

The Study of *Mycobacterium tuberculosis* Cytochrome P450 Enzymes by Experimental and Theoretical Methods

A thesis submitted to the University of Manchester for the degree of
Doctor of Philosophy (PhD)
In the Faculty of Engineering and Physical Sciences

2008

BY
PHILIP R. BALDING

School of Chemical Engineering and Analytical Science

ProQuest Number: 10997050

All rights reserved

INFORMATION TO ALL USERS

The quality of this reproduction is dependent upon the quality of the copy submitted.

In the unlikely event that the author did not send a complete manuscript and there are missing pages, these will be noted. Also, if material had to be removed, a note will indicate the deletion.



ProQuest 10997050

Published by ProQuest LLC (2018). Copyright of the Dissertation is held by the Author.

All rights reserved.

This work is protected against unauthorized copying under Title 17, United States Code
Microform Edition © ProQuest LLC.

ProQuest LLC.
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106 – 1346

(ET7NH)

~~X~~

Th 30953

THE
JOHN RYLANDS
UNIVERSITY
LIBRARY

CONTENTS

ABBREVIATIONS	7
LIST OF FIGURES	10
LIST OF TABLES	15
ABSTRACT	18
DECLARATION	19
COPYRIGHT STATEMENT	20
ACKNOWLEDGEMENTS	21
CHAPTER 1: INTRODUCTION.....	22
1.1 Haem and Haemoproteins.....	22
1.1.1 Haem structures in nature	22
1.1.2 Biosynthesis of haem	25
1.1.3 Haemoproteins.....	27
1.2 Cytochromes P450	31
1.3 Tuberculosis – A ‘Global Emergency’	38
1.3.1 <i>Mtb</i> P450s, novel targets in the fight against TB	40
1.4 Aim	41
CHAPTER 2: MATERIALS AND METHODS.....	44
2.1 Experimental Materials	44
2.1.1 Liquid media.....	44
2.1.2 Antibiotics.....	44
2.1.2.1 Ampicillin	44
2.1.2.2 Carbenicillin	44
2.1.2.3 Rifampicin.....	45
2.1.2.4 Chloramphenicol.....	45
2.1.2.5 Tetracycline	45
2.1.2.6 Kanamycin.....	45
2.1.3 LB agar plates.....	45
2.1.4 Buffers	45
2.1.4.1 Tris and Tris/EDTA buffers	45
2.1.4.2 Potassium phosphate (KPi) buffers.....	46
2.1.4.3 MOPS buffer.....	46

2.1.5 Isopropyl- β -D-thiogalactopyranoside (IPTG)	46
2.2 Experimental Methods	46
2.2.1 Molecular biology	46
2.2.1.1 Cloning of Rv1394c into pET11a and pET15b	46
2.2.1.2 Cloning of Rv1394c into pET42a	47
2.2.2 Transformation of DNA into <i>E. coli</i> cells	47
2.2.3 Mini-prep DNA	48
2.2.4 Midi-prep DNA	48
2.2.5 0.8 % Agarose gels	48
2.2.6 Verifying plasmid transformation	49
2.2.7 Expression of CYP132	50
2.2.8 Plasmid stability test.....	50
2.2.9 SDS PAGE gel.....	51
2.2.10 Determination of amount of soluble protein by SDS PAGE analysis.....	52
2.2.11 Purification of CYP132	53
2.2.11.1 Protein extraction.....	53
2.2.11.2 Ammonium sulphate purification	53
2.2.11.3 Column purification.....	53
2.2.12 Expression of <i>E. coli</i> flavodoxin reductase	54
2.2.13 Purification of <i>E. coli</i> flavodoxin reductase	55
2.2.14 UV/Vis spectrometry	55
2.2.15 Determination of extinction coefficient of CYP132 using pyridine haemochromogen method	55
2.2.16 Mass spectrometry of CYP132.....	56
2.2.17 Binding titrations.....	56
2.2.18 Electron Paramagnetic Resonance (EPR) spectroscopy	56
2.2.19 Circular Dichroism (CD)	57
2.2.20 Turnover kinetics.....	57
2.2.21 Mass spectrometry of turnover products	57
2.3 Theoretical Methods	58
2.3.1 Quantum mechanics methods.....	58
2.3.1.1 Hartree-Fock (HF)	59
2.3.1.2 Basis functions	60
2.3.1.3 Electron correlation.....	61

2.3.2 Density Functional Theory.....	62
2.3.2.1 Open and closed shell systems	63
2.3.2.2 Basis Set Superposition Error (BSSE)	63
2.3.3 Potential Energy Surfaces.....	63
2.3.4 Theoretical methods used in this study	66
2.3.4.1 DFT methods.....	66
2.3.4.2 Basis sets	66
2.3.4.3 The haem model.....	67
CHAPTER 3: THE EXPRESSION AND CHARACTERISATION OF <i>MYCOBACTERIUM</i> <i>TUBERCULOSIS</i> CYP132	68
3.1 Introduction.....	68
3.1.1 <i>Mycobacterium tuberculosis</i>	68
3.1.2 <i>Mtb</i> P450s.....	68
3.1.3 <i>Mtb</i> CYP132, a putative fatty acid hydroxylase	70
3.1.4 <i>Aim</i>	71
3.2 Results and Discussion	72
3.2.1 Expression and purification of CYP132	72
3.2.1.1 Gene expression over time	73
3.2.1.2 Effect of growth medium on CYP132 production.....	74
3.2.1.3 Attempted purification of CYP132.....	75
3.2.2 Expression and purification of Polyhistidine-tagged CYP132	78
3.2.2.1 Expression of polyhistidine-tagged CYP132	78
3.2.2.2 Plasmid stability test	79
3.2.2.3 Purification of polyhistidine tagged CYP132	80
3.2.2.4 Investigation of other column resins and buffer conditions for CYP132 isolation	83
3.2.3 Characterisation of CYP132 expression from the <i>Rv1394c</i> /pET15b Plasmid	86
3.2.3.1 Preliminary analysis of the HMS174(DE3) <i>Rv1394c</i> /pET15b expression products	86
3.2.4 Analysis of cleaved CYP132	95
3.2.4.1 Binding of CO and NO	95
3.2.4.2 EPR spectroscopy of cleaved CYP132.....	97
3.2.4.3 Analysis of cleaved CYP132 by circular dichroism.....	99
3.2.4.4 Attempted removal of bound substrate with solvent.....	100

3.2.4.5 Attempted analysis of bound substrate by mass spectrometry.....	104
3.2.4.6 Binding of imidazole.....	104
3.2.4.7 The binding of azole anti-fungal drugs	107
3.2.4.8 Titrations with putative fatty acid substrates.....	109
3.2.4.9 Turnover kinetics of cleaved CYP132 with fatty acids.....	111
3.2.4.10 Analysis of the products of the turnover studies by mass spectrometry	116
3.2.5 Expression and purification of GST tagged CYP132	122
3.2.5.1 Cloning of <i>Rv1394c</i> into the pET42a vector	122
3.2.5.2 Expression of GST tagged CYP132.....	124
3.3 Conclusion.....	129
3.3.1 Future work	131
CHAPTER 4: INVESTIGATION OF THE P450 REACTION CYCLE BY DFT METHODS	133
4.1 Introduction.....	133
4.1.1 The P450 catalytic cycle.....	134
4.1.2 Aim.....	140
4.2 Results and Discussion	141
4.2.1 Water-bound resting state	141
4.2.2 Penta-coordinated ferric porphyrin species	151
4.2.3 Penta-coordinated ferrous porphyrin species	156
4.2.4 Dioxygen-bound ferrous porphyrin	159
4.2.5 Dioxygen-bound reduced ferrous porphyrin.....	163
4.2.6 Hydroperoxo ferric porphyrin – Compound 0.....	166
4.2.7 Compound I	171
4.3 Summary	176
4.4 Conclusion.....	179
4.4.1 Future Work	182
CHAPTER 5: ANALYSIS OF AZOLE BINDING TO P450S BY DFT METHODS.....	184
5.1 Introduction.....	184
5.1.1 P450 inhibitors	186
5.1.1.1 Irreversible and quasi irreversible inhibitors	186
5.1.1.2 Reversible inhibitors	187
5.1.1.3 Azole drugs	189

5.1.1.4 Aim	191
5.1.1.5 Previous studies on azoles binding to P450s and haems	192
5.1.1.6 The model used.....	193
5.2 Results and Discussion	194
5.2.1 1D binding curves	194
5.2.2 2D Potential Energy Surfaces	199
5.2.3 Transition states.....	201
5.2.4 Two minima.....	204
5.2.5 Comparing the affinity of the P450 haem for water and azoles.....	207
5.2.6 Varying basis set.....	210
5.2.7 Varying cysteine model	211
5.2.8 Rotation energy.....	212
5.2.9 Counterpoise Correction calculations.....	213
5.2.9.1 Basis Set Superposition Error.....	213
5.2.9.2 Counterpoise Correction Calculations.....	214
5.2.10 Solvation models.....	218
5.2.10.1 Implicit solvent.....	218
5.2.11 Explicit solvent models	220
5.2.12 Dissociation calculations	224
5.2.13 Modelling CYP121 Serine-237	227
5.3 Conclusion.....	229
CHAPTER 6: CONCLUSIONS AND FUTURE PROSPECTS	232
6.1 Introduction.....	232
6.2 Conclusions.....	233
6.2.1 The expression and characterisation of <i>Mtb</i> CYP132.....	233
6.2.2 Investigation of the P450 reaction cycle by DFT methods	236
6.2.3 Analysis of azole binding in P450s by DFT methods	238
6.3 Future work	239
6.3.1 The expression and characterisation of <i>Mtb</i> CYP132.....	239
6.3.2 The P450 reaction cycle.....	241
6.3.3 Azole binding to P450s.....	242
6.4 Concluding remarks.....	243
APPENDIX: Paper Submitted to J. Phys. Chem. For Publication	245

Word Count: 60200

ABBREVIATIONS

1D	1 Dimensional
2D	2 Dimensional
3D	3 Dimensional
ALAS	δ -ALA Synthase
B3LYP	Becke Exchange, Lee, Yang Parr Correlation Hybrid DFT/HF Functional
BCG	<i>Bacille</i> Calmette-Guérin
BLYP	Becke Exchange, Lee Yang Parr Correlation Functional
bp	Base Pairs
BP86	Becke Exchange, Perdew Correlation Functional
BS1	Basis set 1 - LANL2DZ, with pseudo potentials (Fe, S), 6-31G* (O, N), 6-31g (C, H)
BS2	Basis set 2 - LANL2DZ, with pseudo potentials (Fe, S), 6-31G** (O, N, C, H)
BS3	Basis set 3 - LANL2DZ, with pseudo potentials (Fe), 6-31G** (S, O, N, C, H)
BS4	Basis set 4 - LACVP, with pseudo potentials (Fe), 6-31G** (S, O, N, C, H)
BS5	Basis set 5 - 6-31G(d,f), no pseudo potentials (Fe, S, O, N, C, H)
BS6	Basis set 6 - 6-31G*, no pseudo potentials (Fe, S, O, N, C, H)
BS7	Basis set 7 - LANL2DZ, with pseudo potentials (Fe), 3-21G (S, O, N, C, H)
BSSE	Basis Set Superposition Error
CCP	Cytochrome <i>c</i> Peroxidase
CD	Circular Dichroism
CO	Carbon Monoxide
CPO	Coproporphyrinogen Oxidase
CPR	Cytochrome P450 Reductase
CYP	Cytochrome P450
DCM	Dichloromethane

ddH₂O	Distilled Deionised Water
DEAE-Sepharose	Diethylaminoethyl-Sepharose
DFT	Density Functional Theory
DMSO	Dimethyl Sulfoxide
ENDOR	Electron Nuclear Double Resonance
EPR	Electron Paramagnetic Resonance
Ether	Diethyl Ether
FAD	Flavin Adenine Dinucleotide
FLD	Flavodoxin
FLDR	Flavodoxin Reductase
FMN	Flavin Mononucleotide
GdmCl	Guanidinium Chloride
GST	Glutathione S-Transferase
HF	Hartree-Fock
HPLC	High Performance Liquid Chromatography
HRP	Horseradish Peroxidase
IPTG	Isopropyl-β-D-Thiogalactopyranoside
IRC	Intrinsic Reaction Coordinate
kb	Kilobase Pairs
K_d	Dissociation Constant
KIE	Kinetic Isotope Effect
KPi	Potassium Phosphate
LB	Luria-Bertani
LC	Liquid Chromatography
LCAO	Linear Combination of Atomic Orbitals
<i>m</i>-CPBA	<i>meta</i> -Chloroperoxybenzoic Acid
MD	Molecular Dynamics
MDR	Multidrug-Resistant
MIC	Minimal Inhibitory Concentration
mM	Milli Molar
MM	Molecular Mechanics
MOPS	3-Morpholinopropanesulphonic Acid
MS	Mass Spectroscopy
<i>Mtb</i>	<i>Mycobacterium tuberculosis</i>

NADPH	Nicotinamide Adenine Dinucleotide Phosphate
NMR	Nuclear Magnetic Resonance
NO	Nitric Oxide
P450	Cytochrome P450
PBD	Porphobilinogen Deaminase
PBG	Porphobilinogen
PCM	Polarisable Continuum Model
PCR	Polymerase Chain Reaction
PES	Potential Energy Surface
PMSF	Phenylmethanesulphonyl Fluoride
His-Tag	Polyhistidine Tag
PPO	Protoporphyrinogen Oxidase
QM	Quantum Mechanics
RHF	Restricted Hartree-Fock
SCF	Self Consistent Field
SDS-PAGE	Tris-Glycine SDS-Polyacrylamide Gel Electrophoresis
SRS	Substrate Recognition Site
TB	Tuberculosis (Disease)
TB	Terrific Broth
TEMED	N,N,N',N'-Tetramethylethylenediamine
Tris	Tris(Hydroxymethyl) Aminomethane Buffer
UHF	Unrestricted Hartree-Fock
UPD	Uroporphyrinogen Decarboxylase
UPS	Uroporphyrinogen III Synthase
WHO	World Health Organization
ZPE	Zero Point Energy
δ-ALA	δ -Aminolevulinic Acid
μM	Micro Molar

LIST OF FIGURES

Figure 1.1:	Structures of the haem cofactors found in nature	23
Figure 1.2:	Generic haem biosynthetic pathway.....	26
Figure 1.3:	Reaction scheme for the decomposition of hydrogen peroxide by cytochrome c peroxidase (CCP)	30
Figure 1.4:	X-ray diffraction crystal structure of cytochrome c peroxidase Compound I	30
Figure 1.5:	Crystal structure of <i>Pseudomonas putida</i> P450cam.....	34
Figure 1.6:	UV/Vis spectra of P450 Biol in the high and low spin states.....	35
Figure 1.7:	The consensus catalytic cycle for the activation of molecular oxygen by P450s and the hydroxylation of a substrate.....	36
Figure 1.8:	Schematic representation of the rebound mechanism (Groves and McClusky, 1976, Groves, 1985).....	37
Figure 2.1:	NEB comparison chart showing the pattern and range of sizes of DNA in the 1kb DNA ladder.....	49
Figure 2.2:	NEB comparison chart showing the pattern and masses in kDa of the proteins in the protein marker used in this study	51
Figure 2.3:	1D PES for a diatomic molecule, or larger system relative to a single geometric parameter.....	64
Figure 3.1:	0.8 % Agarose gels of the <i>Rv1394c</i> /pET11a plasmid.....	72
Figure 3.2:	SDS-PAGE gels showing the total production of protein with time from the <i>Rv1394c</i> /pET11a plasmid in <i>E. coli</i> strain HMS174(DE3).	74
Figure 3.3:	SDS-PAGE gels showing the total production of protein with time from the <i>Rv1394c</i> /pET11a plasmid transformed in HMS174(DE3), using LB, TB and YT media.	75
Figure 3.4:	UV/Vis spectrum of the cell extract from the expression of <i>Rv1394c</i> /pET11a in HMS174(DE3).....	77
Figure 3.5:	SDS-PAGE gels showing the expression with time for transformants of <i>Rv1394c</i> /pET15b plasmid in the HMS174 (DE3) cell strain.....	78
Figure 3.6:	UV/Vis spectra showing the relative purity of the CYP132 protein expressed from the <i>Rv1394c</i> /pET15b plasmid in HMS174(DE3) transformants.....	82

Figure 3.7:	SDS-PAGE gels of the CYP132 product of expression from the <i>Rv1394c</i> /pET15b plasmid in HMS174(DE3) cells	83
Figure 3.8:	UV/Vis spectra of the cell extract from the expression of CYP132 from the HMS174(DE3)/ <i>Rv1394c</i> /pET15b transformant	84
Figure 3.9:	UV/Vis spectra of the cell extract from the expression of CYP132 from the HMS174(DE3)/ <i>Rv1394c</i> /pET15b transformant.	85
Figure 3.10:	Electrospray time of flight mass spectroscopy of purified protein from the expression of CYP132 from the <i>Rv1394c</i> /pET15b plasmid	87
Figure 3.11:	Electrospray time of flight mass spectroscopy of the small amount of purified protein from the expression of CYP132 from the <i>Rv1394c</i> /pET15b plasmid.	88
Figure 3.12:	UV/Vis spectra showing binding of CO to the ferrous CYP132 products	90
Figure 3.13:	Protein alignment of polyhistidine tagged CYP132 and the P450 BM3 haem domain.	92
Figure 3.14:	Crystal structure of the P450 BM3 haem domain with bound palmitoleic acid.	93
Figure 3.15:	UV/Vis spectrum of cleaved CYP132	94
Figure 3.16:	Binding of CO to cleaved, ferrous CYP132 under anaerobic conditions.	96
Figure 3.17:	Binding of NO to cleaved CYP132	96
Figure 3.18:	EPR spectrum of cleaved CYP132	98
Figure 3.19:	Far UV CD spectra for cleaved CYP132	100
Figure 3.20:	UV/Vis spectra showing the effect of the addition of solvents to cleaved CYP132.	103
Figure 3.21:	UV/Vis spectra showing the change in Soret band absorbance from 416 nm to 424 nm with the addition of imidazole to cleaved CYP132.	105
Figure 3.22:	Graphs used in the calculation of the K_d value for the binding of imidazole to the cleaved form of CYP132	106
Figure 3.23:	Structures ofazole inhibitors titrated with CYP132	107
Figure 3.24:	Structures of fatty acids titrated with CYP132	109
Figure 3.25:	UV/Vis spectra of CYP132 and arachidonic acid	110
Figure 3.26:	UV/Vis spectra of CYP132 and SDS	111

Figure 3.27:	SDS-PAGE gel of the enzymes used in the investigation of the kinetics of turnover of cleaved CYP132 with fatty acids	112
Figure 3.28:	Kinetics studies of CYP132 and palmitoleic acid.....	115
Figure 3.29:	LC/MS spectra of the products of the turnover of myristoleic acid by cleaved CYP132	119
Figure 3.30:	LC/MS spectra of the products of the control reaction.....	120
Figure 3.31:	MS/MS analysis of the LC/MS peaks at 4 and 5 minutes from the products of the reaction of myristoleic acid with CYP132.	121
Figure 3.32:	LC/MS chromatogram of the products of the reaction of palmitoleic acid with cleaved CYP132.....	122
Figure 3.33:	Agarose gel showing diagnostic restriction digests of <i>Rv1394c</i> /pET42a and <i>Rv1394c</i> /Tr-pET42a plasmids.....	123
Figure 3.34:	SDS-PAGE gel analysis of the expression of CYP132 from the <i>Rv1394c</i> /pET42a plasmid in HMS174(DE3)	126
Figure 3.35:	UV/Vis spectra of progressive dilutions of the cell extract from the expression of <i>Rv1394c</i> /Tr-pET42a in HMS174(DE3)	127
Figure 3.36:	SDS-PAGE gel analysis of expression of proteins expressed from <i>Rv1394c</i> /Tr-pET42a transformants and purified on immobilised glutathione	127
Figure 3.37:	UV/Vis analysis of the expression product from HMS174(DE3) / <i>Rv1394c</i> /pET42a transformants after purification on immobilised glutathione	128
Figure 4.1:	Iron anti-bonding d orbitals, shown here for the doublet spin state of the water bound resting state.....	135
Figure 4.2:	The consensus catalytic cycle for the activation of molecular oxygen by P450s and the hydroxylation of a substrate.....	138
Figure 4.3:	Schematic representation of the rebound mechanism proposed by Groves <i>et al.</i>	139
Figure 4.4:	Structures and relative energies of doublet, quartet and sextet spin states of water-bound ferric porphyrin	142
Figure 4.5:	Optimisation convergence of upright water structure.....	143
Figure 4.6:	The water-bound resting state.....	145
Figure 4.7:	Iron anti-bonding d orbitals for the doublet spin upright structure of the water-bound ferric porphyrin species	148

Figure 4.8:	Possible α orbitals responsible for hydrogen bonding between hydrogen atoms of the bound water and porphyrin nitrogen atoms.....	149
Figure 4.9:	Structures and relative energies of doublet, quartet and sextet spin states of penta-coordinated ferric porphyrin.....	151
Figure 4.10:	Iron anti-bonding d orbitals for the doublet spin state of the substrate-bound ferric porphyrin species.	154
Figure 4.11:	Structures and relative energies of singlet, triplet and quintet spin states of penta-coordinated ferrous porphyrin.	157
Figure 4.12:	Structures and electronic distributions of the various spin states calculated for the dioxygen-bound ferrous complex.	159
Figure 4.13:	Structures and relative energies of doublet, quartet, sextet and octet spin states of dioxygen-bound reduced ferrous porphyrin.....	163
Figure 4.14:	Structures and relative energies of doublet, quartet and sextet spin states of hydroperoxo ferric porphyrin.	166
Figure 4.15:	Hydroperoxo ferric porphyrin.....	168
Figure 4.16:	Structures and relative energies of doublet, quartet and sextet spin states of Compound I.	171
Figure 4.17:	Compound I.	173
Figure 4.18:	Relative energies of the low, intermediate and high spin states of each species in the P450 reaction cycle	177
Figure 5.1:	Structures of various azole antifungal drugs currently in clinical use. ...	190
Figure 5.2:	The models used in the calculation of the 1D and 2D potential energy surfaces.....	194
Figure 5.3:	Graphs of energy in kcal mol ⁻¹ against Fe-N _{azole} distance in Å for the binding of methyl imidazole, methyl triazole and pyridine to haem using B3LYP	195
Figure 5.4:	Graphs of energy in kcal mol ⁻¹ against Fe-N _{azole} distance in Å for the binding of methyl imidazole, methyl triazole and pyridine to haem using BP86.....	196
Figure 5.5:	PESs of the displacement of water by azoles.....	200
Figure 5.6:	Images showing the lowest energy pathway between the two minima in the 2D PES of the displacement of water by pyridine.....	201
Figure 5.7:	Picture illustrating a transition state.....	202

Figure 5.8:	Transition state in the displacement of water bound to haem by triazole	204
Figure 5.9:	Two minima in the displacement of water by azoles.....	205
Figure 5.10:	Bridged-binding and direct binding observed in the fluconazole-bound CYP121 crystal structure	206
Figure 5.11:	Rotation energies of water bound to haem and triazole bound to haem.....	213
Figure 5.12:	Graphs of counterpoise corrected relative energies.	216
Figure 5.13:	Graphs of counterpoise corrected relative energies	217
Figure 5.14:	Models of the solvation of triazole and water binding to haem	221
Figure 5.15:	Dissociation curves of low, intermediate and high spin water and azoles.	226
Figure 5.16:	Fluconazole-bound CYP121 structure.....	228
Figure 5.17:	Structure of the imidazole-water bridged complex with an ethanol molecule fixed in the position of the Serine-237 residue in CYP121.....	229

LIST OF TABLES

Table 2.1:	Components of the bacterial growth media used per litre of ddH ₂ O	44
Table 3.1:	Plasmid stability test on a sample taken from a culture of HMS174(DE3) transformants containing the <i>Rv1394c</i> /pET15b plasmid	80
Table 3.2:	The maximum concentration of various solvents tolerated by CYP132	101
Table 3.3:	The K_d values for the binding of azole anti-fungal drugs to cleaved CYP132	108
Table 3.4:	Kinetics of oxidation of myristoleic, palmitoleic and arachidonic acids by CYP132.	116
Table 4.1:	Summary of the significant bond distances, bond angles, electron spin densities and atomic charges of the water-bound resting state	144
Table 4.2:	Summary of the energy and significant geometries of the water-bound resting state as found by various different groups	146
Table 4.3:	Iron d orbital energies for the doublet, quartet and sextet spin species of the water-bound resting species	147
Table 4.4:	Summary of the significant bond distances, bond angles, electron spin densities and atomic charges in the substrate-bound iron III species...	152
Table 4.5:	Iron d orbital energies for the penta-coordinated iron III species	153
Table 4.6:	Summary of the energy and significant geometries of the penta-coordinated ferric porphyrin, as found by various different groups	155
Table 4.7:	Iron d orbital energies for the penta-coordinated iron II species	157
Table 4.8:	Summary of the significant bond distances, bond angles, electron spin densities and atomic charges in the substrate-bound iron II species....	158
Table 4.9:	Summary of the energy and significant geometries of the penta-coordinated ferrous porphyrin, as found by various different groups	158
Table 4.10:	Summary of the significant bond distances, bond angles, electron spin densities and atomic charges in the dioxygen-bound ferrous porphyrin.	160
Table 4.11:	Summary of the energies and significant geometries found by other groups for the various spin states of the dioxygen-bound ferrous porphyrin.....	162
Table 4.12:	Summary of the significant bond distances, bond angles, electron spin densities and atomic charges in the O ₂ -bound iron III species	164

Table 4.13:	Summary of the energy and significant geometries of the dioxygen-bound reduced ferrous porphyrin, as found by various different groups	165
Table 4.14:	Summary of the significant bond distances, bond angles, electron spin densities and atomic charges in the hydroperoxo ferric porphyrin species	168
Table 4.15:	Summary of the energy and significant geometries of the protonated dioxygen-bound ferric porphyrin, as found by various groups	169
Table 4.16:	Iron d orbital energies for the hydroperoxo ferric porphyrin species	170
Table 4.17:	Summary of the significant bond distances, bond angles, electron spin densities and atomic charges in Compound I	173
Table 4.18:	Iron d orbital energies for Compound I	174
Table 4.19:	Summary of the energy, significant geometries and spin densities of Compound I, as found by various different groups	175
Table 5.1:	Fe-N _{azole} , Fe-O _{water} distances, relative energies and spin contamination values for the reaction pathways in the displacement of water by azoles	197
Table 5.2:	Fe-N _{azole} , Fe-O _{water} distances, relative energies and spin contamination values for the reaction pathways in the displacement of water by azoles	198
Table 5.3:	Geometry, energy and imaginary frequency of the transition states of the displacement of haem-water by azoles	203
Table 5.4:	Binding affinities of azoles and water for the P450 haem	207
Table 5.5:	Relative energies of triazole and water binding to haem using a range of basis sets	211
Table 5.6:	Relative energies of the binding of water and azoles to Fe-porphyrin-SCH ₃	211
Table 5.7:	Counterpoise corrected relative energies and BSSE values for the binding of haem to water and azoles.	215
Table 5.8:	The axial bond distances (Å) in the counterpoise corrected structures of the water- and azole-bound haem species	215
Table 5.9:	PCM calculations of haem binding to triazole and water	219
Table 5.10:	Results of the calculations of the binding of water and triazole to haem, using the solvation models 1 and 2	222

Table 5.11:	Relative energies of water and triazole binding to iron-porphyrin, relative to the separated species, with the water or triazole molecules bound to a cluster of 9 water molecules.....	223
Table 5.12:	Dissociation energies of water- and azole-bound haem.	225
Table 5.13:	Relative energies of water and imidazole binding to haem in the presence of an ethanol molecule fixed in the position of the Serine-237 residue in CYP121..	229

ABSTRACT

The University of Manchester
Faculty of Engineering and Physical Sciences

**ABSTRACT OF THESIS submitted by Philip R. Balding
For the degree of Doctor of Philosophy (PhD)
and entitled "The Study of *Mycobacterium tuberculosis* Cytochrome P450 Enzymes by
Experimental and Theoretical Methods"**

Date of submission 31/03/2008

Cytochromes P450 (P450s) are ubiquitous haemoproteins that are found in every family of organisms. P450s have a wide range of functions in nature, from roles in protection, in the metabolism of xenobiotics, to the synthesis of critical signalling molecules and the conversion of fatty acids to biologically active molecules. *Mycobacterium tuberculosis* (*Mtb*), the pathogen responsible for the disease TB, is responsible for more human deaths than any other infectious agent. With 9.2 million new cases and 1.7 million deaths from TB worldwide each year, and the increasing prevalence of multidrug-resistant strains, there is an urgent need for new treatments to combat *Mtb*. The genome of *Mtb* was solved in 1998 (Cole *et al.* (1998) *Nature* **393**, 537-544) and revealed the presence of 20 P450s in *Mtb*. P450s are known targets for azole inhibitors, and presented exciting novel approaches to *Mtb* treatments.

In order to develop new treatments for *Mtb* P450s, and understand their mode of activity in the bacterium, it is important to understand how P450s bind azole anti-fungals, and the mechanism by which they turn over substrates. To this end, this thesis describes quantum mechanical calculations, using Density Functional Theory (DFT) into the catalytic cycle of P450s, and the binding of azoles to the P450 haem. These processes were also studied experimentally in the *Mtb* P450 CYP132.

CYP132 shows significant homology at the amino acid level with various eukaryotic CYP4 family fatty acid hydroxylases and has been shown to be important to the virulence of *Mtb*. CYP132 was expressed in *E. coli* and purified. The purified enzyme was found to have been proteolytically cleaved, with the loss of around 10 kDa (~1/5 of the total enzyme). The cleaved enzyme was found to retain its P450 character, however. Cleaved CYP132 bound azole inhibitors, and was found to carry out a range of oxidation reactions on unsaturated fatty acids.

DFT calculations were carried out on each spin state of all haem species in the P450 catalytic cycle. This provided insights into the fine tuning of the haem spin states, and how small changes in the haem environment may alter the behaviour of the protein.

DFT calculations on the binding of azoles to haem revealed a reaction pathway for the displacement of bound water by azoles, involving a transition state with both ligands dissociated from the haem. Also, the role of unfavourable interactions of azoles with bulk water solvent in the reaction was revealed.

DECLARATION

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

COPYRIGHT STATEMENT

- i. The author of this thesis (including any appendices and/or schedules to this thesis) owns any copyright in it (the "Copyright") and s/he has given The University of Manchester the right to use such Copyright for any administrative, promotional, educational and/or teaching purposes.
- ii. Copies of this thesis, either in full or in extracts, may be made **only** in accordance with the regulations of the John Rylands University Library of Manchester. Details of these regulations may be obtained from the Librarian. This page must form part of any such copies made.
- iii. The ownership of any patents, designs, trade marks and any and all other intellectual property rights except for the Copyright (the "Intellectual Property Rights") and any reproductions of copyright works, for example graphs and tables ("Reproductions"), which may be described in this thesis, may not be owned by the author and may be owned by third parties. Such Intellectual Property Rights and Reproductions cannot and must not be made available for use without the prior written permission of the owner(s) of the relevant Intellectual Property Rights and/or Reproductions.
- iv. Further information on the conditions under which disclosure, publication and exploitation of this thesis, the Copyright and any Intellectual Property Rights and/or Reproductions described in it may take place is available from the Head of School of (*insert name of school*) (or the Vice-President) and the Dean of the Faculty of Life Sciences, for Faculty of Life Sciences' candidates.

ACKNOWLEDGEMENTS

I would like to thank all those who have helped me during my PhD. To my supervisors, Andy Munro and Mike Sutcliffe for their continuing patience, to all the members of both research groups for making research the pleasurable experience it should be, especially Kirsty McLean for help with the experimental work and J-D Marechal for help with the theoretical work. Also, thank you to Sam de Visser for assistance with the theoretical work and writing of publications, to Martin Warren and Andrew Lawrence for their help with mass spectrometry data collection.

CHAPTER 1: INTRODUCTION

1.1 *Haem and Haemoproteins*

1.1.1 Haem structures in nature

Haemoproteins are essential for respiration, oxygen metabolism and electron transfer in a wide range of prokaryotes and eukaryotes (O'Brian, 1999). Haem is the active moiety (cofactor) of proteins that carry out a wide range of biological functions, including haemoglobins, hydroxylases, catalases, peroxidases and cytochromes. Haem is one of the most versatile of protein cofactors (Munro *et al.*, 2007) and is involved in steroid and bioactive lipid synthesis, energy transduction, gene regulation cellular signalling, oxygen transport and antibiotic biosynthesis. Haem has roles as a biosensor of signal transduction pathways that respond to dioxygen (O₂), carbon monoxide (CO) and nitric oxide (NO) (Gilles-Gonzalez *et al.*, 1991, Lowenstein and Snyder, 1992, O'Brian, 1999, Verma *et al.*, 1993).

There are a number of haem structures in nature, which are shown in Figure 1.1. The most common haem structure in nature is protohaem, haem *b* (Yamanaka, 1988). This is the prosthetic group of haemoglobin, myoglobin, catalase, cytochromes *b*, and some haem-containing oxygenases, including cytochromes P450, which are the subject of this thesis. Haem *b* is the starting point for the synthesis of the other haem structures.

Haem *a* (Caughey *et al.*, 1975) is formed from haem *b* by the addition of a 17-hydroxyethylfarnesyl moiety by haem *o* synthase (haem *o* is formed as a stable intermediate), followed by modification of a methyl group to a formyl group by haem *a* synthase (Brown *et al.*, 2004, Mogi *et al.*, 1994). Haem *a* is found in the terminal oxidases of aerobic respiratory chains (Munro *et al.*, 2007). This includes cytochrome *c* oxidase, which is the terminal oxidase in all plants, animals, aerobic yeasts, and some bacteria (Brown *et al.*, 2004).

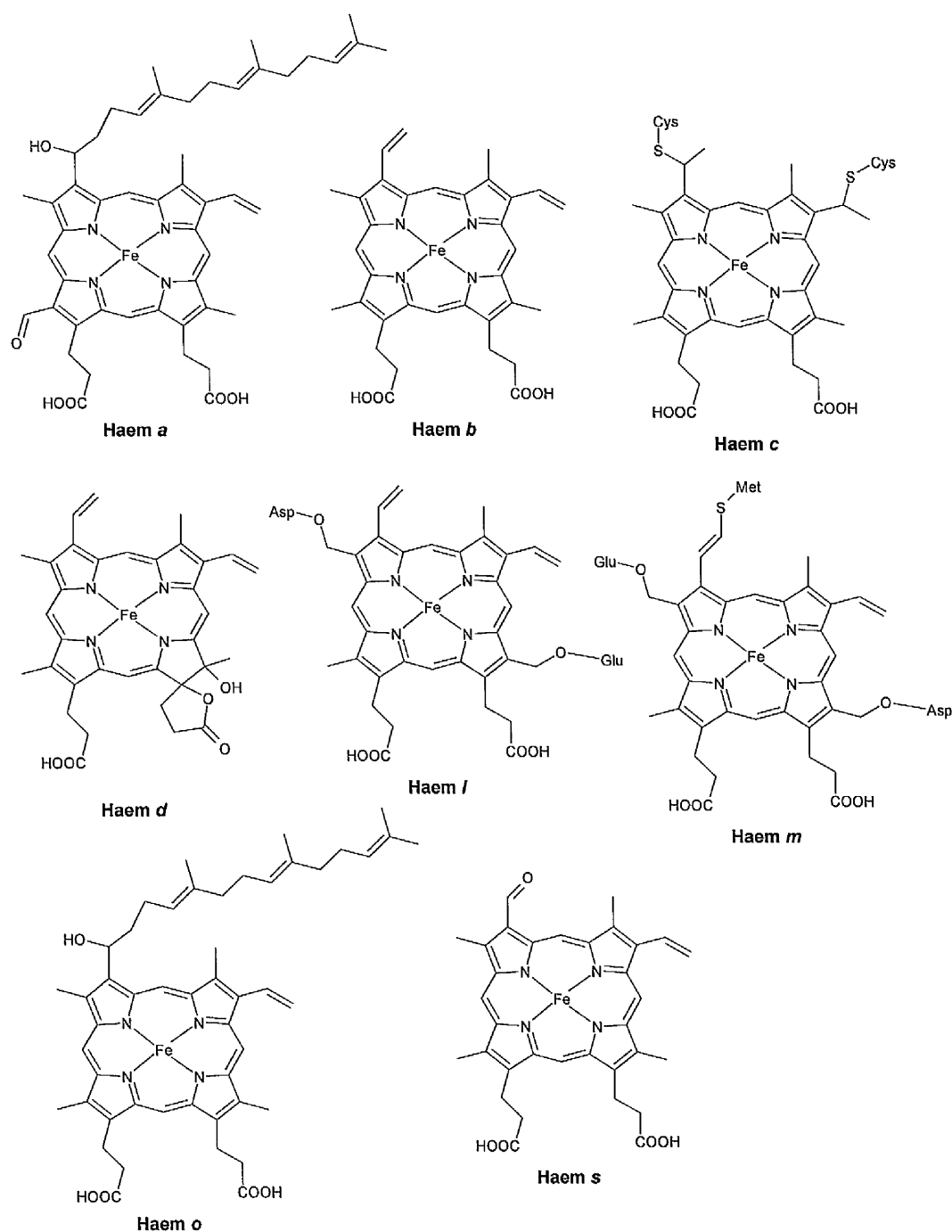


Figure 1.1: Structures of the haem cofactors found in nature, adapted from (Munro *et al.*, 2007). Haem c is covalently bound to the host protein via thioether linkages to cysteine residues. Haem I is covalently linked to glutamate/aspartate residues in the protein through esterification of two methyl group substituents. Haem m is also covalently linked to glutamate/aspartate residues in the protein, and forms a sulphonium bond between the haem 2-vinyl group and a methionine residue.

Haem *c* forms covalent bonds with the apoprotein via thioether linkages (Yeh and Han, 1998). Haem *c* is found in various bacterial respiratory proteins (Munro *et al.*, 2007). A number of cytochromes, mainly from single-cell organisms, have been found to contain numerous haem *c* cofactors, e.g. the dimeric octahaem cytochrome *c*₃ protein from *Desulfovibrio desulfuricans* (Czjzek *et al.*, 1996). There is some speculation as to why the covalent linkages are necessary in haem *c*. Possible explanations include the covalent bonding providing an additional structural element in the protein, and leading to more stable binding of the haem and/or changes in its redox properties. Also the covalent bonding allows a greater number of haems per length of polypeptide, leading to more efficient electron transfer systems that can “harvest” electrons using distinct surface-exposed haems as points of exit and entry for electrons (Barker and Ferguson, 1999).

In haem *d*, modification of one of the propionate side chains causes a fifth ring in the haem macrocycle (Figure 1.1) (Murshudov *et al.*, 1996). Haem *d* is found in certain catalase enzymes and is the site for oxygen reduction to water in some bacterial terminal respiratory oxidases (Munro *et al.*, 2007). Haem *d*₁ is a variant ferric-dioxoisobacteriochlorin structure found in the cytochrome *cd*₁ nitrite reductase class of enzymes (Chang and Wu, 1986).

Haem *l* is found in mammalian-type peroxidases (Oxvig *et al.*, 1999, Rae and Goff, 1998). It is covalently linked to glutamate/aspartate residues from the protein via esterification of two haem methyl groups (Munro *et al.*, 2007).

Haem *m* is found in the respiratory burst enzyme myeloperoxidase (Fenna *et al.*, 1995). In addition to forming covalent bonds to the protein glutamate/aspartate residues via esterification of the haem methyl groups, as seen in haem *l*, haem *m* forms a sulphonium ion linkage between a methionine side chain and a haem vinyl group (Fiedler *et al.*, 2000). Myeloperoxidase is thought to be a key component of the respiratory burst (i.e. oxygen radical production) in host defence mechanisms against infection and may be important in inflammatory diseases (Munro *et al.*, 2007).

Haem *o* is formed as a stable intermediate in the biosynthesis of haem *a*. It is found in the *E. coli* cytochrome *bo* complex, which has a role as a terminal quinol oxidase in the aerobic respiratory chain under highly aerated growth conditions (Mogi *et al.*, 1994).

Haem *s* is found in marine worm haemoglobins (Munro *et al.*, 2007). It has a formyl group in place of a haem *b* vinyl group (Figure 1.1) (Caughey *et al.*, 1975).

1.1.2 Biosynthesis of haem

The haem biosynthetic pathway is summarised in Figure 1.2. The intermediate products from δ -aminolevulinic acid (δ -ALA) to protohaem are conserved across all organisms studied to date (Yamanaka, 1988), although there is variability among organisms in the properties of the enzymes that catalyse the steps, their regulation and their compartmentalisation within the cell. δ -ALA is synthesised either from glutamate, via the C_5 pathway, or from succinyl CoA and glycine by δ -ALA synthase (ALAS). ALAS is found in animals, yeast and some bacteria. The C_5 pathway is observed in plants, algae and other bacteria (Yamanaka, 1988).

The biosynthetic pathway of haem is shared with that of chlorophylls, with the biosynthesis of chlorophylls diverging from that of haem at the iron inclusion step (Munro *et al.*, 2007). Two molecules of δ -ALA are brought together to form porphobilinogen (PBG) catalysed by δ -ALA dehydratase (also known as PBG synthase). The condensation of four PBG molecules by porphobilinogen deaminase (PBD) produces the linear tetrapyrrole, hydroxymethylbilane. Next, uroporphyrinogen III synthase (UPS) inverts the pyrrole D ring and effects cyclisation, leading to uroporphyrinogen III. Decarboxylation of the four acetate side chains by uroporphyrinogen decarboxylase (UPD) follows, with evolution of four molecules of CO_2 , and giving the product coproporphyrinogen III. The enzyme coproporphyrinogen oxidase (CPO) then performs oxidative decarboxylation, transforming two propionate side chains into vinyl groups and forming protoporphyrinogen IX. The synthesis of the haem skeleton is completed with the formation of protoporphyrin IX, by the oxidation of methylene bridges connecting the pyrrole moieties into methenyl bridges, catalysed by protoporphyrinogen oxidase (PPO). The insertion of iron by ferrochelatase (also known as haem synthase) produces protohaem (or haem *b*).

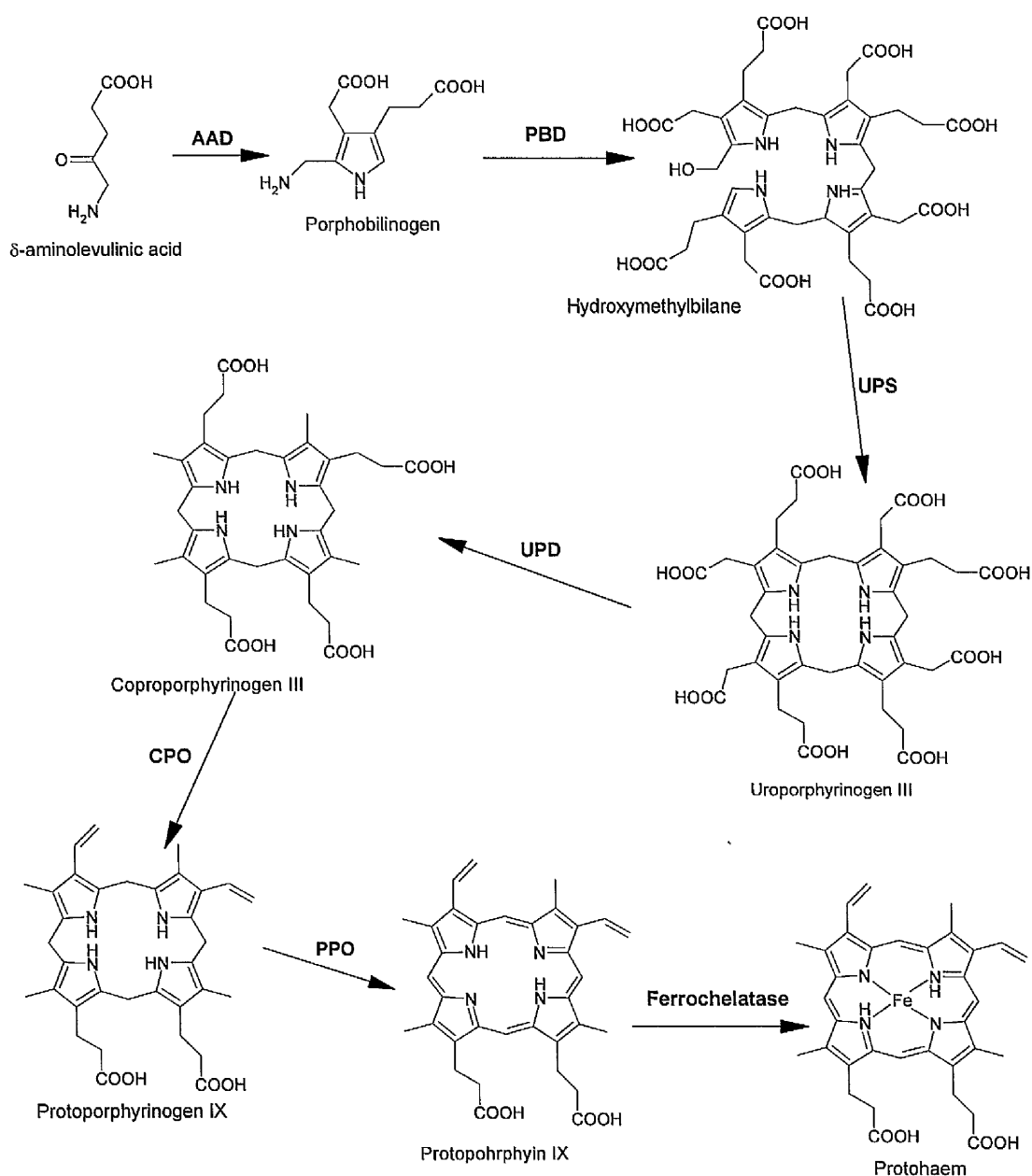


Figure 1.2: Generic haem biosynthetic pathway, adapted from (Munro *et al.*, 2007) and (Yamanaka, 1988). Origin of δ -aminolevulinic acid (δ -ALA) varies between organisms. Enzymes involved in each step are: δ -ALA dehydratase (AAD, also known as PBG synthase), porphobilinogen deaminase (PBD, also known as hydroxymethylbilane synthase), uroporphyrinogen III synthase (UPS), uroporphyrinogen decarboxylase (UPD), coproporphyrinogen oxidase (CPO), protoporphyrinogen oxidase (PPO) and ferrochelatase (also known as haem synthase).

1.1.3 Haemoproteins

There are a great variety of haem-containing proteins, with various structures, functions and chemistries (Chapman *et al.*, 1997). There are a number of axial ligands to haem observed in nature. These play an important role in the modulation of the haem iron reduction potential, and the redox chemistry of the associated haemoprotein (Munro *et al.*, 2007). The haemoproteins that are involved in the binding of and/or the reactions with a ligand to the haem iron are typically pentacoordinated, with a sixth coordination site available for substrate binding. This is observed in e.g. the oxygenases and peroxidases, as well as haemoglobin and myoglobin. The haemoproteins that are only involved in electron transfer reactions are typically hexacoordinated. Reasons for this include the ability to poise the redox potential at the required level and to avoid interactions with inhibitors and unwanted reduction of oxygen at a vacant coordination site on the haem iron.

Haemoproteins may bind ligands for transport or storage, or as sensors of a particular material. Haemoglobin and myoglobin bind molecular oxygen (Lukin and Ho, 2004, Perutz, 1978). Haemoglobin is found in red blood cells and is responsible for oxygen transfer around the host organism. Myoglobin is found in red muscle, binding oxygen and transferring it to the mitochondria. Haemoglobin consists of four polypeptide subunits, each containing a single haem. Myoglobin is similar in structure to a single haemoglobin subunit. The haem cofactors in ferrous haemoglobin and myoglobin are penta-coordinate, with a single histidine axial ligand, called the proximal histidine. On the opposite side of the haem from the proximal histidine, the 'distal histidine' and 'distal valine' residues are close to the haem, but do not bind to it. Instead oxygen coordinates to the ferrous haem iron.

Nitrophorins are haem *b* containing proteins found in the salivary glands of blood sucking insects such as bedbugs (*Cimex lectularius*) (Berry *et al.*, 2007). The haem is coordinated by either a proximal histidine or cysteine ligand, with a vacant axial position allowing binding of other ligands. Nitrophorins carry NO, which is synthesised by a nitric oxide synthase (NOS). In the cysteine-coordinated form type of nitrophorins (and at high NO concentrations), the NO may be carried as a S-nitroso conjugate of the proximal cysteine (as well as on the haem iron) (Weichsel *et al.*, 2005). Upon injection into the tissues of the victim, NO dissociates from the nitrophorin and causes vasodilation,

allowing more blood to be transported to the site of the wound. Some histidine ligated nitrophorins also bind histamine, which is released into the the victim and causes swelling and itching. This prevents detection of the insect for a period of time, which allows it to obtain a sufficient blood meal.

Pentacoordinated haems are also found in haemoproteins that are biosensors of diatomic gases. In *Mycobacterium tuberculosis* the haemoproteins DevS (also known as DosS) and DosT are sensory kinases that control the activity of the protein DevR (also known as DosR) (Sousa *et al.*, 2007, Yukl *et al.*, 2007). These proteins are considered to provide the gene regulation that switches the mycobacterium from a replicating to a non-replicating, persistent infective state. The kinase activity of DevS and DosT is coupled to the coordination state of the haem, such that the binding of carbon monoxide (CO) or NO causes the kinase activity to be 'switched on', whereas the oxygen-bound or pentacoordinated states are 'switched off'. When their kinase function is active, DevS and DosT phosphorylate DevR, which gains enhanced affinity for its DNA recognition sequence, causing it to activate the hypoxic response at the transcriptional level.

Haemoproteins that only carry out electron transfer are mainly involved in energy transduction, such as in photosynthesis or respiration (Chapman *et al.*, 1997). These include mitochondrial and other cytochromes *c*, cytochromes *bc₁*, and cytochrome *f* (Munro *et al.*, 2007). In cytochromes *c*, the haem is typically hexacoordinated, with two histidine residues as axial ligands (Skulachev, 1998). Cytochrome *c* transports electrons from cytochrome *c₁* to cytochrome oxidase. It has also been found to have a role in apoptosis (Liu *et al.*, 1996). Other haemoproteins with two histidine axial ligands to the haem include cytochromes *b₅*, which carry out electron transfer in a wide range of eukaryotes, as well as in some prokaryotes (Schenkman and Jansson, 2003). Cytochromes *bc₁* have histidine and methionine axial ligation to the haem (Xia *et al.*, 1997). This forms part of an electron transport chain, transferring electrons to the membrane protein cytochrome *c*, which in turn transfers electrons to cytochrome *c* oxidase. Cytochrome *f* has a histidine axial haem ligand. The sixth coordination site to the haem iron is taken up by a nitrogen ligand from the terminal amine of the peptide (which is a tyrosine amino acid). This protein is involved in electron transfer in photosynthesis.

The driving force for electron transfer reactions is determined by the redox potential of the donor and acceptor, which can be modulated across a range of more than 0.7 V by the protein environment (Chapman *et al.*, 1997). Also, the electron transfer rate has been found, generally, to decrease exponentially with the distance between the donor and acceptor molecules (often protein cofactors), at an average rate of 1.4 \AA^{-1} (Beratan and Onuchic, 1996). This can also be used to control the electron transfer reactions in nature.

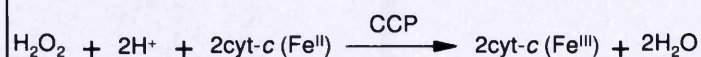
Many haemoproteins incorporate electron transfer mechanisms that are linked to catalysis. These include catalases, peroxidases and P450s, which are the subject of this thesis.

Most catalases contain haem *b* bound to a tyrosine residue. Catalases catalyse the decomposition of harmful hydrogen peroxide to water and dioxygen, as shown below (Ko *et al.*, 2000).



Some peroxidases catalyse the decomposition of alkyl peroxides as well as hydrogen peroxide (Dunford and Stillman, 1976). Most peroxidases contain haem *b* bound to a histidine residue (Fulop *et al.*, 1994, Rodriguez-Lopez *et al.*, 2001). The reaction scheme for the decomposition of hydrogen peroxide by peroxidases, as shown for cytochrome *c* peroxidase (CCP) from yeast mitochondria (Fulop *et al.*, 1994), is illustrated in Figure 1.3. First CCP reacts with hydrogen peroxide, generating a doubly oxidised oxy-ferryl enzyme intermediate, termed Compound I, and a semi-stable free radical associated with Trp191. The free radical reacts with ferrocytochrome *c*, forming Compound II. Compound II then receives a second electron from another ferrocytochrome *c*, returning CCP to the ground state. The crystal structure of CCP Compound I is shown in Figure 1.4. It was found that only slight structural changes occur on the conversion from the ground state to Compound I.

Overall reaction:



Three major steps:

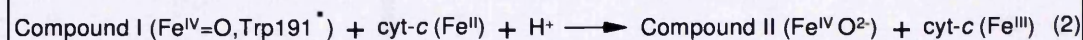
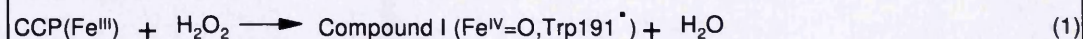


Figure 1.3: Reaction scheme for the decomposition of hydrogen peroxide by cytochrome *c* peroxidase (CCP), adapted from (Fulop *et al.*, 1994). The overall reaction can be broken down into three major steps. Cyt-*c* (Fe^{II}) = ferrocycytochrome *c*, cyt-*c* (Fe^{III}) = ferricytochrome *c*

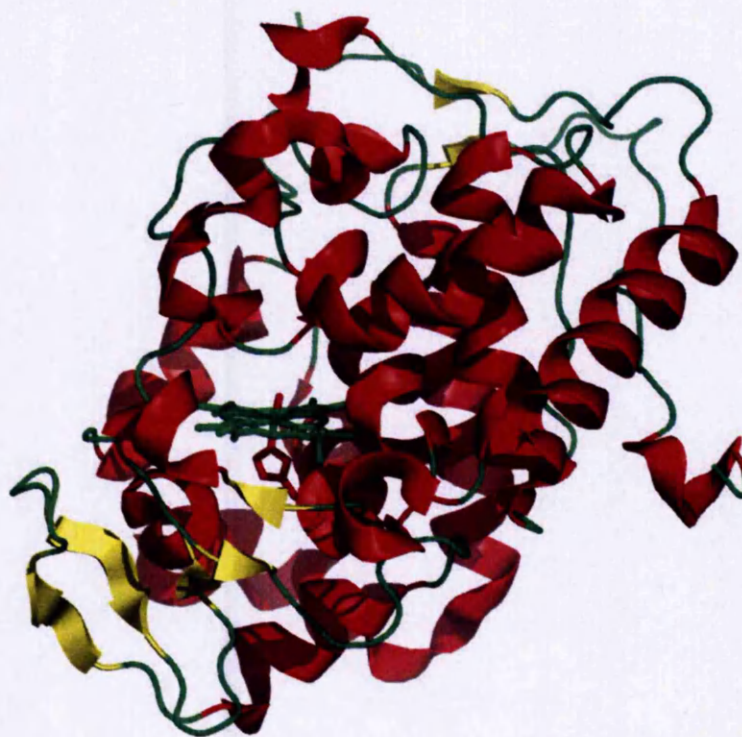


Figure 1.4: X-ray diffraction crystal structure of cytochrome *c* peroxidase Compound I. PDB number 1EBE (Fulop *et al.*, 1994). Haem and His-175, which ligates the haem, are shown, along with the bound oxygen atom. The structure of the protein and haem were found to change little upon conversion to Compound I, with respect to the resting state structure.

1.2 Cytochromes P450

Cytochromes P450 (P450s) are haemoproteins that share many similarities with peroxidases and catalases, mentioned above. P450s are ubiquitous, being found in all domains of life, from bacteria and fungi, through to plants, birds and mammals (Groves, 2005). P450s catalyse a range of reactions, these are predominantly oxidations and hydroxylations, but dealkylation, dehydrogenation and carbon-carbon cleavage reactions have also been observed (Ortiz de Montellano and de Voss, 2005).

P450s contain a haem *b* cofactor, with a cysteine axial ligand (Poulos and Johnson, 2005). In the P450 resting state, the sixth ligand position is occupied by a water molecule, this is replaced by oxygen in various oxidation states during the catalytic cycle, including O_2 , O_2^{2-} and OOH^- (Makris *et al.*, 2005). Various other species can also bind to the sixth position of the haem, including CO, NO, imidazole and other azoles. The binding of CO to P450s gives rise to a characteristic absorption at ~450 nm (Klingenberg, 1958), which gave P450s their name (Omura and Sato, 1964).

P450s carry out a range of functions in nature. These mostly fall into two broad categories, the metabolism of xenobiotics, and the biosynthesis of critical signalling molecules used for control of development and homeostasis (Denisov *et al.*, 2005). In mammals these functions include the metabolism of drugs and other xenobiotics, the synthesis of steroid hormones and fat-soluble vitamin metabolism, and the conversion of polyunsaturated fatty acids to biologically active molecules.

In plants P450s are utilised to synthesise hormones and degrade herbicides. Similarly, in insects P450s control development through hormone biosynthesis, and provide insecticide resistance.

There are 57 P450s encoded in the human genome (Guengerich, 2005). These have roles in steroid production, metabolism of vitamins A and D, fatty acid metabolism and transformation of xenobiotic chemicals. The most abundant human P450 is CYP3A4. This is responsible for the metabolism of around 50 % of the pharmaceuticals currently on the market (Denisov *et al.*, 2005).

P450s have varying substrate specificities, from those that bind a relatively small number of substrates, such as P450eryF, which catalyses the 6S-hydroxylation of 6-deoxyerythronolide B, in the synthesis of the polyketide antibiotic erythromycin (Cupp-Vickery and Poulos, 1995), to the human CYP3A4, which binds and metabolises around 50 % of the pharmaceuticals currently in use (Guengerich, 2005). Upon binding of a substrate to P450s, the bound water molecule is typically displaced, causing the haem iron to move out of the haem plane, giving rise to a shift from the low spin to the high spin state (Keller *et al.*, 1972).

Mixed function oxidation by enzymes in the liver were first reported in the late 1940's (Brodie *et al.*, 1958, Guengerich, 2001, Mueller and Miller, 1948). The enzymology behind these observations was not defined for a number of years. Klingenberg reported the spectral changes that occurred upon binding of CO to pigments in microsomes from rat liver (Klingenberg, 1958). Sato and Omura investigated the behaviour of the microsomes upon reduction and oxidation, and first used the term 'P-450' (Omura and Sato, 1962), referring to the characteristic absorption at 450 nm upon binding of CO to the reduced haem-containing enzyme that we now know to be cytochrome P450. Estabrook and Cooper established a P450 as the terminal oxidase in an adrenal steroid hydroxylation (Estabrook *et al.*, 1963). Induction experiments in rats during the 1960's led to the conclusion that multiple P450s might be present in liver tissues (Alvares *et al.*, 1967, Hildebrandt *et al.*, 1968).

The first P450 x-ray crystal structure to be solved was for *Pseudomonas putida* P450cam (Poulos *et al.*, 1987). Since then the structures of numerous P450s have been solved. These include the bacterial P450s P450 BM3 from *Bacillus megaterium* (Ravichandran *et al.*, 1993), P450eryF from *Saccharopolyspora erythraea* (Cupp-Vickery and Poulos, 1995), *Mycobacterium tuberculosis* (*Mtb*) CYP121 (Leys *et al.*, 2003), CYP51B1 (Podust *et al.*, 2001) and recently CYP130 (Ouellet *et al.*, 2008), and the human P450s CYP3A4 (Yano *et al.*, 2004), CYP2C9 (Williams *et al.*, 2003) and CYP1A2 (Sansen *et al.*, 2007). The overall fold and topology is conserved across the P450s for which structures have been solved to date (Denisov *et al.*, 2005). This is despite only 20 % amino acid sequence identity being shared across the P450 superfamily (Hasemann

et al., 1995). The P450 fold is also unique; no non-P450 structure has yet been found that shares the P450 fold (Poulos and Johnson, 2005).

The crystal structure of P450cam (Poulos *et al.*, 1987) is shown in Figure 1.5. The α helix and β sheet regions are labelled A-L and 1-4, respectively, according to the labelling system employed by Poulos *et al.* The conserved P450 structural core consists of the three parallel helices D, L and I, and the antiparallel helix E (Denisov *et al.*, 2005). The haem *b* group is held between the distal I helix and proximal L helix, and the haem iron is bound to a cysteine residue, which is conserved across all P450s. The cysteine typically forms two hydrogen bonds with amides in the protein backbone (Denisov *et al.*, 2005).

Despite the conserved P450 fold, there is considerable diversity in the size and structure of the substrates bound by different P450s. Six 'substrate recognition sites' (SRSs) have been identified that are largely responsible for P450 substrate specificity (Gotoh, 1992). These are flexible protein regions that line the P450 active site (Denisov *et al.*, 2005). The SRSs move upon substrate binding to accommodate the substrate, leading to their optimal binding for the catalytic reaction.

A simple and effective method for following changes in binding of substrates and ligands to P450s is by measuring changes in the UV/Vis spectrum. P450s have a characteristic, haem *b* UV/Vis spectrum, as shown for P450 BioI (Lawson *et al.*, 2004) in Figure 1.6. The low spin resting state has α and β bands at ~570 nm and ~540 nm, respectively, with a γ (Soret) band at 418 nm. Upon substrate binding, the water bound to the P450 haem iron is typically displaced, causing the iron to shift to the high spin state. This gives rise to a shift in the Soret band from ~420 nm to ~395 nm, which is called a type I shift (Schenkman *et al.*, 1967, Schenkman and Jansson, 2006). Upon binding of an inhibitor that ligates the haem iron, such as imidazole, the Soret band shifts to higher wavelengths, usually between 425 nm and 435 nm. This is called a type II shift. Certain types of inhibitor may cause the P450 to shift from the high spin towards the low spin state. This is called a 'reverse type I' shift. This was observed e.g. for the addition of various steroids to P450 BioI, and is observed for many types of P450 inhibitors (Lawson *et al.*, 2004)

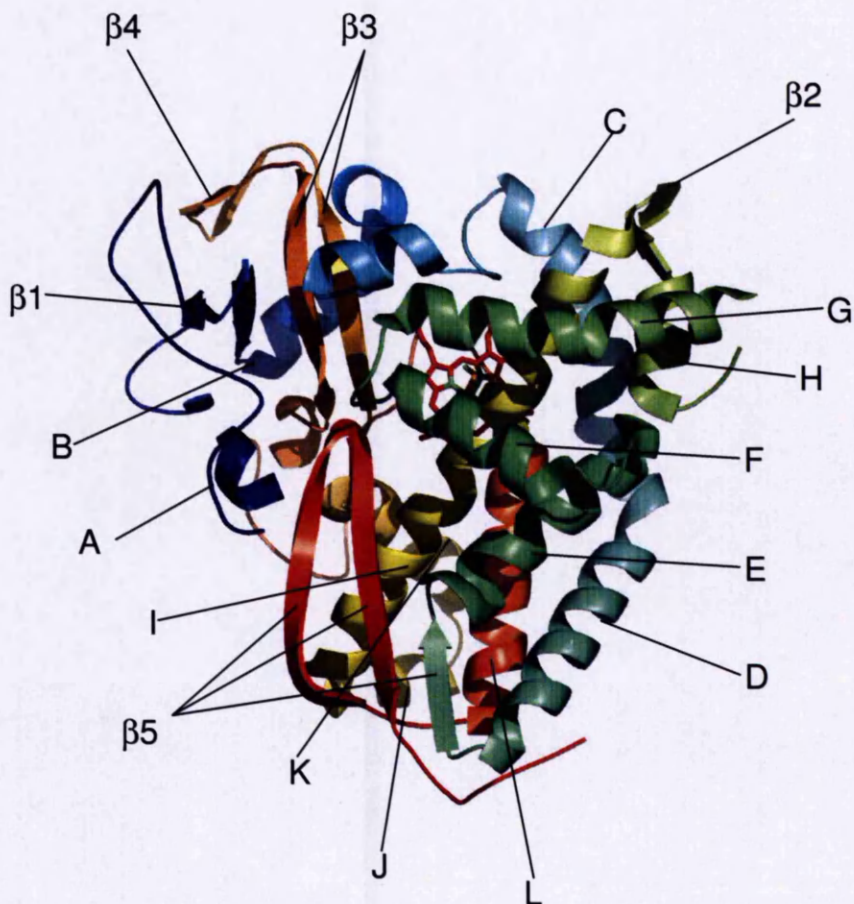
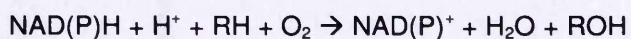


Figure 1.5: Crystal structure of *Pseudomonas putida* P450cam (Poulos *et al.*, 1987). The α helix regions are labelled A-L and the β -sheet regions 1-4. Labels are taken from (Poulos *et al.*, 1987). The conserved P450 structural core consists of the four helices, D, L, I and E.

. The most common reaction catalysed by P450s is the insertion of a single oxygen atom, often into a C-H bond (hydroxylation) or across a C=C double bond (epoxidation). The activation of molecular oxygen by P450s requires two electrons, which are usually obtained from either NADPH or NADH. The general reaction for these processes is shown below (Guengerich, 2001).



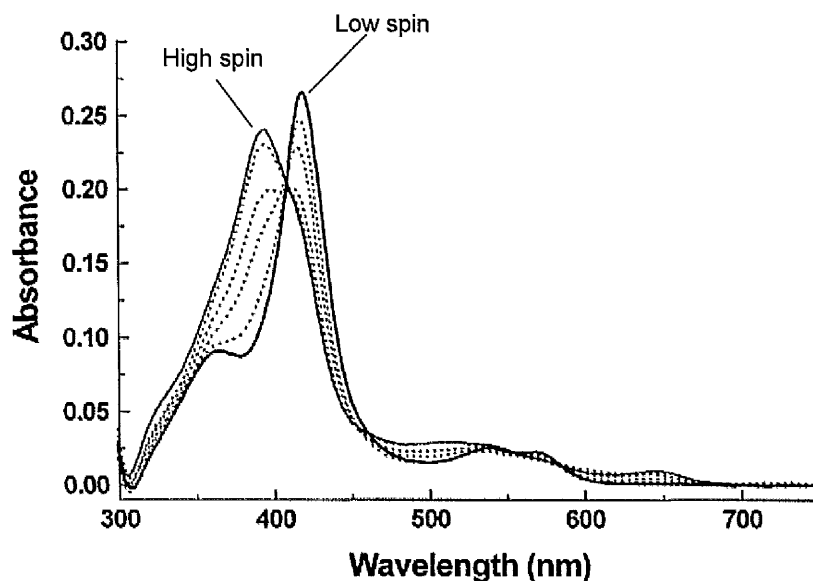


Figure 1.6: UV/Vis spectra of P450 Biol, in the high and low spin states, taken from (Lawson *et al.*, 2004). The Soret band peak shifts from 418 nm to 394 nm as the protein shifts from the low spin water-bound state to the high spin substrate-bound state.

The generally accepted catalytic cycle for the activation of molecular oxygen by P450s is shown in Figure 1.7 (Shaik and de Visser, 2005). The sequential two-electron reduction of cytochrome P450, and the existence of multiple intermediates, was first proposed in P450cam in the late 1960s (Conrad *et al.*, 1965, Denisov *et al.*, 2005, Hedegaar and Gunsalus, 1965, Katagiri *et al.*, 1968). The binding of a substrate usually displaces the water molecule bound as the 6th ligand to the haem iron, inducing a shift to the high spin state. The reduction potential of high spin P450 is more positive than for the low spin form. This allows the transfer of the first electron from a redox partner enzyme and the reduction of haem to the ferrous species. Oxygen binding leads to an oxy-P450 complex. This is the last relatively stable intermediate in the cycle. The addition of the second electron from NAD(P)H (again via a redox partner protein), and two protons, leads to formation of the highly reactive Compound I. Unlike peroxidases, such as cytochrome *c* peroxidase, for which the crystal structure of Compound I was shown in the previous section (Figure 1.4), Compound I of P450s is a very short lived species, and has only been putatively observed by flash photolysis studies (Newcomb *et al.*, 2006). There has been intense debate for many years over the identity of the 'true' oxidant species in the P450 reaction cycle (Shaik *et al.*, 2005a), with various other alternatives being put forward, such as an iron(V)oxo species, or iron-peroxo species.

This will be discussed further in Chapter 4. Compound I is still generally accepted as being the major oxidant species in the P450 catalytic cycle.

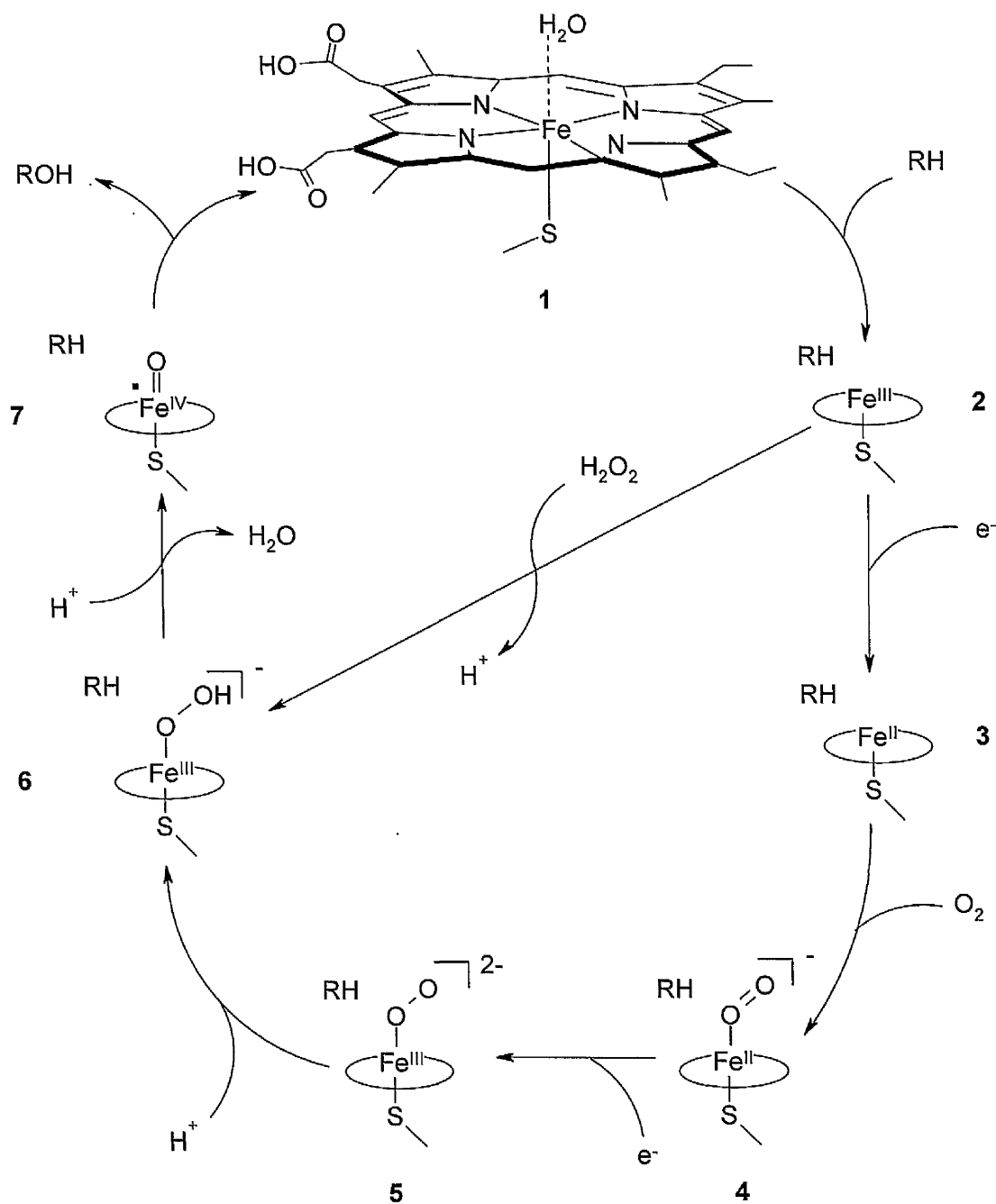


Figure 1.7: The consensus catalytic cycle for the activation of molecular oxygen by P450s and the hydroxylation of a substrate, adapted from (Shaik and de Visser, 2005). Except for the resting state, the haem macrocycle is represented by an ellipse. The substrate is represented by RH.

Compound I inserts an oxygen atom into the substrate molecule. The mechanism for this is generally believed to be the 'rebound mechanism' proposed by Groves *et al.* (Groves and McClusky, 1976, Groves, 1985). This is shown in Figure 1.8.

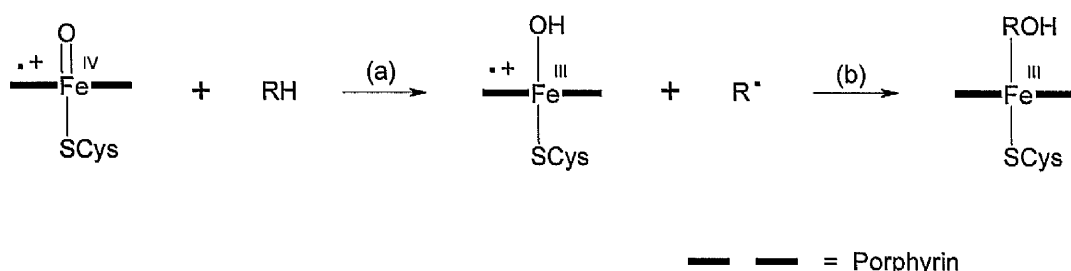


Figure 1.8: Schematic representation of the rebound mechanism proposed by Groves *et al.* (Groves and McClusky, 1976, Groves, 1985).

Most P450s can be divided into two classes, with respect to the electron transfer system used to supply the electrons for the activation of oxygen (Paine *et al.*, 2005). Class I P450s include bacterial and mitochondrial proteins, receiving electrons from a two-component shuttle system, consisting of a ferredoxin (an iron-sulphur cluster-containing protein) and a flavoprotein ferredoxin reductase. Electrons are donated from NAD(P)H to the flavin adenine dinucleotide (FAD)-containing ferredoxin reductase, and then shuttled through the iron-sulphur centre of the ferredoxin one at a time to the P450 haem.

Class II P450s utilise a single membrane-bound redox partner. This is NADPH cytochrome P450 reductase (CPR), which contains two flavin cofactors, FAD and flavin mononucleotide (FMN). In *Bacillus megaterium* P450 BM3 (CYP102A1) soluble CPR and P450 domain are fused together in a single 119 kDa polypeptide. This arrangement enables efficient electron transfer between the adjacent domains and results in P450 BM3 having the fastest rate of substrate oxidation yet determined for a P450 (Noble *et al.*, 1999).

The *E. coli* flavodoxin (FLD) and flavodoxin reductase (FLDR) redox partners have been used in place of ferredoxin and ferredoxin reductase proteins to drive catalysis in a number of P450 systems, including a range of mammalian P450s expressed in *E. coli*

(Jenkins and Waterman, 1998). This system was utilised in the reaction of CYP132 with fatty acids in Chapter 3 of this thesis.

1.3 Tuberculosis – A ‘Global Emergency’

Tuberculosis (TB), from the bacterium *Mycobacterium tuberculosis* (*Mtb*) is a disease of antiquity, with spinal disease being found in Egyptian mummies (Harries and Dye, 2006). It was thought that *Mtb* evolved from the bovine bacterium *Mycobacterium bovis* and was acquired by humans during the development of agriculture around 9000 years ago (Russell, 2007). More recent findings suggest *Mtb* and *M. bovis* originated from a common ancestor, and have been with humans much longer than this.

During the Industrial Revolution TB became particularly prevalent, gaining notoriety as ‘the captain of all these men of death’ (Harries and Dye, 2006). By the 19th century TB was the greatest killer across Europe, with incidence of disease reaching eight cases in a thousand. In 1921, the *Bacille Calmette-Guérin* (BCG) vaccine was developed, and with the introduction of freeze-drying in 1943, campaigns of mass BCG vaccination could be embarked on (Colditz *et al.*, 1994).

In the 1940s TB drug treatments also began to be discovered, with streptomycin being introduced in 1943, isoniazid in 1952 and later rifampicin in 1965. The combination of the vaccination program and drug treatments led to a marked decline in incidence of TB in developed countries in the 1970s and 1980s, although it remained prevalent in developing countries. With this decline in incidence, TB slipped off the radar for many people.

The coming of the HIV/AIDS epidemic in the early 1980s brought a marked increase in TB (Aaron *et al.*, 2004). The treatment of TB takes at least 6 months, and wherever there is a breakdown in healthcare, such as with the dissolution of the former Soviet Union, the incidence of TB increases.

In 1990 it was estimated that there were 8 million new cases of TB. In 2006, this had risen to 9.2 million, with 1.7 million people dying from the disease during the year (WHO, 2008). One third of the world is infected with latent TB infection, with around 10 % of

those expected to develop TB (Cardona and Ruiz-Manzano, 2004, WHO, 2007). In 1993 the World Health Organization declared TB a 'global emergency' (WHO, 1993).

There have been no new classes of drugs for TB developed and delivered to patients in more than thirty years (Duncan and Barry, 2004). Multidrug-resistant (MDR) strains of *Mtb* are becoming a serious concern in some countries (Espinal, 2003, Ormerod, 2005). These are defined as strains that are resistant to at least the two most effective TB drugs (isoniazid and rifampicin) (Johnson *et al.*, 2006). More recently, XDR *Mtb* strains (eXtreme or eXtensively drug resistant) have been characterised (Blower and Supervie, 2007, WHO, 2008, Zager and McNerney, 2008). Treatment of *Mtb* that is not drug resistant takes 6-8 months. There is an urgent need for new TB treatments in order to tackle drug resistance, and to lower the treatment time to two months or less (Duncan and Barry, 2004). Novel drug targets are required, hopefully leading to drugs that are not susceptible to development of bacterial resistance (Fair *et al.*, 2007).

Infection with *Mtb* usually occurs via inhalation of droplets from the atmosphere (Russell, 2007). Considering that exhaled droplets remain in the atmosphere for several hours, and the infection dose is estimated to be a single bacterium, it is perhaps not surprising that around one third of the world's human population are infected with *Mtb* (WHO, 2007). In the lungs the bacterium is surrounded by macrophages from the alveoli. This induces inflammation, leading to the recruitment of mononuclear cells from neighbouring blood vessels (Russell, 2007). These cells form a granuloma, or tubercle, which gives the disease its name. The bacterium can remain in this containment phase indefinitely, and only 5-10 % of infected individuals progress to showing symptoms of the tuberculosis disease (WHO, 2007).

A weakening of the immune system, such as from HIV, in old age, or through malnutrition, can lead to the granuloma decaying into a structureless mass of cellular debris, spilling thousands of viable, infectious bacilli into the organism. This leads to the subject showing symptoms of the disease (Idemyor, 2007, Russell, 2007).

Mtb is difficult to kill for a number of reasons (Berger and Knodel, 2003). Antibiotics are prevented from entering the cell by a dense, lipid-rich cell envelope surrounding the organism. This contains an abundance of mycolipids consisting of very long chain (C60-

C90) α -branched, β -hydroxylated fatty acids (McLean *et al.*, 2007). *Mtb* is able to enter a very slow growth phase, in which many biochemical targets are down-regulated. During this phase the bacteria shift from metabolising sugars to the β -oxidation of fatty acids. In order to cure TB, it is necessary to have one or more agents that are active against both the acute and slow growing phases. This is one reason why treatment of *Mtb* relies on a combination of several drugs.

1.3.1 *Mtb* P450s, novel targets in the fight against TB

The genome of *Mtb* was solved in 1998 (Cole *et al.*, 1998). This revealed the presence of 20 putative P450 enzymes encoded by the bacterium. P450s are known targets for azole antifungal drugs, and it was speculated that they could be useful targets in the fight against TB (Munro *et al.*, 2003, Souter *et al.*, 2000). The azoles econazole, miconazole and clotrimazole were found to inhibit the growth of *Mycobacterium smegmatis*, a relation of *Mtb*, with minimal inhibitory concentrations (MICs) lower than that for isoniazid (McLean *et al.*, 2002b). Ahmad *et al.* (Ahmad *et al.*, 2006) recently studied the effect of econazole on murine *Mtb*. They found econazole to be effective in clearing *Mtb* from the lungs of mice, reporting that it is able to replace the more traditional treatments rifampicin and isoniazid.

A number of *Mtb* P450s have been studied by our group. The first were CYP51B1, the product of gene *Rv0764c*, and CYP121, the product of *Rv2276* (McLean *et al.*, 2002b, McLean *et al.*, 2006b). CYP51B1 shows close homology to eukaryotic 14α -sterol demethylases involved in steroid hormone synthesis pathways. *Mtb* CYP51B1 was purified and shown to catalyse sterol demethylation (Bellamine *et al.*, 1999). Although *Mtb* does not have a sterol biosynthetic pathway, cholesterol was found to be necessary for *Mtb* macrophage entry (Gatfield and Pieters, 2000). A likely role for *Mtb* CYP51B1 could be in modification of host cholesterol or sterols/steroids, as opposed to its participation in a pathway for *de novo* steroid synthesis. *Mtb* CYP51B1 was shown to bind tightly to a range of azole anti-fungal drugs (Bellamine *et al.*, 1999, McLean *et al.*, 2002b).

CYP121 shows 28 % amino acid sequence identity with *Streptomyces griseolus* CYP105A1 (McLean *et al.*, 2002a), which is involved in detoxification of sulphonylurea

herbicides. It also has 29 % identity with *Saccharopolyspora erythraea* P450 eryF, which catalyses the hydroxylation of the polyketide 6-deoxyerythronolide B, in the synthesis of the antibiotic erythromycin. CYP121 was found to bind more tightly to the azole anti-fungal drugs than does CYP51B1, and to have the same order of binding affinities (K_d values) as the MIC values for these drugs for *M. smegmatis* (McLean *et al.*, 2002b). It was speculated that CYP121 may be the true target for the azoles instead of CYP51B1, although other *Mtb* P450 enzymes are also potential targets for these drugs.

While relatively little is known about the physiological and biochemical properties of the *Mtb* P450 systems, crystal structure data are now available for four of the P450s, namely CYP51B1 (Podust *et al.*, 2001), CYP121 (Leys *et al.*, 2003), CYP125 (McLean *et al.*, unpublished data) and CYP130 (Ouellet *et al.*, 2008). The CYP130 gene is absent from the genome of the less virulent, *M. bovis* BCG strain, which is used for TB vaccinations. These structures provide insights into the differing nature of the active sites in these enzymes and could be informative for drug design (Terwilliger *et al.*, 2003). CYP121 appears to be an essential gene for survival of *Mtb* (McLean *et al.*, submitted), and the (as yet uncharacterised) CYP128 was also reported to be essential for viability of *Mtb* (Sasseti *et al.*, 2001, Sasseti *et al.*, 2003). Other *Mtb* P450 systems of interest include CYP144, which was also shown to bind tightly to a variety of azole drugs (McLean *et al.*, 2006a) and CYP132, which has been implicated in virulence of *Mtb* (Recchi *et al.*, 2003). CYP132 and the characterisation of the P450 protein is the subject of Chapter 3 of this thesis.

1.4 Aim

P450s are ubiquitous haemoproteins that are found in every family of organisms (Groves, 2005). There are a wide variety of P450s, with a range of substrate specificities, from those that catalyse a single reaction on a specific substrate, to those that metabolise many substrates (McLean *et al.*, 2005). P450s have been identified as possible drug targets in the fight against *Mtb* (McLean *et al.*, 2006a, Munro *et al.*, 2003), the genome for which encodes 20 putative P450 enzymes (Cole *et al.*, 1998).

Mtb CYP132 shows significant homology at the amino acid level with various eukaryotic CYP4 family fatty acid hydroxylases. The expression of CYP132 is induced by the

protein encoded by *Rv1395* (Recchi *et al.*, 2003). Gene knockout studies revealed that when *Rv1395* is not present in *Mtb*, the growth of the bacterium is reduced (Camacho *et al.*, 1999). The link with the growth rate of the bacterium, and likelihood that it binds fatty acids, made CYP132 a good candidate for further investigation. *Mtb* CYP132 was expressed in *E. coli*, purified and characterised by a range of analytical techniques. CYP132 was found to be a fully functional cytochrome P450 enzyme. It was found to bind tightly to a range of azole drugs. Also CYP132 was found to hydroxylate unsaturated fatty acids, with possible C-C bond cleavage reactions also taking place. This is discussed in Chapter 3 of this thesis.

In order to investigate further the phenomena seen in CYP132, and other P450s, the binding of the P450 haem region to a range of ligands was investigated extensively by density functional theory (DFT) methods. This is discussed in Chapters 4 and 5 of this thesis.

Although the overall P450 fold is conserved across all P450s (Denisov *et al.*, 2005), the exact structure of the active site varies across the many P450 enzymes. The structure of the haem can also be altered, such as a distortion of $\sim 30^\circ$ at one of the four pyrrole rings observed in the crystal structure of CYP121 (Leys *et al.*, 2003). The structure of the P450 haem also changes as the axial ligand and spin state changes as the P450s progress through the catalytic cycle (Meunier *et al.*, 2004).

In order to ascertain the influence of the P450 protein environment on the haem structure, it is necessary to know the structure of the haem without the protein environment present. Also, molecular mechanics methods, such as the docking of drug molecules into P450s, rely on an accurate representation of the P450 haem structure. It was decided to carry out a comprehensive survey of the P450 haem structure and of all spin states of each species in the P450 catalytic cycle, using density functional theory (DFT) methods. From this study, conclusions could be drawn about the influence of the P450 protein environment on the haem structure, and spin states of the haem iron. This is discussed in Chapter 4 of this thesis. The theoretical methods involved will be discussed further in Chapter 2.

P450s are known targets for azole antifungal drugs (McLean *et al.*, 2002b), and were found to bind tightly to CYP132, discussed in Chapter 3 of this thesis. In order to be able to develop more effective inhibitors of P450s, it is useful to better understand the mode of binding of azole drugs to the P450 haem. DFT methods were used to study the potential energy surface as azole drugs approach the P450 haem and displace the bound water molecule. Azole binding was revealed to be a stepwise process, whereby water is first released as the sixth ligand of the haem, followed by coordination of the azole nitrogen to the haem iron. Specific elements of the CYP121 structure that enable binding of fluconazole to haem via a bridging water molecule (Seward *et al.*, 2006) were also investigated. This is discussed in Chapter 5 of this thesis.

CHAPTER 2: MATERIALS AND METHODS

2.1 *Experimental Materials*

2.1.1 Liquid media

Media for bacterial growth was made up to the specifications shown in Table 2.1. All media were sterilised by autoclaving at 120°C, 15 lb/inch² for 20 minutes prior to use.

Table 2.1: Components of the bacterial growth media used per litre of ddH₂O

	<u>Luria- Bertani (LB)</u>	<u>YT</u>	<u>Terrific Broth (TB)</u>	<u>SOC</u>
Bacto-Tryptone	10g	16g	12g	20g
Bacto-Yeast extract	5g	10g	24g	5g
Sodium Chloride	10g	5g	-	0.5g
Glycerol	-	-	4 ml	-
KH ₂ PO ₄ (0.17 M)	-	-	100 ml added to	-
K ₂ HPO ₄ (0.72 M)	-	-	900ml of the above	-
KCl (250 mM)	-	-	-	10 ml pH adjusted to 7.0 with NaOH (5 N)
MgCl ₂ (2 M)	-	-	-	5 ml
Glucose (1 M)	-	-	-	20 ml

2.1.2 Antibiotics

Antibiotics were used according to the following specifications, where appropriate.

2.1.2.1 Ampicillin

Ampicillin was made up to 50 mg/ml in distilled deionised water (ddH₂O), filter sterilised and used at 1 µl/ml in media cultures and 2 µl/ml in agar plates. It was stored at -20°C.

2.1.2.2 Carbenicillin

Carbenicillin was made up to 50 mg/ml in ddH₂O, filter sterilised and used at 1 µl/ml. It was stored at -20°C.

2.1.2.3 Rifampicin

Rifampicin was made up to 10 mg/ml in methanol (670 μ l/ml), sodium hydroxide (1M, 170 μ l/ml) and ddH₂O (160 μ l/ml). It was used at 200 μ l/ml in 5ml cultures, 100 μ l/ml in 100 ml cultures. It was stored at -20°C and replaced each week.

2.1.2.4 Chloramphenicol

Chloramphenicol was made up to 34 mg/ml in ethanol. It was used at 1 μ l/ml and stored at -20°C.

2.1.2.5 Tetracycline

Tetracycline was made up to 5 mg/ml in ddH₂O. It was used at 2.5 μ l/ml. It was stored at -20°C and replaced each day.

2.1.2.6 Kanamycin

Kanamycin was used at 30 μ g/ml, diluted from a filter sterilised 30 mg/ml stock in ddH₂O.

2.1.3 LB agar plates

Luria-Bertani (LB) media was prepared with agar (15 g/l) added. The suspension was sterilised by autoclaving at 120°C, 15 lb/inch² for 20 minutes. The dissolved solution was left to cool until comfortable to hold (~ 50°C) and the appropriate antibiotics or other additives incorporated. The solution was poured into petri dishes (25-30 ml per plate), flaming the bottle between plates to maintain sterility, and left to solidify. The plates were stored at 5°C for up to a month.

2.1.4 Buffers

All buffers were stored between 2 and 7°C and filtered prior to loading onto any column.

2.1.4.1 Tris and Tris/EDTA buffers

The pH of tris(hydroxymethyl) aminomethane (Tris) buffer is affected by temperature. It was necessary to make up Tris buffers with cold water (between 0 and 10°C) in order to ensure the pH of the buffers would not change significantly when stored and used at low temperatures.

The required amount of Tris and EDTA were made up to almost to the total required volume with cold ddH₂O. The solution was brought to the required pH using concentrated hydrochloric acid. Once at the required pH, the solution was made up to the appropriate volume with ddH₂O.

2.1.4.2 Potassium phosphate (KPi) buffers

Mono basic (KH₂PO₄), pH ~4.2, and dibasic (K₂HPO₄), pH ~9.2, potassium phosphate (KPi) buffers were made up to the required concentration and one added to the other until the required pH was reached.

2.1.4.3 MOPS buffer

3-Morpholinopropanesulphonic acid (MOPS) buffer was used at 25 mM concentration, pH 7.4 with the addition of KCl (100 mM).

2.1.5 Isopropyl-β-D-thiogalactopyranoside (IPTG)

IPTG was made up to the required concentration in ddH₂O, stored at -20°C and filter-sterilised prior to use.

2.2 Experimental Methods

2.2.1 Molecular biology

2.2.1.1 Cloning of *Rv1394c* into pET11a and pET15b

The cloning of *Rv1394c*, the gene encoding CYP132, into the expression vectors pET11a and the histidine (His)-tagged pET15b was carried out by Dr. Kirsty McLean (University of Manchester). *Rv1394c* was amplified by polymerase chain reaction (PCR) from cosmid DNA kindly supplied by Prof. Stewart Cole at the Pasteur Institute. The gene was cloned into the Promega pGEM-T plasmid for verification and sequencing. The gene was then subcloned into the expression vectors pET11a and the His-tagged pET15b, incorporating a His-tag onto the C-terminal of CYP132. The plasmids were transformed into XL Blue for verification by agarose gel analysis. All molecular biology was performed using standard protocols.

2.2.1.2 Cloning of Rv1394c into pET42a

A *NcoI* restriction site was engineered into the *Rv1394c*/pET11a plasmid using the primers: Forward: 5'-GAAGGAGATATACCCCATGGCCACCGCCACCACCC-3', Reverse: 5'-CTTCCTCTATATGGGTACCGGTGGCGGTGGTGGG-3'. The *Rv1394c* gene was excised from the *Rv1394c*/pET11a(*NcoI*) plasmid using the *NcoI*/*Bam*HI restriction enzymes and the *Rv1394c* fragment ligated into the T7lac promoter vector pET42a (Novagen Nottingham, UK) pre-digested with the same restriction enzymes. The ligation mix was transformed into *E. coli* strain novablue and plasmid preparations made from transformant colonies, leading to identification of *Rv1394c* clones by restriction digests.

A second *SpeI* restriction site was engineered into the *Rv1394c*/pET42a plasmid using the primers: Forward: 5'- CCAGATCTGGGTACTAGTGGTGGCTCCGGTATTGAGG-3', Reverse: 5'- GGTCTAGACCCATGATCACCACCGAGGCCATAACTCC-3'. The *Rv1394c*/pET42a(*SpeI*) plasmid was truncated by digestion with *SpeI* followed by ligation. The ligation mix was transformed into *E. coli* strain novablue and plasmid preparations made from transformant colonies, leading to identification of truncated *Rv1394c*/pET42a plasmids (termed *Rv1394c*/Tr-pET42a) by restriction digests. All molecular biology was performed using standard protocols.

2.2.2 Transformation of DNA into *E. coli* cells

50-100 µl of a 5 ml overnight culture of the required *E. coli* strain in LB, with the appropriate antibiotics added, was added to fresh LB (5 ml) with the appropriate antibiotics added. The culture was incubated at 37°C, shaking at 200-250 rpm until OD₆₀₀ ~0.5. 1 ml samples of the culture were pelleted by micro centrifugation at 13000 rpm and the supernatant discarded. The pellets were resuspended in calcium chloride (CaCl₂) (50 mM, 0.5 ml) at 5°C, and pelleted again at 13000 rpm, the supernatant discarded and the pellet resuspended again in CaCl₂ (50 mM, 0.5 ml). The suspension was kept on ice for 1 hour. The required DNA from a mini or midi prep was added (2-5 ng) and the suspension left on ice for a further hour. One sample was left without DNA as a control. The samples were heat shocked at 42°C for 1 minute, returned to the ice for 2 minutes and SOC media (0.8 ml) added. The cultures were incubated at 37°C with shaking for 30-60 minutes. 150 µl of each culture was added to LB agar ampicillin plates and incubated overnight at 37°C. The control containing only *E. coli* should not contain any bacterial colonies after incubation overnight.

2.2.3 Mini-prep DNA

Mini-prep DNA was prepared using the Sigma-Aldrich™ Genelute™ plasmid mini-prep kit. A 5 ml culture of the required *E. coli* strain (transformed with the relevant plasmid) in LB with the appropriate antibiotics added was incubated at 37°C with shaking at 200-250 rpm until OD₆₀₀ ~0.5. 3-5 ml of the culture were pelleted for each mini-prep and the supernatant discarded. The DNA was extracted following the procedure set out in the kit manual.

2.2.4 Midi-prep DNA

Midi-prep DNA was prepared from a 50 ml overnight culture of XL1 Blue transformant cells containing the relevant plasmid, and carried out according to the kit manual.

2.2.5 0.8 % Agarose gels

Agarose gel analysis allows the estimation of the lengths of double stranded DNA molecules in a sample mixture, by comparison with a standard containing a range of DNA fragments of known length in nucleotide base pairs (bp). The standard used in this study was NEB 1kb DNA ladder and is shown in Figure 2.1. The sizes of the pKM3b and pKM3c plasmids used were 7028 and 7087 bp, respectively.

To agarose (0.2 g) was added 40 mM Tris-acetate and 1 mM EDTA (TAE) (25 ml). The suspension was brought to the boil in a microwave oven and shaken until the solid was dissolved. The solution was allowed to cool until comfortable to hold and 10 mg/ml ethidium bromide solution (3 µl of stock) added. The solution was poured into a horizontal gel apparatus and left until solid.

Samples were made up with mini-prep DNA (10 µl, ~10 ng) and 6 X DNA loading buffer (2 µl) added. Marker was made up with New England Biolabs 1kb DNA ladder (1 µl), ddH₂O (9 µl) and 6 X DNA loading buffer (2 µl). The samples and marker were loaded onto the gel. The gel was covered with TAE and the gel run at 80V until the bands were sufficiently separated. The gels were viewed under UV light.

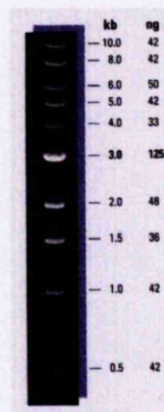


Figure 2.1: NEB comparison chart showing the pattern and range of sizes of DNA (kb, kilobase pairs) in the 1kb DNA ladder, elucidated by electrophoresis on a 0.8 % agarose, ethidium bromide-stained gel

2.2.6 Verifying plasmid transformation

To verify that the plasmid had been transformed into the *E. coli* cell strain effectively, the mini-prep DNA was digested with the restriction enzyme *AvrII*, and with a mixture of *NdeI/BamHI*. *AvrII* cuts in the centre of *Rv1394c*, giving a single linearised segment of DNA, which is likely to run true to size according to the linear marker fragments used. By contrast, circular (uncut) plasmid can adopt a range of conformations and its length is difficult to determine in a similar experiment. *BamHI/NdeI* restriction enzymes cut at either end of *Rv1394c*, separating the gene from the pET vector. This gives fragments of 5642 and 1386 bp for pKM3b, and 5701 and 1386 bp for pKM3c.

Mini-prep DNA was first run on a horizontal agarose gel to verify that the sample contained DNA at a sufficient concentration to obtain a satisfactory digestion and fragment resolution with the amounts used. To a 1.5 ml Eppendorf tube was added mini-prep DNA (12.5 μ l), NEB Buffer 2 (1.5 μ l), and *AvrII* (1 μ l). To a second Eppendorf was added mini-prep DNA (10 μ l), NEB buffer Bam (1.5 μ l), BSA (1.5 μ l), *NdeI* (1 μ l) and *BamHI* (1 μ l). Final enzyme concentrations used were typically 1-2 units of enzyme activity. The tubes were incubated at 37°C for at least 1 hour. The digested DNA was then run on a horizontal agarose gel and the bands obtained compared with that expected for the DNA used, and with reference to a set of marker fragments.

2.2.7 Expression of CYP132

To sterile bacterial growth media (600 ml) in 2 L conical flasks were added 6 ml aliquots of a 100 ml overnight culture of the appropriate *E. coli* strain containing the pKM3b or pKM3c plasmid, as appropriate. The appropriate antibiotics were added, typically only ampicillin (50 mg/ml, 600 μ l) was used in order to select for cells containing the cloned plasmid. The cultures were shaken at 37°C, 210-240 rpm until OD₆₀₀ ~0.7. The temperature was lowered to 25°C or 18°C, according to the requirements for gene expression. At OD₆₀₀ ~0.8, the temperature was lowered further to 18°C. At OD₆₀₀ ~0.9 the cultures were induced with the appropriate concentration of IPTG. 1 ml samples were taken after induction and at appropriate time intervals throughout expression. These samples were pelleted at 13000 rpm, the supernatant discarded and the pellets stored at -20°C for further analysis. The total culture growth time was approximately 24 hours.

Bacterial cells from the cultures were pelleted by centrifugation at 5000-7000 rpm for 15-30 minutes and the supernatant discarded. The pellets were resuspended in Tris (50 mM), EDTA (1 mM), pH 7.2 buffer (Buffer A), and re-pelleted at 5000-7000 rpm for 20-30 minutes. The pellets were stored at -20°C.

2.2.8 Plasmid stability test

In order to confirm that the plasmid was not expelled from the cells during expression, a plasmid stability test was performed. Two LB agar plates containing each of the following were prepared.

2 x LB only

2 x LB containing ampicillin only

2 x LB containing IPTG (1 mM) only

2 x LB containing both IPTG (1 mM) and ampicillin

A 1 ml sample was taken from the cell culture immediately prior to inducing with IPTG. The sample was diluted by a factor of 10⁵ and 150 μ l aliquots added to one plate of each type. As a control, a sample of ddH₂O was diluted in the same way and 150 μ l aliquots added to the control plates. The plates were incubated at 37 °C overnight. If the plasmid

is unstable (due to expression of a toxic gene product), there should be growth on the plates containing LB only and LB containing ampicillin, but possibly not on the plates containing IPTG. A reduction in growth on the plate containing ampicillin alone (compared to plasmid-free cells and perhaps compared to LB alone) might also suggest some cells have lost the plasmid due to leaky expression of a toxic gene.

2.2.9 SDS PAGE gel

The expression and purity of the expressed protein was determined by Tris-glycine SDS-Polyacrylamide Gel Electrophoresis (SDS-PAGE) analysis. Proteins migrate through the gel at varying rates when an electric field is applied, depending on their mass and their folding state. Comparison with a 'Protein marker' containing proteins of known mass gives an indication of the mass of the proteins in the mixture being analysed. The range of protein masses in the marker used in this study is shown in Figure 2.2. The expected mass of CYP132 is 52.2 kDa.

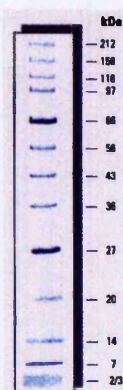


Figure 2.2: NEB comparison chart showing the pattern and masses in kDa of the proteins in the protein marker used in this study, resolved on a 10 % SDS PAGE gel with Coomassie blue staining.

A 12 % resolving gel containing ddH₂O (3.3 ml), a solution of acrylamide (29 % w/v) and N,N'-methylenebisacrylamide (1 % w/v) (4.0 ml), 1.5 M Tris (pH 8.8, 2.5 ml), 10 % (w/v) SDS (0.1 ml), N,N,N',N'-tetramethylethylenediamine (TEMED) (4 µl) and 10 % (w/v) ammonium persulphate (0.1 ml) was aliquoted into the pre-prepared gel plates to a depth of two thirds, ethanol was added and the gel allowed to set. The ethanol was then poured off and a 5 % stacking gel containing ddH₂O (3.4 ml), 30 % acrylamide mix (0.83

ml), 1.0 M Tris (pH 6.8, 0.63 ml), 10 % SDS (0.05 ml), TEMED (0.005 ml) and 10 % ammonium persulphate (0.05 ml) added and allowed to set, with a set of wells incorporated to allow sample loading. The wells were washed with water and the apparatus set up using 25 mM Tris, 250 mM glycine, 0.1 % SDS in ddH₂O, pH 8.3 as running buffer.

The pellets taken during expression were resuspended in tris (50 mM) EDTA (1 mM) (Buffer A) (150 µl). 20 µl aliquots were removed and 10 µl of a solution of NEB 3x SDS sample buffer:DTT, 9:1 added. Marker was prepared containing NEB protein marker (24 µl), NEB 3xSDS sample marker/DTT (22 µl) and Buffer A (20 µl). The samples and marker were boiled at 95 °C for 3 minutes. Thereafter, 15 µl aliquots of each sample were loaded onto the gel. The gel was run at 150 V until the leading marker dye band reached the resolving gel. The voltage was then increased to 200 V until the leading marker band reached the bottom of the plates.

The protein was visualised by staining with Coomassie Brilliant Blue stain (1.25 g per 90 ml methanol:H₂O 1:1 v/v, 10 ml acetic acid) for 5 minutes before destaining with H₂O (85 %), methanol (10 %), acetic acid (5 %) overnight. The gels were washed in a drying solution of methanol (40 %), glycerol (10 %), acetic acid (7.5 %) and dried between Promega™ Gel Drying film.

2.2.10 Determination of amount of soluble protein by SDS PAGE analysis

Pellets taken during expression were resuspended in Buffer A (150 µl) and 20 µl aliquots removed and labelled 'total protein'. The remaining sample was pelleted by centrifuging at 13000 rpm and the supernatant discarded. The pellet was resuspended in Bugbuster™ and the tubes shaken at room temperature for 20 minutes before being centrifuged at 13000 rpm at 4 °C for 25 minutes. The supernatant was removed to a fresh tube, 20 µl were taken and labelled 'soluble protein', and the pellet was resuspended in Buffer A (150 µl). Thereafter, 20 µl were removed and labelled 'insoluble protein'. NEB 3xSDS sample marker/DTT (10 µl) was added to each tube and the samples run on a 12 % SDS PAGE gel as above.

2.2.11 Purification of CYP132

2.2.11.1 Protein extraction

The frozen pellets were resuspended in ice cold Buffer A unless otherwise stated. Glycerol ~5 %, phenylmethylsulphonyl fluoride (PMSF) (~200 µl to a final concentration of ~1 mM) and benzamidine hydrochloride (~1 spatula full, to a final concentration of ~2 mM) was added. The suspension was kept on ice and sonicated using a Bandelin Sonoplus GM2600™ for 10 seconds at 50 % power and for ten times over 1-2 hours. The suspension was centrifuged at 18000 rpm at 4 °C for 15-30 minutes. The supernatant was removed and centrifuged again until the supernatant was visibly free of solid contaminants. Any semi-solid portion after centrifugation was stored at -20°C and added to the next batch of pellets to be sonicated.

2.2.11.2 Ammonium sulphate purification

Initial purification of the protein could be performed by the step-wise addition of ammonium sulphate. Different proteins precipitate out of solution at different concentrations of ammonium sulphate. The appropriate amount of ammonium sulphate was added slowly as solid with stirring of the cell lysate at 0 °C to take the concentration first from 0 to 25 %, and then from 25 to 40 %. The resulting suspension was centrifuged at 18000 rpm between each addition to the percentage indicated, to separate the precipitated protein from the supernatant. The P450 content in the pellet and supernatant was determined by UV/Vis spectrometry. If necessary, the concentration was raised further from 40 to 60 % and beyond. Fractions (solid, resuspended in a small volume of Buffer A, or liquid from supernatant) containing P450 were loaded onto a Phenyl Sepharose column, equilibrated in 1.5 M ammonium sulphate in Buffer A. The protein was washed off the column using a decreasing concentration gradient of ammonium sulphate in the same buffer (typically a 500 ml gradient was applied).

2.2.11.3 Column purification

The protein was purified using a series of ion exchange and other affinity columns. The column resins used were diethylaminoethyl (DEAE)-Sepharose, Q-Sepharose, hydroxyapatite, nickel and mono-Q. The running buffer used with the DEAE, Q-Sepharose and mono-Q resins was buffer A, with a salt gradient of KCl in buffer A, up to 500 mM KCl. The P450 typically eluted from the DEAE column resin without the need for

KCl (i.e. through continuous washing with buffer A), and with KCl concentrations of 50-100 mM being required for the elution of the P450 from the Q-Sepharose and mono-Q column resins. The running buffer used with the hydroxyapatite and nickel columns was KPi (25 mM, pH 6.5), with an elution gradient of 25 – 500 mM KPi, pH6.5. The P450 typically eluted slowly from the hydroxyapatite through continuous washing with the 25 mM running buffer, otherwise a KPi concentration of 25 – 75 mM was required. No binding of the P450 to the nickel column was detected.

The column was first equilibrated with a minimum of two column volumes of the appropriate buffer. The protein was then loaded onto the column dissolved in the same buffer, and if it bound sufficiently (easily observed due to the red colour of the protein) was washed with a further 100 ml of the buffer to remove other weakly-bound and non-bound proteins. A concentration gradient was used as necessary (e.g. increasing salt to displace protein from the ion exchange resins) to wash the protein off the column. The presence and quantity of the desired protein in each fraction was determined by UV/Vis spectrometry.

The fractions containing the P450 protein were concentrated using Amicon Series 8000 stirred cells under nitrogen and Centriprep® Centrifugal Filter Units, depending on the volume to be concentrated. The concentrated protein was decanted into dialysis tubing and dialysed into the appropriate solvent ready to be loaded onto the next column. Protease inhibitors (as described above) were maintained in all solutions. The process was repeated with the appropriate columns until the protein had reached the required purity. This was determined from the ratio between the UV/Vis absorption at 416 nm and 280 nm, reflecting the heme-specific absorption by comparison to total protein absorption.

Once the protein had reached an acceptable level of purity, it was dialysed into Tris (10 mM), glycerol (50 %) buffer and stored at -80°C.

2.2.12 Expression of *E. coli* flavodoxin reductase

E. coli flavodoxin reductase (FLDR) was prepared as described by McIver *et al.* (McIver *et al.*, 1998). To sterile LB (600 ml) were added 6 ml aliquots of a 100 ml overnight culture of HMS174 (DE3) containing the pCL21 plasmid encoding the FLDR protein.

Ampicillin (50 mg/ml, 600 μ l) was added and the cultures were shaken at 37 °C, 240 rpm until the $OD_{600} = 1$, at which point *FLDR* gene expression in the cultures was induced with IPTG (100 μ M), and cultures allowed to grow for a further 3 hours.

The cultures were pelleted by centrifugation and the supernatant discarded. The pellets were resuspended in sodium phosphate buffer (10 mM, pH 7.5), and re-pelleted. The pellets were stored at -20°C.

2.2.13 Purification of *E. coli* flavodoxin reductase

The frozen cell pellet was resuspended in ice cold sodium phosphate buffer (10 mM, pH 7.5). Glycerol (~5 %), phenylmethylsulphonyl fluoride (PMSF) (~200 μ l to a final concentration of 1 mM) and benzamidine hydrochloride (~2 mM) was added. The suspension was kept on ice and sonicated using a Bandelin Sonoplus GM2600™ on 50 % power for 30 seconds every minute for 5 minutes. The remaining solid debris was removed by centrifugation at 18000 rpm and the solute purified by Q-Sepharose and hydroxyapatite column purification, as described previously, using a salt gradient of KCl (0-500 mM) in sodium phosphate buffer (10 mM, pH 7.5) with Q-Sepharose, and a concentration gradient of sodium phosphate (10-500 mM, pH 7.5) with hydroxyapatite. The pure protein was stored at -80°C in Tris (10 mM), glycerol (50 %) buffer.

2.2.14 UV/Vis spectrometry

(UV/Vis) spectroscopy was carried out on a Cary UV-50 UV-Visible spectrophotometer (Varian) over a wavelength range of 250-800 nm using a quartz cell of path length 1 cm.

2.2.15 Determination of extinction coefficient of CYP132 using pyridine haemochromogen method

The determination of the extinction coefficient of CYP132 was carried out as described by Berry and Trumpower (Berry and Trumpower, 1987). The UV/Vis spectrum of a solution of pyridine (200 μ l), NaOH (100 mM) and $K_3Fe(CN)_6$ (300 μ M) made up to 1 ml in water was used as a baseline. CYP132 (3 μ l, ~0.8 mM) was added and the UV/Vis spectrum measured. Sodium dithionite (2-5 mg) was added and the UV/Vis spectrum for the reduced heme measured. A difference spectrum was calculated between the two spectra. Using the known coefficient between the peaks at 557 and 540 nm in the

difference spectrum (22.1 mM^{-1}), the extinction coefficient of CYP132 at 416 nm was calculated as $92 \text{ mM}^{-1} \text{ cm}^{-1}$.

2.2.16 Mass spectrometry of CYP132

The mass spectrum of CYP132 was measured by electrospray time of flight mass spectrometry. Data were collected at the Biosciences facility at the University of Birmingham.

2.2.17 Binding titrations

To a solution of the protein (typically $\sim 3 \mu\text{M}$, 1 ml) in 10 mM Tris, pH 7.0 buffer were added aliquots of a known concentration and volume of the ligand in question, and the UV/Vis spectrum of the solution recorded after each addition. Where solids dissolved in organic solvents were used, it was necessary to adjust the concentration of the ligand solution so that maximum optical shift of the P450 haem was achieved with between 5 and $10 \mu\text{l}$, in order to avoid denaturation of the protein and interference with the results.

After correcting for dilution and baseline effects as appropriate, difference spectra between the protein alone and the protein with ligand were created by subtraction of the spectrum for the ligand-free enzyme from those for each enzyme sample with ligand bound. For each set of difference spectra so generated, a peak and trough value were identified, and the maximum difference induced by ligand addition in each case determined by the total absorption difference between these two wavelengths. A graph of the ligand-induced absorption against ligand concentration was plotted and the curve fitted iteratively to an appropriate binding equation using MicrocalTM Origin® software, to give a value of the dissociation constant (K_d) for the binding of the ligand to the protein.

2.2.18 Electron Paramagnetic Resonance (EPR) spectroscopy

Electron Paramagnetic Resonance (EPR) spectroscopy was carried out by Dr. Harriet Seward (University of Leicester) at the University of East Anglia on a Bruker Spectrospin spectrometer. Spectra were recorded at 9.67 GHz, at a temperature of 10.5 K.

2.2.19 Circular Dichroism (CD)

Circular dichroism (CD) spectra were recorded at 20 °C on a JASCO J600 spectropolarimeter. Far UV CD spectra were recorded over the wavelength range 180–260 nm in a quartz cell of 0.02-cm pathlength.

In order to assess protein susceptibility to denaturation by guanidinium chloride (GdmCl), CYP132 (8 μ M) was incubated in solutions of GdmCl (0–6 M) in KPi (50 mM, pH 7.0) at 37 °C for 15 minutes prior to CD analysis. A graph of the CD against concentration of GdmCl was plotted.

2.2.20 Turnover kinetics

Turnover kinetics were measured on a Cary UV-50 UV-Visible spectrophotometer (Varian) in kinetics mode, by calculating the rate of change of the absorbance at 340 nm, corresponding to oxidation of NADPH as it was converted into NADP⁺, and using an extinction coefficient of $\Delta\epsilon_{340} = 6.21 \text{ mM}^{-1} \text{ cm}^{-1}$. Data were collected using suspensions of different forms of the CYP132 protein, with or without potential fatty acid substrates, and in the presence of different redox partner mixtures that would transfer electrons to the P450 from NADPH (see Chapter 3 *Results* section for full details).

2.2.21 Mass spectrometry of turnover products

A solution of CYP132 (6.67 μ M), *E. coli* flavodoxin (FLD, 1.65 μ M), *E. coli* flavodoxin reductase (FLDR, 1.65 μ M), fatty acid (66.7 μ M), NADPH (1 mM), glucose dehydrogenase (7.5 units), glucose (20 mM) in MOPS buffer (25 mM, pH 7.4) was stirred at 30 °C overnight. The solution was acidified with HCl and extracted twice with dichloromethane (DCM). The organic layers were combined, dried over MgSO₄ and evaporated to dryness under vacuum. The dry samples were stored in the dark at -80°C.

The extracted product was dissolved in methanol and the mass spectrum measured on an Agilent 1100 series HPLC/MSD ion trap mass spectrometer. An ACE 5A, 250 x 4.6 mm column and a gradient of methanol (A) and acetic acid (0.2 %) in water (B) was applied, as described in (Okita *et al.*, 1991). Saturated fatty acids were run at 62 % A, 38 % B for 18 minutes, followed by 100 % B for 12 minutes. Unsaturated fatty acids were run at 66 % A, 34 % B for 28 minutes, followed by 100 % B for 22 minutes.

2.3 Theoretical Methods

There are a range of methods available for the computational modelling of atoms, molecules, proteins and other systems in order to make theoretical predictions about real life processes. As a general rule, the computational time required increases greatly as the accuracy of the method increases (Cramer, 2004). The accuracy required from a calculation must always be balanced against the computational resources available. Broadly speaking, computational methods can be divided into two categories, quantum mechanics and molecular mechanics. Quantum mechanics methods consider both the electrons and nuclei of a system, while molecular mechanics consider the interaction of the atoms as a whole.

2.3.1 Quantum mechanics methods

Quantum mechanics methods seek to solve the time independent Schrödinger equation (Nelson, 1966, Schrödinger, 1926):

$$H\Psi = E\Psi \quad (2.1)$$

Where H is the Hamiltonian operator, Ψ is the wavefunction and E is the energy. The exact solution of the Schrödinger equation can only be solved for single electron atoms and diatomics. The mathematical equations behind how quantum mechanics seeks to come to a solution to the Schrödinger equation for a multi-electron system become very involved. The mathematics behind quantum mechanics methods can not be meaningfully presented without presenting the whole progression of the arguments. This would result in large volumes of mathematical equations. In order to avoid becoming bogged down in voluminous equations, only the theory behind the mathematics, and the applications it brings are discussed here. A clearly presented review of quantum mechanics methods and the mathematics involved is presented in (Leach, 2001) and (Lewars, 2003). A more in depth discussion may be found in (Cramer, 2004).

The typical form of the Hamiltonian operator takes into account five contributions to the total energy of a system - the kinetic energies of the electrons, the kinetic energies of the nuclei, the attraction of the electrons to the nuclei, and the interelectronic and internuclear repulsions, disregarding external influences such as external electronic or

magnetic fields (Kato, 1951, Leach, 2001). The Born-Oppenheimer approximation decouples the motion of the electrons from that of the nuclei (Born and Oppenheimer, 1927, Kiselev, 1970). This allows the Schrödinger equation to be solved for the electrons alone, in the field of the nuclei, for each arrangement of the nuclei.

The total electronic energy of the system is made up of the kinetic energy due to the movement of the protons and electrons, the potential energy resulting from the electrostatic interactions (proton-proton, proton-electron, electron-electron), and the 'exchange interaction' (Leach, 2001). The exchange interaction is a quantum effect, and refers to the finite probability of two electrons of opposite spin occupying the same space at the same time, whereas due to the Pauli exclusion principle, two electrons of equal spin can not occupy the same space at the same time. As a result electrons of the same spin tend to 'avoid' each other, giving rise to a lower Coulombic repulsion, and thus a lower energy.

2.3.1.1 Hartree-Fock (HF)

Hartree-Fock (HF) methods form the basis of most other quantum mechanics methods (Dykstra *et al.*, 2005, Fock, 1930, Hartree, 1928). These calculate the Coulombic interaction of each electron with the average field of all the other electrons. Solving the Schrödinger equation for any one electron thus has an effect on the solution for all of the other electrons in the system. In order to solve the wavefunction of the whole system, the 'self consistent field' (SCF) approach is employed (Leach, 2001). This makes use of the 'variation theorem', which states that the energy from an approximation to the true wavefunction will always be greater than the true energy; thus the better the wavefunction, the lower the energy. The HF equations are solved iteratively until progressive iterations produce no change for any of the electrons. A set of trial solutions to the HF eigenvalue equations are first obtained. These are used to calculate the Coulomb and exchange operators. The HF equations are then solved, giving a second set of solutions, which are used in the next iteration, and so on.

In the HF equations the energy has the form (Frisch *et al.*, 2003):

$$E_{\text{HF}} = V + \langle hP \rangle + 1/2 \langle PJ(P) \rangle - 1/2 \langle PK(P) \rangle \quad (2.2)$$

Where V is the nuclear repulsion energy, $\langle hP \rangle$ is the one-electron (kinetic plus potential) energy, $1/2\langle PJ(P) \rangle$ the coulomb repulsion of the electrons and $1/2\langle PK(P) \rangle$ the exchange energy.

2.3.1.2 Basis functions

It is not practical to solve exactly the HF equations for molecules. The most popular solution to this problem is to employ the Linear Combination of Atomic Orbitals (LCAO) (Lewars, 2003). Using LCAO, each molecular orbital is written as a combination of single electron orbitals. The single electron orbitals, called basis functions, often correspond to atomic orbitals, e.g. of s, p or d character. N basis functions would give rise to N molecular orbitals. The smallest number of basis functions that can represent a molecular system is that which can just accommodate all of the electrons in the molecule. More sophisticated calculations use more basis functions than this minimal set, and give rise to lower energies, which, according to the variation theorem, will be closer to the true energy of the system. At the HF limit, the energy of the system can no longer be reduced by the addition of more basis functions.

Quantum mechanics programs typically employ sets of basis functions, called basis sets, to model the electronic orbitals. The Slater type atomic orbitals (Slater, 1930), such as the 1s, 2s, 2p, 3d orbitals etc would be an obvious choice as basis functions to model the molecular orbitals. Unfortunately the mathematics required for the combination of Slater orbitals is often prohibitively complicated. The product of two gaussian functions, of the type $\exp(-\alpha r^2)$, is another gaussian (Leach, 2001). A number of gaussian functions can be combined in order to closely model a Slater orbital. The minimum number of gaussians required to adequately represent a Slater orbital is three.

The minimal basis set STO-3G employs just the Slater orbitals required to fill the outer atomic shell, i.e. 1s for hydrogen and helium, 1s, 2s, and 2p for lithium - neon, each represented by three gaussians (Leach, 2001). Minimal basis sets can represent molecular geometries fairly well, but encounter difficulties, especially with elements at the end of periods, such as oxygen or fluorine, as these are represented by the same number of functions as lithium or beryllium. Also, minimal basis sets can not describe non-spherical aspects of the electronic distribution.

Double zeta basis sets employ twice the number of basis functions than the minimal basis set. A linear combination of a contracted function and a diffuse function gives an overall result that is intermediate between the two, and that can vary depending on the requirements of the environment. The number of gaussians used to describe each Slater orbital can also be increased. The 6-31G basis represents the core orbitals with six gaussians, while the valence orbitals are described by three contracted gaussians and one diffuse gaussian.

The charge distribution about an atom in a molecule may be perturbed in comparison with the isolated atom. In order to represent this, polarisation functions may be employed. These use an orbital with higher angular momentum quantum number, such as d orbitals mixing with p orbitals, to model polarisation of the orbitals. Polarisation functions are denoted by a *, such as 6-31G*, referring to polarisation functions on the heavy atoms, and 6-31G** referring to polarisation functions on the hydrogen atoms also. In order to model systems with lone pairs, with electron density away from the nucleus, diffuse functions can also be added. These are denoted by a "+" symbol.

2.3.1.3 Electron correlation

A major flaw with HF methods is that they do not take into account electron correlation (Cramer, 2004). HF calculates the energy of each electron relative to the average of all the other electrons. This means the instantaneous position of each electron is not influenced by the position of a neighbouring electron. Electron correlation means electron motions are correlated, and they avoid each other more than HF would suggest. This gives rise to lower energies in reality than HF calculates.

There are various methods based on HF methods that take electron correlation into account (Leach, 2001). HF is an *ab initio* method, meaning it calculates the energies of all electrons in the system. There are various post-HF methods, which are advanced *ab initio* methods that build on HF methods and take electron correlation into account. These include configuration interaction and many-body perturbation theory. Semi-empirical methods use parameters to model the core electrons and *ab initio* methods to model the valence electrons. The parameters used also take into account electron correlation.

2.3.2 Density Functional Theory

Density Functional Theory (DFT) is an approach to the electronic structure of atoms and molecules that has seen a dramatic surge of interest since the late 1980s (Hohenberg and Kohn, 1964, Kohn and Sham, 1965, Leach, 2001). DFT depends on the relationship between the total electronic energy and the overall electronic density. HF expresses the wavefunction as a Slater determinant, constructed from a set of N single-electron wavefunctions. DFT also considers single-electron functions. However, instead of calculating the full N -electron wavefunction, DFT only calculates the total electronic energy and the overall electronic density distribution.

The energy in DFT calculations takes the form (Frisch *et al.*, 2003):

$$E_{KS} = V + \langle hP \rangle + 1/2 \langle PJ(P) \rangle + E_x[P] + E_c[P] \quad (2.3)$$

Where V is the nuclear repulsion energy, $\langle hP \rangle$ is the one-electron (kinetic plus potential) energy, $1/2 \langle PJ(P) \rangle$ the coulomb repulsion of the electrons, $E_x[P]$ is the exchange functional, and $E_c[P]$ is the correlation functional. Thus, HF can be considered a special case of DFT, with $E_x[P]$ given by the exchange integral $1/2 \langle PK(P) \rangle$ and $E_c=0$.

In order to model the system accurately, DFT must account for the electronic exchange and correlation in the system. One method of modelling the exchange and correlation is to use 'gradient corrections'. These depend on the gradient of the electron density at each point in space rather than just on the density value. A variety of gradient corrections have been proposed in the literature. These are typically divided into separate exchange and correlation contributions. A combination of the exchange functional proposed by Becke (Becke, 1993a) and the Lee-Yang-Parr correlation functional (Lee *et al.*, 1988) (BLYP) is commonly used.

Hybrid HF-DFT methods can also be used. DFT incorporates the electron correlation from the beginning, whereas HF methods give an exact means of treating the electron exchange. Becke proposed a method of combining DFT and HF methods (Becke, 1993a, Becke, 1993b). This is combined with the Lee-Yang-Parr correlation functional, giving the B3LYP method.

2.3.2.1 Open and closed shell systems

Systems containing unpaired electrons may require different treatment to systems where the electrons are paired. Restricted HF (RHF) methods model the system by a combination of singly and doubly occupied molecular orbitals. Unrestricted HF (UHF) methods use two distinct sets of molecular orbitals for the α and β orbitals, resulting in molecular orbitals that are not exactly equivalent in energy.

2.3.2.2 Basis Set Superposition Error (BSSE)

If the energies of two molecules are compared when they are far apart, and when they are close together, this can result in an overestimation of the resultant stabilisation energy (Cramer, 2004). This is because, as well as being represented by their own basis sets, each molecule overlaps with the basis functions of the other molecule. This is called the basis set superposition error (BSSE). The size of the BSSE can be calculated by calculating the energy of each molecule in the presence of the basis sets for both molecules.

2.3.3 Potential Energy Surfaces

The potential energy surface (PES) is a central concept in computational chemistry (Lewars, 2003). A PES is the relationship, presented either graphically or mathematically, between the energy of a molecule, or collection of molecules, and its geometry. If the atoms in a system are fixed in a given position, they will possess no kinetic energy, but instead their energy will be entirely potential energy. In reality, due to the Heisenberg uncertainty principle (Heisenberg, 1927), an atom can never be completely fixed in space, as this would require it to have both fixed energy, and position in space. Instead, atoms always have some vibrational energy, above the potential energy. The minimum energy a system can have is called the zero point energy (ZPE) (Cramer, 2004).

PESs are expressed as the potential energy of a system relative to one or more geometric parameters (Agmon and Rabinovich, 1992). PESs relative to one geometric parameter, such as a single bond length, called a 1 dimensional (1D) PES, or relative to two geometric parameters, can be presented graphically on a 2D or 3D graph, respectively. An example 1D PES is shown in Figure 2.3. 1D PESs were produced for

the binding of azoles to water bound haem, in Chapter 5 of this thesis, relative to the Fe-N_{azole} distance. In this case, the other geometric parameters were optimised at each Fe-N_{azole} distance studied. This is referred to as a relaxed PES. If the rest of the geometry had remained fixed, this would be an unrelaxed, or rigid PES (Lewars, 2003). 2D PESs for the binding of azoles to haem were also produced in Chapter 5, fixing the Fe-N_{azole} and Fe-O_{water} bond distances. PESs relative to more than two geometric parameters can not be represented graphically, as it is not possible to represent more than two dimensions on a 3D graph. PESs with more than two dimensions can be represented mathematically, however. Such a PES is referred to as a hypersurface (Agmon and Rabinovich, 1992).

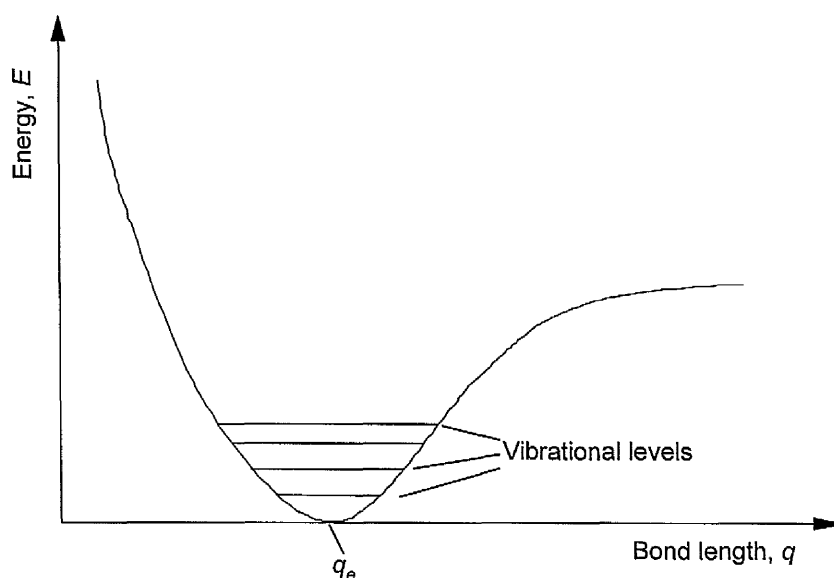


Figure 2.3: 1D PES for a diatomic molecule, or larger system relative to a single geometric parameter. The potential energy increases if the bond length q is increased or decreased from the equilibrium value q_e . In reality, atoms can never be fixed in space and the molecule will have some vibrational energy, occupying one of the vibrational levels, shown above the PES.

Some of the main tasks of computational chemistry, including all of the calculations described in this thesis, are concerned with the structure and energy of molecules, and the transition states of chemical reactions (Senn and Thiel, 2007). The equilibrium structure and energy of molecules (neglecting vibrations), and the transition states of

chemical reactions, occur at stationary points on the PES. A stationary point has a gradient of 0, and is defined mathematically as a point at which the first derivative of the potential energy, E with respect to each geometric parameter, q is 0.

$$\frac{\partial E}{\partial q_1} = \frac{\partial E}{\partial q_2} = \dots = 0 \quad (2.4)$$

Stationary points at the bottom of an energy well are minima, whereas transition states occur at a point where the potential energy, with respect to at least one geometric parameter, is a maximum. The lowest energy pathway linking two minima, via a transition state, is referred to as the reaction coordinate, or intrinsic reaction coordinate (IRC).

Geometry optimisation is the procedure by which a stationary point on the PES is located. Optimisation to a minimum is referred to as minimisation. A program such as Gaussian 03 carries out geometry optimisation on an inputted starting structure, believed to be close to the stationary point, using an algorithm that utilises the first and second derivatives of the potential energy with respect to each geometric parameter.

The zero point energy of a molecule at a stationary point can be calculated by calculating the normal mode frequencies (Floquet *et al.*, 2006). These are the simplest vibrations of the molecule, which combine to form the complex vibrations of the real molecule. A non-linear molecule with n atoms has $3n - 6$ normal modes (each atom can move in the x , y , or z direction, giving $3n$ normal modes, the six modes relating to translation or rotation of the whole molecule are neglected) and a linear molecule $3n - 5$ (only two rotational modes of the whole molecule are geometrically relevant).

The frequency, ν , of a normal mode is given by:

$$\nu = \frac{1}{2\pi c} \left(\frac{k}{\mu} \right)^{1/2} \quad (2.5)$$

where c is the velocity of light, k the force constant of the vibration, and μ the reduced mass of the molecule.

The force constant of each normal mode depends on the second derivative of the energy with respect to the change in geometry (Lewars, 2003). At a minimum, all force constants are positive, as the gradient of the PES increases in every direction. At the transition stage, however, one of the normal modes takes the system along the reaction coordinate, towards the reactants in one direction and the products in the other. This results in a decrease in the gradient of the PES, and thus a negative second derivative, and negative force constant. The square root of a negative number is an imaginary number. Thus, a transition state is characterised by having one normal mode with an imaginary frequency.

2.3.4 Theoretical methods used in this study

2.3.4.1 DFT methods

The software used for the DFT calculations in this thesis was Gaussian 03 (Frisch *et al.*, 2004). This incorporates all of the theoretical methods discussed above. Optimisation calculations were carried out to minima and transition states. These used the Berny algorithm, with redundant internal coordinates (Peng *et al.*, 1996, Reed and Weinhold, 1983). Transition states were confirmed by normal mode (frequency) calculations, with a single imaginary frequency confirming that a transition state had been reached. For the DFT calculations, the BP86 (Becke, 1988, Perdew, 1986) and B3LYP (Becke, 1993a, Becke, 1993b, Lee *et al.*, 1988) functionals were used.

Basis set superposition error was calculated using counterpoise calculations (Boys and Bernardi, 1970, Simon *et al.*, 1996). Implicit solvent models were calculated using the polarisable continuum model (PCM) (Cances *et al.*, 1997, Cossi *et al.*, 1998, Mennucci and Tomasi, 1997).

2.3.4.2 Basis sets

The following basis sets were used in this study:

BS1 - LANL2DZ, with pseudo potentials (Fe, S), 6-31G* (O, N), 6-31g (C, H)

BS2 - LANL2DZ, with pseudo potentials (Fe, S), 6-31G** (O, N, C, H)

BS3 - LANL2DZ, with pseudo potentials (Fe), 6-31G** (S, O, N, C, H)

BS4 - LACVP, with pseudo potentials (Fe), 6-31G** (S, O, N, C, H)

BS5 - 6-31G(d,f), no pseudo potentials (Fe, S, O, N, C, H)

BS6 - 6-31G*, no pseudo potentials (Fe, S, O, N, C, H)

BS7 - LANL2DZ, with pseudo potentials (Fe), 3-21G (S, O, N, C, H)

The basis set used for the majority of the calculations was BS1. This was based on the basis set used by Filatov *et al.* (Filatov *et al.*, 1999), who used 6-31G* for the coordination of iron, and STO-3G for the carbon and hydrogen porphyrin atoms. In order to determine the optimum basis set, the fatty acid structures lauric acid and myristic acid were optimised with the basis sets STO-3G, 3-21G, 6-31G, 6-31G* and 6-31G**. A significant difference in energy was observed between the STO-3G, 3-21G and 6-31G basis sets. Little difference was found between the optimised energies with the 6-31G* and 6-31G** basis sets, whereas the time required for the calculations was increased significantly. The 6-31G basis set was used for the porphyrin carbon and hydrogen atoms, and the 6-31G* basis used for the iron coordination sphere (nitrogen and oxygen atoms). The LANL2DZ and LACVP basis sets (Hay and Wadt, 1985a, Hay and Wadt, 1985b, Wadt and Hay, 1985) have been used extensively for the haem iron in the literature (Shaik *et al.*, 2005a), and these were employed for iron and sulphur in this study.

2.3.4.3 The haem model

The P450 haem was modeled by a ferrous/ferric protoporphyrin IX molecule. The cysteine ligand to the haem was modeled either by SH or SCH₃, as described in the text. This model has been used by other groups to model the P450 haem (Ogliaro *et al.*, 2000c, Schöneboom and Thiel, 2004). Ogliaro *et al.* report that a SH group gives the best representation of the cysteine ligand in gas phase DFT calculations.

CHAPTER 3: THE EXPRESSION AND CHARACTERISATION OF *MYCOBACTERIUM TUBERCULOSIS* CYP132

3.1 Introduction

3.1.1 *Mycobacterium tuberculosis*

As was discussed in Chapter 1, there is an urgent need for new treatments for *Mycobacterium tuberculosis* (*Mtb*). In 2006 there were 9.2 million new cases and 1.7 million deaths from TB worldwide (WHO, 2008). Of these, 0.2 million were in HIV positive people. Multidrug-resistant strains of *Mtb* have developed and are a considerable problem in some countries, particularly in Asia and Africa (Espinal, 2003). Where drug resistance is not present, treatment of TB takes 6-8 months.

Around one third of the world's human population are infected with *Mtb* (WHO, 2007). Of these, between 5 and 10 % will develop symptoms of the disease. A dense, lipid rich, waxy cell wall, and the ability to enter a slow growth phase, in which biochemical targets are down-regulated, allows *Mtb* to survive in infected organisms for a long time (Berger and Knodel, 2003). In order to cure TB, it is necessary to have agents that are active against both the acute and slow growing phases. For this reason, *Mtb* treatments rely on a combination of several drugs.

3.1.2 *Mtb* P450s

The genome of *Mtb* was solved in 1998 (Cole *et al.*, 1998). This revealed the presence of 20 putative P450 enzymes in *Mtb*. It was suggested that one or more of the *Mtb* P450s may be useful drug targets, particularly as P450s are known drug targets for azole anti-fungal drugs, and other inhibitors (Munro *et al.*, 2003, Souter *et al.*, 2000). Ahmad *et al.* found the azole antifungal drug econazole to be effective against murine TB (Ahmad *et al.*, 2006). Econazole has been found to bind tightly to several *Mtb* P450s (McLean *et al.*, 2006a), including CYP132, which is studied in this chapter.

The first of the *Mtb* P450s to be investigated was CYP51B1 (Aoyama *et al.*, 1998, Bellamine *et al.*, 1999). This had substantial similarity to eukaryotic CYP51 enzymes involved in sterol 14 α -demethylation. The fungal CYP51 was known to be the target for the azole anti-fungal drugs (Odds *et al.*, 2003). *Mtb* CYP51B1 was expressed in *E. coli* and shown to catalyse the 14 α -demethylation of various sterols (Bellamine *et al.*, 1999). The x-ray crystal structure of CYP51B1, in the resting state and bound to 4-phenylimidazole and fluconazole was solved (Podust *et al.*, 2001). Although there is no evidence for a complete endogenous pathway for sterol synthesis in *Mtb*, it has been speculated that CYP51B1 could be involved in catalysis of host sterol/steroid conversions (McLean *et al.*, 2007).

Mtb CYP121 shows similarities to polyketide metabolising P450 isoforms (McLean *et al.*, 2002a), including *Saccharopolyspora erythraea* P450 eryF, which catalyses the hydroxylation of the polyketide 6-deoxyerythronolide B in the synthesis of erythromycin (Cupp-Vickery *et al.*, 1996). CYP121 was expressed and purified and found to bind tightly to a range of azole antifungal drugs. The crystal structure of CYP121 was determined to 1.06 Å (Leys *et al.*, 2003). Later, the crystal structure of CYP121 bound to fluconazole was solved (Seward *et al.*, 2006). This revealed novel binding modes of the azole to the haem, which will be discussed further in Chapter 5 of this thesis. The CYP121 haem was found to be bound in two conformations, related to one another by a 180° rotation about the CH $_{\alpha}$ -Fe-CH $_{\beta}$ axis. There was also a distortion of ~30° of the haem macrocycle at one of its four pyrrole rings (Leys *et al.*, 2003).

Comparison of the genome sequences of *Mtb* and the vaccine strain *M. bovis* reveals that *Mtb* CYP130 and CYP141 are not present in the *M. Bovis* genome (McLean *et al.*, 2007). This can be interpreted two ways in terms of these enzymes' importance to *Mtb*. Firstly, it suggests that the P450s are not vital for the survival of *Mtb*. Conversely, it can be inferred that they may be important for the virulence or infectivity of *Mtb*. CYP130 and CYP141 have not been characterised to date. Some information about their possible functions may be inferred by their position in the *Mtb* genome. CYP130 is adjacent to an oxidoreductase, and the genes surrounding CYP141 are involved in molybdenum cofactor biosynthesis.

Mtb CYP128 was identified by genome-wide transposon libraries as required for optimal growth of the bacterium. Also, it was found to be upregulated in cell starvation (Sassetti *et al.*, 2003). CYP128 has not been characterised to date, but is adjacent to lipoprotein genes, as well as being close to *CYP121* and *CYP124* encoding genes (McLean *et al.*, 2007).

The expression of *Mtb* CYP123 and CYP138 have been found to be induced at elevated temperatures (Stewart *et al.*, 2002). The function of CYP123 may be related to that of CYP51B1 (McLean *et al.*, 2007).

CYP125 was found to be both induced in infected murine macrophages, and to be essential for *Mtb* infection in mice (Kendall *et al.*, 2004). CYP125 has been found to bind tightly to azole drugs. It is adjacent to genes involved in lipid degradation in the *Mtb* genome (McLean *et al.*, 2007). CYP125 and CYP142 from *Rhodococcus* were found to be part of a pathway necessary for cholesterol and steroid degradation (Van der Geize *et al.*, 2007). *Mtb* has been shown to grow on cholesterol as a carbon source. Also, cholesterol has been found to be necessary for *Mtb* macrophage entry (Gatfield and Pieters, 2000). This suggests CYP125 and CYP142 are also important to *Mtb*.

CYP144 has been found to bind tightly azole drugs, (McLean *et al.*, 2006a), and has a possible role in *Mtb* virulence (McLean *et al.*, 2007). It is adjacent to genes that are likely to be involved in membrane metabolism in the *Mtb* genome.

3.1.3 *Mtb* CYP132, a putative fatty acid hydroxylase

Mtb CYP132, encoded by the gene *Rv1394c*, shows significant homology at the amino acid level with various eukaryotic CYP4 family fatty acid hydroxylases, and was postulated to have such a role in *Mtb* (Munro *et al.*, 2003). *Mtb* has a thick, lipid rich, cell wall, consisting of unusual long-chain fatty acids with hydroxyl, methyl and cyclopropyl substitutions (Berger and Knodel, 2003). It is possible that CYP132 is involved in the synthesis of these fatty acids. Alternatively, CYP132 may be involved in the metabolism of fatty acids during the slow growth phase of the bacterium.

Recchi *et al.* (Recchi *et al.*, 2003) found that the expression of CYP132 is induced by the protein encoded by *Rv1395*. This is an AraC class transcriptional regulator (Plano, 2004). Gene regulation is pivotal to host-microbe interactions. The trigger for the expression of *Rv1395* is unknown, but it may be that CYP132 is involved in maintaining *Mtb* in the intracellular environment, and is expressed in response to a defence mechanism of the host organism.

Gene knockout studies by Camacho *et al.* (Camacho *et al.*, 1999) using signature-tagged transposon mutagenesis revealed that when *Rv1395* is not present in *Mtb*, the growth of the bacterium was reduced. As CYP132 has been shown to be induced by the *Rv1395*-encoded regulator protein, this suggests CYP132 is important to the virulence of *Mtb*. This, coupled with the likelihood that it is a fatty acid hydroxylase, make CYP132 a good candidate for further study from the group of *Mtb* P450s that have not yet been characterised. Potentially, CYP132 has a virulence-associated role and could be a target for novel anti-tubercular drugs.

3.1.4 Aim

Mycobacterium tuberculosis, the bacterium responsible for the disease tuberculosis (TB), has seen a massive increase in the past two decades, with nine million new cases each year worldwide, and two million deaths. Multidrug-resistance and treatment times of at least six months mean there is a desperate need for new treatments for TB.

The genome of *Mtb* was solved in 1998 (Cole *et al.*, 1998). This revealed the presence of 20 putative P450 (CYP) enzymes, of which several have now been characterised (McLean *et al.*, 2007). CYP132 is induced by the transcriptional regulator that is the product of the *Rv1395* gene (Recchi *et al.*, 2003). Studies of a *Rv1395* gene knockout strain of *Mtb* showed decreased virulence in the lungs of mice (Camacho *et al.*, 1999). The link with virulence of the bacterium, and likelihood of finding ligands that will bind to the P450 (e.g. azoles), make CYP132 a possible anti-TB drug target, and an excellent candidate for protein characterisation studies.

In this chapter, the expression of the *Rv1394c* gene in *E. coli*, and the purification and characterisation of CYP132, will be described. The interaction of CYP132 with various

possible fatty acid substrates and azole inhibitors was investigated, leading to some unexpected results, which were investigated further and are reported in this chapter.

The purified CYP132 was found to have been digested during the expression and purification process, leading to a fragment 4/5 of the length of the complete enzyme. Attempts were made to protect the protein from digestion by the attachment of a GST protein tag.

3.2 Results and Discussion

3.2.1 Expression and purification of CYP132

The *Mtb Rv1394c* gene that had been cloned into the pET11a vector was obtained from Dr. K. McLean, University of Manchester. The restriction enzymes *NdeI* and *BamHI* had been used to excise the gene from a storage vector and to enable its ligation into the pET11a plasmid. An *AvrII* restriction site was also created during the cloning process. The *AvrII* and *NdeI/BamHI* restriction enzymes were used as diagnostic tests of the cloning process.

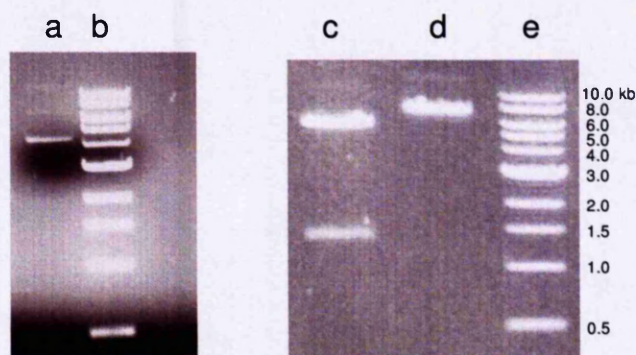


Figure 3.1: 0.8 % Agarose gels of the *Rv1394c/pET11a* plasmid. (a) Plasmid only, (b) NEB 1kb DNA ladder, (c) *Rv1394c/pET11a* digested with *NdeI* and *BamHI*. The two fragments correspond to the *Rv1394c* gene, length ~1.3 kb and the pET11a vector, length ~5.5 kb, (d) *Rv1394c/pET11a* digested with *AvrII*, the single band corresponds to the linearised plasmid DNA, length 7.028 kb, (e) NEB 1kb DNA ladder. The lengths of the marker bands are shown in kb.

The DNA sequence of the *Rv1394c*/pET11a plasmid was confirmed by automated dideoxy DNA sequencing. Correct transformation into the appropriate expression strain prior to protein expression was confirmed by digestion with *AvrII* and *NdeI*/*BamHI* restriction enzymes. Agarose gels of one such digestion are shown in Figure 3.1.

3.2.1.1 Gene expression over time

The expression protocol of McLean *et al.* (McLean *et al.*, 2002a) was used as a starting point in the expression of CYP132. In order to produce properly folded, soluble protein, a low growth temperature of 18 °C was used. This decreases the growth rate of the cells, the transcription and translation rates and hence the production of the protein, and thus may aid enhance correct protein folding. A similar strategy was employed successfully in the production of the CYP121 enzyme (McLean *et al.*, 2002a).

The expression of CYP132 was carried out as described in Chapter 2, taking samples at regular time intervals after induction with IPTG, up to 18 h. In order to assess the amount of IPTG required to induce a suitable level of target gene expression, a range of concentrations of IPTG were investigated. SDS-PAGE gels of the samples taken following different induction regimes are shown in Figure 3.2.

It can be seen in Figure 3.2 that a large band of mass ~36 kDa is present in the cell samples that had been induced with IPTG, but not in the sample left uninduced. At the time it was considered that this might correspond to CYP132 that had not run to size on SDS-PAGE. It was later found that CYP132 does run true to size, and that the protein contained in the band at ~36 kDa does not contain haem. It is thus likely that the ~36 kDa band is due either to a truncated version of CYP132 (i.e. possibly reflecting an alternative stop codon, perhaps through mutation of the gene), and that has thus not incorporated haem, or (more likely) to CYP132 that has been expressed as an intact protein, but has subsequently been proteolytically cleaved and has lost the haem-binding region and the cofactor.

The band at ~36 kDa is more intense in the samples grown for 18h after induction than in the samples grown for 4h. This suggests that it is beneficial to use long growth times in order to express a greater amount of protein. The amount of IPTG used (from 0.1-1

mM) seems to make little difference to amounts of CYP132 protein produced, except for when none is used. Following this observation 0.1 mM IPTG was used in the expression from the *Rv1394c/pET11a* transformed cells.

3.2.1.2 Effect of growth medium on CYP132 production

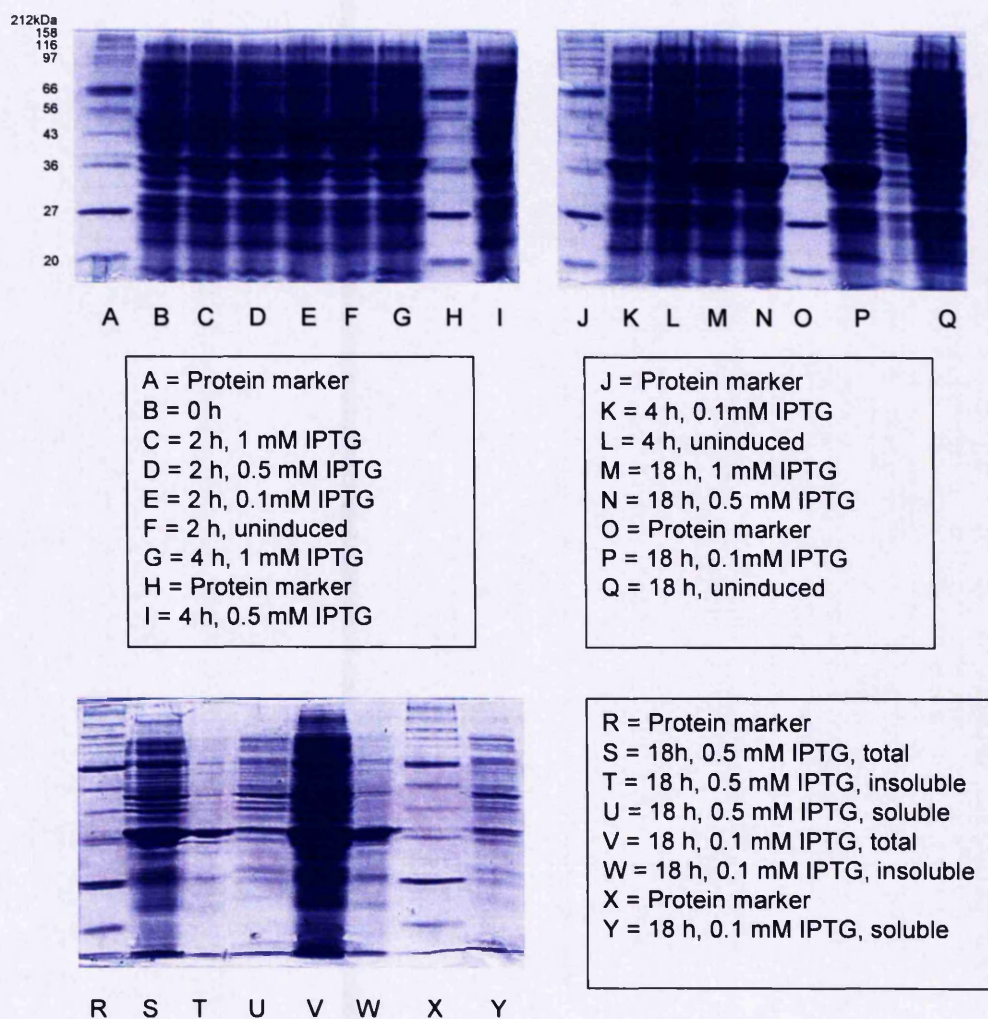


Figure 3.2: Expression of CYP132 protein (Top) SDS-PAGE gels showing the total production of protein with time from the *Rv1394c/pET11a* plasmid in *E. coli* strain HMS174(DE3), after induction with a range of concentrations of IPTG. Key shows time after induction and concentration of IPTG. Masses of protein marker bands are shown, in kDa. (Bottom) SDS-PAGE gels of samples after separation into total protein, insoluble and soluble fractions using Bugbuster solution.

In order to assess the best growth medium to use to obtain the optimum yield of CYP132, the *Rv1394c/pET11a* plasmid transformant in HMS174(DE3) was expressed in LB, YT, and TB media. Samples were taken at times of 0, 5 and 18 h after induction with IPTG. SDS-PAGE gels of the samples are shown in Figure 3.3.

There was little apparent difference in CYP132 expression between the three media used according to the SDS-PAGE gel results. YT and TB media gave slightly higher final cell densities (as established from OD₆₀₀ measurements), and the cells grew faster in these media, as judged by shorter times for the cultures to reach OD₆₀₀=0.9. This also allowed for longer protein expression times following IPTG addition to the cultures at this point. As there was no apparent benefit to using TB over YT, the cheaper YT medium was used for the expression of CYP132.

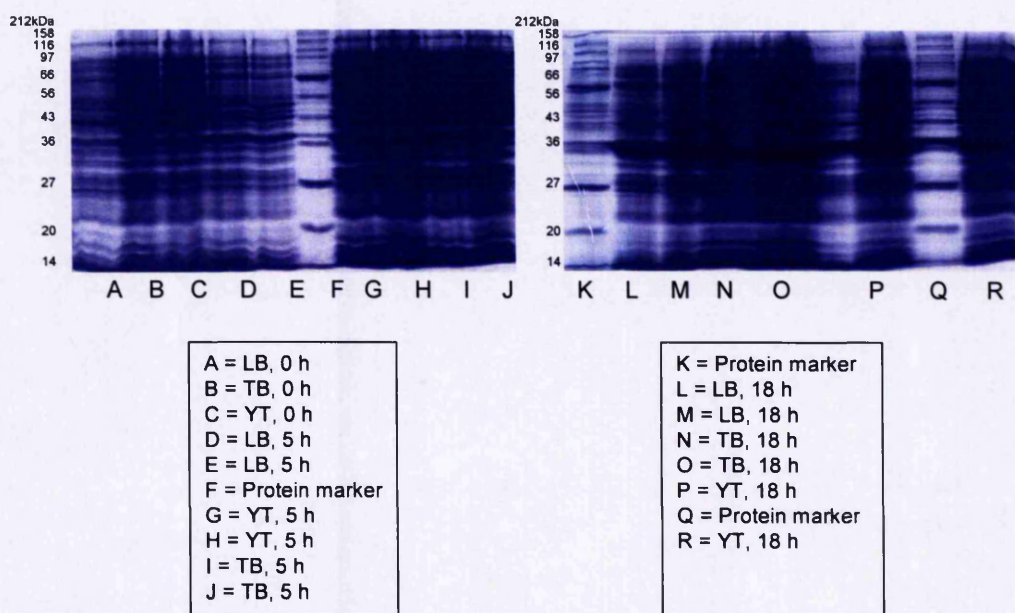


Figure 3.3: SDS-PAGE gels showing the total production of protein with time from the *Rv1394c/pET11a* plasmid transformed in HMS174(DE3), using LB, TB and YT media. Key shows time after induction and medium used. Masses of protein marker bands are shown, in kDa.

3.2.1.3 Attempted purification of CYP132

The pelleted cells from the expression of *Rv1394c/pET11a* in HMS174(DE3) and Rosetta(DE3) were broken open by sonication as described in Chapter 2. Purification of

the expressed protein was attempted by DEAE and phenyl Sepharose column purification. Similar results were found with Rosetta(DE3) and HMS174(DE3) cells. The P450 component of the cell extract was found not to bind to the DEAE Sepharose column. UV/Vis spectra of the cell extract and the eluent from the phenyl Sepharose column are shown in Figure 3.4.

It can be seen in Figure 3.4 that the cell extract from the HMS174(DE3) cells contains a small amount of haem-binding protein, as shown by the small feature around 420 nm (i.e. at the position of the haem Soret band). This becomes slightly more pronounced after phenyl Sepharose purification. The SDS-PAGE gels in Figure 3.3 show a large band at mass ~36 kDa appearing over time and that is not present in the uninduced sample.

The small amount of haem-containing protein present in the cell extract suggests that the ~36 kDa protein seen on the SDS-PAGE gels does not contain haem. The band on the gels indicates considerable amounts of the ~36 kDa protein. If this protein all contained haem, it might be expected to give a much larger haem peak around 420 nm in the UV/Vis spectrum. As the expression of this protein is induced by the addition of IPTG it is likely that it is expressed by the *Rv1394c*/pET11a plasmid. The predicted mass of CYP132 is 52.2 kDa.

It is likely that the source of the ~36 kDa protein, seen on the SDS-PAGE gels in Figure 3.3, is either truncated form of CYP132 resulting from incomplete expression of the gene from the *Rv1394c*/pET11a plasmid (i.e. likely a C-terminal truncation removing haem-binding sections and meaning the protein is not able to incorporate the cofactor), or proteolytic cleavage of the expressed protein following expression, resulting in the loss of the haem-binding region.

There was not sufficient haem-containing protein in the eluent from the phenyl Sepharose column to merit further purification of the protein from this material. The low level expression, if any, of CYP132 from these cells meant it was desirable to be able to purify the protein in a single step. For this reason it was decided to attempt the expression of CYP132 with a polyhistidine tag. This would allow purification by selective binding of the tag region to a nickel column. Also, the addition of a polyhistidine tag to

the protein would be a convenient way to ascertain whether the observed ~36 kDa protein was the result of incomplete expression, as if this were the case it would be expected to gain six histidine residues, resulting in an increase in mass and an affinity for nickel-containing resin.

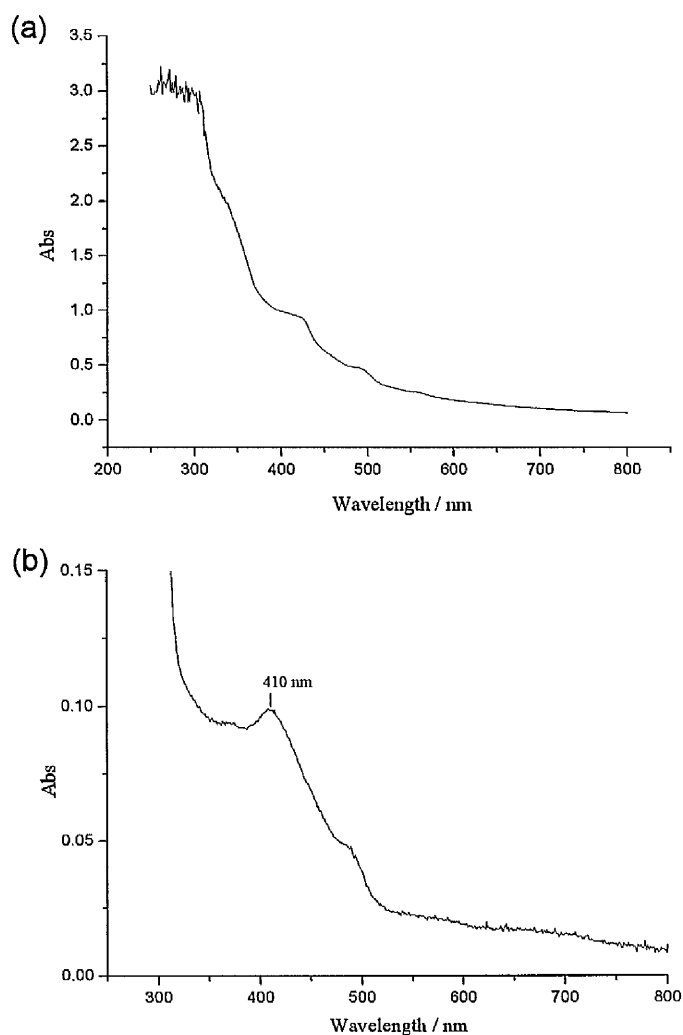
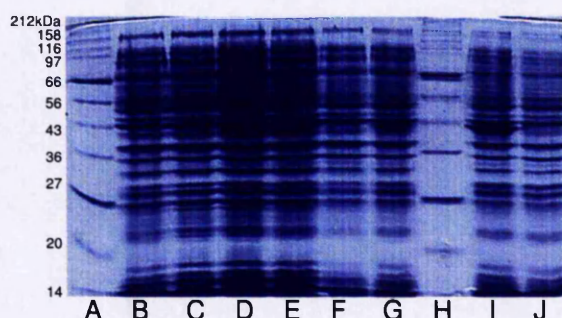


Figure 3.4: (a) UV/Vis spectrum of the cell extract from the expression of *Rv1394c/pET11a* in HMS174(DE3). (b) UV/Vis spectrum of the eluent from phenyl Sepharose purification of the cell extract from the expression of *Rv1394c/pET11a* in HMS174(DE3)

3.2.2 Expression and purification of Polyhistidine-tagged CYP132

3.2.2.1 Expression of polyhistidine-tagged CYP132

The use of a polyhistidine tag to aid purification of proteins was developed by Hochuli *et al.* (Hochuli *et al.*, 1988). They fused a poly-histidine peptide to mouse dihydrofolate reductase. It was found that the protein could be purified by metal ion affinity chromatography using Cu^{2+} , Ni^{2+} , Co^{2+} or Zn^{2+} ions. The histidine peptide could be subsequently removed with the proteolytic enzyme carboxypeptidase A.



A = Protein marker
B = 0 h
C = 0 h
D = 5 h
E = 5 h
F = Uninduced, 5h
G = 18h
H = Protein marker
I = 18h
J = Uninduced, 18h

Figure 3.5: SDS-PAGE gels showing the expression with time for transformants of *Rv1394c*/pET15b plasmid in the HMS174 (DE3) cell strain using YT growth medium. Key shows time after induction, in hours. The masses of the protein marker bands are shown on the left hand side, in kDa.

The *Rv1394c* gene that had been cloned into the NEB pET15b vector was obtained from Dr. K. J. McLean, University of Manchester. This incorporates a six histidine tag at the N-terminus of the protein. The DNA sequence of the *Rv1394c*/pET15b plasmid was confirmed by automated dideoxy DNA sequencing (MWG Biotech DNA sequencer).

The *Rv1394c*/pET15b plasmid was transformed into HMS174(DE3) and Rosetta(DE3) cell strains and the expression carried out as described in Chapter 2. Samples were taken at regular intervals. SDS-PAGE gels showing the expression with time after induction are shown in Figure 3.5.

The first thing that is apparent from the SDS-PAGE gel from the expression of the *Rv1394c*/pET15b plasmid in HMS174(DE3), shown in Figure 3.5, is that the large band at ~36 kDa, seen in the expression of *Rv1394c*/pET11a, does not appear to be present. The bands above and below ~36 kDa seen in lanes G and I of Figure 3.5 are not significantly more intense than those than those present in the uninduced cells, lane J, suggesting they arise from the *E. coli* cell strain rather than being encoded by the plasmid.

There is a band at 42 kDa, which is especially pronounced in lane I of the gel, which is not present in the uninduced sample. This was found to be the major product from the expression of *Rv1394c*/pET15b in HMS174(DE3), and was later found to correspond to CYP132 that has been cleaved, but still incorporates haem.

3.2.2.2 Plasmid stability test

HMS174(DE3) was found to express His-tagged CYP132 very well, providing that the cells grew well and reached an $OD_{600} \sim 0.9$. On several occasions, however, the OD_{600} was found to stop increasing at around 0.3 and then to decrease back towards zero. This suggested that the cells ceased growth at an early stage, possibly due to sensitivity to the antibiotic used. This brought into doubt the stability of the plasmid in the cells, which conferred ampicillin resistance. The plasmid stability test, as described in Chapter 2, was carried out on samples taken at $OD_{600} = 0.9$ and immediately prior to gene induction with IPTG. The results are shown in Table 3.1.

The data in Table 3.1 show that there were few colonies on plates containing IPTG, and that there were none on the plate containing both LB and ampicillin. Larger numbers of colonies grew on plates in the absence of inducer or antibiotic. This suggested that (over)expression of the *Rv1394c* gene product (i.e. CYP132) might be toxic to cells, that "leaky" expression of the gene in absence of inducer might occur, and that there was

likely adaptive plasmid loss in these cells, consistent with variable successes in culture growth.

Table 3.1: Plasmid stability test on a sample taken from a culture of HMS174(DE3) transformants containing the *Rv1394c/pET15b* plasmid, at a culture $OD_{600} = 0.9$ immediately prior to inducing with IPTG. The numbers are the number of colonies counted on each LB agar plate and following overnight incubation at 37 °C.

Plate	Test Sample
LB only	~ 80
LB Ampicillin	~ 50
LB IPTG	3
LB IPTG Ampicillin	0

The HMS174(DE3) and Rosetta(DE3) transformant cell strains were found to express similar amounts of (cleaved) CYP132, as defined by SDS-PAGE gel analysis. Rosetta(DE3) cells proved to be less useful as an expression host, since cell extracts consistently became turbid, and this led to failures in cell extract proteins binding to columns used for initial purification steps. To overcome this problem, an additional ammonium sulphate fractionation purification step was required, increasing the purification time, and the risk of protein denaturing/cleaving. Because of this difficulty in purifying CYP132 protein expressed using the Rosetta(DE3) cell strain, HMS174(DE3) was chosen as the preferred cell strain for the expression of CYP132 from the *Rv1394c/pET15b* plasmid.

3.2.2.3 Purification of polyhistidine tagged CYP132

The pelleted cells from the expression of *Rv1394/pET15b* in the HMS174(DE3) and Rosetta(DE3) cell strains were disrupted by sonication and by using a French press, as described in Chapter 2. UV/Vis spectrometry showed a much more pronounced peak around 416 nm than was observed in cell extracts from the expression of the *Rv1394c/pET11a* construct transformant, indicating a much higher concentration of haem-containing protein in the sample.

Purification was attempted by binding CYP132 to a nickel affinity column. The protein was found not to bind to the column under a variety of conditions. This suggested that the polyhistidine tag was either not present in the protein, or it had become buried within

the protein due to the protein folding. It was later found that the histidine tag was not present as it was part of the section of protein (at the N-terminal) that had been cleaved from the P450 during cellular expression.

Purification of the protein expressed by *Rv1394c*/pET15b was carried out using a succession of DEAE Sepharose, hydroxyapatite and Q-Sepharose column steps. Further purification, for applications requiring high enzyme purity, was carried out using a monoQ column. UV/Vis spectra showing the progressive increase in purity following each column purification step for one batch of protein are shown in Figure 3.6. SDS-PAGE gels showing the relative purity after each stage are shown in Figure 3.7.

Figure 3.6 and Figure 3.7 show the progressive increase in CYP132 purity with each column purification step used. The need for three columns to obtain acceptable purity (for e.g. spectroscopic analysis), or four for close to total purity (for e.g. protein crystallographic studies), is not ideal, since the purification regime is laborious and clearly leads to progressive proteolysis of the protein. The binding of CYP132 to the DEAE-Sepharose and hydroxy apatite columns was not strong, with the P450 often gradually eluting from the column without the need for salt gradients, or else in low salt concentrations. It was found that acceptable purity could not be obtained if either of these columns were omitted, however. Within the time available, a superior purification regime could not be developed. The monoQ column could potentially purify CYP132 to near-homogeneity in a single step, but as this high quality, expensive column had rather low capacity, it proved necessary for the protein to be considerably purified from the cell extract prior to loading onto monoQ in order to avoid requirement for multiple loadings of the same column.

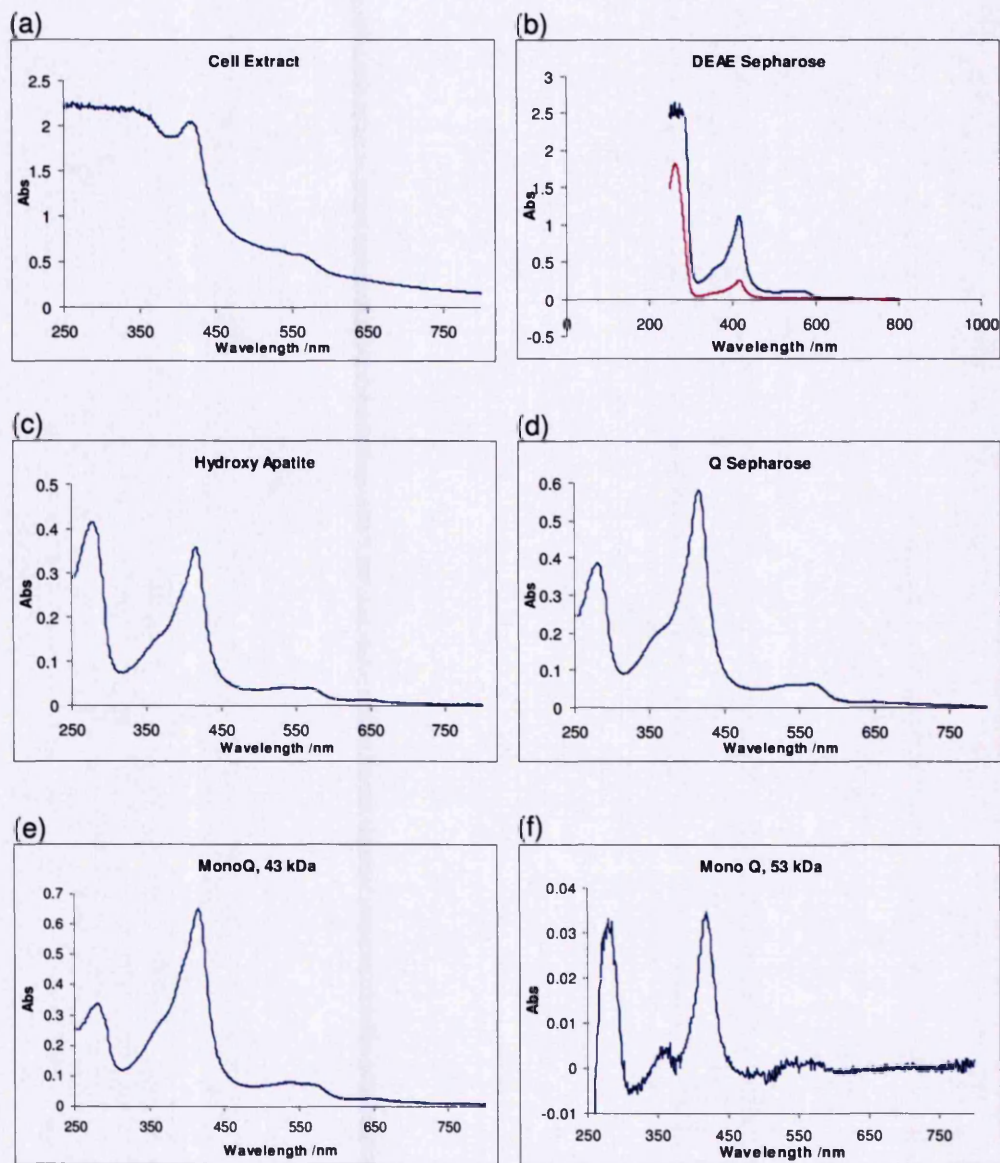


Figure 3.6: UV/Vis spectra showing the relative purity of the CYP132 protein expressed from the *Rv1394*/pET15b plasmid in HMS174(DE3) transformants after various stages of protein purification. (a) Cell extract, (b) after DEAE-Sepharose purification, spectrum in magenta is with a four fold dilution, (c) after hydroxy apatite purification, (d) after Q-Sepharose purification (e) after monoQ purification, major product, 43 kDa (f) after monoQ, trace product, 53 kDa. Poor quality of spectrum (f) is due to the low protein concentration for this species.

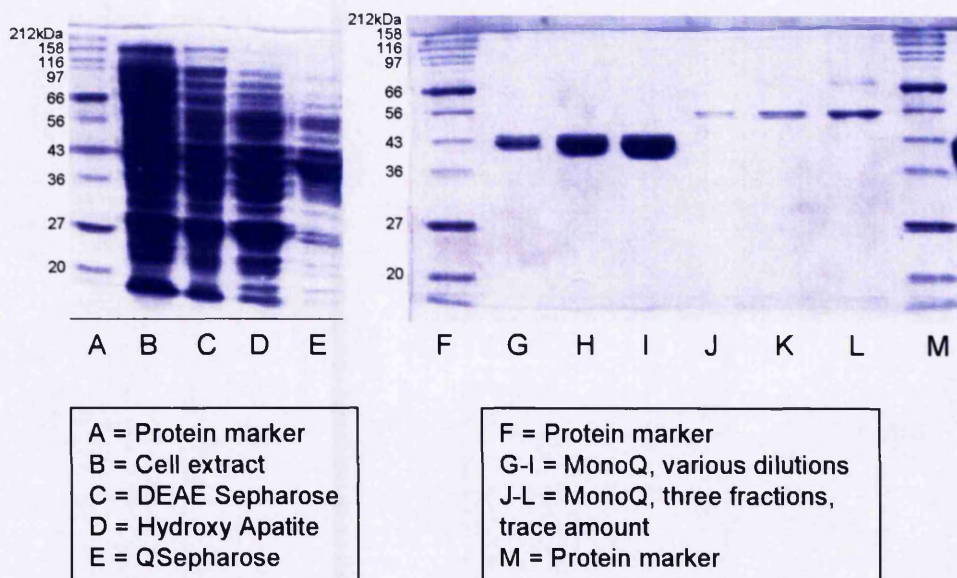


Figure 3.7: SDS-PAGE gels of the CYP132 product of expression from the *Rv1394c/pET15b* plasmid in HMS174(DE3) cells, showing relative purity after each stage of purification. Samples G, H and I correspond to three quantities of the monoQ purification product (ca 1, 5 and 10 μ g). Samples J, K and L correspond to three fractions containing a trace amount of haem-containing protein, revealed by monoQ purification and likely representing intact CYP132.

3.2.2.4 Investigation of other column resins and buffer conditions for CYP132 isolation

It had been found that application of four different column purification steps were required to obtain pure CYP132 protein. This is not ideal, as the lengthy purification regimen means that the protein is exposed to the possibility of proteolysis or other denaturation over the extensive period required for its isolation. In order to reduce the number of purification steps required to obtain pure protein, the use of other resins and alternative buffer pH conditions was investigated.

1 ml samples of cell extract were mixed with 1.5 ml of various different types of column resins in appropriate buffer of varying pH. UV/Vis spectra of the solutions after their mixing with the various resins are shown in Figure 3.8 and Figure 3.9. The loss of the 416 nm absorption feature from the eluted extract indicates that the P450 has bound to the resin.

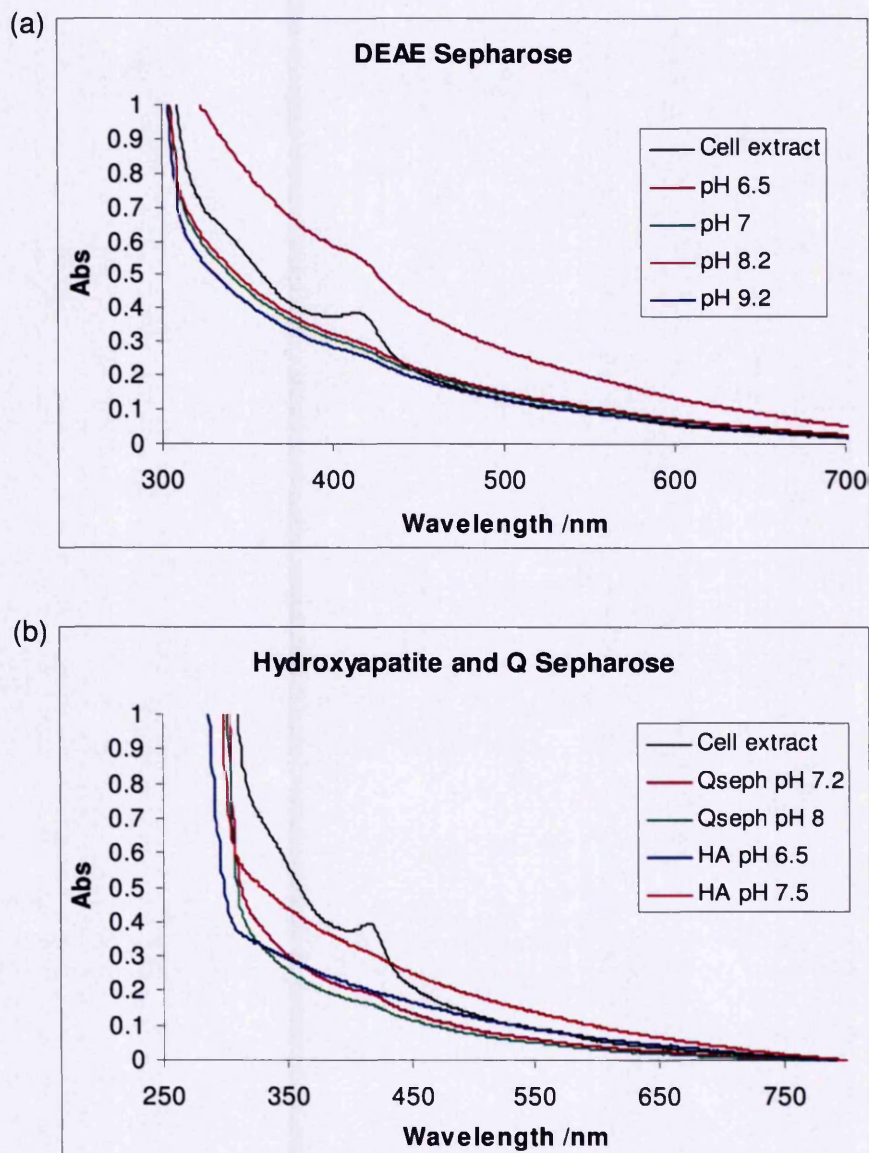


Figure 3.8: UV/Vis spectra of the cell extract from the expression of CYP132 from the HMS174(DE3)/*Rv1394c*/pET15b transformant before and after mixing with DEAE Sepharose, Q Sepharose and hydroxy apatite column resins. The 800 nm absorbance has been fixed at zero absorption in each spectrum. Buffer used is Tris (10 mM) for the DEAE and Q Sepharose resins, and potassium phosphate (KPi, 10 mM) for hydroxy apatite. pH of buffers is shown in the key.

It can be seen in Figure 3.8 and Figure 3.9 that the CM Sepharose and SP Sepharose resins did not bind any significant amount of the P450 under any of the pH conditions used. By contrast, DEAE, Q Sepharose and hydroxy apatite all bound CYP132. There

was little difference in apparent binding efficiency between the different pH conditions used for the binding of CYP132 to hydroxy apatite and Q Sepharose resins. However, DEAE Sepharose showed reduced binding of CYP132 at pH 6.5, with the three higher pH values investigated producing roughly equal binding of the P450 to the resin.

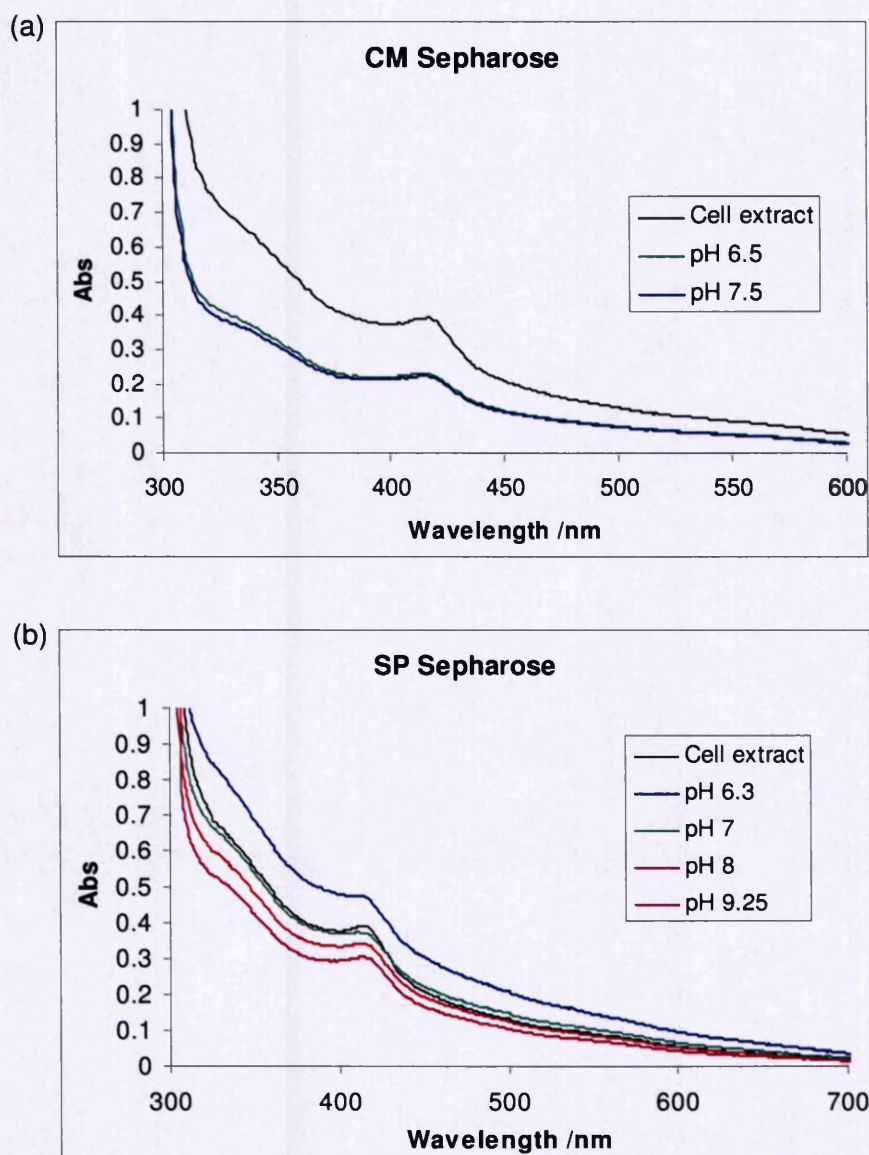


Figure 3.9: UV/Vis spectra of the cell extract from the expression of CYP132 from the HMS174(DE3)/*Rv1394c*/pET15b transformant before and after mixing with CM Sepharose and SP Sepharose column resins. The 800 nm absorbance has been fixed at zero absorption in each spectrum. Buffer used was KPi (10 mM). pH of buffer is shown in the key.

3.2.3 Characterisation of CYP132 expression from the *Rv1394c*/pET15b Plasmid

3.2.3.1 Preliminary analysis of the HMS174(DE3) *Rv1394c*/pET15b expression products

The 43 kDa and 53 kDa CYP132 protein products from the expression of the *Rv1394c* gene in HMS174(DE3)/*Rv1394c*/pET15b were analysed by mass spectrometry. The mass spectra are shown in Figure 3.10 and Figure 3.11. The major product was found to have a mass of 43.1 kDa, and the trace product 53.3 kDa. Analysis of the two samples by UV/Vis spectrometry revealed them to have typical P450 spectra, with features such as the Soret band at 416 nm, due to π to π^* electronic transitions in the haem macrocycle, and with the less intense α and β bands at around 550 nm. These are shown as the black lines in Figure 3.12.

The 43 kDa and 53 kDa protein products were reduced with sodium dithionite and bubbled with carbon monoxide (CO). This was designed to enable formation of the Fe(II)CO complex of the enzyme, which (for thiolate-coordinated P450s) has a diagnostic Soret absorption shift to ~450 nm (McLean *et al.*, 2006c). The UV/Vis spectra for the Fe(II)CO species are shown as the red lines in Figure 3.12. The 43 kDa protein gave a shift in the Soret band predominantly to 418 nm, with a very small absorbance feature at 446 nm. The 53 kDa protein gave an absorption shift to mostly 449 nm, with only a small absorbance at 420 nm.

The binding of reduced P450s with carbon monoxide to give a characteristic absorption at 450 nm was first observed by Klingenberg (Klingenberg, 1958) in microsomes prepared from rat liver. As a result of this observation, the name 'P-450' was first used by Omura and Sato in 1964 (Omura and Sato, 1964). It is now clear that the thiolate form of the cysteine (i.e. cysteinate) proximal ligand is required for the Fe(II)CO complex to have an absorption maximum near 450 nm (i.e. P450). If the cysteinate ligand to the haem is lost, usually as a consequence of its protonation to the thiol, binding of CO to the ferrous iron will instead give an absorbance at ~420 nm (Perera *et al.*, 2003). This is termed P420.

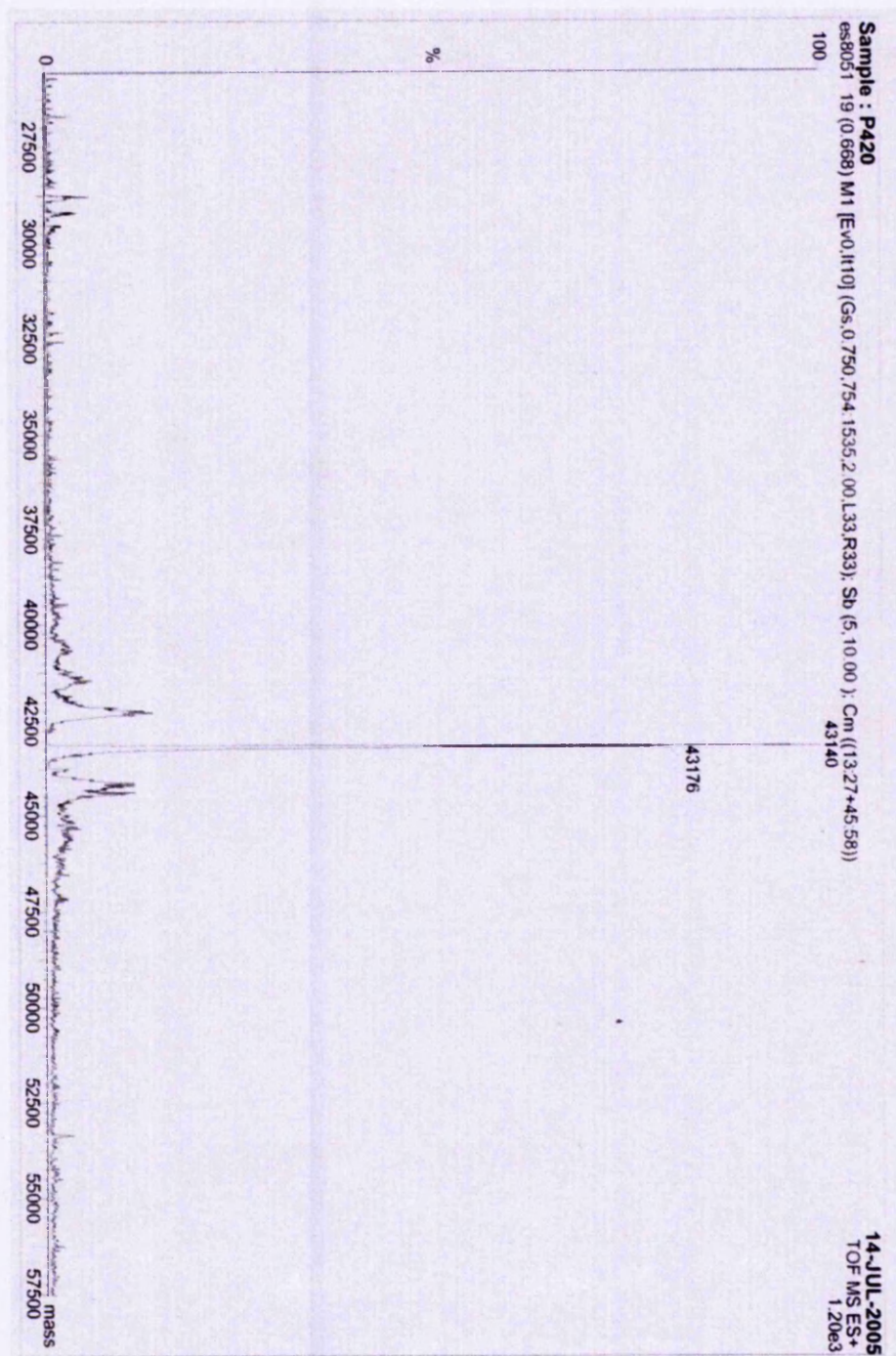


Figure 3.10: Electrospray time of flight mass spectroscopy of purified protein from the expression of CYP132 from the *Rv1394c*/pET15b plasmid, Figure 3.7 sample G. The molecular mass of 43176 Da is consistent with a proteolytically cleaved version of CYP132. Data were collected at the Biosciences facility at the University of Birmingham.

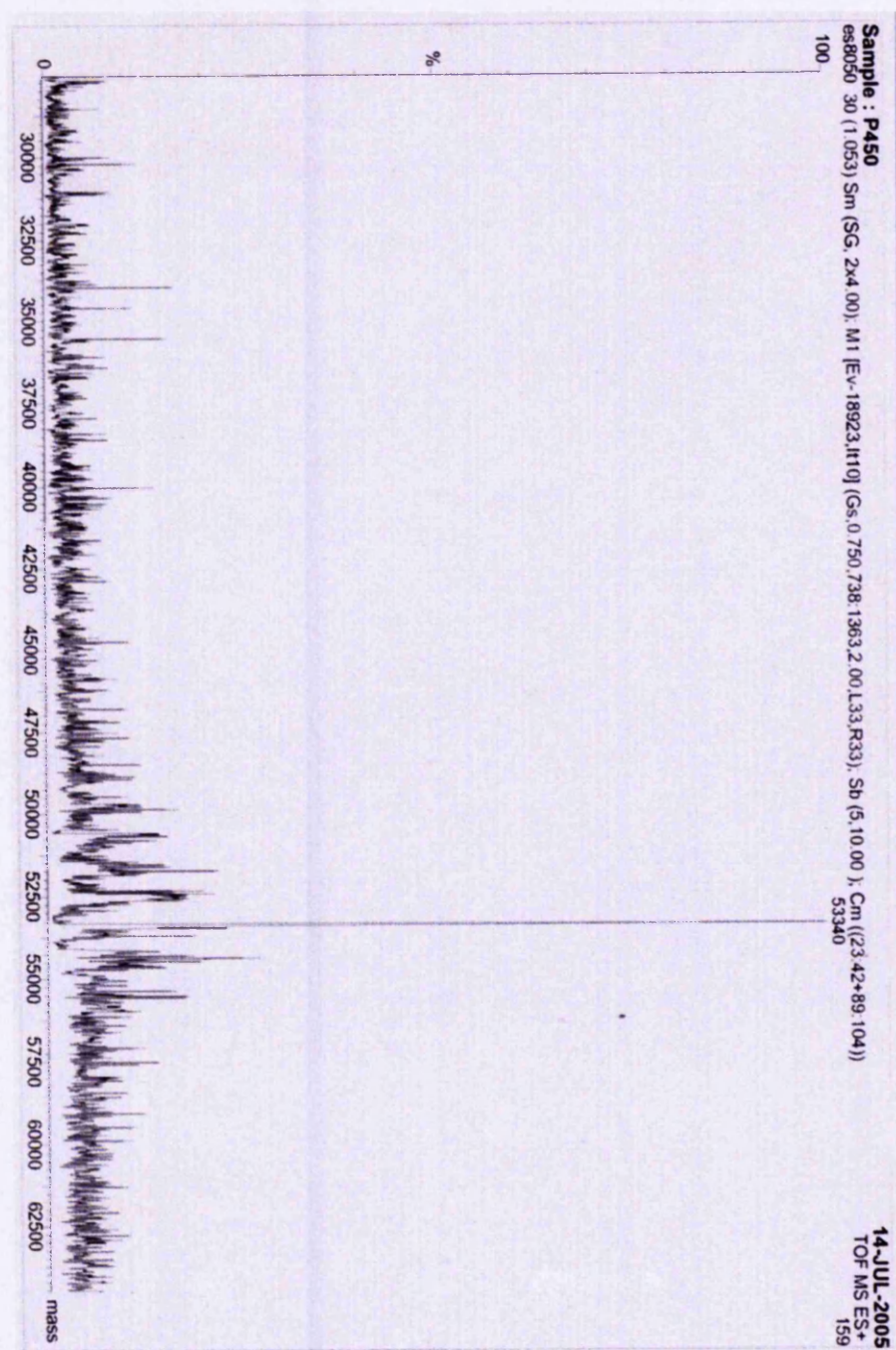


Figure 3.11: Electrospray time of flight mass spectrometry of the small amount of purified protein from the expression of CYP132 from the *Rv1394c*/pET15b plasmid, Figure 3.7 sample I. The molecular mass of 53340 Da is consistent with an intact version of CYP132. Data were collected at the Biosciences facility at the University of Birmingham.

Binding of CO to the reduced forms of the different CYP132 products from the expression of *Rv1394c/pET15b* in HMS174(DE3) revealed the major product, the 43 kDa protein, to be almost entirely P420. By contrast, the trace product, with a mass of 53 kDa was found to be mostly P450 in its Fe(II)CO complex. There are no P450 proteins encoded in the *E. coli* genome. This means that both the 43 kDa and the 53 kDa proteins must be associated with CYP132 and derived from the *Rv1394c/pET15b* expression system. The mass of 53.3 kDa is very close to the expected mass of polyhistidine tagged CYP132 (53.4 kDa). This suggests that the 53.3 kDa protein is the intact polyhistidine tagged CYP132.

The UV/Vis spectrum of full-length CYP132 shows it to be predominantly in the low-spin, doublet, spin state. The shift to 450 nm following the binding of CO to the ferrous haem shows the cysteinate ligation to the haem to be intact. The trace amounts of full-length CYP132 isolated meant that more detailed analysis was not possible. It is not clear why the full-length polyhistidine tagged CYP132 did not bind to the nickel column. However, it may be that the tag becomes buried in the protein, or binds to other residues in the protein, and is thus unavailable to bind to the nickel column.

The 43 kDa protein has some P450 character, specifically in that it is able to form a Fe(II)CO complex with at least a small absorbance at ~450 nm, which reflects the presence of a cysteine residue in the proximal position. Thus, this is consistent with this protein being a shortened (i.e. proteolysed) version of CYP132. McLean *et al.* (McLean *et al.*, 2006b) found that *Mtb* CYP51B1 gave an absorbance at 450 nm when bound to CO, which collapsed to 420 nm over around 30 minutes. This shows that the binding of CO can destabilise cysteinate ligation in some P450s (leading to thiolate protonation and P420 formation). It may be that the cysteinate ligand in the truncated CYP132 is intact, but that it is liable to protonation upon CO binding.

The P420 Soret feature is in the same spectral region as might be expected for a His-Fe-CO adduct. However, features in the long wavelength region of the spectrum are generally distinct for these two forms. Structural studies of P450s do not reveal histidine residues that could readily replace cysteinate as the ligand, and other techniques provide compelling evidence for the retention of a sulphur ligand.

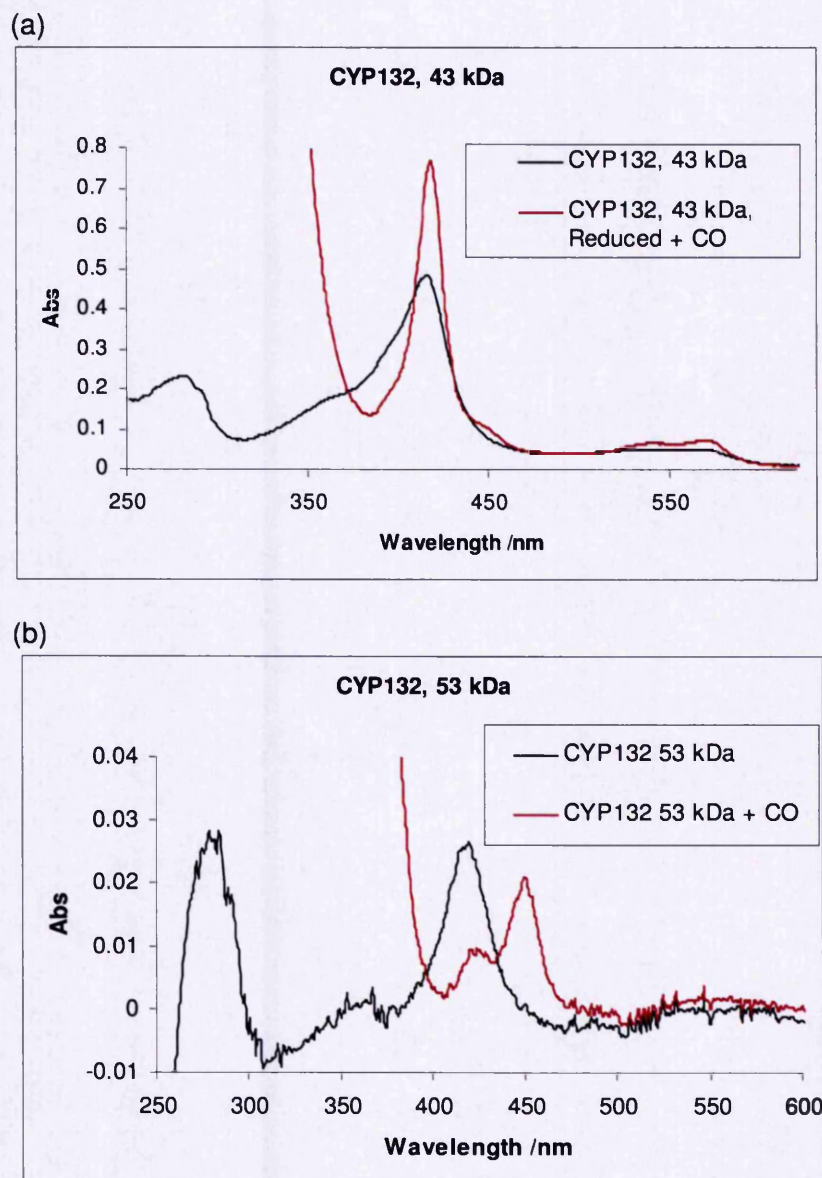


Figure 3.12: UV/Vis spectra showing binding of CO to the ferrous CYP132 products (cleaved, 43 kDa, and intact, 53 kDa) purified following protein production using the HMS174(DE3)/*Rv1394c*/pET15b expression system and using a monoQ resin to achieve high quality protein purification.

In order to assess the potential impact of the loss of ~10 kDa from CYP132, a protein sequence alignment was carried out with the haem domain of P450 BM3, using the Geneious 3.5.6 program (Figure 3.13). This P450 has 22.6 % amino acid sequence identity with CYP132 and a similar mass, of 53.5 kDa. Various crystal structures are available for the BM3 haem domain in ligand-free and ligand-bound forms. The structure of the BM3 haem domain with the substrate palmitoleic acid bound, solved by Li and Poulos (Li and Poulos, 1997), is shown in Figure 3.14.

With only 22.6 % amino acid identity between the CYP132 and P450 BM3 amino acid sequences, the structures will not be near-identical, but the cysteine residue that ligates the haem iron, and many of the residues surrounding the haem, are conserved. Moreover, the P450 fold in general is conserved among those P450s for which crystal structures are known. Many of the conserved residues around the haem, including the proximal cysteine, are within 10 kDa (i.e. ~90 amino acids) of the C-terminal, and would be missing if the cleavage to produce the 43 kDa protein was from this end of the protein. It is thus clear (since the truncated CYP132 binds haem) that the missing region of the cleaved CYP132 must be at the N-terminal of the protein.

The portion of the P450 BM3 haem domain that is aligned with the missing region of CYP132 is coloured blue in Figure 3.14 (a), and has been removed in (b). It can be seen that the residues around the haem active site are intact, but a large portion of the substrate binding region is missing. The similarity between the sequences is much lower away from the haem-binding region than in the area surrounding the haem. As a result, it can not be absolutely certain that the substrate binding site is in the same region of the CYP132 protein sequence as it is in P450 BM3. However, given that diverse regions throughout the P450 sequences (so called SRS sequences) are implicated in substrate interactions, it is likely that substrate binding is altered in the cleaved CYP132 compared to the full-length enzyme.

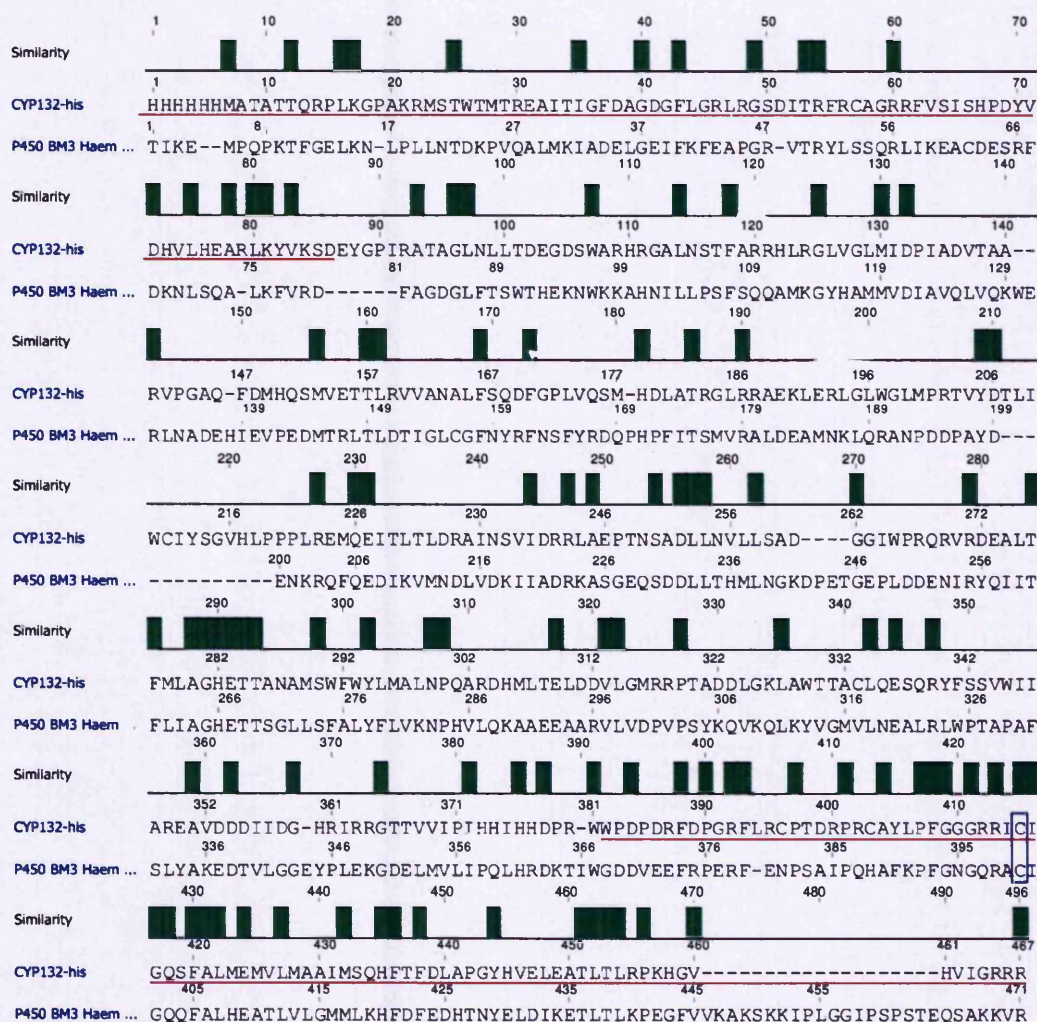


Figure 3.13: Protein alignment of polyhistidine tagged CYP132 and the P450 BM3 haem domain. Green boxes above the sequence indicate conserved residues. The portions underlined in red (~90 amino acids) at the beginning and end of the CYP132 sequence are the sections of protein which would be missing if the protein was cleaved at the N-terminal or the C-terminal, leading to a mass of 43.1 kDa in the cleaved CYP132 protein. The cysteine ligand to the haem is indicated by a blue box, and would be removed if a C-terminal proteolytic event occurred.

It can be seen in Figure 3.14 that, although the residues surrounding the haem are intact, the haem is likely much more exposed when the blue area is removed. Whatever the exact structure of CYP132, it is clear that the loss of ~10 kDa of protein from the N-terminal region will mean that the cofactor is more solvent exposed and vulnerable to

cleavage/protonation of the iron-sulphur bond. This likely explains why only a very small proportion of P450 was observed upon binding of CO to the cleaved, ferrous CYP132.

Examination of the UV/Vis spectrum of the cleaved CYP132 reveals that it is predominantly in the low-spin, doublet spin state, with a Soret band absorbance at 416 nm. There is a shoulder on the Soret band at ~395 nm, however. A Soret band absorbance at 395 nm usually signifies the type I binding of a substrate, and can be diagnostic for the replacement of a water ligand to the haem iron. This type of spectral perturbation is shown for the binding of palmitoleic acid to P450 BM3 in Figure 3.15 (Girvan, 2003). Also, there is a small, broad peak around 650 nm, which is similar to that seen upon binding of palmitoleic acid to P450 BM3 and is also characteristic of substrate binding to P450 and/or formation of high-spin haem iron. This is highlighted in Figure 3.15.

These features in the UV/Vis spectra suggest that CYP132 may be expressed in *E. coli* bound to a substrate. Lawson *et al.* (Lawson *et al.*, 2004) found that P450 Biol (CYP107H1) is expressed in *E. coli* with palmitic acid bound. As CYP132 is predicted to bind fatty acids, it is possible that it is also expressed with palmitic acid bound. Attempts to extract a bound substrate will be discussed later in this chapter.

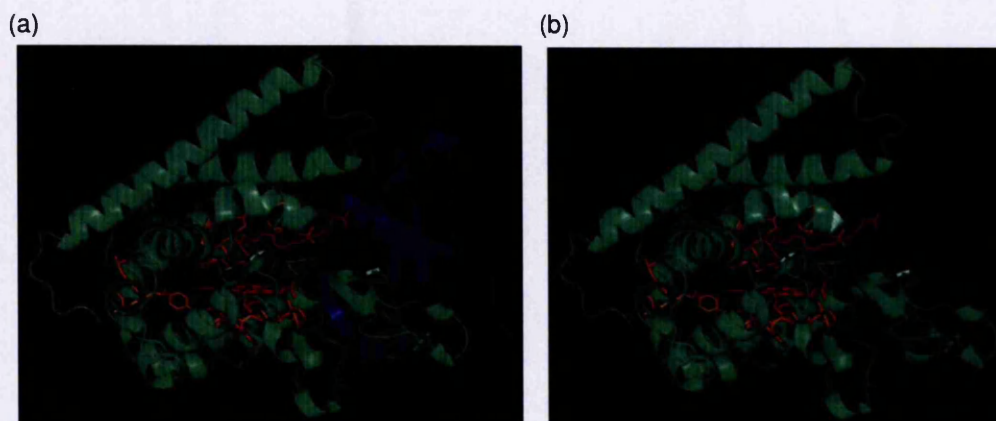


Figure 3.14: Crystal structure of the P450 BM3 haem domain with bound palmitoleic acid. The haem is shown in red, with the palmitoleic acid in magenta. Clusters of three or more conserved residues between P450 BM3 and CYP132 are shown in orange. The N-terminal residues aligned with the CYP132 region that would be lost in order to obtain a mass of 43.1 kDa are shown in blue in panel (a) and are missing in panel (b).

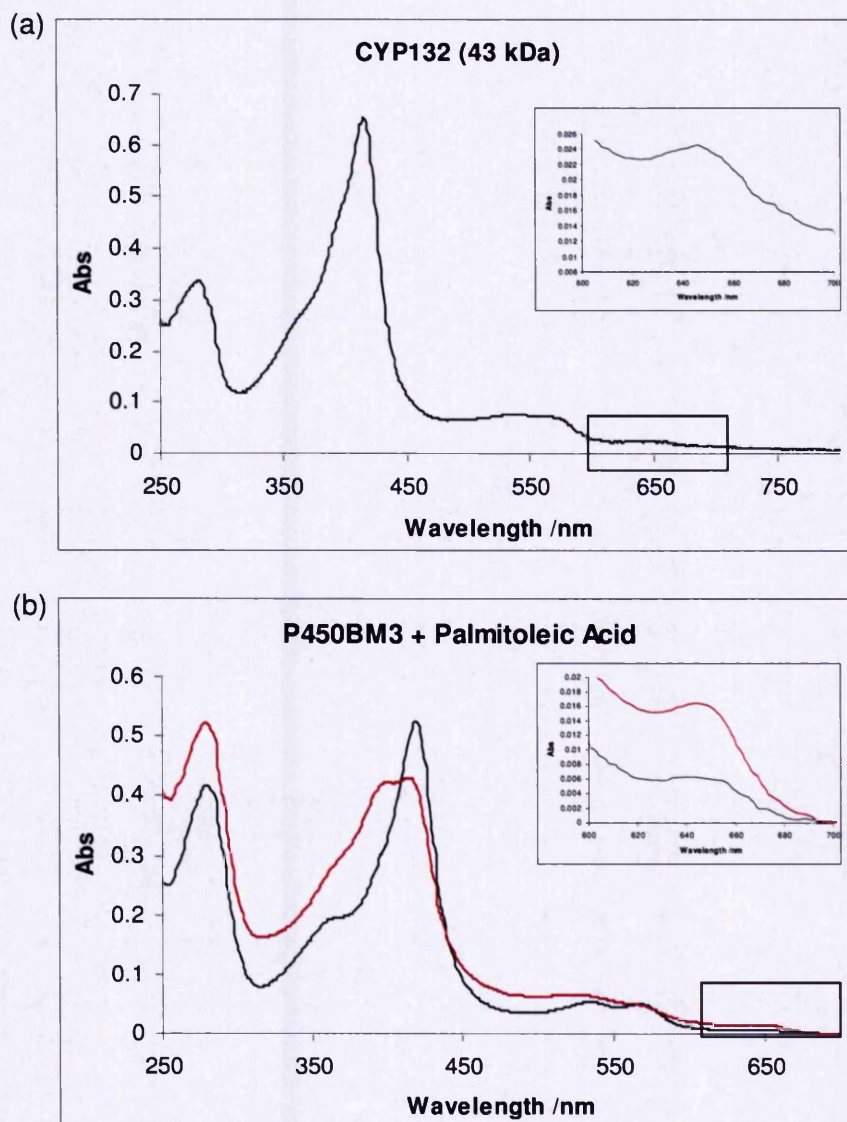


Figure 3.15: (a) UV/Vis spectrum of cleaved CYP132 (~8 μ M). The area from 600-700 nm is highlighted, in which a small peak can be diagnostic of a substrate bound to the P450. Peaks are at 277 (protein), 360 (shoulder, delta band of haem), 395 (shoulder, high-spin ferric form), 415 (haem Soret), 539, and 559 nm (beta and alpha bands, respectively). (b) UV/Vis spectrum of P450 BM3 haem domain (black, ~6 μ M) and bound to palmitic acid (ca 20 μ M, red). The area from 600-700 nm is highlighted, again indicating the substrate-bound feature.

The UV/Vis spectrum of full-length CYP132, as shown in Figure 3.12, appears to be entirely for the low-spin haem species, with no shoulder at 395 nm, or peak at 650 nm that might indicate binding of a substrate. There are a number of possible reasons for

this difference between the spectra of full-length and cleaved CYP132. The more exposed binding site of cleaved CYP132 may bind substrates/ligands that the full-length enzyme does not. Alternatively, the perturbed nature of the cleaved CYP132 haem-binding region may induce a change in the UV/Vis spectrum that is not due to the binding of a substrate (although spectra are suggestive that thiolate coordination is retained).

Due to the small amounts of full length CYP132 available, extensive processing of the protein was not realistic in view of requirements for preliminary spectroscopic and SDS-PAGE/mass spectrometry characterisation using the available material. The true reason for the difference in the UV/Vis spectra of proteolytically cleaved and full-length CYP132 will only be resolved when sufficient quantities of the full-length CYP132 are produced for further study. The preliminary work presented here provides the basis for the successful isolation of the intact CYP132, leading to more detailed structural, spectroscopic and catalytic studies.

3.2.4 Analysis of cleaved CYP132

There were not sufficient quantities of full-length CYP132 for further study. However, examination of cleaved CYP132 could give valuable insights into the structural and spectroscopic properties of CYP132, and will allow important comparisons to be made with full-length CYP132 once sufficient quantities of the latter are produced.

3.2.4.1 Binding of CO and NO

Binding of CO to cleaved, ferrous CYP132 was found to give a shift in the Soret band to predominantly 418 nm, with a small feature at 446 nm. In order to obtain a greater proportion of P450 (i.e. in view of the possibility that agitation of the solution caused by gas bubbling resulted in a greater proportion of P420 formed), CYP132 was added to a buffered solution pre-saturated with CO, under anaerobic conditions, and then a small amount of concentrated sodium dithionite solution was added. This avoids unnecessary agitation of the protein due to bubbling of CO gas through the solution. The anaerobic environment prevents re-oxidation of the enzyme by oxygen in the air. The UV/Vis spectra obtained from this experiment are shown in Figure 3.16.

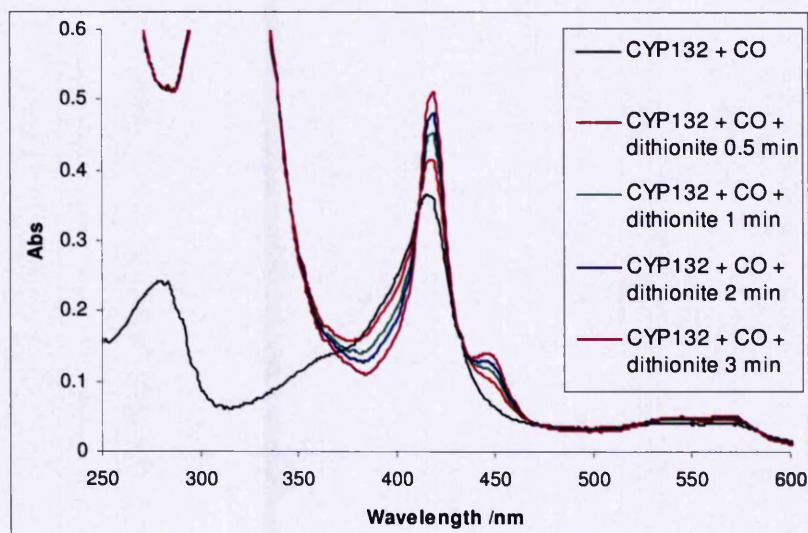


Figure 3.16: Binding of CO to cleaved, ferrous CYP132 under anaerobic conditions. Cleaved CYP132 (ca 4 μM) was first added to a buffered solution saturated with CO gas. A small volume of a concentrated solution of sodium dithionite was then added. Distinctive haem spectral features in the reduced, CO-bound spectrum are at 315, 418, 446, 545 and 568 nm. A small amount of P450 (ca 15 %) is formed along with P420 (ca 85 %). The proportions of the P450/P420 forms remain stable once formed.

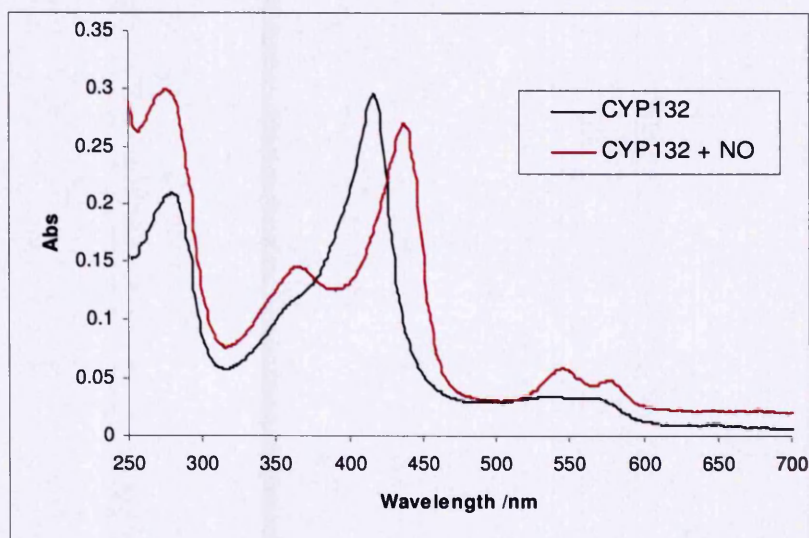


Figure 3.17: Binding of NO to cleaved CYP132 (ca 3.5 μM). Protein was prepared in buffer as described in Figure 3.16. Thereafter, a few bubbles of NO gas were released into the solution to form the Fe(III)NO (formally Fe(II)NO^+) complex. Peaks in the NO-bound spectrum are at 275, 365, 437, 544 and 576 nm.

As can be seen in Figure 3.16, a greater proportion of P450 over P420 is obtained by CO complexation under anaerobic conditions. The major product is still P420, however. The conversion to the 420/450 nm absorption occurs over several minutes, likely as a result of the slow kinetics of CYP132 haem iron reduction by dithionite (followed by relatively rapid binding of CO to the ferrous haem iron). It is notable that there is negligible P450-to-P420 Fe(II)CO complex conversion during this time and under these conditions. Thus, it appears likely that the cleaved CYP132 may exist as a stable equilibrium of species with thiol- and thiolate-coordinated haem iron in the Fe(II)CO complex.

The binding of nitric oxide (NO) to CYP132 gives spectral changes similar to those found for the well characterised P450cam (Franke *et al.*, 2004) and P450 BM3 enzymes (Quaroni *et al.*, 2004), and for the *Mtb* CYP121 P450 (McLean *et al.*, 2002a). The CYP132 NO complex has its Soret band shifted to 437 nm, with prominent shifts also observed in the Q-band region. The shift in the Soret band absorbance to ~435 nm is characteristic of NO-bound P450s. This is compared to e.g. the NO-bound histidine-coordinated nitrophorins (Andersen *et al.*, 2000), in which the Soret band has been shifted to ~420 nm. The UV/Vis spectrum for the NO-bound, cysteinate-coordinated form of nitrophorin (Weichsel *et al.*, 2005) is P450-like, with Soret band at 437 nm.

The fact that a single spectral species is obtained upon the addition of NO to the cleaved CYP132 indicates that the mixture of P450 and P420 forms found upon the addition of CO to ferrous enzyme likely results from partial protonation of the cysteinate ligand. A single species (likely thiolate coordinated) in the NO complex indicates that the resting form of the (cleaved) CYP132 has a unique (thiolate) proximal ligand state. The position of the haem Soret feature in the resting state of the enzyme is consistent with those for several other P450 enzymes and thus also indicative that this form of the enzyme has a proximal thiolate ligand.

3.2.4.2 EPR spectroscopy of cleaved CYP132

Cleaved CYP132 was analysed by electron paramagnetic resonance (EPR) spectroscopy at 10.5 K (Figure 3.18). This revealed a rhombic trio of g-tensor elements,

at 2.47, 2.25 and 1.90. These are typical of a low-spin, ferric haem iron. There were also minor signal at $g = 6.0$, indicating a small amount of high-spin ferric haem, and $g = 4.25$, due to adventitious Fe^{III} ion. Similar levels of high-spin haem iron were also detected in the EPR spectra of CYP121 (McLean *et al.*, 2002a), crystal structures of which did not reveal any bound substrate (Leys *et al.*, 2003), suggesting that CYP132 may not be expressed with a substrate bound after all.

The cleaved CYP132 EPR spectrum is very similar to those of the well characterised P450cam ($g = 2.46, 2.26$ and 1.91) and P450 BM3 ($g = 2.42, 2.26$ and 1.92) (Dawson *et al.*, 1982, Miles *et al.*, 1992), as well as the *Mtb* CYP121 ($g = 2.47, 2.25, 1.90$). The EPR spectra for ferric P450 haems are readily distinguishable from those for proteins which do not have cysteinate coordination of the haem. For instance, values of $g = 2.93, 2.28$ and 1.51 are associated with bis-histidine coordinated low-spin ferric haems in the *Shewanella frigidimarina* NCIMB400 tetraheme flavocytochrome c_3 (Pealing *et al.*, 1995).

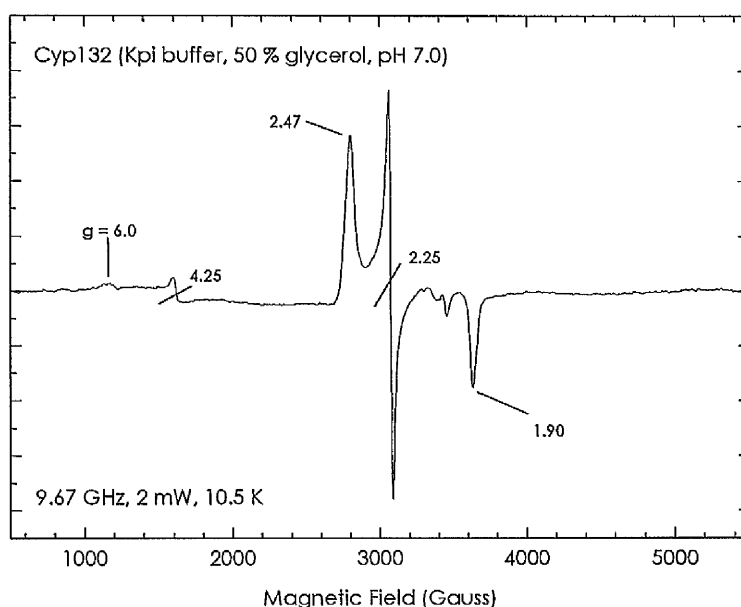


Figure 3.18: EPR spectrum of cleaved CYP132 (200 μM), dissolved in KPi (25 mM), 50 % glycerol, pH 7.0 buffer taken at 9.67 GHz, 10.5 K.

3.2.4.3 Analysis of cleaved CYP132 by circular dichroism

Cleaved CYP132 was analysed by circular dichroism spectroscopy. The far UV (180-260 nm) spectrum is shown in Figure 3.19. This is similar to CD spectra for P450 BM3 and CYP121 (McLean *et al.*, 2002a) and reveals a mainly alpha helical structure, in line with other P450 enzymes characterised in this way (Poulos *et al.*, 1987).

It is likely that the loss of around 90 amino acids from the N-terminal of CYP132 lowers the structural stability of the enzyme. In order to assess the susceptibility of cleaved CYP132 to unfolding, samples of cleaved CYP132 were mixed with increasing concentrations of guanidinium chloride (GdmCl) and incubated at 37 °C for 15 minutes. The far UV CD spectra of the samples were then measured. The "flattening" of the CD spectrum (i.e. the loss of the ellipticity typical of folded protein as shown in Fig. 3.19, top) indicates the unfolding of the protein.

A graph of the circular dichroism against concentration of GdmCl at a wavelength of 223 nm is shown in Figure 3.19 (bottom). This reveals that cleaved CYP132 is almost completely unfolded after mixing with 0.75 M GdmCl. This is a fairly low concentration of GdmCl required for complete unfolding, and this reveals that cleaved CYP132 has a quite low structural stability, and is vulnerable to unfolding. This can be explained by studying the crystal structure of the P450 BM3 haem domain, with the 86 residues at the N-terminal removed, as shown in Figure 3.14. Assuming CYP132 has a similar tertiary structure to the P450 BM3 haem domain, it can be seen that the loss of the residues at the N-terminal could lead to a large cavity in the protein structure. This might allow GdmCl molecules to enter the active site and other internal regions of the protein more readily, and to cause it to unfold much more easily than if the other 86 residues were present. Similar chaotrope stability analysis of full-length CYP132 would likely reveal much higher structural stability than was found for the proteolytically cleaved enzyme.

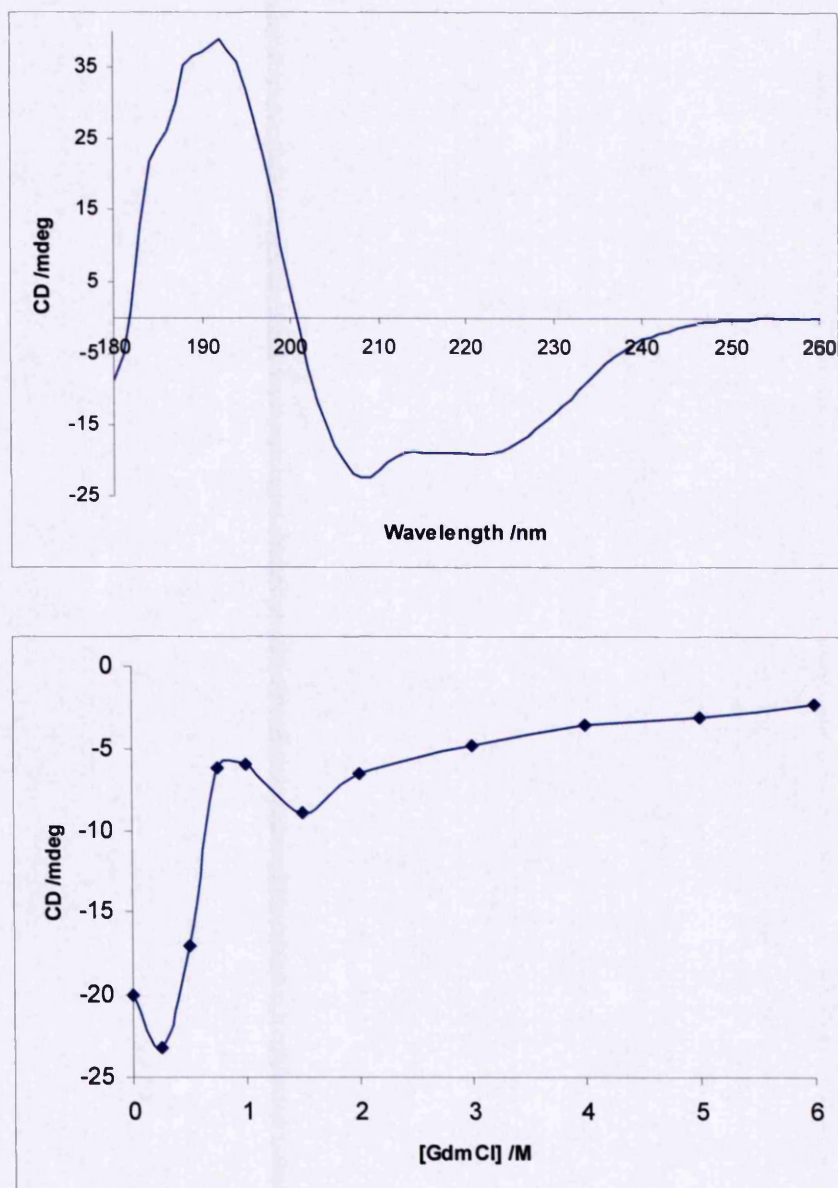


Figure 3.19: (Top) Far UV CD spectra for cleaved CYP132 (8 μ M), recorded at 20 $^{\circ}$ C in KPi (50 mM, pH 7.0). (Bottom) Circular dichroism at 223 nm vs concentration of GdmCl. Samples of cleaved CYP132 (8 μ M) were mixed with GdmCl solution and incubated at 37 $^{\circ}$ C for 15 minutes prior to CD spectra being taken.

3.2.4.4 Attempted removal of bound substrate with solvent

The UV/Vis spectra of cleaved CYP132 suggest that it might be expressed in *E. coli* bound to a substrate. In attempts to determine the true binding constants (i.e. K_d values)

of substrates and inhibitors with CYP132 it is necessary to remove any substrate that is already bound to the expressed protein. Solvents have been used previously to extract bound substrates from P450s (Lawson *et al.*, 2004). Various solvents were added in 50 μ l aliquots to a solution of CYP132 (9 μ M) in Tris (10 mM, pH 7.2) buffer (750 μ l). In Table 3.2 the maximum concentration of each solvent that is tolerated by the protein is tabulated, along with whether the high-spin shoulder is removed by the solvent before denaturation occurs. Denaturation was defined as a significant loss of the wild-type like haem Soret signal, with large increases in the absorbance at 280-350 nm. This assignment was a little arbitrary, but provided a useful, general indication of how well the cleaved CYP132 was able to tolerate each solvent.

Table 3.2: The maximum concentration of various solvents tolerated by CYP132 before denaturation and the effect on the high-spin shoulder in the UV/Vis spectra. Denaturation was defined as a significant loss of the general structure of the UV/Vis spectrum, with a large increase in the absorbance at 280-350 nm.

Solvent	Max. concentration tolerated (% v/v)	HS Shoulder removed (y/n)
Acetonitrile	14	y
Acetone	17	n
Chloroform	12	y
Ethyl acetate	6	n
Ether	>50	n
Ethanol	17	n
DMSO	40	y
DCM	6	y
Methanol	25	y
n-butanol	6	n

It can be seen in Table 3.2 that the only solvents that both remove the high-spin shoulder from CYP132 and are tolerated in high concentrations are DMSO and methanol. The effects on the UV/Vis spectra are shown in Figure 3.20.

A larger sample of CYP132 (200 μ l, 0.8 mM) was treated with 25 % DMSO. The high-spin shoulder was observed to have been removed by UV/Vis spectrometry. The solution was then dialysed back into Buffer A. It was hoped that any bound substrate would have been removed with the solvent during dialysis. UV/Vis spectrometry showed the absorption shoulder to have returned after dialysis, however. The absorption shoulder could be removed once more by the addition of DMSO.

If a bound substrate (prospectively a fatty acid) is present, it is likely that it is not very soluble in water. If this is the case, it is unlikely that it would dialyse into buffer. Instead it is more likely to rebind to the protein. To counteract this problem, a solution of CYP132 in 25 % DMSO was dialysed into 25 % DMSO for 3 hours. It was then dialysed back into Buffer A. UV/Vis spectrometry of the dialysed product showed it to be identical to the starting material.

Diethyl ether (ether) was also found to be tolerated by CYP132 in high concentrations, but it did not remove the haem high-spin shoulder. This was possibly due to the high immiscibility of ether with water preventing it from getting into the active site of the protein and removing any substrate. It was thought that ether might be able to extract the substrate from 25 % DMSO, however. A solution of CYP132 in 25 % DMSO was extracted with ether. However, subsequent analysis of the protein product by UV/Vis spectrometry showed the high-spin shoulder to remain. Thus, data collected to this point indicated that a "natural" balance of high-spin/low-spin haem iron might be present in the CYP132 sample in aqueous buffer, or else the solvents used did not effectively remove any bound substrate, but did perturb the haem optical signal.

It was next attempted to remove any bound substrate by gel filtration of a DMSO treated sample. A solution of CYP132 (25 μ M) in 25 % DMSO (1 ml) was loaded onto a BioRad 10 ml desalting column. UV/Vis spectrometry of the eluent after dialysis back into buffer showed the high-spin shoulder to remain, again suggesting that high-spin optical features were either a natural feature of the enzyme, or that displacement of any substrate-like molecule was inefficient.

Methanol was also found to be tolerated by CYP132 in fairly high concentrations and to remove the high-spin shoulder. A solution of CYP132 in 10 % methanol was dialysed into 10 % methanol for 2 hours, followed by dialysis into Tris bufer (10 mM, pH 7.2). The high-spin shoulder was also found to remain following this treatment.

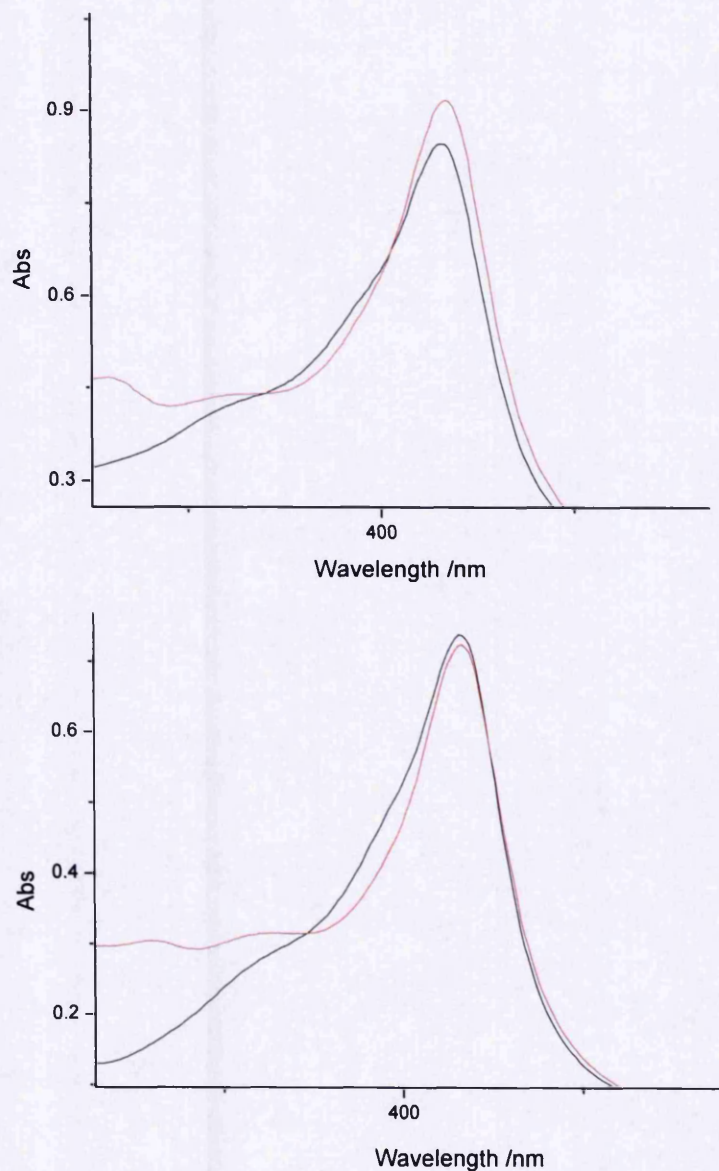


Figure 3.20: UV/Vis spectra showing the effect of the addition of solvents to cleaved CYP132, in order to attempt to remove any bound substrate. The black lines correspond to CYP132 alone ($\sim 8 \mu\text{M}$), in Tris buffer (10 mM, pH 7.2) and the red line with solvent added. Peaks of samples containing solvent have been adjusted to account for dilution and turbidity effects. The top panel shows effects of addition of 21 % DMSO. The bottom panel shows effects of addition of 25 % methanol.

It was considered that the most likely substrate to have been co-purified with CYP132 was palmitic acid (the major fatty acid in *E. coli*, and representative of a class of substrates that bind to CYP132's closest eukaryotic P450 relatives – the CYP4 enzymes). Solubility tests with commercially available palmitic acid showed it to be

insoluble in either 25 % DMSO or 10 % methanol. The fact that the high-spin shoulder is removed by the addition of these concentrations of solvent, yet upon dialysis back into buffer the high-spin shoulder returns, suggests that the solvents might themselves be interacting directly with the P450 haem to perturb its optical spectrum, rather than acting to remove a bound substrate. If there is a substrate bound to the expressed CYP132, it seems evident that attempts to remove it with a selection of solvents were not successful, and thus another approach should be sought.

3.2.4.5 Attempted analysis of bound substrate by mass spectrometry

Extraction of the potential bound substrate for mass spectrometry with DCM was attempted following the method set out by Lawson *et al.* (Lawson *et al.*, 2004). A sample of CYP132 (80 μ M, 2ml) in high salt Buffer A, (containing 500 mM KCl), was acidified with HCl (0.15 M, pH 1.3). It is necessary for the solution to be acidified in order to protonate the fatty acid, causing it to be uncharged, and thus soluble in the organic solvent. The now turbid suspension was extracted with DCM (2 ml), the organic layer removed and evaporated to dryness under vacuum. The residue was dissolved in methanol and analysed by mass spectrometry. However, it was not possible to identify any bound substrate.

The repeated failure to remove the high-spin shoulder by treatment with solvents, and to identify any substrate by mass spectrometry, brought into doubt the presence of any lipid substrate co-purified with the protein. The UV/Vis spectrum of full-length (intact) CYP132 does not contain high-spin haem optical features indicative of any bound substrate (i.e. it is predominantly low-spin). It is thus concluded that the high-spin shoulder at 395 nm and small peak at 650 nm in the UV/Vis spectra of cleaved CYP132 are a result of perturbations in the protein structure induced by the loss of ~90 amino acids from the N-terminal, and are not due to any bound substrate.

3.2.4.6 Binding of imidazole

Various azole anti-fungal drugs, including econazole, clotrimazole and ketoconazole have been shown to inhibit the growth of *Mycobacterium smegmatis*, a fast-growing mycobacterial relative of *Mtb* (McLean *et al.*, 2002b). Econazole was recently shown to

be effective in clearing *Mtb* from mice (Ahmad *et al.*, 2006) and was found to be able to replace more traditional treatments (rifampicin and isoniazid) in successful treatment of murine *Mtb*. *Mtb* CYP51B1 (Guardiola-Diaz *et al.*, 2001) and CYP121 (McLean *et al.*, 2002a) have been shown to bind these drugs tightly. The K_d values for the binding to CYP121 were found to be in line with the MIC values of the drugs for *M. Smegmatis*, leading to the suggestion that CYP121 may be a true target for the azole drugs (McLean *et al.*, 2002b).

As was discussed earlier, CYP132 has been found to be important to the virulence of *Mtb* in mice and may be a useful drug target in the fight against *Mtb* (Recchi *et al.*, 2003). The first step in the development of possible treatments targeting CYP132 is to study the interaction of azole inhibitors, in order to gauge the specificity of the CYP132 active site for particular azole drugs and to determine the K_d values for their binding.

Imidazole is the functional group that binds the iron in a number of the azole antifungal drugs. The binding of imidazole alone to cleaved CYP132 was first investigated in order to gauge (by comparison) the effects of the remaining “scaffold” structures of the clinically used azole drugs on their affinity of binding to CYP132.

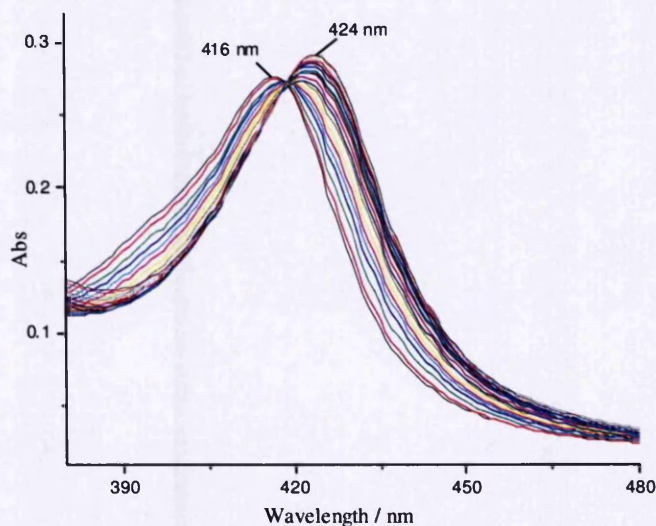


Figure 3.21: UV/Vis spectra showing the change in Soret band absorbance from 416 nm to 424 nm with the addition of imidazole to cleaved CYP132 (3 μ M).

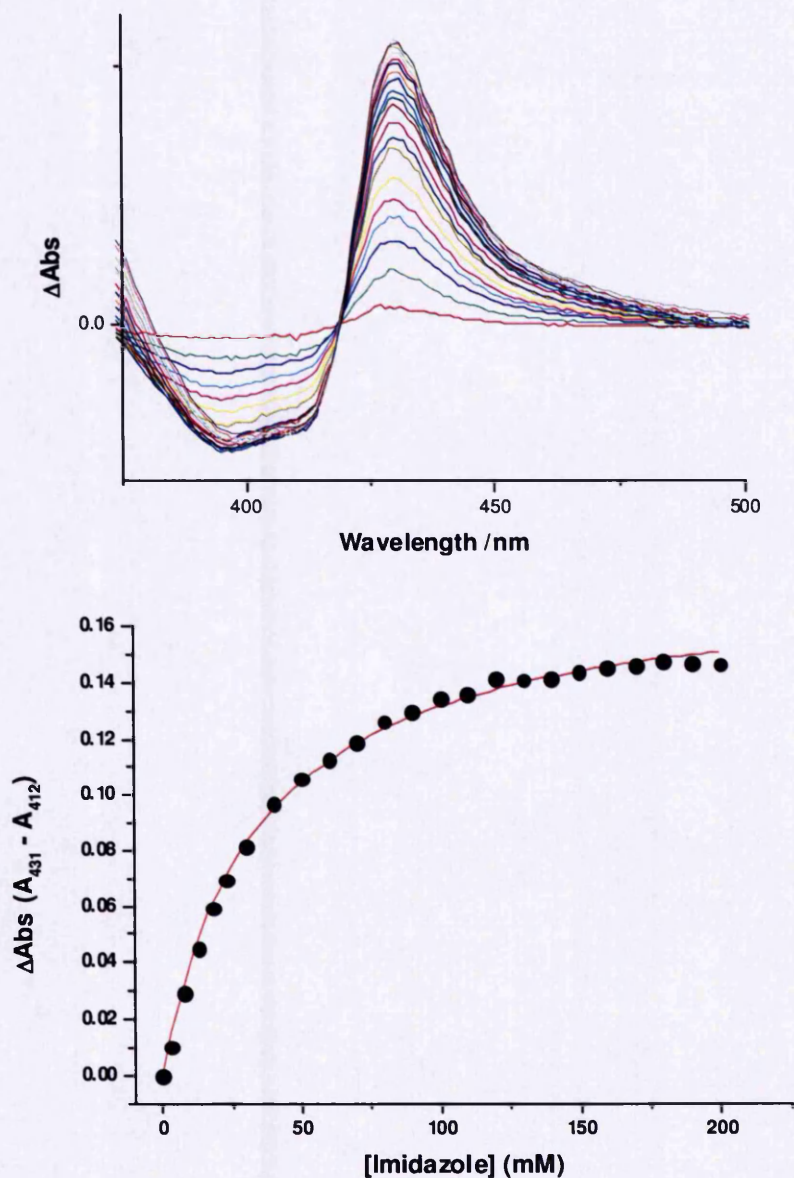


Figure 3.22: Graphs used in the calculation of the K_d value for the binding of imidazole to the cleaved form of CYP132. The top graph shows difference spectra for the titration of cleaved CYP132 with imidazole, and for imidazole concentrations ranging from 0 to 200 mM. Difference spectra were generated by subtraction of the absolute spectrum for the ligand-free form of the protein from each successive absolute spectrum collected during the imidazole titration. The bottom graph shows imidazole-induced absorption change (computed by subtracting the absorption value at the difference spectrum minimum from that at the maximum for each individual spectrum, using the values at 395 nm and 430 nm, respectively) plotted against [imidazole]. Data points are fitted using a hyperbolic function, described in the text, yielding an imidazole K_d value of 34.9 ± 1.2 mM.

Cleaved CYP132 was titrated with imidazole, as described in Chapter 2. This gave a type II shift in the Soret band from 416 nm to 424 nm. This is shown in Figure 3.21. Difference spectra were obtained by subtracting the original (ligand-free) spectrum from that obtained following each successive imidazole addition. The difference between the maximum at 430 nm and the minimum at 395 nm (in each difference spectrum) was plotted against the relevant imidazole concentration, and the data were fitted using equation 3.1, where P_1 is the maximum absorption at ligand saturation and P_2 is the K_d . Parameter "y" is the absorption change at ligand concentration "x". This gave a K_d for imidazole of 34.9 ± 1.2 mM.

$$y = \frac{P_1 x}{P_2 + x} \quad (3.1)$$

3.2.4.7 The binding of azole anti-fungal drugs

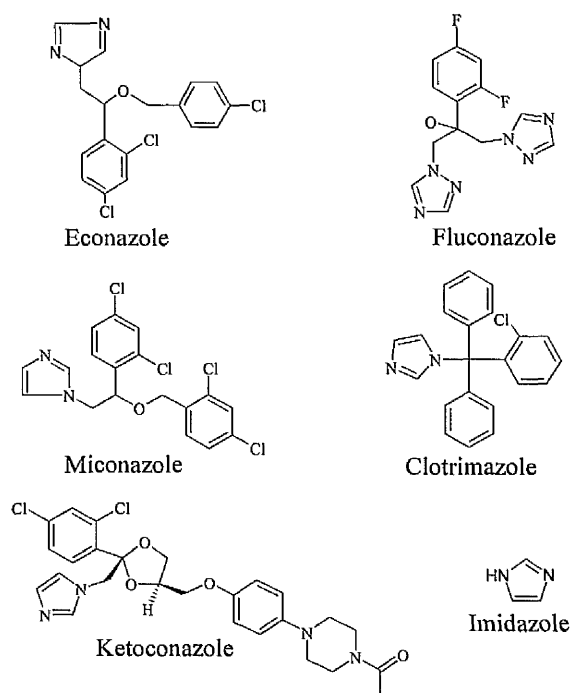


Figure 3.23: Structures of azole inhibitors titrated with CYP132

The azole inhibitors econazole, clotrimazole, ketoconazole, miconazole and fluconazole were titrated against CYP132, using the same procedure as that used for imidazole (see

Figure 3.23 for structures of the azoles). Due to the tight binding of the azole anti-fungals, it was necessary to fit the data using the quadratic equation 3.2, which incorporates the additional parameter E_t , which is the concentration of the enzyme. The K_d values obtained are shown in Table 3.3, along with the K_d values for the binding to *Mtb* CYP51B1 and CYP121 and the MIC values of these drugs for *M. smegmatis*, taken from (McLean *et al.*, 2002b).

$$y = \frac{A_{\max}}{2E_t} * \left((x + E_t + K_d) - \sqrt{(x + E_t + K_d)^2 - 4x E_t} \right) \quad (3.2)$$

The data in Table 3.3 show that the binding of the azole anti-fungal drugs tested is weaker for cleaved CYP132 than for CYP121 or CYP51B1. The binding of fluconazole is especially weak, with a K_d value of $719 \pm 16 \mu\text{M}$. It is not clear whether the K_d values for cleaved CYP132 are a true reflection of the drugs' affinities for binding to the full-length enzyme, or if the loss of part of the substrate recognition portion of CYP132 has led to a decrease in the strength of binding to cleaved CYP132. The fact that cleaved CYP132 binds tightly to four of the five azole anti-fungals tested, with only slightly weaker binding than that of CYP121 or CYP51B1, further confirms that the haem region of the truncated enzyme is intact. The weak binding of fluconazole is in line with the high MIC value of fluconazole for *M. smegmatis*. This is in contrast with the tighter binding of fluconazole to CYP121 and CYP51B1 (although this is the weakest binding azole for each of these P450s). These data confirm CYP132 as a possible candidate mycobacterial drug target for these anti-fungals.

Table 3.3: The K_d values for the binding of azole anti-fungal drugs to cleaved CYP132, along with the MIC values for *M. smegmatis* and the K_d values for *Mtb* CYP121 and CYP51B1. Values for CYP132 are from this study, other values are taken from (McLean *et al.*, 2002b)

Drug	MIC value for <i>M. smegmatis</i> (μM)	K_d value (μM)		
		<i>Mtb</i> CYP132 (cleaved)	<i>Mtb</i> CYP121	<i>Mtb</i> CYP51B1
Econazole	<0.2	9.0 ± 0.6	<0.2	1.2 ± 0.1
Clotrimazole	0.3	<1.3	<0.2	0.18 ± 0.02
Miconazole	2.6	2.7 ± 0.2	<0.2	0.91 ± 0.10
Ketoconazole	38	9.2 ± 0.7	3.3 ± 0.3	2.03 ± 0.08
Fluconazole	>325	719 ± 16	9.7 ± 0.2	5.82 ± 0.50

3.2.4.8 Titrations with putative fatty acid substrates

CYP132 has a higher level of amino acid similarity to various CYP4a fatty acid hydroxylase P450s than to other P450 enzymes. This suggests that CYP132 may also bind fatty acids. The range of fatty acids that can be simply titrated with P450s is limited by their poor solubility in water, due to the molecules' long hydrophobic chains. Cleaved CYP132 was titrated with the saturated fatty acids lauric (C_{12}), myristic (C_{14}) and palmitic (C_{16}) acid, as well as with the unsaturated fatty acids arachidonic (C_{20}), palmitoleic (C_{16}) and myristoleic (C_{14}) acid (Figure 3.24). No change in the UV/Vis spectra of cleaved CYP132 was observed upon the addition of the saturated fatty acids, within the limits of solubility, except for increased turbidity.

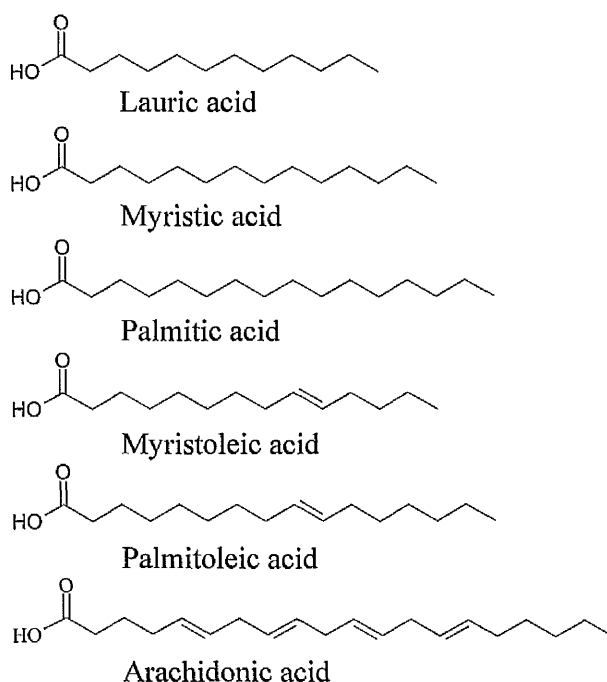


Figure 3.24: Structures of fatty acids titrated with CYP132

It was found that a gradual change in the UV/Vis spectrum of cleaved CYP132 occurred when the unsaturated fatty acids were added at a concentration of 120 μM or more, using a CYP132 concentration between 5-10 μM . This is shown in Figure 3.25 for the addition of arachidonic acid (330 μM) to CYP132 (5 μM) at 30 $^{\circ}\text{C}$. The wavelength

maximum of the Soret band was observed to shift from 416 nm to 419 nm over around 15 minutes, before decreasing back to 416 nm over the next 30 minutes. During this time the absorbance of the Soret band decreased, with an increase in the absorbance at 365 nm. Similar behaviour was observed with palmitoleic and myristoleic acids. The rate at which the change occurred increased with temperature, but a slow change was still observed at 10 °C. The process also occurred under anaerobic conditions.

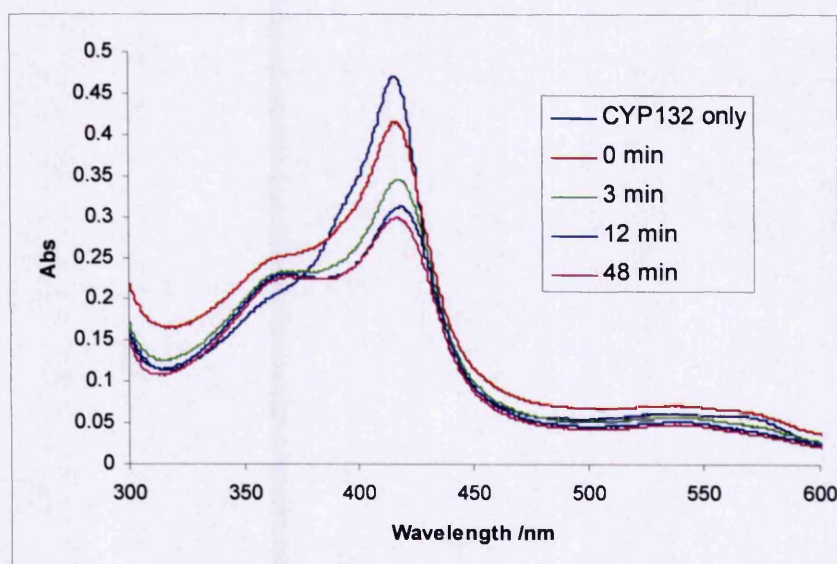


Figure 3.25: UV/Vis spectra of CYP132 (5 μM) and arachidonic acid (330 μM) in MOPS buffer (25 mM, pH 7.4) at 30 °C, and at various time intervals after the addition of arachidonic acid. The Soret band absorbance was maximal at 416 nm initially, shifting to 419 nm over 12 min, before decreasing back to 416 nm over the following 30 min. Absorbance at 416–419 nm decreases in this time, with increased absorbance at 365 nm (possibly reflecting haem dissociated from the protein matrix). The spectra remained constant after 50 min for up to several hours, until aggregation of the protein caused turbidity in the solution.

It was found that when fatty acid concentrations below $\sim 80 \mu\text{M}$ were used, the gradual decrease in the Soret band absorbance was not observed. The need for such a large excess of fatty acid in order for the process to occur suggested that the observed spectral changes were not due to specific effects of the fatty acid binding to the CYP132 active site, but were instead due to the detergent effect induced by having a large quantity of unsaturated fatty acid in the solution. This was confirmed when similar changes in the UV/Vis spectra over time were observed when 300 μM SDS was used instead of fatty acids. This is shown in Figure 3.26.

The origin of the observed loss in intensity of the Soret band in the UV/Vis spectra of cleaved CYP132 is not clear. It is likely that the detergent causes the protein to partially unfold, resulting in a loss of P450 character. The reason cleaved CYP132 is especially susceptible to detergent in the solution is probably the loss of the N-terminal amino acids disrupting the folding of the protein. Studying Figure 3.14 (b) it can be seen that the loss of ~90 amino acids from the N-terminal of the P450 BM3 haem domain may leave a large cavity (assuming compensatory structural rearrangements do not occur), through which detergent molecules could enter to facilitate further disruption of the haem environment and/or some unfolding of the protein.

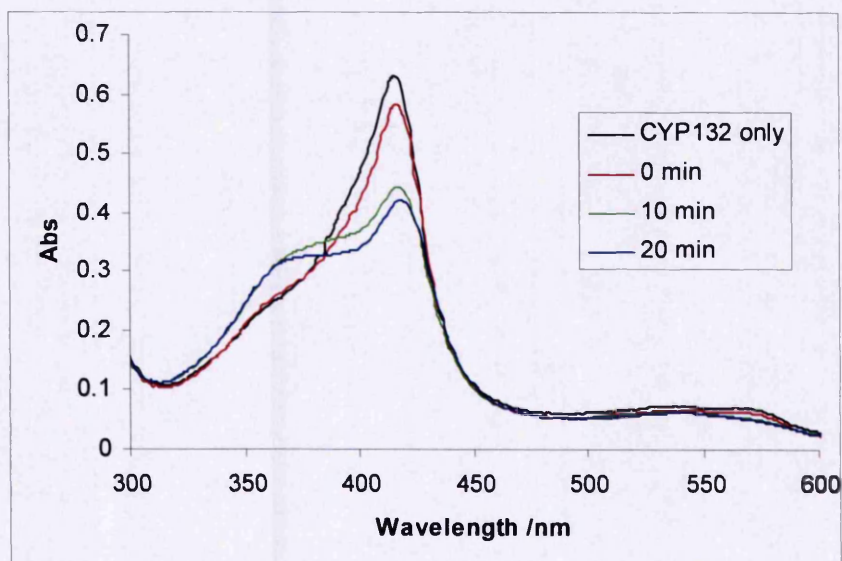


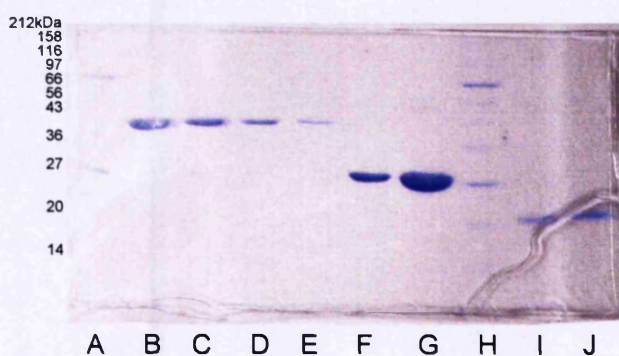
Figure 3.26: UV/Vis spectra of CYP132 (7 μ M) and SDS (300 μ M) in MOPS buffer (25 mM, pH 7.4) at 30 °C, at various time intervals after addition of SDS. The Soret band absorbance was maximal initially at 416 nm, shifting to 418 nm over 20 min. The absorbance at 416–418 nm decreased in this time, with an increase in the absorbance at 365 nm (possibly reflecting haem dissociated from the protein matrix).

3.2.4.9 Turnover kinetics of cleaved CYP132 with fatty acids

The shift to the high-spin state in the UV/Vis spectra, indicative of type I binding of substrate-like molecules, was not observed upon the addition of fatty acids to CYP132. Changes in the spectra, possibly due to detergent effects, were observed when high concentrations of unsaturated fatty acids were added. It was considered possible that

any spectral changes due to type I binding of unsaturated fatty acids to CYP132 were masked by the changes due to their detergent effects. Alternatively, the fatty acids may have bound to CYP132 without displacing the bound water molecule from the haem iron, and thus did not induce a type I spectral shift. In order to determine whether cleaved CYP132 was able to turn over fatty acids, despite apparent lack of type I binding, the kinetics of turnover of CYP132 with a range of fatty acids was investigated using a heterologous redox partner system.

In order to turn over substrates, most P450s require redox partner enzymes to provide the electrons for P450-dependent oxygen activation. The *E. coli* flavodoxin (FLD) and flavodoxin reductase (FLDR) proteins were used to transfer electrons from NADPH to CYP132. This system has been used successfully for catalysis with other P450s (Jenkins and Waterman, 1998). FLD and FLDR were prepared according to the protocol set out by McIver *et al.* (McIver *et al.*, 1998). SDS-PAGE gel analysis of the cleaved CYP132, FLD and FLDR proteins used in the kinetic studies is shown in Figure 3.27.



A = Protein marker
B-E = CYP132
F-G = *E. coli* FLDR
H = Protein marker
I-J = *E. coli* FLD

Figure 3.27: SDS-PAGE gel of the enzymes used in the investigation of the kinetics of turnover of cleaved CYP132 with fatty acids. Protein marker was from NEB, with relevant molecular mass data indicated (kDa). Lanes B-E contain CYP132 at 2, 1, 0.5 and 0.25 μ g. Lanes F and G contain 2 and 4 μ g of FLDR, while lanes I and J contain FLD at 0.25 and 0.5 μ g, respectively.

NADPH absorbs strongly at 340 nm with an extinction coefficient of $6.21 \text{ mM}^{-1}\text{cm}^{-1}$. This absorbance is lost as NADPH is oxidised to NADP^+ . This process would occur if FLDR oxidised the NADPH, passing electrons to CYP132 via the FLD protein. The rate of turnover of CYP132 was calculated by measuring the rate of absorbance decrease at 340 nm (compared to a background rate in samples not containing P450 and/or substrate) and calculating the corresponding quantity of NADPH oxidised. CYP132, FLD, FLDR and NADPH were used at a ratio of 1:5:5:20. The fatty acids were used in excess in order maintain steady-state conditions and to prolong any catalytic reaction for a sufficient period to enable accurate measurement of rate of NADPH oxidation.

As was discussed above, a decrease in absorbance at 416 nm occurs with the addition of unsaturated fatty acids to cleaved CYP132, but there is little change in the absorbance at 340 nm. By following the change in the haem absorbance at 340 nm following the addition of palmitoleic acid to cleaved CYP132, it was found that it was necessary to allow 2 minutes for the haem-dependent absorbance at 340 nm to equilibrate. This equilibration period was allowed prior to addition of NADPH to initiate any catalytic reaction dependent on CYP132.

The rate of absorbance decrease at 340 nm for a solution of FLD ($37.5 \text{ }\mu\text{M}$), FLDR ($37.5 \text{ }\mu\text{M}$) and NADPH ($150 \text{ }\mu\text{M}$) at $30 \text{ }^\circ\text{C}$ was measured over one minute in duplicate as the background rate (Figure 3.28, black). This was repeated with the addition of CYP132 ($7.5 \text{ }\mu\text{M}$) (Figure 3.28, red). A solution of FLD ($37.5 \text{ }\mu\text{M}$), FLDR ($37.5 \text{ }\mu\text{M}$), CYP132 ($7.5 \text{ }\mu\text{M}$) and palmitoleic acid (1.6 mM) was prepared. The absorbance at 340 nm was monitored following fatty acid addition. Once the absorbance had become stable NADPH ($150 \text{ }\mu\text{M}$) was added and the change in absorption at 340 nm was measured over 1 minute. This process was repeated with myristoleic acid (1.6 mM) and arachidonic acid (1.65 mM and 3.3 mM) as potential lipid substrates for CYP132. The experiments were carried out in duplicate in each case.

Graphs of the absorbance change at 340 nm (reflecting NADPH oxidation) against time after the addition of arachidonic acid, palmitoleic acid and myristoleic acid to cleaved CYP132, *E. coli* FLD and FLDR and NADPH, as well as controls without fatty acid substrate added, are shown in Figure 3.28. The average ΔA_{340} initial gradients with each

fatty acid, converted to the number of turnovers of CYP132 per minute (i.e. mol of NADPH oxidised), are tabulated in Table 3.4.

It can be seen in Table 3.4 that palmitoleic and myristoleic acids stimulate CYP132-dependent NADPH oxidation at rates of 13 and 11 turnovers per minute (expressed relative to the concentration of CYP132) respectively. This is a fairly low rate for a P450 enzyme. P450cam oxidises d-camphor at a rate greater than 1000 turnovers per minute per enzyme (Watanabe, 2002). It is not unusually low for the reaction with a substrate that is not the natural target of the enzyme, however. For example, Bell *et al.* found turnover rates of 81.5 and 4.0 turnovers per minute per enzyme for the oxidation by P450cam of (+)- α -pinene and (S)-limonene, respectively (Bell *et al.*, 2001).

Although this is not a particularly high rate for a P450 enzyme, this enhanced turnover appears to confirm that CYP132 binds and reacts with these unsaturated fatty acids. The rate of turnover with arachidonic acid was lower. Using a similar lipid concentration to that used with palmitoleic and myristoleic acids (1.65 mM), an increase in NADPH oxidation rate of only ~1 turnover per minute was observed compared to CYP132 and redox partners alone. Increasing the concentration of arachidonic acid to 3.3 mM gave a rate of ~2 turnovers per minute. Repeating the reactions using lauric and palmitic acids did not give any significant increase in the rate of oxidation of NADPH.

It would appear from the above results that CYP132 may be able to oxidise unsaturated fatty acids, without appearing to give any notable type I spectral shift when mixed with them in its oxidised form. As discussed previously, it is possible that type I binding was masked by the detergent effects on the CYP132 spectrum. However, examination of the spectra in Figure 3.25 suggests that this is not the case. It is more likely that cleaved CYP132 binds the unsaturated fatty acids tested, but without giving a shift to the high-spin state. This may be because the bound water ligand is not displaced upon binding of the fatty acids, possibly as a result of the more exposed substrate binding region in the cleaved protein.

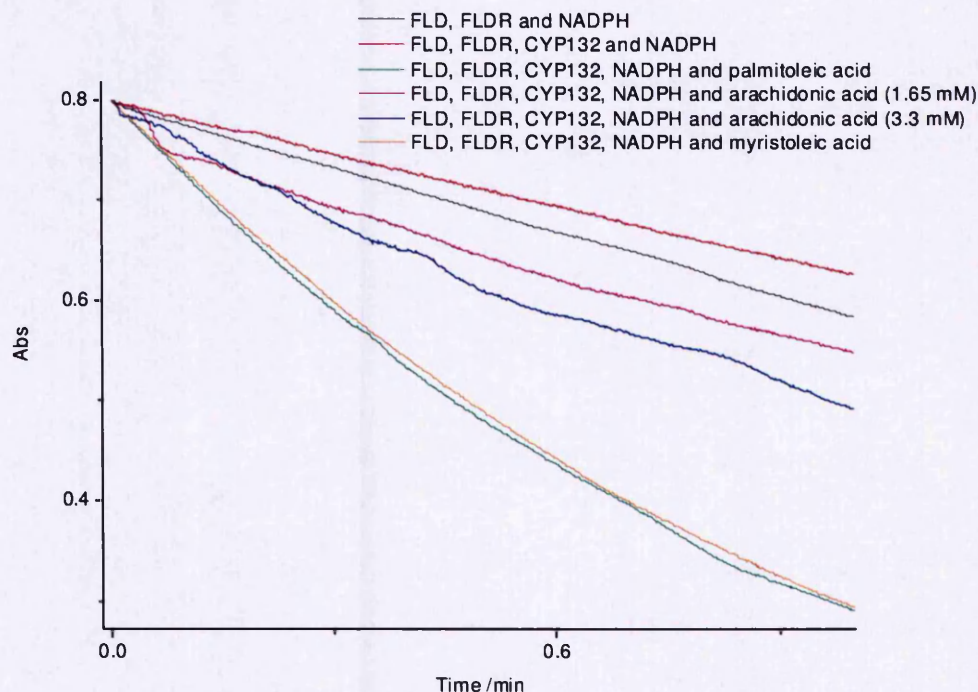


Figure 3.28: Absorbance at 340 nm measured over 1 minute after the addition of NADPH (150 μ M) to solutions of *E. coli* FLD (37.5 μ M) and FLDR (37.5 μ M) in MOPS buffer (25 mM, pH 7.4) at 30 $^{\circ}$ C containing a mixture of CYP132 (7.5 μ M), palmitoleic acid (1.6 mM), myristoleic acid (1.6 mM) and arachidonic acid (1.65 mM and 3.3 mM) according to the key. Samples containing fatty acid were allowed to equilibrate for 2 minutes prior to addition of NADPH. Graphs have been normalised to have the same starting absorbance.

The fact that cleaved CYP132 may be able to oxidise unsaturated fatty acids (and that these fatty acids can stimulate NADPH oxidation using a redox chain of which CYP132 is the terminal component) is further evidence that the P450 character is conserved in CYP132, despite the loss of a large portion of the protein from the N-terminal. The rate of turnover with myristoleic and palmitoleic acids is considerably higher than with arachidonic acid. This may be because CYP132 favours shorter chain fatty acids, or perhaps the more constrained structure of arachidonic acid, caused by the higher number of double bonds, clashes with residues in the substrate binding region of CYP132, preventing it from entering the binding site or from adopting a catalytically relevant binding mode.

Table 3.4: Initial rate of A_{340} decrease (reporting NADPH oxidation) for solutions containing a mixture of *E. coli* FLD (37.5 μ M), FLDR (37.5 μ M), CYP132 (7.5 μ M), palmitoleic acid (1.6 mM), or myristoleic acid (1.6 mM), or arachidonic acid (1.65 mM and 3.3 mM) and NADPH (150 μ M) in MOPS buffer (25 mM, pH 7.4) at 30 °C. Turnover number for CYP132 is calculated as the increased rate of NADPH oxidation (in mM NADPH oxidised/min, see relevant column) over that determined for the redox partners (FLDR/FLD) and CYP132 only, and expressed as mol NADPH oxidised per mol CYP132.

Contents of assay solution	Average gradient ($\Delta A_{340}/\text{min}$)	Average rate of NADPH oxidation (mM/min)	Average number of turnovers of CYP132/min
FLD, FLDR, NADPH	-0.2	0.03	
FLD, FLDR, CYP132, NADPH	-0.2	0.03	
FLD, FLDR, CYP132, palmitoleic acid	-0.8	0.1	13
FLD, FLDR, CYP132, myristoleic acid	-0.7	0.1	11
FLD, FLDR, CYP132, arachidonic acid (1.65 mM)	-0.2	0.04	1
FLD, FLDR, CYP132, arachidonic acid (3.3 mM)	-0.3	0.05	2

3.2.4.10 Analysis of the products of the turnover studies by mass spectrometry

Kinetic studies on the turnover of fatty acids by cleaved CYP132 (together with a heterologous redox partner system) showed an increase in the rate of oxidation of NADPH following the addition of palmitoleic acid or myristoleic acid. In order to investigate the processes further, the products of the reactions were analysed by liquid chromatography – tandem mass spectrometry (LC/MS).

The reactions studied above were carried out in a larger scale in order to identify the product by mass spectrometry. Glucose and glucose dehydrogenase were used to regenerate NADP^+ back into NADPH at the expense of glucose. The reactions were carried out at 30 °C for 16 hours. Controls were also run containing only glucose, glucose dehydrogenase, NADPH and each of the fatty acids.

The HPLC chromatogram, along with the mass spectra of the important product fractions from the turnover of myristoleic acid, is shown in Figure 3.29. The chromatogram and mass spectra for the myristoleic control are shown in Figure 3.30. The predicted identities of the species responsible for the spectra are shown alongside the mass spectra in the two figures.

The large peaks on the LC chromatogram, with $m/z = 225$, are due to intact myristoleic acid. The chromatogram from the turnover of myristoleic acid by CYP132 contains two peaks, at 10 and 12 minutes, each with $m/z = 241$. There is also a small peak at 10 minutes, with $m/z = 241$, for the myristoleic acid control, without CYP132. It is likely that the peaks at 10 minutes are due to epoxidation of the myristoleic acid double bond, and the peak at 12 minutes due to hydroxylation of myristoleic acid. The low intensity of the 10 minute peak in the chromatogram from the control reaction shows that epoxidation of myristoleic acid without a catalyst present is slow. The peak from the reaction with CYP132 present is much larger. This suggests that CYP132 catalyses both epoxidation and hydroxylation of (at least) unsaturated fatty acids.

There are another three large peaks in the LC chromatogram from the products of the turnover of myristoleic acid by CYP132 that are not in the control sample. The peaks at ~4 and 5 minutes were analysed by MS/MS. These are shown in Figure 3.31. The masses of these peaks were $m/z = 183$ and 197 . Further fragmentation of the ions gave losses of 28 and 44 mass units from each species. These are most likely due to the loss of C_2H_4 or CO , and CO_2 . The origin of these peaks is not clear. The loss of C_2H_4 or CO , and CO_2 suggests that they contain an acid group. The difference in mass between the two peaks is 14 mass units, suggesting they differ by one CH_2 group. The most sensible assignments for the identity of the $m/z=197$ and 183 peaks is the loss of 2 and 3 carbon atoms, and their associated hydrogen atoms, respectively, from the fatty acid. These are shown beneath the mass spectra in Figure 3.31.

The LC/MS chromatograms for the reaction with palmitoleic acid, and the control reaction, are shown in Figure 3.32. These contained the same features as those for myristoleic acid. Masses are not available for the peaks at 3 and 4 minutes as these were overlooked when carrying out the mass spectrometry. The fact they eluted from the HPLC column slightly earlier than the similar peaks on the myristoleic acid chromatogram shows that they are different structures, meaning they must be derived from the fatty acids, and are not a result of contamination from the other species in the reaction.

The species that came off the HPLC column at 3-5 minutes are particularly interesting. If the assignment for the identity of the species responsible for these peaks is correct, then CYP132 is catalysing the removal of the ω -1 and ω -2 carbon atoms of the fatty acids. This is an uncommon reaction for P450s as it does not involve the addition of an oxygen atom. Although further analysis is required to confirm the products of the reactions, it would appear that CYP132 is able to carry out hydroxylation, epoxidation, and C-C bond cleavage reactions, all on the same substrate.

Further analysis of the various species produced by the turnover of myristoleic acid and palmitoleic acid, by CYP132, would be necessary to determine their identity with any certainty. Whatever their origin, it appears clear that CYP132 catalyses a number of reactions on the unsaturated fatty acids. This further confirms the activity of cleaved CYP132. Also it begins to shed light on the possible role of CYP132 in *Mtb*. If CYP132 carries out a number of reactions on the same molecule, it seems unlikely that it should be involved in the synthesis of a larger biomolecule. In such a case, only one of the products would be beneficial, while the others were wasted. It would be more likely that CYP132 is involved in the digestion of fatty acids to produce more soluble products, or possibly in the neutralisation of toxins or lipid mediators produced by the immune system of the host organism. However, more work is clearly needed prior to any firm conclusion being reached in this area.

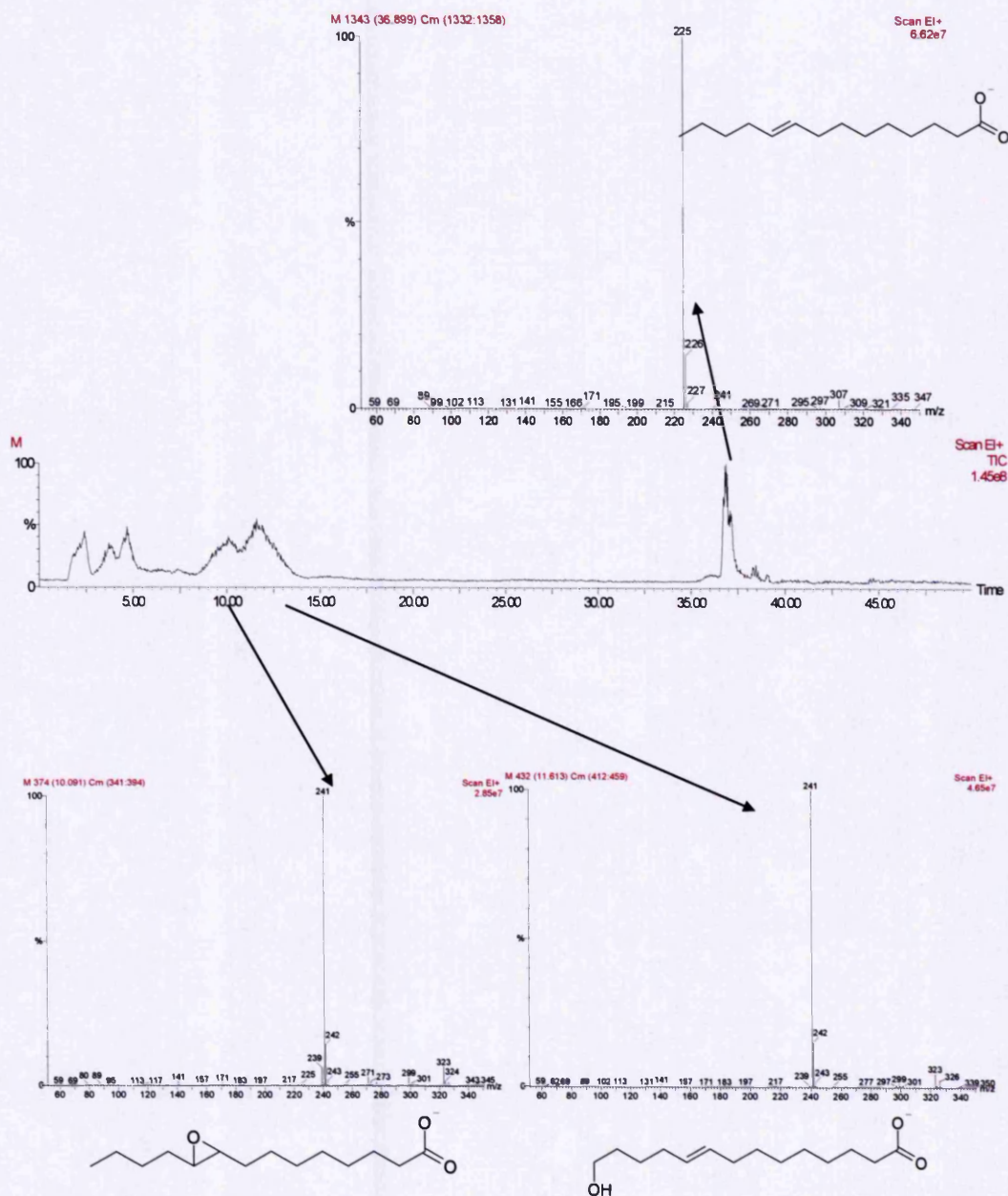


Figure 3.29: LC/MS spectra of the products of the turnover of myristoleic acid by cleaved CYP132, *E. coli* FLD and FLDR and NADPH. Suggested assignments for the main peaks on the mass spectra are shown. LC/MS was carried out at Queen Mary, University of London, on an Agilent 1100 series HPLC/MSD ion trap mass spectrometer, with an ACE 5A, 250 by 4.6 mm column and a gradient of methanol (A) and acetic acid (0.2 %) in water (B).

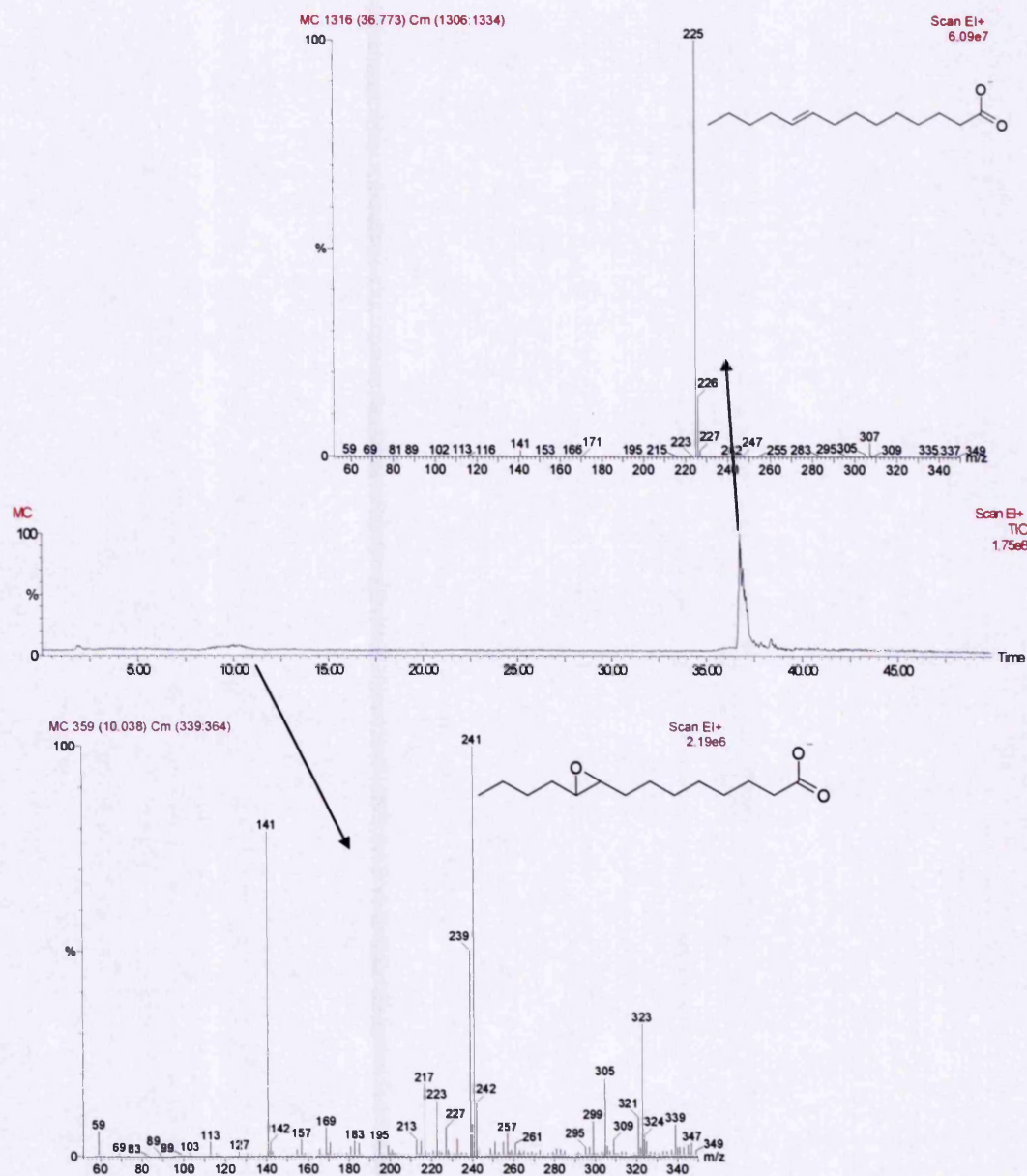


Figure 3.30: LC/MS spectra of the products of the control reaction, containing myristoleic acid, NADPH, glucose and glucose dehydrogenase only. Suggested assignments for the main peaks on the mass spectra are shown. LC/MS was carried out at Queen Mary, University of London, on an Agilent 1100 series HPLC/MSD ion trap mass spectrometer, with an ACE 5A, 250 by 4.6 mm column and a gradient of methanol (A) and acetic acid (0.2 %) in water (B).

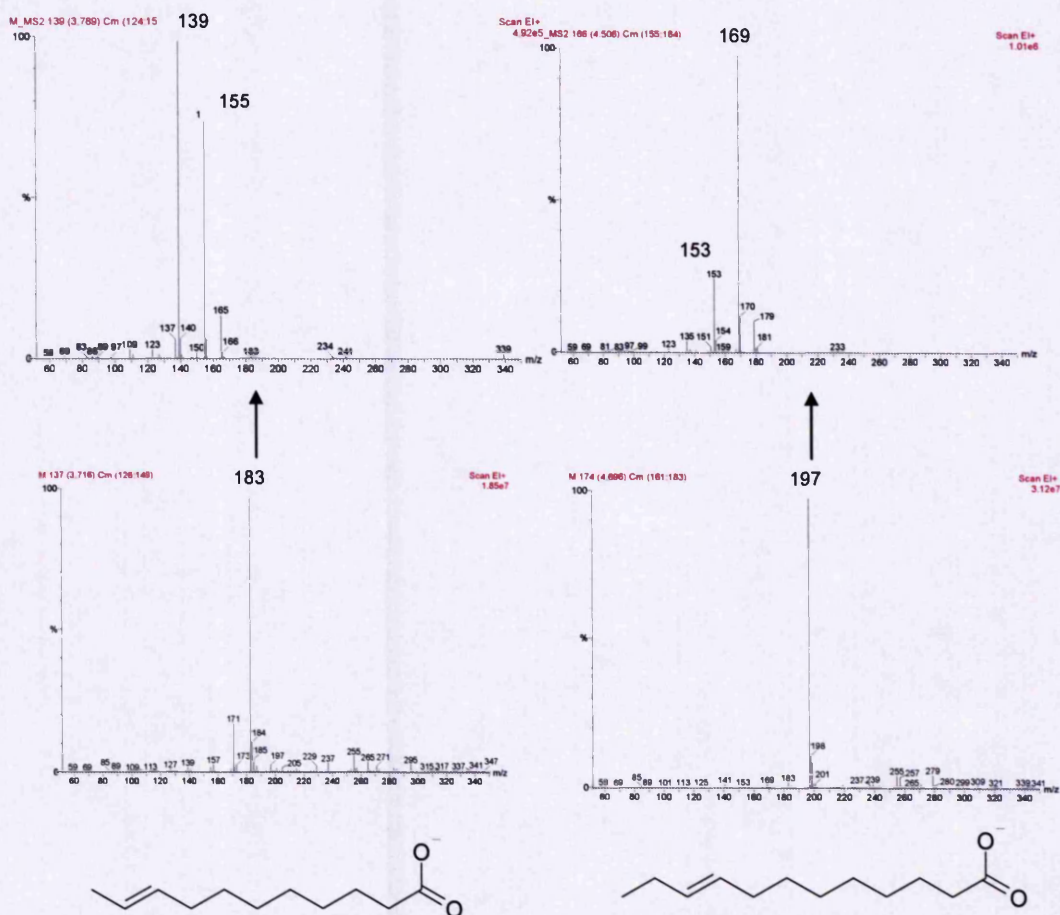


Figure 3.31: MS/MS analysis of the LC/MS peaks at 4 and 5 minutes from the products of the reaction of myristoleic acid with CYP132, FLD, FLDR and NADPH. Possible assignments for the species responsible for the mass spectra are shown below the graphs. Fragmentation of both the $m/z=183$ and $m/z=197$ leads to the loss of 28 and 44 mass units, suggesting the loss of C_2H_4 or CO , and CO_2 . This suggests the species both contain an acid functional group. The masses of 183 and 197 are 14 mass units apart, suggesting they differ by one CH_2 group. LC/MS was carried out at Queen Mary, University of London, on an Agilent 1100 series HPLC/MSD ion trap mass spectrometer, with an ACE 5A, 250 by 4.6 mm column and a gradient of methanol (A) and acetic acid (0.2 %) in water (B).

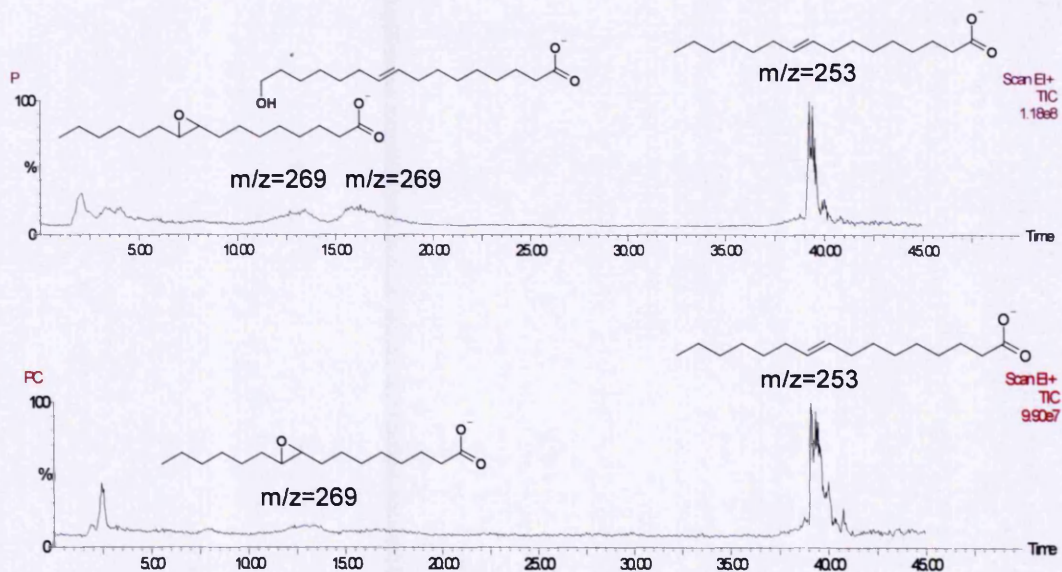


Figure 3.32: (Top) LC/MS chromatogram of the products of the reaction of palmitoleic acid with cleaved CYP132, FLD, FLDR and NADPH, with glucose and glucose dehydrogenase to replenish the NADPH (Bottom) LC/MS chromatogram of the control reaction, containing palmitoleic acid, NADPH, glucose and glucose dehydrogenase only. Masses of the species responsible for each peak are shown, along with suggested identities of the molecules responsible for them. LC/MS was carried out at Queen Mary, University of London, on an Agilent 1100 series HPLC/MSD ion trap mass spectrometer, with an ACE 5A, 250 by 4.6 mm column and a gradient of methanol (A) and acetic acid (0.2 %) in water (B).

3.2.5 Expression and purification of GST tagged CYP132

The use of a glutathione S-transferase (GST) tag for the purification of polypeptides by affinity chromatography on immobilised glutathione was developed by Smith and Johnson (Smith and Johnson, 1988). This utilised the affinity of a 26 kDa GST encoded by the parasitic helminth *Schistosoma japonicum* for glutathione to enable the purification of polypeptides from crude bacterial lysates in a single step.

The *Rv1394c* gene was cloned into the Novagen pET42a plasmid. It was thought that the presence of the large GST tag at the N-terminal of CYP132 would protect the P450 from proteases, which had caused cleavage of polyhistidine tagged CYP132.

3.2.5.1 Cloning of *Rv1394c* into the pET42a vector

A *NcoI* restriction site was engineered into the *Rv1394c*/pET11a plasmid using primers designed from the genomic sequence. These were as follows:

5'-GAAGGAGATATACCCATGCCACCGCCACCACCC-3'

5'-CTTCCTCTATATGGGTACCGGTGGCGGTGGTGGG-3'

The underlined region indicates the engineered *NcoI* restriction site. The *Rv1394c* gene was excised from the *Rv1394c/pET11a* plasmid using *NcoI* and *BamHI* and ligated into the pET42a expression vector, which had been pre-cut with the same enzymes. Clones of the expression plasmid *Rv1394c/pET42a* were verified by automated dideoxy DNA sequencing (MWG Biotech). Diagnostic restriction digests are shown in Figure 3.33.

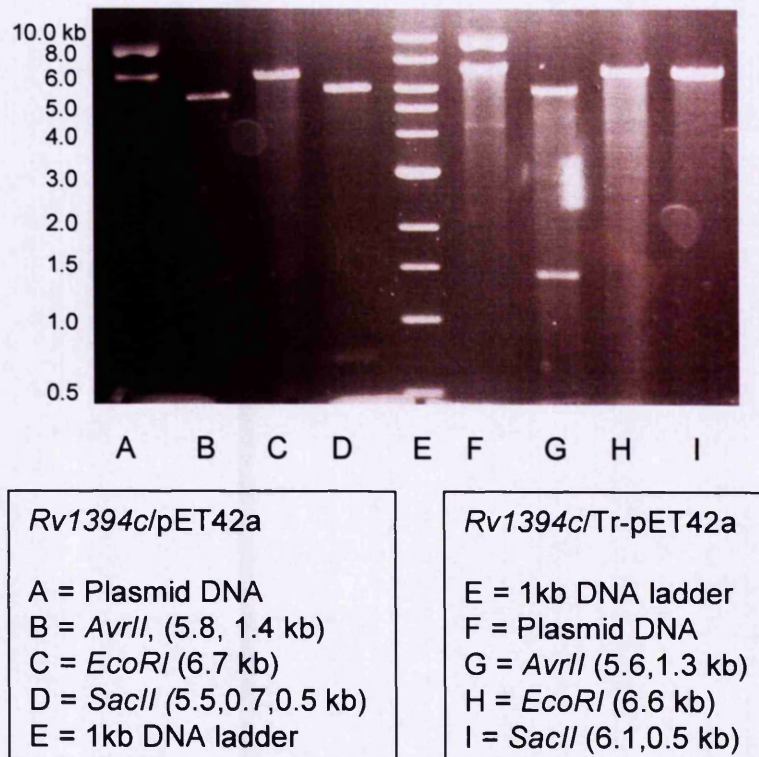


Figure 3.33: Agarose gel showing diagnostic restriction digests of *Rv1394c/pET42a* and *Rv1394c/Tr-pET42a* plasmids. The size of the bands in the NEB 1kb DNA ladder are shown to the left of the gel. Key shows restriction enzyme, and expected length of fragments. Of particular interest are the *SacII* digestion fragments (D and I), for which an extra band, at 0.7 kb, is expected for *Rv1394c/pET42a* that is not present for the truncated *Rv1394c/Tr-pET42a*.

pET42a encodes a region of 58 amino acids between the GST tag and the *NcoI* restriction site. In order to provide greater protection against proteolysis of CYP132, it was decided to truncate this region. A second *SpeI* restriction site was engineered five amino acids before the Factor Xa cleavage site using the following primer:

5'- CCAGATCTGGGTACTAGTGGTGGCTCCGGTATTGAGG-3'

5'- GGTCTAGACCCATGATCACCACCGAGGCCATAACTCC-3'

The underlined region shows the engineered *SpeI* restriction site. The mutated *Rv1394c/pET42a* plasmid was truncated by digestion with *SpeI* followed by ligation. The plasmid obtained was termed *Rv1394c/Tr-pET42a*. Clones were verified by automated dideoxy DNA sequencing (MWG Biotech). Diagnostic restriction digests are shown in Figure 3.33.

3.2.5.2 Expression of GST tagged CYP132

The *Rv1394c/pET42a* and *Rv1394c/Tr-pET42a* plasmids were transformed into the BL21, HMS174(DE3) and Rosetta(DE3) expression vectors and the expression of GST tagged CYP132 carried out following the protocol established for the expression of the *Rv1394c/pET15b* plasmid, above.

SDS-PAGE gel analysis of the expression products in the Rosetta(DE3) and BL21 cell strains showed no protein bands in the samples induced with IPTG that weren't in the samples that were left uninduced. SDS-PAGE gel analysis of the expression of *Rv1394c/pET42a* in HMS174(DE3) (Figure 3.34) showed a small band at ~62 kDa in the sample induced with IPTG that is not in the uninduced sample. This band was not detected in SDS-PAGE gel analysis of the expression of *Rv1394c/Tr-pET42a* in any of the cell strains used.

The pelleted cells from the expression of *Rv1394c/pET42a* and *Rv1394c/Tr-pET42a* in HMS174(DE3) were disrupted by sonication and by using a French press. UV/Vis analysis of the cell extracts showed high turbidity, with small haem shoulders around

420 nm. This is shown for progressive dilutions of the *Rv1394c*/Tr-pET42a expression cell extract in Figure 3.35.

The cell extracts were dialysed into 1 x PBS buffer (pH 7.4) and loaded onto a FPLC column containing immobilised glutathione. The bound protein was eluted in Tris (50 mM), containing reduced glutathione (20 mM, pH 8). SDS-PAGE gel analysis of the fractions obtained from *Rv1394c*/Tr-pET42a expression is shown in Figure 3.36. Similar SDS-PAGE data were obtained from the immobilised glutathione-purified expression products of *Rv1394c*/pET42a. UV/Vis analysis of the material in fraction 4 (lane E in Figure 3.36) is shown in Figure 3.37. This has a peak at 395 nm. It was considered possible (but unlikely) that this may correspond to CYP132 in the high-spin, substrate bound state. Addition of NO, sodium dithionite and CO, and imidazole (up to a concentration of 100 mM), had no effect on the UV/Vis spectra, except to increase the turbidity. It was concluded that the species bound to the immobilised glutathione was not CYP132.

Purification of the products of the expression of the *Rv1394c*/pET42a and *Rv1394c*/Tr-pET42a plasmids was attempted using DEAE Sepharose and Q Sepharose columns. No fractions corresponding to P450 were detected.

In order to attempt to aid the expression of CYP132 from the *Rv1394c*/pET42a and *Rv1394c*/Tr-pET42a plasmids, the haem precursor d-aminolaevulinic acid (150 μ M) was added to the expression media shortly before induction with IPTG (1 mM). CYP132 was still not detected in the expressed protein.

The attempted expression of GST-tagged CYP132 did not produce any material with the characteristics of a P450. Purification on immobilised glutathione produced two proteins, with masses ~27 kDa and ~62 kDa. UV/Vis analysis of a mixture of these proteins revealed them to contain a small proportion of high-spin haem. The low proportion of haem in the mixture suggests that this belonged to the ~62 kDa protein, as the minor component of the eluent. The addition of NO, sodium dithionite and CO, and imidazole (up to a concentration of 100 mM) had no effect on the UV/Vis spectra. This indicated that the high-spin haem species did not exhibit properties typical of a native (i.e. cysteinate-coordinated) P450.

The identities of the ~27 kDa and ~62 kDa proteins that bound to the immobilised glutathione column are unclear. The expression of *Rv1394c/pET11a*, discussed in Section 3.2.1 above, gave a protein of mass ~36 kDa, which did not contain haem. If this protein were bound to the truncated-GST tag, this would give a mass of ~63 kDa. If this were the identity of the ~62 kDa protein expressed by the *Rv1394c/pET42a* and *Rv1394c/Tr-pET42a* plasmids, however, the *Rv1394c/pET42a* plasmid would be expected to express a ~67 kDa protein, whereas both plasmids expressed proteins of the same mass. It is possible that the ~27 kDa protein expressed by the plasmids is the GST tag, which has been proteolytically cleaved from CYP132. Once again, if this were the case, however, the two plasmids would be expected to encode proteins 4.5 kDa different in mass (unless some further proteolytic event occurred), whereas both plasmids expressed proteins of equal mass.

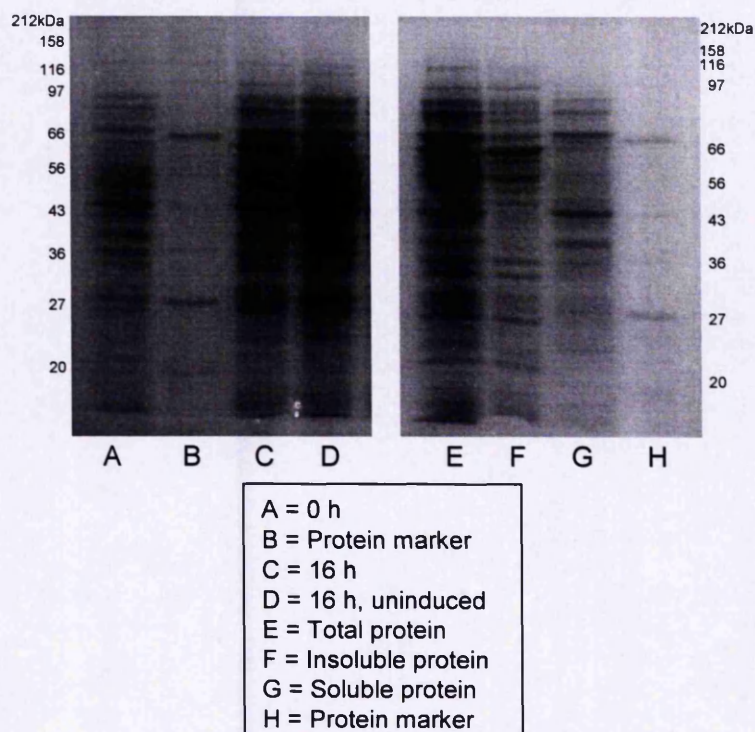


Figure 3.34: SDS-PAGE gel analysis of the expression of CYP132 from the *Rv1394c/pET42a* plasmid in HMS174(DE3). The only distinct band detected in the sample induced with IPTG (lane C) that was not present in the sample left uninduced (lane D) is that at ~62 kDa. This was found not to be a P450.

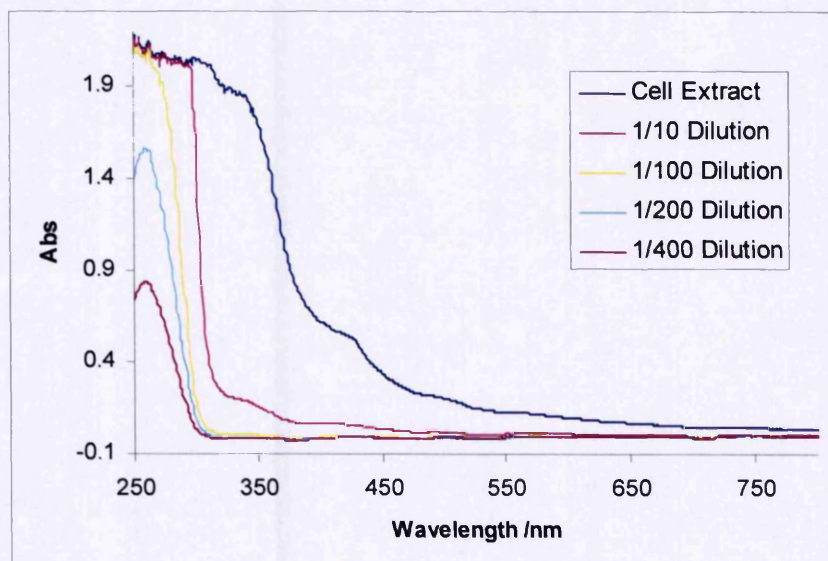


Figure 3.35: UV/Vis spectra of progressive dilutions of the cell extract from the expression of *Rv1394c/Tr-pET42a* in HMS174(DE3)

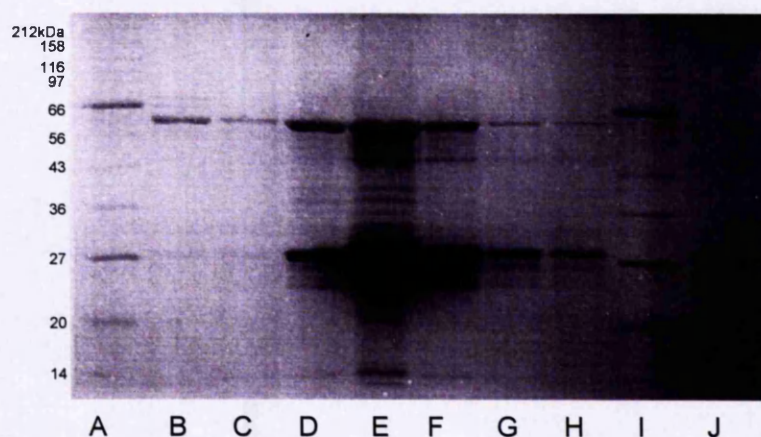


Figure 3.36: SDS-PAGE gel analysis of expression of proteins expressed from *Rv1394c/Tr-pET42a* transformants and purified on immobilised glutathione. Fractions were eluted from immobilised glutathione with Tris (50 mM), containing reduced glutathione (20 mM, pH 8). Proteins with similar masses were detected by SDS-PAGE analysis of proteins produced from *Rv1394c/pET42a* transformants and purified on immobilised glutathione, perhaps suggesting they were not expressed from the plasmids, as a difference in mass of 4.5 kDa would be expected between proteins produced from these constructs, unless a proteolytic or other event had occurred.

The expression of the ~62 kDa protein, from the *Rv1394c/pET42a* transformants, was observed to be induced by the addition of IPTG, indicating it was expressed from the plasmid, and did not result from the *E. coli* cell strain. The ~62 kDa protein was not

detected in the samples from SDS-PAGE analysis of gene expression in the HMS174(DE3)/*Rv1394c*/Tr-pET42a transformants, but was detected after partial purification from the cell extract, suggesting only low level expression. It is possible that the ~62 kDa protein purified from the transformants expressing genes from both these plasmids was not the same as the band at ~62 kDa detected in the expression samples from *Rv1394c*/pET42a expression, and that a lower mass protein, with mass ~58 kDa, was expressed from the plasmid *Rv1394c*/Tr-pET42a that was not detected by SDS-PAGE analysis. Analysis of the purification products with immobilised glutathione, DEAE-Sepharose and Q-Sepharose columns did not detect such a protein, however.

Whatever is the identity of the proteins detected in the SDS-PAGE gel and UV/Vis spectrometry analyses of the expressed proteins, it is clear that observable levels of GST-tagged CYP132 were not expressed by the *Rv1394c*/pET42a and *Rv1394c*/Tr-pET42a plasmid transformants on SDS-PAGE gels. Further investigation is required in order to obtain intact, full-length CYP132, so that the validity of the results already obtained with cleaved CYP132 can be established.

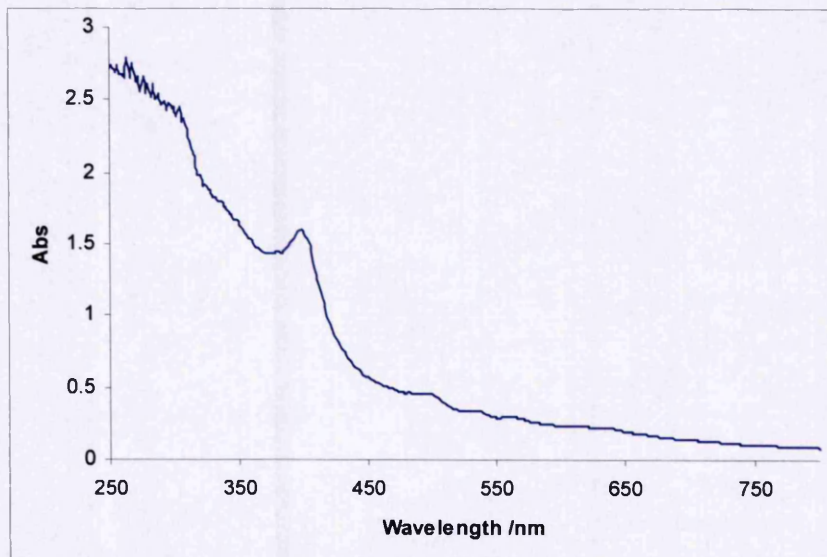


Figure 3.37: UV/Vis analysis of the expression product from HMS174(DE3)/*Rv1394c*/pET42a transformants after purification on immobilised glutathione. Haem Soret absorbance is at ~395 nm. Addition of NO, sodium dithionite and CO, and imidazole did not induce any significant change in the haem region of the spectrum, except for increased turbidity, indicating that the product bound to the immobilised glutathione is either not a P450, or is a P450 without normal cysteinyl ligation of its haem iron.

3.3 Conclusion

Around one third of the earth's population is estimated to be infected with *Mtb* (Ahmad *et al.*, 2006). There are around 8 million newly diagnosed cases and 2-3 million deaths from TB worldwide each year (Cardona and Ruiz-Manzano, 2004). With the increasing prevalence of multidrug-resistant TB there is a desperate need for novel strategies, and to shorten the duration of TB chemotherapy (Zhang, 2005).

The genome of *Mtb* H37Rv was solved in 1998 (Cole *et al.*, 1998). This revealed the presence of 20 predicted P450 enzymes in *Mtb*. P450s are known targets of azole anti-fungal drugs, and presented exciting new prospective drug targets in the fight against *Mtb* (McLean *et al.*, 2002b).

Mtb CYP132, encoded by the gene *Rv1394c* and regulated by the AraC transcriptional regulator encoded by *Rv1395* (Recchi *et al.*, 2003) was found in gene knockout studies to be important for the virulence of *Mtb* in mice (Camacho *et al.*, 1999). CYP132 shows significant amino acid similarity to eukaryotic CYP4A family fatty acid hydroxylase P450s, and was predicted to have a similar function (Munro *et al.*, 2003). The importance of CYP132 to the virulence of *Mtb*, and the likelihood of finding molecules that acted as substrates, made CYP132 an excellent candidate for investigation, in the search for new protein targets in the fight against *Mtb*.

Mtb Rv1394c, the gene that encodes CYP132, was cloned into the vectors pET11a, for wild type CYP132, pET15b, for polyhistidine tagged CYP132, and pET42a, which encodes GST-tagged protein. *Rv1394c*/pET11a was found to express only a small amount of CYP132, and no P450 expression was convincingly detected from *Rv1394c*/pET42a. *E. coli* transformants of *Rv1394c*/pET15b were found to produce predominantly CYP132 that had been proteolytically cleaved, with a mass of 43.1 kDa, but with small amounts of intact, full-length, 53.3 kDa CYP132 also produced. Intact CYP132 was confirmed to have P450 character, with a dominant shift to 449 nm on the addition of CO to the reduced enzyme.

Cleaved CYP132 was also confirmed to have P450-like character. Binding of CO to the reduced (ferrous) haem gave absorbance predominantly at 418 nm, with a small absorbance at 446 nm, indicative of formation of the P420 species and hence likely thiol coordination of the haem iron. Analysis of CYP132 by UV/Vis, EPR and CD spectroscopy also showed character consistent with other P450s.

CYP132 was found to bind NO, imidazole, and a range of azole anti-fungal drugs. The binding of clotrimazole and miconazole was particularly strong, with ketoconazole and econazole also showing strong binding. Binding to fluconazole was weaker than to some other *Mtb* P450s, but in line with higher MIC values of fluconazole for *Mtb*.

Cleaved CYP132 was found to be susceptible to disruption of the haem environment and unfolding of the protein. Analysis by CD showed that only 0.75 M GdmCl at 30 °C for 15 minutes was required for extensive unfolding of the protein. Additionally, it was found that in the presence of unsaturated fatty acids, or SDS, at concentrations ~120 mM or more, the haem environment was disrupted, leading to a loss of absorbance at 416 nm, and an increase at 365 nm, possibly reflecting haem dissociated from the protein matrix.

Type I binding shifts were not observed upon the addition of fatty acids to cleaved CYP132. However, when combined with a heterologous redox partner system, cleaved CYP132 was found to turn over (i.e. oxidise) unsaturated fatty acids, with a rate of up to 13 fatty acid turnovers per CYP132 per minute. This confirmed the interaction of CYP132 with unsaturated fatty acids, despite the apparent lack of type I binding. No interaction of cleaved CYP132 with saturated fatty acids was detected, however.

The products of the turnover of unsaturated fatty acids by CYP132, together with a heterologous redox partner system, were analysed by LC/MS. This revealed that cleaved CYP132 is able to hydroxylate unsaturated fatty acids, oxidising in at least one position, and it may also catalyse epoxidation. In addition, the LC/MS/MS results suggested that CYP132 may also catalyse C-C bond cleavage at the ω -2 and ω -3 positions. Further investigation is needed to confirm this result.

CYP132 has been shown to be an exciting new P450, which has not been characterised previously. Expression of CYP132 is challenging, especially the full-length enzyme, and further work is required to perfect the expression regime. Binding specificities of CYP132 are slightly different from those of some of the other *Mtb* P450s, with cleaved CYP132 showing a broader range of K_d values for azole anti-fungal drugs compared to the other *Mtb* P450s studied to date. Cleaved CYP132 turns over unsaturated fatty acids, but shows no interaction with saturated fatty acids of the same length.

CYP132 has been confirmed as a fatty acid hydroxylase, as predicted by its homology with eukaryotic CYP4As. CYP132 also appears to catalyse epoxidation as well as hydroxylation reaction. In addition, C-C bond cleavage, a reaction not previously known in family 4 P450s, may also be catalysed.

3.3.1 Future work

The first priority in the study of CYP132 is to express and purify useable quantities of intact, full-length CYP132. Cleaved CYP132, with a mass of 43.1 kDa, has been expressed, purified and characterised. Full-length, 53.3 kDa CYP132 has been expressed and purified in small amounts.

The use of a GST-tag, with the pET42a expression vector, does not appear to be a viable option for the expression of full-length (i.e. intact) CYP132. A small amount of full-length enzyme was expressed with the pET15b expression vector, and this could be a sensible starting point in the expression of larger quantities, although a range of conditions have already been investigated with this expression system.

Once an appropriate quantity of full-length CYP132 has been obtained, a range of analyses can be carried out, and the results compared with those for the cleaved CYP132. Of particular importance is the binding of substrates and inhibitors to full-length CYP132. Cleaved CYP132 showed lower affinity for azole anti-fungal drugs than some other *Mtb* P450s, and it is important to know whether full-length CYP132 has binding affinities more in line with these other P450s.

Preliminary results suggest some very interesting behaviour of CYP132 with fatty acids. Further investigation is necessary to confirm the exact nature of the reactions catalysed by CYP132. If a sufficient quantity of hydroxylated fatty acid could be produced, nuclear magnetic resonance (NMR) spectrometry could be used to confirm the position of the hydroxylations. NMR could also confirm the identity of the species that were tentatively assigned as C-C bond cleavage products, giving a much clearer assignment where LC/MS could not.

Much greater insight into the CYP132 secondary and tertiary protein structure could be obtained by x-ray crystallography studies. This would also confirm, or otherwise, the presence of a substrate co-purified with cleaved CYP132, which could not be extracted by the methods used in this study. If a substrate-bound structure could be obtained, this would give insight into any change in the CYP132 structure upon substrate binding, and give greater insight into the binding specificity, which would aid the design of new drugs, and possibly new treatments that could target this virulence-associated P450 and might be active against *Mtb*.

CHAPTER 4: INVESTIGATION OF THE P450 REACTION CYCLE BY DFT METHODS

4.1 Introduction

Since their discovery in 1958 (Garfinkel, 1958, Klingenberg, 1958) there has been increasing interest in the reactions of cytochromes P450 (Groves, 2005). P450s are the major catalysts involved in the oxidation of xenophobic chemicals (Guengerich, 2005) and catalyse a wide range of reactions, from oxidation, hydroxylation and epoxidation to reduction, ester cleavage, aldehyde scission and dehydration (Guengerich, 2001).

The soluble bacterial P450s have been particularly useful in the development of knowledge about the P450 reaction cycle. One of the most widely studied reactions is the hydroxylation of camphor by P450cam (Davydov *et al.*, 2001). Many crystal structures of P450cam are available (Hishiki *et al.*, 2000, Poulos *et al.*, 1987, Schlichting *et al.*, 2000) and this has led to it being the basis of many theoretical studies on P450s (Guallar and Friesner, 2004, Harris *et al.*, 2001, Schöneboom and Thiel, 2004). The *Bacillus megaterium* flavocytochrome P450BM3 has also been characterised extensively (Munro *et al.*, 2002). This is due to it being similar to eukaryotic members of the CYP4A family of fatty acid hydroxylases, but also soluble, as opposed to the eukaryotic P450s, which are membrane-bound. P450BM3 has been shown to hydroxylate a range of fatty acids close to the ω terminus, and shows the highest turnover rate of any of the P450 mono-oxygenases discovered to date.

The most abundant P450 in humans is CYP3A4 (Guengerich, 2005). This is involved in xenobiotic metabolism and contributes to the metabolism of around 50% of drugs on the market. CYP3A4 and the various other human P450s involved in xenobiotic metabolism have a broad range of substrates and catalyse a variety of reactions. As well as the standard hydroxylation reactions, such as testosterone 6 β -hydroxylation (Guengerich *et al.*, 1986), CYP3A4 has been found to catalyse some atypical reactions (Guengerich, 2001), such as the desaturation of lovastatin (Wang *et al.*, 1991), oxidation of a nitrile to an amide in the metabolism of pinacidil (Zhang, Z. P. *et al.*, 2002) and oxidative

carboxylic acid ester cleavage of the Hantzsch pyridine ester of pranidipine (Kudo *et al.*, 1999).

The first crystal structure of a P450 was that of P450cam, solved by Poulos *et al.* (Poulos *et al.*, 1986). This revealed the P450 active site to consist of a ferriprotoporphyrin IX (haem *b*) prosthetic group, with the iron bound to the sulphur atom of a cysteine residue. MCD and EPR studies by McKnight *et al.* revealed the sixth ligand to be a strongly hydrogen bonded water molecule. ENDOR and ESEEM spectroscopy studies by Goldfarb *et al.* (Goldfarb *et al.*, 1996) confirmed this.

P450s have been shown to effectively activate molecular oxygen (O₂) and insert one of its atoms into a variety of substrates (Guengerich, 2001). P450s require two electrons, which they usually receive from NADPH, giving the general reaction shown in equation 4.1.



In 1955, Hayaishi *et al.* (Groves, 2005, Hayaishi *et al.*, 1955) demonstrated the activation and transfer of molecular oxygen into its substrate by an iron-containing enzyme. Using oxygen isotopic measurements they showed that both the inserted oxygen atoms in the conversion of catechol to *cis*-muconic acid came from molecular oxygen, and not from water. Prior to this it had been believed that oxygen in biological molecules was derived exclusively from water via hydration processes (Makris *et al.*, 2005). This led to further investigation into oxygen activation, and by 1965 at least eight variations of structurally different metal-oxygen complexes had been discovered in nature.

4.1.1 The P450 catalytic cycle

The generally accepted mechanism for the activation of molecular oxygen by P450s is shown schematically in Figure 4.2 (Shaik and de Visser, 2005). Except for at the top of the figure, the porphyrin macrocycle is represented by the ellipse around the iron. This was developed based on experimental and theoretical methods, and by analogy with

similar enzymes, such as peroxidases. There has been much debate over the nature of the catalytic cycle over the past decades, which will be summarised briefly here.

The resting state of the enzyme is the water-bound ferric (iron III) complex. Upon binding of a substrate, the bound water molecule is displaced. The removal of the water ligand lowers the interaction with the five d orbitals of the iron. This often leads to a change in spin to the high spin state, and makes the iron a good electron acceptor (Shaik and de Visser, 2005). In particular, the loss of the water ligand removes the anti-bonding interaction between the water p_z and iron d_{z^2} orbitals. Also the removal of the water ligand allows the iron to move out of the plane of the porphyrin more easily, lowering the anti-bonding interaction between the nitrogen p_x/p_y orbitals and the iron d_{xy} orbital. These are shown in Figure 4.1.

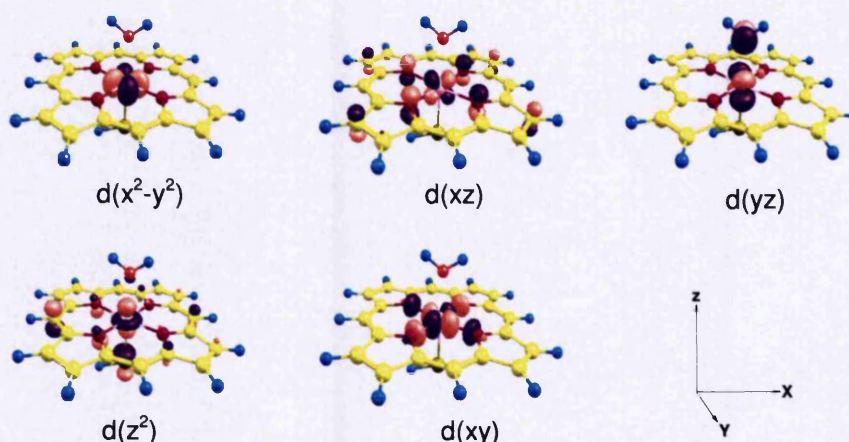


Figure 4.1: Iron anti-bonding d orbitals, shown here for the doublet spin state of the water bound resting state.

An electron is provided by NADPH and transported by an associated reductase, leading to a reduction to high spin ferrous P450 (Groves, 2005). Ferrous porphyrin is a good dioxygen binder. The binding of molecular oxygen to the ferrous P450 increases the interaction with the iron d_{z^2} and d_{xy} orbitals, more than the interaction with the iron d_{xz} , d_{yz} and $d_{x^2-y^2}$ orbitals. This increases the difference in energy between the two sets of orbitals, inducing a shift to low spin oxygen-bound ferrous P450 (Shaik and de Visser, 2005). A second electron is now added. The resulting complex is generally considered to be the ferric-peroxo species, with the two negative charges residing on the bound

oxygen atoms. Protonation leads to the Fe-O-O-H species, also known as Compound 0. A second protonation leads to Fe^{IV}=O, also called Compound I, and water. Compound I inserts an oxygen atom into the substrate, leading to the water-bound resting state once more, and the oxygenated product.

The first three steps of the cycle, the substrate binding, first electron reduction and oxygen binding, have been observed spectroscopically (Groves and Han, 1995, White and Coon, 1980), the second electron transfer, the O-O bond cleavage and the oxidation of the substrate occur too rapidly for observation by standard spectroscopic techniques. Davydov *et al.* (Davydov *et al.*, 2001) using gamma radiation as a source of electrons, at cryogenic temperatures were able to isolate 4, 5, and 6, and identify them by electron paramagnetic resonance (EPR) and electron nuclear double resonance (ENDOR) spectroscopy. They were not able to detect Compound I (7) however.

Kellner *et al.* (Kellner *et al.*, 2002) used stopped-flow rapid mixing techniques to study the reaction of CYP119 with *meta*-chloroperoxybenzoic acid (*m*-CPBA). They observed absorptions at 370, 610 and 690 nm, indicative of the presence of Compound I, for around 10 ms.

The identity of the remaining species in the reaction cycle was initially inferred from other results. It was found that in addition to the activation of molecular oxygen, hydrogen peroxide, alkyl hydroperoxides, peroxy acids, periodate and iodosylbenzene were also functional with P450s. This suggested that the activation of oxygen in P450s was the two electron reduction of molecular oxygen to hydrogen peroxide (Groves, 2005).

The nature of Compound I was not known for many years (Groves and Han, 1995). It was observed that the addition of hydrogen peroxide to horseradish peroxidase (HRP) gave a green compound with a unique optical spectrum, this was named Compound I (George, 1952). Low temperature optical spectra and EPR spectroscopy indicated an S = 3/2 spin state (Peisach *et al.*, 1968). Isotopic labelling (Schonbaum and Lo, 1972), and ENDOR (Roberts *et al.*, 1981) studies showed HRP Compound I to contain one atom of oxygen. X-ray absorption spectroscopic studies revealed the presence of a short, terminal iron-oxygen bond of 1.6 Å (Pennerhahn *et al.*, 1983). The crystal structure of

HRP Compound I was solved by Berglund *et al.* (Berglund *et al.*, 2002) This indicated a Fe-O distance of 1.7 Å.

The crystal structure of yeast cytochrome *c* peroxidase Compound I was solved by Edwards *et al.* in 1987 (Edwards *et al.*, 1987) and a higher resolution structure of *Saccharomyces cerevisiae* Compound I was solved by Bonagura *et al.* in 2003 (Bonagura *et al.*, 2003). The structure of Bonagura *et al.* showed a slightly longer Fe-O distance than expected of 1.87 Å. They concluded that cytochrome *c* peroxidase Compound I comprises a Fe-OH single bond, with the bound hydroxyl stabilised by hydrogen bonding with surrounding residues.

The nature of P450 Compound I, and the identity of the active oxidant species, is still a matter of contention (Altun *et al.*, 2007). Egawa *et al.* (Egawa *et al.*, 1994) studied the reaction of P450cam with *meta*-chloroperoxybenzoic acid (*m*-CPBA) using rapid scan absorption spectrometry. They observed a species formed within 10 ms of the start of the reaction that had a similar absorption spectrum to that of chloroperoxidase Compound I. Stopped-flow studies by Spolidakis *et al.* (Spolidakis *et al.*, 2005) showed a pH dependence, with a tyrosyl radical-bound Compound I species being formed at low pH. The reaction of *m*-CPBA with CYP119 was studied by Kellner *et al.* (Kellner *et al.*, 2002), giving reaction kinetics of $3.20 \times 10^5 \text{ M}^{-1} \text{ s}^{-1}$ at pH 7.0, 4 °C.

Schlichting *et al.* (Schlichting *et al.*, 2000), using cryogenic, time-lapse x-ray crystallographic techniques were able to solve structures of substrate-bound ferrous and ferric P450cam, dioxygen-bound ferric P450cam, and a structure containing iron bound to a single oxygen atom, which they believed to be Compound I. This conclusion was challenged by Davydov *et al.* (Davydov *et al.*, 2001) who found, using EPR and ENDOR spectroscopy, that the rate constant for the oxidation of the unactivated C-H bond in camphor exceeds 1000 s^{-1} , and the reaction is thus too fast for the accumulation of Compound I.

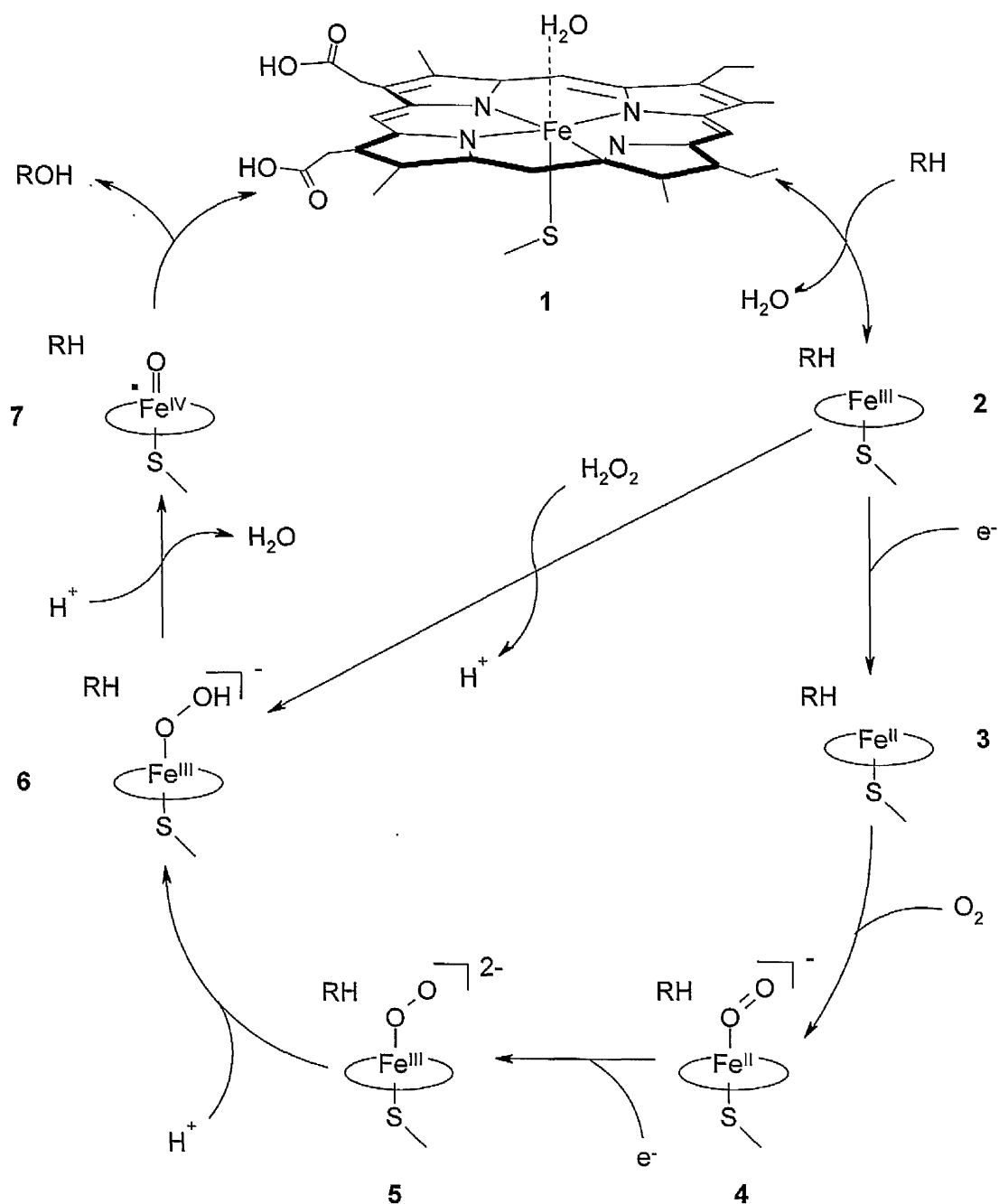


Figure 4.2: The consensus catalytic cycle for the activation of molecular oxygen by P450s and the hydroxylation of a substrate, adapted from (Shaik and de Visser, 2005). Except for the resting state, the haem macrocycle is represented by an ellipse. The substrate is represented by RH.

The consensus mechanism of the oxidation step for many years was the 'rebound mechanism' proposed by Groves *et al.* (Groves and McClusky, 1976, Groves, 1985), this is illustrated in Figure 4.3. Step (a) involves hydrogen abstraction from the alkane (RH). This is followed by hydroxyl rebound onto the radical to generate the ferric-alcohol complex in step (b) (Harris *et al.*, 2000). Evidence for the rebound mechanism includes rearranged alcohol products, which indicate the presence of a free radical with a finite lifetime (Meunier, 1992, Woggon, 1997). Also large kinetic isotope effects (KIE) (Manchester *et al.*, 1997) suggest that C-H bond breaking is the rate-determining step.

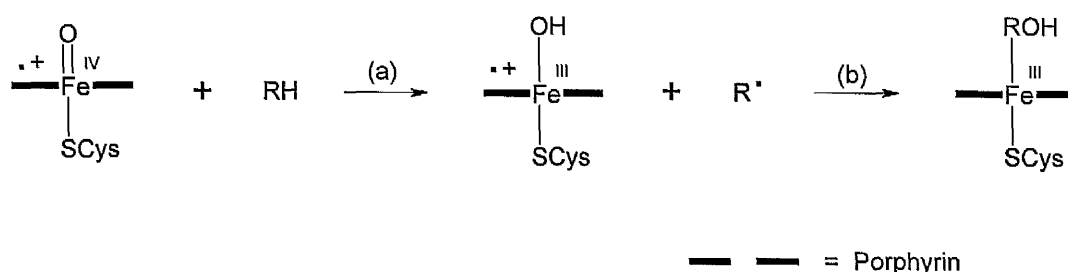


Figure 4.3: Schematic representation of the rebound mechanism proposed by Groves *et al.*

The rebound mechanism is contradicted, however, by evidence from experiments by Newcomb *et al.* using 'radical clocks' to probe the rate of the rebound step (Newcomb *et al.*, 1995a, Newcomb *et al.*, 1995b, Newcomb *et al.*, 2000, Newcomb and Toy, 2000). In a series of studies using radical probes which rearrange at a known rate, they found that the lifetimes of the intermediates in the reaction are in the order of 80-200 fs and are thus too short to correspond to free radical intermediates. They hypothesised the presence of two oxidants, one involving a free radical mechanism and the other a cationic intermediate.

The conflicting findings of Groves and Newcomb were summarised by Shaik *et al.* as "radicals are both present and absent and the rebound mechanism is both right and wrong" (Shaik *et al.*, 2002). Shaik *et al.* used DFT to model the mechanism of hydroxylation by Compound I (Harris *et al.*, 2000, Shaik *et al.*, 2002). They found that the doublet and quartet spin states of Compound I are very close in energy. The doublet spin state was found to lead to a concerted hydroxylation step, which did not give time for the production of radicals. The quartet spin state, however, was found to proceed via a stepwise mechanism, with the incursion of radical intermediates and the possibility of

substrate rearrangement and the loss of stereospecificity. Shaik *et al.* attribute the conflicting findings of the Groves and Newcomb groups to this 'two-state reactivity'.

More recently Newcomb *et al.* (Newcomb *et al.*, 2006) used laser flash photolysis to produce a species assigned as a CYP119 Compound I derivative. They found that the rate of decay of this transient was unchanged in the presence of laurate or styrene, known substrates of CYP119, and suggested that Compound I was not the oxidant species. They suggested that the high energy iron(V)-oxo species could be the true oxidant. This suggestion is refuted by Altun *et al.* (Altun *et al.*, 2007), who cite a previous DFT study (Ogliaro *et al.*, 2001) which found that iron(V)-oxo species are too high in energy to be the primary oxidant. In the same article Altun *et al.* found that the sextet spin state of Compound I, which they found to have an energy 12 kcal mol⁻¹ above the ground state, had a comparable barrier to hydroxylation to the lower spin states. They suggested that the sextet spin state could also be a viable oxidation species. And so the debate goes on.

4.1.2 Aim

It has been observed that the structure of the haem region in P450s changes depending on the axial ligand bound to the haem and on the spin state of the iron (Schlichting *et al.*, 2000). The protein environment can also vary greatly between different P450s. The P450 haem is conserved throughout the various P450 structures. DFT studies allow the changes in the haem structure and electronic configuration to be examined without the influence of the protein environment, in order to give information that can be applied to all P450s.

At the time this study was carried out, a complete DFT study of each haem spin state of all the species in the P450 reaction cycle had not been carried out. A systematic study had been carried out by Goller *et al.* (Göller and Clark, 2001) using semi-empirical calculations. Some of the results from this study did not correlate well with other studies (Shaik and de Visser, 2005). For example the Fe-O distances in the Compound I structures were shorter than those from *ab initio* calculations, giving a different electronic structure. For this reason it was decided to carry out a systematic survey of the P450 catalytic cycle using DFT. This would give invaluable information about the catalytic

cycle which could be used in further studies, such as in the docking of drugs using MM methods.

The work described in this chapter was carried out before a number of the references cited above were written. At the time there had not been a DFT study of the low, intermediate and high spin states of each species in the P450 catalytic cycle. Since the work in this chapter was started a number of papers have superseded it, including the systematic survey of the catalytic cycle of haemoglobin and myoglobin, peroxidases, P450s and haem catalases using DFT by Rydberg *et al.* (Rydberg *et al.*, 2004). The various DFT studies on the different species in the catalytic cycle in P450s are summarised in the review articles and book chapters by Shaik *et al.* (Shaik and de Visser, 2005, Shaik *et al.*, 2005a).

In this chapter, the calculation of the low, intermediate and high spin states of each species in the consensus P450 catalytic cycle using DFT, with the B3LYP functional and BS1 will be discussed. The haem region was modelled by a porphyrin group with a thiol axial ligand, as used by Shaik *et al.* (Ogliaro *et al.*, 2000a), who report that a SH group gives the best representation of the cysteine ligand in gas phase DFT calculations. Each species in the catalytic cycle will be discussed in turn, and, where available, the results will be compared to those reported by others in the literature.

4.2 Results and Discussion

4.2.1 Water-bound resting state

This study of the reaction cycle of P450s started by looking at the water-bound resting state without substrate present, 1. This is the most widely studied species in the reaction cycle. As has been seen with CYP132 in the previous chapter, the experimentally observed spin state of the P450 resting state is predominantly low spin. This was first observed by Tsai *et al.* in 1970 (Tsai *et al.*, 1970), who performed EPR spectroscopy on P450cam. They found the resting state to be around 93% low (doublet) spin at temperatures below 20 K, the rest being high spin. They found that upon titration with camphor the spin state became around 60% high spin. Later, in 1976, Sharrock *et al.* (Sharrock *et al.*, 1976) carried out Mössbauer spectroscopy on P450cam, again finding

the ground state to be doublet spin, and the substrate-bound ground state to be the high spin, sextet state.

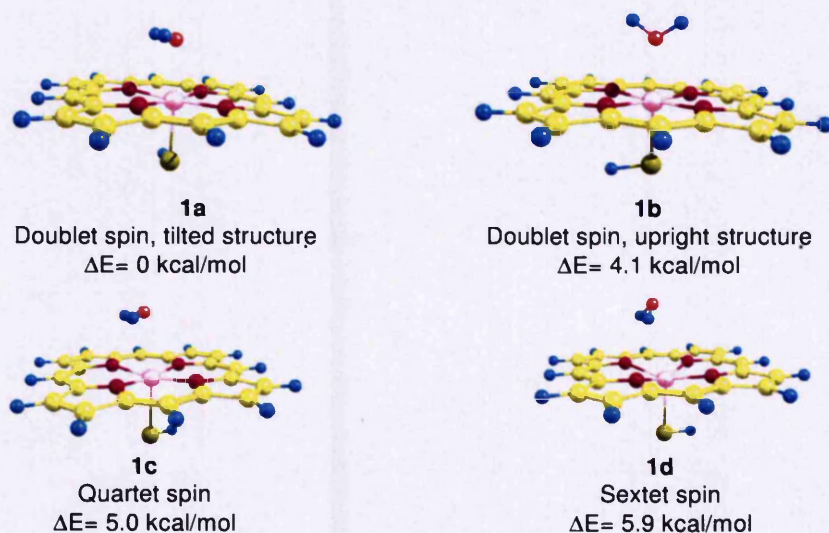


Figure 4.4: Structures and relative energies of doublet, quartet and sextet spin states of water-bound ferric porphyrin. Both the tilted and upright structures of the doublet spin state are shown. The upright structure was calculated from a single point calculation on an unconverged structure, see Figure 4.5

The three spin states, doublet, quartet and sextet, of the water-bound, Fe^{III} resting state (1) were optimised using Gaussian 03, with the B3LYP functional and basis set BS1, as described in Chapter 2. In the course of investigations that will be described in Chapter 5, the three spin states of the resting state were also optimised using BP86 and basis set BS4 and will be discussed briefly here. Shaik *et al.* (Filatov *et al.*, 1999), reported two structures for the doublet spin state of the water-bound resting state. In the lower energy of these, termed 'tilted' in a later review by the same group (Shaik *et al.*, 2005b), the bound water is almost parallel to the porphyrin ring. A higher energy species, termed 'upright', was also found. This has the plane of the bound water molecule at right angles with the porphyrin ring.

It was found when an upright structure was used as a starting point for the optimisation with Gaussian 03, that the water fell down to the tilted structure during the optimisation. Attempts to rectify this by first optimising the structure with the Fe-O-H angles fixed, in order to obtain a starting structure close to the local minimum, did not meet with

success. This was repeated with basis sets BS2 [3-21G], B3[6-31G**] and BS4[6-31G] with both the B3LYP and BP86 functionals. The water was still found to fall down to the tilted structure.

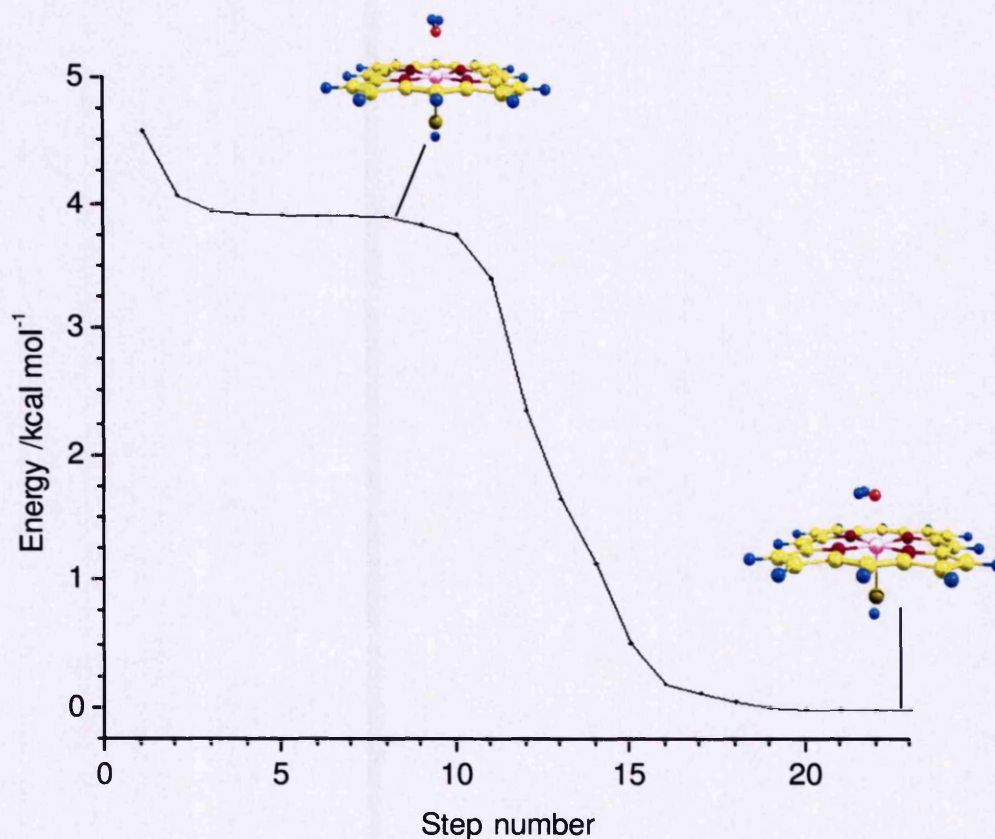


Figure 4.5: Optimisation convergence of upright water structure from starting geometry provided by Dr. Sam de Visser, University of Manchester. The upright starting geometry converged to the tilted structure. A single point calculation from step 8 (structure shown on graph) was used to represent the upright structure in the following results.

The geometry optimised by Shaik *et al.* (de Visser, Personal communication) was kindly provided by Dr. Sam de Visser, University of Manchester. Using this structure as the starting point for the optimisation with B3LYP and BP86, with basis sets BS1 and BS4, the water was still found to fall down to the tilted conformation. With B3LYP and the BS1 basis set the optimisation convergence graph shown in Figure 4.5 was obtained. A single point calculation was carried out on step 7, the lowest energy upright structure

before the water begins to fall down to the tilted structure. This structure is used in the following results.

Table 4.1: Summary of the significant bond distances, bond angles, electron spin densities and atomic charges in the doublet, quartet and sextet spin species of the water-bound resting state. Distances are in Å and angles in degrees. The atom numbering is shown in Figure 4.6

	Doublet (tilted)	Doublet (upright)	Quartet	Sextet
Fe-O distance	2.17	2.15	2.66	3.19
Fe-S distance	2.29	2.29	2.47	2.40
Fe-N₅ distance	2.03	2.02	2.04	2.10
Fe-N₆ distance	2.03	2.02	2.02	2.10
Fe-N₇ distance	2.01	2.02	2.01	2.08
Fe-N₈ distance	2.01	2.02	2.02	2.09
N₅-H₃ distance	2.42	2.96	2.42	2.37
N₆-H₄ distance	2.41	2.96	2.55	2.38
Fe deviation from plane	0.099	0.081	0.19	0.43
Fe-O-H₃ angle	98.9°	125.3°	83.2°	63.9°
Fe-O-H₄ angle	98.7°	124.7°	86.3°	63.2°
Fe spin density	1.09	1.13	2.51	3.91
S spin density	-0.0032	-0.041	0.49	0.49
N₅ spin density	-0.023	-0.027	-0.010	0.13
N₆ spin density	-0.021	-0.024	-0.019	0.12
N₇ spin density	-0.024	-0.026	-0.020	0.13
N₈ spin density	-0.22	0.025	-0.013	0.14
O spin density	-0.00063	0.0043	0.019	0.0037
Fe atomic charge	0.28	0.36	0.48	0.65
S atomic charge	-0.092	-0.010	-0.25	-0.21
N₅ atomic charge	-0.82	-0.80	-0.85	-0.90
N₆ atomic charge	-0.82	-0.81	-0.86	-0.91
N₇ atomic charge	-0.80	-0.81	-0.84	-0.89
N₈ atomic charge	-0.81	-0.80	-0.83	-0.88
O atomic charge	-0.71	-0.73	-0.74	-0.80
H₃ atomic charge	0.45	0.44	0.44	0.43
H₄ atomic charge	0.45	0.44	0.43	0.43

Schöneboom and Thiel (Schöneboom and Thiel, 2004) carried out QM/MM studies on the resting state of P450cam. Using the camphor-bound crystal structure as a starting point, but replacing the bound camphor with six water molecules, they found a water orientation intermediate between the upright and tilted structures of Shaik *et al.*, with Fe-O-H angles of 108.9° and 110°. This is compared to angles of 100.4° and 102.3° in their gas phase calculations, and 98.9° and 98.7° in this study. ESEEM studies on P450cam by Goldfarb *et al.* (Goldfarb *et al.*, 1996) gave Fe-H distances of 2.6 Å, which is in good agreement with the QM/MM studies of Schöneboom and Thiel.

The resulting structures from the present study, and their relative energies, are shown in Figure 4.4. Table 4.1 summarises the important bond distances and angles, as well as the significant spin and charge densities. A low spin, doublet ground state was found, with the quartet and sextet species being 5 and 5.9 kcal mol⁻¹ higher in energy, respectively. Fe-O bond distances of 2.17, 2.66, and 3.19 Å were found, with the hydrogen atoms of the water molecule being progressively more tilted towards the porphyrin with increasing spin state. The energies and structures obtained by various groups using various different methods are summarised in Table 4.2.

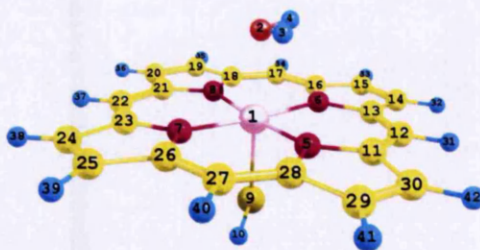


Figure 4.6: The water-bound resting state **1a**, showing the numbering system used in Table 4.1.

Table 4.2: Summary of the energy and significant geometries of the doublet (tilted and upright), quartet and sextet spin species of the water-bound resting state as found by various different groups using various methods. (a) – (d) are from this study (a) B3LYP, BS1, (b) B3LYP BS4, (c) BP86, BS1, (d) BP86 BS4, (e) Shaik *et al.* B3LYP (Shaik *et al.*, 2005a), (f) Filatov *et al.* (Filatov *et al.*, 1999), (g) Schöneboom and Thiel B3LYP/CHARMM (Schöneboom and Thiel, 2004), (h) Rydberg *et al.* B3LYP (Rydberg *et al.*, 2004), (i) Göller *et al.* semi-empirical (Göller and Clark, 2001), (j) Guallar and Friesner QM/MM (Guallar and Friesner, 2004)

Spin	Group	$\Delta E /$ kcal mol ⁻¹	r(Fe-O) /Å	r(Fe-S) /Å	r(Fe-N) (av) /Å	$\theta(\text{F-O-H})$ (av) /Å
Doublet (tilted)	a	0	2.17	2.29	2.02	98.8
	b	0	2.10	2.29	2.02	108.5
	c	0	2.15	2.25	2.02	95.6
	d	0	2.11	2.26	2.01	103.4
	e	0	2.11	2.29	2.02	108.4
	f	0	2.25	2.178	-	-
	g	0	2.14	2.27	2.04	109.5
	h	0	2.28	2.20	2.02	-
	i	0	2.09	2.25	2.02	-
	j	0.0	2.10			
Doublet (Upright)	a	4.1	2.14	2.29	2.02	125.0
	b	1.0	2.06	2.30	2.02	122.6
	c	5.4	2.12	2.26	2.02	125.1
	d	2.0	2.04	2.26	2.01	122.7
	e	1.1	2.07	2.29	2.02	122.6
	f	6.6	2.16	2.18	-	-
	a	5.0	2.66	2.47	2.02	84.8
	b	5.9	2.67	2.48	2.02	85.3
	c	12.1	2.76	2.45	2.02	75.1
	d	13.2	2.64	2.46	2.02	82.1
Quartet	e	-	-	-	-	-
	f	13.5	8.204	2.34	-	-
	g	2.4	2.48	2.49	2.04	108.1
	h	13.4	-	-	-	-
	i	2.6	2.09	2.22	1.93	-
	j	9.5				
	a	5.9	3.19	2.40	2.09	63.6
	b	7.7	3.17	2.41	2.09	63.8
	c	22.6	2.72	2.41	2.10	76.1
	d	24.0	3.00	2.40	2.09	64.5
Sextet	f	'higher E'	-	-	-	-
	g	3.3	2.47	2.42	2.10	106.9
	h	0.36	3.73	2.30	2.10	-
	i	9.66	2.09	2.22	1.93	-
	j	20.8	2.46			

Table 4.3: Iron d orbital energies for the doublet, quartet and sextet spin species of the water-bound resting species, 1, in atomic units. Orbitals in red are unoccupied.

		$d(x^2-y^2)$	$d(xz)$	$d(yz)$	$d(z^2)$	$d(xy)$
Doublet, (upright)	α	-0.28	-0.29	-0.32	-0.077	-0.036
	β	-0.25	-0.22	-0.088	-0.056	-0.0069
Doublet, (tilted)	α	-0.28	-0.29	-0.33	-0.077	-0.041
	β	-0.26	-0.23	-0.063	-0.045	-0.013
Quartet	α	-0.33	-0.37	-0.37	-0.30	-0.093
	β	-0.27	-0.069	-0.064	-0.076	-0.020
Sextet	α	-0.44	-0.45	-0.44	-0.43	-0.25
	β	-0.12	-0.082	-0.12	-0.10	-0.061

As can be seen from Table 4.3, in the quartet spin state the anti-bonding $d(z^2)$ orbital is populated, and in the sextet spin state both the $d(z^2)$ and $d(xy)$ orbitals are populated. These orbitals are illustrated in Figure 4.7. The $d(z^2)$ and $d(xy)$ orbitals mix strongly with the $p(z)$ orbitals of the sulphur and oxygen, and with the $p(x)$ and $p(y)$ orbitals of the porphyrin nitrogen atoms respectively, forming anti-bonding σ^* orbitals. These destabilise the quartet and sextet spin states relative to the doublet spin state, in which only the $d(xz)$, $d(yz)$ and $d(x^2-y^2)$ orbitals are populated. These lower energy orbitals point between the coordinating nitrogen, sulphur and oxygen atoms, mixing less strongly and forming more stable π^* orbitals.

The populating of the $\sigma^*(dz^2)$ orbital in the quartet and sextet spin states pushes the oxygen of the bound water molecule away from the iron, causing the Fe-O distance to increase, as can be seen in Table 4.1 and Table 4.2. The additional population of the $\sigma^*(dxy)$ orbital in the sextet spin state causes the iron to move out of the plane of the porphyrin, leading to a deviation from plane of 0.43 Å, compared to just 0.19 Å for the quartet spin state.

The water is prevented from moving further from the porphyrin in the quartet and sextet spin states by the interaction between the water hydrogen atoms and the closest nitrogen atoms of the porphyrin. Hydrogen bonds between the nitrogen and hydrogen atoms cause the water molecule to tilt so that the interacting atoms are closer together. It is not possible to attribute this interaction to any one molecular orbital; rather the interaction is spread across several of the occupied orbitals, as seen by the higher charges on the two nitrogen atoms closest to the water hydrogen atoms in each of the tilted structures. Two such orbitals are illustrated for each spin state in Figure 4.8. As

can be seen in Figure 4.4, the tilting becomes more pronounced with increase in spin state as the interaction between the iron and oxygen becomes less favourable and the hydrogen bonding becomes the major factor holding the water close to the porphyrin.

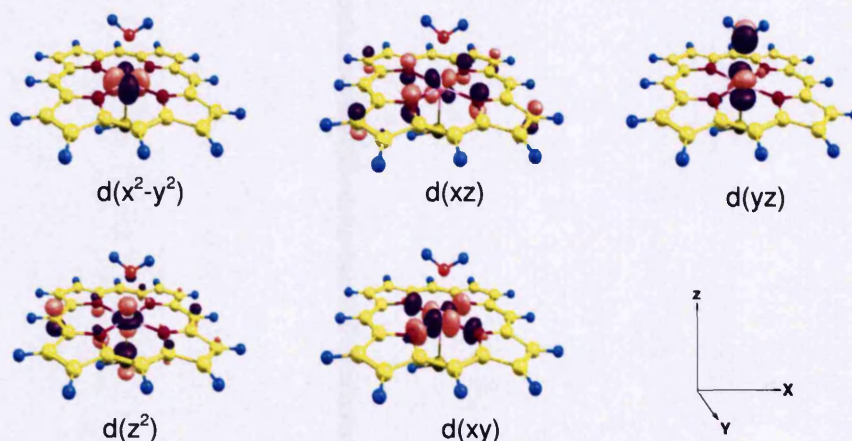


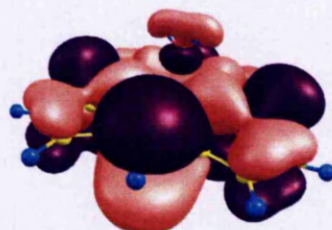
Figure 4.7: Iron anti-bonding d orbitals shown here for the doublet spin upright structure of the water-bound ferric porphyrin species, **1**. Relative energies of the orbitals vary between structures and spin states depending on the electronic environment of the porphyrin.

In their QM/MM studies of the whole reaction cycle Guallar and Friesner (Guallar and Friesner, 2004) found the protein environment decreases the Fe-O distance from 2.18 to 2.10 Å and from 2.65 to 2.46 Å in the doublet and sextet spin states, respectively. They attributed this to repulsive steric interactions with residues in the active site.

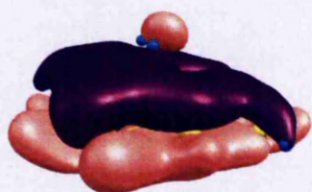
Unlike the Fe-O distance, the Fe-S distance is longest in the quartet state (2.47 Å compared with 2.29 and 2.40 Å in the doublet and sextet spin states respectively). This can be attributed to the population of the iron $d(z^2)$ orbital in the quartet spin species pushing the sulphur away from the porphyrin, whereas in the sextet spin species the $d(xy)$ orbital is also populated, which pushes the iron out of the plane of the porphyrin towards the sulphur.



Doublet spin upright structure
 α_{66}



Doublet spin upright structure
 α_{97}



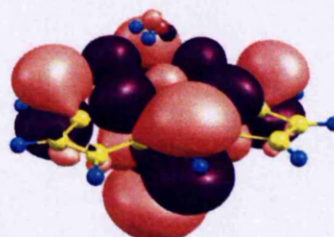
Doublet spin tilted structure
 α_{63}



Doublet spin tilted structure
 α_{97}



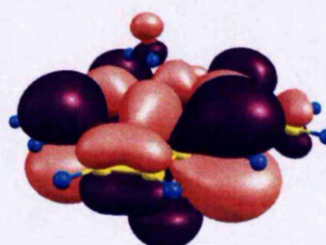
Quartet spin α_{62}



Quartet spin α_{98}



Sextet spin α_{64}



Sextet spin α_{99}

Figure 4.8: Possible α orbitals responsible for hydrogen bonding between hydrogen atoms of the bound water and porphyrin nitrogen atoms in **1**. Diagrams were produced using Chemcraft with a contour value of 0.01.

The unpaired electrons in the structures are predominantly located on the iron, resulting in the spin density being almost entirely on the iron in the low spin state. This can be seen in Table 4.1. The singly occupied $\sigma^*d(z^2)$ orbital in the quartet and sextet spin states mixes with the sulphur $p(z)$ orbital, inducing a spin of 0.49 in both of these spin states. The $d(xy)$ orbital, which is only populated in the sextet spin state, mixes with the porphyrin nitrogen $p(x,y)$ orbitals, producing a spin of 0.1-0.2, which is not present in the lower spin states.

The calculated spin densities on the iron and sulphur atoms match closely those found by Schöneboom and Thiel (Schöneboom and Thiel, 2004) and Rydberg *et al.* (Rydberg *et al.*, 2004). There has been little experimental investigation of the spin delocalisation in P450s. Yang *et al.* (Yang *et al.*, 2006) carried out ENDOR studies on cyanide-bound P450cam, finding an upper limit for the spin density on the sulphur of 0.14. DFT studies on cyano-iron-porphyrin-cysteine gave spin densities of 0.3 – 0.4 on the sulphur atom, confirming that the spin density on the sulphur of cyanide-bound P450 haem can be expected to be higher than that for the water-bound P450 haem.

It can be seen in Table 4.1 that with increasing spin state the charge on the iron becomes more positive and the charge on the oxygen, sulphur and nitrogen atoms more negative. This can be attributed to the higher energy $d(z^2)$ and $d(xy)$ iron orbitals, which are populated in the quartet and sextet spin states, donating some electron density to the coordinating atoms in order to lower the energy of the system.

Harris and Loew (Harris and Loew, 1993), using an INDO/ROHF method, concluded the enzyme environment is necessary for a low spin ground state. The reason for this discrepancy with the results presented here is not clear, but it is probably due to the decreased accuracy of the semi-empirical method used by Harris and Loew, compared with the DFT methods used here. It is worth noting that they report the iron to be out of plane in the low spin state, which suggests that there is some high spin character in the species they report to be low spin, which would destabilise the structure and may explain the discrepancy in the results.

Single point calculations by Green (Green, 1998), on a model system similar to that used in this study, but using methyl mercaptide to model the cysteine ligand instead of SH,

contradicted the conclusions of Harris and Loew. He carried out DFT calculations with the Gaussian 94 program, the B3LYP functional, and a range of basis sets from STO3G to 6-311+G(2d). He found that with increasing basis set size there was increased overlap between the sulphur p(z) and iron d(z^2) orbitals in the e_g and t_{2g} molecular orbitals, which caused the d(z^2) and d(xy) orbitals to be destabilised, resulting in a low spin ground state. They concluded that the enzyme environment was not necessary to give a low spin ground state, in agreement with the results of this study, but in disagreement with Harris and Loew.

4.2.2 Penta-coordinated ferric porphyrin species

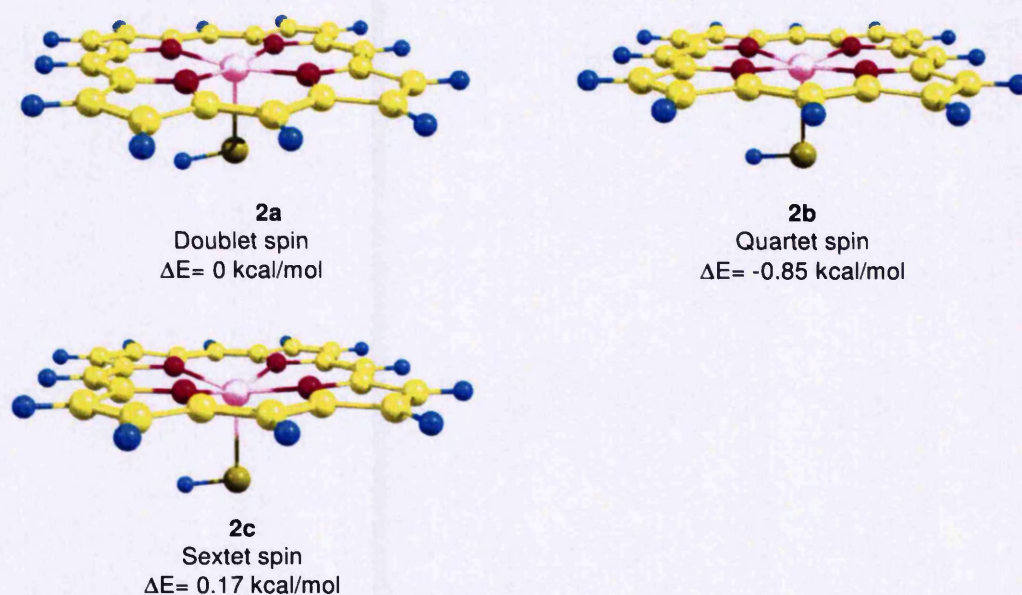


Figure 4.9: Structures and relative energies of doublet, quartet and sextet spin states of penta-coordinated ferric porphyrin.

The first step in the P450 catalytic cycle is the binding of the substrate in the active site. The bound water molecule is displaced from the haem, along with a number of other water molecules in the active site. The resulting entropy increase has been suggested to drive the process (Griffin and Peterson, 1972, Meunier *et al.*, 2004). The crystal structure of P450cam with camphor bound was solved by Poulos *et al.* (Poulos *et al.*, 1987). This gave insight into the structure of the substrate-bound P450.

The experimental observation is that upon substrate binding the P450 goes from low, doublet, to high, sextet spin. This is commonly observed using UV/Vis spectroscopy as a decrease in the peak at 420 nm and an increase at 390 nm. Cammer *et al.* used EPR spectroscopy to first define the observed change in the UV/Vis spectra as a change from doublet to sextet spin (Cammer *et al.*, 1966). Later, this was confirmed by Peterson (Peterson, 1971), using magnetic susceptibility and EPR measurements and Schünemann *et al.* (Schunemann *et al.*, 2002) using a combination of Mössbauer and EPR spectroscopy.

Table 4.4: Summary of the significant bond distances, bond angles, electron spin densities and atomic charges in the doublet, quartet and sextet spin species of the substrate-bound iron III species. Distances are in Å and angles in degrees.

	Doublet	Quartet	Sextet
Fe-S distance	2.31	2.47	2.41
Fe-N_{av} distance	2.01	2.02	2.10
Fe deviation from plane	0.19	0.24	0.46
Fe spin density	1.35	2.46	3.90
S spin density	-0.17	0.50	0.48
N₃ spin density	-0.052	-0.005	0.14
N₄ spin density	-0.052	-0.001	0.14
N₅ spin density	-0.048	-0.007	0.14
N₆ spin density	-0.048	-0.003	0.14
Fe atomic charge	0.42	0.58	0.77
S atomic charge	-0.13	-0.27	-0.25
N₃ atomic charge	-0.82	-0.85	-0.89
N₄ atomic charge	-0.82	-0.84	-0.89
N₅ atomic charge	-0.82	-0.84	-0.89
N₆ atomic charge	-0.82	-0.85	-0.89

The rationale for the change in spin state on substrate binding, is that once the axial water ligand is removed, the $d(z^2)$ anti-bonding orbital is more favoured. Also, the iron is able to move out of the plane of the porphyrin, causing the $d(xy)$ orbital to be more stable.

DFT optimisations using B3LYP and BS1 gave a value of 12.1 kcal mol⁻¹ for the removal of the bound water from the resting state. This is in good agreement with Griffin and Peterson (Griffin and Peterson, 1972) who found that camphor binding to P450cam is

entropy driven, driven by the increase in entropy on binding of camphor and displacement of the water molecules from the active site.

Optimisation with B3LYP and BS1 gave very similar energies for the three spin states. These are shown in Figure 4.9. The occupied orbitals in each spin state are listed in Table 4.5 and the iron d orbitals are displayed in Figure 4.10. Some results of other groups are shown in Table 4.6. Ogliaro *et al.* (Ogliaro *et al.*, 2002b) and Altun and Thiel (Altun and Thiel, 2005) (Table 4.6d and e) also found similar energies for the three spin states when using B3LYP and comparable basis sets to BS1 in the gas phase. Both groups found that the use of higher basis sets gave energies more in line with experimental values, with sextet ground states. QM/MM calculations by Altun and Thiel gave a quartet ground state unless a high basis set was used. QM/MM studies by Guallar and Friesner (Guallar and Friesner, 2004) (Table 4.6f) also gave a quartet ground state. Studies with the slightly higher basis set BS2 (Table 4.6b) did not favour a sextet ground state. Instead the quartet state was stabilised further.

Table 4.5: Iron d orbital energies for the doublet, quartet and sextet spin states of the penta-coordinated iron III species, **2**, in atomic units. Orbitals in red are unoccupied.

		$d(x^2-y^2)$	$d(xz)$	$d(yz)$	$d(z^2)$	$d(xy)$
Doublet	α	-0.28712	-0.23533	-0.24379	-0.11071	-0.04311
	β	-0.24114	-0.22396	-0.05631	-0.06114	-0.00954
Quartet	α	-0.32454	-0.36656	-0.37194	-0.31223	-0.09059
	β	-0.26567	-0.06654	-0.06349	-0.08880	-0.02073
Sextet	α	-0.43559	-0.44192	-0.43734	-0.20324	-0.09025
	β	-0.11121	-0.07745	-0.08076	-0.10132	-0.05651

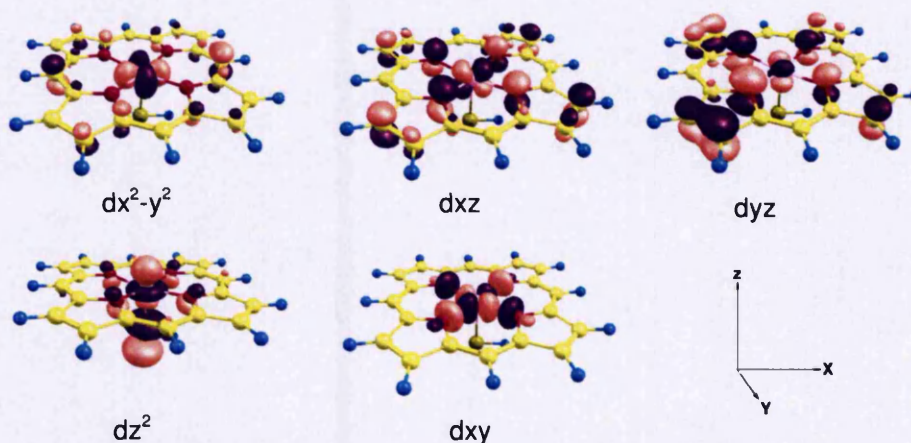


Figure 4.10: Iron anti-bonding d orbitals shown here for the doublet spin state of the substrate-bound ferric porphyrin species.

Studies by Rydberg *et al.* (Rydberg *et al.*, 2004) (Table 4.6c) gave results that favoured the sextet spin state. They found the quartet to be strongly disfavoured, by 12 kcal mol⁻¹. They used a large, all electron basis set for iron, in contrast to the other results cited in Table 4.6, which may be the reason for the difference in results.

The results of this study, and all of the groups cited except Rydberg *et al.*, suggest that the protein environment is necessary for the sextet spin state to be favoured. It was observed in Chapter 3 of this thesis that CYP132 was able to oxidise fatty acids without giving an apparent shift to the high spin state. This suggests that certain P450 structures do not give a shift to the high spin state upon substrate binding. It appears that the high spin state is not strongly favoured in the substrate bound P450 haem, and the balance between high and low spin can be shifted by changes in the protein structure between different P450s.

It is not clear why the quartet spin state is not observed in substrate-bound P450s. A quartet spin ground state was found in this study, as well as the QM/MM studies of Altun and Thiel (Altun and Thiel, 2005) and Guallar and Friesner (Guallar and Friesner, 2004). Guallar and Friesner claim this result to be "in perfect agreement with experimental observations", yet the sextet spin state is consistently observed experimentally for substrate bound P450. For example, the EPR studies of Schünemann *et al.*

(Schunemann *et al.*, 2002) on substrate-bound P450cam only show features attributed to the doublet and sextet spin states.

Table 4.6: Summary of the energy and significant geometries of the doublet, quartet and sextet spin species of the penta-coordinated ferric porphyrin, as found by various different groups using various methods. (a) and (b) are from this study. (a) B3LYP, BS1, (b) B3LYP, BS2, (c) Rydberg *et al.* (Rydberg *et al.*, 2004) B3LYP, Turbomole (d) Ogliaro *et al.* (Ogliaro *et al.*, 2002b), out of brackets - B3LYP, LACVP/6-31G, in brackets – single point calculation with B3LYP, LACV3P1*2/6-3111G*/LACVP/6-31G, (e) Altun and Thiel (Altun and Thiel, 2005), out of brackets - gas phase, B3LYP, in brackets – QM/MM (f) Guallar and Friesner (Guallar and Friesner, 2004), out of brackets - gas phase, inside brackets - QM/MM.

Spin	Group	ΔE /kcal mol ⁻¹	r(Fe-S) /Å	r(Fe-N) (av) /Å
Doublet	a	-0.2	2.31	2.01
	b	0.5	2.31	2.00
	c	2.3	-	-
	d	-0.3 (+4.2)	2.29	2.01
	e	0.6 (3.7)	2.23 (2.30)	2.01 (2.03)
	f	4.0 (3.6)	-	-
Quartet	a	-1.0	2.47	2.02
	b	-1.8	2.48	2.01
	c	12.0	-	-
	d	-1.7 (+4.2)	2.49	2.01
	e	0.9 (-0.7)	2.41 (2.46)	2.02 (2.03)
	f	-0.3 (-2.0)	-	-
Sextet	a	0.0	2.41	2.10
	b	0.0	2.41	2.09
	c	0.00	2.30	2.10
	d	0.0 (0.0)	2.40	2.09
	e	0.0 (0.0)	2.35 (2.40)	2.10 (2.10)
	f	0.0 (0.0)	-	-

The structures of the three spin states are shown in Figure 4.9 and the important bond distances and angles summarised in Table 4.4. In the doublet spin state the iron is close to the plane of the porphyrin, with a relatively short Fe-S distance. This is due to only the low energy $d(x^2-y^2)$, $d(xz)$ and $d(yz)$ orbitals being occupied. In the quartet spin state the $d(z^2)$ orbital is also occupied, giving a longer Fe-S distance, and causing the iron to move slightly out of the porphyrin plane. In the sextet spin state the $d(xy)$ orbital is also occupied, which pushes the iron further out of the plane of the porphyrin, giving longer Fe-N distances and a slightly shorter Fe-S distance.

There was little difference with the geometries found by the other groups cited in Table 4.6, although the gas phase calculations of Altun and Thiel gave consistently shorter

Fe-S distances. This is probably due to the inclusion of diffuse orbitals on the sulphur atom. Examination of the supporting information given by Altun and Thiel reveals that the Fe-S distance decreases with increasing basis set complexity.

4.2.3 Penta-coordinated ferrous porphyrin species

The next step in the catalytic cycle is the reduction of the substrate-bound iron III species to give a substrate-bound iron II species. This again has a high spin ground state. The singlet, triplet and quintet spin states of the penta-coordinated iron II were optimised using B3LYP and BS1. The resulting structures and energies are shown in Figure 4.11 and the significant geometries, spins and charges shown in Table 4.8. The energies and occupation of the iron d anti-bonding orbitals are shown in Table 4.7. The results of some other groups are summarised in Table 4.9.

The energy of the quintet spin ground state of penta-coordinated ferrous porphyrin, **3c**, relative to the sextet spin ground state of the ferric porphyrin, **2c** is $-53 \text{ kcal mol}^{-1}$. This represents a large stabilisation on the reduction of penta-coordinated ferric porphyrin to iron II. The relative energies of all species in this study are summarised in Figure 4.18, at the end of the chapter. Ogliaro *et al.* (Ogliaro *et al.*, 2002b) studied this process with a range of axial ligands in place of cysteine and found that iron II is even more favoured when cysteine is not present. They concluded that cysteine gates the process, preventing auto-reduction of the system without an appropriate electron donation system.

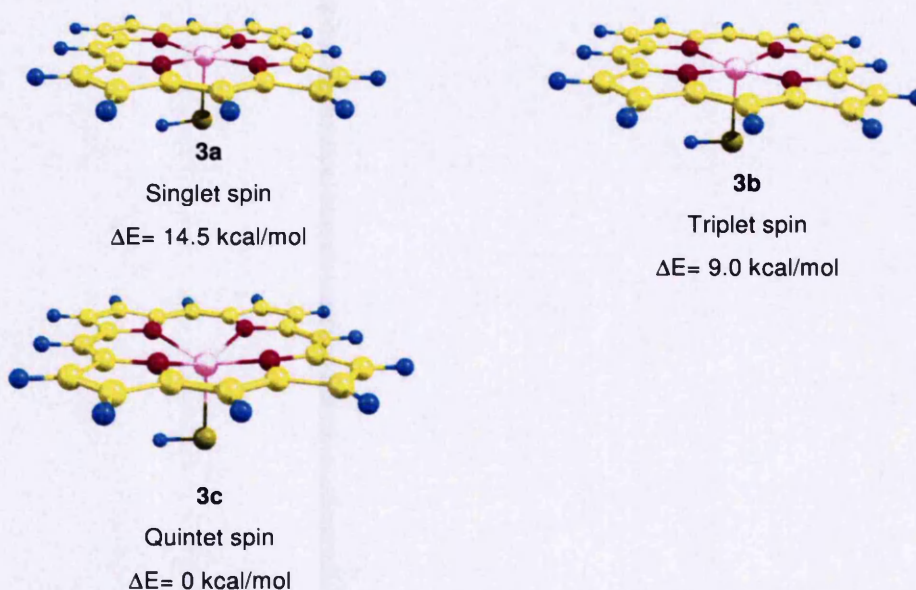


Figure 4.11: Structures and relative energies of singlet, triplet and quintet spin states of penta-coordinated ferrous porphyrin.

The energies of the three spin states follow the trend singlet>triplet>quintet. This is consistent across all of the results in Table 4.9. Again, the reason for this is the lack of a sixth ligand to destabilise the $d(z^2)$ orbital and hold the iron in the plane of the porphyrin. The Fe-S distances are also fairly consistent for each spin state between the results from the groups. From Table 4.8 it appears that the additional electron in the iron II species is shared among the iron and sulphur, with slightly less electron density residing on the nitrogen atoms.

Table 4.7: Iron d orbital energies for the singlet, triplet and quintet spin states of the penta-coordinated iron II species, **3**, in atomic units. Orbital in red are unoccupied.

		$d(x^2-y^2)$	$d(xz)$	$d(yz)$	$d(z^2)$	$d(xy)$
Singlet	α	-0.07280	-0.05878	-0.03063	0.09692	0.16069
	β	-0.07280	-0.05878	-0.03063	0.09692	0.16069
Triplet	α	-0.12460	-0.08757	-0.11148	-0.14390	0.10745
	β	-0.08174	-0.04363	0.11678	0.11957	0.15430
Quintet	α	-0.12873	-0.15333	-0.15887	-0.18334	-0.06812
	β	-0.04892	0.11679	0.12015	0.10771	0.20903

Table 4.8: Summary of the significant bond distances, bond angles, electron spin densities and atomic charges in the singlet, triplet and quintet spin species of the substrate-bound iron II species. Distances are in Å and angles in degrees.

	Singlet	Triplet	Quintet
Fe-S distance	2.38	2.55	2.47
Fe-N_{av} distance	2.02	2.04	2.16
Fe deviation from plane	0.15	0.22	0.64
Fe spin density	-	1.92	3.65
S spin density	-	0.20	0.14
N₃ spin density	-	-0.012	0.042
N₄ spin density	-	-0.07	0.043
N₅ spin density	-	-0.012	0.042
N₆ spin density	-	-0.008	0.043
Fe atomic charge	0.26	0.41	0.53
S atomic charge	-0.33	-0.48	-0.45
N₃ atomic charge	-0.78	-0.80	-0.83
N₄ atomic charge	-0.78	-0.81	-0.82
N₅ atomic charge	-0.78	-0.80	-0.82
N₆ atomic charge	-0.78	-0.81	-0.83

Table 4.9: Summary of the energy and significant geometries of the singlet, triplet and quintet spin species of the penta-coordinated ferrous porphyrin, as found by various different groups using various methods. (a) This study, B3LYP, BS1, (b) Rydberg *et al.* (Rydberg *et al.*, 2004) B3LYP, Turbomole, (c) Ogliaro *et al.* (Ogliaro *et al.*, 2002b), out of brackets - B3LYP, LACVP/6-31G, in brackets - single point calculation with B3LYP, LACV3P1*2/6-3111G*/LACVP /6-31G, (d) Altun and Thiel (Altun and Thiel, 2005) out of brackets - gas phase, B3LYP, in brackets - QM/MM

Spin	Group	ΔE /kcal mol ⁻¹	r(Fe-S) /Å	r(Fe-N) (av) /Å
Singlet	a	14.5	2.38	2.02
	b	11.3	-	-
	c	12.1(14.1)	2.38	2.02
	d	15.5 (15.9)	2.34 (2.37)	2.02 (2.04)
Triplet	a	9.0	2.55	2.04
	b	7.6	-	-
	c	5.9 (9.4)	2.59	2.03
	d	10.1 (5.6)	2.53 (2.70)	2.04 (2.05)
Quintet	a	0.0	2.47	2.16
	b	0.0	2.39	2.17
	c	0.0 (0.0)	2.50 (2.47)	2.02
	d	0.0 (0.0)	2.44 (2.48)	2.17 (2.15)

4.2.4 Dioxygen-bound ferrous porphyrin

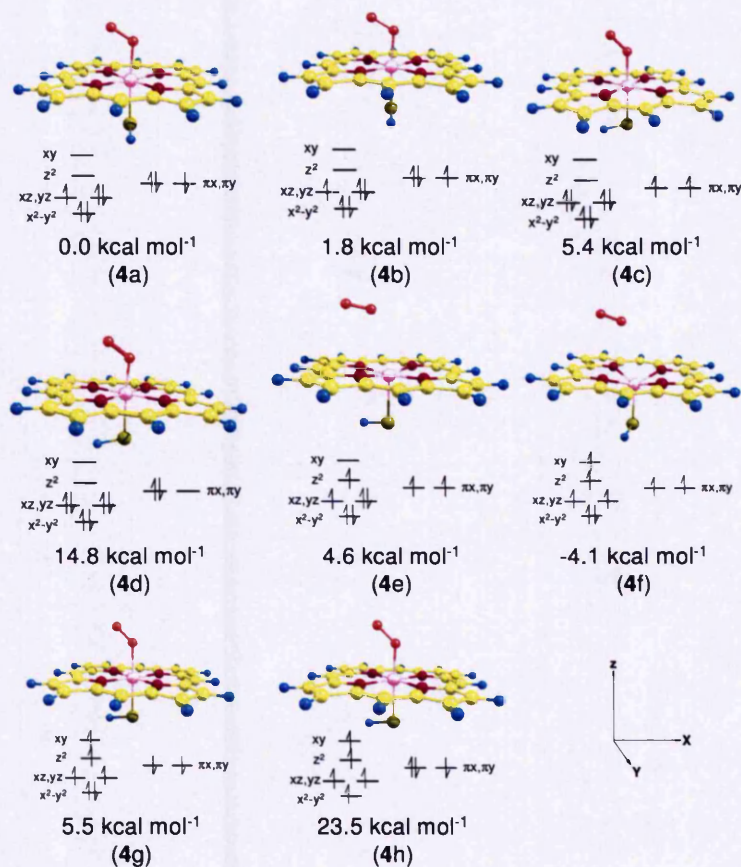


Figure 4.12: Structures and electronic distributions of the various spin states calculated for the dioxygen-bound ferrous complex.

The next step in the catalytic cycle is the binding of dioxygen to the penta-coordinated iron II haem. This is the last metastable species in the catalytic cycle and has been analysed by a number of experimental methods. Resonance Raman studies on model compounds (Chottard *et al.*, 1984) show the binding of dioxygen to be end-on. This is confirmed by an x-ray crystal structure of a model compound by Schappacher *et al.* (Schappacher *et al.*, 1987). The ground state is considered to be a low spin singlet (Shaik *et al.*, 2005a). This can not easily be determined experimentally as all spin states of the complex are EPR silent.

As was seen in the previous section, the ground state of iron II deoxyheme is a high spin quintet. The ground state of dioxygen is a triplet. This means in order to produce a singlet complex, spin crossing must occur. Chemical reactions cannot normally change the spin state of an electron, such reactions are said to be spin forbidden and therefore occur slowly. If the potential energy surfaces of two spin states cross, electrons can jump from one state to another, this is called spin crossing.

Table 4.10: Summary of the significant bond distances, bond angles, electron spin densities and atomic charges in the various spin species studied of the dioxygen-bound ferrous porphyrin. The structure and electronic distribution of each species is shown in Figure 4.12. Distances are in Å and angles in degrees.

	(4a)	(4b)	(4c)	(4d)	(4e)	(4f)	(4g)	(4h)
Fe-O distance	2.06	1.95	2.02	1.78	3.42	3.65	2.05	1.95
O-O distance	1.28	1.30	1.25	1.28	1.22	1.22	1.25	1.32
Fe-S distance	2.45	2.41	2.40	2.40	2.54	2.47	2.41	2.42
Fe-N av distance	2.03	2.03	2.03	2.03	2.04	2.16	2.03	2.09
Fe deviation from plane	0.062	0.036	0.029	0.0086	0.22	0.63	0.033	0.019
Fe-O-O angle	117.9	117.3	136.0	122.5	116.0	118.7	135.7	132.8
Fe spin density	-1.43	1.08	0.36	0	1.93	3.64	3.52	2.83
S spin density	-0.07	0.045	0.077	0	0.21	0.14	0.26	-0.23
O₁ spin density	0.67	0.37	0.78	0	0.99	1.01	1.01	0.54
O₂ spin density	0.74	0.57	0.84	0	0.97	0.97	1.01	0.75
Fe atomic charge	0.43	0.44	0.33	0.40	0.39	0.51	0.39	0.56
S atomic charge	-0.35	-0.30	-0.31	-0.31	-0.47	-0.46	-0.39	-0.24
N av atomic charge	-0.79	-0.79	-0.78	-0.78	-0.81	-0.82	-0.81	-0.83
O₁ atomic charge	-0.17	-0.23	-0.093	-0.16	0.017	0.033	0.013	-0.21
O₂ atomic charge	-0.18	-0.23	-0.12	-0.20	-0.025	-0.027	0.018	-0.17

Jensen and Ryde (Jensen and Ryde, 2004) carried out DFT calculations on the binding of dioxygen to a model myoglobin haem, using histidine as the axial ligand. Fully relaxed potential energy curves of the seven lowest spin states revealed a broad crossing region at a Fe-O distance around 2.5 Å, with five spin states within 4 kcal mol⁻¹, with almost parallel potential energy surfaces. They estimated that the combined effects of the haem and axial ligand on the spin crossing ability of the complex may accelerate the binding of dioxygen to haem compared with iron alone by a factor of up to 10¹¹. It is likely that the P450 haem has a similar effect on the probability of spin crossing.

DFT calculations by Harris *et al.* (Harris *et al.*, 1998) and Rydberg *et al.* (Rydberg *et al.*, 2004) on the dioxygen-bound ferrous porphyrin revealed an open shell singlet ground

state with unpaired electrons in the π^* iron d(yz) and π^* OO orbitals, these have opposite spins, giving an overall spin of 0.

There are a large number of possible electronic distributions in the dioxygen-bound ferrous porphyrin complex. The charge distribution can be either $\text{Fe}^{2+}(\text{O}_2)^0$ or $\text{Fe}^{3+}(\text{O}_2)^-$. The two oxidation states of iron each have three possible spin states, and the unpaired electrons on dioxygen may or may not line up with those on the iron. Due to this range of possibilities it was considered prudent to allow Gaussian to find the electronic configurations of lowest energy, with some 'tweaking' to guide it towards other important spin states.

The spin states and relative energies which were solved are shown in Figure 4.12. Using total spins of 0, 2, 4 and 6 unpaired electrons (spin multiplicity 1, 3, 5 and 7) and allowing Gaussian to assign the spin distribution gave the structures (4b), (4d), (4e) and (4f). 4d is a closed shell singlet, with identical electronic distributions in the α and β electrons.

In order to obtain the open shell singlet 4a, as found by Harris *et al.* (Harris *et al.*, 1998) and Rydberg *et al.* (Rydberg *et al.*, 2004), the Guess=(Read,Alter) Gaussian keyword was used, starting from the triplet spin structure (4b) and swapping the alpha and beta d(xz) orbitals.

The remaining three spin states, 4c, 4g and 4h, were obtained using the counterpoise correction function in Gaussian 03. This allows the spin multiplicity of individual fragments to be defined. The fragments used were the dioxygen molecule, and the iron-porphyrin. This is not a perfect method, as it is not possible to draw a direct comparison between values obtained using different methods. The energies obtained using the counterpoise keyword will be expected to be slightly too high for a number of reasons. Firstly, the true wavefunction will not be able to be obtained as the spin multiplicity on each fragment is fixed as an integer. Also the structures are converged in order to minimise the counterpoise corrected energy. From other counterpoise calculations carried out for Chapter 5, it is expected that this will give an energy that is around 1 kcal mol⁻¹ higher than that obtained without counterpoise correction.

Attempts to obtain structures for **4c**, **4g** and **4h** without the counterpoise correction failed to converge, or fell down to the lower energy structures with the same spin multiplicity. This was attempted both starting from the counterpoise corrected wavefunctions, using the vshift keyword to prevent falling down to a lower energy wavefunction, and using guess=alter from a similar structure, also with a vshift. Further work is needed in order to obtain reliable energies for these three structures. Time did not allow for this however.

Table 4.11: Summary of the energies and significant geometries found by other groups for the various spin states of the dioxygen-bound ferrous porphyrin. (a) This study (b) Rydberg *et al.* (Rydberg *et al.*, 2004), B3LYP, turbomole (c) Harris *et al.* (Harris *et al.*, 1998) BLYP.

Spin	Group	ΔE /kcal mol ⁻¹	r(Fe-O) /Å	r(Fe-S) /Å	r(O-O) /Å	r(Fe-N) (av) /Å
Open shell singlet	a	0.0	2.06	2.45	1.28	2.03
	b	0.0	1.95	2.32	1.31	2.02
	c	0.0	1.83	2.37	1.32	2.05
Singlet	a	14.8	1.78	2.40	1.28	2.03
	b	18.2	-	-	-	-
	c	-	-	-	-	-
Triplet	a (4e)	4.6	3.42	2.54	1.22	2.04
	b	3.2	-	-	-	-
	c	0.8	-	-	-	-
Quintet	a (4f)	-4.1	3.65	2.47	1.22	2.16
	a (4g)	5.5	2.05	2.41	1.25	2.03
	b	14.5	-	-	-	-
	c	-	-	-	-	-

The structures, energies and electronic distributions are shown in Figure 4.12. The lowest energy species was the heptet spin state, **4f**. It can be seen in Table 4.10 that the Fe-O distance in **4f** is 3.65 Å. This means the oxygen molecule is essentially dissociated from the porphyrin, giving the ground state **3c**, and triplet oxygen. The lowest energy structure in which the dioxygen is bound to the porphyrin is the open shell singlet structure, **4a**. This is in agreement with the EPR results mentioned above (Davydov *et al.*, 1991, Kobayashi *et al.*, 1994), and the DFT calculations of Harris *et al.* (Harris *et al.*, 1998) and Rydberg *et al.* (Rydberg *et al.*, 2004). The triplet spin structure **4d** is 1.8 kcal mol⁻¹ above this. This value falls between those obtained by Harris *et al.* (0.8 kcal mol⁻¹) (Harris *et al.*, 1998) and Rydberg *et al.* (3.2 kcal mol⁻¹) (Rydberg *et al.*, 2004), see Table 4.11.

Structures **4c** and **4g** are only 5.4 and 5.5 kcal mol⁻¹ higher in energy than **4a** respectively. These two structures were obtained using the counterpoise correction method, and so it is likely that their relative energies are in fact even lower, perhaps by 2 or 3 kcal mol⁻¹. These four spin states close in energy suggests that the true spin state of the dioxygen-bound ferrous species may be an equilibrium between a number of spin states, with electron transfer taking place between the porphyrin and bound oxygen.

4.2.5 Dioxygen-bound reduced ferrous porphyrin

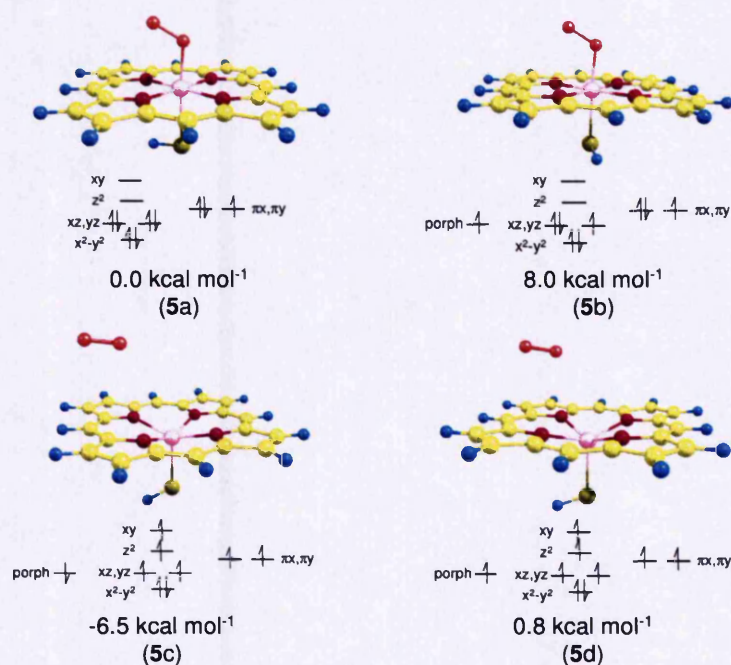


Figure 4.13: Structures and relative energies of doublet, quartet, sextet and octet spin states of dioxygen-bound reduced ferrous porphyrin.

The next step in the catalytic cycle is the reduction of the dioxygen-bound ferrous porphyrin, **4**, to give the reduced dioxygen ferrous porphyrin species, **5**. The reduction of **4** is the rate determining step in the cycle in at least the case of P450cam (Davydov *et al.*, 2001), this means **5** cannot be observed under normal experimental conditions. Davydov *et al.*, using cryo-reduction techniques on a D251N mutant of P450cam (Davydov *et al.*, 1999), and found that the impaired proton delivery system of this mutant

allowed **5** to be observed at temperatures below 77 K. Later they found that **5** could be trapped in wild type P450cam at ~6 K (Davydov *et al.*, 2001).

Shaik *et al.* (Shaik *et al.*, 2005a) reported the reduction of **4** to be highly endothermic in small gas phase systems, such as in this study, but to become exothermic when a dielectric medium is used. They found that when more of the surrounding residues are taken into account the reduction becomes more exothermic, suggesting the hydrogen bonding interactions with the surrounding residues and water molecules are important for the stabilisation of **5**.

Table 4.12: Summary of the significant bond distances, bond angles, electron spin densities and atomic charges in the doublet, quartet and sextet spin species of the O₂-bound iron III species. Distances are in Å and angles in degrees.

	Doublet	Quartet	Sextet	Octet
Fe-O distance	1.95	1.96	4.85	3.90
O-O distance	1.32	1.30	1.26	1.22
Fe-S distance	2.59	2.45	2.54	2.54
Fe-N _{av} distance	2.03	2.04	2.16	2.17
Fe deviation from plane	0.01	0.03	0.58	0.62
Fe-O-O angle	118.4	118.5	121.5	118.2
Fe spin density	0.01	1.13	3.66	3.66
S spin density	-0.00	0.01	0.11	0.11
N ₅ spin density	0.00	-0.04	0.04	0.17
N ₆ spin density	0.00	0.10	-0.02	0.03
N ₇ spin density	0.00	-0.05	0.05	0.18
N ₈ spin density	0.00	0.10	0.00	0.03
O ₁ spin density	0.45	0.38	0.85	0.98
O ₂ spin density	0.57	0.55	0.80	0.94
Fe atomic charge	0.33	0.46	0.54	0.52
S atomic charge	-0.53	-0.38	-0.53	-0.53
N ₅ atomic charge	-0.75	-0.79	-0.83	-0.84
N ₆ atomic charge	-0.74	-0.81	-0.84	-0.83
N ₇ atomic charge	-0.75	-0.80	-0.82	-0.84
N ₈ atomic charge	-0.74	-0.81	-0.83	-0.82
O ₁ atomic charge	-0.27	-0.23	-0.12	0.00
O ₂ atomic charge	-0.31	-0.25	-0.20	-0.06

EPR and ENDOR studies by Davydov *et al.* reveal **5** to have a low spin ground state (Davydov *et al.*, 2001). DFT studies by Harris *et al.* (Harris *et al.*, 1998) showed reduction of **4** to fill the oxygen $\pi^*(yz)$ orbital to give the doublet spin species. Rydberg *et al.* also find a doublet ground state (Rydberg *et al.*, 2004). Both groups find an increased

energy gap between the low and intermediate spin states compared with **4**. This is more pronounced in the results of Harris *et al.*, who found the intermediate spin state to have an energy of 11.3 kcal mol⁻¹ relative to the doublet. Neither groups report results for the sextet spin state.

Optimisation of the doublet and quartet spin states with B3LYP and BS1 produced a doublet spin state, in agreement with the work of Harris *et al.* and Rydberg *et al.*, and with the findings of Davydov *et al.* The structures, spin states and relative energies of these states are shown in Figure 4.13. The ground state, **5a** is 44.2 kcal mol⁻¹ higher in energy than **4a**. This is in agreement with the findings of Shaik *et al.* for small systems without a dielectric medium (Shaik *et al.*, 2005a).

Table 4.13: Summary of the energy and significant geometries of the doublet and quartet spin states of the dioxygen-bound reduced ferrous porphyrin, as found by various different groups using various methods. (a) This study, B3LYP, BS1, (b) Rydberg *et al.* (Rydberg *et al.*, 2004) B3LYP, Turbomole, (c) Harris *et al.* (Harris *et al.*, 1998) BLYP. The sextet/octet spin states are not reported by other groups.

Spin	Group	ΔE /kcal mol ⁻¹	r(Fe-O) /Å	r(Fe-S) /Å	r(O-O) /Å	r(Fe-N) (av) /Å
Doublet	a	0.0	1.95	2.59	1.32	2.03
	b	0.0	1.99	2.40	1.33	2.02
	c	0.0	1.95	2.60	1.35	2.03
Quartet	a	8.0	1.96	2.45	1.30	2.04
	b	4.9	-	-	-	-
	c	11.3	-	-	-	-

The relative energy of the quartet spin state of 8.0 kcal mol⁻¹ is between those found by Harris *et al.* and Rydberg *et al.* It can be seen in Figure 4.13 that, rather than the iron d(z²) α orbital being occupied, an additional orbital on the porphyrin has been occupied. This is evidenced by increased spin density on a number of carbon atoms. It appears that there is a strong interaction between the superoxide molecule and the iron d(z²) orbital, and it is more energetically favourable to occupy an additional orbital on the porphyrin. This was seen in **4h**, with a high energetic cost to occupying the d(z²) orbital and with three electrons in the dioxygen π^* orbitals.

Many attempts to optimise the sextet spin state failed to converge. A single point calculation was carried out starting from the lowest energy, non-converged structure. This is shown in Figure 4.13. The octet spin state was also optimised. The two high spin

states have similar wavefunctions, with opposite spins on the porphyrin. In both structures the dioxygen has dissociated from the porphyrin. The ground state dioxygen-bound structure is thus the doublet spin state.

4.2.6 Hydroperoxo ferric porphyrin – Compound 6

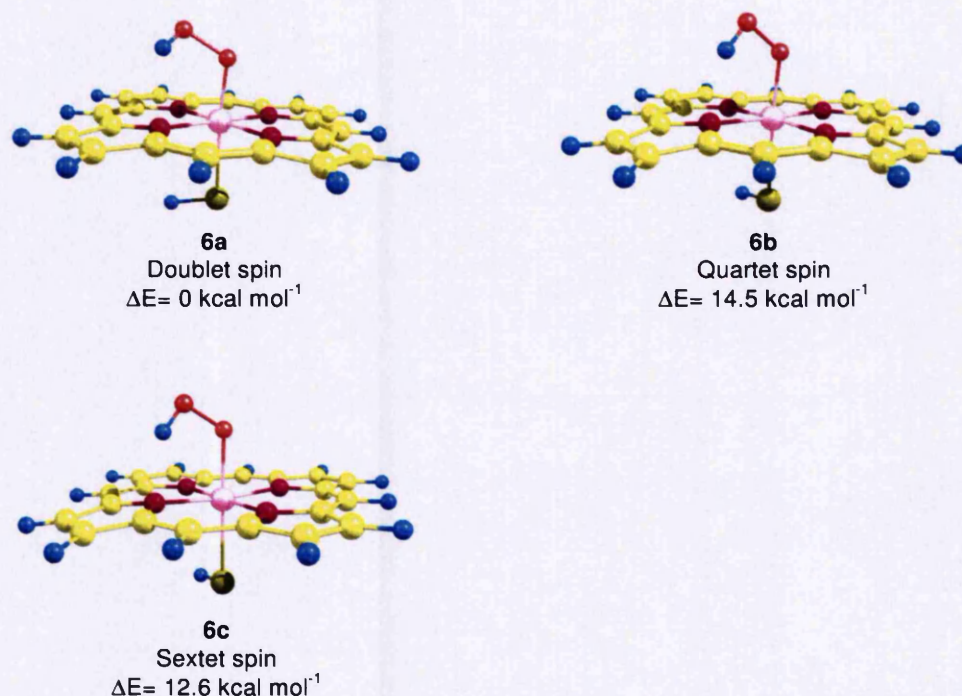


Figure 4.14: Structures and relative energies of doublet, quartet and sextet spin states of hydroperoxo ferric porphyrin.

The next step in the catalytic cycle is the protonation of **5** to form **6**. This is the last species that has been observed experimentally before the hydroxylated product is produced. Davydov *et al.* (Davydov *et al.*, 1999, Davydov *et al.*, 2001) found that **6** was formed from **5** when the cryoreduced D251N mutant of P450cam, mentioned in the previous section, was transiently warmed to 180 K. Upon further heating, the 5-exo-hydroxycamphor product was produced. Compound **7**, which will be discussed in the next section, was not observed, but can be inferred from the nature of the hydroxylation.

Davydov *et al.* found that the T252A mutant of P450cam also formed **6**, but was unable to produce hydroxylated product. Instead hydrogen peroxide is produced. They postulated that **6** is a key intermediate, at or near to the branch point, leading either to product formation, or to uncoupling and the production of hydrogen peroxide.

The ongoing debate as to the role of **6** in the catalytic cycle, and whether it is the true oxidant species, possibly in addition to Compound I, was discussed in section 4.1.1. In the course of this debate, many theoretical and experimental studies have been carried out on **6**. Guallar *et al.* and Harris *et al.* (Guallar *et al.*, 2002, Harris and Loew, 1998, Harris, 2002) studied the hydroxylation of **5** to form **6** in P450eryF, using a combination of DFT and MM techniques. Later Kamachi *et al.* and Zheng *et al.* (Kamachi and Yoshizawa, 2003, Zheng *et al.*, 2006) used B3LYP to study the process in P450cam. Newcomb *et al.* suggested that two oxidant species may be involved in the oxidation reaction, and that **6** may one of them (Newcomb *et al.*, 2003, Newcomb and Chandrasena, 2005). Shaik *et al.* hold that the second oxidant species is an alternative spin state of Compound I (Altun *et al.*, 2007, Kumar *et al.*, 2005, Shaik *et al.*, 2004). Bach *et al.* (Bach and Dmitrenko, 2006) suggested that the true oxidant species may be a 'somersault' species, which is an isomer of **6**.

Various protonation pathways have been studied by DFT and MM methods, with differing arrangements of water molecules and other groups in the active site (Guallar and Friesner, 2004, Kamachi and Yoshizawa, 2003, Zheng *et al.*, 2006). The involvement of a range of acid and alcohol groups has been considered. There is some debate as to which groups are involved, but the consensus is that the protonation is exothermic only when the alcohol or acid groups are included in the model (Shaik *et al.*, 2005a).

Harris and Loew (Harris and Loew, 1998) studied the effect of protonating the distal and proximal oxygen atoms of the bound oxygen molecule. They found that distal protonation weakens the O-O bond and upon a second protonation leads to formation of Compound I, whereas proximal protonation weakens the Fe-O bond and leads to decoupling with the formation of hydrogen peroxide. For this reason only distal protonation was considered in this study.

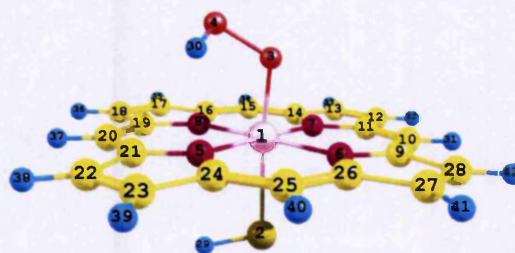


Figure 4.15: Hydroperoxo ferric porphyrin, showing the numbering used in Table 4.14

Table 4.14: Summary of the significant bond distances, bond angles, electron spin densities and atomic charges in the doublet, quartet and sextet spin species of the hydroperoxo ferric porphyrin species. Distances are in Å and angles in degrees. The atom numbering is shown in Figure 4.15

	Doublet	Quartet	Sextet
Fe-O distance	1.83	1.89	2.03
O-O distance	1.45	1.43	1.40
Fe-S distance	2.42	2.48	2.55
Fe-N₅ distance	2.03	2.19	2.07
Fe-N₆ distance	2.04	1.99	2.09
Fe-N₇ distance	2.04	2.13	2.13
Fe-N₈ distance	2.02	1.99	2.11
N₅-H₃₀ distance	2.16	1.93	1.95
Fe deviation from plane	0.015	0.054	0.088
Fe-O-O angle	116.1	118.5	119.5
Fe spin density	0.88	2.62	3.86
S spin density	0.031	0.030	0.26
N₅ spin density	-0.013	0.041	0.068
N₆ spin density	-0.098	-0.031	0.083
N₇ spin density	-0.0098	0.059	0.085
N₈ spin density	-0.012	-0.032	0.085
O₁ spin density	0.15	0.27	0.49
O₂ spin density	0.0082	0.066	0.13
Fe atomic charge	0.41	0.52	0.66
S atomic charge	-0.32	-0.35	-0.42
N₅ atomic charge	-0.81	-0.85	-0.90
N₆ atomic charge	-0.77	-0.81	-0.85
N₇ atomic charge	-0.77	-0.79	-0.86
N₈ atomic charge	-0.77	-0.81	-0.84
O₁ atomic charge	-0.36	-0.36	-0.29
O₂ atomic charge	-0.45	-0.45	-0.42

Table 4.15: Summary of the energy and significant geometries of the doublet, quartet and sextet spin states of the protonated dioxygen-bound ferric porphyrin, as found by various groups using various methods. (a) This study, B3LYP, BS1, (b) Harris *et al.* (Harris and Loew, 1998), BPW91-DZ, DGauss, (c) Ogliaro *et al.* (Ogliaro *et al.*, 2002a) B3LYP LACVP, (d) Rydberg *et al.* (Rydberg *et al.*, 2004) B3LYP Turbomole, (e) Kamachi *et al.* (Kamachi and Yoshizawa, 2003) B3LYP, LANL2DZ, porphyrin and active site residues and waters included (f) Guallar *et al.* (Guallar and Friesner, 2004) B3LYP, LACVP QM/MM (g) Shaik *et al.* (Shaik *et al.*, 2005a) B3LYP, LACVP

Spin	Group	ΔE /kcal mol ⁻¹	r(Fe-O) /Å	r(Fe-S) /Å	r(O-O) /Å	r(Fe-N) (av) /Å
Doublet	a	0.0	1.83	2.42	1.45	2.04
	b	-	1.89	2.32	1.46	2.03
	c	0.0	1.86	2.41	1.52	2.03
	d	0.0	1.88	2.34	1.46	2.03
	e	0.0	1.85	2.59	1.54	2.01
	f	-	1.81	2.42	1.48	-
	g	-	1.87	2.42	1.52	-
Quartet	a	14.5	1.89	2.48	1.43	2.08
	b	-	-	-	-	-
	c	16.8	2.03	2.69	1.51	2.03
	d	33.6	-	-	-	-
	e	7.7	2.04	2.88	1.53	2.02
	f	-	-	-	-	-
	g	-	-	-	-	-
Sextet	a	12.6	2.03	2.55	1.40	2.10
	b	-	-	-	-	-
	c	14.9	2.03	2.69	1.51	2.03
	d	75.8	-	-	-	-
	e	-	-	-	-	-
	f	-	-	-	-	-
	g	-	-	-	-	-

The doublet, quartet and sextet spin states of the distal protonated dioxygen-bound ferric porphyrin, **6**, were calculated using B3LYP and BS1. The structures and relative energies of the three spin states are shown in Figure 4.14. The significant bond distances, angles, spin and charge densities are shown in Table 4.14. These are compared with results in the literature in Table 4.15. The ground state was found to be the doublet spin, with quartet and sextet spin states 14.5 and 12.6 kcal mol⁻¹ higher in energy respectively. This is in good agreement with Ogliaro *et al.* (Ogliaro *et al.*, 2002a) who found relative energies for the quartet and sextet spin of 16.8 and 14.9 kcal mol⁻¹ respectively. The high energies of the higher spin states found by Rydberg *et al.* (Rydberg *et al.*, 2004) do not fit with the results of this study, or with those of other groups. The origin of this discrepancy is not clear as the other results of that group cited previously fit well with the rest of the literature.

The Fe-O distance increases slightly with the increase in spin state, but the hydroperoxo group remains bound in all calculations. This is confirmed by the results of Ogliaro *et al.* (Ogliaro *et al.*, 2002a) and Kamachi *et al.* (Kamachi and Yoshizawa, 2003). The hydroperoxo group hydrogen atom forms a hydrogen bond with a porphyrin nitrogen atom, which aids in holding the hydroperoxo group close to the porphyrin in the higher spin states.

Table 4.16: Iron d orbital energies for the doublet, quartet and sextet spin states of hydroperoxo ferric porphyrin species, 6, in atomic units. Orbitals in red are unoccupied

		$d(x^2-y^2)$	$d(xz)$	$d(yz)$	$d(z^2)$	$d(xy)$
Doublet	α	-0.14014	-0.10888	-0.12233	0.09583	0.09886
	β	-0.11527	-0.09673	0.07318	0.11537	0.12272
Quartet	α	-0.16098	-0.20168	-0.25235	-0.05092	0.05019
	β	-0.10064	0.07947	0.06699	0.10928	0.15064
Sextet	α	-0.28401	-0.26519	-0.26055	-0.21647	-0.09955
	β	0.04648	0.05930	0.05366	0.10391	0.10955

It is not clear why the sextet spin state is lower in energy than the quartet spin state, both in this study and in the results of Ogliaro *et al.* (Ogliaro *et al.*, 2002a). The slightly increased Fe-O distance probably plays a part. Also, the increase in charge on the iron, and corresponding decrease in charge on the oxygen, nitrogen and sulphur atoms may strengthen the nitrogen-hydrogen hydrogen bond, although the N-H distance is longer in the sextet spin state. The Fe-S bond is longer in the high spin state. In the protein it would be more difficult for the Fe-S bond distance to increase. This may mean the stability of the sextet spin state compared to the quartet is not representative of the situation in the protein environment.

The orbitals assigned to the iron d anti-bonding orbitals are shown in Table 4.16. The assignment of the quartet spin orbitals was difficult as the d orbital electron density was spread over several orbitals. It was found with other calculations that 'tidier' wavefunctions, with lower energies, could be obtained by optimising with BP86 followed by re-optimisation with B3LYP. However, this did not help in this case. Also using higher basis sets prior to re-optimisation with BS1 did not produce lower energies. Despite not being able to produce a lower energy 'tidier' wavefunction, it may be that one exists, and this is the reason that the quartet spin state was higher in energy than the sextet spin state.

4.2.7 Compound I

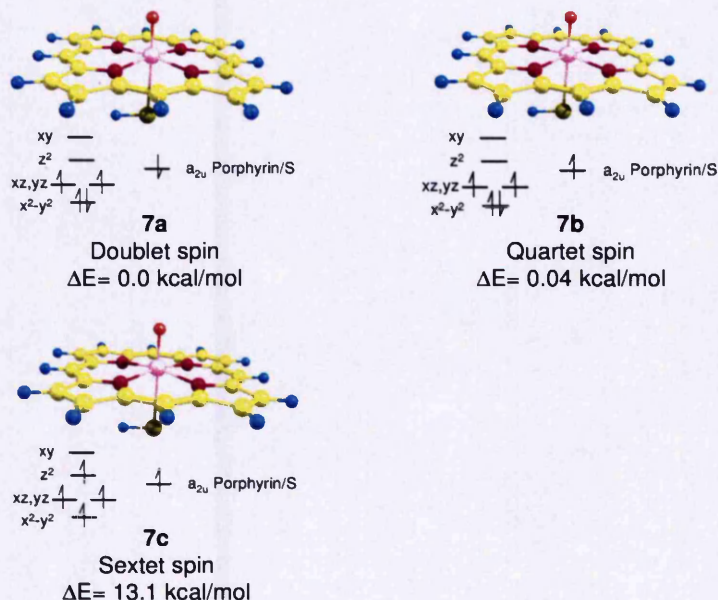


Figure 4.16: Structures and relative energies of doublet, quartet and sextet spin states of Compound I.

The last species in the reaction cycle which will be covered in this study is the high valent, oxygen-bound species commonly referred to as Compound I (**7**). This is generally accepted to be the primary, if not the sole oxidant species. This was first observed as a green compound produced upon the addition of hydrogen peroxide to horseradish peroxidase (George, 1952).

Compound I is thought to be formed from the hydroperoxo species, **6**, following the second protonation of **6** and subsequent release of water. The remaining oxygen atom receives single electrons from the iron, giving iron IV, and from the porphyrin, giving a positive radical, and forms an iron-oxygen double bond. It is perhaps unsurprising that this triradical species is extremely reactive, being a strong oxidant, and it has not been possible to isolate this species experimentally for more than a few milliseconds.

Compound I in P450s has proved elusive. Cryogenic studies by Davydov *et al.* (Davydov *et al.*, 2001), using gamma radiation as the source of electrons, revealed the dioxygen-bound species 5, and the hydroperoxo species 6, but Compound I, 7, was not observed. Egawa *et al.* (Egawa *et al.*, 1994), using rapid scan absorption spectrometry, observed a spectrum similar to that of HRP Compound I 10 ms after the addition of *m*-chloroperbenzoate (*m*-CPBA) to P450cam. Kellner *et al.* (Kellner *et al.*, 2002), used stopped-flow rapid mixing techniques with CYP119 and *m*-CPBA. They also observed the characteristic broad absorption at 370 nm, with additional absorptions at 610 and 690 nm, indicative of the presence of Compound I.

Because of the elusive nature of P450 Compound I experimentally, and its perceived importance in the catalytic cycle, there has been substantial interest in studying Compound I computationally. It has been found that the electronic ground state returned by quantum mechanics calculations varies slightly, depending on the model used for the cysteinate ligand (Shaik *et al.*, 2005a). In all cases the electronic configuration of the iron *d* anti-bonding orbitals is $d(x^2-y^2)^2 d(xz)^1 d(yz)^1$. There is also a third unpaired electron. Calculations modelling the cysteinate ligand with a SMe group (Antony *et al.*, 1997, Green, 1999, Rydberg *et al.*, 2004) or SCys, without internal hydrogen bonding (Ohta *et al.*, 2000) found this third unpaired electron to be in a sulphur π^* orbital. Calculations, such as those in this study, modelling the cysteinate as SH (Ogliaro *et al.*, 2000b, Ogliaro *et al.*, 2000c) or cysteinate with internal hydrogen bonding (Schöneboom *et al.*, 2002) found the third unpaired electron to be in an $a(2u)$ orbital on the porphyrin, giving a mix of spin density on the porphyrin and sulphur atoms.

A closed shell porphyrin, with spin density only on the iron and sulphur, would cause Compound I to be red in colour. In contrast, species with spin density on the porphyrin are green in colour (Shaik *et al.*, 2005a). It has been found that chloroperoxidase Compound I, which also has a cysteinate axial ligand, is green in colour (Rutter *et al.*, 1984). It would appear, therefore, that the electronic configuration given by calculations using an SH axial ligand, or by modelling the whole cysteinate ligand with hydrogen bonding, is the correct one.

The unpaired electrons on the iron and porphyrin/sulphur can have either the same, or opposite spins. This gives rise in all cases to doublet and quartet spin states with similar energies.

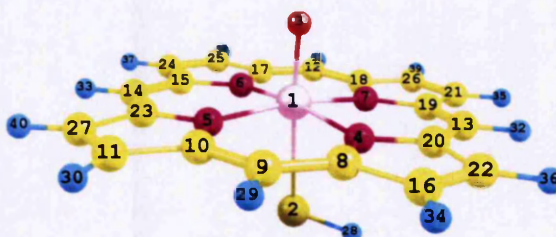


Figure 4.17: Compound I, showing the numbering used in the text.

Table 4.17: Summary of the significant bond distances, bond angles, electron spin densities and atomic charges in the doublet, quartet and sextet spin species of Compound 1. Distances are in Å and angles in degrees.

	Doublet	Quartet	Sextet
Fe-O distance	1.62	1.63	1.63
Fe-S distance	2.61	2.60	2.58
Fe-N _{av} distance	2.02	2.02	2.09
Fe deviation from plane	-0.15	-0.15	-0.22
Fe spin density	1.19	1.06	3.02
S spin density	-0.60	0.55	0.41
N ₅ spin density	-0.077	0.084	0.18
N ₆ spin density	-0.072	0.075	0.17
N ₇ spin density	-0.10	0.11	0.20
N ₈ spin density	-0.15	0.16	0.25
O spin density	0.90	0.94	0.70
Fe atomic charge	0.58	0.57	0.74
O atomic charge	-0.38	-0.37	-0.40
S atomic charge	-0.18	-0.18	-0.18
N ₄ atomic charge	-0.79	-0.79	-0.83
N ₅ atomic charge	-0.79	-0.78	-0.83
N ₆ atomic charge	-0.79	-0.79	-0.83
N ₇ atomic charge	-0.78	-0.78	-0.82

The doublet, quartet and sextet spin states of Compound 1, **7**, were optimised using B3LYP and BS1. The structures, electronic configurations and relative energies are shown in Figure 4.16, and the geometries, spin densities and atomic charges summarised in Table 4.17. The geometries, energies and spin densities of the three species are compared with results in the literature in Table 4.19. Due to the large

amount of literature data available on Compound I, only a selection of the results is presented in Table 4.19.

Table 4.18: Iron d orbital energies for the doublet, quartet and sextet spin states of Compound I, in atomic units. Orbitals in red are unoccupied.

		$d(x^2-y^2)$	$d(xz)$	$d(yz)$	$d(xy)$	$d(z^2)$
Doublet	α	-0.30067	-0.24478	-0.24546	-0.06236	-0.04786
	β	-0.27141	-0.06875	-0.06996	-0.02832	-0.02599
Quartet	α	-0.29957	-0.24817	-0.24577	-0.06133	-0.04770
	β	-0.27205	-0.06940	-0.06813	-0.02880	-0.02496
Sextet	α	-0.37336	-0.26829	-0.26801	-0.23295	-0.09936
	β	-0.09811	-0.07423	-0.07337	-0.03907	-0.02832

It was found that the doublet and quartet spin states had close to identical energies, with the sextet spin state 13.1 kcal mol⁻¹ higher in energy. This is in good agreement with the literature values, all of which recorded a negligible difference in energy between the doublet and quartet spin states.

The doublet and quartet spin states had very similar electronic configurations, shown in Figure 4.16, with two unpaired electrons on the iron, and a third unpaired electron on the porphyrin/sulphur. In the quartet spin state the third unpaired electron has the same spin as those on the iron, whereas in the doublet spin state the spins are opposite. The very close energy of the two spin states would suggest that there is little interaction between the unpaired electrons on the iron and porphyrin/sulphur. This is compared, for example, to the structures **4a** and **4b**, where the reversal of the spin of an electron on the iron $d(xz)$ orbital gave rise to a difference in energy of 1.8 kcal mol⁻¹.

The iron spin densities are slightly lower than expected as the iron binds strongly to the oxygen and the spin density is shared between these two atoms. There is also spin density on the sulphur atom, and on carbon atoms 9, 12, 13 and 14 (see Figure 4.17 for numbering), coming to totals of -1.1 and 1.2 for the doublet and quartet spin states respectively. This spin density on the sulphur and porphyrin corresponds to the unpaired electron in the $a(2u)$ orbital. This is in line with the observation of Shaik *et al.* (Shaik *et al.*, 2005a) that DFT studies modelling the cysteinate axial ligand as SH find a porphyrin/sulphur-based radical with $a(2u)$ symmetry.

The spin densities on the sulphur of -0.60 and 0.55 in the doublet and quartet spin states respectively are very similar to those found by Ogliaro *et al.* (Ogliaro *et al.*, 2000a), who used a similar model and method to that used in this study. The sulphur spin densities found by Schöneboom *et al.*, (Schöneboom *et al.*, 2002) using QM/MM were lower, with higher spin densities on the porphyrin compared to those found by the same group using QM methods. This suggests that in the protein environment the second electron is donated almost entirely from the porphyrin, rather than the sulphur. The DFT calculations modelling the cysteinate as SH give a wavefunction closer to reality than calculations using an SCH₃ group, but the calculations cannot match perfectly the electronic effects of the whole protein environment.

Table 4.19: Summary of the energy, significant geometries and spin densities of the doublet, quartet and sextet spin states Compound I, as found by various different groups using various methods. (a) This study, B3LYP, BS1, (b) Rydberg *et al.* (Rydberg *et al.*, 2004) B3LYP, Turbomole, SCH₃ (c) Green (Green, 1999), B3LYP, Gaussian 94, SCH₃, (d) Ohta *et al.* (Ohta *et al.*, 2000) B3LYP, Gaussian 98, LANL2DZ, SCys (e) Ogliaro *et al.* (Ogliaro *et al.*, 2000a) Gaussian 98, LACVP/6-31G, SH (f) Schöneboom *et al.* (Schöneboom *et al.*, 2002) B3LYP/CHARMM QM/MM, full cysteine in QM region, LACVP/6-31G, 6-31G* on O, N and S.

Spin	Group	ΔE /kcal mol ⁻¹	r(Fe-O) /Å	r(Fe-S) /Å	r(Fe-N) (av) /Å	Fe Spin	S Spin	O Spin
Doublet	a	0.0	1.62	2.61	2.02	1.19	-0.60	0.90
	b	0.21	1.63	2.61	2.02	1.23	-0.75	0.86
	c	0.0	1.65	2.69	2.02	-	-	-
	d	0.0	1.64	2.70	2.02	1.15	-0.87	0.93
	e	0.0	1.65	2.60	2.02	1.17	-0.59	0.92
	f	0.0	1.63	2.61	2.03	Fe+O 2.12	-0.36	
Quartet	a	0.04	1.63	2.60	2.02	1.06	0.55	0.94
	b	0.0	1.63	2.62	2.02	1.15	0.72	0.88
	c	-	-	-	-			
	d	0.087	1.64	2.70	2.02	1.08	0.83	0.95
	e	-	1.65	2.58	2.02	1.04	0.54	0.98
	f	0.04	1.63	2.59	2.03	Fe+O 2.02	0.31	
Sextet	a	13.12	1.63	2.58	2.08	3.02	0.41	0.70
	b	11.63	-	-	-	-	-	-
	c	-	-	-	-	-	-	-
	d	-	-	-	-	-	-	-
	e	-	-	-	-	-	-	-
	f	-	-	-	-	-	-	-

There is also spin density of 0.9 on the oxygen atom in the doublet and quartet spin states. This arises from coupling with the iron d(xz) and d(yz) orbitals, with the spin

density on the iron of 1.1 and 1.2 being lower than expected for species with two unpaired electrons on the iron d orbitals.

The sextet spin state maintains the unpaired electron on the porphyrin/sulphur $a(2u)$ orbital, with four unpaired electrons on the iron d orbitals. The strong Fe-O bond means the iron $d(z^2)$ orbital is highly disfavoured, and it is the $d(xy)$ orbital that is populated in contrast to the high spin states in the other species in the cycle. Because the high energy $d(xy)$ orbital is populated, and the iron is unable to move far out of the plane, the sextet spin state is higher in energy than the doublet and quartet spin states.

The geometries of the three spin states remain fairly constant. The Fe-O and Fe-S distances change little between the three spin states. Unlike the other species in the catalytic cycle, the Fe-O bond is stronger than the Fe-S bond, leading to the iron being above the plane of the porphyrin rather than below it. This is slightly more pronounced in the sextet spin state, with a higher iron deviation from the plane of the porphyrin, and slightly longer Fe-N distances.

4.3 Summary

In this study, the three spin states, high, intermediate and low, of each species in the generally accepted P450 reaction cycle, have been studied. Where several spin states are possible, the most plausible wavefunctions have been considered.

The energies of the high, intermediate and low spin states of each species in the reaction cycle are shown in Figure 4.18. In this figure the energies are given relative to the low spin, water-bound resting state, **1a**. The values have been normalised by adding in the values of the missing components. Ethanoic acid was chosen as the proton source. This gives approximate values for the relative energies of species **6** and **7**. A larger study, including the waters in the active site, and at least the protein residues around the active site would be necessary to obtain accurate values for the relative energies of the protonation steps.

Comparing the lowest energy spin state of each species, the removal of water from the water-bound resting state, **1**, to form **2** requires 12 kcal mol⁻¹. This value does not take

into account the additional stabilisation due to the binding of the substrate, or the solvation of the dissociated water molecule, but it does suggest that the dissociation of the bound water molecule will not be favoured without the binding of a substrate.

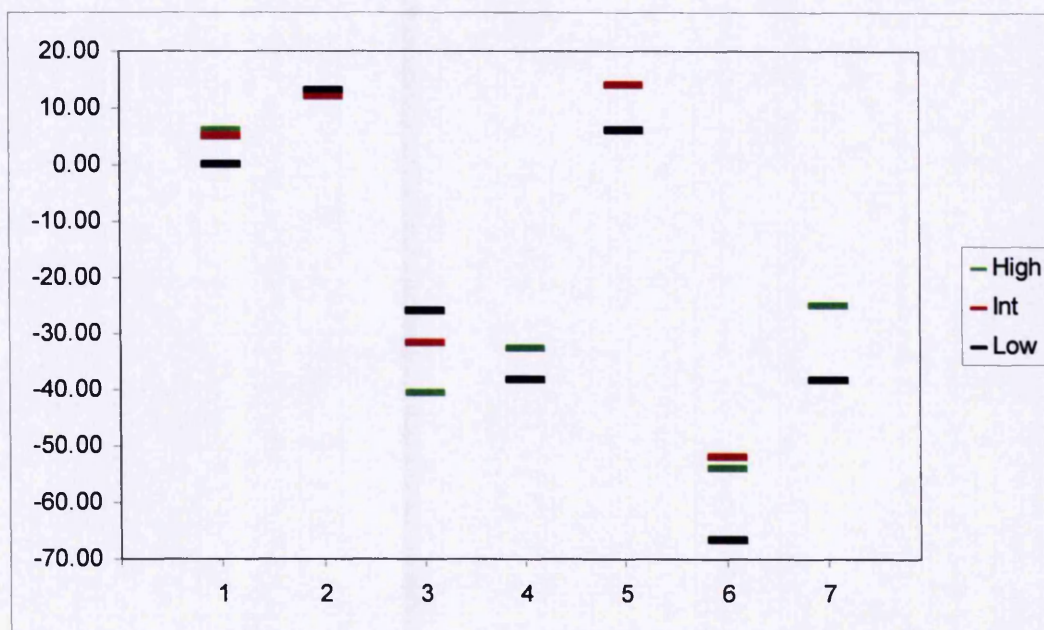


Figure 4.18: Relative energies of the low, intermediate and high spin states of each species in the P450 reaction cycle in kcal mol⁻¹ relative to the low spin, water-bound ground state 1. Energies are of the sum of components as follows: (a) 1 + O₂ + 2CH₃COOH (b) 2 + H₂O + O₂ + 2CH₃COOH (c) 3 + H₂O + O₂ + 2CH₃COOH (d) 4 + H₂O + 2CH₃COOH (e) 5 + H₂O + 2CH₃COOH (f) 6 + H₂O + CH₃COOH + CH₃COO⁻ (g) 7 + 2H₂O + 2CH₃COO⁻. Oxygen-bound species where the oxygen is dissociated are not shown. Where several wavefunctions were found with the same spin multiplicity on the iron, only the lowest energy wavefunction is shown.

The reduction of Fe^{III}(porph)SH, 2 to form Fe^{II}(porph)SH, 3 is highly favoured, by 53 kcal mol⁻¹. This is in close agreement with Shaik *et al.* (Ogliaro *et al.*, 2002b), who found a stabilisation of 51.4 kcal mol⁻¹ on reduction of 2 to form 3. Shaik *et al.* studied the 'push effect' of the thiolate ligand in P450s by DFT. They found that without the thiolate ligand, the stabilisation on reduction of 2 to 3 is much greater, with an energy of 152.3 kcal mol⁻¹. This would make the reduction of 2 much more facile, leading to non-specific reductions by any nearby reductants. Shaik *et al.* surmise that the thiolate ligand gates the reaction cycle, preventing facile, non-specific reductive processes.

The binding of **3**, which has a high spin ground state, with molecular oxygen, which has a triplet spin, to form **4**, which has an open shell singlet ground state, requires a change in spin state, called spin crossing. Jensen and Ryde (Jensen and Ryde, 2004) found a broad spin crossing region at a Fe-O distance around 2.5 Å in the binding of molecular oxygen to myoglobin. They estimated that the combined effects of the haem and axial ligand on the spin crossing ability of the complex may accelerate the binding of dioxygen to haem in myoglobin compared with iron alone by a factor of up to 10^{11} . In this study several spin states with similar structures and energies were found, suggesting that the P450 haem may have a similar effect on the probability of spin crossing.

The reduction of **4** to form **5** requires 38 kcal mol^{-1} . This is the second highest energy species in the reaction cycle, after the penta-coordinated ferric porphyrin species, **2**. It is not surprising, therefore, that the reduction of **4** is the rate determining step in the reaction cycle (Davydov *et al.*, 2001). The electron is added to the iron d(xz) orbital, forming $\text{Fe}^{\text{II}}(\text{O}_2)^{\cdot -}$.

The protonation of **5** occurs rapidly. The energy of this process has been estimated by employing ethanoic acid as a remote proton source, adding in the energy of either ethanoic acid or ethanoate as appropriate (see Figure 4.18 and Figure 4.17 and legend). This gives an approximate energy difference of 72 kcal mol^{-1} . It is not possible to calculate an accurate energy for the protonation step without taking into account the true protonation source. This would vary for each individual P450 and is outside the scope of this study.

The final step in the reaction cycle, before turnover of the substrate and production of the product, is the protonation of **6** to form Compound I, **7**. An approximate energy increase for this step of $28.3 \text{ kcal mol}^{-1}$ was calculated, using ethanoic acid as a protonation source once more. Various QM/MM studies on the formation of Compound I, most recently by Zheng *et al.* (Zheng *et al.*, 2006), found that the relative energies of **6** and **7** depend on the method used, ranging from Compound I being less stable by 10 kcal mol^{-1} to it being more stable by 80 kcal mol^{-1} . As Compound I is extremely elusive, and has not been isolated in P450s for more than a few milliseconds, it is clear that it reacts much faster than it is produced. This would suggest that an energy 28 kcal mol^{-1} higher than that of **6** is reasonable.

The subsequent steps in the P450 catalytic cycle are specific to the individual P450s and substrates, and are beyond the scope of this study.

4.4 Conclusion

P450s are ubiquitous enzymes, found in every domain of life and in the vast majority of living organisms. P450s catalyse a wide range of reactions, most of which are oxidations, including hydroxylations, epoxidations, oxidation of heteroatoms and dealkylations. P450s have a variety of structures and substrate specificities, ranging from the more substrate specific P450cam, and CYP4 fatty acid hydroxylases, to the rather unspecific drug metabolism enzymes such as CYP3A4.

In order to understand the reactivity of P450s it is important to understand the core of the enzyme, the haem active site. This is conserved in all P450s. It is important to understand how the electronic structure of the haem changes as the enzyme turns over substrates, independently of the protein environment.

DFT can provide accurate information on the electronic structure of systems such as haem. In this study, DFT has been used to model the haem as it goes through the P450 reaction cycle. All of the most likely spin states of each species in the reaction cycle have been studied, in order to obtain a survey of how the structure and electronic configuration of the haem changes as it turns over substrates. This was done independently of individual protein environments or substrates, in order to obtain generic results that can be applied to all P450s.

Towards the end of the study being carried out, similar surveys of the whole reaction cycle were published by other groups. The results of this study have been compared with the results of these groups.

The study began with the doublet, quartet and sextet spin states of the water-bound resting state, **1**. A low spin ground state was found. This was in disagreement with some groups, who said that the protein environment was necessary to obtain a doublet ground state (Harris and Loew, 1993). Other groups (Filatov *et al.*, 1999) had reported a doublet

spin structure in which the water had an 'upright' structure, rather than the 'tilted' structure returned by unrestrained DFT calculations. This was found to be a local minimum, which could not be converged upon with the methods used in this study. Comparison with QM/MM studies (Schöneboom and Thiel, 2004) and ESEEM data (Goldfarb *et al.*, 1996) showed the true state to be intermediate between the 'upright' and 'tilted' structures. It is clear that the natural binding mode when binding to haem is a tilted structure, which tends to be more upright in the protein environment.

The Fe-O distance was found to increase with increase in spin state, with the water effectively dissociated in the sextet spin state. This is due to the population of the iron $d(z^2)$ anti-bonding orbital clashing with the water $p(z)$ orbitals.

The penta-coordinated ferric species, **2**, was found to have very similar energies for all three of the spin states. It was found that the relative energies of the spin states are sensitive to the method used in the calculations, with all electron basis sets for the iron giving results that were most in line with the experimental observation of a high spin ground state.

The penta-coordinated ferrous species, **3**, had a quintet spin ground state, in line with experimental observations. This is due to the lack of a second axial ligand releasing the constraints on the $d(z^2)$ orbital, lowering it in energy, and the ability of the iron to move out of the plane of the porphyrin causing the iron $d(xy)$ orbital to also be stabilised. The energy of **3** was found to be 53 kcal mol⁻¹ lower than that of **2**. This is considerably less than the difference in energy would be if the sulphur atom was not present. The cysteine ligand in P450s is thought to gate the catalytic cycle, preventing unspecific reduction by less powerful reductants.

Eight possible spin states were calculated for the oxygen-bound ferrous porphyrin species, **4**. The lowest energy structure was found to be heptet spin structure, **4f**, in which the dioxygen was dissociated from the porphyrin. The lowest energy structure in which the dioxygen remains bound to the porphyrin was the open shell singlet, **4a**. It was found that **4** is in equilibrium between Fe^{II}(O₂) and Fe^{III}(O₂)⁻. The quintet ground state of **3** and triplet spin molecular oxygen do not combine to form open shell singlet **4a**. There are a number of spin states of **4** with similar energies. This suggests that spin crossing

occurs as molecular oxygen approaches **3**, allowing oxygen to bind much faster than to iron alone.

It was found that 38 kcal mol^{-1} were required for the reduction of **4** to form **5**. The electron is added to the iron $d(xz)$ orbital. There are also low lying spin states in which an electron has been transferred from the iron $d(yz)$ orbital to the porphyrin. One such species has a dissociated oxygen molecule, and is $6.5 \text{ kcal mol}^{-1}$ more stable than the low spin, oxygen-bound, ground state. This is the product of reduction of the dissociated heptet species **4f**. The transfer from the low spin species **4a** to the high spin, dissociated species **5c** is spin forbidden, so it is likely to be too slow to be significant in this transient reactive intermediate.

The hydroperoxo ferric porphyrin species **6**, known as Compound 0, was found to have a doublet ground state. Hydrogen bonding between the hydroperoxo group and porphyrin nitrogen atoms prevents it from dissociating in the higher spin states. Some have suggested that this may be the true oxidant in the reaction cycle. This hydrogen bonded, relatively stable species seems an unlikely candidate compared to the electronically constrained, highly reactive Compound I.

Compound I was found to have two spin states, the doublet and quartet states, with almost identical energies. Shaik *et al.* (Shaik *et al.*, 2002) have suggested these two spin states give rise to reaction pathways in the subsequent steps and different stereoselectivities. With two very similar spin states so close in energy it seems likely that both would be involved in catalysis, and feasible that they react on slightly different timescales.

There is ongoing debate over the true nature of the oxidant species in the P450 catalytic cycle. In addition to Compound I, which is widely accepted to be the oxidant species, the hydroperoxo, Compound 0, and derivatives of Compound I have been put forward as possible candidates. Although other derivatives of Compound I have not been studied in this study, Compound 0 was found to be the most stable species in the catalytic cycle. The protonation sources used in this study were not sufficiently accurate to enable quantitative predictions of the reactivity of Compound 0, but it would appear to be too stable to carry out the extremely powerful oxidation reactions characteristic of P450s.

This study has provided a number of useful insights into the P450 reaction cycle. The most likely structure for drugs targeting P450s to bind to is the water-bound resting state. This has been found to be low spin, independent of the protein environment. The ground state of the substrate bound structure has been found to depend on the protein environment, with similar energies for the three spin states without protein present.

The binding of oxygen to the ferrous haem is energetically unfavourable, with the lowest energy species found to have Fe-O distance 3.65 Å. As the rate of reaction from this point is reportedly extremely fast, this does not present a problem to the function of the catalytic cycle. Also, the protein environment would be expected to stabilise the binding of oxygen. The unfavourable nature of oxygen binding to haem does illustrate, however, how a competitive inhibitor, such as the azole-antifungal drugs studied in Chapter 3, and investigated further in Chapter 5, would prevent the binding of oxygen to the haem, and thus prevent the catalytic function of the enzyme.

Of the species studied, Compound I was confirmed to be the most likely candidate as the oxidant species in the P450 catalytic cycle. Two spin states with identical energy were found, showing how Compound I may carry out reactions with a range of energy barriers and stereoselectivities. This should be taken into account when designing any treatments targeting P450s, as rates of reaction may be affected by the variation in spin states possible for Compound I.

4.4.1 Future Work

This study has provided a brief survey of the most likely spin states of each structure in the P450 catalytic cycle, without the involvement of any specific P450's protein structure, or substrates. These data may now be used to refine other models, such as for docking of drug structures into specific P450s.

During the course of the study, further questions and areas that remain unclear have been raised. The origin of the high spin ground state in **3**, and the reason it only appears to be obtained when higher basis sets are used, should be addressed in future work. Also the spin crossing mechanism by which triplet spin molecular oxygen is able to bind

with the quintet spin penta-coordinated haem species **3**, to form the open shell singlet **4a**, needs to be studied further.

There is still debate over the protonation source in the two protonation steps. This is likely to vary between individual P450s. The blocking of the protonation pathways could be an effective method of inhibiting the action of P450s, for instance in the treatment of *Mtb*. Also, the true relative energies of Compound I and Compound 0 are still unclear.

Much work has been done on Compound I in P450s, yet there is still debate over the true identity of the oxidant species in the P450 catalytic cycle, and experimental evidence of Compound I in P450s remains sparse. The majority of the theoretical work on Compound I has been based on hydroxylation reactions in P450cam. Further study is needed on the mechanisms of the reactions of other P450s, to ascertain how consistent the catalytic cycle determined for P450cam is across the range of P450s in nature.

CHAPTER 5: ANALYSIS OF AZOLE BINDING TO P450S BY DFT METHODS

5.1 Introduction

As was stated in Chapter 1, P450s are a ubiquitous family of enzymes found in a very broad range of organisms, from bacteria and fungi to plants and mammals (Denisov *et al.*, 2005). P450s catalyse a broad array of reactions, the most common of which are oxidations and hydroxylations. They can also catalyse other reactions, including reduction, desaturation, ester cleavage, ring expansion, dehydration and one-electron oxidation (Guengerich, 2001, Ortiz de Montellano and de Voss, 2005). The generally accepted reaction cycle for the majority of P450 reactions was studied in depth in Chapter 4.

There is a high degree of variability in the structures of P450s (Verras and Ortiz de Montellano, 2006), ranging from those that have fairly substrate specific active sites, such as P450cam binding camphor (Keller *et al.*, 1972, Poulos *et al.*, 1987), to those that have broad specificities and relatively undefined substrates, such as CYP3A4 (Ekroos and Sjogren, 2006) and CYP2C9 (Guengerich, 2005, Williams *et al.*, 2003), which are responsible for the bulk of human xenobiotic metabolism.

P450s in nature have a large number of functions. One role is protection, in the metabolism of xenobiotics. The major example of this in humans is CYP3A4 (Guengerich, 2005, Guengrich, 1999), which is the most abundant P450 in the human body, and contributes to the metabolism of around half the drugs in use today. This is also likely to be the role of some of the P450s in *Mtb*, for example it has been suggested that CYP51 is involved in the inactivation of host sterols and steroids in order to prolong the infective state (Gamieldien *et al.*, 2002, McLean *et al.*, 2006a).

Another role of P450s is in the biosynthesis of important biomolecules, for use for example as signal molecules used for control of development and homeostasis. Examples of this in humans include CYP1B1 and 7A1, which are involved in the

synthesis of sterols (Guengerich, 2005). This may be the role of CYP132 in *Mtb*, which shows close homology to CYP4 fatty acid hydroxylases and may be involved in the synthesis of the lipid rich cell wall (Recchi *et al.*, 2003).

P450s play important roles in many pathogens and pests, and are possible drug targets for future treatments. For example *Mycobacterium tuberculosis* contains 20 P450 enzymes, some of which have been shown to be important for the development or virulence of the bacterium (McLean *et al.*, 2006a), such as CYP132 (Recchi *et al.*, 2003), which was studied in Chapter 3. P450s are also important drug targets in fungal infections (Odds *et al.*, 2003), they are involved in insecticide resistance in the mosquito (Feyereisen, 2006), and there is one remaining P450 in the genome of *Mycobacterium leprae*, the bacterium responsible for the disease leprosy. P450s are also drug targets in some cancer treatments. For instance CYP17 and CYP19 are important in the synthesis of androgens and oestrogen and are useful targets in the treatment of hormone-dependent breast and prostate cancers (Baston and Leroux, 2007).

As well as being important drug targets, there are 57 P450s in the human genome that play vital roles in biosynthesis and metabolism, and which must be prevented from being attacked by any treatments aimed at other P450s. A detailed review of human P450s can be found in the book *Cytochrome P450: Structure, Mechanism, and Biochemistry*, 3rd edition (Guengerich, 2005). The functions of P450s in humans include the transformation of xenobiotic chemicals, steroidogenesis, metabolism of vitamins, and fatty acid metabolism (Guengerich, 2005).

If the drug metabolising P450s in the liver are inadvertently inhibited by treatments for other conditions, this can alter the biological activity of the drug, leading to the potential for harmful drug-drug interactions (Hollenberg, 2002). If two different drugs compete for metabolism by the same P450, this can lead to unexpected elevations in the plasma concentrations of one or both drugs. This can ultimately result in a variety of minor as well as serious adverse effects, even fatalities. This has led to several prominent drugs being withdrawn from the market, including clozapine and fluoxetine (Ferslew *et al.*, 1998) and triazolam and amitriptyline (Kudo *et al.*, 1997). Limited utility, or even withdrawal from the market, of a given drug due to significant drug interactions may result in considerable economic losses for the pharmaceutical companies. This has

caused the pharmaceutical companies to develop and utilise many new approaches that can be used to predict the possibility that new drug candidates will cause significant drug interactions with the human P450s (Hollenberg, 2002). This will be especially true for any new drugs for which the drug target is itself a P450.

5.1.1 P450 inhibitors

P450 inhibitors can be divided into three mechanistically distinct classes (Correia and Ortiz de Montellano, 2005, Lin and Lu, 1998). These are those that bind reversibly, those that form quasi-irreversible complexes with the haem iron atom and those that bind irreversibly to the protein or the haem moiety. Reversible binders generally interfere in the catalytic cycle prior to the actual oxidative event (Hollenberg, 2002). Irreversible and quasi irreversible binders usually act during or subsequent to the oxygen transfer step. These are often mechanism based inactivators, relying on the P450 to transform them into species which denature or form irreversible complexes with the P450. There are other, less common methods of inhibition, such as agents that interfere with the electron supply to the haemoprotein by accepting electrons directly from P450 reductase. As these agents do not interact directly with the P450 they are generally non-specific with respect to the P450s they inhibit (Hollenberg, 2002).

5.1.1.1 Irreversible and quasi irreversible inhibitors

There are multiple known classes of compounds that are activated by P450s to form reactive intermediates that irreversibly or quasi-irreversibly inactivate the P450 responsible for their activation (Correia and Ortiz de Montellano, 2005). This catalysis-dependent inactivation can be highly enzyme-specific because the compound must both bind to the active site of the P450, and undergo catalytic activation to a product that irreversibly alters the enzyme and permanently removes it from the catalytic pool. Chapter 7 of *Cytochrome P450: Structure, Mechanism, and Biochemistry, 3rd edition* (Correia and Ortiz de Montellano, 2005) discusses catalysis-dependent inactivation at length and these will be summarised briefly here.

There are four general classes of catalysis-dependent P450 inactivators (Correia and Ortiz de Montellano, 2005). Agents that bind covalently to the protein, agents that bind quasi-irreversibly to the haem iron atom, agents that alkylate or arylate the porphyrin

framework of the haem and agents that degrade the prosthetic haem to products that can themselves modify the protein.

Agents whose activated products bind covalently to the protein include various sulphur compounds (e.g. carbon disulphide (Dalvi, 1987) and isothiocyanates (Elhawari and Plaa, 1979)), halogenated species (e.g. chloramphenicol (Halpert and Neal, 1980)), alkyl and aryl olefins and acetylenes (e.g. 1-ethynylpyrene and phenylacetylene (Chan *et al.*, 1993)) and furanocoumarins (e.g. 8-methoxypsoralen (Letteron *et al.*, 1986)).

Agents that form products that coordinate to the haem quasi-irreversibly include dioxymethylene compounds, such as the drug paroxetine (Bloomer *et al.*, 1992) and nitrogen compounds, such as amines that are converted to nitroso metabolites (e.g. erythromycin (Larrey *et al.*, 1983)).

There are a number of ways that species can form products that covalently bind to the P450 haem. Species with terminal monosubstituted double bonds or terminal acetylenes at the site of catalytic oxidation can alkylate the nitrogen atoms of the haem, inactivating the enzyme (e.g. 4-ethyl-1-hexene and 1-heptene (Ortiz de Montellano and Mico, 1980) and phenylacetylene (Ortiz de Montellano and Komives, 1985). Substituted dihydropyridines and dihydroquinolines can form alkyl radicals which can also bind to nitrogen atoms on the haem or protein residues (e.g. 3,5-bis(carbethoxy)-2,4,6-trimethyl-1,4-dihydropyridine (Gayarthri and Padmanaban, 1974) and 2,4-diethyl-1,2-dihydro-2-methylquinoline (Lukton *et al.*, 1988)).

Some P450s are inactivated by certain substrates, which cause fragments of the haem to bind irreversibly to the protein. This is mainly seen in P450s with large active sites, such as CYP3A4 and CYP2C6. Examples of inactivators of this type include carbon tetrachloride, spironolactone and 3,5-dicarbethoxy-2,6-dimethyl-4-ethyl-1,4-dihydropyridine. Studies are ongoing into the features that predispose an enzyme to cross-linking of haem fragments to the protein (Masubuchi and Horie, 2007).

5.1.1.2 Reversible inhibitors

The remainder of this chapter will study reversible inhibitors and how they bind to the P450 haem region. In order to be activated by the P450 and form species that bind to

the protein or haem, irreversible inhibitors must first bind reversibly to the haem region, so the work discussed in this chapter will also be relevant to some irreversible inhibitors.

Reversible inhibitors inhibit the action of the enzyme by competing with substrates for occupancy of the active site (Correia and Ortiz de Montellano, 2005). They differ in where they bind to the active site and include agents that bind to hydrophobic regions of the active site, coordinate to the haem iron atom, or enter into specific hydrogen bonding or ionic interactions with active-site residues.

The competition for binding to lipophilic domains of the active site is often responsible for the inhibition observed when two substrates compete for oxidation by a single P450 isoform, such as the mutual inhibition of benzene and toluene metabolism (Sato and Nakajima, 1979). This form of inhibition is not usually highly effective but can give physiologically relevant metabolic changes and clinically significant interactions (Correia and Ortiz de Montellano, 2005).

Cyanide (Kitada *et al.*, 1977) and nitric oxide (Wink *et al.*, 1993) can bind tightly to ferric haem, and carbon monoxide binds tightly to ferrous haem (Hanson *et al.*, 1976, McLean *et al.*, 2002a), giving the characteristic peak in the UV/Vis spectrum at 450 nm. Use of these highly toxic small molecules for P450 directed treatments would be a little drastic however and larger, more selective agents must be used for *in vivo* treatments of P450-dependent conditions.

Reversible inhibitors that bind to both the lipophilic domains of the active site and the haem iron are much more effective P450 inhibitors than agents that only employ one of these methods (Correia and Ortiz de Montellano, 2005). How effective inhibitors of this type are depends on the strength of the bond between their heteroatomic lone pair and the haem iron and how this is moderated by steric interactions with substituents on the inhibitor, the interactions with the lipophilic domains and how well the geometry of the agent fits into the enzyme active site.

Structures which contain an oxygen atom that coordinates the iron, such as alcohols, ethers, ketones and lactones, bind relatively poorly, and are mostly weak inhibitors (Schenkman *et al.*, 1981). In contrast, structures containing aliphatic or aromatic

nitrogen atoms can bind strongly to the haem iron and are often strong inhibitors (Murray and Reidy, 1990). Pyridine, imidazole and triazole derivatives have proven particularly useful as P450 inhibitors (Saber *et al.*, 2006, Testa and Jenner, 1981).

5.1.1.3 Azole drugs

There are a number of reversible inhibitors that target P450s which contain azole functional groups as the haem iron binding component (although not strictly accurate, for simplicity pyridine will be referred to as an azole in this chapter). Many of these are licensed as antifungal drugs, although they do have other uses (Odds *et al.*, 2003). One of the first P450 inhibitors to be widely employed was metyrapone, which targeted the 11 β -hydroxylase P450 (CYP11B1) in the treatment of Cushing's syndrome (Temple and Liddle, 1970) and has a pyridine functional group. The binding of several azole antifungal drugs to *Mtb* CYP132 was reported in Chapter 3 of this work. Several other *Mtb* P450s have also been shown to bind azole antifungals, including CYP51B1, CYP121 and CYP140 (Guardiola-Diaz *et al.*, 2001, McLean *et al.*, 2002b). Of the azoles studied in Chapter 3, only fluconazole is generally used systemically due to side effects from interactions with human P450s leading to drug-drug interactions (Zhang, W. *et al.*, 2002), although systemic use of ketoconazole has also been reported (Vantwout *et al.*, 1988, Yazdanpanah *et al.*, 2007), clotrimazole, econazole and miconazole are used to treat fungal infections extraneously (Kyle and Dahl, 2004).

In the last few years three new azole antifungal drugs, all with triazole functional groups, have been licensed for clinical use (Odds *et al.*, 2003). These were developed to tackle oropharyngeal *Candida* infections in AIDS patients and invasive aspergillosis due to *Aspergillus fumigatus*, but they have extremely broad antifungal specificities. Of these new drugs, voriconazole and ravuconazole are structurally related to fluconazole, whereas posaconazole bears a close resemblance to the previously used itraconazole (see Figure 5.1 for structures).

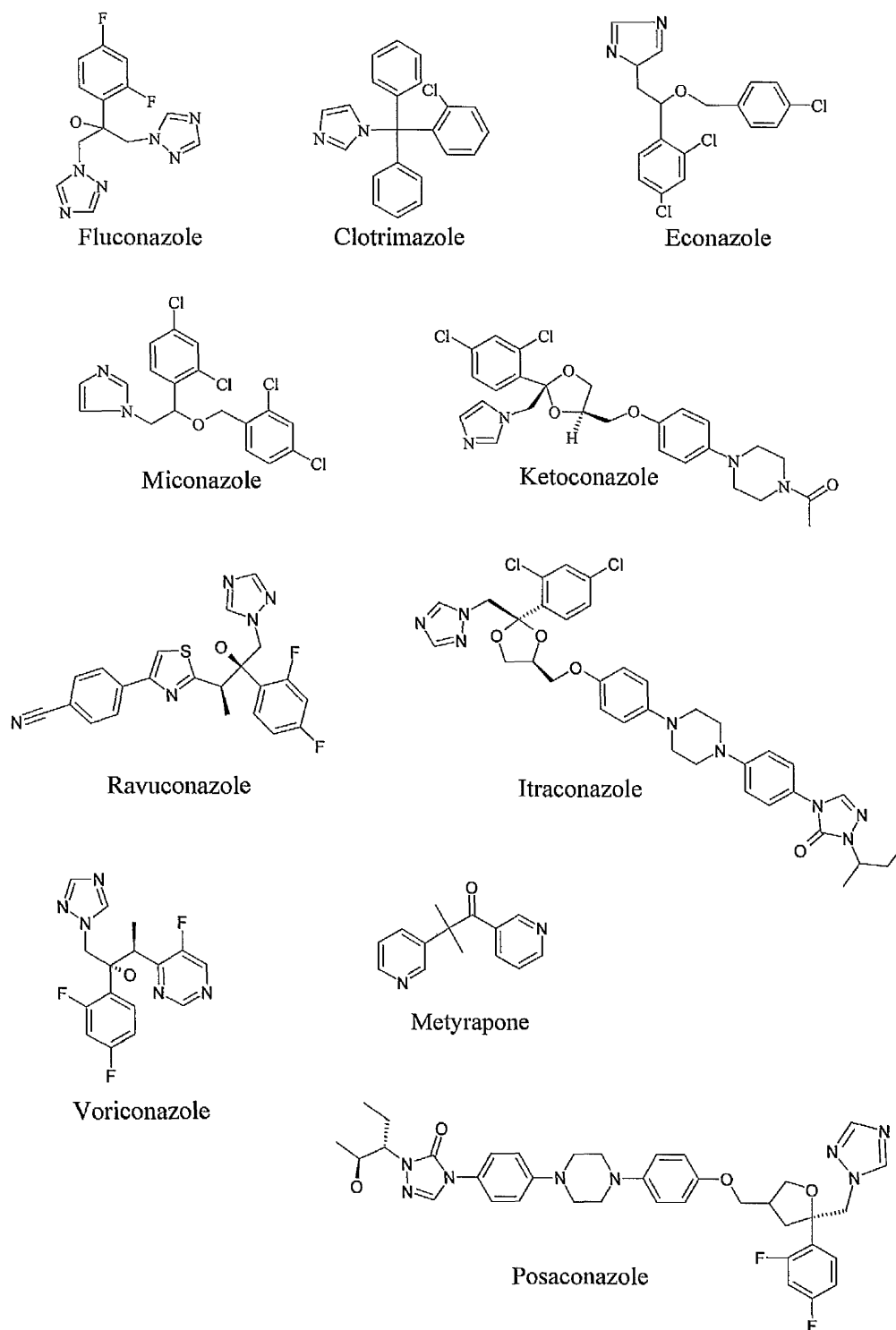


Figure 5.1: Structures of various azole antifungal drugs currently in clinical use. Inactivation of P450 targets occurs through haem iron coordination by one or more of the imidazole/triazole functional groups on the molecules.

The focus throughout this study has been on the *Mtb* P450 enzymes as possible new drug targets in the fight against TB. *Mtb* has 20 P450 enzymes, an unprecedented number of P450s in a non-mycobacterial bacterium discovered to date (Souter *et al.*, 2000). Several azole antifungal drugs have been shown to inhibit the growth of *Mycobacterium smegmatis* and *Mycobacterium bovis*, which are closely related to *Mtb* (Guardiola-Diaz *et al.*, 2001, McLean *et al.*, 2002b) and thus azole antifungals have been suggested to be good starting points for the development of new treatments for *Mtb*. Also, the presence of multiple targets in *Mtb* for the azoles could reduce the likelihood of strains becoming resistant to a new azole based *Mtb* therapy.

5.1.1.4 Aim

As was stated above, and in the previous chapters, there has been a great increase in the incidence of tuberculosis in recent years, such that *Mycobacterium tuberculosis* (*Mtb*) now causes more deaths than any other infectious agent (WHO, 2002). With the development of multidrug-resistant strains, there is an urgent need for new treatments to tackle TB. Azole antifungals have been shown to inhibit the growth of *Mycobacterium smegmatis* and *Mycobacterium bovis*, which are closely related to *Mycobacterium tuberculosis*, and thus these seem a sensible starting point for the development for new treatments for *Mtb*.

There are 20 P450s in the *Mtb* genome, several of which have been suggested to be potential drug targets for the treatment of TB (McLean *et al.*, 2002b, Munro *et al.*, 2003, Recchi *et al.*, 2003). Each P450 has a different structure, and a different active site. How the various azole drugs bind with the P450 active sites will differ, depending on the interactions between the drug structure and the residues in the active site. In contrast, in all of the different P450s the structure of the haem remains the same. Similarly there are currently only three azole functional groups in use in azole antifungals, these are imidazole, triazole and pyridine (for simplicity pyridine will be referred to as an azole in this chapter). In the development of new azoles to bind to the *Mtb* P450s, it is important to better understand the energetic and electronic driving force behind the binding of the azole functional group and the P450 haem.

Density functional theory (DFT) can provide an invaluable insight into the electronic and energetic driving forces behind fairly small systems such as haem (Harris and Loew, 1998, Shaik *et al.*, 2005a). In this chapter, the binding of imidazole, triazole and pyridine to the P450 haem is investigated using DFT methods in order to gain a better understanding of the driving force behind azole binding to P450s. The pseudo reaction coordinate method is used (Young, 2001), fixing the Fe-N_{azole} distance in order to elucidate the reaction pathway by which azoles dissociate the bound water from haem. The energetic driving force behind the process is also investigated.

5.1.1.5 Previous studies on azoles binding to P450s and haems

There has been fairly extensive work done on the binding of azole drugs to P450s. As early as 1967 Estabrook *et al.* (Schenkman *et al.*, 1967) reported type II spectral shifts from the binding of nitrogen-containing compounds, such as pyridine and nicotine, to human liver P450s. Vanden Bossche (Vanden Bossche, 1985) postulated that the target of the azole antifungal drugs was the P450 14 α -demethylase, later to be called CYP51. Schuster (Schuster, 1985) studied the effect of azoles on human P450s, concluding that the metabolism of xenobiotic substrates would be impeded by azole drugs. Since then there have been innumerable reports of azole antifungals binding to P450s, including many studies on drug-drug interactions involving binding to CYP3A4 (Murayama *et al.*, 2007, Park *et al.*, 2006, Varhe *et al.*, 1994), and a number of studies of the binding of azole antifungals to *Mtb* P450s (Banfi *et al.*, 2006, Guardiola-Diaz *et al.*, 2001, McLean *et al.*, 2002b).

The first crystal structures of azole bound P450s were fluconazole and pheynylimidazole bound *Mtb* CYP51 (Podust *et al.*, 2001) and show the nitrogen atom of either fluconazole or phenylimidazole coordinated to the haem iron. Since then structures have been solved of ketoconazole bound to P450eryF (Cupp-Vickery *et al.*, 2001), P450cam bound to imidazole (Verras *et al.*, 2006) and CYP3A4 bound to ketoconazole (Ekroos and Sjogren, 2006). Seward *et al.* also solved the structure of CYP121 bound to fluconazole (Seward *et al.*, 2006). They found a novel binding mode, involving fluconazole binding via a bridging water molecule to the haem iron. This will be discussed further in the Results and Discussion section.

There have been a number of studies on imidazole binding to iron-porphyrin in order to model haemoglobin (Lamar and Walker, 1972, Nappa *et al.*, 1977, Pasternack *et al.*, 1978, Uno *et al.*, 1991, Walker *et al.*, 1976). Useful information can be gained from these studies, such as that imidazole is a strong σ and π donor, (Johnson *et al.*, 1991) and pyridine is a better π acceptor than imidazole (Johnson and Shepherd, 1983). These studies do not relate directly to the binding of imidazole to P450 haem, however.

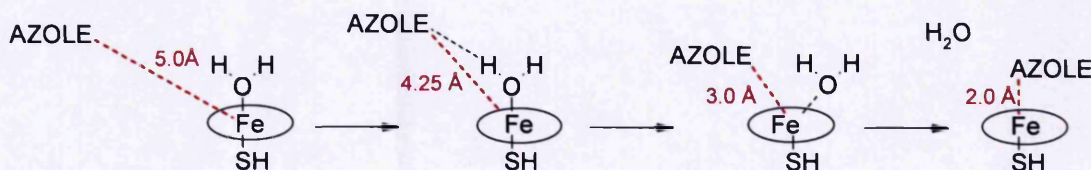
There is little available literature on DFT studies of the binding of azoles to P450 haem. Harris *et al.* (Harris *et al.*, 2004) carried out DFT studies on the binding of benzphetamine, a tertiary amine, its metabolite desmethylbenzphetamine (DMBP), a secondary amine, 4-phenylimidazole and water to iron-porphyrin-methyl mercaptide using the B3LYP functional and the Jaguar program. They found haem-ligand interaction energies of -15, -14.4 and -18.6 kcal mol⁻¹ for water, DMBP and phenylimidazole, respectively. Imidazole is often used in DFT studies to model histidine-bound haems, such as haemoglobin (Du and Loew, 1992, Rydberg *et al.*, 2004), but there have been few DFT studies on azoles binding to haem-thiolates.

In this chapter the binding of the azoles imidazole, triazole and pyridine to the P450 haem is investigated using DFT methods, in order to determine the binding mode and driving force behind azole binding to P450s. Calculations carried out based on the fluconazole-bound CYP121 crystal structure (Seward *et al.*, 2006) are also described, in order to shed light on novel binding modes observed in the structure.

5.1.1.6 The model used

The models used in the calculations of the potential energy surfaces (PES) for the displacement of water by azole are illustrated in Figure 5.2. In the calculation of the 1D PES the distance from the haem iron to the incoming azole nitrogen atom was fixed at progressively smaller distances, each time minimising the rest of the system and using the geometry and wavefunction from the previous calculation as a starting point for the next. The 1D PES was not found to give a smooth path from the water-bound structure to the azole-bound structure. For this reason the 2D PES was calculated. This used a similar method to the calculation of the 1D PES, this time fixing both the Fe-N_{azole} and Fe-O_{water} distances over a range of lengths.

1D Reaction Surface



2D Reaction Surface

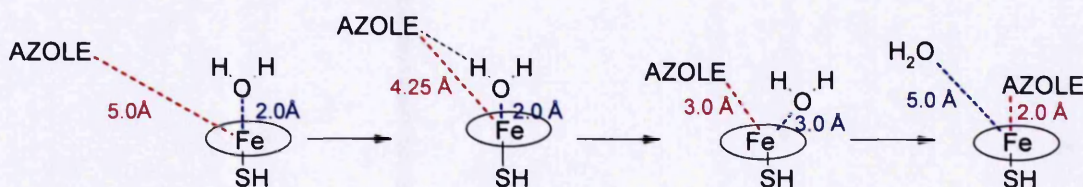


Figure 5.2: The models used in the calculation of the 1D and 2D potential energy surfaces (PES). In the 1D PES the Fe-N_{azole} distance was fixed over a range of values and the rest of the system allowed to optimise freely. In the 2D PES the Fe-N_{azole} and Fe-O_{water} distances were fixed over a range of values. The fixed distances for some of the minima and transition states on the PESs are shown above. The porphyrin is represented by an ellipse.

5.2 Results and Discussion

5.2.1 1D binding curves

Reaction pathways were calculated for the displacement of water from low spin iron porphyrin by each of the three azoles, imidazole, triazole and pyridine. For the purpose of simplicity, iron-porphyrin will be referred to as haem in this chapter. Basis set BS1 and both the B3LYP and BP86 functionals were used. The optimised low spin water bound porphyrin species from Chapter 4 was used as a starting point, with the azoles added in manually, the Fe-N_{azole} distance being fixed at 5 Å. The reaction path was followed along the Fe-N_{azole} pseudo reaction coordinate, the Fe-N_{azole} distance being decreased in 0.5 Å increments, and the structures optimised at each point (Young, 2001). Two minima were found for each azole, one at Fe-N_{azole} \approx 4.25 Å and one at \approx 2 Å. The distance constraint was removed from the structures closest to these minima and the structures minimised to give the true minima.

It was observed that the energy fell sharply after the $\text{Fe-N}_{\text{azole}} = 3.0 \text{ \AA}$ point. Searching around this point in smaller increments elucidated two reaction pathways, labelled A and B in Figure 5.3 and Figure 5.4. Path A begins with water bound to the haem, with the azole binding first to the bound water, forming a bridged complex, and then binding to the iron, forming a hepta-coordinated iron complex. As the $\text{Fe-N}_{\text{azole}}$ distance was decreased in small increments from 3.0 \AA the species fell down to pathway B. This starts with azole bound to the iron, with the water binding to the azole. As the $\text{Fe-N}_{\text{azole}}$ distance increases the water also moves away from the haem, remaining hydrogen bonded to the azole.

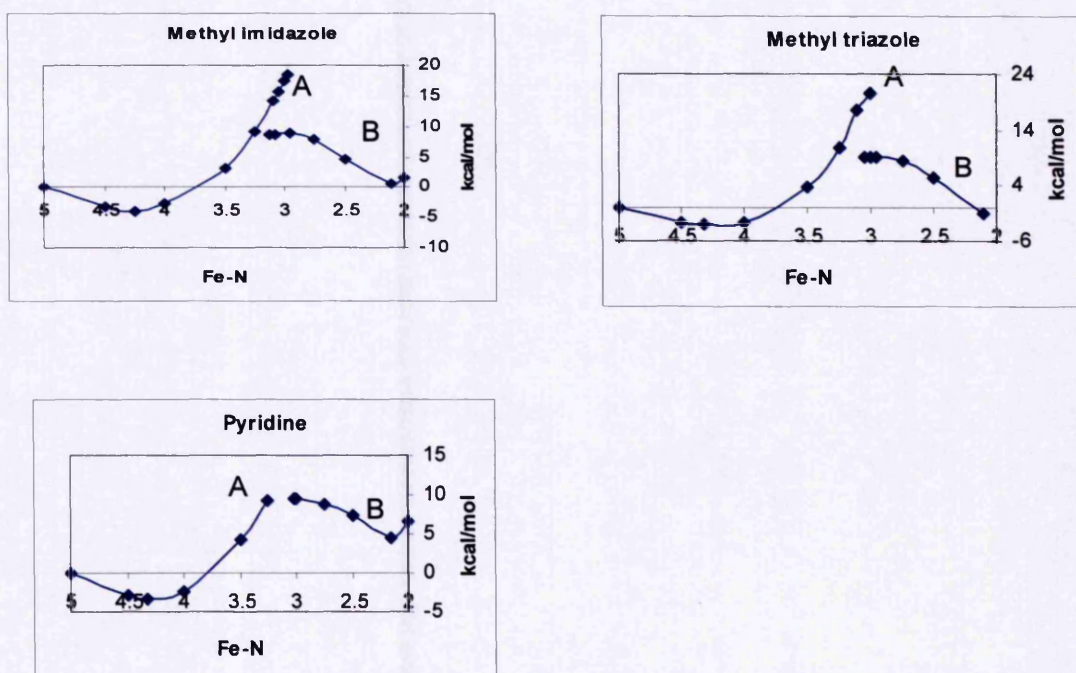


Figure 5.3: Graphs of energy in kcal mol^{-1} against $\text{Fe-N}_{\text{azole}}$ distance in \AA for the binding of methyl imidazole, methyl triazole and pyridine to haem using B3LYP and BS1. Two reaction paths were observed for each azole, labelled A and B

The relative energies and $\text{Fe-N}_{\text{azole}}$ and $\text{Fe-O}_{\text{water}}$ distances are shown in Table 5.1 and Table 5.2 for the results using the B3LYP and BP86 functionals, respectively. The S^2 values using the B3LYP functional, shown in Table 5.1, rise to 0.96 at $\text{Fe-N}_{\text{azole}}$ distances around 3.0 \AA for the three azoles. This is compared to expected values of 0.75. Using the BP86 functional the S^2 values do not go above 0.82. These higher levels of spin contamination with B3LYP suggest that there is some intermediate spin or high spin

character in the calculations, probably due to the stabilisation of higher spin states associated with the B3LYP functional. There is little difference in the energies and geometries between the results obtained with the different functionals, suggesting that the extra spin contamination with B3LYP does not have a significant effect.

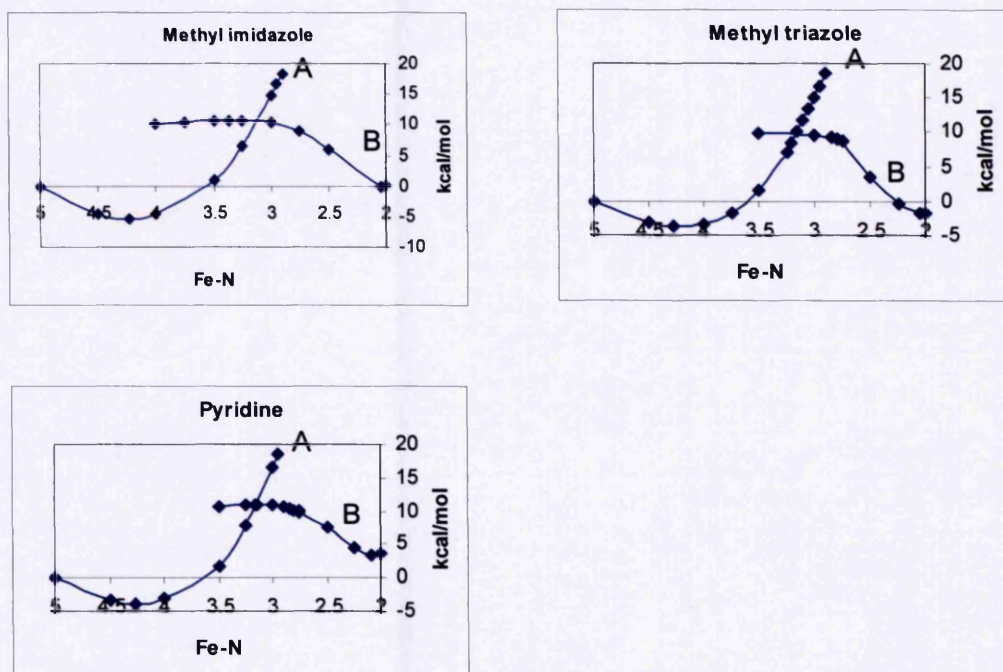


Figure 5.4: Graphs of energy in kcal mol^{-1} against $\text{Fe-N}_{\text{azole}}$ distance in Å for the binding of methyl imidazole, methyl triazole and pyridine to haem using BP86 and BS1. Two reaction paths were observed for each azole, labelled A and B

The binding curves for imidazole displacing water bound to haem, shown in Figure 5.3 and Figure 5.4, and Table 5.1 and Table 5.2, are very similar using the B3LYP and BP86 functionals. Both have two minima, the first with the water forming a bridge between the imidazole and iron, and the second with the imidazole bound directly to the iron. The bridged-binding mode is 4.6 and 5.2 kcal mol^{-1} more stable than the direct binding mode using the B3LYP and BP86 functionals respectively, with barrier heights of 13.0 and 15.5 kcal mol^{-1} .

It was more difficult to converge on high energy pathways using B3LYP than with BP86, hence the two curves do not carry on as far with the former, as the structures tended to fall down to the lower energy pathway. Although the two curves have equal energy at

Fe-N_{azole} ~ 3.25 Å, the structures on the two curves around this point are quite different, with Fe-O_{water} distances of 2.1 and 4.1 Å on paths A and B, respectively.

Table 5.1: Fe-N_{azole}, Fe-O_{water} distances, relative energies and spin contamination values for the reaction pathways in the displacement of water by (a) methyl imidazole, (b) methyl triazole, and (c) pyridine using BS1 and the B3LYP functional. The Fe-N_{azole} distances were fixed and all other bonds allowed to minimise freely.

Fe-N /Å imid	Fe-O /Å	ΔE/ kcal mol ⁻¹	S ² (expected 0.75)
5.00	2.21	0.0	0.84
4.50	2.12	-3.3	0.82
4.24	2.08	-3.9	0.81
4.00	2.05	-2.8	0.81
3.50	2.03	3.0	0.80
3.25	2.08	9.1	0.81
3.00	2.24	17.6	0.87
3.00	4.25	9.1	0.94
2.75	4.35	7.9	0.91
2.50	4.48	4.7	0.86
2.10	4.67	0.7	0.79
2.00	4.65	1.5	0.78

(b)

Fe-N /Å tri	Fe-O /Å	ΔE/ kcal mol ⁻¹	S ² (expected 0.75)
5.00	2.19	0.0	0.83
4.50	2.13	-2.8	0.82
4.32	2.09	-3.0	0.81
4.00	2.07	-2.5	0.81
3.50	2.04	3.8	0.81
3.25	2.07	10.9	0.82
3.00	2.42	20.6	0.98
3.00	4.41	9.1	0.96
2.75	4.32	8.4	0.93
2.50	4.51	5.5	0.87
2.10	5.26	-1.0	0.79

(c)

Fe-N /Å pyr	Fe-O /Å	ΔE/ kcal mol ⁻¹	S ² (expected 0.75)
5.00	2.21	0.0	0.83
4.50	2.13	-3.0	0.82
4.33	2.10	-3.4	0.81
4.00	2.06	-2.4	0.81
3.50	2.03	4.3	0.81
3.25	4.15	9.3	0.96
3.00	4.15	9.5	0.94
2.75	4.26	8.9	0.91
2.50	4.59	7.4	0.86
2.18	5.38	3.2	0.80

Table 5.2: Fe-N_{azole}, Fe-O_{water} distances, relative energies and spin contamination values for the reaction pathways in the displacement of water by (a) methyl imidazole, (b) methyl triazole, (c) pyridine calculated using BS1 and the BP86 functional. Fe-N_{azole} distances were fixed and all other bonds allowed to minimise freely.

(a)	Fe-N /Å imi	Fe-O /Å	ΔE/ kcal mol ⁻¹	S ² (expected 0.75)
	5.00	2.19	0.0	0.79
	4.50	2.11	-4.3	0.79
	4.23	2.06	-5.1	0.78
	4.00	2.04	-4.4	0.78
	3.50	2.01	1.0	0.76
	3.25	2.05	6.7	0.78
	3.00	2.13	15.0	0.79
	3.50	4.05	10.7	0.83
	3.25	4.09	10.7	0.82
	3.00	4.10	10.4	0.82
	2.75	4.14	9.2	0.81
	2.50	4.48	6.1	0.80
	2.25	4.50	3.4	0.79
	2.05	4.66	0.1	0.76
(b)	Fe-N /Å tri	Fe-O /Å	ΔE/ kcal mol ⁻¹	S ² (expected 0.75)
	5.00	2.17	0.0	0.79
	4.50	2.13	-3.2	0.82
	4.27	2.09	-3.7	0.78
	4.00	2.07	-3.4	0.77
	3.50	2.05	1.5	0.78
	3.25	2.07	7.1	0.77
	3.00	2.16	15.1	0.79
	3.00	4.12	9.6	0.82
	2.75	4.13	8.8	0.81
	2.50	5.38	3.4	0.80
	2.25	5.25	-0.4	0.79
	2.05	5.25	-1.7	0.77
(c)	Fe-N /Å pyr	Fe-O /Å	ΔE/ kcal mol ⁻¹	S ² (expected 0.75)
	5.00	2.18	0.0	0.79
	4.50	2.12	-3.2	0.79
	4.26	2.07	-3.8	0.78
	4.00	2.05	-3.1	0.78
	3.50	2.04	1.8	0.78
	3.25	2.07	8.0	0.76
	3.00	2.16	16.6	0.79
	3.50	3.96	10.7	0.82
	3.25	4.04	11.1	0.82
	3.00	4.07	11.0	0.82
	2.75	4.12	10.0	0.81
	2.50	4.45	7.7	0.80
	2.25	4.65	4.5	0.79
	2.09	5.30	1.6	0.77

The binding curves for triazole are also similar when B3LYP and BP86 are used. The bridged-binding structure is $2.0 \text{ kcal mol}^{-1}$ more stable than the direct binding mode with both functionals. The barrier height going from bridged to direct binding is 12.1 and 13.3 kcal mol^{-1} with the B3LYP and BP86 functionals, respectively. These numbers are close to those for imidazole binding, which is perhaps unsurprising as the two substrates differ only by the replacement of one carbon atom with a nitrogen. The bridged-binding mode is slightly less favoured with triazole binding than imidazole binding, and the barrier heights are slightly lower with triazole than imidazole.

The order of stability of the direct binding modes for the three azoles is triazole>imidazole>pyridine. When the structures of the azoles alone, without any other species present, were minimised with Gaussian 03 and basis set BS1, the Mulliken atomic charges on the nitrogen atoms that bind to the iron for each azole were found to be -0.593, -0.587 and -0.553 with B3LYP and -0.560, -0.556 and -0.514 with BP86 for triazole, imidazole and pyridine, respectively. The trend in these values matches the trend in the stabilities of the direct binding structures. It is likely that the trend in the stabilities of the direct binding structures is due to the increasing negative charge on the nitrogen atom causing a greater affinity for the positively charged iron atom.

5.2.2 2D Potential Energy Surfaces

A smooth pathway between the two minima could not be obtained simply by fixing the Fe-N_{azole} distance. A grid of results was obtained by fixing the Fe-N_{azole} distance between 2.0 and 5.0 Å and the Fe-O_{water} distance between 2.0 and 5.2 Å, using the BP86 functional. These results are shown in the 2D potential energy surfaces (PESs) in Figure 5.5. It can be seen in Figure 5.5 that the two reaction paths obtained by fixing the Fe-N distance correspond to opposite sides of the 2D surface, path A with the Fe-O_{water} distance fixed close to 2.0 Å, and path B with the Fe-O_{water} distance around 5 Å. Using the 2D potential surfaces, it is now possible to trace the geometries as the bound water is displaced by the azole. This is shown in Figure 5.6 for pyridine.

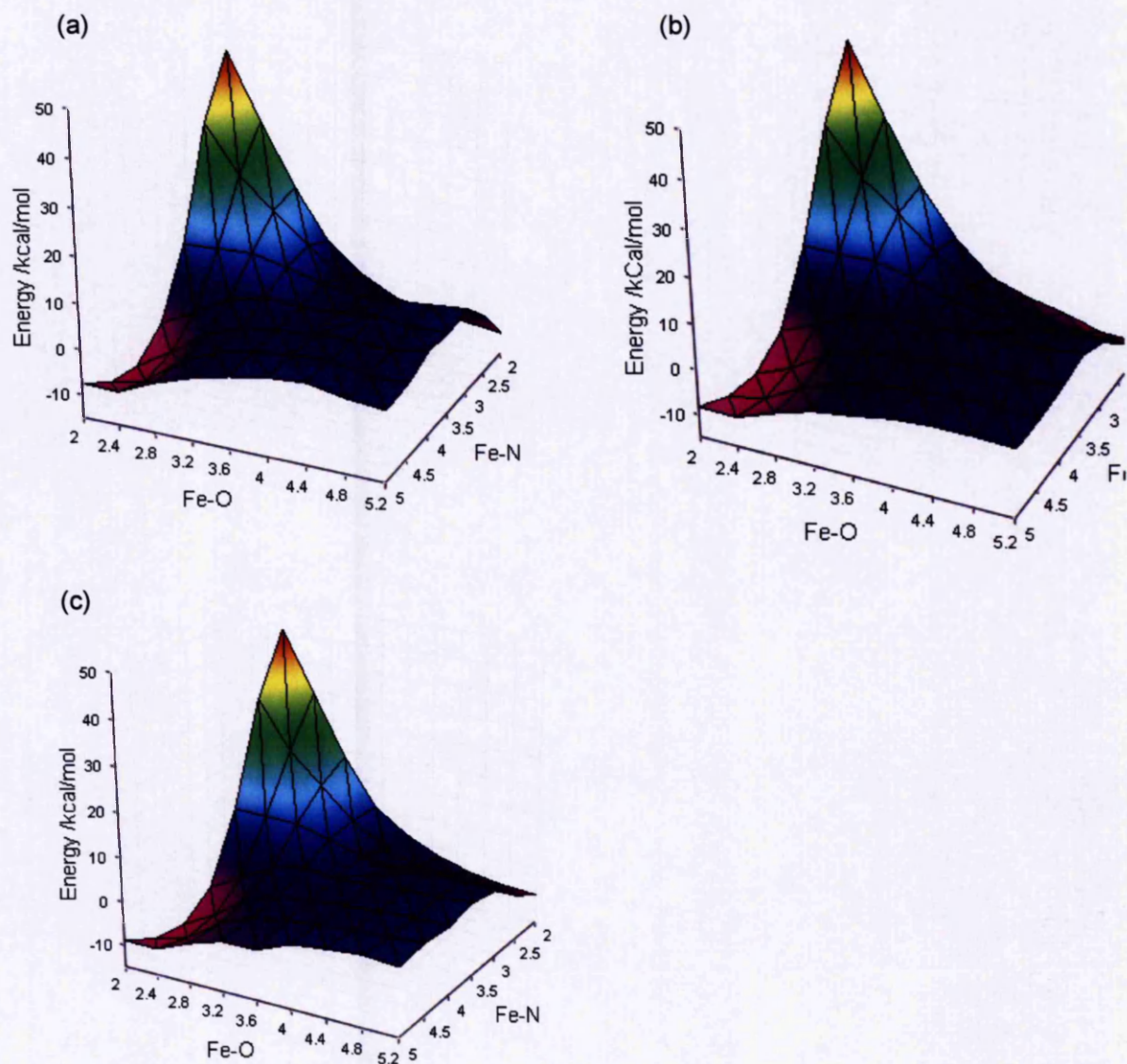


Figure 5.5: PESs of the displacement of water by (a) methyl imidazole, (b) methyl triazole and (c) pyridine from the haem, fixing the $\text{Fe-N}_{\text{azole}}$ and the $\text{Fe-O}_{\text{water}}$ distances, using the BP86 functional and BS1. Distances are given in Å. Transition states are marked for each reaction at $(\text{Fe-N}_{\text{azole}}, \text{Fe-O}_{\text{water}})$ coordinates: imidazole - (3.38, 4.12), triazole - (3.34, 4.30), pyridine (3.16, 4.03).

The 2D PESs, fixing both the $\text{Fe-N}_{\text{azole}}$ and $\text{Fe-O}_{\text{water}}$ distances, shed more light on the nature of the reaction. Examining the surfaces it becomes apparent that a smooth pathway between the two minima can not be obtained by fixing only the $\text{Fe-N}_{\text{azole}}$ distance. Taking the pyridine binding PES as an example, shown in Figure 5.5, for $\text{Fe-N}_{\text{azole}}$ distances of 2.0, 2.5 and 3.0 Å, the lowest energy structure is that with an $\text{Fe-O}_{\text{water}}$

distance of 5.2 Å; but for Fe-N_{azole} distances of 3.5, 4.0, 4.5 and 5.0 Å the lowest energy structures are those with Fe-O_{water} distances of 2.0 and 2.4 Å. This explains why there appear to be two reaction pathways when only the Fe-N_{azole} distance is fixed, when in reality the two apparent pathways are two opposite sides of the PES.

Figure 5.6 shows the structures on the PES closest to the lowest energy reaction path between the two minima in the binding of pyridine. Although the choice of the structures along the reaction path is fairly arbitrary, and other structures could have been selected to illustrate the reaction path, the pictures do show a smooth transition between the two minima, with the water gradually moving to the side of the haem as the pyridine moves closer to the iron, whilst rotating from an angle of 31° with the plane of the haem to 90°.

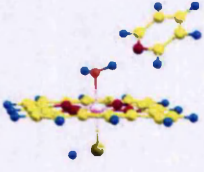
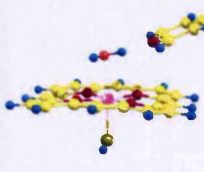
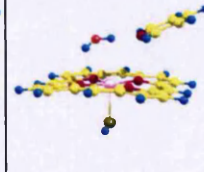
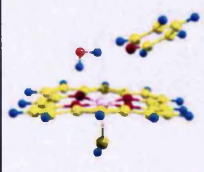
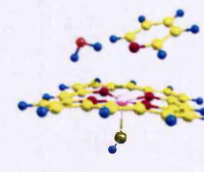
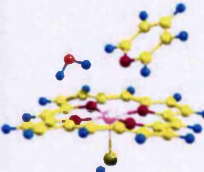
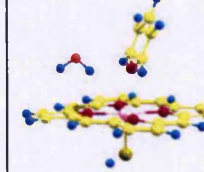
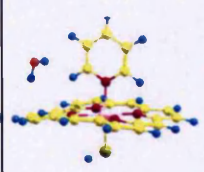
				
Fe-N _{azole} / Å	4.0	4.0	3.5	3.5
Fe-O _{water} / Å	2.0	2.4	2.8	3.2
S ²	0.7785	0.7993	0.8173	0.8221
				
Fe-N _{azole} / Å	3.16	3.0	2.5	2.0
Fe-O _{water} / Å	4.03	3.6	4.0	5.2
S ²	0.8183	0.8161	0.7976	0.7690

Figure 5.6: Images showing the lowest energy pathway between the two minima in the 2D PES of the displacement of water by pyridine.

5.2.3 Transition states

The transition state is the structure at the top of the potential energy curve in a reaction coordinate. It is important to be able to find the transition state in the reaction in order to

be able to determine the energy barrier to reaction. It also gives important information on the reaction mechanism. Mathematically, as the maximum of a curve, a transition structure is defined as the geometry that has a zero derivative of energy with respect to moving in any direction, and a positive second derivative for moving in any direction except along the reaction pathway, for which it is negative. This is illustrated in Figure 5.7.

There are various methods available for finding the transition state of a reaction. As the reaction surface had already been calculated, the simplest method was to take the structures at the top of the reaction surface as starting points, and to carry out a Gaussian OPT=TS calculations. This optimises to a transition state rather than to a local minimum.

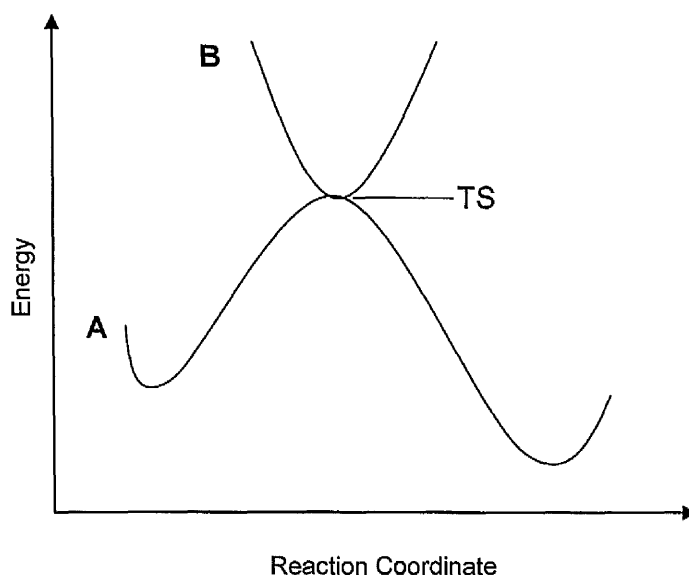


Figure 5.7: Picture illustrating a transition state (TS) on a reaction pathway (A). At the TS the gradient (1st derivative with respect to energy) is 0, and the second derivative is positive for movement in every direction, such as along curve B, except for movement along the reaction pathway A, for which it is negative

Transition state calculations were carried out for each azole starting from each of 9 points across the top of the energy surface. The geometry, relative energy and imaginary frequency of the transition states obtained are shown in Table 5.3 and marked

on the 3D graphs. The imaginary frequencies of the transition states were 20i, 23i and 32i for imidazole, triazole and pyridine, respectively.

The imaginary frequencies of the transition states for the three azoles are very low. Imaginary frequencies of several hundred i cm^{-1} are seen for most reactions involving P450s (Bathelt *et al.*, 2004, de Jong *et al.*, 2004, de Visser, 2006, Hirao *et al.*, 2005), which puts into doubt whether they are true transition states. A comparable imaginary frequency of 29.4i cm^{-1} was found by Kamachi and Yoshizawa (Kamachi and Yoshizawa, 2003) for the rebound step in the hydroxylation of camphor by Compound I in the doublet state. The reported rebound step is virtually barrierless, with an activation energy of $0.7 \text{ kcal mol}^{-1}$. The displacement of water by the three azoles is far from barrierless, with activation energies of 15 or 16 kcal mol^{-1} . The tops of the PESs, shown in Figure 5.5, are very flat. This is due to no bonds being formed or broken at the transition state. In fact it can be seen in Figure 5.8 that the imaginary frequencies only involve rotation of the azole and water molecules in space, with little stretching of the hydrogen bond between them. This could explain the low imaginary frequencies.

Table 5.3: Geometry, energy and imaginary frequency of the transition states of the displacement of haem-water by imidazole, triazole and pyridine, calculated using BS1 and the BP86 functional. Energies are given relative to the water-azole bridged-binding complex.

	Fe-N _{azole} /Å	Fe-O _{water} /Å	Relative Energy / kcal mol^{-1}	Imaginary frequency / cm^{-1}
Imidazole	3.38	4.12	15.9	20i
Triazole	3.34	4.30	14.8	23i
Pyridine	3.16	4.03	14.9	32i

Examining the structures of the three transition states, shown for the pyridine reaction in Figure 5.6, it can be seen that their structures fit with the other structures in the reaction pathway. Also, the transition states are located at the apex of the lowest energy pathway in each of the 2D PESs, as expected. The imaginary frequencies, shown in Figure 5.8 for methyl triazole, involve the azole rocking towards the iron as the water moves away. If the motion is amplified, using the ChemCraft program, a bond is formed between the iron and nitrogen, as shown in Figure 5.8. This again is how the transition state would be expected to behave. In the light of these arguments it would appear that the transition states obtained are indeed true transition states, despite their relatively low imaginary frequencies.

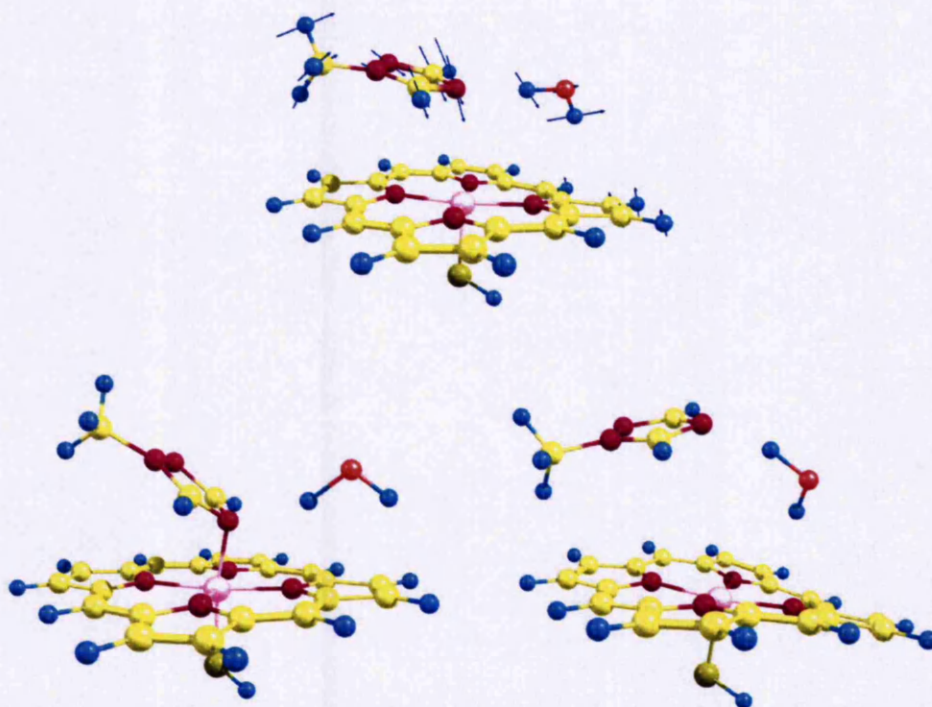


Figure 5.8: Transition state in the displacement of water bound to haem by triazole. Fe-N1 = 3.34 Å and Fe-O = 4.30 Å. The top picture shows the displacement vectors of the imaginary frequency motion with frequency $23i\text{ cm}^{-1}$, the bottom pictures are the extremes of the amplified motion of the $23i\text{ cm}^{-1}$ motion showing the rocking of the triazole towards the haem and the formation of the Fe-N azole bond.

5.2.4 Two minima

There are two minima in the reaction pathways for each azole, shown in Figure 5.9. The first has water bound to the iron, forming a bridge between the haem and the azole. This will be referred to as bridged-binding. The second minimum for each azole has the azole binding directly to the haem. This will be referred to as direct binding. The bridged-binding mode of the three azoles with haem is between 2.0 and 6.6 kcal mol⁻¹ more stable than the corresponding direct binding mode. On first reflection, this seemed a surprising result. When this study was begun there were structures of three P450-antifungal azole complexes available, *Mtb* CYP51 with fluconazole (Podust *et al.*, 2001), P450eryF (CYP107A1) with ketoconazole (Cupp-Vickery *et al.*, 2001) and CYP2B4 with bifonazole (Zhao *et al.*, 2006). Also the structures of *Mtb* CYP51 with 4-phenylimidazole

(Podust *et al.*, 2001) and P450cam with imidazole (Verras *et al.*, 2006) were available. In each of these structures only the direct binding mode is observed, suggesting that if the bridged-binding mode was possible it should be less stable than the direct binding mode.

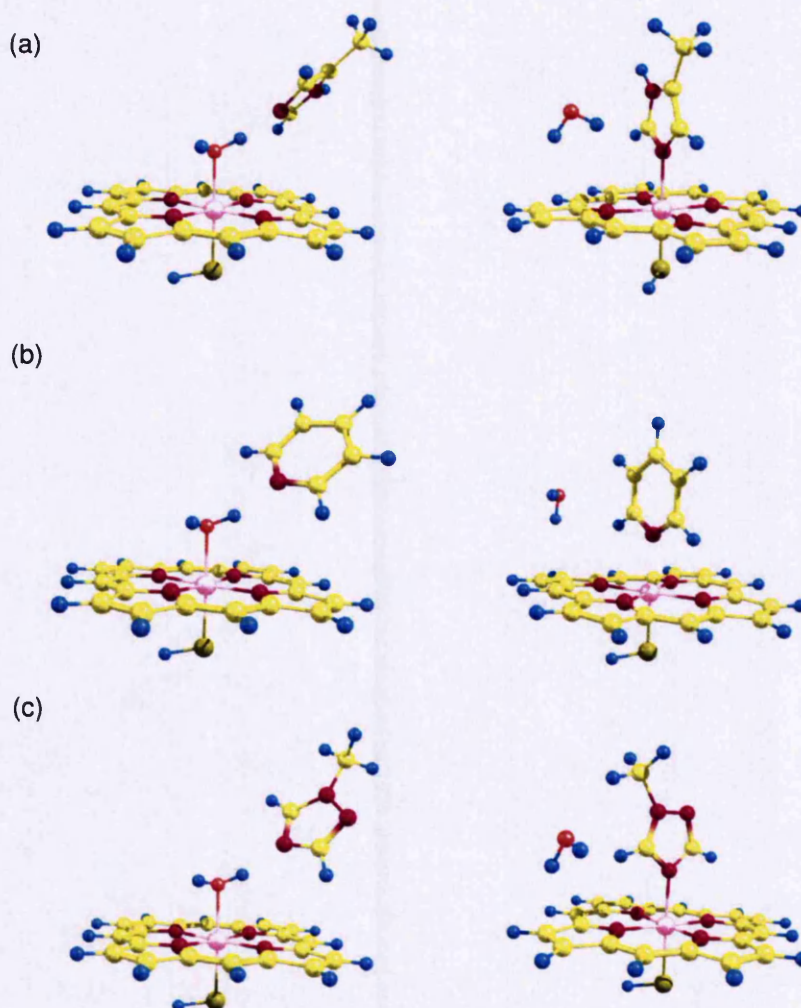


Figure 5.9: Two minima in the displacement of water by (a) methyl imidazole, (b) methyl triazole, (c) pyridine obtained using the B3LYP functional.

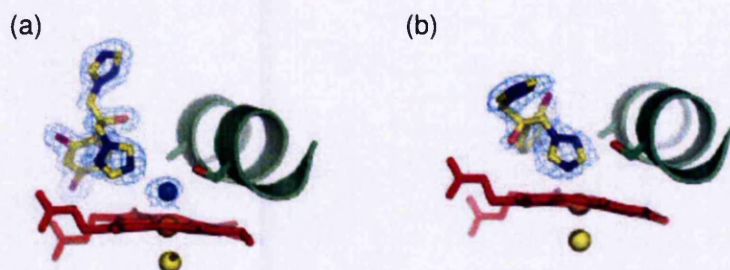
There are a number of factors which could be contributing to the enhanced stability of the bridged-binding mode over that of the direct binding mode. The calculations were carried out in the gas phase, meaning there is nothing else for the species in the calculations to bind to. In a protein, or in solution, the displaced water molecule would

bind to a polar residue, or to other water molecules. This is likely to give rise to additional stability over that of the direct binding mode here, with the water molecule binding to the bound azole.

Around the same time as this work was being carried out, Seward *et al.* (Seward *et al.*, 2006) determined the structure of CYP121 bound to fluconazole. They found a mixture of direct binding and bridged-binding between the fluconazole and the P450 haem both in the crystal structure and in solution, shown in Figure 5.10. This is the only example to date of such a binding mode in P450s. They found that the I-helix in CYP121 is closer to the haem than in other P450s and found a direct correlation between the proportion of bridged-binding and the distance between the I-helix residues and the haem iron. They found that the close proximity of the I-helix to the haem caused the direct bound fluconazole to bind to the iron at an angle 25° removed from the ideal 90° binding angle.

The fluconazole-bound CYP121 structures reported by Seward *et al.* (Seward *et al.*, 2006) agree with the results of this study in that they confirm that the bridged-binding mode is possible, but they do not agree with the finding that the bridged-binding mode is more stable than the direct binding mode. Their results, and all the other azole bound structures available, suggest that the direct binding mode is the most stable, but the close proximity of the I-helix in CYP121 destabilises the direct binding mode, causing the bridged-binding mode to be observed. We will discuss later how the residues in the I-helix may also stabilise the bridged-binding mode in CYP121.

Figure 5.10: (a) Bridged-binding and (b) direct binding observed in the fluconazole-bound CYP121 crystal structure, reproduced from (Seward *et al.*, 2006). The I helix is shown in green.



5.2.5 Comparing the affinity of the P450 haem for water and azoles

The first thing to ascertain is whether the stability of the bridged-binding complex comes from the affinity of the water for the haem, or from the azole-water bond, or another factor. Gaussian optimisation calculations were carried out on complexes of haem in the doublet, quartet and sextet spin states with each of the three azoles without water present, and on the water bound structure, using BS1 and both the B3LYP and BP86 functionals. Also, each component, the haem, the azoles and water, was optimised using the same method. The affinities of the substrates for iron-haem were calculated by subtracting the gas phase values of the separate components from the energy of the complex. As the unbound haem is generally considered to be in the sextet spin state, this was used as the reference value for the haem throughout. The affinities of the azoles and water for haem are shown in Table 5.4.

Table 5.4: Binding affinities of imidazole, pyridine, triazole and water for the P450 haem in the doublet, quartet and sextet spin state using BS1 and the B3LYP and BP86 functionals. Energies are given in kcal mol⁻¹ and are calculated relative to the sum of the separate components, using high spin ferric porphyrin to represent the unbound haem.

	Relative Energies /kcal mol ⁻¹					
	B3LYP			BP86		
	High	Intermediate	Low	High	Intermediate	Low
Imidazole	-3.8	-5.5	-13.9	-3.3	-14.1	-31.3
Haem -Water- Imidazole	-16.7	-18.5	-25.2	-16.3	-27.2	-42.6
Triazole	-2.9	-4.7	-12.6	-2.3	-13.2	-29.7
Haem-Water- Triazole-	-15.5	-17.1	-23.8	-15.0	-25.8	-40.3
Pyridine	-2.4	-4.6	-11.6	-2.3	-13.4	-29.5
Haem-Water- Pyridine	-15.5	-17.1	-23.7	-15.2	-25.8	-40.4
Water	-7.2	-8.1	-13.1	-14.1	-18.1	-30.1

In order to confirm that summing the gas phase values of the separate components gave an accurate representation of the separated components, a calculation was carried out taking the converged low spin haem and imidazole structures and fixing them a distance of 40 Å apart. Upon optimisation the energies were found to be within 0.1 kcal mol⁻¹ of the sum of their gas phase values. This shows that the sum of the gas phase values of the separate components is an accurate representation of the energy of the separated components.

It can be seen in Table 5.4 that the affinity of water for low spin haem is higher than those of pyridine and triazole, and only slightly lower than that of imidazole, with both B3LYP and BP86. The affinity of water for intermediate and high spin haem is significantly higher than those of the three azoles, especially using the BP86 functional. The very large stabilisation of the low spin states using the BP86 functional is symptomatic of the destabilisation of high spin states by the BP86 functional referred to earlier.

As the calculations were carried out on species in the gas phase, without the presence of a protein, there are no available experimental results with which they can be directly compared. The binding affinities of substrates for P450s can be calculated from the K_d value using the equation:

$$\Delta G^\circ = -RT \ln K_d$$

where ΔG° is the free energy of the reaction, R is the gas constant and T is the temperature in K. Binding affinities vary widely between the different azoles and the different P450s (Gustafsson *et al.*, 2004, Lawson *et al.*, 2004, McLean *et al.*, 2002b). Taking the results from Chapter 3 for CYP132 as an example, this gives a value of 2.0 kcal mol⁻¹ for the dissociation of imidazole and 7.6 kcal mol⁻¹ for the dissociation of miconazole.

The protein environment can have a great effect on the binding of substrates, and so comparisons with these figures must not be taken too far. Perhaps a closer comparison can be made with the work of Lamar and Walker (Lamar and Walker, 1972) who carried out NMR analyses on *meso*-tetraphenylporphyrin iron bis N-methylimidazole. Calculations from NMR line broadening values gave a dissociation energy of a single N-methylimidazole of 17 kcal mol⁻¹. For comparison purposes, calculations were carried out on mono and bis imidazole iron-porphyrin using the BP86 functional. This gave a value of -29.6 kcal mol⁻¹ for the dissociation of a single imidazole from bis-imidazole iron-porphyrin. This is compared to -31.3 kcal mol⁻¹ for dissociation of imidazole from the thiolate-bound iron-porphyrin. This suggests that the expected value for the dissociation of imidazole from thiolate bound haem should be slightly less than 17 kcal mol⁻¹, which is in good agreement with the values obtained using the B3LYP functional.

Harris *et al.* (Harris *et al.*, 2004), using the B3LYP functional with the Jaguar program, found that water binding to haem was 0.6 kcal mol⁻¹ more favourable than that of DMBP, a secondary amine, but 3.6 kcal mol⁻¹ less favourable than phenylimidazole. Neither of these structures is exactly the same as those studied here, but one would expect the results from the three azoles used in this study to be closer to those of phenylimidazole than to DMBP. The absolute water binding energy of -15 kcal mol⁻¹ is slightly different to that obtained here. This is to be expected as Harris *et al.* calculated their binding energies relative to a single point calculation with the ligand displaced 20 Å, whereas the values in this study are calculated relative to optimised structures.

The similar or higher affinity for water compared to that of the azoles is a surprising result. Azoles are known to bind to P450s preferentially over water, even when water is used as a solvent. This was shown in Chapter 3 of this work with the binding of a range of azoles to CYP132. Azole binding in the presence of water has also been shown for numerous other P450s, including CYP51, CYP121 (McLean *et al.*, 2002b), P450BM3 (Smith *et al.*, 2003) and CYP3A4 (Lamb *et al.*, 2000).

Strickland and Harvey (Strickland and Harvey, 2007) recently carried out *ab initio* and DFT studies on the binding of CO, NO and H₂O to ferrous deoxy-haem. Comparing the values from the DFT studies with those obtained with the high level *ab initio* calculations, they found that pure functionals such as BP86 overestimate the binding energy of H₂O and CO, but hybrid functionals such as B3LYP give accurate results. They were unable to carry out *ab initio* calculations on NO binding, but found that the binding energies obtained with B3LYP were significantly lower than experiment. They postulate that B3LYP may not be accurate for NO binding.

In the light of these findings it may be that the binding energy of azoles is also underestimated by hybrid functionals such as B3LYP. However, the fact that the binding energies for water and the azoles are also similar when using the BP86 functional would suggest that the trend is not due to errors associated with the functional used.

The bridged-binding mode is between 11 and 12 kcal mol⁻¹ more stable than the azole binding alone for each azole. This stability is more pronounced than in the 2D binding

curves in the previous section, due to the direct binding mode here not including a water molecule and thus not being stabilised by the water binding to the azole. There are various ways in which the binding of the displaced water molecule can be modelled. These will be discussed in section 5.2.11

5.2.6 Varying basis set

The size and quality of the basis set used can have a significant effect on the results of any DFT calculation. In order to assess the accuracy of the basis set used, the energy of water and triazole binding was calculated using a selection of basis sets. The basis sets used were as follows:

BS1 - LANL2DZ, with pseudo potentials (Fe, S), 6-31G* (O, N), 6-31g (C, H)

BS2 - LANL2DZ, with pseudo potentials (Fe, S), 6-31G** (O, N, C, H)

BS3 - LANL2DZ, with pseudo potentials (Fe), 6-31G** (S, O, N, C, H)

BS4 - LACVP, with pseudo potentials (Fe), 6-31G** (S, O, N, C, H)

BS5 - 6-31G(d,f), no pseudo potentials (Fe, S, O, N, C, H)

BS6 - 6-31G*, no pseudo potentials (Fe, S, O, N, C, H)

Basis sets BS1, 2, and 3 use the LANL2DZ basis with pseudo potentials to describe the core electrons for iron, with varying sizes of basis sets on the remaining atoms. BS4 uses the LACVP basis for iron, which was copied from the Jaguar program. This is very similar to LANL2DZ, with one fewer s orbital than LANL2DZ. BS5 and 6 used versions of the 6-31G basis for all atoms. Although 6-31G basis sets are not usually used for iron, they have been developed to be applicable for all first row transition metals (Rassolov *et al.*, 1998).

The results of the calculations with the various basis sets are shown in Table 5.5. The general trend in the results is similar for all basis sets. With basis sets BS1-4 the relative energies of triazole and water binding are similar, with the bridged-binding complex stabilised by around 10 kcal mol⁻¹. It is interesting that basis sets BS5 and 6 give the more expected result that triazole binding is more favourable than water binding, by around 1 kcal mol⁻¹ with B3LYP, and by 2 kcal mol⁻¹ with BP86. 1 or 2 kcal mol⁻¹ is still not a significant difference between the binding energies, however. Further work

comparing the results with high level *ab initio* calculations would be necessary to assess the relative accuracies of the different basis sets for this system.

Table 5.5: Relative energies of triazole and water binding to haem using a range of basis sets. BS1 - LANL2DZ, with pseudo potentials (Fe, S), 6-31G* (O, N), 6-31g (C, H), BS2 - LANL2DZ, with pseudo potentials (Fe, S), 6-31G** (O, N, C, H), BS3 - LANL2DZ, with pseudo potentials (Fe), 6-31G** *(S, O, N, C, H), BS4 - LACVP, with pseudo potentials (Fe), 6-31G** *(S, O, N, C, H), BS5 - 6-31G(d,f), no pseudo potentials (Fe, S, O, N, C, H), BS6 - 6-31G*, no pseudo potentials (Fe, S, O, N, C, H). Energies are given relative to separated components.

Functional	Basis Set	Relative Energies /kcal mol ⁻¹		
		Triazole	Water	Water-Triazole Bridged
B3LYP	BS1	-12.6	-13.1	-23.9
	BS2	-13.4	-13.4	-24.3
	BS3	-13.3	-14.0	-24.7
	BS4	-13.4	-14.0	-24.7
	BS5	-8.9	-7.6	-19.0
	BS6	-7.9	-7.2	-18.5
BP86	BS1	-29.7	-30.1	-40.3
	BS2	-28.7	-29.1	-39.6
	BS3	-30.3	-31.2	-41.4
	BS4	-30.6	-31.2	-41.4
	BS5	-29.7	-27.3	-38.3
	BS6	-28.5	-26.8	-37.6

5.2.7 Varying cysteine model

In the previous calculations the thiolate ligand to the haem was modelled as a SH group, as this was found by Ogliaro *et al.* (Ogliaro *et al.*, 2000b) to best model the P450 haem when using gas phase calculations. Others have used SCH₃ to model the thiolate ligand. In order to determine whether there is a significant difference in the results given by the two methods, the relative energies of the binding of the three azoles and water to Fe-porphyrin-SCH₃ were calculated using BS1 and the BP86 functional.

Table 5.6: Relative energies of the binding of water, imidazole, triazole and pyridine to Fe-porphyrin-SCH₃ using BS1 and the BP86 functional. Note, direct binding refers to binding without a water molecule present.

Cysteine model	Relative Energies /kcal mol ⁻¹						
	Water	Imidazole		Triazole		Pyridine	
		Direct	Bridged	Direct	Bridged	Direct	Bridged
-SCH ₃	-30.7	-31.6	-42.8	-30.2	-40.6	-29.6	-40.8
-SH	-30.1	-31.3	-42.6	-29.7	-40.3	-29.5	-40.4

The results are shown in Table 5.6. It can be seen that using SCH₃ rather than SH to model the thiolate ligand makes almost no difference to the relative binding energies. This shows that any discrepancy between the calculations and experimental values are not due to the model used for the cysteine ligand. In the light of these findings it was decided to use SH for all other calculations as it was not necessary to increase computation time by using SCH₃.

5.2.8 Rotation energy

It was observed that the dihedral angle between the SH bond and the bound substrates was not consistent throughout the reaction pathways. In order to assess the importance of the rotation of the axial ligands, calculations were carried out by increasing the dihedral angle between Fe-N_{porph} and the axial ligands in 15° increments and minimising at each point. The triazole-bound and water-bound structures were used, rotating around the iron-substrate bond. Figure 5.11 shows graphs of the results of the rotations.

It can be seen that the rotation of the bound water has a total energy cost of around 1 kcal mol⁻¹, and the rotation of triazole requires 0.4 kcal mol⁻¹. The lowest energy in the rotation of water occurs when the water hydrogen atoms line up with the porphyrin nitrogen atoms, forming hydrogen bonds; the highest energy structures being those where the O-H bonds point between the porphyrin nitrogen atoms. There is a small interaction between the H_{water} and the H_S atoms, causing slightly higher energies when the hydrogen atoms are lined up than when they are pointing away from each other. This only increases the energy by 0.2 kcal mol⁻¹, however, and is thus negligible.

In contrast to the rotation of water, steric hindrance means the triazole has the lowest energy when it is pointing between the porphyrin nitrogen atoms, with a little variation depending on the position of the sulphur atom. These calculations on the rotation energies show that the position of the sulphur atom relative to the bound substrate makes little difference. During the course of the minimisation, the bound substrate will be lined up with the nitrogen atoms in the lowest energy conformation, and it makes little difference to the calculation which sector the sulphur atom is in.

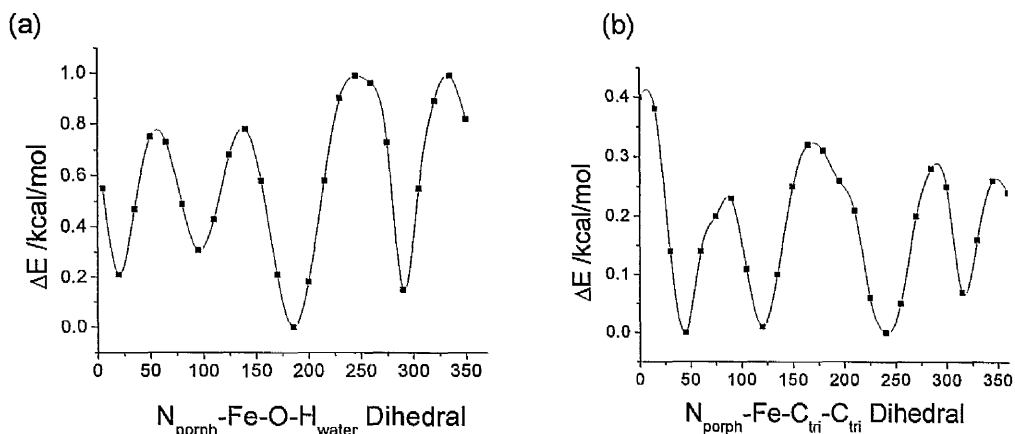


Figure 5.11: Rotation energies of (a) water bound to haem, (b) triazole bound to haem. Angles were increased in 15° increments and the structure minimised at each point

5.2.9 Counterpoise Correction calculations

5.2.9.1 Basis Set Superposition Error

A possible source of error in the calculations is Basis Set Superposition Error (BSSE) (Jensen, 1999). This occurs when the energy of a complex is compared to that of its separated components, such as in the binding of haem to the azole and water substrates described in this study. No basis set will ever completely describe the electronic orbitals of a system, but the larger the basis set becomes, the more closely it will resemble the true orbitals. When calculations are carried out on two monomers in a complex, the basis sets on each monomer partially overlap those on the other monomer, meaning each atom is covered by more orbitals than if the monomers were considered individually. This enlarging of each basis set by the overlapping with the basis set on the other monomer causes a stabilisation compared with the separated monomers. This stabilisation is referred to as the BSSE.

A common method of determining the size of the BSSE is to use a counterpoise correction (Boys and Bernardi, 1970, van Duijneveldt *et al.*, 1994). This involves comparing the energy of each monomer both with its own basis set and with the combined basis set of the complex; the basis sets of the other monomers being fixed in place without the nuclei of the monomers being present, such that for a complex of two monomers A and B, (Jensen, 1999):

$$\Delta E_{CP} = E(A)_{ab} + E(B)_{ab} - E(A)_a - E(B)_b$$

Where ΔE_{cp} is the counterpoise correction, $E(A)_{ab}$ and $E(B)_{ab}$ are the energy of monomers A and B calculated with the basis set of the whole complex AB, and $E(A)_a$ and $E(B)_b$ are the energies of monomers A and B calculated with their own basis sets.

5.2.9.2 Counterpoise Correction Calculations

Counterpoise correction calculations were carried out on the low, intermediate and high spin states of haem bound to each of the azoles and to water, as well as the bridged-binding imidazole structure. The B3LYP and BP86 functionals were used with BS1, as well as BP86 with BS2. The results of the calculations showing the BSSE values and the counterpoise corrected energies are shown in Table 5.7. Figure 5.12 and Figure 5.13 show graphs of the counterpoise corrected relative energies calculated with the B3LYP and BP86 functionals. The energies without counterpoise correction are also included for comparison. The high spin water-imidazole bridged species failed to converge. As the high spin states have little importance in this study it was decided not to spend more time trying to obtain values for them. The geometries of the complexes also changed slightly when the counterpoise correction was used. Table 5.8 shows the Fe-S, Fe-O and Fe-N distances for the structures studied.

The effect of BSSE is to reduce all of the binding energies, as seen in Figure 5.12 and Figure 5.13. The BSSE is slightly higher for the calculations involving the binding of water - both for water only and for bridged-binding. This has the effect of increasing the stability of azole binding slightly relative to that of water so that azole binding becomes 2.2, 0.7 and 0.3 kcal mol⁻¹ more stable than water binding using the B3LYP functional and BS1, and 2.6, 1.2 and 1.4 kcal mol⁻¹ more stable with BP86. Although these differences are still not large, the trend is in the right direction, and azole binding is now slightly more stable than water binding. The relative stability of the bridged-binding mode is also slightly reduced. The relative stability of the water-imidazole bridged complex compared to that of imidazole binding directly to haem decreases from 11.3 to 7.1 kcal mol⁻¹ using both B3LYP and BP86 with BS1. Again, the trend in the stability of the bridged-binding complex is in the right direction.

Table 5.7: Counterpoise corrected relative energies and BSSE values for the binding of low, intermediate and high spin haem to water, imidazole, triazole, pyridine, and the water-imidazole bridged complex. Energies are in kcal mol⁻¹ and were calculated relative to the separate components. The high spin water-imidazole bridged-binding species failed to converge.

		Relative Energies /kcal mol ⁻¹									
		Water		Imidazole		Triazole		Pyridine		Fe-water-imidazole	
		BSSE	E	BSSE	E	BSSE	E	BSSE	E	BSSE	E
B3LYP BS1	Low	7.6	-5.2	6.6	-7.4	6.4	-5.9	5.9	-5.5	10.2	-14.5
	Int	3.3	-2.8	4.5	-0.8	3.9	-0.4	3.8	-0.6	8.1	-9.8
	High	3.4	-2.3	5.1	1.6	4.2	2.0	0.9	-0.8		
BP86 BS1	Low	8.0	-21.9	6.6	-24.5	6.4	-23.1	6.1	-23.3	10.7	-31.6
	Int	3.7	-12.1	4.4	-9.6	3.8	-9.1	3.8	-9.4	8.4	-18.2
	High	3.6	-1.9	4.9	1.9	4.3	2.4	4.3	2.3		
BP86 BS2	Low	8.2	-21.7	7.0	-21.4	6.8	-23.4	-20.3	6.5	11.1	28.1
	Int	3.5	-12.5	4.7	-6.7						
	High	3.6	-1.8	5.3	5.4						

Table 5.8: The axial bond distances (Å) in the counterpoise corrected structures of the water- and azole-bound haem species, and the water-imidazole bridged-binding structure using the B3LYP and BP86 functionals and basis sets BS1 and BS2. The high spin water-imidazole bridged-binding species failed to converge.

		Water		Imidazole		Triazole		Pyridine		Fe-water-imidazole		
		Fe-S	Fe-O	Fe-S	Fe-N	Fe-S	Fe-N	Fe-S	Fe-N	Fe-S	Fe-N	Fe-O
B3LYP BS1	Low	2.30	2.23	2.32	2.14	2.31	2.16	2.32	2.23	2.32	4.42	2.14
	Int	2.46	3.82	2.51	2.68	2.50	2.82	2.49	2.91	2.51	4.84	2.54
	High	2.40	3.88	2.45	2.61	2.43	2.85	2.40	7.42			
BP86 BS1	Low	2.26	2.20	2.28	2.10	2.28	2.10	2.29	2.15	2.28	4.35	2.11
	Int	2.44	3.70	2.49	2.61	2.48	2.72	2.48	2.78	2.50	4.73	2.50
	High	2.39	3.78	2.46	2.54	2.44	2.66	2.44	2.71			
BP86 BS2	Low	2.27	2.22	2.29	2.09	2.29	2.10	2.29	2.15	2.29	4.31	2.11
	Int	2.44	3.68	2.50	2.62							
	High	2.39	3.76	2.45	2.55							

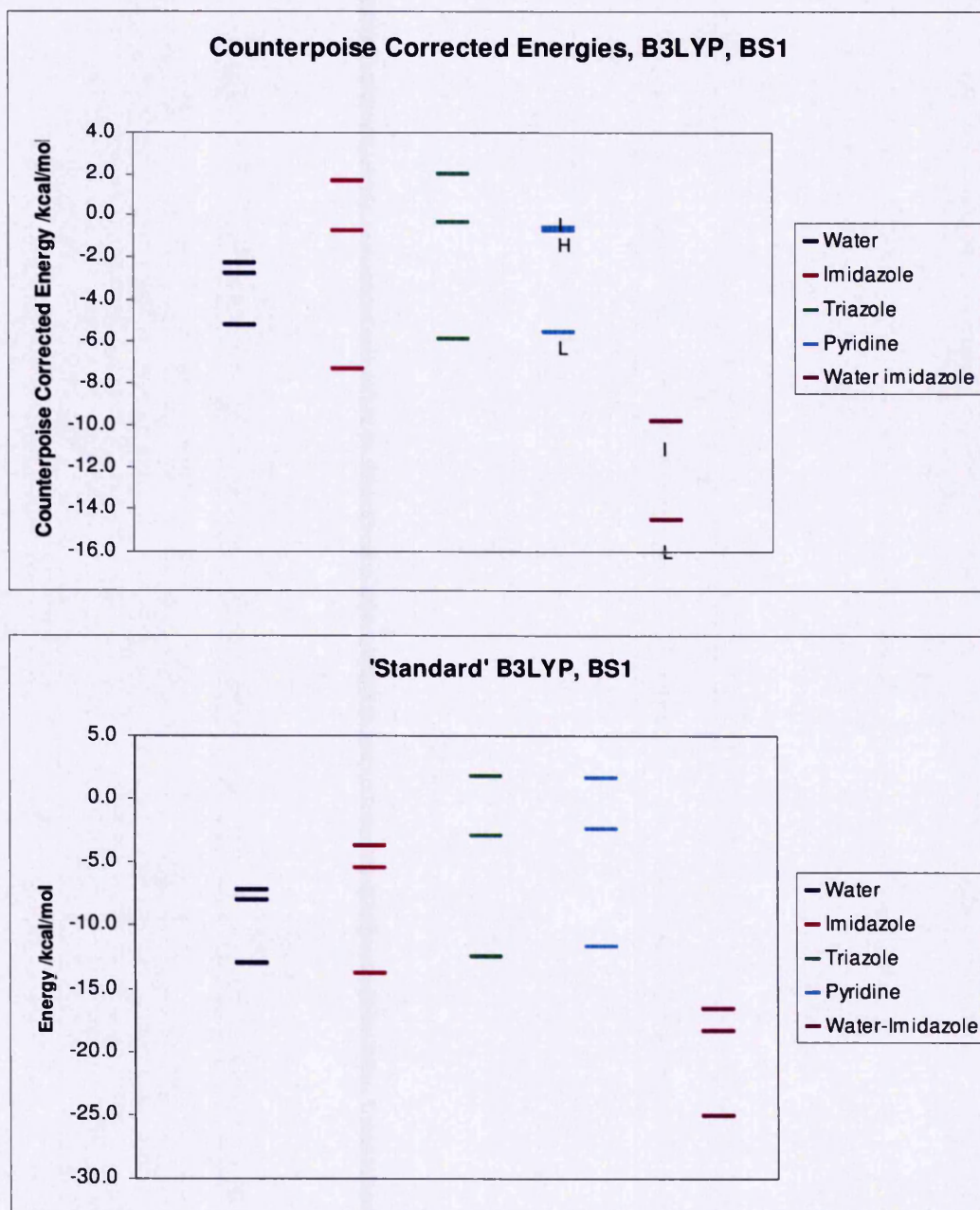


Figure 5.12: Graphs of counterpoise corrected relative energies (top) and 'standard' energies with no correction in kcal mol⁻¹ for the binding of low (L), intermediate (I) and high (H) spin haem with water, imidazole, triazole, pyridine and the water-imidazole bridged complex using the B3LYP functional and BS1. Energies are calculated relative to separated components. The order of the energy of the spin states is low<intermediate<high throughout, except where labelled otherwise. The water-imidazole bridged high spin species failed to converge.

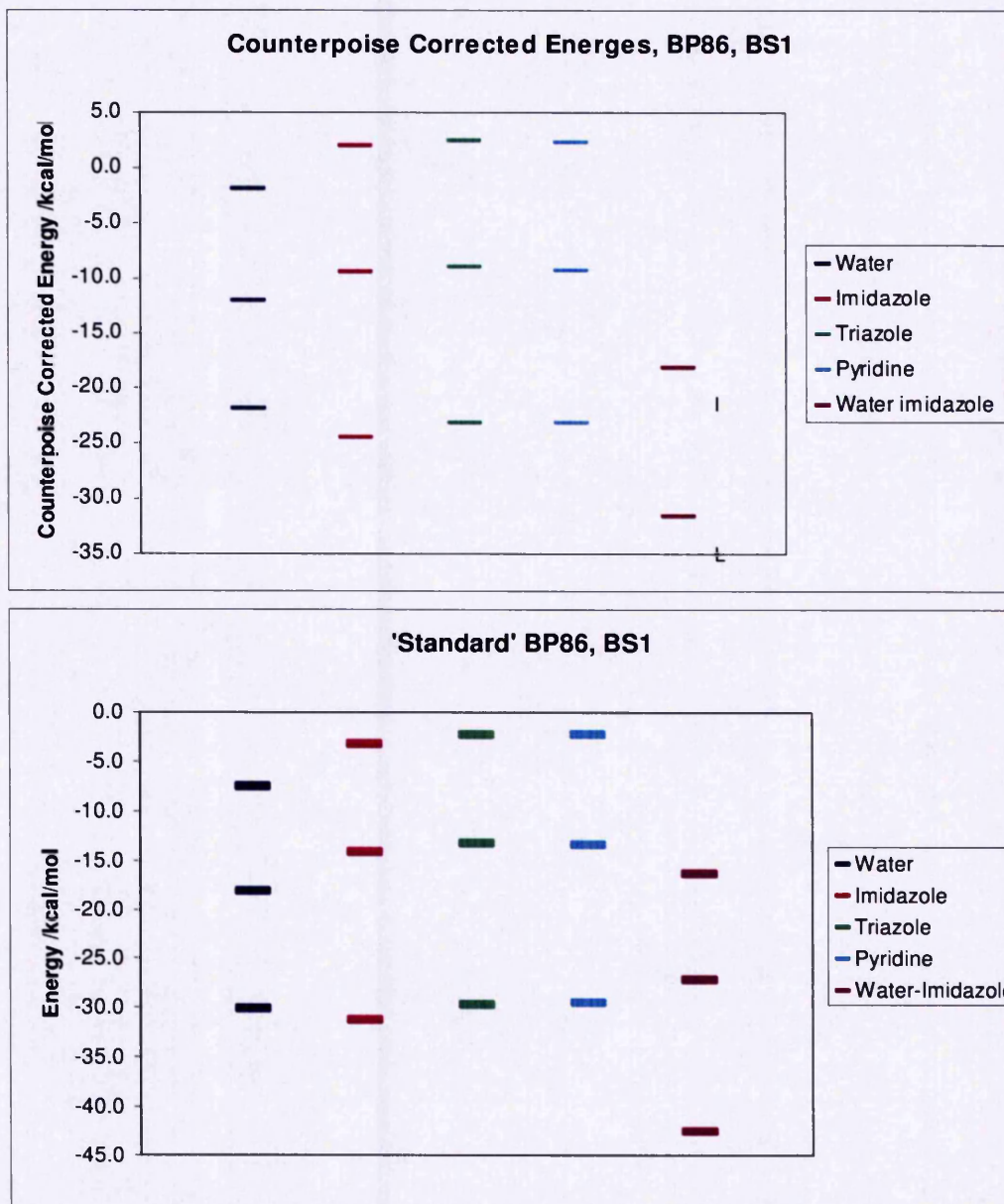


Figure 5.13: Graphs of counterpoise corrected relative energies (top) and 'standard' energies with no correction in kcal mol⁻¹ for the binding of low (L), intermediate (I) and high (H) spin haem with water, imidazole, triazole, pyridine and the water-imidazole bridged complex using the BP86 functional and BS1. Energies are calculated relative to separated components. The order of the energy of the spin states is low<intermediate<high throughout, except where labelled otherwise. The water-imidazole bridged high spin species failed to converge.

5.2.10 Solvation models

The calculations of the binding of the azoles and water to haem do not take into account the effect of the displaced substrates binding to other protein residues or to bulk water. This could have a significant impact on the relative stability of the bound complexes, especially on that of the direct binding complex in the 1D binding curves described in Section 5.2.1 of this chapter, as the water molecule in the direct binding mode is unlikely to be in an optimum binding position when the azole is bound to the haem. In contrast in the bridged-binding mode the water binds to the electropositive iron atom, and forms a hydrogen bond with the incoming azole nitrogen atom.

It was difficult to devise a system that will account for the solvation of the displaced substrates adequately using DFT methods. Implicit solvent methods, using the PCM keyword in Gaussian, account for the average effect of the solvent, but can not account for the effect of hydrogen bonding with water solvent. Various explicit solvation models are described below, with varying effects on the relative energies. These small scale models were designed to account for specific binding between polar groups, especially hydrogen bonding.

5.2.10.1 Implicit solvent

The calculations carried out so far were all in the gas phase. In order to assess the effect of carrying out the reaction in a solvent, or in the protein environment, an implicit solvent was included, using the Polarizable Continuum Model (PCM) keyword in Gaussian 03 (www.gaussian.com). This is based on the work of Barone and coworkers (Cossi *et al.*, 1996, Cossi *et al.*, 1999, Cossi and Barone, 2000, Cossi and Barone, 2001, Cossi *et al.*, 2001, Cossi *et al.*, 2002, Cossi *et al.*, 2003, Miertus *et al.*, 1981, Miertus and Tomasi, 1982) and Tomasi, Mennucci and coworkers (Cammi *et al.*, 1999, Cammi *et al.*, 2000, Mennucci *et al.*, 1997, Mennucci and Tomasi, 1997). The water solvent model was used to model the haem reactions carried out in water, and chlorobenzene was also used as this is often used to model the non-polar protein environment (Ogliaro *et al.*, 2002b, Wang *et al.*, 2006).

The results of the PCM calculations for triazole and water binding are shown in Table 5.9. It can be seen that in water solvent, water binding is more favourable than triazole

binding by 2.5 kcal mol⁻¹. This is compared with a stabilisation of just 0.5 kcal mol⁻¹ in the gas phase. In chlorobenzene the difference is just 1.1 kcal mol⁻¹. This falls to 0.4 kcal mol⁻¹ when the larger BS2 is used. This compares to 0 kcal mol⁻¹ with BS2 in the gas phase.

Using the BP86 functional and BS2, triazole binding in chlorobenzene is favoured by 0.4 kcal mol⁻¹. This is compared to -0.4 kcal mol⁻¹ in the gas phase.

Table 5.9: PCM calculations of haem binding to triazole and water using parameters for water and chlorobenzene. Relative energies are compared to separate components.

Functional and Basis Set	Implicit Solvent	Relative Energies /kcal mol ⁻¹		
		Triazole	Water	H ₂ O-Triazole bridged
B3LYP, BS1	Water	-7.1	-9.6	-15.7
	Chlorobenzene	-10.5	-11.6	-21.8
B3LYP, BS2	Chlorobenzene	-8.9	-9.3	-17.3
BP86, BS2	Chlorobenzene	-10.3	-9.9	-17.9

The relative stability of the water-triazole bridged-binding complex is slightly decreased, being only 8.6 kcal mol⁻¹ more stable than triazole direct binding in water solvent using B3LYP and BS1, and 8.4 kcal mol⁻¹ more stable in chlorobenzene solvent, using BS2. This is compared to 11.3 and 10.9 kcal mol⁻¹ using the B3LYP functional and BS1 and BS2, respectively, in the gas phase. Using B3LYP and BS1 in chlorobenzene, the gap remains the same at 11.3 kcal mol⁻¹.

Upon examination of the stabilisation of the individual components by the implicit solvent it becomes apparent that it is the stabilisation of the free triazole that drives the increased stability of water binding over that of triazole. In water solvent, free triazole is stabilised by 8.0 kcal mol⁻¹ compared to 6.5 kcal mol⁻¹ stabilisation of free water, and in chlorobenzene solvent the figures are 3.4 and 2.6 kcal mol⁻¹, respectively. This additional stabilisation of triazole over that of water can be attributed to the larger size of triazole compared to water, presenting a larger surface area to make contact with the implicit solvent.

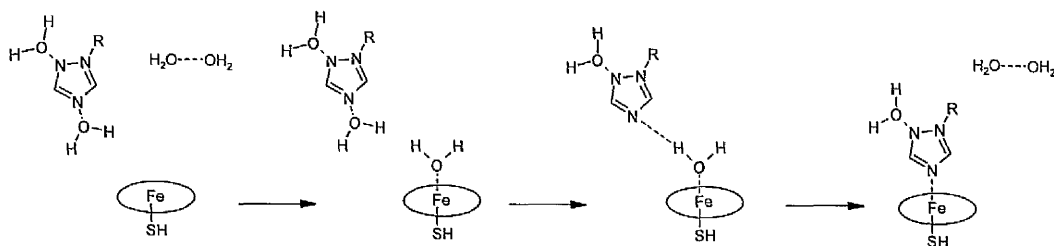
These values show that there is little qualitative difference to the results using an implicit solvent compared to those in the gas phase, especially using chlorobenzene, which mimics the protein environment. The trend in the results is taken slightly further from experiment in the additional stability of water binding over triazole, but is brought closer to experimental observations with the slight destabilisation of the bridged-binding mode compared to direct binding with certain methods. The bridged-binding mode still remains significantly more stable than direct binding using implicit solvent methods, and thus it is necessary to consider other factors to account for the experimental observation of direct binding being the dominant binding mode unless environmental effects dictate otherwise.

5.2.11 Explicit solvent models

In the previous section it was seen that the use of an implicit solvent in the calculations could not account for the discrepancies between the calculated relative energies of azole and water binding, and experimental observations. It seemed likely that the primary reason for the stability of azole-water bridged-binding to haem over that of direct azole binding was that the calculations did not adequately account for the stabilisation of the free water molecule once it had dissociated from the haem.

Implicit solvent methods do not take into account specific hydrogen bonding interactions between the free water molecule and the bulk solvent or protein residues. In order to more accurately determine the stabilisation of the individual components by specific interactions with water solvent, or protein residues, various explicit solvation models were used. The two models shown in Figure 5.14 were used to model interactions between the components and water molecules in the solvent. Calculations were also carried out on clusters of 9 and 10 water molecules, and triazole surrounded by 9 water molecules in order to assess the effect of specific interactions between water solvent and the free components.

Model 1



Model 2

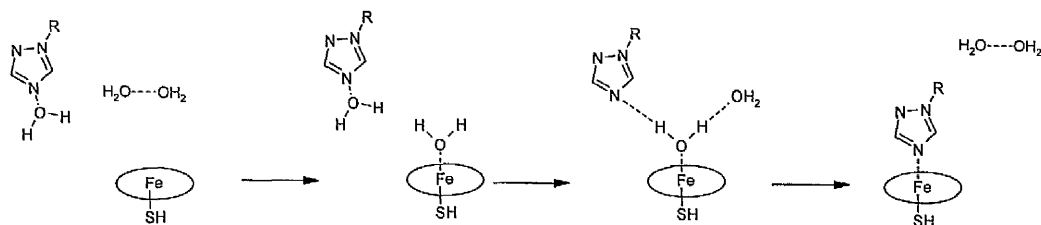


Figure 5.14: Models of the solvation of triazole and water binding to haem. The haem macrocycle is represented by an ellipse around the iron atom. Where no bond is shown, fragments were minimised independently of each other, and their absolute energies added together. The diagrams for each model show, from left to right, (a) the free, 'solvated' components, relative to which the energies were calculated, (b), the water bound haem, and solvated triazole, (c) the bridged-binding species, (d) the triazole bound haem, and solvated water.

It is not possible to accurately model explicit solvation using DFT models of this size. The few interactions modelled in Models 1 and 2 in Figure 5.14 will not take into account the solvation of the whole molecule, and the clusters of 9 and 10 water molecules, which firstly do not take into account the wider effect on the structure of the water solvent, contain so many degrees of freedom that the energies given by the DFT calculations are too dependent on the conformation of the water molecules to give an accurate description of the effect on the energy of the system of adding a water or triazole molecule to a cluster of water molecules. Molecular dynamics (MD) calculations on a much larger cluster of water molecules would be necessary to measure accurately the energy of solvating either a water or triazole molecule, in order to ensure the lowest energy conformation of water molecules had been obtained, and that long range effects had been taken into account. QM/MM MD calculations of this type are beyond the scope of this study.

Keeping these provisos in mind, the results shown in Table 5.10 may be considered. It can be seen that using Model 1, the energy of solvation of triazole is greater than that of water, causing the energy of triazole binding to haem to become much less favourable, with an energy of $\pm 1 \text{ kcal mol}^{-1}$, using the B3LYP or BP86 functionals. The stabilisation of free water and triazole due to solvation means the bridged-binding mode is considerably destabilised in Model 1, with an energy of -1.7 or $-3.0 \text{ kcal mol}^{-1}$, compared to energies of -24.3 and $-39.6 \text{ kcal mol}^{-1}$, using the B3LYP and BP86 functionals, respectively.

Using Model 2, the free water and triazole molecules are both only solvated by one water molecule, meaning the stabilisation of free triazole is less, leading to a more favourable binding energy of triazole to haem. The relative energies of triazole and water binding to haem are now within 1 kcal mol^{-1} of each other, in line with the previous results without explicit solvent. The bridged-binding mode is now stabilised by 15 kcal mol^{-1} relative to direct binding of triazole, using both functionals, compared to a stabilisation of 11 kcal mol^{-1} over the direct binding mode without explicit solvent.

Table 5.10: Results of the calculations of the binding of water and triazole to haem, using the solvation models 1 and 2, shown in Figure 5.14, with BS2 and the B3LYP and BP86 functionals. Energies are given relative to the starting material shown for each model. The energy of a single water molecule was added where required.

BS2		Relative Energy / kcal mol^{-1}		
		Water	Triazole	Water-Triazole Bridged
B3LYP	Model 1	-5.2	+0.8	-1.7
	Model 2	-5.2	-4.4	-19.3
BP86	Model 1	-6.7	-0.9	-3.0
	Model 2	-6.7	-6.4	-21.2

The results for Model 2 showed that free water and triazole are stabilised by a similar extent when hydrogen bonded to a single water molecule. In Model 1, free triazole formed hydrogen bonds with two water molecules, whereas free water formed only one, leading to a larger stabilisation of triazole by the water solvent than that of water. In reality, the free water molecule would form hydrogen bonds with several (up to four) water molecules, which would be expected to stabilise free water to a greater extent than triazole in solution.

In order to assess the level to which each species is stabilised by the surrounding water solvent, clusters of 9 water molecules, 10 water molecules and a triazole surrounded by 9 water molecules were optimised using B3LYP and BS1. Comparing the clusters to the cluster of 9 water molecules (adding in the energy of the free water or triazole molecule as appropriate) gave a stabilisation of 17.9 kcal mol⁻¹ for the binding of a water molecule to a cluster of 9 waters, and a destabilisation of 34.4 kcal mol⁻¹ for the binding of triazole.

As stated above, there were a large number of degrees of freedom involved in the optimisation of the clusters of water and triazole molecules. Calculations on slightly different conformations gave differences in energy of up to 60 kcal mol⁻¹, the lowest energy conformation in each case being used to calculate the values above. The large difference in energy between binding water and triazole molecules to a cluster of 9 waters could be due to the optimisation of the 9 waters + triazole cluster not finding the lowest energy conformation. Alternatively, the disruption to the hydrogen bonding between water molecules caused by the placing of a triazole in the cluster may be the reason for the unfavourable interaction.

Table 5.11: Relative energies of water and triazole binding to iron-porphyrin, relative to the separated species, with the water or triazole molecules bound to a cluster of 9 water molecules, using B3LYP and BS1.

BS1	Relative Energy /kcal mol ⁻¹		
	Water	Triazole	Water-Triazole Bridged
B3LYP	+4.8	-16.8	-10.1

The relative energies of water and triazole binding, and the triazole-water bridged-binding mode, using the energies from the water cluster calculations are shown in Table 5.11. Using these data triazole binding becomes much more favourable than water binding, with water binding now slightly unfavourable, and 6.7 kcal mol⁻¹ more favourable than bridged-binding via water. The exact values of the relative energies are probably not reliable, due to the high degree of freedom in the conformation of the cluster molecules, as discussed above. However, the qualitative message of the results is clear. When the relative solvation energies of water and azole are included in the calculation,

direct binding of azole to the haem iron becomes significantly more favourable than water binding, or bridged-binding via a water molecule.

It would appear from these data that the relative affinities for the water solvent for water, compared to that for azole, is a large driving factor in the preferential binding of azoles to haem over that of water.

The results shown in Table 5.11 do not include the relative stabilisation of water- and triazole-bound haem by the surrounding water. There are always more calculations that could be carried out. A true representation of the overall binding energies would only be obtained by calculations including either the protein environment or a large quantity of surrounding water molecules, using QM/MM methods. This study is concerned only with the effects of the haem system, and demonstrating that those results are feasible by brief consideration of the environmental effects.

5.2.12 Dissociation calculations

The relative energies of binding of water and the three azoles to haem had been found to be similar. This did not appear to correlate with the experimental observation that azoles readily bind to haem in the presence of water as a solvent, and would therefore be expected to have a higher affinity for haem than that of water. There are other factors involved in the stability of a bound ligand. The ease with which a bound ligand can be removed once it is bound is also a factor. If it could be shown that azoles are more difficult to remove once they are bound, then this could explain why they bind haem iron preferentially to water.

In order to investigate the ease with which azoles and water can be removed from haem, DFT calculations were carried out, starting with the ligand bound and then carrying out further minimisations with the iron-ligand distance fixed at progressively increasing distances. In order to assess the effect of changing spin states on the dissociations, low, intermediate and high spin haem was used with water and imidazole bound, with only low spin used for the other ligands. The calculations were also carried out with the imidazole-water bridged-binding complex.

Table 5.12: Dissociation energies of water- and azole-bound haem at low, intermediate and high spin using BS1 and the B3LYP (top) and BP86 (bottom) functionals. The Fe-O_{water} distance is used for the imidazole-water bridged-binding complex.

B3LYP	Relative energy /kcal mol ⁻¹								
	Fe-O _{water} /Fe-N _{azole} distance /Å	Water			Imidazole			Triazole	Pyridine
		Low	Int	High	Low	Int	High	Low	Low
2.25		-12.9	-6.1	-5.2	-13.2	-3.5	-2.7	-11.7	-11.4
2.5		-11.2	-8.1	-6.7	-9.7	-5.6	-3.4	-8.7	-9.0
2.75		-9.5	-8.3	-6.9	-6.7	-5.0	-3.1	-5.9	-6.4
3		-8.4	-8.2	-7.1	-4.6	-4.3	-2.1	-4.1	-4.5
3.25		-7.6	-7.9	-7.0	-3.2	-3.5	-1.5	-3.1	-3.2
3.5		-6.8	-7.3	-6.7	-2.6	-3.1	-1.2	-2.6	-2.4
3.75		-5.9	-6.5	-6.1	-1.7	-2.5	-0.6	-2.4	-2.0
4		-5.0	-5.7	-5.3	-1.6	-3.3	-1.9	-2.2	-1.7

BP86	Relative energy /kcal mol ⁻¹								
	Fe-O _{water} /Fe-N _{azole} distance /Å	Water			Imidazole			Triazole	Pyridine
		Low	Int	High	Low	Int	High	Low	Low
2.25		-29.8	-15.8	-5.7	-29.7	-12.7	-2.4	-28.1	-28.6
2.5		-27.8	-17.8	-7.4	-25.5	-14.2	-2.9	-24.4	-25.2
2.75		-25.9	-18.1	-7.5	-21.9	-13.6	-2.3	-21.1	-21.9
3		-24.4	-17.9	-7.4	-19.5	-12.6	-1.2	-19.0	-19.5
3.25		-23.3	-17.4	-7.1	-18.0	-11.9	-0.5	-17.9	-18.0
3.5		-22.1	-16.6	-6.5	-17.3	-11.5	-0.2	-17.4	-17.1
3.75		-21.1	-15.6	-5.7	-16.3	-10.9	0.3	-17.1	-16.6
4		-20.1	-14.8	-4.8	-18.1	-11.9	-0.9	-17.1	-16.2

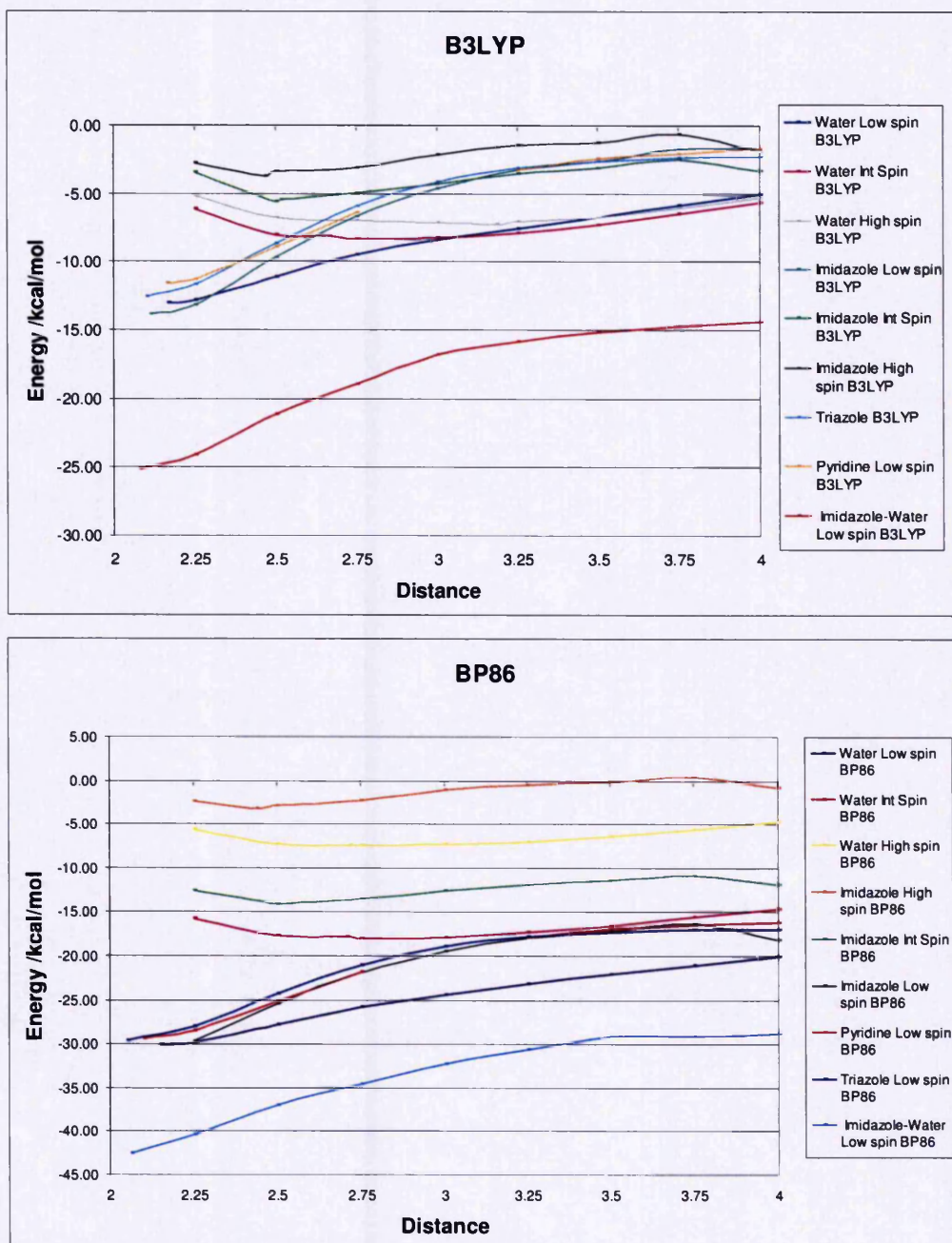


Figure 5.15: Dissociation curves of low, intermediate and high spin water and imidazole, and low spin triazole and pyridine, and the imidazole water bridged-binding complex using BS1 and the B3LYP (top) and BP86 (bottom) functionals. Energies are relative to separate components and distances are in Å.

The results of the dissociation calculations are shown in Table 5.12 and Figure 5.15. It can be seen that although the water and azole curves start at approximately the same relative energy, the gradient of the curve for the removal of water is less steep than that for the removal of the azoles, such that, at a distance of 3 Å, the difference between the water and azole curves is 4 kcal mol⁻¹ with B3LYP, and 5 kcal mol⁻¹ with BP86. Once the water has been removed 3 Å away from the iron, it can be postulated that another molecule, such as imidazole, would be able to bind to the iron and displace the water. If it is 4 or 5 kcal mol⁻¹ easier to remove water 3 Å from haem than one of the azoles, this could explain why azole binding is favoured over water binding, as once they are bound to haem the azoles are more likely to stay bound

5.2.13 Modelling CYP121 Serine-237

The x-ray crystal structure of fluconazole-bound CYP121 solved by Seward *et al.* (Seward *et al.*, 2006) shows the serine residue 237 to form a hydrogen bond with the water molecule bound to the iron, reproduced in Figure 5.16. Seward *et al.* found this residue to be much closer to the haem in CYP121 than for comparable residues in any of the other P450s for which structures are available. It is likely that the ability of this residue to form hydrogen bonds with the bound water is an important factor in the observation of the bridged-binding mode in CYP121.

In order to assess the impact of the presence of Serine-237 close to the haem in CYP121, DFT minimisations were carried out by fixing an ethanol molecule in the same position as Serine-237 in CYP121. The hydrogen atom of the SH group and the ethanol methyl group were fixed in space and the rest of the system allowed to move freely. In order to decrease computing time, each system was first minimised using the 3-21G basis set on all atoms other than iron, and this structure then used as the starting point for minimisation using BS2. The structure obtained for the water-imidazole bridged complex is shown in Figure 5.17.

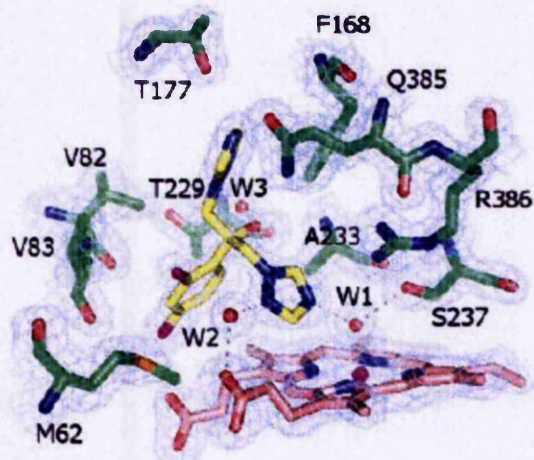


Figure 5.16: Fluconazole-bound CYP121 structure. Reproduced from (Seward *et al.*, 2006), the fluconazole bound CYP121 x-ray crystallography structure, showing Serine-237 forming a hydrogen bond with the bound water molecule.

The results for water and imidazole binding to low, intermediate and high spin haem, using BS2 and the B3LYP and BP86 functionals, are shown in Table 5.13. It can be seen that the presence of the ethanol molecule, fixed in the same position as Serine-237 in CYP121, stabilises the binding of water to low spin haem by 9 kcal mol⁻¹ compared to that of imidazole. The bridged-binding mode is stabilised by between 15 and 20 kcal mol⁻¹, relative to imidazole binding, compared to just 10 kcal mol⁻¹ in the absence of serine. A similar effect is seen with intermediate and high spin haem, with water binding stabilised by around 10 kcal mol⁻¹ compared to that of imidazole throughout.

These results suggest that the Serine-237 residue is indeed important in the stabilisation of the bridged-binding mode, and in the destabilisation of the direct binding mode. Coupled with the steric effects, which prevent the direct bound azole from binding at the ideal 90°, this adds to the postulate of Seward *et al.* that the closeness of the I helix to the haem in CYP121 is the reason for the observation of the bridged-binding mode when it has not been observed in other P450s.

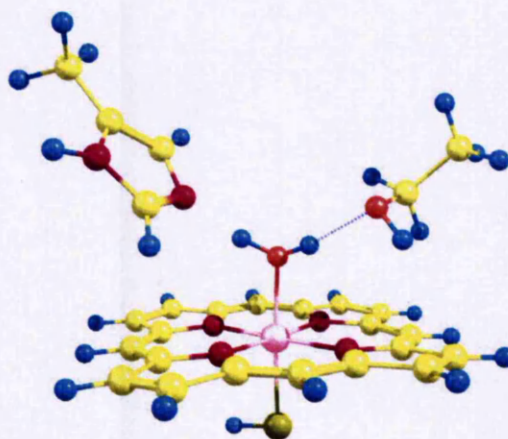


Figure 5.17: Structure of the imidazole-water bridged complex with an ethanol molecule fixed in the position of the Serine-237 residue in CYP121.

Table 5.13: Relative energies of water and imidazole binding to low, intermediate and high spin haem in the presence of an ethanol molecule fixed in the position of the Serine-237 residue in CYP121. Energies are given relative to the separated water, imidazole and haem-ethanol components.

		Relative Energy /kcal mol ⁻¹		
		Water	Imidazole	Water-Imidazole Bridged
B3LYP	Low	-19.0	-10.2	-31.9
	Int	-15.3	-5.9	-24.0
	High	-11.0	-3.7	-21.5
BP86	Low	-38.2	-29.2	-44.8
	Int	-24.9	-14.8	-33.1
	High	-13.6	-3.2	-21.5

5.3 Conclusion

Mycobacterium tuberculosis causes more deaths than any other infectious agent. Multidrug-resistant strains, and the long treatment time with current therapies mean there is an urgent need for the development of new drugs. The genome of *Mtb* genome was solved in 1998 (Cole *et al.*, 1998). This showed that there are 20 P450 enzymes encoded in the *Mtb* genome. Several of these P450s have been shown to bind azole antifungals (McLean *et al.*, 2002b). The structures of the P450s vary greatly. The haem and cysteine residue remain constant through all the P450s, however.

Little work has been done previously to understand the driving force behind azole binding to the P450 haem. In order to gain a greater understanding into how azole antifungals interact with P450s it was decided to investigate the fundamental driving force behind azole binding to P450s. The binding of the three functional groups found in azole antifungals (imidazole, triazole and pyridine) with haem-thiolate was investigated.

The pseudo reaction coordinate method was used to determine the reaction pathway by which the water bound resting state of P450s is displaced by the incoming azole. It was found that following the Fe-N_{azole} pseudo reaction coordinate was not sufficient to provide a smooth pathway from water binding to azole binding. In order to obtain a smooth pathway between the two products, a 2D PES was calculated by fixing both the Fe-N_{azole} distance and the Fe-O_{water} distance.

Using these two pseudo reaction coordinates, a reaction surface was elucidated and transition states were obtained. It was found that there were two minima on the reaction surface. As the azole approached the water-bound haem it first bound to the bound water, giving a bridged-binding complex. It then displaced the water to give a second minimum with the azole bound directly to the iron. The reaction barriers going from bridged bound to direct bound structures were 15.9, 14.8 and 14.9 kcal mol⁻¹ for imidazole, triazole and pyridine, respectively, using the B3LYP functional.

The bridged-binding mode of the three azoles with haem was between 2.0 and 6.6 kcal mol⁻¹ more stable than the corresponding direct binding mode. This result appeared to be contradictory with the experimental observation that azoles bind to P450s in water as solvent, and therefore the direct bound structure would be expected to have the lowest energy. Determination of the binding energy of water and the azoles to haem revealed that it was the additional stability of water binding to haem, over that of azole, that was causing the bridged-binding mode to appear to be more stable.

Various calculations were carried out by varying the basis set and starting structures used to carry out the calculations. These adjustments had very little effect on the results. The inclusion of counterpoise correction energy in the calculations shifted the results slightly so that azole binding became 2.2, 0.7 and 0.3 kcal mol⁻¹ more stable than water binding for imidazole, triazole and pyridine, respectively, using the B3LYP functional.

Including models of the solvation of the free water and azole molecules in the calculations further increased the stability of azole binding. This suggested that part of the driving force behind azole binding to haem is that azoles disrupt the structure of the water they are dissolved in, so it is more favourable for them to bind to haem than to be dissolved in water. In contrast, water binds strongly to other water molecules so the binding of water to haem is less favourable than it appears to be in the gas phase.

Seward *et al* solved the structure of fluconazole-bound CYP121 (Seward *et al.*, 2006). They found a novel bridged-binding mode to be present in the structure. They postulated that the close proximity of the I helix to the haem hindered the direct binding of azoles to the haem, instead favouring binding via a bridging water molecule. In particular, the Serine-237 residue was found to be closer to the haem than corresponding residues in other structures. It was suggested that the serine residue may stabilise the bound water molecule via hydrogen bonding.

Calculations were carried out, fixing an ethanol molecule in the position of the serine residue in the crystal structure in order to assess the impact of the serine on the water and azole binding. It was found that the serine residue both stabilises the bound water molecule and destabilises direct azole binding to the haem. This further explains why the bridged-binding mode is observed in CYP121 but has not been observed in other P450s.

CHAPTER 6: CONCLUSIONS AND FUTURE PROSPECTS

6.1 Introduction

P450s are ubiquitous haemoproteins that are found in every family of organisms, ranging from bacteria and fungi, through to plants, birds and mammals (Groves, 2005). P450s carry out a range of functions, these include the metabolism of drugs and other xenobiotics, metabolism of vitamins, the biosynthesis of critical signalling molecules and the conversion of fatty acids to biologically active molecules (Ortiz de Montellano and de Voss, 2005). P450s catalyse a range of reactions, these are predominantly oxidations and hydroxylations, but dealkylation, dehydrogenation and carbon-carbon bond cleavage reactions have also been observed (Ortiz de Montellano and de Voss, 2005).

In the pathogenic bacterium *Mycobacterium tuberculosis* (*Mtb*), an unexpectedly large number of P450 enzymes were found to be encoded following the determination of the genome sequence of *Mtb* H37Rv (Munro *et al.*, 2003). There is an urgent need for new treatments for *Mtb*, in order to combat multidrug-resistant strains, and to provide shorter treatment times. There are around 9.2 million new cases and 1.7 million deaths from TB worldwide each year, and a third of the world's human population is estimated to be infected with *Mtb* (WHO, 2008). The presence of a CYP51 isoform in *Mtb* H37Rv (CYP51B1) raised the possibility that azole drugs could be useful as novel antitubercular antibiotics, since this P450 is the fungal target for the azoles. Azoles have since proved to be effective in clearing *Mtb* from mice, although the target P450(s) are likely to be other systems than CYP51B1. CYP121, CYP128 and CYP132 are prime candidates, and CYP132 (a virulence-associated *Mtb* P450) was studied as part of the work presented in this thesis.

The P450 catalytic cycle and the mode of P450 binding to inhibitors are two important factors for consideration when developing new treatments targeting P450s. To this end, the P450 catalytic cycle and the mode of binding to azole inhibitors have been studied using DFT calculations. These factors have been investigated experimentally in *Mtb* CYP132, which has also been characterised by a range of spectroscopic techniques.

6.2 Conclusions

6.2.1 The expression and characterisation of *Mtb* CYP132

CYP132 shows significant homology at the amino acid level with various eukaryotic CYP4 family fatty acid hydroxylases. The expression of CYP132 is induced by the gene regulator protein encoded by *Rv1395*, and it was concluded that CYP132 was a bacterial virulence associated P450 protein (Recchi *et al.*, 2003). Gene knockout studies by Camacho *et al.* (Camacho *et al.*, 1999) revealed that when *Rv1395* is not present in *Mtb*, the growth of the bacterium was slower. These factors made CYP132 a good candidate as a potential drug target for treatment of *Mtb*. Data reported in this thesis are the first relating to the expression and characterisation of *Mtb* CYP132.

Expression and purification protocols were developed for *Mtb* CYP132. It was found that the major product of expression of the *CYP132* gene was a truncated form of CYP132, with mass 43.1 kDa, compared with a predicted mass of 53.4 kDa. This protein was found to have been proteolytically cleaved at the N-terminal. A small amount of intact CYP132 was also produced. Attempts to express GST-tagged CYP132, in order to protect the N-terminal from proteases, from the *Rv1394c*/pET42a plasmid, did not produce detectable quantities of material with P450 character.

Binding of CO to reduced, intact CYP132 was found to give a shift in the Soret band in the UV/Vis spectrum, from 418 nm to 449 nm. This confirmed the P450 character of CYP132. Cleaved CYP132 was found to give a shift from 416 nm to predominantly 418 nm, with a small feature at 446 nm. This showed that the cysteine ligation to the haem is retained in cleaved CYP132, but that the proteolysed enzyme is more prone to protonation of the S-Fe bond (at least in the ferrous-CO form) than is the full length enzyme.

Features in the UV/Vis spectra of cleaved CYP132 suggested that it may have been co-purified bound to a substrate like molecule. Attempts to extract any bound material and analyse it by mass spectrometry did not meet with success. Also, attempts to remove the features indicative of substrate binding in the UV/Vis spectrum by removal of any bound material with solvents were not successful, and the spectral signal was restored on solvent removal. It was concluded that the features indicative of substrate binding

observed in the UV/Vis spectra were due to the environment of the heme and its cysteinate ligand, and not to a bound substrate co-purified with the P450. That is, a small amount of high-spin heme iron was likely a consequence of incomplete coordination of heme iron by a water molecule in the 6th position in the cleaved CYP132 enzyme, rather than being due to bound lipid.

Cleaved CYP132 was found to bind tightly a range of azole anti-fungal drugs. Titrations with fatty acids did not result in a type I shift in the UV/Vis spectrum, however, which would be indicative of binding to the P450 active site and the removal of the water ligand to the haem iron. Kinetics studies, using a heterologous redox partner system from *E. coli*, revealed that cleaved CYP132 is able to oxidise unsaturated fatty acids, with a rate of up to 13 fatty acid turnovers per CYP132 per minute. Analysis of the oxidation products by LC/MS revealed that cleaved CYP132 is able to hydroxylate unsaturated fatty acids, oxidising in at least one position, and may also catalyse epoxidation. In addition, the LC/MS/MS results suggested that CYP132 may also catalyse C-C bond cleavage at the ω -2 and ω -3 positions, an activity that likely requires three successive oxidation reactions at adjacent carbon atoms on the substrate.

The expression of CYP132 in *E. coli* cell strains was found to not be trivial. Using the pET15b vector, predominantly cleaved CYP132 was produced, and with the pET42a vector no significant level of P450 expression was detected. Both the intact and cleaved CYP132 forms expressed from the *Rv1394c*/pET15b plasmid were found to have P450 character. This raises the question whether the cleavage of CYP132 could be a natural process, and one that also occurs naturally in *Mtb*. This is perhaps unlikely, but in the light of the ability of cleaved CYP132 to oxidise fatty acids, it is worth considering. In this respect, it is notable that CYP132 is one of the largest of the *Mtb* P450 cohort, and it is thus feasible that processing of a pre-protein is involved in developing a mature form of CYP132. However, further work would be required to confirm such a theory.

An alternative method that might produce higher levels of intact CYP132 is to use an alternative affinity purification method, such as immobilised lipids. If near-total purity could be obtained in a single step, this would reduce the time available for proteases to cleave CYP132. If the site of the proteolytic cleavage could be identified by N-terminal sequencing of cleaved CYP132, a mutation could be produced to remove this cleavage

site. As long as the mutation contains amino acids with similar properties to wild type CYP132, the structure of CYP132 should not be significantly altered.

It has not yet been conclusively determined whether cleaved CYP132 is co-purified with a substrate like material. Attempts to remove any such material did not meet with success, but this might be due to lack of solubility of the bound material in the solvents used, or to the isolation of only undetectably low quantities of lipid. The UV/Vis spectra of full length CYP132 did not contain features indicative of a bound substrate (i.e. any significant high-spin character or long wavelength absorption feature). This further suggests that the features observed in the UV/Vis spectrum of cleaved CYP132 resulted from structural changes in the heme coordination sphere due to the cleavage, and not to a bound substrate molecule. Alternatively, the cleaved CYP132 could have a rather different substrate specificity profile to the full length CYP132, resulting from structural perturbations in the active site and causing it to bind a different type of substrate available in the *E. coli* cells (e.g. a long chain fatty acid) that does not bind to intact CYP132 (or does not bind with comparable affinity). Another possible, but perhaps less likely, alternative is that the binding of a substrate triggers a structural change that causes CYP132 to be more prone to proteolytic cleavage. Structural changes associated with substrate binding are well established in many P450 isoforms. It is difficult to envision a situation where this would be beneficial to the organism, however.

The binding of cleaved CYP132 to azole anti-fungal drugs confirms CYP132 as another potential target for the azole drugs in *Mtb*. However, it remains unclear as to whether or not there are multiple P450 azole targets in the bacterium. It is likely that any azole-based treatment developed for *Mtb* would result in binding to many P450 targets. It is likely that only a small number of these P450s would be essential for bacterial viability, but inhibition of others could contribute to azole-induced cell death through perturbations to important metabolic processes. This multiple-targeting should be advantageous as it would reduce the incidence of drug resistance. However, any such strategy would also require the avoidance of toxic interactions with host P450s and other key proteins.

The LC/MS of the turnover products of cleaved CYP132 with unsaturated fatty acids indicated that CYP132 can carry out multiple reactions on a single substrate, including hydroxylation, possibly at multiple positions along the fatty acid chain, epoxidation

across the unsaturated linkage, and perhaps C-C cleavage at the ω -2 and ω -3 positions. This suggests that CYP132 is probably involved in fatty acid metabolism in *Mtb*. Lipid metabolism is pivotal in *Mtb* biochemistry, and an extraordinarily large amount of the genome is given over to proteins involved in lipid metabolism. The *Mtb* envelope structure contains large amount of highly complex, long chain lipids. It is thus feasible that CYP132 participates in important biosynthetic processes, or perhaps (given its association with *Mtb* virulence) in oxidation of host lipids or cell signalling molecules. Further investigation is required in order to determine the exact nature of the reactions carried out by cleaved CYP132 on unsaturated fatty acids.

6.2.2 Investigation of the P450 reaction cycle by DFT methods

P450s carry out the oxidation of C-H bonds with apparent ease. This reaction is extremely difficult to achieve with standard laboratory methods. There is a high level of variability between P450 structures, particularly with respect to their substrate specificity, although the general fold is conserved. The common factor in all P450s is the haem and its cysteinate coordination. In order to understand the basic driving force behind all P450 reactions, DFT calculations were carried out on the P450 haem for each species in the catalytic cycle.

A number of spin states may be involved in P450 catalysis. It is important not to overlook the alternative spin states for the various species in the P450 catalytic cycle. In many cases, the difference in energy between the spin states is small, in the order of a few kcal mol⁻¹, allowing more than one spin state of a particular species in the catalytic cycle to be involved in the reaction. Effects of the protein environment may even alter the ground state of a particular species.

The water bound species was found to have a low spin ground state, independent of the protein environment. The intermediate and high spin states were found to have similar energies. This raises the question whether the spectroscopic features commonly attributed to the high spin state may in fact be due to a mixture of sextet and quartet spins. There is no evidence for the presence of the quartet spin state by EPR, however. Mixtures of quartet and sextet spin states have been observed in *Chromatium*

cytochrome *c* (Fujiia *et al.*, 1995), indicating that if the quartet species was present it would be detected.

The ground state of the substrate-bound haem was found to depend on the protein environment, with similar energies for all spin states for the penta-coordinated ferric haem without protein present. Again, there is little apparent reason why the quartet spin state should not be observed along with the doublet and sextet states.

Spin crossing is required for the binding of quintet spin penta-coordinated ferrous haem, **3c**, to triplet spin dioxygen. A number of spin states with similar energies were found, suggesting the P450 haem aids spin crossing, as has been found for the binding of oxygen to myoglobin (Jensen and Ryde, 2004). The binding of oxygen to ferrous haem was found to be energetically slightly unfavourable. The protein environment would be expected to stabilise oxygen binding, and the extremely high rates at which the rest of the catalytic cycle takes place would keep any equilibrium between bound and unbound oxygen firmly on the side of oxygen binding.

Compound I was found to have doublet and quartet spin states with identical energies. This shows the highly versatile nature of Compound I, and may explain observations of mixed stereospecificities of some P450 reactions, as has been discussed at length by Shaik *et al.* (Hirao *et al.*, 2005). Compound I was found to be a more likely candidate than Compound 0 as the true oxidant species in the catalytic cycle. Compound 0 was considered to be too stable to carry out the powerful oxidation reaction observed in P450s.

The P450 reaction cycle is a delicate blend of varying spin states and haem structures. Small changes in the environment, or method used to observe it, can lead to changes in electronic ground states, and associated changes in haem structure. Finely tuned protein structures can harness the extreme oxidising power of Compound I, to give oxidation reactions with great specificity. Varying types of P450 structures allow a wide variety of toxins and xenobiotics to be metabolised and safely excreted, as well as for metabolically important transformations (e.g. on fatty acids and steroids) to occur.

6.2.3 Analysis of azole binding in P450s by DFT methods

P450s are the target for a range of azole-type inhibitors. The *Mtb* P450s CYP51B1 and CYP121 have been shown to bind tightly to azole inhibitors (McLean *et al.*, 2002b), as does CYP132, as demonstrated in this thesis. The azoles econazole, miconazole and clotrimazole were found to inhibit the growth of *Mycobacterium smegmatis* (McLean *et al.*, 2002b) and econazole was found to be effective in clearing *Mtb* from the lungs of mice (Ahmad *et al.*, 2006). The binding mode of azoles to P450s has not been studied by DFT methods previously. In order to investigate the driving force behind azole binding to P450s, the binding of the azole-type functional groups imidazole, triazole and pyridine to the P450 haem were studied using DFT.

A reaction surface for the displacement of water from the P450 haem by azoles was calculated by fixing Fe-N_{azole} and Fe-O_{water} distances over a range of distances. These revealed that, on approaching the water-bound haem, the azole first forms a hydrogen bond with the bound water molecule. The bound water then dissociates from the haem, with a transition state containing Fe-N_{azole} and Fe-O_{water} distances of between 3 and 4 Å, before the azole binds to the haem.

The gas phase binding energies of water and the azoles for the P450 haem were found to be similar, with water showing a slightly higher affinity for haem than the azoles. This trend was reversed slightly when a counterpoise correction was used, in order to account for the BSSE.

The bound water molecule would be expected to have a higher affinity for bulk water than the incoming azole. This would push the equilibrium in the direction of azole binding. The calculations show that the binding of azoles to P450 haem is not due to an inherent higher affinity of azoles for haem over that of water, but rather to the greater affinity of the bound water molecule for the bulk water solvent, compared to the incoming azole, the hydrophobic regions of which make binding to bulk water unfavourable.

An x-ray crystal structure of CYP121 bound to fluconazole (Seward *et al.*, 2006) revealed both direct binding of fluconazole to haem iron, and bridged-binding via a water molecule to the haem. DFT calculations on the direct binding and bridged-binding modes, with an ethanol molecule fixed in the position of CYP121 Serine-237, revealed

that the serine residue close to the haem iron both stabilises the binding of water, and destabilises direct azole binding to the haem.

P450s are important drug targets in conditions such as TB. P450s also play crucial roles in human biology. With the development of novel treatments involving azole drugs, it is vital to understand the basic driving force behind azole binding to P450s. This study has revealed how azoles first bind to the bound water molecule, before the water dissociates and the azole binds to the iron. Also, the driving force behind azole binding to iron has been revealed to be only partially due to inherent affinity of azoles for haem, over that of water. Rather, the increased affinity of the bound water molecule for bulk water, over that of azole, also plays an important role.

6.3 Future work

6.3.1 The expression and characterisation of Mtb CYP132

In the expression and characterisation of CYP132 it was found that predominantly cleaved CYP132 had been produced. The first priority of any future work on CYP132 is to express and purify useable quantities of intact (full length) CYP132. A possible approach to obtain full length CYP132 would be to speed up the purification process, by achieving near total purity in a single step. This would reduce the time available for proteases to proteolytically cleave the protein. Other possibilities would be to ascertain the position of the proteolytic cleavage by N-terminal sequencing, and mutate one or two residues at the point of cleavage. Also, shorter growth times would reduce the time available for cleavage, although this could potentially decrease the amount of protein expressed. However, using a powerful T7 polymerase/promoter system, an expression regime combining high production and short induction time should be feasible.

Cleaved CYP132 was found to oxidise unsaturated fatty acids. Further investigation is required to fully characterise the nature of the reactions carried out. Analysis of the palmitoleic acid oxidation products that eluted from the HPLC column at short timescales by mass spectroscopy, and comparison with the masses associated with the similar peaks on the LC chromatogram for myristoleic acid, would confirm whether or not these were oxidation products of the unsaturated fatty acids, resulting from C-C bond cleavage

reactions, or in some way artefacts of the methods used. However, regardless of the requirement to study in more detail the spectrum of products resulting from CYP132's turnover with unsaturated fatty acids, it appears clear that at least some epoxidation and hydroxylation products are formed.

If sufficient quantities of fatty acid could be oxidised, NMR could be used to further identify the reaction products of the oxidation of unsaturated fatty acids by CYP132. This would further confirm that hydroxylation, epoxidation, and/or another reaction types had occurred, and the position of any hydroxylation reactions. Establishing the chirality of oxidation products should also be a priority of future studies in this area.

Further optimisation of the turnover of fatty acids by CYP132 could be attempted using homologous redox partners from *Mtb*, such as the flavin-containing, NAD(P)H-dependent FprA and the 3Fe-4S ferredoxin Fdx, the gene encoding the latter protein found adjacent to the *CYP51B1* gene in the *Mtb* genome. The rate of reaction with these homologous redox partners (and the nature of products formed) could then be compared to the same reaction carried out using the heterologous redox partners, or to the parameters/products obtained using a mixture of heterologous/homologous partners. Such studies would doubtless shed further light on the catalytic capabilities of CYP132 and its preferred type(s) of redox partner system. Obviously, a further long term goal is also the crystallisation and solution of the crystal structure of the CYP132 protein in ligand-free, azole-bound and substrate-bound forms. This work would be essential to rationalise its catalytic activities and to enable optimal inhibitor design.

CYP132 was established as being important to the growth of *Mtb* by gene knockout studies on the *Rv1395* gene, which encodes an AraC/XylS-type gene regulator for CYP132. The essentiality of CYP132 to *Mtb* growth and virulence could further be established by gene knockout studies on *Rv1394c* itself. Examination of the lipid profiles of *Mtb* with, and without, the *CYP132* gene present (assuming CYP132 is not essential for growth *per se*) would give a much clearer idea of the physiological activity of CYP132, and would suggest possible substrates, which could then be tested for binding and turnover by CYP132 together with homologous/heterologous redox partner systems. Collectively, this programme of future study would enable a much clearer picture of the

structure/function and physiological/pathogenic importance of *Mtb* CYP132 to be developed.

6.3.2 The P450 reaction cycle

Since the DFT calculations on the P450 catalytic cycle detailed in this thesis were begun, many further calculations on the P450 catalytic cycle have been carried out, including analyses using DFT and QM/MM methods. There still remain various questions unanswered, however, which warrant further study.

The origin of the high spin ground state of substrate-bound P450 remains unclear. The spin state of penta-coordinated ferric haem that is calculated as the ground state depends on the method used for the calculations, and also on the protein environment. Some P450s, such as cleaved CYP132, do not appear to give a shift to the high spin state upon substrate binding (assuming that unsaturated fatty acids are good substrates in this case), further suggesting that the ground state depends on the protein environment. In this case, these experimental observations are more likely to be due to the substrate not displacing the water molecule bound as the 6th haem ligand, rather than the penta-coordinated ferric haem adopting a low spin ground state.

The binding of dioxygen to the P450 haem requires spin crossing. It is likely that the P450 haem aids spin crossing in a similar manner to that found for the myoglobin haem (Jensen and Ryde, 2004). Further study is required to confirm the role of the P450 haem in spin crossing upon oxygen binding.

The protonation events in the catalytic cycle depend on the protein environment. Some QM/MM studies have been carried out, mainly on P450cam, but little work has been done on characterisation of proton sources in other P450s. Accurate determination of proton sources in P450s would in turn allow the relative energies of Compound I and Compound 0 to be calculated. This would aid in the determination of the true oxidant species in the P450 catalytic cycle for each P450 examined.

Much work has been done on Compound I in P450s, yet there is still debate over the true identity of the oxidant species in the P450 catalytic cycle. The majority of the studies

involving analysis of the protein environment have been carried out on P450cam, so it is not really known whether the oxidant species in other P450s is the same as in P450cam. Also, the electronic structure of Compound I is likely to vary in other P450s, as the protein environment around the haem is altered. Investigation of Compound I in other P450s would give an idea of the consistency of the Compound I electronic structure across the range of P450s, and this is clearly a priority area for research aimed at determining more completely the relevance of different oxidants in oxidative catalysis in different P450 frameworks.

6.3.3 Azole binding to P450s

The driving forces behind azole binding to P450 haem have been investigated, and these were found to depend only partially on the greater affinity of the haem for azoles rather than water. The protein environment is likely to play an important role in this process. QM/MM studies on the binding of azoles to a range of P450s, including CYP121 and other *Mtb* P450s, would give a clearer idea of the driving force behind azole binding to P450s. The larger azoles, such as clotrimazole and econazole, which have specificity for P450s, will have much greater affinity for the protein active site than for small model azoles (e.g. imidazole) as a result of multiple hydrogen bonding and other interactions with residues in the active site. The driving force behind the binding of these azoles would be expected to be almost entirely protein-driven, and thus will depend much less on relative affinities for bulk water.

The mode of binding found in this study could have important implications for the design of azole drugs targeting P450s. It would appear to be necessary to have enough space in the P450 active site for the azole and water complex to move away from the haem, before binding of the azole to the haem iron. QM/MM studies of azole drugs binding to, and displacing water, from P450 structures, such as CYP51B1 and CYP121, would give a clearer insight into the importance of having enough room in the active site for this process, and whether azole binding is diminished in structures that require the bound water to dissociate away from the haem iron, before azoles can bind. Studies on the dissociation of water and azole from haem in this study revealed that azoles are less likely to dissociate once bound (i.e. have a low k_{off} value), so the rate of binding of azoles

(i.e. the k_{on} value) may not be a particularly important factor in the affinity of azoles for P450s.

This study showed how the bridged-binding mode of azoles to P450s can occur if constraints in the protein structure are present. Studies by Seward *et al.* (Seward *et al.*, 2006) showed that the structural characteristics that stabilise bridged-binding in CYP121 are unique among the P450s for which structures are available to date, but it is likely that there are other P450s with similar structural constraints and for which the structures have not yet been characterised. The bridged-binding mode is an important consideration if designing azole inhibitors targeting CYP121, and will also be important if P450s with similar structural constraints are revealed through subsequent P450 structural studies.

6.4 Concluding remarks

In this thesis the catalytic cycle and binding of azole inhibitors to P450s have been discussed, with particular reference to *Mtb* P450s, and especially relating to *Mtb* CYP132. CYP132 was expressed, purified and characterised. It was found to have been proteolytically cleaved during expression and purification, but was still found to bind tightly to various azole anti-fungal drugs, and to oxidise unsaturated fatty acids. Cleaved CYP132 was found to carry out a range of reactions on unsaturated fatty acids when provided with electrons from NADPH and delivered using an *E. coli* class I-like P450 redox system. Preliminary analysis of the reaction products suggested them to include hydroxylation, epoxidation and C-C bond cleavage products.

The P450 catalytic cycle was studied using DFT calculations on the P450 haem. The electronic ground state upon substrate binding was found to depend on the protein environment, and the computation methods used. The quartet spin state was found to be much more prominent than experimental observations would suggest. The binding of oxygen to haem requires spin crossing. A number of spin states with similar energies were found for oxygen-bound ferric haem. This suggested that the P450 haem plays an important role in spin crossing in the binding of oxygen. P450 Compound I was found to contain two spin states of equal energy. This indicates the versatile nature of Compound I, as both spin states are likely to be involved in oxidation of bound substrates.

DFT studies on azole binding to P450 haem revealed a bridged-binding intermediate via water, and a transition state with both water and azole removed from the haem. Studies on the active site of CYP121 revealed how structural elements in the active site stabilise the binding of water, and destabilise the direct binding of azoles, causing the observation of the bridged-binding mode in the fluconazole-bound CYP121 crystal structure.

Since their discovery in the 1930s, 40s and 50s, a vast amount of knowledge about P450s has been obtained, with thousands of P450-encoding genes being discovered in nature. From the metabolism of drugs and other xenobiotics in humans, to the synthesis of important biomolecules in bacteria, P450s play a whole host of different roles, in every family of organisms. With more P450s being discovered seemingly every day, revealing exciting new structures and functionalities, P450s can be expected to be an important area of research for many years to come.

**APPENDIX: Paper Submitted to J. Phys. Chem. For
Publication**

How Do Azoles Inhibit Cytochrome P450 Enzymes? A Density Functional Study[†]Philip R. Balding,^{*,§} Cristina S. Porro,^{*,§} Kirsty J. McLean,^{*,||} Michael J. Sutcliffe,^{*,§} Jean-Didier Maréchal,^{*,§,||} Andrew W. Munro,^{*,*,||} and Sam P. de Visser^{*,*,§}

Manchester Interdisciplinary Biocentre, School of Chemical Engineering and Analytical Science, and Faculty of Life Sciences, The University of Manchester, 131 Princess Street, Manchester M1 7DN, United Kingdom

Received: March 10, 2008; Revised Manuscript Received: April 3, 2008

To examine how azole inhibitors interact with the heme active site of the cytochrome P450 enzymes, we have performed a series of density functional theory studies on azole binding. These are the first density functional studies on azole interactions with a heme center and give fundamental insight into how azoles inhibit the catalytic function of P450 enzymes. Since azoles come in many varieties, we tested three typical azole motifs representing a broad range of azole and azole-type inhibitors: methylimidazolate, methyltriazolate, and pyridine. These structural motifs represent typical azoles, such as econazole, fluconazole, and metyrapone. The calculations show that azole binding is a stepwise mechanism whereby first the water molecule from the resting state of P450 is released from the sixth binding site of the heme to create a pentacoordinated active site followed by coordination of the azole nitrogen to the heme iron. This process leads to the breaking of a hydrogen bond between the resting state water molecule and the approaching inhibitor molecule. Although, formally, the water molecule is released in the first step of the reaction mechanism and a pentacoordinated heme is created, this does not lead to an observed spin state crossing. Thus, we show that release of a water molecule from the resting state of P450 enzymes to create a pentacoordinated heme will lead to a doublet to quartet spin state crossing at an Fe–OH₂ distance of approximately 3.0 Å, while the azole substitution process takes place at shorter distances. Azoles bind heme with significantly stronger binding energies than a water molecule, so that these inhibitors block the catalytic cycle of the enzyme and prevent oxygen binding and the catalysis of substrate oxidation. Perturbations within the active site (e.g., a polarized environment) have little effect on the relative energies of azole binding. Studies with an extra hydrogen-bonded ethanol molecule in the model, mimicking the active site of the CYP121 P450, show that the resting state and azole binding structures are close in energy, which may lead to chemical equilibrium between the two structures, as indeed observed with recent protein structural studies that have demonstrated two distinct azole binding mechanisms to P450 heme.

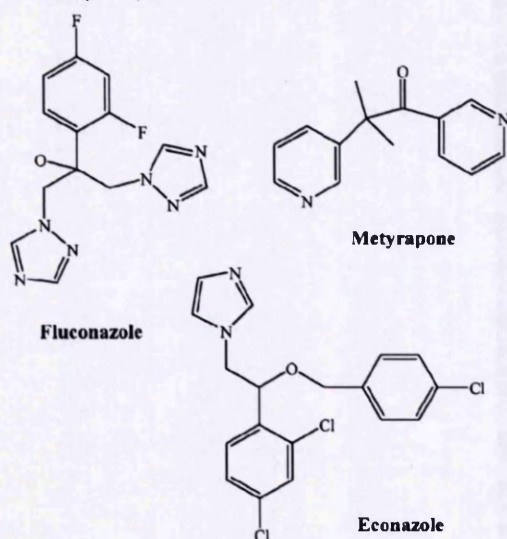
Introduction

The cytochromes P450 (P450s) are a ubiquitous family of enzymes found in a very broad range of organisms, from bacteria and fungi through to plants and mammals.¹ They catalyze a large variety of reactions, of which the most common are oxidations and hydroxylations. Other P450 reactions including reduction, desaturation, ester cleavage, ring expansion, dehydration, and one-electron oxidation have also been reported.² P450s in nature have a large number of important biochemical functions. A key role is protection, for example, in the metabolism of xenobiotics.³ A major example of this in humans is the enzyme CYP3A4, which is the most abundant P450 in the human body and contributes to the metabolism of around half of all drugs in use today.⁴ Another role of P450s is in the biosynthesis of important biomolecules, for use as, for example, signal molecules in control of development and homeostasis and regulation of physiological processes.⁵ P450s are important drug targets in fungal infections⁶ and are, for example,

responsible for insecticide resistance in the mosquito.⁷ Furthermore, multiple P450s are present in the genome of *Mycobacterium tuberculosis*, the pathogenic bacterium responsible for the disease tuberculosis.⁸ In this particular case, azole- and triazole-based P450 inhibitors were shown to coordinate tightly to P450 enzymes from the bacterium (including CYP121 and CYP51B1) and to inhibit mycobacterial growth.^{9,10}

Many drug-metabolizing P450s are located in the liver and can inadvertently be inhibited by treatments given to the patient for specific medical conditions. Alternatively, P450-mediated oxidation of certain drugs can alter their biological activity and subsequently lead to potentially harmful drug–drug interactions.¹¹ Due to this, several prominent drugs have been withdrawn from the market, such as clozapine and fluoxetine, as well as triazolam and amitriptyline.¹² There are several reversible inhibitors with azole functional groups that target the P450s. One of the first P450 inhibitors to be widely employed was metyrapone, which targeted the 11 β -hydroxylase P450 (CYP11B1) in the treatment of Cushing's syndrome.¹³ Scheme 1 shows chemical structures of metyrapone, fluconazole, and econazole, all of which are regarded as azole inhibitors. Metyrapone has a pyridine functional group, rather than an imidazole or triazole, but due to structural and functional similarities to the azole class, we will include these types of inhibitors here. Fluconazole is a systemically tolerated azole

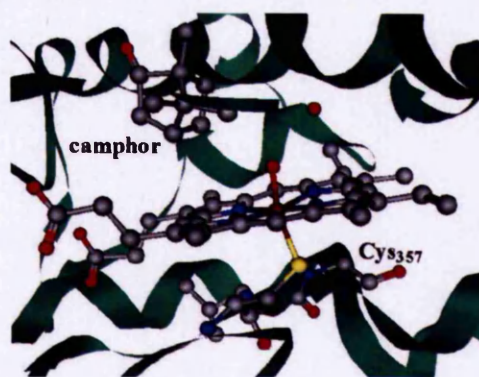
[†] Part of the "Sason S. Shaik Festschrift".^{*} To whom correspondence should be addressed. E-mail: sam.devisser@manchester.ac.uk (S.P.V.); andrew.munro@manchester.ac.uk (A.W.M.).[†] Manchester Interdisciplinary Biocentre.[§] School of Chemical Engineering and Analytical Science.^{||} Faculty of Life Sciences.^{||} Present address: Chemistry Department, University Autonomas of Barcelona, 08193 Bellaterra, Catalonia, Spain.

SCHEME 1: Examples of Three Typical Azole Inhibitors of P450 Enzymes, As Studied in This Work

widely used to inhibit the sterol demethylase CYP51 and to treat *Candidiasis*,¹⁴ while econazole is a topical treatment for yeast and fungal infections and was recently shown to be effective in clearing *M. tuberculosis* from infected mice.¹⁵

The binding of azole drugs to P450s has been extensively studied. As early as 1967, Estabrook et al.¹⁶ reported type-II spectral shifts arising from the binding of nitrogen-containing compounds, such as pyridine and nicotine, to human liver P450s. Van den Bossche postulated that the target of the azole antifungal drugs was the P450 sterol 14 α -demethylase, later to be called CYP51.¹⁷ Schuster studied the effect of azoles on human P450s, concluding that the metabolism of xenobiotic substrates would be impeded by azole drugs.¹⁸ Since then, there have been innumerable reports of the binding of azole antifungals to P450s, including many studies on drug–drug interactions involving azole binding to CYP3A4¹⁹ and a number of studies of the binding of azole antifungals to *Mycobacterium tuberculosis* (Mtb) P450s.²⁰ The first crystal structures of azole-bound P450s were for the fluconazole and phenylimidazole complexes of Mtb CYP51B1, and these showed a nitrogen atom of fluconazole or 4-phenylimidazole coordinated to the heme iron.²¹ The inhibitor binds directly to the iron and forms an Fe–N bond. The bulky fluconazole induced structural changes in the active site and in putative ligand entry/exit regions of the P450. Crystal structures have also been solved for the ketoconazole complex with *Saccharopolyspora erythraea* P450_{eryF}, for the *Pseudomonas putida* camphor hydroxylase P450_{cam} in complex with imidazole, and for human CYP3A4 bound to ketoconazole.²² Recently, a high-resolution crystal structure was solved for the *M. tuberculosis* CYP121 bound to fluconazole.²³ Nitric oxide synthase, an oxidase enzyme with heme coordination similar to that of the P450s, is also inhibited by imidazole, which binds to form a low-spin complex.^{24,25}

The active center of P450 enzymes contains a central heme moiety that is linked to the protein backbone via an Fe–S linkage from a cysteinate residue.^{1,26} Scheme 2 shows the active site of a typical P450 isozyme, namely P450_{cam} (CYP101A1) as taken from the 1DZ9 pdb file.²⁷ This axial cysteinate ligand exerts a “push” effect on the iron center that has been correlated with the monooxygenase capacity of the enzyme.²⁸ On the distal side of the heme, the sixth ligand site on the ferric iron is

SCHEME 2: Active Site Structure of the Camphor-Bound Form of P450_{cam}^a

^a Image was created using MOE and using the atomic coordinates from the 1DZ9 pdb file.

occupied by a water molecule in the resting state. When the substrate enters the active site pocket, this triggers the catalytic cycle that starts with displacement of the water molecule, a shift of the heme iron spin state equilibrium towards high spin, a reduction step, binding of molecular oxygen, a further reduction, and two protonation steps to create an oxo-iron active species.¹ Electrons are delivered via the redox partner iron–sulfur protein putidaredoxin. P450_{cam} regioselectively hydroxylates camphor at the C5 position and has long been used as a model for the structure and function of the P450 class of enzymes. Another structurally and biotechnologically important member of the P450 superfamily is P450_{BM3} (CYP102A1) from *Bacillus megaterium*, in which the P450 is fused to a eukaryotic-like redox partner enzyme—a FAD—and FMN-containing cytochrome P450 reductase (CPR). This makes for a highly efficient electron-transport system, and P450_{BM3} hydroxylates its long-chain fatty acid substrates at rates faster than those of any other known P450 oxidase.²⁹

Azole inhibitors occupy the active site space designed for the P450 substrates and bind directly to the heme center, thereby rendering it inactive to oxygen chemistry. There is little available literature on theoretical studies of the binding of azoles to P450 heme. Many azoles contain an imidazole functional group that is expected to bind to the iron center. Imidazole is often used in theoretical studies to model histidine-bound hemes, such as peroxidases, although in those cases, it is located in the proximal axial ligand position occupied by the cysteinate in P450s.³⁰ In order to gain insight into how azoles bind to P450 enzymes, we have performed a series of density functional theory (DFT) studies. The questions which will be addressed in this work are: why is it that these inhibitors bind so well to the active center and what determines the efficiency of binding? In order to answer these questions, we have performed a series of DFT studies into the mechanisms of azole binding to the heme group of P450 enzymes and have used three typical functional groups representing most azole or azole-like inhibitors.

Methods

All calculations were performed using well-established procedures in the field.³¹ We studied the binding of azole models to the active center of cytochrome P450 enzymes. Initial calculations in the gas phase used a minimal model of the resting state of P450 with iron inside of protoporphyrin IX that is ligated to a thiolate axial (proximal) ligand and a water molecule on the distal site. Subsequently, we added an azole mimic and

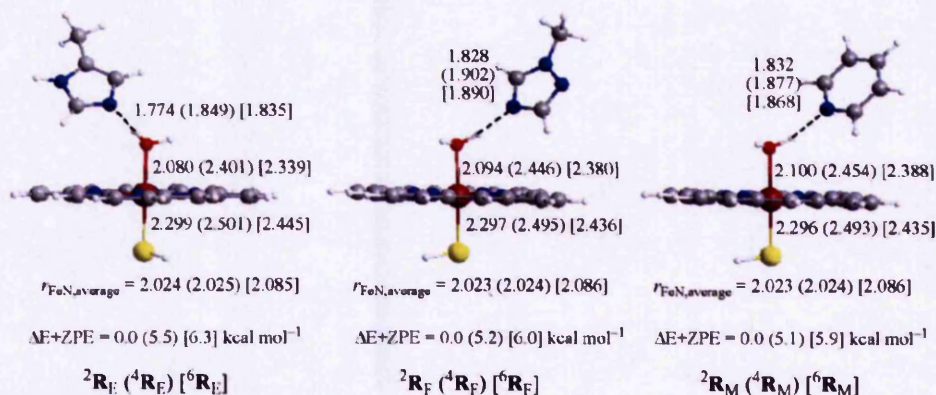


Figure 1. Optimized geometries of \mathbf{R}_E , \mathbf{R}_F , and \mathbf{R}_M resting state structures in the doublet (quartet) [sextet] spin states. All bond lengths are in Å, and relative energies are in kcal mol⁻¹.

studied the replacement of the water molecule by the azole. Detailed geometry scans with one degree of freedom fixed, namely the distance between the metal and the atom of the azole that binds to iron (Fe–N distance), were performed in Gaussian 03.³² This procedure, however, did not displace the water molecule from the sixth coordination position of the heme and led to a high-energy pathway. Therefore, an alternative two-dimensional scan was performed where the Fe–OH₂ and Fe–N_{azole} distances were simultaneously scanned. These two-dimensional scans were the result of a full geometry optimization with two degrees of freedom constrained (the Fe–O and Fe–N_{azole} distances).

All calculations described in this work utilize the unrestricted hybrid density functional method UB3LYP (Method M1).³³ Detailed comparative studies on iron porphyrin systems with various density functional methods show that, sometimes, deviations in spin state ordering are observed.^{30c} Nevertheless, test calculations (see Supporting Information) using the BP86 functional (Method M2) gave virtually the same results as the ones described in this paper. However, for economy of space, these data are relegated to the Supporting Information.³⁴ We used a Los Alamos type double- ζ quality basis set that contains a core potential on iron (LANL2DZ), while oxygen and nitrogen are described with 6-31G* and all other atoms with the 6-31G: basis set BS1.³⁵ Subsequent single-point calculations with a triple- ζ -type LACV3P+ basis set on iron and 6-311+G* on the rest of the atoms (basis set BS2) were performed in Jaguar 7.0.³⁶ Further test calculations (see Supporting Information) with larger/alternative basis sets showed minimal changes in the binding energies of the various substrates. For all local minima, full optimizations (without constraints) followed by an analytical frequency calculation were performed in Gaussian 03. All local minima described here had real frequencies only, while the transition states were characterized by a single imaginary frequency for the correct mode. The local minima were calculated in the lowest-lying doublet, quartet, and sextet spin states. However, since the doublet spin state was generally the lowest-lying state, we only performed geometry scans for the doublet spin state.

The three inhibitors econazole, fluconazole, and metyrapone were abbreviated to methylimidazolate, methyltriazolate, and pyridine, respectively.

To test the effect of the environment on the azole binding affinities, we tested three different perturbations: (i) by applying a dielectric constant of $\epsilon = 5.7$ to the system, (ii) with the addition of two hydrogen-bonded ammonia molecules mimicking the peptide environment in the vicinity of the cysteine

ligand, and (iii) a combination of both effects. The calculations with a dielectric constant were performed with the self-consistent reaction field model as implemented in Jaguar 7.0 with a probe radius of 2.7 Å.³⁶ Previous studies on P450 systems showed that hydrogen bonding toward the cysteine ligand changes the electronegativity of the sulfur atom and thereby influences the Fe–S bonding interactions strongly.³⁷ These hydrogen-bonding interactions were shown to be critical for a correct description of the oxo-iron species and were shown to change the regioselectivity of substrate oxidative catalysis.³⁸ We used the coordinates of the ammonia molecules as described in ref 37 and ran single-point calculations in Jaguar using BS1.

Results

The work described here focuses on the nature of azole binding to the active center of P450 enzymes and how these chemicals work as inhibitors. We studied the binding of three typical azole drugs, each with characteristic features, namely, econazole, fluconazole, and metyrapone. These three inhibitors have different chemical structures that interact with the heme, as shown in Scheme 1 above. Thus, econazole binds the heme with the imidazolate group, which we abbreviated in our model with methylimidazolate. Fluconazole, by contrast, has a characteristic triazolate group that binds the heme, which we abbreviated with methyltriazolate. Finally, metyrapone binds the heme with a pyridine group. The studies started from the resting state of P450, where a water molecule occupies the sixth binding site of iron. Figure 1 shows the optimized geometries of the resting state of P450 (labeled \mathbf{R}) with an approaching inhibitor molecule. The structures are labeled as based on the inhibitor molecule in the model, with a subscript next to the structure label: E for econazole, F for fluconazole, and M for metyrapone. We calculated all structures in the lowest-lying doublet, quartet, and sextet spin states, as indicated with a superscript. As can be seen from Figure 1, the Fe–S distance is hardly influenced by the nature of the approaching inhibitor but is different for each spin state. This is as expected since in ${}^2\mathbf{R}$, the system has $\delta^2 \pi_{xz}^* \pi_{yz}^*$ occupation, while in ${}^4\mathbf{R}$ and ${}^6\mathbf{R}$, the orbital occupation is $\delta^2 \pi_{xz}^* \pi_{yz}^* \sigma_{z2}^*$ and $\delta^1 \pi_{xz}^* \pi_{yz}^* \sigma_{z2}^* \sigma_{xz}^*$, respectively.^{29c} Thus, single occupation of the σ_{z2}^* orbital, which is antibonding along the Fe–S bond, will elongate this distance. Relative energies between the doublet, quartet, and sextet spin states are not influenced by the approaching inhibitor molecule either.

The approaching inhibitor molecule forms a hydrogen bond with the water molecule. The Fe–OH₂ distance is influenced

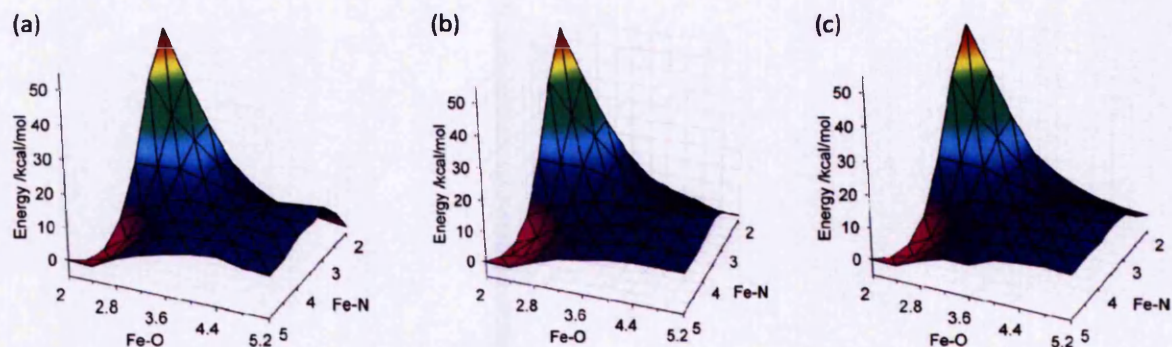


Figure 2. Two-dimensional geometry scans for the simultaneous displacement of water from the active center and for binding of the azole. All data points represent a full geometry optimization in Gaussian 03 with two degrees of freedom fixed (the Fe–OH₂ and Fe–N_{azole} distances) using method M2 and basis set BS1. All energies are in kcal mol⁻¹ relative to the resting state. (a) ²R_E potential energy surface. (b) ²R_F potential energy surface. (c) ²R_M potential energy surface.

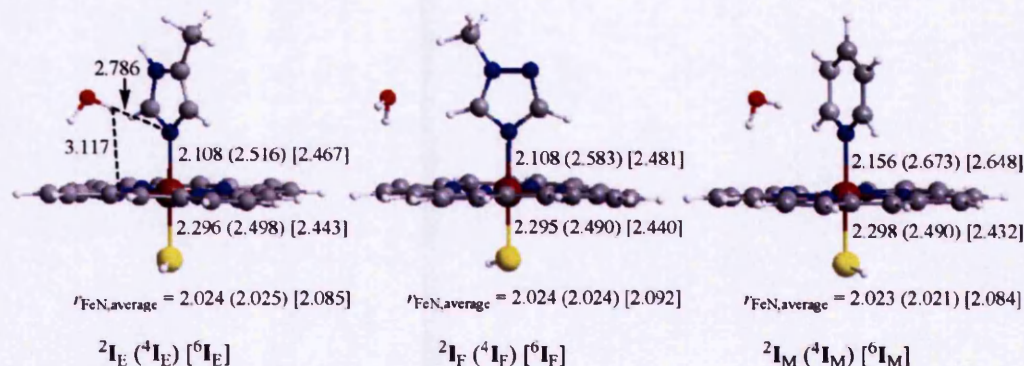


Figure 3. Optimized geometries of I_E, I_F, and I_M inhibitor complexes in the doublet (quartet) [sextet] spin states. All bond lengths are in Å, and relative energies are in kcal mol⁻¹.

by the strength of the hydrogen bond of the inhibitor molecule and is shortest for R_E and longest for R_M. The resting state has been extensively studied with experimental as well as theoretical procedures.^{30c,39,40} Thus, electron spin-echo envelope modulation (ESEEM) spectroscopy characterized the resting state as a doublet spin state.³⁹ Detailed quantum mechanical/molecular mechanics (QM/MM) calculations predicted the doublet spin state to be 3.3 and 2.4 kcal mol⁻¹ lower in energy than the lowest quartet and sextet states, respectively.^{40c} Our calculations shown in Figure 1 support the experimental and theoretical assignments of the resting state from the literature.

To find out how azoles bind to the active center of hemes, we ran extensive geometry scans starting from ²R_E, ²R_F, and ²R_M for the approach of the azoles to the heme by shortening the Fe–N distance between the heme and the azole. These geometry scans, however, failed to displace the water molecule from the iron center and gave a high-energy pathway. Subsequently, we ran two-dimensional geometry scans for the simultaneous approach of the azole to the heme and the displacement of the water molecule from the heme, and the results are shown in Figure 2. The three inhibitors give the same potential energy surface for the displacement of water by the inhibitor molecule. As follows from these geometry scans, the reaction mechanism is not accomplished with a simultaneous replacement of water by inhibitor. Instead, a stepwise mechanism takes place, with an initial water displacement from the heme site to create a pentacoordinated heme followed by binding of inhibitor to the empty ligand site on the iron. Displacement of water creates a pentacoordinated iron center with a nearby water and inhibitor molecule and with barriers for displacement of the water molecule from ²R_E, ²R_F, and ²R_M of 15.2, 13.0,

and 13.6 kcal mol⁻¹, respectively. These barriers lead to local minima representing a long-range complex between heme, water, and inhibitor and are endothermic by about 13 kcal mol⁻¹. From these long-range complexes, a barrier to the inhibitor-bound complexes (I) is crossed. These barriers are 2.3, 1.3, and 0.5 kcal mol⁻¹ for econazole, fluconazole, and metyrapone, respectively. As such, the second barrier, that is, for inhibitor binding, is the largest of the two and is likely to be rate-determining. The azole-bound inhibitor complexes are labeled I_E, I_F, and I_M, and their optimized geometries are shown in Figure 3. Binding of inhibitor hardly affects the Fe–S and Fe–N distances in the heme, which are also similar in length. Moreover, these distances are comparable to the ones observed for the resting state complexes (R) shown above in Figure 1. The water molecule in the inhibitor-bound complexes (I_E, I_F, and I_M) moves away from the reaction center and out of reach of hydrogen-bonding donors. In most structures, it moves to the meso position of the heme.

The potential energy landscapes shown in Figure 2 seem to imply that inhibitor binding is thermodynamically unfavorable and that the resting state (R) is lower in energy than the inhibitor-bound complex (I). Of course, this would contradict experiment and would suggest that inhibitors will not be able to displace the water molecule in the resting state. Careful examination of the differences between the structures R_E, R_F, and R_M, on the one hand, and I_E, I_F, and I_M reveals an extra hydrogen bond in the resting state geometries between the water molecule and the approaching inhibitor, which is missing in the inhibitor-bound complexes. Thus, the loss of the hydrogen bond between the water molecule and the inhibitor that is characteristic for the complexes R_E, R_F, and R_M artificially

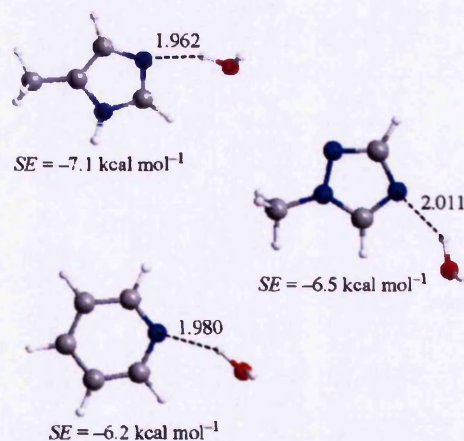


Figure 4. Hydrogen bond energy and hydrogen bond length of methylimidazolate–water, trimethyltriazolate–water, and pyridine–water complexes.

destabilizes the inhibitor-bound complexes in energy. As a result, we find ${}^2\text{I}_\text{E}$, ${}^2\text{I}_\text{F}$, and ${}^2\text{I}_\text{M}$ to be less stable in energy by 3.9, 1.6, and 6.5 kcal mol $^{-1}$ with respect to ${}^2\text{R}_\text{E}$, ${}^2\text{R}_\text{F}$, and ${}^2\text{R}_\text{M}$. This would imply that the inhibitor molecules are less likely to bind the iron active center of P450 enzymes than is a water molecule. However, crystal structures of inhibitor–heme complexes have been clearly resolved, indicating that they are more stable than the resting state (water-bound) structures. Thus, several crystal structures are available with azoles directly bound to the heme iron, that is, in the **I** configuration.^{21–23,41}

P450–antifungal azole complexes have been crystallized for Mtb CYP51B1 and CYP121 with fluconazole, for *S. erythraea* P450_{eryF} with ketoconazole, and for rabbit CYP2B4 with bifonazole. Moreover, structures of Mtb CYP51B1 with 4-phenylimidazole, as well as for *P. putida* P450_{cam} with imidazole, have been reported. All show direct coordination of heme iron by a nitrogen atom from the azole drug. On the other hand, the Mtb CYP121 structure with fluconazole revealed two different azole-bound species, one with fluconazole directly bound to the iron and the other with the azole coordinating the heme iron via a bridging water molecule. The indirect binding mode was predominant and was consistent with a constrained CYP121 active site architecture.²³ Thus, these data for CYP121 indicated that alternative binding modes for azole drugs are accessible should steric constraints prevent direct heme iron coordination by the azole.

To estimate the energetic strength of this hydrogen bond between the water molecule and the inhibitor, we ran additional calculations of the inhibitor–water complex and compared the energies with isolated water and inhibitor. The results are shown in Figure 4. In the free inhibitor–water complexes, the hydrogen bond is between 1.962 and 2.011 Å in length, which is significantly longer than that in the resting state complexes in Figure 1, where values between 1.774 and 1.832 Å were found. Therefore, our estimated value of the hydrogen bond strength will be somewhat larger than is actually the case in Figure 1. Thus, the hydrogen bond energies shown in Figure 4 indicate an extra stabilization energy (SE) of the water–inhibitor complexes by at least SE = 6.2–7.1 kcal mol $^{-1}$. Thus, subtracting the stabilization energy values of the water–inhibitor complexes, as shown in Figure 4 from the reactants complexes, gives a relative energy between the resting state and inhibitor complexes of −3.2 (I_E), −4.9 (I_F), and +0.3 (I_M) kcal mol $^{-1}$. These corrected relative energies of the inhibitor-bound com-

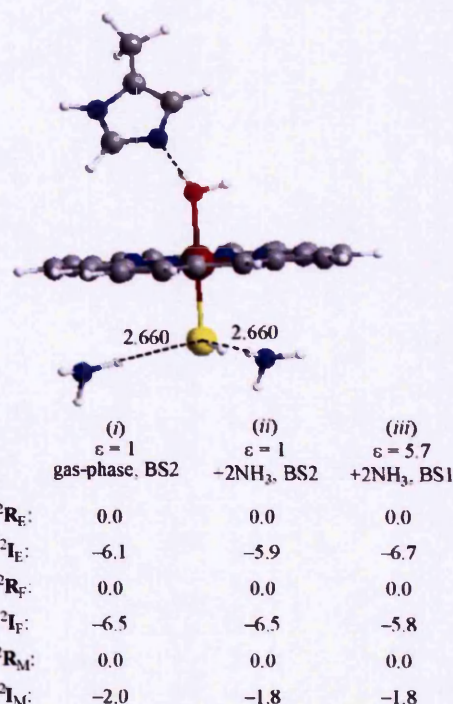


Figure 5. The effect of the environment on the ordering of the resting state and azole-bound structures: (i) gas-phase results without external perturbations, (ii) gas-phase model with two extra hydrogen-bonded ammonia molecules, and (iii) system with two hydrogen-bonded ammonia molecules in a dielectric constant of $\epsilon = 5.7$. All calculations represent single-point energy calculations with either basis set BS1 or BS2. Zero-point energy (ZPE) and stabilization energy (SE) corrections have been included in the data.

plexes support experimental findings that the inhibitor-bound complex is more stable than the resting state complexes. However, it appears here that the energy difference is very small; therefore, it may be anticipated that small perturbations within the active pocket may influence the efficiency of inhibitor binding. In the actual enzyme, the broken hydrogen bond between the water molecule and the approaching inhibitor molecule will be replaced by another hydrogen bond to the water inside the active site pocket.

Furthermore, the barriers for displacement of the water molecule by the inhibitor molecules from Figure 2 are also overestimated by SE, with values of approximately 6.2–7.1 kcal mol $^{-1}$. Therefore, the barriers for the displacement of the water molecules from ${}^2\text{R}_\text{E}$, ${}^2\text{R}_\text{F}$, and ${}^2\text{R}_\text{M}$ estimated from the scans in Figure 2 should be lowered by the same amount and are expected instead to be in the range of 8.8, 6.5, and 7.4 kcal mol $^{-1}$, respectively, for the three processes. Thus, displacement of the water molecule in the resting state of P450 enzymes with an inhibitor molecule will be dependent on the availability of hydrogen-bonding interactions in the active site pocket that can stabilize the leaving water molecule. Moreover, the size and shape of the substrate pocket will determine whether a “perfect fit” of the inhibitor molecule in the pocket is possible.

To find out whether inhibitor binding is influenced by external perturbations, such as solvent polarity or hydrogen bonding, we tested the effect of the local environment on the energy differences of the resting state and azole-bound structures in the doublet spin state. We tested several effects, such as a dielectric constant of value $\epsilon = 5.7$ and the addition of the hydrogen-bonded ammonia molecules toward the thiolate axial

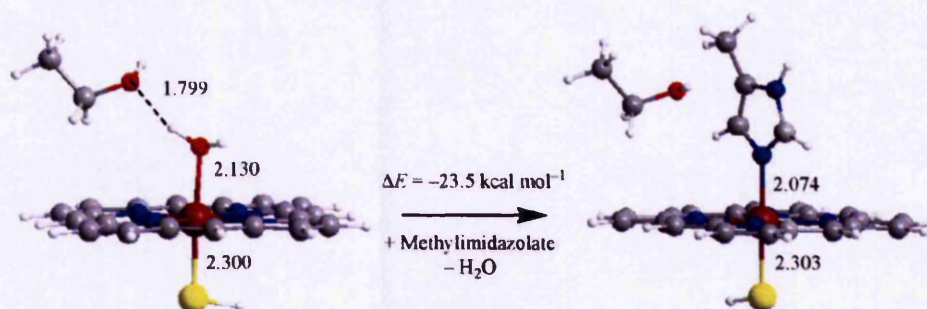


Figure 6. Methylimidazolate-bound P450 structures with a nearby ethanol molecule mimicking Ser₂₃₇ in CYP121. All calculations were performed in Gaussian with a fixed ethanol molecule. Bond lengths shown are in Å.

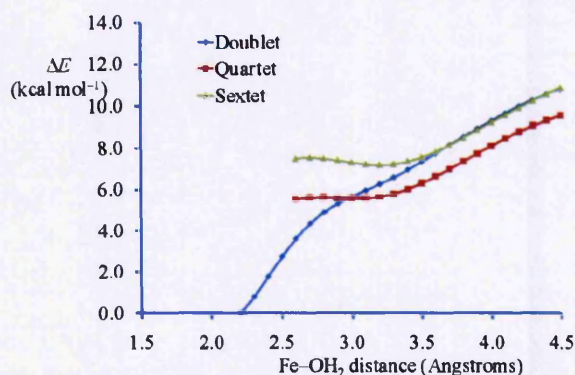


Figure 7. Geometry scans starting from the resting state models (without azole groups) for the dissociation of water from the heme. All data points correspond to a full geometry optimization in Jaguar 7.0 with one degree of freedom (the Fe–O distance) fixed. All energies are in kcal mol⁻¹ relative to the optimized geometry of the resting state in the doublet spin state.

ligand; see Figure 5. The latter effect was shown previously to be important for the correct description of the polarity and electronegativity of the iron–sulfur group.³⁷

As follows from Figure 5, the azole binding energy is hardly influenced by environmental effects such as a dielectric constant or hydrogen-bonding interactions toward the thiolate group. The calculations do show, however, that pyridine is a much weaker inhibitor than methylimidazolate and methyltriazolate. Therefore, P450 enzymes may not bind pyridine-containing inhibitor molecules strongly. Destabilizing effects of the substrate binding pocket should weaken the inhibitor binding complex by at least 6 kcal mol⁻¹ in order to make the two structures comparable in energy. Thus, recent crystallographic evidence showed two alternative inhibitor-bound complexes, one in which the inhibitor is directly bound to the heme (cf. structures **I** above) and one with a water molecule bridging the inhibitor (cf. structures **R** above).²³ The substrate binding pocket of this particular P450 isozyme (CYP121), however, is characterized by a nearby serine (Ser₂₃₇) amino acid within hydrogen-bonding distance of the heme. To find out how this serine amino acid influences inhibitor binding, we ran calculations with an extra ethanol molecule in our model in the position of Ser₂₃₇ of the CYP121 P450, and the results are shown in Figure 6.

As follows from Figure 6, the resting state structure enjoys a hydrogen bond with the ethanol molecule mimicking Ser₂₃₇. This, however, does not stabilize the resting state but in fact weakens the Fe–OH₂ bond considerably, which is increased in length from 2.080 Å in ²R_E to 2.130 Å in Figure 6. As a result of this change, the inhibitor-bound complex is now more stable by 23.5 kcal mol⁻¹. Therefore, hydrogen-bonding interactions

within the protein pocket determine the relative stability of the resting state and inhibitor-bound complexes. Moreover, the size and shape of the substrate binding pocket will also clearly determine whether the inhibitor fits easily or not.

Discussion

In this paper, we presented studies on azole inhibition of P450 enzymes. Thus, extensive geometry scans on a resting state model and the approach of azole molecules show that subtle hydrogen-bonding interactions within the active site pocket guide the azole binding process. Azole molecules approach the heme group and are pulled toward it through an accepting hydrogen bond of the resting state water molecule. When the azole molecule approaches the heme closely, the water molecule is released from the heme, and a pentacoordinated heme group remains with nearby azole and water. The azole then replaces the water in the sixth ligand position of the heme, thereby blocking dioxygen binding and rendering the heme centre inactive. The DFT models show that the process is subtle and influenced by hydrogen-bonding donor and acceptor groups within the active site, leading to release of the distal water molecule from the heme iron.

Azole groups bind the heme group and form a covalent linkage with the iron atom with distances of 2.108–2.156 Å in the low-spin states. These distances are similar to those that one would expect for an Fe–N distance in a heme enzyme and have been observed before, for example, for imidazole-ligated iron porphyrin systems mimicking the active site of peroxidases.^{30,42} Although, the Fe–N bond between the metal and the azole group is weak, it is of sufficient strength to keep the metal ion in the center (or close to the center) of the porphyrin ring. Thus, previous studies showed that removal of the water molecule from the resting state leads to a displacement of the iron atom well below the porphyrin ring.^{28c} As a consequence of this, the orbital mixing patterns are influenced as well. In particular, in strongly distorted systems where the metal is displaced well below the center of the porphyrin ring, the σ^*_{-2} orbital on the metal, representing the antibonding interactions of the metal with the axial cysteinyl ligand, mixes with heme-type π^* orbitals, giving the system extra stabilization.⁴³ Especially in high-spin states where the σ^*_{-2} orbital is singly occupied, this effect is stronger, and, hence, high spin states are stabilized over the low-spin states in pentacoordinated iron porphyrin systems. This means that removal of a water molecule from the resting state should lead to a low-spin to high-spin conversion of the metal center, as indeed observed using, for example, electron paramagnetic resonance (EPR) studies.⁴⁴ To find out whether a spin state crossing is expected along the azole binding mechanism, we ran additional geometry scans, but

without the azole molecules removed from the model. In this model, the water molecule is withdrawn from the resting state structure of the heme in small steps. The results are shown in Figure 7 and plotted versus the Fe–O distance for the doublet, quartet, and sextet spin states. As follows from Figure 7, although in the resting state the energy separation between the different spin states is large and well in favor of the doublet spin at Fe–O distances of 2.8 Å or more, the three spin state surfaces approach each other to within 1–2 kcal mol⁻¹ at larger distances. In our particular case, the quartet spin state is the ground state of the pentacoordinated complex, and the doublet–quartet spin state crossing appears at around 3.0 Å. Thus, since the azole replacement of the water molecule (Figure 2) leads to a pentacoordinate heme as an intermediate complex, it may very well be that a spin crossing occurs. However, as shown in Figures 2 and 7, no spin crossing is observed in the azole replacement mechanism. Thus, the azole-bound structures are low-spin (doublet) complexes and so are the resting state structures. During the water removal step in the reaction mechanism, the Fe–OH₂ distance is elongated to 3.0–3.5 Å before the azole group moves in. In this region, the quartet and doublet spin states are close in energy, but the system will remain on the doublet since the azole-bound complexes are the lowest in energy on this spin state surface.

Summary and Conclusion

Density functional calculations on inhibitor binding to the active site of P450 enzymes have been studied. It is shown that the relative energies between the resting state and the inhibitor-bound complexes are subtly dependent on small external perturbations. In particular, hydrogen-bonding interactions with either the water molecule in the resting state or with the inhibitor molecule can lead to extra stabilization of one of these states. These factors, together with the “fit” of the inhibitor molecule into the substrate pocket, determine the ordering and the capacity of the azoles to act as inhibitors.

Acknowledgment. The research was supported by a studentship to P.R.B. from the Medical Research Council (MRC, U.K.) and one to C.S.P. from a Doctoral Training Award (DTA) from the Engineering and Physical Sciences Research Council (EPSRC, U.K.). In addition, research grant funding from the European Union (New Medicines for Tuberculosis, NM4TB) and from the Biotechnology and Biological Sciences Research Council (BBSRC, U.K., Grant Codes BB/F002521/1 and BB/C511305/1) is acknowledged. The Jaguar calculations were done with CPU time provided by the National Service of Computational Chemistry Software (NSCCS).

Supporting Information Available: Cartesian coordinates of all structures described in this work, figures with geometry scans. This material is available free of charge via the Internet at <http://pubs.acs.org>.

References and Notes

- (1) (a) Sono, M.; Roach, M. P.; Coulter, E. D.; Dawson, J. H. *Chem. Rev.* **1996**, *96*, 2841–2887. (b) Ortiz de Montellano, P. R., Ed. *Cytochrome P450: Structure, Mechanism and Biochemistry*, 3rd ed.; Kluwer Academic/Plenum Publishers: New York, 2004. (c) Meunier, B.; de Visser, S. P.; Shaik, S. *Chem. Rev.* **2004**, *104*, 3947–3980. (d) Denisov, I. G.; Makris, T. M.; Sligar, S. G.; Schlichting, I. *Chem. Rev.* **2005**, *105*, 2253–2277. (e) Groves, J. T. *J. Inorg. Biochem.* **2006**, *100*, 434–447.
- (2) (a) Guengerich, F. P. *Chem. Res. Toxicol.* **2001**, *14*, 611–650. (b) Ortiz de Montellano, P. R.; De Voss, J. J. Substrate Oxidation by Cytochrome P450 Enzymes. In *Cytochrome P450: Structure, Mechanism, and Biochemistry*, 3rd ed.; Ortiz de Montellano, P. R., Ed.; Kluwer Academic/Plenum Publishers: New York, 2004; Chapter 6, pp 183–245.
- (3) Guengerich, F. P. *Drug Metab. Rev.* **2004**, *36*, 159–197.
- (4) (a) Guengerich, F. P. *Annu. Rev. Pharmacol. Toxicol.* **1999**, *39*, 1–17. (b) Guengerich, F. P. Human Cytochrome P450 Enzymes. In *Cytochrome P450: Structure, Mechanism, and Biochemistry*, 3rd ed.; Ortiz de Montellano, P. R., Ed.; Kluwer Academic/Plenum Publishers: New York, 2004; Chapter 10, pp 377–530.
- (5) Capdevila, J. H.; Falcik, J. R.; Imig, J. D. *Kidney Int.* **2007**, *72*, 683–689.
- (6) Odds, F. C.; Brown, A. J.; Gow, N. A. *Trends Microbiol.* **2003**, *11*, 272–279.
- (7) Feyerisen, R. *Biochem. Soc. Trans.* **2006**, *34*, 1252–1255.
- (8) Cole, S. T. *Microbiology* **2002**, *148*, 2919–2928.
- (9) McLean, K. J.; Dunford, A. J.; Neeli, R.; Driscoll, M. D.; Munro, A. W. *Arch. Biochem. Biophys.* **2007**, *464*, 228–240.
- (10) McLean, K. J.; Warman, A. J.; Seward, H. E.; Marshall, K. R.; Girvan, H. M.; Cheesman, M. R.; Waterman, M. R.; Munro, A. W. *Biochemistry* **2006**, *45*, 8427–8443.
- (11) Hollenberg, P. F. *Drug Metab. Rev.* **2002**, *34*, 17–35.
- (12) (a) Ferslew, K. E.; Hagardorn, A. N.; Harlan, G. C.; McCormick, W. F. *J. Forensic Sci.* **1998**, *43*, 1082–1085. (b) Kudo, K.; Imamura, T.; Jitsufuchi, N.; Zhang, X. X.; Tokunaga, H.; Nagata, T. *Forensic Sci. Int.* **1997**, *86*, 35–41.
- (13) Temple, T. E.; Liddle, G. W. *Annu. Rev. Pharmacol. Toxicol.* **1970**, *10*, 199–218.
- (14) Zhang, W.; Ramamoorthy, Y.; Kilicarslan, T.; Nolte, H.; Tyndale, R. F.; Sellers, E. M. *Drug Metab. Dispos.* **2002**, *30*, 314–318.
- (15) Ahmad, Z.; Sharma, S.; Khuller, G. K. *FEMS Microbiol. Lett.* **2006**, *262*, 181–186.
- (16) Schenkman, J. B.; Remmer, H.; Estabrook, R. W. *Mol. Pharmacol.* **1967**, *3*, 113–123.
- (17) Van den Bossche, H. *Curr. Top. Med. Mycol.* **1985**, *1*, 313–351.
- (18) Schuster, I. *Xenobiotica* **1985**, *15*, 529–546.
- (19) (a) Varhe, A.; Olkkola, K. T.; Neuvonen, P. J. *Clin. Pharmacol. Ther.* **1994**, *56*, 601–607. (b) Park, J. Y.; Shon, J. H.; Kim, K. A.; Jung, H. J.; Shim, J. C.; Yoon, Y. R.; Cha, I. J.; Shin, J. G. *J. Clin. Psychopharmacol.* **2006**, *26*, 135–142. (c) Murayama, N.; Imai, N.; Nakane, T.; Shimizu, M.; Yamazaki, H. *Biochem. Pharmacol.* **2007**, *73*, 2020–2026.
- (20) (a) Guardiola-Diaz, H. M.; Foster, L. A.; Mushrush, D.; Vaz, A. D. *Biochem. Pharmacol.* **2001**, *61*, 1463–1470. (b) McLean, K. J.; Cheesman, M. R.; Rivers, S. L.; Richmond, A.; Leys, D.; Chapman, S. K.; Reid, G. A.; Price, N. C.; Kelly, S. M.; Clarkson, J.; Smith, W. E.; Munro, A. W. *J. Inorg. Biochem.* **2002**, *91*, 527–541. (c) Banfi, E.; Scialino, G.; Zampieri, D.; Mamolo, M. G.; Vio, L.; Ferrone, M.; Ferneglia, M.; Paneni, M. S.; Pricl, S. *J. Antimicrob. Chemother.* **2006**, *58*, 76–84.
- (21) (a) Poulos, T. L.; Howard, A. J. *Biochemistry* **1987**, *26*, 8165–8174. (b) Podust, L. M.; Poulos, T. L.; Waterman, M. R. *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98*, 3068–3073.
- (22) (a) Cupp-Vickery, J. R.; Garcia, C.; Hofacre, A.; McGee-Estrada, K. J. *Mol. Biol.* **2001**, *311*, 101–110. (b) Verras, A.; Alian, A.; Ortiz de Montellano, P. R. *Protein Eng., Des. Sel.* **2006**, *19*, 491–496. (c) Ekroos, M.; Sjogren, T. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 13682–13687.
- (23) Seward, H. E.; Roujeinikova, A.; McLean, K. J.; Munro, A. W.; Leys, D. *J. Biol. Chem.* **2006**, *281*, 39437–39443.
- (24) Goren, A. C. F.; Schmidt, K.; Mayer, B. *Biochemistry* **2002**, *41*, 7819–7829.
- (25) (a) Das, A.; Grinkova, Y. V.; Sligar, S. G. *J. Am. Chem. Soc.* **2007**, *129*, 13778–13779. (b) Denisov, I. G.; Grinkova, Y. V.; McLean, M. A.; Sligar, S. G. *J. Biol. Chem.* **2007**, *282*, 26865–26873.
- (26) Poulos, T. L.; Raag, R. *FASEB J.* **1992**, *6*, 674–679.
- (27) Schlichting, I.; Berendzen, J.; Chu, K.; Stock, A. M.; Maves, S. A.; Benson, D. E.; Sweet, R. M.; Ringe, D.; Petsko, G. A.; Sligar, S. G. *Science* **2000**, *287*, 1615–1622.
- (28) (a) Dawson, J. H.; Holm, R. H.; Trudell, J. R.; Barth, G.; Linder, R. E.; Bunnenberg, E.; Djerassi, C.; Tang, S. C. *J. Am. Chem. Soc.* **1976**, *98*, 3707–3709. (b) Poulos, T. L. *J. Biol. Inorg. Chem.* **1996**, *1*, 356–359. (c) Shaik, S.; Kumar, D.; de Visser, S. P.; Altun, A.; Thiel, W. *Chem. Rev.* **2005**, *105*, 2279–2328.
- (29) (a) Paine, M. J. I.; Scrutton, N. S.; Munro, A. W.; Gutierrez, A.; Roberts, G. C. K.; Wolf, C. R. Electron Transfer Partners of Cytochrome P450. In *Cytochrome P450: Structure, Mechanism, and Biochemistry*, 3rd ed.; Ortiz de Montellano, P. R., Ed.; Kluwer Academic/Plenum Publishers: New York, 2004; Chapter 4, pp 115–148. (b) Wannan, A. J.; Roitel, O.; Neeli, R.; Girvan, H. M.; Seward, H. E.; Murray, S. A.; McLean, K. J.; Joyce, M. G.; Toogood, H.; Holt, R. A.; Leys, D.; Scrutton, N. S.; Munro, A. W. *Biochem. Soc. Trans.* **2005**, *33*, 747–753.
- (30) (a) Green, M. T. *J. Am. Chem. Soc.* **2000**, *122*, 9495–9499. (b) de Visser, S. P.; Shaik, S.; Sharma, P. K.; Kumar, D.; Thiel, W. *J. Am. Chem. Soc.* **2003**, *125*, 15779–15788. (c) Rydberg, P.; Sigfridsson, E.; Ryde, U. *J. Biol. Inorg. Chem.* **2004**, *9*, 203–223. (d) Derat, E.; Shaik, S. *J. Am. Chem. Soc.* **2006**, *128*, 8185–8198. (e) Groenhof, A. E.; Swart, M.; Ehlers, A. W.; Lammertsma, K. J. *Phys. Chem. A* **2005**, *109*, 3411–3417.

- (31) (a) de Visser, S. P. *J. Am. Chem. Soc.* **2006**, *128*, 9813–9824. (b) de Visser, S. P. *J. Am. Chem. Soc.* **2006**, *128*, 15809–15818. (c) Aluri, S.; de Visser, S. P. *J. Am. Chem. Soc.* **2007**, *129*, 14846–14847.
- (32) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Montgomery, J. A., Jr.; Vreven, T.; Kudin, K. N.; Burant, J. C.; Millam, J. M.; Iyengar, S. S.; Tomasi, J.; Barone, V.; Mennucci, B.; Cossi, M.; Scalmani, G.; Rega, N.; Petersson, G. A.; Nakatsuji, H.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Klene, M.; Li, X.; Knox, J. E.; Hratchian, H. P.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Ayala, P. Y.; Morokuma, K.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Zakrzewski, V. G.; Dapprich, S.; Daniels, A. D.; Strain, M. C.; Farkas, O.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Ortiz, J. V.; Cui, Q.; Baboul, A. G.; Clifford, S.; Cioslowski, J.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Challacombe, M.; Gill, P. M. W.; Johnson, B.; Chen, W.; Wong, M. W.; Gonzalez, C.; Pople, J. A. *Gaussian 03*, Revision C.02; Gaussian, Inc.: Wallingford, CT, 2003.
- (33) (a) Becke, A. D. *J. Chem. Phys.* **1993**, *98*, 5648–5652. (b) Lee, C.; Yang, W.; Parr, R. G. *Phys. Rev. B* **1988**, *37*, 785–789.
- (34) (a) Becke, A. D. *Phys. Rev. A* **1988**, *38*, 3098–3100. (b) Perdew, J. P. *Phys. Rev. B* **1986**, *33*, 8822–8824.
- (35) Hay, P. J.; Wadt, W. R. *J. Chem. Phys.* **1985**, *82*, 270–283.
- (36) *Jaguar 7.0*; Schrödinger, LLC.: New York, NY, 2007.
- (37) Ogliaro, F.; Cohen, S.; de Visser, S. P.; Shaik, S. *J. Am. Chem. Soc.* **2000**, *122*, 12892–12893.
- (38) (a) de Visser, S. P.; Ogliaro, F.; Sharma, P. K.; Shaik, S. *Angew. Chem., Int. Ed.* **2002**, *41*, 1947–1951. (b) de Visser, S. P.; Ogliaro, F.; Sharma, P. K.; Shaik, S. *J. Am. Chem. Soc.* **2002**, *124*, 11809–11826. (c) de Visser, S. P.; Shaik, S. *J. Am. Chem. Soc.* **2003**, *125*, 7413–7424. (d) de Visser, S. P. *Chem.—Eur. J.* **2006**, *12*, 8168–8177.
- (39) Thomann, H.; Bernardo, M.; Goldfarb, D.; Kroneck, P. M. H.; Ullrich, V. *J. Am. Chem. Soc.* **1995**, *117*, 8243–8251.
- (40) (a) Green, M. T. *J. Am. Chem. Soc.* **1998**, *120*, 10772–10773. (b) Filatov, M.; Harris, N.; Shaik, S. *J. Chem. Soc., Perkin Trans. 2* **1999**, *3*, 399–411. (c) Schöneboom, J. C.; Thiel, W. *J. Phys. Chem. B* **2004**, *108*, 7468–7478.
- (41) Zhao, Y. H.; White, M. A.; Muralidhara, B. K.; Sun, L.; Halpert, J. R.; Stout, C. D. *J. Biol. Chem.* **2006**, *281*, 5973–5981.
- (42) de Visser, S. P. *J. Phys. Chem. A* **2005**, *109*, 11050–11057.
- (43) de Visser, S. P.; Ogliaro, F.; Gross, Z.; Shaik, S. *Chem.—Eur. J.* **2001**, *7*, 4954–4960.
- (44) (a) Sligar, S. G. *Biochemistry* **1976**, *15*, 5399–5406. (b) Auclair, K.; Moënné-Loccoz, P.; Ortiz de Montellano, P. R. *J. Am. Chem. Soc.* **2001**, *123*, 4877–4885.

JP802087W

REFERENCES

- Aaron, L., Saadoun, D., Calatroni, I., Launay, O., Memain, N., Vincent, V., Marchal, G., Dupont, B., Bouchaud, O., Valeyre, D. and Lortholary, O. (2004) *Clin. Microbiol. Infect.* **10**, 388-398.
- Agmon, N. and Rabinovich, S. (1992) *J. Chem. Phys.* **97**, 7270-7286.
- Ahmad, Z., Sharma, S. and Khuller, G. K. (2006) *FEMS Microbiol. Lett.* **261**, 181-186.
- Altun, A. and Thiel, W. (2005) *J. Phys. Chem. B* **109**, 1268-1280.
- Altun, A., Shaik, S. and Thiel, W. (2007) *J. Am. Chem. Soc.* **129**, 8978-8987.
- Alvares, A. P., Schillin, G., Levin, W. and Kuntzman, R. (1967) *Biochem. Biophys. Res. Commun.* **29**, 521-526.
- Andersen, J. F., Ding, X. D., Balfour, C., Shokhireva, T. K., Champagne, D. E., Walker, F. A. and Montfort, W. R. (2000) *Biochemistry* **39**, 10118-10131.
- Antony, J., Grodzicki, M. and Trautwein, A. X. (1997) *J. Phys. Chem. A* **101**, 2692-2701.
- Aoyama, Y., Horiuchi, T., Gotoh, O., Noshiro, M. and Yoshida, Y. (1998) *J. Biol. Chem.* **124**, 694-696.
- Bach, R. D. and Dmitrenko, O. (2006) *J. Am. Chem. Soc.* **128**, 1474-1488.
- Banfi, E., Scialino, G., Zampieri, D., Mamolo, M. G., Vio, L., Ferrone, M., Fermeglia, M., Paneni, M. S. and Pricl, S. (2006) *J. Antimicrob. Chemother.* **58**, 76-84.
- Barker, P. D. and Ferguson, S. J. (1999) *Structure* **7**, 281-290.
- Baston, E. and Leroux, F. R. (2007) *Rec. Pat. Anti-Can. Drug Disc.* **2**, 31-58.
- Bathelt, C. M., Ridder, L., Mulholland, A. J. and Harvey, J. N. (2004) *Org. Biomol. Chem.* **2**, 2998-3005.
- Becke, A. D. (1988) *Phys. Rev. A* **38**, 3098-3100.
- Becke, A. D. (1993a) *J. Chem. Phys.* **98**, 5648-5652.
- Becke, A. D. (1993b) *J. Chem. Phys.* **98**, 1372-1377.
- Bell, S. G., Sowden, R. J. and Wong, L.-L. (2001) *Chem. Commun.* 635-636.
- Bellamine, A., Mangla, A. T., Nes, W. D. and Waterman, M. R. (1999) *Proc. Natl. Acad. Sci. USA* **96**, 8937-8942.
- Beratan, D. N. and Onuchic, J. N. (1996) *Protein electron transfer*, Bios Scientific Publishers, Oxford, edited by Bendall, D. S. pp. 23-42.

- Berger, B. J. and Knodel, M. H. (2003) *BMC Microbiol.* **3** 1-13.
- Berglund, G. I., Carlsson, G. H., Smith, A. T., Szoke, H., Henriksen, A. and Hajdu, J. (2002) *Nature* **417**, 463-468.
- Berry, E. A. and Trumpower, B. L. (1987) *Anal. Biochem.* **161**, 1-15.
- Berry, R. E., Shokhireva, T. K., Filippov, I., Shokhirev, M. N., Zhang, H. J. and Walker, F. A. (2007) *Biochemistry* **46**, 6830-6843.
- Bloomer, J. C., Woods, F. R., Haddock, R. E., Lennard, M. S. and Tucker, G. T. (1992) *Br. J. Clin. Pharmacol.* **33**, 521-523.
- Blower, S. and Supervie, V. (2007) *Lancet Infect. Dis.* **7**, 443-443.
- Bonagura, C. A., Bhaskar, B., Shimizu, H., Li, H. Y., Sundaramoorthy, M., McRee, D. E., Goodin, D. B. and Poulos, T. L. (2003) *Biochemistry* **42**, 5600-5608.
- Born, M. and Oppenheimer, J. R. (1927) *Ann. Physik.* **84**, 457.
- Boys, S. F. and Bernardi, F. (1970) *Mol. Phys.* **19**, 553-566.
- Brodie, B. B., Gillette, J. R. and Ladu, B. N. (1958) *Annu. Rev. Biochem.* **27**, 427-454.
- Brown, K. R., Brown, B. M., Hoagland, E., Mayne, C. L. and Hegg, E. L. (2004) *Biochemistry* **43**, 8616-8624.
- Camacho, L. R., Ensergueix, D., Perez, E., Gicquel, B. and Guilhot, C. (1999) *Mol. Microbiol.* **34**, 257-267.
- Cammer, W., Schenkman, J. B. and Estabrook, R. W. (1966) *Biochem. Biophys. Res. Commun.* **23**, 264-268.
- Cammi, R., Mennucci, B. and Tomasi, J. (1999) *J. Phys. Chem. A* **103**, 9100-9108.
- Cammi, R., Mennucci, B. and Tomasi, J. (2000) *J. Phys. Chem. A* **104**, 5631-5637.
- Cances, E., Mennucci, B. and Tomasi, J. (1997) *J. Chem. Phys.* **107**, 3032-3041.
- Cardona, P. J. and Ruiz-Manzano, J. (2004) *Eur. Resp. J.* **24**, 1044-1051.
- Caughey, W. S., Smythe, G. A., O'Keeffe, D. H., Maskasky, J. E. and Smith, M. L. (1975) *J. Biol. Chem.* **250**, 7602-7622.
- Chan, W. K., Sui, Z. and Ortiz de Montellano, P. R. (1993) *Chem. Res. Toxicol.* **6**, 38-45.
- Chang, C. K. and Wu, W. (1986) *J. Biol. Chem.* **261**, 8593-8596.
- Chapman, S. K., Daff, S. and Munro, A. W. (1997) *Metal Sites in Proteins and Models*, Springer Berlin / Heidelberg pp 39-70.

Chottard, G., Schappacher, M., Ricard, L. and Weiss, R. (1984) *Inorg. Chem.* **23**, 4557-4561.

Colditz, G. A., Brewer, T. F., Berkey, C. S., Wilson, M. E., Burdick, E., Fineberg, H. V. and Mosteller, F. (1994) *J. Am. Med. Assoc.* **271**, 698-702.

Cole, S. T., Brosch, R., Parkhill, J., Garnier, T., Churcher, C., Harris, D., Gordon, S. V., Eiglmeier, K., Gas, S., Barry, C. E., Tekai, F., Badcock, K., Basham, D., Brown, D., Chillingworth, T., Connor, R., Davies, R., Devlin, K., Feltwell, T., Gentles, S., Hamlin, N., Holroyd, S., Hornsby, T., Jagels, K., Krogh, A., McLean, J., Moule, S., Murphy, L., Oliver, K., Osborne, J., Quail, M. A., Rajandream, M.-A., Rogers, J., Rutter, S., Seeger, K., Skelton, J., Squares, R., Squares, S., Sulston, J. E., Taylor, K., Whitehead, S. and Barrell, B. G. (1998) *Nature* **393**, 537-544.

Conrad, H. E., Lieb, K. and Gunsalus, I. C. (1965) *J. Biol. Chem.* **240**, 4029-4037.

Correia, M. and Ortiz de Montellano, P. R. (2005) *Cytochrome P450: Structure, Mechanism, and Biochemistry*, 3rd edition, Plenum Press, New York pp. 247-322.

Cossi, M., Barone, V., Cammi, R. and Tomasi, J. (1996) *Chem. Phys. Lett.* **255**, 327-335.

Cossi, M., Barone, V., Mennucci, B. and Tomasi, J. (1998) *Chem. Phys. Lett.* **286**, 253-260.

Cossi, M., Barone, V. and Robb, M. A. (1999) *J. Chem. Phys.* **111**, 5295-5302.

Cossi, M. and Barone, V. (2000) *J. Chem. Phys.* **112**, 2427-2435.

Cossi, M. and Barone, V. (2001) *J. Chem. Phys.* **115**, 4708-4717.

Cossi, M., Rega, N., Scalmani, G. and Barone, V. (2001) *J. Chem. Phys.* **114**, 5691-5701.

Cossi, M., Scalmani, G., Rega, N. and Barone, V. (2002) *J. Chem. Phys.* **117**, 43-54.

Cossi, M., Rega, N., Scalmani, G. and Barone, V. (2003) *J. Comput. Chem.* **24**, 669-681.

Cramer, C. J. (2004) *Computational Chemistry, Theories and Models*, 2nd edition, Wiley

Cupp-Vickery, J. R. and Poulos, T. L. (1995) *Nat. Struct. Biol.* **2**, 144-153.

Cupp-Vickery, J. R., Han, O., Hutchinson, C. R. and Poulos, T. L. (1996) *Nat. Struct. Biol.* **3**, 632-637.

Cupp-Vickery, J. R., Garcia, C., Hofacre, A. and McGee-Estrada, K. (2001) *J. Mol. Biol.* **311**, 101-110.

Czjzek, M., Guerlesquin, F., Bruschi, M. and Haser, R. (1996) *Structure* **4**, 395-404.

- Dalvi, R. R. (1987) *Arch. Toxicol.* **61**, 155-157.
- Davydov, R., Kappl, R., Huttermann, J. and Peterson, J. A. (1991) *FEBS Lett.* **295**, 113-115.
- Davydov, R., Macdonald, I. D. G., Makris, T. M., Sligar, S. G. and Hoffman, B. M. (1999) *J. Am. Chem. Soc.* **121**, 10654-10655.
- Davydov, R., Makris, T. M., Kofman, V., Werst, D. E., Sligar, S. G. and Hoffman, B. M. (2001) *J. Am. Chem. Soc.* **123**, 1403-1415.
- Dawson, J. H., Andersson, L. A. and Sono, M. (1982) *J. Biol. Chem.* **257**, 3606-3617.
- de Jong, G. T., Sola, M., Visscher, L. and Bickelhaupt, F. M. (2004) *J. Chem. Phys.* **121**, 9982-9992.
- de Visser, S. P. (2006) *J. Biol. Inorg. Chem.* **11**, 168-178.
- de Visser, S. P. (Personal communication) Resting state upright conformation geometry kindly provided by Sam de Visser, University of Manchester.
- Denisov, I. G., Makris, T. M., Sligar, S. G. and Schlichting, I. (2005) *Chem. Rev.* **105**, 2253-2277.
- Du, P. and Loew, G. H. (1992) *Int. J. Quantum Chem.* **44**, 251-261.
- Duncan, K. and Barry, C. E. (2004) *Curr. Opin. Microbiol.* **7**, 460-465.
- Dunford, H. B. and Stillman, J. S. (1976) *Coord. Chem. Rev.* **19**, 187-251.
- Dykstra, C. E., Frenking, G., Kim, K. S. and Scuseria, G. E. (2005) *Theory and Applications of Computational Chemistry: The First Forty Years*, Elsevier
- Edwards, S. L., Xuong, N. H., Hamlin, R. C. and Kraut, J. (1987) *Biochemistry* **26**, 1503-1511.
- Egawa, T., Shimada, H. and Ishimura, Y. (1994) *Biochem. Biophys. Res. Commun.* **201**, 1464-1469.
- Ekroos, M. and Sjogren, T. (2006) *Proc. Natl. Acad. Sci. USA* **103**, 13682-13687.
- Elhawari, A. M. and Plaa, G. L. (1979) *Toxicol. Appl. Pharmacol.* **48**, 445-458.
- Espinal, M. A. (2003) *Tuberculosis* **83**, 44-51.
- Estabrook, R. W., Rosenthal, O. and Cooper, D. Y. (1963) *Biochem. Z.* **338**, 741-755.
- Fair, E., Hopewell, P. C. and Pai, M. (2007) *Exp. Rev. Anti-Infect. Ther.* **5**, 61-65.
- Fenna, R., Zeng, J. and Davey, C. (1995) *Arch. Biochem. Biophys.* **316**, 653-656.

- Ferslew, K. E., Hagardorn, A. N., Harlan, G. C. and McCormick, W. F. (1998) *J. Forensic. Sci.* **43**, 1082-1085.
- Feyereisen, R. (2006) *Biochem. Soc. Trans.* **34**, 1252-1255.
- Fiedler, T. J., Davey, C. A. and Fenna, R. E. (2000) *J. Biol. Chem.* **275**, 11964-11971.
- Filatov, M., Harris, N. and Shaik, S. (1999) *J. Chem. Soc., Perkin Trans. 2* 399-410.
- Floquet, N., Marechal, J. D., Badet-Denisot, M. A., Robert, C. H., Dauchez, M. and Perahia, D. (2006) *Febs Lett.* **580**, 5130-5136.
- Fock, V. (1930) *Z. Physik.* **61**, 126
- Franke, A., Stochel, G., Jung, C. and van Eldik, R. (2004) *J. Am. Chem. Soc.* **126**, 4181-4191.
- Frisch, E., Frisch, M. J. and Trucks, G. W. (2003) *Gaussian 03 User's Reference*, Gaussian, Inc.
- Frisch, M. J., Trucks, G. W., Schlegel, H. B., Scuseria, G. E., Robb, M. A., Cheeseman, J. R., Montgomery, J., J. A., Vreven, T., Kudin, K. N., Burant, J. C., Millam, J. M., Iyengar, S. S., Tomasi, J., Barone, V., Mennucci, B., Cossi, M., Scalmani, G., Rega, N., Petersson, G. A., Nakatsuji, H., Hada, M., Ehara, M., Toyota, K., Fukuda, R., Hasegawa, J., Ishida, M., Nakajima, T., Honda, Y., Kitao, O., Nakai, H., Klene, M., Li, X., Knox, J. E., Hratchian, H. P., Cross, J. B., Bakken, V., Adamo, C., Jaramillo, J., Gomperts, R., Stratmann, R. E., Yazyev, O., Austin, A. J., Cammi, R., Pomelli, C., Ochterski, J. W., Ayala, P. Y., Morokuma, K., Voth, G. A., Salvador, P., Dannenberg, J. J., Zakrzewski, V. G., Dapprich, S., Daniels, A. D., Strain, M. C., Farkas, O., Malick, D. K., Rabuck, A. D., Raghavachari, K., Foresman, J. B., Ortiz, J. V., Cui, Q., Baboul, A. G., Clifford, S., Cioslowski, J., Stefanov, B. B., Liu, G., Liashenko, A., Piskorz, P., Komaromi, I., Martin, R. L., Fox, D. J., Keith, T., Al-Laham, M. A., Peng, C. Y., Nanayakkara, A., Challacombe, M., Gill, P. M. W., Johnson, B., Chen, W., Wong, M. W., Gonzalez, C. and Pople, J. A. (2004) *Gaussian 03, Revision C.02* Gaussian, Inc., Wallingford CT.
- Fujiia, S., Yoshimuraa, T., Kamadaa, H., Yamaguchib, K., Suzukib, S., Shidarac, S. and Takakuwa, S. (1995) *Biochim. Biophys. Acta* **1251**, 161-169.
- Fulop, V., Phizackerley, R. P., Soltis, S. M., Clifton, I. J., Wakatsuki, S., Erman, J., Hajdu, J. and Edwards, S. L. (1994) *Structure* **2**, 201-208.
- Gamielidien, J., Ptitsyn, A. and Hide, W. (2002) *Trends Gen.* **18**, 5-8.
- Garfinkel, D. (1958) *Arch. Biochem. Biophys.* **77**, 493-509.
- Gatfield, J. and Pieters, J. (2000) *Science* **288**, 1647-1650.
- Gayarthri, A. K. and Padmanaban, G. (1974) *Biochem. Pharmacol.* **23**, 2713-2725.
- George, P. (1952) *Nature* **169**, 612-613.

- Gilles-Gonzalez, M. A., Ditta, G. S. and Helinski, D. R. (1991) *Nature* **350**, 170-172.
- Girvan, H. M. (2003) *Unpublished results*
- Goldfarb, D., Bernardo, M., Thomann, H., Kroneck, P. M. H. and Ullrich, V. (1996) *J. Am. Chem. Soc.* **118**, 2686-2693.
- Göller, A. H. and Clark, T. (2001) *J. Mol. Struct. (Theochem.)* **541**, 263-281.
- Gotoh, O. (1992) *J. Biol. Chem.* **267**, 83-90.
- Green, M. T. (1998) *J. Am. Chem. Soc.* **120**, 10772-10773.
- Green, M. T. (1999) *J. Am. Chem. Soc.* **121**, 7939-7940.
- Groves, J. T. and McClusky, G. A. (1976) *J. Am. Chem. Soc.* **98**, 859-861.
- Groves, J. T. (1985) *J. Chem. Educ.* **62**, 928-931.
- Groves, J. T. and Han, Y. Z. (1995) *Cytochrome P450: Structure, Mechanism, and Biochemistry, 2nd edition* Plenum Press, New York pp. 3-48.
- Groves, J. T. (2005) *Cytochrome P450: Structure, Mechanism, and Biochemistry, 3rd edition*, Plenum Press, New York pp. 1-43.
- Guallar, V., Harris, D. L., Batista, V. S. and Miller, W. H. (2002) *J. Am. Chem. Soc.* **124**, 1430-1437.
- Guallar, V. and Friesner, R. A. (2004) *J. Am. Chem. Soc.* **126**, 8501-8508.
- Guardiola-Diaz, H. M., Foster, L. A., Mushrush, D. and Vaz, A. D. (2001) *Biochem. Pharmacol.* **61**, 1463-1470.
- Guengerich, F. P., Martin, M. V., Beaune, P. H., Kremers, P., Wolff, T. and Waxman, D. J. (1986) *J. Biol. Chem.* **261**, 5051-5060.
- Guengerich, F. P. (2001) *Chem. Res. Toxicol.* **14**, 611-650.
- Guengerich, F. P. (2005) *Cytochrome P450: Structure, Mechanism, and Biochemistry, 3rd edition*, Plenum Press, New York pp. 377-530.
- Guengrich, F. P. (1999) *Annu. Rev. Pharmacol. Toxicol.* **39**, 1-17.
- Gustafsson, M. C. U., Roitel, O., Marshall, K. R., Noble, M. A., Chapman, S. K., Pessegueiro, A., Fulco, A. J., Cheesman, M. R., von Wachenfeldt, C. and Munro, A. W. (2004) *Biochemistry* **43**, 5474-5487.
- Halpert, J. and Neal, R. A. (1980) *Mol. Pharmacol.* **17**, 427-431.
- Hanson, L. K., Eaton, W. A., Sligar, S. G., Gunsalus, I. C., Gouterman, M. and Connell, C. R. (1976) *J. Am. Chem. Soc.* **98**, 2672-2674.

- Harries, A. D. and Dye, C. (2006) *Annals Trop. Med. Parasitol.* **100**, 415-431.
- Harris, D. and Loew, G. (1993) *J. Am. Chem. Soc.* **115**, 8775-8779.
- Harris, D., Loew, G. and Waskell, L. (1998) *J. Am. Chem. Soc.* **120**, 4308-4318.
- Harris, D., Loew, G. and Waskell, L. (2001) *J. Inorg. Biochem.* **83**, 309-318.
- Harris, D. L. and Loew, G. H. (1998) *J. Am. Chem. Soc.* **120**, 8941-8948.
- Harris, D. L. (2002) *J. Inorg. Biochem.* **91**, 568-585.
- Harris, D. L., Park, J. Y., Gruenke, L. and Waskell, L. (2004) *Proteins* **55**, 895-914.
- Harris, N., Cohen, S., Filatov, M., Ogliaro, F. and Shaik, S. (2000) *Angew. Chem., Int. Ed.* **39**, 2003-2007.
- Hartree, D. R. (1928) *Proc. Cambridge Phil. Soc.* **24**, 89.
- Hasemann, C. A., Kurumbail, R. G., Boddupalli, S. S., Peterson, J. A. and Deisenhofer, J. (1995) *Structure* **3**, 41-62.
- Hay, P. J. and Wadt, W. R. (1985a) *J. Chem. Phys.* **82**, 270-283.
- Hay, P. J. and Wadt, W. R. (1985b) *J. Chem. Phys.* **82**, 299-310.
- Hayaishi, O., Katagiri, M. and Rothberg, S. (1955) *J. Am. Chem. Soc.* **77**, 5450-5451.
- Hedegaar, J. and Gunsalus, I. C. (1965) *J. Biol. Chem.* **240**, 4038-4043.
- Heisenberg, W. (1927) *Z. Phys.* **43**, 172.
- Hildebrandt, A., Remmer, H. and Estabrook, R. W. (1968) *Biochem. Biophys. Res. Commun.* **30**, 607-612.
- Hirao, H., Kumar, D., Thiel, W., Shaik, S., de Visser, S. P., Sharma, P. K. and Derat, E. (2005) *J. Am. Chem. Soc.* **127**, 13007-13018.
- Hishiki, T., Shimada, H., Nagano, S., Egawa, T., Kanamori, Y., Makino, R., Park, S. Y., Adachi, S., Shiro, Y. and Ishimura, Y. (2000) *J. Biochem.* **128**, 965-974.
- Hochuli, E., Bannwarth, W., Dobeli, H., Gentz, R. and Stuber, D. (1988) *Nat. Biotechnol.* **6**, 1321-1325.
- Hohenberg, P. and Kohn, W. (1964) *Phys. Rev.* **136**, B864 - B871.
- Hollenberg, P. F. (2002) *Drug Met. Rev.* **34**, 17-35.
- Idemyor, V. (2007) *J. Natl. Med. Assoc.* **99**, 1414-1419.

- Jenkins, C. M. and Waterman, M. R. (1998) *Biochemistry* **37**, 6106-6113.
- Jensen, F. (1999) *Introduction to Computational Chemistry*, John Wiley & Sons pp. 172-173.
- Jensen, K. P. and Ryde, U. (2004) *J. Biol. Chem.* **279**, 14561-14569.
- Johnson, C. R. and Shepherd, R. E. (1983) *Inorg. Chem.* **22**, 3506-3513.
- Johnson, C. R., Jones, C. M., Asher, S. A. and Abola, J. E. (1991) *Inorg. Chem.* **30**, 2120-2129.
- Johnson, R., Streicher, E. M., Louw, G. E., Warren, R. M., van Helden, P. D. and Victor, T. C. (2006) *Curr. Iss. Mol. Biol.* **8**, 97-111.
- Kamachi, T. and Yoshizawa, K. (2003) *J. Am. Chem. Soc.* **125**, 4652-4661.
- Katagiri, M., Ganguli, B. N. and Gunsalus, I. C. (1968) *J. Biol. Chem.* **243**, 3543-3546.
- Kato, T. (1951) *Trans. Am. Math. Soc.* **70**, 195-211.
- Keller, R. M., Debrunner, P. G. and Wuthrich, K. (1972) *Proc. Natl. Acad. Sci. USA* **69**, 2073-2075.
- Kellner, D. G., Hung, S. C., Weiss, K. E. and Sligar, S. G. (2002) *J. Biol. Chem.* **277**, 9641-9644.
- Kendall, S. L., Rison, S. C. G., Movahedzadeh, F., Frita, R. and Stoker, N. G. (2004) *Trends Microbiol.* **12**, 537-544.
- Kiselev, A. A. (1970) *J. Phys. B: Atom. Molec. Phys.* **3**, 904-915.
- Kitada, M., Chiba, K., Kamataki, T. and Kitagawa, H. (1977) *Jpn. J. Pharmacol.* **27**, 601-608.
- Klingenberg, M. (1958) *Arch. Biochem. Biophys.* **75**, 376-386.
- Ko, T. P., Safo, M. K., Musayev, F. N., Di Salvo, M. L., Wang, C. Q., Wu, S. H. and Abraham, D. J. (2000) *Acta Crystallograph. D* **56**, 241-245.
- Kobayashi, K., Iwamoto, T. and Honda, K. (1994) *Biochem. Biophys. Res. Commun.* **201**, 1348-1355.
- Kohn, W. and Sham, L. J. (1965) *Phys. Rev.* **140**, A1133 - A1138.
- Kudo, K., Imamura, T., Jitsufuchi, N., Zhang, X. X., Tokunaga, H. and Nagata, T. (1997) *Forensic Sci. Int.* **86**, 35-41.
- Kudo, S., Okumura, H., Miyamoto, G. and Ishizaki, T. (1999) *Drug. Metab. Dispos.* **27**, 303-308.

- Kumar, D., de Visser, S. P. and Shaik, S. (2005) *Chem. Eur. J.* **11**, 2825-2835.
- Kyle, A. A. and Dahl, M. V. (2004) *Am. J. Clin. Dermatol.* **5**, 443-451.
- Lamar, G. N. and Walker, F. A. (1972) *J. Am. Chem. Soc.* **94**, 8607-8608.
- Lamb, D. C., Kelly, D. E., Baldwin, B. C. and Kelly, S. L. (2000) *Chem. Biol. Interact.* **125**, 165-175.
- Larrey, D., Tinel, M. and Pessayre, D. (1983) *Biochem. Pharmacol.* **32**, 1487-1493.
- Lawson, R. J., Leys, D., Sutcliffe, M. J., Kemp, C. A., Cheesman, M. R., Smith, S. J., Clarkson, J., Smith, W. E., Haq, I., Perkins, J. B. and Munro, A. W. (2004) *Biochemistry* **43**, 12410-12426.
- Leach, A. R. (2001) *Molecular Modelling, Principles and Applications*, 2nd edition, Pearson Prentice Hall
- Lee, C. T., Yang, W. T. and Parr, R. G. (1988) *Phys. Rev. B* **37**, 785-789.
- Letteron, P., Descatoire, V., Larrey, D., Tinel, M., Geneve, J. and Pessayre, D. (1986) *J. Pharmacol. Exp. Ther.* **238**, 685-692.
- Lewars, E. (2003) *Computational Chemistry: Introduction to the Theory and Applications of Molecular and Quantum Mechanics*, Kluwer Academic Publishers
- Leys, D., Mowat, C. G., McLean, K. J., Richmond, A., Chapman, S. K., Walkinshaw, M. D. and Munro, A. W. (2003) *J. Biol. Chem.* **278**, 5141-5147.
- Li, H. Y. and Poulos, T. L. (1997) *Nat. Struct. Biol.* **4**, 140-146.
- Lin, J. H. and Lu, A. Y. (1998) *Clin. Pharmacokinet.* **35**, 361-390.
- Liu, X. S., Kim, C. N., Yang, J., Jemmerson, R. and Wang, X. D. (1996) *Cell* **86**, 147-157.
- Lowenstein, C. J. and Snyder, S. H. (1992) *Cell* **70**, 705-707.
- Lukin, J. A. and Ho, C. (2004) *Chem. Rev.* **104**, 1219-1230.
- Lukton, D., Mackie, J. E., Lee, J. S., Marks, G. S. and Ortiz de Montellano, P. R. (1988) *Chem. Res. Toxicol.* **1**, 208-215.
- Makris, T., Denisov, I., Schlichting, I. and Sligar, S. (2005) *Cytochrome P450: Structure, Mechanism, and Biochemistry*, 3rd edition, Plenum Press, New York pp. 149-182.
- Manchester, J. I., Dinnocenzo, J. P., Higgins, L. A. and Jones, J. P. (1997) *J. Am. Chem. Soc.* **119**, 5069-5070.
- Masubuchi, Y. and Horie, T. (2007) *Crit. Rev. Toxicol.* **37**, 389-412.

- McIver, L., Leadbeater, C., Campopiano, D. J., Baxter, R. L., Daff, S. N., Chapman, S. K. and Munro, A. W. (1998) *Eur. J. Biochem.* **257**, 577-585.
- McLean, K. J., Cheesman, M. R., Rivers, S. L., Richmond, A., Leys, D., Chapman, S. K., Reid, G. A., Price, N. C., Kelly, S. M., Clarkson, J., Smith, W. E. and Munro, A. W. (2002a) *J. Inorg. Biochem.* **91**, 527-541.
- McLean, K. J., Marshall, K. R., Richmond, A., Hunter, I. S., Fowler, K., Kieser, T., Gurucha, S. S., Besra, G. S. and Munro, A. W. (2002b) *Microbiology* **148**, 2937-2949.
- McLean, K. J., Sabri, M., Marshall, K. R., Lawson, R. J., Lewis, D. G., Clift, D., Balding, P. R., Dunford, A. J., Warman, A. J., McVey, J. P., Quinn, A. M., Sutcliffe, M. J., Scrutton, N. S. and Munro, A. W. (2005) *Biochem. Soc. Trans.* **33**, 796-801.
- McLean, K. J., Clift, D., Lewis, D. G., Sabri, M., Balding, P. R., Sutcliffe, M. J., Leys, D. and Munro, A. W. (2006a) *Trends Microbiol.* **14**, 220-228.
- McLean, K. J., Dunford, A. J., Sabri, M., Neeli, R., Girvan, H. M., Balding, P. R., Leys, D., Seward, H. E., Marshall, K. R. and Munro, A. W. (2006b) *Biochem. Soc. Trans.* **34**, 1178-1182.
- McLean, K. J., Warman, A. J., Seward, H. E., Marshall, K. R., Girvan, H. M., Cheesman, M. R., Waterman, M. R. and Munro, A. W. (2006c) *Biochemistry* **45**, 8427-8443.
- McLean, K. J., Dunford, A. J., Neeli, R., Driscoll, M. D. and Munro, A. W. (2007) *Arch. Biochem. Biophys.* **464**, 228-240.
- Mennucci, B., Cancès, E. and Tomasi, J. (1997) *J. Phys. Chem. B* **101**, 10506-10517.
- Mennucci, B. and Tomasi, J. (1997) *J. Chem. Phys.* **106**, 5151-5158.
- Meunier, B. (1992) *Chem. Rev.* **92**, 1411-1456.
- Meunier, B., De Visser, S. P. and Shaik, S. (2004) *Chem. Rev.* **104**, 3947-3980.
- Miertus, S., Scrocco, E. and Tomasi, J. (1981) *Chem. Phys.* **55**, 117-129.
- Miertus, S. and Tomasi, J. (1982) *Chem. Phys.* **65**, 239-245.
- Miles, J. S., Munro, A. W., Rospendowski, B. N., Smith, W. E., McKnight, J. and Thomson, A. J. (1992) *Biochem. J.* **288**, 503-509.
- Mogi, T., Saiki, K. and Anraku, Y. (1994) *Mol. Microbiol.* **14**, 391-398.
- Mueller, G. C. and Miller, J. A. (1948) *J. Biol. Chem.* **176**, 535-544.
- Munro, A. W., Leys, D. G., McLean, K. J., Marshall, K. R., Ost, T. W., Daff, S., Miles, C. S., Chapman, S. K., Lysek, D. A., Moser, C. C., Page, C. C. and Dutton, P. L. (2002) *Trends Biochem. Sci.* **27**, 250-257.

- Munro, A. W., McLean, K. J., Marshall, K. R., Warman, A. J., Lewis, G., Roitel, O., Sutcliffe, M. J., Kemp, C. A., Modi, S., Scrutton, N. S. and Leys, D. (2003) *Biochem. Soc. Trans.* **31**, 625-630.
- Munro, A. W., Girvan, H. M., McLean, K. J., Cheesman, M. J. and Leys, D. (2007) *Tetrapyrroles: Birth, Life and Death, Landes Bioscience, Edited by Warren, M. J. and Smith, A. G.* pp. 411-434.
- Murayama, N., Imai, N., Nakane, T., Shimizu, M. and Yamazaki, H. (2007) *Biochem. Pharmacol.* **73**, 2020-2026.
- Murray, M. and Reidy, G. F. (1990) *Pharmacol. Rev.* **42**, 85-101.
- Murshudov, G. N., Grebenko, A. I., Barynin, V., Dauter, Z., Wilson, K. S., Vainshtein, B. K., Melik-Adamyanyan, W., Bravo, J., Ferrán, J. M., Ferrer, J. C., Switala, J., Loewen, P. C. and Fita, I. (1996) *J. Biol. Chem.* **271**, 8863-8868.
- Nappa, M., Valentine, J. S. and Snyder, P. A. (1977) *J. Am. Chem. Soc.* **99**, 5799-5800.
- Nelson, E. (1966) *Phys. Rev.* **150**, 1079-1085.
- Newcomb, M., Le Tadic-Biadatti, M. H., Chestney, D. L., Roberts, E. S. and Hollenberg, P. F. (1995a) *J. Am. Chem. Soc.* **117**, 12085-12091.
- Newcomb, M., Letadic, M. H., Putt, D. A. and Hollenberg, P. F. (1995b) *J. Am. Chem. Soc.* **117**, 3312-3313.
- Newcomb, M., Shen, R., Choi, S. Y., Toy, P. H., Hollenberg, P. F., Vaz, A. D. N. and Coon, M. J. (2000) *J. Am. Chem. Soc.* **122**, 2677-2686.
- Newcomb, M. and Toy, P. H. (2000) *Acc. Chem. Res.* **33**, 449-455.
- Newcomb, M., Hollenberg, P. F. and Coon, M. J. (2003) *Arch. Biochem. Biophys.* **409**, 72-79.
- Newcomb, M. and Chandrasena, R. E. P. (2005) *Biochem. Biophys. Res. Commun.* **338**, 394-403.
- Newcomb, M., Zhang, R., Chandrasena, R. E. P., Halgrimson, J. A., Horner, J. H., Makris, T. M. and Sligar, S. G. (2006) *J. Am. Chem. Soc.* **128**, 4580-4581.
- Noble, M. A., Miles, C. S., Chapman, S. K., Lysek, D. A., MacKay, A. C., Reid, G. A., Hanzlik, R. P. and Munro, A. W. (1999) *Biochem. J.* **339**, 371-379.
- O'Brian, M. R. (1999) *Iron Metabolism*, Wiley-VCH pp. 1-13.
- Odds, F. C., Brown, A. J. and Gow, N. A. (2003) *Trends Microbiol.* **11**, 272-279.
- Ogliaro, F., Cohen, S., de Visser, S. P. and Shaik, S. (2000a) *J. Am. Chem. Soc.* **122**, 12892-12893.

- Ogliaro, F., Cohen, S., Filatov, M., Harris, N. and Shaik, S. (2000b) *Angew. Chem., Int. Ed.* **39**, 3851-3855.
- Ogliaro, F., Harris, N., Cohen, S., Filatov, M., de Visser, S. P. and Shaik, S. (2000c) *J. Am. Chem. Soc.* **122**, 8977-8989.
- Ogliaro, F., de Visser, S. P., Groves, J. T. and Shaik, S. (2001) *Angew. Chem., Int. Ed.* **40**, 2874-2878.
- Ogliaro, F., de Visser, S. P., Cohen, S., Sharma, P. K. and Shaik, S. (2002a) *J. Am. Chem. Soc.* **124**, 2806-2817.
- Ogliaro, F., de Visser, S. P. and Shaik, S. (2002b) *J. Inorg. Biochem.* **91**, 554-567.
- Ohta, T., Matsuura, K., Yoshizawa, K. and Morishima, I. (2000) *J. Inorg. Biochem.* **82**, 141-152.
- Okita, R. T., Clark, J. E., Okita, J. R. and Masters, B. S. S. (1991) *Methods Enzymol.* **206**, 432-441.
- Omura, T. and Sato, R. (1962) *J. Biol. Chem.* **237**, 1375-1376.
- Omura, T. and Sato, R. (1964) *J. Biol. Chem.* **239**, 2370-2378.
- Ormerod, L. P. (2005) *Brit. Med. Bull.* **73 and 74**, 17-24.
- Ortiz de Montellano, P. R. and Mico, B. A. (1980) *Mol. Pharmacol.* **18**, 128-135.
- Ortiz de Montellano, P. R. and Komives, E. A. (1985) *J. Biol. Chem.* **260**, 3330-3336.
- Ortiz de Montellano, P. R. and de Voss, J. (2005) *Cytochrome P450: Structure, Mechanism, and Biochemistry, 3rd edition*, Plenum Press, New York pp. 183-245.
- Ouellet, H., Podust, L. M. and Ortiz de Montellano, P. R. (2008) *J. Biol. Chem.* **283**, 5069-5080.
- Oxvig, C., Thomsen, A. R., Overgaard, M. T., Sorensen, E. S., Hojrup, P., Bjerrum, M. J., Gleich, G. J. and Sottrup-Jensen, L. (1999) *J. Biol. Chem.* **274**, 16953-16958.
- Paine, M., Scrutton, N. S., Munro, A. W., Gutierrez, A., Roberts, G. and Wolf, C. (2005) *Cytochrome P450: Structure, Mechanism, and Biochemistry, 3rd edition*, Plenum Press, New York pp. 115-148.
- Park, J. Y., Shon, J. H., Kim, K. A., Jung, H. J., Shim, J. C., Yoon, Y. R., Cha, I. J. and Shin, J. G. (2006) *J. Clin. Psycho. Pharmacol.* **26**, 135-142.
- Pasternack, R. F., Gillies, B. S. and Stahlbush, J. R. (1978) *J. Am. Chem. Soc.* **100**, 2613-2619.
- Pealing, S. L., Cheesman, M. R., Reid, G. A., Thomson, A. J., Ward, F. B. and Chapman, S. K. (1995) *Biochemistry* **34**, 6153-6158.

- Peisach, J., Blumberg, W. E., Wittenberg, B. A. and Wittenberg, J. B. (1968) *J. Biol. Chem.* **243**, 1871-1880.
- Peng, C., Ayala, P. Y., Schlegel, H. B. and Frisch, M. J. (1996) *J. Comp. Chem.* **17**, 49-56.
- Pennerhahn, J. E., McMurry, T. J., Renner, M., Latosgrazynsky, L., Eble, K. S., Davis, I. M., Balch, A. L., Groves, J. T., Dawson, J. H. and Hodgson, K. O. (1983) *J. Biol. Chem.* **258**, 12761-12764.
- Perdew, J. P. (1986) *Phys. Rev. B* **33**, 8822-8824.
- Perera, R., Sono, M., Sigman, J. A., Pfister, T. D., Lu, Y. and Dawson, J. H. (2003) *Proc. Natl. Acad. Sci. U. S. A.* **100**, 3641-3646.
- Perutz, M. F. (1978) *Sci. Am.* **239**, 92-125.
- Peterson, J. A. (1971) *Arch. Biochem. Biophys.* **144**, 678-693.
- Plano, G. V. (2004) *Mol. Microbiol.* **54**, 287-290.
- Podust, L. M., Poulos, T. L. and Waterman, M. R. (2001) *Proc. Natl. Acad. Sci. USA* **98**, 3068-3073.
- Poulos, T. L., Finzel, B. C. and Howard, A. J. (1986) *Biochemistry* **25**, 5314-5322.
- Poulos, T. L., Finzel, B. C. and Howard, A. J. (1987) *J. Mol. Biol.* **195**, 687-700.
- Poulos, T. L. and Johnson, T. (2005) *Cytochrome P450: Structure, Mechanism, and Biochemistry, 3rd edition*, Plenum Press, New York pp. 87-114.
- Quaroni, L. G., Seward, H. E., McLean, K. J., Girvan, H. M., Ost, T. W. B., Noble, M. A., Kelly, S. M., Price, N. C., Cheesman, M. R., Smith, W. E. and Munro, A. W. (2004) *Biochemistry* **43**, 16416-16431.
- Rae, T. D. and Goff, H. M. (1998) *J. Biol. Chem.* **273**, 27968-27977.
- Rassolov, V. A., Pople, J. A., Ratner, M. A. and Windus, T. L. (1998) *J. Chem. Phys.* **109**, 1223-1229.
- Ravichandran, K. G., Boddupalli, S. S., Hasemann, C. A., Peterson, J. A. and Deisenhofer, J. (1993) *Science* **261**, 731-736.
- Recchi, C., Sclavi, B., Rauzier, J., Gicquel, B. and Reytrat, J. M. (2003) *J. Biol. Chem.* **278**, 33763-33773.
- Reed, A. E. and Weinhold, F. (1983) *J. Chem. Phys.* **78**, 4066-4073.
- Roberts, J. E., Hoffman, B. M., Rutter, R. and Hager, L. P. (1981) *J. Am. Chem. Soc.* **103**, 7654-7656.

- Rodriguez-Lopez, J. N., Lowe, D. J., Hernandez-Ruiz, J., Hiner, A. N. P., Garcia-Canovas, F. and Thorneley, R. N. F. (2001) *J. Am. Chem. Soc.* **123**, 11838-11847.
- Russell, D. G. (2007) *Nat. Rev. Microbiol.* **5**, 39-47.
- Rutter, R., Hager, L. P., Dhonau, H., Hendrich, M., Valentine, M. and Debrunner, P. (1984) *Biochemistry* **23**, 6809-6816.
- Rydberg, P., Sigfridsson, E. and Ryde, U. (2004) *J. Biol. Inorg. Chem.* **9**, 203-223.
- Saberi, M. R., Vinh, T. K., Yee, S. W., Griffiths, B. J. N., Evans, P. J. and Simons, C. (2006) *J. Med. Chem.* **49**, 1016-1022.
- Sansen, S., Yano, J. K., Reynald, R. L., Schoch, G. A., Griffin, K. J., Stout, C. D. and Johnson, E. F. (2007) *J. Biol. Chem.* **282**, 14348-14355.
- Sassetti, C. M., Boyd, D. H. and Rubin, E. J. (2001) *Proc. Natl. Acad. Sci. USA* **98**, 12712-12717.
- Sassetti, C. M., Boyd, D. H. and Rubin, E. J. (2003) *Mol. Microbiol.* **48**, 77-84.
- Sato, A. and Nakajima, T. (1979) *Toxicol. Appl. Pharmacol.* **48**, 249-256.
- Schappacher, M., Ricard, L., Fischer, J., Weiss, R., Bill, E., Montielmontoya, R., Winkler, H. and Trautwein, A. X. (1987) *Eur. J. Biochem.* **168**, 419-429.
- Schenkman, J. B., Remmer, H. and Estabrook, R. W. (1967) *Mol. Pharmacol.* **3**, 113-123.
- Schenkman, J. B., Sligar, S. G. and Cinti, D. L. (1981) *Pharmacol. Ther.* **12**, 43-71.
- Schenkman, J. B. and Jansson, I. (2003) *Pharmacol. Ther.* **97**, 139-152.
- Schenkman, J. B. and Jansson, I. (2006) *Cytochrome P450 Protocols: Spectral Analyses of Cytochromes P450* Humana Press pp. 11-18.
- Schlichting, I., Berendzen, J., Chu, K., Stock, A. M., Maves, S. A., Benson, D. E., Sweet, R. M., Ringe, D., Petsko, G. A. and Sligar, S. G. (2000) *Science* **287**, 1615-1622.
- Schonbaum, G. and Lo, S. (1972) *J. Biol. Chem.* **247**, 3353-3360.
- Schöneboom, J. C., Lin, H., Reuter, N., Thiel, W., Cohen, S., Ogliaro, F. and Shaik, S. (2002) *J. Am. Chem. Soc.* **124**, 8142-8151.
- Schöneboom, J. C. and Thiel, W. (2004) *J. Phys. Chem. B* **108**, 7468-7478.
- Schrödinger, E. (1926) *Phys. Rev. A* **28**, 1049 - 1070.
- Schunemann, V., Trautwein, A. X., Jung, C. and Turner, J. (2002) *Hyperfine Interactions* **141**, 279-284.

- Schuster, I. (1985) *Xenobiotica* **15**, 529-546.
- Senn, H. M. and Thiel, W. (2007) *Curr. Opin. Chem. Biol.* **11**, 182-187.
- Seward, H. E., Roujeinikova, A., McLean, K. J., Munro, A. W. and Leys, D. (2006) *J. Biol. Chem.* 39437-39443.
- Shaik, S., de Visser, S. P., Ogliaro, F., Schwarz, H. and Schroder, D. (2002) *Curr. Opin. Chem. Biol.* **6**, 556-567.
- Shaik, S., de Visser, S. P. and Kumar, D. (2004) *J. Biol. Inorg. Chem.* **9**, 661-668.
- Shaik, S. and de Visser, S. P. (2005) *Cytochrome P450: Structure, Mechanism, and Biochemistry*, 3rd edition, Plenum Press, New York pp. 45-85.
- Shaik, S., Kumar, D., de Visser, S. P., Altun, A. and Thiel, W. (2005a) *Chem. Rev.* **105**, 2279-2328.
- Shaik, S., Kumar, D., de Visser, S. P., Altun, A. and Thiel, W. (2005b) *Chem Rev* **105**, 2279-328.
- Sharrock, M., Debrunner, P. G., Schulz, C., Lipscomb, J. D., Marshall, V. and Gunsalus, I. C. (1976) *Biochim. Biophys. Acta* **420**, 8-26.
- Simon, S., Duran, M. and Dannenberg, J. J. (1996) *J. Chem. Phys.* **105**, 11024-11031.
- Skulachev, V. P. (1998) *FEBS Lett.* **423**, 275-280.
- Slater, J. C. (1930) *Phys. Rev.* **36**, 57-64.
- Smith, D. B. and Johnson, K. S. (1988) *Gene* **67**, 31-40.
- Smith, S. J., Munro, A. W. and Smith, W. E. (2003) *Biopolymers* **70**, 620-627.
- Sousa, E. H. S., Tuckerman, J. R., Gonzalez, G. and Gilles-Gonzalez, M. A. (2007) *Prot. Sci.* **16**, 1708-1719.
- Souter, A., McLean, K. J., Smith, W. E. and Munro, A. W. (2000) *J. Chem. Technol. Biotechnol.* **75**, 933-941.
- Spolitak, T., Dawson, J. H. and Ballou, D. P. (2005) *J. Biol. Chem.* **280**, 20300-20309.
- Stewart, G. R., Wernisch, L., Stabler, R., Mangan, J. A., Hinds, J., Laing, K. G., Young, D. B. and Butcher, P. D. (2002) *Microbiology* **148**, 3129-3138.
- Strickland, N. and Harvey, J. N. (2007) *J. Phys. Chem. B* **111**, 841-852.
- Temple, T. E. and Liddle, G. W. (1970) *Annu. Rev. Pharmacol. Toxicol.* **10**, 199-218.

Terwilliger, T. C., Park, M. S., Waldo, G. S., Berendzen, J., Hung, L. W., Kim, C. Y., Smith, C. V., Sacchettini, J. C., Bellinzoni, M., Bossi, R., De Rossi, E., Mattevi, A., Milano, A., Riccardi, G., Rizzi, M., Roberts, M. M., Coker, A. R., Fossati, G., Mascagni, P., Coates, A. R., Wood, S. P., Goulding, C. W., Apostol, M. I., Anderson, D. H., Gill, H. S., Eisenberg, D. S., Taneja, B., Mande, S., Pohl, E., Lamzin, V., Tucker, P., Wilmanns, M., Colovos, C., Meyer-Klaucke, W., Munro, A. W., McLean, K. J., Marshall, K. R., Leys, D., Yang, J. K., Yoon, H. J., Lee, B. I., Lee, M. G., Kwak, J. E., Han, B. W., Lee, J. Y., Baek, S. H., Suh, S. W., Komen, M. M., Arcus, V. L., Baker, E. N., Lott, J. S., Jacobs, W., Jr., Alber, T. and Rupp, B. (2003) *Tuberculosis* **83**, 223-249.

Testa, B. and Jenner, P. (1981) *Drug Met. Rev.* **12**, 1-117.

Tsai, R., Yu, C. A., Gunsalus, I. C., Peisach, J., Blumberg, W., Ormejohn, W. and Beinert, H. (1970) *Proc. Natl. Acad. Sci. USA* **66**, 1157-1163.

Uno, T., Hatano, K., Nawa, T., Nakamura, K., Nishimura, Y. and Arata, Y. (1991) *Inorg. Chem.* **30**, 4322-4327.

Van der Geize, R., Yam, K., Heuser, T., Wilbrink, M. H., Hara, H., Anderton, M. C., Sim, E., Dijkhuizen, L., Davies, J. E., Mohn, W. W. and Eltis, L. D. (2007) *Proc. Natl. Acad. Sci. USA* **104**, 1947-1952.

van Duijneveldt, F. B., van Duijneveldt-van de Rijdt, J. G. C. M. and van Lenthe, J. H. (1994) *Chem. Rev.* **94**, 1873-1885.

Vanden Bossche, H. (1985) *Curr. Top. Med. Mycol.* **1**, 313-351.

Vantwout, J. W., Mattie, H. and Vanfurth, R. (1988) *J. Antimicrob. Chemother.* **21**, 665-672.

Varhe, A., Olkkola, K. T. and Neuvonen, P. J. (1994) *Clin. Pharmacol. Ther.* **56**, 601-607.

Verma, A., Hirsch, D. J., Glatt, C. E., Ronnett, G. V. and Snyder, S. H. (1993) *Science* **259**, 381-384.

Verras, A., Alian, A. and Ortiz de Montellano, P. R. (2006) *Prot. Eng. Des. Sel.* **19**, 491-496.

Verras, A. and Ortiz de Montellano, P. R. (2006) *Biochem. Soc. Trans.* **34**, 1170-1172.

Wadt, W. R. and Hay, P. J. (1985) *J. Chem. Phys.* **82**, 284-298.

Walker, F. A., Lo, M. W. and Ree, M. T. (1976) *J. Am. Chem. Soc.* **98**, 5552-5560.

Wang, R. W., Kari, P. H., Lu, A. Y. H., Thomas, P. E., Guengerich, F. P. and Vyas, K. P. (1991) *Arch. Biochem. Biophys.* **290**, 355-361.

Wang, Y., Wang, H. M., Wang, Y. H., Yang, C. L., Yang, L. and Han, K. L. (2006) *J. Phys. Chem. B* **110**, 6154-6159.

- Watanabe, Y. (2002) *Curr. Opin. Chem. Biol.* **6**, 208-219.
- Weichsel, A., Maes, E. M., Andersen, J. F., Valenzuela, J. G., Shokhireva, T. K., Walker, F. A. and Montfort, W. R. (2005) *Proc. Natl. Acad. Sci. U. S. A.* **102**, 594-599.
- White, R. E. and Coon, M. J. (1980) *Annu. Rev. Biochem.* **49**, 315-356.
- World Health Organization (1993) <http://www.who.int/tb/en/>
- WHO (2002) <http://www.who.int/mediacentre/factsheets/who104/en/>
- World Health Organization fact sheet on tuberculosis (2007) www.who.int/mediacentre/factsheets/fs104/en/index.html
- World Health Organization Report (2008) *Global Tuberculosis Control, Surveillance, Planning, Financing*, www.who.int/tb/publications/global_report/2008/pdf
- Williams, P. A., Cosme, J., Ward, A., Angova, H. C., Vinkovic, D. M. and Jhoti, H. (2003) *Nature* **424**, 464-468.
- Wink, D. A., Osawa, Y., Darbyshire, J. F., Jones, C. R., Eshenaur, S. C. and Nims, R. W. (1993) *Arch. Biochem. Biophys.* **300**, 115-123.
- Woggon, W. D. (1997) *Top. Curr. Chem.* **184**, 39-96.
- Xia, D., Yu, C. A., Kim, H., Xian, J. Z., Kachurin, A. M., Zhang, L., Yu, L. and Deisenhofer, J. (1997) *Science* **277**, 60-66.
- Yamanaka, T. (1988) *Bioactive molecules*, Elsevier pp. 95-99.
- Yang, T. C., McNaughton, R. L., Clay, M. D., Jenney, F. E., Krishnan, R., Kurtz, D. M., Adams, M. W. W., Johnson, M. K. and Hoffman, B. M. (2006) *Journal of the American Chemical Society* **128**, 16566-16578.
- Yano, J. K., Wester, M. R., Schoch, G. A., Griffin, K. J., Stout, C. D. and Johnson, E. F. (2004) *J. Biol. Chem.* **279**, 38091-38094.
- Yazdanpanah, M. J., Azizi, H. and Suizi, B. (2007) *Mycoses* **50**, 311-313.
- Yeh, S. R. and Han, S. (1998) *Acc. Chem. Res.* **31**, 727-735.
- Young, D. (2001) *Computational Chemistry: A Practical Guide for Applying Techniques to Real World Problems*, John Wiley & Sons Chapter 17 - Finding Transition States.
- Yukl, E. T., Ioanoviciu, A., Ortiz de Montellano, P. R. and Moenne-Loccoz, P. (2007) *Biochemistry* **46**, 9728-9736.
- Zager, E. M. and McNerney, R. (2008) *BMC Infect. Dis.* **8**, 1-5.
- Zhang, W., Ramamoorthy, Y., Kilicarslan, T., Nolte, H., Tyndale, R. F. and Sellers, E. M. (2002) *Drug Metab. Dispos.* **30**, 314-318.

Zhang, Y. (2005) *Annu. Rev. Pharmacol. Toxicol.* **45**, 529-564.

Zhang, Z. P., Li, Y., Stearns, R. A., Ortiz de Montellano, P. R., Baillie, T. A. and Tang, W. (2002) *Biochemistry* **41**, 2712-2718.

Zhao, Y. H., White, M. A., Muralidhara, B. K., Sun, L., Halpert, J. R. and Stout, C. D. (2006) *J. Biol. Chem.* **281**, 5973-5981.

Zheng, J. J., Wang, D. Q., Thiel, W. and Shaik, S. (2006) *J. Am. Chem. Soc.* **128**, 13204-13215.

THE
JOHN RYLANDS
UNIVERSITY
LIBRARY