

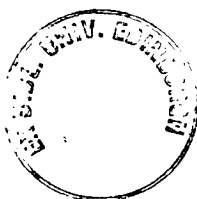
Structure, Modelling and Biotransformations of the  
Cytochrome P-450cam Y96A  
Mutant Enzyme

by ADAM G STAINES

Submitted for Doctor of Philosophy

University of Edinburgh

2000



**For My Parents...**

**"Anyone who has never made a mistake has never tried anything new."**

**"Science is a wonderful thing if one does not have to earn one's living at it."**

**Albert Einstein.**



## **Declaration.**

This thesis is submitted in part fulfilment of the requirement for the degree of Doctor of Philosophy, at the University of Edinburgh. The following thesis is an original work composed by the undersigned author unless cited otherwise. This work has not been submitted in part, or in whole for any degree.

## Abstract.

Cytochrome P-450 monooxygenases catalyse the insertion of molecular oxygen into unactivated carbon-hydrogen bonds. This reaction is difficult to reproduce synthetically; therefore cytochrome P-450 enzymes are of great interest as potential biocatalysts. Of all the cytochrome P-450 systems investigated to-date cytochrome P-450cam from, *Pseudomonas Putida*, is probably the best characterized. The use of cytochrome P-450cam, as a biocatalyst, is hampered by the narrow substrate range. This was overcome by use of the cytochrome P-450cam Y96A mutant. To identify the substrate range of the enzyme a model of the cytochrome P-450cam binding site was generated and potential substrates were modelled. However, discrepancies between predicted and observed results necessitated a more rigorous study of the mutant binding site. Conditions were developed and the cytochrome P-450cam mutant enzyme was crystallized in the presence of a number of substrates. Six different substrate bound structures were then solved by X-ray crystallography. With a crystal structure of the binding site, it was then possible to perform more rigorous molecular modelling; this was achieved with the modelling program DOCK. Biotransformations and binding studies were performed with a range of substrates, to test the validity of the modelling and to characterize the substrate specificity of the cytochrome P-450cam Y96A mutant.

## Acknowledgements.

I would like to thank the following people.

Professor Sabine Flitsch for all her support and invaluable advice over the past 4 years and for correcting my English. Professor Malcolm Walkinshaw for providing me with the opportunity to do X-ray crystallography.

I would like to thank Dr. Luet-Lok Wong for providing the P-450cam C334A, P-450cam C334-Y96A and Pd clones and also for providing the purified PdR. I also would like to thank Professor Stephen Sligar for providing the original wild-type P-450cam, Pd and PdR clones. I wish to thank Marina Alexeeva, Dr. Lisa McIver and Dr. Andy Munro for providing reductases. I wish to thank Suzanne Aitken and Dr. Gideon Grogan and Dr. Andy Willetts for providing me with substrates.

For the crystallography I wish to thank Professor Malcolm Walkinshaw for his advice over the last two years, Dr. Paul Taylor for his help and for solving the first data set. I would also like to thank George, Jackie and all the members of the structural biology group for their assistance. A special thankyou to Dr. Dima Alexeev for the many late Friday nights, helping me collect the data and to Marina for letting him stay out past his bed time.

A big thank you to Suzanne, Gareth and Gideon for proof-reading the thesis, without their help it might never have got finished.

Thanks to all the members of Lab 120, 34 and 209 for all their help over the years especially Gareth, Gideon, Marina, Nick and Suzanne.

A special thank you to Aileen for tolerating me whilst I was writing up.

Big thank you to my parents for many years of funding and convincing me not to do biology.

I wish to thank the BBSRC for funding.

Thank you serendipity for without you I would be lost...

# Contents.

<b>TITLE</b> .....	<b>I</b>
<b>DECLARATION</b> .....	<b>II</b>
<b>ABSTRACT</b> .....	<b>III</b>
<b>ACKNOWLEDGEMENTS</b> .....	<b>IV</b>
<b>CONTENTS</b> .....	<b>V</b>
<b>ABBREVIATIONS</b> .....	<b>XI</b>
<b>1. INTRODUCTION</b> .....	<b>1</b>
1.1 OXYGENASES.....	1
1.2 MONOOXYGENASES.....	2
1.3 CYTOCHROME P-450.....	3
1.4 CYTOCHROME P-450CAM MONOOXYGENASE.....	7
1.4.1 <i>Why Do We Study Cytochrome P-450cam?</i> .....	9
1.4.2 <i>P-450cam Catalytic Cycle</i> .....	10
1.4.2.1 Mechanism of Oxidation.....	10
1.4.2.2 Radical Rebound Mechanism.....	12
1.4.2.3 Alternative Theories of the Catalytic Cycle (Recent Advances).....	15
1.4.2.3.1 Role of Sixth Water Ligand.....	15
1.4.2.3.2 Oxidation State of the Iron and the Active Species.....	15
1.4.2.3.3 The Alternative Catalytic Pathway: the Concerted Carbocation Mechanism.....	16
1.4.2.4 Peroxide Shunt.....	17
1.4.2.5 Electron Transfer Mechanism.....	18
1.4.2.5.1 Understanding Methods of Electron Transfer.....	18
1.4.3 <i>Structure of Cytochrome P-450cam</i> .....	21
1.4.3.1 Structural Studies of Electron Transfer.....	24
1.4.4 <i>The P-450cam Binding Site</i> .....	26
1.4.4.1 Thermodynamic Factors that Control Substrate Binding.....	27
1.4.4.2 Substrate Specificity of P-450cam.....	28
1.4.4.3 Mutagenesis Effects on P-450cam Substrate Binding.....	29
1.5 PROJECT AIMS.....	33
1.6 REFERENCES.....	35
<b>2. EXPERIMENTAL</b> .....	<b>45</b>
2.1 MATERIALS.....	45
2.1.1 <i>General Chemicals</i> .....	45
2.1.2 <i>Proteins</i> .....	45
2.1.3 <i>Protein Expression and Purification</i> .....	46
2.1.4 <i>Protein Analysis</i> .....	46
2.1.5 <i>Product Analysis</i> .....	46
2.1.6 <i>Crystallography</i> .....	47
2.1.7 <i>Modelling</i> .....	47

2.2 GENERAL EXPERIMENTAL CONDITIONS. ....	48
2.2.1 GC Method.....	48
2.3 GROWTH AND EXPRESSION OF CYTOCHROME P-450CAM. ....	49
2.3.1 General Growth Conditions. ....	49
2.3.2 Large Scale Growth, in a 50 L Fermenter. ....	49
2.4 PURIFICATION OF P-450CAM. ....	49
2.4.1 Preparation of the Cell-Free Extract.....	49
2.4.2 Anion Exchange Chromatography on a DEAE Sepharose™ Column. ....	50
2.4.3 Concentration and Desalting by Ultrafiltration.....	50
2.4.4 Anion Exchange Chromatography with SourceQ™ Media and a HPLC System. ....	51
2.4.5 Preparation of Protein Pure Enough for Crystallographic Studies. ....	51
2.4.6 Removal of Camphor and Glycerol by Gel Filtration. ....	52
2.5 GROWTH AND EXPRESSION OF PUTIDAREDOXIN. ....	53
2.5.1 General Growth Conditions. ....	53
2.5.2 Large Scale Growth, in 1 L Shake Flasks. ....	53
2.6 PURIFICATION OF PUTIDAREDOXIN. ....	54
2.6.1 Preparation of the Cell-free Extract.....	54
2.6.2 Anion Exchange Chromatography on a DEAE Sepharose Column. ....	54
2.6.3 Desalting by Dialysis. ....	54
2.6.4 Anion Exchange Chromatography with SourceQ Media and a Waters Protein HPLC System. ....	55
2.6.5 Final Purification Step of Pd.....	55
2.7 PROTEIN QUANTIFICATION.....	56
2.8 BINDING ASSAYS.....	57
2.8.1 Preparation of Protein Solution for Binding Experiments. ....	57
2.8.2 Assay to Determine Maximum Binding via Shift in the Soret Band. ....	57
2.8.3 Assay to Determine Binding Constants.....	57
2.8.4 Assay to Determine Binding at a Maximum Soret Shift for Oxygen Ligation. ....	58
2.9 SOLVENT STABILITY OF CYTOCHROME P-450CAM. ....	58
2.10 OPTIMIZATION OF THE ELECTRONIC TRANSFER AGENTS.....	58
2.10.1 Preparation of Enzymes and Chemicals. ....	58
2.10.2 Biotransformation Conditions. ....	59
2.11 BIOTRANSFORMATIONS WITH P-450CAM.....	59
2.12 CRYSTALLOGRAPHY OF THE CYTOCHROME P-450CAM Y96A C334A MUTANT.....	60
2.12.1 Preparation of Protein. ....	60
2.12.2 Crystallization Screens.....	60
2.12.3 Freezing Condition Screening.....	61
2.12.4 Final Conditions for Crystal Growth. ....	61
2.12.5 Final Conditions for Cryofreezing of Cytochrome P-450cam Crystals.....	61
2.12.6 Crystal Screening.....	62
2.12.7 Data Collection.....	62
2.12.8 Data Reduction and Refinement.....	62
2.12.8.1 Substrate Building.....	63
2.12.8.2 4-Hydroxydiphenylmethane Structure.....	63
2.12.8.3 General Technique.....	64
2.12.8.4 Camphor Structure.....	64
2.12.8.5 Phenylcyclohexane Structure.....	64
2.12.8.6 Diphenylmethane Structure.....	65
2.12.8.7 Bicyclohexyl Structure.....	65
2.12.8.8 N-Cbz-2,6-dimethylpiperidine.....	65
2.12.8.9 Structural Verification.....	66
2.12.8.10 Comparison of Structures.....	66
2.13 MOLECULAR MODELLING.....	67
2.13.1 Initial Modelling.....	67
2.13.2 Remodelling of Substrates.....	67

2.13.3 Use of DOCK.....	68
2.13.3.1 Ligand Preparation.....	68
2.13.3.2 Site Characterization.....	69
2.13.3.3 Generating Spheres.....	70
2.13.3.4 Creating Scoring Grids.....	70
2.13.3.5 Running DOCK.....	71
2.14 REFERENCES.....	72
<b>3. PROTEIN EXPRESSION AND PURIFICATION.....</b>	<b>74</b>
3.1 CYTOCHROME P-450CAM.....	74
3.1.1 Expression of Cytochrome P-450cam in <i>E. coli</i> .....	74
3.1.2 Purification of P-450cam.....	75
3.1.2.1 Protein Quantification and Purity Measurements.....	76
3.1.2.2 Extraction of Cytochrome P-450cam from <i>E. coli</i> .....	77
3.1.2.3 Purification by Anion Exchange Chromatography.....	78
3.1.2.4 Desalting and Concentration of DEAE Fractions.....	79
3.1.2.5 Anion Exchange Chromatography Using HPLC with SourceQ™ Media.....	79
3.1.2.6 Removal of Camphor and Glycerol by Gel Filtration.....	80
3.1.2.7 Production of Protein of Sufficient Quality for Crystallography.....	80
3.1.2.8 Summary.....	81
3.2 PUTIDAREDOXIN.....	82
3.2.1 Expression of Putidaredoxin in <i>E. coli</i> .....	82
3.2.2 Purification of Pd.....	83
3.2.2.1 Protein Quantification and Purity Measurements.....	83
3.2.2.2 Extraction of Pd from <i>E. coli</i> .....	84
3.2.2.3 Purification by Anion Exchange Chromatography.....	84
3.2.2.4 Desalting and Concentration of DEAE Fractions.....	85
3.2.2.5 Anion Exchange Chromatography Using HPLC with SourceQ™ Media.....	85
3.2.2.6 Final Purification Step of Pd.....	86
3.2.2.7 Summary.....	87
3.3 PUTIDAREDOXIN REDUCTASE.....	87
3.3.1 PdR Quantification:.....	87
3.4 SUMMARY.....	88
3.5 REFERENCES.....	89
<b>4. BIOCHEMICAL ANALYSIS OF CYTOCHROME P-450CAM.....</b>	<b>91</b>
4.1 THE SOLVENT STABILITY OF P-450CAM Y96A MUTANT.....	91
4.1.1 Introduction.....	91
4.1.2 Results of Solvent Stability Analysis.....	92
4.2 ALTERNATIVE ELECTRON TRANSFER AGENTS.....	93
4.2.1 Introduction.....	93
4.2.2 Results of Electron Transfer Optimization.....	94
4.2.2.1 Enzymatic Systems.....	94
4.2.2.2 PMS/ Enzymatic System.....	96
4.2.2.3 Use of Crude Enzyme Samples.....	97
4.2.3 Conclusions.....	97
4.3 BINDING ASSAYS AND SUBSTRATE TURNOVER.....	98
4.3.1 Introduction.....	98
4.3.1.1 Calculation of Binding Constants.....	99
4.3.1.2 Binding with Heme Ligation.....	99
4.3.1.3 Percentage High-spin.....	100
4.3.1.4 Substrate Turnover.....	100
4.3.2 Substrates Analysis.....	101

4.3.2.1 Camphor.....	101
4.3.2.1.1 Product Analysis of Camphor.....	102
4.3.2.1.2 Effect of the P-450cam Y96A Mutation on R-Camphor Hydroxylation.....	108
4.3.2.2 Hydroxylation of Camphor/ Adamantane Analogues.....	109
4.3.2.3 Diphenylmethane Analogues.....	114
4.3.2.4 Naphthalene Analogues.....	118
4.3.2.5 Chlorinated Aromatic Compounds.....	120
4.3.2.6 Aliphatic Substituted Aromatics.....	122
4.3.2.7 Aromatically Substituted Piperidines.....	124
4.3.2.8 Oxygen Heterocycles with Aromatic Sidechains.....	126
4.4 CONCLUSIONS.....	128
4.5 REFERENCES.....	130
<b>5. CRYSTALLIZATION OF THE P-450CAM Y96A MUTANT.....</b>	<b>132</b>
5.1 INTRODUCTION.....	132
5.2 PROTEIN EXPRESSION AND PURIFICATION.....	133
5.3 CRYSTALLIZATION SCREENS.....	133
5.3.1 <i>Using Wild-Type Conditions</i> .....	134
5.3.2 <i>Development of New Conditions</i> .....	135
5.3.2.1 pH Dependence.....	135
5.3.2.2 Substrate Co-crystallization Effects.....	135
5.3.2.3 PEG and Protein Concentration.....	136
5.3.2.4 Temperature.....	139
5.3.2.5 Substrate Concentration.....	139
5.4 FINAL CRYSTALLIZATION CONDITIONS.....	140
5.5 DATA COLLECTION.....	140
5.6 PHASE DETERMINATION.....	140
5.7 THE NEW UNIT CELL.....	141
5.7.1.1 Differences in the Asymmetric Units.....	142
5.8 CONCLUSIONS.....	142
5.9 REFERENCES.....	147
<b>6. CRYSTAL STRUCTURE OF THE CYTOCHROME P-450 Y96A-C334A MUTANT. ....</b>	<b>148</b>
6.1 INTRODUCTION.....	148
6.2 EFFECTS OF THE C334A MUTATION.....	148
6.3 EFFECTS ON THE BINDING SITE BY THE Y96A MUTATION.....	150
6.3.1 <i>Effects on the Y96A Binding Site Based on the Camphor Bound Structure</i> .....	151
6.3.1.1 Modelling the Iron-Sulfur Bond.....	154
6.3.2 <i>Effects on the Y96A Binding Site by Phenyl Substituted Substrates</i> .....	154
6.3.2.1 Differences in the 4-Hydroxydiphenylmethane Bound Structure Compared to Camphor Bound Structure.....	154
6.3.2.2 Differences in the Diphenylmethane, Phenylcyclohexane and <i>N</i> -cbz-2,6-dimethylpiperidine Bound Structures.....	155
6.3.3 <i>Effects on the Y96A Binding Site with a Non-Aromatic Substrate</i> .....	155
6.3.3.1 Differences in the Bicyclohexyl Bound Structure.....	155
6.4 SUBSTRATE ORIENTATIONS IN THE Y96A MUTANT BINDING SITE.....	156
6.4.1 <i>Orientation of Camphor in the Y96A Mutant</i> .....	156
6.4.1.1 Main Orientation of Camphor.....	156
6.4.1.2 Alternative Orientations of Camphor.....	156
6.4.1.3 Addition of a Water Molecule to Camphor Bound Binding Site.....	157
6.4.1.4 Comparison of Camphor in Wild-Type and Y96A Mutant P-450cam.....	158
6.4.2 <i>Orientation of 4-Hydroxydiphenylmethane in the Y96A Mutant</i> .....	160
6.4.3 <i>Orientation of Diphenylmethane in the Y96A Mutant</i> .....	166

6.4.3.1 Diphenylmethane Orientation.....	166
6.4.3.2 Binding Site Water in the Diphenylmethane Bound Structure.....	167
6.4.4 Orientation of Phenylcyclohexane in the Y96A Mutant.....	168
6.4.4.1 Orientation of Phenylcyclohexane.....	168
6.4.4.2 Phenylcyclohexane Active Site Bound Water.....	170
6.4.5 Orientation of Bicyclohexyl in the Y96A Mutant.....	171
6.4.6 Binding of N-cbz-2,6-dimethylpiperidine.....	173
6.5 COMPARISON OF THE BINDING ORIENTATIONS.....	176
6.5.1 Orientation of the 'Upper' Phenyl Rings.....	176
6.5.2 Comparison of the 'Lower' Rings of the 4-Hydroxydiphenylmethane, Diphenylmethane, Phenylcyclohexane and Bicyclohexyl Bound Orientations.....	177
6.5.2.1 4-Hydroxydiphenylmethane vs. Diphenylmethane.....	177
6.5.2.2 Phenylcyclohexane Compared to 4-Hydroxydiphenylmethane/ Diphenylmethane.....	178
6.5.2.3 Comparison of Bicyclohexyl to 4-Hydroxydiphenylmethane/ Diphenylmethane and Phenylcyclohexane Structures.....	179
6.6 CONCLUSIONS.....	180
6.7 REFERENCES.....	181
<b>7. FEATURES OF THE CYTOCHROME P-450 Y96A MUTANT.....</b>	<b>182</b>
7.1 IMPLICATIONS OF THE SUBSTRATE BOUND DATA.....	182
7.1.1 Enzyme Complimentarily Towards Camphor.....	182
7.1.2 Implications of the 4-Hydroxydiphenylmethane Structure for Cytochrome P-450 Mechanisms-Unusual 251/252 Configuration in the 4-Hydroxydiphenylmethane Structure.....	183
7.2 COMMON FEATURES OF THE Y96A MUTANT STRUCTURES.....	195
7.2.1 The Hydrophobic Binding Region (the 'Phenyl Pocket').....	195
7.2.2 First Potassium Binding Site.....	197
7.2.3 A Second Potassium Binding Site.....	200
7.2.4 Aqua Ligand.....	201
7.2.4.1 Camphor Bound Structural Water.....	203
7.2.4.2 Phenylcyclohexane Bound Structural Water.....	203
7.2.4.3 Water Occupancy of Other P-450cam Y96A Mutant Structures.....	204
7.2.5 Uncoupling.....	204
7.3 CONCLUSIONS.....	205
7.4 REFERENCES.....	206
<b>8. MOLECULAR MODELLING.....</b>	<b>209</b>
8.1 INTRODUCTION.....	209
8.2 CONFIRMATION OF THE MODEL BY MODELLING KNOWN SUBSTRATES.....	210
8.3 MODELLING OF NOVEL SUBSTRATES.....	211
8.3.1 Diclofenac.....	211
8.3.2 Studies of 2,3-Epoxy cyclohexyl-1-benzoate.....	213
8.4 MODELLING OF PREVIOUSLY STUDIED SUBSTRATES.....	215
8.5 REMODELLING THE SUBSTRATES IN THE Y96A CRYSTAL STRUCTURE.....	217
8.5.1 Camphor and Adamantane Analogues (2, 10, 23-28).....	217
8.5.2 Diphenylmethane Analogues (13, 18, 29, 30).....	218
8.5.3 Naphthalene Analogues (31-36).....	218
8.5.4 Chlorinated Aromatics (38-40).....	218
8.5.5 Other Aromatically Substituted Substrates (41-57).....	219
8.6 CONCLUSIONS FROM THE SUBSTRATE MODELLING STUDIES.....	219
8.7 SUBSTRATE ENTRY IN P-450CAM BINDING SITE.....	219
8.8 REFERENCES.....	226



<b>9. MODELLING WITH THE PACKAGE DOCK.....</b>	<b>229</b>
9.1 INTRODUCTION.....	229
9.2 PREVIOUS STUDIES WITH P-450CAM AND DOCK.....	232
9.3 DIFFERENCES IN OUR STUDIES WITH DOCK.....	232
9.3.1 <i>Best Binding Site Model</i> .....	234
9.3.2 <i>Selection of the Best Forcefield</i> .....	236
9.3.3 <i>Substrates Analysis</i> .....	237
9.3.4 <i>Understanding DOCK Output Files</i> .....	238
9.4 RESULTS OF DOCK SIMULATIONS.....	239
9.4.1 <i>Camphor and Adamantane Analogues</i> .....	239
9.4.1.1 Camphor.....	239
9.4.1.1.1 DOCK Orientations of Camphor.....	239
9.4.1.2 Camphor Analogues.....	240
9.4.1.2.1 DOCK Orientations of 24, 25 and 26.....	241
9.4.1.3 Adamantane and Adamantanols.....	242
9.4.1.3.1 DOCK Orientations of 10, 27 and 28.....	243
9.4.2 <i>General Orientations of Phenyl Substituted Substrates</i> .....	243
9.4.3 <i>Diphenylmethane Analogues</i> .....	244
9.4.3.1 DOCK Orientations of 13, 18, 29 and 30.....	245
9.4.4 <i>Naphthalene Analogues</i> .....	246
9.4.4.1 DOCK orientations of 31, 32, 33, 35 and 36.....	247
9.4.5 <i>Chlorinated Aromatics</i> .....	249
9.4.5.1 DOCK orientations of 38, 39, 37, 60 and 61.....	250
9.4.6 <i>Aliphatic Substituted Aromatics</i> .....	250
9.4.6.1 DOCK orientations of 41, 42 and 43.....	251
9.4.7 <i>Aromatically Substituted Nitrogen Heterocycles</i> .....	252
9.4.7.1 DOCK orientations of 44, 45, 46, 47, 48, 49, 50 and 51.....	254
9.4.8 <i>Oxygen-Heterocycles with Aromatic Sidechains</i> .....	257
9.4.8.1 DOCK orientation of 52, 53, 54, 55, 56 and 57.....	258
9.5 CONCLUSIONS.....	260
9.5.1 <i>Validation of the DOCK Output Data</i> .....	260
9.6 REFERENCES.....	264
<b>10. FUTURE WORK.....</b>	<b>266</b>
10.1 EXPANSION OF CURRENT STUDIES.....	266
10.1.1 <i>X-ray Crystallography</i> .....	266
10.1.2 <i>Molecular Modelling</i> .....	267
10.2 SCALE-UP OF CYTOCHROME P-450CAM BIOTRANSFORMATION.....	267
10.3 FURTHER MUTAGENESIS OF CYTOCHROME P-450CAM.....	268
10.4 NMR STUDIES.....	269
10.5 MOLECULAR DYNAMICS.....	270
10.6 REFERENCES.....	271
<b>APPENDIX 1: EXAMPLE DATASETS FOR THE CALCULATION OF THE MAXIMUM PERCENTAGE BINDING AND <math>K_D</math>.....</b>	<b>273</b>
<b>APPENDIX 2: THE DOCK INSTRUCTION FILE.....</b>	<b>274</b>
<b>APPENDIX 3: EXAMPLE OF DOCK OUTPUT FILE.....</b>	<b>276</b>

## Abbreviations.

AS	Ammonium sulfate
BM-3R	P-450BM-3 reductase
buffer AS	Ammonium sulfate buffer (1 M, pH 7.4) in buffer Pi
buffer Pi	Phosphate buffer (40mM, pH 7.4)
buffer Pi-cam	Phosphate buffer (40mM, pH 7.4) with 1 mM camphor
buffer T	Tris buffer (50 mM, pH 7.2)
DEAE	Diethylaminoethyl
DTT	Dithiothreitol
$\epsilon$	Extinction coefficient
FLD	Flavodoxin from <i>E.coli</i>
FLDR	Flavodoxin reductase from <i>E.coli</i>
GC	Gas chromatography
GC/MS	Gas chromatography with mass spectral analysis
H-bond	Hydrogen bond
HPLC	High pressure liquid chromatography
IPA	Isopropyl alcohol
IPTG	Isopropylthiogalactoside
JM109	<i>E.coli</i> strain JM109
kbp	Thousand base pairs
kDa	Thousand daltons
$\pi$ - $\pi$	Ligand $\pi$ orbital to $\pi$ orbital charge transfer
MWt	Molecular weight
MD	Molecular dynamics
mol2	SYBYL mol2 data format
NADH	Nicotinamide adenosine dinucleotide (reduced form)
NMR	Nuclear magnetic resonance spectroscopy

ABBREVIATIONS

OYE	Old yellow enzyme
P-450	Cytochrome P-450 monooxygenase
P-450BM-3	Cytochrome P450 fatty acid monooxygenase from <i>Bacillus megaterium</i>
P-450cam	Cytochrome P-450 camphor monooxygenase from <i>Pseudomonas putida</i>
P-450eryF	Cytochrome P-450 monooxygenase from <i>Saccharopolyspora erythraea</i>
P-450terp	Cytochrome P-450 terpene monooxygenase from <i>Pseudomonas spp.</i>
PAGE	Polyacrylamide gel electrophoresis
pAW7292	Plasmid with code pAW7292
Pd	Putidaredoxin
pdb	Protein databank
PdR	Putidaredoxin reductase
PEG	Polyethylene glycol
PMS	Phenazine methylsulfate
OD	Optical density
$R_{\text{factor}}$	$R_{\text{factor}} = \frac{\sum  F_{\text{obs}}  -  F_{\text{calc}} }{\sum  F_{\text{obs}} }$
$R_{\text{free}}$	$R_{\text{factor}}$ calculated for 5 % omitted reflections
rmsd	Root mean squared deviation
$R_{\text{sym}}$	$R_{\text{sym}} = \frac{\sum  I - \langle I \rangle }{\sum I}$
Sd	Spinach ferredoxin
SdR	Spinach ferredoxin reductase
SDS	Sodiumdodecylsulphate
TB	'Terrific broth' <sup>1</sup>
TBamp	TB with 1 mM ampicillin
TBcm	TB with 1 mM chloroamphenicol
UV/Vis	Ultraviolet and visible range

Y96A                      Cytochrome P-450cam Y96A mutant

Amino acid one letter codes used:

Alanine (A), Arginine (R), Asparagine (N), Aspartate (D), Cysteine (C), Glutamate (E), Glutamine (Q), Glycine (G), Histidine (H), Isoleucine (I), Leucine (L), Lysine (K), Methionine (M), Phenyl Alanine (F), Proline (P), Serine (S), Threonine (T), Tryptophan (W), Tyrosine (Y), Valine (V).

Amino acid three letter codes used:

Alanine (Ala), Arginine (Arg), Asparagine (Asn), Aspartate (Asp), Cysteine (Cys), Glutamate (Glu), Glutamine (Gln), Glycine (Gly), Histidine (His), Isoleucine (Iso), Leucine (Leu), Lysine (Lys), Methionine (Met), Phenyl Alanine (Phe), Proline (Pro), Serine (Ser), Threonine (Thr), Tryptophan (Trp), Tyrosine (Tyr), Valine (Val).

Mutant protein coding system used:

Y96A, F87A etc. : The first letter for the original amino acid and the last for the amino acid inserted, the number corresponds to the residue number.

1) Sambrook, J.; Fritsch, E.; Maniatis, T. *Molecular Cloning*; 2nd ed.; Cold Spring Harbour Laboratory Press, 1989.

# 1 Introduction.

---

## 1.1 Oxygenases.

The direct incorporation of molecular oxygen into a molecule is a type of oxidation known as oxygenation and it is catalyzed by enzymes called oxygenases.<sup>1</sup> There are three general mechanisms for oxygenation:

- 1) Monooxygenation is the incorporation of one oxygen atom from molecular oxygen into a substrate, the other oxygen being reduced at the expense of a proton donor (usually NADH or NADPH), to form water. Monooxygenases are then subdivided into two classes, internal and external,<sup>2</sup> depending on whether the source of hydrogen for the water molecule comes from an 'external' donor or the substrate itself (internal).
- 2) Di-oxygenation, the transfer of both molecular oxygen atoms to a substrate, (enzymes catalyzing these reactions are known as oxygen transferases as well as di-oxygenases).
- 3) Oxidation catalyzed by oxidases, which involves the transfer of two or four electrons utilizing either water or hydrogen peroxide as the acceptor.

These reactions proceed *via* general mechanisms, figure 1.1.

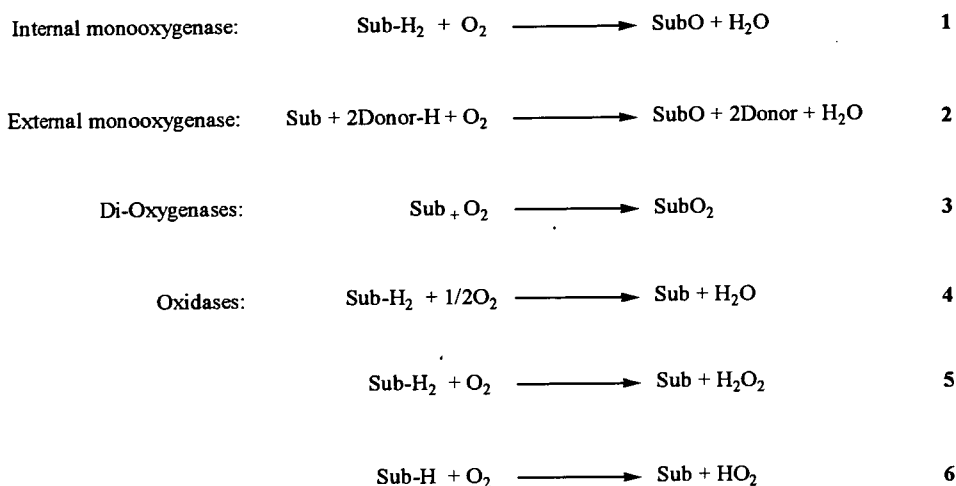


Figure 1.1: Types of oxygenation reaction (Sub = Substrate). Examples: 1- L-lysine monooxygenase; 2- cytochrome P-450; 3- tryptophan pyrrolase; 4- cytochrome oxidase; 5- D-amino acid oxidase; 6- xanthine oxidase.<sup>1,3</sup>

Of these types of oxygenases the group of interest in this study is that of the monooxygenases.

## 1.2 Monooxygenases.

Monooxygenases can catalyze a range of oxidations such as the epoxidation of alkenes,<sup>4</sup> sulfoxidation,<sup>5</sup> Baeyer-Villiger reactions,<sup>6</sup> dehalogenation<sup>7</sup> and C-C bond scission.<sup>8</sup> Monooxygenases can contain a heme cofactor as the catalytic centre (e.g. cytochrome P-450<sup>9</sup> and catalase<sup>10</sup>), although they can utilize an iron centre without heme. Of all these enzyme systems, cytochrome P-450 is the one that has been most extensively studied and is of interest here.

### 1.3 Cytochrome P-450.

Cytochrome P-450 (or P-450 monooxygenase) is a class of monooxygenase first isolated from mammalian cells in the adrenal cortex and in liver microsomes.<sup>11,12</sup> Their name arises from the b-type heme group, which has an absorbance (or pigment, P) at 450 nm when bound to carbon monoxide.<sup>13</sup> Their natural substrate range varies in size from ethene (MWt of 28) to cyclosporin (MWt of 1201). They react with many diverse substrates, including: fatty acids; steroids; prostaglandins; drugs; anesthetics; organic solvents; alcohols; pesticides and carcinogens. They occur in nearly all types of living organism, from mammals<sup>14-16</sup> and yeast<sup>17</sup> to plants and bacteria.<sup>18,19</sup> P-450 monooxygenases were first correctly identified, in mammals, in the 1960's<sup>20</sup> and they have been shown to be responsible for a wide range of biological functions from metabolism of xenobiotics (especially drugs, steroids and lipids) to detoxification,<sup>21</sup> which in some cases leads to the unwanted production of carcinogens. Examples of the wide range of reactions catalysed by P-450 monooxygenases are shown in figure 1.2.

The nomenclature for cytochrome P-450 was standardized in 1991<sup>22</sup> and by 1995 the cytochrome P-450 gene super-family has been classified into 36 subgroups and many subfamilies. The P-450 monooxygenase family is now classified by the letters CYP followed by the number of the gene family, the letter of the subfamily and lastly by a number for the individual enzyme (e.g. CYP11A1- family 11, subfamily A, enzyme number 1; formally called P-450<sub>sc</sub>). The simpler eukaryotes have numbers starting at 50, e.g. yeast. However for bacterial P-450 monooxygenases, all numbered with gene family numbers over 100, the traditional names are still commonly in use (e.g. P-450<sub>cam</sub>, P-450<sub>terp</sub> etc.).

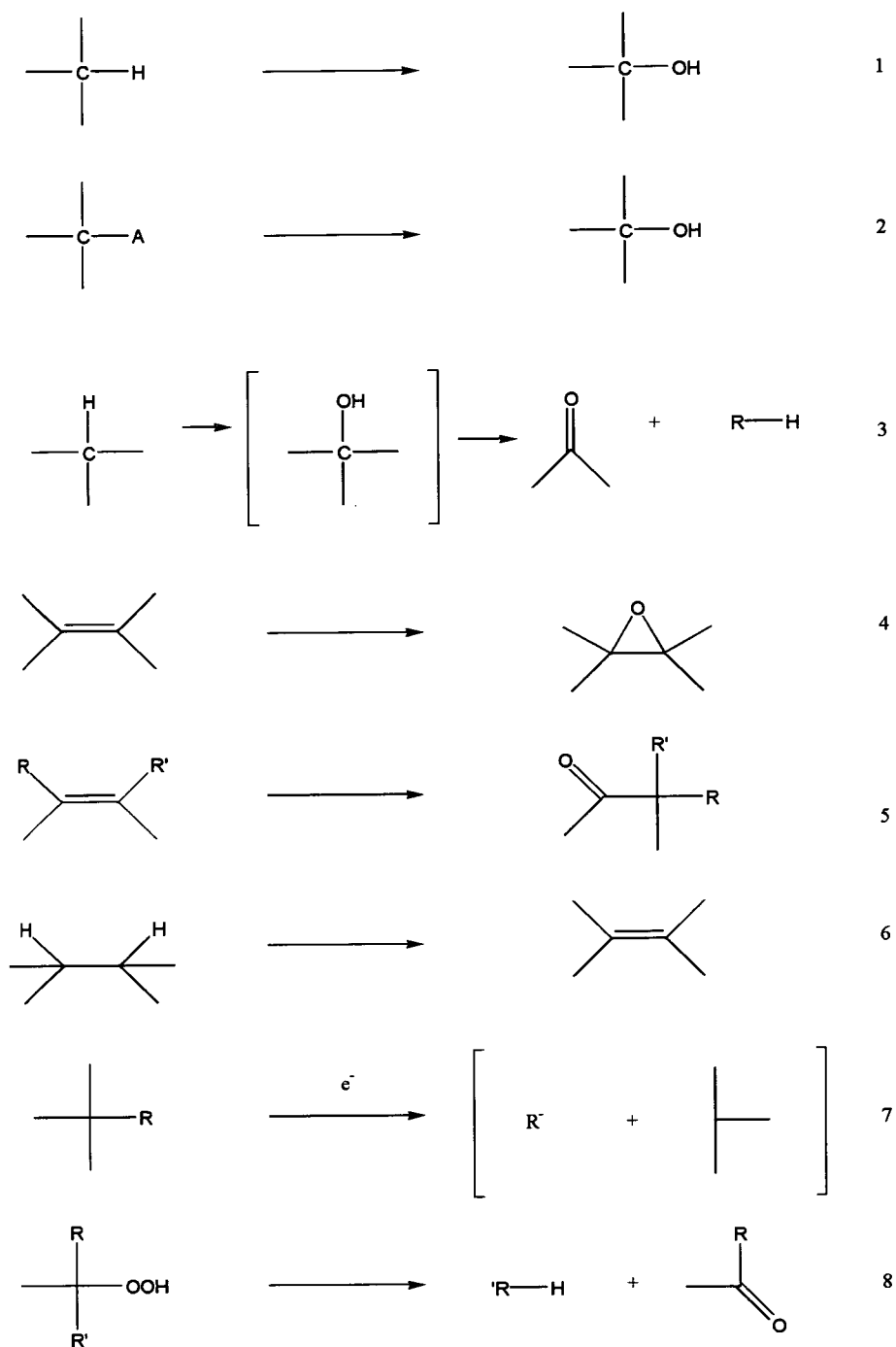


Figure 1.2: Examples of reactions catalyzed by cytochrome P-450, where A is a hetero atom and R is a hydrocarbon.<sup>23</sup> 1-Carbon hydroxylation,<sup>24</sup> 2-Heteroatom hydroxylation; 3-Dealkylation; 4-Exoxidation;<sup>25</sup> 5-NIH shift,<sup>26</sup> 6-Dehydrogenation,<sup>27</sup> 7-Reduction<sup>28</sup> and 8-Hydroperoxide cleavage.<sup>29</sup>



Due to this potential confusion the modern CYP numbers and traditional cytochrome P-450 names of the most common bacterial P-450 monooxygenases are tabulated below.

Table 1.1: Common bacterial P-450 monooxygenases with their traditional and modern names and with the host organism:<sup>9</sup>

CYP number	P-450x, where x=	original bacterial source
101	cam	<i>Pseudomonas putida</i>
102	BM-3	<i>Bacillus megaterium</i>
105A1/B1	SU1/2	<i>Streptomyces griseolus</i>
105C1	choP	<i>Streptomyces sp.</i>
105D1	soy	<i>Streptomyces griseus</i>
106A1	BM-1	<i>Bacillus megaterium</i>
106A2	meg	<i>Bacillus megaterium</i>
107A1	eryF	<i>Saccharopolyspora erythraea</i>
108	terp	<i>Pseudomonas sp.</i>

The gene identity between all P-450 monooxygenases is less than 20 % however between families it can rise to 40 % and between subfamilies greater than 55 % identity is observed. These percentages are biased by two areas of strong sequence identity, one for the proposed oxygen binding pocket the other for the heme binding site, figure 1.3.

341	ILS	Q	I	H	F	M	R	P	L	A	K	L	P450eryF				
367	ILG	Q	W	A	H	M	L	Q	H	L	A	K	L	P450terp			
347	H	T	T	C	H	S	H	L	L	G	Q	H	L	A	R	R	P450cam
390	E	K	P	E	N	Q	R	A	I	G	Q	Q	F	A	L	H	P450BM-3

Figure 1.3: Section of strong sequence identity of the heme binding region in a number of P-450 monooxygenases.<sup>30</sup>

In humans the function of cytochrome P-450 is of interest because of their role in drug metabolism and activation of procarcinogens.<sup>31</sup> Human and eukaryotic P-450 monooxygenases generally are present in very small quantities, are hard to purify and are often membrane bound; therefore the focus of cytochrome P-450 research has been on the soluble, bacterial P-450 monooxygenases and the more readily available rat liver cytochrome P-450, P-450<sub>LM2</sub>. Recently however eukaryotic P-450 monooxygenases have been expressed in mammalian<sup>32</sup> and insect<sup>33,34</sup> cell lines, yeast<sup>35</sup> and *E. coli*;<sup>36</sup> the reconstitution of these membrane bound species into active forms is rapidly advancing.<sup>37</sup>

Cytochrome P-450 has the potential to be a useful catalysts in synthetic chemistry because they can oxidize unactivated carbon-hydrogen bonds. Eukaryotic P-450 monooxygenases involved in detoxification can have a wide substrate specificity and low stereo selectivity<sup>38</sup> however prokaryotic P-450 monooxygenases tend to be stereoselective as well as having greater availability of genes, making them of potential interest as catalysts.

The bacterial P-450 monooxygenases studied to date are able to hydroxylate a wide variety of substrates: *Pseudomonas putida* (ATCC 17453) produces a cytochrome P-450 that can utilize camphor as its sole carbon source, P-450cam (see later). *Pseudomonas putida* (PpG777) has a cytochrome P-450 which hydroxylates linalool [**1**] (P-450lin), figure 1.4.

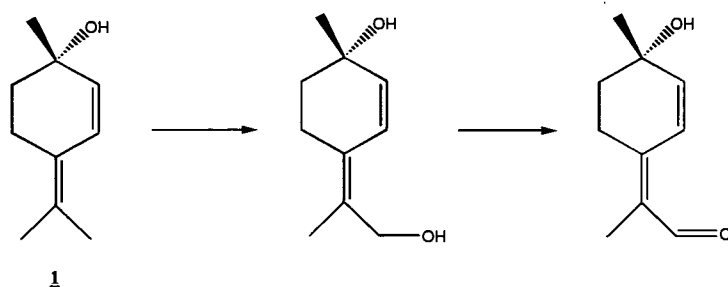


Figure 1.4: Schematic of Linalool [**1**] hydroxylation by P-450lin.<sup>39</sup>

Another *Pseudomonas* sp. hydroxylates  $\alpha$ -terpine (P-450terp). *Bacillus megaterium* (ATCC 14581) has two fatty acid hydroxylases: P-450BM-1<sup>40</sup> which has 27 % sequence homology to P-450cam from *P.putida* and, the better characterized P-450BM-3, which has greater sequence homology to mammalian P-450 monooxygenases than even to P-450BM-1. In addition to these fatty acid hydroxylases another *Bacillus megaterium* (ATCC 13368) has a steroid hydroxylase P-450meg.<sup>41,42</sup> Other well characterized P-450 monooxygenases can hydroxylate substrates from cyclohexane (*Xanthobacter* sp.<sup>40</sup>), hexadecane (*Acinetobacter calcoaceticus*<sup>43</sup>) and 2-ethoxyphenol (*Moraxella* sp.). A large number of the P-450 monooxygenases characterized to date belong to the actinomycetes of which P-450eryF (*Sacchropolyspora erythraea*); P-450sca-1, P-450sca-2 (*Streptomyces carbophilus*); P-450choP (*Streptomyces* sp.); P-450su-1 P-450su-2 (*Streptomyces griseolus*) and P-450soy (*Streptomyces griseus*) are the most well known.

The importance of research into cytochrome P-450 can be demonstrated when recombinant DNA technology became available; a bacterial P-450 monooxygenase, cytochrome P-450cam, was one of the first proteins whose DNA was cloned and sequenced.<sup>44</sup> Cytochrome P-450cam consequently became the most extensively studied<sup>45</sup> and was the first cytochrome P-450 to have its crystal structure solved.<sup>46</sup>

#### 1.4 Cytochrome P-450cam Monooxygenase.

Cytochrome P-450cam,<sup>9,47</sup> is a class II monooxygenase and was isolated from *Pseudomonas putida* (*P.putida*- ATCC 17453)<sup>48</sup> which metabolizes the monoterpene camphor to ultimately produce acetate and isobutyrate, thus allowing *P.putida* to utilize camphor as the sole carbon source,<sup>47,48</sup> figure 1.5.

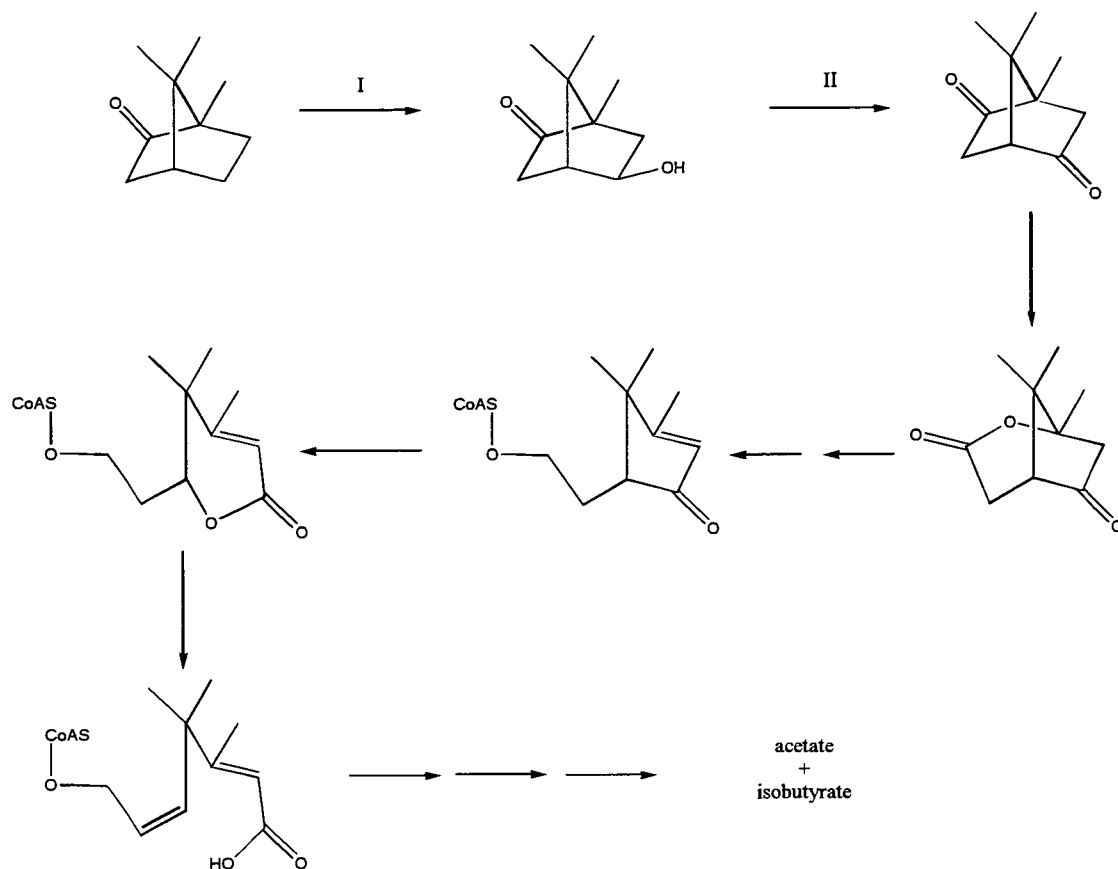


Figure 1.5: Metabolism of Camphor by *Pseudomonas Putida*.

I: Cytochrome P-450cam

II: 5-Hydroxycamphor dehydrogenase

Cytochrome P-450cam catalyses the first step in the catalytic cycle i.e. conversion of camphor [2] to produce 5-*exo*-hydroxycamphor [3],<sup>45</sup> figure 1.6.

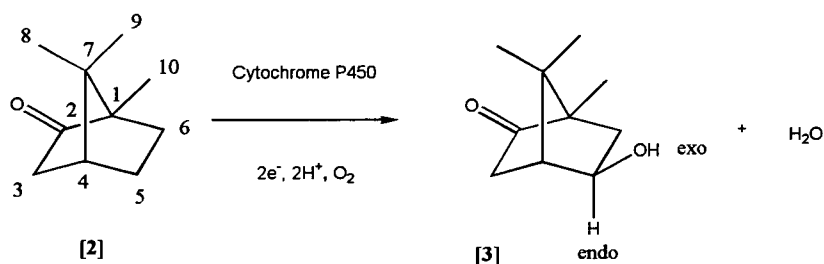


Figure 1.6: The hydroxylation of camphor by cytochrome P-450cam with the standard numbering of camphor is shown for clarity.

### 1.4.1 Why Do We Study Cytochrome P-450cam?

The oxidation of unactivated carbon centres is difficult synthetically, requiring harsh conditions such as high temperatures or extreme pH. An additional problem that is the stereoselectivity is hard to control. There are a number of reasons for using cytochrome P-450cam to catalyze a monooxidation; the oxidation is highly regio- and stereo-selective and can be performed under mild conditions. The enzyme has been extensively studied because its gene has been cloned and over-expressed in the heterologous host *E. coli*<sup>45</sup> allowing substantial amounts of soluble protein to be purified to homogeneity.<sup>49</sup> However these conditions can be equally applied to some of the other prokaryotic P-450 monooxygenases discussed previously. There is strong homology between both the cytochrome P-450 mechanism and structures<sup>30,50</sup> and since we are studying the complex interactions between substrates and enzyme the availability of structural information is also essential and, to date, only six different P-450 monooxygenases have solved crystal structures. Of these P-450 monooxygenases the enzymes best studied are P-450cam and P-450BM-3. The choice of cytochrome P-450 therefore comes down to substrate range which is either a fatty acid, in the case of P-450BM-3 or a small hydrophobic molecule for P-450cam. It was determined that the P-450cam could yield the type of hydroxylated substrates we desired.

1.4.2 P-450cam Catalytic Cycle.

1.4.2.1 Mechanism of Oxidation.

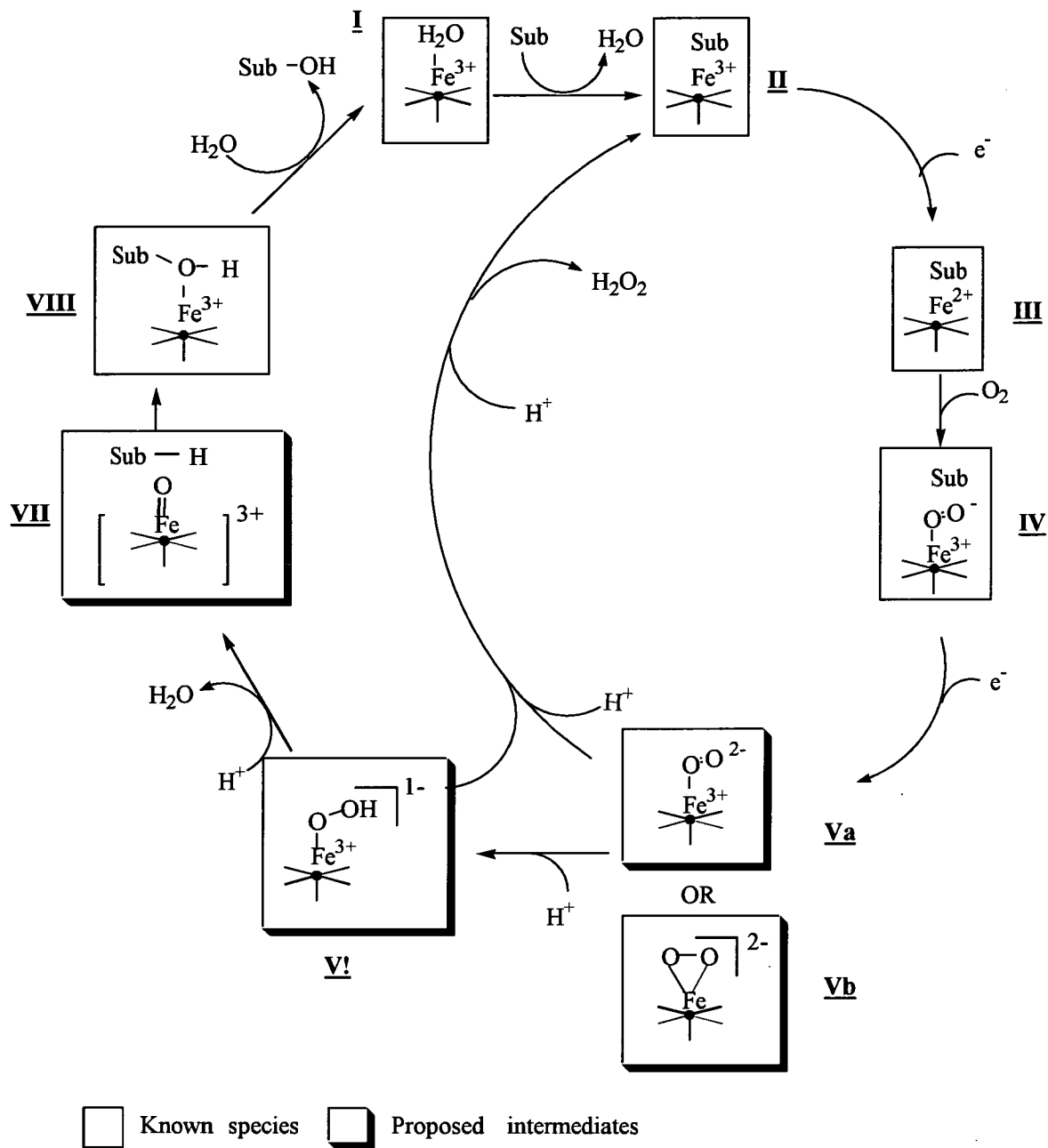


Figure 1.7: The P-450cam oxidation cycle. 45,51-53

This water molecule and its subsequent displacement holds the key to the oxidation process.<sup>56</sup> As a substrate enters the binding site the water ligand is displaced converting the hexa-coordinate low-spin iron ( $S=1/2$ ) into a penta-coordinate high-spin ( $S=5/2$ ) state,<sup>57</sup> ii (although evidence suggests the presence of substrate near the heme also plays a role in spin state change).<sup>58</sup> This change in the electronic configuration of the iron has an effect of increasing the reduction potential of the heme from -300 mV to -170 mV.<sup>59</sup>

Putidaredoxin (Pd) is an electron transfer protein required in the P-450cam pathway. Putidaredoxin has a reduction potential of -240 mV which changes to -196 mV upon binding to P-450cam. Thus when putidaredoxin interacts with the cytochrome P-450, containing heme in a high-spin state, it can transfer its electron to P-450cam and consequently to the heme, thus changing the penta-coordinate iron from Fe(III) to Fe(II), iii.

This change in oxidation state of the iron allows binding of molecular oxygen, which is known to bind to iron(II).<sup>60</sup> Oxygen is believed to be present in the binding site and possibly binds to a cleft around threonine 252.<sup>60</sup> Upon binding, the iron transfers the electron to the molecular oxygen thus reducing the iron back to give a hexa-coordinate Fe(III) species, iv.<sup>53</sup> This is known because the bound O-O<sup>-</sup> stretching frequency is detectable by Raman spectroscopy.<sup>61</sup> This species is the last fully detectable intermediate in the catalytic cycle and is unstable with a half life of ~90 ns.

The electrode potential of this oxygen bound iron is unknown however it is sufficiently low to allow the transfer of a second electron from Pd. The electron transfer of the second electron step is faster than the first thus in high concentrations of Pd the first electron transfer is rate limiting. However at physiological concentrations the second step becomes limiting. This additional electron converts the hexa-coordinate iron (III) to an iron-oxy species, v, only recently detected utilizing Raman spectroscopy by Sligar and coworkers.<sup>51</sup> Although it has not been confirmed, current experiments have indicated that species va is more likely to be the intermediate species.<sup>62</sup>

## 1.4.2.2 Radical Rebound Mechanism.

The next few stages in the catalytic cycle are still unknown due to the relatively short lifetime of the intermediate species; the second electron transfer to the heme occurs at the same rate as the substrate turnover (5 ms) therefore none of these intermediates build up.<sup>63</sup> Most of the pathway has been deduced from experimental results, energetic calculations and analogies to peroxidases which have similar excited heme bound oxygen molecules. Kinetic isotope studies have shown that two protons are involved in the next stages of the catalytic cycle and it has been deduced that their role is to remove one of the bound oxygen atoms as water. This theory of double protonation *via* a hypothesized peroxide unit, vi and the subsequent loss of water, should give a highly reactive iron-oxy species, vii<sup>60</sup> indeed this oxygen species is known to favour proton binding.<sup>64</sup>

It is this highly reactive iron-oxy species, vii, analogous to the peroxidase reactive species, that carries out the hydroxylation. It has been proposed that this oxygen is a charged radical which inserts into a carbon hydrogen bond of the substrate. This is known as the radical (or oxygen) rebound mechanism, figure 1.8.

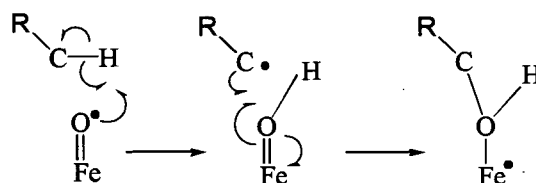


Figure 1.8: The radical rebound mechanism for insertion of oxygen into a carbon hydrogen bond.

The reactive oxygen radical extracts the nearest hydrogen, creating an iron with a bound hydroxyl and a carbon radical. This carbon radical then abstracts the hydroxyl group from the heme returning the iron to a penta-coordinate Fe(III) state, viii. The hydroxylated substrate is only weakly ligated to the heme and can easily be displaced. The substrate then leaves and is replaced by water, returning the iron to a hexa-coordinate iron (III) state, i.



The strongest evidence for this reaction of the carbon-hydrogen bond is found in experiments conducted by Sligar on deuteriated substrate,<sup>65</sup> e.g. hydroxylation of deuteriated norcamphor [4], figure 1.9.

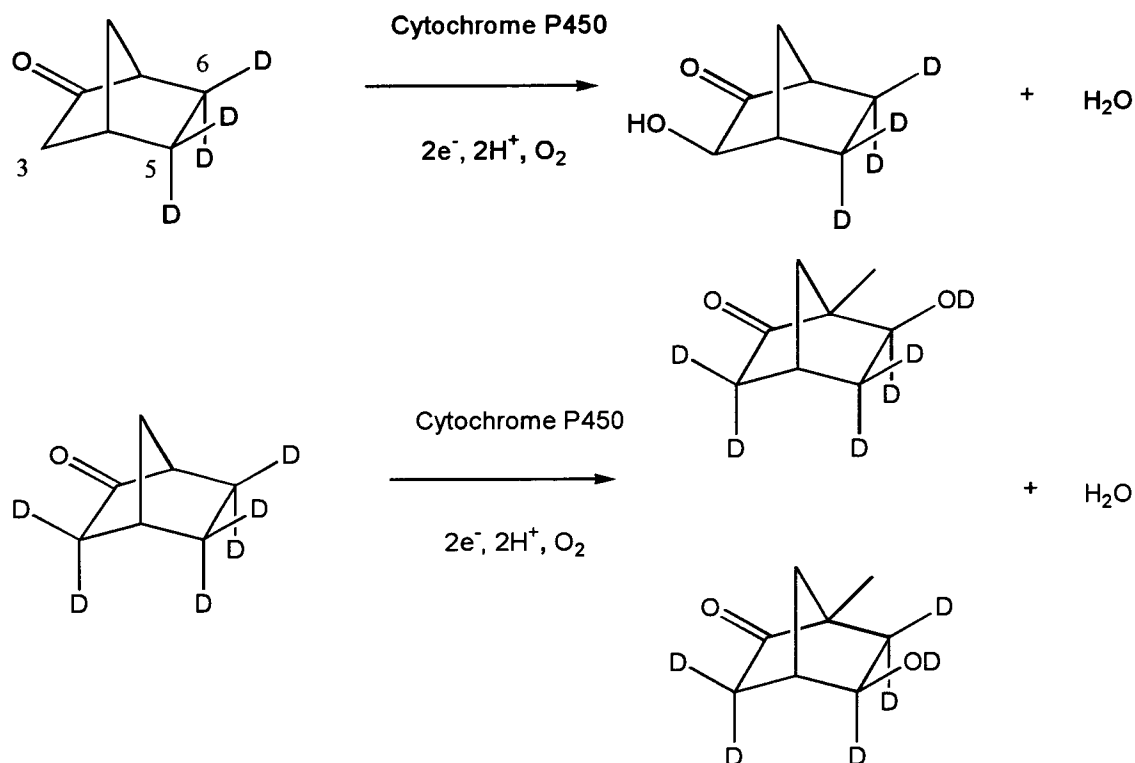


Figure 1.9: Hydroxylation of deuteriated norcamphor by cytochrome P-450cam.

The product profile is dependent on the deuteriation site. 3-*exo* hydroxylation (the thermodynamically favoured product) occurs when the 6- and 5- positions are deuteriated, however, when all three positions are deuteriated normal product profile is observed, albeit at a lower rate. This would indicate that the reaction proceeds *via* the sterically unfavourable 3- position rather than react with a deuteriated carbon, clearly indicating an isotope effect in the reaction cycle.

The evidence for a substrate radical intermediate is extensive,<sup>66</sup> again confirmed by experimentation on different deuteriated norcamphor,<sup>67</sup> figure 1.10.

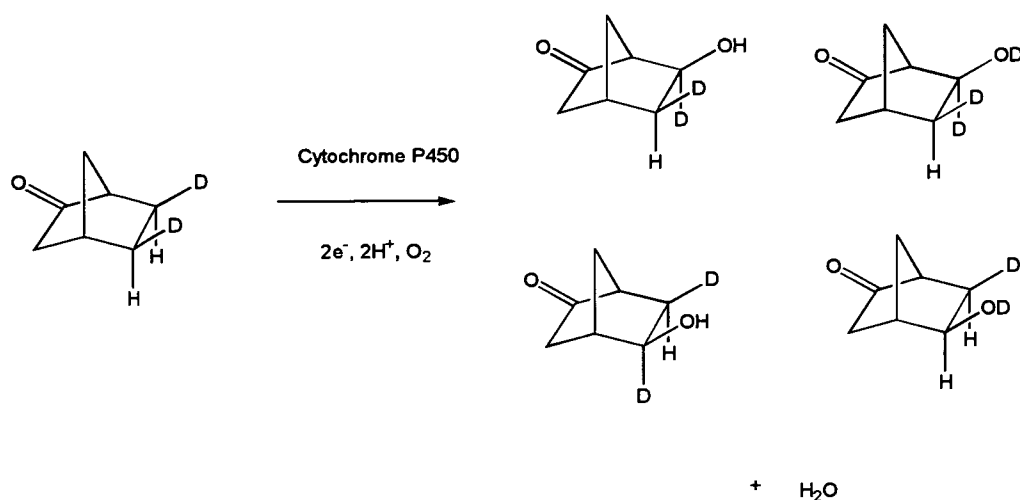


Figure 1.10: Hydroxylation of 5,6-*exo*-deutero-norcamphor by cytochrome P-450cam.

This data showed that, although *exo* abstraction is preferred *endo* abstraction is additionally possible. However, when analyzing product, entirely *exo* recombination is observed thus indicating a non-concerted mechanism. Chemical mimics of cytochrome P-450 systems also indicate radical intermediates.<sup>4</sup>

It has been proposed that a tetrahedral carbanion intermediate is present during the reaction, figure 1.11.

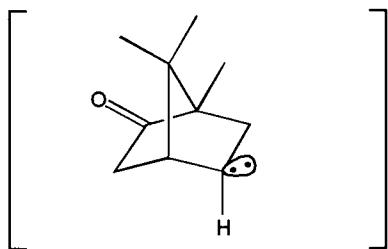


Figure 1.11 The proposed tetrahedral carbanion intermediate of camphor.

However, this can be discounted by experiments on pericyclocamphanone, [5], figure 1.12.<sup>68</sup>

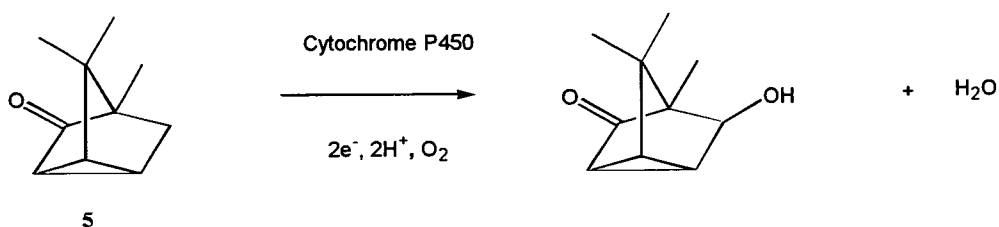


Figure 1.12: Hydroxylation of pericyclocamphanone by cytochrome P-450cam.

The 5-hydroxy product should still be possible *via* the proposed tetrahedral intermediate. However when hydroxylating this substrate the 6-hydroxy product is formed due to the impossible ring strain forced on a planar 5-hydroxy intermediate. Even if the 6- position is blocked no 5-hydroxylation is observed. This would indicate a planar rather than a tetrahedral intermediate, thus ruling out a carbanion.

#### 1.4.2.3 Alternative Theories of the Catalytic Cycle (Recent Advances).

##### 1.4.2.3.1 *Role of Sixth Water Ligand.*

Recent quantum mechanical calculations by Loew and coworkers on the electronic mechanism between the transition between low and high-spin have revealed that the removal of the aqua ligand may not be sufficient to induce the change from low to high-spin.<sup>54,55</sup> These energy calculations indicate that the heme must undergo a small change in the electronic distribution of its electrons in addition to the loss of the aqua ligand. This would suggest that the substrate's influence on the electrostatic field of the protein has a role to play in changing the electronic environment of the heme to give a high-spin species.

##### 1.4.2.3.2 *Oxidation State of the Iron and the Active Species.*

Evidence has been tending towards a ferryl iron (IV) species by analogy to horseradish peroxidase 'compound I'.<sup>69</sup> However some evidence would indicate that the iron is in a perferryl, iron (V) state;<sup>70</sup> either a iron (IV) or iron (V) species would have a mechanism that utilises the stepwise mechanism suggested by Sligar.<sup>71</sup> Recent

evidence from model complexes of the cytochrome P-450 system has indicated that the hydroxylation could proceed *via* a combination of a hydroxy- bound Fe (IV) species and a hydroxyl radical. In addition to this, peroxidase reactions have suggested that bound peroxide is the catalytic intermediate.<sup>72,73</sup>

#### 1.4.2.3.3 *The Alternative Catalytic Pathway: the Concerted Carbocation Mechanism.*

Although the radical rebound mechanism has been the more favoured over the years, recent advances have cast doubt on this mechanism. Newcomb and co-workers designed experiments to probe the mechanism of the cytochrome P-450 reaction through use of substrates (known as radical clocks), that could rearrange, as a radical, to give products indicative of the intermediate radicals.<sup>74-76</sup> These probes are usually substituted cyclopropane rings which undergo radical rearrangement with known half-lives. If the cytochrome P-450 system creates a radical and the substrate rearranges, then the lifetime of the unstable cytochrome P-450 intermediate must be greater than the radical rearrangement time scale. Thus, this way the intermediate lifetime can be estimated.

A variety of substrates seemed to confirm the radical pathway<sup>74,75,77-79</sup> until inconsistencies appeared,<sup>76</sup> suggesting a cationic species might be involved in the reaction pathway, figure 1.13.

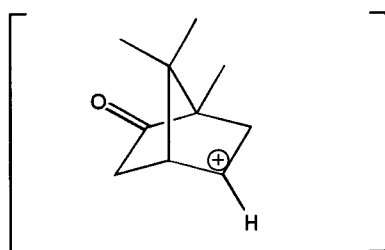


Figure 1.13: Proposed cationic intermediate of camphor.

Reappraisal of the previous experiments confirmed a cationic mechanism could explain many, but not all, of the observations. Recently, substrates whose

rearrangements occurred at rates faster than that possible for a radical rebound mechanism were found <sup>75</sup> suggesting the theory of a cationic species is more likely.

Further studies indicated that the life time of the 'radical' species is shorter than 70 fs. In this short life time it is impossible for two distinct species to exist, thus the mechanism cannot be regarded as non-concerted. It would also appear that the intermediate reaction contains both radical and cationic character. More recent studies have confirmed both the cationic <sup>80</sup> and a non-cationic nature of the species.<sup>81</sup> This apparent paradigm might be explained by a theory by Shaik <sup>82-84</sup> and confirmed again by Newcomb,<sup>85</sup> which indicates two species due to a competition between a one and two state reactivity. It therefore must be concluded that, although the radical rebound mechanism does not fully describe the catalytic process, a better alternative theory is not currently available.

#### 1.4.2.4 Peroxide Shunt.

A side reaction known as the 'peroxide shunt' also occurs within the catalytic cycle, figure 1.7, in which a breakdown of the cycle removes peroxide from intermediates v and vi, converting them back to species ii <sup>45,86</sup> The 'peroxide shunt' is believed to be caused by unfavorable interactions between the iron bound oxygen and protic solvents, due to poor binding of substrate, figure 1.14. <sup>86,87</sup>

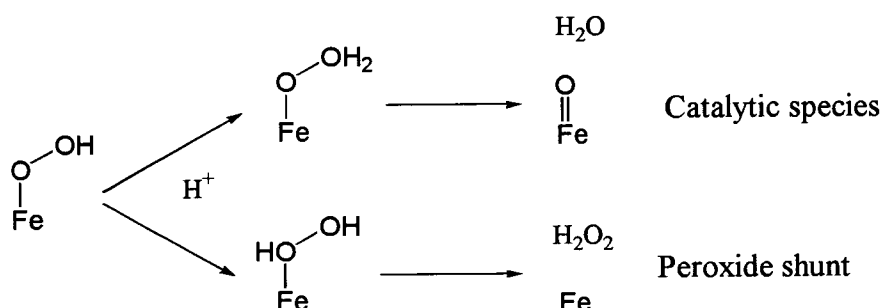


Figure 1.14: Alternative protonation mechanisms of the peroxide species.

This suggests that protonated water accesses the unprotonated oxygen of the reactive oxygen species, thus creating peroxide rather than a double protonated oxygen, which can be lost as water. The peroxide shunt is of particular interest when discussing unnatural substrates or protein mutants, as they have a higher propensity towards generation of peroxide. This can be seen when comparing camphor hydroxylation, which produces no peroxide, compared to norcamphor which can have as much as 60 % peroxide production,<sup>67</sup> or ethyl benzene which has 95 %.<sup>88,89</sup>

#### 1.4.2.5 Electron Transfer Mechanism.

The two electrons required for this catalytic cycle are provided by a cofactor NADH which is oxidized to NAD<sup>+</sup>.<sup>53,90</sup> Two co-enzymes, putidaredoxin (molecular weight of 11.5kDa)<sup>91</sup> and putidaredoxin reductase (PdR, molecular weight of 43.5kDa)<sup>92</sup> transfer the two electrons to the P-450cam,<sup>93,94</sup> figure 1.14. The rate determining step in the catalytic cycle is known to be after the formation of species v<sup>51</sup> but before formation of species ii (see figure 1.7). However the rate determining step in the whole P-450cam turnover cycle is the electron transfer between the Pd and P-450cam.<sup>51</sup>

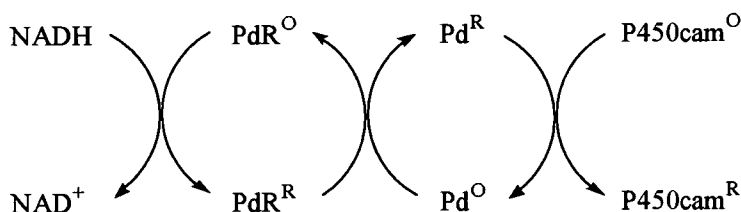


Figure 1.14: Electron transfer process from NADH to P-450cam (O and R indicate the oxidized and reduced species respectively).<sup>90</sup>

##### 1.4.2.5.1 Understanding Methods of Electron Transfer.

Microsomal P-450 monooxygenases (type I) are usually membrane bound and accept electrons from a reductase domain containing FAD and FMN<sup>3</sup> and they are reduced by both NADH and NADPH. Cytochrome P-450BM-3 is a unique example

of a bacterial cytochrome P-450 of this type containing both the heme and reductase domains in one protein.<sup>95</sup> The crystal structure for the heme domain of P-450BM-3, has been solved<sup>96</sup> and extensive study of the flavo domain has been performed,<sup>97-98</sup> making this the standard model for all microsomal P-450 monooxygenases, figure 1.15.

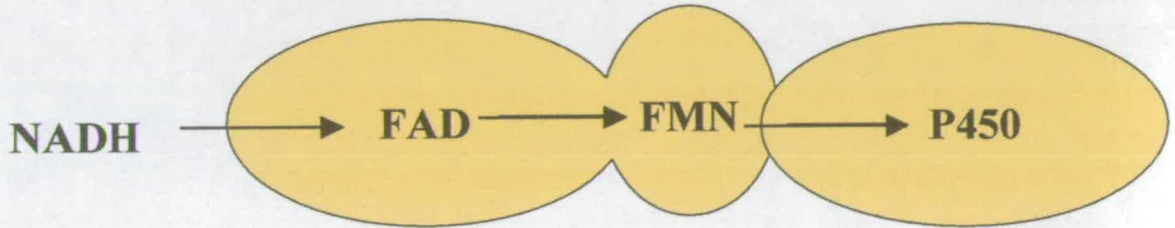


Figure 1.15: Microsomal cytochrome P-450 structure, arrows indicate direction of electrons.

Mitochondrial<sup>†</sup> type P-450 monooxygenases (type II), e.g. P-450cam, are distinct from microsomal P-450 monooxygenases in that they use a 2Fe2S centre (not the FMN unit) as an electron transfer medium to the cytochrome P-450, figure 1.16. These electron transfer units are generally expressed as separate proteins, although they may be bound together in membranes. Their reductases are specific to NADH, although some lower activity towards NADPH can be observed. There is a strong recognition factor between the 2Fe2S protein and the cytochrome P-450<sup>3</sup> which has only recently been characterized.<sup>94,99-106</sup>

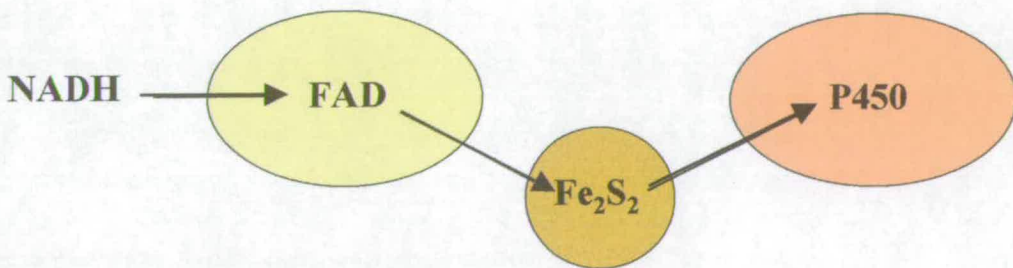


Figure 1.16: Mitochondrial cytochrome P-450 structure with the associated coproteins, arrows indicate direction of electrons.

<sup>†</sup> A mitochondrial type cytochrome P-450 is so named due to its similarity to mitochondrial cytochrome P-450 but is not necessarily from a mitochondrial source.

P-450cam has PdR as its 'Flavoprotein', i.e. containing the Flavin group, figure 1.17, of a sub class known as FAD.

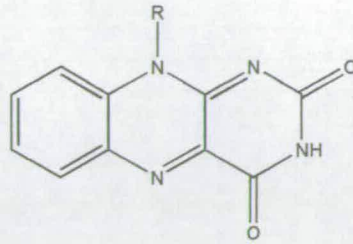


Figure 1.17: Structure of oxidized flavin unit, R is different for FMN or FAD proteins.

The FAD ring is incorporated into the protein substructure and is connected to the surface through a series of charged residues, figure 1.18.

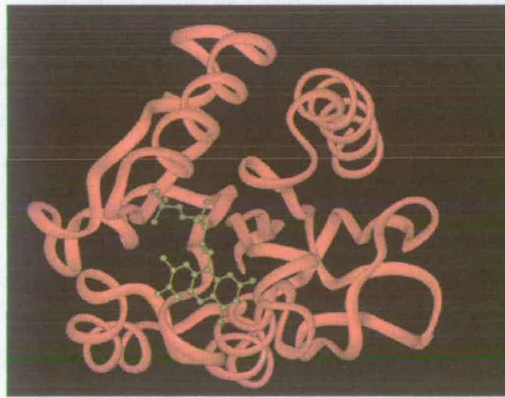


Figure 1.18: Model structure of PdR based on homologous structures, the FAD group is highlighted in yellow.<sup>107</sup>

Putidaredoxin is the second transfer protein in the cytochrome P-450 system, containing a bridged 2Fe2S as the key transfer agent; these proteins are known collectively as ferredoxins, figure 1.19.

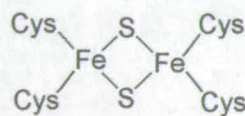


Figure 1.19: The structure of the 2Fe2S sub-unit of Putidaredoxin.



The 2Fe2S sub-unit is sufficiently near the surface of Pd to easily transfer electrons *via* one residue to P-450cam, although the exact electron transfer pathway is not known, figure 1.20.

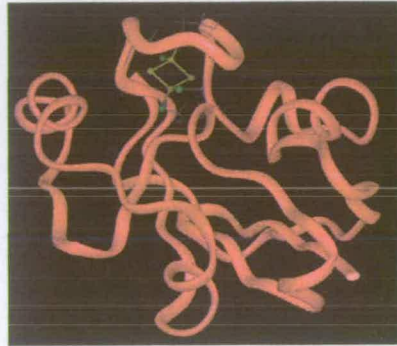
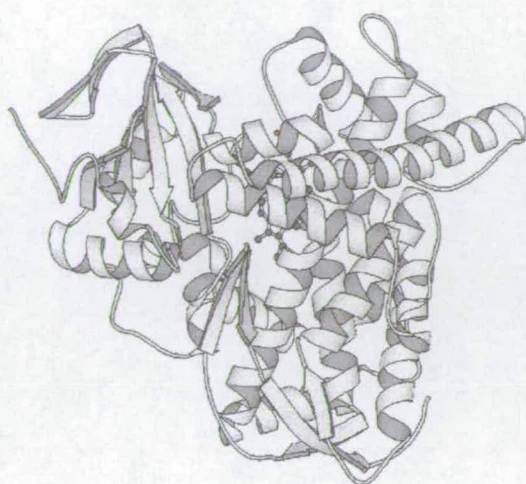


Figure 1.20: NMR structure of Pd with the 2Fe2S highlighted in yellow, the bound cysteines in green.<sup>108</sup>

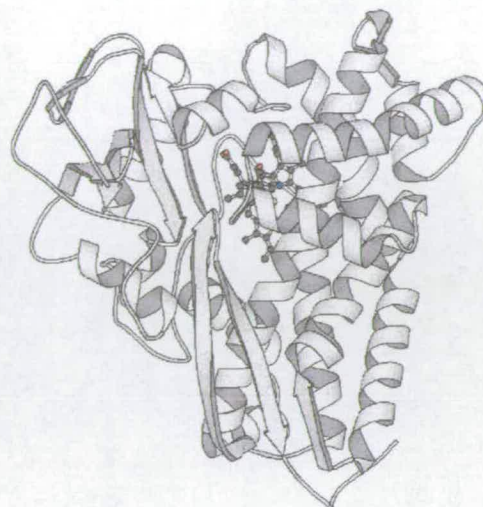
These 2Fe2S species are present in the transfer proteins of bacterial systems such as P-450cam and P-450terp (e.g. putidaredoxin, terpredoxin), in mammalian systems (e.g. adrenodoxin) and in plants (spinach ferredoxin).

### 1.4.3 Structure of Cytochrome P-450cam.

The crystal structure of P-450cam has been solved with bound R-camphor<sup>46,109</sup> and substrate-free<sup>110</sup> and additionally with 5-*exo*-hydroxycamphor,<sup>111</sup> S-camphor<sup>112</sup> and with many unnatural substrates.<sup>57,113,114</sup> P-450cam is one of only six cytochrome P-450 crystal structures to be solved to date, others being: P-450terp;<sup>115,116</sup> P-450BM-3;<sup>96,98</sup> P-450eryF;<sup>117</sup> P-450nor;<sup>118</sup> and P-450sca.<sup>119,120</sup> Whether P-450nor should be classified as a cytochrome P-450 is still under debate as it does not actually insert oxygen into a carbon hydrogen bond. The structure of P-450sca is under commercial restrictions and is not available. The other four proteins show extensive structural homology,<sup>30</sup> figure 1.21, which supports the theory that P-450 monooxygenases exhibit common structural features.

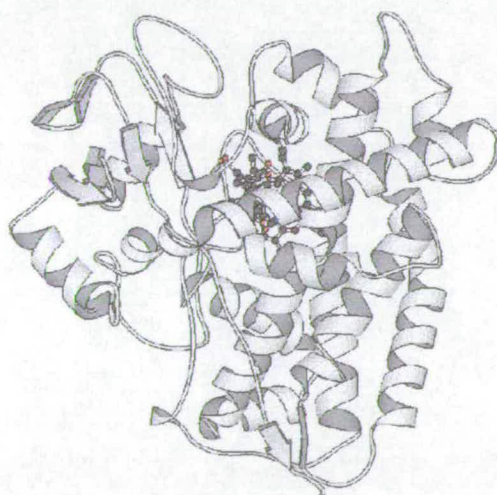


2HPD



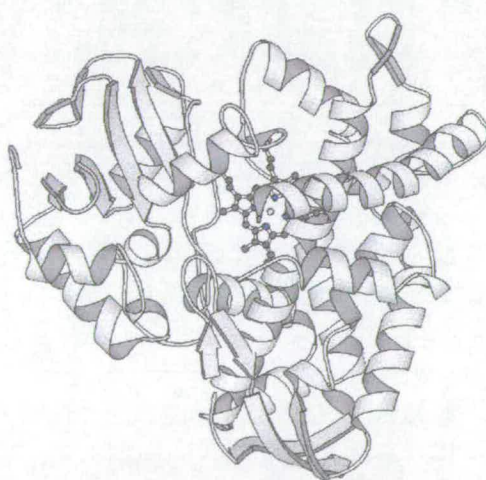
2CPP

Cytochrome P-450BM-3



10XA

Cytochrome P-450cam



1CPT

Cytochrome P-450eryF

Cytochrome P-450terp

Figure 1.21: Structures of four known P-450 monooxygenases showing their strong structural homology.



Cytochrome P-450cam is a 47 kDa b-type<sup>‡</sup> heme protein possessing the common cytochrome P-450 shape of a triangular prism about 30 Å thick with sides of approximately 60 Å, figure 1.22.

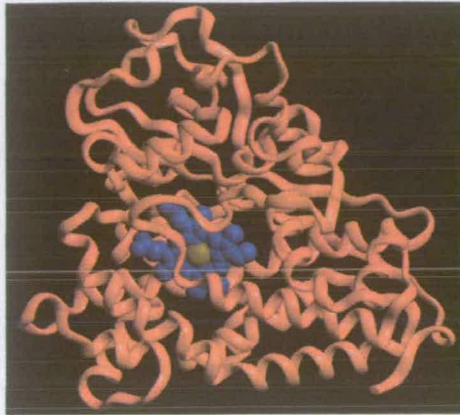


Figure 1.22: Structure of P-450cam. The protein is represented by the red ribbon along the backbone, the heme group is shown in blue and the iron centre is highlighted in yellow.

The protein contains a helical rich region consisting of 12 helices (A-L) which make up 45 % of the structure and a smaller  $\beta$  sheet rich region, figure 1.23.

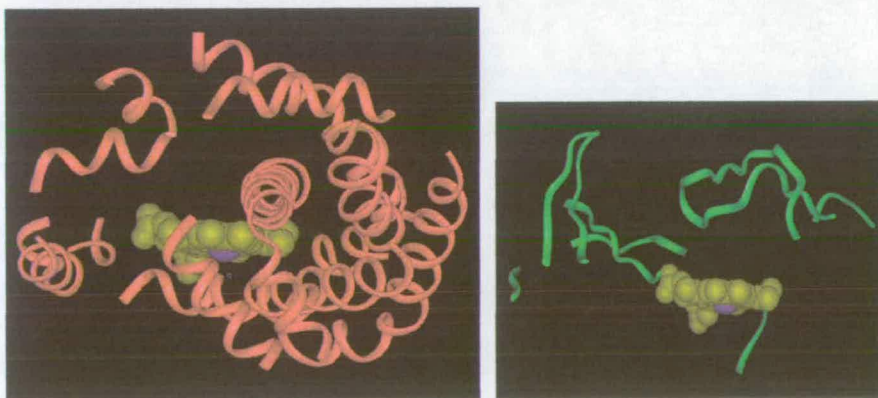


Figure 1.23: Sub structures of P-450cam showing the  $\alpha$ -helix region (left) in red and the  $\beta$ -sheet region (right) in green, the heme group is shown in yellow and the iron centre is highlighted in blue.

<sup>‡</sup> b-type refers to the way in which the heme is attached to the protein

The heme group is embedded between two parallel helices (proximal side to helix L, distal side to helix I) with the catalytic heme centre buried deeply in the active site, figure 1.24.

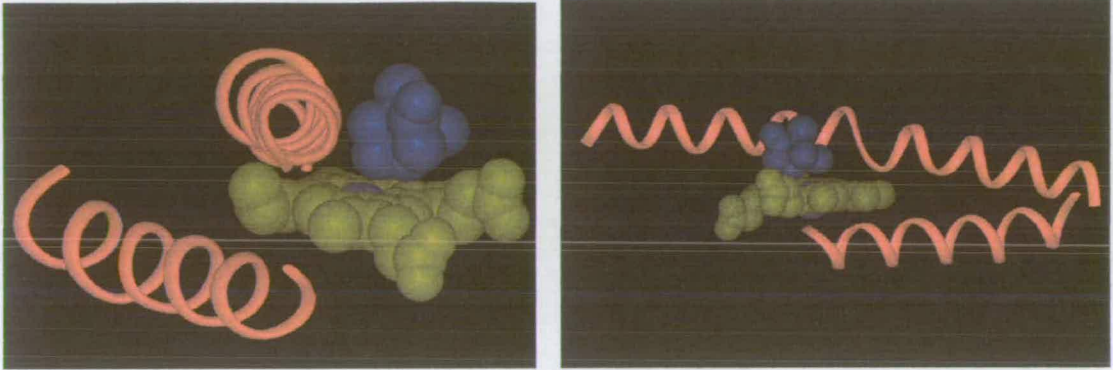


Figure 1.24: Heme orientation held by the helices L and I, the heme group is indicated in yellow, camphor in blue.

#### 1.4.3.1 Structural Studies of Electron Transfer.

It is still unknown which residues are responsible for the transfer of electrons from putidaredoxin to the heme, however Sligar and Poulos have identified a number of key residues that might be involved.<sup>121,122</sup> They propose that four basic residues which lie at the proximal side of the heme: Arg 72; Arg 112; Lys 344 and Arg 364 interact with acidic residues on the surface of putidaredoxin<sup>123</sup> to form the electron transfer chain, figure 1.25.

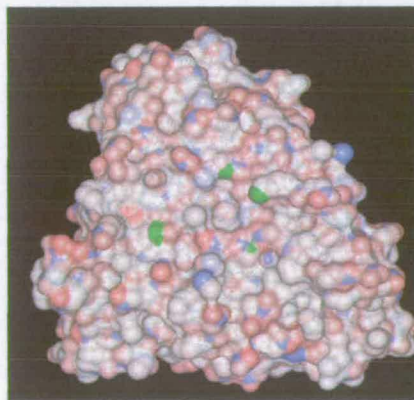


Figure 1.25: Proximal surface of P-450cam, the green residues represent the proposed Pd binding residues, other residues are represented by charge (red negative, blue positive).



The surface of the protein, on the proximal side of the heme, is more hydrophobic than the rest of the protein, allowing the putidaredoxin to bind with only a number of charged residues which could transfer electrons to the heme, figure 1.26.

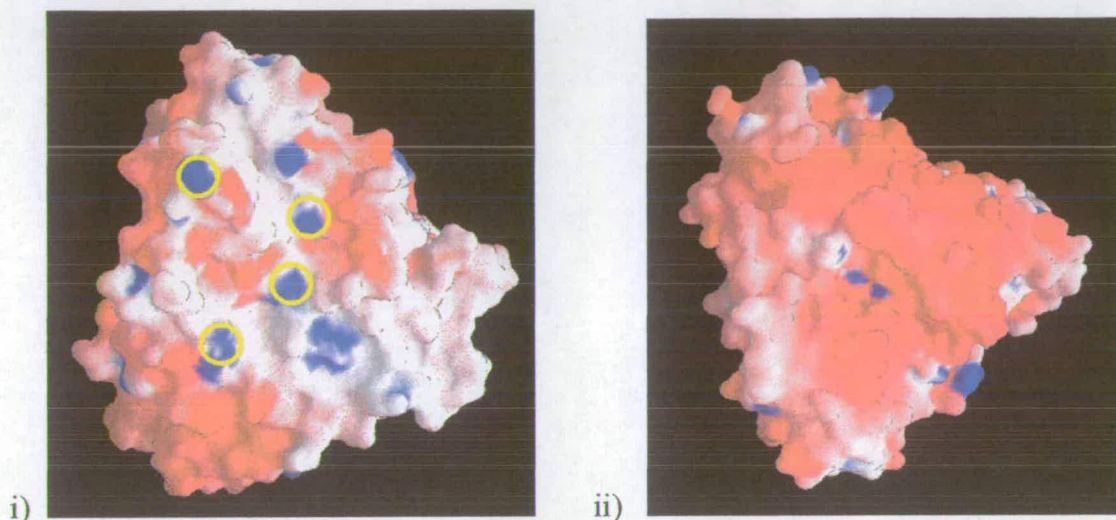


Figure 1.26: Comparison of the surface of proximal side of the protein (i) and distal side (ii), with absolute charges of surface (red for positive, blue negative and white neutral) the proposed binding residues are highlighted with yellow rings.

Evidence that binding occurs *via* these charged residues is found from mutagenesis of Arg 72 or Lys 344, which results in slight changes to the protein-protein binding affinity. However, mutagenesis of Arg 112 both substantially reduces the electron transfer rate and the binding affinity.<sup>103</sup>

The interaction between putidaredoxin and P-450cam can be mimicked by attaching P-450cam to an electrode,<sup>124</sup> with the negative charges on the electrode mimicking putidaredoxin.

#### 1.4.4 The P-450cam Binding Site.

The substrate binding site of cytochrome P-450cam contains a large number of hydrophobic residues, figure 1.27.

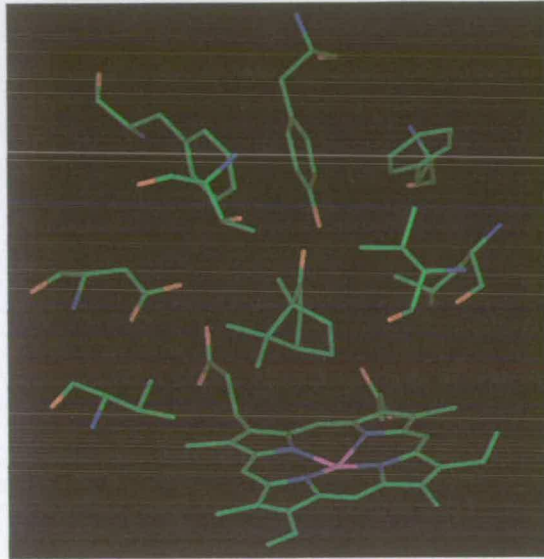


Figure 1.27: Structure of the P-450cam binding site with camphor bound; the heme group is at the bottom of the picture, the camphor in the centre, amino acid residues in a clockwise manner from the heme are Val 295, Asp 297, Thr 185, Phe 87, Tyr 96, Phe 98, Val 247 and Lys 244 (for picture clarity residues Ile 395 and Phe 193 have not been shown).

There are only three hydrophilic residues in the binding site of P-450cam: Thr 252, which may have a role in oxygen binding; Thr 101 (which may stabilize the orientation of Tyr 96) and Tyr 96 itself, which hydrogen bonds to the camphor. The hydrogen bond between Tyr 96 and the carbonyl on the camphor, figure 1.28a, orientates the camphor so that the C-5 carbon is in the closest proximity to the iron centre, figure 1.28b.



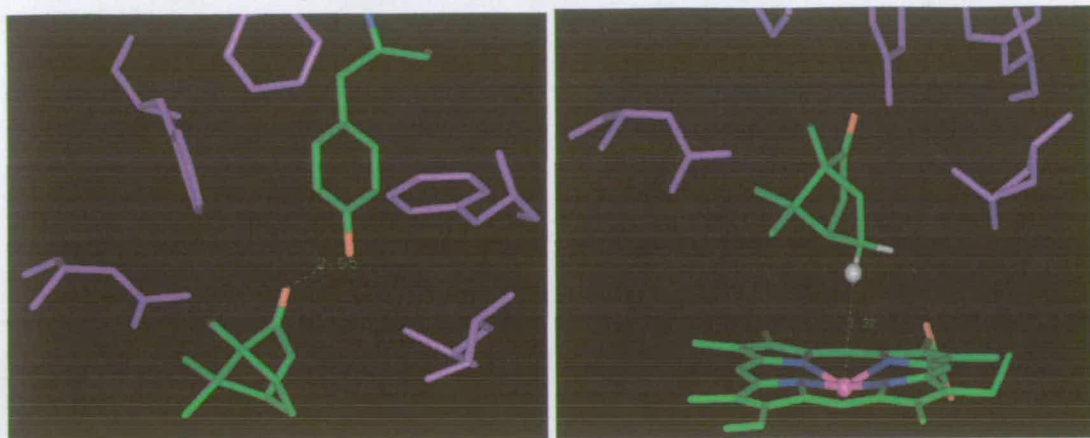


Figure 1.28a: Structure of the binding site showing the hydrogen bond between camphor and tyrosine 96. (Camphor and Y96 are indicated by atom colour code; green for carbon, red for oxygen, the other residues are in purple).

Figure 1.28b: Structure of the binding site showing the orientation of the carbon at position 5 on the camphor to the iron. (Camphor and iron are indicated by atom colour code; green for carbon, red for oxygen and iron, the other residues are in purple).

The hydrophobic residues in the binding site appear to be responsible for orientating the camphor molecule to prevent rotation within the binding site; this is especially true of the interactions between the geminal methyl groups of the camphor and Val 395 and the C1 methyl of camphor and Val 247.

#### 1.4.4.1 Thermodynamic Factors that Control Substrate Binding.

The hydrophobic residues in the binding site, combined with hydrophobic residues in the entrance channel, create a strong attraction for lipophilic substrates. The substrate-free structure has essentially the same binding site as the camphor bound structure except that six water molecules are present in the cavity, one of which is ligated to the heme.<sup>110</sup> The expulsion of these six water molecules from a hydrophobic binding site could provide the energy required for substrate binding. Apart from the expulsion of these water molecules other water molecules may remain in the binding site and play a role in the substrate binding.<sup>125</sup> Osmotic studies

have indicated 28 water molecules are involved in the catalytic process, although some are certainly involved in the Pd/P-450 reaction interface.<sup>126</sup>

#### 1.4.4.2 Substrate Specificity of P-450cam.

For P-450cam many substrate analogues have been studied over the years, figure 1.29.

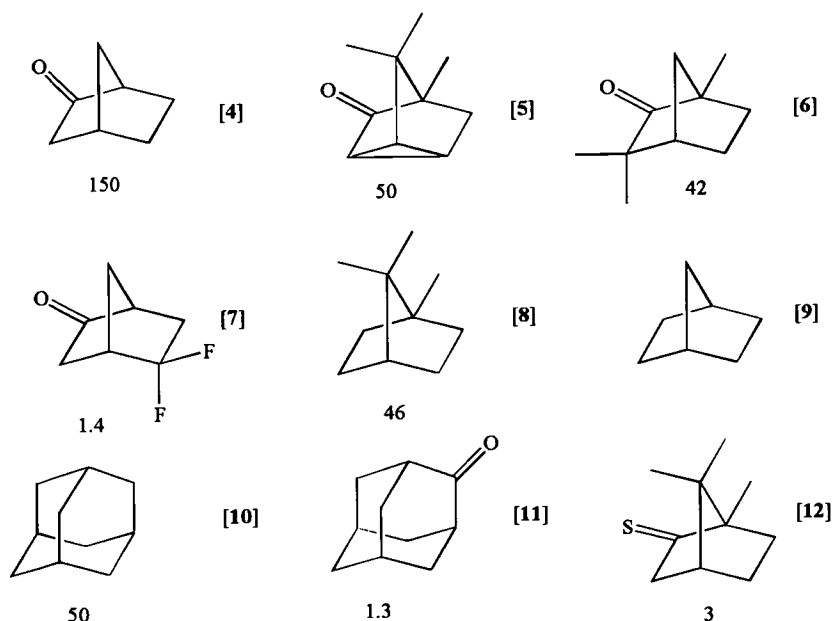


Figure 1.29: Substrates studied with P-450cam, numbers underneath are binding constants: norcamphor [4], pericyclocamphor [5], fenchone [6], 5,5-difluoronorcamphor [7], camphane [8], norbornane [9], adamantane [10], adamantanone [11], thiocamphor [12].<sup>127-129</sup>

Despite high specificity for camphor, a range of substrates have been shown to bind to wild-type P-450cam, including camphor analogues such as thiocamphor [12], adamantane [10], norcamphor [4] and camphane [8], for which the substrate bound crystal structures have been solved.<sup>57,113</sup>

Studies of camphor analogues indicate that removal of the carbonyl group (e.g. adamantane [10], camphane [8] and norbornane [9]) results in a loss of product selectivity, due to the loss of the hydrogen bond to the tyrosine 96 hydroxyl group;



for camphane the products are the 5-*exo*, 6-*exo* and 3-*exo*-hydroxylated compounds. Removal of side chain methyl groups from the substrate has a similar effect (e.g. fenchone [6] and norcamphor [4]) also showing 5-, 6- and 3-hydroxy products. Hydroxylation of adamantane [10] and norbornane [9] indicates that even without a carbonyl group or side chain methyl groups hydroxylation is still possible. By inhibition of H-abstraction from the camphor C-5 position, e.g. 5,5-difluorocamphor [7], other products are observed, e.g. the 9-hydroxy product.<sup>130</sup>

Hydrophobic hydrocarbons such as styrene,<sup>131</sup> pentane<sup>132</sup> and small aromatics compounds<sup>133</sup> (e.g. phenyldiazene, which interacts with the heme group<sup>134,135</sup>) can bind to the active site. Small halomethane and haloethane derivatives have also been shown to bind to P-450cam, the affinity depending on the nature of the halogen atoms.<sup>136-139</sup> Dithiotreitol (DTT) has also been shown to have a binding affinity, in this case with sulfur ligation to the heme.<sup>110</sup> However the binding affinity of all these compounds is considerably lower than that of camphor. The only compounds with a higher affinity for the wild-type enzyme than camphor are inhibitor molecules such as metyrapone and phenylimidazole<sup>114</sup> where binding is enhanced by nitrogen ligation to the iron. Thus the substrate range of P-450cam is limited to molecules smaller in shape and size than camphor.

#### 1.4.4.3 Mutagenesis Effects on P-450cam Substrate Binding.

Mutation of tyrosine 96 to phenylalanine (Y96F)<sup>128,140,141</sup> reduces the selective oxidation at the C-5 position on camphor from 100 % to 92 %. Other products are the C-3, C-6 and C-9 hydroxylated camphor,<sup>142,143</sup> thus mutating the enzyme has the same effect as removing the hydroxyl on the substrate (see above). Mutation of the binding site hydrophobic residues, Val 295 and Val 247 has a similar effect as the removal of the substrate methyl groups (see above).

Selectivity of the enzyme for camphor and the oxidation of the C-5 position is thus due to a combination of the Van der Waals contacts, facilitated by the overall shape of the active and the hydrogen bond with tyrosine 96.

Single point mutations and multiple mutations indicate that the hydrophobicity of the binding site has a role to play in preventing peroxide production when using unnatural substrates, since they effect the active site hydration and thus water access to the reactive oxygen species. Mutagenesis of specific residues believed to be associated with the hydroxylation process, e.g. Thr 252 and Asp 251 has a significant role in disrupting the cytochrome P-450 catalytic cycle.<sup>87,144</sup>

Mutagenesis of the protein has been performed over the years to selectively hydroxylate new substrates, (e.g. ethyl benzene<sup>52</sup>). Although most of these mutations have affected the substrate hydroxylation, none had a significant effect on the range of substrates.

In order to expand the binding site of P-450cam, to hydroxylate a wider range of substrates, a mutation was made to the P-450cam binding site by Flitsch and co-workers.<sup>145</sup> This mutation altered Tyr 96 to an alanine, the Y96A mutant, figure 1.30.

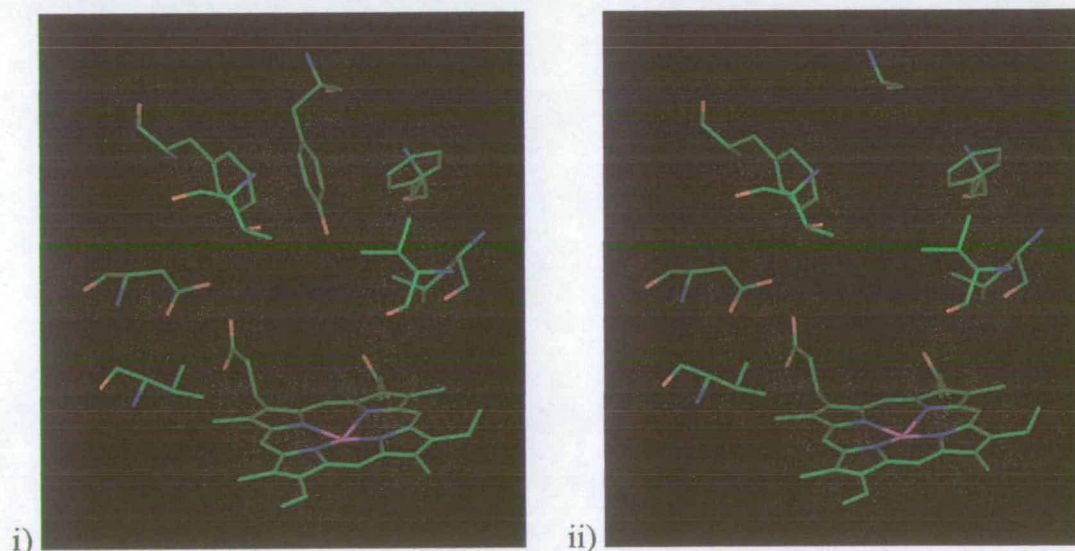


Figure 1.30: Binding site of P-450cam showing the change in creation of the Y96A mutant; wild-type (i) and Y96A mutant model (ii). (Note: this is not based on a X-ray crystal structure, the mutant was generated by deletion of the phenyl and hydroxyl groups of Y96 from the pdb file for wild-type).

This mutation has a number of effects. Firstly it removes the hydroxyl group from the binding site, which has the advantage of making the binding site more

hydrophobic thus changing the potential substrate range. Secondly this mutation potentially created a phenyl pocket between Phe 87 and Phe 98, where the tyrosine ring was previously located. This mutation has the potential to allow hydroxylation of new, more hydrophobic substrates thus substantially expanding the substrate range. Examples of some of the compounds found to be substrates for the new P-450cam mutant are shown below, figure 1.31.

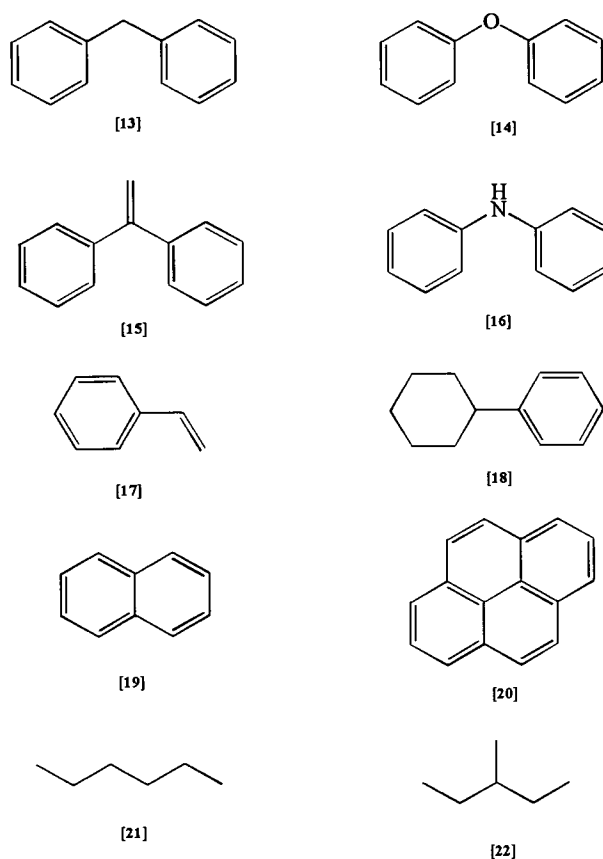


Figure 1.31: Diphenylmethane [13],<sup>145</sup> diphenylether [14], 1,1-diphenylethylene [15], diphenylamine [16],<sup>146</sup> ethenylbenzene [17] (styrene),<sup>131</sup> phenyl cyclohexane [18],<sup>147,148</sup> naphthalene [19], pyrene [20],<sup>149</sup> hexane [21] and 3-methyl pentane [22].<sup>132</sup>

Substrates that fit in the 'aromatic pocket' were believed to be stabilized by  $\pi$ -stacking between the aromatic side chains of the substrate and the aromatic rings on either side of the pocket, i.e. Phe 87 and Phe 98, figure 1.32.

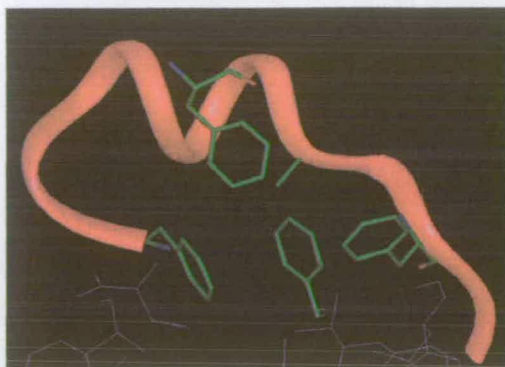


Figure 1.32: Estimated interaction of phenyl substituted substrates with the Y96A mutant binding site, showing a proposed phenyl ring of a substrate interacting with Phe 87 Phe 193 and Phe 98.

The first of these substrates studied was diphenylmethane which is hydroxylated regioselectively by the Y96A mutant, to give 4-hydroxydiphenylmethane, figure 1.33.

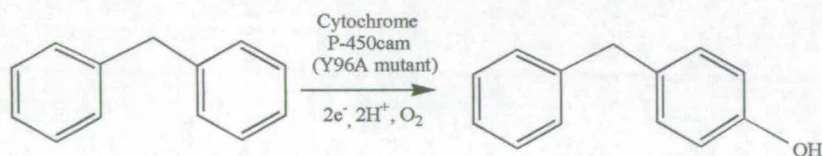


Figure 1.33: Hydroxylation of diphenylmethane [13] by P-450cam mutant Y96A.<sup>145</sup>

No hydroxylation of diphenylmethane was observed with the wild-type enzyme; thus, by a single mutation the substrate specificity of the P-450cam was significantly altered. After these initial studies the hydroxylation of a non-aromatic centre was attempted.<sup>147,149</sup> It was hoped that the phenyl ring of phenylcyclohexane would still confer regioselectivity by fitting into the aromatic pocket. This substrate was hydroxylated with some degree of selectivity yielding three products, figure 1.34.

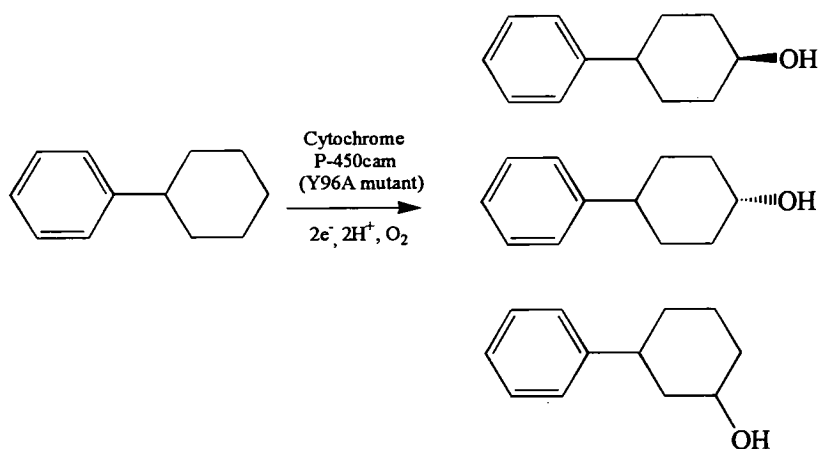


Figure 1.34: Hydroxylation of phenylcyclohexane [18] by P-450cam mutant Y96A.<sup>147</sup>

In this case no aromatic hydroxylation was observed, suggesting that the aromatic ring is fitting in the proposed aromatic binding site.

To explore the substrate specificity further, other aromatic substrates were investigated, figure 1.31. These substrates are small molecules with an aromatic side chain which would have an affinity for the aromatic pocket. These all show similar selectivity to the diphenylmethane and the phenylcyclohexane.

## 1.5 Project Aims.

The current purification protocol allows for the production of small quantities of pure protein and this was deemed to be insufficient for the experiments necessary in our study. Therefore one objective of this project was the further refinement of the purification protocol for P-450cam, to improve yields and protein purity.

With sufficient protein it was hoped that a range of compounds could be tested for their potential to be substrates for the cytochrome P-450cam Y96A mutant. With the biochemical characterization of these compounds the substrate specificity of the enzyme could then be explored.

Using molecular modelling it should be possible to combine the biochemical data on substrates for the Y96A mutant with what is currently known about the wild-

type binding site; thus creating a model of substrate enzyme interactions that could allow the prediction of the binding capacity and products of any new substrates.

A more thorough investigation of the cytochrome P-450cam Y96A binding site, by analysis using X-ray crystallography, should provide a greater understanding of the relationship between the substrate and binding site residues. This will facilitate the understanding of the elements of the binding site responsible for both substrate selectivity and specificity. The X-ray structures could also be used to validate and refine the modelling studies.

## 1.6 References.

- 1)Faber, K. *Bio-transformations in Organic Chemistry*; Springer-Verlag: New York, 1992.
- 2)Hayaishi, D.; Nozaki, M. *Science* **1969**, *164*, 389-396.
- 3)Bernhardt, R. *Reviews of Physiology Biochemistry and Pharmacology* **1996**, *127*, 137-221.
- 4)White, P. W. *Bioorganic Chemistry* **1990**, *18*, 440-456.
- 5)Fruetel, J.; Chang, Y. T.; Collins, J.; Loew, G.; DeMontellano, P. R. O. *Journal of the American Chemical Society* **1994**, *116*, 11643-11648.
- 6)Carnell, A. J.; Roberts, S. M.; Sik, V.; Willetts, A. J. *Journal of the Chemical Society-Chemical Communications* **1990**, 1438-1439.
- 7)Li, S. Y.; Wackett, L. P. *Biochemistry* **1993**, *32*, 9355-9361.
- 8)Vaz, A. D. N. *FASEB Journal* **1994**, *8*, A100.
- 9)DeMontellano, P. R. O. ed. *Cytochrome P-450, Structure, Mechanism and Biochemistry*; 2nd ed.; Plenum Press, New York and London:, 1995.
- 10)Vainshtein, B. K.; Melik-Adamyanyan, W. R.; Barynin, V. V.; Vagin, A. A.; Grebenko, A. I. *Nature* **1981**, *293*, 411-412.
- 11)Garfinkel, D. *Archives of Biochemistry and Biophysics* **1957**, *71*, 111-120.
- 12)Klingenberg, M. *Archives of Biochemistry and Biophysics* **1958**, *75*, 376-386.
- 13)Raag, R.; Poulos, T. L. *Biochemistry* **1989**, *28*, 7586-7592.
- 14)Omura, T.; Sato, R. *Journal of Biological Chemistry* **1962**, *237*, 1375-1376.
- 15)Orrenius, S.; Ernster, L. *Biochemal and Biophysical Research Communications* **1964**, *16*, 60.
- 16)Schenkman, J. B.; Remmer, H.; Estabrook, R. W. *Molecular Pharmacology* **1963**, *3*, 113.
- 17)Muller, H.; Schunck, W.; Kargel, E. *Frontiers in Biotransformation*; Taylor and Francis: London, 1991; Vol. 4.
- 18)Appleby, C. A. *Biochimica et Biophysica Acta* **1969**, *172*, 71.



- 19)Murphy, P. J.; West, C. A. *Archives of Biochemistry and Biophysics* **1970**, *133*, 395.
- 20)Estabrook, R. W.; Cooper, D. Y.; Rosenthal, O. *Biochemistry* **1963**, *338*, 741.
- 21)Shimada, T.; Yamazaki, H.; Mimura, M.; Inui, Y.; Guengerich, F. P. *J. Pharmacological Experimental Therapeutics* **1994**, *270*, 414-423.
- 22)Nerbert, D. *DNA Cell Biology* **1991**, *12*.
- 23)Sariaslani, F. S. *Advances in Applied Microbiology* **1991**, *36*, 133-178.
- 24)DeMontellano, P. R. O. *Trends in Pharmacological Sciences* **1989**, *10*, 354-359.
- 25)Guengerich, F. P.; Macdonald, T. L. *Accounts of Chemical Research* **1984**, *17*, 9-16.
- 26)Miller, R. E.; Guengerich, F. P. *Biochemistry* **1982**, *21*, 1090-1097.
- 27)Bocker, R. H.; Guengerich, F. P. *Journal of Medicinal Chemistry* **1986**, *29*, 1596-1603.
- 28)Pohl, L. R.; Schulick, R. D.; Highet, R. J.; George, J. W. *Molecular Pharmacology* **1984**, *25*, 318-321.
- 29)Vaz, A. D. N.; Coon, M. J. *Federation Proceedings* **1987**, *46*, 1956.
- 30)Hasemann, C. A.; Kurumbail, R. G.; Boddupalli, S. S.; Peterson, J. A.; Deisenhofer, J. *Structure* **1995**, *3*, 41-62.
- 31)Shimada, T.; Iwasaki, M.; Martin, M. V.; Guengerich, F. P. *Cancer Research* **1989**, *49*, 3218-3228.
- 32)Clark, B. J.; Waterman, M. R. *Methods in Enzymology* **1991**, *206*, 100-108.
- 33)Gonzalez, F. J.; Aoyama, T.; Gelboin, H. V. *Methods in Enzymology* **1991**, *206*, 85-92.
- 34)Gonzalez, F. J.; Kimura, S.; Tamura, S.; Gelboin, H. V. *Methods in Enzymology* **1991**, *206*, 93-99.
- 35)Guengerich, F. P.; Brian, W. R.; Sari, M. A.; Ross, J. T. *Methods in Enzymology* **1991**, *206*, 130-145.
- 36)Barnes, H. J. *Methods in Enzymology* **1996**, *272*, 3-14.
- 37)Porter, T. D.; Larson, J. R. *Methods in Enzymology* **1991**, *206*, 108-116.
- 38)Lindberg, R. L. P.; Negishi, M. *Nature* **1989**, *339*, 632-634.



- 39)Ullah, A. J. H.; Murray, R. I.; Bhattacharyya, P. K.; Wagner, G. C.; Gunsalus, I. C. *Journal of Biological Chemistry* **1990**, *265*, 1345-1351.
- 40)Schwalb, H.; Narhi, L. O.; Fulco, A. J. *Biochimica et Biophysica Acta* **1985**, *838*, 302-311.
- 41)Berg, A.; Carlstrom, K.; Gustafson, J. *Biochemical Biophysical Research Communications* **1975**.
- 42)Simgen, B.; Rudiger, V.; Bernhardt, R. *FASEB Journal* **1997**, *11*, 83.
- 43)Asperger, O.; Naumann, A.; Kleber, H. P. *Febs Microbiology Letters* **1981**, *11*, 309-312.
- 44)Koga, H.; Aramaki, H.; Yamaguchi, E.; Horiuchi, T. *Journal of Pharmacobio-Dynamics* **1985**, *8*, S31.
- 45)Mueller, E. J.; Loida, P. J.; Sligar, S. G. *Twenty-five Years of P-450cam Research*; Mueller, E. J.; Loida, P. J.; Sligar, S. G., Ed.; Plenum Press, New York and London, 1995, pp 83-124.
- 46)Poulos, T. L.; Finzel, B. C.; Gunsalus, I. C.; Wagner, G. C.; Kraut, J. *Journal of Biological Chemistry* **1985**, *260*, 16122-16130.
- 47)DeMontellano, P. R. O. *Cytochrome P-450, Structure, Mechanism and Biochemistry*; 1st ed.; Plenum Press, New York and London:, 1986.
- 48)Henggaard, J.; Gunsalus, I. C. *Journal of Biological Chemistry* **1965**, *240*, 4038-4043.
- 49)O'Keeffe, D. H.; Ebel, R. E.; Peterson, J. A. *Methods in Enzymology* **1978**, *52*, 151-157.
- 50)Unger, B. P.; Gunsalus, I. C.; Sligar, S. G. *Journal of Biological Chemistry* **1986**, *261*, 1158-1163.
- 51)Benson, D. E.; Suslick, K. S.; Sligar, S. G. *Biochemistry* **1997**, *36*, 5104-5107.
- 52)Loida, P. J.; Sligar, S. G. *Protein Engineering* **1993**, *6*, 207-212.
- 53)Eble, K. S.; Dawson, J. H. *Biochemistry* **1984**, *23*, 2068-2073.
- 54)Harris, D.; Loew, G. *Journal of the American Chemical Society* **1993**, *115*, 8775-8779.

- 55)Harris, D.; Loew, G. *Journal of the American Chemical Society* **1993**, *115*, 5799-5802.
- 56)Griffin, B. W.; Peterson, J. A. *Journal of Biological Chemistry* **1975**, *250*, 6445-6451.
- 57)Raag, R.; Poulos, T. L. *Biochemistry* **1989**, *28*, 917-922.
- 58)Lambeth, J.; Kriengsiri, S. *Journal of Biological Chemistry* **1985**, *260*, 8810-8816.
- 59)Sligar, S. G. *Biochemistry* **1976**, *15*, 5399-5406.
- 60)Martinis, S. A.; Atkins, W. M.; Stayton, P. S.; Sligar, S. G. *Journal of the American Chemical Society* **1989**, *111*, 9252-9253.
- 61)Hu, S.; Kincaid, J. R. *Journal of the American Chemical Society* **1991**, *113*, 2843-2850.
- 62)Davydov, R.; Macdonald, I. D. G.; Makris, T. M.; Sligar, S. G.; Hoffman, B. M. *Journal of the American Chemical Society* **1999**, *121*, 10654-10655.
- 63)Hintz, M. J.; Mock, D. M.; Peterson, L. L.; Tuttle, K.; Peterson, J. A. *Journal of Biological Chemistry* **1982**, *257*, 4324-4332.
- 64)Rein, I; *Biological Activation of Oxygen*; Schuster, I. Ed.; Gordon and Breach: London, 1986; Vol. 2.
- 65)Atkins, W. M.; Sligar, S. G. *Biochemistry* **1988**, *27*, 1610-1616.
- 66)White, R. E. *Pharmacology & Therapeutics* **1991**, *49*, 21-42.
- 67)Gelb, M. H.; Heimbrook, D. C.; Malkonen, P.; Sligar, S. G. *Biochemistry* **1982**, *21*, 370-377.
- 68)Sligar, S.; Gelb, M.; Heimbrook, D. *Microsomes and Drug Oxidation and Drug Toxicity*; Wiley-Interscience: New York, 1982.
- 69)Ballou, D. P.; Coulter, E. D.; Isaac, I. S.; Beharry, Z. M.; Dawson, J. H. *FASEB Journal* **1997**, *11*, 72.
- 70)Kadkhodayan, S.; Coulter, E. D.; Maryniak, D. M.; Bryson, T. A.; Dawson, J. H. *Journal Of Biological Chemistry* **1995**, *270*, 28042-28048.

- 71) Gelb, M. H.; Toscano, W. A.; Sligar, S. G. *Proceedings of the National Academy of Sciences of the United States of America-Biological Sciences* **1982**, *79*, 5758-5762.
- 72) Roberts, E. S.; Vaz, A. D. N.; Coon, M. J. *Proceedings of the National Academy of Sciences of the United States of America* **1991**, *88*, 8963-8966.
- 73) Pratt, J. M.; Ridd, T. I.; King, L. J. *Journal of the Chemical Society-Chemical Communications* **1995**, 2297-2298.
- 74) Newcomb, M. *Abstracts of Papers of the American Chemical Society* **1996**, *211*, 153-ORGN.
- 75) Newcomb, M.; Letadicbiadatti, F. H.; Chestney, D. L.; Roberts, E. S.; Hollenberg, P. F. *Journal of the American Chemical Society* **1995**, *117*, 12085-12091.
- 76) Newcomb, M.; Chestney, D. L. *Journal of the American Chemical Society* **1994**, *116*, 9753-9754.
- 77) Atkinson, J. K.; Hollenberg, P. F.; Ingold, K. U.; Johnson, C. C.; Letadic, M. H.; Newcomb, M.; Putt, D. A. *Biochemistry* **1994**, *33*, 10630-10637.
- 78) Newcomb, M. *FASEB Journal* **1997**, *11*, 61.
- 79) Newcomb, M. *Abstracts of Papers of the American Chemical Society* **1997**, *214*, 137-INOR.
- 80) Toy, P. H.; Dhanabalasingam, B.; Newcomb, M.; Hanna, I. H.; Hollenberg, P. F. *Journal of Organic Chemistry* **1997**, *62*, 9114-9122.
- 81) Toy, P. H.; Newcomb, M.; Hollenberg, P. F. *Journal of the American Chemical Society* **1998**, *120*, 7719-7729.
- 82) Filatov, M.; Shaik, S. *Journal of Physical Chemistry* **1998**, *102*, 3835-3846.
- 83) Filatov, M.; Harris, N.; Shaik, S. *Angewandte Chemie-International Edition* **1999**, *38*, 3510-3512.
- 84) Shaik, S.; Filatov, M.; Schroder, D.; Schwarz, H. *Chemistry-a European Journal* **1998**, *4*, 193-199.
- 85) Toy, P. H.; Newcomb, M.; Coon, M. J.; Vaz, A. D. N. *Journal of the American Chemical Society* **1998**, *120*, 9718-9719.
- 86) Loida, P. J.; Sligar, S. G. *Biochemistry* **1993**, *32*, 11530-11538.

- 87) Raag, R.; Martinis, S. A.; Sligar, S. G.; Poulos, T. L. *Biochemistry* **1991**, *30*, 11420-11429.
- 88) Filipovic, D.; Paulsen, M. D.; Loida, P. J.; Sligar, S. G.; Ornstein, R. L. *Biochemical and Biophysical Research Communications* **1992**, *189*, 488-495.
- 89) DeMontellano, P. R. O.; Fruetel, J. A.; Collins, J. R.; Camper, D. L.; Loew, G. H. *Journal of the American Chemical Society* **1991**, *113*, 3195-3196.
- 90) Gunsalus, I. C.; Wagner, G. C. *Methods in Enzymology* **1978**, *52*, 166-188.
- 91) Peterson, J. A.; Lorence, M. C.; Amarneh, B. *Journal of Biological Chemistry* **1990**, *265*, 6066-6073.
- 92) Koga, H.; Yamaguchi, E.; Matsunaga, K.; Aramaki, H.; Horiuchi, T. *Journal of Biochemistry* **1989**, *106*, 831-836.
- 93) Brewer, C. B.; Peterson, J. A. *Federation Proceedings* **1986**, *45*, 1506.
- 94) Pochapsky, T. C.; Lyons, T. A.; Kazanis, S.; Arakaki, T.; Ratnaswamy, G. *Biochimie* **1996**, *78*, 723-733.
- 95) Miles, J. S.; Munro, A. W.; Rospendowski, B. N.; Smith, W. E.; McKnight, J.; Thomson, A. J. *Biochemical Journal* **1992**, *288*, 503-509.
- 96) Ravichandran, K. G.; Boddupalli, S. S.; Hasemann, C. A.; Peterson, J. A.; Deisenhofer, J. *Science* **1993**, *261*, 731-736.
- 97) Li, H. Y.; Poulos, T. L. *Acta Crystallographica Section D-Biological Crystallography* **1995**, *51*, 21-32.
- 98) Black, S. D. *Biochemical and Biophysical Research Communications* **1994**, *203*, 162-168.
- 99) Yasukochi, T.; Okada, O.; Hara, T.; Sagara, Y.; Sekimizu, K.; Horiuchi, T. *Biochimica et Biophysica Acta-Protein Structure and Molecular Enzymology* **1994**, *1204*, 84-90.
- 100) Unno, M.; Shimada, H.; Toba, Y.; Makino, R.; Ishimura, Y. *Journal of Biological Chemistry* **1996**, *271*, 17869-17874.
- 101) Shimada, H.; Nagano, S.; Ariga, Y.; Unno, M.; Egawa, T.; Hishiki, T.; Ishimura, Y. *Journal of Biological Chemistry* **1999**, *274*, 9363-9369.

- 102)Roitberg, A. *Abstracts of Papers of the American Chemical Society* **1997**, *213*, 200-BIOT.
- 103)Koga, H.; Sagara, Y.; Yaoi, T.; Tsujimura, M.; Nakamura, K.; Sekimizu, K.; Makino, R.; Shimada, H.; Ishimura, Y.; Yura, K.; Go, M.; Ikeguchi, M.; Horiuchi, T. *FEBS Letters* **1993**, *331*, 109-113.
- 104)Honeychurch, M. J.; Hill, H. A. O.; Wong, L. L. *FEBS Letters* **1999**, *451*, 351-353.
- 105)Christian, J. F.; Unno, M.; Kumar, A. T. N.; Macdonald, I. D. G.; Benson, D. E.; Sligar, S. G.; Champion, P. M. *Biophysical Journal* **1998**, *74*, A133.
- 106)Unno, M.; Christian, J. F.; Benson, D.; Gerber, N.; Sligar, S. G.; Champion, P. M. *Biophysical Journal* **1996**, *70*, WP335.
- 107)Wang, M.; Roberts, D. L.; Paschke, R.; Shea, T. M.; Masters, B. S. S.; Kim, J. J. P. *Proceedings of the National Academy of Sciences of the United States of America* **1997**, *94*, 8411-8416.
- 108)Pochapsky, T. C.; Ye, X. M.; Ratnaswamy, G.; Lyons, T. A. *Biochemistry* **1994**, *33*, 6424-6432.
- 109)Poulos, T. L.; Finzel, B. C.; Howard, A. J. *Journal of Molecular Biology* **1987**, *195*, 687-700.
- 110)Poulos, T. L.; Finzel, B. C.; Howard, A. J. *Biochemistry* **1986**, *25*, 5314-5322.
- 111)Li, H. Y.; Narasimhulu, S.; Havran, L. M.; Winkler, J. D.; Poulos, T. L. *Journal of the American Chemical Society* **1995**, *117*, 6297-6299.
- 112)Schlichting, I.; Jung, C.; Schulze, H. *FEBS Letters* **1997**, *415*, 253-257.
- 113)Raag, R.; Poulos, T. L. *Biochemistry* **1991**, *30*, 2674-2684.
- 114)Poulos, T. L.; Howard, A. J. *Biochemistry* **1987**, *26*, 8165-8174.
- 115)Peterson, J. A.; Lu, J.-Y.; Geisselsoder, J.; Graham-Lorence, S.; Carmona, C.; Witney, F.; Lorence, M. C. *Journal of Biological Chemistry* **1992**, *267*, 14193-14203.
- 116)Hasemann, C. A.; Ravichandran, K. G.; Peterson, J. A.; Deisenhofer, J. *Journal of Molecular Biology* **1994**, *236*, 1169-1185.

- 117) Poulos, T. L.; Cupp-Vickery, J.; Huiying, L. *The Structure of Cytochrome P450*; 2nd ed.; DeMontellano, P. R. O, Ed.; Plenum Press, New York and London, 1995, pp 125-150.
- 118) Shiro, Y.; Fujii, M.; Isogai, Y.; Adachi, S.-I.; Iizuka, T.; Obayashi, E.; Makino, R.; Nakahara, K.; Shoun, H. *Biochemistry* **1995**, *34*, 9052-9058.
- 119) Ito, S.; Matsuoka, T.; Watanabe, I.; Kagasaki, T.; Serizawa, N.; Hata, T. *Acta Crystallographica Section D-Biological Crystallography* **1999**, *55*, 1209-1211.
- 120) Ito, S.; Hata, T.; Watanabe, I.; Matsuoka, T.; Serizawa, N. *FASEB Journal* **1997**, *11*, A776.
- 121) Stayton, P. S.; Poulos, T. L.; Sligar, S. G. *Biochemistry* **1989**, *28*, 8201-8205.
- 122) Stayton, P. S.; Sligar, S. G. *Biochemistry* **1990**, *29*, 7381-7386.
- 123) Geren, L.; Tuls, J.; Obrien, P.; Millett, F.; Peterson, J. A. *Journal of Biological Chemistry* **1986**, *261*, 5491-5495.
- 124) Kazlauskaitė, J.; Westlake, A. C. G.; Wong, L.-L.; O'Hill, H. A. *Journal of the Chemical Society-Chemical Communications* **1996**, 2189-2190.
- 125) Wade, R. C. *Journal of Computer-Aided Molecular Design* **1990**, *4*, 199-204.
- 126) Diprimo, C.; Sligar, S. G.; Hoa, G. H. B.; Douzou, P. *FEBS Letters* **1992**, *312*, 252-254.
- 127) Diprimo, C.; Hoa, G. H. B.; Douzou, P.; Sligar, S. G. *European Journal of Biochemistry* **1992**, *209*, 583-588.
- 128) Atkins, W. M.; Sligar, S. G. *Journal Of Biological Chemistry* **1988**, *263*, 18842-18849.
- 129) Fisher, M. T.; Sligar, S. G. *Biochemistry* **1987**, *26*, 4797-4803.
- 130) Eble, K. S.; Dawson, J. H. *Journal of Biological Chemistry* **1984**, *259*, 4389-4393.
- 131) Nickerson, D. P.; Harford-Cross, C. F.; Fulcher, S. R.; Wong, L. L. *FEBS Letters* **1997**, *405*, 153-156.
- 132) Stevenson, J. A.; Westlake, A. C. G.; Whittock, C.; Wong, L. L. *Journal of the American Chemical Society* **1996**, *118*, 12846-12847.

- 133) Raag, R.; Swanson, B. A.; Poulos, T. L.; DeMontellano, P. R. O. *Biochemistry* **1990**, *29*, 8119-8126.
- 134) Tuck, S. F.; Peterson, J. A.; DeMontellano, P. R. O. *Journal of Biological Chemistry* **1992**, *267*, 5614-5620.
- 135) Tuck, S. F.; Peterson, J. A.; DeMontellano, P. R. O. *FASEB Journal* **1992**, *6*, A321.
- 136) Castro, C. E.; Wade, R. S.; Belser, N. O. *Biochemistry* **1985**, *24*, 204-210.
- 137) Koe, G. S.; Vilker, V. L. *Biotechnology Progress* **1993**, *9*, 608-614.
- 138) Manchester, J. I.; Ornstein, R. L. *Protein Engineering* **1995**, *8*, 801-807.
- 139) Paulsen, M. D.; Ornstein, R. L. *Journal of Computer-Aided Molecular Design* **1994**, *8*, 389-404.
- 140) Atkins, W. M.; Sligar, S. G. *Journal of the American Chemical Society* **1989**, *111*, 2715-2717.
- 141) Atkins, W. M.; Sligar, S. G. *Biochemistry* **1990**, *29*, 1271-1275.
- 142) Diprimo, C.; Huibonhoa, G.; Douzou, P.; Sligar, S. *European Journal of Biochemistry* **1990**, *193*, 383-386.
- 143) Diprimo, C.; Hoa, G. H. B.; Douzou, P.; Sligar, S. *Journal of Biological Chemistry* **1990**, *265*, 5361-5363.
- 144) Vidakovic, M.; Sligar, S. G.; Li, H. Y.; Poulos, T. L. *Biochemistry* **1998**, *37*, 9211-9219.
- 145) Fowler, S. M.; England, P. A.; Westlake, A. C. G.; Rouch, D. R.; Nickerson, D. P.; Blunt, C.; Braybrook, D.; West, S.; Wong, L. L.; Flitsch, S. L. *Journal of the Chemical Society-Chemical Communications* **1994**, 2761-2762.
- 146) Bell, S. G.; Rouch, D. A.; Wong, L. L. *Journal of Molecular Catalysis B-Enzymatic* **1997**, *3*, 293-302.
- 147) England, P. A.; Rouch, D. A.; Westlake, A. C. G.; Bell, S. G.; Nickerson, D. P.; Webberley, M.; Flitsch, S. L.; Wong, L. L. *Journal of the Chemical Society-Chemical Communications* **1996**, 357-358.
- 148) Jones, N. E.; England, P. A.; Rouch, D. A.; Wong, L. L. *Journal of the Chemical Society-Chemical Communications* **1996**, 2413-2414.

149)England, P. A.; HarfordCross, C. F.; Stevenson, J. A.; Rouch, D. A.; Wong, L. L. *FEBS Letters* **1998**, *424*, 271-274.



## 2 Experimental.

---

### 2.1 Materials.

#### 2.1.1 General Chemicals.

All chemical were purchased from Sigma, UK, at analytical grade, unless stated otherwise. Solvents (ethanol, methanol, iso-propanol and ethyl acetate) were purchased from Acros chemical company and 30 % acrylamide solutions from Stratagene. Water was purified on a Millipore MilliRO 5+ with MilliQ plus.

#### 2.1.2 Proteins.

The cytochrome P-450cam C334A-Y96A gene, pAW7292, in *E.coli* cell line JM109, the cytochrome P-450cam C334A gene in *E.coli* cell line JM109, the putidaredoxin gene in *E.coli* cell line JM109 and purified putidaredoxin reductase were provided by Dr. Luet-Lok Wong, Oxford University, UK. Genes for wild-type P-450cam, pUS250, Pd, pRM534 and PdR all in *E.coli* cell line TB1 were provided by Prof. Stephen Sligar, University of Illinois (Urban-Champagne), USA. Spinach ferredoxin (Sd) and spinach ferredoxin reductase (SdR) were purchased from Sigma. Flavodoxin (FLD) and flavodoxin reductase (FLDR) from *E.coli* were provided by Dr. Lisa McIver, University of Edinburgh, UK. Cytochrome P-450BM-3 flavo

domains (BM-3R) were provided by Dr. Andy Munro, University of Edinburgh, UK. Old yellow enzyme (OYE) was provided by Marina Alexeeva, University of Edinburgh, UK.

### **2.1.3 Protein Expression and Purification.**

Luria-Bertani broth (LB) and 'terrific broth' (TB) media and LB agar were provided by in house service using standard recipes.<sup>1</sup> Centrifugation was conducted using a Sorvall RC-4B centrifuge and sonication was performed with the Soniprep™ 150 (MSE). Large scale concentration was achieved with a Millipore 8200 concentrator (300 ml) with membranes YM30 (Amicon), small scale concentration with centriprep™, centricon™, microcon™ and centriplus™ (Amicon) with MWt cut-off of 10 kDa or 30 kDa. HPLC was performed on a Waters 650 protein HPLC system. Protein purification columns and media were purchased from Pharmacia, except DEAE sepharose™, which was purchased from Sigma. Anion exchange column was pumped through with a P1™ pump (Pharmacia) and protein was loaded on the HPLC system with a 50 ml superloop™ (Pharmacia).

### **2.1.4 Protein Analysis.**

Spectral studies were conducted on a Hewlett Packard 8452A UV/Vis spectrometer using 1 ml UV grade disposable cuvettes from Kartell. SDS PAGE was performed with the Mini-protean™ II system (Biorad).

### **2.1.5 Product Analysis.**

GC/MS analysis was performed with a crosslinked 5 % siloxane column (30 m x 0.25 mm, HP-5MS, Hewlett Packard), with a Hewlett Packard 6890 GC and 5973 mass spectrum analyzer.

### 2.1.6 Crystallography.

Polyethyleneglycol (PEG) was obtained from Fluka, UK. Dimethyldichloro silane (2 %), in 1,1,1-trichloroethane was purchased from Fisons, UK. Cryo freezing containers and mounting loops were purchased from Hampton Research, US; Limbro™ plates were purchased from, ICN and cover slips were bought from BDH. Crystallographic computation was performed on Silicon Graphics Indy™, O<sup>2</sup>™ and IndigoII™ workstations, with processing performed on a 200 MHz R10000 Challenge L from Silicon Graphics. Software used included InsightII™ (Biosym), Macromodel™, ccp4™,<sup>2</sup> SCALEPACK™ and DENZO™,<sup>3</sup> O™<sup>4</sup> and SHELX 97™.<sup>5</sup>

### 2.1.7 Modelling.

Molecular modelling was performed on Silicon Graphics Indy™, O<sup>2</sup>™ and IndigoII™ workstations, the package used was Insight II™ (Biosym).

DOCK and all its supplemental programs were provided by Irwin Kuntz, University of California (San Francisco), USA. Substrates for DOCK were built using InsightII™ (Biosym) and Sybyl™ (Tripos associates).

## 2.2 General Experimental Conditions.

Long-term storage of protein solutions was generally performed by snap freezing in liquid nitrogen and storage at  $-80\text{ }^{\circ}\text{C}$ . Solutions contained 30 % glycerol, to enhance protein stability during the freezing process. During experiments protein solutions were stored on ice prior to use. All protein buffer solutions were filtered with cellulose membranes, 20  $\mu\text{m}$  pore size, from Altech. Columns used for glycerol and substrate removal by gel filtration were PD-10<sup>TM</sup> columns (Pharmacia) with Sephadex<sup>TM</sup> G25 resin.

### 2.2.1 GC method.

Organic metabolites from the biotransformations were extracted by addition of 500  $\mu\text{l}$  of ethyl acetate, agitated and centrifuged at 10,000 x g. The organic phase was extracted and analyzed on a GC/MS, with a crosslinked 5 % siloxane column (30 m x 0.25 mm, HP-5MS). Elution was achieved using a gradient program from 100 to 290  $^{\circ}\text{C}$ , at flow rate 1 ml/min, table 2.1.

Table 2.1: GC temperature gradient used to elute biotransformation products.

time	temperature ( $^{\circ}\text{C}$ )
0	100
5	100
9	200
14	200
18	290
28	290

## **2.3 Growth and Expression of Cytochrome P-450cam.**

### **2.3.1 General Growth Conditions.**

The growth characteristics of the JM109 *E.coli* strain containing the pAW7292 plasmid was determined previously <sup>6</sup> and optimized to give a maximal growth in TBcm (TB broth with 1 mM chloroamphenicol) medium at 30 °C. All cultures were inoculated from a single colony on a TBcm agar plate, scaled up, first to a 5 ml overnight culture, then to a 150 ml overnight culture. This culture was then split into the appropriate volume of TB and grown until induction. Protein expression was induced at an optimal OD<sub>600</sub> (optical density of 600 nm) of 1.0 by raising the temperature to 37 °C and then allowing the cells to grow for at least a further 6-16 hours. Camphor was also added at induction, to a final concentration of 1 mM, as a stabilizing agent for the P-450cam.<sup>7</sup> Expression was confirmed by SDS PAGE (12.5 % polyacrylamide).

### **2.3.2 Large Scale Growth, in a 50 L Fermenter.**

1 L of starter culture was pooled and transferred to a fermenter containing 40 L TBcm. Protein expression was induced as above and the culture was grown for 16 hours. Cells were harvested with the built-in centrifuge, at 5000 x g, for 30 minutes, with continuous extraction of supernatant. Typical yields were ~9 g/L of cell pellet containing ~37 mg/L of P-450cam protein.

## **2.4 Purification of P-450cam.**

### **2.4.1 Preparation of the Cell-Free Extract.**

Revised from previous small scale protocol,<sup>6</sup> about 40-50 g of cell pellet was thawed from -20 °C, resuspended in 320 ml of Pi-cam (40 mM potassium phosphate

buffer with 1 mM camphor, pH 7.4), then divided into 8 x 40 ml portions. The samples were then sonicated, on ice, in a Soniprep™ 150, at medium power, for 7 x 1 minute. The suspension was harvested at 20000 x g, for 30 minutes, at 4 °C. The supernatant was decanted and stored at 4 °C. The cell debris was then resuspended in 2 x 40 ml of Pi-cam with gentle mixing. The suspension was then subjected to a further 2 x 1 minute sonication, with the above conditions. The solution was then centrifuged at 20000 x g for 10 minutes at 0 °C. The supernatant was decanted and pooled with the previous supernatant at 4 °C.

#### **2.4.2 Anion Exchange Chromatography on a DEAE Sepharose™ Column.**

A XK50™ column, 50 mm x 150 mm filled with DEAE sepharose™ fast-flow, vol. ~300 ml, was equilibrated with 7 column volumes of Pi-cam (~2 L) and the cell free extract (~200 ml) was loaded, using a P1™ pump at a flow rate of 6 ml/min. The column was washed through with 3 column volumes of Pi-cam (~1 L), the protein was eluted with a stepwise increasing salt gradient of 25 mM KCl every 30 minutes, at the same flow rate as before. Elution of the P-450cam (as seen by movement in the red/brown coloured band), began at 100 mM KCl. The gradient was increased to 150 mM and maintained until all the protein had eluted. Fractions were analyzed on a UV/Vis spectrometer at 417 nm. The fractions with  $OD_{417} > 0.5$  were pooled and stored at 4 °C.

#### **2.4.3 Concentration and Desalting by Ultrafiltration.**

The fractions from the DEAE column (~150-200 ml) were concentrated in a 300 ml pressure concentrator cell with a YM30 membrane (30 kDa cut-off). The sample was desalted by initial dilution with Pi-cam to 300 ml and two successive dilution/concentration steps from 300-70 ml. The final protein solution (~20-50 ml) was stored at 4 °C (< 24 hours).

#### 2.4.4 Anion Exchange Chromatography with SourceQ™ Media by HPLC.

A HR 16/10™ column (16 x 100 mm), containing 14 ml of Source15Q™ media, was equilibrated with buffer Pi-cam. Protein (30-40 ml) was loaded by means of a 50 ml superloop onto the column with a P1 pump™ at a flow rate of 7 ml/min. The optical density of the column effluent was measured at 280 nm, until waste protein was removed. The P-450cam protein was eluted with an increasing salt gradient, table 2.2.

Table 2.2: The HPLC program used to elute the P-450cam.

Time (minutes)	Flow rate (ml/min)	% Buffer A	% Buffer B
(loading)	7	100	0
0	7	100	0
1	10	100	0
3	10	95	5
15	10	75	18
22	10	50	50
25	10	0	100
26	10	0	100
26.5	10	100	0
28	10	100	0
29	7	100	0
30	0	100	0

Buffer A= Pi-cam

Buffer B= Pi-cam, 1 M potassium chloride

Eluent was analyzed at 417 nm and fractions containing protein were pooled based on their purity. Low purity fractions (i.e. 280 nm: 417 nm ratio < 1.2) were concentrated and desalted and the HPLC step was repeated.

#### 2.4.5 Preparation of Protein Pure Enough for Crystallographic Studies.

A HR 16/10™ column (16 x 100 mm), containing 10 ml of phenyl sepharose™, was equilibrated with buffer AS (1 M ammonium sulfate, 40 mM potassium phosphate pH 7.4). Protein solution was made up to 1 M ammonium sulfate

concentration, with a saturated ammonium sulfate solution, at 4 °C. Protein solution (20 ml) was loaded by means of a 50 ml superloop, onto the column at a flow rate of 5 ml/min. The P-450cam protein was then eluted with a reducing salt gradient, table 2.3.

Table 2.3: The HPLC program used to elute the protein.

Time (minutes)	Flow rate (ml/min)	% Buffer A	% Buffer B
(loading)	5	100	0
0	5	100	0
1	8	100	0
3	8	100	0
5	8	50	50
15	8	30	70
18	8	0	100
20	8	0	100
22	8	100	0
26	8	100	0
29	5	100	0
30	0	100	0

Buffer A= AS

Buffer B= Pi-cam

Eluent was analyzed at 417 nm and was pooled and stored at 4 °C. Low purity fractions (i.e. 280 nm: 417 nm ratio < 1.3) were concentrated, desalted and the HPLC step was repeated. Protein was then snap frozen in liquid nitrogen stored in 30 % glycerol at -80 °C.

#### 2.4.6 Removal of Camphor and Glycerol by Gel Filtration.

Protein that had been stored in 30 % glycerol, at -80 °C, was allowed to thaw to -20 °C overnight and then allowed to warm to 4 °C. Protein solution (1 ml) was loaded on a prepacked 10 ml Sephadex™ G25 gel filtration column (PD-10™). The protein solution was initially allowed to pass through the column under gravity. Once the head solution was depleted, the protein was washed through with buffer Pi (40



mM potassium phosphate buffer, pH 7.4). The fractions containing P-450 were then collected and pooled, fractions were analyzed by UV/Vis to make sure all camphor was removed. This step was repeated for fractions which showed the presence of camphor.

## **2.5 Growth and Expression of Putidaredoxin.**

### **2.5.1 General Growth Conditions.**

The Pd gene was provided in *E.coli* strain JM109. The growth characteristics of the JM109 *E.coli* containing the Pd gene were the same as for the P-450cam, pAW7292, but were optimized to give maximal growth in TBamp (TB broth with 1 mM ampicillin) media at 37 °C. All growths were started from a single colony on a TBamp agar plate. Protein expression was induced at an optimal OD<sub>600</sub> of 1.0 by addition of 0.3 mM isopropylthiogalactoside (IPTG) and the cells were grown for 4 hours. Expression was confirmed by SDS PAGE (12.5 % polyacrylamide).

### **2.5.2 Large Scale Growth, in 1 L Shake Flasks.**

From 5 ml overnight culture the *E.coli* was grown overnight in 4 x 150 ml TBamp at 37 °C. The cultures were then pooled and transferred to 15 x 1 L baffled shake flasks, each containing 300 ml of TBamp. After reaching an OD<sub>600</sub> of 1 the culture was induced by addition of 0.3 mM IPTG. The culture was then grown for 4 hours. The cell solution was centrifuged at 10000 x g for 20 minutes at 4 °C. Typical yields were ~7 g/L of cell pellet containing ~ 30 mg/L of Pd.

## **2.6 Purification of Putidaredoxin.**

### **2.6.1 Preparation of the Cell-free Extract.**

Revised from previous small scale protocols,<sup>8</sup> the cell pellet (~30 g) was thawed from -20 °C, resuspended in 400 ml of buffer T (50 mM Tris buffer, pH 7.2), then divided into 10 x 40 ml aliquots. The samples were then sonicated on ice, at medium power, for 5 x 1 minute. The suspension was centrifuged at 20,000 x g for 30 minutes at 4 °C. The supernatant was decanted and stored at 4 °C. The pellets were then resuspended in 2 x 40 ml of buffer T with gentle mixing. These pellets were then subjected to a further 2 x 1 minute sonication with the above conditions. The solution was then centrifuged again at 20000 x g for 20 minutes, at 0 °C. The supernatant was decanted and pooled with the previous supernatant, at 4 °C.

### **2.6.2 Anion Exchange Chromatography on a DEAE Sepharose Column.**

A XK50™ column, 50 mm x 150 mm filled with DEAE sepharose fast flow, (~300 ml), was equilibrated with 7 column volumes of buffer T (~2 L) and the cell free extract (~150 ml) was loaded. The column was washed through with 2 column volumes of buffer T and the protein was then eluted with an increasing salt gradient of 25 mM KCl every 30 minutes, using a P1™ pump at a flow rate of 6 ml/min. At 175 mM KCl when the protein starts to elute (as seen by movement in the dark brown coloured band), the gradient was increased to 200 mM and maintained until the protein eluted. Fractions were analyzed on a UV/Vis spectrometer at 405 nm. The fractions with OD<sub>405</sub> > 0.1 were pooled and stored at 4 °C.

### **2.6.3 Desalting by Dialysis.**

The fractions from the DEAE column (~150-200 ml) were dialyzed against buffer T with a 8 kDa cut-off membrane at 4 °C, overnight.

### 2.6.4 Anion Exchange Chromatography with SourceQ Media and a Waters Protein HPLC System.

A HR 16/10™ column (16 x 100 mm), containing 14 ml of Source15Q™ media was equilibrated with buffer T. Protein (20-40 ml) was loaded onto the column by means of a 50 ml superloop, at a flow rate of 7 ml/min. The OD of the column effluent was measured at 280 nm, until waste protein was removed. The Pd protein was then eluted with an increasing salt gradient, table 2.4.

Table 2.4: The HPLC program used to elute the Pd.

Time (minutes)	Flow rate (ml/min)	% Buffer A	% Buffer B
(loading)	7	100	0
0	7	100	0
1	10	100	0
5	10	85	15
20	10	10	90
21	10	0	100
22	10	0	100
23	10	100	0
23.50	10	100	0
24	7	100	0
25	0	100	0

Buffer A= buffer T

Buffer B= buffer T, 1M potassium chloride

Eluent was analyzed at 405 nm and samples with Pd with purity (405 nm: 280 nm) of greater than 0.3 were stored at 4 °C.

### 2.6.5 Final Purification Step of Pd.

A HR 16/10™ column (16 x 100 mm), containing 10 ml of Phenyl Sepharose™, was equilibrated with buffer AS (1 M ammonium sulfate, 40 mM potassium phosphate pH 7.2). Protein solution was made up to 1 M ammonium

sulfate concentration with a saturated ammonium sulfate solution, at 4 °C. Protein (10 ml) was loaded onto the column by means of a 50 ml superloop, at a flow rate of 5 ml/min. The Pd protein was then eluted with a reducing salt gradient, table 2.5.

Table 2.5: The HPLC program used to elute the Pd.

Time (minutes)	Flow rate (ml/min)	% Buffer A	% Buffer B
(loading)	7	100	0
0	7	100	0
2	7	90	10
10	7	50	50
23	7	0	100
25	7	0	100
26	7	0	100
28	7	100	0
29	7	100	0
30	0	100	0

Buffer A= 100 % AS in buffer T

Buffer B= buffer T

Eluent was analyzed at 405 nm and fractions with purity (405 nm: 280 nm) of greater than 0.4 were snap frozen in liquid nitrogen and stored in 30 % glycerol at -80 °C.

## 2.7 Protein Quantification.

P-450cam content and purity was quantified by UV/Vis spectroscopy by measurement of OD at the Soret band maximum (at 390 nm for wild type protein in the presence of camphor and 417 nm for unbound wild type protein and Y96A mutant). Pd content and purity was quantified by UV/Vis spectroscopy by measurement of OD at the maximum of 405 nm. PdR content or purity was quantified by UV/Vis spectroscopy by measurement of OD at the maximum of 454 nm. Protein concentration was measured on a diode array spectrometer with a 1 ml cuvette and solutions with a  $\lambda$ -max over 1 were diluted.

## **2.8 Binding Assays.**

### **2.8.1 Preparation of Protein Solution for Binding Experiments.**

Protein that had been stored in 30 % glycerol, at -80 °C, was allowed to thaw at -20 °C overnight and then allowed to warm to 4 °C. Glycerol and camphor were removed using the same procedure as for the preparation of crystallographic quality protein. Protein was then diluted to a concentration of ~1.7  $\mu$ M and stored on ice until use.

### **2.8.2 Assay to Determine Maximum Binding via Shift in the Soret Band.**

A UV/Vis spectrum of 1 ml of protein solution (~1.7  $\mu$ M, OD<sub>417</sub> of 0.2), equilibrated to 30 °C for 5 minutes, was measured on a diode array spectrometer, in the range of 270-500 nm and the absorbance at 390 nm, 404 nm and 417 nm were monitored. 20  $\mu$ g of substrate (from a 10 mg/ml stock solution in ethanol), was added and the solution was allowed to equilibrate to 30 °C for 1 minute. If binding was not observed then addition of substrate was repeated until maximum binding was achieved or precipitation of the substrate occurred.

### **2.8.3 Assay to Determine Binding Constants.**

A UV/Vis spectrum of 1 ml of protein solution (~10-20  $\mu$ M), equilibrated to 30 °C for 5 minutes, was measured on a diode array spectrometer, in the range of 270-500 nm and the absorbencies at 390 nm, 404 nm and 417 nm were monitored. Substrate, 0.2  $\mu$ M, was added, (from a 0.1 mg/ml stock solution in ethanol), and the spectrum recorded as before. If no significant shift of the Soret band towards the known maximum binding shift was observed the concentration was increased by

additions of 1 mg/ml stock solution. This data collection continued until at least 10 different, non extreme, data points were gathered.

#### **2.8.4 Assay to Determine Binding at a Maximum Soret Shift for Oxygen Ligation.**

DTT (dithiotreitol) was added to the substrate free protein solution at a concentration of 50 mM, from a 1 M stock solution. Substrate was then added in incremental amounts, as above, except the absorbance at 460 nm and 417 nm were measured.

### **2.9 Solvent Stability of Cytochrome P-450cam.**

The turnover of camphor by P-450cam was performed with 1  $\mu$ M P-450cam, 1  $\mu$ M PdR, 5  $\mu$ M Pd, 4 mM NADH and 10 mM camphor in a 1 ml reaction volume. The reaction was performed in a sealed 50 ml vessel and shaken continuously for 3 hours, at 30 °C. Product was extracted with standard procedures.

### **2.10 Optimization of the Electronic Transfer Agents.**

#### **2.10.1 Preparation of Enzymes and Chemicals.**

P-450cam was used from a 40 mM potassium phosphate buffer, (pH 7.4), stock solution, PdR and Pd were in 50 mM Tris (pH 7.2). Crude P-450cam extract was a  $\sim$ 1  $\mu$ M P-450cam concentration in buffer Pi. Spinach ferredoxin and spinach ferredoxin reductase were made up in 50 mM Tris buffer pH 7.2, flavodoxin and flavodoxin reductase were provided in 40 mM Tris buffer, pH 7.4, P-450BM-3 flavo domain and OYE were provided in 40 mM Tris buffer pH 7.2. Phenazinemethylsulphate (PMS) stock solution was made up to 10 mM in water.

### 2.10.2 Biotransformation Conditions.

The standard for the experiment was the turnover of camphor by P-450cam with 1  $\mu\text{M}$  P-450cam, 1  $\mu\text{M}$  of PdR, 5  $\mu\text{M}$  Pd, 500  $\mu\text{M}$  NADH and 1 mM of camphor, in a 1 ml reaction volume. The reaction was performed in a 10 ml vessel and shaken continuously for 3 hours, at 30 °C.

Reactions with alternate reductases were performed in similar conditions with 1  $\mu\text{M}$  of the reductase and 5  $\mu\text{M}$  of the ferredoxin or flavodoxin. P-450BM-3 flavo-domain was made up to 2  $\mu\text{M}$  as it contains both reductase and flavodoxin. OYE was treated as a reductase and used at 1  $\mu\text{M}$  concentration. Crude P-450cam extract was at an estimated P-450cam level of at least 2  $\mu\text{M}$ . PMS concentrations varied between 10-500  $\mu\text{M}$ . Products were analyzed as before. The combinations used of the various electron transfer agents used is presented in the results section.

### 2.11 Biotransformations with P-450cam.

Substrate turnovers were performed at 30 °C with 1  $\mu\text{M}$  P-450cam, 5  $\mu\text{M}$  Pd, 1  $\mu\text{M}$  PdR, 50  $\mu\text{g/ml}$  catalase in buffer Pi containing 100 mM KCl. The reaction was initiated by the addition of 5 mg NADH and 5  $\mu\text{l}$  of substrate (from ~10 mg/ml ethanol stock). Reaction mixtures were shaken for 1 hour and the reaction was stopped by addition of ethyl acetate.

Authentic standards of 5-*exo*-hydroxycamphor were synthesized by turnover of wild type enzyme. 6-*exo* and 6-*endo*-hydroxycamphor were isolated from fermentation of (1*R*)-(+)-camphor with *Rhodococcus* species NCIMB 9784 and provided by Dr. Gideon Grogan, University of Edinburgh, UK.

## **2.12 Crystallography of the Cytochrome P-450cam Y96A C334A Mutant.**

### **2.12.1 Preparation of Protein.**

Purified P-450cam from a 1 ml, 20  $\mu$ M stock solution, was removed from cryostorage at -80 °C and allowed to thaw at 4 °C. Camphor was removed with standard protocols and its removal and protein concentration was confirmed by UV/Vis spectroscopy. This solution was then concentrated by a factor of ten by ultrafiltration (from ~2 ml - 200  $\mu$ l) using a Centricon™ 30 concentrator, spun at 6000 x g for 2 x 20 minutes, at 4 °C. The solution was then diluted with 1 ml of buffer Pi containing 1 mM of relevant substrate. The concentration step and dilution steps were repeated with an incremental addition of 1 mM substrate until a concentration of 4 mM substrate concentration was achieved. Protein solution was then concentrated to a final concentration of 1 mM (45 mg/ml).

### **2.12.2 Crystallization Screens.**

The P-450cam mutant was co-crystallized in the presence of substrate using vapor diffusion with the hanging drop technique.<sup>9</sup> Drops consisted of 2  $\mu$ L well solution and 2  $\mu$ L protein solution. Screening plates contained 24 wells of ~3-4 ml in a 6 x 4 grid. Cover slips were prepared prior to use with silicanization, by vapor diffusion by leaving the cover slips in a vacuum in the presence 2 % dimethyldichloro silane in 1,1,1-trichloroethane for 2 hours. Screens were performed with variation in precipitant solution, either ammonium sulfate (AS) or different sized polyethyleneglycol (PEG) and their relative concentrations. Additionally, variations were made in protein concentration, pH and in the presence or absence of substrate (details are presented in results section). In all wells, solutions consisted of 40 mM potassium phosphate and 200 mM potassium chloride. Screening was performed at 18, 14 or 4 °C.



### 2.12.3 Freezing Condition Screening.

The freezing solution consisted of the well solution from which the crystal came from, plus glycerol. The amount of glycerol was varied and the crystals were screened by X-ray diffraction. The criterium for the selected glycerol concentration was a value that was sufficiently high to prevent the formation of ice, as observed by the detection of ice rings on the image plate. However, the glycerol concentration must be low enough to prevent premature fracturing of the crystal, as observed under a microscope.

### 2.12.4 Final Conditions for Crystal Growth.

Wells consisted of 40 mM potassium phosphate buffer (pH 7.4), 200 mM potassium chloride, 4 mM substrate. Precipitant was PEG 4000 with wells from 16-24 % or 14-26 %. Drops consisted of 2  $\mu$ l of 45 mg/ml protein solution in 40 mM potassium phosphate buffer (pH 7.4) and 2  $\mu$ l of well solution. Trays were left at 4 or 14  $^{\circ}$ C, crystals formed between 3 days and 2 weeks.

### 2.12.5 Final Conditions for Cryofreezing of Cytochrome P-450cam Crystals.

Crystals were transferred at 18  $^{\circ}$ C, by mounting loop, to a cryo-solution consisting of well solution and 21 % glycerol (pH 7.4) and allowed to diffuse. The crystal was then removed from the cryo-solution in a 0.3-0.5 mm mounting loop, so that the crystal lies head-tail in the loop (figure 2.1). The crystal was then snap frozen at -80  $^{\circ}$ C and stored in liquid nitrogen.

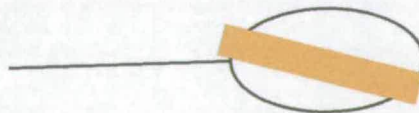


Figure 2.1: Orientation of the crystal in the mounting loop.

### 2.12.6 Crystal Screening.

Crystals were screened in-house on a 30 mm Mar imaging plate with a Enraf-Nonius FR571 rotating anode generator, which generated CuK  $\alpha$ -radiation at a wavelength of  $\lambda=1.5412 \text{ \AA}^{-1}$ . Suitable crystals were then run for one frame of collection for 20 minutes, with the orientation of the beam towards the centre of the crystal. The crystal was then repositioned off-centre at either end of the crystal and data collected again for 20 minutes. Crystals with diffraction better than  $2.5 \text{ \AA}$  were then stored for later data collection.

### 2.12.7 Data Collection.

For the camphor and *N*-Cbz-2,6-dimethylpiperidine bound protein crystals diffraction data was collected at the Daresbury Synchrotron Facility, by Prof. Malcolm Walkinshaw and Dr. Paul Taylor, University of Edinburgh, UK, on a 30 mm Mar image plate, at station 7.2, with wavelength of  $1.488 \text{ \AA}^{-1}$ . For the 4-hydroxydiphenylmethane, diphenylmethane, phenylcyclohexane and bicyclohexyl bound protein crystals data was collected in-house, as above.

For in-house data collection crystals were mounted in a continuous stream of liquid nitrogen. Collection was at a psi-angle of  $1.5^\circ$  or  $1.0^\circ$  depending on mosaicity. Data collection was for 30 minutes per frame, with the number of frames collected being dependent on time restraints, but always at least 100 frames.

### 2.12.8 Data Reduction and Refinement.

The data was indexed and reduced with DENZO™ and SCALEPACK™. The structure of the P-450cam mutant with 4-hydroxydiphenylmethane bound was solved by molecular replacement with AMoRe™, (by Dr. Paul Taylor, University of Edinburgh, UK), using the camphor bound wild type structure as a model. The camphor bound structure was solved with the 4-hydroxydiphenylmethane bound structure as a model. Subsequent structures were solved based upon the camphor

bound model. Refinement was carried out using SHELX™ and model building with O. Over refinement was monitored by following the  $R_{\text{free}}$  value at a level of 5 %.

#### 2.12.8.1 Substrate Building.

Substrates were built in InsightII and minimized with a Discover™ (Biosym) force field. For substrates with multiple low energy conformations, all conformations were built. Substrate conformations were stored as pdb format (protein databank format).

Substrate parameters for SHELX™ were built by hand, 'DFIX' and 'DANG' parameters were based on the measured distances of the energy minimized structure. For substrates with multiple orientations the 'DANG' parameters that configure between these conformers were omitted from the SHELX™ input file. Aromatic rings were additionally restrained with the 'FLAT' restraint set at a level of 0.1.

#### 2.12.8.2 4-Hydroxydiphenylmethane Structure.

The individual residues of the Y96A protein were modelled into the  $\alpha$ -weighted 2Fo-Fc difference map contoured between a level of 1.5 and 2.0. Subsequent adjustments of the protein were made until the  $R_{\text{factor}}$  dropped below 0.25. Water molecules were added by SHELXWAT™ and by eye, using  $\sigma$ -weighted 2Fo-Fc maps. Once there was approximately 1 water molecule for every protein residue (~800 in total) the substrate was modelled into the binding site. The 4-hydroxydiphenylmethane molecule was modelled into the 2Fo-Fc difference map by initially placing the oxygen of the hydroxyl group above the heme, as indicated by the binding data. The remainder of the substrate was modelled into the residual density. Structure was refined until no significant difference in the  $\sigma$  weighted 2Fo-Fc difference map was observed at the a  $3\sigma$  level. The 4-hydroxydiphenylmethane orientation with the hydroxyl group away from the heme was also modelled.

### 2.12.8.3 General Technique.

The other refinements were essentially carried out in the same way except for the differences noted below. Structurally conserved waters, 1-500, were refined in the new structures before the SHELXWAT™ refinement and the waters with high temperature factors were removed.

Since there are two molecules per asymmetric unit they were refined with non-crystallographic restraints (the NCSY parameter set at 0.1). This was only performed once the structure was refined to a  $R_{\text{factor}}$  of less than 0.22. The NCSY restraint was removed once refinement was complete and any change to the  $R_{\text{factor}}$  and  $R_{\text{free}}$  was observed.

### 2.12.8.4 Camphor Structure.

Camphor was initially placed in the same orientation as the wild-type structure and then allowed to refine freely. The alternate conformations of camphor were investigated by placing the carbon/hydrogen bond known to be hydroxylated near the heme and fitting the rest of the camphor molecule into the remaining density. A water molecule was modelled in the binding site, ligated to the iron, in addition to the camphor molecule with both fixed and free occupancy. Starting distances and bond restraints for both the iron-sulfur and the iron-aqua bonds were altered and the effect on the refined structure was noted.

### 2.12.8.5 Phenylcyclohexane Structure.

Phenylcyclohexane was modelled into the density map by placing the phenyl ring in the same orientation as the non-hydroxylated phenyl ring from the 4-hydroxydiphenylmethane structure; alternate conformations with the aliphatic ring in this position and the other geometric isomers were additionally modelled.

#### 2.12.8.6 Diphenylmethane Structure.

Diphenylmethane was initially modelled into the density based upon the 4-hydroxydiphenylmethane structure and refined. Diphenylmethane was modelled with variable occupancies and in cases where the occupancy was ~50 % water molecules were additionally modelled in the binding site.

#### 2.12.8.7 Bicyclohexyl Structure.

Bicyclohexyl was fitted into the 2Fo-Fc density map initially with the separate refinement of the two cyclohexane rings. The first ring was fitted into the density in both chair and the boat conformation. The other ring was modelled after optimization of their first ring, again in both chair and boat conformations.

#### 2.12.8.8 *N*-Cbz-2,6-dimethylpiperidine.

The low energy orientations of the *N*-Cbz-2,6-dimethylpiperidine was additionally modelled in Macromodel. One energy minimized structure was entered and a Monte Carlo minimization run was performed (in collaboration with Suzanne Aitken, University of Edinburgh); the lowest energy forms were then noted.

For this structure the averaged  $\alpha$ -weighted difference map was used to model the substrate. After initial refinement the average map was generated in the ccp4 modules; f2mtz, mapman, 6d\_ave and 6d\_mapman, with rotational matrix from O. The map was generated in the second asymmetric unit (residues 510-914). The *N*-Cbz-2,6-dimethylpiperidine was modelled in two halves. The first was to model the phenyl ring in the orientation occupied by the phenyl ring of all the other substrates. The piperidine ring was orientated in the density and refined separately. The link between the two molecules was then modelled into the remaining density and the whole molecule was refined. After completion of the model building the ligand was translated to the other asymmetric unit, by overlapping the two units with ccp4 module 'match'. Refinement of the structure was then carried out as above.

The final orientation was, with the protein coordinates removed, subjected to energy minimization in InsightII with a discover force field and the minimized form compared to the crystal structure.

#### 2.12.8.9 Structural Verification.

The validity of the structures was verified initially by following the  $R_{\text{free}}$  during the refinement and not allowing it to rise from its lowest refined level. Final structures were verified by processing the structure in Procheck and examining the Ramachandran plots and the analysis of the individual residues. After any adjustments the final structure was again checked with Procheck.

#### 2.12.8.10 Comparison of Structures.

Comparison of the various ligand bound structures and the sub-units of the asymmetric units was performed with the ccp4 suite of programs. For the asymmetric sub-units, the second molecule (510-914) was overlapped onto the first, with the residue numbers being altered with 'transform'. 'Match' was used to overlay the structures on top of each other and 'Compar' was used to compare the data between the different structures. In every case the structures only contained the heteroatoms of the heme group and no water molecules. Substrates were only included for analysis when analyzing the overlaid substrate positions. For the wild type structure the 4cp4 (camphor bound) structure was used.<sup>10</sup> The heme group and camphor molecule were altered manually with a text editor to conform to the numbering convention used in the Y96A structures (i.e. 451 and 461 respectively).

## 2.13 Molecular Modelling.

### 2.13.1 Initial Modelling.

Without the available P-450cam Y96A mutant structure modelling was initially performed on a model structure generated by manipulation of the camphor bound wild-type structure (2cpp4).<sup>11</sup> This structure was generated by deletion of the phenyl and hydroxyl groups of the tyrosine 96 residues (atoms 761-768) to generate alanine. It was believed that this model was a good approximation of the actual mutant structure.<sup>8</sup>

Substrates were generated using InsightII™. To obtain a realistic conformation for the substrates, the general structure was entered and then minimized with a Discover force field. Multiple low energy ligand conformations were explored by manual manipulation of the structure to an alternate conformation followed by energy minimization.

Substrates were modelled into the binding site by placement of the phenyl ring of the substrate in the location previously occupied by the tyrosine phenyl ring. The rest of the substrate was then placed in the binding site above the phenyl ring. The substrates were initially moved in the binding site, to maintain the phenyl ring orientation and to minimize any Van der Waals interactions between the protein and the substrate. This meant maintaining a greater than 3 Å distance for all carbon-carbon interactions; if this was not possible small torsional adjustments were made to the molecule. Once a satisfactory orientation was achieved the nearest carbon hydrogen bonds to the heme iron were measured.

### 2.13.2 Remodelling of Substrates.

Once the crystal structure of the Y96A mutant was solved this was used for modelling studies. The studies performed with the model were repeated with the 4-hydroxydiphenyl methane structure. The unhydroxylated phenyl ring position from the



4-hydroxydiphenylmethane structure was used as a starting point for placement of the substrate aromatic rings and the substrate was then modelled as before.

### 2.13.3 Use of DOCK.

The DOCK suite of programs has a series of steps involved in substrate modelling (figure 2.2):<sup>12-17</sup>

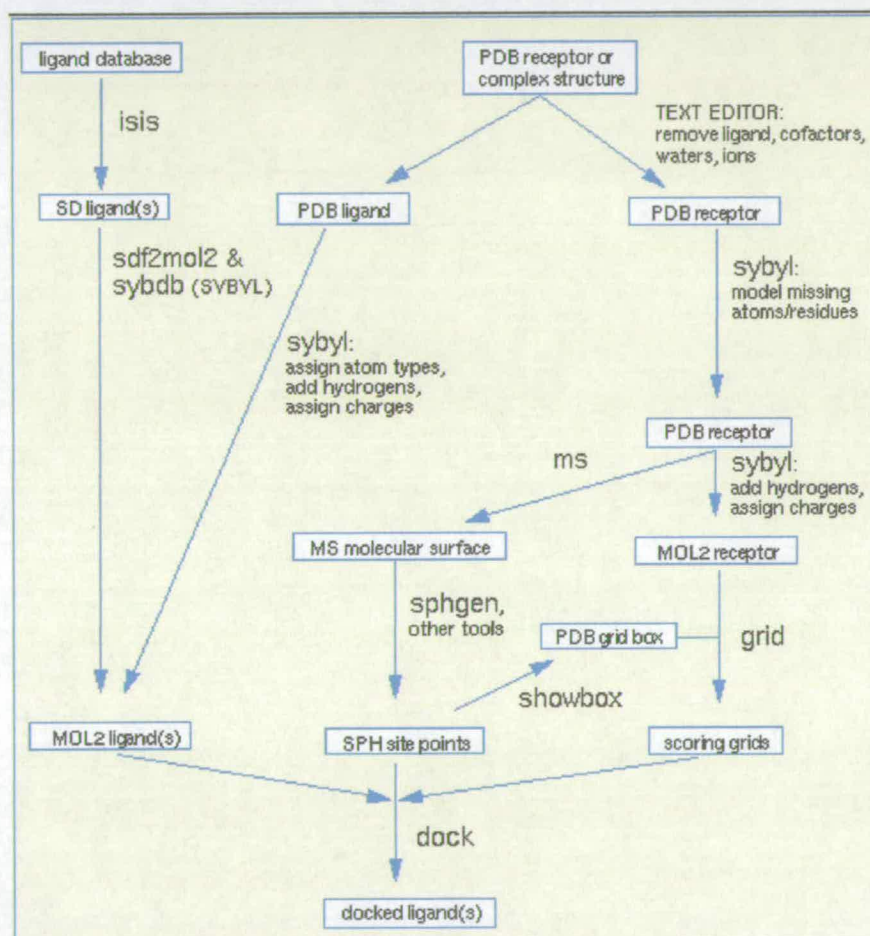


Figure 2.2: Overview of DOCK suite of programs

#### 2.13.3.1 Ligand Preparation.

Ligands were built in InsightII and energy minimized with a Discover forcefield using steepest decent, conjugate decent and VA09A minimization routines



with 1000 iterations of minimization. These ligand files then had charges added using SYBYL utilizing the Gasteiger Marsili charge assignment method. Ligand files were output as mol2 format files (mol2 format is a file format compatible with the SYBYL suite of modelling packages which contains information on the charges and atom types as well as atom position).

#### 2.13.3.2 Site Characterization.

Of the Y96A crystal structures available the camphor and 4-hydroxydiphenylmethane bound structures have the best resolution. The camphor structure was the best choice for a DOCK template since it did not have the change to Thr 252 orientation (see chapter 7). However, the Phe 98 orientation is slightly different in the camphor structure than in structures with aromatically substituted substrates. Since many of the substrates modelled would be larger than camphor, it was not thought appropriate to have this Phe 98 ring orientation. The solution was to use the camphor bound structure with orientation of the Phe 98 ring from the 4-hydroxydiphenylmethane bound structure. This was accomplished by overlapping the 4-hydroxydiphenylmethane bound structure with the camphor bound structure, using ccp4 program 'match', as before. The coordinates for the Phe 98 residue were then manually substituted into the pdb file for camphor bound structure, using a text editor; thus generating the enzyme binding site with the 'normal' Thr 252 orientation and the Phe 98 orientation consistent with the presence of a phenyl substituted substrate.

The substrate and water coordinates were removed manually. To minimize the computational time only the binding site was modelled. This was accomplished by using the 'get\_near\_res' program to define all residues which have any atom within 8 Å of camphor carbon C1. This included residues: Pro 86, Phe 87, Ala 96, Asp 97, Phe 98, Thr 101, Ser 102, Thr 181, Met 184, Thr 185, Phe 193, Gly 243, Leu 244, Val 247, Gly 248, Asp 251, Thr 252, Leu 294, Val 295, Ala 296, Asp 297, Gln 322, Gly 394, Ile 395, Val 396 and hem 451.

The molecular surface was created with 'autoMS' using a surface density of 1 dot/Å<sup>2</sup> and a probe radius of 1.2 Å. The sphere centres were created in 'sphgen',

using the surface file generated by 'autoMS'. Parameters were set with the maximum and minimum radii at 5.0 Å and 1.2 Å respectively.

#### 2.13.3.3 Generating Spheres.

Generated sphere clusters were converted to pdb format using 'showsphere' and examined using InsightII. Due to the limitations of sphgen, additional spheres were added manually. In order to determine possible ligand positions, which may have been overlooked by sphgen, a ligand file was created containing all the previous P-450cam co-crystallized substrate orientations (1akd, S-camphor; 1cp4, phenyl radical; 1noo, 5-exo-hydroxycamphor; 1pha, Pfizer + inhibitor; 1phb, Pfizer - inhibitor; 1phd and 1phe, 2-phelylimidazole; 1phf, 4-phenylimidazole; 1phg, metyrapone; 4cp4, camphor; 4cpp, adamantane; 5cpp, adamantanone; 6cpp, camphane; 7cpp, norcamphor; 8cpp, thiocamphor)<sup>11,18-25</sup> with the six Y96A structures, overlapped with ccp4's 'match' as before. The sphere centres generated with sphgen were then examined to see if any previous ligand atom locations were not characterized by a sphere (no sphere centre within 0.5 Å). Spheres were then added manually, with a text editor, to the sphgen sphere file to cover these missing sites.

#### 2.13.3.4 Creating Scoring Grids.

The binding site file used to create the spheres was used to generate the grid with a number of modifications. 'Grid' uses mol2 format input files so the binding site pdb file had to be converted. Hydrogens were added with InsightII and charges with SYBYL. Charge parameters for the amino acids were based on the SYBYL bipolymer database using Kollman charges. The heme group was defined by Gasteiger Marsili charge calculation method, except for the iron, which cannot be defined by SYBYL. The iron was redefined as a planar carbon atom by attaching it via two double bonds to NA and NC heme nitro groups. The charge on the carbon was manually defined as + 0.9 based on the *ab initio* based charge calculation for unbound

heme.<sup>26</sup> In order to check the possibility of using oxygen species bound to the iron, as a means of better characterizing the binding site, an oxygen was added to the heme group at a distance of 1.3 Å from the heme. The charge on the oxygen was defined as -0.6, based on the *ab initio* calculated value.<sup>26</sup> A separate binding site file was then created for this oxygen bound site. Files were saved as mol2 format.

To define the grid calculation area, a box was created with 'showbox' to include all points within 5 Å of the bound 4-hydroxydiphenylmethane orientation. Grids were calculated for both energy and contact scoring, with grid spacing of 0.1 Å and distance cut-off at 5.0 Å, while other values were as default.

#### 2.13.3.5 Running DOCK.

The final DOCK parameter file used is presented in appendix 2. Notable differences from the default values are: the maximum orientations of the ligand and the number of orientations to be scored, which are significantly higher here than the defaults of 500 and 100 respectively. The lowest energy scored ligand orientation was redocked in DOCK. Output files included the lowest scored ligand orientations for both the contact and energy scored grid, plus the redocked values after re-inputting the calculated energy conformation and the DOCK output file containing the details of the run. For an example see appendix 3.

Docking was performed initially on camphor until suitable file parameters were established. The binding site grid used for calculations was the one without bound heme-oxygen. The ligand orientations were then viewed with InsightII.

## 2.14 References.

- 1) Sambrook, J.; Fritsch, E.; Maniatis, T. *Molecular Cloning*, 2nd ed.; Cold Spring Harbour Laboratory Press; 1989.
- 2) Bailey, S. *Acta Crystallographica Section D-Biological Crystallography* **1994**, *50*, 760-763.
- 3) Otwinowski, Z.; Minor, W. *Methods in Enzymology* **1997**, *276*, 307-326.
- 4) developed by T.A Jones; Department of Molecular Biology, B., Uppsala University, Sweden & Morten Kjeldgaard; Department of Chemistry, Aarhus University, Denmark.
- 5) developed by George M Sheldrick, G. U., Germany, .
- 6) Staines, A.G. *Monooxidations Using a Cytochrome P450cam Mutant*; MSc Thesis; University of Edinburgh, 1996.
- 7) O'Keeffe, D. H.; Ebel, R. E.; Peterson, J. A. *Methods in Enzymology* **1978**, *52*, 151-157.
- 8) Wong, L-L., Unpublished Results.
- 9) Weber, P. C. *Methods in Enzymology* **1997**, *276*, 13-22.
- 10) Vidakovic, M.; Sligar, S. G.; Li, H. Y.; Poulos, T. L. *Biochemistry* **1998**, *37*, 9211-9219.
- 11) Poulos, T. L.; Finzel, B. C.; Gunsalus, I. C.; Wagner, G. C.; Kraut, J. *Journal of Biological Chemistry* **1985**, *260*, 16122-16130.
- 12) Shoichet, B. K.; Kuntz, I. D. *Protein Engineering* **1993**, *6*, 723-732.
- 13) Shoichet, B. K.; Bodian, D. L.; Kuntz, I. D. *Journal of Computational Chemistry* **1992**, *13*, 380-397.
- 14) Shoichet, B. K.; Kuntz, I. D. *Journal of Molecular Biology* **1991**, *221*, 327-346.
- 15) Kuntz, I. D. *Science* **1992**, *257*, 1078-1082.
- 16) Kuntz, I. D.; Blaney, J. M.; Oatley, S. J.; Langridge, R.; Ferrin, T. E. *Journal of Molecular Biology* **1982**, *161*, 269-288.
- 17) Kuntz, I. *DOCK Version 4.0*; Regents of the University of California; 1998.
- 18) Li, H. Y.; Narasimhulu, S.; Havran, L. M.; Winkler, J. D.; Poulos, T. L. *Journal of the American Chemical Society* **1995**, *117*, 6297-6299.

- 19) Poulos, T. L.; Howard, A. J. *Biochemistry* **1987**, *26*, 8165-8174.
- 20) Poulos, T. L.; Finzel, B. C.; Howard, A. J. *Journal of Molecular Biology* **1987**, *195*, 687-700.
- 21) Raag, R.; Poulos, T. L. *Biochemistry* **1989**, *28*, 7586-7592.
- 22) Raag, R.; Swanson, B. A.; Poulos, T. L.; DeMontellano, P. R. O. *Biochemistry* **1990**, *29*, 8119-8126.
- 23) Raag, R.; Poulos, T. L. *Biochemistry* **1991**, *30*, 2674-2684.
- 24) Raag, R.; Li, H. Y.; Jones, B. C.; Poulos, T. L. *Biochemistry* **1993**, *32*, 4571-4578.
- 25) Schlichting, I.; Jung, C.; Schulze, H. *FEBS Letters* **1997**, *415*, 253-257.
- 26) Zakharieva, O.; Grodzicki, M.; Trautwein, A. X.; Veeger, C.; Rietjens, I. *Journal of Biological Inorganic Chemistry* **1996**, *1*, 192-204.

## 3 Protein Expression and Purification.

---

### 3.1 Cytochrome P-450cam.

#### 3.1.1 Expression of Cytochrome P-450cam in *E. coli*.

The *Pseudomonas* bacterium has a well documented nutritional diversity.<sup>1</sup> In 1959, the isolation from soil of a *Pseudomonad* (*P. putida*), known to be able to metabolize camphor as its sole carbon source, was achieved.<sup>2</sup> This led to an investigation of the enzymes involved in camphor metabolism. In 1965 the pathway for the camphor metabolism was established (figure 1.5) thus leading to the identification and isolation of the first enzyme of the camphor metabolic pathway, cytochrome P-450cam.<sup>3</sup>

Large scale expression and purification of cytochrome P-450cam was attempted in the early 1970s but a typical 130 L fermentation yielded only 1 mg of protein.<sup>4</sup> The genes encoding the cytochrome P-450cam and many of the other enzymes involved in the camphor metabolic pathway were found to reside on a 230 kbp plasmid <sup>5</sup> known as the 'CAM' plasmid (PpG1). This plasmid encodes four proteins camA, B, C and D under one regulator, camR. The operon was arranged in order DCAB where camD is putidaredoxin, camC is putidaredoxin reductase, camA is cytochrome P-450cam and camB is 5-exohydroxycamphor dehydrogenase. In 1982,

with the advent of modern cloning techniques, the P-450cam gene was cloned and overexpressed in *E. coli* (using the broad range pKT240 vector,<sup>6</sup> creating plasmid pKG201). For the first time, expression of the P-450cam above the normal levels expressed with the CAM plasmid was achieved. This expression was still poor due to the accidental removal of the cytochrome P-450cam promoter in the cloning<sup>7</sup> and in 1986, with the elucidation of the sequence for P-450cam<sup>8</sup> the gene was subcloned into a high copy number pUC vector.<sup>9</sup> This allowed over-expression of the P-450cam under a Lac promoter, to a level of 10 % of the total soluble protein content.<sup>8</sup>

Although expression of cytochrome P-450cam has been optimized in *P. putida*, this expression system has a number of drawbacks compared to expression in *E. coli*:<sup>10</sup> i) the growth of *E. coli* is about 5-fold faster than *P. putida*, ii) the yields of cytochrome P-450cam are higher in *E. coli*; iii) *E. coli* only express P-450cam without the related co-proteins whereas the *P. putida* also produce the two co-proteins putidaredoxin (Pd) and putidaredoxin reductase (PdR), which results in turnover of camphor and prevents the accurate study of the reaction kinetics. Importantly mutagenesis of a *P. putida* plasmid is not possible as no *P. putida* host vector currently exists. Due to the advantages of the *E. coli* system, the cytochrome P-450cam gene was incorporated into a heterologous expression system, constructed using plasmid pAW7292 in *E. coli* cell line JM109.<sup>11</sup>

The induction was controlled by a temperature rise from 30 °C to 37 °C. This plasmid also included antibiotic resistance genes for chloramphenicol.<sup>11</sup> It is important to note that the Y96A mutant and the 'wild-type' protein studied here contain the C334A mutation (to remove the surface cysteine), though it has been shown that this mutation has no effect on catalytic activity.<sup>12,13</sup>

### 3.1.2 Purification of P-450cam.

Many purification methodologies for P-450cam from bacterial systems have been established,<sup>14</sup> consisting of post-lysis purification with a series of anion exchange columns. For over-expressed proteins the most common purification technique used is

the method of Sligar,<sup>15</sup> but this had to be adapted due to unsatisfactory preliminary results.<sup>10</sup> The general protocol used for the production of P-450cam was then adapted from one produced by Flitsch and Wong.<sup>16</sup> The following purification protocols were initially applied to the Y96A mutant purification but wild type protocols only differ in the OD at which the protein is assayed.

### 3.1.2.1 Protein Quantification and Purity Measurements.

The heme groups of P-450 monooxygenases have a characteristic absorbance, known as the Soret band, at about 400 nm caused by a  $\pi$ - $\pi$  (ligand to ligand) charge transfer. As discussed previously, (chapter 1), the spin state of the iron centre alters upon substrate binding, increasing the  $\pi$ - $\pi$  transition energy, thus changing the energy (i.e. wavelength) of the Soret band.<sup>15</sup> In the case of the P-450cam, the Soret band shifts from a maximum absorbance ( $\lambda_{max}$ ) absorbance of 417 nm (water bound) to an absorbance of 390 nm (no water bound), figure 3.1. This provides a method of determining both protein concentration and purity and additionally a means of detecting substrate binding, all by measuring the change in  $\lambda_{max}$ , spectroscopically.

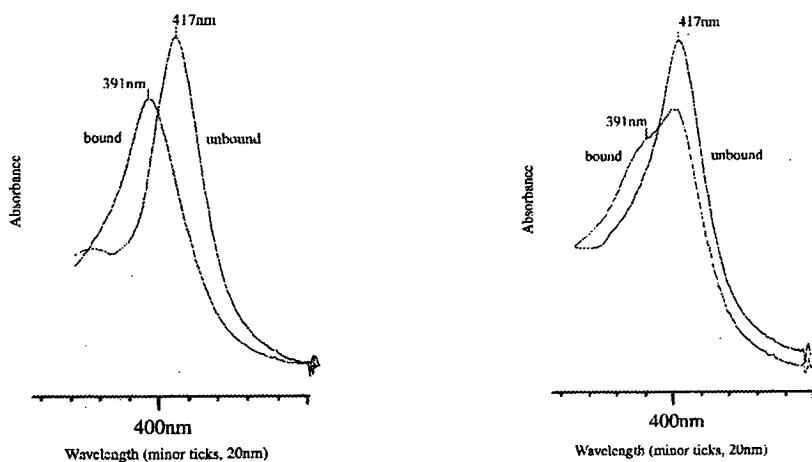


Figure 3.1: Soret band shift upon binding of 1 mM camphor to wild type (i) and Y96A mutant (ii) P-450cam.



For protein concentration determination, the value of the Soret band maximum is used. For almost all stages of the purification the protein is stabilized with camphor, therefore the protein is in its substrate bound state. For wild type the maximum of the Soret band is therefore at 390 nm. For the Y96A mutant the maximum absorbance is at 417 nm, but at lower energy than the unbound maximum (i.e. lower extinction coefficient) because the camphor bound Y96A mutant has only 55 % of the heme in the low-spin form, figure 3.1ii.

Table 3.1: Extinction Coefficients of wild type P-450cam.

wavelength	camphor free	camphor bound
280nm	68.3 mM <sup>-1</sup> cm <sup>-1</sup> <sup>17</sup>	63.3 mM <sup>-1</sup> cm <sup>-1</sup> <sup>17</sup>
390nm	44 mM <sup>-1</sup> cm <sup>-1</sup> <sup>18</sup>	102 mM <sup>-1</sup> cm <sup>-1</sup> <sup>17</sup>
417nm	115 mM <sup>-1</sup> cm <sup>-1</sup> <sup>17</sup>	56 mM <sup>-1</sup> cm <sup>-1</sup> <sup>18</sup>

Table 3.2: Extinction Coefficients of mutant Y96A, P-450cam. <sup>18</sup>

wavelength	camphor free	camphor bound
280nm	68.3 mM <sup>-1</sup> cm <sup>-1</sup>	63 mM <sup>-1</sup> cm <sup>-1</sup>
390nm	44 mM <sup>-1</sup> cm <sup>-1</sup>	70 mM <sup>-1</sup> cm <sup>-1</sup>
417nm	115 mM <sup>-1</sup> cm <sup>-1</sup>	88 mM <sup>-1</sup> cm <sup>-1</sup>

From the extinction coefficients tables 3.1 and 3.2, the ratio of 417:280 nm for pure camphor bound protein should be 1.2-1.3 for the mutant protein and 1.5-1.6 for wild type protein.

### 3.1.2.2 Extraction of Cytochrome P-450cam from *E. coli*.

Cells were harvested by centrifugation at 4 °C and stored at -20 °C. The P-450cam was released by cell lysis, through sonication at 4 °C. After sonication and two centrifugation cycles the cell pellet was an 'off-white' colour, indicating complete cell lysis.

## 3.1.2.3 Purification by Anion Exchange Chromatography.

The DEAE sepharose column is a crude first protein purification step that equilibrates the cytochrome P-450cam into phosphate buffer containing camphor and potassium (40 mM), both of which stabilize the protein.<sup>15</sup> The movement of the protein, due to elution with a salt gradient, on the DEAE column was easily monitored by the movement of the coloured band on the column. The coloured DEAE column fractions were analyzed by UV/Vis spectroscopy at 417 nm, figure 3.2.

Fractions collected (~200 ml) were all above an OD at 417 nm of 0.1. Other fractions below OD 0.1, were not selected as their purity is not sufficiently high. The purification was also confirmed by SDS PAGE, figure 3.5.

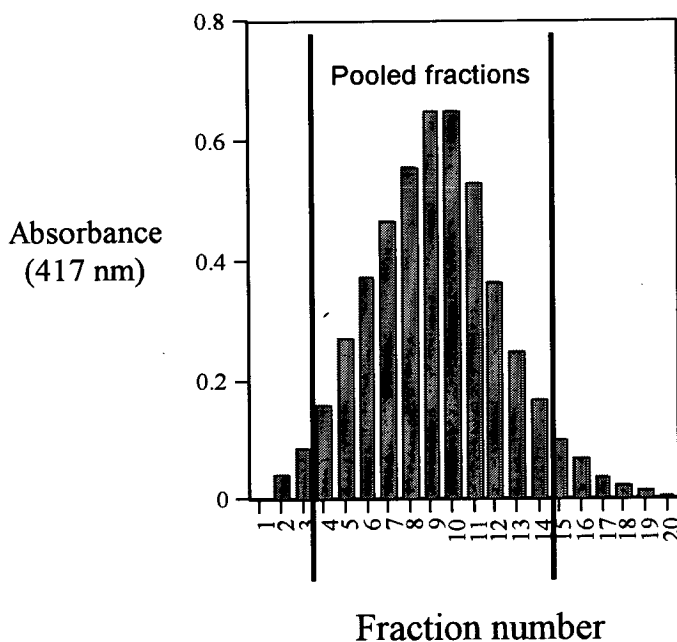


Figure 3.2: Selected DEAE column fractions, from an initial cell free extract load of 150 ml (only the coloured fractions were selected and spectroscopically analyzed).

### 3.1.2.4 Desalting and Concentration of DEAE Fractions.

The DEAE fractions were desalted before the HPLC purification otherwise the protein would not bind to the SourceQ™ anion exchange resin. The fractions also needed to be concentrated as the large fraction volume would prevent efficient binding to the HPLC column.

Two successive concentration and dilution steps in a 300 ml concentrator (30 kDa cut-off membrane) desalted the solution and concentrated it by a factor of ten, additionally protein contaminants below 25 kDa were removed, as confirmed by SDS PAGE, figure 3.5.

### 3.1.2.5 Anion Exchange Chromatography Using HPLC with SourceQ™ Media.

The final step in the protein purification consists of high resolution Source15Q™ anion exchange chromatography using a HPLC system. A significant amount of impurities do not bind to the SourceQ media and are washed off in the loading stage, figure 3.3. The protein elutes from the column at ~9.5 minutes, at a salt concentration of 7-12 % 1 M KCl.

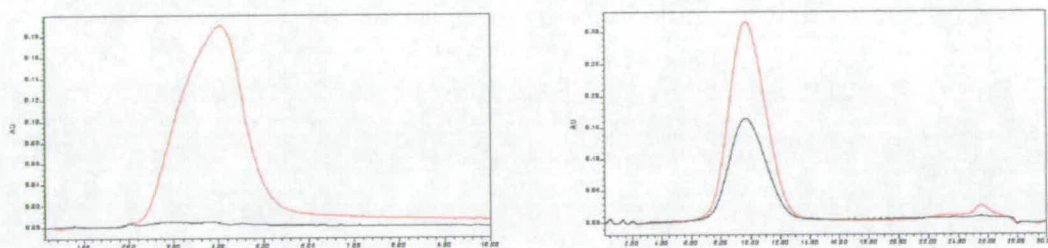


Figure 3.3: HPLC trace of Y96A protein sample, run at 10ml/min with a KCl gradient, monitored by absorbance at 280 nm and 417 nm. Left trace (red 280 nm, black 417 nm) is the sample loading stage (i.e. flow-through) and the right trace (red 417 nm, black 280 nm) is trace from the gradient elution.

The purity of the fraction can be directly ascertained by spectroscopic assay; the ratio of 280 nm : 417 nm was 1:1.2 for Y96A mutant for fractions of almost pure P-450cam. Protein purity was confirmed by SDS PAGE, figure 3.5.

### 3.1.2.6 Removal of Camphor and Glycerol by Gel Filtration.

The protein is stored in camphor and glycerol which was removed before any spectroscopic assays could be performed. A Sephadex G25 gel filtration column removed these components without excessive dilution (~2x dilution factor). As in the case of the DEAE column, the elution of the protein could be observed by the movement of the coloured band on the column. Fraction purity and concentration was measured by UV/Vis spectroscopy.

### 3.1.2.7 Production of Protein of Sufficient Quality for Crystallography.

For crystallography studies, the level of purity of the protein is vital. An extra HPLC step was performed using the hydrophobic interaction media phenyl sepharose. This removed any remaining protein contaminants (at high salt concentration), figure 3.4 and any denatured P-450cam.

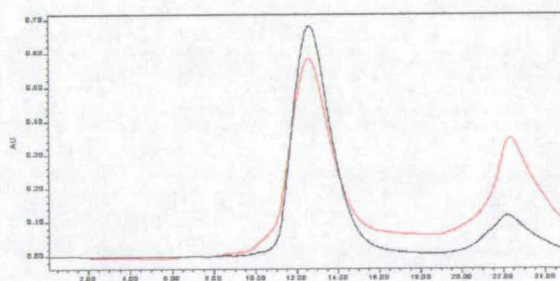


Figure 3.4: HPLC trace of Y96A protein sample run on the phenyl sepharose hydrophobic interaction column with a decreasing ammonium sulfate gradient, monitoring at absorbance at 280 nm (red) and 417 nm (black).

## 3.1.2.8 Summary.

Purification of cytochrome P-450cam was achieved in five steps with an overall yield of about 45 %, table 3.3.

Table 3.3: Typical protein purification data of Y96A mutant.

	mass (mg) Y96A <sup>(a)</sup>	ratio 417:280 nm <sup>(b)</sup>
Pellet weight	46.3 g	-
Cell free extract	(c)	(c)
DEAE fractions	130	0.35
Concentration and desalting	125	0.6
HPLC fraction <sup>(d)</sup>	100	1.2
Crystallography grade protein <sup>(e)</sup>	90	1.3
G-25 fractions <sup>(f)</sup>	80	1.7 <sup>(g)</sup>

a) Protein concentration determined by spectroscopic assay at 417 nm .

b) The ratio of 417 nm:280 nm is used as a measure of purity .

c) It is impossible to measure the cell free extract concentration accurately due to contamination from DNA and other E. coli impurities.

d) Anion exchange HPLC was performed on 1/5 of the total sample, per run.

e) Protein after the phenyl sepharose column performed on 1/5 of the total sample, per run.

f) Samples loaded on the Sephadex G-25 column were about 1/8 of the total sample.

g) Ratio is substrate free enzyme.



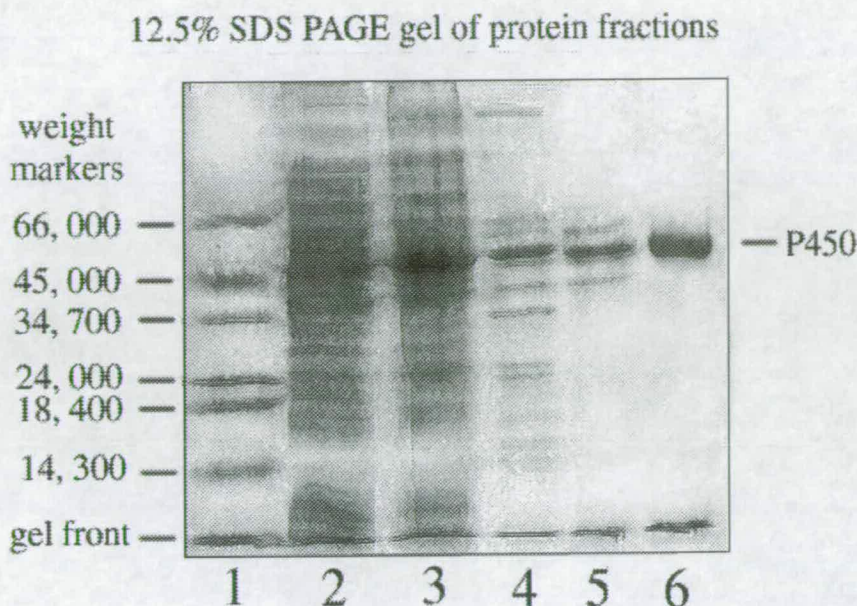


Figure 3.5: SDS PAGE gel of protein fractions; lanes- 1: Molecular weight markers, 2: Pre-induced cell suspension, 3: Induced cell suspension, 4: DEAE column fraction, 5: Concentration and desalting step, 6: HPLC fraction.

After the HPLC step, the protein was pure enough for binding studies, as judged by SDS PAGE, figure 3.5. The only protein to be further purified was protein for the crystallography experiments.

## 3.2 Putidaredoxin.

### 3.2.1 Expression of Putidaredoxin in *E. coli*.

The gene for putidaredoxin was discovered to be associated with the genes encoding P-450cam and putidaredoxin reductase on the CAM plasmid.<sup>19</sup> Part of the putidaredoxin sequence was identified in 1986<sup>20</sup> with the sequencing of P-450cam, since the Pd gene lies downstream of P-450cam. In 1989 the Pd gene was sequenced by Koga<sup>21</sup> and Peterson,<sup>22</sup> working independently. The Pd gene was then engineered into the *E. coli* plasmid pIBI24 (derived from pEMBL) and expression

was placed under the control of the *lacZ* promoter. This gave yields of protein of 4.8 mg/g of wet cells in *E. coli* strain TB1.

### 3.2.2 Purification of Pd.

Putidaredoxin was first purified in 1977 by Mock, in the laboratory of Peterson, and since then many purification methodologies have been established.<sup>23,24</sup> For over-expressed proteins the purification is essentially the same as that developed by Sligar<sup>17</sup> which consists of an ammonium sulfate cut followed by gel filtration and anion exchange chromatography.

Attempts to reproduce this technique proved unsuccessful due to high protein loss during the ammonium sulfate cut (~50 % loss). It was therefore decided to develop new protein purification methods.

#### 3.2.2.1 Protein Quantification and Purity Measurements.

The 2Fe2S core of the putidaredoxin gives rise to a number of electronic transitions that are detectable by UV/Vis spectroscopy. For protein concentration determination, the peak at 405 nm is used because for impure protein solution, protein absorbance at 280 nm interferes with the 325 nm peak.

Table 3.4: Extinction Coefficients of Pd.

wavelength	extinction coefficients <sup>(a)</sup>
280 nm	22.8 mM <sup>-1</sup> cm <sup>-1</sup>
325 nm	15.6 mM <sup>-1</sup> cm <sup>-1</sup>
405 nm	11.1 mM <sup>-1</sup> cm <sup>-1</sup>
455 nm	10.4 mM <sup>-1</sup> cm <sup>-1</sup>

a) Extinction coefficients from published data.<sup>17</sup>

From the extinction coefficients, table 3.4, the ratio of 405:280 nm for pure protein should be 0.49.

### 3.2.2.2 Extraction of Pd from *E. coli*.

The protein is unstable at 37 °C, the cells were harvested by centrifugation at 4 °C and stored at -20 °C. The Pd was released by cell lysis, through sonication at 4 °C. After sonication and two centrifugation cycles the cell pellet was an 'off-white' colour, indicating complete cell lysis.

### 3.2.2.3 Purification by Anion Exchange Chromatography.

The DEAE sepharose column is a crude first protein purification step that equilibrates the Pd into Tris buffer, which stabilizes the protein.<sup>15</sup> The movement of the protein, due to elution with a salt gradient, on the DEAE column was easily monitored by the shift of the coloured band on the column. The coloured DEAE column fractions eluted were analyzed by UV/Vis spectroscopy at 405 nm (figure 3.2).

Fractions collected (~200ml) were all above an OD at 405 nm of 0.1. Other fractions below OD 0.1, were not selected as their purity is not sufficiently high. The purification was also confirmed by SDS PAGE (figure 3.6)



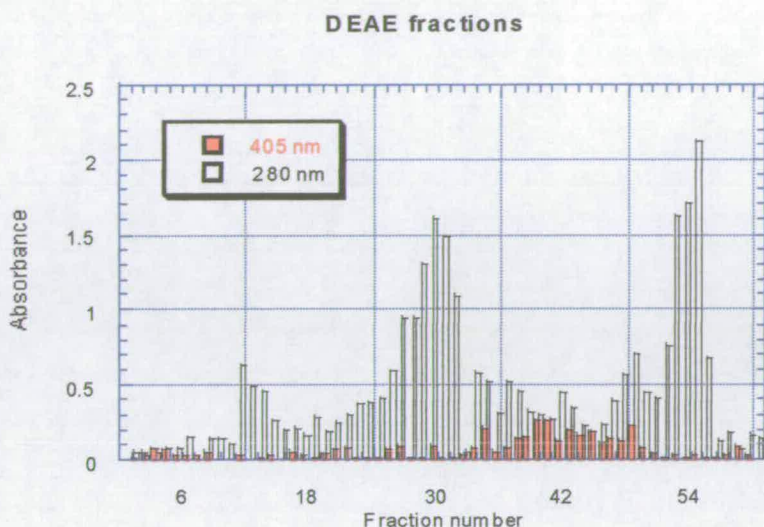


Figure 3.6: Selected DEAE column fractions, from an initial cell free extract load of 150 ml (only the coloured fractions were selected and spectroscopically analyzed). Fractions were collected in 10 ml volumes.

#### 3.2.2.4 Desalting and Concentration of DEAE Fractions.

The DEAE fractions were desalted before the HPLC purification otherwise the protein would not bind to the SourceQ™ anion exchange resin. Desalting and concentration through use of a concentrator cell was not possible for Pd as it was for P-450 because with the use of a smaller molecular weight cut-off membrane the concentration took too long (>2 days), therefore it was decided to use dialysis to remove the salt from the DEAE column. Concentration was not necessary as the Pd eluted from the DEAE column more efficiently than the P-450cam. The resulting solution was sufficiently concentrated to pump directly on the SourceQ column. Dialysis also removed any small molecular weight proteins (<8 kDa).

#### 3.2.2.5 Anion Exchange Chromatography Using HPLC with SourceQ™ Media.

The final step in the protein purification consists of high resolution Source15Q™ anion exchange chromatography using an HPLC system. A significant

amount of impurities did not bind to the SourceQ media and were washed off in the loading stage (figure 3.3) The protein elutes from the column at ~19 minutes, at a salt concentration of 50 % 1 M KCl (figure 3.7).

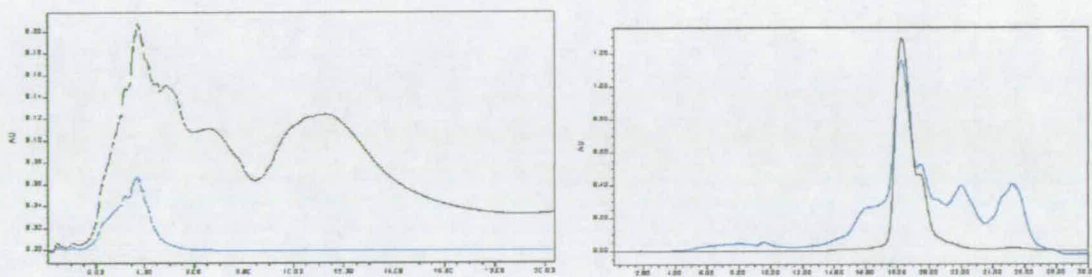


Figure 3.7: HPLC trace of Y96A protein sample, run at 10ml/min with a KCl gradient, monitoring at absorbance at 280 nm and 417 nm. Left trace (black 280 nm) is the loading of the sample and the right trace (black 405 nm) is the gradient elution. The absorbance for 405 nm (i.e. black) is normalized on the graph by a factor of 2.

From the dual wavelength output spectra the purity of the fraction can be directly ascertained without the need for a separate spectroscopic assay. The ratio of 280 nm: 405 nm was 1:0.4. Protein purity was confirmed by SDS PAGE.

### 3.2.2.6 Final Purification Step of Pd.

An extra HPLC step was performed using the hydrophobic interaction media phenyl sepharose. This removed any residual protein contaminants at high salt concentration along with any denatured Pd.

## 3.2.2.7 Summary.

Purification of cytochrome Pd was achieved in five steps with an overall yield of about 45% (table 3.5).

Table 3.5: Typical protein purification data of Y96A mutant.

	mass (mg) Pd <sup>(a)</sup>	ratio 405:280 nm <sup>(b)</sup>
Pellet weight	25 g	-
Cell free extract	84	0.17
DEAE fractions	60	0.19
Dialysis	55	0.27
Anion exchange HPLC <sup>(c)</sup>	50	0.45
Hydrophobic interaction HPLC <sup>(d)</sup>	45	0.50

a) Protein concentration determined by spectroscopic assay at 405 nm .

b) The ratio of 405 nm:280 nm is a used as a measure of purity.

c) Anion exchange HPLC was performed on 1/10 of the total sample, per run.

d) Protein after the phenyl sepharose column performed on 1/8 of the total sample, per run.

### 3.3 Putidaredoxin Reductase.

#### 3.3.1 PdR Quantification:

The flavin group in the PdR give the protein a characteristic absorbance spectrum which can be used to quantitate the protein. From the previously published extinction coefficients, the protein absorbance at 454 nm is used to measure protein concentration, table 3.6.<sup>17</sup>

Table 3.6: Extinction coefficients of PdR.

wavelength	extinction coefficients
280 nm	72 mM <sup>-1</sup> cm <sup>-1</sup>
378 nm	9.7 mM <sup>-1</sup> cm <sup>-1</sup>
454 nm	10 mM <sup>-1</sup> cm <sup>-1</sup>
480 nm	8.5 mM <sup>-1</sup> cm <sup>-1</sup>

### 3.4 Summary.

The purification of both the wild-type cytochrome P-450cam, the Y96A mutant and the putidaredoxin was thus improved to give large quantities of pure protein. The purity of the protein used for crystallographic studies was of sufficient quality to give crystals that diffracted well ( $< 2 \text{ \AA}$ ), (chapter 5). Enough pure putidaredoxin reductase was provided by Dr. Luet Lok Wong that no expression and purification was necessary.

### 3.5 References

- 1) Stainer, R. Y.; Palleroni, N. J.; Doudoroff, M. *Journal of Genetics and Microbiology* **1966**, *43*, 159-271.
- 2) Bradshaw, W. H.; Conrad, H. E.; Corey, J. E.; Gunsalus, I. C.; Lednicer, D. *Journal of the American Chemical Society* **1959**, *81*, 5007.
- 3) Henggaard, J.; Gunsalus, I. C. *Journal Of Biological Chemistry* **1965**, *240*, 4038-4043.
- 4) Peterson, J. A. *Archives of Biochemistry and Biophysics* **1971**, *144*, 678-693.
- 5) Chakrabarty, A. M. *Annual Review Genetics* **1976**, *10*, 7-30.
- 6) Bagdasarian, M. M.; Amann, E.; Lurz, R.; Ruckert, B. *Gene* **1983**, *26*, 273-282.
- 7) Demontellano, P.R.O.; *Cytochrome P-450, Structure, Mechanism and Biochemistry*; 1st ed.; Plenum Press, New York and London.; 1986.
- 8) Unger, B. P.; Gunsalus, I. C.; Sligar, S. G. *Journal Of Biological Chemistry* **1986**, *261*, 1158-1163.
- 9) Viera, J.; Messing, J. *Gene* **1982**, *19*, 259-268.
- 10) Braybrook, D. PhD thesis, *Studies of P-450cam in Non-aqueous Media.*; Oxford University, 1994.
- 11) Wong, L., Unpublished Results.
- 12) Nickerson, D. P.; Wong, L. L. *Protein Engineering* **1997**, *10*, 1357-1361.
- 13) Nickerson, D.; Wong, L. L.; Rao, Z. H. *Acta Crystallographica Section D-Biological Crystallography* **1998**, *54*, 470-472.
- 14) O'Keeffe, D. H.; Ebel, R. E.; Peterson, J. A. *Methods in Enzymology* **1978**, *52*, 151-157.
- 15) Sligar, S. G. *Biochemistry* **1976**, *15*, 5399-5406.
- 16) Fowler, S. M.; England, P. A.; Westlake, A. C. G.; Rouch, D. R.; Nickerson, D. P.; Blunt, C.; Braybrook, D.; West, S.; Wong, L. L.; Flitsch, S. L. *Journal Of the Chemical Society-Chemical Communications* **1994**, 2761-2762.
- 17) Gunsalus, I. C.; Wagner, G. C. *Methods in Enzymology* . **1978**, *52*, 166-188.

- 18)Staines, A.G.; MSc thesis, *Monooxidations using a cytochrome P450cam mutant*, University of Edinburgh, 1996.
- 19)Koga, H.; Gunsalus, I. C. *Proc. 55th Annual Meeting Japanese Biochemical Society* **1982**, *10*, 82-84.
- 20)Unger, B.; Jollie, D.; Atkins, W.; Dabrowski, M.; Sligar, S. *Federation Proceedings* **1986**, *45*, 1874.
- 21)Koga, H.; Yamaguchi, E.; Matsunaga, K.; Aramaki, H.; Horiuchi, T. *Journal of Biochemistry* **1989**, *106*, 831-836.
- 22)Peterson, J. A.; Lorence, M. C.; Amarneh, B. *Journal of Biological Chemistry* **1990**, *265*, 6066-6073.
- 23)Roitberg, A. E.; Holden, M. J.; Mayhew, M. P.; Kurnikov, I. V.; Beratan, D. N.; Vilker, V. L. *Journal of the American Chemical Society* **1998**, *120*, 8927-8932.
- 24)Aoki, M.; Ishimori, K.; Morishima, I. *Biochimica Et Biophysica Acta-Protein Structure and Molecular Enzymology* **1998**, *1386*, 168-178.

## **4 Biochemical Analysis of Cytochrome P-450cam.**

---

### **4.1 The Solvent Stability of P-450cam Y96A Mutant.**

#### **4.1.1 Introduction.**

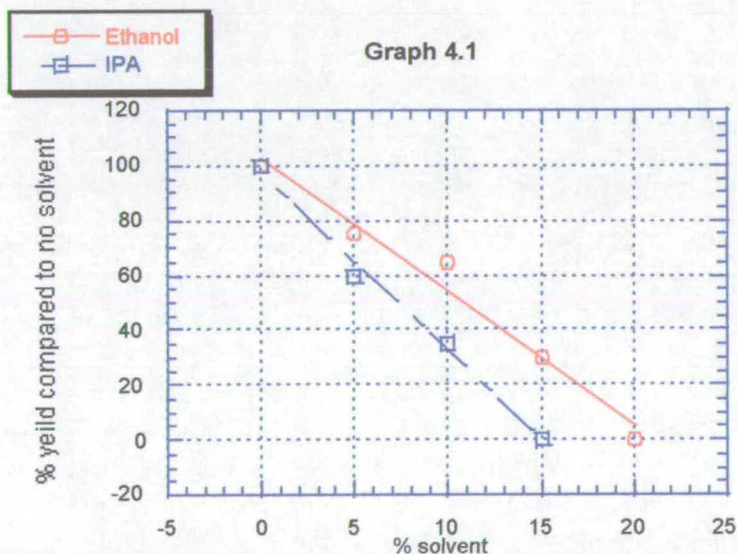
The P-450cam Y96A binding site is extremely lipophilic, therefore successfully binding substrates that are hydrophobic. This leads to a solubility problem when using the reconstituted enzyme system which is in aqueous media. Higher substrate solubility would allow a higher substrate concentration and therefore could give higher product yields. Camphor is probably one of the most soluble of the P-450cam substrates with a saturation concentration of 7 mg/ml. <sup>1</sup> Other substrates have significantly lower solubility than this, normally below 1 mg/ml. For this reason, it was decided to investigate the tolerance of the P-450cam reconstituted system with increasing amounts of organic solvent. A number of limitations were placed on the type of solvent used. As the products were to be analyzed by GC/MS the solvent must not have a high boiling point (e.g. DMSO), as they could interfere with substrate traces. Equally, solvents must be aqueously soluble to avoid generating a biphasic system. Substrates have, in the past, been added to P-450cam systems from ethanol stock solutions, therefore it was known that the reconstituted P-450cam system could



tolerate ~1 % ethanol. For these reasons it was decided to explore the use of alcohols to increase the substrate solubility.

#### 4.1.2 Results of Solvent Stability Analysis.

By addition of small amounts of alcohol to a reaction of camphor with cytochrome P-450cam any detrimental effects can be observed by measurement of the product formation. Thus, by plotting a graph of the loss of turn over vs the alcohol concentration, the effect of the alcohol can be assessed. The results of the solvent stability experiments with ethanol and isopropanol are presented in graph 4.1.



Analysis would indicate that the P-450 reconstituted system does not have a high tolerance for organic solvent. Ethanol would appear to be tolerated more than IPA however, even at a 12 % ethanol concentration the product yield is reduced by half. Although substrate solubility is marginally increased, the loss of product turn over is too great to justify the use of alcohols. The range of other possible solubilizing agents is extensive but due to time restraints, it was decided not to pursue this area of research any further, although this is a valid area for future study.



## 4.2 Alternative Electron Transfer Agents.

### 4.2.1 Introduction.

Previous standard conditions for a biotransformation with the P-450cam system uses 1  $\mu\text{M}$  P-450cam, 1  $\mu\text{M}$  PdR, 10  $\mu\text{M}$  Pd, 10 mg NADH and 1 mg of substrate, in a 1 ml reaction volume.<sup>2</sup> Thus, in order to produce enough material for spectroscopic analysis of products ( $\sim 5$  mg for NMR Spectroscopy), assuming 100 % turn over, at least a 5 ml reaction must be performed. This would mean that a large amount of the co-proteins, especially Pd, will be utilized in every biotransformation, which has financial implications. For this reason it was decided to explore the possibility of utilizing alternate electron transfer agents.

Several electron transfer systems were available to us, either from commercial sources, or through project collaborations with other research groups. Spinach has reductases of the mitochondrial type and both a ferredoxin and a ferredoxin reductase are commercially available, (these are the cheapest commercially available ferredoxin and ferredoxin reductase). The possibility of using *E. coli* (containing a P-450cam gene) as a whole cell biotransformant had to be considered and for this reason it was decided to see if the natural flavodoxin and reductase from *E. coli* were capable of performing the electron transfer.<sup>3,4</sup> P-450BM-3 is potentially a more efficient electron transfer system than P-450cam, as only one protein needs to be produced instead of three. The P-450BM-3 flavo domain (BM-3R) is not commercially available but was made available to us. Old yellow enzyme (OYE) is a universal electron transfer agent present in bakers yeast and was provided by a collaborator. For the electron source, NADH was used in the chemical experiments and when using PdR; for all other protein systems NADPH was used (the other reductases have a higher affinity to NADPH than to NADH). Since one of the principal costs in the biotransformation is the purification of the proteins, crude extract of P-450cam was used to see if the electron transfer was still possible in a complex mixture of proteins. Chemical electron transfer agents such as PMS<sup>5</sup> have been used previously to perform biotransformations<sup>6</sup> and to turn over P-450cam in absence of co-proteins,<sup>7</sup> therefore

its use was also explored. Since the primary goal was high product yield, at a cheaper cost, the experimental measurements were performed based on product yield and not the rate of turn over.

#### **4.2.2 Results of Electron Transfer Optimization.**

The various combinations of reaction mixes are presented based on their success relative to the standard conditions. The product analysis was based on the production of 5-*exo*-hydroxycamphor.

##### **4.2.2.1 Enzymatic Systems.**

No system was able to improve on the electron transfer rate of the reconstituted P-450cam system, indicating the high efficiency of the system. The only systems that were in any way effective were when the reductases were combined with Pd. This is supported by previous studies with adrenodoxin which, despite 37 % homology to Pd, does not react well with P-450cam.<sup>8</sup> For any enzymatic system not containing Pd there was no observable product. This evidence supports previous experimental data that suggests that the putidaredoxin/P-450cam binding has specific interactions, which prevent the P-450cam being reduced by any other system (chapter 1). This is an essential part of the P-450cam gated system, in that it prevents the reduction of the heme when not in the presence of substrate.

Table 4.1: Results of the optimization of electron transfer agents.

Enzymatic controls	Systems without camphor or NADH/NADPH or P-450cam showed no products.	
Enzymatic systems	Pd + PdR	100 %
	Pd + SdR	90 %
	Pd + FLDR	10 %
	Pd + OYE	<5 %
	Sd + PdR	no detectable product
	FLD + PdR	no detectable product
	Sd + SdR	no detectable product
	Sd + OYE	no detectable product
	FLD + FLDR	no detectable product
	Pd + FLD <sup>a</sup>	no detectable product
	BM-3R <sup>a</sup>	no detectable product
Chemical controls	PMS does not react with camphor or P-450cam	
Successful PMS/enzymatic systems (50 $\mu$ M PMS)	Pd + PMS	50 %
	PMS	30 %
Unsuccessful PMS/enzymatic systems	PdR + PMS	30 % <sup>b</sup>
	FLD + PMS	30 % <sup>b</sup>
	Sd + PMS	20 % <sup>b</sup>
Crude extract controls	unlysed pellet, or pellets containing no camphor or NADH gave no product	
Crude extract systems	lysed pellet from a wild-type cytochrome P-450cam expression	10 %

<sup>a</sup> product peaks were detected but could not be identified.

<sup>b</sup> percentage not above level with PMS alone.

It would appear therefore from the data, table 4.1, that the reductases are less specific in their binding interaction with their respective ferredoxins or flavodoxins, as both flavodoxin reductase and ferredoxin reductase are capable of interacting with Pd. The spinach reductase is understandably better than the *E. coli* reductase since its role is to interact with ferredoxins.

#### 4.2.2.2 PMS/ Enzymatic System.

Results from previous experiments have indicated that PMS can transfer electrons to the P-450cam molecule from NADH alone and with an increased rate in the presence of Pd.<sup>6</sup> This is supported by our data which places a more efficient transfer to the P-450cam when PMS is present with Pd than using PMS alone. The PMS/Pd system is, however, still inferior to the fully reconstituted system. PMS combined with any reductase, including PdR, was not capable of enhancing the product yield above that of PMS alone.

Since PMS is clearly capable of transferring the electron from NADH to P-450cam, the improved efficiency upon addition of the Pd must be due to the high affinity between the Pd and the P-450cam, indicating that this affinity is not only higher than other enzyme systems, but better than an interaction from the PMS.

The concentration of PMS used at 50  $\mu\text{M}$  is based on the value used previously <sup>7</sup> and was not fully optimized. Some crude optimization of the protocol was carried out by varying the concentration. It was known that increasing the PMS concentration has the unwanted effect of interfering with the GC assay, as the PMS concentration becomes significantly higher than any other molecule. This aside, increasing the concentration to 100  $\mu\text{M}$  improved the product yield by 10 % without significant interference with the 5-*exo*-hydroxycamphor detection. PMS concentrations above 100  $\mu\text{M}$  did not appear to improve the overall yield. Reducing the PMS concentration to 10  $\mu\text{M}$  reduced the efficiency of the system five-fold, implying that 100  $\mu\text{M}$  is close to the optimal concentration when using these conditions.

PMS can successfully transfer electrons to the P-450cam without any co-proteins, albeit at a lower yield, as discussed above. Only one other chemical system, titanium (III) citrate, has been previously employed in the reduction of P-450cam in anaerobic conditions and upon examination it was found that titanium (III) citrate is capable of converting oxygen to peroxide in aqueous media.<sup>5</sup>

#### 4.2.2.3 Use of Crude Enzyme Samples.

The use of crude enzyme samples is attractive as it would significantly reduce the time taken to purify the enzymes. Here the P-450cam protein levels for the crude extract, plus the Pd and PdR concentrations were in fact two times higher than the purified recombinant system, however, the yield of hydroxylated product was 10 %. The only other system with the crude extract to yield any product was the combination of Pd with PMS. In all cases however, there were unassigned peaks in the GC trace which can be attributed to the milieu of competing reactions that take place in a crude cell extract.

#### 4.2.3 Conclusions.

It has thus been demonstrated that it is possible to activate catalysis with the P-450cam system using alternative electron transfer agents. It is clear from the data that, unless using PMS, it is not viable to perform the reaction without the presence of Pd. The P-450cam/Pd affinity would appear to be too specific to allow the use of other ferredoxins. For the PMS system alone, the yield is sufficient to consider it as an alternative, however, this data was for camphor, which has 100 % conversion to product. For most unnatural substrates, the yield is significantly reduced thus making the additional use of a low yielding electron transfer system unviable. Crude extracts, whilst saving the time involved in purification, have the drawback of producing many side products. With the drop in yield and these competing reactions it is not thought viable to perform these biotransformations with our crude enzyme mixture.

Both the PMS/Pd and the PdR/Pd system give acceptable yields (50 % and 100 % respectively). For their use, the lower yield of the PMS/Pd system and potential problems in extraction will have to be assessed vs the need to purify PdR, which is potentially problematic as it is known to express with low yields, (unpublished data).

### 4.3 Binding Assays and Substrate Turn over.

#### 4.3.1 Introduction.

In the study of enzymes knowing whether a new substrate binds to the active site is crucial. In many cases this must rely on biochemical assays, the detection of products or structural elucidation *via* NMR or X-ray crystallography. In the case of cytochrome P-450cam, the binding of substrate can be measured directly *via* UV/Vis spectroscopy. The heme group has a characteristic absorbance, known as the Soret band,<sup>9</sup> at about 400 nm caused by a  $\pi$ - $\pi^*$  (ligand to ligand) charge transfer. As discussed previously (chapter 1) a substrate that enters the binding site displaces the heme bound water changing the electronic configuration of the heme. This change of electronic configuration shifts the Soret band from 417 nm (water bound) to 390 nm (substrate bound) known as a type 1 shift. Thus, the binding of substrate to P-450cam can be measured by observing this shift in the Soret band, providing a quick assay for determining if a substrate will bind.

Extensive studies have been made of the Soret band at various stages of substrate binding.<sup>10,11</sup> Carbon or hydrogen substituents which are in the vicinity of the heme do not interact with the iron, which gives a Soret band at 390 nm. Nitrogen, sulfur and oxygen, however, bind as ligands to the iron centre (e.g. in the cases of imidazole, metyrapone, dithiothreitol, cyclohexanol or indeed water).<sup>10</sup> Where one of these heteroatoms is present, ligand binding can clearly be observed (for nitrogen ligand interaction a Soret band at ~423 nm is observed, for oxygen ~417 nm, and for sulfur two bands are observed at ~374 nm and 461 nm)<sup>10</sup> This binding of heteroatoms also means that the change from hexa-coordinate low spin to penta-coordinate high-

spin does not occur and the enzyme is not capable of catalyzing the hydroxylation of the substrate.

It is important to note that although the binding assay is a good guide to potential substrates it is not a guarantee that substrates causing this shift will be oxidized, they may act as an inhibitor or the competing peroxide shunt pathway (chapter 1) could prevent product formation. More accurate information is obtained from NADH consumption assays and product isolation (see later). The binding assay does however indicate whether a compound enters the binding site.

#### 4.3.1.1 Calculation of Binding Constants.

The binding constants can be calculated by the method of Dawson <sup>12</sup> using the following equation, figure 4.1.

$$\log \left[ \frac{\Delta A}{\Delta A_{\max} - \Delta A} \right] = \log [S] - \log K_d$$

Figure 4.1 Where  $\Delta A$  is the change in absorbance of the Soret band,  $\Delta A_{\max}$  is the maximum possible change in absorbance for this substrate,  $[S]$  is the substrate concentration and  $K_D$  is the dissociation constant.

Thus by plotting  $\log (\Delta A / \Delta A_{\max} - \Delta A)$  vs.  $\log [S]$ , a graph with intercept of  $-\log K_D$  is produced; an example dataset is shown in appendix 1

#### 4.3.1.2 Binding with Heme Ligation.

When substrates bind with ligation to the heme it is not possible to measure the change in the Soret band (nitrogen ligation occurs too close to the 417 nm to get

accurate measurements and oxygen ligation occurs at 417 nm so no change in absorbance is observed). The binding properties of one of these substrates can be determined by observing the displacement of a ligand, rather than a heme bound water. Strong binding substrates would interfere with the assay therefore DTT is used. DTT is a weak substrate for P-450cam ( $K_D > 100 \mu\text{M}$ ) with sulfur ligation that gives characteristic peaks at 370 nm and 460 nm. Its displacement can thus be monitored spectroscopically and, due to its low binding affinity, binding constants (or more correctly DTT displacement constants) can be calculated.

#### 4.3.1.3 Percentage High-Spin.

The maximum percentage high-spin can be measured by addition of a saturating concentration of substrate and measuring the relative absorbance at 390 nm and 417 nm, for an example graph appendix 1. This is a measure of the water accessibility to the binding site in the presence of substrate, therefore should reflect how tightly the substrate will bind. This could thus be used as a guide to stereoselectivity and peroxide production, since substrates that have a low percentage high-spin will be more mobile therefore have lower product stereoselectivity and higher peroxide production. It should be noted that the percentage high-spin may not be related to water occupancy (see chapter 7).

#### 4.3.1.4 Substrate Turn-over.

Substrate turn over conditions were designed with excess NADH to provide maximum product formation, even with the high levels of peroxide production previously exhibited with the P-450cam Y96A mutant.<sup>13-15</sup> The products of the turn over can be easily extracted into organic solvent and analyzed by GC/MS. Peaks not present in the substrate GC/MS spectrum were analyzed as potential products. Peaks exhibiting mass ions increased by 16 (i.e. oxygen addition) were assumed to be hydroxylated products if the breakdown peaks were consistent with a hydroxylated



product. Similarly products with mass ions increased by 14, (monooxygenation followed by double hydrogen loss), and fitting the correct mass spectrum profile, were assumed to be keto-products. If possible, the oxidized product were further characterized by analysis of the mass spectrum breakdown profile.

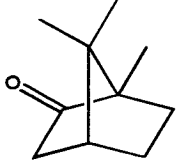
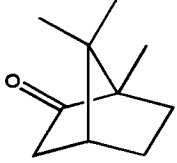
### 4.3.2 Substrates Analysis.

The substrates studied can be divided into a number of classes based upon analogies to known substrates, these are: camphor and adamantane analogues; diphenylmethane analogues; naphthalene analogues; chlorinated aromatic molecules; aromatic molecules with carbocycles; aromatic molecules with nitro-heterocycles; aromatic molecules with oxygen heterocycles.

#### 4.3.2.1 Camphor.

R-(+)-Camphor [2], the natural substrate for P-450cam, has a strong binding affinity to wild-type P-450cam ( $K_D=2.0 \mu\text{M}$ ) and near complete water dissociation (98 % high-spin).<sup>16</sup> The P-450cam Y96A mutant has a very similar binding affinity to the wild-type (table 4.2), however, the percentage high-spin is considerably lower.

Table 4.2: Camphor hydroxylation by P450cam mutants Y96F and Y96A.

	Structure	% high-spin	$K_D$ ( $\mu\text{M}$ )	Products
Y96A		48	2.5	5 hydroxylated 1 keto product
Y96F		59 <sup>17,18</sup>	-	5 hydroxylated 1 unidentified

This would imply that the factors controlling the P-450cam/camphor recognition are maintained, but the substrate has greater mobility in the binding site. It is possible that this is solely due to the loss of the hydrogen bond to camphor. This can be confirmed by comparison to the Y96F hydroxylation (table 4.2). While it is clear that the Y96F and Y96A mutants have similar substrate properties, i.e. lower % high-spin and the same number of products, the P-450cam Y96A mutation has obviously had a greater effect on the product distribution than could be attributed to the loss of the hydrogen bond alone.

#### 4.3.2.1.1 *Product Analysis of Camphor.*

The products of the P-450cam Y96A mutant reaction with camphor were analyzed by GC and these results are presented in figure 4.2.

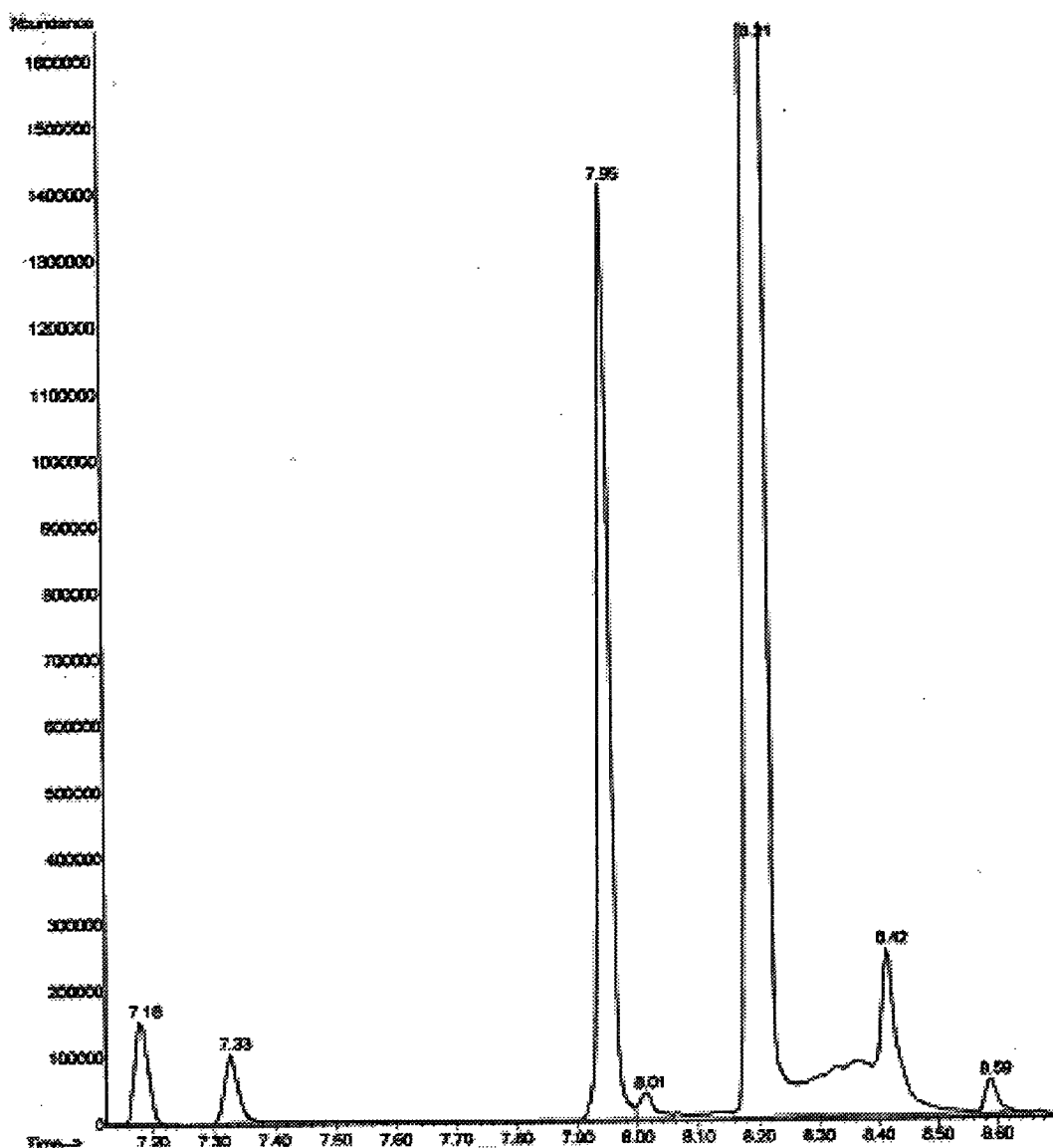


Figure 4.2: GC trace of the hydroxylation of camphor by the P-450cam Y96A mutant.

Hydroxylation of R-camphor by wild-type enzyme gives one product, the 5-*exo*-hydroxy camphor. The GC trace of the P-450cam Y96A mutant hydroxylation clearly shows the presence of other metabolites in addition to the 5-*exo*-hydroxycamphor (8.21 min). This is consistent with previous studies which indicated that the removal of the hydrogen bond, by making the Y96F mutant, results in some loss of selectivity.<sup>17,18</sup> The percentage products of the Y96A and Y96F mutant turn overs are presented in table 4.3.

Table 4.3: Comparison of the products of biotransformations of camphor with P-450cam Y96 mutants.

	wild-type (%)	Y96F mutant <sup>17,18</sup> (%)	Y96A mutant (%)
5- <i>exo</i> -hydroxycamphor	99	92	74
5- <i>endo</i> -hydroxycamphor		1	4
6- <i>exo</i> -hydroxycamphor		2	15
6- <i>endo</i> -hydroxycamphor		0 <sup>a</sup>	1 <sup>a</sup>
9-hydroxycamphor		<1	4
3- <i>exo</i> -hydroxycamphor		4	1
2,5-di-ketocamphor		0 <sup>b</sup>	1
% High-spin	99	59	45

<sup>a</sup> 6-*exo* and 6-*endo*-hydroxycamphor can undergo inter-conversion *via* a keto intermediate therefore 6-*endo* levels might be due to conversion from 6-*exo* and not by direct hydroxylation by the enzyme.  
19

<sup>b</sup> Although 2,5-diketocamphor was not positively identified a peak can be seen in the GC trace at the correct retention time.

The six hydroxylated camphor molecules can all be distinguished by their retention times and mass spectral profiles, table 4.4.

Table 4.4: MS profiles of hydroxycamphors (for numbering see chapter 1).

Hydroxy Camphor	Retention time	Highest mass ions
3- <i>exo</i>	7.33	83, 84, 55, 125, 69, 168
6- <i>exo</i>	7.95	153, 108, 107, 93, 70, 55
6- <i>endo</i>	8.01	153, 70, 55, 97, 108, 69
5- <i>exo</i>	8.21	111, 168, 125, 69, 55, 83
5- <i>endo</i>	8.42	153, 111, 107, 108, 93, 55
9-	8.59	95, 108, 109, 93, 67, 137

In order to determine the products of the hydroxylation detailed mass spectral analysis was necessary. The camphor molecule has a number of 'high mass' ions in the mass spectrum (137, 109, 95). Careful analysis of the possible breakdown peaks for camphor indicate how these products are formed, figure 4.3.

From this, the identity of peaks in the GC-MS spectrum was determined, since the breakdown pathway will be affected differently depending on the site of hydroxylation.

The 3-hydroxy products of camphor can be distinguished from the others because they have a peak at + 95, due to the loss of the hydroxyl group from next to the carbonyl, figure 4.4.

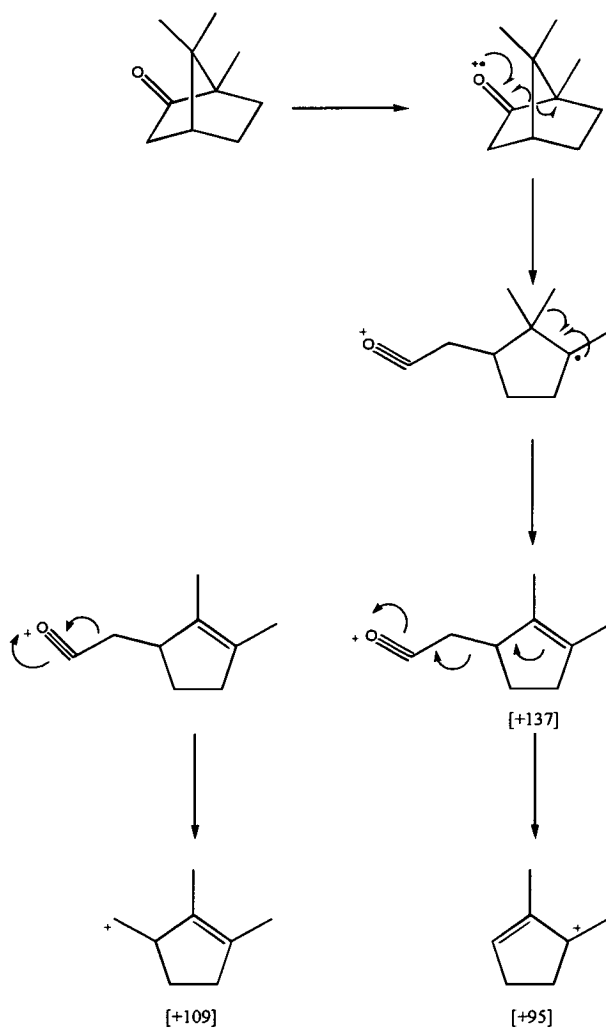


Figure 4.3: Possible camphor breakdown products formed by EI mass spectrometry.

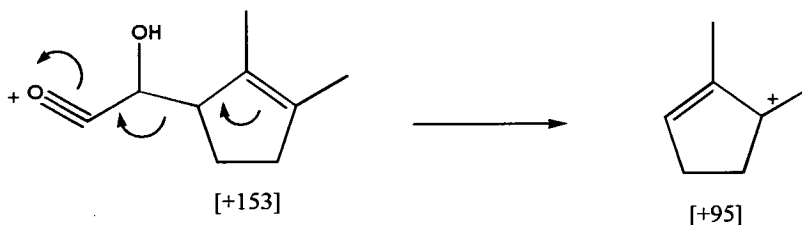


Figure 4.4: Possible formation of the +95 species in 3-hydroxycamphor.

In addition the 3-hydroxy products have significantly lower retention times on the column, due to an internal H-bond to the camphor carbonyl, figure 4.5.

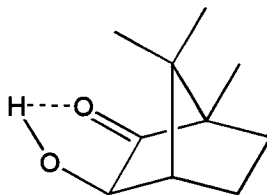


Figure 4.5: Intramolecular hydrogen bond present in 3-*exo*-hydroxycamphor.

The 3-*exo* species is characterized by a significant peak at -28 (mass ion 140) corresponding to the loss of carbon monoxide. This is distinctly different from the 3-*endo* product which has loss of water, -18 (mass ion 150), as a major peak since the 3-*exo* product has no access to a readily extractable hydrogen for the alcohol moiety, figure 4.6.<sup>20</sup>

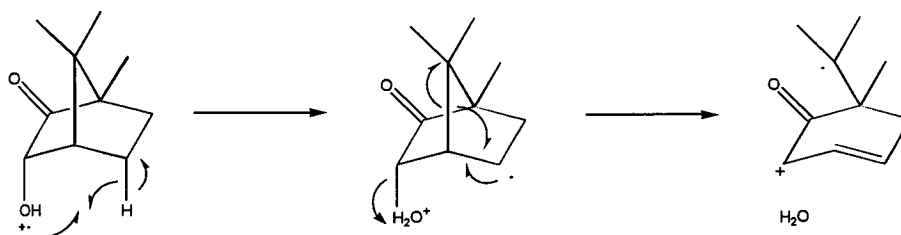


Figure 4.6: Possible loss of water from 3-*endo*-hydroxycamphor.

The 9-*exo*-hydroxycamphor can be characterized by the loss of methanol as described previously.<sup>20</sup>

The 5- and 6-hydroxy products could not be identified by the mass spectrum analysis alone, therefore had to be determined from authentic standards. The 5-*exo*-hydroxycamphor was made by hydroxylation of R-camphor by wild-type P-450cam. The 5-*endo*-hydroxycamphor was provided by Prof. Sabine Flitsch. The 6-*exo* and 6-*endo*-hydroxycamphor was provided by Dr. Gideon Grogan, University of Edinburgh.

The 2,5 di-keto product was identified by an authentic standard although it was not previously identified in the hydroxylation of camphor by Y96F mutant.<sup>17,18</sup> When performing the biotransformation reaction with the Y96F mutant the same 2,5 di-keto product was produced. The mechanism of formation of this di-keto species is not known, however, it only appears in biotransformations that have a non-saturated amount of camphor and may therefore could be due to a second 5-hydroxylation of the 5-*exo*-hydroxycamphor followed by loss of water, figure 4.7; alternately it could be due to the presence of an alcohol dehydrogenase.

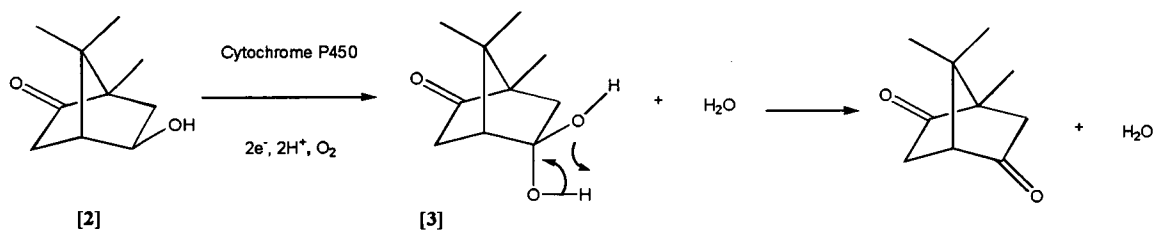


Figure 4.7: Possible formation of 2,5-diketocamphor.

When the biotransformation was repeated in the presence of alcohol dehydrogenase the keto product increased to 30 % of the total product distribution, with the corresponding decrease in the 5-*exo*-hydroxy product. This could imply that the keto product was formed due to contamination of one of the protein preparations by an alcohol dehydrogenase.

#### 4.3.2.1.2 *Effect of the P-450cam Y96A Mutation on R-Camphor Hydroxylation.*

The additional removal of the phenyl ring (i.e. the difference between the Y96A and Y96F mutant) increases the movement of camphor in the binding site beyond that of the removal of the tyrosine-camphor hydrogen bond, resulting in an increase of other hydroxy-camphor products. The significant presence of 5-*exo*-hydroxycamphor indicates that the movement of the camphor molecule is still restricted, despite the removal of the hydrogen bond to the tyrosine. This is also found in previous studies with the wild-type enzyme which showed that substrate with no geminal methyl groups, e.g. norcamphor, had increased movement in the binding site. This indicates both the hydrogen bond to tyrosine 96 and the hydrophobic interactions with the geminal methyl groups are essential to maintain 100 % product specificity.<sup>17,18</sup> The 6-*ex*-hydroxy camphor is the primary secondary product as seen for the Y96F mutant, however, the percentage of product is significantly increased from 2 % to 15 %. The percentage of other products essentially remains the same as for the Y96F mutant. These results would indicate that the general rotation of the camphor is still inhibited by hydrophobic interactions between the geminal methyl groups and the binding site but the molecule has some minimal increase in movement about the C5 carbon that additionally allows the C6 position to be hydroxylated.

The binding data on P-450cam Y96A mutant indicated that the enzyme rests 48 % in the high-spin state, at 390 nm, when saturated with camphor. This partial spin state is probably due to the presence of a water molecule in addition to the camphor. It has been shown with the wild-type enzyme that the camphor must be able to move off the enzyme binding site to allow the oxygen to bind to the heme and it is thought that the hydroxy group of the tyrosine helps to push the camphor back towards the iron. This hypothesis would be supported by our data for the Y96A and Y96F mutants, as the increased space at the 'top' of the binding site allows the camphor molecule to exist in a place that partially allows water to enter the binding site and interact with the heme. The extra space in the Y96A mutant compared to the Y96F mutant is reflected in the lower percentage high-spin.



Other P-450 monooxygenases that hydroxylate camphor include P-450LM2 (63 % 5-*endo*, 14 % 5-*exo* and 16 % 3-*endo*), and P-450soy (60 % 6-*endo*, 9 % 5-*exo*, 5 %-*endo* and 2 % 3- *endo*) but only P-450camr from *rhodococcus sp.* hydroxylates to give a stereo selective product, 100 % 6-*endo*-hydroxycamphor.<sup>19</sup>

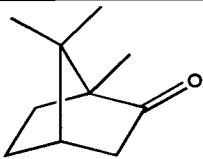
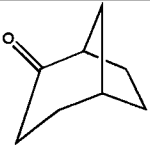
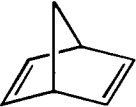
#### 4.3.2.2 Hydroxylation of Camphor/ Adamantane Analogues.

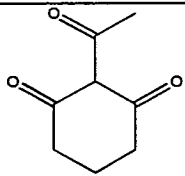
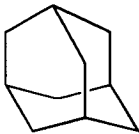
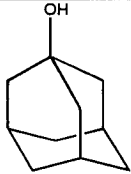
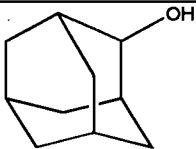
Although camphor hydroxylation is less stereoselective with the use of the P-450cam Y96A mutant, it was still of interest to examine compounds known to be substrates for the wild-type enzyme.

S-camphor [23], is hydroxylated to produce mostly 5-*exo* product (40 %), but with significantly greater amounts of the other hydroxy products (6-*exo*=33 %, 5-*endo*=12 %, 3-*exo*=15 %). The lower  $K_D$  is indicative of the selectivity of the enzyme for R-camphor, however, the % high-spin is similar, thus suggesting that once the substrate is in the binding site the poor complementarity between the enzyme and the S-camphor allows partial water access to the heme, as seen with the R-camphor.

Bicyclo[3.2.1]octan-2-one [24], is structurally similar to camphor and it was of interest to see if the extra carbon in the bicyclic ring would confer greater stereoselectivity. As can be seen from the data, table 4.5, oxidation of substrate 24 occurs with lower stereoselectivity than camphor.

Table 4.5: Biochemical data of compounds 23-28.

Compound	Structure	% high-spin	$K_D$ ( $\mu$ M)	Products	Retention time and percentage
23		45	10	starting material	5.72
				hydroxy product	7.33
				hydroxy product	8.01
				hydroxy product	8.21
				hydroxy product	8.42
24		32	-	starting material	4.90
				keto product	6.07 (trace)
				hydroxy product	6.15 (3%)
				hydroxy product	6.57(18%)
				hydroxy product	6.67 (25%)
				hydroxy product	7.15 (1%)
				hydroxy product	7.21 (trace)
				keto product	7.53 (2%)
				keto product	7.82 (50%)
				hydroxy product	8.11 (10%)
				hydroxy product	8.23 (1%)
25		56	a	a	

26		Oxygen binding	-	starting material	7.33
10		45	-	starting material	5.49
				hydroxy product	7.10 (trace)
				hydroxy product	7.18 (trace)
				dihydroxy product	7.56 (1%)
				dihydroxy product	8.20 (96.5%)
				dihydroxy product	8.62 (2.5%)
27		Oxygen binding	-	starting material	7.10
				dihydroxy product	7.56 (1%)
				dihydroxy product	8.20 (96.5%)
				dihydroxy product	8.62 (2.5%)
28		Oxygen binding	-	starting material	7.18
				dihydroxy product	7.56 (1.5%)
				dihydroxy product	7.95 (47%)
				dihydroxy product	8.20 (9%)
				dihydroxy product	8.62 (6.5%)
				dihydroxy product	8.78 (4%)
				dihydroxy product	8.95 (5%)
				keto-hydro product	9.11 (18%)
				keto-hydro product	9.15 (1.5%)

<sup>a</sup> - Compound could not be analyzed due to excessive evaporation

The low percentage high-spin indicates a high degree of mobility of the substrate in the binding site. This is possibly due to the absence of the geminal carbons in **24** which, as with other camphor analogues, have been shown to decrease stereoselectivity when absent (chapter 7).

Bicyclo[2.2.1]hepta-2,5-diene [**25**] is a norbornane [**9**] analogue, and it was thought that, due to the greater movement exhibited by camphor analogues in the P-450cam Y96A mutant binding site that dihydroxylation might be possible to give a di-epoxide species. Unfortunately, due to the volatility of the compound, a biotransformation could not be successfully performed to test this. Attempts to prevent this proved unsuccessful. Since the compound binds, table 4.5, it could be of interest to pursue methods of performing the turn over without evaporation of substrate.

2-Acetyl-1,3-cyclohexanedione [**26**] was of similar size to the other compounds studied, however, it contains additional carbonyl groups and therefore it was of interest to examine the effect this would have on the enzyme/substrate recognition. Data, table 4.5, clearly shows that compound **26** is a substrate for the enzyme. The carbonyl group would not be expected to interact with the heme therefore the O-ligation might be caused by a bound water hydrogen bonded to the heme and the substrate

Adamantane [**10**] is a known substrate for wild-type P-450cam<sup>20</sup> giving the 1-adamantanol [**27**] as the product. This is not indicative of high regioselectivity, as the adamantane molecule only has two distinct carbons, of which the tertiary carbon is the more favoured for hydroxylation, on thermodynamic grounds. The binding constant for adamantane with wild-type enzyme is 50  $\mu\text{M}$  with a percentage high-spin of 99 %. For the P-450cam Y96A mutant, as for camphor, the percentage high-spin is lower than the wild-type, indicative of the increased movement of the adamantane in the mutant binding site. This is also confirmed by the increase in the number of products, table 4.5. Through commercially available standards, the two hydroxylated peaks in the product mixture were identified as the 1-adamantanol (1 %) and the 2-adamantanol [**28**] (<1 %). Three peaks at a  $M^+$  of 168 (i.e. + 32) were found. It was

possible that these products were due to the biohydroxylation of 1- and 2-adamantanol. To confirm this **27** and **28** were hydroxylated with the P-450cam Y96A mutant. The 1-adamantanol gave three products similar to the unidentified peaks from the adamantane hydroxylation. The 2-adamantanol gave six di-hydroxylated products and three keto-products. By comparison of the products to literature values and analysis of the products produced by each adamantanol the di-hydroxylated species could be identified. The 1-adamantanol produces, in order of retention time, the 1,2-diol (1 %) the 1,3-diol (96.5 %) and the 1,4-diol (2.5 %). The 1,2-diol is distinguishable from the other products by the lower retention time caused by the internal hydrogen bond. The 1,3-diol is distinct by its absence from the 2-adamantanol hydroxylation. The 2-adamantanol produces, in order of retention time, the 1,2-diol (1.5 %), the 1,4-diol (47 %), *cis* and *trans* 2,4-diol (9 %, 6.5 %) *cis* and *trans* 2,6-diol (4 %, 5 %) and three keto products (18 %, 1.5 % and 7 %), figure 4.8.

As for the camphor, it is believed that the keto products are due to the presence of alcohol dehydrogenase.

The hydroxylation of adamantane is significant because it is the first example of a double hydroxylation by P-450cam. In addition to this the hydroxylation of adamantanol is the first example of alcohol hydroxylation by P-450cam.

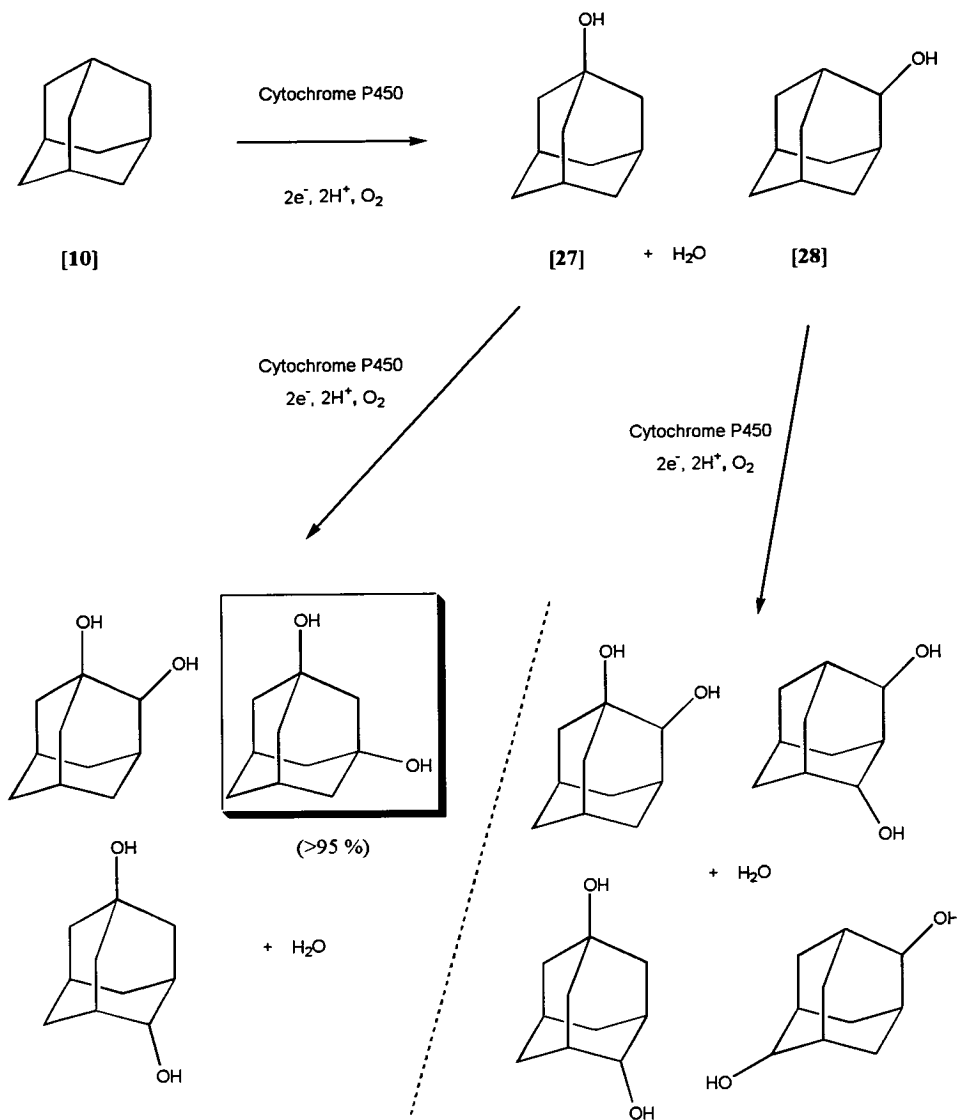
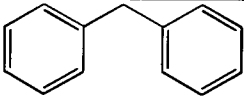
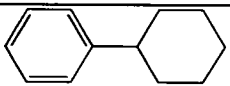
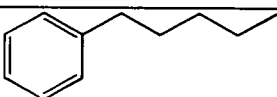
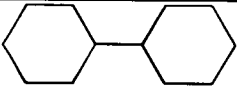


Figure 4.8: Hydroxylation of adamantane by P-450cam Y96A mutant, the major product (1,3-adamantan-diol) is boxed.

#### 4.3.2.3 Diphenylmethane Analogues.

Both diphenylmethane [13] and phenylcyclohexane [18] are known substrates for P-450cam. It was therefore of interest to explore the binding of a phenyl substituted pentane chain, amyl benzene [29], table 4.6.

Table 4.6: Biochemical data of compounds 13, 18, 29, 30.

Compound	Structure	% high- spin	$K_D$ ( $\mu\text{M}$ )	GC peaks	Retention time and percentage (in min, (% total))
13		55	1.7 <sup>b</sup>	starting material hydroxy product	8.67 10.96 (8%)
18		85 <sup>a</sup>	1.8	starting material hydroxy product	7.72 9.52 (2 %)
29		85	3	starting material hydroxy product	5.64 8.7 (9.5 %) 10.8 (15 %)
30		45	3.5	starting material hydroxy product	6.34 8.56 (<1%)

<sup>a, b</sup>-this is verified by previous data <sup>21,22</sup>

The binding constant for amyl benzene is only slightly higher than for diphenylmethane and phenylcyclohexane, showing that the recognition of the P-450cam mutant binding site is still maintained with aliphatic chains on the molecule. The percentage high-spin is the same for amyl benzene as for phenylcyclohexane and both are significantly higher than diphenylmethane. There are a number possible reasons for this difference. The increased size of the aliphatic molecules could cause greater displacement of the iron bound water. An alternative to this theory would be, that due to the increased degrees of mobility allowed with an aliphatic ring or chain they have greater movement and displace the iron bound water. It is possible that the percentage high-spin is purely related to the orientation of the substrate in the binding site, thus each of the above substrates has a different orientation that causes the changes in percentage high-spin. One alternative theory is more controversial, and

states that it is the effect of the substrate on the heme environment that causes the change and not a direct change in the iron bound water.<sup>23</sup>

The binding of diphenylmethane, phenylcyclohexane and amylbenzene implies that the P-450cam Y96A mutation has created a binding affinity for phenyl rings i.e. a 'handle', possibly in the proposed phenyl binding pocket (see chapter 7). To test this theory the binding of a non-aromatic substrate, of similar shape to the other substrates studied, was explored. Bicyclohexyl [30] binds with similar efficiency to diphenylmethane, phenylcyclohexane and amyl benzene (table 4.6) and with only a slightly lower percentage binding. This would imply that the phenyl ring on the substrates enhances binding but is by no means essential. The slightly lower percentage high-spin here would suggest that, of the above theories, the substrate orientation might play the most significant role as the bicyclohexyl is bulkier and more mobile and thus should have a higher percentage high-spin if these are the controlling factors.

Diphenylmethane has been previously shown to produce the 4-hydroxy product,<sup>22</sup> figure 4.9, although this probably results from formation of the epoxide, followed by ring opening to form the more favoured para- substituted product and by a hydrogen shift.<sup>24</sup>

The phenylcyclohexane has been shown to have similar stereoselectivity,<sup>21</sup> producing 3- and 4-hydroxy products. These products may all be present in the biotransformation but may not be separated by the GC method used here and hence appear as one product. The product of the amyl benzene biotransformation could be identified from the mass spectral breakdown peaks which showed the presence of  $M^+$  peak of 118 corresponding to the loss of ethanol (46) and no peaks for loss of methanol or propanol, suggesting the 1-benzylpentan-4-ol product.



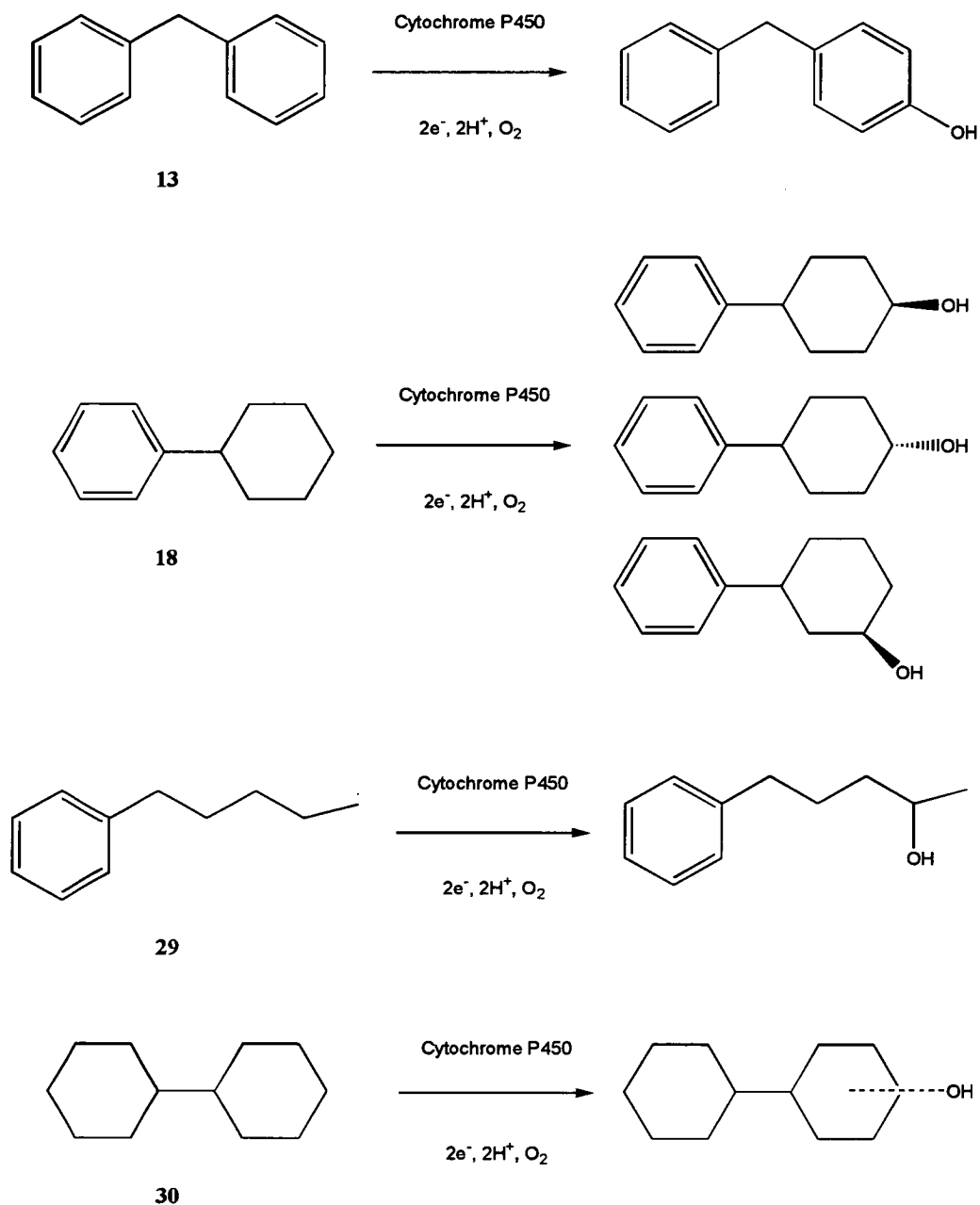


Figure 4.9: Known products of the biotransformations of compounds 13, 18, 29, 30.

The bicyclohexyl product could not be readily identified and as with the phenylcyclohexane, probably contains multiple hydroxylated products that did not separate on the GC.

For these substrates a consensus about the binding properties of the P-450cam Y96A mutant towards molecules with a hydrophobic 'handle' can be seen. The

binding site orientates the substrate in a position so that the hydroxylation takes place opposite the hydrophobic handle. The region that binds this handle has a strong preference for aromatic rings although it still binds aliphatic groups.

#### 4.3.2.4 Naphthalene Analogues.

Naphthalene [19] is a known substrate of the P-450cam Y96A mutant <sup>25</sup> and it is thought that it binds similarly to diphenylmethane and the phenylcyclohexane, with one of the naphthalene rings in the hydrophobic pocket.

Naphthalene analogues 1,2,3,4-tetrahydronaphthalene [31] (tetralin) and 1,2-dihydronaphthalene [32] both bind to the Y96A mutant enzyme, table 4.7, with percentage high-spins comparable to the more mobile substrates such as adamantane.

Tetralin has previously been shown to be a substrate <sup>26</sup> for the wild-type enzyme producing the 1,2,3,4-tetrahydronaphth-1-ol [33]. Substrate 31 produces one hydroxylated product. Both 1,2,3,4-tetrahydronaphth-1-ol and 1,2,3,4-tetrahydronaphth-2-ol [34] were commercially available and upon analysis the product was determined to be the 1,2,3,4-Tetrahydronaphth-1-ol, figure 4.10.

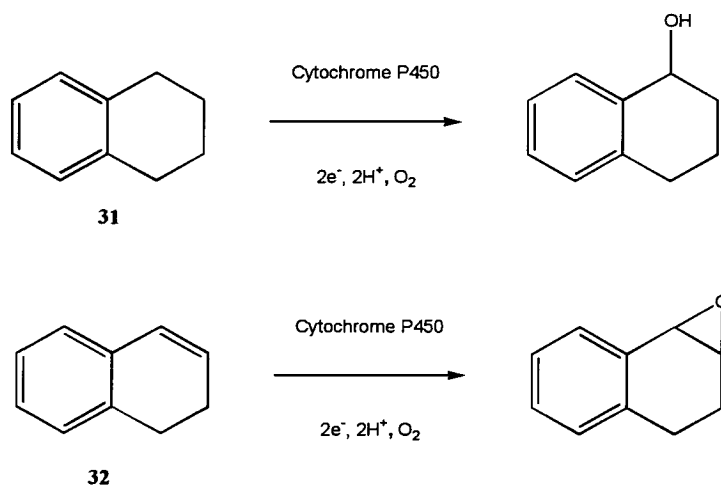
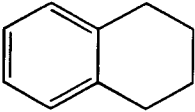
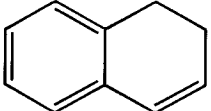
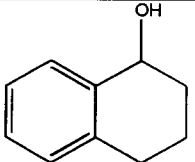
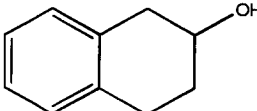
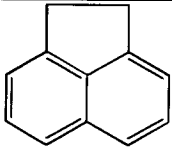
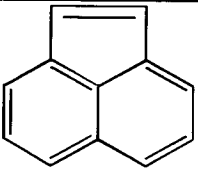


Figure 4.10: Products of the biotransformation of compounds 31 and 32.

When analyzing the alcohols as substrates they were shown to bind to the heme *via* oxygen ligation and produce no product upon biotransformation. Substrate **32** produces an epoxide, which exhibits similar fragmentation patterns to **33** and **34** which suggests it is 1,2-epoxy-1,2,3,4-tetrahydronaphthlene.

Table 4.7: Biochemical data of compounds 31-36.

Compound	Structure	% high-spin	$K_D$ ( $\mu\text{M}$ )	GC peaks	Retention time and percentage in min, (% total)
<b>31</b>		55	-	starting material hydroxy product	5.73 8.18 (8%)
<b>32</b>		32	-	starting material epoxide product	5.73 (98.5 %) 7.92 (1.5 %)
<b>33</b>		Oxygen binding	-	starting material	8.18
<b>34</b>		Oxygen binding	-	starting material	8.66
<b>35</b>		a	b	b	-
<b>36</b>		a	b	b	-

a-Compound was too insoluble for accurate readings

b-Compound could not be analyzed due to excessive evaporation

Acenaphthene [35] and acenaphthylene [36] were both potential substrates for the P-450cam mutant however due to poor aqueous solubility and high levels of evaporation during the biotransformation their substrate potential could not be confirmed.

#### 4.3.2.5 Chlorinated Aromatic Compounds.

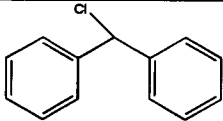
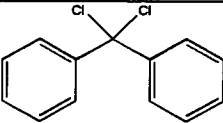
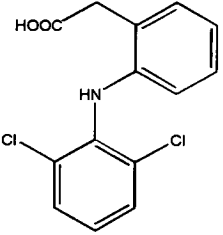
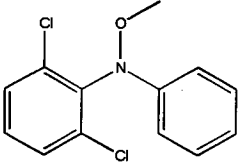
Halogenated substrates have been shown to be substrates for wild-type P-450cam,<sup>27-30</sup> sometimes exhibiting dehalogenation,<sup>6,29,31</sup> which is of great interest in studies of biodegradation of polyhalohydrocarbons.<sup>30,32-34</sup> Here we were interested in chlorinated substrates because of their potential for crystallographic studies, as heavy atoms in the substrate would make interpretation of the electron density maps easier.

Diclofenac [37] is an important anti-inflammatory agent and it has been suggested that hydroxylation of either of the aromatic rings could enhance biological activity. Selective hydroxylation of diclofenac is chemically difficult and it was hoped that hydroxylation of diclofenac or a synthetic intermediate might be achieved using the P-450cam Y96A mutant enzyme system.

Monochlorodiphenylmethane [38] and dichlorodiphenylmethane [39] are of potential interest due to their similarity to a known substrate, diphenylmethane, and are thus of potential use as heavy atom substrates in crystallographic studies, table 4.8.

Of these two compounds only the monosubstituted diphenylmethane bound to the enzyme, with a high percentage high-spin, suggesting minimal movement in the binding site. The biotransformation produced one hydroxylated product for the monochlorinated diphenylmethane (this product could not be identified) and none for the dichlorinated molecule.

Table 4.8: Biochemical data of compounds 38–40.

Compound	Structure	% high- spin	$K_D$ ( $\mu\text{M}$ )	GC peaks	Retention time and percentage in min, (% total)
38		81	-	starting material hydroxy product	18.35 19.63
39		a	-	starting material	17.98
37		a	a	a	-
40		93	-	starting material product	15.49 16.92 (1%)

a-Compound was too insoluble for accurate readings.

Unfortunately diclofenac could not be studied directly because it was insoluble in any solvent at usable concentrations. A synthetic precursor to diclofenac, [40], was synthesized in-house. This bound to the enzyme with a high value for the percentage high-spin. A binding constant could not be determined accurately due to compound insolubility. The biotransformation produced one product which could not be readily identified. The fragmentation pattern showed the characteristic peaks for dichlorinated molecules, but the absence of the mass ion make identification difficult.

## 4.3.2.6 Aliphatic Substituted Aromatics.

It had been known from previous data that phenyl substituted aliphatic ring are potential substrates for the P-450cam Y96A mutant.<sup>13-15,21,22,25</sup> A range of derivatives were tested with different linkages to the aromatic 'handle', including keto, ester, and ether linkages. One of these compounds, 1-benzoyl-cyclohex-2,3-ene [41], was known to have a good binding affinity but the identification of the product had not been determined, table 4.9.

From the GC/MS data, 41 is oxidized to the epoxide, figure 4.11.

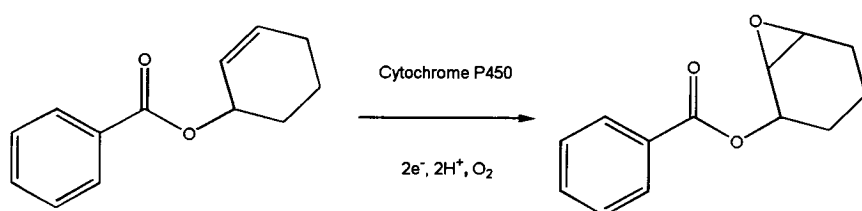
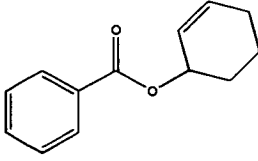
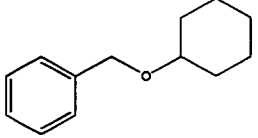
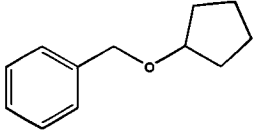


Figure 4.11:Hydroxylation of 41 by the P-450cam Y96A mutant.

This is consistent with the previous results, where hydroxylation on the 3 or 4 position was observed. Despite a binding constant better than camphor and high percentage high-spin, the yield for this substrate was very low with < 5 % of the reaction mixture being product, compared to camphor and adamantane which give 100 % product under the same conditions. An obvious reason for this can not be found since peroxide production would not be expected due to the high percentage high-spin. It might be possible that the product is acting as an inhibitor by binding too strongly to the heme thus reducing the efficiency of turn over.

Table 4.9: Biochemical data of compounds 41-43.

Compound	Structure	% high- spin	$K_D$ ( $\mu\text{M}$ )	GC peaks	Retention time and percentage in min, (% total)
41		52	1	starting material epoxide product	10.05 (84 %) 11.73 (16 %)
42		95	3	starting material hydroxy product	9.14 (86 %) 11.69 (14 %)
43		95	25	starting material hydroxy product	8.85 (99 %) 11.51 (<1 %)

Two other substrates were tested to examine whether a more flexible ether linkage would improve the substrate binding and product formation. Benzyl-oxy-cyclohexane [42] and benzyl-oxy-cyclopentane [43] were examined as potential substrates, table 4.9. As can be seen from the data both substrates bound to the enzyme, albeit with poor binding constants. What could not be understood from the data was why neither substrate appeared to produce product. An explanation could be that similar to 41 or the product yield is low and this might not be detected by our GC/MS assay.

#### 4.3.2.7 Aromatically Substituted Piperidines.

Hydroxylated piperidines are of potential interest as synthetic intermediates for the production of a range of pharmaceuticals. Previously, a range of aromatically substituted piperidines were found to be substrates for the P-450cam Y96A mutant.<sup>2</sup> All these piperidines are attached to the phenyl group through the piperidine nitrogen to avoid nitrogen ligation to the heme. Here we further explore their use as substrates for the mutant enzyme, in the hope of providing a greater insight into the effect of substitution on the piperidine ring and the nature of the linkage between the piperidine and phenyl rings.

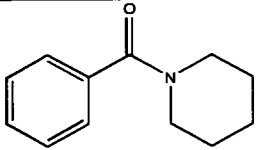
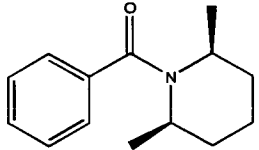
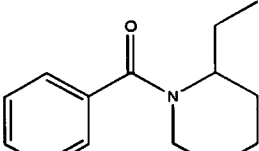
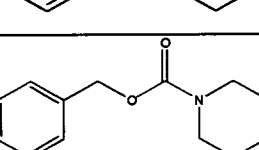
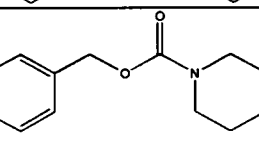
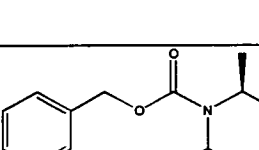
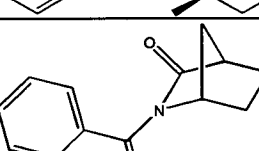
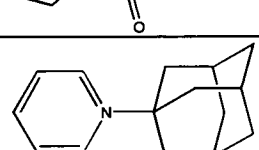
Three substrates contained the benzoyl group bound to an unsubstituted, dimethylated and ethylated piperidine ring (N-benzoylpiperidine [44], N-benzoyl-2,6-dimethylpiperidine [45] and N-benzoyl-2-ethylpiperidine [45], respectively). These substrates all bound to the enzyme, with the substituted piperidine rings showing greater affinity and displacement of the water, table 4.10.

All three substrates produced hydroxylated products with only 46 giving two products. However, for reasons discussed previously, this method only detects multiple products that are distinguishable by our GC/MS method. The yields of products were better here than for the similar aliphatic substrate (41), with 13 % for 44, 33 % for 45. For 46, the major peak was 37 % and the minor peak was 6 % of the total compounds present. It is clear therefore that with these substrates increased steric bulk on the piperidine ring enhances all aspects of the binding and turn over to the mutant enzyme.

Initial modelling studies indicated that a longer linkage between the aromatic and piperidine rings might enhance substrate binding. Since the aromatic 'handle' was designed for the enzyme and not because of any pharmaceutical interest, it was decided to use the benzoylcarbonyl group (*Cbz*) linked to the piperidine nitrogen, which could easily be cleaved from any hydroxylated product.



Table 4.10: Biochemical data of compounds 44-51.

Compound	Structure	% high-spin	K <sub>D</sub> (μM)	GC peaks	Retention time and percentage in min, (% total)
44		76	>20 <sup>a</sup>	starting material hydroxy product	10.77 (87 %) 13.52 (13 %)
45		95	10	starting material hydroxy product hydroxy product	11.71 (63 %) 14.65 (33 %) 14.89 (4 %)
46		100	0.5	starting material hydroxy product hydroxy product hydroxy product	12.07 (57 %) 14.68 (1 %) 15.07 (5 %) 15.16 (37 %)
47		45	>20 <sup>a</sup>	starting material hydroxy product	11.45 (88 %) 14.64 (12 %)
48		90	>20 <sup>a</sup>	starting material hydroxy product  keto product	12.06 (64 %) 14.58 (36 %) 15.30 (<1 %)
49		55	40	starting material hydroxy product	12.65 (52 %) 15.51 (48 %)
50		Oxygen ligation	-	starting material	11.48
51		a	a	a	-

a-Compound was too insoluble for accurate readings

As before, three substrates with *Cbz* substituted piperidines were studied, with an unsubstituted, methyl and dimethyl piperidine ring (*N-Cbz*-piperidine, [47], *N-Cbz*-

4-methylpiperidine, [48], and *N*-Cbz-2,6-dimethylpiperidine [49]). The three substrates all bound to the enzyme, table 4.10, but with lower binding affinities than the benzoyl substituted piperidines. Only 48 generated a good percentage high-spin, with 47 and 49 being comparable to camphor, indicating a degree of motion. All three substrates produced products with a yields of 15 %, 33 % and 50 % for 47, 48, and 49 respectively.

Comparisons of the data for 44-49 gives us a further insight into the P-450cam mutant enzyme. For the linker, the longer *Cbz* group reduces the binding affinity for the enzyme but slightly enhances product yields, (possibly due to faster substrate release). Substitution of the piperidine ring enhances all aspects of the substrate towards the enzyme, the greater the degree of substitution the better the binding/turn over.

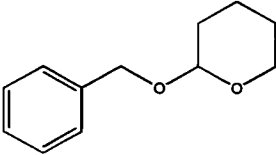
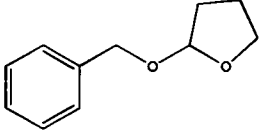
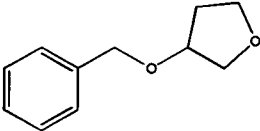
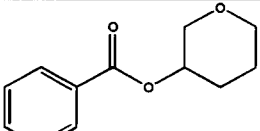
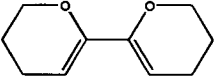
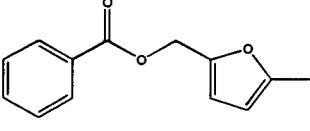
Two other phenyl substituted nitrogen heterocycles were studied here. *N*-benzoyl azabicyclo[2,2,1] pentan-2-one [50] was provided by colleagues at the University of Exeter and was of interest because the substrate was a phenyl substituted camphor analogue. Data, table 4.10, indicates that the substrate binds with oxygen ligation to the heme, with no product upon hydroxylation.

The other substrate studied was a substituted adamantane, adamantyl pyridinium (bromide salt) [51]. However, 51 proved difficult to handle due to high insolubility.

#### 4.3.2.8 Oxygen Heterocycles with Aromatic Sidechains.

The final category of substrates studied was a series of oxygen heterocycles. It was of interest to use cyclic acetals as an aromatic handle, as hydroxylation and subsequent cleavage of the acetal would yield diols. Two substrates tested were 2-benzyloxy-2,3,5,6-tetrahydropyran, [52] and 2-benzyloxy-2,3,4,5-tetrahydrofuran, [53]. Both substrates exhibited O ligation to the heme with relatively poor binding constants, table 4.11.

Table 4.11: Biochemical data of compounds 52-57.

Compound	Structure	% high-spin	$K_D$ ( $\mu\text{M}$ )	GC peaks	Retention time and percentage in min, (% total)
52		Oxygen ligation	40	starting material	11.74
53		Oxygen ligation	30	starting material	9.01
54		Oxygen ligation	40	starting material	8.85
55		Oxygen ligation	40	starting material	8.34
56		38	-	none	-
57		77	<sup>a</sup>	none	-

a-Compound was too insoluble for accurate readings

b-Compound could not be analyzed due to excessive evaporation

The isomer of **53**, 3-benzyloxy-2,3,4,5-tetrahydrofuran, [**54**], gave a similar effect, with a slightly higher binding constant. When the linker was changed to a benzoyl group as in 3-benzoyloxy-2,3,4,5-tetrahydrofuran, [**55**], again oxygen ligation was observed with an improvement binding. It would appear therefore that with an available free oxygen, ligation to the heme is observed, with the orientation of the oxygen and nature of the linker only affecting the binding affinity.

It was of interest to examine if an alkyl substitution next to the heterocyclic oxygen would prevent heme ligation. A substrate of this type was commercially

available, 5-methyl-2-furylmethylbenzoate [57]. This substrate did indeed prevent oxygen ligation to the heme with a percentage high-spin better than many of the previously studied substrates. However no product was detected in the biotransformation. The reason for this could not be determined but may be due to positioning of the methyl group near the heme, which would provide the active species with an unfavourable primary carbon to hydroxylate.

Another approach to preventing heme ligation was to use a substrate with rigid linkage to the 'handle'. One substrate was commercially available that could fall into this category, 6,6-bi-3,4-dihydro-2H-pyran, [56]. This compound was confirmed to be a substrate, table 4.11, although due to evaporation binding constants and products could not be determined.

#### 4.4 Conclusions.

These studies on the cytochrome P-450cam Y96A mutant have revealed some patterns in the properties of its substrates.

All wild-type substrates would appear to be substrates for the mutant enzyme indicating no loss in binding site volume. All these substrates undergo biotransformations with poor stereoselectivity.

The use of an aromatic ring in the substrate enhances the binding. If the other part of the substrate is an aliphatic ring, the degree of substitution and the linkage between the aromatic handle and the aliphatic ring are important. Increasing amounts of substitution on the aliphatic ring increases binding. However, if an oxygen group is present in the aliphatic ring the binding of the oxygen to the heme centre prevents turn over. Flexible linkers (e.g. ethers) and long linkers (e.g. *cbz*) reduce the binding affinity of the mutant enzyme. For the greatest binding, the benzoyl group should be used preferentially attached to an aliphatic ring with aliphatic side-chains.

In order to understand the binding properties of these substrates it is necessary to examine the substrate/enzyme interactions in more detail. Molecular modelling and

structural analysis was performed to better understand the cytochrome P-450cam Y96A mutant binding site and how substrates interact with it (chapters 5-9).

## 4.5 References

- 1)Grogan, G. Exeter University, unpublished results.
- 2)Aitken, S. J. MSc Thesis *Cytochrome P-450 Monooxygenases: Their Application in Biocatalysis*; University of Edinburgh, 1997.
- 3)McIver, L.; Leadbeater, C.; Campopiano, D. J.; Baxter, R. L.; Daff, S. N.; Chapman, S. K.; Munro, A. W. *European Journal of Biochemistry* **1998**, *257*, 577-585.
- 4)Jenkins, C. M.; Waterman, M. R. *Biochemistry* **1998**, *37*, 6106-6113.
- 5)Zehnder, A. J. B.; Wuhrmann, K. *Science* **1976**, *194*, 1165-1166.
- 6)Li, S. Y.; Wackett, L. P. *Biochemistry* **1993**, *32*, 9355-9361.
- 7)Eble, K. S.; Dawson, J. H. *Biochemistry* **1984**, *23*, 2068-2073.
- 8)Geren, L.; Tuls, J.; O'Brien, P.; Millett, F.; Peterson, J. A. *Journal of Biological Chemistry* **1986**, *261*, 5491-5495.
- 9)Mitani, F.; Horie, S. *Journal of Biological Chemistry* **1969**, *66*, 139-149.
- 10)Dawson, J. H.; Andersson, L. A.; Sono, M. *Journal of Biological Chemistry* **1982**, *257*, 3606-3617.
- 11)Sligar, S. G. *Biochemistry* **1976**, *15*, 5399-5406.
- 12)Eble, K. S.; Dawson, J. H. *Journal of Biological Chemistry* **1984**, *259*, 4389-4393.
- 13)Wong, L. L.; Westlake, C. G.; Nickerson, D. P. *Structure and Bonding* **1997**, *88*, 175-207.
- 14)Stevenson, J. A.; Bearpark, J. K.; Wong, L. L. *New Journal of Chemistry* **1998**, *22*, 551-552.
- 15)Bell, S. G.; Rouch, D. A.; Wong, L. L. *Journal of Molecular Catalysis B-Enzymatic* **1997**, *3*, 293-302.
- 16)DeVoss, J. J.; Sibbesen, O.; Zhang, Z. P.; DeMontellano, P. R. O. *Journal Of the American Chemical Society* **1997**, *119*, 5489-5498.
- 17)Atkins, W. M.; Sligar, S. G. *Journal of Biological Chemistry* **1988**, *263*, 18842-18849.

- 18) Atkins, W. M.; Sligar, S. G. *Journal of the American Chemical Society* **1989**, *111*, 2715-2717.
- 19) Grogan, G.; University of Edinburgh, unpublished results.
- 20) White, R. E.; McCarthy, M. B.; Egeberg, K. D.; Sligar, S. G. *Archives Of Biochemistry and Biophysics* **1984**, *228*, 493-502.
- 21) Jones, N. E.; England, P. A.; Rouch, D. A.; Wong, L. L. *Journal of the Chemical Society-Chemical Communications* **1996**, 2413-2414.
- 22) Fowler, S. M.; England, P. A.; Westlake, A. C. G.; Rouch, D. R.; Nickerson, D. P.; Blunt, C.; Braybrook, D.; West, S.; Wong, L. L.; Flitsch, S. L. *Journal of the Chemical Society-Chemical Communications* **1994**, 2761-2762.
- 23) Harris, D.; Loew, G. *Journal of the American Chemical Society* **1993**, *115*, 8775-8779.
- 24) Mueller, E. J.; Loida, P. J.; Sligar, S. G. *Twenty-five Years of P-450cam Research*; DeMontellano, P.R.O. Ed.; Plenum Press, New York and London, 1995, pp 83-124.
- 25) England, P. A.; HarfordCross, C. F.; Stevenson, J. A.; Rouch, D. A.; Wong, L. L. *FEBS Letters* **1998**, *424*, 271-274.
- 26) Grayson, D. A.; Tewari, Y. B.; Mayhew, M. P.; Vilker, V. L.; Goldberg, R. N. *Archives of Biochemistry and Biophysics* **1996**, *332*, 239-247.
- 27) Castro, C. E.; Wade, R. S.; Belser, N. O. *Biochemistry* **1985**, *24*, 204-210.
- 28) Hur, H. G.; Sadowsky, M. J.; Wackett, L. P. *Applied and Environmental Microbiology* **1994**, *60*, 4148-4154.
- 29) Manchester, J. I.; Ornstein, R. L. *Protein Engineering* **1995**, *8*, 801-807.
- 30) Wackett, L. P.; Sadowsky, M. J.; Newman, L. M.; Hur, H. G.; Li, S. Y. *Nature* **1994**, *368*, 627-629.
- 31) Koe, G. S.; Vilker, V. L. *Biotechnology Progress* **1993**, *9*, 608-614.
- 32) Horowitz, J. B.; Vilker, V. L. *Biotechnology and Bioengineering* **1994**, *44*, 248-255.
- 33) Ornstein, R. L.; Paulsen, M. D.; Bass, M. B. *Abstracts of Papers of the American Chemical Society* **1991**, *202*, 118-BIOT.
- 34) Wackett, L. P. *Annals of the New York Academy of Sciences* **1998**, *864*, 142-152.

## **5 Crystallization of the P-450cam Y96A Mutant.**

---

### **5.1 Introduction.**

Crystals are highly ordered three dimensional arrays of molecules made up from a basic unit cell. Each unit cell is related to another by a well defined symmetry operation, known as the space group. The unit cell is repeated many thousands of times in a crystal, thus forming a regular lattice. Due to this regular arrangement of atoms, when X-ray radiation is focused through a crystal, the X-rays diffract to produce an ordered diffraction pattern. Using a number of mathematical processes, the diffraction pattern can be extrapolated to give the positions of the atoms in the unit cell and thus a structure can be determined.

For any protein larger than 30 kDa, X-ray crystallography has become the fundamental method to determine the structures at atomic resolution. Consequently, a great deal of time has been spent trying to develop new and faster techniques. Despite standard crystallization methodologies and advances in computer technology, the crystallization and subsequent X-ray structural determination of proteins can be a long cumbersome process.



The production of an X-ray structure can be broken up into six stages:

- Expression and purification of the protein, in sufficient amounts and of very high purity.
- Screening for crystallization conditions.
- Crystallization of protein.
- Data measurements.
- Phase determination.
- Electron density map computation and interpretation.
- Model refinement.

Each of these stages has its own associated problems. With improvements in computational methodologies, however, the generation of suitable protein crystals has become the bottleneck to crystallographic studies.

## **5.2 Protein Expression and Purification.**

Protein purification protocols for P-450cam have been established (chapter 2). A scale-up allowed for the production of sufficient protein from one set of purification runs for many crystal screening experiments, thus maintaining consistency in the protein batch used in different screens. The extra purification step for the P-450cam gave protein of greater purity than can be determined by SDS PAGE, which could yield a better quality crystal.

## **5.3 Crystallization Screens.**

P-450cam was first crystallized in 1974,<sup>1</sup> for purification studies and not for use in X-ray structural determination. Cytochrome P-450cam crystals for structural studies were first made in 1982.<sup>2</sup> Three types of crystals were produced, of differing unit cell dimensions and space group. One crystal form, orthorhombic II, was shown

to have better diffraction than the other two, diffracting to 3.0 Å. This crystal form was used in the determination of the wild-type structure<sup>3</sup> which later was refined to 2.6 Å.

Crystallization conditions for the P-450cam protein have now been standardized,<sup>4</sup> which has led to 11 different substrate bound structures and two mutant structures.<sup>3-15</sup> Crystals were grown by the free interface diffusion technique with the protein solution layered over a high concentration ammonium sulfate solution in a narrow bore glass tube. Crystals of cytochrome P-450cam appear at the interface between the two phases. DTT is added to the protein solution to prevent dimerization caused by disulfide bond formation through free cysteine C334. After crystallization, DTT can be removed by washing the crystals. Ligands are then soaked, over a number of weeks, into the crystal. Although this technique has been used for over ten years there are problems of reproducibility.<sup>4</sup> P-450cam crystallization has been shown to be sensitive to protein concentration, ammonium sulfate concentration and temperature.<sup>4</sup>

As discussed previously, the Y96A protein studied here is, in fact, a double mutant, where cysteine 334 has been mutated to an alanine (C334A). The removal of the free cysteine facilitates handling of the protein in solution, negating the need for reducing agents such as DTT or mercaptoethanol to prevent dimerization. This is beneficial for the crystallization studies as there is no need to co-crystallize the protein with the reducing agents and then consequently remove them. This means that the conditions for crystallizing wild-type protein will possibly not apply to this double mutant, hence the need to screen for crystallization conditions.

### 5.3.1 Using Wild-Type Conditions.

Initial crystallization conditions for the wild-type protein were followed.<sup>2</sup> This consisted of a screen of ammonium sulfate concentration vs. pH. Ammonium sulfate was screened in 6 x 7 % increments from 35 % to 70 % ammonium sulfate at a pH of 7.4 and 7.0, with a protein concentration of 20 mg/ml and temperature of 4 °C. All screens were conducted using vapor diffusion by the hanging drop method. For all the

wells of ammonium sulfate concentration greater than 42 %, protein precipitated almost immediately (10 - 20 minutes). In the other wells no crystallization occurred. This was repeated with higher protein concentration (30 mg/ml) with the same result. Rapid crystallization does not lead to the formation of crystals suitable for crystallography, therefore no further attempts to use ammonium sulfate as a crystallizing agent were performed.

### 5.3.2 Development of New Conditions.

Polyethylene glycol (PEG) was then screened as a potential precipitant to examine conditions which produced crystals slower than the conditions using ammonium sulfate.

#### 5.3.2.1 pH Dependence.

The pH was not varied because literature on the P-450cam wild-type enzyme suggests that protein crystallizes within the range of pH that the protein is stable (~pH 7.4).<sup>4</sup> Small changes (i.e. 7.2-7.4) to pH do not affect the crystallization, however larger changes in pH degrade the protein.<sup>16</sup>

#### 5.3.2.2 Substrate Co-crystallization Effects.

Crystallization studies were initially performed, in parallel, using wells both with and without substrate. At camphor concentrations of 4 mM no appreciable difference was noticed in the crystallization pattern. At high substrate concentrations, precipitation of the substrate resulted in premature precipitation of the protein. It is known that camphor confers stability upon P-450cam, especially at higher temperatures, therefore it was thought better to perform all initial screens in the presence of substrate. Camphor was chosen because it has reasonable aqueous solubility and was readily available.

## 5.3.2.3 PEG and Protein Concentration.

Crystallization screens were performed with varying PEG size (weight 2000-20000) and concentration (10-30 %) and with different protein concentrations (5-80 mg/ml). The initial results of these screens are summarized in table 5.1.

Table 5.1: Crystallization screens of PEG.

Protein concentration/ Precipitant	10 mg/ml	20 mg/ml	30 mg/ml
PEG (MWt 2000)	no crystals	no crystals	no crystals
PEG (MWt 4000)	crystals formed (26-28% PEG)	not done	crystals formed (22-26% PEG)
PEG (MWt 6000)	crystals formed (22-28% PEG)	crystals formed (14-30% PEG)	crystals formed (14-22% PEG)
PEG (MWt 8000)	crystals formed (22-28% PEG)	crystals formed (14-30% PEG)	crystals formed (14-22% PEG)
PEG (MWt 10000)	crystals formed (22-28% PEG)	crystals formed (14-26% PEG)	precipitation
PEG (MWt 12000)	crystals formed (22-28% PEG)	crystals formed (14-26% PEG)	precipitation
PEG (MWt 20000)	crystals formed (20-24% PEG)	crystals formed (14-30% PEG)	precipitation

N.B. Screens performed at lower protein concentration yielded no crystals.

It was clear from preliminary results that a relationship between the PEG size and concentration exists with the protein concentration. Crystals only grew when, either the protein concentration, or the PEG concentration was sufficiently high, but low enough not to cause precipitation. A variety of crystal forms were observed for P-450cam (figure 5.1), the predominant crystal form present in the crystallization trials was a medium sized crystal of a wedge like nature figures 5.1-ii, 5.2-d, e. These were found to produce diffraction of sufficient size to yield basic structural information (resolution  $<2.8 \text{ \AA}$ ) but exhibited a considerable amount of twinning.

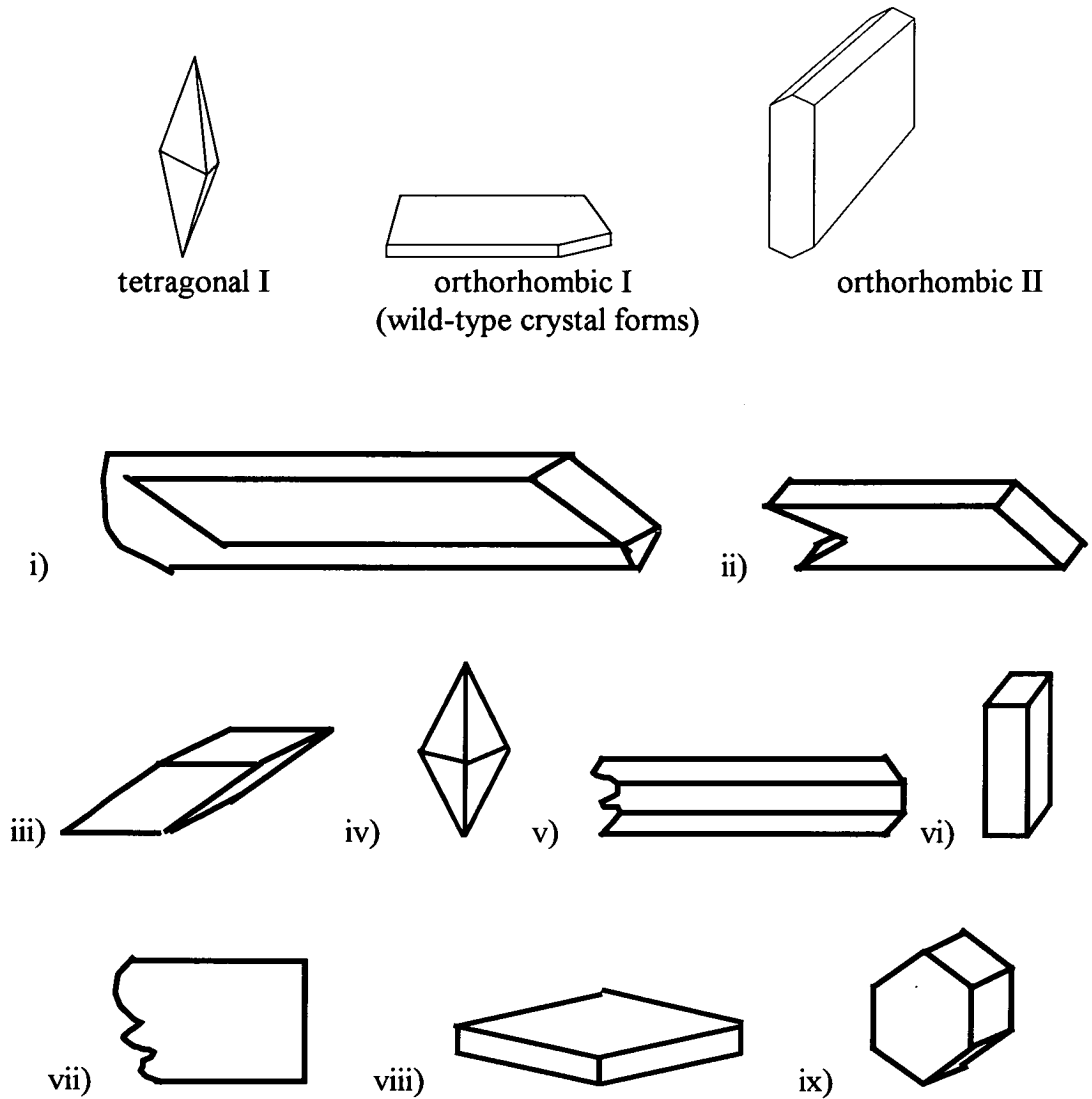


Figure 5.1: Structure of the crystals previously observed <sup>2</sup> for wild-type P-450cam compared to the new Y96A mutant crystal forms (i-ix).

After conditions were established for the production of crystals it was necessary to optimize these conditions, to produce crystals of sufficient size and quality for good diffraction. It was found that crystals that grew in the larger PEG solutions were small and of poor diffraction quality. The best crystals grew in PEG (MWt 6000) at a concentration between 14 % and 20 % (figure 5.2-d, e), with a protein concentration of 20 mg/ml. These conditions produced 3-10 medium sized crystals per well.

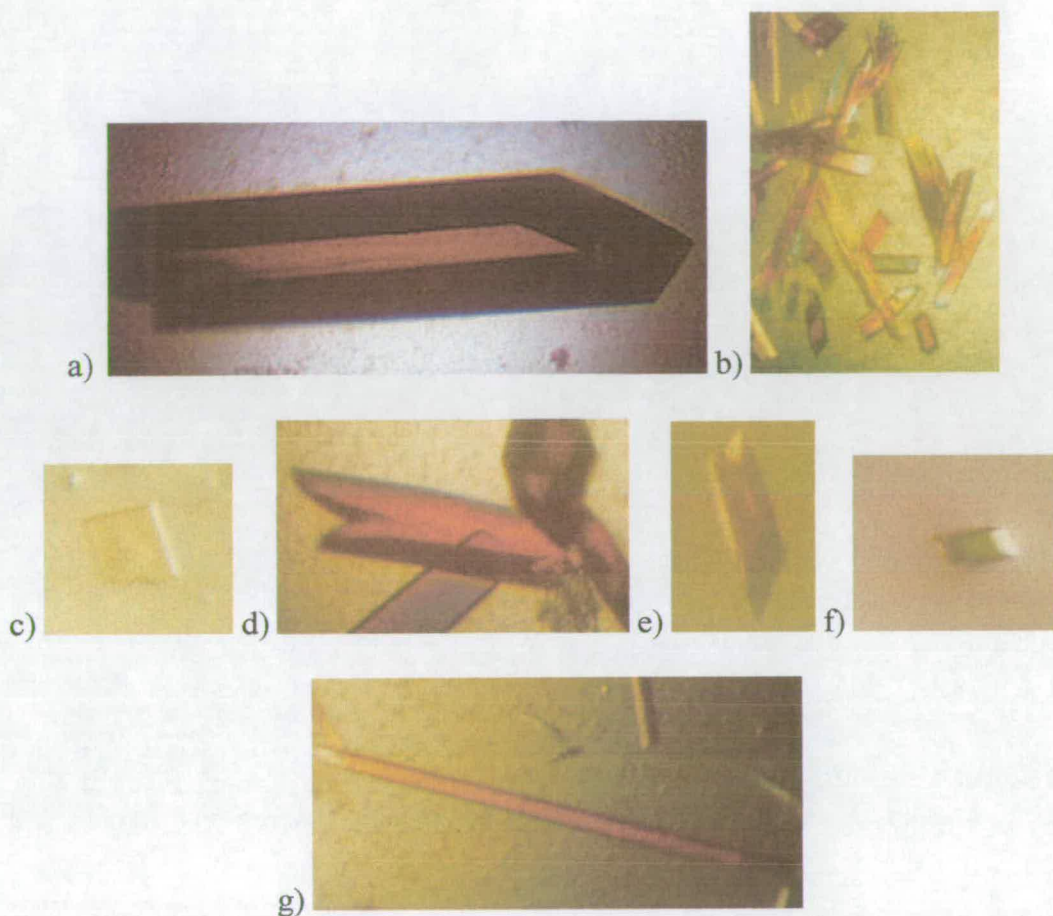


Figure 5.2: Photographs of the different crystal types of P-450cam Y96A C334A mutant.

At 14 % PEG (MWt 6000) larger crystals grew (figure 5.2-a), but the diffraction was poor due to the splitting of the diffraction signal by fractures in the crystal. At higher than 20 % PEG (MWt 6000) smaller crystals grew in about 48 hours (figure 5.2-b, f). For most other PEG conditions needles formed or the crystals were small and not uniform. In one case for 24 % PEG (MWt 10000) rose coloured plates formed (figure 5.2-c), possibly similar to orthorhombic I crystals previously described.<sup>2</sup> Crystals from the 14 and 18 % PEG (MWt 6000) wells with protein concentration of 20 mg/ml were used in further X-ray crystallography studies, figure 5.2-d, e.

Unfortunately crystals grown at the 'optimal' conditions for growth, produced poor diffraction with considerable amounts of twinning. This was possibly due to the length of time it took to produce these crystals, (about 6 weeks). It was decided to find a way to speed up the crystal formation whilst maintaining good quality crystals. At higher PEG concentrations, small crystals formed suggesting that it was better to increase the protein concentration and not the PEG concentration.

With further screening, it was found that 45 mg/ml protein induces faster crystallization without loss of crystal quality. Higher concentrations of protein than this resulted in precipitation and denaturing of the protein. Once it was established that 45 mg/ml protein could give the desired speed of crystal growth, the well conditions were reoptimized based on this protein concentration.

#### 5.3.2.4 Temperature.

Initially the temperature was kept constant at 4 °C throughout the screening, as P-450cam is unstable at room temperature in an oxygen atmosphere. However, stable crystals can form at higher temperatures, therefore screening was performed at a variety of different temperatures. It was found that crystals formed more slowly at 4 °C than 14 °C. However at 18 °C the quality of the crystals formed was not sufficient for data collection, probably due to protein degradation. The difference between plate incubation temperature and crystal quality at 4 and 14 °C was found to be variable so both temperatures were used in crystal generation.

#### 5.3.2.5 Substrate Concentration.

After suitable conditions were found to induce crystallization, the effects of the substrate were explored. At 45 mg/ml (1 mM) to maintain a substrate to enzyme ratio of 1:1, at least 2 mM substrate is required. This caused problems with some of the substrates because of their poor aqueous solubility. However, so long as the substrate did not precipitate, crystal formation was essentially unaffected.

## 5.4 Final Crystallization Conditions.

PEG 6000 and PEG 4000 gave good quality crystals between 14-26 % PEG concentration (figure 5.2-g). PEG 4000 was the preferred choice since its lower viscosity meant it was easier to handle in solution. Final conditions used six wells with PEG concentrations from 14-26 %, at both 4 °C and 14 °C.

## 5.5 Data Collection.

With the selection of good crystals only a number of small problems exist with the data collection of the mutant crystals. Even with cryo-freezing P-450cam crystals degrade over time, usually ~50-75 hours, on the in-house rotating anode generator. These time limits are such that collection of sufficient frames for complete processing was not always possible. The only other problem with the data collection of the crystals is the crystal shape. As mentioned previously (chapter 2) the crystals that give the best diffraction are long needles, this means that with the long axis of the crystal lying in the plane of the loop, the X-ray beam passes through the small axes. Thus, to get sufficient diffraction for structural determination large forms of these needles are required.

## 5.6 Phase Determination.

The phase for the first structure, 4-hydroxydiphenylmethane, was solved by molecular replacement, by Paul Taylor, University of Edinburgh.



## 5.7 The New Unit Cell.

With the new crystallization conditions and the use of the C334A mutant, the P-450cam crystallized with new unit cell parameters. This new crystal form has cell angles close to orthorhombic, however, refinement in the orthorhombic space group gave significantly higher  $R_{\text{sym}}$  values than the corresponding monoclinic refinement. Comparison of the unit cell dimensions of other forms of wild-type P-450cam (table 5.2) show that the double mutant protein has a different number of molecules per unit cell and significantly different unit cell dimensions.

Table 5.2: Summary of the unit cells of known P-450cam crystals.

Cell dimensions Crystal type	a (Å)	b(Å)	c(Å)	Space group	No. molecules per unit cell
Wild-type orthorhombic I <sup>2</sup>	174	64	39	P2 <sub>1</sub> 2 <sub>1</sub> 2	4
Wild-type orthorhombic II <sup>2§</sup>	108.73	104.46	36.44	P2 <sub>1</sub> 2 <sub>1</sub> 2 <sub>1</sub>	1
Wild-type tetragonal I <sup>2</sup>	64.2	64.2	255	P4 <sub>1</sub> 22/P 4 <sub>3</sub> 22	8
Y96A C334A mutant orthorhombic III	~62	~61	~96	P2 <sub>1</sub>	4

The unit cell is approximately orthorhombic with angles of 90.00, 91.24, 90.00 °. The cell dimensions are, for example: a = 62.252 Å, b = 61.508 Å and c = 95.657 Å for the 4-hydroxydiphenylmethane structure and the protein belongs to the space group P2<sub>1</sub>. There is only a slight difference between the unit cell dimensions for the other Y96A structures solved here, table 5.3, 5.4, 5.5.

<sup>§</sup> The crystal properties of the orthorhombic II species were recalculated in 1985 to higher resolution 3.

### 5.7.1.1 Differences in the Asymmetric Units.

The two molecules in the asymmetric unit were compared and found to have a root mean squared difference (rmsd) of 0.27 Å for the backbone and 0.45 Å for the side chain atoms (based upon comparison of the units of the camphor bound structure). Residues with rmsd over 1 Å were all surface residues with long, disordered side chains. Average temperature factors between the units differed by rmsd of 1.7 Å<sup>2</sup>. The active site showed no other significant changes in topology, the largest difference being less than 0.1 Å. As in previous P-450cam structures, the assignment of residues 1-9 was not possible due to poor electron density in this region. The summary of the data diffraction statistics are presented in tables 5.3-5.5.

To check the validity of the data, the files were examined by Procheck. Of the information examined the most important is the Ramachandran plot, which shows the  $\phi$  and  $\psi$  angles for every residue in relation to idealized values shown in the plot (based on a large selection of previously solved crystal structures), figure 5.3.

## 5.8 Conclusions.

Reproducible crystallization conditions have been produced using PEG as the precipitant, with the protein at a concentration of 1 mM. These conditions gave co-crystallized mutant crystals that diffract to a resolution of better than 2.4 Å and in some cases to 2.0 Å. The crystals diffracted in a new crystal form with the monoclinic space group P<sub>21</sub>. Although the crystallization conditions are efficient at producing good quality crystals, a problem does exist. At this protein concentration substrates have to be soluble at concentrations of significantly greater than 1 mM. For some of the crystals the presence of possibly poorly soluble substrates clearly reduced the quality of the crystals (e.g. diphenylmethane and *N*-cbz-2,4-dimethylpiperidine). The other problem with poor aqueous solubility is that the binding site might not be fully occupied. This is observed for some of the substrates studied here (see chapter 6). For insoluble substrates it might be worth considering substrate soaks as a means of obtaining the substrate bound structure.

Table 5.2: Data statistics for 4-hydroxydiphenylmethane and camphor bound structures.

	4-hydroxydiphenylmethane	camphor
resolution	2.00	2.00
unit cell	94.51, 61.70, 66.74 90.00, 90.28, 90.00	96.45, 62.78, 66.95 90.00, 90.28, 90.00
no. molecules per asymmetric unit	2	2
total observations	81735	227193
unique reflections	50447	53664
$R_{\text{sym}}$ (overall/ last shell)	6.0/26.1	7.3/ 23.7
% completeness (total)	96.7	96.5
% completeness (last shell)	91.1 (2.03-2.00)	81.7 (2.03-2.00)
reflections used for refinement	45089	50587
$R_{\text{factor}}$	17.39	16.02
$R_{\text{free}}$	24.78	20.74
overall average B	21.6	17.7
no. water molecules	955	1197

$$R_{\text{sym}} = \frac{\sum |I - \langle I \rangle|}{\sum I}$$

$$R_{\text{factor}} = \frac{\sum ||F_{\text{obs}}| - |F_{\text{calc}}||}{\sum |F_{\text{obs}}|}$$

5 % randomly omitted reflections were used for  $R_{\text{free}}$ .

Table 5.2: Data statistics for diphenylmethane and phenylcyclohexane bound structures.

	diphenylmethane	phenylcyclohexane
resolution	2.34	2.54
unit cell	95.32, 62.51, 67.09 90.00, 90.42, 90.00	96.24, 63.04, 66.57 90.00, 90.41, 90.00
no. molecules per asymmetric unit	2	2
total observations	121600	171575
unique reflections	41342	32398
$R_{\text{sym}}$ (overall/ last shell)	9.3/ 21.3	10.7/ 48
% completeness (total)	92.2	98.1
% completeness (last shell)	88.2	90.2
reflections used for refinement	39137	39137
$R_{\text{factor}}$	21.16	17.82
$R_{\text{free}}$	29.79	27.65
overall average B	21.2	32.2
no. water molecules	1871	1906

$$R_{\text{sym}} = \frac{\sum |I - \langle I \rangle|}{\sum I}$$

$$R_{\text{factor}} = \frac{\sum ||F_{\text{obs}}| - |F_{\text{calc}}||}{\sum |F_{\text{obs}}|}$$

5 % randomly omitted reflections were used for  $R_{\text{free}}$ .

Table 5.2: Data statistics for diphenylmethane and phenylcyclohexane bound structures.

	bicyclohexyl	<i>N</i> -cbz-2,4-dimethylpiperidine
resolution	2.30	2.30
unit cell	95.26, 62.30, 67.06 90.00, 90.40, 90.00	95.84, 62.66, 66.78 90.00, 90.41, 90.00
no. Molecules per asymmetric unit	2	2
total observations	338642	214266
unique reflections	35358	34751
$R_{\text{sym}}$ (overall/ last shell)	13/ 36	8.4/ 37.8
% completeness (total)	99.9	98.5
% completeness (last shell)	99.9	97.0
reflections used for refinement	33286	34322
$R_{\text{factor}}$	18.71	17.0
$R_{\text{free}}$	27.85	25.95
overall average B	24.3	24.3
no. water molecules	925	1001

$$R_{\text{sym}} = \frac{\sum |I - \langle I \rangle|}{\sum I}$$

$$R_{\text{factor}} = \frac{\sum ||F_{\text{obs}}| - |F_{\text{calc}}||}{\sum |F_{\text{obs}}|}$$

5 % randomly omitted reflections were used for  $R_{\text{free}}$ .

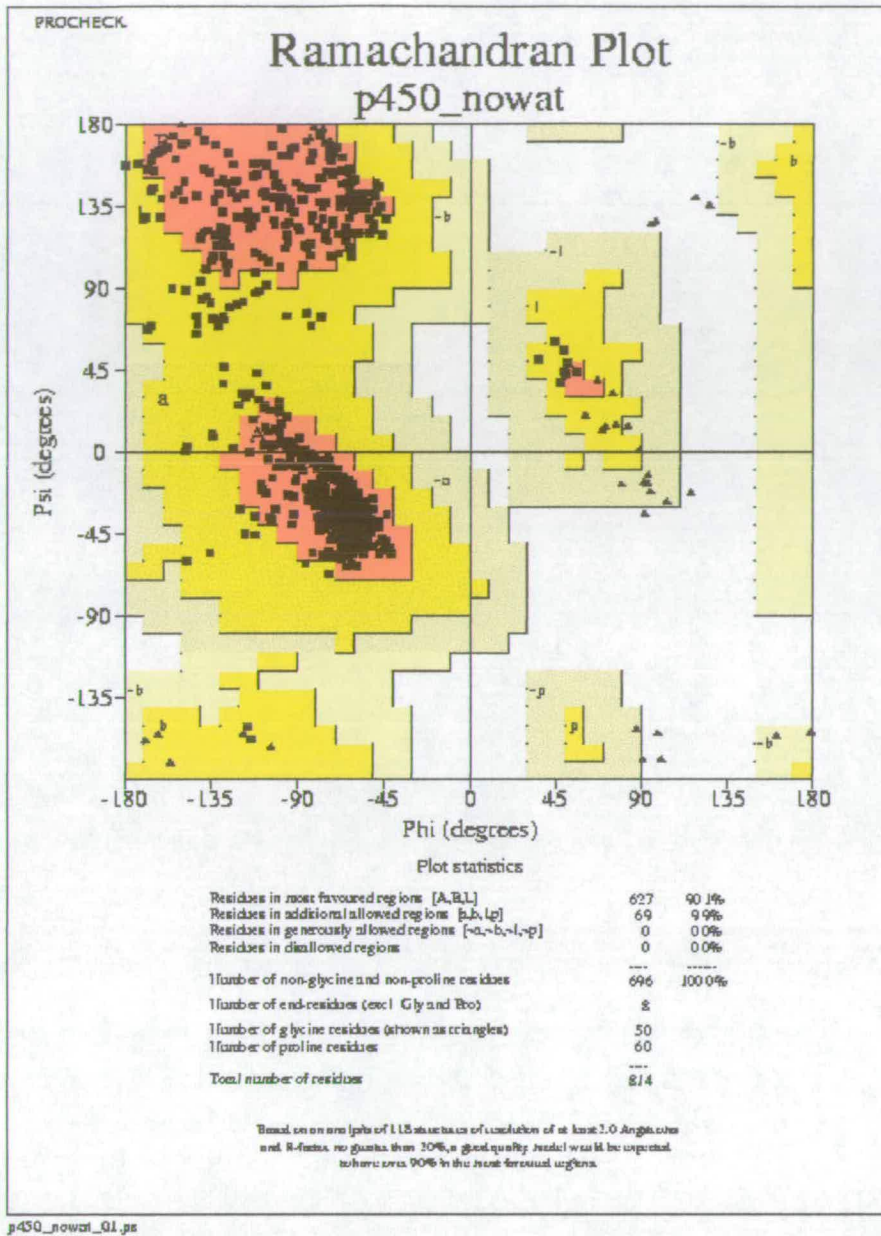


Figure 5.3: Ramachandran plot of the camphor bound structure.

## 5.9 References.

- 1) Yu, C-A.; Gunsalus, I. C. *Journal of Biological Chemistry* **1974**, *249*, 107-110.
- 2) Poulos, T. L.; Perez, M.; Wagner, G. C. *Journal of Biological Chemistry* **1982**, *257*, 10427-10429.
- 3) Poulos, T. L.; Finzel, B. C.; Gunsalus, I. C.; Wagner, G. C.; Kraut, J. *Journal of Biological Chemistry* **1985**, *260*, 16122-16130.
- 4) Poulos, T. L. *Methods in Enzymology*. **1996**, *272?*, 358-368.
- 5) Li, H. Y.; Narasimhulu, S.; Havran, L. M.; Winkler, J. D.; Poulos, T. L. *Journal of the American Chemical Society* **1995**, *117*, 6297-6299.
- 6) Poulos, T. L.; Finzel, B. C.; Howard, A. J. *Biochemistry* **1986**, *25*, 5314-5322.
- 7) Poulos, T. L.; Howard, A. J. *Biochemistry* **1987**, *26*, 8165-8174.
- 8) Poulos, T. L.; Finzel, B. C.; Howard, A. J. *Journal of Molecular Biology* **1987**, *195*, 687-700.
- 9) Poulos, T. L.; Cupp-Vickery, J.; Huiying, L. *The Structure of Cytochrome P450*; 2nd ed.; Demontellano, P. R. O. Ed.; Plenum Press, New York and London, 1995, pp 125-150.
- 10) Raag, R.; Poulos, T. L. *Biochemistry* **1989**, *28*, 7586-7592.
- 11) Raag, R.; Swanson, B. A.; Poulos, T. L.; Demontellano, P. R. O. *Biochemistry* **1990**, *29*, 8119-8126.
- 12) Raag, R.; Martinis, S. A.; Sligar, S. G.; Poulos, T. L. *Biochemistry* **1991**, *30*, 11420-11429.
- 13) Raag, R.; Poulos, T. L. *Biochemistry* **1991**, *30*, 2674-2684.
- 14) Raag, R.; Li, H. Y.; Jones, B. C.; Poulos, T. L. *Biochemistry* **1993**, *32*, 4571-4578.
- 15) Vidakovic, M.; Sligar, S. G.; Li, H. Y.; Poulos, T. L. *Biochemistry* **1998**, *37*, 9211-9219.
- 16) Diprimo, C.; Hoa, G. H. B.; Douzou, P.; Sligar, S. G. *European Journal of Biochemistry* **1992**, *209*, 583-588.

## **6 Crystal Structure of the Cytochrome P-450 Y96A-C334A Mutant.**

---

### **6.1 Introduction.**

The analysis of the crystal structure of the cytochrome P-450cam double mutant is subdivided into a number of categories. Firstly, the effects of the two mutations are discussed; the C334A mutation does not effect the Y96A site so the effect of each mutation is discussed separately. The orientation of each substrate is then discussed and comparisons are made to previous structures or the other substrate structures studied here. Finally a more in-depth discussion of the most important features of the structures is presented.

### **6.2 Effects of the C334A Mutation.**

Cysteine 334 is located on a long loop between the L helix and the  $\beta$ 4 sheet of P-450cam with the free cysteine on the surface, figure 6.1 (nomenclature based on original crystal structure <sup>1</sup>).



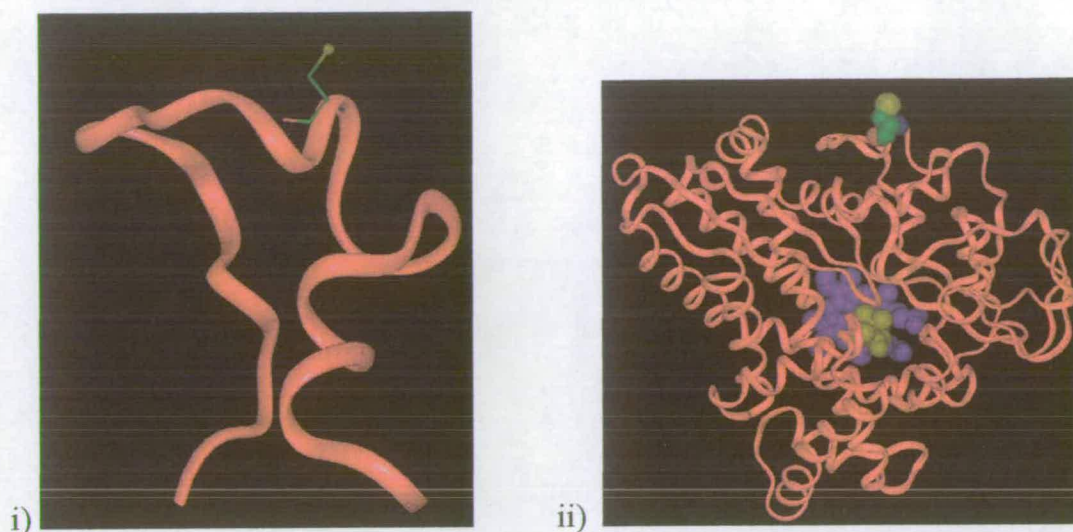


Figure 6.1: i) location of cysteine 357 on the L helix (cysteine is in yellow) ii) cysteine 357 in relation to the whole protein (cysteine is in green, heme in blue and camphor in yellow).

This surface cysteine leads to dimerization of the enzyme in solution.<sup>2</sup> Mutagenesis of the cysteine to alanine shows no effect on the turnover rates of camphor. However, until now the structural effect of this mutation remained unknown.<sup>2,3</sup>

Analysis of the region surrounding the C334A mutation indicates no visible change in the secondary or tertiary structure. Sigma weighted difference maps, contoured at a  $3\sigma$  level, showed no density in this region. When compared to the wild-type cryo-structure (5cp4<sup>4</sup>), the rmsd for the side-chains of the surrounding residues (Ala 333, Pro 335, His 337) is 0.40 Å compared to the total protein rmsd of 0.61 Å. The mutant alanine at 334 adopts the same backbone configuration as the wild-type cysteine with a rmsd of 0.26 Å compared to total protein backbone differences of 0.29 Å. The overall temperature factors of side-chains in this region differ by rmsd of 3.0 Å<sup>2</sup>, compared to the average protein value of 6.1 Å<sup>2</sup>. These results indicate that the C334A mutation has little effect on the protein structure or residue mobility.

### 6.3 Effects on the Binding Site by the Y96A Mutation.

To aid the description of the binding site a number of cartoons have been provided, figures 6.2 and 6.3.

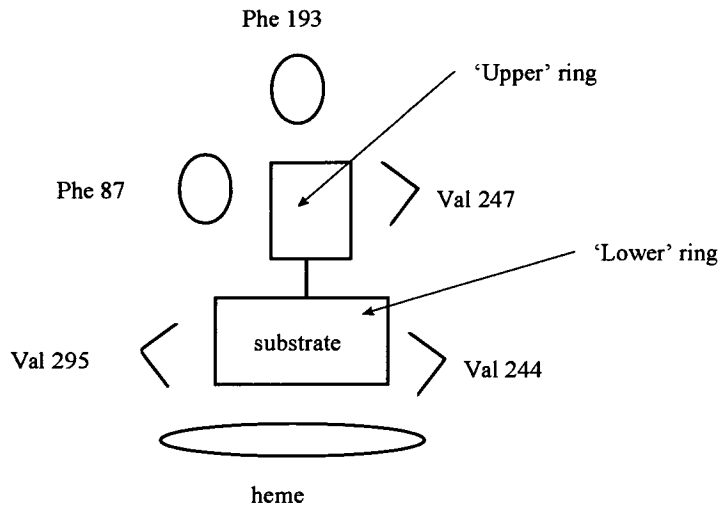


Figure 6.2: Schematic of the P-450cam-Y96A binding site showing important residues mentioned in the following chapter. Picture shows the method of defining the different substrate rings with the term 'upper' and 'lower'.

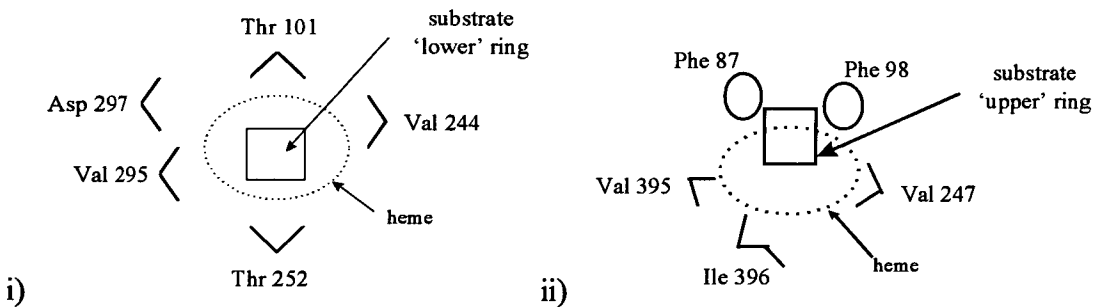


Figure 6.3: Orientation of the P-450cam Y96A binding site showing the orientation of the i) 'lower' and ii) 'upper' rings of the substrate.

### 6.3.1 Effects on the Y96A Binding Site Based on the Camphor Bound Structure.

For the camphor structure, compared to the wild-type camphor bound cryo-structure (5cp4), the total rmsd of the binding site side-chains is less than 0.67 Å, compared to total protein side-chain rmsd of 0.61 Å (residues 87, 98, 101, 185, 244, 247, 252, 297, 395, 396 and heme). This would indicate that there has been no 'significant' alteration of the binding site, other than by the mutation itself. The sigma weighted difference map of the Y96A protein compared to the wild-type confirms the absence of the tyrosine residue caused by the mutation, figure 6.4.

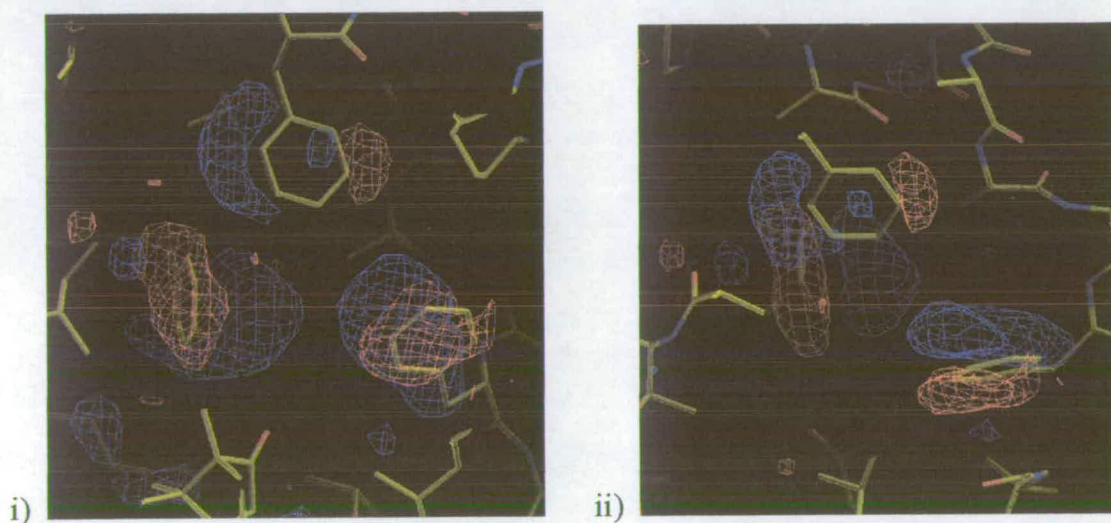


Figure 6.4:  $\sigma$ -weighted difference map showing the result of modelling the wild-type structure (with the tyrosine 96 to alanine change). i) shows, clockwise, camphor, Phe 87, Phe 198 and Phe 98. ii) shows more clearly the Phe 193 (top) and Phe 98 density. The colour code is positive density in blue and negative density in red.

The difference map also shows that the space left by the tyrosine has some positive density in it (blue) and the Phe 87 residue has some corresponding negative density (red) on the phenyl ring. This implies that the Phe 87 residue is incorrectly positioned in its current (wild-type) position and should, in fact, be moved into the space previously occupied by Tyr 96. This map additionally shows that, to a lesser



degree, residues Phe 98 and Phe 193 are also incorrectly positioned in their wild-type positions. Comparison of the final refined structure shows these differences more clearly, figure 6.5.

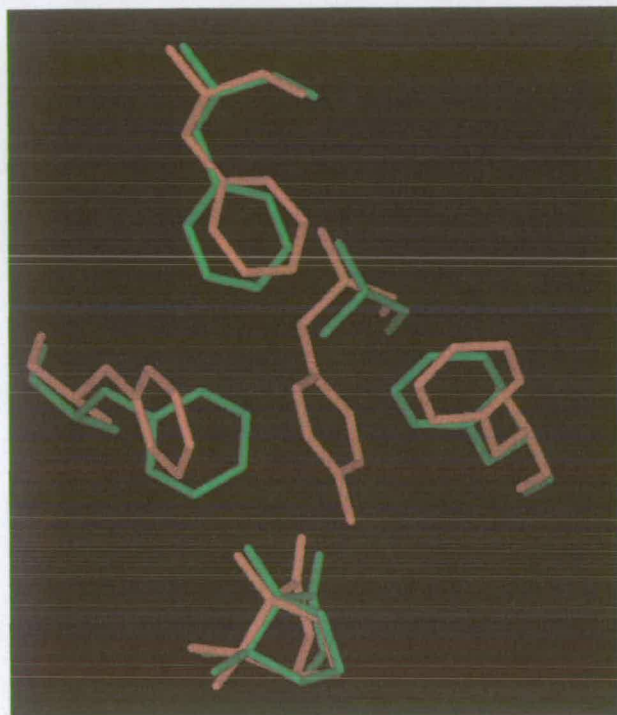


Figure 6.5: Comparison of the wild-type (red) to the Y96A (green) camphor bound structures. Residues are clockwise: Camphor, Phe 87, Phe 193, Phe/Ala 96, Phe 98.

The Phe 87 residue has undergone a  $57^\circ$  rotation and  $1.8 \text{ \AA}$  movement towards the space left by the removal of the tyrosine residue; in addition to this Phe 98 has also shifted  $0.98 \text{ \AA}$ , into this gap. Two residues have moved as a consequence of the Phe 87 movement, Phe 193 has shifted  $1.3 \text{ \AA}$  and rotated  $30^\circ$  towards the wild-type Phe 87 site and the isoleucine 395 CG carbon rotates  $109^\circ$  towards the wild-type Phe 87 site. These changes to this region of the protein will be discussed in more detail later (section 7.2.1).

The only other binding site change is Thr 101 which, in the wild-type structure, is hydrogen bonded to the tyrosine 96 hydroxyl. With the removal of the

tyrosine and thus the hydrogen bond the Thr 101 has adopted a new configuration, by moving  $0.9 \text{ \AA}$  away from the heme centre and rotating  $104^\circ$ , figure 6.6.

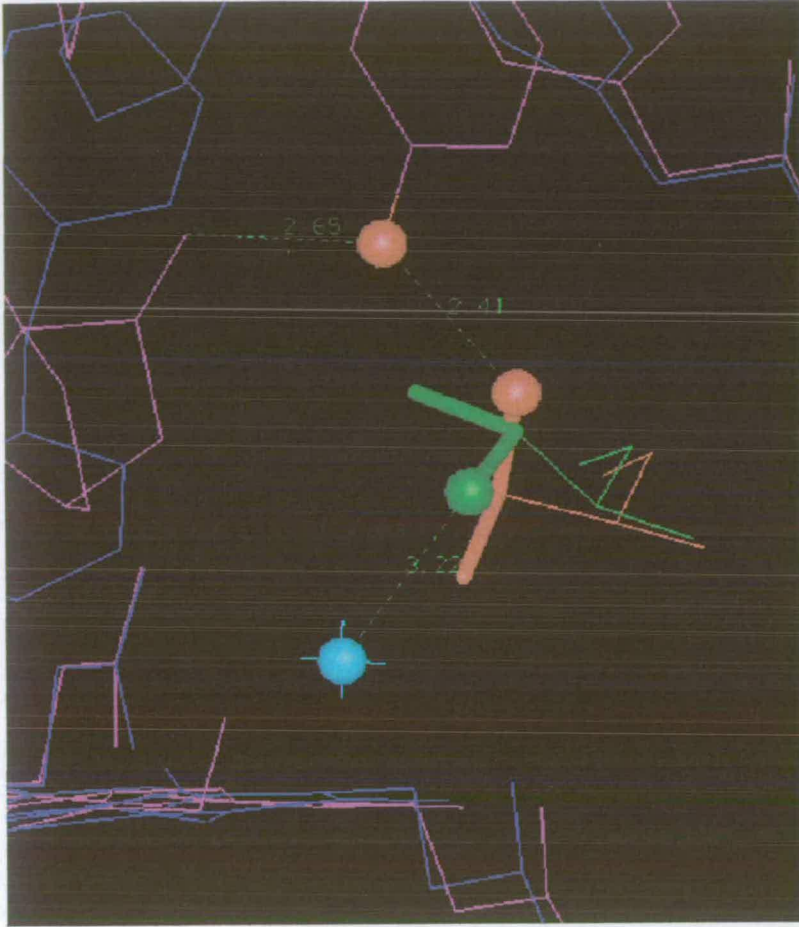


Figure 6.6: Comparison of Thr 101 for the wild-type (red) with the Y96A mutant (green). The Tyr 96-Thr 101 hydrogen bond for the wild-type protein and the new water molecule (cyan) are also shown.

This movement has allowed an extra water molecule to enter the binding site which H-bonds to the newly positioned Thr 101 hydroxyl (OG1), the heme oxygen (O2D) and a water molecule. This new water is sufficiently removed from the heme active centre ( $7.2 \text{ \AA}$ ) that it is unlikely to interfere with the catalytic mechanism of the heme centre, although its presence may alter the relative hydrophobicity of the binding site.

### 6.3.1.1 Modelling the Iron-Sulfur Bond.

The iron-sulfur distance is potentially important in the increasing debate over the cause of the Soret shift (see section 7.4.2) therefore careful analysis of it was made. Variations in the starting distance did not affect the final refined iron-sulfur bond length of 2.34 Å when no bond restraint was used. When using an iron-sulfur distance restraint, (using the 'DFIX' command), the final value for the refined distance was 2.34 Å (without deviating more than 0.2 Å from the restraint), therefore for future refinements no iron-sulfur distance restraints were used. Anisotropic parameters indicate that the shape of the iron and sulfur atoms does not significantly deviate from spherical. This data, combined with the low temperature factors of the atoms, would indicate that there is only one orientation of the iron-sulfur bond and that its distance is 2.34 Å.

## 6.3.2 Effects on the Y96A Binding Site by Phenyl Substituted Substrates.

### 6.3.2.1 Differences in the 4-Hydroxydiphenylmethane Bound Structure Compared to Camphor Bound Structure.

The Y96A 4-hydroxydiphenylmethane bound structure is essentially the same as the Y96A-camphor bound structure (rmsd of the binding site side-chains of 0.46 Å compared to total protein of 0.49 Å). There are however two residues that have alternate orientations. Phe 98 rotates 35 ° to minimize a potential 2.8 Å interaction distance with the unhydroxylated 4-hydroxydiphenylmethane ring, improving it to a minimum interaction of 3.5 Å (discussed in more detail later, section 7.2.1).

The more significant change in this structure is to the key catalytic residue Thr 252, which has rotated 78 ° so that the threonine hydroxyl group forms a hydrogen bond with the substrate hydroxyl (3.0 Å); the Thr 252 CG2 carbon moving into the hydrophobic pocket around Val 294, Val 295 and Val 396. This rotation of the Thr 252 has an additional effect on the Glu 251 backbone, rotating the backbone carbonyl 85 °. This rotation alters an extensive H-bond network behind the binding

site, including residues Asn 255, Gln 186, Glu 251 and a number of water molecules (discussed in more detail later, section 7.1.2).

The new water molecule bound to the Thr 101 is in a slightly different orientation here; instead of being between the heme and the Thr 101 hydroxyl, the water is nearer the substrate, further into the binding site. This change in water orientation is probably caused by the alteration in the binding site hydrophobicity caused by the combination of the rotation of the Thr 252 and the presence of a hydroxyl group on the substrate.

#### 6.3.2.2 Differences in the Diphenylmethane, Phenylcyclohexane and *N*-cbz-2,6-dimethylpiperidine Bound Structures.

These three structures have essentially the same binding site structure as the Y96A-camphor and 4-hydroxydiphenylmethane bound structures. The Thr 252 has the orientation observed in the camphor structure, (i.e. the same as in the wild-type structure). The residues around Phe 87 and Phe 98 adopt the same configuration as the 4-hydroxydiphenylmethane bound structure so that they minimize the contact between the Phe 98 and the phenyl ring of the substrates. The Thr 101 bound water has the same location as in the camphor bound, Y96A-structure. The only significant differences in these structures is in the location of the binding site water molecules (see section 6.4.4.2).

### 6.3.3 Effects on the Y96A Binding Site with a Non-Aromatic Substrate.

#### 6.3.3.1 Differences in the Bicyclohexyl Bound Structure.

The binding site of the bicyclohexyl structure is the same as the three above structures. Despite the absence of an aromatic group in the substrate, the presence of the aliphatic ring near Phe 87 and Phe 98 has the same effect on the Phe 87, Phe 98 and Phe 193 orientation as the phenyl substituted substrates.



## 6.4 Substrate Orientations in the Y96A Mutant Binding Site.

### 6.4.1 Orientation of Camphor in the Y96A Mutant.

#### 6.4.1.1 Main Orientation of Camphor.

The camphor molecule was modelled into the orientation originally taken by camphor in the wild-type structure, then adjusted to fit the density of the Fo-Fc difference map and allowed to refine freely, figure 6.7.

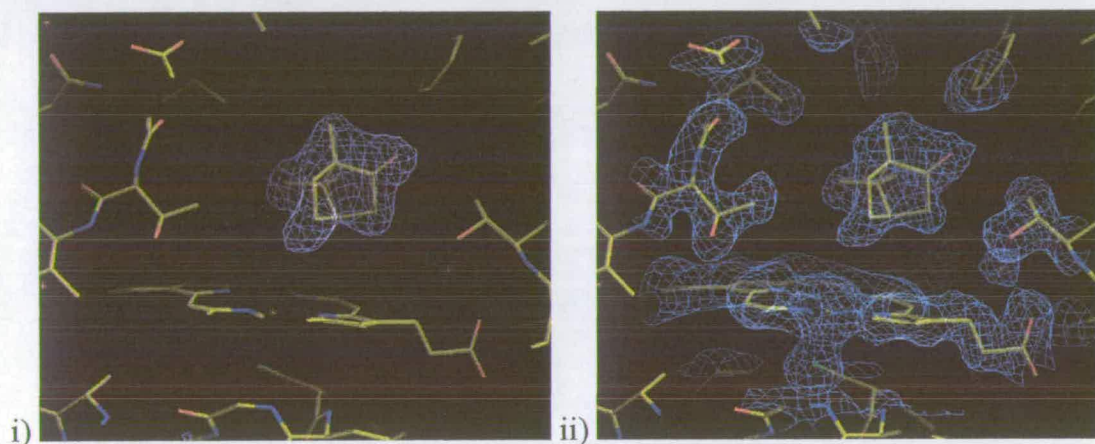


Figure 6.7: i) Fo-Fc map of the camphor bound binding site, with the final orientation of camphor added, contoured at a level of  $3\sigma$ . ii)  $\sigma$ -weighted difference map of the camphor bound binding site

After refinement, the camphor molecule occupied an orientation essentially the same as that initially modelled into the Fo-Fc difference map. A temperature factor of  $26.5 \text{ \AA}^2$ , for the camphor molecule, indicates that the thermal motion of the camphor is more similar to that of previously co-crystallized wild-type camphor analogues (e.g. camphane ( $30.1 \text{ \AA}^2$ ), adamantane ( $24.7 \text{ \AA}^2$ ) and norcamphor ( $33.5 \text{ \AA}^2$ )),<sup>5</sup> than to camphor ( $16.2 \text{ \AA}^2$ ).<sup>6</sup>

#### 6.4.1.2 Alternative Orientations of Camphor.

Since it is known that the Y96A mutant catalyses the formation of products other than 5-*exo*-hydroxycamphor (section 4.3.2.1), the substrate was modelled in



different orientations in order to account for potential hydroxylations of other carbon atoms. This was achieved by placing the hydroxylated carbon closest to the heme and fitting the camphor into the remaining density. These orientations were then allowed to refine with free occupancy, for 5 cycles. All refinements gave the previously optimized position in all but one case, in which the 3-*exo*-hydroxycamphor was the predicted product. In this case, the rotation of the camphor molecule back to the 'normal' orientation is prevented by unfavorable hydrophobic contacts with the methyl groups on the camphor, however, the occupancy was too low and the temperature factor was too high to definitely assign this as an orientation. Although other orientations of the camphor molecule must exist, it is not possible, at this resolution, to definitely assign any of them. These alternative orientations probably account for the higher temperature factor of the camphor with the Y96A mutant compared to the wild-type structure.

#### 6.4.1.3 Addition of a Water Molecule to Camphor Bound Binding Site.

Since it was known that there is a water molecule in the presence of camphor within the Y96A mutant binding site (section 4.3.2.1), the area between the heme and the camphor molecule was examined. The  $\sigma$ -weighted difference map, contoured at  $3\sigma$ , clearly shows density corresponding to a molecule above the heme with the centre of this density being 1.5 Å from the iron, figure 6.8.

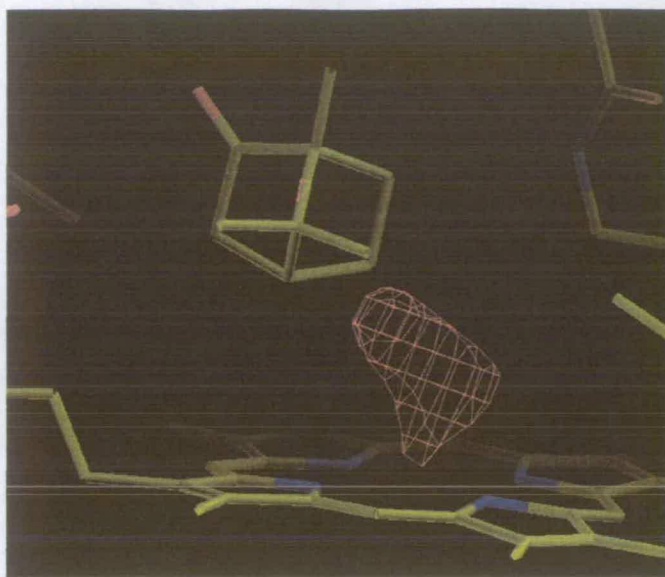


Figure 6.8:  $\sigma$ -weighted difference map showing the positive density (red) for a molecule above the heme.

This molecule was modelled as water and was initially placed at a distance of 1.5 Å from the iron which was consistent with previously reported water-iron distances in the presence of substrates.<sup>5</sup> This water was allowed to refine with a Fe-O bond set with no distance restraint and a freely refining occupancy. The final refined value of the occupancy was 58 % which is consistent with the binding data, which indicates a water molecule present at 52 % occupancy. The Fe-O starting distance was altered and the structure refined to investigate the effect on the final orientation, but the distance repeatedly refined to 1.45 Å. The starting occupancy was also varied and again refined to 58 %. When the occupancy of the water was forced to be 100 % the temperature factor increased from 24.5 Å<sup>2</sup> to 34.5 Å<sup>2</sup>.

#### 6.4.1.4 Comparison of Camphor in Wild-Type and Y96A Mutant P-450cam.

The final orientation of the camphor molecule is almost identical to the wild-type structure with a rmsd of only 0.28 Å, compared to the average rmsd between the two structures of 0.45 Å, figure 6.9.

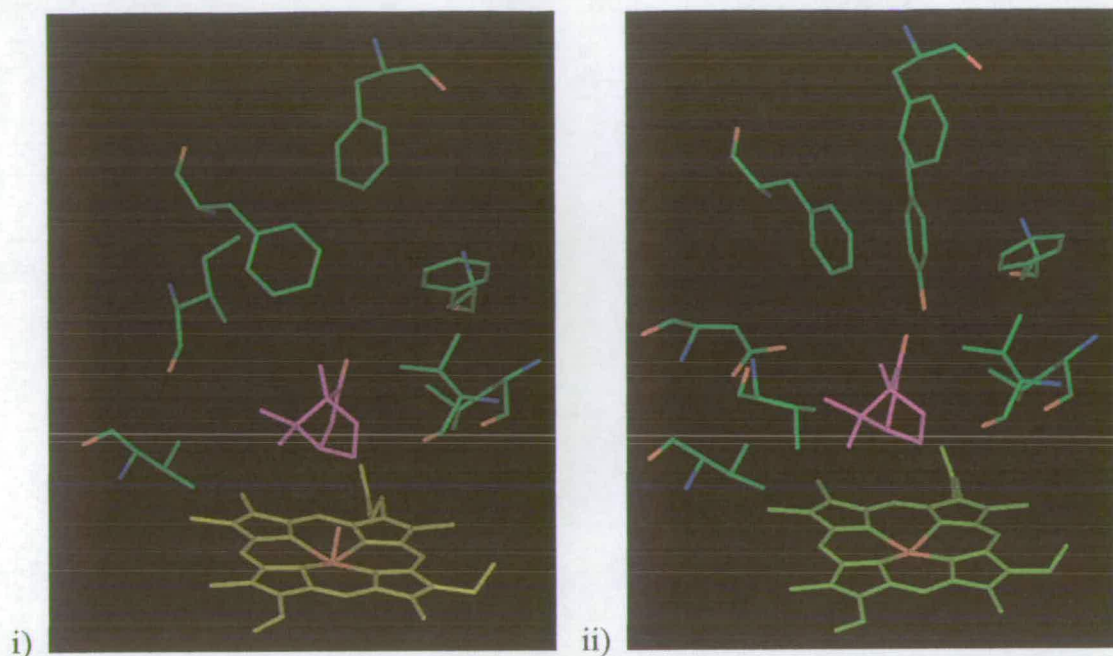


Figure 6.9: i) the wild-type camphor bound and ii) the Y96A camphor bound binding sites.

These similarities can be seen more clearly when the two structures were overlapped, figure 6.10.

The temperature factor of the camphor is lower in the wild-type structure ( $16.2 \text{ \AA}^2$  vs.  $26.5 \text{ \AA}^2$ ) as would be expected from the binding data, which indicates increased movement of the camphor molecule in the Y96A mutant, resulting in the entry of a water molecule and loss of stereoselectivity.

The largest difference in the orientation of the camphor molecule in the Y96A mutant vs. the wild-type enzyme is  $0.69 \text{ \AA}$  for the carbonyl oxygen, the other differences ranging from  $0.65 \text{ \AA}$ - $0.35 \text{ \AA}$ .



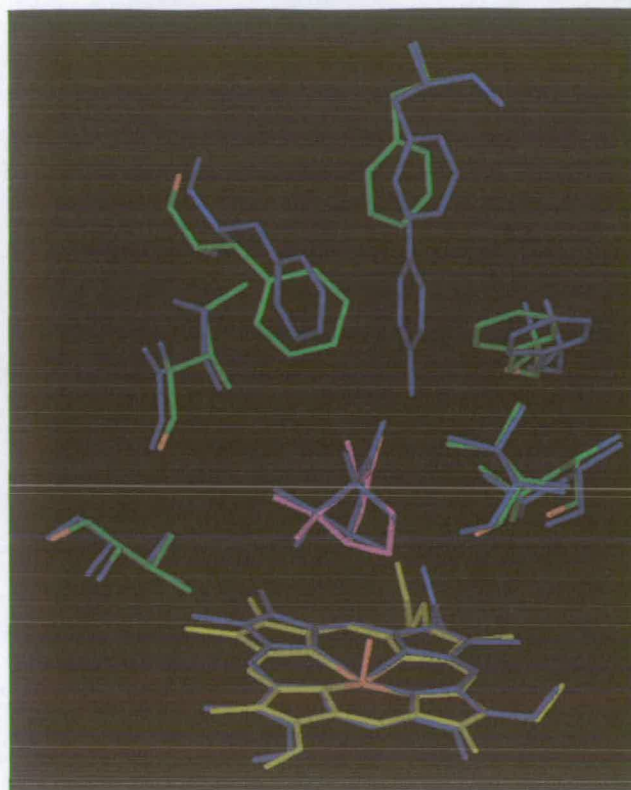


Figure 6.10: Overlap of the wild-type (blue) and Y96A (coloured by atom type) camphor bound binding sites.

This indicates that the camphor carbonyl to tyrosine 96 hydrogen bond does not fully control the orientation of the camphor molecule within the binding site. The camphor molecule, with the loss of the hydrogen bond, has rotated slightly towards Thr 101 ( $22.5^\circ$ ), however, the additional restraints on the substrate maintain the C5 carbon (hydroxylation site) within  $0.34 \text{ \AA}$  of the wild-type position.

#### 6.4.2 Orientation of 4-Hydroxydiphenylmethane in the Y96A Mutant.

4-Hydroxydiphenylmethane is the sole product of the biotransformation of diphenylmethane with the cytochrome P-450cam Y96A mutant.<sup>7</sup> This product was chosen for initial crystallization studies because, as an alcohol, it has higher aqueous solubility, facilitating co-crystallization; additionally the hydroxy group of the 4-hydroxydiphenylmethane ligates to the heme as indicated by a Soret band maximum at 417 nm. It was hoped that this additional constraint on the substrate orientation

would make the molecule less mobile and therefore facilitate interpretation of the substrate electron density maps. The density of the Fo-Fc difference map, at a level of  $2\sigma$ , shows clearly the orientation of the 4-hydroxydiphenylmethane in the binding site, figure 6.11.

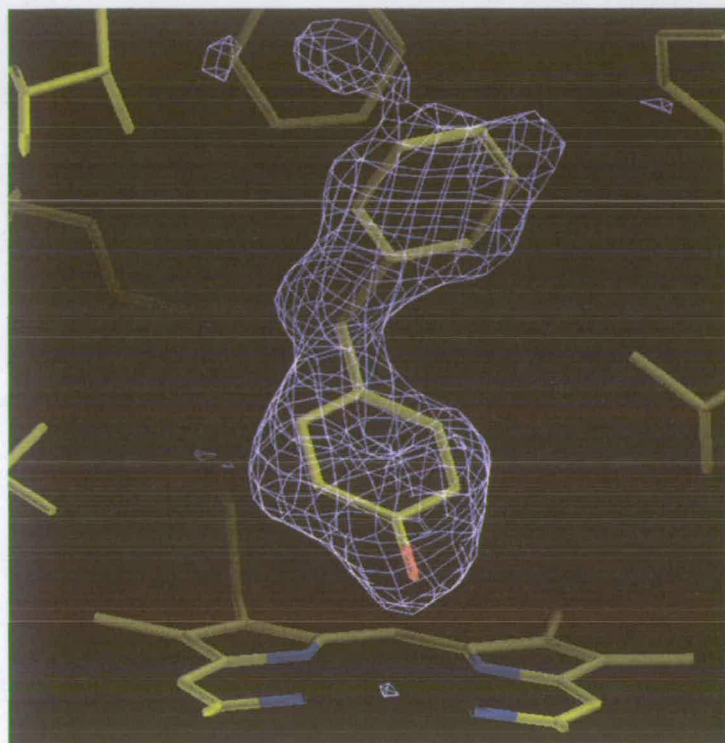


Figure 6.11: The positive Fo-Fc difference map of the binding site of the 4-hydroxydiphenylmethane structure, with the final refined 4-hydroxydiphenylmethane orientation added for clarity.

The final refined position of the 4-hydroxydiphenylmethane molecule can be seen below, figure 6.12.

This orientation of the 4-hydroxydiphenylmethane 'upper' (unhydroxylated) ring is significantly different from the formal location of the wild-type tyrosine ring where it was initially envisaged that the phenyl ring might be positioned. The 'upper' ring has, rather than fit between the Phe 87 and Phe 98 residues, orientated between Phe 87 and Val 247, stabilized by a series of hydrophobic contacts from Phe 87, Phe 98, Phe 193 and Val 395, to the upper 4-hydroxydiphenylmethane phenyl ring, figure 6.13.



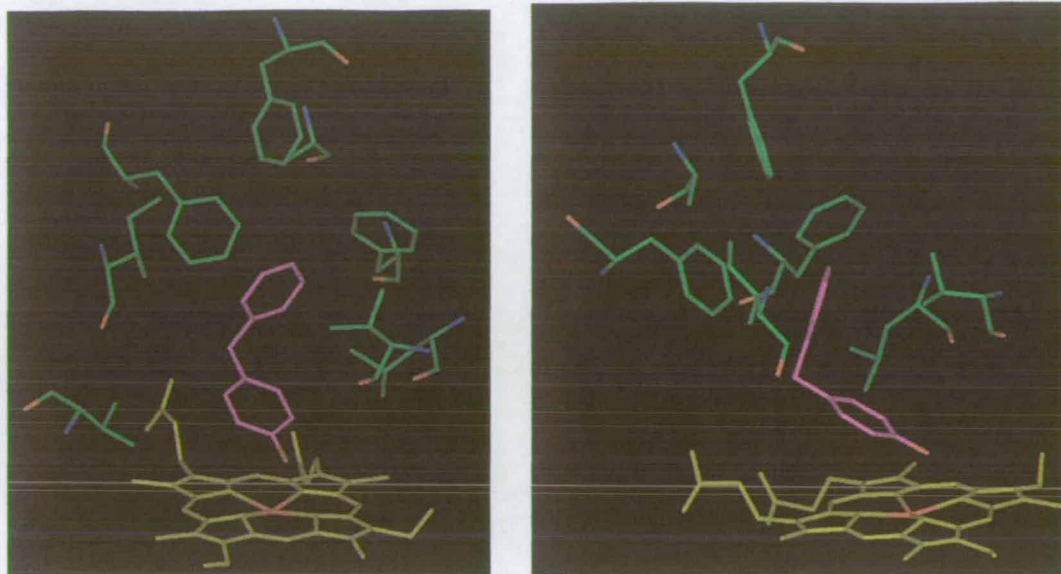


Figure 6.12: Binding site of the 4-hydroxydiphenylmethane bound Y96A structure from two different perspectives.

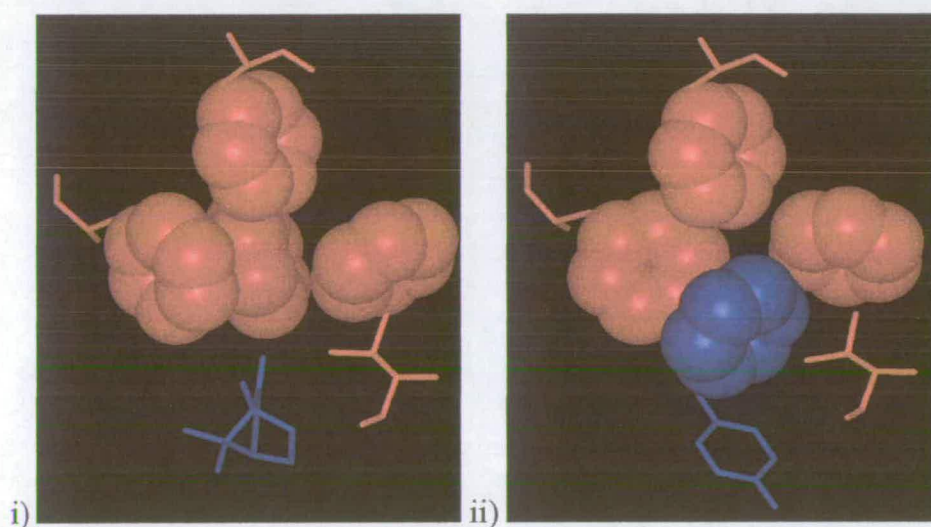


Figure 6.13: Comparison of the phenyl ring orientations of Phe 87, Phe 193, Phe 98 and Tyr 96 for the wild-type camphor bound structure (i), with the 4-hydroxydiphenylmethane structure (ii). Val 247 is also shown. Rings are shown in CPK format.

The 'lower' (hydroxylated) ring has no close ( $<3.5 \text{ \AA}$ ) interactions with any protein residues except for the hydroxyl group which ligates to the iron and Thr 252, figure 6.14.

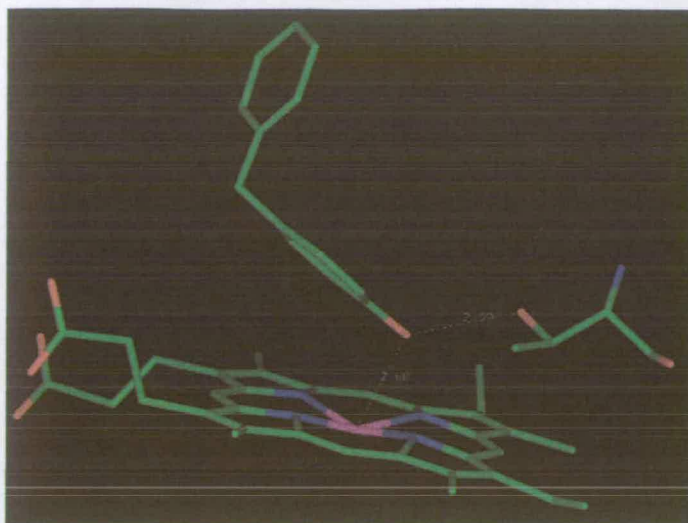


Figure 6.14: The hydrogen bonds between 4-hydroxydiphenylmethane and Thr 252.

The 4-hydroxydiphenylmethane hydroxyl oxygen is 2.39 Å from the heme iron with a 154 ° angle between the oxygen, heme iron and the Cys 357 sulfur (SG). A 180 ° angle for an octahedrally co-ordinated system would be expected, however, the deviation from linear is likely to be caused by a hydrogen bond to Thr 252 (3.0 Å), which is indicated in the  $\sigma$ -weighted map at a level of  $2\sigma$ , figure 6.15.

The temperature factor for the hydroxyl group is higher ( $40 \text{ \AA}^2$ ) than the rest of the ligand ( $35 \text{ \AA}^2$ ) probably due to improper optimization of the oxygen electron density by SHELX (because of the electron displacement of the oxygen over the Fe-O bond and the hydrogen bond to Thr 252).

Previous structures of P-450cam have ligand temperature factors ranging from  $16.2 \text{ \AA}^2$  (camphor)<sup>6</sup> to  $33.5 \text{ \AA}^2$  (norcamphor).<sup>5</sup> This would indicate that, by comparison, the 4-hydroxydiphenylmethane is relatively mobile in the active site.



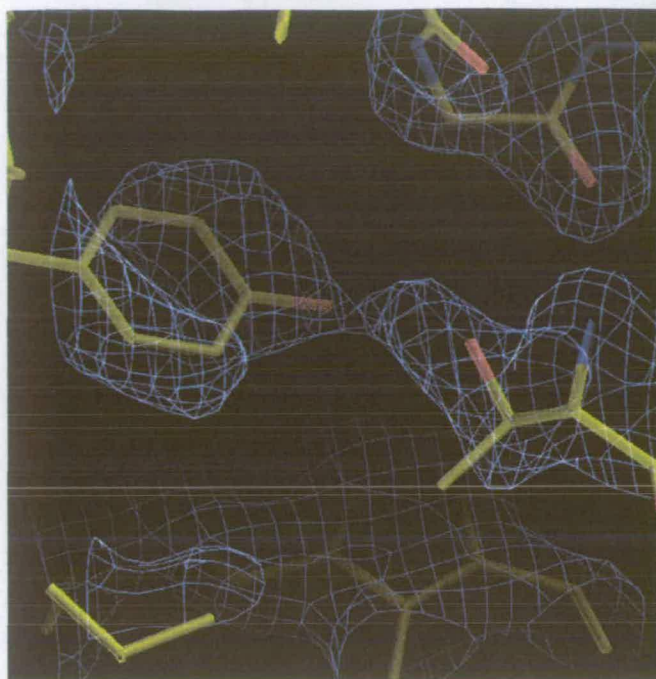


Figure 6.15:  $\sigma$ -weighted density map of the 4-hydroxydiphenylmethane structure showing the density between the hydroxy group and the threonine side-chain oxygen.

The density of the 'upper' ring indicates that the hydroxydiphenylmethane is able to move in the plane of the phenyl ring only. Motion perpendicular to the ring is prevented by interactions with Phe 87 and Val 395 and motion 'upwards', away from the heme, is prevented by residue Phe 193 and restrained by the hydroxy ligation to the heme. This rocking motion of the 'upper' ring is effectively a 0.5 Å movement to either side of the optimized structure (figure 6.16), this motion bringing the phenyl ring either closer to residues Val 385 or Phe 98. The nearest the phenyl ring gets to either residue is 3.3 Å, the optimized location being 4.0 Å from Val 385 and 3.5 Å from Phe 98.

Some small positive density was noted in the weighted 2Fo-Fc map by the upper phenyl ring, so the 4-hydroxydiphenylmethane was modelled with the phenyl rings reversed such that the hydroxy group was on the 'upper' ring; the occupancy of the ligand was then modified.



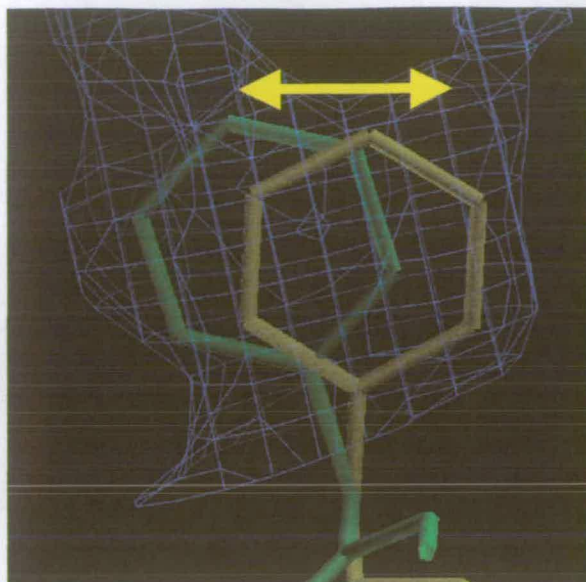


Figure 6.16:  $\sigma$ -weighted density map showing the area of motion covered by the 'upper' phenyl ring.

The newly refined structure indicated that if an oxygen exists in this location it must have an occupancy of less than 10 %. This clearly is inconsistent with the occupancy of the other atoms of the ligand and is therefore unlikely to be a major ligand orientation. However, a potential hydrogen bond to the backbone carbonyl of Met 184, to the substrate hydroxyl group, could stabilize this orientation of the substrate so it cannot be ruled out as a minor substrate orientation. If this orientation did exist in any significant amount we would expect to find the dihydroxylated product, which is not detected in the biotransformation.

There also exists the possibility that the 4-hydroxydiphenylmethane is not fully occupied in the binding site. This is plausible due to the low solubility of the substrate, although the relatively good binding constant makes this less likely. The combination of these factors could explain why the temperature factor of this ligand is higher than for the previous structures.

### 6.4.3 Orientation of Diphenylmethane in the Y96A Mutant.

#### 6.4.3.1 Diphenylmethane Orientation.

The electron density in the Fo-Fc difference map, for the diphenylmethane, was not as clear to that of the hydroxydiphenylmethane, however, the difference map suggested it has a similar orientation as the 4-hydroxydiphenylmethane. This can be seen more clearly with the  $\sigma$ -weighted map, figure 6.17.

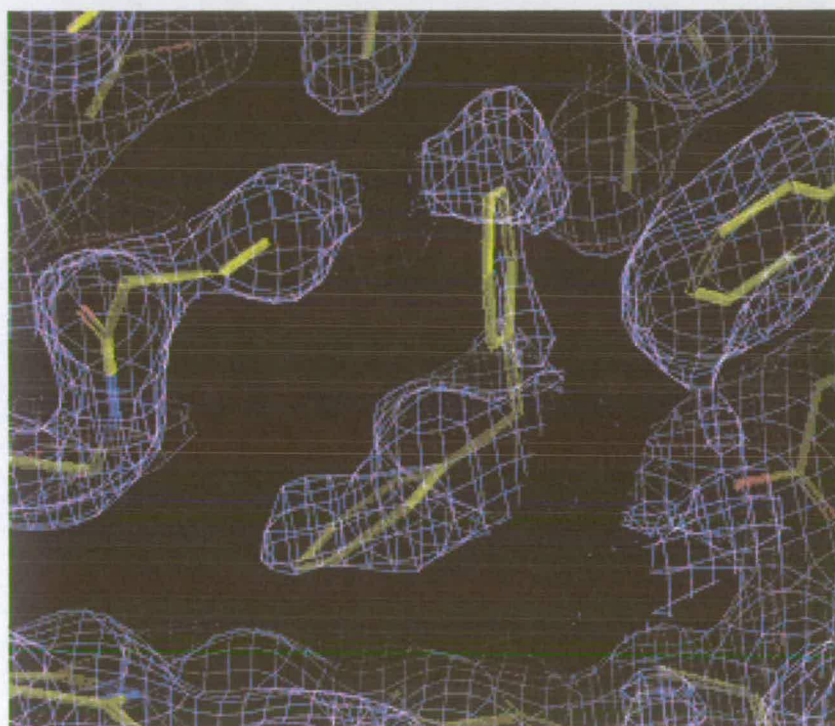


Figure 6.17:  $\sigma$ -weighted difference map of diphenylmethane in the binding site.

The optimized orientation of the diphenylmethane does appear to be similar to the 4-hydroxydiphenylmethane, with one phenyl ring between the Phe 87 and Val 244 and the other ring orientated above the heme, figure 6.18.

The diphenylmethane is in an orientation that holds the C4 carbon on the 'lower' ring above the heme (2.86 Å), the next nearest carbon is the C3 carbon at 3.39 Å. This supports the experimental result, in which the P-450cam Y96A mutant produces only 4-hydroxydiphenylmethane.



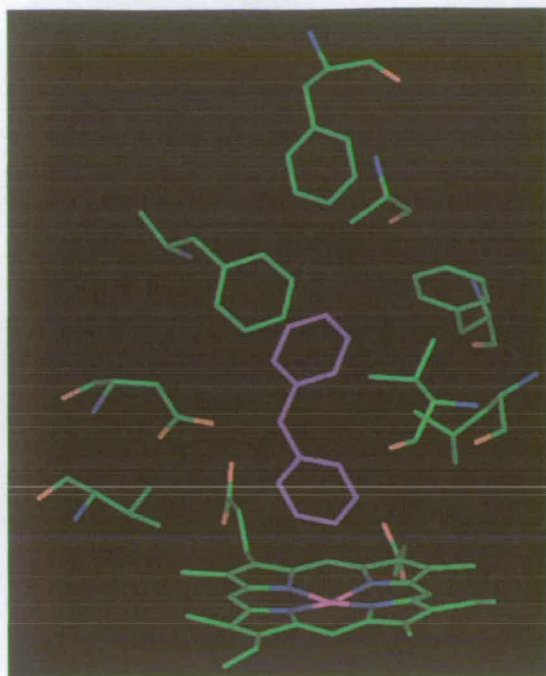


Figure 6.18: Diphenylmethane (purple) binding to the Y96A mutant.

#### 6.4.3.2 Binding Site Water in the Diphenylmethane Bound Structure.

The density around the diphenylmethane was not as clear as in other structures. It was thought that this was due to partial occupancy of the binding site by the diphenylmethane ligand. For the wild-type structure, with no substrate present, six water molecules occupy the binding site.<sup>8</sup> Therefore, if the diphenylmethane is only partially occupied, these water molecules could occupy the binding site and distort the density map. When the ligand was refined based on this premise, the diphenylmethane occupancy was refined in addition to four water molecules occupying the unassigned density of the substrate bound  $\sigma$ -weighted  $2F_o-F_c$  map. When this was done the water and substrate refined to 50 % each. The orientations of these bound waters are shown in figure 6.19.

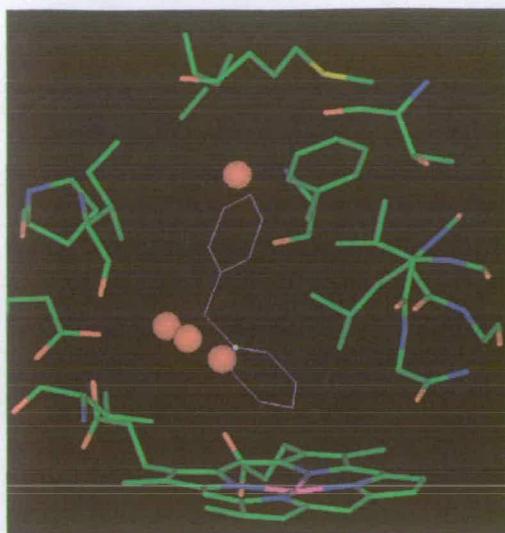


Figure 6.19: Binding site of the diphenylmethane structure showing the location of the partially bound waters (red) in relation to the substrate (purple).

To improve the electron density maps another co-crystallized structure for diphenylmethane was solved. This structure was of lower resolution ( $2.4 \text{ \AA}$ ) than the previous diphenylmethane structure, however, it showed the same 50 % occupancy of the diphenylmethane. The diphenylmethane has a  $1.7 \text{ }\mu\text{M}$  binding constant, therefore, the partial occupancy is probably due to the low aqueous solubility. This is a problem due to the use of co-crystallization as a method of substrate addition. The solubility of the diphenylmethane is probably not sufficient to give the maximum occupancy of the ligand. This aside, it was still possible to assign the density to a fixed orientation, although we cannot be confident of the temperature factors.

#### 6.4.4 Orientation of Phenylcyclohexane in the Y96A Mutant.

##### 6.4.4.1 Orientation of Phenylcyclohexane.

From the density of the  $F_o - F_c$  difference map of the phenylcyclohexane structure, figure 6.20, it was clear that the phenylcyclohexane was in a similar orientation to 4-hydroxydiphenylmethane and diphenylmethane.



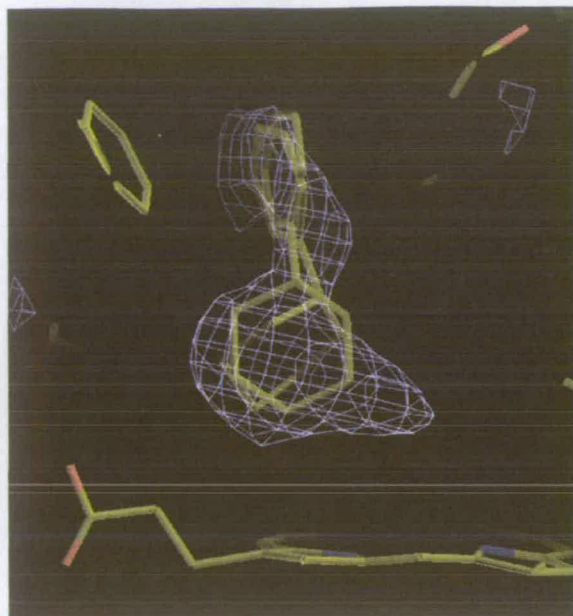


Figure 6.20: The  $F_o-F_c$  difference map of the phenylcyclohexane structure contoured at a  $\sigma$ -level of 2.

The phenyl ring was easily modelled in the space between Phe 87 and Val 244 with clear density. It became obvious that the cyclohexane ring could not be orientated to give one refined position that covered the density shown in the difference map. It was found that two phenylcyclohexane orientations could fit into the density, figure 6.21.

The phenyl rings were essentially in the same position,  $\sim 0.5$  Å difference. The main difference between the two orientations is in the configuration of the cyclohexane rings. The two cyclohexane rings are in opposite chair configurations. One is orientated towards Val 295, the other towards Val 244. In both cases the cyclohexane ring is oriented away from the heme centre towards the Thr 101/ Asp 297 edge of the heme. The nearest hydrogens for potential hydroxylation, for the molecule near Val 295, are the equatorial C3 (2.2 Å) and both the C4 axial (3.3 Å) and equatorial (3.5 Å) hydrogens; for the other orientation, near Val 244, the equatorial C2 (4.6 Å) and C3 (4.4 Å) and the axial C4 (4.5 Å) hydrogens are nearest to the iron. For hydroxylation to take place, the cyclohexane molecule would have to move closer to the centre of the heme.

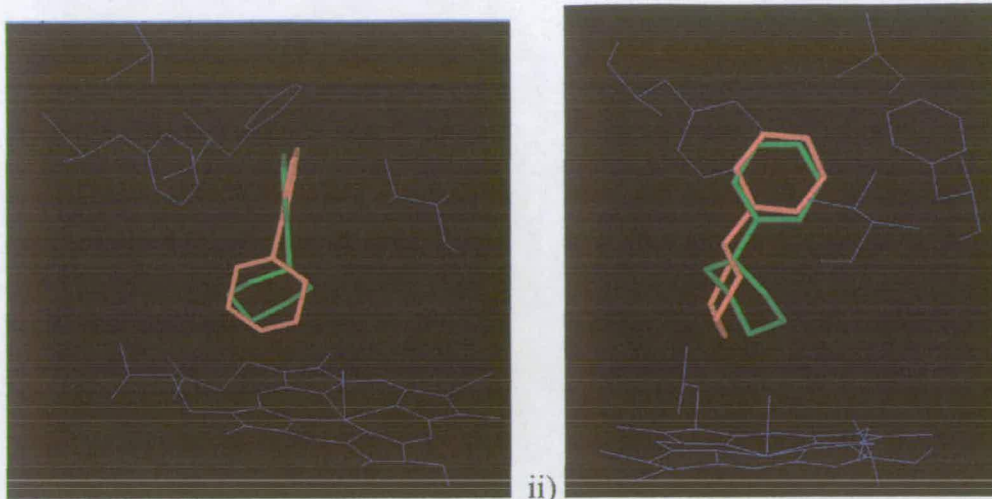


Figure 6.21: Orientations of the phenylcyclohexane structures from two perspectives.

If we assume that the phenyl ring maintains its position then the cyclohexane ring must move, thus changing the positions of the nearest hydrogens. When this is taken into account, the movement of the cyclohexane ring would move the C4 and C3 hydrogens nearer but move the C2 hydrogen further away from the iron centre. This is consistent with the experimental evidence which shows 4-axial, 4-equatorial and 3-equatorial hydroxylation, see chapter 1.

The density between the two orientations of the phenylcyclohexane would suggest there is a degree of mobility between the two orientations, rather than two distinct cyclohexane orientations. This explains the higher temperature factor of phenylcyclohexane ( $60 \text{ \AA}^2$ ) than previous structures.

#### 6.4.4.2 Phenylcyclohexane Active Site Bound Water.

As for the Y96A camphor bound structure the difference map for the binding site has unexplained density above the heme iron that could not be accounted for by the ligand. A water molecule was modelled into this position and refined with varying occupancy. The final refined value was 64 % occupancy and a temperature factor of  $28 \text{ \AA}^2$ . There is a large difference in this high occupancy for the water molecule compared to the low occupancy suggested by the percentage high-spin of 85 %.

The water molecule is in close contact ( $< 3\text{\AA}$ ) with some of the hydrogens on the substrates, which would suggest that the water cannot be present in this position when the substrate is in its crystal (average) orientation. This implies that the occupancy of the water molecule here is too high (if the phenylcyclohexane is fully occupied) and that the water molecule can bind to the heme only when the cyclohexane ring moves way from the iron center. If, however, the phenylcyclohexane is not fully occupied, that could explain why the water occupancy is higher than expected; although there is no unexplained density as in the case of the diphenylmethane structure.

Examination of wild-type P-450cam structures with bound water indicates that the water ligand is always  $< 3\text{\AA}$  from the substrate.<sup>5,9</sup> It is possible that the water molecule can exist in tandem with a close substrate. For this to happen the water must have a smaller Van der Waals radii than normal, due to delocalization of the electrons towards the heme, thus allowing interactions of  $< 3\text{\AA}$ .

#### **6.4.5 Orientation of Bicyclohexyl in the Y96A Mutant.**

As with the previous structures the  $F_o-F_c$  difference map showed clear density for the bicyclohexyl rings, figure 6.23.

One of the cyclohexane rings lies in the chair conformation between Val 247 and Phe 87, in approximately the similar orientation as the 'upper' phenyl ring in 4-hydroxydiphenylmethane, diphenylmethane and phenylcyclohexane, figure 6.23.



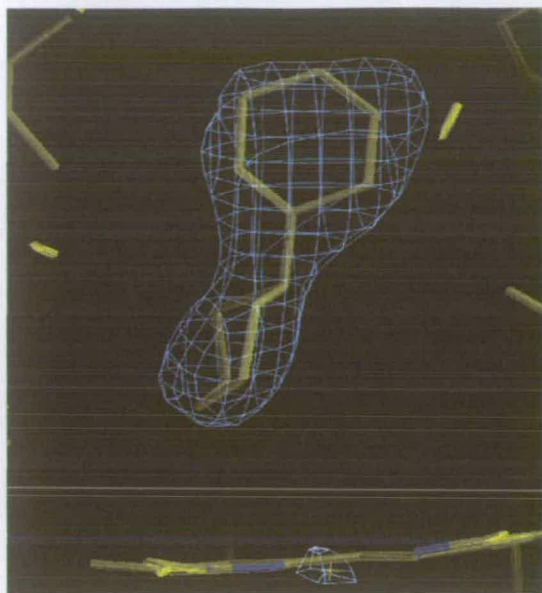


Figure 6.22: The  $F_o - F_c$  difference map of the bicyclohexyl structure contoured at a  $\sigma$ -level of 2.

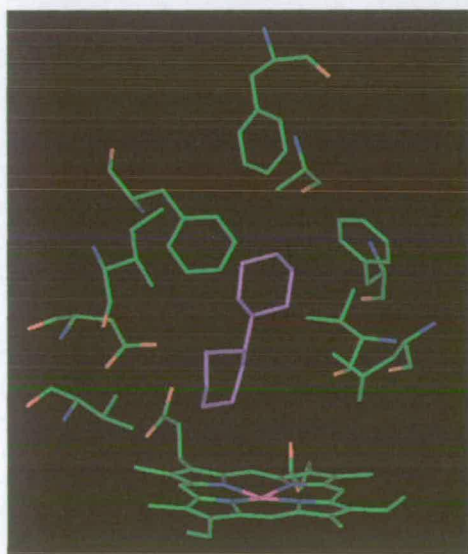


Figure 6.23: Bicyclohexyl in the Y96A binding site.

This ring is slightly flattened compared to the optimized chair conformation with angles of  $110^\circ - 111^\circ$  (vs.  $111^\circ - 115^\circ$  for the energy minimized structure). The other cyclohexane ring lies in a chair conformation with the C4 carbon on the ring pointing towards Val 295. The C3 axial hydrogen on the 'lower' ring is nearest to the iron centre, with the next nearest hydrogen being the C4 axial, although neither carbon is in a close position for hydroxylation ( $\sim 4.5 \text{ \AA}$ ). In order to be hydroxylated



the cyclohexane ring would have to move towards the iron centre. This would place the C3 equatorial and axial and the C4 axial hydrogens within hydroxylation distance. There is no evidence in the  $\sigma$ -weighted difference map of an iron bound water.

#### 6.4.6 Binding of *N*-cbz-2,6-dimethylpiperidine.

From the  $F_o - F_c$  difference map for the *N*-cbz-2,6-dimethylpiperidine [49] it was not obvious what the exact orientation was, although the density of the 'upper' ring suggested it was occupied by a phenyl ring. After some manipulation the final refined structure did fit the density, figure 6.24.

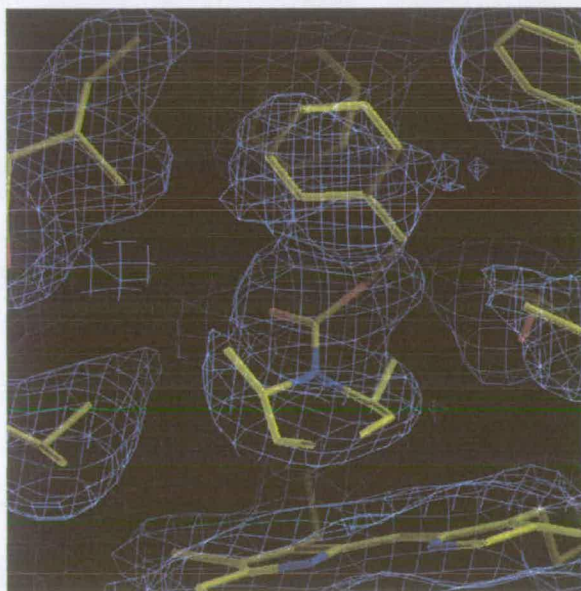


Figure 6.24:  $\sigma$ -weighted difference map of *N*-cbz-2,6-dimethylpiperidine contoured at  $2\sigma$ .

The reason for the problem in fitting **49** to the density was that none of the energy minimized structures seemed to be consistent with the electron density map. It was not until the structure was refined separately, (i.e. with the phenyl ring, linker and piperidine ring all placed into the density then joined together), that a satisfactory orientation could be found, figure 6.25.

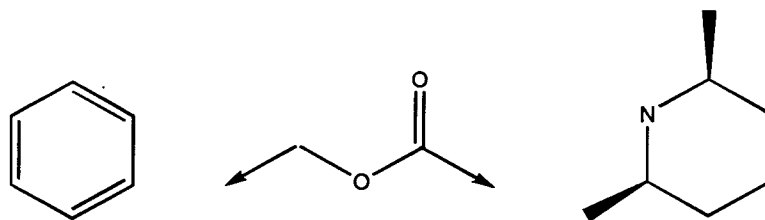


Figure 6.25: Separate parts of **49** modelled into the electron density.

The orientation of the final refined **49** was not close to any of the energy minimized forms. A Monte Carlo search of various low energy forms of the molecule did not reveal this orientation either. However, when the final refined value for the orientation was subjected to an energy minimization routine in InsightII, it maintained its orientation. This shows that crystal orientation of **49** is indeed in a local energy minimum and not thermodynamically impossible.

The final binding orientation of **49** has the 'upper' phenyl ring in the same position as for the other aromatically substituted substrates, with the lower ring sitting above the heme, slightly towards Asp 297, figure 6.26.

The methyl groups of the piperidine ring could not be satisfactorily placed in the density in any one orientation, therefore, they were left out of the final refinement, although they are in a *cis* position with regards to the cbz group. This lack of clear density for the methyl groups suggests a high degree of movement of the piperidine ring. The only other feature of note in the binding of **49** is a hydrogen bond from the carbonyl to the Asp 297 (OG2), at a distance of 2.6 Å, figure 6.27.

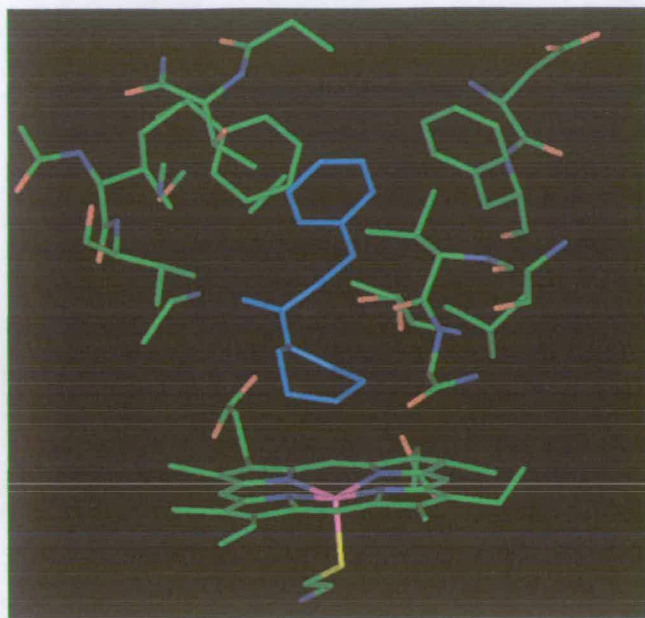


Figure 6.26: Substrate **49** in the Y96A binding site.

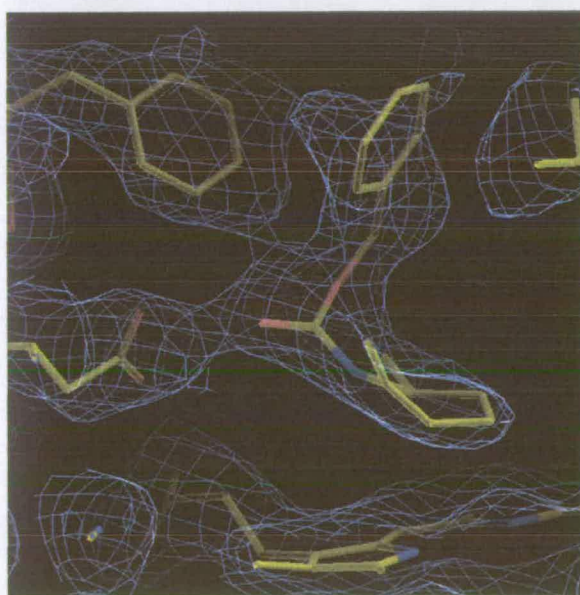


Figure 6.27:  $\sigma$ -weighted difference map of *N*-cbz-2,6-dimethylpiperidine showing the hydrogen bond to Asp 297 contoured at  $2\sigma$ .

The temperature factors of this substrate are higher than for the other substrates. The reasons for this are probably a combination of the movement of the substrate and a lower occupancy. The binding constant for **49** is poorer than for any of the other crystal bound substrates and it has poor aqueous solubility. This would

imply that the substrate may not be fully bound. The substrate was finally modelled with an occupancy of 70 % which gave the piperidine ring and the ether linkage a similar temperature factor to the other substrates ( $\sim 40 \text{ \AA}^2$ ). It was found however that the temperature factor of the phenyl ring was significantly higher ( $50 \text{ \AA}^2$ ) and had to be refined at an occupancy of 50 % to maintain consistent temperature factors over the molecule. The reason for this cannot be immediately determined, however, it may be that the poor binding of this substrate is due to the strain placed upon the molecule by this binding orientation. Thus due to this, the phenyl ring is not in a low energy conformation and is more mobile in an attempt to reduce the intramolecular energy.

## 6.5 Comparison of the Binding Orientations.

### 6.5.1 Orientation of the 'Upper' Phenyl Rings.

The phenyl rings of 4-hydroxydiphenylmethane, diphenylmethane phenylcyclohexane and all sit in similar orientations. The phenyl rings all sit in the same plane ( $< 20^\circ$  difference) between Phe 87 and Val 247. The differences between the structures lie in the distance of the centre of the phenyl ring to the iron centre. The 'upper' phenyl rings of diphenylmethane and hydroxydiphenylmethane are almost identical, the diphenylmethane ring being  $\sim 0.5 \text{ \AA}$  further from the heme than the hydroxydiphenylmethane ring.

When comparing the bicyclohexyl 'upper' ring orientation, it is orientated slightly further away from the heme ( $1.0 \text{ \AA}$ ) and nearer Val 247 ( $3.8 \text{ \AA}$  vs.  $4.3 \text{ \AA}$ ) than the 4-hydroxydiphenylmethane phenyl ring.



## 6.5.2 Comparison of the 'Lower' Rings of the 4-Hydroxydiphenylmethane, Diphenylmethane, Phenylcyclohexane and Bicyclohexyl Bound Orientations.

### 6.5.2.1 4-Hydroxydiphenylmethane vs. Diphenylmethane.

In the case of 4-hydroxydiphenylmethane the hydroxylated phenyl ring was orientated so that the C4 carbon was 'pulled' towards Thr 252 by the hydrogen bond to the 4-hydroxydiphenylmethane hydroxyl group, to give an angle between the Cys 357, sulfur heme iron and the C4 carbon of  $157^\circ$ , figure 6.28.

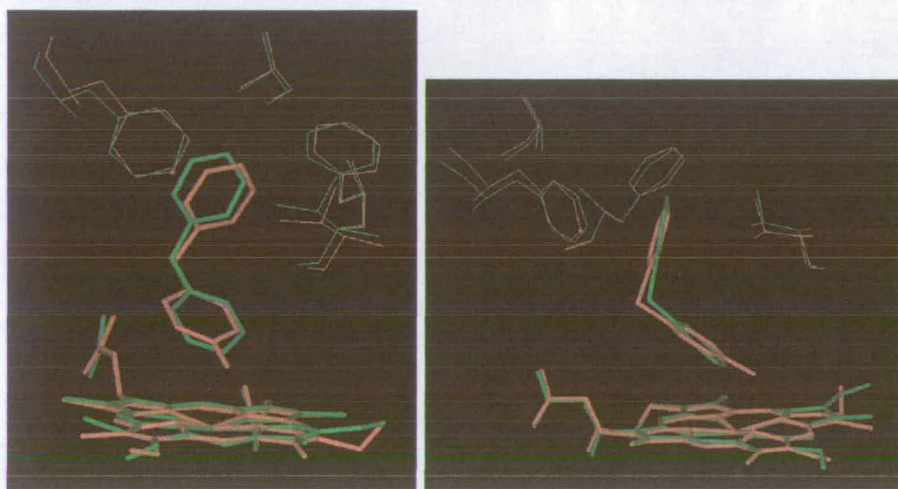


Figure 6.28: Comparison of diphenylmethane (green) and hydroxydiphenylmethane (red) from two perspectives.

However, for the diphenylmethane there is no Thr 252 hydrogen bond and the diphenylmethane lies slightly closer to the heme ( $2.71 \text{ \AA}$  vs.  $3.17 \text{ \AA}$  for 4-hydroxydiphenylmethane) to give a Cys (SG) to iron to 4-hydroxydiphenylmethane (O1) angle of close to linear,  $171.84^\circ$ . Other than this, the hydroxydiphenylmethane and diphenylmethane molecules are in essentially the same orientation.

### 6.5.2.2 Phenylcyclohexane Compared to 4-Hydroxydiphenylmethane/ Diphenylmethane.

The fundamental difference in structure between the phenylcyclohexane and 4-hydroxydiphenylmethane/ diphenylmethane bound structures is the fact the 4-hydroxydiphenylmethane/ diphenylmethane had a 'bridging' functional group between its two rings. As stated previously, the 'upper' phenyl rings are essentially in the same orientation, thus the absence of the bridging carbon for phenylcyclohexane orientates the cyclohexane ring in a different orientation to the 'lower' phenyl rings of 4-hydroxydiphenylmethane and diphenylmethane. This orientation of the phenylcyclohexane is away from the centre of the heme towards Val 295, figure 6.29.

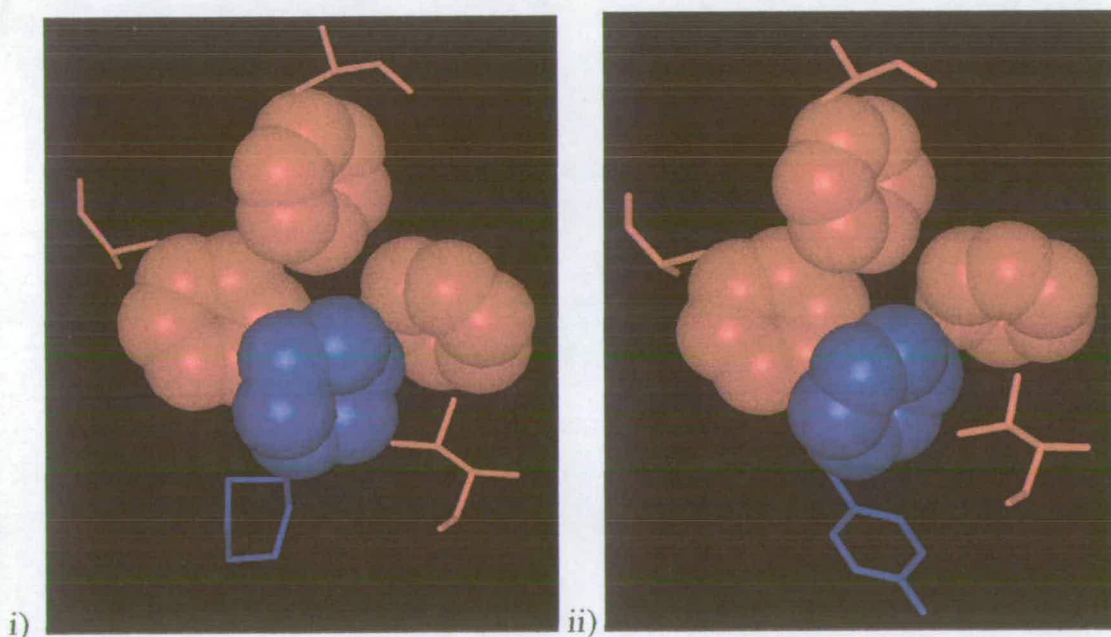


Figure 6.29: Showing one orientation of phenylcyclohexane (i) compared to hydroxydiphenylmethane (ii), showing the similar orientation of the 'upper' phenyl rings but different orientations of the 'lower' rings.



### 6.5.2.3 Comparison of Bicyclohexyl to 4-Hydroxydiphenylmethane/ Diphenylmethane and Phenylcyclohexane Structures.

The 'lower' cyclohexane ring of bicyclohexyl is in a significantly different orientation to the hydroxylated 4-hydroxydiphenylmethane ring which lies above the heme ring. The lower bicyclohexyl ring lies away from the heme center towards Val 295, figure 6.30.

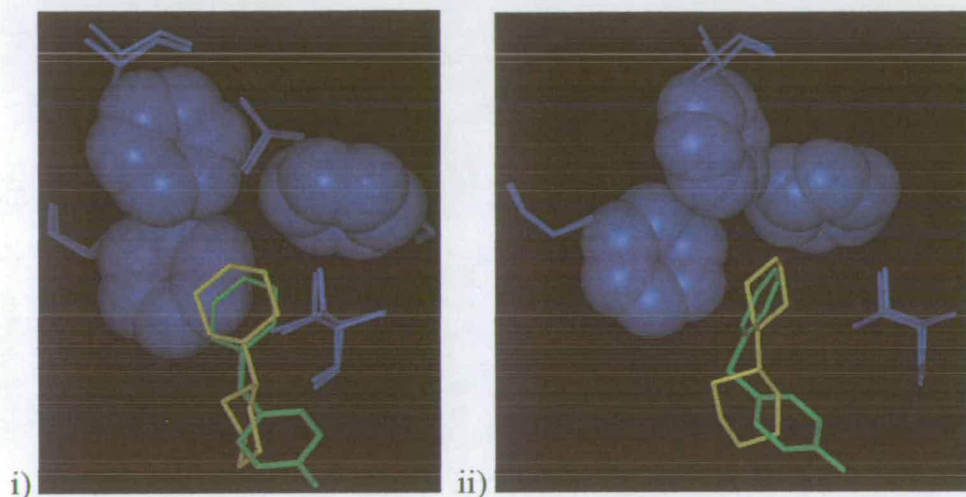


Figure 6.30: Comparison of bicyclohexyl (yellow) with 4-hydroxydiphenylmethane (green) in the binding site, from two different perspectives.

The bicyclohexyl, like the phenylcyclohexane molecule, has no 'bridging' methylene group to orientate the molecule towards the heme centre. There is a strong correlation between one of the phenylcyclohexane cyclohexane orientations and the bicyclohexyl 'lower' cyclohexane ring. Both cyclohexane rings lie away from the heme centre towards Val 295 (3.7 Å for bicyclohexyl and 4.0 Å for phenylcyclohexane).

Despite the similarities the two cyclohexane rings are in opposite chair configurations with the C1, C2 and C6 carbons overlapping but the C3, C4 and C5 carbons of the phenylcyclohexane ring are nearer the heme centre. The orientation of these two substrates close to Val 295 implies that there is a strong hydrophobic binding affinity in this region that these substrate preferentially bind to, rather than

occupy the large space above the heme. In fact for camphor binding, both Y96A mutant and wild-type, the two bridging methyl groups (C8, C9) are orientated into this region. Thus substrates that have some degree of motion in the binding site seem to position hydrophobic centers close to the Val 295 side-chain.

## 6.6 Conclusions.

Here we have determined the structures of the cytochrome P-450cam Y96A mutant co-crystallized with the natural substrate, camphor, 4-hydroxydiphenylmethane, diphenylmethane, phenylcyclohexane, bicyclohexyl and *N*-cbz-2,6-dimethylpiperidine.

The orientation of the 4-hydroxydiphenylmethane and diphenylmethane structures was similar to that expected, with the one phenyl ring in the upper aliphatic pocket and the other ring pointing towards the iron at the centre of the heme. The phenylcyclohexane and bicyclohexyl structures however bound with the lower ring towards Val 295 and not above the iron centre. This must be due to a strong aliphatic attraction to the region of the binding site around Val 295. Binding in this region is also seen in the camphor bound structure (both Y96A and wild-type) which has both the geminal methyl groups in this pocket. The *N*-cbz-2,6-dimethylpiperidine bound orientation was not predicted, although the poor binding was indicated by the experimental data (chapter 4). The important properties of these structures are discussed in the next chapter.



## 6.7 References.

- 1) Poulos, T. L.; Finzel, B. C.; Gunsalus, I. C.; Wagner, G. C.; Kraut, J. *Journal of Biological Chemistry* **1985**, *260*, 16122-16130.
- 2) Nickerson, D. P.; Wong, L. L. *Protein Engineering* **1997**, *10*, 1357-1361.
- 3) Lo, K. K. W.; Wong, L. L.; Hill, A. O. *FEBS Letters* **1999**, *451*, 342-346.
- 4) Vidakovic, M.; Sligar, S. G.; Li, H. Y.; Poulos, T. L. *Biochemistry* **1998**, *37*, 9211-9219.
- 5) Raag, R.; Poulos, T. L. *Biochemistry* **1991**, *30*, 2674-2684.
- 6) Poulos, T. L.; Finzel, B. C.; Howard, A. J. *Journal of Molecular Biology* **1987**, *195*, 687-700.
- 7) Fowler, S. M.; England, P. A.; Westlake, A. C. G.; Rouch, D. R.; Nickerson, D. P.; Blunt, C.; Braybrook, D.; West, S.; Wong, L. L.; Flitsch, S. L. *Journal of the Chemical Society-Chemical Communications* **1994**, 2761-2762.
- 8) Poulos, T. L.; Finzel, B. C.; Howard, A. J. *Biochemistry* **1986**, *25*, 5314-5322.
- 9) Raag, R.; Poulos, T. L. *Biochemistry* **1989**, *28*, 917-922.

## 7 Features of the Cytochrome P-450 Y96A Mutant.

---

### 7.1 Implications of the Substrate Bound Data.

#### 7.1.1 Enzyme Complementarity Towards Camphor.

The results presented here reflect the role of a number of parameters controlling the efficiency of substrate binding. Comparison of data from the biotransformation of camphane and norcamphor and the data from the Y96F and Y96A biotransformations of camphor indicate the importance of the various controlling factors. The increase in space in the binding site, by making a mutation, increases the production of the 3-*exo* product (the thermodynamically favoured hydroxylated product)<sup>1</sup> which can only result from significant rotation of the camphor molecule. The same result is gained by removal of the geminal carbons on the substrate (i.e. norcamphor).<sup>2</sup>

Camphane has no carbonyl oxygen available for hydrogen bond formation with the Tyr 96 hydroxyl, however the substrate/enzyme complementarity remains essentially the same as for camphor, i.e. no significant rotation of the camphane molecule was possible and thus little 3-*exo* product is produced.<sup>3</sup> Quantum mechanical calculations on the camphor have indicated that hydroxylation of the 5-

and 6- positions are energetically similar, therefore in cases where the motion of camphor is only slightly increased (e.g the Y96A mutant) a significant amount of the 6-*exo* product is produced. This shows that there is not enough selectivity in the wild-type enzyme to confer specificity without the hydrogen bond and additionally some hydrophobic interaction with the substrate.

### 7.1.2 Implications of the 4-Hydroxydiphenylmethane Structure for Cytochrome P-450 Mechanisms: an Unusual 251/252 Configuration in the 4-Hydroxydiphenylmethane Structure.

One of the remaining areas of uncertainty in cytochrome P-450 research is the mechanism for proton transfer to the active oxygen species. Unlike peroxidases, which have an acid-base catalyzed reaction supported by a binding site histidine,<sup>4</sup> P-450 monooxygenases have no obvious proton delivery mechanism.

Thr 252 is a conserved residue in all known P-450 monooxygenases (except P-450eryF<sup>5</sup> where its function is replaced by a binding site water). The role of this conserved residue is thought to be in stabilizing the iron bound oxygen and as part of a possible proton delivery channel to the heme.<sup>6-8</sup> Thr 252 is present in an unusual kink in the I helix in both the wild-type and Y96A mutant structures, figure 7.1. It has been proposed that this kink between Thr 252 and Glu 248 could provide space for an oxygen binding pocket.<sup>6,9</sup> This would thus allow molecular oxygen to access the iron during the catalytic cycle.

Glutamate 251 (next to threonine 252) has also been shown to have a role in the catalytic cycle. Mutagenesis of Glu 251 results in reduction of the turnover rate by an order of magnitude, the effect occurring after the second electron transfer.<sup>10</sup> Additionally the effect of the mutation of Glu 251 can be linked to the solvent pH.<sup>11</sup>

Kinetic isotope effects indicate that two protons are involved in the catalytic cycle, probably double protonating one of the bound dioxygen molecules so that it can be lost as water. The exact mechanism of this proton transfer is still unclear but the most likely mechanism, proposed recently by Sligar, has proton transfer *via* a catalytic

triad near a solvent rich cleft of residues Arg 186, Asp 182 and Lys 178 ; the proton then entering the binding site *via* this triad followed by an interaction with Glu 251.<sup>4,12</sup>

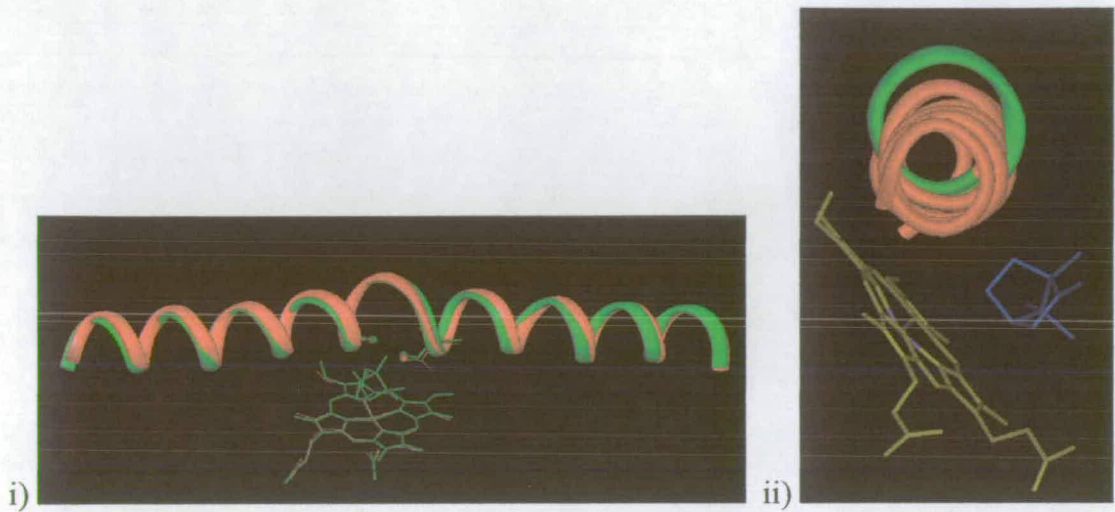


Figure 7.1: i) The I helix of wild-type (red) and Y96A (green) showing the kink due to Thr 252 in the middle (heme, Thr252 and Glu 248 have been shown). ii) The I helix (red) viewed by looking 'through' it, the ribbon from Glu 248- Thr 252 has been shown in green. Camphor and the heme group have been shown for clarity.

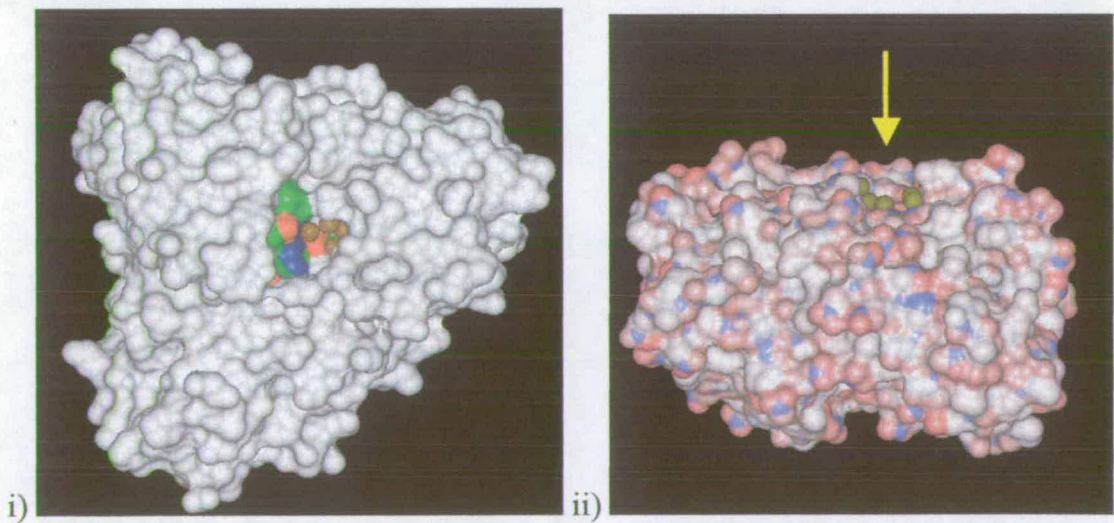


Figure 7.2: i) The protein Connolly surface showing the solvent cleft with residues Arg 186, Asp 182, Lys 178 (coloured by atom) and a number of water molecules (brown). ii) The side view of the cleft containing water molecules (yellow). The protein surface is coloured by charge (red-positive, blue negative).

Examination of the protein surface shows the cleft and the accessible charged residues with the bound water molecules, figure 7.2.

The pathway from the binding site to this solvent rich cleft is quite clear, as shown in figure 7.3.



Figure 7.3: Showing the proposed residues of the proton delivery pathway (green) in relation to the heme (yellow) and substrate (blue). The arrows indicate the proposed proton pathway.

It is postulated that once protons reach the end of the pathway, i.e. glutamate 251, the Glu side-chain acts as a carboxylate switch, transferring protons *via* two water molecules and Thr 252, to the bound dioxygen. This transfer of protons is only made possible because of an intricate hydrogen bonding network leading from the water rich surface cleft to the binding site, figure 7.4.

Mutagenesis of any of the residues in this pathway has been shown to lead to serious disruption in P-450cam turnover.<sup>4,11-13</sup> It has additionally been previously postulated that the Thr 252 could rotate in the binding site and stabilize the bound dioxygen, however, this would disrupt a strong hydrogen bond to Glu 248, figure 7.5.



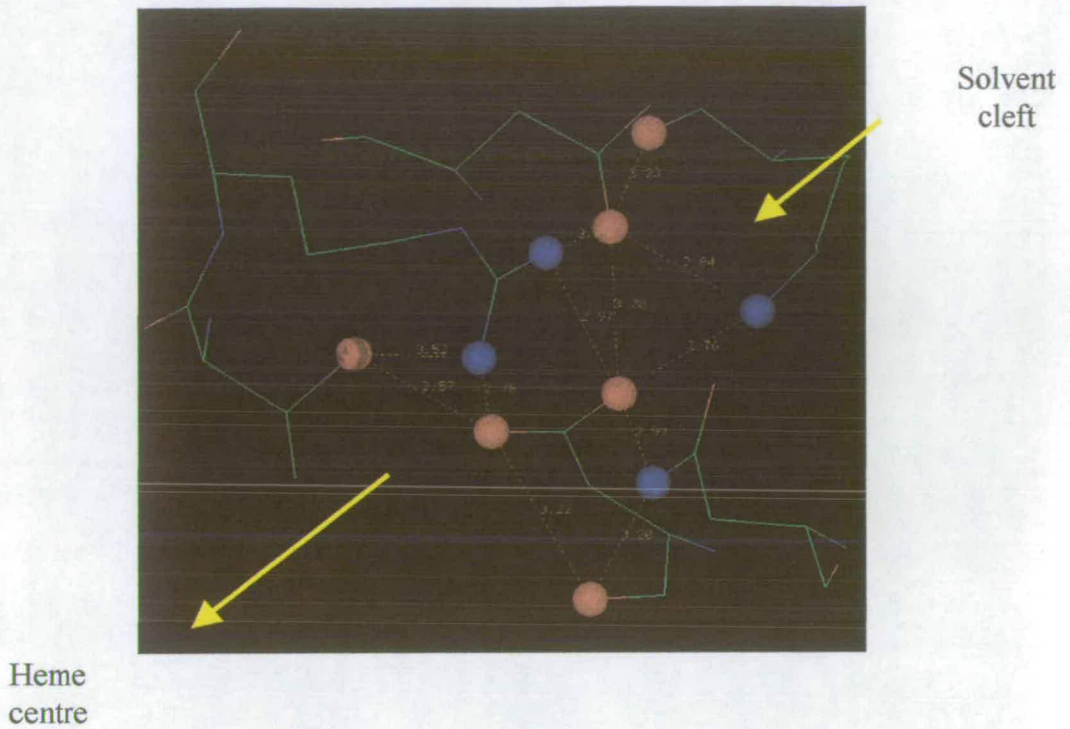


Figure 7.4: The hydrogen bonding network around Glu 251 (centre) with, clockwise, Thr 185, Arg 186, Asp 182, Lys 178, Asn 255 and Glu 251.

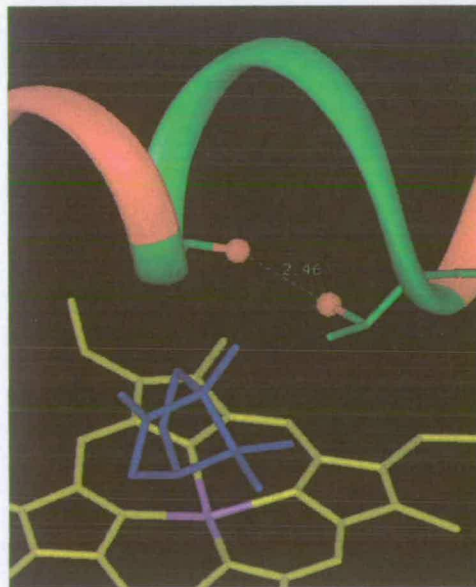


Figure 7.5: Section of the I helix showing the hydrogen bond between Glu 248 and Thr 252, camphor and heme have been included for clarity. The 'kink' is indicated in green.

In addition to this, Thr 252 has a hydrogen bond to water 687 and rotation would disrupt a hydrogen bonding network that extends to highly conserved Glu 366, figure 7.6.

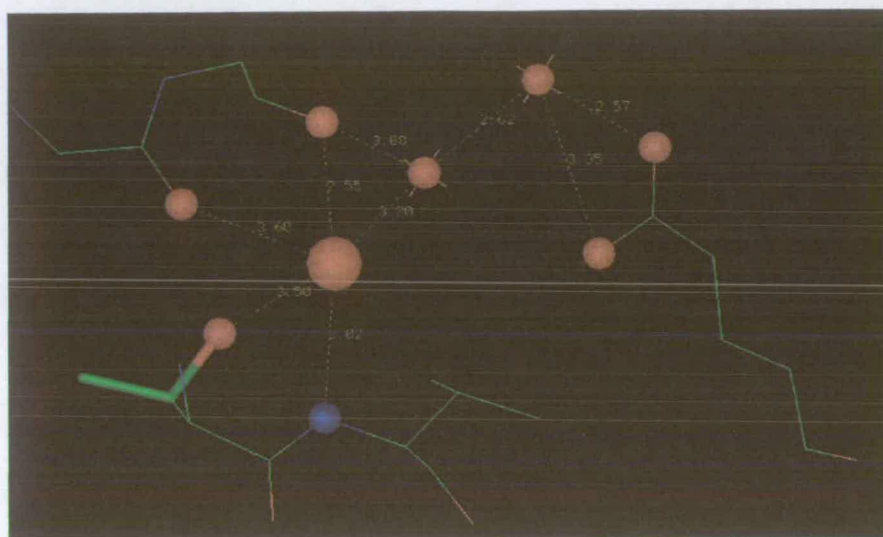


Figure 7.6: Hydrogen bonding network 'behind' the Thr 252 residue. Water 687 is enlarged and Thr 252 is highlighted. Residues include, clockwise, Thr 252, Glu 248, Gly 249, Glu 366.

For these reasons it was thought that the rotation of Thr 252 was energetically unfeasible. The exact role of the Thr 252 therefore still remains in some question. In the P-450cam Y96A 4-hydroxydiphenylmethane bound structure we have discovered the first evidence that the threonine can indeed rotate into a hydrogen bonding position towards a heme bound oxygen (figure 7.7), and this has lead us to a new model of proton transfer, which slightly differs from that recently proposed.

In this case, the rotation was initiated by ligand binding, however it can be postulated that a similar effect could occur upon the binding of dioxygen to the iron (II) species (after conversion of the molecular oxygen to superoxide). The Thr 252 could then stabilize the dioxygen in one orientation so that only one of the oxygen atoms is in a position to be double protonated *via* the proton channel, thus preventing the peroxide shunt (chapter 1).

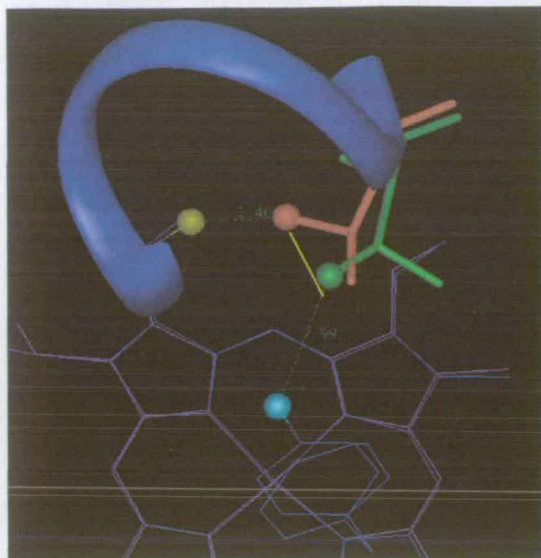


Figure 7.7: Comparison of wild-type (red) and Y96A, 4-hydroxydiphenylmethane (green) Thr 252 orientations. Glu 248 (yellow) - Thr 252 (wild-type) and Thr 252 - 4-hydroxydiphenylmethane hydroxyl group (cyan) hydrogen bonds are indicated.

This orientation of the threonine, thus the orientation of the bound di-oxygen, is ideal for proton transfer to the unligated oxygen *via* the proposed protonation mechanism, figure 7.8.

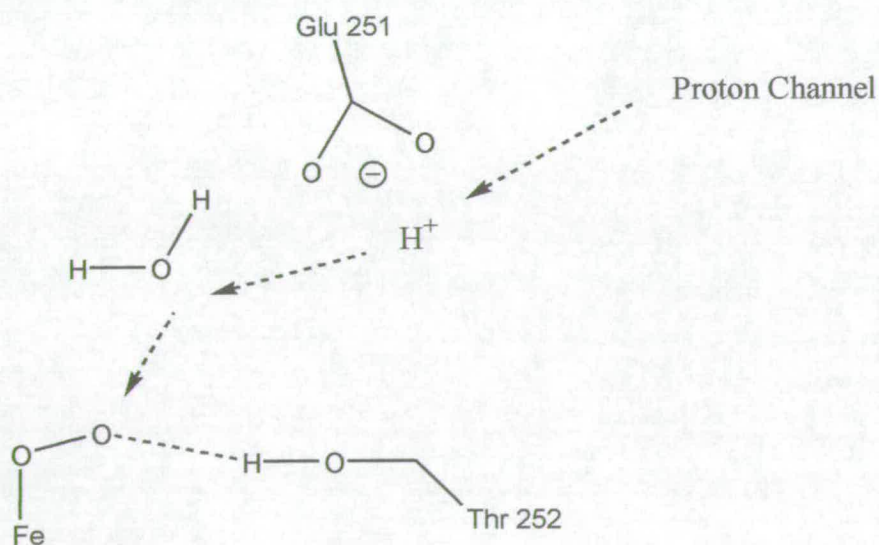


Figure 7.8: Schematic of the proposed proton delivery mechanism.



An additional consequences to the movement of the threonine, beyond that of rotating into a hydrogen bonding position towards the heme, is the disruption to the residues surrounding the Thr 252 main chain. The most significant change is the rotation of the backbone carbonyl of Glu 251, figure 7.9.

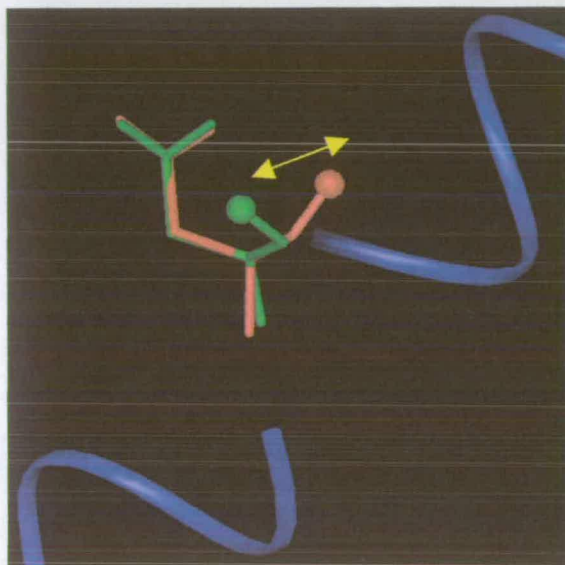


Figure 7.9: Comparison of wild-type (green) with Y96A 4-hydroxydiphenylmethane (red) orientation of the Glu 251 backbone carbonyl.

This movement of Glu 251 backbone has previously been observed for the T252A mutant structure and three other ligand bound structures, that of adamantane, an unnamed inhibitor and one of the 2-phenylimidazole orientations.<sup>6,14-16</sup> Each of these structures has a different cause for this movement (table 7.1), but in every case the intricate H-bonding network around water 687 was disrupted.

The data from the T252A mutant suggests that this orientation of the backbone is the low energy position and the strained configuration observed in the wild-type enzyme is only stabilized *via* the Thr 252-Glu 248 and Thr 252-Water 687 hydrogen bonds.

Table 7.1: Comparison of structures showing a Glu 251 backbone rotation.

structure	possible reason for rotation of the Glu 251 carbonyl
T252A mutant <sup>6</sup>	Thr 252 has lost H-bonding potential to water 687
adamantane <sup>14</sup>	Water 687 has moved into binding site
Pfizer unnamed (-) inhibitor <sup>15</sup>	Inhibitor carbon C20 comes into contact with Thr 252 CG2 (2.8Å) and it moves to compensate (3.2Å) disrupting H-bond to water 687
one of the 2-phenylimidazole orientations <sup>16</sup>	New binding site water (802) H-bonds to Glu 251 and disrupts the H-bonding pattern to water 687
4-hydroxydiphenylmethane	Thr252 moves to H-bonds to substrate and therefore removes bond to water 687

Disruption to this hydrogen bonding network *via* mutagenesis, interference with water 687 binding pattern or rotation of the threonine side-chain allows the more thermodynamically favourable orientation.

One major consequence of the rotation of the Glu 251 backbone, in this structure, is the effect on the hydrogen bond between it and the proposed carboxylate switch (Glu 251). Upon movement of the Glu 251 backbone carbonyl the hydrogen bond between it and the Glu 251 oxygen (OD1) is broken and a new one is formed, to the other Glu 251 carbonyl oxygen (OD2), figure 7.10.

The hydrogen bond from the backbone carbonyl to the Asn 255 (ND2) nitrogen is also lost, which is of importance mechanistically as Asn 255 is one of the residues stabilizing the proposed Glu 251 carboxylate switch. The movement of the Thr 252 backbone carbonyl induces a movement in Asn 255, thus affecting the strength of its hydrogen bond to Glu 251. It is thought that these changes to the environment of the carboxylate switch, especially the change in direct hydrogen bonding from the backbone oxygen, could be sufficient to change the pKa and initiate the proton transfer.

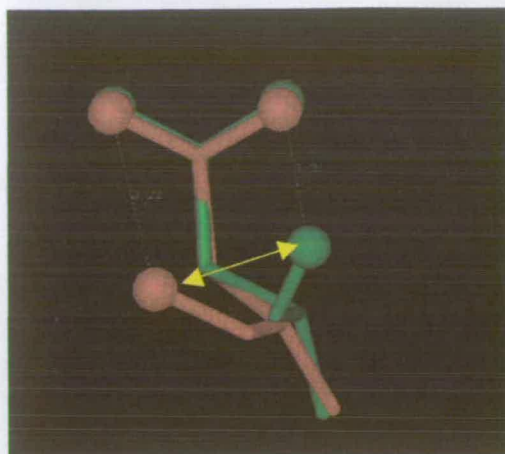


Figure 7.10: Movement of Glu 251 backbone carbonyl from wild-type orientation (red) to 4-hydroxydiphenylmethane orientation (green), showing the change in H-bonding to the glutamate side-chain.

Another important factor in stabilizing threonine 252, in the thermodynamically unfavored position, is an intricate hydrogen bonding network ‘behind’ the threonine (i.e. away from the binding site). As well as a hydrogen bond to the backbone of glutamate 248, threonine 252 hydrogen bonds to a water molecule, previously assigned as 687. This water molecule has a rigid conformation with hydrogen bonds to Gly 249, Val 253 and to Glu 366 *via* another water molecule, figure 7.11.

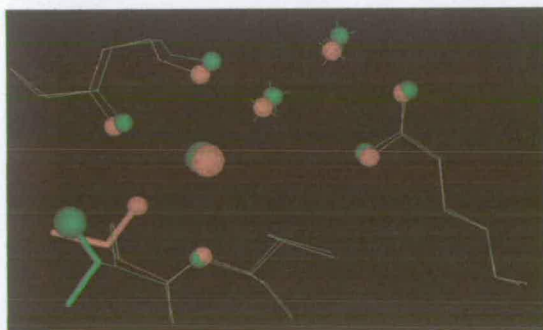


Figure 7.11: comparison of hydrogen bonding network behind Thr 252 for wild-type (red) and the Y96A 4-hydroxydiphenylmethane structure (green).

Glu 366 is another highly conserved residue in P-450 monooxygenases and it had been previously thought that the intricate hydrogen bonding around water 687



might have a role in the proton delivery mechanism; however mutagenesis experiments and no accessibility to the bulk solvent ( figure 7.12) have tended to discount this.<sup>6</sup>



Figure 7.12: Showing the Glu 366 (red) in relation to the protein. The arrow indicates how the protons would have to pass through the protein from Glu 366, a pathway not supported by a proton channel.

Another possible role for the region between the Glu 251 backbone and Glu 366 is in providing water to the binding site. In either model for the proton transfer, a water molecule is needed to transfer the proton from the Glu 251 carboxylate to the heme bound oxygen. No water molecules are readily observed in this region of the binding site in most of the crystal structures, which implies that water enters during the catalytic cycle. Indeed there is a putative binding site discovered by electrostatic modelling that is in the region of our proposed water molecule, figure 7.13.<sup>17</sup>

A water bound in this site should be capable of hydrogen bonding to Glu 251. It was additionally proposed that this water can only bind to Thr 252 if it underwent rotation.<sup>6</sup> The movement of the Glu 251 backbone carbonyl in fact enhances the

propensity for a water to bind in this site, by providing additional space for the water molecule and an additional hydrogen bond to Glu 251 side-chain carbonyl.

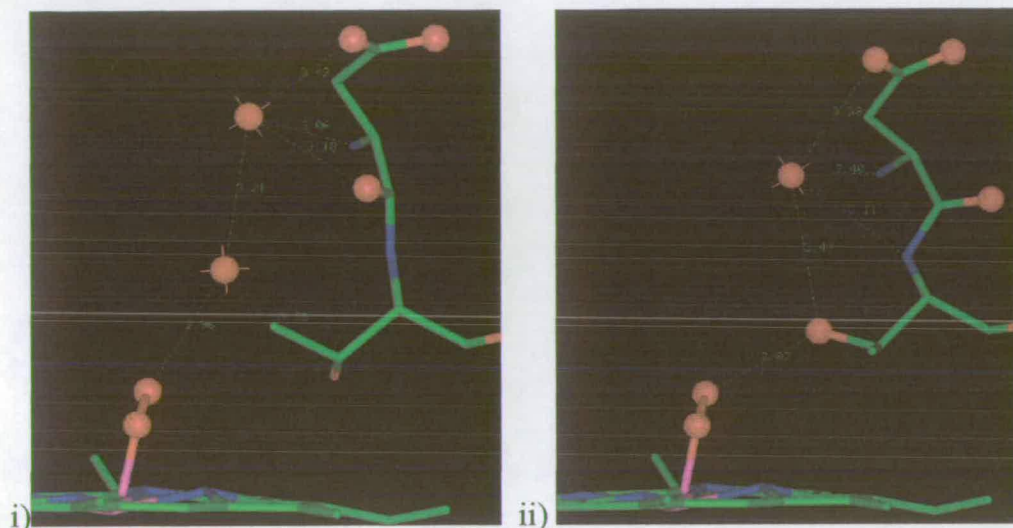


Figure 7.13: Showing the position of the proposed water molecules for the Sligar model (i) and our model (ii).

This results in slight differences in the location of the proposed water molecules, figure 7.14.

The only case when a water is observed in this proposed site is the 2-phenyl imidazole inhibitor structure.<sup>16</sup> It is important to note that not only does this water bind to Glu 251 (the carboxylate switch), but the backbone carbonyl of Glu 251 has rotated to the orientation observed in our structure. There are few instances where no water 687 is present in the binding site of previous crystal structure (e.g. camphane and adamantane and substrate free), however in these cases a binding site water is present, perhaps as a consequence of water 687 moving into the binding site.



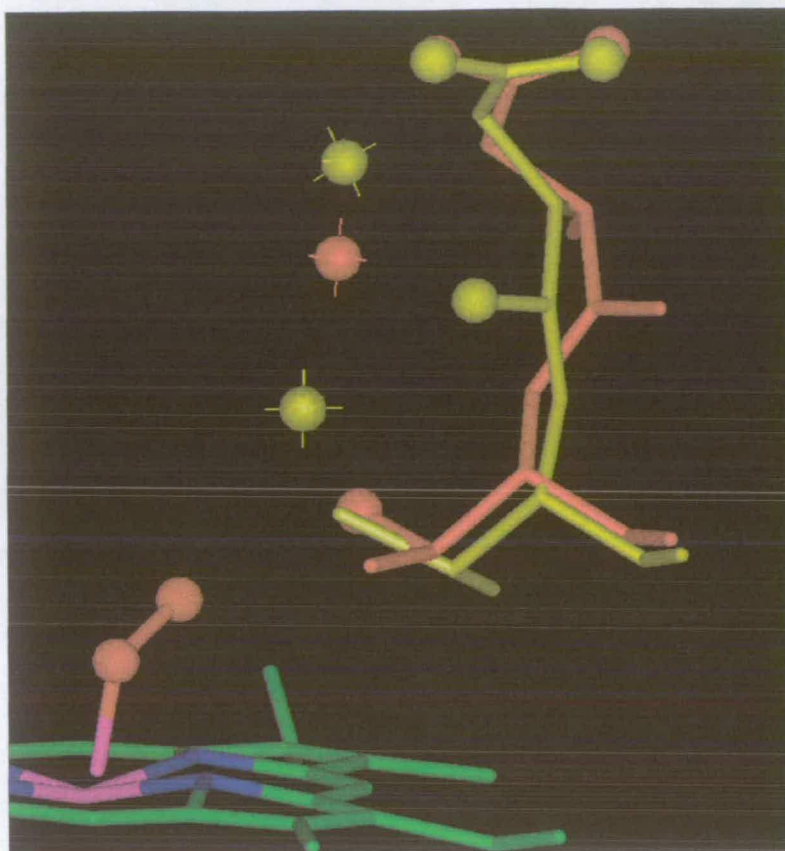


Figure 7.14: The differences in the proposed proton delivery mechanisms, showing Thr 252, Glu 251 the proposed water molecules and the dioxygen (red) bound to the heme (green/blue). The Sligar model is in yellow, <sup>4</sup> our alternate model is in red.

Thus we can propose a possible mechanism of proton transfer based on these disruptions to the binding site:

- After substrate binding and electron transfer, molecular oxygen binds to the reduced heme.
- After the second electron transfer Thr 252 rotates to H-bond to the bound dioxygen/ Glu 251 backbone rotates and Wat 687 enters the binding site (in unknown sequence).
- The carboxylate switch can then transfer a proton to this water which, in turn, transfers it to the bound dioxygen.
- A further proton transfer to the dioxygen.

- The double protonated oxygen is expelled as water leaving the reactive iron (V) species.

Thus the 4-hydroxydiphenylmethane structure might have given us an important insight into the cytochrome P-450 binding site and a possible explanation for the role of the Thr 252 residue in the catalytic cycle. The only questions unanswered would therefore be whether there is one or two water molecules in the pathway and how water 687 enters the binding site.

## **7.2 Common Features of the Y96A Mutant Structures.**

### **7.2.1 The Hydrophobic Binding Region (the 'Phenyl Pocket').**

In the wild-type structure a close aromatic interaction exists between Phe 87, Phe 98 and Phe 193 to the Tyr 96, figure 7.15.

It was hoped that the Y96A mutation would maintain this interaction by replacing the tyrosine phenyl ring with one from the substrate. This assumption was the basis for some of the initial modelling (chapter 8), however it is clear from the camphor bound structure that when no substrate phenyl ring is present, phenylalanines 87, 98 and 193 rearrange to a new aromatic stacking arrangement (to minimize the residues around the hole left by the removal of the tyrosine). When the substrate contains a phenyl ring this aromatic stacking arrangement is maintained, with the substrate aromatic ring supporting this new configuration, figure 7.16.



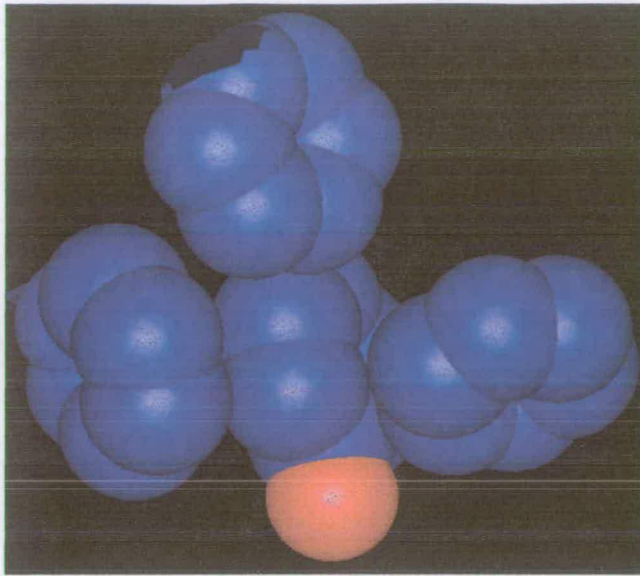


Figure 7.15: A CPK representation of the phenyl rings surrounding tyrosine 96, showing, clockwise: Phe 87; Phe 193 and Phe 98 with Tyr 96 in the centre.

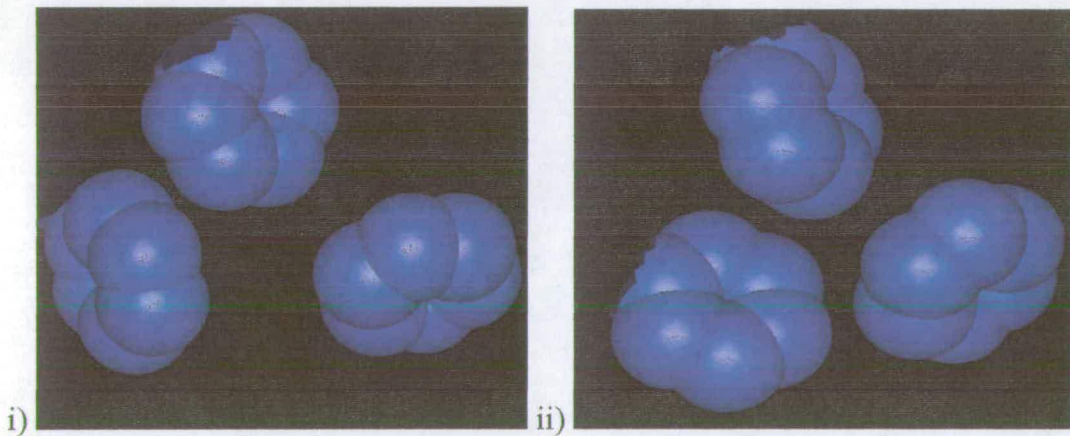


Figure 7.16: Comparison of the arrangement in the Y96A mutant of the phenylalanines around residue 96 for the proposed model (i) and the actual crystal structure (ii).

Comparison of the temperature factors for the aromatic residues (table 7.2) indicates that the Y96A camphor bound mutant orientation is the more stable, with the wild-type and Y96A- 4-hydroxydiphenylmethane structure being of similar stability.

Table 7.2: Comparison of the temperature factors of wild-type and Y96A camphor and 4-hydroxydiphenylmethane in the region around Tyr 96.

residue/ structure	Temperature factor ( $\text{\AA}^2$ )				
	Phe 87	Phe 98	Phe 193	Tyr 96	protein average
wild-type (camphor bound)	15.0	20.2	31.4	16.62	19
Y96A mutant (camphor bound)	12.0	12.5	13.3	-	17
Y96A mutant (4-hydroxy diphenylmethane bound)	17.0	19.4	16.6	-	22

The temperature factors indicate that the mutation has made this area of the structure less mobile than the wild-type enzyme. The reason for this reduction of mobility is probably the removal of the strain on the residue 96 backbone which causes the strained distortion at the end of the B' helix, figure 7.18. When phenyl substituted substrates are bound, the region has greater mobility (as mobile as the wild-type orientation). The Y96A camphor bound structure has greater potential mobility than the wild-type structure, therefore the lower temperature factor must be due to greater stability of this region caused by the Y96A mutation.

### 7.2.2 First Potassium Binding Site.

Peterson discovered in the 1970's that there is a link between monovalent cation concentration and the efficiency of substrate binding.<sup>18</sup> The role of this cation has been explored both by biochemical studies and by crystallography.<sup>19,20</sup> This cation has been shown to be potassium,<sup>19</sup> although its role is unclear. There is a relationship between the cation and the state of the Tyr 96 residue; mutation from tyrosine to phenylalanine increases the level of potassium required to induce maximum

binding.<sup>19,21</sup> Recently Sligar and Polous have identified the previously hypothesized<sup>9</sup> location of the cation binding site.<sup>4</sup> This binding site is formed between the backbone carbonyl groups of residues Gly 93, Glu 94, Ala 96 and Glu 84 and two water molecules, figure 7.17.

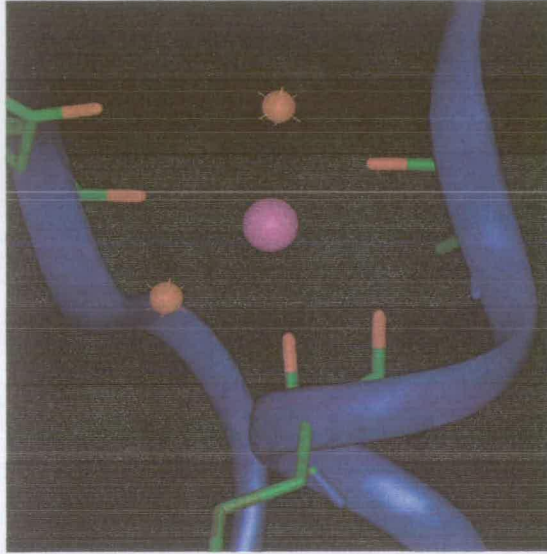


Figure 7.17: The orientation of the potassium ion in the wild-type structure, showing from top, clockwise, Phe 96, Glu 94, Gly 93, Phe 87

It is believed that this potassium ion stabilizes an unusual orientation of the Tyr 96 and the end of a helix starting at residue 89 (helix B'), figure 7.18.

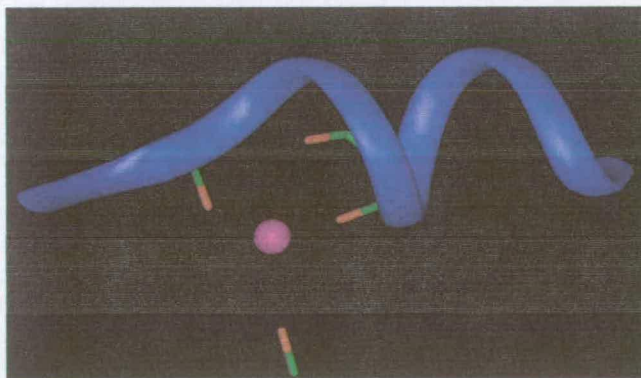


Figure 7.18: The B' helix showing the potassium ion (purple) and how it relates to the unusual termination of the B' helix.

In the Y96A mutant the potassium ion exists in this site, in an identical position to the wild-type enzyme. The density can be assigned as potassium from the  $\sigma$ -weighted 2Fo-Fc map which indicates it is too big to be a water molecule (SHELX refines the atom to 0 occupancy as an oxygen, indicating an error in the assignment). The oxygen-cation distances are also consistent with potassium, rather than calcium or any other similarly sized cation. The temperature factor of the potassium is lower here ( $14 \text{ \AA}^2$ ) than the wild-type cryo structure, 5cp4, ( $22 \text{ \AA}^2$ ) and comparison of the average temperature factors for the side-chains involved in the potassium binding shows similar differences ( $11 \text{ \AA}^2$  for our structure compared to  $23 \text{ \AA}^2$  for the wild-type cryo-structure, table 7.3).

Table 7.3: Temperature factor of the potassium and its bound residues.

	WT-cryo <sup>4</sup>	Y96A
K	22	14
84	22.0	12.8
93	19.3	11.5
94	23.7	12.2
96	21.0	13.3
total protein	19	17

This decreased temperature factor, in this region, could indicate that without the camphor tyrosine hydrogen bond, the tyrosine (now alanine) backbone, adopts a more stable orientation. The overall temperature factors of the total protein backbone are comparable for both structures ( $\sim 16 \text{ \AA}^2$ ) therefore the  $12 \text{ \AA}^2$  difference in average backbone temperature factors is probably a real reflection of decreased mobility of the potassium bound residues and more particularly the B' helix. Thus the B' helix has adjusted its thermal motion due to the mutation which has resulted in a less mobile orientation of the both the cation and the B' helix.

It has been shown previously that a dependence between potassium concentration and substrate binding affinity exists, even after removal of the tyrosine



H-bond to the camphor carbonyl (Y96F and camphor), indicating that the role of the potassium bound ligand to residue 96 goes far beyond that of stabilizing the camphor tyrosine hydrogen bond.<sup>3,14,19,21-23</sup> Since three octahedrally co-ordinated O donors are on the B' helix (89-96) it is possible that the role of the potassium is in controlling the movement of the B' helix. This small helix forms a significant portion of the proposed entrance channel and therefore potassium may play a role in controlling the size of the substrate entrance channel and consequently cause the observed effects on substrate binding.

### 7.2.3 A Second Potassium Binding Site.

Another binding site for a potassium ion can be identified in the Y96A structure where potassium lies in a surface groove near the N-terminus, figure 7.18. This potassium binds the backbone carbonyl groups of residues Pro 15, Pro 16, Val 18, side-chain oxygen OE1 of Glu 20 and a surface water. This potassium exhibited similar electron density characteristics to the first potassium site. This potassium was identified as water 706 in the cryo structure (5cp4)<sup>4</sup> but was not present in the other P-450cam structures.

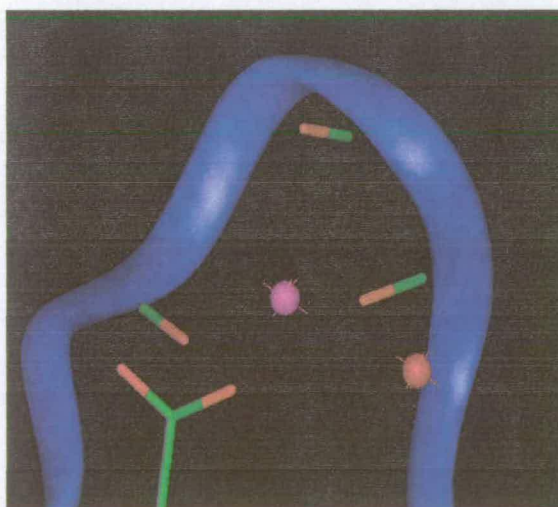


Figure 7.18: Second potassium binding site. The potassium is in purple, H-bonded residues are indicated in red, starting left and going anticlockwise: Glu 20, Val 18, Pro 16 and Pro 15.

The structural significance of this potassium is not known, although it appears to stabilize the last  $\beta$ -turn at the N terminus, figure 7.19.

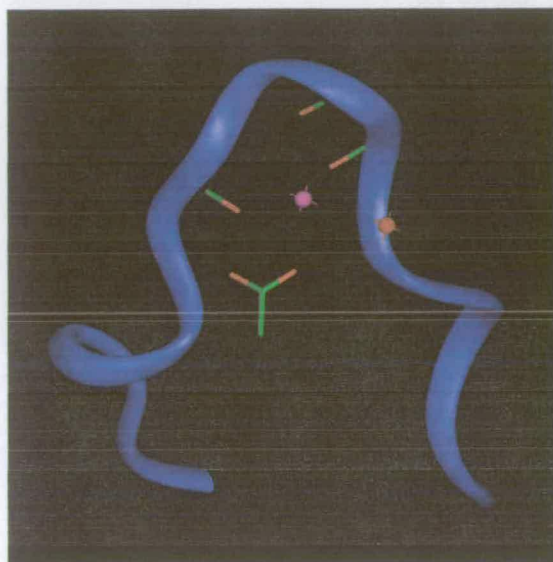


Figure 7.19: Showing the location of the second potassium binding site in relation to the  $\beta$ -turn.

This potassium has only been observed in the cryo-structures; at room temperature these residues may be too mobile to bind potassium. This potassium thus might be an artifact of the freezing process (although the room temperature structures have significantly fewer assigned water molecules). The overall temperature factors for our camphor structure ( $17.9 \text{ \AA}^2$ ) is comparable to the wild-type room temperature structure, 2ccp, ( $18.8 \text{ \AA}^2$ ). However, the average temperature factor for the residues binding the potassium are  $17.6 \text{ \AA}^2$  for our structure compared to  $29.0 \text{ \AA}^2$  for the free residues in the wild-type structure. This shows that the potassium has the effect of stabilizing the orientation of the N-terminus and is not just an artifact of the crystallization.

#### 7.2.4 Aqua Ligand.

With no substrate bound, the heme iron exists as a hexacoordinate low spin species. Addition of substrate displaces the water and gives pseudo-pentacoordinate

geometry onto the heme, resulting in a high-spin species, as characterized by the shift in  $\pi$ - $\pi^*$  bands in the UV spectrum ('Soret' shift). It is known that for non-natural substrates, e.g. camphane, norcamphor and adamantane, a state between high and low spin can appear.<sup>3</sup> It has been postulated from the crystallographic evidence that the water may be present 100 % of the time, but in two forms, as both hydroxyl and water. It is thus proposed that it is these two distinct forms that cause of the differences in the spectra.<sup>14</sup> Recent *ab initio* studies have shown that the mixed spin state that occurs when substrates are bound is an effect caused by the shortening of the iron sulfur bond and not by the nature of ligated water.<sup>24-26</sup> This combines with recent analysis of the P-450 system, which suggests that the shift of water from the iron is not sufficient to induce the change in spin state and the substrate induced changes to the heme environment plays a significant role.<sup>27,28</sup>

During the recent *ab initio* analysis, a number of key conclusions were missed. The shift from high to low spin at iron sulfur bond lengths of over 2.2 Å is governed solely by the presence or absence of the aqua ligand and is not related to the length of the bond. However when the iron sulfur bond is reduced towards 2.05 Å a mixed state is induced between the high and low spin species. Theoretical calculations showed that this mixed state is independent of the presence of the aqua ligand (although the aqua ligand may be inducing the shortening of the bond). All the camphor analogues that have these shorter iron sulfur distances exhibit mixed spin states.<sup>14</sup>

From previous data on camphor analogues it is clear that more 'mobile' substrates will allow water to enter the binding site and thus potentially change the iron-sulfur bond. Therefore the length of the iron-sulfur bond is dependent on the substrate and its effect on the heme environment and not on the nature of the aqua ligand itself. The factors that control the substrate/heme relationship are therefore more complicated than simple displacement of the water to induce a change in spin state. A combination of substrates hydrophobicity, steric bulk and mobility probably affect the heme environment and consequently the spin state of the iron.



#### 7.2.4.1 Camphor Bound Structural Water.

Since camphor in the Y96A mutant has only 48 % high-spin, it was crucial to measure the correct iron-sulfur distance in the camphor bound structure and to establish if this mixed spin state is caused by shortening of the iron-sulfur bond or by some other mechanism. The result of a fully optimized iron-sulfur distances of 2.34 Å is consistent with the camphor bound wild-type structure. This suggests that in the mixed spin states of the Y96A camphor bound structure is a direct result of the occupancy of the water ligand and not caused by shortening of the iron sulfur bond. This is confirmed by the refined occupancy of the water, at 58 %. This is the first of the P-450cam structures to have a partial water occupancy and would support the idea that mixed spin states can be caused by shortening of the iron-sulfur bond or partial occupancy of the aqua ligand.

#### 7.2.4.2 Phenylcyclohexane Bound Structural Water.

The iron sulfur distance for the phenylcyclohexane bound structure is 1.93 Å which is longer than for the camphor bound structure but consistent with other previous bound water such as adamantane (1.95 Å).<sup>14</sup> The similarity of the distance between the iron and the bound water to the wild-type adamantane structure was of interest as the adamantane structure has the anomalous data of 100 % water occupancy combined with a high percentage high-spin. This is indeed the case for the phenylcyclohexane structure which has a percentage high-spin of 85 % and a water occupancy of 65 %. The adamantane structure additionally has a iron sulfur distance of 2.14 Å, shorter than the camphor bound structure, but comparable to the phenylcyclohexane bound structure of 2.16 Å. It can thus be concluded that unlike the camphor bound structure the phenylcyclohexane bound water does induce a change in the spin state due to the shortening of the iron sulfur bond.

#### 7.2.4.3 Water Occupancy of Other P-450cam Y96A Mutant Structures.

For the obvious reason that the ligand ligates to the heme, the 4-hydroxydiphenylmethane structure has no bound water. The bicyclohexyl structure showed no sign of water in the binding site. As for the diphenylmethane and substrate **49** the data suggests the poor substrate solubility has caused a partial occupancy of the ligands. Thus because of the partial occupancy of the water ligands it is difficult to determine if any bound water is present in the presence of the substrate.

#### 7.2.5 Uncoupling.

When unnatural substrates are used with P-450 monooxygenases, uncoupling can occur resulting in the production of peroxide rather than hydroxylated substrate. Accessibility of water to the active oxygen species during the catalytic cycle is believed to be the primary causal factor. Uncoupling is a problem for more commercial applications of P-450 monooxygenases as it utilizes NADH in the production of the peroxide. A number of substrates for the Y96A mutant have been studied over the years and they all indicate that there is a high degree of uncoupling for this mutant.<sup>29-32</sup> This may in part be due to greater water accessibility to the binding site due to poor substrate enzyme interactions, (especially when relying on the phenyl ring to provide substrate specificity). Another factor might be the effect of the Tyr 96 removal, as it has been postulated that the Y96 residue pushes camphor towards the iron centre maintaining a close iron C-5 distance.<sup>33</sup> A possible structural cause of uncoupling has been found in the Y96A mutant.

With the removal of the Tyr96-Thr101 hydrogen bond, Thr 101 has rotated 105 ° creating a hydrophilic pocket with the Thr 101 OE1 and a heme propionate group. This pocket allows binding of an extra water molecule, not present in any published structures. Although this water molecule is 8 Å from the iron, the presence of this water in the binding site may explain the increases in uncoupling with the Y96A mutant. Simple mutation of the threonine to valine could remove this problem without significantly altering the binding site volume.

### 7.3 Conclusions.

The six substrate bound structures of the P-450cam Y96A mutant enzyme have yielded a number of important results.

Firstly the camphor bound structure has indicated the elegant way in which the binding site has been designed for camphor and how one mutation cannot significantly disrupt the binding specificity of the enzyme towards its natural substrate.

Analysis of the interaction of the other substrates with the binding site revealed the subtle relationship between the 'phenyl pocket' and the aromatic rings on the substrates. Bicyclohexyl showed that non-aromatic substrates can interact with the enzyme in a similar way.

The 4-hydroxydiphenylmethane bound structure also provided an insight in to the relationship between the binding site and the possible mechanism of proton transfer.

Overall, the six structures showed how the potassium binding site is stabilized by the mutation of tyrosine to alanine. In addition to this a new potassium binding site was discovered.

These six structures also provided a possible explanation for the increased peroxide production witnessed in the P-450cam Y96A mutant turnover.

## 7.4 References.

- 1) Harris, D.; Loew, G. *Journal of the American Chemical Society* **1995**, *117*, 2738-2746.
- 2) Atkins, W. M.; Sligar, S. G. *Journal of the American Chemical Society* **1989**, *111*, 2715-2717.
- 3) Atkins, W. M.; Sligar, S. G. *Journal of Biological Chemistry* **1988**, *263*, 18842-18849.
- 4) Vidakovic, M.; Sligar, S. G.; Li, H. Y.; Poulos, T. L. *Biochemistry* **1998**, *37*, 9211-9219.
- 5) Cuppvickery, J. R.; Poulos, T. L. *Nature Structural Biology* **1995**, *2*, 144-153.
- 6) Raag, R.; Martinis, S. A.; Sligar, S. G.; Poulos, T. L. *Biochemistry* **1991**, *30*, 11420-11429.
- 7) Harris, D. L.; Loew, G. H. *Journal of the American Chemical Society* **1994**, *116*, 11671-11674.
- 8) Kimata, Y.; Shimada, H.; Hirose, T.; Ishimura, Y. *Biochemical and Biophysical Research Communications* **1995**, *208*, 96-102.
- 9) Poulos, T. L.; Finzel, B. C.; Howard, A. J. *Journal of Molecular Biology* **1987**, *195*, 687-700.
- 10) Gerber, N. C.; Sligar, S. G. *Journal of the American Chemical Society* **1992**, *114*, 8742-8743.
- 11) Gerber, N. C.; Sligar, S. G. *Journal of Biological Chemistry* **1994**, *269*, 4260-4266.
- 12) DiPrimo, C.; Deprez, E.; Sligar, S. G.; Hoa, G. H. B. *Biochemistry* **1997**, *36*, 112-118.
- 13) Benson, D. E.; Suslick, K. S.; Sligar, S. G. *Biochemistry* **1997**, *36*, 5104-5107.
- 14) Raag, R.; Poulos, T. L. *Biochemistry* **1991**, *30*, 2674-2684.
- 15) Raag, R.; Li, H. Y.; Jones, B. C.; Poulos, T. L. *Biochemistry* **1993**, *32*, 4571-4578.

- 16) Poulos, T. L.; Howard, A. J. *Biochemistry* **1987**, *26*, 8165-8174.
- 17) Wade, R. C. *Journal of Computer-Aided Molecular Design* **1990**, *4*, 199-204.
- 18) Peterson, J. A. *Arch. Biochem. Biophys.* **1971**, *144*, 678-693.
- 19) Deprez, E.; Diprimo, C.; Hoa, G. H. B.; Douzou, P. *FEBS Letters* **1994**, *347*, 207-210.
- 20).
- 21) Diprimo, C.; Hoa, G. H. B.; Douzou, P.; Sligar, S. *Journal of Biological Chemistry* **1990**, *265*, 5361-5363.
- 22) Schulze, H.; Hoa, G. H. B.; Helms, V.; Wade, R. C.; Jung, C. *Biochemistry* **1996**, *35*, 14127-14138.
- 23) Bass, M. B.; Paulsen, M. D.; Ornstein, R. L. *Proteins-Structure Function and Genetics* **1992**, *13*, 26-37.
- 24) Segall, M. D.; Payne, M. C.; Ellis, S. W.; Tucker, G. T.; Boyes, R. N. *Xenobiotica* **1998**, *28*, 15-20.
- 25) Segall, M. D.; Payne, M. C.; Ellis, W.; Tucker, G. T.; Boyes, N. *Chemical Research In Toxicology* **1998**, *11*, 962-966.
- 26) Segall, M. D.; Payne, M. C.; Ellis, S. W.; Tucker, G. T.; Boyes, R. N. *European Journal of Drug Metabolism and Pharmacokinetics* **1997**, *22*, 283-289.
- 27) Harris, D.; Loew, G. *Journal of the American Chemical Society* **1993**, *115*, 8775-8779.
- 28) Harris, D.; Loew, G. *Journal of the American Chemical Society* **1993**, *115*, 5799-5802.
- 29) Wong, L. L.; Westlake, C. G.; Nickerson, D. P. *Structure and Bonding* **1997**, *88*, 175-207.
- 30) Stevenson, J. A.; Westlake, A. C. G.; Whittock, C.; Wong, L. L. *Journal of the American Chemical Society* **1996**, *118*, 12846-12847.
- 31) Stevenson, J. A.; Bearpark, J. K.; Wong, L. L. *New Journal of Chemistry* **1998**, *22*, 551-552.
- 32) Bell, S. G.; Rouch, D. A.; Wong, L. L. *Journal of Molecular Catalysis B-Enzymatic* **1997**, *3*, 293-302.

33)Poulos, T. L.; Cupp-Vickery, J.; Huiying, L. *The Structure of Cytochrome P-450cam*; 2nd ed.; DeMontellano, P.R.O. Ed.; Plenum Press, New York and London, 1995, pp 125-150.

## 8 Molecular Modelling.

---

### 8.1 Introduction.

In recent years, molecular modelling has been used for a wide range of applications within cytochrome P-450 research. Unsolved structures have been hypothesized using X-ray crystal structures of known P-450s<sup>1-5</sup> and available chemical databases are used to screen for potential substrates.<sup>6-10</sup> Substrate entry into the binding site has been modelled,<sup>11</sup> as well as calculations of the maximum binding pocket space for potential substrates.<sup>12</sup> Interactions between P-450 monooxygenases and their electron transfer co-proteins can also be modelled.<sup>13-15</sup> Modelling studies have led to the creation of many mutant enzymes,<sup>16</sup> in an attempt to understand the importance of various binding site residues, for example mutants Y96F,<sup>17</sup> T252A,<sup>18-20</sup> D251N<sup>21</sup> Y96H, F87W and V396W.<sup>22</sup>

Without an available crystal structure for the Y96A mutant, a crude model was created by deletion of the tyrosine side chain to give alanine. It was hoped that substrates with an aromatic ring would interact with the mutant protein (due to, energetically favourable  $\Pi$ -stacking within the aromatic pocket between phenylalanine residues Phe 87 and Phe 98). This gives a potential starting orientation for the substrate modelling, i.e. in the position previously occupied by the phenyl ring of the tyrosine side-chain.



If the aromatic ring fits into this aromatic pocket, the orientation of the substrate within the binding site will be limited by the phenylalanine residues on either side, (Phe 87 and Phe 98) and by the interactions with other residues of the binding site. This meant that a crude attempt at modelling the substrates could be made without the need for extensive computation. This was accomplished by placing the aromatic ring of the substrate in the aromatic pocket, in the position previously occupied by the tyrosine 96 aromatic ring.

## **8.2 Confirmation of the Model by Modelling Known Substrates.**

Diphenylmethane and phenylcyclohexane are known substrates for the P-450cam Y96A mutant and it was hoped that modelling them in the binding site could confirm the validity of the model system. Diphenylmethane was modelled with one phenyl ring in the location previously occupied by the tyrosine and the other positioned to avoid unfavourable interactions with the binding site residues, figure 8.1.

It was found that the orientation of the second phenyl ring placed the C4 carbon in the closest hydroxylation position to the heme. This agrees with the experimental result which produces 4-hydroxydiphenylmethane as the sole product of the biotransformation with the Y96A mutant.<sup>23</sup> The same process was applied to phenylcyclohexane and it was found that both the C4 hydrogens and the trans (axial) C3 hydrogen were in close proximity to the heme. This is again in agreement with the experimental result which yields 4- and 3-hydroxy products.<sup>24</sup> It appeared from these observations that the mutant model could adequately predict substrates and hydroxylation sites.

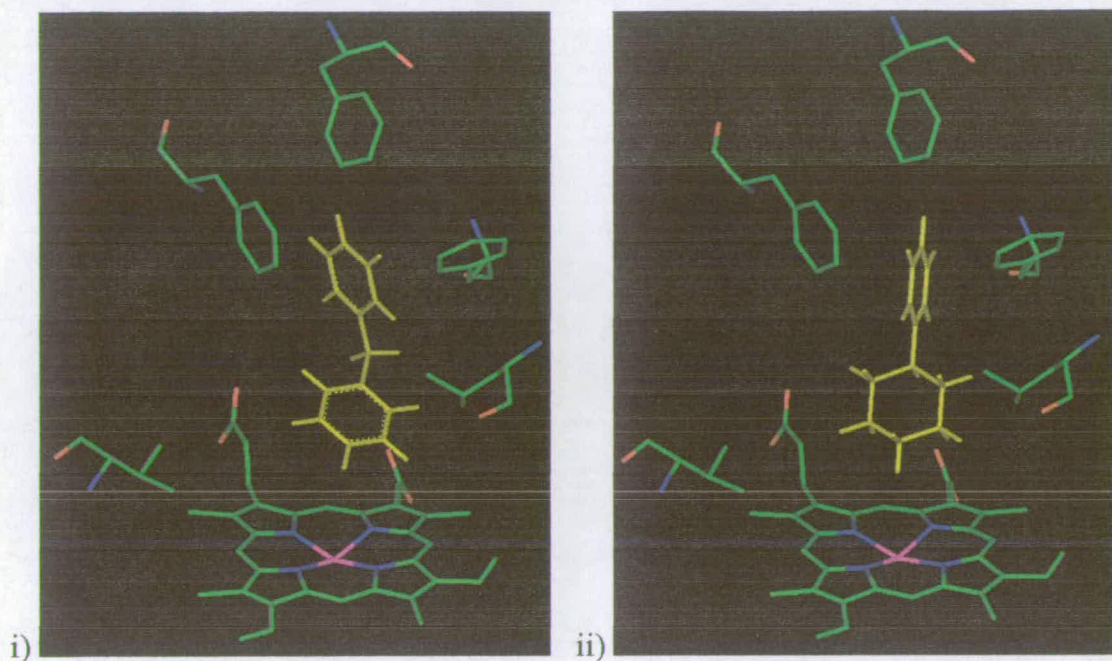


Figure 8.1: Diphenylmethane (i) and phenylcyclohexane (ii) in the binding site model.

### 8.3 Modelling of Novel Substrates.

#### 8.3.1 Diclofenac.

Diclofenac [37] is an important anti-inflammatory agent and it has been suggested that hydroxylation of either of the aromatic rings could enhance biological activity. Selective hydroxylation of diclofenac is chemically difficult and it was hoped that hydroxylation of diclofenac or a synthetic intermediate might be achieved using the P-450cam mutant enzyme system. To this end, diclofenac and some of its synthetic precursors (58, 40, 59, 60, 61) were modelled in the binding site of the Y96A mutant, figure 8.2.

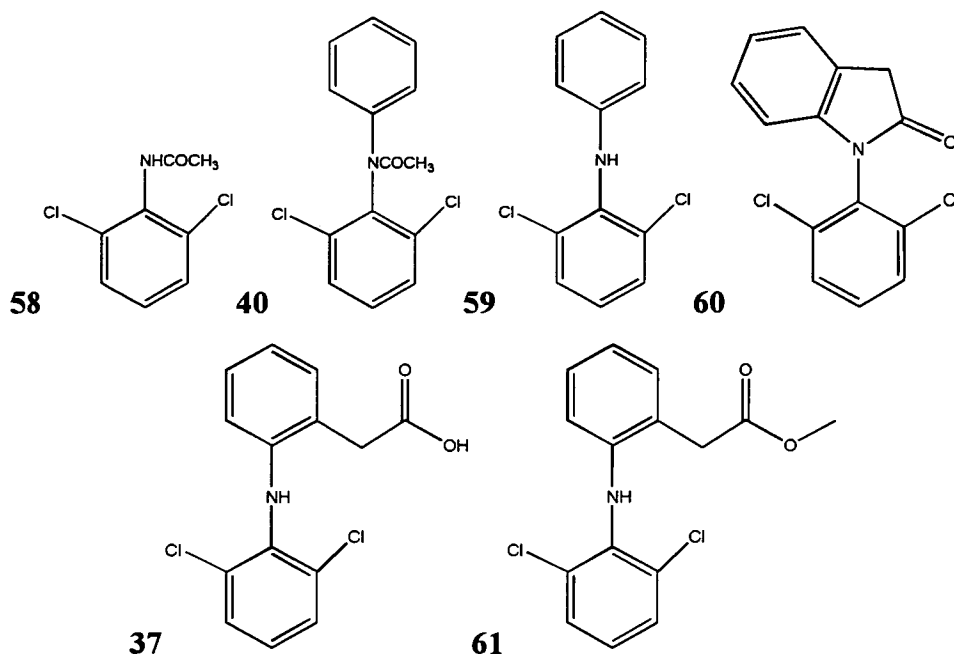


Figure 8.2: Diclofenac [37] and some of its synthetic precursors.

Compound **58** is too small to fit into both the aromatic pocket and above the heme. It is thought that the free amine would preferentially bind to the iron therefore making turnover impossible. For the other compounds, the chlorinated ring was found to have unfavourable Van der Waals interactions with the protein, so the non-chlorinated ring of each compound was modelled in the aromatic pocket.

Compounds **60**, **37**, **61** are all predicted to be too big to fit into the binding site. When the aromatic ring is orientated into the aromatic pocket there are unfavourable Van der Waal contacts with the substrate chlorines and the surrounding residues. These three molecules should only bind if there is a good degree of flexibility in the amide linkage between the two aromatic rings, which is unlikely.

Compounds **59** and **40** appear to fit into the binding site but compound **59** has a different orientation of the two aromatic rings (perpendicular to each other) which does not lead to favourable Van der Waals interactions between the chlorine and some of the binding site residues. Compound **40** however binds well. The orientation of the rings is better due to the coplanarity of the aromatic rings with the amide. A small rotation of the chlorinated ring gives almost a perfect fit in the binding site,

figure 8.3i. There is also a potential hydrogen bond between the carbonyl on the diclofenac and the backbone carbonyl of phenylalanine 98, figure 8.3ii.

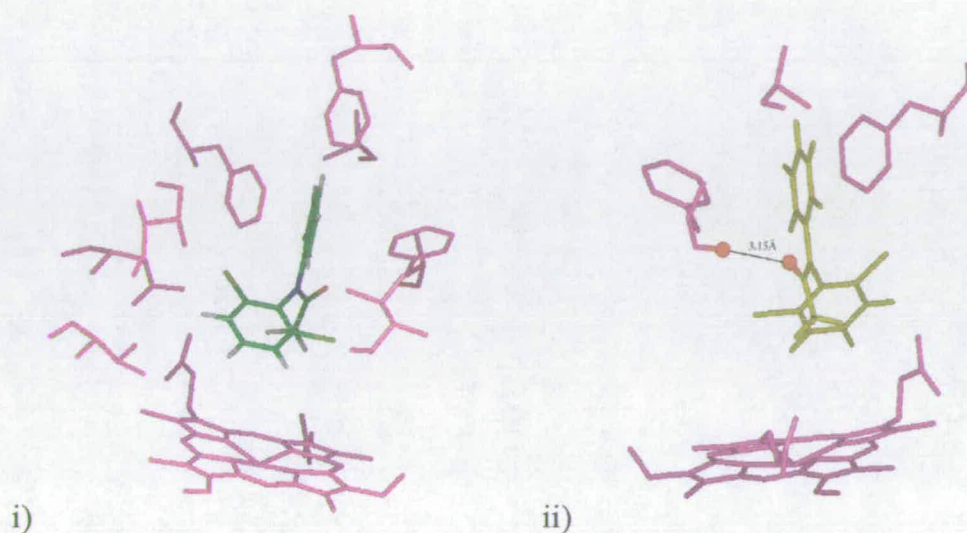


Figure 8.3: i) Model of compound **40** (in green) in the P-450cam Y96A binding pocket, ii) Model of compound **40** (yellow) showing possible hydrogen bond with phenylalanine 98.

Binding studies have indicated that compound **40** does bind whereas the other compounds do not show any signs of binding. In this case, the modelling correctly predicted the binding of potential substrates to the P-450cam mutant enzyme.

### 8.3.2 Studies of 2,3-Epoxy cyclohexyl-1-benzoate.

Another set of compounds, from previous work with substrates containing benzoyl attached to a non-aromatic six membered ring,<sup>25</sup> were modelled into the active site of the Y96A model. The substrates were four diastereomers of 2,3-epoxycyclohexyl-1-benzoate(**61-64**), figure 8.4.

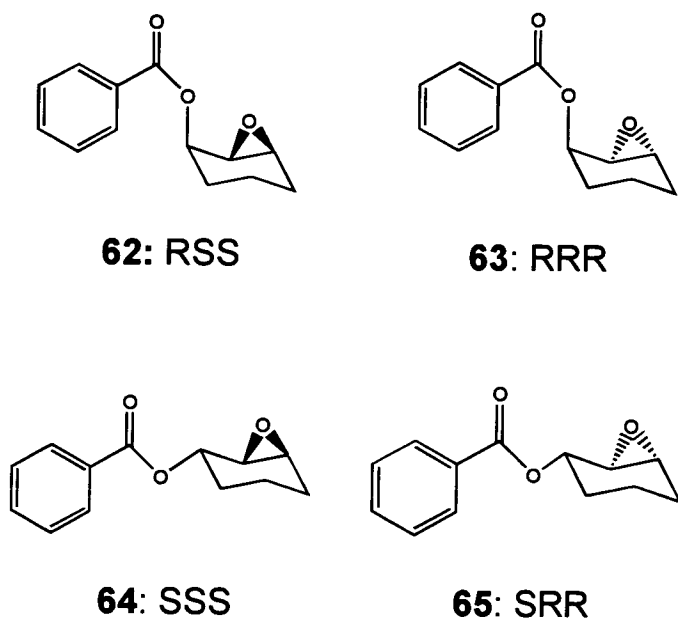


Figure 8.4: The four diastereomers of 2,3-epoxycyclohexyl-1-benzoate.

All four diastereomers were modelled to discover if the enzyme could distinguish between them. In all cases the 4- position on the aliphatic ring was the favoured location for hydroxylation. There was very little difference in orientation of the substrate in each case, figure 8.5i.

For the RRR and SSS form (compounds **63** and **64**), the epoxide was in close contact to the iron, figure 8.5ii. This might cause the epoxide to react with the iron centre, if it does not, it may still be possible to hydroxylate them. For the two enantiomers RSS (**62**) and SRR (**65**), the epoxide is not in close contact with the iron centre and therefore could be hydroxylated. Taking into account the movement of the substrate within the binding site, the major product in all cases should be the 4-hydroxylated product but minor products might also be produced. This was later confirmed by experimental results, which gave no hydroxylated products.



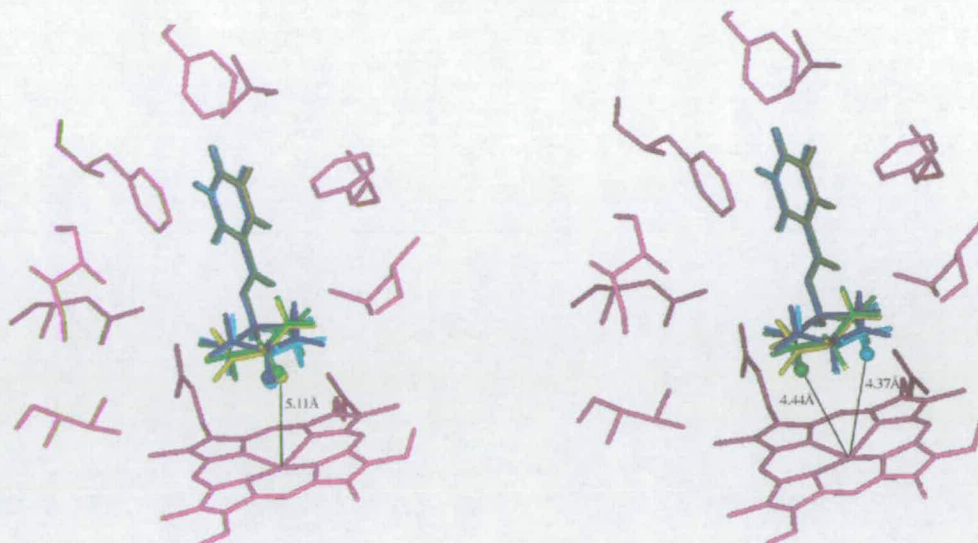


Figure 8.5: i) All four diastereomers (compounds **62-65**) modelled in the active site of the P-450cam Y96A mutant. The heme distance from the 4-axial position on the substrates is marked. ii) All four diastereomers (compounds **62-65**) modelled in the active site of the P-450cam Y96A mutant. The heme distance from the axial epoxides of compounds **63** and **64**. Colour scheme: Purple for Y96A enzyme, compounds: **62** dark blue; **63** light blue; **64** green and **65** yellow.

In this case it was not possible to distinguish significant differences in the binding capacity of the compounds (**62-65**). It is possible that a more refined model using Molecular Dynamics (MD) might find greater differences in the binding of these compounds. In these two distinct cases, both the potential benefits and limitation of molecular modelling have been demonstrated.

#### 8.4 Modelling of Previously Studied Substrates.

A number of *Cbz*-substituted piperidines have been previously studied as substrates for the Y96A mutant, figure 8.6.<sup>26</sup>

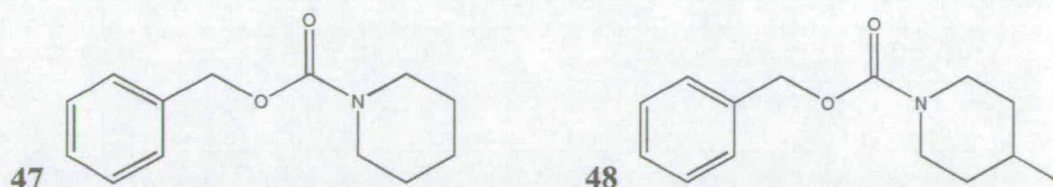


Figure 8.6: Two *Cbz*-substituted piperidines used in the modelling study.

From experimental results it was known that both substrates bind to the enzyme. However only substrate **47** produces a significant amount of hydroxylated product (5 %), compound **48** produced less than 1 % of product. On the other hand, binding studies indicated that compound **48** binds better than compound **47**. In order to investigate this paradigm, modelling was performed with both substrates by placing the phenyl ring of the *cbz* group in the aromatic pocket, figure 8.7.

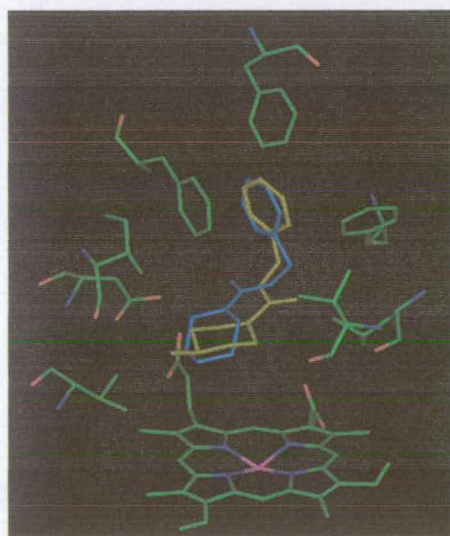


Figure 8.7: Picture of compounds **48** (yellow) and **47** (blue) in the Y96A mutant model.

The modelling studies gave almost identical orientations for both compounds, it was not therefore possible to establish why these two compounds had different binding and turnover properties. This demonstrates that the modelling studies are clearly not sufficient to fully predict the binding properties of compounds for the cytochrome P-450cam Y96A mutant.



Due to these areas of uncertainty about the Y96A model it was decided a more thorough investigation of the substrate binding was necessary. The modelling studies discussed so far were performed with a Y96A structure based on manipulation of the wild-type enzyme, not on the actual structure. In addition to this, even if the model of the Y96A structure is accurate, the orientation of the substrates in the binding site is unknown. For these reasons it was decided to investigate the structure of the P-450cam Y96A mutant and how substrates bind to it. To date X-ray crystallography is the only technique able to solve the structure of proteins of this size (~45 kDa). It was therefore decided that the only way in which we could satisfactorily explain the substrate binding to the Y96A mutant was to solve a number of substrate bound crystal structures.

## **8.5 Remodelling the Substrates in the Y96A Crystal Structure.**

When modelling substrates in the crystal structure similar assumptions were made to the previous modelling studies, i.e. using the location of the phenyl ring from known substrates to model the location of the aromatic handle of new substrates.

### **8.5.1 Camphor and Adamantane Analogues (2, 10, 23-28).**

All the camphor and adamantane molecules can be easily placed into the binding site, however, there is no way to assign the possible orientation without a frame of reference. Simple overlay of the molecule with the orientation of the camphor molecule in the crystal structure is not sufficient as the molecule can then freely rotate. Therefore without an aromatic handle or greater steric bulk these substrates can not be modelled accurately.

### 8.5.2 Diphenylmethane Analogues (13, 18, 29, 30).

There was no need to model some of these substrates as the crystal structure of diphenylmethane, phenylcyclohexane and bicyclohexyl bound to the P-450cam Y96A mutant has been solved, although instinctively the cyclohexane rings of phenylcyclohexane and bicyclohexyl would be modelled nearer the heme (i.e. like the diphenylmethane structure) and not towards Val 295 as in their crystal structures. The only substrate in this series without crystal data is amyl benzene, although attempt to co-crystallize it were made. Modelling cannot be easily made here as, after placing the phenyl ring in the aromatic pocket, the pentane chain has a high degree of potential mobility in the center of the binding site and no single orientation can be assigned.

### 8.5.3 Naphthalene Analogues (31-36).

When placing the aromatic groups of these analogues in the phenyl pocket the distance between the other sections of the molecule and the heme is too great for hydroxylation to occur ( $>5$  Å). Since it is known that naphthalene is a wild-type P-450cam substrate it is entirely possible that these substrates all bind in a similar way to the Y96A mutant. A possible mode of interaction could be partial  $\pi$ -stacking between the substrate and the heme, similar to the phenyl radical studied previously.<sup>27</sup> There is plenty of room for this to occur however the energetic feasibility is unknown and no obvious products would be produced.

### 8.5.4 Chlorinated Aromatics (38-40).

It was extremely difficult to place these molecules in the binding site due to unfavourable Van der Waals interactions with the binding site residues. All substrates required extensive manipulation from the energy minimized structure to get them to fit inside the binding site. Even when a binding orientation could be found, no obvious hydroxylation sites could be elucidated.

### 8.5.5 Other Aromatically Substituted Substrates (41-57).

All the other substrates could be successfully modelled in to the binding site by placing the phenyl ring in the aromatic pocket however they all have slight differences depending on the linker between the aromatic and non-aromatic rings.

The compounds that gave the best fit were those with a rigid single atom length linker (e.g. 44-46). The ones with the worst fit were those with the *cbz* group (e.g. 47-49). All the other molecules fit inside the binding site with similar degrees of success. The oxygen substituted rings all fit in an orientation that places the oxygen near the heme, thus preventing turnover.

## 8.6 Conclusions from the Substrate Modelling Studies.

With this basic method of modelling, substrates can be successfully modelled into the binding site of the enzyme. What is clear is that without a aromatic handle it is impossible to predict any kind of information on the binding orientation of the substrates. With the aromatic handle the most that can be concluded is whether or not the substrate is too big for the binding site. Prediction of the site of hydroxylation is not valid as most substrates have a degree of potential movement in the binding site and this cannot be predicted with this methodology. However this technique can be used to access potential substrates before embarking on any practical experimentation. If modelling is to be used to predict substrates a more sophisticated method must be used (see chapter 9).

## 8.7 Substrate Entry in P-450cam Binding Site.

It became apparent from the first crystal structure of P-450cam that there is no obvious entrance channel.<sup>28</sup> Upon closer examination a gap in the helical arrangement, above the heme was found, figure 8.7.

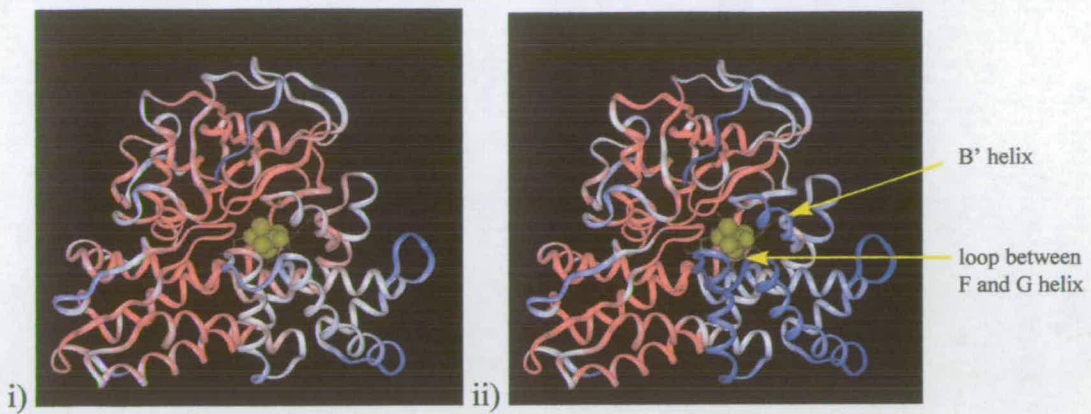


Figure 8.10: The ribbon structure of the protein coloured by temperature factor (red-static, blue-mobile) for substrate bound (i) and substrate free enzyme (ii).

This chain appeared to stick into the entrance channel interfering with residues Phe 193, Phe 87 and Tyr 96 and in this 'open' configuration: the Phe 193 ring points up towards the protein surface; the Phe 87 ring adopting the position formally occupied by the Phe 193 ring. When examining this binding channel it is apparent that entrance pocket is lined with hydrophobic residues, figure 8.11.

A substrate binding must pass the Phe 87, Phe 98 and Tyr 96 residues on its way to the centre of the binding site; this must apply to our Y96A mutant as well. Thus with the mutation to the Tyr 96 residue the substrate entrance should be larger, thus facilitating easier binding of substrate.

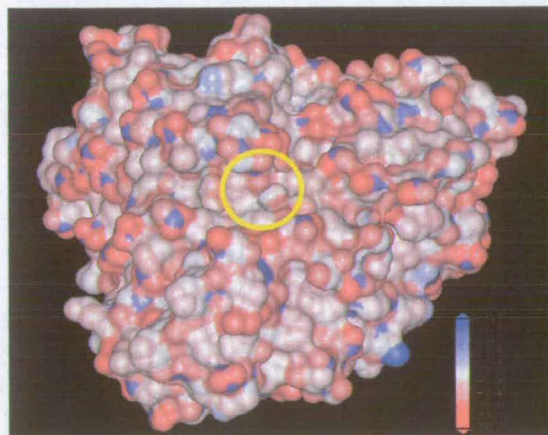


Figure 8.11: Surface of proposed entrance channel of P-450cam, coloured by charge (red-positive, blue-negative). Proposed entrance channel is indicated.



Due to the limited range of thermal motions possible in a protein crystal and the increased size (compared to typical wild type substrates) of our substrates it was decided to co-crystallize these substrates with the enzyme. Co-crystallization of course is only possible because mercaptans are not needed in the crystallization solution, through use of the C334A mutant.

Thus substrate binding with P-450cam is only possible with significant movements of significant portions of the protein. With the availability of the substrate free structure the areas of the protein with increased thermal motions could be identified, figure 8.9.

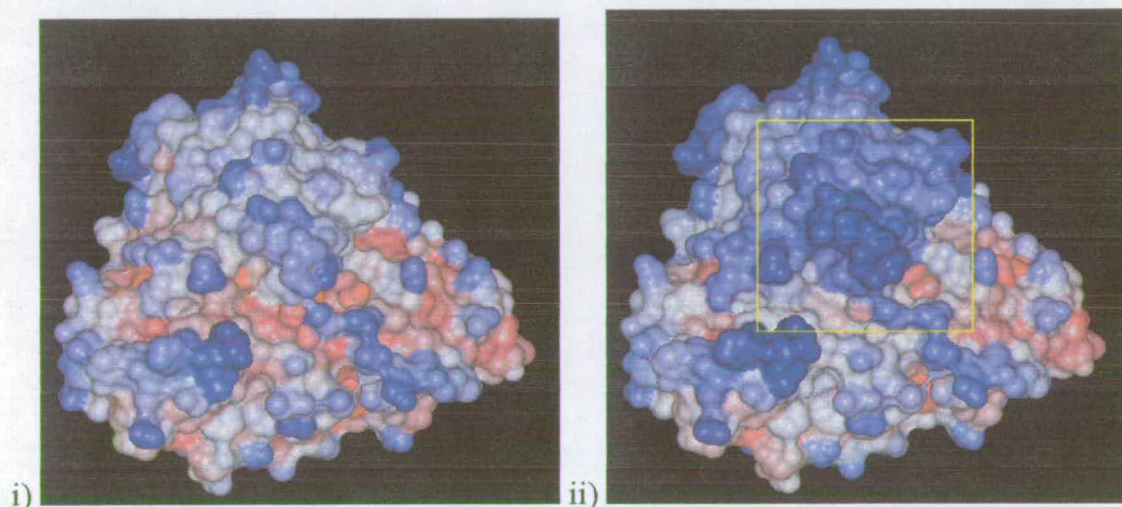


Figure 8.9: The Connolly surface of the protein coloured by temperature factor (red-static, blue-mobile) for substrate bound (i) and substrate free enzyme (ii). Area of greatest change is boxed.

As can be seen in the above figure there is a region of the protein that has significantly increased thermal motion upon removal of the substrate. Three main areas are affected by substrate binding, the B' helix between residues 87 and 96, the loop joining helices F and G (185-192) and region around residue 251, figure 8.10.

The first of these is the region around the Y96 residue, the area of the protein that was proposed as the binding entrance. This view was supported when Poulos crystallized a enzyme inhibitor with a long aliphatic chain.<sup>30</sup>

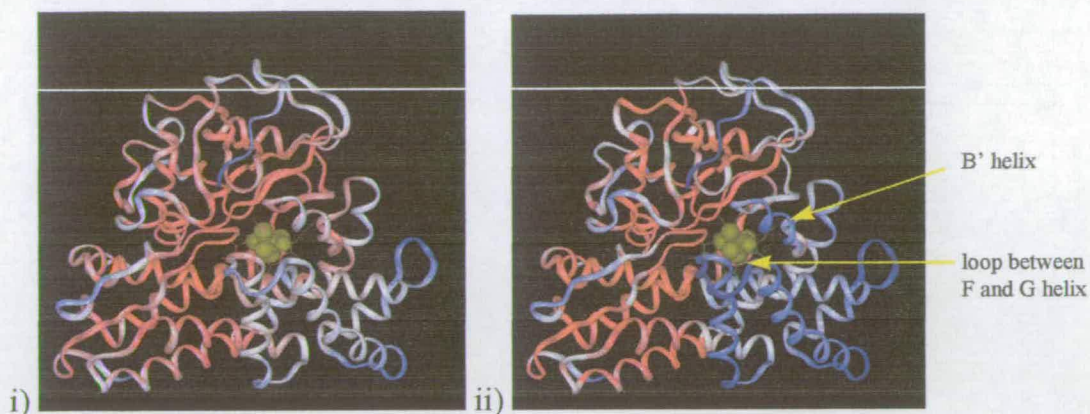


Figure 8.10: The ribbon structure of the protein coloured by temperature factor (red-static, blue-mobile) for substrate bound (i) and substrate free enzyme (ii).

This chain appeared to stick into the entrance channel interfering with residues Phe 193, Phe 87 and Tyr 96 and in this 'open' configuration: the Phe 193 ring points up towards the protein surface; the Phe 87 ring adopting the position formally occupied by the Phe 193 ring. When examining this binding channel it is apparent that entrance pocket is lined with hydrophobic residues, figure 8.11.

A substrate binding must pass the Phe 87, Phe 98 and Tyr 96 residues on its way to the centre of the binding site; this must apply to our Y96A mutant as well. Thus with the mutation to the Tyr 96 residue the substrate entrance should be larger, thus facilitating easier binding of substrate.

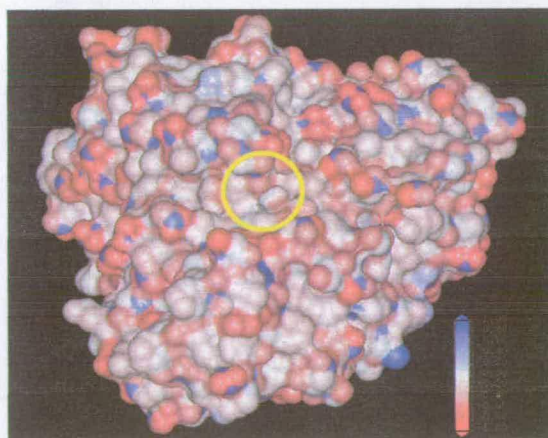


Figure 8.11: Surface of proposed entrance channel of P-450cam, coloured by charge (red-positive, blue-negative). Proposed entrance channel is indicated.

Unfortunately this also applies to the dissociation of the substrate which should also be greater as recent studies with photoacoustic calorimetry have indicated that substrate can be expelled rapidly from the binding site by the photolysis of a bound carbon monoxide.<sup>31</sup> This could explain the good binding constants for P-450cam Y96A mutants but why they give poor turnover. If the substrate enters and leaves the binding site quickly the catalytic cycle could be initiated but the substrate could leave the binding site before being hydroxylated. This would thus explain the increased peroxide production caused by the Y96A mutation.<sup>32</sup>

The other region to move, when substrate is removed, is the loop between the F and G helices. These helices form a 'lid' above the binding site with the residues of the loop between them lining the 'upper' region of the binding site (e.g. Phe 193), figure 8.12.

It is clear from this picture that the substrate bound structure has less thermal motion than the substrate free structure, this area of instability being centered at the loop between the helices. This region is more significant, not for the binding site residues it processes but for its role in the proton delivery pathway (see chapter 7).



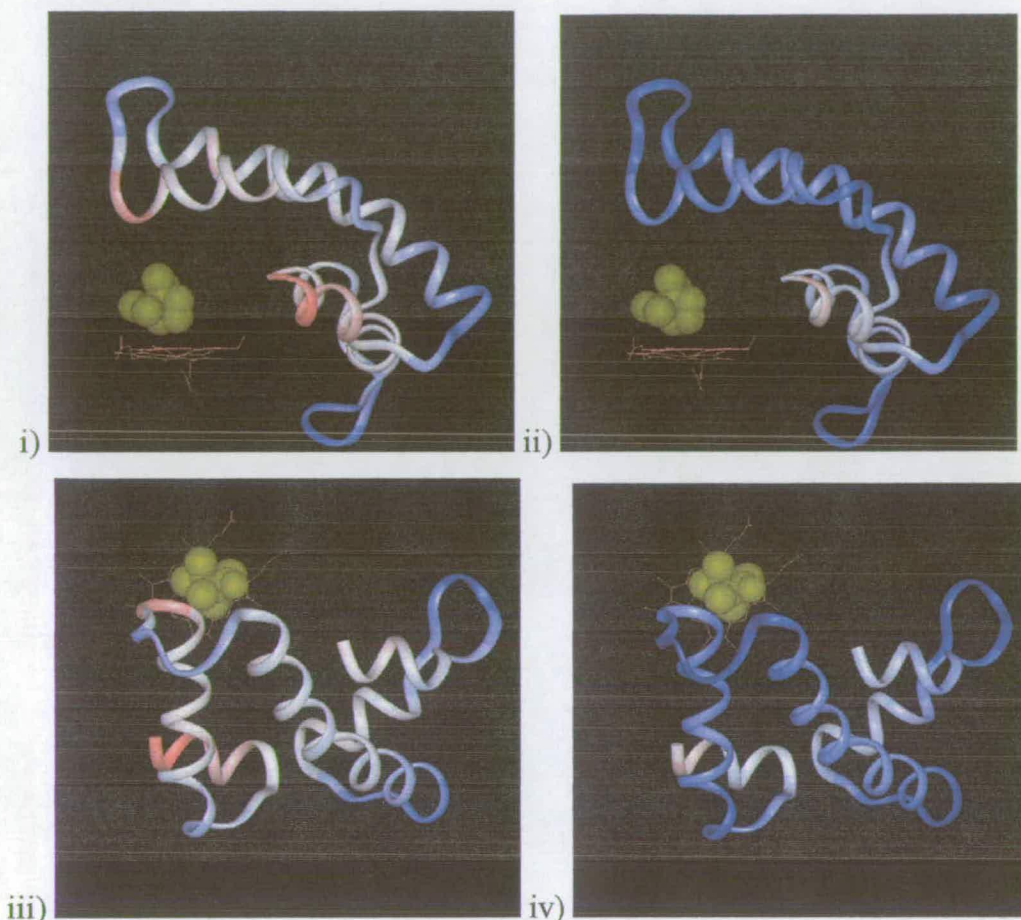


Figure 8.12: The F and G helices for substrate bound (i)/ (iii) and substrate free (ii)/ (iv) enzyme from two perspectives. Colour scheme as before. Camphor and heme group are shown for clarity.

A number of residues believed to be involved in this pathway lie in the region of this loop, (e.g. Met 184, Arg 186, Asp 182, Lys 178). Indeed the third region of movement is around Asp 251 which is the key catalytic residue in this model for the proton pathway. It would appear that these movements of the protein might be associated with salt bridges as it has been demonstrated that salt bridges must be broken upon substrate entry.<sup>33</sup> Though these salt bridges have not been identified, Poulos<sup>34</sup> and Sligar<sup>35</sup> separately have proposed that a bridge between Asp 251 and Lys 178 might be responsible. Site directed mutagenesis of some of these residues would also seem to indicate that they have a role in regulation of substrate control.

This would lend credence to the theory proposed earlier (chapter 7), that the substrate entry to the binding site has an effect on the residues of the proton pathway. The link between these two regions can be seen more clearly when they are shown on the same picture, figure 8.13.

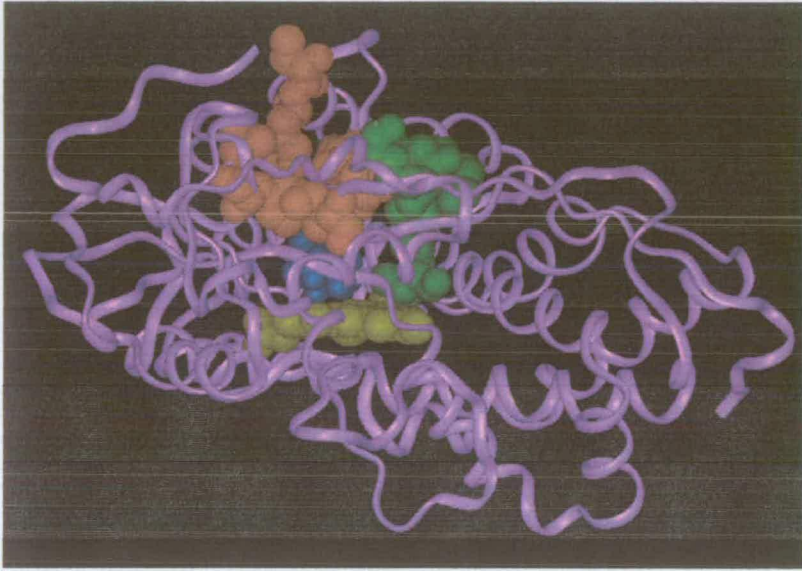


Figure 8.13: Showing the entrance channel (red) and the proton pathway (green) in relation to the substrate (blue) and heme (yellow).

Thus it is clear from this that both the 'upper' region of the binding pocket and the proton delivery pathway are linked by their close proximity. This common region between them is also the more mobile area upon substrate disassociation. The P-450cam system therefore has developed a method of associating the substrate entry with the proton pathway independent of the interactions with the heme. This relationship should be explored further, as mutagenesis to the binding site to expand the substrate range, may have an adverse effect on the proton pathway. Although this does not seem to have happened with the Y96A mutation.

**8.8 References.**

- 1)Chang, Y. T.; Loew, G. H. *Protein Engineering* **1996**, *9*, 755-766.
- 2)Chang, Y. T.; Stiffelman, O. B.; Vakser, I. A.; Loew, G. H.; Bridges, A.; Waskell, L. *Protein Engineering* **1997**, *10*, 119-129.
- 3)Chang, Y. T.; Stiffelman, O. B.; Loew, G. H. *Biochimie* **1996**, *78*, 771-779.
- 4)Modi, S.; Paine, M. J.; Sutcliffe, M. J.; Lian, L. Y.; Primrose, W. U.; Wolf, C. R.; Roberts, G. C. K. *Biochemistry* **1996**, *35*, 4540-4550.
- 5)Zvelebil, M.; Wolf, C. R.; Sternberg, M. J. E. *Protein Engineering* **1991**, *4*, 271-282.
- 6)DeVoss, J. J.; DeMontellano, P. R. O. *Journal of the American Chemical Society* **1995**, *117*, 4185-4186.
- 7)DeVoss, J. J.; DeMontellano, P. R. O. *Methods in Enzymology* **1996**, *272*, 336-347.
- 8)DeVoss, J. J.; Sibbesen, O.; Zhang, Z. P.; DeMontellano, P. R. O. *Journal of the American Chemical Society* **1997**, *119*, 5489-5498.
- 9)Belkina, N. V.; Dubanov, A. V.; Kolesanova, E. F.; Petushkova, N. A.; Skvortsov, V. S.; Ivanov, A. S. *FASEB Journal* **1997**, *11*, 82.
- 10)Belkina, N. V.; Skvortsov, V. S.; Ivanov, A. S.; Archakov, A. I. *Voprosy Meditsinskoi Khimii* **1998**, *44*, 464-473.
- 11)Li, H. Y.; Poulos, T. L. *Acta Crystallographica Section D-Biological Crystallography* **1995**, *51*, 21-32.
- 12)Helms, V.; Deprez, E.; Gill, E.; Barret, C.; Hoa, G. H. B.; Wade, R. C. *Biochemistry* **1996**, *35*, 1485-1499.
- 13)Davies, M. D.; Qin, L.; Beck, J. L.; Suslick, K. S.; Koga, H.; Horiuchi, T.; Sligar, S. G. *Journal of the American Chemical Society* **1990**, *112*, 7396-7398.
- 14)Davies, M. D.; Sligar, S. G. *Biochemistry* **1992**, *31*, 11383-11389.
- 15)Stayton, P. S.; Sligar, S. G. *Biochemistry* **1990**, *29*, 7381-7386.
- 16)Sligar, S. G.; Filipovic, D.; Stayton, P. S. *Methods in Enzymology* **1991**, *206*, 31-49.

- 17) Atkins, W. M.; Sligar, S. G. *Journal of Biological Chemistry* **1988**, *263*, 18842-18849.
- 18) Martinis, S. A.; Atkins, W. M.; Stayton, P. S.; Sligar, S. G. *Journal of the American Chemical Society* **1989**, *111*, 9252-9253.
- 19) Raag, R.; Martinis, S. A.; Sligar, S. G.; Poulos, T. L. *Biochemistry* **1991**, *30*, 11420-11429.
- 20) Kimata, Y.; Shimada, H.; Hirose, T.; Ishimura, Y. *Biochemical and Biophysical Research Communications* **1995**, *208*, 96-102.
- 21) Vidakovic, M.; Sligar, S. G.; Li, H. Y.; Poulos, T. L. *Biochemistry* **1998**, *37*, 9211-9219.
- 22) Manchester, J. I.; Ornstein, R. L. *Biochimie* **1996**, *78*, 714-722.
- 23) Fowler, S. M.; England, P. A.; Westlake, A. C. G.; Rouch, D. R.; Nickerson, D. P.; Blunt, C.; Braybrook, D.; West, S.; Wong, L. L.; Flitsch, S. L. *Journal of the Chemical Society-Chemical Communications* **1994**, 2761-2762.
- 24) Jones, N. E.; England, P. A.; Rouch, D. A.; Wong, L. L. *Journal of the Chemical Society-Chemical Communications* **1996**, 2413-2414.
- 25) Staines\_AG MSc Thesis *Monooxidations using a cytochrome P450cam mutant*; University of Edinburgh, 1996.
- 26) Aitken, S. J. MSc Thesis *Cytochrome P-450 Monooxygenases: Their Application in Biocatalysis*; University of Edinburgh, 1997.
- 27) Raag, R.; Swanson, B. A.; Poulos, T. L.; Demontellano, P. R. O. *Biochemistry* **1990**, *29*, 8119-8126.
- 28) Poulos, T. L.; Finzel, B. C.; Gunsalus, I. C.; Wagner, G. C.; Kraut, J. *Journal of Biological Chemistry* **1985**, *260*, 16122-16130.
- 29) Poulos, T. L.; Finzel, B. C.; Howard, A. J. *Journal of Molecular Biology* **1987**, *195*, 687-700.
- 30) Raag, R.; Li, H. Y.; Jones, B. C.; Poulos, T. L. *Biochemistry* **1993**, *32*, 4571-4578.
- 31) Diprimo, C.; Hoa, G. H. B.; Deprez, E.; Douzou, P.; Sligar, S. G. *Biochemistry* **1993**, *32*, 3671-3676.



- 32) Wong, L. L.; Westlake, C. G.; Nickerson, D. P. *Structure and Bonding* **1997**, *88*, 175-207.
- 33) Deprez, E.; Gerber, N. C.; DiPrimo, C.; Douzou, P.; Sligar, S. G.; Hoa, G. H. B. *Biochemistry* **1994**, *33*, 14464-14468.
- 34) Poulos, T. L.; Finzel, B. C.; Howard, A. J. *Biochemistry* **1986**, *25*, 5314-5322.
- 35) DiPrimo, C.; Deprez, E.; Sligar, S. G.; Hoa, G. H. B. *Biochemistry* **1997**, *36*, 112-118.

## 9 Modelling with the Package DOCK.

---

### 9.1 Introduction.

DOCK is a suite of programs designed to find favourable orientations of a ligand in a receptor. First devised in 1982 by Irwin Kuntz,<sup>1</sup> at the University of California and later updated and revised.<sup>2-14</sup> DOCK has a unique method of modelling the receptor/ligand interactions, without extensive computational requirements. The DOCK suite consists of three basic programs, sphgen, grid and DOCK, figure 9.1.

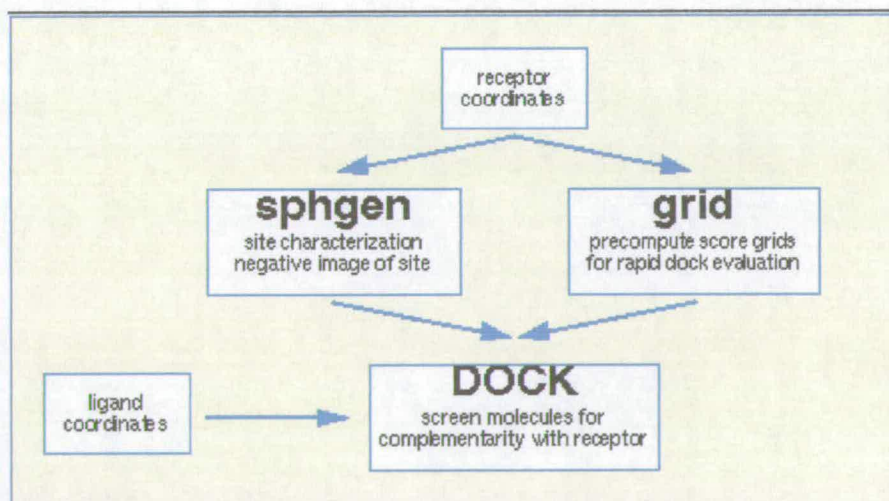


Figure 9.1: Overview of the DOCK suite.

To complement these three programs other supplemental programs are available which enhance the DOCK suite or allow conversion of output files to more accessible formats. To reduce the computational time necessary to process the myriad of receptor/ligand interactions DOCK has two unique components.

- 1) The first is the use of sphere centers to characterize the binding site. These centers are created with sphgen and consist of a series of partially overlapping spheres, which fit completely inside the binding site of a ligand, touching the surface but without intersecting the Van der Waals radii of the binding site atoms. To find a starting orientation for the ligand a number, usually three or four, of its atoms are overlapped with the sphere centers creating sphere-atom pairs. This gives the computer a starting orientation for the ligand. Alternate methodologies would use every possible orientation of the ligand, which can utilize considerable computer resources. DOCK then minimizes the ligand in this orientation and filters the ligand position for no 'bad contacts' (bumps) with any receptor atoms. These bumps are user defined minimum distances between the ligand and receptor atoms. An important point to note is the limitation put upon DOCK by this approach, since it does not allow for any movement of the ligand binding site. A good quote is 'what is good about DOCK is that it uses spheres; what is bad about DOCK is that it uses spheres...'.<sup>2</sup> It is not sufficient to rely on the spheres generated by sphgen to characterize the binding site. It is essential to view and manipulate the sphere centers to better characterize the binding site. This approach requires some level of intuition on the part of the user therefore mistakes can easily be made without careful analysis of the location of the sphere centers.
- 2) The second aspect to the DOCK program suite is the scoring mechanism. Once a suitable ligand orientation has been found it must be characterized relative to other suitable orientations. The method by which DOCK does this is to use a predefined grid of points to describe the binding site, created with 'grid'. The suitable ligand orientations are then evaluated (scored) by their interaction with points on the grid.



grid created can be based on the distance between the receptor atoms, (contact scoring), or a more complicated grid based on the electrostatic parameters of the atoms (energy scoring). Contact scoring is based upon a predetermined value for the atom-atom minimum allowed distance, usually about 3 Å. The energy scored function is based on the Lennard-Jones function, figure 9.2.

$$E = \sum_{i=1}^{ng} \sum_{j=1}^{nc} \left( \frac{A_{ij}}{r_{ij}^a} - \frac{B_{ij}}{r_{ij}^b} + 332 \frac{q_i q_j}{D r_{ij}} \right)$$

Figure 9.2 Lennard-Jones function, used to calculate the energy scored grid. Where: E=intermolecular interaction energy;  $r_{ij}$ =distance between atoms i and j;  $A_{ij}$  and  $B_{ij}$ = van der Waals repulsion and attraction parameters; a and b=van der Waals repulsion and attraction exponents;  $q_i$  and  $q_j$ =point charges on atoms i and j; D=dielectric function; 332=factor to convert electrostatic energy to kcal/mol.<sup>2</sup>

The score of the ligand orientation is stored and a new orientation is considered, figure 9.3.

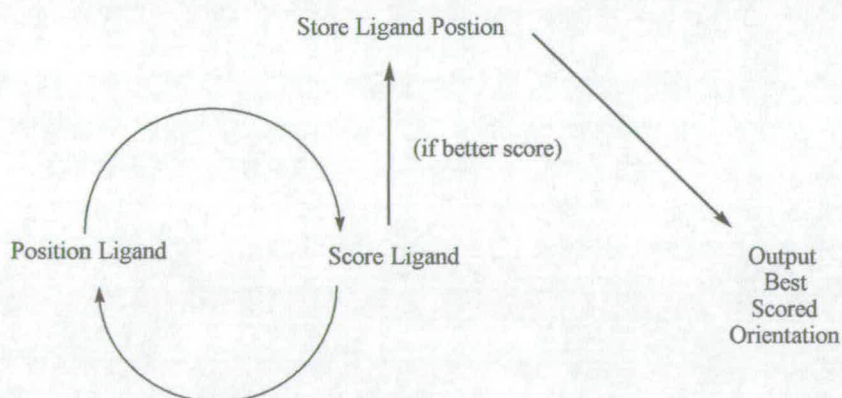


Figure 9.3: General schematic of scoring mechanism.

Once this process is finished DOCK outputs the coordinates of the ligand orientation with the best score, its score and details of the number of orientations scored.

## 9.2 Previous studies with P-450cam and DOCK

A number of previous studies of substrates with DOCK have been performed on P-450cam by deMontellano and coworkers.<sup>15-18</sup> The emphasis here was utilizing a search facility in DOCK to trawl through a database for possible ligands. This was attempted with both wild-type P-450cam and latterly with the L244A mutant.<sup>15,16</sup> In each case deMontellano and coworkers selected a dozen of the best scored potential substrates and then physically tested them on the enzyme for binding and turnover. A number of compounds not chosen by DOCK to be substrates were additionally tested to validate the methodology. This technique proved to have some limited success however one major drawback occurred. It was apparent that the ability of DOCK to successfully predict real substrates was dependent on parameters for the polar and non-polar cut-off distances. If these distances were set too high, actual substrates would not be predicted as such and if it was set too low some non-substrates would be selected as candidates. Later, it was found that these distances also changed with the use of the L244A mutant. It is also important to note that camphor is not predicted to be a substrate with this method as its close contacts to the protein binding site give it a poor score.

## 9.3 Differences in Our Studies with DOCK.

One difference to our approach is that here we do not use a database of randomly chosen potential substrates. Instead compounds are speculatively selected by eye from similarities to known substrates. This approach more rigorously examines individual compounds, however, it does not provide a means of exhaustively identifying all the P-450cam Y96A potential substrates.

The key difference in our study is the due to the version of DOCK used. Previous studies with DOCK have used versions up to version 3.5 which all have similar drawbacks, namely the strong dependence on the non-polar and polar cut-off distances. In version 4.0 the contact scoring system was made more uniform with a contact distance and an allowed overlap (bump overlap) rather than different values

for non-polar or polar atoms. This reduction in control was due to the significant improvement in ligand orientation parameters. Up to DOCK version 3.5, ligands were placed in the binding site in a rigid form, which meant reduction of the allowed contact distances was necessary to allow for potential reorientation of the substrate within the binding site environment. Additionally, this previous method could not compensate for multiple low energy conformations of ligands such as the two chair and boat forms of a cyclohexane ring. The significant improvement to DOCK is the ligand orientation and conformation parameters. With the 'orientation' parameters one can now search all the possible orientations of the compound in the binding site. Previously the ligand orientation was solely dependent on the location of the sphere centers and compounds without 4 atoms overlapping the sphere centers were missed out. Now the sphere centers are used as a guide, with only one sphere center necessary to orientate the molecule and greater deviation from the sphere centers allowed. This does however significantly increase the computational time used by DOCK.

The other new DOCK algorithm is the 'conformational' search. This allows the all rotatable bonds to be adjusted *in situ*, in the binding site, until the best scored orientation is found (e.g. figure 9.4).

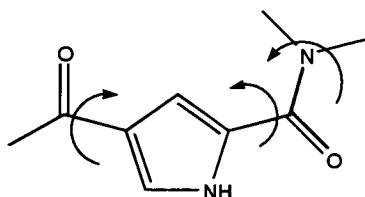


Figure 9.4: Example of the parts of a molecule altered by the 'conformational' algorithm.

This alters the limited method that DOCK searched for potential ligands. Now DOCK performs a sophisticated cycle of adjustments to find the best ligand orientation, figure 9.5.



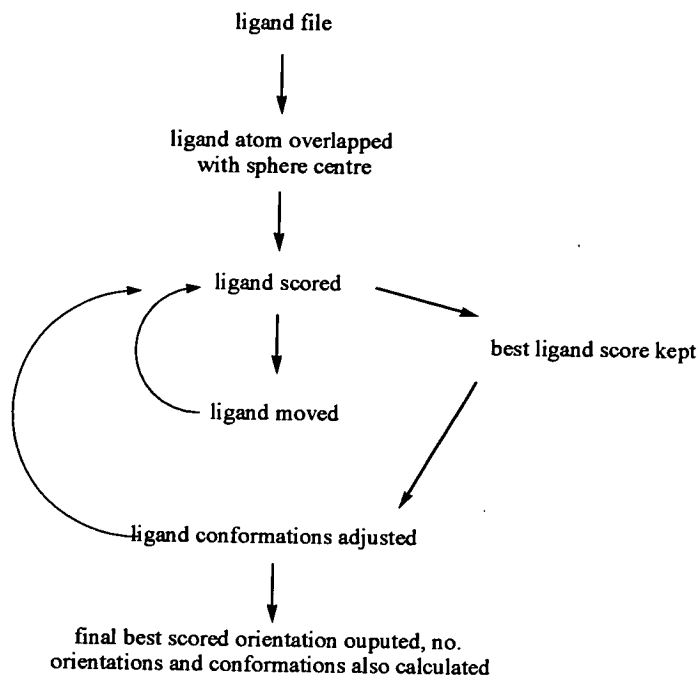


Figure 9.5: Overall schematic of the ligand binding process.

Thus the new version of DOCK reduces the previous two main drawbacks to its use, namely the strong reliance on sphere centres to orientate the ligand and the use of rigid ligands. These improvements are made at the expense of computational resources.

### 9.3.1 Best Binding Site Model.

The two basic binding site models that were explored in our studies are the substrate free binding site of the cytochrome P-450cam Y96A mutant both with and without a bound single oxygen species. Since camphor is not predicted as a substrate with the previous DOCK methodologies it was decided to use camphor as a robust test of the new DOCK suite.

When using camphor, the heme bound oxygen artificially pushed the camphor away from the heme towards Val 244 and rotates it so that the camphor carbonyl points towards the heme, figure 9.6. This is clearly a camphor orientation that is inconsistent with the crystal structure and binding data.

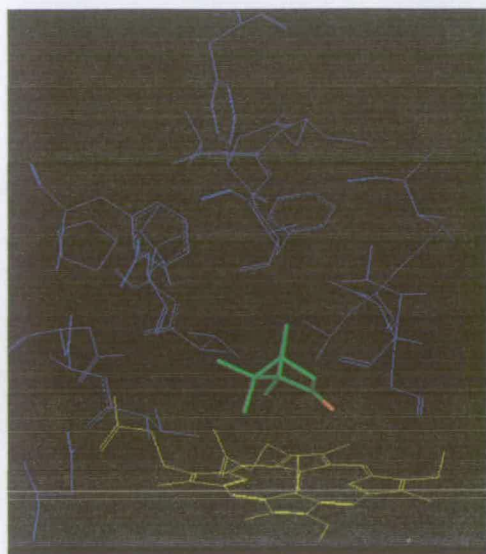


Figure 9.6: Orientation of camphor outputted by DOCK using a heme bound oxygen in the binding site.

However, when the oxygen species is removed and new scoring grids were calculated it was found that camphor adopted a realistic orientation, albeit slightly different from the orientation in our Y96A mutant structure, figure 9.7.

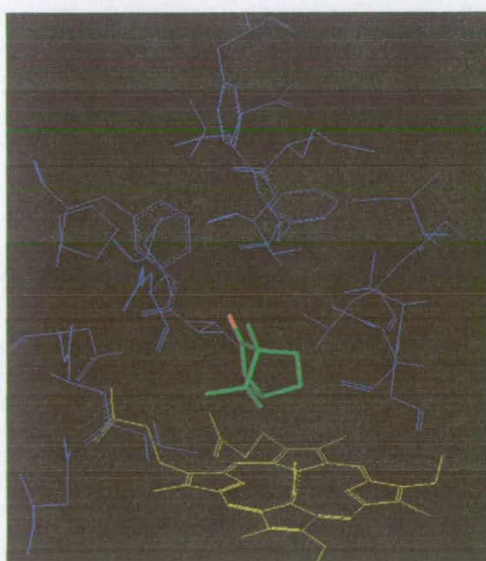


Figure 9.7: Orientation of camphor outputted by DOCK using a no heme bound oxygen in the binding site.

The use of an oxygen species in the binding site would appear to be more appropriate as it should more closely resemble the binding site prior to hydroxylation (see chapter 1). The non-oxygen system, although a better model for the initial substrate binding of the catalytic cycle, is not a realistic representation of the later stages and is thus not a good model for the hydroxylation step. This would suggest that the system without oxygen would be a good model for substrate binding prediction but that the active site with bound oxygen should be used in calculation of the potential hydroxylation sites. This however proves not to be entirely accurate as demonstrated above by the camphor system. The reason for this is probably not that the oxygen bound model is not a good representation of the hydroxylating species, but due to a number of inaccuracies in this model. Firstly it is not known if the binding site undergoes minor rearrangement during the catalytic cycle which could reorientate the substrate back towards the repulsive Van der Waals interactions with the oxygen. Secondly, and more importantly, the oxygen species modelled by DOCK is based on Van der Waals radii of 'normal' oxygen atoms. The species here is of unknown charge and it is likely that some of the electron density assigned to the oxygen by DOCK is distributed towards the iron and possibly the heme ring. This would thus make the radii of the oxygen atom less than that used by DOCK and therefore less repulsive to the substrate. Additionally to this DOCK cannot account for the interaction between a radical (if the oxygen species is indeed a radical) and the substrate. For these reasons it was decided that the oxygen free system is a better model for both the substrate binding and for calculating the potential hydroxylation site. For purposes of calculation of oxygen-substrate distances, a bound oxygen was added to the post calculation model but it was not present during the calculations performed in the DOCK run.

### 9.3.2 Selection of the Best Forcefield.

Although it increases the computational time used by DOCK both energy and contact scoring were used. To examine the reliability of the scoring mechanism it was

decided to perform multiple DOCK runs with camphor. It was found that in all cases the lowest energy scored species was always in the same orientation irrespective of starting orientation. However when using the contact scoring system, different final orientations were produced (of equivalent score). All the orientations, upon examination, looked like plausible substrate binding orientation, however, they all significantly differed, both from each other and the low energy scored orientation.

The reason behind the discrepancy between the contact and energy scoring is the reliance on fixed distance parameters for calculation of the atom-atom interactions. This can be seen in a plot of the score vs. the interatomic distance, figure 9.8.

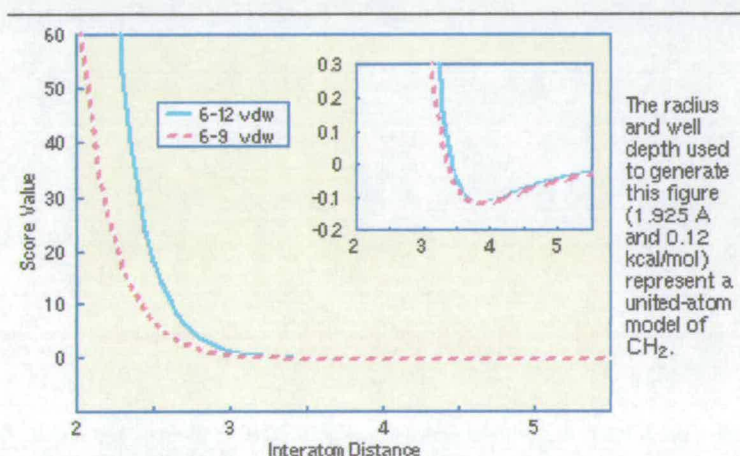


Figure 9.8: Picture of the dependence on the distance for the ligand score.

This effect was also examined for the other substrates and it was found that many of the contact scored orientations had binding orientations incompatible with known crystal data. However in some cases the contact scores had more feasible orientations therefore it was decided to score and model both methods.

### 9.3.3 Substrates Analysis.

In order to understand the docking orientations of the molecules studied it is necessary to describe the important binding regions of the protein relative to some



fixed point of origin. For the purposes of clarity the binding site is always referred to similarly to the orientations used previously for the description of the crystal structure, figure 9.9.

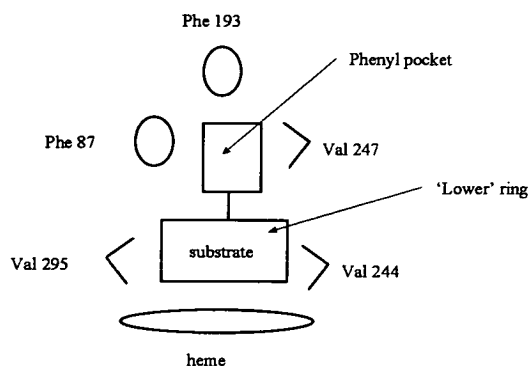


Figure 9.9: Schematic of the P-450cam Y96A binding site showing important residues mentioned in the following chapter. Picture shows the method of defining the different substrate rings with the term 'upper' and 'lower'.

A number of common substrate orientations exist in the binding site, identified by the DOCK simulations. The first is the hydrophobic binding pocket between Phe 87 and Val 247, called 'the phenyl pocket' the other binding regions are a pocket between Val 295 and Asp 297 referred to as the 295/297 pocket and a region in front of Val 244 referred to as the 244 pocket.

### 9.3.4 Understanding DOCK Output Files.

In addition to providing the best ligand orientation for the contact scored and energy scored molecule, DOCK outputs a number of pieces of information which need to be explained if the result of the DOCK run is to be understood. The 'total conformations' reflects the number of unique substrate conformations that the DOCK routine could determine. Each one of these conformations was placed in the binding site to give between one and two million orientations per conformation ('orientations tried'). Of these orientation only a small number met the scoring criteria ('orientations scored'). The best scores are displayed for the contact and energy scores and the

difference between the orientation of the lowest energy scored molecule, for each scoring method, is displayed as an rmsd value.

The orientation data gives an insight into the possible motion of the substrate in the binding site, as a low number of scored orientations would indicate low mobility. The scores for the substrates are only comparative but can be used to access the enzyme substrate recognition. The rmsd value for the two scoring methods assesses their complementarity.

It should be noted that, with the exception of camphor, all other molecules were examined and are discussed without reference to their binding data.

## 9.4 Results of DOCK Simulations.

### 9.4.1 Camphor and Adamantane Analogues.

#### 9.4.1.1 Camphor.

Camphor is a very inflexible molecule, indicated by only one conformation. The number of orientations scored for camphor was lower than for any of the other of the camphor/adamantane analogues, table 9.1, indicating the strong enzyme substrate complementarity towards camphor. The contact score for camphor was higher here than for most other compounds, reflecting its smaller size, the energy score reflecting very favourable camphor interactions.

Table 9.1: DOCK data for camphor (2).

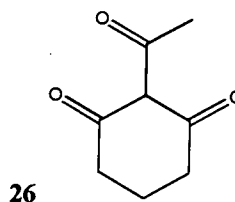
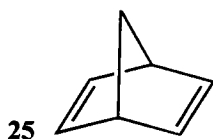
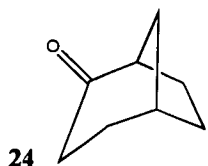
compound	total conformations	orientations tried	orientations scored	contact score	energy score	RMSD
2	1	789401	10000	-186.00	-22.18	1.25

#### 9.4.1.1.1 DOCK Orientations of Camphor.

For the actual proposed position, the energy scored molecule occupies an orientation similar to that observed by the crystal structure except that the geminal methyl groups are slightly 'lower' i.e. nearer the heme. This places their carbonyl pointing towards Ala 96, however the C5 and C6 positions are pulled slightly towards Leu 244, the geminal methyls pointing towards Val 295. The 5-*exo* or the 4 position are the likely hydroxylation sites. For the contact scored molecule the carbonyl points towards Thr 101 forming a potential H-bond with the side chain threonine oxygen. The rest of the molecule is slightly away from the heme centre and towards the 295/297 pocket. In this case the C6 position could be the most likely to be hydroxylated.

From this data it can be seen that DOCK successfully predicts camphor as a substrate for the enzyme. As stated previously the energy scored orientation is closer to the crystal data than the contact scored one, although the contact scored orientation does predict the second major product of the camphor hydroxylation (i.e. 6-*exo*-hydroxycamphor).

#### 9.4.1.2 Camphor Analogues.



Compounds, 24, 25 and 26 are all smaller than camphor reflected in their low contact scores, table 9.2.

Table 9.2: DOCK data for compounds 24-26

compound	total conformations	orientations tried	orientations scored	contact score	energy score	RMSD
<b>24</b>	1	309673	20000	-179.00	-18.58	2.98
<b>25</b>	1	99748	20000	-144.00	-12.11	3.63
<b>26</b>	6	11214652	108141	-232.00	-24.03	1.23

The energy scores are comparable for camphor, except for **25**, which has a very poor score, probably due to its smaller size and lack of functionality. Compound **26** is obviously more flexible than the others, since it is not a bicyclic system, and its large number of scored orientations could indicate that without a hydrogen bond it might exhibit poor stereoselectivity in its hydroxylation.

#### 9.4.1.2.1 DOCK Orientations of 24, 25 and 26.

Compound **24** occupies a central position just above the heme. The energy scored molecule is positioned towards Val 297 and the contact scored molecule is towards Val 244, a theme seen in many more of the molecules studied. For the energy scored molecule C2, C3 and C4 are all equally near the heme but not in a particularly good orientation for hydroxylation. The energy scored molecule's carbonyl would seem to be able to hydrogen bond with Asp 297 however the contact scored molecule's carbonyl hydrogen bonds to Thr 252. All the hydrogens on the *endo* side of the ring are potential hydroxylation sites for both energy scored and contact scored models.

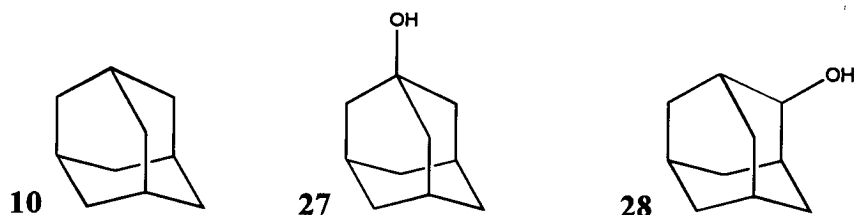
Compound **25** is essentially the same as for **24**, with the energy scored molecule towards the Val 295 and the contact scored molecule towards Val 244. For the contact scored molecule the orientation of the bicyclic ring is such that the bridgehead carbon is further most from the heme leaving all main ring carbons equally

positioned near the iron centre and thus none are in a sufficiently good hydroxylating position.

For **26**, both the scored options had the same basic orientation however with the alternate position of the side-chain carbonyl. The ring is towards the lower pocket between Val 295 and Asp 297. The molecule lies across the heme with the side chain near Leu 244. The nearest hydrogens are the single hydrogen on C1 or the hydrogen of the side chain methyl group. There are no discernable hydrogen bonds from the carbonyl groups to the enzyme.

Thus from this data it can be concluded that, whilst all three compounds are probably substrates, the lack of complementarity of **26** and smaller size of **25** would make **24** the best of these compounds.

#### 9.4.1.3 Adamantane and Adamantanols.



Although adamantane [**10**] is a rigid molecule, the lack of any functionality gives it multiple binding orientations, table 9.3.

Table 9.3 DOCK scores for compounds **10**, **27** **28**.

	total conformations	orientations tried	orientations scored	contact score	energy score	RMSD
<b>10</b>	1	1772878	100000	-173.00	-17.39	3.16
<b>27</b>	3	9188320	300000	-189.00	-22.18	3.30
<b>28</b>	3	9281708	300000	-190.00	-22.09	3.27

This lack of functionality gives adamantane a poor contact score but a relatively worse energy score. Both adamantanols [27,28], have marginally improved scores compared to adamantane and both molecule share essentially the same DOCK profiles.

#### 9.4.1.3.1 *DOCK Orientations of 10, 27 and 28.*

For compound **10** the scored molecules are in slightly different orientations with the energy scored molecule slightly towards the 295/297 pocket and the contact scored molecule in the other direction towards Val 244. The three axial hydrogens on one 'hexane' ring are the most accessible to the heme.

For **27** the orientations of the contact scored and energy scored molecules are virtually the same. The adamantane ring sits above the heme with minimal side chain interactions. The hydroxyl group is towards Glu 248 with the C2, C3, and C4 axial hydrogens all readily accessible to the heme for hydroxylation

Compound **28** is in essentially the same orientation as **10**, with the energy scored molecule slightly towards the 295/297 pocket and the contact scored molecule in the other direction towards Val 244. In both instances the hydroxyl group is pointing towards the heme ring at near Val 295 with the C1-hydrogen and the axial C2 and C3 hydrogens are readily accessible by the heme.

#### 9.4.2 **General Orientations of Phenyl Substituted Substrates.**

Most of the following substrates have a aromatic ring, designed to fit into the pocket between Phe 87, Phe 98 and Val 247. For most of these substrates they do indeed have this orientation, figure 9.10.



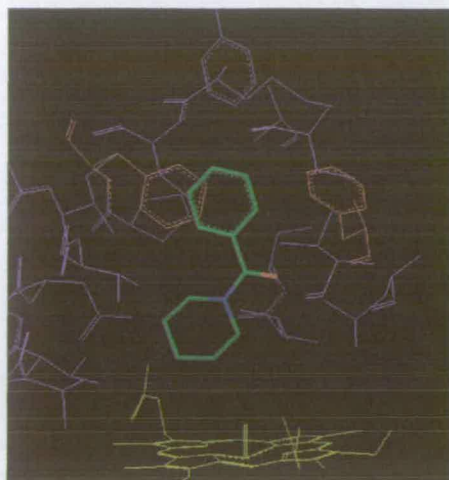
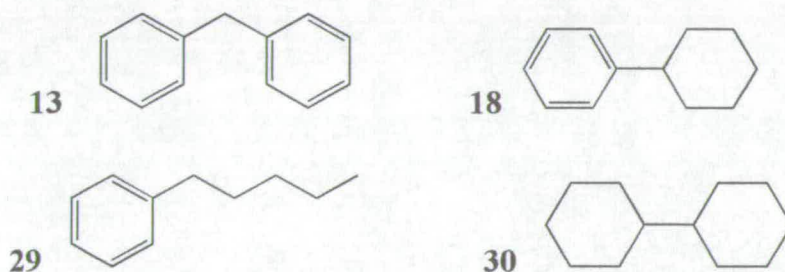


Figure 9.10 Compound 44, showing typical orientation of phenyl substituted substrates.

This is a typical substrate orientation that can be used to visualize most of the orientations described below, with the aromatic ring in the 'upper' phenyl pocket and in this case the non-aromatic ring in the 295/297 pocket.

#### 9.4.3 Diphenylmethane Analogues.



Diphenylmethane [13] binds with a relatively good score for both the contact and energy scored molecules, table 9.4. The scores for phenylcyclohexane [18] are essentially the same however the rmsd value would suggest the energy and contact scored molecules are in different orientations. Amyl benzene [29], again, has essentially the same scores as 13 and 18 although the higher number of conformations implies that the stereoselectivity could be poor. Bicyclohexyl [30] has a worse contact score than the others however the energy score is essentially the same.

Table 9.4: DOCK scores of 13, 18, 29, 30.

compound	total conformations	orientations tried	orientations scored	contact score	energy score	RMSD
13	20	20000000	13231	-230.00	-19.41	2.87
18	6	6000000	11224	-222.00	-19.78	5.21
29	40	40000000	543167	-233.00	-19.50	2.99
30	3	3000000	2855	-196.00	-20.21	5.10

#### 9.4.3.1 DOCK Orientations of 13, 18, 29 and 30.

For diphenylmethane [13], for both the energy scored and contact scored molecules, one ring occupies the upper phenyl pocket. For the energy scored molecule the lower ring points down towards Thr 101, this leaves the C3 hydrogen the only one in any reasonable orientation towards the iron centre. The contact scored molecule is in a similar orientation to the crystal structure, with the lower ring pointing towards the heme so that the C3 and C4 positions could be hydroxylated. This is the first instance of the contact scored molecule having a more realistic orientation than the energy scored molecule.

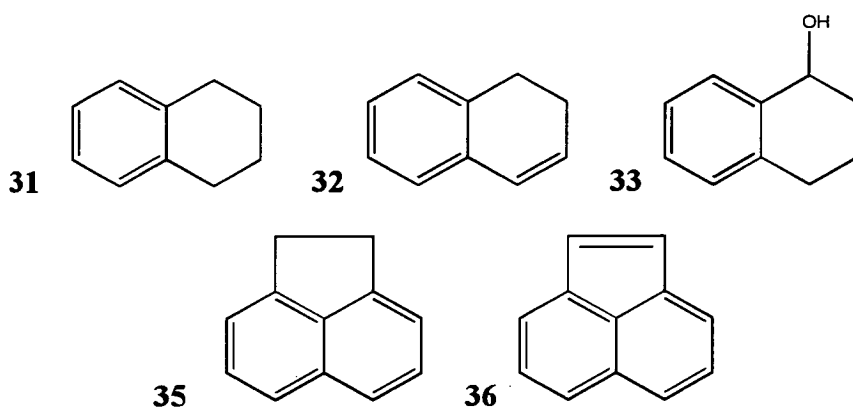
In the case of phenylcyclohexane [18], the energy scored molecule has the phenyl ring in the normal upper phenyl pocket and the aliphatic ring pointing down towards the heme, slightly towards Asp 297. This is an almost identical orientation to the crystal structure. However the contact scored orientation bizarrely has the two rings in the opposite positions with the aliphatic ring in the phenyl pocket. The reason for this orientation is unknown and this is the first time DOCK has preferentially placed an aliphatic ring in the phenyl pocket over a aromatic one.

The orientations of amylbenzene [29], for both the energy scored and contact scored molecules are very different here. The energy scored molecule has the phenyl ring in the phenyl pocket with the aliphatic chain pointing down and towards Leu 244 and Gly 248. The nearest hydrogen is on the C10 carbon, i.e. second from the end of

the aliphatic chain. For the contact scored molecule the phenyl ring occupies the same position as the aliphatic chain of the energy scored molecule orientation. The aliphatic chain then coils round above the heme ring. Again the C10 carbon is the nearest to the iron.

For bicyclohexyl [30], one of the cyclohexane rings adopts the same configuration as a typical ring in the phenyl pocket, for both the contact scored and energy scored molecules. However the other ring is in a significantly different orientation in each case. The energy scored molecule lower ring sits towards Val 295 with the axial-C4 carbon nearest the heme centre. The contact scored molecule lower ring however adopts an orientation on the other side of the pocket towards Leu 244. In this case both the axial-C4 and the equatorial-C3 are in a hydroxylating position.

#### 9.4.4 Naphthalene Analogues.



Both the tetrahydro [31] and dihydro [32] naphthalene have similar scores suggesting the extra hydrogens do not significantly alter the molecules, in this case, table 9.5. However, 33 has a better score for both the energy and contact scores suggesting the alcohol moiety actually enhances binding. 35 and 36 both have excellent contact scores suggesting a tight fit to the enzyme.

Table 9.5: DOCK scores for **31-36**.

compound	total conformations	orientations tried	orientations scored	contact score	energy score	RMSD
<b>31</b>	1	1477989	20000	-211.00	-18.54	4.18
<b>32</b>	1	1256840	20000	-210.00	-18.19	0.91
<b>33</b>	3	6000000	37784	-227.00	-20.90	0.64
<b>35</b>	1	2000000	8219	-240.00	-20.20	3.04
<b>36</b>	1	2000000	9324	-246.00	-20.00	3.11

#### 9.4.4.1 DOCK orientations of **31**, **32**, **33**, **35** and **36**.

Both energy scored and contact scored molecules are the same for **31**. The orientation is not similar to any of the other previous substrates, i.e. instead of the aromatic ring orientating itself into the upper phenyl pocket, the substrate lies flat, in the same plane as the heme ring, figure 9.11.

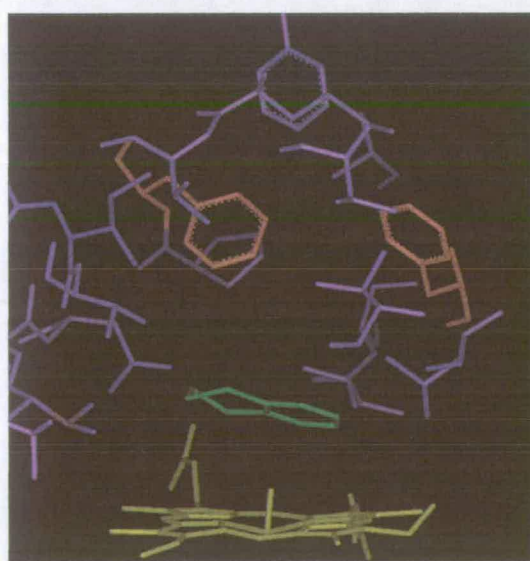


Figure 9.11: Orientation of **31** (green) in the binding site, with the heme group (yellow), Phe 87 and Phe 98 indicated (red).

The aliphatic ring is slightly towards Val 295. There are no obvious hydrogens for hydroxylation, although many of the aliphatic hydrogens could be in position with a small movement of the molecule away from Val 295.

The orientation of **32** is similar to the **31** orientation, being planar and above the heme. The phenyl ring is towards Leu 244 in both cases. For the energy scored orientation the aliphatic ring is slightly towards Val 295, similar to the orientation of **31**, for the contact scored orientation, towards Thr 101. The orientation of the acyl bond is in both cases away from the binding site side chains. No hydrogens are in any obvious hydroxylation position, as for **31**.

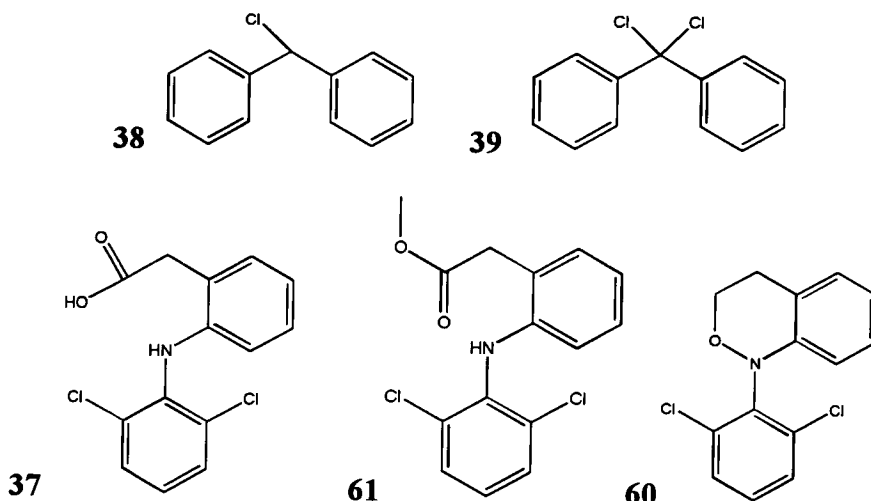
Both the contact scored and energy scored molecules for **33** are the same as for the **31/32** with the phenyl ring towards Val 244 and the aliphatic towards Val 295. In this case the hydroxyl group points towards the heme on the Val 295 side of the binding site.

Both energy scored and contact scored molecules are the same for **35**, again the rings are in the same plane as the heme with the aliphatic ring towards Thr 252. This means both the aliphatic protons on one side are the only ones in hydroxylation range but the orientation is not particularly suitable for hydroxylation.

The orientation of **36** is almost the same as for the unsaturated version, **35**, except the aliphatic ring is further from the heme, more towards Val 396. In this case it is even more unlikely that hydroxylation would occur.

The DOCK predictions for this range of substrates is that although they bind they do not fit into the aromatic pocket. This is consistent with the data that shows that these compounds are substrates for wild-type protein.

## 9.4.5 Chlorinated Aromatics.



As discussed previously **38** and **39** are potentially of interest for crystallization studies although the tolerance of the enzyme for halogens is of interest. Both compounds fit inside the binding site, although with an extremely tight fit, shown from the low contact scores and the minimal number of orientations that bind, table 9.6.

Table 9.6: DOCK orientations of **37-38**, **60**, **61**.

compound	Total conformations	orientations tried	orientations scored	contact score	energy score	RMSD
<b>38</b>	20	40000000	3444	-242.00	-20.17	5.21
<b>39</b>	20	40000000	532	-249.00	-19.82	0.57
<b>37</b>	50	50000000	0	*	*	*
<b>61</b>	50	50000000	0	*	*	*
<b>60</b>	1	1000000	0	*	*	*



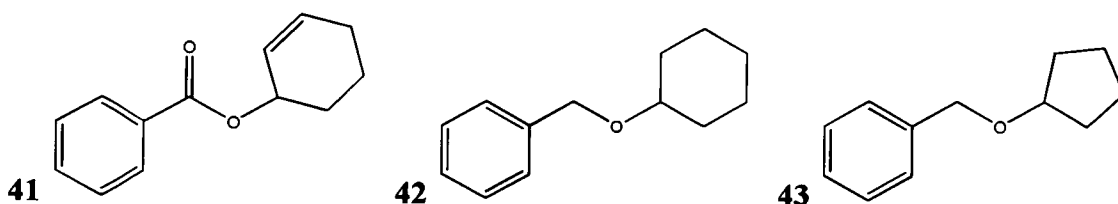
9.4.5.1 DOCK orientations of **38**, **39**, **37**, **60** and **61**.

Both energy scored and contact scored orientations are the same for **38**. One phenyl ring sits in the upper phenyl pocket, in the standard orientation. With the constraints placed on the molecule by the chlorine (needing a large space), the lower phenyl ring occupies the space in front of Leu 244, with a very close interaction between the ring and Leu 244 side chains. In both cases the C3 and C4 hydrogens are nearest but not in a particularly good orientation with the heme.

Again for **39** both scored orientations are the same and are similar to **38** with a slight difference in angle of the lower ring, relative to the first, to allow for the extra chlorine.

Molecules **37**, **60** and **61** are all predicted by DOCK not to be substrates for the P-450cam Y96A mutant. These compounds are too large for the binding pocket and this is the first instance, in our study, of DOCK failing compounds as potential substrates.

## 9.4.6 Aliphatic Substituted Aromatics.



DOCK suggests that these three molecules, **41**, **42**, and **43**, are substrates for the Y96A mutant enzyme, table 9.7. The contact scores all have some of the best scores seen in any of the compounds studied, the energy scores being comparable. Despite a large number of potential orientations the number that could actually bind with the enzyme is comparatively low, indicating a good fit between the enzyme and these substrates

Table 9.7: DOCK scores for 41-43.

compound	total conformations	orientations tried	orientations scored	contact score	energy score	RMSD
<b>41</b>	10	20000000	3175	-260.00	-21.38	2.43
<b>42</b>	30	60000000	3098	-261.00	-19.08	4.96
<b>43</b>	30	60000000	20863	-267.00	-18.92	4.99

#### 9.4.6.1 DOCK orientations of 41, 42 and 43.

For **41**, the contact scored and energy scored molecules are almost mirror images of each other. Both have the phenyl ring in the upper pocket, in the standard orientation. The energy scored molecule has the lower ring pointing towards Thr 101 and the carbonyl group towards Val 396. The contact scored molecule has the aliphatic ring pointing towards Val 295 with the carbonyl towards Leu 244. Neither carbonyl would seem to have a preferred H-bond to the enzyme. The nearest hydrogen on the contact scored molecule is the axial-C4 however on the energy scored the equatorial-C5 (C3) hydrogen is nearest to the heme.

In the case of **42** the energy scored molecule sits with the phenyl ring in the upper phenyl pocket which orientates the cyclohexane ring towards the iron centre, in a boat conformation. Both the C3 hydrogens are potential hydroxylation sites for the iron. The contact scored molecule has a different orientation with the cyclohexane ring towards the 295/297 pocket. This places all the C3 and C4 hydrogens within hydroxylation range of the iron.

The contact scored orientation for **43** looks very strange at first glance with the molecule 'doubled up' in an implausible orientation so that the phenyl ring is next to Leu 244 and the cyclopentane ring in the 295/297 pocket, figure 9.12.

The energy scored orientation however has the phenyl ring in the upper phenyl pocket with the cyclopentane ring slightly towards the 295/297 pocket. Both the C2 hydrogens are potential hydroxylation sites.

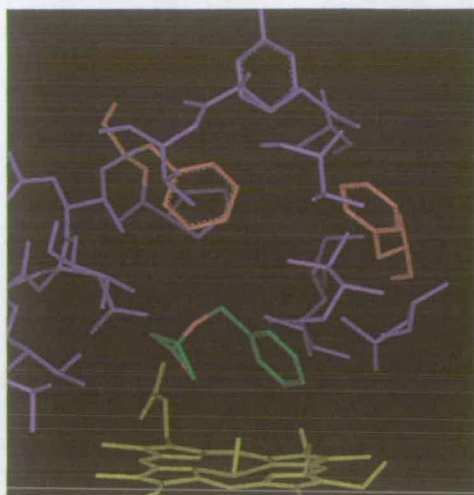
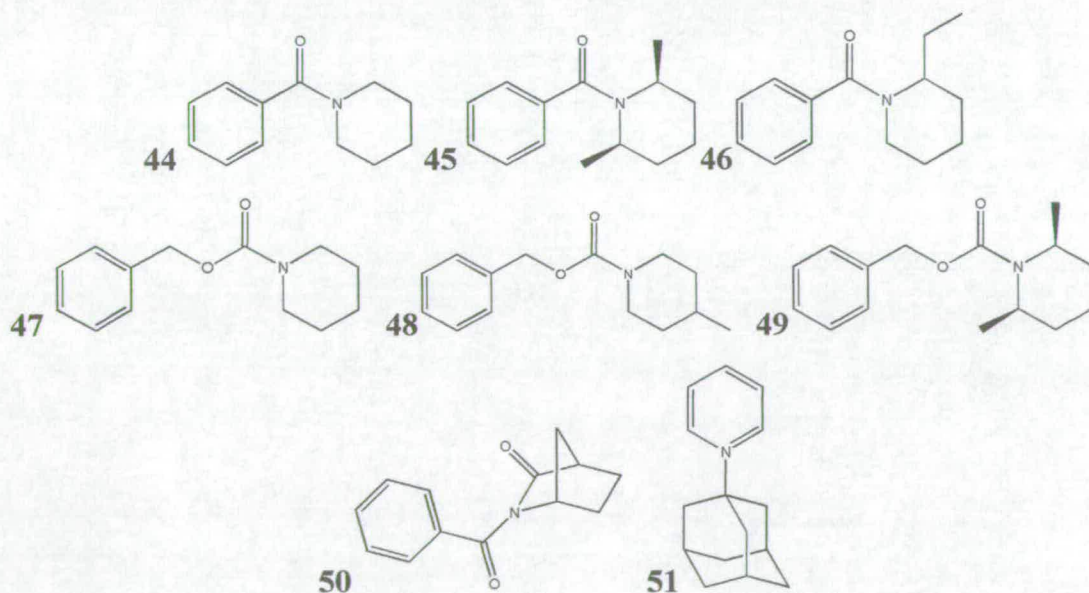


Figure 9.12: The contact scored orientation of **43** (green) in the binding site, with the heme group (yellow) Phe 87 and Phe 98 (red) indicated

This is another example of the contact scored molecule scoring producing an unusual binding orientation.

#### 9.4.7 Aromatically Substituted Nitrogen Heterocycles.



Compounds **44-49** all exhibit similar characteristics, with low contact scores and better than average energy scores, with a number of notable exceptions. Both **46**

and **49** both have lower scores than the other compounds, indeed the energy score is considerably lower than most other compounds studied. A reason for this cannot immediately be ascertained. For **49**, three orientations were modelled, two isomers and another structure from the crystal structure, all molecules showed a very tight fit to the enzyme with a very low number of orientations that bound. Although the contact scores for **49** are good the energy scores are some of the lowest seen for any substrate, suggesting that this would be a poor substrate for the enzyme. The only other molecule of note is **47** which has a slightly poorer energy score than the others but more importantly the rmsd between the two different scored molecules is significantly larger than the other molecules. Compounds **50** and **51** are structurally different from the other piperidine molecules studied however both bind extremely well with low scores for both the energy and contact scoring methods. For all these molecules, despite a large number of possible orientations, a low number of scored orientations were observed, indicating limited movement in the binding site and thus potentially high stereoselectivity.

Table 9.8: DOCK scores for 44-51.

compound	total conformations	orientations tried	orientations scored	contact score	energy score	RMSD
<b>44</b>	40	80000000	142	-217.00	-19.45	0.55
<b>45</b>	4	8000000	292	-272.00	-23.85	0.51
<b>46</b>	8	16000000	402	-173.00	-11.36	1.26
<b>47</b>	40	80000000	2196	-285.00	-17.77	5.72
<b>48</b>	40	80000000	142	-217.00	-22.18	0.55
<b>49-A</b>	40	40000000	2	-187.00	-11.81	0.56
<b>B</b>	40	80000000	3	-202.00	-7.43	0.51
<b>C</b>	40	80000000	4	-215.00	-6.13	1.55
<b>50</b>	4	8000000	1516	-254.00	-21.89	2.37
<b>51</b>	3	6000000	2157	-264.00	-24.24	2.31

9.4.7.1 DOCK orientations of **44**, **45**, **46**, **47**, **48**, **49**, **50** and **51**.

The energy scored and the contact scored orientations are essentially the same for **44**. The phenyl ring occupies the upper phenyl pocket, the piperidine ring sits towards the 295/297 pocket, with the contact scored molecule being slightly nearer Asp 297 than the energy scored one. In both cases the C2, C3 and C4 hydrogens are near the heme, the C3 ones being the most likely candidate for hydroxylation.

For **45** the energy scored and contact scored molecules are in identical orientations. The phenyl ring is in the upper phenyl pocket with the carbonyl pointing towards Leu 244, with no discernable hydrogen bond. The piperidine ring is orientated towards the 295/297 pocket so that the methyl groups sit above the heme ring. The nearest hydroxylation site for the iron is one of the methyl groups, none of the ring hydrogens are in range.

Again for **46** both orientation are similar. The phenyl ring is in the upper pocket, the lower ring points towards the 297/295 pocket so that the ethyl group can orientate towards Thr 101 and Val 244. No apparent H-bond to the carbonyl exist. The nearest hydrogen bonds to the heme are the C3 hydrogens and the hydrogens on the first carbon of the ethyl group, the C3 hydrogens being in a slightly better orientation for hydroxylation.

Both the contact scored and energy scored orientation are the same for **47**. Again **47** is another instance of the aliphatic ring preferentially in the upper phenyl pocket over a aromatic ring. With the piperidine ring in the standard upper ring position. Neither of the oxygens in the ester linkage have potential hydrogen bonds. The phenyl ring orientates itself in the 295/297 pocket. The C2 on the phenyl ring is the closest hydrogen to the heme centre but is not in a good hydroxylating position.

Both the **48** orientations look very strange at first glance, similar to that observed for **43**. The energy scored molecule is doubled back on its self, with both the phenyl and piperidine rings pointing towards the heme. The intra molecular distances are too close for this to be an orientation. The contact scored molecule has the phenyl ring towards Asp 297 and the piperidine pointing up towards the upper phenyl

pocket. Again the molecule is bent back upon itself such that there are implausible intramolecular distances.

There were three orientations modelled for **49**, the pseudo stereoisomers about the amide bond (around the partial nitrogen-carbon double bond) and the configuration from the crystal structure, figure 9.13.

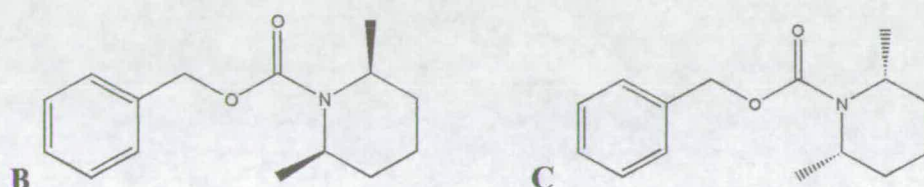


Figure 9.13: Stereo isomers B and C of substrate **49**.

The conformation from the crystal structure, A, (same amide bond conformation as B ) has the same orientation for contact scored and energy scored molecules, with the phenyl ring in the upper phenyl pocket. The piperidine ring sits towards the 297/295 pocket, so that the methyl groups point towards Val 396, figure 9.14.

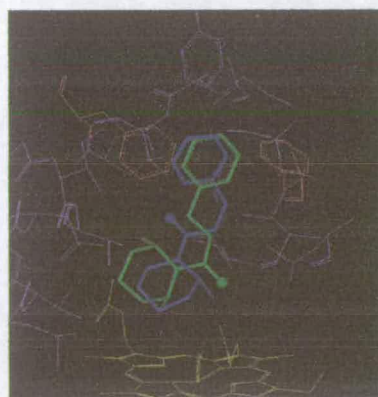


Figure 9.14: Orientations of isomer 1 and 2 for compound **49**.

The C2 and C2 methyl hydrogens are the nearest to the iron centre. The isomer with the same original stereochemistry as the crystal structure configuration, B, under went a rearrangement to adopt the same orientation as A. For the other

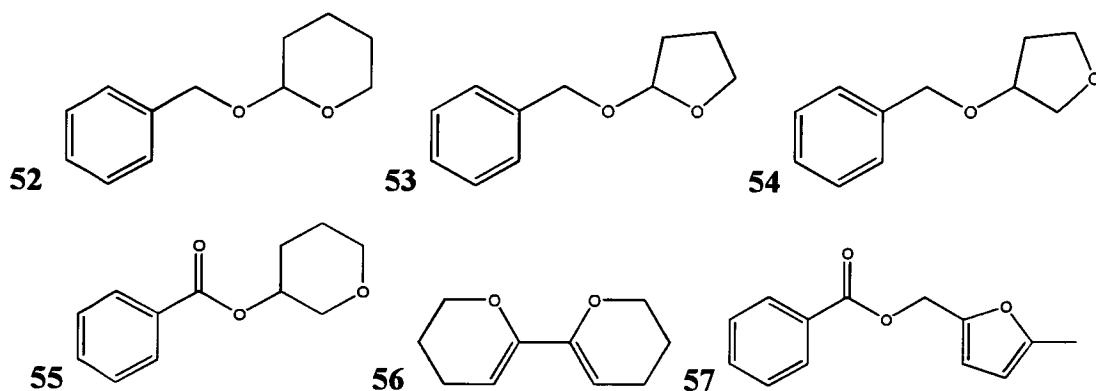


isomer, C, the phenyl ring in the phenyl pocket however the methyl groups pointed towards the heme, with the carbonyl group pointing towards Phe 87. This placed the methyl or trans-C5 hydrogen nearest the heme. It is interesting to note that stereo isomer C actually under went a rotation around the pseudo double bond to change to isomer B (DOCK treats partial double bonds as single ones when performing orientation adjustments). The reason the molecule has not orientated itself to the same location as for B is because DOCK performs the rotation *in situ* and it is not possible to move the molecule into the orientation for isomer B. This would suggest that conformation B binds significantly better to the enzyme than conformation C.

The phenyl ring of the energy scored molecule of 50 fits into the usual position in the upper pocket. This orientates the bicyclic ring towards the 295/297 pocket. The carbonyl groups are not hydrogen bonded to anything. The bicyclic ring carbonyl is the nearest group to the heme and no hydroxylation would seem possible. The contact scored molecule has the same upper ring conformation however the bicyclic ring is pointing over the heme and not towards the lower pocket. This orientation is the equivalent of the camphor position of the bicyclic ring and the nearest carbon-hydrogen bond for hydroxylation is the equivalent of the 5-*exo* position on camphor.

In both cases for 51 the phenyl ring is in the phenyl pocket in the normal orientation. There are slight differences in the adamantamine ring orientation with the energy scored molecule slightly towards the Asp 297 pocket the contact scored molecule is more centrally orientated. For the contact scored molecule the C6, C8 and C10 axial hydrogens are all accessible however for the energy scored molecule both the hydrogens on the C6 carbon are the nearest to the heme.

## 9.4.8 Oxygen-Heterocycles with Aromatic Sidechains.



Compounds **52-55** essentially have the same scores, table 9.9. Both the energy and contact scores imply that these compounds would make good substrates, with **55** having the one of the best energy scores of any compound. However there is a high number of scored orientations suggesting a potential lack of stereoselectivity. For **56** the scores are slightly lower than for the other oxygen heterocycles however it would still appear to be a good substrate. **57**, however, has a lower number of scored orientations despite a high number of conformations. This must mean a good substrate/enzyme recognition, which is supported by a good contact score. The energy score for **57** is extremely poor due to intramolecular clashes in this orientation. This would suggest that despite the close similarity to other substrates **57** could have very poor binding potential.

Table 9.9: DOCK scores for 52-57.

Compound	total conformations	orientations tried	orientations scored	contact score	energy score	RMSD
<b>52</b>	30	60000000	16871	-250.00	-24.23	2.35
<b>53</b>	30	60000000	44199	-269.00	-30.43	4.60
<b>54</b>	30	60000000	23071	-275.00	-25.45	4.98
<b>55</b>	12	24000000	5673	-265.00	-28.47	2.89
<b>56</b>	2	4000000	10799	-205.00	-18.85	4.90
<b>57</b>	40	80000000	1606	-206.00	38.98	5.51

#### 9.4.8.1 DOCK orientation of 52, 53, 54, 55, 56 and 57.

Both the contact scored and the energy scored molecules for **52** have the phenyl ring in the upper pocket, however the pyran rings differ in orientation significantly. The energy scored molecule is orientated towards the 295/297 pocket so that the pyran oxygen and the trans-C3 hydrogen are nearest the heme. The contact scored molecule locates the pyran ring towards the Leu 244 pocket with the pyran oxygen and the trans-C6 hydrogen near the heme. With the closer orientation of the oxygen this should not produce a product.

The energy scored molecule for **53** has its phenyl ring in the upper phenyl pocket. This orientates the furan ring towards the heme, and slightly towards Leu 244. This places the trans-C5 hydrogen and to a lesser extent the trans C3 and C4 hydrogens in hydroxylation distance from the iron. This visually has a very good substrate binding for hydroxylation. The contact scored molecule has a strange orientation, seen for **43** and **48** with the phenyl ring doubled back on the furan ring, both above the heme, with the phenyl group towards Leu 244 and the furan towards the 295/297 pocket. The contact scored molecule does not seem to be a realistic orientation as the intramolecular distances are too short.

Again the orientations of **54** are similar to **43**, **48** and **53**. The contact scored molecule is in the 'doubled up' position with the phenyl towards Leu 244 and the

furan towards 295/297. The energy scored molecule orientates the furan ring towards Val 295 but not Asp 297 so that the cis-C4 hydrogen lies above the heme. However the energy scored molecule also orientates the furan oxygen to a position very close to the iron centre suggesting strong oxygen binding might be possible.

The contact scored and the energy scored molecules of **55** look here like mirror images of each other with both molecules having the phenyl ring in the upper phenyl pocket. The energy scored molecule has the carbonyl pointing towards Val 396 with the furan ring in the 295/297 pocket. This places the trans-C4 hydrogen near the iron centre. The furan oxygen is slightly further from the iron than the previous case but still might interfere with turnover. The contact scored molecule is orientated to place the carbonyl towards Leu 244 so that the furan ring sits in the 244 pocket. This places the trans-C4 hydrogen nearest the heme, again with the furan oxygen close to the iron.

The energy scored molecule for **56** lies with one ring in the upper pocket and the other ring near Thr 101. The oxygen on the upper ring is nearest to Ile 395 and in the lower ring to Leu 244. There are no easily discernable hydrogen bonds to either oxygen. The nearest carbon to the iron center is on C5 on the lower ring, i.e. in the para position. If the lower ring was to move into a hydroxylating distance to the heme centre it should hydroxylate the equatorial hydrogen in preference to the axial. No other hydroxylation sites would seem to be preferred. The contact scored molecule lies perpendicular to the energy scored molecule (one parallel to the heme) in the orientation seen for the naphthalene derivatives

The energy scored orientation of **57** has the phenyl ring in a lower position than the more normal orientation in the upper pocket. The furan ring is then bent round almost back double on the phenyl ring so that the methyl group is pointing towards Asp 251. This is a bizarre orientation possibly due to the length of the inflexible region between the furyl methyl group and the ether linkage. The C4 hydrogen is the only one in hydroxylation range of the iron. The contact scored orientation is equally strange with the furyl ring in the upper pocket and the phenyl ring above the heme, parallel to it. In this case the C2 and C3 positions on the phenyl ring are in the best position for hydroxylation.

## 9.5 Conclusions.

The results here clearly show that DOCK can predict substrates and dismiss non-substrates for the cytochrome P-450cam Y96A mutant system. The main question that remains is whether or not these predictions are valid. With the availability of binding and turnover data we can verify the DOCK scores and the importance of the number of orientations that each molecule has. For the prediction of substrate orientations we have available six crystal structures that we can use to test the reliability of this aspect of the DOCK program.

### 9.5.1 Validation of the DOCK Output Data.

The comparison of the data for the DOCK runs and the experimental data is presented in table 9.10.

For the prediction of whether a compound was a substrate for the P-450cam mutant system or not would seem to be supported by the data from the DOCK runs. Substrates **37**, **60**, and **61** have been shown not to be substrates for the mutant. From this and previous studies it has been shown that these compounds are indeed not substrates for the P-450cam mutant enzyme. None of the other compounds predicted as substrates have been shown to be non-substrates for the enzyme, unlike camphor in previous studies.<sup>15-18</sup> Thus DOCK appears a suitable method for the prediction of substrates for this mutant enzyme.

The compounds with the largest number of scored orientations are the camphor and adamantane derivatives (>20,000 orientations scored) which all show poor product stereoselectivity. The compounds with the lowest number of scored orientations are the aromatically substituted piperidines (**44-49**) (<1000 orientations scored); these molecules show good stereoselectivity. The compounds with orientations between these values have stereoselectivity that is appropriate to their scores. This would thus suggest that DOCK can successfully predict the stereoselectivity of substrates based on the number of scored orientations.

The energy scoring method produces very similar scores for most molecules ( $\sim -20$ ) thus suggesting that the method is not able to distinguish the binding properties based on the electrostatic interactions with the binding site. However molecules with a low energy score ( $>-15$ ) proved to have high binding constants. The contact scoring produced a range of values for the compounds showing that DOCK is able to distinguish them from each other.

The contact score would seem to reflect the size of the molecules with scores of  $>-200$  for small molecules,  $-200$  to  $-250$  for the medium sized molecules and scores of  $<-250$  for the molecules that fit well into the binding site. Once molecules become too big their scores reflect this by increasing to values  $>-200$ . The contact score can thus be used to determine how tight the fit is between the molecule and the binding site; parameters such as binding constants rely on other factors such as the substrate hydrophobicity etc. therefore cannot be determined from this data.

The more useful information produced by the scoring is the lowest energy orientation. The validity of this can be measured against the six available crystal structures and the known hydroxylation sites. DOCK was able to predict the crystal orientations of all the molecules studied, either with the energy scored or the contact scored molecules. In the cases where the substrate was found to sit away from the heme towards the Val 295 pocket, DOCK successfully predicted this (phenylcyclohexane and bicyclohexyl). When the substrate was found to lie over the heme center again, DOCK predicts this (diphenylmethane). When it came to the more difficult molecule, **49**, DOCK again managed to predict the binding orientations of the molecule.



Table 9.10: Comparison of DOCK predictions with experimental data.

Compound	binding prediction from DOCK	experimental binding data	prediction from <sup>a</sup> DOCK of product	experimental product data
2	good	good	5- <i>exo</i> -hydroxy	mostly 5- <i>exo</i> --hydroxy
24	good	-	many	many
25	poor	-	1-epoxide	-
26	good	good	no product	no product
10	poor	poor	many	many
27	poor	poor	many	many
28	poor	poor	many	many
13	good	good	3,4-epoxide	3,4-epoxide
18	good	good	3,4-hydroxy	3,4-hydroxy
29	good	good	10-hydroxy	10-hydroxy
30	good	good	3,4-hydroxy	1 product
31	OK	OK	none	1 product
32	OK	OK	none	3,4-epoxide
33	good	good	no product	no product
35	good	-	no product	-
36	good	-	no product	-
37	poor	poor	none	1 product
38	poor	poor	none	no product
39	no	no	no product	no product
60	no	no	no product	no product
61	no	no	no product	no product
41	good	good	2,3-epoxide	1 epoxide
42	good	good	3,4-hydroxy	no product
43	good	OK	no product	no product
44	good	good	3-hydroxy	1 product
45	good	good	none	1 product
46	OK	good	3-hydroxy	2 products
47	good	OK	none	1 product
48	good	OK	none	1 product
49	OK	OK	5-hydroxy	1 product
50	good	good	5- <i>exo</i> -hydroxy	no product
51	good	-	many	-
52	good	OK	no product	no product
53	good	OK	no product	no product
54	good	OK	no product	no product
55	good	OK	no product	no product
56	OK	OK	5-hydroxy	no product
57	poor	-	no product	no product

a- for prediction of product: none is used when data is unclear if any product will be produced, no product is used when that is the predicted result.

The data from known hydroxylation sites or oxygen ligation would also seem to be supported by the DOCK orientations. The two adamantanols (27, 28) show the oxygen away from the heme which is consistent with the production of the diols. The oxygen heterocycles (52-56), which show oxygen ligation to the heme, are supported from DOCK data which shows that the oxygen molecules in the ring come into close proximity to the iron. Most molecules bind in the DOCK simulation with the para position of the non-aromatic ring nearest the heme which is consistent with the known binding properties and product data for these molecules.

Thus it would seem that DOCK can predict the binding of substrates, to some extent predict stereo- and regioselectivity of the oxidation and what the orientation of the molecule is in the binding site. This makes DOCK a far superior method for molecular modelling than the method used in the previous chapter (chapter 8).

## 9.6 References.

- 1)Kuntz, I. D.; Blaney, J. M.; Oatley, S. J.; Langridge, R.; Ferrin, T. E. *Journal of Molecular Biology* **1982**, *161*, 269-288.
- 2)Kuntz, I. *DOCK Version 4.0*; Regents of the University of California:, 1998.
- 3)Sun, Y.; Ewing, T. J. A.; Skillman, A. G.; Kuntz, I. D. *Journal of Computer-Aided Molecular Design* **1998**, *12*, 597-604.
- 4)Oshiro, C. M.; Kuntz, I. D. *Proteins-Structure Function and Genetics* **1998**, *30*, 321-336.
- 5)Makino, S.; Kuntz, I. D. *Journal of Computational Chemistry* **1997**, *18*, 1812-1825.
- 6)Ewing, T. J. A.; Kuntz, I. D. *Journal of Computational Chemistry* **1997**, *18*, 1175-1189.
- 7)Knegtel, R. M. A.; Kuntz, I. D.; Oshiro, C. M. *Journal of Molecular Biology* **1997**, *266*, 424-440.
- 8)Briem, H.; Kuntz, I. D. *Journal of Medicinal Chemistry* **1996**, *39*, 3401-3408.
- 9)Gschwend, D. A.; Good, A. C.; Kuntz, I. D. *Journal of Molecular Recognition* **1996**, *9*, 175-186.
- 10)Gschwend, D. A.; Kuntz, I. D. *Journal of Computer-Aided Molecular Design* **1996**, *10*, 123-132.
- 11)Oshiro, C. M.; Kuntz, I. D.; Dixon, J. S. *Journal of Computer-Aided Molecular Design* **1995**, *9*, 113-130.
- 12)Good, A. C.; Ewing, T. J. A.; Gschwend, D. A.; Kuntz, I. D. *Journal of Computer-Aided Molecular Design* **1995**, *9*, 1-12.
- 13)Shoichet, B. K.; Kuntz, I. D. *Protein Engineering* **1993**, *6*, 723-732.
- 14)Meng, E. C.; Gschwend, D. A.; Blaney, J. M.; Kuntz, I. D. *Proteins-Structure Function and Genetics* **1993**, *17*, 266-278.
- 15)DeVoss, J. J.; Sibbesen, O.; Zhang, Z. P.; DeMontellano, P. R. O. *Journal of the American Chemical Society* **1997**, *119*, 5489-5498.
- 16)DeVoss, J. J.; DeMontellano, P. R. O. *Methods in Enzymology* **1996**, *272*, 336-347.

17) DeVoss, J. J.; DeMontellano, P. R. O. *Journal of the American Chemical Society* **1995**, *117*, 4185-4186.

18) Zhang, Z. P.; Sibbesen, O.; Johnson, R. A.; DeMontellano, P. R. O. *Bioorganic & Medicinal Chemistry* **1998**, *6*, 1501-1508.

## 10 Future Work.

---

### 10.1 Expansion of Current Studies.

#### 10.1.1 X-ray Crystallography.

Since we have discovered a large number of substrates for the cytochrome P-450cam Y96A mutant, it could be of interest to expand the number of crystal structures. The current substrates are of a similar type and shape and it would be of interest to expand the knowledge of the cytochrome P-450cam binding site with new substrates. This might give an insight into the binding of halogenated compounds or how the naphthalene analogues bind.

For the substrates with low solubility, crystallization by soaking the substrate into the crystal, rather than co-crystallization, might be a preferred method and conditions would need to be developed to perform this.

If other mutants are made it would be essential to have a x-ray crystal structure of the protein; as this study has shown the difficulty in predicting the changes made to the binding site upon mutagenesis.

### 10.1.2 Molecular Modelling.

DOCK has been shown to be relatively accurate in predicting the binding properties of a variety of substrates. It should prove of interest to examine other potential cytochrome P-450cam substrates with this technique. DOCK has, additionally, a search function that can examine commercial databases for substrates. If this was employed, a complete range of the Y96A substrates could be determined and then further analysis could be made of any that have potentially interesting products.

## 10.2 Scale-up of Cytochrome P-450cam Biotransformation.

A potential problem in the large scale production of hydroxylated products by the P-450cam is the fact that its catalytic cycle relies on NADH and two co-proteins, thereby tripling protein expression and purification costs. There are three possible solutions to this problem, which have not been extensively studied to date. One is the use of electrochemistry, that is attachment of the cytochrome P-450cam directly to an electrode thus removing the need for NADH, Pd or PdR. This has been attempted by research groups in both Oxford and Japan<sup>1-5</sup> however turnover is low and the protein denatures after a short time, preventing large scale production of hydroxylated products.

An alternative is to use different electron transfer agents.<sup>6</sup> NADH turnover in the absence of Pd and PdR has been attempted previously<sup>7</sup> and been reported here. In all cases however the product yields are significantly smaller than the reconstituted system. If a new chemical agent becomes available then this technique might become viable although currently it is not commercially viable.

The third alternative to enhance product turnover is the use of whole cell systems. Whole cell systems have the advantage of no protein purification of NADH addition is necessary as the host organism provides all necessary electron transfer agents. Whole cell hydroxylation can be performed with simple addition of substrate to a growing/resting cell culture and leaving it to metabolize the substrate. Product

isolation can be performed by organic extraction and one step separation from starting materials.

It would seem from this that whole cell systems are better than purified systems however there are disadvantages to their use. In order to generate mutants of the required protein the gene must first be transferred to *E.coli* to perform the mutagenesis before being placed back into the host organism. This would therefore require shuttle vectors that are compatible with both the host and *E.coli*, which is severely limiting. The most efficient method would be to find a well characterized host system that has generic electron transfer proteins that will be compatible with all added cytochrome P-450s. The a suitable shuttle vector should be designed that is compatible with both *E.coli* and the host organism. This should be the most commercially viable method, although will require much exploratory experimentation. Even if this system was devised whole cell systems have one other disadvantage over the use of recombinant systems in that the required products might be degraded by other enzymes in the host organism, especially problematic are esterases.

A alternative to transferring the gene for P-450cam back and forward between expression systems would be to use *E.coli* as the whole cell system. Our experiments with the *E.coli* reductases indicate that suitable co-proteins are not available in *E.coli*. One solution to this would be to express the P-450cam gene as a triple fusion protein (mimicking cytochrome P450BM-3). This was attempted by Montellano and co-workers, although product yields were low.<sup>8</sup> Further experimentation with this system might allow successful product formation with *E.coli* as a biotransformant.

### 10.3 Further Mutagenesis of Cytochrome P-450cam.

Further mutagenesis of P-450cam should expand the substrate range, could prevent peroxide production or enhance stereoselectivity. For further expansion of the substrate range Phe 87 is a good residue to pick since it has an effect on both the binding site and substrate entry. Phe 193 is also a potential target as it would allow longer substrates that stick out of the binding site similar to the Pfizer inhibitors.<sup>9</sup> As



discussed earlier Thr 101 could be responsible for the increased peroxide production in the Y96A mutant and thus mutagenesis could remove this problem. Other mutations to the binding site around Thr 252 could also be made to prevent peroxide production however change to the delicate proton transfer pathway is potentially disruptive to turnover rates. If a particular stereoselective product is desired, small subtle changes to the binding site, especially the lower portion nearer the heme, could be made. Amino acids Val 244, and Val 295 are especially good targets for this as both control the hydrophobic pockets either side of the heme.

If a particular target substrate is desired one technique that can be applied is directed evolution. One experiment that has recently been applied to cytochrome P-450cam is random mutagenesis. Arnold and coworkers enhanced the ability of P-450cam to hydroxylate naphthalene by use of directed evolution.<sup>10</sup> The main disadvantage of this methodology is the screening protocols. In order to find an effective mutant a quick simple screening protocol must be developed preferably using a colorimetric assay. This currently limits the substrate range to compounds that change colour upon hydroxylation and ones that can be easily converted to coloured compounds.

#### **10.4 NMR Studies.**

Although, due to its size, the NMR structure of the P-450cam Y96A mutant cannot be determined, other NMR experiments are possible which will provide useful data on the substrate in the binding site. Structural information on the P-450cam binding site by NMR has been attempted as long ago as 1975, when water was detected in the vicinity of the heme.<sup>11</sup> Since then experiments have been performed on P-450's to detect the binding of substrate<sup>12,13</sup> and to observe changes to the binding site after mutagenesis.<sup>14</sup> It is also possible to detect the movement of residues in the binding site upon substrate binding.<sup>15-17</sup> Using information from the crystal structure of the mutant enzyme with bound substrates and from NMR protein-substrate experiments, it should be possible to obtain dynamic information not accessible by X-

ray crystallography. Although preliminary experiments would suggest that this approach is not as quick and simple as literature experiments might imply (data not shown).

### **10.5 Molecular Dynamics.**

Another area of potential research is extensive computational studies of the P-450cam mutant protein. Molecular dynamics (MD) has been extensively used over the last five years to model aspects of the protein that cannot be determined by the crystal structure alone. These have included MD, explaining difference between predicted results (from X-ray structure) and experimental results of the thiocamphor hydroxylation by P450cam.<sup>18</sup> Extensive attempts have been made to predict products and product distribution from MD alone<sup>19,20</sup> and with the benefit of experimental verification.<sup>21-25</sup> It is this potential prediction of products that is of major interest to us. The modelling studies to date have only given crude details of the fit of a substrate within the binding site. MD can be used to give more realistic data on the enzyme substrate interaction, by allowing for bond flexibility of both the enzyme binding site residues and the substrate. Thus MD, in time, could accurately predict the products of mutant P-450cam hydroxylation without the need for any experimentation.

Initial studies have indicated that this approach is successful at predicting the interaction between the substrate and the P-450cam Y96A mutant enzyme (data not shown). Although further analysis in this area is needed.

## 10.6 References.

- 1) Estabrook, R. W.; Faulkner, K. M.; Shet, M. S.; Fisher, C. W. *Methods in Enzymology* **1996**, *272*, 44-51.
- 2) Sugihara, N.; Ogoma, Y.; Abe, K.; Kondo, Y.; Akaike, T. *Polymers for Advanced Technologies* **1998**, *9*, 307-313.
- 3) Zhang, Z.; Nassar, A. E. F.; Lu, Z. Q.; Schenkman, J. B.; Rusling, J. F. *Journal of the Chemical Society-Faraday Transactions* **1997**, *93*, 1769-1774.
- 4) Sugihara, N.; Ogoma, Y.; Abe, K.; Murakami, Y.; Kondo, Y.; Akaike, T. *Polymers for Advanced Technologies* **1998**, *9*, 858-860.
- 5) Lvov, Y. M.; Lu, Z. Q.; Schenkman, J. B.; Zu, X. L.; Rusling, J. F. *Journal of the American Chemical Society* **1998**, *120*, 4073-4080.
- 6) Sligar, S. G. *Biochemistry* **1976**, *15*, 5399-5406.
- 7) Eble, K. S.; Dawson, J. H. *Biochemistry* **1984**, *23*, 2068-2073.
- 8) Sibbesen, O.; DeVoss, J. J.; DeMontellano, P. R. O. *Journal of Biological Chemistry* **1996**, *271*, 22462-22469.
- 9) Raag, R.; Li, H. Y.; Jones, B. C.; Poulos, T. L. *Biochemistry* **1993**, *32*, 4571-4578.
- 10) Joo, H.; Lin, Z. L.; Arnold, F. H. *Nature* **1999**, *399*, 670-673.
- 11) Griffin, B. W.; Peterson, J. A. *Journal of Biological Chemistry* **1975**, *250*, 6445-6451.
- 12) Banci, L.; Bertini, I.; Marconi, S.; Pierattelli, R. *European Journal of Biochemistry* **1993**, *215*, 431-437.
- 13) Banci, L.; Bertini, I.; Marconi, S.; Pierattelli, R.; Sligar, S. G. *Journal of the American Chemical Society* **1994**, *116*, 4866-4873.
- 14) Wakasugi, K.; Ishimori, K.; Morishima, I. *Biochimie* **1996**, *78*, 763-770.
- 15) Modi, S.; Primrose, W. U.; Boyle, J. M. B.; Gibson, C. F.; Lian, L. Y.; Roberts, G. C. K. *Biochemistry* **1995**, *34*, 8982-8988.
- 16) Modi, S.; Boyle, W. U. P. J. M. B.; Gibson, C. F.; Lian, L.-Y.; Roberts, G. C. K. *Biochemistry* **1995**, *34*, 8982-8988.

- 17) Modi, S.; Paine, M. J.; Sutcliffe, M. J.; Lian, L. Y.; Primrose, W. U.; Wolf, C. R.; Roberts, G. C. K. *Biochemistry* **1996**, *35*, 4540-4550.
- 18) Paulsen, M. D.; Ornstein, R. L. *Protein Engineering* **1993**, *6*, 359-365.
- 19) Harris, D.; Loew, G. *Journal of the American Chemical Society* **1995**, *117*, 2738-2746.
- 20) DeVoss, J. J.; DeMontellano, P. R. O. . *Journal of the American Chemical Society* **1995**, *117*, 4185-4186.
- 21) Demontellano, P. R. O.; Fruetel, J. A.; Collins, J. R.; Camper, D. L.; Loew, G. H. *Journal of the American Chemical Society* **1991**, *113*, 3195-3196.
- 22) Fruetel, J. A.; Collins, J. R.; Camper, D. L.; Loew, G. H.; Demontellano, P. R. O. *Journal of the American Chemical Society* **1992**, *114*, 6987-6993.
- 23) Loida, P. J.; Sligar, S. G.; Paulsen, M. D.; Arnold, G. E.; Ornstein, R. L. *Journal of Biological Chemistry* **1995**, *270*, 5326-5330.
- 24) Jones, J. P.; Trager, W. F.; Carlson, T. J. . *Journal of the American Chemical Society* **1993**, *115*, 381-387.
- 25) Manchester, J. I.; Ornstein, R. L. *Protein Engineering* **1995**, *8*, 801-807.

## 11 Appendix 1: Example datasets for the calculation of the maximum percentage binding and $K_D$ .

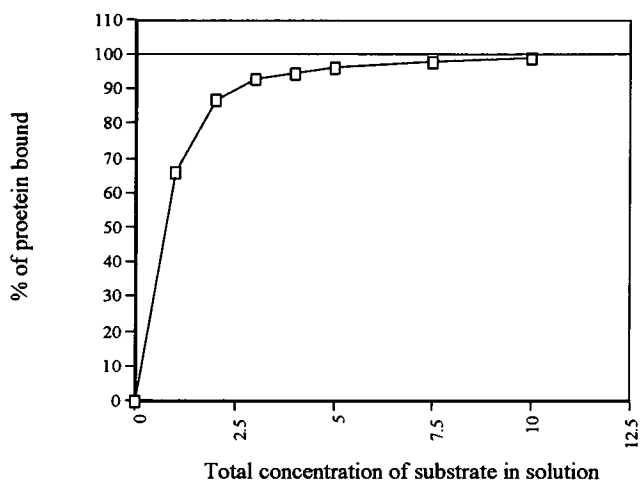


Figure 11.1: An example of the percentage binding curve generated by the method outlined in Chapter 4. In this case the substrate was compound 46.

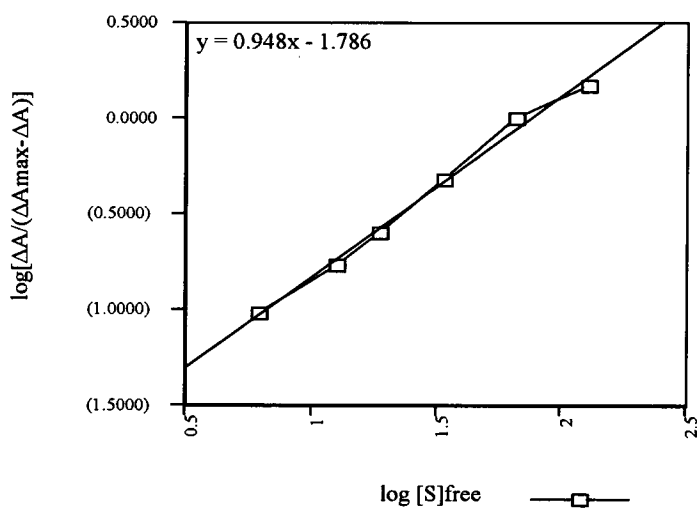


Figure 11.2: Example graph for calculation of  $K_D$ , in this case for compound 43.

## 12 Appendix 2: The DOCK instruction file.

```

flexible_ligand      yes
orient_ligand       yes
score_ligand        yes
minimize_ligand     yes
multiple_ligands    no
random_seed         0
anchor_search       no
torsion_drive       yes
clash_overlap       0.5
conformation_cutoff_factor 10
torsion_minimize    yes
match_receptor_sites yes
random_search       yes
ligand_centers      no
automated_matching  yes
maximum_orientations 20000
write_orientations  yes
rank_orientations   yes
rank_orientation_total 100
write_configurations yes
write_configuration_total 10
intramolecular_score yes
intermolecular_score yes
gridded_score       yes
grid_version        4
bump_filter         yes
bump_maximum        0
contact_score       yes
contact_cutoff_distance 4.5
contact_clash_overlap 0.75
contact_clash_penalty 50
chemical_score      no
energy_score        yes
energy_cutoff_distance 10
distance_dielectric yes
dielectric_factor   4
attractive_exponent 6
repulsive_exponent  12
atom_model          u
vdw_scale           1
electrostatic_scale 1
rmsd_score          no

```

contact_minimize	yes
energy_minimize	yes
initial_translation	1
initial_rotation	0.1
initial_torsion	10
maximum_iterations	100
contact_convergence	0.1
energy_convergence	0.1
maximum_cycles	1
ligand_atom_file	lig.mol2
receptor_site_file	near.sph
score_grid_prefix	grid
vdw_definition_file	/hosts/gordon/usr6/adams/dock_4.0.1/parameter/vdw.defn
flex_definition_file	/hosts/gordon/usr6/adams/dock_4.0.1/parameter/flex.defn
flex_drive_file	/hosts/gordon/usr6/adams/dock_4.0.1/parameter/flex_drive.tbl
ligand_contact_file	lig_contact.mol2
ligand_energy_file	lig_energy.mol2



### 13 Appendix 3: Example of DOCK output file.

```

UUUUUUUUU   CCCCCC   SSSSSS   FF/ FFF/
UU/         UU/  CC/    CC/  SS/  SS/  FF/ FFF/
UU/         UU/  CC/    CC/  SS/          FFFFF/
UU/         UU/  CC/    CC/  SS/          FF/ FF\
UU/         UU/  CC/    CC/  SS/  SS/  FF/ FF\
UUUUUUUUU/  CCCCCC/  SSSSSS/  FF/  FF\

```

University of California at San Francisco, DOCK 4.0.1  
DOCK 4.0.1 was released on May 17, 1998.

#### Job Information

```

launch_time   Fri Dec 24 11:43:00 1999
host_name     gaffer
memory_limit  156712960
working_directory /hosts/gordon/usr6/adams/dock_4.0.1/new/camphor
user_name     adams

```

#### General Parameters

```

flexible_ligand  yes
orient_ligand    yes
score_ligand     yes
minimize_ligand yes
multiple_ligands no
random_seed      0

```

#### Flexible Ligand Parameters

```

anchor_search    no
torsion_drive    yes
clash_overlap    0.5
conformation_cutoff_factor 10
torsion_minimize yes

```

#### Orient Ligand Parameters

```

match_receptor_sites  yes
random_search         yes
ligand_centers        no
automated_matching    yes
maximum_orientations  10000
write_configurations  yes
write_configuration_total 10

```

#### Scoring Parameters

intramolecular_score	yes
intermolecular_score	yes
gridded_score	yes
grid_version	4
bump_filter	yes
bump_maximum	0
contact_score	yes
contact_cutoff_distance	4.5
contact_clash_overlap	0.75
contact_clash_penalty	50
chemical_score	no
energy_score	yes
energy_cutoff_distance	10
distance_dielectric	yes
dielectric_factor	4
attractive_exponent	6
repulsive_exponent	12
atom_model	u
vdw_scale	1
electrostatic_scale	1

---

Minimization\_Parameters

---

contact_minimize	yes
energy_minimize	yes
initial_translation	1
initial_rotation	0.1
initial_torsion	10
maximum_iterations	100
contact_convergence	0.1
energy_convergence	0.1
maximum_cycles	1

---

File\_Input

---

ligand_atom_file	lig.mol2
receptor_site_file	near.sph
score_grid_prefix	grid
vdw_definition_file	/hosts/gordon/usr6/adams/dock_4.0.1/parameter/vdw.defn
flex_definition_file	/hosts/gordon/usr6/adams/dock_4.0.1/parameter/flex.defn
flex_drive_file	/hosts/gordon/usr6/adams/dock_4.0.1/parameter/flex_drive.tbl

---

File\_Output

---

ligand_contact_file	lig_contact.mol2
ligand_energy_file	lig_energy.mol2

Reading general grid info from grid.bmp  
 Reading bump grid from grid.bmp  
 Reading contact grid from grid.cnt  
 Reading energy grids from grid.nrg  
 VDW grids use a 6-12 Lennard-Jones potential with a united atom model.  
 Reading attractive VDW energy grid.  
 Reading repulsive VDW energy grid.  
 Reading electrostatic energy grid.

### Docking\_Results

---

Name : CYHX\_B0  
 Description : \*\*\*\*  
 Total conformations : 1  
 Orientations tried : 788786  
 Orientations scored : 10000  
  
 Best contact score : -182.00  
   Intramolecular contact score : 0.00  
   Intermolecular contact score : -182.00  
 RMSD of best contact scorer (A) : 64.88  
  
 Best energy score : -22.17  
   Intramolecular energy score : 0.00  
   Intermolecular energy score : -22.17  
 RMSD of best energy scorer (A) : 64.46  
  
 Elapsed cpu time (sec) : 206.51

Finished processing molecule in 206.545 seconds.