The phenoxybenzyl group of cypermethrin is also surrounded by an aromatic network formed by F116, and F106 in the BC loop and Y210 in the FG loop. In the model these residues may hold the phenoxybenzyl group distant, allowing the halogens to orient towards the heme, with this mode occurring as the highest ranked modes for both isomers of α-cypermethrin. In addition, cypermethrin H-bonds to the sidechains of Y210 and T372 and the main chain of Q373 (Figure 4.17). However, in lower scoring modes the methyls are close to the iron which is inconsistent with the lack of activity found in CYP6B1. As the network of residues occurs along an access channel, they may affect access to the active site rather than positioning within the active site.
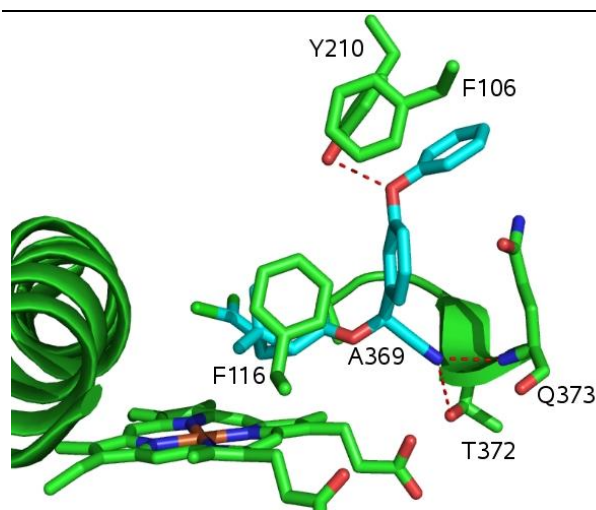
Figure 4.17 Cypermethrin binding in CYP6B1 with H-bonds to Q373, T372 and Y210 (red dotted lines).

4.3 CYP6 Metabolisers and Non-metabolisers comparisons

4.3.1 BC loop

The CYP6 metabolisers share a number of conservations. In the BC loop two aromatic residues are conserved. These correspond to F110 and F123 in CYP6M2 (Figure 4.18, A). In CYP6B8 an insert occurs but F108 aligns with the template F102 to place it within the active site. The known non-metabolisers of pyrethroids also share the conserved aromatic residues in the BC loop, but there are substitutions in the FG and SRS5 regions. Similarities in the active site between metabolisers and non-metabolisers could indicate that there may be other factors affecting metabolism such as binding in an unproductive mode or ligand access.

4.3.2 SRS5

In SRS5, the metabolisers have a conserved aliphatic residue that aligns with V372 of CYP6M2 (Figure 4.18, B). While metabolisers tend to have bulky aliphatic residues projecting into the site, the non-metabolisers may have small or polar residues at this position (Figure 4.18, C). The larger active site produced by Ala and Ser may allow positioning in non-productive modes or the substrate may not be stabilised above the heme. In metabolisers this bulky aliphatic residue may form hydrophobic interactions that may be absent in some non-metabolisers.

147

Figure 4.18 (A) Alignment of the BC loops of CYP6 metabolisers, the positions of F110 and F123 of CYP6M2 are identified. (B) Alignment of metabolisers SRS5 region, the position of V372 in CYP6M2 is identified. (C) Alignment of the SRS5 region of non-metabolisers, the position of L365 in CYP6Z2 is identified and is in an equivalent position to V372 in CYP6M2. (D) Alignment of the FG region of non-metabolisers the positions of F212 and F222 in CYP6Z2 are identified.

4.3.3 FG loop

In the CYP6Z2 model a number of residues were predicted to affect ligand access. One of these, F212, was in a position to π-stack with the alcohol moiety. An aromatic residue is conserved in this location in the other CYP6 non-metabolisers CYP6CB1 and

148

CYP6B1 and may be a feature in affecting access in these also. CYP6Z2 and CYP6CB1 have an additional aromatic residue F222 (Figure 4.18, D). Both of these residues are substituted in CYP6D1, CYP6M2 and CYP6P3.

4.4 *Helicoverpa zea* CYP321A1

CYP321A1 has not been considered to be part of the CYP6 family as it arose independently from a different progenitor gene, but phylogenetic analysis indicates that CYP321A1 forms a clade with the CYP6Bs that is separate from the other CYP6 clades (Li et al., 2003). It has been shown to metabolise α-cypermethrin at a lower rate than CYP6B8, 1.6 µmol/µmol P450 per minute for CYP321A1 compared to 2.44 µmol/µmol P450 per minute for CYP6B8 (Rupasinghe et al., 2007).
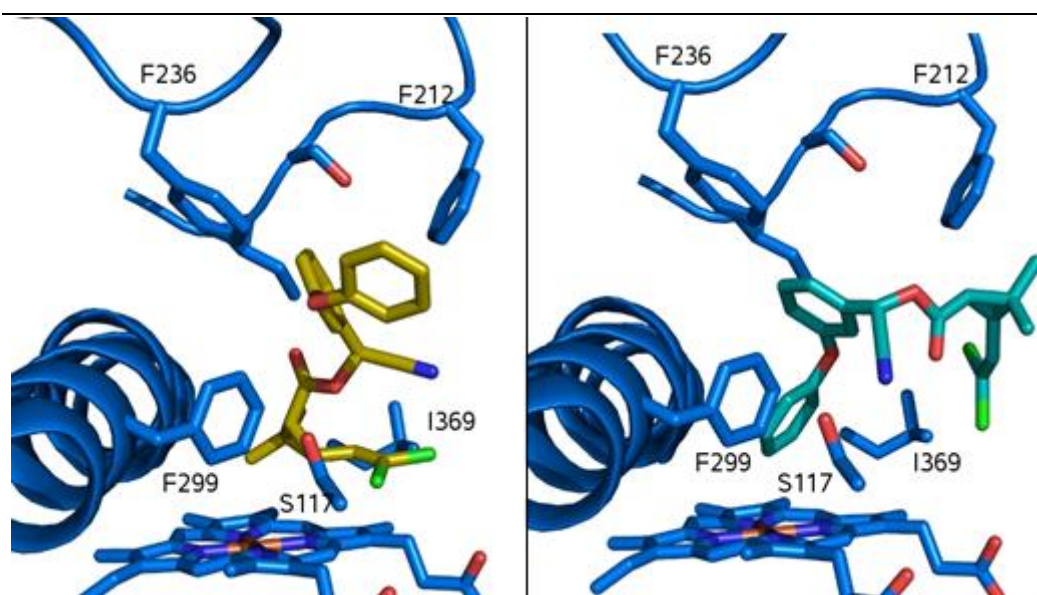
Figure 4.19 Cypermethrin RcisS binding in two modes in CYP321A1.

CYP321A1 differs from CYP6B8 in that it is missing the aromatic residues in the BC loop that are present in CYP6B8. These are substituted for S117 and L106. However, it contains aromatic residues in other locations that may be involved in binding. Cypermethrin binds for metabolism on the dimethyls by stacking with F299 of the I helix and may form aromatic interactions with F236 of the GG' loop. The CYP321A1 model can also bind cypermethrin for 4' metabolism. In this position there may be aromatic interactions between the phenyl ring and F299 (Figure 4.19). The aromatic residues F299 and F236 may have a similar role as F110 and F123 of CYP6M2, and may represent an alternative mechanism of pyrethroid binding than the CYP6s. The placement of the aromatic residues may be non-optimal positions for π-interactions that may contribute to the lower metabolism in CYP321A1.

149

4.5 *Aedes aegypti* **CYP9 family**

CYP9J26, 9J32, 9J28 and 9J24 are capable of metabolising pyrethroids, CYP6CB1 and CYP9J19 are non metabolisers. Compared to CYP6P9, it was shown that deltamethrin is metabolised preferentially over permethrin (Figure 4.20).
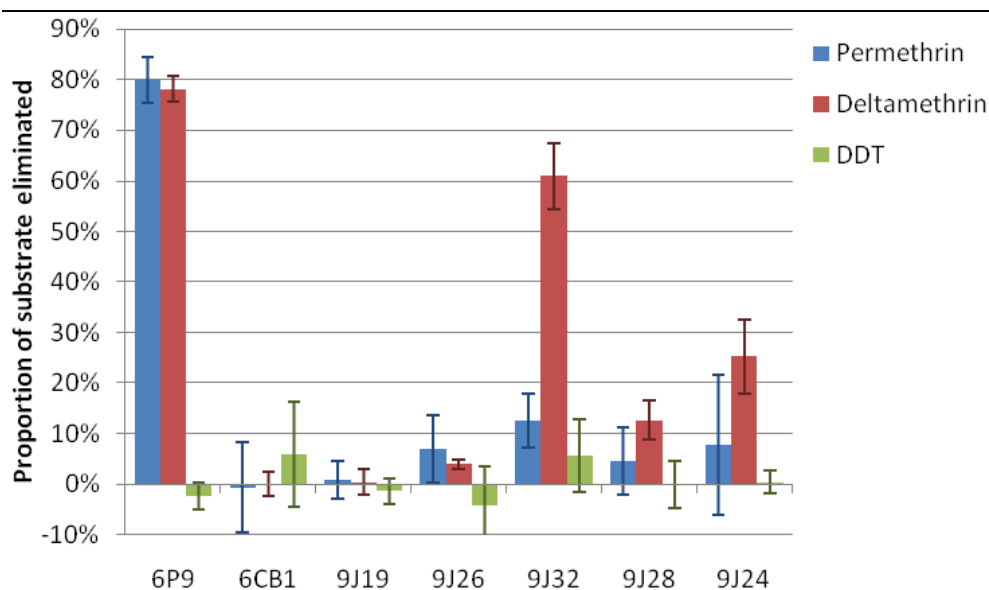
Figure 4.20 The experimentally determined (McLaughlin, unpublished) activities of P450s towards permethrin, deltamethrin and DDT.

4.5.1 Pyrethroid metabolisers

4.5.1.1 CYP9J24

In CYP9J24, both pyrethroids bind for metabolism of the alcohol group with the benzyl ring stacking with F126. H-bonds occur between K224 and both pyrethroids, but the deltamethrin cyano group forms additional H-bonds with the backbone and has a higher H-bond score. The presence of a additional stabilising H-bond may contribute to the preference for deltamethrin (Figure 4.21, A and B).

4.5.1.2 CYP9J26

Both pyrethroids bind for metabolism of the acid group with the benzyl ring stacking with F111. CYP9J26 shows similar clearance for both permethrin and deltamethrin and deltamethrin docks with a poorer score. However, the highest scoring deltamethrin mode shows a high H-bond score while the permethrin Rcis isomer does not show a H-bond. With the acid bound above the heme, H-bonds are possible between the cyano group and K228 (Figure 4.21, C and D). Rcis permethrin binds in a similar mode but no H-bonds occur, although other isomers form a H-bond with S131.
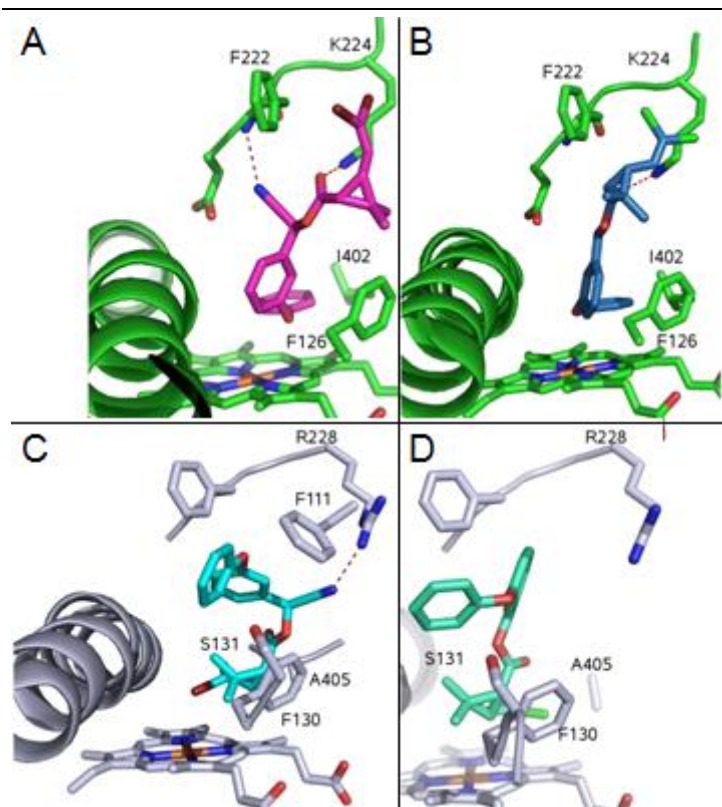
Figure 4.21 CYP9J24 binding deltamethrin (RcisS) (A) and Rcis permethrin (B). CYP9J26 binding deltamethrin (C) and Rcis permethrin (D).

4.5.1.3 CYP9J28
CYP9J28 is similar to CYP9J26 in that similar modes occur for both deltamethrin and permethrin with the benzyl ring stacking with F109. A H-bond occurs with deltamethrin that does not occur with Rcis permethrin (Figure 4.22, A and B). However, the other isomers bind in alternate modes and can form H-bonds.

4.5.1.4 CYP9J32
With CYP9J32 a large difference in metabolism is seen between deltamethrin and permethrin but the scores for deltamethrin are poor. Both pyrethroids bind with the acid group above the heme with the benzyl ring stacking with F111. The best ranked modes of both permethrin and deltamethrin form single H-bonds, the permethrin forming H-bonds with S128 while deltamethrin forms H-bonds with the main chain of the FG loop F226 (Figure 4.22, C and D). As there are similar H-bond scores, the presence of an additional stabilising H-bond may not explain the higher metabolism of deltamethrin.
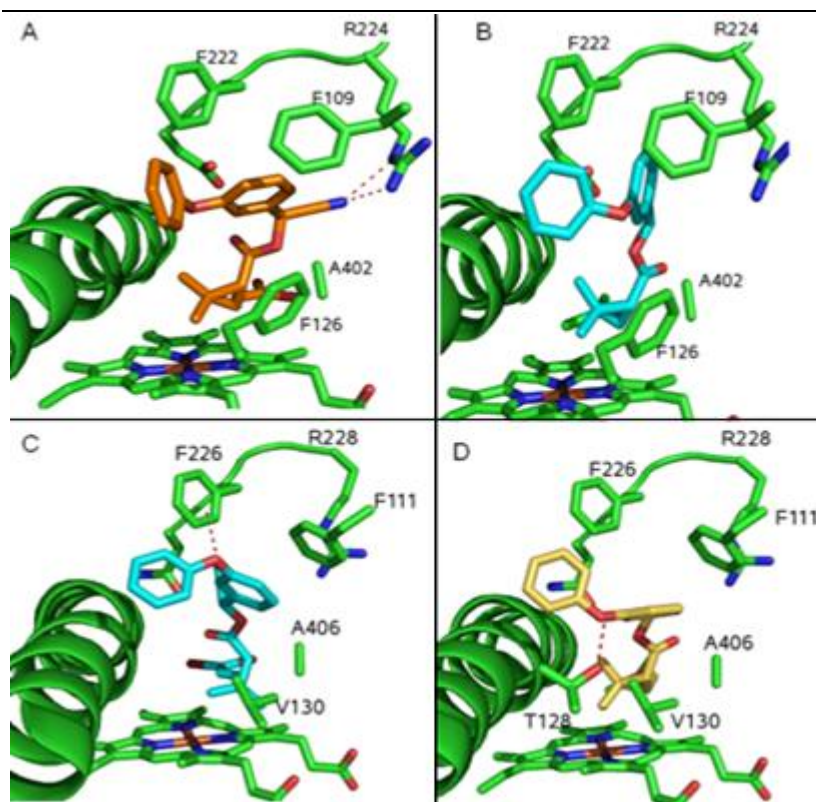
Figure 4.22 CYP9J28 binding deltamethrin (A) and permethrin (B). CYP9J32 binding deltamethrin (C) and permethrin Rcis (D).

4.5.2 Remaining questions

The models do not offer an explanation for the difference in clearance between permethrin and deltamethrin nor an explanation for why this difference occurs in the CYP9s but not the CYP6s. In all models, permethrin has similar modes and scores higher than deltamethrin. This is inconsistent with the experimental results that show that permethrin is metabolised either similarly or poorly. The possibility that the larger bulk of deltamethrin provides an increase in stability is not supported as deltamethrin tended to show lower vdW scores than the highest scoring permethrin isomer. Both the CYP9s and CYP6s show higher scores for the permethrin isomers than deltamethrin. This may be due to a lower rotatable bond freezing term for permethrin compared to deltamethrin as deltamethrin has 10 rotatable bonds compared to 9 in permethrin.

An alternative possibility could be the existence of additional H-bonds. Unlike permethrin, deltamethrin can form H-bonds with the cyano group, It may be deltamethrins ability to form additional stabilising H-bonds that allows for a greater clearance. However, this was not found as the H-bond score for deltamethrin was not consistently higher than that of permethrin and the cyano group does not offer

additional H-bonds except in CYP9J24. Additionally, the H-bonding pattern differs between P450s with no donor shared between the metabolisers. It is possible that factors external to the active site may play a role in substrate selectivity.

4.5.3 Non-metaboliser CYP9J19

CYP9J19 has a number of substitutions of residues conserved in the metabolisers. Both BC loop Phes are substituted and the aliphatic residue in SRS5 is substituted with a Thr (CYP6CB1 also has a similar substitution for a Ser) (Figure 4.23). Although CYP9J19 is a non metaboliser, the best ranked modes bind for metabolism, although unproductive modes occur involving H-bonding to the SRS5 Thr. The productive mode is low scoring (Table 4.7) and this appears to be due to a low vdW/hydrophobic score which may be due to the substitution of a Phe for Ser, but may also be due to the substitution of a Ile conserved in the I helix of the other CYP9s for a smaller V337 that gives poor vdW contacts.

Table 4.7 ChemScores (kJ/mol) of the best ranked mode of deltamethrin.

	Score	ΔG	S(hbond)	S(metal)	S(lipo)	H(rot)	ΔE(clash)	ΔE(int)
9J19	37.1	-40.4	0.0	0.0	334.4	1.6	2.8	0.6
9J24	48.4	-52.6	1.7	0.0	389.1	1.6	0.4	3.8
9J26	44.8	-48.6	1.4	0.0	365.7	1.6	1.1	2.8
9J28	49.7	-53.7	1.5	0.0	405.8	1.6	0.9	3.2
9J32	44.0	-50.7	1.0	0.0	394.1	1.6	4.5	2.2

Both non metabolisers CYP6CB1 and CYP9J19 both have a polar residue in SRS5 that can H-bond with deltamethrin in non-productive modes. Alternatively it may be the substitution of a non polar for a polar residue that affects the hydrophobicity of the active site and its affinity for hydrophobic ligands.

4.5.4 Comparison of CYP9 metabolisers and non metabolisers

The CYP9J metabolisers share conservations with the CYP6 metabolisers CYP6P3, CYP6P9 and CYP6M2 such as two conserved aromatic residues in the BC loop and an aliphatic residue in SRS5. This suggests common binding interactions between these metabolisers. The CYP9J2s also have similarities that are particular to this family such as conserved aliphatic residues on the I helix (I331), a Ser in the BC loop (S127) and a basic residue in the FG loop (R224) (Figure 4.23). While CYP9J28, CYP9J26 and CYP9J24 have similar active sites, CYP9J32 and CYP9J19 differ from this group.
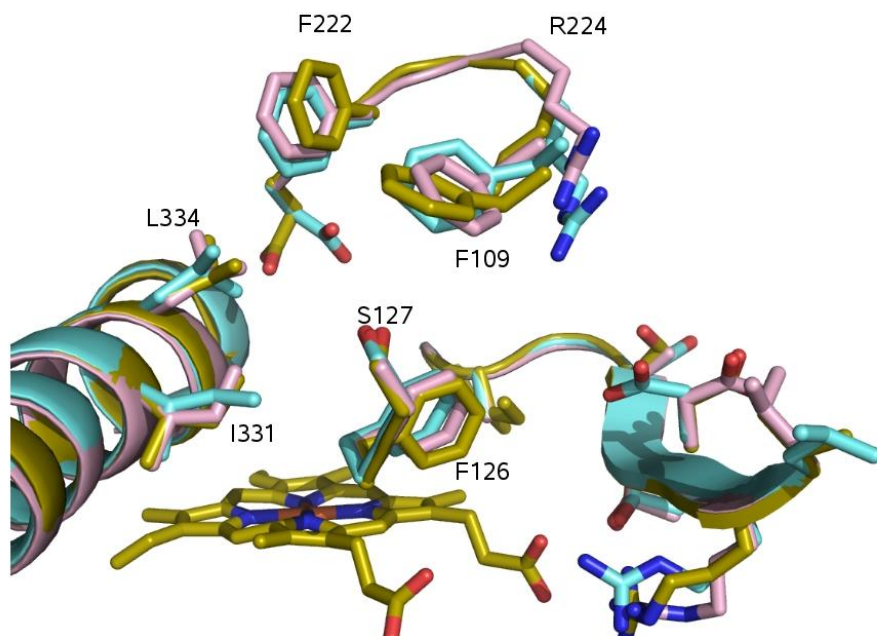
Figure 4.23 An overlay of CYP9J24, CYP9J28 and CYP9J26 showing active site conservations. The numbering corresponds to CYP9J28.
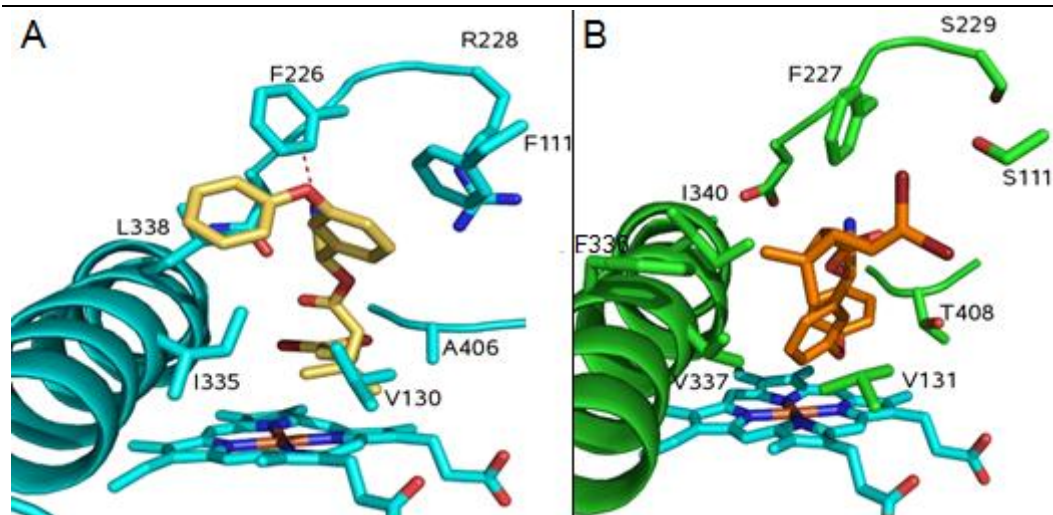
Figure 4.24 The binding modes and active sites of CYP9J32 (A) and CYP9J19 (B).

CYP9J32 and CYP9J19 both differ from the 9J2 group and differ from the other metabolisers. In both CYP9J32 and CYP9J19 the conserved BC loop Phe above the heme is substituted for a Val, but while CYP9J32 still shares the second conserved Phe, in CYP9J19 this is substituted for a Ser. While CYP9J32 has one of the conserved residues substituted, it still shares the second conserved BC loop Phe and is able to bind deltamethrin in a similar mode as CYP6M2, and the Phe is in a position to form aromatic interactions with the benzyl ring of deltamethrin. This residue may act as a π-anchor for this particular binding mode. The Phe above the heme is not in a position to form aromatic interactions in this binding mode and may provide only hydrophobic

interactions and the substitution of this Phe for an aliphatic residue may be tolerated (Figure 4.24).

4.6 Pharmacophore and pyrethroid binding fingerprint

4.6.1 'Typical' metabolisers

All of the metabolisers CYP9J24, CYP9J26 and CYP9J28 are closely related and share active site similarities. These similarities are also shared with the CYP6 metabolisers CYP6P9, CYP6P3, CYP6M2 CYP6B8 and CYP6D1 and may be termed 'typical' metabolisers. The conservations between this group of metabolisers may indicate common ligand binding interactions. In all of these metabolisers, there are two conserved aromatic residues that both project into the active site (Figure 4.25, A). Within the BC loop there are few other conservations and large inserts occur in the CYP9J2 group. There is also the conservation of an aliphatic residue in SRS5 although this varies in size (Figure 4.25, B). Few conservations occur in the FG loop, but a conserved polar residue tends to occur above the heme and project into the active site (Figure 4.25, C).

4.6.2 Non-typical metabolisers and non-metabolisers

CYP9J32 and CYP321A1 are metabolisers but differ from the 'typical' metabolisers in that the conserved aromatic residues are substituted. In CYP9J32 the conserved BC loop aromatic residue above the heme is substituted by V/G but still shares the second conserved residue. In CYP321A1 both of the conserved residues substituted but other aromatic residues on the I helix and FG loop may act as substitutes. Some non-metabolisers resemble the metabolisers, CYP6B1 and CYP6CB1 both share the conserved residues that occur in the metabolisers, while CYP9J19 has both of these residues substituted (Figure 4.26, A). In SRS5, both non metabolisers CYP6CB1 and CYP9J19 have the conserved aliphatic residue substituted with a polar Thr or Ser (Figure 4.26, B).
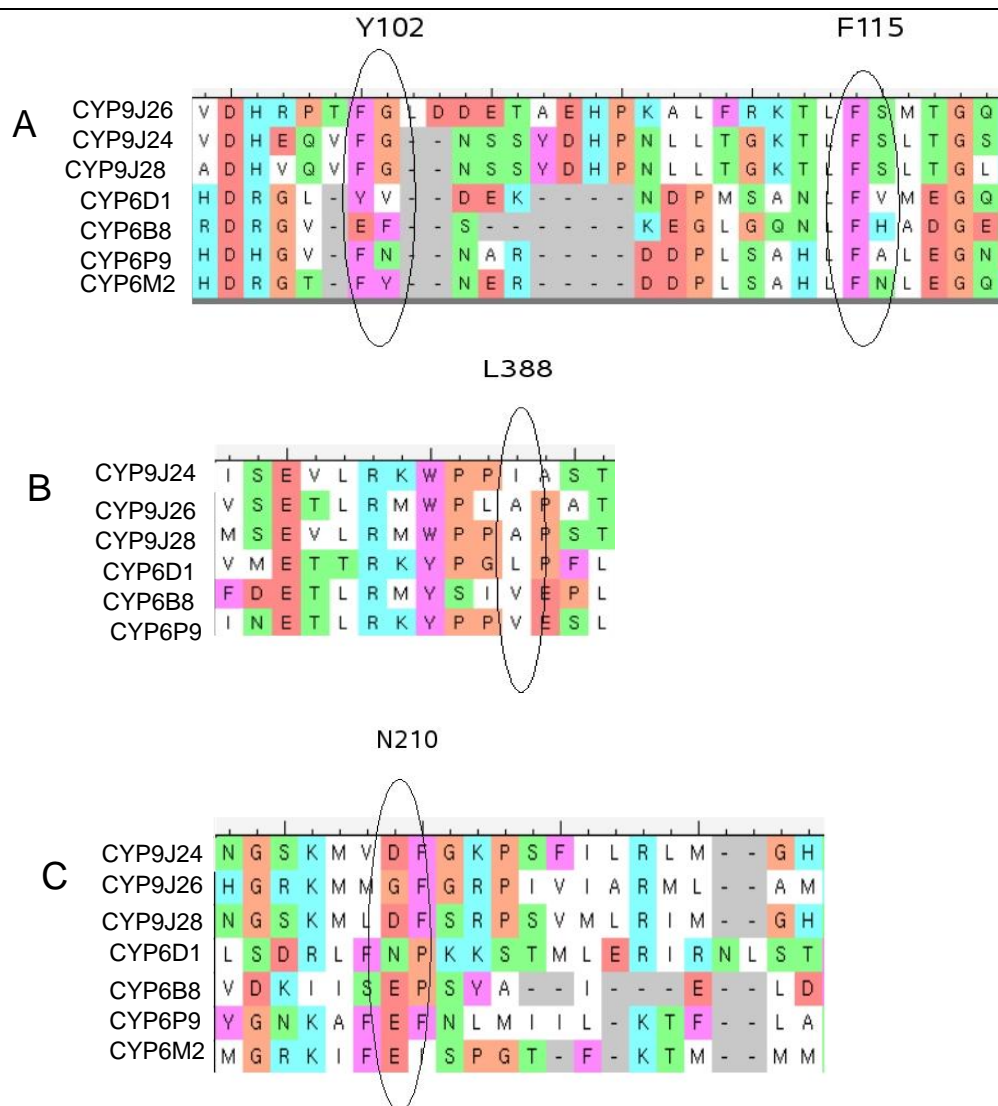
Figure 4.25 (A) Conserved aromatic residues of the BC loop identified by alignment with CYP6D1. (B) A conserved aliphatic residue in SRS5, with L388 in CYP6D1 identified.(C) A conserved polar residue in the FG loop alignment with N210 in CYP6D1 identified.

The use of docking and pharmacophores were used here to suggest a pyrethroid binding fingerprint to explain the ability of both the 'typical' and 'non-typical' metabolisers to bind pyrethroids. This also suggests an explanation for the lack of metabolism in P450s that show substitutions that differ from the metabolisers. However, the active sites of some non-metabolisers share the same conservations that occur in the 'typical' metabolisers and the lack of metabolism cannot be attributed to differences in the active site.
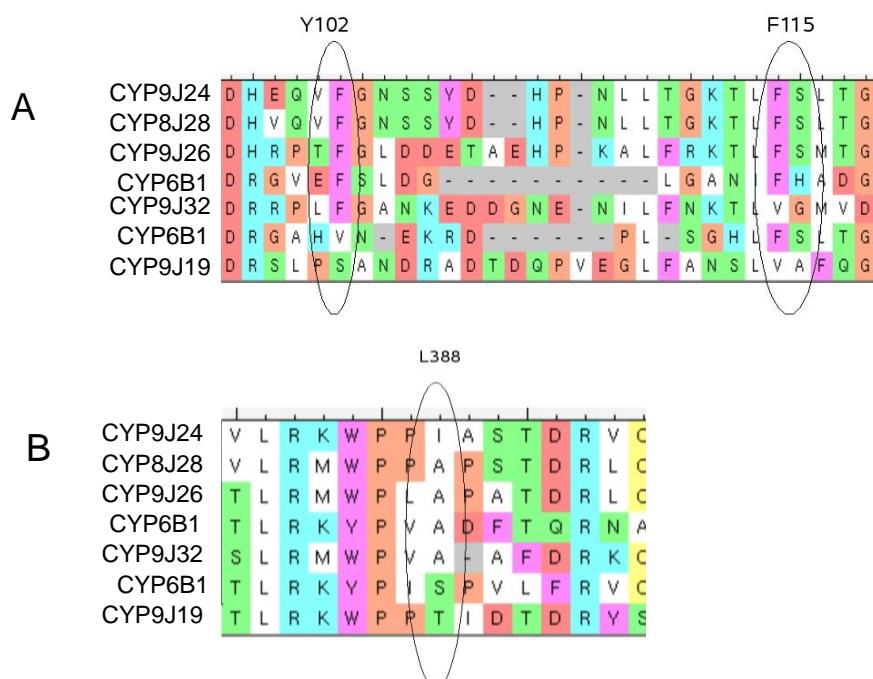
Figure 4.26 (A) The BC loop alignment of metabolisers and non-metabolisers with the BC loop residues aligning with CYP6D1 Y102 and F115 identified. (B) Alignment of metaboliser and non-metaboliser SRS5 region with residues aligning with housefly L388 identified.

4.6.3 CYP6/9 clade pharmacophore

The metabolisers with bound ligands (CYPs 6M2, 6P3, 6P9, 6B1, 6B8, 321A1, 9J24, 9J26, 9J28, 9J32) were structurally aligned using MOE, which overlapped the bound ligands. A consensus pharmacophore was produced from the overlapped ligands to identify common interactions with the receptor. From the structural alignment of the receptors the bound ligands were well overlaid despite having differing binding modes, and the pharmacophore features of the ligands could be grouped into a consensus pharmacophore (consensus threshold of 30%) (Figure 4.27, A). The pharmacophores were produced in MOE. Structural features are represented as points in space and annotated encoding its structural features. The label 'Aro' is applied to an aromatic centre, 'Hyd' to hydrophobic regions, 'ML' to metal ligands, and 'Acc' to H-bond acceptors.
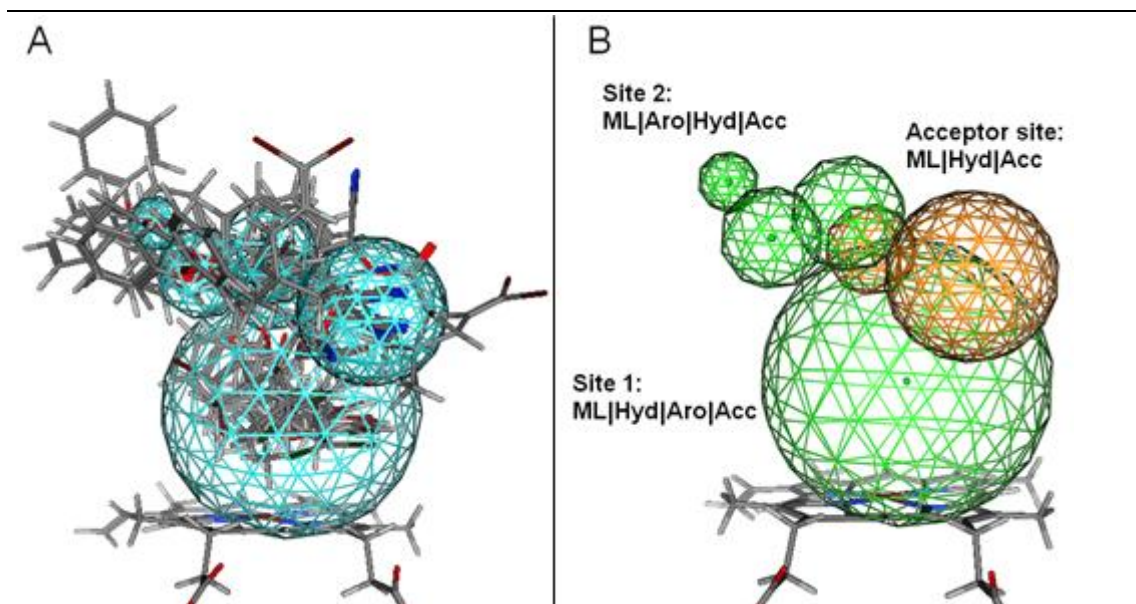
Figure 4.27 The pyrethroids bound in metabolising CYPs were aligned based on a structural alignment of receptors. The pharmacophore was produced from a consensus of this alignment (A). The pharmacophore was coloured according to the presence (green) and absence (orange) of aromatic features and grouped into sites (B). The label 'Aro' is applied to an aromatic centre, 'Hyd' to hydrophobic regions, 'ML' to metal ligands, and 'Acc' to H-bond acceptors.

Features of this consensus pharmacophore could be grouped into sites (Figure 4.27, B) consisting of two groups of hydrophobic and aromatic features (sites 1 and 2) separated by a group of hydrophobic and H-bond acceptor features (acceptor site). Pyrethroids bind with either the acid or alcohol groups above the heme. These are clustered into a single bulky feature (site1). This feature implies an interaction with the heme and hydrophobic and aromatic residues of the BC loop, I helix and SRS5 above the heme. Above this are two clusters consisting of hydrophobic and H-bond acceptor features (acceptor site), which were identified due to clustering of the ester bond and cyano groups. Above the ester bond are other acceptor and aromatic features that correspond to the overlaid phenoxybenzyl groups (site2).

From all of the metabolisers two general modes are identified: dockings with either the acid or alcohol groups above the heme, and a pharmacophore consisting of two sets of hydrophobic and aromatic features separated by an acceptor/hydrophobic feature may account for the two docking modes found. Both docking modes can be overlaid onto the pharmacophore, indicating that similar interactions may occur with both modes. With

docking modes placing the acid group above the heme, such as with CYP6M2 (Figure 4.28, A and B), the bulky acid group occupies the area of site 1 while the phenoxybenzyl group occupies site 2. The ester bond and cyano groups occupy the acceptor features between the sites. The ligands bound within CYP321A1 6B8, 9J26, 2J28 and 9J32 overlay with the pharmacophore in a similar way as CYP6M2. With docking modes where the alcohol group binds above the heme such as in CYP6D1 (Figure 4.28, C and D), the alcohol group spans both aromatic sites with the phenyl ring within site 1 and the benzyl ring within site 2. The acceptors of the phenoxybenzyl and ester oxygens occupy the acceptor area.

As both binding modes are overlaid into a common pharmacophore, they may indicate common binding interactions. The common features of both modes include a π-stacking between the conserved aromatic residue in the BC loop (site 2) and the benzyl ring, and hydrophobic or aromatic interactions with the residues immediately above the heme (site 1). In both modes the acceptors of the ester bond and cyano groups are overlaid but no common interactions with the protein occur with these acceptors.

In both of these possible modes, both aromatic and hydrophobic interactions may be important in docking, and the placement of the pharmacophore within the active sites of the metabolisers shows the common active site features that may be involved in binding. The two conserved BC loop aromatic rings are complementary to the position of the pharmacophore aromatic sites 1 and 2. The large aromatic and hydrophobic site 1 is surrounded by complementary aromatic and hydrophobic residues on the BC loop I helix and SRS5. The position of the aromatic site 2 corresponds to the position of the conserved aromatic residue on the BC loop aligned with Y102 (CYP6D1). However, the acceptor region only shares complementary electrostatic sites with a few metabolisers and H-bonding may have a minor role in individual enzymes.
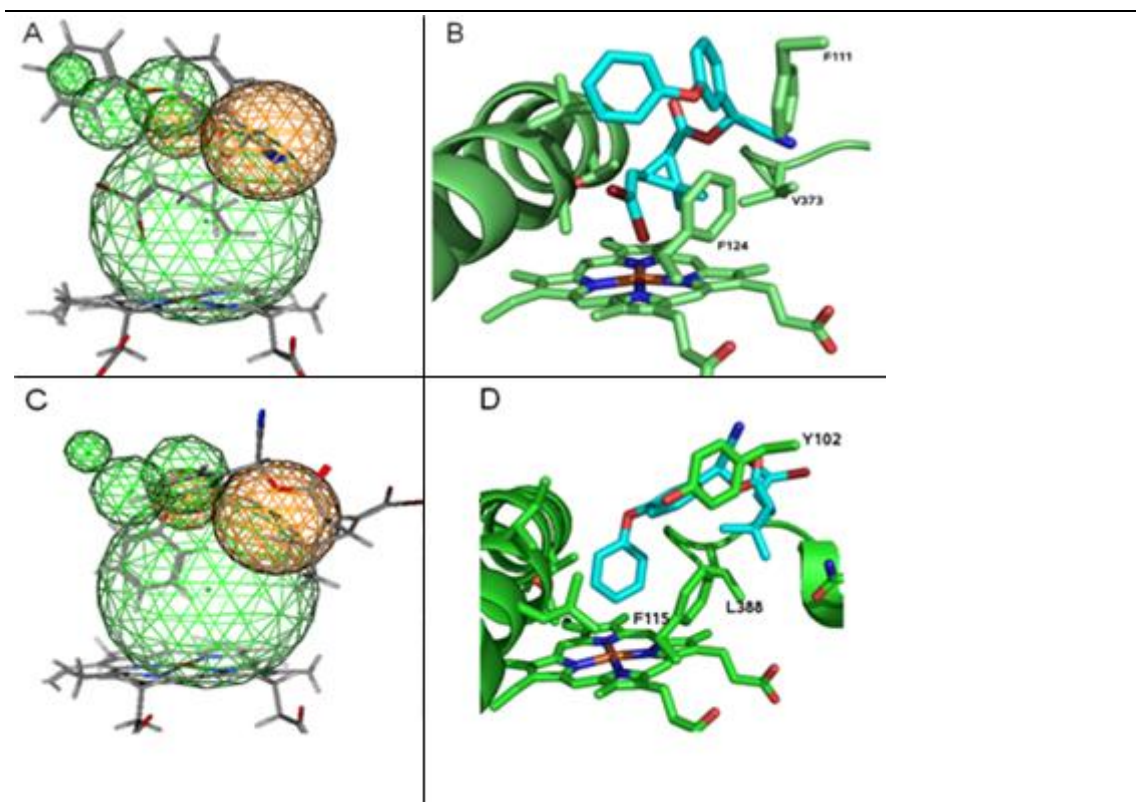
Figure 4.28 (A) The pharmacophore overlaid over the CYP6M2 docking (viewed towards the I helix) and (B) the ligand bound in the CYP6M2 active site. (C) The pharmacophore is overlaid onto the docking of deltamethrin into CYP6D1 (viewed towards the I helix) and (D) the ligand bound in the CYP6D1 active site.

4.6.4 Non-'typical' metabolisers pharmacophore comparison

The non-typical metaboliser CYP321A1 lacks both of the conserved aromatic residues on the BC loop but the placement of the pharmacophore into the active site shows two Phes in complementary positions for interaction with aromatic sites 1 and 2, as well as hydrophobic residues surrounding site 1 (Figure 4.29). The ligand also overlays into the pharmacophore with the bulky acid occupying site 1 and the phenoxybenzyl group occupying site 2 (Figure 4.30). Despite the substitution of the conserved BC loop residues, there may be an occurrence of similar interactions in the non-'typical' metabolisers. This indicates that there may be a common binding fingerprint for both the 'typical' and non-'typical' metabolisers.
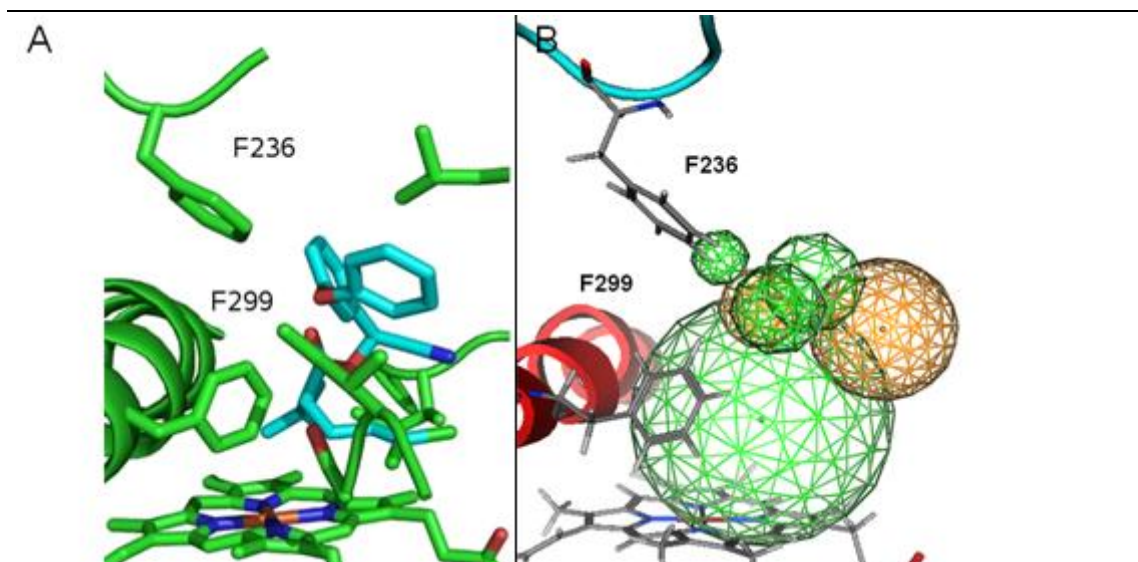
Figure 4.29 RcisS cypermethrin bound in CYP321A1 (A) and the pharmacophore positioned within the active site (B), the aromatic residues F236 and F299 are in positions that are complementary to the aromatic sites of the pharmacophore.
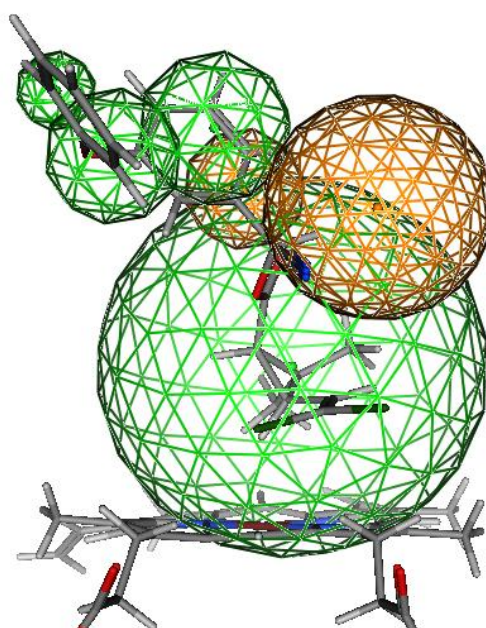
Figure 4.30 The pharmacophore overlaid onto the docking of deltamethrin into CYP321A1.

4.6.5 A pyrethroid binding fingerprint

Based on the pharmacophore and commonalities between the binding interactions in the metabolisers, a pyrethroid binding fingerprint is suggested. This offers an explanation for the occurrence of alternative but superimposable binding modes as well general explanations for the occurrence of mutations that have P450 and substrate specific affects. This fingerprint consists of two primary aromatic/hydrophobic features that may

be common to all metabolisers and a minor H-bond feature that may be P450 specific:

1) Hydrophobic/vdW or aromatic interactions above the heme (pharmacophore site1).
2) A π-anchor (pharmacophore site 1 or 2).
3) A H-bond with either the ester or cyano groups (pharmacophore acceptor site).

4.6.5.1 Hydrophobic/aromatic site 1

The aromatic or hydrophobic residues above the heme in the BC loop and SRS5 tend to be conserved among metabolisers. This may provide a hydrophobic cavity for the binding of the acid group as in CYP6M2. It has been demonstrated that mutations that increase the hydrophobicity of the active site increase turnover. In P450cam the mutations Y96F and Y96A removed the hydroxyl group to give a more hydrophobic active site and improved activity. It has been proposed that hydrophobic side chains interact more strongly with hydrophobic substrates and promote the displacement of water molecules from the active site upon substrate binding (Nickerson et al., 1997). This hydrophobic cavity may be necessary for the binding of hydrophobic substrates such as deltamethrin.

In the metabolisers, SRS5 has a conserved aliphatic residue of varying size which may contribute to the hydrophobicity of the cavity above the heme. In addition to this, residues in this position are known to have effects on regiospecificity. The differences in the size of this residue could produce the preference for the docking of either the acid or alcohol above the heme. The BC loop Phe above the heme may have a similar role as the aliphatic residue in SRS5 in providing a hydrophobic environment, or could provide an additional π-anchor for the binding of the alcohol group.

4.6.5.2 π-anchor

Aromatic interactions may act as a π-anchor to position the substrate. The presence of a π-anchor appears to be a common interaction as it occurs among the 'typical' metabolisers but also occurs in the 'non-typical' metabolisers such as CYP9J32 and CYP321A1. The presence of two conserved aromatic residues in the BC loop allow for the possibility of two π-anchors, and the role each plays could be determined by the preference for binding mode, and mutations of these residues could have P450 specific effects depending on the importance of π-stacking. In CYP6D1 both of these residues could be involved in π-stacking, while in the best ranked mode of CYP6M2 only one

may be involved in stacking. The presence of two possible π-anchors in the active site may allow deltamethrin to bind in alternative modes and to produce a range of metabolites.

A role for a π-anchor has been previously proposed in CYP2C9 by Haining et al. (1999) when it was found that mutation of an aromatic residue, the π-anchor, in the active site affected the metabolism of aromatic substrates but not aliphatic ones. Haining et al. (1999) suggested that F114 could act as a source of π-stacking interactions for certain aromatic substrates. They found that the effects of the F114L mutation were substrate selective with a graded response. The mutation had no affect on the aliphatic substrates lauric acid and arachidonic acid metabolism but affected both efficiency and metabolite profiles for the aromatic warfarin with novel metabolites produced, with intermediate affects for dichlofenac. The F114L mutation retained 16% of the warfarin activity, but 70% of the lauric acid activity. The substrate dependant affects of this mutation may vary with the importance of π-stacking as a determinant of binding for a given substrate.

A number of studies on CYP2C9 have shown that the F114L mutation decreased the efficiency of metabolism, or abolished metabolism for the aromatic substrates warfarin, flurbiprofen, diclofenac and tienilic acid. However, the F114W mutation preserved the aromaticity and maintained wildtype activity for warfarin and flurbiprofen. Mosher et al. (2008) suggested that the aromatic interaction between substrates and F114 could be important for normal catalytic activity.

Mutagenesis studies in a number of P450s have shown that aromatic residues in these positions in other P450s have important roles in substrate selectivity, regiospecificity and rates, but mutations have been shown to have P450 and substrate specific affects. In CYP3A4 mutagenesis of F108 showed a significant change in activity and a shift in preference for regiospecificity. Mutagenesis for a larger residue F108W, showed a higher preference for 1'-OH of midazolam while mutation for a smaller residue F108A showed no change in regiospecificity (Khan et al., 2002). However, in CYP2C9 mutations of this position (F100L) were silent. From the dockings in this study, while the residues are conserved they show different interactions with the substrate, P450 specific effects of mutations could be expected between metabolisers.

Apart from substrate binding this Phe may have other roles such as access or egress. Li

et al. (2007) used steered MD to accelerate the unbinding of metyrapone from the active site of CYP3A4 (1W0G) to identify the unbinding pathways. They found that F108 and I120 act as gate keepers to prevent the ligand from exiting the active site by forming hydrophobic interactions. To allow the ligand to exit, F108 rotated to expand the volume of the channel.

4.6.5.3 H-bond

This may be a minor factor in determining metabolism as the occurrence of H-bonds may be P450 specific and no common H-bond donors occur.

4.6.6 Non-metabolisers pharmacophore comparison

Three non metabolisers (CYP6CB1, 6B1 and 9J19) were also aligned but fewer consensus features were found. Hydrophobic and aromatic features above the heme were the only consensus features. This may be due to having only a small number of non metabolisers identified, but they lack a consistent binding mode this may indicate that they lack the interactions present in the metabolisers (Figure 4.31, A). All of the ligands bound within the metabolisers overlap the metaboliser pharmacophore well. By contrast, the non-metaboliser CYP9J19 overlaps poorly with the metaboliser pharmacophore with only the phenoxybenzyl group overlapping with site 1 (Figure 4.31, B).

The placement of the pharmacophore into the non-metaboliser CYP9J19 shows few complementarities to the active site (Figure 4.31, D). Like CYP321A1, the aromatic residues of the BC loop are substituted but only F227 in the FG loop is near to the aromatic site 2. The positioning of this residue may be in a non-optimal position as it does not form aromatic interactions with deltamethrin in modes that allow metabolism (Figure 4.31, C), but can form interactions in non-productive modes. The occurrence of both a reduced hydrophobic interaction and a reduced ability to form π-interactions may contribute to the lack of metabolism in CYP9J19.
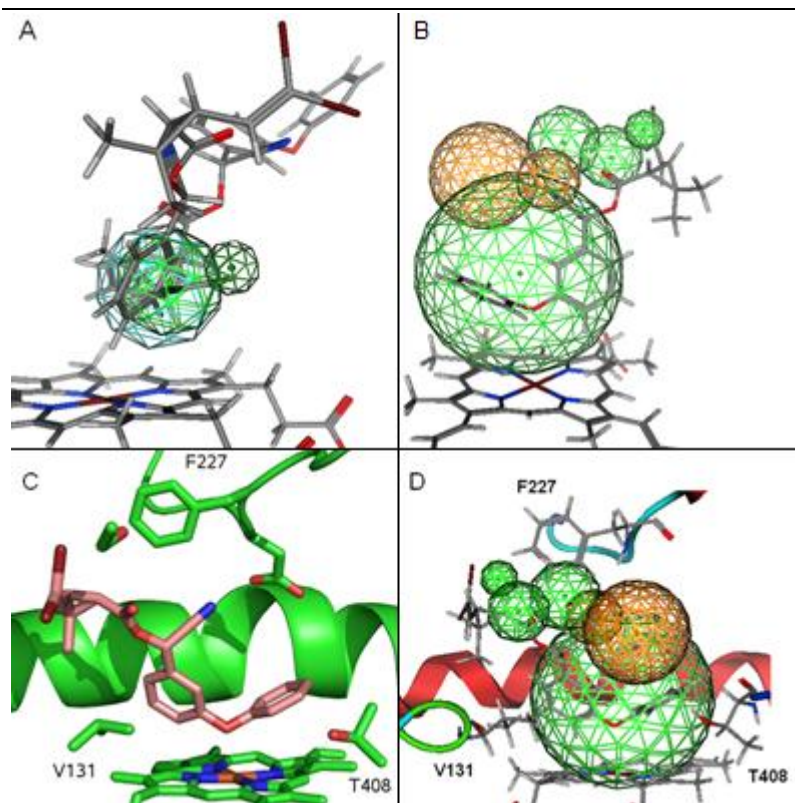
Figure 4.31 A Consensus pharmacophore based on the overlap of deltamethrin bound within the active sites of non-metabolisers (A). The metaboliser pharmacophore overlaid onto the docking of deltamethrin in CYP9J19 (B). Best ranked mode of deltamethrin binding to CYP9J19 (C). The pharmacophore overlaid in to the CYP9J19 active site (D).

4.6.7 Regiospecificity

As similar interactions occur with both modes, differences in regiospecificity may not be due to differences in interactions with the protein but may be due to steric factors. CYP6D1 which has bulky residues above the heme, shows metabolism on the alcohol while those with smaller residues bind also for metabolism on the acid group. The difference in regiospecifitiy may be due to impairing the space available for the docking of the acid group in this region while permitting the binding of the alcohol, but similar interactions occur with both modes.

4.6.8 Pharmacophore summary

A pharmacophore was produced based on the overlap of deltamethrin or cypermethrin bound within the active sites of the CYP6/9 metabolisers. This pharmacophore consists of one large aromatic/hydrophobic site (site 1), a collection of smaller aromatic sites (site 2) and a collection of acceptor sites (acceptor region). There is a close agreement

between this pharmacophore and active sites of the metabolisers with the aromatic residues of the BC loop complementing the aromatic sites 1 and 2 and hydrophobic residues surrounding site 1. Based on this there appears to be two primary factors determining the binding of deltamethrin, an aromatic and a vdW/hydrophobic interaction, with other minor factors being P450 specific.

A π-anchor hypothesis is suggested for the anchoring of the aromatic alcohol group. The occurrence of two possible π-anchors in the BC loop could explain the alternative binding modes of deltamethrin as similar interactions could occur in alternative modes. With similar interactions the regiospecificity of deltamethrin may be determined by other P450 specific characteristics such as steric restrictions. The two π-anchors may have different roles in the binding of different substrates, this could indicate a mechanism of how the mutagenesis of similar residues in other P450s have been shown to be substrate specific. The second steric and hydrophobic interaction is determined by aromatic/aliphatic residues immediately above the heme. Both of these interactions appear to be important in deltamethrin binding in all of the CYP6/9 metabolisers, and the non-metaboliser CYP9J19 has substitutions that affect both of these interactions.

4.7 Comparisons of P450s to the fingerprint

A range of known pyrethroid metabolising P450 families from rat and human were also analysed. There may be different factors in each family that determine metabolism but the stacking of the pyrethroid ligand with aromatic residues are a factor across clades, the metabolisers tend to contain aromatic groups in range to contact the ligand in productive poses. In the CYP6, CYP9 CYP2 and CYP3 families, two aromatic residues tend to be conserved on the BC loop that may be determinants of pyrethroid binding as they can act as π-anchors to bind the phenoxybenzyl group. However, non typical metabolisers can occur in each family that differ from this pattern, but which have their own alternate pattern of aromatic residues to act as π-anchors. These proteins may represent alternate solutions to the problem of pyrethroid binding.

While the metabolisers share commonalities in pyrethroid binding, the non metabolisers are more diverse. There may be a number of reasons why the non metabolisers fail to metabolise pyrethroids, some P450s lack the aromatic residues found in the metabolisers such as CYP9J19, and CYP2D1. Some non metabolisers have aromatic networks that may bind the ligand in non-productive poses such as CYP6Z2, CYP6CB1

CYP2B6 and CYP2C12. Others may have structural architecture that prevents efficient binding such as in CYP2D1 and CYP2D2 that have small active sites that provide poor contacts with the ligand.

4.8 Unknown P450s comparisons to the fingerprint

The fingerprint may be used to predict metabolism in candidate P450s. A number of P450s have been linked to resistance but experiments have not yet been undertaken to determine if they are able to metabolise pyrethroids. CYP6E1 was identified in pyrethroid resistant *Culex quinquefasciatus* (Kasai et al., 1998). CYP6F1 is over expressed in deltamethrin resistant *Culex pipiens pallens* (Gong et al., 2005). CYP6AA3 and 6P7 are correlated with deltamethrin resistance in *Anopheles minimus* (Rodpradit et al., 2005). The expression of CYP6D3 is also increased in LPR housefly compared to the susceptible CS strain and is 78% identical to CYP6D1 (Kasai and Scott, 2001). Kamiya et al. (2001) also found CYP6D1, 6D3v2 and CYP6A24 over expressed in the pyrethroid resistant housefly strain (YPER). CYP6A36 is over expressed in both larvae and adults of a pyrethroid resistant strain (ALHF), by contrast CYP6A37 does not show any expression differences (Zhu et al., 2008). Zhu et al. (2008) found a differential expression of two alleles in a pyrethroid resistant strain (ALHF). CYP6A5 showed no difference in expression between susceptible and resistant strains, but CYP6A5v2 showed a 1000 fold over expression in ALHF compared to the susceptible CS strain.

The P450 candidates for pyrethroid metabolism can be compared to the 'fingerprint' to predict if they are capable of pyrethroid metabolism, this could also prioritise candidates for experimental studies. Metabolisers tend to have two aromatic residues in the BC loop. A multiple alignment of a number of candidates shows that an aromatic residue at the position of Y102 in CYP6D1 is conserved across all candidates (Figure 4.32, A). The position of F115 is conserved as aromatic in most but not all of the candidates. However, CYP6A36 that is linked to resistance shows a substitution for an aliphatic residue while CYP6A37 which is not linked to metabolism shows conservation of both aromatic residues. The known metabolisers tend to lack aromatic residues in the FG loop and contain a polar residue. Most of the candidates are missing aromatic residues that could align with F212 of CYP6Z2 and restrict metabolism (Figure 4.32, B). The P450s identified as candidates by over expression are similar to the fingerprint and may be suggested for further experimental characterisation.
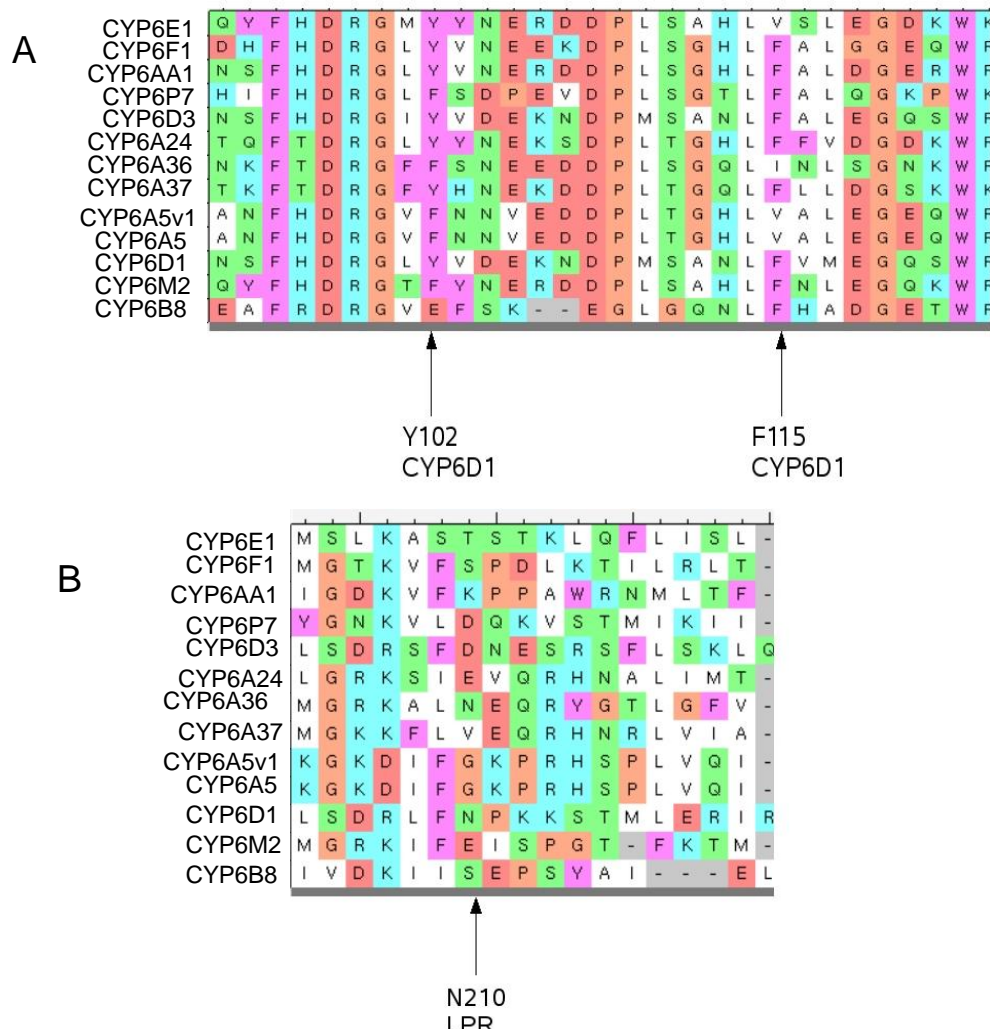
Figure 4.32 (A) Multiple alignment of the BC loops of P450 candidates. (B) Alignment of the FG loops of candidates.

4.9 Conclusion

The metabolism pathway of CYP6M2 has been experimentally determined and the regiospecificities of metabolism can be explained by the homology model produced in this study. The initial metabolism of the native deltamethrin involved a major pathway to metabolise the 4' and a minor metabolite on the trans methyl. This regiospecificity is recreated in the model and the preference for the 4' metabolite can be explained by its greater reactivity. Similarly reasons for the further successive reactions can also be rationalised.

A number of P450s have been identified as being involved in pyrethroid resistance and some have had their activities characterised. Homology models of both pyrethroid

metabolisers and non-metabolisers have been produced to identify commonalities between metabolisers and differences to the non-metabolisers. The modelling suggests that the metabolisers may have common aromatic or hydrophobic interactions with pyrethroids, while the non-metabolisers are diverse and may have a number of factors preventing metabolism.

The commonalities between metabolisers were used to suggest a 'fingerprint' that could be used to predict metabolism in other P450s. This 'fingerprint' primarily consists of one or more aromatic residue in the active site that may act as π-anchors for aromatic substrates such as pyrethroids. Aromatic interactions appear consistently in pyrethroid metabolising enzymes across P450 clades and across species. In most metabolisers identified, these aromatic residues occur on the BC loop, but some non typical metabolisers have aromatic residues on the I helix or FG loop.

Non metabolisers are diverse, some resemble the metabolisers and share active site conservations, but have other factors preventing productive binding, while other non-metabolisers are lacking in the residues conserved in the metabolisers.

5. Substrate specificities

5.0 Preface

This chapter covers studies to understand the substrate preferences of insect CYP6 pyrethroid metabolisers. Their activity towards substrates, fluorescent markers and inhibitors may give some insight into the structure of the active site.

5.1 Fluorescent markers

Resorufins and luciferins are fluorescent markers used to determine P450 metabolism. Both markers consist of a fluorescent body with an alkyl chain bonded to an oxygen. In this form the marker does not fluoresce but when the oxygen is dealkylated, the body is released and becomes fluorescent. The alkyl chain can be released by the hydroxylation of the carbon adjacent to the oxygen to produce an unstable hemiacetal or hemiketal that decomposes to cleave the bond between the carbon and oxygen and yield the O-dealkylated body and an aldehyde or ketone chain.

5.2 Experimental results

Six luciferin compounds were tested (Figure 5.1) with CYPs 6M2, 6P3, 6P9 and 6Z2 and their activities determined (McLaughlin, unpublished). It is assumed that metabolism needs to occur on the carbon next to the oxygen to break the C-O bond and release the fluorescent luciferin.
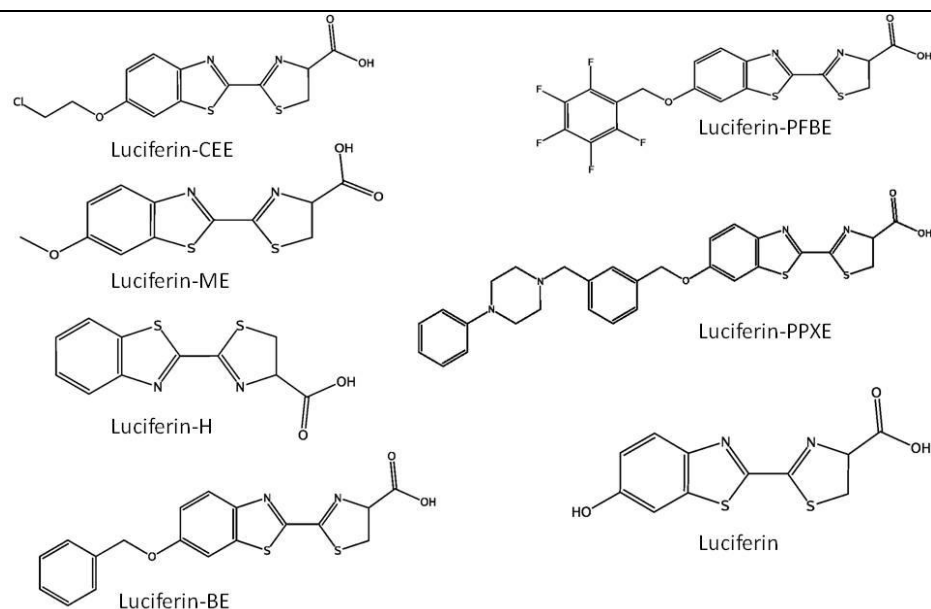
Figure 5.1 The luciferin compounds tested: 6′-deoxyluciferin (-H), luciferin 6′-methyl ether (-ME), luciferin 6′-chloroethyl ether (-CEE), luciferin 6′ benzyl ether (-BE), luciferin 6′-pentafluorobenzyl ether (-PFBE) and luciferin-6' phenylpiperazinylyl ether (-PPXE)
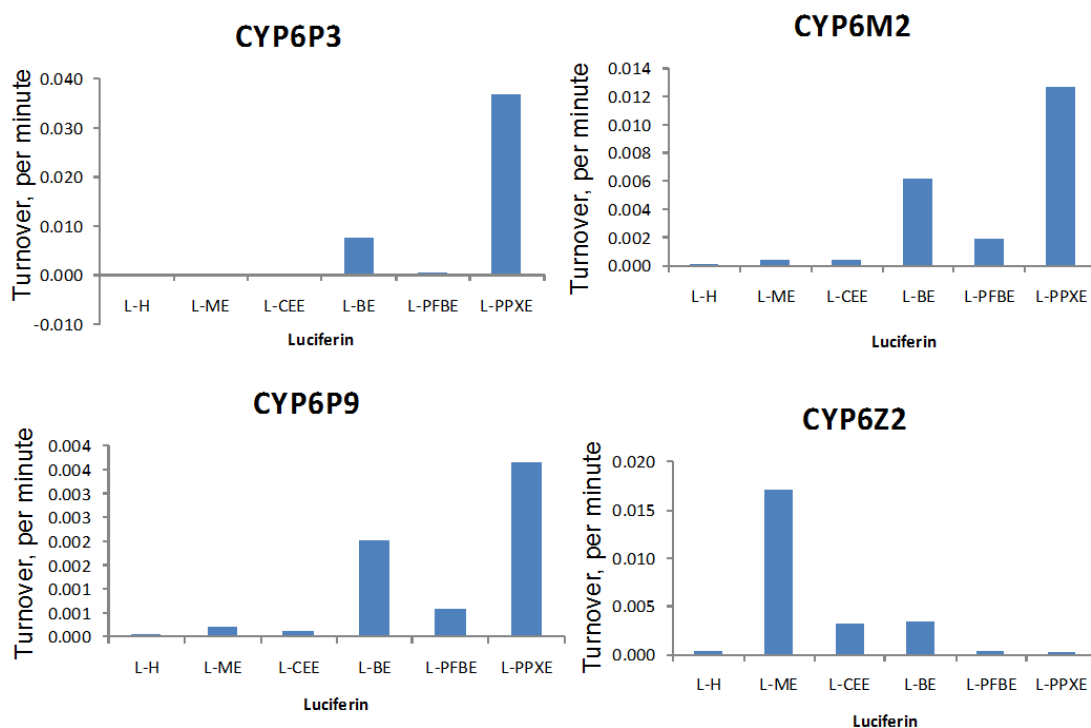
Figure 5.2 Experimentally determined activity towards luciferins by CYPs 6P3, 6P9, 6M2 and 6Z2 (McLaughlin, unpublished).

CYPs 6P3, 6P9 and 6M2 all prefer to metabolise the high molecular weight compounds -BE and -PPXE while CYP6Z2 prefers to metabolise the low weight compound -ME (Figure 5.2). A comparison of the active sites of these enzymes and an analysis of the docking poses of the ligands may be used to explain the observed substrate preferences.

5.3 Active site comparison

5.3.1 SRS5

In SRS5 of all the P450s studied, an aliphatic residue projects into the site that aligns with A370 in the CYP3A4 (1TQN) template, this is conserved as a V372 in CYP6M2 and CYP6P9/3 but in CYP6Z2 this is substituted for a Leu which reduces the size of the active site in CYP6Z2. There may be structural differences between CYP6M2 and CYP6P3/9 that could affect substrate binding. In all three proteins a Val projects into the active site from SRS5. Although this is conserved, residues either side of this differ. In CYP6M2, V372 is flanked on either side by prolines, while in CYP6P3 and CYP6P9 the Val is flanked by a Pro only on the N terminal side (Figure 5.3). Proline can have an effect on the preceding residue as the bulk of the pyrrolidine ring restricts the conformational space. The space available for the preceding residue is restricted by steric conflicts between the Cδ attached to the imide nitrogen and the NH and Cβ of the preceding residue (MacArthur and Thornton, 1991).

The presence of preceding prolines in CYP6M2 and CYP6Z2 could affect backbone dihedrals and sidechain rotamers, and this is seen in the models. An overlay of CYP6M2 and CYP6P3 shows that the backbone conformation in CYP6P3 cannot occur in CYP6M2 due to steric restrictions between the proline and the carbonyl oxygen. The backbone position of CYP6P3 and CYP6P9 differs from that of CYP6M2 and this affects the sidechain conformation of the Val. The methyls of the Val project to a greater degree into the active site of CYP6M2, while in CYP6P3/9 they project to the rear of the site (Figure 5.4). CYP6Z2 has a similar backbone conformation as CYP6M2 and the Leu projects into the active site.

Figure 5.3 SRS5 region of an alignment of CYPs 6M2, 6P3, 6P9 and 6Z2.
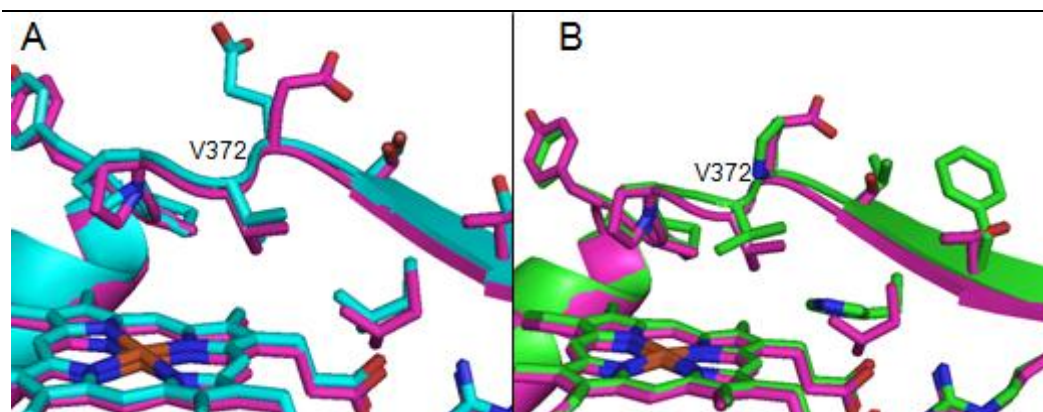
Figure 5.4 A structural alignment of the SRS5 region of CYP6P3 (Pink), with (A) CYP6P9 (blue) and (B) CYP6M2 (green) the Val of CYP6M2 projects further into the active site. The Val aligned with CYP6M2 V372 is identified.

5.3.2 FG loop

In the CYP3A4 crystal structure 1TQN, a channel through the FG loop is closed off by a network of aromatic residues (Yano et al., 2004). Aside from a conserved aromatic residue aligning with F108 in the BC loop, these tend to be substituted in CYP6P3 and CYP 6M2 but both CYP6Z2 and CYP6P9 have aromatic residues in the FG loop that align with members of the aromatic network in CYP3A4 (Table 5.1). In CYP6P9 one of these residues, F216, aligns with a member of this network and is in a position that could affect access through access channel 3 (Figure 5.5).

Table 5.1 Substitutions of residues closing the FG loop in CYP3A4.

P450	Residues				
CYP3A4[1]	F108	F213	F215	F241	F304
CYP6P3	F110	L216	E/L216	T242	L313
CYP6P9	F110	F216	L218	T242	L313
CYP6Z2[2]	Y102	R210	F212	L236	I298
CYP6Z2[3]	Y102	N211	F212	L236	I298
CYP6M2	F108	I215	I215/S216	T240	L303

[1] residues closing the channel through the FG loop in the CYP3A4 structure 1TQN, determined by Yano et al. (2004). [2] model produced here. [3] model produced by McLaughlin et al. (2008).
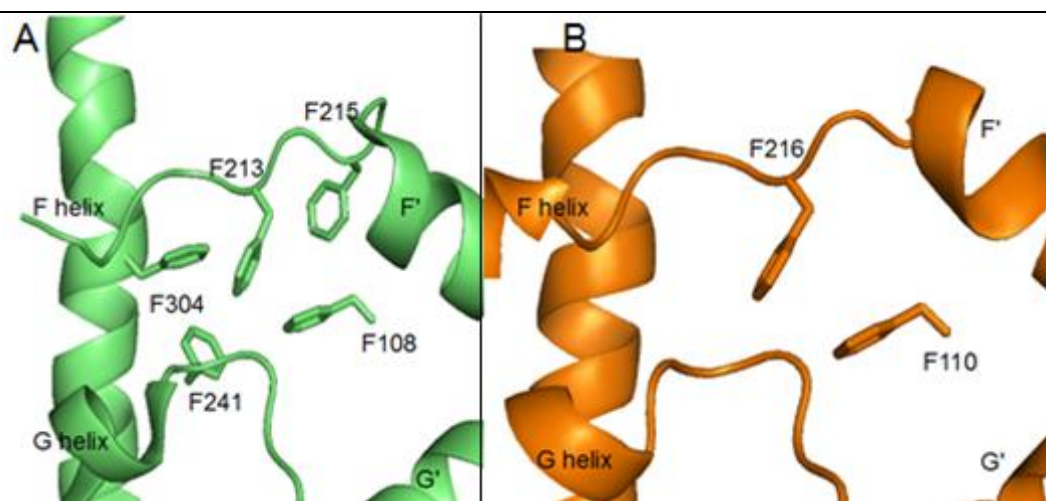
Figure 5.5 (A) The FG loop position of the aromatic network in CYP3A4, (B) position of aromatic network in CYP6P9. In CYP6P9, the F216 in the FG loop stacks with F110 of the BC loop.

5.3.3 BC loop

In the BC loop all four proteins have conserved aromatic residues but CYP6P9/3 and CYP6M2 differ from CYP6Z2 in that they have an additional His while in CYP6Z2 this is substituted for a Asn. As this Asn can form H-bonds with luciferin, it may have an affect on specificity.

5.4 Luciferin Dockings

5.4.1 -H, -ME, -CEE

In experiments CYP6M2, CYP6P3/9 show no activity for the small luciferins -H, -ME, and -CEE while CYP6Z2 shows activity for -ME and -CEE. In the dockings, while productive binding modes occur for -ME and -CEE for all proteins, the scores for these tend to be higher for CYP6Z2 due to improved vdW interactions (Table 5.2). There may be improved vdW interactions between the small luciferins and the smaller CYP6Z2 active site, and poorer packing in the larger active sites of CYP6M2, CYP6P3 and

174

CYP6P9 that could account for the activity profile. -H is not well metabolised in any P450 tested and the scores for -H are low in all proteins although the highest ranked modes are productive modes. In CYP6Z2, -ME is metabolised more efficiently than -CEE. The first ranked dockings of both allow metabolism but the preference is not explained as -CEE has a higher score than -ME, but non productive modes of -CEE show a steric screening of the site of metabolism that could not occur with -ME. All proteins except CYP6P9 bind -ME, -CEE and -H in a similar mode. The luciferin is perpendicular to the heme and lies between the BC loop Phe and the I helix (Figure 5.6, B). In his mode part of the luciferin lies within access channel 3. In CYP6P9 as channel 3 is closed and this mode is not possible, -ME binds in a different orientation between V380 and F123 (Figure 5.6, A).

Table 5.2 ChemScores (kJ/mol) of the productive modes of luciferins.

P450	ligand	rnkScore	ΔG	S(hbond)	S(metal)	S(lipo)	H(rot)	ΔE(clash)	ΔE(int)
6P3	ME	1	33.9	-35.1	0.9	0.0	264.9	1.8	0.5	0.6
	CEE	1	38.2	-39.0	0.9	0.0	307.1	2.1	0.3	0.6
	H	1	34.9	-35.4	0.9	0.0	260.6	1.5	0.3	0.0
6P9	ME	1	34.2	-35.3	2.8	0.0	215.6	1.8	0.2	0.8
	CEE	5	36.6	-38.5	2.9	0.0	246.9	2.1	0.9	0.9
	H	1	37.9	-38.5	2.7	0.0	238.2	1.5	0.5	0.2
6Z2	ME	1	38.3	-38.9	1.1	0.0	293.8	1.8	0.2	0.4
	CEE	1	42.5	-43.4	1.2	0.0	337.0	2.1	0.1	0.8
	H	9	35.9	-36.6	1.0	0.0	270.3	1.5	0.1	0.6
6M2	ME	1	33.8	-42.4	1.7	0.0	306.6	1.8	8.2	0.4
	CEE	1	38.8	-45.8	1.6	0.0	344.4	2.1	6.7	0.4
	H	1	33.7	-41.0	1.8	0.0	283.9	1.5	7.1	0.3
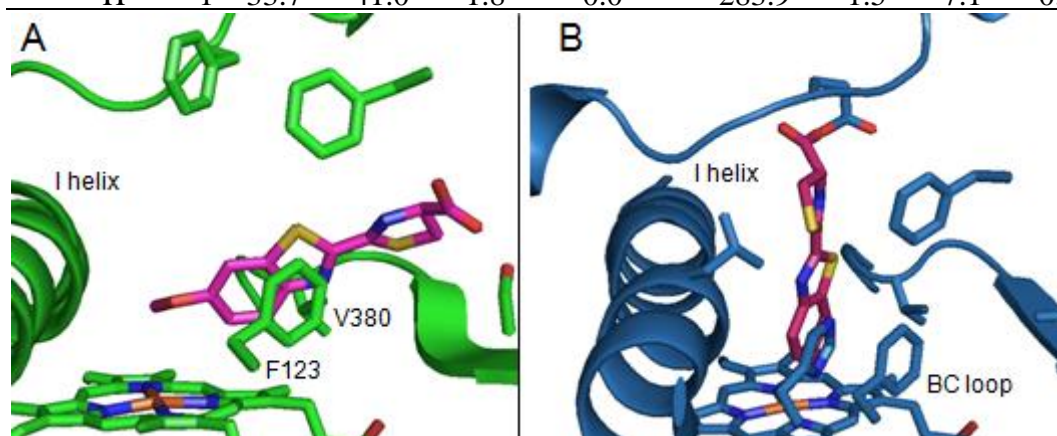
Figure 5.6 CYP6P9 (A) and CYP6P3 (B) binding -ME showing differing orientations.

5.4.2 -PPXE

-PPXE is the largest of the luciferins tested. CYP6P3 shows the greatest activity towards this while CYP6M2 and CYP6P9 show reduced activity. CYP6Z2 shows no activity, this is not reflected in the docking scores (Table 5.3) though may be explained

by the poses. In CYP6P3, -PPXE can bind for metabolism in a similar mode as -ME. The body of luciferin lies perpendicular to the heme. The tail packs into the active site between F123 and V380 and this mode shows a good vdW score, and despite large clashes this mode has a high score. In this mode the ligand partly lies within access channel 3 between the F and G loops (Figure 5.7, A).

Table 5.3. ChemScores (kJ/mol) of modes that allow metabolism of -PPXE.

		Score	ΔG	S(hbond)	S(metal)	S(lipo)	H(rot)	ΔE(clash)	ΔE(int)
6P3	4th	51.9	-60.8	0.0	0.0	523.7	2.3	6.4	2.6
6M2[1]	1st	52.4	-57.9	0.5	0.0	485.0	2.3	2.5	3.0
6P9[1]	1st	54.3	-66.4	1.0	0.0	542.5	2.3	9.8	2.2

[1] the docking pose placed the site of metabolism over 5 Å.
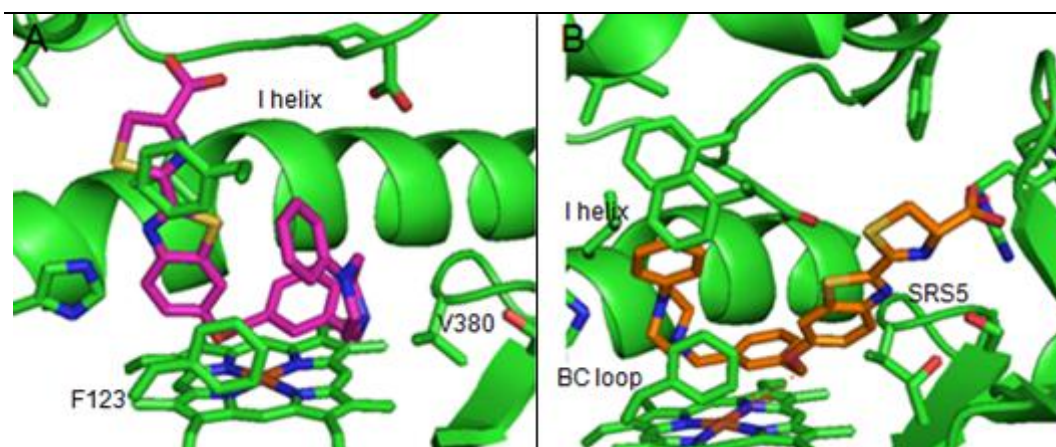
Figure 5.7 (A) Productive mode (54th ranked mode) of -PPXE in CYP6P3. (B) Unproductive binding mode of -PPXE in CYP6P9.

In CYP6P9, channel 3 may be closed due to an aromatic network as seen in CYP3A4. This affects the binding of luciferin as it cannot lie within the access channel as in CYP6P3 but instead -PPXE binds between SRS5 and the BC loop (Figure 5.7, B). In this mode the tail group of -PPXE cannot be accommodated, and the site of metabolism is distant from the Fe which could affect the efficiency of metabolism. This may explain the different activities of CYP6P3 and CYP6P9 despite sequence conservation. In CYP6M2 the positioning of Val in SRS5 may sterically restrict the accommodation of the tail and it binds in a similar mode as in CYP6P9. The site of metabolism is also distant at 5.2 Å. This may account for the lower activity in CYP6M2. In CYP6Z2, no modes were found that allow metabolism. This may be due to the smaller size of the active site compared to the other three P450s.

5.4.3 -BE, -PFBE

Luciferin-BE and -PFBE have similar structures differing only in that the hydrogens of the benzyl group are replaced by fluorines in -PFBE. Despite these similarities, all CYPs tested metabolise -BE well and -PFBE poorly. As fluorines are only slightly larger than hydrogens it is unlikely that the increase in the steric bulk of the ring that affects binding. In CYPs 6M2 6P3 and 6P9 the first ranked docking allows metabolism of -BE. -PFBE can also bind in similar positions that allow metabolism (Figure 5.8) but with a lower score due to clashes (Table 5.4). The addition of fluorines affects the positioning of the ring to alter the interaction with the BC loop Phe and this may be producing poorer vdW interactions to give lower scores. In addition, the halogens have other effects. Electron withdrawing halogen groups can stabilise an aromatic ring to oxidative attack. Fluorine is electronegative and C-F bonds are strongly polarised. Fluorine can affect the acidity or basicity of neighbouring functional groups and strengthen neighbouring bonds (Jeschke, 2010) and the effects of the fluorine may prevent metabolism by P450s.

Table 5.4 ChemScores (kJ/mol) of modes that allow metabolism.

			Score	ΔG	S(hbond)	S(metal)	S(lipo)	H(rot)	ΔE(clash)	ΔE(int)
6P9	BE	1st	47.8	-49.2	2.8	0.0	339.1	2.1	1.2	0.2
	PFBE	6th	34.2	-40.5	1.5	0.0	313.8	2.7	0.7	5.6
6M2	BE	1st	47.1	-49.0	1.0	0.0	389.5	2.1	0.7	1.2
	PFBE	1st	36.2	-44.3	1.0	0.0	361.9	2.7	6.2	1.8
6P3	BE	1st	47.4	-50.0	0.9	0.0	399.51	2.1	1.5	1.2
	PFBE	1st	37.2	-46.4	0.7	0.0	388.09	2.7	5.4	3.8
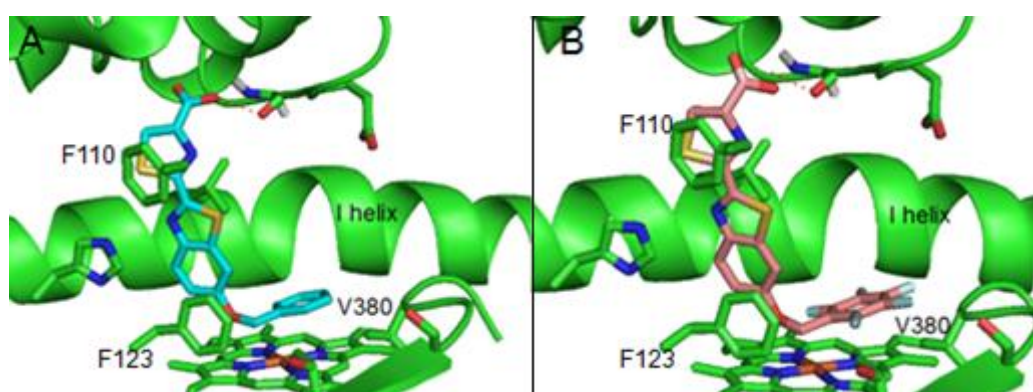
Figure 5.8 CYP6P3 binding -BE (A) and -PFBE (B).

With deltamethrin both F110 and F123 are involved in π-stacking, in the luciferins, F123 is involved in stacking while F110 is either distant or forms hydrophobic contacts. While either F123 or F110 could act as a π-anchor for deltamethrin, F123 may act as a

π-anchor for the luciferins.

5.5 Resorufin metabolism in CYP6D1

The activity of CYP6D1 towards a number of substrates was determined by Wheelock and Scott (1992) by treating housefly microsomes with anti 6D1lpr. CYP6D1 was found to be the major cytochrome involved in metabolising methoxyresorufin and ethoxyresorufin. Ethoxycoumarin activity was only inhibited at the highest concentration of antiserum and may only be a poor substrate for CYP6D1 while pentoxyresorufin was not identified as a substrate (Figure 5.9) (Wheelock and Scott, 1992).
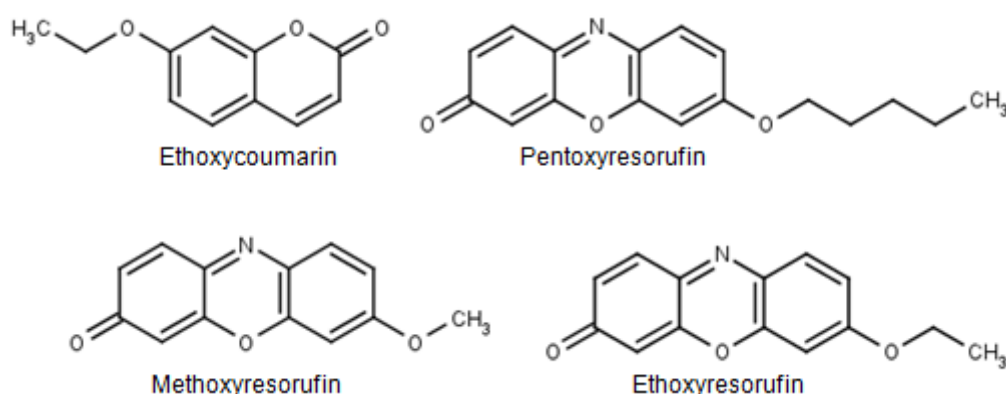
Figure 5.9 Structures of ligands tested in CYP6D1.

5.5.1 Dockings

Methoxyresorufin and ethoxyresorufin can both bind in positions that allow O dealkylation. In the docking modes of methoxy- and ethoxyresorufin, a face on face stacking may be possible with F115, and an edge on face stacking could occur with Y102 (Figure 5.10, A and B). Similarly, in the best ranked mode of ethoxycoumarin there is a face on face π-stacking between F115 and the coumarin aromatic ring (Figure 5.10, C), there is also an H-bond with N113 of the BC loop and these interactions position the ligand for metabolism. However, the score is lower than the resorufins (Table 5.5), as it is missing the second aromatic ring that occurs in the resorufins. It makes fewer contacts with the protein that may account for its lower metabolism in CYP6D1. The preference for the resorufins over the coumarin may be due to the additional aromatic group in the resorufins. With the resorufins a larger or more stable aromatic network may be formed with F115 and Y102. In the model these ligands also have higher ChemScores than ethoxycoumarin, indicating a stronger binding. Alternatively, with the absence of a second ring, the coumarin may stack with F115 in a position that would not allow metabolism. Pentoxyresorufin does not dock in a position

that would allow metabolism. The large alkoxy tail may sterically prevent a docking close to the heme or may sterically screen the site of metabolism (Figure 5.10, D).

In CYP6D1 the preference for alkoxyresorufins decreases as the alkoxy substituent increases in length. The preference may be due to the steric restrictions of the active site, in the model F115 and L388 project into the site. A long hydrocarbon tail may not be accommodated that may prevent metabolism of pentoxyresorufin while resorufins with shorter tails may be accommodated and metabolised.

Table 5.5. ChemScores (kJ/mol) of CYP6D1 substrates.

	Score	ΔG	S(hbond)	S(metal)	S(lipo)	H(rot)	ΔE(clash)	ΔE(int)
ethoxycoumarin	28.5	-30.9	1.9	0.0	186.5	1.1	2.1	0.3
methoxyresorufin	33.6	-35.6	1.9	0.0	225.9	1.0	2.0	0.0
ethoxyresorufin	34.6	-35.1	0.0	0.0	277.3	1.1	0.4	0.0
pentoxyresorufin	38.6	-40.0	0.9	0.0	298.7	1.3	1.2	0.2
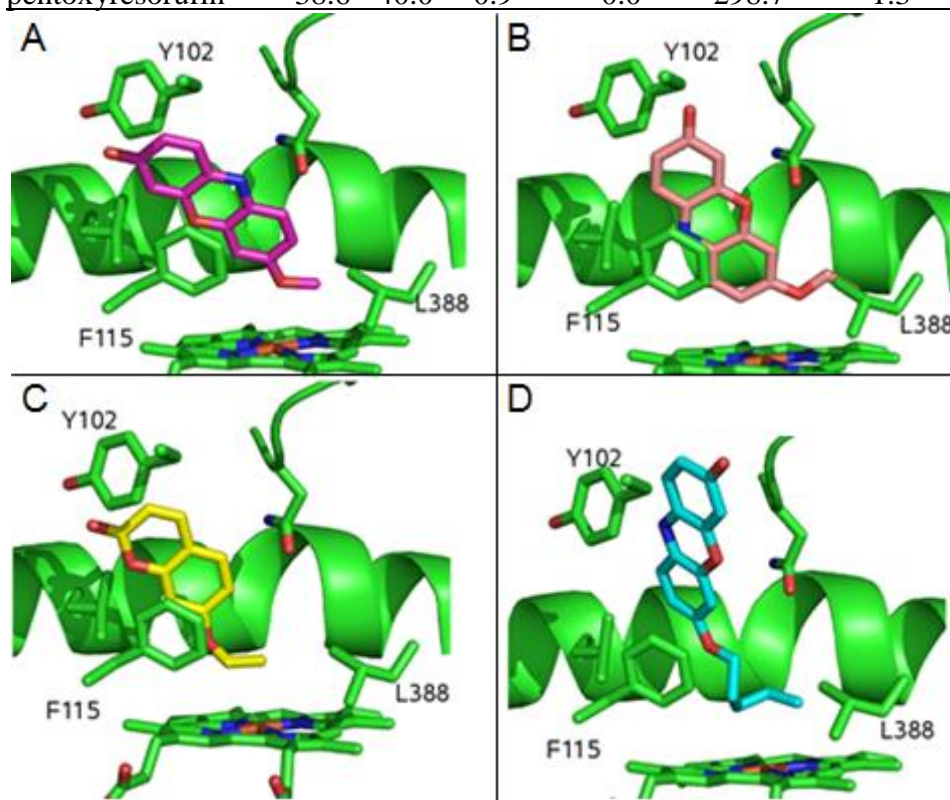
Figure 5.10 The best ranked docking modes of methoxyresorufin (A), ethoxyresorufin (B), ethoxycoumarin (C) and pentoxyresorufin (D) bound in CYP6D1.

5.5.2 Substrate preference

A number of hypotheses have been previously suggested to explain the preference of P450s towards particular fluorescent markers such as resorufins that differ in their chains. The rate of alkoxyresorufin dealkylation has been shown to be determined by three factors: substrate binding affinity, substrate-binding induced reduction of the heme

by water displacement, and binding orientation and mobility with respect to the heme. In addition, other factors such as substrate access/exit and electron transfer may also have a role.

The steric bulk of residues at equivalent positions to L388 in CYP6D1 have previously been shown to affect metabolism of such substrates. CYP1A1 and CYP1A2 have been shown to differ in their substrate preferences for alkoxyresorufins, with CYP1A1 metabolising ethoxyresorufin and CYP1A2 metabolising methoxyresorufin. These enzymes differ in the SRS5 position corresponding to L388 in CYP6D1, with CYP1A1 having a Val and CYP1A2 having a Leu and the size of the aliphatic residue affects metabolism of resorufins with differing lengths of alkoxy tail. Reciprocal mutations of this one residue have been shown to inter-convert the function of these enzymes. CYP1A2 L382V showed reduced MROD activity but unchanged EROD activity. From a 5 ps MD, it was suggested that in this case methoxyresorufin was less stable than in the wild type as the increase in active site volume occurs at the position where the methyl group would bind. A change in the hydrophobic pocket destabilises an interaction allowing less productive orientations to occur. A L382A mutation showed that this occurred for both resorufins as the pocket is further expanded (Liu et al., 2004).

In CYP1A1, V382 in SRS5 is in a position to interact with resorufin substrates. Liu et al. (2003) found that in the wild type CYP1A1, 7-methoxy, 7-ethoxy and 7-pentoxy bound in similar positions with the planar part of the molecules superimposed and the alkoxy chain in the space close to the heme and V382. In the wild type, activity was highest towards 7-ethoxy and lowest to 7-pentoxy resorufin. While these resorufins could all bind without steric hindrance in CYP1A1, during a 5 ps MD simulation only 7-ethoxyresorufin remained 3 Å from the heme, 7-pentoxy and 7-methoxy moved away from the site of metabolism. In addition the site of metabolism in 7-pentoxy could become sterically screened by the other atoms of the alkoxy chain. When V382 is mutated to an Ala, activities for 7-methoxy and 7-ethoxy fell by 10 fold while 7-pentoxy activity increased 2 fold. The V382A mutation enlarged the active site leading to increased mobility of 7-ethoxy and 7-methoxy resorufin, with the oxidation site of 7-ethoxy remaining further from the heme. In contrast, the 7-pentoxyresorufin could now approach the heme allowing metabolism. When V382 is substituted for a Leu there was a decrease in activity for all substrates. The position of the smaller resorufins were dramatically affected with increased distances from the heme. The increased bulk of the

side chain pushed the resorufins out of productive binding orientations.

A similar situation has been identified in CYP2B1 by Kobayashi et al. (1998) with alkoxycoumarins with chains of differing length. While other mutations had little effect, mutation of V363 in SRS5 produced alterations in activity towards alkoxycoumarins. The wild type showed a particular preference for length of alkoxy chain:

C2 > C4 > C3 > C1 > C5 > C6 = C7

They found that the C2 could be docked into the active site with no vdW overlaps, but increasing the length of the chain caused overlaps to occur with V363. The decreased activity towards C1 was explained by the increased mobility of the substrate, while C2 could fill the active site optimally.

Mutations to Ala and Leu alter preference for chain length. The V363A mutation increases the volume of the active site, allowing C4 to dock in an orientation that allows metabolism (Figure 5.11, A). In addition, it showed a greater activity towards 7-butoxycoumarin analogues with branched alkyl chains, and they were a better fit to the active site than in the wild type. Smaller substrates could exhibit higher mobility while larger substrates have vdW overlaps lowering activity.

The mutation to Leu was similar to the wild type in having a high activity towards C2, but unlike the wild type, activity towards chains greater than 2 atoms was reduced (Figure 5.11, A), and activity towards analogues with branched chains was undetectable. It was suggested that the larger Leu prevented the substrates form binding in productive orientations, and that this was indicated by the appearance of vdW overlaps. Other residues were identified as involved in resorufin binding, the mutation F206L in the BC loop was associated with very low rates of activity (Figure 5.11, B). It was suggested substituting the Phe for a smaller more flexible Leu could increase the size of the active site and lead to substrate mobility. This was indicated with a decrease in coupling of NADPH to product formation and uncoupling to water (Kobayashi et al., 1998).
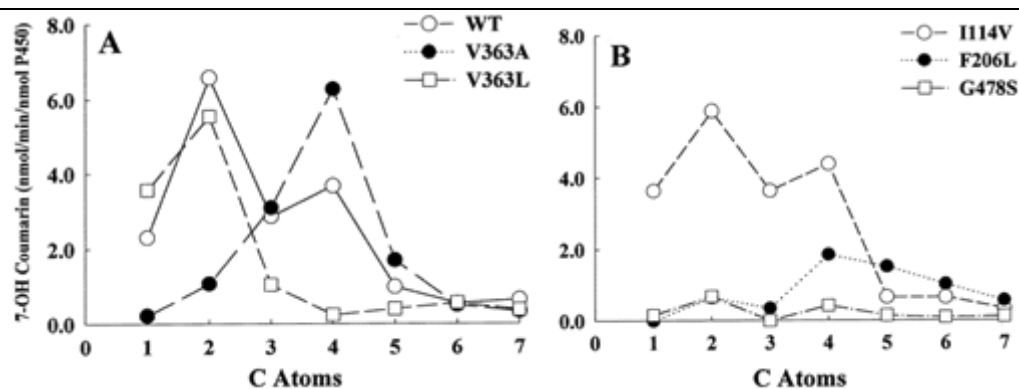
Figure 5.11 Taken from Kobayashi et al. (1998). The preference for coumarins with varying chain length in CYP2B1 wild type and mutants.

5.5.3 Luciferin preference summary

The activity profiles of the CYP6s could be explained by steric and vdW interactions. There may be optimal filling of the active site by the preferred ligands, steric clashes with larger ligands and a greater mobility of smaller ligands. The P450s with larger active sites prefer larger substrates while the smaller active sites prefer the smaller substrates. CYP6M2, CYP6P3/9 have a Val in SRS5 and a preference for substrates with large chains. The lack of activity of CYPs 6P3/9 and 6M2 towards the smaller substrates be explained by poor vdW interactions or increased mobility, while the higher activity towards the larger -PPXE may be explained by greater vdW interactions. The bulk of L388 in CYP6D1 may produce the preference for methoxy- over ethoxyresorufin, and the Leu in CYP6Z2 may produce the preference for the small chain luciferins.

This study has shown that the candidate P450s have particular preferences for fluorescent probes, this preference will be used in assays to study the binding of inhibitors and to design acetylene probes specific to the active sites of these enzymes. Luciferins can also be used as selective probes of P450 architecture. They are sensitive to mutagenesis of the active site and may be used to study the function of a particular residue.

5.6 Acetylene Inhibitors

5.6.1 Aryl acetylenes as suicide inhibitors

Acetylenes are P450 suicide inhibitors that cause a time dependant destruction of P450s. Suicide inhibition takes place by a regiospecific oxidation of the terminal carbon of the

triple bond to generate a ketene intermediate. The oxygen of this ketene is derived from molecular oxygen and so is formed by transfer of enzymatically activated ferryl oxygen (Chan et al., 1993).

There are two mechanisms of inhibition by these acetylenes. N-alkylation of the heme can occur leading to the loss of the heme by the formation of an adduct where the acetylene is bound to the nitrogen as a 2-oxoalkyl group. The second mechanism involves modification of the protein, nucleophilic amino acids within the P450 can react with the ketenes to form a covalent bond. One of the characteristic reactions of ketenes is their reaction with nucleophiles at the electrophilic β-carbon. The insertion of an oxygen at the terminal carbon results in protein modification while insertion at the internal carbon results in heme modification. Alternatively, the ketene can be hydrolysed to form carboxylic acid and fail to inhibit the enzyme (Figure 5.12). The conserved I helix threonine has been proposed as the site of covalent binding to the protein. A modified CYP2B4 lacking the T302 is inactivated more slowly than the wild type (Roberts et al., 1996).
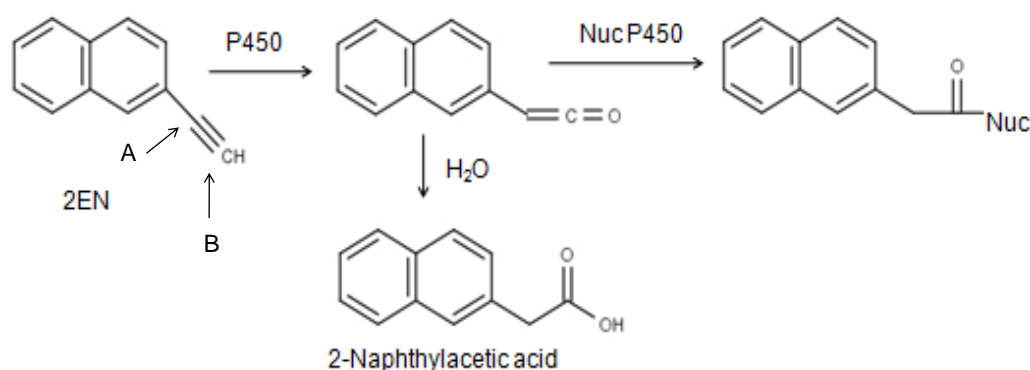
Figure 5.12 Taken from Roberts et al. (1996). Oxidation of the acetylene can occur at the internal (A) carbon or external or terminal carbon (B). The insertion of oxygen at the terminal carbon can lead to the modification of nucleophilic amino acids. Alternatively it can react with water to form an acid.

5.6.2 Substrate specificity

Acetylenes can act as selective inhibitors. Hopkins et al. (1992) found that the shape of the aromatic ring system and the placement of the triple bond of these compounds affect the type and potency of inhibition. Suicide inhibition varied between P450s with some inhibitors selective for certain enzymes. 1EP, 2EPh, 3EPh 3PPh 2PN and 4EAA only caused suicide inhibition in CYP1A1 but not CYP1A2, CYP2B1 or CYP2B2. The

enzyme system was also found to have a large affect on the type of inhibition. Some compounds were effective suicide inhibitors in purified rat CYP1A2 but were less effective inhibitors in purified rabbit CYP1A2. As the suicide inhibition reaction is dependent on regiospecific oxidation, the correct orientation of the ligand within the active site is needed, producing a selectivity of inhibition for different enzymes.

The size of the aromatic ring system also affected inhibition. Compounds with large ring systems such as biphenyl, pyrene and phenanthrene tended to be suicide inhibitors in the CYP1A family, while those with smaller aryl rings were inhibitors of CYP2B1. Hopkins et al. (1992) reported structure activity relationships for inhibitors of CYP1A1 and CYP2B1. They showed that in addition to size and shape, the positioning of the carbon-carbon triple bond is important in selectivity of suicide inhibition. In CYP2B1, only 1EN, 2EN, 1EA, 9EPh were found to be suicide inhibitors (Figure 5.13). It was suggested that in CYP2B1 aryl acetylenes must be small and compact for the proper orientation of the ethynyl group. 1EP and 2EP were considered too large to bind in an orientation that could lead to suicide inhibition. 1EP was a better inhibitor of CYP1A1 because the larger aromatic system produced a more favourable binding interaction to the large active site, but 2EP did not produce suicide inhibition indicating that the correct placement of the acetylene group is needed.
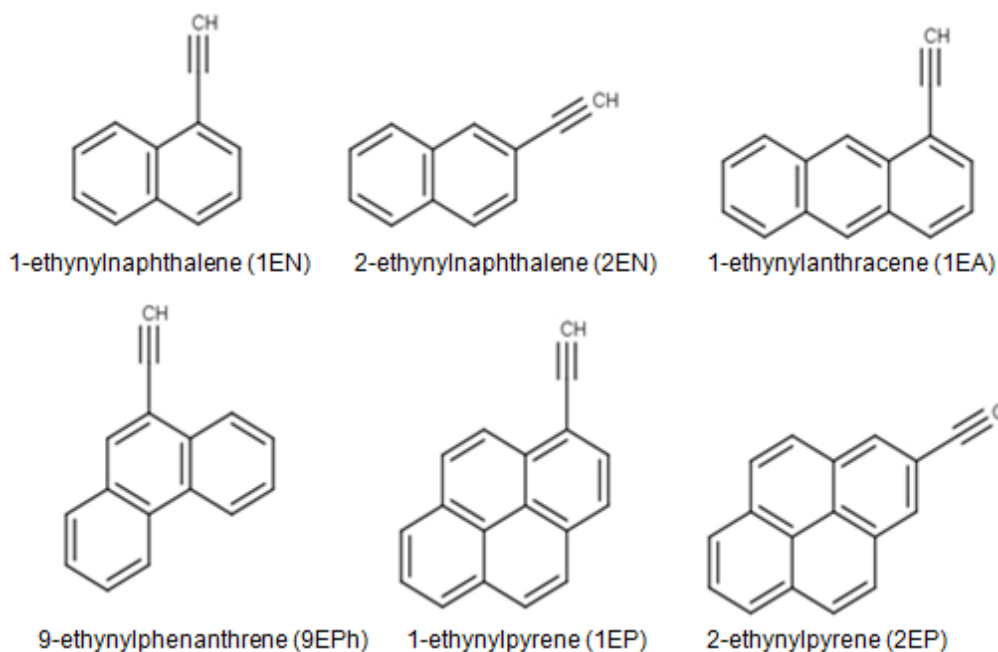
Figure 5.13 Taken from Hopkins et al. (1992). The structures of suicide inhibitors of P450s.

5.6.3 CYP6P3 acetylene inhibitors

A number of acetylene inhibitors were tested in CYP6P3 (Figure 5.14, Table 5.6). The experimental results show that some acetylene ligands act as mechanism inhibitors while others do not. The mechanism inhibitors were found to covalently bond to the protein rather than destroy the heme. A number of the smaller ligands (I1-I4 and I9) dock in modes that are inconsistent with the experimental results. This may be a limitation of the docking algorithm in placing small hydrophobic molecules that make few contacts with the protein. I1 has been found to be a mechanism inhibitor that does not destroy the heme but binds with the internal carbon close to the heme, which is inconsistent. I2 does not show mechanism inhibition but the best scoring modes place the internal and external carbon above the heme. I3 is a mechanism inhibitor and while the best ranked mode places the internal carbon close to the heme, no modes occur that place the external carbon for metabolism. I4 is not a mechanism inhibitor but the best ranked mode places the internal carbon above the heme (Figure 5.15).
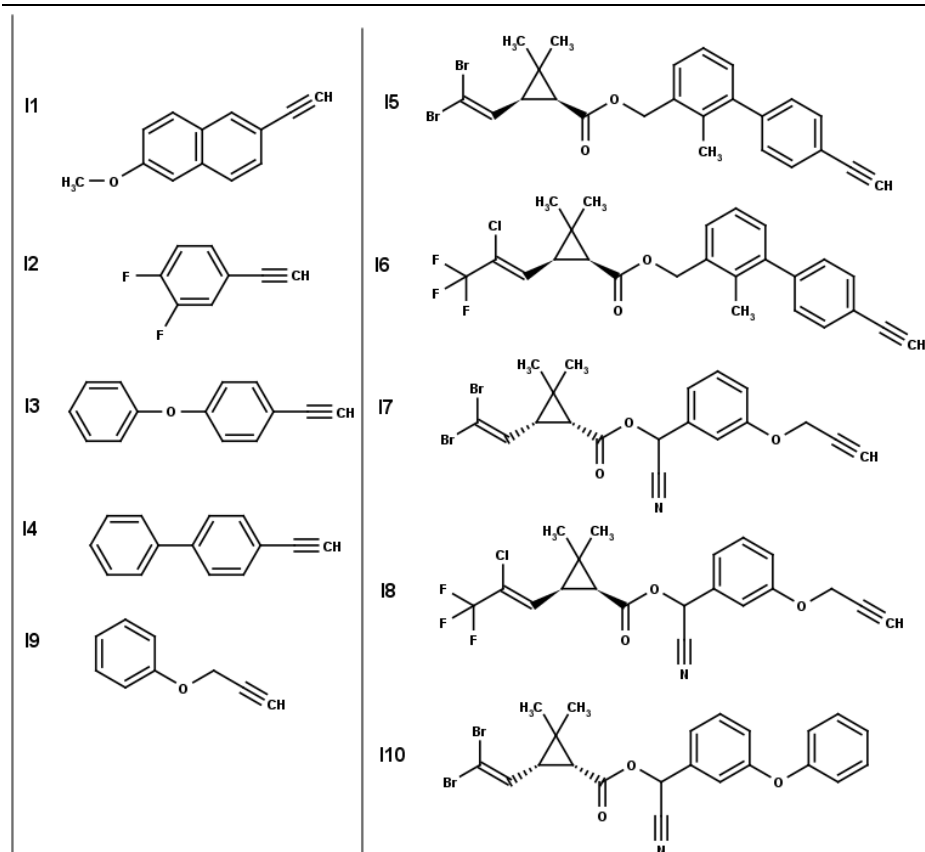
Figure 5.14 The acetylene inhibitors tested in CYP6P3. I10 is deltamethrin used as a comparison.
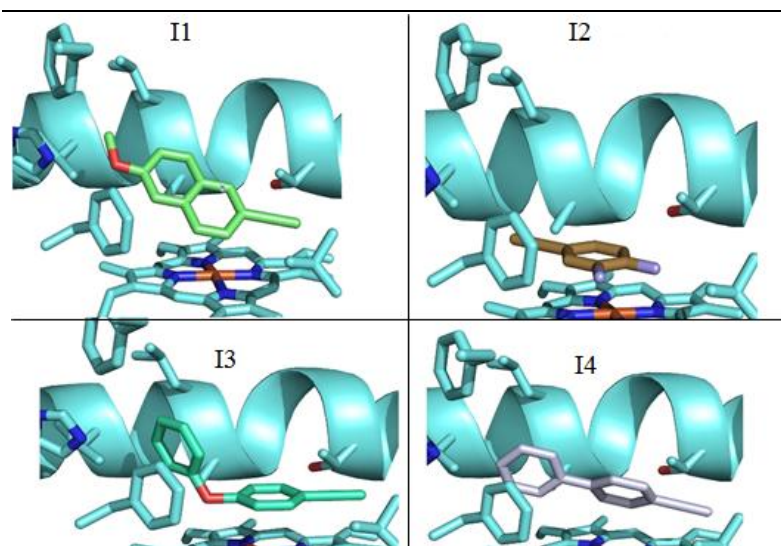
Figure 5.15 The docking modes that allow metabolism of the acetylene group I1-4 in CYP6P3.

Table 5.6 The remaining activity of CYP6P3 towards L-PPXE substrate in the presence of acetylene inhibitors (McLaughlin, unpublished). Pink highlights the inhibitors that act as suicide inhibitors as they decrease activity in the presence of NADPH.

Compounds	Log p value	Concentration uM	Remaining activity after 30min preincubation - /+NADPH		NADPH Factor
			-NADPH	+NADPH	
I1		10uM	70	22	3
	3.06	1uM	147	56	3
I2		10 uM	104	114	1
	2.51	1 uM	109	120	1
I3		10 uM	64	6	11
	3.73	1 uM	66	69	1
I4		10 uM	84	92	1
	3.87	1 uM	88	101	1
I5		10 uM	146	106	1
	7.36	1 uM	78	95	1
I6		10 uM	70	52	1
	6.68	1 uM	88	65	1
I7		10 uM	1	0	0
	5.02	1 uM	50	15	3
I8		10 uM	3	0	0
	4.8	1 uM	21	13	2
I9		10 uM	85	21	4
	2.12	1 uM	102	126	1
I10		10 uM	37	65	1
	6.93	1 uM	110	96	1
No Inhibitor		0	100	100	1

With both I5 and I6, the planar alcohol binds perpendicular to the heme, stacking with F110 and F123. The acid group occupies the cavity above the heme and forms hydrophobic contacts with F123, V380 and T318. In this position, the acetylene group is positioned distant from the heme. No productive modes were found for either I5 or I6 which is consistent with their lack of mechanism inhibition. I5 and I6 are found to be weak competitive inhibitors. This may be due to an inability to access the active site due to their higher logP or their hook shape.

I7 and I8 were shown to be both mechanism inhibitors and to be strong competitive inhibitors. For I7 the best ranked mode does not allow metabolism on the acetylene but binds with the acid above the heme. Lower scoring modes occur that allow metabolism on the acetylene group external carbon, and that position the cyano group for coordination (Table 5.7 and Figure 5.16). These modes may explain both the high levels of inhibition in the absence of NADPH and mechanism inhibition in the presence of NADPH. In a free docking, I8 binds for coordination in a similar mode as I7, but no modes are found that allow metabolism on the acetylene. I7 and I8 are good inhibitors in the presence and absence of NADPH. This may be because they contain a cyano group and their strong inhibition in both the presence and absence of NADPH may be due to type II coordination.

Table 5.7 ChemScores (kJ/mol) of modes of binding for I7 and I8 in CYP6P3, a mode positioning the acid above the heme is the best ranked mode. Modes placing the external carbon and cyano group also occur.

	Score	ΔG	S(hbond)	S(metal)	S(lipo)	H(rot)	ΔE(clash)	ΔE(int)
I7 (1)	38.7	-39.5	0.0	0.0	328.3	1.7	0.2	0.7
I7 external (20)	34.8	-38.2	0.0	0.0	317.0	1.7	1.5	1.9
I7 coord (23)	34.2	-41.1	0.0	1.0	291.0	1.7	3.3	3.6
I8 (1)	33.8	-37.7	1.5	0.0	278.2	2.1	1.2	2.6
I8 coord (7)	32.9	-36.6	0.0	1.0	261.4	2.1	1.5	2.3
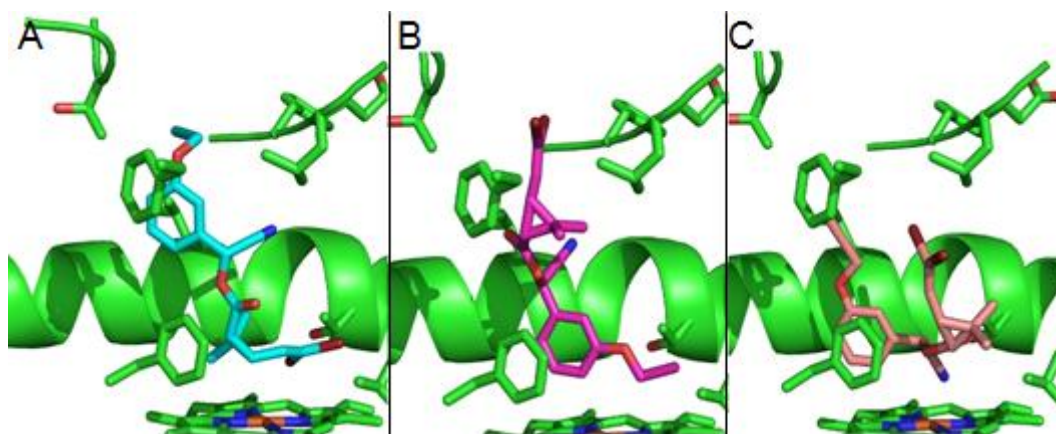
Figure 5.16 Three binding modes for I7 in CYP6P3, placing the acid above the heme (A), binding for mechanism inhibition (B), and for coordination (C).

5.6.4 Pyrethroid-like Probes

In addition to the acetylene probes tested in CYP6P3, a pyrethroid-like probe was tested in CYP6M2 and CYP6P3 (Figure 5.17, Table 5.8). CYP6P3 shows a NADPH dependant decrease in activity in the presence of NADPH while CYP6M2 does not, indicating that CYP6P3 metabolises the probe acetylene while CYP6M2 does not. The best ranked modes in both CYP6M2 and CYP6P3 do not bind productively. However, a high scoring mode of CYP6P3 is able to bind productively while a low scoring mode of CYP6M2 binds productively. This may be consistent with the experimental results in that CYP6P3 may be more likely to bind the probe for metabolism than CYP6M2. In CYP6P3 the binding mode of the probe is similar to I5 and I6 as the planar naphthalene of the probe stacks with F123 and F110 (Figure 5.18). Although the probe binds in a similar mode as I5/6, it is able to be metabolised on the acid acetylene to produce mechanism inhibition.
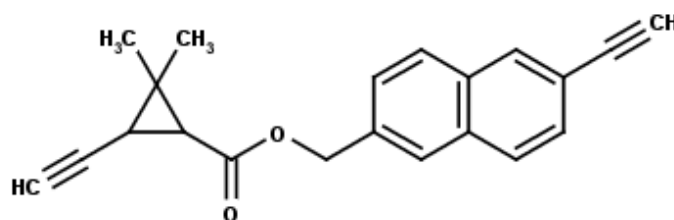
Figure 5.17 Pyrethroid-like probe.

Table 5.8 Remaining activity towards L-PPXE of CYP6M2 and CYP6P3 after treatment with different concentrations of probe (McLaughlin, unpublished).

CYP6P3	Concentration uM*	Remaining activity after 30min preincubation -/+NADPH		NADPH Factor
		-NADPH	+NADPH	
	20	70	25	2.7
	10	83	34.5	2.4
	1	93	66	1.4

CYP6M2	Concentration uM*	Remaining activity after 30min preincubation -/+NADPH		NADPH Factor
		-NADPH	+NADPH	
	20	57	42	1.4
	10	58.8	59.5	1
	1	71	80	1

<u>5.6.5 CYP6P3 vs CYP6M2</u>

The best ranked mode of CYP6M2 does not allow mechanism inhibition while a low scoring mode similar to that in CYP6P3 allows metabolism on the internal carbon. The difference in score between CYP6P3 and CYP6M2 may be due to steric differences between the active sites. In CYP6M2 the V372 projects further into the active site in a position to cause clashes with the acetylene. This is caused by the presence of two flanking prolines in CYP6M2. The presence of a proline preceding the SRS Val in CYP6M2 affects both the backbone dihedrals and sidechain rotamer so that the Val of CYP6M2 projects further into the active site. The difference in rotamer affects the steric restrictions imposed by the Val. An overlay of CYP6M2 and CYP6P3 shows that the Val in CYP6M2 is in a position to cause clashes with the ligand. Productive modes in CYP6M2 differ from those in CYP6P3 as the ligand must bind in a less favourable orientation for metabolism to occur (Figure 5.19).
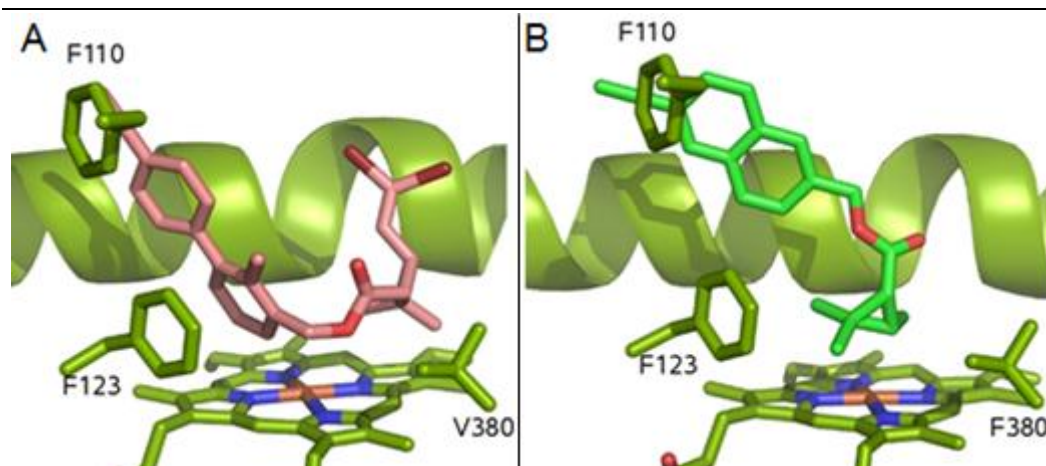
Figure 5.18 First ranked mode of I5 (A) and 3rd ranked docking of the probe (B) in
CYP6P3.
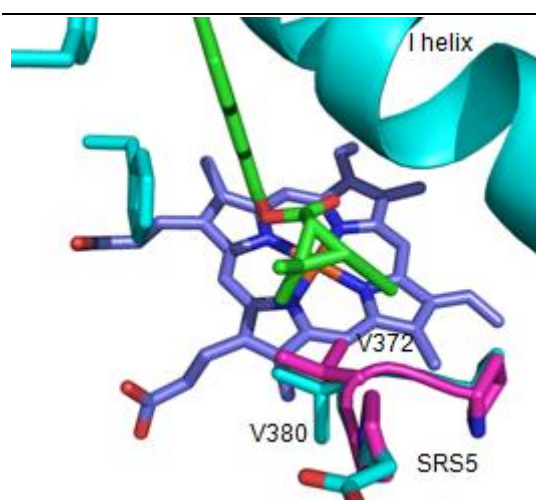
Figure 5.19 The probe bound in CYP6P3 (blue) with CYP6M2 (pink) overlaid. The
rotamer of the valine in SRS5 differs between proteins.

5.6.6 Nucleophilic target

For mechanism inhibition to occur, the inhibitor reacts with the protein, nucleophilic
amino acids such as threonine, serine and cysteine are available for attack by the ketene.
In CYP6P3, the position of the acetylene group of the probe is consistent with
modification of the I helix T308. This is consistent with previous studies showing that
covalent binding to the I helix conserved Thr occurs in CYP2B4 (Roberts et al., 1996).
If T308 is the site of attack this could be demonstrated by mutation to an Ala. To
determine if such a mutation could have any other affect on ligand binding, a T308A
model was used for docking. The T308A mutant was not found to affect the productive
mode. An overlay of the mutant and wild type shows that the ligand binds in a similar
position with a similar score. However, T308 occurs on the I helix at the electron
transfer groove and may have an effect on enzyme function.

5.7 Prediction of novel substrates- database screening

The models of CYPs have been shown to be able to explain the observed experimental results. They may also be used for receptor based virtual screening to predict novel substrates to guide further experimental studies. Virtual screening of a database is complementary to experimental high throughput screening of compounds and is used in the lead identification stage of drug discovery. Virtual screening can be either ligand based or receptor-based. For ligand based methods, information about known ligands can be used to identify other compounds with a similar structure or properties, and this can be done using similarity or substructure searching. Where there are no known ligands or where a structure of the target is available, receptor based screening can be done. This involves the explicit docking of each ligand in a database into the target to give a measure of the quality of fit. The compounds can be ranked by score to select a small subset for experimental testing.

Although compounds could be selected based on a simple ranked order, this tends not be advisable as virtual screening is inaccurate at predicting the affinity of a ligand for a receptor due to inadequacies of the scoring functions. Post-analysis strategies can be used to minimise false positives in the selected subset to give a list of true hits. A common post-analysis strategy is to use consensus scoring. After compounds are docked, a number of scoring methods are used to re-score the poses. Only compounds that score well with each scoring function are chosen. Geometric analysis is another post-analysis strategy that can involve calculating surface complementarity between protein and ligand, or can be knowledge based where particular interactions are known to be desirable (Lyne, 2002).

5.7.1 Structure based virtual screening caveats

A number of problems are known to occur with virtual screening, Kitchen et al. (2004) reviewed the caveats. Firstly docking scores scale poorly with molecular mass and the number of rotatable bonds. This is because large compounds can form many interactions with the active site and so generate better scores than small compounds. The number of rotatable bonds also has an affect because immobilisation of a rotatable bond leads to a loss of torsional or rotational entropy, torsional entropy penalties need to be correctly weighted or artefacts may occur. When a ligand binds to a protein and becomes less mobile the loss of torsional entropy can oppose the attractive force that drives the binding. If such entropy penalties are included in the scoring function, large flexible

molecules score lower than rigid molecules. It has also been shown that structure based virtual screening tends to select compounds that are biologically promiscuous and are termed 'frequent hitters' which show non-specific inhibition. This has been attributed to their hydrophobic character that favours their detection in docking simulations.

It is not just the treatment of the ligands that can affect virtual screening. The characteristics of the active site has an effect. Hydrophobic active sites produce better results than hydrophilic sites as hydrophobic sites can be approximated by shape complementarity, while electrostatic interactions are not precisely modelled. The placement of water molecules that are directly involved in binding and the rigidity of side-chains can dramatically influence the posing of ligands. And where conformational changes upon binding occur, rigid active sites are limited in their ability to predict poses (Kitchen et al., 2004).

The majority of database screening tends to be carried out using DOCK but consensus scoring using two or three independent scoring methods has been shown to outperform single scoring. It is suggested that consensus scoring is able to combine information from different scores to balance errors in single scores, and to identify 'true' ligands. However, consensus scoring has little potential if scoring functions are correlated as this could amplify errors (Kitchen et al., 2004).

In this study, Dock and GOLD were selected for consensus scoring as they use different methods to dock the ligand into the active site. GOLD uses a random/stochasic approach. Dock uses a systematic approach. DOCK and GOLD also use different scoring methods. Dock uses a forcefield method while ChemScore is an empirical free energy scoring function. While there are advantages and disadvantages with both methods, the differences between them may avoid amplification of errors.

DOCK uses a systematic approach to docking by exploring all of the degrees of freedom of the molecule. To avoid a large search space, it grows the ligand into the active site incrementally by anchoring a rigid fragment and growing the flexible sidechains one bond at a time with pruning to remove unfavourable conformations. GOLD uses a random search using a genetic algorithm. A 'chromosome' is stochastically varied. The best chromosomes are subjected to crossover and mutations to produce the next generation, but may fail to find the optimal pose. Force field based

scoring such as the DOCK energy score does not include any solvation or entropic terms and non bonded interaction cutoffs are arbitrarily chosen. Empirical scoring functions such as ChemScore are fitted to reproduce experimental data such as binding energies by summing individual uncorrelated terms, and unlike the force field scores there are explicit H-bonds and penalties for rotational entropy. The disadvantage of empirical scoring is that they are dependent on data sets for fitting.

Kitchen et al, (2004) suggested that despite all of the approximations involved in database screening, screenings are successful because they are an enrichment process. Rather than selecting good binding ligands, they deselect inappropriate ones to enrich a short list and deselecting is more easily achieved with the limitations of the calculations than selecting good candidates. Also the ranking of ligands is unimportant as long as actives are present within the shortlist. Kitchen et al, (2004) states that it is this enrichment that explains the success of virtual screening despite false positives and negatives.

5.7.2 Screening Methods

The ZINC database (Irwin and Shoichet, 2004) subset of 100, 000 lead-like compounds was chosen for screening and converted to mol2 format using MOE. Both DOCK and GOLD were used. DOCK and GOLD tend to be used together. As they use different docking strategies and scoring functions the results of GOLD can support those of DOCK. The first calculations were done using DOCK. From this initial DOCK screening, the top 500 and bottom 500 scoring compounds were used as the input for GOLD. For each compound, 10 docking runs were used. From these only the top and bottom scoring ligands were selected for use. Therefore the top set scored highly with both methods while the bottom set scored poorly with both methods.

As many of the hits are similar, a diverse subset was determined using MOE. To produce the diverse subset the top and bottom 100 were fingerprinted using MDL MACCAS where each key describes a substructure of 1-10 non hydrogen atoms. A MACCAS fingerprint then indicates if one of 166 substructure keys are present in a molecule. Distance is then calculated between molecules using the Tanimoto similarity coefficient and molecules are then ranked by assigning 1 to the first entry and then repeatedly choosing the farthest from those already ranked.

5.7.3 Screening Results

From a database screening into the CYP6M2 model, the top hits from DOCK tend to have many flexible side chains and tend to not to have many rings (Figure 5.20, A). The presence of side chains may be due to the lack of a torsional entropy penalty in the DOCK scoring method. The top scoring hits tend to fill the crevices of the active site (Figure 5.21). This may give a high surface complementarity and may be responsible for their high DOCK scores, but may be spurious as torsional entropy is not taken into account. Also, no H-bonds were identified in many of the top hits and this may be due to a lack of an explicit H-bond function in DOCK. The low scoring ligands from DOCK tended to be large and containing rings similar to the known substrates for CYP6M2, and also steroids similar to testosterone (Figure 5.20, B).

From the DOCK screening the top 500 and bottom 500 were selected to be docked using GOLD. The top and bottom ranked compounds from GOLD were different from those of DOCK. Unlike DOCK, the top ranked ligands tend not to have side chains, which may be due to the torsional entropy penalty, and they tend to be large hydrophobic planar ligands with aromatic rings (Figure 5.22, A). A number of the high scoring ligands bind with similar interactions to the known ligands of CYP6M2. The second ranked ligand binds with similar interactions to the luciferins, with the ligand perpendicular to the heme, forming aromatic or hydrophobic interactions with F123 and F110 of the BC loop, and forming H-bonds with Q241 of the G'G loop (Figure 5.23, A, D). The 5th ranked ligand binds in a similar mode to deltamethrin. The structure of the ligand resembles the structure of the alcohol group and binds in a similar position with hydrophobic interactions with F110, H121 and F123 (Figure 5.23, B, E). A number of ligands also bind in a similar position to luciferin-PPXE with the ligands binding against F376 of SRS5 (Figure 5.23, C, F).
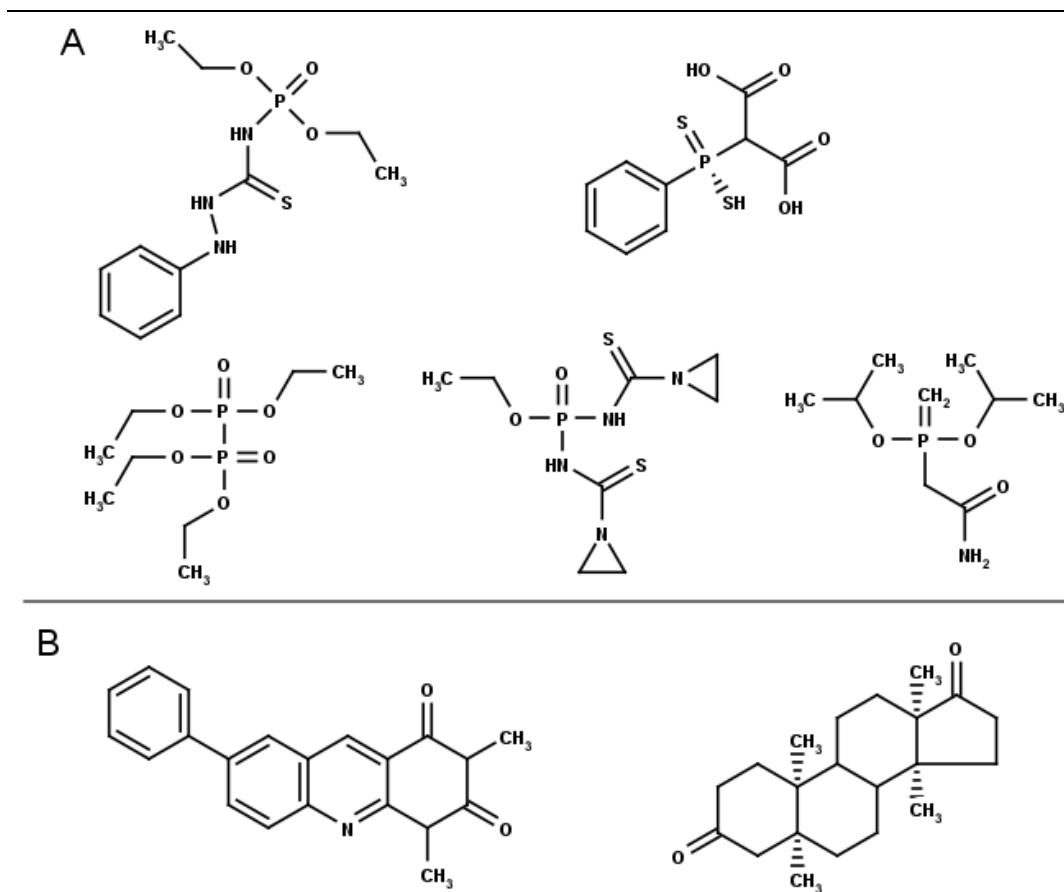
Figure 5.20 (A) The top five ligands produced by the Dock screening. (B) Examples of low scoring ligands from DOCK.
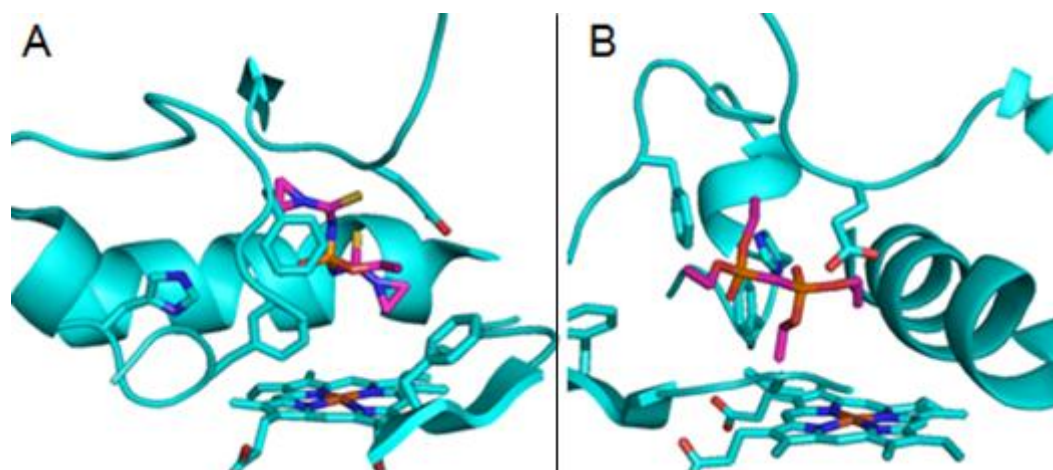
Figure 5.21 The first (A) and second (B) scoring compounds identified by DOCK bound in CYP6M2.
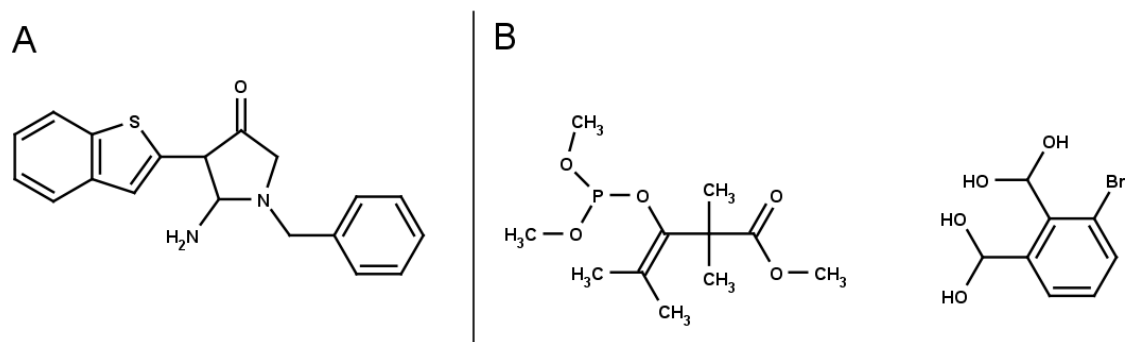
Figure 5.22 Examples of a high scoring (A) and low scoring (B) ligands from GOLD.
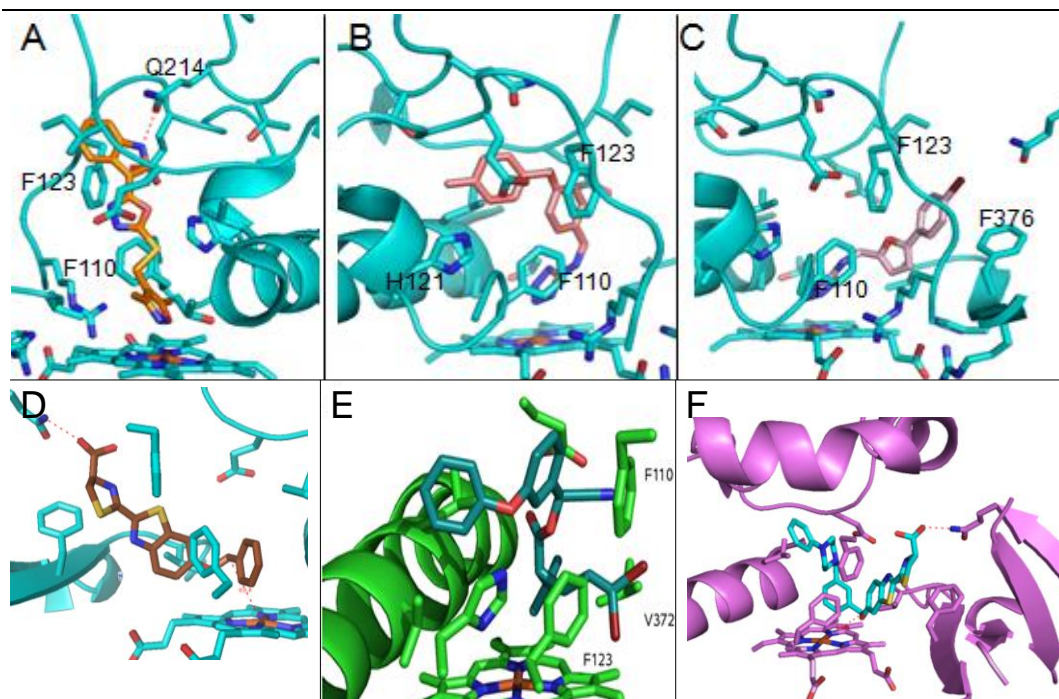
Figure 5.23 (A) The second highest scoring ligand from the GOLD docking, bound in CYP6M2. (B) The fifth highest scoring ligand from the GOLD docking binding in a similar mode as the deltamethrin acid group. (C) A high scoring ligand from the GOLD docking binding in a similar mode as luciferin-BE. CYP6M2 binding L-BE (D), deltamethrin (E) and L-PFBE (F).

By contrast, the low scoring ligands tend to be small ligands with few rings and containing polar groups such as NO2 and COOH (Figure 5.22, B). Ligands with many flexible sidechains are also low scoring, this may be due to the torsional entropy penalty. The identification of large ligands as high scorers and small ligands as poor scorers is consistent with the caveats identified by Kitchen et al, (2004) and may be due to large ligands having many contacts with the protein while small ligands have few contacts. However, a docking score that increases with ligand size is also consistent with findings that the free energy of binding increases with ligand size. Previous studies

196

have shown that binding free energy increases with increasing numbers of nonhydrogen atoms, with a contribution of up to 1.5 kcal/mol per nonhydrogen. However, each additional atom contributes less as the mass of the ligand increases and no improvement occurs in ligands with around 15 heavy atoms (Kuntz et al., 1999).

5.7.4 MOE analysis

A set of the top 100 and bottom 100 compounds were selected for an analysis of properties. To determine differences between the top and bottom scoring compounds and to identify common features within each set, molecular descriptors were calculated using MOE which were then correlated with ChemScore.

Among simple 2D descriptors there were a number of correlations. There was a correlation between ChemScore and the number of aromatic atoms ($R^2 = 0.62$), and number of hydrophobic atoms ($R^2 = 0.5$). However, this may be due to a bias for smaller ligands being poor scorers and large ligands being good scorers as there was also a correlation with the number of heavy atoms ($R^2 = 0.61$) and with molecular weight ($R^2 = 0.61$). There was a negative correlation for the number of oxygen atoms ($R^2 = 0.64$) while there was a positive correlation for the number of nitrogen atoms ($R^2 = 0.55$). There was also a positive correlation with total positive partial charge ($R^2 = 0.1$) and a negative correlation with total negative partial charge ($R^2 = 0.03$).

Ligands that were hydrophobic tended to produce good scores. There was a strong correlation with SlogP ($R^2 = 0.56$) and a negative correlation with logS, the log aqueous solubility ($R^2 = 0.49$). There was a correlation with the sum of vdW surfaces of hydrophobic atoms ($R^2 = 0.46$) and a negative correlation with that of polar atoms ($R^2 = 0.21$). This is constant with a hydrophobic active site for CYP6M2 but also with the caveat of Kitchen et al. (2004) that hydrophobic ligands tend to be high scorers. There was a positive correlation between ChemScore and the number of rings ($R^2 = 0.55$) indicating that aromatic interactions may be favourable. This is consistent with the presence of aromatic rings in both the known ligands and the CYP6M2 active site. The top scoring ligands are similar to the known substrates for CYP6M2 as they were of similar size, hydrophobicity, and tended to contain ring structures (Figure 5.24). The top 100 compounds had on average 23 heavy atoms and a molecular weight of 328.6. They contained on average 13.6 aromatic atoms, 14.14 hydrophobic atoms and contained 3.4 rings. The average logP (o/w) 3.02 and SlogP (3.15) and the logS (-4.76) indicated that

the top hits were large hydrophobic molecules.
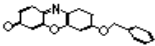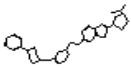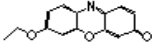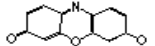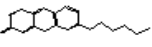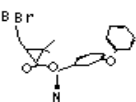
	mol	logS	SlogP	logP(o/w)	Weight	rings
1		−5.7356	4.0197	3.5410	303.3170	4.0000
2		−7.3148	4.1142	5.2690	543.6920	6.0000
3		−4.2949	2.5730	2.0940	241.2460	3.0000
4		−3.9677	2.1829	1.7530	227.2190	3.0000
5		−5.5271	3.7433	3.5920	283.3270	3.0000
6		−7.6351	6.5857	6.1164	505.2060	3.0000

Figure 5.24 MOE calculated properties of the known substrates for CYP6M2.

5.8 Conclusion

The homology models of CYPs 6M2, 6P3, 6P9, 6D1 and 6Z2 produced in this study can be used to explain the preferences towards the luciferins and acetylene inhibitors, and give some insight into the active sites of the pyrethroid metabolisers by identifying residues involved in producing these preferences. The models can also be used to predict novel substrates for testing that may be further used for the development of novel pyrethroids or inhibitors.

The CYP6 metabolisers show individual preferences towards luciferins and resorufins with varying substituents. CYPs 6P3, 6P9 and 6M2 preferred large substrates while CYPs 6D1 and 6Z2 preferred small substrates. A number of factors were found to affect the preference, the size and structure of the active site could be a major determinant.

198

The P450s with larger active sites prefer larger substrates while the smaller active sites prefer the smaller substrates. In particular, the size of the residue in SRS5 appears to determine the preference for chain length. Those with a Val had a preference for substrates with large chains, those with a Leu had a preference for the small chain luciferins. Sequence difference alone may not account for differences in preference, the sidechain rotamer may also have an effect. Rotamers are affected by the presence of prolines, and differences in rotamers may have a steric affect on preference for substituent size. The activity profiles could further be explained by optimal filling of the active site by the preferred ligands, steric clashes with larger ligands and greater mobility by smaller ligands.

There are similarities between the binding of luciferins and the binding of pyrethroids. These similarities could indicate a preference for aromatic substrates that occupy the volume above the heme. While there are similarities, the binding interactions of the luciferins differ from deltamethrin. With deltamethrin, two BC loop aromatic residues are involved in π-stacking while in the luciferins, only one is involved. This suggests that while the substrates share similarities, their interactions with the active site may differ and active site residues may have different ligand specific roles.

The structure of the active site can also be tested by the use of acetylene probes. These are selective for P450s by differing in size and shape and the positioning of the acetylene. The probes that produced the most potent inhibition in CYP6P3 were large and hydrophobic further indicating a preference of this enzyme for such substrates. Some probes were poor inhibitors, which may be due to their high logP or shape preventing their entry into the active site indicating that other factors external to the active site may have an effect. There were differences in the metabolism of pyrethroid-like probes in CYP6M2 and CYP6P3. The difference in preference was similar to that of the luciferins with the residue in SRS5 having a steric affect on binding and metabolism.

As the models were able to explain the observed experimental results, they may be used for database screening. Database screening can be used to identify candidate substrates but the results may be spurious due to the nature of the docking and scoring methods. This effect can be minimised where two different scoring methods are used to score the same ligands. DOCK and GOLD differ in both their docking and scoring methods and

may be used for consensus scoring. Receptor based screening using the CYP6M2 model identified candidate substrates that are similar to the known substrates of CYP6M2, hydrophobic ligands with aromatic rings, but further identified novel candidate ligands. Screening in this way gives an indication of the nature of the active site, as the ligands identified were a good structural and chemical match to the CYP6M2 active site.

6. Factors external to the active site affecting metabolism

6.0 Preface

This chapter addresses factors external to the P450 active site that can affect substrate preference and metabolism. In section 6.1 the role of cytochrome b5 is discussed. The presence of b5 is known to affect pyrethroid metabolism by P450s and may exert its role in stabilising the protein or acting as a second electron donor. In section 6.2-6.3 the role of protein flexibility is discussed. Flexibility needs to be taken into account in docking, and b5 may also have an influence on stabilising protein conformation. In section 6.4 the role of b5 as a second electron donor is discussed in relation to DDT. In section 6.5-6.6 membrane interactions are discussed, the interaction of the P450 with the membrane may have an influence on ligand access. In section 6.7 the occurrence of SNPs are discussed.

6.1 Cytochrome b5

Cytochrome b5 is a 17 kDa membrane anchored heme protein that participates in a range of electron transfer reactions. It has a role in augmenting P450 reactions by direct transfer of the second electron to the P450, and by allosteric stimulation without electron transfer. Although the overall structure of cytochrome b5 is not known, it is described as having 3 domains, a water soluble heme containing domain, a transmembrane anchor domain and a linker region which connects the two (Figure 6.1) (Durr et al., 2007b). b5 interacts with P450s but in a P450 specific way. Some P450s show an enhanced activity; some show no affect while others are inhibited; the interaction can also be substrate specific. The strength of binding also varies with some P450s showing strong K_d values while in others is it undetectable (Durr et al., 2007a).
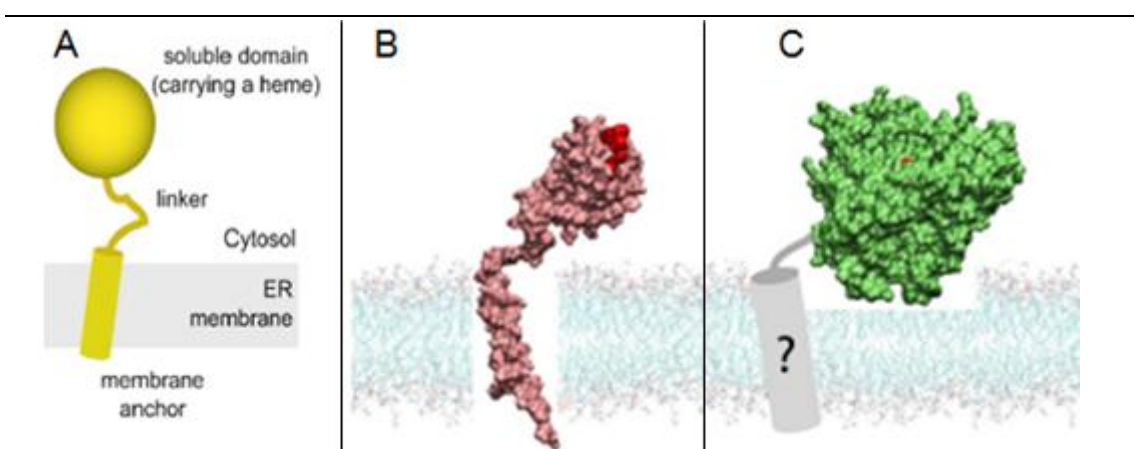
Figure 6.1 (A) Topology of b5, (B) model of b5, (C) CYP2B4 (1SUO). Taken from Durr et al. (2007a).

A number of possible binding sites have been proposed for b5. These sites include the distal site with interactions with the A, F and I helices. A second site perpendicular to the heme plane on the B' helix, and a third site on the proximal face with the C, K and L helices and meander region. It has been demonstrated that b5 could share an overlapping but non-identical binding site with reductase on the proximal surface as at high ratios of b5 to reductase, b5 inhibits the activity of reductase. b5 has been shown to be required for the CYP6D1 mediated metabolism of cypermethrin. This was suggested to be due to its role as a second electron donor although other influences were not investigated (Zhang and Scott, 1996).

6.1.1 The effects of b5 on CYP6M2

In CYP6M2, the presence of b5 shifts the ratio of metabolites for DDT, and the metabolism of permethrin and deltamethrin was increased 2.5 fold in the presence of b5. Experimental studies also suggest that b5 increases the stability of CYP6M2 (McLaughlin, unpublished). To determine the effects of b5 on CYP6M2, the site of binding of b5 needs to be determined. From the literature, mutagenesis studies affecting b5 binding can be used to predict the binding site on CYP6M2 and understand how b5 binds with different strengths to different P450s. This may give some insight into the effects of b5 binding to CYP6M2.

As both b5 and CYP6M2 are membrane bound, the orientation of interaction with the membrane may have an influence on the interaction with b5. The interaction of CYP6M2 with the membrane was estimated by using MAPAS. In the top scoring mode the hydrophobic patch on the A', F' and G' helices is in contact with the membrane. The heme iron is around 24 Å from the membrane surface similar to the model proposed for CYP2C5, the protein is also tilted which could expose its proximal surface for the binding of redox partners (Clarke et al., 2004).

6.1.2 b5 binding site

Tamburini et al, (1985) characterised the interaction between b5 and P450s using chemical modification of b5. b5 was modified by amidination to alter the structure of the ε-amino group of lysine but not the charge, acetylation to remove the cationic charge, and methylamidation to remove the carboxyl anionic charge. Native b5 bound tightly with P450s RLM3 and LM2, and the presence of b5 increased *p*-nitroanisole

demethylation activity. The amidinated and acetylated b5 exhibited similar patterns to those of the native b5 showing that neither positive charge or sidechain structure of the lysene residues were involved in complex formation with the P450s. By contrast modification of carboxyl groups interfered with *p*-nitroanisole demethylation activity stimulation. This demonstrates that amino groups were unimportant but carboxyl groups were essential for electron transfer between b5 and P450s. They suggested that there may be a need for charge pairing for efficient electron transfer. Tamburini et al. (1986) also showed that the heme propionates are important for b5 binding. Esterification of heme propionates to produce DME-b5 (dimethyl ester b5) causes anionic charge neutralisation. DME-b5 showed a weaker affinity for P450 RLM5 than native b5.

A number of possible binding sites have been proposed for b5. Mutagenesis has indicated roles for the basic residues on the P450 proximal face. Using chemical cross linking of ion pairs, the electrostatic interactions between CYP2E1 and b5 have been identified, these have also been confirmed using mutagenesis (Gao et al., 2006) (Table 6.1).

Table 6.1 From mass spectrometry two cross linked ion pairs were identified by Gao et al. (2006).

Region	CYP2E1	b5
meander	K428	E53
B bulge	K434	E56 *D60*

These ion pairs place the binding site on b5 on the surface where the heme propionates project into the solvent, while the binding site on the proximal face of CYP2E1 occurs where the heme is closest to the surface of the P450. Mutagenesis suggested that CYP2E1 K428 and K434 are structurally responsible for the stimulatory affects of b5. Mutagenesis of either of these reduced the stimulatory affect of b5 and simultaneous mutagenesis decreased the stimulatory affect by up to 82%. K434A has the greatest effect indicating it may be involved in a second ion pair interaction, this was predicted to be with b5 D60 (Gao et al., 2006). From modelling, Gao et al. (2006) also predicted other ion pairs (Table 6.2).

Table 6.2 Other ion pairs identified by Gao et al. (2006).

Region	CYP2E1	b5
C helix	R126	D60
J helix	K342	E43
J helix	R344	E48 E44
L helix	R444	E48
meander	K422	E48

The mutagenesis by Gao et al. (2006) places the site of binding on b5 on the helices 3 and 4 (Figure 6.2, A). The site on CYP2E1 was on the proximal surface (Figure 6.2, B). Site directed mutagenesis in CYP2B4 has also identified possible binding sites for b5 and reductase on this P450. In a study using both mutagenesis and protein-protein docking, Bridges et al. (1998) and Zhang et al, (2005) identified seven residues involved in b5 binding. The seven residues were identified to be involved in binding b5 by the lack of b5 stimulation of metabolism and by an increase in b5 dissociation constant. These are R122, R126, R133, F135, M137, K139 and K433, and occur along the C helix (Figure 6.3).
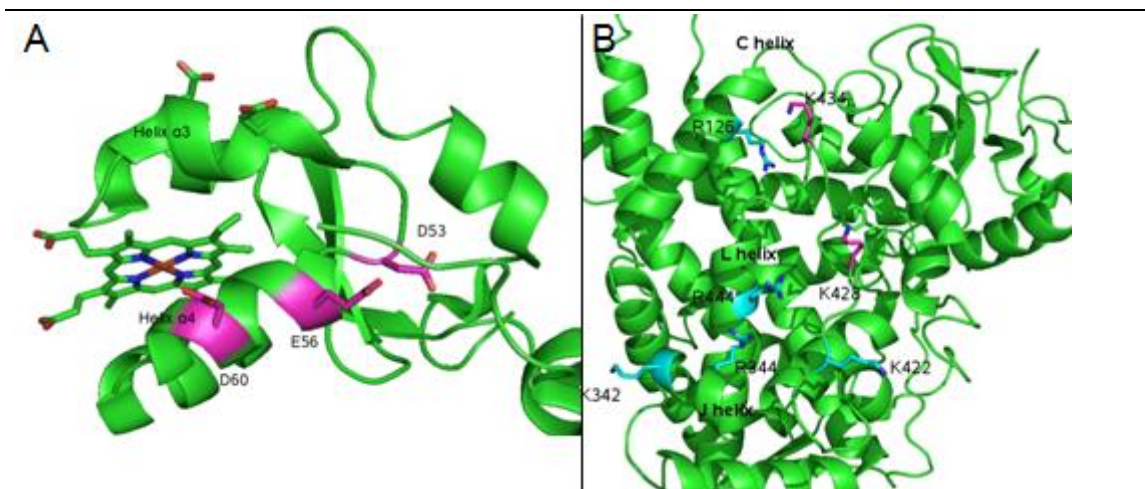
Figure 6.2 (A) Rat b5 (1AW3) with the residues involved in the two ion pairs identified by mass spectrometry and the predicted pair highlighted in pink. (B) CYP2E1(3E6I) proximal surface. The ion pairs identified experimentally are highlighted in pink. The pairs identified by modelling (Gao et al., 2006) are highlighted in blue.
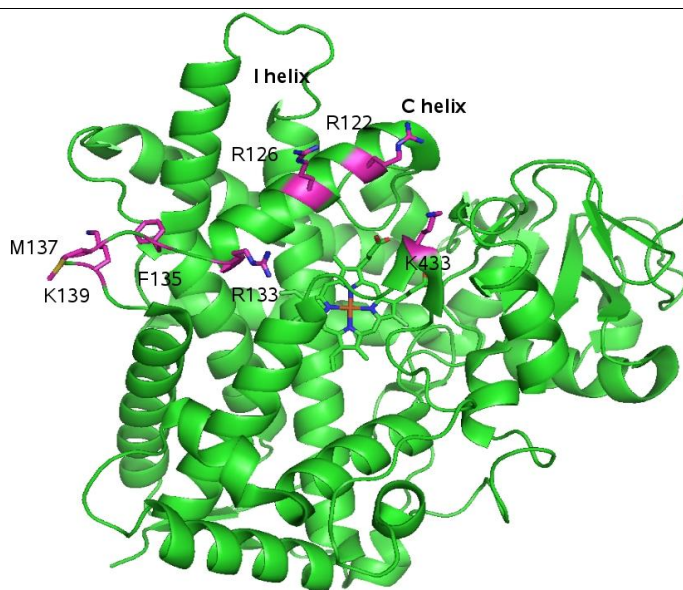
Figure 6.3 The seven residues identified by Zhang et al, (2005) to be involved in b5 binding highlighted in pink.

The site of b5 binding is similar in both CYP2E1 and CYP2B4. This may be due to sequence conservations. The seven residues of CYP2B4 are either conserved or are conservatively substituted in CYP2E1. Similarly some of the basic residues of the J and L helices identified in CYP2E1 are conserved in CYP2B4. Of the residues identified as involved in b5 binding there were only substitutions in the J helix (the K342 of CYP2E1 replaced by D341 in CYP2B4) and meander (the CYP2E1 R428 replaced by M427 in CYP2B4).

6.1.3 b5 interaction differences

It has been shown that different P450s show different strengths of interaction with b5 (Shimada et al., 2005) (Table 6.3). An alignment of these sequences was produced with Tcoffee to determine if differences in b5 interaction could be attributed to residue substitutions, and the mutagenic studies in CYP2B4 and CYP2E1 were used to identify positions that have been shown to be involved in b5 binding.

Table 6.3 Kd values of b5 interaction with P450s taken from Shimada et al. (2005).

Enzyme	Kd (µM)
Rat P450 1A2	0.09 ± 0.02
Rabbit P450 1A2	5.4 ± 1.9
Human P450 1A2	ND
P450 2A6	0.13 ± 0.07
P450 2C19	ND
P450 2D6	0.14 ± 0.02
P450 2E1	0.55 ± 0.45
P450 3A4	0.05 ± 0.01

Kd ± SE

ND, not detectable
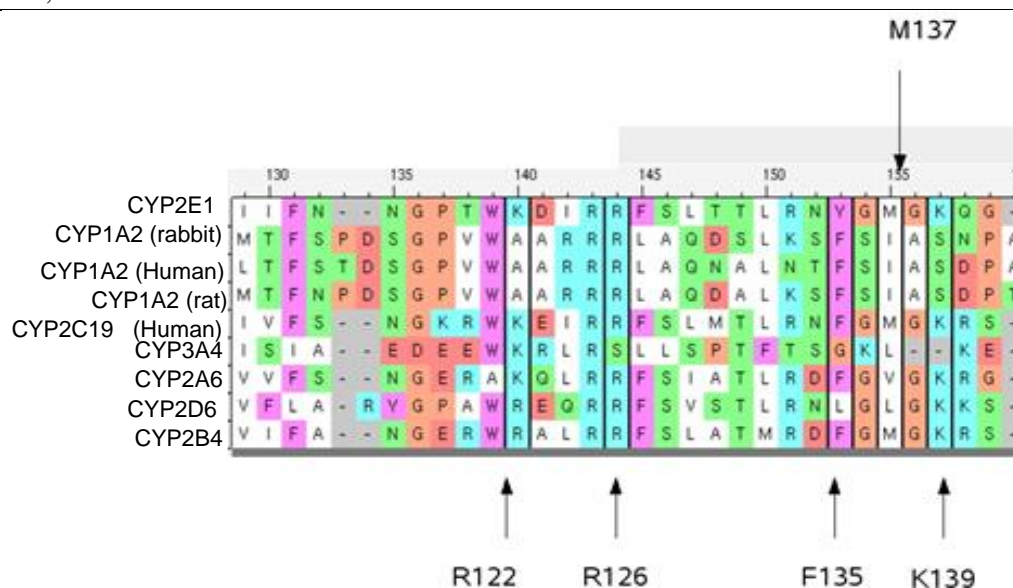
Figure 6.4 The positions of residues identified as involved in b5 binding in CYP2B4.

At the proximal surface, there is a high degree of sequence conservation. The residues involved in b5 binding identified in CYP2B4, tend to be conserved in other P450s with the exception of the CYP1A2 group (Figure 6.4) which may explain their poor binding to b5. Non conservative substitutions also occur in CYP3A4 and CYP2D6 although they have strong interactions with b5, and other substitutions may compensate for these. There is an additional basic residue (K96). The CYP3A4 K96 is in a position to interact with b5 as it is at the C terminal of the B helix. In the CYP1A2 group, substitutions of some of the b5 binding residues occur. While human CYP1A2 has no detectable interaction with b5, rat and rabbit CYP1A2 showed stronger interactions. In the rabbit and rat sequences an additional basic residue occurs that substituted for N145 in human CYP1A2. The position of this residue in the crystal structure of human CYP1A2 shows that this residue is in a position to interact with b5 (Figure 6.5) and its substitution in human CYP1A2 may contribute to the reduced interaction with b5. Shimada et al. (2005) found that CYP2C19 interacted weakly with b5. CYP2C19 does not have any substitutions of the residues identified to be involved with b5 interaction, and has an additional basic residue on the N terminus of the C helix. As the lack of interaction in this case cannot be understood basis on sequence in this case, there may be other factors such as structural differences or membrane interaction that affects b5 binding.
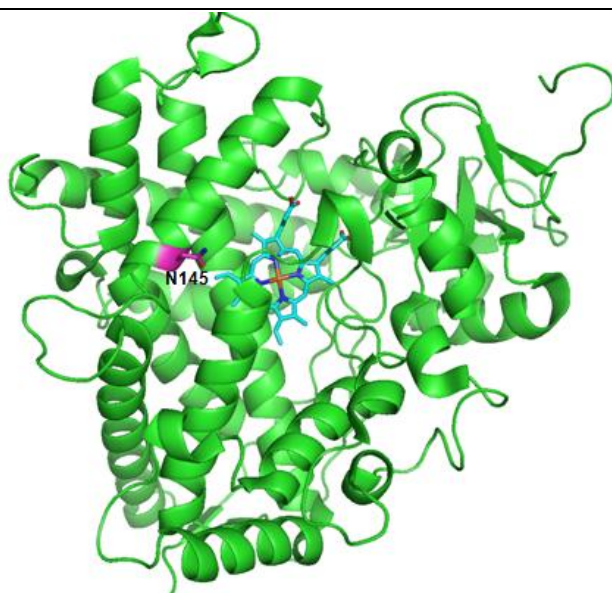
Figure 6.5 The position of N145 on the proximal surface of human CYP1A2 (2HI4).

6.1.4 Proximal surface of CYP6M2

The proximal surface of P450s is suggested to be the site of binding for b5, with basic residues involved in the interaction. The proximal surface of CYP6M2 can be analysed

for these features to identify residues that may be involved in binding to b5. Although in CYP2E1 two basic residues have been identified as important for b5 binding, both of these are substituted in CYP6M2, of the other basic residues identified by modelling (Gao et al, 2006) some are substituted while others are conserved.

The two basic residues identified on the CYP2E1 J helix are substituted, but a basic residue on the CYP6M2 K helix occurs in a similar position. Although the basic residue on the CYP2E1 L helix is substituted, a basic residue occurs on the CYP6M2 L helix in an adjacent position that is shifted towards the B and C helices similar to CYP3A4. The basic residue on the CYP2E1 C helix is conserved, although this may be conserved due to its role in binding a propionate.

In addition to the residues identified in CYP2B4 and CYP2E1, a number of basic residues occur on the CYP6M2 proximal face that may indicate a b5 binding position (Figure 6.6). Two basic residues occur on each of the B and C helices in CYP6M2 and CYP3A4, that are also conserved in the other pyrethroid metabolising CYP6s (Table 6.4). A basic residue is also conserved on the L helix, although its position is shifted towards the C helix in the CYP3s relative to its position the CYP2s. This along with the absence of the basic residues on the J helix may act to shift the binding position of b5 towards the BC loop in CYP6M2.

Table 6.4 Basic residues in the CYP3 clade.

	C helix		L helix		B helix		D helix
6M2	R131	R134	R450	R457	K94	K99	K147
3A2	K	R	R	K	K	K	K
3A5	K	R	R	K	R	K	K
3A4	K	R	R	K	K	K	K
321A1	K	R	R	L	Q	S	K
6B8	R	R	R	L	K	K	K
6D1	R	R	R	K	R	T	K
6P3	R	R	R	K	K	K	K
2E1	R122	R126			K87	K94	
2B4	K122	R125			R85	A92	
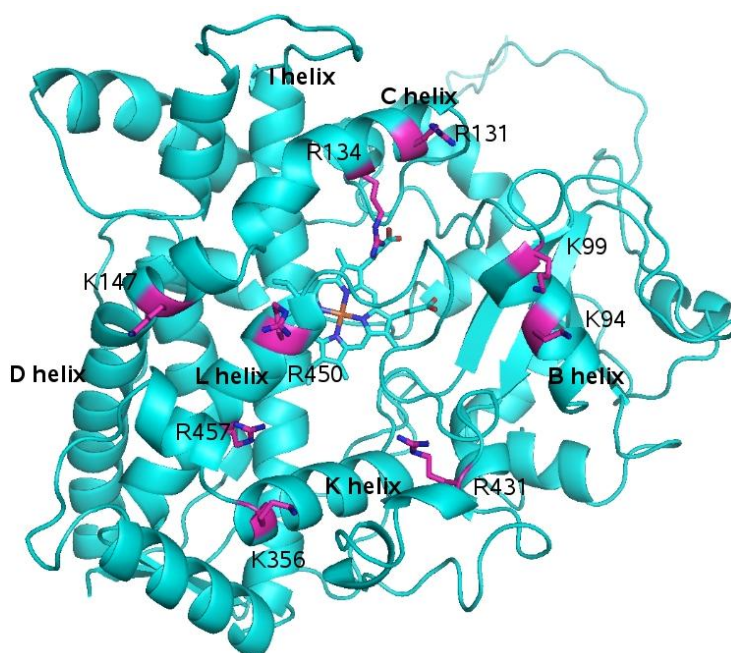
Figure 6.6 CYP6M2 proximal face with basic residues highlighted in pink.

6.1.5 Docking predictions for b5 binding sites in CYP6M2

Protein-protein docking of b5 with CYP6M2 may also be used to identify binding sites. GRAMM can be used for prediction of protein interactions of ultralow resolution structures (7 Å) to give general preferences for sites of interaction and therefore can be used with homology models. It does not produce real coordinates but a distribution of positions around the binding site. An ultralow resolution docking was used where structural details are averaged up to 6.8 Å to predict the gross binding features. From 100 output conformations a primary cluster occurred on the distal surface while a much smaller cluster occurred on the proximal surface. The top 50 dockings consisted only of dockings on the distal surface.

A ClusPro docking was performed using either DOT or ZDOCK and retaining 2000 of the electrostatic hits. Using DOT two main clusters occur on the proximal and distal surfaces. There are also some dockings on the lateral surface on the BC loop although they are diffuse. A number of the dockings on the proximal surface places the b5 heme close to the heme of CYP6M2. Similar to the DOT output, with ZDOCK there is a large cluster on the proximal surface and diffuse dockings on the lateral surfaces. Unlike DOT there are two clusters on the distal surface although they occur close together. This difference may be due to differences between the algorithms. Zdock has a desolvation component that DOT lacks and this additional cluster may be due to a favourable

209

desolvation score where b5 binds to the hydrophobic patch.

A single cluster on the proximal surface is consistent between both DOT and Zdock and may indicate an actual binding site. However, the orientation of the b5 with the P450 varies within the cluster with some orientations positioning the b5 heme away from the proximal surface while others position it towards it. As the b5 heme propionates and residues close to the heme have been shown to be involved in the binding of b5 to P450s, b5 dockings with the heme at the binding interface were assumed to be the naturally occurring orientation. In addition a proximity of the hemes may be needed for electron transfer.

The second ranked mode of Zdock places the hemes of b5 and CYP6M2 close to each other, the distance between the irons of the two heme groups is 16.9 Å while the distance between the propionates is 5.6 Å. As it has been found that electrons can travel up to 14 Å between redox centres (Page et al., 1999), this pose may be close to that needed for electron transfer. This docking places b5 at the rim of the concavity binding with the B, C, and N terminal parts of both the D and L helices (Figure 6.8). In this binding mode b5 partly occupies the concavity of the proximal surface, as it has been suggested that other redox partners also bind at this surface, this binding mode may only occur in the absence of the reductase.

In this docking mode the start of the b5 linker region was 57 Å from the membrane surface as predicted by MAPAS. Human b5 has a 15 amino acid linker region. If this were extended like a beta sheet its length could be 51 Å whereas if it were an α-helix its length would be 22 Å. It has been found that if the linker region is shortened to 7 amino acids, this causes it to be unable to interact with CYP2B4 as the predicted binding surface was 15-20 Å above the membrane (Clarke et al., 2004). The predicted binding site in CYP6M2 is further from the membrane than that predicted in CYP2B4. However, *Anopheles* b5 has an insert in the linker region that may extend the region to 19 amino acids (Figure 6.7). The increased length of the linker region may allow b5 to bind in this position.
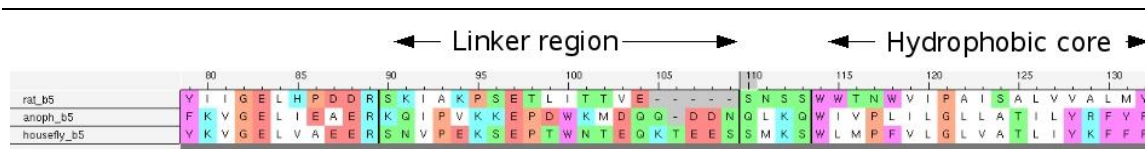
Figure 6.7 Alignment of *Anopheles*, rat and housefly b5. Positions of the linker region and transmembrane domain were identified according to Clarke et al. (2004). The insect b5s have an insert in the linker region.
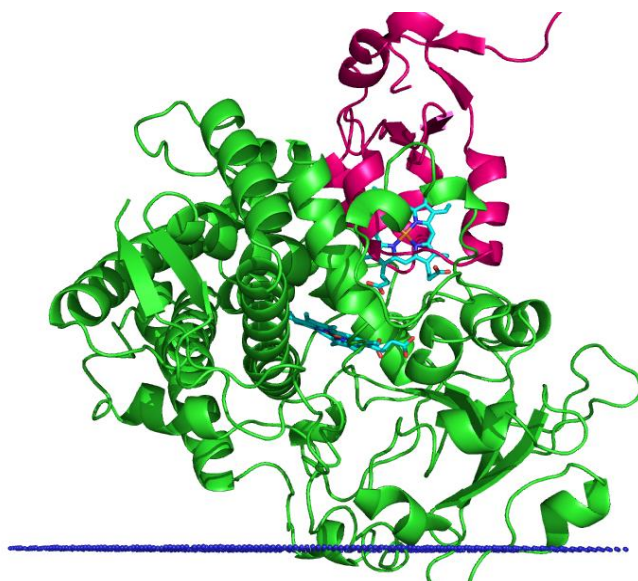
Figure 6.8 Second ranked docking using ClusPro Zdock. The heme of b5 (pink) is close to the heme of CYP6M2 (green) and b5 is distant from the membrane surface (blue plane).

The dockings of b5 at the distal surface may be an artefact due to the presence of the hydrophobic membrane binding patch and a concavity. As protein docking algorithms assume that the protein is surrounded by water, the binding on the distal surface may be due to a favourable desolvation score. Also there is a concavity on both the distal and proximal surfaces, as b5 is a similar size to both of these cavities, docking algorithms may be identifying a good shape complementarity between the surfaces. The interaction between CYP6M2 and the membrane was predicted using MAPAS. If this is the correct orientation, dockings on the distal face place the water soluble heme binding domain of b5 into the membrane (Figure 6.9). As this binding mode may not be possible, the dockings on the distal surface were regarded as an artefact.
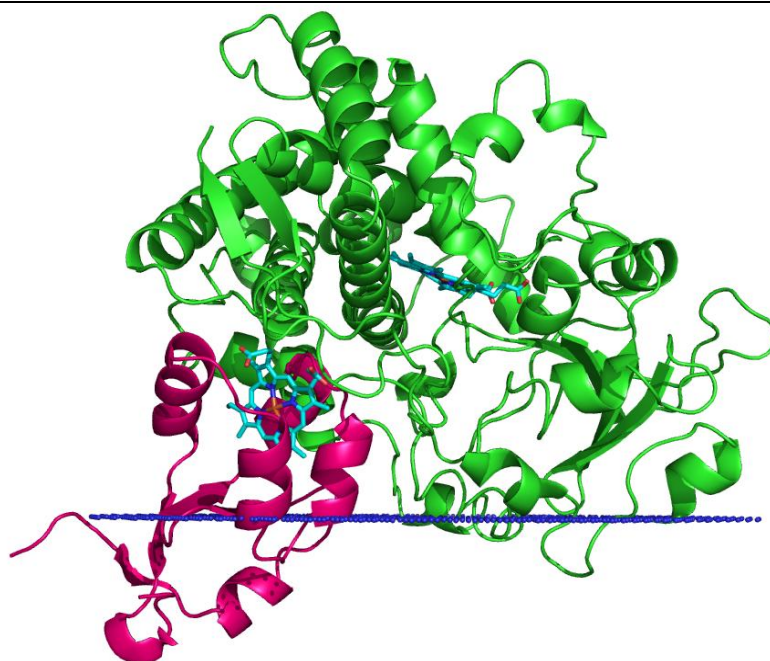
Figure 6.9 The 3rd ranked Cluspro Zdock docking of CYP6M2 and b5 with the position of the membrane predicted (blue plane).

A FireDock refinement of the ClusPro-Zdock poses was carried out. Of the top 10 scoring modes, the dockings ranked 1-5 and 10 occurred on the lateral surfaces. Although these modes were high scoring, they were diffuse and may not represent actual binding sites. Modes 6,7,8 were clustered on the proximal surface with modes 6 and 7 orienting the heme of b5 towards the heme of CYP6M2. The 9th ranked mode was on the distal surface. The scores indicated that the dockings on the proximal surface involved strong electrostatic interactions as these scores were high, this is consistent with the binding of b5 involving ion pairs. The dockings on the lateral surface had higher van der waals scores indicating that these dockings may be driven by surface complementarity (Table 6.5).

Table 6.5 Scores for the cluspro-Zdock FireDock refinement. Position indicates the surface of binding: L: lateral ; P: proximal; D: distal.

	position	glob	aVdW	rVdW	ACE	inside	aElec	rElec	laElec	lrElec	HB	piS	catpiS	aliph
1	L	-102.9	-52.11	18.48	-4.9	18.67	-122.3	0	-22.98	3.8	-6.62	0	0	-5.5
2	L	-101.78	-54.62	16.09	-6.03	15.31	-96.53	10.48	-29.72	4.07	-6.07	0	0	-4.5
3	L	-98.67	-48.32	17.05	2.98	15.42	-187.2	24.6	-27.9	0	-10	-0.5	0	-3.5
4	L	-94.02	-60.71	22.31	-2.77	22	-131.75	26.29	-55.95	14.17	-14	-1	-0.5	-0.5
5	L	-88.07	-49.93	11.1	-7.01	13.67	-101.82	10.48	-22.13	12.17	-3.64	-1.5	0	-1.5
6	P	-82.51	-44.47	19.71	6.72	11.82	-218.76	17.41	-33.98	4.13	-5.78	-0.5	0	-1.5
7	P	-77.55	-57.46	19.85	10.31	11.76	-186.98	59.27	-34.96	17.08	-5.78	-2.5	0	-3
8	P	-75.81	-48.87	16.11	2.36	16.04	-132.71	0	-36.49	4.71	-4.82	0	0	0
9	D	-69.64	-67.26	19.04	14.91	13.35	-146.4	93.05	-29.54	17.67	-10.97	-5	0	0
10	L	-68.72	-55.97	21.05	-4.47	14.43	-85.39	56.19	-18.62	31.11	-11.79	-2	0	-1.5

Zdock was also used alone. Unlike the ClusPro output, the primary cluster was on the distal surface on the hydrophobic patch while a much smaller cluster occurred on the

proximal surface. Also unlike ClusPro, the docking on the proximal surface is shifted further onto the rim (helices B and C) away from the concave surface, this was also suggested by the position of basic residues on these helices. Docking on the rim of the proximal surface may be in a position to stabilise the interaction with P450 reductase.

The binding position of P450 reductase can be predicted from the crystal structure of the P450BM3-reductase complex 1BVY. To model the CYP6M2-FMN complex, 1BVY was aligned to CYP6M2. In the mammalian P450 reductase, the FMN domain is buried within the structure but as the domain is tethered by a single flexible linkage, it may be possible for it to approach the heme in this way as suggested by Williams et al. (2000). In the CYP6M2-FMN complex the FMN occupies the concave surface of CYP6M2 and would exclude b5 from this site. However, the Zdock mode positioning the b5 further onto the rim of the proximal surface could allow the reductase to bind and may stabilise the interaction. However, there was some overlap between the FMN and b5 (Figure 6.10). Alternatively, b5 binding to the lateral surface could also stabilise the binding of the reductase although docking modes on the lateral surface are diffuse and do not indicate a particular binding site.

This stabilising of the reductase has been previously suggested as b5 has been shown to affect the stability of CYP3A4. Voice et al. (1999) found that in E. coli cells expressing CYP3A4 and NADPH in the presence of testosterone lead to a decrease in spectrally active CYP3A4, but in cells also expressing b5 the levels of active CYP3A4 were unchanged. In this case it was suggested that this was a result of b5 improving the coupling of CYP3A4 with NADPH reductase to reduce the levels of damaging reactive oxygen species. It was also found that in the absence of substrate, b5 led to a loss of spectrally active CYP3A4. They suggested that an increased coupling in the absence of substrate could produce a futile cycle that could destroy the enzyme.

The site of b5 binding on CYP6M2 is predicted to be on the proximal surface where it may bind in a position that allows it to donate electrons or alternative poses that may stabilise the interaction with the reductase. This indicates roles for b5 in increasing CYP6M2s metabolism towards pyrethroids.
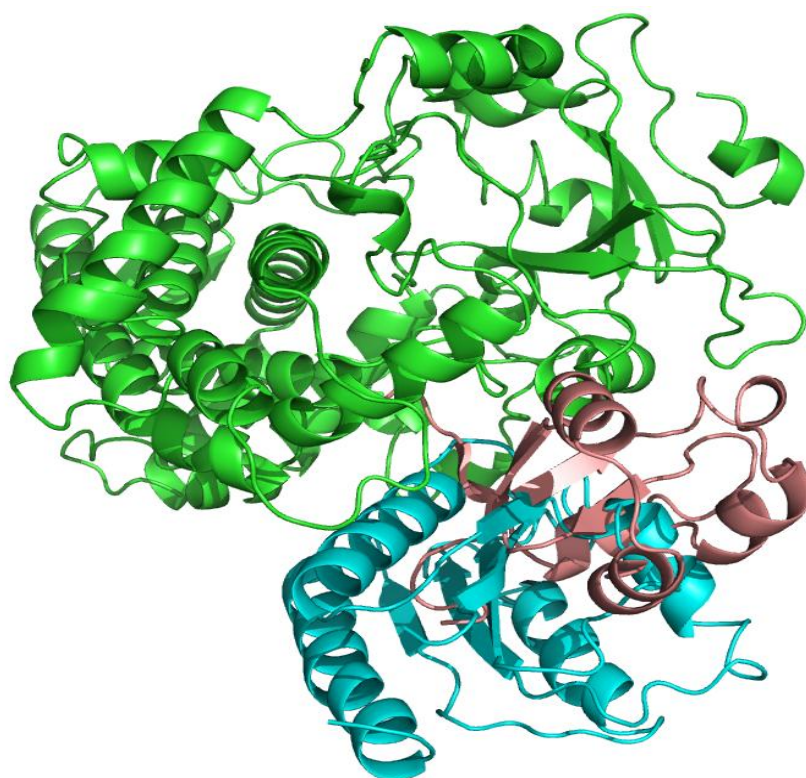
Figure 6.10 CYP6M2 (green) with the best ranked binding mode in Zdock (pink) and the superimposed position of FMN from 1BVY (blue).

6.2 Protein flexibility

The binding of b5 may influence the stability of CYP6M2, to identify a mechanism for this stability, the flexibility of CYP6M2 was investigated. From the normal mode analysis, the position of binding of b5 on the proximal face could stabilise a hinge bending motion to increase the rigidity of the protein.

6.2.1 P450 flexibility

The use of X-ray structures to understand the flexibility of P450s is limited. X-ray structures do not provide direct evidence of protein flexibility, which are based on temperature factors or comparisons of ligand bound and unbound structures. X-ray data may be affected by the non-physiological conditions and by crystal packing to give non-native contacts.

Zhao et al. (2006) identified five plastic regions (PR) from the crystal structure of CYP2B4. PR1 consists of the A helix; PR2 includes the B', B'C loop, helix C and the CD loop; PR3 consists of the C-terminal half of helix E. PR4 consists of the C-terminal half of helix F through to the N-terminal half of helix I. PR5 consists of part of the C-

terminal loop. Zhao et al. (2006) suggested that the most conformationally flexible regions of CYP2B4, the BC loop and FG regions, are embedded in the membrane and may be dependent on the lipid environment. They suggested that this region could act to recognise and deliver diverse substrates to the active site. Skopalik et al. (2008) found that P450s have regions of low or high malleability. The heme-binding core is rigid. The proximal side where the redox partners bind is moderately malleable. The distal side that binds with the membrane is highly malleable and may contribute to the substrate specificity, as substrate entry and release can occur without disrupting the active site arrangement or interaction with the redox partner. This may be the case in CYP6M2 as the FG loop in contact with the membrane spans three domains and contains hinge regions.

It has been demonstrated in CYP4A7 that b5 can have a conformational role rather than an electron transfer role as heme-depleted apo b5 increased turnover of laurate and myristate indicating it plays a conformational role as it is not involved in electron transfer (Loughran et al., 2001). Loughran et al. (2001) characterised CYP4A7 in the presence and absence of b5, they suggested that it enhances the interaction between P450 reductase and the P450. They determined the flexibility of the active site on fatty acid substrates with different chain lengths, CYP4A7 prefers shorter chain lengths in the presence or absence of b5, but apo b5 was required for the metabolism of longer chain lengths. The greater the chain length the greater the dependence on b5. They suggested a conformational role for b5 which may dock at the proximal surface to open a channel for the access of substrates with varying bulk. A spectral characterisation showing a 3 fold increase in the spectral binding constant of lauric acid in the presence of b5 indicates a change in how the substrate binds to the protein. They suggested that b5 acts as a conformational modifier that causes an increase in overall activity as there is a three fold increase in hydroxylation in the presence of b5 and heme depleted b5. Yamazaki et al. (1996) also found that while b5 is traditionally considered to be involved in electron transfer, heme depleted b5 (apo-b5) can replace b5 for the oxidation of substrates of CYP3A4.

6.2.2 Normal modes

Normal mode analysis (NMA) predicts the movements of a molecule and has been used with proteins to predict the kinds of conformational changes that occur during normal function. Half of the known movements can be modelled by at most two normal modes

215

(Suhre and Sanejouand, 2004). Normal modes are harmonic oscillations around a local energy minimum, the starting point for normal mode analysis is one particular stable conformation that represents a potential energy minimum. The potential energy landscape of a protein is multiscale. At the large scale a stable conformation corresponds to a local minimum. If several local minima exist they represent different conformations. At a smaller scale the potential well is not smooth but has many local minima that represent conformational sub-states that may be different arrangements of sidechains.

In proteins, conformational change usually involves the relative movement of rigid structural elements. The degree of collectivity is the fraction of residues that are affected by a given mode. For maximal collective movements the degree of collectivity tends to be a value of 1. For localised motions the value approaches 0. While low frequency modes are expected to be collective, they can be localised and usually correspond to the movement of extended parts of the system such as loops. As these are meaningless they can be ignored and the extended parts are usually removed before normal mode computation. ElNémo was used for normal mode analysis. The model of CYP6M2 was submitted to the elNémo server. In all of the first 5 models there was low collectivity. Low frequency modes are expected to be collective while localised motions may correspond to extended parts of the system such as N or C termini. The cause of the localised motions was identified as the transmembrane domain. With the transmembrane sequence removed only one of the modes show low collectivity. The lowest frequency mode is number 7. The first six correspond to translation and rotation of the protein. The vector field representation of mode 7 identifies the B and C helices as moving in opposite directions (Figure 6.11). In this mode there is collective motion of three domains: the beta sheet region, the FG region and the I,J,K helices. The B helix appears to be part of the beta sheet domain and shows a collective motion along with this domain. The BC loop is acting as a hinge region between domains.
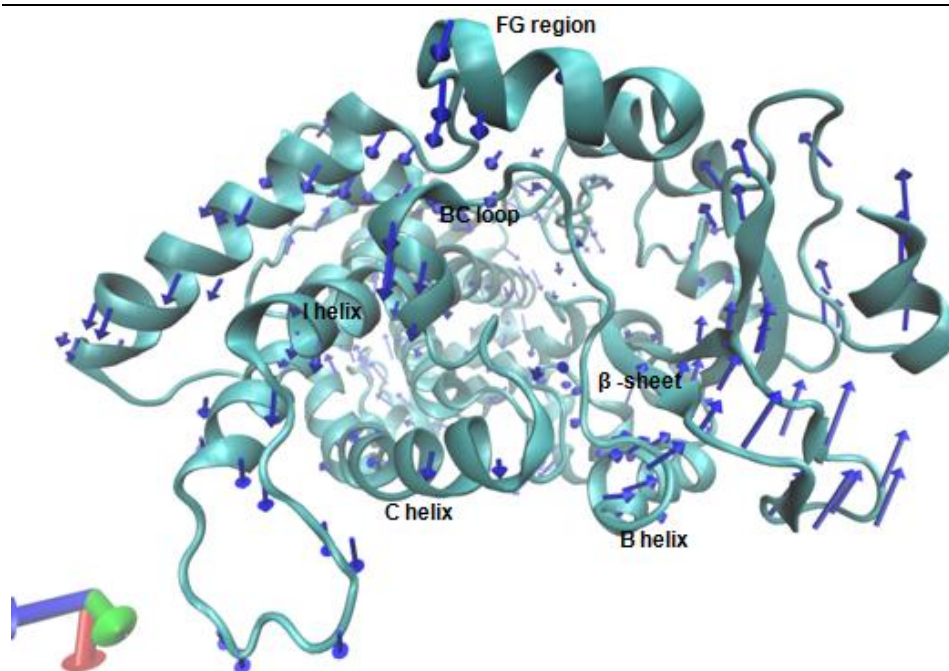
Figure 6.11 The vector field representation of the displacements of mode 7 of CYP6M2. Vectors represent the direction of the domain displacements.

6.2.3 Domain analysis

DynDom was used to identify domains. In the CYP6M2 elNémo normal mode 7, three domains are identified (Figure 6.12). The domains identified correspond with those identified by Arnold et al. (1997) in P450BM3:

β   B sheet region

α'   FG, N terminal part of I helix

α"   C terminal part of I helix, JK and L helices C terminal loop.

With helices B', C, D and E acting as transition regions between domains.
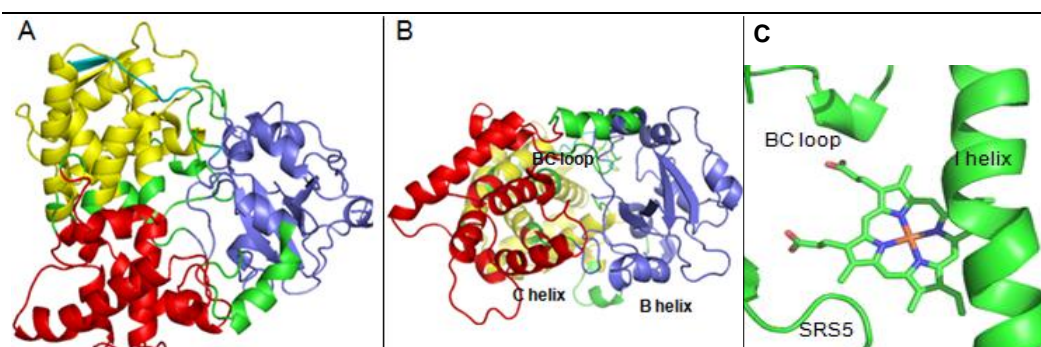
Figure 6.12. The CYP6M2 domains of mode 7 highlighted. (A) distal surface, (B) lateral surface. β domain (blue). α' domain (red). α" domain (yellow), between these domains are hinge regions (green). Positions of the BC loop, I helix and SRS5 in the active site.

In normal mode 7, the B helix forms part of domain β while the C helix forms part of domain α' with a hinge region in the BC loop. The BC hinge region consists of residues F110-E113. In this normal mode there is a bending of the I helix. The I helix is divided, with one part in α' domain and the other part in the α'' domain separated by a hinge at the electron transfer groove. The I helix straightens and flexes along with the domain movement. ElNémo Normal Mode 8 is in agreement with mode 7 in identifying 3 large domains. Unlike mode 8 it identifies the G' helix as being a fourth domain. Also unlike mode 7 the BC loop hinge region is extended to the C helix.

While CYP6M2 shows 3 domains, CYP6P3 shows 6 domains. However some of these domains correspond to surface loops, excluding these there are three large domains similar to that found in CYP6M2. The BC loop acts as an interdomain region between the α' and β domains in both CYP6M2 and CYP6P3. The template CYP3A4 (1TQN) also shows three main domains and a similar domain motion. In both CYP6P3 and CYP6M2 there is a hinge in the I helix, next to this hinge is a second hinge in helix E and a third in the D helix. Together, these hinges occur between two domains and may form an interdomain region between domains α' and α''.

The F and G helices are in the α' domain, but the FG loop is part of the β domain. The α' domain moves in the opposite direction to the β domain and the FG loop acts as a hinge between the two domains. Along with the domain motion, the FG loop bends laterally. This can act to change the dimensions of the access channel 3 that runs through the loop. From ElNémo, it was found that the C helix occurs in one domain while the B helix occurs in another. Both domains move in opposite directions with a hinge region in the BC loop. The predicted b5 binding site occurs at the B and C helices. Binding in this position could stabilise the domain motions to increase the rigidity of the P450.

In the non metaboliser CYP6B1, as with the metabolisers and template, there are also three domains. The domain motion of the FG loop is similar to CYP6M2 and CYP6P3, with a change in the dimensions of the FG loop, but as there are bulky residues in the FG loop the domain motion may act to partly open and close the channel.

6.3 b5 effects on isomers
In the template CYP3A4, the topology of the active site has been shown to be affected by the binding of b5. Methyl phenyldiazene carboxylate azo ester is used to obtain

topological information about the active site. When oxidised by a P450, the phenyl ring of the compound attacks an available nitrogen on one of the four heme pyrrole rings A, B, C or D to produce one of four regioisomers (N_A, N_B, N_C, N_D). The ratio of these regioisomers reflects the space available above each pyrrole ring. In the absence of b5 and P450 reductase CYP3A4 was found to primarily produce the N_A regioisomer. In the presence of b5 formation of all regioisomers was decreased with the effect smaller for N_A. There were changes in regioisomer profile for SRS mutants, in the absence of b5, the BC loop mutation S119W significantly enhanced all regioisomers but in the presence of b5 these increases were suppressed indicating that b5 binding causes a conformational change in the BC loop. This is consistent with the binding of b5 to residues of the C helix (Yamaguchi et al., 2004).

It is possible that the interaction with b5 could affect metabolism of different stereoisomers of permethrin in CYP6M2. From the dockings of the isomers it was found that the cis/trans positioning of the halovynal group affects binding of the isomers, while the R/S has less of an affect. With the trans isomers the halovynal tends to lie between the BC loop and the I helix while the cis isomers tend to lie between the BC loop and SRS5 or between SRS5 and I helix. In these modes, the alcohol moiety binds in a similar position and the structure of the acid moiety affects the position of binding of the halovynal group. However, with lower scoring modes, the halovinyl group can bind in alternative positions (Table 6.6 and Figure 2.12).

Table 6.6 ChemScores (kJ/mol) of the best ranked dockings and lower scoring alternative modes that differ in the position of the acid group.

Isomer	Position	score	rnk
(R)trans	between BC and I helix	46.4	1
(R)trans	between BC and SRS	43.3	7
(S)trans	between BC and I helix	44.4	1
(S)trans	between BC and SRS5	42.5	9
(R)cis	between BC and SRS5	42.3	3
(R)cis	between BC and I helix	42.0	4
(S)cis	between SRS5 and I helix	43.1	1
(S)cis	between BC and I helix	41.6	2

6.3.1 Normal mode and docking

To determine if the effect of b5s on domain stabilisation could have an affect on the

metabolism of permethrin isomers, the isomers were docked into normal mode perturbed models of CYP6M2. This method of docking has been previously suggested to represent receptor flexibility (Cavasotto et al., 2005). The flexibility of the receptor has been one of the problems in computational drug design. Dealing with receptor flexibility may be crucial to accurately predicting the binding orientation and interactions within the active site, the use of a single rigid target may reduce the chances of finding the correct binding mode of ligands. A number of methods have been proposed to address this problem including allowing side chain flexibility during docking (Leach, 1994), the use of multiple receptor conformations (Teodoro and Kavraki, 2003) or using an ensemble of structures collected from molecular dynamics (Lin et al., 2002) and normal mode analysis (Cavasotto et al., 2005).

There is evidence that there is a relationship between the pre-existing conformations of the receptor in the unbound state and structural changes upon ligand binding. In solution proteins exist in a range of conformations and their populations follow statistical distributions. Upon ligand binding the equilibrium may shift in favour of the bound conformation (Ma et al., 1999). Ma et al. (1999) proposed an alternative to the induced fit mechanism. They suggested that there may be pre existing subpopulations of conformations that preferentially bind to corresponding ligands. This proposal is an extension of the concept of energy landscapes and folding funnels. The energy landscapes in protein folding are represented by hills corresponding to high energy landscapes and valleys corresponding to favourable conformations and protein folding can be viewed as an ensemble of molecules going downhill through an energy funnel. If the bottom of the valley is smooth then the protein is expected to have only small changes in conformations, while if it is rugged there may be an ensemble of conformations.

Ma et al. (1999) suggested that this ensemble of conformations may be the origin of the induced fit upon binding. The 'lock' may exist as a range of conformations some of which fit the 'key'. As a ligand binds to the correct conformation the equilibrium may shift in favour of the bound conformation. Ma et al. (1999) suggested that these motions are primarily hinge bending domain motions. The domains being structural units that move with respect to another structural entity around a swivelling point, although the domains are not entirely rigid they are relatively rigid compared to the swivelling region. There may be selection at the hinge or at interdomain boundaries limiting the

extent of the motions and producing preferred motions. However, side chain flexibility or local motions may not be uncoupled from ligand binding and ligand binding can also be seen as a conformational selection stage of partly fitting structures followed by structural rearrangements, an induced fit stage.

As equilibrium conformations could be represented using normal mode analysis (NMA), Cavasotto et al. (2005) proposed using low frequency NMA to generate multiple receptor conformations to incorporate ligand flexibility in ligand docking. They selected modes relevant to ligand binding and perturbed along these modes. Floquet et al. (2006) found that docking using a single energy minimised structure produced incorrect binding poses. They generated a number of intermediate structures for a normal mode and found that ligands could be correctly docked. They showed that NMA could provide structures that may be more appropriate for docking.

6.3.2 Normal mode perturbed models

The CYP6M2 normal mode perturbed models were used for docking of the isomers. The normal modes define only the direction but not amplitude of the conformational change, reasonable amplitudes needs to be selected to prevent distortions and to reflect motions at room temperature (Suhre and Sanejouand, 2004). In all of the structures similar binding modes occurred although there were differences in scores (Table 6.7). Some structures show higher scores than others and may indicate a sub-population with a preference for permethrin as proposed by Ma et al. (1999). The effects of b5 to increase metabolism may be due to the stability of a conformation that improves binding of pyrethroids.

Table 6.7 ChemScores (kJ/mol) of dockings into normal mode perturbed structures (1 and 11) from modes 7 and 8. Isomers of permethrin were docked in each model.

Isomer	6M2	Mode 7_(1)	Mode 7_(11)	Mode 8_(1)	Mode 8_(11)
(R)cis	42.5	43.7	40.6	42.7	41.7
(S)cis	43.1	41.6	40.8	44.0	42.8
(R)trans	46.4	46.3	45.3	46.9	44.6
(S)trans	44.4	43.9	45.0	45.4	43.1

6.4 DDT metabolism

The preference of P450s to de-halogenate aliphatic-halogenated hydrocarbons occurs in the order iodine> bromine> chlorine >fluorine, with the strength of the halogen-carbon

bond increasing in the same order. This F-C bond is considered to be inert and difficult to break and has been used to prevent biodegredation of compounds. With aromatic-halogenated hydrocarbons, the order of elimination is reversed, fluorine > chlorine > bromine > iodine. This may be due to the higher electronegativity of fluorine. Dehalogenation can occur either by a one or two electron reductive pathway or by oxidative hydroxylation followed by a loss of the halogen by non-enzymatic destruction of an unstable intermediate (Cnubben et al., 1995).

DDT can be dechlorinated to DDD under anaerobic conditions. Both ligand binding and metabolism has been shown to be enhanced in sodium dithionite reduced P450. DDT may directly coordinate at the 6th coordination site with the reduced heme iron to accept electrons directly. This reductive metabolism of DDT may be due to a non-enzymatic interaction as reduced heme in solution can catalyse the formation of DDD (Baker and Van Dyke, 1984) and dehalogenation is increased in boiled microsomes as there may be greater access to the heme. It has been shown that under aerobic conditions halogen compounds bind to the protein part of the P450, but under anaerobic conditions they change to type II substrates after reduction by NADPH (Fujii, 1995).

In rats, the pathway of DDT metabolism involves the non-enzymatic reduction of DDT to DDD. This is the first and rate limiting step in DDT metabolism. Human CYPs 2B1, 3A1, 2B6 and 3A4 have been shown to dechlorinate DDT. Some activity is also shown with CYPs 2C9, 1A1/2, 2A6 and 2D6 (Kitamura et al., 2002). The P450 is reduced by P450 reductase, while under normal conditions molecular oxygen would bind to the iron. During reductive dehalogenation a second electron reduction can occur with a rapid transfer of an electron to the halogenated hydrocarbon. This eliminates a halogen ion and a radical. Diclofol is a metabolite of DDT in which there is hydroxylation of the methylene bridge carbon. In rat liver microsomes, unlike DDD formation, this process is oxidative (Kitamura et al, 2002).

A number of residues have been identified as being involved in DDT metabolism. While the wild type *Drosophila melanogaster* CYP6A2 does not metabolise DDT, the point mutations R335S, L336V and V476L increase both the affinity for, and metabolism of DDT. This CYP6A2vSVL mutant produces the metabolites dichlorodiphenyldichloroethane (DDD), dichlorodiphenyl acetic acid (DDA) and dicofol. As these point mutations are far from the active site and are clustered around

222

the distal end of the I helix, It was suggested that these mutations affect the structure of the protein (Amichot et al, 2004).

6.4.1 CYP6M2 DDT metabolism

CYP6M2 has been shown to metabolise DDT to produce the metabolites dicofol and DDE (Figure 6.13), the ratio of the metabolites are about 2:1 without b5 and 20:1 with b5. DDT is metabolised to DDE by dehydrochlorination and dicofol by hydroxylation. Assuming that dechlorination to produce DDE is reductive, reductive dechloriantion involves a single electron transfer provided by the reductase, while hydroxylation to produce dicofol would involve a second electron from b5. The effect of b5 on metabolite ratio could be due to its role as a second electron donor. Alternatively, as DDT stacks with F123, any conformational change of the BC loop could affect the binding of DDT and regiospeicifity.
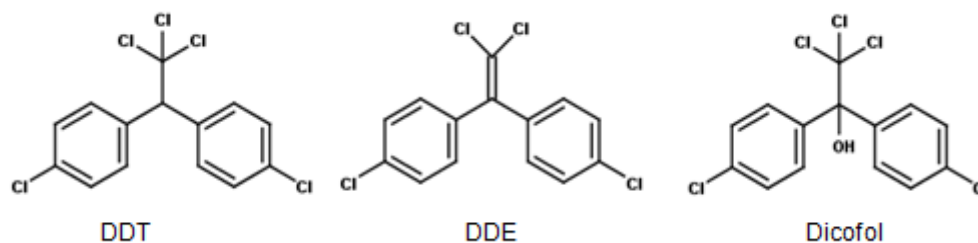
Figure 6.13 Structures of DDT, DDE and Dicofol.

6.4.2 CYP6Z2 and CYP6P3 non-metabolisers

CYP6Z2 and CYP6P3 have been shown to be incapable of metabolising DDT. The CYP6Z2 model binds DDT distantly due to restrictions in the active site due to L365 and F115. DDT can also stack with F112 which may contribute to holding the ligand distantly (Figure 6.14, A). CYP6P3 is able to bind DDT for metabolism (Figure 6.14, B), but this score is reduced compared to CYP6M2 due to a low lipophilic score (Table 6.8). The position of V380 in CYP6P3 increases the space within the active site that reduces the contacts with DDT while L365 in CYP6Z2 reduces the space available for docking.
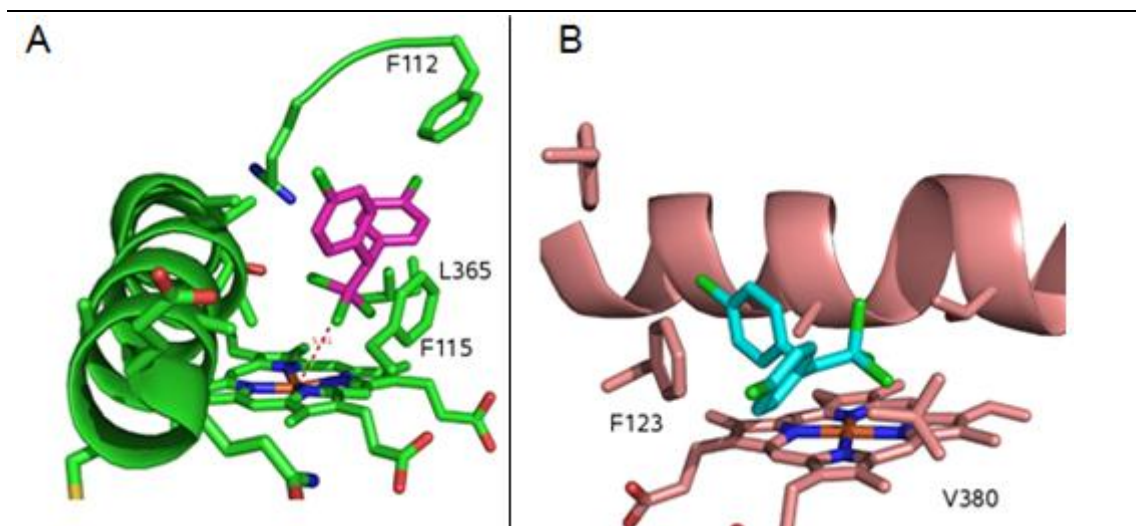
Figure 6.14 CYP6Z2 (A) and CYP6P3 (C) binding DDT.

Table 6.8 ChemScores (kJ/mol) for DDT in CYP6M2, CYP6Z2 and CYP6P3.

	Score	ΔG	S(hbond)	S(metal)	S(lipo)	H(rot)	ΔE(clash)	ΔE(int)
6Z2	39.8	-40.7	0.0	0.0	322.6	1.0	0.1	0.8
6P3	42.6	-46.1	0.0	0.0	369.4	1.0	2.6	0.9
6M2	47.0	-47.9	0.0	0.0	384.2	1.0	0.5	0.4

6.4.3 P450 activation

P450 activation can occur by several mechanisms. Huang et al. (1981) found that flavanoids activate B[a]P hydroxylation by enhancing the interaction between P450 and the reductase facilitating the flow of electrons. Lee et al. (1997) found that caffeine and 7,8-benzoflavone activates CYP3A2 and that this can be inhibited by replacing NADPH with CHP, but the activation by caffeine can be blocked by b5 antibodies. The activation by caffeine is mediated by b5 transfer of the second electron. The combination of caffeine and 7,8-benzoflavone caused less activation than when they were included separately and indicates that they operate by different and antagonistic mechanisms with simultaneous binding preventing the optimal interaction of either molecule with the protein.

P450s can also be activated by substrates. Shou et al. (1994) found that phenanthrene metabolism can be activated by 7,8-benzoflavone and that this also acts as a substrate. Both occupy the active site simultaneously and have access to the oxygen as neither competitively inhibits the other. The Km of phenanthrene was not affected by the presence of 7,8-benzoflavone indicating that one does not affect the binding affinity of

the other and do not displace each other. Shou et al. (1994) suggested that the activator altered the dimensions of the active site.

6.4.4 Sodium cholate activation of CYP6M2

In CYP6M2 DDT metabolism only occurs in the presence of sodium cholate (Figure 6.15), but the chemically similar cholesterol does not activate CYP6M2. Cholate has an optimal concentration of 1mM and high concentrations of cholate, over 10mM, inhibit both the metabolism of DDT and deltamethrin (McLaughlin, unpublished). The presence of sodium cholate may activate CYP6M2 to metabolise DDT but the mechanism of activation is unclear.
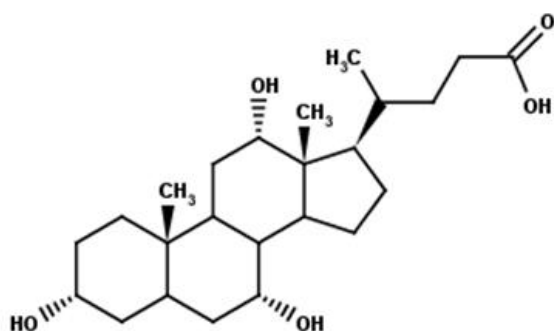
Figure 6.15 Structure of sodium cholate.

Sodium cholate has been found to show similar effects in other enzymes. Its effects on phospholipase C are biphasic. It enhances rates up to 0.3-0.6 mM but is inhibitory above this rate. The lag time decreases monotonically with concentration. Ruiz-Arguello et al. (1999) hypothesised that this effect may be caused by a stimulation of enzyme activity during the lag phase and an inhibition of the post lag active form. Sotiroudis et al. (1983) found that the effect of cholate on phosphorylase b at first increases than decreases activity as the concentration is raised. This effect was specific as other bile acids and detergents were ineffective. They suggested that this effect was caused by an allosteric shift towards an active conformation with the cholate binding at hydrophobic pockets. They suggested two classes of cholate sites. The first class are saturated at a low concentration and are responsible for activation. A second class are saturated at higher concentrations. Jacobson et al. (1990) suggested that bile salts can cause a conformational change in pancreatic cholesterol esterases, possibly by binding at specific sites in the enzyme. At concentrations of 70-100 nm there was a conformational change. Higher concentrations caused a second conformational change

Sodium cholate may also have an effect on CYP6M2 by altering the membrane. Cholate

225

has been suggested to stabilise the lamellar phase over the hexagonal phase of a monolayer, but also alter the curvature in the direction of the hexagonal structure. Its effects could be attributed to the perturbation of the lamellar phase or be modifying the electric charges in the bilayer (Ruiz-Arguello et al., 1999). While sodium cholate stimulates metabolism by *Anopheles* CYP6M2, it is not an endogenous compound produced by the insect. However, there are other derivatives of cholesterol, such as ecdysone, that are produced by insects that may have similar endogenous roles.

6.5 Membrane interactions

The membrane bound members of the P450 superfamily have been proposed to interact with the membrane by a common membrane binding domain, the N-terminal trans membrane helix. Additional membrane attachment surfaces have been previously proposed because P450s lacking the transmembrane helix are still able to associate with membranes.

In CYP2C5 the membrane binding surface has been predicted to be formed by N terminal residues 30-45, 60-69, and by β-strand 2-2 residues 376-379 and by the C terminal end of the FG loop. If this is the case, this would place the opening of a substrate channel into the membrane that could be favourable for the entry of lipophilic substrates (Williams et al., 2000a). Residues in these regions are more hydrophobic than those of the soluble microsomal enzymes. Compared to the soluble BM3 the FG loop of CYP2C5 has 12 hydrophobic residues verses the 5 in BM3. The binding of a P450 to a membrane has an affect on substrate metabolism, in CYP3A4, there are differences in binding kinetics between soluble and membrane bound CYP3A4 (Nath et al., 2007). Substrate access channels may have an affect on substrate selectivity. CYP2C9 prefers small acidic lipophilic substrates while the related CYP2C19 shows no preference for acidic compounds. CYP2C9 has no basic residues in the active site to account for this preference but in an access channel CYP2C9 has Lys72 while CYP2C19 has a glutamate (Williams et al., 2003).

Pyrethroids are hydrophobic ligands and P450s that metabolise pyrethroids may have a preference for such ligands. As it has been suggested that access channel or membrane binding features may affect ligand preference, in this study such features are investigated to explain preferences towards hydrophobic ligands. Lewis et al. (2004) reviewed the preferences of P450s towards their substrates and found that the logP of a

substrate can give an indication of its P450 selectivity (Table 6.9). P450s such as CYP2A6 and CYP1A2 prefer hydrophilic ligands while others such as CYP3A4 and CYP2C8/9 prefer hydrophobic substrates. Although substrate access may have an influence on selectivity, Lewis et al. (2004) suggested that the preference for logP may be due to variations in hydrophobicity of each active site and could provide a measure of the degree of hydrophobicity in the heme environment.

Table 6.9 The average logPs of substrates of P450s.

CYP	average logP (Lewis, 2000)	average logP (Lewis et al., 2004)
1A1	4.51	3.41
1A2	1.57	2.01
2A6	1.66	1.41
2C8	Not determined	3.38
2C9	3.15	3.2
2C19	2.12	2.56
2D6	3.18	3.08
2E1	0.63	2.07
3A4	2.94	3.10

Table 6.10 Membrane interaction predictions from the OPM database. The average logP of substrates determined by Lewis et al. (2004).

	PDB	log P	depth in membrane (Å)	ΔG transfer (kcal/mol)	tilt angle (°)
2A6	1Z10	1.44	6.3 ± 1.0	-10.3	38 ± 12
1A2	2HI4	2.01	6.2 ± 1.0	-16.9	68 ± 5
2D6	2F9Q	3.08	8.7 ± 1.4	-16.8	34 ± 4
3A4	1TQN	3.1	10.2± 1	-20.7	61 ± 11
2C9	1R90	3.2	7.5 ± 1.0	-10.7	57 ± 10
2C8	1PQ2	3.38	7.6 ± 0.9	-9.4	20 ± 10

Table 6.11 MAPAS membrane interaction predictions.

		Memb.	MRS	MAS	Kmpha	
2A6	1Z10	21	3.82	40.20	1.13	
1A2	2HI4	19	3.78	40.09	1.42	*
3A4	1TQN	41	4.03	38.00	0.93	*
	1W0F	35	4.20	52.49	1.27	
2C9	1OG5	33	4.11	47.08	1.25	*
	1OG2	32	4.07	45.57	1.24	-
2C8	2NNI	32	4.42	52.63	1.32	-
	1PQ2	19	4.37	54.67	1.37	*
2D6- none found						

Memb: the number of residues in contact with the membrane (within +- 0.5 Å of the membrane). MRS: membranephilic residues score. MAS: membranephilic area score. Kmpha: membranephilic asymmetry.
*Visually superimposable to the opm prediction.- no prediction available in OPM. These predictions are for the crystal structures where the N terminus is deleted or replaced by hydrophilic residues.

The OPM (Orientations of Proteins in Membranes) (Lomize et al., 2006) database was accessed for data on the binding of the crystal structures to membranes (Table 6.10) and MAPAS was used to predict the number of residues in contact with the membrane and to predict the plane of attachment (Table 6.11).

6.5.1 MAPAS and OPM validation

To determine how suitable MAPAS and OPM are for the prediction of membrane interactions for P450, their outputs were compared to experimental results. CYP2C5 has a hydrophobic patch formed by residues 30-45, 60-69, 376-379. Proteolysis liberates a peptide 75-C terminus, and this region clustered around P30 corresponds to a cluster of epitopes that only bind with antibodies with soluble CYP2s but not the membrane bound proteins. In addition residue 46 of CYP1A2 has been identified as embedded in the membrane. This corresponds to residue 37 in CYP2C5 (Williams et al., 2000a). The OPM server identified these hydrophobic residues as contacting the membrane with both the CYP2C5 structures (1DT6 and 1NR6) but with different orientations in the membrane. As the two structures are similar, the difference in the prediction between the structures may occur because the 1DT6 structure is missing the F' and G' helices that are present in 1NR6. The HotPatch prediction for lipid interaction using both the neural network and hydrophobicity identified a hydrophobic patch buried by the OPM membrane for both 1NR6 and 1DT6.

6.5.2 Correlation between membrane attachment and logP

While there is no clear correlation between logP and depth of transfer or tilt, P450s with a preference for hydrophobic substrates have a larger number of residues in contact with the membrane, and are predicted to have a larger area of contact on the membrane than those that prefer hydrophilic substrates (Table 6.12).

Table 6.12 Correlations between membrane interaction scores and logP.

Score	R
#residues	0.83
MRS	0.85
MAS	0.59
Kmpha	-0.1
depth	0.67
ΔG	-0.9
tilt	-0.22

Table 6.13 Number of residues and surface area in contact with the membrane predicted by HotPatch.

P450	Number of residues	Area of patch (Å^2)
2A6	11	772
1A2	14	1293
2D6	7	687
3A4	25	1641
2C9	28	1894
2C8	25	1857

HotPatch was also used to identify hydrophobic patches on the surfaces of the proteins. In HotPatch, for each atom, a value of the property score x is assigned and patches are identified by a clustering algorithm to find patches of high x. HotPatch indicated that P450s that prefer hydrophobic ligands also have a larger hydrophobic patch involving a larger number of residues (Table 6.13).

The surface of the hydrophobic patches identified by HotPatch were consistent with the membrane interaction predicted with MAPAS as the hydrophobic surface is buried by the MAPAS membrane surface (Figure 6.16). However, HotPatch was unable to differentiate between the surface of the protein and the interior active site and access channels. The hydrophobic surface identified in CYP3A4 was not restricted to the surface but also lined the channel between the F' and G' helices and the active site (Figure 6.16, A and B). Similarly, in both CYP2C8 and CYP2C9 the hydrophobic patch extends through the access channel. By contrast, the hydrophobic patches of CYP1A2 and CYP2A6 are restricted to the area of membrane contact and does not extend along the access channel (Figure 6.16, C and D). These results may indicate a failure of HotPatch to correctly differentiate between the hydrophobic membrane binding surface and interior and may be unsuitable for such an analysis. However, the identification by HotPatch that the surface hydrophobic patch extends through the access channel in P450s that bind hydrophobic ligands, may indicate adaptations within the access channel for the transfer of these ligands to the active site. P450s that prefer hydrophilic ligands lack this feature.
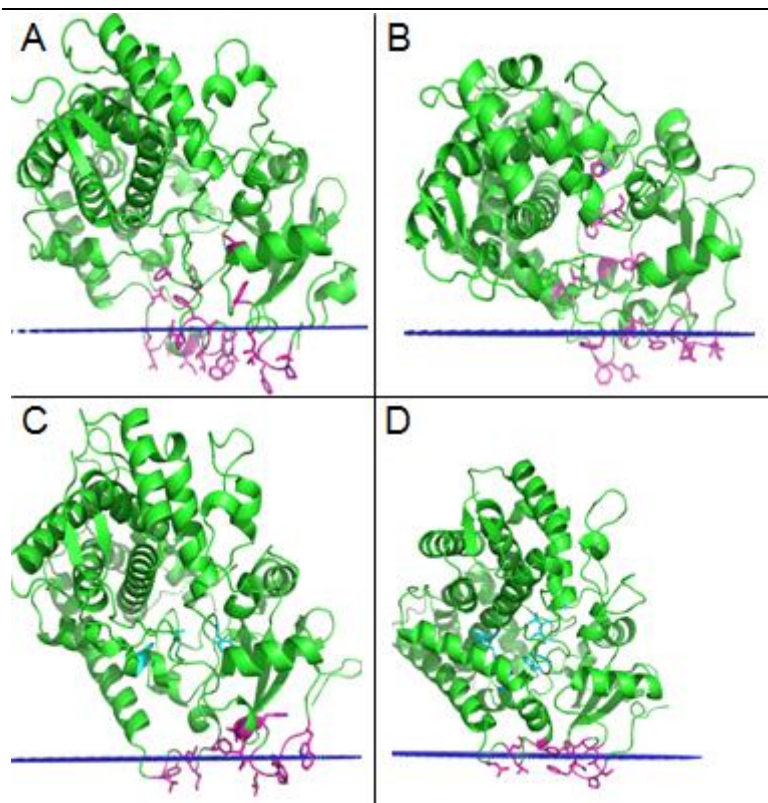
Figure 6.16 (A) CYP3A4 (1TQN), (B) CYP2C9 (1OG5), (C) CYP1A2 (2HI4) and (D) CYP2A6 (1Z10) with the MAPAS predicted membrane surface (blue line) and HotPatch predicted hydrophobic patch (pink). In both A and B, the hydrophobic patch extends into the active site. In C and D the hydrophobic patch is restricted to the site of membrane contact and is separate from the hydrophobic patch within the active site (pale blue).

6.5.3 Structural Features

There may be structural features affecting the interaction with the membrane which could affect ligand access. The FG loop membrane contacts of CYP2A6 (1Z10) and CYP1A2 (2HI4) are small with openings of the access channel exposed to the solvent. The structure of the F and G helices, such as the disruption of the F helix in CYP1A2, may be a structural adaptation that acts to reduce the contact with the membrane to facilitate ligand access from the solvent (Figure 6.17, A). The structures of CYP3A4, CYP2C8/9 have larger contacts between the membrane and FG loop and the opening of the access channel is exposed to the membrane, the structures of the FG region may be adapted for this placement. In CYP2C8/9 the disruption of the F and G helices into the F' and G' helices orients the mouth of the access channel towards the membrane (Figure 6.17, B), while in CYP3A4 the F' and G' helices are in contact with the membrane and part of the access channel opens directly onto the membrane (Figure 6.17, C).

230
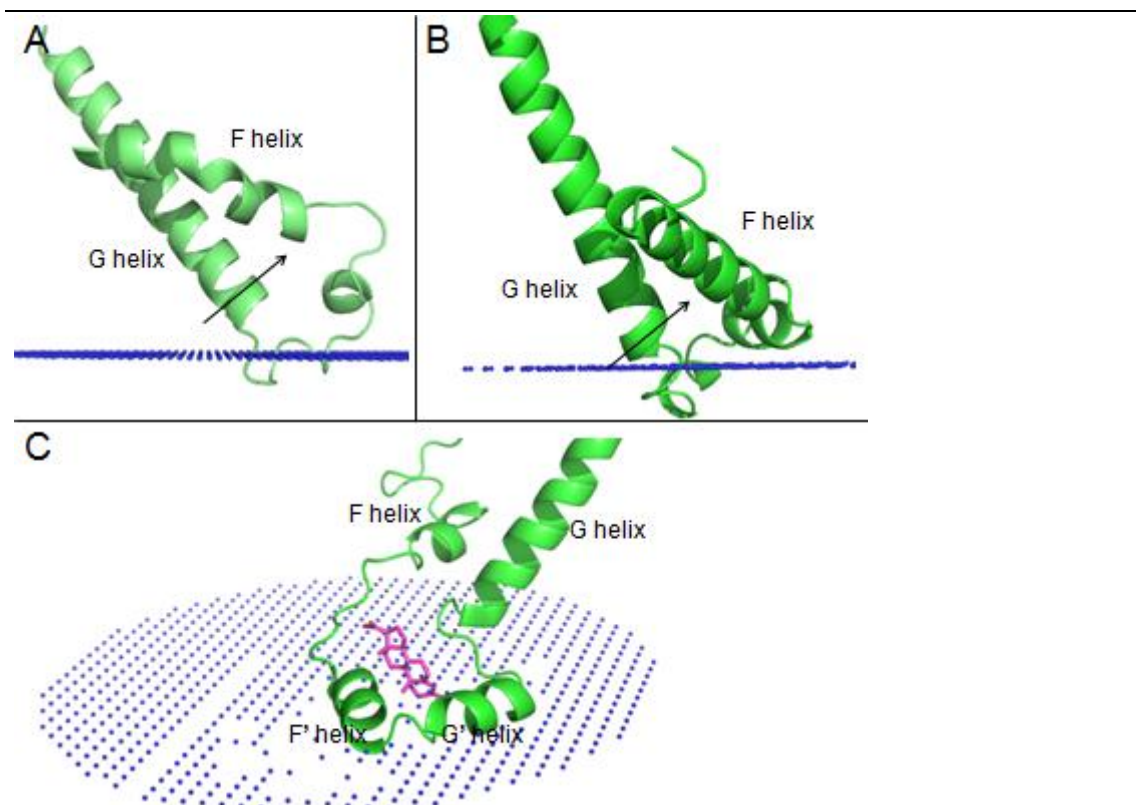
Figure 6.17 The FG regions of CYP1A2 (2HI4) (A) and CYP2C8 (2NNI) (B), the entrance to the channel is between the F and G helices (arrow), in CYP1A2 the entrance is oriented towards the solvent, in CYP2C8 it is oriented towards the membrane. (C) The FG region of CYP3A4 (1W0F) binding progesterone (pink). The F' G' helices place the mouth of the access channel onto the membrane. The plane of the membrane is predicted using MAPAS (blue plane).

This is consistent with the suggestion by Williams et al. (2004) that the FG loop acts to orient the substrate access channel towards the membrane surface. Williams et al. (2004) found that in the crystal structure of CYP3A4 (1W0F) the binding of progesterone may be at a substrate recognition site as it lies along an access channel along which substrates from the can move directly from the membrane into the active site (Figure 6.17, C). A similar recognition site has also been proposed in P450cam by Ludemann et al. (2000). It was suggested that before being able to enter the active site, camphor will weakly associate with hydrophobic crevices on the surface of the protein, such as the hydrophobic patch near pw2a and pw3. Ligand access from the membrane has also been proposed by Zhao et al. (2006). Their crystal structure of CYP2B4 was predicted to have the FG and BC loops embedded in the membrane and had an access channel connecting the hydrophobic core of the membrane and the active site. Zhao et al. (2006) suggested that the most conformationally flexible regions of CYP2B4, the BC

231

loop and FG regions, are embedded in the membrane and that the conformations of the membrane binding regions could be dependent on the lipid environment of the membrane, and the conformational plasticity of the membrane binding region could encounter, recognise and deliver substrates to the active site.

6.5.4 FG loop mutations

The interaction between the FG loop and ligand access from the membrane is consistent with previous studies where the FG region has been shown to be simultaneously a site of membrane binding and ligand access where the substrate is not soluble in water. This has been demonstrated in P450s that metabolise cholesterol (CYPs 27A1, 11A1 and 7A1). In CYP27A1 and CYP11A1, mutations of the F and G helices have been shown to alter the metabolites produced. The CYP27A1 mutants I211K and F215K produced a new metabolite but were located outside of the active site as mutations did not alter dissociation constants (Kd) (Pikuleva et al., 2001). This mutagenic data is consistent with an access channel passing between the F and G helices and the residues lining the channel orienting the substrate before it arrived in the active site. It was suggested by Pikuleva et al. (2001) that these residues may have a structural role with steric restrictions only allowing access to the ligand in certain orientations producing regio- and stereospecificities. The physicochemical properties of the FG region have also been shown to affect ligand access. Replacements in CYP27A1 of residues lining the FG region access channel (F207, I211 and F215) with polar residues had a larger affect than mutations to alanine. Pikuleva et al. (2001) suggested that this may be due to steric hindrance, or may be due to making the hydrophobic access channel more polar with the substrate (cholesterol) less likely to partition into the active site.

Non conservative mutations of membrane binding residues (L219N, Y220A, Y220S, F223A, K226R, R229A) of the CYP27A1 FG loop impaired activity, showing that surface residues outside the active site can affect catalysis inside the active site. Murtazina et al. (2002) suggested two explanations for this, firstly that the weakening of the membrane interaction increases water access to the active site. Water acts as a proton donor and can affect the proton transfer pathway to the bound oxygen affecting catalysis. Secondly, they suggested that alteration of protein-membrane interactions could slow product exit to decrease Kcat.

Nakayama et al. (2001) found that there may be two types of residues involved in

membrane binding, those that interact with the membrane and those that prevent deep insertion into the membrane with the size of the side chains having a role. In CYP7A1 mutations (F215A, L218A, L218N, L218V, I224A, F227A) resulted in a tighter binding to the membrane. They suggested that these residues also control the depth of insertion into the membrane. Three of the mutants that showed either no activity or reduced activity (L218N, I224A, F227A) had altered membrane interactions as they were bound more tightly than the wild type and the decrease in activity may be due reduced amounts of substrate reaching the active site.

6.5.5 Candidate membrane interactions

Pyrethroid metabolising and non metabolising P450s were also tested with MAPAS. Similar to CYP3A4, the FG loops of all candidate P450s were in contact with the membrane with the mouth of the access channel oriented towards the membrane.

Table 6.14 MAPAS scores for pyrethroid metabolising and non-metabolising (*) P450s.

P450	Memb.	MRS	MAS	Kmpha
6M2	38	4.43	35.54	1.02
6B8	18	4.80	42.25	1.14
6P3	43	4.20	37.77	1.10
6P9	43	4.39	45.64	1.28
9J32	39	4.12	39.50	1.18
*6B1	16	4.03	42.12	1.09
*6CB1	40	4.01	33.26	1.04

Memb: the number of residues in contact with the membrane (within +- 0.5 Å of the membrane). MRS: membranephilic residues score. MAS: membranephilic area score. Kmpha: membranephilic asymmetry.

From the MAPAS scores (Table 6.14), there is no relationship between the number of residues in contact with the membrane and ability to metabolise pyrethroids. However, the MRS score is higher in the metabolisers than in the non-metabolisers indicating a more membranephilic surface. This may indicate that binding a hydrophobic substrate such as deltamethrin may require other general adaptations for the binding of hydrophobic ligands such as an increased contact with the membrane.

6.6 Access channels

As pyrethroids are hydrophobic ligands, they may access the P450 from the membrane. P450s that prefer to metabolise hydrophobic ligands may have features of their access channels to promote ligand access form the membrane, pyrethroid metabolising P450s may share these features for the access of hydrophobic ligands such as pyrethroids. Examining the orientations of access channels with respect to the membrane for P450s

233

with a range of logP preferences can be used to identify these features.

Access channels were initially named based on a Thermal Motion Pathway (TMP) analysis of P450cam. In TMP, chains of atoms with elevated temperature factors are used to indicate there channels could open. Access channels have also been identified and named by Wade et al. (2004). To determine how membrane interactions could facilitate the binding of hydrophobic ligands from the membrane, the location of the access channels with respect to the membrane were examined. MOLE was used to identify the locations of access channels and named according to Wade et al. (2004).

The three channels identified in CYP2A6 open into the solvent, but with channel 2c closest to the membrane. For a hydrophobic ligand dissolved in the membrane to pass into the active site, it must first pass into the solvent (Figure 6.18, A). This is consistent with the CYP2A6 preference for hydrophilic ligands. In CYP1A2, the prediction for the plane of the membrane differ between MAPAS and OPM, but with either membrane position, the channels open into the solvent (Figure 6.18, B) also consistent with its preference for hydrophilic ligands. In CYP2C8, the solvent channel opens close to the membrane (Figure 6.18, C), and in CYP3A4, MOLE identifies the pathway Pw2b as opening onto the membrane (Figure 6.18, D). This is consistent with the preference of CYP3A4 and CYP2C8 for hydrophobic ligands.

There appears to be a correlation between the placement of access channels and the preference for the logP of substrates. CYP2A6 and CYP1A2 prefer hydrophilic substrates and have access channels that open into the solvent while CYP2C8 and CYP3A4 that prefer more hydrophobic substrates have channels that open into the membrane. This preference for logP may also explain the ability or lack of ability of a P450 to metabolise hydrophobic ligands such as pyrethroids. CYP2A6 and CYP1A2 are also poor or non metabolisers of pyrethroids while CYP2C8 and CYP3A4 are metabolisers. In the CYP6 pyrethroid metabolisers, a channel is possible through the FG loop. This channel in CYP6M2 opens onto the membrane and may contribute towards its ability to metabolise pyrethroids by allowing its access from the membrane to the active site.
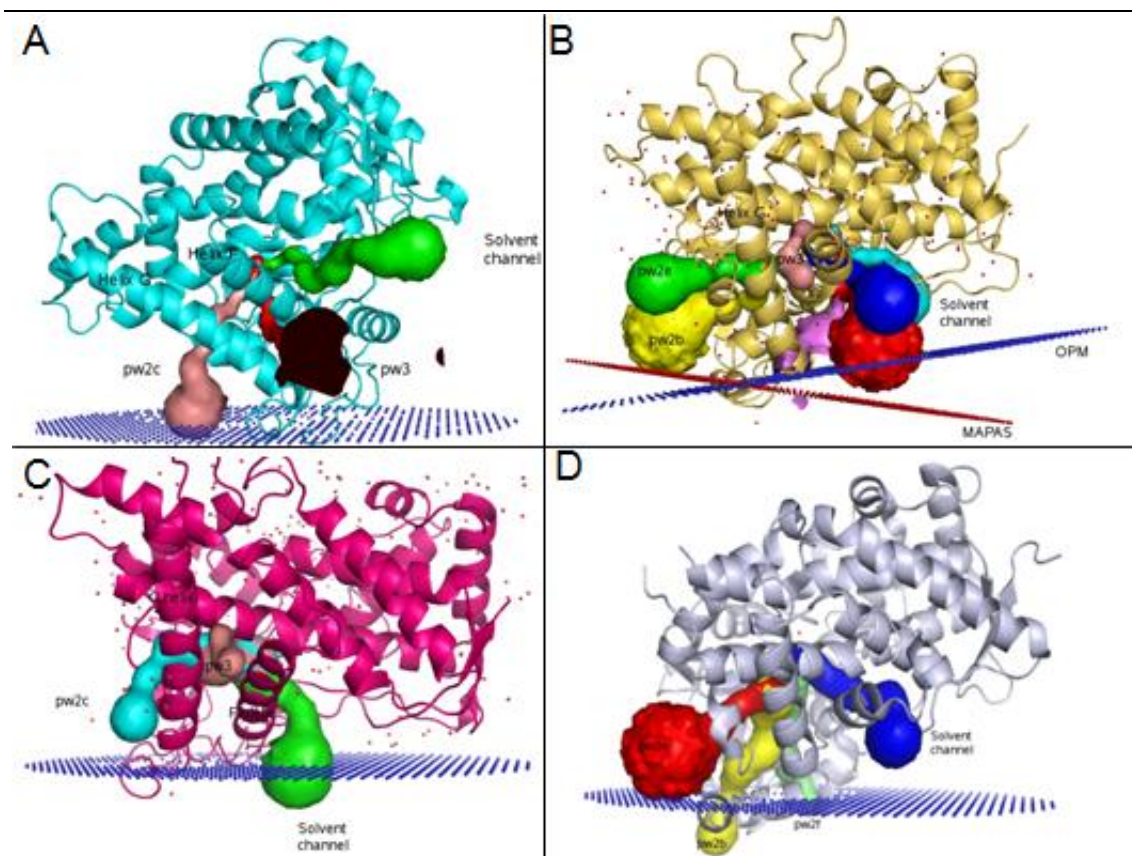
Figure 6.18 (A) CYP2A6, (B) CYP1A2 (2HI4), (C) CYP2C8, (D) CYP3A4 (1TQN) with the membrane position identified by MAPAS and OPM (dotted plane), and the access channels identified by MOLE.

These results are consistent with previous studies suggesting entry from the membrane of hydrophobic ligands. Pw2a has been suggested to be a channel for hydrophobic substrates as the protein around the entrance of pw2a, the FG loop, is hydrophobic and dips into the membrane. Egress pathways are likely to differ as the water soluble products are likely to be released into the cytoplasm (Winn et al., 2002). Schleinkofer et al. (2005) proposed that there may be alternative mechanisms of ligand access depending on the cellular location of the substrate. The substrates of membrane bound P450s are hydrophobic and are likely to come from the membrane. As the products are hydrophilic they are likely to be released into the aqueous environment. The predominant pathway for access and exit are likely to differ between P450s. In soluble bacterial P450s, access was found to be most common through pw2a. In the CYP2C5 chimera, pw2c was found to be the predominant pathway. They suggested that in CYP2C5, lipophilic substrates could enter from the membrane through pw2a and leave by pw2c, while soluble substrates could enter and leave by pw2c (Figure 6.19). The extent to which pathways are used could depend on the nature of the substrate and

235

the interactions of the membrane that could affect the opening of channels.
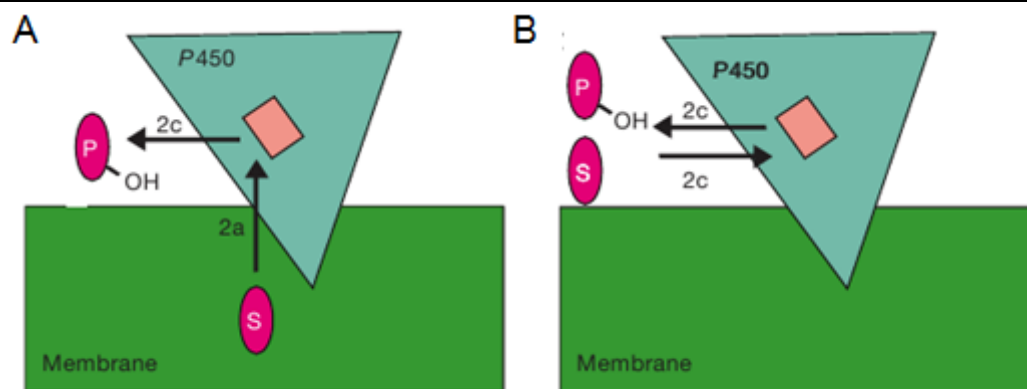
Figure 6.19 Taken from Schleinkofer et al. (2005), (A) lipophilic substrates enter from the membrane through pw2a and products leave through pw2c. (B) soluble substrates enter and products leave through pw2c.

6.6.1 Normal modes and channel gating

Channels in the protein allow substrates to enter and products to leave the buried active site. Protein motions are known to be essential for substrate access and product egress as they can cause the opening of channels as well as changing the dimensions of the channels. This has been shown by Winn et al. (2002) using random expulsion MD. They found that in both P450cam and P450eryF breathing motions of the enzyme and induced fit opens channel pw2a. The FG region is lined by the most commonly observed channels (2c, 2a, 2f, solvent channel) a number of studies have shown that channels in this region can open to allow egress, with opening dependant on the position of the FG region. Cojocaru et al. (2007) reviewed the mechanisms of channel gating and suggested that channels could be classified based on gating mechanism. The opening of channels 2c, 2ac, 2a, 2b, and 2f require the relative movement of secondary structure elements, while channels passing through loops such as 2e, 2d and 4 require a conformational change of the flexible BC or FG loop they pass through. Cojocaru et al. (2007) suggested the opening of channels surrounding the FG region can merge to provide a larger channel and that the insertion of the FG region into the membrane could facilitate opening of channel pw2a. There may also be structural rearrangements upon membrane binding as different CYP2B4 crystals show differing secondary structures with plasticity of the FG and BC loops. In one structure the F' and G' helices are combined into a single helix

From the NMA of the CYP6M2 model, it was found that the opening and closing of

channels occurs during normal mode 7. There is an opening of pw2f, pw4 and the solvent channel with a narrowing of pw2b. The use of algorithms such as MOLE may not be accurate at identifying possible access channels in P450s as this does not account for dynamic gating mechanisms.

6.7 CYP6M2 SNP Mutants

A number of single nucleotide polymorphisms (SNPs) have been found in wild *A. gambiae* populations (McLaughlin, unpublished). Although most SNPs are disadvantageous, it is possible that some SNPs may have a selective advantage in metabolism. The positions of the SNPs and possible affects were predicted based on the location on the model (Table 6.15).

Table 6.15 List of SNPs identified in CYP6M2 (McLaughlin, unpublished), their position on the model and their possible roles.

SNP[1]	Position[2]	Possible role
11[V/M]	N terminal loop	Membrane interactions
65[D/A]	A helix	
68[G/V]	A helix – β1-1 loop	
240[T/I]	G'G loop	Access channel (channel 3)
285[D/G]	H-I loop	
328[E/Q]	J helix	
359[I/V]	K helix	
382[H/Y]	β1-4 β1-3 loop	Access channel (channel 2b)
392[A/S]	β1-4 β1-3 loop	Access channel (channel 2b)
407[P/L]	K' K" loop	
409[V/E]	K'K" loop	
412[N/T]	K'K" loop	
474[V/I]	C terminal loop	

[1] position on the wild type CYP6M2 database sequence AAAB01008964.
[2] position in the model.
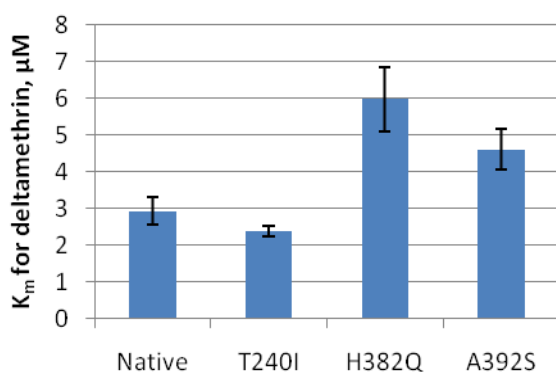
Figure 6.20 The effects of CYP6M2 SNPs on the Km of deltamethrin (McLaughlin, unpublished).

The positions of access channels were predicted using the MOE sitefinder tool to identify cavities within the model. These were compared to the positions of access

channels identified by Wade et al. (2005). The SNPs T240I and A392S occur at the mouths of substrate access channels 3 and 2b and one of these may project into the active site. Both of these SNPs involve a transition from a polar to a non-polar residue. The third, H382Y is close to an access channel. All but one of the SNPs are located at the surface of the protein, while I359 is buried. The three SNPs located near to the access channel were tested by mutagenesis and were found to alter the kinetics (Figure 6.20). All mutants appear to be active and exhibit the same instability in membrane preparations as observed for native CYP6M2. H382Q and A392S appear to have reduced the affinity for deltamethrin while T240I increases affinity.

6.7.1 T240I

T240I occurs on the G' G loop and is situated at a putative access channel that is identified by MOE sitefinder (Figure 6.21). This channel, corresponding to the channel 3 identified by Wade et al. (2005), occurs between the F and G helices. In the CYP3A4 crystal structure 1TQN a channel through the FG loop is closed off by a network of aromatic residues (Yano et al., 2004) but in CYP6M2 this network is substituted and a channel is possible (Table 6.16).
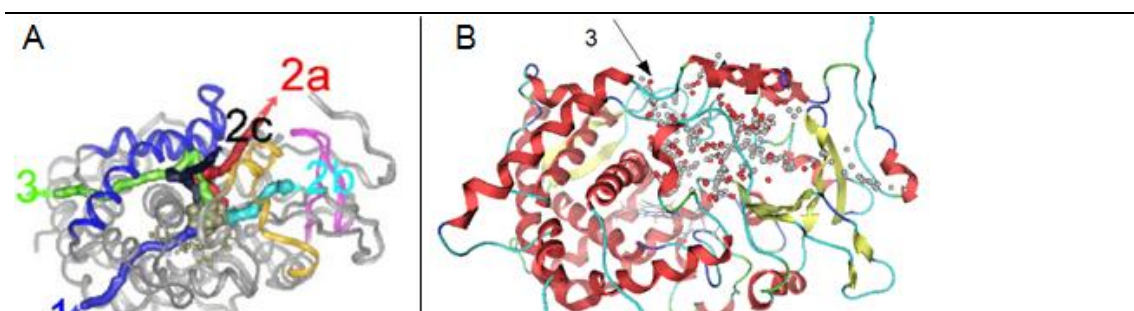
Figure 6.21 (A) egress routes identified by Wade et al. (2005). (B) Access channel 3 (arrow) identified by MOE alpha site finder tool.

Table 6.16 Residues closing the FG loop in CYP3A4 that are substituted in CYP6M2 to create an additional access channel.

P450	Residues				
CYP3A4[1]	F108	F213	F215	F241	F304
CYP6M2[2]	F108	I215	I215/S216	T240	L303

[1] residues closing the channel through the FG loop in the CYP3A4 structure 1TQN, determined by Yano et al. (2004).
[2] residues of CYP6M2 aligning with CYP3A4 determined from a pairwise alignment and an alignment of structures.

In an alignment of CYP6M2 and CYP3A4, T240I aligns with F241 of CYP3A4. In the crystal structure this residue packs closely with other aromatic residues to close off the channel in the crystal structure 1TQN (Yano et al., 2004) while in the model this channel is open. This residue occurs on an access channel which opens onto the

membrane, and the position of T240 could be involved in the transfer of hydrophobic ligands from the membrane into the active site (Figure 6.22). F241 of CYP3A4 projects into the active site of the crystal and therefore T240I is also predicted to project into the active site. In the model T240I is involved in H-bonding with luciferins, and may be involved in ligand or product entry or egress, substrate recognition, or may be involved in forming an intermediate binding position. The mutation of a small polar T for a larger aliphatic I may affect access though this channel.
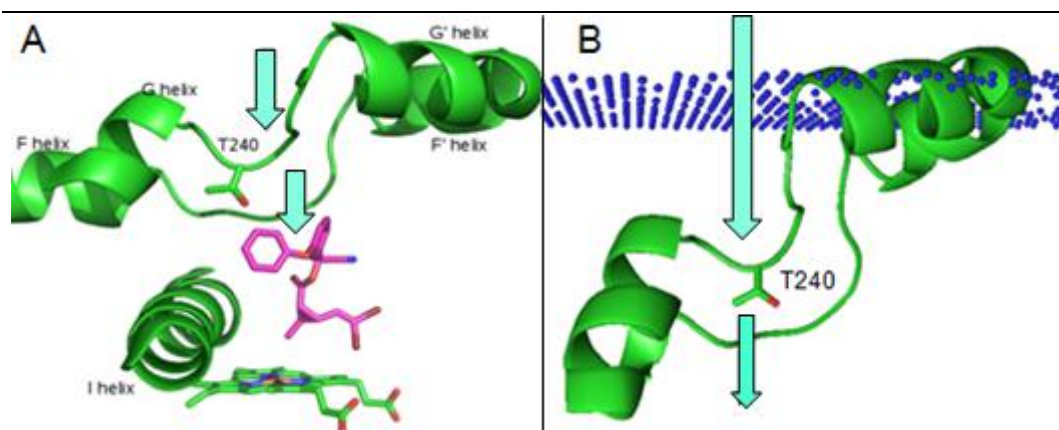
Figure 6.22 (A) CYP6M2 binding deltamethrin. The SNP T240I occurs in the G'G loop and lines access channel 3 in a position to affect access of ligands (arrows). (B) The position of T240 and access channel 3 (arrows), the channel opens onto the MAPAS membrane (blue plane) to allow access of ligands.

This sidechain also projects into the active site in a position to contact the substrate. Alterations in kinetics could reflect either changes in access or a change in substrate binding in the active site. Deltamethrin binds in a similar mode in both the mutant and wildtype. The increase in lipophilic score of the mutant (Table 6.17) may reflect the substitution of a polar residue for an aliphatic or may reflect increased vdW interactions of a larger side chain (Figure 6.23) and may account for the increase in affinity. However, the effects of the mutant could be due to other structural factors. T240 aligns with the template F241. In the CYP3A4 template, F241 occupies a hydrophobic pocket and π-stacks with a Phe cluster that closes the roof of the active site. This alignment also places T240 within a hydrophobic pocket. While the mutant I240 may be well positioned within this pocket, the more polar T240 may not be well positioned within this pocket and its ability to H-bond may cause an alteration in structure of the GG' loop.

Table 6.17 ChemScores (kJ/mol) of wildtype and mutant CYP6M2 deltamethrin dockings.

	Score	ΔG	S(hbond)	S(metal)	S(lipo)	H(rot)	ΔE(clash)	ΔE(int)
Wild	41.0	-45.5	0.0	0.0	377.0	1.6	2.0	2.5
T240I	42.2	-47.0	0.0	0.0	390.7	1.6	2.6	2.2
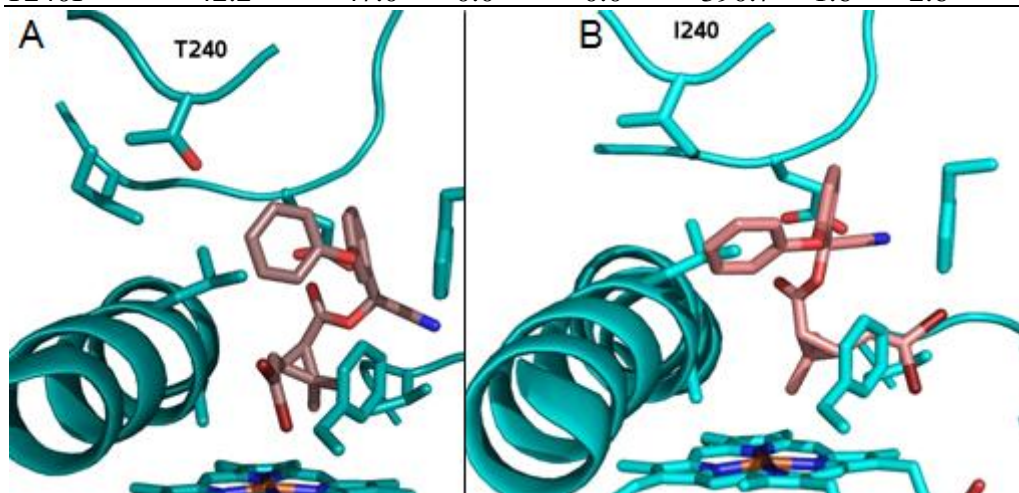
Figure 6.23 Deltamethrin binds in a similar mode in both the wild type (A) and mutant (B).

6.7.2 H382Q and A392S

H382Q and A392S also occur at the mouth of an access channel. This channel is a large channel that is also present in the CYP3A4 structure 1TQN (Yano et al., 2004). It extends along the surface of sheet β-1 and between the F' helix and B-C loop, and corresponds to channel 2b identified by Wade et al. (2005). In 1TQN, R106 extends across this channel to participate in a stabilising H-bonding network. This residue also divides the access channel into two. In CYP6M2, this residue is substituted by a Gly and therefore the entrance to this channel is larger. Also in CYP6M2, many of the residues forming the H-bonding network in CYP3A4 are substituted by aliphatic or aromatic residues (Table 6.18). This may make this access channel a possible entrance point for hydrophobic ligands such as deltamethrin. As with channel 3, this channel opens near to the membrane but does not open directly onto the membrane as channel 3 does (Figure 6.24, A). The positions of A392S and H382Y could also be in positions to affect the access of hydrophobic ligands along the surface of the membrane.

Table 6.18 Residues of the H-bonding network in CYP3A4 and the substitutions in CYP6M2.

P450	Residue					
CYP3A4[1]	R106	E374	D76	R372	D61	Y53
CYP6M2[2]	G106	F374	Q77	V372	K60	I49

[1] residues involved in the H-bonding network in the CYP3A4 structure 1TQN determined by Yano et al. (2004).
[2] residues of CYP6M2 aligning with CYP3A4 determined from a pairwise alignment and an alignment of structures.
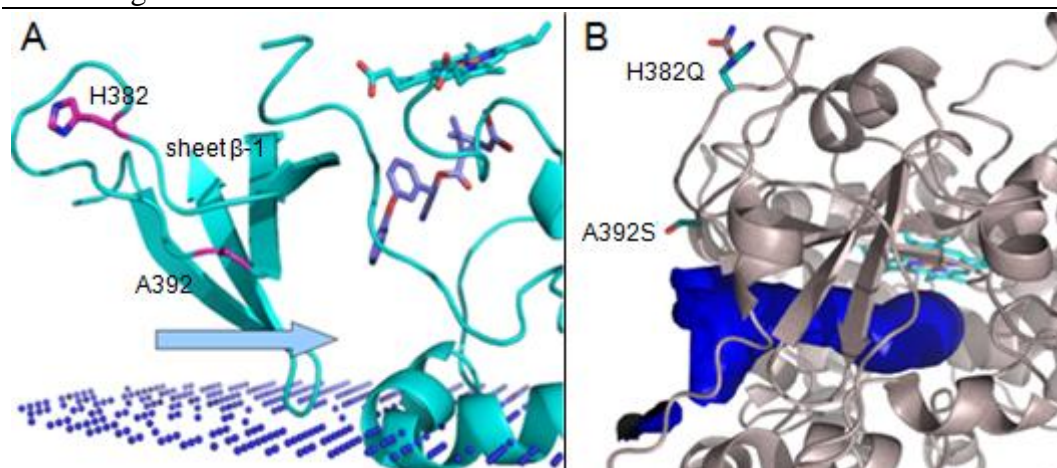
Figure 6.24 (A) CYP6M2 with the position of access channel 2b (arrow) and the positions of SNPs H382Q and A392S (pink), with the position of the membrane predicted by MAPAS (blue dotted plane). (B) Positions of the mutants relative to the access channel 2b identified by MOLE (blue).

The normal mode analysis of CYP6M2 also identifies the β-sheet (sheet β-1) as being mobile (Figure 6.11), particularly in normal mode 7. In this normal mode, this domain moves away and towards the membrane and may be involved in promoting ligand access. The changes in K_m for these SNPs may be due to changes in substrate access, product egress or may form a peripheral binding site. It has been previously shown that surface substitutions increased K_m but not Kd (Pikuleva, 2006). A392S is located at the mouth of access channel 2b and is in a position to affect entry or egress. However, H382Q is distant from the access channel and is unlikely to have a direct influence on this channel (Figure 6.24, B). It is possible that a peripheral intermediate binding site could occur and either of these mutants could be involved. Dockings at the entrance of the access channel place deltamethrin within clefts on the surface. Deltamethrin can dock in a cleft formed by the β1-4 β1-3 loop and A'' helix which is lined by A392 (Figure 6.25). The high scoring modes are distant from this residue but the low scoring modes dock close to this residue which could have an effect on peripheral binding.

A392S produces similar docking modes to the wildtype but a H-bond occurs that is absent in the wildtype.
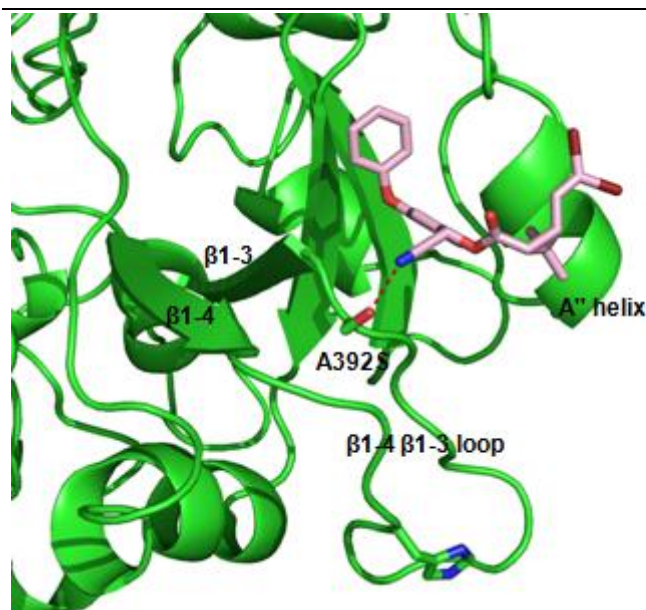
Figure 6.25 A surface docking of deltamethrin to CYP6M2 forming a H-bond with the mutant A392S. The surface docking was constrained within 10 Å of the mutant residue.

Mutations outside the active site can affect metabolism by causing changes in reductase binding, altering substrate channels, affecting conformational changes that occur during catalysis, and producing global changes that affect the active site. This has been demonstrated in CYP1A2 where mutations distant from the site affected metabolism. In this case the mutations were also distant from access channels and the reductase binding site, and were suggested to cause changes in global folding or domain motions (Kim and Guengerich, 2004).

The SNP mutants occur in the β1-4 β1-3 loop. In the template this loop is ordered into a beta sheet, and a beta sheet is also predicted in CYP6M2 by secondary structure predictors, but in CYP6M2 an insert occurs in the loop and the secondary structure is disordered to form a single loop. This loop has been found to vary between crystal structures of the same protein. For CYP2C9 it occurs as a loop in the 1OG2 and 1OG5 structures but occurs as an anti-parallel sheet in 1R90. However, the sheet region in 1R90 was identified as poorly ordered and may not be precisely modelled. The position of the sheet also differs between P450s. In CYP2D6 it closely packs with the underside of sheet 1 due to the presence of small hydrophobic sidechains while in CYP2C9 there is looser packing.

The mutations could be affecting the conformation of the loop and ModLoop was used to remodel the β1-4 β1-3 loop. Both of the H382Q and A392S mutants produced similar loops that differed from the wild type (Figure 6.26). These mutants could have their effects by the conformational changes on this loop. Alternatively, as this loop occurs on the proximal surface it may have an affect on reductase binding. Both H382Q and A392S were shown to increase Km and are close to an access channel. As other studies have demonstrated that the mutations of access channel residues can increase Km, the increase in Km for the SNPs may also indicate a role for access. Overall, the effects of the A392S could be explained by a possible affect on access. The effect of T240I could be explained by its affect on substrate binding but also access. However, the effects of H382Q could not be fully explained but may be due to changes in conformation, reductase binding or the presence of a peripheral binding site.
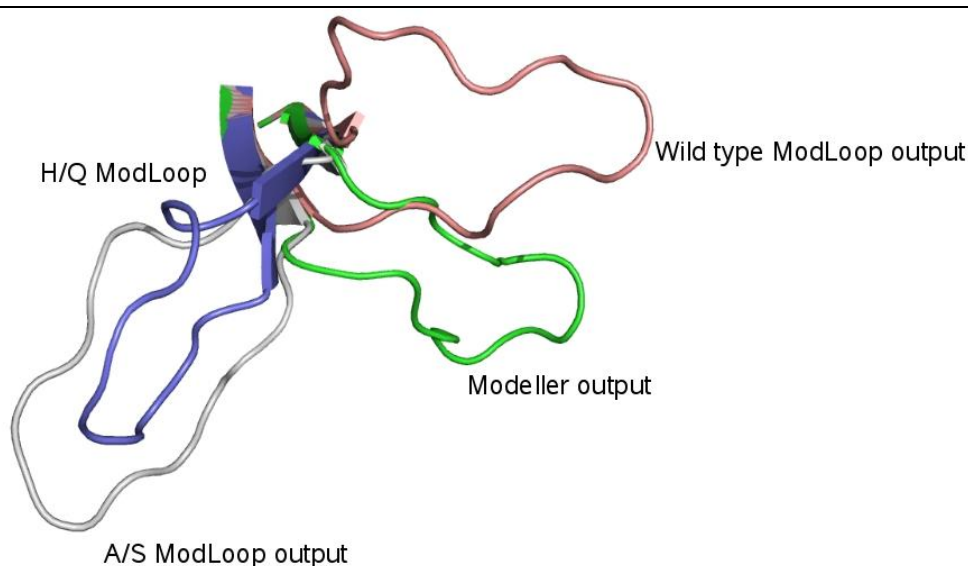
Figure 6.26 ModLoop re-modelling of the β1-4 β1-3 loop. All of the remodelled loops differ from the Modeller output, but both of the mutants produce similar loops that differ from the wild type.

6.8 SNPs within populations

In P450s SNPs tend to occur within hot spots or hyper-variable regions. The SNPs occurring in CYP6M2 were compared to those occurring in other pyrethroid metabolising P450s to determine any commonalities. Six alleles of CYP6D1 have been identified in 6 strains of housefly (Table 6.19) and polymorphism has been identified at some positions. The alleles found in the insecticide resistant strains have some unique substitutions compared to the susceptible strains and may represent adaptations for metabolism. Some of these substitutions occur in the same areas as the substitutions in

CYP6M2. In both CYP6M2 and CYP6D1 substitutions occur in the C terminal loop, A helix and the FG loop. In CYP6D1 the cluster of substitutions in the FG loop also line an access channel (Table 6.20). Sequence variations also occur in CYP6B8 at similar positions to CYP6M2, with SNPs at the β1-4 and β1-3 loop and G' helix (Table 6.21). Although these positions may have some functional significance, these locations may represent hyper-variable regions common to all P450s and variations in these regions may not affect function.

Table 6.19 CYP6D1 alleles identified from 6 strains, adapted from Scott et al. (1999).

Strain	allele	strain characteristic
LPR	v1	pyrethroid resistant strain
aabys	v3	insecticide susceptible
CS	v2	insecticide susceptible
ISK	v4	insecticide susceptible
OCR	v5	pyrethroid susceptible, cyclodiene resistant
Cornell-R	v6	pyrethroid susceptible, organophosphate resistant.

Table 6.20 Residue differences between the 6 alleles of CYP6D1 identified in Swissprot.

Position	Strain LPR	AABYS	CS	ISK	OCR	C-R		Location[1]
2	L	-	L	L	L	L		C terminus
55	M	L	M	M	L	L		A helix
150	D	A	A	A	A	A	*	D helix
153	I	L	L	L	L	L	*	D helix
165	T	S	S	S	S	S	*	D helix
182	I	I	I	I	N	I		E helix
218	E	Q	Q	Q	E	Q	^	FG loop
220	I	F	F	F	F	F	*	FG loop
225	T	N	N	N	T	N	^	FG loop
227	M	I	I	I	M	I	^	FG loop
262	K	K	K	K	T			GH loop
266	R	P	R	R	R			GH loop
447	D	D	D	D	N			K" L loop
469	I	M	M	I	I			L helix

* substitutions unique to LPR; ^ substitutions in LPR and OCR
[1] position in the CYP6D1 model.

Table 6.21 Locations of SNPs in variants of CYP6B8.

Position	Variant				Location
	6B8v1	6B8v4	6B8v3	6B8v2	
226	L	L	L	S	G' helix
283	D	D	G	D	HI loop
322	S	S	S	T	I helix C terminus
367	M	M	V	M	K helix C terminus
375	Q	Q	Q	R	β1-4
377	K	K	K	R	β1-4
379	T	S	T	T	β1-4
388	D	D	G	D	β1-4 and β1-3 loop

Of the SNPs that occur in CYP6B8, only Q375R is in a position to directly affect the binding of a substrate in the active site. The substitutions on beta strand 1-4 occur on the side of the strand that lines access channel 2b, L226S occurs on access channel 3. Overall, in the pyrethroid metabolising CYPs 6M2, 6D1 and 6B8, SNPs tend to occur around channels 3 and 2b.

6.9 Conclusion

As the structure of the active site and conservations of active site residues can not alone explain preferences towards substrates, factors external to the active site were explored and a number of factors were found to have an effect on substrate preference, metabolites produced and activity.

The presence of b5 has been found by our collaborators to enhance the activity of CYP6M2. In CYP6M2 the site of b5 binding was predicted to be on the proximal surface near the BC loop. NMA identifies the BC loop as a hinge region between structural domains and the binding of b5 to this region may stabilise the conformation of the protein. It may also stabilise the interaction with the reductase. b5 also has a role as a second electron donor and its presence can shift metabolite profiles. In the absence of b5, DDT is reductively dehalogenated involving a single electron transfer. In the presence of b5, a second election enables DDT to be hydroxylated.

There may be a number of factors other than binding within the active site that affect ligand binding. P450s may have adaptations that promote the binding of particular ligands. In this study, it was shown that P450s that metabolise hydrophobic ligands tend to have a larger area of contact with the membrane and have structural adaptations for the binding of hydrophobic ligands, such as placing the openings of access channels

into the membrane. As pyrethroids are hydrophobic, metabolisers may also be expected to have sequence or structural features that facilitate access of hydrophobic ligands. The pyrethroid metabolisers were shown to have larger surface areas in contact with the membrane and the openings of access channels oriented towards the membrane. In addition, the hydrophobic patches in contact with the membrane extended through the access channel into the active sites, this may represent adaptations for the delivery of hydrophobic ligands. Non metabolisers of pyrethroids tend to prefer hydrophilic ligands and tend to lack these features, with access channels oriented towards water and poorer contacts with the membrane. Some non metabolisers of pyrethroids tend to have similar active site characteristics to the metabolisers and activity cannot be explained by the active site alone. Features required for the binding of hydrophobic ligands may also determine activity towards hydrophobic pyrethroids, and active site characteristics alone may not be the determinant of activity.

CYP6M2 SNPs occur in wild populations and were shown to affect the kinetics of pyrethroid binding. Some of these occur around access channels and their effects on kinetics could reflect their effects on ligand access. One SNP, T240I, increases affinity for pyrethroids, it occurs on an access channel and projects into the site to contact the bound ligand, and may be an adaptation for pyrethroid binding. The locations of these SNPs correspond to locations of SNPs also identified in other pyrethroid metabolising P450s, and may be general adaptations for pyrethroid metabolism or common hyper-variable regions.

7. Design of novel pyrethroids

7.0 Preface

Pyrethroids are the only class of insecticide used on bed nets to prevent malaria, but metabolism by insect P450s can lead to a failure of control. This chapter explores the requirements of pyrethroids for toxicity and how changes in structure can affect both toxicity and metabolism. The understanding of how pyrethroids bind to metabolising P450s can be used to design novel pyrethroids that are capable of binding to the sodium channel target but avoid metabolism by P450s.

7.1 Designing pyrethroids

A number of methods have been used to design pyrethroids that have increased toxicity or decreased metabolism. The masking of sites of metabolism has been shown to shift metabolism to other parts of the molecule. The addition of groups that act as mechanism inhibitors can reduce resistance. Altering the structure of either the acid or alcohol can affect metabolism. As CYP6M2 binds deltamethrin for metabolism on the dimethyl group and 4', understanding the mode of binding can be used to design novel structures that bind poorly by steric clashes or by poor interactions, or fail to bind for metabolism.
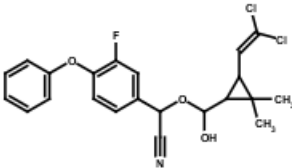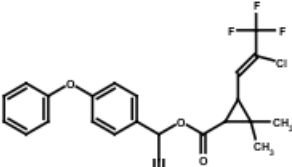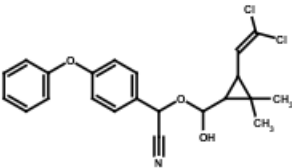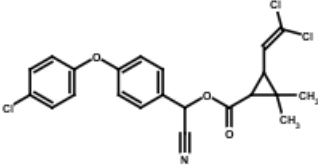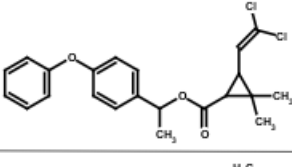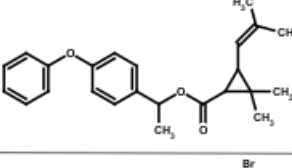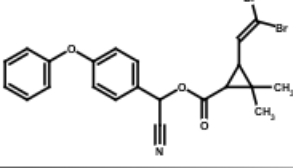
7.1.1 The effects of pyrethroid structure on metabolism

As metabolism by P450s has been shown to be the primary mechanism of resistance in a number of insects. Altering the structure of pyrethroids may affect binding and metabolism and therefore resistance, and the structure of the alcohol moiety in particular has been shown to greatly affect pyrethroid toxicity. Scott and Georghiou (1986) found that pyrethroid metabolism in CYP6D1 may be restricted to specific sites on the phenoxybenzyl group particularly the 4' site. Similarly CYP6M2 metabolism also primarily occurs at the 4' site. This restriction suggests that P450 mediated resistance may not confer resistance to a variety of insecticides and that structural modifications can reduce or eliminate P450 mediated resistance.

Changes to the pyrethroid structure have been shown to affect toxicity. To determine the relationship between pyrethroid structure and resistance, Scott et al. (1986) tested pyrethroids with diverse structures. These were tested on a Learn-PyR strain that was selected with permethrin and had a 6073 fold resistance to permethrin compared to the susceptible NAIDM strain. Scot et al. (1986) identified a number of factors concerning resistance: the presence of an α-CN group did not affect resistance, and had no clear affect on toxicity. The presence of an unsubstituted phenoxybenzyl alcohol moiety was

always associated with high levels of resistance while modifications of the acid affected toxicity but not resistance ratio (Table 7.1). This pattern was also seen in bollworms as similar changes to the structure of the alcohol group also affects toxicity (Yang et al., 2005). The oxygen of the phenoxybenzyl group may not be important for high resistance as some pyrethroids lacking this group have high RRs, while others have low resistance ratios.

Table 7.1. Taken from Scot et al. (1986). Toxicity of pyrethroids in housefly[a].

Pyrethroid	Structure	Fly LD50	(RR)
Cyfluthrin		9.9	240
Cyhalothrin		33	4700
Cypermethrin		36	5100
CGA		52	400
Permethrin		150	6200
Phenothrin		>200	3400
Deltamethrin		>200	(10^5)

[a](RR) Resistance ratio: LD50 of LPR strain /LD50 of NAIDM strain. LD50: µg/fly.

Although different pyrethroid structures have been shown to affect LD50 and resistance ratio, it is difficult to determine whether this is due to metabolism or due to other factors
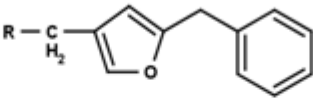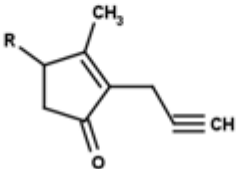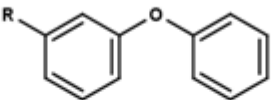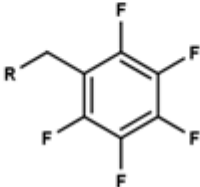
such as kinetics and target site interactions. To determine the role of P450s, Scott et al. (1986) tested some pyrethroids with and without piperonyl butoxide (PBO). They found that PBO reduced the LD50s (µg/fly) of cyhalothrin from 33 to 0.11, of fenfluthrin from 2 to 0.38 and of RU 38702 from 120 to 0.13. They indicated that P450 metabolism is the primary mechanism of resistance in housefly, and this mechanism could be overcome by changes to the alcohol moiety, but not the only mechanism and metabolism alone may not explain differences in toxicity.

7.1.2 Alcohol substitution

Scot et al. (1986) found that the presence of an unsubstituted phenoxybenzyl alcohol moiety was always associated with high levels of resistance. Bioallethrin, bioresmethrin and phenothrin have identical acid groups but different alcohol groups and differ in their toxicity to houseflies (Table 7.2). The difference in toxicity may be due to the availability of sites of metabolism. Phenothrin was shown to have the lowest toxicity. This may be due to sites of metabolism on both the unsubstituted isobutenyl methyls and the phenoxybenzyl group. Bioallethrin had a higher toxicity than phenothrin that may be due to fewer sites available for metabolism, as bioallethrin metabolism was shown to be exclusively on the acid group and does not occur on the alcohol group. The only sites of metabolism were on the isobutenyl group. Bioallethrin, that has neither a phenyl nor a benzyl group also had a low resistance ratio. This indicates that there may be preferential attack on benzyl, phenyl or similar groups. The alcohol group of bioresmethrin is similar to bioallethrin in that it contains a hydrophobic ring, and similar to phenothrin in that it contains an aromatic ring. Although it has additional sites for hydroxylation, it has a higher toxicity than either phenothrin or bioallethrin indicating that the availability of sites of metabolism may not be a primary determinant of resistance.

The difference in toxicity between pyrethroids may be due to the alcohol group structure. Changes to the alcohol structure have been shown to have the greatest affect on LD50 compared to changes to the acid group. Scott et al. (1986) found that pyrethroids with un-substituted phenoxybenzyl groups were always associated with low levels of toxicity while substitutions of the aromatic rings increased toxicity, and this has been found to be due primarily to P450 detoxification.

Table 7.2 A comparison of the structure of the alcohol group and toxicity (adapted from Scott et al, 1986).

Compound	LD50(µg/fly)	Alcohol structure
Bioresmethrin	45	
Bioallethrin	110	
Phenothrin	>200	
Fenfluthin	2.0	

Altering the orientation of the phenoxybenzyl rings also affects toxicity. The fluorine added to the benzyl ring of cyfluthrin causes the 3-phenoxy ring to become twisted relative to the benzyl ring, due to the electrostatic interaction between the fluorine and the 3-phenoxy ring. There is also a relationship between insecticidal activity and this twisted structure (Jeschke, 2004). The increase in toxicity of cyfluthrin may be due to this twist producing a less preferred binding position with CYP6D1. In the CYP6D1 model cyfluthin binds for metabolism at the 4' position and the twisted orientation of the alcohol group may produce less preferred π interactions with the BC loop aromatic residues (F115, Y102) or poorer contacts (Figure 7.1) as cyfluthrin has a poorer score compared to permethrin (Table 7.3). However, pyrethroids with non-coplanar aromatic rings may have an improved interaction with the sodium channel, and this increase in toxicity may be due to target site interactions. Nakagawa et al. (1982) found that the position of the phenoxy relative to the benzyl ring could have an affect on neurotoxicity. There was the highest affect when a benzyl group was attached to the benzyl ring in the meta position, and lowest in the ortho position.

Table 7.3 The ChemScores (kJ/mol) of cyfluthrin compared to permethrin.

	Score	ΔG	S(hbond)	S(metal)	S(lipo)	H(rot)	ΔE(clash)	ΔE(int)
cyfluthrin	36.1	-37.0	0.8	0.0	286.0	1.7	0.2	0.8
Rtrans	39.8	-45.1	1.0	0.0	342.5	1.5	3.6	1.7
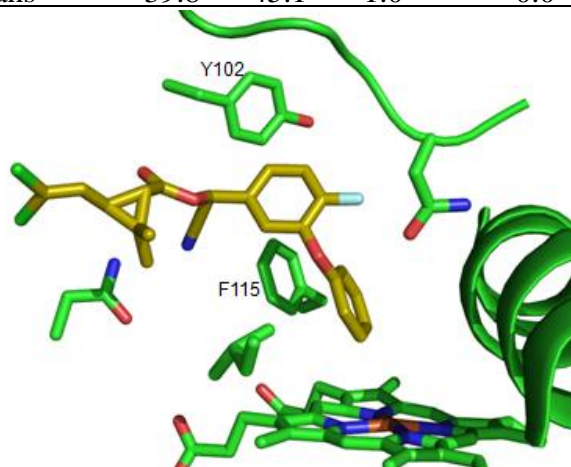
Figure 7.1 Cyfluthrin binding in CYP6D1.

7.1.3 Masking

Scott et al. (1986) found that the substitution of the permethrin phenoxybenzyl alcohol with a pentafluorophenyl as in fenfluthrin reduced resistance from 6073 to 41 fold. This resistance could not be further lowered by PBO indicating that the remaining resistance could be due to target site insensitivity or decreased penetration. The high oxidative activity of the Lean-PyR strain did not confer resistance to fenfluthrin. The masking of all sites of hydroxylation on the alcohol group almost entirely eliminates P450 metabolism. A number of factors could affect the metabolism of fenfluthrin. The 3-phenoxy ring is absent and the benzyl ring is replaced by a pentofluorophenyl ring. This ring may prefer to bind distant from the heme with the inert chorines of the acid towards the heme in an unproductive binding mode (Figure 7.2, B). Where the ring is close to the heme metabolism may also be prevented by the absence of a phenoxy ring and the masking of any sites of metabolism on the benzyl ring by fluorines.

The addition of a chlorine to the 4' position of permethrin to produce CGA decreased resistance from 6073 to 400 fold. CGA has a similar mode of binding to permethrin. The presence of the fluorine in the 4' position may prevent metabolism at this position, leading to the lower resistance shown. The additional bulk of the fluorine may also cause the pyrethroid to dock further from the heme than those with a hydrogen in this position (Figure 7.2, A). This may also prevent metabolism at the 2' or 6 position, further contributing to a decreased resistance.
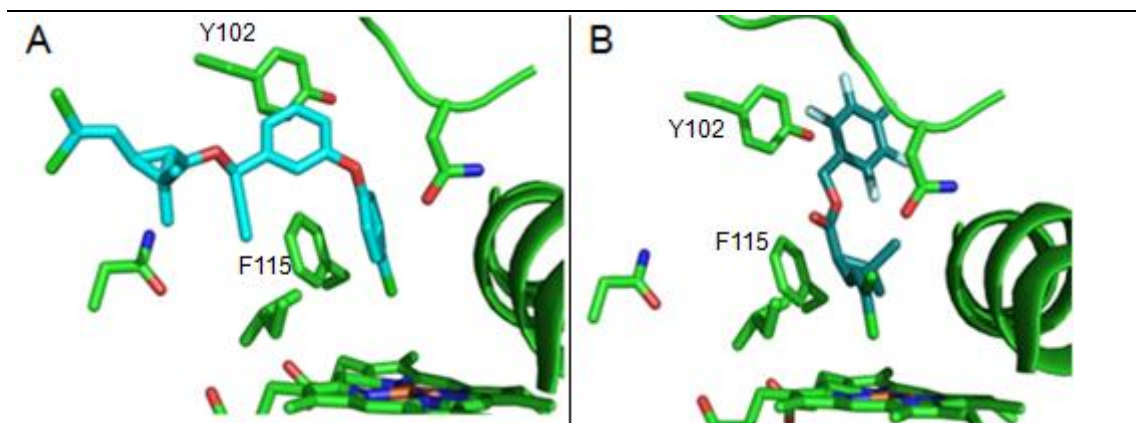
Figure 7.2 CYP6D1 binding CGA (A), and fenfluthin (B).

Interactions with P450s is only one factor affecting toxicity, pyrethroids primarily have their toxic effects by interactions with the sodium channel target site. Bioresmethrin has a 5-benzyl-3-furylmethyl alcohol group. Pyrethroids with this group have been found to have high degrees of toxicity despite a high degree of P450 metabolism (Berteau and Casida, 1969). Bioresmethrin may have a high toxicity due to target site interactions. Sheppard and Norton (1980) identified a trend in synthetic pyrethroids where pyrethroids containing the 5-benzyl-3-furylmethyl alcohol groups were 2-3 fold more potent than those with the phenoxybenzyl alcohol. However, Berteau and Casida (1969) found that in housefly treatment with PBO could significantly synergise the activity of 5-benzyl-3-furylmethyl pyrethroids, indicating that despite increase toxicity there is also a high degree of metabolism limiting its activity.

Toxicity effects could be due to interactions not only at the sodium channel but also at alternative targets. Burr et al. (2004) have found that while all pyrethroids have a similar mode of action on the sodium channel, some pyrethroids can also have an additional target, the chloride ion channel, where they reduce the chance of the channel opening. Bioallethrin and cyfluthrin, along with permethrin, cypermethrin and deltamethrin, were found to interact with chloride channels to reduce open channel probability, while cyhalothrin, bioresmethrin and non ester pyrethroids were found not to decrease the probability.

7.1.4 Toxicity to *Anopheles*

Hougard et al. (2003) compared a susceptible *Anopheles gambiae* strain and a strain homozygous for the kdr gene for susceptibilities to a range of pyrethroids. Differences in pyrethroid structure affect toxicity, but as the resistant strain had a mutation in the

sodium channel it is unclear what is producing the resistance. Cypermethrin and deltamethrin were identified as the most effective pyrethroids (Table 7.4).

Table 7.4 Mortality rates of *A. gambiae* susceptible and resistant strains taken from Hougard et al. (2003).

	(mg/m^2)	WHO recommended concentration Susceptible (%)	Resistant (%)	25% recommended concentration Susceptible (%)	Resistant (%)
Alpha-cypermethrin	40	100	94	100	70
Bifenthrin	25	61	8	24	7
Cyfluthrin	50	74	0	47	8
Deltamethrin	25	100	14	92	4
Etofenprox	200	94	6	56	4
Lambdacyhalothrin	20	43	0	43	2
Permethrin	500	60	2	44	2

7.1.5 CYP6M2 dockings

The pyrethroids with toxicities determined by Hougard et al. (2003) were docked and the dockings show some correlations with the known toxicity of the pyrethroids (Table 7.5). Lambdacyhalothrin (RcisS and ScisR) has low toxicity in *A. gambiae*. It binds with a low score but it is able to bind in a position for metabolism at the 4' and the extension of the fluoromethyl to the halovynal does not prevent the binding for metabolism of the trans methyl (Figure 7.3, A and B).

Table 7.5 ChemScores (kJ/mol) of pyrethroids in the CYP6M2 model.

		Score	ΔG	S(hbond)	S(metal)	S(lipo)	H(rot)	ΔE(clash)	ΔE(int)
Cypermethrin	RcisS	39.7	-45.2	0.0	0.0	375.2	1.6	1.9	3.5
Cypermethrin	ScisR	41.4	-44.3	0.0	0.0	367.4	1.6	0.3	2.7
Bifenthrin		41.9	-47.9	0.0	0.0	396.7	1.6	3.1	2.8
Cyfluthrin		41.7	-47.6	0.0	0.0	397.4	1.7	4.6	1.2
Deltamethrin		41.5	-46.6	0.0	0.0	386.9	1.6	2.7	2.4
Etofenprox		50.1	-52.6	0.8	0.0	409.8	1.4	0.2	2.2
Cyhalothrin	RcisS	39.2	-43.4	0.0	0.0	368.6	2.1	2.2	2.0
Cyhalothrin	ScisR	38.4	-40.6	0.0	0.0	344.9	2.1	0.1	2.1
Permethrin	Rcis	42.4	-45.8	0.0	0.0	375.9	1.5	2.3	1.0
Permethrin	Rtrans	45.7	-48.5	0.0	0.0	399.7	1.5	0.8	2.1
Permethrin	Scis	42.1	-47.3	0.0	0.0	389.2	1.5	2.6	2.6
Permethrin	Strans	44.7	-46.6	0.0	0.0	383.6	1.5	0.9	1.1
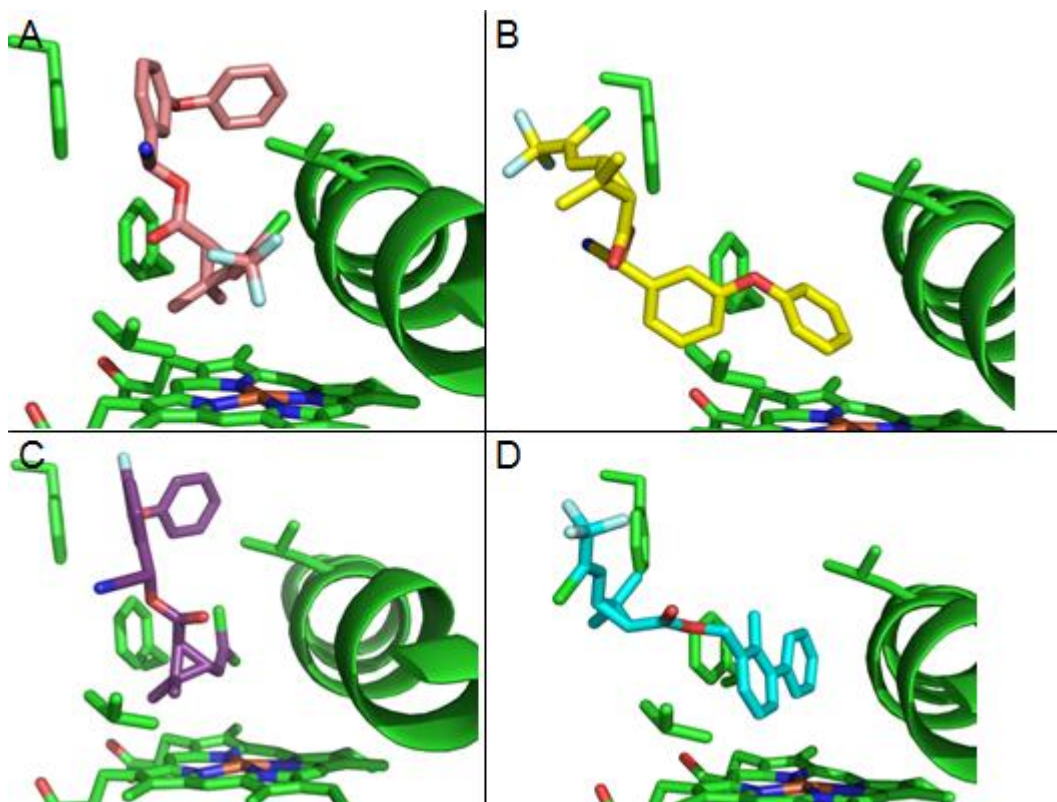
Figure 7.3 CYP6M2 binding ScisR cyhalothrin, the first ranked mode(A) and the 3rd ranked mode(B). The first ranked modes of cyfluthrin(C) and bifenthrin(D).

Bifenthrin and cyfluthin also have low toxicity in *A. gambiae*. Both of these bind with high scores and in modes that allow metabolism indicating that they may be substrates for CYP6M2 (Figure 7.3, C and D). However, some pyrethroids with a high toxicity also have higher scores, deltamethrin, permethrin and cypermethrin bind with high scores with etofenprox giving the highest score. Although the scores of permethrin and deltamethrin are consistent with them being substrates with a high turnover in CYP6M2, the binding scores of pyrethroids to CYP6M2 do not entirely correlate with their recorded toxicity. As toxicity is affected by a large range of factors, the ability of pyrethroids to bind metabolising enzyme may not be used to predict toxicity.

In addition, a structurally diverse range of other untested pyrethroids were docked to determine the requirements for binding. Modifications to the acid group had effects on the binding score. Large aliphatic extensions tended to reduce the score while the addition of aromatic rings tended to increase the score (Table 7.6).

Table 7.6 The ChemScores (kJ/mol) of pyrethroids docked in CYP6M2.

	Score	ΔG	S(hbond)	S(metal)	S(lipo)	H(rot)	ΔE(clash)	ΔE(int)
Acaranthrin	32.2	-38.8	0.0	0.0	354.1	3.2	1.1	5.5
Flucythrinate	43.6	-48.1	0.0	0.0	414.7	2.3	0.4	4.1
Fluvalinate	43.2	-45.6	0.0	0.0	385.6	2.0	0.4	2.0
Fenvalerate	47.9	-53.6	0.0	0.0	447.8	1.7	1.5	4.3
Allethrin	35.9	-36.8	0.0	0.0	298.8	1.4	0.1	0.8
Phalthrin	33.6	-34.5	0.0	0.0	288.6	1.9	0.2	0.6
Dimethrin	40.0	-41.6	0.0	0.0	335.2	1.2	0.8	0.8
Fenfluthrin	29.2	-31.1	0.0	0.0	261.0	1.9	1.5	0.4

Acrianthrin has a low score due to a large clash score with the extended acid group. However, the best scoring mode places it for metabolism of the 4' position. High scoring pyrethroids flucythrinate, fluvalinate and fenvalerate had additional aromatic rings in the acid group and also productive modes (Figure 7.4). The higher score may be due to additional interactions with the BC loop aromatic residues (F110, F123). Low scoring pyrethroids tend to be the natural pyrethroids that lack aromatic groups such as allethrin and phalthrin, or small pyrethroids with only a single aromatic group such as dimethrin or fenfluthrin. This indicates that binding is affected by the presence of aromatic groups in the ligand.
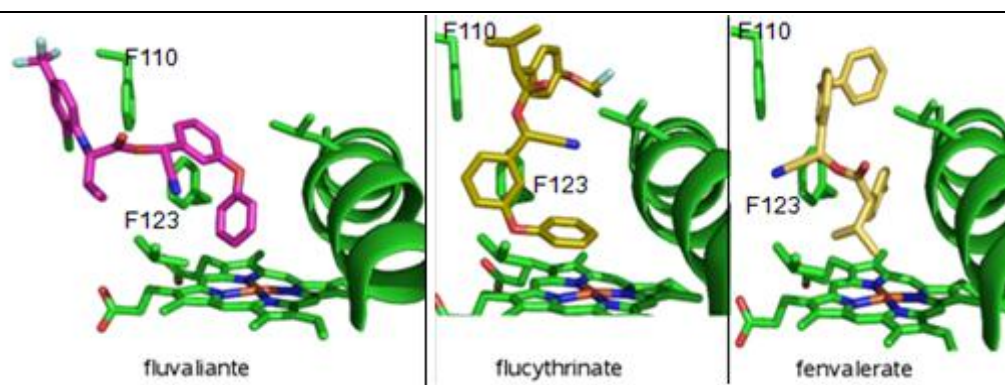
Figure 7.4 CYP6M2 binding fluvalinate, flucythrinate and fenvalerate.

7.2 Pyrethroid requirement for activity at the sodium channel

The sodium channel is the target site for pyrethroids. Pyrethroids bind to site 7 on the sodium channel to stabilise the open state and prevent its closing. The movement of sodium in to the neuron produces persistent depolarisation and repetitive nerve firing that causes paralysis and death of the insect (O'Reilly et al., 2006).

7.2.1 Alcohol group

O'Reilly et al. (2006) produced homology models of the insect sodium channel. They

determined that some of the pyrethroid isomers were inactive due to an inability to bind to the sodium channel. The inactive isomer of fenvalerate 1R could not bind to the channel in the same way as the active 1S isomer due to steric hindrance. O'Reilly et al. (2006) also determined some features required by pyrethroids for binding to the channel. The key requirement was for the alcohol group of the pyrethroid to match in volume the hydrophobic cavity of the binding site. They suggested that the maximum dimensions of the alcohol were represented by two benzene rings with a bridging atom, but the nature and dimensions of the rings are important for activity. A cyclic group has a higher activity when attached at the meta rather than the para position. The rings should also be non-coplanar due to steric restrictions.

Naumann et al. (1998) attempted to map the pyrethroid binding site by using analogues to determine the structure activity space and to assess the maximum volume of the acid and alcohol components before activity is lost. For this, a range of pyrethroid structures were studied for their toxicity in insects and a number of structural constraints have been suggested for pyrethroid design. From this study it was found that a positive charge at the 4' position abolishes activity (Figure 7.5, A compound i) but replacement of the CN in cypermethrin for an N-(n'-methyl)piperazinylmethyl retained activity (Figure 7.5, A compound ii). Unlike O'Reilly et al. (2006), they found that large extensions of the alcohol group could be tolerated but there were spatial restraints with a limited space available for extension of the para position, as large extensions in this direction cause a loss in activity, but other extensions are tolerated (Figure 7.5 B and C).

7.2.2 Acid group

Ford et al. (1989) identified a number of factors required for toxicity including a gemdimethyl group on the cyclopropane ring of the acid group. Byberg et al. (1987) further developed a pharmacophore for pyrethroid acid and alcohol groups by mapping the ether oxygen, methyl groups and aromatic ring. O'Reilly et al. (2006) suggested that the acid group is tolerant to a large degree of steric and electronic substitution before activity is lost and that these acid substitutions could be accommodated in the binding pocket. The extension of the acid group as occurs in acrianthrin was suggested to account for its activity.

7.2.3 Ester bond

The ester bond has been found to be a requirement for activity as replacement with

other groups always leads to a loss of activity. O'Reilly et al. (2006) suggested a role of the ester for binding to the sodium channel as the sodium channel T929 could form H-bonds with the carbonyl oxygen of the ester group.
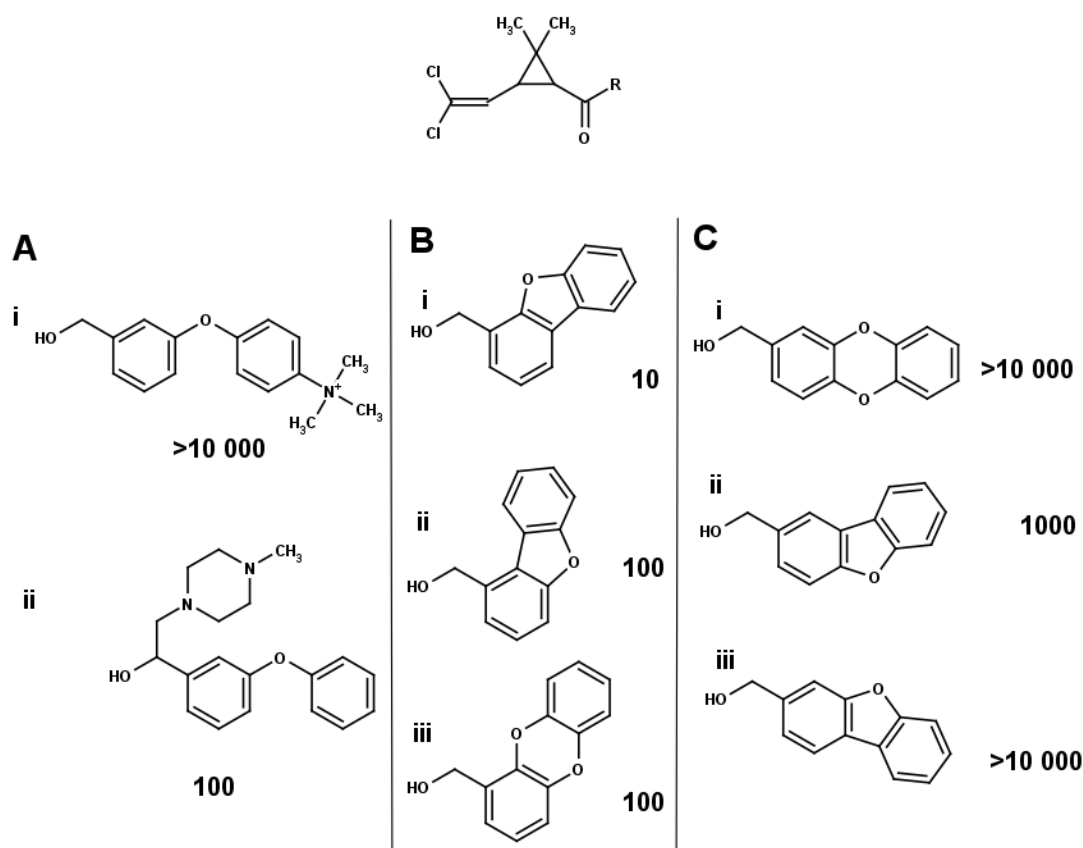
Figure 7.5 (A) A range of alcohol variants tested in *Plutella maculipennis* larvae (values LC95 mg/litre). Activities of flattened permethrin with extensions in the meta (B) and para (C) positions, against *Plutella maculipennis* (values LC95 mg/litre) taken from Naumann et al. (1998).

7.3 Designing pyrethroids to overcome resistance

The CYP6M2 model was used to design a number of pyrethroids that may overcome metabolic resistance.

7.3.1 Good binding P450 inhibitors

Conventional drug design approaches have been used to develop P450 inhibitors. Inhibitor design has usually been focused around aromatic nitrogen heterocycles such as imidazole as they can coordinate with the iron to give a type II inhibition. While there have been reports of imidazole fungicides acting as pyrethroid synergists, there have been no commercial pyrethroids containing imidazole groups.
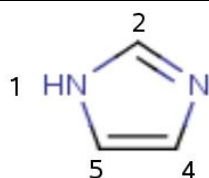
Figure 7.6 Imidazole nomenclature.

Imidazole inhibitors have certain requirements to produce inhibition. Their inhibition is correlated to their lipophilicity and steric factors. Imidazoles with substituents at the 2 or 4 positions (Figure 7.6) show little inhibitory activity. As inhibition is derived from the direct interaction between the non-bonded electrons of the 3-nitrogen and the sixth ligand of the heme, the 3-nitrogen must be able to freely approach the heme. Substitutions of the 2 or 4 positions may introduce steric clashes with the heme that prevent coordination, but substitutions are allowed at positions 1 and 5 and allow for the addition of active site contacts (Rogerson et al., 1977; Verras et al., 2004).

In this study, pyrethroids with imidazole substituents were designed to determine if a pyrethroid P450 inhibitor could be effective. The phenyl ring was replaced with an imidazole ring with the coordinating 3-nitrogen at a similar position as the 4' of the phenyl ring. As CYP6M2 prefers to metabolise deltamethrin at the 4' position, the nitrogen at this position may coordinate to prevent metabolism. To prevent steric hindrance to type II coordination, the imidazole was added at the 5 position. A number of pyrethroid ligands were designed and produced using MOE and docked into the CYP6M2 model (Figure 7.7). The direct replacement of the phenyl ring by an imidazole ring (imi8) produces modes that allow metabolism of the acid group, with lower scoring modes allowing metabolism on the imidazole ring. One of these variants (imi10) could bind with the 3 nitrogen above the heme in a high scoring mode. This binding mode may not allow coordination, but metabolism at the 4' may be prevented (Figure 7.8).

7.3.2 Good binding non metabolised pyrethroids

An attempt was made to design pyrethroids that bind in non productive modes as occurs in some non metabolisers such as CYP6Z2. The molecule was designed to have a high affinity for parts of the active site and access channel that were remote from the heme, to give high scoring non productive modes. However, the access channels of metabolisers are more variable than the active sites so ligands that are able to bind non productively within the active site may be more suitable as candidates than those that bind within access channels to avoid selectivity for a particular P450.
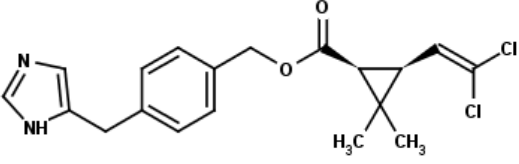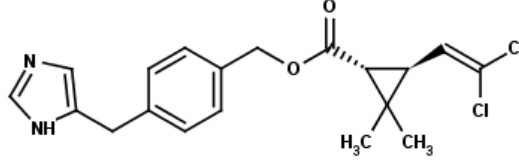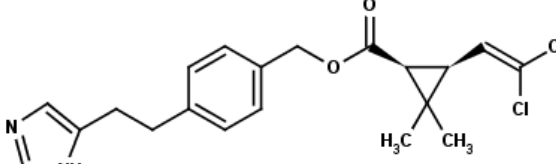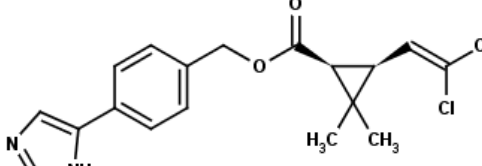
Name	Structure	ChemScore (kj/mol)
Imi8 (Rcis)		37.3
Imi10 (Strans)		40.4
Imi9		38.7
Imi11		none

Figure 7.7 The ChemScores (kJ/mol) of pyrethroid variants with imidazole substitutions in modes that place the imidazole above the heme.
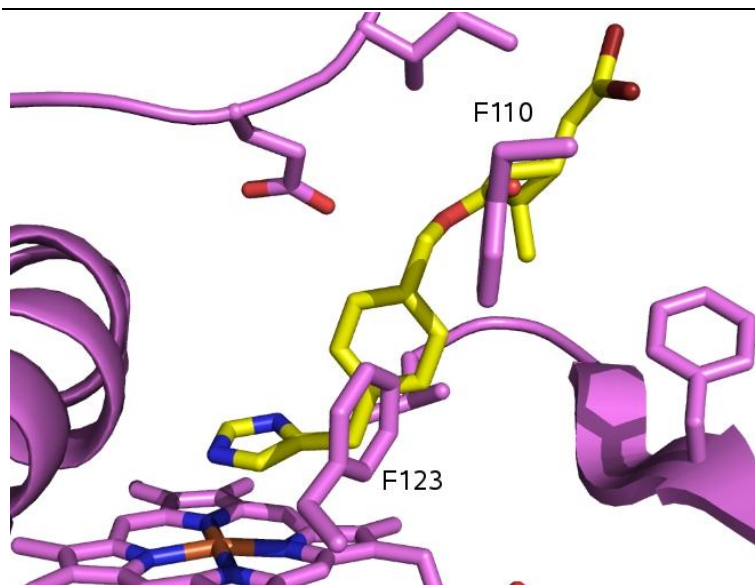
Figure 7.8 Imi10 binding with the imidazole ring above the heme.

Table 7.7 ChemScores (kJ/mol) for the productive mode (trans methyl) and non productive mode (non) of deltamethrin.

	Score	ΔG	S(hbond)	S(metal)	S(lipo)	H(rot)	ΔE(clash)	ΔE(int)
trans methyl	40.7	-43.2	0.0	0.0	357.9	1.6	0.9	1.6
non	35.3	-39.0	0.0	0.0	322.6	1.6	1.0	2.7

Non productive modes occur for deltamethrin but are low scoring. The deltamethrin productive modes have scores of around 40.6 kJ/mol, while the non productive modes have low scores below 36 kJ/mol (Table 7.7). A variety of deltamethrin variants were designed and screened with 10 dockings to identify the occurrence of non productive modes. A primary amine rather than the tertiary cyano group was used because as it was able to make H-bonds with the protein in a larger variety of poses. The majority of deltamethrin variants only produced productive modes, but a small number gave high scoring non-productive modes. It was found that the addition of an amine onto the benzene ring gave two high scoring non-productive modes. The first of these places the variant in a non productive mode as occurs in CYP6Z2. The amine on the benzene ring is in a position to H-bond with the semi conserved acidic residue in the BC loop (E216) to form a H-bond that positions the ligand in a non-productive mode (Figure 7.9). This mode has a high score of 41.2 kJ/mol. The second non productive mode places the variant within the access channel with the alpha carbon amine H-bonding the FG acidic residue and the benzene ring amine bonding with an acidic residue on the A'A loop. A residue in this position was suggested to have a role in stabilising the non productive binding mode in CYP6Z2.
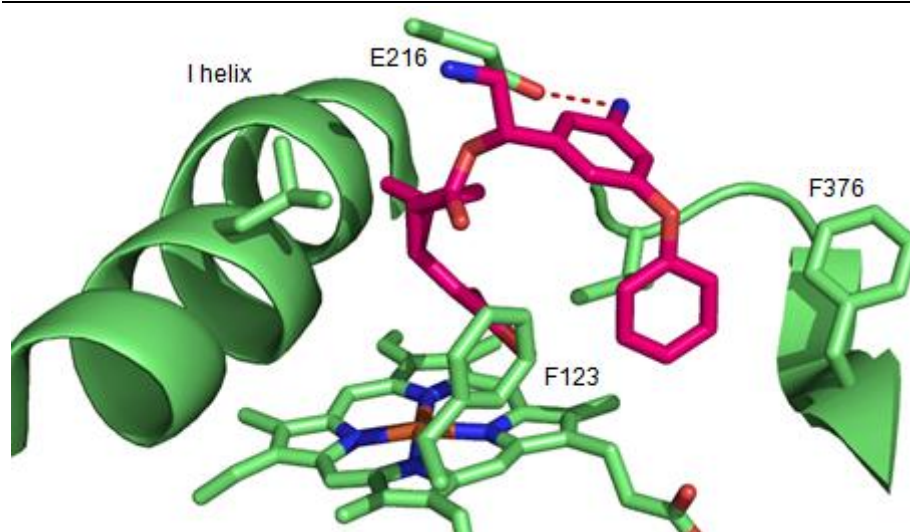
Figure 7.9 The addition of an amine on the benzene ring provides a H-bond (dotted line) to position the ligand in a non-productive mode.

The replacement of the phenyl ring with a cyclohexane ring also produced non productive modes. These molecules bind non productively as the cyclohexane binds within a hydrophobic pocket between the BC loop and the I helix while the benzyl ring stacks with the conserved Phe (F110) in the BC loop. This hydrophobic pocket is formed by F302, V302, L305 of the I helix and L118 and H121 of the BC loop, the binding of the cyclohexane within this pocket places the gem dimethyls distant and the bromines of the halovynal group above the heme (Figure 7.10). However, productive modes also occur that place the cyclohexane for metabolism at the 4' position. This unproductive mode is not found with the native deltamethrin and may occur due to improved hydrophobic contacts between the cyclohexane and the hydrophobic pocket. The best ranked modes of the cyclohexanes bind non productively but also with high scores similar to or higher than the best scoring mode of native deltamethrin (Table 7.8).
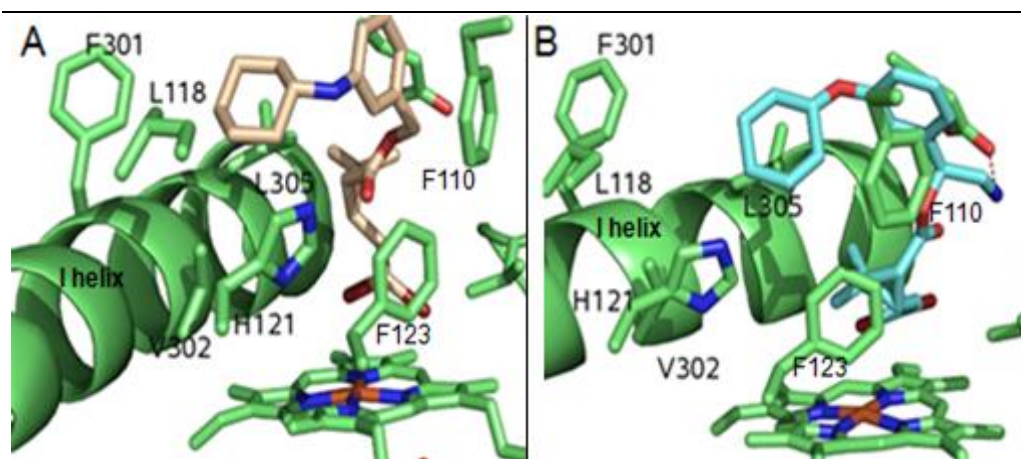
Figure 7.10 The best ranked modes were unproductive for 2 cyclohexane deltamethrin variants Hex1 (A) and Hex2 (B).

Table 7.8 ChemScores (kJ/mol) of pyrethroids with cyclohexane substitutions.

	Score	ΔG	S(hbond)	S(metal)	S(lipo)	H(rot)	ΔE(clash)	ΔE(int)
Hex1	43.4	-49.4	0.9	0.0	376.5	1.3	2.5	3.6
Hex2	40.3	-47.1	0.9	0.0	370.7	1.9	4.3	2.6

7.3.3 Poor or transient binders

Drug discovery usually involves finding the best possible binder by either screening molecular libraries or by rational drug design based on knowledge of the target. Conventionally the tightest binding drugs have been assumed to have the highest potency and selectivity and weak binding drugs have been overlooked as they are assumed to have low specificity or be unable to produce a response. However, there has been some interest in designing weak binders as it has been shown that transient binders

can show a high specificity, specificity is a ratio of affinities between desired and non-desired interactions which can be higher for a weak binder depending on the affinity of cross-binding. Weak binders can also have as high an efficacy as strong binders. While the binding is weak, if local concentrations of a weak binder are high, it can drive equilibrium and fill receptor sites. A number of drugs such as alcohol, aspirin and ibuprofen are transient binding drugs. Fragment based drug design can be used for the design of transient binders. While conventional design involves finding good binding fragments to grow them into a high affinity compound, the design of transient binders involves finding fragments with very weak affinities (Ohlson, 2008). This method was used to design pyrethroids that weakly bind CYP6M2 that may show limited metabolism.

The database screen of CYP6M2 was used to select weak binding compounds. SARvision was used to identify scaffolds in these poor scoring ligands, and the scaffolds were then used for constructing deltamethrin variants. As a large number of fragments were identified in the weak binders, fragments were selected to retain the characteristics of pyrethroids and were used to substitute either the alcohol or acid moiety. They were selected for use based on the criteria for pyrethroid activity, due to the spacial restraints they were restricted in size in the para position and a positive charge at the 4' was avoided but allowed elsewhere. In addition, as O'Reilly et al. (2006) found a π-stacking interaction between the benzene ring and the sodium channel. Fragments were selected that retained a benzene ring. The fragments were also similar in dimensions as two benzene rings connected by a bridging atom, to complement the binding site on the sodium channel. To identify poor binders an arbitrary cut off of 30 kJ/mol was used as described by Kemp et al. (2004). In general, the substitution of the acid scored well with some ligands having higher scores than the native deltamethrin, while substitution of the alcohol scored poorly. It may be that the phenoxybenzyl group a major determinant of pyrethroid binding in CYP6M2 and its substitution affects binding.

7.3.3.1 Deltamethrin variants

The phenoxybenzyl group was substituted for the low scoring fragments that maintain the requirement of an aromatic ring in the position of the benzyl ring, and 12 fragments were selected. A number of pyrethroid variants were produced with the fragment attached at the meta position as this is the active structure (O'Reilly et al, 2006).

Secondly, these fragments were then used to substitute the deltamethric acid, these were attached at the para position as occurs in other commercially produced pyrethroids (Figure 7.11). These substitution variants were docked and scored.
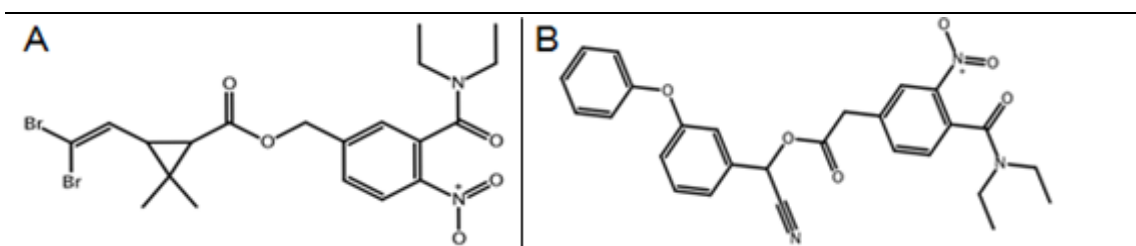
Figure 7.11 Examples of pyrethroids with substitutions for low scoring fragments. (A) deltamethrin with the alcohol group substituted for a fragment (P3) attached at the meta position. (B) deltamethrin with the acid group substituted for the fragment attached at the para position.
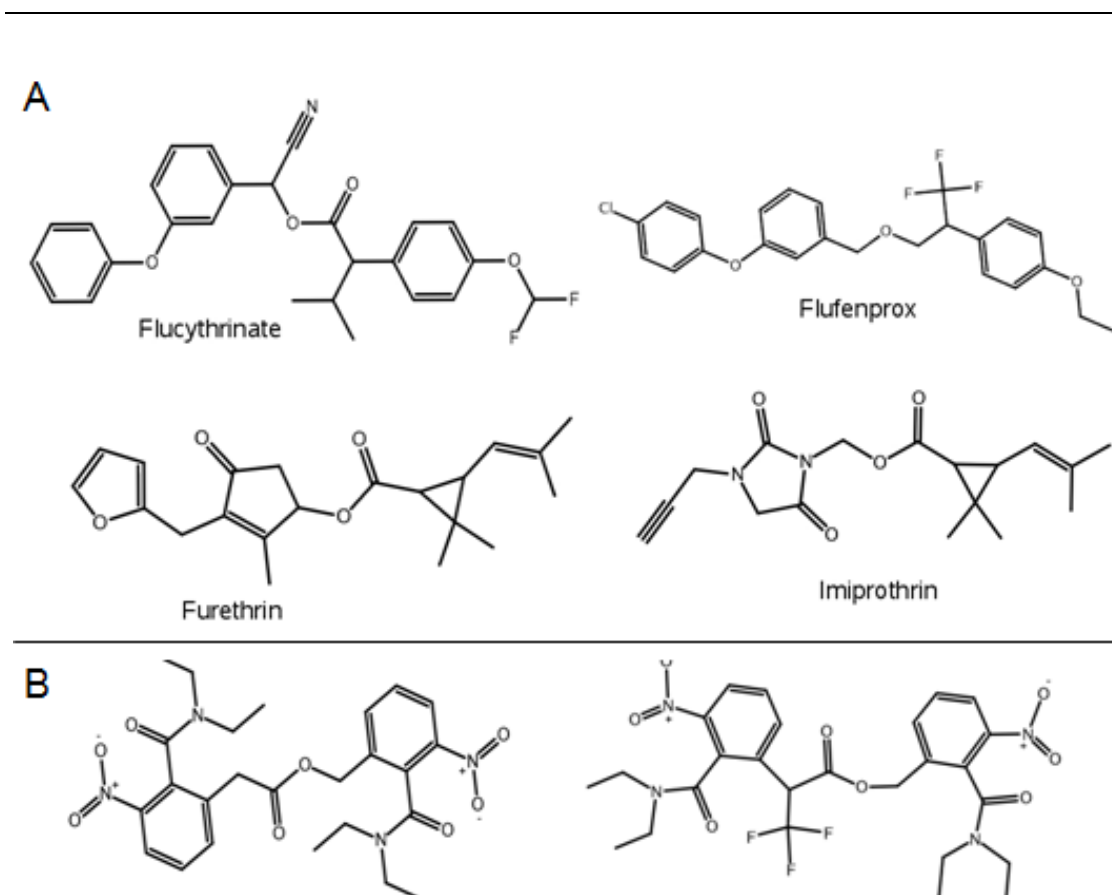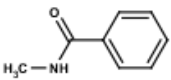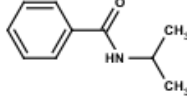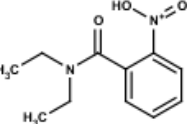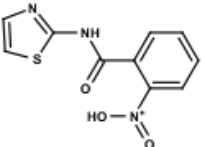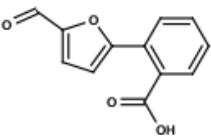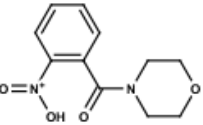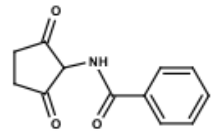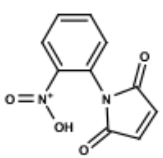
Figure 7.12 (A) Examples of commercially produced pyrethroids with alternative acid and alcohol groups. (B) Examples of a pyrethroid with both acid and alcohol groups substituted for a low scoring fragment (P3).

Table 7.9 ChemScores (kJ/mol) of dockings of acid, alcohol and dual substitutions.

Name	Fragment	Alcohol	Acid	Dual	Duel flu
P1		34.4	39.7	31.5	30.4
P5		37.0	44.5		
P6		33.7	37.3	29.3	33.9
P7		34.9	39.1	35.6	30.6
P8		40.5	44.2		
P9		34.8	35.5	29.7	28.5
P10		35.8	39.4		
P14		34.5	35.0	25.3	20.2

As none of the phenoxybenzyl or acid substitutions scored below the cutoff, a second set was constructed containing only the required ester bond with both the acid and alcohol groups substituted with low scoring scaffolds to produce dual substitutions. A number of commercially produced pyrethroids differ in having alternative acid and alcohol groups. These dual substitutions resemble commercial pyrethroids such as flufenprox and flucythrinate by having aromatic acid and alcohol groups, or imiprothrin and furethrin in having a penta heterocycle (Figure 7.12, A). An additional set were

based on the flufenprox structure with the dimethyl group substituted for a CF3 and an aromatic group substituting the deltamethric acid (Figure 7.12, B). This addition also acts to match the acid pharmacophore suggested by Byberg et al. (1987).

All alcohol substitution variants scored lower than the native deltamethrin while one scored just above the cutoff of 30 kJ/mol. However, variants with substituted acid groups scored consistently higher than those with substituted alcohol groups, further indicating that the phenoxybenzyl group has an important role in binding (Table 7.9).

7.3.3.2 Acid and alcohol dual substitution

A number of the dual substitutions scored below the threshold for binding and others scored just above the threshold (Figures 7.13). Aside from a poor score, some of these also bind non-productively. Two non-productive modes occur. One of these occurs with P14 and involves two π-stacking interactions between the rings of both groups and the BC loop F110 and the FG loop F215 (Figure 7.14, A). In a second mode, π-stacking also occurs with rings of both groups and F110 and F376 of SRS5. H-bonds also occur and may contribute but differ between ligands (Figure 7.14, B).
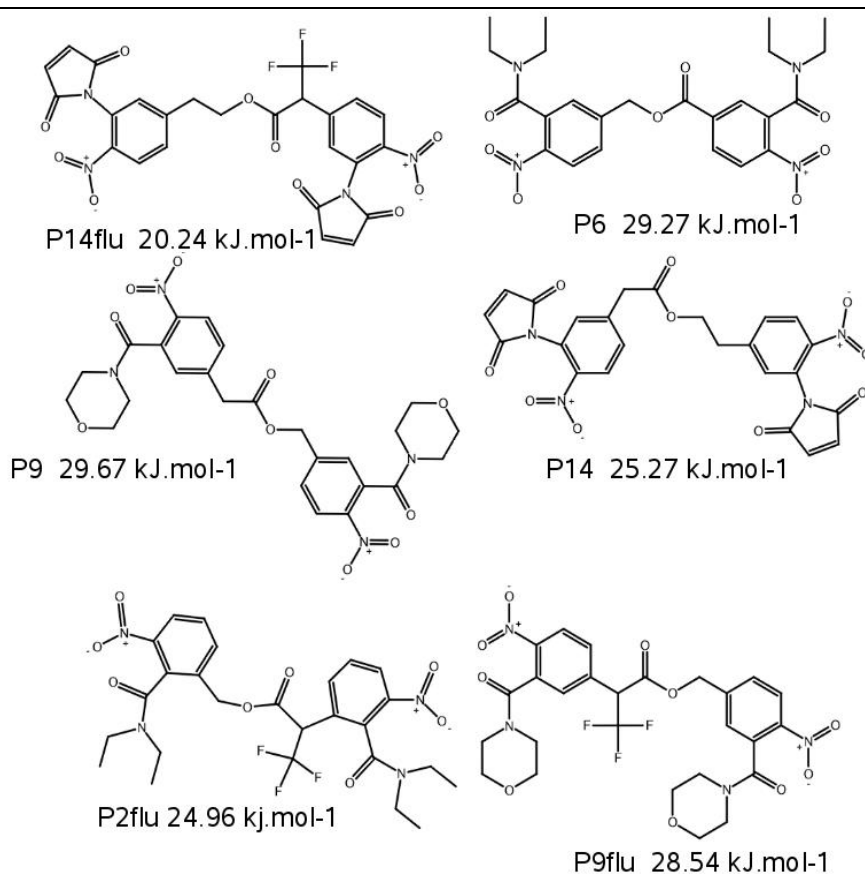
Figure 7.13 The structures of poor binders of CYP6M2.
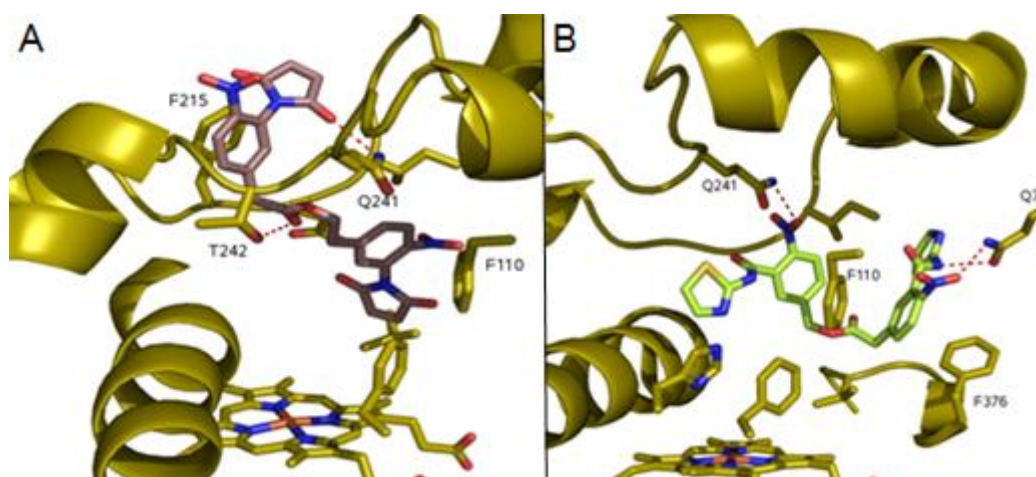
Figure 7.14 P14 (A) and I7 (B) binding in a non productive modes in CYP6M2.

As it is known that the size of the ligand has an affect on binding score as larger ligands make more contacts, the docking scores were correlated with molecular weight and number of heavy atoms to determine if a poor score was due to a smaller size of ligand. A correlation plot shows a correlation between ChemScore and either molecular weight or heavy atom number with the larger ligands scoring poorly. The low score could not be attributed to a reduced size of the ligand (Figure 7.15). There was no correlation between score and number of rotatable bonds indicating that the low score could not be attributed to the torsional entropy penalty (Figure 7.16).
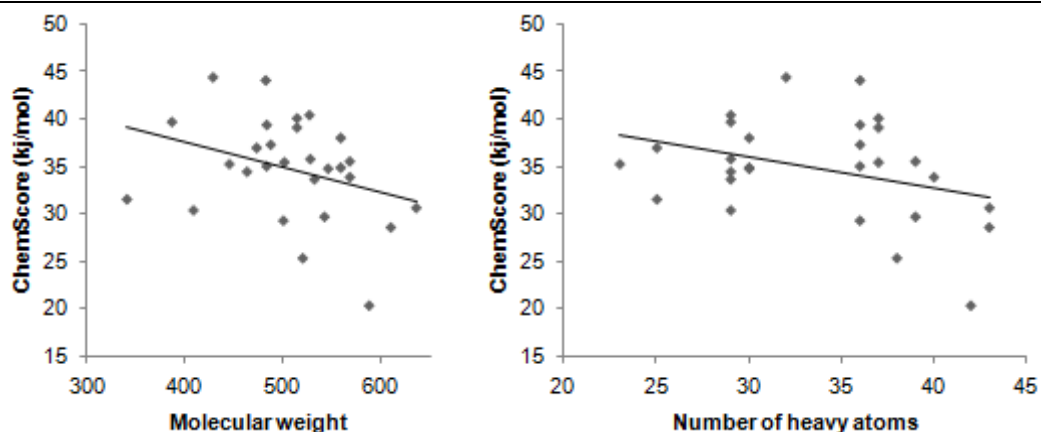
Figure 7.15 Correlation between number of heavy atoms, molecular weight and score.

Lipinski et al. (1997) suggested a 'rule of 5' that predicts drug absorption. There is poor absorption when there are more than 5 H-bond donors 10 H-bond acceptors, a molecular weight above 500, and a logP greater than 5. The dual substituted pyrethroids fit the criteria of having no more than 5 H-bond donors, not more than 10 H-bond acceptors and a logP less than 5. However, they have molecular weights above 500 daltons.
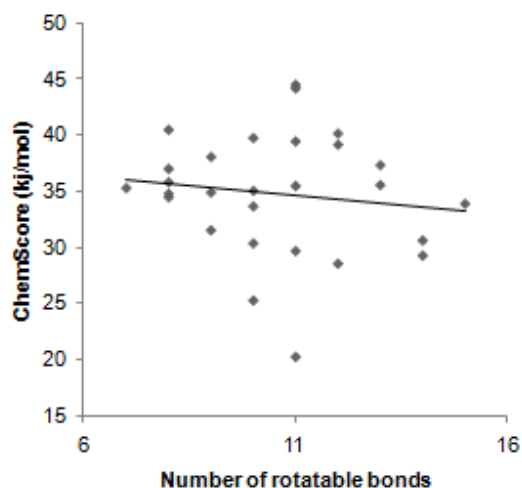
Figure 7.16 Correlation between the number of rotatable bonds and score.

7.3.4 Pyrethroid activation

Some insecticides are non toxic but become toxic by activation by P450s. This has also been previously attempted with pyrethroids. Scott et al. (1986) designed Pyr-III to be both a pyrethroid and synergyst by incorporating a methylinedioxyphenyl (oxidase inhibiting) moiety, the addition of this mechanism inhibitor group lead to a low resistance.

Other activated components could be incorporated into pyrethroids so that their metabolism could yield toxic metabolites rather than inactive products. The design of such insecticides is similar to the design of anti cancer drugs such as cyclophosphamide. Cyclophosphamide requires activation at the 4 position to form 4-hydroxycyclophosphamide which is a precursor for the metabolites acrolein and phosphoramide mustard (Figure 7.17). In this study, deltamethrin variants with a cyclophosphamide moiety were designed to place the 4 position of activation at equivalent positions as the sites of metabolism of the native deltamethrin so that the variant could be activated upon metabolism. The toxic metabolites would be produced where the P450s are expressed, in the brain and midgut of the insect.
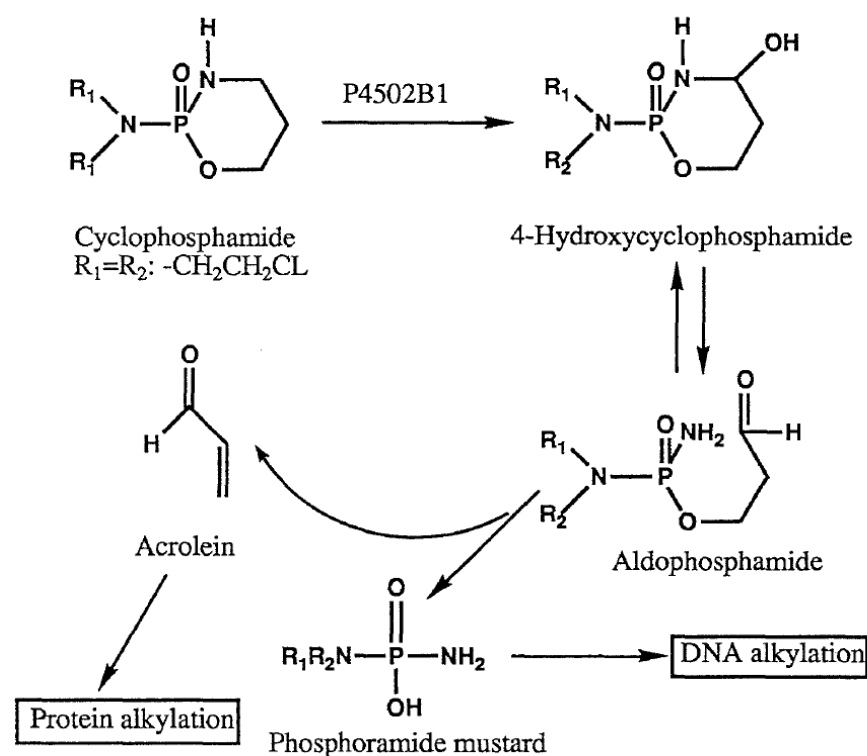
Figure 7.17 Cyclophosphamide activation by CYP2B1 taken from Doehmer et al. (1993).

Substitution of either the acid or alcohol with a cyclophosphamide group produces modes that allow metabolism, but modes that allow activation also occur. The alcohol substituted deltamethrin prefers to bind with the acid group above the heme and the mode of binding placing the 4 position is low scoring. An alternate structure substituting the acid group for a cyclophosphamide group was also produced but this also preferentially binds for metabolism on the 4' and only binds for activation in a low scoring mode (Figure 7.18 and Table 7.10). However, as both of these variants are capable of binding CYP6M2 for metabolism at the 4 position required for production of the toxic metabolite, the substitution for a cyclophosphamide group could be used to produce a toxic metabolite upon metabolism of the pyrethroid.

Table 7.10 ChemScores (kJ/mol) of dockings of the alcohol substituted (alc) and acid substituted (acid) binding for metabolism and activation.

	rnk	Score	ΔG	S(hbond)	S(metal)	S(lipo)	H(rot)	ΔE(clash)	ΔE(int)
alc(metabolism)	1st	43.6	-47.8	2.6	0.0	322.1	1.5	2.4	1.8
alc(activation)	10th	33.8	-37.0	0.0	0.0	303.5	1.5	0.2	3.0
acid(metabolism)	1st	37.0	-39.3	2.7	0.0	255.2	1.9	0.1	2.2
acid(activation)	8th	31.4	-33.4	0.0	0.0	281.3	1.9	0.5	1.6
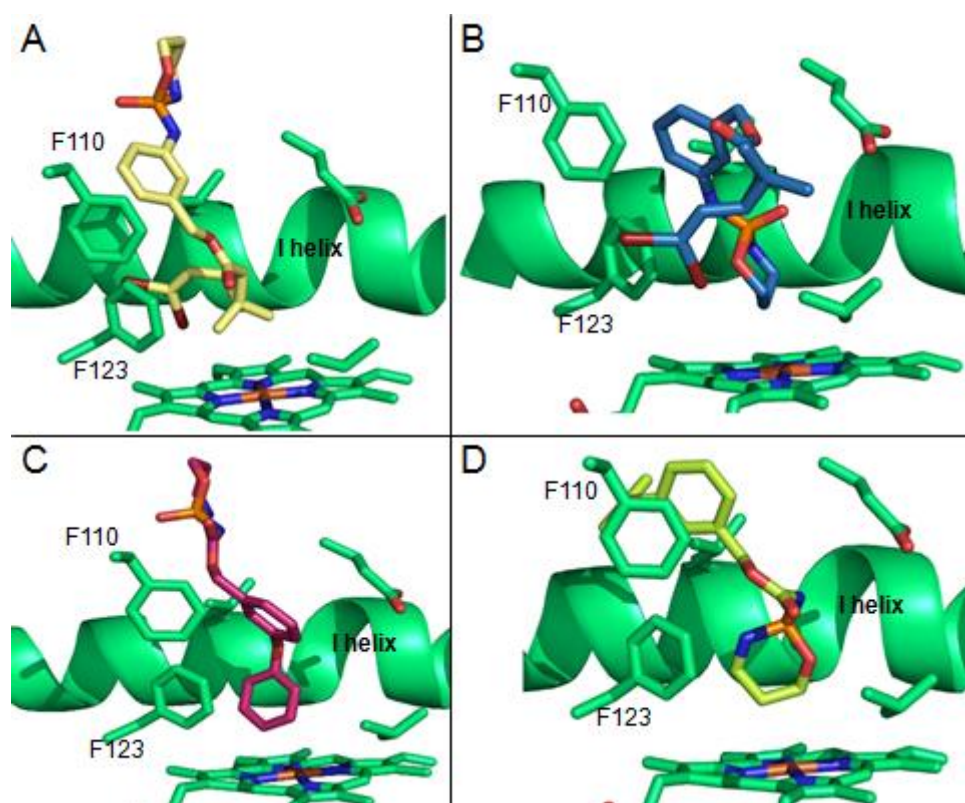
Figure 7.18 The cyclophosphamide alcohol substituted pyrethroid binds preferentially for metabolism on the acid (A), but a low scoring mode allowing activation occurs (B). The acid substituted pyrethroid prefers to bind for metabolism at the 4' (C), but a low scoring mode allowing activation occurs (D).

7.4 Probe Design

Understanding the binding of pyrethroids can be used to design activity-based probes that can label pyrethroid metabolising P450s. A pyrethroid like probe that contains an acetylene group can covalently bond to the protein upon metabolism. A second acetylene group could bind to a label to allow detection (Wright and Cravatt, 2007). Deltamethrin mimics with two acetylene groups were designed and docked into P450 models to determine an optimum design (Figure 7.19). These results were compared to the experimentally derived binding data (Figure 7.20).

Code	Chemical structure	Click handle Position	Notes
P1		Terminal	General P450s Probe
P2		Terminal	Permethrin mimic probe Type I pyrethroid
P3R		Terminal	Type II pyrethroid mimic probe
P4S		Terminal	Type II pyrethroid mimic probe
P5R		Terminal	Type II pyrethroid mimic probe
P6S		Terminal	Type II pyrethroid mimic probe
P7RS		Middle	Deltamethrin mimic Probe
P8RS		Middle	Deltamethrin mimic Probe

Figure 7.19 The deltamethrin probes tested (McLaughlin, unpublished).
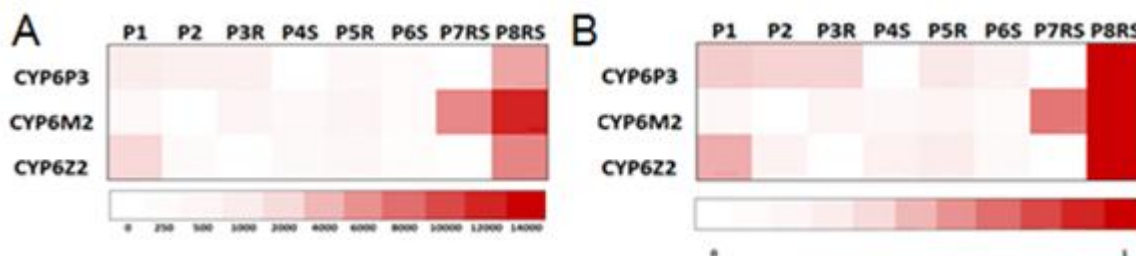
Figure 7.20 Heat maps (McLaughlin, unpublished) illustrating probe labelling profiles for individual P450 enzymes. (A) Absolute fluorescence signals of probe labelling events.(B) Normalized fluorescence signals of probe labelling events, where data for each P450 enzyme are shown as a ratio of the strongest labelling signal for that enzyme. "1" is the strongest binding event for an individual P450.

Most of the probes tested show little labelling, while probes P7 and P8 show the highest labelling. This is reflected in the docking scores where probes P7 and P8 show the highest binding scores.

In CYP6P3, P7 and P8 probes can bind in modes that allow the metabolism of the acetylene and also place the ketene intermediate in a position that facilitates nucleophilic attack on I helix Thr. The modes that bind close to the heme for efficient metabolism score higher in P8 than in P7. This may indicate that P8 may be preferred due to a better fit to the active site. The poses also differ in the position of the click handle. The poses for P7RS place the click handle projecting out of the active site in a position that may allow less labelling. The P8RS place the click handle projecting into the cavity of the active site where it may be in a better position for labelling (Figure 7.21).
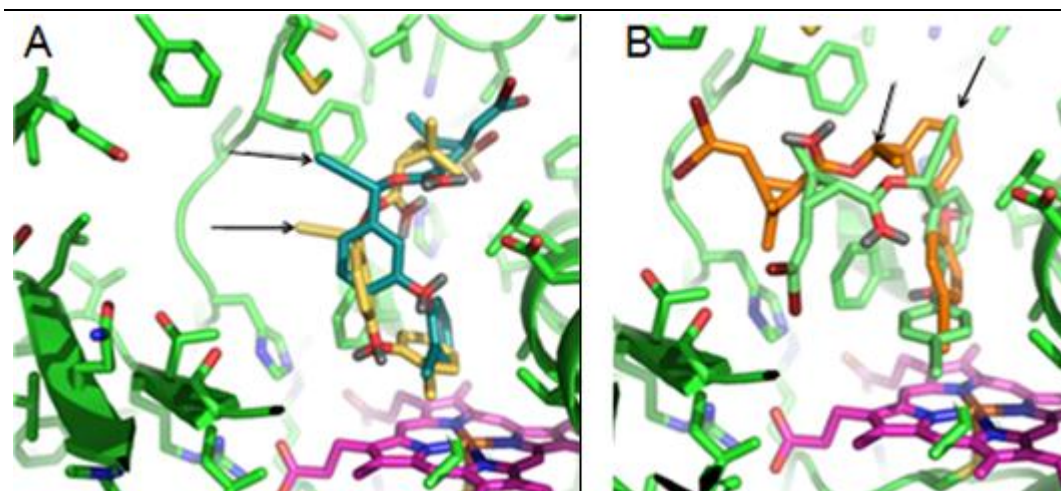
Figure 7.21 P8R and P8S bind with their click handles projecting into the active site (A). P7S and P7R bind with their click handles projecting out of the active site (B).

Probe P7 shows a selectivity for CYP6M2 with a higher labelling in this isoform. The selectivity for P7 in CYP6M2 cannot be fully explained in the dockings alone. The scores for P7 are lower than P8. However, P7SR docks in a different mode in CYP6M2 than in CYP6P3 that places the click handle in a different position. In CYP6P3 the click handle of P7 is oriented out of the active site into the protein which may affect labelling. In CYP6M2 the click handle is oriented into the active site that might improve labelling.

In CYP6Z2 the labelling of P7 and P8 was not found to be due to metabolism. In CYP6Z2 although productive modes occur, modes that are similar to the unproductive modes of the pyrethroids also occur. The best ranked modes place the ligand in non productive poses in a pocket distant from the heme. This pocket is formed by the FG loop F222, F212 and the BC loop Y102 (Figure 7.22). These residues form aromatic interactions with the phenoxybenzyl group. Other hydrophobic interactions occur with the A helix and beta sheet. These residues may form a hydrophobic pocket to hold the

ligand distant from the heme. This may explain the experimental results showing labelling in the absence of metabolism.
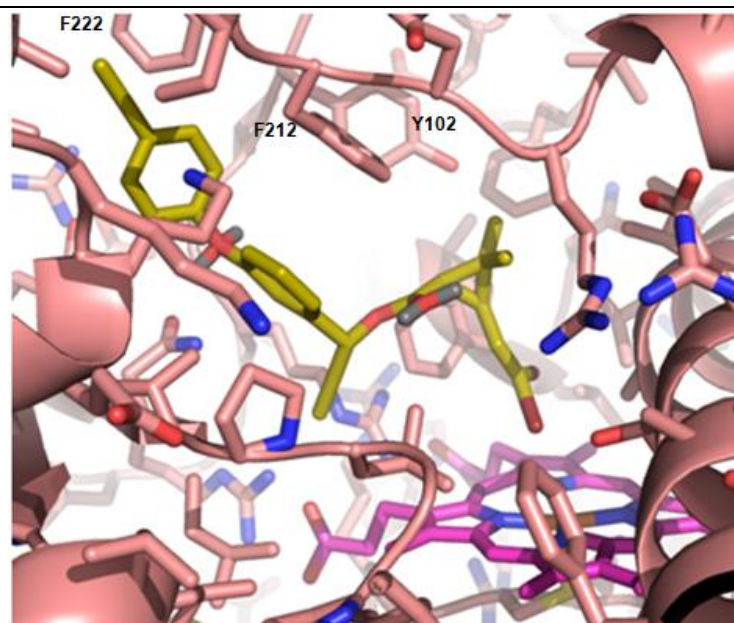
Figure 7.22 The positioning of the probe distant from the heme in CYP6Z2.

7.5 Conclusion

Insects use P450s to detoxify pyrethroids, and the homology models of the P450s involved in metabolism can be used to guide the development of novel pyrethroids that are resistant to metabolism.

Altering the structure of a pyrethroid is known to affect its metabolism by P450s, with a number of factors affecting metabolism such as the presence of a cyano group and the structure of the acid group. However, the structure of the alcohol group was found to have the largest effect on resistance, with the phenoxybenzyl group associated with resistance. The homology models produced in this study also indicate that the structure of the alcohol group has an affect on binding within the active site that may explain the changes in resistance towards pyrethroids with modified alcohol groups.

Identifying or predicting the site of metabolism of pyrethroids is useful as it can guide a number of developments. Sites of metabolism can be masked by the addition of halogens to produce pyrethroids that are metabolism resistant. Knowing sites of metabolism can guide the design of pyrethroids that are activated by metabolism into toxic products or to produce mechanism based inhibitors. Dockings of these designs into the homology models can screen for those that bind in the best positions or with the

best scores.

Identifying the binding interactions with active site residues can be used to produce pyrethroids that bind unproductively. Database screening can be used for scaffold based ligand design. Unlike conventional drug design, the design of poor binders requires novel methods such as the identification of scaffolds that are associated with poor binding.

Pyrethroids mediate their toxicity by binding to the sodium channel. An understanding of the requirements of pyrethroids for toxicity or binding to the sodium channel is required to guide the design of novel pyrethroids that have activity. The design of the pyrethroids in this study was restricted to contain the features shown to be required for toxicity.

Understanding the binding of pyrethroids can be used to design activity-based probes that can label pyrethroid metabolising P450s. Such probes can be used on undefined mixtures of P450s taken from pyrethroid resistant populations and used to identify the metabolising isoforms. The models showed that they could explain the experimental results and can be predictive in designing probes.

8. The role of Mal in malaria

8.0 Preface

Infection by malaria can lead to a severe inflammatory response in the host and the adaptor protein Mal is involved in the inflammatory pathway. Variations in Mal affect both the susceptibility and the hosts response to severe malaria and an understanding of the role of this protein can identify targets for drug development.

8.1 Inflammation

Inflammation is one of the mechanisms by which the innate immune system prevents infection. A normal inflammatory response localises infection but an unchecked inflammatory response can lead to inflammatory disorders. Severe malaria can develop into a diffuse encephalopathy called cerebral malaria (CM) that is a major contributor to mortality.

Toll like receptors (TLRs) are immune sensors that mediate the activation of the host innate immune response. TLRs are expressed at the cell surface and have an ectodomain that is involved in ligand recognition, a transmembrane domain, and a cytoplasmic domain that contains a Toll-IL-1 receptor domain (TIR) for signal transduction. Bacterial, viral or parasitic pathogen-associated molecular patterns are recognised at the cell surface and create changes in docking platforms within the TIR domains, allowing recruitment of adaptor proteins. The glycosylphosphatidylinositol (GPI) from the malaria protozoan parasite act as a ligand for both TLR2 and TLR4. Mal (MyD88-adapter-like), encoded by the gene TIRAP (SWISSPROT Acc. No. P58753) acts as a bridging adaptor in TLR2 and TLR4 signalling. Bridging adaptors act by delivering cytoplasmic signalling adaptors to the cell surface TLRs. After stimulation of TLRs, the signalling cascade causes the activation of NF-κB (nuclear factor that binds the kappa immunoglobulin light chain gene enhancer) and the activation of pro-inflammatory genes (Piao et al., 2008).

Mal has a central position in the TLR2 and TLR4 pathways and genetic variation of Mal affects susceptibility to disease. Mal S180L hetrozygosity has been shown to be protective against malaria as well as tuberculosis and bacteremia. The wild type Mal Ser180 is able to activate NF-κB while Leu180 was inactive. An excessive host inflammatory response causes individuals to be more susceptible to severe forms of malaria and this mutation is suggested to be involved in attenuating the interaction with TLR2 and reducing the NF-κB activation. However, too little signalling leads to an

inadequate response and heterozygotes have an optimal protection (Khor et al., 2007). The understanding the regulation of TLR signalling can lead to new treatments for infectious diseases.

8.1.1 Mal Cycle overview

The Mal Cycle has been proposed by Sheedy et al. (2007) (Figure 8.1). Signalling by TLR2 and TLR4 involves the adaptor protein Mal. Mal acts as a bridge to recruit the adaptor MyD88 (myeloid differentiation factor 88) leading to NF-κB activation.
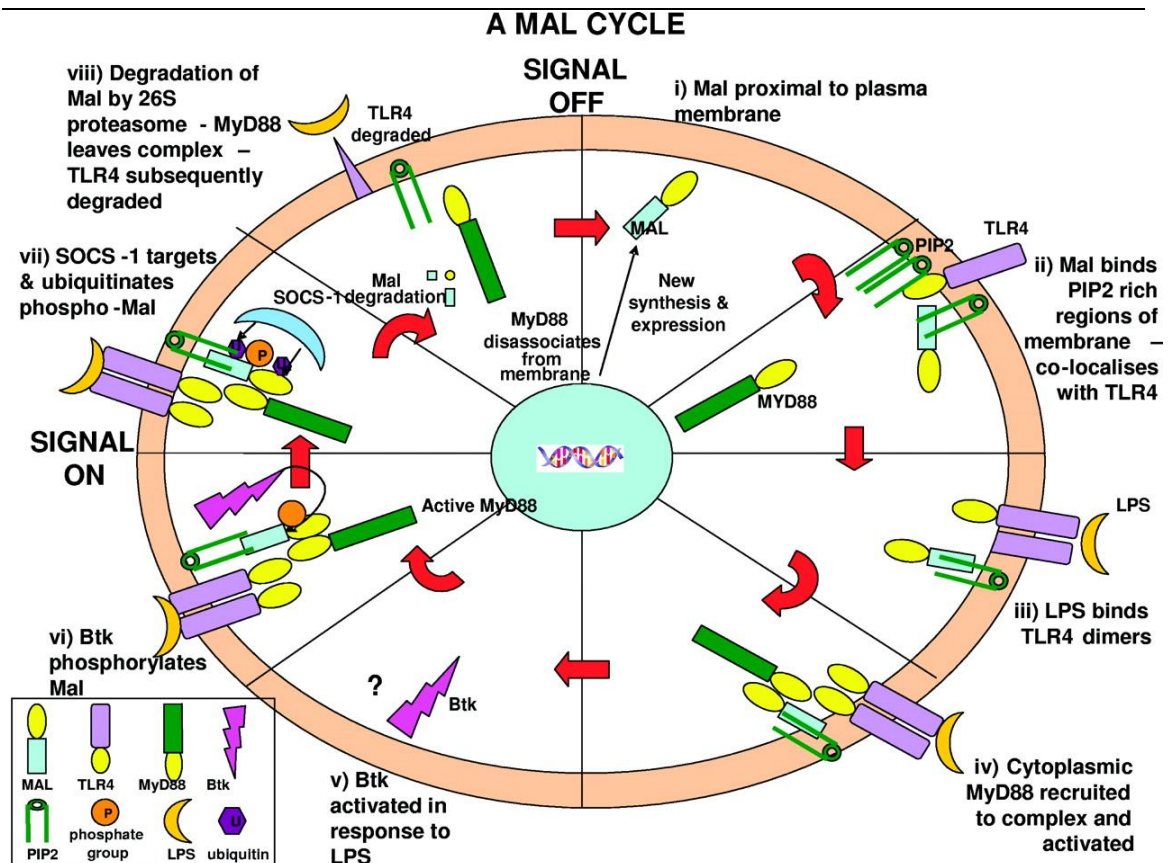
Figure 8.1 Taken from Sheedy et al. (2007). The Mal Cycle.

After activation, the TLRs interact with the adaptor proteins by their TIR domain and the adaptor recruits downstream molecules. Mal contains a TIR domain in its C terminus that interacts with the TIR domains of TLR4, TLR2 and MyD88. Mal also contains a PIP2 (phosphatidylinositol 4,5-bisphosphate) binding domain at the N-terminus (residues 15 to 35) which binds to PIP2, this recruits Mal to PIP2 rich regions in the plasma membrane and functions to recruit MyD88 to the membrane. Mal therefore acts as a bridging adaptor bringing MyD88 into the TLR signalling complex, MyD88 acts as a signalling adaptor leading to downstream events. Mal undergoes modification by tyosine phosphorylation by Bruton's tyrosine kinase (Btk) upon

activation by TLR2 and TLR4. This is required for downstream signalling to activate NF-κB. The phosphorylation of Mal is also required for its degradation and acts to terminate its signalling. Following phosphorylation, Mal is degraded by SOCS-1 that recognises a proline, glutamic acid, serine, and threonine rich area (PEST) N-terminal to its TIR domain.

8.1.2 Mal subcellular localisation

TLR2 and TLR4 require Mal to induce MyD88-dependant signalling and it has been proposed that Mal functions to recruit MyD88 to TLR2/4. Kagan and Medzhitov (2006) found that tagged MyD88 occurred as foci throughout the cytosol, but when expressed with Mal it was relocalised to the cell periphery and became enriched at the plasma membrane. Mal was therefore shown to deliver MyD88 to the membrane and in contrast, MyD88 was not required to deliver Mal to the cell surface. A Mal point mutation in the BB loop (P125H) to produce a non functional TIR domain was ineffective at delivering MyD88 showing that recruitment is TIR dependant. In addition, the PIP2 domain was found to be necessary as chimeras lacking this domain were also ineffective. Mal knockout cells are deficient in TLR4 signalling. This requirement for Mal in signalling could be bypassed by a MyD88 chimera with an added PIP2 binding specificity, indicating that Mal's function is to recruit MyD88 to PIP2 containing membranes for the initiation of TLR signalling. Monie et al. (2009) suggested that as Mal is not needed. MyD88 may bind to the receptor with a low affinity and that Mal only enhances the sensitivity of the signalling, or may be a 'sorting' adaptor and may not form a complex with the receptor.

8.1.3 Mal phosphorylation

Mal is phosphorylated during TLR signalling and tyrosines at positions 86 and 106 have been identified as possible sites. Wild type Mal strongly activates NF-κB while the mutations Y86F Y106F and Y187F are less active and have been identified as critical residues for NF-κB activity. Mal forms with mutations at either Y86 or Y187 act as dominant negative inhibitors of signalling (Gray et al., 2006). Piao et al. (2008) also found that the Y86A, Y106A and Y159A mutants increase the interaction with TLR4 allowing the mutants to act as negative inhibitors of TL4 activation. Gray et al. (2006). identified tyrosines at positions 86, 106 and 187 as possible phospho-accepting sites but not at position 159; while Piao et al. (2008) have identified tyrosines at positions 86, 106 and 159 as possible phospho-accepting sites and have suggested that mutagenesis

of tyrosine 187 did not affect Mal tyrosine phosphorylation. Both studies have shown tyrosines 86 and 106 as possible phosphorylation sites, while they do not agree on the roles of tyrosines 159 and 187. Therefore, the roles of Tyr86 and Tyr106 were investigated.

8.1.4 The role of Mal phosphorylation

Mal contains a TIR domain similar to that of CheY bacterial chemotaxis protein. TIR domains tend to consist of a central five stranded parallel β-sheet surrounded by five helices. The conserved residues tend to be buried in the core but some are solvent exposed. The TIR fold is similar to that of CheY, which is transiently phosphorylated at an Asp. This induces a conformational change in a Tyr residue in a loop between the fourth β-sheet and fourth α-helix (Dunne et al., 2003). Gray et al. (2006) suggested that as Mal is similar to CheY. A structural conformational change may occur following phosphorylation.

Mal is phosphorylated upon activation of the receptors and this phosphorylation is required for the activation of Mal. However, the roles of Mal and its phosphorylation are unclear. Piao et al. (2008) suggested that Mal could act exclusively at the level of the plasma membrane, associating with TLR4 to recruit cytoplasmic MyD88. In this case, Y86 could have a role in initiating conformational changes to the PIP2 domain. They also suggested an alternative mechanism where Mal acts as a shuttle to recruit MyD88 to TLR4 from the cytoplasm to the membrane. As phosphorylation of TRAM initiates its translocation from the membrane, phosphorylation of Mal could affect its translocation. The phosphorylation of Mal leads to a decreased association with TLR4. The signal-incompetent Y86A Mal with stronger associations may be retained at the membrane. Mal phosphorylation may confer the ability of Mal to shuttle from the membrane to bind non-activated downstream intermediates and deliver them to TLR4 and initiate assembly of the signalosome. They suggested that dephosphorylated Mal would be incapable of translocating from the membrane or not able to bind inactive, non-oligomerised cytoplasmic intermediates to recruit them to TLR4.

8.1.5 TLR binding

The initial step in signal transduction by TLR4 involves its dimerisation induced by binding to LPS. LPS binding causes the conformational changes in the receptor to allow dimerisation of the TIR domains to provide a new scaffold that allows recruitment of

adaptor proteins to form a signalling complex. The BB loop of TLR4 was proposed as the site of dimerisation as the BB loop peptide prevents dimerisation by competing with the receptor BB loops. The BB loop peptides from Mal, MyD88, TEAM and TRIF all block both Myd88 dependent and independent signalling by TLR4 indicating a 'receptor knockout' rather than an 'adaptor knockout' and indicating that the BB loops of these proteins interact with TLR4. The Mal BB loop peptides were weak inhibitors of TLR2 signalling, indicating that Mal does not use its BB loop in the formation of the TLR2 signalling complex (Toshchakov et al., 2005). Khor et al. (2007) proposed that Mal interacts with TLR2 via the DD loop on the opposite side of the protein to the BB loop. They mutated Ser180 that occurs on a surface exposed loop close to the DD loop and found that unlike the wild type, the S180L failed to bind to TLR2 but had no affect on its interaction with itself or with MyD88. Mal may therefore have different interactions with TLR2 and TLR4.

Mal could engage directly with the TLR4 dimer to create a new surface for the recruitment of MyD88. Nunez Miguel et al. (2007) suggested that Mal binds at the interface of the TLR4 homodimer at a site that overlaps the TRAM binding site. Although the Mal phosphorylated on the tyrosines is the active form of Mal needed for NF-κB activation, the phospho forms of Mal do not bind TLR4. The role of phosphorylation is not clear, the activation of NF-κB by LPs is strongly inhibited by Mal Y86F showing that tyrosine phosphorylation of this residue is critical for TLR4 signalling (Nunez Miguel et al., 2007).

8.1.6 MyD88

Activation of the TLR receptors involves dimerisation and recruitment of MyD88 and the protein kinase IRAK (interleukin 1 receptor-associated kinase). MyD88 is modular with a TIR domain and a protein interaction domain termed the death domain. The TIR domain interacts with the receptor and the death domain interacts with the IRAK N-terminus death domain. Post receptor signalling leads to the activation of the transcription factor NF-κB. MyD88 forms homodimers and activates IRAK-4 and IRAK-1. MyD88 binds to IRAK-4 and promotes phosphorylation of IRAK-1 by IRAK-4. In response to this IRAK-1 autophosphorylates and its interaction with TRAF6 (tumour necrosis factor receptor associated factor 6) leads to the activation of the inhibitory κB kinase (IKK) and mitogen-activated protein kinases and p38. These

kinases activate transcription factors such as NF-κB that produce the proteins for inflammatory responses (Loiarro et al., 2005).
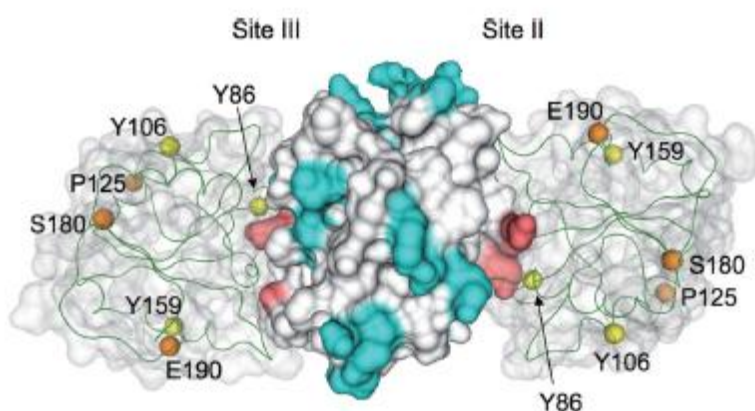
Figure 8.2 Taken from Ohnishi et al. (2009). The Proposed complexes of Mal and MyD88 with the critical residues in red.

Dunne et al. (2003) suggested non-overlapping binding sites on TLR4 for Mal and MyD88. Mal is predicted to interact with TLR4 through its DD and DE loops and MyD88 is predicted to bind via its AA, CD DD loop surfaces. They suggested that Mal and MyD88 also interact with each other to give a symmetrical arrangement via the hydrophobic area of the BB-loop and polar residues of the fourth α-helix. Ohnishi et al. (2009) solved the solution structure of MyD88 showed through mutagenesis that the TIR domain of MyD88 has two binding sites for Mal and it is possible that two Mal-TIR can make simultaneous contact with one MyD88 TIR (Figure 8.2). MyD88 also forms homodimers mediated by a direct TIR-TIR interaction. The BB loop is conserved in TIR domains and is suggested to be at the interface of the interaction. This was demonstrated by Loiarro et al. (2005) because a peptide consisting of the BB loop was able to inhibit MyD88 dimerisation and a mutation in the BB loop abolishes the ability of TLR2 to interact with MyD88.

8.2 Mal Modelling

Structural studies of Mal were based on two homology models produced by our collaborator (Niazi, unpublished) who found that BLAST indicated a number of TIR containing proteins. For modelling the TIR domain of human TLR2 (1FYW), TIR of human TLR1 (1FYV), and TIR domain of Myd88 (2JS7, 2Z5V) were selected as templates.
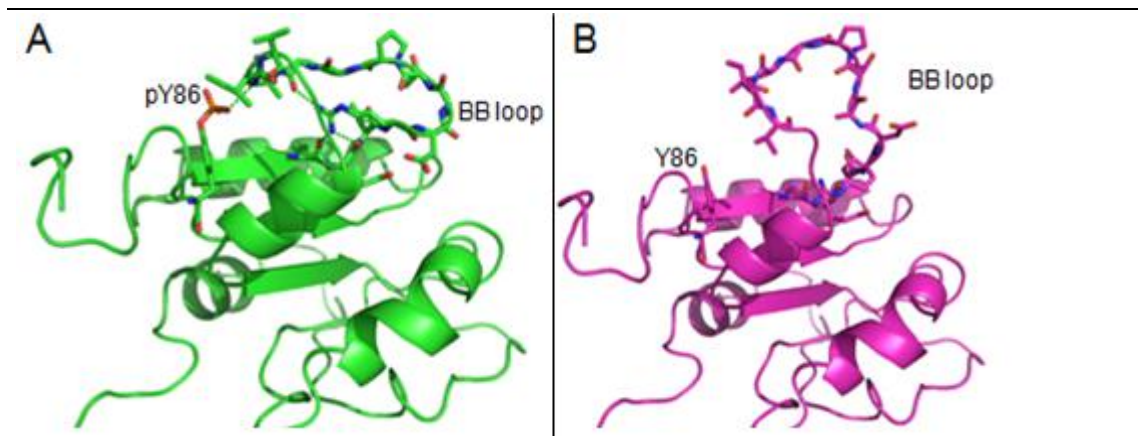
Figure 8.3 The Mal TIR domain with Y86 phosphorylated (A), and normal Y86 (B). In the phosphorylated Mal the BB loop is in a closed conformation while in the wild type the BB loop is open.

<u>8.2.1 Mal Models</u>

The two models produced by Niazi (unpublished), comprised the wild type with a tyrosine and a phosphorylated form with the Tyr mutated to phosphotyrosine (P-Mal). The model of the wild type was produced by MODELLER, and the tyrosines mutated to pY in the phosphorylated form. The BB loop area was subjected to loop refinement with the highest scoring structures selected. It was observed that the phosphorylation of Tyr allowed it to interact with the BB loop and could stabilise the BB loop in a closed conformation that may provide stable interactions with other proteins. The normal Tyr has less negative charge and hydrogen bonding potential which may allow the BB loop to have an open and more mobile conformation, which may not provide a stable interface for binding (Figure 8.3). As phosphorylation is required for function this may represent the active conformation while the native form is inactive.

8.3 Zdock docking of MyD88

Ohnishi et al. (2009) mutated residues of MyD88 at putative binding sites and identified two residues on opposite sides of the protein, R196 (Site II) and R288 (Site III), that decreased the affinity of MyD88 for Mal, and double mutations of these abolished affinity. They also found that the contributions from site II and III were comparable. Mutations at site I had no effect (Figure 8.4). MyD88 was docked with the active phosphorylated Mal and the inactive wild type Mal.
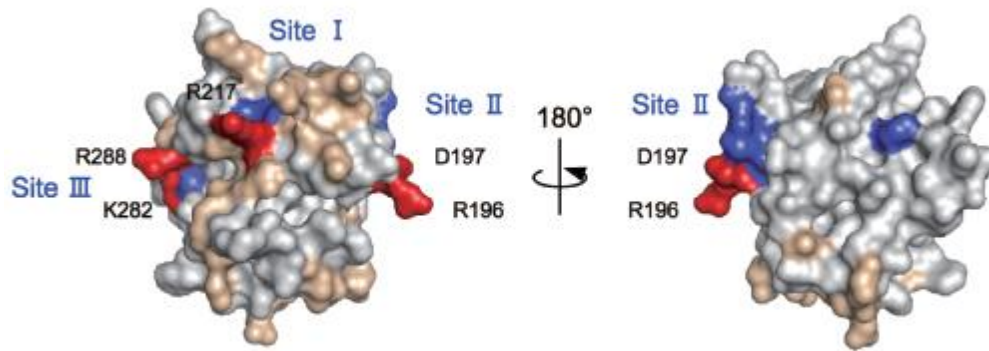
Figure 8.4 Taken from Ohnishi et al. (2009). Residues indicated in red are required for interaction with Mal.
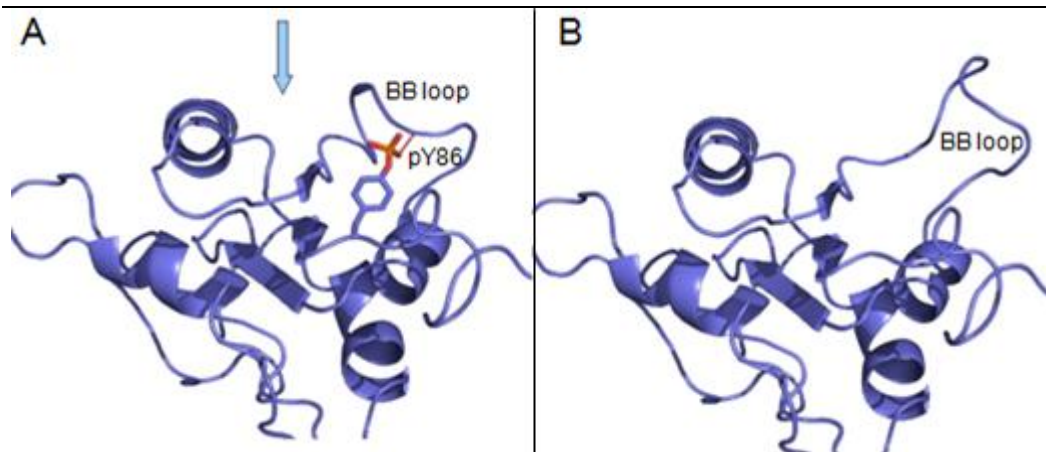
Figure 8.5 (A) P-Mal active conformation, pY86 forms H-bonds (red line) with the BB loop to form a deep concavity (arrow). (B) The native, inactive form with a shallow concavity.

8.3.1 Mal structure affects binding to MyD88

The active and inactive conformations of Mal differ in that the active conformation of the BB loop provides a more concave surface for binding while the surface of the inactive Mal is flatter (Figure 8.5) and this was found to affect the docking poses.

In the dockings with P-Mal, a cluster of dockings place the MyD88 into the concave surface produced by the BB-loop. This Mode 1 places P-Mal with its BB loop residues in contact with the BB loop of MyD88. The BB loop of MyD88 binds within the concave surface formed by the active conformation of the P-Mal BB loop. In this pose, R196 at Site II in the BB loop of MyD88 forms the contact surface with Mal. Loiarro et al. (2005) suggested that the BB loop of MyD88 can be subdivided into a conserved charged portion (RD) a conserved B-turn portion (PG) and a central linker. In the docking, the B-turn residues PG lie within the concavity formed by the BB loop of P-

Mal (Figure 8.6).
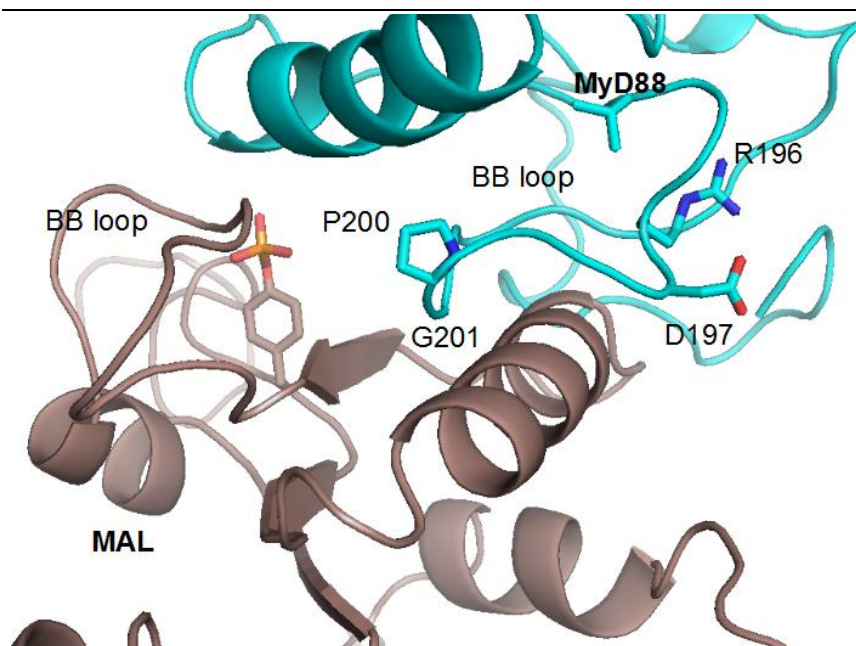
Figure 8.6 The PG beta turn (P200 and G201) of MyD88 projects into the Mal cavity formed by the BB loop.

An alternative docking pose also occurs with R217 of Site III forming the contact surface with P-Mal. In this Mode 2, R217 is in a position to form H-bonds with the Mal helix αA, while helix αB lies within the concavity produced by the P-Mal BB loop (Figure 8.7). In both poses the arginines from Sites II and III form the contact surface which is supported by the findings of Ohnishi et al. (2009).

The wild type Mal was also docked with MyD88, and MyD88 could bind with the BB loop of Mal in two modes. MyD88 could bind in the shallow concavity formed by the Mal BB loop (Figure 8.8, B), and also on the surface of the BB loop (Figure 8.8, A). In both cases the important arginines also form the interaction site.

To evaluate the docking poses, FireDock was modified to recognise the phosphotyrosine so that the poses could be refined and scored. FireDock showed that the most favourable conformation is the docking placing the BB loop of MyD88 in the interaction surface with the BB loop of P-Mal. While this is high scoring, the equivalent pose with the wild type Mal was low scoring (Table 8.1).

Table 8.1 The FireDock scores of the ZDOCK complexes.

cmp	glob	aVdW	rVdW	ACE	inside	aElec	rElec	laElec	lrElec	hb	piS	catpiS	aliph
pY 1	-44.8	-45.13	31.77	-7.83	19.16	0	5.08	0	7.21	-7.88	0	0	0
pY 2	-34.59	-36.3	17.99	4.23	16.8	-57.68	5.65	-19.31	6.65	-1.87	-4.5	0	0
Wild 1	-19.36	-26.43	10.06	-2.48	13.77	0	14.03	0	10.26	-2.6	0	-0.5	0
Wild 2	-39.48	-33.55	14.32	-5.06	14.26	-19.04	11.46	-10.5	8.94	-2.27	-1.5	-0.5	-1
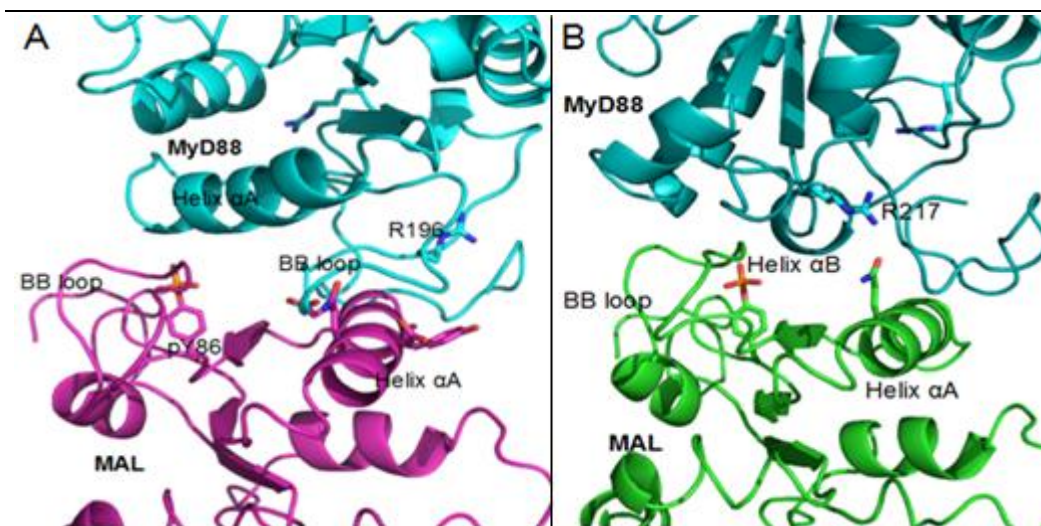
Figure 8.7 Alternative binding modes of MyD88 with P-Mal. Two surfaces of MyD88 can contact Mal, site II containing R196 (Mode 1, A) and Site III containing R217 (Mode 2, B).
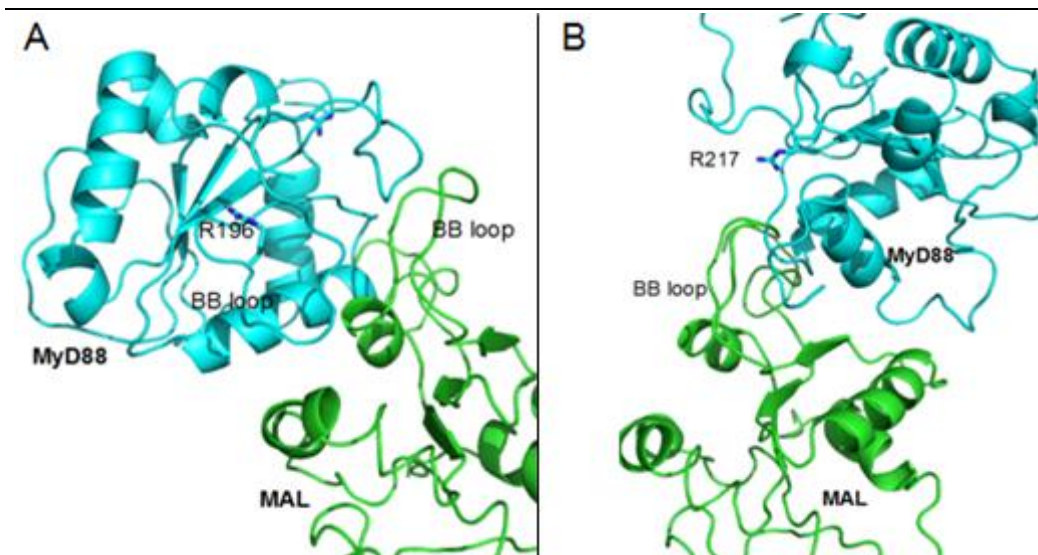
Figure 8.8 Alternative binding modes of MyD88 with wild type Mal. MyD88 can bind onto the surface of the BB loop (Mode 2, A) or in the concavity formed by the Mal BB loop (Mode 1, B).

The differences in score between these similar poses may be due partly to the 'inside' score. Insideness is scored favourably by FireDock as concave interfaces are more

likely than flat interfaces and significantly increases the success rate of the program. The active conformation of the BB loop in Mal creates a more concave surface for the binding of MyD88 than the inactive conformation. The binding of MyD88 into a concavity on the surface of Mal may act to stabilise the interaction and may be reduced in the inactive conformation. In addition, the area of contact for the wild type appears to be less than in the active conformation with a difference in the vdW score.

8.4 Dynamics

Molecular dynamics was carried out on the wild type and phosphorylated Mal to determine a role for pY86 in the dynamics of the BB loop. A 10 ns simulation was carried out for both models. Molecular dynamics simulations were performed with the molecular dynamics program NAMD 2.6 (Phillips et al., 2005) using the CHARMM22 force field and the TIP3P model for all water in the system.

To calculate the RMSD, the VMD RMSD trajectory tool was used. This tool calculates the RMSD from the starting frame. The BB loop was identified as residues 119 to 134 and the loop RMSD was calculated and found to be consistently higher in the wild type than in the pY model (Figure 8.9). In addition, in the wild type the BB loop moves into an open conformation after 8 ns.
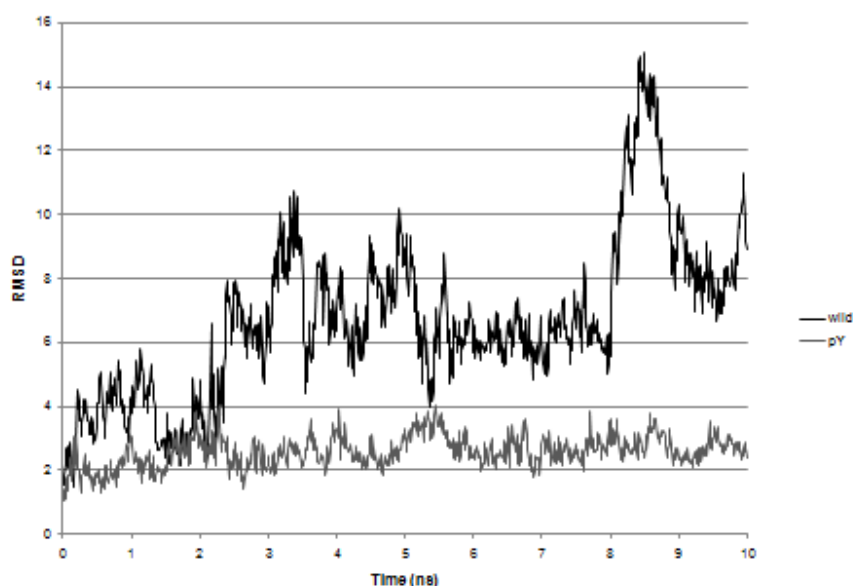
Figure 8.9 The RMSD of the BB loops was found to be higher in the wild type indicating a greater mobility.

8.4.1 Water Bridge

Water bridges have been shown to have a functional role in protein folding, architecture and conformational stability. As no adequate programs were found to identify water bridges, a custom script was written for this purpose (Appendix B). A water bridge was searched for between the BB loop and pY residue for the pY model and between the BB loop and Tyr for the wildtype. The BB loop was defined as residues 119 to 134.

Four types of water bridge were searched for:

where the water acts as an acceptor-donor (AD bridge):

Y O-----HO------HO protein

           H
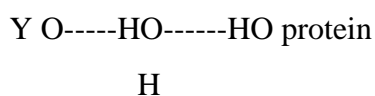
where the water acts as a donor-donor (DD bridge):

Y O-----HOH----O protein
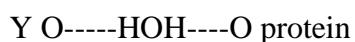
where water acts as a donor-acceptor (DA bridge):

Y OH-----OH-----O protein

          H
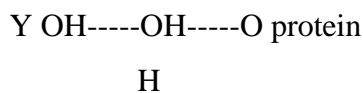
and where the water act as an acceptor-acceptor (AA bridge):

       H

Y OH-----O-----HO protein

       H
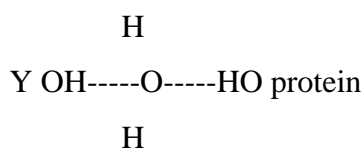
The percent occupancy of a bridge was defined as the number of frames with the bridge present divided by the number of total analysis frames. The lifetime of a specific bridge was defined as the number of consecutive frames from its first appearance until it is broken, The average lifetime is an average over all lifetimes.

For the pY, the AD bridge is present consistently throughout the trajectory with generally a consistent occupancy at each ns. An additional DD bridge occurs at a low frequency throughout most of the trajectory but towards the end of the 10 ns increases in occupancy and lifetime (Table 8.2). Throughout most of the trajectory a water forms

a AD bridge between the BB loop and pY. At 9 ns the BB loop approaches and begins to H-bond directly with the pY and the DD bridge increasingly occurs. The DD bridge occurs due to the closer proximity of the pY with the BB loop, the bridge forms a sharper angle than the AD bridge. The pY water bridge may stabilise the position of the BB loop and facilitate its movement towards the pY for direct H-bonding.

Table 8.2 The occupancy (Occ), maximum (Max) and average lifetime (Ave) of the AD and DD water bridges per ns of trajectory of the P-Mal. H-bonds were defined with a distance cutoff of 3.5 Å and an angle cutoff of 60º.

	P-Mal						Mal					
	AD			DD			DA			DD		
	Occ	Max	Ave	Occ	Max	Ave	Occ	Max	Ave	Occ	Max	Ave
Ns	(%)	(ps)	(ps)	(%)	(ps)	(ps)	(%)	(ps)	(ps)	(%)	(ps)	(ps)
1	27.6	48	6.1	2	3	1.3	3.1	12	3.4	1.4	5	2
2	57.3	39	5.3	6	13	1.7	2.6	7	2.4	2.5	5	1.1
3	53.9	48	9.6	5.6	8	1.8	7.8	22	4.3	0.3	1	1
4	29	30	5.2	1	1	1	0	0	0	0	0	0
5	55.5	27	3.7	8.8	8	1.5	0	0	0	0	0	0
6	52.5	61	5.4	6.8	11	1.6	0	0	0	0	0	0
7	67.9	40	5	7.6	6	1.7	0	0	0	0	0	0
8	63.6	36	5.1	9.8	15	2.1	0	0	0	0	0	0
9	56.9	38	4.8	11.5	11	2.1	0	0	0	0	0	0
10	60.5	76	5.7	13.3	12	3	0	0	0	0	0	0

For the wildtype a bridge initially occurs with residue E132, at 4ns the BB loop moves away from the Tyr and no further bridges are found (Table 8.2). Although a water bridge can form transiently, unlike the pY it is not as stable with a low occupancy and lifetime. This water bridge may fail to stabilise the BB loop as it has a higher mobility and towards the end of 3ns the BB loop moves away from the Tyr.

8.4.2 Summary

The pY bonds with the BB loop via a stable water bridge and by a direct H-bond towards the end of the trajectory. This may stabilise the BB loop shown with a low RMSD. In the wildtype a bridge is initially present but is unstable and disappears as the BB loop moves away from the Tyr and has a higher flexibility.

A similar anchoring of the BB loop to the rest of the Tir domain occurs in both TLR10

and TLR2. In the crystal structure of the TIR domain of TLR2 (1FYW) the BB loop contains the motif $RDx\Phi_1\Phi_2G$ (Φ is any hydrophobic residue and x any residue) the conserved R and D form an ion pair interaction with each other and the R forms an ion pair with the E at the end of the helix αA. Xu et al. (2000) suggested that this ion pair between the BB loop R and helix αA E stabilised the conformation of the BB loop. TLR10 was crystallised as a dimer involving the BB loops. Nyman et al. (2008) suggested that the residue F672 acts to anchor the BB loop to the rest of the TIR domain as well as making contacts within the dimer.

Mal acts as a bridging adaptor to bind MyD88 and bring it into a complex with the TLRs. Phosphorylation of Mal is required for its function and the results of this study suggest that phosphorylation of Y86 has a role in this. pY86 affects the stability and conformation of the BB loop that may act to bind MyD88.

8.5 SOCS1 (JAB, SSI-1, TIP3)

8.5.1 SOCS1 and Mal

Although Mal is a homologue of MyD88, it differs in having an N-terminal PEST domain, short lived proteins contain a PEST domain and undergo ubiquitination and degradation. Mal also undergoes rapid degradation within 30 min of activation of TLR2/4. Mal has been demonstrated to bind to SOCS1 and over expression of SOCS1 induces degradation of Mal. Mal also fails to be degraded in SOCS1 deficient macrophages. SOCS1 therefore negatively regulates TLR4 cytokine signalling by interacting with Mal. As Btk inhibitors inhibit the degradation of Mal, phosphorylation of Mal by Btk is followed by the binding of Mal to the SH2 domain of SOCS leading to ubiquitination (Kobayashi et al., 2006). Mal binds to SOCS1 to be polyubiquitinated on two N-terminal lysines which leads to its degradation by the 26S proteasome (Mansell et al., 2006).

8.5.2 SOCS structure

The SOCS (suppressor of cytokine signalling) family of proteins (SOCS1-7 and CIS1) contain three domains, a variable N-terminal domain, a central SH2 domain and a conserved SOCS box. The SOCS family are targeted to their substrate by means of the SH2 interaction domain, which interacts with phosphorylated tyrosines on the substrate. The SOCS box recruits the elongin BC complex for ubiquitination and subsequent degradation.

The SH2 structure is comprised of a central four stranded β sheet (βA-D) flanked by two α helices (αA and αB). The SH2 substrate binds across the central βD strand which separates the phosphotyrosine pocket (pY) from the hydrophobic +3 site where ligand specificity is determined (Figure 8.10). From the crystal structure of SOCS3 (2HMH) it was shown that the phosphate group of the tyrophosphate of the substrate binds in the phosphate binding pocket salt bridged to R71. Most of the contacts are made by the pTyr and residues C-terminal to the pTyr with the C-terminal enveloped by the EF and BG loops. The structures have been solved for SOCS2 (2C9W) and SOCS3 (2HMH), SOCS3 was used as a template for SOCS1 due to the similarities in function and substrate (Niazi, unpublished).

8.5.3 SOCS1 and Mal binding

The protein-protein interactions of Mal with SOCS1 were studied using ZDOCK carried out in this study and by our collaborator (Niazi, unpublished). The residues of the BB loop of Mal were blocked from the interface as that part is believed to be involved in an interaction with MyD88.

The SOCS1 sequence has the SH2 domain specific phosphotyrosine binding motif FLVRDS which forms electrostatic and hydrogen bonding contacts with pY containing ligands. In the SOCS3 crystal structure (2HMH), the Arg in this motif has a strong interaction with the phosphotyrosine of the ligand (GP130 antigen). The Y106 in Mal is a site of phosphorylation and it was noted that the region surrounding Y106 had sequence conservations with the SOCS3 ligand GP130.
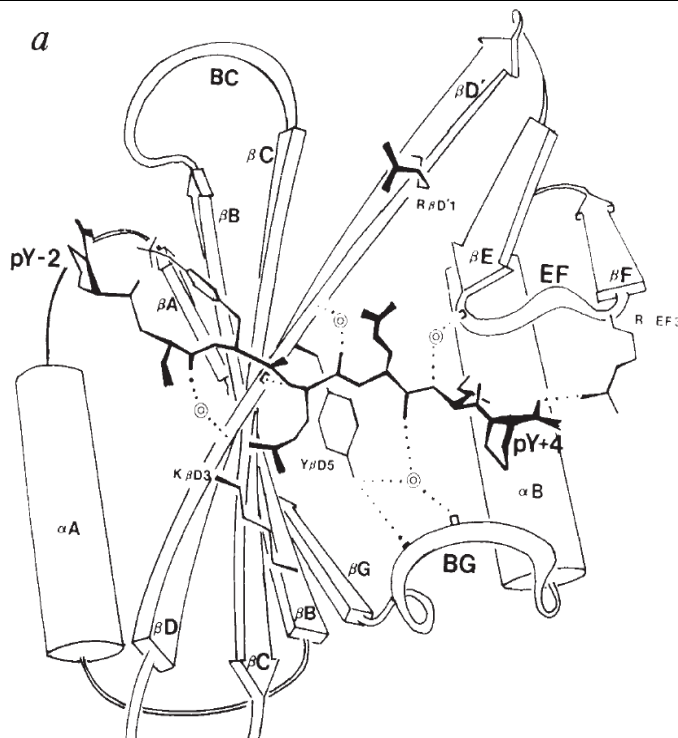
Figure 8.10 Nomenclature of the SH2 domain taken from Eck (1993).

In 2HMH the ligand interactions occur primarily through the pY, the pY can make 7 direct H-bonds to SOCS3 with a salt bridge to the R71 in the canonical phosphate binding site. In the SOCS1-Mal docking, the pY is also in a position to make H-bonds with the phosphate bindng site residues equivalent to those in SOCS3. In SOCS 1, arginines surrounding the phosphate binding site are in a position to interact with acidic residues on the Mal helix αA (Figure 8.11). This interaction may explain the specificity of SOCS1 for Mal, The basic residues in SOCS1 are in positions to interact with the acidic residues of Mal, and these residues are substituted in SOCS2/3 (Figure 8.12).

8.5.4 Charge complementarity

There are some sequence differences between SOCS1 and SOCS2/3. SOCS1 has an additional network of arginine residues providing a strong positively charged region surrounding the phosphate binding pocket (Figure 8.12). The dockings indicated that these arginines form a contact surface with Mal.
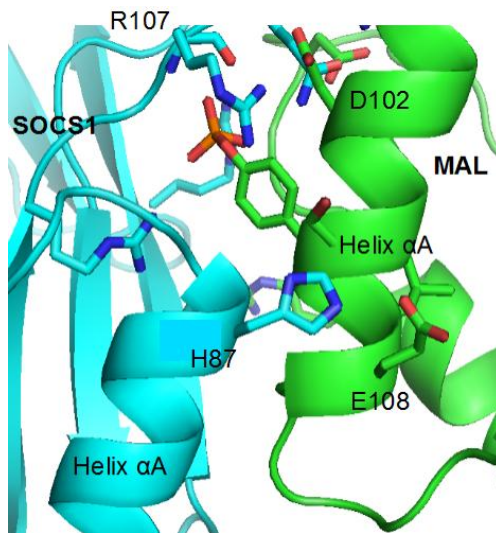
Figure 8.11 The ZDOCK docking of Mal and SOCS1. There is a charge complementarity between SOCS1 and Mal with the basic residues H87 and R107 on SOCS1 complementary to E108 and D102 on Mal.
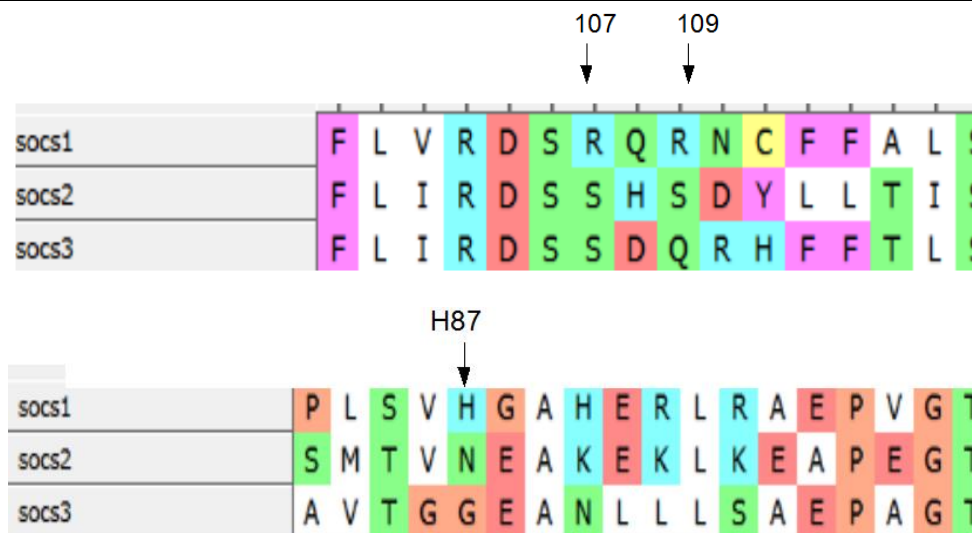
Figure 8.12 The alignment of SOCS1/2/3 with arrows indicating residues in SOCS1 at the binding interface with Mal.

Mal is an acidic protein and the surface surrounding Y106 is strongly negatively charged. There are complementary basic residues at the contact surface of SOCS1. H87, R107 and R109 only occur in SOCS1 and are in positions to interact with acidic residues (E108 and D102) that surround the pY106 when it is bound within the phosphate binding pocket. As these residues are substituted in SOCS2 and 3 this may contribute to the specificity of SOCS1 for Mal. The sequence differences between SOCS1 and SOCS2 and 3 affect the electrostatic surfaces as produced by DelPhi (Figure 8.13). The phosphate binding pocket in SOCS1 is surrounded by basic residues while in SOCS2 and 3 the region surrounding the pocket is less positively charged. The

positively charged surface of SOCS1 has a complementarity to the negatively charged surface surrounding Y106, in particular the residues surrounding the phosphate binding pocket.
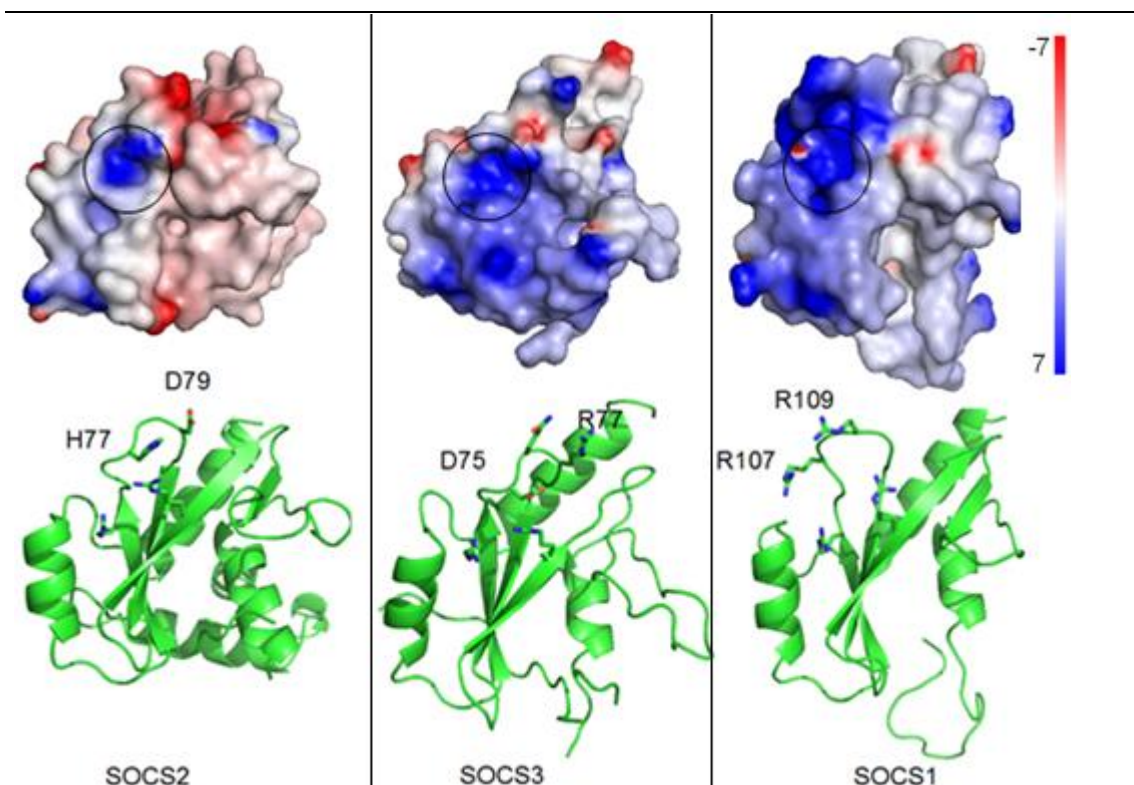
Figure 8.13 A comparison of the electrostatic surfaces of SOCS1, 2 and 3 (Blue: positive, red: negative, white: uncharged in units of kT/e). The loop of SOCS1 contains basic residues contributing to a positively charged surface surrounding the phosphate binding pocket (circle).

8.6 Conclusion

Toll like receptors are involved in the innate immune system and provides an inflammatory response to pathogens such as the malaria parasite. TLR2 and 4 are involved in sensing this parasite and require the adaptor protein Mal for signalling. Mal has a central role in the immune response to malaria and its variations affect the host response. During TLR2 and TLR4 signalling, Mal is activated by phosphorylation. The role of phosphorylation has not been clearly determined and this study has proposed suggestions for the roles of phosphotyrosines.

Mal binds with MyD88 and this interaction is thought to involve the BB loops and a role for Y86 in this interaction is proposed in this study. Y86 is located close to the BB loop and the pY86 is able to form water bridges and H-bonds with this loop that do not

form with the native tyrosine. These bonds may stabilise a conformation of the loop that promotes the interaction with MyD88.

Mal is rapidly degraded by SOCS1 upon phosphorylation and Y106 is proposed in this study to be the site of SOCS1 binding. This site is proposed based on similarities of the region of Y106 to the native substrate GP130, and due to the exposed nature of this tyrosine on the surface of the protein, and also by a strong charge complementarity between the acidic surface of Mal and the basic surface of SOCS1. SOCS1 has a specificity towards Mal and this study suggests that this is due to this charge complementarity, the SOCS1 phosphate binding site is surrounded by basic residues that are substituted in SOCS2 and SOCS3 which provides a positively charged surface surrounding the site.

By modulating the interactions of Mal, the hyper inflammatory responses that lead to severe malaria can be controlled. Understanding the nature of Mal with its binding partners is essential to identify targets for drug development.

9 Conclusions and Future Perspectives

9.0 Concluding Remarks

The work in this thesis involved the application of the computational techniques of computer-aided molecular design to give atomic level insight into the structure of the P450s under investigation, and their interaction with substrates. This work aims to complement experimental studies to understand the structure of the enzymes and identify a relationship between its structure and its activity towards its substrates. This work may be applied for the design of novel insecticides or to guide further experimental studies.

In this thesis, for a number of insecticide metabolisers and non-metabolisers homology modelling was used to build structural models. For a number of insect P450s experimental kinetic and metabolite data were obtained by collaboration with the Liverpool University School of Tropical Medicine, and for others data were obtained from the literature. Docking studies were carried out on these models using the experimental data to assess the validity of the models. Models were deemed to be valid if they could replicate the observed data by showing a correlation between the computed docking scores and the experimental IC50 values, and also if the docking poses of ligands within the active site replicated the known regiospecificity of metabolism.

The docking poses produced by the models gives an insight into the atomic level structure of the active site and the interactions between the ligand and protein. This gives an indication of the mechanism of substrate recognition and identifies residues involved in this interaction. By the comparison of active sites of P450s capable of pyrethroid metabolism and those that are incapable, it is possible to identify the interactions that are specific for the binding of pyrethroids. It was observed that there were commonalities between pyrethroid metabolisers and this could act as a fingerprint to identify these features in other P450s to predict their ability as a metaboliser. This fingerprint consists of aromatic residues in the active site that are in a position to bind with the aromatic moieties of the pyrethroid when bound in a productive mode. These π-anchors occur consistently in P450s across clades and appear to be a primary feature of substrate recognition. These aromatic residues occur in the BC loop for most of the metabolising enzymes identified and these were termed typical metabolisers. In a small number of metabolisers these aromatic residues occurred in other parts of the active site

but were still in positions to contact the substrate, these were termed non-typical metabolisers. These fingerprint residues act as a chemical and structural match for the features of the pyrethroid, and appear to be one of the requirements for the binding of the ligand to the active site. This study proposes the role of a pi anchor in the binding of pyrethroids. By identifying these features, it is possible to predict if a novel P450 is capable of metabolising pyrethroids. This can be used to guide experimental studies towards likely candidates.

Non metabolisers of pyrethroids were found to be diverse. Some non metabolisers contained the fingerprint residues found in the metabolisers, but had additional factors that prevented metabolism. These additional features include the presence of aromatic networks distant from the heme that act to prevent the ligand from accessing the heme. Other non metabolisers lacked the fingerprint residues; for these there were fewer complementarities between the pyrethroid and their active sites. Aromatic residues in the active site may be needed for the binding of pyrethroids and the lack of these residues is associated with a lack of metabolism. In addition, the presence of polar residues in the active site is also associated with a lack of activity. A hydrophobic active site may be required for the binding of hydrophobic pyrethroids.

The structure of the active site was investigated by the use of fluorescent ligands with alkyl chains of differing lengths. The preference towards particular chains was shown to be caused by residue substitutions within the substrate recognition site 5 (SRS5) and by the structure of this region. This site is also involved in determining the docking pose of pyrethroids and in determining its regiospecificity of metabolism. The CYP6M2 and CYP6D1 models aided an understanding of the regiospecifity of metabolism. They both primarily metabolise the 4' position and the dockings show that this preference is due partly to the structure of the active site placing this site for metabolism, and partly due to its high reactivity. Understanding how P450s bind and metabolise pyrethroids can be used to design novel pyrethroids that are resistant to metabolism by masking the sites of metabolism, or designing pyrethroids that bind poorly or are activated by metabolism.

Features of the active site alone may not be sufficient for the metabolism of pyrethroids as non-metabolisers may also share these conservations. Factors external to the active site were found to have an effect on substrate preference, including the interaction of P450s with the membrane. The orientation of the protein with respect to the membrane could affect the access of hydrophobic ligands from the membrane to the active site and may have an effect on its interaction with its binding partners. The interaction with b5 has been found by our collaborators to affect the metabolism of pyrethroids, and in this study it was proposed that this may be due to its role as a second electron donor or due to a stabilising of the P450 conformation.

An excessive inflammatory response is involved in the development of the severe forms of malaria and Mal is an adaptor protein involved in this inflammatory pathway. Phosphorylation of Mal is required for its activation and the Mal model and its interaction with its binding partners was used to give an insight into the roles of phosphorylation. The role of Y86 was suggested to be involved with its interaction with the signalling adaptor protein MyD88. The phosphorylated Y86 is able to form bonds with the BB loop that are absent in the wild type. In the wild type the tyrosine has a lower H-bonding propensity and only forms an unstable transient bridge to the BB loop, which is more mobile and adopts an open conformation. The pY86 is able to form stable bridges and H-bonds to the BB loop which stabilised the loop in a closed conformation. This conformation may be preferred for the interaction with MyD88.

The phosphorylation of Y106 was suggested to be involved with its binding to SOCS1 which targets Mal for degradation and acts to limit the inflammatory response. Y106 is exposed on the surface of Mal and is surrounded by acidic residues. This position has a charge complementarity with the positively charged surface surrounding the SOCS1 phosphate binding pocket. In addition, this complementarity may be used to explain the specificity of SOCS1 for Mal. SOCS2 and 3 lack the basic residues that occur in SOCS1 with a less basic surface surrounding their binding pockets. This characterisation of Mal activation and degradation may be used to identify drug targets for the design of ligands that could be used to attenuate the inflammatory response.

9.1 Future perspectives

The models and results obtained in this thesis have given an insight into the structure of a range of P450s and their activity towards pyrethroids. The models of metabolisers such as CYP6M2 have given an understanding of the structure of the active site and how insecticides bind to produce the regiospecific metabolism. The identification of a pyrethroid binding fingerprint lays the groundwork for the further study of pyrethroid metabolising enzymes. This identification of residues involved in metabolism can be used to design new experimental studies using mutagenesis to conform these findings. This fingerprint may also act as a basis for the further identification of P450s involved in insecticide resistance.

The understanding of how insecticides bind to the active sites of P450s can be used for the design the next generation of insecticides that are resistant to metabolism. An understanding of binding can lead to insecticides that bind poorly to the P450s responsible for resistance, or to design inhibitors specific for these enzymes. An understanding of how the regiospecificity of metabolism is achieved can lead to the development of insecticides with the site of metabolism masked by inert groups, or may be used to place chemical groups that are activated by metabolism to become suicide inhibitors or to produce toxic products. The structures of the active sites of these models can be used to aid the design of new insecticides. This has implications as the increase of insecticide resistance in mosquito affects the ability to control the spread of malaria.

The models can be used for in silico screening of compound databases to find novel substrates and inhibitors that can be used to guide experimental work. The identification of structural scaffolds from these screens can be used to guide the development of novel insecticides. Such screens are cost effective as only the ones selected from in silico results need to be tested. The design of a probe that is specific to pyrethroid metabolising P450s can be used to identify novel P450s involved in resistance and to monitor the levels of resistance in populations, and to suggest changes in the use of currently used insecticides to control resistant populations.

Infection by malaria can progress into cerebral malaria, a severe form of malaria caused by an excessive inflammatory response in the host that is a cause of mortality. The adapter protein Mal is central to the inflammatory pathway and mutations in this protein can affect the susceptibility of the host to the severe forms of malaria. Malaria is sensed by the toll like receptors TLR2 and TLR4 that cause signalling to produce an inflammatory response. During this signalling Mal is activated by phosphorylation on tyrosines although the role this plays is not fully understood. In this thesis the predicted interactions between models of Mal and its binding partners as well as molecular dynamics simulations have offered explanations for the importance of phosphorylation. An understanding of these interactions is important for the identification of targets for drug development to modulate the immune response. This has implications for the design of drugs to prevent the progress of malaria into its severe forms, but also as a number of other inflammatory disorders involve the same pathway, this may have implications for a range of disorders. Mal is a target for drug design and this study identifies its binding interfaces with other members of the cascade. These findings may be used to target drug discovery.

The results from this thesis highlight the importance of experimental and computational studies to work in tandem to fully understand observations. Computational studies can explain experimental observations and be used to guide further experimental work.

References

Altschul, S.F., W. Gish, W. Miller, E.W. Myers, and D.J. Lipman. 1990. Basic local alignment search tool. *J Mol Biol*. 215:403-10.

Altschul, S.F., T.L. Madden, A.A. Schaffer, J. Zhang, Z. Zhang, W. Miller, and D.J. Lipman. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res*. 25:3389-402.

Andersen, J.F., J.G. Utermohlen, and R. Feyereisen. 1994. Expression of house fly CYP6A1 and NADPH-cytochrome P450 reductase in Escherichia coli and reconstitution of an insecticide-metabolizing P450 system. *Biochemistry*. 33:2171-7.

Anderson, A.C. 2003. The process of structure-based drug design. *Chem Biol*. 10:787-97.

Anderson, R.J., K.G. Adams, and C.A. Henrick. 1985. Synthesis and insecticidal activity of the stereoisomers of .alpha.-cyano-3-phenoxybenzyl 2-[2-chloro-4-(trifluoromethyl)anilino]-3-methylbutanoate (fluvalinate) and related analogs. *J. Agric. Food Chem*. 33:508-514.

Andrusier, N., R. Nussinov, and H.J. Wolfson. 2007. FireDock: fast interaction refinement in molecular docking. *Proteins*. 69:139-59.

Arnold, K., L. Bordoli, J. Kopp, and T. Schwede. 2006. The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling. *Bioinformatics*. 22:195-201.

Baker, M.T., and R.A. Van Dyke. 1984. Metabolism-dependent binding of the chlorinated insecticide DDT and its metabolite, DDD, to microsomal protein and lipids. *Biochem Pharmacol*. 33:255-60.

Baxter, C.A., C.W. Murray, D.E. Clark, D.R. Westhead, and M.D. Eldridge. 1998. Flexible docking using Tabu search and an empirical estimate of binding affinity. *Proteins*. 33:367-82.

Bennett-Lovsey, R.M., A.D. Herbert, M.J.E. Sternberg, and L.A. Kelley. 2008. Exploring the extremes of sequence/structure space with ensemble fold recognition in the program Phyre. *Proteins: Structure, Function, and Bioinformatics*. 70:611-625.

Berman, H.M., J. Westbrook, Z. Feng, G. Gilliland, T.N. Bhat, H. Weissig, I.N. Shindyalov, and P.E. Bourne. 2000. The Protein Data Bank. *Nucleic Acids Res*. 28:235-42.

Bohm, H.J. 1994. The development of a simple empirical scoring function to estimate the binding constant for a protein-ligand complex of known three-dimensional structure. *J Comput Aided Mol Des*. 8:243-56.

Braks, M.A., N.A. Honorio, R. Lourencqo-De-Oliveira, S.A. Juliano, and L.P. Lounibos. 2003. Convergent habitat segregation of Aedes aegypti and Aedes albopictus (Diptera: Culicidae) in southeastern Brazil and Florida. *J Med Entomol*. 40:785-94.

Bridges, A., L. Gruenke, Y.T. Chang, I.A. Vakser, G. Loew, and L. Waskell. 1998. Identification of the binding site on cytochrome P450 2B4 for cytochrome b5 and cytochrome P450 reductase. *J Biol Chem*. 273:17036-49.

Brown, R.P., C.M. McDonnell, M.R. Berenbaum, and M.A. Schuler. 2005. Regulation of an insect cytochrome P450 monooxygenase gene (CYP6B1) by aryl hydrocarbon and xanthotoxin response cascades. *Gene*. 358:39-52.

Casida, J.E. 1980. Pyrethrum flowers and pyrethroid insecticides. *Environ Health Perspect*. 34:189-202.

Cavasotto, C.N., J.A. Kovacs, and R.A. Abagyan. 2005. Representing receptor flexibility in ligand docking through relevant normal modes. *J Am Chem Soc*. 127:9632-40.

Chan, W.K., Z. Sui, and P.R. Ortiz de Montellano. 1993. Determinants of protein modification versus heme alkylation: inactivation of cytochrome P450 1A1 by 1-ethynylpyrene and phenylacetylene. *Chem Res Toxicol*. 6:38-45.

Chen, R., L. Li, and Z. Weng. 2003. ZDOCK: an initial-stage protein-docking algorithm. *Proteins*. 52:80-7.

Chen, R., and Z. Weng. 2003. A novel shape complementarity scoring function for protein-protein docking. *Proteins*. 51:397-408.

Chiu, T.L., Z. Wen, S.G. Rupasinghe, and M.A. Schuler. 2008. Comparative molecular modeling of Anopheles gambiae CYP6Z1, a mosquito P450 capable of metabolizing DDT. *Proc Natl Acad Sci U S A*. 105:8855-60.

Clarke, T.A., S.C. Im, A. Bidwai, and L. Waskell. 2004. The role of the length and sequence of the linker domain of cytochrome b5 in stimulating cytochrome P450 2B4 catalysis. *J Biol Chem*. 279:36809-18.

Cnubben, N.H., J. Vervoort, M.G. Boersma, and I.M. Rietjens. 1995. The effect of varying halogen substituent patterns on the cytochrome P450 catalysed dehalogenation of 4-halogenated anilines to 4-aminophenol metabolites. *Biochem Pharmacol*. 49:1235-48.

Colovos, C., and T.O. Yeates. 1993. Verification of protein structures: patterns of nonbonded atomic interactions. *Protein Sci*. 2:1511-9.

Comeau, S.R., D.W. Gatchell, S. Vajda, and C.J. Camacho. 2004a. ClusPro: a fully automated algorithm for protein-protein docking. *Nucleic Acids Res*. 32:W96-9.

Comeau, S.R., D.W. Gatchell, S. Vajda, and C.J. Camacho. 2004b. ClusPro: an automated docking and discrimination method for the prediction of protein complexes. *Bioinformatics*. 20:45-50.

Cosme, J., and E.F. Johnson. 2000. Engineering Microsomal Cytochrome P450 2C5 to Be a Soluble, Monomeric Enzyme. *Journal of Biological Chemistry*. 275:2545-2553.

Cruciani, G., E. Carosati, B. De Boeck, K. Ethirajulu, C. Mackie, T. Howe, and R. Vianello. 2005. MetaSite: understanding metabolism in human cytochromes from the perspective of the chemist. *J Med Chem*. 48:6970-9.

de Groot, M.J. 2006. Designing better drugs: predicting cytochrome P450 metabolism. *Drug Discov Today*. 11:601-6.

Djouaka, R.F., A.A. Bakare, O.N. Coulibaly, M.C. Akogbeto, H. Ranson, J. Hemingway, and C. Strode. 2008. Expression of the cytochrome P450s, CYP6P3 and CYP6M2 are significantly elevated in multiple pyrethroid resistant populations of Anopheles gambiae s.s. from Southern Benin and Nigeria. *BMC Genomics*. 9:538.

Duhovny, D., R. Nussinov, and H.J. Wolfson. 2002. Efficient unbound docking of rigid molecules. *In Proceedings of the Fourth International Workshop on Algorithms in Bioinformatics*. Vol. 2452.

Dunne, A., M. Ejdeback, P.L. Ludidi, L.A. O'Neill, and N.J. Gay. 2003. Structural complementarity of Toll/interleukin-1 receptor domains in Toll-like receptors and the adaptors Mal and MyD88. *J Biol Chem*. 278:41443-51.

Durr, U.H., L. Waskell, and A. Ramamoorthy. 2007a. The cytochromes P450 and b5 and their reductases--promising targets for structural studies by advanced solid-state NMR spectroscopy. *Biochim Biophys Acta*. 1768:3235-59.

Durr, U.H., K. Yamamoto, S.C. Im, L. Waskell, and A. Ramamoorthy. 2007b. Solid-state NMR reveals structural and dynamical properties of a membrane-anchored electron-carrier protein, cytochrome b5. *J Am Chem Soc*. 129:6670-1.

Eldridge, M.D., C.W. Murray, T.R. Auton, G.V. Paolini, and R.P. Mee. 1997. Empirical scoring functions: I. The development of a fast empirical scoring function to estimate the binding affinity of ligands in receptor complexes. *J Comput Aided Mol Des*. 11:425-45.

Elliott, M. 1971. The relationship between the structure and the activity of pyrethroids. *Bull World Health Organ*. 44:313–324.

Elliott, M. 1980. Established pyrethroid insecticides. *Pesticide Science*. 11:119-128.

Ewing, T.J., S. Makino, A.G. Skillman, and I.D. Kuntz. 2001. DOCK 4.0: search strategies for automated molecular docking of flexible molecule databases. *J Comput Aided Mol Des*. 15:411-28.

Feyereisen, R. 2006. Evolution of insect P450. *Biochem Soc Trans*. 34:1252-5.

Fiser, A., and A. Sali. 2003. Modeller: generation and refinement of homology-based protein structure models. *Methods Enzymol*. 374:461-91.

Floquet, N., J.-D. Marechal, M.-A. Badet-Denisot, C.H. Robert, M. Dauchez, and D. Perahia. 2006. Normal mode analysis as a prerequisite for drug design: Application to matrix metalloproteinases inhibitors. *FEBS Letters*. 580:5130-5136.

Friesner, R.A., J.L. Banks, R.B. Murphy, T.A. Halgren, J.J. Klicic, D.T. Mainz, M.P. Repasky, E.H. Knoll, M. Shelley, J.K. Perry, D.E. Shaw, P. Francis, and P.S. Shenkin. 2004. Glide: a new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy. *J Med Chem*. 47:1739-49.

Fujii, K. 1995. Isoflurane acts as an inhibitor of oxidative dehalogenation while acting as an accelerator of reductive dehalogenation of halothane in guinea pig liver microsomes. *Toxicology*. 104:123-8.

Gao, Q., C.E. Doneanu, S.A. Shaffer, E.T. Adman, D.R. Goodlett, and S.D. Nelson. 2006. Identification of the interactions between cytochrome P450 2E1 and cytochrome b5 by mass spectrometry and site-directed mutagenesis. *J Biol Chem*. 281:20404-17.

Garnier, J., J.F. Gibrat, and B. Robson. 1996. GOR method for predicting protein secondary structure from amino acid sequence. *Methods Enzymol*. 266:540-53.

Gigon, P.L., T.E. Gram, and J.R. Gillette. 1968. Effect of drug substrates on the reduction of hepatic microsomal cytochrome P-450 by NADPH. *Biochemical and Biophysical Research Communications*. 31:558-562.

Gillam, E.M., L.M. Notley, H. Cai, J.J. De Voss, and F.P. Guengerich. 2000. Oxidation of indole by cytochrome P450 enzymes. *Biochemistry*. 39:13817-24.

Gilson, M.K., and B. Honig. 1988. Calculation of the total electrostatic energy of a macromolecular system: solvation energies, binding energies, and conformational analysis. *Proteins*. 4:7-18.

Gilson, M.K., K.A. Sharp, and B.H. Honig. 1988. Calculating the electrostatic potential of molecules in solution: Method and error assessment. *Journal of Computational Chemistry*. 9:327-335.

Giraudo, M., G.C. Unnithan, G. Le Goff, and R. Feyereisen. Regulation of cytochrome P450 expression in Drosophila: Genomic insights. *Pesticide Biochemistry and Physiology*. 97:115-122.

Gohlke, H., M. Hendlich, and G. Klebe. 2000. Knowledge-based scoring function to predict protein-ligand interactions. *J Mol Biol*. 295:337-56.

Gong, M.Q., Y. Gu, X.B. Hu, Y. Sun, L. Ma, X.L. Li, L.X. Sun, J. Sun, J. Qian, and C.L. Zhu. 2005. Cloning and overexpression of CYP6F1, a cytochrome P450 gene, from deltamethrin-resistant Culex pipiens pallens. *Acta Biochim Biophys Sin (Shanghai)*. 37:317-26.

Goodsell, D.S., H. Lauble, C.D. Stout, and A.J. Olson. 1993. Automated docking in crystallography: analysis of the substrates of aconitase. *Proteins*. 17:1-10.

Gotoh, O. 1992. Substrate recognition sites in cytochrome P450 family 2 (CYP2) proteins inferred from comparative analyses of amino acid and coding nucleotide sequences. *J Biol Chem*. 267:83-90.

Graham, S.E., and J.A. Peterson. 1999. How Similar Are P450s and What Can Their Differences Teach Us? *Archives of Biochemistry and Biophysics*. 369:24-29.

Gray, P., A. Dunne, C. Brikos, C.A. Jefferies, S.L. Doyle, and L.A. O'Neill. 2006. MyD88 adapter-like (Mal) is phosphorylated by Bruton's tyrosine kinase during TLR2 and TLR4 signal transduction. *J Biol Chem*. 281:10489-95.

Gubler, D.J. 2006. Dengue/dengue haemorrhagic fever: history and current status. *Novartis Found Symp*. 277:3-16; discussion 16-22, 71-3, 251-3.

Guerrucci, M.A., and R. Belle. 1995. Characterisation of protein structure/function relationship by sequence analysis without previous alignment: distinction between sub-groups of protein kinases. *Biosci Rep*. 15:161-71.

Halbwirth, H., and K. Stich. 2006. An NADPH and FAD dependent enzyme catalyzes hydroxylation of flavonoids in position 8. *Phytochemistry*. 67:1080-7.

Hannemann, F., A. Bichet, K.M. Ewen, and R. Bernhardt. 2007. Cytochrome P450 systems--biological variations of electron transport chains. *Biochim Biophys Acta*. 1770:330-44.

Hayward, S., and H.J. Berendsen. 1998. Systematic analysis of domain motions in proteins from conformational change: new results on citrate synthase and T4 lysozyme. *Proteins*. 30:144-54.

Heringa, J. 2000. Computational methods for protein secondary structure prediction using multiple sequence alignments. *Curr Protein Pept Sci*. 1:273-301.

Herrmann, T., P. Güntert, and K. Wüthrich. 2002. Protein NMR Structure Determination with Automated NOE Assignment Using the New Software CANDID and the Torsion Angle Dynamics Algorithm DYANA. *Journal of Molecular Biology*. 319:209-227.

Hiroya, K., Y. Murakami, T. Shimizu, M. Hatano, and P.R. Ortiz de Montellano. 1994. Differential roles of Glu318 and Thr319 in cytochrome P450 1A2 catalysis supported by NADPH-cytochrome P450 reductase and tert-butyl hydroperoxide. *Arch Biochem Biophys*. 310:397-401.

Honig, B., and A. Nicholls. 1995. Classical electrostatics in biology and chemistry. *Science*. 268:1144-9.

Ingwall, J.S. 1982. Phosphorus nuclear magnetic resonance spectroscopy of cardiac and skeletal muscles. *Am J Physiol*. 242:H729-44.

Irwin, J.J., and B.K. Shoichet. 2004. ZINC - A Free Database of Commercially Available Compounds for Virtual Screening. *Journal of Chemical Information and Modeling*. 45:177-182.

Ishigooka, M., T. Shimizu, K. Hiroya, and M. Hatano. 1992. Role of Glu318 at the putative distal site in the catalytic function of cytochrome P450d. *Biochemistry*. 31:1528-31.

Jeschke, P. 2004. The unique role of fluorine in the design of active ingredients for modern crop protection. *Chembiochem*. 5:571-89.

Jeschke, P. 2010. The unique role of halogen substituents in the design of modern agrochemicals. *Pest Manag Sci*. 66:10-27.

Jones, G., P. Willett, R.C. Glen, A.R. Leach, and R. Taylor. 1997. Development and validation of a genetic algorithm for flexible docking. *J Mol Biol*. 267:727-48.

Kasai, S., and J.G. Scott. 2001. Expression and regulation of CYP6D3 in the house fly, Musca domestica (L.). *Insect Biochem Mol Biol*. 32:1-8.

Kasai, S., T. Shono, and M. Yamakawa. 1998. Molecular cloning and nucleotide sequence of a cytochrome P450 cDNA from a pyrethroid-resistant mosquito, Culex quinquefasciatus say. *Insect Mol Biol*. 7:185-90.

Katchalski-Katzir, E., I. Shariv, M. Eisenstein, A.A. Friesem, C. Aflalo, and I.A. Vakser. 1992. Molecular surface recognition: determination of geometric fit between proteins and their ligands by correlation techniques. *Proc Natl Acad Sci U S A*. 89:2195-9.

Kay, B., and S.N. Vu. 2005. New strategy against Aedes aegypti in Vietnam. *Lancet*. 365:613-7.

Kemp, C.A., J.U. Flanagan, A.J. van Eldik, J.D. Marechal, C.R. Wolf, G.C. Roberts, M.J. Paine, and M.J. Sutcliffe. 2004. Validation of model of cytochrome P450 2D6: an in silico tool for predicting metabolism and inhibition. *J Med Chem*. 47:5340-6.

Khan, K.K., Y.Q. He, T.L. Domanski, and J.R. Halpert. 2002. Midazolam oxidation by cytochrome P450 3A4 and active-site mutants: an evaluation of multiple binding sites and of the metabolic pathway that leads to enzyme inactivation. *Mol Pharmacol*. 61:495-506.

Khor, C.C., S.J. Chapman, F.O. Vannberg, A. Dunne, C. Murphy, E.Y. Ling, A.J. Frodsham, A.J. Walley, O. Kyrieleis, A. Khan, C. Aucan, S. Segal, C.E. Moore, K. Knox, S.J. Campbell, C. Lienhardt, A. Scott, P. Aaby, O.Y. Sow, R.T. Grignani, J. Sillah, G. Sirugo, N. Peshu, T.N. Williams, K. Maitland, R.J.O. Davies, D.P. Kwiatkowski, N.P. Day, D. Yala, D.W. Crook, K. Marsh, J.A. Berkley, L.A.J. O'Neill, and A.V.S. Hill. 2007. A Mal functional variant is associated with protection against invasive pneumococcal disease, bacteremia, malaria and tuberculosis. *Nat Genet*. 39:523-528.

Kim, D., and F.P. Guengerich. 2004. Selection of human cytochrome P450 1A2 mutants with enhanced catalytic activity for heterocyclic amine N-hydroxylation. *Biochemistry*. 43:981-8.

Kirton, S.B., C.A. Kemp, N.P. Tomkinson, S. St.-Gallay, and M.J. Sutcliffe. 2002. Impact of incorporating the 2C5 crystal structure into comparative models of cytochrome P450 2D6. *Proteins: Structure, Function, and Bioinformatics*. 49:216-231.

Kitamura, S., Y. Shimizu, Y. Shiraga, M. Yoshida, K. Sugihara, and S. Ohta. 2002. Reductive metabolism of p,p'-DDT and o,p'-DDT by rat liver cytochrome P450. *Drug Metab Dispos*. 30:113-8.

Kitchen, D.B., H. Decornez, J.R. Furr, and J. Bajorath. 2004. Docking and scoring in virtual screening for drug discovery: methods and applications. *Nat Rev Drug Discov*. 3:935-49.

Kneller, D.G., F.E. Cohen, and R. Langridge. 1990. Improvements in protein secondary structure prediction by an enhanced neural network. *J Mol Biol*. 214:171-82.

Kobayashi, T., G. Takaesu, and A. Yoshimura. 2006. Mal-function of TLRs by SOCS. *Nat Immunol*. 7:123-4.

Kobayashi, Y., X. Fang, G.D. Szklarz, and J.R. Halpert. 1998. Probing the active site of cytochrome P450 2B1: metabolism of 7-alkoxycoumarins by the wild type and five site-directed mutants. *Biochemistry*. 37:6679-88.

Kopp, J., and T. Schwede. 2004. Automated protein structure homology modeling: a progress report. *Pharmacogenomics*. 5:405-16.

Korytko, P.J., F.W. Quimby, and J.G. Scott. 2000. Metabolism of phenanthrene by house fly CYP6D1 and dog liver cytochrome P450. *J Biochem Mol Toxicol*. 14:20-5.

Kramer, M.A., and T.S. Tracy. 2008. Studying cytochrome P450 kinetics in drug metabolism. *Expert Opinion on Drug Metabolism & Toxicology*. 4:591-603.

Kuntz, I.D., K. Chen, K.A. Sharp, and P.A. Kollman. 1999. The maximal affinity of ligands. *Proceedings of the National Academy of Sciences of the United States of America*. 96:9997-10002.

Ladbury, J.E. 1996. Just add water! The effect of water on the specificity of protein-ligand binding sites and its potential application to drug design. *Chem Biol*. 3:973-80.

Laskowski, R.A., M.W. MacArthur, D.S. Moss, and J.M. Thornton. 1993. PROCHECK: a program to check the stereochemical quality of protein structures. *Journal of Applied Crystallography*. 26:283-291.

Leach, A.R. 1994. Ligand docking to proteins with discrete side-chain flexibility. *J Mol Biol*. 235:345-56.

Lesk, A.M., and C. Chothia. 1980. How different amino acid sequences determine similar protein structures: the structure and evolutionary dynamics of the globins. *J Mol Biol*. 136:225-70.

Lewis, D.F. 2000. Structural characteristics of human P450s involved in drug metabolism: QSARs and lipophilicity profiles. *Toxicology*. 144:197-203.

Lewis, D.F., M.N. Jacobs, and M. Dickins. 2004. Compound lipophilicity for substrate binding to human P450s in drug metabolism. *Drug Discov Today*. 9:530-7.

Lewis, D.F., and J.M. Pratt. 1998. The P450 catalytic cycle and oxygenation mechanism. *Drug Metab Rev*. 30:739-86.

Lewis, D.F., E. Watson, and B.G. Lake. 1998. Evolution of the cytochrome P450 superfamily: sequence alignments and pharmacogenetics. *Mutat Res*. 410:245-70.

Li, W., M.A. Schuler, and M.R. Berenbaum. 2003. Diversification of furanocoumarin-metabolizing cytochrome P450 monooxygenases in two papilionids: Specificity and substrate encounter rate. *Proc Natl Acad Sci U S A*. 100 Suppl 2:14593-8.

Li, X., J. Baudry, M.R. Berenbaum, and M.A. Schuler. 2004. Structural and functional divergence of insect CYP6B proteins: From specialist to generalist cytochrome P450. *Proc Natl Acad Sci U S A*. 101:2939-44.

Lin, J.H., A.L. Perryman, J.R. Schames, and J.A. McCammon. 2002. Computational drug design accommodating receptor flexibility: the relaxed complex scheme. *J Am Chem Soc*. 124:5632-3.

Liu, J., S.S. Ericksen, M. Sivaneri, D. Besspiata, C.W. Fisher, and G.D. Szklarz. 2004. The effect of reciprocal active site mutations in human cytochromes P450 1A1 and 1A2 on alkoxyresorufin metabolism. *Arch Biochem Biophys*. 424:33-43.

Lloyd, L.S. 2003. Best practices for dengue prevention and control in the Americas. US Agency for International Development, Washington, DC. . http://www.ehproject.org/PDF/Strategic_papers/SR7-BestPractice.pdf

Locuson, C.W., J.M. Hutzler, and T.S. Tracy. 2007. Visible Spectra of Type II Cytochrome P450-Drug Complexes: Evidence that "Incomplete" Heme Coordination Is Common. *Drug Metabolism and Disposition*. 35:614-622.

Loiarro, M., C. Sette, G. Gallo, A. Ciacci, N. Fanto, D. Mastroianni, P. Carminati, and V. Ruggiero. 2005. Peptide-mediated interference of TIR domain dimerization in MyD88 inhibits interleukin-1-dependent activation of NF-{kappa}B. *J Biol Chem*. 280:15809-14.

Lomize, M.A., A.L. Lomize, I.D. Pogozheva, and H.I. Mosberg. 2006. OPM: orientations of proteins in membranes database. *Bioinformatics*. 22:623-5.

Loughran, P.A., L.J. Roman, R.T. Miller, and B.S. Masters. 2001. The kinetic and spectral characterization of the E. coli-expressed mammalian CYP4A7: cytochrome b5 effects vary with substrate. *Arch Biochem Biophys*. 385:311-21.

Lyne, P.D. 2002. Structure-based virtual screening: an overview. *Drug Discov Today*. 7:1047-55.

Ma, B., S. Kumar, C.J. Tsai, and R. Nussinov. 1999. Folding funnels and binding mechanisms. *Protein Eng*. 12:713-20.

MacArthur, M.W., and J.M. Thornton. 1991. Influence of proline residues on protein conformation. *J Mol Biol*. 218:397-412.

MacArthur, M.W., and J.M. Thornton. 1993. Conformational analysis of protein structures derived from NMR data. *Proteins*. 17:232-51.

Mandell, J.G., V.A. Roberts, M.E. Pique, V. Kotlovyi, J.C. Mitchell, E. Nelson, I. Tsigelny, and L.F. Ten Eyck. 2001. Protein docking using continuum electrostatics and geometric fit. *Protein Eng*. 14:105-13.

Mansell, A., R. Smith, S.L. Doyle, P. Gray, J.E. Fenner, P.J. Crack, S.E. Nicholson, D.J. Hilton, L.A. O'Neill, and P.J. Hertzog. 2006. Suppressor of cytokine signaling 1 negatively regulates Toll-like receptor signaling by mediating Mal degradation. *Nat Immunol*. 7:148-55.

Marechal, J.D., C.A. Kemp, G.C. Roberts, M.J. Paine, C.R. Wolf, and M.J. Sutcliffe. 2008. Insights into drug metabolism by cytochromes P450 from modelling studies of CYP2D6-drug interactions. *Br J Pharmacol*. 153 Suppl 1:S82-9.

Meller, J., and R. Elber. 2001. Linear programming optimization and a double statistical filter for protein threading protocols. *Proteins*. 45:241-61.

Meng, E.C., B.K. Shoichet, and I.D. Kuntz. 1992. Automated docking with grid-based energy evaluation. *Journal of Computational Chemistry*. 13:505-524.

Najera, J.A., and M. Zaim. 2002. Malaria Vector Control. Decision Making Criteria and Procedures for Judicious Use of Insecticides. Geneva: World Health Organization.
http://whqlibdoc.who.int/hq/2003/WHO_CDS_WHOPES_2002.5_Rev.1.pdf

Nickerson, D.P., C.F. Harford-Cross, S.R. Fulcher, and L.L. Wong. 1997. The catalytic activity of cytochrome P450cam towards styrene oxidation is increased by site-specific mutagenesis. *FEBS Lett*. 405:153-6.

Nordqvist, M., D.R. Thakker, K.P. Vyas, H. Yagi, W. Levin, D.E. Ryan, P.E. Thomas, A.H. Conney, and D.M. Jerina. 1981. Metabolism of chrysene and phenanthrene to bay-region diol epoxides by rat liver enzymes. *Mol Pharmacol*. 19:168-78.

Nunez Miguel, R., J. Wong, J.F. Westoll, H.J. Brooks, L.A. O'Neill, N.J. Gay, C.E. Bryant, and T.P. Monie. 2007. A dimer of the Toll-like receptor 4 cytoplasmic domain provides a specific scaffold for the recruitment of signalling adaptor proteins. *PLoS One*. 2:e788.

O'Reilly, A.O., B.P. Khambay, M.S. Williamson, L.M. Field, B.A. Wallace, and T.G. Davies. 2006. Modelling insecticide-binding sites in the voltage-gated sodium channel. *Biochem J*. 396:255-63.

Ohlson, S. 2008. Designing transient binding drugs: a new concept for drug discovery. *Drug Discov Today*. 13:433-9.

Osguthorpe, D.J. 2000. Ab initio protein folding. *Curr Opin Struct Biol*. 10:146-52.

Page, C.C., C.C. Moser, X. Chen, and P.L. Dutton. 1999. Natural engineering principles of electron tunnelling in biological oxidation-reduction. *Nature*. 402:47-52.

Petrek, M., P. Kosinova, J. Koca, and M. Otyepka. 2007. MOLE: a Voronoi diagram-based explorer of molecular channels, pores, and tunnels. *Structure*. 15:1357-63.

Pettit, F.K., E. Bare, A. Tsai, and J.U. Bowie. 2007. HotPatch: a statistical approach to finding biologically relevant features on protein surfaces. *J Mol Biol*. 369:863-79.

Phillips, J.C., R. Braun, W. Wang, J. Gumbart, E. Tajkhorshid, E. Villa, C. Chipot, R.D. Skeel, L. Kale, and K. Schulten. 2005. Scalable molecular dynamics with NAMD. *J Comput Chem*. 26:1781-802.

Piao, W., C. Song, H. Chen, L.M. Wahl, K.A. Fitzgerald, L.A. O'Neill, and A.E. Medvedev. 2008. Tyrosine phosphorylation of MyD88 adapter-like (Mal) is critical for signal transduction and blocked in endotoxin tolerance. *J Biol Chem*. 283:3109-19.

Pikuleva, I.A. 2006. Cholesterol-metabolizing cytochromes P450. *Drug Metab Dispos*. 34:513-20.

Pikuleva, I.A., A. Puchkaev, and I. Bjorkhem. 2001. Putative helix F contributes to regioselectivity of hydroxylation in mitochondrial cytochrome P450 27A1. *Biochemistry*. 40:7621-9.

Poirot, O., K. Suhre, C. Abergel, E. O'Toole, and C. Notredame. 2004. 3DCoffee@igs: a web server for combining sequences and structures into a multiple sequence alignment. *Nucleic Acids Res*. 32:W37-40.

Poulos, T.L., B.C. Finzel, I.C. Gunsalus, G.C. Wagner, and J. Kraut. 1985. The 2.6-A crystal structure of Pseudomonas putida cytochrome P-450. *J Biol Chem*. 260:16122-30.

Rarey, M., B. Kramer, T. Lengauer, and G. Klebe. 1996. A Fast Flexible Docking Method using an Incremental Construction Algorithm. *Journal of Molecular Biology*. 261:470-489.

Ravichandran, K.G., S.S. Boddupalli, C.A. Hasermann, J.A. Peterson, and J. Deisenhofer. 1993. Crystal structure of hemoprotein domain of P450BM-3, a prototype for microsomal P450's. *Science*. 261:731-6.

Ray, D.E., and P.J. Forshaw. 2000. Pyrethroid insecticides: poisoning syndromes, synergies, and therapy. *J Toxicol Clin Toxicol*. 38:95-101.

Reichard, G.A. 2008. SARVision Plus. *J Chem Inf Model*. 48:1287-8.

Roberts, E.S., S.J. Pernecky, W.L. Alworth, and P.F. Hollenberg. 1996. A role for threonine 302 in the mechanism-based inactivation of P450 2B4 by 2-ethynylnaphthalene. *Arch Biochem Biophys*. 331:170-6.

Rocchia, W., S. Sridharan, A. Nicholls, E. Alexov, A. Chiabrera, and B. Honig. 2002. Rapid grid-based construction of the molecular surface and the use of induced surface charge to calculate reaction field energies: applications to the molecular systems and geometric objects. *J Comput Chem*. 23:128-37.

Rodpradit, P., S. Boonsuepsakul, T. Chareonviriyaphap, M.J. Bangs, and P. Rongnoparut. 2005. Cytochrome P450 genes: molecular cloning and overexpression in a pyrethroid-resistant strain of Anopheles minimus mosquito. *J Am Mosq Control Assoc*. 21:71-9.

Rodriguez, R., G. Chinea, N. Lopez, T. Pons, and G. Vriend. 1998. Homology modeling, model and software evaluation: three related resources. *Bioinformatics*. 14:523-8.

Rogerson, T.D., C.F. Wilkinson, and K. Hetarski. 1977. Steric factors in the inhibitory interaction of imidazoles with microsomal enzymes. *Biochem Pharmacol*. 26:1039-42.

Ruiz-Arguello, M.B., M.P. Veiga, A. Alonso, and F.M. Goni. 1999. Effect of Sublytic Concentrations of Sodium Cholate on Phospholipase C Hydrolysis of Phospholipid Bilayers. *J Colloid Interface Sci*. 219:163-167.

Rupasinghe, S., J. Baudry, and M.A. Schuler. 2003. Common active site architecture and binding strategy of four phenylpropanoid P450s from Arabidopsis thaliana as revealed by molecular modeling. *Protein Eng*. 16:721-31.

Rupasinghe, S.G., Z. Wen, T.L. Chiu, and M.A. Schuler. 2007. Helicoverpa zea CYP6B8 and CYP321A1: different molecular solutions to the problem of metabolizing plant toxins and insecticides. *Protein Eng Des Sel*. 20:615-24.

Ruzo, L.O., and J.E. Casida. 1977. Metabolism and toxicology of pyrethroids with dihalovinyl substituents. *Environ Health Perspect*. 21:285-92.

Saitou, N., and M. Nei. 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol Biol Evol*. 4:406-25.

Sali, A., and T.L. Blundell. 1993. Comparative protein modelling by satisfaction of spatial restraints. *J Mol Biol*. 234:779-815.

Sanchez, R., and A. Sali. 2000. Comparative protein structure modeling. Introduction and practical examples with modeller. *Methods Mol Biol*. 143:97-129.

Sander, C., and R. Schneider. 1991. Database of homology-derived protein structures and the structural meaning of sequence alignment. *Proteins*. 9:56-68.

Sasabe, M., Z. Wen, M.R. Berenbaum, and M.A. Schuler. 2004. Molecular analysis of CYP321A1, a novel cytochrome P450 involved in metabolism of plant allelochemicals (furanocoumarins) and insecticides (cypermethrin) in Helicoverpa zea. *Gene*. 338:163-75.

Schleinkofer, K., Sudarko, P.J. Winn, S.K. Ludemann, and R.C. Wade. 2005. Do mammalian cytochrome P450s show multiple ligand access pathways and ligand channelling? *EMBO Rep*. 6:584-9.

Scott, J.G., and G.P. Georghiou. 1986. Mechanisms responsible for high levels of permethrin resistance in the house fly. *Pesticide Science*. 17:195-206.

Scott, J.G., N. Liu, Z. Wen, F.F. Smith, S. Kasai, and C.E. Horak. 1999. House-fly cytochrome P450 CYP6D1: 5' flanking sequences and comparison of alleles. *Gene*. 226:347-353.

Segall, M.D., M.C. Payne, W. Ellis, G.T. Tucker, and N. Boyes. 1998. Evidence for stabilization of the low-spin state of cytochrome P450 due to shortening of the proximal heme bond. *Chem Res Toxicol*. 11:962-6.

Shaik, S., D. Kumar, S.P. de Visser, A. Altun, and W. Thiel. 2005. Theoretical perspective on the structure and mechanism of cytochrome P450 enzymes. *Chem Rev*. 105:2279-328.

Sharikov, Y., R.C. Walker, J. Greenberg, V. Kouznetsova, S.K. Nigam, M.A. Miller, E. Masliah, and I.F. Tsigelny. 2008. MAPAS: a tool for predicting membrane-contacting protein surfaces. *Nat Methods*. 5:119.

Shi, J., T.L. Blundell, and K. Mizuguchi. 2001. FUGUE: sequence-structure homology recognition using environment-specific substitution tables and structure-dependent gap penalties. *J Mol Biol*. 310:243-57.

Shimada, T., R.L. Mernaugh, and F.P. Guengerich. 2005. Interactions of mammalian cytochrome P450, NADPH-cytochrome P450 reductase, and cytochrome b(5) enzymes. *Arch Biochem Biophys*. 435:207-16.

Simard, F., E. Nchoutpouen, J.C. Toto, and D. Fontenille. 2005. Geographic distribution and breeding site preference of Aedes albopictus and Aedes aegypti (Diptera: culicidae) in Cameroon, Central Africa. *J Med Entomol*. 42:726-31.

Spatzenegger, M., and W. Jaeger. 1995. Clinical importance of hepatic cytochrome P450 in drug metabolism. *Drug Metab Rev*. 27:397-417.

Suhre, K., and Y.H. Sanejouand. 2004. ElNemo: a normal mode web server for protein movement analysis and the generation of templates for molecular replacement. *Nucleic Acids Res*. 32:W610-4.

Tang, J., Y. Cao, R.L. Rose, and E. Hodgson. 2002. In vitro metabolism of carbaryl by human cytochrome P450 and its inhibition by chlorpyrifos. *Chem Biol Interact*. 141:229-41.

Teodoro, M.L., and L.E. Kavraki. 2003. Conformational flexibility models for the receptor in structure based drug design. *Curr Pharm Des*. 9:1635-48.

Thompson, J.D., D.G. Higgins, and T.J. Gibson. 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res*. 22:4673-80.

Thompson, J.D., J.C. Thierry, and O. Poch. 2003. RASCAL: rapid scanning and correction of multiple sequence alignments. *Bioinformatics*. 19:1155-61.

Toshchakov, V.U., S. Basu, M.J. Fenton, and S.N. Vogel. 2005. Differential involvement of BB loops of toll-IL-1 resistance (TIR) domain-containing

adapter proteins in TLR4- versus TLR2-mediated signal transduction. *J Immunol*. 175:494-500.

Vakser, I.A. 1995. Protein docking for low-resolution structures. *Protein Eng*. 8:371-7.

Verdonk, M.L., G. Chessari, J.C. Cole, M.J. Hartshorn, C.W. Murray, J.W. Nissink, R.D. Taylor, and R. Taylor. 2005. Modeling water molecules in protein-ligand docking using GOLD. *J Med Chem*. 48:6504-15.

Verras, A., I.D. Kuntz, and P.R. Ortiz de Montellano. 2004. Computer-assisted design of selective imidazole inhibitors for cytochrome p450 enzymes. *J Med Chem*. 47:3572-9.

Wade, R.C., P.J. Winn, I. Schlichting, and Sudarko. 2004. A survey of active site access channels in cytochromes P450. *J Inorg Biochem*. 98:1175-82.

Wagner, G., S.G. Hyberts, and T.F. Havel. 1992. NMR structure determination in solution: a critique and comparison with X-ray crystallography. *Annu Rev Biophys Biomol Struct*. 21:167-98.

Weber, P.C., J.J. Wendoloski, M.W. Pantoliano, and F.R. Salemme. 1992. Crystallographic and thermodynamic comparison of natural and synthetic ligands bound to streptavidin. *Journal of the American Chemical Society*. 114:3197-3200.

Werck-Reichhart, D., and R. Feyereisen. 2000. Cytochromes P450: a success story. *Genome Biol*. 1:REVIEWS3003.

Wheelock, G.D., and J.G. Scott. 1992. Anti-P450lpr antiserum inhibits specific monooxygenase activities in LPR house fly microsomes. *J Exp Zool*. 264:153-8.

WHO. WHO Pesticides Evaluation Scheme: "WHOPES" http://www.who.int/whopes/en/

Wilder-Smith, A., and E. Schwartz. 2005. Dengue in travelers. *N Engl J Med*. 353:924-32.

Williams, P.A., J. Cosme, V. Sridhar, E.F. Johnson, and D.E. McRee. 2000a. Mammalian microsomal cytochrome P450 monooxygenase: structural adaptations for membrane binding and functional diversity. *Mol Cell*. 5:121-31.

Williams, P.A., J. Cosme, V. Sridhar, E.F. Johnson, and D.E. McRee. 2000b. Microsomal cytochrome P450 2C5: comparison to microbial P450s and unique features. *J Inorg Biochem*. 81:183-90.

Williams, P.A., J. Cosme, A. Ward, H.C. Angove, D. Matak Vinkovic, and H. Jhoti. 2003. Crystal structure of human cytochrome P450 2C9 with bound warfarin. *Nature*. 424:464-8.

Winn, P.J., S.K. Ludemann, R. Gauges, V. Lounnas, and R.C. Wade. 2002. Comparison of the dynamics of substrate access channels in three cytochrome P450s reveals different opening mechanisms and a novel functional role for a buried arginine. *Proc Natl Acad Sci U S A*. 99:5361-6.

Wlodawer, A., W. Minor, Z. Dauter, and M. Jaskolski. 2008. Protein crystallography for non-crystallographers, or how to get the best (but not more) from published macromolecular structures. *FEBS J*. 275:1-21.

Wolber, G., and T. Langer. 2005. LigandScout: 3-D pharmacophores derived from protein-bound ligands and their use as virtual screening filters. *J Chem Inf Model*. 45:160-9.

Wright, A.T., and B.F. Cravatt. 2007. Chemical Proteomic Probes for Profiling Cytochrome P450 Activities and Drug Interactions In Vivo. 14:1043-1051.

Yamaguchi, Y., K.K. Khan, Y.A. He, Y.Q. He, and J.R. Halpert. 2004. Topological changes in the CYP3A4 active site probed with phenyldiazene: effect of interaction with NADPH-cytochrome P450 reductase and cytochrome b5 and of site-directed mutagenesis. *Drug Metab Dispos*. 32:155-61.

Yang, E., Y. Yang, S. Wu, and Y. Wu. 2005. Relative contribution of detoxifying enzymes to pyrethroid resistance in a resistant strain of Helicoverpa armigera. *Journal of Applied Entomology*. 129:521-525.

Yano, J.K., M.R. Wester, G.A. Schoch, K.J. Griffin, C.D. Stout, and E.F. Johnson. 2004. The structure of human microsomal cytochrome P450 3A4 determined by X-ray crystallography to 2.05-A resolution. *J Biol Chem*. 279:38091-4.

Yu, H. 1999. Extending the size limit of protein nuclear magnetic resonance. *Proceedings of the National Academy of Sciences of the United States of America*. 96:332-334.

Zhang, L., and J. Hermans. 1996. Hydrophilicity of cavities in proteins. *Proteins*. 24:433-8.

Zhang, M., and J.G. Scott. 1996. Cytochrome b5 Is Essential for Cytochrome P450 6D1-Mediated Cypermethrin Resistance in LPR House Flies. *Pestic Biochem Physiol*. 55:150-6.

Zhu, F., J.N. Feng, L. Zhang, and N. Liu. 2008. Characterization of two novel cytochrome P450 genes in insecticide-resistant house-flies. *Insect Mol Biol*. 17:27-37.

Appendix

Appendix A

Python script for Modeller V 8.0 to build the models.

```
# Homology modelling by the automodel class

from modeller.automodel import *    # Load the automodel class

log.verbose()    # request verbose output
env = environ()  # create a new MODELLER environment to build this
model in

# directories for input atom files
env.io.atom_files_directory = '/home/momlejb2/bin/modeller8v0/'
env.io.hetatm = True
env.libs.topology.read(file='/home/momlejb2/bin/modeller8v0/modlib/top
_heav.lib')
env.libs.parameters.read(file='/home/momlejb2/bin/modeller8v0/modlib/p
ar.lib')



class mymodel(automodel):

   def special_patches(self,aln):
      # a patch for FE2S  (position of C, position of heme)
      self.patch(residue_type='FE2S', residue_ids=('441', '499'))



a = mymodel(env,
              alnfile  = '/home/momlejb2/bin/modeller8v0/target and
template.ali',    # alignment filename
              knowns   = ( '1TQN', ), # codes of the templates
              sequence = 'target'
           )               # code of the target
a.starting_model= 1                   # index of the first model
a.ending_model  = 30                   # index of the last model
                                       # (determines how many models to
calculate)
a.make()                              # do the actual homology modelling
```

Appendix B

The .tcl script to find water bridges.

```
##########################################################
#Finds Water Bridges between two selections
#############################################

#####################################################
#Select the sections of protein to find bridges between


#Set the solvent
set closewater [atomselect top "water"]


#First selection #NOT CARBON- will find any bond to any atom with a H
set resnameN27 [atomselect top "sidechain residue 2 and not carbon"]

#Second selection
set resid97 [atomselect top "residue 35 to 50 and not carbon"]
#output files generated

#debugging
set file1 [open "close8.dat" w]
set file2 [open "close9.dat" w]

#output files
set file3 [open "PROD-10_AD.dat" w]
set file4 [open "PROD-10_DD.dat" w]
set file5 [open "PROD-10_DA.dat" w]
set file6 [open "PROD-10_AA.dat" w]

###########################################################
#Read in each frame of the trajectory in top
set n [molinfo top get numframes]
for { set i 0} {$i < $n} {incr i} {

#update selection in each frame
$closewater frame $i
$closewater update
$resid97 frame $i
$resid97 update
$resnameN27 frame $i
$resnameN27 update

#find hbond between pY and surrounding water pY is the acceptor
set nhb [measure hbonds 3.5 60 $closewater $resnameN27]

#find hbond between surounding water and SER- Ser is donor
set SERdon [measure hbonds 3.5 60 $resid97 $closewater ]

#find hbond between surounding water and SER, ser is the aceptor
set SERacc [measure hbonds 3.5 60 $closewater $resid97  ]

#find H-bond with Y as a donor
set Ydon [measure hbonds 3.5 60  $resnameN27 $closewater  ]
puts $file1 "$Ydon || $nhb  || $SERdon"

#hbonds creates 3 lists:
#0 ontains the indices of the donors
#1 contains the indices of the acceptors
#2 contains the index of the hydrogen atom in the hydrogen bond
```

```
####################################################

#find the following bridge:
#
# pY O--HO--HO SER
#         H
#find a common water
set my_AD 0
set my_DD 0
set my_DA 0
set my_AA 0

foreach donor [lindex $nhb 0] {
#read through water oxygen id acting as a donor for pY
        foreach acceptor [lindex $SERdon 1] {
        #read through each water id acting as an acceptor for SER

        #set my_AD 0

                if  {$donor == $acceptor} {;

                    #puts $file3 "$1"

                    set my_AD 1
                    puts $file2 "O-HO-HO $i $donor"
                    #puts $file2 "$i $my_AD"
                }

        }
}

#find the following bridge:
#
#pY O--HOH--O SER
#
#
foreach donor_two [lindex $nhb 0] {
        foreach acceptor_two [lindex $SERacc 0] {

        #set my_DD 0

                if  {$donor_two == $acceptor_two} {;
                    #puts $file3 "2"


                    set my_DD 1
                    puts $file2 "O-OHO-O $i $donor_two"
                }
        #puts "my_DD $my_DD"
        }
}

#find the following bridge:
#Y-donor  SER-acceptor

#pY OH--OH--O SER
#      H
#
#for each water acceptor
foreach donor_three [lindex $Ydon 1] {
        #Ser donor
```

315

```
        foreach acceptor_three [lindex $SERacc 0] {

        #set my_DD 0

                if {$donor_three == $acceptor_three} {;
                    #puts $file3 "2"


                    set my_DA 1
                    puts $file2 "DA YOH-OH-O $i $donor_three"
                }
        #puts "my_DD $my_DD"
        }
}


#find the following bridge:
#Y-donor Ser-donor
#        H
#pY OH--O--HO SER
#        H

#for each water acceptor
foreach donor_four [lindex $Ydon 1] {
        #for each ser water acceptor
        foreach acceptor_four [lindex $SERdon 1] {

        #set my_DD 0

                if {$donor_four == $acceptor_four} {;
                    #puts $file3 "2"


                    set my_AA 1
                    puts $file2 "YOH-OH-O $i $donor_four"
                }
        #puts "my_DD $my_DD"
        }
}


#do some stats
set ADoccupancy [expr 0 + $my_AD]
set DDoccupancy [expr 0 + $my_DD]

#output the bonds found in current frame
puts "$i AD $my_AD DD $my_DD DA $my_DA AA $my_AA"
puts $file3 "$i $my_AD"
puts $file4 "$i $my_DD"
puts $file5 "$i $my_DA"
puts $file6 "$i $my_AA"

}
close $file1
close $file2
close $file3
close $file4
close $file5
close $file6

set final_occ [expr $ADoccupancy/$i]
puts "$i occupancy is $ADoccupancy $final_occ %"
```