

# **STRUCTURAL STUDIES OF CYTOCHROME**

## **P450 BM3 AND CPRK**

**Thesis submitted for the degree of Doctor of Philosophy**

**at the University of Leicester**

**by**

**Michael Gordon Joyce, B.Sc. (Hons)**

**Department of Biochemistry**

**Faculty of Medicine and Biological Sciences**

**University of Leicester**

**December 2005**

UMI Number: U203781

All rights reserved

INFORMATION TO ALL USERS

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

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

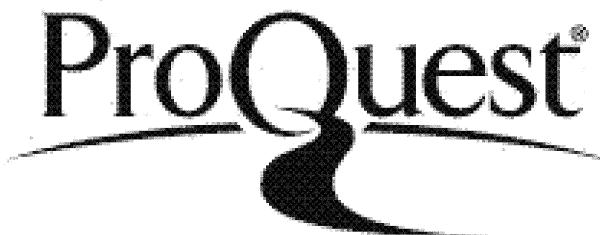


UMI U203781

Published by ProQuest LLC 2013. Copyright in the Dissertation held by the Author.

Microform Edition © ProQuest LLC.

All rights reserved. This work is protected against unauthorized copying under Title 17, United States Code.



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

---

## **ACKNOWLEDGEMENTS**

I would like to sincerely thank my supervisor Dr. David Leys for allowing me to carry out my PhD studies in his laboratory. His expertise and instruction during the PhD was critical in the successful determination of the protein structures discussed herein. Apart from this were the many adventures in Hamburg and Grenoble. I am also most grateful for the advice and instruction during the final writing up period.

I would like to thank the BBSRC and the University of Leicester for providing the funding for my PhD studies without which I would not have been able to carry out this research.

I would like to thank Dr. Jaswir Basran for instruction in the fluorescence quenching experiments. I thank Dr. Parvinder Hothi and Mr. Phil Balding for their help in working with Origin. I must also thank Dr. Kirsty McLean for instruction and advice in the determination of binding constants for both the wild type and mutant forms of CprK.

I am grateful to Prof. Andrew Munro for collaborations and highly expert advice on BM3 and the many interesting conversations on this topic throughout my PhD studies. I thank all of his group members for their company on many evenings during the years and especially Mr. Joe McVey for introducing me to rock climbing in the Peak District. I must thank Dr. Olivier Roitel, Dr. Rajasekhar Neeli and Dr. Hazel Girvan for their contributions to the successful structure determination of the FAD domain and the A264E mutant of the heme domain of BM3 without whose help, might not have been possible.

I have to thank Dr. Jonas Emsley, Dr. Alexandre Gingras and Dr. Vagelis Papagrigoriou for many good nights out having fun and adventures around the city of Leicester. There will be many stories from my times in Leicester with you guys.

I must thank all of the members of the Leys lab who made the day to day work bright and enjoyable. Thanks for all of their help go to Dr. Helen Toogood, Dr. Gergely Katona, Mr Ralph Wigley, Dr. Stephen Fullerton and Dr. Anna Roujeinikova who were always happy to answer questions and have a laugh with throughout the years. I must single out Mr. Adam van Thiel for many major discussions on hugely diverse topics.

I am very grateful to Prof. Nigel Scrutton and Dr. Peter Moody for use of their labs and equipment and also for many useful discussions during the three years of the PhD.

I must thank our collaborators Dr.'s van de Pas, de Vos, van de Oost, Smidt and Miss Krisztina Gabor from Wageningen University in the Netherlands for the kind donation of the CprK clone and also many discussions and interest throughout this project.

A special mention must go to Anne who supported me throughout the years of my PhD. She was always encouraging and ever forgiving for the late nights and weekends spent purifying protein or working on data even when things didn't work.

Finally I would like to thank my parents Michael and Veronica who are at all times a source of strength, love and understanding in all of my endeavors.

---

# **STRUCTURAL STUDIES OF CYTOCHROME P450 BM3 and CPRK**

**Michael Gordon Joyce**

## **ABSTRACT**

This work presents the crystal structure determination of the transcriptional regulator CprK and of individual domains of the multidomain cytochrome P450-BM3. The latter is a paradigm system for study of the mechanism of eukaryotic P450 systems as it presents a natural linkage between a diflavin reductase and a cytochrome P450. The crystal structure of the A264E mutant heme domain was determined with and without substrate present. Surprisingly, the structures reveal the protein to exhibit a substrate bound conformation regardless of the presence of substrate. This has provided further evidence that substrate binding leads to a dramatic shift in the equilibrium of conformational states available to the protein. In addition, the crystal structure of the C773A mutant flavin binding domain has been determined both in presence and absence of NADP<sup>+</sup>. Together with the already available structures of the other domains, this now allows both modelling and further solution studies of the full length cytochrome P450-BM3 structure.

Dehalogenans sp. are capable of using a range of chlorophenolic compounds as terminal electron acceptors in a respiratory metabolism known as halorespiration. This process is under transcriptional control by CprK, a member of the CRP-FNR family of transcriptional regulators. The crystal structure of *D. hafniense* CprK in complex with o-chlorophenolacetic acid (OPA) reveals tightly bound effector molecules. Binding of OPA is analysed through both mutagenesis and fluorescence quenching binding studies. The results have led to the hypothesis that CprK uses the bound phenolic compound pK<sub>a</sub> as an additional mechanism to sense the presence of the chloride atom. In addition, the unexpected interaction between individual DNA-binding domains suggests a possible physiological role for this association in the absence of an effector molecule. The structure presents a structural framework for further studies of the mechanism of this family of transcriptional regulators and of CprK homologues in particular.

---

# Contents

<b>Acknowledgements.....</b>	<b>i</b>
<b>Abstract.....</b>	<b>ii</b>
<b>Table of Contents.....</b>	<b>iii</b>
<b>Abbreviations.....</b>	<b>xi</b>
<b>CHAPTER 1 Introduction .....</b>	<b>1</b>
<b>1.1. Introduction .....</b>	<b>2</b>
<b>1.2. Principles of ligand binding and protein motion.....</b>	<b>2</b>
<b>1.2.1. Entropic and Enthalpic factors .....</b>	<b>2</b>
<b>1.2.2. Ligand Binding Specificity: Non-covalent interactions .....</b>	<b>4</b>
<b>1.2.2.1. London-van der Waals interactions .....</b>	<b>4</b>
<b>1.2.2.2. Hydrogen bonding.....</b>	<b>5</b>
<b>1.2.2.3. Electrostatic interactions .....</b>	<b>6</b>
<b>1.2.2.4. Hydrophobic interactions .....</b>	<b>6</b>
<b>1.3. Cytochromes P450 .....</b>	<b>7</b>
<b>1.3.1. Cytochrome P450 catalytic cycle .....</b>	<b>10</b>
<b>1.3.2. Cytochrome P450 structure .....</b>	<b>13</b>
<b>1.3.3. P450-Substrate complexes.....</b>	<b>14</b>
<b>1.3.4. Electron Transfer Partners of Cytochrome P450s .....</b>	<b>17</b>
<b>1.4. Cytochrome P450 BM3 .....</b>	<b>18</b>
<b>1.4.1. Active site mutagenesis .....</b>	<b>18</b>
<b>1.4.2. Electron transfer in cytochrome P450 BM3 .....</b>	<b>18</b>
<b>1.4.2.1. Structural analysis.....</b>	<b>20</b>
<b>1.5. Degradation of Halogenated phenolic compounds.....</b>	<b>22</b>
<b>1.5.1. Halorespiration .....</b>	<b>23</b>
<b>1.5.2. Desulfitobacterium sp.....</b>	<b>25</b>
<b>1.6. DNA binding proteins.....</b>	<b>26</b>
<b>1.6.1. Transcriptional Regulation .....</b>	<b>26</b>
<b>1.6.2. Helix-turn helix family of regulators.....</b>	<b>29</b>
<b>1.7. Aims of the project .....</b>	<b>32</b>

---

<b>CHAPTER 2 Introduction to crystallography .....</b>	<b>34</b>
<b>2.1. Introduction.....</b>	<b>35</b>
<b>2.2. Crystal growth.....</b>	<b>35</b>
<b>2.3. Cryo-protection .....</b>	<b>39</b>
<b>2.4. Experimental setup.....</b>	<b>39</b>
<b>2.4.1. Sources of X-rays .....</b>	<b>40</b>
<b>2.4.1.1. Home Sources .....</b>	<b>40</b>
<b>2.4.1.2. Synchrotron Radiation .....</b>	<b>42</b>
<b>2.4.2. Detection of X-rays .....</b>	<b>43</b>
<b>2.4.2.1. Image plate detectors.....</b>	<b>44</b>
<b>2.4.2.2. Charge-coupled devices (CCDs).....</b>	<b>45</b>
<b>2.5. X-ray Diffraction.....</b>	<b>45</b>
<b>2.5.1. Crystals .....</b>	<b>46</b>
<b>2.5.2. Symmetry .....</b>	<b>47</b>
<b>2.6. Basic Diffraction physics.....</b>	<b>47</b>
<b>2.6.1. Diffraction by one electron.....</b>	<b>47</b>
<b>2.6.2. Diffraction of X-rays from a lattice .....</b>	<b>48</b>
<b>2.6.3. Scattering by a system of two electrons .....</b>	<b>48</b>
<b>2.6.4. Scattering by an atom .....</b>	<b>50</b>
<b>2.6.5. Scattering by a crystal.....</b>	<b>51</b>
<b>2.6.6. X-ray Diffraction Data Collection and Processing.....</b>	<b>52</b>
<b>2.6.7. Calculation of the electron density .....</b>	<b>55</b>
<b>2.7. Structure Determination.....</b>	<b>57</b>
<b>2.7.1. The Phase problem .....</b>	<b>57</b>
<b>2.7.2. The Patterson function.....</b>	<b>58</b>
<b>2.7.3. Molecular Replacement .....</b>	<b>59</b>
<b>2.7.3.1. Rotation Function.....</b>	<b>60</b>
<b>2.7.3.2. Translation function .....</b>	<b>61</b>
<b>2.7.4. Isomorphous Replacement .....</b>	<b>62</b>
<b>2.7.4.1. Preparation of heavy atom derivatives .....</b>	<b>66</b>
<b>2.7.5. Anomalous Dispersion .....</b>	<b>67</b>
<b>2.7.6. Ab initio methods .....</b>	<b>70</b>
<b>2.7.7. Radiation induced phasing.....</b>	<b>71</b>
<b>2.7.8. Density modification .....</b>	<b>71</b>

---

---

2.7.8.1. Solvent flattening .....	72
2.7.8.2. Histogram matching .....	72
2.7.8.3. Non Crystallographic Symmetry.....	73
2.8. Model building and structure refinement .....	73
2.8.1. Least squares refinement .....	74
2.8.2. Maximum-likelihood based refinement.....	75
2.9. Structure Validation .....	75
2.10. High Throughput Technologies.....	76

---

<b>CHAPTER 3 Materials and Methods .....</b>	<b>77</b>
<b>3.1. Materials.....</b>	<b>78</b>
3.1.1. Chemicals and Reagents .....	78
3.1.2. Bacterial Strains and Media.....	78
3.1.3. Chromatographic Media and Membranes .....	78
3.1.4. Oligonucleotides and Sequencing .....	79
3.1.5. Crystallisation Screens and Plates .....	79
<b>3.2. Molecular Biology Methods .....</b>	<b>79</b>
3.2.1. DNA purification.....	79
3.2.2. Transformation of <i>E. coli</i> Cells with DNA.....	79
3.2.3. Agarose Gel Electrophoresis .....	80
<b>3.3. Protein Methods .....</b>	<b>80</b>
3.3.1. SDS Polyacrylamide Gel Electrophoresis .....	80
3.3.2. Refinement and analysis of structures .....	81
3.3.3. Analysis of protein sequences .....	81
3.3.4. Analysis of structural homologues .....	82
<b>3.4. Methods for Analysis of A264E mutant of Cytochrome P450 BM3 Heme domain.....</b>	<b>82</b>
3.4.1. Crystallization of P450 BM3 A264E.....	82
3.4.2. Data collection and processing .....	82
3.4.3. Structure solution of A264E mutant of Cytochromes P450 BM3 .....	83
3.4.4. Modelling of the active site mutants.....	84
<b>3.5. Methods for Analysis of FAD domain of Cytochrome P450 BM3 .....</b>	<b>84</b>
3.5.1. Crystallisation of the FAD domain of Cytochrome P450 BM3 .....	84
3.5.2. Data collection .....	85
3.5.3. Structure solution of FAD domain of Cytochrome P450 BM3 .....	85
<b>3.6. Methods for Analysis of CprK.....</b>	<b>85</b>
3.6.1. CprK Growth and Expression.....	85
3.6.2. CprK Purification .....	86
3.6.3. Determination of Protein Concentration.....	86
3.6.4. Crystallisation of CprK.....	86
3.6.5. Structure solution of CprK .....	87
3.6.5.1 Heavy atom derivitisation .....	87
3.6.5.2 Data collection .....	87

---

---

3.6.5.3. Scaling.....	87
3.6.5.4. Structure determination.....	88
3.6.6. Map improvement.....	88
3.6.7. Fluorimetry .....	88
3.6.8 Site-Directed Mutagenesis.....	89

---

<b>Chapter 4 Crystal structures of Cytochrome P450 BM3 A264E mutant heme domain .....</b>	<b>91</b>
<b>4.1. Introduction.....</b>	<b>92</b>
<b>4.2. Results .....</b>	<b>94</b>
<b>4.2.1. A264E mutagenesis, Enzyme Preparation and Crystallisation .....</b>	<b>94</b>
<b>4.2.2. Data collection and processing .....</b>	<b>95</b>
<b>4.2.3. Structure Determination by Molecular Replacement .....</b>	<b>96</b>
<b>4.2.4. Model Building and Refinement .....</b>	<b>98</b>
<b>4.2.4.1. Building and Refinement of BM3 A264E mutant .....</b>	<b>98</b>
<b>4.2.4.2. Building and Refinement of BM3 A264E mutant palmitoleate complex</b>	<b>99</b>
<b>4.3. Analysis of the BM3 A264E mutant and palmitoleate complex structures</b>	<b>99</b>
<b>4.3.1. BM3 A264E mutant.....</b>	<b>99</b>
<b>4.3.2. BM3 A264E mutant palmitoleate complex .....</b>	<b>101</b>
<b>4.4. Comparison of the BM3 A264E mutant with the wild type BM3 structure .....</b>	<b>104</b>
<b>4.4.1. Analysis of the active site .....</b>	<b>108</b>
<b>4.5. Implications for P450 mechanism .....</b>	<b>110</b>
<b>4.5.1. Analysis of the spin state conversion mechanism .....</b>	<b>111</b>
<b>4.6. Examination of the substrate-bound form .....</b>	<b>114</b>
<b>4.7. Possibility of covalent heme ligand generation .....</b>	<b>116</b>
<b>4.8. Conclusion .....</b>	<b>120</b>

---

<b>Chapter 5 Crystal structure of the cytochrome P450 BM3 FAD domain.....</b>	<b>122</b>
<b>5.1. Introduction.....</b>	<b>123</b>
<b>5.2. Results.....</b>	<b>124</b>
<b>5.2.1. BM3 FAD domain C773A mutant Preparation and Crystallisation.....</b>	<b>124</b>
<b>5.2.2. Data collection and processing .....</b>	<b>124</b>
<b>5.2.3. Structure Determination by Molecular Replacement .....</b>	<b>125</b>
<b>5.2.4. Model Building and Refinement .....</b>	<b>126</b>
<b>5.2.4.1. Building and Refinement of BM3 FAD domain.....</b>	<b>126</b>
<b>5.2.4.2. Building and Refinement of BM3 FAD domain NADP<sup>+</sup> complex.....</b>	<b>127</b>
<b>5.3. Analysis of the BM3 FAD domain C773A mutant and NADP<sup>+</sup> complex structures .....</b>	<b>127</b>
<b>5.3.1. BM3 FAD domain C773A mutant structure .....</b>	<b>127</b>
<b>5.3.2. BM3 FAD domain C773A mutant NADP<sup>+</sup> complex structure .....</b>	<b>128</b>
<b>5.4. Discussion .....</b>	<b>130</b>
<b>5.4.1. Overall structure .....</b>	<b>130</b>
<b>5.4.2. C773A mutation .....</b>	<b>132</b>
<b>5.4.3. Differences between molecules A and B in the asymmetric unit.....</b>	<b>135</b>
<b>5.4.4. FAD cofactor binding region.....</b>	<b>138</b>
<b>5.4.5. NAD(P)H binding region .....</b>	<b>141</b>
<b>5.4.6. Domain organisation.....</b>	<b>146</b>
<b>5.5. Conclusions and Future Work .....</b>	<b>152</b>

---

<b>Chapter 6 Crystal structure of the transcriptional regulator CprK ...</b>	<b>154</b>
<b>6.1. Introduction.....</b>	<b>155</b>
<b>6.2. Experimental Results .....</b>	<b>156</b>
<b>6.2.1. Purification of CprK .....</b>	<b>156</b>
<b>6.2.2. Crystal growth .....</b>	<b>157</b>
<b>6.2.2.1. CprK + OCPA.....</b>	<b>157</b>
<b>6.2.2.2. CprK.....</b>	<b>158</b>
<b>6.2.3. Heavy atom derivatisation.....</b>	<b>159</b>
<b>6.2.4. X-ray data collection and processing.....</b>	<b>159</b>
<b>6.2.4.1. CprK + OCPA.....</b>	<b>159</b>
<b>6.2.4.2. CprK crystals in absence of OCPA.....</b>	<b>161</b>
<b>6.2.5. Structure Solution.....</b>	<b>162</b>
<b>6.2.5.1. CprK+OCPA .....</b>	<b>162</b>
<b>6.2.5.2. Location and Refinement of Heavy Atom Sites.....</b>	<b>163</b>
<b>6.2.5.3. Map improvement .....</b>	<b>165</b>
<b>6.2.6. Building and refinement of CprK protein structure model.....</b>	<b>165</b>
<b>6.2.6.1. Interpretation of the Electron Density Map .....</b>	<b>165</b>
<b>6.2.6.2. Model Building and Refinement .....</b>	<b>167</b>
<b>6.2.7. Analysis of the CprK structure .....</b>	<b>167</b>
<b>6.2.7.1. Quality of the model.....</b>	<b>167</b>
<b>6.2.8. Binding studies .....</b>	<b>169</b>
<b>6.2.9. Mutagenesis of active site residues .....</b>	<b>171</b>
<b>6.3. Discussion .....</b>	<b>172</b>
<b>6.3.1. Overall fold and Secondary structure .....</b>	<b>172</b>
<b>6.3.2. Comparison with structurally identified homologues .....</b>	<b>175</b>
<b>6.3.3. OCPA binding by CprK .....</b>	<b>178</b>
<b>6.3.4. Binding studies .....</b>	<b>180</b>
<b>6.3.5. CprK DNA binding domain .....</b>	<b>182</b>
<b>6.4. Conclusions and Future Work .....</b>	<b>189</b>
<b>Bibliography.....</b>	<b>191</b>

---

---

## Abbreviations

Amino acids are abbreviated according to the three and single letter codes recommended by the I.U.P.A.C. Joint Commission on Biochemical Nomenclature (1985).

°	Degree
°C	Degrees celcius
αCTD	alpha-carboxy terminal domain of RNA polymerase
arac	arachidonate
BSβ	Cytochrome P450 from <i>Bacillus subtilis</i>
C-α	Carbon alpha
C-β	Carbon beta
cAMP	cyclic adenosine monophosphate
CPR	Cytochrome P450 reductase
CPR	chlorophenol reductase
CprK	Chlorophenol reductase protein K
CRP-FNR	cAMP receptor protein-Fumarate and nitrate reduction regulator
CYP, P450	Cytochrome P450 mono-oxygenase enzyme
Da	Dalton units
DP	degree of polymerization
DTT	Dithiotreitol
DMZ	4-methyl-N-methyl-N-(2-phenyl-2H-pyrazol-3-yl) benzenesulfonamide
EDTA	Ethylenediaminetetraacetic acid
ε	dielectric constant
e-	electron
EPR	Electron paramagnetic resonance
ESRF	European Synchrotron Radiation Facility

---

<b>Et</b>	<b>Protein concentration</b>
<b>ETF</b>	<b>Electron transferring flavoprotein</b>
<b>FAD</b>	<b>Flavin Adenine Dinucleotide</b>
<b>Fe</b>	<b>Iron</b>
<b>FMN</b>	<b>Flavin Mononucleotide</b>
<b>FNR</b>	<b>Ferredoxin-NADP<sup>+</sup> reductase</b>
<b>G</b>	<b>Gibbs energy</b>
<b>H</b>	<b>enthalpy</b>
<b>HPA</b>	<b>hydroxyphenylacetic acid</b>
<b>HTH</b>	<b>helix turn helix</b>
<b>IPTG</b>	<b>isopropyl-βD-thiogalactopyranoside</b>
<b>kb</b>	<b>kilobase</b>
<b>LB</b>	<b>Luria-Bertani</b>
<b>min</b>	<b>minute</b>
<b>mV</b>	<b>millivolt</b>
<b>MME</b>	<b>monomethylether</b>
<b>MCAD</b>	<b>medium chain Acyl CoA dehydrogenase</b>
<b>NADP<sup>+</sup></b>	<b>nicotinamide adenine dinucleotide phosphate, oxidized form</b>
<b>NADPH</b>	<b>nicotinamide adenine dinucleotide phosphate, reduced form</b>
<b>NHE</b>	<b>Na<sup>+</sup>/H<sup>+</sup> exchanger</b>
<b>Ni-NTA</b>	<b>Nickel nitrilotriacetic acid</b>
<b>NirI-NosR</b>	<b>Nitrite reductase inducer- Nitric oxide reductase Regulator</b>
<b>NOS</b>	<b>Nitric oxide synthase</b>
<b>PEG</b>	<b>Polyethylene glycol</b>
<b>UV</b>	<b>ultra violet</b>
<b>ω</b>	<b>omega</b>
<b>OCPA</b>	<b><i>ortho</i>-chlorophenolacetic acid</b>
<b>PAGE</b>	<b>Polyacrylamide gel electrophoresis</b>

---

---

palm	palmitoleate
P450 <sub>cam</sub>	Cytochrome P450 from <i>Pseudomonas putida</i> that acts on camphor
P450 <sub>terp</sub>	Cytochrome P450 enzyme that acts on alpha-terineol
P450 <sub>eryF</sub>	Cytochrome P450 enzyme that acts on deoxyerythronolide B
PCBs	polychlorinated biphenyls
PCE	polychlorinated tetrachloroethene
PCP	pentachlorophenol
pH	$\log 1/[H^+]$
pI	isoelectric point
pK <sub>a</sub>	negative logarithm of the association constant, $K_a$
q	charge
r	distance
rpm	revolutions per minute
S	entropy
SDS	sodium dodecyl sulphate
sec	seconds
T	temperature
WT	wild type

## Crystallography-Related Symbols and Abbreviations

Å	Ångström units
a, b, c, $\alpha$ , $\beta$ , $\gamma$	real space cell dimensions and angles
$\alpha$	phase angle
$\alpha_{\text{calc}}$	calculated phase angle
$\alpha_{\text{PH}}$ , $\alpha_{\text{P}}$ , $\alpha_{\text{H}}$	derivative, protein and heavy atom phase angles
$\alpha$ , $\beta$ , $\gamma$	Eulerian angles

---

<b>ADC</b>	analog-to-digital converter
<b>AMORE</b>	automated package for molecular replacement
<b>ARP</b>	Automated Refinement Procedure
<b>B</b>	isotropic temperature factor
<b>C</b>	rotation matrix
<b>CCD</b>	charge-coupled device
<b>CCP4</b>	Collaborative Computational Project No. 4
<b>DQE</b>	Detective quantum efficiency
<i>d</i>	spacing of the reflecting planes (resolution)
<i>E</i>	root-mean-square value of $\epsilon$
$\epsilon(\alpha)$	lack of closure
<i>f</i>	atomic scattering factor
$f_0$	Thomson scattering factor
$F_R, F_i$	real and imaginary anomalous scattering correction terms
$F$	structure factor
$F_o$ ( $F_{obs}$ ), $F_c$ ( $F_{calc}$ )	amplitudes of observed and calculated structure factors
$F_{PH}, F_P$	structure factors for derivative and native crystals
$F_H$	heavy atom structure factor
<b>FID</b>	mean fractional isomorphous difference
<b>FOM</b>	figure of merit
$\varphi, \psi,$	phi and psi torsion angles of protein backbone
<b>h</b>	reciprocal space vector
$h, k, l$	reciprocal lattice indices
$I_o, I_c$	observed and calculated intensity
<b>ITC</b>	isothermal titration calorimetry
<b>K</b>	scale factor
<b>K<sub>D</sub></b>	equilibrium constant for dissociation
<b>L</b>	likelihood

---

---

LS	least squares
$\lambda$	wavelength
$m$	figure of merit
MAD	multiple wavelength anomalous dispersion
MIRAS	multiple isomorphous replacement using anomalous scattering
ML	maximum likelihood
$N_o, N_i$	Noise output and input
NCS	non-crystallographic symmetry
$N_{symm}$	number of equivalent positions in the unit cell
OCPA	<i>ortho</i> -chlorophenolacetic acid
PA	phenylacetic acid
$P(\alpha)$	probability distribution
$P(A;B)$	conditional probability distribution
$P(\mathbf{u})$	Patterson function
PDB	Protein Data Bank
$\mathbf{r}$	position vector in real space
R	crystallographic agreement factor
RF, R(C)	rotation function
RMSD	root mean square deviation
$\rho$	electron density
$S_o, S_i$	signal output and signal input
$\mathbf{S}$	scattering vector
SC	surface complementarity
$T(\mathbf{x}), T_1(\mathbf{x})$	translation function
$t_x, t_y, t_z$	fractional translation parameters
$\theta$	Bragg angle
$u$	amplitude of vibration of an atom
$\mathbf{u}$	position vector in Patterson space
$u, v, w$	fractional coordinates of a heavy atom

---

---

V	volume of a crystal
x, y, z	real space coordinates
Z-score	strength of structural similarity in standard deviations above expected

# **Chapter 1**

## **Introduction**

## 1.1. Introduction

X-ray crystallography of macromolecules has generated unprecedented insights into the detailed molecular understanding of biology over the last 50 years. This includes atomic level insights into protein-DNA interactions, allosteric changes involved in enzyme mechanism and more recently design and optimisation of lead compounds for pharmaceutical development. Ligand binding can induce significant structural changes in proteins and both co-operative and non-cooperative binding can be observed for multiple binding sites. The induced fit hypothesis of Koshland (1958) proposes that proteins have multi-conformational states that are in equilibrium with each other and that this equilibrium is easily affected by ligand binding (Tsou, 1998). More subtle “nanoallostery” has also been proposed to occur in monomeric proteins upon ligand binding with recent studies utilising residue mutations in order to create novel equilibrium states (Youssef *et al.*, 2004).

Analysis of the structural changes brought about by ligand binding can be of critical importance in gaining a full understanding of protein function and specificity. This thesis describes structural studies that were carried out on two distinct proteins that bind small molecules. These are cytochrome P450 BM3 from *Bacillus megaterium* involved in fatty acid hydroxylation and the transcriptional regulator CprK that regulates halorespiration in the bacterium *Desulfobacterium hafniense*.

## 1.2. Principles of ligand binding and protein motion

Several studies have shown that proteins are highly dynamic molecules that exist in an equilibrium of multiple states (Hammes, 1964) with equilibrium changes induced by a range of stimulants such as substrate or inhibitor binding (Bruice and Pandit, 1960). The structure determination of these proteins in a ligand-bound and ligand-free form has allowed visualisation of structural changes that can occur upon ligand binding (Ostermann *et al.*, 2000; Joseph *et al.*, 1990; Hansen *et al.*, 2005).

### 1.2.1. Entropic and Enthalpic factors

The laws of thermodynamics state that energy is conserved within a system and this system tends towards greater entropy. The entropy of an isolated system can decrease

concomitant with an increase in the entropy of its surroundings. Gibb's free energy is a measure of chemical energy and all systems tend towards a minimum Gibb's free energy at a constant pressure and temperature. Calculation of the free energy of a system can indicate if an event such as protein-ligand binding is likely to occur i.e. it is thermodynamically favourable and will describe the equilibrium the system will go towards.

$$G = H - TS$$

**Equation 1. 1**

where G is the Gibb's free energy; H is the enthalpy or heat constant; T is temperature and S is entropy.

The molecular interpretation of entropy (S) is a measure of disorder and with increased disorder, the greater the entropic value. A number of factors influence the entropy of a system, e.g. when the temperature is increased within a system there is an increase in entropy due to an increase in molecular motions. Enthalpy (H) describes the amount of heat or energy absorbed by a system. Whereas enthalpy is largely dependent upon electronic internal energies, entropy values are dependent upon translational, vibrational and rotational internal energies. Proteins can exist as an equilibrium of distinct conformations which  $\Delta G$  for conversion between these near or equal to zero.

The binding of a small molecule may lead to significant changes in this equilibrium thus leading to a change in the preferred protein conformation. The association of two molecules leads to a loss in entropy for the molecules involved due to the new interactions that are created. However, there is an increase in overall entropy within the system due to the release of hydrogen bonded water molecules (Fersht, 1985). While entropy changes influence many of the interactions between proteins and ligands there is also a contribution from enthalpic effects which influences ligand binding. In the case of ligand binding, where initial unfavourable protein interactions such as repulsive charged residues are replaced with a more favourable protein-ligand interaction, the enthalpic effect is significant. The enthalpic effects due to the interaction of protein and ligand, structural changes caused by ligand binding and any protonation/deprotonation events that occur must all be analysed (Luque and Friere, 2002). Enthalpic effects thus may in fact contribute significantly to the thermodynamics of ligand binding and optimisation of these effects may lead to tighter binding of ligands (Christensen *et al.*, 2003).

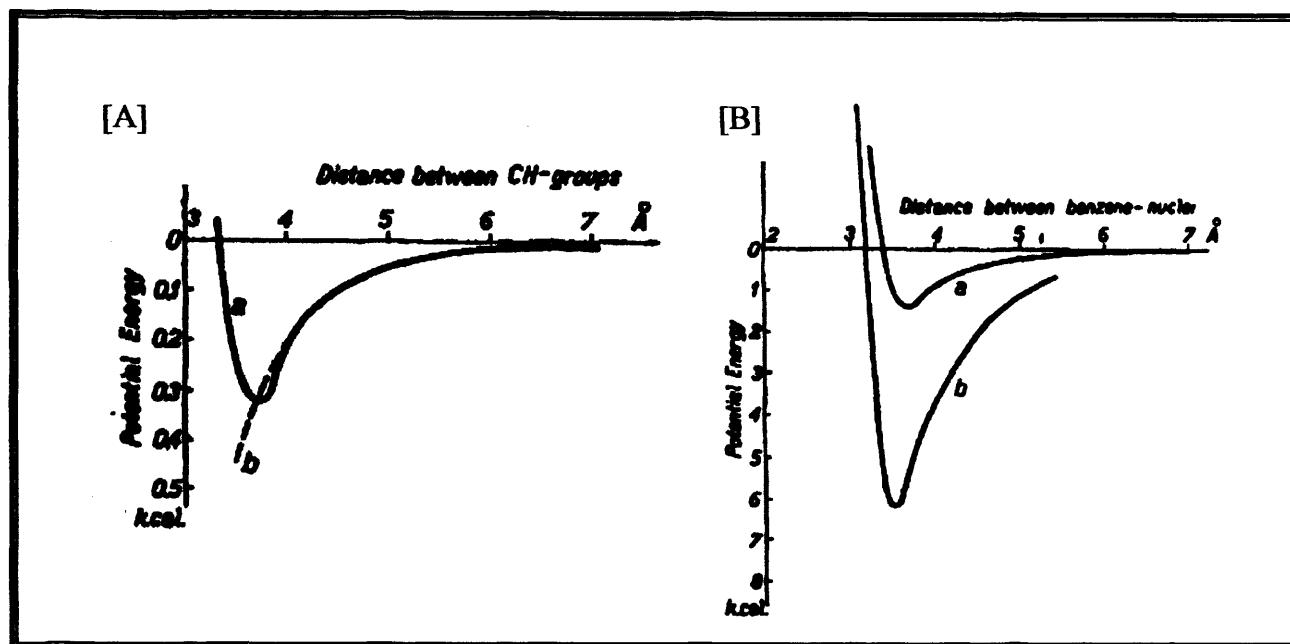
### 1.2.2. Ligand Binding Specificity: Non-covalent interactions

Ligand binding depends on direct protein interactions with high specificity derived from the formation of multiple non-covalent interactions. Ligand differentiation is critical in ensuring optimal activity within an ordered biological system. The major non-covalent forces and interactions are van der Waals interactions, hydrogen bonds, electrostatic interactions and hydrophobic interactions. These interactions are dictated by changes in entropic and enthalpic effects determining binding affinity.

#### 1.2.2.1. London-van der Waals interactions

Van der Waals interactions are a combination of short-range attractive (London dispersion) forces caused by the generation of a temporary dipole between atoms and short-range van der Waals repulsive forces due to electrostatic repulsion between occupied orbitals of non-bonded atoms that are very close together in space (Tinocco *et al.*, 2003). The temporary dipole in London dispersion forces induces the electron distribution of an adjacent atom to polarize in order to minimize electron-electron repulsion between the two atoms. As the distance decreases below 3 Å the energy of the system increases sharply, thus it is reasonable to think of atoms as hard spheres defined by their van der Waals radii. The van der Waals radius of an atom is defined where the net force between two atoms is zero and the van der Waals potential is at its minimum (De Boer, 1936, Figure 1.1).

The van der Waals forces are weak but very important for molecular recognition phenomena. The van der Waals bond energy of a pair of atoms is about 1 kcal/mol which is only a little more than the average thermal energy of molecules at room temperature (0.3 kcal/mol). However, these forces are additive and the contribution of the van der Waals interactions to the binding energy becomes significant when numerous atoms in one molecule simultaneously bind to many atoms of a second molecule.



**Figure 1.1** [A] Potential energy of two aromatic CH groups as a function of the distance between the groups (curve a). Curve b gives the course of the van der Waals attractive energy. [B] Potential energy of two benzene molecules that lie on parallel planes which vary depending on the distance between the molecules (curve b). Curve a represents the reciprocal energy of the six pairs of CH groups which lie opposite each other (Taken from De Boer, 1936)

### 1.2.2.2. Hydrogen bonding

The hydrogen bond is a special case of molecular dipole force which forms a favourable interaction between a hydrogen atom which has a slight electropositive charge bonding to electronegative atoms such as oxygen, nitrogen and sulphur. The acidity of the proton bound to the donor and the basicity of the acceptor both correlate roughly with the strength of the hydrogen bond.

The strength of hydrogen bonds between uncharged groups is 1.0–1.4 kcal/mol, whereas that involving a single charged group is slightly higher at 1.5–2.8 kcal/mol. The energy of a hydrogen bond between charged donor and charged acceptor is about 4 kcal/mol (Jeffrey and Saenger 1991, Fersht *et al.*, 1985). More recent estimates of the energetic contributions of individual hydrogen bonds in proteins confirm these values and are generally in the range 2–5 kcal/mol (Perrin & Nielson, 1997). An important feature of the hydrogen bond is its conformational constraint. The optimal alignment of donor-H to acceptor is a head to tail linear arrangement and deviation from linearity causes energy loss (Nelson & Cox, 2000).

### 1.2.2.3. Electrostatic interactions

The force of an electrostatic interaction between charged atoms is given by Coulomb's law:

$$F = \frac{q_1 \cdot q_2}{\epsilon r^2}$$

#### Equation 1. 2

in which  $q_1$  and  $q_2$  are the charges of the two atoms,  $r$  is the distance between them and  $\epsilon$  is the dielectric constant of the medium. The effect of the dielectric constant  $\epsilon$  on electrostatic interactions is the most important aspect of Coulomb's law for biological systems. In the interior of a protein, the dielectric microenvironment is variable, with less shielding of charges in regions of hydrocarbon side chains and greater shielding in regions of polar side chains (Bashford & Karplus, 1991). The dielectric constant is a macroscopic property therefore it is difficult to apply to a system of very small volume, such as the interior of the protein. In macromolecular modelling, a protein dielectric constant of 4.0 is commonly used because this value gives the best agreement between calculated and experimental pK<sub>a</sub>'s in proteins (Lancaster *et al.*, 1996).

A positively charged lysine or arginine residue can form a strong interaction with a negatively charged aspartate or glutamate group known as a salt bridge and inorganic metal ions may play a crucial role in the mechanism of a protein or as a stabilising structural element. However, these charge interactions are highly sensitive to changes in pH.

### 1.2.2.4. Hydrophobic interactions

The term hydrophobic interaction describes the tendency of non-polar compounds to transfer from an aqueous solution to an organic phase. This is not due to the repulsion of these non-polar entities to water but is due to the much higher attraction that water exhibits to itself as opposed to forming a water-non polar interaction. Classic theory describes the hydrophobic interaction as a mainly entropic phenomenon, arising not from intrinsic attraction between non-polar moieties, but from the system achieving the lowest entropic value by minimizing the number of ordered water molecules required to surround hydrophobic regions of the

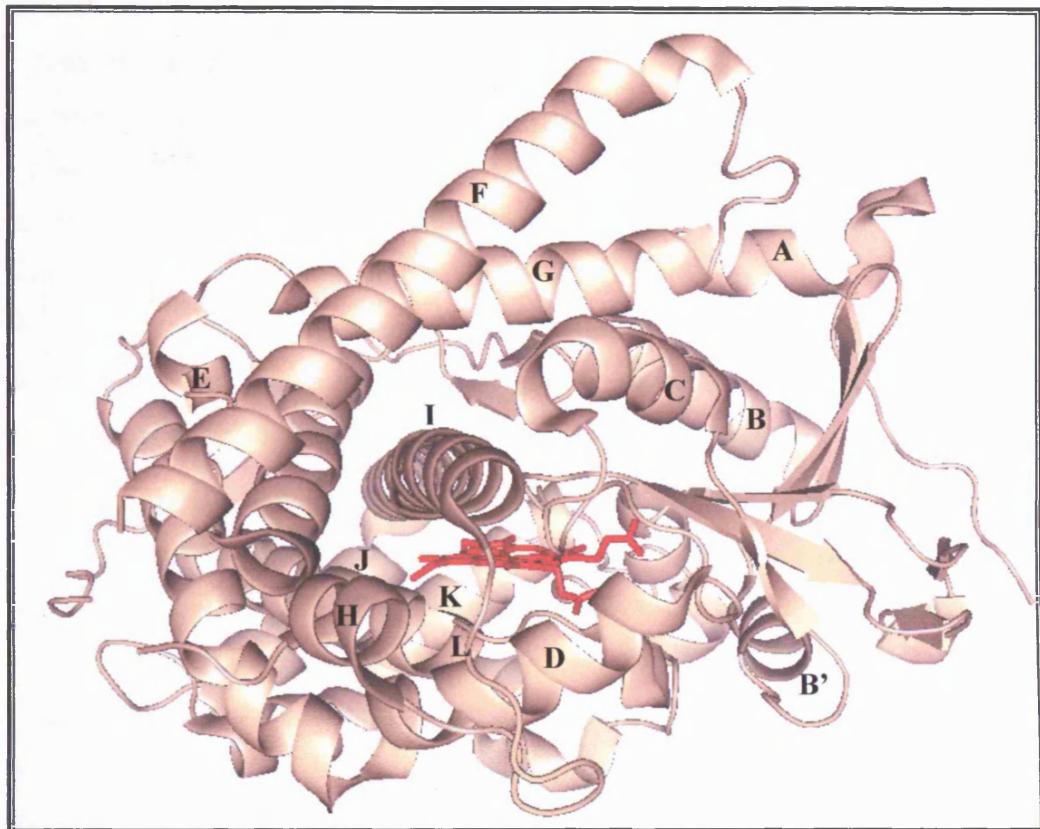
molecule (Tinocco *et al.*, 2002). The measurement of the hydrophobic interaction between dissolved non-polar molecules is quite difficult to carry out mainly due to problems with solubility. Tucker *et al.* (1981) reported values of -8.4 and -11.3 kJ/mol (-2.0 and -2.7 kcal/mol) for the free energies of dimerisation of benzene-benzene and cyclohexane-cyclohexane, respectively, and Ben-Naim *et al.* (1973) deduced a value of about -8.5 kJ/mol (-2.0 kcal/mol) for two methane molecules. The evaluation of hydrophobic interaction energies in proteins is difficult to calculate. This is because there are fundamental differences between the hydrophobic interactions of an organic solvent and water compared to the binding of hydrophobic substrates to a protein. The transfer of a solute from the aqueous to the organic phase may be divided into three hypothetical steps: formation of a cavity in the organic phase; transfer of the solute to the cavity; and closing of the cavity left in the aqueous phase. The transfer of a hydrophobic substrate to a hydrophobic region in an enzyme involves the occupation of a pre-formed cavity, and probably the removal of water from this cavity (Muller *et al.*, 1990).

### 1.3. Cytochromes P450

The cytochromes P450 family is made up of a vast group of heme thiolate enzymes which activates molecular dioxygen into a highly reactive oxygen species and a subsequent mono-oxygenation reaction with a huge array of substrates. Cytochromes P450 have been one of the most studied enzyme systems in the world (Lewis, 2001). There are over 1200 distinct P450 enzymes revealed through biochemical and genomic studies to date (Nelson, 1999). Cytochromes P450 are found in all five biological kingdoms thus partly explaining their wide range of activities and their importance. They have a conserved trigonal structure that fold around a *b* heme group as shown in Figure 1.2.

They have been shown to have important implications in research areas ranging from the study of genetic polymorphisms involved in the conversion of herbicides and insecticides to the biotechnological development of biosensors (Lewis, 2001). Cytochromes P450 commonly act on a wide range of compounds including steroids, fatty acids and xenobiotics such as pharmaceutical drugs and environmental chemicals. Furthermore, in eukaryotes, these enzymes play critical roles in the synthesis and inter-conversion of steroids, while in mammals hepatic cytochromes P450 are vital for the detoxification of many drugs (Guengerich and Shimada, 1991). A detailed understanding of these enzymes has been a slow

and rigorous process with difficulties arising in comparisons between different P450s due to the low sequence identities between P450 families.



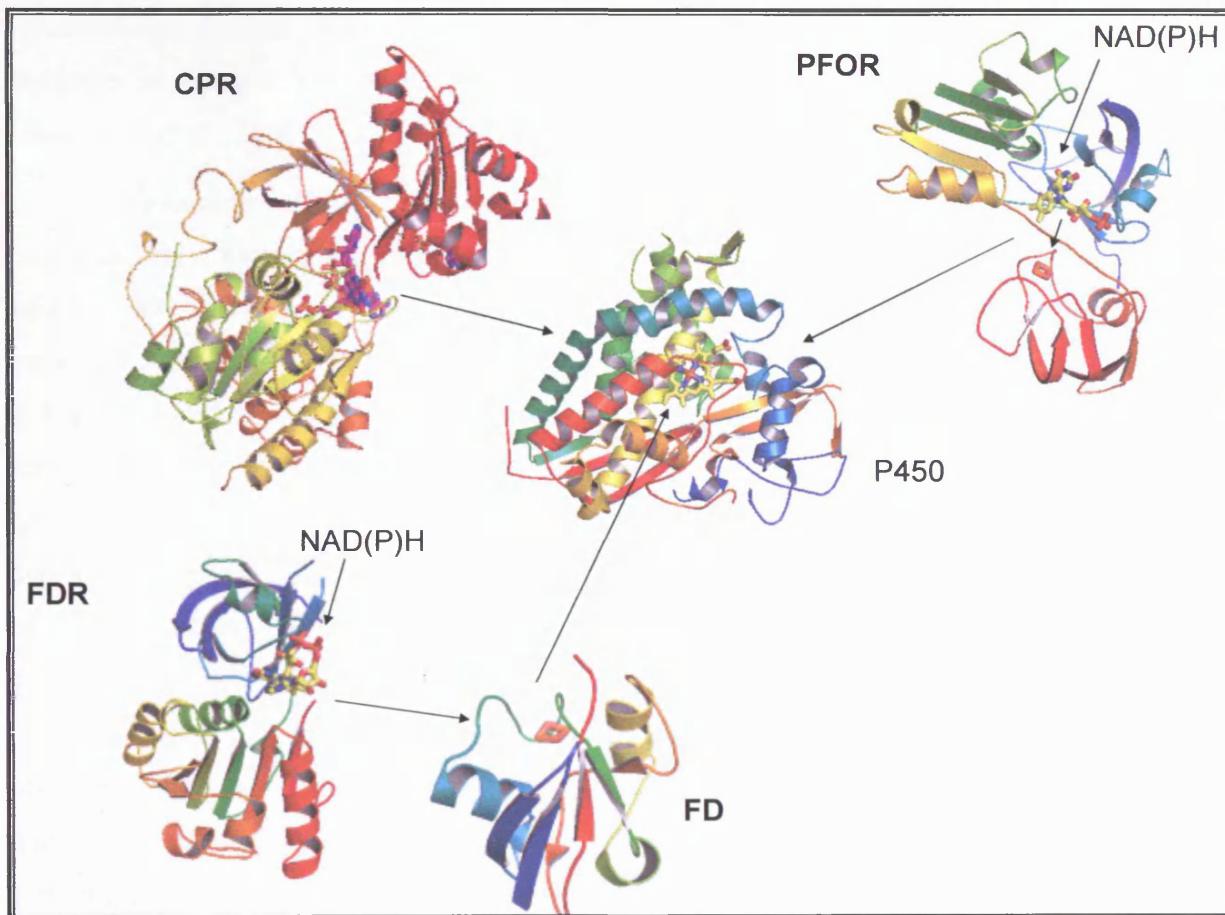
**Figure 1.2** Cartoon representation of cytochrome P450 BM3 heme domain (PDB code: 1JPZ) with the *b* heme (red) shown in stick representation. Each helix is labelled with a single letter indicating its position within the primary protein sequence.

Cytochromes P450 are of particular interest to the pharmaceutical industry due to their importance in drug breakdown and possible conversion to toxic chemicals that can cause serious side-effects (Ingelman-Sundberg, 2002) and even death (Gonzalez, 2005). This conversion of pharmaceuticals leads to a large number of drugs being rejected prior to clinical trials (Williams *et al.*, 2003). The recent structural identification of mammalian cytochromes P450 may lead to a greater insight in the design of modern pharmaceuticals leading to lower drug dosages and reduced toxicity (Congreve *et al.*, 2005). There is also significant opportunity for the biotechnological use of these enzymes for *in-vitro* drug metabolic studies, biodegradation and bio-remediation. Optimisation of catalytic activity and substrate specificity can be carried out by site-directed mutagenesis and some success has been reported to date (Glieder *et al.*, 2002).

Due to the large number of cytochrome P450s, a system was developed to gain order in the nomenclature of the multiple P450 families. A number is given to each group of P450s based on their function or the species that they originate from and this system is described in Nelson *et al.* (1996).

Initially, the structural insights into P450 enzymes was obtained from the structure of P450<sub>cam</sub> (Poulos *et al.*, 1987) and a number of other bacterial structures determined in the 1990s (P450<sub>nor</sub>, Park *et al.*, 1997; P450-BM3, Ravichandran *et al.*, 1993; P450<sub>eryF</sub>, Cupp-Vickery and Poulos, 1995; P450<sub>terp</sub>, Hasemann *et al.*, 1994). Since then, there has been an explosion in the number of structures determined and their biological sources with 18 bacterial P450 structures and 6 mammalian structures presently deposited in the PDB. Ultimately, these and subsequent structures will allow us to understand the complex mechanisms and parameters involved in cytochrome P450 biology.

Most bacterial, and all of the eukaryotic mitochondrial P450 systems, are three component systems composed of an NADH-dependent FAD containing ferredoxin reductase and a ferredoxin (Figure 1.3). Eukaryotic microsomal P450 systems (those associated with the endoplasmic reticulum) are two component systems made up a NADPH-dependent diflavin (FAD and FMN containing) reductase. A third class of which the representative system is the bacterial P450 BM3 (CYP 102A1), is a one component system similar to class II but where the two components are covalently linked into a single polypeptide (Miles *et al.*, 2000). BM3 has been extensively studied because of its high levels of recombinant soluble expression and its similarity to the eukaryotic P450 systems. Further classes of the cytochrome P450 systems have been identified with a new system recently identified in *Rhodococcus* sp. which is made up of a cytochrome P450 covalently attached to a phthalate family oxygenase reductase (PFOR) module (containing an FMN and 2Fe-2S coenzyme groups; De Mot and Parret, 2002).



**Figure 1.3** Electron transfer systems in the Cytochrome P450 family. Representative structures shown are the *Pseudomonas putida* P450 (camphor hydroxylase CYP101; PDB code 2CPP); rat CPR (PDB code 1AMO); *Escherichia coli* FDR (PDB code 1FDR); *P. putida* FD (putidaredoxin; PDB code 1PDX); and *Burkholderia cepacia* PFOR (phthalate dioxygenase reductase; PDB code 2PIA). The protein structures are shown in cartoon format with the bound cofactors shown in stick representation.

### 1.3.1. Cytochrome P450 catalytic cycle

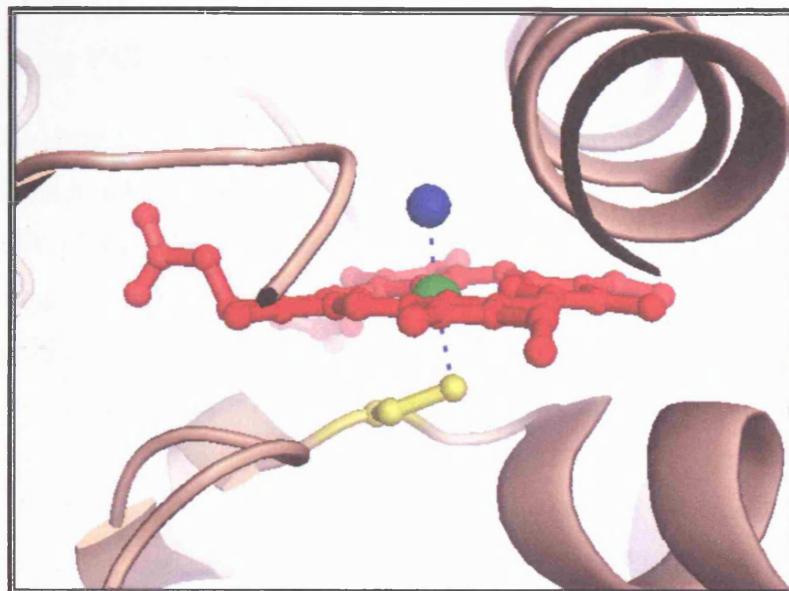
An outline of the common catalytic cycle in P450s was outlined as early as 1968 (Gunsalez *et al.*, 1975). Examination of the catalytic cycle of cytochrome P450s has been carried out by structural, spectroscopic, mechanistic, computational and chemical modelling techniques resulting in a detailed understanding of the process but there are still factors in the mechanism that are under investigation.

The cytochrome P450 reaction mechanism utilises an iron atom ligated to a protoporphyrin IX macrocycle with two additional axial ligands shown in Figure 1.4. The proximal ligand is a conserved cysteine thiolate group and the distal ligand is a variable ligand that changes during the catalytic cycle. The proximal thiolate ligand is a strong electron donor and this effect is thought to have a significant effect in the catalytic cycle. The recent production of a synthetic stable iron atom with a thiolate ligand in a core

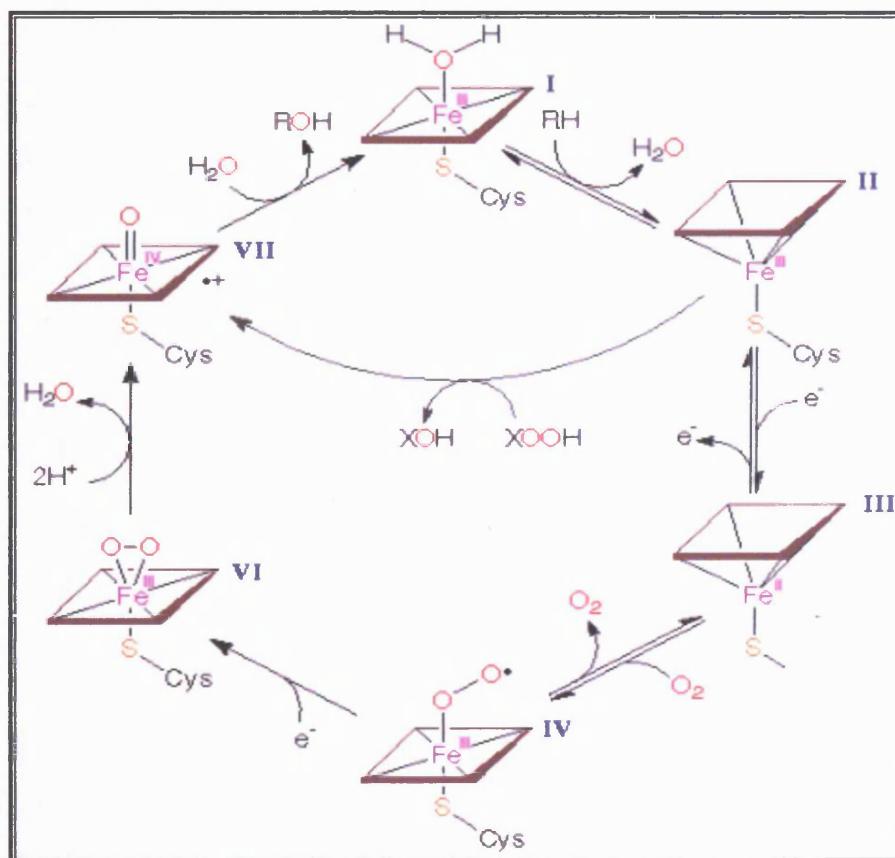
cyclotetradecane unit allowed reactivity studies to be carried out. These indicate that hydrogen abstraction is preferred over oxygen transfer in the presence of reducing substrates (Bukowski *et al.*, 2005).

Cytochromes P450 carry out a mono-oxygenation reaction producing a water molecule and a product with an incorporated oxygen atom. Besides the substrate molecule and oxygen, the reaction mechanism requires two protons and two electrons which are provided by NAD(P)H in a two step process using either one or two redox partners (either an iron-sulphur redoxin, a flavoprotein or cytochrome  $b_5$  depending on the cytochrome P450 involved). However, there is a complex set of components involved in this unique chemistry of cytochrome P450s. There is a rich variation in the chemistry afforded by the heme-oxygen centre including nucleophilic, electrophilic and hydrogen abstraction mechanisms. A diagram of the catalytic cycle is presented in Figure 1.5 below. The resting state (I) of the enzyme is the ferric ( $\text{Fe}^{\text{III}}$ ) complex with a water molecule as a weak distal ligand. The introduction of a substrate displaces the water molecule causing a penta-coordinated ferric species (II). This leads to displacement of the iron atom from the plane of the porphyrin ring concomitant with a shift in the position of the d-orbital electrons from low spin to a high spin state. This causes the complex to become a better electron acceptor due to the concomitant change in the redox potential of the iron and electron transfer is initiated from a specific redox partner causing the reduction of the iron to the ferrous form ( $\text{Fe}^{\text{II}}$ ; III). The high spin ferrous form binds dioxygen forming a low spin semi-stable hexa-coordinate ferrous-O<sub>2</sub> adduct (IV). A second electron is then transferred from the redox partner leading to a ferric-dioxy species (VI). This species undergoes two protonation events and the release of a water molecule leading to an oxy-ferryl species ( $\text{Fe}^{\text{IV}}=\text{O}$ ; VII). This is the reactive species which transfers the distal oxygen atom to the substrate. The oxygenated substrate is released and replaced by a water molecule returning the P450 to the resting state. An important factor thought to contribute to the efficiency of the catalytic cycle is the hydrogen bonding network with the thiolate ligand on the proximal side of the heme (Yoshioka *et al.*, 2002).

It is proposed that a fundamental part of the P450 reaction is the modulation of the spin state coupled to the iron redox potential which controls the catalytic rate of substrate metabolism (Blanck *et al.*, 1983). However, the situation in the microsomal systems is not as simple with some P450s appearing to be in the high spin form permanently. These systems may thus not require spin and redox equilibrium changes to initiate catalytic activity (Reed and Hollenberg, 2003).



**Figure 1.4** Iron (green sphere) co-ordination within the cytochrome P450 heme *b* cofactor with a distal water molecule shown in blue sphere representation and a proximal cysteine shown in yellow stick representation. The protoporphyrin IX ring is shown in red ball and stick representation. The protein molecule is shown in brown cartoon format.



**Figure 1.5** The Cytochrome P450 catalytic cycle. Substrate is shown as RH with the spin state of the Fe indicated by the raised and flat rectangle. (Taken from Gustaffson, 2000)

### 1.3.2. Cytochrome P450 structure

Using the large number of structures identified to date, it is possible to state that the overall P450 fold is much conserved and also unique to the P450 family (Poulos and Johnson, 2005). This fold allows for the specific heme-thiolate chemistry involved in oxygen activation, binding of redox partners and the stereochemical requirements involved in substrate recognition. Differentiation in enzyme specificity is dictated by the precise positioning of various structural elements. However, structural differences are less evident in the environment of the heme group especially helices I and L which directly interact with the protoporphyrin IX group. The B' helix is involved in substrate determination and this is where significant differences are observed. In the case of P450<sub>eryF</sub> the B' helix is oriented approximately 90° from the orientation observed for the B' helix in P450<sub>cam</sub> resulting in substantial differences in the local environment. The most conserved element between different P450s is the β bulge segment housing the heme ligating cysteine (Poulos and Johnson, 2005).

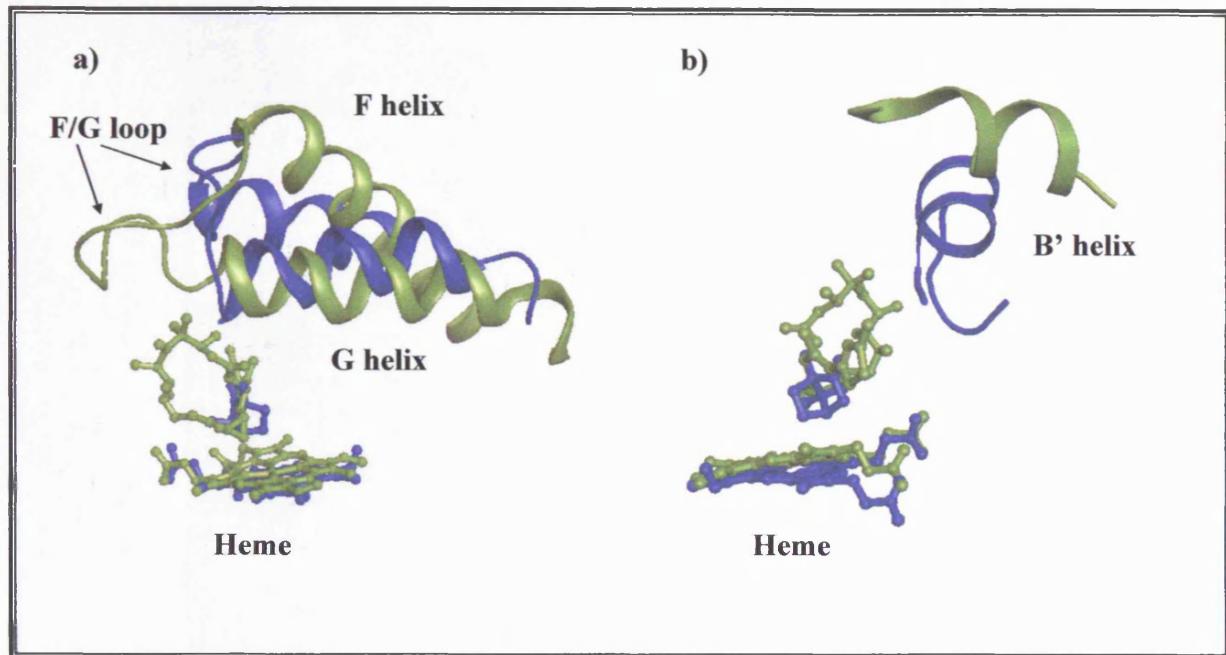
The importance of the cysteine proximal ligand to the heme group was emphasised by the C357H mutation in P450<sub>cam</sub> which reduced activity to almost undetectable levels. The cysteine is also involved in hydrogen bonding with peptide NH groups and these interactions regulate the heme-iron redox potential. In their absence the redox potential would be too low for reduction by redox partners (Ueyama *et al.*, 1996). The second highly conserved region is a portion of the I helix where a small distortion is observed and this region is involved in oxygen activation. An acid-alcohol side chain pair is observed in the majority of P450 active sites. These are typically a threonine or serine as the alcohol side chain and glutamate or aspartate as the conserved acid group (Nebert *et al.*, 1989; 1991). In cytochrome P450<sub>eryF</sub> this alcohol group is missing from the protein and is instead provided by the substrate making deoxyerythronolide B, the only acceptable substrate for this P450. Introduction of an A254T mutation into the protein allows the protein to act on a range of novel substrates (Xiang *et al.*, 2000). Multiple site directed mutagenesis studies in P450<sub>cam</sub> have suggested a role for these residues in stabilising the distal substrate binding pocket, H-bond networks and contributing to proton delivery during enzyme reaction. Ordered solvent has been proposed to act as the direct proton donor to the iron linked dioxygen. However, the details of proton transfer may differ between one P450 and the next, furthermore the positioning of solvent in the active site by protein and substrate may also be critical in relaying protons critical for dioxygen bond cleavage.

### 1.3.3. P450-Substrate complexes

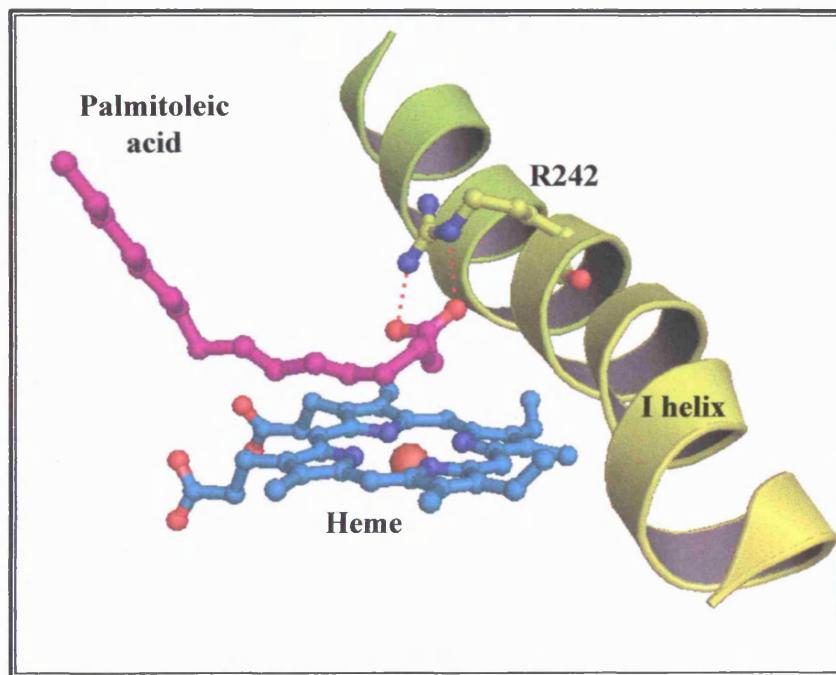
To date, the structure of five prokaryotic P450-substrate complexes have been determined (Cupp-Vickery and Poulos, 1995; Poulos *et al.*, 1987; Li and Poulos, 1997; Nagano *et al.*, 2003; Lee *et al.*, 2003). Within this group of P450 substrate complexes is a diverse range of substrates that differ substantially both in size and composition. From this group it is however possible to gain some understanding of the parameters that are important in defining substrate binding. The regio and stereo specificity in P450 binding is achieved by substrate interaction with active site protein residues locating the substrate in the correct position. All substrates are located with the atom to be oxygenated approximately 4-5 Å from the heme iron. The exception to this is cytochrome P450 BM3 in which the P450-substrate complex places the substrate 7-8 Å from the heme iron (Li and Poulos, 1997; Haines *et al.*, 2001). However, NMR results indicate the substrate relocates to within 3 Å of the iron upon reduction from Fe<sup>III</sup> to the Fe<sup>II</sup> form indicating the importance redox regulation plays in the catalytic mechanism (Modi *et al.*, 1996).

P450<sub>cam</sub> and P450<sub>eryF</sub> are representative of the possible differences in substrate size and shape that are seen within the P450 family. Structural alignment of these structures indicates that the F, G, B' helices and the F/G loop differ the most between the two structures with the B' helix rotated ~90° in P450<sub>eryF</sub> opening the substrate binding site therefore allowing the large thiazole ring to enter (Figure 1.6). In the case of P450 BSβ an arginine residue located on the I helix electrostatically stabilises the carboxyl head of the palmitic acid substrate demonstrating the adaptability of the P450 family to specific substrates (Lee *et al.*, 2003; Figure 1.7).

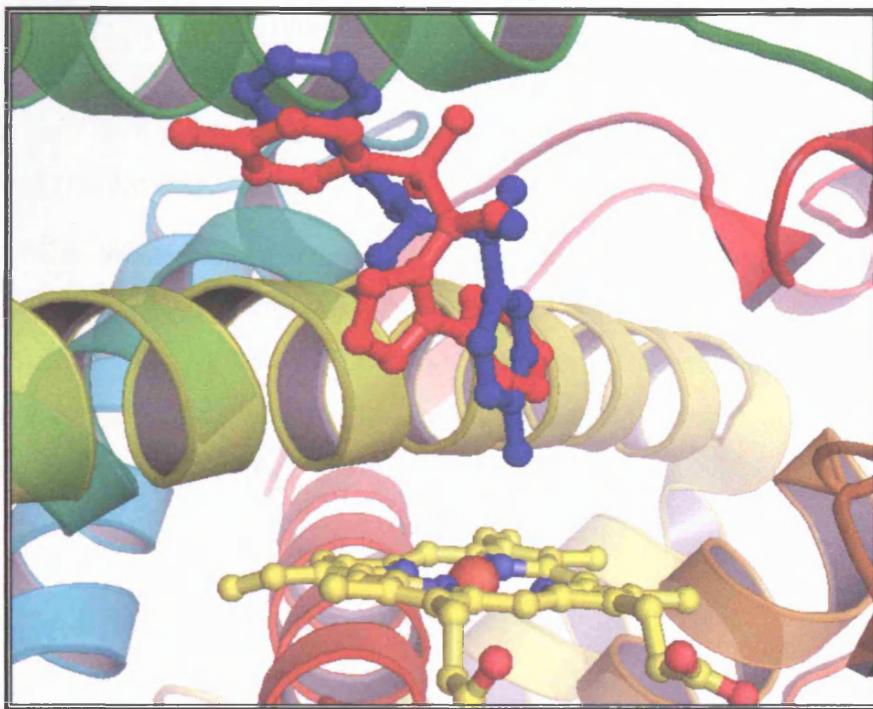
CYP119 has been determined in the presence of phenyl-imidazole and imidazole and in the case of imidazole the F helix unwinds causing the active site to change to an optimal configuration for ligand binding (Yano *et al.*, 2000). In the case of CYP2C5, numerable different substrates can bind with large differences in their size, flexibility and polarity by structural adaptations upon substrate binding (Wester *et al.*, 2003a; b). The major structural differences exist in the region of the B' helix, F/G helices and the loop between the F/G helices. Furthermore multiple sites of oxidation are frequently seen for substrates of the mammalian drug metabolising enzymes (Figure 1.8). This substrate promiscuity may be due to the large dimensions of the active sites leading to activities against multiple substrates (Yano *et al.*, 2005).



**Figure 1.6** Structural changes brought about due to substrate size. Cytochrome P450<sub>epok</sub> is shown in green cartoon representation and P450<sub>cam</sub> is shown in blue cartoon representation overlaid with an RMSD of 2.951 Å. The heme groups and the substrates camphor (blue) and epitholine (green) are shown in stick representation. a) F and G helices and F/G loop. b) B' helices.



**Figure 1.7** Substrate binding in cytochrome P450 BS $\beta$  (PDB code: 1IZO). The heme molecule, palmitoleic acid substrate and residue R242 are shown in ball and stick format while the I helix is shown in cartoon representation.



**Figure 1. 8** Multiple binding of 4-methyl-N-methyl-N-(2-phenyl-2H-pyrazol-3-yl) benzenesulfonamide (DMZ) to cytochrome P450 2C5 (PDB code: 1N6B). Protein is shown in cartoon representation with the heme shown in ball and stick format with yellow coloured carbons and alternate DMZ molecules shown in ball and stick format coloured red and blue.

In the case of P450<sub>cam</sub> and P450<sub>epok</sub> there is no significant structural difference between the substrate bound and substrate free forms (Poulos *et al.*, 1985, 1987; Nagano *et al.*, 2003) However, in P450<sub>cam</sub> the B', F and G helices show higher thermal motion in the substrate free structure indicating the possibility of an unobserved “open” and transient conformation. The first clear structural changes seen upon substrate binding were observed for the heme domain of cytochrome P450 BM3 with the F/G helices sliding over the surface of the I helix (Li and Poulos, 1997). This motion closes off the entry channel indicating the substrate enters near the F/G loop region. In the structures of OxyB (Zerbe *et al.*, 2002) and CYP154C1 (Podust *et al.*, 2003) the enzymes are in a similar open conformation to cytochrome P450 BM3 with the F/G helices rotated out of the active site of the structures. The region between helices B and C separates from the G helix to form a cleft that forms along helix I. In the case of CYP51 there is an unprecedented break in the I helix creating a new substrate entry point that runs roughly parallel to the heme and is indicative that multiple substrate entry points may be possible (Podust *et al.*, 2001). In contrast to the prokaryotic P450s eukaryotic P450s have a longer loop region between the F and G helices that interacts more closely with the  $\beta$ -sheet region of the protein and this region is also thought to interact with the membrane (Li and Poulos, 2004). The B' helix is however less closely associated with the protein and substrate entry is thought to occur between the B' helix, G helix and the

B'/C loop. This substrate entry pathway is supported by the crystal structure of CYP2B4 in an ‘open’ conformation (Scott *et al.*, 2003) and the fact that progesterone can be soaked into CYP2C5 crystals even though there are severe constraints on the F and G helices due to crystal packing (Poulos and Johnson, 2005).

The P450 substrate complexes describe some of the structural changes that are required in substrate binding within the P450 family and that multiple binding modes may exist within a single P450 enzyme. These parameters are still being interpreted and further biochemical and structural information will only lead to a better understanding of the diverse components that are critical in the P450 enzyme family.

#### 1.3.4. Electron Transfer Partners of Cytochrome P450s

A supply of electrons is required for the P450 mono-oxygenation reaction. These are derived from NAD(P)H (Figure 1.9) and supplied via a redox partner. Cytochromes P450 redox systems are made up of essentially two major classes, (i) bacterial or mitochondrial and (ii) microsomal. The majority of bacterial and mitochondrial cytochromes P450 utilise a dual component system made up of a ferredoxin and a ferredoxin reductase as redox partners. The microsomal P450s utilise a single NAD(P)H dependant P450 reductase (CPR). CPR is the prototype of the diflavin reductase family made up of FAD and FMN containing electron transfer domains (Paine *et al.*, 2000). There are three other mammalian diflavin reductases Methionine Synthase Reductase (Wolthers *et al.*, 2004), human oxidoreductase NR1 (Finn *et al.*, 2003) and the nitric oxide synthases (Stuehr, 1999). To date, 21 CPR genes have been identified from genome sequencing initiatives which in general lead to the identification of only a single *cpr* gene within an organism. The exception is in plant genomes where multiple *cpr* genes are observed (Benveniste *et al.*, 1991, Koopman and Hahlbrock, 1997; [http://www.icgeb.org/~p450srv/P450\\_Reductases.html](http://www.icgeb.org/~p450srv/P450_Reductases.html)). The CPR structure is made up of three domains, an FMN domain, a NAD(P)H/FAD domain and a linker domain. The linker domain sequence is important in interflavin orientations as shown by mutagenesis studies (Hubbard *et al.*, 2001).

## 1.4. Cytochrome P450 BM3

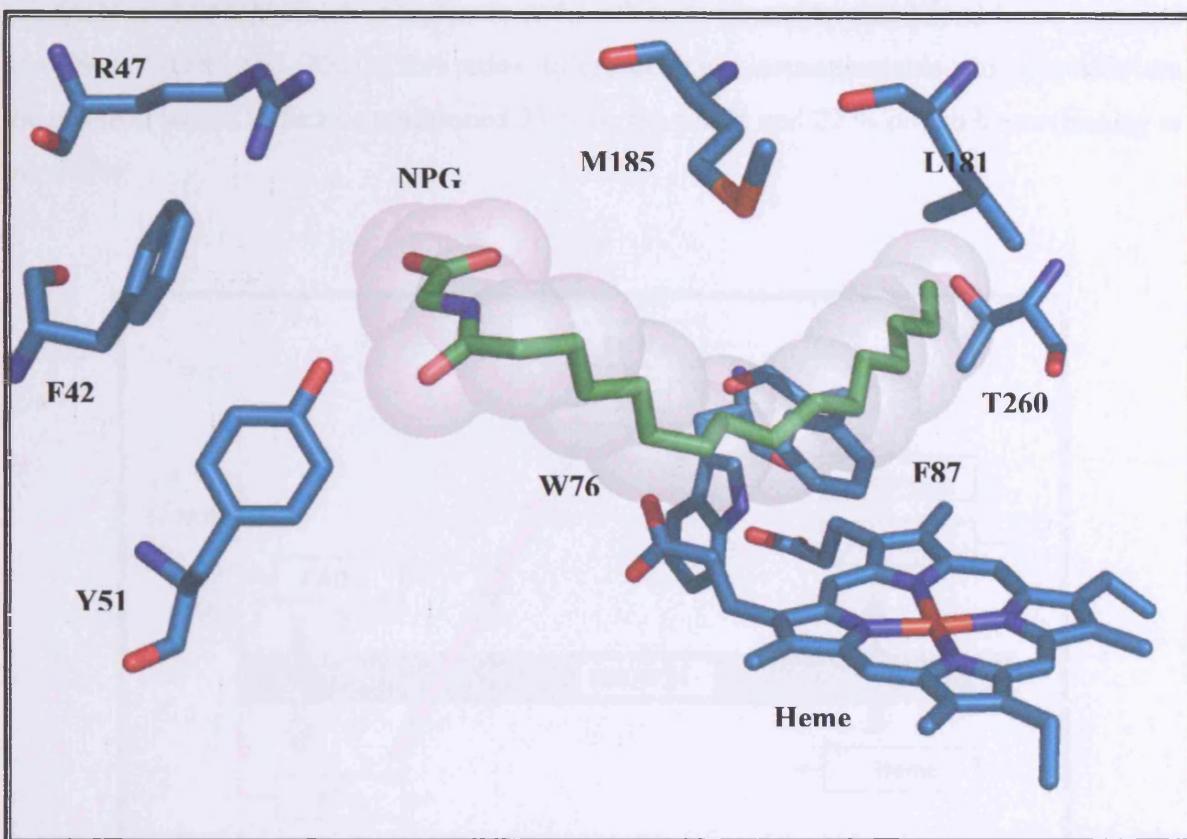
Cytochrome P450 BM3 is made up of a cytochrome P450 heme domain that is covalently attached to a CPR like diflavin reductase domain. This enzyme exhibits the highest catalytic rate of any P450 mono-oxygenase with rates of  $17,000 \text{ min}^{-1}$  seen with arachidonate (Noble *et al.*, 1999). This holoenzyme is highly amenable to biochemical characterisation due to high levels of protein expression in bacterial systems and the solubility of the holoenzyme. The enzyme hydroxylates the  $\omega$  terminal end of long chain fatty acids most frequently at the  $\omega$ -2 position (Narhi and Fulco, 1986).

### 1.4.1. Active site mutagenesis

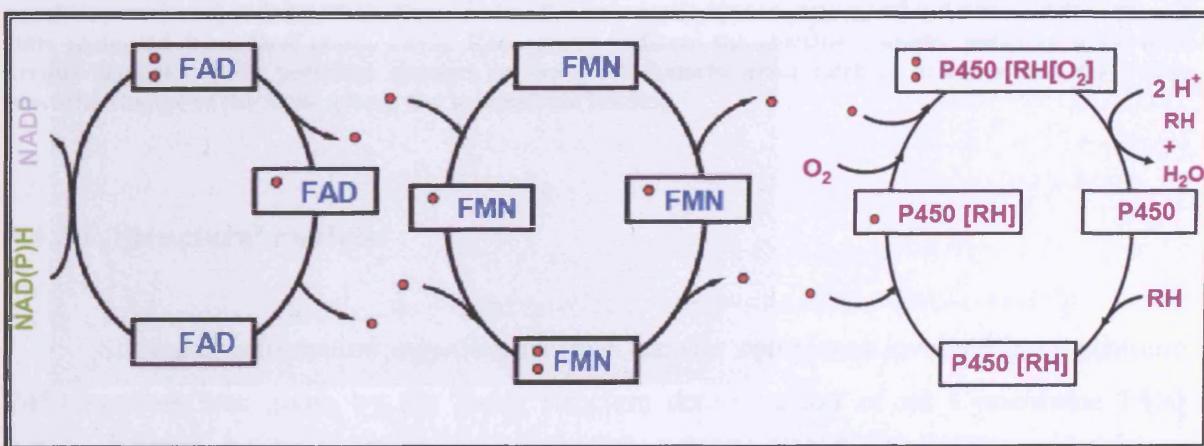
Multiple studies have been carried out on active site residues in P450 BM3 (Figure 1.9) in an effort to understand their role in substrate specification and catalysis. W96 has been shown to be important in heme binding (Munro *et al.*, 1994). The F87A mutant indicated that this residue was important in preventing hydroxylation at the  $\omega$ -terminus of the substrate. An F87G mutant increased the rate of polycyclic aromatic oxidation while an F87V mutation allowed further sole product formation of 14S, 15R-epoxyeicosatrienoic acid from arachidonic acid as opposed to a mixture with 18R-hydroxyeicosatetraenoic acid formed by wild-type P450 BM3 (Graham-Lorence *et al.*, 1997). Further mutations at R47 have changed substrate specificity (Oliver *et al.*, 1997) indicating the potential of this protein in specific product formation which could be utilised in biotechnological roles.

### 1.4.2. Electron transfer in cytochrome P450 BM3

The electron transfer in P450 BM3 can be measured spectroscopically as the absorption spectra of the flavin cofactors are altered by changes in their reduction state. This allows the measurement of the redox properties and subsequent electron transfer from NAD(P)H across the flavins and to the P450 heme and has led to invaluable understanding in the electron transfer mechanisms (Figure 1.10) and the elucidation of the path of electron transfer. In addition, it has led to the identification of the redox potential changes that occur within the system to allow efficient catalysis to take place (Daff *et al.*, 1997; Hanley *et al.*, 2004).



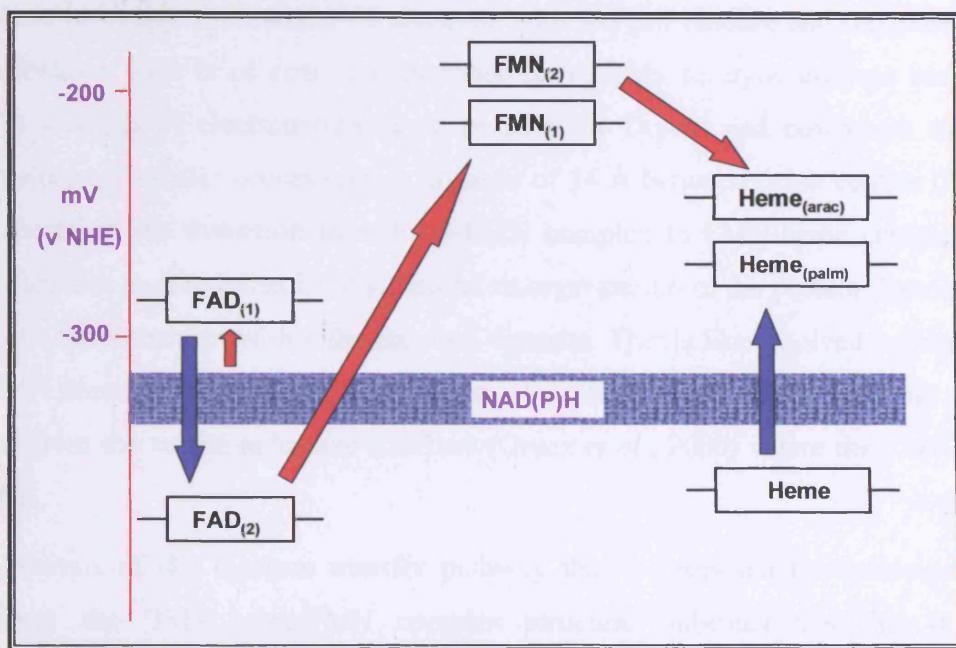
**Figure 1.9** Active site residues of P450 BM3 heme domain (PDB code: 1JPZ, Haines *et al.*, 2001). All molecules are shown in stick representation with the substrate N-palmitoleate glycine (NPG) also shown in space filling format.



**Figure 1.10** Schematic representation of the electron transfer process that occurs in cytochrome P450 BM3. Electrons represented by red circles (Adapted from Daff *et al.*, 1997).

This analysis has shown that there are little structural changes in the direct environment of the flavins throughout the electron transfer process (Daff *et al.*, 1997) and that the dissection of the individual domains leads to little change in the redox potentials of these domains. In cytochrome P450 BM3 the FAD domain is the low potential flavin and electron

transfer proceeds based on the redox potentials as seen in Figure 1.11. The redox potential of the FMN is -240 mV which is higher than the substrate (arachidonate) bound heme potential of -289 mV (Ost *et al.*, 2001). This redox difference is not insurmountable and at equilibrium the electron would in fact be partitioned 73 % on the flavin and 27 % on the heme (Hanley *et al.*, 2004)



**Figure 1. 11 Comparison of the individual reduction potentials for FMN, FAD (both from the diflavin reductase domain), and P450 domain both before and after saturation with palmitate (palm) and arachidonate (arac) in P450 BM3. Number 1 indicates a semiquinone state and 2 indicates a fully reduced state (Adapted from Daff *et al.*, 1997).**

#### 1.4.2.1. Structural analysis

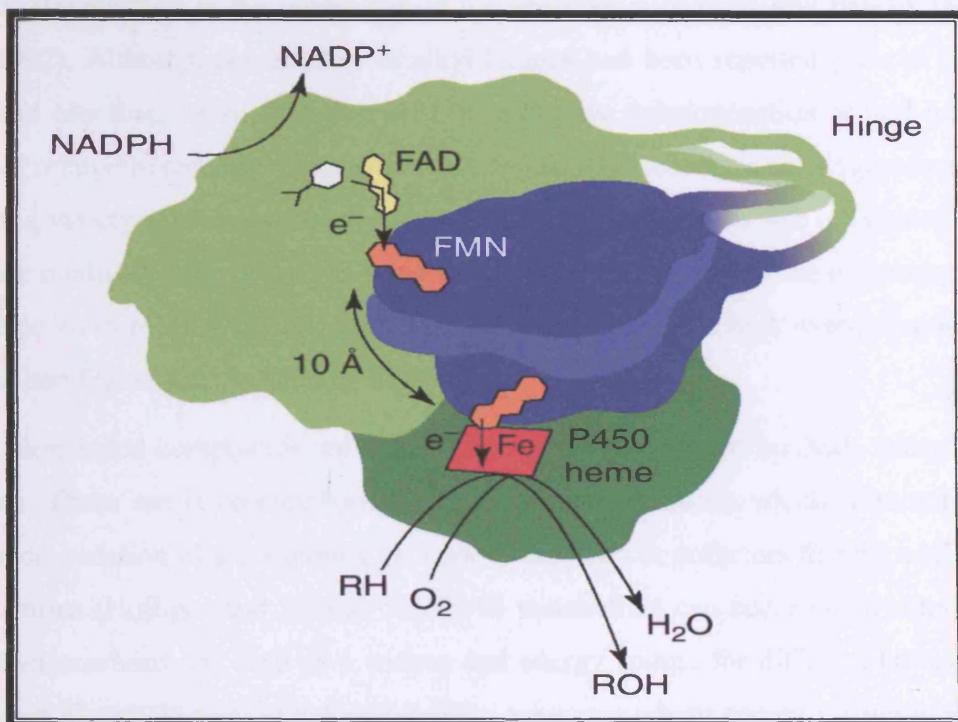
Structural information regarding electron transfer complexes involved in cytochrome P450 systems was given by the X-ray structure determination of rat Cytochrome P450 reductase (CPR; Wang *et al.*, 1997) and the subsequent structure of FAD/NADPH-binding Domain of Rat Neuronal Nitric-oxide Synthase (nNOS, Zhang *et al.*, 1999). These structures provide information about electron transfer from NAD(P)H to FAD and subsequently to the FMN prosthetic group. An FMN-Heme structural complex from BM3 was determined by Sevrioukova *et al.*, 1999a. This structure showed the FMN domain interacting with the proximal surface of the P450 domain. However, the linker region between the heme-FMN

domains was proteolysed and there are two heme molecules per single FMN molecule questioning the biological validity of the structure. The rate of electron transfer between the domains in this orientation would also be rather inefficient (Munro *et al.*, 2002) However, FMN interaction on the proximal surface had been proposed from previous electrostatic surface analysis and mutagenesis studies (Sevrioukova *et al.*, 1999b).

Electron transfer between NAD(P)H to heme must be highly regulated and controls are in place to couple the transfer of electrons with oxygen binding and oxygenation of the bound substrate. This is of critical importance in a highly catalytic enzyme such as P450 BM3. An analysis of electron transfer in proteins by Dutton and coworkers showed that efficient electron transfer occurs up to a distance of 14 Å between redox centres (Page *et al.*, 1999). Therefore the transition from FAD-FMN complex to FMN-heme complex presents some difficulties as it requires large structural re-organisation of the protein domains in order to bring the redox centres within the required distance. This is likely solved by high mobility of the FMN domain. Evidence for the mobility of FMN domains within reductase domains is presented from the sulfite reductase structure (Gruez *et al.*, 2000) where the FMN domain is not visible.

Analysis of the electron transfer pathway that is proposed by Sevrioukova *et al.*, 1999a from the BM3 heme/FMN complex structure indicates that this is a highly unfavourable pathway. This protein complex places the FMN cofactor group 18.4 Å away from the heme macrocycle and the electron transfer is proposed to take an elaborate pathway through a number of peptide atoms. Theoretical calculations of the electron transfer rates that would occur through the proposed 50 σ bond pathway between the heme and FMN cofactors give values of 0.02 year <sup>-1</sup> clearly ruling out this pathway as an efficient electron transfer mechanism (Munro *et al.*, 2002). The closest distance between the two cofactors in this structure is in the order of 18 Å which, using standard tunnelling expressions (Darrouzet *et al.*, 2001), gives values of ~ 12 s<sup>-1</sup> which does not agree well with the recorded 200 s<sup>-1</sup> for BM3 catalysis. This would indicate that further structural rearrangement is required and modelling indicates that the heme and FMN cofactors can indeed come within 8 Å of each other which would result in the higher catalytic rates that have been experimentally measured (Munro *et al.*, 2002). The FAD and FMN domains reside on distinct domains linked by a large hinge region which is proposed to move between a closed conformation in which the two flavin domains are juxtaposed for interflavin electron transfer, and an open conformation, after the FMN domain has moved ~10 Å to interact with and reduce the P450 (Figure 1.12). A similar hinged domain motion with comparable displacement has been characterized in the

cytochrome *bc<sub>1</sub>* complex (Xiao *et al.*, 2000) and in the human electron transferring flavoprotein: medium chain Acyl-CoA dehydrogenase complex (ETF: MCAD) structures (Toogood *et al.*, 2004, 2005).



**Figure 1. 12** Flavin mononucleotide (FMN) domain motion in flavocytochrome P450 BM3 (Taken from Munro *et al.*, 2002).

### 1.5. Degradation of Halogenated phenolic compounds

Past and present industrial and agricultural activities have led to the ever increasing presence of haloorganic compounds such as chlorophenols and chlorinated ethenes in the environment (Hileman, 1993). Due to both toxicity and their recalcitrant nature, increasing amounts of these xenobiotics threaten the integrity of the environment and human health. Some of the most toxic chemicals that exist are organochlorine pesticides, alkyl solvents, and aryl halides (Mohn and Tiedje, 1992). Efficient breakdown and clean-up of these chemicals presents a formidable challenge to both present day and future scientists and environmentalists. These chemicals are termed recalcitrant because microbiological flora found in soil is usually unable to break down these unnatural compounds. However, several reports detailing degradation have shown the dehalogenating potential of a range of fermentative, sulfidogenic, and methanogenic iron-reducing microbial communities (Kuhn and Suflita, 1989; Mohn and Tiedje, 1992; Holliger and Schrein, 1994). These bacteria have been shown to have high dehalogenation rates and result in low residual concentration of the contaminant following the introduction of these bacteria. This offers the possibility of a viable

and efficient means of removing these persistent compounds from nature as dehalogenation renders most xenobiotics less toxic and more readily degradable by other organisms (El Fantroussi *et al.*, 1998).

Dehalogenation is the initial step in the breakdown of most aryl halides (Mohn and Tiedje, 1992). Although degradation of alkyl halides had been reported prior to the 1980's (Raghu and MacRae, 1966; Malone 1970) the reductive dehalogenation of aryl halides was only noted relatively recently (Horowitz *et al.*, 1983). Aryl reductive dehalogenation has been studied in a variety of sites, however pure bacterial cultures that are able to dehalogenate aryl halides are relatively rare. The C-Cl bond is difficult to break due to the electronegativity of the chloride atom making the covalent bond polar resulting in a high energy requirement to break this bond (Zumdahl and Zumdahl, 2003).

Halogenated compounds are metabolised by three different methods under anaerobic conditions. These are i) co-metabolism i.e., a series of reactions which fortuitously brings about the degradation of a compound by a set of enzymes or cofactors that normally catalyse other reactions (Holliger and Schraa, 1994), ii) metabolism can occur when chlorinated C1 and C2 hydrocarbons are used as a carbon and energy source for different bacterial strains (Messmer *et al.*, 1993) and iii) halorespiration, a process where energy for microbial growth is generated using halogenated compounds as electron acceptor (Holliger and Schumacher, 1994).

### 1.5.1. Halorespiration

Halorespiration is a relatively novel anaerobic respiratory pathway linked to a chemiosmotic process as shown in Figure 1.13a. This highly efficient anaerobic process provides energy to the organism. Halorespiration remains the only known biodegradation method for the removal of certain significant pollutants including polychlorinated biphenyls (PCBs), tetrachloroethene (PCE) and pentachlorophenol (PCP; Bouwer and McCarthy, 1983; Brown *et al.*, 1987; Chaudhry and Chapalamuga, 1991; Copley 1998). These compounds are highly toxic with long half lives and their existence in nature is highly undesirable (Vogel *et al.*, 1987). The key enzyme in the halorespiratory pathway responsible for reduction of the halogenated terminal electron acceptor is a member of a novel class of corrinoid-containing reductases (a corrinoid is a tetrapyrrole ring system with a cobalt atom bound in the centre by four nitrogen atoms). One of the best studied member of the family is the *Desulfitobacterium dehalogenans* o-chlorophenol dehalogenase (van de Pas *et al.*, 1999, Smidt *et al.*, 2000). This

enzyme is a heterodimer with a catalytic subunit (CprA) that contains one [4Fe-4S] cluster, one [3Fe-4S] cluster, and one cobalamin-derivative per monomer (Figure 1.13b). CprA is attached to the membrane by association with the small transmembrane subunit CprB.

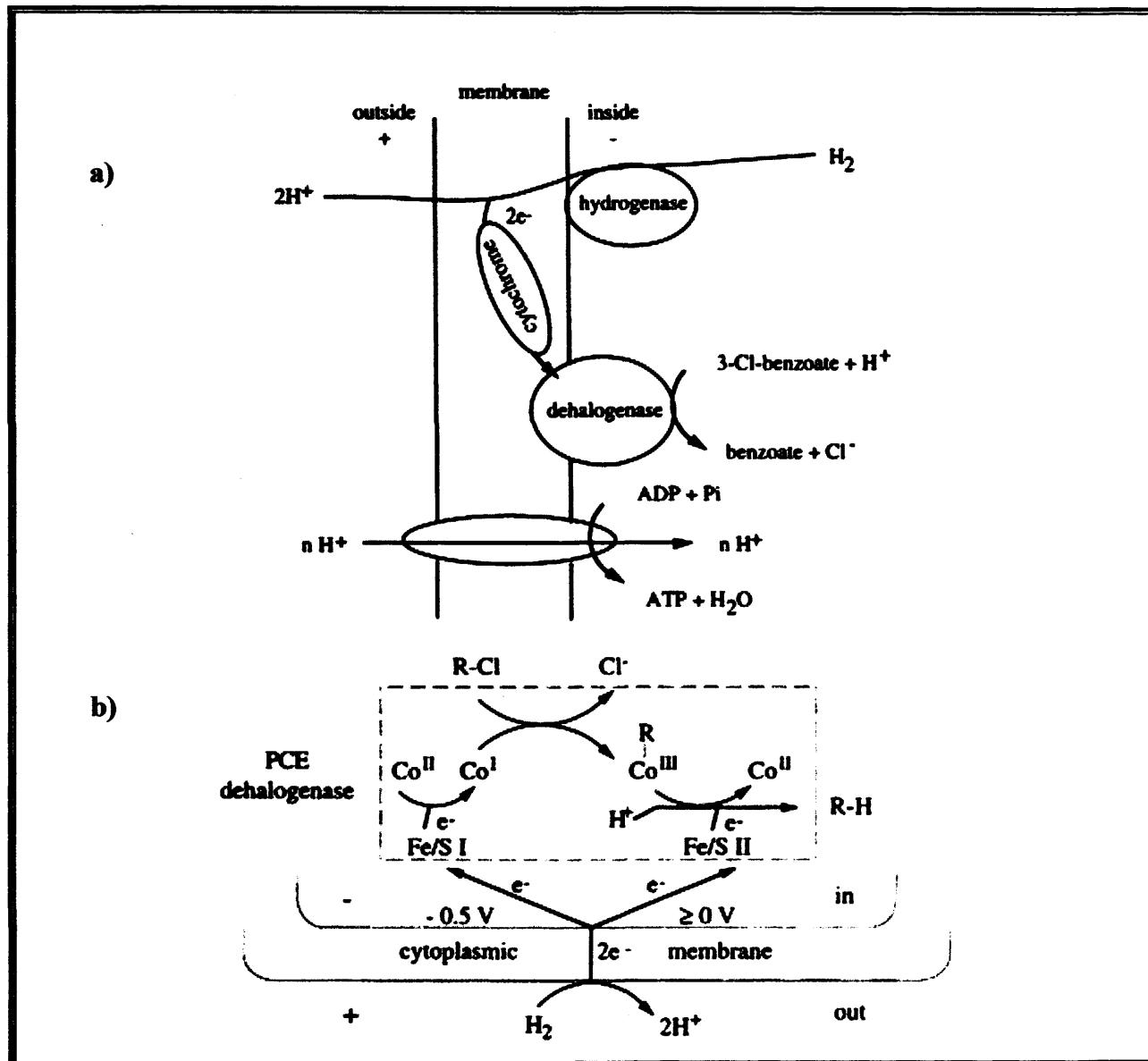


Figure 1.13 a) Schematic diagram of the chemiosmotic linkage between reductive dehalogenation and energy production (Taken from El Fantroussi *et al.*, 1998). b) Schematic diagram of the electron transfer process involved in the dehalogenase system (Taken from Neumann *et al.*, 1996)

### 1.5.2. *Desulfitobacterium* sp.

*D. dehalogenans* was isolated from a methanogenic lake sediment (Utkin *et al.*, 1994) and shown to remove the ortho position chloride from the compounds 2, 4-dichlorophenol and also 3-chloro-4-hydroxyphenylacetate (OCPA). Using OCPA as an electron acceptor and either pyruvate, lactate, formate or hydrogen as electron donors the bacterium was able to reductively dechlorinate the compounds and consequently use this energy for growth (Utkin *et al.*, 1994).

In contrast to a number of previously studied strains, dechlorination activity is highly specific for the ortho-position (Utkin *et al.*, 1995; Van Dort and Bedard, 1991). In order to identify the complex regulatory control and interactions at a molecular level, studies have focused not only on the reductive dehalogenase but also on the regulatory components of the electron transport chain identified in this halorespiring bacterium (Holliger *et al.*, 1999).

Characterisation of an 11.5 kb gene locus which flanked the dehalogenase gene was carried out in order to understand the expression and possible regulation of reductive dehalogenation (Smidt *et al.*, 2000). Sequence analysis of the *cpr* operon from *D. dehalogenans* responsible for the ortho-chlorophenol respiration revealed the presence of eight genes designated *cprTKZEBACD*. CprD and CprE are predicted to be molecular chaperones of the GroEL type, whereas *cprT* may encode a homologue of the trigger factor folding catalysts. The *cprC* and *cprK* gene products belong to the NirI/NosR and CRP-FNR families of transcription regulatory proteins, respectively. In *D. dehalogenans*, transcription of *cprBA* is induced 15-fold within 30 min of addition of OCPA with a concomitant increase in dehalogenation activity and halorespiration-specific transcriptional induction of the monocistronic *cprT* gene and the dicistronic *cprBA* and *cprZE* genes. The induction of the system is thought to be heavily reliant on the transcriptional regulator CprK for response to o-chlorophenol presence and expression of the multiple proteins involved in this system. This is due to the identification of palindromic DNA binding regions situated 35 bp upstream of the coding regions of the genes that are similar to the FNR type promoter regions.

Analysis of the *D. hafniense* genome further identified the presence of 3 *cprA*-like gene fragments which also contained other genes related to the *cpr* gene locus with three predicted ORFs with greater than 35% sequence identity to CprK (Frishman *et al.*, 2003). Several gene clusters have been identified in other members of the *Desulfitobacterium* sp. that are similar to the *cpr* gene cluster indicating that these bacteria use distinct but related

enzymes to achieve the dehalogenation of several chlorinated compounds (Villemur *et al.*, 2002). The understanding of how the chlorophenol regulation system is controlled would allow further utilisation of these bacteria as a biotechnological tool and environmental aid in the future.

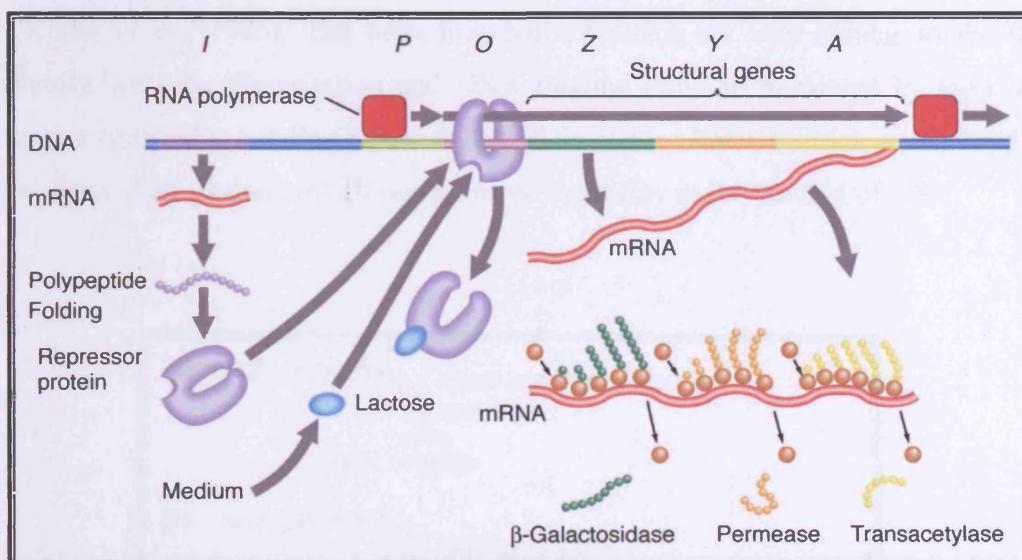
## 1.6. DNA binding proteins

DNA binding proteins are involved in all aspects of genetic control within cells including transcriptional control, replication and DNA damage recognition and repair mechanisms. Analysis of the genome sequences in PEDANT (Frishman *et al.*, 2003) indicates that 2-3 % of prokaryotic genomes and 6-7 % of eukaryotic genomes are made up of DNA binding proteins. The most common members of this family are the Zinc fingers (Rubin *et al.*, 2000; Sarai and Kono, 2005), the Helix-turn-helix proteins (Brennan and Matthews, 1989; Aravind *et al.*, 2005) and the Leucine zippers (Vinson *et al.*, 1989, 2002). These are the largest representatives of transcriptional regulators that have been structurally identified. A full analysis of the protein-DNA complexes that are present in the PDB is presented in a review by Luscombe *et al.*, 2000 which gives a clear division of the different DNA binding protein families with 54 structurally distinct families identified.

### 1.6.1. Transcriptional Regulation

Transcriptional regulation is quite different in eukaryotes compared to prokaryotes due to the presence of a eukaryotic nuclear membrane, which prevents the simultaneous transcription and translation system that occurs in prokaryotes. Thus, in prokaryotes the level of control is solely at the transcriptional level, while within eukaryotes there are multiple points of control (Jackson, 2005). In prokaryotes, initiation of gene expression is controlled at two sites, 35 and 10 bases upstream of the gene, known as promoter regions recognized by RNA polymerase. Activity at these promoter regions is regulated by transcriptional regulators that activate and repress gene transcription. Transcription factors are regulated by a host of methods including dimerisation, phosphorylation and ligand binding. Thus it can be seen that transcription is regulated by a series of synergistic or inhibitory mechanisms. Upon DNA binding, the transcriptional regulator influences gene expression either through interaction with other transcriptional factors or direct interaction with the transcription polymerase.

An example is the *lac* operon made up of one regulator and  $\beta$  galactosidase, transacetylase and permease, three proteins that are involved in the breakdown of sugar for energy production. This operon is controlled by negative regulation i.e. transcription occurs except in the presence of a regulator molecule that binds to the promoter region of the enzyme genes and prevents transcription. However, in the presence of a repressor binding molecule, the repressor can no longer bind to the promoter region and transcription and expression of the sugar metabolising proteins occur (Figure 1.14).



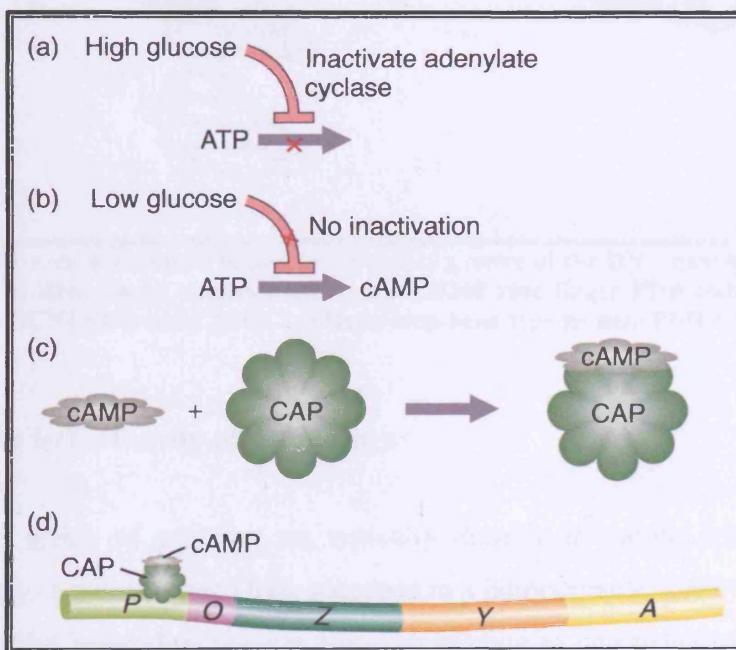
**Figure 1. 14 Schematic representation of the lac operon with expression repressed in the absence of lactose. Binding of lactose inhibits DNA binding by the lac repressor allowing the RNA polymerase to bind.**

(Taken from [http://www.mun.ca/biology/scarr/4241Apr\\_Regulation\\_of\\_Gene\\_Transcription\\_Robyn\\_Page.html](http://www.mun.ca/biology/scarr/4241Apr_Regulation_of_Gene_Transcription_Robyn_Page.html))

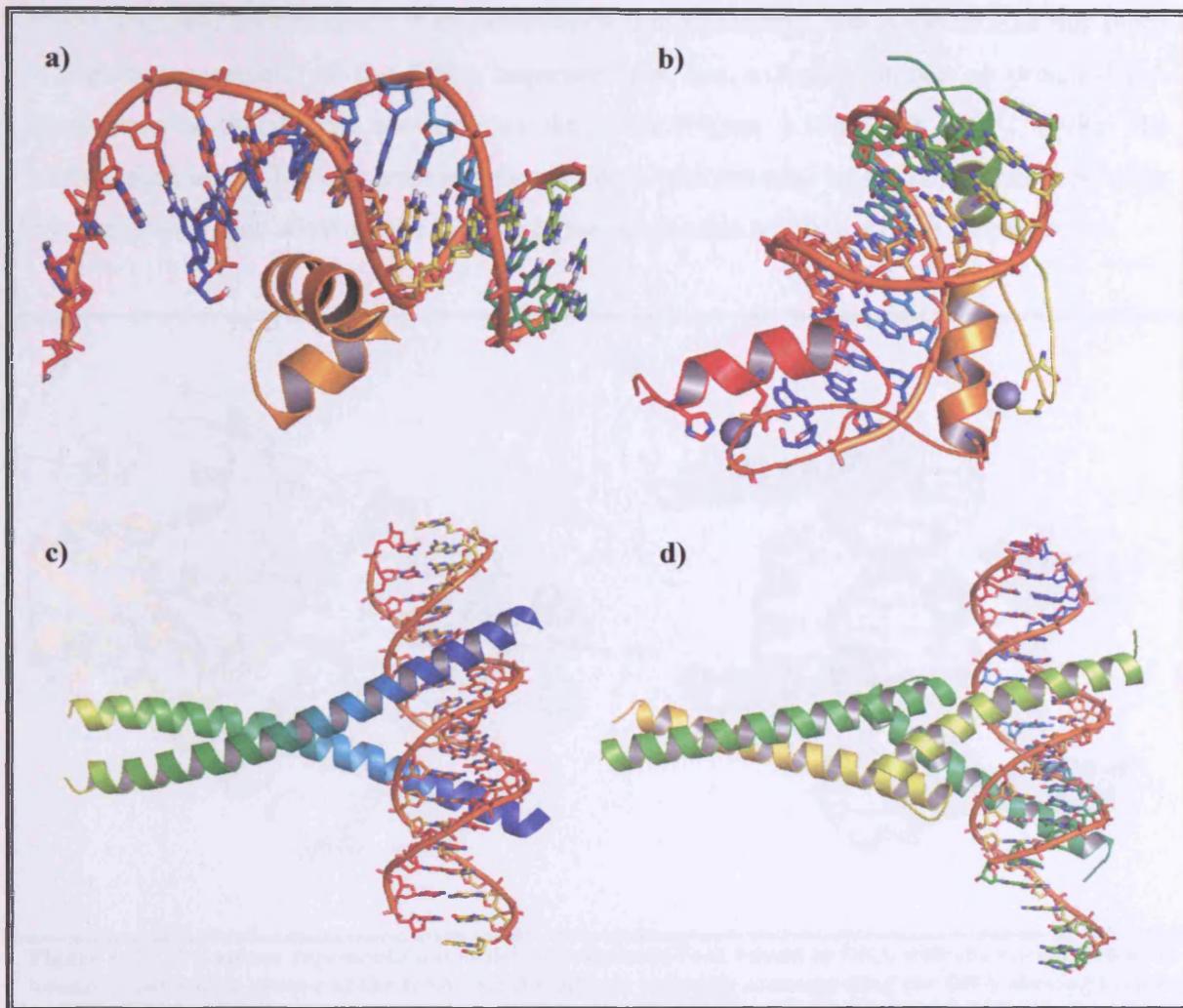
The expression from this operon is inefficient and is positively regulated by the transcriptional regulator cAMP receptor protein. When glucose is present, there are low levels of cAMP and as glucose levels fall the levels of cAMP increase. The cAMP-CRP complex binds to a specific DNA region that leads to improved RNA polymerase binding (Figure 1.15). This ultimately leads to a 20-50 fold increase in the transcription of  $\beta$  galactosidase, transacetylase and permease (Kolb *et al.*, 1993).

The transcriptional regulator group of proteins is made up of a vast array of individual members of different protein families of which the most common are the helix-turn-helix (HTH) family of proteins, Zinc coordinating proteins and the leucine zipper family (Luscombe *et al.*, 2000). The HTH motif is made up of two alpha helices held at a specific angle to each other separated by a  $\beta$  turn (Figure 1.16a). The zinc finger transcription factors require zinc in order to bind to DNA and two motifs have been identified within this family;

one contains a 30 amino acid stretch that contains two invariant histidine cysteine pairings that form a tetrahedral zinc binding site (Figure 1.16b, PDB code: 1AAY, Miller *et al.*, 1985, Elrod-Erickson *et al.*, 1996). In the second example, cysteine residues form interactions that allow zinc binding as seen in the steroid binding hormones (Schwabe and Rhodes, 1991). The leucine zipper motif is seen in a number of proto-oncogene products such as Myc, Jun and Fos and is made up of leucine residues that present every seven residues along the hydrophobic face of an  $\alpha$  helix. Dimerisation occurs along these hydrophobic faces mediated by the leucine residues and causes DNA binding by two sets of C-terminal basic residues that bind to the major groove of the DNA helix in opposing directions (Figure 1.16c, PDB code: 2DGC, Keller *et al.*, 1995). The helix-loop-helix proteins are very similar to the leucine zipper family with the dimerisation and DNA binding domains separated by loop regions resulting in a four helix bundle (Figure 1.16d, PDB code: 1AM9, Parraga *et al.*, 1995). The presence of two distinct domains allows for more flexibility in the binding of DNA.



**Figure 1. 15 Diagram of the influence of glucose levels on cAMP production leading to increased expression within the lac operon. CRP is denoted as CAP in this figure. (Taken from [http://www.mun.ca/biology/scarr/4241Apr\\_Regulation\\_of\\_Gene\\_Transcription\\_Robyn\\_Page.html](http://www.mun.ca/biology/scarr/4241Apr_Regulation_of_Gene_Transcription_Robyn_Page.html))**

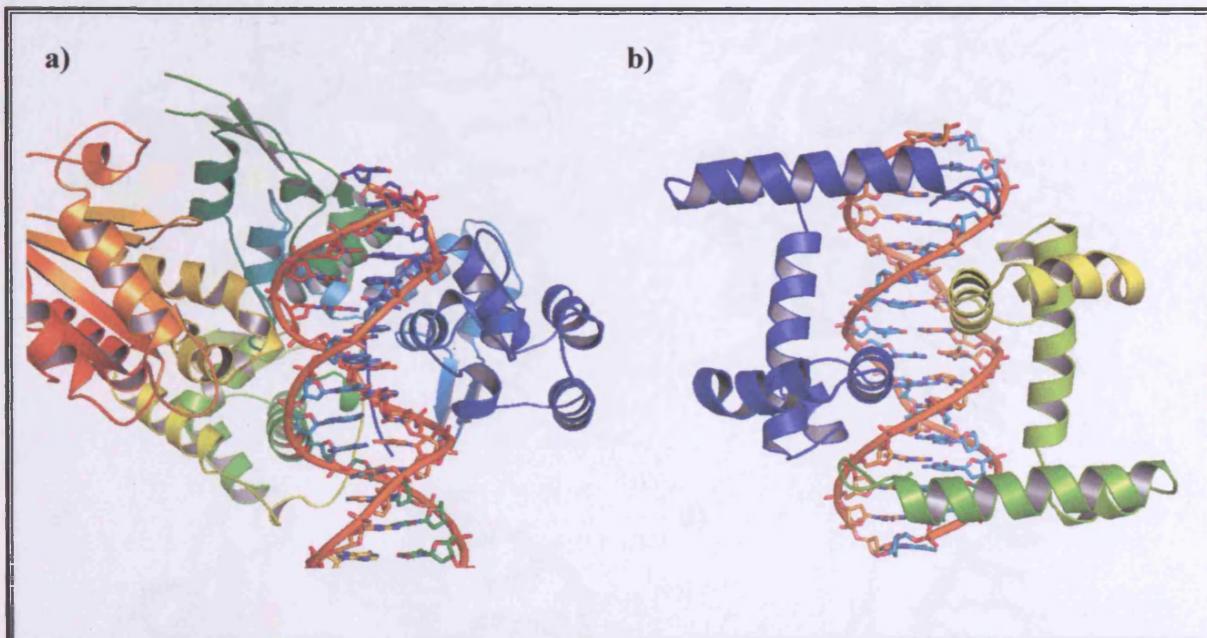


**Figure 1.16** a) Helix turn helix motif bound in the major groove of the DNA double helix (part of CRP, PDB code: 1CGP). b) Zinc finger transcription factor Zif268 zinc finger PDB code: 1AAY. c) Leucine zipper type protein GCN4 PDB code: 2DGC. d) Helix-loop-helix type protein PDB code: 1AM9

### 1.6.2. Helix-turn helix family of regulators

The HTH group of proteins are typically dimeric in nature with the prokaryotic examples binding to a palindromic DNA sequence in a homodimeric form whereas eukaryotic members also exhibit heterodimerisation allowing binding to non-palindromic sequences (Li *et al.*, 1995, PDB code: 1YRN). This family of transcriptional regulators bind in the major groove of the DNA helix, with direct contacts between protein side chains of the second helix known as the recognition helix and nucleotide bases (Figure 1.17). Additional contacts are made by the first helix and the linker region. The precise positioning of the recognition helix within the DNA major groove varies within the different proteins structurally determined to date and alignment of the helix with the base-pair edges of the nucleotides has been observed alongside other examples of alignment with the sugar-phosphate backbone (Suzuki *et al.*,

1995). The FokI endonuclease is an exception within this family and is a monomer that binds in the major groove of the DNA sequence and has a large number of protein-DNA interactions as the protein encompasses the DNA (Figure 1.17a; Wah *et al.*, 1997). The binding seen in the Trp repressor is unique to date with the edge of the helix almost pointing into the groove thus allowing the binding of two molecules within the DNA major groove.

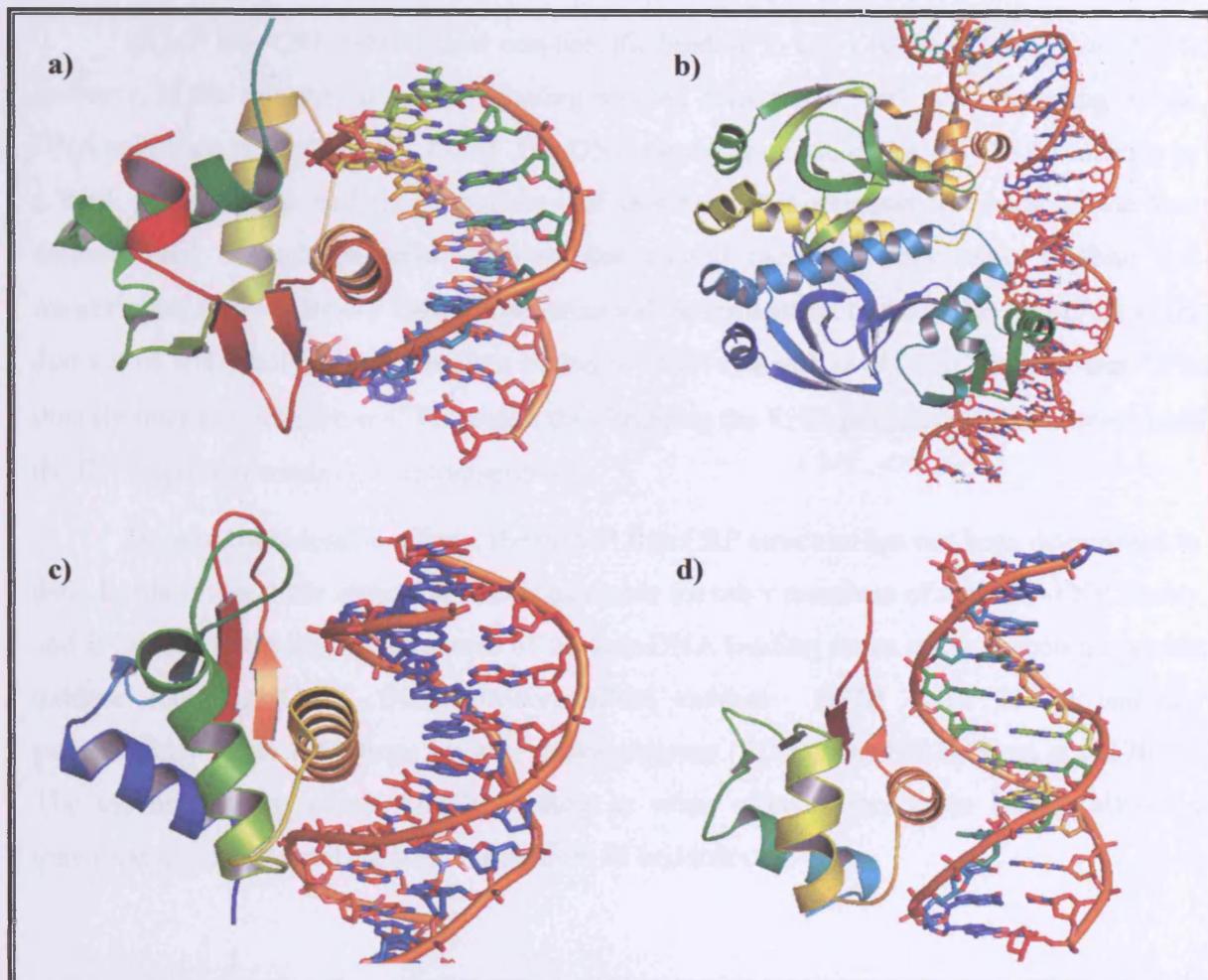


**Figure 1.17** a) Cartoon representation of the endonuclease FokI bound to DNA with the recognition helix bound in the major groove of the DNA and the protein molecule encompassing the DNA eliciting further specific interactions. b) DNA binding by the tryptophan repressor molecule with the two recognition helices shown face on interacting with the major groove of the DNA helix. Protein molecules are shown in cartoon format with the DNA bases shown in stick representation.

The ‘winged’ HTH family is an extension of the HTH family which has a conserved domain containing a third  $\alpha$ -helix and an adjacent  $\beta$  sheet which are considered to be part of the DNA binding motif. The addition of these secondary structural elements allows additional protein contacts to be made with the DNA. Structurally identified protein-DNA complexes of this family identified to date include the Interferon regulatory factor (PDB code: 2IRF, Fujii *et al.*, 1999), the transcription factor family (PDB code: 3HTS, Littlefield and Nelson, 1999) and the Ets domain family (PDB code: 1BC8, Mo *et al.*, 1998) as shown in Figure 1.18.

Within this ‘winged’ HTH family is the CRP-FNR group of proteins of which cAMP receptor protein (CRP) is extensively studied and has attained textbook status. This family of regulators stand out by virtue of their ability to respond to a broad spectrum of intracellular and exogenous signals such as cAMP, redox state, stress, nitric oxide, carbon monoxide, 2-oxoglutarate, or temperature (Korner *et al.*, 2003). These proteins are made up of two domains, an inducer binding domain and a DNA binding domain. To accomplish their roles,

CRP-FNR members have highly specific binding sites allowing the binding of allosteric effector molecules or contain prosthetic groups for interaction with signal molecules. Significant structural reorganisation has been proposed to take place upon inducer binding allowing DNA binding to occur and subsequent induction of transcription (Kolb et al., 1993).



**Figure 1.18 Members of the HTH family of transcriptional regulators in their DNA bound form.** a) CRP PDB code 1CGP b) interferon regulatory factor: PDB code 2IRF, c) transcription factor: PDB code 3HTS, d) Ets domain family: PDB code 1BC8. The protein is shown in cartoon format with the DNA bases shown in stick format.

From the available HTH DNA binding protein structures, CprK has most sequence similarity to cAMP receptor protein (CRP). To date, crystal structures of the cAMP-bound CRP (Passner et al., 2000, PDB code: 1G6N), the CRP-DNA complex (PDB code: 1J59, Parkinson et al., 1996) and the CRP- $\alpha$ CTD-DNA complex (Benoff et al., 2002, PDB code: 1LB2) are available, providing atomic insights into the cAMP-, DNA- and RNA-polymerase-binding interactions. The cAMP bound CRP has two cAMP molecules bound which bind in a region close to the subunit interface of the dimer. Analysis of the binding indicates that it is sequential and exhibits negative co-operativity (Heyduk and Lee, 1989). From the crystal

structure and the binding studies, CRP is thought to exist in three distinct conformations. These are i) cAMP free structure, ii) CRP: cAMP<sub>1</sub> and iii) CRP: cAMP<sub>2</sub>. At physiological concentrations however, only the cAMP free protein and CRP: cAMP<sub>1</sub> structures are thought to exist and coupled with this data is the *in-vitro* measurement of a 1:1 binding stoichiometry between CRP and cAMP in a DNA bound form (Hudson *et al.*, 1990).

cAMP free CRP exhibits low non-specific binding to DNA (Giraud-Panis *et al.*, 1994) however, in the presence of cAMP binding reaches between  $10^7$ - $10^{10}$  M<sup>-1</sup> depending on the DNA sequence (Ebright *et al.*, 1989). The DNA bound structure shows the DNA molecule in a bent conformation and the importance of this bent conformation for transcription was demonstrated through sequence changes that caused increased polymerase binding and transcription (Bracco *et al.*, 1989). The structural determination of the CRP: cAMP:  $\alpha$ -CTD domain of RNA polymerase complex bound to DNA (Benoff *et al.*, 2002) shows that CRP directly interacts with the  $\alpha$ -CTD region thus bringing the RNA polymerase into contact with the DNA leading to subsequent transcription.

Despite considerable efforts, the cAMP free CRP structure has not been determined to date. Furthermore, little structural data is available for other members of the CRP-FNR family and is limited to the crystal structures of the non-DNA binding states of the carbon monoxide oxidase inducing CooA from *Rhodospirillum rubrum* (PDB code: 1FT9) and the pathogenicity factor PrfA from *Listeria monocytogenes* (PDB code: 1OMI, Vega *et al.*, 2004). The mechanism by which cAMP binding or other effector molecules induce allosteric transition to a DNA-binding state is therefore ill understood.

## 1.7. Aims of the project

Studies to date on cytochrome P450 BM3 have identified this holoenzyme as a highly efficient catalytic enzyme that is capable of mono-oxygenation reactions on fatty acid substrates in the order of 17,000 min<sup>-1</sup>. Due to the solubility and the high levels of catalysis this enzyme is of great interest for biotechnological uses. The high level of homology of the cytochrome P450 domain to human P450 enzymes also allows studies on this enzyme to have relevance in understanding eukaryotic P450 enzyme forms. Protein engineering of this system for novel enzymatic functions requires extensive understanding of the protein. Determination of the full cytochrome P450 BM3 holoenzyme structure has been elusive to date. However significant structural information has been obtained for substrate bound forms of this enzyme

and the substrate free form. Chapter 4 of this thesis describes an A264E mutant of the P450 domain in both the presence and absence of substrate which gives valuable information on structural changes that are seen upon substrate binding and catalysis. Chapter 5 of this thesis describes the structural determination of the FAD domain of the cytochrome P450 BM3 holoenzyme. The structural identification of this domain allows rationalisation of mutagenesis studies that have been carried out to date. Together these structures give valuable information on the electron transfer mechanism and substrate binding within this complex protein system.

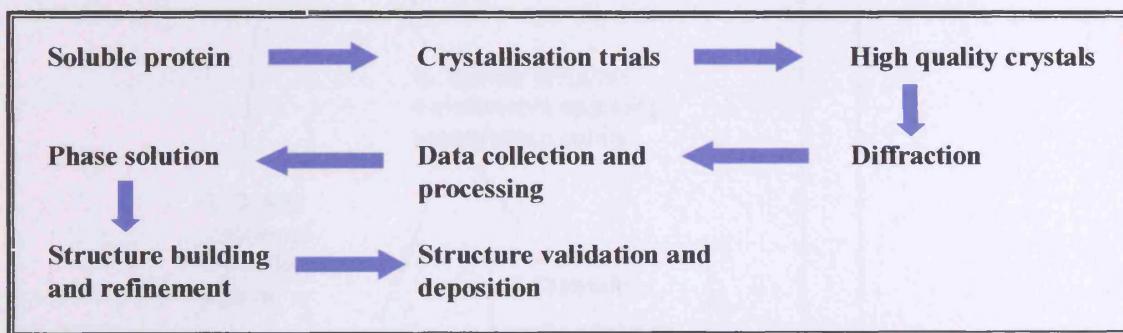
Chapter 6 describes the structural determination and ligand binding studies of a HTH transcriptional regulator CprK from *D. hafniense*. These studies serve to dissect the process of ligand binding and subsequent transcriptional regulation of a dehalogenating enzyme. It also serves to improve on the understanding of the HTH family of proteins and how small ligands translate effector binding to transcriptional regulation. These studies are of further relevance to BM3 which is also regulated by a HTH transcriptional regulator and improved understanding of this family of proteins would be of benefit in understanding BM3 regulation.

## **CHAPTER 2**

### **INTRODUCTION TO CRYSTALLOGRAPHY**

## 2.1. Introduction

The exponential growth in the amount of X-ray crystal structures available is due to a number of factors. These range from the increase in high throughput crystallography consortiums (Pusey *et al.*, 2005), the development of efficient crystallographic computing programs (The CCP4 suite, 1994), the relatively inexpensive computing power that is presently available plus the availability of third generation synchrotron sources (Helliwell, 2001). This chapter describes some of the important components of x-ray crystallography for protein structure determination as shown in Figure 2.1. Further reviews which deal with the fundamentals and important aspects of the process of structure solution of proteins and macromolecules through crystallography can be found elsewhere (Blundell & Johnson, 1976; Drenth, 1996; Stout & Jensen, 1989; Woolfson, 1970).

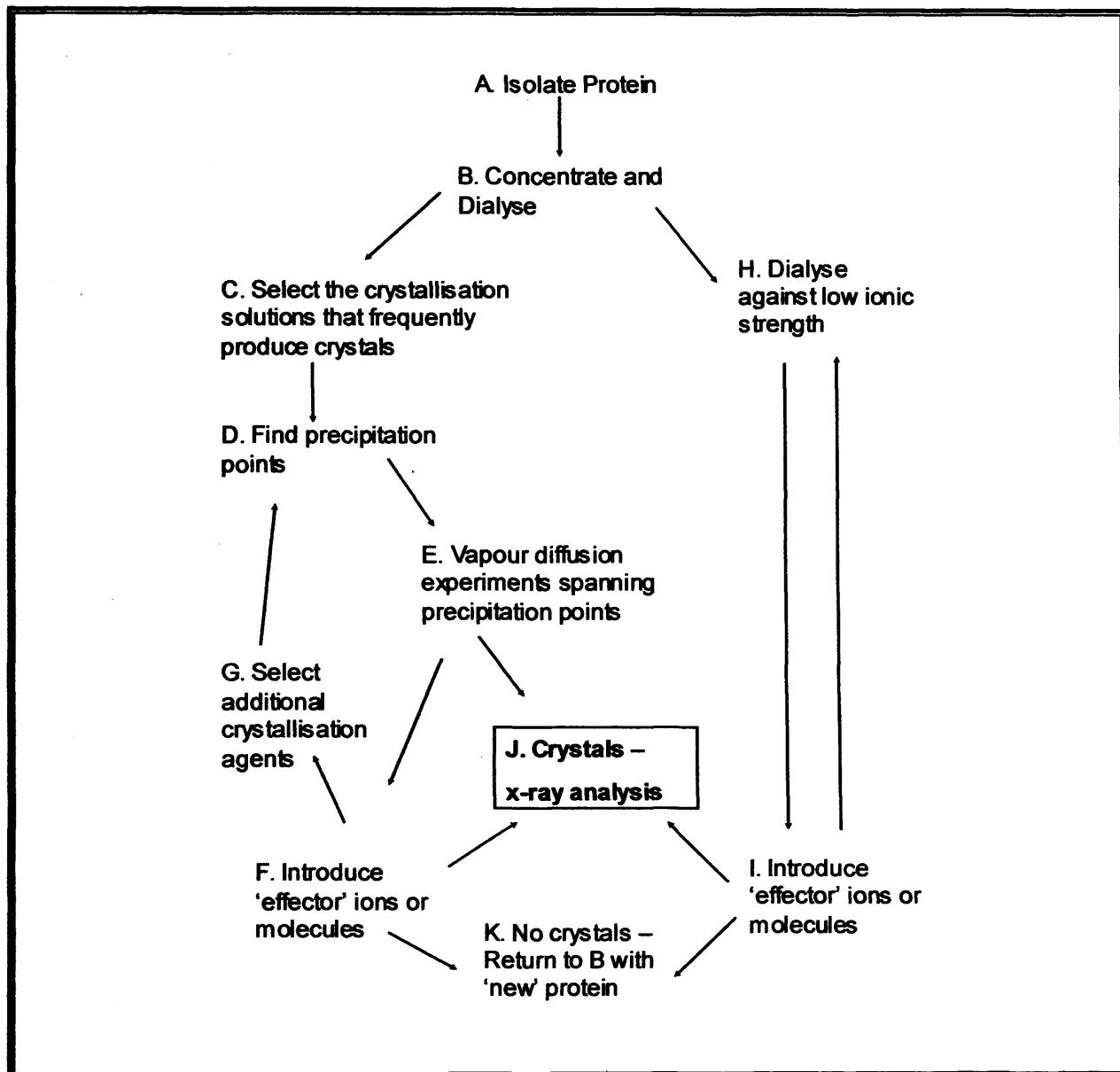


**Figure 2. 1 Overall schematic diagram of the steps involved in successful structure determination**

## 2.2. Crystal growth

The initial step towards X-ray crystallographic studies requires the production of sufficient quantities of pure protein from a natural source or by recombinant means. Recombinant technology has led to the production of mammalian and human proteins in insect and bacterial cells that are capable of producing large amounts of the protein of interest. Crystallisation of biological macromolecules is often considered unpredictable but has been shown to follow the same characteristics as the crystallisation of small molecules (Giege *et al.*, 1995; McPherson *et al.*, 1995) and is a multi-factorial process. A crystal growth strategy based on analysis of crystal conditions of over 600 biological molecules has been proposed by Gilliland (1988) and is shown in Figure 2.2; however, ultimately crystallisation is still an empirical process. This is due to protein molecules having large irregular surfaces with protrusions, cavities and a variety of hydrophobic and hydrophilic regions. This prevents

protein molecules from easily interacting with each other in an ordered manner which is critical in crystal formation.



**Figure 2. 2 A general strategy for exploring growth conditions of soluble proteins (taken from Gilliland, 1988).**

During a typical crystallisation experiment, protein samples are combined with a number of reagents that cause changes in the protein solubility. These solubility changes are described by the phase separation diagrams as shown in Figure 2.3 (Ries-Kautt and Ducruix, 1992). For crystal growth to occur, the protein solution must reach the nucleation zone, which under optimum conditions leads to only a small number of crystal nuclei being produced. The concomitant decrease in the protein solution concentration brings the system to the metastable region where crystal growth proceeds in the absence of further nucleation.

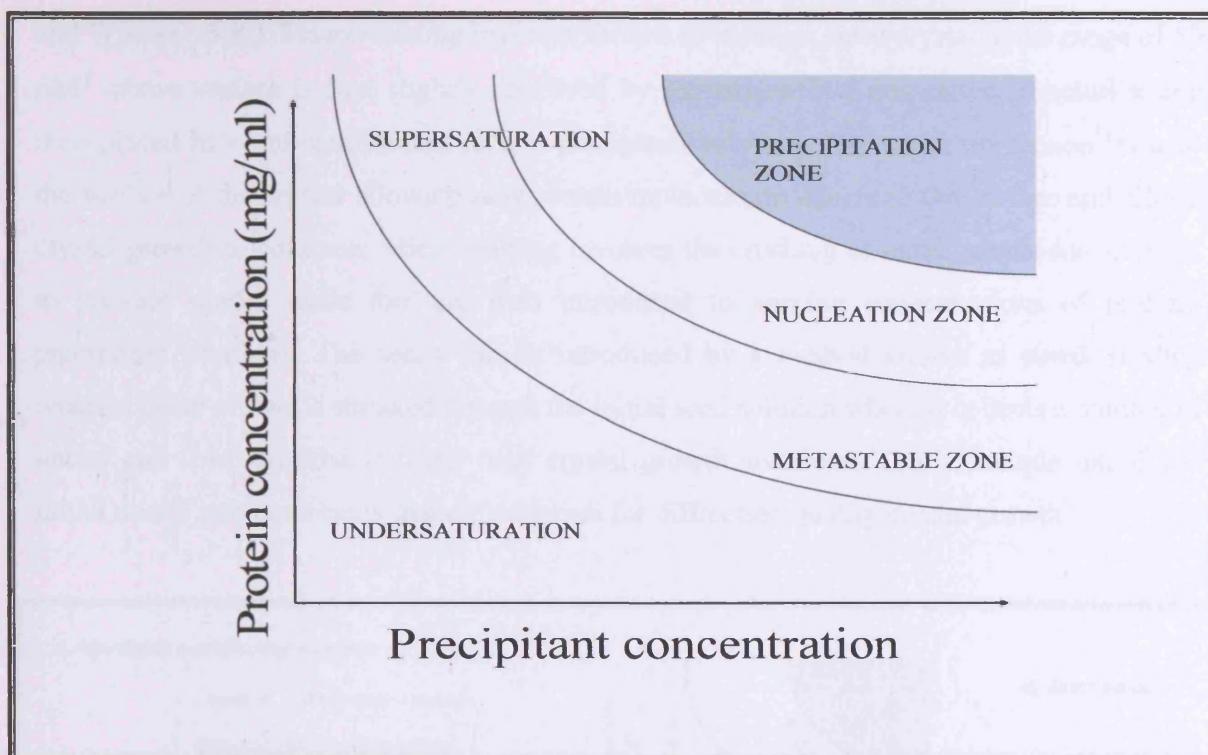
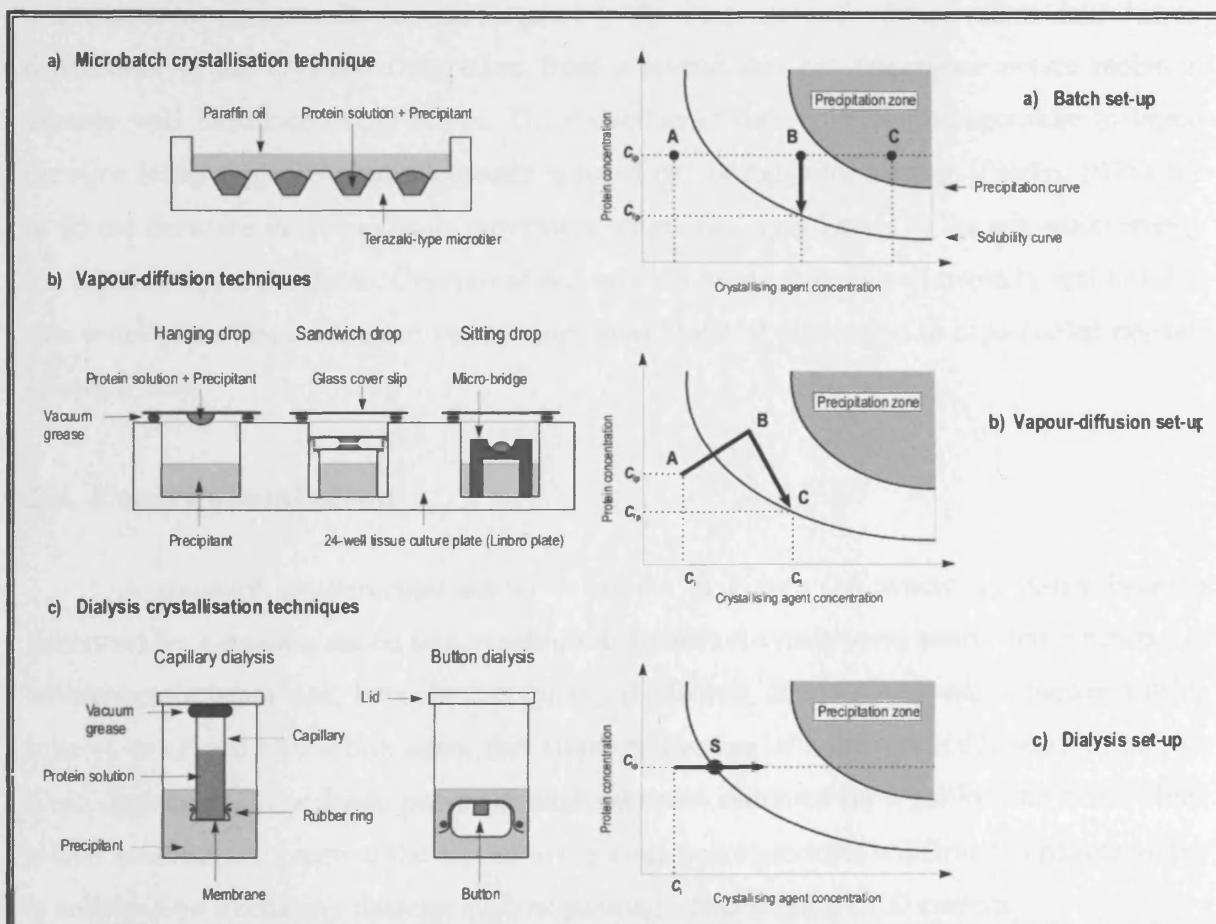


Figure 2.3 Theoretical phase separation diagram of a protein in solution.

On occasion, crystal conditions may only give small or imperfect crystals which still require further optimisation prior to diffraction data collection. Thus, a further screening process is carried out to optimise the crystallisation conditions. As macromolecules are labile and readily lose their structure they grow in a relatively narrow range of pH, temperature and ionic strength. Locating and optimising the conditions required to produce good crystals can be difficult in view of the large number of conditions that need to be analysed. Typically, 300 crystallisation conditions are initially screened with each one of these conditions made up of 2-4 components. A statistical approach is recommended to solving this “needle in a haystack” problem (Ducruix and Giege, 1999). This premise works on the basis of an incomplete factorial design where random combinations of experimental conditions are selected that range over the entire crystal growth parameter space. Through averaging the results of non-identical experiments that share particular features a common trend may be detected and a possible lead may be exposed for assessment via further trials. There are furthermore a range of methods for crystallisation of the protein of interest (Figure 2.4) and this may be another important parameter in crystallisation screening.

Due to the difference in growth conditions and nucleation conditions, improved crystals may be developed by seeding techniques. This utilises initial crystals in two seeding methods known as macro-seeding (Thaller *et al.*, 1985) and micro-seeding procedures (Stura

and Wilson, 1990). Macro-seeding involves the use of medium sized crystal in the range of  $30 \mu\text{M}^3$  whose surface is first slightly dissolved by immersion in a non-saturated solution and then placed in a pre-equilibrated protein-precipitant solution. The initial immersion ‘cleans’ the surface of the crystal allowing new protein molecules to adhere to the surface and allows crystal growth to continue. Micro-seeding involves the crushing of initial needle-like crystals to provide nuclei seeds that are then introduced to varying concentrations of protein-precipitant solutions. The seeds can be introduced by a method known as streak seeding where a piece of hair is streaked through the initial seed solution where it collects a number of nuclei and then streaked into the final crystal growth solutions. This technique introduces initial nuclei into conditions that are optimum for diffraction quality crystal growth.



**Figure 2.4** Schematic representations of protein crystallisation techniques and their corresponding solubility phase diagram and correlation between protein and crystallising agent in a) microbatch, b) vapour-diffusion and c) dialysis crystallisation experiments.  $C_p$  and  $C_1$  are the initial concentrations of protein and crystallising agent respectively,  $C_{fp}$  and  $C_f$  are the final concentrations. Figures taken from <http://daffy.bioc.cam.ac.uk/~dima/whitepapers/xtal-in-action/node3.html>

## 2.3. Cryo-protection

Once the optimum crystals have been obtained, crystals are transferred to a solution of cryoprotectant and flash cooled by rapidly plunging the crystal (suspended in a loop) into liquid nitrogen. The introduction of cryo-crystallography has developed since the late 1980's from stumbling experimentation into a routine procedure (Hope, 2001). The x-ray beam often readily causes radiation damage to protein crystals and cryoprotection can drastically reduce this damage. Radiation damage is related to the formation of free radicals within the protein crystal that react with the protein molecules causing damage and rapid decay in diffraction power. This damage can often be so rapid that a number of different crystals are required to get a full data set and in some cases data collection may be impossible. The use of more than one crystal always results in inaccuracies in the measurement of equivalent data due to differences in the crystals. Data taken from a crystal that has undergone severe radiation damage will introduce decay errors. The reduction of data collection temperature to liquid nitrogen levels typically shows a greatly reduced rate of radiation damage (Petsko, 1975) due to (i) the decrease in free radicals movement within the crystal and (ii) the activation energy for reaction is not available. Cryoprotectant solution needs to be experimentally optimised as this usually has large influence on the resolution limits of diffraction in cryo-cooled crystals (Hope, 1988).

## 2.4. Experimental setup

A standard experimental set-up is shown in Figure 2.5 where an X-ray beam is generated by a rotating anode or a synchrotron source. A synchrotron source has a number of advantages in beam size, intensity and quality (Helliwell, 2001). The beam is focussed using mirrors or crystal monochromators that allow production of a characteristic beam of narrow fixed dimensions. The beam passes through a crystal mounted on a goniometer head which allows rotation. Exposure of the crystal to the x-ray beam produces a diffraction pattern which is collected on a sensitive detector such as an image detector or a CCD camera.

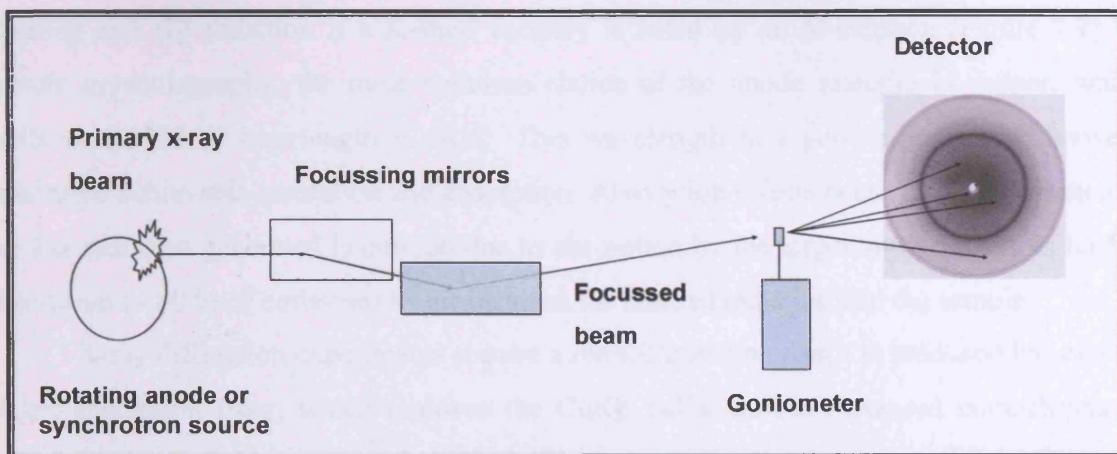


Figure 2. 5 Diagram of the components involved in a typical x-ray diffraction experimental set-up

### 2.4.1. Sources of X-rays

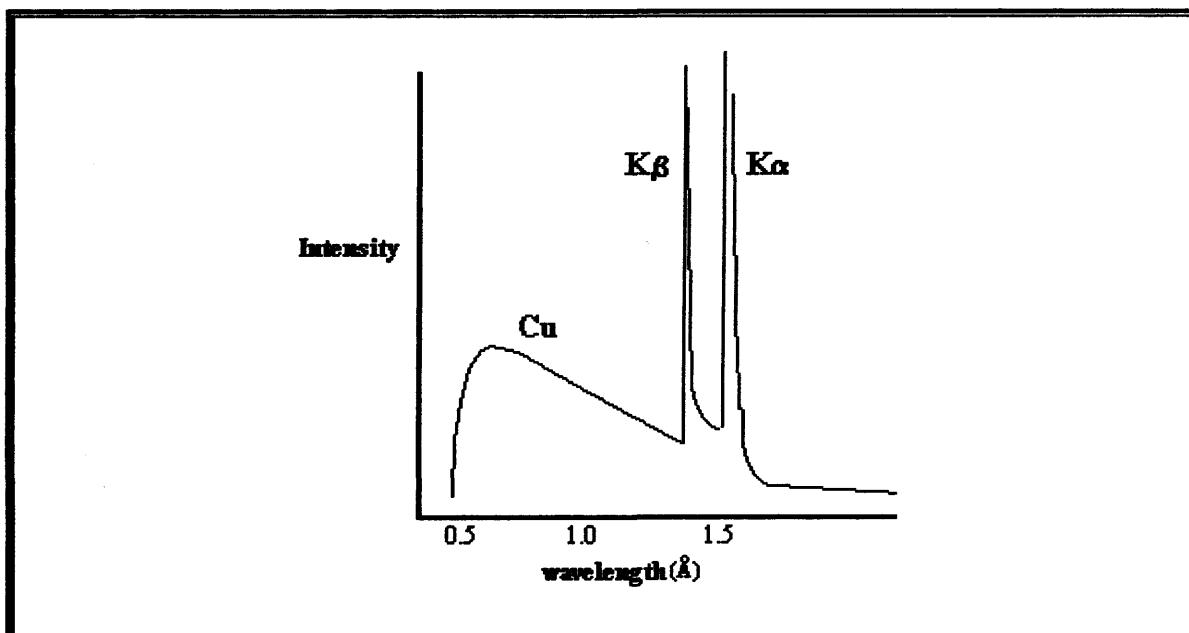
X-rays are electromagnetic radiation with wavelengths of 0.1-1000 Å. X-rays used in protein crystallography are produced by two methods. The first method is to accelerate electrons at high voltage against a metal target as used in home sources, and the second is to use synchrotron radiation emitted by electrons and positrons in high-energy storage rings.

#### 2.4.1.1. Home Sources

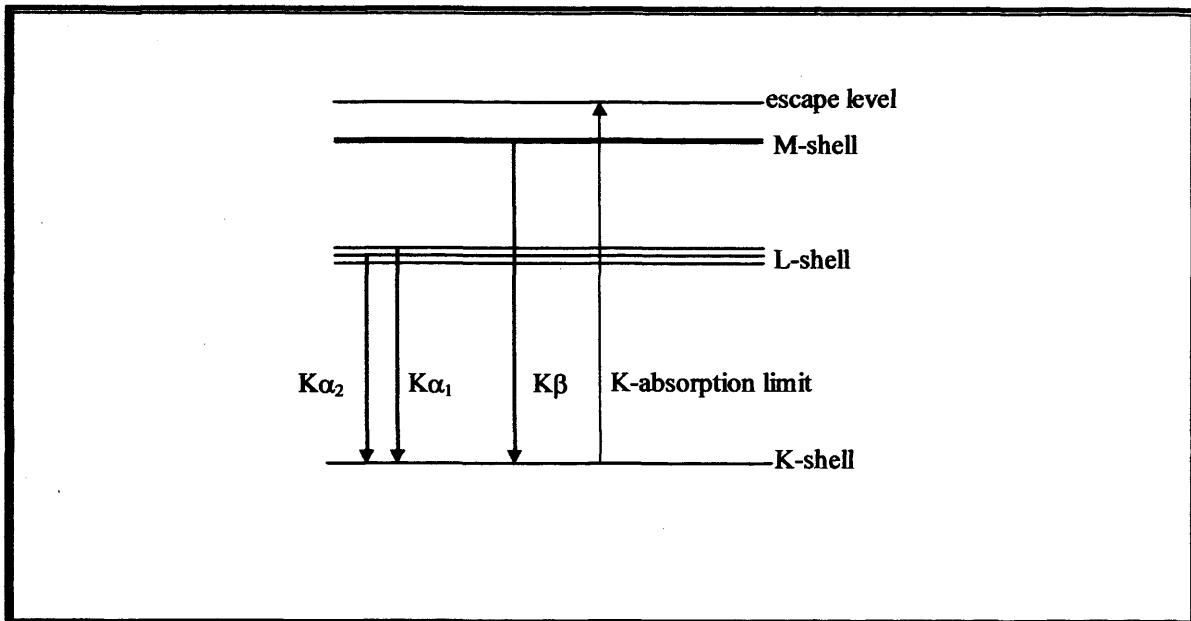
In the case of laboratory x-ray generators, electrons are emitted from a filament-heated cathode and accelerated through a high vacuum. These electrons strike the anode with the emission of x-rays in two different ways: a continuous x-ray spectrum known as *Bremsstrahlung*, which is accompanied by a series of discrete spectral lines characteristic of the target material (Wilson and Prince, 1999) as shown in Figure 2.6. The continuous region is due to the phenomenon that acceleration of charged particles leads to electromagnetic radiation. The sharp peaks in the spectrum are due to electron transitions between inner orbitals in the atoms of the anode material. On impact with the metal, high energy electrons shoot electrons out of the K shells of the respective atoms (in the case of Cu), leaving them in an excited state. After excitation of the K-shell, an electron from the outer shell may fall into the vacancy in the K-shell accompanied by the simultaneous emission of X-rays of a characteristic wavelength: this is called  $K\alpha$  radiation if an L-shell electron fills the K-shell

vacancy and K $\beta$  radiation if a K-shell vacancy is filled by an M-electron (Figure 2.7). In protein crystallography, the most common choice of the anode material is copper, which emits at 1.5418 Å wavelength (CuK $\alpha$ ). This wavelength is a good compromise between maximum-achievable resolution and absorption. Absorption effects occur as only a fraction of the K $\alpha$  radiation generated is emitted due to absorption by the target anode. There is further absorption (~ 80 % of emission) by air between the emitted radiation and the sample.

X-ray diffraction experiments require a monochromator which is produced by use of a nickel absorption filter, which removes the CuK $\beta$  radiation. An increased monochromatic beam can also be achieved by reflecting the X-ray beam from a single crystal (e.g. diamond) set at the appropriate Bragg angle for the characteristic radiation (Freund, 1996). As the majority of collisions between the electrons and the metal anode only result in heat that can eventually damage the anode, the limiting factor in the power at which the X-rays source can operate is the rate at which heat can be removed from the anode. The method of rotating the anode to diminish target heating by distributing the heat over a large area is widely used. In the rotating anode generator the anode is a hollow cylinder of copper (copper has superior heat-conducting properties) that is internally cooled with water and rotated as a focused incident beam of electrons strikes it. Effective heat dispersion increases the maximum input power and thus the emitted intensity of the beam.



**Figure 2.6** Intensity curve for X-rays from a copper target: K $\alpha$ 1=1.5405 Å, K $\alpha$ 2=1.5443 Å (the average value K $\alpha$ =1.5418 Å); K $\beta$ =1.3922 Å (taken from <http://saf.bio.caltech.edu/www/bi170/lecture4.pdf>).



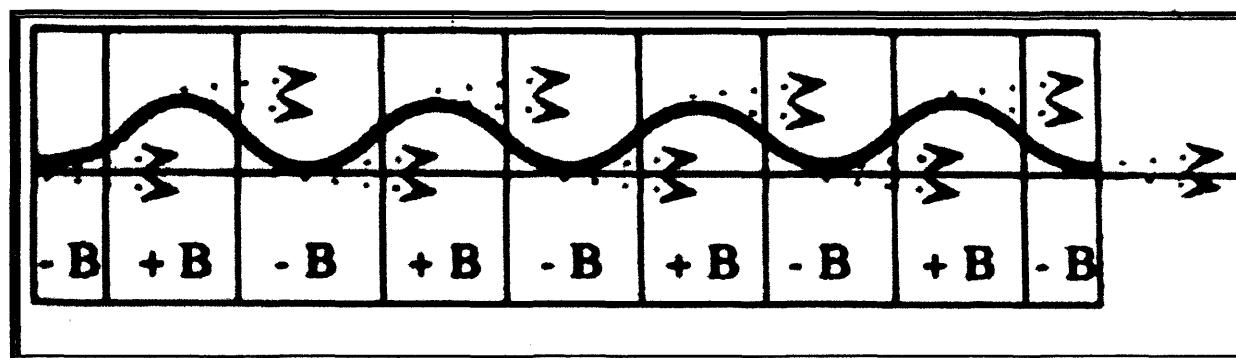
**Figure 2. 7 Schematic representation of the atomic energy levels and transitions causing characteristic X-ray wavelengths (taken from <http://saf.bio.caltech.edu/www/bi170/lecture4.pdf>).**

#### 2.4.1.2. Synchrotron Radiation

Use of a synchrotron radiation source has the advantage of high intensity x-ray beams that allow much shorter data collection experiments. The higher intensity allows the study of larger structures, the use of small crystals and even time resolved structural studies are possible (Helliwell, 2001). In addition, highly parallel rays can be finely collimated to produce a focused beam, which may be essential in the case of very small crystals or crystals with large unit cells resulting in closely spaced reflections as seen in the case of the structural elucidation of the nucleosome (Luger *et al.*, 1997). Another advantage of a synchrotron source is the tunability of the wavelength. This allows the use of short wavelengths, avoiding the effects of absorption and associated radiation damage, which can be severe with CuK $\alpha$  radiation at 1.54 Å. The spectral region of choice is 0.8-1.0 Å for most experiments. The tunability of synchrotron radiation can be further exploited by using X-ray absorption edges (multiple wavelength anomalous dispersion (MAD) phasing experiments) where the anomalous scattering effects provide experimental phase information (Helliwell, 1992).

A synchrotron accelerates charged particles, such as electrons, to speeds close to that of light. When the particle beam changes direction, the electrons are accelerated toward the centre of the ring and emit electromagnetic radiation. To keep the electrons travelling in a nearly circular path, a lattice of bending magnets is set up around the storage ring. As the

electron beam crosses each magnet, the path of beam is altered and synchrotron radiation is emitted. Other types of magnets - wigglers and undulators - can be set up in straight sections of the storage ring to generate radiation of much higher intensity. Both wigglers and undulators are comprised of two sets of small magnets of alternating polarity. As the electron beam passes through the magnetic fields of these insertion devices the electrons are deviated in their path. There are multiple deviations from the path, and the flux of the radiation emitted is increased proportionally as shown in Figure 2.8. The radiation caused by bending magnets and wigglers has a continuous spectrum. In contrast, the radiation produced by use of an undulator has a discontinuous spectrum, as the smaller deviations of the beam in an undulator lead to the phenomenon of interference, which results in a great enhancement of the brilliance at specific wavelengths (Helliwell, 2001).



**Figure 2. 8 Diagram of a multipole wiggler. The electron path is described by the thick black line and the subsequent radiation is described by the dotted lines and arrows with B as the magnetic field strength (Taken from <http://saf.bio.caltech.edu/www/bi170/lecture4.pdf>).**

#### 2.4.2. Detection of X-rays

Experimental measurement of the reflection intensity is accomplished by quantifying the intensity surrounding each reflection. The strength of an X-ray field may be measured either by counting X-ray photons or by integrating energy deposition during a time interval. The choice of detector depends on a large number of factors, which include the experimental situation (home source as opposed to synchrotron source), the sample (whether it is highly radiation sensitive) and the availability and cost of detector. The *detective quantum efficiency* (DQE) is an overall measure of the efficiency and noise performance of a detector (Gruner *et al.*, 1978).

The DQE is defined as:

$$DQE = (S_o/N_o)^2/(S_i/N_i)^2$$

**Equation 2. 1**

where S is the signal, N is the noise, and the subscripts o and i refer to the output and input of the detector respectively.

The DQE is used to measure the degradation caused by the detector in the signal to noise ratio. The ideal detector introduces no additional noise in the detection process therefore preserving the incident to signal ratio *i.e.* DQE = 1. Real detectors always have DQE values < 1 as some noise is always added in the detection process. However, a single value does not characterise the DQE of a system which varies depending on spot size, the length of exposure, the rate of signal accumulation and the x-ray energy. Noise in the detector system will limit the DQE at low dose, while the inability to remove all systematic non-uniformities will limit the high dose behaviour. Measuring the accuracy of the system determines the output noise relative to the signal.

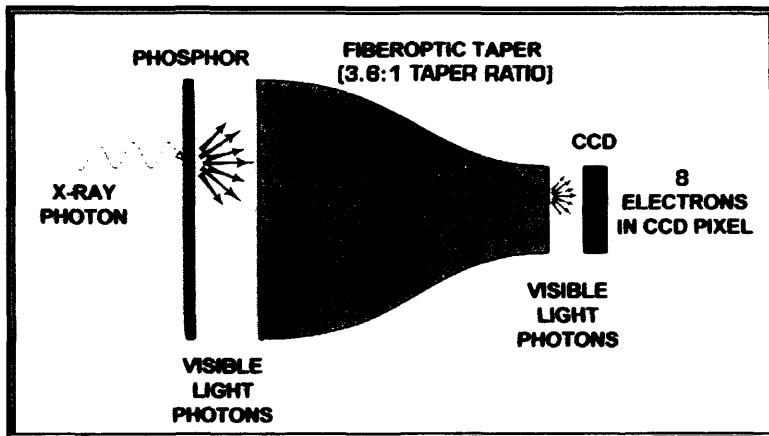
#### **2.4.2.1. Image plate detectors**

These detectors are used in the majority of home laboratories and at many synchrotron radiation beam lines. The major components of the image plate detector are the image plate itself which is fixed to a circular platter mounted normal to the incident X-ray beam, a He-Ne laser/photomultiplier scanning head configuration which may be translated radially along the active surface of the image plate, and an amplifier/ADC (analog-to-digital converter) unit which converts the analogue photomultiplier signal into a digital one ready for storage on hard disk and immediate use by image display software.

Such integrating devices have virtually no upper limits for count rates because they measure the total energy deposited during an integration period (although individual pixels can become saturated if the signal exceeds its storage capacity). Moreover, the “hotter” the source the better an image plate detector performs as the ratio of signal/noise increases with the number of X-ray photons incident on each pixel during the integration period. However, image plate detectors suffer from a major limitation. Reading out the image takes a substantial time (0.5-1.5 min for the *mar345* Image Plate Detector) thus becoming the rate limiting step at synchrotron radiation beam line where the exposure time is as short as 0.5-20 seconds .

#### 2.4.2.2. Charge-coupled devices (CCDs)

Charge-coupled devices have overcome the read-out delay problem (Westbrook & Naday, 1997). These high resolution devices offer excellent data quality and a rapid readout system in the order of a few seconds. Another big advantage of CCDs compared to image plate detectors is the absence of any mechanical parts in the detector box. The three main components of the Quantum CCD detector are phosphor screen (to convert X-rays to light), fiber-optic taper (to reduce the light image down to the size of the CCD chip), and 1 x 1 inch semiconductor chip to detect the light image as an electric charge image (Figure 2.9). The electric charge image is read out of the CCD chip and digitised, then fed into a computer. After geometric and intensity corrections are applied, the resulting data quality is similar to data from other types of X-ray detectors and can be processed by standard data reduction software.



**Figure 2. 9 Principle of a CCD detector.**

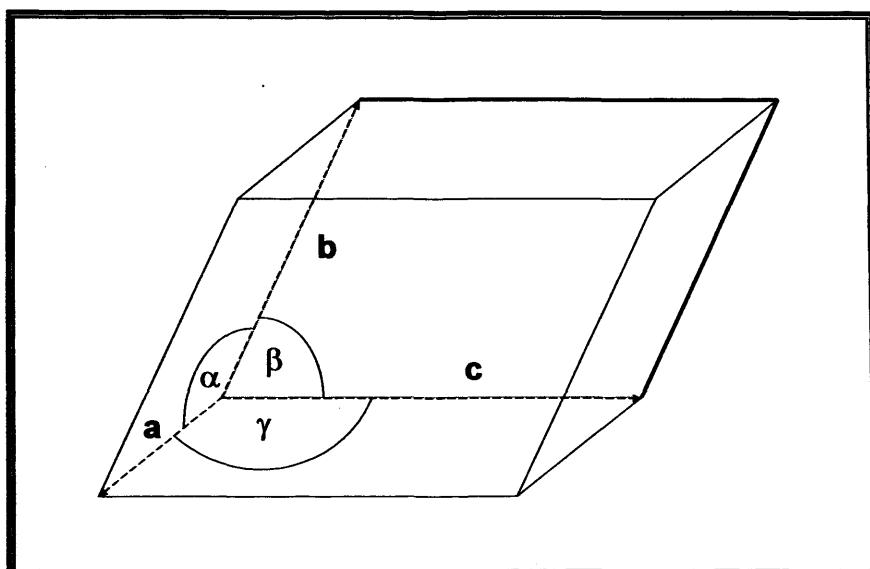
## 2.5. X-ray Diffraction

Visualisation of the structure of protein molecules is one of the fundamental prerequisites in gaining a full understanding of their physical and chemical properties. As atomic bonds are in the order of Å in size, it is not possible to use electromagnetic radiation in the visible region to visually observe a protein molecule. X-ray wavelengths are of the same order as atomic bonds but in order to visualise a magnified image of the atomic bonds, X-ray lenses would need to be developed. Unfortunately, X-ray lenses have never been engineered despite countless efforts. The discovery that protein crystals diffract X-rays giving a regular diffraction pattern opened the possibility of protein structure visualisation (Bernal and

Crowfoot, 1934). Using the Fourier theorem whereby any periodical function can be described by the summation of an infinite set of harmonic functions, i.e. a protein structure in a crystal lattice is described by its diffracted X-rays, allows the elucidation of large macromolecular structures.

### 2.5.1. Crystals

Crystals are made up of a repeating unit known as the unit cell which has basis vectors **a**, **b** and **c** and angles  $\alpha$ ,  $\beta$  and  $\gamma$  between them (Figure 2.10)



**Figure 2. 10 One unit cell with axes **a**, **b** and **c** and angles  $\alpha$ ,  $\beta$  and  $\gamma$  between the axes.**

The large number of unit cells in a crystal are arranged in a specific ordered fashion with the origins of the cells forming a lattice. The planes in a unit cell are named with the plane containing the unit cell vectors **a** and **b** called (001), **b** and **c** (100), **a** and **c** (100). Many more sets of parallel and equidistant planes can be drawn through the lattice points. The naming of the lattice point is based on the number of divisions the unit cell vectors are divided i.e.  $hkl$  of 15,5,30 divides the **a** vector 15 times, **b** vector 5 times and the **c** vector 30 times. Furthermore the planes are not limited to a single unit cell but extend throughout the crystal. The choice of the unit cell is not unique and therefore guidelines have been established for choosing the standard basis vectors and the origin.

A lattice is defined as an array of equivalent points in one, two or three dimensions. The environment of an atom placed on any one of these lattice points would be identical to that placed on any other lattice point. Therefore the lattice locates equivalent positions and

shows the translational symmetry. Primitive cells have only one lattice point per cell whereas a centred cell has an additional lattice point on a particular face. If the extra lattice point is body centred it is called an I cell (body centred, from German Innenzentriert) or if there is an extra lattice point at each of the cell faces it is known as an F type cell (face centred, from German Flächenzentriert).

### 2.5.2. Symmetry

Within a unit cell, the organisation of macromolecular molecules can exhibit symmetry. A symmetry operation is an operation that once applied results in a form that is indistinguishable from the original. Therefore repetition along axes **a**, **b** and **c** represent translational symmetry. Rotational symmetry can also exist but for 3D crystals is restricted to rotational angles of 60, 90, 120, 180 and 360°. Combination of a rotational and translational symmetry element is known as a screw axis. In a two-fold screw axis, the translation is over half of the unit cell length in the direction of the axis combined with a 180 degree rotation along the axis. It can be proven that 230 different symmetry combinations are possible that allow 3D crystals, however, as biological molecules are made up of non-symmetrical building blocks, mirror planes and inversion centres are not allowed for these molecules and therefore only 65 enantiomeric (enantiomeric means the molecule is not superimposable on its mirror image) space groups are possible.

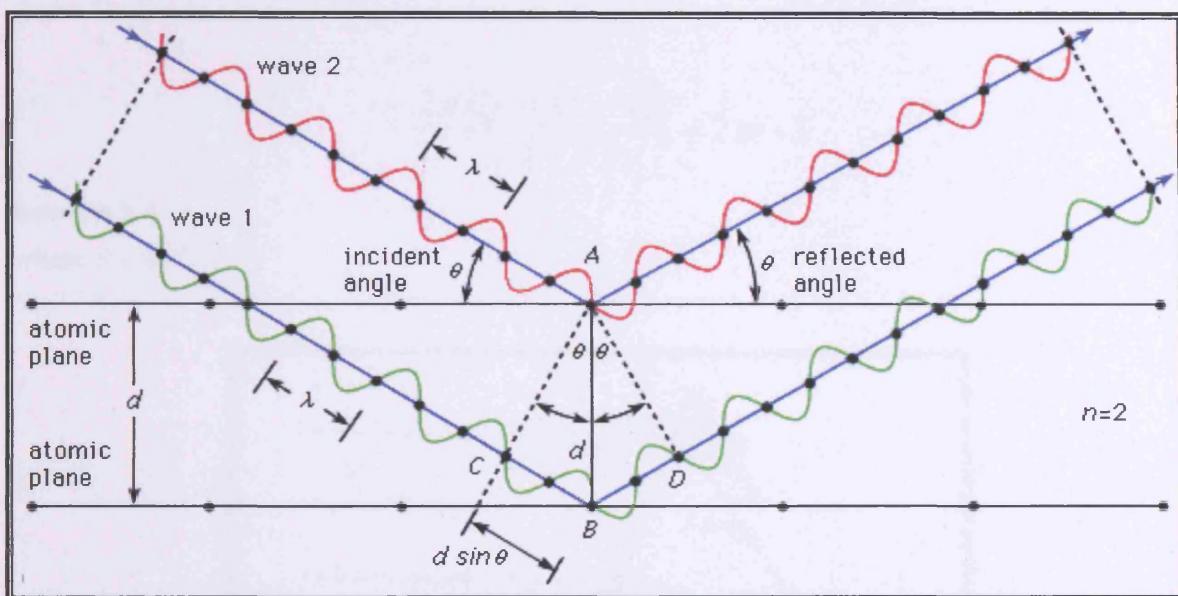
## 2.6. Basic Diffraction physics

### 2.6.1. Diffraction by one electron

Upon exposure to X-rays, a large number of atoms (electrons) in the protein crystal scatter the incident X-ray beam. Radiation scattered by materials consists of two parts, a coherent part known as Thomson scattering and an incoherent part known as Compton scattering. The Thomson scattering gives diffraction in a co-operative manner resulting in the diffraction image while the incoherent scattering leads to destructive interference and is ignored in protein X-ray crystallography.

### 2.6.2. Diffraction of X-rays from a lattice

In 1913, WL Bragg showed that x-ray diffraction behaved in the same way as light reflecting from a mirror plane with the incident and the reflected X-rays lying on one plane with the angle of incidence and angle of reflection equal (Figure 2.11; Bragg and Bragg, 1913). Each reflected beam can be regarded as the superimposition of reflections from sets of parallel planes passing through lattice points of the crystal. In order to refer to the 3D-lattice planes, the Miller notation ( $hkl$ ) is used.



**Figure 2. 11 Diagram showing Bragg's law with the incident beam angle equal to the reflected beam angle**  
(taken from <http://www.britannica.com/nobel/cap/obragg1001a4.html>)

From this observation the Bragg's equation can be determined as:

$$d_{hkl} = \frac{\lambda}{2 \sin \theta}$$

**Equation 2. 2 Bragg's Equation**

### 2.6.3. Scattering by a system of two electrons

In Figure 2.12, a system of two electrons which scatter independently of each other is given with the origin at electron 1 and electron 2 at position  $\mathbf{r}$ . The direction of the x-ray

beam is along the vector  $s^o$ . They scatter the incident beam in a direction given by the vector  $s$ . The length of the vectors is  $1/\lambda$ . The amplitudes of the scattered beams 1 and 2 are equal, but there is a phase difference between the diffracted x-ray beams due to the difference in position between electron 1 and electron 2. The path difference can be expressed as:

$$p + q = \lambda[r \cdot (s - s^o)]$$

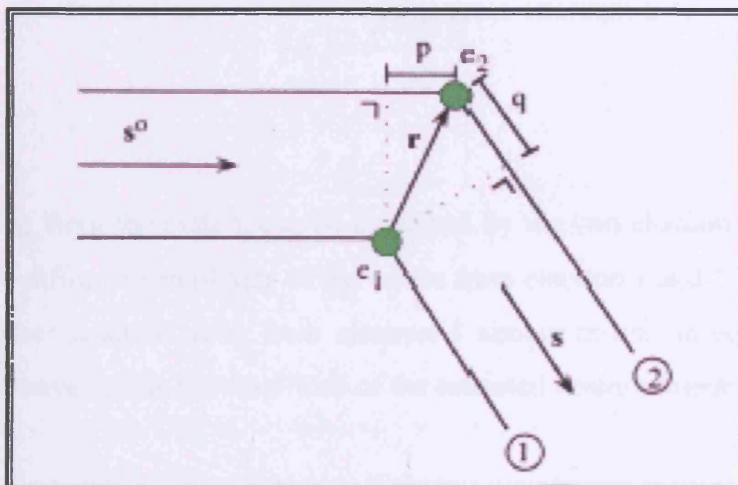
**Equation 2. 3**

Wave 2 has a phase difference relative to wave 1 and in relation to wave 1 its phase angle is:

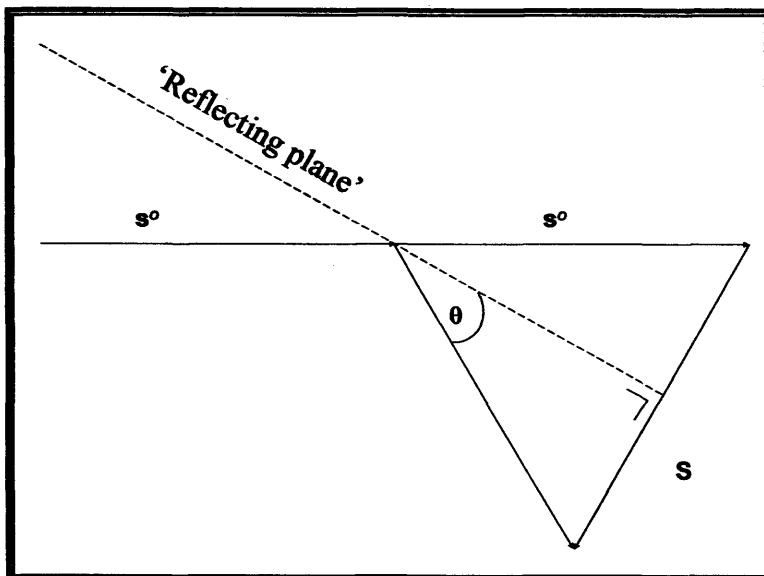
$$\frac{-2\pi\lambda[r \cdot (s^o - s)]}{\lambda} = 2\pi r \cdot S$$

**Equation 2. 4**

where  $S = s - s^o$ .



**Figure 2. 12 An x-ray beam diffracted by two electrons (Adapted from Drenth, 1999)**



**Figure 2.13 The reflecting plane.** A plane that makes equal angles with  $s$  and  $s^\circ$  can be seen as a mirror plane that reflects the incident beam (Adapted from Drenth, 1999).

From Figure 2.12, the vector  $S$  is perpendicular to an imaginary plane that reflects the incident beam at an angle  $\theta$  and the length of  $S$  is given by:

$$|S| = \frac{2 \sin \theta}{\lambda}$$

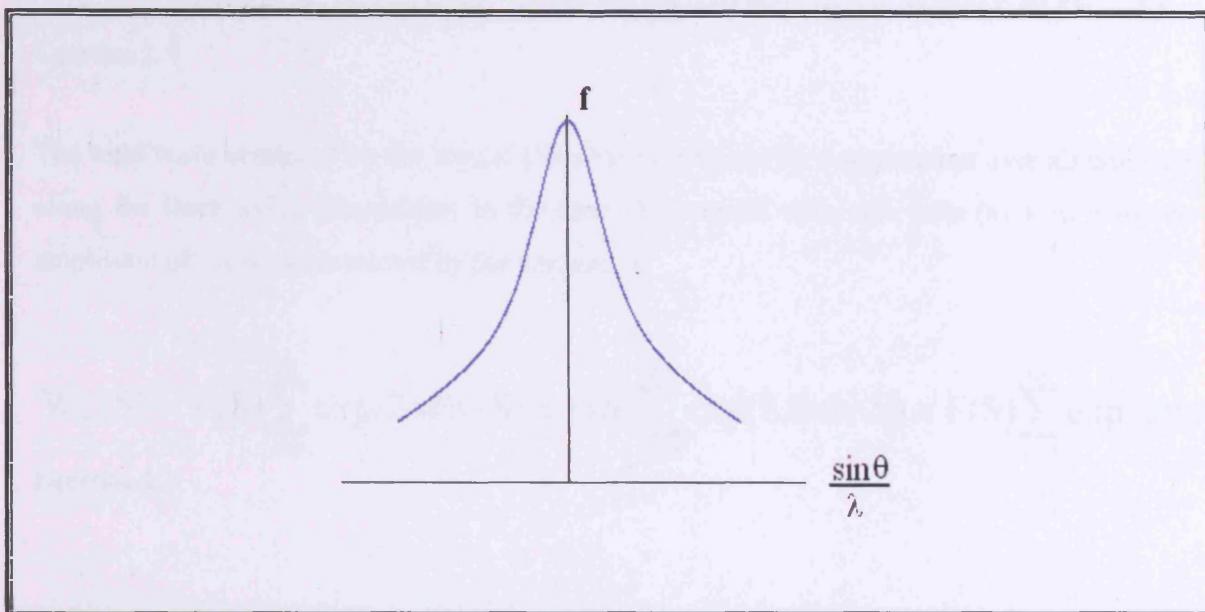
**Equation 2.5**

The total scattering from the system can be described by the two electron system as  $1 + 1 \times \exp(2\pi i r \cdot S)$  if the diffracted amplitude of the waves from electron 1 and 2 is set to 1. Moving the origin to another position away from electron 1 simply means an equal change in the phase angle of all waves while the amplitude of the reflected beams remains unchanged.

#### 2.6.4. Scattering by an atom

The scattering amplitude of an atom is called the atomic scattering factor  $f$ . It is possible from the known electron density functions for electronic shells to evaluate the variation in scattered amplitude with incident wavelength and scattering angle. A typical profile of an atomic scattering factor as a function of  $(\sin \theta)/\lambda$  is shown in Figure 2.14. However, as atoms in crystalline solids are not stationary their natural vibration must be accounted for by a temperature factor  $B$ . The thermal motion of atoms cause a decrease in the scattered intensity by a factor  $\exp[-(B \sin^2 \theta)/\lambda^2]$ , which is known as the Debye-Waller factor. Here, the temperature factor  $B$  equals  $8\pi^2 U$ , and  $U$  is the mean-square amplitude of

vibration of the atom from its equilibrium position. A large temperature factor implies a rapid fall off in intensity with angle  $\theta$  and hence limits the resolution of the structure. This is not the same for all atoms of the same element within a protein structure as they generally have different temperature factors depending on their atomic interactions and location. A mean value of  $B$  for all atoms in the crystal structure can be assessed from a Wilson plot.



**Figure 2. 14** A profile of an atomic scattering factor as a function of  $(\sin\theta)/\lambda$ . At  $\theta=0$  the value of  $f$  is equal to the atomic number  $Z$ . The variation of intensity with angle is called the atomic scattering  $f(\theta)$ .

### 2.6.5. Scattering by a crystal

Due to the large number of atoms and thus electrons in a unit cell, there is a large amount of interference between the scattered waves diffracted by the electrons. The structure factor can be regarded as the sum of the scattering by the atoms in the cell dependant on their positions and the phase differences between the scattered waves. For  $n$  atoms in a single unit cell then

$$\mathbf{F}(\mathbf{S}) = \sum_{j=1}^n f_j \exp(2\pi i \mathbf{r}_j \cdot \mathbf{S})$$

**Equation 2. 6**

where  $\mathbf{S}$  is a vector perpendicular to the plane reflecting the incident beam at an angle  $\theta$ ; the length of  $\mathbf{S}$  is given by  $|\mathbf{S}| = 2\sin\theta/\lambda$  (Eq. 2.5). Atom  $j$  is at position  $\mathbf{r}_j$  relative to the origin. When considering a crystal, with a unit cell with its origin at position  $t\cdot\mathbf{a} + u\cdot\mathbf{b} + v\cdot\mathbf{c}$ , in which  $t, u$  and  $v$  are whole numbers and  $\mathbf{a}, \mathbf{b}$  and  $\mathbf{c}$  are the crystal axes, the scattering is:

$$F(\mathbf{S}) = \exp[2\pi it\mathbf{a} \cdot \mathbf{S}] \times \exp[2\pi iu\mathbf{b} \cdot \mathbf{S}] \times \exp[2\pi iv\mathbf{c} \cdot \mathbf{S}]$$

**Equation 2. 7**

The total wave scattered by the crystal ( $W_{cr}(\mathbf{S})$ ) is obtained by a summation over all unit cells along the three axial dimensions. In the case of a crystal with unit cells ( $n_1 \times n_2 \times n_3$ ) the amplitude of the wave scattered by the crystals is:

$$W_{cr}(\mathbf{S}) = F(\mathbf{S}) \sum_{t=0}^{n_1} \exp(2\pi it\mathbf{a} \cdot \mathbf{S}) \times F(\mathbf{S}) \sum_{u=0}^{n_2} \exp(2\pi iu\mathbf{b} \cdot \mathbf{S}) \times F(\mathbf{S}) \sum_{v=0}^{n_3} \exp(2\pi iv\mathbf{c} \cdot \mathbf{S})$$

**Equation 2. 8**

For a 3D array of unit cells which is defined by vectors  $\mathbf{a}, \mathbf{b}$  and  $\mathbf{c}$ , scattering will only be observed when the phase difference between the waves scattered by successive unit cells is equal to an integral multiple of  $2\pi$ , i.e.

$$2\pi\mathbf{a} \cdot \mathbf{S} = 2\pi h$$

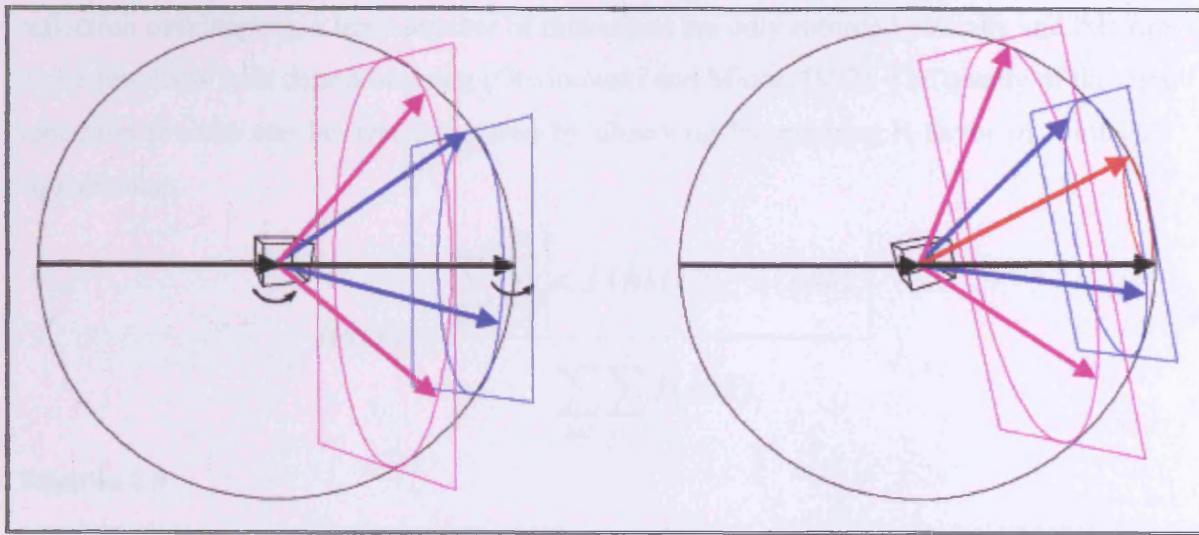
i.e.  $\mathbf{a} \cdot \mathbf{S} = h$ ,  $\mathbf{b} \cdot \mathbf{S} = k$ ,  $\mathbf{c} \cdot \mathbf{S} = l$  where  $h, k$  and  $l$  is an integer, either positive, negative or zero. These are known as the Laue conditions and if they are fulfilled, all of the unit cells diffract in phase and the amplitude of the wave scattered by the crystal is proportional to the amplitude of the structure factor  $F$

## 2.6.6. X-ray Diffraction Data Collection and Processing

Strategy of data collection is guided by unit cell dimensions and also internal symmetry within a unit cell (Rossman, 2001). If there is high symmetry, certain sets of crystal orientations produce the exact same reflections reducing the number of crystal orientations required to obtain all unique reflections. All reciprocal lattices possess a symmetry element called a centre of symmetry or point of inversion at the origin with the equivalence of  $I_{hkl}$  and

$L_{h+k+l}$  known as Friedel's law. This means that half of the diffracted reflections are redundant and data collection that covers  $180^\circ$  about any reciprocal lattice axis will capture all unique reflections.

The Ewald construction, a geometric equivalent of the Bragg law, is shown in Figure 2.15 (Drenth, 1999). The radiation of wavelength  $\lambda$  is represented by a circle of radius  $1/\lambda$  known as the Ewald sphere centred on the crystal. The reciprocal lattice is represented with its origin at the point on the Ewald sphere where the direct beam leaves it. Whenever a reciprocal-lattice point with indices  $(hkl)$  lies exactly on the Ewald sphere, the Bragg equation is satisfied and a diffraction maximum occurs. For any particular crystal orientation, only a few reciprocal lattice points are in a diffracting position. The contribution of an atom is greatest to the reflections whose indices correspond to lattice planes that intersect that atom, so a specific atom contributes to some reflections strongly and to some weakly or not at all. To observe a larger number of reflections, one has to rotate the crystal (thus rotating the reciprocal lattice) in order to move more reciprocal lattice points through the Ewald sphere. As a reciprocal lattice point moves through the Ewald sphere, diffraction is initially weak, reaches a peak when the centre of the point lies exactly on the sphere and weakens again until it is eventually extinguished. In general, the crystal is rotated  $0.2 - 2^\circ$  to bring a large number of reciprocal lattice points into contact with the Ewald sphere surface and a large number of reflections are measured on a single image.



**Figure 2. 15** The Ewald construction. a) a crystal oriented so that planes in the reciprocal lattice are perpendicular to the X-ray beam. The planes, intersecting circles and sample diffracted beams are shown in a different colour for each plane. b) When the crystal is rotated, the reciprocal lattice planes rotate as well resulting in a new diffraction beam (red) appearing from behind the beam stop (taken from [http://www-structmed.cimr.cam.ac.uk/Course/Basic\\_diffraction/Diffraction.html](http://www-structmed.cimr.cam.ac.uk/Course/Basic_diffraction/Diffraction.html)).

The first step in the processing of the raw images is the indexing: the interpretation of the pattern of reflections and allocation of  $hkl$  indexes to each observed diffraction spot. Determination of the space group is carried out by analysis of one or more diffraction images. In the DENZO program (Otwinowski & Minor, 1997), the crystal and detector orientation parameters are established and the unit cell parameters are determined according to International Table conventions and all of these parameters are then refined by a least-squares procedure, and the place of each reflection in three-dimensional reciprocal space is determined. A set of indices ( $hkl$ ) is thus assigned to each reflection.

The next step, data processing, is the conversion of observed intensities ( $I_o$ ) to observed structure amplitudes  $|F_o|$  and has several components. The data reduction process includes the merging and averaging of identical and symmetry equivalent measurements in order to produce a corrected and scaled set of data. This calculation affords a numerical measure of the agreement among multiple determinations of equivalent reflections to allow correction of errors in data measurements and gives an indication of the quality of the data. Obtaining this consistency in measurements is carried out by scaling the intensities from multiple blocks of data to give identical reflections, identical intensity values. Post-refinement involves the estimation of the partiality of each reflection that is not fully recorded on a diffraction image and can be used as a correction factor to convert a measured intensity of a partial reflection to an estimate of the total reflections intensity (Rossman, 1979). In the case of large unit cells where smaller oscillation angles are employed to reduce

reflection overlapping, a large number of reflections are only recorded partially and this can cause problems with data processing (Otwinowski and Minor, 1997). The quality of the data reduction process can be best monitored by observing the merging R-factor of symmetry equivalents:

$$R_{\text{sym}} = \frac{\sum_{hkl}^N \sum_{i=1}^N \left| \langle I(hkl) \rangle - I(hkl)_i \right|}{\sum_{hkl}^N \sum_{i=1}^N I(hkl)_i}$$

**Equation 2. 9**

where  $I(hkl)$  is the  $i$ th measurement of reflection  $hkl$  and  $\langle I(hkl) \rangle$  is the mean value of the N equivalent reflections.

### 2.6.7. Calculation of the electron density

The electron density in a unit cell is the reverse Fourier transform of the diffraction pattern obtained. The summation is performed over all values of  $h$ ,  $k$  &  $l$  i.e. all of the reflections in the diffraction pattern contribute to the electron density at point xyz.

If  $x$ ,  $y$  and  $z$  are fractional coordinates in the unit cell,  $\mathbf{r} \cdot \mathbf{S} = (\mathbf{a} \cdot \mathbf{x} + \mathbf{b} \cdot \mathbf{y} + \mathbf{c} \cdot \mathbf{z}) \cdot \mathbf{S} = \mathbf{a} \cdot \mathbf{S} \cdot \mathbf{x} + \mathbf{b} \cdot \mathbf{S} \cdot \mathbf{y} + \mathbf{c} \cdot \mathbf{S} \cdot \mathbf{z} = hx + ky + lz$ , and an expression of the electron density is:

$$\rho(x, y, z) = \frac{1}{V} \sum_{hkl} F(hkl) \cdot \exp[-2\pi i(hx + ky + lz)]$$

**Equation 2. 10**

The electron density at a point xyz is equal to the summation of the correct phases and amplitudes multiplied by a phase shift appropriate to each geometrical position in the image relative to the unit cell origin.  $V$  is the volume of the cell. Instead of expressing  $F(S)$  as a summation over the atoms (Equation 2.5), it can be expressed as an integration over the electron density in the unit cell:

$$F(hkl) = V \int_{x=0}^1 \int_{y=0}^1 \int_{z=0}^1 \rho(xyz) \exp[2\pi i(hx + ky + lz)] dx dy dz$$

**Equation 2. 11**

Because a vector  $F(hkl)$  can be expressed as a vector in an Argand diagram with an amplitude  $|F(hkl)|$  and a phase angle  $\alpha(hkl)$  then:

$$F(hkl) = |F(hkl)| \exp[i\alpha(hkl)]$$

**Equation 2. 12**

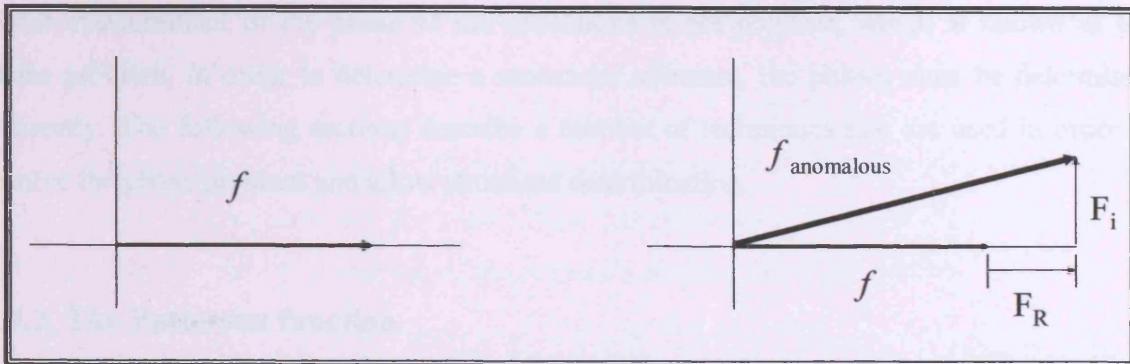
and

$$\rho(xyz) = \left(\frac{1}{V}\right) \sum_h \sum_k \sum_l |F(hkl)| \exp[-2\pi i(hx + ky + lz) + i\alpha(hkl)]$$

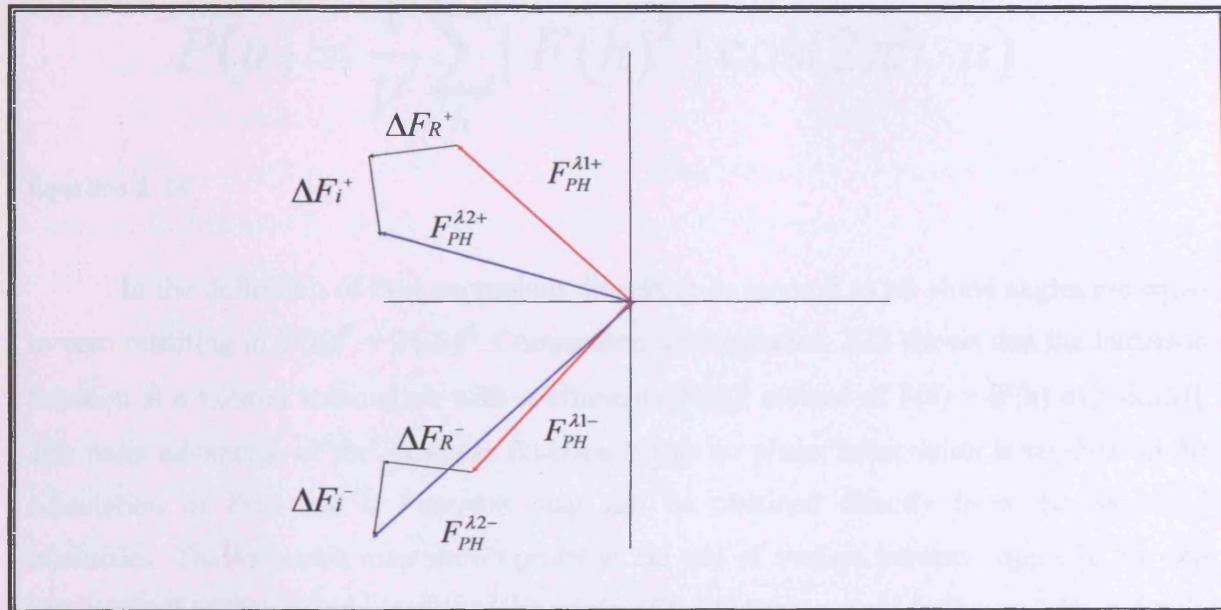
**Equation 2. 13**

Thus the electron density distribution can be calculated for the unit cell provided that the amplitude and the phase angle are known. The amplitude can be calculated from the intensities of the diffracted rays however the phase angle is more difficult to determine and various methods to determine protein phases are discussed in Section 2.7.

From Equation 2.11 it can be seen that the reflections  $hkl$  and  $-h-k-l$  have the same value for their structure factor amplitudes and for their intensities, but have opposite values for their phase angles. The two reflections  $hkl$  and  $-h-k-l$  are known as Friedel or Bijvoet pairs. In the case of anomalous dispersion, the Friedel pairs are no longer equal with differences in their structure factor amplitudes and also their phase angles are no longer equal and opposite to one another. This is caused by the imaginary component of the anomalous scattering. An anomalous structure factor is made up of three components as shown in Figure 2.16 and 2.17 with a real  $F_R$  and imaginary  $F_i$  correction component that must be applied to the structure factor.



**Figure 2.16** The atomic scattering factor as a vector in the Argand diagram. a) No anomalous scattering  
b) Anomalous scattering where there is a anomalous scattering with a real anomalous contribution ( $F_R$ ) and an imaginary contribution ( $F_i$ ).



**Figure 2.17** Argand diagram showing relationship between structure factors of direct ( $F=F(hkl)$ ) and inverse ( $-F=F(-h,-k,-l)$ ) reflections from an anomalously scattering derivative crystal.

## 2.7. Structure Determination

### 2.7.1. The Phase problem

From equation 2.13 the calculation of the electron density is based on the identification of the intensity of diffraction as well as the phases associated with the reflections. Determination of the intensity is easily carried out by measurement. However,

direct measurement of the phase of the reflections is not possible, which is known as the phase problem. In order to determine a molecular structure, the phases must be determined indirectly. The following sections describe a number of techniques that are used in order to resolve the phase problem and allow structural determination.

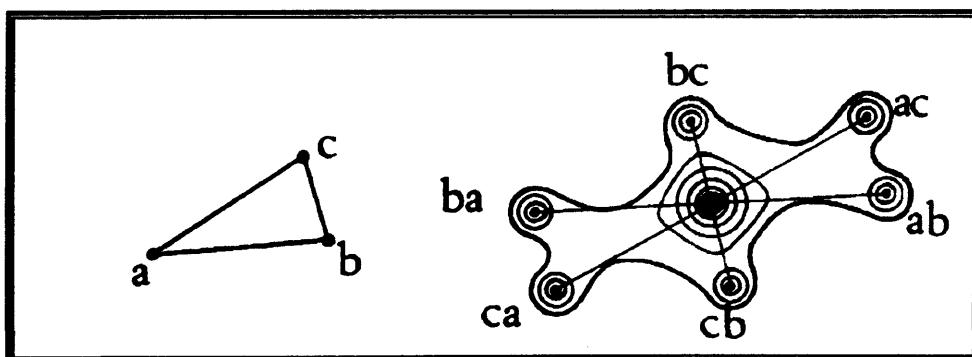
### 2.7.2. The Patterson function

In 1934, A.L. Patterson outlined a technique which allowed the determination of the atomic positions of small molecules without knowing the phase angles (Patterson, 1934). This involves the determination of the Patterson function,  $P(u,v,w) = P(u)$ :

$$P(u) = \frac{1}{V} \sum_h |F(h)|^2 \cos(2\pi h \cdot u)$$

**Equation 2. 14**

In the definition of  $P(u)$  anomalous dispersion is ignored as all phase angles are equal to zero resulting in  $|F(h)|^2 = |F(-h)|^2$ . Comparison with equation 2.13 shows that the Patterson function is a Fourier summation with coefficients  $|F(h)|^2$  instead of  $F(h) = |F(h)| \exp[i\alpha(h)]$ . The main advantage of the Patterson function is that no phase information is required in the calculation of  $P(u)$  and a Patterson map can be obtained directly from the measured intensities. The Patterson map shows peaks at the end of vectors between atoms in the unit cell, with all of the vectors starting at the origin of the Patterson map. In Figure 2.18, a simple molecule is shown containing only three atoms (a, b and c). The Patterson function of these atoms is made up of six peaks which equate to the end of the vector between each atom.



**Figure 2. 18 Structure and Patterson function of a simple molecule. Simple object consisting of three**

atoms and the idealised Patterson function of this object with the letters identifying the corresponding vectors (Taken from Taylor, 1967).

The number of peaks in a Patterson function increases with the number of atoms and in the case of  $n$  atoms there are  $n^2-n$  non origin Patterson peaks (Figure 2.18) and with large amounts of atoms, the interpretation of the Patterson map becomes impossible.

### 2.7.3. Molecular Replacement

Due to the number of structures presently available, this technique has become the most used method for structure determination in recent years. If the structure of a protein similar to that of the protein of interest is known, then molecular replacement may allow structure solution. This technique uses the phases calculated from a previously solved structure placed in the new crystal lattice on the approximate position of the molecule of interest in order to gain initial phase estimates for the new structure. There are a number of pre-requisites in order to successfully solve a structure using molecular replacement: the observed data should be as complete as possible and the homologous protein requires at least 20 % sequence identity.

In theory, it should be possible to calculate the structure factors  $|F_{\text{calc}}|$  of the model and compare with the  $|F_{\text{obs}}|$ . In order to calculate initial model phases, it is necessary to determine the position and orientation of the model structure to correspond with the unknown structure. However, this determination of the model within the unit cell is a six-parameter problem. The search can be divided into an initial three dimensional angular search using a rotation function and a subsequent positional search using a translation function (Figure 2.19). Using this procedure, model phases can be calculated and combined with the observed amplitudes to give an approximate Fourier synthesis of the structure's electron density map which is used as a starting point for model rebuilding and refinement. If the phasing model and the new protein are isomorphous as is the case with ligand soaks, then the phases of the free protein can be used directly to compute  $\rho(x, y, z)$  from native intensities of the new structure (e.g. protein-ligand complex).

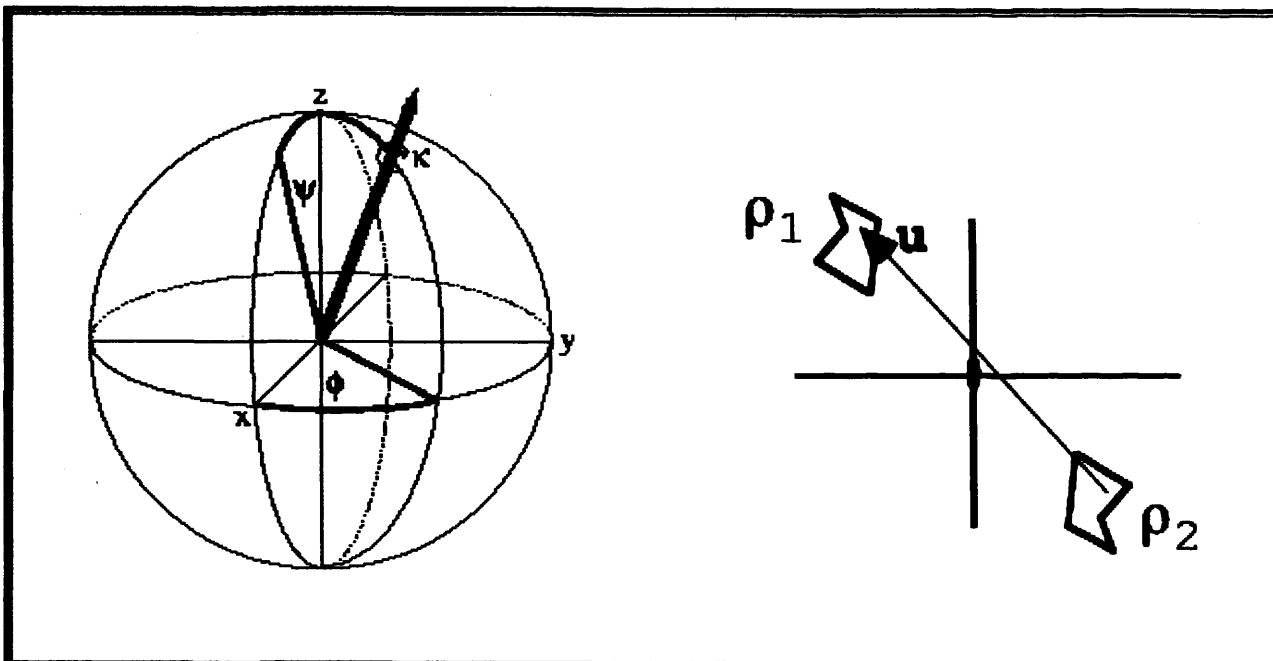
### 2.7.3.1. Rotation Function

Using the Patterson function it is possible to search for the best orientation of the search molecule independently of location. For intramolecular vectors, the Patterson function only differs with the orientation of the molecule therefore; it is possible to determine the best orientation of the model by correlation between theoretical and observed Patterson maps. If the search model is orientated correctly it should give rise to similar Patterson maps. During a rotational search, large correlation values of the model Patterson function  $\rho_{\text{model}}(u, v, w)$  to the observed data Patterson function likely correspond to a correct orientation of the model. A function to evaluate this agreement was originally defined by Rossmann and Blow (1962) as

$$R(C) = \int_U P_1(x)P_2(Cx)dV$$

#### Equation 2. 15

Where  $P_1$  and  $P_2$  are the two Patterson functions,  $C$  is a matrix defining a rotation of the coordinate system of  $P_2$  with respect to  $P_1$ , and  $U$  is a volume of integration, usually spherical, centred at the origin. It is usually over a radius of integration at about 75 - 80% of the radius of the molecule, in the case of an approximately spherical molecule. This value is chosen to minimise the number of inter-molecular cross-vectors. A maximum in the rotation function  $R(C)$  indicates a potentially correct orientation for the search molecule in the target crystal.



**Figure 2.19** Schematic diagram describing the orientation and translation operations involved in molecular replacement. a) the orientation of the molecule is given by defining an axis from the centre of the molecule to the points  $x$ ,  $y$  and  $z$  respectively and angles around this axis  $\phi$ ,  $\psi$ , and  $\kappa$ . b) the translation of the molecule from an origin point to the search site is given by the vector  $u$  (taken from <http://eagle.mmid.med.ualberta.ca/tutorials/MR/MR.html>)

### 2.7.3.2. Translation function

The translation function assesses possible models by moving properly oriented molecules within the unit cell. For each position of the test molecule, say  $x$ , one can calculate the intensities for the resultant model and see how they agree with the  $I_{obs}(\mathbf{h})$ . Crowther and Blow (1967) initially defined the translation function as

$$T(\mathbf{x}) = \sum_{\mathbf{h}} F_c^2(\mathbf{h}, \mathbf{x}) I_{obs}(\mathbf{h})$$

**Equation 2.16**

For given rotation  $R(\alpha, \beta, \gamma)$  and translation  $x$ , the calculated structure factor is given by

$$F_H(\alpha, \beta, \gamma, \mathbf{x}) = \sum_s f(\mathbf{H}\mathbf{M}_s \mathbf{D}R(\alpha, \beta, \gamma) \mathbf{O}) \exp(2\pi i \mathbf{H}t_s) \exp(2\pi i \mathbf{H}\mathbf{M}_s \mathbf{x})$$

**Equation 2.17**

where  $\mathbf{M}_s$  and  $\mathbf{t}_s$  are the transformation matrix and translation vector corresponding to the  $s$ th symmetry operation,  $\mathbf{D}$  is a matrix to transform orthogonal into fractional coordinates in the crystal cell and  $\mathbf{O}$  is a matrix to transform fractional into orthogonal coordinates in the model cell.

The most commonly used translation search indicator is based on the correlation between the observed and the calculated Patterson map (Tong, 1993). Further agreement between the calculated structure factor amplitudes of the model in a given trial location and the actual measured amplitudes (correlation coefficient) and R factor also serve to evaluate the possible translation search solutions.

#### 2.7.4. Isomorphous Replacement

Utilisation of isomorphous replacement to solve the phase problem requires diffraction data from a native crystal and data from either a single or multiple heavy atom derivatised crystals. Of critical importance is that the heavy metal atoms are introduced to the protein without changing the overall structure of the protein, crystal packing or unit cell size allowing accurate measurements of the reflection intensity changes between the data sets. Isomorphous replacement is based on the premise that each structure factor  $F_{hkl}$  for a certain reflection is a summation of all atomic scattering factors:

$$F_{hkl} = \sum_{j=1} f_j \exp 2\pi i(hx_j + ky_j + lz_j)$$

#### Equation 2. 18

Heavy metals are used as they diffract the x-ray beams strongly thus giving a detectable change in the diffraction pattern that is large enough to be measured accurately. These differences can then be used to initially estimate protein phases using the equation:

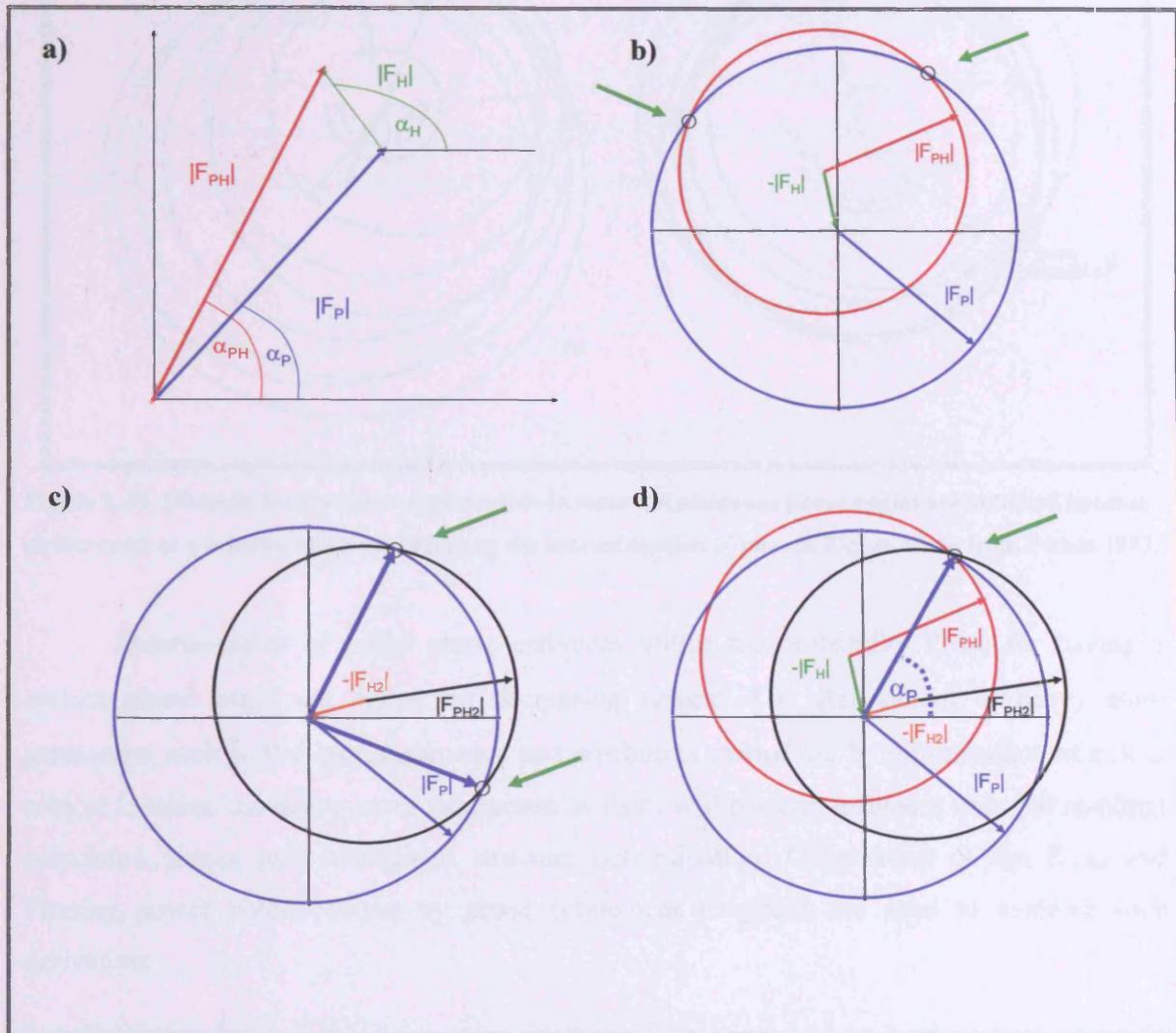
$$F_{PH} = F_P + F_H$$

#### Equation 2. 19

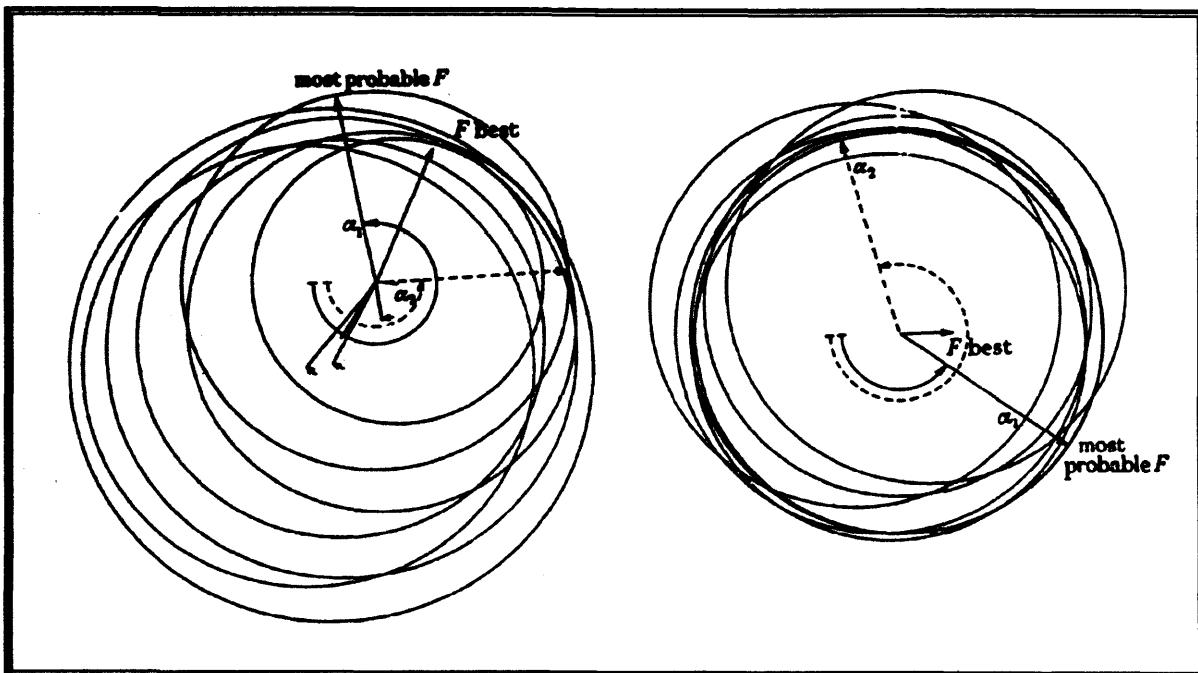
where  $F_{PH}$  is the structure factor of a reflection for the heavy-atom-bound protein and  $F_P$  is the contribution for the protein alone.

Subtraction of the amplitude  $|F_P|$  in the native data from the corresponding reflection of amplitude  $|F_{PH}|$  from the heavy atom derivative will allow determination of the diffraction pattern of the heavy atom alone in the unit cell ( $|F_H|$ ). The position of the heavy atoms in the unit cell is relatively easy to determine by use of Patterson maps using the estimated  $|F_H|$ . From this it is possible to identify a set of atomic positions for heavy atoms that have been introduced into the crystal and determination of the phases of  $\alpha_H$  by calculation.

The corresponding phase angle,  $\alpha_P$ , for the protein can then be calculated from the vector construction shown in Figure 2.20b. A circle of radius equal to  $F_{PH}$  is drawn centred on the end of the vector  $-F_H$ . The points of intersection with a circle of radius equal to  $F_P$  centred on the origin allow the construction of two vectors which represent the two possibilities for  $F_P$  shown by the green arrows, thus leaving an ambiguity for the protein phase angle. This ambiguity can be resolved by the use of further, unrelated derivatives. It can be illustrated by the vector diagram (Figure 2.20d) which incorporates a second derivative, where the circle with radius  $F_{PH2}$  is centred on the end of the vector  $-F_{H2}$ . In the ideal case of two perfectly isomorphous derivatives all three circles will coincide at a single point, thus giving a clear, unique solution. In the ideal situation,  $\alpha_P$  for both  $F_{PH1}=F_P + F_{H1}$  and  $F_{PH2}=F_P + F_{H2}$  summations will give only one value. However, this is not the case with experimentally determined phases and in real situations the observed and calculated values of  $F_{PH}$  differs by a lack of closure error ( $\varepsilon$ ; Figure 2.21).



**Figure 2.20** a) Vector diagram showing the relationship between the structure factors of native protein ( $F_P$ ), heavy atom ( $F_H$ ) and heavy atom-protein derivative ( $F_{PH}$ ); b) Vector construction for a single isomorphous derivative; c) Vector construction for a second isomorphous derivative; d) combination of the two derivatives to determine the  $F_P$  phases as used in MIR.



**Figure 2.21 Multiple isomorphous replacement in reality. Ambiguous phase angles are obtained because circles cross at a number of points requiring the best estimation of phases. Figure taken from Perutz 1992.**

Determination of initial phase estimates utilise the probability  $P(\alpha_p)$  for having a correct phase angle are higher on decreasing values of  $\varepsilon$ . Refinement of heavy atom parameters such as B factor, occupancy and position is carried out by minimisation of  $\varepsilon$ . It is critical to refine the heavy atom parameters as these will have an influence over the resulting calculated phases and subsequent structure determination. Observation of the  $R_{cullis}$  and Phasing power values output by phase refinement programs are used to evaluate each derivative:

$$R_{cullis} = \sum \frac{|F_{PH} \pm F_P| - |F_{Hcalc}|}{|F_{PH} \pm F_P|}$$

**Equation 2.20**

$$\text{Phasing power} = \sqrt{\sum |F_H| \sum (|F_{PH}|_{obs} - |F_{PH}|_{calc})^2}$$

**Equation 2.21**

An additional aid in determining protein phases is that reflections can be divided into two forms, centric reflections and acentric reflections. Centric reflections are much fewer in number and are seen under specific restraints specific to each space group. Protein molecules

are non-centrosymmetric structures since only one enantiomorph is present (that comprising L-amino acids). A crystal structure will appear centrosymmetric when projected down any even-fold rotation axis. Thus, space groups P<sub>2</sub><sub>1</sub> and C<sub>2</sub> have one centric zone  $h\ 0\ 1$  if **b** is the unique axis. In a similar way P<sub>2</sub><sub>2</sub><sub>2</sub> have three centric zones which are  $h\ 0\ 1$ ,  $h\ k\ 0$  and  $0\ k\ l$ . The phase angle of centric reflections is either  $\pi$  or 0 i.e. the  $F_P$  and  $F_{PH}$  are collinear to each other, resulting in an aid to the determination of the location of the introduced heavy atoms and also in the estimation of errors in their positions.

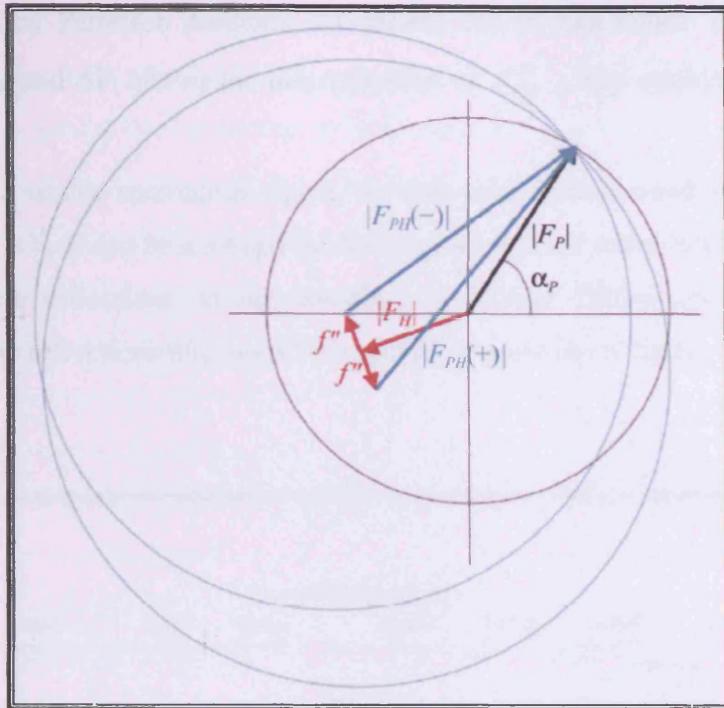
#### **2.7.4.1. Preparation of heavy atom derivatives**

A crystal is soaked in a solution of heavy atoms that penetrate the crystal via solvent channels and interact with the protein molecule. The heavy atom may be chosen to exploit the physical and chemical properties of the protein as they frequently bind to well defined sites in the native protein. For example, mercurial compounds will react specifically with the thiol groups of cysteine side chains, and platinum salts with the side-chain atoms of methionine and histidine residues. However, these residues must be solvent exposed to allow the interactions to occur. Traditionally, crystals were soaked for a number of hours to days to produce derivatives (Blundell and Johnson, 1976), however recent experiments by Dauter (2000) have introduced the idea of very short soaks of less than one minute to introduce halide atoms into the ordered solvent region surrounding the protein molecule. A compromise between these two techniques is to use relatively high concentrations of heavy atom reagents and soak times of 5-30 minutes which leads to improved data quality and resolution limits compared to longer soaked crystals (Sun *et al.*, 2002). Determination of which heavy atoms can react with the protein of interest may also be carried out in solution using mass spectrometry avoiding the need of collecting multiple putative heavy atom data sets (Sun and Hammer, 2000). Use of this technique may provide a heavy atom that may also be suitable for co-crystallisation if derivatisation by soaking methods has proved unsatisfactory. Heavy atom substituted protein ligands can also be used to produce a derivative of a ligand bound protein structure. In addition, Noble gases (Xe for instance) can be used quite successfully to obtain derivatives (Vitali, 1991). With the advent of site directed mutagenesis, specific residues in the protein sequence can be engineered to facilitate binding of heavy atoms (Sun *et al.*, 1987). Thus, random introduction of cysteine residues, aimed to produce isomorphous mercury derivatives, has become a possibility.

Replacing the sulphur in the methionine side chains with selenium during expression is a technique commonly used to obtain suitable crystals for a MAD experiment. Selenium substitution is achieved by growing the host cell containing the vector with the gene of interest in a minimal media devoid of methionine but supplemented with selenomethionine. Generally a Met<sup>-</sup> strain of the host cell is used so the host is unable to synthesise methionine *de novo* ensuring an improved incorporation of Selenium within the protein.

### 2.7.5. Anomalous Dispersion

Resolving the phase ambiguity inherent in a single derivative is possible by using the anomalous scattering component of the diffraction as outlined in Section 2.6.7. This takes advantage of a heavy atoms capacity to absorb x-rays of specified wavelength. During a diffraction experiment, atoms absorb x-rays as well as emit them and this absorption drops sharply at wavelengths just below their characteristic emission wavelength  $K_{\beta}$ . This sudden change in absorption as a function of wavelength is called an absorption edge. An element thus exhibits maximal anomalous scattering when the x-ray wavelength is near the elements absorption edge. Absorption edges for C, N and O are not near the wavelength of X-rays used in X-ray diffraction experiments. The contributions of  $\Delta F_R$  and  $\Delta F_i$  are significant for heavy atoms in the case of acentric reflections. Depending on the sign of the phase angle, the  $\Delta F_i$  component either increases or decreases the intensity of a diffracted ray. Since the phase angles of reflections from opposite sides of the same crystal have opposite signs, anomalous scattering increases the intensities of one and decreases that of the other. The signs of these intensity changes provide the signs of the phase angles. Figure 2.22 shows the modified Harker construction for a single heavy atom derivative and anomalous scattering, which explains how the protein phase angle problem can be solved with one isomorphous derivative if anomalous scattering is incorporated. However, the changes in the intensities caused by the anomalous scattering are very small and differences of a few percent must be measured very accurately between  $|F(hkl)|$  and its Friedel opposite  $|F(-h-k-l)|$  at the same wavelength (Bijvoet differences).



**Figure 2.22 Harker construction for a single derivative using anomalous scattering (SIRAS) Taken from Taylor, 2003.**

A further way of utilising anomalous dispersion for phase determination involves carrying out diffraction experiments at different wavelengths known as multi-wavelength anomalous dispersion (MAD, Fanchon and Hendrickson, 1990). Data sets from a heavy atom derivative at different wavelengths are in many respects like those from distinct heavy atom derivatives thus allowing phase determination. When the x-ray wavelength is near the heavy atom absorption edge, a fraction of the radiation is absorbed by the heavy atom and re-emitted with altered phase.

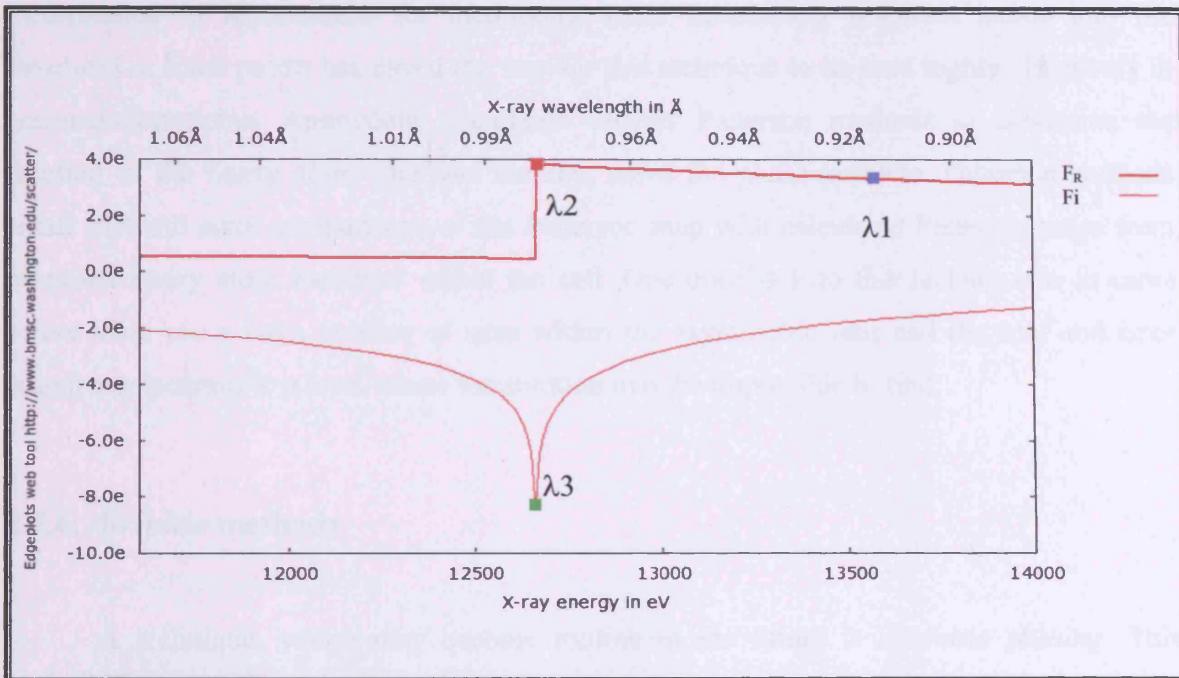
$$F_{PH}^{\lambda 1} = F_{PH}^{\lambda 2+} - \Delta F_R^+ - \Delta F_i^+$$

**Equation 2.22**

$F_{PH}^{\lambda 1}$  represents the structure factor for the heavy atom derivative measured at  $\lambda 1$  where anomalous scattering does not occur.  $F_{PH}^{\lambda 2}$  is the same structure factor measured at a second  $\lambda$  near the absorption edge of the heavy atom. The vectors representing anomalous scattering contributions are  $\Delta F_R$  and  $\Delta F_i$ . The magnitude of anomalous scattering contributions  $\Delta F_R$  and  $\Delta F_i$  for a given element are constant and roughly independent of reflection angle  $\theta$ . The phases of  $\Delta F_R$  and  $\Delta F_i$  depend only upon the position of the heavy atom in the unit cell so

once it is located by Patterson methods, the phases can be calculated. The resulting full knowledge of  $\Delta F_R$  and  $\Delta F_i$  allows the determination of  $F_{PH}^{\lambda_1}$ , thus establishing the protein phases.

To obtain a usable anomalous signal, the data must be measured with a significantly lower noise level, which can be assessed by deriving the Bijvoet ratios from measured centric reflections (**centric** reflections do not contain any Bijvoet differences, and the merging statistics for centric reflections thus reveal the statistical noise in the data).



**Figure 2.23** X-ray fluorescence scan of a seleno-methionine containing crystal and the derived  $F_R$  and  $F_i$  values. Three wavelengths are chosen to maximize dispersive and anomalous signal: remote (blue), peak (red) and the infliction point (green). Figure adapted from Walsh *et al.*, 1998.

Especially in the vicinity of the absorption maximum of the heavy atom, the scattering factors  $\Delta F_R$  and  $\Delta F_i$  vary greatly with x-ray wavelength (Figure 2.23). At the absorption max ( $\lambda_2$ )  $\Delta F_i$  reaches its maximum value, whereas at the ascending inflection point or edge ( $\lambda_3$ ),  $\Delta F_R$  reaches a minimum and then increases further from the absorption peak. Data sets taken at the absorption maximum, the edge and at a wavelength distant from the maximum all have distinct values for the  $F_R$  and  $F_i$  of anomalous dispersion. Therefore each measurement of a Friedel pair at a specific wavelength provides the components of distinct sets of phasing

equations. In addition, individual intensities vary slightly with wavelength and these differences also have phase information.

Using the tunability of synchrotron radiation sources it is possible to collect data under conditions that maximise anomalous scattering by the heavy atom. However, the measured difference is still relatively small and the accuracy of measurement is critical. Experimental error is minimised by highly redundant data collection, the use of high quality x-ray beams maintained at specific wavelengths and cryo-crystallography, which preserves the crystal. Multi-wavelength anomalous dispersion was initially only used for small proteins with functional metal cofactors. However, the ability to produce proteins with seleno-methionine incorporated in replacement for methionine (thus introducing selenium atoms into the structure) at fixed points has paved the way for this technique to be used highly effectively in non-metalloproteins. Anomalous dispersion utilises Patterson methods to determine the location of the heavy atom sites and thereby, solve the phase problem. Patterson methods entail trial and error comparisons of the Patterson map with calculated Patterson maps from proposed heavy atom locations within the cell. One drawback to this technique is in cases where there are a large number of sites within the asymmetric unit and the trial and error search can increase to a level where the solution may be impossible to find.

#### 2.7.6. *Ab initio* methods

A technique, which may become routine in the future is *ab initio* phasing. This technique has been used in small molecule crystallography and has only been recently brought into the domain of macromolecular crystallography. It relies on solving the phase problem using the diffraction amplitudes alone. The development of a direct methods program by Weeks *et al.*, (1993) in the form of the *Shake and Bake* program has led to the solving of structures with high-resolution data approaching 1 Å resolution. This program works by a repeated and unconditional alternation of reciprocal space phase refinement with a complementary real space process that seeks to improve phases by applying constraints. This technique can also be used in conjunction with isomorphous or anomalous scattering information to get phase information. Using this second approach, the dual space *Shake and Bake* procedure was used to solve the anomalously scattering seleno-methionine derivative of an epimerase enzyme that has 70 selenium sites (Deacon and Ealick, 1999). Recent work by Giacovazzo to extend his small molecule program SIR to applications in macromolecular crystallography has shown some promising results in solving macromolecular structures that

diffract to 1.4 Å (Giacovazzo and Siliqi, 2002). Further extension of the resolution limits and the macromolecular size limits would lead to widespread use of this technique.

### 2.7.7. Radiation induced phasing

Although radiation damage is generally avoided to maintain the quality of the data collection over a full data set. A recent technique has been proposed to utilise this radiation damage which utilises the fact that radiation damage can cause specific structural changes. Accurate and redundant datasets are collected from a crystal a number of times with the crystal exposed to an accumulated dose of  $1 \times 10^7$  Gy in between collected datasets. In order to determine the sites of structural damage the data is treated as a case of SIR with quality of initial phases unrelated to resolution limits. Some of the specific structural changes that are observed are disulphide bond cleavage, carboxyl group and carbonyl oxygen removal as well as removal of water molecules (Ravelli and McSweeney, 2000). The amenability of this technique to a large number of protein crystals may in the future render the problem of derivatisation of crystals a less significant barrier to structure determination.

### 2.7.8. Density modification

Initial phase information that is obtained from experimental phasing is usually not of high enough quality to produce interpretable electron density maps. Therefore, prior to model building and refinement, phase improvement techniques are carried out to provide a map that will allow iterative model building. Errors in the experimental phases are due to experimental errors, questionable isomorphism in heavy atom derivatives and a poor initial starting model in the case of molecular replacement. Density modification methods use known features of a protein electron density map to improve the phasing obtained from experimental data. These include solvent flattening (Bricogne, 1974), non crystallographic symmetry averaging (Crowther, 1969) and histogram matching (Zhang and Main, 1990) that can improve the electron density maps.

### **2.7.8.1. Solvent flattening**

Due to the dynamic nature of the disordered solvent, the electron density has a low constant value in the solvent region of the crystal. Inaccuracies in the initial experimental phases lead to noise peaks in the solvent region of the electron density maps. Solvent flattening is accomplished by constructing a molecular envelope or mask around the protein region by dividing the unit cell into a three dimensional grid and replacing the density at each grid point with a new value which is proportional to the weighted sum of the densities within a sphere of radius  $R$  with the centre in that grid point.  $R$  is typically on the order of 10 Å.

$$\rho'_j = K \sum_i^R w_i \rho_i$$

**Equation 2. 23**

with  $w_i = 0$  for  $\rho_i < 0$  and  $w_i = 1 - (r_{ij}/R)$  for  $\rho_i > 0$ . The summation is over the grid points  $i$  within the sphere.  $K$  is an arbitrary constant and  $r_{ij}$  is the distance between the grid points  $i$  and  $j$ . Structure factors and phases calculated from the solvent flattened map are combined with the initial phases and weighted on the basis of how well the modified magnitudes match the observed values, then phase combination with the experimental phases is carried out with this process repeated until convergence. Solvent flattening reduces the noise in the solvent region to a constant low value, which results in significant improvements in the quality of density in the protein region (Wang, 1985; Leslie, 1987).

### **2.7.8.2. Histogram matching**

Histogram matching can be applied to the protein electron density map. A histogram is produced that covers the range of positive and negative electron density values throughout the protein. Parameters such as resolution limits and overall B factor allow a predicted histogram to be constructed with high precision. Differences between the predicted histogram and the one calculated from experimental phases must be due to phasing errors. Histogram matching converts the electron density values to a value consistent with the predicted distribution of the

structure. This adjustment allows the calculation of enhanced phase angles which allow a more accurate representation of the protein structure (Zhang and Main, 1990).

### 2.7.8.3. Non Crystallographic Symmetry

Non crystallographic symmetry averaging utilises the presence of identical copies of a subunit within the asymmetric unit. These molecules can be aligned by symmetry elements that do not coincide with the symmetry elements of the entire crystal. Averaging of the densities for molecules related by non crystallographic symmetry can be applied and allows the signal of identical molecules to be combined thus increasing the signal to noise ratio and an improvement of the overall electron density map.

## 2.8. Model building and structure refinement

From an atomic model it is possible to calculate all of the structure factors and the phases of the reflections via a Fourier transform. This back transformation is critical in accurately determining the phases of a structure. Initial estimates of the phases are introduced and improved by iterative steps in the phase estimates. An estimated  $\rho(xyz)$  is computed with observed intensities  $I_o$  and estimated phases ( $\alpha_{calc}$ ) from which map interpretation can be carried out leading to structure building. Determination of new structure factors  $F_c$  using the improved  $\rho(xyz)$  derived from the new model can then be carried out. Using the new estimation of  $F_c$ , an improved estimate of  $\alpha_{calc}$  and an iterative process of structure building and phase refinement are carried out until structure completion.

Map fitting is carried out by use of interactive map building programs. Structure factors calculated ( $F_{calc}$ ) on the basis of the initial model are generally in rather poor agreement with the observations ( $F_{obs}$ ). Subsequent model building is observed by use of a crystallographic R factor which allows structure changes to be evaluated.

$$R = \frac{\sum_{\mathbf{h}} |F_{obs} - kF_{calc}|}{\sum_{\mathbf{h}} F_{obs}}$$

**Equation 2.24**

where  $k$  is a scale factor.

Structure refinement is the process of adjusting the model parameters (the coordinates and the temperature factors for all atoms in the structure) to find a closer agreement between the calculated and observed structure factor, and hence lower the R-factor and improve upon the accuracy of the structural model.

However, over refinement of a protein structure has been shown to occur with the introduction of excessive model parameters leading to incorrect structures (Brünger, 1992). This has led to the introduction of a free R factor parameter where refinement can be assessed in an unbiased manner. A random selection of 5-10 % of the observed reflections is selected as the ‘test set (T)’ and is not included in refinement. Refinement is carried out and the free R-factor is calculated with the test set T only:

$$R_T^{\text{free}} = \frac{\sum_{hkl \subset T} ||F_{\text{obs}}| - k|F_{\text{calc}}||}{\sum_{hkl \subset T} |F_{\text{obs}}|}$$

**Equation 2.25**

where  $k$  is a scale factor and  $hkl \subset T$  are all reflections belonging to test set T. There is a high inverse correlation between the free R-factor and the accuracy of the atomic model phases.

### 2.8.1. Least squares refinement

The most common macromolecular crystallographic refinement involves restrained optimisation of the agreement between diffraction amplitudes calculated from an atomic model and those that are derived from the experimental data. In the TNT program (Tronrud, 1992), stereochemical restraints are introduced by optimising the agreement with ideal geometries (bond lengths, bond angles, dihedral angles etc) using the least-squares (LS) method. Effectively, this means that ‘extra’ data are added to the experimental observations in the form of geometrical and stereochemical restraints on the model, thus making the overall number of ‘observations’ sufficient for an unambiguous solution of the minimisation problem to be found using the classical LS method. The restrained LS

refinement method has some drawbacks. In particular, when the model is very incomplete, density for the missing parts tends to disappear rather than improve during LS refinement.

### 2.8.2. Maximum-likelihood based refinement

In the maximum-likelihood (ML) based refinement method, the model structure factors are handled as probability distributions (Bricogne, 1997). These are integrated over the phase to yield predicted distributions of model amplitudes; substituting the observed values of these amplitudes yields the likelihood  $L$  of the model. All parameters are then refined by maximization of the likelihood function  $L$ :

$$L = \prod_{hkl} P(|\mathbf{F}_o|; |\mathbf{F}_c|)$$

**Equation 2. 26**

which is done by minimization of the logarithm  $-\log(L) = -\sum_{hkl} \log[P(|\mathbf{F}_o|; |\mathbf{F}_c|)]$ , as sums are easier to deal with than products.

It has been shown that various sources of random error in the model have equivalent effects on the probability distribution for the true structure factor, whether the errors are in atomic positions or temperature factors or whether there are missing or extra atoms; in each case, the distribution of the true structure factor is well approximated by a Gaussian distribution. Thus, the maximum-likelihood (ML) method takes into account the uncertainty of the phases associated with model incompleteness and imperfection. ML refinement allows the refinement of an incomplete model by using the structure factor statistics of randomly distributed atoms to represent the effect of the missing atoms. Due to the general phase improvement through ML refinement, the final map provides indications about the location of these missing atoms.

## 2.9. Structure Validation

Due to resolution limits and phase calculation errors, building and refinement of a structure may lead to errors in the structure which need to be corrected prior to analysis, publication and deposition of the structure (Kleywegt, 2000). Errors that have been introduced during model building in the past include total model or subunit errors (Ammon *et al.*, 1988), incorrect local or main chain conformations (Kleywegt and Jones, 1998) or frame shift of a

number of residues (Kleywegt *et al.*, 1997). Quality indicators of a model that can be assessed include real space fit of the model to the density (Jones *et al.*, 1991, Ramachandran plot (Ramachandran and Sasisekharan, 1968), hydrogen bonding analysis (Hooft *et al.*, 1996). These indicators can be used primarily to highlight specific residues that may require modification.

## 2.10. High Throughput Technologies

A recent development in X-ray crystallography is the arrival of high throughput structural genomics groups whose aim is the structural elucidation of a large amount of known proteins available from genomics data as quickly as possible (Pusey *et al.*, 2005). However, as for all protein structure determination, the limiting step is the crystallisation of the proteins, with most structural genomics initiatives having a low conversion rate from purified proteins to solved protein structures. Due to the large resources invested in these programs considerable results are required and expected. However, traditional expertise in molecular biology and protein chemistry are abandoned in the high throughput procedures. However the progress in the high throughput technologies that have been implemented cannot be ignored and this includes nano-crystallisation techniques, sample mounting robotics (Adams *et al.*, 2003). Automatic sample mounting and crystal focusing facilities may lead to unattended X-ray diffraction experiments with greater time efficiency at synchrotron beamlines (Muchmore *et al.*, 2000). On a structural level, examination of the US PSI initiative gives an investment of \$270 million resulting in 1,100 structures (<700 sharing 30% homology with previously solved structures) in the pilot phase.

## **Chapter 3**

### **Materials and Methods**

### 3.1. Materials

#### 3.1.1. Chemicals and Reagents

Ampicillin, kanamycin, *o*-chlorophenylhydroxyacetic acid, 4-hydroxyphenylacetic acid, isopropyl- $\beta$ D-thiogalactopyranoside (IPTG), cacodylic acid and polyethylene glycol (all MW ranges) were from Sigma chemical company. All other chemicals were supplied by Sigma-Aldrich, Fisher or BDH and were of analytical grade wherever possible. Water used was double deionised. 1 kb DNA marker was from New England Biolabs.

#### 3.1.2. Bacterial Strains and Media

Bacterial *E. coli* strains employed were BL21 DE3 and DH5 $\alpha$ . *E.coli* strains were grown in Terrific broth media, which contains per litre: 24 g tryptone, 12 g yeast extract and 4 ml glycerol. Solid media used was Luria-Bertani (LB) medium which contains, per litre: 10 g yeast extract, 5 g NaCl and 15 g bacto-agar. All media components were of laboratory standard supplied by OXOID Ltd., Basingstoke, Hampshire. Solid media was produced by the addition of 1.5 % bacto-agar (w/v) to the liquid media. Transformed bacteria were grown in LB media supplemented with 50  $\mu$ g/ml ampicillin or 30  $\mu$ g/ml kanamycin as required. Glycerol stocks were made by adding 0.5 ml of a late exponential phase culture to 0.5 ml sterile glycerol. The stocks were then stored at -20 °C or -70 °C.

#### 3.1.3. Chromatographic Media and Membranes

Hi-Trap metal chelating columns and HiTrap MonoQ columns were obtained from Amersham Pharmacia. Desalting columns used were the Hi-Prep 26/10 Desalting column from Amersham Pharmacia and Biorad EconoPac 10 DG disposable chromatography columns. Centricon 10 and Centricon 100 were supplied by Amicon Inc. Dialysis tubing (width 25mm, 1.98 ml/cm, molecular weight cut-off 12,000-14,000 kDa) was obtained from CelluSep.

### 3.1.4. Oligonucleotides and Sequencing

All oligonucleotides were synthesised by Invitrogen Corp., UK. All sequencing was carried out by Lark Technologies, Essex, UK.

### 3.1.5. Crystallisation Screens and Plates

Initial screening of protein for crystallisation conditions was carried out using commercial screens from Hampton Research, California, USA; Molecular Dimensions, Cambridgeshire, UK; and Emerald Biosystems, Washington, USA. Plates used for sitting drop vapour evaporation experiments were 24 well Cryschem plates from Hampton Research and 96 well Combi-clover trays from Emerald Biosciences. Heavy atom reagents used were from Hampton Research Pt screen and Hg screen for the platinum and mercury derivative respectively.

## 3.2. Molecular Biology Methods

### 3.2.1. DNA purification

For small-scale plasmid DNA isolation (miniprep), a single *E. coli* colony was picked from an agar plate and used to inoculate 10 ml of LB medium, containing 50 µg/ml ampicillin or kanamycin. Cultures were grown overnight at 37 °C with vigorous shaking. DNA was extracted from the overnight culture using the Qiagen Mini-prep kit. This involves the alkaline lysis of the bacterial cells and the removal of the cell lysis and the subsequent adsorption of the DNA plasmid material onto a silica membrane in the presence of high salt and the subsequent elution of the plasmid in a concentrated buffered solution (Vogelstein and Gillespie, 1979; Sambrook *et al.*, 1989).

### 3.2.2. Transformation of *E. coli* Cells with DNA

Competent cells from Stratagene were removed from -80 °C freezer and immediately placed on ice allowing the cells to thaw. 1 µl of a ligation reaction or purified plasmid DNA was added directly to the cells in a 1 ml eppendorf tube. The tube was placed on ice for 10 minutes and then in a water bath at 42 °C for exactly 30 sec. The tube was then placed on ice

for 10 minutes. 80 µl of SOC buffer (2 % tryptone, 0.5 % yeast extract, 10 mM sodium chloride, 2.5 mM potassium chloride, 10 mM magnesium chloride, magnesium sulphate, 20 mM glucose per litre of water) was added to the solution and incubated at 37 °C for 1 hour at 250 rpm. Selection for successful transformants was accomplished by plating on medium containing antibiotic for the plasmid encoded drug resistance.

### 3.2.3. Agarose Gel Electrophoresis

Agarose gel electrophoresis was carried out according to the method of Sambrook *et al.*, (1989). Agarose gels (0.8 % w/v, 11 cm x 12 cm) containing 0.6 µg/ml ethidium bromide were cast and electrophoresis was performed in 1x TAE buffer (50x TAE buffer is 242 g Tris base, 18.6 g EDTA and 57.1 ml glacial acetic acid in 1 litre), at 150 V for 40 min. Samples (20-25 µl) were loaded onto the gel in 1x loading buffer (6x loading buffer is composed of 0.25 % bromophenol blue, 0.25 % xylene cyanol and 30 % v/v glycerol in water). DNA markers (1 kb marker) were also loaded onto the gel. DNA was visualised by exposure to long wavelength UV light on a UVP transilluminator.

## 3.3. Protein Methods

### 3.3.1. SDS Polyacrylamide Gel Electrophoresis

Electrophoresis was carried out using NuPAGE Novex 4-12% Bis-Tris gels in MES SDS Buffer (50 mM MES, pH 7.2, 50 mM Tris base, 0.1 % SDS, 1 mM EDTA) from Invitrogen Corp., UK at 200 V for 30-40 min with 120 mA at start of run and 70 mA at the end of the run. See Blue Plus II from Invitrogen was used as pre-stained molecular weight standards. Protein samples (10-20 µl) were loaded onto the gel in 1x loading buffer (4x loading buffer is 200 mM Tris, pH 8, 8 % SDS, 20% glycerol, 1 mg/ml Coomassie Brilliant Blue R250 in water). Gels were stained using Coomassie Brilliant Blue R250 (0.25 % in 5:1:5 solution of methanol: acetic acid: water) for approximately 30 min and then destained in 5:1:5 v/v methanol: acetic acid: water for 1 hour.

### 3.3.2. Refinement and analysis of structures

The REFMAC5 program was used for restrained refinement using maximum likelihood, where both the X-ray residual which reflects the agreement between the observed and the calculated Fs, and the geometric residual which reflects the fit between the expected and the observed geometry were minimised at the same time. The relative weighting of these two terms can be adjusted depending on the quality of the data with a low weighting giving more emphasis on geometric correctness and a higher weighting allowing the observed data to be more important. REFMAC5 produces an output file (containing weighted coefficients for SigmaA weighted Fo-Fc and 2Fo-Fc maps (Maps calculated using partial structure phases are biased towards the partial structure. Sigma A is an estimation of the mean coordinate error of the model and suppresses the model bias leading to more accurately calculated electron density maps). In the case of the low resolution substrate-bound crystal form of the A264E mutant of BM3, strong NCS-restraints of level 1 (tight restraint of both side chain and main chain atoms) were imposed throughout refinement. Electron density maps were displayed in TURBO-FRODO (Roussel and Cambillau, 1992) to guide model building. In the final stages of refinement molecular geometry is analysed using PROCHECK (Laskowski *et al.*, 1993) to analyse and validate the atomic model. A Ramachandran plot is constructed which analyses the phi and psi angles of the peptide bonds between amino acid residues. Regions in the map are known to be the most likely angles in the case of common secondary structure such as  $\alpha$ -helices and  $\beta$ -sheets. Outliers in the Ramachandran plots are analysed for their fit to the electron density maps. Analysis by the program WHATIF (Hooft *et al.*, 1996) can also indicate alternate rotamers of charged residue side chains to give an improved hydrogen bonding network within the protein. Final refinement can be carried out after these changes are made to ensure accurate fit of the model to the electron density.

Comparison of the structure with homologues is carried out by overlaying these structures using LSQKAB (Kabsch, 1976). This aligns the structures on a least squares basis over a user defined number of structure residues.

### 3.3.3. Analysis of protein sequences

All proteins and homologues were aligned using ClustalW 1.8 using the website <http://www.ebi.ac.uk/clustalw/> with default parameters.

### 3.3.4. Analysis of structural homologues

Structural homologues of the determined protein structures were identified by submitting the determined protein structure coordinates to a database search of the PDB using the webpage <http://www.ebi.ac.uk/dali/Interactive.html> which is part of the EMBL-EBI website.

## 3.4. Methods for Analysis of A264E mutant of Cytochrome P450 BM3 Heme domain

### 3.4.1. Crystallization of P450 BM3 A264E

The P450 BM3 A264E heme domain was crystallized in both the palmitoleic acid bound and substrate-free forms by the sitting drop method. Sitting drops were prepared by adding 2 µl of mother liquor to 2 µl of 15 mg/ml enzyme. Palmitoleic acid-bound crystals were obtained by co-crystallizing the enzyme with 1.2 µM palmitoleic acid (approximately 6 x the  $K_D$  value obtained from spectral binding titration data), and a well solution of 100 mM magnesium acetate, 20 % PEG 2000 MME and 100 mM cacodylic acid at pH 6.3 at 4 °C. Substrate-free crystals were obtained with a well solution of 10mM manganese sulphate, 20% PEG 2000MME and 100 mM cacodylic acid at pH 6.3 at 4 °C. Crystals were formed in both cases after approximately 7 days. Crystals were immersed in 10 % PEG 200 in mother liquor to act as a cryoprotectant, before being mounted on a nylon loop and flash-cooled in liquid nitrogen. Crystallisation was carried out by Hazel M. Girvan at the University of Leicester.

### 3.4.2. Data collection and processing

Data used for refinement were collected at ESRF, Grenoble, France on ID14-EH1 using an ADSC Q4R CCD detector. Crystals were cooled at 100 K and diffraction data collected in oscillations of 1°.

Data were processed and scaled using the HKL package programs DENZO and SCALEPACK. In the DENZO program (Otwinowski & Minor, 1997), the indexing procedure is automatic. The crystal and detector orientation parameters and the unit cell parameters are refined by a least-squares procedure, and the position of each reflection in three-dimensional reciprocal space is determined. To increase the precision of measurements, profile fitting is used. First, the profile of the Bragg peak is predicted based on the profiles of other reflections

within a chosen radius. The predicted profile in DENZO is an average of profiles shifted by the predicted separation between the spots, so that they are put on top of each other. If the predicted positions are in error, then the average profile will be broadened and/or displaced from the actual profile of the reflection. In the second step, the information from the actual and the predicted profile is combined to effectively match the shape of the spot and its nearest neighbours. To calculate the diffraction intensity, the detector background is estimated by calculation of an average detector signal in the neighbourhood of a specific reflection and then subtracted from the reflection profile.

The program Scalepack scales all data together and allows for the analysis of the quality of the diffraction data collected based on the statistics that are obtained from the scaling and merging of the raw diffraction data. A correction factor is applied to the measured data in order to compensate for weak data and give reasonable estimates. Parameters such as error scale factor and R factor values are adjusted to give an overall  $\chi^2$  value or goodness of fit for the data close to 1 which allows for a proper estimation of the correction factor that should be applied to the diffraction data. Outlier reflections are rejected from the diffraction data. The ratio of the measured intensity of a reflection to the error of the intensity ( $I/\sigma$ ) within a resolution shell allows the identification of the resolution limits of the diffraction experiment with a reasonable resolution cut off value for useful data where  $I/\sigma < 2$ . A complete dataset should contain  $\geq 90\%$  of the total measurements within the resolution limits of the diffraction experiment.  $R_{\text{merge}}$  values and data completeness are indicative of the data quality.

TRUNCATE (French & Wilson, 1978) was used to convert averaged intensities to mean values of amplitudes. There are two ways to calculate the amplitudes from the intensities in TRUNCATE. The simplest is just to take the square root of the intensities, setting any negative ones to zero. Alternatively, the TRUNCATE procedure calculates a best estimate of the structure factors ( $F$ ) from measured intensities ( $I$ ), standard deviation of  $I$ , and the distribution of intensities in resolution shells. This has the effect of forcing all negative observations to be positive and inflating the weakest reflections, because an observation significantly smaller than the average intensity is likely to be underestimated.

### 3.4.3. Structure solution of A264E mutant of Cytochromes P450 BM3

The substrate-free crystal structures were solved via molecular replacement using the program AMoRE and the high-resolution P450 BM3 wild-type crystal structure (PDB code

2HPD) as a search model. The wild-type palmitoleate-bound crystal structure was used a starting model for the palmitoleate-bound mutant form.

AMoRE calculates the structure factors for a given model structure on a fine grid with all subsequent searches calculated by interpolating into this data. The rotation search is carried out looking for correlation between Patterson maps centred on the origin. The correlation between the modelled and calculated Patterson maps is used as an indicator for the correctness of the solution. Similarly, the correlation coefficients and R factors between the  $I_{\text{obs}}$  and  $|F_{\text{obs}}|$  and the calculated values from the model indicate correct solutions when these values are clearly above noise levels. Translation searches are carried out with each individual set of rotation parameters and the results are added together in subsequent searches (in the case of multiple copies of the protein molecule in the asymmetric unit). Solutions are output in Crowther angles and are used to re-orientate the search model for use in subsequent building and refinement. Rigid body refinement is carried out on the final solutions to give an optimised fit of the reoriented model to the data.

#### 3.4.4. Modelling of the active site mutants

The modelled A264E mutant of the native BM3 cytochrome P450 domain (PDB code: 2HPD) in different orientations as shown in Figure 4.14 was generated using the residue mutation option in Pymol. The introduced glutamate residue side chain was shown in two conformations that are regularly seen in nature.

The mutations shown in Figure 4.18 are also generated similarly with the newly introduced residue side chain shown in the orientation that is most often seen in nature.

### 3.5. Methods for Analysis of the FAD domain of Cytochrome P450 BM3

#### 3.5.1. Crystallisation of the FAD domain of Cytochrome P450 BM3

Initially crystallisation of BM3 FAD domain was attempted by Dr. Olivier Roitel (University of Leicester), however the crystals were poorly diffracting (3.5-4 Å) and the crystallisation conditions were not reproducible. The process of screening for crystallisation conditions was repeated. Protein was stored at 12 mg/ml concentration at -20 °C in 50 % glycerol and buffer exchange for crystallisation trials into 10 mM Tris pH 7 performed using a Biorad EconoPac 10 DG column. The eluted protein was concentrated to 5-15 mg/ml for crystallisation trials using the sitting drop vapour diffusion technique. Upon initial

crystallisation conditions being identified the condition was optimised by varying the pH, salt concentration and precipitant concentration until large reproducible crystals were obtained. Sitting drops were prepared by adding 2 µl of mother liquor to 2 µl of 12 mg/ml FAD domain. In order to form a cofactor complex with NADP<sup>+</sup>, crystals were soaked in a 10 mM NADP<sup>+</sup> solution for 5-10 minutes. Native crystals and crystals soaked with NADP<sup>+</sup> were immersed in 10 % polyethylene glycol in mother liquor to act as cryoprotectant, prior to mounting on a nylon loop and flash cooling in liquid nitrogen.

### **3.5.2. Data collection**

The data used for refinement was collected on a Mar CCD detector on ID14-EH3 at the European Synchrotron Radiation Facility (Grenoble, France). Crystals were cooled at 100 K and diffraction data was collected in 0.5 ° oscillations on a Mar 165 mm CCD detector.

### **3.5.3. Structure solution of FAD domain of Cytochrome P450 BM3**

The crystal structure of the NADP<sup>+</sup> bound form was solved using the molecular replacement program AMoRE (Section 3.4.3) and the FAD domain of rat cytochrome P450 reductase (Protein Data Bank code 1AMO) as the search model. The uncomplexed FAD domain was solved using difference Fourier methods.

## **3.6. Methods for Analysis of CprK**

### **3.6.1. CprK Growth and Expression**

The CprK gene cloned into pET vector 21a+ was kindly donated by Dr. J. van der Oost (Wageningen, The Netherlands). This vector was transformed into BL21 (DE3) *E.coli* (Novagen) cells. Cells were grown in Luria-Bertani (LB) media at 37 °C in an orbital shaker at 250 rpm until cell density reached an OD<sub>600</sub> of 0.8. Protein expression was induced by adding 0.5 mM IPTG and cells were grown for an additional 6 hours at 30 °C. Cells were harvested by centrifugation at 5000 rpm x 15 min and stored at -80 °C prior to purification.

### 3.6.2. CprK Purification

Cells were broken by sonication in the presence of EDTA free protease inhibitors (Roche). Cell debris was removed by centrifugation at 15,000 rpm for 30 min and the supernatant was loaded onto a HiTrap Chelating column (Amersham) pre-charged with Nickel ( $\text{Ni}^{2+}$ ) ions and pre-equilibrated in 0.1 M potassium phosphate buffer pH7, 0.5 M NaCl (Buffer A). Two wash steps were carried out using Buffer A plus 20 mM and 50 mM imidazole for 10 column volumes each. CprK was then eluted with Buffer A plus 400 mM imidazole. Fractions containing CprK were pooled together and desalting into 20 mM Tris pH 8 using a HiPrep 26/10 Desalting column. The sample was concentrated to 10 mg/ml and stored at 4 °C.

### 3.6.3. Determination of Protein Concentration

Protein concentration was determined using the BCA protein assay kit from Pierce which is based on the bicinchoninic acid (BCA) technique of the colorimetric detection and quantification of total protein. The enhanced protocol procedure was carried out with 2 ml of working reagent added to 0.1 ml of each standard and unknown sample and incubated at 60 °C for 30 minutes. Absorbance of the solutions was measured at 562 nm and the sample concentration determined from a standard curve.

### 3.6.4. Crystallisation of CprK

CprK was dialysed against 10 mM Tris pH 7.5 and concentrated to 7-10 mg /ml. This solution was then used to screen for suitable crystallisation conditions using Hampton Research crystal screens I and II, Molecular Dimensions clear strategy screen I and II and Structure screen I and II, Emerald Biosciences Cryo I and II and Wizard I and II screens. Screening solution (500 µl) was placed into a well of the Cryschem box, 2 µl of protein solution was placed onto the ledge above the well and 2 µl of the well solution was added to the protein drop and allowed to mix. The well was sealed with crystal clear tape and duplicate trays were prepared and stored at 4 °C and 20 °C respectively. 1 µl of a saturated solution of OCPA was added to 200 µl of protein solution and allowed to equilibrate for 30 min. The inducer bound protein solution was also screened for crystallisation conditions as described above. The conditions that allowed crystallisation were further optimised by screening against

a range of pH, salt conditions and precipitant concentration close to the initial crystallisation condition. Details of crystallisation conditions are given in Chapter 6.

### **3.6.5. Structure solution of CprK**

#### ***3.6.5.1 Heavy atom derivitisation***

Heavy atom derivitisation of the crystals was carried out by soaking the crystals in mother liquor solutions containing 10 mM potassium tetrachloroplatinate (II) and 10 mM mercury (II) acetate for 10 minutes. These crystals were then transferred to cryoprotectant solution as given above for 10-20 s and cryo cooled by plunging the crystals into liquid nitrogen.

#### ***3.6.5.2 Data collection***

Data used for refinement and phase determination were collected at EMBL DESY, Hamburg on BW7A using a MAR 165mm CCD detector. Crystals were cooled at 100 K using an Oxford Cryostream and diffraction data was collected in oscillations of 1°. Data were processed and scaled using the HKL package programs DENZO and SCALEPACK.

#### ***3.6.5.3. Scaling***

Before isomorphous differences can be calculated, the derivative data must be scaled together with the native data. The initial scale factor, which is refined later in the isomorphous replacement procedure, is calculated using the program FHSCAL (CCP4, 1994). This program scales the data in reciprocal space using the Kraut method of scaling which is inherently more accurate than the Wilson or Least squares procedure used by SCALEIT (CCP4, 1994). FHSCAL divides the data into resolution shells, smoothes the shell scale factors and then interpolates to get the final scale factor for each reflection using the equation given.

$$K = \frac{\sum (M+1)F_p^2}{\sum M F_p F_{PH} + \sqrt{(\sum M F_p F_{PH})^2 - \sum (M+1)F_p^2 \cdot \sum (M-1)F_{PH}^2}}$$

**Equation 3. 1 The Kraut scale factor is generalised to include both centric and acentric data and applied to the derivative data with  $F_{PH}$ ,  $F_p$  and  $F_H$  as derivative, native and heavy-atom amplitudes respectively and  $K$  is the derivative scale factor to be determined and  $M = 1$  for centric and 2 for acentric data**

### 3.6.5.4. Structure determination

Patterson difference maps were calculated from the three data sets using FFT (Ten Eyck, 1973). Harker sections were displayed using XPLOR84DRIVER (CCP4, 1994) at the u, v and w=0 sections respectively. Initial SIR phases were refined using MLPHARE (Otwinowski, 1991); further sites were determined through difference Fourier analysis and used to produce cross Patterson difference maps for other derivatives. All of the sites were thus determined relative to the same origin. Once all of the sites had been determined and refined, anomalous occupancies were introduced and used to further refine the heavy atom parameters and improve the experimental phases (MIRAS).

### 3.6.6. Map improvement

The experimental electron density maps were further improved by density modification and NCS averaging in DM. Non-crystallographic symmetry (NCS) averaging matrices were generated by positioning 2 copies of the inducer binding domain and the DNA binding domains of CRP into the experimental map manually. Matrices between each set of dimers were generated using LSQKAB.

### 3.6.7. Fluorimetry

All fluorescence experiments were carried out using a Cary Eclipse Fluorescence Spectrophotometer with constant temperature at 25 °C maintained using a Cary single cell peltier regulator. Samples were measured using a Hellma precision quartz cuvette with a 10 mm light path. Excitation was carried out at 295 nm with a 5 nM wavepath and emission fluorescence measured at 370 nm with a 10nM wavepath. Initial baseline was performed

using a 50 mM potassium phosphate buffer pH 7.2 solution with all subsequent experiments carried out using this buffer. 0.5 mM, 1mM and 10 mM solutions of OCPA were prepared in the phosphate buffer and adjusted to pH 7.2 using NaOH. Titrations were carried out by addition of OCPA up to a maximum volume added of 30  $\mu$ l to minimise any changes through dilution effects. The compound was added to a 1 ml solution of 50  $\mu$ M wild-type or mutant CprK. To ensure quenching observed was due to binding of the protein and not due to any interaction between the compound and the tryptophan residue causing the fluorescence; OCPA was titrated against a solution containing tryptophan that gave a fluorescence signal equivalent to the initial CprK fluorescence value.

To determine the binding constant of HPA (hydroxyphenylacetic acid) to CprK, titrations versus OCPA were carried out in the presence of 2.5 mM to 10 mM HPA. The degree to which the fluorescence signal was quenched was determined using the equation:

$$\% \text{ quenched} = \left( \frac{F_1 - F}{F_1} \right) * 100$$

where  $F_1$  and  $F$  represent the Fluorescence intensity of CprK before and after titration of OCPA (with or without PA). The initial value for the titrations with PA added was that after the addition of this compound. In order to determine the binding of OCPA the following quadratic equation was used.

$$\{(A_{\max}/(2*ET)) * ((X+ET+K_D) - (((X+ET+K_D)^2) - (4*X*ET))^{0.5})\}$$

In order to determine the binding constant of HPA the following equation was used

$$Y = (A_{\max}*x)/(x+K_D*(1+(Inh/K_i)))$$

### 3.6.8 Site-Directed Mutagenesis

Mutagenesis was carried out using the Stratagene Quik-Change mutagenesis kit with oligonucleotide primers synthesised by Invitrogen Corp. UK. To create the Y76F mutant the following primers were used:

GGAAGTAAAAGCTTCTC **TTT** TATGCCGGCGGCAATTCC and  
 GGAATTGCCGCCGGCATA **AAAGAGAAGCTTTCACTTCC** ( $T_m = 68$ ,  $2^O T_m = 47$ , length = 39, 49 % GC content) encompassing the mutagenic codon highlighted in blue.

In the case of the K133A mutant the following primers were used:

TTGAAATCTTAAAAACTACCTTACCGCAGTGGCITATTATGCACGACAGGTT and  
AACCTGTCGCGCATAATAAGCCACTGGGGTAAGGTAGTTAAAGATTCAA  
( $T_m = 68$ ,  $2^{\circ} T_m = 50$ , length = 53, 40 % GC content) encompassing the mutagenic codon highlighted in blue.

The PCR reaction was carried out using *pFu* turbo DNA polymerase (Invitrogen) in a 50  $\mu$ l reaction mixture. To assess whether a PCR product was generated, 20  $\mu$ l of the reaction mixture was analysed by Agarose electrophoresis and visualised as previously described in Section 3.2.3. The amplified PCR reaction was digested with DpnI for 1 hour to remove parental DNA and a sample was sent for sequencing by Lark Technologies.

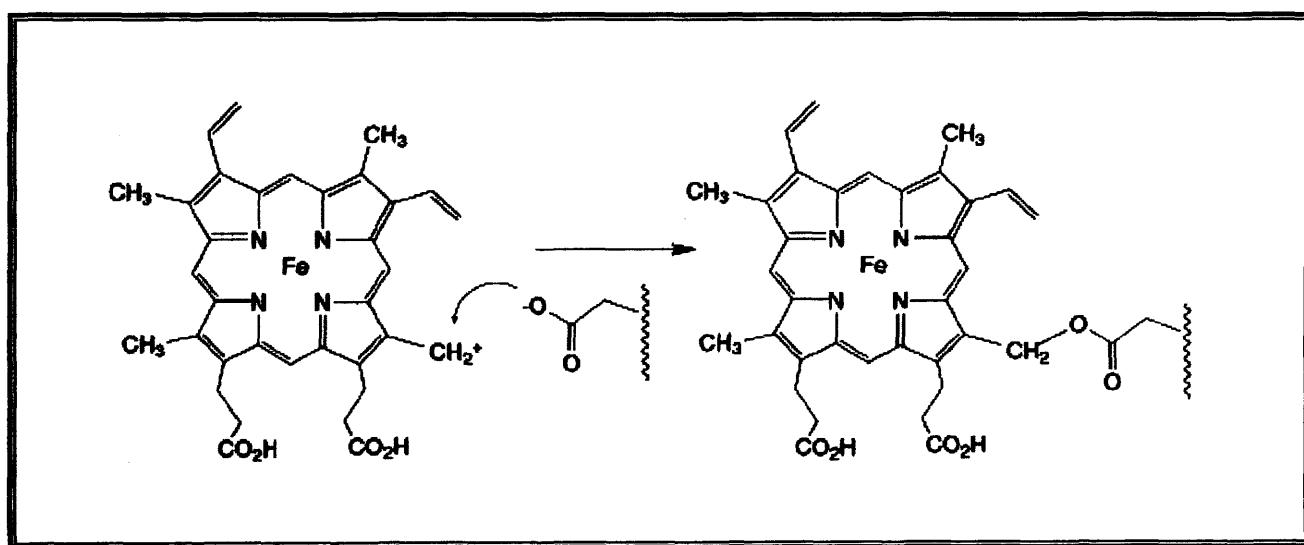
## Chapter 4

### **Crystal structures of Cytochrome P450 BM3 A264E mutant heme domain**

#### 4.1. Introduction

Cytochrome P450 BM3 has been studied as a paradigm system for the membrane-associated, eukaryotic microsomal P450 systems. This is due to the fact that BM3 is soluble, uses a eukaryotic-like CPR as the redox partner (as opposed to the two component ferredoxin reductase and ferredoxin systems found in many other bacterial systems) and is a convenient, catalytically self-sufficient fusion protein enzyme (Munro *et al.*, 2002; Narhi and Fulco, 1987; Fulco, 1991). Cytochrome P450 BM3 is furthermore of great interest due to its function as a fatty acid hydroxylase displaying an extremely high rate of oxygenation of long chain fatty acids (e.g. >15,000 turnovers/min with arachidonic acid; Noble *et al.*, 1999). This high catalytic rate is likely due to the efficient electron transfer between the different redox modules afforded by their covalent linkage and hence close spatial organization (Munro *et al.*, 2002).

Cytochrome P450 BM3 heme domain exhibits strong similarity to the fatty acid hydroxylases from the P450 family 4. In several members of this family of fatty acid oxygenase P450s, a conserved glutamate residue in the I helix is known to covalently ligate the heme macrocycle through an autocatalytic, turnover-dependent attachment to the 5-methyl group of the porphyrin ring (Henne *et al.*, 2001; Hoch and Ortiz de Montellano, 2001; Figure 4.1).



**Figure 4. 1 Site of covalent binding between heme and protein in family 4 cytochromes P450 and the proposed mechanism for formation of the ester bond (Taken from LeBrun *et al.*, 2002).**

An alignment of amino acid sequences from various members of the CYP4 family of fatty acid-oxygenating P450 enzymes indicates the corresponding residue in P450 BM3 is alanine 264 (Figure 4.2).

<b>CYP 102</b>	( <i>B. megaterium</i> )	253	<b>IRYQI ITFL<u>I</u>A<u>G</u>HETTSGLLSFALYFLVK</b>	282
<b>CYP 4A4</b>	(Rabbit)	310	<b>LRAEVDTFM<u>F</u>E<u>G</u>HDTTASGVSNIFYALAT</b>	339
<b>CYP 4A1</b>	(Rat)	309	<b>LRAEVDTFM<u>F</u>E<u>G</u>HDTTASGVSNIFYALAT</b>	338
<b>CYP 4B1</b>	(Human)	304	<b>LRAEVDTFM<u>F</u>E<u>G</u>HDTTSGISWFLYCMAL</b>	333
<b>CYP 4C1</b>	(Cockroach)	303	<b>IREEVDTFM<u>F</u>E<u>G</u>HDTTSAGICWALFLLGS</b>	332
<b>CYP 4D1</b>	( <i>Drosophila melanogaster</i> )	305	<b>IREEVDTFM<u>F</u>E<u>G</u>HDTTSSALMFFFYNIAT</b>	334

**Figure 4. 2 Alignment of P450 amino acid sequences in the region of the heme-ligating glutamate in the CYP4 family. The relevant amino acids at this position in the alignment are shown in boldface and are underlined. The amino acid sequences shown are for flavocytochrome P450 BM3 from *B. megaterium* (CYP102A1), rabbit CYP4A4, rat CYP4A1, human CYP4B1, cockroach CYP4C1, *Drosophila melanogaster* CYP4D1.**

In order to investigate the possibility of engineering a similar protein-heme link in the related P450 BM3 heme domain, the A264E mutant was created. Covalent attachment of the heme group to the protein would be of substantial interest to the biotechnological community. The P450 enzymes ability to carry out specific and often unique reactions coupled with a more stable enzyme that is able to tolerate harsh production conditions could lead to a range of industrial P450 usages. Laboratory evolution cycles of random mutagenesis of BM3 carried out by Arnold and co-workers led to the production of a highly catalytic alkane hydroxylase that has industrial potential (Glieder *et al.*, 2002). More recently, random and rational mutagenesis has allowed the production of further novel enzymatic functions for BM3 that may have other biotechnological uses (Munzer *et al.*, 2005; Li *et al.*, 2005; Sulistyaningdyah *et al.*, 2005). A novel use of BM3 as part of a bi-enzyme insecticide sensor has also recently been developed (Waibel *et al.*, 2005). These developments indicate the large potential that BM3 may represent in satisfying specific industrial requirements. However, the stability of cytochrome P450s is likely insufficient due to relatively easy denaturation by chemical, thermal or pressure alterations that are commonly utilised in industrial processes. Loss of activity may be directly linked to heme destabilisation or removal (Perera *et al.*, 2003). The significant similarity between BM3 and the fatty-acid hydroxylases of the CYP4 family suggests that a covalently attached heme may be possible to generate in BM3. The likely increase in the stability of BM3 by this method could have important ramifications in the development and application of useful biotechnological processes.

In this chapter the crystal structures of both substrate-free and palmitoleate bound forms of the A264E mutant in the P450 BM3 heme domain are presented (Joyce *et al.*, 2004). Examination of these protein structures indicate that alanine 264 occupies a key and critical position in the I helix of the P450 with mutation leading to interesting and unexpected

structural changes. The influence of the observed structural changes in the overall understanding of the enzyme mechanism is discussed.

## 4.2. Results

### 4.2.1. A264E mutagenesis, Enzyme Preparation and Crystallisation

The A264E mutant forms of full-length flavocytochrome P450 BM3 and its heme domain were constructed by Hazel M. Girvan (University of Leicester) by overlapping mutagenic PCR as described in Girvan et al., (2004). The A264E mutant of cytochrome P450 BM3 was overproduced in *Escherichia coli* strain TG1 and purified essentially as previously described (Miles et al., 1992).

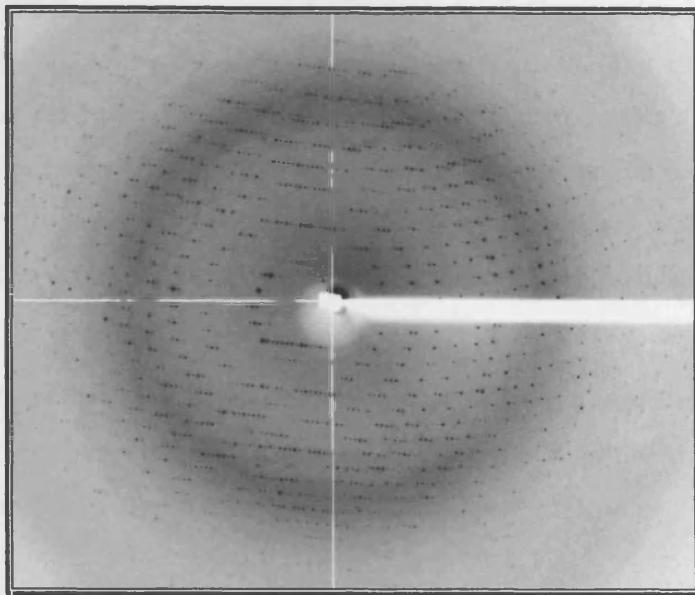
Crystallisation of the palmitoleate bound and unbound A264 heme domain were carried out by Hazel M. Girvan as described in section 3.4.1. Initial crystallisation trials using the published conditions for the wild type BM3 heme domain (Ravichandran et al., 1993) were unsuccessful. The use of manganese sulphate in place of magnesium acetate led to conditions that generated large diffraction-quality crystals. In contrast, crystallisation of the A264E mutant of BM3 heme domain palmitoleate complex was reproducible in the previously published conditions (Li and Poulos, 1997). Large crystals of both the palmitoleate bound and unbound structures were formed after ~7 days with an example of a palmitoleate A264E complex crystal shown in Figure 4.3. Crystals were immersed in 10% polyethylene glycol 200 in mother liquor as a cryoprotectant, before being mounted on a nylon loop and flash-cooled in liquid nitrogen.



**Figure 4. 3 Crystal of the BM3 A264E mutant palmitoleate complex grown in 100 mM magnesium acetate, 20% polyethylene glycol 2000MME, and 100 mM cacodylic acid at pH 6.3.**

#### 4.2.2. Data collection and processing

Data was collected and processed as described in Section 3.4.2 with a typical diffraction image for the substrate free BM3 A264E mutant shown in Figure 4.4. Analysis of the diffraction data from the substrate free BM3 A264E mutant crystal by the auto-indexing routine in DENZO was consistent with a primitive orthorhombic crystal system, class 222, with cell dimensions  $a = 61.2$   $b = 118.9$   $c = 146.3$  Å. Scaling and merging of the diffraction data was carried out using SCALEPACK (Otwinowski and Minor, 1997). Diffraction data analysis of the palmitoleate bound BM3 A264E crystals was consistent with the C centred orthorhombic space group class 222 with cell dimensions  $a = 106.5$   $b = 165.7$   $c = 224.7$  Å. Scaling and merging of the data was carried out using SCALEPACK in space group C222<sub>1</sub> identical to that of the wild-type BM3 palmitoleate complex structure (Li and Poulos, 1997). The data collection statistics are summarized in Table 4.1 overleaf.



**Figure 4. 4** A representative 1° rotation frame of data collected from a BM3 A264E mutant crystal on a 300 mm MAR image-plate system on beamline 14.1 at the ESRF, Grenoble. Reflections at the edge of the diffraction image show diffraction to 2.0 Å.

#### 4.2.3. Structure Determination by Molecular Replacement

Molecular replacement was carried out using AMORE (Navazza, 2001) as described in Section 3.4.3 using PDB code: 2HPD (Ravichandran *et al.*, 1993) in order to determine the structure of the substrate free BM3 A264E mutant heme domain. Determination of the number of molecules in the asymmetric unit was carried out using the Matthew's coefficient. The Table below shows the Matthew's values and solvent content depending on the number of molecules in the asymmetric unit. From these calculations the strong indication was that there are probably two molecules in the asymmetric unit.

**Table 4. 1** Calculated solvent content and Matthew's coefficient

No. of molecules/asymmetric unit	Matthew's coefficient	Solvent content
1	5.9	79.1
2	3.0	58.1
3	2.0	37.2
4	1.5	16.2

Searches were carried out in space group P222 and all other permutations within this Laue group. Initial assessment of the systematic absences following data processing indicated that space group P<sub>2</sub><sub>1</sub>2<sub>1</sub>2<sub>1</sub> was the most probable space group. The initial rotation search results are given in Table 4.2 for space group P<sub>2</sub><sub>1</sub>2<sub>1</sub>2<sub>1</sub> with the correct solutions markedly distinct from random noise. The search results for the two translation searches are shown in Table 4.3 for space group P<sub>2</sub><sub>1</sub>2<sub>1</sub>2<sub>1</sub>. The final correlation coefficient between F<sub>o</sub> and F<sub>c</sub> and R<sub>factor</sub> for the solutions after the rigid body refinement module of AMORE was 60.1 and 37.1 respectively.

The A264E mutant of BM3 palmitoleate complex has the same space group as the native palmitoleate complex previously determined and the structure was simply solved by difference Fourier methods giving rise to a starting R factor of 32.7

**Table 4. 2 Data collection, scaling and model refinement statistics for the BM3 A264E mutant with and without substrate.**

<b>Data Collection</b>	<b>Protein</b>	
	A264E mutant of BM3	A264E mutant palmitoleate complex
Space group	P 2 <sub>1</sub> 2 <sub>1</sub> 2 <sub>1</sub>	C 222 <sub>1</sub>
Cell parameters (Å)		
<i>a</i>	61.2	106.5
<i>b</i>	118.9	165.7
<i>c</i>	146.3	224.7
X-ray source	ESRF ID14-EH1	ESRF ID14-EH1
Resolution	15.0–2.0	15.0–2.75
No. of observations		
Total	686,004	787,601
Unique	68,550	53,600
Completeness (%)	94.0 (97.1)	99.9 (99.6)
R <sub>merge</sub>	5.1 (44.1)	6.9 (49.1)
I/σI	28.2 (2.41)	30.9 (3.7)
<b>Refinement</b>		
R <sub>work</sub>	0.227	0.257
R <sub>free</sub>	0.288	0.338
RMSD from ideal		
Bonds (Å)	0.018	0.034
Angle (°)	1.719	2.933

**Table 4. 3 Rotation search results for BM3 A264E mutant using AMORE. Solution 1 and 2 are clearly above background**

Solution no.	alpha	beta	gamma	Correlation of $F_o$ and $F_c$	R factor	Correlation of $I_o$ and $I_c$	Correlation coefficient of Patterson maps
1	125.34	72.76	331.75	12.6	53.0	23.6	8.3
2	113.82	55.08	320.07	12.3	52.8	22.1	12.2
3	136.65	88.03	265.18	8.0	54.2	14.7	6.3
4	125.40	54.38	215.32	8.0	54.1	14.7	6.2
5	64.24	66.53	141.75	8.0	54.3	14.8	7.1

**Table 4. 4 Translation search results for BM3 A264E mutant using AMORE in space group P212121. Solution 1 is clearly above noise level in both cases.****Translation search 1**

Solution number	Alpha	Beta	Gamma	X	Y	Z	Correlation of $F_o$ and $F_c$	R factor	Correlation of $I_o$ and $I_c$
1	125.34	72.76	331.75	0.1189	0.3416	0.1098	34.2	48.0	32.1
2	125.34	72.76	331.75	0.1196	0.3399	0.4434	24.3	50.5	25.6
3	125.34	72.76	331.75	0.1195	0.3404	0.2938	23.3	51.2	24.8

**Translation search 2**

1	113.82	55.08	320.07	0.1100	0.8947	0.0578	58.2	38.1	58.4
2	113.82	55.08	320.07	0.6096	0.3941	0.0577	34.2	46.7	32.7
3	113.82	55.08	320.07	0.6051	0.8938	0.5582	33.6	46.9	33.2

**4.2.4. Model Building and Refinement****4.2.4.1. Building and Refinement of BM3 A264E mutant**

An  $R_{\text{free}}$  test set containing 5 % of the measured reflections was selected for use in validation of refinement as discussed in Section 2.8 using FREERFLAG (Brünger, 1992). Using REFMAC5 (Murshudov *et al.*, 1997) for initial refinement of the molecular replacement atom coordinates an output file was generated with coefficients for sigmaA weighted  $F_o$ - $F_c$  and  $2F_o$ - $F_c$  difference maps. The maps and molecules were visualised using TURBO-FRODO (Roussel and Cambillau, 1991) and model building and map fitting were

carried out. Manual rebuilding was alternated with multiple cycles of refinement using Refmac5 with the weighting for stereochemical restraints versus crystallographic data set at 0.15. In the final stages of refinement, addition of solvent molecules at positions of the peaks which were at least  $3\sigma$  in the sigmaA weighted  $F_o$ - $F_c$  difference map was carried out using Arp/Warp (Perrakis *et al.*, 2001). Manual removal of all water molecules with B values larger than  $60\text{ \AA}^2$  and greater than  $3.5\text{ \AA}$  distance from hydrogen bond donor or acceptor atoms was carried out. While the majority of the protein molecule was built; there were regions of poor electron density where no protein model could be accurately built.

#### **4.2.4.2. Building and Refinement of BM3 A264E mutant palmitoleate complex**

$F_o$ - $F_c$  and  $2F_o$ - $F_c$  maps were generated using REFMAC5 and the maps and molecules were visualised using TURBO-FRODO and model building and optimal map fitting were carried out. Initial examination of the density in the heme region showed the E264 residue ligating to the heme iron atom in all four molecules in the asymmetric unit. Manual rebuilding was alternated with multiple cycles of refinement using Refmac5 with the weighting for stereochemical restraints versus crystallographic data set at 0.10 due to the relatively low resolution of the data. Because of the low resolution no water molecules were introduced into the model.

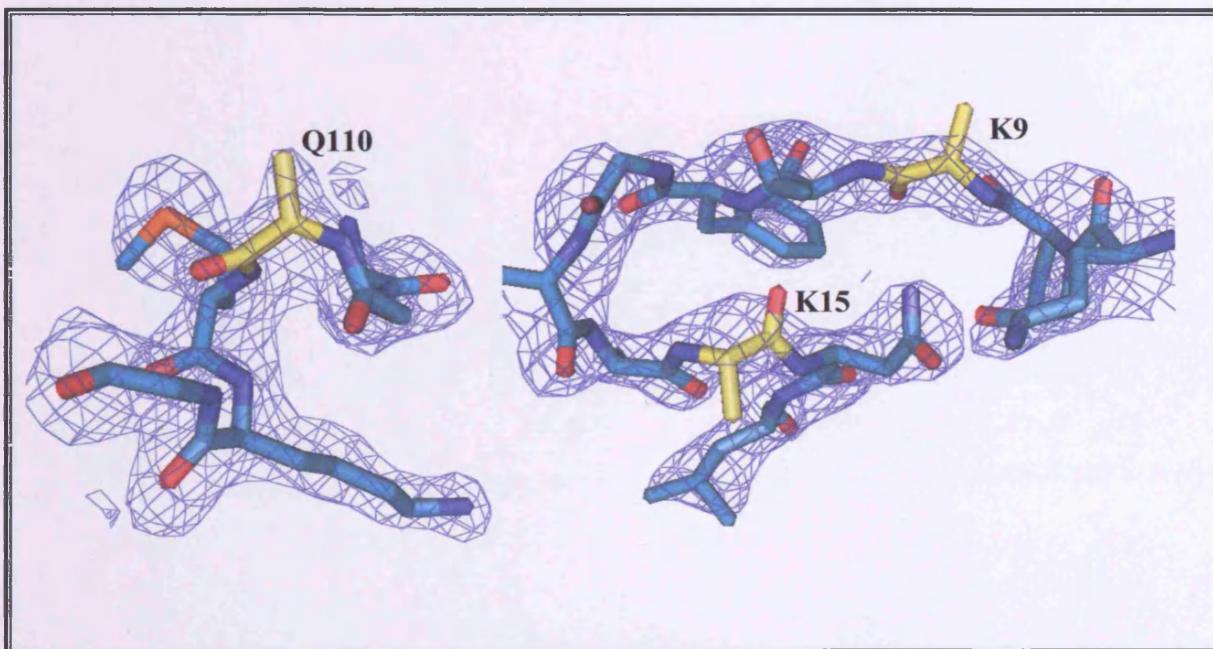
### **4.3. Analysis of the BM3 A264E mutant and palmitoleate complex structures**

#### **4.3.1. BM3 A264E mutant**

The final model of the substrate free BM3 A264E mutant contains two molecules per asymmetric unit. Molecule A contains 441 of the expected 458 amino acid residues. There is no interpretable density for residue 1, residues 193-203 and residues 226-230. The side chains for several residues have been modelled up to the beta carbon because of poor side chain electron density. Molecule B contains 455 of the expected 458 amino acid residues. There is no interpretable density for residue 1 or residues 227 and 228. This model gives a final R-factor of 0.227 and a  $R_{\text{free}}$ -factor of 0.288. The overall conformation of both molecules is similar (RMSD of  $0.383\text{ \AA}$  for all C- $\alpha$  atoms). Omitting residues 1-30 and residues 105-110

due to conformational differences in these loop regions gives an RMSD of 0.335. The average B factor for the 3500 protein atoms in molecule A is  $41.86 \text{ \AA}^2$  and for the 3624 protein atoms in molecule B is  $50.682 \text{ \AA}^2$ . The difference in B factor is due to increased crystal packing restraints for molecule A. The average B factor for water molecules is  $50 \text{ \AA}^2$ . Overall B factor for the 7702 atoms is  $46.410 \text{ \AA}^2$ .

Analysis of the stereochemical quality of the models was achieved using PROCHECK (Laskowski *et al.*, 1993). The overall stereochemistry for the BM3 A264E mutant structure gives values for the RMSD from standard values of the bond lengths of  $0.018 \text{ \AA}$  and angle of  $1.719^\circ$ , respectively. The Ramachandran plot given in Figure 4.7 shows that 99.6 % of non-glycine residues have conformational angles ( $\phi, \psi$ ) in permitted regions and 88.9 % of these in most favoured regions. Three residues are highlighted (Q110 in chain A and K9 and K15 in chain B) as existing in disallowed regions of the Ramachandran plot; examination of these residues however shows that they are a reasonable fit to the electron density in these areas (Figure 4.5). All of these residues have their side chains curtailed at the C- $\beta$  indicative of the poor electron density.

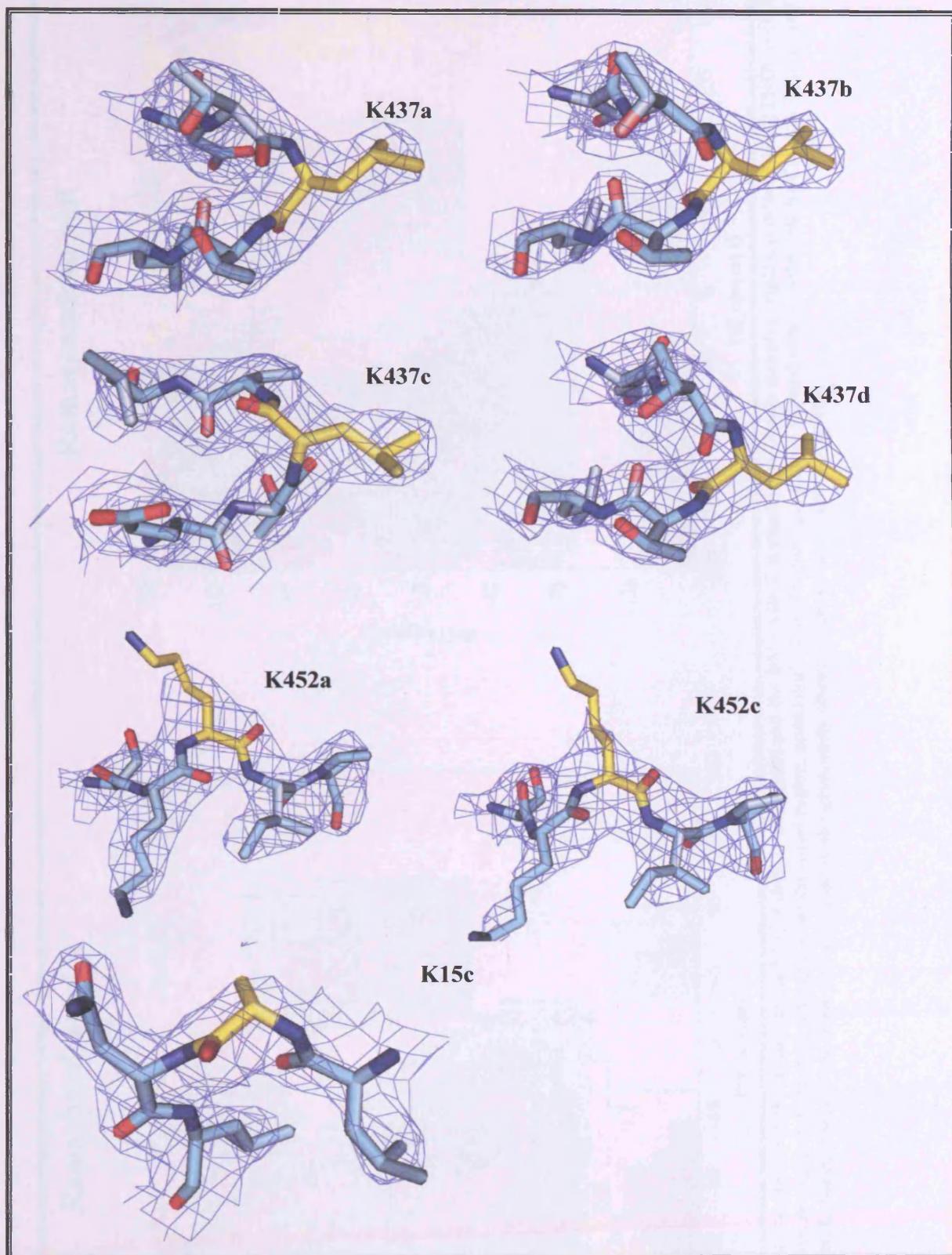


**Figure 4.5** Electron density surrounding the Ramachandran plot outliers for the A264E mutant of BM3 heme domain as calculated by PROCHECK. Residues are shown in stick format with the Ramachandran outlier residues shown with yellow carbon atoms. A  $2Fo - Fc$  map contoured at  $1\sigma$  is shown in blue.

#### 4.3.2. BM3 A264E mutant palmitoleate complex

The final model of the BM3 A264E mutant palmitoleate complex structure contains four molecules per asymmetric unit. Molecules A-D contains all of the expected 458 amino acid residues. This model gives a final R-factor of 0.257 and a free R-factor of 0.338. The average B factor for molecule A is  $47.4 \text{ \AA}^2$ , molecule B is  $59.1 \text{ \AA}^2$ , molecule C is  $58.8 \text{ \AA}^2$  and in molecule D is  $62.1 \text{ \AA}^2$ . The overall B factor for the 14,913 atoms is  $56.6 \text{ \AA}^2$ .

Analysis of the stereochemical quality of the protein model was carried out using PROCHECK (Laskowski *et al.*, 1993). The overall statistics for the RMSD from standard values of the bond lengths of  $0.034 \text{ \AA}$  and angle of  $2.933^\circ$ , respectively. The Ramachandran plot (Figure 4.7) show that 98.4 % of non-glycine residues have conformational angles ( $\phi, \psi$ ) which fall within permitted regions. From all four molecules/asymmetric unit only seven residues are highlighted (K452 from molecule A, K15 and K452 from molecule C and K437 from all four molecules) as existing in disallowed regions of the Ramachandran plot; examination of these residues however shows that they have a good fit to the electron density (Figure 4.6).



**Figure 4.6** Electron density surrounding the Ramachandran plot outliers for the A264E mutant of BM3 heme domain-palmitoleate complex as calculated by PROCHECK. Residues are shown in stick format with the outlier residues shown with yellow carbon atoms. A sigmaA weighted  $2F_o - F_c$  map contoured at  $1\sigma$  is shown in blue.

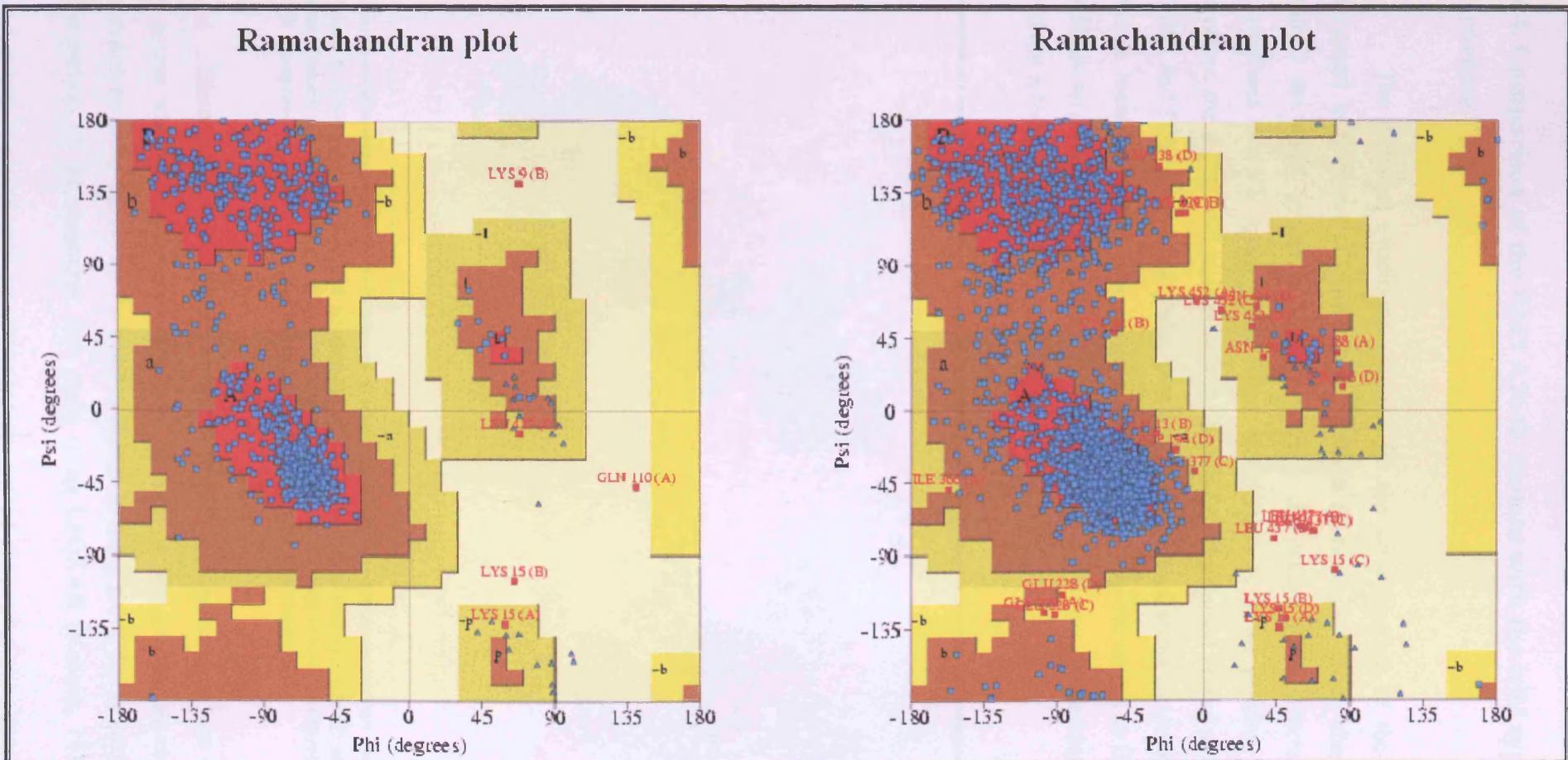
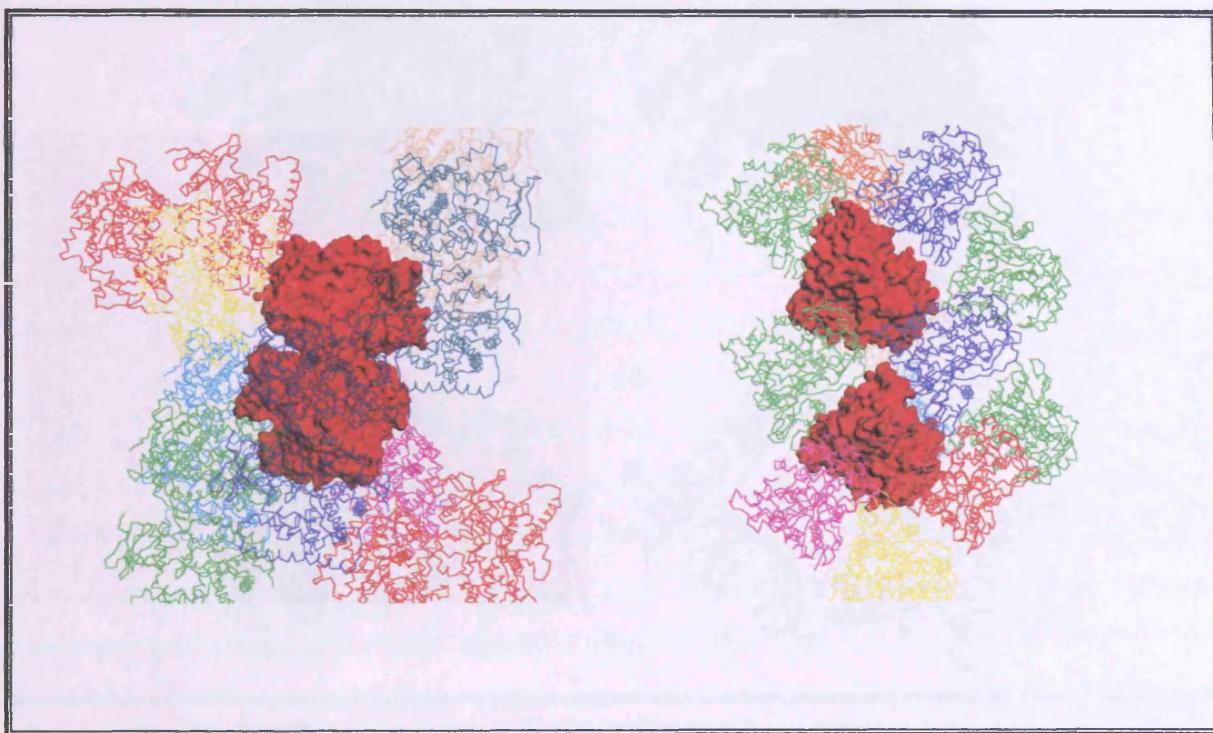


Figure 4.7 Ramachandran plots for the models of the BM3 A264E mutant (left) and the BM3 A264E mutant-palmitoleate complex (right), produced using PROCHECK. The coloured areas indicate, in order of decreasing intensity, most favoured region, additional allowed region, generously allowed region and disallowed region. Glycine residues are shown as triangles. Labelled residues are those that are found in the generously allowed region and disallowed regions.

#### 4.4. Comparison of the BM3 A264E mutant with the wild type BM3 structure

The different conditions required for the crystallisation of the A264E mutant as discussed in Section 4.2.1. can be rationalised by the space group change. In the A264E mutant the space group is  $P_{2_1}2_12_1$  whereas the native BM3 structure was previously crystallised in a  $P_{2_1}$  space group. There are differences in the packing between the native structure and the A264E mutant with the previously determined crystal structures (PDB code: 2HPD and PDB code: 1JPZ) exhibiting a  $132 \text{ \AA}^2$  interface along the proximal surface of the protein between the two protein molecules in the asymmetric unit. The BM3 A264E mutant exhibits no such interface as shown in Figure 4.8 which shows the packing in both the native and the A264E mutant of the BM3 heme domain.

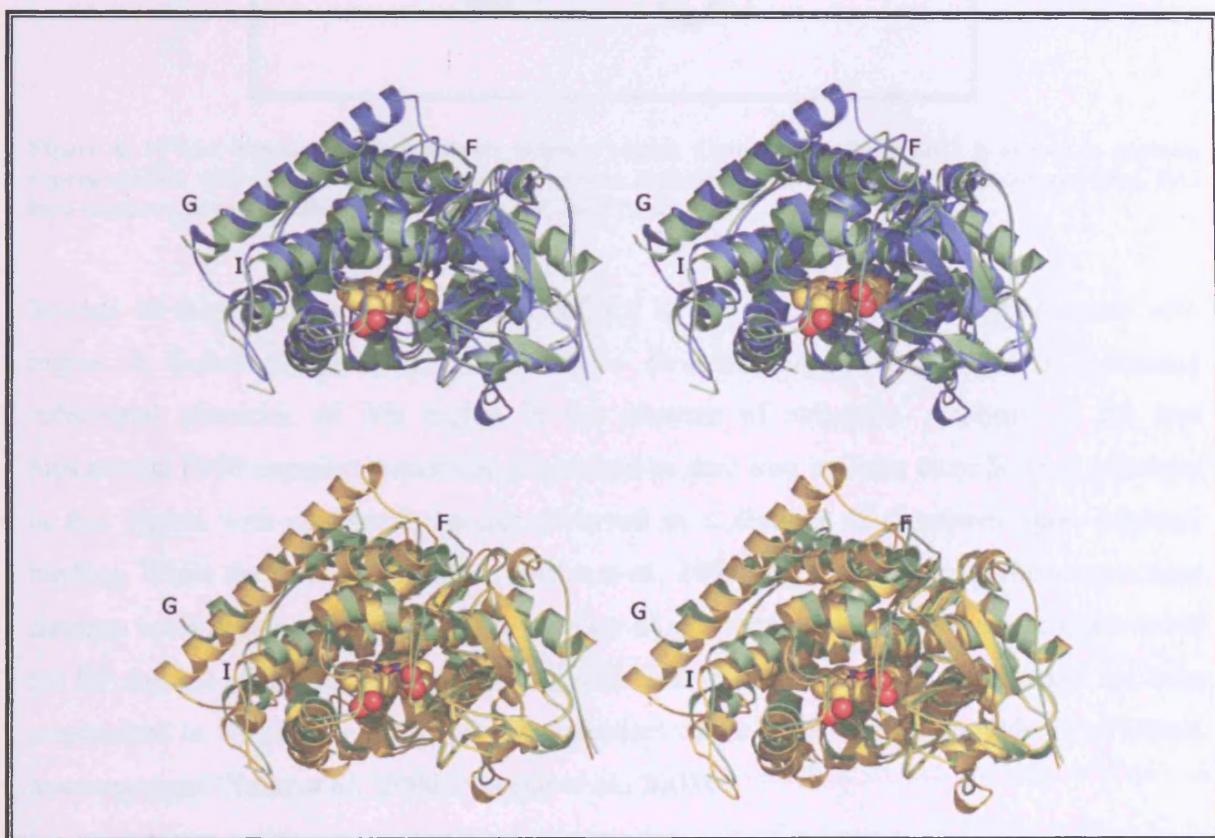


**Figure 4. 8 Symmetry generated neighbours of the native (left) and the A264E mutant (right) of BM3 heme domain. The molecules in the asymmetric unit are shown with a surface representation coloured red with the symmetry generated neighbours shown in ribbon representations.**

Comparison of the A264E mutant with the wild type enzyme was carried out in order to assess whether any overall structural changes in the protein structure caused by the mutation was responsible for the marked change in space group and crystallisation conditions. Comparison of the structures was made using LSQKAB (Kabsch, 1976) using only the

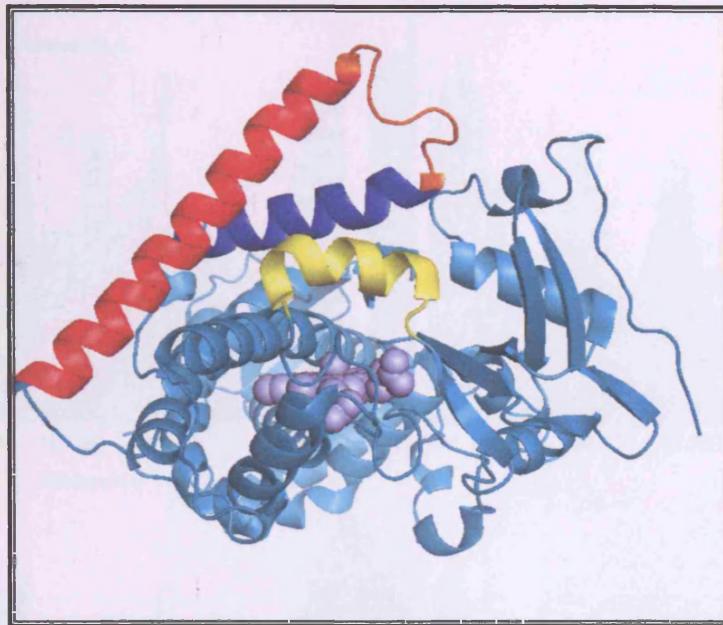
structurally invariant residues, as described by Haines *et al.*, (2002) which represent approximately 62% of the structure.

This overlay of the structures (Figure 4.9) led to an unexpected finding that the A264E structure is structurally similar (RMSD 0.54 Å for all C $\alpha$ ) with the substrate-bound form of the wild-type enzyme (PDB code: 1FAG; Li and Poulos, 1997) and shows a significant difference (RMSD 1.37 Å for all C- $\alpha$ ) with the substrate-free wild-type structure (PDB code: 2HPD, Ravichandran *et al.*, 1993). No substrate was added to the A264E mutant during either purification or crystallization, nor could any substrate be observed in the electron density maps indicating that the A264E mutation had in some way altered the overall protein structure.



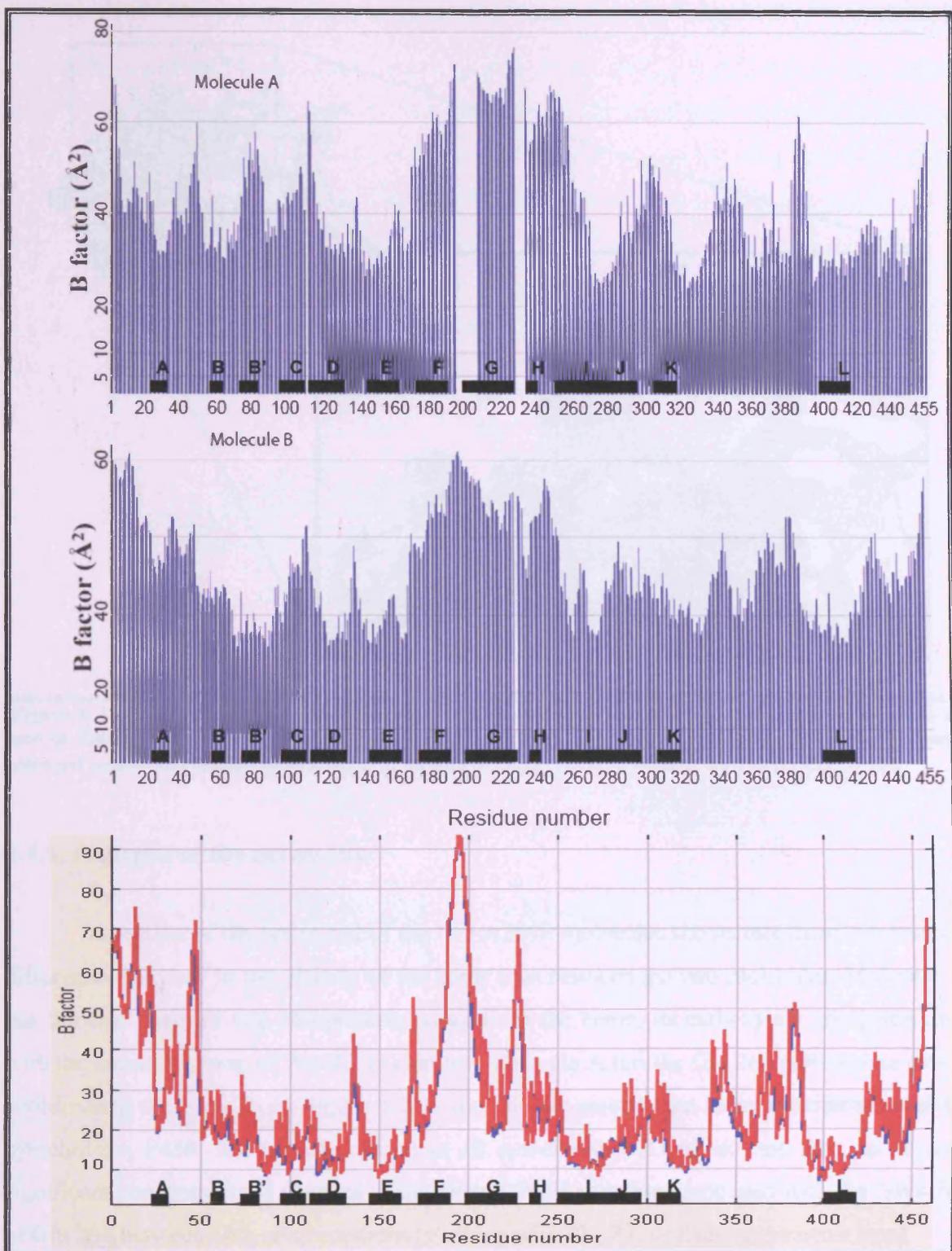
**Figure 4.9** Stereo view of an overlay of the substrate-free A264E heme domain (green ribbons). *Upper*, substrate-free form (blue) of the wild-type P450 BM3 heme domain; *lower*, substrate-bound form (yellow) of the wild-type P450 BM3 heme domain. The heme cofactor is shown in sphere representation.

Similar to the changes seen upon substrate-binding in wild-type P450 BM3, the majority of residues that are in significantly different positions between the substrate-free structures of A264E and the wild-type heme domain are located in the so called "lid domain" of the substrate access channel, which consists of the F and G helices, the loop between them and the B'-helix (Figure 4.10; Li and Poulos, 1997).

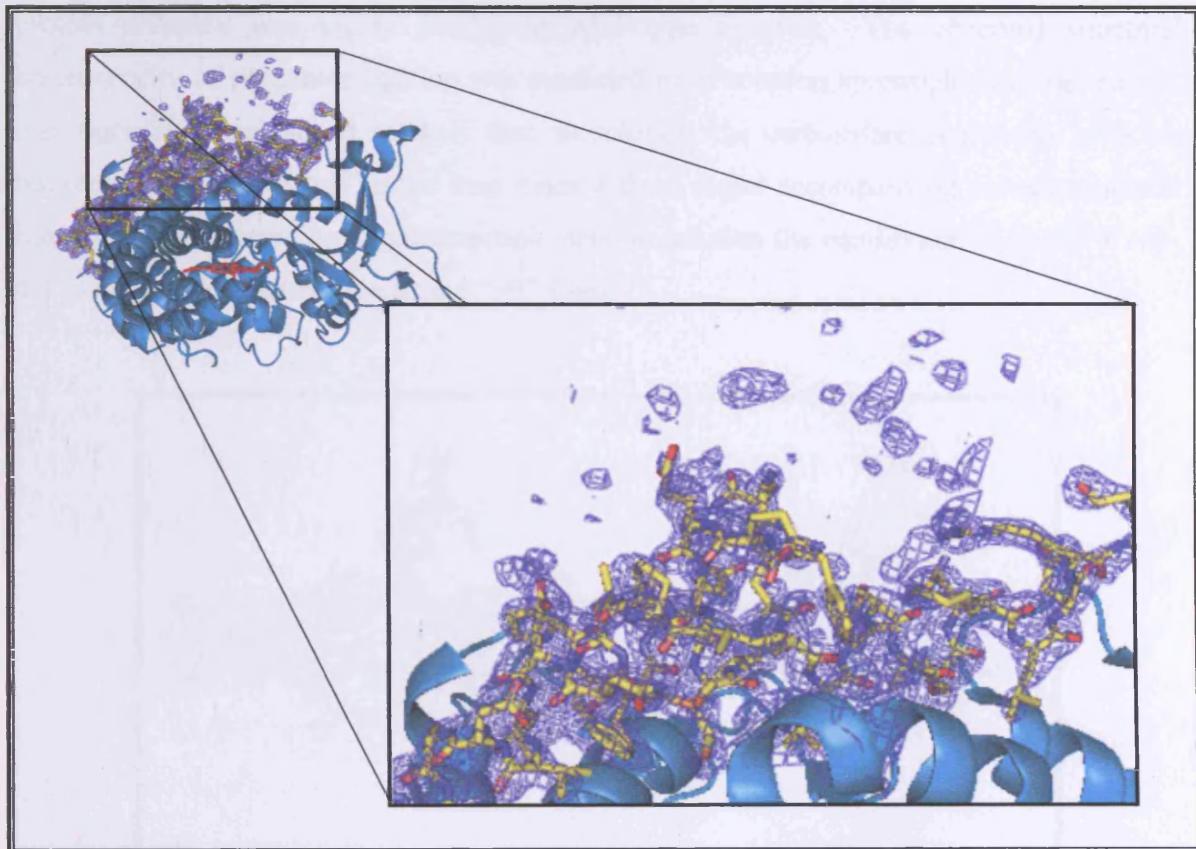


**Figure 4. 10 Lid domain of the substrate access channel. Cytochrome P450 BM3 is shown in cartoon representation with the heme group shown in sphere representation with the F helix coloured blue, F/G loop coloured orange, G helix coloured red and the B' helix coloured yellow.**

Several of these residues are less well defined in the BM3 A264E mutant structure with higher B factors (Figure 4.11) and/or poor electron density (Figure 4.12), indicating substantial plasticity of this region in the absence of substrate. Analysis of the five prokaryotic P450 complex structures determined to date also indicate there is large plasticity in this region with structural changes observed in a number of structures upon substrate binding. While the P450<sub>cam</sub> structure (Poulos *et al.*, 1985) does not show significant structural changes upon substrate binding, in the absence of substrate, higher B factors are observed in the lid domain of the substrate access channel. The importance of this lid domain has been emphasised in the other substrate bound structures where these regions do undergo structural rearrangement (Yano *et al.*, 2000; Nagano *et al.*, 2003).



**Figure 4.11** Graph of the B factor values of molecule A and B of the A264E mutant of BM3 heme domain. The B factor values of the native heme domain of cytochrome P450 BM3 (PDB code: 2HPD) is shown for comparison. The helix boundaries are outlined by black blocks and the helix nomenclature is given. The lid domain is made up of the B' helix (residues 72-83) and the F/G helical region (residues 170-230).



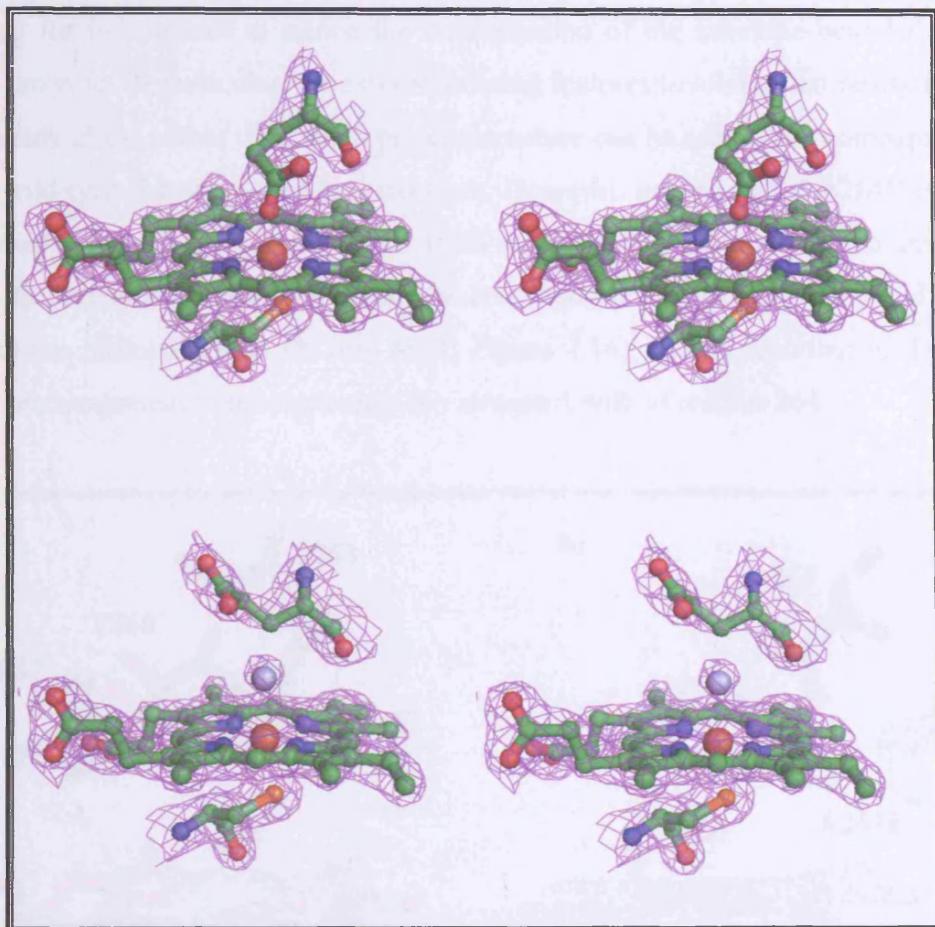
**Figure 4.12** Structure of the A264E mutant of BM3 shown in cartoon format in the same orientation as seen in Figure 1.2 with a close up of the F/G helices shown in stick representation with carbon atoms coloured yellow surrounded with a sigma A weighted  $2F_o - F_c$  electron density map contoured at  $1\sigma$ .

#### 4.4.1. Analysis of the active site

Inspection of the active site of the two A264E molecules shows that there is a marked difference observed in the vicinity of the heme iron between the two molecules. Molecule B has the side chain of Glu 264 pointing away from the heme, its carboxylate group stacking with the aromatic group of Phe 87. In contrast, molecule A has the Glu 264 carboxylate group coordinating the heme iron (Figure 4.13). The thiolate-carboxylate heme ligation is novel to cytochromes P450, and to date unique in all cytochromes. There do not seem to be any significant conformational changes to the overall P450 structure associated with the “switch” of Glu 264 between both conformations (stacking with Phe 87, or ligating the heme iron).

Extensive solution analysis of the A264E mutant full length and heme domain of BM3 were carried out by Hazel M. Girvan. Spectroscopic analysis of the A264E mutant indicated that there was additional protein interaction with the heme group as evidenced by a change in the Soret band maximum. Substrate binding caused different spectral changes than the wild type enzyme indicating structurally different substrate complexes (Girvan *et al.*, 2004). However, inhibitor binding studies gave identical results demonstrating that the inhibitor

protein complex was highly similar to wild type enzyme. The observed structural heterogeneity of glutamate ligation was predicted from solution spectrophotometric studies, and from EPR analysis. It is likely that, in solution, the carboxylate continually switches between the “on” and “off” heme iron states without major accompanying protein structural reorganisation. According to spectroscopic data, in solution the equilibrium is poised at ~ 3-4:1 in favour of the heme iron ligand “off” form.



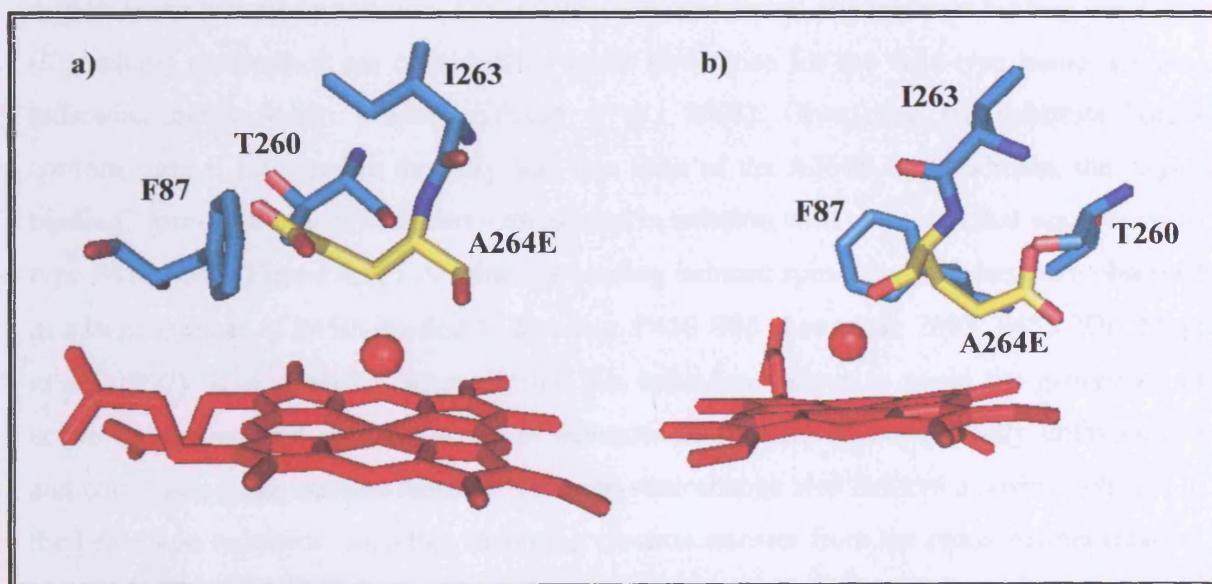
**Figure 4.13** Stereo view of the conformations of Glu264 in molecules A and B of the substrate-free A264E heme domain of Flavocytochrome P450 BM3. The heme macrocycle along with Cys400 and Glu264 is displayed for molecule A (*upper*) and molecule B (*lower*) with the  $2F_o - F_c$  map contoured at  $1\sigma$  superposed over the ball and stick models.

The observed glutamate stacking interaction with Phe 87 is particularly interesting, given the fact that this residue is absent from the CYP4 enzymes in which the covalent heme ligation has been observed previously. In wild-type P450 BM3 Phe 87 interacts with the  $\omega$ -methyl group of fatty acid substrate(s), and is considered to be a critical regulatory residue that controls regioselectivity of substrate oxygenation (Li and Poulos, 1997; Oliver *et al.*, 1997; Munro *et al.*, 2002). A particular difference in the behaviour of P450 BM3 with respect

to eukaryotic CYP4 enzymes is the inability of the former to hydroxylate at the  $\omega$ -position (Okita and Okita, 2001).

#### 4.5. Implications for P450 mechanism

Close examination of the A264E structure reveals no clear direct structural explanation for this mutant to mimic the conformation of the substrate-bound form of the wild-type enzyme. In particular, no extra stabilizing features involving the newly introduced glutamate side chain within the overall protein structure can be seen when compared with the available wild-type substrate complex structures. However, modelling the A264E mutation in the previously determined substrate free BM3 wild type structures leads to severe steric clashes with several other residues within the active site (e.g. Phe 87, Thr260 and Ile263, in addition to the protoporphyrin IX ring itself; Figure 4.14), clearly resulting in the need for structural rearrangement to accommodate the increased bulk of residue 264.



**Figure 4. 14** Stick representation of the active site of native BM3 (PDB code: 2HPD) with modelled A264E mutation without any change to the position of the main chain atoms. Two distinct typical glutamate side-chain rotamer conformations of the modelled E264 mutation are shown indicating the steric clashes with F87 (a) and T260 (b, view rotated by 90°) that prevent accurate modelling of the mutation.

It therefore seems likely that the A264E mutation does not particularly stabilise the substrate bound conformation, but rather destabilises the substrate free conformation of the enzyme, to the extent that the substrate bound conformation, even in absence of the substrate, is preferred. In light of the above observations, it is interesting to note that in the substrate-

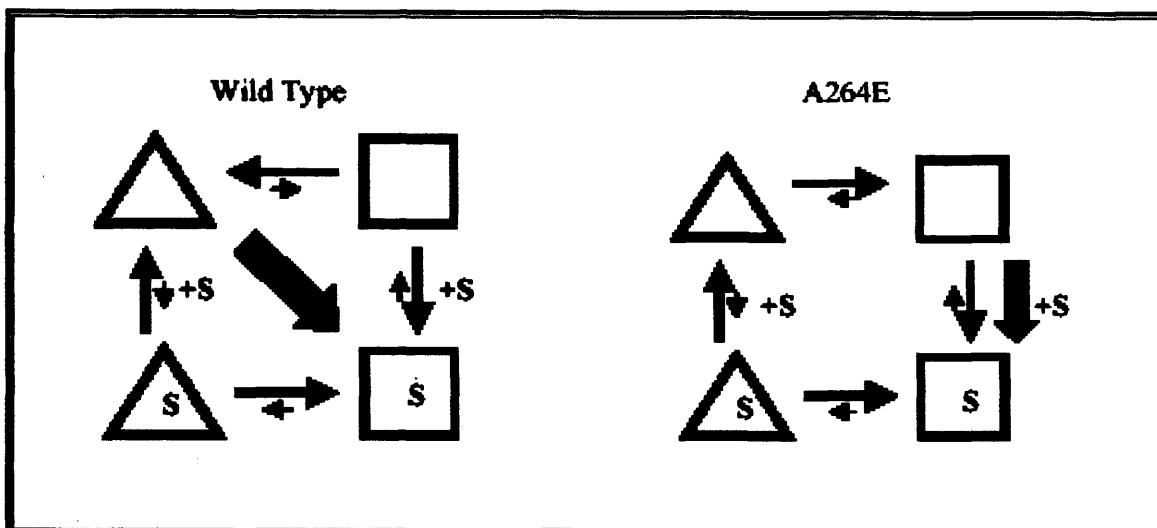
bound wild-type P450 BM3 structures the substrates are in close contact with the side chain of Ala 264. The main chain backbone of the alanine hydrogen bonds to the distal heme ligand water molecule in the unbound form and this water molecule is substantially shifted upon substrate binding (Li and Poulos, 1997; Ravichandran *et al.*, 1993). This suggests the mechanism by which substrate-binding switches P450 BM3 to a different conformational state does not simply involve the expulsion of water molecules from the substrate-binding cavity, but in addition exploits the force exerted by the substrate on Ala 264, a residue that acts as a sensitive trigger for the conformational conversion. In the A264E mutant, unfavourable steric interactions of the glutamate side chain induce the switch to the substrate bound conformation, without the necessity for substrate interaction.

An alternative explanation is that the P450 BM3 heme domain is in rapid conformational equilibrium between substrate bound and substrate free forms, with the equilibrium being strongly favoured towards the substrate free form for wild-type P450 BM3 in the absence of substrate. In this model, the substrate binds preferentially to the substrate bound conformation, effectively shifting the equilibrium towards this form as substrate concentration is increased. This model is also consistent with the behaviour observed for the A264E heme domain in solution. For several substrates tested, the apparent binding constants ( $K_D$  values) determined are considerably lower than those for the wild-type heme domain, indicating much tighter binding (Girvan *et al.*, 2004). Given that the substrate bound conformation is favoured in the fatty acid-free form of the A264E heme domain, the “tight-binding” form of the protein is over-represented in solution with respect to that seen for wild-type P450 BM3 (Figure 4.15). A substrate-binding induced spin-state shift has been observed in a large number of P450s studied to date (e.g. P450 BS $\beta$ , Lee *et al.*, 2003; P450 2D6, Modi *et al.*, 1997). It is generally accepted that this behaviour serves to avoid the generation of active oxygen species in the absence of substrate that would be energetically unfavourable and could also cause cellular damage. The spin-state change also induces a positive change in the heme iron reduction potential, favouring electron transfer from the redox partner (Daff *et al.*, 1997; Sligar, 1996) and subsequent catalytic activity.

#### **4.5.1. Analysis of the spin state conversion mechanism**

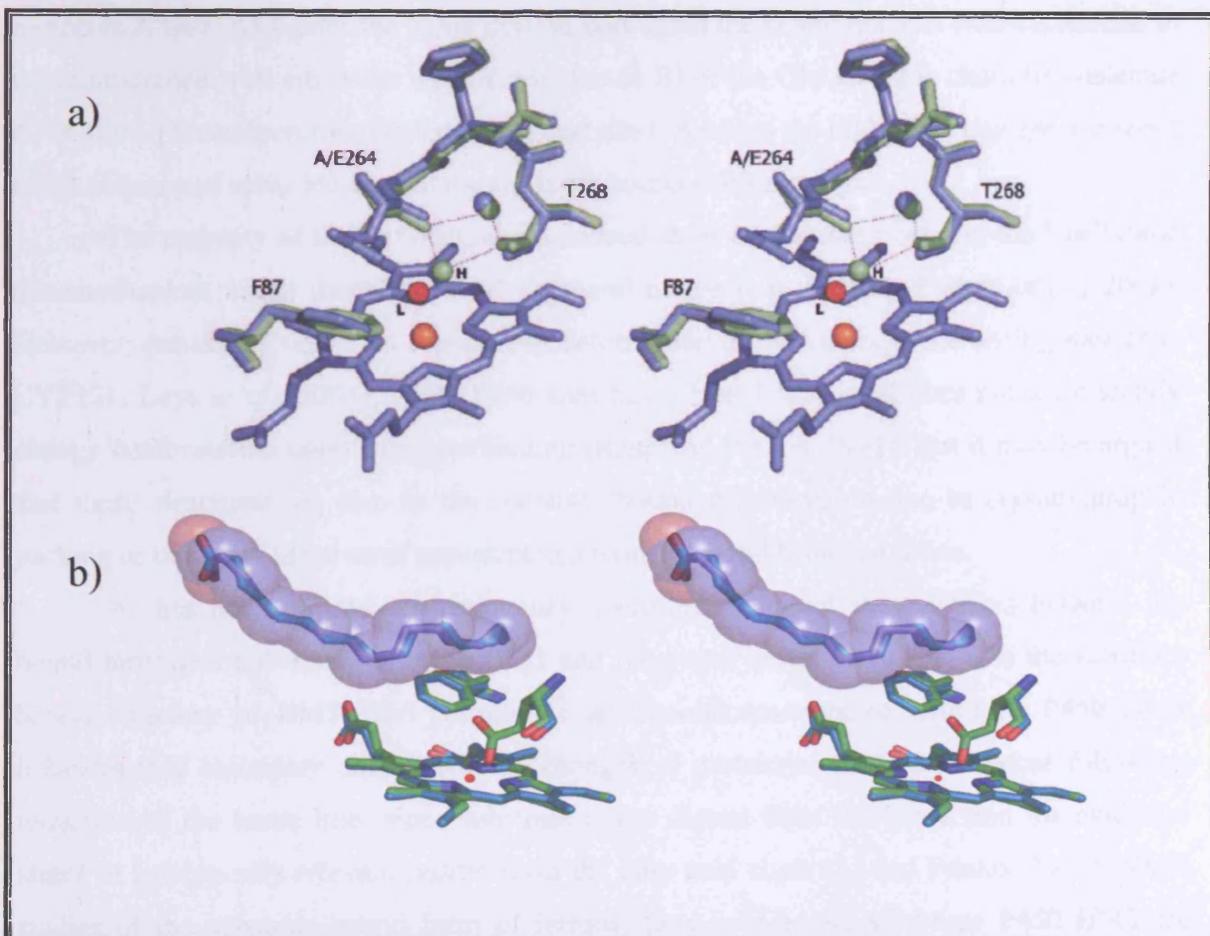
The molecular mechanism whereby substrate-binding induces this shift seems to be variable in P450s studied to date, although the substrate-binding induced displacement of water as the sixth ligand to the heme iron is a common feature (e.g. Li and Poulos, 1997; Cupp-Vickery *et al.*, 2000). The displacement of the distal ligating water in P450<sub>cam</sub> is

specifically due to direct steric hindrance with the camphor molecule (Poulos *et al.*, 1987). In contrast, P450 BM3 undergoes large-scale conformational changes upon binding of fatty acids, and these changes have been proposed to drive the conversion from a hexa- to a penta-coordinated heme group without direct substrate interaction.



**Figure 4. 15** Scheme representing a model for the conformational equilibrium of P450 BM3 in solution and the influence of substrate binding. *Left*, wild-type equilibrium; *right*, A264E equilibrium. The equilibrium in the absence of substrate is drastically changed by introduction of the A264E mutation. In both cases, the *large grey arrows* indicate the apparent changes upon substrate binding.

Heme iron spin-state may be due to either a direct displacement of the distal ligand water molecule (as in P450 cam) or a more indirect displacement via substrate-induced changes in the protein structure (as for P450 BM3, Haines *et al.*, 2001). The structure of fatty acid-free A264E shows all the structural hallmarks of a substrate-bound BM3 enzyme, but in the absence of substrate. An intriguing difference regarding A264E is the fact that, in contrast to substrate-bound wild type-enzyme, A264E does not show any significant high-spin character, in either solution or crystalline state. A low-spin configuration would be expected for the glutamate-ligated species, but both spectroscopic and crystallographic studies show that there is a large population of non glutamate-ligated protein in substrate-free A264E heme domain and that this species does not lose the aqua ligand and convert to high-spin despite the change to the substrate bound conformation. This suggests strongly that even in wild-type P450 BM3 the spin-state shift is a direct, rather than an indirect, result of substrate binding.



**Figure 4.16** Stereo view of a stick diagram of the active site of A264E mutant of BM3. a) Overlay of the active site of A264E (blue) and N-palmitoylglycine-bound wild-type P450 BM3 (green, PDB code: 1JPZ). For clarity, the heme macrocycle is displayed only for the A264E mutant. The pattern of hydrogen bonding to site H is indicated by the dotted lines. The ligating water molecule occupying site L in the A264E mutant is colored red. b) Active-site structure of the palmitoleic acid-bound form of the A264E heme domain shown with yellow carbon atoms with the native BM3 heme domain active site shown for comparison with grey carbon atoms with the bound palmitoleate given in purple stick and sphere representation.

The high-resolution N-palmitoylglycine bound wild-type structure (PDB: 1JPZ) has led to the suggestion that substrate-binding induced protein rearrangement creates a new water-binding site (designated site H) adjacent to the heme (Haines *et al.*, 2001). The proposed higher water affinity of site H over the heme iron ligation site L, and the fact that these sites are mutually exclusive (in that only a single water molecule can bind at either site L or H at any given time) leading to the proposal that site H effectively pulls the water molecule away from site L, leading to the observed shift in heme iron coordination state and hence conversion to the high spin form (Haines *et al.*, 2001). The A264E mutant heme domain structure shows all ligands to the H site to be in identical position as for the N-palmitoyl-glycine bound wild-type P450 BM3 structure as seen in Figure 4.16a. However, the water molecule is still occupying site L and ligates to the heme iron in molecule B, while in

molecule A Glu 264 ligates the heme iron. In both cases the H site remains unoccupied due to steric hindrance with either the water (in molecule B) or the Glu 264 side chain (in molecule A) at site L. It can therefore be concluded that site L remains the higher affinity site for water in the absence of substrate even in the substrate bound conformation.

The majority of the P450 structures indeed show a conserved bend in the I helix and this mechanism might therefore be of a general nature (e.g. CYP51, Podust *et al.*, 2001). However, certain P450s do not contain any deformation of the I helix in the resting state (e.g. CYP121, Leys *et al.*, 2003), while P450 cam has a bent I helix that does not significantly change conformation upon substrate binding (Raag and Poulos, 1991). But it may be argued that these structures are also in the substrate bound conformation due to crystallographic packing or this form is the most prevalent in a dynamic equilibrium situation.

As has been pointed out previously, there are no direct steric clashes between the bound fatty acid substrates of P450 BM3 and the water molecule at site L in the substrate bound structure of BM3. The palmitoleic acid-bound structure of wild-type P450 BM3 indicates that secondary conformational changes of protein/substrate must occur following reduction of the heme iron, since substrate is too distant from the heme iron for oxidative attack at catalytically relevant positions on the fatty acid chain (Li and Poulos, 1997). NMR studies of the substrate-bound form of ferrous, fatty acid-bound wild-type P450 BM3 are consistent with a significant reorientation of the substrate in this enzyme form (Modi *et al.*, 1996). It is, however, clear that upon binding of the fatty acid analogue N-palmitoylglycine the surroundings of site L become more hydrophobic, decreasing the water affinity and ultimately shifting the water molecule to predominantly occupy site H. Spectroscopic studies of wild-type P450 BM3 at catalytically-relevant temperatures (by both electronic absorption and resonance Raman) have shown that, even in the presence of apparently saturating concentrations of substrate, an equilibrium exists between the high-spin, five-coordinate, and the low-spin, six-coordinate states of the heme iron. Depending on the nature of the fatty acid substrate used, varying amounts of low-spin heme iron are detected, with shorter chain saturated fatty acids (e.g. lauric acid) being less effective than longer chain ones (e.g. palmitic acid) at effecting the shift in spin state equilibrium towards high-spin (Miles *et al.*, 1992; Boddupalli *et al.*, 1992). This may be due to the hydrophobic effect that these substrates can have on the local environment of the heme iron. This clearly indicates how, even in the presence of substrate, the possibility exists for water remaining bound at the heme iron (site L), and the water affinity of this site is strongly dependent on the nature of the substrate, although all substrate binding ultimately drives the L-H equilibrium towards the H site.

#### 4.6. Examination of the substrate-bound form

In contrast to the requirement for screening novel conditions to obtain suitable crystals of the substrate-free A264E heme domain, the palmitoleate-bound form of A264E was found to crystallise in the same space group as that reported previously for the substrate-bound wild-type P450 BM3 (Li and Poulos, 1997). As observed for the wild-type enzyme, the resolution and the quality of the data obtained for this particular crystal form is rather poor in comparison with the substrate-free enzyme structure. Nevertheless, electron density clearly indicates no major changes between the palmitoleate-bound A264E and palmitoleate-bound wild-type P450 BM3 structures. The single exception is the fact that in all four molecules in the asymmetric unit of the A264E structure, the Glu 264 side chain ligates the heme iron as shown in Figure 4.16b. This finding indicates that the presence of substrate induces movement of the glutamate onto the iron to replace water as the sixth heme ligand, and is completely consistent with the spectroscopic studies reported (Girvan *et al.*, 2004). Addition of long chain fatty acids perturbs the UV-visible absorption spectrum of the A264E heme domain, inducing red shifts of the heme Soret band towards 426 nm at apparent substrate saturation. This is in contrast to the blue shifts (towards 390 nm) observed following substrate addition to wild-type P450 BM3 and reflects the increasing proportion of the glutamate ligated, low-spin form of the A264E mutant following substrate addition, as opposed to the penta-coordinated, high-spin form of the wild-type that accumulates through substrate-induced displacement of the water ligand from the heme iron (Girvan *et al.*, 2004; Oliver *et al.*, 1997).

From the substrate-free A264E structure, it is clear that two conformations are possible for the Glu 264 side chain and solution studies indicate the equilibrium ratio between both states to be strongly dependent on the solution conditions. Upon palmitoleate binding, the substrate effectively occupies the volume of the non heme iron-ligating conformation, driving the enzyme towards a completely ligated state, as observed in solution studies. Specifically, palmitoleate interacts with Phe 87, preventing Glu 264 from occupying the position observed in molecule A in the substrate-free A264E structure. Despite any significant further structural rearrangement induced following palmitoleate binding to the substrate bound conformation of A264E, the substrate does influence directly the heme iron ligation state by minimising the degrees of freedom available to the Glu 264 side chain. Strong heme ligands such as azoles function as potent inhibitors for P450s, and many are used as antifungal drugs to inactivate the sterol demethylase P450 CYP51 (Georgopapadakou, 1998). It is therefore surprising to note that, despite the fact that the substrate-bound oxidised A264E structure shows fully hexa-coordinated heme iron, catalytic turnover for this mutant can still

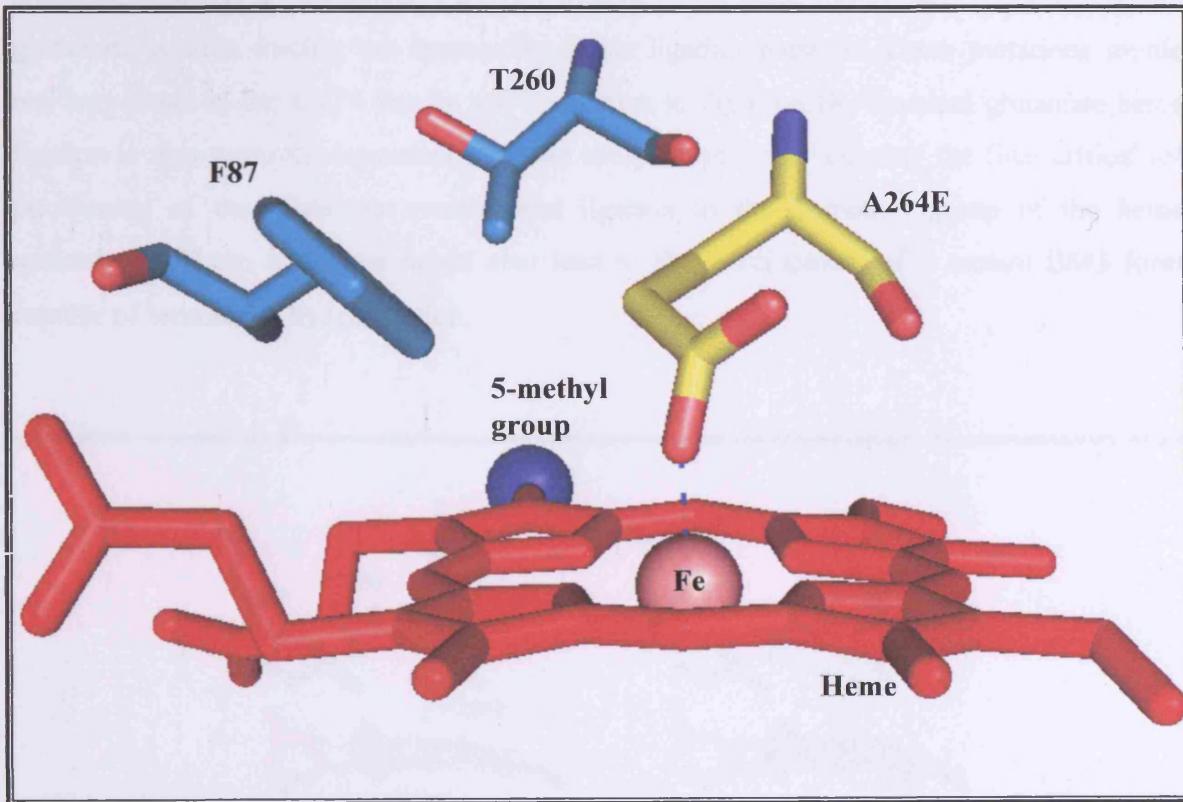
be observed, albeit at lower levels than those observed for wild-type P450 BM3 (Girvan *et al.*, 2004). This is a further indication that upon reduction of ferric iron to the ferrous form the position of the substrate with respect to the heme changes dramatically. A further consideration is that the structure of the enzyme itself changes dramatically, releasing the strong conformational lock on the glutamate side-chain and allowing oxygen to bind to the iron. Inhibitor binding is also reduced but still possible, indicating that the direct heme ligation by 4-phenylimidazole is also feasible.

Presumably reduction of the enzyme commits the enzyme to its “regular” catalytic cycle, and prevents coordination of the glutamate to the iron until it returns to a ferric form following product formation. An interesting aspect of this study arises from the structural change and its effect on thermodynamic properties of the P450. In both the substrate-free and arachidonate-bound forms of the A264E mutant, the reduction potential of the heme iron is  $\sim -315$  mV (Girvan *et al.*, 2004). By contrast, the reduction potential of the conformationally different substrate-free form of P450 BM3 is  $-427$  mV, rising to  $-289$  mV on binding of arachidonate and an extensive switch in spin-state equilibrium towards high spin (Ost *et al.*, 2001). In A264E, the heme iron remains predominantly low-spin in both substrate-free and substrate-bound states. The apparent near identical heme-iron reduction potentials are possibly a consequence of the same species undergoing reduction/oxidation in both cases. That is, the Cys-aqua is reduced preferentially and the Cys-Glu form is drawn towards this state as the reduction titration is carried out leading to the identical potential values. This conformational change and its effects on the electronics of the heme system are of considerable importance in controlling the reduction potential of the heme iron. This is under further study, using A264 variants in which the side chain of the novel amino acid does not ligate the heme iron in substrate-free or substrate-bound forms.

#### 4.7. Possibility of covalent heme ligand generation

Further scrutiny of both the substrate-free and substrate-bound A264E structures reveal other important features of the P450 BM3 structure that relate to attempts to engineer covalent ligation of the heme macrocycle via the interaction of Glu 264 with the heme 5-methyl group. The active site organization in P450 BM3 is such that Phe 87, and likely the I helix residue Thr 260, obstruct access of Glu 264 to the relevant position on the porphyrin ring. The failure to obtain any significant degree of covalent ligation in the A264E mutant may thus be explained by steric restrictions in the active site (Figure 4.17). The possibility of

mutating another residue to form the mutation was assessed but residues neighbouring the 5 methyl heme group are too distant (3.5 Å for I401; 3.65 Å for F261; 4.16 Å for I401).

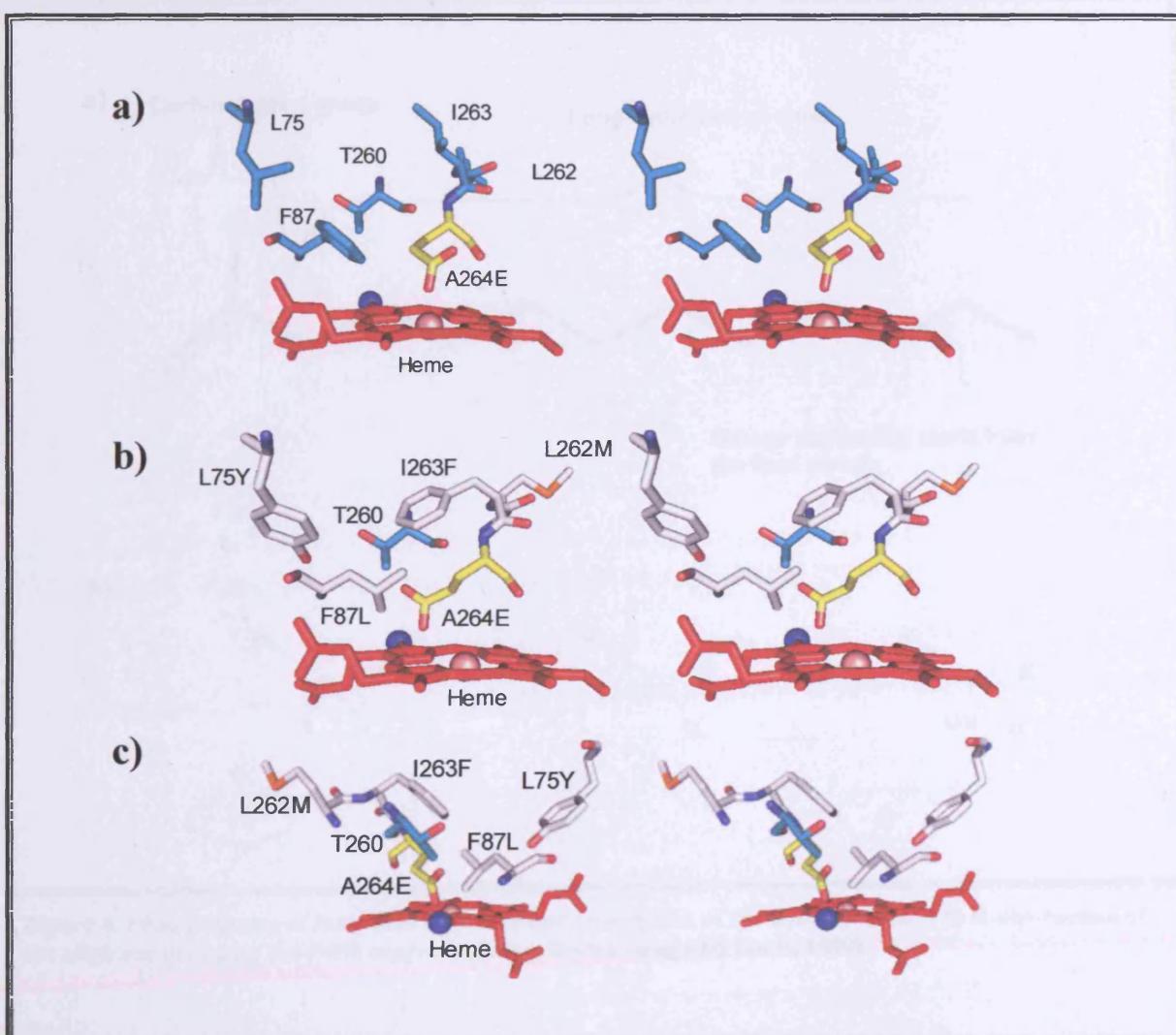


**Figure 4.17** Stick representation of the active site of the A264E mutant of BM3 with the 5-methyl group of the heme shown as a blue sphere.

Analysis of a model of CYP4A11 by Chang and Loew 1999 indicates that there are a large number of steric restraints in the active site caused by Y120, L132, T325, V386 (L75, F87, T260 and A328 in BM3) with major steric pressures exerted on the ligating glutamate residue by Y120. In comparison, the BM3 active site is relatively spacious and analysis of the I helix sequence alignment in Figure 4.2 indicates a conserved FMF sequence prior to the equivalent E264 residue. These bulky residues could contribute to the steric orientation of the ligating glutamate in the CYP4 family (Figure 4.18).

Even if the CYP4 model based on BM3 is inaccurate, common sense dictates that the introduction of multiple bulky residues in the active site will have significant effects and influence the heme ligation. To address the active site restrictions, and in work to enable covalent heme ligation by Glu 264 and produce a more robust and biotechnologically exploitable form of P450 BM3, secondary mutations may facilitate access of Glu 264 to the relevant methyl group allowing the autocatalytic covalent heme linkage to occur. These

mutations include mutations to remove the major hindering residue in a F87L mutation and the introduction of other bulkier residues i.e. L262M and I263F mutants which may cause a slight re-positioning of the I helix and thus move the glutamate residue closer to the 5 methyl group of the heme. Introduction of a L75Y mutation may place further steric pressures on the glutamate residue forcing the appropriate heme ligating position. These mutations mimic residues found in the CYP4 family and are shown in Figure 4.18. Covalent glutamate heme ligation is also turnover dependent with the substrate probably exerting the final critical re-positioning of the glutamate residue and ligation to the 5 methyl group of the heme macrocycle. These mutations might also lead to the development of a mutant BM3 form capable of terminal  $\omega$  hydroxylation.



**Figure 4. 18** Stereo diagrams with stick representations of the proposed mutations to BM3 in order to allow covalent heme ligation. a) active site of A264E mutant of BM3 molecule A. b) and c) Modelled mutations of the active site residues with no positional changes at two different views. L87L, L262M, L75Y and I263F are shown in preferred rotamers positions, the A264E mutant is shown in a possible rotamer with carbons coloured yellow. The heme macrocycle is shown in red, iron atom in a salmon sphere and the 5 methyl group shown as a blue sphere.

These residues are also thought to play a role in the terminal  $\omega$  hydroxylation reaction that is specific to the CYP4 family (Figure 4.19) in contrast to other P450 enzymes. This specific hydroxylation is highly unfavourable in comparison to the  $\omega$ -n ( $n > 0$ ) reactions performed by BM3. Steric and thermodynamic criteria must be satisfied in determining the site of hydroxylation. Experimental and computational studies have demonstrated that it is thermodynamically more difficult to break the stronger C-H bond of a terminal methyl group than from the weaker C-H bonds of the remaining methylene groups (Collins *et al.*, 1991; Harris and Loew, 1995). However the site of hydroxylation may be explained by the distance of the H atom from the ferryl oxygen with only the terminal C-H bond close enough to undergo the hydroxylation reaction.

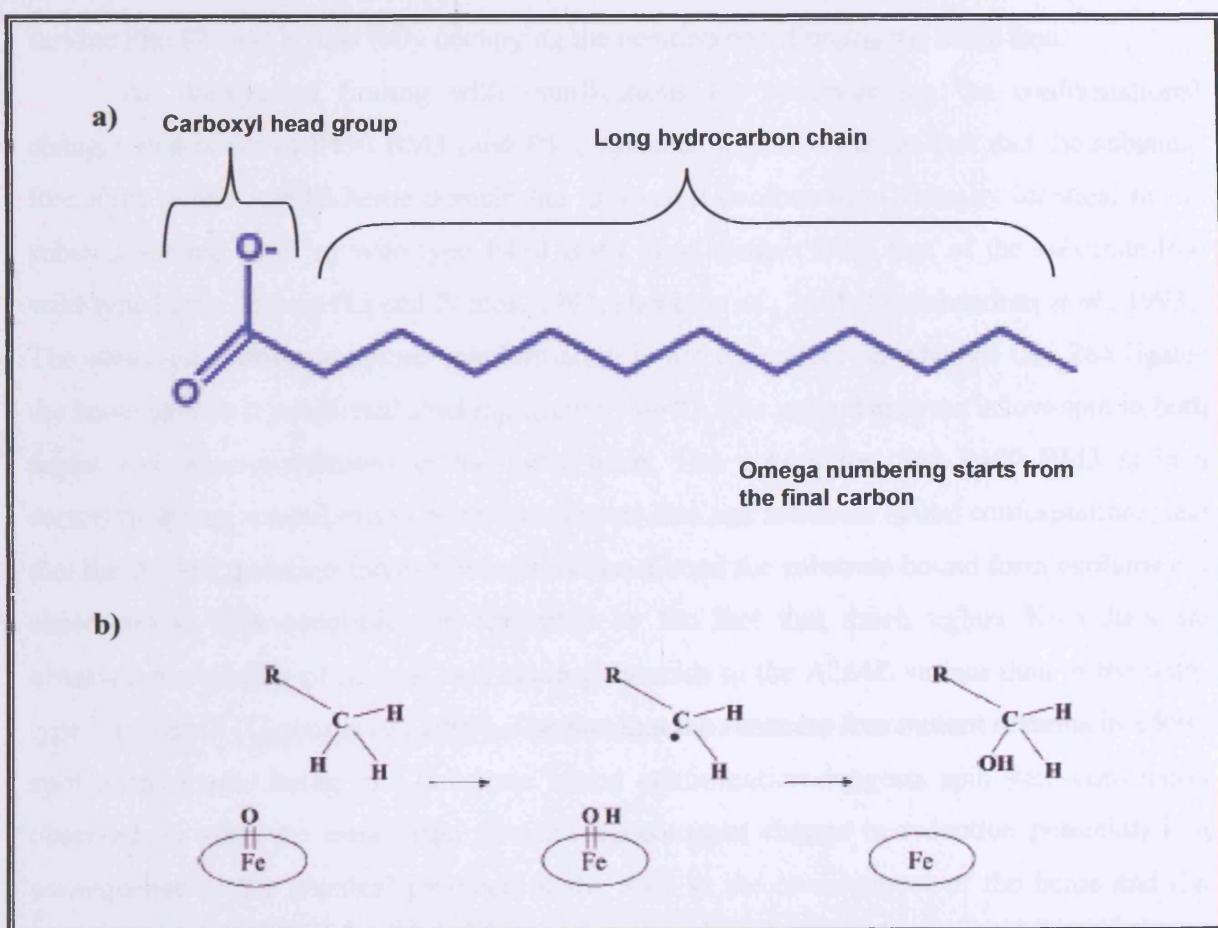


Figure 4.19 a) Diagram of fatty acid structure and description of the naming system. b) H-abstraction of the aliphatic group by the P450 enzymes (Taken from Chang and Loew, 1999)

#### 4.8. Conclusion

Crystallographic studies of the A264E variant of P450 BM3 confirm that the glutamate is able to ligate to the ferric heme iron of the mutant in the substrate free form creating a previously unobserved Cys-Fe-Glu ligation. Substrate binding shifts the equilibrium of the novel ligand set to a 100% heme ligated form producing a completely hexa-coordinated low-spin species, as opposed to the predominantly high-spin penta-coordinated form previously observed for the wild-type P450 BM3. Our studies explain clearly why substrate binding has this effect in the A264E enzyme, since palmitoleate occupies one of two possible positions for the Glu 264 sidechain. In the palmitoleate bound form, Glu 264 can no longer form an interaction with the key regiospecificity-determining residue Phe 87, and is thus fully occupying the position coordinating the heme iron.

An unexpected finding with ramifications for understanding the conformational changes that occur in P450 BM3 (and P450 systems in general) is the fact that the substrate free form of the A264E heme domain has an overall conformation virtually identical to the substrate-bound form of wild-type P450 BM3, thus distinct from that of the substrate-free wild-type heme domain (Li and Poulos, 1997; Haines *et al.*, 2001; Ravichandran *et al.*, 1993). The observed “substrate bound” conformation is not dependent on whether Glu 264 ligates the heme iron or is positioned stacking against Phe 87. The mutant enzyme is low-spin in both forms and aqua-coordinated in the latter form. The assumption that P450 BM3 is in a continual dynamic equilibrium between substrate free and substrate bound conformations, and that the A264E mutation forces this equilibrium toward the substrate bound form explains our observations. This conclusion is supported by the fact that much tighter  $K_D$  values are observed for binding of several long chain fatty acids to the A264E variant than to the wild-type P450 BM3 (Girvan *et al.*, 2004). The fact that the substrate free mutant remains in a low-spin form despite being in a substrate bound conformation suggests spin-state conversion observed on substrate association (and the concomitant change in reduction potential) is a consequence of the physical presence of the lipid in the environment of the heme and the significant change in the hydrophobic environment of the heme, and not a direct result of the adoption of the substrate bound conformation. Further work, in order to validate the hypotheses that arise from these findings would involve creation of other variants at position 264 – specifically investigating A264 variants that induce the conformational switch to the substrate bound conformation, but which do not however give rise to coordination of the heme iron due either to their size or their chemical makeup. Elucidation of the crystal structure of a reduced BM3 structure in the presence of substrate has to date remained elusive.

This structure would allow full understanding of the structural changes that occur upon reduction of the enzyme-substrate complex. The A264E mutant is active despite the heme-glutamate ligation observed indicating that structural rearrangement must take place in order to allow direct substrate interaction with the ferryl oxygen and subsequent enzymatic reaction.

In the interests of creating an autocatalytic 5-methyl covalent heme ligand, additional mutations are required to orient the glutamate residue in the right orientation for covalent bond formation to the heme macrocycle.

## **Chapter 5**

### **Crystal structure of the cytochrome P450 BM3 FAD domain**

## 5.1. Introduction

P450 BM3 from *Bacillus megaterium* is a paradigm model system in the P450 superfamily. This soluble bacterial fatty acid hydroxylase P450 is fused to a eukaryotic-like cytochrome P450 reductase (CPR) in a single holoenzyme polypeptide. Members of this family include cytochromes P450 reductase, nitric-oxide synthase (Stuehr, 1999), methionine synthase reductase (Wolthers *et al.*, 2004), and nitrite reductase 1 (Finn *et al.*, 2003). P450 BM3 has the highest reported hydroxylase activity for any P450 enzyme, oxygenating arachidonic acid in a NADPH-dependent manner at a rate of  $>17,000 \text{ min}^{-1}$  (Noble *et al.*, 1999). Crystallization of the intact, multi-domain flavocytochrome has been elusive; however, structures of the heme domain in both substrate-bound and substrate-free forms have been determined (Li and Poulos, 1997; Ravichandran *et al.*, 1993). More recently, the structure of the heme and FMN domains (the first two domains of the 1048 amino acid flavocytochrome) was reported in a non-stoichiometric ratio of two heme domains to one FMN domain (Sevrioukova *et al.*, 1999).

P450 BM3 and related flavoproteins show preference for NADPH over NADH as the reducing coenzyme. An aromatic residue stacks over the FAD isoalloxazine ring in each enzyme, and in some cases it is important in controlling coenzyme specificity (Dohr *et al.*, 2001). In P450 BM3, the aromatic residue inferred from sequence alignments to stack over the FAD is Trp-1046 and recent mutation of this residue led to a switch in coenzyme specificity (Neeli *et al.*, 2005).

The last remaining domain of the protein to be structurally elucidated is the FAD/NADPH binding domain; this domain interacts with NADPH, accepting two electrons (as a hydride ion) and shuttling these, one at a time, through the FMN domain onto the P450. The FAD/NADPH domain is related to the respective domain in eukaryotic CPRs, and to the NADPH-dependent ferredoxin/flavodoxin reductases. The mechanism that enables the FMN domain to communicate both with the FAD domain (to receive electrons) and then with the cytochrome P450 to transfer electrons to the heme centre is still little understood (Munro *et al.*, 2002). Structural characterization of the BM3 FAD domain will enable an advanced understanding of the roles of several amino acids in interactions with NADPH and FAD, and in interfacial recognition and docking with the FMN domain to facilitate inter-flavin electron transfer.

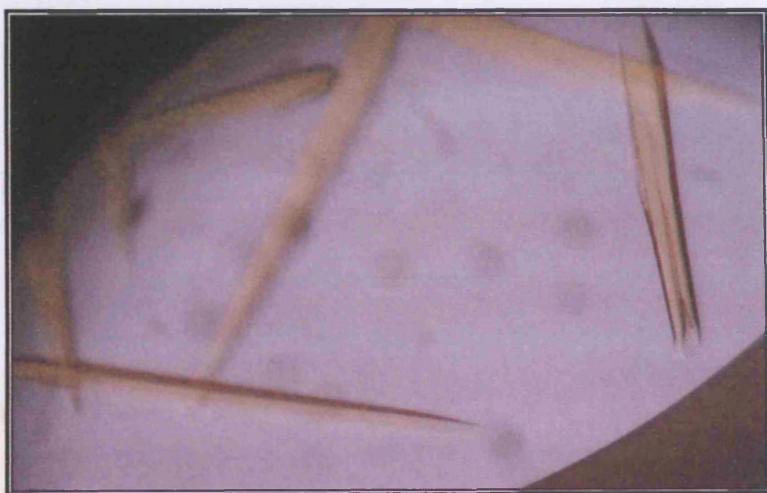
In this chapter the structure of the C773A mutant FAD-domain in both the presence and absence of  $\text{NADP}^+$  at a resolution of 2.15 Å and 2.4 Å, respectively is presented. Modelling of the complete BM3 structure using the previously determined heme and FMN domain structures is now possible.

## 5.2. Results

### 5.2.1. BM3 FAD domain C773A mutant Preparation and Crystallisation

Site directed mutagenesis of the FAD domain consisting of residues 655 – 1049 of the full length cytochrome P450 holoenzyme BM3 was carried out using the Stratagene Quik Change kit. Protein was produced and purified as previously described (Munro *et al.*, 1996; Govindaraj and Poulos, 1997). These procedures were carried out by Dr. Olivier Roitel (University of Leicester).

Crystallisation of the C773A mutant of the FAD domain was carried out as described in Section 3.5.1. Crystals were obtained using a well solution of 28 % polyethyleneglycol 8000, 0.3 M ammonium sulphate, cacodylic acid pH 6.5 at room temperature using the sitting drop vapour diffusion technique. Large hexagonal crystals with dimensions 70 x 70 x 900 µm formed after 4-7 days as shown in Figure 5.1.



**Figure 5. 1** Large hexagonal FAD domain crystals grown using the sitting drop vapour diffusion technique in 28 % polyethyleneglycol 8000, 0.3 M ammonium sulphate, cacodylic acid pH 6.5 after 4-7 days of equilibration.

### 5.2.2. Data collection and processing

Data was collected as described in Section 3.5.2. Analysis of the diffraction data from a single BM3 FAD domain crystal using the auto indexing routine in DENZO was consistent with a hexagonal space group with cell dimensions  $a = b = 191.387$   $c = 74.211$  Å. Scaling and merging of the diffraction data was carried out using SCALEPACK (Otwinowski and Minor, 1997) in space group  $P3_{1}2_1$  with systematic absences indicating that this or  $P3_{2}2_1$  was the

correct space group. Diffraction analysis of the NADP<sup>+</sup> bound BM3 FAD domain crystal was consistent with the native crystal form with a slight difference in the unit cell dimensions.

Scaling and merging of this data was carried out using SCALEPACK and data collection statistics are summarized in Table 5.1 below.

**Table 5. 1 Data refinement and processing statistics**

<b>Data Collection</b>	<b>Protein</b>	
	FAD domain	FAD domain + NADP <sup>+</sup>
Space group	P3 <sub>1</sub> 21	P3 <sub>1</sub> 21
Cell dimensions		
a (Å)	191.3	190.
b (Å)	191.3	190.6
c (Å)	74.2	74.3
X-ray source	ESRF ID14-EH3	ESRF ID14-EH3
Resolution Å	30-2.15	30-2.4
No. of observations		
Total	507,921	906,276
Unique	60932	83,817
Completeness (%)	99.9 (99.5)	99.3 (97)
I/σI	18.6 (2.1)	16.5 (2.0)
R <sub>merge</sub>	0.071 (0.455)	0.07 (0.431)
<b>Refinement</b>		
R <sub>work</sub>	0.211	0.191
R <sub>free</sub>	0.256	0.225
r.m.s deviations from ideal		
Bonds (Å)	0.013	0.010
Angle (°)	1.419	1.242

### 5.2.3. Structure Determination by Molecular Replacement

Molecular replacement was carried out using AMORE (Navazza, 2001) as described in Section 3.4.3 using a modified model of the FAD/NADPH domain of rat CPR (PDB code: 1OMI; Wang *et al.*, 1997) in order to determine the BM3 FAD domain C773A structure. The initial rotation search results for the substrate free structure are given in Table 5.2 below with the correct solutions distinct from random noise. The search results for the two translation searches are shown in Table 5.3. The translation searches were carried out in both P3<sub>1</sub>2<sub>1</sub> and P3<sub>2</sub>2<sub>1</sub>. The final correlation coefficient between F<sub>o</sub> and F<sub>c</sub> and R<sub>factor</sub> for the solutions after the rigid body refinement module of AMORE was 69.4 and 35.4 respectively.

Using the phases of the FAD domain C773A structure it was possible to determine the structure of the NADP<sup>+</sup> bound form using difference Fourier techniques.

**Table 5. 2 Rotation search results for BM3 FAD domain C773A mutant using AMORE. Solution 1 and 2 are clearly above noise levels.**

Solution no.	alpha	beta	gamma	Correlation of $F_o$ and $F_c$	R factor	Correlation of $I_o$ and $I_c$	Correlation coefficient of Patterson maps
1	27.40	12.15	92.39	27.5	54.4	35.9	15.7
2	10.83	32.24	46.34	26.5	54.7	34.7	13.0
3	77.12	74.78	102.73	23.6	55.6	30.4	6.9
4	10.06	64.53	46.53	23.4	55.6	30.1	6.9
5	23.44	54.18	214.62	23.3	55.7	30.6	6.2

**Table 5. 3 Translation search results for BM3 FAD domain C773A mutant using AMORE in space group P3<sub>1</sub>2<sub>1</sub>. Solution 1 is clearly above noise level in both cases.**

Solution number	Alpha	Beta	Gamma	X	Y	Z	Correlation of $F_o$ and $F_c$	R factor	Correlation of $I_o$ and $I_c$
1	27.40	12.15	92.39	0.2310	0.4801	0.4091	29.5	50.6	30.7
2	27.40	12.15	92.39	0.8971	0.8126	0.1349	20.3	53.4	21.5
3	27.40	12.15	92.39	0.8967	0.8133	0.4314	20.0	53.3	21.4

### Translation search 2

1	10.83	32.24	46.34	0.5304	0.3544	0.4630	47.5	44.6	49.9
2	10.83	32.24	46.34	0.5298	0.3539	0.1522	28.5	50.8	29.5
3	10.83	32.24	46.34	0.7298	0.4697	0.9213	23.8	52.2	27.5

## 5.2.4. Model Building and Refinement

### 5.2.4.1. Building and Refinement of BM3 FAD domain

An R<sub>free</sub> test set containing 5 % of the measured reflections was selected for use in validation of refinement as discussed in Section 2.8. using FREERFLAG (Brünger, 1992). Using REFMAC5 (Murshudov *et al.*, 1997) for initial refinement of the molecular replacement atom coordinates an output file was generated with coefficients for sigma A weighted  $F_o$ - $F_c$  and 2 $F_o$ - $F_c$  difference maps. The maps and molecules were visualised using

TURBO-FRODO (Roussel and Cambillau, 1992) and COOT (Emsley and Cowtan, 2004) and model building and map fitting was carried out. Manual rebuilding was alternated with multiple cycles of refinement using Refmac5 (Murshudov *et al.*, 1997) with the weighting for stereochemical restraints versus crystallographic data set at 0.20. In the final stages of refinement, addition of solvent molecules at positions of the peaks which were at least  $3\sigma$  in an  $F_o - F_c$  difference map was carried out using Arp/Warp (Perrakis *et al.*, 2001). Manual removal of all water molecules with B values larger than  $60 \text{ \AA}^2$  and greater than 3.5 Å distance from hydrogen bond donor or acceptor atoms was carried out. Analysis of the stereochemical quality of the protein model was accomplished using the PROCHECK (Hooft *et al.*, 1996) and WHATIF (Laskowski *et al.*, 1993) structure validation programs.

#### **5.2.4.2. Building and Refinement of BM3 FAD domain NADP<sup>+</sup> complex**

Using the final FAD domain C773A model as an initial model,  $F_o - F_c$  and  $2F_o - F_c$  maps were generated. The maps and molecules were visualised using TURBO-FRODO and small amounts of model building and map fitting were carried out. The vast majority of the protein molecule fitted the electron density perfectly. Manual rebuilding using COOT was alternated with multiple cycles of refinement using Refmac5 with the weighting for stereochemical restraints versus crystallographic data set at 0.25 due to the higher resolution of the data. Analysis of the stereochemical quality of the protein model was accomplished using the PROCHECK and WHATIF structure validation programs.

### **5.3. Analysis of the BM3 FAD domain C773A mutant and NADP<sup>+</sup> complex structures**

#### **5.3.1. BM3 FAD domain C773A mutant structure**

The final refined model of the BM3 FAD domain C773A mutant contains two molecules per asymmetric unit. Molecule A contains 384 of the expected 394 amino acid residues. There is no interpretable density for residues 655-658, residues 732-736 and residue 1049. The side chains for several residues have been modelled up to the beta carbon because of poor side chain electron density. Molecule B contains 376 of the expected 394 amino acid residues. There is no interpretable density for residue 655-659, residues 731-742 and residue 1049 while side chains for several residues have been modelled up to the beta carbon because

of poor side chain electron density. This model gives a final R-factor of 0.211 and the free R-factor of 0.256.

The overall statistics for the BM3 FAD domain C773 mutant structure gives values for the RMSD from standard values of the bond lengths of 0.013 Å and angle of 1.419°, respectively. The Ramachandran plot given in Figure 5.2 shows that 100 % of non-glycine residues have conformational angles ( $\phi, \psi$ ) in permitted regions and 92.5 % of these are within the most favoured regions.

### 5.3.2. BM3 FAD domain C773A mutant NADP<sup>+</sup> complex structure

The final refined model of the BM3 FAD domain C773A mutant NADP<sup>+</sup> complex is very similar to the non-complexed BM3 FAD domain C773A structure. This model gives a final R-factor of 0.191 and the free R-factor of 0.225.

The overall statistics for the BM3 FAD domain C773A mutant + NADP<sup>+</sup> structure gives values for the RMSD from standard values of the bond lengths of 0.010 Å and angle of 1.242°, respectively. The Ramachandran plot given in Figure 5.2 shows that 100 % of non-glycine residues have conformational angles ( $\phi, \psi$ ) in permitted regions and 93.2 % of these are within the most favoured regions.

**Table 5. 4 B factor analysis of the BM3 FAD domain C773A mutant structure in the presence and absence of NADP<sup>+</sup>**

	Protein	
	BM3 FAD domain	BM3 FAD domain + NADP <sup>+</sup>
<b>Molecule A</b>	44.11	32.99
<b>Molecule B</b>	44.63	34.18
<b>Water molecules</b>	42.45	38.74
<b>Overall</b>	44.86	34.34

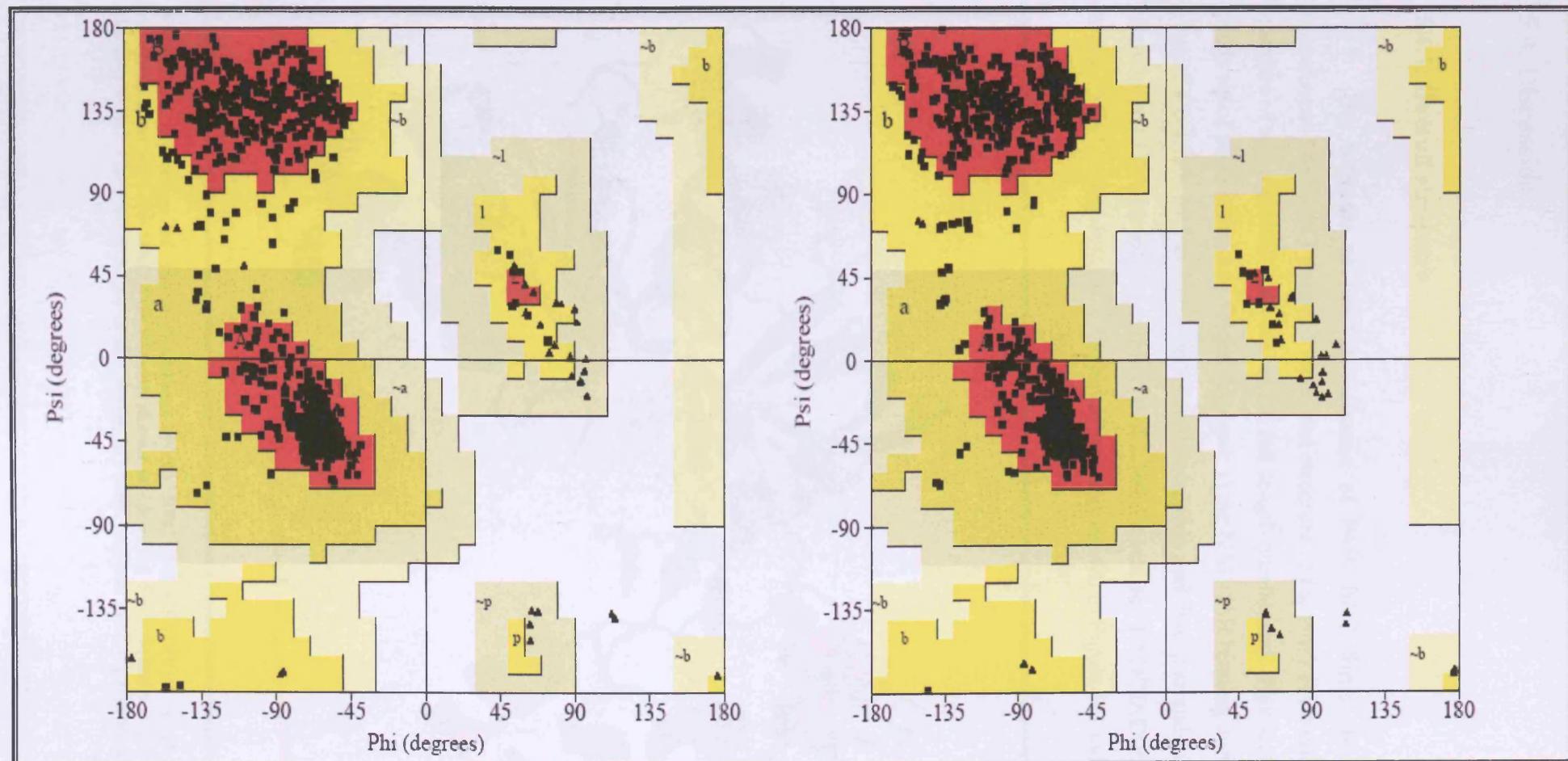
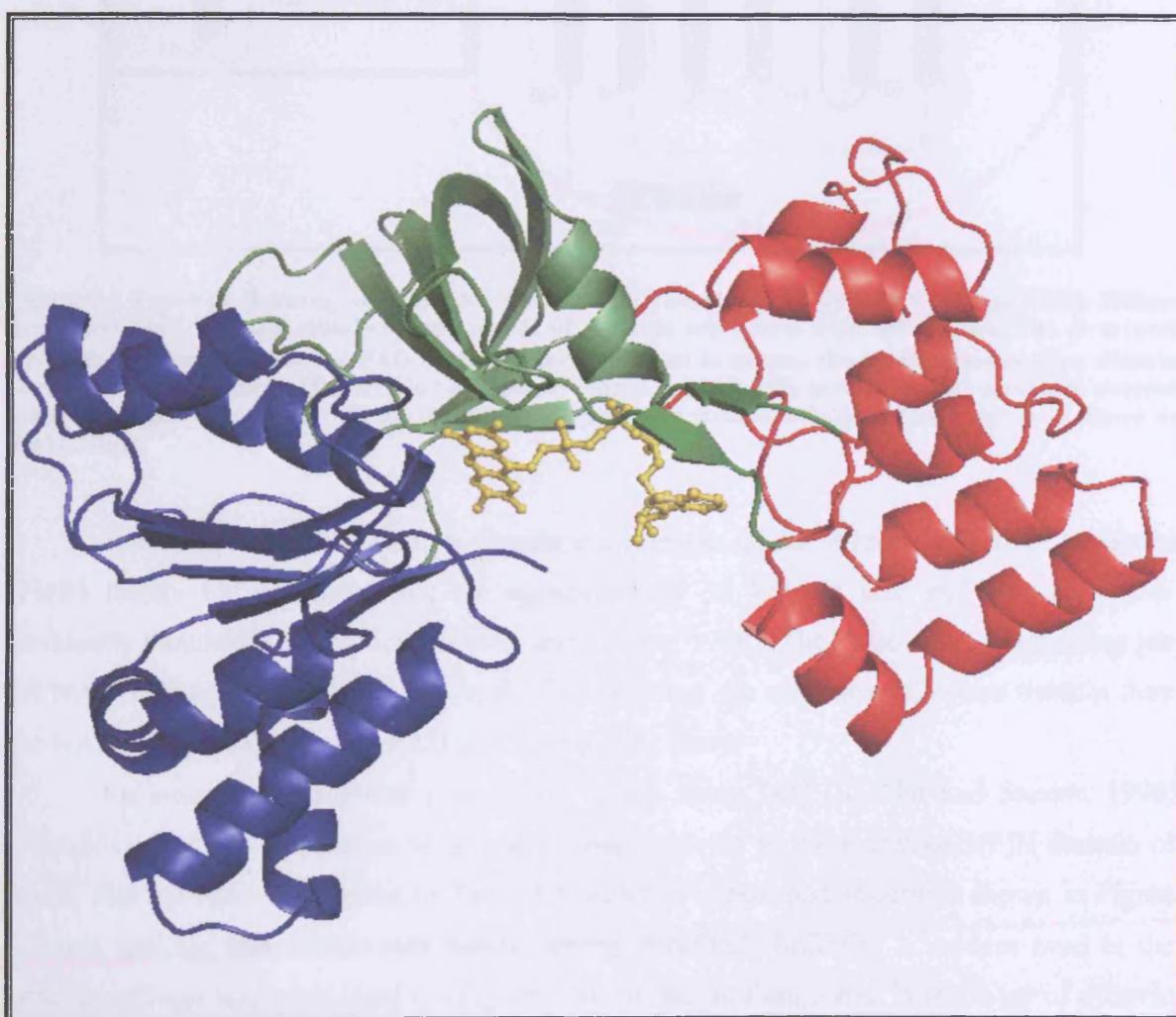


Figure 5. 2 Ramachandran plots for the models of the BM3 FAD domain C773A mutant (left) and the BM3 FAD domain C773A mutant-NADP<sup>+</sup> complex (right), produced using PROCHECK. The coloured areas indicate, in order of decreasing intensity, most favoured region, additional allowed region, generously allowed region and disallowed region. Glycine residues are shown as triangles.

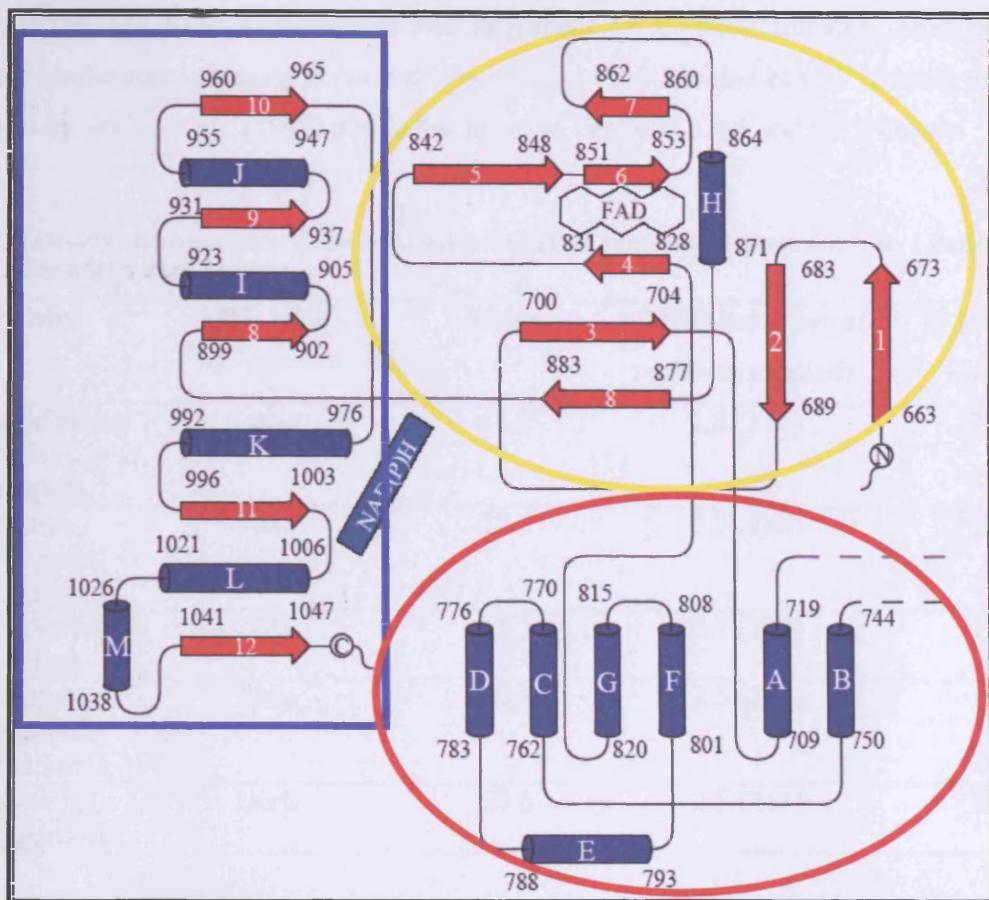
## 5.4. Discussion

### 5.4.1. Overall structure

The structure of the FAD domain of P450 BM3 forms the final piece in the cytochrome P450 BM3 jigsaw. The crystal structure of the BM3 FAD domain C773A mutant consists of residues 658-1048 (using the full length numbering). This section of the protein is made up of three individual folded domains: i) the NAD(P)H binding domain is made up of a typical Rossman fold constructed of five  $\alpha$ -helices and five  $\beta$  strands ii) the FAD-binding domain and iii) ‘connecting domain’ that interconnects the FAD/NAD(P)H binding domains to the FMN binding domain in the holoenzyme as shown in Figure 5.3 and 5.4 below.



**Figure 5.3 Structure of the Cytochrome P450 BM3 FAD/NAD(P)H domain shown in cartoon representation with the NAD(P)H domain shown in blue, the FAD domain in green and the connecting domain in red. The FAD cofactor is also shown in yellow ball and stick representation.**



in the FAD/FMN domain interaction and alignment for electron transfer. Analysis of the sequence shows that sequence conservation occurs at the FAD and NADP binding regions as described by Wang *et al.*, (1997) and is discussed in Sections 5.4.4 and 5.4.5 below.

**Table 5. 5 Structural homologues of the BM3 FAD/NAD(P)H domain as determined from a database search of the EMBL DALI server.**

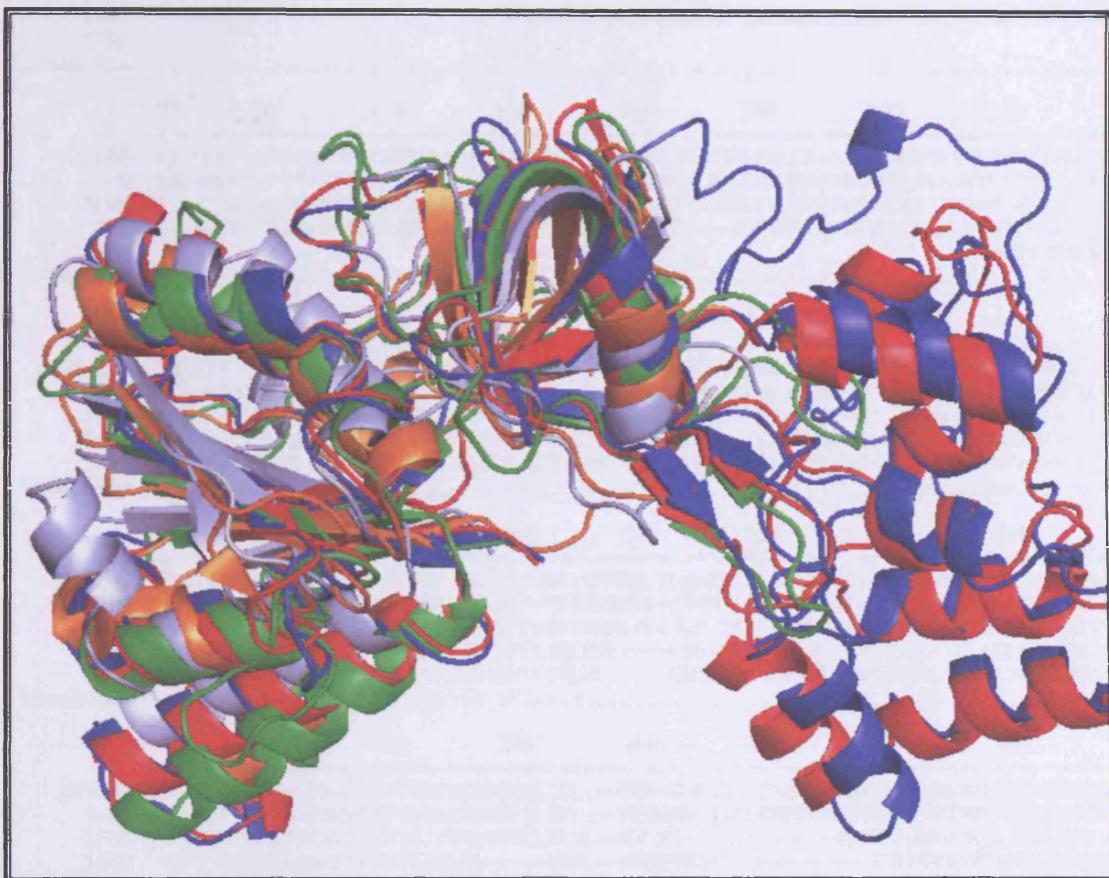
Protein	PDB code	Z score	RMSD (number of residues aligned)	% sequence identity
Rat Cytochrome p450 reductase	<b>1amo</b>	44.1	1.8 (376)	35
Spinach ferredoxin reductase	<b>1fnc</b>	28.3	2.2 (264)	28
Flavodoxin reductase	<b>1fdr</b>	21.5	2.7 (220)	17
Phthalate dioxygenase reductase	<b>2pia</b>	20.8	2.5 (211)	18
Benzoate 1,2-dioxygenase	<b>1krh</b>	20.6	2.5 (211)	25

#### 5.4.2. C773A mutation

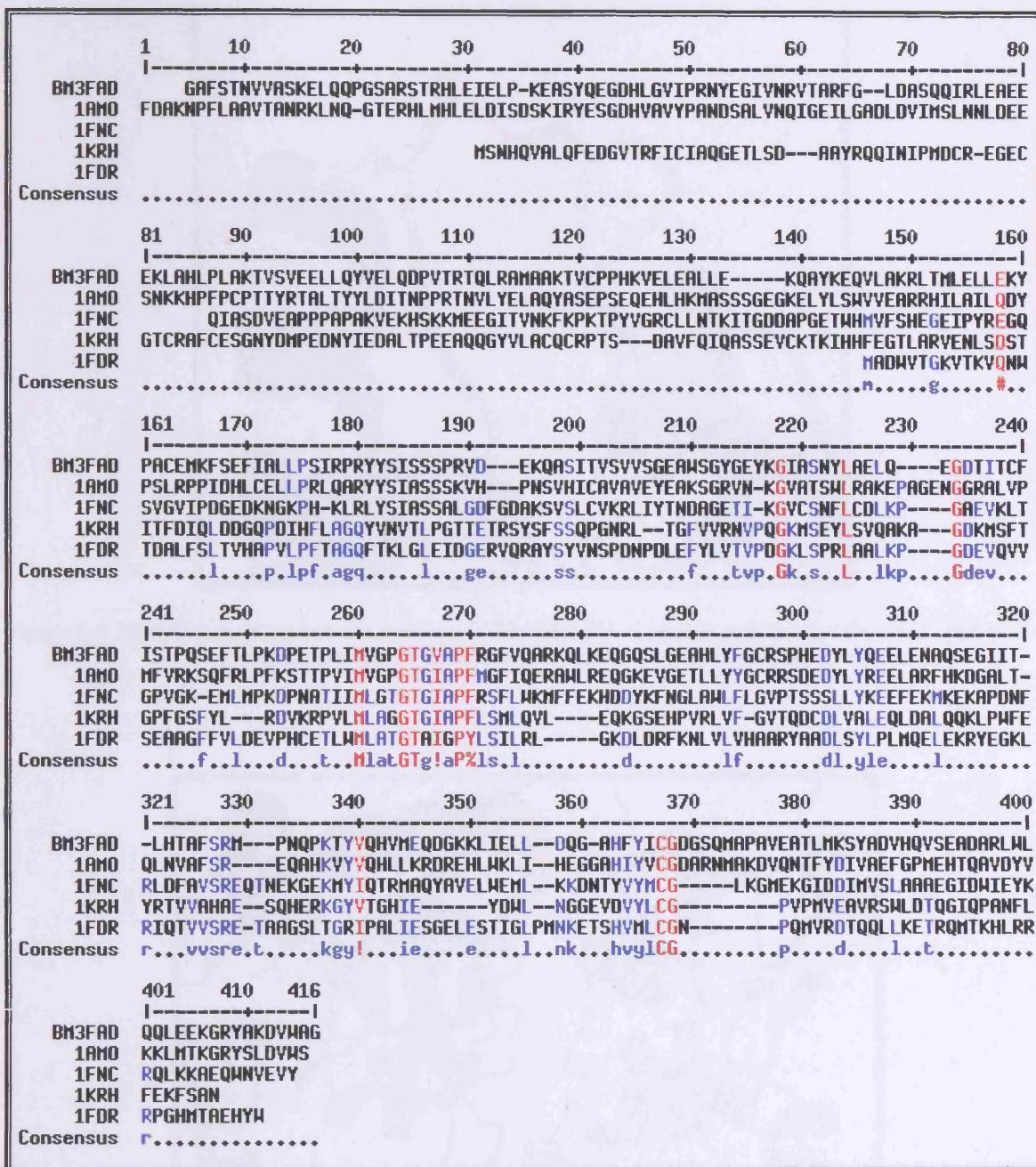
The initial rationale for the C773A mutation was to create a monomeric form of the reductase and FAD domains that might be more amenable to crystallization attempts. The C773A mutation is located at the surface in the loop region between helices C and D. Residue 773 is located near to C811 in the loop region between helices F and G. Due to the close proximity of these cysteines, putative disulphide bond formation may occur and only slight remodelling of the loop regions is required to allow formation of a disulphide bond between C773 and C811 (Figure 5.7). The presence of a disulphide bond on the surface of the protein molecule could lead to the disulphide bond exchange with a second BM3 molecule leading to dimer formation as previously described (Govindaraj and Poulos, 1997). The presence of a non-uniform protein population in the crystallisation sample would be a strong hindrance to crystal nucleation and growth. The remaining C811 does not cause dimer formation and this residue is seen facing away from the solvent in the crystal structure which may indicate that this residue is not important in inter-protein disulphide bond formation.

In addition, the BM3 FAD domain crystal structure shows that the C773A mutation occurs at a crystal contact position between the ‘connecting’ domains of two molecules

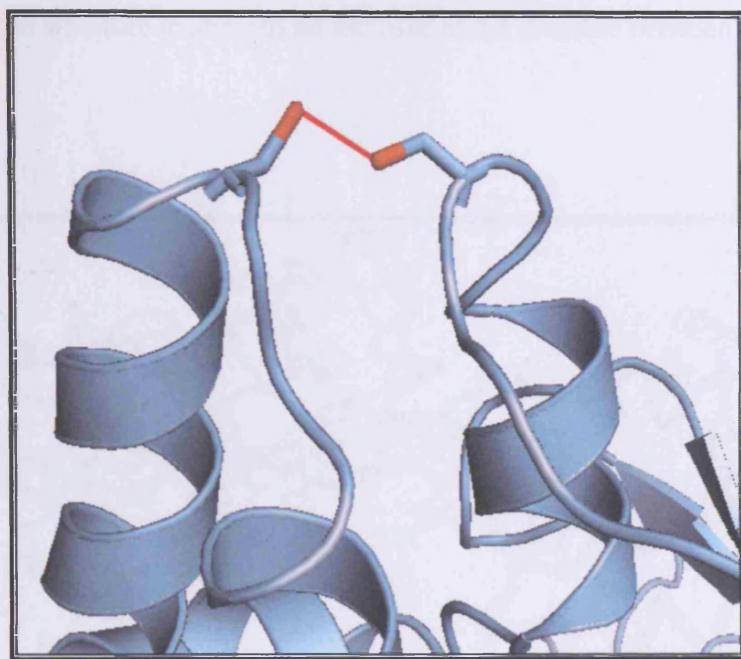
(Figure 5.8). The formation of an intermolecular disulphide bond at this position might prevent this crystallisation interaction.



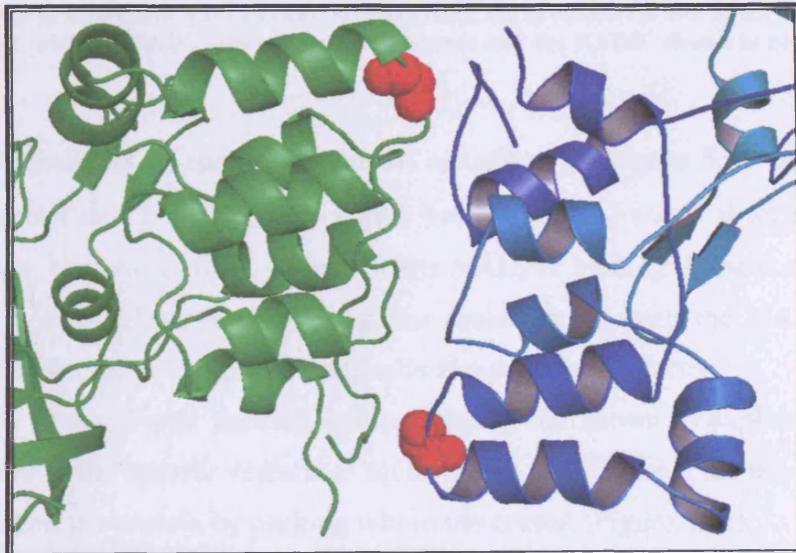
**Figure 5.5** Overlap of the structurally homologous proteins identified by DALI shown in cartoon representation. The structures shown are rat CPR pdb code: 1AMO (blue), spinach ferredoxin reductase pdb code: 1FNC (green), Benzoate 1,2-Dioxygenase Reductase pdb code: 1KRH (orange) and E.coli Flavodoxin reductase pdb code: 1FDR (light blue) overlapping the BM3 FAD domain shown in red.



**Figure 5.6 Alignment of the protein sequence of BM3 FAD/NAD(P)H domain with structural homologues using MULTALIN (Corpet, 1988). These include rat CPR PDB code: 1AMO, spinach ferredoxin reductase, PDB code: 1FNC, Benzoate 1, 2-Dioxygenase reductase, PDB code: 1KRH and *E.coli* Flavodoxin reductase, PDB code: 1FDR. The residues coloured red indicate greater than 90 % consensus and those coloured blue indicate greater than 50 % consensus.**



**Figure 5.7** Modelled disulphide bond between C773 and C811. Cysteine residues are shown in stick mode with the protein structure shown in cartoon representation.

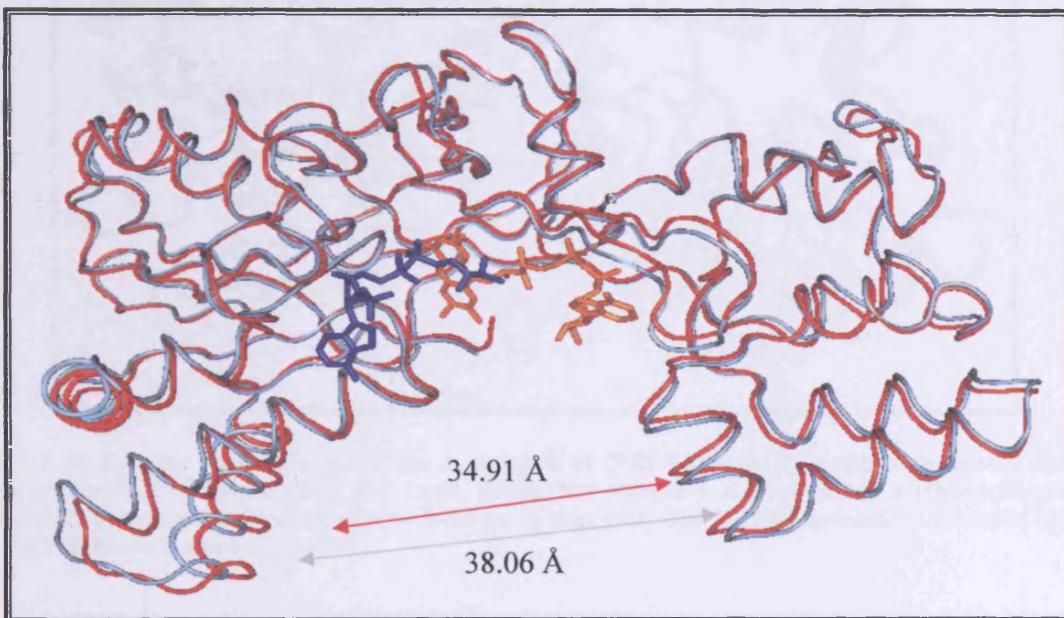


**Figure 5.8** Crystal contact surface between BM3 FAD molecule and a symmetry related molecule shown in cartoon representation with the C773A mutation shown as red spheres.

#### 5.4.3. Differences between molecules A and B in the asymmetric unit

Comparison of the two molecules in the asymmetric unit reveals little structural differences between the two molecules except for a movement between FAD/NAD(P)H domain and the ‘connecting’ domains, as shown in Figure 5.9, corresponding to an overall

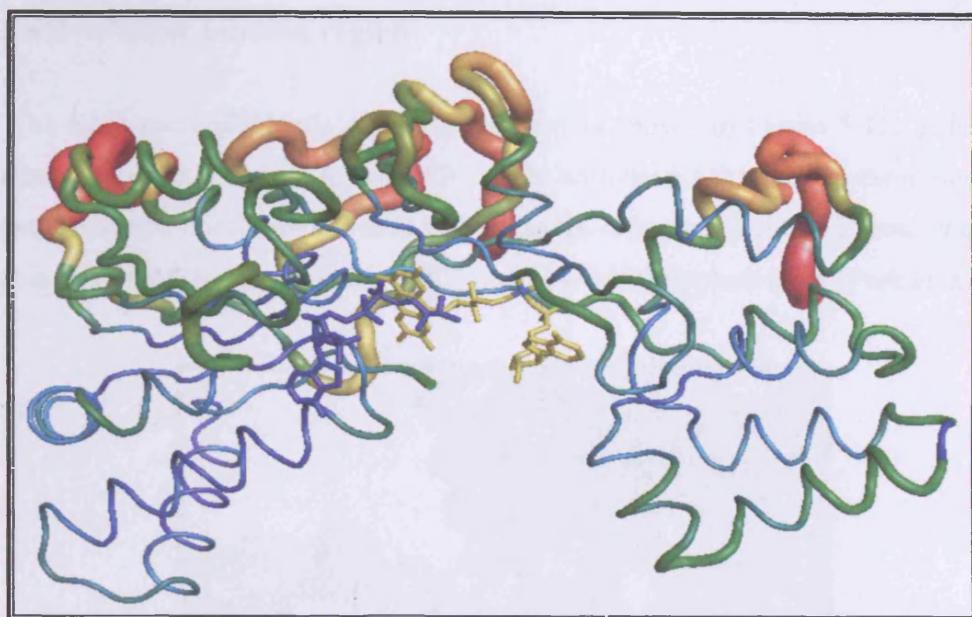
displacement in the structure leading to an increase in the distance between residues 1027 and 787 of over 3 Å.



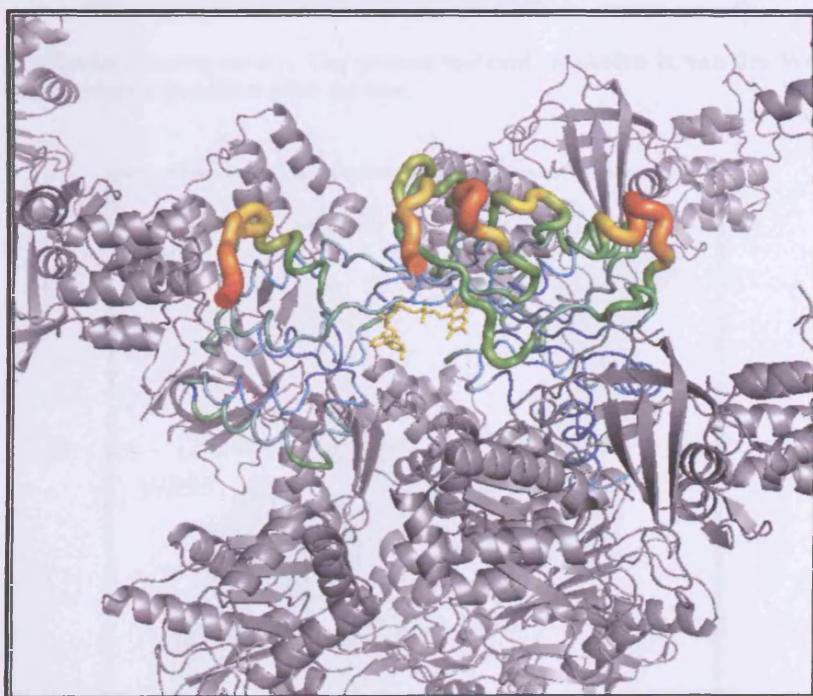
**Figure 5.9** Overlay of molecules A (cyan) and B (red) from the asymmetric unit of the NADP<sup>+</sup>-bound BM3 FAD domain, with the FAD molecule shown in orange and the NADP<sup>+</sup> shown in blue as stick representations.

B factor analysis of the FAD domain visualised in Figure 5.10 indicates that the regions of β sheet 1 and 2 and the loop region between helix I and β sheet 9 exhibit high B factors. However, the entire distal surface of the NADPH binding domain exhibit B factors that are high in comparison to the rest of the molecule. In both the NADP<sup>+</sup> bound and unbound forms molecule A exhibits a marginally lower overall B factor.

Analysis of the crystal packing was performed and shown to explain some of the B factor differences in the specific regions of molecules A and B which are due to the restraints that are imposed on the protein by packing within the crystal (Figure 5.11).



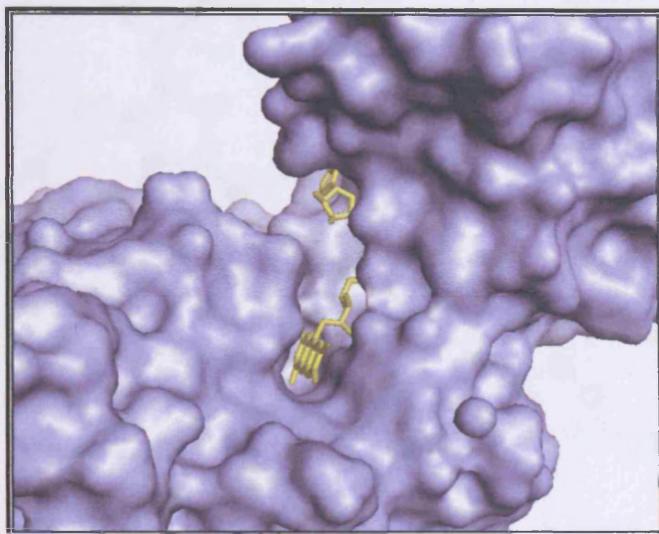
**Figure 5. 10** B-factor putty diagram of the A molecule of BM3 FAD-NADP<sup>+</sup> complex generated using the preset mechanism of Pymol v0.98 (De Lano, 2002). The structure is exhibited in a ribbon format that increases in size depending on the atomic B factor in that area. The NADP<sup>+</sup> molecule is shown in blue and the FAD molecule is shown in yellow.



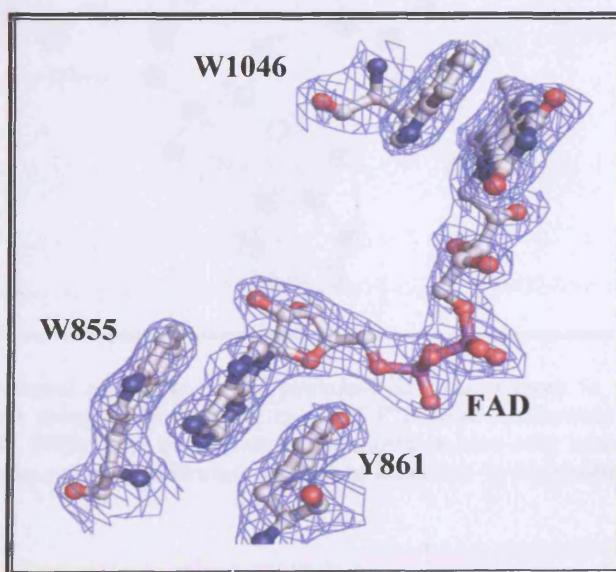
**Figure 5. 11** Molecule A of the BM3 FAD domain shown in B factor putty preset mode generated by Pymol v0.98 and coloured according to B factor with its symmetry equivalents within 6 Å shown in grey cartoon representation. The molecule is rotated 180° along the y axis to that shown in Figure 5.10 in order to allow easier viewing of the protein that has less packing restraints and thus higher overall B factors.

#### 5.4.4. FAD cofactor binding region

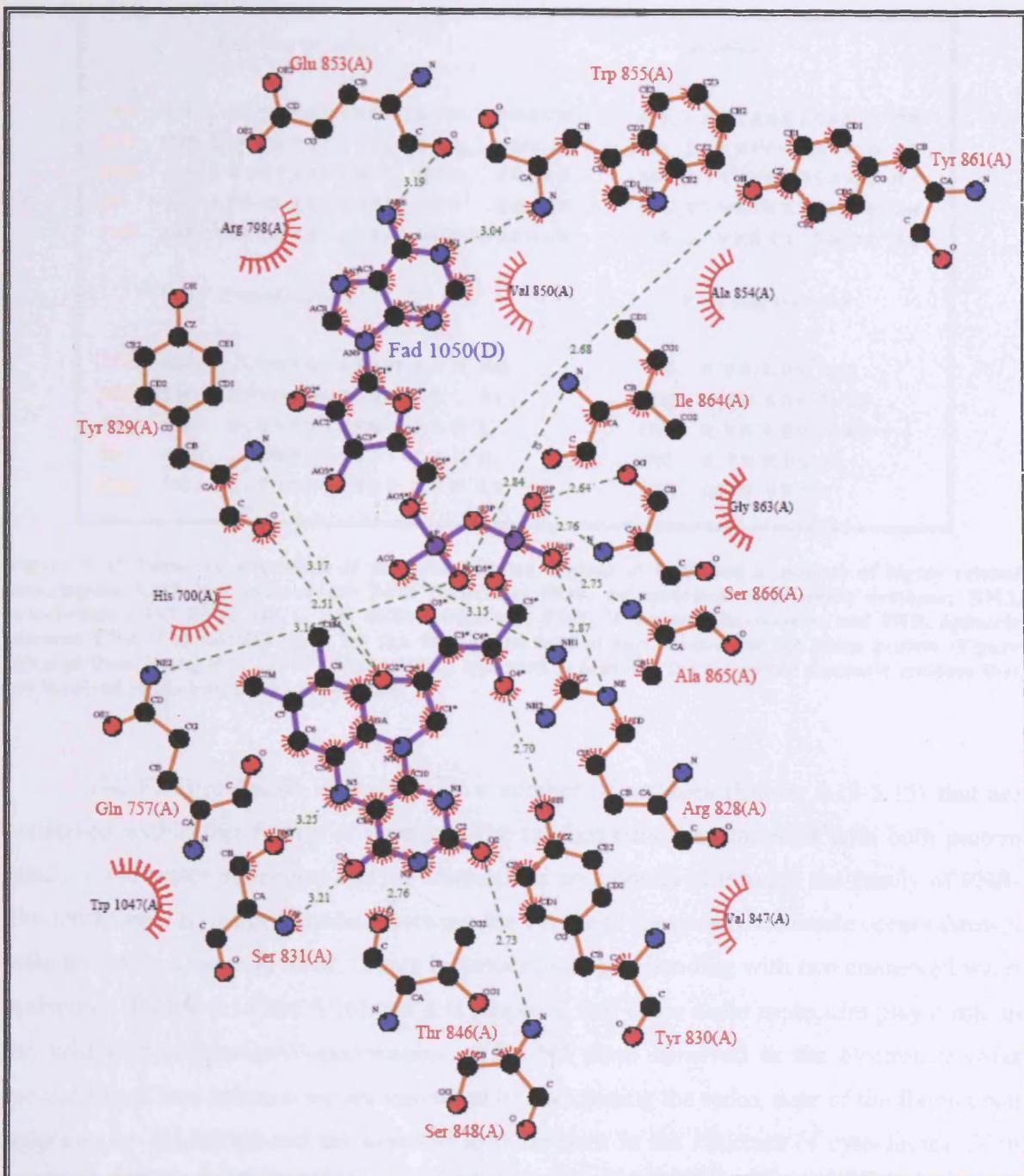
The FAD molecule binds in a large groove as shown in Figure 5.12. This region is highly similar to other members of the FNR family with the FAD bound in the elongated form as opposed to a bent conformation seen in ferredoxin reductase (Sridhar Prasad *et al.*, 1998), flavodoxin reductase (Ingelman *et al.*, 1997) and also in DNA photolyase (Park *et al.*, 1995).



**Figure 5. 12 FAD cofactor binding cavity.** The protein molecule is shown in van der Waals surface mode with the FAD molecule shown in yellow stick format.



**Figure 5. 13 Aromatic residues interacting with the FAD molecule.** SigmaA weighted 2Fo-Fc map surrounding the FAD cofactor and W1046, Y851 and W884 in BM3 FAD domain shown in blue with the residues and cofactor shown in ball and stick format coloured according to atom type.



**Figure 5. 14** Two dimensional overview of the protein-FAD interactions in molecule A of the C773A mutant of BM3 generated using the automatic mode of PDBsum (Laskowski, 2001) and its LIGPLOT component (Wallace *et al.*, 1995). The FAD molecule is shown in blue with hydrogen bonding interactions between the protein and the cofactor indicated by dashed lines and hydrophobic interactions indicated by spiked red semi-circles.

FAD ring (si face)		Adenine	
CPR	450 L Q A R Y Y S I A S S S . K V H . . P N S V H	470 I C A V A V E Y E A K S G R	
NOS	1170 L Q P R Y Y S I S S S P . D M Y . . P D E V H	1190 L T V A I V S Y H T R D G E	
BM3	824 I R P R Y Y S I S S S P . R V D . . E K Q A S	844 I T V S V V S G E A W S G Y	
SR	382 L T P R L Y S I A S S Q . A E V . . E N E V H	402 V T V G V V R Y D V . E G R	
FNR	145 H K L R L Y S I A S S A L G D F G D A K S V S	168 L C V K R L I Y T N D A G E	
Pyrophosphate		FAD ring (re face)	
CPR	484 . . V N K G V A T S W L R A K E	670 R Y S L D V W S	
NOS	1208 G P V H H G V C S S W L N . R I	1389 R Y H E D I F G V . .	
BM3	858 G . E Y K G I A S N Y L A E L .	1040 R Y A K D V W A G	
SR	415 . . A R A G G A S S F L A D R .	592 R Y Q R D V Y	
FNR	182 . . T I K G V C S N F L C D L K	363 Q W N V E V Y	

**Figure 5. 15 Sequence alignment of the FAD binding regions of BM3 and a number of highly related homologues.** CPR, rat cytochrome P450 reductase; NOS, rat neuronal nitric-oxide synthase; BM3, cytochrome P450 BM3; SR, *E. coli* sulfite reductase; FDX, *D. vulgaris* flavodoxin; and FNR, *Spinacia oleracea* FNR. The numbering is for the first amino acid of each region for the given protein (Figure adapted from Wang *et al.*, 1997). The residues coloured in blue are the conserved aromatic residues that are involved in binding the FAD cofactor.

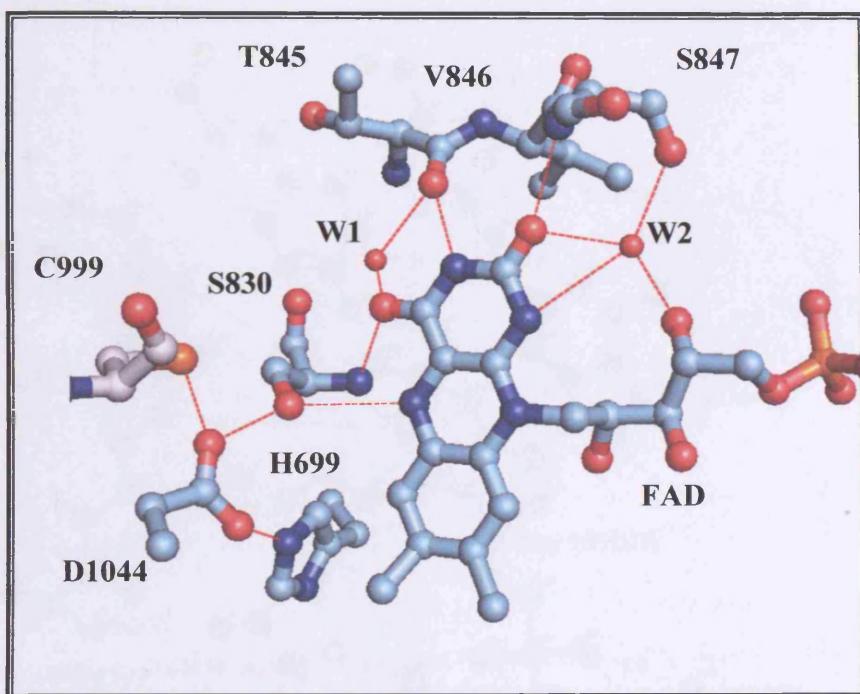
The FAD molecule interacts with a number of residues (Figure 5.13-5.15) that are conserved within this family of proteins. The isoalloxazine ring interacts with both protein residues and water molecules and the interactions are conserved through the family of FNR-like reductases. Hydrogen bonding between the FAD and the protein molecule occurs through residues S831, T846 and S848. There is further hydrogen bonding with two conserved water molecules (Figure 5.14 and 5.16) and it is proposed that these water molecules play a role in the acid/base protonation/deprotonation of the N5 atom involved in the electron transfer mechanism. These interactions are important in maintaining the redox state of the flavin upon reduction by NAD(P)H and are identical to those seen in the structure of cytochrome P450 reductase (Wang *et al.*, 1997).

A conserved triad of residues (Cys-999, Ser-830, and Asp-1044, P450 BM3 numbering) within the diflavin reductase family of flavoproteins was identified as important in hydride transfer (Govindaraj and Poulos, 1997; Shen *et al.*, 1999). The C999A mutant of BM3 FAD domain was studied by detailed spectroscopic characterisation and was identified as being crucial in the hydride transfer mechanism from NAD(P)H to the FAD cofactor (Roitel *et al.*, 2003). In rat CPR, an equivalent catalytic triad was identified and mutation of the C680 residue also led to large decreases in the rate of reduction (Shen *et al.*, 1999). These low rates are not caused by changes in the thermodynamic properties of the flavins but are attributed to the formation of a reduced enzyme-NADP<sup>+</sup> charge transfer species which is formed due to the slow release of NADP<sup>+</sup> from the two electron reduced enzyme. Hydride

transfer is reversible with the observed rate of oxidation of the two electron reduced C999A BM3 reductase and FAD domain hyperbolically dependant on NADP<sup>+</sup> concentration.

There is no protein residue interacting with the N1 atom of the isoalloxazine ring which is involved in electron transfer. However, the N1 makes a hydrogen bond with W2 (as seen in homologous protein structures) which is in turn hydrogen bonded to S847.

The adenosine moiety of the FAD cofactor hydrogen bonds with the side chain of E853 and the main chain nitrogen of W855. There are a number of van der Waals interactions from R798, V850 and A854 and aromatic stacking interactions with Y861 and W855. The pyrophosphate oxygen atoms are hydrogen bonded to the protein through the side-chains of R828 and Q757 and main chain nitrogen atoms of A865, I864, Y829 and Y830. The pyrophosphate binding motive has been identified as being the most strongly conserved element involved in binding (Dym and Eisenberg, 2001).

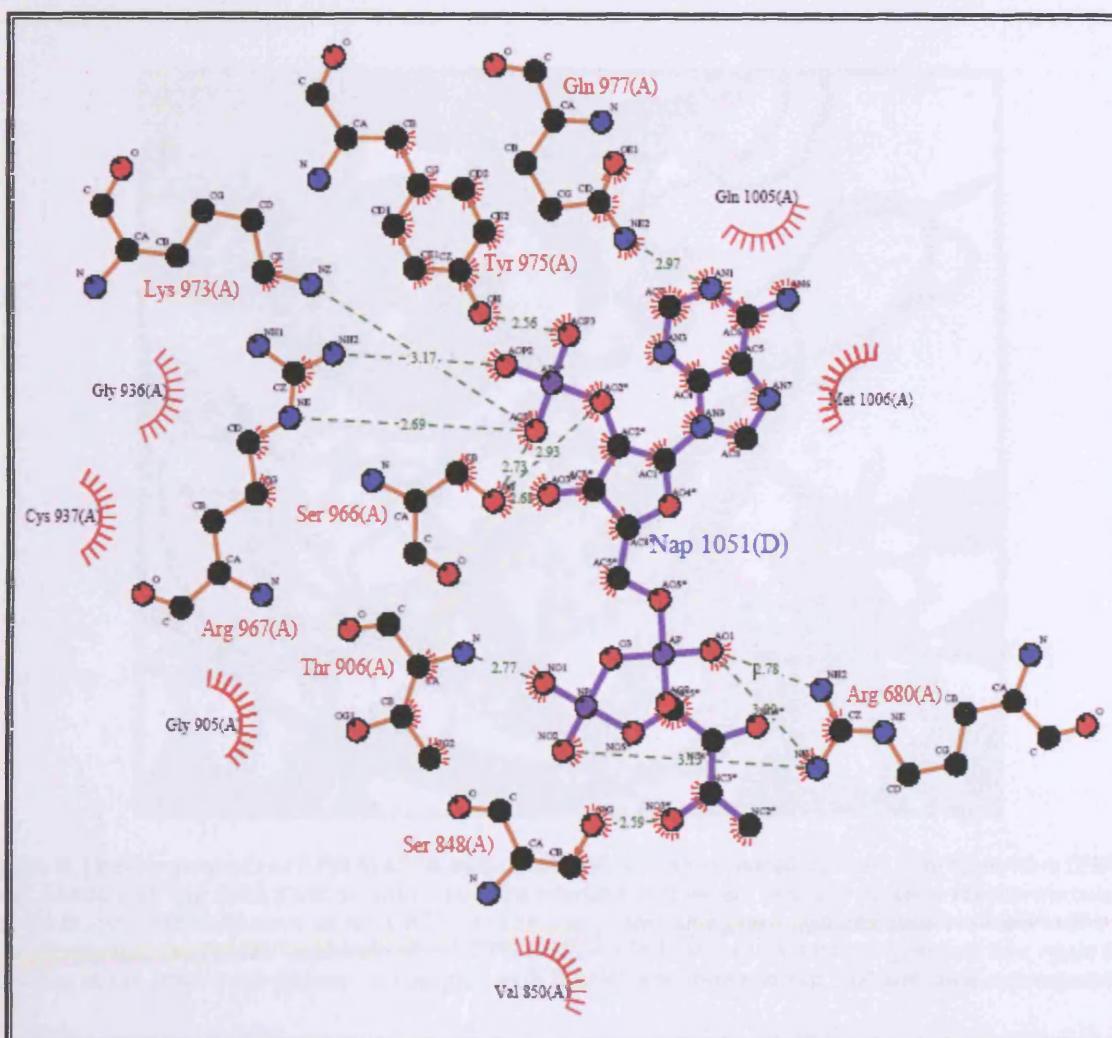


**Figure 5.16** Ball and stick representation of the isoalloxazine ring with surrounding residues and water molecules in Molecule A of the NADP<sup>+</sup>-free BM3 FAD domain. Carbon atoms are coloured white with nitrogen atoms coloured blue and oxygen atoms coloured red and the sulphur and phosphate atoms are coloured green.

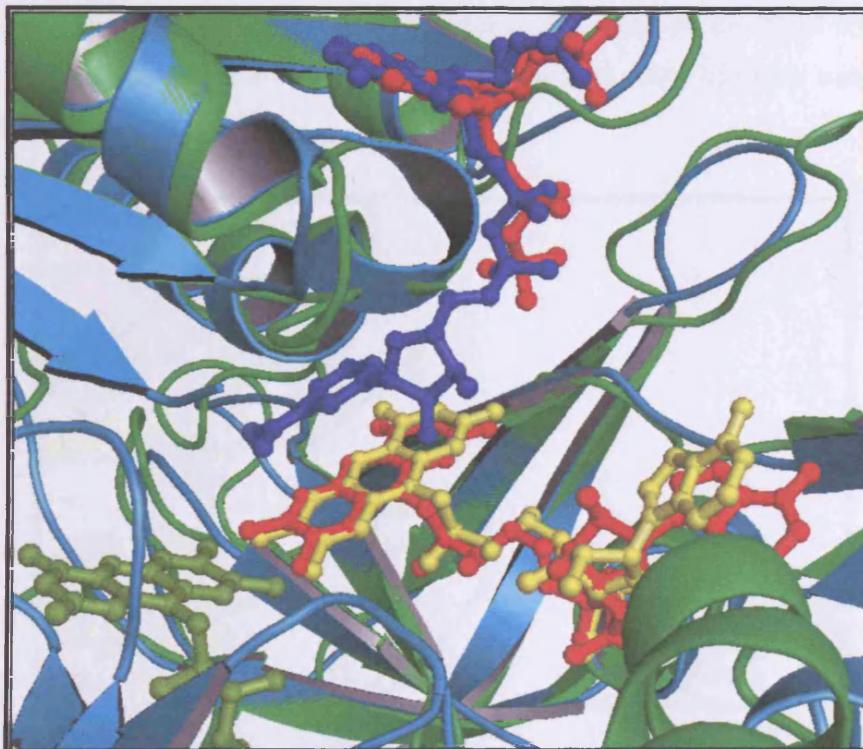
#### 5.4.5. NAD(P)H binding region

No significant structural rearrangement is seen upon NADP<sup>+</sup> binding to the BM3 FAD structure with the exception of the removal of a number of water molecules in the NADP<sup>+</sup> binding pocket. The RMSD between the native and complexed protein structure is 0.206 Å which is of a similar order as the RMSD between molecules A and B in the asymmetric unit.

The binding site of the NADP<sup>+</sup> is made by a highly conserved domain within the diflavin reductase families. There are sixteen protein interactions in the BM3 FAD/NAD(P)H domain with the NADP<sup>+</sup> in total (Figure 5.17). The adenine moiety is hydrogen bonded to Q977 and has further hydrophobic interactions with Q1005 and M1006. There is stacking of the adenine ring with the aromatic side chain of Y974 (as seen in other members of the family). The pyrophosphate moiety is hydrogen bonded by the main chain of T906 and has multiple interactions with the side-chain of R680. The ribosyl phosphate oxygen molecules are hydrogen bonded to Y974, R967 and K973. This particular set of interactions is highly conserved within the family and is involved in the specification of NADPH versus NADH as the electron donating group. In the nNOS reductase structure (Garcin *et al.*, 2004, PDB code: 1TLL), there is a further interaction with nNOS R1400 and this has been identified as being critical to the specification of cofactor type. This arginine lies on a helix not present in the BM3 FAD/NAD(P)H domain.



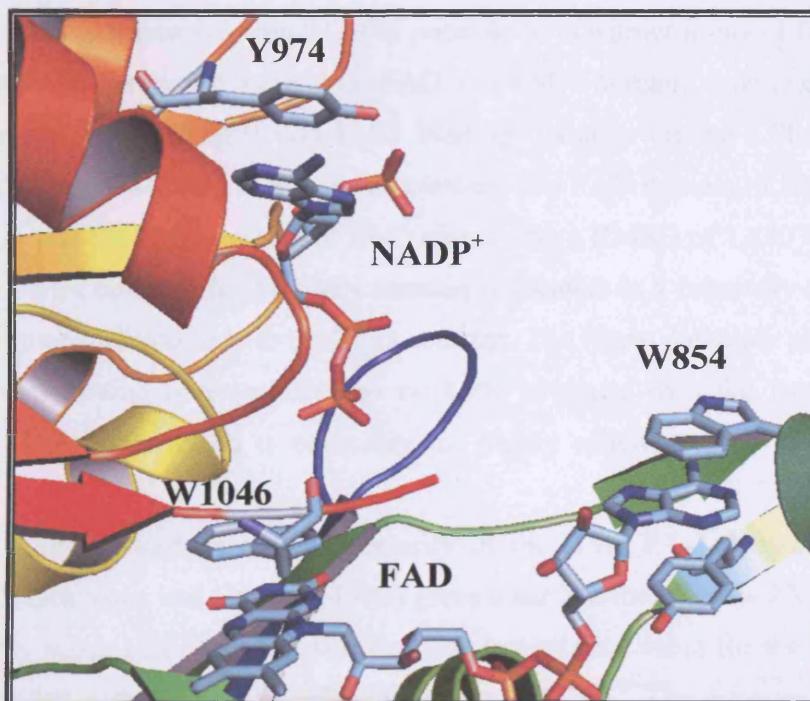
Electron density for the ribose nicotinamide moiety is disordered; there is however some electron density at the beginning of this moiety that suggests hydrogen bonding to S848. The lack of electron density around the NMN group indicates there is high mobility of the nicotinamide when W1046 is stacked with the FAD. This disorder of the nicotinamide mononucleotide (NMN) moiety is seen previously in the NADP<sup>+</sup> complex structures of CPR (Wang *et al.*, 1997), NOS FAD domain (Zhang *et al.*, 2001) and spinach FNR (Bruns *et al.*, 1995). The nicotinamide moiety is however visible in the structure of the spinach FNR molecule Y308S mutant (Deng *et al.*, 1999) where the nicotinamide stacks against the isoalloxazine ring and also in the W677X mutant of rat CPR (Hubbard *et al.*, 2001, Figure 5.18). Sequence alignment indicates that an aromatic residue stacking over the FAD is conserved within the family of proteins and it is understood that this residue needs to undergo a large movement in order to allow the stacking of the nicotinamide group on the FAD cofactor and subsequent hydride transfer to take place.



**Figure 5.18 Structure of rat CPR W677X mutant shown in cyan coloured cartoon representation (PDB code: 1JA0) with the BM3 FAD domain structure overlaid and shown in green cartoon representation. The FAD and FMN cofactors of rat CPR are shown in yellow and green ball and stick representation respectively and the NADP<sup>+</sup> molecule of rat CPR is shown in blue ball and stick representation while the cofactors of the BM3 FAD domain in complex with NADP<sup>+</sup> are shown in red ball and stick representation.**

An area of potential importance in the biotechnological exploitation of BM3 is the cofactor specificity of the reductase. Cofactor specificity is of great importance because of the large cost of the natural cofactor NADPH and the relatively low levels of activity with NADH in the wild type enzyme. Changing the specificity of a reductase from a NADPH dependant to

a NADH dependant type was first successfully carried out by Scrutton *et al.*, (1990) which involved a large number of mutations. More recently, a single amino acid substitution has achieved the same coenzyme specificity switch in the rat CPR (Gutierrez *et al.*, 2000), BM3 (Neeli *et al.*, 2005) and in nNOS (Dunford *et al.*, 2004). These mutations involve the aromatic residue that is stacking against the FAD isoalloxazine ring (Figure 5.19) specifically W1046A in BM3 whose direct interaction with the FAD cofactor is confirmed by the FAD domain structure. Mutation of this residue in BM3 also results in the observation of tighter binding of both NADPH and NADH indicating that the tryptophan residue prevents interaction of the NMN domain with the FAD cofactor. This tighter binding is probably critical in allowing the NADH molecule to interact with the FAD molecule and allow electron transfer to the FAD molecule. The ribosyl phosphate of NADPH binding hydrogen bonds to a number of highly conserved positively charged residues (R967 and K973) and these interactions may strengthen the binding of the NADPH molecule in the native protein enough to allow the NMN group to ‘force’ the flipping of W1046 and lead to efficient electron transfer. In the case of NADH binding to the native protein the binding is weaker and may be too short-lived to induce replacement of the aromatic stacking residue which would allow electron transfer from the NADH to the flavin to occur.



**Figure 5.19** NADP<sup>+</sup> interactions with the conserved aromatic residues and the FAD cofactor. The conserved aromatic residues and the FAD cofactor and bound NADP<sup>+</sup> are shown in stick representation. The protein molecule is shown in cartoon representation.

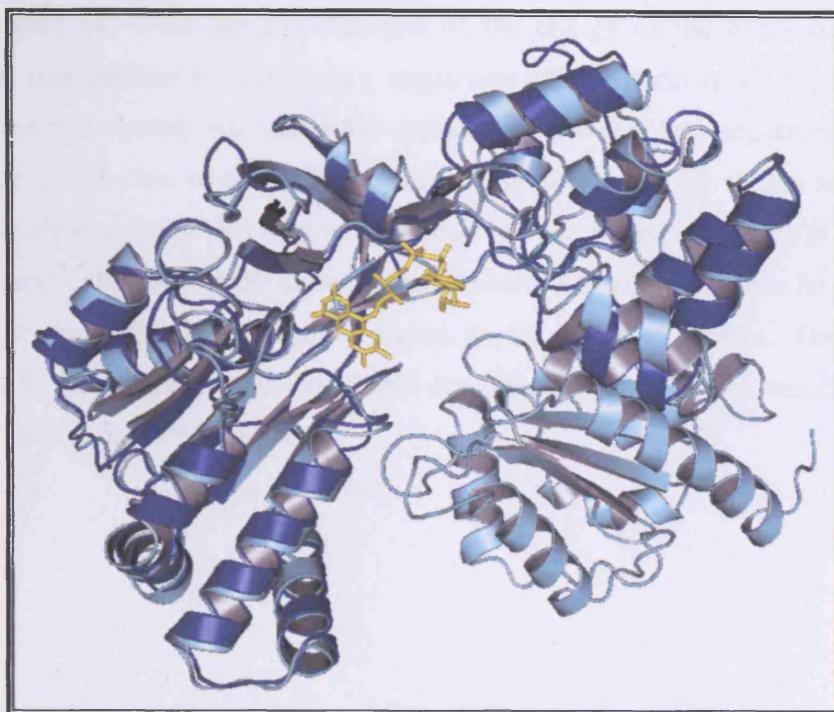
The BM3 FAD domain NADP<sup>+</sup> complex structure also provides information on the mechanism of electron transfer and the number of NADP<sup>+</sup> molecules that interact with the

FAD domain that has recently been debated. The hydride transfer process can be modelled on a two site binding model in which two NADPH coenzyme molecules bind to the catalytic site and a second regulatory site within the FAD domain of BM3 or regions within a bipartite coenzyme binding site. Possible location of these sites within close proximity to each other and the release of NADPH being sterically hindered by NADPH molecules bound in the regulatory region may explain the slow release of NADP<sup>+</sup> (Roitel *et al.*, 2003). However, recent analysis (Daff, 2004) calls this reasoning into question and indicates the hyperbolic dependence on NADP<sup>+</sup> is due to a reversal of the hydride transfer process declaring that a single binding site models the data accurately. The NADP<sup>+</sup> bound structure of BM3 FAD/NAD(P)H shows only a single NADP<sup>+</sup> binding site (at 10 mM NADP<sup>+</sup>) and provides no evidence for a two binding site model indicating that the single binding site model is correct. The hyperbolic dependence on NADP<sup>+</sup> concentration was also seen in the reductase FprA (McLean *et al.*, 2003) whose protein structure also reveals a single NADP<sup>+</sup> molecule bound (Bossi *et al.*, 2002).

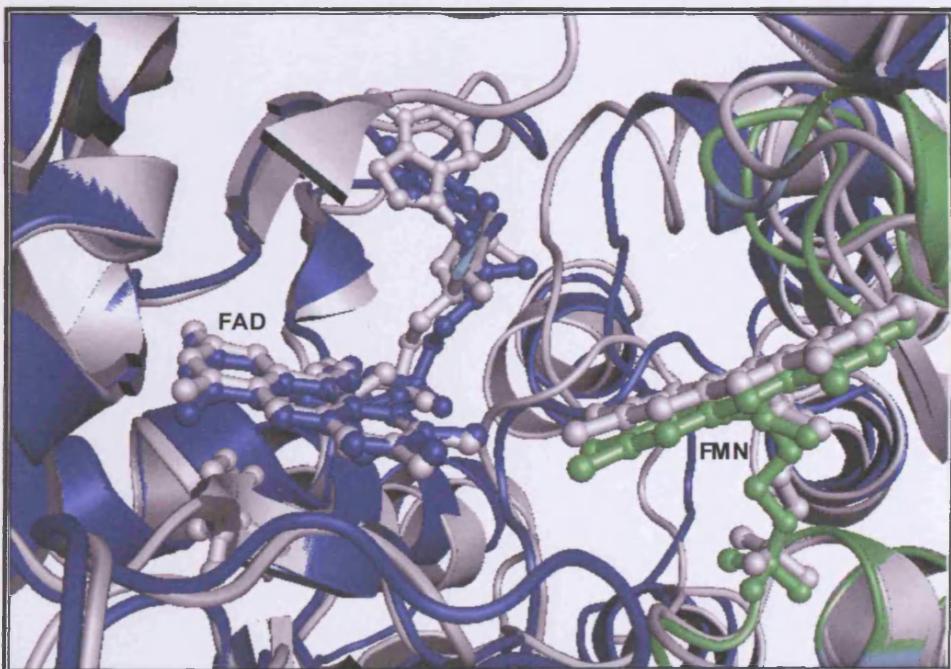
#### 5.4.6. Domain organisation

Using CPR as a template model, it is possible to construct a model for the full BM3 reductase component using the individual FAD and FMN domain structures (Figure 5.20). Superimposing the FAD/NADPH and FMN binding domains on the CPR template using LSQKAB (Kabsch, 1976) leads to little or no clashes. The FAD domain of BM3 aligns with a RMSD of 1.636 and the FMN domain of BM3 aligns with a RMSD of 1.420 with the rat CPR structure. The FMN cofactor of the FMN domain is located in a relatively solvent exposed position that allows interaction with the FAD cofactor. The flavin cofactors align in an almost identical position to that observed for the rat CPR structure with the isoalloxazine rings approximately 3 Å apart which is necessary for highly efficient electron transfer (Figure 5.21).

Analysis of the surface complementarity of the BM3 FAD domain and the FMN domain by SC (Lawrence and Colman, 1993) gives a surface interface of 270 Å<sup>2</sup> and surface complementarity value of 0.309. The surface complementarity value for the FMN and FAD domains of rat CPR is 0.641 with a surface interface of 640 Å<sup>2</sup>. The difference in the surface complementarity is likely explained by the crude generation of the model and that individual residues may require local reorientation that would lead to an improved surface complementarity.



**Figure 5.20** Overlay of rat CPR shown in cyan cartoon representation with the BM3 FAD domain structure shown in blue. The FAD cofactor for BM3 is shown in yellow stick representation.



**Figure 5.21** Modelled structure of the BM3 reductase domain with the FAD and FMN prosthetic groups shown in stick format with the stacking aromatic residues also shown in stick format. The protein domains are shown in cartoon representation (FAD: blue and FMN: green). The modelled structure is generated by a least squares structural alignment with the rat CPR structure shown in grey cartoon representation with FMN and FAD cofactors shown in ball and stick representation.

The overall surface charge of the BM3 FAD domain is shown in Figure 5.22A. The structure is largely negatively charged with an area of positively charged surface seen where the NAD(P)H binding site is located and also in the region where the FMN domain comes

into contact with the FAD domain. Analysis of the charge of the FMN domain of BM3 indicates that this surface is also largely negatively charged shown in Figure 5.22B. The regions that interact directly with the FAD domain are however less negatively charged. The overall charge distribution of the cofactor domains is highly similar to that seen in rat CPR indicating that charge interaction may be important in the correct orientation of the domains.

Analysis of the BM3 P450 domain also indicates that this domain is largely negatively charged but it displays a large positive region on its proximal surface. This would allow interaction with the negatively charged FMN domain allowing electron transfer to the heme group of the cytochrome P450 module.

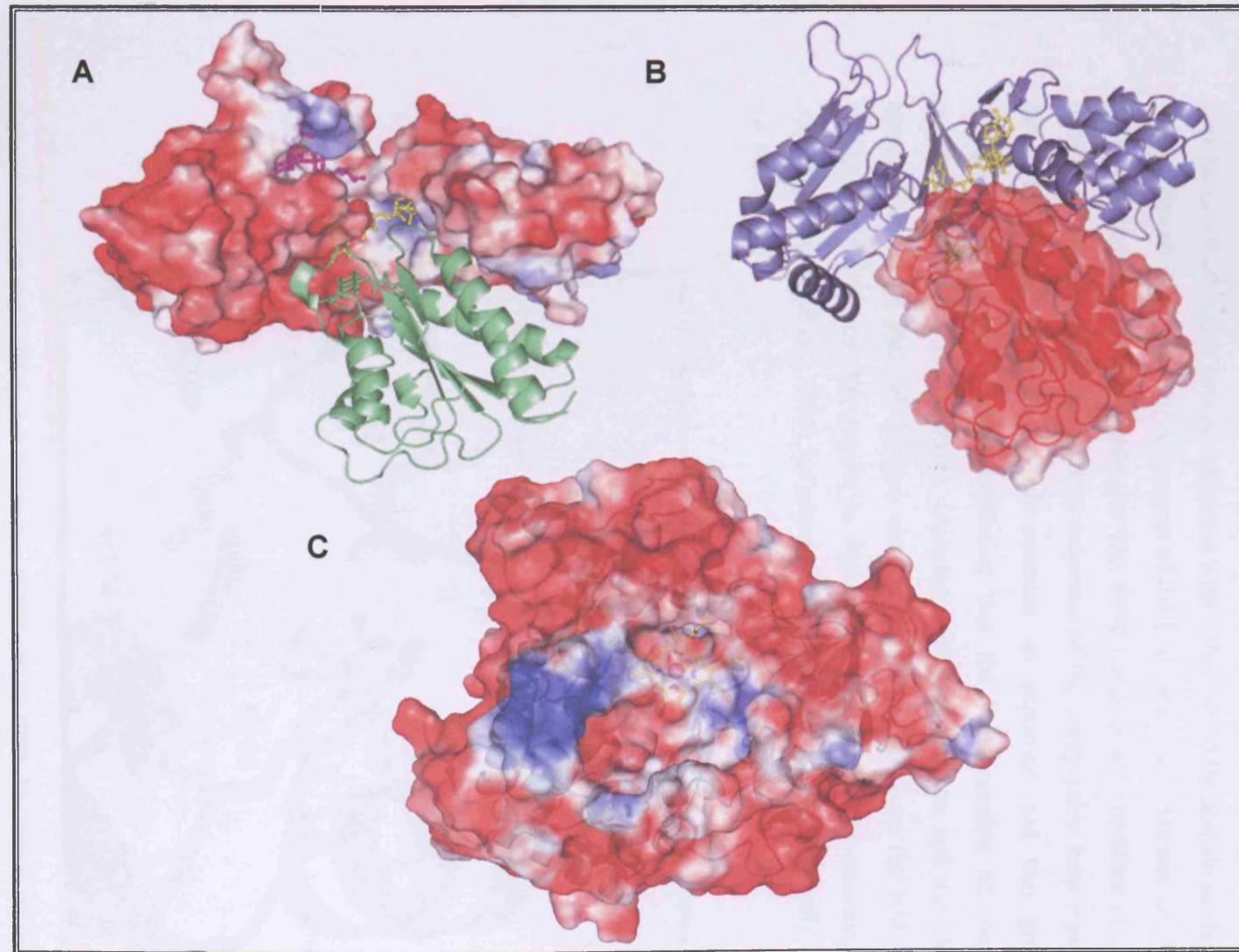
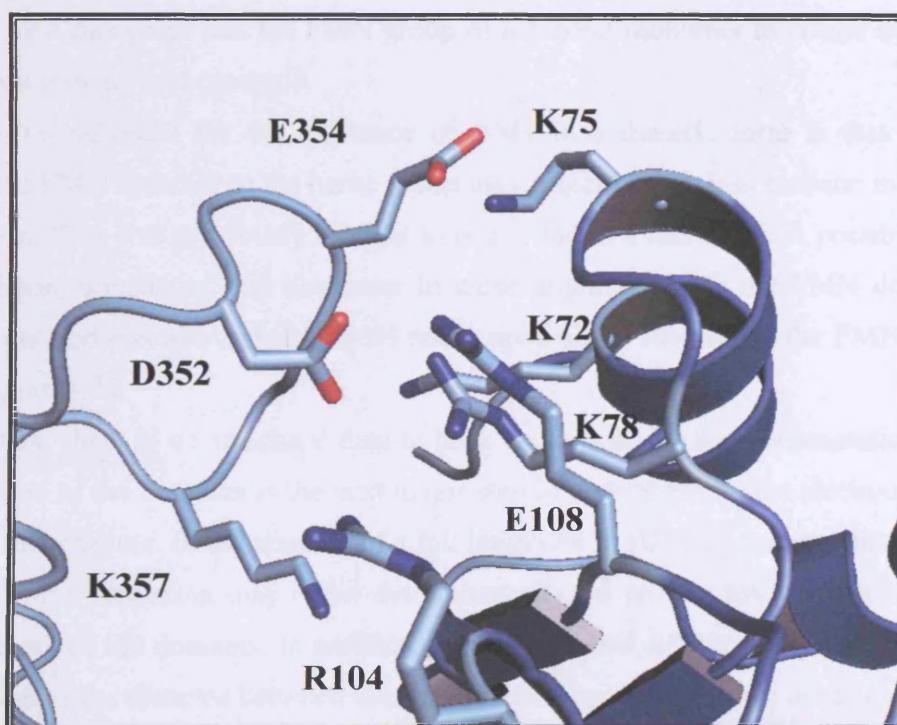


Figure 5.22 A) Surface representation of BM3 FAD domain shown in surface representation generated using APBS (ref) in complex with BM3 FMN domain as generated using rat CPR as a model. B) Surface representation of BM3 FMN domain shown in complex with FAD domain. C) Surface representation of the BM3 heme domain (PDB code: 2HPD) generated using APBS shown from an orientation proximal to the heme. The surface is shown in transparent mode with the cartoon representation of the protein also displayed and the heme cofactor shown in stick representation.

The residues that are missing are in BM3 FAD A molecule are residues 732-736, molecule B residues 731-742 and in the NADP<sup>+</sup> complexed structure A molecule 731-743 and in the B molecule 730-742 indicating a considerable amount of mobility in this loop region. The sequence from 729-744 is **LEAEEEKLAHLPLAKT**. This sequence contains four glutamate residues and two lysine residues indicating that this region is highly charged and may be involved in salt bridge formation with either the FMN domain or the heme domain in the full length BM3 protein. Alignment of the FAD and FMN domains of BM3 using the rat CPR structure as a template indicates that these residues are important in aligning the FMN domain as seen in Figure 5.23. The sequence of the comparable loop region in rat CPR is **NNLDEESNKHKPFPCPT** which contains an aspartate and two glutamate residues alongside two lysine residues indicating that the large number of charged residues is conserved within the two proteins. Utilising CPR as a template and the BM3 FMN domain structure indicates that salt bridges can be constructed between the BM3 FAD and FMN domains. Analysis of this region in the structurally similar rat neuronal NOS reductase structure (Garcin *et al.*, 2004) indicated this loop region is not structurally conserved within the family.



**Figure 5. 23 FAD and FMN domain interacting residues of rat CPR.** The loop at the left of the picture is the homologous region that is missing in the BM3 FAD domain. Charged residues are shown in stick representation with the protein molecule shown in cartoon representation.

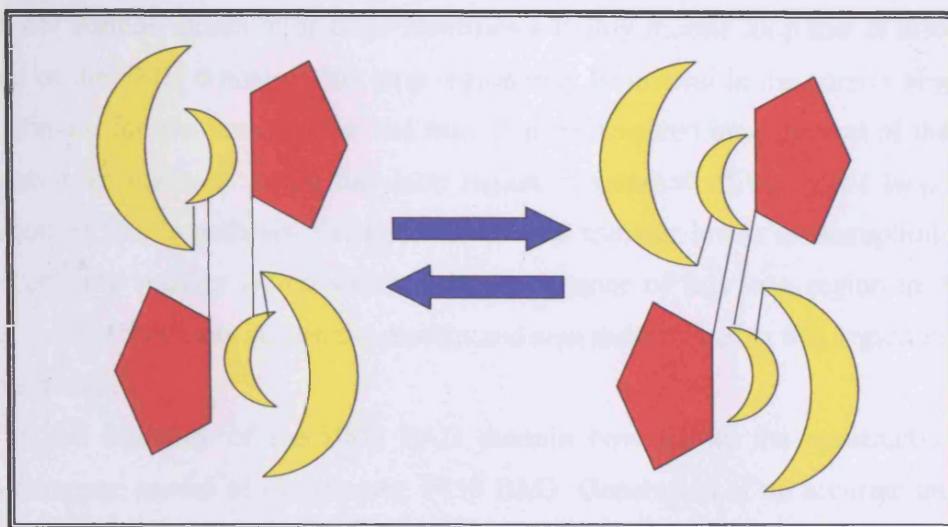
The transfer of electrons from the FMN cofactor to the heme of the cytochrome P450 was previously thought to occur within a monomer with significant movement of the FMN

domain as proposed by Munro et al., 2003. This proposed domain motion is required due to the alignment of the FMN cofactor within 14 Å of both the FAD cofactor and the heme cofactor. This spatial requirement is required to allow efficient electron transfer to occur with values in the order of the experimentally determined values. Large scale mobility of flavin domains have been previously seen in a number of flavoproteins as discussed in Section 1.4.2.1. Although all of the components of the BM3 holoenzyme have now been determined, the linker regions between these domains have not been determined and in the case of heme to FMN domain the linker region is 21 residues in length and from the FMN domain to the FAD domain the linker is 29 residues which does not impose large restraints on the positions of the domains.

A recent study by Neeli *et al.* (2005) has shown that the dimeric form of the protein is catalytically functional and the physiological form of the enzyme. At low enzyme concentrations the protein becomes a monomer and in this state the protein's fatty acid hydroxylase activity is dramatically reduced indicating that the functional form of the protein is a dimer. At the lower enzyme concentration efficient electron transfer occurs between the FAD and FMN cofactors but transfer to the heme group is not catalytically efficient. Analysis of mutant enzyme forms indicate that the normal mode of electron transfer is between the heme group of one monomer and the FMN group of a second monomer molecule leading to efficient electron transfer and catalysis.

A possible rationale for the existence of BM3 in a dimeric form is that electron transfer from the FMN cofactor to the heme group may require much less extreme movement in a dimeric form than was previously thought to occur within a monomer. A possible model positions the heme domain of one monomer in close alignment with the FMN domain of another monomer and electron transfer could occur upon slight rotation of the FMN domain as shown in Figure 5.24.

At present, there is no structural data to back up this model and determination of the correct orientation of the domains is the next major step in understanding the electron transfer mechanism of this enzyme. In the absence of a full length P450 BM3 crystal structure, the use of small angle x-ray scattering may allow determination of a protein fold that would allow accurate placement of the domains. In addition, the use of novel EPR techniques could allow the determination of the distance between spin labels resulting in the correct orientation of the BM3 domains relative to one another within the dimer.



**Figure 5.24 Putative model of the holoenzyme BM3.** The heme domains are shown as red pentagons and the FAD domain is shown by the large yellow moon shaped object and the FMN domain by the smaller moon shaped segment.

## 5.5. Conclusions and Future Work

The crystal structure of the C773A mutant of the BM3 FAD/NAD(P)H domain in the presence and absence of NADP<sup>+</sup> was determined. This allows the identification of the residues that are important in FAD and NAD(P)H binding to cytochrome P450 BM3. It also confirms that the structure is highly similar to previously solved crystal structures of members of the FNR family and in particular the rat CPR structure. This structure validates the change in enzyme activity and electron transfer that was observed following the C999A mutant in both the reductase domain and the full length enzyme.

Crystallisation of the FAD domain was only possible after generation of the C773A mutant. Structure determination of the full reductase domain of BM3 which incorporates this mutant may also be possible and would be valuable as comparison with the rat CPR structure and the putative reductase model that is discussed in Section 5.4.6.

The protein structure of the FAD domain furthermore identifies the possibility of disulphide bond formation between C773 and C811 as a reason for the as yet elusive crystallisation of the full length P450 BM3. The dual C773A/C811A mutant should be pursued in an attempt to generate the crystal structure of the full length BM3 holoenzyme.

Structure determination of the native FAD domain and the W1046H/A mutants in complex with NAD<sup>+</sup> would allow elucidation of the direct role of the tryptophan residue in cofactor differentiation and the role of the NADP<sup>+</sup> ribosyl phosphate in binding and electron transfer to BM3.

The FAD domain structure of BM3 identifies a highly mobile loop that is disordered in the absence of the FMN domain. This loop region may be critical in the correct alignment of the FMN domain for electron transfer and may also be involved in alignment of the heme domain. Mutation of residues within this loop region or removal of the entire loop would allow validation of this hypothesis. Changes in electron transfer levels or disruption of the high rates of enzyme activity would validate the importance of this loop region in domain orientation of the FMN domain within the protein and also assess whether this region interacts with the heme domain.

The crystal structure of the BM3 FAD domain now allows the construction of a complete holoenzyme model of cytochrome P450 BM3. Generation of an accurate model in the absence of the crystal structure of the holoenzyme requires determination of the orientation of the domains relative to one another. The determination of domain orientation may be carried out by small angle x-ray scattering to determine the overall fold of the protein. Novel EPR techniques could then allow the domains to be correctly placed within the determined protein fold outline. It would also be possible to assess structural changes in individual domain orientation prior to and after NADPH binding and subsequent electron transfer and could lead to a complete structural picture of the protein throughout the three electron transfer steps.

## **Chapter 6**

# **Crystal structure of the transcriptional regulator CprK**

## 6.1. Introduction

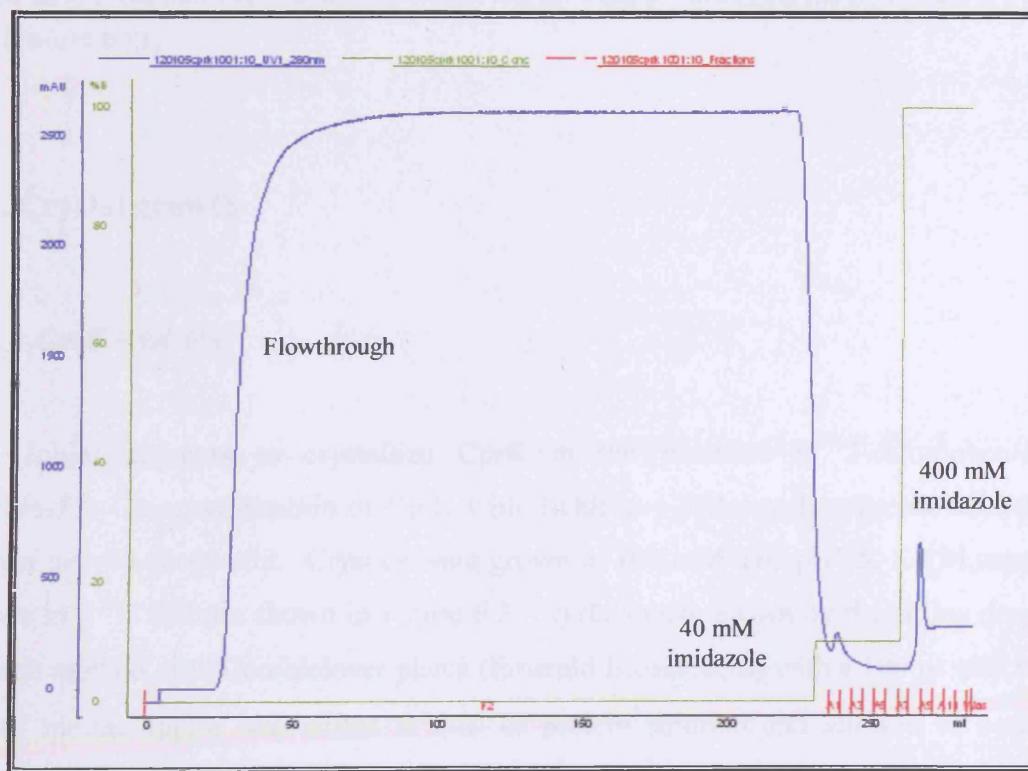
Halogenated phenolic compounds are significant pollutants that have been introduced into the environment through use in industry and as by products during chemical synthesis (van Pee and Unversucht, 2003). Contrary to common belief that halogenated phenolic compounds are entirely man made in origin, there is evidence that similar compounds are also synthesized by micro-organisms (Gribble, 1998). It thus comes as no surprise that organisms exist that effectively break down halogenated phenolic compounds. Recent advances in the analysis of such organisms that colonize polluted regions have led to the identification of a number of bacterial species that can carry out dehalogenation (Jansson, 1994). This process is generally considered to be more efficient in anaerobic bacteria (Jansson *et al.*, 2001) and is coupled to a process known as halorespiration. The dehalogenase enzyme (the terminal electron acceptor in the respiratory pathway) from *D. dehalogenans* has been purified and characterized (van de Pas *et al.*, 1999) and shown to exist in a wide range of *Desulfitobacterium* sp. Further genetic characterization led to the identification of an operon containing the dehalogenase gene in addition to a number of proteins that are involved in the regulation and expression of this enzyme (Smidt *et al.*, 2000). Sequence comparison of the CprK protein indicated that this protein was from the CRP-FNR family of transcriptional regulators and subsequent characterization of this protein indicated its role as the transcriptional regulator of the *cpr* operon. It was shown to bind to a DNA region at the -10 and -35 region upstream of the dehalogenase gene only in the presence of OCPA (*ortho*-chlorophenolacetic acid; Pop *et al.*, 2004).

Further analysis of the genome of the highly related bacteria *D. hafniense* has shown that three similar operons exist in this bacterium (Frishman *et al.*, 2003) each containing a CprK homologue. In order to understand how CprK translates the presence of OCPA into DNA binding we determined the CprK crystal structure. This chapter describes the 2.2 Å crystal structure of CprK from *D. hafniense* as a covalent dimer in complex with OCPA determined under aerobic conditions. This structure provides detailed atomic resolution information on the OCPA binding mechanism. In combination with solution data on WT CprK and mutant forms, this structure reveals the molecular basis for halogen binding and the possible redox-regulation of CprK.

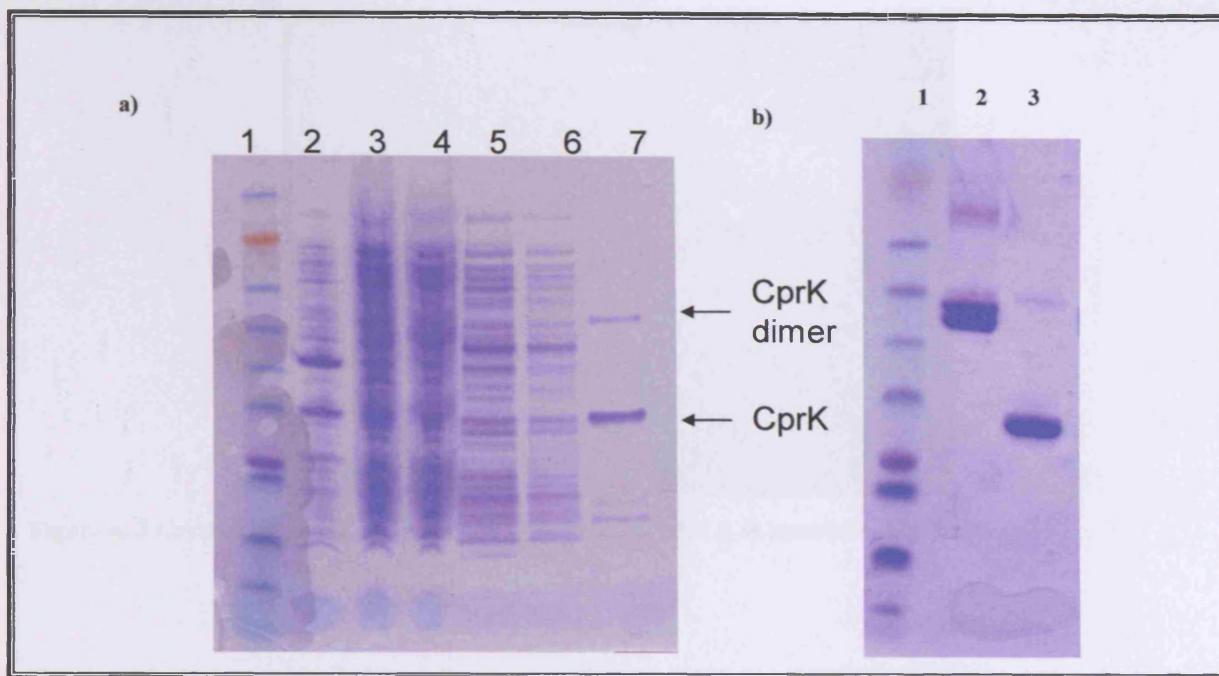
## 6.2. Experimental Results

### 6.2.1. Purification of CprK

Protein was purified from *E. coli* BL21 (DE3) cells containing a pET vector 21a+ with *cprK* gene which was kindly donated by Dr. J. van der Oost (Wageningen, The Netherlands) as described in Section 3.6.2 with a typical chromatographic purification profile using a Ni-NTA column shown below (Figure 6.1). Samples from the purification procedure were analyzed by SDS-PAGE as shown in Figure 6.2a. In the presence of 1 mM DTT (incubated for 5 min) two bands are visible indicating the presence of a dimer and monomer, increased concentrations of DTT show just monomer (Figure 6.2b) demonstrating that CprK is likely a disulphide linked dimer. Upon elution from the Ni-NTA column, the sample was desalted using an Amersham HiPrep 26/10 Desalting column. If the protein was left overnight in the purification buffer, large amounts of the purified protein aggregated and became insoluble.



**Figure 6.1** Purification profile of CprK from Akta Purifier 100 with absorption values at 280 shown in blue and imidazole concentration shown in green.

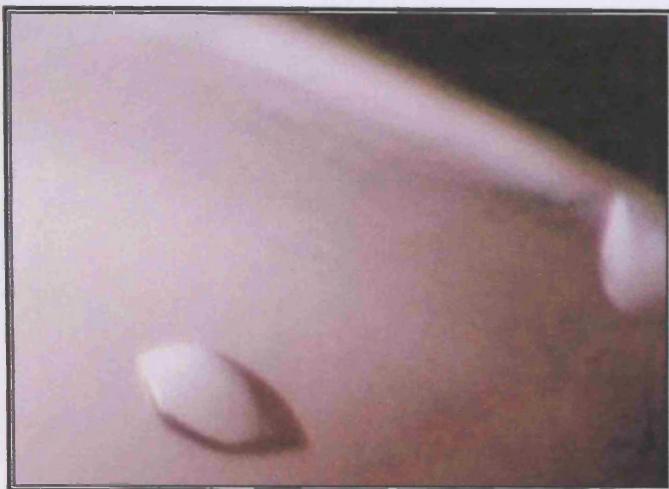


**Figure 6. 2 a)** SDS-PAGE gel of purification steps for CprK Lane 1: SeeBlue Plus 2 molecular weight marker Lane 2: Pellet sample Lane 3: Sample applied Lane 4: Flowthrough Lane 5: Initial wash Lane 6: Low imidazole wash Lane 7: Eluted protein. All samples were loaded in 5 µl loading buffer with 1 mM DTT. **b)** Lane 1: SeeBlue Plus 2 molecular weight marker Lane 2: Pure CprK (no DTT) Lane 3: Pure CprK (10 mM DTT)

### 6.2.2. Crystal growth

#### 6.2.2.1. *CprK + OCPA*

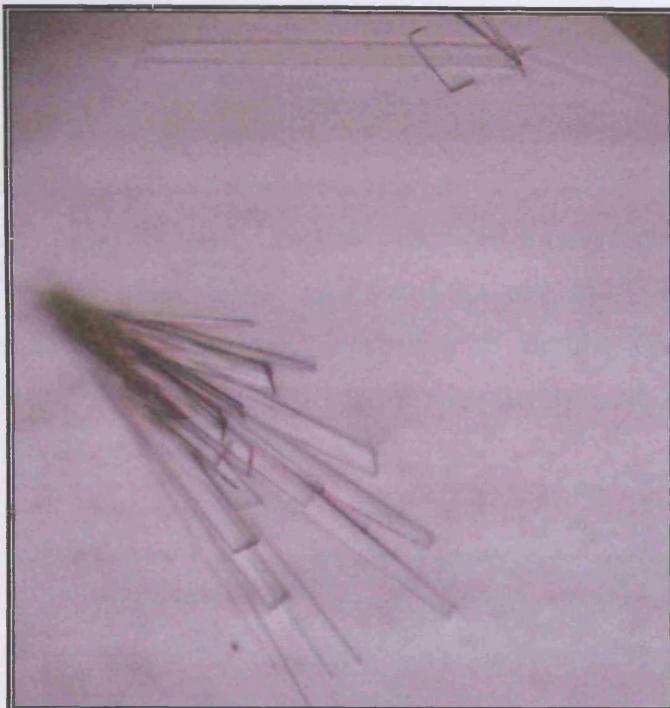
Initial attempts to crystallize CprK in the presence of 2-chlorophenol were unsuccessful. Co-crystallisation of CprK with 3-chloro-4-hydroxyphenylacetic acid (OCPA) however proved successful. Crystals were grown in 100 mM Tris pH7.5, 1.8 M ammonium sulphate at 4 °C and are shown in Figure 6.3. Crystals were grown by the sitting drop vapor diffusion method with CombiClover plates (Emerald Biosciences) with a 100 µl well volume. 2 µl of mother liquor was added to 2 µl of protein solution and allowed to equilibrate. Crystals of the CprK-OCPA complex grew within 5-7 days and typically grew to dimensions 0.15 x 0.15 x 0.3 mm. The crystals were cryoprotected using mother liquor plus 20 % glycerol and flash-cooled by plunging into liquid nitrogen.



**Figure 6. 3 Crystals of CprK grown in 100 mM Tris pH 7.5, 1.8 M ammonium sulphate.**

#### 6.2.2.2. *CprK*

Initial attempts to crystallize CprK in absence of chlorinated ligand with a range of commercially available screens failed despite variations in pH, temperature, salt and precipitant in conjunction with the use of a range of protein concentrations from 5-20 mg/ml. Improved purification of CprK over a shorter time period and with desalting and protein concentration carried out immediately upon elution from the Ni-NTA column led to growth of needle shaped crystals. Crystallisation was carried out in a sitting drop vapor diffusion experiment with a protein concentration of 10 mg/ml and mother liquor containing 10 % PEG 3000, 100 mM Magnesium acetate pH 5.5. Optimization of these crystal conditions was difficult and ultimately improvement was observed with the addition of 2,5 hexanediol leading to crystal plates with dimensions 10 x 50 x 300  $\mu\text{M}$  in size as shown in Figure 6.4. Macro and micro-seeding techniques failed to improve crystal size. Crystals were cryoprotected using mother liquor plus 20 % glycerol and flash-cooled by plunging into liquid nitrogen.



**Figure 6. 4 Crystals of CprK which were grown in 0.2 M Calcium Chloride dihydrate, 20% w/v Polyethylene Glycol 3350, pH 5.1, 4 % v/v 2, 5 hexanediol.**

### 6.2.3. Heavy atom derivatisation

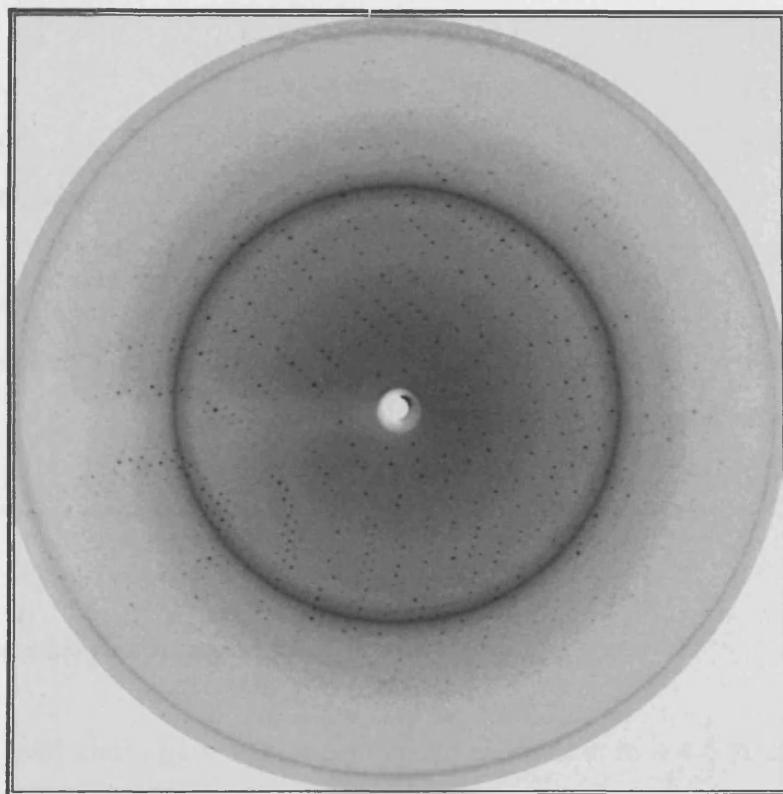
Heavy atom derivatisation of OCPA – CprK crystals was attempted with a number of derivatives with a range soaking times. In the case of 5-10 mM solutions of potassium dicyanoaurate (I) and trimethyl lead (IV) acetate, crystal cracking was observed within 5-10 minutes. Soaking crystals in mother liquor containing 10 mM potassium tetrachloroplatinate (II) and 10 mM mercury (II) acetate for 10 minutes gave no visible change in crystal quality. These crystals were transferred to mother liquor plus 20 % glycerol to act as cryoprotectant.

### 6.2.4. X-ray data collection and processing

#### 6.2.4.1. *CprK + OCPA*

A complete native data set to 2.2 Å was collected on a single crystal at 100K using an Oxford Cryostream cooling device, with oscillations of 1° at DESY, Hamburg, beamline BW7A. All images were collected using a Mar 165mm CCD detector (Figure 6.5). The data was processed and scaled using the HKL2000 programs DENZO and SCALEPACK

(Otwinowski and Major, 1997) and found to belong to the orthorhombic space group I 222 or I<sub>2</sub>I<sub>2</sub>I<sub>2</sub> with unit cell dimensions  $a=104.4 \text{ \AA}$ ,  $b=112.2 \text{ \AA}$ ,  $c=119.5 \text{ \AA}$ ,  $\alpha=\beta=\gamma=90^\circ$ . Data statistics for data processing and scaling are given in Table 6.1. Assuming two molecules are present per asymmetric unit gives a Matthews's coefficient of  $3.5 \text{ \AA}^3 \text{ Da}^{-1}$  and a solvent content of 64.5 % (Matthews, 1968). Initial diffraction data from putative platinum and mercury derivatised crystals was compared to the native data set and shown to contain sufficient differences to warrant collection of complete data sets. Two complete data sets to 2.5 and 3.1  $\text{\AA}$  for a platinum and mercury derivatised crystal respectively were collected at DESY, Hamburg, beamline BW7A.



**Figure 6. 5** Typical 1° oscillation image obtained during data collection of CprK with diffraction spots extending to ~2 Å at BW7A DESY-Hamburg.

**Table 6.1 Data scaling and processing statistics**

	<b>Native</b>	<b>potassium tetrachloro-platinate (II) soak</b>	<b>mercury (II) acetate soak</b>
<b>Soak time/Concentration</b>		10min/10mM	10min/10mM
<b>Space group</b>	I 222	I 222	I 222
<b>Cell dimensions</b>			
<b>a (Å)</b>	104.4	104.9	105.2
<b>b (Å)</b>	112.1	113.0	112.3
<b>c (Å)</b>	119.4	119.6	117.8
<b>X-ray source</b>	DESY BW7A	DESY BW7A	DESY BW7A
<b>Resolution (Å)</b>	50-2-2	50-2.5	50-3.1
<b>No. of observations</b>			
<b>Total</b>	766,373	313,230	423,430
<b>Unique</b>	35,512	24,333	12,468
<b>Completeness (%)</b>	99.2 (100)	97.6 (98.6)	95.8 (99.8)
<b>I/σI</b>	16.6 (2.5)	17.08 (2.906)	8.97 (2.31)
<b>R<sub>merge</sub></b>	0.102 (0.573)	0.078 (0.427)	0.147 (0.415)

#### 6.2.4.2. CprK crystals in absence of OCPA

CprK crystal plates gave very weak crystal diffraction to ~ 4-5 Å at a number of 3<sup>rd</sup> generation synchrotron beamlines. However, using the Microfocus beamline ID-13 at ESRF-Grenoble with a 10 μM x-ray beam, it was possible to observe diffraction to ~3 Å with a 2 second exposure over a 1° oscillation. Due to the highly intense beam, radiation damage was severe with diffraction limits decreasing to ~5 Å after seven degrees of data collection. The crystal was then translated to expose a fresh region of the crystal unaffected by radiation damage and data collection continued. An overlap of 2 ° between data collection segments was used in order to improve data scaling. Thirty segments of data could be collected from a single crystal.

Segments of 5 ° were indexed and processed using DENZO and shown to be of the orthorhombic space group of C222<sub>1</sub> or C222 with unit cell dimensions of  $a= 79.95$  Å,  $b=286.7$  Å,  $c=65.279$  Å,  $\alpha=\beta=\gamma=90^\circ$ . Data was scaled using SCALEPACK and subsequently

also scaled using SADABS (Sheldrick, 1996) with the help of Prof. GM Sheldrick which resulted in improved scaling statistics for the data. Data statistics for the processing and scaling are given in Table 6.2.

Attempts to solve this structure using the CprK-OCPA complex structure by molecular replacement using both Amore and Phaser were unsuccessful. This may be due to the poor quality and completeness of the data and errors introduced with scaling.

**Table 6.2 Data statistics of CprK data collection and scaling using SCALEPACK and SADABS**

Scaling program	SCALEPACK	SADABS
Cell dimensions		
a (Å)	79.9	79.9
b (Å)	286.7	286.7
c (Å)	65.2	65.2
X-ray source	ID13-ESRF	ID13-ESRF
Resolution (Å)	40-3.3 (3.42-3.3)	40-3.5 (3.6-3.5)
No. of observations		
Total	31,209	28,954
Unique	10,887	11,893
Overall redundancy	2.9	2.4
% Completeness	90.2 (73.3)	(80)
I/σI	5.966 (1.41)	5.93 (2.48)
R <sub>merge</sub>	0.140 (0.290)	0.1234 (0.371)

## 6.2.5. Structure Solution

### 6.2.5.1. *CprK+OCPA*

Several attempts to solve the OCPA-CprK structure by molecular replacement were carried out with a number of different models but yielded no clear solutions. This is not surprising as the best model available is CRP with only 21 % sequence identity to CprK. Heavy atom derivitisation was thus carried out as described above to gain experimental phase information.

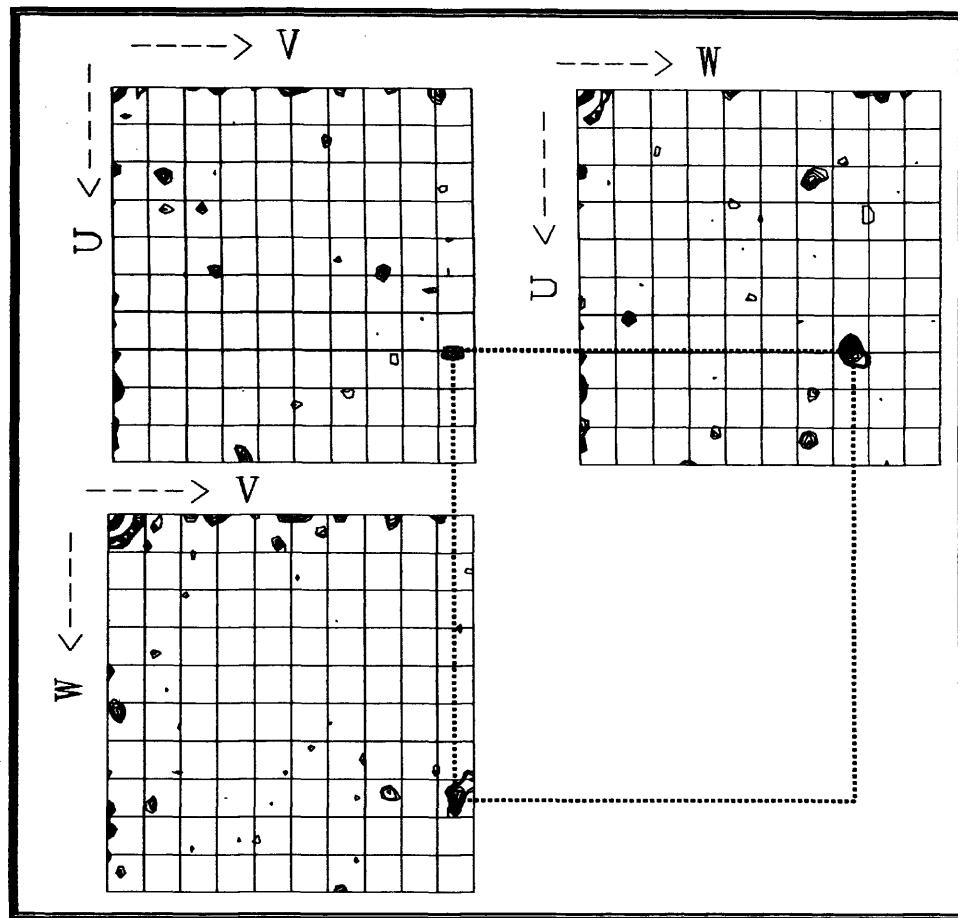
### **6.2.5.2. Location and Refinement of Heavy Atom Sites**

There are 4 symmetry operations that generate the 222 symmetry in an icosahedral orthorhombic lattice (International Tables for Crystallography, volume A; Hahn (ed), 1996). Thus in space group I 222, the heavy atom located at the position (u, v, w) within the unit cell must also be located at positions (-u, -v, w), (-u, v, -w) and (u, -v, -w). The Patterson map will therefore contain peaks at (2u, 2v, 0), (2u, 0, 2w) and (0, 2v, 2w). These peaks fall into three Harker sections with respectively x=0, y=0 and z=0. Patterson difference maps were calculated using FFT (Ten Eyck, 1973) and these show peaks corresponding to a first platinum site as indicated in Figure 6.6 on the three Harker planes. These peaks identify a single heavy atom binding site with fractional coordinates u = 0.7, v = 0.95, w = 0.75.

The position of additional platinum sites was determined using difference Fourier analysis using the initial SIR phases generated by MLPHARE on the basis of the first platinum site (Otwinowski, 1991). The position of mercury atoms was found using cross difference Fourier analysis. Coordinates and real and anomalous occupancies were refined for all heavy atom position and final MIRAS phases calculated using MLPHARE with phasing statistics shown in Tables 6.3 and 6.4. The overall figure of merit (FOM) of the resulting phase set was 0.464 for 35,391 reflections between 20 and 3.5 Å. Experimental electron density maps were produced that revealed clear solvent boundaries and secondary structure features.

**Table 6.3 Phasing statistics for the mercury and platinum derivatives**

	<b>Mercury derivative</b>	<b>Platinum derivative</b>
<b>R<sub>calcs</sub> (acentric/centric)</b>	0.82/0.74	0.85/0.74
<b>Phasing power (acentric/centric)</b>	1.17/0.96	0.99/0.85



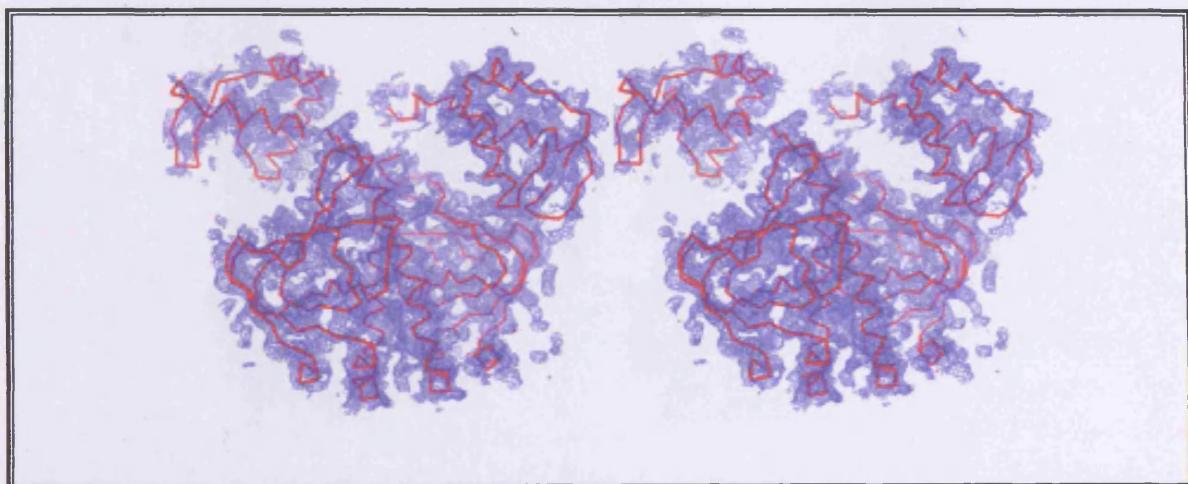
**Figure 6.6** Harker sections of the difference Patterson map for the Platinum derivative of CprK at 3.0 Å resolution. The sections are contoured starting at 0.5  $\sigma$  and increasing in 0.1  $\sigma$  steps.

**Table 6.4** The fractional coordinates and occupancies for the heavy atoms of the two derivatives used in the generation of the MIR phase set.

Derivative/Atom	x	y	z	Relative occupancy	
				Real	Anomalous
Hg					
Atom 1	0.291	0.059	0.881	0.413	0.331
Atom 2	0.282	0.145	0.139	0.419	0.319
Atom 3	0.432	0.101	0.787	0.292	0.208
Atom 4	0.239	0.146	0.919	0.196	0.163
Atom 5	0.243	0.153	0.340	0.231	0.179
Pt					
Atom 1	0.330	0.474	0.370	0.179	0.222
Atom 2	0.256	0.089	0.284	0.183	0.198
Atom 3	0.297	0.228	0.500	0.095	0.090
Atom 4	0.505	0.038	0.216	0.063	0.062

### 6.2.5.3. Map improvement

An improved electron density map was obtained by combined solvent flattening and histogram matching as implemented in DM (Cowtan, 1994), using a solvent content of 64%. Molecular masks for use in non-crystallographic symmetry (NCS) averaging were calculated by manual positioning of both the DNA binding domains and inducer binding domains of CRP in the density map as shown in Figure 6.7. Matrices describing the spatial correlation between individual domains were calculated using LSQKAB (Kabsch, 1976). Initial map correlation for the inducer binding domains was 0.273 and 0.313 for the DNA binding domains. Following density modification, the final values were 0.901 and 0.876 respectively. Following NCS averaging in combination with solvent flattening and histogram matching, the resultant FOM for the data from DM was 0.753. A stepwise and marked improvement in the electron density maps using these methods is illustrated in Figure 6.8.



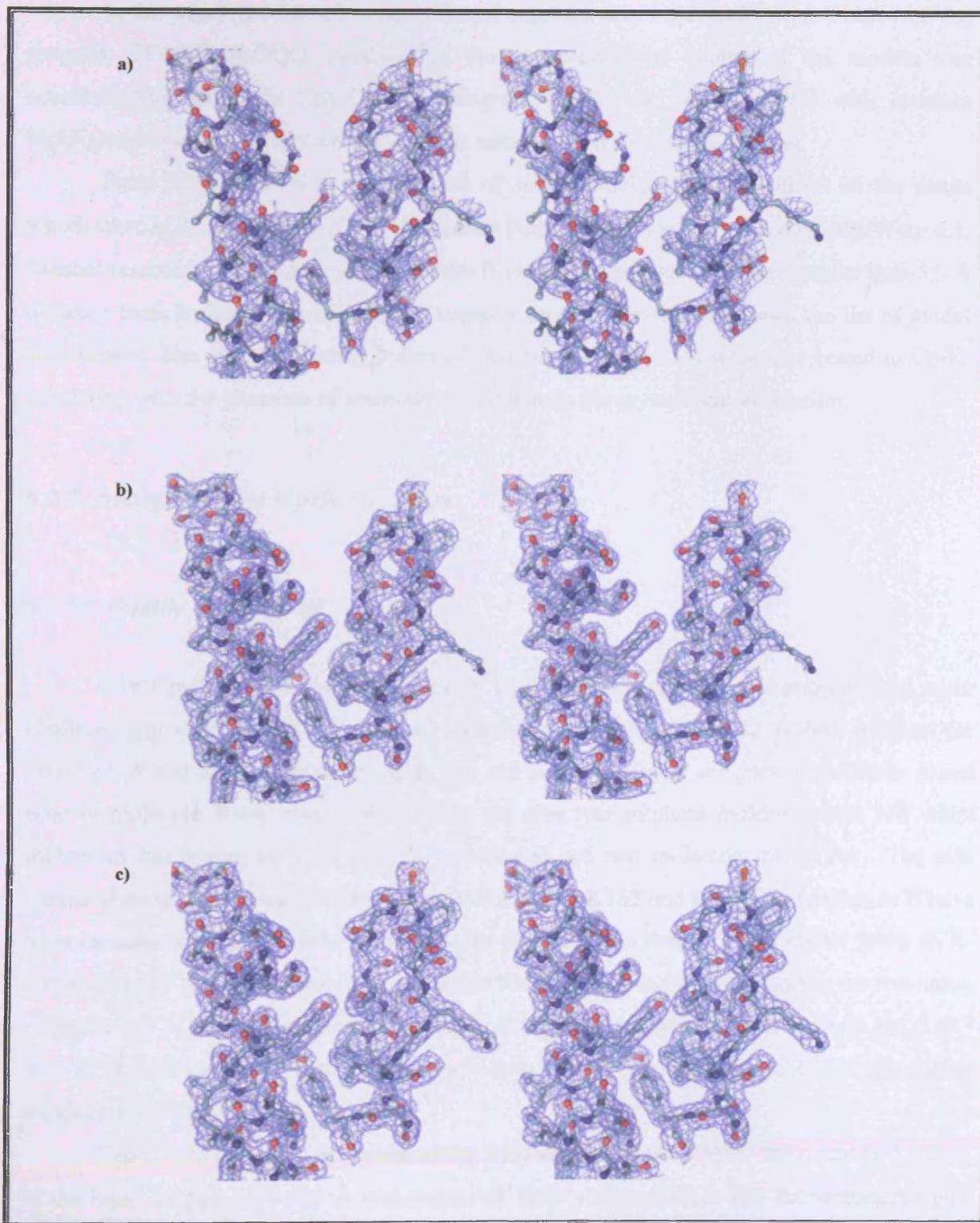
**Figure 6.7 Generation of the NCS averaging matrices. Stereo diagram of the electron density of the CprK-OCPA complex structure shown with overlaid domains of CRP shown in ribbon representation.**

### 6.2.6. Building and refinement of CprK protein structure model

#### 6.2.6.1. Interpretation of the Electron Density Map

Initial interpretation of the electron density map allowed manual placing of the CRP inducer binding domain. Part of the DNA binding domain could also be positioned in the density. However, the density in this region was of lower quality particularly in the region connecting the DNA and inducer binding domains. Following the density modification procedures described in Section 6.2.5.3. electron density was much improved. Using the CRP structure as a template, large segments of the structure were manually built with TURBO-

FRODO. After an initial model was built, the model was refined against crystallographic data extended to 2.2 Å.



**Figure 6.8** Stereo diagrams of the improvement in the electron density of CprK. a) Initial electron density b) density following density modification and NCS averaging c) final electron density. The final built protein model of residues 125–141 (left) and 2–15 (right) from Molecule A shown in stick representation is superimposed over the electron density for reference.

### **6.2.6.2. Model Building and Refinement**

In the course of the refinement manual adjustments of the model were made with the program TURBO-FRODO. Analysis of the stereo-chemical quality of the models was accomplished using the PROCHECK program (Laskowski *et al.*, 1993) with residues highlighted assessed for agreement with the calculated electron density maps.

Final refinement with the addition of solvent molecules at positions of the peaks which were at least  $3\sigma$  in an Fo-Fc difference Fourier map was carried out by Arp/Warp 6.1. Manual removal of all water molecules with B values larger than  $60\text{ \AA}^2$  and greater than  $3.5\text{ \AA}$  distance from hydrogen bond donor or acceptor atoms were removed from the list of model coordinates. The electron density indicated that two sulphate ions were also bound to CprK, consistent with the presence of ammonium sulphate in the crystallisation solution.

### **6.2.7. Analysis of the CprK structure**

#### **6.2.7.1. Quality of the model**

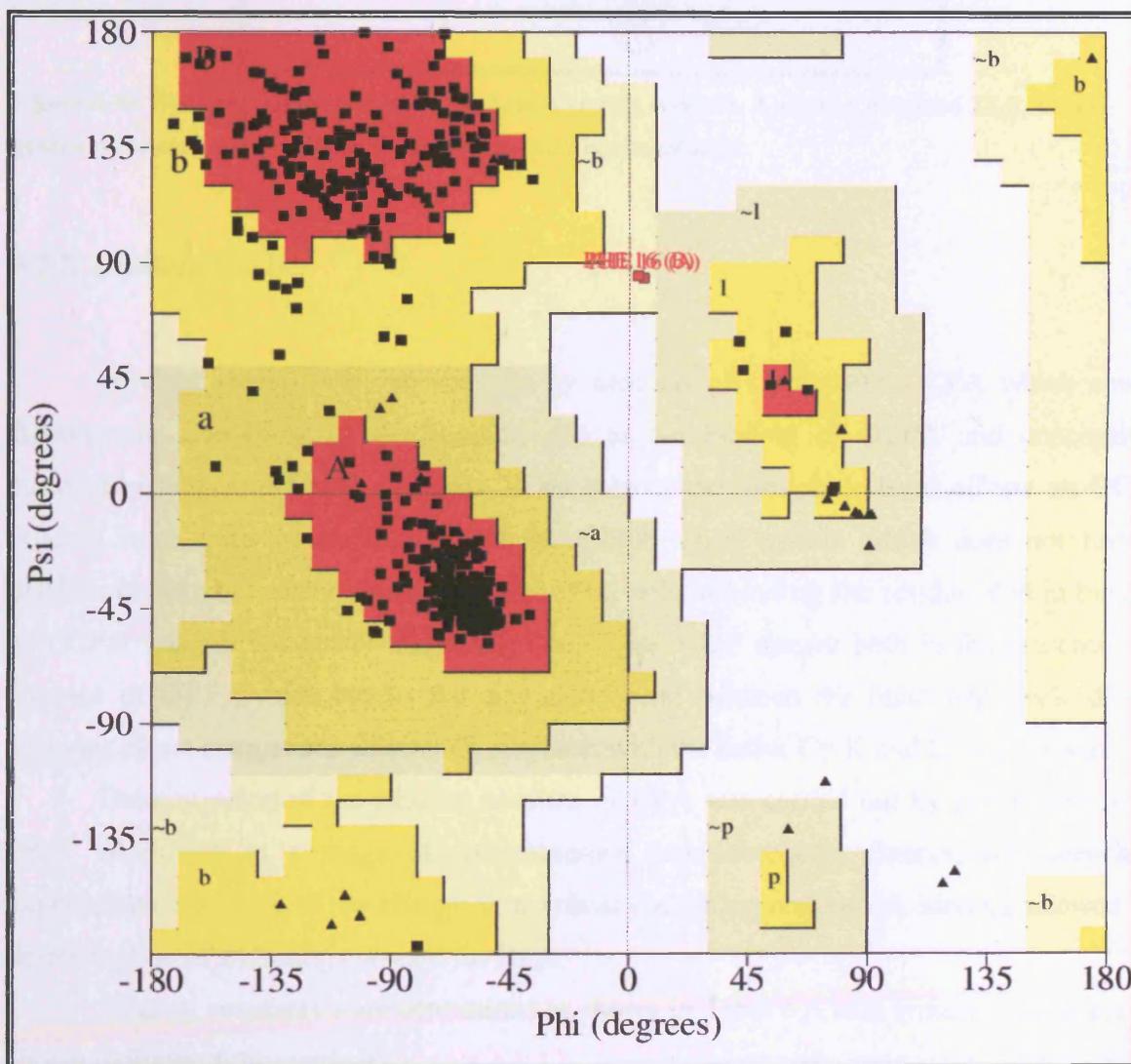
The CprK gene was cloned with an additional linker of sixteen residues and a six histidine segment. The crystal structure contained 230 out of the 232 protein residues for molecule A and all 232 for molecule B. An additional ten linker residues in molecule A and nine in molecule B are also visible. There are also two sulphate molecules and 368 water molecules interacting with the protein structure as are two molecules of OCPA. The side chains of residues E45 and K209 from molecule A and K162 and K209 from molecule B have been modeled by a single beta carbon due to poor electron density. This model gives an R-factor of 0.193 for the working data and 0.239 for the test data (free R-factor) in the resolution range  $20\text{--}2.2\text{ \AA}$  (44,535 reflections in the working data set and 2,369 in the test set; 5,483 atoms). B factor analysis of the CprK model was carried out using BAVERAGE and values are shown in Table 6.5.

The model has good stereochemistry with values for the RMSD from standard values of the bond lengths of  $0.016\text{ \AA}$  and angles of  $1.68^\circ$ , respectively. The Ramachandran plot (Figure 6.9) shows that all the non-glycine residues have conformational angles ( $\phi, \psi$ ) in permitted regions with 89% of these in most favored regions. Phenylalanine 16 in both the A and B molecule is highlighted in the generously allowed region. However, the modeled Phe

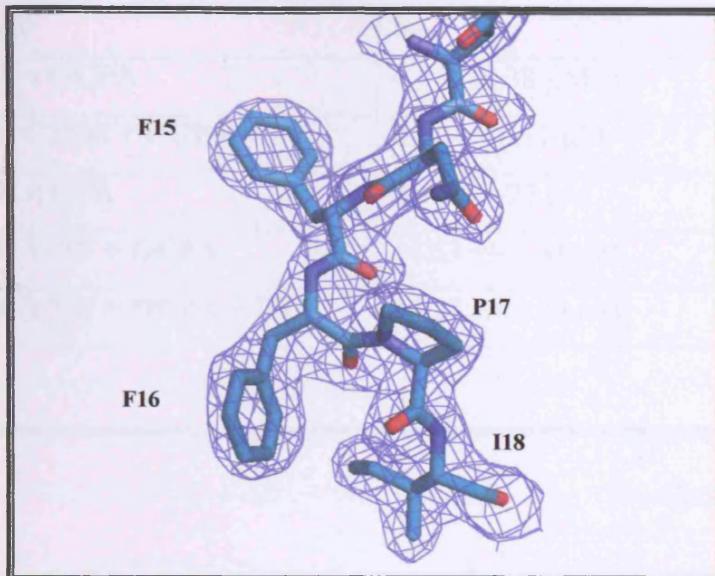
16 residues fit the  $2F_o - F_c$  electron density map well and there is no  $F_o - F_c$  density visible in this region (Figure 6.10).

**Table 6.5 B-factor analysis of CprK model**

	B factor ( $\text{\AA}$ )
Protein atoms	37.3
Sulfate ions	57.5
OCPA molecules	42.8
Water molecules	27.9



**Figure 6.9** Ramachandran plots for the model of the CprK-OCPA complex produced using PROCHECK. The coloured areas indicate, in order of decreasing intensity, most favoured region, additional allowed region, generously allowed region and disallowed region. Glycine residues are shown as triangles.



**Figure 6. 10 Stick representation of F16 and surrounding residues. A sigmaA weighted  $2F_oF_c$  map is generated around these residues shown in blue wire representation.**

#### 6.2.8. Binding studies

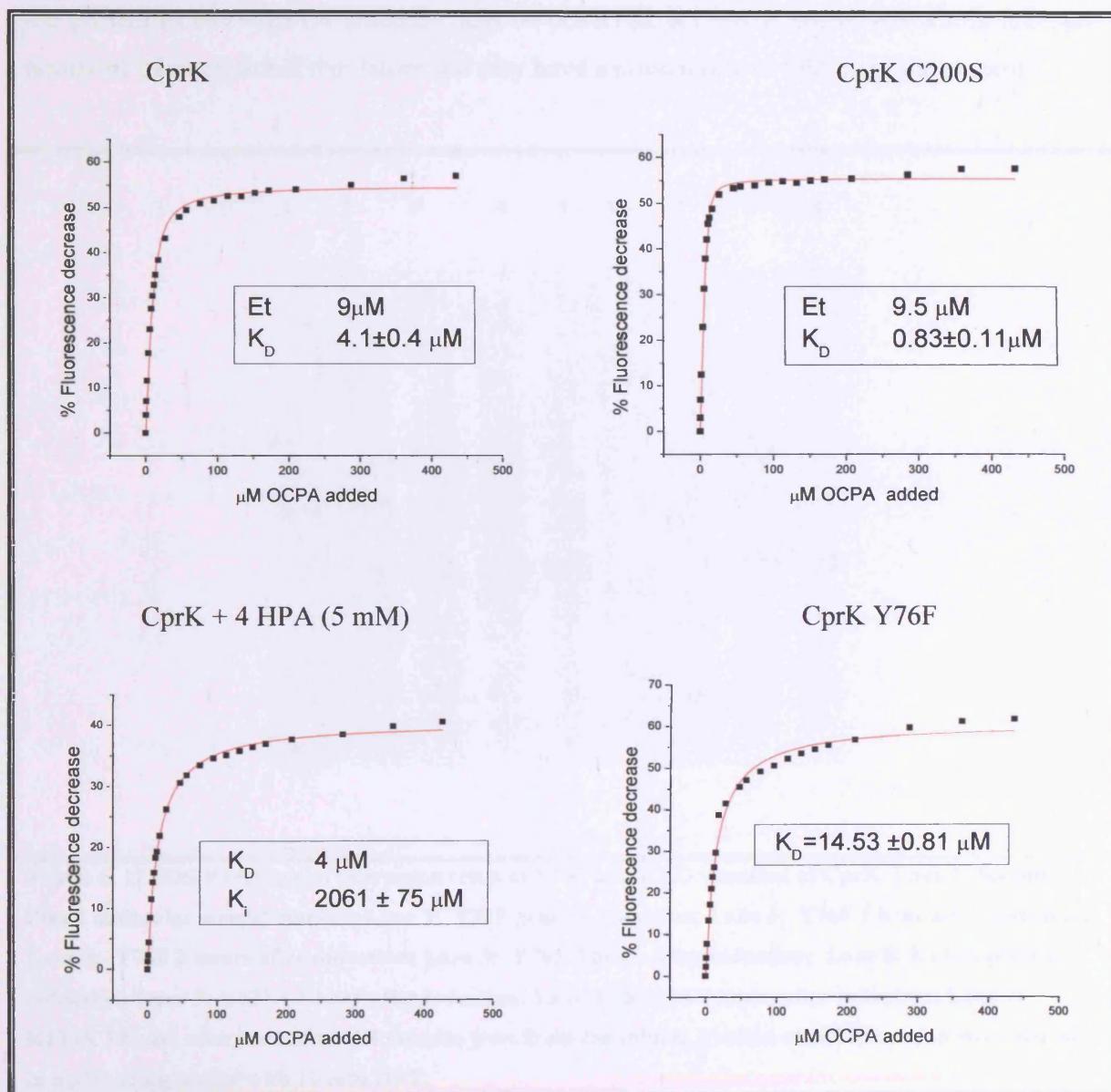
Binding studies were carried out by titration of CprK with OCPA which caused fluorescence quenching of the solution due to the binding of OCPA and concomitant structural changes of CprK. Analysis of the role of the disulphide bond effects on OCPA binding were analysed by titration of the C200S CprK mutant which does not form a disulphide linked covalent dimer. Analysis of the role of binding site residue Y76 in binding of OCPA was carried out by the generation of the Y76F mutant both in the presence and absence of DTT (which breaks the disulphide bond between the functional CprK dimer) allowing direct comparison of binding constants with the native CprK and C200S mutant.

Determination of the binding constant of HPA was carried out by pre-incubation of CprK with HPA at a range of concentrations and subsequent fluorescence quenching experiments. Analysis of the change in maximal quenching and OCPA binding allowed the determination of a binding constant for HPA.

Binding constants were determined as shown in Table 6.6 with typical binding curves shown in Figure 6.11 overleaf.

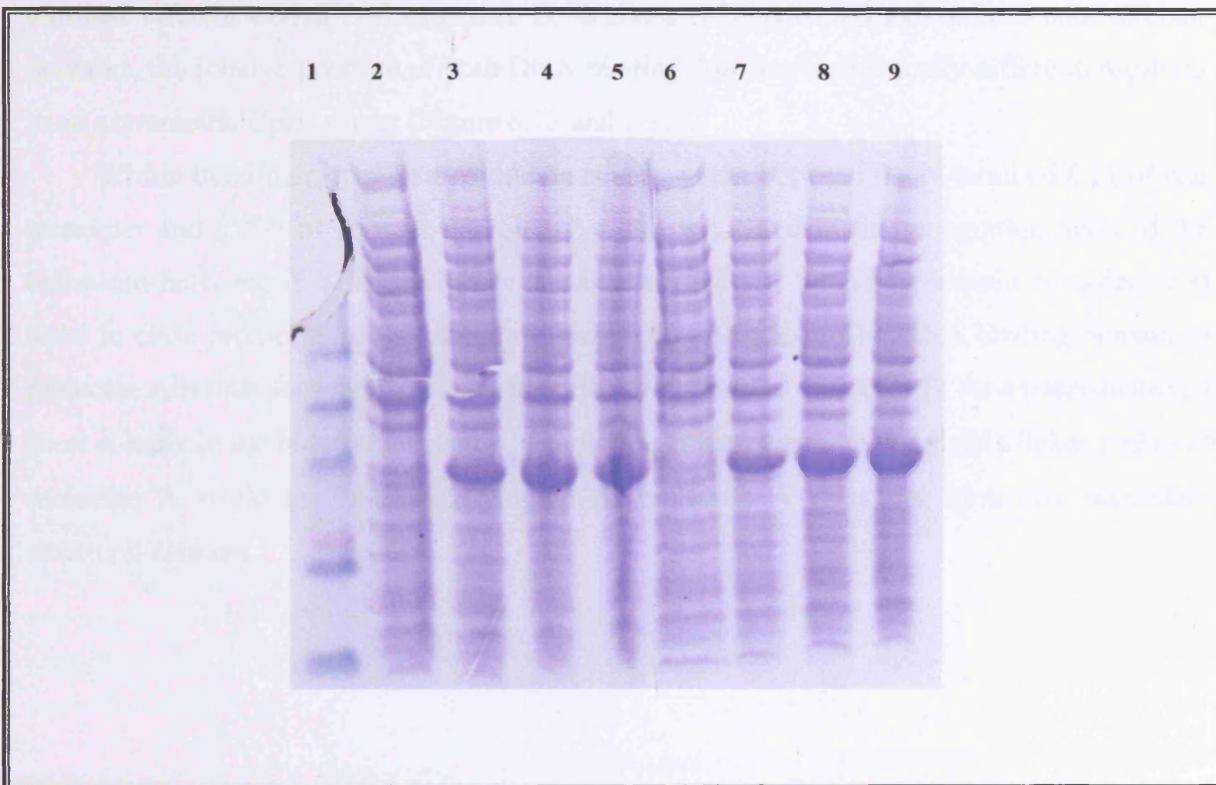
**Table 6.6 Summary of the CprK samples analysed by fluorescence and subsequent determined  $K_D$  values**

Sample	$K_D$
CprK + OCPA	4.1 +/- 0.38 $\mu\text{M}$
CprK C200S + OCPA	0.81 +/- 0.11 $\mu\text{M}$
CprK + HPA	2061 +/- 75 $\mu\text{M}$
CprK Y76F + OCPA	14.53 +/- 0.81 $\mu\text{M}$
CprK Y76F + OCPA + DTT	10.24 +/- 0.49 $\mu\text{M}$

**Figure 6. 11 Typical graphs of the fluorescence quenching binding experiments on CprK.**

### 6.2.9. Mutagenesis of active site residues

Mutagenesis was carried out as described in Section 3.6.8. Y76 and K133 were chosen for mutagenesis due to their presence in the binding site of OCPA. These mutations assess the importance of the residues in binding of OCPA. Sequencing by Lark technologies indicated that the Y76F mutation and the K133A mutation were successfully carried out. Expression and purification of the Y76F mutant proceeded as described for wild-type CprK. The K133A mutant was shown to express soluble protein (Figure 6.12, lanes 6-9). However, no binding of the protein to the Ni-NTA column could be observed. K133A is conserved within this subfamily of transcriptional regulators and may have a critical role in folding of the protein.



**Figure 6. 12 SDS-PAGE gel of expression trials of Y76F and K133A mutant of CprK. Lane 1: SeeBlue Plus 2 molecular weight marker; Lane 2: Y76F prior to induction; Lane 3: Y76F 1 hour after induction; Lane 4: Y76F 2 hours after induction; Lane 5: Y76F 3 hours after induction; Lane 6: K133A prior to induction; Lane 7: K133A 1 hour after induction; Lane 8: K133A 2 hours after induction; Lane 9: K133A 3 hours after induction. All samples were from the soluble fraction of cell lysate and were loaded in 5 µl loading buffer with 10 mM DTT.**

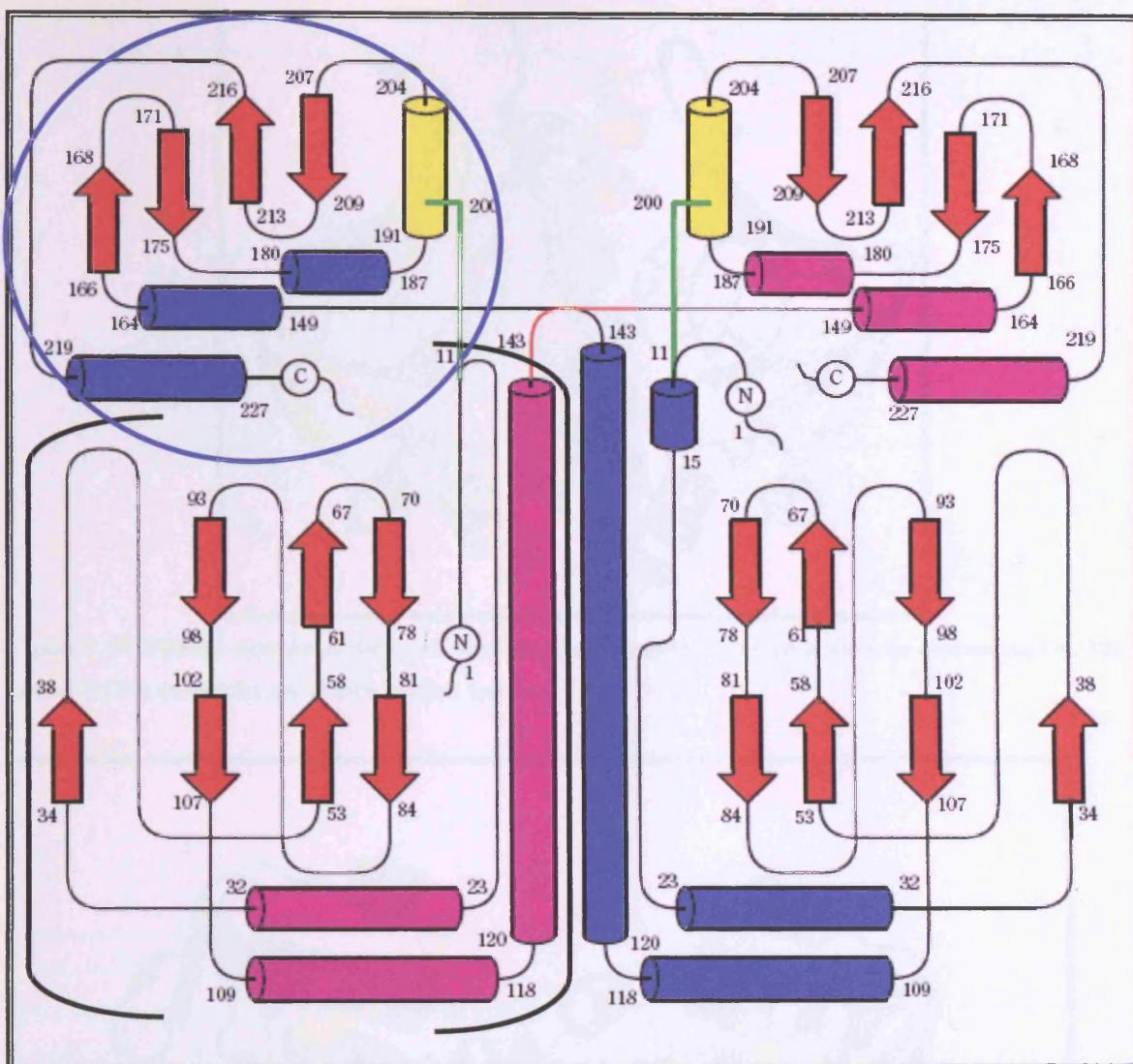
## 6.3. Discussion

### 6.3.1. Overall fold and Secondary structure

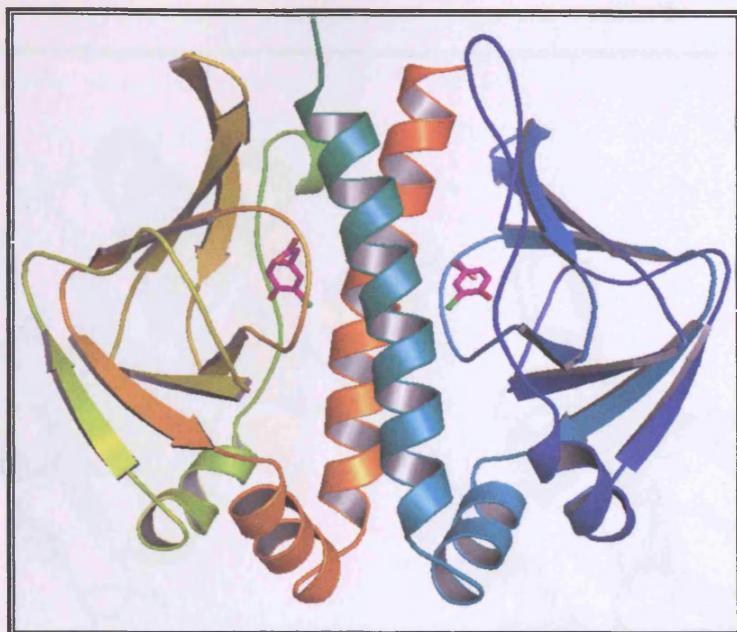
The asymmetric unit of the crystal contains a single CprK dimer with each monomer folded in two distinct domains as seen for members of the CRP-FNR family of winged HTH transcriptional regulators. These domains are the effector binding domain and the DNA binding domain. The relative position of the effector domains within the CprK dimer is similar to CRP, with the dimer interface predominantly made up by the central  $\alpha$ -helices connecting the effector domain with the DNA-binding domain (Figure 6.13 and 6.14).

However, in contrast to CRP and other crystal structures of related regulators, CprK is a domain swapped dimer, with the DNA-binding domain of monomer A interacting with the N-terminal effector domain of monomer B. While a NCS two-fold axis relates both effector domains, the relative position of both DNA-binding domains is drastically different, resulting in an asymmetric CprK dimer (Figure 6.15 and 6.16).

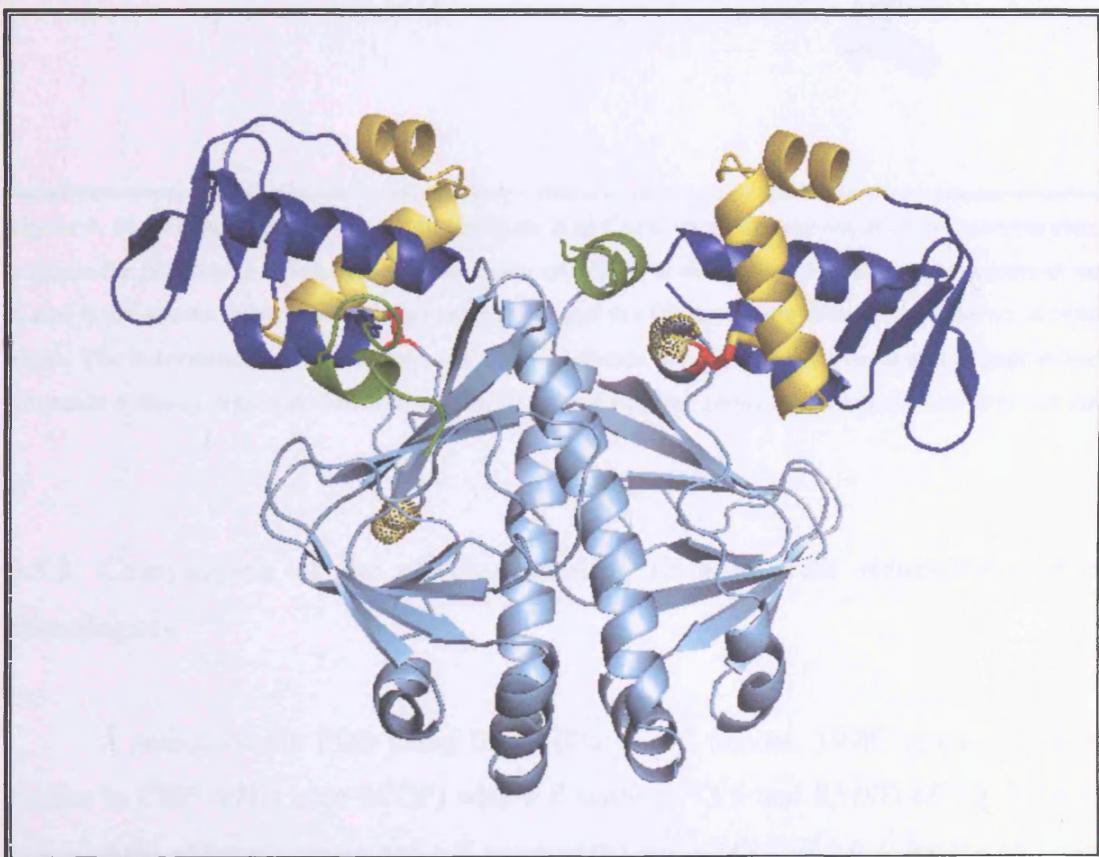
Within the dimer, two disulphide bonds are present between the N-terminal C11 of one monomer and C200 of another monomer. C200 is located in the recognition helix of the helix-turn-helix motif. This disulphide bond causes the DNA binding domain of molecule B to be in close proximity to the effector domain of molecule A. The DNA binding domain of molecule A is only loosely contacting the effector domain of molecule B. As a consequence, a short  $\alpha$ -helix in the N-terminal region of molecule B interacts with the visible linker region of molecule A while the N-terminal region of molecule A does not form any secondary structural element.



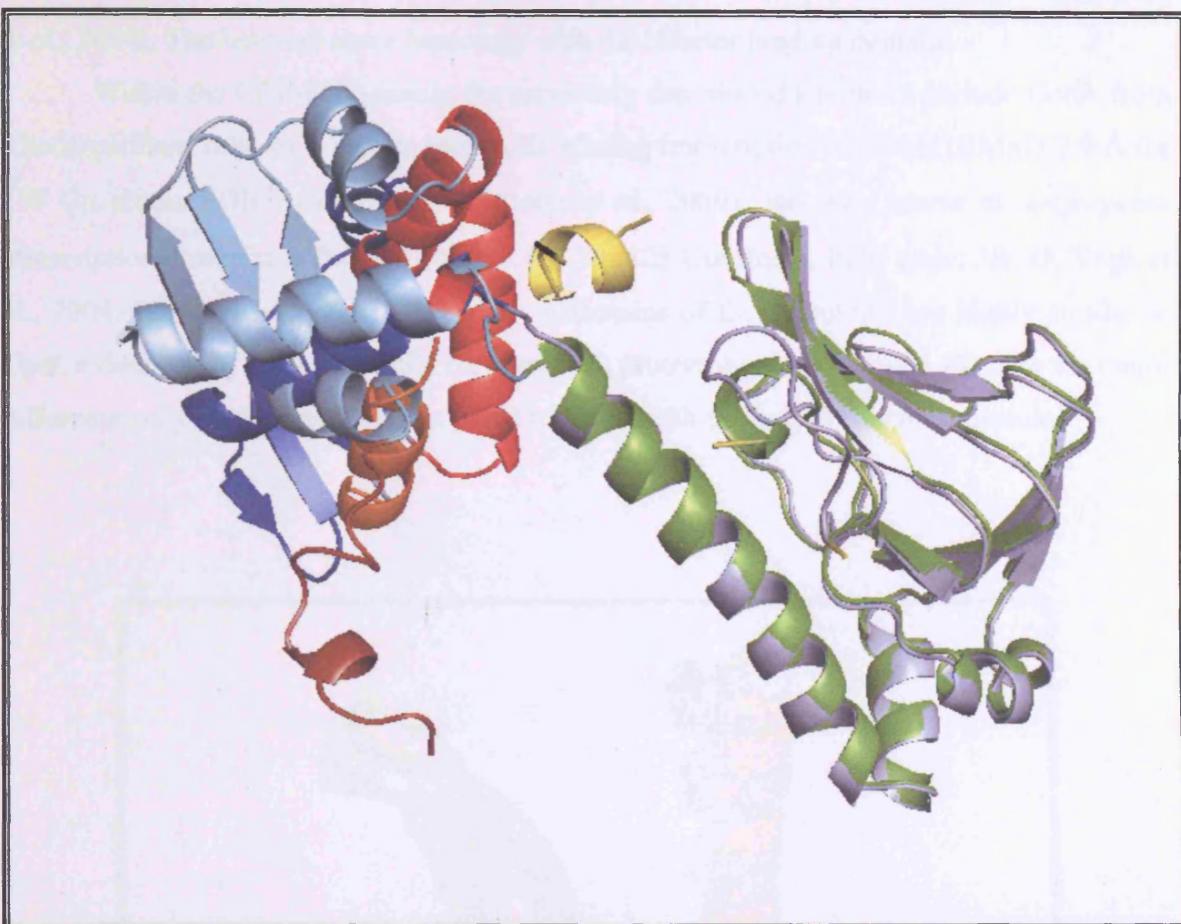
**Figure 6. 13** Topology drawing of CprK constructed using TOPDRAW (Bond, 2003). Helices are represented by magenta (molecule a) and blue (molecule b) cylinders and strands of  $\beta$ -sheets are shown with red arrows. A single effector domain is enclosed by the black brackets and a single DNA binding domain is indicated by the blue circle. The recognition helix involved in direct DNA interaction is shown as a yellow cylinder and the disulphide bond is shown by a thick green line.



**Figure 6. 14** Effector domain of the CprK transcriptional regulator shown in cartoon representation. The bound OCPA molecules are shown in stick format.



**Figure 6. 15** Structure of the CprK transcriptional regulator shown in cartoon representation. The DNA binding domain is highlighted in blue and the disulphide bond between the effector domain and the recognition helix is shown in red stick format. The N-terminus in each molecule is indicated by a yellow dotted area and the additional His-tag regions visible are shown in green. The recognition helix is shown in yellow.



**Figure 6.16** Overlay of molecule A and molecule B of CprK structure shown in cartoon representation to indicate the structural differences between each monomer in the dimer. The effector domains of molecules A and B are shown in grey and green respectively and the DNA binding domains are shown in blue and white. The N-terminal regions of molecule A and molecule B are shown in brown and orange respectively. Molecule A linker region is shown in yellow. The DNA binding recognition helix is shown in red colouring.

### 6.3.2. Comparison of the effector binding domain with structurally identified homologues

A search of the PDB using DALI (Holm and Sander, 1998) reveals CprK is most similar to CRP (PDB code 2CGP) with a Z score of 13.9 and RMSD of 3.0 Å for 124 Ca atoms of the effector domain and a Z score of 9.3 and RMSD of 1.9 Å for the 68 aligned Ca atoms of the DNA-binding domain. Other structures that are identified as structurally homologous outside of the CRP-FNR family include the effector binding domains of the guanine nucleotide exchange factor Epac2 (PDB code: 1O7F, Rehmann *et al.*, 2002), the fatty acid responsive transcription factor (FadR) from *E. coli* (PDB code: 1E2X, van Aaltan *et al.*,

2000) and also the cyclic nucleotide regulated potassium channel (PDB code: 1VP6, Clayton *et al.*, 2004). The latter all share homology with the effector binding domain.

Within the CRP-FNR family the previously determined structures include CooA from *Rhodospirillum rubrum*, a carbon monoxide sensing transcription regulator (RMSD 2.9 Å for 118 C $\alpha$  atoms PDB code: 1FT9, Lanzilotta *et al.*, 2000), and the *Listeria monocytogenes* transcriptional regulator PrfA (RMSD: 2.4 Å for 125 C $\alpha$  atoms, PDB code: 2BEO, Vega *et al.*, 2004, Eiting *et al.*, 2005). The effector domains of these proteins are highly similar to CprK evidenced by the low RMSD between each protein domain (Figure 6.17) with the major difference only in the length of a loop that interacts with the bound effector molecule.

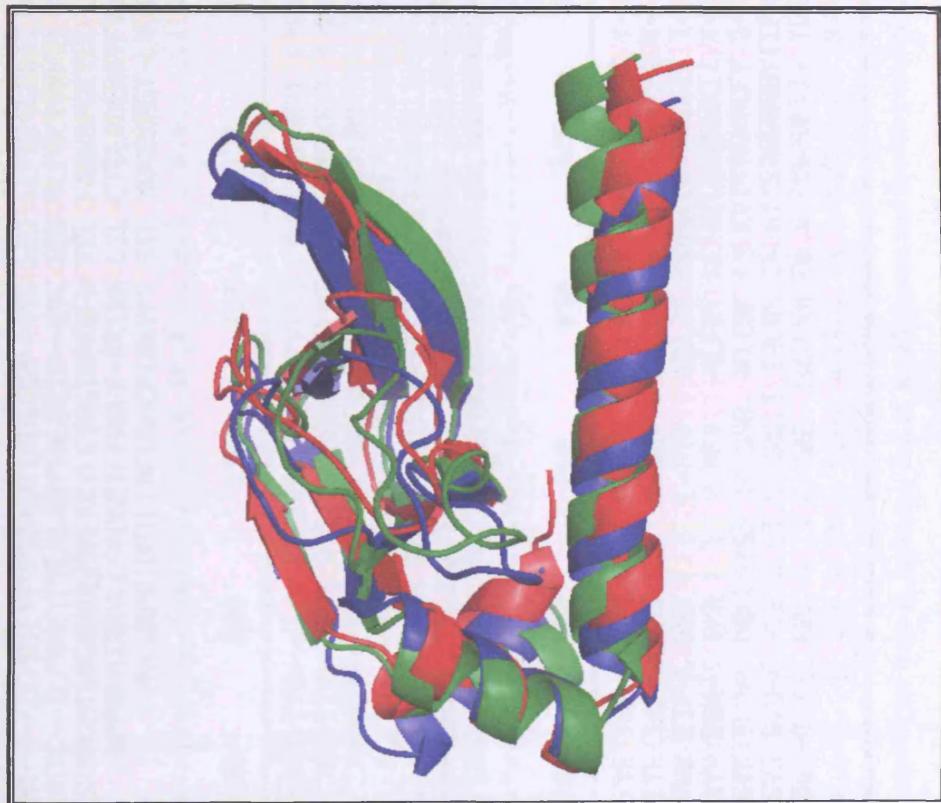
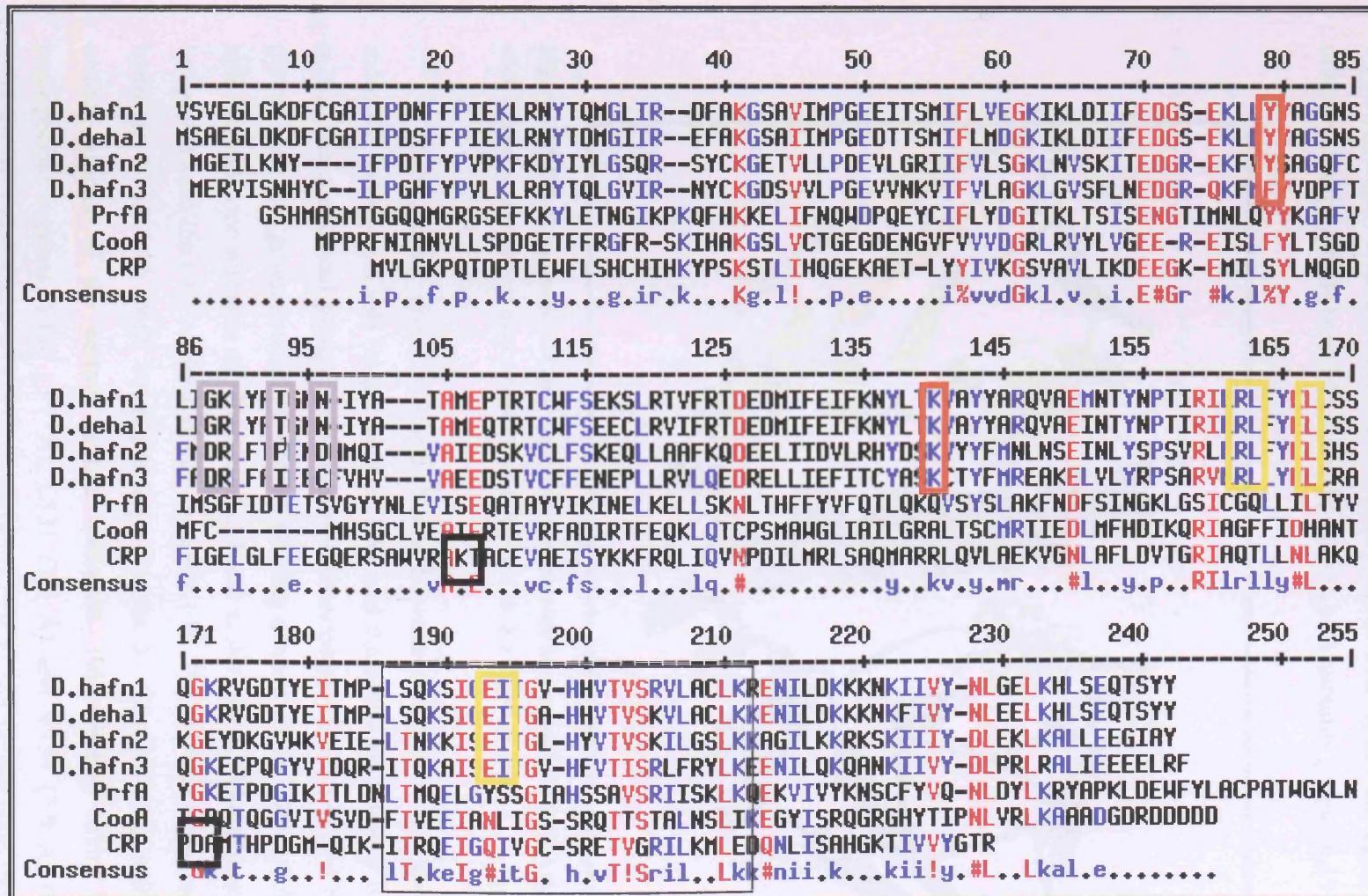


Figure 6.17 Overlay of the effector domains of CprK (green), CRP (red) and CooA (blue).

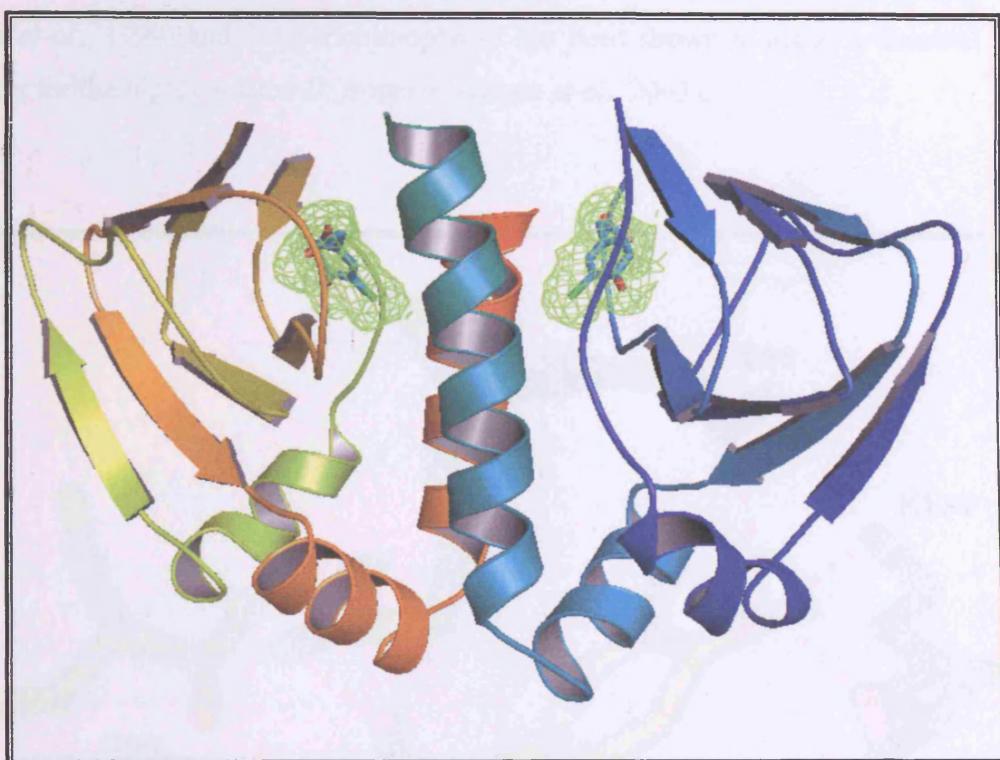
Sequence alignment of CprK with its homologues from *D.hafniense* and *D.dehalogenans* and members of the CRP-FNR family for which a structure is available indicate the high conservation within the family and specifically within the halorespiration regulators (Figure 6.18).



**Figure 6. 18 Sequence alignment of CprK homologues from *D. hafniense* and *D. dehalogenans* and homologues PrfA, CooA and CRP. Residues are colour coded based on homology (Red: >70 %; Blue: >40 %) with a homology sequence is also shown. The residues highlighted in brown and grey boxes are involved in effector binding and the residues highlighted in yellow boxes are involved in DNA domain interactions. K89 and D155 from CRP are highlighted in thick black boxes. The helix turn helix motif is indicated by the thin black box.**

### 6.3.3. OCPA binding by CprK

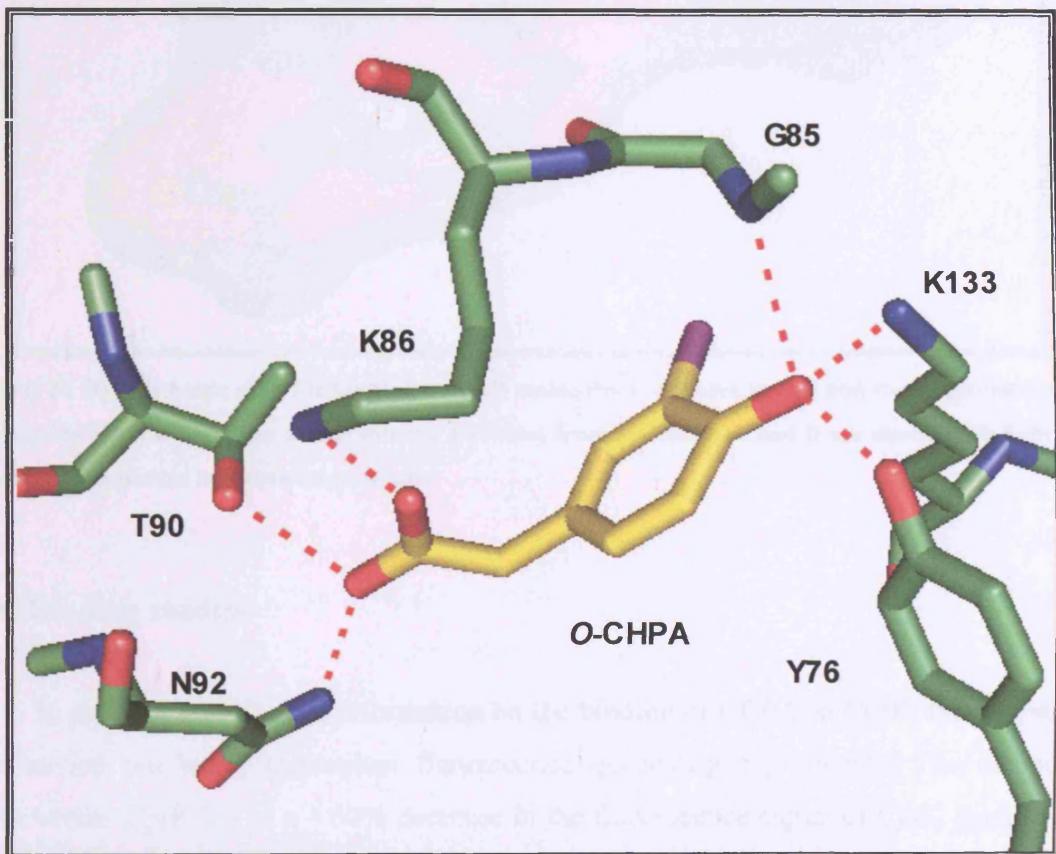
OCPA binds in the highly conserved jelly roll  $\beta$ -barrel motif of the effector domain in a position highly similar to the cAMP binding site in CRP with clear  $F_o$ - $F_c$  electron density observed for the tightly bound OCPA molecules in both subunits (Figure 6.19).



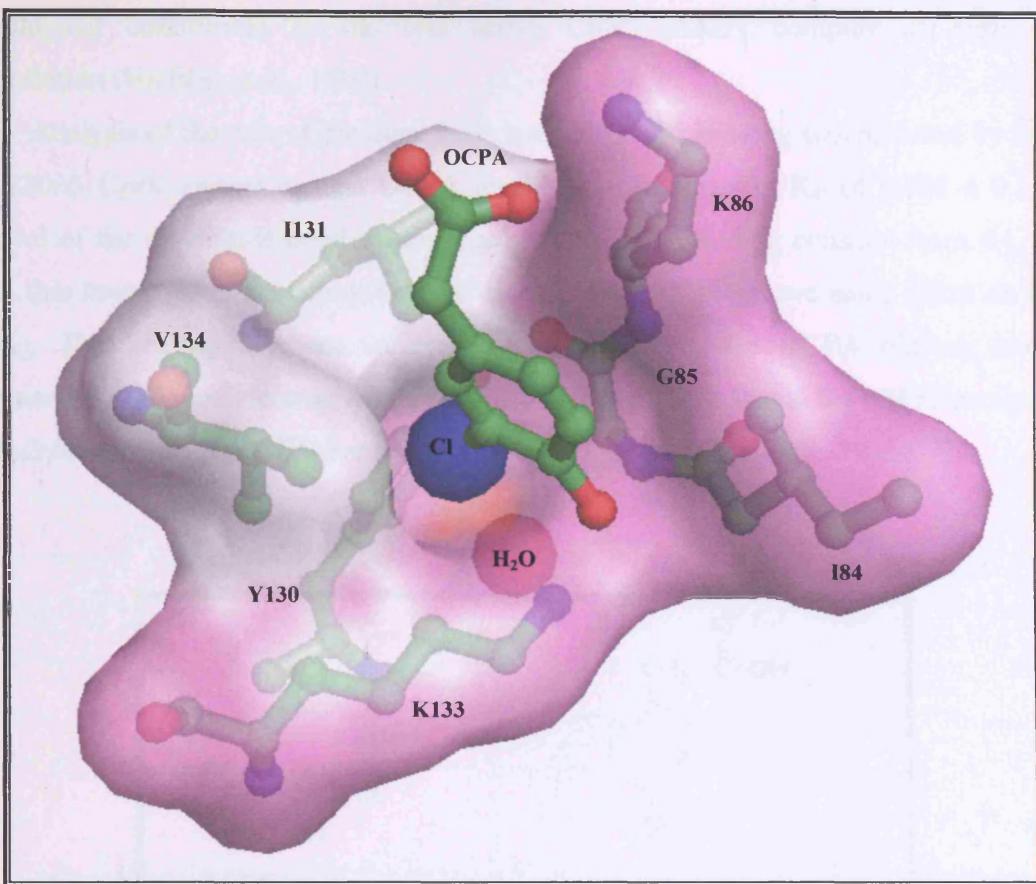
**Figure 6. 19 Effector domain of CprK represented in cartoon format with the OCPA molecules shown in stick format superimposed over an omit map calculated at  $3\sigma$  shown in green .**

There are six direct hydrogen bonds between the protein and each bound OCPA molecule including a salt bridge between K86 and the acetic acid moiety of OCPA (Figure 6.20). The acetic acid group is also hydrogen bonded to T86 (2.5 Å) and N88 (2.7 Å). The phenol hydroxyl group is within hydrogen bonding distance from Y76 (2.4 Å), K133 (2.7 Å) and the backbone nitrogen of G85 (3.0 Å). There is thus the additional possibility for a salt bridge between the OCPA phenolate group and K133. The chloride atom does not make any hydrogen bonds with CprK but is positioned in van der Waals contact with residues of the central  $\alpha$ -helix of the partner protein molecule. This binding cavity is made up of hydrophobic residues Y130 (3.7 Å), L131 (3.8 Å) and V134 (3.9 Å; from the partner molecule) in addition to a buried water molecule (3.9 Å) and several main chain atoms of G85 (3.7, 3.9, 4.0 Å) and also with the K133 amino terminal group (3.6 Å; Figure 6.21). The residues involved in the binding of the phenol moiety K133 and Y76 (one homologue has an

F at this position) are conserved within the sub-family of CprK homologues (Figure 6.18 brown boxes) while the residues involved in binding the acetic acid group N92, G85, K86 and T90 are specific to the *D. hafniense* CprK (Figure 6.18 grey boxes). These sequence differences might allow other transcription regulators to bind alternative chlorinated halogens such as 2, 4-dichlorophenol and/or 2,4,6-trichlorophenol. The terminal dehalogenase enzyme has been shown to have significant activity against these compounds in *D. dehalogenans* (van de Pas *et al.*, 1999) and 2,4,6-trichlorophenol has been shown to act as a terminal electron acceptor for the highly related *D. frappieri* (Boyer *et al.*, 2003).



**Figure 6. 20** Hydrogen bonding between OCPA and CprK with the specific residues shown in stick representation.



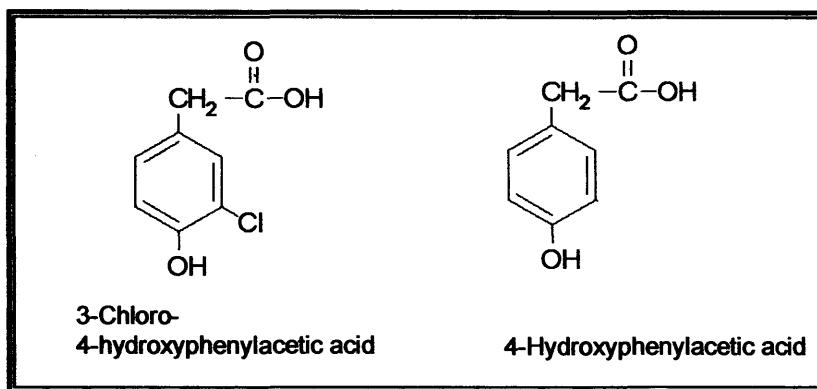
**Figure 6. 21 Hydrophobic chloride binding site.** All molecules are shown in ball and stick representation. The chloride atom is shown as a blue sphere. Residues from molecules A and B are shown with light blue and white transparent surfaces respectively.

#### 6.3.4. Binding studies

In order to gain further information on the binding of OCPA to CprK, binding studies were carried out using tryptophan fluorescence quenching experiments. The titration of OCPA versus CprK led to a ~ 60% decrease in the fluorescence signal of CprK leading to an apparent  $K_D$  of  $4.1 \pm 0.38 \mu\text{M}$  for OCPA (using a single binding site model; Fig 6.11). A single tryptophan (W106) is present in CprK and is situated at the bottom of the  $\beta$ -barrel effector binding domain thus not in close contact with the OCPA binding site. Therefore quenching of the tryptophan fluorescence is likely due to an allosteric structural change rather than the measurement of direct binding of OCPA. Binding studies carried out on the *D. dehalogenans* CprK (Pop *et al.*, 2004) using isothermal calorimetry measurements revealed a binding stoichiometry of  $0.55 \pm 0.01$  OCPA per protein dimer indicating only one site within the dimer is occupied for conformational changes to occur leading to DNA binding. This is also highly similar to binding studies carried out on CRP that is seen to undergo sequential conformational changes from the unbound form to the active CRP: cAMP<sub>1</sub> complex (under

physiological conditions) to the less active CRP: cAMP<sub>2</sub> complex at high cAMP concentration (Hudson *et al.*, 1990).

Analysis of the role of the disulphide bond in OCPA binding was assessed by titration of a C200S CprK mutant against OCPA resulting in an apparent  $K_D$  of  $0.808 \pm 0.11 \mu\text{M}$ . Removal of the disulphide bond causes a reduction of the binding constant from  $4.1 \pm 0.38 \mu\text{M}$  to this lower value indicating that the disulphide bond does have some effect on OCPA binding. The binding constant values are comparable to the OCPA binding constants determined by ITC and electrophoretic mobility assays of  $3.5 \mu\text{M}$  and  $0.4 \mu\text{M}$  respectively for the *D. dehalogenans* CprK (Pop *et al.*, 2004).



**Figure 6. 22 Chemical formula of 3-chloro-4hydroxyphenylacetic acid and 4-hydroxyphenylacetic acid.**

The product of the dehalogenation reaction phenyl acetic acid (PA) was also assessed for binding to CprK, but even at mM concentrations negligible fluorescence quenching was observed. However, binding of PA to CprK does occur as seen through the inhibition of fluorescence quenching by OCPA titration (Figure 6.11c). It was thus possible to determine an observed  $K_i$  for PA binding of  $2.06 \pm 0.07 \text{ mM}$  for the wild type oxidized CprK. Despite the relatively small difference in structure (Figure 6.22) and the apparent lack of protein to chloride substituent interaction (Figure 6.20-21), oxidized CprK thus achieves a remarkable ~500-fold specificity for OCPA over PA. In addition, PA binding does not induce any allosteric structural change in CprK and this must be part of the efficient induction and regulation of CprAB transcription and ultimately halorespiration within *D. hafniense*.

While the PA-bound form of CprK does not undergo the allosteric transition, the binding of the PA within the effector domain binding pocket may well be similar to the obtained OCPA CprK crystal structure. It thus seems unlikely that the additional binding energy of the chloride atom present only in OCPA can fully account for the observed 500-fold preference for OCPA.

While the pKa for the phenol group of PA is approximately 10, this value is lowered to 8.4 in OCPA due to the chloride substituent at the ortho-position (<http://www.speclab.com/compound/c95578.htm>). It is possible that CprK can distinguish between OCPA and PA not only on the additional steric bulk provided by the chloride atom, but also by the ability of the compound to form the phenolate ion and consequently interact more strongly with both Y76 and K133, ultimately leading to the allosteric structural changes observed in CprK. This system of regulation has been reported previously where substrate deprotonation at the phenolic group of p-hydroxybenzoate bound to p-hydroxybenzoate hydroxylase led to flavin movement and initiation of the catalytic cycle (Palfey *et al.*, 1999).

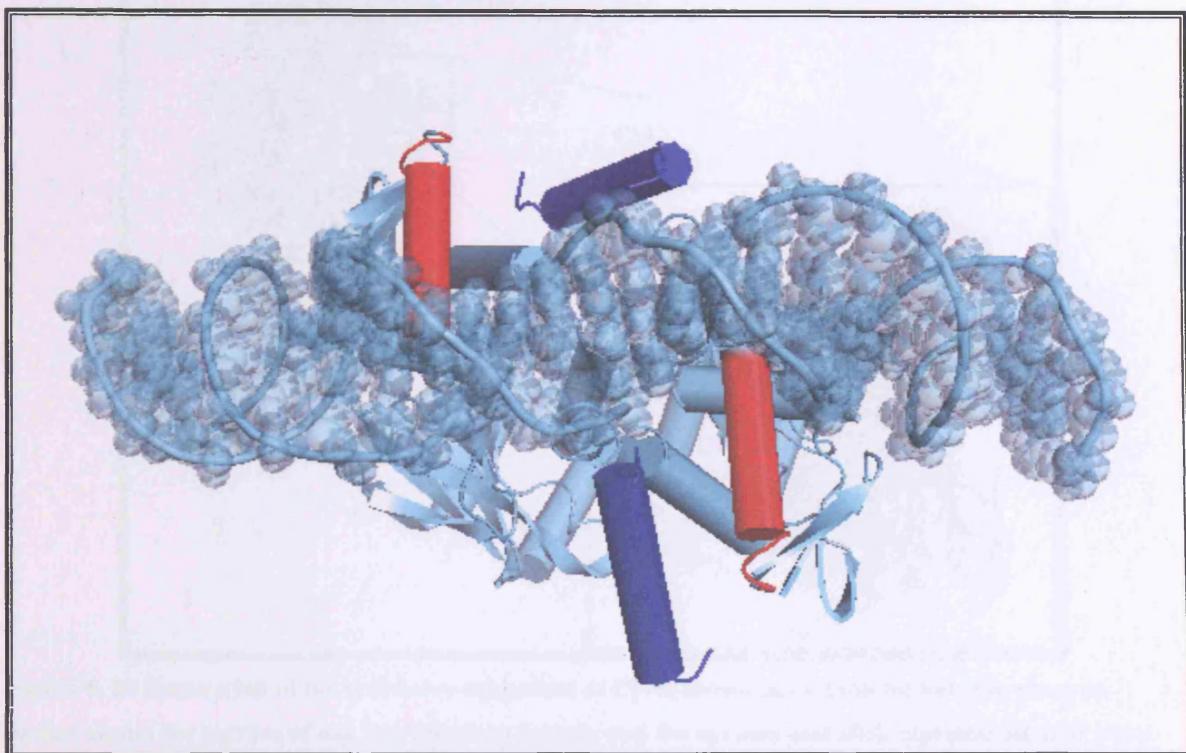
The role of Y76 in OCPA binding was assessed by generation of the Y76F mutation and subsequent binding studies. This mutation resulted in lowering the binding affinity of CprK for OCPA giving a  $K_D$  of  $14.53 \pm 0.81 \mu\text{M}$ . OCPA titration in the presence of 1 mM DTT to break the disulphide bond gave a binding constant of  $10.24 \pm 0.49 \mu\text{M}$ . The 10-fold change observed in binding strength can be explained by the lack of a single hydrogen bond in the Y76F mutant OCPA complex, it is however clear that Y76 is not essential to OCPA sensing.

All of the CprK homologues identified to date reveal that K133 is conserved and this residue may be pivotal to “pKa interrogation” of the bound molecule. Within the other CprK homologues that may bind 2,4,6-trichlorophenol (pKa of 6.2), 2,4-dichlorophenol (pKa of 7.8) and 4-chlorophenol (pKa of 9.3), this mechanism of pKa interrogation is also a feasible mode of action.

### 6.3.5. CprK DNA binding domain

In the previously determined CRP structures, both of the DNA binding helices are positioned roughly parallel to one another along the surface of the DNA binding domain. Interactions are observed between the recognition helices and the DNA specifically between R180, E181, T182 and R185 of CRP. Analysis of the sequence alignment in Figure 6.18 shows that CprK contains residues HHT and R at equivalent positions indicating that there is differences in DNA binding specificity but the recognition helix is still made up of charged or hydrophobic residues as seen in CRP. CprK binds to a DNA sequence of TTAATacgcACTAA in contrast to the CRP specific DNA binding sequence of TGTGAtctagaTCACA (DNA bases in lower case have no interaction with the transcriptional regulator).

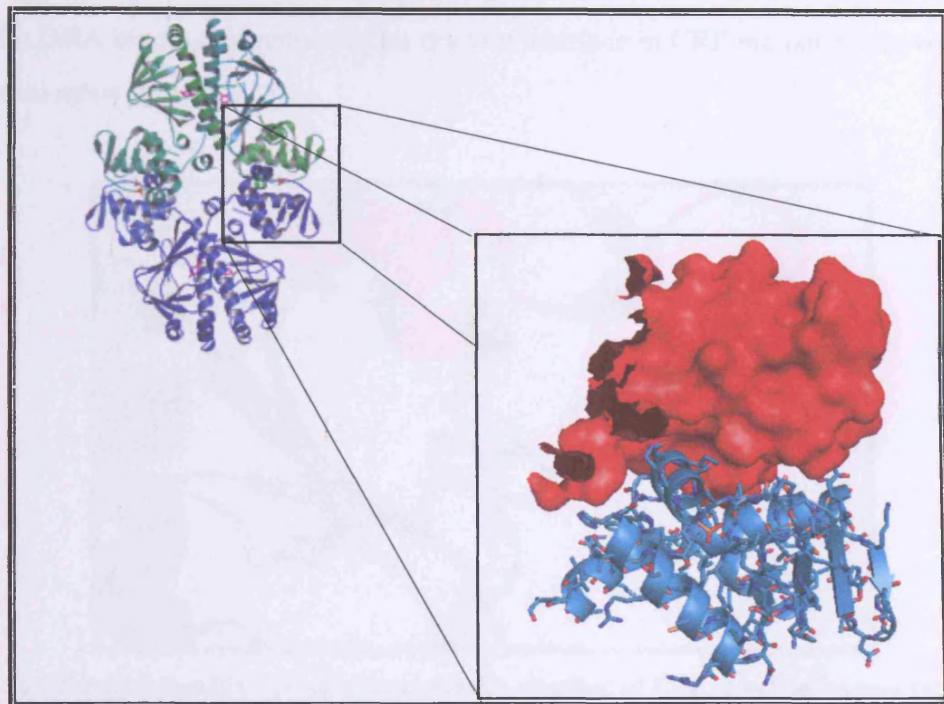
A least squares structural alignment of the DNA bound CRP structure and CprK shows that the CprK DNA-binding helices are situated on opposite sides of the protein. This conformation is clearly not compatible with tight binding to a palindromic DNA sequence (Figure 6.23).



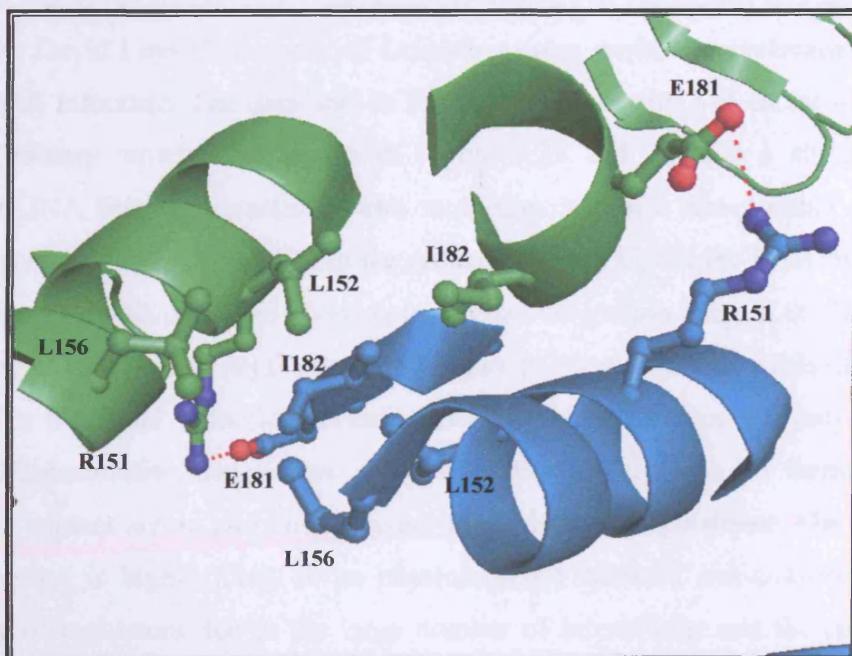
**Figure 6.23** CRP structure (grey) in complex with DNA (transparent spheres and cartoon representation) with the DNA binding recognition helix shown in solid red cylindrical representation (PDB code: 1CGP) viewed from above the DNA. CprK is structurally aligned with CRP and the recognition helices of the HTH motif of CprK are shown in solid blue cylindrical format.

Reorientation of both domains to a position similar to CRP requires the C11-C200 disulphide bond to be broken. DNA binding studies on the *D. dehalogenans* CprK have demonstrated that the formation of a disulphide bond under aerobic conditions abolishes CprK DNA-binding (Pop *et al.*, 2004) and this phenomenon has been implicated in the possible redox control of the halorespiration system. Analysis of the role that crystallographic packing might play in the misalignment of the DNA binding helices was also assessed. However, upon generation of symmetry neighbours, a very close association between the DNA-binding domains of both molecules of the dimer and its symmetry neighbour was observed (Figure 6.24). The interface made between the two DNA binding domains is reasonably large ( $422 \text{ \AA}^2$ ). It is made up of salt bridges between E181 and R151 from both molecules with these residues also conserved in other CprK homologues. These salt bridges are situated on the periphery of the domains while there are further hydrophobic interactions

between I182 and L152 and L156 residues at the centre of the molecular interface (Figure 6.25). Analysis of the surface complementarity using SC (CCP4, 1994) gives a value of 0.67 (a value of ~0.7 is given for physiologically relevant contacts, Lawrence and Colman, 1993).



**Figure 6. 24 Generation of the symmetry equivalent of CprK shown in cartoon format. The close up section shows the surface of one DNA binding domain and the cartoon and stick representation of the other DNA binding domain that interacts with high complementarity.**



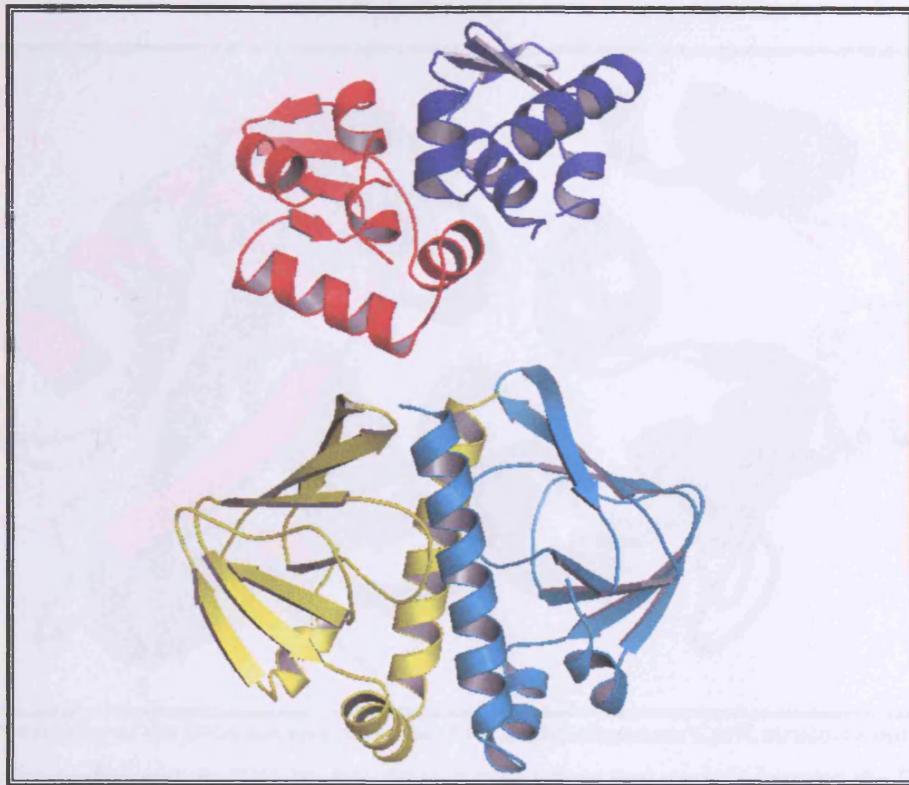
**Figure 6. 25 Salt bridges and hydrophobic residues between interacting symmetry related DNA binding domains. The DNA binding domains are shown in cartoon format with the interacting residues shown in ball and stick format.**

In CRP, this region of the DNA binding domain interacts closely with the effector domain with salt bridges between K89 and D155 observed at this interface (Figure 6.26). In contrast, little contact is made between the DNA-binding and effector domains of CprK when placed in positions similar to the corresponding CRP domains. Furthermore, the salt bridges seen at the DNA binding domain effector domain interface in CRP are not conserved in CprK or its homologues (Figure 6.18).



**Figure 6. 26 Effector (blue)-DNA binding (red) domain interface of CRP shown in cartoon representation with the K89: D155 salt bridge shown in ball and stick format.**

X-ray diffraction data collected from a reduced, non-OCPA bound *D. dehalogenans* CprK crystal was given by Dr. S.W. Ragsdale (University of Nebraska) and the structure was solved by Dr. David Leys (University of Leicester) using molecular replacement and the *D. hafniense* CprK structure. The data was to 3.5 Å resolution with a B factor of 54.9 Å. The present preliminary structure is shown in Figure 6.27 and exhibits a similar interaction between the DNA binding domains of two molecules within a dimer rather than within a tetramer (Figure 6.28, RMSD 0.882). In the reduced form of CprK, the DNA binding domain is not constrained by the disulphide bond between the recognition helix of the HTH motif and the N-terminus of the protein (as illustrated in Figure 6.29) and can make this interface within the dimer. The S-S bond formation probably locks a tetrameric form that only occurs under high protein concentration that is seen in the crystallisation trials and is formed by making DNA domain contact across two dimers rather than within a single dimer. The DNA binding domains interface is highly likely to be physiologically relevant and conserved within the CprK family of regulators due to the large number of interactions and the conservation of interacting residues within the CprK homologues.



**Figure 6. 27 Structure of the reduced non-OCPA bound structure of CprK from *D. dehalogenans* shown in cartoon representation. The effector domains are coloured yellow and cyan and the DNA binding domains are shown in blue and red.**



Figure 6.28 Overlay of the DNA binding domains of the *D. dehalogenans* CprK structure shown in red cartoon representation and the DNA binding domain interactions that are seen between the *D. hafniense* CprK and a symmetry neighbour shown in blue and light blue cartoon format respectively.

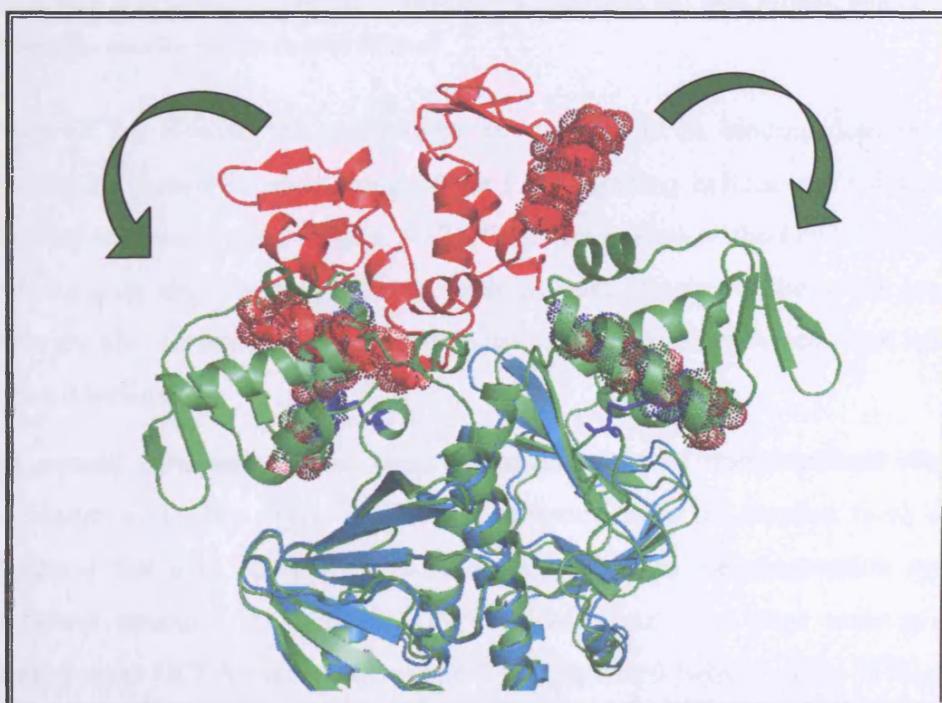


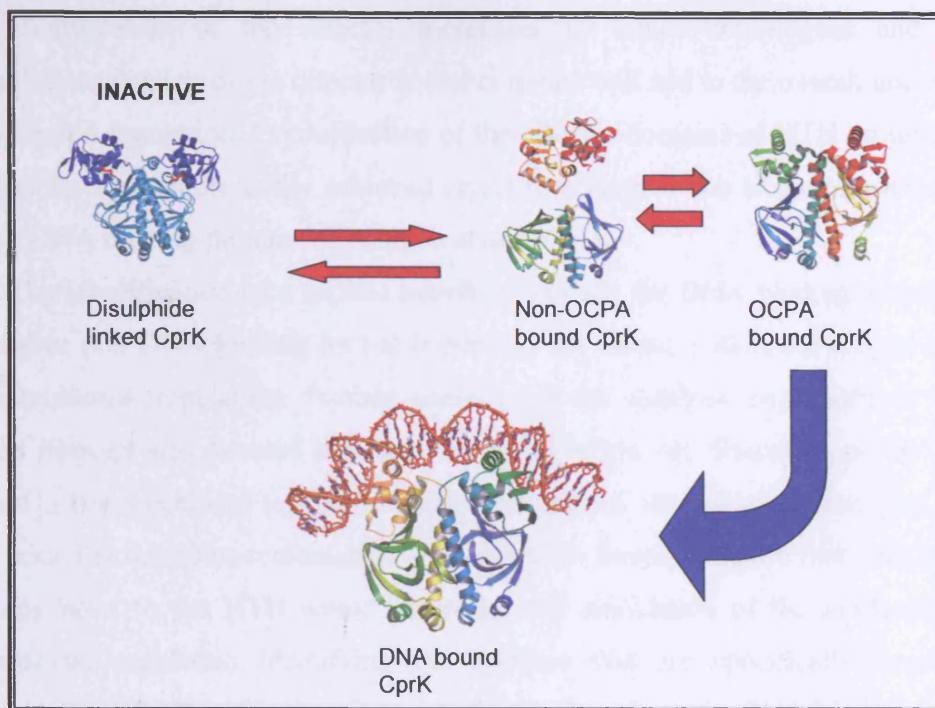
Figure 6.29 Overlay of the oxidised and reduced structures of CprK. The reduced structure is shown in green cartoon format with the disulphide bond shown in blue stick format with the recognition helix outlined by dot surface representation. The reduced non OCPA bound effector domains are coloured grey and the DNA binding domains are coloured red with the DNA binding recognition helices indicated by dot surface representation. The large structural restraints of the DNA binding domains due to the disulphide bonds are indicated by the green arrows.



**Figure 6. 30 DNA bound model of CprK generated by superimposition on CRP (PDB code: 1CGP). CprK is shown in cartoon representation and the DNA is shown in cartoon and stick format. Bound OCPA and sulphate molecules are also shown in stick format.**

Upon OCPA binding the interactions between the DNA binding domains are likely broken leading to favorable positioning of the DNA binding helices and subsequent DNA binding leading to transcription (Figure 6.30). Superimposition of the DNA binding domains on CRP shows good alignment with the recognition helix aligning in the major groove of the DNA. There are also no structural clashes with the effector domains which align well with the CRP effector domains.

An overall schematic of the proposed mechanism of transcriptional regulation is shown in Figure 6.31. The crystallized CprK structure is in an inactive form due to the disulphide bond that may be relevant to redox control of the dehalogenation system. The OCPA unbound structure is in a non DNA binding form that must undergo structural reorganization upon OCPA binding to enable the recognition helices of the HTH domain to align allowing subsequent DNA binding and transcription which ultimately leads to halorespiration of the organism.



**Figure 6. 31** Proposed schematic of CprK regulation.

#### 6.4. Conclusions and Future Work

The structure of CprK in complex with OCPA allows identification of specific residues important in the regulation of transcription and effector binding. This provides a structural framework for study of the *D. hafniense* CprK and homologous regulators in other halorespiring bacteria. The binding studies that were carried out using fluorescence quenching indicate strong differentiation between substrate and product of the halorespiration reaction by the transcriptional regulator. Further assessment of the importance of individual residues in OCPA binding through site directed mutagenesis must be carried out in order to gain a complete understanding of the binding mechanism. Assessment of the binding of OCPA over a pH gradient using more amenable techniques such as ITC in order to ascertain the role of protonation in the binding mechanism will provide insights into the possible “pKa interrogation” mechanism proposed.

Improved crystal size and quality of the non-OCPA complexed CprK would lead to better quality data that might allow structure determination by molecular replacement. This structure could then be compared and contrasted to the OCPA bound CprK structure which would allow the identification of specific structural changes that take place in the effector domain upon effector binding.

Identification of the effector molecules for other homologues and subsequent structural identification of the effector bound complex will add to the overall understanding of halorespiration regulation. Crystallisation of the effector domains of HTH proteins solely has been identified as a more easily achieved crystallisation goal due to the removal of the more dynamic DNA binding domains (Spraggan *et al.*, 2004).

The identification of a surface interface between the DNA binding domains of CprK in a putative non DNA binding format is perhaps not unique within the winged HTH family of transcriptional regulators. Further analysis of the residues implicated in the surface interface through site directed mutagenesis would allow the dissection of the role of this interface in transcriptional regulation within this family. Higher resolution data of the non-OCPA bound structure in combination with an OCPA bound structure that does not contain a disulphide bond to the HTH would allow the full elucidation of the mechanism of DNA transcriptional regulation identifying the residues that are specifically involved in the communication of OCPA binding from the effector domain to the DNA binding domains. Co-crystallisation of OCPA bound CprK with the specific DNA sequence involved in transcription would give a complete picture of the regulation mechanism and indicate the specific residues that are critical for DNA binding within the recognition helix of the DNA binding domain.

## **BIBLIOGRAPHY**

Adams, P. D., Grosse-Kunstleve, R. W., Brunger, A. T., (2003) Computational aspects of high-throughput crystallographic macromolecular structure determination. *Meth. Biochem. Anal.* **44**:75-87.

Ammon, H. L., Weber, I. T., Wlodawer, A., Harrison, R. W., Gilliland, G. L., Murphy, K. C., Sjolin, L., Roberts, J., (1988) Preliminary crystal structure of *Acinetobacter glutaminasificans* glutaminase-asparaginase. *J Biol Chem* **263**:150-156.

Aravind, L., Anantharaman, V., Balaji, S., Babu, M. M., Iyer, L. M., (2005) The many faces of the helix-turn-helix domain: transcription regulation and beyond. *FEMS Microbiol Rev* **29**:231-262.

Baker, N., Sept, D., Joseph, S., Holst, M., McCammon, J.A., (2001) Electrostatics of nanosystems: Application to microtubules and the ribosome. *Proc. Natl. Acad. Sci.* **98**:10037-10041.

Bashford, D., Karplus, M., (1991) Multiple-site titration curves of proteins – an analysis of exact and approximate methods for their calculation. *J. Phys. Chem.* **95**:9556-9561.

Ben-Naim, A., Wilf, J., Yaacobi, M., (1973) Hydrophobic interaction in light and heavy water. *J. Phys. Chem.*, **77**:95-102

Benoff, B., Yang, H., Lawson, C. L., Parkinson, G., Liu, J., Blatter, E., Ebright, Y. W., Berman, H. M., Ebright, R. H., (2002) Structural basis of transcription activation: the CAP-alpha CTD-DNA complex. *Science* **297**:1562-1566.

Benveniste, I., Lesot, A., Hasenfratz, M. P., Kochs, G., Durst, F., (1991) Multiple forms of NADPH-cytochrome P450 reductase in higher plants. *Biochem Biophys Res Commun* **177**:105-112.

Bernal, J.D., Crowfoot, D., (1934) X-ray photographs of crystalline pepsin. *Nature* **794**: 133-134.

Blanck, J., Rein, H., Sommer, M., Ristau, O., Smettan, G., Ruckpaul, K., (1983) Correlations between spin equilibrium shift, reduction rate, and N-demethylation activity in liver microsomal cytochrome P-450 and a series of benzphetamine analogues as substrates. *Biochem Pharmacol* **32**:1683-1688.

Blundell, T.L., Johnson, L.N., (1976). Protein Crystallography. Academic Press, London.

Boddupalli, S. S., Pramanik, B. C., Slaughter, C. A., Estabrook, R. W., Peterson, J. A., (1992) Fatty acid monooxygenation by P450BM-3: product identification and proposed mechanisms for the sequential hydroxylation reactions. *Arch Biochem Biophys* **292**:20-28.

Bond, C. S., (2003) TopDraw: a sketchpad for protein structure topology cartoons. *Bioinformatics* **19**:311-312.

Bossi, R. T., Aliverti, A., Raimondi, D., Fischer, F., Zanetti, G., Ferrari, D., Tahallah, N., Maier, C. S., Heck, A. J., Rizzi, M., Mattevi, A., (2002) A covalent modification of NADP<sup>+</sup> revealed by the atomic resolution structure of FprA, a *Mycobacterium tuberculosis*

oxidoreductase. *Biochemistry* **41**:8807-8818.

Bouwer, E. J., McCarty, P. L., (1983) Transformations of halogenated organic compounds under denitrification conditions. *Appl Environ Microbiol.* **45**:1295-129.

Boyer, A., Page-BeLanger, R., Saucier, M., Villemur, R., Lepine, F., Juteau, P., Beaudet, R., (2003) Purification, cloning and sequencing of an enzyme mediating the reductive dechlorination of 2,4,6-trichlorophenol from *Desulfitobacterium frappieri* PCP-1. *Biochem J* **373**:297-303.

Bracco, L., Kotlarz, D., Kolb, A., Diekmann, S., Buc, H., (1989) Synthetic curved DNA sequences can act as transcriptional activators in *Escherichia coli*. *EMBO J* **8**:4289-4296.

Bragg, W. H., Bragg, W. L., (1913) The reflection of X-rays in crystals. *Proc. R. Soc. London Ser. A* **88**:428-438.

Brennan, R. G., Matthews, B. W., (1989) The helix-turn-helix DNA binding motif. *J Biol Chem* **264**:1903-1906.

Bricogne, G., (1974) Geometric sources of redundancy in intensity data and their use for phase determination. *Acta Cryst A* **30**:395-405.

Bricogne, G., (1997) Bayesian statistical viewpoint on structure determination: basic concepts and examples. In Carter, C.W. and Sweet, R.M. (ed.) *Methods in Enzymology* vol 276, Academic Press, Orlando, Fl: pp. 361-423.

Brown, J. F., Bedard, M. J., Brennan, M. J., Carnahan, J. C., Feng, H., Wagner, R. E., (1987) Polychlorinated biphenyl dechlorination in aquatic sediments. *Science* **236**:709-712.

Bruice, T. C., Pandit, U. K., (1960) Intramolecular Models Depicting the Kinetic Importance of "Fit" in Enzymatic Catalysis. *Proc Natl Acad Sci USA* **46**:402-404

Brünger, A.T., (1992). The free *R* value: a novel statistical quantity for assessing the accuracy of crystal structures. *Nature* **355**:472-475.

Bruns, C. M., Karplus, P. A., (1995) Refined crystal structure of spinach ferredoxin reductase at 1.7 Å resolution: oxidized, reduced and 2'-phospho-5'-AMP bound states. *J Mol Biol* **247**:125-145.

Bukowski, M. R., Koehntop, K. D., Stubna, A., Bominaar, E. L., Halfen, J. A., Munck, E., Nam, W., Que, L., Jr., (2005) A thiolate-ligated nonheme oxoiron(IV) complex relevant to cytochrome P450. *Science* **310**:1000-1002.

Chang, Y. T., Loew, G. H., (1999) Homology modeling and substrate binding study of human CYP4A11 enzyme. *Proteins* **34**:403-415.

Chaudry, G. R., Chapalamaduga, S., (1991) Biodegradation of halogenated organic compounds. *Microbiol Rev.* **55**:59-79.

Christensen, T., Gooden, D. M., Kung, J. E., Toone, E. J., (2003) Additivity and the physical basis of multivalency effects: a thermodynamic investigation of the calcium EDTA interaction. *J Am Chem Soc* **125**:7357-7366.

Clayton, G. M., Silverman, W. R., Heginbotham, L., Morais-Cabral, J. H., (2004) Structural basis of ligand activation in a cyclic nucleotide regulated potassium channel. *Cell* **119**:615-627.

Collaborative Computational Project, Number 4. (1994) The CCP4 Suite: Programs for Protein Crystallography. *Acta Cryst. D* **50**:760-763.

Collins, J. R., Camper, D. L., Loew, G. H., (1991) Valproic Acid Metabolism by Cytochrome P450: A Theoretical Study of Stereoelectronic Modulators of Product Distribution. *J. Am. Chem. Soc.* **113**:2736-2743.

Congreve, M., Murray, C. W., Blundell, T. L., (2005) Structural biology and drug discovery. *Drug Discov Today* **10**:895-907.

Copley, S. D., (1998) Microbial dehalogenases: enzymes recruited to convert xenobiotic substrates. *Curr Opin Chem Biol* **2**:613-617.

Corpet, F., (1988) Multiple sequence alignment with hierarchical clustering. *Nucleic Acids Res* **16**:10881-10890.

Crowther, R. A., (1969) The use of non-crystallographic symmetry for phase determination. *Acta Cryst. B* **25**:2571-2580.

Crowther, R. A., Blow, D. M. (1967) A method of positioning a known molecule in an unknown crystal structure. *Acta Cryst.* **23**:544-548.

Cupp-Vickery, J. R., Poulos, T. L., (1995) Structure of cytochrome P450eryF involved in erythromycin biosynthesis. *Nat Struct Biol* **2**:144-153.

Cupp-Vickery, J., Anderson, R., Hatziris, Z., (2000) Crystal structures of ligand complexes of P450eryF exhibiting homotropic cooperativity. *Proc Natl Acad Sci U S A* **97**:3050-3055.

Daff, S. N., Chapman, S. K., Turner, K. L., Holt, R. A., Govindaraj, S., Poulos, T. L., Munro, A. W., (1997) Redox control of the catalytic cycle of flavocytochrome P-450 BM3. *Biochemistry* **36**:13816-13823.

Daff, S., (2004) An appraisal of multiple NADPH binding-site models proposed for cytochrome P450 reductase, NO synthase, and related diflavin reductase systems. *Biochemistry* **43**:3929-3932.

Darrouzet, E., Moser, C. C., Dutton, P. L., Daldal F., (2001) Large scale domain movement in cytochrome *bc*<sub>1</sub>: a new device for electron transfer in proteins. *Trends Biochem. Sci.* **26**:445-451.

Dauter, Z., Dauter, M., Rajashankar, K. R., (2000) Novel approach to phasing proteins: derivatization by short cryo-soaking with halides. *Acta Cryst. D* **56**:232-7.

De Boer, J. H., (1936) The influence of van der waals' forces and primary bonds on binding energy, strength and orientation, with special reference to some artificial resins. *TRANSACTIONS OF THE FARADAY SOCIETY* **32**:10-36

- De Mot, R., Parret, A.H., (2002) A novel class of self-sufficient cytochrome P450 monooxygenases in prokaryotes. *Trends Microbiol.* **10**:502-508.
- Deacon, A.M., Ealick, S.E., (1999) Selenium-based MAD phasing: setting the sites on larger structures. *Structure Fold Des.* **7**:161-166.
- DeLano, W.L., (2002) The PyMOL Molecular Graphics System DeLano Scientific, San Carlos, CA, USA. <http://www.pymol.org>
- Deng, Z., Aliverti, A., Zanetti, G., Arakaki, A. K., Ottado, J., Orellano, E. G., Calcaterra, N. B., Ceccarelli, E. A., Carrillo, N., Karplus, P. A., (1999) A productive NADP<sup>+</sup> binding mode of ferredoxin-NADP<sup>+</sup> reductase revealed by protein engineering and crystallographic studies. *Nat Struct Biol* **6**:847-853.
- Dohr, O., Paine, M. J., Friedberg, T., Roberts, G. C., Wolf, C. R., (2001) Engineering of a functional human NADH-dependent cytochrome P450 system. *Proc Natl Acad Sci U S A* **98**:81-86.
- Drenth, J. (1999). Principles of Protein X-ray Crystallography, 2nd ed. (Canton, C.R., ed.) New York: Springer-Verlag.
- Ducruix, A., Giege, R., (1999) Crystallisation of proteins and nucleic acids: a practical approach, 2nd ed., IRL Press, Oxford.
- Dunford, A. J., Marshall, K. R., Munro, A. W., Scrutton, N. S., (2004) Thermodynamic and kinetic analysis of the isolated FAD domain of rat neuronal nitric oxide synthase altered in the region of the FAD shielding residue Phe1395. *Eur J Biochem* **271**:2548-2560.
- Dym, O., Eisenberg, D., (2001) Sequence-structure analysis of FAD-containing proteins. *Protein Sci* **10**:1712-1728.
- Ebright, R. H., Ebright, Y. W., Pendergrast, P. S., Gunasekera, A., (1990) Conversion of a helix-turn-helix motif sequence-specific DNA binding protein into a site-specific DNA cleavage agent. *Proc Natl Acad Sci U S A* **87**:2882-2886.
- Eiting, M., Hageluken, G., Schubert, W. D., Heinz, D. W., (2005) The mutation G145S in PrfA, a key virulence regulator of Listeria monocytogenes, increases DNA-binding affinity by stabilizing the HTH motif. *Mol Microbiol* **56**:433-446.
- El Fantroussi, S., Naveau, H., Agathos, S. N., (1998) Anaerobic dechlorinating bacteria. *Biotechnol Prog* **14**:167-188.
- Elrod-Erickson, M., Rould, M. A., Nekludova, L., Pabo, C. O., (1996) Zif268 protein-DNA complex refined at 1.6 Å: a model system for understanding zinc finger-DNA interactions. *Structure* **4**:1171-1180.
- Emsley, P., Cowtan, K., (2004) Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr* **60**:2126-2132.
- Fanchon, E., Hendrickson, W. A., (1990) Effect of the anisotropy of anomalous scattering on the MAD phasing method. *Acta Crystallogr A* **46**:809-820.

Fersht, A. R., Shi, J. P., Knill-Jones, J., Lowe, D. M., Wilkinson, A. J., Blow, D. M., Brick, P., Carter, P., Waye, M. M. Y., Winter, G., (1985) Hydrogen bonding and biological specificity analyzed by protein engineering. *Nature* **314**:235-238.

Finn, R. D., Basran, J., Roitel, O., Wolf, C. R., Munro, A. W., Paine, M. J., Scrutton, N. S., (2003) Determination of the redox potentials and electron transfer properties of the FAD- and FMN-binding domains of the human oxidoreductase NR1. *Eur J Biochem* **270**:1164-1175.

French G.S., Wilson K.S., (1978) On the treatment of negative intensity observations. *Acta Cryst. A* **34**:517.

Freund, A. K., (1996) Third-generation synchrotron radiation X-ray optics. *Structure* **4**:121-125.

Frishman D. et al., (2003). The PEDANT genome database *Nucleic Acids Res.*, **31**, 207-211.

Fujii, Y., Shimizu, T., Kusumoto, M., Kyogoku, Y., Taniguchi, T., Hakoshima, T., (1999) Crystal structure of an IRF-DNA complex reveals novel DNA recognition and cooperative binding to a tandem repeat of core sequences. *EMBO J* **18**:5028-5041.

Fulco, A. J., (1991) P450BM-3 and other inducible bacterial P450 cytochromes: biochemistry and regulation. *Annu Rev Pharmacol Toxicol.* **31**:177-203.

Garcin, E. D., Bruns, C. M., Lloyd, S. J., Hosfield, D. J., Tiso, M., Gachhui, R., Stuehr, D. J., Tainer, J. A., Getzoff, E. D., (2004) Structural basis for isozyme-specific regulation of electron transfer in nitric-oxide synthase. *J Biol Chem* **279**:37918-37927.

Georgopapadakou, N. H., (1998) Antifungals: mechanism of action and resistance, established and novel drugs. *Curr Opin Microbiol* **1**:547-557.

Giacovazzo, C., Siliqi, D., (2002) The method of joint probability distribution functions applied to SIR-MIR and to SIRAS-MIRAS cases. *Acta Crystallogr A* **58**:590-597.

Giege, R., Drenth, J., Ducruix, A., McPherson, A., Saenger, W., (1995) Crystallogenesis of biological macromolecules. Biological , microgravity , and other physico-chemical aspects. *Prog. Cryst. Growth Charact.* **30**:237-281.

Gilliland, G. L., (1988) A biological crystallization database: a basis for a crystallization strategy. *J. Cryst. Growth* **90**:51-59.

Giraud-Panis, M. J., Toulme, F., Blazy, B., Maurizot, J. C., Culard, F., (1994) Fluorescence study on the non-specific binding of cyclic-AMP receptor protein to DNA: effect of pH. *Biochimie* **76**:133-139.

Girvan, H. M., Marshall, K. R., Lawson, R. J., Leys, D., Joyce, M. G., Clarkson, J., Smith, W. E., Cheesman, M. R., Munro, A. W., (2004) Flavocytochrome P450 BM3 mutant A264E undergoes substrate-dependent formation of a novel heme iron ligand set. *J Biol Chem* **279**:23274-23286.

Glieder, A., Farinas, E. T., Arnold, F. H., (2002) Laboratory evolution of a soluble, self-

- sufficient, highly active alkane hydroxylase. *Nat Biotechnol* **20**:1135-1139.
- Gonzalez, F. J., (2005) Role of cytochromes P450 in chemical toxicity and oxidative stress: studies with CYP2E1. *Mutat Res* **569**:101-110.
- Govindaraj, S., Poulos, T. L., (1997) The domain architecture of cytochrome P450BM-3. *J Biol Chem* **272**:7915-7921.
- Graham-Lorence, S., Truan, G., Peterson, J. A., Falck, J. R., Wei, S., Helvig, C., Capdevila, J. H., (1997) An active site substitution, F87V, converts cytochrome P450 BM-3 into a regio- and stereoselective (14S,15R)-arachidonic acid epoxygenase. *J Biol Chem* **272**:1127-1135.
- Green, J., Sharrocks, A. D., Green, B., Geisow, M., Guest, J. R., (1993) Properties of FNR proteins substituted at each of the five cysteine residues. *Mol. Microbiol.* **8**:61-68.
- Gribble, G. W., (1998) Naturally Occurring Organohalogen Compounds. *Acc. Chem. Res.* **31**:141-152
- Gruez, A., Pignol, D., Zeghouf, M., Coves, J., Fontecave, M., Ferrer, J. L., Fontecilla-Camps, J. C., (2000) Four crystal structures of the 60 kDa flavoprotein monomer of the sulfite reductase indicate a disordered flavodoxin-like module. *J Mol Biol* **299**:199-212.
- Gruner, S. M., Milch, J. R., Reynolds, G. T., (1978) Evaluation of area photon detectors by a method based on detective quantum energy (DQE). *IEEE Trans. Nucl. Sci.* **NS-25**, 562-565.
- Guengerich, F. P., and Shimada, T., (1991) Oxidation of toxic and carcinogenic chemicals by human cytochrome P-450 enzymes. *Chem. Res. Toxicol.* **4**, 391-407.
- Gunsalus, I. C., Pederson, T. C., Sligar, S. G., (1975) Oxygenase-catalyzed biological hydroxylations. *Annu Rev Biochem.* **44**:377-407.
- Gustaffson, M., (2000) Bacterial cytochromes P450 Studies on cytochrome P450 102A2 and P450 102A3 of *Bacillus subtilis*. Ph.D. thesis, Department of Microbiology, Lund University, Sweden.
- Gutierrez, A., Doehr, O., Paine, M., Wolf, C. R., Scrutton, N. S., Roberts, G. C., (2000) Trp-676 facilitates nicotinamide coenzyme exchange in the reductive half-reaction of human cytochrome P450 reductase: properties of the soluble W676H and W676A mutant reductases. *Biochemistry* **39**:15990-15999.
- Haines, D.C., Tomchick, D.R., Machius, M., Peterson, J.A., (2001) Pivotal role of water in the mechanism of P450BM-3. *Biochemistry* **40**:13456-13465.
- Hammes, G. G., (1964) Mechanism of enzyme catalysis. *Nature* **204**:342-343.
- Hanley, S. C., Ost, T. W., Daff, S., (2004) The unusual redox properties of flavocytochrome P450 BM3 flavodoxin domain. *Biochem Biophys Res Commun* **325**:1418-1423.
- Hansen, S. B., Sulzenbacher, G., Huxford, T., Marchot, P., Taylor, P., Bourne, Y., (2005) Structures of Aplysia AChBP complexes with nicotinic agonists and antagonists reveal distinctive binding interfaces and conformations. *EMBO J* **24**:3635-3646.

Harris, D., Loew, G.H., (1995) Prediction of regiospecific hydroxylation of camphor analogs by cytochrome P450cam. *J Am Chem Soc* **117**:2837-2746.

Hasemann, C. A., Ravichandran, K. G., Peterson, J. A., Deisenhofer, J., (1994) Crystal structure and refinement of cytochrome P450terp at 2.3 Å resolution. *J Mol Biol* **236**:1169-1185.

Helliwell, J. R., (1992) Macromolecular crystallography with synchrotron radiation. Cambridge University Press.

Helliwell, J. R., Synchrotron-radiation instrumentation, methods and scientific utilization In: Rossman, M.G., Arnold, E., (2001) International Tables for Crystallography Volume F Crystallography of biological macromolecules 1<sup>st</sup> ed. Kluwer Academic Publishers Dordrecht, pp 155-164

Henne, K. R., Kunze, K. L., Zheng, Y. M., Christmas, P., Soberman, R. J., Rettie, A. E., (2001) Covalent linkage of prosthetic heme to CYP4 family P450 enzymes. *Biochemistry* **40**:12925-12931.

Heyduk, T., Lee, J. C., (1989) Escherichia coli cAMP receptor protein: evidence for three protein conformational states with different promoter binding affinities. *Biochemistry* **28**:6914-6924.

Hileman, B., (1993). Concerns broaden over chlorine and chlorinated hydrocarbons. *Chem. Eng. News*, **71**, 11-20.

Hoch, U., Ortiz De Montellano, P. R., (2001) Covalently linked heme in cytochrome p4504a fatty acid hydroxylases. *J Biol Chem* **276**:11339-11346.

Holliger, C., Schraa, G., (1994) Physiological meaning and potential for application of reductive dechlorination by anaerobic bacteria. *FEMS Microbiol. Rev.* **2**:297-305.

Holliger, C., Schumacher, W., (1994) Reductive dehalogenation as a respiratory process. *Antonie Van Leeuwenhoek* **66**:239-46.

Holliger, C., Wohlfarth, G., Diekert, G., (1999) Reductive dechlorination in the energy metabolism of anaerobic bacteria. *FEMS Microbiol. Rev.* **22**:383-398.

Holm, L., Sander, C., (1998) Touring protein fold space with Dali/FSSP. *Nucleic Acids Res* **26**:316-319.

Hooft, R. W., Sander, C., Vriend, G., (1996) Positioning hydrogen atoms by optimizing hydrogen-bond networks in protein structures. *Proteins* **26**:363-376.

Hope, H., (1988) Cryocrystallography of biological macromolecules: a generally applicable method. *Acta Crystallogr B* **44**:22-26.

Hope, H., Cryocrystallography In: Rossman, M.G., Arnold, E., (2001) International Tables for Crystallography Volume F Crystallography of biological macromolecules 1<sup>st</sup> ed. Kluwer Academic Publishers Dordrecht, pp 197-201.

- Horowitz, (1983) Isolation and characterization of Tn-Dha1, a transposon containing the tetrachloroethene reductive dehalogenase of *Desulfitobacterium hafniense* strain TCE1. *Appl Environ Micro*. **45**:1459-1465
- Hubbard, P. A., Shen, A. L., Paschke, R., Kasper, C. B., Kim, J. J., (2001) NADPH-cytochrome P450 oxidoreductase. Structural basis for hydride and electron transfer. *J Biol Chem* **276**:29163-29170.
- Hudson, J. M., Crowe, L. G., Fried, M. G., (1990) A new DNA binding mode for CAP. *J Biol Chem* **265**:3219-3225.
- Ingelman, M., Bianchi, V., Eklund, H., (1997) The three-dimensional structure of flavodoxin reductase from *Escherichia coli* at 1.7 Å resolution. *J Mol Biol* **268**:147-157.
- Ingelman-Sundberg, M., (2002) Polymorphism of cytochrome P450 and xenobiotic toxicity. *Toxicology* **181-182**:447-452.
- Isserman, I., Green, S., (1990) Activation of a member of the steroid hormone receptor superfamily by peroxisome proliferators. *Nature* **347**:645-649.
- Jackson, D. A., (2005) The amazing complexity of transcription factories. *Brief Funct Genomic Proteomic* **4**:143-157.
- Janssen, D. B., Oppentocht, J. E., Poelarends, G. J., (2001) Microbial dehalogenation. *Curr Opin Biotechnol* **12**:254-258.
- Janssen, D. B., Pries, F., van der Ploeg, J. R., (1994) Genetics and biochemistry of dehalogenating enzymes. *Annu Rev Microbiol* **48**:163-191.
- Jeffrey, G.A., Saenger, W., (1991) Hydrogen Bonding in Biological Structures, pp. 27-29. Berlin: Springer-Verlag.
- Jones, T. A., Zou, J. Y., Cowan, S. W., Kjeldgaard, (1991) Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallogr A* **47**:110-119.
- Joseph, D., Petsko, G. A., Karplus, M., (1990) Anatomy of a conformational change: hinged "lid" motion of the triosephosphate isomerase loop. *Science* **249**:1425-1428.
- K. Cowtan (1994), Joint CCP4 and ESF-EACBM Newsletter on Protein Crystallography, 31, p34-38.
- Kabsch, W., (1976) A solution for the best rotation to relate two sets of vectors. *Acta Cryst. A* **32**:922-923.
- Keller, W., Konig, P., Richmond, T. J., (1995) Crystal structure of a bZIP/DNA complex at 2.2 Å: determinants of DNA specific recognition. *J Mol Biol* **254**:657-667.
- Kleywegt, G. J., Jones, T. A., (1998) Databases in protein crystallography. *Acta Crystallogr D Biol Crystallogr* **54**:1119-1131.
- Kleywegt, G. J., Zou, J. Y., Divne, C., Davies, G. J., Sinning, I., Stahlberg, J., Reinikainen,

T., Srisodsuk, M., Teeri, T. T., Jones, T. A., (1997) The crystal structure of the catalytic core domain of endoglucanase I from *Trichoderma reesei* at 3.6 Å resolution, and a comparison with related enzymes. *J Mol Biol* **272**:383-397.

Kleywegt, G. J., Structure validation In: Rossman, M.G., Arnold, E., (2001) International Tables for Crystallography Volume F Crystallography of biological macromolecules 1<sup>st</sup> ed. Kluwer Academic Publishers Dordrecht, pp 497-506

Kolb, A., Busby, S., Buc, H., Garges, S., Adhya, S., (1993) Transcriptional regulation by cAMP and its receptor protein. *Annu Rev Biochem* **62**:749-795.

Koopmann, E., Hahlbrock, K., (1997) Differentially regulated NADPH:cytochrome P450 oxidoreductases in parsley. *Proc Natl Acad Sci U S A* **94**:14954-14959.

Korner, H., Sofia, H. J., Zumft, W. G., (2003) Phylogeny of the bacterial superfamily of Crp-Fnr transcription regulators: exploiting the metabolic spectrum by controlling alternative gene programs. *FEMS Microbiol Rev* **27**:559-592.

Kuhn, E. P., Suflita, J. M., (1989) Sequential reductive dehalogenation of chloroanilines by microorganisms from a methanogenic aquifer. *Environ Sci. Technol.* **23**:848-852.

Lancaster, C.R.D., Michel, H., Honig, B., Gunner, M.R., (1996). Calculated coupling of electron and proton transfer in the photosynthetic reaction center of *Rhodopseudomonas viridis*. *Biophys. J.* **70**:2469-2492.

Lanzilotta, W. N., Schuller, D. J., Thorsteinsson, M. V., Kerby, R. L., Roberts, G. P., Poulos, T. L., (2000) Structure of the CO sensing transcription activator CooA. *Nat Struct Biol* **7**:876-880.

Laskowski, R. A., (2001) PDBsum: summaries and analyses of PDB structures. *Nucleic Acids Res* **29**:221-222.

Laskowski, R. A., MacArthur, M. W., Moss D. S., Thornton, J. M., (1993) PROCHECK: a program to check the stereochemical quality of protein structures. *J. App. Cryst.* **26**:283-291.

Lawrence, M. C., Colman, P. M., (1993) Shape complementarity at protein/protein interfaces. *J Mol Biol* **234**:946-950.

LeBrun, L. A., Hoch, U., Ortiz de Montellano, P. R., (2002) Autocatalytic mechanism and consequences of covalent heme attachment in the cytochrome P4504A family. *J Biol Chem* **277**:12755-12761.

Lee, D. S., Yamada, A., Sugimoto, H., Matsunaga, I., Ogura, H., Ichihara, K., Adachi, S., Park, S. Y., Shiro, Y., (2003) Substrate recognition and molecular mechanism of fatty acid hydroxylation by cytochrome P450 from *Bacillus subtilis*. Crystallographic, spectroscopic, and mutational studies. *J Biol Chem* **278**:9761-9767.

Leslie, A. G. W. (1987). A reciprocal space method for calculating a molecular envelope using the algorithm of Wang, B.C. *Acta Cryst. A* **43**:134-136.

Lewis, D. F. V., (2001) Guide to cytochrome P450 structure and function, 1<sup>st</sup> ed., Taylor and

Francis, London and New York.

Leys, D., Mowat, C. G., McLean, K. J., Richmond, A., Chapman, S. K., Walkinshaw, M. D., Munro, A. W., (2003) Atomic structure of *Mycobacterium tuberculosis* CYP121 to 1.06 Å reveals novel features of cytochrome P450. *J Biol Chem* **278**:5141-5147.

Li, H., Poulos, T. L., (1997) The structure of the cytochrome p450BM-3 haem domain complexed with the fatty acid substrate, palmitoleic acid. *Nat Struct Biol* **4**:140-146.

Li, H., Poulos, T. L., (2004) Crystallization of cytochromes P450 and substrate-enzyme interactions. *Curr. Top. Med. Chem.* **4**:1789-1802.

Li, Q. S., Ogawa, J., Schmid, R. D., Shimizu, S., (2005) Indole hydroxylation by bacterial cytochrome P450 BM-3 and modulation of activity by cumene hydroperoxide. *Biosci Biotechnol Biochem* **69**:293-300.

Li, T., Stark, M. R., Johnson, A. D., Wolberger, C., (1995) Crystal structure of the MATα1/MAT alpha 2 homeodomain heterodimer bound to DNA. *Science* **270**:262-269.

Littlefield, O., Nelson, H. C., (1999) A new use for the 'wing' of the 'winged' helix-turn-helix motif in the HSF-DNA cocrystal. *Nat Struct Biol* **6**:464-470.

Luger, K., Mader, A. W., Richmond, R. K., Sargent, D. F., Richmond, T. J., (1997) Crystal structure of the nucleosome core particle at 2.8 Å resolution. *Nature* **389**:251-60.

Luque, I., Freire, E., (2002) Structural parameterization of the binding enthalpy of small ligands. *Proteins* **49**:181-190.

Luscombe, N. M., Austin, S. E., Berman, H. M., Thornton, J. M., (2000) An overview of the structures of protein-DNA complexes. *Genome Biology* **1**:1-37.

Malone, T. C., (1970) In vitro conversion of DDT to DDD by the intestinal microflora of the Northern anchovy, *Engraulis mordax*. *Nature* **227**:848-849.

Martinis, S. A., Blanke, S. R., Hager, L. P., Sligar, S. G., Hoa, G. H., Rux, J. J., Dawson, J. H., (1996) Probing the heme iron coordination structure of pressure-induced cytochrome P420cam. *Biochemistry* **35**:14530-14536.

Matthews, B. W., (1985) Determination of protein molecular weight, hydration and packing from crystal density. *Methods Enzymol.* **114**, 176-187.

McLean, K. J., Scrutton, N. S., Munro, A. W., (2003) Kinetic, spectroscopic and thermodynamic characterization of the *Mycobacterium tuberculosis* adrenodoxin reductase homologue FprA. *Biochem J* **372**:317-327.

McPherson, A., Malkin, A. J., Kuznetsov, Y. G., (1995) The science of macromolecular crystallisation. *Structure* **3**:759-768.

Messmer, M., Wohlfarth, G., Diekert, G., (1993) Methyl chloride metabolism of the strictly anaerobic methyl-utilizing homoacetogen strain MC. *Arch Microbiol.* **160**:383-387.

Miles, C. S., Ost, T. W., Noble, M. A., Munro, A. W., Chapman, S. K., (2000) Protein

engineering of cytochromes P-450. *Bio. Biophys Acta* **1543**, 383-407.

Miles, J. S., Munro, A. W., Rospodowski, B. N., Smith, W. E., McKnight, J., Thomson, A. J., (1992) Domains of the catalytically self-sufficient cytochrome P-450 BM-3. Genetic construction, overexpression, purification and spectroscopic characterization. *Biochem J* **288**:503-509.

Miller, J., McLachlan, A. D., Klug, A., (1985) Repetitive zinc-binding domains in the protein transcription factor IIIA from Xenopus oocytes. *EMBO J* **4**:1609-1614.

Mo, Y., Vaessen, B., Johnston, K., Marmorstein, R., (1998) Structures of SAP-1 bound to DNA targets from the E74 and c-fos promoters: insights into DNA sequence discrimination by Ets proteins. *Mol Cell* **2**:201-212.

Modi, S., Gilham, D. E., Sutcliffe, M. J., Lian, L. Y., Primrose, W. U., Wolf, C. R., Roberts, G. C., (1997) 1-methyl-4-phenyl-1,2,3,6-tetrahydropyridine as a substrate of cytochrome P450 2D6: allosteric effects of NADPH-cytochrome P450 reductase. *Biochemistry* **36**:4461-4470.

Modi, S., Sutcliffe, M. J., Primrose, W. U., Lian, L. Y., Roberts, G. C., (1996) The catalytic mechanism of cytochrome P450 BM3 involves a 6 Å movement of the bound substrate on reduction. *Nat Struct Biol* **3**:414-417.

Mohn, W. W., and Tiedje, J. M., (1990) catabolic thiosulfate disproportionation and carbon dioxide reduction in strain DCb-1, a reductively dechlorinating anaerobe. *J. Bacteriol.* **172**: 2065-2070.

Moy, J. P., (1999) Large area X-ray detectors based on amorphous silicon detector. *Thin Solid Films* **337**:213.

Muchmore, S. W., Olson, J., Jones, R., Pan, J., Blum, M., Greer, J., Merrick, S. M., Magdalinos, P., Nienaber, V. L., (2000) Automated crystal mounting and data collection for protein crystallography. *Structure Fold Des* **8**:R243-246.

Muller, N., (1990) Search for a realistic view of hydrophobic effects. *Acc. Chem. Res.* **23**:23-28

Munro, A. W., Malarkey, K., McKnight, J., Thomson, A. J., Kelly, S. M., Price, N. C., Lindsay, J. G., Coggins, J. R., Miles, J. S., (1994) The role of tryptophan 97 of cytochrome P450 BM3 from *Bacillus megaterium* in catalytic function. Evidence against the 'covalent switching' hypothesis of P-450 electron transfer. *Biochem J* **303**:423-428.

Munro, A. W., Daff, S., Coggins, J. R., Lindsay, J. G., Chapman, S. K., (1996) Probing electron transfer in flavocytochrome P450 BM3 and its component domains. *Eur. J. Biochem.* **239**, 403-409.

Munro, A. W., Leys, D. G., McLean, K. J., Marshall, K. R., Ost, T. W., Daff, S., Miles, C. S., Chapman, S. K., Lysek, D. A., Moser, C. C., Page, C. C., Dutton, P. L., (2002) P450 BM3: the very model of a modern flavocytochrome. *Trends Biochem. Sci.* **27**, 250-257.

Munzer, D. F., Meinhold, P., Peters, M. W., Feichtenhofer, S., Griengl, H., Arnold, F. H., Glieder, A., de Raadt, A., (2005) Stereoselective hydroxylation of an achiral

cyclopentanecarboxylic acid derivative using engineered P450s BM-3. *Chem Commun (Camb)* 2597-2599.

Murshudov, G. N., Vagin, A. A., Dodson, E. J., (1997) Refinement of Macromolecular Structures by the Maximum-Likelihood Method. *Acta Cryst. D* 53:240-255.

Nagano, S., Li, H., Shimizu, H., Nishida, C., Ogura, H., Ortiz de Montellano, P. R., Poulos, T. L., (2003) Crystal structures of epothilone D-bound, epothilone B-bound, and substrate-free forms of cytochrome P450epoK. *J Biol Chem* 278:44886-44893.

Narhi, L. O., Fulco, A. J., (1986) Characterization of a catalytically self-sufficient 119,000-dalton cytochrome P-450 monooxygenase induced by barbiturates in *Bacillus megaterium*. *J Biol Chem* 261:7160-7169.

Narhi, L. O., Fulco, A. J., (1987) Identification and characterization of two functional domains in cytochrome P-450BM-3, a catalytically self-sufficient monooxygenase induced by barbiturates in *Bacillus megaterium*. *J Biol Chem* 262:6683-6690.

Navaza, J., (1994) AMoRe: an automated package for molecular replacement. *Acta Cryst. A* 50:157-163.

Nebert, D. W., Nelson, D. R., Coon, M. J., Estabrook, R. W., Feyereisen, R., Fujii-Kuriyama, Y., Gonzalez, F. J., Guengerich, F. P., Gunsalus, I. C., Johnson, E. F., et al., (1991) The P450 superfamily: update on new sequences, gene mapping, and recommended nomenclature. *DNA Cell Biol* 10:1-14.

Nebert, D. W., Nelson, D. R., Feyereisen, R., (1989) Evolution of the cytochrome P450 genes. *Xenobiotica* 19:1149-1160.

Neeli, R., Roitel, O., Scrutton, N. S., Munro, A. W., (2005) Switching pyridine nucleotide specificity in P450 BM3: mechanistic analysis of the W1046H and W1046A enzymes. *J Biol Chem* 280:17634-17644.

Nelson, D. R., Koymans, L., Kamataki, T., Stegeman, J. J., Feyereisen, R., Waxman, D. J., Waterman, M. R., Gotoh, O., Coon, M. J., Estabrook, R. W., Gunsalus, I. C., Nebert, D. W., (1996) P450 superfamily: update on new sequences, gene mapping, accession numbers and nomenclature. *Pharmacogenetics* 6:1-42.

Nelson, D. L. & Cox, M. M. (eds.) (2000). *Lehninger Principles of Biochemistry*. Worth Publishers, New York.

Nelson, D. R., (1999) Cytochrome P450 and the individuality of species. *Arch Biochem Biophys* 369:1-10.

Neumann, A., Wohlfarth, G., Diekert, G., (1996) Purification and characterization of tetrachloroethene reductive dehalogenase from *Dehalospirillum multivorans*. *J Biol Chem* 271:16515-16519.

Noble, M. A., Miles, C. S., Chapman, S. K., Lysek, D. A., MacKay, A. C., Reid, G. A., Hanzlik, R. P., Munro, A. W., (1999) Roles of key active-site residues in flavocytochrome P450 Bm3. *Biochem. J.* 339:371-379.

- Okita, R. T., Okita, J. R., (2001) Cytochrome P450 4A fatty acid omega hydroxylases. *Curr Drug Metab* **2**:265-281.
- Oliver, C. F., Modi, S., Sutcliffe, M. J., Primrose, W. U., Lian, L. Y., Roberts, G. C., (1997) A single mutation in cytochrome P450 BM3 changes substrate orientation in a catalytic intermediate and the regiospecificity of hydroxylation. *Biochemistry* **36**:1567-1572.
- Ost, T. W., Miles, C. S., Munro, A. W., Murdoch, J., Reid, G. A., Chapman, S. K., (2001) Phenylalanine 393 exerts thermodynamic control over the heme of flavocytochrome P450 BM3. *Biochemistry* **40**:13421-13429.
- Ostermann, A., Waschikpy, R., Parak, F. G., Nienhaus, G. U., (2000) Ligand binding and conformational motions in myoglobin. *Nature* **404**:205-208.
- Otwinowski, Z., (1991). Daresbury Study Weekend proceedings.
- Otwinowski, Z., Minor, W., (1997) Processing of x-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276**:307-326.
- Page, C. C., Moser, C. C., Chen, X., Dutton, P. L., (1999) Natural engineering principles of electron tunnelling in biological oxidation-reduction. *Nature* **402**:47-52.
- Paine, M. J., Garner, A. P., Powell, D., Sibbald, J., Sales, M., Pratt, N., Smith, T., Tew, D. G., Wolf, C. R., (2000) Cloning and characterization of a novel human dual flavin reductase. *J Biol Chem* **275**:1471-1478.
- Palfey, B. A., Moran, G. R., Entsch, B., Ballou, D. P., Massey, V., (1999) Substrate recognition by "password" in p-hydroxybenzoate hydroxylase. *Biochemistry* **38**:1153-1158.
- Park, H. W., Kim, S. T., Sancar, A., Deisenhofer, J., (1995) Crystal structure of DNA photolyase from Escherichia coli. *Science* **268**:1866-1872.
- Park, S. Y., Shimizu, H., Adachi, S., Shiro, Y., Iizuka, T., Nakagawa, A., Tanaka, I., Shoun, H., Hori, H., (1997) Crystallization, preliminary diffraction and electron paramagnetic resonance studies of a single crystal of cytochrome P450nor. *FEBS Lett* **412**:346-350.
- Parkinson, G., Wilson, C., Gunasekera, A., Ebright, Y. W., Ebright, R. E., Berman, H. M., (1996) Structure of the CAP-DNA complex at 2.5 angstroms resolution: a complete picture of the protein-DNA interface. *J Mol Biol* **260**:395-408.
- Parraga, A., Bellsolell, L., Ferre-D'Amare, A. R., Burley, S. K., (1998) Co-crystal structure of sterol regulatory element binding protein 1a at 2.3 Å resolution. *Structure* **6**:661-672.
- Passner, J. M., Schultz, S. C., Steitz, T. A., (2000) Modeling the cAMP-induced allosteric transition using the crystal structure of CAP-cAMP at 2.1 Å resolution. *J Mol Biol* **304**:847-859.
- Patterson, A. L., (1934) A Fourier Series Method for the Determination of the Components of Interatomic Distances in Crystals. *Physics Review* **46**: 372-376
- Perera, R., Sono, M., Sigman, J. A., Pfister, T. D., Lu, Y., Dawson, J. H., (2003) Neutral thiol as a proximal ligand to ferrous heme iron: implications for heme proteins that lose

- cysteine thiolate ligation on reduction. *Proc Natl Acad Sci U S A* **100**:3641-3646.
- Perrakis, A., Harkiolaki, M., Wilson, K. S., Lamzin, V. S., (2001) ARP/wARP and molecular replacement. *Acta Crystallogr D Biol Crystallogr* **57**:1445-1450.
- Perrin, C.L. & Nielson, J.B. (1997). "Strong" hydrogen bonds in chemistry and biology. *Annu. Rev. Phys. Chem.* **48**, 511-544.
- Petsko, G.A., (1975) Protein crystallography at sub-zero temperatures: cryo-protective mother liquors for protein crystals. *J Mol Biol.* **96**:381-392.
- Podust, L. M., Poulos, T. L., Waterman, M. R., (2001) Crystal structure of cytochrome P450 14alpha-sterol demethylase (CYP51) from *Mycobacterium tuberculosis* in complex with azole inhibitors. *Proc Natl Acad Sci U S A* **98**:3068-3073.
- Pop, S. M., Kolarik, R. J., Ragsdale, S. W., (2004) Regulation of anaerobic dehalorespiration by the transcriptional activator CprK. *J Biol Chem* **279**:49910-49918.
- Porter, T. D., Kasper, C. B., (1986) NADPH-cytochrome P-450 oxidoreductase: flavin mononucleotide and flavin adenine dinucleotide domains evolved from different flavoproteins. *Biochemistry* **25**:1682-1687.
- Poulos, T. L., Finzel, B. C., Gunsalus, I. C., Wagner, G. C., Kraut, J., (1985) The 2.6-A crystal structure of *Pseudomonas putida* cytochrome P-450. *J Biol Chem* **260**:16122-16130.
- Poulos, T. L., Finzel, B. C., Howard, A. J., (1987) High-resolution crystal structure of cytochrome P450cam. *J Mol Biol* **195**:687-700.
- Poulos, T. L., Johnson, E. F., Structures of Cytochrome P450 enzymes. Ortiz de Montellano, P. R., (2005) In *Cytochrome P450: structure, mechanism, and biochemistry*. 3rd ed. Kluwer Academic/Plenum, pp 87-113.
- Pusey, M. L., Liu, Z. J., Tempel, W., Praissman, J., Lin, D., Wang, B. C., Gavira, J. A., Ng, J. D., (2005) Life in the fast lane for protein crystallization and X-ray crystallography. *Prog Biophys Mol Biol* **88**:359-386.
- Raag, R., Poulos, T. L., (1991) Crystal structures of cytochrome P-450CAM complexed with camphane, thiocamphor, and adamantane: factors controlling P-450 substrate hydroxylation. *Biochemistry* **30**:2674-2684.
- Raghu, K., MacRae, I. C., (1966) Biodegradation of the gamma isomer of benzene hexachloride in submerged soils. *Science* **154**:263-264.
- Ramachandran, G. N., Sasisekharan V., (1968) Conformation of polypeptides and proteins. *Adv Protein Chem.* **23**:283-438.
- Ravelli, R. B., McSweeney, S. M., (2000) The 'fingerprint' that X-rays can leave on structures. *Structure Fold Des* **8**:315-328.
- Ravichandran, K. G., Boddupalli, S. S., Hasermann, C. A., Peterson, J. A., Deisenhofer, J., (1993) Crystal structure of hemoprotein domain of P450BM-3, a prototype for microsomal

P450's. *Science* **261**:731-736.

Reed, J. R., Hollenberg, P. F., (2003) Comparison of substrate metabolism by cytochromes P450 2B1, 2B4, and 2B6: relationship of heme spin state, catalysis, and the effects of cytochrome b5. *J Inorg Biochem* **93**:152-160.

Rehmann, H., Prakash, B., Wolf, E., Rueppel, A., De Rooij, J., Bos, J. L., Wittinghofer, A., (2003) Structure and regulation of the cAMP-binding domains of Epac2. *Nat Struct Biol* **10**:26-32.

Ries-Kautt, M., Ducruix, A., (1992) Phase diagrams. In Crystallization of Nucleic Acids and Proteins. Eds. Ducruix, A., Giegé, R. IRL press, Oxford pp. 195-218

Roitel, O., Scrutton, N. S., Munro, A. W., (2003) Electron transfer in flavocytochrome P450 BM3: kinetics of flavin reduction and oxidation, the role of cysteine 999, and relationships with mammalian cytochrome P450 reductase. *Biochemistry* **42**:10809-10821.

Rossmann, M. G. Data Processing. In: Rossmann, M.G., Arnold, E., (2001) International Tables for Crystallography Volume F Crystallography of biological macromolecules 1<sup>st</sup> ed. Kluwer Academic Publishers Dordrecht, pp 209-245

Rossmann, M. G., (1979). Processing oscillation diffraction data for very large unit cells with an automatic convolution technique and profile fitting. *J. Appl. Cryst.* **12**:225-238.

Rossmann, M. J., Blow, D. M., (1962). The detection of subunits within the crystallographic asymmetric unit. *Acta Cryst.* **15**:24-31.

Roussel, A., Cambillau, C., (1991) Silicon Graphics Geometry Partners Directory 86, Silicon Graphics, Mountain View, CA.

Rubin, G. M., Yandell, M. D., Wortman, J. R., Gabor Miklos, G. L., Nelson, *et al.*, (2000) Comparative genomics of the eukaryotes. *Science* **287**:2204-2215.

Sambrook, J., Sambrook, J., Maniatis, T., (1989) Molecular Cloning: A Laboratory Manual

Sarai, A., Kono, H., (2005) Protein-DNA recognition patterns and predictions. *Ann. Rev. Biophys. Biomol. Struct.* **34**:379-398.

Schwabe, J. W., Rhodes, D., (1991) Beyond zinc fingers: steroid hormone receptors have a novel structural motif for DNA recognition. *Trends Biochem Sci* **16**:291-296.

Scott, E. E., He, Y. A., Wester, M. R., White, M. A., Chin, C. C., Halpert, J. R., Johnson, E. F., Stout, C. D., (2003) An open conformation of mammalian cytochrome P450 2B4 at 1.6-A resolution. *Proc Natl Acad Sci U S A* **100**:13196-13201.

Scrutton, N. S., Berry, A., Perham, R. N., (1990) Redesign of the coenzyme specificity of a dehydrogenase by protein engineering. *Nature* **343**:38-43.

Sevrioukova I. F., Li H., Zhang H., Peterson J. A., Poulos T. L., (1999) Structure of a cytochrome P450-redox partner electron-transfer complex. *Proc. Natl. Acad. Sci. U S A* **96**:1863-1868.

- Sevrioukova, I. F., Hazzard, J. T., Tollin, G., Poulos, T. L., (1999) The FMN to heme electron transfer in cytochrome P450BM-3. Effect of chemical modification of cysteines engineered at the FMN-heme domain interaction site. *J Biol Chem* **274**:36097-36106.
- Sheldrick, G. M., (1996) SADABS. Program for Empirical Absorption Correction of Area Detector Data. University of Gottingen: Germany.
- Shen, A. L., Sem, D. S., Kasper, C. B., (1999) Mechanistic studies on the reductive half-reaction of NADPH-cytochrome P450 oxidoreductase. *J Biol Chem* **274**:5391-5398.
- Smidt, H., van Leest, M., van der Oost, J., de Vos, W.M., (2000) Transcriptional regulation of the cpr gene cluster in ortho-chlorophenol-respiring *Desulfotobacterium dehalogenans*. *J. Bacteriol.* **182**:5683-5691.
- Sridhar Prasad, G., Kresge, N., Muhlberg, A. B., Shaw, A., Jung, Y. S., Burgess, B. K., Stout, C. D., (1998) The crystal structure of NADPH:ferredoxin reductase from *Azotobacter vinelandii*. *Protein Sci* **7**:2541-2549.
- Stout, G. H. & Johnson, L. H. (1989). X-ray Structure Determination, a Practical Guide. New York: John Wiley and Sons.
- Stuehr, D. J., (1999) Mammalian nitric oxide synthases. *Biochim Biophys Acta* **1411**:217-230.
- Stura, E.A., Wilson, I.A., (1990) Analytical and Production Seeding Techniques. *Methods* **1**:38-49.
- Sulistyaningdyah, W. T., Ogawa, J., Li, Q. S., Maeda, C., Yano, Y., Schmid, R. D., Shimizu, S., (2005) Hydroxylation activity of P450 BM-3 mutant F87V towards aromatic compounds and its application to the synthesis of hydroquinone derivatives from phenolic compounds. *Appl Microbiol Biotechnol* **67**:556-562.
- Sun, D. P., Alber, T., Bell, J. A., Weaver, L. H., Matthews, B. W., (1987) Use of site-directed mutagenesis to obtain isomorphous heavy-atom derivatives for protein crystallography: cysteine-containing mutants of phage T4 lysozyme. *Protein Eng* **1**:115-123.
- Sun, P. D., Hammer, C. H., (2000) Mass-spectrometry assisted heavy-atom derivative screening of human FcγRIII crystals. *Acta Cryst D* **56**:161-168.
- Sun, P. D., Radaev, S., Kattah, M., (2002) Generating isomorphous heavy-atom derivatives by a quick-soak method. Part I: test cases. *Acta Cryst. D* **58**:1092-1098.
- Suzuki, M., Brenner, S. E., Gerstein, M., Yagi, N., (1995) DNA recognition code of transcription factors. *Protein Eng* **8**:319-328.
- Taylor, G., (2003) The phase problem. *Acta Cryst. D* **59**:1881-90.
- Ten Eyck, L.F., (1973) Crystallographic fast Fourier transforms *Acta Cryst A* **29**:183-191.
- Thaller, C., Eichele, G., Weaver, L. H., Wilson, E., Karlsson, R., Jansonius, J. N., (1985) Diffraction methods for biological macromolecules. Seed enlargement and repeated seeding.

*Methods Enzymol.* **114**:132-135.

Tinocco, I., Sauer, K., Wang, J. C., Puglisi, J. D., (2003) Physical Chemistry. Principles and Applications in Biological Sciences 4<sup>th</sup> ed. Prentice Hall, New Jersey, pp 505-508.

Tong, L., (1993) *REPLACE*, a suite of computer programs for molecular-replacement calculations. *J. Appl. Cryst.* **26**:748-751

Toogood, H. S., van Thiel, A., Basran, J., Sutcliffe, M. J., Scrutton, N. S., Leys, D., (2004) Extensive domain motion and electron transfer in the human electron transferring flavoprotein.medium chain Acyl-CoA dehydrogenase complex. *J Biol Chem* **279**:32904-32912.

Toogood, H. S., van Thiel, A., Scrutton, N. S., Leys, D., (2005) Stabilization of non-productive conformations underpins rapid electron transfer to electron-transferring flavoprotein. *J Biol Chem* **280**:30361-30366.

Tronrud, D. E., (1992) Conjugate-direction minimization: an improved method for the refinement of macromolecules. *Acta Crystallogr A* **48**:912-916.

Tsou, C. L., (1998) The role of active site flexibility in enzyme catalysis. *Biochemistry (Mosc)* **63**:253-258.

Tucker, E. E., Lane, E. H., Christian, S. D., (1981) Vapor pressure studies of hydrophobic interactions. Formation of benzene-benzene and cyclohexane-cyclohexanol dimers in dilute aqueous solutions. *J. Solut. Chem.* **10**:1-20.

Ueyama, N., Nishikawa, N., Yamada, Y., Okamura, T., Nakamura, A., (1996) Cytochrome P-450 Model (Porphinato)(thiolato)iron(III) Complexes with Single and Double NH-S Hydrogen Bonds at the Thiolate Site. *J. Am. Chem. Soc.* **118**:12826-12827.

Utkin, I., Dalton, D. D., Wiegel, J., (1995) Specificity of reductive dehalogenation of substituted ortho-chlorophenols by Desulfitobacterium dehalogenans JW/IU-DC1. *Appl. Environ. Microbiol.* 1995 **61**:346-351. Erratum in: *Appl Environ Microbiol* **61**:1677.

Utkin, I., Woese, C., Wiegel, J., (1994) Isolation and characterization of Desulfitobacterium dehalogenans gen. nov., sp. nov., an anaerobic bacterium which reductively dechlorinates chlorophenolic compounds. *Int. J. Syst. Bacteriol.* **44**:612-619.

van Aalten, D. M., DiRusso, C. C., Knudsen, J., Wierenga, R. K., (2000) Crystal structure of FadR, a fatty acid-responsive transcription factor with a novel acyl coenzyme A-binding fold. *EMBO J* **19**:5167-5177.

van de Pas, B. A., Smidt, H., Hagen, W. R., van der Oost, J., Schraa, G., Stams, A. J., de Vos, W. M., (1999) Purification and molecular characterization of ortho-chlorophenol reductive dehalogenase, a key enzyme of halorespiration in Desulfitobacterium dehalogenans. *J Biol Chem* **274**:20287-20292.

Van Dort, H. M., Bedard, D. L., (1991) Reductive ortho and meta dechlorination of a polychlorinated biphenyl congener by anaerobic microorganisms. *Appl. Environ. Microbiol.* **57**:1576-1578.

- van Pee, K. H., Unversucht, S., (2003) Biological dehalogenation and halogenation reactions. *Chemosphere* **52**:299-312.
- Vega, Y., Rauch, M., Banfield, M. J., Ermolaeva, S., Scotti, M., Goebel, W., Vazquez-Boland, J. A., (2004) New Listeria monocytogenes prfA\* mutants, transcriptional properties of PrfA\* proteins and structure-function of the virulence regulator PrfA. *Mol Microbiol* **52**:1553-1565.
- Villemur, R., Saucier, M., Gauthier, A., Beaudet, R., (2002) Occurrence of several genes encoding putative reductive dehalogenases in Desulfobacterium hafniense/frappieri and Dehalococcoides ethenogenes. *Can. J. Microbiol.* **48**:697-706. Erratum in: *Can J Microbiol.* **48**:945.
- Vinson, C. R., Sigler, P. B., McKnight, S. L., (1989) Scissors-grip model for DNA recognition by a family of leucine zipper proteins. *Science* **246**:911-916.
- Vinson, C., Myakishev, M., Acharya, A., Mir, A. A., Moll, J. R., Bonovich, M., (2002) Classification of human B-ZIP proteins based on dimerization properties. *Mol Cell Biol* **22**:6321-6335.
- Vitali, J., Robbins, A.H., Almo, S.C., Tilton, R.F., (1991) Using xenon as a heavy atom for determining phases in sperm whale met-myoglobin. *J. Appl. Cryst.* **24**:931-935.
- Vogel, T. M., Criddle, C. S., McCarty, K. S., (1987) Transformations of halogenated aliphatic compounds. *Environ.Sci.Techol.* **21**:722-736.
- Vogelstein, B., Gillespie, D., (1979) Preparative and analytical purification of DNA from agarose. *Proc Natl Acad Sci USA* **76**:615-619.
- Wah, D. A., Hirsch, J. A., Dorner, L. F., Schildkraut, I., Aggarwal, A. K., (1997) Structure of the multimodular endonuclease FokI bound to DNA. *Nature* **388**:97-100.
- Waibel, M., Schulze, H., Huber, N., Bachmann, T. T., (2005) Screen-printed bienzymatic sensor based on sol-gel immobilized Nippostrongylusbrasiliensis acetylcholinesterase and a cytochrome P450 BM-3 (CYP102-A1) mutant. *Biosens Bioelectron*.
- Wallace, A. C., Laskowski, R. A., Thornton, J. M., (1995) LIGPLOT: a program to generate schematic diagrams of protein-ligand interactions. *Protein Eng* **8**:127-134.
- Walsh, M. A., Schneider, T. R., Sieker, L. C., Dauter, Z., Lamzin, V. S., Wilson, K. S., (1998) Refinement of triclinic hen egg-white lysozyme at atomic resolution. *Acta Crystallogr D Biol Crystallogr* **54**:522-546.
- Wang, B.C. (1985). Resolution of phase ambiguity in macromolecular crystallography. *Meth Enzymol.* **114**:187-196.
- Wang, M., Roberts, D. L., Paschke, R., Shea, T. M., Masters, B. S., Kim, J. J., (1997) Three-dimensional structure of NADPH-cytochrome P450 reductase: prototype for FMN- and FAD-containing enzymes. *Proc Natl Acad Sci USA* **94**:8411-8416.
- Weeks, C. M., DeTitta, G. T., Miller, R., Hauptman, H. A., (1993) Applications of the minimum principle to peptide structures. *Acta Cryst. D* **49**:179-181.

Westbrook, E. M., Naday, I., (1997). Charge-coupled device-based area detectors. *Meth. Enzymol.* **276**:244-268.

Wester, M. R., Johnson, E. F., Marques-Soares, C., Dansette, P. M., Mansuy, D., Stout, C. D., (2003) Structure of a substrate complex of mammalian cytochrome P450 2C5 at 2.3 Å resolution: evidence for multiple substrate binding modes. *Biochemistry* **42**:6370-6379.

Wester, M. R., Johnson, E. F., Marques-Soares, C., Dijols, S., Dansette, P. M., Mansuy, D., Stout, C. D., (2003) Structure of mammalian cytochrome P450 2C5 complexed with diclofenac at 2.1 Å resolution: evidence for an induced fit model of substrate binding. *Biochemistry* **42**:9335-9345.

Williams, P. A., Cosme, J., Ward, A., Angove, H. C., Mata Vinkovic, D., Jhoti, H., (2003) Crystal structure of human cytochrome P450 2C9 with bound warfarin. *Nature* **424**:464-468.

Wilson, A. J. C., Prince, E., (1999) International Tables for Crystallography Volume C: Mathematical, physical and chemical tables, 2<sup>nd</sup> ed., Kluwer Academic Publishers, Dordrecht.

Wolthers, K. R., Basran, J., Munro, A. W., Scrutton, N. S., (2003) Molecular dissection of human methionine synthase reductase: determination of the flavin redox potentials in full-length enzyme and isolated flavin-binding domains. *Biochemistry* **42**:3911-3920.

Woolfson, M. M. (1970) An Introduction to X-ray Crystallography. Cambridge University Press.

Xiang, H., Tscherret-Guth, R. A., Ortiz De Montellano, P. R., (2000) An A245T mutation conveys on cytochrome P450eryF the ability to oxidize alternative substrates. *J Biol Chem* **275**:35999-36006.

Xiao, K. H., Yu, L., Yu, C. A., (2000) Confirmation of the involvement of protein domain movement during the catalytic cycle of the cytochrome *bc*<sub>1</sub> complex by the formation of an intersubunit disulfide bond between cytochrome *b* and the iron-sulphur protein. *J. Biol. Chem.* **275**, 38597-38604.

Yano, J. K., Hsu, M. H., Griffin, K. J., Stout, C. D., Johnson, E. F., (2005) Structures of human microsomal cytochrome P450 2A6 complexed with coumarin and methoxsalen. *Nat Struct Mol Biol* **12**:822-823.

Yano, J. K., Koo, L. S., Schuller, D. J., Li, H., Ortiz de Montellano, P. R., Poulos, T. L., (2000) Crystal structure of a thermophilic cytochrome P450 from the archaeon *Sulfolobus solfataricus*. *J Biol Chem* **275**:31086-31092.

Yoshioka, S., Tosha, T., Takahashi, S., Ishimori, K., Hori, H., Morishima, I., (2002) Roles of the proximal hydrogen bonding network in cytochrome P450cam-catalyzed oxygenation. *J Am Chem Soc* **124**:14571-14579.

Yousef, M. S., Baase, W. A., Matthews, B. W., (2004) Use of sequence duplication to engineer a ligand-triggered, long-distance molecular switch in T4 lysozyme. *Proc Natl Acad Sci U S A* **101**:11583-11586.

Zerbe, K., Pylypenko, O., Vitali, F., Zhang, W., Rouset, S., Heck, M., Vrijbloed, J. W., Bischoff, D., Bister, B., Sussmuth, R. D., Pelzer, S., Wohlleben, W., Robinson, J. A., Schlichting, I., (2002) Crystal structure of OxyB, a cytochrome P450 implicated in an

oxidative phenol coupling reaction during vancomycin biosynthesis. *J Biol Chem* **277**:47476-47485.

Zhang, J., Martasek, P., Paschke, R., Shea, T., Siler Masters, B. S., Kim, J. J., (2001) Crystal structure of the FAD/NADPH-binding domain of rat neuronal nitric-oxide synthase. Comparisons with NADPH-cytochrome P450 oxidoreductase. *J Biol Chem* **276**:37506-37513.

Zhang, K. Y. J., Main, P., (1990) The use of Sayre's equation with solvent flattening and histogram matching for phase extension and refinement of protein structures. *Acta Cryst. A* **46**:377-381.

Zumdahl S. S., Zumdahl, S. L., (2003) Chemical Principles, 6th Ed., Houghton-Mifflin, Boston.

Zumft, W. G., (1997) Cell biology and molecular basis of denitrification. *Microbiol Mol Biol Rev.* **61**:533-616.

## **Publications**

# A Single Mutation in Cytochrome P450 BM3 Induces the Conformational Rearrangement Seen upon Substrate Binding in the Wild-type Enzyme\*

Received for publication, February 17, 2004, and in revised form, March 12, 2004  
Published, JBC Papers in Press, March 12, 2004, DOI 10.1074/jbc.M401717200

M. Gordon Joyce†, Hazel M. Girvan§, Andrew W. Munro¶, and David Ley¶||

From the Department of Biochemistry, University of Leicester, Leicester LE1 7RH, United Kingdom

The multidomain fatty-acid hydroxylase flavocytochrome P450 BM3 has been studied as a paradigm model for eukaryotic microsomal P450 enzymes because of its homology to eukaryotic family 4 P450 enzymes and its use of a eukaryotic-like diflavin reductase redox partner. High-resolution crystal structures have led to the proposal that substrate-induced conformational changes lead to removal of water as the sixth ligand to the heme iron. Concomitant changes in the heme iron spin state and heme iron reduction potential help to trigger electron transfer from the reductase and to initiate catalysis. Surprisingly, the crystal structure of the substrate-free A264E heme domain mutant reveals the enzyme to be in the conformation observed for substrate-bound wild-type P450, but with the iron in the low-spin state. This provides strong evidence that the spin-state shift observed upon substrate binding in wild-type P450 BM3 not only is caused indirectly by structural changes in the protein, but is a direct consequence of the presence of the substrate itself, similar to what has been observed for P450cam. The crystal structure of the palmitoleate-bound A264E mutant reveals that substrate binding promotes heme ligation by Glu<sup>264</sup>, with little other difference from the palmitoleate-bound wild-type structure observable. Despite having a protein-derived sixth heme ligand in the substrate-bound form, the A264E mutant is catalytically active, providing further indication for structural rearrangement of the active site upon reduction of the heme iron, including displacement of the glutamate ligand to allow binding of dioxygen.

Cytochrome P450 enzymes are among the most studied enzymes, in no small part because of the pivotal roles that hepatic

P450 enzymes play in mammalian drug metabolism (1). Recent years have seen an explosion in the structural data available for these systems, a substantial proportion of which is on cytochrome P450 BM3 (2). This multidomain enzyme is isolated from *Bacillus megaterium* and contains an N-terminal fatty acid-binding P450 domain fused to its redox partner, an NADPH-dependent diflavin-cytochrome P450 reductase (3). It has been used as the paradigm model for studying the similar, but membrane-associated eukaryotic microsomal P450 systems. This has been primarily because its heme domain is similar to the eukaryotic fatty-acid hydroxylases from P450 family 4, because the enzyme is soluble and uses a eukaryote-like cytochrome P450 reductase as the redox partner (as opposed to the two-component ferredoxin reductase and ferredoxin systems found in many other bacterial systems), and because it is a convenient catalytically self-sufficient fusion protein enzyme (2–4). Cytochrome P450 BM3 is a fatty-acid hydroxylase that displays an unusually high rate of oxygenation of long-chain fatty acids (e.g. >15,000 turnovers/min with arachidonic acid) (5), likely because of the efficient electron transfer between the different redox modules afforded by their covalent linkage and hence close spatial organization (2). A sophisticated mechanism to avoid the unwanted generation of reactive oxygen species through futile cycling has been found in many P450 enzymes studied to date. The binding of oxygen occurs only with the reduced (ferrous) heme, and the reduction of the ferric heme iron by electron transfer from the redox partner is, in turn, dependent on the binding of substrate, effectively gating initiation of the reaction by substrate binding (6, 7). In P450 BM3 and the *Pseudomonas putida* camphor hydroxylase P450cam (the most intensively studied P450 enzyme), substrate binding induces a heme iron spin-state shift and a concomitant increase in the reduction potential of the heme iron, favoring the 1-electron reduction that commits the enzyme to the catalytic cycle (6, 7). The molecular mechanism whereby substrate binding induces this shift seems to be somewhat different in the P450 enzymes studied to date, although the substrate binding-induced displacement of water as the sixth ligand to the heme iron is a common feature (see Refs. 8 and 9). In P450cam, the binding of substrate does not effect any large-scale changes in the protein structure, and the displacement of water is due to direct steric hindrance with the camphor molecule (10). In contrast, P450 BM3 undergoes large-scale conformational changes upon binding of fatty acids, and these changes have been proposed to drive the conversion of a six- to a five-coordinate heme group. Indeed, there is no direct interaction observed between bound fatty acids and the water molecule in the sixth ligand position in the available crystal structures of the P450 BM3 heme domain. The conformational change in P450 BM3 involves a reorganization of the I helix, and it has been proposed that this creates a new water binding position (11).

\* This work was supported in part by the European Community Access to Research Infrastructure Action of the Improving Human Potential Program to the European Molecular Biology Laboratory Hamburg Outstation under Contract HPRI-CT-1999-00017. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

† The atomic coordinates and structure factors (codes 1SMI and 1SMJ) have been deposited in the Protein Data Bank, Research Collaboratory for Structural Bioinformatics, Rutgers University, New Brunswick, NJ (<http://www.rcsb.org/>).

‡ Supported by the Biotechnology and Biological Sciences Research Council (BBSRC).

§ Supported by the Engineering and Physical Sciences Research Council (EPSRC).

¶ Supported by the BBSRC and the EPSRC.

|| Royal Society University Research Fellow. To whom correspondence should be addressed: Dept. of Biochemistry, University of Leicester, Adrian Bldg., University Rd., Leicester LE1 7RH, UK. Tel.: 44-116-252-3484; Fax: 44-116-252-3369; E-mail: dl37@le.ac.uk.

## Crystal Structure of Cytochrome P450 BM3 A264E

TABLE I  
Crystallographic statistics

Data collection	Protein	
	A264E mutant of BM3	A264E mutant with substrate bound
Space group	P2 <sub>1</sub> 2 <sub>1</sub> 2 <sub>1</sub>	C222 <sub>1</sub>
Cell parameters (Å)		
<i>a</i>	61.2	106.5
<i>b</i>	118.9	165.7
<i>c</i>	146.3	224.7
X-ray source	ESRF ID14-EH1	ESRF ID14-EH1
Resolution	15.0–2.0	15.0–2.75
No. of observations		
Total	686,004	787,601
Unique	68,550	53,600
Completeness (%)	94.0 (97.1)	99.9 (99.6)
<i>R</i> <sub>merge</sub>	5.1 (44.1)	6.9 (49.1)
<i>I</i> / <i>I</i> <sub>0</sub>	28.2 (2.41)	30.9 (3.7)
Refinement		
<i>R</i> <sub>work</sub>	0.227	0.257
<i>R</i> <sub>free</sub>	0.288	0.338
r.m.s. deviations from ideal		
Bonds (Å)	0.018	0.034
Angle	1.719°	2.933°

<sup>a</sup> ESRF, European Synchrotron Radiation Facility; r.m.s., root mean square.

This position is mutually exclusive with the sixth ligand binding position and is proposed to have greater affinity, leading to an effective switch in heme coordination by transfer of the water to the new position. The majority of the P450 structures indeed show a conserved bend in the I helix, and this mechanism might therefore be of a general nature (see Ref. 12). However, certain P450 enzymes do not contain any deformation of the I helix in the resting state (see Ref. 13), whereas P450cam has a bent I helix that does not significantly change conformation upon substrate binding (14).

We present here the crystal structures of both the substrate-free and palmitoleic acid-bound forms of the A264E mutant in the P450 BM3 heme domain. The alanine occupies a key position in the I helix of the P450 enzyme. The backbone of this residue hydrogen bonds to the sixth ligand water molecule in the unbound form and is substantially shifted upon substrate binding (8, 15). Interestingly, in several of the CYP4 family of fatty-acid oxygenase P450 enzymes, a conserved glutamate residue at position 264 (BM3 numbering) is known to covalently ligate the heme macrocycle through autocatalytic turnover-dependent attachment to the 5-methyl group of the porphyrin (16). To investigate the possibility of creating a similar protein-heme link in the related P450 BM3 heme domain, the A264E mutant was created. Although no covalent modification of the heme was observed, this mutant has several unique features (see accompanying article (17)). In the fatty acid-free form, Glu<sup>264</sup> ligates the heme iron in a proportion of the molecules, creating a novel thiolate-carboxylate ligation that is pushed toward full ligation by binding of the substrate. Surprisingly, the crystal structure of the substrate-free form of this mutant reveals the protein to be in the conformation previously considered to be induced by substrate binding, despite the fact that the P450 enzyme is still in a low-spin state and free of fatty acid. We show that binding of fatty acid does not then introduce any further gross conformational change in the protein structure, although a change in the proportion of molecules in which glutamate ligates the heme iron is effected. The implications of these observations for the mechanism of the substrate binding-induced heme iron spin-state shift of P450 BM3 and P450 enzymes in general are discussed, along with the ramifications for understanding conformational equilibria in P450 enzymes and how the binding of substrate impacts on these equilibria.

## EXPERIMENTAL PROCEDURES

**Mutagenesis, Expression, and Purification of Mutant P450 BM3—**The A264E mutant of the P450 BM3 heme domain was created, ex-

pressed, and purified as described in the accompanying article (17). A final purification step by fast protein liquid chromatography using Q-Sepharose resin (under the same conditions as used for the low-pressure chromatographic purification described in the accompanying article (17)) was used to produce homogeneous enzyme for crystallographic studies.

**Crystallization**—The P450 BM3 A264E heme domain was crystallized in both the palmitoleic acid-bound and substrate-free forms by the sitting drop method at 4 °C. Sitting drops were prepared by adding 2 μl of mother liquor to 2 μl of 15 mg/ml enzyme. Palmitoleic acid-bound crystals were obtained by co-crystallizing the enzyme with 1.2 μM palmitoleic acid (~6 times the *K*<sub>d</sub> value obtained from spectral binding titration data) and a well solution of 100 mM magnesium acetate, 20% polyethylene glycol 2000MME, and 100 mM cacodylic acid at pH 6.3. Substrate-free crystals were obtained with a well solution of 10 mM manganese sulfate, 20% polyethylene glycol 2000MME, and 100 mM cacodylic acid at pH 6.3. Crystals were formed in both cases after ~7 days. Crystals were immersed in 10% polyethylene glycol 200 in mother liquor as a cryoprotectant, before being mounted on a nylon loop and flash-cooled in liquid nitrogen.

**Data Collection, Structure Elucidation, and Refinement**—The data used for refinement were collected at the European Synchrotron Radiation Facility (Grenoble, France) on ID14-EH1 using an ADSC Q4R CCD detector. Crystals were cooled at 100 K, and diffraction data were collected in oscillations of 1°. Data were processed and scaled using the HKL package programs DENZO and SCALEPACK (18). The substrate-free crystal structures were solved via molecular replacement using the program AMoRE and the high-resolution wild-type P450 BM3 crystal structure (Protein Data Bank code 2HPD) (19) as a search model. The palmitoleate-bound wild-type crystal structure was used a starting model for the palmitoleate-bound mutant form. In both cases, positional and *B*-factor refinement was carried out using REFMAC5 with manual rebuilding of the model at regular intervals in TURBO-FRODO (20, 21). Only in the case of the low-resolution substrate-bound crystal form were strong NCS restraints imposed throughout refinement. Data collection and final refinement statistics are given in Table I. The atomic coordinates and structure factors for both crystal structures have been deposited in the Protein Data Bank with codes 1SMI (substrate-free A264E mutant) and 1SMJ (substrate-bound A264E mutant).

## RESULTS AND DISCUSSION

**Crystal Structure of the Substrate-free Form**—Initial crystallization trials for the substrate-free form of the A264E heme domain of P450 BM3 using the published conditions for the wild-type heme domain proved unsuccessful. The use of manganese sulfate in place of magnesium acetate led to related conditions that generated large diffraction-quality crystals. The structure was solved to 2.0 Å and contained two molecules in the asymmetric unit. The overall conformation of both molecules is similar (root mean square deviation of 0.46 Å for all C-α atoms), with molecule B having a significantly higher

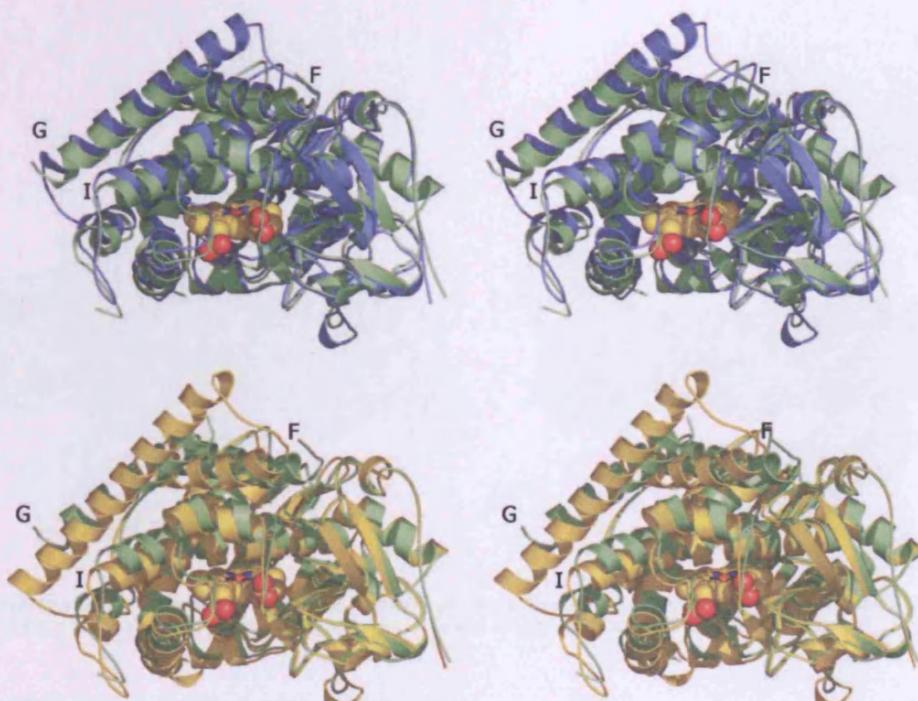


FIG. 1. Stereo view of an overlay of the substrate-free A264E heme domain (green ribbons). Upper, substrate-free form (yellow) of the wild-type P450 BM3 heme domain; lower, substrate-bound form (blue) of the wild-type P450 BM3 heme domain.

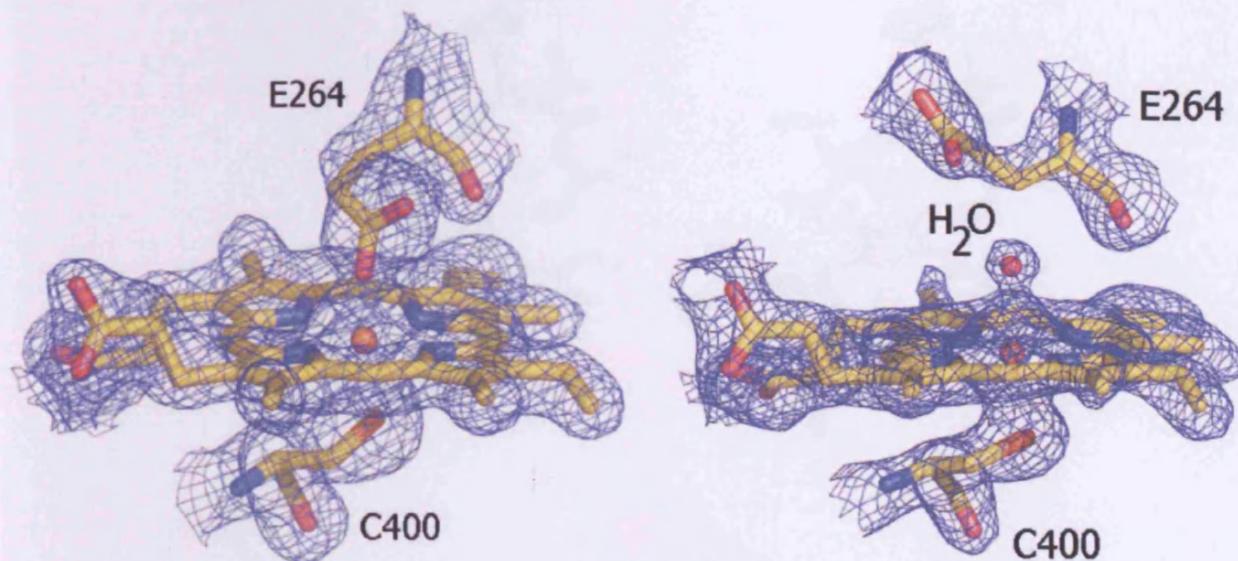
average *B*-factor ( $39.1 \text{ \AA}^2$  compared with  $47.6 \text{ \AA}^2$ ) due to fewer packing constraints. In this study, molecule A will be used for discussion and calculations unless mentioned otherwise.

Because of the different space groups of the orthorhombic A264E mutant and the wild-type monoclinic crystal form, comparison of the structures was made after overlay using the structurally invariant residues, as described by Haines *et al.* (11), representing ~62% of the structure. Surprisingly, it was found that the A264E structure is structurally similar (root mean square deviation of  $0.54 \text{ \AA}$  for all C- $\alpha$  atoms) to the substrate-bound form of the wild-type enzyme (hereafter referred to as the SB<sup>1</sup> conformation) and shows a significant difference (root mean square deviation  $1.37 \text{ \AA}$  for all C- $\alpha$  atoms) from the substrate-free wild-type structure (hereafter referred to as the SF conformation) (Fig. 1, upper and lower) (8, 15). However, no substrate was added to the A264E mutant during either purification or crystallization, and no substrate could be observed in the electron density maps. Similar to the changes seen upon substrate binding in wild-type P450 BM3, the majority of residues that are in significantly different positions between the substrate-free structures of A264E and the wild-type heme domain are located in the "lid domain" of the substrate access channel, which consists of the F and G helices, the loop between them, and the B' helix (8). Several of these residues are less well defined in A264E molecule B, indicating substantial plasticity of this region in the absence of substrate. However, there is no significant large-scale difference between both A264E molecules in the asymmetric unit.

In contrast, upon closer inspection of the active site of the A264E enzyme, there is a marked difference in the vicinity of the heme iron between the two molecules in the asymmetric unit. Molecule B has the side chain of Glu<sup>264</sup> pointing away from the heme, its carboxylate group stacking with the aromatic group of Phe<sup>87</sup> (Fig. 2, left). In contrast, molecule A has the carboxylate group coordinating the heme iron (Fig. 2, right). This heterogeneity of glutamate ligation was predicted from solution spectrophotometric studies and from EPR analysis (see accompanying article (17)), and the thiolate-carboxylate ligation is novel to P450 enzymes and (as far as we are aware to date) to cytochromes in general. There do not seem to be any significant conformational changes to the overall P450 structure associated with the "switch" of Glu<sup>264</sup> between its two detected conformations on (a) the heme iron or (b) stacking with Phe<sup>87</sup>. It is likely that, in solution, the carboxylate continually switches between the "on" and "off" heme iron states without major accompanying protein reorganization. According to spectroscopic data, in the solution state, the equilibrium is poised at ~3–4:1 in favor of the heme iron ligand off form (17). The interaction with Phe<sup>87</sup> is particularly interesting, given the fact that this residue is absent from the CYP4 enzymes in which covalent ligation of the heme methyl group has been demonstrated. Phe<sup>87</sup> interacts with the  $\omega$ -methyl group of fatty acid substrate(s) in wild-type P450 BM3 and is considered to be a critical regulatory residue that controls regioselectivity of substrate oxygenation (8, 22, 23). A particular difference in the behavior of P450 BM3 with respect to eukaryotic CYP4 enzymes is the inability of the former to hydroxylate at the  $\omega$ -position (24).

Close examination of the A264E structure reveals no clear direct structural explanation for the fact that this mutant mimics the conformation of the substrate-bound form of the wild-type enzyme. In particular, no extra stabilizing features involving the newly introduced glutamate side chain can be found in comparison with available wild-type structures. We therefore tried to place the Glu<sup>264</sup> side chain in the corresponding substrate-free wild-type structure. All of the conformations available to the glutamate result in severe steric clashes with several other residues nearby (e.g. Phe<sup>87</sup>, Thr<sup>260</sup>, and Ile<sup>263</sup>, in

<sup>1</sup> The abbreviations used are: SB, substrate-bound; SF, substrate-free.

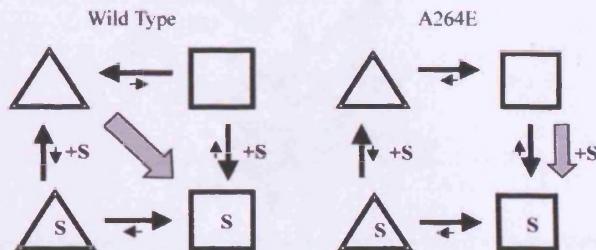


**FIG. 2. Conformations of Glu<sup>264</sup> in molecules A and B of the substrate-free A264E heme domain of flavocytochrome P450 BM3.** The heme macrocycle along with Cys<sup>400</sup> and Glu<sup>264</sup> is displayed for molecule B (left) and molecule A (right) with the  $2F_o - F_c$  map contoured at  $1\sigma$  superposed.

addition to the heme macrocycle itself), clearly resulting in the need for a protein conformational rearrangement to accommodate for the increased bulk of residue 264. It therefore seems likely that the A264E mutation does not particularly stabilize the SB conformation, but rather destabilizes the SF conformation of the enzyme to the extent that the SB conformation, even in the absence of substrate, is preferred.

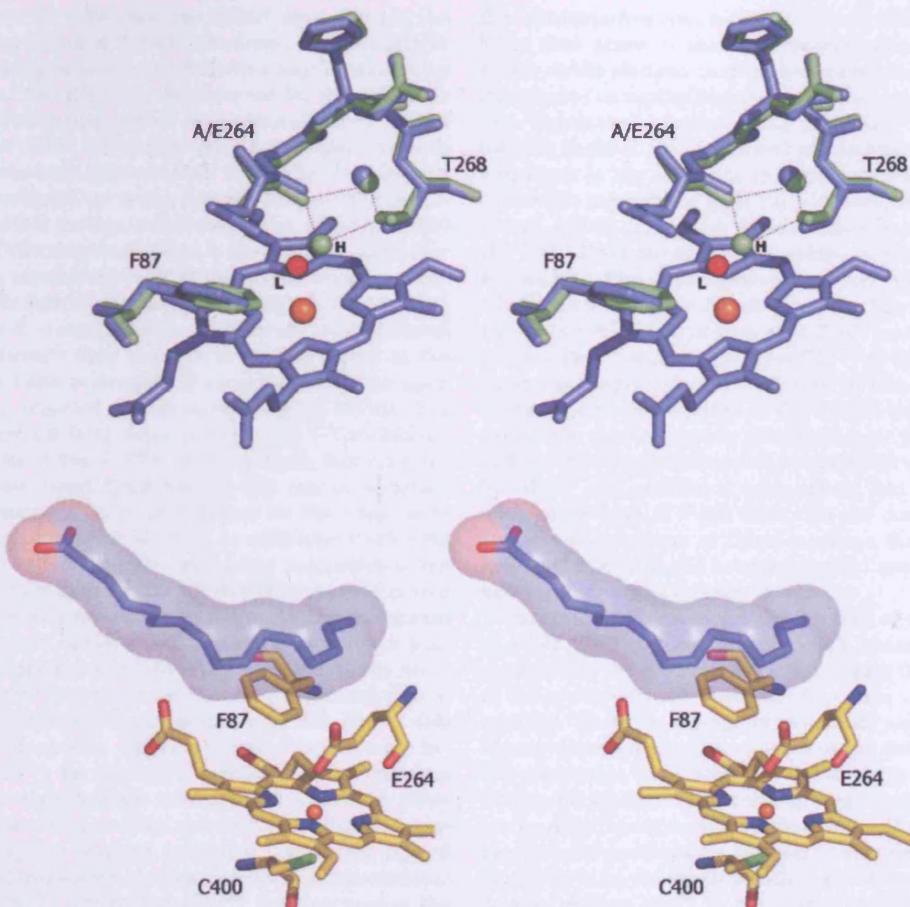
In light of the above observations, it is interesting to note that, in the substrate-bound wild-type P450 BM3 structures, the substrates are in close contact with the side chain of Ala<sup>264</sup>. This suggests that the mechanism by which substrate binding switches P450 BM3 to a different conformational state does not involve simply the expulsion of water molecules from the substrate-binding cavity, but in addition exploits the force exerted by the substrate on Ala<sup>264</sup>, a residue that acts as a sensitive trigger for the conformational conversion. In the A264E mutant, unfavorable steric interactions of the glutamate side chain induce the switch to the SB conformation, without necessity for the substrate interaction. An alternative explanation that could be put forward is that the P450 BM3 heme domain is in rapid conformational equilibrium between the SF and SB conformations, with the equilibrium being strongly favored toward the SF conformation for wild-type P450 BM3 in the absence of substrate. In this model, the substrate binds preferentially to the SB conformation, effectively shifting the equilibrium toward this form as the substrate concentration is increased. This model is also consistent with the behavior observed for the A264E heme domain. For several substrates tested, the apparent binding constants ( $K_d$ ) determined are considerably lower than those for the wild-type heme domain, indicating much tighter binding (17). Given that the SB conformation is favored in the fatty acid-free form of the A264E heme domain, the “tight-binding” form of the protein is overrepresented in solution with respect to that seen for wild-type P450 BM3 (Fig. 3).

A substrate binding-induced spin-state shift has been observed in a large number of P450 enzymes studied to date (see Refs. 25 and 26). It is generally accepted that this behavior serves to avoid the potentially dangerous generation of active oxygen species that would occur through binding and subsequent



**FIG. 3. Scheme representing a model for the conformational equilibrium of P450 BM3 in solution and the influence of substrate binding.** Left, wild-type equilibrium; right, A264E equilibrium. The equilibrium in the absence of substrate is drastically changed by introduction of the A264E mutation. In both cases, the large gray arrows indicate the apparent shift upon substrate binding.

reduction of molecular oxygen in the absence of substrate. Spin-state change induces a positive change in the heme iron reduction potential, favoring electron transfer from the redox partner (6, 7). The molecular mechanism underlying this change in the heme iron spin state has been proposed to be either a direct displacement of the sixth ligand water molecule (as in P450cam) or a more indirect displacement via substrate-induced changes in the protein structure (as in P450 BM3) (11). We have shown that the fatty acid-free A264E structure does show all the structural hallmarks of a substrate-bound P450 BM3 enzyme, but in the absence of substrate. An intriguing difference regarding A264E is the fact that, in contrast to the substrate-bound wild-type enzyme, A264E does not show any significant high-spin character in either the solution or crystalline state. A low-spin configuration would be expected for the glutamate-ligated species, but both spectroscopic and crystallographic studies show that there is a large population of non-glutamate-ligated protein in the substrate-free A264E heme domain and that this species does not lose the water ligand and convert to the high-spin form despite the change to the SB conformation. This suggests strongly that, even in wild-type P450 BM3, the spin-state shift is a direct (rather than an indirect) result of substrate binding. The high-resolution *N*-palmitoylglycine-bound wild-type structure has led to the sug-



**FIG. 4.** Stereo view of the active site of A264E. *Upper*, overlay of the active site of A264E (blue) and *N*-palmitoylglycine-bound wild-type P450 BM3 (green) (Protein Data Bank code 1JPZ). For clarity, the heme macrocycle is displayed only for the A264E mutant. The pattern of hydrogen bonding to site H is indicated by the dotted lines. The ligating water molecule occupying site L in the A264E mutant is colored red. *Lower*, active-site structure of the palmitoleic acid-bound form of the A264E heme domain. Residues are colored according to residue type; the substrate is depicted in purple.

gestion that substrate binding-induced protein rearrangement creates a new water-binding site (designated site H) adjacent to the heme (11). The proposed higher water affinity of site H over the heme iron ligation site (designated site L) and the fact that these sites are mutually exclusive (so that only a single water molecule can bind at either site L or H at any given time) led to the proposal that site H effectively pulls the water molecule away from site L, leading to the observed shift in the heme iron coordination state and hence conversion to the high-spin form (11). The A264E heme domain structure shows all ligands to site H to be in identical positions compared with the *N*-palmitoylglycine-bound wild-type P450 BM3 structure (Fig. 4 *upper*). However, the water molecule still occupies site L (the heme iron) for molecule B, whereas in molecule A, Glu<sup>264</sup> ligates the heme iron. In both cases, site H remains unoccupied due to steric hindrance with either the water (in molecule B) or the Glu<sup>264</sup> side chain (in molecule A) at site L. It can therefore be concluded that site L (heme iron) remains the higher affinity site for water in the absence of substrate in the SB conformation.

As has been pointed out previously (11), there are no direct steric clashes between the bound fatty acid substrates of P450 BM3 and the water molecule at site L. The palmitoleic acid-bound structure of wild-type P450 BM3 indicates that secondary conformational changes in the protein/substrate must occur following reduction of the heme iron because the substrate

is too distant from the heme iron for oxidative attack at catalytically relevant positions on the fatty acid chain (8). NMR studies of the substrate-bound form of ferrous, fatty acid-bound wild-type P450 BM3 are consistent with a significant reorientation of the substrate in this enzyme form (27). It is clear, however, that, upon binding of the fatty acid analog *N*-palmitoylglycine, the surroundings of site L become more hydrophobic, decreasing the water affinity and ultimately shifting the water molecule to predominantly occupy site H. Spectroscopic studies of wild-type P450 BM3 at catalytically relevant temperatures (by both electronic absorption and resonance Raman) have shown that, even in the presence of apparently saturating concentrations of substrate, an equilibrium exists between the high-spin five-coordinate and the low-spin six-coordinate states of the heme iron. Depending on the nature of the fatty acid substrate used, varying amounts of low-spin heme iron are detected, with shorter chain saturated fatty acids (*e.g.* lauric acid) being less effective than longer chain ones (*e.g.* palmitic acid) at effecting the shift in the spin-state equilibrium toward the high-spin state (28, 29). This clearly indicates how, even in the presence of substrate, the possibility exists for water remaining bound at the heme iron (site L), and the water affinity of this site is strongly dependent on the nature of the substrate, although all drive the L-H equilibrium toward site H.

*Crystal Structure of the Substrate-bound Form*—In contrast to the requirement for screening novel conditions to obtain

suitable crystals of the substrate-free A264E heme domain, the palmitoleate-bound form of A264E was found to crystallize in the same space group as that reported previously for substrate-bound wild-type P450 BM3 (8). As observed for the wild-type enzyme, the resolution and quality of the data obtained for this particular crystal form are rather poor by comparison with those for the substrate-free enzyme. Nevertheless, electron density clearly indicates no major changes between the palmitoleate-bound A264E and palmitoleate-bound wild-type P450 BM3 structures. The single exception is the fact that, in all four molecules in the asymmetric unit of the A264E structure, the Glu<sup>264</sup> side chain ligates the heme iron (Fig. 4, *lower*). This finding indicates that the presence of substrate induces movement of the glutamate onto the iron to replace water as the sixth heme ligand and is completely consistent with the spectroscopic studies reported in the accompanying article (17). Addition of long-chain fatty acids perturbs the UV-visible absorption spectrum of the A264E heme domain, inducing red shifts of the heme Soret band toward 426 nm at apparent substrate saturation. This is in contrast to the blue shifts observed following substrate addition to wild-type P450 BM3 (toward 390 nm) and reflects the increasing proportion of the glutamate-ligated low-spin form of the A264E mutant following substrate addition as opposed to the high-spin five-coordinate form of the wild-type enzyme that accumulates through substrate-induced displacement of the water ligand from the heme iron (17, 28). From the substrate-free A264E structure, it is clear that two conformations are possible for the Glu<sup>264</sup> side chain, and solution studies indicate the equilibrium ratio between both states to be strongly dependent on the solution conditions. Upon palmitoleate binding, the substrate effectively occupies the volume of the non-heme iron-ligating conformation, driving the enzyme toward a completely ligated state, as observed in solution studies. Specifically, palmitoleate interacts with Phe<sup>87</sup>, preventing Glu<sup>264</sup> from occupying the position observed in molecule A in the substrate-free A264E structure. Despite any significant further structural rearrangement induced following palmitoleic acid binding to the SB conformation of A264E, the substrate does influence directly the heme iron ligation state by minimizing the degrees of freedom available to the Glu<sup>264</sup> side chain. Strong heme ligands such as azoles function as potent inhibitors for P450 enzymes, and many are used as antifungal drugs to inactivate the sterol demethylase P450 CYP51 (30). It is therefore surprising to note that, despite the fact that the substrate-bound oxidized A264E structure shows fully six-coordinate heme iron, catalytic turnover for this mutant can still be observed, albeit at lower levels than those observed for wild-type P450 BM3 (17). This is a further indication that, upon reduction of ferric to ferrous iron, the position of the substrate with respect to the heme and perhaps the structure of the enzyme itself change dramatically, releasing the strong conformational lock on the glutamate side chain and allowing oxygen to bind to the iron. Presumably, this commits the enzyme to its "regular" catalytic cycle and prevents coordination of the glutamate to the iron until it returns to a ferric form following product formation. A further interesting aspect of this study arises from the structural change and its effect on the thermodynamic properties of the P450. In both the substrate-free and arachidonate-bound forms of the A264E mutant, the reduction potential of the heme iron is approximately -315 mV (see accompanying article (17)). By contrast, the reduction potential of the conformationally different substrate-free form of P450 BM3 is -427 mV, rising to -289 mV upon binding arachidonate and an extensive switch in the spin-state equilibrium toward the high-spin form (31). In A264E, the heme iron remains predominantly low-spin in both

the substrate-free and substrate-bound states. Thus, a possibility that arises is that the conformational change and its effects on the electronics of the heme system are of considerable importance in controlling the reduction potential of the heme iron. This is under further study using Ala<sup>264</sup> variants in which the side chain of the introduced amino acid does not ligate the heme iron in the substrate-free and substrate-bound forms.

Further scrutiny of both the substrate-free and substrate-bound A264E structures reveals other important features of the P450 BM3 structure that relate to attempts to engineer covalent ligation of the heme macrocycle via the interaction of Glu<sup>264</sup> with the heme 5-methyl group. The active-site organization in P450 BM3 is such that Phe<sup>87</sup> and likely the I helix residue Thr<sup>260</sup> obstruct access of Glu<sup>264</sup> to the relevant position on the porphyrin ring. The failure to obtain any significant degree of covalent ligation in the A264E mutant may thus be explicable through steric restrictions in the active site. To address these restrictions and to enable covalent heme ligation by Glu<sup>264</sup> and produce a more robust and biotechnologically exploitable form of P450 BM3, we are currently generating secondary mutations at these locations that might facilitate access of Glu<sup>264</sup> to the relevant methyl group and could thus allow autocatalytic linkage to occur.

**Conclusion**—Crystallographic studies of the A264E variant of P450 BM3 confirm the proposals based on spectroscopic studies that the glutamate is able to ligate the ferric heme iron of the mutant in the substrate-free form and that substrate addition "forces" on the ligand, producing a completely low-spin six-coordinate species, as opposed to the extensively high-spin five-coordinate form seen for the wild-type P450 BM3 (17). Structural studies explain clearly why substrate has this effect on the A264E enzyme since palmitoleic acid occupies one of the two favored positions for the Glu<sup>264</sup> side chain. Glu<sup>264</sup> can no longer form an interaction with the key regiospecificity-determining residue Phe<sup>87</sup> in the palmitoleate-bound form and is thus induced to move toward its only other acceptable position, coordinating to the heme iron.

An unexpected finding, but one with enormous ramifications for understanding the conformational changes that occur in P450 BM3 (and P450 systems in general) and their consequences, is the fact that both the substrate-free and palmitoleic acid-bound forms of the A264E heme domain have overall structural conformations that are virtually identical to those found for the substrate-bound forms of wild-type P450 BM3, but are distinct from that of the substrate-free wild-type heme domain (8, 11, 15). This SB conformation is not dependent on whether Glu<sup>264</sup> ligates the heme iron or is positioned against Phe<sup>87</sup>, and the enzyme is low-spin in both forms and water-coordinated in the latter form for substrate-free A264E. The most obvious explanations are 1) that the SB conformation in wild-type P450 BM3 is a consequence of substrate-induced deformation of the I helix in the region of Ala<sup>264</sup> and that the A264E mutation favors this conformational rearrangement independent of the substrate due to steric restrictions to movement of the glutamate side chain in the SF conformation, and/or 2) that P450 BM3 is in a continual dynamic equilibrium between the SF and SB conformations and that the A264E mutation forces this equilibrium toward the SB conformation. For both cases, the fact that the mutant remains in a low-spin form in the SB conformation in the fatty acid-free structure suggests that the spin-state conversion observed in wild-type P450 BM3 upon substrate association (and the concomitant change in the reduction potential) is a consequence of the physical presence of the lipid in the environment of the heme, and not a result of the adoption of the SB conformation *per se*. Moreover, the fact that the SB conformation is clearly accessible

ble in the substrate-free A264E mutant also suggests that the binding of fatty acid might not be essential for inducing this conformational rearrangement in the wild-type enzyme and that the adoption of the SB conformation in the palmitoleate-bound wild-type structure could merely be a consequence of favorable binding of the substrate to this conformer. This conclusion is supported by the fact that much lower  $K_d$  values are observed for binding of several long-chain fatty acids to the A264E variant than to the wild-type P450 BM3 (17). The SB conformation predominates in the mutant. In ongoing work, we aim to validate further the hypotheses that arise from these findings through creation of other variants at position 264, specifically investigating Ala<sup>264</sup> variants that induce the conformational switch to the SB conformation, but that do not, in addition, give rise to coordination to the heme iron.

**Acknowledgments**—We are grateful for access to beamlines X11 and BW7A at the European Molecular Biology Laboratory Hamburg Outstation (Deutsches Elektronen-Synchrotron), which were essential to crystal improvement, and for access to ID14.1 at the European Synchrotron Radiation Facility, where the final data were collected.

## REFERENCES

- Guengerich, F. P. (2002) *Drug Metab. Rev.* **34**, 7–15
- Munro, A. W., Leys, D., McLean, K. J., Marshall, K. R., Ost, T. W., Daff, S., Miles, C. S., Chapman, S. K., Lysek, D. A., Moser, C. C., Page, C. C., and Dutton, P. L. (2002) *Trends Biochem. Sci.* **27**, 250–257
- Narhi, L. O., and Fulco, A. J. (1987) *J. Biol. Chem.* **262**, 6683–6690
- Fulco, A. J. (1991) *Annu. Rev. Pharmacol. Toxicol.* **31**, 177–203
- Noble, M. A., Miles, C. S., Chapman, S. K., Lysek, D. A., MacKay, A. C., Reid, G. A., Hanzlik, R. P., and Munro, A. W. (1999) *Biochem. J.* **339**, 371–379
- Daff, S. N., Chapman, S. K., Turner, K. L., Holt, R. A., Govindaraj, S., Poulos, T. L., and Munro, A. W. (1997) *Biochemistry* **36**, 13816–13823
- Sligar, S. G. (1976) *Biochemistry* **15**, 5399–5406
- Li, H., and Poulos, T. L. (1997) *Nat. Struct. Biol.* **4**, 140–146
- Cupp-Vickery, J., Anderson, R., and Hatziris, Z. (2000) *Proc. Natl. Acad. Sci. U. S. A.* **97**, 3050–3055
- Poulos, T. L., Finzel, B. C., and Howard, A. J. (1987) *J. Mol. Biol.* **195**, 687–700
- Haines, D. C., Tomchick, D. R., Machius, M., and Peterson, J. A. (2001) *Biochemistry* **40**, 13456–13465
- Podust, L. M., Poulos, T. L., and Waterman, M. R. (2001) *Proc. Natl. Acad. Sci. U. S. A.* **98**, 3068–3073
- Leys, D., Mowat, C. G., McLean, K. J., Richmond, A., Chapman, S. K., Walkinshaw, M. D., and Munro, A. W. (2003) *J. Biol. Chem.* **278**, 5141–5147
- Raag, R., and Poulos, T. L. (1991) *Biochemistry* **30**, 2674–2684
- Ravichandran, K. G., Boddupalli, S. S., Hasermann, C. A., Peterson, J. A., and Deisenhofer, J. (1993) *Science* **261**, 731–736
- Colas, C., and Ortiz de Montellano, P. R. (2003) *Chem. Rev.* **103**, 2305–2332
- Girvan, H. M., Marshall, K. R., Lawson, R. J., Leys, D., Joyce, M. G., Clarkson, J., Smith, W. E., Cheesman, M. R., and Munro, A. W. (2004) *J. Biol. Chem.* **279**, 23274–23286
- Otwinoski, Z., and Minor, W. (1997) *Methods Enzymol.* **276**, 307–326
- Navazza, J. (2001) *Acta Crystallogr. Sect. D Biol. Crystallogr.* **57**, 1367–1372
- Murashudov, G. N., Vagin, A. A., and Dodson, E. J. (1997) *Acta Crystallogr. Sect. D Biol. Crystallogr.* **53**, 240–255
- Roussel, A., and Cambillau, C. (1992) *TURBO-FRODO*, Biographics, Architecture and Fonction des Macromolécules Biologiques (AFMB), Marseilles, France
- Oliver, C. F., Modi, S., Sutcliffe, M. J., Primrose, W. U., Lian, L. Y., and Roberts, G. C. K. (1997) *Biochemistry* **36**, 1567–1572
- Munro, A. W., Noble, M. A., Miles, C. S., Daff, S. N., Green, A. J., Quaroni, L., Rivers, S., Ost, T. W., Reid, G. A., and Chapman, S. K. (1999) *Biochem. Soc. Trans.* **27**, 190–196
- Okita, R. T., and Okita, J. R. (2001) *Curr. Drug Metab.* **2**, 265–281
- Lee, D. S., Yamada, A., Sugimoto, H., Matsunaga, I., Ogura, H., Ichihara, K., Adachi, S., Park, S. Y., and Shiro, Y. (2003) *J. Biol. Chem.* **278**, 9761–9767
- Modi, S., Gilham, D. E., Sutcliffe, M. J., Lian, L. Y., Primrose, W. U., Wolf, C. R., and Roberts, G. C. K. (1997) *Biochemistry* **36**, 4461–4470
- Modi, S., Sutcliffe, M. J., Primrose, W. U., Lian, L. Y., and Roberts, G. C. K. (1996) *Nat. Struct. Biol.* **3**, 414–417
- Miles, J. S., Munro, A. W., Roependowski, B. N., Smith, W. E., McKnight, J., and Thomson, A. J. (1992) *Biochem. J.* **288**, 503–509
- Boddupalli, S. S., Pramanik, B. C., Slaughter, C. A., Estabrook, R. W., and Peterson, J. A. (1992) *Arch. Biochem. Biophys.* **292**, 20–28
- Georgopapadakou, N. H. (1998) *Curr. Opin. Microbiol.* **1**, 547–557
- Ost, T. W., Miles, C. S., Munro, A. W., Murdoch, J., Reid, G. A., and Chapman, S. K. (2001) *Biochemistry* **40**, 13421–13429

## Flavocytochrome P450 BM3 Mutant A264E Undergoes Substrate-dependent Formation of a Novel Heme Iron Ligand Set\*

Received for publication, February 17, 2004  
Published, JBC Papers in Press, March 12, 2004, DOI 10.1074/jbc.M401716200

Hazel M. Girvan<sup>‡§</sup>, Ker R. Marshall<sup>‡</sup>, Rachel J. Lawson<sup>‡</sup>, David Leys<sup>‡</sup>, M. Gordon Joyce<sup>‡</sup>, John Clarkson<sup>¶</sup>, W. Ewen Smith<sup>¶</sup>, Myles R. Cheesman<sup>¶</sup>, and Andrew W. Munro<sup>‡\*\*</sup>

From the <sup>‡</sup>Department of Biochemistry, University of Leicester, Leicester LE1 7RH, United Kingdom,  
the <sup>¶</sup>Department of Pure and Applied Chemistry, University of Strathclyde, Glasgow G1 1XL, Scotland, United Kingdom,  
and the <sup>¶</sup>Department of Chemical Sciences, University of East Anglia, Norwich NR4 7TJ, United Kingdom

A conserved glutamate covalently attaches the heme to the protein backbone of eukaryotic CYP4 P450 enzymes. In the related *Bacillus megaterium* P450 BM3, the corresponding residue is Ala<sup>264</sup>. The A264E mutant was generated and characterized by kinetic and spectroscopic methods. A264E has an altered absorption spectrum compared with the wild-type enzyme (Soret maximum at ~420.5 nm). Fatty acid substrates produced an inhibitor-like spectral change, with the Soret band shifting to 426 nm. Optical titrations with long-chain fatty acids indicated higher affinity for A264E over the wild-type enzyme. The heme iron midpoint reduction potential in substrate-free A264E is more positive than that in wild-type P450 BM3 and was not changed upon substrate binding. EPR, resonance Raman, and magnetic CD spectroscopies indicated that A264E remains in the low-spin state upon substrate binding, unlike wild-type P450 BM3. EPR spectroscopy showed two major species in substrate-free A264E. The first has normal Cys-aqua iron ligation. The second resembles formate-ligated P450cam. Saturation with fatty acid increased the population of the latter species, suggesting that substrate forces on the glutamate to promote a Cys-Glu ligand set, present in lower amounts in the substrate-free enzyme. A novel charge-transfer transition in the near-infrared magnetic CD spectrum provides a spectroscopic signature characteristic of the new A264E heme iron ligation state. A264E retains oxygenase activity, despite glutamate coordination of the iron, indicating that structural rearrangements occur following heme iron reduction to allow dioxygen binding. Glutamate coordination of the heme iron is confirmed by structural studies of the A264E mutant (Joyce, M. G., Girvan, H. M., Munro, A. W., and Leys, D. (2004) *J. Biol. Chem.* 279, 23287–23293).

The cytochrome P450 enzymes constitute a superfamily of heme *b*-containing monooxygenase enzymes responsible for a huge variety of physiologically and biotechnologically impor-

tant transformations (see Refs. 1 and 2). The P450 enzyme heme iron is ligated to the protein by a conserved cysteinate (the proximal ligand), with a water molecule usually present as the distal axial ligand (3, 4). The P450 enzymes are found throughout nature and typically catalyze the reductive scission of dioxygen bound to the heme iron, frequently resulting in hydroxylation of an organic substrate. Two successive 1-electron transfers to the P450 enzyme are required for oxygenation reactions. The source of the electron is reduced pyridine nucleotides (NADPH or NADH), and electron transfer is usually mediated by one or more redox partner enzymes. In hepatic P450 systems, the redox partner is the NADPH-dependent dihydroflavin-cytochrome P450 reductase, which contains FAD and FMN cofactors (1, 5). In mammalian adrenal systems and many bacterial P450 enzymes, electrons are delivered via a two-protein redox system comprising an FAD-containing reductase (adrenodoxin reductase or ferredoxin reductase) and an iron-sulfur protein (ferredoxin) (6, 7). However, other forms of redox systems supporting P450 catalysis are known to exist, including the direct interaction of P450 with hydrogen peroxide to facilitate fatty acid hydroxylation in the *Bacillus subtilis* P450 BS $\beta$  enzyme (8). In recent years, the flavocytochrome P450 BM3 enzyme from *Bacillus megaterium* has been studied intensively as a “model” P450 enzyme (9). P450 BM3 is a natural soluble fusion enzyme in which a fatty-acid hydroxylase P450 enzyme (N-terminal) is joined to a cytochrome P450 reductase (C-terminal), creating an efficient electron transfer system capable of hydroxylating a wide range of fatty acids at rates of several thousand events/min (10, 11). In certain P450 enzymes (including P450 BM3), the binding of substrate facilitates the first 1-electron transfer from the redox partner by increasing the reduction potential of the heme iron through removal of the distal water ligand (12, 13). Dioxygen then binds to the ferrous P450 heme iron. Further reduction (by another 1-electron transfer from the redox partner) accompanied by proton transfer (from solvent, mediated by active-site amino acids) facilitates dioxygen cleavage, production of a molecule of H<sub>2</sub>O, and creation of a highly oxidizing ferryl species that ultimately attacks the substrate to facilitate its oxygenation. Dissociation of the oxygenated product completes the catalytic cycle (14).

Cytochrome P450 enzymes have been isolated from organisms in all domains of life. P450 enzymes from thermostable prokaryotes have been studied to understand the structural adaptations that facilitate the thermal resistance of the protein and heme system, with a view toward potential biotechnological exploitation (see Ref. 15). However, aside from the thermostable isoforms, P450 enzymes from mammals and other mesophilic organisms are not particularly stable enzymes. In particular, the heme site is readily denatured by chemical,

\* The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked “advertisement” in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

§ Supported by the Engineering and Physical Sciences Research Council.

\*\* Supported by the Biotechnology and Biological Sciences Research Council and the Engineering and Physical Sciences Research Council. To whom correspondence should be addressed: Dept. of Biochemistry, University of Leicester, Adrian Bldg., University Rd., Leicester LE1 7RH, UK. Tel.: 44-116-252-3464; Fax: 44-116-252-3369; E-mail: awm9@le.ac.uk.

CYP 102	( <i>B. megaterium</i> )	253	<b>IRYQIITFLIA<u>G</u>HETTSGLLSFALYFLVK</b>	282
CYP 4A4	(Rabbit)	310	<b>LRAEVDTFMF<u>E</u>GHDTTASGVSWIFYALAT</b>	339
CYP 4A1	(Rat)	309	<b>LRAEVDTFMF<u>E</u>GHDTTASGVSWIFYALAT</b>	338
CYP 4B1	(Human)	304	<b>LRAEVDTFMF<u>E</u>GHDTTSGISWFLYCMAL</b>	333
CYP 4C1	(Cockroach)	303	<b>IREEVDTFMF<u>E</u>GHDTTSAGICWALFLLGS</b>	332
CYP 4D1	( <i>Drosophila</i> )	305	<b>IREEVDTFMF<u>E</u>GHDTSSALMFFFYNIAT</b>	334
CYP 101	( <i>P. putida</i> )	238	<b>AKRMCGLLLVL<u>G</u>LDTVVNFLSFSMEFLAK</b>	267

FIG. 1. Alignment of P450 amino acid sequences in the region of the heme-ligating glutamate in the CYP4 family. An alignment of amino acid sequences from various members of the CYP4 family of fatty acid-oxygenating P450 enzymes was performed using the ClustalW sequence alignment program via the European Bioinformatics Institute (available at [www.ebi.ac.uk/clustalw/](http://www.ebi.ac.uk/clustalw/)). The sequence alignment shown is for residues in the I helix region of the P450 enzymes, surrounding the conserved glutamate residue shown to participate in covalent linkage of the heme group in eukaryotic CYP4 enzymes (23, 24). The relevant amino acids at this position in the alignment are shown in **boldface** and are underlined. The amino acid sequences shown are for flavocytochrome P450 BM3 from *B. megaterium* (CYP102A1), rabbit CYP4A4, rat CYP4A1, human CYP4B1, cockroach CYP4C1, *Drosophila melanogaster* CYP4D1, and *Pseudomonas putida* P450cam (CYP101).

thermal, or pressure treatment (see Refs. 16 and 17) to generate one or more inactive forms with altered electronic spectral properties, the so-called "P420" species. In the P420 enzyme, the normal ligation of the heme iron is lost, giving rise to a carbon monooxy complex with Soret absorption maximum shifted from close to 450 nm (in the native form) to ~420 nm (in the inactive form). Protonation of the thiolate ligand is thought to underlie this spectral change and concomitant loss of activity (18). Heme can also be dissociated from P450 enzymes under harsh conditions, and its re-incorporation may not be facile (19). Given the attention to the biotechnological exploitation of P450 enzymes in, for example, diagnostics, biosensing, and fine chemical synthesis (see Ref. 20), there is strong interest in this area. Recent studies have highlighted the power of mutagenesis in the evolution of cytochrome P450 enzymes with altered catalytic properties (21, 22), but their operational stability remains a major problem.

Recent studies on mammalian family 4 (fatty-acid hydroxylase) P450 enzymes have revealed unusual stability in the binding of heme in various members of the family. It was shown that SDS-PAGE resolution of several CYP4 isoforms did not separate heme from the protein, suggesting that there was covalent ligation between the heme macrocycle and the protein (23, 24). In studies of rat liver CYP4A1, a conserved glutamate residue (Glu<sup>318</sup> in this isoform) was found to be the amino acid residue to which the heme had become esterified. The importance of the covalent linkage of heme to the activity of this P450 enzyme was demonstrated through mutagenesis of key residues, although the activity in other CYP4A isoforms may not be so sensitive to the extent of covalent ligation of the heme to the protein (24, 25). Studies on mammalian CYP4A1, CYP4A3, and CYP4A11 confirmed that covalent heme linkage to the protein is autocatalytic and occurs because of esterification with the glutamate via the heme 5-methyl group (25). The formation of a porphyrin carbocation species was postulated as an intermediate in the reaction with the conserved glutamate that facilitates covalent heme ligation (26). Covalent heme linkage has not been recognized in any other P450 isoform, but the ability to stabilize the heme cofactor may have important ramifications regarding the exploitation of P450 enzymes in such areas as biotransformations and toxicological applications.

Flavocytochrome P450 BM3 has been recognized as perhaps the most important P450 enzyme with respect to its capacity to perform biotechnologically exploitable chemical transformations (9). The wild-type enzyme has been shown to catalyze regio- and stereoselective hydroxylation and epoxidation of long-chain saturated and unsaturated fatty acids of varying chain length (11, 21), and mutants generated both rationally and by forced evolution have exhibited novel properties such as hydroxylation at the  $\omega$ -position on fatty acids (rather than the

preferred  $\omega$ -1 to  $\omega$ -3 positions) and oxygenation of polycyclics, substituted fatty acids, indole, alkanes, and short-chain fatty acids and alcohols (see Refs. 21, 22, 27, and 28). However, P450 BM3 exhibits the same structural instability observed in other P450 isoforms regarding the tendency to undergo inactivation at the heme site through P420 formation (16). In this respect, generation of a more stable P450 derivative through covalent attachment of the heme macrocycle is an attractive proposition. The fact that the heme domain of P450 BM3 is strongly related to fatty-acid hydroxylases of the CYP4 family suggests that covalent heme linkage might also be feasible in this isoform, and the fact that it has not been observed for wild-type P450 BM3 could be explained by the presence of an alanine rather than a glutamate in the respective position of the conserved I helix region in this P450 enzyme (Fig. 1) (29).

In this study, we have generated and characterized the A264E variant of P450 BM3 (in both the full-length flavocytochrome and the heme domain) and determined the effect of the mutation on the catalytic, spectroscopic, thermodynamic, and structural properties of the enzyme. In contrast to the results with the mammalian P450 isoforms, the introduction of a glutamate residue does not result in turnover-dependent covalent linkage of the heme macrocycle. Instead, the glutamate becomes a sixth axial ligand to the ferric heme iron, producing a completely novel heme iron ligand set (Cys-Fe-Glu), with occupancy of the glutamate (instead of water) promoted by the binding of substrates to the P450 enzyme.

## EXPERIMENTAL PROCEDURES

### Expression and Purification of Wild-type and Mutant P450 BM3 Proteins

Expression and purification of the mutant (A264E) and wild-type full-length flavocytochrome P450 BM3 proteins and heme domains (amino acids 1–472) were performed essentially as described previously (30, 31). Expression plasmids pBM20 and pBM23 (wild-type P450 BM3 heme domain and intact flavocytochrome P450, respectively) and pHMG1 and pHMG2 (the respective heme domain and intact flavocytochrome P450 A264E clones) were expressed in *Escherichia coli* strain TG1 grown in Terrific Broth and 50  $\mu$ g/ml ampicillin (typically 5 liters of cells) for ~36 h following inoculation from an overnight culture of the relevant transformant. Cells were collected by centrifugation, resuspended in buffer A (50 mM Tris-HCl (pH 7.2) and 1 mM EDTA), and broken using a French press (three passes at 950 p.s.i.), followed by sonication of the resulting suspension on ice in a Bandelin Sonoplus sonicator (5  $\times$  20-s pulses at 50% full power, with adequate cooling time between pulses). Extract was exchanged by dialysis into ice-cold buffer A containing the protease inhibitors benzamidine hydrochloride and phenylmethylsulfonyl fluoride (both at a 1 mM final concentration) prior to loading onto a DEAE column pre-equilibrated in the same buffer. Enzymes were eluted in a linear gradient of 0–500 mM KCl in buffer A. The most intensely red-colored fractions were retained, concentrated by ultrafiltration (Centriprep 30, Millipore Corp.), and dialyzed extensively into buffer B (25 mM potassium phosphate (pH 6.5)) containing

benzamidine and phenylmethylsulfonyl fluoride. Intact flavocytochrome P450 BM3 and its A264E mutant were loaded onto a mimetic yellow column ( $2 \times 15$  cm) and washed extensively with buffer B prior to elution with 25 mM 2'- and 3'-AMP (mixed isomers, Sigma) containing 500 mM potassium chloride. The wild-type heme domain and the A264E mutant were loaded onto a hydroxylapatite column in buffer B and eluted in a linear gradient of buffer B to 500 mM potassium phosphate (pH 6.5) plus protease inhibitors. The most intensely red-colored fractions were pooled, concentrated, and exchanged by dialysis into buffer A at 4 °C prior to loading onto a Q-Sepharose column and eluting as described for the DEAE resin. All flavocytochrome and heme domains were pure at this stage as judged by SDS-PAGE analysis and were concentrated by ultrafiltration to  $\geq 500 \mu\text{M}$  prior to dialysis into buffer A and 50% (v/v) glycerol and storage at -80 °C. The A264E heme domain used in crystallographic trials was exchanged instead into 10 mM Tris-HCl (pH 7.2) and used directly for crystallogenesis.

#### Site-directed Mutagenesis of the CYP102A1 Gene

The A264E mutant forms of full-length flavocytochrome P450 BM3 and its heme domain were constructed by overlapping mutagenic PCR. Three PCRs (A-C) were carried out in total using the heme domain construct plasmid pBM20 (30) as template in PCRs A and B. The pBM20 construct contains the ~1.4-kb gene encoding the heme domain of P450 BM3 (amino acids 1-472 of the flavocytochrome) in the expression vector pUC118 (30). PCR A used primers MfeF (5'-CGCTTGATACAATTGGTCTTGCGG-3', incorporating an MfeI restriction site (underlined) 350 base pairs upstream of the mutation) and BmutR (5'-CACTTGTTGTTCGTGTCCCTCAATTAAAG-3', encompassing the mutagenic codon, with the mutated nucleotide underlined). PCR B used the same template DNA and primers BMutF (5'-CTTAATTGAGGGACCACGAAACAACAAGTGG-3', encompassing the mutagenic codon, with the mutated nucleotide underlined) and BamR (5'-CAGGTCCGACTCTAGAGGATCTATTAGCG-3', incorporating a BamHI site (underlined) 335 bp downstream of the mutation).

The final PCR (PCR C) combined the products of PCRs A and B using primers MfeF and BamR. The product of PCR C was A-tailed using *Taq* DNA polymerase and ligated into pGEM-T (Promega pGEM-T Easy Vector Systems) according to the manufacturer's instructions. This yielded plasmid pGEM-A264E, which was completely sequenced by the dideoxy chain termination method at the Protein and Nucleic Acid Chemistry Laboratory of the University of Leicester. An MscI site upstream of the mutation was used in combination with MfeI to excise a 518-bp fragment containing the A264E mutated region. The relevant fragment was resolved on a 1% agarose gel and purified using a QIAquick gel extraction kit (QIAGEN Inc.). The MfeI/MscI restriction fragment was religated into the backbones of plasmids pBM20 and pBM23, which had been digested with the same restriction enzymes and gel-purified in the same way as the insert fragment. Correct insertion was verified by restriction enzyme digestion.

#### SDS-PAGE Analysis of A264E

SDS-PAGE was performed using a Bio-Rad Mini-PROTEAN II apparatus and either 10 or 6% polyacrylamide gels. SDS-PAGE was used to establish the purity of the wild-type and mutant flavocytochrome P450 BM3 enzymes and heme domains and also to resolve the A264E flavocytochrome P450 BM3 enzyme and heme domain prior to testing for covalent attachment between the heme and P450 protein. This was done using a heme staining procedure to establish whether heme remained associated with the protein following denaturation in SDS (32).

The A264E heme domain was resolved on 10% SDS-polyacrylamide gels in substrate-free and arachidonate-bound forms (~50  $\mu\text{M}$  fatty acid) both in the absence of and following incubation with excess dithionite reductant. Intact A264E flavocytochrome P450 BM3 was run on 6% SDS-polyacrylamide gels in the same forms as for the heme domain and also following incubation with excess NADPH (500  $\mu\text{M}$ ) in the presence and absence of arachidonate and following treatment with hydrogen peroxide (20 mM). Cytochrome c (horse heart, type IV, Sigma) was used as a positive control for covalent heme ligation on 10% gels, and *Shewanella frigidimarina* flavocytochrome  $c_3$  was used on 6% gels (33). A total of 2  $\mu\text{g}$  of enzyme was run in each lane of the SDS-polyacrylamide gels along with broad-range protein molecular weight markers (New England Biolabs Inc.). Gel preparation and electrophoresis were performed as described previously (34). To determine whether heme remained bound to the enzymes following denaturing gel electrophoresis (i.e. was covalently linked to the protein backbone), gels were stained in the dark with 70 ml of 0.25 M sodium acetate (pH 5.0) to which was added 30 ml of 6.3 mM 3,3',5,5'-tetramethylbenzidine in

methanol. Hydrogen peroxide (30 mM final concentration) was then added, and the gel was incubated until light blue bands appeared, indicating the presence of heme in certain proteins. The reaction was then stopped by the addition of ~10% (v/v) isopropyl alcohol. The gels were photographed and then restained with Coomassie Blue to ensure the presence of control and sample proteins at appropriate concentrations.

#### Binding of Substrates and Ligands to the Wild-type and A264E P450 BM3 Enzymes

The binding of fatty acids and heme-coordinating inhibitors to the wild-type and A264E heme domains was analyzed by optical titrations using a Cary UV-50 Bio scanning spectrophotometer (Varian Instruments). The fatty acids arachidonate, palmitoleate, palmitate, myristate, and laurate were used in the titrations. The spectra for the substrate-free wild-type and A264E enzymes (typically 5–8  $\mu\text{M}$  protein) were recorded at 30 °C in assay buffer (1-ml total volume of 20 mM MOPS<sup>1</sup> (pH 7.4) and 100 mM KCl) prior to additions of the fatty acids in 0.1–0.5- $\mu\text{l}$  aliquots (using a Hamilton syringe) up to a final volume of not more than 1% of the total volume of the solution. Fatty acids were prepared as concentrated stock solutions (3–25 mM) in ethanol (arachidonate, palmitate, and myristate) or Me<sub>2</sub>SO (palmitoleate) or as a saturated aqueous stock solution (~900  $\mu\text{M}$ ) in assay buffer (laurate) (31). Spectra were recorded after each addition of substrate, and a difference spectrum was computed by subtraction of the starting (substrate-free) spectrum from those generated at each point in the titration. The maximum apparent absorption change induced at each point in the titration was determined by subtraction of the minimum absorption value at the trough in each difference spectrum from the maximum value at the peak (using data at the same wavelengths in each titration). The maximum changes in absorption determined in this way were plotted versus the relevant fatty acid concentration. A similar approach was taken to determine the apparent binding constant for the inhibitor 4-phenylimidazole. Data were fitted either to a rectangular hyperbola or, in cases where fatty acids bound very tightly to the P450 enzyme, to a quadratic function that accounts for the quantity of enzyme consumed in the enzyme-substrate complex at each point in the titration (Equation 1),

$$A_{\text{obs}} = (A_{\text{max}}/2E_t) \times (S + E_t + K_d) - (((S + E_t + K_d)^2 - (4 \times S \times E_t))^{0.5}) \quad (\text{Eq. 1})$$

where  $A_{\text{obs}}$  is the observed absorption change at the substrate/ligand concentration  $S$ ,  $E_t$  is the total enzyme concentration, and  $K_d$  is the dissociation constant for the enzyme-ligand/substrate complex. All fitting of data was done using Origin software (Microcal).

The "P450" forms of the wild-type and A264E P450 BM3 enzymes were generated by the addition of a few grains of fresh sodium dithionite to the proteins (typically 4–8  $\mu\text{M}$  in assay buffer), followed by slow bubbling of the solution with carbon monoxide gas for ~1 min. The nitric oxide adduct of the P450 enzymes was generated by release of approximately five small bubbles into a similarly buffered solution of ferric enzyme.

#### Studies of the Effects of pH, Ionic Strength, and Temperature on the Optical Properties of the A264E Heme Domain

UV-visible spectra were recorded for the substrate-free form of the A264E heme domain (4  $\mu\text{M}$ ) in 50 mM potassium phosphate (pH 5.0–9.0) at 0.5 pH unit intervals. Spectral perturbations were observed, and absorption data reflecting the maximum overall change between the low- and high-pH spectra ( $\Delta A_{417} - \Delta A_{423}$  with reference to the spectrum collected at pH 7.0) were plotted against pH and fitted to a sigmoid to derive an apparent  $pK_a$  value accompanying the spectral conversion. Further UV-visible spectra for the substrate-free form of the A264E heme domain (4  $\mu\text{M}$ ) were collected in 20 mM MOPS (pH 7.4) and in the same buffer containing potassium chloride at 0–1 M. Finally, spectra were recorded for the A264E heme domain (4  $\mu\text{M}$ ) in assay buffer at temperatures between 18 and 66 °C in 2 °C intervals. The sample was heated using a Peltier system on the Varian spectrophotometer, with temperature controlled via a circulating water bath attached to the Peltier system. A 2-min equilibration time was allowed at each temperature point in the titration prior to collection of the spectrum.

<sup>1</sup> The abbreviations used are: MOPS, 4-morpholinepropanesulfonic acid; MCD, magnetic circular dichroism; CT, charge-transfer.

### Redox Potentiometry

All redox titrations were carried out in an anaerobic glove box (Belle Technology, Portesham, United Kingdom) under a nitrogen atmosphere, with oxygen levels maintained at <5 ppm. Redox titrations were carried out for both the wild-type and A264E P450 BM3 heme domains (typically 6–10  $\mu\text{M}$ ) in both the presence and absence of arachidonic acid (~70  $\mu\text{M}$ ,  $K_d < 5 \mu\text{M}$ ) following the method of Dutton (36) and essentially as described previously (12, 35). Redox titration buffer (100 mM potassium phosphate (pH 7.0)) was deoxygenated by bubbling extensively with  $\text{O}_2$ -free argon and degassed prior to transfer to the glove box. Oxygen was removed from P450 samples by passing concentrated stock solutions through a Bio-Rad Econo-Pac 10DG gel filtration column in the glove box, which had been pre-equilibrated with redox titration buffer containing 5% (v/v) glycerol. Enzyme (typically 6–10  $\mu\text{M}$ ) was titrated with sodium dithionite as a reductant and potassium ferricyanide as an oxidant. Dithionite and ferricyanide were delivered in ~1–2- $\mu\text{l}$  aliquots from concentrated stock solutions (typically 10–50 mM). Titrations were performed in both the reductive and oxidative directions to ensure lack of hysteresis. Absorption changes during the titrations were monitored via a fiber optic absorption probe (Varian Instruments) immersed in the P450 solution in an anaerobic environment and connected to a Cary UV-50 Bio UV-visible spectrophotometer (Varian Instruments) outside the glove box. Potentials were measured using a Hanna pH 211 meter coupled to a platinum/calomel electrode (ThermoRussell Ltd.) at  $25 \pm 2^\circ\text{C}$ . The electrode was calibrated using the  $\text{Fe}^{3+}/\text{Fe}^{2+}$ -EDTA couple as a standard (+108 mV). A factor of +244 mV was used for correction relative to the standard hydrogen electrode. Mediators were added to facilitate electrical communication between enzyme and electrode prior to titration. Typically, 2  $\mu\text{M}$  phenazine methosulfate, 5  $\mu\text{M}$  2-hydroxy-1,4-naphthoquinone, 0.5  $\mu\text{M}$  methyl viologen, and 1  $\mu\text{M}$  benzyl viologen were included to mediate in the range between +100 and –480 mV as described previously (12, 35). The electrode was allowed to stabilize between each addition of reductant/oxidant prior to spectral acquisition and recording of the potential.

Data were analyzed by plotting the absorbance at an appropriate wavelength, corresponding to the maximum absorbance change between oxidized and reduced forms, against the potential. A 1-electron Nernst function was then fitted to the data to describe the transition between ferric and ferrous heme iron, and the midpoint potential was calculated from this data fit. For the substrate-bound titrations, arachidonate was added from a 33 mM stock solution in ethanol until no further spectral shift was observed. Data generated from the fitting procedures with the wild-type P450 BM3 heme domain were in close agreement with those from previous studies (37).

### Steady-state Kinetics

The apparent rates of fatty acid-dependent NADPH oxidation catalyzed by the wild-type and A264E flavocytochrome P450 BM3 enzymes were determined essentially as described previously (35). All measurements were carried out in a 1-cm path length quartz cuvette in assay buffer at 30 °C at a final enzyme concentration of 0.1  $\mu\text{M}$ . Reaction rates were determined over a range of fatty acid concentrations (using laurate, myristate, palmitate, palmitoleate, and arachidonate as described for the optical titrations), typically up to ~100  $\mu\text{M}$  fatty acid. The reaction was started by the addition of 200  $\mu\text{M}$  NADPH, and activity was monitored by  $\Delta A_{340}$  ( $\Delta\epsilon_{340} = 6210 \text{ M}^{-1} \text{ cm}^{-1}$ ). Rates at individual substrate concentrations were measured in triplicate. Rate versus substrate concentration data were fitted to a rectangular hyperbola to define the  $K_m$  and  $k_{cat}$  parameters.

Measurement of reductase domain-dependent reduction of cytochrome *c* by the wild-type and A264E flavocytochrome P450 BM3 enzymes was performed essentially as described above for fatty acid turnover. However, the enzyme concentration was 7 nM, and rates were determined from the accumulation of reduced products at appropriate wavelengths: for ferricyanide, at 420 nm ( $\Delta\epsilon_{420} = 1010 \text{ M}^{-1} \text{ cm}^{-1}$ ); and for cytochrome *c*, at 550 nm ( $\Delta\epsilon_{550} = 22640 \text{ M}^{-1} \text{ cm}^{-1}$ ) (38).

### Determination of Oxygenated Fatty Acid Products

Turnover experiments were carried out by incubating 0.4  $\mu\text{M}$  wild-type or A264E flavocytochrome P450 BM3 with 200  $\mu\text{M}$  myristic acid, palmitic acid, or palmitoleic acid and 600  $\mu\text{M}$  NADPH in a final volume of 5 mL. The reaction was allowed to proceed for 4 h at 30 °C with stirring before halting the reaction by acidification to pH 2.0 with hydrochloric acid. Fatty acids were extracted from the aqueous environment into 1 volume of dichloromethane. The remaining aqueous material was removed by the addition of excess solid magnesium sulfate to the mixture. Following filtration, to remove magnesium sulfate, the

dichloromethane was evaporated under a stream of nitrogen gas, and lipids were resuspended in a small volume of methanol. Thereafter, 20- $\mu\text{l}$  samples were analyzed by electrospray mass spectrometry (70-eV ionization) using a Micromass Quattro triple-quadrupole mass spectrometer. Samples from the aqueous layer were also run to ensure complete extraction, and negative controls (in which no NADPH was added to the enzyme/fatty acid mixtures) were also examined to ensure that the products resulted from NADPH-dependent enzyme activity.

### Spectroscopic Analysis

**EPR Spectroscopy**—Perpendicular mode X-band EPR spectra were recorded on an EPR spectrometer comprising an ER-200D electromagnet and microwave bridge interfaced to an EMX control system (Bruker Spectrospin) and fitted with a liquid helium flow cryostat (ESR-9, Oxford Instruments) and a dual-mode X-band cavity (Bruker type ER4116DM). Spectra were recorded at 10 K with 2-milliwatt microwave power and a modulation amplitude of 1 milliTesla. Protein samples were in assay buffer.

EPR spectra were recorded for the wild-type (~300  $\mu\text{M}$ ) and A264E (~400  $\mu\text{M}$ ) heme domains of P450 BM3 in the absence of added substrate, in the presence of the substrate arachidonic acid (500  $\mu\text{M}$ ), and in the presence of the azole inhibitor 4-phenylimidazole (500  $\mu\text{M}$ ). Saturation with substrate/inhibitor was verified by acquisition of UV-visible spectra (using a 0.1-cm path length cuvette).

**Circular Dichroism**—CD spectra were recorded in the far-UV (190–260 nm) and near-UV-visible (260–600 nm) ranges using a Jasco J715 spectropolarimeter at ambient temperature. Spectra were recorded for both the substrate-free and substrate (arachidonate)-bound forms of the A264E heme domain (0.25  $\mu\text{M}$  P450 in the far-UV region and 20  $\mu\text{M}$  P450 in the near-UV-visible region) in assay buffer. Spectra were recorded in quartz cells of 0.1- and 1.0-cm path length for the far-UV and near-UV-visible regions, respectively, at scan rates of 10 nm/min for the far-UV region and 20 nm/min for the near-UV-visible region. Arachidonic acid (100  $\mu\text{M}$ ) was delivered in ethanol (<0.5% (v/v) final concentration). Control spectra were recorded for A264E P450 BM3 in the presence of ethanol alone to ensure that no effect was induced by the solvent.

**Magnetic Circular Dichroism Spectroscopy**—Magnetic circular dichroism (MCD) spectra were recorded for the wild-type and A264E P450 BM3 heme domains on circular dichroographs (Jasco models J-810 and J-730 for the UV-visible and near-infrared regions, respectively) using a 0.1-cm path length quartz cuvette and the same sample concentrations used for the EPR studies. An Oxford Instruments superconducting solenoid with a 25-mm ambient bore was used to generate a magnetic field of 6 Tesla for the room temperature MCD measurements. At room temperature, MCD intensities are linearly dependent on magnetic field and are plotted normalized to magnetic fields as  $\Delta\epsilon/H$  ( $\text{M}^{-1} \text{ cm}^{-1} \text{ Tesla}^{-1}$ ). To record MCD through the 1400–2000 nm region, interference from absorption caused by vibrational overtones was minimized by preparing samples in deuterium oxide solutions (39). Samples were prepared in 20 mM HEPES made up in deuterium oxide (pH 7.5).

**Resonance Raman Spectroscopy**—Resonance Raman spectra for the A264E P450 BM3 heme domain were obtained in 0.1 M potassium phosphate (pH 7.0) using 15-milliwatt 406.7 nm radiation at the sample from a Coherent Innova 300 krypton ion laser. Spectra were acquired using a Renishaw micro-Raman system 1000 spectrometer. The sample (~50  $\mu\text{M}$ ) was held in a capillary under a microscope, with 8 × 30-s exposure for data collection. Spectra were recorded in the absence of substrate and following the addition of arachidonic acid (100  $\mu\text{M}$ ) to the sample.

### RESULTS

**Molecular Biology, Expression, and Heme Binding Properties**—The A264E mutation was introduced into wild-type constructs of full-length flavocytochrome P450 BM3 and its heme domain by overlapping PCR mutagenesis as described under “Experimental Procedures.” In the final constructs, the presence of the correct mutation (and the absence of exogenous mutations) was verified by complete sequencing of the genes by the dideoxy chain termination method. The levels of expression of the A264E flavocytochrome P450 BM3 enzyme and heme domain were not significantly perturbed by the I helix mutation A264E, with typically 20–30 mg of protein produced per liter of *E. coli* transformants. The mutant flavocytochrome P450 enzyme and heme domain were purified as described

previously (21, 31), and preliminary characterization by UV-visible spectroscopy indicated that the oxidized form of the pure enzymes incorporated heme fully and had Soret/ $A_{280}$  ratios very similar to those of the wild-type forms. To establish whether the enzymes purified from *E. coli* had covalently attached heme, wild-type and A264E protein samples (heme domain and full-length flavocytochrome) were resolved by SDS-PAGE and stained using 3,3',5,5'-tetramethylbenzidine following the methods described by Thomas *et al.* (32). In contrast to the results of Rettie and co-workers (23, 26) and of Ortiz de Montellano *et al.* (24, 25) with various CYP4 isoforms, there was no discernible staining of the P450 BM3 heme with 3,3',5,5'-tetramethylbenzidine that would indicate the presence of covalently bound heme. In view of previous results indicating turnover-dependent formation of the covalent linkage in CYP4 isoforms, the A264E heme domain and full-length flavocytochrome were preincubated both with and without excess arachidonic acid in the presence or absence of excess reductant (dithionite or NADPH). Additional samples (in the presence or absence of arachidonate) were exposed to hydrogen peroxide (which can drive P450 catalysis via the "peroxide shunt" pathway). All samples were incubated at ambient temperature or at 30 °C for periods between 5 and 30 min for reactions to occur. Samples thus treated were again resolved by SDS-PAGE and stained for heme. However, again there was no evidence of any detectable amount of covalently bound heme. Controls with wild-type enzymes produced similar results. By contrast, strong blue bands were observed for the two control hemoproteins used: horse heart cytochrome c and *S. frigidimaria* flavocytochrome *c*<sub>3</sub>. Both of these enzymes contain covalently attached *c*-type heme groups (33). Subsequent staining of these gels with Coomassie Blue confirmed the presence of large amounts of the relevant P450 proteins, providing further proof that the *b*-type heme did not remain bound to the proteins in either the wild-type or A264E forms.

**UV-visible Spectroscopy**—Despite the apparent lack of covalent heme ligation, close examination of the electronic spectrum of the oxidized forms of the A264E enzymes indicated that the heme signals were slightly shifted with respect to the wild-type forms. The Soret band maximum for both the oxidized substrate-free A264E flavocytochrome and heme domain was shifted to a longer wavelength by ~2 nm with respect to the wild-type forms (from ~418 to 420.5 nm), and there were similar small perturbations in the  $\alpha/\beta$ -band region (from ~568 to 571 nm for the  $\alpha$ -band and from 534 to 538 nm for the  $\beta$ -band). In the dithionite-reduced form, the spectral properties of the ferrous forms of the substrate-free wild-type and A264E mutants were virtually indistinguishable, with the Soret band shifted to 411 nm and the  $\alpha/\beta$ -bands apparently fused with a maximum at 546 nm (Fig. 2A). The spectral properties of the ferrous-carbon monooxy complexes were also determined, showing a Soret shift to ~449 nm for the A264E heme domain compared with ~448 nm for the wild-type enzyme (30, 40). To analyze further the spectral properties of the A264E mutant, a nitrosyl complex was generated by bubbling the A264E heme domain with NO gas (five small bubbles). A Soret band shift to 435 nm was observed (compared with 434 nm for the wild-type P450 BM3 heme domain). As with the wild-type enzyme, there was a marked increase in intensity of the  $\alpha$ - and  $\beta$ -bands, with their maxima shifted to 575 and 543.5 nm, respectively. The binding of a tight-binding azole inhibitor (4-phenylimidazole) was also performed. In this case, the spectral features of the complex were virtually indistinguishable from those of wild-type P450 BM3, with the Soret band shifted to 425.5 nm and with changes in  $\alpha/\beta$ -band intensity and shifts to 575/544 nm (Fig. 2B).

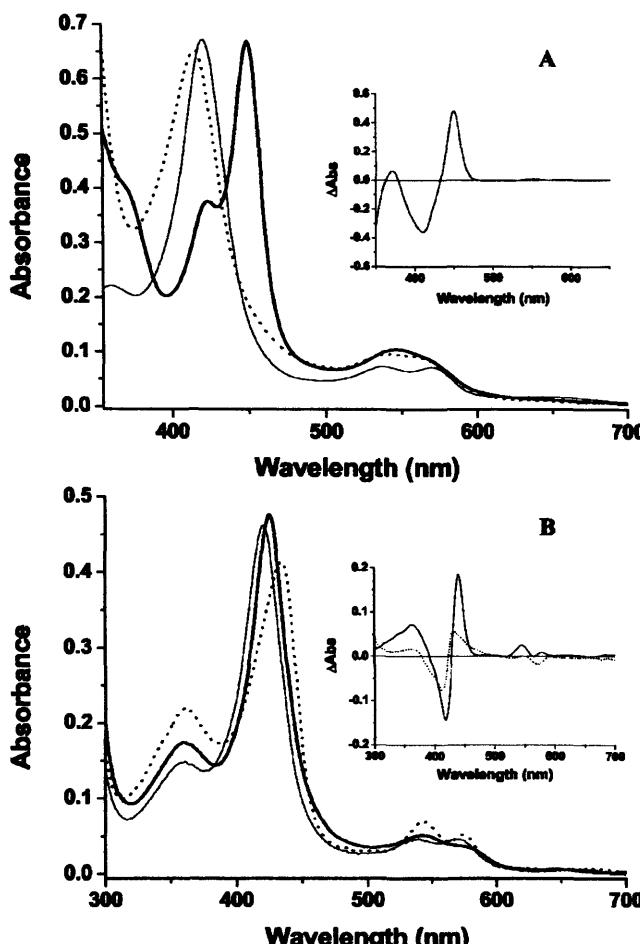


FIG. 2. Absorption properties of the A264E heme domain of flavocytochrome P450 BM3. A, spectra for the oxidized A264E heme domain (7  $\mu$ M; thin solid line) and for the protein following the addition of excess sodium dithionite reductant (dotted line) and the formation of the carbon monooxy complex (thick solid line). The inset shows the difference spectrum generated by subtraction of the spectrum for the reduced form from that of the CO complex. B, spectra for the oxidized A264E heme domain (5  $\mu$ M; thin solid line) and for the protein following saturation with the azole ligand 4-phenylimidazole (thick solid line) and complexation with nitric oxide (dotted line).  $Abs$ , absorbance. The inset shows the difference spectra generated by subtraction of the spectrum for the oxidized form from those of the nitrosyl complex (solid line) and the 4-phenylimidazole complex (dotted line).

To examine the spectral effects on the binding of fatty acid substrates, optical binding titrations were done for the A264E heme domain using a number of fatty acids (lauric acid, myristic acid, palmitic acid, palmitoleic acid, and arachidonic acid) known to bind tightly to the wild-type enzyme. Surprisingly, spectral changes observed upon the addition of fatty acids to A264E were distinct from those observed previously for wild-type P450 BM3 and various mutant forms (see Ref. 31). Neither lauric acid nor myristic acid ( $C_{12}$  and  $C_{14}$  saturated fatty acids, respectively) induced any significant change in the spectral properties of the A264E mutant, even at concentrations near their solubility limits (950 and 250  $\mu$ M, respectively). By contrast, the longer chain fatty acids palmitic acid ( $C_{16}$  saturated), palmitoleic acid ( $C_{16}$  monounsaturated), and arachidonic acid ( $C_{20}$  polyunsaturated) all induced spectral changes. However, rather than inducing a type I optical change with a shift in the Soret maximum to a shorter wavelength, a Soret shift to a final wavelength maximum of ~426 nm was observed in all cases. The final extent of the spectral shift induced was dependent on the particular fatty acid used, with arachidonic acid being the

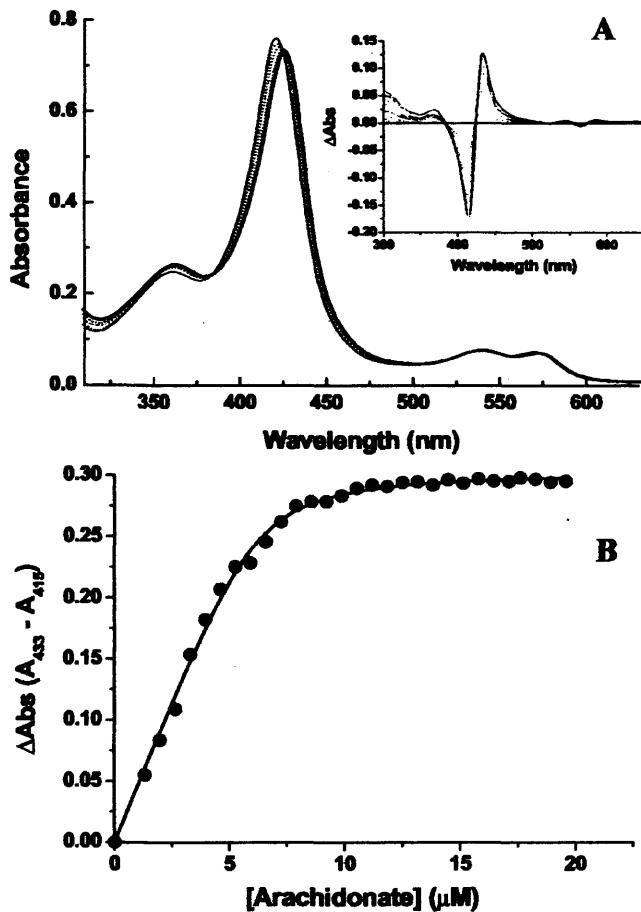


FIG. 3. Optical transitions associated with the binding of arachidonic acid to the A264E heme domain. A, absolute spectrum showing the substrate-free A264E heme domain (thin solid line) with the Soret maximum at 420.5 nm, with successive spectra recorded following the addition of 1.65, 2.97, and 4.95  $\mu$ M arachidonic acid (dotted lines), showing progressive deviation from the starting spectrum. The final spectrum shown was recorded at 8.25  $\mu$ M arachidonic acid (thick solid line), with the Soret maximum at 426 nm. The inset shows nested difference spectra from the arachidonic acid titration of the A264E heme domain generated by subtraction of the starting absolute spectrum for the fatty acid-free enzyme from those produced at subsequent points in the titration. Selected difference spectra (with progressive deviation from the base line) are shown for data recorded at 0.66, 0.99, 1.32, 1.65, 1.98, 2.31, 2.97, 3.3, 3.96, 5.28, 5.61, 6.6, and 6.93  $\mu$ M. An isosbestic point is located at  $\sim$ 424.5 nm. B, plot of the arachidonic acid-induced spectral absorption change ( $A_{433} - A_{415}$ ) from the difference spectra versus [arachidonic acid]. The data were fitted to Equation 1, yielding  $K_d = 0.29 \pm 0.05 \mu$ M. Abs, absorbance.

most effective among those tested. This aspect of the mutant's behavior is discussed in more detail under "Discussion." Thus, rather than undergoing a substrate-dependent optical transition typical of increased high-spin heme iron content, the A264E mutant showed instead a type II transition usually observed upon ligation of inhibitors to the heme iron (e.g. imidazoles) (Fig. 2B). This type of optical transition likely indicates reinforcement of the low-spin form of the cytochrome. Data for the optical titration with arachidonic acid are shown in Fig. 3.

Notwithstanding the unusual spectral changes observed, the apparent binding constants ( $K_d$ ) were determinable from plots of induced spectral changes versus fatty acid concentration as described under "Experimental Procedures." The  $K_d$  values determined for the binding of the fatty acids arachidonic acid, palmitic acid, and palmitoleic acid to the A264E heme domain are lower than those determined from optical titrations with

the same fatty acids and the wild-type heme domain (Table I). Although myristic acid and lauric acid failed to produce spectral changes of sufficient magnitude to facilitate any accurate determination of their apparent  $K_d$  values for the A264E heme domain, the same fatty acids did induce spectral conversion of the wild-type enzyme. However, it should be noted that the extent of laurate- and myristate-induced spin-state conversion in the wild-type P450 BM3 heme domain was lower than that produced with the other fatty acids tested here. It should also be noted that the apparent  $K_d$  values determined for palmitate, palmitoleate, and arachidonate with the A264E heme domain are all considerably tighter than those for the wild-type heme domain (Table I).

An observation from studies of the pH dependence of the UV-visible electronic spectrum of the A264E mutants was that the absorption maximum of the Soret band was very sensitive to pH changes in the range between 5 and 9.5 (in which there was not any significant destruction of the heme). Spectra for the A264E heme domain were recorded in potassium phosphate at several pH values across this range. At pH 7.5, the Soret maximum was located at 420.5 nm, as seen in the buffer used for fatty acid binding titrations and kinetic studies (Fig. 2). As the pH was lowered, there was a consistent shift in the absorption maximum of the Soret band toward shorter wavelengths, reaching 418 nm at pH 5. At higher pH values, the Soret maximum moved to longer wavelengths, positioning at 423 nm at pH 9 and 9.5. A plot of the peak position of the Soret band versus the solution pH described a sigmoid, suggesting the importance of a single ionizable group and possibly reflecting the  $pK_a$  value for a protonatable amino acid side chain. The apparent  $pK_a$  value determined for the optical transition observed was  $8.0 \pm 0.2$ . To examine further the sensitivity of the spectrum of the A264E heme domain to solution conditions, spectra were recorded for the wild-type and A264E enzymes in assay buffer and in the same MOPS buffer with KCl at 0–1 M in 100 mM intervals. Increasing the ionic strength produced negligible effects on the spectrum of the wild-type P450 BM3 heme domain, but induced spectral changes in the A264E heme domain similar to those observed at high pH, i.e. a shift in the Soret band to a longer wavelength. In 20 mM MOPS (pH 7.4), the A264E Soret band was centered at  $\sim$ 419 nm, moving to 420.5 nm in the typical P450 BM3 assay buffer (i.e. plus 100 mM KCl) and to progressively longer wavelengths at higher [KCl]. By 500 mM KCl, the Soret band had shifted as far as 422 nm, and at 1 M KCl, it was further red-shifted to  $\sim$ 423.5 nm. Thus, both high pH and elevated ionic strength produced type II optical shifts in the A264E heme domain (and intact flavocytochrome) that are comparable with those induced by the addition of long-chain fatty acids to the mutant and that are indicative of a novel ligation state of the heme iron.

**Steady-state Kinetic Parameters**—The kinetic properties of the wild-type and A264E flavocytochrome P450 BM3 enzymes were determined with respect to their capacity to catalyze the fatty acid-dependent oxidation of NADPH and the reduction of an exogenous electron acceptor (cytochrome c). Preliminary studies of the pH dependence of the reaction of both the wild-type and A264E enzymes with arachidonic acid indicated that the rate was maximal at pH  $7.5 \pm 0.2$ ; thus, the kinetics were studied in assay buffer, which we have used in several previous studies of P450 BM3. As expected, the A264E mutant catalyzed rapid NADPH-dependent reduction of cytochrome c (mediated via the FMN domain of the enzyme), indicating that the diflavin reductase domain of the enzyme is catalytically unimpaired by the mutation in the heme domain. However, the rates of fatty acid-dependent NADPH oxidation were considerably slower than those for wild-type P450 BM3 (Table I). In the

## Novel Heme Ligation in a Cytochrome P450

TABLE I

Shown are the comparative kinetic ( $k_{cat}/K_m$ ) and substrate/inhibitor binding ( $K_d$ ) constants for the wild-type and A264E P450 BM3 enzymes. Binding constants were determined by optical titrations of fatty acids and heme ligands with the wild-type and A264E heme domains. Kinetic constants were derived from steady-state assays with the wild-type and A264E full-length flavocytochromes. All data were collected as described under "Experimental Procedures." Data were fitted to rectangular hyperbolae, except where indicated. Selected values of  $k_{cat}/K_m$  and  $K_d$  for the wild-type enzyme from previous studies are as indicated.

	Wild-type BM3			A264E BM3		
	$K_d$	$k_{cat}$	$K_m$	$K_d$	$k_{cat}$	$K_m$
	$\mu M$	$min^{-1}$	$\mu M$	$\mu M$	$min^{-1}$	$\mu M$
Laurate	100 ± 15 <sup>a</sup>	2770 ± 120	87 ± 8	ND	744 ± 21	114 ± 12
Myristate	6.9 ± 0.4	4835 ± 295	37.3 ± 11.5	ND	1020 ± 28	10.9 ± 1.3
Palmitate	11.3 ± 0.4	4590 ± 407	11.5 ± 4.9	0.045 ± 0.021 <sup>b</sup>	962 ± 50	16.4 ± 3.6
Arachidonate	3.5 ± 0.17	17100 ± 190 <sup>c</sup>	4.7 ± 0.25 <sup>c</sup>	0.154 ± 0.040 <sup>b</sup>	1710 ± 120	0.65 ± 0.15
Palmitoleate	0.55 ± 0.05 <sup>b</sup>	6980 ± 430	0.75 ± 0.25	0.214 ± 0.019 <sup>b</sup>	3090 ± 90	1.7 ± 0.2
Cytochrome c		11050 ± 630	16.1 ± 4.1		15785 ± 890	28.7 ± 6.5
4-Phenylimidazole	0.85 ± 0.45 <sup>b</sup>			2.5 ± 0.5 <sup>b</sup>		

<sup>a</sup> Data are from Noble *et al.* (71).

<sup>b</sup> Data were fitted to Equation 1.

<sup>c</sup> Data are from Noble *et al.* (31).

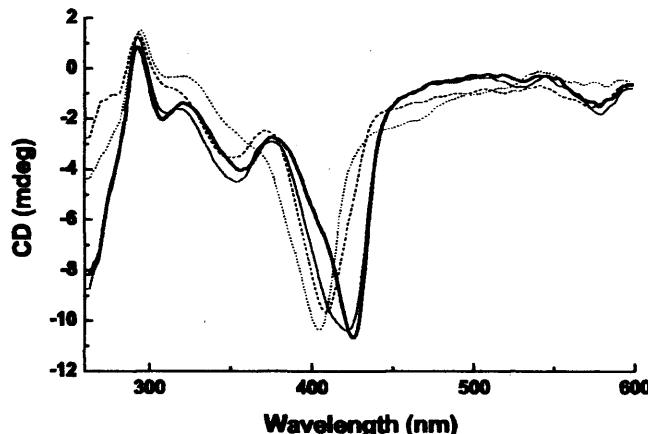
absence of fatty acids, both the wild-type and A264E flavocytochrome P450 BM3 enzymes oxidized NADPH at a slow rate (~5 min<sup>-1</sup>). In the presence of either lauric acid or myristic acid, there was considerable stimulation of NADPH oxidase activity, despite the fact that the addition of these fatty acids did not induce any considerable changes in the optical spectrum of the A264E mutant. Enzyme activity in the presence of palmitic acid, palmitoleic acid, or arachidonic acid was even higher, albeit rather less than that observed with wild-type P450 BM3 (Table I). Thus, it appears that A264E flavocytochrome P450 BM3 retains considerable levels of activity with long-chain fatty acids, despite the fact that these fatty acids induce either negligible change toward the high-spin state or inhibitor-like optical change to the heme spectrum.

**Fatty Acid Oxygenation**—In view of the apparently conflicting data indicating high levels of fatty acid-dependent NADPH oxidation despite type II optical shifts induced by the same fatty acids, we undertook studies to establish whether NADPH oxidation is linked to oxygenation of myristic acid, palmitic acid, and palmitoleic acid. Turnover studies were performed as described under "Experimental Procedures," and products were examined by mass spectrometry. Products were evident from turnover of each of these substrates. In the case of myristic acid, there were similar levels of production of monooxygenated product for both the wild-type and A264E flavocytochrome P450 BM3 enzymes. With palmitic acid, mass spectrometry showed the presence of both mono- and dioxygenated products for both the wild-type and A264E enzymes, consistent with previous data and indicative that a primary hydroxylated product can act as a substrate for a second round of oxidation (11, 41). With the monounsaturated fatty acid palmitoleic acid (*cis*-9-hexadecenoic acid) as the substrate, both mono- and dioxygenated products were observed for both the wild-type and A264E enzymes, with rather lower amounts of products with the mutant enzyme. In previous studies, Fulco and co-workers (42, 43) demonstrated that both hydroxylation (close to the  $\omega$ -terminus) and epoxidation (across the C<sub>9</sub>–C<sub>10</sub> double bond) of palmitoleic acid are catalyzed by wild-type P450 BM3. Under the same experimental conditions, the amounts of products generated from palmitic acid and palmitoleic acid by the A264E mutant were ~30–40% lower than those produced by wild-type P450 BM3, consistent with the differences in steady-state kinetics shown in Table I. In parallel studies of peroxide production during fatty acid turnover, there was no significant difference between the wild-type and A264E enzymes, suggesting that both enzymes couple NADPH oxidation to fatty acid oxygenation tightly, but that the A264E mutant is a much slower hydroxylase than wild-type P450 BM3 (44). Thus, despite the

unusual spectral conversions produced upon binding long-chain fatty acids, the A264E flavocytochrome retains the capacity to oxygenate fatty acids.

**Spectroscopic Analysis**—Optical binding studies for the A264E heme domain indicate that the resting oxidized form of the enzyme has a perturbed UV-visible spectrum compared with that seen in the wild-type enzyme and that substrate addition, basic pH, and high ionic strength induce optical changes indicative of a change in the heme iron coordination state. Although no detectable covalent ligation of the heme macrocycle was detectable with A264E flavocytochrome P450 BM3, it appeared likely that the glutamate may instead interact with the heme macrocycle or, more likely, coordinate to the heme iron itself to produce the spectral perturbation observed. For this reason, we undertook spectroscopic analysis to examine structural features of the mutant protein and its bound heme cofactor.

**Circular Dichroism**—The far-UV CD spectra (190–260 nm) for the substrate-free A264E heme domain indicate no significant change in secondary structure compared with the wild-type heme domain. The predominantly  $\alpha$ -helical A264E heme domain has a far-UV CD spectrum virtually identical to that of the wild-type enzyme, with minima at 223 and 209 nm and an abscissa at 202 nm. The far-UV CD spectrum for A264E (as with the wild-type enzyme) was not significantly altered by the addition of arachidonic acid, showing that secondary structural content is not significantly altered upon the binding of substrate. However, in the near-UV-visible CD region (260–600 nm), the spectrum of the A264E heme domain in its resting oxidized form is rather different from that of the wild-type heme domain. This portion of the spectrum is dominated by a sharp Soret feature with negative ellipticity. For the substrate-free wild-type heme domain, this is centered at ~409 nm and shifts to ~404 nm upon arachidonic acid binding (Fig. 4). Thus, the direction of movement of the Soret band (to shorter wavelengths) mimics the optical transition seen in the electronic absorption spectrum. However, the near-UV-visible CD spectrum of the substrate-free A264E heme domain has its Soret band at ~422 nm. Upon the addition of arachidonic acid, this feature sharpened and shifted to 426 nm, the same wavelength seen for the Soret maximum in the optical spectrum of the arachidonate-bound A264E heme domain (Fig. 4). Thus, near-UV-visible CD spectroscopy shows much more marked differences in the spectra of the wild-type and A264E heme enzymes than does optical absorption spectroscopy. A large spectral difference is seen between the near-UV-visible CD spectra of the substrate-free forms of the wild-type and A264E heme

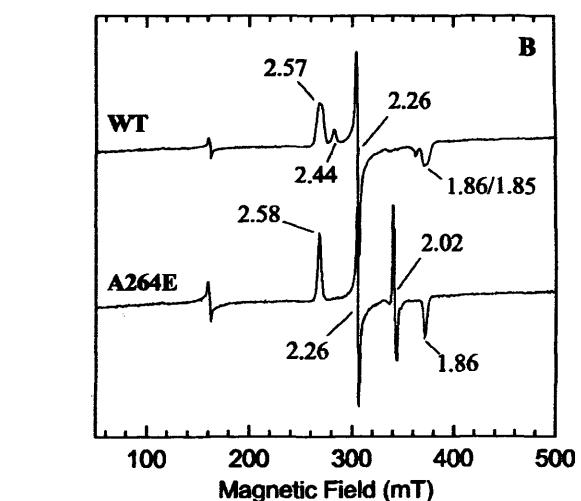
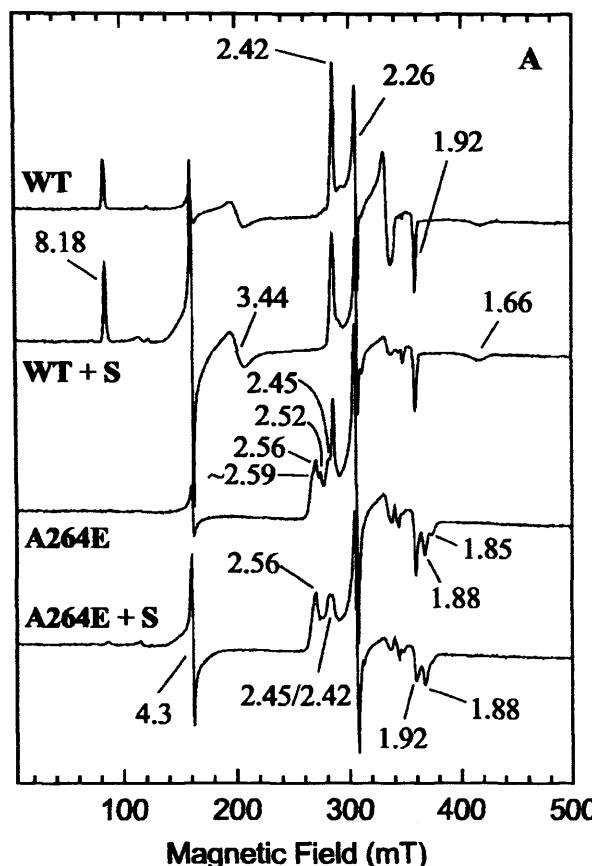


**FIG. 4.** Near-UV-visible CD spectra for the wild-type and A264E P450 BM3 heme domains in the presence and absence of fatty acid substrate. Spectra were recorded for the wild-type (20  $\mu$ M) and A264E (20  $\mu$ M) P450 BM3 heme domain as described under "Experimental Procedures." Following spectral acquisition, arachidonic acid (100  $\mu$ M) was added to both the wild-type and A264E proteins, and spectra were again recorded. All spectra show negative ellipticity in the region of the heme Soret band, with minima located at 409 nm for the substrate-free wild-type P450 (dashed line), shifting to 404 nm in the arachidonate-bound form (dotted line), and at 422 nm for the substrate-free A264E heme domain (thin solid line), shifting to 426 nm for the arachidonate-bound form (thick solid line). mdeg, millidegrees.

domains, whereas (under identical conditions) optical absorption shows a shift of only  $\sim 1.5$  nm in the Soret band.

**Resonance Raman Spectroscopy**—Resonance Raman spectra were recorded for the substrate-free and arachidonic acid-bound forms of the A264E heme domain. For wild-type P450 BM3, considerable spectral perturbations were observed upon the addition of fatty acids, including shifts in position and intensity of the spin-state marker bands  $\nu_2$  and  $\nu_{10}$ , reflecting the change in heme iron spin-state equilibrium toward the high-spin form (30). By contrast, there were no significant changes in any of the oxidation (*e.g.* intensity of  $\nu_4$  remained constant at  $1371\text{ cm}^{-1}$ ) or spin-state (*e.g.* intensity of  $\nu_3$  remained constant at  $1501\text{ cm}^{-1}$ ) marker bands for the A264E heme domain following the addition of arachidonic acid. Resonance Raman spectroscopy confirmed that the substrate-free enzyme is essentially completely low-spin and that the addition of arachidonate reinforces the low-spin state. The only significant differences following fatty acid addition to the A264E heme domain were small increases in intensity of the  $\nu_{C=C}$  band at  $1621\text{ cm}^{-1}$ , relating to the heme vinyl groups, and of the  $\nu_{11}$  band at  $1560\text{ cm}^{-1}$ . The  $\nu_{11}$  band is affected by the degree of conjugation of the heme vinyl groups and their planarity with the heme macrocycle, and these small perturbations should reflect alterations in the positioning of the vinyl groups relative to the heme plane following fatty acid binding. However, resonance Raman spectroscopy did not provide specific vibrational information on the proposed switch of distal ligands to the heme iron that occurred upon arachidonic acid binding to the A264E heme domain. For such information, we turned instead to EPR and MCD spectroscopies.

**EPR Spectroscopy**—EPR spectra of the arachidonate-bound and substrate-free forms of the wild-type and A264E heme domains were recorded and are shown in Fig. 5A. The EPR spectrum for wild-type P450 BM3 is as previously reported (45) and is typical for low-spin ferric P450 enzymes, which all give rise to spectra with  $g_z$  in the range 2.40–2.45 (46–49). Upon the binding of substrate, the ferric spin equilibrium was perturbed, and the iron became a mixture of low- and high-spin forms. The latter appears in the EPR spectra with features at



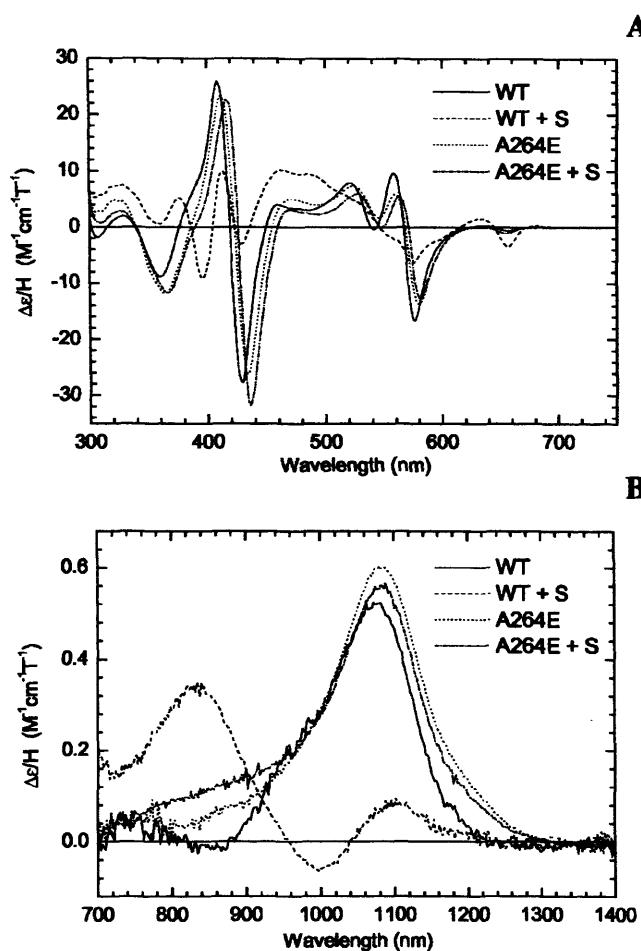
**FIG. 5.** EPR spectra for the wild-type and A264E P450 BM3 heme domains in the substrate-free, arachidonate-bound, and 4-phenylimidazole-bound forms. Spectra were recorded for the wild-type and A264E P450 BM3 heme domains as described under "Experimental Procedures." A, EPR spectra are shown for the oxidized substrate-free forms of the wild-type (WT) and A264E heme domains and following the binding of arachidonate substrate (S). Relevant  $g$ -values are assigned in each spectrum. B, EPR spectra are shown for the wild-type and A264E mutant heme domains in their 4-phenylimidazole complexes. Relevant  $g$ -values are assigned in each spectrum. mT, milliTesla.

$g = 8.18, 3.44$ , and  $1.66$ , which originate in the lowest ( $m_s = \pm \frac{1}{2}$ ) Kramers doublet of the  $S = 5/2$  ferric ion. In the "low-field limit," where the axial zero field-splitting parameter is greater than the Zeeman splittings ( $D \gg g\beta H$ ), these values correspond to a rhombicity of  $E/D = 0.11$ . In ferric hemopro-

teins, such substantial rhombicities are found only with thiolate ligation. The EPR spectra of the substrate-free and arachidonate-bound A264E P450 BM3 enzymes suggest the presence of several low-spin ferric species, all with g-values that indicate that thiolate coordination has been maintained. A264E appears to contain two distinct forms with  $g_z \sim 2.56$  and 2.43, respectively. Each feature shows structure indicative of further minor heterogeneities. There is no example of cysteinate proximal to a neutral oxygen ligand giving a  $g_z$ -value of  $>2.45$ . The  $g_z$  value of  $\sim 2.56$  must therefore indicate ligation different from that of the wild-type enzyme and is strong evidence for coordination of the distal glutamate. In support of this, the formate derivatives of P450cam and chloroperoxidase give  $g_z = 2.55$  and 2.59, respectively (49, 50). The feature at  $g_z \sim 2.43$  suggests that a proportion of the sample retains a distal water ligand, although it appears that the distal mutation causes some heterogeneity in this subpopulation. In contrast to the wild-type P450 BM3 enzyme, when substrate was bound to the A264E mutant, there was no significant switch to a high-spin form. Instead, changes in the EPR spectrum for arachidonate-bound A264E indicate differences in the distribution of the low-spin species, with a diminution of the contribution from a Cys-aqua-ligated form and a simultaneous increase in the proportion of Cys-Glu-coordinated species, indicating that substrate binding promotes the ligation of Glu<sup>264</sup> to the heme iron.

A comparison of the EPR spectra for the 4-phenylimidazole-bound forms of the wild-type and A264E heme domains is shown in Fig. 5B. There is some heterogeneity in the wild-type complex, with the major triplet of g-values at 2.57, 2.26, and 1.86/1.85. The  $g_z$  signal is broadened, suggesting that there may be a split population of two conformers with a slightly different  $g_z$  component and paired with the two subtly different  $g_x$  components. Minor signals at  $g = 2.44$  and 1.92 may reflect a small proportion of unligated enzyme. In the azole-bound A264E complex, there appears to be one predominant species with g-values at 2.58, 2.26, and 1.86. The homogeneity of this spectrum is in part due to the apparently complete ligation of the azole to the heme iron in this sample (removing the residual aqua-ligated components seen in the wild-type enzyme spectrum). However, it appears that a single conformational form of azole-bound heme is present in A264E, whereas there may be two distinct species in wild-type P450.

**MCD**—The MCD spectra for the wild-type P450 BM3 heme domain are consistent with those we have reported previously (45). The room temperature near-UV-visible MCD spectra of the substrate-free wild-type and A264E P450 BM3 enzymes and the arachidonate-bound A264E enzyme (Fig. 6A) each show a pattern of bands typical for low-spin ferric hemes with a thiolate ligand. The unusually low MCD intensity in both the Soret band (400–420 nm) and the  $\alpha/\beta$ -region (500–600 nm) is also characteristic of such species (47, 51–59). A small additional negative feature at  $\sim 655$  nm is part of a derivative-shaped charge-transfer band and arises from the presence of a low level ( $\leq 15\%$ ) of the high-spin form (see Ref. 54). The spectrum of substrate-bound wild-type P450 BM3 is very different and shows that the low-spin heme is now the minority species ( $\sim 25\%$ ). The charge-transfer (CT) band near 655 nm has increased in intensity, and other high-spin bands are evident at 360–405 nm and as a shoulder at  $\sim 555$  nm. Low-spin ferric hemes also give rise to a porphyrin-to-ferric charge-transfer transition at longer wavelengths. This appears as a positive signed band in the MCD spectra and has been located in the room temperature near-infrared MCD spectra for wild-type P450 BM3 minus substrate and for the A264E mutant in both the absence and presence of substrate at  $\sim 1080$  nm (Fig. 6B), as was previously reported for wild-type P450 BM3 at low



**FIG. 6. MCD studies of the A264E heme domain.** MCD spectra in the near-UV-visible (A) and near-infrared (B) regions are shown for both heme domains of wild-type P450 BM3 and the A264E mutant. Protein concentrations used were as follows: wild-type heme domain (WT), 300  $\mu\text{M}$  (near-infrared region) and 30  $\mu\text{M}$  (near-UV-visible region); wild-type heme domain + substrate (S), 550 and 45  $\mu\text{M}$ ; A264E domain, 690 and 33  $\mu\text{M}$ ; and A264E domain + substrate, 660 and 63  $\mu\text{M}$ . T, Tesla.

temperature (45). Close inspection of these three spectra reveals qualitative and quantitative differences. The peak position for the transition shifts from  $\sim 1075$  nm for the substrate-free wild-type enzyme through to  $\sim 1080$  nm for the substrate-free A264E mutant and to  $\sim 1085$  nm for the arachidonate-bound A264E mutant. Differences in both the breadth and intensity of the CT band are discernible (Fig. 6B). The exact energy of this CT transition is generally diagnostic of the two heme axial ligands (60, 61), but the influence of the second ligand is somewhat reduced in the presence of thiolate, as illustrated here by these three examples. Changing water for carboxylate at the distal side of the heme results in only very minor band shifts. For imidazole-bound P450, native CooA, and the hemes in subunit I of SoxAX, all of which have a nitrogenous ligand distal to cysteinate, the CT transitions are at 1180 nm (45), 1120 nm (60), and 1150 nm (62), respectively.

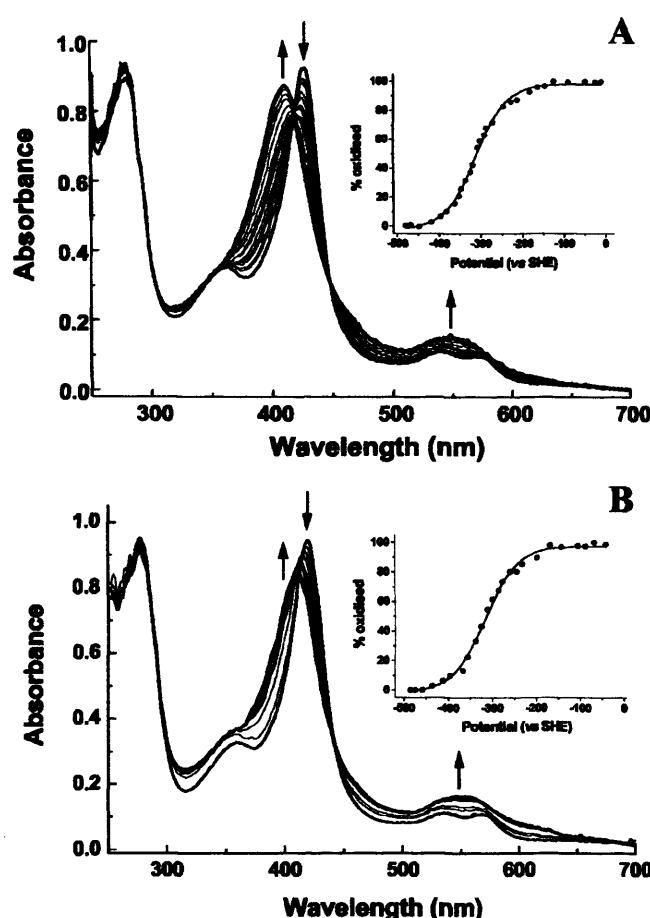
The addition of arachidonic acid substrate to wild-type P450 BM3 resulted in a marked change in the near-infrared MCD spectrum. Consistent with the switch to predominantly high-spin heme that was observed at UV-visible wavelengths, the low-spin CT band near 1080 nm was significantly diminished. The derivative-shaped MCD band centered at  $\sim 900$  nm is the CT band characteristic of high-spin ferric heme and is ex-

tremely similar to that reported for substrate-bound cytochrome P450cam (63).

**Potentiometric Analysis.** Previous studies have shown that fatty acid binding to wild-type P450 BM3 is accompanied by loss of the aqua ligand to the heme iron and a shift in the heme iron spin-state equilibrium toward the high-spin form (21, 30). In P450 BM3 (as in P450cam), this is accompanied by a change in the heme iron reduction potential of  $\sim 130$ – $140$  mV (from  $-427$  to  $-289$  mV for P450 BM3) (12, 37). In view of the markedly different effect of fatty acid binding to the A264E variant, we undertook potentiometric studies of the substrate-free and arachidonic acid-bound forms of this mutant. For both forms of the enzyme, complete dithionite-dependent reduction of the heme iron proved facile under anaerobic conditions; and even under aerobic conditions, the A264E heme iron was almost stoichiometrically reduced by the addition of excess dithionite. By contrast, it was difficult to reduce the substrate-free wild-type heme domain completely using dithionite and impossible under aerobic conditions because of its negative potential and rapid reoxidation of the ferrous form. Upon the addition of near-saturating arachidonate, the A264E Soret band was located at  $\sim 426$  nm and shifted to a final position of  $\sim 410$  nm upon complete reduction of the heme iron during the redox titration (Fig. 7A). In the substrate-free form, the Soret band of the oxidized A264E heme domain was located at 419.5 nm and shifted progressively to a final position of  $\sim 410$  nm upon complete reduction of the heme iron (Fig. 7A). Absorption versus potential data were plotted at 410 nm and were fitted to a 1-electron Nernst equation to define the midpoint potential for the A264E heme iron in the presence and absence of arachidonate. These values were  $-316 \pm 3$  mV in the substrate-free form and  $-314 \pm 4$  mV in the arachidonate-bound form. Thus, substrate binding (which does not increase the high-spin content in the A264E enzyme) does not induce any significant alteration of the apparent reduction potential in the A264E heme domain under these conditions.

#### DISCUSSION

The capacity of the eukaryotic family 4 cytochrome P450 enzymes to link their heme macrocycle covalently to the protein backbone has been one of the most significant discoveries in P450 research in recent years (see Refs. 23 and 24). From a biotechnological perspective, the ability to covalently link the porphyrin to the P450 protein matrix is attractive for at least two reasons. First, the presence of the glutamate and the glutamate ligation process were shown to enhance catalytic rate is clearly a desirable feature to endow on an enzyme. Second (and the major reason), the capacity of P450 enzymes to undergo conversion to the inactive P420 form (in which native cysteinate heme ligation is lost) is well recognized, and heme can even be dissociated completely from the P450 enzyme under moderately denaturing conditions (see Ref. 16). Thus, the ability to covalently tether heme should promote longevity of P450 activity, particularly since the P450-to-P420 conversion has been shown to be reversible in selected P450 enzymes, including P450 BM3 (see Ref. 19). From a perspective of exploitation of P450 enzymes, the bacterial enzymes P450cam and P450 BM3 have been the most intensively studied. Rational mutagenesis of P450cam has produced variants of the camphor hydroxylase that are able to oxygenate molecules such as butane and propane (64). Rational mutagenesis of the fatty-acid hydroxylase P450 BM3 has produced variants that catalyze oxygenation of fatty acids at different positions compared with the wild-type form and in which substrate selectivity has been converted toward short-chain fatty acids and polycyclic aromatic hydrocarbons such as phenanthrene (21, 28). Forced



**FIG. 7. Redox potentiometric analysis of the A264E P450 BM3 heme domain.** The midpoint reduction potential of the heme iron of the A264E heme domain was determined in the presence and absence of arachidonic acid as described under “Experimental Procedures.” *A*, selected spectra from the titration of the arachidonate-bound form of the A264E mutant. Arrows indicate the direction of the spectral change at selected points during the reductive phase of the titration. The spectral maximum shifts from 426 to 410 nm upon conversion between the oxidized (ferric) and reduced (ferrous) forms of the heme iron. The inset shows a plot of the proportion of oxidized enzyme (from  $A_{410}$  data) versus applied potential, fitted to the Nernst function and producing a midpoint reduction potential of  $-314 \pm 4$  mV for the heme iron. *B*, selected spectra from the titration of the substrate-free form of the A264E mutant. Arrows indicate the direction of the spectral change at selected points during the reductive phase of the titration. For clarity, fewer spectra are shown than in *A* since, in this case, the spectral change upon reduction occurs over a more compressed wavelength range (from  $\sim 419.5$  to 410 nm). The inset shows a data fit (as in *A*), producing a heme iron midpoint reduction potential of  $-316 \pm 3$  mV. SHE, standard hydrogen electrode.

evolution of P450 BM3 has also produced an efficient alkane hydroxylase enzyme (22). It is clear that P450 BM3 has great biotechnological potential for production of functionalized hydrocarbons. In view of this potential and the close relationship between P450 BM3 and the eukaryotic family 4 fatty-acid hydroxylases, we mutagenized P450 BM3 to introduce the glutamate conserved in the eukaryotic family 4 P450 enzymes at the same I helix position in P450 BM3 (*i.e.* Ala<sup>264</sup>) to examine whether a similar covalent heme linkage could be induced in P450 BM3. It is clear that the heme linkage does not occur in this point mutant. However, unusual spectral perturbations were evident in the A264E variant, and these suggested that the glutamate residue might instead interact directly with the heme iron. This theory was validated by a combination of spectroscopic methods and ultimately by obtaining the atomic

structure of the A264E heme domain in the substrate-free and palmitoleate-bound forms (see accompanying article (72)).

The capacity of long-chain fatty acids to induce a type II spectral transition of the P450 heme in the A264E heme domain and flavocytochrome was the first indication that the introduced glutamate might replace water as the distal ligand to the heme iron in the long-chain fatty acid-bound forms of the enzyme. In turn, this suggested that the reason for the perturbed spectrum for the substrate-free form of the A264E enzyme (Soret band at 420.5 nm compared with 419 nm for wild-type P450 BM3) was that a proportion of the heme iron was glutamate-ligated in the resting enzyme and that the equilibrium was forced toward the glutamate-ligated form in the presence of long-chain fatty acids (the preferred substrates for P450 BM3). Increasing ionic strength and basic pH had a similar effect on the optical spectra; and in the latter case, an apparent  $pK_a$  of  $8.0 \pm 0.2$  may be assigned to the protonation of Glu<sup>264</sup> in its hydrophobic location in the mutant active site. The fact that the shorter chain fatty acids (myristic acid and lauric acid) had an almost negligible effect on the ligation state is consistent with their weaker binding to wild-type P450 BM3 and is also explicable in terms of their positioning in the active site of the enzyme. Li and Poulos (65) determined the atomic structure of oxidized palmitoleate-bound wild-type P450 BM3 and highlighted the interaction of the carboxylate group with the Tyr<sup>51</sup>/Arg<sup>47</sup> motif near the mouth of the substrate entry channel. The  $\omega$ -terminus of the longer chain fatty acids will extend further toward the heme iron in the oxidized form of P450 BM3 and the A264E variant, and this is predicted to afford their interaction (directly or indirectly through the influence of Phe<sup>87</sup>) with the glutamate and to induce its repositioning to interact with the heme iron. The shorter chain fatty acids, with their carboxylate still tethered at the active-site mouth, are predicted not to extend far enough toward the heme to enable interaction with the glutamate. Again, these predictions are supported by structural data (see accompanying article (72)).

Spectroscopic studies were undertaken to provide further proof that a novel form of heme iron ligation occurred in the A264E P450 BM3 variant. Resonance Raman spectroscopy provided data consistent with those obtained from optical titrations, confirming that the low-spin form is reinforced upon substrate binding. However, it could not provide convincing data regarding the proposed ligand switch in A264E. EPR and MCD spectroscopies could provide such data. EPR studies of the wild-type P450 BM3 heme domain are consistent with previous work on the enzyme and show that the binding of fatty acid induces a large change in spin-state equilibrium toward the high-spin form. However, the spectra for the A264E mutant are considerably different from those for the wild-type heme domain. In the absence of substrate, the EPR spectrum of A264E is typical for a low-spin heme, but shows substantial heterogeneity in the signal that we consider to result from an ensemble of the forms present. The observation of  $g$ -values identical to those of the substrate-free wild-type heme domain indicates that a significant proportion of the mutant persists in a "native-like" water-bound form, although small shoulders on these features suggest some minor heterogeneity (Fig. 5A). The other three  $g_z$  components clustered around  $g = 2.56$  are assigned to glutamate-ligated forms of the protein. This 3-fold heterogeneity may result from different orientations of the glutamate distal ligand and possibly from interactions between the two different oxygen atoms of the glutamate carboxylate and the heme iron.

Atomic structural data show that, in the aqua-ligated form of A264E, the glutamate residue interacts with the side chain of

Phe<sup>87</sup> (see accompanying article (72)). Phe<sup>87</sup> is known to be a conformationally flexible amino acid, which is important in controlling substrate selectivity and for interaction with the  $\omega$ -terminal carbon of fatty acid substrates of P450 BM3, preventing hydroxylation at this position (65, 66). Potentially, interactions between Phe<sup>87</sup> and Glu<sup>264</sup> in its ligated form provide further explanations for the heterogeneity in the substrate-free EPR spectrum. In the arachidonate-bound form of A264E, the heterogeneity is reduced, due partly to the loss of much of the aqua-ligated form as the fatty acid "drives" on the glutamate ligand. It is likely that this occurs because of the substrate displacing unligated glutamate from its other favored conformation, in which it interacts with the phenyl group of Phe<sup>87</sup>. However, the apparent degree of heterogeneity in the EPR spectrum is further reduced because of alterations in the proportions of the different glutamate-ligated forms of A264E. Potentially, this also reflects the repositioning of Phe<sup>87</sup> following the binding of substrate and the position(s) adopted by the fatty acid itself. The EPR data for the glutamate-ligated form(s) of P450 BM3 are consistent with previous studies of P450cam in complex with oxygen donor ligands (e.g. formate, acetate, and propionate), and there is particularly strong similarity to formate-bound P450cam, which displays  $g$ -values at 2.55, 2.25, and 1.88 (49, 52).

A limited number of heme iron ligand sets are observed in natural cytochromes. These include His-Met (e.g. in cytochrome  $cd_1$  from *Pseudomonas aeruginosa*), Met-Met (bacterioferritin from the same bacterium), and His-His (e.g. eukaryotic cytochrome  $b_5$ ) (67–69). Cysteinate-ligated heme enzymes include the Cys-aqua-ligated cytochrome P450 enzymes and the nitric-oxide synthases (see Refs. 45 and 49), and the Cys-His-ligated heme in the SoxAX protein from *Rhodovulum sulfidophilum*, with a role in thiosulfate oxidation (62). However, the Cys-Glu ligation observed in the A264E mutant of P450 BM3 is an unprecedented heme iron ligand set. Spectroscopic studies indicate that it is present as a minor component of the substrate-free form of the A264E P450 BM3 enzyme, with the predominant form being the "normal" Cys-aqua-ligated P450. However, the binding of long-chain fatty acids forces the equilibrium heavily in favor of the novel Cys-Glu ligand set in a similar fashion as the binding of these substrates induces aqua ligand displacement and formation of a high-spin five-coordinate heme iron species in wild-type P450 BM3. An intriguing aspect evident from this study is that the A264E flavocytochrome P450 BM3 enzyme retains considerable fatty-acid oxygenase activity, at least toward those fatty acids that are efficient in inducing the switch to Cys-Glu coordination. This indicates that the Glu ligand must be displaced (at least in a proportion of the enzyme) following the first reduction step in the catalytic cycle. This should allow oxygen to bind, reductive scission of dioxygen to occur following the second electron reduction of the iron, and the tightly bound fatty acid substrate to be oxygenated. In reporting their structure of the palmitoleate-bound form of the wild-type P450 BM3 heme domain, Li and Poulos (65) noted that the distance between the  $\omega$ -carbon of the bound substrate and the heme iron is too great for oxidative attack of the substrate and that further structural change resulting in repositioning of the substrate closer to the heme is required following reduction of the ferric iron. NMR studies of the ferric and ferrous P450 enzymes indicate that there should be a 6-Å movement of substrate subsequent to heme iron reduction (70). Thus, it appears clear that the Cys-Glu coordination is broken following heme iron reduction, enabling binding of oxygen and catalysis to ensue, albeit with lower catalytic rate than in the wild-type enzyme. Potentiometric analysis (Fig. 7) demonstrated that the reduction potentials of the sub-

strate-bound and substrate-free forms of the A264E heme domain are similar to one another ( $-315 \pm 5$  mV) and to that of the fatty acid-bound wild-type enzyme ( $-289$  mV), consistent with the capacity of the reductase domain to mediate electron transfer to the heme iron in the A264E flavocytochrome enzyme. The reduction potential of the substrate-free form of the wild-type P450 BM3 heme domain is  $-427$  mV (37). The apparent near-identical heme iron reduction potentials of the substrate-free and arachidonate-bound forms of the A264E variant are possibly a consequence of the same species undergoing reduction/oxidation in both cases. A likely scenario is that the Cys-aqua species is reduced preferentially (i.e. has a more positive reduction potential than the Cys-Glu species) and that the equilibrium between the Cys-aqua and Cys-Glu ferric forms is drawn toward the former as the reductive titration progresses. The altered heme iron reduction potential of the A264E mutant is discussed further in the accompanying article (72) in light of major structural changes observed for the mutant.

In conclusion, the A264E variant of P450 BM3 produces an unprecedented heme iron ligand set (Cys-Fe-Glu), with the binding of long-chain fatty acids forcing the equilibrium between the "native" Cys-aqua-ligated form and the novel Cys-Glu-ligated form toward the new ligation state. Structural reasons underlying this substrate-induced equilibrium perturbation are described in the accompanying article (72). Differences in active-site architecture between P450 BM3 and the mammalian CYP4 enzymes likely explain the failure of the A264E P450 BM3 variant to covalently link the heme macrocycle to the glutamate side chain. Such covalent attachment should still be feasible by altering the position of the introduced glutamate to avoid steric restrictions or by combining A264E with secondary mutations. However, the results with the A264E variant open new avenues in the study of P450 heme coordination chemistry. Having created a completely novel Cys-Glu ligand set to the heme iron, the opportunity arises that other mutations at Ala<sup>264</sup> could generate additional sets of proteinaceous heme ligands that have not yet been observed in nature, possibly preempting the discovery of such cytochromes. This offers exciting possibilities for structural and spectroscopic study. In ongoing work, we are generating a series of Ala<sup>264</sup> mutants in P450 BM3 in attempts to produce and characterize additional new heme iron coordination states.

**Acknowledgments**—We are grateful to Prof. A. J. Thomson (Fellow of the Royal Society) and Prof. D. J. Richardson for use of the Centre for Metalloprotein Spectroscopy and Biology facilities at the University of East Anglia.

#### REFERENCES

- Munro, A. W., and Lindsay, J. G. (1996) *Mol. Microbiol.* **20**, 1115–1125
- Guengerich, F. P. (1991) *J. Biol. Chem.* **266**, 10019–10022
- Poulos, T. L., Finzel, B. C., and Howard, A. J. (1986) *Biochemistry* **25**, 5314–5322
- Leys, D., Mowat, C. G., McLean, K. J., Richmond, A., Chapman, S. K., Walkinshaw, M. D., and Munro, A. W. (2003) *J. Biol. Chem.* **278**, 5141–5147
- Wang, M., Roberts, D. L., Paschke, R., Shea, T. M., Masters, B. S. S., and Kim, J.-J. P. (1997) *Proc. Natl. Acad. Sci. U. S. A.* **94**, 8411–8416
- Pochapsky, T. C., Jain, N. U., Kuti, M., Lyons, T. A., and Heymont, J. (1999) *Biochemistry* **38**, 4681–4690
- Muller, J. J., Lapko, A., Bourenkov, G., Ruckpaul, K., and Heinemann, U. (2001) *J. Biol. Chem.* **276**, 2786–2789
- Lee, D. S., Yamada, A., Sugimoto, H., Matsunaga, I., Ogura, H., Ichihara, K., Adachi, S., Park, S. Y., and Shiro, Y. (2003) *J. Biol. Chem.* **278**, 9761–9767
- Munro, A. W., Leys, D. G., McLean, K. J., Marshall, K. R., Ost, T. W. B., Daff, S., Miles, C. S., Chapman, S. K., Lysek, D. A., Moser, C. C., Page, C. C., and Dutton, P. L. (2002) *Trends Biochem. Sci.* **27**, 250–257
- Narhi, L. O., and Fulco, A. J. (1987) *J. Biol. Chem.* **262**, 6683–6690
- Fulco, A. J. (1991) *Annu. Rev. Pharmacol. Toxicol.* **31**, 177–203
- Daff, S. N., Chapman, S. K., Turner, K. L., Holt, R. A., Govindaraj, S., Poulos, T. L., and Munro, A. W. (1997) *Biochemistry* **36**, 13816–13823
- Sigar, S. G., and Gunsalus, I. C. (1976) *Proc. Natl. Acad. Sci. U. S. A.* **73**, 1078–1082
- Groves, J. T., and Han, Y.-Z. (1995) in *Cytochrome P450: Structure, Mechanism and Biochemistry* (Ortiz de Montellano, P. R., ed) pp. 3–48, Plenum Publishing Corp., New York
- McLean, M. A., Maves, S. A., Weiss, K. E., Krelich, S., and Sligar, S. G. (1998) *Biochem. Biophys. Res. Commun.* **252**, 166–172
- Munro, A. W., Lindsay, J. G., Coggins, J. R., Kelly, S. M., and Price, N. C. (1996) *Biochim. Biophys. Acta* **1296**, 127–137
- Hui Bon Hoa, G., McLean, M. A., and Sligar, S. G. (2002) *Biochim. Biophys. Acta* **1598**, 297–308
- Perera, R., Sono, M., Sigman, J. A., Pfister, T. D., Lu, Y., and Dawson, J. H. (2003) *Proc. Natl. Acad. Sci. U. S. A.* **100**, 3641–3646
- Modi, S., Primrose, W. U., and Roberts, G. C. K. (1995) *Biochem. J.* **310**, 939–943
- Bell, S. G., Chen, X., Xu, F., Rao, Z., and Wong, L. L. (2003) *Biochem. Soc. Trans.* **31**, 558–562
- Ost, T. W. B., Miles, C. S., Murdoch, J., Cheung, Y., Reid, G. A., Chapman, S. K., and Munro, A. W. (2000) *FEBS Lett.* **486**, 173–177
- Glieder, A., Farinas, E. T., and Arnold, F. H. (2002) *Nat. Biotechnol.* **11**, 1135–1139
- Henne, K. R., Kunze, K. L., Zheng, Y.-M., Christmas, P., Soberman, R. J., and Rettie, A. E. (2001) *Biochemistry* **40**, 12925–12931
- Hoch, U., and Ortiz de Montellano, P. R. (2001) *J. Biol. Chem.* **276**, 11339–11346
- LeBrun, L., Hoch, U., and Ortiz de Montellano, P. R. (2002) *J. Biol. Chem.* **277**, 12755–12761
- Zheng, Y.-M., Baer, B. R., Kneller, B., Henne, K. R., Kunze, K. L., and Rettie, A. E. (2003) *Biochemistry* **42**, 4601–4606
- Oliver, C. F., Modi, S., Primrose, W. U., Lian, L. Y., and Roberts, G. C. K. (1997) *Biochem. J.* **327**, 537–544
- Carmichael, A. B., and Wong, L. L. (2001) *Eur. J. Biochem.* **268**, 3117–3125
- Ravichandran, K. G., Bodduvalli, S. S., Hasemann, C. A., Peterson, J. A., and Deisenhofer, J. (1993) *Science* **261**, 731–736
- Miles, J. S., Munro, A. W., Rospendowski, B. R., Smith, E. W., McKnight, J., Thomson, A. J. (1992) *Biochem. J.* **288**, 503–509
- Noble, M. A., Miles, C. S., Chapman, S. K., Lysek, D. A., Mackay, A. C., Reid, G. A., Hanzlik, R. P., and Munro, A. W. (1999) *Biochem. J.* **339**, 371–379
- Thomas, P. E., Ryan, D., and Levin, W. (1976) *Anal. Biochem.* **75**, 168–176
- Mowat, C. G., Moyses, R., Miles, C. S., Ley, D., Doherty, M. K., Taylor, P., Walkinshaw, M. D., Reid, G. A., and Chapman, S. K. (2001) *Biochemistry* **40**, 12292–12298
- Sambrook, J., Fritsch, E. F., and Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual*, 2nd Ed., Cold Spring Harbor Laboratory, Cold Spring Harbor, NY
- Munro, A. W., Noble, M. A., Robledo, L., Daff, S. N., and Chapman, S. K. (2001) *Biochemistry* **40**, 1956–1963
- Dutton, P. L. (1978) *Methods Enzymol.* **54**, 411–435
- Ost, T. W. B., Miles, C. S., Munro, A. W., Murdoch, J., Reid, G. A., and Chapman, S. K. (2001) *Biochemistry* **40**, 13421–13429
- Daff, S., Sharp, R. E., Short, D. M., Bell, C., White, P., Manson, F. D. C., Reid, G. A., and Chapman, S. K. (1996) *Biochemistry* **35**, 6351–6357
- Thomson, A. J., Cheesman, M. R., and George, S. J. (1993) *Methods Enzymol.* **226**, 199–232
- Munro, A. W., Daff, S., Coggins, J. R., Lindsay, J. G., and Chapman, S. K. (1996) *Eur. J. Biochem.* **239**, 403–409
- Ruettinger, R. T., and Fulco, A. J. (1981) *J. Biol. Chem.* **256**, 5728–5734
- Buchanan, J. F., and Fulco, A. J. (1978) *Biochem. Biophys. Res. Commun.* **85**, 1254–1260
- Narhi, L. O., and Fulco, A. J. (1986) *J. Biol. Chem.* **261**, 7160–7169
- Macheroux, P., Massey, V., Thiele, D. J., and Volokita, M. (1991) *Biochemistry* **30**, 4612–4619
- McKnight, J., Cheesman, M. R., Thomson, A. J., Miles, J. S., and Munro, A. W. (1993) *Eur. J. Biochem.* **213**, 683–687
- Dawson, J. H., Andersson, L. A., and Sono, M. (1982) *J. Biol. Chem.* **257**, 3606–3617
- Lipscomb, J. D. (1980) *Biochemistry* **19**, 3590–3599
- Tsai, A. L., Berka, V., Chen, P.-F., and Palmer, G. (1996) *J. Biol. Chem.* **271**, 32563–32571
- Sono, M., and Dawson, J. H. (1982) *J. Biol. Chem.* **257**, 5496–5502
- Sono, M., Hager, L. P., and Dawson, J. H. (1991) *Biochim. Biophys. Acta* **1078**, 351–359
- Cheesman, M. R., Greenwood, C., and Thomson, A. J. (1991) *Adv. Inorg. Chem.* **36**, 201–255
- Berka, V., Palmer, G., Chen, P.-F., and Tsai, A. L. (1998) *Biochemistry* **37**, 6136–6144
- Shimizu, T., Nozawa, T., Hatano, M., Imai, Y., and Sato, R. (1975) *Biochemistry* **14**, 4172–4178
- Vickery, L., Salmon, A., and Sauer, K. (1975) *Biochim. Biophys. Acta* **386**, 87–98
- Dawson, J. H., Sono, M., and Hager, L. P. (1983) *Inorg. Chim. Acta* **79**, 184–186
- Svastis, E. W., Alberta, J. A., Kim, I.-C., and Dawson, J. H. (1989) *Biochem. Biophys. Res. Commun.* **165**, 1170–1176
- Sigman, J. A., Pond, A. E., Dawson, J. H., and Lu, Y. (1999) *Biochemistry* **38**, 11122–11129
- Rux, J. J., and Dawson, J. H. (1991) *FEBS Lett.* **290**, 49–51
- Vickery, L., Nozawa, T., and Sauer, K. (1976) *J. Am. Chem. Soc.* **98**, 351–357
- Dhawan, I. K., Shelver, D., Thorsteinsson, D., Roberts, G. P., and Johnson, M. K. (1999) *Biochemistry* **38**, 12805–12813
- Gadsby, P. M. A., and Thomson, A. J. (1990) *J. Am. Chem. Soc.* **112**, 5003–5011
- Cheesman, M. R., Little, P. J., and Berka, B. C. (2001) *Biochemistry* **40**, 10562–10569
- Nozawa, T., Shimizu, T., Hatano, M., Shimada, S., Izuka, T., and Ishimura, Y. (1978) *Biochim. Biophys. Acta* **534**, 285–294
- Bell, S. G., Stevenson, J. A., Boyd, H. D., Campbell, S., Riddle, A. D., Orton,

*Novel Heme Ligation in a Cytochrome P450*

- E. L., and Wong, L. L. (2002) *Chem. Commun.* **7**, 490–491
65. Li, H., and Poulos, T. L. (1997) *Nat. Struct. Biol.* **4**, 140–146
66. Oliver, C. F., Modi, S., Sutcliffe, M. J., Primrose, W. U., Lian, L. Y., and Roberts, G. C. K. (1997) *Biochemistry* **36**, 1567–1572
67. Sutherland, J., Greenwood, C., Peterson, J., and Thomson, A. J. (1986) *Biochem. J.* **233**, 893–898
68. Cheesman, M. R., Thomson, A. J., Greenwood, C., Moore, G. R., and Kadir, F. (1990) *Nature* **346**, 771–773
69. Arnesano, F., Banci, L., Bertini, I., and Felli, I. C. (1998) *Biochemistry* **37**, 173–184
70. Modi, S., Sutcliffe, M. J., Primrose, W. U., Lian, L. Y., and Roberts, G. C. K. (1996) *Nat. Struct. Biol.* **3**, 414–417
71. Noble, M. A., Quaroni, L., Chumanov, G. D., Turner, K. L., Chapman, S. K., Hanzlik, R. P., and Munro, A. W. (1998) *Biochemistry* **37**, 15799–15807
72. Joyce, M. G., Girvan, H. M., Munro, A. W., and Leys, D. (2004) *J. Biol. Chem.* **279**, 23287–23293

# Flavocytochrome P450 BM3: an update on structure and mechanism of a biotechnologically important enzyme

A.J. Warman\*, O. Roitel\*, R. Neeli\*, H.M. Girvan\*, H.E. Seward\*, S.A. Murray\*, K.J. McLean\*, M.G. Joyce\*, H. Toogood\*, R.A. Holt†, D. Leys\*, N.S. Scrutton\* and A.W. Munro\*<sup>1</sup>

\*Department of Biochemistry, University of Leicester, The Henry Wellcome Building, Lancaster Road, Leicester LE1 9HN, U.K., and †Avecia Biotechnology, Belasis Avenue, Billingham, Cleveland TS23 1YN, U.K.

## Abstract

Since its discovery in the 1980s, the fatty acid hydroxylase flavocytochrome P450 (cytochrome P450) BM3 (CYP102A1) from *Bacillus megaterium* has been adopted as a paradigm for the understanding of structure and mechanism in the P450 superfamily of enzymes. P450 BM3 was the first P450 discovered as a fusion to its redox partner – a eukaryotic-like diflavin reductase. This fact fuelled the interest in soluble P450 BM3 as a model for the mammalian hepatic P450 enzymes, which operate a similar electron transport chain using separate, membrane-embedded P450 and reductase enzymes. Structures of each of the component domains of P450 BM3 have now been resolved and detailed protein engineering and molecular enzymology studies have established roles for several amino acids in, e.g. substrate binding, coenzyme selectivity and catalysis. The potential of P450 BM3 for biotechnological applications has also been recognized, with variants capable of industrially important transformations generated using rational mutagenesis and forced evolution techniques. This paper focuses on recent developments in our understanding of structure and mechanism of this important enzyme and highlights important problems still to be resolved.

## Introduction

Since their discovery in the 1950s, the P450s (cytochromes P450) have been studied in enormous detail due to their involvement in a plethora of crucial cellular roles – from human drug and steroid metabolism through to the bacterial catabolism of unusual compounds as energy sources [1,2]. The P450s are haem *b*-containing mono-oxygenase enzymes, often catalysing hydroxylation of hydrophobic substrate molecules. Molecular oxygen is bound to the haem iron and reductively activated to produce an oxyferryl radical cation intermediate. This ‘active oxygen’ compound I species attacks a substrate bound adjacent to the haem. Recent studies suggest that a preceding intermediate in the catalytic cycle (the ferric peroxy form or compound 0) may also be catalytically competent, at least with respect to epoxidation across carbon–carbon double bonds [3]. However, computational studies indicate that compound I is a superior oxidant even for alkene epoxidation [4]. The key to the P450s’ ability to cleave molecular oxygen is the axial ligation of haem iron by a conserved cysteine residue in the protein. The thiolate ligation is shared among the P450, NOS (nitric oxide synthase) and chloroperoxidase enzyme families – members of which all activate oxygen [5]. The structure and cellular roles of P450

enzymes are discussed in detail elsewhere [1,2,6], and recent developments in characterization of P450 redox partner proteins are discussed in the accompanying paper by McLean et al. [7].

The prokaryotic P450s are soluble enzymes and interact with soluble redox partners [7]. In contrast, eukaryotic P450s are, almost without exception, integral membrane proteins. The prokaryotic forms have proven easier to express and purify, and reconstitution of their activities with their soluble redox partners is also a simpler proposition than for the eukaryotic forms. The major redox partner for the eukaryotic P450s is the diflavin enzyme NADPH-CPR (where CPR stands for cytochrome P450 reductase), which is also an integral membrane protein [8]. Many early studies on structure, spectroscopy and enzymology of cytochrome P450 focused on the *Pseudomonas putida* cytochrome P450 cam enzyme (CYP101). P450 cam is an essential component of a plasmid-borne catabolic pathway for breakdown of *D*-camphor as an energy source for the bacterium [9]. A huge volume of the fundamental work establishing the nature of cytochrome P450 was done with P450 cam, including the solution of the first P450 atomic structure and pivotal studies on the thermodynamic regulation of P450 catalysis [10,11]. A problem perceived with respect to extrapolation of data from the class I P450 cam system to the medically relevant class II drug-metabolizing P450 forms was that the redox partners used were distinctly different. P450 cam is the terminal oxidase in a three-component redox chain, with electrons delivered from NADH to the FAD-containing

**Key words:** biocatalysis, coenzyme selectivity, flavocytochrome, fusion enzyme, haem ligand, P450 BM3.

**Abbreviations used:** BM3, *Bacillus megaterium* flavocytochrome P450 BM3; CPR, cytochrome P450 reductase; NR1, novel reductase 1; P450, cytochrome P450; SB, substrate (palmitoleate)-bound; SF, substrate-free.

<sup>1</sup>To whom correspondence should be addressed (email awm9@le.ac.uk).

putidaredoxin reductase, and on to the P450 by the 2Fe-2S ferredoxin putidaredoxin [9]. The eukaryotic class II P450s interact with the FAD- and FMN-containing CPR, which sources its electrons from NADPH [12]. Questions regarding e.g. structural differences in class I and II P450s to accommodate binding of the distinct redox partners and potential differences in electron transfer mechanism gave rise to nagging doubts as to the universal applicability of P450 cam. What might be more desirable as a tractable model for the class II P450s would be a soluble bacterial P450 with a CPR redox partner. The prokaryotic P450 that ultimately attained the status of a 'model system' for the class II eukaryotic P450s fulfilled this requirement, and had the additional advantage that the CPR was fused to the P450 in a single polypeptide chain. This paper will focus on the BM3 (*Bacillus megaterium* flavocytochrome P450 BM3) fusion system, highlighting recent advances in the structural and mechanistic characterizations of this intriguing enzyme.

### Flavocytochrome P450 BM3: isolation, structural and catalytic properties

Miura and Fulco from the University of California at Los Angeles first isolated P450 BM3 as a fatty acid oxygenating enzyme activity from the cytosol of the soil bacterium *B. megaterium* [13]. The enzyme was shown to catalyse hydroxylation of saturated fatty acids of various chain lengths, close to the  $\omega$ -terminal (typically  $\omega - 1$ ,  $\omega - 2$  and  $\omega - 3$  positions). The enzyme was isolated and named P450 BM3 – in recognition of the fact that two previous P450 enzymes have been identified in the bacterium. Subsequent studies demonstrated the multidomain nature of the enzyme and ultimately proved that the enzyme was a *bona fide* fusion of soluble P450 (N-terminal) to CPR, as predicted on the basis of the gene sequence [14,15]. P450 BM3 is reduced by NADPH, with electrons transferred to the FAD cofactor in its CPR domain. Electrons are transferred singly from FAD to FMN within the CPR domain and then on to the haem iron in the SB (substrate-bound) P450 domain. Enzyme turnover studies demonstrated that flavocytochrome P450 BM3 had very high oxygenase activity with a range of fatty acids ( $> 15\,000\text{ min}^{-1}$  with arachidonic acid). Transient kinetic studies revealed that rapid rates of electron transfer both within the reductase domain and between the reductase and the P450 underlie the catalytic efficiency of the enzyme [16]. Molecular dissection of the enzyme produced haem-containing (P450) and diflavin (reductase) domains that retained UV-visible absorption and other spectroscopic properties characteristic of the flavocytochrome [17,18]. However, molecular interactions between isolated domains and the reductase-to-P450 electron transfer rate were disrupted such that catalytic rates of fatty acid hydroxylation were diminished several fold [17].

Thereafter, the race was on to determine structures for the enzyme and its domains. Atomic structures of the haem domain of the enzyme were solved in SF (substrate-free) and palmitoleic acid-bound forms [19,20]. These structures indicated that a major structural rearrangement occurs be-

tween the two forms. However, whether this rearrangement is a direct consequence of substrate docking to the P450 is now uncertain given recent structural data on an A264E (Ala<sup>264</sup> → Glu) mutant form of the haem domain ([21] and see below). In the SB form, the substrate is tethered at the mouth of the active site, held by electrostatic and hydrogen-bonding interactions between its carboxylate group and the side chains of residues Arg<sup>47</sup> and Tyr<sup>51</sup>. The remainder of the alkyl chain extends into the active site and interacts with several hydrophobic residues [20]. The  $\omega$ -terminus of the fatty acid is located distant from the haem iron, and movement of substrate closer to the haem iron was inferred from NMR studies on a sodium dithionite-reduced form of the fatty acid-bound haem domain [22]. Such movement is essential to relocate the substrate in order to position the relevant  $\omega - 1$  to  $\omega - 3$  target atoms in the immediate vicinity of compound I. However, it is not immediately obvious why the simple reduction of ferric iron to ferrous at the haem core should be sufficient to induce substrate relocation. Perhaps a more likely scenario is that the haem domain of the flavocytochrome undergoes conformational rearrangements, and that the equilibrium between conformational forms is affected by substrate binding and, potentially, by interaction with the redox partner. Substrate may re-position in different conformers or else be mobile within the active site. This hypothesis would also be consistent with recent structural studies on P450 BM3 mutants that exhibit novel haem iron ligand sets [21,23].

To date, crystallization and resolution of the atomic structure of the full-length flavocytochrome P450 BM3 has proven impossible, probably because of the multidomain nature of the enzyme and the presence of flexible linker regions between domains. A construct encoding the linked haem and FMN domains was created, and the purified FMN-haem domain protein crystallized. The structure was determined, but revealed proteolytic scission between the haem and FMN domains, and a non-stoichiometric complex of one FMN domain to two haem domains [24]. An extended sigma bond pathway for electron transfer between FMN and haem iron was hypothesized based on the structure, but distances involved are large and a more direct intercofactor electron transfer pathway is likely [24,25]. Notwithstanding the problems of proteolysis and non-stoichiometry of domains, the FMN domain is seen to dock in a depression in the haem domain on the proximal face of the haem iron [6,24]. In addition, the structure of the FMN domain was critical for explaining why the neutral, blue semiquinone form of the flavin is not stabilized by the BM3 FMN cofactor. The BM3 FMN is closely related to prokaryotic flavodoxins, which strongly stabilize the semiquinone form [25]. The semiquinone in the P450 BM3 FMN cofactor is destabilized in favour of the quinol form, probably as a result of altered charge distribution around the flavin and rigidity of the FMN-binding loop region by comparison with flavodoxins [24]. FMNs' tendency to become 'over-reduced' to the quinol form has interesting connotations with respect to catalytic activity of the enzyme, as discussed below.

## The reductive inactivation enigma

Studies by Narhi and Fulco [15] revealed that fatty acid hydroxylase activity of P450 BM3 was inhibited progressively following preincubation of the enzyme with NADPH. No such inhibition occurs if fatty acid and NADPH are mixed simultaneously with the enzyme. Moreover, no significant inhibition of cytochrome *c* reduction rate (a process that requires electron transfer from the FMN cofactor) was observed on prereduction of P450 BM3 by NADPH, and fatty acid hydroxylation activity could be restored by oxidation of the prereduced enzyme [25]. The hypothesis developed was that the inactivation of fatty acid hydroxylation resulted from the development of FMN quinol as a 3-electron (or 4-electron) reduced form of the reductase domain accumulated. EPR analysis of P450 BM3 samples taken during active fatty acid oxygenation indicated that two semiquinone species (one a blue, neutral form and the other a red, anionic semiquinone) were present [26]. Potentiometric studies had revealed that the FAD flavin stabilized a blue semiquinone, thus the red semiquinone was proposed to form on the FMN [27]. Recently, stopped-flow studies were performed to analyse the dithionite-dependent reduction of the isolated BM3 FMN domain. Dithionite is a one-electron donor, and clear spectral evidence was obtained for transient formation of the red semiquinone on the FMN, before electronic disproportionation to form the thermodynamically favoured quinol [28]. It appears likely, then, that the formation of FMN quinol is at least partly responsible for the NADPH-dependent inactivation of P450 BM3 fatty acid hydroxylation. However, the phenomenon is by no means well understood. Further detailed kinetic and thermodynamic studies are required to ascertain whether, e.g. the FMN quinol is simply an inefficient electron donor to the P450 haem iron, or whether conformational alterations occur as a consequence of 'over-reduction', perhaps as a means of regulating enzyme activity if availability of fatty acid substrates is limited.

## Completing the 'structural jigsaw' of flavocytochrome P450 BM3

With the structure of the haem domain of P450 BM3 solved in various forms [19,20,23,29] and the FMN domain also determined in complex with the haem domain [24], the last remaining segment of the flavocytochrome to be structurally characterized was the FAD/NADPH binding domain. The FAD domain is related to ferredoxin reductases and to the FAD domain of rat CPR, which has been structurally characterized [8]. In recent work, our group has crystallized the isolated FAD domain of P450 BM3 and resolved the atomic structure in the presence and absence of NADP<sup>+</sup> ligand (M.G. Joyce, R. Neeli, O. Roitel, A.W. Munro and D. Leys, unpublished work). The structure details the key residues involved in binding both FAD and NADPH, and (by analogy with the rat CPR structure) the interaction surface with the FMN domain. Modelling of the FAD domain's interface with the FMN domain indicates that the cofactors can be placed in close proximity without any significant

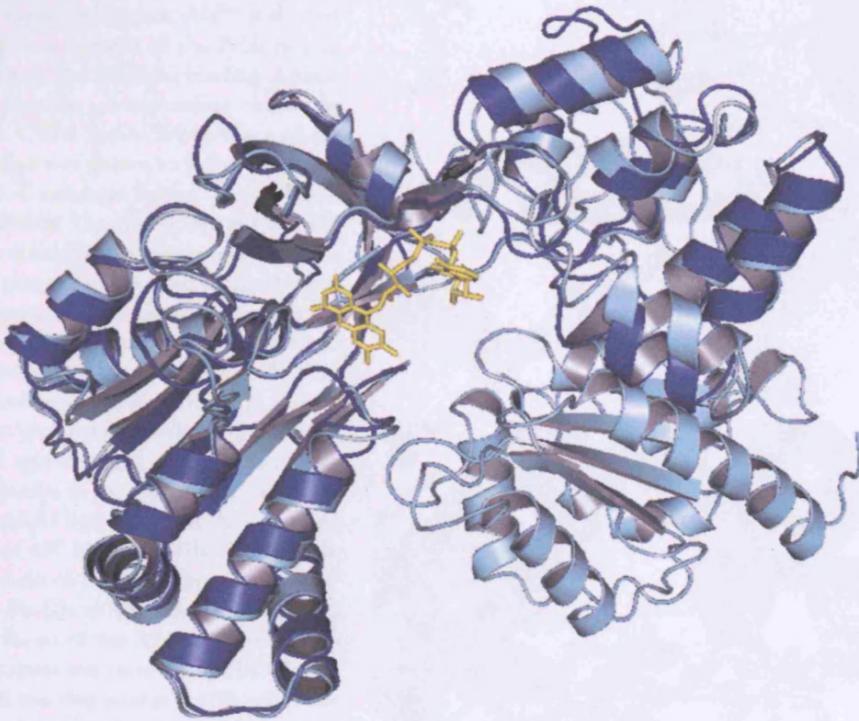
steric obstruction or unfavourable chemical interaction, and the structure overlays well with the rat CPR FAD domain (Figure 1). The likelihood is that electron transfer occurs directly between FAD and FMN cofactors in P450 BM3 (as would be expected for this highly efficient electron transfer system). Structural reorganization is almost certainly required to facilitate interaction of the FMN domain with the haem domain. Flexible linker regions between haem and FMN domains, and between FAD and FMN domains, probably facilitate interdomain communication [30].

## Deconvoluting enzyme mechanism using protein engineering

Rational mutagenesis has been used extensively to probe roles of various amino acids in substrate/ligand binding and catalysis. In the first mutagenesis study on BM3 (before structural characterization of the haem domain), the role of Trp<sup>96</sup> was investigated. This conserved residue was suggested to have a key role in electron transfer across the P450s [31]. Analysis of various P450 BM3 Trp<sup>96</sup> point mutants revealed that the residue was instead important for haem binding and for control of haem iron spin-state [32]. The subsequent atomic structure of the haem domain demonstrated a hydrogen bonding interaction between Trp<sup>96</sup> and haem propionate [19]. Phe<sup>393</sup>, another phylogenetically conserved P450 residue, was shown to be a regulator of the thermodynamic properties of the haem iron. The more positive reduction potential measured for the haem iron in F393A and F393H mutants was consistent with the faster flavin-to-haem electron transfer rates measured by stopped-flow methods. However, reductive activation of oxygen was severely compromised in the mutants, explaining diminished fatty acid hydroxylation rates and enabling rationalization of similar effects observed on catalytic turnover in comparable mutants of other P450 enzymes [33]. Another important amino acid in P450 BM3 is Phe<sup>87</sup>, the side chain of which lies above the haem plane and in proximity to the site for oxygen binding to the haem iron. The atomic structure of palmitoleate-bound BM3 haem domain indicated that the interaction of the terminal methyl group of substrate with the phenyl side chain of Phe<sup>87</sup> protected the former from oxidative attack by compound I, resulting in hydroxylation away from the  $\omega$ -methyl group [20]. NMR studies of lauric acid turnover in a F87A mutant suggested that the effect of the mutation was to convert the enzyme into a near-exclusive fatty acid  $\omega$ -hydroxylase (i.e. with negligible hydroxylation further down the lipid chain) [34]. However, later studies suggested that the position of oxidative attack was not at the  $\omega$ -terminal and may in fact be removed further down the alkyl chain by replacement of Phe<sup>87</sup> with non-aromatic residues [35]. The latter results are perhaps more consistent with the notion that Phe<sup>87</sup> is a tether for the  $\omega$ -methyl of fatty acid substrates. In this model, greater lateral motion of the fatty acid chain can occur when the tether is removed, enabling oxidative attack at positions further from the  $\omega$ -terminal of the molecule. As discussed in the next section, an A264E

**Figure 1 | Overlay of the structure of flavocytochrome P450 BM3 FAD domain with rat CPR**

The recent structural solution of the FAD domain of P450 BM3 allows for structural comparisons with the homologous domain of rat CPR (M.G. Joyce, R. Neeli, O. Roitel, A.W. Munro and D. Leys, unpublished work). The BM3 FAD domain (dark blue) is overlaid with the rat CPR structure (light blue), and the similar positioning of the major structural elements is clear. The non-overlaid section (light blue) is that of the rat CPR FMN domain. The BM3 FAD cofactor is seen in yellow at the centre of the image. Structural analysis of rat CPR demonstrated that the FAD and FMN cofactors were juxtaposed only 4 Å (1 Å = 0.1 nm) apart, consistent with direct electron transfer between the cofactors [8]. The P450 BM3 FAD and FMN domains can be positioned in similar conformations to those of the respective rat CPR domains without significant steric clashes, consistent with the established rapidity of the redox reactions in this enzyme [37].



mutant in the P450 BM3 haem domain has also had far-reaching consequences for our mechanistic understanding of the enzyme.

In the reductase domain of P450 BM3, several mutagenic studies have revealed, e.g. key residues involved in cofactor and coenzyme binding [36,37]. A 'catalytic triad' of residues (Ser<sup>830</sup>, Cys<sup>999</sup> and Asp<sup>1044</sup>) in P450 BM3 is extensively conserved in the diflavin reductase family to which the CPR enzymes belong. Recent research on C999A variants of BM3 reductase and FAD domains showed that the rate of hydride transfer from NADPH to FAD was substantially decreased, but that thermodynamic properties of the flavin were not significantly altered. Also, a stable charge-transfer species formed on reaction between NADPH and the wild-type FAD domain was not observed for the C999A FAD domain. These results were consistent with previous studies of mammalian CPR, and indicated a role for Cys<sup>999</sup> both in enhancement of hydride transfer rate to FAD cofactor and in stabilizing the

FADH<sub>2</sub>-NADP<sup>+</sup> complex of P450 BM3 [37]. Interestingly, the triad cysteine is absent from the wild-type form of human NR1 (novel reductase 1), a cancer-related diflavin reductase [38]. NR1 is the slowest of all the diflavin reductases, with a limiting rate of NADPH-to-FAD hydride transfer ( $k_{lim}$ ) of only approx. 1 s<sup>-1</sup> at 25°C (compared with 237 s<sup>-1</sup> for P450 BM3 reductase at 5°C). In a recent work, we have engineered a triad cysteine replacement (A549C) variant of NR1 FAD domain and demonstrated a dramatic increase in hydride transfer in this mutant (O. Roitel, J. Basran, R.D. Finn, M.J. Paine, N.S. Scrutton and A.W. Munro, unpublished work). In further recent engineering studies of BM3 reductase and FAD domains, an aromatic amino acid that stacks over the FAD cofactor (Trp<sup>1046</sup>) was replaced with both histidine and alanine residues. The mutants showed a huge switch in coenzyme specificity from NADPH towards NADH (~6000-fold with the W1046A FAD domain) [39]. NADH is a substantially cheaper coenzyme than NADPH, and these

mutants have potential for exploitation in biotechnological processes (see below).

### Conformational equilibria and catalytic control in the BM3 haem domain

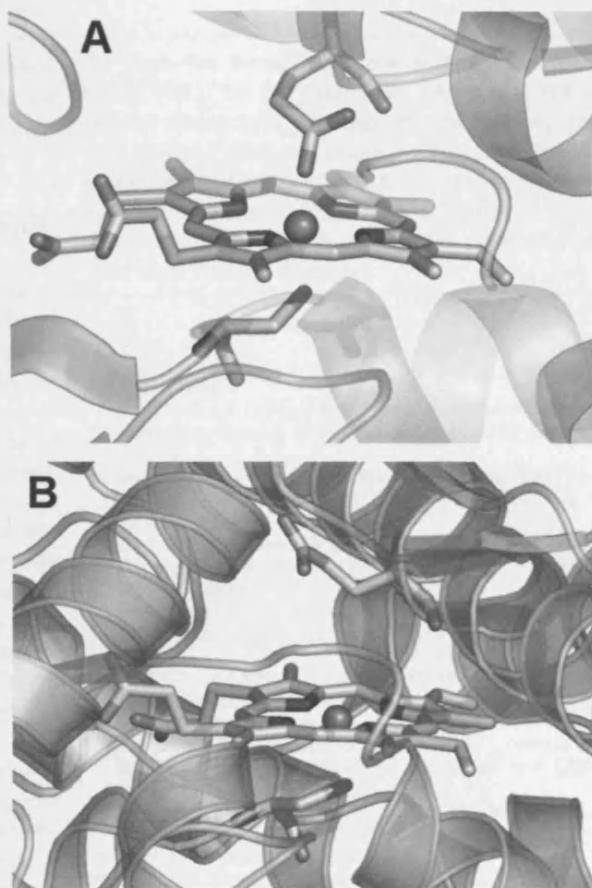
Structures of SF and SB forms of the BM3 haem domain revealed a major structural rearrangement in the P450 [19,20]. A substrate-induced conformational switch was inferred. However, recent structural analysis of the A264E mutant haem domain has offered a different explanation. Ala<sup>264</sup> is located in the I helix of P450 BM3, a segment of the P450 rich in residues with roles in catalysis and substrate binding. Amino acid alignments showed that the corresponding residue in several related eukaryotic CYP4 family P450s was a glutamate. This glutamate residue was shown to form (in a turnover-dependent manner) a covalent bridge to the haem 5-methyl group in these P450s. The A264E mutant of P450 BM3 did not form a covalent bridge, but spectroscopic studies indicated instead that the glutamate side chain replaced water as an axial ligand to the haem iron in a proportion of the SF form of the A264E haem domain. Addition of fatty acids to the mutant forced this glutamate ligand on/off equilibrium towards a completely glutamate-ligated form [21]. Atomic structures were solved for SF and SB A264E haem domains, confirming the results of spectroscopic analysis. In the SF form, there were two molecules in the asymmetric unit. One had the 'normal' Cys-Fe-H<sub>2</sub>O ligand set and the other had Cys-Fe-Glu. In the 'ligand off' form, the Glu<sup>264</sup> side chain interacted with the side chain of Phe<sup>87</sup>. In the palmitoleate-bound form, a single Cys-Fe-Glu set was observed (Figure 2) [23]. In both SF and SB forms of the A264E haem domain (regardless of whether the haem was co-ordinated by Glu<sup>264</sup>), the conformation adopted was that seen in the SB wild-type enzyme. Fatty acid substrates were found to bind tighter to the A264E haem domain compared with wild-type BM3, suggesting that the mutation itself had altered the conformational equilibrium between SF and SB forms in solution (regardless of the presence of substrate) and that fatty acids bind more tightly to (and stabilize) the SB conformation. Thus we conclude that both SF and SB conformations are present in solutions of the fatty acid-free BM3 haem domain, and that the equilibrium between these forms can be altered by protein mutations. The SB conformation is that which binds more tightly to substrates, rather than one that forms following substrate association [23]. The fact that the A264E enzyme retains considerable fatty acid hydroxylase activity points strongly to further structural reorientations (that displace Glu<sup>264</sup> and enable oxygen binding to haem iron) that occur subsequent to interaction with the CPR domain and/or haem iron reduction.

### Biotechnological exploitation of flavocytochrome P450 BM3

Several P450s (particularly the bacterial forms) are region- and stereoselective mono-oxygenases. As such, they have the potential to catalyse the production of high-value oxygenated

**Figure 2 | Active site structure in the SF A264E haem domain of P450 BM3**

The structure reveals two molecules (**A**, **B**) in the asymmetric unit cell. (**A**) Axial co-ordination of the haem iron by both Cys<sup>400</sup> (proximal) and the Glu<sup>264</sup> side chain (distal) and (**B**) the interaction of glutamate instead with the phenyl side chain of residue Phe<sup>87</sup>. In (**B**) water is the distal ligand to the haem iron. Both molecules adopt the protein conformation previously observed for SB wild-type P450 BM3 haem domain [23].



organic molecules for use as, e.g. chiral synthons in synthetic pathways. Enzyme-mediated product formation is a relatively 'clean' technique that has obvious advantages over 'traditional' organic chemistry methods. Large excesses of desired chiral products can be formed enzymatically. P450 BM3 is an obvious target for such applications, since it is a very high activity, single-component enzyme. Numerous protein engineering studies have demonstrated the pliability of its active site and excellent structural data provide the basis for further engineering to promote recognition of desired substrates [30]. Recent work producing NADH-specific variants make P450 BM3 even more attractive and cost-effective for biotechnological processes [39]. While the field is still in its relative infancy, significant steps have been made to create biotechnologically more applicable variants. For instance, rational mutagenesis of the central portion of the BM3

substrate-binding site produced mutants with considerably enhanced activities such as short chain ( $C_4$ – $C_8$ ) alkanoic acid hydroxylases [40]. A forced evolution approach generated a high activity alkane hydroxylase with several mutations scattered throughout the protein scaffold [41]. In recent work, our group has characterized reactivity of wild-type and Phe<sup>87</sup> mutant forms of P450 BM3 with several alkenes. Rates of more than 1000 turnovers · min<sup>-1</sup> were obtained with a number of substrates, and different product profiles were obtained with wild-type and mutant forms (S.A. Murray, R.A. Holt, N.S. Scrutton and A.W. Munro, unpublished work). P450 BM3 clearly has several applications in the biotechnology arena, and its potential will be increasingly tapped in the coming years.

### 'We are not alone' – BM3 homologues and other P450:reductase fusion enzymes

For several years, P450 BM3 stood alone as the only characterized P450:reductase fusion enzyme – casting some doubt on the broader importance of this type of fusion enzyme in biology. However, the discovery that NOS, a key cell-signalling enzyme in eukaryotes, was also a flavocytochrome of similar construction to P450 BM3 (cysteinate-co-ordinated haem-containing oxygenase domain fused to a CPR module), provided the first clue that this type of P450:CPR fusion would be widespread in nature [42]. Subsequently, the BM3-type enzyme has been recognized in various prokaryotes and lower eukaryotes, with genome sequences continuing to reveal new members for the CYP102 family of BM3 (e.g. from *Bacillus anthracis*, *Bradyrhizobium japonicum*, *Streptomyces avermitilis* and *Gibberella moniliformis*). The only other direct homologues characterized at the protein level are CYP102A2 and CYP102A3 from *Bacillus subtilis*. Both are also NADPH-dependent fatty acid hydroxylases, but display distinct differences in substrate specificity compared with P450 BM3 – with strong selectivity for long-chain unsaturated and branched-chain fatty acids [43]. Branched-chain fatty acids are prevalent in the *B. subtilis* membrane. A P450 BM3 homologue is present in the heavy metal-tolerant bacterium *Ralstonia metallidurans*, which also encodes a novel type of P450:reductase fusion enzyme. The reductase in this new P450 class resembles the FMN- and 2Fe–2S-containing phthalate dioxygenase reductase, and the P450 has significant similarity with a *Rhodococcus* sp. P450 known to degrade thiocarbamate herbicides [44]. Doubtless, genome sequencing studies will continue to reveal new types of P450:reductase fusion enzymes that have evolved to provide efficient P450 oxygenase activities tailored to the specific physiological requirements of the host organism.

### Summary and future prospects

P450 BM3 has been, and continues to be, an important model system for the comprehension of structure–function–mechanism relationships in the P450 superfamily. In addition, BM3 has clear biotechnological potential [30,40,41]. Despite advances in structural biology of mammalian cytochromes

P450 and their reductase systems, BM3 remains a tractable and valuable model system due to its convenient modular construction, soluble nature and the detailed understanding of its structural and catalytic properties. Several issues pertaining to, e.g. conformational dynamics and thermodynamic control over electron transfer remain to be resolved. These and other topics of broad general relevance in the P450 field maintain BM3 as a pivotal member of its superfamily.

We thank the Biotechnology and Biological Sciences Research Council (U.K.), the EC (Marie Curie Fellowship to O.R. and X-TB network Postdoctoral Fellowship to K.J.M.) and the Engineering and Physical Sciences Research Council (U.K.)/ProBio Faraday partnership (studentship to S.A.M.). D.L. is a Royal Society University Research Fellow and acknowledges the EMBO for a YIP award. A.W.M. thanks the Royal Society for the award of a Leverhulme Trust Senior Research Fellowship.

### References

- Guengerich, F.P. (2005) in Cytochrome P450: Structure, Mechanism and Biochemistry (Ortiz de Montellano, P.R., ed.), pp. 377–530, Plenum, New York
- Munro, A.W. and Lindsay, J.G. (1996) Mol. Microbiol. **20**, 1115–1125
- Vaz, A.D.N., McGinnity, D.F. and Coon, M.J. (1998) Proc. Natl. Acad. Sci. U.S.A. **95**, 3555–3560
- Shaik, S. and De Visser, S.P. (2005) in Cytochrome P450: Structure, Mechanism and Biochemistry (Ortiz de Montellano, P.R., ed.), pp. 45–85, Plenum, New York
- Groves, J.T. (2005) in Cytochrome P450: Structure, Mechanism and Biochemistry (Ortiz de Montellano, P.R., ed.), pp. 1–43, Plenum, New York
- Poulos, T.L. and Johnson, E.F. (2005) in Cytochrome P450: Structure, Mechanism and Biochemistry (Ortiz de Montellano, P.R., ed.), pp. 87–114, Plenum, New York
- McLean, K.J., Sabri, M., Marshall, K.R., Lawson, R.J., Lewis, D.G., Clift, D., Balding, P., Dunford, A.J., Warman, A.J., McVey, J.P. et al. (2005) Biochem. Soc. Trans. **33**, 796–801
- Wang, M., Roberts, D.L., Paschke, R., Shea, T.M., Masters, B.S. and Kim, J.J. (1997) Proc. Natl. Acad. Sci. U.S.A. **94**, 8411–8416
- Sligar, S.G. (1999) Essays Biochem. **34**, 71–83
- Poulos, T.L., Finzel, B.C. and Howard, A.J. (1987) J. Mol. Biol. **195**, 687–700
- Sligar, S.G. and Gunsalus, I.C. (1976) Proc. Natl. Acad. Sci. U.S.A. **73**, 1078–1082
- Iyanagi, T. and Mason, H.S. (1973) Biochemistry **12**, 2297–2308
- Miura, Y. and Fulco, A.J. (1974) J. Biol. Chem. **249**, 1880–1888
- Narhi, L.O. and Fulco, A.J. (1996) J. Biol. Chem. **261**, 7160–7169
- Narhi, L.O. and Fulco, A.J. (1987) J. Biol. Chem. **262**, 6683–6690
- Munro, A.W., Daff, S., Coggins, J.R., Lindsay, J.G. and Chapman, S.K. (1996) Eur. J. Biochem. **239**, 403–409
- Miles, J.S., Munro, A.W., Rospondowski, B.N., Smith, W.E., McKnight, J.E. and Thomson, A.J. (1992) Biochem. J. **288**, 503–509
- Li, H.Y., Darwish, K. and Poulos, T.L. (1991) J. Biol. Chem. **266**, 11909–11914
- Ravichandran, K.G., Boddupalli, S.S., Hasemann, C.A., Peterson, J.A. and Deisenhofer, J. (1993) Science **261**, 731–736
- Li, H. and Poulos, T.L. (1997) Nat. Struct. Biol. **4**, 140–146
- Girvan, H.M., Marshall, K.R., Lawson, R.J., Leys, D., Joyce, M.G., Clarkson, J., Smith, W.E., Cheesman, M.R. and Munro, A.W. (2004) J. Biol. Chem. **279**, 23274–23286
- Modi, S., Sutcliffe, M.J., Primrose, W.U., Lian, L.-Y. and Roberts, G.C.K. (1996) Nat. Struct. Biol. **3**, 414–417
- Joyce, M.G., Girvan, H.M., Munro, A.W. and Leys, D. (2004) J. Biol. Chem. **279**, 23287–23293
- Seviroukova, I.F., Li, H., Zhang, J.A., Peterson, J.A. and Poulos, T.L. (1999) Proc. Natl. Acad. Sci. U.S.A. **94**, 8411–8416

- 25 Lawson, R.J., von Wachenfeldt, C., Haq, I., Perkins, J. and Munro, A.W. (2004) *Biochemistry* **43**, 12390–12409
- 26 Murataliev, M.B. and Feyereisen, R. (2000) *Biochemistry* **39**, 5066–5074
- 27 Daff, S.N., Chapman, S.K., Turner, K.L., Holt, R.A., Govindaraj, S., Poulos, T.L. and Munro, A.W. (1997) *Biochemistry* **36**, 13816–13823
- 28 Hanley, S.C., Ost, T.W. and Daff, S. (2004) *Biochem. Biophys. Res. Commun.* **325**, 1418–1423
- 29 Haines, D.C., Tomchick, D.R., Machius, M. and Peterson, J.A. (2001) *Biochemistry* **40**, 13456–13465
- 30 Munro, A.W., Leys, D., McLean, K.J., Marshall, K.R., Ost, T.W., Daff, S., Miles, C.S., Chapman, S.K., Lysek, D.A., Moser, C.C. et al. (2002) *Trends Biochem. Sci.* **27**, 250–257
- 31 Baldwin, J.E., Morris, G.M. and Richards, W.G. (1991) *Proc. R. Soc. Lond. B Biol. Sci.* **245**, 43–51
- 32 Munro, A.W., Malarkey, K., McKnight, J., Thomson, A.J., Kelly, S.M., Price, N.C., Lindsay, J.G., Coggins, J.R. and Miles, J.S. (1994) *Biochem. J.* **303**, 423–428
- 33 Ost, T.W., Miles, C.S., Munro, A.W., Murdoch, J., Reid, G.A. and Chapman, S.K. (2001) *Biochemistry* **40**, 13421–13429
- 34 Oliver, C.F., Modi, S., Sutcliffe, M.J., Primrose, W.U., Lian, L.Y. and Roberts, G.C. (1997) *Biochemistry* **36**, 1567–1572
- 35 Graham-Lorence, S., Truan, G., Peterson, J.A., Falck, J.R., Wei, S., Helvig, C. and Capdevila, J.H. (1997) *J. Biol. Chem.* **272**, 1127–1135
- 36 Klein, M.L. and Fulco, A.J. (1993) *J. Biol. Chem.* **268**, 7553–7561
- 37 Roitel, O., Scrutton, N.S. and Munro, A.W. (2003) *Biochemistry* **42**, 10809–10821
- 38 Finn, R.D., Basran, J., Roitel, O., Wolf, C.R., Munro, A.W., Paine, M.J. and Scrutton, N.S. (2003) *Eur. J. Biochem.* **270**, 1164–1175
- 39 Neeli, R., Roitel, O., Scrutton, N.S. and Munro, A.W. (2005) *J. Biol. Chem.* **280**, 17634–17644
- 40 Ost, T.W., Miles, C.S., Murdoch, J., Cheung, Y., Reid, G.A., Chapman, S.K. and Munro, A.W. (2000) *FEBS Lett.* **486**, 173–177
- 41 Peters, M.W., Meinhold, P., Glieder, A. and Arnold, F.H. (2003) *J. Am. Chem. Soc.* **125**, 13442–13450
- 42 Stuehr, D.J., Santolini, J., Wang, Z.Q., Wei, C.C. and Adak, S. (2004) *J. Biol. Chem.* **279**, 36167–36170
- 43 Gustafsson, M.C., Roitel, O., Marshall, K.R., Noble, M.A., Chapman, S.K., Pesseguero, A., Fulco, A.J., Cheesman, M.R., von Wachenfeldt, C. and Munro, A.W. (2004) *Biochemistry* **43**, 5474–5487
- 44 De Mot, R. and Parret, A.H.A. (2002) *Trends Microbiol.* **10**, 502–508

Received 13 April 2005

# **Structure of BM3 FAD-domain: The Last Piece in the Jigsaw**

**M.G. Joyce, R. Neeli, O. Roitel, A.W. Munro and D. Leys**

*Henry Wellcome Building, University of Leicester, UK*

Cytochrome P450 BM3 is one of the few cytochromes P450 that is covalently attached to its redox partner. We here report the structure of the C773A FAD-domain in presence and absence of NADP<sup>+</sup> to 2.15 Å and 2.4 Å, respectively. Modelling of the complete BM3 structure using the previously determined heme and FMN domain structures is now possible. The model allows for the identification of putative domain-domain contacts regions. A conserved set of residues is postulated to be involved in aligning the FMN domain for efficient electron transfer.

## **Introduction**

P450 BM3 from *Bacillus megaterium* is a paradigm model system in the P450 superfamily. This soluble bacterial fatty acid hydroxylase P450 is fused to a eukaryotic-like cytochrome P450 reductase (CPR) in a single polypeptide (1). P450 BM3 has the highest reported hydroxylase activity for any P450, oxygenating arachidonic acid in a NADPH-dependent manner at >17,000 min<sup>-1</sup>. The structure of the intact, multidomain flavocytochrome has been elusive, but structures of the heme domain in both substrate-bound and substrate-free forms have been determined (2, 3). More recently, the structure of the heme and FMN domains (the first two domains of the 1048 amino acid flavocytochrome) was reported (4).

The last remaining domain of the protein to be structurally elucidated is the FAD/NADPH binding domain; this domain interacts with NADPH, accepting two electrons (as a hydride ion) and shuttling these, one at a time, through the FMN domain onto the P450. The FAD/NADPH domain is related to the respective domain in eukaryotic CPRs, and to the NADPH-dependent ferredoxin/flavodoxin reductases. A major problem to be solved for the CPRs is the mechanism of molecular dynamics that enables the FMN domain to communicate both with the

FAD domain (to receive electrons) and then with the P450 to transfer electrons to the heme centre. Structural characterization of the BM3 FAD-domain will enable an advanced understanding of the roles of several amino acids in interactions with NADPH and FAD, and in interfacial recognition and docking with the FMN domain to facilitate inter-flavin electron transfer.

In this paper we report the crystal structure of the BM3 FAD/NADPH binding domain in both ligand-free and NADP<sup>+</sup>-bound forms. Aspects of the interaction with coenzyme and flavin are discussed, as well as the proposed mode of docking with the FMN domain in the catalytically relevant electron transfer complex.

## Materials and Methods

**Mutagenesis**—The C773A mutation was prepared to prevent aggregation and protein dimerization. Oligonucleotide-directed mutagenesis of the P450 BM3 FAD domain was performed using the Stratagene QuikChange kit. Purification was carried out as previously described (5).

**Crystallisation**—The BM3 C773A FAD domain was crystallised using the sitting drop vapour diffusion method. Drops were prepared by adding 2 µl of 28 % polyethylene glylcol 8000, 0.3 M ammonium sulphate, cacodylate buffer pH 6.5 to 2 µl of 12 mg/ml protein solution. Crystals of dimensions 70 x 70 x 900 mm formed after 4-7 days. The NADP<sup>+</sup> complex was obtained by briefly soaking crystals in a 10 mM NADP<sup>+</sup> mother liquor solution.

**Data Collection, Structure Determination**—Diffraction data was collected at ID14-EH3 at the ESRF (Grenoble, France). Data were processed and scaled using the HKL package programs DENZO and SCALEPACK (6). The crystal structure was solved using molecular replacement with the FAD domain of rat cytochrome P450 reductase (PDB code 1AMO, 7) as the search model. Model building was carried out in TURBO-FRODO (8) with positional and B-factor refinement carried out using REFMAC5 (9). Final refinement statistics for the ligand-free BM3 FAD structure are  $R_{\text{free}} = 0.256$ ,  $R_{\text{factor}} = 0.211$ , and  $R_{\text{free}} = 0.228$  and  $R_{\text{factor}} = 0.202$  for the NADP<sup>+</sup> complex.

## Results and Discussion

**Overall structure**—The crystal structure of the BM3 FAD domain consists of residues 658-1048. The overall fold is highly similar to the corresponding domain of CPR (rmsd of 1.8 Å) and consists of three sub-domains: an FAD-binding domain, a NADP(H)-binding domain and a connecting domain. Comparison of the two molecules in the asymmetric unit reveals little difference except for a movement between

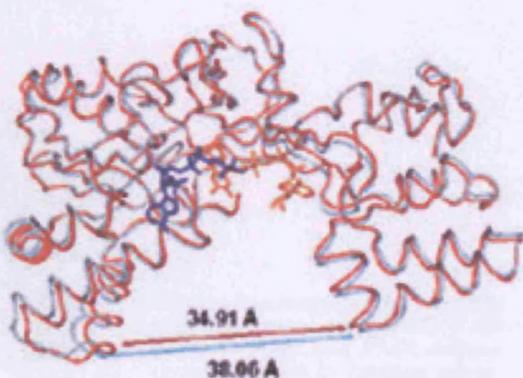


Figure 1. (Top left) Overlay of both molecules from the asymmetric unit of the NADP<sup>+</sup>-bound BM3 FAD domain, with FAD group (orange) and NADP<sup>+</sup> group (blue) in sticks.

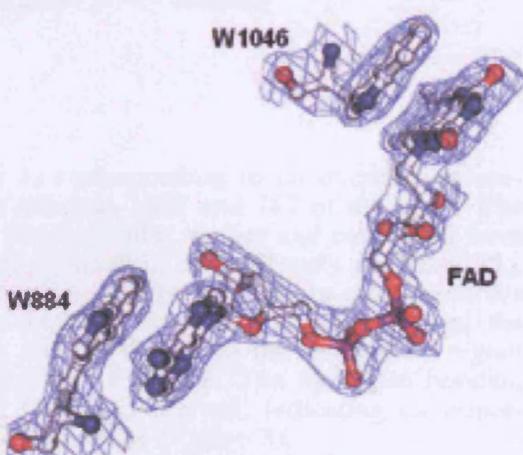


Figure 2. (Top right) SigmaA weighted 2Fo-Fc map surrounding the FAD cofactor and W1046 and W884 in ligand-free FAD domain.

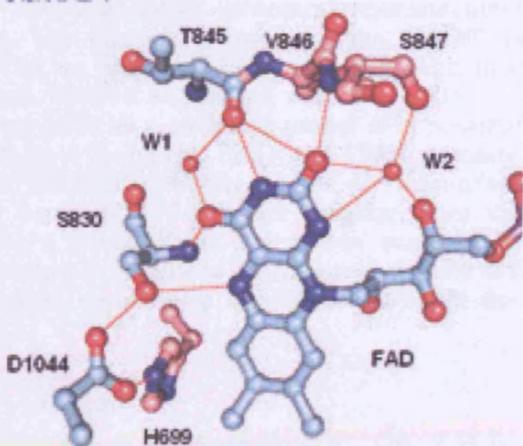


Figure 3. (Bottom left) Ball and stick representation of the isoalloxazine ring and the surrounding residues and water molecules in ligand-free BM3 FAD domain.



*Figure 4. (Bottom right) Overlay of CPR (cyan) with BM3 FAD domain (blue) and FMN domain (green) shown in cartoon representation.*

domains, as shown in Figure 1, corresponding to an overall displacement in the structure between residues 1027 and 787 of over 3 Å. The C773A mutation is located at the molecular surface and could thus form a disulfide bond between BM3 molecules, as previously described (5).

**FAD binding region**—This region is highly similar to other members of the FNR family with the FAD bound in the elongated form, the isoalloxazine ring interacting with W1046 and the nucleotide region interacting with W884, as shown in Figure 2. The hydrogen bonding network previously described (10) is conserved, indicating its importance in the hydride transfer mechanism (Figure 3).

**NADP(H) binding region**—No significant structural rearrangement is seen upon NADP<sup>+</sup> binding. The electron density for the NADP<sup>+</sup> is lacking for nicotinamide moiety, as seen in CPR, indicating high mobility of the nicotinamide when W1046 is stacked with the FAD.

**Domain organisation**—Using CPR as a template model it is possible to construct the full BM3 reductase using the FAD and FMN domains (Figure 4). Superimposing the different domains on the CPR template leads to little or no clashes. Residues 731-744 are missing from the BM3 FAD structure indicating a considerable amount of mobility in this loop region. In CPR this region is stabilised by interactions with the FMN domain and the loop region connecting the FAD and FMN domains.

## References

- 1) Munro, A.W., Leys, D.G., McLean, K.J., Marshall, K.R., Ost, T.W.B.,

- Daff, S., Miles, C.S., Chapman, S.K., Lysek, D.A., Moser, C.C., Page, C.C. and Dutton, P.L. *Trends Biochem. Sci.* 27, 250-257, (2002).
- 2) Li, H. and Poulos, T.L. *Nat. Struct. Biol.* 4, 140-146, (1997).
  - 3) Ravichandran, K.G., Bodupalli, S.S., Hasemann, C.A., Peterson, J.A. and Deisenhofer, J. *Science* 261, 731-736, (1993).
  - 4) Sevrioukova IF, Li H, Zhang H, Peterson JA, Poulos TL. *Proc. Natl. Acad. Sci. USA.* 96, 1863-1868, (1999).
  - 5) Govindaraj, S., and Poulos, T. L. *J. Biol. Chem.* 272, 7915–7921, (1997).
  - 6) Otwinowski, Z., and Minor, W. *Methods Enzymol.* 276, 307-326, (1997).
  - 7) Wang, M., Roberts, D.L., Paschke, R., Shea, T.M., Masters, B.S. and Kim, J.J. *Proc. Natl. Acad. Sci. USA.* 94, 8411-8416, (1997).
  - 8) Roussel, A., and Cambillau, C. TURBO-FRODO, Biographics, AFMB. Marseille, France, (1992).
  - 9) Murshudov, G.N., Vagin, A.A., and Dodson, E.J. *Acta Cryst. D* 53, 240-255, (1997).
  - 10) Zhang, J., Martasek, P., Paschke, R., Shea, T., Masters, B.S.S. and Kim, J.J.P. *J. Biol. Chem.* 276, 37506-37513, (2001).