

Structural and Mechanistic Studies of Quinone Oxidoreductase II: NQO2

Rana Hosny Refaey

B.Sc. (Pharmacy)

**GEORGE GREEN LIBRARY OF
SCIENCE AND ENGINEERING**

**Thesis submitted to the University of Nottingham for the
degree of the Doctor of Philosophy**

September 2010

Dedicated to my family

Abstract

Flavoenzymes are a widely diverse group of enzymes that are able to catalyze a variety of different chemical reactions. A growing interest in flavoenzymes capable of reducing aromatic nitro groups may be attributed to their ability to reduce anticancer prodrugs such as CB1954. Similarly, quinone oxidoreductases are able to activate anticancer prodrugs such as mitomycin derivatives by reducing the quinone derivative to the hydroxyl derivative. Some oxidoreductases are able to reduce nitro aromatic compounds. These enzymes may be either mammalian such as NQO1 and NQO2 which are over expressed in certain cancerous cells or bacterial which can be introduced into the cancerous growth. The bacterial nitroreductases have been previously well characterized and were found to share several similarities. On the other hand, other nitro reducing enzymes have not been investigated to determine the presence of any shared similarities. Another aim of this thesis is to identify the differences between the structure of the oxidized and reduced forms of NQO2.

The similarities shared between the flavoenzymes which are known to reduce nitroaromatic compounds to their hydroxylamine or amine derivative were investigated in this thesis. Both the sequences and 3D structures of the selected proteins were compared and investigated using a variety of bioinformatics tools such as sequence and structure alignment, in addition to homology modelling. It was discovered that there are several aromatic residues conserved in different positions, relative to the flavin cofactor, in the catalytic sites of the proteins.

The 3D structure of the reduced NQO2 was determined using X-ray crystallography. The oxidized and reduced protein structures were compared to determine the structural differences. The expression, purification and crystallization of the NQO2 protein, as well as the reduction of the protein crystals are described in this work. The determination of the structure of the reduced protein provided new insights into the mechanism of electron transport to and from the flavin cofactor. Moreover, kinetics studies were undertaken on the purified NQO2 protein using fluorescence spectroscopy. These resulted in the identification of several new substrates for NQO2.

Acknowledgments

Firstly, I would like to thank my supervisor Dr. Cristina DeMatteis for her guidance and understanding through out my PhD. I would also like to thank the International Office and the School of Pharmacy at the University of Nottingham for funding.

I would also like to add my sincere appreciation for all the help provided by all my colleagues in the crystallography and modelling labs. Special thanks to Dr. Elizabeth Hooley for her help in collecting data at Diamond Light Source and to Dr. Paul McEwan for his guidance during data processing. I would also like to add my appreciation for the technical support given by Dr. Ian Withers, Dr. Huw Williams, Dr. Jed Long and Lee Hibbet. I would also like to thank Graham Coxhill from the School of Chemistry for help provided on mass spectroscopy work carried out.

I would also like to thank my friends for their good company, encouragement, and laughs, as well as many lunches, coffees and greatly needed chocolate breaks. I would also like to thank my friends abroad for their encouragement and many long phone conversations at the oddest times of the day.

Last but not least I would like to express my gratitude to my parents and brother for all their financial, moral and emotional support. I would also like to thank them for their continuous encouragement and never ending belief in me without which I would have not been able to carry out my PhD.

Rana H. Refaey

Table of Contents

Abstract.....	i
Acknowledgments	iii
Table of Contents	iv
List of Figures.....	x
List of Tables	xiii
Abbreviations	xv

1. Introduction to Flavoenzymes and their Role in Bio-reductive Cancer Chemotherapy..... 1

1.1 Flavoenzymes	1
1.1.1 Regulation of Flavoenzymes	3
1.1.2 Recurrent Characteristics in Flavoproteins.....	5
1.1.3 Classification of Flavoproteins	6
1.2 Quinone Oxidoreductases	7
1.2.1 Quinones	8
1.2.1.1 Sources of Quinones	8
1.2.1.2 Cytotoxicity of Quinones.....	8
1.2.3 Primary Structure of NQOs	9
1.2.4 Secondary Structure of NQOs.....	10
1.2.5 Catalytic Site of NQOs	12
1.2.6 Mechanism of Action of NQOs.....	13
1.2.7 Cosubstrates, Substrates and Inhibitors of NQOs.....	16
1.2.7.1 Cosubstrates	16
1.2.8 Distribution of NQOs	17
1.2.9 Physiological functions of NQOs	17
1.2.10 Medical Applications of NQOs.....	18
1.3 Bacterial Nitroreductases.....	19
1.4 Cancer	20
1.4.1 The Development of Cancer	20
1.4.2 Hypoxia in Tumours.....	20
1.5 Cancer Treatment.....	22
1.5.1 Anticancer Targeted Prodrug Therapy	22
1.5.1.1 ADEPT.....	23
1.5.1.2 GDEPT	23
1.5.2 Bioreductive Anticancer Drugs	24
1.5.2.2 Nitroaromatics	28
1.6 Aims of the Thesis.....	32

2. Introduction to X-ray Crystallography	33
2.1 Introduction.....	33
2.2 Protein Crystallization	35
2.2.1 Protein crystallization techniques	36
2.2.2. Cryocrystallography.....	37
2.2.3 Nature of the Crystal	38
2.2.3.1 <i>Lattice planes and Miller indices</i>	<i>40</i>
2.2.4 X-ray Diffraction by Crystals	41
2.2.4.1 <i>Bragg's Law</i>	<i>41</i>
2.2.4.2 <i>Reciprocal lattice and Ewald's Sphere.....</i>	<i>42</i>
2.3 Data collection.....	43
2.4 Data Processing.....	44
2.4.1 Indexing and Integration of Images.....	44
2.4.2 Scaling and Merging of Reflections.....	45
2.4.2.1 <i>Assessing Data Quality.....</i>	<i>46</i>
2.4.3 The Structure Factors and the Phase Problem.....	47
2.4.3.1 <i>Molecular Replacement</i>	<i>48</i>
2.4.4 Electron Density Maps	49
2.4.5 Refinement of the Model	49
2.4.5.1 <i>Refinement Indicators</i>	<i>50</i>
2.4.6 Model Quality.....	51
3. Sequence analysis and Structural Studies of Nitroaromatic Reducing Enzymes	53
3.1 Introduction.....	53
3.2 Methods and Materials.....	56
3.2.1 Summary of Methods Used.....	56
3.2.2 Identification of Enzymes	57
3.2.3 Strategy for Obtaining Sequences and 3D Structural Data.....	57
3.2.4 Extraction of Active Site Residues from 3D Structures	57
3.2.5 Sequence Alignment	58
3.2.5.1 <i>Alignment parameters used in alignments with Clustal W2</i>	<i>59</i>
3.2.6 The 3D Structural Alignment of the Active Site Residues.....	59
3.2.7 Sequence Alignment for Enzymes within Different Structural Groups	60
3.2.7.1 <i>Calculation of percentage identity & similarity between sequences of group members</i>	<i>60</i>
3.2.8 Homology Modelling	61
3.2.8.1 <i>Choice of model</i>	<i>61</i>
3.2.8.2 <i>Template examination and group identification.....</i>	<i>62</i>
3.2.8.3 <i>Percentage identity and similarity calculation for modelled proteins</i>	<i>62</i>
3.2.8.4 <i>Validation of group identification</i>	<i>63</i>
3.2.9 Structural and Sequence Alignment for Proteins Groups.....	63
3.2.9.1 <i>Structural alignment</i>	<i>63</i>
3.2.9.2 <i>Sequence alignment</i>	<i>63</i>
3.2.9.3 <i>Secondary sequence alignment.....</i>	<i>63</i>
3.2.10 Identification of Conserved Residues.....	64

3.3 Results	64
3.3.1 Obtaining the Sequences and 3D Structural Data	64
3.3.2 Active Site Residue Extraction from 3D Structures	64
3.3.3 Sequence Alignment for All Enzymes	67
3.3.4 The 3D Structural Alignments of Active Site Residues	67
3.3.4.1 <i>Group I active site structural alignment</i>	68
3.3.4.2 <i>Group II active site structural alignment</i>	69
3.3.5 Sequence Alignment for Enzymes within Identified Groups	70
3.3.5.1 <i>Calculation of identity and similarity within identified groups</i>	70
3.3.5.2 <i>Sequence alignment of group I</i>	71
3.3.5.3 <i>Sequence alignment of group II</i>	72
3.3.6 Homology Modelling	73
3.3.6.1 <i>Template examination and group identification</i>	74
3.3.6.2 <i>Comparison of models produced by homology modelling</i>	75
3.3.6.3 <i>Percentage identity and similarity calculation between modelled and previously identified group I members</i>	76
3.3.6.4 <i>Percentage identity and similarity calculation between modelled and previously identified group II members</i>	77
3.3.7 Structural and Sequence Alignment for Group I Proteins	77
3.3.7.1. <i>Structural alignment</i>	77
3.3.7.2 <i>Sequence alignment</i>	78
3.3.7.3 <i>Secondary Structure prediction and alignment</i>	80
3.3.8 Structural and Sequence Alignment for Group II Proteins	81
3.3.8.1 <i>Sequence alignment</i>	81
3.3.8.2 <i>Structural alignment</i>	82
3.3.8.3 <i>Secondary Structure prediction and alignment:</i>	82
3.3.9 Structural and Sequence Alignment for Group III Proteins	83
3.3.9.1 <i>Identification of group III members</i>	83
3.3.9.2 <i>Percentage identity and similarity between group III members</i>	83
3.3.9.3 <i>Sequence alignment</i>	85
3.3.9.4 <i>Secondary Structure prediction and alignment</i>	85
3.3.10 Structural and Sequence Alignment for Group IV Proteins ... 86	
3.3.10.1 <i>Structural alignment</i>	86
3.3.10.2 <i>Sequence alignment:</i>	88
3.3.10.3 <i>Secondary Structure predication and alignment</i>	89
3.4 Discussion and Conclusion	89
4. Kinetic Studies for NQO2 I: Methods and Materials	96
4.1 Introduction	96
4.1.1 Enzyme Reactions	96
4.1.2 Enzyme Kinetic Constants	97
4.1.3 Potential NQO2 Substrates	100
4.1.3.1 <i>Novel substrate, GC201</i>	100
4.1.3.2 <i>Coenzyme Q0, Coenzyme Q1 and Coenzyme Q2</i>	100
4.1.4 Experimental Methods for Determining Michealis-Menten Constants	101
4.2 Experimental Methods and Materials	102
4.2.1 Cloning of NQO2	103
4.2.2 Expression and Harvesting of NQO2	103
4.2.3 Purification of NQO2	104

4.2.4 Assessment of Protein Purity	105
4.2.4.1 <i>Preparation of gel</i>	106
4.2.4.2 <i>Preparation of samples and running of the gel</i>	107
4.2.5 Concentrating and Storage of NQO2	107
4.2.6 Optimizing Enzyme Kinetics Protocol	108
4.2.6.1 <i>UV visible absorbance scans</i>	108
4.2.6.2 <i>Fluorescence emission and excitation spectrums</i>	108
4.2.6.3 <i>Effect of air exposure on the fluorescence of NBDH</i>	108
4.2.6.4 <i>Fluorescence emission controls for other reaction components</i>	109
4.2.7 Calibration Curve Relating Concentration of NBDH to its Fluorescence	109
4.2.8 Optimization of Published Reaction Conditions	110
4.2.8.1 <i>Enzyme concentration optimization</i>	111
4.2.8.2 <i>Temperature choice</i>	111
4.2.8.3 <i>Reproducibility testing</i>	112
4.2.9 K_m Determination for Novel Substrates	112
4.2.9.1 <i>Determination of linear relationship between enzyme concentration and reaction rate</i>	112
4.2.9.2 <i>K_m determination</i>	113
4.2.9.3 <i>Substrate-cosubstrate reaction controls</i>	114
5. Kinetic Studies for NQO2 II: Results and Discussion	115
5.1 Protein Preparation	115
5.1.1 Assessing Protein Purity Using SDS-PAGE	115
5.2 Optimizing Enzyme Kinetics Protocol	115
5.2.1 UV Visible Spectrum of NBDH	115
5.2.2 Fluorescence Emission Spectrum of NBDH	117
5.2.3 Effect of Air Oxidation on the Fluorescence of NBDH	117
5.2.4 Fluorescence Emission Measurements for Other Reaction Components	118
5.2.5 Calibration Curve for NBDH	119
5.3 Protocol Development	121
5.3.1 Choice of Enzyme Concentration	121
5.3.2 Effect of temperature on reaction of menadione and K_m	122
5.3.3 Determining the Reproducibility of Results Using the Optimized Parameters	123
5.4 Enzyme Kinetics for Substrates	124
5.4.1 Substrate/cosubstrate Mixture Fluorescence Monitoring	124
5.4.2 Relationship between enzyme and reaction rate	125
5.4.3 K_m determination for Substrates	127
5.4.3.1 <i>Coenzyme Q0</i>	127
5.4.3.2 <i>Coenzyme Q1</i>	129
5.4.3.3 <i>Coenzyme Q2</i>	130
5.4.3.4 <i>GC201</i>	132
5.5 Discussion and Conclusion	133

6. Determination of the Crystal Structure of Reduced NQO2 I: Methods and Material	135
6.1 Introduction.....	135
6.2 Methods and Materials.....	136
6.2.1 Protein Production	137
6.2.2 Protein Crystallization	137
6.2.3 Crystal Reduction & Cryoprotectant Solution Optimization .	139
6.2.4 Data Collection.....	140
6.2.4.1 <i>Dataset one</i>	140
6.2.4.2 <i>Dataset two</i>	141
6.2.4.3 <i>Datasets three, four and five.....</i>	141
6.2.5 Data Processing.....	142
6.2.5.1 <i>Indexing and integration of images</i>	142
6.2.5.2 <i>Scaling and merging of images.....</i>	142
6.2.5.3 <i>Molecular replacement.....</i>	143
6.2.5.4 <i>Refinement of model</i>	144
6.2.6 Identification of Unknown Electron Density in the Binding Pocket.....	145
6.2.6.1 <i>Automated ligand identification</i>	145
6.2.6.2 <i>Refinement of probable molecules into the electron density</i>	145
6.2.7 Model Validation	147
7. Determination of the Crystal Structure of Reduced NQO2 II: Results and Discussion.....	149
7.1 Protein Production.....	149
7.2 Protein Crystallization	149
7.3 Crystal Reduction and the Choice of Cryoprotectant.....	150
7.3.1 <i>Crystal Reduction</i>	150
7.3.2 <i>Cryoprotection Solution Choice.....</i>	152
7.4 Data Processing.....	152
7.4.1 <i>Space Group and Cell Dimension Determination</i>	152
7.4.2 <i>Data Processing Statistics for Different Datasets</i>	153
7.5 Molecular Replacement.....	154
7.6 Model Examination	155
7.7 Refinement and Ligand Identification.....	159
7.7.1 <i>Refinement Statistics After Manual Adjustments in COOT... ..</i>	159
7.7.2 <i>Identification of Unknown Electron Density.....</i>	159
7.7.2.1 <i>Automated ligand identification</i>	163
7.7.2.2 <i>Refinement of probable molecules into the electron density</i>	164
7.7.2.3 <i>Analytical work</i>	166
7.8 Discussion	166
7.8.1 <i>Structural Differences Between Oxidized & Reduced NQO2. .</i>	166
7.8.2 <i>Unidentified Electron Density</i>	171
8. Conclusions and Future work.....	173

Bibliography	178
Appendices	xviii
<i>A1. Amino acid abbreviations</i>	<i>xviii</i>
<i>A2. Composition of solutions used in experimental procedures</i>	<i>xix</i>
<i>A3. Sequence alignments</i>	<i>xx</i>
<i>A4. Ramachandran plots</i>	<i>xxviii</i>

List of Figures

Figure 1. 1 Structure of flavins and numbering of isoalloxazine ring.	2
Figure 1. 2 Different oxidation states of isoalloxazine ring	3
Figure 1. 3 The chemical properties of the isoalloxazine ring	5
Figure 1. 4 Structures of different quinones	9
Figure 1. 5 Sequence alignment of NQO1 and NQO2.....	10
Figure 1. 6 Superimposed structures of hNQO1 and hNQO2.....	11
Figure 1. 7 Cartoon representation of the structure of NQO2.....	12
Figure 1. 8 The overlaid catalytic sites of hNQO1 and hNQO2	13
Figure 1. 9 Aromatic residues in the catalytic sites of hNQO1 and hNQO2 ...	14
Figure 1. 10 Schematic representation of the mechanism of action of NQOs .	14
Figure 1. 11 Proposed mechanism of hydride transfer in NQO1.	15
Figure 1. 12 Stepwise reduction of nitroaromatic compounds.....	20
Figure 1. 13 Cancer development models	21
Figure 1. 14 Representation of ADEPT.....	24
Figure 1. 15 Representation of GDEPT.....	25
Figure 1. 16 Reduction of MMC and formation of DNA cross links.....	26
Figure 1. 17 Structures of EO9 and AZQ.....	27
Figure 1. 18 Release of active drug upon reduction of indolequinone.....	28
Figure 1. 19 Structure and bioactivation of CB1954.....	30
Figure 2. 1 Processes involved in X-ray Crystallography	34
Figure 2. 2 Phase diagram for crystallization	36
Figure 2. 3 Schematic representations of sitting and hanging drop.....	37
Figure 2. 4 Schematic representation of crystal parameters	39
Figure 2. 5 Lattice planes and their designated Miller indices	40
Figure 2. 6 Constructive interference as defined by Bragg's law.....	41
Figure 2. 7 Construction of Ewald's sphere	43
Figure 2. 8 Schematic representation of varying degrees of crystal packing ...	45
Figure 2. 9 Backbone of protein, showing Φ and Ψ angles.	52
Figure 3. 1 Overlaid extracted active sites of group I members.....	68
Figure 3. 2 Overlaid extracted active sites of group II members.....	69
Figure 3. 3 T-Coffee sequence alignment of identified group I members.....	72
Figure 3. 4 Clustal W2 sequence alignment of identified group I members	72

Figure 3. 5 T-Coffee sequence alignment of identified group II member.....	73
Figure 3. 6 Clustal W2 sequence alignment of identified group II members...	73
Figure 3.7 Overlaid active site of all group I members	78
Figure 3. 8 Sequence alignment of all group I using T-Coffee	79
Figure 3. 9 Sequence alignment of all group I using Clustal W2.....	79
Figure 3. 10 Secondary structure alignment of all group I.	80
Figure 3.11 Overlaid active site of all group II members.....	81
Figure 3.12 Sequence alignment of all group II using T-Coffee.....	82
Figure 3.13 Sequence alignment of all group II using Clustal W2	82
Figure 3. 14 Secondary structure alignment of all group II.....	83
Figure 3. 15 Overlaid active site of all III members.....	84
Figure 3. 16 Sequence alignment of all group III using T-Coffee.....	85
Figure 3. 17 Sequence alignment of all group III using Clustal W2	85
Figure 3. 18 Secondary structure alignment of all group III.	86
Figure 3. 19 Overlaid cartoon representation of the structures of group IV.....	87
Figure 3. 20 Overlaid active site of all group IV members.	87
Figure 3. 21 Sequence alignment of all group IV using T-Coffee	88
Figure 3. 22 Sequence alignment of all group IV using Clustal W2.....	88
Figure 3. 23 Secondary structure alignment of all group IV	89
Figure 3. 24 Residues of mechanistic importance in group I members.....	91
Figure 3. 25 Residues of mechanistic importance in group IV members.....	94
Figure 3. 26 The orientations of the conserved residues in different groups ...	94
Figure 4. 1 Progress curve of an enzyme reaction.....	97
Figure 4. 2 Representation of Michealis-menten & Lineweaver-Burke plot ..	99
Figure 4. 3 Reduction of GC201.....	100
Figure 4. 4 Structures of coenzymes Q0, coenzyme Q1 and coenzyme Q2 ..	100
Figure 4. 5 Structure of NBDH.....	101
Figure 5. 1 Example of SDS-PAGE	116
Figure 5. 2 UV/visible absorbance spectrum of NBDH.....	116
Figure 5. 3 Emission spectrum of NBDH	117
Figure 5. 4 Decrease in fluorescence of NBDH due to air oxidation	118
Figure 5. 5 Fluorescence of NQO2, coenzyme Q series and menadione.....	119
Figure 5. 6 Calibration curve of NBDH at low concentration.....	120
Figure 5. 7 Calibration curve of NBDH at high concentration.....	121

Figure 5. 8 Effect of enzyme concentration on the menadione reaction rate .	122
Figure 5. 9 Michealis-Menten plots for NQO2/menadione reaction.....	123
Figure 5. 10 Reproducibility testing of NQO2/menadione reaction.....	124
Figure 5. 11 Fluorescence decay of cosubstrate/substrate mixtures.....	125
Figure 5. 12 Effect of enzyme concentration on reaction rate.....	126
Figure 5. 13 Michealis-Menten plot for NQO2/coenzyme Q0 reaction.....	127
Figure 5. 14 Hanes-Woolf plot for NQO2/coenzyme Q0 reaction.....	128
Figure 5. 15 Lineweaver-Burke plot for NQO2/coenzyme Q0 reaction	128
Figure 5. 16 Michealis-Menten plot for NQO2/coenzyme Q1 reaction.....	129
Figure 5. 17 Lineweaver-Burke plot of NQO2/coenzyme Q1 reaction.....	130
Figure 5. 18 Michealis-Menten plot for NQO2/coenzyme Q2 reaction.....	131
Figure 5. 19 Lineweaver-Burke plot of NQO2/coenzyme Q2 reaction.....	131
Figure 5. 20 Michealis-Menten plot for reaction of GC201 with NQO2.....	132
Figure 6. 1 Side view of isoalloxazine ring showing bend N5-N10.....	136
Figure 6. 2 Oxidation of DTT	146
Figure 7. 1 Photographs of NQO2 crystals.....	151
Figure 7. 2 Reduced NQO2 crystal mounted on goniometer.	151
Figure 7. 3 Cartoon representation of NQO2, showing mobile loop.....	156
Figure 7. 4 Active site of NQO2 showing the unidentified electron density.	157
Figure 7. 5 Different orientations of the active site electron density maps of dataset 4	158
Figure 7. 6 The active site electron density maps of data set 5	159
Figure 7. 7 Side view of unidentified electron density of different data sets.	161
Figure 7. 8 Top view of unidentified electron density of different data sets..	162
Figure 7. 9 Overlaid active site electron density maps of data sets 1,2 & 3...	163
Figure 7. 10 Structures of 4- hydroxycinnamic acid and palmitic acid.....	164
Figure 7. 11 Different molecules refined in the unidentified electron density	165
Figure 7. 12 Overlaid cartoon representation of oxidized & reduced NQO2.	167
Figure 7. 13 Overlaid FAD molecules of the oxidized & reduced NQO2	168
Figure 7. 14 Overlaid structures of oxidized & reduced thioredoxin reductase	169
Figure 7. 15 Hydrogen bond between Tyr-155 residue and O2 of FAD.....	170
Figure 7. 16 Overlaid Tyr-155 & FAD of the oxidized & reduced NQO2	170

List of Tables

Table 1.1 Reported substrates and inhibitors of NQO1 and NQO2.	17
Table 3. 1 Different parameter combinations used for Clustal W2 alignments	59
Table 3. 2 Final list of enzymes, accession code & associated species.....	65
Table 3. 3 Final list of enzymes, accession & pdb codes & associated species	66
Table 3. 4 Conserved residues of group I.....	69
Table 3. 5 Conserved residues of group II.....	70
Table 3.6 Percentage similarity between group I members.....	70
Table 3. 7 Percentage similarity between group II members	71
Table 3. 8 Symbols & colours used in Clustal W2 & T-Coffee alignments	72
Table 3. 9 List of templates used in homology modelling	74
Table 3.10 RMSD values between the different homology models.....	76
Table 3.11 Percentage similarities between identified & modelled group I members.....	76
Table 3. 12 Percentage similarities between identified & modelled group II members.....	77
Table 3.13 Conserved residues in modelled group I proteins	78
Table 3.14 Conserved aromatic residues in all group III members.....	83
Table 3.15 Percentage similarity between all group III members	84
Table 3.16 Conserved residues in group IV members.....	88
Table 3. 17 Percentage similarity between group IV members.....	88
Table 4. 1 Elution protocol of NQO2 from Ni-NTA resin.....	105
Table 4. 2 Preparation of resolving and stacking gels of SDS-PAGE.....	106
Table 4. 3 Samples in gel lanes	107
Table 4. 4 Optimum reaction conditions.	112
Table 6. 1 Previously published crystallization conditions	138
Table 6. 2 Parameters used for collection of dataset 1.	141
Table 6. 3 Parameters used for collection of dataset 2.	141
Table 6. 4 Parameters used for collection of datasets 3, 4 and 5.....	141
Table 6. 5 Combined parameters for determination of nominal resolution....	142
Table 6. 6 Structures of probable molecules refined into the unidentified electron density.....	147
Table 7. 1 Optimized conditions for NQO2 crystallization.....	150

Table 7. 2 Average cell dimensions for the different cell units.....	152
Table 7. 3 Data processing statistics for crystals with chunky morphology ..	153
Table 7. 4 Data processing statistics for crystals with plate like morphology	154
Table 7. 5 Space group, solvent content, Z-scores & LLG values	154
Table 7. 6 Refinement statistics for datasets one, two and three.....	160
Table 7. 7 Refinement statistics for datasets four and five.....	160

Abbreviations

°	Degree
2xYT	Two times yeast-trytone media
3D	Three dimensional
Å	Angstrom
ADEPT	Antibody directed enzyme prodrug therapy
APS	Ammonium persulphate
C	Celsius
CB1954	5-(azrindin-1-yl)-2,4-dinitrobenzamide
CCP4	Collabrative computing project number 4
DMSO	Dimethylsulphoxide
DNA	Deoxyribonucleic acid
DTT	Dithiothreitol
E.cloacae	<i>Enterobacter cloacae</i>
E.coli	<i>Escherichia coli</i>
ESRF	European Synchrotron Radiation Facility
FAD	Flavin adenine dinucleotide
FMN	Flavin mononucleotide
FRase	Flavin oxidoreductase of <i>Vibrio fischeri</i>
FRP	NADPH:FMN Ooxidoreductase of <i>Vibrio harveyi</i>
FrxA	NAD(P)H flavin oxidoreductase of <i>Helicobacter pylori</i>
FUDR	5-fluorodeoxyuridine
GDEPT	Gene Directed Enzyme Prodrug Therapy
GSH	Glutathione
GTN	Glycerol trinitrate reductase
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
HPLC	High pressure (performance) liquid chromatography
K	degree kalvin
kDa	Kilodalton
l	Litre
m	Milli
M	Molar

MAO	Monoamine oxidase
MMC	Mitomycin C
n	Nano
NADPH	Nicotinamide adenine dinucleotide phosphate
Na-HEPES	4-(2-Hydroxyethyl)piperazine-1-ethanesulfonic acid sodium salt
NBDH	N-benzylidihydronicotinamide
NbzA	Nitrobenzene nitroreductase
NEC	<i>E. cloacae</i> nitroreductase
NemA	N-ethylmaleimide reductase
NfsA	Major nitroreductase of <i>Escherichia coli</i>
NfsB	Minor nitroreductase of <i>Escherichia coli</i>
Ni-NTA	Nickel-nitrilotriacetic acid
NMR	Nuclear Magnetic Resonance
NOX	NADH oxidase of <i>Thermus thermophilis</i>
NQO1	NAD(P)H:quinone oxidoreductase
NQO2	Quinone reductase 2
NRH	N-ribosyl-1,4-dihydronicotinamide
OD	Optical Density
OYE	Old yellow enzyme
PDB	Protein database
PETN	Pentoerythritol tetranitrate reductase
PnrA	<i>Pseudomonas putida</i> nitroreductase
RBCs	Red blood cells
RMSD	Root mean square deviation
Rnr	Retro-nitroreductase
RPM	Revolution per minute
SAXS	Small angle X-ray scattering
SDS	Sodium dodecyl sulphate
SDS-PAGE	sodium dodecyl sulphate polyacrylamide gel electrophoresis
sec	Second
SnrA	Oxygen-insensitive NADPH reductase
TEMED	N,N,N',N'-Tetramethylethylenediamine

TNT	Trinitrotoluene
Tris	Tris(hydroxymethyl)aminomethane
UV	Ultra violet
V	Volts
VDEPT	Virus Directed Enzyme Prodrug Therapy
XenA	Xenobiotic reductase A
XenB	Xenobiotic reductase B
μ	Micro

Chapter One

Introduction to Flavoenzymes and their Role in Bio-reductive Cancer Chemotherapy

1.1 Flavoenzymes

Flavoenzymes are very versatile and catalyze a wide variety of biochemical reactions, including energy production, oxidation, reduction, light emission and biodegradation. They depend on the presence of a flavin cofactor to be able to catalyze enzymatic reactions. Flavins are derivatives of vitamin B₂, riboflavin; which can be synthesized *de novo* in plants and bacteria, but not in animals, and thus must be provided from dietary sources (De Colibus and Mattevi, 2006, Sinnott et al., 1997). The two most common flavins utilized by proteins are flavin mononucleotide (FMN) and flavin adenine dinucleotide (FAD), where FAD is the phosphoryl adenosine derivative of FMN. Flavin cofactors consist of two moieties, a tricyclic isoalloxazine ring and a side chain (Figure 1. 1) (Sinnott et al., 1997). The side chain is responsible for anchoring the cofactor, usually via non-covalent bonds to the protein, while the isoalloxazine ring is the reactive moiety of the flavin cofactor.

The isoalloxazine ring is responsible for the chemical versatility of the flavin cofactor. The flavin cofactor is able to undergo both one and two electron reduction to produce the semiquinone and fully reduced flavin, respectively (Miura, 2001). Flavoprotein redox reactions typically involve two separate steps, a first reduction step followed by an oxidation step. The reduction step in

many flavoenzymes occurs in one step involving the transfer of two electrons to the flavin from an electron donor to produce the fully reduced flavin. The second oxidation step may occur in either one step involving the transfer of two electrons to produce the fully oxidized cofactor directly, or two steps involving the transfer of one electron per step producing the fully oxidized cofactor via the semiquinone derivative (Figure 1. 2) (Massey, 2000).

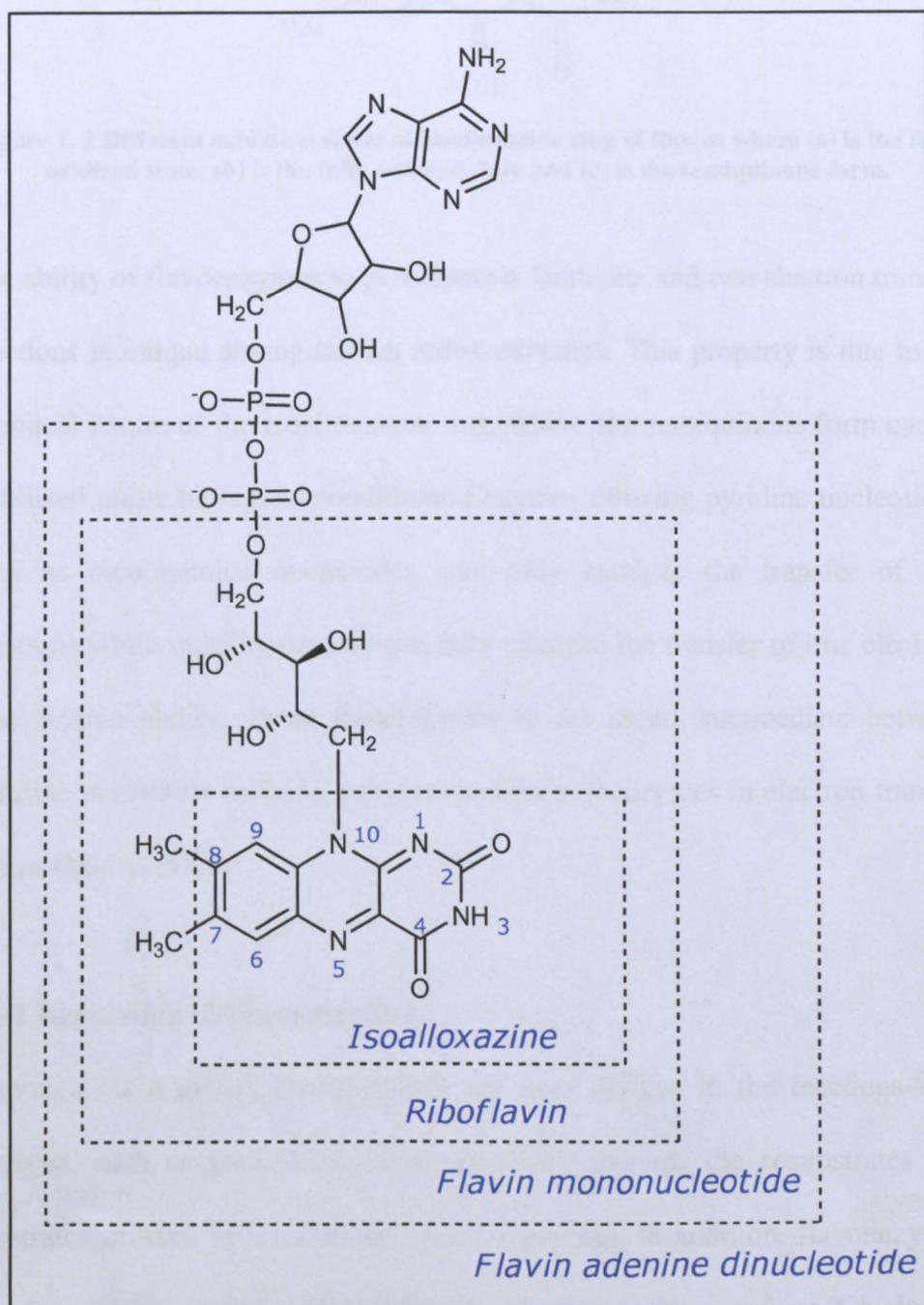


Figure 1. 1 Structure of flavins and numbering of isoalloxazine ring.

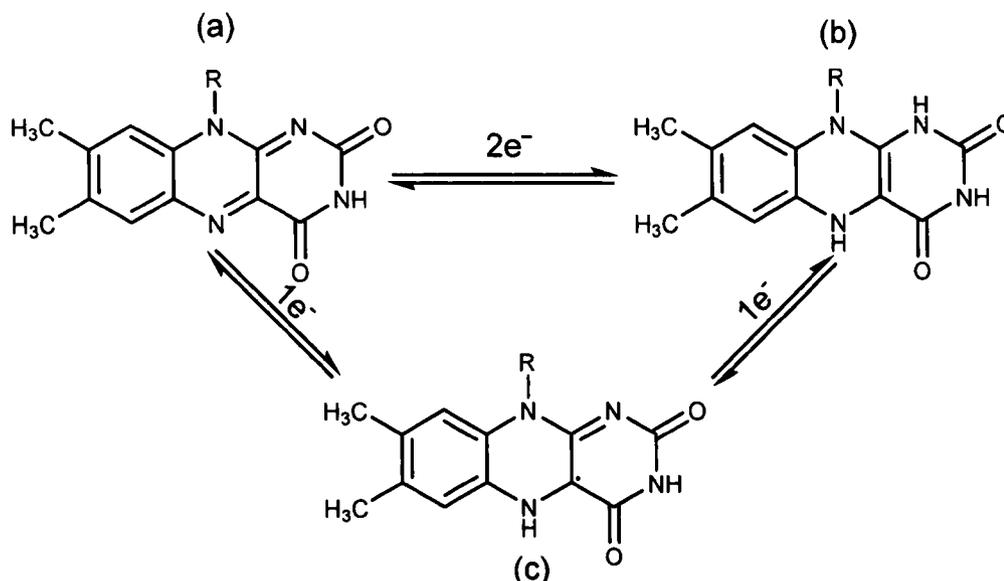


Figure 1. 2 Different oxidation states of isoalloxazine ring of flavins where (a) is the fully oxidized state; (b) is the fully reduced state and (c) is the semiquinone form.

The ability of flavoenzymes to participate in both one and two electron transfer reactions is unique among known redox enzymes. This property is due to the chemical nature of the isoalloxazine ring, where the semiquinone form can be stabilized under biological conditions. Enzymes utilizing pyridine nucleotides, such as nicotinamide nucleotides, can only catalyze the transfer of two electrons while metalloenzymes can only catalyze the transfer of one electron. This unique ability allows flavoproteins to act as an intermediate between pyridine nucleotide utilizing enzymes and metalloenzymes in electron transfer chains (Miura, 2001).

1.1.1 Regulation of Flavoenzymes

Although, as a group, flavoenzymes are very diverse in the reactions they catalyze, each enzyme shows high specificity towards the cosubstrates and substrates utilized in the electron transfer process. In addition, flavoenzymes are also selective with regards to the type of reaction they catalyse (Ghisla and

Massey, 1986, Miura, 2001). This suggests the presence of several mechanisms by which the reactions of flavoenzymes are regulated and controlled. These may include hydrogen bonding between the enzyme and the isoalloxazine ring of the flavin cofactor, the electrostatic environment in the catalytic site of the enzyme, the formation of charge-transfer complexes between the flavin and the substrate, as well as the regulation of the position of the substrate in the binding pocket in relation to the flavin cosubstrate (Sinnott et al., 1997, Miura, 2001).

The electrostatic environment in the catalytic site of the enzyme is regulated by controlling the accessibility of the solvent to the binding site and thus maintaining the hydrophobicity of the binding site. In several proteins catalyzing dehydrogenation reactions the catalytic site is shielded from the solvents, by a mobile loop, which allows the entrance of the substrate to the catalytic site and then closes. Alternatively, the catalytic site of the enzyme may be narrow and hence the binding of the substrate expels any solvent molecules previously present in the binding site (Fraaije and Mattevi, 2000, Miura, 2001, Sinnott et al., 1997).

The amphipathic chemical nature of the isoalloxazine ring allows many different interactions. The xylene like portion of the molecule is hydrophobic and thus is more likely to interact with hydrophobic portions of the protein and substrates. On the other hand, the pyrimidine moiety is hydrophilic and electron deficient and thus more likely to interact with electron rich portions of the enzyme and substrates (Figure 1. 3). The chemical properties of the

isoalloxazine ring are affected by the redox state of the flavin, where the basicity of the N5 atom increases upon the reduction of the flavin from the oxidized state to the semiquinone state, and shows a dramatic increase upon further reduction to the fully reduced state. Likewise, the pyrimidine moiety becomes electron rich when in the reduced state compared to being electron deficient in the oxidized state (Ghisla and Massey, 1986). Moreover, there are several sites on the flavin that could participate in hydrogen bonds as donors or acceptors, which are also affected by the redox state of the cofactor. In the oxidized state N1, O2, O4 and N5 act as hydrogen bond accepting sites, while N3 acts as both hydrogen bond donating and accepting site. On the other hand, in the reduced state O2 and O4 act as hydrogen bond accepting sites, while N1 and N5 can act as both hydrogen bond accepting and donating sites (Figure 1.1 and Figure 1.3) (Miura, 2001).

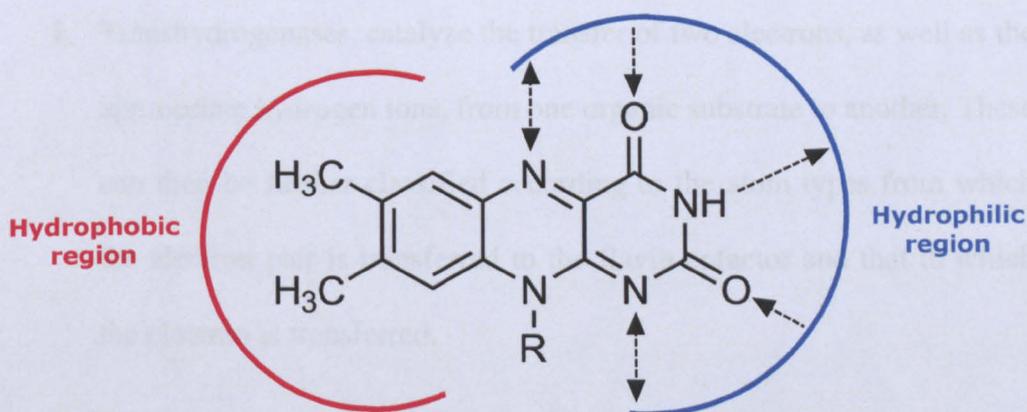


Figure 1.3 The isoalloxazine ring of flavins, showing the chemical properties of the ring and the position of hydrogen bond interactions, represented by dashed lines.

1.1.2 Recurrent Characteristics in Flavoproteins

Although flavoenzymes are very diverse in their functions, they share several recurrent characteristics. FMN binding proteins tend to exhibit an $(\alpha/\beta)_8$ barrel folding topology; on the other hand, FAD binding proteins generally exhibit

more diverse binding topologies (Fraaije and Mattevi, 2000, Sinnott et al., 1997). Flavoenzymes catalyzing dehydrogenation reactions share several similarities in both the flavin and substrate binding pocket. Firstly, the presence of a positively charged residue in proximity to N1 and C2 flavin atoms stabilizes the anionic form of the reduced flavin. Secondly, the presence of the N5 atom of the flavin in close proximity to an enzyme hydrogen donor atom, typically a nitrogen atom from a side chain or protein backbone (Fraaije and Mattevi, 2000).

1.1.3 Classification of Flavoproteins

Flavoproteins can be classified according to their biological functions and the nature of the substrates involved in both the reduction and oxidation reactions into five major classes;

1. Transhydrogenases, catalyze the transfer of two electrons, as well as the appropriate hydrogen ions, from one organic substrate to another. These can then be further classified according to the atom types from which the electron pair is transferred to the flavin cofactor and that to which the electron is transferred.
2. Dehydrogenases/oxidases, transfer electrons from an organic substrate to molecular oxygen generating hydrogen peroxide without any measurable intermediates.
3. Dehydrogenases/oxygenases, transfer electrons from an organic substrate to molecular oxygen but unlike oxidases, they generate two water molecules and incorporate oxygen into another organic substrate,

forming the hydroxide derivative. In addition, a flavin hydroperoxide intermediate is observed.

4. Dehydrogenases/electron-transferases, transform a two-electron transfer into a one-electron transfer, where the reduction step involves the transfer of two electrons simultaneously to the flavin cofactor. On the other hand, the oxidation step occurs in two sequential steps, each involving the transfer of one electron to a different substrate.
5. Pure electron-transferases, involving only the transfer of one electron between an electron donor and an electron acceptor. This group of flavoenzymes have the ability to strongly stabilize the semiquinone form of the flavin cofactor (Massey and Hemmerich, 1980, Massey, 1995).

1.2 Quinone Oxidoreductases

Quinone oxidoreductases are flavoenzymes involved in the detoxification of quinones through a strict two electron reduction to the hydroquinone derivative, and thus avoiding the semiquinone derivative. Quinone oxidoreductases can be divided into three classes according to their sequence, the first class contains an extended C-terminal domain, such as NQO1, the second class have a shorter sequence, lacking the extended C-terminal domain, such as NQO2. The third class, which in addition to lacking the extended C-terminal domain, also have a deletion of approximately 18 residues between residues 64 and 82, such as *Escherichia coli* and *Saccharomyces cerevisiae* quinone reductases. Bacterial flavodoxins also have a similar deletion pattern

and lack the same amino acids and thus these quinone reductases are known as flavodoxin-like (Bianchet et al., 2004, Deller et al., 2008).

1.2.1 Quinones

1.2.1.1 Sources of Quinones

Quinones are abundant naturally occurring components of the electron transport chain, present in plants, bacteria and fungi. Moreover, quinones may be produced by the *in vivo* metabolism of naturally occurring compounds such as dopamine and adrenaline, as well as xenobiotics. Quinones are also present in the air as a result of pollution caused by cigarette smoke, car exhaust and burnt organic material. Moreover, quinones are also widely used as drugs for the treatment of various infections and cancer. Thus humans are regularly exposed to high amounts of these chemicals either through food intake or inhalation. Quinones may occur in different forms including monocyclic structures, such as benzoquinones, extended structures, such as diethylstilbestrone or condensed forms, such as benzo[*a*]pyrene-1,6-dione (Figure 1. 4) (Monks et al., 1992, Obrien, 1991).

1.2.1.2 Cytotoxicity of Quinones

The one electron reduction of quinones, such as that carried out by cytochrome P450, results in the production of the semiquinone derivatives, which mostly exist as radicals at physiological pH. The semiquinone radicals may then react with molecular oxygen resulting in the production of oxygen radicals, and regenerating the oxidized quinone, which can be reduced to the semiquinone derivative once again. The accumulation of the radical oxygen species interferes with the cellular oxidative defensive mechanisms by depleting

cellular stocks of glutathione (GSH), resulting in oxidative stress. Furthermore, oxygen radicals may also react with protein thiol and amine groups as well as DNA functional groups. This leads to the disruption of biological functions. The complete two electron reduction of the quinone results in the production of the hydroquinone derivative which is then conjugated with UDP-glucuronic acid, glutathione or sulphate ions and thus avoiding cellular toxicity caused by the semiquinone form (Monks et al., 1992, Obrien, 1991).

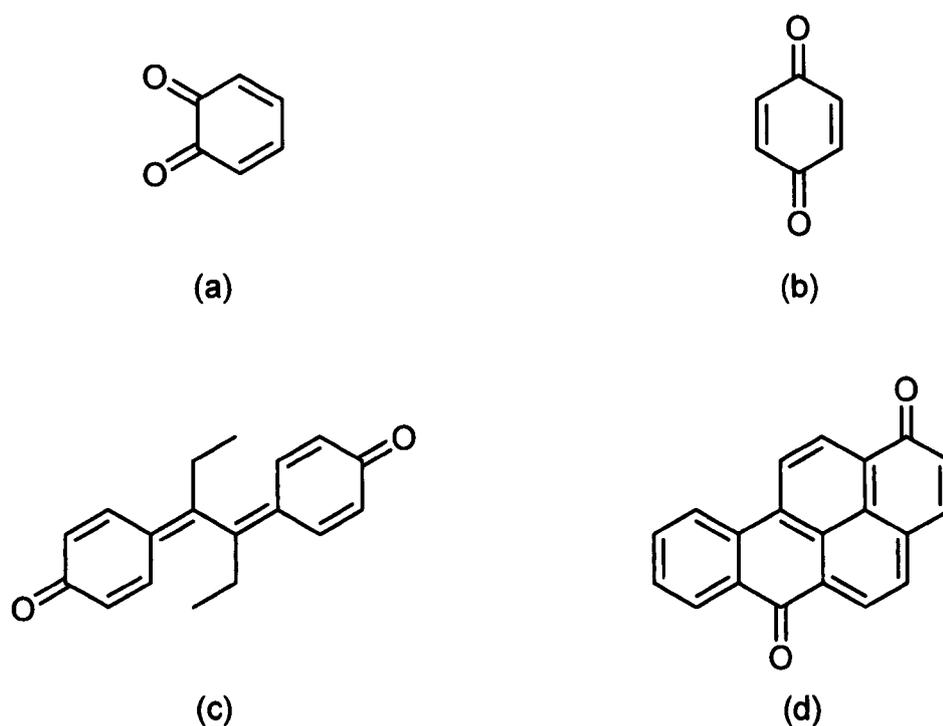


Figure 1. 4 Structures of benzoquinones (a, b), diethylstilbestrol (c) and benzo[a]pyrene-1,6-dione (d).

1.2.3 Primary Structure of NQOs

NQO1 and NQO2 share a sequence homology of approximately 50 per cent. The main structural difference between the two enzymes is the presence of an additional 44 residues forming the extended C-domain tail in NQO1. The alignment of NQO1 and NQO2 sequences shows a region that is highly

to contain Zn ions, but it is proposed to contain Cu ions *in vivo* due to the nature of the co-ordinating residues, which is most similar to type I Cu sites (Foster et al., 1999).

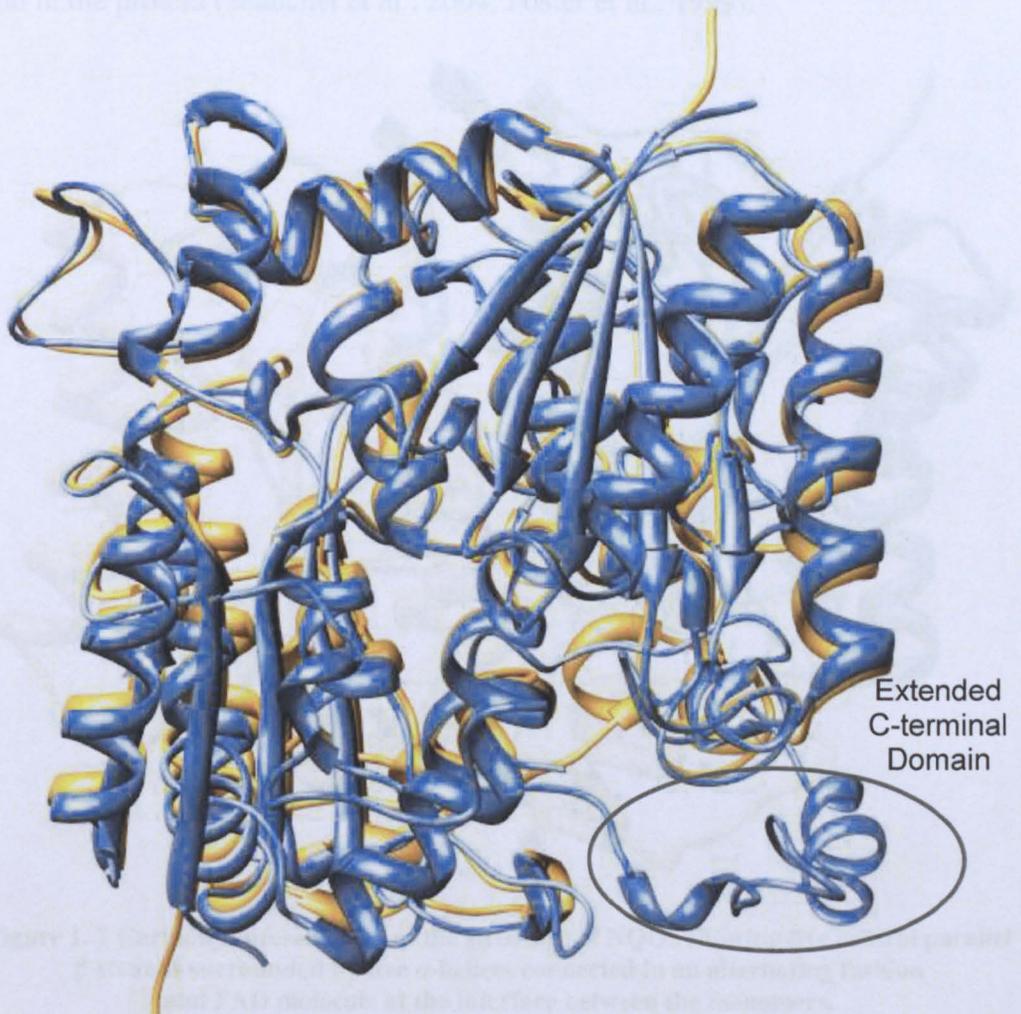


Figure 1.6 Superimposed structures of hNQO1 (pdb code:1D4A (Faig et al., 2000), yellow) and hNQO2 (pdb code:1QR2 (Calamini et al., 2008), blue)[†], showing the overall structures of the NQO1 and NQO2 dimers and the extended C-terminal domain of NQO1. Figure was created using UCSF Chimera (Pettersen et al., 2004).

The structure of the NQO2 monomer consists of two domains an N-terminal catalytic domain, which is similar to that of NQO1 and a C-terminal domain. The N-terminal is composed of residues 1 to 220 and consists of five central parallel β -strands surrounded by five α -helices that connect the consecutive

[†] Structures were overlaid using Sybyl 8.0 using “align structures by homology” command.

strands in an alternating α/β topology. This fold may be considered as a modified Rossman fold (Figure 1.7). The C-terminal domain of NQO2 consists of 10 residues which form a well defined loop, that co-ordinate with the metal ion in the protein (Bianchet et al., 2004, Foster et al., 1999).

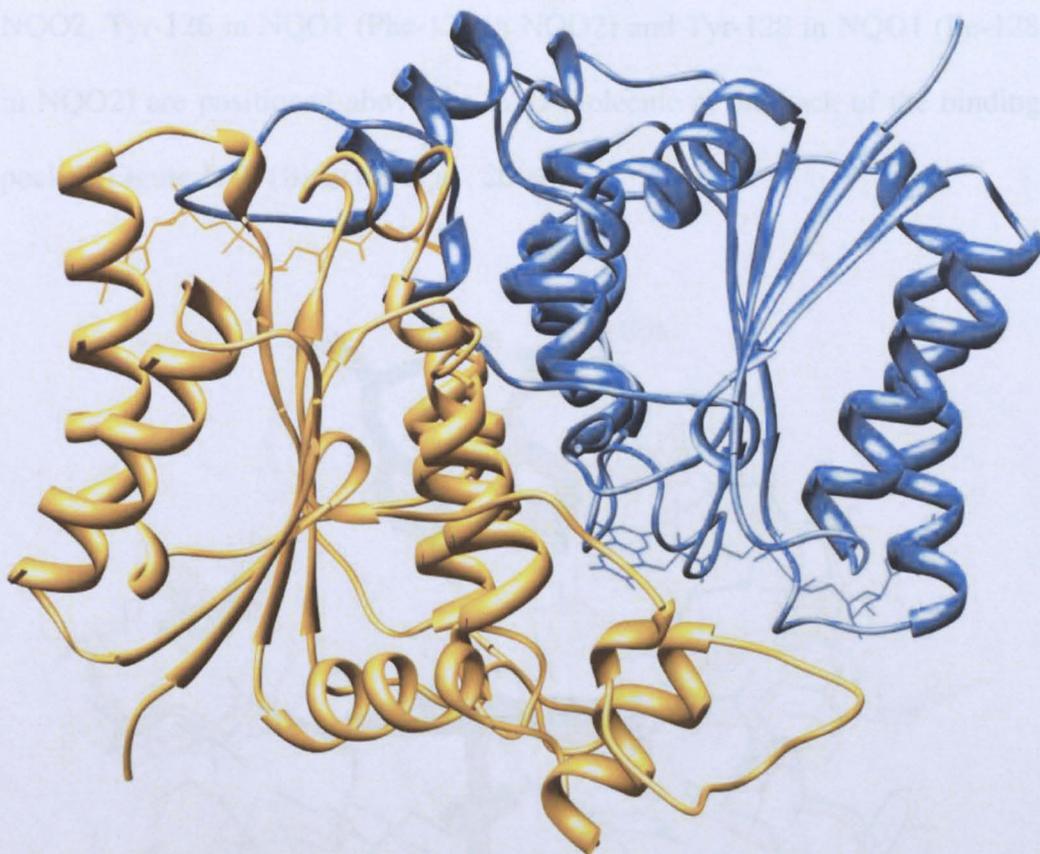


Figure 1.7 Cartoon representation of the structure of NQO2 showing five central parallel β -strands surrounded by five α -helices connected in an alternating fashion and FAD molecule at the interface between the monomers.
Figure was created using UCSF Chimera (Pettersen et al., 2004).

1.2.5 Catalytic Site of NQOs

The catalytic sites of both NQO1 and NQO2 are pockets located at the dimer interface (Figure 1.7). Both enzymes have large catalytic sites that are able to accommodate a wide variety of substrates, as is characteristic of enzymes that catalyze the detoxification of xenobiotics. The catalytic sites of both enzymes are very similar, with differences seen in only three residues; His-161, Tyr-126 and Tyr-128 in NQO1 which are replaced by Asn-161, Phe-126 and Ile-128

respectively in NQO2. The substitution of Tyr-128 in NQO1 with Ile-128 in NQO2 renders the catalytic site of NQO2 slightly wider and more polar (Figure 1. 8). Both binding pockets are lined with aromatic residues; Trp-105 and Phe-106 are at the bottom of the pockets. Phe-178, in both NQO1 and NQO2, Tyr-126 in NQO1 (Phe-126 in NQO2) and Tyr-128 in NQO1 (Ile-128 in NQO2) are positioned above the FAD molecule at the back of the binding pocket (Figure 1. 9) (Bianchet et al., 2004).

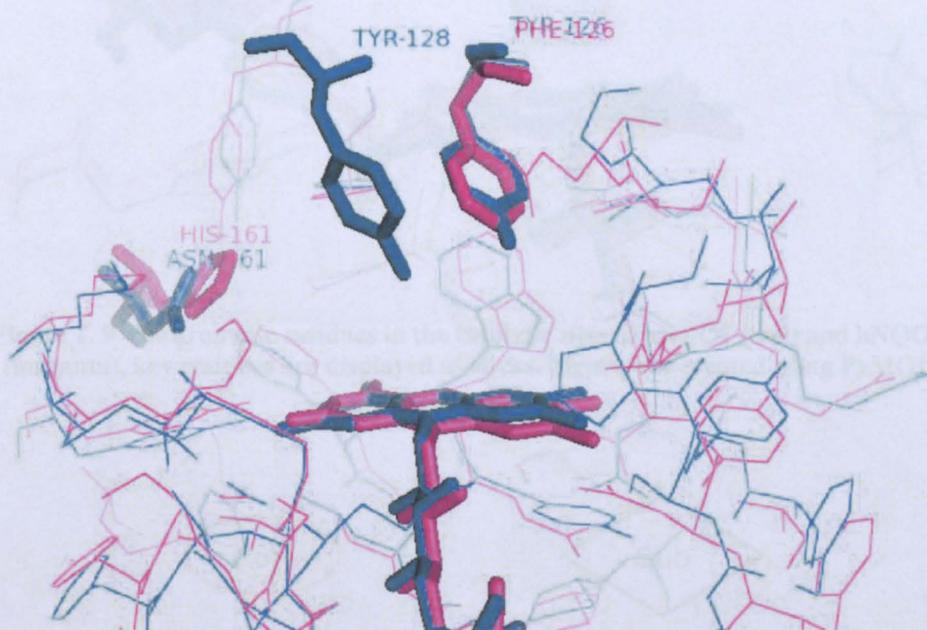


Figure 1. 8 The overlaid catalytic sites of hNQO1 (teal) and hNQO2 (magenta) showing the three residue difference between the active sites (displayed as sticks). The Ile-128 of NQO2 is not included in the image as it is 10.4 Å from the FAD molecule. Figure was created using PyMOL (Delano, 2008).

1.2.6 Mechanism of Action of NQOs

Quinone oxidoreductases are believed to function by a ping-pong mechanism, where two hydride transfer reactions occur without the cofactor leaving the catalytic site. The first hydride transfer is between the oxidized cofactor and the cosubstrate (electron-donor); the second is between the reduced cofactor and the substrate. Both the cosubstrate and the substrate bind in the same

catalytic site, and thus the cosubstrate must leave the catalytic site before the substrate can be reduced (Figure 1. 10) (Bianchet et al., 2004).

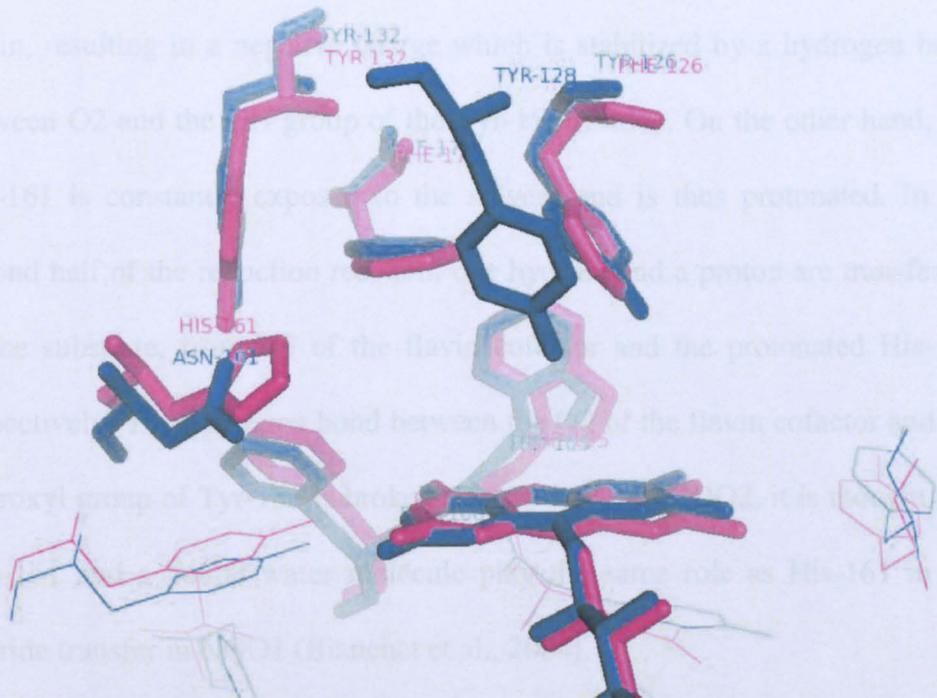


Figure 1. 9 The aromatic residues in the catalytic sites of hNQO1 (teal) and hNQO2 (magenta), key residues are displayed as sticks. Figure was created using PyMOL (Delano, 2008).

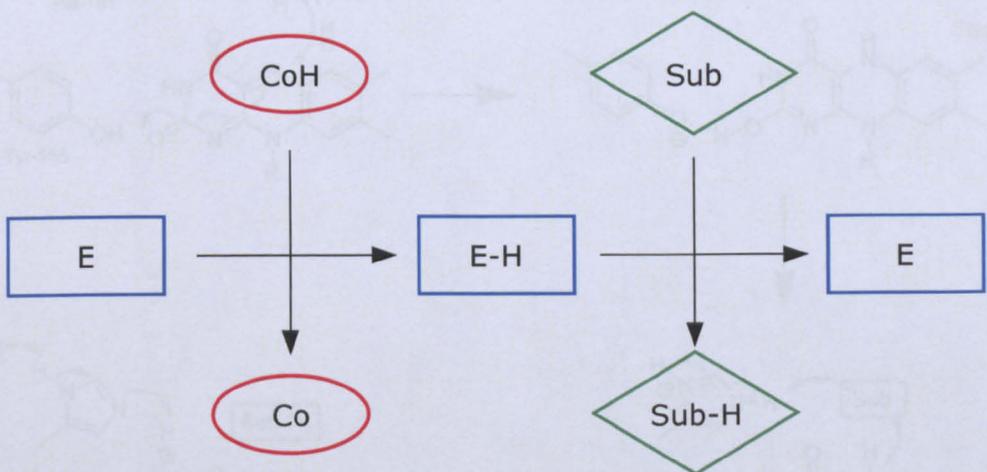


Figure 1. 10 Schematic representation of the ping-pong mechanism of reaction between enzyme, cosubstrate and substrate, where E represents the oxidized enzyme and E-H represents the reduced enzyme, CoH represents the reduced cosubstrate and Co represents the oxidized cosubstrate, Sub represents the substrate, while Sub-H represents the reduced substrate.

The mechanism of hydride transfer in NQO1 is believed to involve the N5 of the flavin cofactor, in addition to two residues in the catalytic site, His-161 and Tyr-155. The hydride is first transferred from the cosubstrate to N5 of the flavin, resulting in a negative charge which is stabilized by a hydrogen bond between O2 and the OH group of the Tyr-155 residue. On the other hand, the His-161 is constantly exposed to the solvent and is thus protonated. In the second half of the reduction reaction, one hydride and a proton are transferred to the substrate, from N5 of the flavin cofactor and the protonated His-161 respectively. The hydrogen bond between the O2 of the flavin cofactor and the hydroxyl group of Tyr-155 is broken (Figure 1.11). In NQO2, it is thought that Asn-161 and a bound water molecule play the same role as His-161 in the hydride transfer in NQO1 (Bianchet et al., 2004).

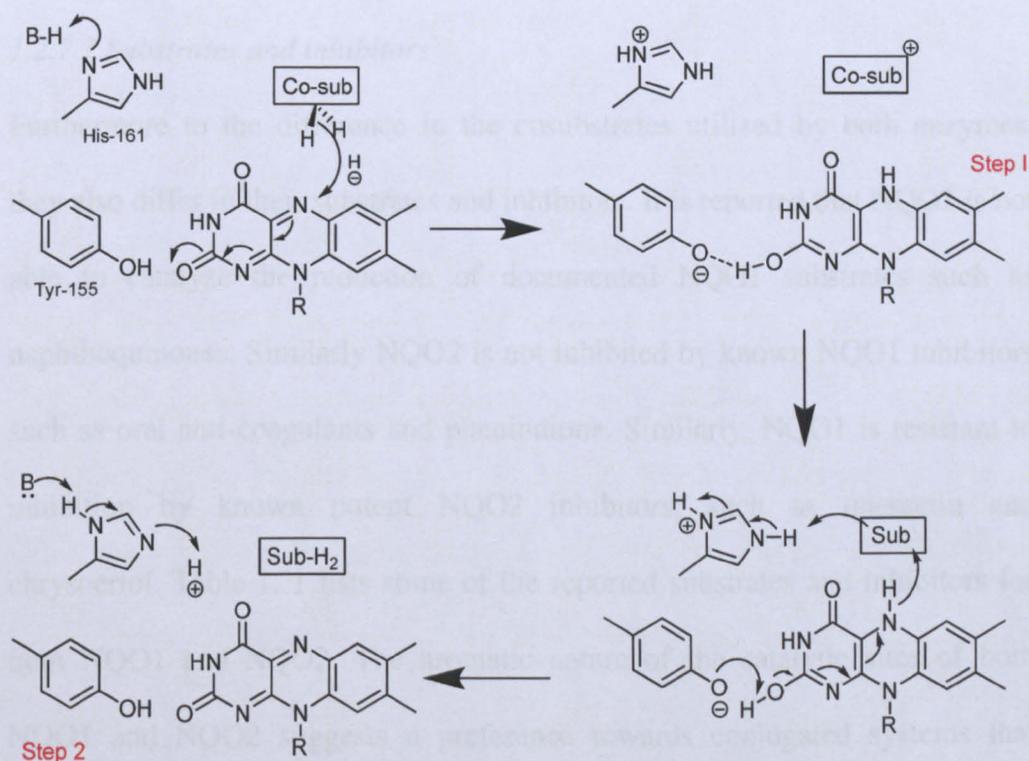


Figure 1. 11 Proposed mechanism of hydride transfer in NQO1.
Adapted from (Deller et al., 2008)

1.2.7 Cosubstrates, Substrates and Inhibitors of NQOs

Although NQO1 and NQO2 share a high sequence identity and have similar 3D structures and catalytic sites, they differ in the cosubstrates that they are able to utilise, in addition to different substrates and inhibitors.

1.2.7.1 Cosubstrates

NQO1 has the ability to utilize both nicotinamide adenine dinucleotide (NADH) and nicotinamide adenine dinucleotide phosphate (NADPH) *in vivo*. On the other hand, NQO2 utilizes either N-ribosyl- or N-alkyl-dihyronicotinamide as cosubstrates *in vivo* (Skelly et al., 2001, Deller et al., 2008). It has also been reported that NQO2 may also be able to utilize tetrahydrofolic acid as a cosubstrate (Boutin et al., 2005).

1.2.7.2 Substrates and inhibitors

Furthermore to the difference in the cosubstrates utilized by both enzymes, they also differ in their substrates and inhibitors. It is reported that NQO2 is not able to catalyze the reduction of documented NQO1 substrates such as naphthoquinones. Similarly NQO2 is not inhibited by known NQO1 inhibitors such as oral anti-coagulants and phenindione. Similarly, NQO1 is resistant to inhibition by known potent NQO2 inhibitors, such as quercetin and chrysoeriol. Table 1. 1 lists some of the reported substrates and inhibitors for both NQO1 and NQO2. The aromatic nature of the catalytic sites of both NQO1 and NQO2 suggests a preference towards conjugated systems that would be able to form additional π - π interactions.

	NQO1	NQO2
Substrates	Quinones Vitamin E Anti-cancer drugs (e.g. CB1954 and mitomycin C)	Ubiquinones Menadione Anticancer drugs (e.g. CB1954 and mitomycin C)
Inhibitors	Oral anticoagulants (e.g. dicoumarol and warfarin) Cibacron blue Phenindone	Antimalarials (e.g. chloroquine and quinacrine) Benzo(a)pyrene Flavone derivatives (e.g. Quercetin) Resveratrol

Table 1. 1 Reported substrates and inhibitors of NQO1 and NQO2(Boutin et al., 2005) .

1.2.8 Distribution of NQOs

NQO1 is widely distributed in the human body with the exception of the liver and blood cells, particularly red blood cells (RBCs), where NQO1 is expressed at very low levels. It has been observed that there are large variations in the levels of expressed NQO1 between individuals. A higher concentration of NQO1 was reported in certain tumour tissues such as breast, brain, colon, lung and liver tumours compared to normal tissues. On the other hand, NQO2 is less widely distributed in the human body and is more selectively expressed in kidney, lung, liver, heart, skeletal muscles and RBCs. It has also been reported that NQO2 is more expressed in colorectal and liver tumours compared to non cancerous cells. Moreover, the levels of NQO2 expressed in known chemosensitive tissues such as bone marrow, were found to be very low (Wu et al., 1997, Knox and Chen, 2004, Vella et al., 2005).

1.2.9 Physiological functions of NQOs

The primary role of NQO1 is thought to be the detoxification of xenobiotics, such as quinones via two electron reduction. It is also thought to have a role in

the metabolism of endogenous quinones, such as ubiquinone and vitamin E quinone. NQO1 is thought to be responsible for generating the vitamin E hydroquinone, which is considered to have stronger antioxidant properties than vitamin E (Ross and Siegel, 2004). On the other hand, the function of NQO2 is less well understood and there are different opinions and evidence in relation to its function. The selective expression of NQO2 in certain tissues suggests a tissue-specific action. NQO2 knockout mice are less susceptible to toxic doses of menadione when compared to the wild type. However, they are more susceptible to the carcinogenic effect of polycyclic aromatic hydrocarbons (Vella et al., 2005). NQO2 is also thought to carry the MT3 binding site for melatonin and is inhibited by melatonin (Calamini et al., 2008). NQO2 polymorphism has been linked to Parkinson's disease as well as schizophrenia (Harada et al., 2001, Harada et al., 2003). The ability of NQO2 to reduce coenzyme Q0 suggests that coenzyme Q10 might be a substrate of NQO2 and that NQO2 might play a role in the respiratory chain, by maintaining high levels of reduced coenzyme Q10 in the cells (Boutin et al., 2005). Both NQO1 and NQO2 are thought to play a role in the reduction of vitamin K to its hydroquinone derivative. Oral anti-coagulants were found to be potent inhibitors of NQO1 but not of NQO2, hence it is thought that NQO2 is a dicoumarol-insensitive vitamin K reductase (Chen et al., 2000).

1.2.10 Medical Applications of NQOs

Antimalarial drugs such as chloroquine, primaquine and quinacrine are potent inhibitors of NQO2 and thus inhibit its activity in RBCs. This leads to the accumulation of quinones and consequently oxidative stress, caused by the

formation of semiquinone and their auto-oxidation producing reactive oxygen species. The absence of NQO1 in the RBCs increases their susceptibility to the oxidative stress caused by the inhibition of NQO2 (Kwiek et al., 2004). Quinone oxidoreductases may also be used to reduce anticancer prodrugs such as CB1954 and mitomycin C to their active metabolites, which is discussed in more detail in section 1.5.2.

1.3 Bacterial Nitroreductases

Polynitroaromatic compounds are usually synthetic compounds that are toxic and mutagenic to several forms of living organisms. Consequently, some bacterial species have developed the ability to metabolise polynitroaromatic compounds by reducing the nitro groups (Esteve-Nunez et al., 2001, Spain, 1995). Bacterial nitroreductases are divided into two groups, depending on their response to the presence of oxygen, and the number of electrons they utilise during the reduction process. Oxygen sensitive (type II) bacterial nitroreductases utilise the transfer of one electron to the nitro group to produce the nitro anion radical, which is then reoxidized in the presence of oxygen to regenerate the aromatic nitrocompound (Esteve-Nunez et al., 2001). Oxygen insensitive (type I) bacterial nitroreductases are FMN dependant flavoenzymes that transfer two electrons from the cofactor to reduce the aromatic nitro group via the nitroso or/and the hydroxylamine derivative into the amino derivative; in some cases the hydroxylamine may be the final product of the reduction (Figure 1. 12) (Dolores Roldan et al., 2008, Esteve-Nunez et al., 2001). Bacterial nitroreductases are able to bioactivate anticancer prodrugs and thus are of clinical interest.

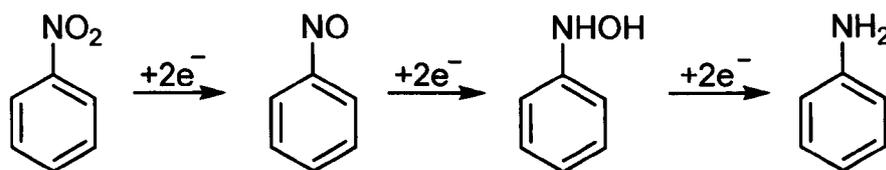


Figure 1. 12 Stepwise reduction of nitroaromatic compounds to the amino derivative.

1.4 Cancer

Cancer can be broadly defined as “a collection of diseases characterized by unregulated cell growth leading to the invasion of surrounding tissues and metastasis to other parts of the body (King and Robins, 2006).” The causes of cancer are not known, but there is a correlation between the incidence rates of certain types of cancers and environmental factors such as urban air pollution, and lifestyle choices such as smoking, excessive alcohol intake and fatty diets (Goordarz Danaei, 2005).

1.4.1 The Development of Cancer

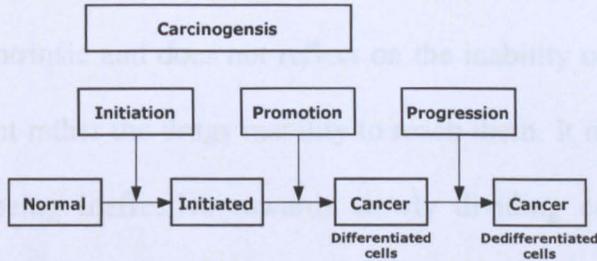
All cancers are monoclonal in origin, i.e. arise from the proliferation of one abnormal cell. There are different models describing the sequence of events in the development of cancer, which depend on the source of the data used in the model development, which may range from animal experiments through to cell and molecular biology, and to clinical studies (Figure 1. 13) (King and Robins, 2006).

1.4.2 Hypoxia in Tumours

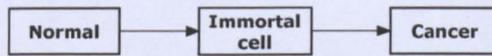
Tumours can be divided into three histologically distinct regions. The first is the anoxic region which is composed of dead necrotic cells; the second is the

avascular hypoxic region which is characterized by hypoxic cells. The third region is known as the vascularised axoic region

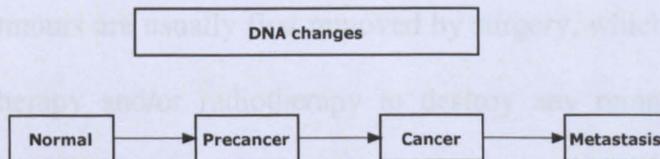
Model 1: Animal experiments



Model 2: Cell biology



Model 3: Molecular biology



Model 4: Clinical data

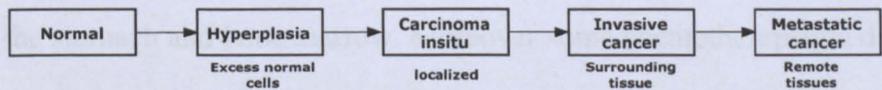


Figure 1. 13 Cancer development models, adapted from (King and Robins, 2006)

There are several mechanisms by which the hypoxic areas are formed in tumours. The growth of cancer cells requires the development of a blood vessel supply which occurs by a process known as angiogenesis. The resultant blood vessels are highly irregular with leaky walls and sluggish flow. This results in perfusion-related (acute) hypoxia due to the inadequate supply of oxygen and nutrients. Diffusion-related (chronic) hypoxia is caused by the growth of the tumour leading to the inadequate supply of oxygen to cells distant from blood

vessels supplying nutrients and oxygen. Finally, anaemic hypoxia is caused by the decreased capacity of the blood to transport oxygen. Hypoxic regions of tumours are more resistant to treatment using ionizing radiation, as well as to most anticancer drugs. The resistance of the hypoxic areas to anticancer drugs is not intrinsic and does not reflect on the inability of the drug to attack these areas but rather the drugs inability to reach them. It may also be caused by the drugs being ineffective towards slowly dividing cells, as seen in hypoxic regions of solid tumours (Brown, 2000, Martin Brown and Giaccia, 1998, Vaupel and Harrison, 2004).

1.5 Cancer Treatment

Solid tumours are usually first removed by surgery, which is then followed by chemotherapy and/or radiotherapy to destroy any remnants of the tumour. Chemotherapeutic agents are generally non specific towards cancerous cells and hence have a variety of side effects, usually towards fast dividing cells such as the stomach and bone marrow. Moreover some chemotherapeutic drugs may have specific toxicity towards vital organs such as the liver and kidney (King and Robins, 2006).

1.5.1 Anticancer Targeted Prodrug Therapy

Prodrugs are inactive molecules that are converted *in vivo* into their active forms. Anticancer targeted prodrugs can selectively target tumours by several mechanisms. One such mechanism is the activation of the prodrug by exogenous enzymes delivered to the tumour cells by either Anti-body-directed enzyme prodrug therapy (ADEPT) or gene directed enzyme prodrug therapy

(GDEPT). The prodrug may also be selectively activated by endogenous enzymes overexpressed in cancerous cells. The selective activation of anticancer targeted prodrugs in cancer cells lead to higher therapeutic effectiveness and a lower incidence of side effects compared to conventional drugs.

1.5.1.1 ADEPT

In this approach, the drug activating enzyme is conjugated with an antibody that specifically binds to an antigen expressed onto the surface of tumour cells. Once the prodrug is administered, it is activated by the conjugated enzyme at the tumour site. The efficiency of this technique is enhanced by the use of small drug molecules generated from the activation of the prodrug which are able to diffuse into tumour cells that do not express the specific antigen (Figure 1. 14) (Knox and Connors, 1995, Singh et al., 2008).

1.5.1.2 GDEPT

In gene directed enzyme prodrug therapy, a gene encoding the activating enzyme is delivered inside the tumour cell using vectors. The choice of the vector used to deliver the gene into the cells is the most challenging step in GDEPT. Viral vectors are the most effective method for gene delivery, as utilized by virus-directed enzyme prodrug therapy (VDEPT), despite the fact that they may cause a severe immunogenic response. Other possible vectors used for gene delivery include bacterial, natural and synthetic vectors. The expressed enzyme should not be present in normal cells to ensure the selectivity of the treatment. The enzyme is then expressed by transfected

tumour cells[‡] and the prodrug is activated. The prodrug must be stable and sufficiently lipophilic to be able to diffuse into the untransfected cells of the tumour. As with ADEPT, the diffusion of the activated drug to neighbouring untransfected cells increases the efficiency of the treatment (Figure 1. 15) (Singh et al., 2008, Hedley et al., 2007).

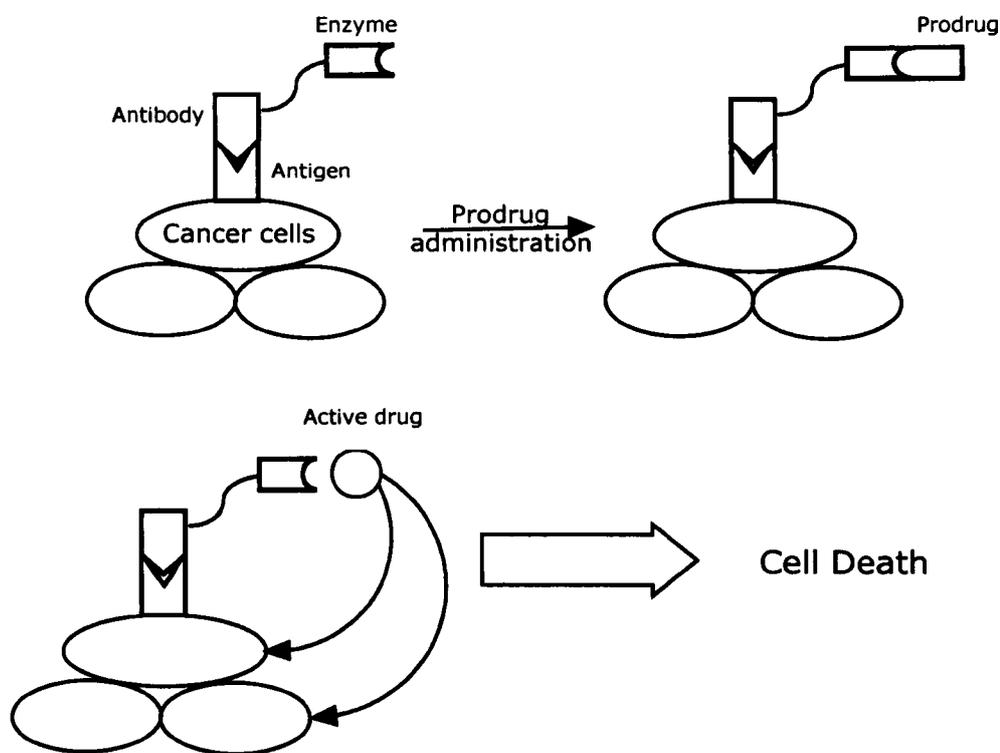


Figure 1. 14 Representation of ADEPT. Adapted from (Singh et al., 2008).

1.5.2 Bioreductive Anticancer Drugs

Bioreductive anticancer drugs are prodrugs that are reduced to their active derivative using enzymes, where the active reduced product then leads to cell death. The bioreduction of the drug may occur by exogenous enzymes delivered to the tumour cells or endogenous enzymes expressed at higher concentrations in cancer cells. The bioreduction of the drug may also occur by

[‡] Cells containing the delivered gene and expressing the enzyme of interest.

endogenous enzymes active only under hypoxic conditions thus minimizing the side effects towards normal non-hypoxic cells (Chen and Hu, 2009). Since only a proportion of the tumour is hypoxic, drugs activated under hypoxic conditions may not be effective against the whole of the tumour. However, the activated drug may diffuse out of the hypoxic cells killing the neighbouring non hypoxic population. This phenomena is known as the ‘bystander effect’ (Denny, 2001).

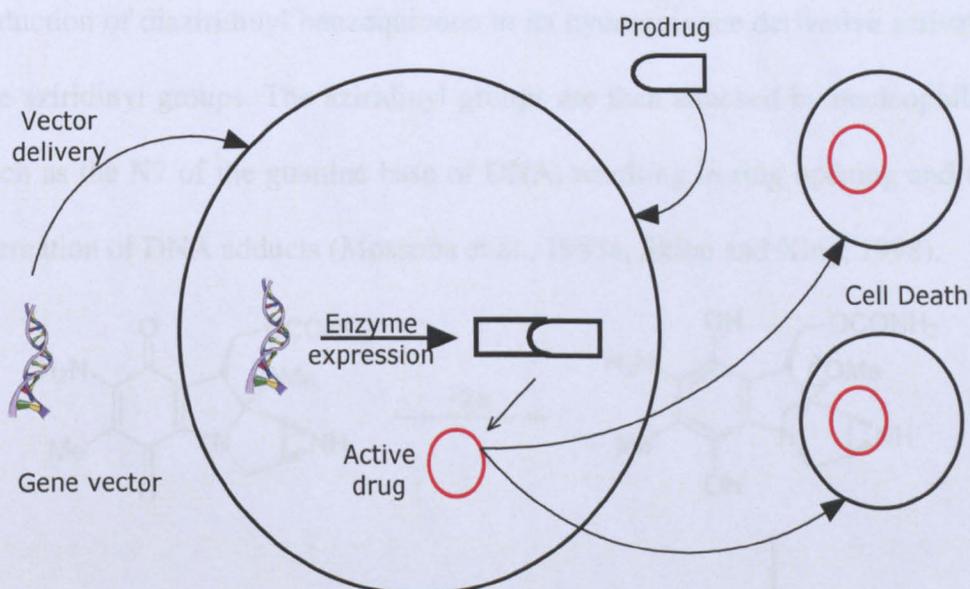


Figure 1. 15 Representation of GDEPT. Adapted from (Singh et al., 2008).

There are four main classes of anticancer prodrugs that are designed for bioreductive activation; quinones, nitroaromatics, N-oxides and metal complexes (Chen and Hu, 2009). Both quinones and nitroaromatics can be reduced *in vivo* to their active form by quinone oxidoreductases such as NQO1 and NQO2, which are over expressed in certain cancer tumours. Bacterial nitroreductases are able to reduce nitroaromatics to their active metabolite, and may be delivered to cancer cells using either GDEPT or VDEPT techniques.

1.5.2.1 Quinones

Mitomycin C (MMC) is an example of a quinone anticancer prodrug that is currently being clinically used. It is reduced by both NQO1 and NQO2 to the hydroquinone derivative, which is then attacked by two nucleophilic centres in the DNA, to form DNA cross links (Figure 1. 16) (Chen and Hu, 2009, Colucci et al., 2008, McGuinness et al., 1991, Lin et al., 1980). Other examples of bio-reductive drugs activated by both NQO1 and NQO2 include indolequinones such as EO9 and diaziridinyl benzoquinones such as AZQ (Figure 1. 17). The reduction of diaziridinyl benzoquinone to its hydroquinone derivative activates the aziridinyl groups. The aziridinyl groups are then attacked by nucleophiles, such as the N7 of the guanine base of DNA, resulting in ring opening and the formation of DNA adducts (Mossoba et al., 1985a, Skibo and Xing, 1998).

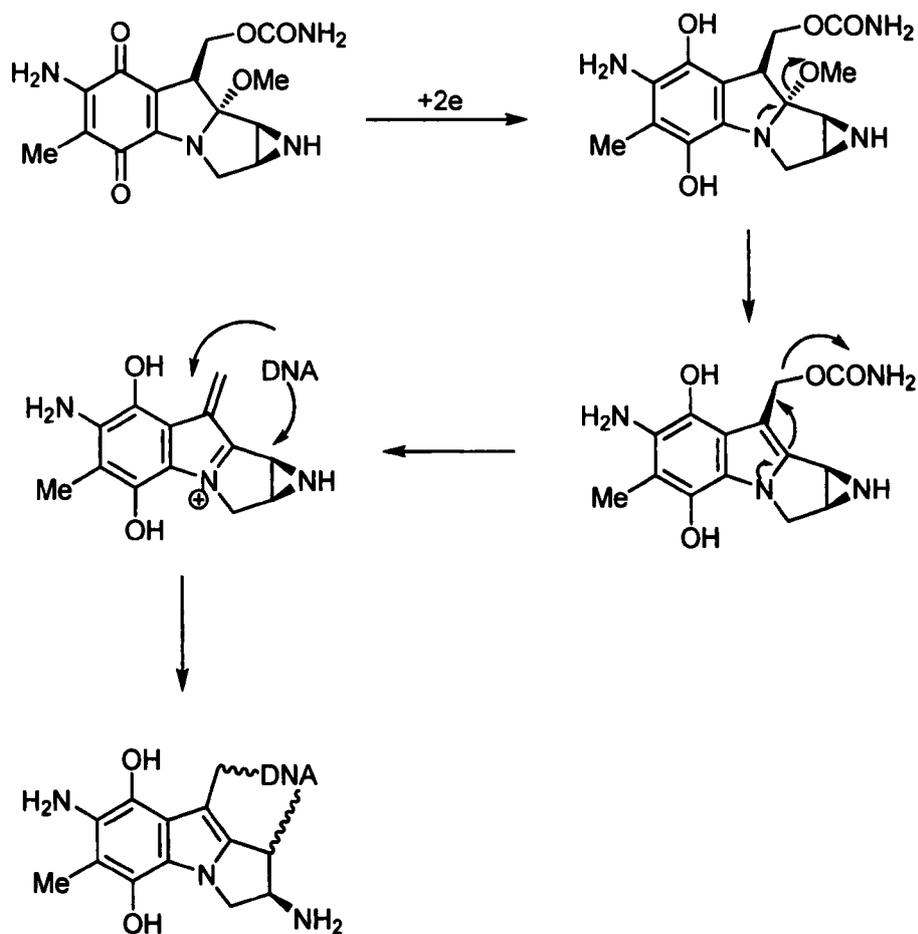


Figure 1. 16 Reduction of MMC and formation of DNA cross links

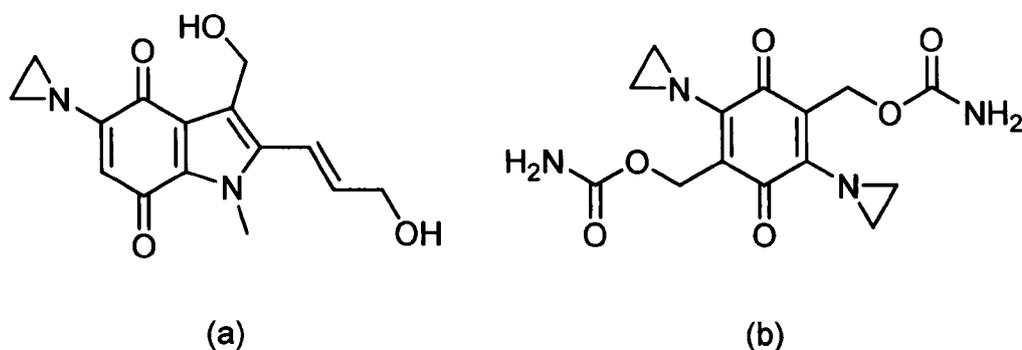


Figure 1. 17 Structures of (a) EO9 and (b) AZQ

An alternative use for the quinone moiety is the release of the active drug molecule upon its bioreduction in the cancer cells (Figure 1. 18). This technique can be applied to indolequinones, naphthoquinones and benzoquinones (Chen and Hu, 2009, Jaffar et al., 1998). EO9 has been proposed as such a delivery system, where upon its reduction the active metabolite of EO9 and another anticancer drug are both produced (Chen and Hu, 2009). Furthermore, the chemical association of some drugs with a quinone moiety has produced promising results by increasing bioavailability and/or decreasing the incidence of side effects. The chemical association of 5-fluorodeoxyuridine (FUDR) with indolequinone resulted in increased selectivity and cytotoxicity towards hypoxic cells. The increase in cytotoxicity is thought to be caused by the electrophilic iminium cation resulting from the release of the drug (Tanabe et al., 2005). Camptothecin is a potent DNA topoisomerase inhibitor however it has unfavourable properties such as poor water solubility and stability, in addition to its poor selectivity. Its chemical association with indolequinone was found to increase its selectivity towards

hypoxic cells, in addition to increasing its water solubility and stability (Zhang et al., 2005).

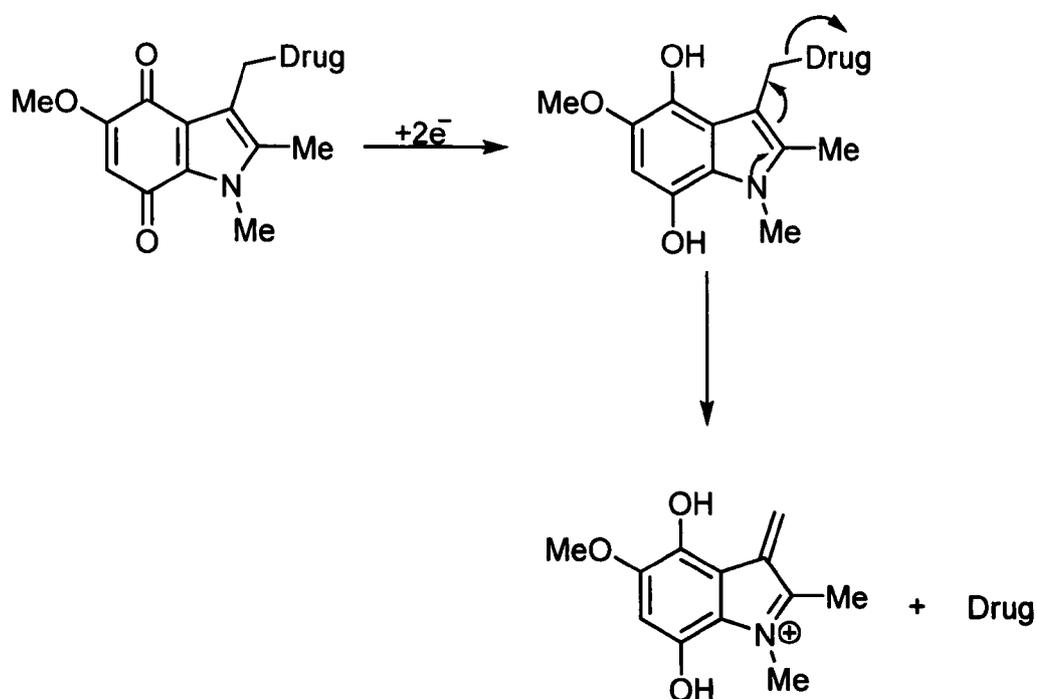


Figure 1. 18 Release of active drug upon reduction of indolequinone derivative.

1.5.2.2 Nitroaromatics

Flavoenzymes reduce nitroaromatic compounds through a stepwise reaction to the amino derivative via the nitroso and hydroxylamine derivatives (Figure 1. 12). In most cases, the hydroxylamine is the final product of the reduction and is the active metabolite. The reduction of the nitro group, a strong electron withdrawing group, to a hydroxylamine, an electron donating group, changes the electronic environment of the aromatic ring. This subsequently can lead to the activation of a latent alkylating species attached to the aromatic ring, as seen with CB1954. As with quinones, nitroaromatic compounds can also be used as delivery systems to release the active drug upon reduction.

1.5.2.2.1 Bioreduction of CB1954

CB1954 (Figure 1.19) is a nitroaromatic prodrug reduced by both bacterial nitroreductases and quinone oxidoreductases to the hydroxylamine derivative. The reduction of the nitro group activates the aziridinyl group allowing nucleophilic attack on the aziridine ring by DNA. The 4-hydroxylamine metabolite then reacts with thioesters such as acetyl-CoA to probably generate 4-(N-acetoxy)-5-aziridin-1-yl)-2-nitrobenzamide. The resultant product then attacks the DNA forming DNA cross links (Knox et al., 1988). *Escherichia coli* and *Enterobacter cloacae* nitroreductases are able to reduce both nitro groups with similar efficiency resulting in the production of equimolar amounts of the 2- and 4- hydroxylamine metabolites, which may be further reduced to produce the 2- and 4- amine derivatives (Knox et al., 1992). Although NQO1 and NQO2 are have been classified as quinone reductases, they are also able to catalyze reduction of some nitroaromatic compounds, such as CB1954. Unlike bacterial nitroreductases, quinone oxidoreductases only reduce the 4-nitro group, generating the 4-hydroxylamine derivative (Figure 1.19) (Knox et al., 1992, Skelly et al., 2001). It has been established that rat NQO1 and human NQO2 are more effective in reducing CB1954 than human NQO1 and mouse NQO1, despite the high degree of homology shared between them. The difference between the catalytic ability of the enzymes is attributed to residue 104, which in rat NQO1 and human NQO2 is a tyrosine and in human and mouse NQO1 is a glutamine (Figure 1. 5) (Chen et al., 1997, Skelly et al., 2001, Knox and Chen, 2004). There have been contradicting discussions on the cytotoxicity of the reduction products. It is thought that the 4-hydroxylamine derivative is more cytotoxic as it is able to form DNA cross links. On the other

hand, recent studies have suggested that the 2-amine derivative may also play a role in the cytotoxic effect of CB1954, due to its superior ability to diffuse into the neighbouring cells. This is due its higher stability and more lipophilic nature compared to the hydroxylamine derivative (Helsby et al., 2004).

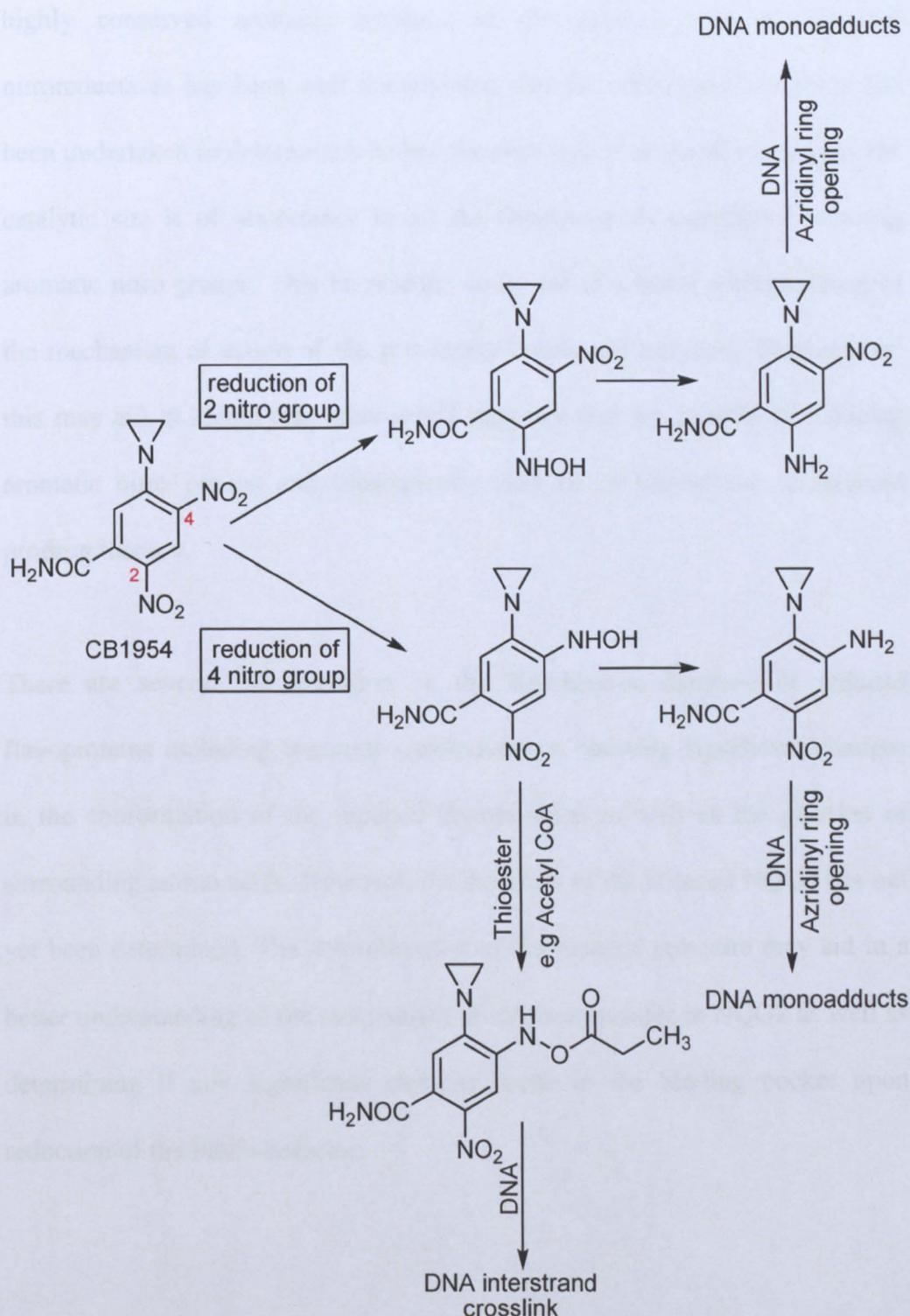


Figure 1. 19 Structure and bioactivation of CB1954 by quinone oxidoreductases.

There is an increasing interest in NQO2 and other enzymes that are able to reduce aromatic nitro groups such as bacterial nitroreductases. This may be attributed to their role in targeted anticancer therapy by activating bioreductive anticancer drugs, such as quinones and nitroaromatics. The importance of highly conserved aromatic residues in the catalytic site of bacterial nitroreductases has been well documented. On the other hand, no study has been undertaken to determine whether the presence of aromatic residues in the catalytic site is of importance in all the flavoenzymes capable of reducing aromatic nitro groups. This knowledge could aid in a better understanding of the mechanism of action of the previously mentioned enzymes. Furthermore, this may aid in identifying other novel enzymes that are capable of reducing aromatic nitro groups and subsequently may be of importance in targeted prodrug therapy.

There are several 3D structures in the Brookhaven database of reduced flavoproteins including bacterial nitroreductases showing significant changes in the conformation of the reduced flavoproteins as well as the position of surrounding amino acids. However, the structure of the reduced NQO2 has not yet been determined. The determination of the reduced structure may aid in a better understanding of the mechanism of electron transfer in NQO2 as well as determining if any significant changes occur in the binding pocket upon reduction of the flavin cofactor.

1.6 Aims of the Thesis

The project aims on utilizing X-ray crystallography, fluorescence spectroscopy and bioinformatics to increase the understanding of the mechanism of action of and identifying the presence of conserved residues between flavoenzymes reducing aromatic nitro groups. The specific aims of the project are listed below:

- The use of bioinformatics computational tools and homology modelling to identify key conserved aromatic residues in the catalytic site of flavoproteins reducing aromatic nitro groups.
- The determination of the kinetic profile of NQO2 with novel substrates using fluorescence spectroscopy.
- The crystallization of NQO2 and reduction of NQO2 crystals.
- Structure determination of reduced NQO2 using X-ray crystallography.
- Analysis of reduced 3D structure of NQO2 to gain more insight into the mechanism of action of NQO2 and probable route of electron transfer between the reduced flavin cofactor and substrate.

Chapter Two

Introduction to X-ray Crystallography

2.1 Introduction

The determination of the 3D structures of proteins helps in gaining valuable knowledge about the biological roles and mechanisms of action of proteins. Moreover, the determination of the structure of a protein ligand complex may aid in identifying key binding sites and interactions between the protein and the ligand. Furthermore, it aids in the identification of any changes in the protein structure upon ligand binding. This information is used in structure-based drug design, where the aim is to design more specific and potent drugs. A number of techniques can be used to determine the 3D structure of biological macromolecules, which include nuclear magnetic resonance spectroscopy (NMR), small angle X-ray scattering (SAXS), and X-ray crystallography.

X-ray crystallography utilises the distinct patterns produced by the diffraction of X-ray beams by the crystalline form of proteins. X-ray crystallography involves several steps; the first is the protein expression and purification. This is followed by the crystallization of the protein. The production of a crystal of suitable size and quality is usually a limiting step in the utilization of this technique and is a process of trial and error. Once a crystal of suitable quality is obtained, it is then mounted onto a rotating goniometer in the path of the X-rays. The diffracted beams are then detected by an X-ray detector and the resulting images are processed using specialized software to determine the protein structure (Figure 2.1).

2.2 Protein Crystallization

A crystal can be defined as an orderly three-dimensional arrangement of identical or different atoms, ions or molecules. The process of forming a crystal from a non-crystalline material is called crystallization.

In a saturated solution, the solute molecules are in a dynamic equilibrium with the solvent. When an increase in the solid phase will be accompanied by equivalent decrease in the liquid phase.

Crystallization is a process in which a solid phase is formed from a liquid phase. It is the opposite of melting. The process of crystallization is a reversible process.

Crystallization is a process in which a solid phase is formed from a liquid phase. It is the opposite of melting. The process of crystallization is a reversible process.

Crystallization is a process in which a solid phase is formed from a liquid phase. It is the opposite of melting. The process of crystallization is a reversible process.

Crystallization is a process in which a solid phase is formed from a liquid phase. It is the opposite of melting. The process of crystallization is a reversible process.

Crystallization is a process in which a solid phase is formed from a liquid phase. It is the opposite of melting. The process of crystallization is a reversible process.

Crystallization is a process in which a solid phase is formed from a liquid phase. It is the opposite of melting. The process of crystallization is a reversible process.

Crystallization is a process in which a solid phase is formed from a liquid phase. It is the opposite of melting. The process of crystallization is a reversible process.

Crystallization is a process in which a solid phase is formed from a liquid phase. It is the opposite of melting. The process of crystallization is a reversible process.

Crystallization is a process in which a solid phase is formed from a liquid phase. It is the opposite of melting. The process of crystallization is a reversible process.

Crystallization is a process in which a solid phase is formed from a liquid phase. It is the opposite of melting. The process of crystallization is a reversible process.

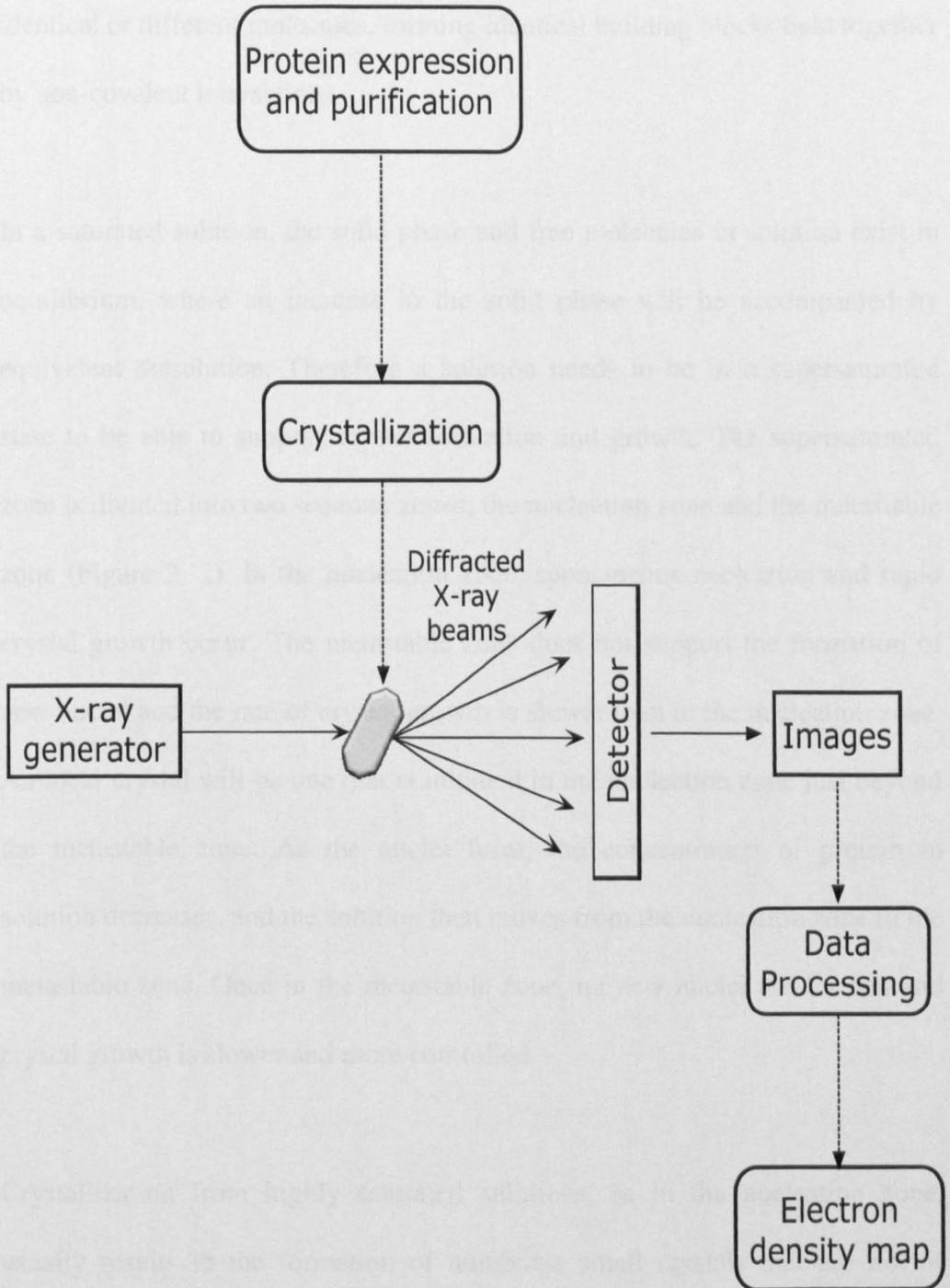


Figure 2. 1 Schematic representation of the processes involved in structure determination using X-ray crystallography.

2.2 Protein Crystallization

A crystal can be defined as an orderly three dimensional arrangement of identical or different molecules, forming identical building blocks held together by non-covalent interactions.

In a saturated solution, the solid phase and free molecules in solution exist in equilibrium, where an increase in the solid phase will be accompanied by equivalent dissolution. Therefore a solution needs to be in a supersaturated state to be able to support crystal formation and growth. The supersaturated zone is divided into two separate zones; the nucleation zone and the metastable zone (Figure 2. 2). In the nucleation zone, spontaneous nucleation and rapid crystal growth occur. The metastable zone does not support the formation of new nuclei and the rate of crystal growth is slower than in the nucleation zone. An ideal crystal will be one that is initiated in the nucleation zone just beyond the metastable zone. As the nuclei form, the concentration of protein in solution decreases, and the solution then moves from the nucleation zone to the metastable zone. Once in the metastable zone, no new nuclei are formed and crystal growth is slower and more controlled.

Crystallization from highly saturated solutions, as in the nucleation zone, usually results in the formation of numerous small crystals that are not of suitable size for X-ray diffraction. Moreover, the rapid growth of the crystals usually results in defective crystals containing impurities (McPherson, 2009). The crystal will then cease to grow once the protein concentration has dropped back to the soluble region. In some cases, the crystal may cease to grow even

when the protein concentration has still not reached the solubility limit. This is thought to be due to the accumulation of impurities or denatured protein onto the crystal surface preventing the further growth of the crystal (Durbin and Feher, 1996).

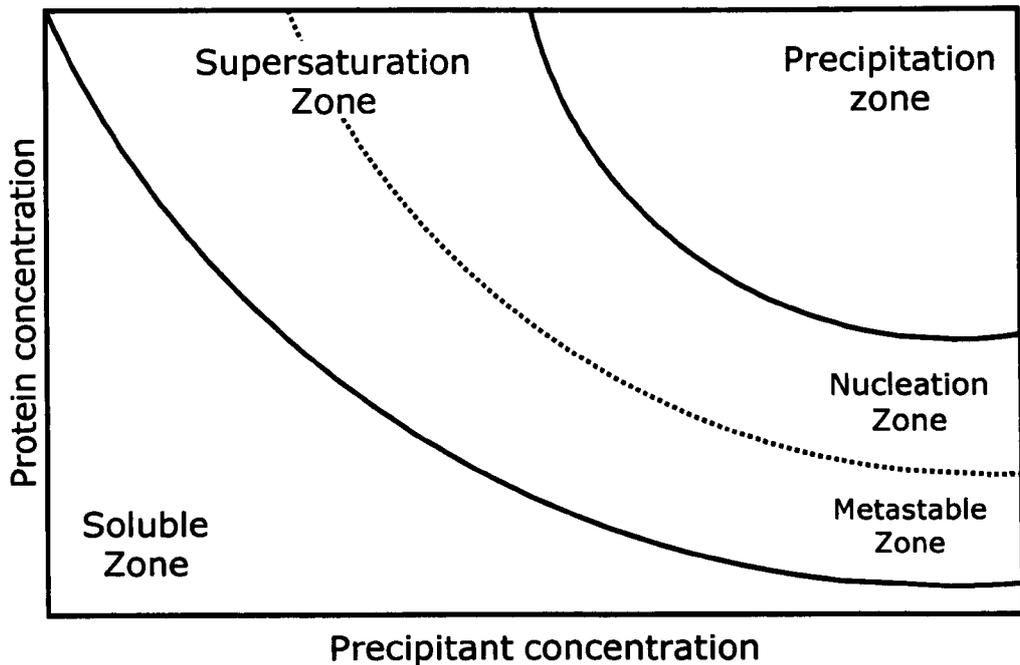


Figure 2. 2 Phase diagram for crystallization, showing precipitation zone, supersaturation zone and soluble zone.

2.2.1 Protein crystallization techniques

The most widely utilized technique in protein crystallization is vapour diffusion. This technique depends on the removal of water from a microdrop, containing a mixture of protein solution and reservoir solution, to a larger more concentrated reservoir solution containing the precipitant. There are two different procedures that rely on the vapour diffusion technique; sitting drop and hanging drop. In the sitting drop technique, the microdrop is placed on a plastic post in the middle of the reservoir solution. The well is then sealed using transparent tape (Figure 2. 3 (a)). In the hanging drop technique, the

microdrop hangs upside down on a microscope cover slip on top of the reservoir solution. The well is sealed using silicon vacuum grease between the coverslip and the crystallization plate (Figure 2. 3 (b)) (McPherson, 2009, Rhodes, 2000).

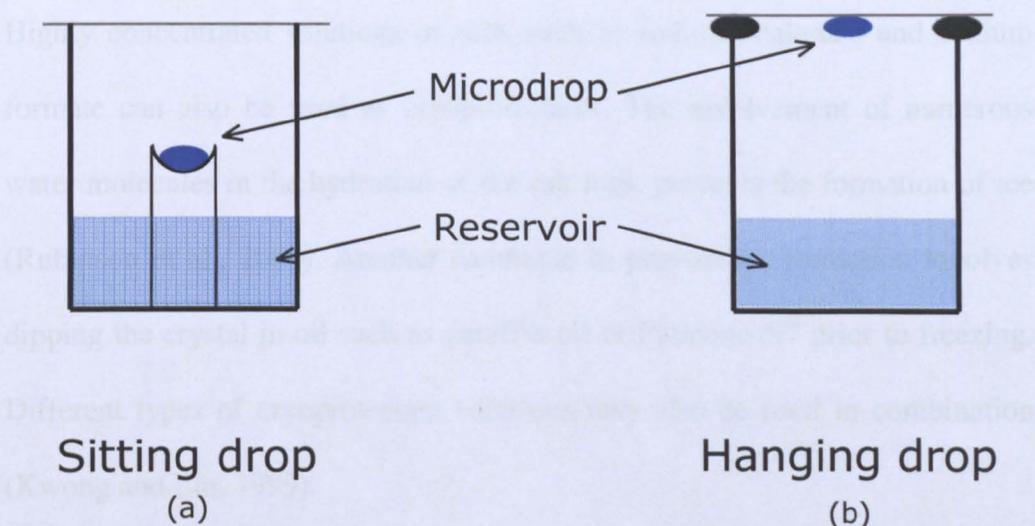


Figure 2. 3 Schematic representations of sitting and hanging drop vapour diffusion methods.

2.2.2. Cryocrystallography

Collecting diffraction data at or near 100K decreases the amount of radiation damage to which the crystal is subjected to. This is due to the decrease in the diffusion rate of the free radicals produced by the X-ray radiation, which decreases the distance they are able to travel in the crystal and thus the amount of damage they cause. This prolongs the crystal lifetime and allows the collection of higher quality datasets and in some cases may allow the collection of higher resolution datasets. Due to the high water content of macromolecular crystals, they must be placed in a protective ‘cryo-protectant’ solution before freezing. This is to prevent the formation of ice which is problematic in macromolecular crystallography as it interferes with the diffraction of the X-

rays and disrupts the crystal order (Garman, 1999, Garman, 2003, Drenth, 2007, Hope, 1988). There are many types of cryoprotectant solutions; these include small organic molecules such as polyethylene glycol, xylitol and glycerol. These act by disrupting the order of water molecules and thus preventing the formation of ice (Garman, 1999, Garman, 2003, Drenth, 2007). Highly concentrated solutions of salts such as sodium malonate and sodium formate can also be used as cryoprotectants. The involvement of numerous water molecules in the hydration of the salt ions, prevents the formation of ice (Rubinson et al., 2000). Another technique to prevent ice formation involves dipping the crystal in oil such as paraffin oil or Paratone-N[®] prior to freezing. Different types of cryoprotectant solutions may also be used in combination (Kwong and Liu, 1999).

2.2.3 Nature of the Crystal

A crystal can be defined by four parameters; the asymmetric unit, the space group, the unit cell and the crystal lattice. The unit cell is the smallest possible volume that when repeated is representative of the whole crystal and acts as the building block of the crystal. It is defined using six values, the lengths of the edges a , b and c and the angles between them α , β and γ . Alternatively, the unit cell may be defined using three vectors \bar{a} , \bar{b} and \bar{c} , which are coincident with the unit cell edges and are of the same length. The unit cell is usually composed of several asymmetric units, which can be defined as the largest volume of atoms that do not possess any symmetry. In the case of proteins, this may be one protein molecule, or if the protein possesses self symmetry, then it may be a fraction of the molecule. The arrangement of the asymmetric units

within the unit cell defines the space group to which the unit cell belongs. The unit cells are then repeated in a regular manner to form the crystal (Figure 2. 4). A lattice is defined as a regular infinite arrangement of similar points. If every molecule in the unit cell is represented by a point, the crystal lattice is generated. There are fourteen different 3D lattices, known as Bravais lattices (McPherson, 2009, Massa, 2004, Ladd and Palmer, 2003).

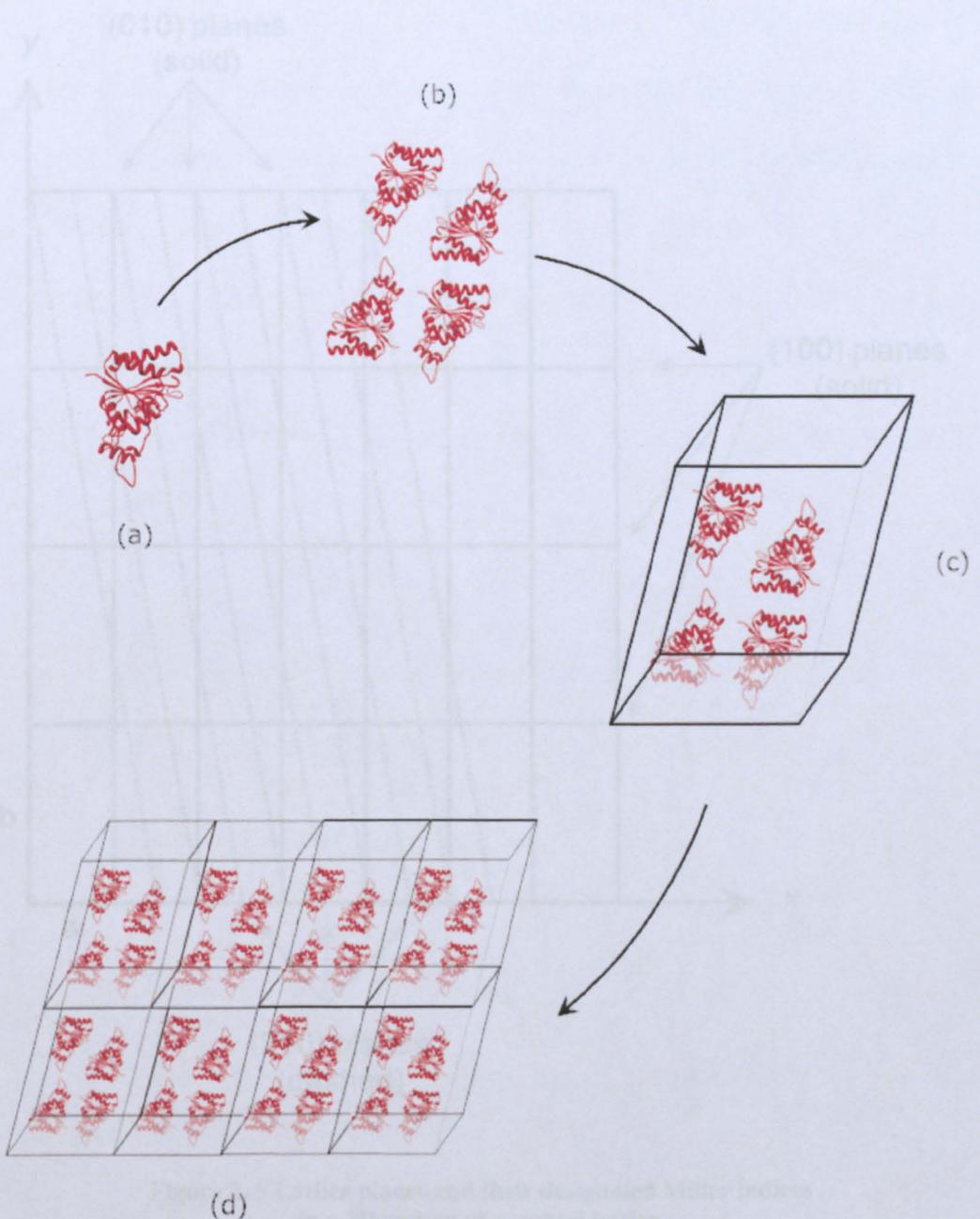


Figure 2. 4 Schematic representation of crystal parameters showing (a) asymmetric unit, (b) the spatial relationship between different asymmetric units, (c) unit cell and (d) crystal. Adapted from (McPherson, 2009)

2.2.3.1 Lattice planes and Miller indices

Lattice planes are families of related planes that intersect the crystal lattice at regular intervals. The planes are defined by three values known as Miller indices (hkl), corresponding to the number of times they intersect \bar{a} , \bar{b} and \bar{c} respectively (Figure 2. 5) (Massa, 2004, McPherson, 2009).

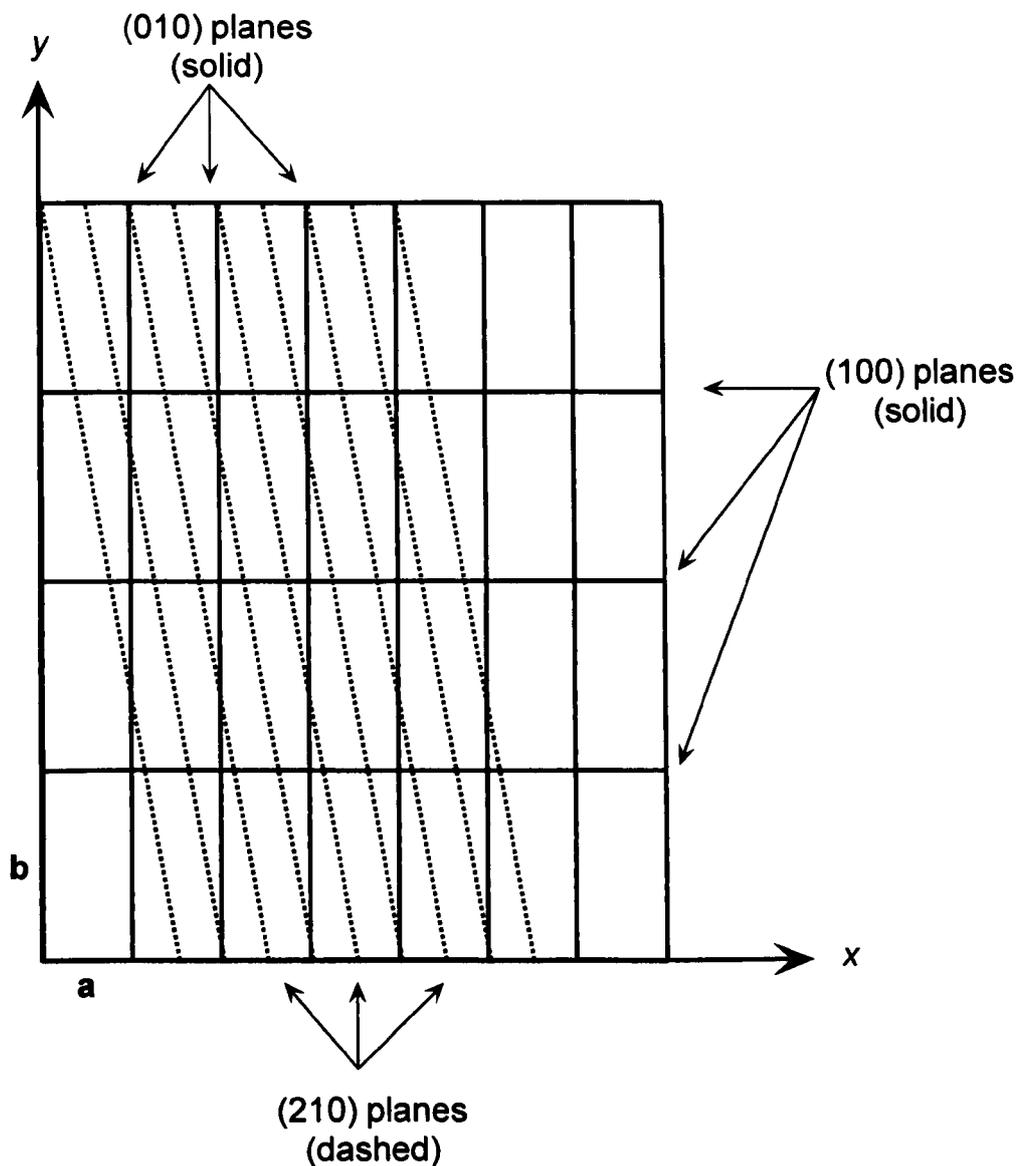


Figure 2. 5 Lattice planes and their designated Miller indices in a 2D section of a crystal lattice.

2.2.4 X-ray Diffraction by Crystals

2.2.4.1 Bragg's Law

Diffraction of X-rays occurs due to the interaction of the X-ray photons with electrons. The excited electrons then emit X-rays in random directions at the same wavelength as the incident beam, as they return to the ground state. Due to the repetitive nature of crystals, the diffracted X-rays from different planes of the crystal may interfere with each other. The scattered X-rays may be *out of phase* and thus interfere with each other destructively cancelling one another. In some cases, the scattered X-rays may be *in phase* resulting in constructive interference, amplifying their intensity. According to Bragg's law, equation (2.1), constructive interference occurs only when the distance travelled by the X-ray beams scattered from parallel planes is an integer of the wavelength of the incident X-ray beams (Figure 2. 6) (Bragg and Bragg, 1913, Blow, 2002).

$$n\lambda = 2d \sin \theta \quad (2.1)$$

Where d is the distance between planes, θ is the angle of incidence of the beam, λ is the wavelength of the incident beam and n is an integer.

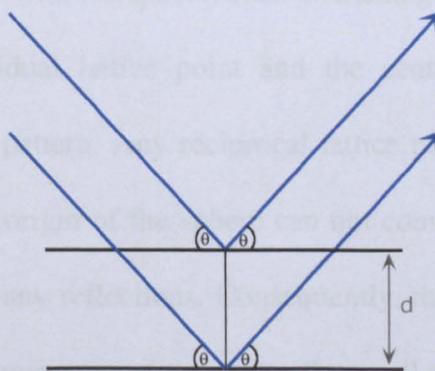


Figure 2. 6 Constructive interference as defined by Bragg's law, where d is the distance between planes and θ is the angle of incidence of the beam.

2.2.4.2 Reciprocal lattice and Ewald's Sphere

The reciprocal lattice is constructed using the end points of vectors representing each family of planes. The vector for each family is perpendicular to the set of planes and has a length of $1/d_{hkl}$. The dimensions of the reciprocal unit cell are defined using a^* , b^* and c^* . The reciprocal lattice is based on $1/d_{hkl}$ and thus the real and reciprocal unit cells dimensions are inversely related. Therefore for a large unit cell, the reciprocal unit cell is small and vice versa. It is important to note that the reciprocal lattice rotates in the same manner as the crystal (Drenth, 2007, McPherson, 2009, Authier, 1981).

The reciprocal lattice aids in determining the best orientation of a crystal, with respect to the incident beam, that satisfies Bragg's law and thus generating strong reflections. This is achieved via the construction of a sphere passing through O, the origin of the reciprocal lattice, with the diffracting crystal at its centre, C. This is known as Ewald's sphere (Ewald, 1921). Ewald's sphere demonstrates that when a reciprocal lattice point passes through the sphere, the associated family of planes satisfy Bragg's law and thus a diffracted ray is produced (Figure 2. 7). As the crystal is rotated, different reciprocal lattice points come in contact with the sphere, each diffracting the beam along the line connecting the individual lattice point and the centre of the sphere, thus forming a diffraction pattern. Any reciprocal lattice point that is not within a distance of $2/\lambda$ of the origin of the sphere can not come in contact with it and thus can not generate any reflections. Consequently, the number of reflections is dependant on the dimensions of the unit cell as well as the wavelength of the X-ray beam. The use of shorter wavelength radiation increases the diameter of the sphere and thus allows for the inclusion of more lattice points and therefore

the generation of more reflections. In addition, due to the inverse relationship between the dimensions of the real unit cell and the reciprocal unit cell, large real unit cells have smaller more densely populated reciprocal unit cells. This is accompanied by an increase in the number of reflections (McPherson, 2009, Rhodes, 2000).

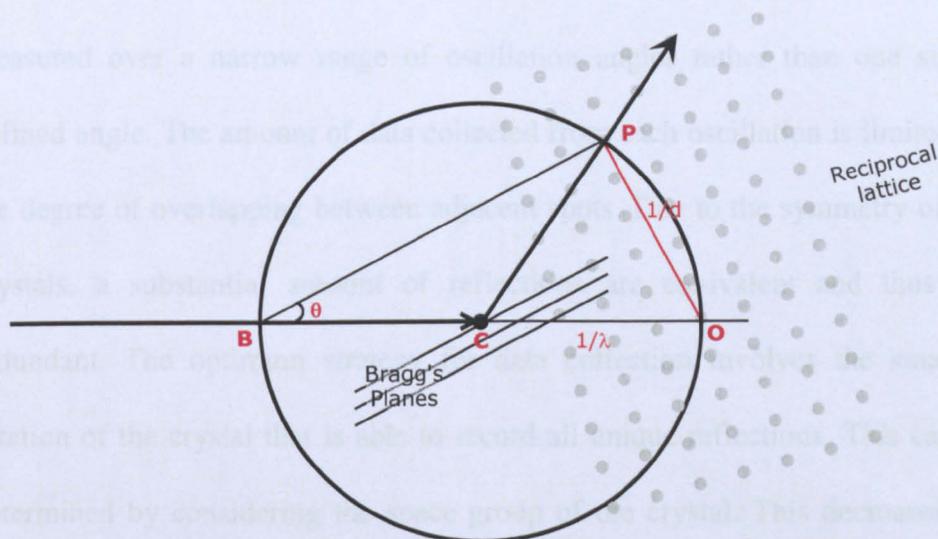


Figure 2.7 Construction of Ewald's sphere, where BC represents the incident beam and CP represents the diffracted beam and O is the origin on the reciprocal lattices. Adapted from (Dauter, 1999).

2.3 Data collection

During data collection the crystals are mounted onto goniometers which allow the correct orientation of the crystal in the path of the X-ray beam. Throughout data collection, the mounted crystal can be rotated to allow different angles of the crystal to be exposed to the X-ray beam. The angle of rotation of the crystal is known as the oscillation angle.

2.3.1 Data Collection Strategy

Crystals are generally not perfectly packed and can thus be represented by a collection of small mosaic blocks, which can be considered as optically

independent each affecting the path of the X-ray beam in a slightly different manner. This is especially true for protein molecules which are flexible and held together by weak forces. As a result of the slight irregularities in the arrangement of the building blocks, the diffracted beam takes the form of a cone rather than a straight line (Figure 2. 8). This results in larger elliptical spots, spread over several diffraction images. Thus each reflection must be measured over a narrow range of oscillation angles rather than one single defined angle. The amount of data collected from each oscillation is limited by the degree of overlapping between adjacent spots. Due to the symmetry of the crystals, a substantial amount of reflections are equivalent and thus are redundant. The optimum strategy for data collection involves the smallest rotation of the crystal that is able to record all unique reflections. This can be determined by considering the space group of the crystal. This decreases the amount of radiation damage a crystal is subjected to due to the decrease in X-ray exposure time (Rhodes, 2000).

2.4 Data Processing

2.4.1 Indexing and Integration of Images

Initial data processing can be divided into three stages; indexing, cell refinement and integration. These can be carried out by a variety of data processing programme suites such as MOSFLM (Leslie, 1992), XDS (Kabsch, 2010) and HKL2000 (Otwinowski and Minor, 1997). Indexing is the process by which the unit cell dimensions and the crystal orientation are determined. This gives an indication of the most likely Laue group of the crystal. This is then followed by the refining of the previously determined

parameters. These parameters include the cell parameters; such as unit cell dimensions, crystal orientation and mosaicity, as well as the detector and beam parameters. Once the correct unit cell parameters have been established, the images can be integrated. This involves predicting the position and calculating the intensity of the spots in the collected images, according to the determined Laue group. For accurate integration, a good estimation of background intensity values must be calculated (Leslie, 2006, Kabsch, 2010).

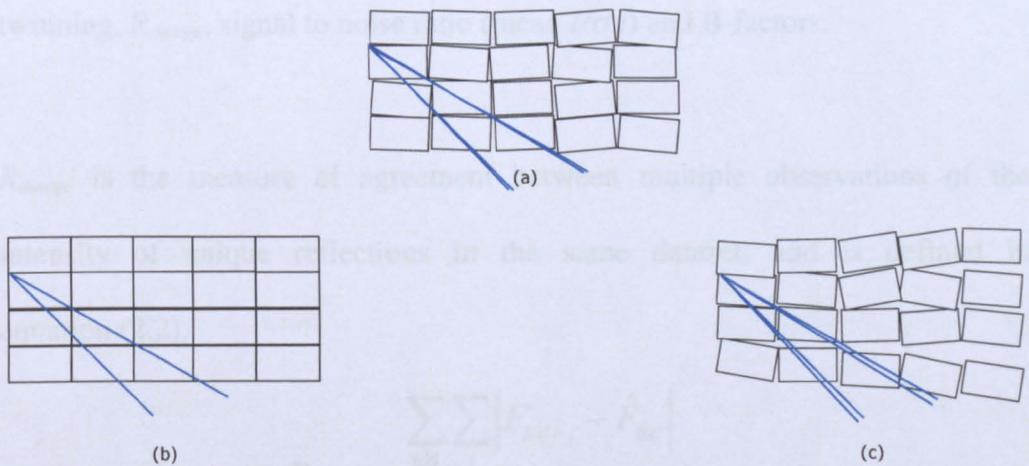


Figure 2. 8 Schematic representation of varying degrees of crystal packing and the associated effect on the shape of the diffracted X-ray beams. (a) Shows a typical crystal, (b) shows perfect arranged crystals resulting in sharp diffraction spots and (c) shows highly irregular crystals resulting in elliptical diffraction spots.

Adapted from (Schneider, 2007).

2.4.2 Scaling and Merging of Reflections

The spot intensities are affected by different experimental factors such as radiation damage and changes in the intensity of the incident beam. Therefore, the measured intensities from different images are not all on the same scale. Consequently, estimated corrections are applied to the intensities to bring them all on to the same scale. This improves the consistency of the data within a dataset. Scaling also gives an estimate of the data quality and identifies parts of the collected data that are in poor agreement with the remainder of the data.

After the intensities are corrected, the intensities are then merged so that an average intensity is calculated for each hkl reflection (Evans, 2006, Rhodes, 2000, Kabsch, 2010).

2.4.2.1 Assessing Data Quality

There are a number of criteria that can be used to assess the quality of the data and can aid in determining whether parts of the data collected should be discarded. These criteria include completeness, presence or absence of twinning, R_{merge} , signal to noise ratio (mean $I/\sigma I$) and B-factors.

R_{merge} is the measure of agreement between multiple observations of the intensity of unique reflections in the same dataset, and is defined in equation (2.2).

$$R_{merge} = \frac{\sum_{hkl} \sum_j |F_{hkl-j} - \hat{F}_{hkl}|}{\sum_{hkl} \hat{F}_{hkl}} \quad (2.2)$$

Where F_{hkl-j} is the j th measurement of F_{hkl} and \hat{F}_{hkl} is the average value of all measurements of F_{hkl} (McPherson, 2009).

The value of R_{merge} increases if the crystal is damaged by X-ray exposure and by any errors that arise during the data collection or integration processes. The use of R_{merge} as the sole indicator for data quality is not reliable, as its value increases with the increase in data redundancy (Evans, 2006, Weiss, 2001, Weiss and Hilgenfeld, 1997). *Completeness* is another data quality indicator and is defined as the fraction of unique reflections that have been measured of

those that could be observed in theory. The nominal resolution limits chosen at the data processing stage to maximise the quality data should take into account the values of R_{merge} , mean $I/\sigma I$ and completeness (Kleywegt, 2000). The examination of the values of both B-factors and R_{merge} of each image provides valuable information about radiation damage and the occurrence of errors during the data collection and integration processes. This then allows for the exclusion of images or groups of images that are significantly worse than others.

2.4.3 The Structure Factors and the Phase Problem

The diffraction images provide atomic positions in reciprocal space which can be assigned by miller indices (hkl) and the intensity of each reflection which is proportional to the square of the amplitude of the structure factor F_{hkl} . The structure factor is the sum of the diffractive contribution of all atoms in the unit cell to the intensity of a single reflection. The structure factor can be described by equation (2.3)

$$F_{hkl} = \sum_{j=1}^n f_j e^{2\pi i(hx_j + ky_j + lz_j)} \quad (2.3)$$

Where n is the number of atoms in the unit cell and f_j is the scattering power of atom j with co-ordinates x_j , y_j and z_j in the unit cell.

The electron density in a unit cell can be calculated by performing a Fourier transformation on all the structure factors to give the Fourier transform in equation (2.4),

$$\rho(x, y, z) = \frac{1}{V} \sum_h \sum_k \sum_l |F_{hkl}| e^{-2\pi i(hx+ky+lz)+i\alpha(hkl)} \quad (2.4)$$

Where $\rho(x, y, z)$ is the electron density at position (x, y, z) , V is the volume of the unit cell, F_{hkl} is the structure factor for each diffraction spot and $\alpha_{(hkl)}$ is the phase angle for the diffracted waves.

As mentioned before the amplitude of the structure factors is proportional to the intensity of the recorded spots, but the measurement of the phase angles is not possible experimentally. This is known as the phase problem and can be overcome by several methods such as; molecular replacement, isomorphous replacement, anomalous scattering and the heavy atom method (Rhodes, 2000, McPherson, 2009).

2.4.3.1 Molecular Replacement

Molecular replacement was developed in 1962 by Rossmann and Blow (Rossmann and Blow, 1962), but its use has become more common as the number of available 3D structures has increased. It is thought that about 70 per cent of the phases of newly deposited structures are determined using molecular replacement (Evans and McCoy, 2008). Molecular replacement utilizes the phases of a known molecular model to provide an initial estimate for a new structure. This involves two steps; rotation and translation of the model. This is repeated until the predicted diffraction pattern of the model best matches the observed reflections. The phases of the model protein are then used with the experimentally determined structure factor amplitudes to produce an initial electron density map. Recently developed molecular replacement software utilise the maximum likelihood probability approach, which takes into

account consistent errors in the data. It is thought that the choice of model is the most important step in molecular replacement. A high sequence identity between the model and protein under study usually indicates high structural similarities, which increases the probability of success of the molecular replacement step (Evans and McCoy, 2008).

2.4.4 Electron Density Maps

There are several different types of electron density maps used to relate the calculated model and the diffraction data and determine the degree of agreement between the built model and the experimental data. The two most commonly used are the difference maps and the composite maps. Traditionally composite maps are contoured at $1\sigma^{\S}$, while difference maps are contoured at $3\sigma^{\S}$ (Tronrud, 2010).

2.4.5 Refinement of the Model

Refinement is defined as the changing of the model parameters, in order to best explain the experimental observations. The model parameters include atomic positions defined by the x, y and z co-ordinates, B-factors and occupancy. B-factors describe the uncertainty of the atom position within the collected data due to either the vibration of the atom or defects in the collected data. The experimental observations include unit cell parameters, structure factor amplitudes and standardized stereochemistry. The number of model parameters to be refined is much larger than the number of experimental observations; this causes the refinement process to be problematic. Therefore stereochemical

[§] Sigma values relate to the probability of finding an electron within the volume enclosed by the electron density map, where 1σ and 3σ correspond to 99% and 68% respectively.

restraints such as bond lengths and angles are included in order to increase the number of observations.

Rigid body refinement is often the first step in the refinement process. The protein is treated as a rigid body, and thus only the overall position and orientation in the unit cell of the protein are refined. This is then usually followed by restrained refinement where each atom is refined separately but restraints are placed on the geometry of the molecule, in order to find the best solution, where the difference between the observed and calculated structure factors are minimal, with idealized geometry (Tronrud, 2004, Drenth, 2007).

2.4.5.1 Refinement Indicators

To decrease the incidence of errors in the model, the refinement process is monitored using a quality indicator known as R_{factor} , which is defined in equation (2.5). It measures the difference between the calculated structure factors and the observed structure factors. A desirable value for R_{factor} is between 0.18 and 0.25 according to the resolution and data quality.

$$R_{factor} = \frac{\sum_{hkl} \left| |F_{obs}| - |F_{calc}| \right|}{\sum_{hkl} |F_{obs}|} \quad (2.5)$$

Where F_{obs} is the observed structure factor and F_{calc} is the calculated structure factor

R_{factor} can be biased towards the model and thus another quality indicator known as R_{free} , which is less prone to overfitting, was introduced (Brunger, 1992). A percentage of unique reflections, usually between 5 to 10 per cent are

set aside and not included in the refinement process. This fraction of data is then used in calculating R_{free} according to equation (2.6). The monitoring of R_{free} as well as R_{factor} helps in preventing the over fitting of the data and model bias (Kleywegt, 2000).

$$R_{free} = \frac{\sum_{hkl(free)} \left| |F_{obs}| - |F_{calc}| \right|}{\sum_{hkl(free)} |F_{obs}|} \quad (2.6)$$

Where F_{obs} is the observed structure factor for the free unrefined data and F_{calc} is the calculated structure factor for the free unrefined data

2.4.6 Model Quality

The peptide back bone of the protein can be defined by two dihedral angles Φ and Ψ (Figure 2. 9), which are greatly restricted due to steric repulsion by the side chains of the amino acids. Glycine can adopt a larger range of Φ and Ψ values as it lacks a side chain. The Ramachandran diagram (Ramachandran et al., 1963), a Φ Ψ scatter plot, defines areas of favourable Φ and Ψ values and is an important indicator of the quality of the protein model. Residues outside the favourable regions should be closely examined and their geometry justified by the presence of strong electron density (Rhodes, 2000, Kleywegt, 2000). Moreover, further aspects of the stereochemistry and geometry of the model should also be examined. The bond lengths and angles should be compared to a library of ‘ideal’ values, and any deviations should be noted.

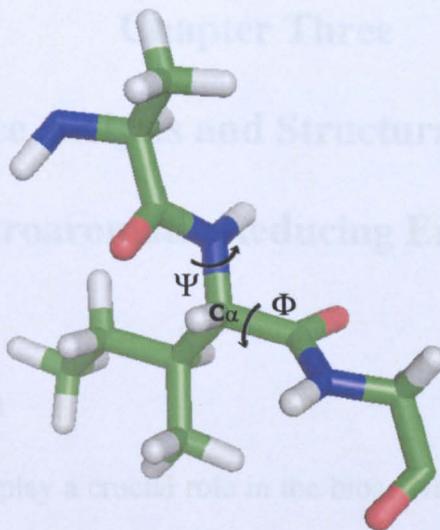


Figure 2.9 Backbone of protein, showing Φ and Ψ angles.

Hydrophobic, electrostatic and hydrogen bonding interactions are the main stabilizing forces in the protein structures. Consequently, hydrophobic residues tend to interact with each other, charged residues tend to form salt bridges and hydrophilic residues prefer to either interact with one another or with the bulk of the solvent. Violations of these rules, such as the presence of a charged residue in a hydrophobic binding pocket, tend to point towards errors in the model. Likewise, the position of added water molecules should be thoroughly checked to establish that they can form hydrogen bonds with neighbouring water molecules or protein atoms (Kleywegt, 2000).

Chapter Three

Sequence analysis and Structural Studies of Nitroaromatic Reducing Enzymes

3.1 Introduction

Reducing enzymes play a crucial role in the bioactivation of chemotherapeutic agents such as mitosenes, diaziridinylbenzoquinones and CB1954, in addition to reducing nitroaromatic and nitro heterocyclic aromatic compounds. The reduction of nitroaromatic and nitro heterocyclic aromatic compounds produces the hydroxylamine derivative which then reacts with thioesters such as acetyl CoA (Skelly et al., 2001). The thioester derivative can attack bio-molecules such as, proteins and DNA. In the case of diaziridinylbenzoquinones, the quinone moiety is reduced to the dihydroxyl benzene derivative hence activating the aziridine group, making it susceptible to nucleophilic attack by the N7 of the guanine base of the DNA (Mossoba et al., 1985b, Rauth et al., 1998, Skibo and Xing, 1998). CB1954 has the unique ability to act as a difunctional-alkylating agent, due to the reduction of the nitro group, in addition to the activation of the aziridine group forming interstrand DNA cross links (Figure 1. 16, 1. 17 and 1.19) (Fu, 1990, Knox et al., 2000, Roldan et al., 2008, Skelly et al., 2001, Skibo and Xing, 1998).

Bacterial nitroreductases are able to reduce nitro compounds to the hydroxyl amine derivative. As previously mentioned in chapter one, they are divided into two groups, oxygen insensitive and oxygen sensitive. Oxygen insensitive

(type I) bacterial nitroreductases are further divided into two groups, group A and group B, according to the percentage sequence identity they share and their phylogenetic origin. Members of both groups share several structural characteristics, where they are dimeric globular proteins with a characteristic $\alpha+\beta$ fold and a conserved binding pocket (Roldan et al., 2008). In addition to bacterial nitroreductases, some mammalian enzymes such as NAD(P)H:quinone oxidoreductase (NQO1) (Boland et al., 1991) and N-ribosyl dihydronicotinamide (NRH): quinone oxidoreductase II (NQO2) (Bianchet et al., 2004) are also able to reduce nitroaromatic compounds to their hydroxylamine derivatives.

Several previously published studies on bacterial nitroreductases showed the conservation of aromatic residues in the binding pocket in bacterial nitroreductases. It is suggested that the aromatic residues present in the binding pocket, play an important role in the interactions of the protein with the substrate, the cosubstrate, and the flavin cofactor (Koike et al., 1998, Parkinson et al., 2000). It was also suggested that phenylalanine residues present in the binding pocket may play an active role in either the transfer of electrons between flavin cofactor and substrate or by stabilizing the intermediate enol form of the cofactor (Koike et al., 1998). All previously published studies on the presence of conserved residues specifically considered bacterial nitroreductases, and did not include any other aromatic nitro group reducing enzymes.

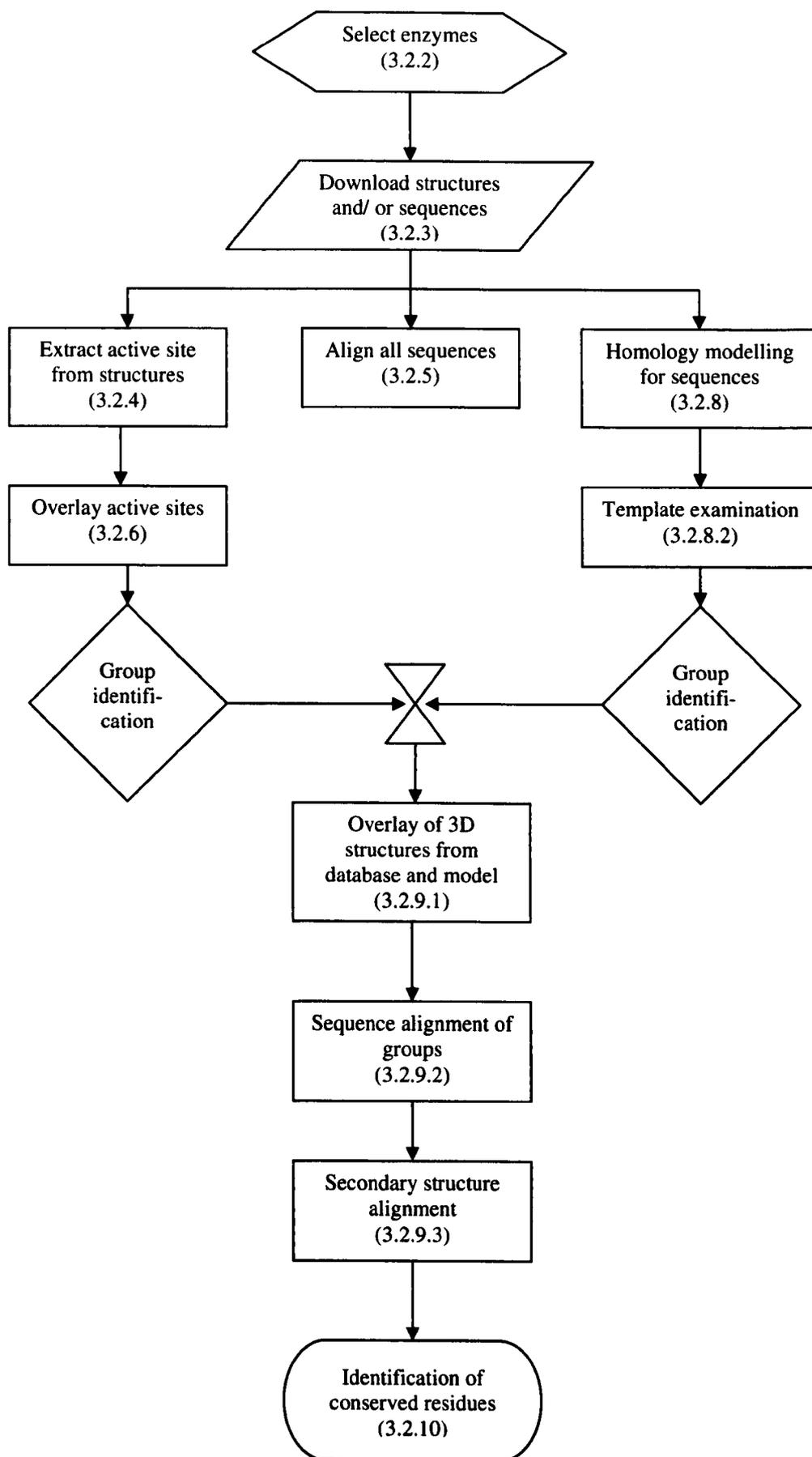
The aim of this work is to use sequence and structural alignments to determine the occurrence of similarities in the active sites of enzymes that are capable of

reducing nitroaromatic compounds to the hydroxylamine derivative as an intermediate or final product. Several criteria were placed for the enzymes used in this study, in addition to the ability of the enzyme to reduce nitroaromatic compounds to their hydroxylamine derivative, the mechanism of action of the enzyme was also considered. Only flavoproteins, utilizing flavin cofactors and are capable of two electron transfers were considered in this study.

The determination of the conservation of residues was done by examining the primary, secondary and tertiary structures to be compared using sequence and 3D structural alignments, in order to detect the conservation of residues in the binding pocket, specifically aromatic residues. Sequence and structural alignments are very useful tools in aiding the identification of conserved regions and the detection of any specific residue patterns in enzymes. Sequence alignment may aid in the identification of the function of proteins, by the identification of closely related proteins. In addition, the comparison of areas of local similarities in protein 3D structures has been used to identify functional sites, and assigning probable function for newly discovered proteins (Jones and Thornton, 2004, Watson et al., 2005).

3.2 Methods and Materials

3.2.1 Summary of Methods Used



3.2.2 Identification of Enzymes

A review of published scientific literature was performed using the ISI web of knowledge to identify enzymes that reduce nitro aromatic compounds to their hydroxylamine derivatives, using the criteria previously mentioned. Glycerol trinitrate reductase (GTN) and 2-cyclohexen-1-one are members of the old yellow enzyme (OYE) family, from which several members are able to reduce nitro aromatic compounds to the hydroxylamine derivative. Therefore they were also included with the selected enzymes due to the similar mechanism of action.

3.2.3 Strategy for Obtaining Sequences and 3D Structural Data

To obtain sequence and 3D structural data for the enzymes of choice, the following strategy was adopted. The RSCB protein databank was searched for 3D structures of the identified proteins. When found they were downloaded with the associated sequence in FASTA format. The first choice was always to download the wild-type protein of the same species identified in the relevant literature and with no substrate in the binding pocket. If no 3D structure was available, the European Bioinformatics Institute website was then used to search the databases and download the appropriate sequences in FASTA format. As with the structure search, the first choice was to download the sequence associated with the species previously identified in the literature.

3.2.4 Extraction of Active Site Residues from 3D Structures

In order to compare the relative position of the amino acid residues in the active site of the different proteins, it was necessary to identify and extract the

residues forming the active site, and separate them from the remaining residues. The active site of the proteins, for which a 3D structure was available, was defined using PyMOL (Delano, 2008). An object containing all the residues within 8 Å of the FAD or FMN cofactor was created; all water molecules, and substrate if present, were then deleted. The resultant structure was then saved as a pdb file.

3.2.5 Sequence Alignment

To identify the presence of conserved regions in the sequences of the 26 chosen proteins, sequence alignment were carried out using, ClustalW2 (Larkin et al., 2007) and T-Coffee (Notredame et al., 2000). Both were accessed through the European Bioinformatics website. The BLOSUM 62 matrix scoring function was used for all alignments. It is a global alignment algorithm, which attempts to align every residue in the sequence. It is based on the likelihood of substitution for a residue by another residue. If the residues are identical, a high score is given, while if the residue is substituted by a similar residue, then a lower positive score is given. If the substituted residues are significantly different, then a negative score is given to the alignment. The scores are then added together to give the overall alignment score between the two proteins (Altschul, 1991). For T-Coffee alignments, the default settings were used, while for Clustal W2, several different parameters were tested, as defined in section 3.2.5.1.

3.2.5.1 Alignment parameters used in alignments with Clustal W2

Clustal W2 has different parameter combinations available, depending on how the penalties are dealt with in the alignment algorithm, thus allowing the optimization of the alignment results. Several different combinations of these parameters were used to try to align the sequences of the all 26 identified proteins, as defined in Table 3. 1

Alignment	GAPOPEN	GAPEXTN	GAPDIST
Alignment 1 (Default parameters)	10	0.2	4
Alignment 2	25	0.2	4
Alignment 3	25	1	6
Alignment 4	5	0.05	2

Table 3. 1 Different parameter combinations used for Clustal W2 alignments where GAPOPEN is the penalty added to the alignment score when a new gap is added to the alignment, while GAPEXTN is the penalty added to the alignment score for each residue added to a pre-existing gap. The GAPDIST is the penalty added to avoid gaps being closely spaced to each other.

3.2.6 The 3D Structural Alignment of the Active Site Residues

Sybyl 8.0 (International) was used to superimpose the active site structures of the chosen proteins, previously created using PyMOL and to compare the positions and types of amino acids in the different active sites. The flavin cofactors in the active site models were overlaid, by selecting the same atoms in the cofactor molecules in the active sites.

The flavin cofactors in the active sites were overlaid by selecting N10, N5, N3 and C8 atoms in the isoalloxazine ring of the cofactor (Figure 1. 1). The structures were then aligned using Sybyl, by the “fit atoms” command in the

“align compounds” menu. On occasion, where the equivalent atoms in the different cofactor molecules did not superimpose, the structural alignment was then repeated using additional atoms from both the ring and tail portion of the cofactor.

From the visual inspection of the overlaid 3D structures of the active sites, it was established that the proteins fall into several different groups showing clear similarities in the residue positioning around the flavin cofactor. The proteins, to which no visual similarities were identified, were reconsidered at a later stage.

3.2.7 Sequence Alignment for Enzymes within Different Structural Groups

For the groups previously identified in section 3.2.6, two sequence alignments were done on the proteins within each group separately using the default settings of T-Coffee and Clustal W2.

3.2.7.1 Calculation of percentage identity and similarity between sequences of group members

For each group, both the identity and similarity between the group members was determined using the Emboss Pairwise Alignment Algorithm (Rice et al., 2000), using the BLOSUM 62 matrix. This was done to ensure the removal of any enzymes that are identical or share more than 75 per cent similarity where large regions of structural similarities would be expected. Enzymes that showed less than 25 per cent similarity to other group members were also excluded, as alignments between sequences with low percentage identities are of low quality (Pei, 2008).

3.2.8 Homology Modelling

3D homology models were created for the sequences for which there were no 3D structures available in the RSCB protein databank; this was to aid in the identification of the structural group to which the enzyme belonged to. Homology modelling is dependent on finding a suitable template protein, with a 3D structure, that is homologous to the required target protein. The structure of the required model is then predicted based upon its similarity with the template protein.

This was done using two separate software packages; SWISS-MODEL (Arnold et al., 2006) and ModWeb (Eswar et al., 2003). SWISS-MODEL was accessed via the ExPASy Proteomics web server (Basel); where model building was done in the fully automated mode. ModWeb was also accessed via its web server (Eswar Narayanan), the models created were scored using the normalised DOPE scoring method (Shen and Sali, 2006) and the best scoring model was selected for output. In addition to the final 3D model, both software packages gave information about the template which was used to create the 3D model.

3.2.8.1 Choice of model

Each of the software packages used produced one model per protein; these were compared, to decide which model would be used for future studies and to determine whether the models produced for the same protein have significant structural differences. The two models for the same protein were aligned using rigid body alignment, based on the sequence identity. This was done using Sybyl 8.0, using the “align structures by homology” command, utilizing only

the backbone atoms of the protein for the alignment. This is done by firstly aligning the sequences of the proteins, and then the backbone atoms of the matching residues are then superimposed. The root mean square deviation (RMSD) is then calculated using the least square fit method. The resulting overlaid structures were then visually examined for differences between the structures.

3.2.8.2 Template examination and group identification

The homology models produced by the modelling software did not contain the cofactor molecule in the structure. Thus extracting the active site residues for structural alignment and subsequent group identification was not possible using the previously used method.

On occasion, some of the template proteins, which were used to build the 3D model, were other proteins used in this study and were identified as members of an identified group. This information is then used to identify the potential group to which the model may belong to. For protein sequences to which no group was identified, the models built were then aligned using Sybyl 8.0, as previously described in section 3.2.8.1, to all the groups to determine the group they belong to.

3.2.8.3 Percentage identity and similarity calculation for modelled proteins

Once the group to which the modelled protein belongs was identified, the percentage identity and similarity of the speculated group members and the

other group members was determined in the same manner described in section 3.2.7.1

3.2.8.4 Validation of group identification

The pdb structures from the RSCB protein databank and those produced by the modelling software were overlaid using Sybyl 8.0 as described in section 3.2.8.1, with the other members of the group identified using the template information. This was done to validate the information gained from the templates and to ascertain that all the group members have similar structural characteristics.

3.2.9 Structural and Sequence Alignment for Proteins Groups

3.2.9.1 Structural alignment

The binding pockets of the previously overlaid proteins in section 3.2.8.4 were visually examined to check that the cofactor molecules and surrounding residues were aligned correctly and identify the conservation of residues.

3.2.9.2 Sequence alignment

After the identification of all group members, two sequence alignments using the default parameters of T-Coffee and Clustal W2 were done using the sequences of all the group members.

3.2.9.3 Secondary sequence alignment

Another sequence alignment was also carried out using PROMALS (Pei and Grishin, 2007), where the software predicts secondary structures of sequences

and uses this information to aid in the sequence alignments. The PROMALS software was accessed via its web server (Pei), where the default parameters were used.

3.2.10 Identification of Conserved Residues

The identification of the conserved residues within the group members was done in two steps. First, the active sites of the overlaid structures from section 3.2.9.1 were visually examined and conserved residues in the area surrounding the flavin cofactor were identified. Afterwards, the sequence and secondary structure alignments from sections 3.2.9.2 and 3.2.9.3 were then examined, to determine whether the identified residues are also conserved in the alignments.

3.3 Results

3.3.1 Obtaining the Sequences and 3D Structural Data

Table 3. 2 and Table 3. 3 show a list of the names and accession codes of the 26 proteins identified in the literature search, in addition to the pdb code if present. All the 3D structures were obtained from X-ray crystallographic experiments with resolutions ranging from 1 to 2.4 Å. Furthermore, the table lists the identified species to which the obtained sequence or 3D structure belongs and the activity of the enzyme.

3.3.2 Active Site Residue Extraction from 3D Structures

The extraction of the active site was successful for all the enzymes to which a 3D structure was available with the exception of NfnB. When attempts were made to define the active site of NfnB, it was found that the crystal structure

did not contain any flavin cofactor, although NfnB is reported to utilise a flavin cofactor to reduce nitroaromatic compounds. In addition, it is also reported to share high sequence homology with other known bacterial nitroreductases such as *E.coli* and *E.cloacae* nitroreductases that utilize FMN (Watanabe et al., 1998). NfnB was found to share a high sequence identity with both NEC and NfsB, 92.6 and 93.5 per cent respectively, suggesting structural similarity. Therefore, it can be discarded from subsequent procedures, and represented by the previously mentioned enzymes.

Protein	Accession code	Organism	Activity
Cyclohexan-1-one reductase	Q9Z3T1	<i>Pseudomonas syringae</i>	Old yellow enzyme family (Williams et al., 2004)
GTN	O31246	<i>Rhizobium radiobacter</i>	Old yellow enzyme family (Williams et al., 2004)
NbzA	Q6DLR9	<i>Pseudomonas pseudoalcaligenes</i>	Reduction of nitrobenzene (Nadeau et al., 2000, Somerville et al., 1995)
NitA	Q97L47	<i>Clostridium acetobutylicum</i>	Reduction of nitroaromatic compounds (Kutty and Bennett, 2005)
NitB	Q97DC3	<i>Clostridium acetobutylicum</i>	Reduction of nitroaromatic compounds (Kutty and Bennett, 2005)
PnrA	Q7B4Y3	<i>Pseudomonas putida</i>	Reduction of TNT (Caballero et al., 2005a, Caballero et al., 2005b)
PnrB	Q8KRI2	<i>Pseudomonas putida</i>	Reduction of TNT (Caballero et al., 2005a, Caballero et al., 2005b)
rnr	Q9EXU3	<i>Enterobacter cloacae</i>	Oxygen insensitive bacterial nitroreductases (Koder et al., 2001)
XenB	Q4KH19	<i>Pseudomonas fluorescens</i>	Reduction of TNT (Pak et al., 2000, Ramos et al., 2005, Williams and Bruce, 2002)
YieF	P0AGE6	<i>Escherichia coli</i>	Activation of CB1954 (Barak et al., 2006)
ywrO	Q65EJ8	<i>Bacillus licheniformis</i>	Activation of CB1954 (Anlezark et al., 2002)

Table 3. 2 Final list of enzymes, accession code and associated species.

Protein	Accession code	Pdb code	Organism	Activity
FRase I	P46072	1VFR	<i>Aliivibrio fischeri</i>	Oxygen insensitive bacterial nitroreductase (Zenno et al., 1994)
FRP	Q56691	1BKJ	<i>Vibrio harveyi</i>	Oxygen insensitive bacterial nitroreductases (Lei et al., 1994, Roldan et al., 2008)
FrxA	Q60049	2H0U	<i>Helicobacter pylori</i>	Reduction of metronidazole (Kwon et al., 2000)
Morphinone reductase	Q51990	1GWJ	<i>Pseudomonas putida</i>	Reduction of TNT (Ramos et al., 2005, Williams et al., 2004)
NEC	Q01234	1NEC	<i>Enterobacter cloacae</i>	Reduction of nitrofurazone (Bryant and Deluca, 1991)
NemA	Q63LY5	3GKA	<i>Burkholderia pseudomallei</i>	Reduction of TNT (Ramos et al., 2005)
NfnB	P15888	3HZN	<i>Salmonella typhimurium</i>	Oxygen insensitive bacterial nitroreductases (Watanabe et al., 1998)
NfsA	P17117	1F5V	<i>Escherichia coli</i>	Oxygen insensitive bacterial nitroreductases (Zenno et al., 1996)
NfsB	P38489	1YLR	<i>Escherichia coli</i>	Oxygen insensitive bacterial nitroreductases and reduction of CB1954 (Knox et al., 1992)
NOX	Q60049	1NOX	<i>Thermus thermophilus</i>	Oxygen insensitive bacterial nitroreductases (Park et al., 1992)
NQO1	P16083	1D4A	<i>Homo sapien</i>	Active against CB1954 (Boland et al., 1991)
NQO2	P15559	1QR2	<i>Homo sapien</i>	Active against CB1954 (Bianchet et al., 2004)
OYE	Q02899	1OYA	<i>Saccharomyces pastorianus</i>	Reduction of TNT (Ramos et al., 2005)
PETN	P71278	2ABA	<i>Enterobacter cloacae</i>	Reduction of TNT (Ramos et al., 2005, Williams and Bruce, 2002)
XenA	Q88NF7	2H8X	<i>Pseudomonas putida</i>	Reduction of TNT (Blehert et al., 1999, Pak et al., 2000, Ramos et al., 2005, Williams and Bruce, 2002)

Table 3. 3 Final list of enzymes, accession and pdb codes and associated species

3.3.3 Sequence Alignment for All Enzymes

Several different sequence alignments for all the sequences, listed in Table 3. 2 and Table 3. 3 were done. The default parameters of T-Coffee were used as well as several different parameter combinations for Clustal W2. The alignments did not show any apparent conserved regions between all the enzymes, rather some conservation of residues could be seen in sub-groups of the enzymes.

3.3.4 The 3D Structural Alignments of Active Site Residues

When Sybyl 8.0 was used to overlay all the active sites defined by PyMOL, it was observed that not all the active sites superimpose, and that they do not share the same orientation of residues within the active site, where the position of amino acids is different relative to the flavin cofactor.

Upon closer visual examination of the overlaid active sites, it was noticed that two distinct groups of proteins were identified, where members of the individual groups shared similar orientation of the flavin cofactor and arrangement of amino acids in the active site. Group I consisted of Morphinone reductase, Nema, OYE, PETN and XenA while group II consisted of FrxA, NEC, FRase I, NOX and NfsB. Furthermore, Frp and NfsA seemed share similar orientation of residues around the flavin cofactor, but did not share any similarity with any of the other proteins. Similarly, NQO1 and NQO2 did not belong to any of the identified groups, but showed structural similarities to each other.

3.3.4.1 Group I active site structural alignment

The overlaid active sites of group I showed that the members of this group have two conserved amino acids, a histidine residue and a tyrosine residue (Figure 3. 1). The positions of the different conserved residues in the protein sequences are listed in Table 3. 4. The position of the conserved tyrosine residue in all sequences ranges between 344 and 375 with the exception of XenA, where it is at position 27. The positions of the conserved histidine residues in all the sequences are very close, ranging from 178 to 191. The active sites of the enzymes showed similar structural characteristics. The binding pocket is wide and is easily accessible to the solvent and is on the *si* face** of the flavin.

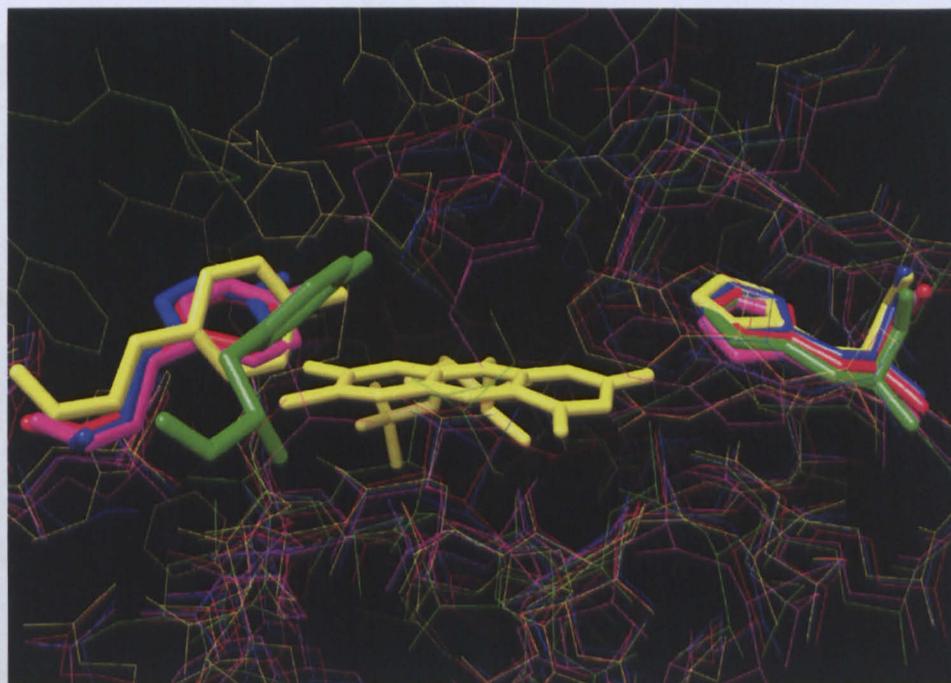


Figure 3. 1 Aligned active sites of members of group I, Morphinone reductase (red), NemaA (blue), OYE (yellow), XenA (Green) and PETN (magenta). The conserved tyrosine residues are represented by bold sticks on the left hand side of the figure while the conserved histidine residues are represented by sticks on the right hand side of the figure. Figure created using USCF Chimera (Pettersen et al., 2004).

** The *si* face of the flavin is viewed when the pyrimidine ring is on the right and the ribotidyl chain is pointing upwards, the *re* face of the flavin is viewed when the pyrimidine ring is on the left.

	Morphinone reductase	NemA	OYE	XenA	PETN
Conserved residue 1	His-186	His-181	His-191	His-178	His-181
Conserved residue 2	Tyr-356	Tyr-344	Tyr-375	Tyr-27	Tyr-351

Table 3. 4 Conserved residues of group I

3.3.4.2 Group II active site structural alignment

The overlaid active sites of members of group II showed one conserved residue, a tyrosine (Figure 3.2). Table 3. 5 shows the position of the conserved residues in the protein sequences, which ranged from position 137 to 144. Unlike group I members, the binding pocket of group II members is on the *re* face of the flavin cofactor. The active site is not buried within the enzyme structure and the flavin cofactor is on the protein surface, at the interface between two identical chains.

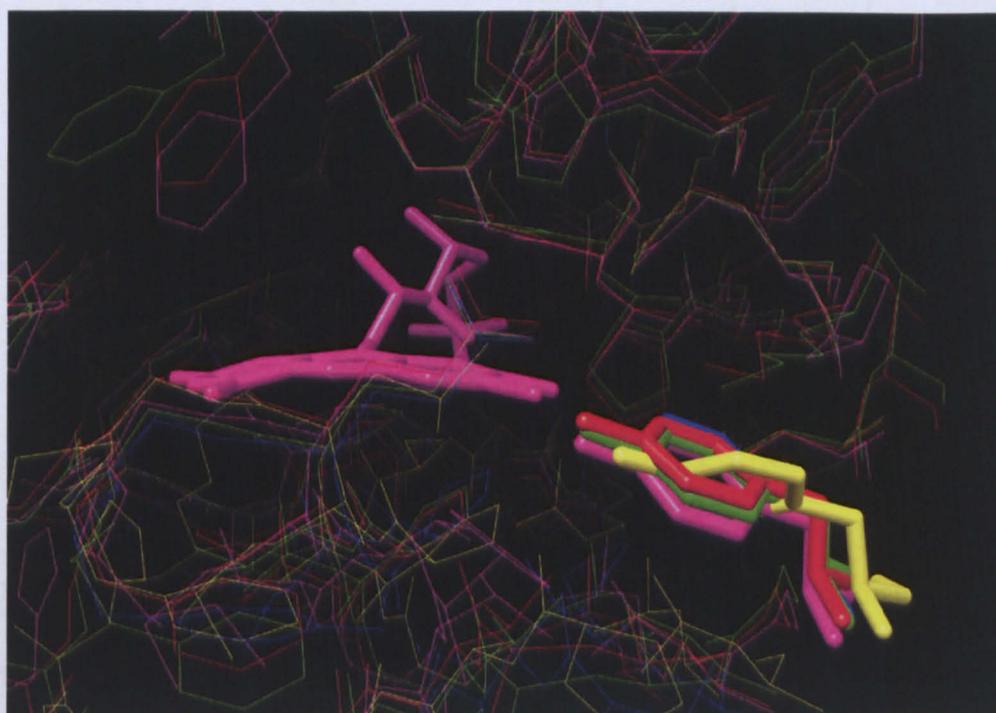


Figure 3. 2 Aligned active sites of group II, NfsB (red), FrxA (blue), NOX (yellow), NEC (green) and FRase I (magenta). The conserved tyrosine residues are represented by bold sticks on the right hand side of the figure.

Figure was created using USCF Chimera (Pettersen et al., 2004).

	NfsB	FrxA	NOX	NEC	FRase I
Conserved residue	Tyr-144	Tyr-143	Tyr-137	Tyr-143	Tyr-144

Table 3. 5 Conserved residues of group II.

3.3.5 Sequence Alignment for Enzymes within Identified Groups

3.3.5.1 Calculation of identity and similarity within identified groups

The percentage identity and similarity between group members were compared in a pair wise manner, to each other. The calculated percentage identities and similarities are shown in Table 3.6 and Table 3. 7 for group I and II members respectively. The members of group I were found to have percentage similarities ranging between 64.0 and 38.2 per cent, which is within the previously decided acceptable range for comparison, Table 3.6.

	NemA	OYE	XenA	PETN
Morphinone reductase	60.4% (49.5%)	51.2% (35.1%)	42.0% (51.2%)	64% (50%)
NemA		46.9 % (33%)	40 % (28.7%)	59.2% (45.5%)
OYE			38.2% (23.6%)	46.4 % (33.6%)
XenA				41.5% (26.0%)

Table 3.6 Percentage similarity between group I members, the percentage identity is shown between parentheses.

On the other hand, two members of group II were found to share a high percentage of similarity, Table 3. 7, where NEC shared 92.6 per cent with NfsB. Due to the high similarity found between the previously mentioned proteins, a decision was taken that only NEC would be used for future comparative purposes, as a representative of both proteins. The remaining

members of the group shared a percentage similarity ranging from 50.2 per cent to 38.5 per cent, which is within the previously decided acceptable range.

	NEC	FRaseI	NfsB	NOX
FrxA	45% (21.8%)	46.2% (25.4%)	45.9% (24.2%)	38.5% (21.4%)
NEC		50.2% (37.4%)	92.2% (87.6%)	41.6% (26.6%)
FRaseI			49.8% (33.5%)	42.1% (27.2%)
NfsB				38.6 % (25.2%)

Table 3. 7 Percentage similarity between group II members, the percentage identity is shown between parentheses.

NfsA and Frp, which were thought to be members of the third group, were found to share 65 per cent identity and 51.7 per cent similarity.

3.3.5.2 Sequence alignment of group I

The alignments of the members of group I using the default settings of both T-Coffee and Clustal W2 showed that the conserved residues in the 3D active site structures were aligned in the sequence alignments, with the exception of Tyr-27 in XenA. The Tyr-27 residue of XenA, was found to be in the same region of the active site with regards to the flavin cofactor as the other conserved tyrosine residues previously mentioned in Table 3. 4 (Figure 3. 1), but was not aligned in the sequence alignments. The conserved histidine and tyrosine residues are highlighted in yellow in the sequence alignments done using both T-Coffee and Clustal W2 (Figure 3. 3 and Figure 3. 4). The full

sequence alignments using Clustal W2 and T-Coffee can be found in appendices A3.1 and A3.2 respectively.

```

OYE      LTKDEIKQYIKEYVQAAKNSIAAGADGVEIHSANGYLLNQFLDPHSNTRTDEYGGSIEN
PETN     LELDEIPGIVNDFRQAVANAREAGFDLVELHSAHGYLELHQLSPSSNQRTDQYGGSVEN
XenA     MTLDDIARVKQDFVDAARRARDAGFEWIELHFAHGYLELHQLSPSSNQRTDQYGGSVEN
NemA     LELDEIPGVVAAFRRGAENARAAGFDGVEVHGANGYLLDQFLQDSANRRTDAYGGSIEN
Morphin  LETDGIPIGIVEDYRQAAQRAKRAAGFDMVEVHAANAACLNPQFLATGTNRRTDQYGGSIEN
-one     :  *  *      :  . . . :  ** :  :* * * . * . . :  : * ** * * * . : *

```

```

OYE      DKRTLIGYGRFFISNPDLVDRLEKGLPLNKYDRDTFYQMSAHGYIDYPTYEEALKLGDW
PETN     GLIDAVAFGRDYIANPDLVARLQKKAELNPQRPESFYGGGAEGYTDYPSL-----
XenA     NQLDLVSVGRAHLADPHWAY-----FAAKELGVEKASWTLPAPY-AHW
NemA     GQADAVAWGKLFIANPDLPRRFKLNAPLNEPNAATFYAQGEVGYTDYPALES-----
Morphin  NTADAVAFGRPFIANPDLPERFRLGAALNEPDPSTFYGGAEVGYTDYPPFLDNHGH---DR
-one     .  * :  . : : * .      :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :

```

Figure 3. 3 Sequence alignment of group I members using T-Coffee

```

OYE      AEQEAkakANNPQHSltkDEIKQYIKEYVQAAKNSIAAGADGVEIHSAN 195
PETN     -ENGNAIRVDtTTPRALELDEIPGIVNDFRQAVANAREAGFDLVELHSAH 184
XenA     IAFG--AHLpkVPREMTLDDIARVKQDFVDAARRARDAGFEWIELHFAH 179
NemA     -----PQRpyVTPRALELDEIPGVVAAFRRGAENARAAGFDGVEVHGAN 184
Morphinone FEDGTAGLHPTSTPRALETdGIPGIVEDYRQAAQRAKRAAGFDMVEVHAAN 189
          :  :  *  *      :  . . . :  ** :  :* * * . * . . :  : * ** * * * . : *

```

```

OYE      GYGRFFISNPDLVDRLEKGLPLNKYDRDTFYQMSAHGYIDYPTYEEALKL 395
PETN     AFGRDYIANPDLVARLQKKAELNPQRPESFYGGGAEGYTDYPSL----- 346
XenA     SVGRAHLADPHWAYFAAKELGVEKAS-----WTLPAPYAHWLE----- 358
NemA     AWGKLFIANPDLPRRFKLNAPLNEPNAATFYAQGEVGYTDYPALESAA-- 361
Morphinone AFGRPFIANPDLPERFRLGAALNEPDPSTFYGGAEVGYTDYPPFLDNHGR 375
          .  * :  . : : * .      :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :

```

Figure 3. 4 Sequence alignment of group I members using Clustal W2.

Red	Small and hydrophobic	*	Residues identical in all aligned sequences
Blue	Acidic	:	Conserved substitution in all aligned sequences
Magenta	Basic	.	Semi-conserved substitution in all aligned sequences
Green	Hydroxyl + amine + basic		

Table 3. 8 Consensus symbols and colours used in Clustal W2 and T-Coffee alignments.

3.3.5.3 Sequence alignment of group II

The sequence alignments using T-Coffee and Clustal W2 of group II members (Figure 3. 5 and Figure 3. 6) showed that the conserved tyrosine residues highlighted in yellow, were aligned in both sequence alignments. The full

sequence alignments using Clustal W2 and T-Coffee can be found in appendices A3.4 and A3.5 respectively.

```

NEC      PEAKAANHKGRTYF--ADMHRVDLKDQWMAKQVYLVNNGNFKLLGKLVGAMGLDAVPIEGF
NOX      AQKQAIQ-R---AF--AAMG---QEARAWASGQSYILLGYLLLLLEAYGLGSPMLGF
FRaseI   EQKEAAF-ASF-K--VELNCDENGEHKAWTKPQAYLALGNALHTLARLNIDSTTMEGI
FrxA     SRFAQII-K---NFQENDMKLNSERSLFDWASKQTYIQMANMMMAAAMLGIDSCPIEGY
          *      :      *      * * : : :      : : : : * *

```

Figure 3. 5 Sequence alignment of group II members using T-Coffee

```

NEC      QVYLVNNGNFKLLGKLVGAMGLDAVPIEGFDAAILDDEEFLKKEKGFSLVVVPV      190
NOX      QS Y ILLGYLLLLLEAYGLGSPMLGFDPERVRAILGLPSR-AAIPALVAL      183
FRaseI   QAYLALGNALHTLARLNIDSTTMEGIDPELLSEIFADELKGYECHVALAI      191
FrxA     QTYIQMANMMMAAAMLGIDSCPIEGYDQEKVEAYLEEKGYLNTAEFGVSV      190
          * * : : :      : : : : * *      :      : : :

```

Figure 3. 6 Sequence alignment of group II members using Clustal W2.

3.3.6 Homology Modelling

SWISS-MODEL and ModWeb modelling servers were used to create 3D structures for the sequences for whom no structures were found in the Brookhaven database. Table 3. 9 shows the templates from which the models were produced by both software, SWISS-MODEL and ModWeb, in addition to the percentage identity the template shares with the modelled protein. A sequence identity above 30 per cent shared between the modelled protein and the template is the generally accepted limit. The low sequence identity decreases the accuracy of the model building process, and produces models with low quality (Elofsson, 2002, Tramontano, 1998). The models that were built for the protein YieF were based on templates that share a sequence identity below 30 per cent with YieF. Hence a decision was made to use these models with caution. The percentage identity shared between the rnr and the template proteins was calculated to be 89 per cent. The remaining sequence identities shared between the templates and the modelled proteins ranged from 69.0 to 31.2 per cent.

Protein	ModWeb template code	SWISS-MODEL template code
2-Cyclohexan-1-one reductase	3GKA (51%)	3GKA (49%)
GTN	3GKA (50%)	2ABA (46%)
Nbza	3GR3 (36%)	3GR3 (35%)
NitB	3E10 (34%)	3E10 (31.2%)
PnrA	3F2V (43%)	2BKJ (35%)
PnrB	3HZN (54%)	1DS7 (51.6%)
nr	3HZN (89%)	1KQB (96%)
XenB	3GKA (69%)	3GKA (69%)
YieF	2Q62 (19%)	2Q62 (20.5%)
ywrO	3FV5 (43%)	No result produced

Table 3. 9 List of templates used to create models and their percentage identity, stated in between parenthesis, to the modelled protein

3.3.6.1 Template examination and group identification

Upon examination of the template pdbs used for building homology models, it was found that some of the template proteins were enzymes that have been selected and have already been assigned to a specific group.

Nema (pdb code: 3GKA) was used as the template protein for building the models of several proteins, 2-Cyclohexan-1-one reductases, XenB and GTN. The model of GTN built by SWISS-MODEL was based on PETN (pdb code: 2ABA). Both Nema and PETN were previously identified to

belong to group I, and thus it is thought that the previously mentioned proteins; 2-Cyclohexan-1-one reductases, XenB and GTN, also belong to the same group and may share the same structural characteristics. Therefore 2-Cyclohexan-1-one reductases, XenB and GTN were allocated to group I.

The models built for rnr and PnrB using WebMod were both based on NfnB (pdb code: 3HZN); this suggests that rnr and PnrB might be structurally related to NfnB. NfnB is reported to be structurally related to members of group II, thus it was thought that rnr and PnrB may be members of group II. Frp (pdb code: 2BKJ) was used by SWISS-MODEL as a template to create the model for PnrA, suggesting that their active sites might align. Hence PnrA may belong to group III. The other templates used for building the remaining models did not provide any insight into the structural information allowing for their allocation to the predetermined group.

3.3.6.2 Comparison of models produced by homology modelling

Alignment of both models for each protein using Sybyl 8.0 showed that both software packages produced similar models, even when different templates were used. Table 3.10 shows the RMSD values between the models produced from both modelling servers.

The models chosen were those produced by the SWISS-MODEL, except for the protein ywrO, where the model produced by WebMod was used since no model was produced for ywrO by SWISS-MODEL.

Enzyme	Cyclohexone reductase	GTN	NbzA	NitB
RMSD (Å)	0.286	0.768	0.406	0.351

Enzyme	PnrA	PnrB	rnr	XenB
RMSD (Å)	1.06	0.538	0.491	0.283

Table 3.10 RMSD values showing the difference between the 2 homology models for each protein (All atoms have been considered in the calculations).

3.3.6.3 Percentage identity and similarity calculation between modelled and previously identified group I members

The new members of group I showed sequence similarity ranging from 69.3 per cent to 40.3 per cent, which is within the previously decided acceptable limit (Table 3.11). The percentage similarity and identity shared between the previously identified group I members is stated in Table 3.6.

	Cyclohexone reductase	GTN	XenB
Morphinone reductase	53.9 % (44.5 %)	58.8 % (45.3 %)	61.9 % (49.3 %)
NemA	62.1 % (48.0 %)	60.8 % (48.2 %)	69.3 % (58.5 %)
OYE	46.9 % (32.3 %)	47.9 % (34.5 %)	48.9% (34.3 %)
PETN	57.3% (43.0 %)	62.0 % (47.0 %)	42.1% (27.2 %)
XenA	44.7% (28.0 %)	40.3 % (26.4 %)	63.2 % (49.5 %)

Table 3.11 Percentage similarities between previously identified and new members of group I, the percentage identity is shown between parentheses.

3.3.6.4 Percentage identity and similarity calculation between modelled and previously identified group II members

The calculated percentage identity and similarity between the modelled proteins and the previously defined group II members are stated in Table 3. 12. The pairwise alignment showed that rnr is almost identical to NEC, sharing 98 per cent similarity, and thus will be excluded from further examination. On the other hand, PnrB shared a percentage similarity within the decided limit with the other members of group II. The percentage similarity and identity shared between the previously identified group II members is stated in Table 3. 7

	rnr	PnrB
FrxA	47.5% (23.8%)	46 % (26.6%)
NEC	98.3% (96.3%)	70% (51.6%)
FRaseI	32.9% (51.1%)	51.4% (34.1%)
NOX	41.2% (26.7%)	45.8% (28.2%)

Table 3. 12 Percentage similarities between previously identified and new members of group II, the percentage identity is shown between parentheses.

3.3.7 Structural and Sequence Alignment for Group I Proteins

3.3.7.1. Structural alignment

The full protein structures of both the experimentally determined structures and the models were aligned using the “homology” in Sybyl 8.0. The alignment of GTN, 2-Cyclohexan-1-one reductase and XenB to the other protein members of group I showed that the two conserved residues the histidine and tyrosine of were also conserved in the newly added protein models (Figure 3.7).

Table 3.13 shows the position of the conserved residues in the sequences of the newly added proteins. The position of the conserved residues in the previously identified members of group I is stated in Table 3. 4.

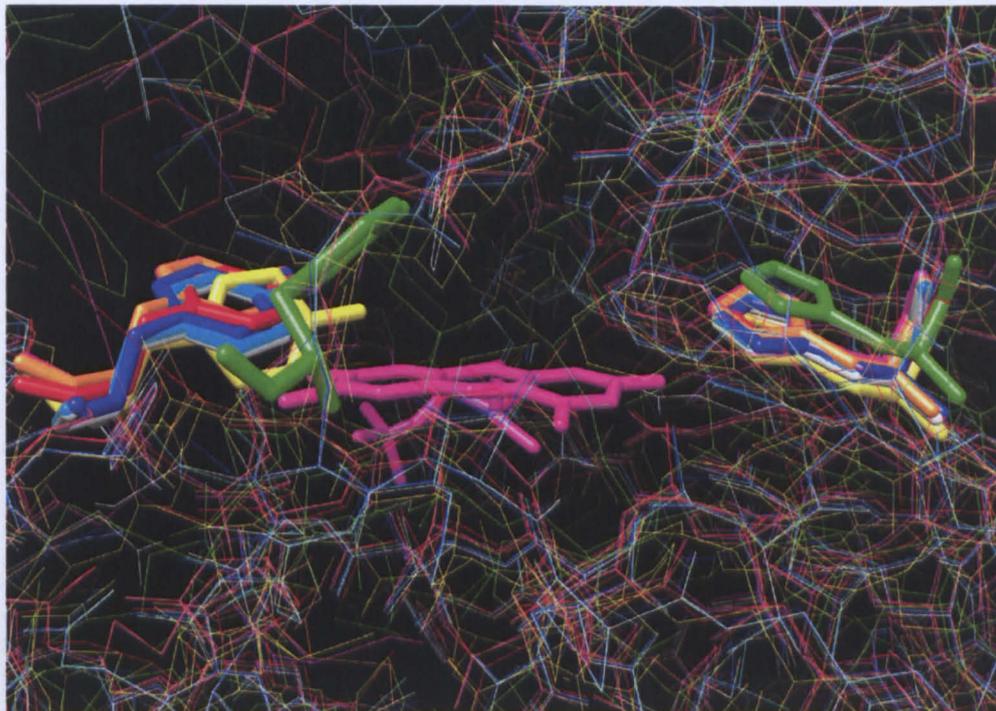


Figure 3.7 Active site of group I members, morphinone reductase (red), XenA (green), NemaA (blue), OYE (yellow), PETN (magenta), GTN (orange), XenB (cyan) and 2-Cyclohexan-1-one reductase (white). The conserved tyrosine residues are represented by bold sticks on the left hand side of the figure while the conserved histidine residues are represented by sticks on the right hand side of the figure.

Figure was created using UCSF Chimera (Pettersen et al., 2004).

Protein	2-Cyclohexan-1-one reductase	GTN	XenB
Conserved residue 1	His-173	His-178	His-173
Conserved residue 2	Tyr-343	Tyr-356	Tyr-343

Table 3.13 Conserved residues in modelled group I proteins.

3.3.7.2 Sequence alignment

The sequence alignments of all group I members showed that the residues conserved in the structural alignment, are also conserved in the sequence alignments, with the exception of Tyr-27 in XenA, as has been previously noted. The Tyr-27 residue in XenA is aligned with the other residues

structurally, but not in the sequence alignments. Figure 3. 8 and Figure 3. 9 show the sequence alignments using both T-Coffee and Clustal W2 respectively, the conserved tyrosine and histidine residues are highlighted in yellow. Full sequence alignments using Clustal W2 and T-Coffee can be found in appendices A3.1 and A3.2 respectively.

```

OYE      TKDEIKQYIKEYVQAAKNSIAAGADGVEIHSANGYLLNQFLDPHSNTRTDEYGGSI
PETN     ELDEIPGIVNDFRQAVANAREAGFDLVELHSAHGILLHQFLSPSSNQRTDQYGGSV
XenA     TLDDIARVKQDFVDAARRARDAGFEWIELHFAHGYLQGSFFSEHSNKRTDAYGGSF
NemA     ELDEIPGVVAAFRRGAENARAAGFDGVEVHGANGYLLDQFLQDSANRRRTDAYGGSI
GTN      TIDDIGLILEDYRSGARAALAEAGFDGVEIHAANGYLIEQFLKSSSTNQRTDDYGGSI
XenB     ETAEIADIVDAYRVGAENAKAAGFDGVEIHGANGYLLDQFLQSSSTNQRTDQYGGSL
Cyclo.   SADEIPRLLNDYEHAAKNAMAAGFDGVQIHAANGYLIDQFLRDNNSNRVDAYGGSI
Morphin- ETDGIPGIVEDYRQAAQRAKRAGFDMVEVHAANA CLPNQFLATGTNRRTDQYGGSI
one      *      :      .      :      **      :      :::*      *      :      *      :      *      *      *      *      *      *

OYE      REEVKDKRTLIGYGRFFISNPDLVDRLEKGLPLNKYDRDTFYQMSAHGYIDYPTYE
PETN     LIG-KGLIDAVAFGRDYIANPDLVARLQKKAELNPQRPESEFYGGGAEGYTDYPS--
XenA     ALQ-ANQLDLVSVGRAHLADPHWAYFAAKELGVEKAS---WTLPPAP--YA---H--
NemA     ALD-AGQADAVAWGKLFIANPDLPRRFKLNAPLNEPNAATFYAQGEVGYTDYPAL
GTN      AVE-SGKVDAVAFGKAFIANPDLVRRLLKNDAPLNAPNQPTFYGGGAEGYTDYPAL-
XenB     WLA-AGKADAVAFGVPFIANPDLPARLKADAPLNEAHPETFYGKGPVGYIDYPT--
Cyclo    ALA-TGEADAITFGRPFLANPDLPHRFAERLPLNKDVMETWYSQGPYVDYPTAD
Morphin- RLD-DNTADAVAFGRPFIANPDLPERFRLGAALNEPDPSTFYGGAEVGYTDYPFLD
one      :      *      .      :::*      :      :      :      :      *

```

Figure 3. 8 Sequence alignment of group I using T-Coffee

```

OYE      -----KAKKANNPQHSLTKEDEIKQYIKEYVQAAKNSIAAGADGVEIHSAN      195
PETN     -ENGNAIRVDTTTPRALELDEIPGIVNDFRQAVANAREAGFDLVELHSAH      184
XenA     ---IAFGAHLPKVPREMTLDDIARVKQDFVDAARRARDAGFEWIELHFAH      179
NemA     -----PQRPYVTPRALELDEIPGVVAAFRRGAENARAAGFDGVEVHGAN      184
GTN      -DDGTGAFABTSEPRALTIDDIGLILEDYRSGARAALAEAGFDGVEIHAAN      181
XenB     -----PLADFPPTRALETAEIADIVDAYRVGAENAKAAGFDGVEIHGAN      176
Cyclo.   -----GKQDYDEARPLSADEIPRLLNDYEHAAKNAMAAGFDGVQIHAAN      176
Morphinone FEDGTAGLHPTSTPRALETGDIPIGIVEDYRQAAQRAKRAGFDMVEVHAAN      189
          :      :      *      :      .      :      **      :      :::*      *      :

```

```

OYE      KRTLIGYGRFFISNPDLVDRLEKGLPLNKYDRDTFYQMSAHGYIDYPTYE      390
PETN     LIDAVAFGRDYIANPDLVARLQKKAELNPQRPESEFYGGGAEGYTDYPSL-      346
XenA     QLDLVSVGRAHLADPHWAYFAAKELGVEKAS---WTLPPAPYAHWLE--      358
NemA     QADAVAWGKLFIANPDLPRRFKLNAPLNEPNAATFYAQGEVGYTDYPAL      358
GTN      KVDVAVAFGKAFIANPDLVRRLLKNDAPLNAPNQPTFYGGGAEGYTDYPALA      370
XenB     KADAVAFGVPFIANPDLPARLKADAPLNEAHPETFYGKGPVGYIDYPTL-      349
Cyclo    EADAITFGRPFLANPDLPHRFAERLPLNKDVMETWYSQGPYVDYPTAD      358
Morphinone TADAVAFGRPFIANPDLPERFRLGAALNEPDPSTFYGGAEVGYTDYPFLD      370
          :      *      .      :::*      :      :      :      :      *      :

```

Figure 3. 9 Sequence alignment of group I using Clustal W2

3.3.7.3 Secondary Structure prediction and alignment

The secondary structure prediction and sequence alignment using PROMALS showed that the conserved histidine residue in all proteins of group I is at the end of a predicted β -sheet, while the conserved tyrosine residue, was not predicted to be in either an α -helix or β -sheet (Figure 3. 10). The conserved tyrosine and histidine residues are highlighted in yellow. The full secondary structure alignment is in appendix A3.3, showed that the secondary structures of the members of group I are highly conserved where the α -helices and β -sheets are generally aligned in many areas.

Conserv.	57 7 759	565 67 57	9977 77 9	96779	797	
NemA	PQRPYVTPRALE LDEI PGVVA AFRRGAENARA AGFDG VEVH GANGY LLDQ FLQ					193
XenB	PLADFPPTRALETAEIADIVDAYRVGAENAKAAGFDGVEIHGANGYLLDQFLQ					185
Cyclo	GKQDYDEARPLS ADEI P RL LN DYEHAAKNAMA AGFDG VQI HAANGY LIDQ FLR					185
GTN	AFAETSEPRALT IDDIGLILE DYRSGARA ALEAGFDGVEI HAANGY LIEQ FLK					190
Morph	GLHPTSTP RALE TDGI PGI VEDYRQAAQRAK RAGFDM VEVHA ANA CLPNQ FLA					198
PETN	IRVDTTTPRALE LDEI PGI VNDFRQAVANARE AGFD LVELH SAHG YLLHQ FLS					193
OYA	AKKANNPQHSLT KDEIKQYI KEY VQAAKNSIA AAGAD GVEI HSANGY LLNQ FLD					204
XenA	GAHLPKVPREMT LDDI ARVK QDFVDAARRARD AGFE WIELH FAHG YLGQS FFS					188
Consensus:	hhhhhhhhhhhhhhhhhhhh	eeee	hhhh			
Conserv.	95 5677977	75 6 77	667 5 77	999	5	
NemA	G KLF IANPDL PRRF KL NAPLN EPNA AAT FYAQGEVGYTDYPAL ESAA -----					361
XenB	GVPFIANPDL PARL KADAPLN EAHP ETFYGKG PGVGI DYPTL-----					349
Cyclo	G RPFL ANPDL LPHR FAERLPL NKDVM ET WYS QGPEGYVDYPTAD QK -----					360
GTN	G KAF IANPDL VRRL KNDAPLN APNQ TFYGGGAEGYTDYPAL AQ -----					371
Morph	G RPFI ANPDL LPER FRLGAAL NEPD PSTFYGGAEVGYTDYPFL DNGH DRLG---					377
PETN	G RDY IANPDL LVAR LQKKAELNPQR PES FYGGGAEGYTDYPS L -----					364
OYA	G RFFI SNPDL VDR LEKGLPL NKYD RDTFYQMSAHGYIDYPT YEEAL KL GW DKK					400
XenA	G RAHL ADPH WAY FAAKELG VEK ---ASWTL PAPY AHW----LE-----					358
Consensus:	hhhhh	hhhhh	hhh			

Figure 3. 10 Secondary structure alignment of group I using PROMALS. The helical regions are colour coded red in the sequence alignment, while the β -sheets are colour coded blue.

If the predicted helices or sheets are aligned in more than 50 per cent of the sequences, than they are denoted by *h* for the α -helices or *e* for the β -sheets in the last line of the alignment. The magenta coloured sequences are used as reference sequences for closely related sequences present in the alignment, which are shown in black and are immediately below the reference sequence.

The number on top of the sequence alignment is the conservation index, which quantifies the conservation of residues at each aligned position. The conservation index is only shown if its value is above five (Pei and Grishin, 2001).

3.3.8 Structural and Sequence Alignment for Group II Proteins

3.3.8.1 Sequence alignment

Figure 3.11 shows the structural alignment of the members of group II and that the Tyr-144 residue of PnrB is present in the same position as the other conserved tyrosine residues of the other group II members. The 3D structure of PnrB showed the same structural characteristics as the other group members, where the active site is found on the surface of the enzyme and the flavin cofactor is easily accessible to the solvent.

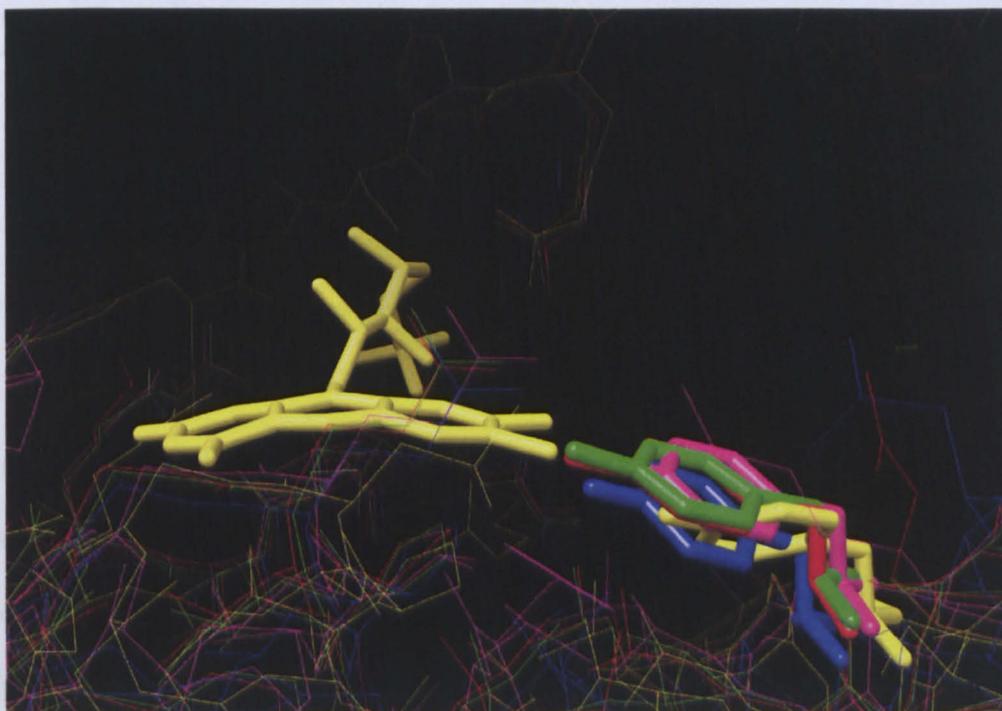


Figure 3.11 Active site of members of group II, NOX (yellow), FRaseI (magenta), NEC (green), FrxA (blue) and PnrB (red). The conserved tyrosine residues are represented by bold sticks on the right hand side of the figure. Figure created using USCF Chimera (Pettersen et al., 2004).

3.3.8.2 Structural alignment

Subsequently, when the sequences of group II members were aligned using both T-Coffee and Clustal W2, the structurally conserved tyrosine residues were also found to be aligned in the sequence alignments (Figure 3.12 and Figure 3.13). The conserved tyrosine residues are highlighted in yellow. The full sequence alignments using Clustal W2 and T-Coffee can be found in appendices A3.4 and A3.5 respectively.

```

NEC      PEAKAANHKGRTYFADMHRVDLKDDDDQWMAKQVYLNVGNFLLGVGAMGLDAVPIEGFDA
NOX      AQQQAIQ--RAFA--AM---GQEARAWASGQSYILLGYLLLLLEAYGLGSPMLGFDP
FRaseI   EQKEAAFASF--KFVELNCDENGEHKAWTKPQAYLALGNALHTLARLNIIDSTTMEGIDP
FrxA     TNSRFAQI IKNFQENDMKLNSERSLFDWASKQTYIQMANMMMAAAMLGIDSCPIEGYDQ
PnrB     EQARAGQNQSRRHVNLHRFDQKDVQHWMEKQTYLALGTALLGAAAHGLDATPIEGFDS
          : . : : : * * * : : : : : : : * *

```

Figure 3.12 Sequence alignment of group II using T-Coffee

```

NEC      WLERVVDQEEADGRFNTPEAKAANHKGRTYFADMHRVDLKDDDDQWMAKQV      142
NOX      HLDEVIHPG-----VQGERREAQQQAIQRAFAAMG---QEARKAWASGQS      136
FRaseI   DYDVVLSKAVADKRITEEQKEAAFASF--FVELNCDENGEHKAWTKPQA      143
FrxA     SDYVKKVMHEVKKRDYDTNSRFAQI IKNFQENDMKLNSERSLFDWASKQT      142
PnrB     HLNAVLVDQEAADGRFRDEQARAGQNQSRRHVNLHRFDQKDVQHWMEKQT      143
          : * *
NEC      YLNVGNFLLGVGAMGLDAVPIEGFDAAILDEEFG---LKEKGFSTLVVV      188
NOX      YILLGYLLLLLEAYGLGSPMLGFDPERVAILG---LPSR-AAIPALV      181
FRaseI   YLALGNALHTLARLNIIDSTTMEGIDPELLSEIFA---DELKGYECHVAL      189
FrxA     YIQMANMMMAAAMLGIDSCPIEGYDQEKVEAYLEEKGYLNTAEFGVSVMA      192
PnrB     YLALGTALLGAAAHGLDATPIEGFDSKVLDAELG---LRERGFSTSVVIL      189
          * : : : : : * * : :

```

Figure 3.13 Sequence alignment of group II using Clustal W2

3.3.8.3 Secondary Structure prediction and alignment:

The conserved tyrosine residue in group II members is aligned in the centre of an α -helix in all the members of group II (Figure 3. 14). The secondary structure alignment revealed that all the members of group II have similar secondary structure, where most of the α -helices or β -sheets were found to be aligned. The full alignment is shown in appendix A3.6

Conserv	9	9	95	6	6	6565	6569	9	5	5	5	
PnrB	DVQHWMEKQTY	LALGTALLGAAA	HGLD	ATPIEGFDSKVLDAELGLR	----	ERGFT						184
NEC	DDDQWMAKQVY	LNVTGNFLLGVG	AMGLD	AVPIEGFDAAILDEEF	GLK----	EKGFT						183
FRaseI	EHKAWTKPQAY	LALGNALHTLAR	LNID	STMEGIDPELLSEI	FADE----	LKGYE						184
FrxA	SLFDWASKQTY	IQMANMMMAAAM	L	GIDSCPIEGYDQEKVEAY	LEEKGYLNTAE	FG						186
NOX	ARKAWASGQSY	IILGYLLLLLEA	YGLG	SVPLGFDP	PERVRAILGL	----	PSRAA					175
Consensus	hhhhhhhhhhhh	hhhhhhhhhh		eeee	hhhhhh		ee					

Figure 3. 14 Secondary structure alignment of group II using PROMALS.
The helical regions are colour coded red in the sequence alignment, while the β -sheets are colour coded blue.

3.3.9 Structural and Sequence Alignment for Group III Proteins

3.3.9.1 Identification of group III members

The structural alignment of Nbza, NitA, NitB, Frp, NfsA and PnrA showed that they shared similar structural characteristics. The binding pocket of the proteins is on the *re* face of the flavin cofactor. The active site is considerably wide and well protected from the solvent. Therefore, they were considered to be members of group III. In addition, an aromatic residue was found to be present in the same position relative to the flavin cofactor (Figure 3.15). The type of residue was not conserved and is varied between histidine (Frp, NfsA and PnrA) tyrosine (NitA) and phenyl alanine (Nbza and Nit B). The position of the aromatic residues in protein sequence is listed in Table 3.14.

	Frp	NfsA	pnrA	Nbza	NitA	NitB
Conserved residue	Tyr-69	His-69	His-68	Phe-122	Tyr-68	Phe-67

Table 3.14 Conserved aromatic residues in group III.

3.3.9.2 Percentage identity and similarity between group III members

The percentage identity and similarity of group III members were calculated in a pair wise manner (Table 3.15). The percentage similarity between the group III members ranged between 65.0 and 25.5 per cent, which is within the previously decided acceptable range.

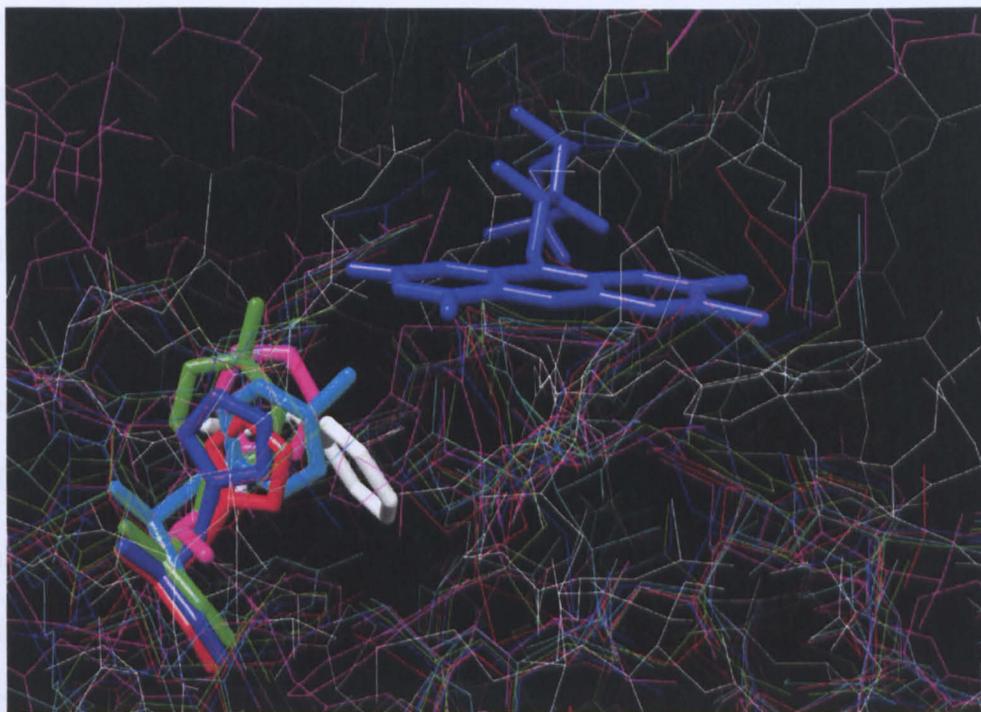


Figure 3.15 Active site of members of group III, PnrA (red), Frp (green), NfsA (blue), Nbza (magenta), NitA (cyan) and NitB (white). The aromatic residues are represented by bold sticks on the left hand side of the figure. Figure was created using USCF Chimera (Pettersen et al., 2004).

	Frp	NfsA	Nbza	NitA	NitB
PnrA	50.9% (32.9%)	50% (30.4%)	25.5% (15.9%)	45.3% (30.1%)	28.8 % (16.3%)
Frp		65% (51.7%)	28.2% (18.2%)	50.2% (32.7%)	34.0 % (19.2%)
NfsA			27.1% (18.0%)	51.2% (34.7%)	31.5% (16.5%)
Nbza				29.3% (15.8%)	35.5% (16.0%)
NitA					34.0 % (20.1%)

Table 3.15 Percentage similarity between group III members, the percentage identity is shown between parentheses.

3.3.9.3 Sequence alignment

When the sequences of group III were aligned using the default settings of T-Coffee and Clustal W2, it was found that the structurally overlapping residues did not align in either of the sequence alignments (Figure 3. 16 and Figure 3. 17). The aromatic residues that are structurally aligned are highlighted in yellow. Moreover, several attempts were made to align the sequences using different parameters in Clustal W2 (Table 3. 1). As with the previous alignments the structurally conserved residues did not align in any of the alignments. The full sequence alignments using the default parameters of Clustal W2 and T-Coffee can be found in appendices A3.7 and A3.8 respectively.

```

NitA -----VGNQEYVAKAPVFLVFMDFYRTYLAGEKTGLKQVI
NitB -----MGSQ-FLKGAPLVIVVEN-----
NfsA -----TGGQKHVAQAAEFWVFCADFNRHLQICP-----DAQ
Frp -----AGNQAYVESAAEFLVFCIDYQRHATINP-----DVQ
Nbza NSFRGQLGDALGIPRSDITLRRRDVERQFRFFDAPVGLIFTMD-----
pnrA -----SGNQRHVEQAPLFLVWLVDWSRLRLLARTLQAPTAG
          * . * . : .

```

Figure 3. 16 Sequence alignment of group III using T-Coffee

```

NitA DEIVKSAQAMPNSINGQQTSVIVVR--DKKKREKLAELVGNQEYVAKAPV 75
NitB DTLLKAALLAPTSMGKRSWEFIAVT--NKNLISELSTARKMGSQFLKGAP 72
NfsA EAIINSARATSSSSFLQCSSIIRIT--DKALREELVTLTGGQKHVAQAAE 75
Frp QTIIQAGLASSSSMLQVVSIVRVT--DSEKRNELAQFAGNQAYVESAAE 75
pnrA SWAIAAAQSASTSSNLQAWSVLAVR--DRERLARLARLSGNQRHVEQAPL 98
Nbza KDILSAAKYAPSSSNTQPWRCYITGEARERITTAAVEAYRAAPEGLPPE 80
          : : . . * : .

```

Figure 3. 17 Sequence alignment of group III using Clustal W2

3.3.9.4 Secondary Structure prediction and alignment

As with the other sequence alignments, the structurally conserved residues did not properly align (Figure 3. 18); however the aligned sequences showed similar predicted secondary structural features. This is demonstrated by the alignment of the α -helices and β -sheets. (Full alignment can be found in appendix A3.9).

Conserv	6	6	6	6	
Nbza	ITVEA-YRAAPEGLPPEY	PFPPQPLH	EPYATRFNSFRGQLGDALGIPRSDITLR		117
NitB	TARKMGSQFLK	GAPLVI	VVVENPE	-----	80
NfsA	TLTGG-QKHVAQAAEFWVFCADFN	RHLQICPDAQLG	-----		94
Frp	QFAGN-QAYVESAAEF	LVFCIDYQRHATINPDVQA	-----D-----		94
pnrA	RLSGN-QRHVEQAPLFLVWLVDWSRLRRLARTLQAP-TAGID	-----			122
NitA	ELVGN-QEYVAKAPVFLVFMDFYR	TYLAGEKTGLK-QVIHE	-----		99
Consensus	hhhhh hhhhhh	eeeeee	hhhhhhhhhhh	hhhhh	

Figure 3. 18 Secondary structure alignment of group III using PROMALS. The helical regions are colour coded red in the sequence alignment, while the β -sheets are colour coded blue.

3.3.10 Structural and Sequence Alignment for Group IV Proteins

3.3.10.1 Structural alignment

When attempts were made to align NQO1, NQO2, YieF and ywrO to the other groups, it was found that they do not align with any of the known groups. On the other hand, the four of them aligned to each other suggesting that they might be all members of a fourth group. Unlike the previous groups, the group IV members shared a lower degree of overall structural similarities. NQO2, YieF and ywrO lacked the tail region present in NQO1. In addition, the 3D structure of ywrO is also lacking another region when compared to NQO1, NQO2 and YieF (Figure 3. 19). On the other hand, the active site of the four enzymes is quite similar.

It was found that there is an aromatic residue present in the same region relative to the flavin cofactor in all four protein structures (Figure 3. 20). The type of residue is not conserved, in NQO1 and NQO2, it is a phenylalanine, while in YieF and ywrO it is a tyrosine. Although the residues show different orientations, this might be due to errors in the modelling of the ywrO and YieF proteins, caused by the low identity shared with the template used for the model building. The position of the aromatic residue in the different protein sequences is listed in Table 3.16.



Figure 3. 19 Overlaid cartoon representations of the structures of group IV members, NQO1 is shown in grey, NQO2 is shown in yellow, ywrO is shown in red and YieF is shown in blue. Figure created using USCF Chimera (Pettersen et al., 2004).

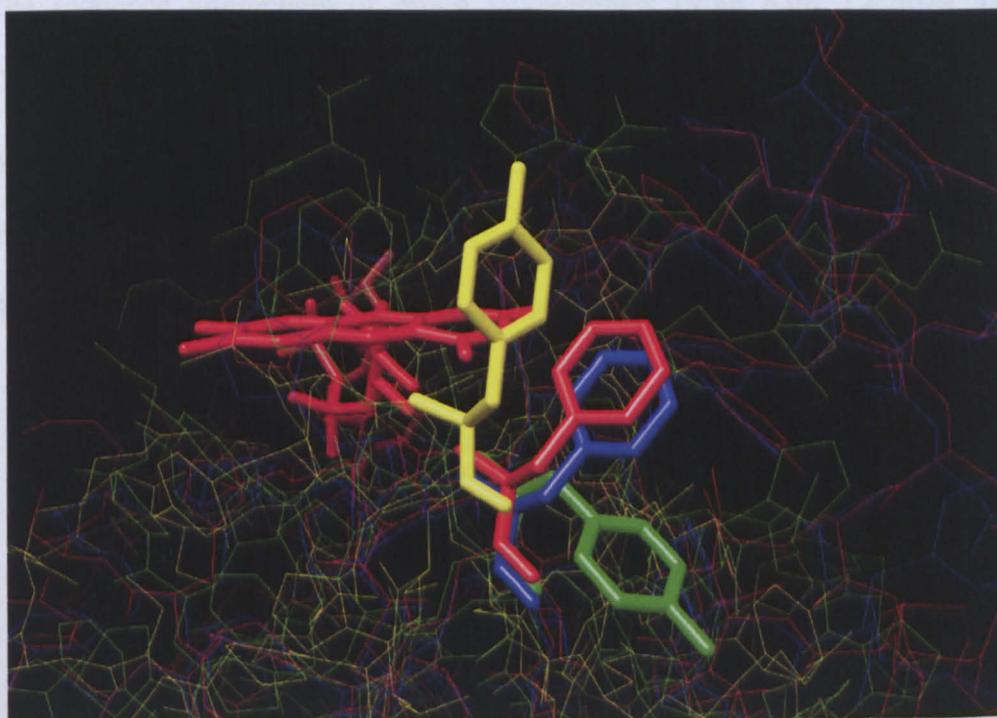


Figure 3. 20 Active site of group IV members, NQO1 (red), NQO2 (blue), ywrO (green) and YieF (magenta). The conserved tyrosine residues are represented by bold sticks. Figure was created using USCF Chimera (Pettersen et al., 2004).

	NQO1	NQO2	YieF	ywrO
Conserved residue	Phe-106	Phe-105	Tyr-81	Tyr-68

Table 3.16 Conserved residues in group IV

The percentage similarity and identity for members of group IV were calculated, and the percentage similarity ranged between 24.6 and 50.4 percent, which are within the acceptable range (Table 3. 17).

	NQO2	YieF	ywrO
NQO1	50.4% (38.5%)	24.6% (13.9%)	31.8% (18.2%)
NQO2		34.5 % (18.4%)	36.7% (22.9%)
YieF			31.7 % (20.6%)

Table 3. 17 Percentage similarity between members of group IV, the percentage identity is shown between parentheses.

3.3.10.2 Sequence alignment:

The sequence alignments of group IV members using both T-Coffee and Clustal W2 showed the alignment of the overlapping phenylalanine and the tyrosine residues (Figure 3. 21 and Figure 3. 22). The aromatic residues of interest are highlighted in yellow. The full sequence alignments using Clustal W2 and T-Coffee can be found in appendices A3.10 and A3.11 respectively.

```

NQO1  TGKLKDPANFQYPAESVLAYKEGHLSPDIVAEQKKLEAADLVIFQFPLQWFGVPAILK
NQO2  TGTLSNPEVFNYGVETHEA-KQRSLASDITDEQKKVREADLVIFQFPLYWFSVPAILK
YieF  -----PLYDADVQQEEGFPATVEALAEQIRQADGVVIVTPEYNYSVPGGLK
ywrO  -----EYP-----DEAIDVEKEQKLCEHDRIVFQFPLYWYSSPPLLK
      :      :      *  ::  *  ..  *  **

```

Figure 3. 21 Sequence alignment of group IV using T-Coffee

```

NQO1  IFQFPPLQWFGVPAILKGWFERVFIGEFAYTYAAMYDKGPFRSKKAVLSIT      147
NQO2  IFQFPLYWFSVPAILKGWMDRVLCQGFAFDIPGFYDSGLLQGLLALLSVT      149
YieF  VIVTPEYNYSVPGGLKNAID-----WLSRLPD----QPLAGKPVLIQTS      116
ywrO  VFQFPLYWYSSPPLLKKWLDHVLLYGWAYGTNG----TALRGKEFMVAVS      106
      ::  *  :.  *  **  ::  :  :  :  *  ::  :

```

Figure 3. 22 Sequence alignment of group IV using Clustal W2

3.3.10.3 Secondary Structure predication and alignment

As with groups I and II, the structural sequence alignment showed that the secondary structures of the group members showed highly aligned helices and sheets. In addition, the residues of interest, tyrosine and phenylalanine were also conserved (Figure 3. 23). The aromatic residues of interest are highlighted in yellow. (Full alignment is found in appendix A3.12)

Conserv.	555	59	555559	5555559	99	5	5	5	
ywrO	EKEQKLCEE	HDRIVFQ	FPLYWY	SSPPLLKKWLDH	VLLYGWAYGTN	---	GTA		95
NQO2	TDEQKVR	EADLVI	FQFPLYWF	SVPAILK	GWMDRVL	CQGF	AFDIPGFYDSGL		139
NQO1	VAEQKLEA	ADLVI	FQFPLQWF	GVPA	ILKGF	ERVF	IGEFAYTYAAMYDKGP		136
YieF	EALAEQIRQ	ADGVVIV	TPEYNY	SVP	GGLKNAID	WLSRLP	-----DQP		105
consensus	hhhhhhhhh	eeee			hhhhhhhhhhh				

Figure 3. 23 Secondary structure alignment of group IV using PROMALS

3.4 Discussion and Conclusion

Two different techniques were used to study the sequence and structural similarities between the enzymes. First a structural alignment of the active sites was undertaken and the residues surrounding the flavin cofactor were examined. This was followed by sequence alignments using Clustal W2 and T-Coffee, as well as PROMALS which utilizes secondary structure prediction. The sequence alignments using PROMALS showed the same results as the alignments using Clustal W2 and T-Coffee. Therefore increasing the confidence of the results, as the use of secondary structure prediction in protein sequence alignment increases the accuracy of the result especially for distantly related proteins with less than 20 per cent sequence identity (Pei, 2008). In addition to increasing the confidence in the results, PROMALS illustrated the fact that all the members of each group share similar secondary structures.

The studied enzymes were found to fall into four distinct groups. Group I consisted of morphinone reductase, XenA, XenB, OYE, NemA, PETN, GTN and 2-Cyclohexan-1-one reductase. These enzymes have two conserved aromatic residues in their binding pocket, a histidine residue within an average distance of 3.3 Å from the O2 of the flavin isoalloxazine ring. In addition, a tyrosine residue is also conserved within 4.8 Å of C7 of the isoalloxazine ring of the flavin cofactor. Group I sequence alignments showed the alignment of the histidine residues in all the sequences. On the other hand, the Tyr-27 residue in XenA did not align with the remaining tyrosine residues in any of the sequence alignments. This is thought to be due to their positions in the protein sequence. The Tyrosine residue of XenA is at position 27, while the tyrosine residues of the other group members are at positions ranging from 343 to 375. The large gap in the position of the residues of interest in the sequences makes it impossible for the tyrosine residue of XenA to align with the other residues in sequence alignments.

OYE is one of the most intensively studied flavoenzymes and thus the amino acid residues in the active site involved in substrate binding and mechanism of action have been well characterized. Both His-191 and Tyr-375 residues in OYE are thought to play a role in the orientation of the substrate in the active site, via hydrogen bonding. The mutation of the histidine residue decreases the binding affinity of substituted phenol substrates to OYE. Tyr-196 is thought to be involved in the mechanism of action of OYE, where it provides one of the hydride ions to be transferred to the substrate. This tyrosine residue was found to be conserved in all the examined enzymes with the exception of morphinone

reductase, where it is replaced by a cysteine residue. Gln-114 and Asn-194 are involved in hydrogen bonds with the flavin cofactor and are thought to affect its orientation. While the Gln-114 was found to be conserved in the sequence alignments of group I the Asn-194 was not and is replaced by a histidine residue in some of the enzymes (Figure 3. 24) (Brown et al., 1998, Fox and Karplus, 1994, Brown et al., 2002, Kohli and Massey, 1998).

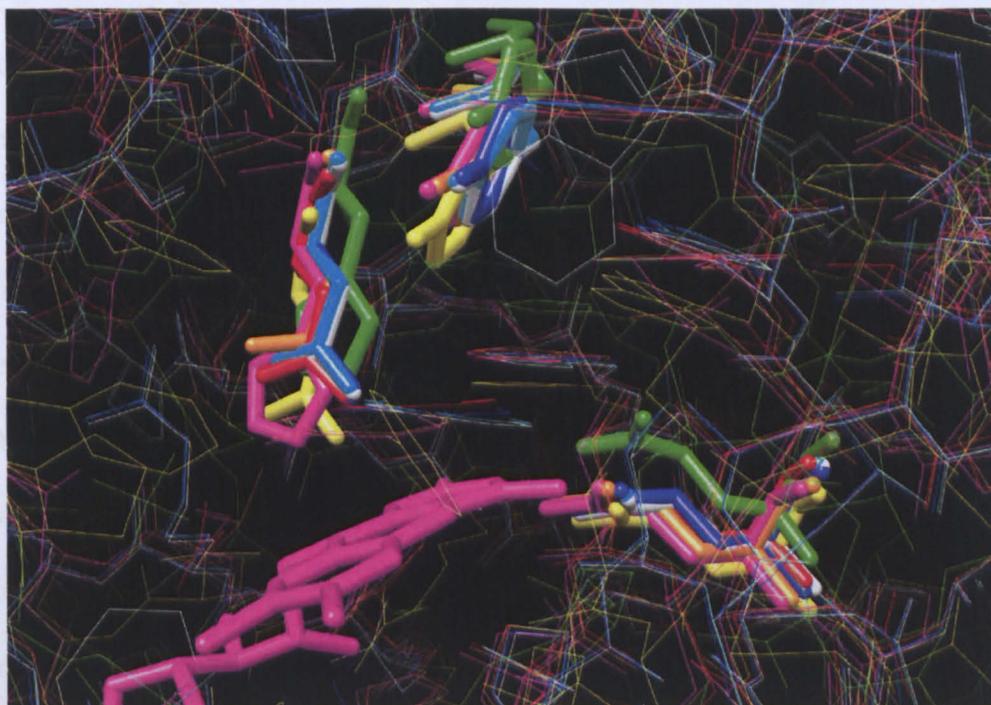


Figure 3. 24 Active site of group I members, morphinone reductase (red), XenA (green), NemaA (blue), OYE (yellow), PETN (magenta) GTN (orange), XenB (cyan) and 2-cyclohexan-1-one reductase (white). The conserved Gln residues are represented as sticks in the bottom left hand side of the figure, while the Tyr 196 and Asn 194 of old yellow enzyme and corresponding aligned residues are represented as sticks on the top of the figure. Figure created using USCF Chimera (Pettersen et al., 2004).

Group II was found to consist of NOX, Nec, FRaseI, FrxA and PnrB which have a tyrosine conserved residues in proximity to C7 of the flavin cofactor. The average distance between the C7 and OH group of tyrosine amino acid is 6.9 Å. Sequence alignments of group II members showed the alignment of all the conserved tyrosine residues. The conserved tyrosine residues in group I and

group II are both in the same region, the hydroxyl groups of the tyrosine residues in both groups are within an average distance 3.2 Å of each other.

Group III members were found to be NfsA, Frp, PnrA, Nbza, Nit A and Nit B. They have different types of aromatic residues, histidine, phenylalanine and tyrosine in the same position, within 3.5 Å of O4 of the cofactor. The sequence alignment of group III members did not result in the alignment of the structurally conserved residues despite the fact that the positions of the residues in the sequences are very close. Group II and III members are classified to belong to group B and A bacterial nitroreductases respectively according to their phylogenetic origin. Members of the same group are known to share high sequence and structural similarities (Koike et al., 1998, Parkinson et al., 2000, Dolores Roldan et al., 2008). The role of the identified conserved amino acids is not yet known, and has not been discussed in the literature. The residues involved in the binding of the flavin cofactor to the bacterial nitroreductases are well studied and they do not include the previously mentioned conserved residues. Thus it is thought that they might play a role in the correct orientation of the substrate or in the mechanism of electron transport.

Group IV members are ywrO, YieF, NQO1 and NQO2, which have either a phenylalanine or tyrosine, conserved 6.4 Å from O2 of the isoalloxazine ring of the flavin. The homology models of YieF produced by both ModWeb and SWISS-MODEL were based on templates sharing a sequence identity of approximately 20 per cent with YieF. The use of low templates sharing a sequence identity less than 30 per cent with the required protein produces

unreliable models (Dalton and Jackson, 2007, Tramontano, 1998). The sequence alignments for group IV showed the conserved residues in all four enzymes perfectly aligned. Consequently the results obtained from the sequence alignment of group IV increased the confidence in the results obtained from the structural alignments. The postulated mechanism of action of NQO1 involves two aromatic residues Tyr-155 and His-161. The tyrosine residue is conserved in all members of the group with the exception of YieF, where it is replaced by an arginine residue, while the histidine residue is not at all conserved between any of the group members (Deller et al., 2008). Residue 104 is known to play an important role in the reduction of nitro groups by NQO1 and NQO2. In human NQO1 this residue is glutamine while in NQO2 and rat NQO1 it is a tyrosine. The presence of the tyrosine residue in that position is thought to increase the efficiency of the enzyme towards the reduction of the CB1954 (Chen et al., 1997, Skelly et al., 2001, Knox and Chen, 2004). This residue is conserved in all group members except for the previously mentioned exception of human NQO1 (Figure 3. 25).

The positions of the conserved residues in the different groups are different with respect to the FAD molecule. Although both the conserved histidine residue of group I members and the conserved aromatic residue of group IV members are within close proximity to O2 of the flavin cofactor, they are on opposite sides of the flavin cofactor. On the other hand, the conserved tyrosine residues of group I and II are in close proximity to C7 of the flavin cofactor. The conserved residues in group III are within close proximity to O4 of the flavin (Figure 3.26).

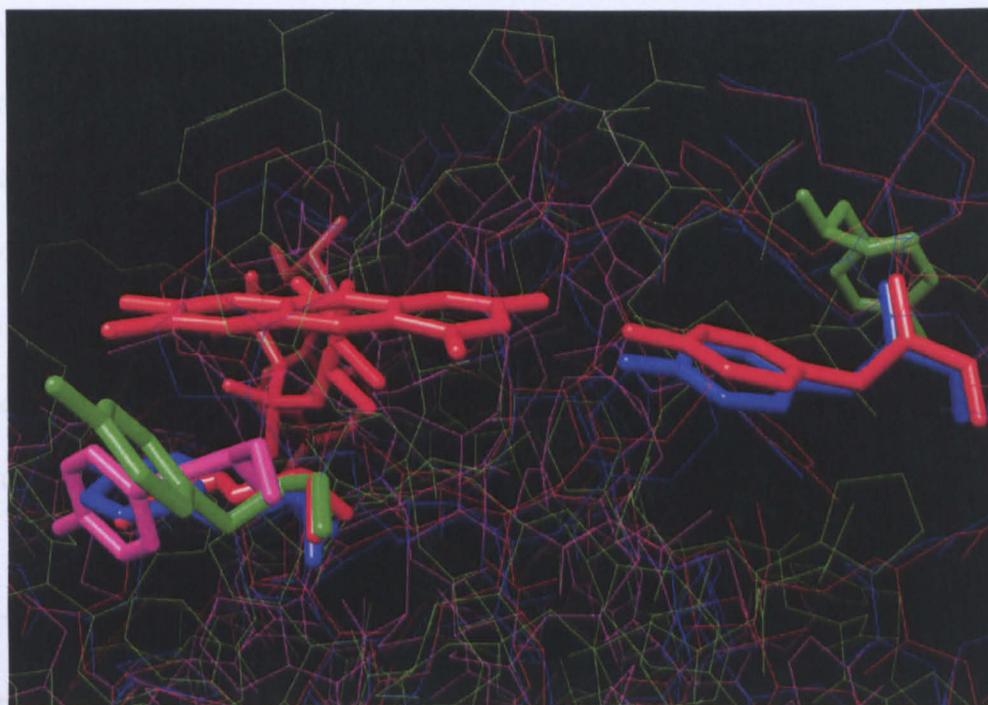


Figure 3. 25 Active site of group IV members, NQO1 (red), NQO2 (blue), ywrO (green) and YieF (magenta). The Glu 104 of NQO1 and aligned tyrosine residues are represented as sticks on the left hand side of the figure. While Tyr 155 of NQO1 and aligned tyrosine r residues are represented as sticks on the right hand side of the figure. Figure created using USCF Chimera (Pettersen et al., 2004).

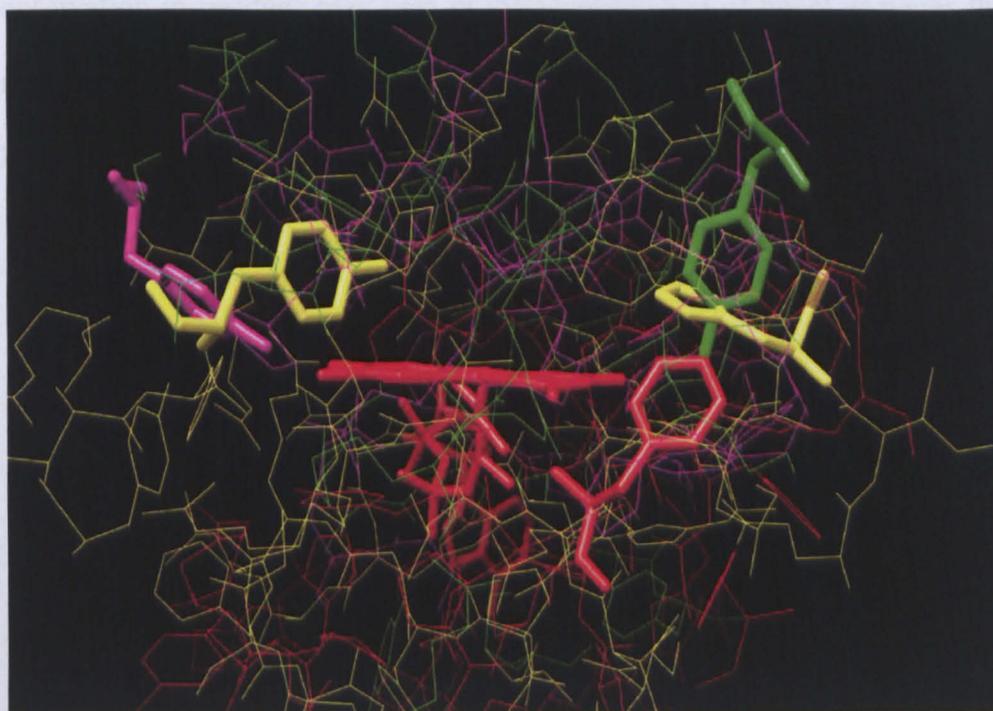


Figure 3. 26 The different orientations of the conserved residues in the aligned active sites around the flavin cofactor, where group I is represented by OYE (yellow), group II is represented by FRaseI (magenta), group III is represented by Frp (green) and group IV is represented by NQO2 (red). Figured created using USCF Chimera (Pettersen et al., 2004).

All groups showed the presence of aromatic residues in the area surrounding the binding pocket suggests that they are of some importance. The aromatic residue may either play a role in the electron transportation or may help in the binding and correct orientation of the substrate or/and cosubstrate in the binding pocket. The difference in the type of conserved aromatic residue may account for the differences in the reduction potential of different enzymes or substrate affinity of the different enzymes (Knox and Chen, 2004).

The identification of trends in the structure and sequences of flavoenzymes may aid in the better understanding of the mechanisms of action by which the anticancer prodrugs such as CB1954 are reduced. This may lead to the designing of more efficient drugs. Moreover, the identification of trends of conservation in residues may help in the identification of novel enzymes that are better able to reduce such drugs which may then be used as part of the GDEPT and ADEPT approaches.

Chapter Four

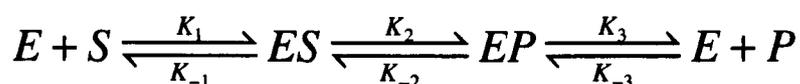
Kinetic Studies for NQO2 I:

Methods and Materials

4.1 Introduction

4.1.1 Enzyme Reactions

In simple enzyme reactions the substrate combines with the enzyme forming an enzyme-substrate complex. An enzyme-product complex is then formed which disassociates to the release of the free product and enzyme. In certain cases, the reaction is more complex and may involve other components such as cosubstrates, cofactors, inhibitors and activators. The rates of reaction are designated by k , which are marked by positive numbers for the forward reactions and negative numbers for the backward reaction (Bisswanger, 2008).



The progress of the reaction can be monitored in two different ways, either by measuring the rate of appearance of the product or the rate of disappearance of the substrate or cosubstrate. A typical progress curve for enzyme catalyzed reactions can be divided into three distinct phases (Figure 4.1):

1. Pre-steady state:

This is the phase with the fastest rate, but is not considered during experiments due to its extremely short duration, (micro to milliseconds)

and very low intensity. This phase can not be detected under normal conditions.

2. Steady state:

During this phase the reaction intermediates are at equilibrium and are thus at constant concentrations. This is a linear region of the curve and therefore the reaction rate can be directly calculated from the slope of the diagram.

3. Substrate depletion:

During this phase the rate of reaction is continuously decreasing due to substrate depletion (Bisswanger, 2004).

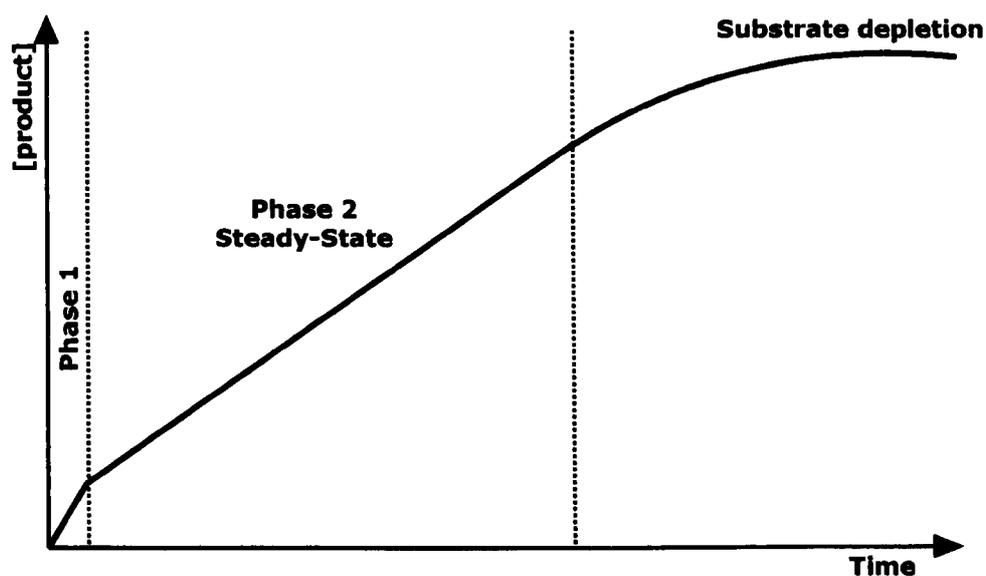


Figure 4.1 Progress curve of an enzyme reaction.
Adapted from (Bisswanger, 2004)

4.1.2 Enzyme Kinetic Constants

Enzyme catalyzed reactions can be described by the Michealis-Menten equation (4.1) within the steady state range of the reaction (Bisswanger, 2004);

$$v = \frac{V_{\max}[S]}{[S] + K_m} \quad (4.1)$$

Where,

v	Reaction rate
V_{\max}	Maximum rate of reaction
$[S]$	Concentration of substrate
K_m	Michealis constant

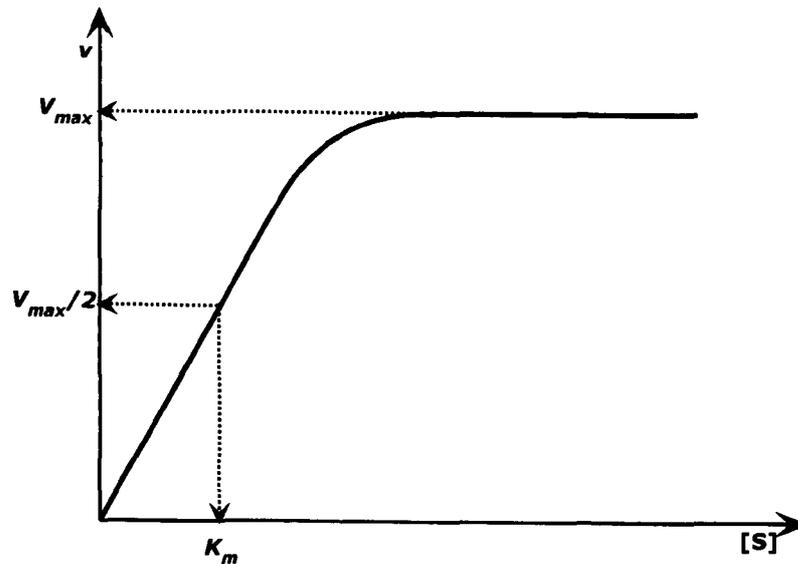
The Michealis constant, K_m , is indirectly proportional to the affinity of the enzyme for the substrate and is specific to the enzyme/substrate combination.

The Lineweaver-Burke equation (4.2) is the double reciprocal of the Michealis-Menten equation and results in a linear relationship between $1/[S]$ and $1/v$. The Lineweaver-Burke plot is prone to errors due to the absence of points at high substrate concentrations resulting in large extrapolations to determine the intercepts.

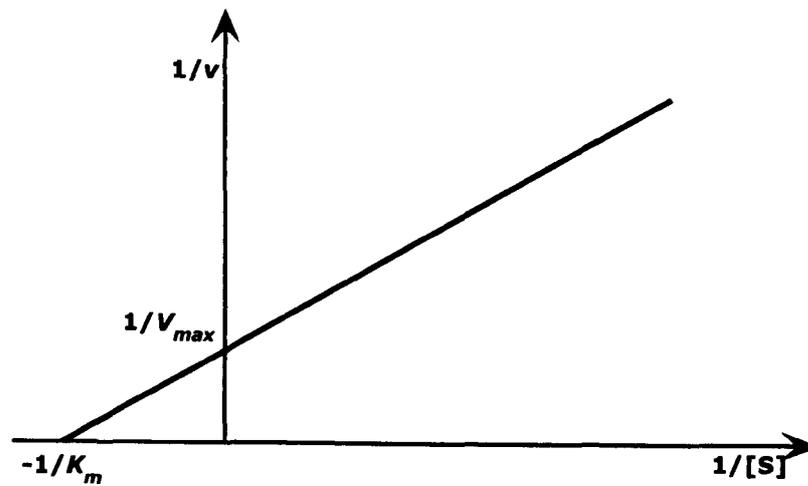
$$\frac{1}{v} = \frac{1}{V_{\max}} + \frac{K_m}{V_{\max}[S]} \quad (4.2)$$

The Hanes-Woolf equation (4.3) is another representation of the Michealis-Menten plot, which overcomes the errors caused by the uneven distribution of the data.

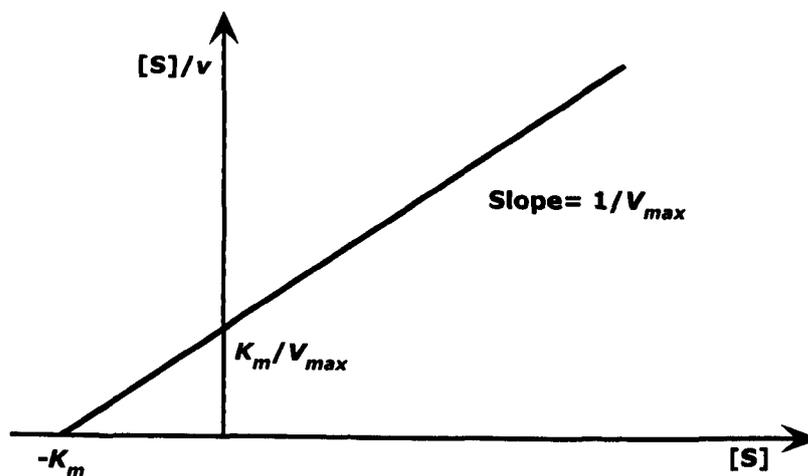
$$\frac{[S]}{v} = \frac{[S]}{V_{\max}} + \frac{K_m}{V_{\max}} \quad (4.3)$$



(a)



(b)



(c)

Figure 4. 2 (a) Classical representation of Michealis-Menten equation (b) Lineweaver-Burke plot, the y-axis intercept is $1/V_{max}$ while the x-axis intercept is $-1/K_m$
 (c) Hanes-Woolf plot, the y-axis intercept is K_m/V_{max} , the x-axis intercept is $-K_m$ while the slope is $1/V_{max}$.

4.1.3 Potential NQO2 Substrates

4.1.3.1 Novel substrate, GC201

GC201 (Figure 1. 19) was designed to be a substrate for NQO2 and was kindly provided by Professor Christopher Moody, School of Chemistry - University of Nottingham. It is a potential NQO2 substrate, where NQO2 is thought to reduce the quinone derivative into the dihydroxyl derivative (Figure 4.3).

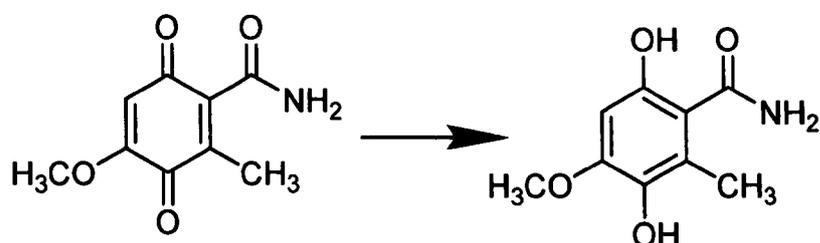


Figure 4.3 Reduction of GC201

4.1.3.2 Coenzyme Q0, Coenzyme Q1 and Coenzyme Q2

Coenzyme Q0 is a natural known substrate for NQO2 and thus it is speculated that other members of the coenzyme Q family (Figure 4. 4) may also be substrates for NQO2 (Boutin et al., 2005).

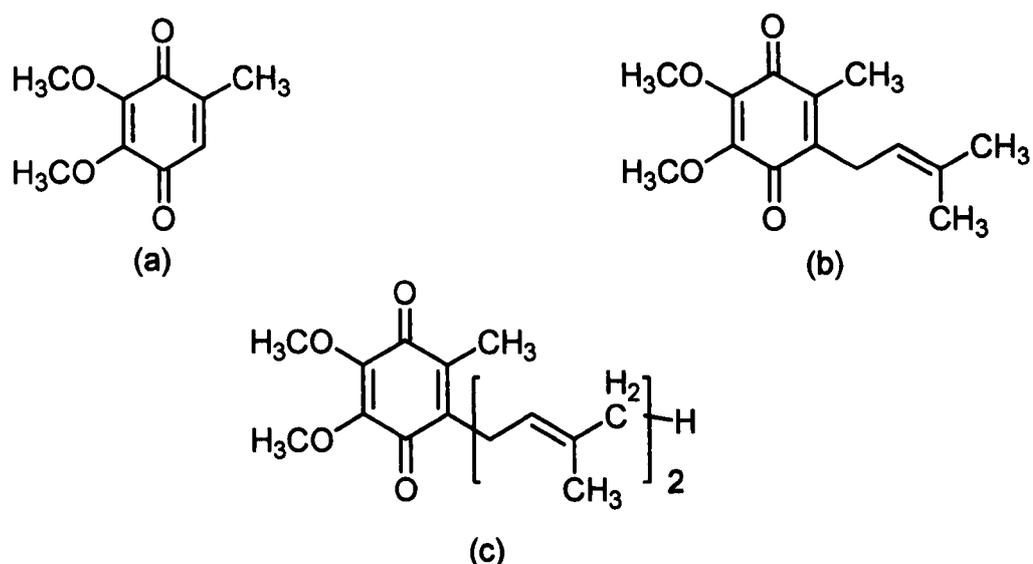


Figure 4. 4 Structures of (a) coenzyme Q0, (b) coenzyme Q1 and (c) coenzyme Q2

4.1.4 Experimental Methods for Determining Michealis-Menten Constants

There are several reported experimental methods based on different techniques used to determine the reaction kinetics constants for NQO2 with different substrates. These techniques included HPLC (Boutin et al., 2005), UV/visible spectroscopy (Calamini et al., 2008, Kwiek et al., 2004, Zhao et al., 1997) and fluorescence spectroscopy (Boutin et al., 2008, Fu et al., 2005, Mailliet et al., 2005, Boutin et al., 2005). The previously reported method based on the measurement of fluorescence was used here to explore the kinetics of reactions between NQO2 and a variety of substrates. This method was chosen over other reported methods as it is more sensitive than methods based on UV/visible spectroscopy. It is also less time consuming than HPLC, and provides the opportunity to monitor the reaction during real time compared to samples taken at regular time intervals from the reaction mixture.

The decrease of the fluorescence of an artificial cosubstrate, N-benzylidihydronicotinamide (NBDH) (Figure 4.5), upon its consumption by NQO2 was monitored during the reaction. For each molecule of NBDH consumed, there is one molecule of substrate consumed, due to the one to one stoichiometry of the reaction.

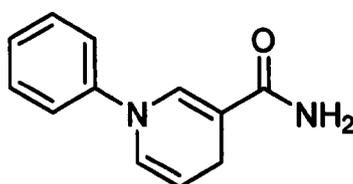


Figure 4.5 Structure of NBDH

Since the fluorescence of a substance is directly proportional to its concentration, equation (4.4) within a linear range. Thus, the decrease of the fluorescence is directly proportional to the decrease in the concentration of the cosubstrate, which is in turn equivalent to the consumption of the substrate (Sharma and Schulman, 1999).

$$I_f = 2.3\Phi_f I_o \epsilon C l \quad (4.4)$$

Where,

I_f	Intensity of emitted light
Φ_f	Quantum yield of fluorescence
I_o	Intensity of incident light
ϵ	Molar absorptivity
C	Molar concentration
l	Pathlength

4.2 Experimental Methods and Materials

All chemicals were purchased from Sigma-Aldrich unless otherwise stated. The fluorescence and UV/visible spectroscopy work was carried out using Perkin Elmer LS-55 fluorescence spectrophotometer. The fluorescence cuvettes used were purchased from Hellma, catalogue number: 114F-QS. The graphs were created using GraFit version 7 (Leatherbarrow, 2009), which was also used for linear regression analysis. The composition of the solutions used can be found in Appendix A2.

4.2.1 Cloning of NQO2

The cloning of the NQO2 gene in TOP10 *E.coli* bacteria was kindly done by Dr. John Heap, Institute of Infection, Immunity and Inflammation - The University of Nottingham. The expression vector used was pMTL1015, which was then transformed into TOP10 *E.coli* bacterial cells. The expressed NQO2 from the clone has a six residue histidine tail, His-Tag.

4.2.2 Expression and Harvesting of NQO2

Before the bacterial cells were grown the working bench surface was sterilized using ethanol. In addition, all glassware and pipette tips used were sterilized by autoclaving.

A few bacterial cells from a glycerol stock were used to inoculate five ml of 2xTY media containing five μl tetracycline solution, (five mg ml^{-1} in ethanol), this was grown overnight at 37°C and 180 rpm. The overnight culture was then used to inoculate 50 ml of 2xYT media containing 50 μl tetracycline solution, (5 mg ml^{-1} in ethanol). This was allowed to grow at 37°C and 180 rpm until the optical density (OD), at $\lambda = 595 \text{ nm}$, reached one. Subsequently it was then transferred to 500 ml of 2xTY media containing 500 μl of tetracycline solution (5 mg ml^{-1} in ethanol), which was allowed to grow overnight at 37°C and 180 rpm. The inserted plasmid confers tetracycline resistance to the bacterial cells, and thus the antibiotic was used to prevent contamination by other bacterial species.

After the bacteria were allowed to grow overnight, the media was then centrifuged at 4500 rpm for 30 minutes to form a pellet of the *E. coli* cells. The supernatant was then discarded and the pellet was resuspended in 20 ml of lysis buffer. The suspension was then sonicated at an amplitude of ten microns to break down the *E. coli* cells, according to the following scheme; *10 seconds on, 10 seconds off, 10 seconds on, 10 seconds off, 10 seconds on, 10 seconds off*. This was repeated three times with a 30 second break between repetitions. The suspension was kept on ice during the sonication to prevent the denaturation of the protein by heat and the action of proteolytic enzymes. The suspension was then centrifuged at 15000 rpm for 45 minutes. The supernatant was carefully collected, avoiding the solid pellet formed at the bottom and centrifuged again at 15000 rpm for one hour. The supernatant was then filtered through a 0.22 μm Minisart™ syringe filter (*Sartorius*), to remove any particulate matter remaining. The resulting yellow solution contained NQO2 in addition to other bacterial enzymes.

The protein content of the filtrate was determined by measuring the absorbance of the protein solution at 288 nm. The extinction coefficient of NQO2 was calculated, using Protein Calculator v3.3 (Putman), to be $1.71(\text{mg/ml})^{-1} \text{cm}^{-1}$ and $44380 \text{M}^{-1} \text{cm}^{-1}$.

4.2.3 Purification of NQO2

The presence of the His-tag allows the use of nickel-nitrilotriacetic acid (Ni-NTA) Superflow resin (*Novagen*), to purify the protein from other proteins produced by the bacterial culture. A gradient elution with increasing

concentrations of imidazole was used to separate NQO2 from the other bacterial proteins. Six to eight ml of Ni-NTA Superflow resin suspension in 30 per cent ethanol at a 1:1 ratio, were placed in a Bio-Rad purifying column. The ethanol was allowed to run through, and the remaining resin was washed with 2 X 5 ml of distilled water to remove all traces of ethanol.

The solution, containing NQO2 in addition to other contaminating proteins, was then added to the column, in addition to imidazole to a final concentration of 10 mM, to reduce non-specific binding. This was then left to equilibrate for one hour by slow rotation, at 4 °C. The flow through (FT), which is the solution in which the protein solution equilibrates with the resin, was allowed to run through. The protein was then eluted using increasing concentration of imidazole solution, from 10 mM to 250 mM, according to the protocol in Table 4. 1.

Elution fraction	1	2	3	3'
Imidazole concentration (mM)	10	50	200	200
Volume of solution (ml)	12	12	10	10

Table 4. 1 Elution protocol of NQO2 from Ni-NTA resin

4.2.4 Assessment of Protein Purity

The purity of the eluted fractions was tested using reducing sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE), where the protein samples are boiled in the presence of a reducing agent to break disulphide linkages and thus denaturing the protein structure.

4.2.4.1 Preparation of gel

The gels were prepared using the Bio-Rad mini gel casting system. The concentration of the polyacrylamide stacking gel used was 5 % and that of the resolving gel was 12.5 %. The resolving gel was first prepared according to the protocol in Table 4. 2, and then poured between the glass plates assembled in the Bio-Rad casting stand to a height of 3 cm from the top of the plates. The gel surface was covered with isopropanol, to remove any air bubbles present and to smooth the gel surface, and the gel was allowed to set for 30 to 45 minutes. The isopropanol was then drained off, and any excess dried with a tissue. The stacking gel was then prepared according to Table 4. 2, and added on top of the resolving gel. The appropriate well comb was then carefully inserted into the solution, avoiding the formation of air bubbles, and the gel was allowed to set for 30 to 45 minutes.

	Resolving gel	Stacking gel
Distilled water	4.7 ml	5.15 ml
10 % SDS	150 μ l	75 μ l
1.5 M Tris-HCl pH 8.8	3.75 ml	0.94 ml
Acrylamide/Bis-acrylamide 30 % (Severn Biotech LTD)	6.25 ml	1.75 ml
10 % w/v Ammonium persulphate	150 μ l	75 μ l
TEMED	7.5 μ l	5.5 μ l

Table 4. 2 Preparation of resolving and stacking gels of SDS-PAGE

The gel was then placed in the electrophoresis apparatus, and the reservoirs were filled with SDS-PAGE running buffer.

4.2.4.2 Preparation of samples and running of the gel

10 μ l of the protein samples were mixed with 5 μ l of 3x loading buffer and then heated to 95° C for five minutes. The samples were then centrifuged at 2000 rpm for one second, to gather the droplets together.

The treated samples and the protein molecular weight marker, either Precision plus protein marker (*Bio-Rad*) or Protein perfect markers (*Merck*) were then loaded into the wells of the gel, according to the order in Table 4. 3. The gel was then run at a constant voltage of 200 V for 50 minutes. The gel was then stained for 30 minutes using Coomassie staining solution, and then destained using Coomassie destaining solution for one hour.

Lane	1	2	3	4	5	6
Elution protein fraction, corresponding to	Protein marker	FT	1	2	3	3'
Imidazole concentration (mM)	–	–	10	50	200	200

Table 4. 3 Samples in gel lanes

4.2.5 Concentrating and Storage of NQO2

The pure eluted fractions were mixed together and concentrated using vivaspin concentrator tubes from Vivascience (*molecular weight cut off: 10,000*). The tubes were centrifuged at a speed of 4500 rpm and 2 x 25 ml of 50 mM Tris-HCl pH 8 were added to remove any traces of imidazole remaining from the elution protocol. The protein solution was then spun to a concentration of 0.4 mM. The concentration of the enzyme solution was calculated as previously mentioned in section 4.2.2. The enzyme solution was stored in 50 mM Tris-HCl pH 8, at 4 °C for future use.

4.2.6 Optimizing Enzyme Kinetics Protocol

4.2.6.1 UV visible absorbance scans

A preliminary UV/visible spectrophotometric scan was carried out on 100 μM NBDH (*TCI-Europe*) in 10 mM octyl- β -D-glucopyranoside in 50 mM Tris-HCl at pH 8.5 to determine the λ_{max} to be used in the fluorescence studies as the excitation wavelength.

4.2.6.2 Fluorescence emission and excitation spectrums

An emission spectrum for NBDH was obtained at $\lambda_{\text{max}} = 361$ nm as the excitation wavelength, to determine the wavelength at which the highest emission occurred. In addition, an excitation spectrum was also obtained, using an emission wavelength of 462 nm, to check the excitation wavelength of this compound. The NBDH solution used was at 100 μM in 10 mM octyl- β -D-glucopyranoside in 50 mM Tris-HCl at pH 8.5.

4.2.6.3 Effect of air exposure on the fluorescence of NBDH

To assess whether NBDH is oxidized upon air exposure, the fluorescence of NBDH in a cuvette was measured over a period of time. The emission of 100 μM NBDH at 462 nm, using 361 nm as the excitation wavelength was measured at zero time, and then after one, four and six hours. The initial stock of 10 mM NBDH was dissolved in dimethylsulphoxide (DMSO), and the subsequent dilution to 100 μM was carried out in 10 mM octyl- β -D-glucopyranoside in 50 mM Tris-HCl at pH 8.5.

4.2.6.4 Fluorescence emission controls for other reaction components

To ensure that none of the other components of the reaction interfere with the fluorescence of the cosubstrate, the fluorescence emission spectrum of 1 mM of menadione, coenzyme Q0, coenzyme Q1 and coenzyme Q2 were measured at the excitation wavelength of 361 nm. In addition, the emission spectrum of 200 μ M of NQO2, at the excitation wavelength of 361 nm was also measured. The substrates were dissolved in DMSO and the appropriate dilutions were then carried out in 10 mM octyl- β -D-glucopyranoside in 50 mM Tris-HCl at pH 8.5. The stock solution of NQO2 was in the storage buffer, 50 mM Tris-HCl pH 8 and was diluted to the appropriate concentration in 10 mM octyl- β -D-glucopyranoside in 50 mM Tris-HCl at pH 8.5.

4.2.7 Calibration Curve Relating Concentration of NBDH to its Fluorescence

The relationship between NBDH concentration and its fluorescence at 462 nm, when excited at 361 nm was explored by measuring the emissions of several concentrations of NBDH ranging between 10 to 350 μ M. For each concentration, three measurements were taken, where each measurement was an accumulation of five scans that were averaged. The NBDH solutions were kept on ice, in closed eppendorf tubes throughout the experiment to decrease the effect of air oxidation. The average fluorescence for each concentration was then plotted against its concentration.

4.2.8 Optimization of Published Reaction Conditions

The substrate used for the optimization of the reaction conditions was menadione, as it is a known substrate for NQO2 and the kinetics of the reaction are well documented (Boutin et al., 2005, Zhao et al., 1997).

The NBDH solution was divided into several 0.5 ml aliquots, which were kept sealed and in ice; the aliquot being used was replaced every hour with a fresh one, to minimize the effect of air oxidation. In an eppendorf tube containing the reaction medium, 10 mM octyl- β -D-glucopyranoside in 50 mM Tris-HCl at pH 8.5, the required volume of the substrate and cosubstrate were added to the reaction medium and were allowed to equilibrate for two minutes, at room temperature. The enzyme in 50 mM Tris-HCl pH 8 was then added to the mixture and the eppendorf was shaken to mix the reaction components. The mixing time was the same for all the eppendorfs, to ensure the consistency of the method used. The reaction mixture was then immediately transferred into a fluorescence cuvette and the decrease in fluorescence emission at 462 nm, using an excitation wavelength of 361 nm, was measured at one second intervals. The rate of decrease in the concentration of NBDH was then calculated from the decrease in fluorescence, using the linear equation obtained from the linear range of the calibration curve. The maximal rate of decrease in the concentration of the cosubstrate for each reaction was then plotted against the corresponding concentration of substrate producing a Michealis-Menten plot, allowing the calculation of K_m .

4.2.8.1 Enzyme concentration optimization

Several enzyme concentrations ranging from 2.5 μM to 37.5 μM were tested in order to determine the range at which the rate of the reaction and enzyme concentration are related by a linear equation. This range was determined by plotting the results during the experiment, and determining the necessary concentration range by examining the resultant curve. The optimal time the reaction took to reach completion was also a factor in deciding the concentration of enzyme to be used in the kinetic experiment. The required duration of the reaction was between 15 to 30 minutes, as reactions taking less than 15 minutes would not allow enough time for monitoring the initial rate of reaction. On the other hand, reactions taking more than 30 minutes would be too lengthy to allow all the required reactions to be completed in one day. The concentration of NBDH used was 100 μM , which is in the upper linear range of the calibration curve. While the concentration of menadione used was 700 μM , which is the highest concentration possible for analysis of menadione, due its aqueous solubility. For each enzyme concentration, triplicate measurements were preformed.

4.2.8.2 Temperature choice

The choice of temperature was determined by carrying out two sets of reactions, one at 25 °C and the other at 37 °C. The concentrations of the enzyme and NBDH were kept constant at 10 μM and 100 μM , respectively. The menadione concentration range was between 10 μM to 150 μM , which was adopted from the published method, where the used range was between 5 μM

and 100 μM (Boutin et al., 2005). For each menadione concentration, triplicate measurements were carried out.

4.2.8.3 Reproducibility testing

Using the optimum conditions, (Table 4. 4) the rate experiment using menadione as the substrate was repeated twice, in order to ensure that the results were reproducible and the calculated values of K_m were comparable to those published in the literature (Boutin et al., 2005, Zhao et al., 1997). The range of concentration used for menadione was again between 10 μM and 150 μM .

Reaction medium	10mM octyl- β -D-glucopyranoside, 50mM Tris-HCl at pH 8.5
Total cuvette volume	1 ml
Enzyme concentration	10 μM
Cosubstrate concentration	100 μM
Excitation wavelength	361 nm
Slit opening	5 mm
Emission wavelength	462 nm
Slit opening	5 mm
Temperature	37 $^{\circ}\text{C}$

Table 4. 4 Optimum reaction conditions chosen.

4.2.9 K_m Determination for Novel Substrates

4.2.9.1 Determination of linear relationship between enzyme concentration and reaction rate

For coenzyme Q0, coenzyme Q1 and coenzyme Q2 the enzyme concentration used in the kinetic experiment was chosen by determining the linear

concentration at which the enzyme operates for each substrate using the method described in section 4.2.8.1. The enzyme concentration range tested for coenzyme Q0 was from 2.5 μM to 10 μM , while that tested for coenzyme Q1 and coenzyme Q2 was from 20 μM to 150 μM . The range used was determined by plotting the results during the experiment and determining the required range by examining the resultant curve. For GC201, the testing of the optimum enzyme concentration was not possible due to the limited quantity available of the substrate; therefore the same optimum concentration for menadione was used. The cosubstrate concentration used was kept constant at 100 μM . The concentrations of coenzyme Q0, coenzyme Q1, coenzyme Q2 and menadione were 150, 70, 350 and 150 μM respectively.

4.2.9.2 K_m determination

The Michealis-Menten plot for each substrate was determined using the same method and conditions described in Table 4. 4 except for the concentration of enzyme. The enzyme concentration was changed according to the linear range relating the enzyme concentration and the reaction rate, for the specific substrate. For GC201 and coenzyme Q0, the enzyme concentration used was 10 μM , while for both coenzyme Q1 and coenzyme Q2, the enzyme concentration used was 100 μM .

In addition to the different concentration of the enzyme used for each substrate, the substrate concentration range used to determine the K_m was also different. For coenzyme Q0, the substrate concentration ranged from 20 μM to 400 μM , while for coenzyme Q1 the range used was between 20 μM to 140 μM . For

coenzyme Q2 the range used was from 30 μM to 1.05mM and the tested range for GC201 was from 25 μM to 1.2 mM. The range of substrate concentration used was determined by plotting Michealis-Menten plot of the results during the experiment and determining the required range by examining the resultant curve. Due to the decrease in the rate of reaction at higher concentrations of the substrates, direct determination of the K_m values from the Michealis-Menten plot using non-linear regression was not possible. Thus the Lineweaver-Burke diagram was plotted for each substrate, and was used to calculate the K_m values using linear regression.

4.2.9.3 Substrate-cosubstrate reaction controls

To verify that the decrease in fluorescence of the cosubstrate is not due to a reaction between the cosubstrate and substrate, but rather due to the action of the enzyme on both the substrate and cosubstrate. Reaction controls were done for all the different concentrations of substrate used in determining the K_m for coenzyme Q0, coenzyme Q1 and coenzyme Q2. For GC201, only a few concentrations, 25 μM , 400 μM , 700 μM and 1200 μM were chosen due to the limited availability of the substrate. This was performed by adding the required volumes of the substrate and the cosubstrate to an eppendorf tube containing the reaction medium, 10mM octyl- β -D-glucopyranoside in 50mM Tris-HCl at pH 8.5. The eppendorf was then shaken and its contents were immediately transferred to the cuvette and the fluorescence of the mixture was monitored for three minutes.

Chapter Five

Kinetic Studies for NQO2 II:

Results and Discussion

5.1 Protein Preparation

5.1.1 Assessing Protein Purity Using SDS-PAGE

SDS-PAGE confirmed that the elution protocol, previously described in section 4.2.3, can be used to successfully to purify NQO2 from other contaminating proteins produced by the *E.coli* culture. The identity of the enzyme was confirmed from its molecular weight, by comparison to the known molecular weights of the protein marker bands. Figure 5. 1 shows an example of the gel used to asses the purity of NQO2 after its purification. Lane one shows the protein marker, the flow-through (FT) is shown in lanes two and three, followed by the elution fractions in increasing imidazole concentrations in lanes four to seven. The fractions eluted using 200 mM imidazole, lanes six and seven, showed a solitary band of protein, identified as NQO2 with the His-tag attached, with a total molecular weight of 27.4 kDa. The absence of other bands, confirms that there were no contaminating proteins eluted with NQO2.

5.2 Optimizing Enzyme Kinetics Protocol

5.2.1 UV Visible Spectrum of NBDH

The UV/visible spectrum of NBDH showed two peaks where strong absorbance occurs, one at 220 nm and the other at 361 nm (Figure 5. 2). Absorbance at 220 nm is not specific, as most aromatic compounds absorb

around that wavelength and is thus subject to interference. Therefore the peak at 361 nm was chosen as the excitation wavelength for the fluorescence assay.

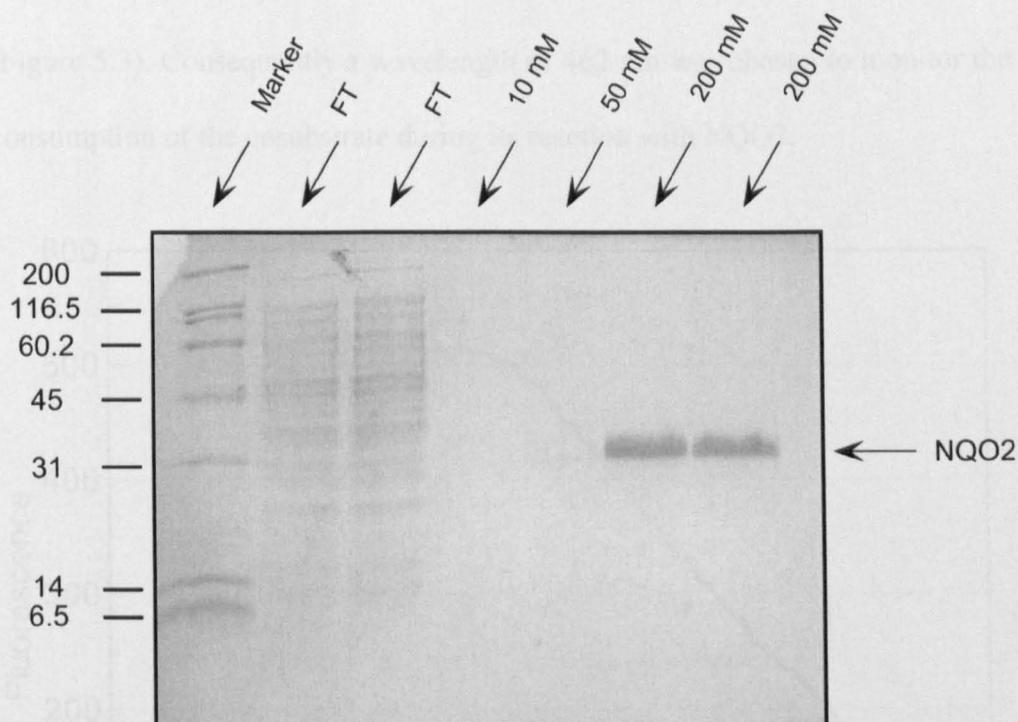


Figure 5. 1 SDS-PAGE showing protein marker, flow-through (FT) and elution samples at increasing imidazole concentrations. Molecular weights of marker bands are in kDa.

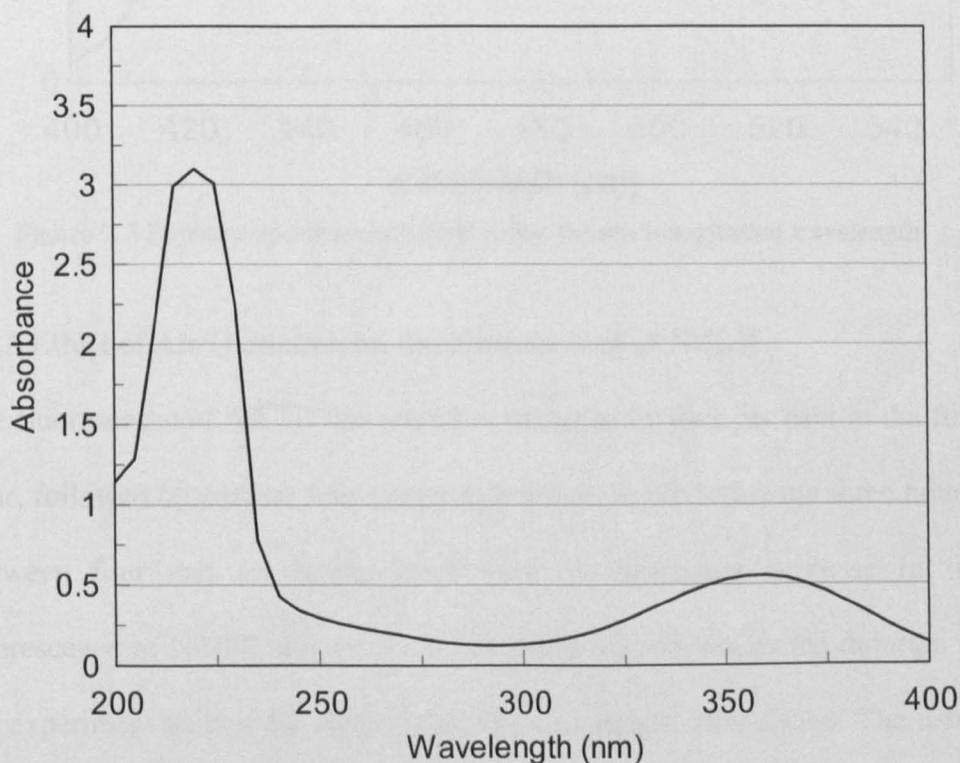


Figure 5. 2 UV/visible absorbance spectrum of NBDH

5.2.2 Fluorescence Emission Spectrum of NBDH

The emission spectrum of NBDH showed that the highest emission fluorescence intensity of NBDH, when excited at 361 nm, occurs at 462 nm (Figure 5.3). Consequently a wavelength of 462 nm was chosen to monitor the consumption of the cosubstrate during its reaction with NQO2.

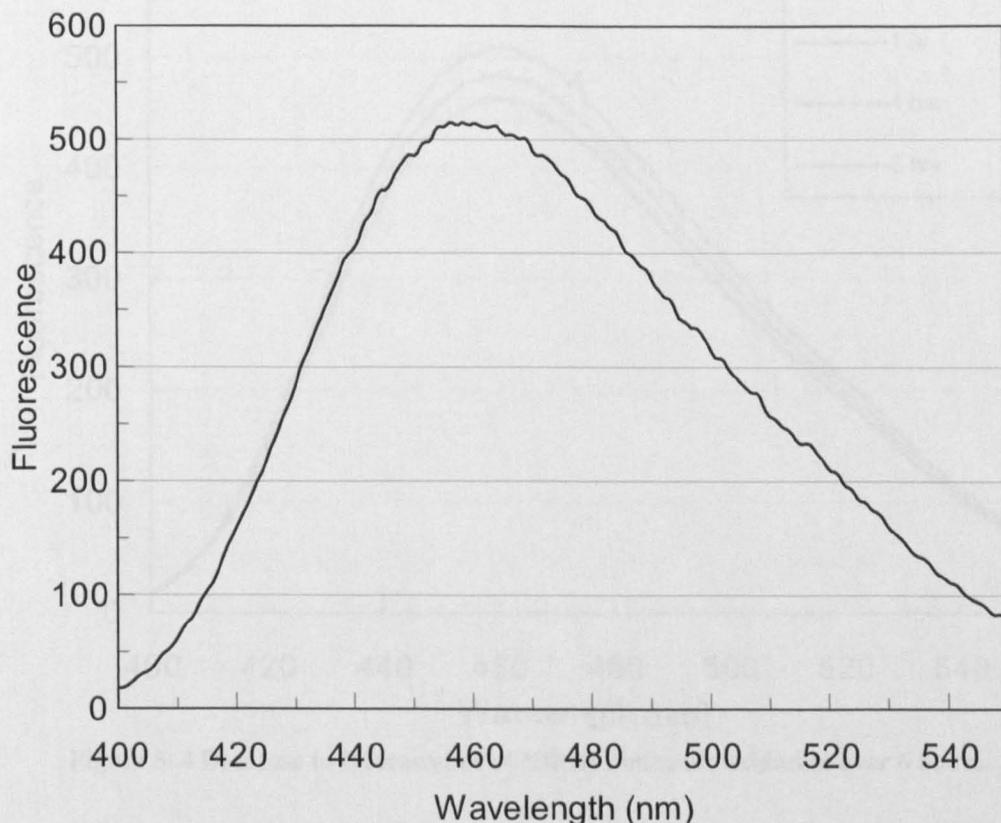


Figure 5.3 Emission spectrum of NBDH using 361 nm as excitation wavelength

5.2.3 Effect of Air Oxidation on the Fluorescence of NBDH

The fluorescence of NBDH was found to decrease by four per cent in the first hour, followed by another four per cent decrease, in the following three hours. Between four and six hours, there was no significant decrease in the fluorescence of NBDH (Figure 5.4). Six hours was chosen as the duration of the experiment as it is far longer than any experiment would take. The usual reaction time ranged between 10 to 15 minutes, thus it was thought that such a

decrease in fluorescence, would not cause any significant interference in the reaction. To minimize any possible effect of air oxidation, the NBDH solution used in the assay was divided into several 0.5 ml allocates which were kept on ice and the opened allocate was replaced with a fresh one every hour.

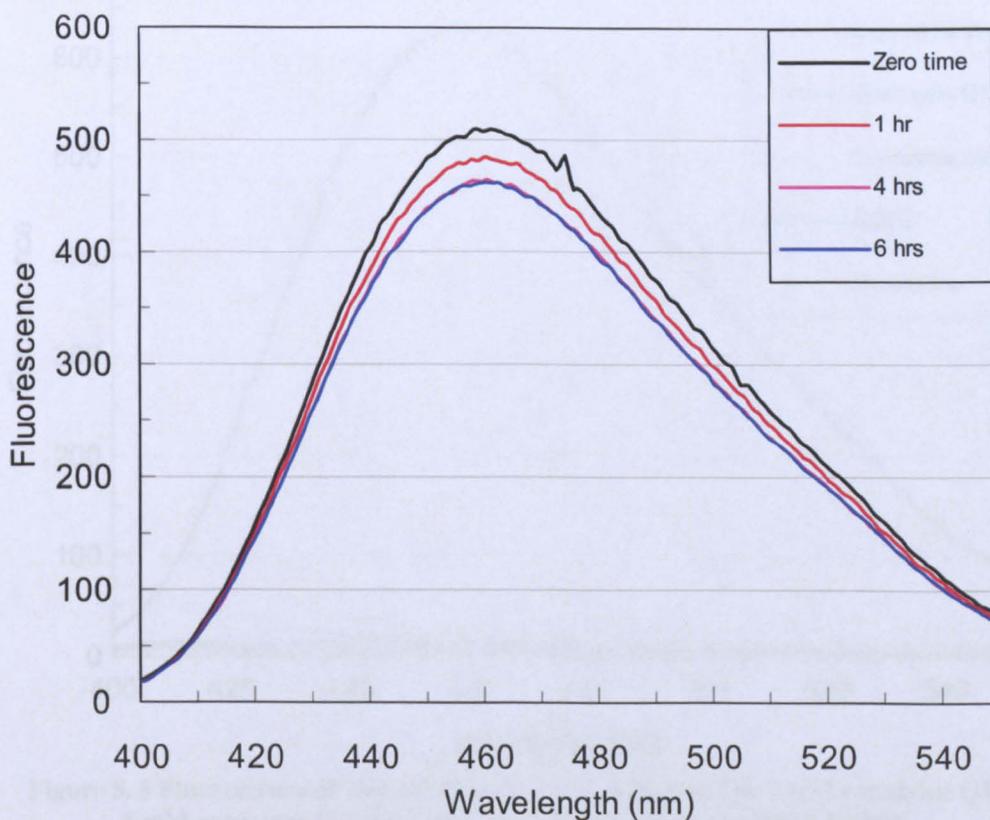


Figure 5. 4 Decrease in fluorescence of NBDH due to air oxidation over 6 hours.

5.2.4 Fluorescence Emission Measurements for Other Reaction Components

Except for NBDH, all the other components of the reaction did not show any significant fluorescence at 462 nm when excited at 361 nm (Figure 5. 5). This indicated that their contribution to fluorescence at this wavelength is negligible and thus would not interfere with the monitoring of the fluorescence of NBDH during the reaction of NQO2 with the substrate. The concentrations used for the measurements of the fluorescence are much larger than those used during

the experimental procedure. The solutions were made up in 10 mM octyl- β -D-glucopyranoside in 50 mM Tris-HCL at pH 8.5, and thus the effect of the buffer solution used in the experiment was also accounted for.

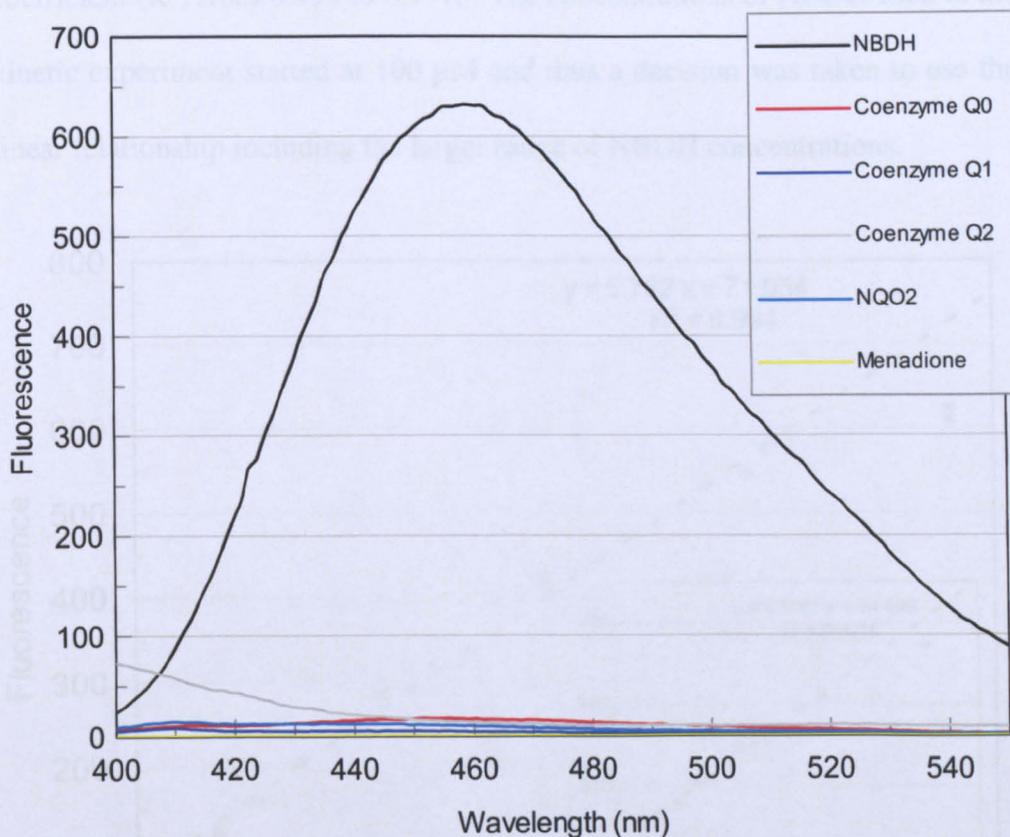


Figure 5. 5 Fluorescence of 200 μ M NQO2, 1 mM coenzyme Q0, 1 mM coenzyme Q1, 1 mM coenzyme Q2 and 1 mM menadione relative to 100 μ M NBDH.

5.2.5 Calibration Curve for NBDH

The calibration curve relating the concentration of the NBDH to its fluorescence showed that up to 120 μ M of NBDH, the relationship between the concentration of NBDH and its fluorescence is linear (Figure 5. 7). The relationship between the concentration of NBDH and its fluorescence can be determined using the following linear equation,

$$y = 5.762x + 71.05 \quad (5.1)$$

The linear relationship between the concentration and fluorescence of NBDH is slightly improved if the concentration of NBDH considered is decreased from 120 μM to 85 μM . This is demonstrated by a small increase in the regression coefficient (R^2) from 0.994 to 0.9975. The concentrations of NBDH used in the kinetic experiment started at 100 μM and thus a decision was taken to use the linear relationship including the larger range of NBDH concentrations.

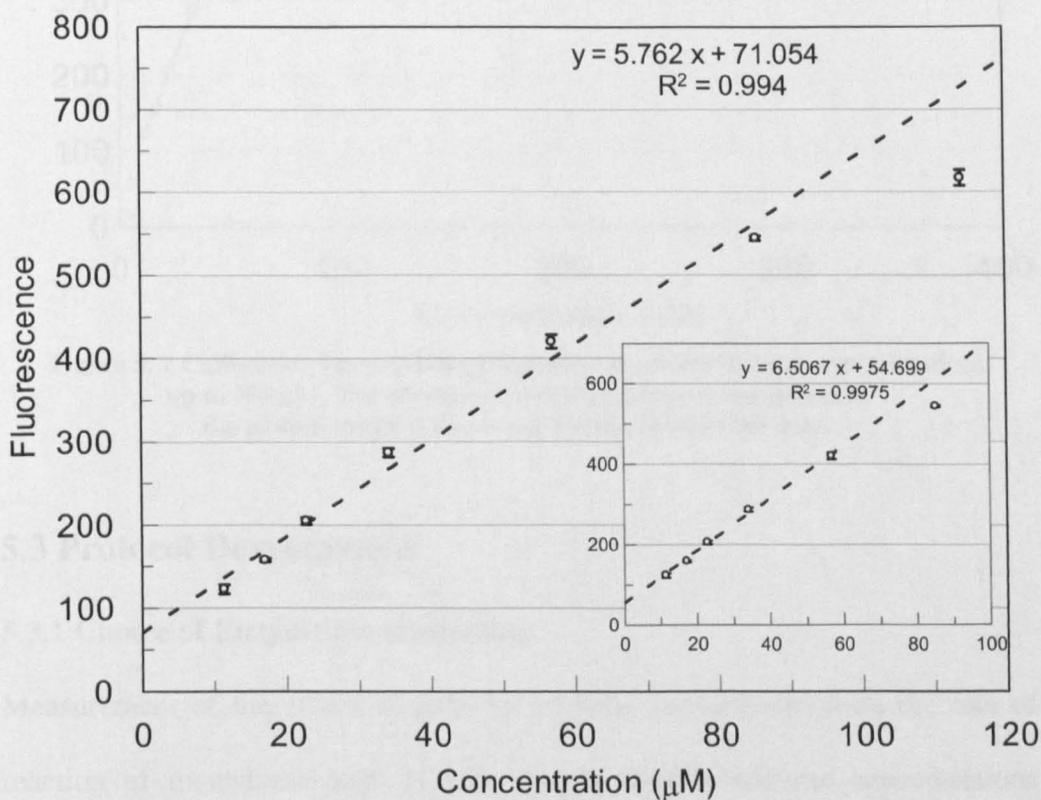


Figure 5. 6 Calibration curve relating fluorescence of NBDH to its concentration, up to 120 μM . Insert shows calibration curve of concentrations of NBDH up to 85 μM . The absence of error bars from some points in the plotted graph is due to the small errors in the data.

Upon the further increase in the concentration of NBDH, the relationship between the fluorescence and concentration is no longer linear, but reaches a maximum level and forms a plateau (Figure 5. 7). Consequently there is no further increase in fluorescence by increasing the concentration of NBDH.

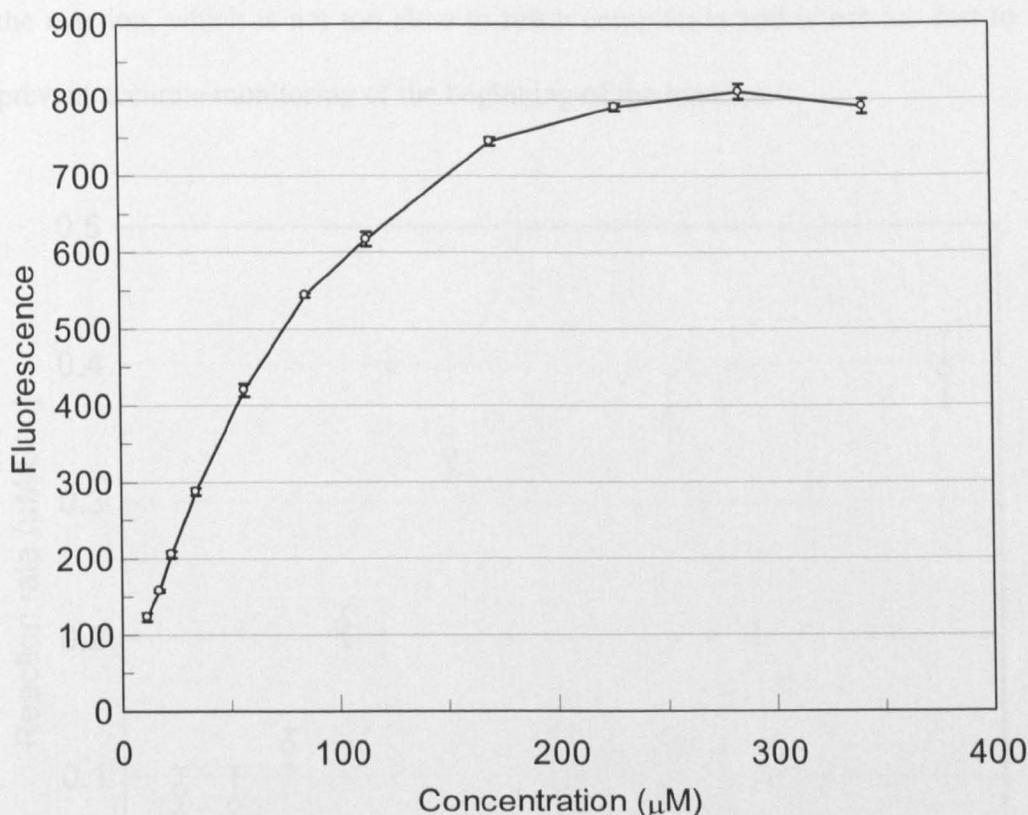


Figure 5.7 Calibration curve relating fluorescence of NBDH to its concentration, up to 350 μM . The absence of error bars from some points in the plotted graph is due to the small errors in the data.

5.3 Protocol Development

5.3.1 Choice of Enzyme Concentration

Measurement of the effect of different enzyme concentrations on the rate of reaction of menadione with NQO2 showed that at enzyme concentrations ranging between 5 μM and 15 μM , there is a linear relationship between the enzyme concentration and reaction rate (Figure 5.8). While at both lower and higher enzyme concentrations, there is no noticeable effect caused by increasing the enzyme concentration. As the enzyme concentration of 10 μM was found to lie within the linear range of the graph, it was chosen to test the remaining conditions for menadione assay, such as the effect of the temperature and reproducibility of results. It also allows appropriate time for

the reaction, which is not too slow to reach completion and is not too fast to prevent accurate monitoring of the beginning of the reaction.

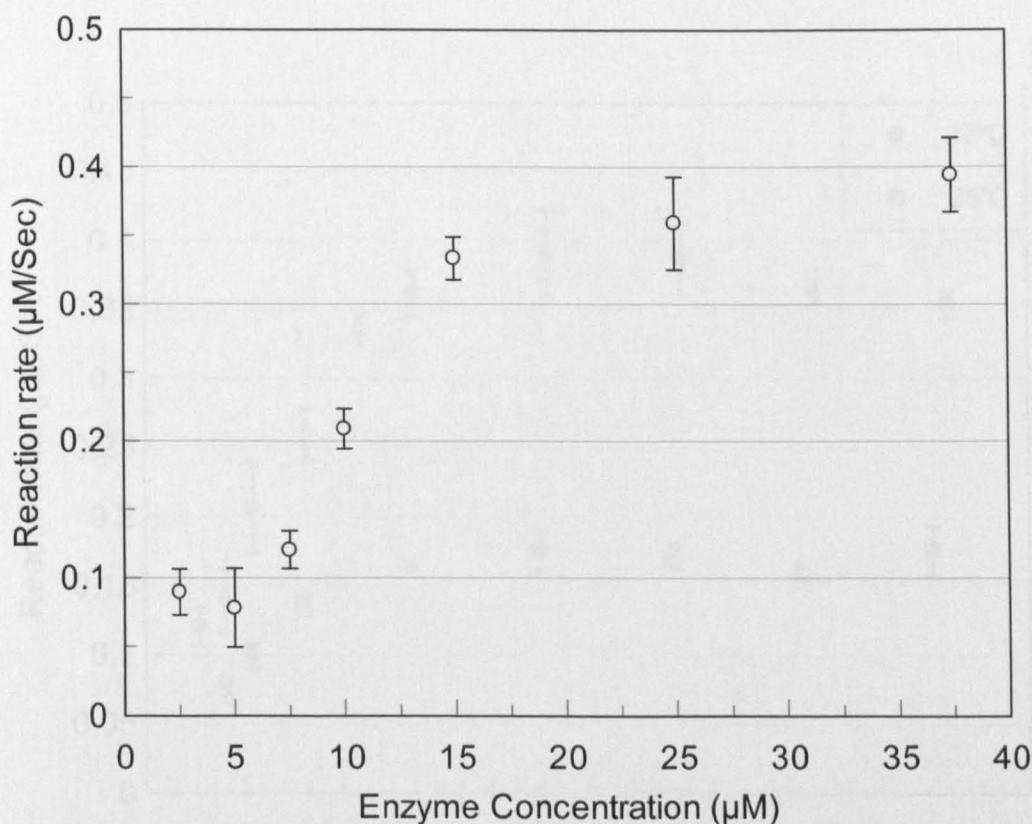


Figure 5. 8 Effect of enzyme concentration on the rate of reaction of NQO2 with menadione

5.3.2 Effect of temperature on reaction of menadione and K_m

When the two Michealis-Menten plots at different temperatures were overlaid, it was found that they share very similar profiles (Figure 5. 9). With the exception of the temperature difference between the two experiments, all other remaining conditions such as enzyme and substrate concentrations were kept identical. The substrate concentration at which V_{max} (maximum velocity of the reaction) occurs is around 85 µM for both temperatures. The values of V_{max} were found to be different, 0.175 µM sec⁻¹ and 0.39 µM sec⁻¹ for 25 °C and 37 °C respectively. The difference in the V_{max} values is caused by the increase in the reaction rate at the higher temperature. The value of K_m calculated from

both curves is approximately 20 μM . The published K_m values for the reaction of menadione and NQO2, using NBDH as a cosubstrate are 6.7 μM and 17 μM (Kwiek et al., 2004).

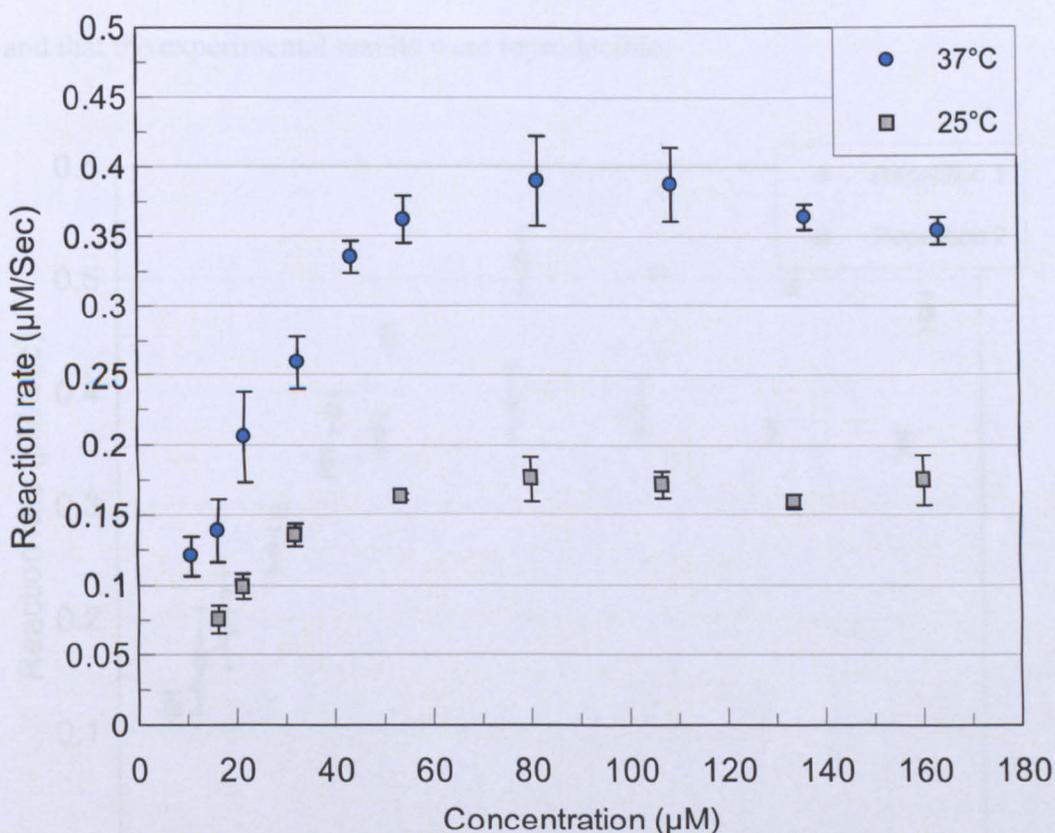


Figure 5.9 Michealis-Menten plots for the reaction between NQO2 and menadione at 25 °C and 37 °C. The absence of error bars from some points in the plotted graph is due to the small errors in the data.

5.3.3 Determining the Reproducibility of Results Using the Optimized Parameters

The same experiment using the optimized tested parameters was repeated twice, to test the reproducibility of the results. Both repetitions showed almost identical profiles (Figure 5. 10). The calculated K_m values of 20 and 26 μM were considered to be similar, as previously published results using the same method showed similar errors (Boutin et al., 2005). On the other hand, the V_{max} values were calculated to be 0.53 $\mu\text{M sec}^{-1}$ and 0.36 $\mu\text{M sec}^{-1}$. The large

difference in the V_{max} values is thought to be caused by experimental and instrumental variations, caused by the experiments being carried out on different days. These variations though, did not affect the K_m values of the experiment, thus it was decided that the chosen conditions were satisfactory and that the experimental results were reproducible.

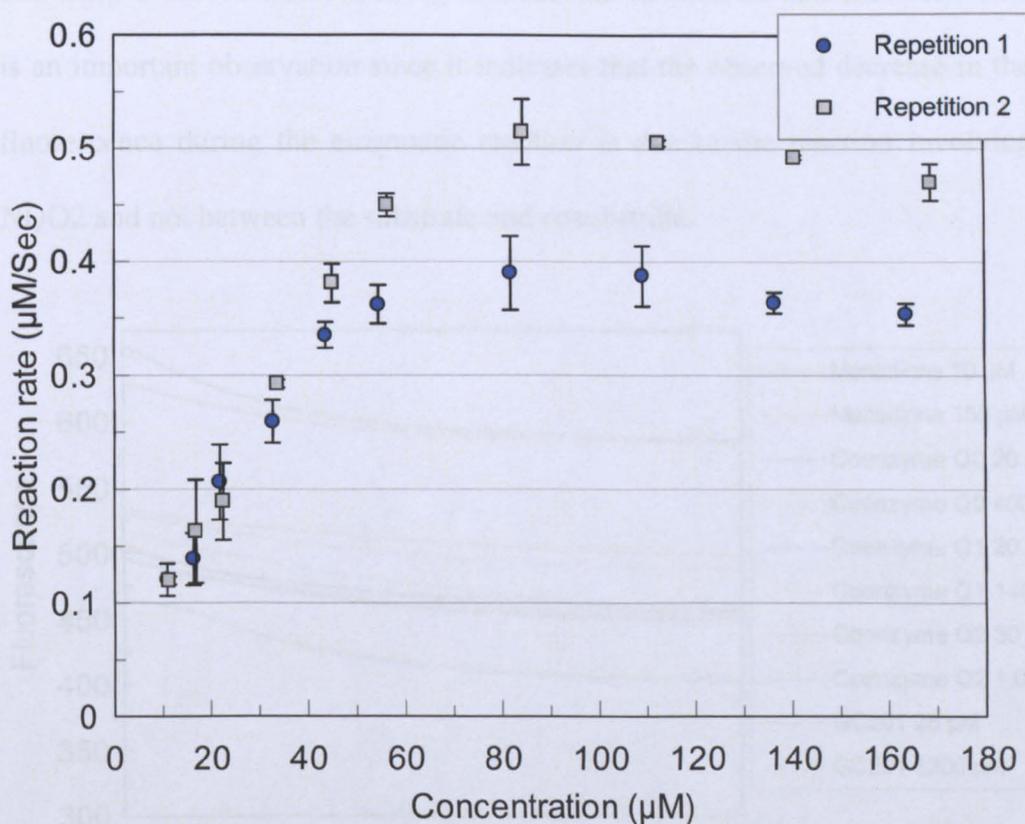


Figure 5. 10 Michealis-Menten plots for two kinetic experiments using menadione. The absence of error bars from some points in the plotted graph is due to the small errors in the data.

5.4 Enzyme Kinetics for Substrates

5.4.1 Substrate/cosubstrate Mixture Fluorescence Monitoring

The decrease in the fluorescence of the mixture of cosubstrate/substrate mixture was measured to determine the effect of the different substrates on the fluorescence of the cosubstrate. The monitoring of the change in fluorescence of the cosubstrate/substrate mixtures showed no decrease in the fluorescence of

the NBDH upon its mixing with any of the tested substrates; menadione, coenzyme Q0, coenzyme Q1 and coenzyme Q2 (Figure 5. 11). The concentrations used were the same as those used in determining the K_m values. For GC201 only a few concentrations were tested due to the limited availability of the substrate. The absence of a large decrease in the fluorescence suggests that there is no reaction occurring between the cosubstrate and substrate. This is an important observation since it indicates that the observed decrease in the fluorescence during the enzymatic reaction is due to the reaction involving NQO2 and not between the substrate and cosubstrate.

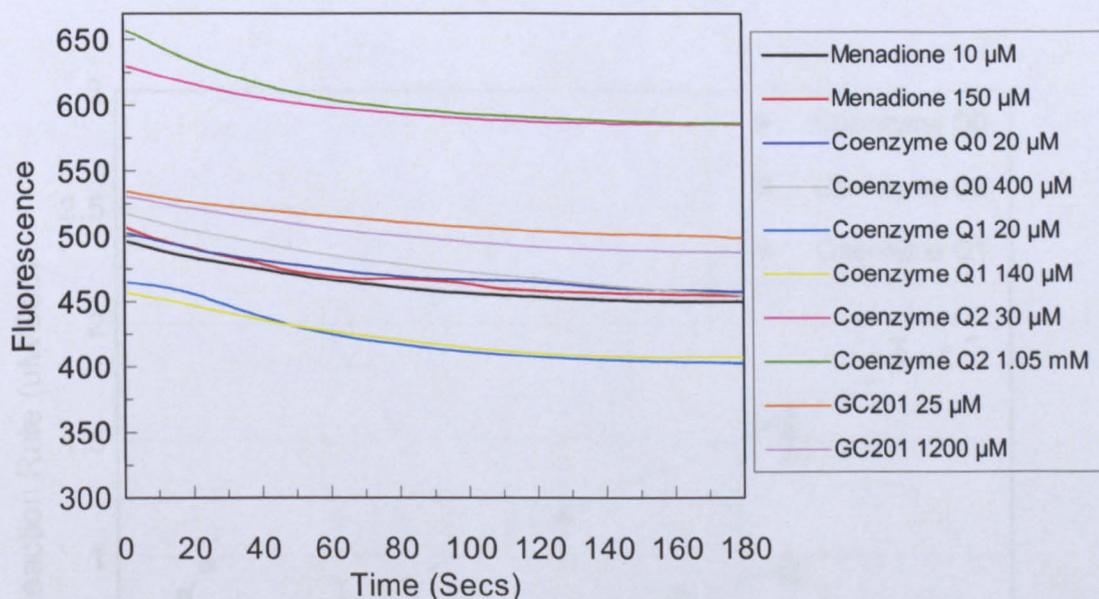


Figure 5. 11 Fluorescence decay over time of cosubstrate/substrate mixture solutions at the highest and lowest substrate concentrations tested.

5.4.2 Relationship between enzyme and reaction rate

The relationship between the concentration of NQO2 and the rates of reactions with coenzyme Q0, coenzyme Q1 and coenzyme Q2 were found to be linear for all substrates, within the concentration range tested (Figure 5. 12). The concentration of NBDH used for all substrates was 100 μM while the

concentrations used for the substrates was differed according to the range of substrate concentration tested.

The concentration of NQO2 that was chosen to determine the K_m value for coenzyme Q0 was 10 μM , since this occurs in the centre of the tested range and was also used in the measurement of the K_m of menadione. In addition, it was also used to asses and optimize the experimental conditions. The concentration used to determine the rate of reaction for both coenzyme Q1 and coenzyme Q2 was 100 μM , since the lower enzyme concentration resulted in very slow reactions.

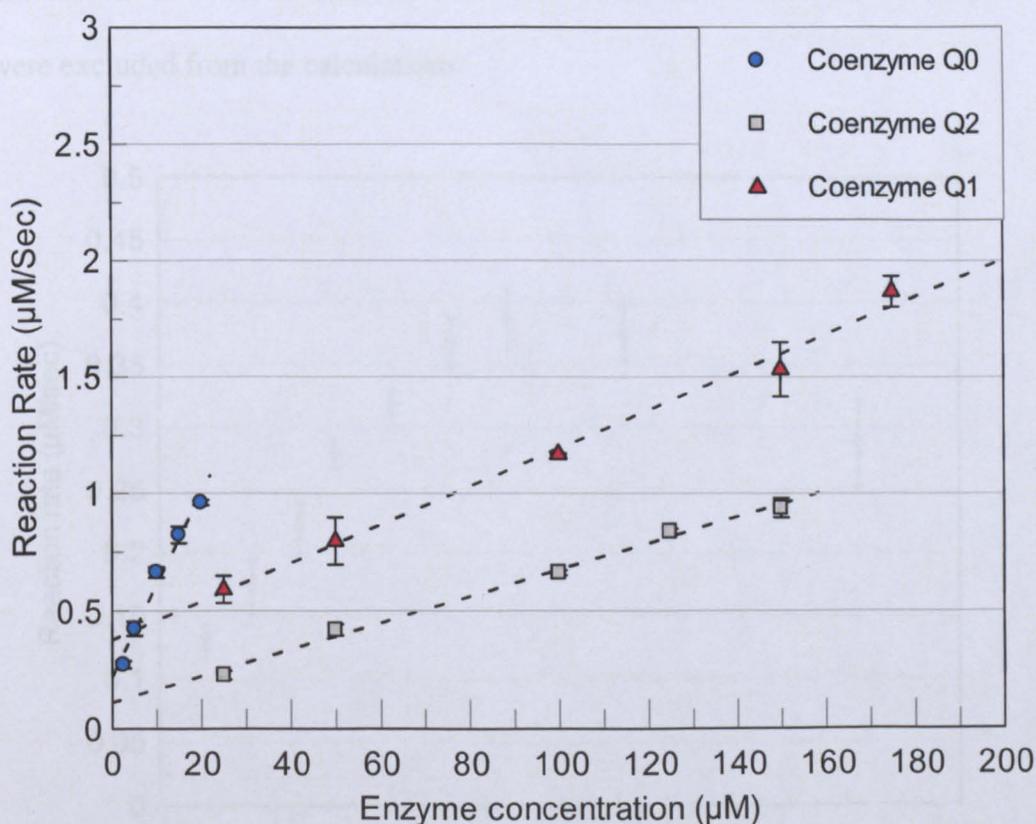


Figure 5.12 Effect of enzyme concentration on reaction rate. The absence of error bars from some points in the plotted graph is due to the small errors in the data.

5.4.3 K_m determination for Substrates

5.4.3.1 Coenzyme Q0

The Michealis-Menten plot shows a decrease in the rate of reaction at concentrations of coenzyme Q0 above 150 μM (Figure 5. 13). This prevented the use of non-linear regression to determine the K_m value directly from the Michealis-Menten plot. The Hanes-Woolf representation of the data was also plotted and it was found not to be linear, preventing its use in determining the K_m and V_{max} (Figure 5. 14). Therefore the Lineweaver-Burke plot was used to determine the kinetic constants. The K_m was calculated to be 54.65 μM , and V_{max} was found to be 0.463 $\mu\text{M sec}^{-1}$ (Figure 5. 15). The points at which a decrease in the rate of reaction was observed in the Michealis-Menten plot were excluded from the calculations.

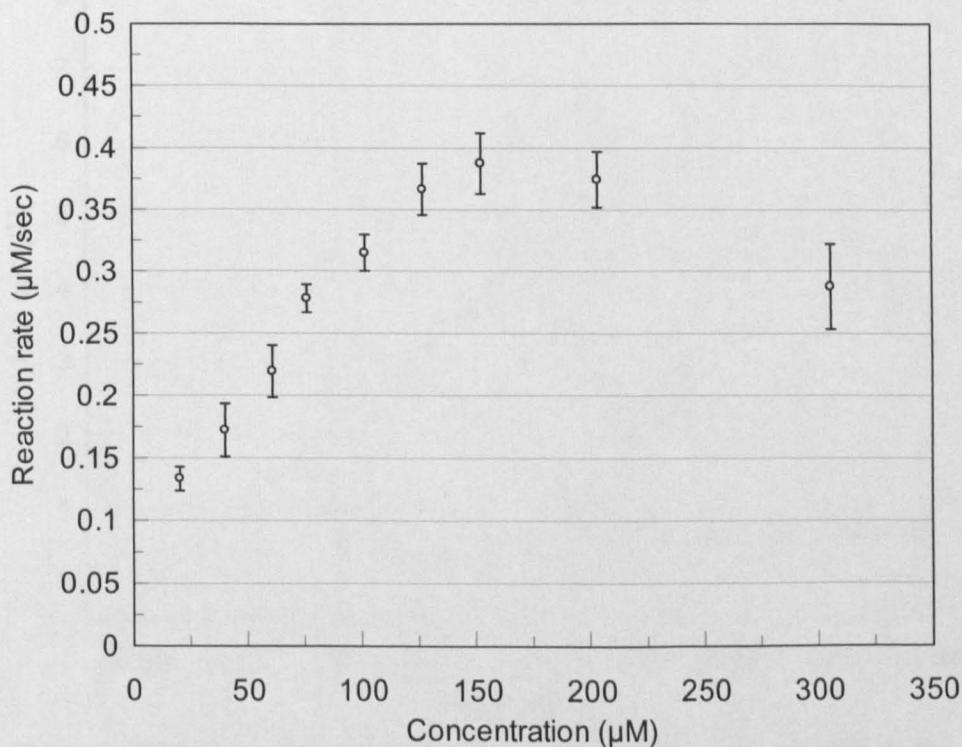


Figure 5. 13 Michealis-Menten plot for the reaction of coenzyme Q0 with NQO2.

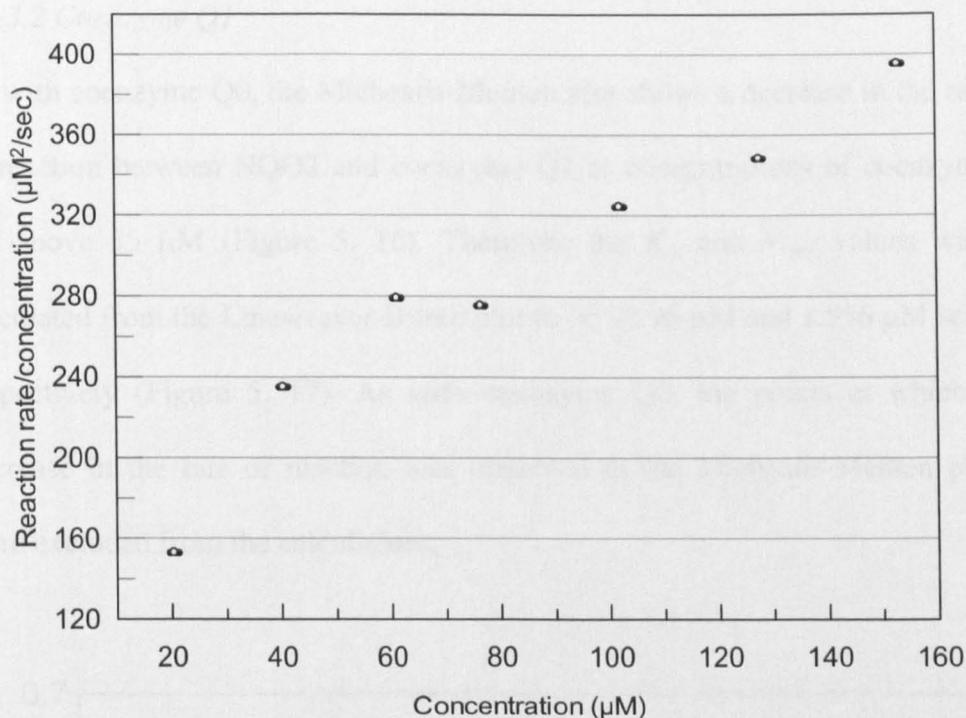


Figure 5. 14 Hanes-Woolf plot for reaction of coenzyme Q0 with NQO2. The absence of error bars from some points in the plotted graph is due to the small errors in the data.

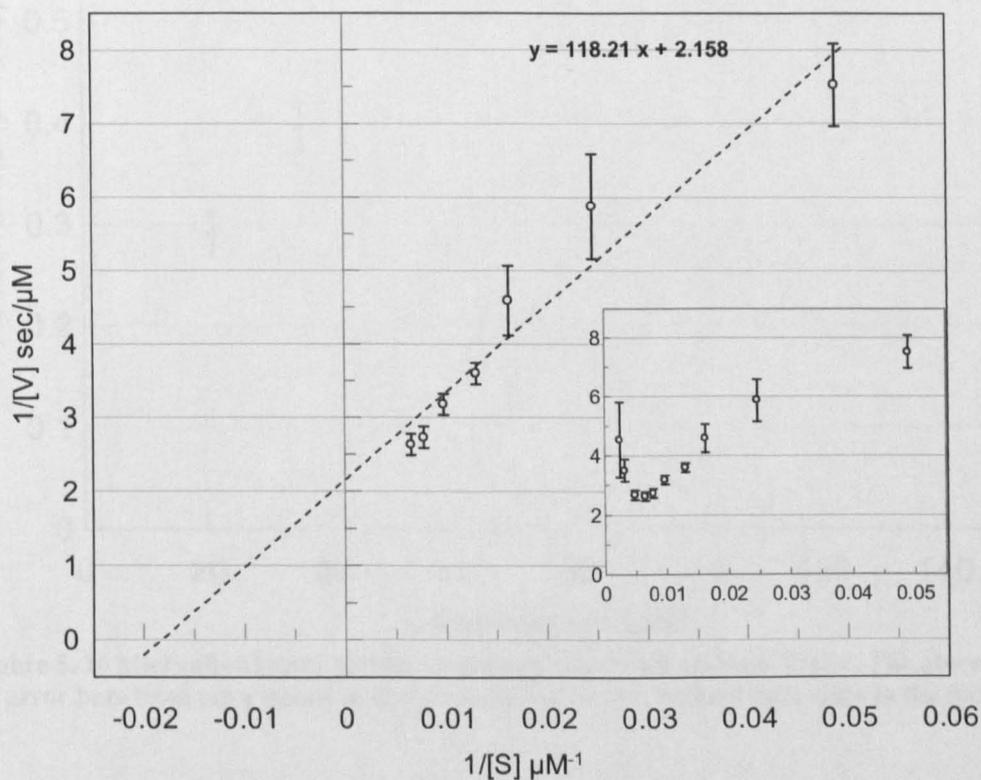


Figure 5. 15 Lineweaver-Burke plot for reaction of coenzyme Q0 with NQO2. The absence of error bars from some points in the plotted graph is due to the small errors in the data. The main graph shows concentrations of coenzyme Q0 up to 150 μM, while the insert shows all data points collected.

5.4.3.2 Coenzyme Q1

As with coenzyme Q0, the Michealis-Menten plot shows a decrease in the rate of reaction between NQO2 and coenzyme Q1 at concentrations of coenzyme Q1 above 55 μM (Figure 5. 16). Therefore the K_m and V_{max} values were calculated from the Lineweaver-Burke plot to be 97.98 μM and 1.556 $\mu\text{M sec}^{-1}$ respectively (Figure 5. 17). As with coenzyme Q0, the points at which a decrease in the rate of reaction was observed in the Michealis-Menten plot were excluded from the calculations.

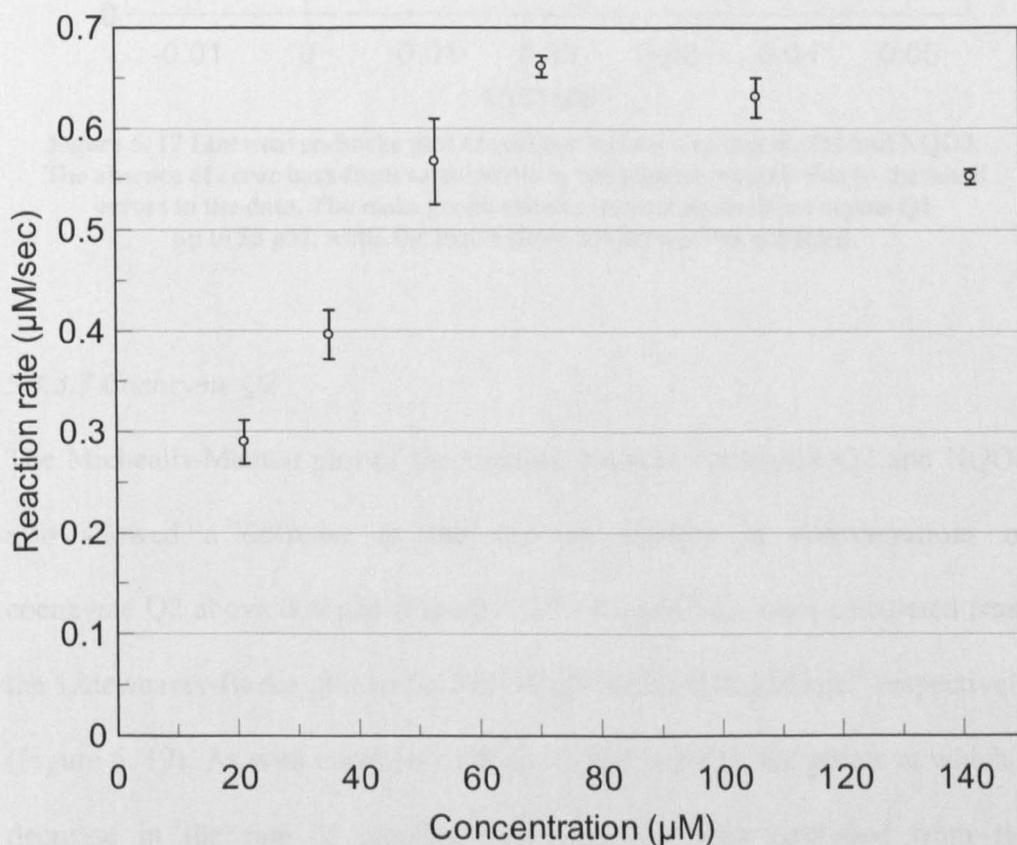


Figure 5. 16 Michealis-Menten plot for reaction of coenzyme Q1 with NQO2. The absence of error bars from some points in the plotted graph is due to the small errors in the data.

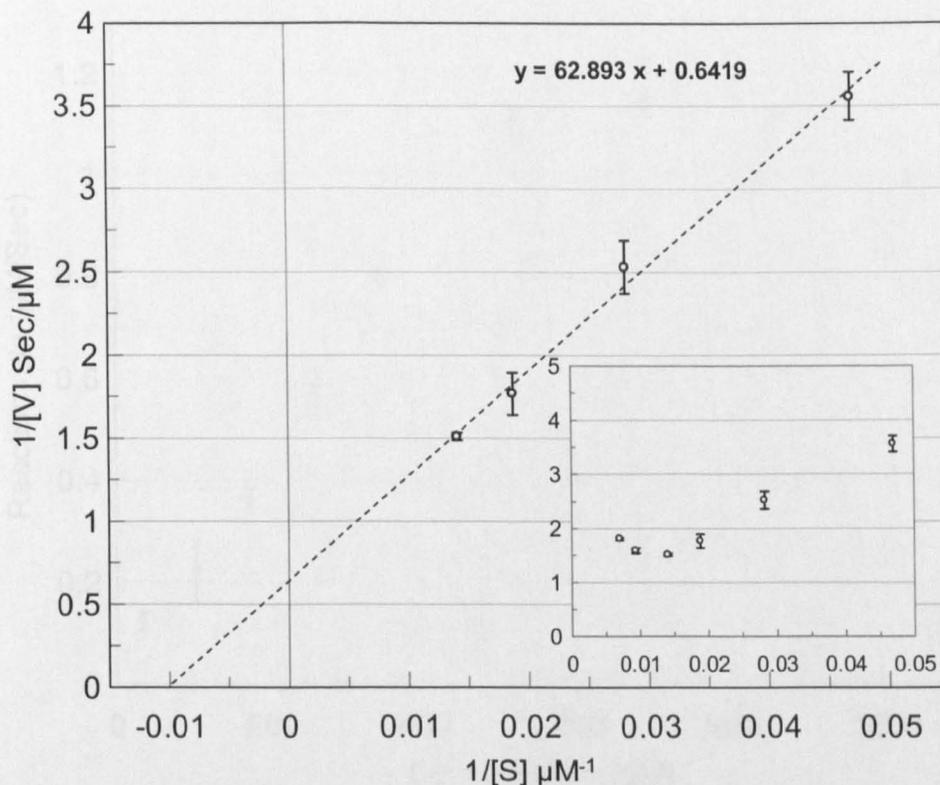


Figure 5. 17 Lineweaver-Burke plot of reaction between coenzyme Q1 and NQO2. The absence of error bars from some points in the plotted graph is due to the small errors in the data. The main graph shows concentrations of coenzyme Q1 up to 55 μM , while the insert shows all data points collected.

5.4.3.3 Coenzyme Q2

The Michealis-Menten plot of the reaction between coenzyme Q2 and NQO2 also showed a decrease in the rate of reaction at concentrations of coenzyme Q2 above 600 μM (Figure 5. 18). K_m and V_{max} were calculated from the Lineweaver-Burke plot to be 283.38 μM and 1.028 $\mu\text{M sec}^{-1}$ respectively (Figure 5. 19). As with coenzyme Q0 and coenzyme Q1, the points at which a decrease in the rate of reaction was observed were excluded from the calculations.

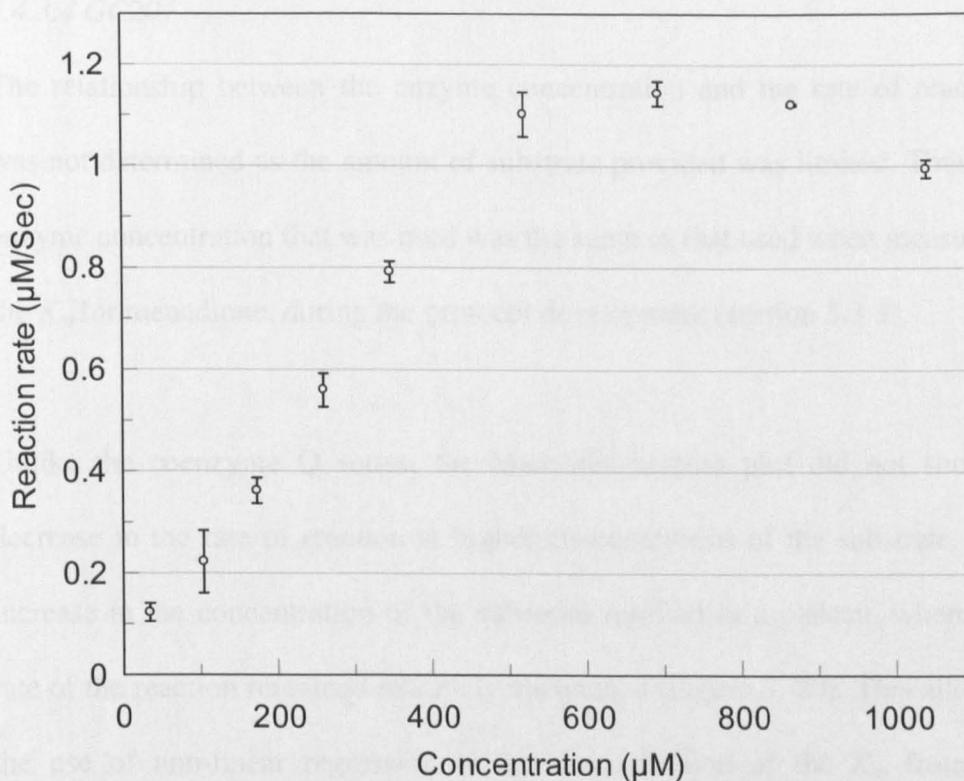


Figure 5. 18 Michealis-Menten plot for reaction of coenzyme Q2 with NQO2. The absence of error bars from some points in the plotted graph is due to the small errors in the data.

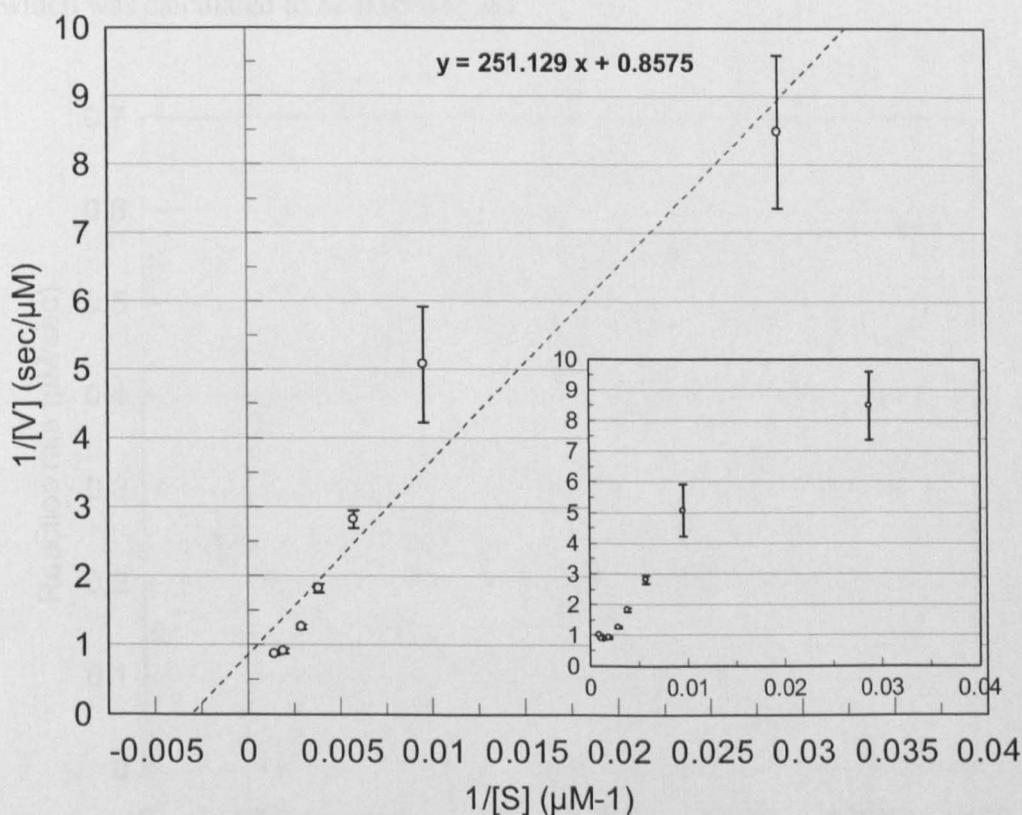


Figure 5. 19 Lineweaver-Burke plot of reaction coenzyme Q2 and NQO2. The absence of error bars from some points in the plotted graph is due to the small errors in the data. The main graph shows concentrations of coenzyme Q2 up to 600 µM, while the insert shows all data points.

5.4.3.4 GC201

The relationship between the enzyme concentration and the rate of reaction was not determined as the amount of substrate provided was limited. Thus the enzyme concentration that was used was the same as that used when measuring the K_m for menadione, during the protocol development (section 5.3.3).

Unlike the coenzyme Q series, the Michealis-Menten plot did not show a decrease in the rate of reaction at higher concentrations of the substrate. The increase in the concentration of the substrate resulted in a plateau, where the rate of the reaction remained relatively unchanged (Figure 5. 20). This allowed the use of non-linear regression in the determination of the K_m from the classical Michealis-Menten plot. The K_m was found to be 168 μM , and the V_{max} which was calculated to be 0.65 $\mu\text{M sec}^{-1}$

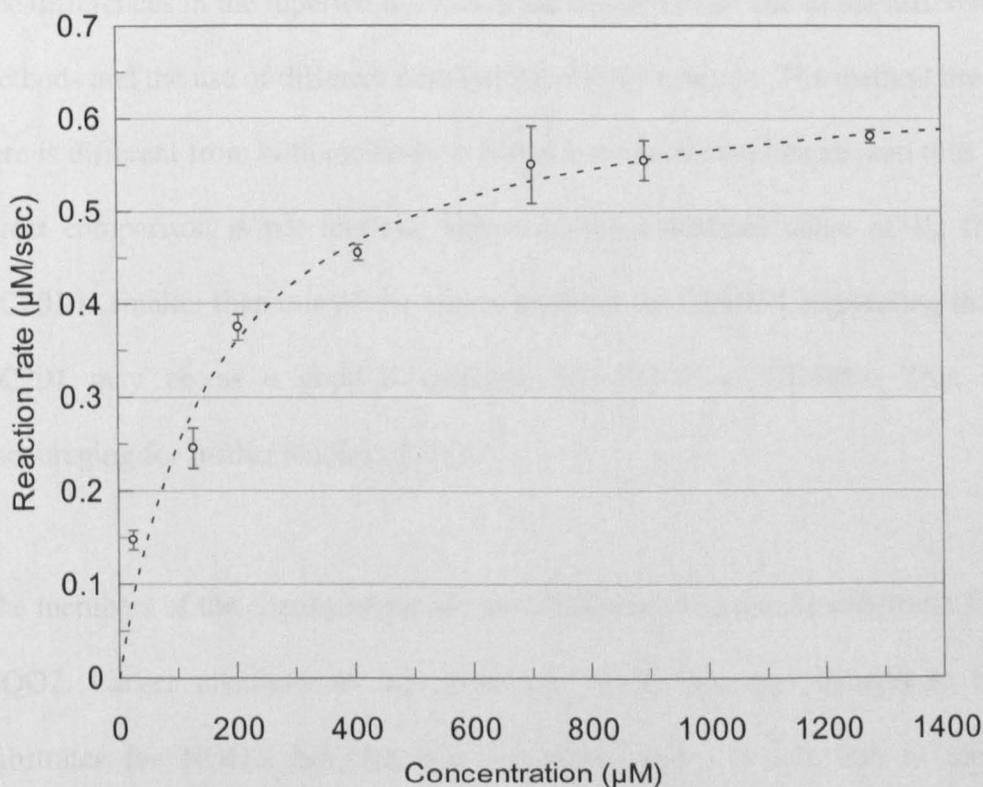


Figure 5. 20 Michealis-Menten plot for reaction of GC201 with NQO2. The absence of error bars from some points in the plotted graph is due to the small errors in the data.

5.5 Discussion and Conclusion

GC201 was found to be a substrate for NQO2, as seen by the decrease in fluorescence of NBDH when mixed with both GC201 and NQO2. GC201 was designed to be a substrate for NQO2. The K_m for GC201 was calculated to be 168 μM . CB1954 is a known anticancer prodrug that can be activated by several enzymes, including NQO2. It was found that there are two widely varied published K_m values for CB1954 from two different experimental techniques. The first was reported as 260 μM , where the method used N-ribosyl-1,4-dihydrionioctinamide (NRH) as the cosubstrate and was dependant on using HPLC to monitor the consumption of CB1954 (Wu et al., 1997). However, in another instance the K_m for CB1954 was reported to be 61 μM , where the method depended on measuring the decrease in the absorbance of an in-house cosubstrate upon its consumption in the reaction (Fu et al., 2005). The differences in the reported K_m values are thought to be due to the different methods and the use of different cosubstrates for the reaction. The method used here is different from both methods and uses a different cosubstrate, and thus a direct comparison is not feasible. However, the calculated value of K_m for GC201 is smaller than one of the values reported for CB1954, suggesting that GC201 may be as a good a substrate for NQO2 as CB1954. This is encouraging for further studies of GC201.

The members of the coenzyme family tested were all found to be substrates for NQO2. Larger members of the coenzyme family are also thought to be substrates for NQO2, but that was not tested here. This is due to their insufficient solubility, in view of the fact that coenzyme Q2 required the

addition of extra DMSO to the media to ensure its complete dissolution. Therefore it was not possible to test whether this was the case and consequently determine the K_m values. The calculated values of K_m were inconclusive and require further investigation. This is due to the use of Lineweaver-Burke plots in their calculation. The Lineweaver-Burke plot results in the uneven distribution of data points due to the absence of data points at high substrate concentrations. Moreover, there is an exaggeration of errors caused by the use of the reciprocal of the reaction rate. This consequently results in errors in the calculation of V_{max} and K_m values due to errors in the fitting of the line and the large extrapolation to calculate the intercepts.

The calculated K_m values for the coenzyme Q series show a clear trend appearing in the ability of NQO2 to turnover the coenzyme Q series. A recent study involving NQO2 and the coenzyme Q series showed a similar trend (Ferry et al.). The increase in size of the side chain is associated with a decrease in the rate of turnover of the substrate by NQO2, demonstrated by the decreasing K_m values. This may be related to the size of the active site and its ability to fit the substrate, where the bulkiness of the side group does not allow proper alignment of the substrate on top of the FAD molecule in the active site.

Chapter Six

Determination of the Crystal Structure of Reduced NQO2 I: Methods and Material

6.1 Introduction

The mechanism of action of NQO2 is dependent on the reduction of the flavin cofactor by a cosubstrate before the subsequent transfer of electrons to the substrate. Thus during catalysis the substrate binds to the 'reduced^{††}' form of the enzyme. To date all the published NQO2 structures are of the 'oxidized^{††}' form of the enzyme and the structure of the 'reduced' form of the enzyme is not known. It is important to determine if the reduction process affects the binding pocket, by changing the geometry of the FAD or the position of amino acids. Moreover, it is also an essential aid in allowing the complete understanding of the mechanism of action of the enzyme.

Several other flavoenzymes sharing the same mechanism of action with NQO2 exhibit changes in the conformation of the flavin cofactor upon its reduction. The flavin cofactor of the *Enterobacter cloacae* nitroreductase, exhibits an increase in the bend of the flavin cofactor along the N5-N10 axis of the flavin isoalloxazine ring, known as a butterfly bend (Figure 6.1), by 9° upon its reduction from an initial bend of 16° in the oxidized form. Residues that interact with the isoalloxazine ring also shift to accommodate the bending of the ring; these include Lys-14, Asn-71 and Lys-74 which interact with the

^{††} Reduced and oxidized forms: refer to the oxidation state of the flavin cofactor bound to the enzyme.

pyrimidine ring and N5 of the flavin. Ile-164, Tyr-144 and Leu-145 which interact with the dimethyl benzene moiety also show comparable movements (Haynes et al., 2002). Likewise, the crystal structure of thioredoxin reductase from *Escherichia coli* shows a much significant bend of 34° along the same axis (Lennon et al., 1999). Moreover, some flavoenzymes have shown positional changes of amino acids in the active site upon reduction of the flavin cofactor. For example, the reduction of monoamine oxidase (MAO) A causes drastic changes in the aromatic environment around the flavin cofactor (Hynson et al., 2004).

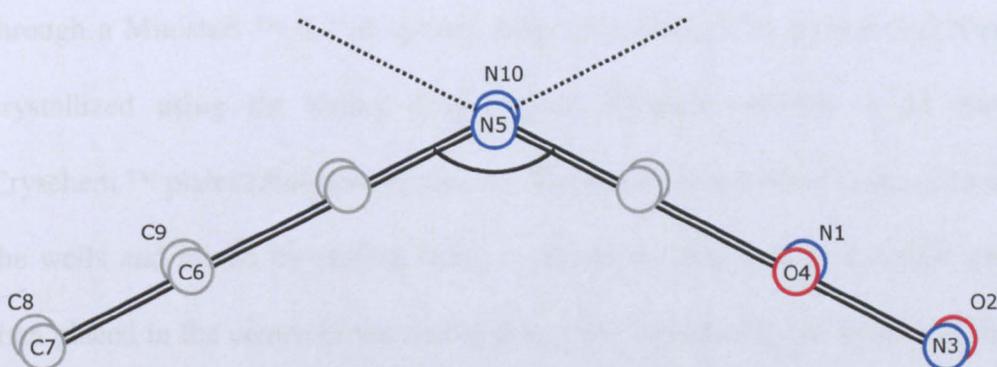


Figure 6. 1 Side view of isoalloxazine ring, showing bend along N5-N10 axis.

Hence an understanding of the changes that occur to the environment of the binding pocket upon the reduction of the FAD cofactor, such as the position of key residues in the binding pocket and any changes to the conformation of the isoalloxazine ring of the cofactor. This might provide useful information about the chemical mechanism of action of the enzyme during reductive catalysis.

6.2 Methods and Materials

All chemicals were purchased from Sigma-Aldrich unless otherwise stated.

6.2.1 Protein Production

The NQO2 protein was expressed, harvested and purified as discussed in chapter four sections 4.2.2 and 4.2.3. The purity of the protein was assessed using SDS-PAGE, described in section 4.2.4. The protein solution was concentrated to several different concentrations ranging between 15 mg ml⁻¹ and 45mg ml⁻¹. The concentration of the protein was calculated as described in chapter four, section 4.2.2.

6.2.2 Protein Crystallization

All solutions used in the crystallization protocol were filtered before use through a Ministart™ 0.2 µl syringe filter (*Sartorius*). The protein was then crystallized using the sitting drop vapour diffusion method in 24 well Cryschem™ plates (*Hampton research*). The reservoir solutions were added to the wells and mixed by stirring using a pipette tip. The protein solution was then placed in the centre of the sitting drop post, followed by an aliquot of the reservoir solution. The plates were then sealed using crystal clear sealing tape (*Hampton research*) and placed at 20°C in an incubator. The plates were then checked at regular intervals for the appearance of crystals.

The conditions for crystallization of NQO2 were adapted from several previously published methods (Table 6. 1). The precipitant used was ammonium sulphate in concentrations ranging from 1.2 to 2.0 M and the pH was adjusted using Na-HEPES (AbuKhader et al., 2005, Calamini et al., 2008, Foster et al., 1999, Fu et al., 2005). The published conditions were manually optimized to yield the best quality crystals. Several different parameters were

varied; the protein concentration was varied from 15 to 45 mg ml⁻¹, the concentration of ammonium sulphate was varied between 1.4 M and 2 M. The pH was also varied between pH 6.5 and 8. The effect of the drop size on crystallization was also investigated; this was done by changing the drop size between two µl and five µl. Finally, different ratios between the protein and reservoir solutions that were added to the crystallization drop were tested (1:1, 1:2 and 2:1). The final reservoir volume was kept constant at one ml. The concentrations of dithiothreitol (DTT) and FAD added to the reservoir solution were also kept constant at 1 mM and 12 µM respectively.

Method Conditions	Foster (Foster et al., 1999)	Calamini (Calamini et al., 2008)	Abu Khader (AbuKhader et al., 2005)	Fu (Fu et al., 2005)
Ammonium sulphate	1.4 M	1.3-1.7 M	1.66 M	1.2-2.0 M
HEPES	100 mM pH 7.0	100mM pH 6.0-7.0	100 mM pH 7.0	100 mM pH 7.0
FAD	12 µM	12 µM	12 µM	10 µM
DTT	1 mM	1 mM	1 mM	-
Other conditions	Final protein concentration of 7.5 mg/ml	1) Final protein concentration of 16 mg/ml 2) 0.1 M NaCl	Final protein concentration of 7.5 mg/ml	-

Table 6. 1 Crystallization conditions used in published method for the crystallization of NQO2.

The quality of the crystals produced was judged using several criteria. The crystals were first visually inspected to ascertain the absence of visible cracks and visible twinning as they lead to interference in the diffraction pattern produced. In addition, the size of the crystals was also considered as an

important criterion since testing has proved that the increase in size of the crystal is accompanied by an increase in diffraction resolution.

6.2.3 Crystal Reduction and Cryoprotectant Solution Optimization

In order to minimize crystal damage caused by the soaking process and the transportation of the crystals between several soaking drops, the reduction of the NQO2 crystals was carried out in the cryoprotectant solution.

Sodium dithionite was successfully used to reduce flavoenzymes in concentrations ranging from 20 mM to 100 mM (Gruber et al., 2009, Haynes et al., 2002, Lennon et al., 1999, Ludwig et al., 1997). Several different concentrations of sodium dithionite were tested, starting from 5 mM to 50 mM. A variety of cryoprotectant solutions were also tested, these included different concentrations of sodium malonate ranging from 0.8 M to 2 M, and sodium formate ranging from 2 M to 3 M in both 0.9 M and 1.8 M ammonium sulphate. In addition, sucrose solutions of concentrations up to 50 per cent were also tested. Successful cryoprotectant solutions should have a glassy appearance when placed in the nitrogen cryostream. In addition, the crystals should not be damaged by showing cracks when soaked in cryoprotectant.

The combined solution of sodium dithionite and cryoprotectant solution was first degassed by bubbling helium gas through the solution. This was done to remove any traces of oxygen, which may hinder the complete reduction of the NQO2 crystals. The crystals were harvested from the crystallization tray using CryoLoops™ (*Hampton research*) ranging in size from 0.2 mm to 0.7 mm,

depending on the size of the crystal. The harvested crystals were then soaked in the combined solution for ten minutes in the crystallization sitting drop plates to limit oxygen exposure during the reduction process. The crystals were visually monitored for complete colour bleaching, indicating complete reduction, as well as the appearance of imperfections such as cracks or rough edges, indicating crystal damage.

Crystals showing complete colour disappearance and no visual signs of damage were once again harvested using CryoLoops™ and quickly transferred to the goniometer into the nitrogen gas stream for in-house data collection or stored in liquid nitrogen for data collection at the synchrotrons.

6.2.4 Data Collection

Data was collected from the in-house X-ray source, equipped with a Rigaku R-Axis IV++ detector using CuK α radiation from a Rigaku Micromax-007 rotating anode and VariMax optics. Data was also collected from several trips to the Diamond Light Source and the European Synchrotron Radiation Facility (ESRF). The parameters used for data collection differed according to the setup used and the quality of the crystals. The number of frames was determined using initial test shoots and strategy optimization in iMosflm (Leslie, 1992).

6.2.4.1 Dataset one

This dataset was collected using the in-house X-ray source; the parameters used are listed in Table 6. 2:

Oscillation	1°
Transmission	100 %
Exposure time	5 minutes
Number of collected frames	200

Table 6. 2 Parameters used for collection of dataset 1.

6.2.4.2 Dataset two

This dataset was collected at the ESRF; the parameters used for the dataset collection are listed in Table 6. 3.

Oscillation	1°
Transmission	100 %
Exposure time	5 seconds
Number of collected frames	200
Beamline ID	ID14/1

Table 6. 3 Parameters used for collection of dataset 2.

6.2.4.3 Datasets three, four and five

These datasets were collected at Diamond Light Source; the different parameters were used for each dataset are listed in Table 6. 4.

	Dataset 3	Dataset 4	Dataset 5
Oscillation (°)	1	1	1
Transmission (%)	90	50	50
Exposure time (seconds)	1	1	1
Number of collected frames	200	120	240
Beam ID	I04	I02	I02

Table 6. 4 Parameters used for collection of datasets 3, 4 and 5.

Other datasets were collected but were discarded due to their relative low quality.

6.2.5 Data Processing

6.2.5.1 Indexing and integration of images

MOSFLM v 7.0.5 (Leslie, 1992) was used to index and integrate the images from the datasets collected. Preliminary unit cell parameters were determined using the first image collected. Subsequently, four images from two different segments 90° apart were then used to refine the unit cell and detector parameters. Once the accurate unit cell parameters were determined, all collected images were then integrated using the default parameters of MOSFLM.

6.2.5.2 Scaling and merging of images

Following the integration step in MOSFLM, SCALA (Evans, 1997) from the CCP4 programme suite (Bailey, 1994) was then used to scale and merge the *hkl* intensities. SCALA was also used to assess the quality of the data collected. The nominal resolution limit was chosen to satisfy all the criteria mentioned in Table 6. 5.

Parameter	Acceptable value
R_{merge}	$\leq 40\%$
Mean $I/\sigma I$	≥ 3
Completeness	$\geq 80\%$

Table 6. 5 Combined parameters for choice of nominal resolution limit.

SCALA is also used to set up a random set of reflections, comprising of five per cent of the data, which will be later used in the calculation of R_{free} during the refinement stage.

6.2.5.3 Molecular replacement

Molecular replacement was done using Phaser (McCoy et al., 2007), through the CCP4 programme suite. Phaser utilizes a maximum likelihood algorithm to determine the orientation and the position of the given model that would result in the closest possible match to the given observed structure factors (Evans and McCoy, 2008). Phaser has two means to measure the success of the molecular replacement, the log-likelihood gain (LLG) and Z-scores. LLG is the difference between the likelihood of a model and the likelihood calculated from the Wilson plot. The likelihood refers to the probability of the data being measured. LLG gives an indication of how well different models resulting from molecular replacement step are in agreement with the data, and thus allows for the comparison of different output models for the same dataset. The Z-score is the number of standard deviations above the mean for a particular LLG score. The RF Z-score is related to the rotational function, while the TF Z-score is related to the translational function. The acceptable values of the Z-scores should be greater than five, and ideally should be greater than eight.

The structure of the oxidized NOQ2 (pdb code: 1QR2) (Foster et al., 1999) from which the water molecules present in the structure have been removed was used as the model for the molecular replacement. This structure has no ligand present in the binding pocket which might cause changes in the position of amino acids in the binding pocket as well as the FAD molecules.

6.2.5.4 Refinement of model

Refinement of the model produced by Phaser was done using REFMAC 5 (Murshudov et al., 1997), through the CCP4 program suite. The acceptable values of R_{factor} and R_{free} depend on the resolution of the data undergoing refinement. The refinement procedure was done in several steps. The first step was a rigid body refinement which was followed by restrained body refinement step. Both R_{factor} and R_{free} were monitored during the refinement procedure. The number of cycles of refinement was chosen to give the lowest values of R_{factor} and R_{free} as well as the lowest difference between R_{factor} and R_{free} . This prevents the possibility of bias and over fitting of data as well as indicating the highest agreement between experimental data and produced model. The weighting term which controls the weight of the X-ray data versus the geometry of the protein was adjusted during the restrained body refinement so that the rms bond distance has a value between 0.01 and 0.02 Å.

COOT (Emsley and Cowtan, 2004) was used for the visual examination of the resulting model and associated maps. In addition, COOT was also used for manual model adjustment; the position of the atoms was checked to assure its agreement with the electron density maps. If the model and the electron density were not in agreement, attempts were made to fit different conformations of the amino acids into the electron density. Moreover, COOT was also used to manually add water molecules where required.

6.2.6 Identification of Unknown Electron Density in the Binding Pocket

During the refinement procedure, an unidentifiable section of electron density was found in the binding pocket on top of the FAD molecule. The densities from several different datasets were compared. This was done by overlaying the different unidentified densities and comparing their shape using the ‘transform map command using LSQ’ in COOT. The cryoprotectant solutions were considered at first as the source of the electron density. This hypothesis was disregarded as the electron densities in different crystals that have been soaked in different cryoprotectant solutions were very similar.

6.2.6.1 Automated ligand identification

The automated ligand identification function in Phenix (Adams et al.) was then used in an attempt to identify the ligand occupying the electron density. This function tries to fit the most common 200 ligands in the protein database into the electron density and then analyzes the fitting of the ligands and provides a ranked list of the best fitting ligands. The ranking of the ligands is based upon the correlation between ligand and surrounding map. This is done through a graphical user interface and is fully automated.

6.2.6.2 Refinement of probable molecules into the electron density

Molecules involved in the protein harvesting, crystallization and reduction were considered as likely sources for the ligand occupying the unidentified electron density. Thus the probable molecules were manually added to the electron density to check if they have the right geometry and size for the unidentified electron density.

Dithiothreitol (DTT) is added to the crystallization mixture to ensure the reduction of the thiols in the crystallized protein. Upon its oxidation, DTT forms a six membered ring, (Cleland, 1964). All probable molecules, including the six membered ring oxidation product of DTT, were placed into the electron density in question, and then visually examined to see if the molecule was of the appropriate size and geometry. If the molecule was found to be of an appropriate size and geometry, the protein with the molecule added into the electron density, was then subjected to one round of restrained body refinement. The R_{factor} and R_{free} were examined to see the impact of the molecule on the model quality; furthermore the resultant files were then once again visually examined to see the fit of the molecule into the electron density. The appearance of large areas of negative red peaks in the difference electron density map indicated that the molecule is bigger than that responsible for the unidentified electron density. On the other hand, the presence of large areas of green electron density after refinement indicates that the molecule is smaller than that responsible for the unidentified electron density.

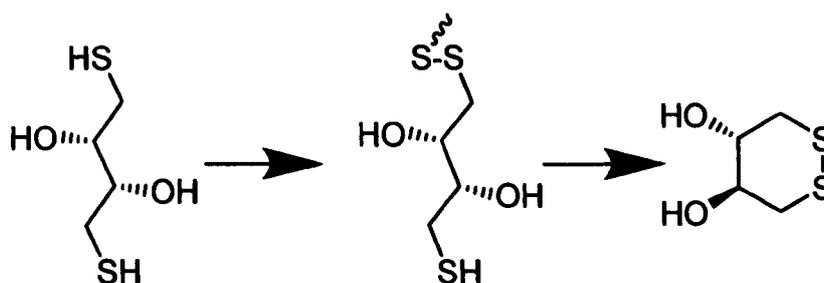


Figure 6. 2 Oxidation of DTT

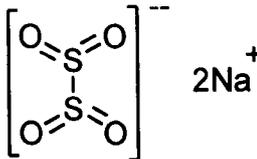
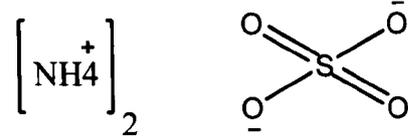
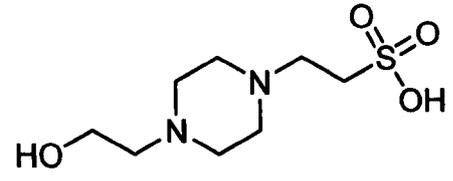
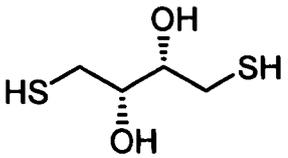
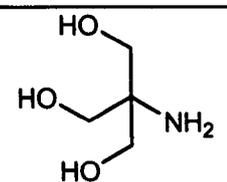
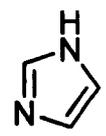
Molecule	Source	Structure
Sodium Dithionite	used for crystal reduction	
Ammonium sulphate	present in crystallization drop	
Na-HEPES	present in crystallization drop	
Dithiothreitol	present in crystallization drop	
Tris	present in crystallization drop	
Imidazole	added to the protein during purification	

Table 6. 6 Structures of molecules with which the crystals have been in contact with.

6.2.7 Model Validation

COOT has several built in validation checks that were used to assess the quality of the model, such as geometry and rotamer analysis and density fit. Any errors that were discovered were manually adjusted in COOT before the final round of the refinement process.

Molprobit (Davis et al., 2007) was used to analyze the steric clashes between the different residues; this was done through the Molprobit web server. Molprobit also analyzes the geometry of the residues. Furthermore, PROCHECK (Laskowski et al., 1993) was also used to assess the stereochemical quality of the structure, and identify any areas of the structure that have geometries with large deviations from average values. PROCHECK produces a Ramachandran plot, in addition to a short summary file of the checks that were run by PROCHECK.

Chapter Seven

Determination of the Crystal Structure of Reduced NQO2 II: Results and Discussion

7.1 Protein Production

NQO2 was expressed, harvested and purified as described in chapter four sections 4.2.2, 4.2.3 and 4.2.4. The results of the SDS-PAGE used to assess the protein purity are described in chapter five, section 5.1.1.

7.2 Protein Crystallization

One of the variables explored during the protein crystallization trials was the protein concentration, as discussed in section 6.2.2. It was found that the use of 7.5 mg ml⁻¹ NQO2 in the crystallization drop as have been previously published (AbuKhader et al., 2005, Foster et al., 1999) did not produce any crystals and it was necessary to increase the concentration of the protein to 15 mg ml⁻¹. The higher concentration of NQO2 was used by one other group in a published method, to produce NQO2 crystals of diffracting quality (Calamini et al., 2008). Trials using concentration of NQO2 higher than 15 mg ml⁻¹ resulted in the production of a large number of crystals that were too small to use to collect diffraction images. Another variable explored was the concentration of ammonium sulphate, which was used as the precipitant. It was found that concentrations ranging 1.6 M to 1.8 M produced crystals of appropriate quality to be used for X-ray diffraction. Lower concentrations of ammonium sulphate did not result in the production of NQO2 crystals. While

higher concentrations resulted in the precipitation of ammonium sulphate, rather than the crystallization of NQO2. It was also found that large deviations of pH from 7.2 caused the formation of distorted crystals. The change in drop size did not affect the process of crystal production or morphology. The final optimized conditions used for the production of NQO2 crystals are listed in Table 7. 1

Ammonium sulphate	1.6 – 1.8 M
Na-HEPES	100 mM
Sodium dithionite	1 mM
FAD	12 μ M
pH	7.2 \pm 0.2

Table 7. 1 Optimized conditions for NQO2 crystal production.

All the NQO2 crystals grown showed a strong yellow colour; this is due to the crystallization of the oxidized FAD cofactor within the NQO2 crystals. Two different and distinct morphologies of NQO2 crystals were observed; one was a very flat plate-like with straight edges and usually grew attached to the bottom of the crystallization plate. Such crystals had a thickness ranging between 0.0025 mm and 0.005 mm (Figure 7. 1, top row). On the other hand, the other crystal morphology showed a chunkier appearance, with tapered ends, with sizes that varied between 0.2 mm to 0.5 mm (Figure 7. 1, bottom row). Both these crystal morphologies showed the ability to diffract X-rays.

7.3 Crystal Reduction and the Choice of Cryoprotectant

7.3.1 Crystal Reduction

A wide range of concentrations of sodium dithionite were tested for their reducing potential. It was found that 25 mM was the optimum concentration of

sodium dithionite to reduce the NQO2 crystals. Higher concentrations of sodium dithionite caused the NQO2 crystals to crack before reduction occurred. On the other hand, lower concentrations of sodium dithionite did not reduce the crystals for sufficient time to allow the mounting of the crystals onto the goniometer or freezing in liquid nitrogen. The reduction process causes a change in colour of the crystal from bright yellow to colourless (Figure 7.2)

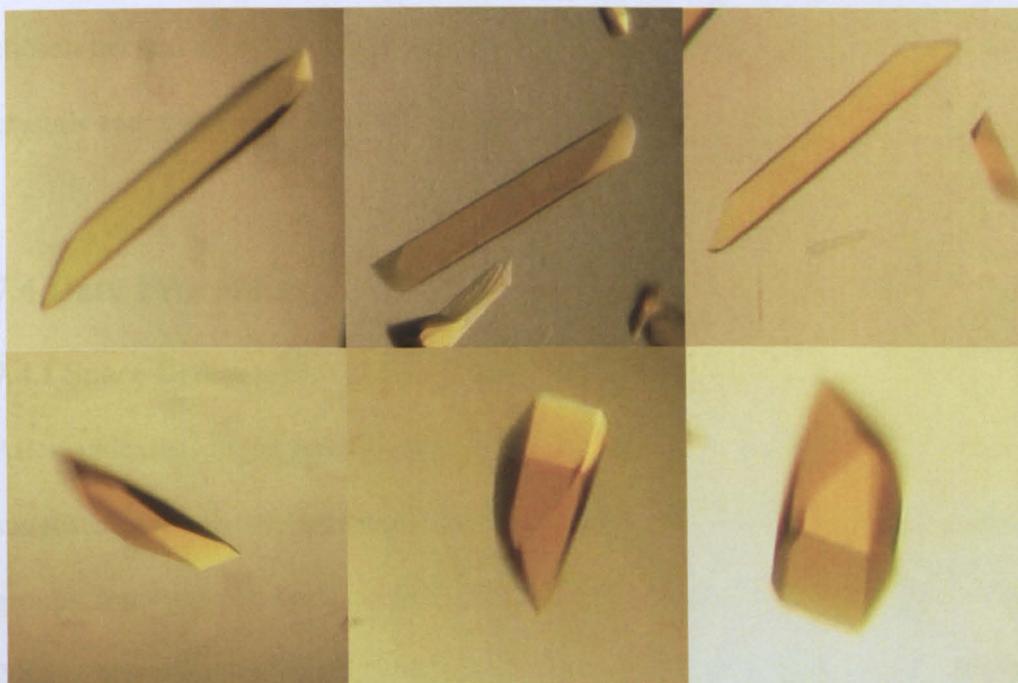


Figure 7. 1 Photographs of NQO2 crystals obtained under varying conditions. The top row shows plate-like morphology, while the bottom row shows chunkier type morphology.

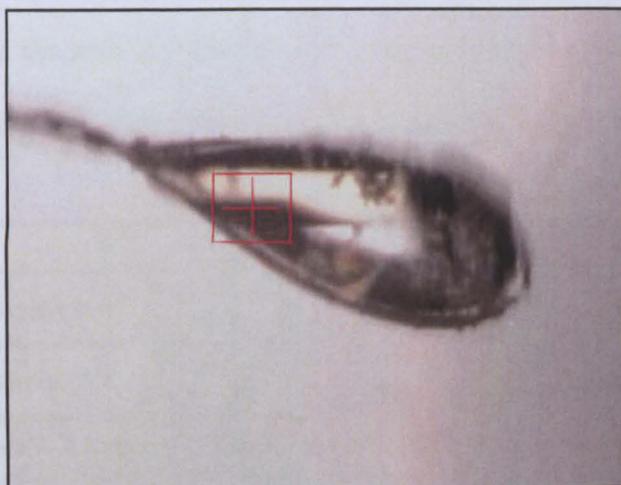


Figure 7. 2 Reduced NQO2 crystal mounted on goniometer.

7.3.2 Cryoprotection Solution Choice

As previously mentioned in section 6.2.3 the crystal reduction was carried out in the cryoprotectant solution to minimize crystal damage. Therefore, 25 mM sodium dithionite were added to the cryoprotectant solutions during testing. The different cryoprotectant solutions used for the collection of datasets were 1.2 M sodium malonate in 1.8 M ammonium sulphate, 2.8 M sodium formate in 1.8 M ammonium sulphate and 50 per cent sucrose. These solutions were chosen because they prevented the formation of ice crystals around the NQO2 crystals and at the same time did not cause the crystals to crack.

7.4 Data Processing

7.4.1 Space Group and Cell Dimension Determination

After indexing the images using MOSFLM, $P2_12_12_1$ was the space group suggested for all the processed datasets, which is consistent with the space groups reported for most published structures of NQO2 (AbuKhader et al., 2005, Calamini et al., 2008, Foster et al., 1999). During data processing, it was observed that there were two distinct unit cell dimensions (Table 7. 2). Three crystals were found to have the unit cell dimensions listed in row one, while two crystals had the unit cell dimensions listed in row two. The different unit cells were not related to the different morphologies of the crystal.

	<i>a</i>	<i>b</i>	<i>c</i>	α	β	γ
Unit cell dimensions 1	61.69 Å (0.179 Å)	80.19 Å (1.28 Å)	106.22 Å (0.257 Å)	90°	90°	90°
Unit cell dimensions 2	56.16 Å (0.742 Å)	83.30 Å (0.106 Å)	106.49 Å (0.007 Å)	90°	90°	90°

**Table 7. 2 Average cell dimensions for the different cell units.
The standard deviation for the length of the unit cell is between parentheses.**

7.4.2 Data Processing Statistics for Different Datasets

During the merging and scaling process carried out by SCALA, several data quality indicators were produced by the software. The data processing statistics for datasets one, two and three which were collected from crystals showing the chunkier morphology are listed in Table 7. 3. The data processing statistics for the datasets four and five which were collected from crystals showing the plate-like morphology are listed in Table 7. 4. Other datasets were also collected, but these were discarded upon the examination of the data processing statistics which suggested that they are of relative poorer quality. Particular interest was given to the resolution collected as well as to R_{merge} , mean I/σ and percentage completeness values.

	Dataset one	Dataset two	Dataset three
Unit cell dimension			
a, b, c (Å)	61.83, 78.75, 106.00	55.63, 83.22, 106.49	56.68, 83.37, 106.48
α, β, γ (°)	90, 90, 90	90, 90, 90	90, 90, 90
Wavelength (Å)	1.54	0.93	0.97
Resolution range (Å)	19.64-1.73 (1.83-1.73)	44.19 -1.65 (1.74-1.65)	53.24-1.6 (1.69-1.6)
R_{merge} (%)	8 (34.5)	7.8 (31.1)	5.6 (34.2)
Number of observations	393101 (44345)	495918 (70897)	261741 (38237)
Number of unique reflections	51471 (6839)	63003 (9088)	61102 (9120)
Mean I/σ	18.4 (3.0)	18.7 (6.7)	14.8 (3.6)
Completeness (%)	98.6 (91.7)	100 (100)	91.1 (94.4)
Multiplicity	7.6 (6.5)	7.9 (7.8)	4.3 (4.2)

Table 7. 3 Data processing statistics for datasets collected from crystals showing chunkier morphology. (Values between parentheses are for higher resolution shell).

	Dataset four	Dataset five
Unit cell dimension a, b, c (Å) α, β, γ (°)	61.76, 80.61, 106.50 90, 90, 90	61.49, 81.20, 106.15 90, 90, 90
Wavelength (Å)	0.97	0.97
Resolution range (Å)	80.58-1.9 (1.9-2.0)	64.50-1.95 (1.95-2.06)
R_{merge} (%)	10 (32.6)	8.2 (40)
Number of observations	173438 (24906)	180880 (26646)
Number of unique reflections	39233 (5541)	38249 (5694)
Mean I/σ	12.8 (4.2)	12.3 (3.8)
Completeness (%)	92.5 (90.8)	97.5 (99.9)
Multiplicity	4.4 (4.5)	4.7 (4.7)

Table 7. 4 Data processing statistics for datasets collected from crystals showing plate-like morphology. (Values between parentheses are for the higher resolution shell).

7.5 Molecular Replacement

The default setting for Phaser produced one solution for each dataset with high LLG and Z-score values (Table 7. 5). This is a strong indication for the success of the molecular replacement step. The solutions found by Phaser for all the datasets had the same space group, $P2_12_12_1$, which is similar to what was suggested by MOSFLM and SCALA.

	Dataset 1	Dataset 2	Dataset 3	Dataset 4	Dataset 5
Space Group	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$
Solvent content	48.20 %	50.53 %	49.26 %	51.83 %	51.82 %
LLG	3669	3553	3405	3303	3581
TF Z-score	47.1	47.2	51.4	54.3	60.0
RF Z-score	29.1	31.7	34.1	31.8	34.0

Table 7. 5 Space group, solvent content, Z-scores and LLG values produced by Phaser.

7.6 Model Examination

After one round of rigid body refinement followed by one round restrained body refinement, the resultant model and associated maps were thoroughly examined in COOT. The examination of the models and the maps for all the datasets in COOT showed no associated electron density for the two terminal residues Ala-1 and Gln-230, in both chains. Similarly, in all the datasets collected the electron density associated with residues Gly-57 to Glu-63 inclusively, in chain B is weak. This is accompanied by high B-factors for this segment when compared to neighbouring residues (Table 7. 6). This suggests that these residues may form a mobile loop, situated at the outer edge of the protein (Figure 7. 3). However, for the remaining residues in both chains there is a good fit between the electron density and the model.

Residue	Gly-57	Thr-58	Leu-59	Ser-60	Asn-61	Pro-62	Glu-63
B-factor	41.04	45.66	47.54	47.98	47.04	45.5	42.44

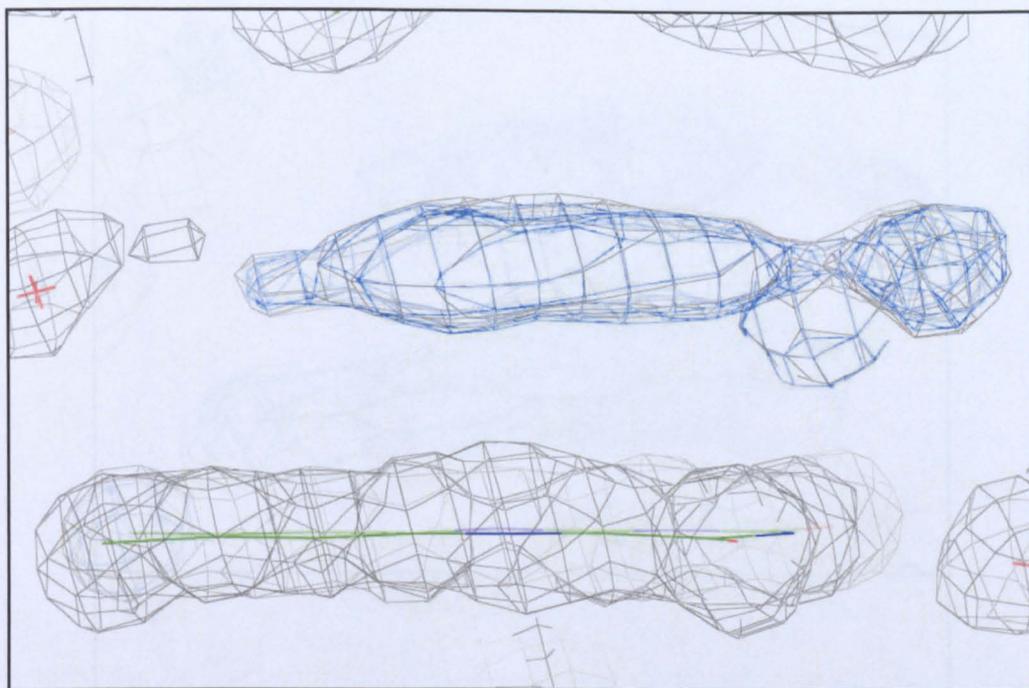
Table 7. 6 The Average B-factors of chain B residues 57 to 63 in all collected datasets.

Upon the examination of the binding pocket it was found that the isoalloxazine ring of the flavin cofactor is flat and did not show any bend along the N5-N10 axis (Figure 6. 1). Moreover, a positive density in the difference map, $F_{\text{obs}} - F_{\text{calc}}$, was present in both binding pockets in all the processed datasets. The electron density is flat and present on top of the FAD molecule. In datasets one to three the unidentified electron density is the same in both binding pockets. It lays parallel to the FAD molecule, on top of the central and pyrimidine rings (Figure 7. 4). These datasets were collected from the NQO2 crystals showing the chunkier morphology. In datasets four and five the

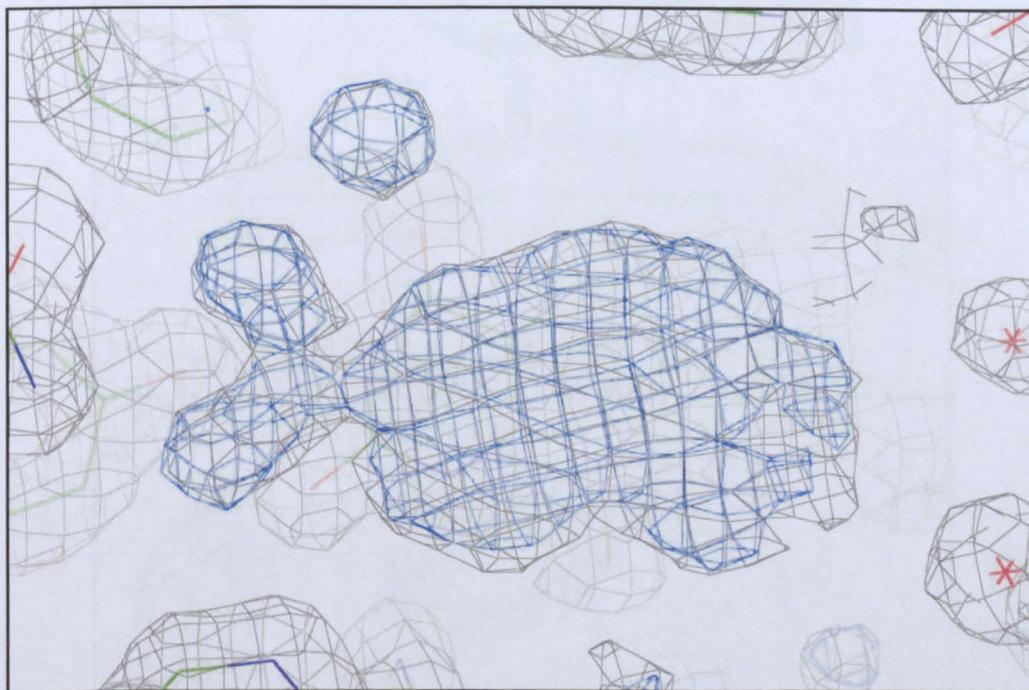
electron density showed a different pattern in the two binding pockets of the protein. In one of the binding pockets, the unidentified electron density is connected to the electron density of the N5 atom (Figure 1. 1) of the isoalloxazine ring of the FAD molecule (Figure 7. 5 and Figure 7. 6). This occurs in the same binding pocket in both data sets. Meanwhile, in the other binding pocket the unidentified electron density is parallel to the FAD molecule with no interconnecting density, similar to what was observed in the binding pockets in datasets one, two and three.



Figure 7. 3 Cartoon representation of NQO2, mobile loop (residues Gly-57 to Glu-63 inclusively) in both monomers is shown in red. Figure created using UCSF Chimera.



(a)



(b)

Figure 7.4 Active site of NQO2 showing the electron density maps of the unidentified electron density in the binding pocket of dataset two (resolution 1.65 Å), where (a) shows the side view and (b) the top view. The difference map, $F_{\text{obs}} - F_{\text{calc}}$ (shown in blue), is contoured at 3σ while the composite map, $2F_{\text{obs}} - F_{\text{calc}}$ (shown in grey) is contoured at 1σ . Red crosses represent water molecules. Figure was created using PyMOL.

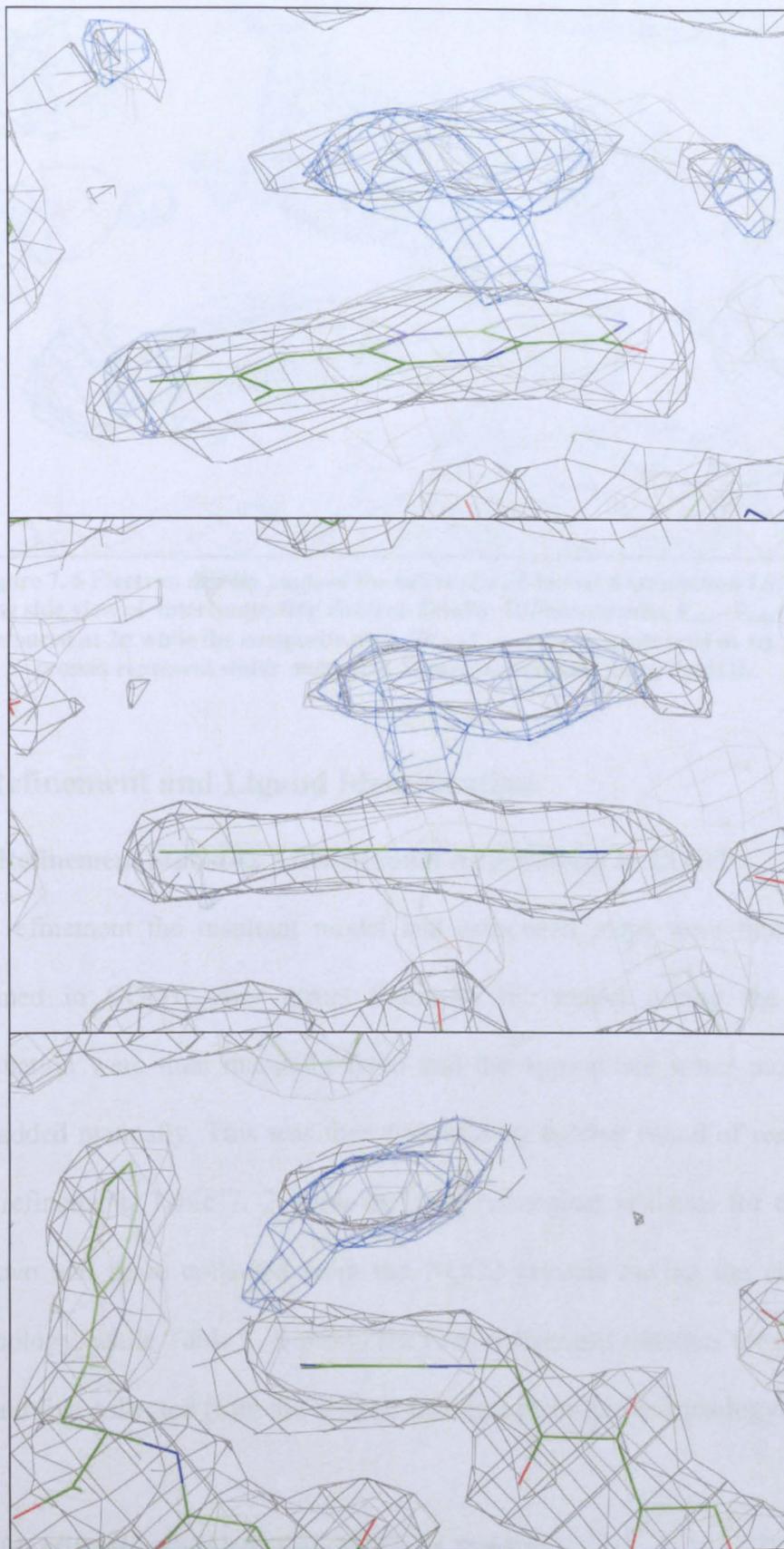


Figure 7.5 Different orientations of electron density maps in the active site of dataset 4 (resolution 1.9 Å) showing interconnecting electron density. Difference map, $F_{\text{obs}} - F_{\text{calc}}$ (blue) is contoured at 3σ while the composite map, $2F_{\text{obs}} - F_{\text{calc}}$ (grey) is contoured at 1σ . Figure was created using PyMOL.

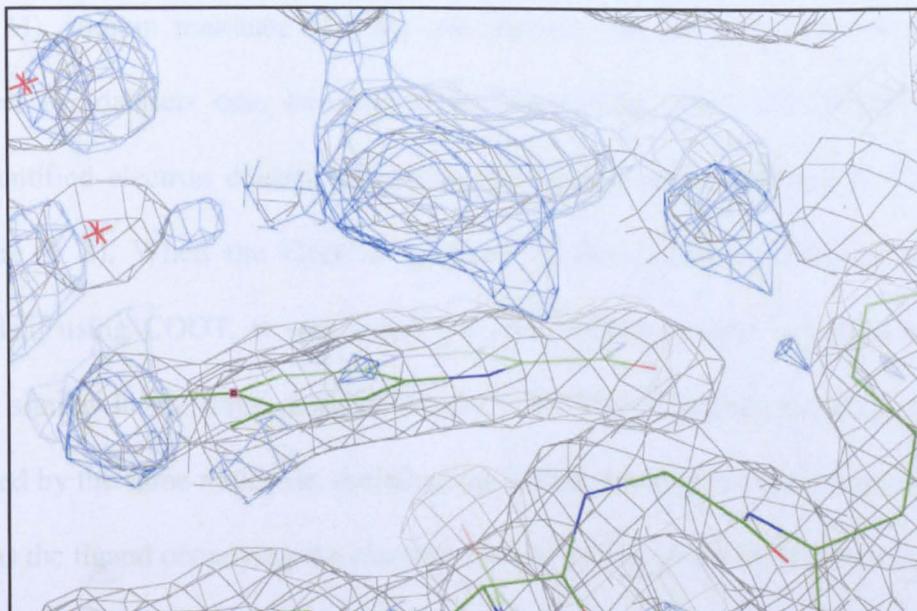


Figure 7. 6 Electron density maps of the active site of dataset 5 (resolution 1.95 Å) showing side view of interconnecting electron density. Difference map, $F_{\text{obs}} - F_{\text{calc}}$ (blue) is contoured at 2σ while the composite map, $2F_{\text{obs}} - F_{\text{calc}}$ (grey) is contoured at 1σ . Red crosses represent water molecules. Figure was created using PyMOL.

7.7 Refinement and Ligand Identification

7.7.1 Refinement Statistics After Manual Adjustments in COOT

After refinement the resultant model and associated maps were thoroughly examined in COOT. Any errors found in the model during the visual examination were then manually fixed and the appropriate water molecules were added manually. This was then followed by another round of restrained body refinement. Table 7. 7 show the final refinement statistics for datasets one, two and three collected from the NQO2 crystals having the chunkier morphology, while Table 7. 8 shows the final refinement statistics for datasets four and five collected from the crystals having a plate-like morphology.

7.7.2 Identification of Unknown Electron Density

A possible interpretation for the unidentified electron density was thought to be the cryoprotectant solution. Different crystals soaked in sodium formate

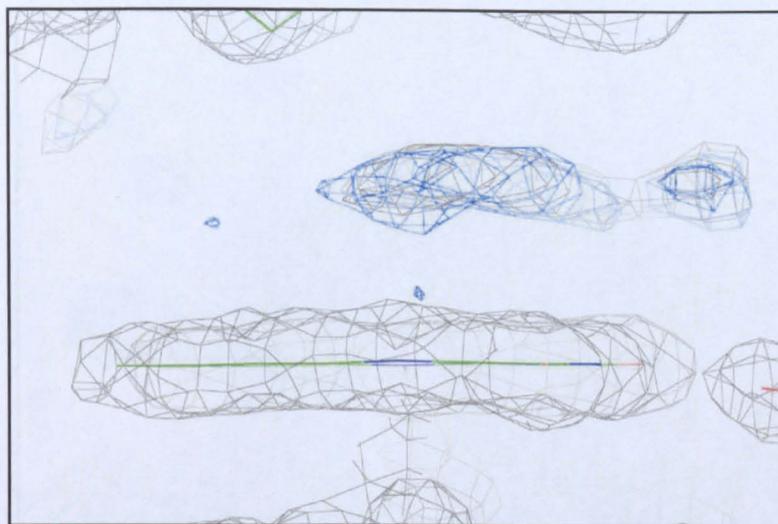
(2.8 M), sodium malonate (1.2 M) and sucrose (50 per cent) were used to collect for datasets one, two and three respectively. They all showed the unidentified electron density present in the binding pocket (Figure 7. 7 and Figure 7. 8). When the electron densities of these different datasets were overlaid using COOT, it was found the unidentified electron densities were very similar in the three datasets (Figure 7. 9). This suggests that they were caused by the same molecule, therefore the different cryoprotectants were ruled out as the ligand occupying the electron density on top of the FAD molecule

	Dataset 1	Dataset 2	Dataset 3
Resolution (Å)	19.64-1.73	44.19-1.65	53.24-1.6
R_{factor}/R_{free}	25.01/20.71	20.90/19.12	22.91/20.35
No of water moelcules	140	183	179
Average <i>B</i> -factors			
All atoms	25.490	12.230	19.839
Protein	25.227	15.976	19.309
Water	23.870	16.278	23.675
Rms Deviations			
Bond length (Å)	0.0159	0.0103	0.0115
Bond Angles (°)	1.554	1.535	1.394

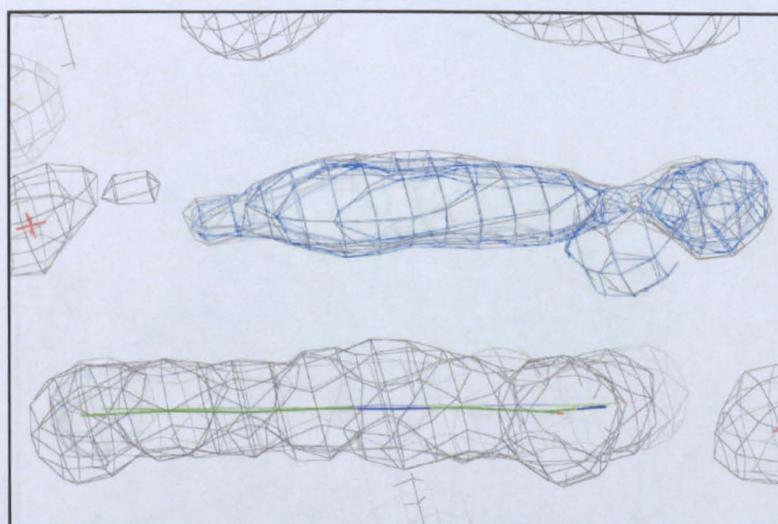
Table 7. 7 Refinement statistics for data sets of crystals having chunkier morphology.

	Dataset 4	Dataset 5
Resolution (Å)	80.58-1.9	64.50-1.95
R_{factor}/R_{free}	25.28/20.81	21.96/18.80
No. of water molecules	166	152
Average <i>B</i> -factors		
All atoms	20.365	21.740
Protein	21.333	21.391
Water	20.025	22.502
Rms Deviations		
Bond length (Å)	0.0175	0.0152
Bond angles (°)	1.676	1.492

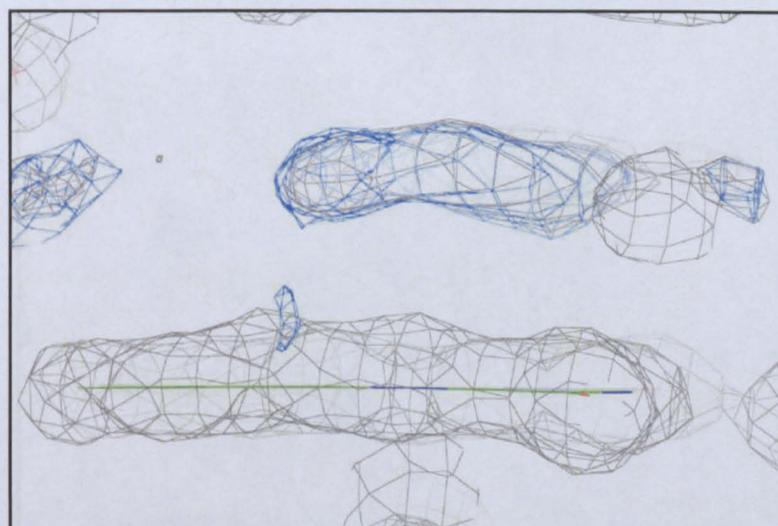
Table 7. 8 Refinement statistics for datasets of crystals having plate like morphology.



(a)



(b)

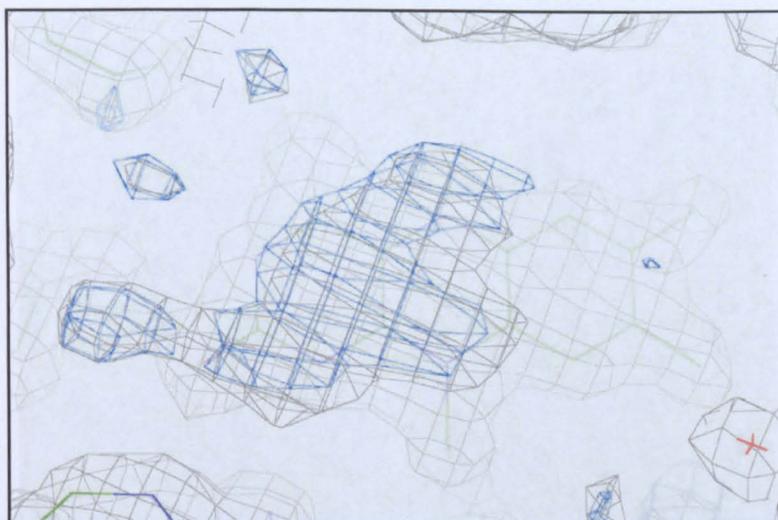


(c)

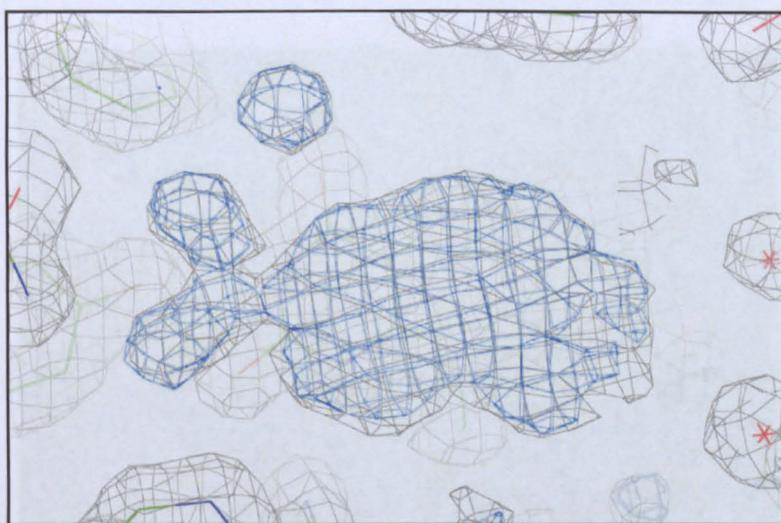
Figure 7.7 Side view of unidentified electron density, where the difference map, $F_{\text{obs}} - F_{\text{calc}}$ (blue), is contoured at 3σ while the composite map, $2F_{\text{obs}} - F_{\text{calc}}$ (grey) is contoured at 1σ .

(a) Shows dataset 1, soaked in 2.8 M Na formate, (b) shows dataset 2, soaked in 1.2 M Na malonate, while (c) shows dataset 3, soaked in 50 % sucrose.

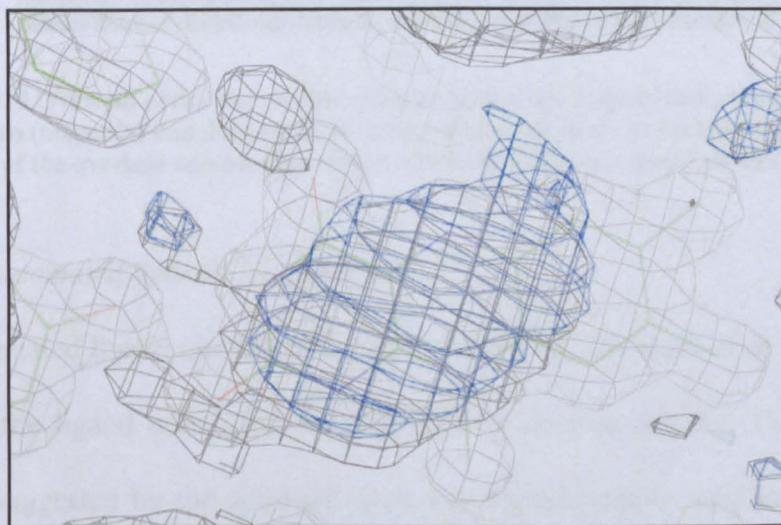
Figure was created using PyMOL.



(a)



(b)



(c)

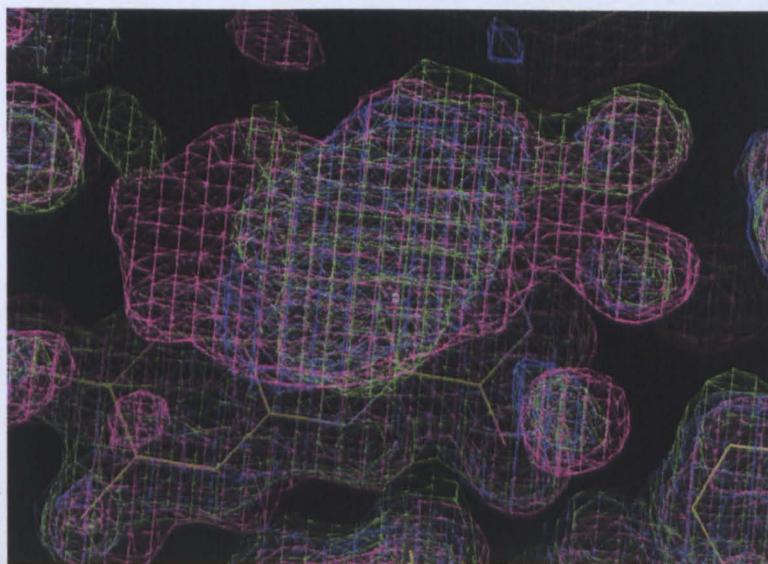
Figure 7.8 Top view of unidentified electron density, where the difference map, $F_{\text{obs}} - F_{\text{calc}}$ (blue), is contoured at 3σ while the composite map, $2F_{\text{obs}} - F_{\text{calc}}$ (grey) is contoured at 1σ .

(a) Shows dataset 1, soaked in 2.8 M Na formate, (b) shows dataset 2, soaked in 1.2 M Na malonate, while (c) shows dataset 3, soaked in 50 % sucrose.

Figure was created using PyMOL



(a)



(b)

Figure 7.9 Overlaid active site electron density difference maps of dataset one (green), dataset two (magenta) and dataset three (blue), where (a) shows the side view and (b) top view of the overlaid unidentified electron densities. Figure created using COOT.

7.7.2.1 Automated ligand identification

The automated ligand identification utility in Phenix was used in an attempt to identify the ligand occupying the unidentified electron density. The top two ligands suggested by the software were 4-hydroxycinnamic acid and palmitic acid (Figure 7. 10). Both ligands had the same scores and thus shared the same probability to be the ligand. The structure of palmitic acid, a long aliphatic chain did not fit with the shape of the density and thus was excluded as a

possible ligand. On the other hand, 4-hydroxycinnamic acid has the right general structure shape, but it was ruled out as the probable ligand as there was no source of 4-hydroxycinnamic acid in the preparation and harvesting of NQO2 enzyme, the crystallization conditions or the preparation of the crystals for data collection.

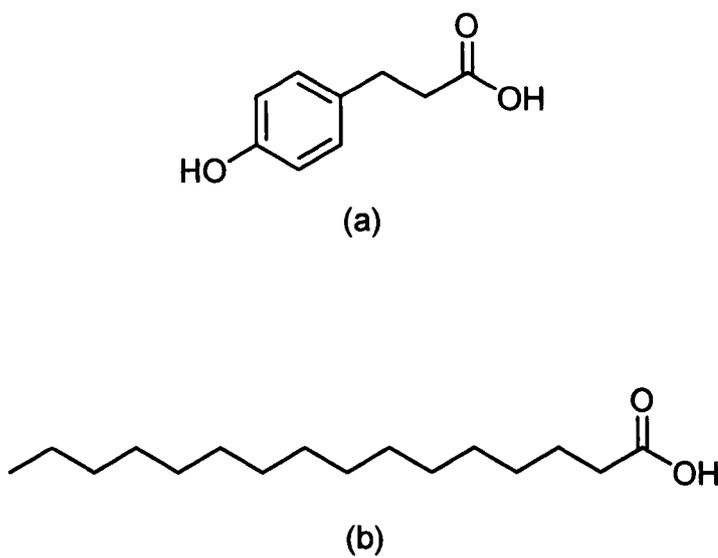


Figure 7. 10 Structures of 4- hydroxycinnamic acid (a) and palmitic acid (b)

7.7.2.2 Refinement of probable molecules into the electron density

A number of molecules that were thought to be possible candidates as the unidentified ligand were refined into the electron density in the binding pocket. It was observed that when sulphur atoms were refined into the electron density, negative peaks appear both above and below the electron density suggesting that the sulphur atom is larger than the atoms of the ligand occupying the unidentified electron density peak. This observation therefore ruled out dithioreitol (Figure 7. 11 (a)) and its reduction product (Figure 7. 11 (b)), in addition to sodium dithionite (Figure 7. 11 (c)), as the source of the unidentified electron density. Ammonium and sulphate ions, from ammonium

sulphate, and imidazole (Figure 7. 11 (d)) were ruled out as the causing molecules for the electron density due to their small size relative to the electron density. Sodium HEPES was ruled out due to the presence of the long side chain which could not be accommodated into the electron density (Figure 7. 11 (e)). Tris was also ruled out due to its geometry, as the shape of the electron density suggested a ring shaped molecule (Figure 7. 11 (f)).

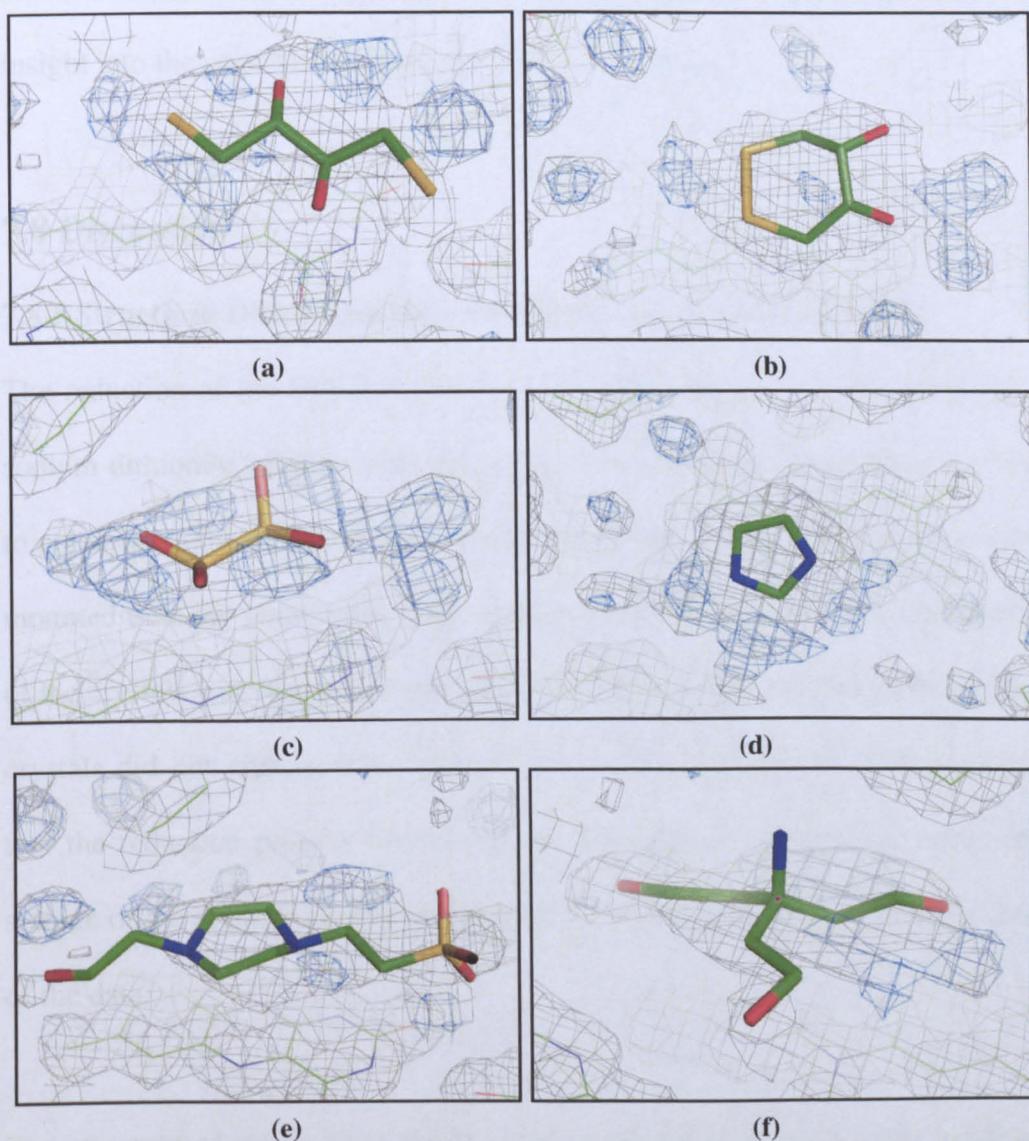


Figure 7. 11 Different molecules refined into the unidentified electron density of dataset 2. (a) is dithioreitol, (b) is the dithioreitol reduction product, (c) is dithionite, (d) is imidazole, (e) is HEPES and (f) is Tris. The blue electron density is the difference map $F_{\text{obs}} - F_{\text{calc}}$, contoured at 3σ , while grey electron density is the composite map, $2F_{\text{obs}} - F_{\text{calc}}$, contoured at 1σ . Figure created using PyMOL.

The R_{factor} and R_{free} values did not show any significant changes when the different ligands were refined into the electron density.

7.7.2.3 Analytical work

Other analytical attempts were done to identify the ligand in the electron density; these included small molecule and protein mass spectrometry, as well as NMR. These results are not mentioned here as they did not provide any insight into the possible identity of the unknown ligand.

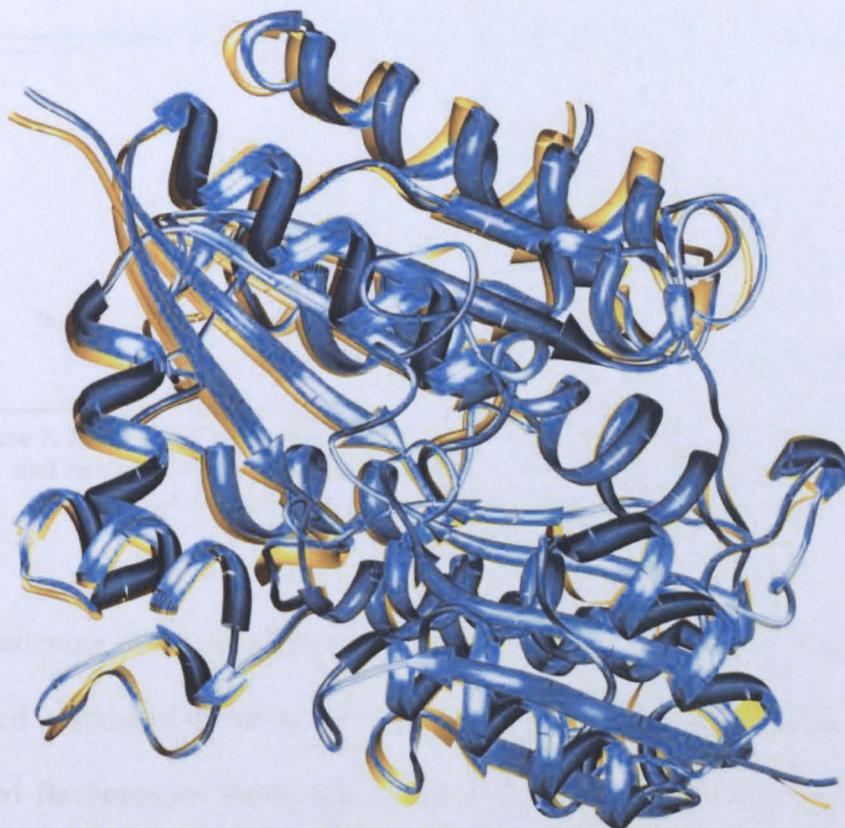
7.8 Discussion

7.8.1 Structural Differences Between Oxidized and Reduced NQO2

The reduction of the NQO2 crystals was done by soaking of the crystals in a sodium dithionite solution until the crystal colour changes from bright yellow to colourless. The colour of the crystals was examined when the crystals were mounted onto the goniometer, prior to data collection. Furthermore, the colour of the crystals was examined once again after data collection. The colour of the crystals did not change once again from colourless to yellow. This suggests that the reduction process was successful and was not affected by either the storage of the crystals or their bombardment with X-rays during the collection of the data.

The structure of the reduced NQO2 was compared to the published oxidized structure of NQO2 (pdb code: 1QR2). Special consideration was given to the residues of the active site around the FAD molecule, as well as to the geometry of the FAD. When the two whole structures were overlaid, the RMSD between

the two structures was calculated to be 0.388 Å using all the atoms of the protein. This suggests that there are no large movements in the position of the chains upon its reduction (Figure 7.12).



**Figure 7. 12 Overlaid cartoon representation of the structures of oxidized NQO2 (pdb code: 1QR2, shown in yellow) and reduced NQO2 (shown in blue).
Figure created using USCF chimera.**

The binding pockets of the two protein molecules were then thoroughly compared to determine if there are any changes in the position of the amino acids surrounding the active site. The geometry of the FAD isoalloxazine ring was also examined to determine if any conformational changes occurred upon its reduction. It was found that the reduction of the FAD molecule caused the flattening of the isoalloxazine ring along the N5-N10 axis (Figure 6. 1). The angle was found to decrease by 6°, from 173° to 179° at N10 and by only 2° at N5, from 176° to 178 (Figure 7. 13). The atomic numbering of the isoalloxazine ring of the flavin cofactor is illustrated in Figure 1. 1.

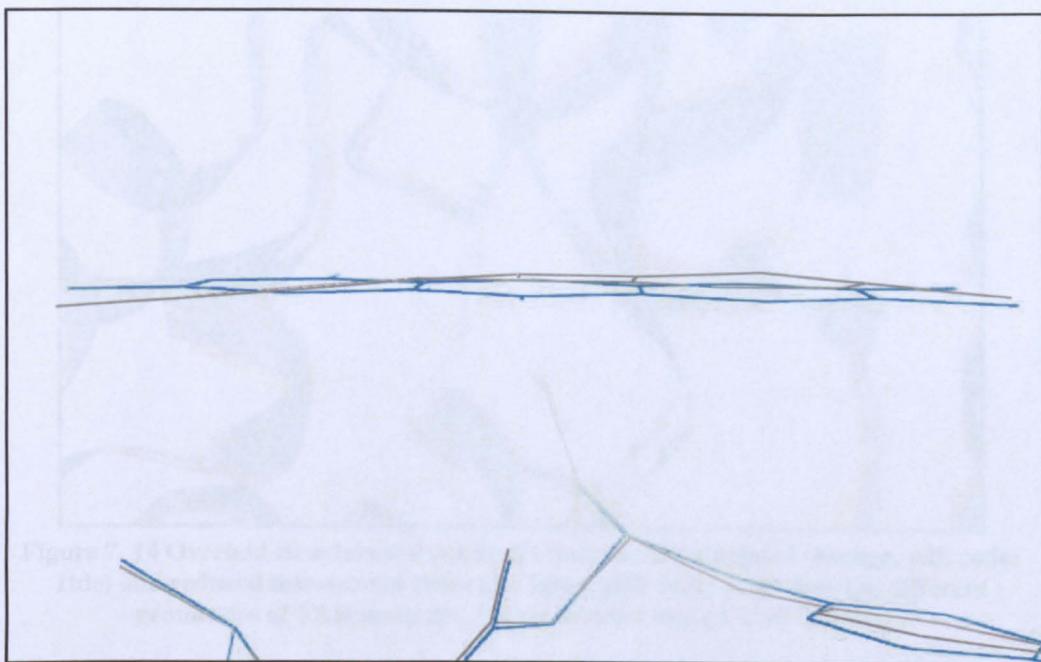


Figure 7. 13 Overlaid FAD molecules of the oxidized NQO2 (pdb code: 1QR2, grey) and reduced NQO2 (teal), showing the relative bending of the FAD in the oxidized structure compared to that of the reduced structure.

Figure created using PyMOL.

The flattening of the isoalloxazine ring is contrary to what was expected and reported in some of the other flavoproteins. Several published structures of the reduced flavoproteins showed an increase in the bend of the flavin cofactor when compared to their respective oxidized structure. Thioredoxin reductase shows an increase in the isoalloxazine bend by 34° upon the reduction of the cofactor (Lennon et al., 1999) (Figure 7. 14). While old yellow enzyme shows an increase of the flavin bend by 13° (Fox and Karplus, 1994), and *Enterobacter cloacae* nitroreductase which shows an increase in bend by 9° upon its reduction (Haynes et al., 2002). The bending of the isoalloxazine ring, results in the movement of the N5 atom of the flavin towards the binding pocket of the enzyme, and consequentially closer to the substrate. This is thought to facilitate the movement of electrons between the flavin cofactor and the substrate.

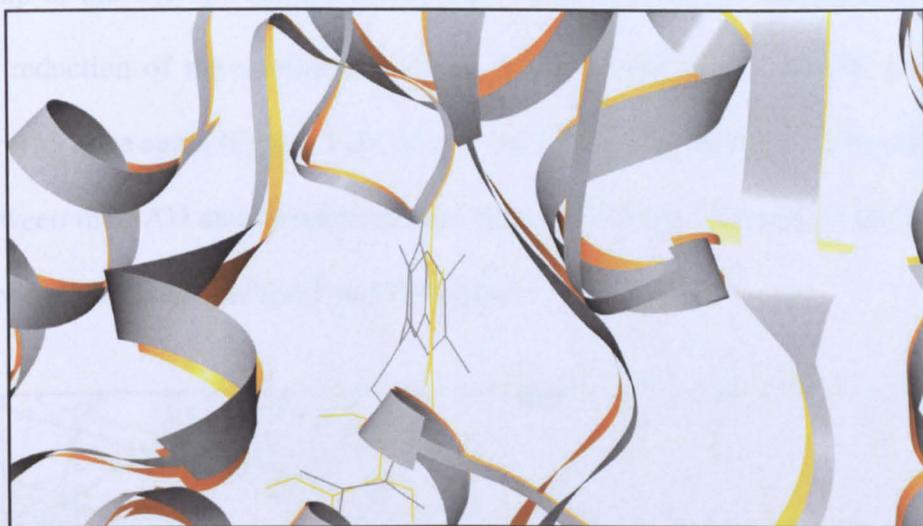


Figure 7. 14 Overlaid structures of oxidized thioredoxin reductase† (orange, pdb code: 1tde) and reduced thioredoxin reductase (grey, pdb code: 1cl0) showing different geometries of FAD molecule. Figure created using UCSF Chimera.

It is also noticed that there is a hydrogen bond between O2 of the flavin (Figure 1. 1) and the hydroxyl group of the Tyr-155 residue in dataset five (Figure 7. 15). The distance between the two atoms was measured to be 2.66 Å in this structure, which is a suitable distance for the formation of a hydrogen bond. The distance between the oxygen of the hydroxyl group of Tyr-155 and O2 of the FAD molecule in the published oxidized structure of NQO2 was measured to be 2.79 Å, which is 0.13 Å more than that measured in the reduced structure (Figure 7. 16). The difference in the distance between the flavin and the Tyr-155 residue between the oxidized and reduced proteins is approximately five per cent of the distance between the flavin and the tyrosine residue. This implies either that there is a slight movement of the tyrosine residue upon the reduction of the FAD, or that the flattening of the FAD molecule caused the movement of the O2 atom closer to the Tyr-155 residue. The Tyr-155 residue is conserved in both NQO1 and NQO2; it is thought to play a role in the electron transport mechanism of NQO1 by stabilizing the FAD molecule. This is through the transfer of a proton from the hydroxyl

group of the Tyr-155 residue to the FAD molecule after its reduction. During the reduction of the substrate, a proton is then transferred from the FAD to Tyr-155 once again (Figure 1. 11) (Li et al., 1995). The decrease in the distance between the FAD and tyrosine residue may facilitate the transfer of the proton between the flavin and the Tyr-155 residue.

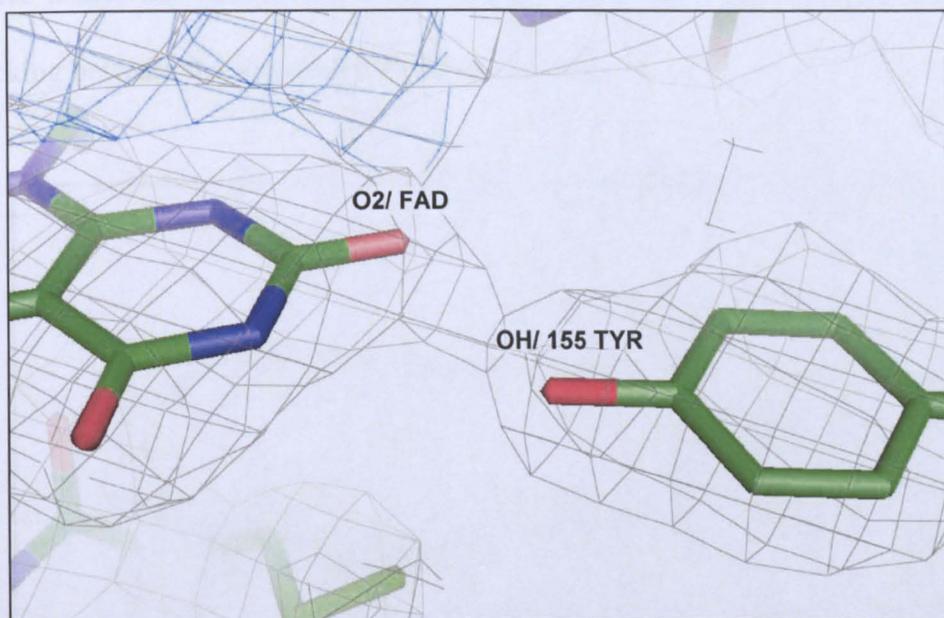


Figure 7. 15 Hydrogen bond between hydroxyl group of Tyr-155 residue and O2 atom of FAD molecule. The difference map $F_{\text{obs}} - F_{\text{calc}}$ (blue) is contoured at 3σ and the composite map, $2F_{\text{obs}} - F_{\text{calc}}$ (grey) is contoured at 1σ . Figure created using PyMOL.

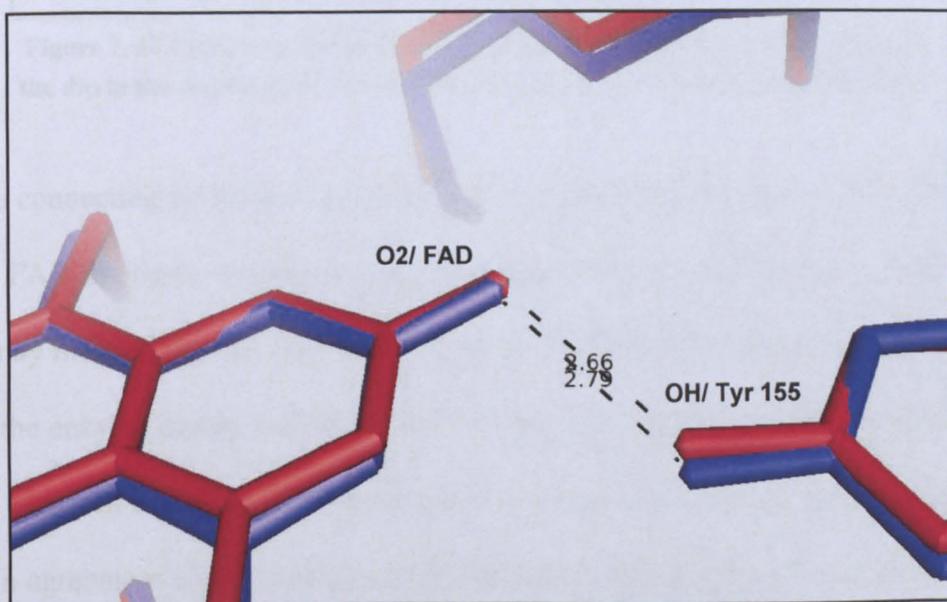


Figure 7. 16 Overlaid structures of the oxidized NQO2 (pdb code:1QR2, blue) and reduced NQO2 (red) overlaid, showing the distance between hydroxyl group of Tyr-155 and O2 oxygen of the FAD cofactor. Figure created using PyMOL.

7.8.2 Unidentified Electron Density

The unidentified electron density above the flavin cofactor is thought to be caused by an aromatic ring structure due to the dip in the middle of the electron density, where the electron density is narrower in the middle than on the edges (Figure 7. 17). Moreover, the orientation of the electron density above the FAD where it lays flat and parallel to the isoalloxazine ring suggests possible π - π stacking between the ligand and the flavin cofactor.

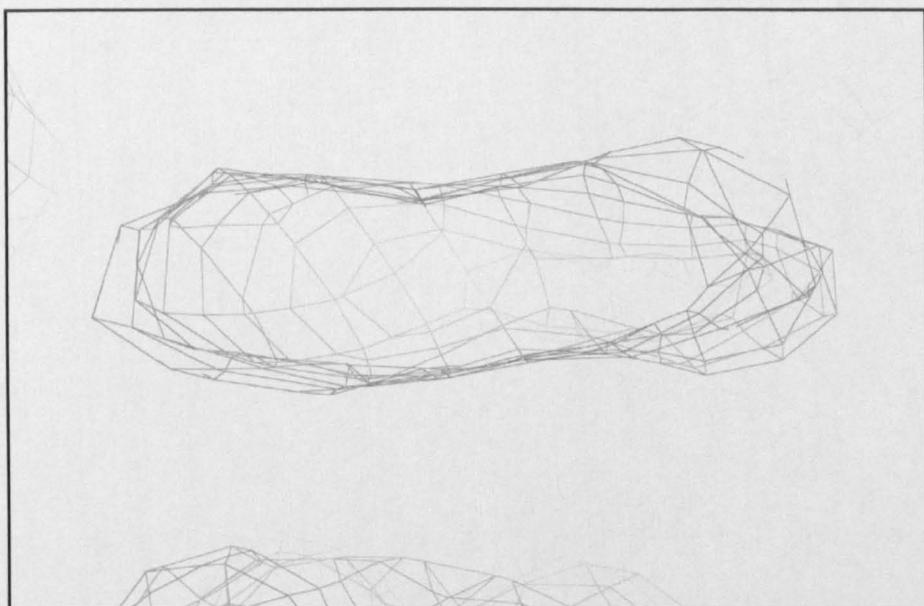


Figure 7. 17 Composite map ($2F_{\text{obs}}-F_{\text{calc}}$) of dataset 2, contoured at 1σ showing the dip in the middle of the electron density. Figure was created using PyMOL.

The connecting electron density between the unidentified ligand and the N5 of the FAD molecule is a noteworthy observation and its significance is unclear. It may indicate that the ligand is a substrate and that it is undergoing reduction by the enzyme during collection of the dataset. The connecting density may be the electrons during the transfer process from the cofactor to the substrate. This is in agreement with the proposed mechanism of electron transfer between the flavin cofactor and the substrate, which suggests that the electron transfer

occurs between the N5 of the flavin cofactor to the substrate (Blankenhorn, 1976). It may also indicate that the ligand is a suicide ligand that binds to the flavin cofactor irreversibly during the reduction process and the connecting density may then be a covalent bond formed between the substrate and the cofactor.

Chapter Eight

Conclusions and Future work

The increased interest in nitroaromatic reducing enzymes can be attributed to their ability to activate bioreductive anticancer drugs. These enzymes are of varied origins, as well as having different biological functions. The bacterial nitroreductases have been well studied and characterized and it has been demonstrated that they share several structural similarities. One of the similarities observed in bacterial nitroreductases is the presence of several conserved aromatic residues within the active site, in the vicinity of the flavin cofactor. These conserved aromatic residues are thought to form interactions with the substrates in the binding pocket and play a role in determining the substrate specificity of the different enzymes (Dolores Roldan et al., 2008). It was also reported that in other nitroaromatic reducing enzymes, such as old yellow enzyme and NQOs, aromatic residues in the active site play different roles in their mechanisms of action (Deller et al., 2008, Brown et al., 1998, Kohli and Massey, 1998).

Although bacterial nitroreductases have been well characterized, other enzymes sharing the same ability have not been as extensively studied. Bioinformatics studies carried in this work on a wide range of enzymes reducing nitroaromatic compounds have shown that they may be divided according to their general structural conformation and the orientation of residues around the active site into four distinct groups. It has been observed that all four groups had different aromatic residues conserved in their active

sites at different positions relative to the flavin cofactor. It is thought that these residues may be involved in the electron transport mechanism from and to the cofactor. They may also be involved in the correct orientation of the cosubstrate and/or the substrate in the binding pocket and consequently may play a role in the cosubstrate or substrate specificity. The identification of the conserved residues may aid in the determination of the different mechanisms of action of the different classes of enzymes that are able to reduce nitroaromatic compounds. Subsequently, this may aid in designing more specific and effective drugs for the different enzymes. Furthermore, it may also aid in the identification of new enzymes that are capable of reducing nitroaromatic compounds and as a result their use in bioreductive anticancer chemotherapy as part of either the ADEPT or GDEPT approaches.

Site directed mutagenesis can be used to determine the exact role of conserved aromatic residues in the four identified groups of nitroaromatic reducing enzymes. For example, enzyme kinetics can be used to determine the effect of the substitution of the aromatic residues on the rate of turnover of various cosubstrates as well as substrates. In addition, calculating the disassociation constant can be used to determine the change of the affinity of the substrates to the enzyme upon the substitution of the aromatic residue. Another approach that can be applied to determine the effect of the substitution of the aromatic residues is the use of molecular modelling software to simulate the reaction between substrates and the mutated enzymes. Similarly, molecular modelling software can be used to predict the possible interactions between the substrates and the mutated enzymes. Moreover, the effect of the different type of aromatic

residues can be also investigated in a similar fashion by replacing the aromatic residue of interest with another aromatic residue.

Previously published results have shown that coenzyme Q0 is a substrate for NQO2 (Boutin et al., 2005). Recently published results have also shown that coenzyme Q1 and coenzyme Q2 are also substrates for NQO2 (Ferry et al.). Kinetic studies carried out in this work on NQO2 also showed that both coenzyme Q1 and coenzyme Q2 are NQO2 substrates. Further testing of other coenzyme Q family members was not possible due to their low aqueous solubility. It was observed that the rate of turn over of the tested coenzyme Q family members is inversely proportional to the length of the aliphatic side chain. This was also reported by the recent study carried out on NQO2 and coenzyme Q series (Ferry et al.). It is thought that the increase in the size of the side chain may affect the binding and correct orientation of the coenzyme Q substrates in the binding pocket of NQO2. In addition, GC201 was also identified as a novel substrate for NQO2. It was found to have an appreciable rate of turn over to CB1954. More studies are needed to determine the potential medicinal use of GC201 as an anti-cancer therapeutic agent upon its reduction.

Several published structures of reduced flavoprotein showed changes in the conformation of the flavin cofactor when compared to their respective oxidized structures. *Enterobacter cloacae* nitroreductase and thioredoxin reductase showed an increase in the bend of the isoalloxazine ring of the flavin cofactor upon its reduction (Haynes et al., 2002, Lennon et al., 1999). The structure of the reduced NQO2 that was determined in this work did not reveal any major

conformational changes when compared to the published structure of the oxidized NQO2. Contrary to what was expected, a slight flattening of the FAD cofactor was observed upon its reduction. Moreover, it was noticed that when compared to the published oxidized structure, the distance between an oxygen atom from the FAD cofactor and the Tyr-155 residue is shorter. This is thought to facilitate the formation of a hydrogen bond between the oxygen of the FAD and the hydroxyl group of the Tyr-155 residue. This observation is in agreement with the postulated mechanism of action of NQO2 (Bianchet et al., 2004). The examination of the catalytic site of the reduced NQO2 structure revealed an electron density connecting the N5 nitrogen of the FAD cofactor with a potential substrate for NQO2. It is thought that the N5 of the FAD cofactor plays an important role in the electron transfer.

Further investigation needs to be undertaken to determine whether the very thin morphology of the plate-like NQO2 crystals is of significant importance in changing the behaviour of the ligand in the binding pocket. Given that it may be a useful aid in providing a better insight into the mechanism of electron transfer between NQO2 and its substrates. This could be done by using X-ray crystallography to determine the structure of several NQO2 crystals, showing plate-like morphology which are soaked in different known cosubstrates and substrates. The binding pocket of the determined structures could then be examined to see if the interconnecting density between the different substrates and the flavin cofactor is present. Similarly, it will be of interest to test out whether the same results could be obtained from other protein crystals showing

the same morphology soaked in known substrates. This could provide us with a valuable tool in determining the mechanisms of action of different enzymes.

The better understanding of the mechanism of action of NQO2 and the determination of the structure of its reduced form may aid in designing and the identification of novel more efficient substrates that can be used in the treatment of cancer. The use of sequence and structural alignments may result in the identification of novel enzymes that are capable of reducing anticancer prodrugs more efficiently or less likely to cause an immunogenic response once delivered into the human body. These can then be used in the activation of anticancer prodrugs by means of either ADEPT or GDEPT.

Bibliography

EMBL-EBI. European Bioinformatics Institute.

ISI Web of Knowledge. Thomson Reuters.

MolProbity web server. The Richardson lab.

RCSB Protein Data bank.

SYBYL 8.0. 8.0 ed. 1699 South Hanley Road, St. Louis, Missouri, 63144 USA, Tripos International.

ABUKHADER, M., HEAP, J., DE MATTEIS, C., KELLAM, B., DOUGHTY, S. W., MINTON, N. & PAOLI, M. (2005) Binding of the anticancer prodrug CB1954 to the activating enzyme NQO2 revealed by the crystal structure of their complex. *Journal of Medicinal Chemistry*, 48, 7714-7719.

ADAMS, P. D., AFONINE, P. V., BUNKOCZI, G., CHEN, V. B., DAVIS, I. W., ECHOLS, N., HEADD, J. J., HUNG, L. W., KAPRAL, G. J., GROSSE-KUNSTLEVE, R. W., MCCOY, A. J., MORIARTY, N. W., OEFFNER, R., READ, R. J., RICHARDSON, D. C., RICHARDSON, J. S., TERWILLIGER, T. C. & ZWART, P. H. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallographica Section D-Biological Crystallography*, 66, 213-221.

ALTSCHUL, S. F. (1991) Amino acid Substitution Matrices from an Information Theoretic Perspective. *Journal of Molecular Biology*, 219, 555-565.

ANLEZARK, G. M., VAUGHAN, T., FASHOLA-STONE, E., MICHAEL, N. P., MURDOCH, H., SIMS, M. A., STUBBS, S., WIGLEYS, S. & MINTON, N. P. (2002) *Bacillus amyloliquefaciens* orthologue of *Bacillus subtilis* ywvO encodes a nitroreductase enzyme which activates the prodrug CB 1954. *Microbiology-Sgm*, 148, 297-306.

ARNOLD, K., BORDOLI, L., KOPP, J. & SCHWEDE, T. (2006) The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling. *Bioinformatics*, 22, 195-201.

AUTHIER, A. (1981) *The Reciprocal lattice*, The International Union of Crystallography.

BAILEY, S. (1994) The CCP4 Suite - Programs for Protein Crystallography. *Acta Crystallographica Section D-Biological Crystallography*, 50, 760-763.

BARAK, Y., THORNE, S. H., ACKERLEY, D. F., LYNCH, S. V., CONTAG, C. H. & MATIN, A. (2006) New enzyme for reductive cancer chemotherapy, YieF, and its improvement by directed evolution. *Molecular Cancer Therapeutics*, 5, 97-103.

BASEL, B. U. Swiss-Model. Swiss institute of Bioinformatics.

BIANCHET, M. A., FAIG, M. & AMZEL, L. M. (2004) Structure and mechanism of NAD[P]H : Quinone acceptor oxidoreductases (NQO). *Quinones and Quinone Enzymes, Pt B*, 382, 144-174.

BISSWANGER, H. (2004) *Practical Enzymology*, Wiley-VCH.

BISSWANGER, H. (2008) *Enzyme Kinetics: Principles and Methods*, Wiley-VCH.

BLANKENHORN, G. (1976) Nicotinamide-Dependent One-Eelectron and 2-Electron (Flavin) Oxidoreduction - Thermodynamics, Kinetics and Mechanism. *European Journal of Biochemistry*, 67, 67-80.

BLEHERT, D. S., FOX, B. G. & CHAMBLISS, G. H. (1999) Cloning and sequence analysis of two *Pseudomonas* flavoprotein xenobiotic reductases. *Journal of Bacteriology*, 181, 6254-6263.

BOLAND, M. P., KNOX, R. J. & ROBERTS, J. J. (1991) The Differences in Kinetics of Rat and Human DT-Diaphorase Result in a Differential Sensitivity of Derived Cell-Lines to CB-1954 (5-(Aziridin-1-yl)-2,4-Dinitrobenzamide). *Biochemical Pharmacology*, 41, 867-875.

BOUTIN, J. A., CHATELAIN-EGGER, F., VELLA, F., DELAGRANGE, P. & FERRY, G. (2005) Quinone reductase 2 substrate specificity and inhibition pharmacology. *Chemo-Biological Interactions*, 151, 213-228.

BOUTIN, J. A., SAUNIER, C., GUENIN, S. P., BERGER, S., MOULHARAT, N., GOHIER, A., DELAGRANGE, P., COGE, F. & FERRY, G. (2008) Studies of the melatonin binding site location onto quinone reductase 2 by directed mutagenesis. *Archives of Biochemistry and Biophysics*, 477, 12-19.

BROWN, B. J., DENG, Z., KARPLUS, P. A. & MASSEY, V. (1998) On the active site of old yellow enzyme - Role of histidine 191 and asparagine 194. *Journal of Biological Chemistry*, 273, 32753-32762.

BROWN, B. J., HYUN, J. W., DUVVURI, S., KARPLUS, P. A. & MASSEY, V. (2002) The role of glutamine 114 in Old Yellow Enzyme. *Journal of Biological Chemistry*, 277, 2138-2145.

BRUNGER, A. T. (1992) Free R-Value - A Novel Statistical Quantity For Assessing The Accuracy of Crystal Structures. *Nature*, 355, 472-475.

BRYANT, C. & DELUCA, M. (1991) Purification and Characterization of an Oxygen-Insensitive NAD(P)H Nitroreductase from *Enterobacter-Cloacae*. *Journal of Biological Chemistry*, 266, 4119-4125.

- CABALLERO, A., ESTEVE-NUNEZ, A., ZYLSTRA, G. J. & RAMOS, J. L. (2005a) Assimilation of nitrogen from nitrite and trinitrotoluene in *Pseudomonas putida* JLR11. *Journal of Bacteriology*, 187, 396-399.
- CABALLERO, A., LAZARO, J. J., RAMOS, J. L. & ESTEVE-NUNEZ, A. (2005b) PnrA, a new nitroreductase-family enzyme in the TNT-degrading strain *Pseudomonas putida* JLR11. *Environmental Microbiology*, 7, 1211-1219.
- CALAMINI, B., SANTARSIERO, B. D., BOUTIN, J. A. & MESECAR, A. D. (2008) Kinetic, thermodynamic and X-ray structural insights into the interaction of melatonin and analogues with quinone reductase 2. *Biochemical Journal*, 413, 81-91.
- CHEN, S., WU, K. B. & KNOX, R. (2000) Structure-function studies of DT-diaphorase (NQO1) and NRH : quinone oxidoreductase (NQO2). *Free Radical Biology and Medicine*, 29, 276-284.
- CHEN, Y. & HU, L. Q. (2009) Design of Anticancer Prodrugs for Reductive Activation. *Medicinal Research Reviews*, 29, 29-64.
- CLELAND, W. W. (1964) Dithiothreitol A New Protective Reagent for SH Groups. *Biochemistry*, 3, 480-481.
- COLUCCI, M. A., MOODY, C. J. & COUCH, G. D. (2008) Natural and synthetic quinones and their reduction by the quinone reductase enzyme NQO1: from synthetic organic chemistry to compounds with anticancer potential. *Organic & Biomolecular Chemistry*, 6, 637-656.
- DALTON, J. A. R. & JACKSON, R. M. (2007) An evaluation of automated homology modelling methods at low target-template sequence similarity. *Bioinformatics*, 23, 1901-1908.
- DAUTER, Z. (1999) Data-collection strategies. *Acta Crystallographica Section D-Biological Crystallography*, 55, 1703-1717.
- DAVIS, I. W., LEAVER-FAY, A., CHEN, V. B., BLOCK, J. N., KAPRAL, G. J., WANG, X., MURRAY, L. W., ARENDALL, W. B., SNOEYINK, J., RICHARDSON, J. S. & RICHARDSON, D. C. (2007) MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. *Nucleic Acids Research*, 35, W375-W383.
- DE COLIBUS, L. & MATTEVI, A. (2006) New frontiers in structural flavoenzymology. *Current Opinion in Structural Biology*, 16, 722-728.
- DELANO, W. (2008) The PyMOL Molecular Graphics System. CA, USA, DeLano Scientific LLC.
- DELLER, S., MACHEROUX, P. & SOLLNER, S. (2008) Flavin-dependent quinone reductases. *Cellular and Molecular Life Sciences*, 65, 141-160.

- DENNY, W. A. (2001) Prodrug strategies in cancer therapy. *European Journal of Medicinal Chemistry*, 36, 577-595.
- DOLORES ROLDAN, M., PEREZ-REINADO, E., CASTILLO, F. & MORENO-VIVIAN, C. (2008) Reduction of polynitroaromatic compounds: the bacterial nitroreductases. *FEMS Microbiology Reviews*, 32, 474-500.
- DRENTH, J. (2007) *Principles of Protein X-ray Crystallography*, Springer.
- DURBIN, S. D. & FEHER, G. (1996) Protein crystallization. *Annual Review of Physical Chemistry*, 47, 171-204.
- ELOFSSON, A. (2002) A study on protein sequence alignment quality. *Proteins-Structure Function and Genetics*, 46, 330-339.
- EMSLEY, P. & COWTAN, K. (2004) Coot: model-building tools for molecular graphics. *Acta Crystallographica Section D-Biological Crystallography*, 60, 2126-2132.
- ESTEVE-NUNEZ, A., CABALLERO, A. & RAMOS, J. L. (2001) Biological degradation of 2,4,6-trinitrotoluene. *Microbiology and Molecular Biology Reviews*, 65, 335-352.
- ESWAR, N., JOHN, B., MIRKOVIC, N., FISER, A., ILYIN, V. A., PIEPER, U., STUART, A. C., MARTI-RENOM, M. A., MADHUSUDHAN, M. S., YERKOVICH, B. & SALI, A. (2003) Tools for comparative protein structure modeling and analysis. *Nucleic Acids Research*, 31, 3375-3380.
- ESWAR NARAYANAN, U. P. A. B. W. ModWeb Server: A Server for Protein Structure Modeling.
- EVANS, P. (1997) SCALA. *Joint CCP4+ ESF-EAMCB Newsletter on Protein Crystallography*, 33, 22-24.
- EVANS, P. & MCCOY, A. (2008) An introduction to molecular replacement. *Acta Crystallographica Section D-Biological Crystallography*, 64, 1-10.
- EWALD, P. P. (1921) The "reciprocal lattice" in structure theory. *Zeitschrift Fur Kristallographie*, 56, 129-156.
- FAIG, M., BIANCHET, M. A., TALALAY, P., CHEN, S., WINSKI, S., ROSS, D. & AMZEL, L. M. (2000) Structures of recombinant human and mouse NAD(P)H : quinone oxidoreductases: Species comparison and structural changes with substrate binding and release. *Proceedings of the National Academy of Sciences of the United States of America*, 97, 3177-3182.

- FERRY, G., HECHT, S., BERGER, S., MOULHARAT, N., COGE, F., GUILLAUMET, G., LECLERC, V., YOUS, S., DELAGRANGE, P. & BOUTIN, J. A. Old and new inhibitors of quinone reductase 2. *Chemico-Biological Interactions*, 186, 103-109.
- FOSTER, C. E., BIANCHET, M. A., TALALAY, P., ZHAO, Q. J. & AMZEL, L. M. (1999) Crystal structure of human quinone reductase type 2, a metalloflavoprotein. *Biochemistry*, 38, 9881-9886.
- FOX, K. M. & KARPLUS, P. A. (1994) Old Yellow Enzyme at 2-Angstrom Resolution - Overall Structure, Ligand-Binding and Comparison With Related Flavoproteins. *Structure*, 2, 1089-1105.
- FRAAIJE, M. W. & MATTEVI, A. (2000) Flavoenzymes: diverse catalysts with recurrent features. *Trends in Biochemical Sciences*, 25, 126-132.
- FU, P. P. (1990) Metabolism of Nitropolycyclic Aromatic Hydrocarbons. *Drug Metabolism Reviews*, 22, 209-268.
- FU, Y., BURYANOVSKYY, L. & ZHANG, Z. T. (2005) Crystal structure of quinone reductase 2 in complex with cancer prodrug CB1954. *Biochemical and Biophysical Research Communications*, 336, 332-338.
- GHISLA, S. & MASSEY, V. (1986) New Flavins for Old - Artificial Flavins as Active-Site Probes of Flavoproteins. *Biochemical Journal*, 239, 1-12.
- GOORDARZ DANAEI, S. V. H., ALAN LOPEZ, CHRISTOPHER MURRAY, MAJID EZZATI AND THE COMPARATIVE RISK ASSESSMENT COLLABORATING GROUP (CANCERS) (2005) Causes of cancer in the world: Comparative Risk Assessment of Nine Behavioural and Environmental Risk Factors. *The Lancet*, 366, 1784-1793.
- GRUBER, T. D., WESTLER, W. M., KIESSLING, L. L. & FOREST, K. T. (2009) X-ray Crystallography Reveals a Reduced Substrate Complex of UDP-Galactopyranose Mutase Poised for Covalent Catalysis by Flavin. *Biochemistry*, 48, 9171-9173.
- HAYNES, C. A., KODER, R. L., MILLER, A. F. & RODGERS, D. W. (2002) Structures of nitroreductase in three states - Effects of inhibitor binding and reduction. *Journal of Biological Chemistry*, 277, 11513-11520.
- HEDLEY, D., OGILVIE, L. & SPRINGER, C. (2007) Carboxypeptidase G2-based gene-directed enzyme-prodrug therapy: a new weapon in the GDEPT armoury. *Nature Reviews Cancer*, 7, 870-879.
- HELSEBY, N. A., FERRY, D. M., PATTERSON, A. V., PULLEN, S. M. & WILSON, W. R. (2004) 2-Amino metabolites are key mediators of CB 1954 and SN 23862 bystander effects in nitroreductase GDEPT. *British Journal of Cancer*, 90, 1084-1092.

- HYNISON, R. M. G., KELLY, S. M., PRICE, N. C. & RAMSAY, R. R. (2004) Conformational changes in monoamine oxidase A in response to ligand binding or reduction. *Biochimica Et Biophysica Acta-General Subjects*, 1672, 60-66.
- INTERNATIONAL, T. Sybyl 8.0. 1699 South Hanley Rd., St. Louis, Missouri, 63144, USA.
- JAFFAR, M., NAYLOR, M. A., ROBERTSON, N., LOCKYER, S. D., PHILLIPS, R. M., EVERETT, S. A., ADAMS, G. E. & STRATFORD, I. J. (1998) 5-substituted analogues of 3-hydroxymethyl-5-aziridinyl-1-methyl-2-[1H-indole-4,7-dione]prop-2-en-1-ol (E09, NSC 382459) and their regioisomers as hypoxia-selective agents: Structure-cytotoxicity in vitro. *Anti-Cancer Drug Design*, 13, 105-123.
- JONES, S. & THORNTON, J. M. (2004) Searching for functional sites in protein structures. *Current Opinion in Chemical Biology*, 8, 3-7.
- KABSCH, W. (2010) Integration, scaling, space-group assignment and post-refinement. *Acta Crystallographica Section D-Biological Crystallography*, 66, 133-144.
- KING, R. J. B. & ROBINS, M. (2006) *Cancer Biology*, Pearson/ Prentice Hall limited.
- KLEYWEGT, G. J. (2000) Validation of protein crystal structures. *Acta Crystallographica Section D-Biological Crystallography*, 56, 249-265.
- KNOX, R. J. & CHEN, S. A. (2004) Quinone reductase-mediated nitro-reduction: Clinical applications. *Quinones and Quinone Enzymes, Pt B*, 382, 194-221.
- KNOX, R. J., FRIEDLOS, F., JARMAN, M. & ROBERTS, J. J. (1988) A New Cytotoxic, DNA Interstrand Crosslinking Agent, 5-(Aziridin-1-yl)-4-hydroxyamino-2-Nitrobenzamide, is Formed from 5-(Aziridin-1-yl)-2,4-dinitrobenzamide (CB-1954) by a Nitroreductase Enzyme in Walker Carcinoma Cells. *Biochemical Pharmacology*, 37, 4661-4669.
- KNOX, R. J., FRIEDLOS, F., SHERWOOD, R. F., MELTON, R. G. & ANLEZARK, G. M. (1992) The Bioactivation of 5-(Aziridin-1-yl)-2,4-dinitrobenzamide (CB1954) .2. A Comparison of an Escherichia Coli Nitroreductase and Walker DT Diaphorase. *Biochemical Pharmacology*, 44, 2297-2301.
- KNOX, R. J., JENKINS, T. C., HOBBS, S. M., CHEN, S. A., MELTON, R. G. & BURKE, P. J. (2000) Bioactivation of 5-(aziridin-1-yl)-2,4-dinitrobenzamide (CB 1954) by human NAD(P)H quinone oxidoreductase 2: A novel co-substrate-mediated antitumor prodrug therapy. *Cancer Research*, 60, 4179-4186.
- KODER, R. L., OYEDELE, O. & MILLER, A. F. (2001) Retro-nitroreductase, a putative evolutionary precursor to Enterobacter cloacae strain 96-3 nitroreductase. *Antioxidants & Redox Signaling*, 3, 747-755.

- KOHLI, R. M. & MASSEY, V. (1998) The oxidative half-reaction of old yellow enzyme - The role of tyrosine 196. *Journal of Biological Chemistry*, 273, 32763-32770.
- KOIKE, H., SASAKI, H., KOBORI, T., ZENNO, S., SAIGO, K., MURPHY, M. E. P., ADMAN, E. T. & TANOKURA, M. (1998) 1.8 angstrom crystal structure of the major NAD(P)H : FMN oxidoreductase of a bioluminescent bacterium, *Vibrio fischeri*: Overall structure, cofactor and substrate-analog binding, and comparison with related flavoproteins. *Journal of Molecular Biology*, 280, 259-273.
- KUTTY, R. & BENNETT, G. N. (2005) Biochemical characterization of trinitrotoluene transforming oxygen-insensitive nitroreductases from *Clostridium acetobutylicum* ATCC 824. *Archives of Microbiology*, 184, 158-167.
- KWIEK, J. J., HAYSTEAD, T. A. J. & RUDOLPH, J. (2004) Kinetic mechanism of quinone oxidoreductase 2 and its inhibition by the antimalarial quinolines. *Biochemistry*, 43, 4538-4547.
- KWON, D. H., EL-ZAATARI, F. A. K., KATO, M., OSATO, M. S., REDDY, R., YAMAOKA, Y. & GRAHAM, D. Y. (2000) Analysis of rdxA and involvement of additional genes encoding NAD(P)H flavin oxidoreductase (FrxA) and ferredoxin-like protein (FdxB) in metronidazole resistance of *Helicobacter pylori*. *Antimicrobial Agents and Chemotherapy*, 44, 2133-2142.
- KWONG, P. D. & LIU, Y. (1999) Use of cryoprotectants in combination with immiscible oils for flash cooling macromolecular crystals. *Journal of Applied Crystallography*, 32, 102-105.
- LADD, M. & PALMER, R. (2003) *Structure Determination by X-ray Crystallography*, Kluwer academic/Plenum publishers.
- LARKIN, M. A., BLACKSHIELDS, G., BROWN, N. P., CHENNA, R., MCGETTIGAN, P. A., MCWILLIAM, H., VALENTIN, F., WALLACE, I. M., WILM, A., LOPEZ, R., THOMPSON, J. D., GIBSON, T. J. & HIGGINS, D. G. (2007) Clustal W and clustal X version 2.0. *Bioinformatics*, 23, 2947-2948.
- LASKOWSKI, R. A., MACARTHUR, M. W., MOSS, D. S. & THORNTON, J. M. (1993) PROCHECK - A Program to Check The Stereochemical Quality of Protein Structures. *Journal of Applied Crystallography*, 26, 283-291.
- LEATHERBARROW, R. J. (2009) GraFit Version 7. 7.0 ed. Horley, UK, Erithacus Software Limited.
- LEI, B. F., LIU, M. Y., HUANG, S. Q. & TU, S. C. (1994) *Vibrio-Harveyi* NADPH-Flavin oxidoreductase - Cloning, Sequencing and Overexpression of the Gene and Purification and Characterization of the Cloned Enzyme. *Journal of Bacteriology*, 176, 3552-3558.

- LENNON, B. W., WILLIAMS, C. H. & LUDWIG, M. L. (1999) Crystal structure of reduced thioredoxin reductase from *Escherichia coli*: Structural flexibility in the isoalloxazine ring of the flavin adenine dinucleotide cofactor. *Protein Science*, 8, 2366-2379.
- LESLIE, A. G. W. (1992) Recent changes to the MOSFLM package for processing film and image plate data. *Joint CCP4+ ESF-EAMCB Newsletter on Protein Crystallography*, 26.
- LESLIE, A. G. W. (2006) The integration of macromolecular diffraction data. *Acta Crystallographica Section D-Biological Crystallography*, 62, 48-57.
- LI, R. B., BIANCHET, M. A., TALALAY, P. & AMZEL, L. M. (1995) The 3-Dimensional Structure of NAD(P)H-Quinone Reductase, a Flavoprotein Involved in Cancer Chemoprotection and Chemotherapy - Mechanism of The 2-Electron Reduction. *Proceedings of the National Academy of Sciences of the United States of America*, 92, 8846-8850.
- LIN, T. S., TEICHER, B. A. & SARTORELLI, A. C. (1980) 2-Methylantraquinone Derivative as Potential Bioreductive Alkylating Agents. *Journal of Medicinal Chemistry*, 23, 1237-1242.
- LUDWIG, M. L., PATTRIDGE, K. A., METZGER, A. L., DIXON, M. M., EREN, M., FENG, Y. C. & SWENSON, R. P. (1997) Control of oxidation-reduction potentials in flavodoxin from *Clostridium beijerinckii*: The role of conformation changes. *Biochemistry*, 36, 1259-1280.
- MAILLIET, F., FERRY, G., VELLA, F., BERGER, S., COGE, F., CHOMARAT, P., MALLET, C., GUENIN, S. P., GUILLAUMET, G., VIAUD-MASSUARD, M. C., YOUS, S., DELAGRANGE, P. & BOUTIN, J. A. (2005) Characterization of the melatonergic MT3 binding site on the NRH : quinone oxidoreductase 2 enzyme. *Biochemical Pharmacology*, 71, 74-88.
- MASSA, W. (2004) *Crystal Structure Determination*, Springer.
- MASSEY, V. (2000) The chemical and biological versatility of riboflavin. *Biochemical Society Transactions*, 28, 283-296.
- MCCOY, A. J., GROSSE-KUNSTLEVE, R. W., ADAMS, P. D., WINN, M. D., STORONI, L. C. & READ, R. J. (2007) Phaser crystallographic software. *Journal of Applied Crystallography*, 40, 658-674.
- MCGUINNESS, B. F., LIPMAN, R., GOLDSTEIN, J., NAKANISHI, K. & TOMASZ, M. (1991) Reductive Alkylation of DNA by Mitomycin-A, a Mitomycin with High Redox Potential. *Biochemistry*, 30, 6444-6453.
- MCPHERSON, A. (2009) *Introduction to Macromolecular Crystallography*, Wiley-Blackwell.

- MIURA, R. (2001) Versatility and specificity in flavoenzymes: Control mechanisms of flavin reactivity. *Chemical Record*, 1, 183-194.
- MONKS, T. J., HANZLIK, R. P., COHEN, G. M., ROSS, D. & GRAHAM, D. G. (1992) Quinone Chemistry and Toxicity. *Toxicology and Applied Pharmacology*, 112, 2-16.
- MOSSOBA, M. M., ALIZADEH, M. & GUTIERREZ, P. L. (1985a) MECHANISM FOR THE REDUCTIVE ACTIVATION OF DIAZQUONE. *Journal of Pharmaceutical Sciences*, 74, 1249-1254.
- MOSSOBA, M. M., ALIZADEH, M. & GUTIERREZ, P. L. (1985b) Mechanism for The Reductive Activation of Diazquinone. *Journal of Pharmaceutical Sciences*, 74, 1249-1254.
- MURSHUDOV, G. N., VAGIN, A. A. & DODSON, E. J. (1997) Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallographica Section D-Biological Crystallography*, 53, 240-255.
- NADEAU, L. J., HE, Z. & SPAIN, J. C. (2000) Production of 2-amino-5-phenoxyphenol from 4-nitrobiphenyl ether using nitrobenzene nitroreductase and hydroxylaminobenzene mutase from *Pseudomonas pseudoalcaligenes* JS45. *Journal of Industrial Microbiology & Biotechnology*, 24, 301-305.
- NOTREDAME, C., HIGGINS, D. G. & HERINGA, J. (2000) T-Coffee: A novel method for fast and accurate multiple sequence alignment. *Journal of Molecular Biology*, 302, 205-217.
- OBRIEN, P. J. (1991) Molecular Mechanisms of Quinone Cytotoxicity. *Chemico-Biological Interactions*, 80, 1-41.
- OTWINOWSKI, Z. & MINOR, W. (1997) Processing of X-ray diffraction data collected in oscillation mode. *Macromolecular Crystallography, Pt A*, 276, 307-326.
- PAK, J. W., KNOKE, K. L., NOGUERA, D. R., FOX, B. G. & CHAMBLISS, G. H. (2000) Transformation of 2,4,6-trinitrotoluene by purified xenobiotic reductase B from *Pseudomonas fluorescens* I-C. *Applied and Environmental Microbiology*, 66, 4742-4750.
- PARK, H. J., REISER, C. O. A., KONDRUWEIT, S., ERDMANN, H., SCHMID, R. D. & SPRINZL, M. (1992) Purification and Characterization of a NADH Oxidase from the Thermophile *Thermus-Thermophilus* HB8. *European Journal of Biochemistry*, 205, 881-885.
- PARKINSON, G. N., SKELLY, J. V. & NEIDLE, S. (2000) Crystal structure of FMN-dependent nitroreductase from *Escherichia coli* B: A prodrug-activating enzyme. *Journal of Medicinal Chemistry*, 43, 3624-3631.
- PEI, J. M. The PROMALS multiple sequence alignment server.

- PEI, J. M. (2008) Multiple protein sequence alignment. *Current Opinion in Structural Biology*, 18, 382-386.
- PEI, J. M. & GRISHIN, N. V. (2001) AL2CO: calculation of positional conservation in a protein sequence alignment. *Bioinformatics*, 17, 700-712.
- PEI, J. M. & GRISHIN, N. V. (2007) PROMALS: towards accurate multiple sequence alignments of distantly related proteins. *Bioinformatics*, 23, 802-808.
- PETTERSEN, E. F., GODDARD, T. D., HUANG, C. C., COUCH, G. S., GREENBLATT, D. M., MENG, E. C. & FERRIN, T. E. (2004) UCSF chimera - A visualization system for exploratory research and analysis. *Journal of Computational Chemistry*, 25, 1605-1612.
- PUTMAN, C. Protein Calculator v 3.3. The Scripps Research Institute.
- RAMACHANDRAN, G. N., RAMAKRISHNAN, C. & SASISEKHARAN, V. (1963) Stereochemistry of Polypeptide Chain Configurations. *Journal of Molecular Biology*, 7, 95-99.
- RAMOS, J. L., GONZALEZ-PEREZ, M. M., CABALLERO, A. & VAN DILLEWIJN, P. (2005) Bioremediation of polynitrated aromatic compounds: plants and microbes put up a fight. *Current Opinion in Biotechnology*, 16, 275-281.
- RAUTH, A. M., MELO, T. & MISRA, V. (1998) Bioreductive therapies: An overview of drugs and their mechanisms of action. *International Journal of Radiation Oncology Biology Physics*, 42, 755-762.
- RHODES, G. (2000) *Crystallography Made Crystal Clear: A Guide for users of Macromolecular Models*.
- RICE, P., LONGDEN, I. & BLEASBY, A. (2000) EMBOSS: The European molecular biology open software suite. *Trends in Genetics*, 16, 276-277.
- ROLDAN, M., PEREZ-REINADO, E., CASTILLO, F. & MORENO-VIVIAN, C. (2008) Reduction of polynitroaromatic compounds: the bacterial nitroreductases. *FEMS Microbiology Reviews*, 32, 474-500.
- ROSS, D. & SIEGEL, D. (2004) NAD(P)H : Quinone oxidoreductase 1 (NQO1, DT-diaphorase), functions and pharmacogenetics. *Quinones and Quinone Enzymes, Pt B*, 382, 115-144.
- ROSSMANN, M. G. & BLOW, D. M. (1962) Detection of Sub-units Within Crystallographic Asymmetric Unit. *Acta Crystallographica*, 15, 24-&.

- RUBINSON, K. A., LADNER, J. E., TORDOVA, M. & GILLILAND, G. L. (2000) Cryosalts: Suppression of ice formation in macromolecular crystallography. *Acta Crystallographica Section D Biological Crystallography*, 56, 996-1001.
- SCHNEIDER, S. (2007) Structural and Biochemical Studies on Novel Bacterial Haem-proteins. *School of Pharmacy, Centre of Biomolecular science*. Nottingham, University of Nottingham.
- SHARMA, A. & SCHULMAN, S. G. (1999) *Introduction to Fluorescence Spectroscopy*, Wiley-Interscience.
- SHEN, M. Y. & SALI, A. (2006) Statistical potential for assessment and prediction of protein structures. *Protein Science*, 15, 2507-2524.
- SINGH, Y., PALOMBO, M. & SINKO, P. J. (2008) Recent trends in targeted anticancer prodrug and conjugate design. *Current Medicinal Chemistry*, 15, 1802-1826.
- SINNOTT, M., GARNER, D., FIRST, E. & DAVIES, G. (1997) *Comprehensive Biological Catalysis: A Mechanistic Reference*, London: Academic Press.
- SKELLY, J. V., KNOX, R. J. & JENKINS, T. C. (2001) Aerobic nitroreduction by flavoproteins: Enzyme structure, mechanisms and role in cancer chemotherapy. *Mini-Reviews in Medicinal Chemistry*, 1, 293-306.
- SKIBO, E. B. & XING, C. G. (1998) Chemistry and DNA alkylation reactions of aziridinyl quinones: Development of an efficient alkylating agent of the phosphate backbone. *Biochemistry*, 37, 15199-15213.
- SOMERVILLE, C. C., NISHINO, S. F. & SPAIN, J. C. (1995) Purification and Characterization of Nitrobenzene Nitroreductase from *Pseudomonas Pseudoalcaligenes* JS45. *Journal of Bacteriology*, 177, 3837-3842.
- TANABE, K., MAKIMURA, Y., TACHI, Y., IMAGAWA-SATO, A. & NISHIMOTO, S. (2005) Hypoxia-selective activation of 5-fluorodeoxyuridine prodrug possessing indolequinone structure: radiolytic reduction and cytotoxicity characteristics. *Bioorganic & Medicinal Chemistry Letters*, 15, 2321-2324.
- TRAMONTANO, A. (1998) Homology modeling with low sequence identity. *Methods-a Companion to Methods in Enzymology*, 14, 293-300.
- TRONRUD, D. (2010) The Wonderful World of Maps. *CCP4 Study Weekend*. Nottingham U.K.
- TRONRUD, D. E. (2004) Introduction to macromolecular refinement. *Acta Crystallographica Section D-Biological Crystallography*, 60, 2156-2168.

- VELLA, F., GILLES, F. A., DELAGRANGE, P. & BOUTIN, J. A. (2005) NRH : quinone reductase 2: An enzyme of surprises and mysteries. *Biochemical Pharmacology*, 71, 1-12.
- WATANABE, M., NISHINO, T., TAKIO, K., SOFUNI, T. & NOHMI, T. (1998) Purification and characterization of wild-type and mutant "Classical" nitroreductases of *Salmonella typhimurium* - L33R mutation greatly diminishes binding of FMN to the nitroreductase of *S*-typhimurium. *Journal of Biological Chemistry*, 273, 23922-23928.
- WATSON, J. D., LASKOWSKI, R. A. & THORNTON, J. M. (2005) Predicting protein function from sequence and structural data. *Current Opinion in Structural Biology*, 15, 275-284.
- WILLIAMS, R. E. & BRUCE, N. C. (2002) 'New uses for an Old Enzyme' - the Old Yellow Enzyme family of flavoenzymes. *Microbiology-Sgm*, 148, 1607-1614.
- WILLIAMS, R. E., RATHBONE, D. A., SCRUTTON, N. S. & BRUCE, N. C. (2004) Biotransformation of explosives by the old yellow enzyme family of flavoproteins. *Applied and Environmental Microbiology*, 70, 3566-3574.
- WU, K. B., KNOX, R., SUN, X. Z., JOSEPH, P., JAISWAL, A. K., ZHANG, D., DENG, P. S. K. & CHEN, S. (1997) Catalytic properties of NAD(P)H:quinone oxidoreductase-2 (NQO2), a dihydronicotinamide riboside dependent oxidoreductase. *Archives of Biochemistry and Biophysics*, 347, 221-228.
- ZENNO, S., KOIKE, H., KUMAR, A. N., JAYARAMAN, R., TANOKURA, M. & SAIGO, K. (1996) Biochemical characterization of NfsA, the *Escherichia coli* major nitroreductase exhibiting a high amino acid sequence homology to Frp, a *Vibrio harveyi* flavin oxidoreductase. *Journal of Bacteriology*, 178, 4508-4514.
- ZENNO, S., SAIGO, K., KANO, H. & INOUE, S. (1994) Identification of the Gene Encoding the Major NAD(P)H-Flavin Oxidoreductase of the Bioluminescent Bacterium *Vibrio-Fischeri* ATCC-7744. *Journal of Bacteriology*, 176, 3536-3543.
- ZHANG, Z., TANABE, K., HATTA, H. & NISHIMOTO, S. (2005) Bioreduction activated prodrugs of camptothecin: molecular design, synthesis, activation mechanism and hypoxia selective cytotoxicity. *Organic & Biomolecular Chemistry*, 3, 1905-1910.
- ZHAO, Q. J., YANG, X. L., HOLTZCLAW, W. D. & TALALAY, P. (1997) Unexpected genetic and structural relationships of a long-forgotten flavoenzyme to NAD(P)H:quinone reductase (DT-diaphorase). *Proceedings of the National Academy of Sciences of the United States of America*, 94, 1669-1674.

Appendices

A1. Amino acid abbreviations

Alanine	Ala	A
Arginine	Arg	R
Asparagine	Asn	N
Aspartic acid (aspartate)	Asp	D
Cysteine	Cys	C
Glutamine	Gln	Q
Glutamic acid (glutamate)	Glu	E
Glycine	Gly	G
Histidine	His	H
Isoleucine	Ile	I
Leucine	Leu	L
Lysine	Lys	K
Methionine	Met	M
Phenylalanine	Phe	F
Proline	Pro	P
Serine	Ser	S
Threonine	Thr	T
Tryptophan	Trp	W
Tyrosine	Try	Y
Valine	Val	V

A2. Composition of solutions used in experimental procedures

2xTY:	16 g Tryptone 10 g Yeast Extract 10 g NaCl to 1L H ₂ O _{MQ}
3X Protein Loading Buffer:	0.135 M Tris-HCl, pH 6.8 0.15 M Dithiothreitol 30% (v/v) Glycerol 3% (w/v) SDS 0.03% (w/v) Bromophenol blue in H ₂ O _{MQ}
Coomassie Destaining Solution:	25% (v/v) Methanol 25% (v/v) Glacial Acetic acid in H ₂ O _{MQ}
Coomassie Staining Solution:	0.25% (v/v) Coomassie Brilliant Blue R250 25% (v/v) Methanol 25% (v/v) Glacial Acetic acid in H ₂ O _{MQ}
Lysis Buffer:	50 mM Tris-HCl, pH 8 300 mM NaCl
SDS Running Buffer:	0.025 M Tris 0.19 M Glycine 0.1 % SDS

A3. Sequence alignments

A3.1 Alignment of group I protein sequences using Clustal-W2

PETN	-----SAEKLF ^T PLKVGAVTAPNRFMAPL ^T RLRSIEPGDIPT	38
Morph	-----MPDTSF ^S NPGLFTPLQLGSL ^S LPN ^R VMAPL ^T RRS--TPDSVPG	42
GTN	-----MTSLFEPQA ^G DIALANR ^I VMAPL ^T RRN--SPGAI ^P N	35
NemA	-----MAHHHHH ^H MPSLFDPL ^T IGDLTLANR ^I IMAPL ^T RRAR-AGDTRTPN	44
XenB	-----MTTLFD ^P IKLGDLELSN ^R IIMAPL ^T RRCR-ADAGRVPN	36
Cyclo	-----MPTLFD ^P LTLGD ^L QSPN ^R VLMA ^P LTRGR-ATREHVPT	36
OYE	MSFVKDFK ^P QALGD ^T NLFK ^P IKIGNNELLHRAV ^I PPLTRMRALHPGN ^I PN	50
XenA	-----ALFEPY ^T LKDV ^T LRNR ^I AIPPM ^C QYM--AEDGMIN	33
	* * * * : : * : :	
PETN	-PLMGEYYRQRAS--AGLI ^I SEATQISAQAKGYAGAPGLHSPEQIAAWKK	85
Morph	-RLQQIYYQRAS--AGLI ^I SEATNISPARGVYTPPGIWTDQAGWKK	89
GTN	-NLNATYYQRAT--AGLIVTEGTPISQQGGYADVPGLYKREAI ^E GWKK	82
NemA	-ALMARYAERAS--AGLI ^I SEATSVTPQGVGYASTPGIWSPEQVDGWR ^L	91
XenB	-ALMAEYVQRAS--AGLILSEATSVTPMGVGYPDTPGIWSNDQVRGWAN	83
Cyclo	-ELMIEYYQRAS--AGLI ^I TEATGITQEGLGWPYAGIWSDEQVEAWK ^P	83
OYE	RDWAVEYYTQRAQRPGTMIITEGAFIS ^P QAGGYDNAPGVWSEEQMV ^E WTK	100
XenA	-DWHHVHLA ^G LARGGAGLLVVEATAVAPEGRITPGCAGIWSDAHAQAFV ^P	82
	: * : : : * : : : : : : : : : : : : : : :	
PETN	ITAGVHAEDGRIAVQLWHTGRISHSSIQGGQAPVSASALNANTRTSLRD	135
Morph	VVEAVHAKGGRIALQLWHVGRVSH ^E LVQPDGQQPVAPSALKAEGAE ^C FVE	139
GTN	ITDGVH ^S AGGKIVAQIWHVGRISH ^T SLQPHGGQPVAPSAITAKSKTY ^I IN	132
NemA	VTDAVHAAGGRI ^F LQLWHVGRVSDPVFLD-GALPVAPSAIAPGGH ^S LVLR	140
XenB	VTKAVHGAGGKI ^F LQLWHVGRISH ^S PSYLN-GETPVAPSA ^L QPKGHV ^S LVLR	132
Cyclo	VTQAVHEAGGRI ^I LQLWHMGR ^T VHSSFLG-GAKPVSSSATRAPGAHT ^Y E	132
OYE	IFNAIHEKKS ^F VVWLQWLWGWA ^A FPDNLARDGLR ^Y DSASDNV ^F MDAEQEA	150
XenA	VVQAIKAAGSVPGIQIAHAGRKASANR ^P WEGDDHIAADDTRGWETIAPSA	132
	: : : : * : * : : : : : : : : : : : : : :	
PETN	-ENGNAIRVDTTTPRALELDEIPGIVNDFRQAVANAREAGFDLVELHSAH	184
Morph	FEDGTAGLHPTSTPRALET ^D GIPGIVEDYRQAAQRAKRA ^G FD ^M VEVHAA ^N	148
GTN	-DDGTGAF ^A ETSEPRAL ^T IDDIGLILEDYRSGARAAL ^E AGFDGVEI ^H AA ^N	181
NemA	-----PQR ^P YVTPRALELDEIPGVVAAFRRGAENARAAGFDGVEV ^H GAN	184
XenB	-----PLAD ^F PTPRALETA ^E IADIVDAYRVGAENAKAAGFDGVEI ^H GAN	176
Cyclo	-----GKQDYDEARPLSADEIPRL ^L NDYEAHAKNAMAAGFDGVQI ^H AA ^N	176
OYE	-----KAKKANNPQHSLTKDEIKQYIKEYVQA ^K NSIAAGADGVEI ^H SAN	195
XenA	---IAFGAHL ^P KVPREMTLDDIARVKQDFVDAARRARDAGFEWIELHFAH	179
	: : * : : : : : : : : : : : : : : : : : : :	
PETN	GYLLHQFLSPSSNQRTDQYGGSVENRARLVLEVVDVAVCN ^E WSADRIGIRV	234
Morph	ACL ^P NQFLATGTNRRTDQYGGSIENRARFPLEVVDVAE ^V FGPERV ^G IRL	239
GTN	GYLIEQFLKSSSTNQRTDDYGGSIENRARF ^L LEVVDVAE ^E IAGRTGIRL	231
NemA	GYL ^L DQFLQDSANRRTDAYGGSIENRARL ^L LEVVDAAIDVWSAARV ^G VHL	234
XenB	GYL ^L DQFLQSSSTNQRTDQYGGSENARL ^L LEVTDAAIEIWAGRV ^G VHL	226
Cyclo	GYLIDQFLRDNSNVRGDAYGGSIENRIR ^L LEVTRRVAETVGA ^E KTV ^R L	226
OYE	GYL ^L NQFLD ^P HSNTRTDEYGGSIENRARF ^T LEVVDALVEAIGHEK ^V GLR ^L	245
XenA	GYL ^G QSF ^F SEHSNKRTDAYGGSF ^D NRSRF ^L LET ^L AAVRE ^V WPEN---LPL	226
	* * * : : * * * * * : : * * : : * : : : : :	
PETN	SP-IGTFQNV ^D -NGPNEEADALYLIEELAKRG-----IAYLHMSE-----	272
Morph	TPFLELFG ^L TD-DEP--EAMAFYLAGELDRG-----LAYLHFNE-----	276
GTN	SPVTPANDIFE-ADP--QPLYNYVVEQLGKRN-----LAFIHVVEGATGG	273
NemA	APRGDAHTMGD-SDP--AATFGHVARELGRRR-----IAFLFARE-----	271
XenB	APRADSHDMGD-ANL--AETFTYVARELGKRG-----IAFIC ^S RE-----	263
Cyclo	SPNGDSQGVND-SNP--EPLFSAAAKALDEIG-----IAHLELREP-----	264
OYE	SPYGVFNSMSGGAETGIVAQYAYVAGELEKRAKAGKRLAFVH ^L VEPR---	292
XenA	TARFVLE ^Y DG-RDEQ ^T LEESIELARRFKAGG-----LDLLSVSVG----	266
	: :	
PETN	-TDLAGG-KPYSEA-FRQKVRER-FHGVIIGAGAYTAEKAEDLIGK---G	315
Morph	-PDWIGGDITYPEG-FREQMRQR-FKGGLIYCGNYDAGRAQARLDD---N	320
GTN	PRDFKQGD ^K PF ^D YASFKAA ^Y RNAGGKGLWIANNNGYDRQSAIEAVES---G	320
NemA	-----SFGGDAIGQQLKAA-FGGPFIVNENFTLDSQAALDA---G	308
XenB	-----KEGADSLG ^P LKEA-FGGPYIANERFTKDSANAWLAA---G	300
Cyclo	---GYEGTFGKADRPVHPVIRQA-FSRTLILNSDYTLETAQAALAT---G	308
OYE	VTNPFLTEGEYEGGNSDFVYSIWKGPVIRAGNFALHPVEVRE ^V EVK---D	340
XenA	-FTIPDTNIPWGPAPFMGPIAERVRREAKLPVTSAWGFGTPQLAEAAALQAN	315
	: :	

PETN	LIDAVAFGRDYIANPDLVARLQKKAELNPQRPFESFYGGGAEGYTDYPSL-	364
Morph	TADAVAFGRPFIANPDLPERFRLGAALNEPDPSTFYGGAEVGYTDYPFLD	370
GTN	KVDVAVFGKAFIANPDLVRRLLKNDAPLNAPNQPTFYGGGAEGYTDYPALA	370
NemA	QADAVAWGKLFIANPDLPRRFKLNAPLNENPNAATFYAQGEVGYTDYPALE	358
XenB	KADAVAFGVPFIANPDLPARLKADAPLNEAHPETFYGKGPVGYIDYPTL-	349
Cyclo	EADAITFGRPFLANPDLPHRFAERLPLNKDVMETWYSQGPGEVVDYPTAD	358
OYE	KRTLIGYGRFFISNPDLDVDRLEKGLPLNKYDRDTFYQMSAHGYIDYPTYE	390
XenA	QLDLVSVGRAHLADPHWAYFAAKELGVKAS-----WTLPAZYAHWLE--	358
	: * : : : * : : : : * : :	
PETN	-----	377
Morph	NGHDLRG---	361
GTN	Q-----	371
NemA	SAA-----	
XenB	-----	360
Cyclo	QK-----	400
OYE	EALKLGWDDK	
XenA	-----	

A3.2 Alignment of group I protein sequences using T-Coffee

Morph	MPDT-----SFSNPGLFTPLQLGSLSPNVRVIMAPLTRSRT--PDSVPGR-LQQIYYG-	
OYE	MSFVKDFKQALGDTNLFKPIKIGNNELLHRAVIPP LTRMRALHPGNIPNRDWAWEYTT-	
PETN	S-----AEKLFTPKLVGAVTAPNRVFMAPLTRLRSIEPGDIPTP-LMGEYRR-	
XenA	-----ALFEPYTLKDVTLRNR IAIPMCMQYMAE-DCMI-NDW-HHVHLAG	
NemA	MAHHHH----HHMPSLFDPLTIGDLTLANRIIMAPLTRARAG-DTRTPNA-LMAYYA-	
GTN	-----MTSLFEPAQAGDIALANRIVMAPLTRNRS--PGAIPNN-LNATYYE-	
XenB	-----MTTLFDPIKLGDLLELSNRIIMAPLTRCRAD-AGRVPNA-LMAEYVV-	
Cyclo	-----MPTLFDPLTLGLDQSPNRVLMAPLTRGRAT-REHVPTL-LMIEYTT-	
	* * * : * : : : : :	
Morph	-QRAS--AGLIISEATNISPTARGYVYTPGIWTD AQEAGWKGVVEAVHAKGGRIALQLWH	
OYE	-QRAQRPGTMIITEGAFISPAQAGGYDNAPGVWSEEQMVEWTKIFNAIHEKKSFWVQLWV	
PETN	-QRAS--AGLIISEATQISAQAKGYAGAPGLHSPEQIAAWKKITAGVHAEDGRIAVQLWH	
XenA	LARGG--AGLVVEATAVAPEGRITPGCAGIWSDAHAQAFVPPVQAIKAAGSVPGIQIAH	
NemA	-ERAS--AGLIISEATSVTPQGVGYASTPGIWSPEQVDGWRVLTDAVHAAGGRIIFLQLWH	
GTN	-QRAT--AGLIVTEGTPISQQGQGYADVPGLYKREAI EGWKKITDGVHSAGGKIVAQIWH	
XenB	-QRAS--AGLILSEATSVTPMGVGYDTPGIWSNDQVRGWANVTKAVHGAGGKIIFLQLWH	
Cyclo	-QRAS--AGLIITEATGITQEGLGWPYAPGIWSDEQVEAWKPVVTQAVHEAGGRIILQLWH	
	* . : : : * : : : : * : : : : : * :	
Morph	VGRVSHLVQ-----PDGQQPVAPSALKAEGAECFVEFEDGTAGLHPTSTPRAL	
OYE	LGWAAPFDNLA-----RDGLRYDSASD-----NVFMDAEQEAkakKANNPQHSL	
PETN	TGRISHSSIIQ-----PGQAPVSASALNANRTSL-RDENGNAIRVDTTTPRAL	
XenA	AGRKASANRPWEGDDHIAADTRGWETIAPSATAFGA-----HLPKVPREM	
NemA	VGRVSDPVF-----LDGALPVAPSAIAPGKHVSL-V-----RPQRPVVTPRAL	
GTN	VGRISHTSLQ-----PHGGQPVAPSATITAKSKTYI-INDDGTGAFAPETSEPRAL	
XenB	VGRISHPSY-----LNGETPVAPSALQPKGHVSL-V-----RPLADFPTPRAL	
Cyclo	MGRTVHSSF-----LGGAKPVSSSATRAPG-----QAHTYEGKQDYDEARPL	
	* * * : * : : * : : : * : : : : * :	
Morph	ETDGPVIGVEDYRQAAQRAKRAGFDMVEVHAANA CLPNQFLATGTNRRTDQYGGSIENRA	
OYE	TKDEIKQYIKEYVQAAKNSIAAGADGVEIHSANGYLLNQFLDPHSNRTDEYGGSIENRA	
PETN	ELDEIPGIVNDFRQAVANAREAGFDLVELHSAHGYYLLHQFLSPSSNQRTDQYGGSVENRA	
XenA	TLDDIARVKQDFVDAARRARDAGFEWIELHFAHGYLGGQFFSEHSNKRTDAYGGSPDNRS	
NemA	ELDEIPGVAAFRRGAENARAAGFDGVEVHGANGYLLDQFLQDSANRRTDAYGGSIENRA	
GTN	TIDDIGLILEDYRSGARAALAGFDGVEIHAANGYLIEQFLKSSSTNQRDDYGGSIENRA	
XenB	ETAETIADIVDAYRVAENAKAAGFDGVEIHGANGYLLDQFLQSSSTNQRDDYGGSIENRA	
Cyclo	SADEIPRLNDYEHAAKNAMAAGFDGVQIHAANGYLIDQFLRDNNSNRGDAYGGSIENRI	
	* : : : * : : : * : : : * : : : * : : * * * * * * * * * * * * :	
Morph	RFPLEVVDVAEVEFGPERVGI RLT PFLFLFGLTDD---EPEAMAFYLAGELDRRG-----	
OYE	RFTLEVVDALVEAIGHEKVGRLRSPYGVFNMSGGAE TGI VAYVAGELEKRAKAGKR	
PETN	RLVLEVVDAVCNESADRIGIRVSPIGTFQNVNDGP---NEEDALYLIIEELAKRG---	
XenA	RFLLETAAVREVWPE-NLPLTA-RFGVL-EYDG-RDEQTLSESI----ELARRFKA-GG	
NemA	RLLEVVDAADVWSAARVGVHLAPRGDAHTMGDS---DPAATFGHVARELGRRR-----	
GTN	RFPLEVVDVAEEI GAGRTGIRLSPVTPANDIFE A---DPQPLYNYVVEQLGKRN-----	
XenB	RLLEVTDAAEI EWGAGRVGVHLAPRADSHDMGDA---NLAETFTYVARELGKRG-----	
Cyclo	RLLEVTTRRAETVGAEKTVGRVLSPNGDSQGVNDS---NPEPLFSAAKALDEIG-----	
	* : * * : : * : : : * : : : * : : : * : : * * * * * * * * * * * * :	

NEC	PEAKAANHKGRTYFADMHRVDLKDQDDQWMAKQVYLVNNGVNFLLGVGAMGLDVAVPIEGFDAA	
NOX	AQKQAIQ--RAFA--AM---GQEARAWASGQSYILLGYLLLLLEAYGLGSPMLGFDPE	
FRase	EQKEAAAFASF--KVELNCDENGEHKAWTKPQAYLALGNALHTLARLNIDSTTMEGIDPE	
FrxA	TNSRFAQI IKNFQENDMKLNSESLFDWASKQTYIQMANMMMAAAMLGIDSCPIEGYDQE	
pnrB	EQARAGQNQRRHYVNLHRFDQKDVQHWMEKQTYLALGTALLGAAAHGLDATPIEGFDSK	
	: : : : * * * : : : : : * *	
NEC	ILDEEFGLKEKGF-----SLVVVPVGH--SVEDFNATLPKSRPLSTIV-TEC	
NOX	RVRAILGLPSRAA-----IPALVALGYP-AEE-----GYPSHRLPLERVV-LWR	
FRaseI	LLSEIFADELKGYE-----CHVALAIGYHHPSEYNASLPKSRKAFEDVI-TIL	
FrxA	KVEAYL--BEKGYLNTAEFGVSVMACFGYR-NQE-----ITPKTRWKTEVIYEVIE	
pnrB	VLDDELGLRERGF-----SVVILSLGYR-SEADFNAGLNKSRPLPASQV-FTL	
	: : : : * * : : *	

A3.6 Alignment of group II protein sequences using PROMALS

Conserv	9 6 66 6 5 6 5696 65699 6 9	
PnrB	-----MDTVSLAKRRYTTKAYDASRRIPQATVDALLEQLRHSPPSSVNSQPWHFIVADTAE GKALLAKS	63
NEC	-----DIISVALKRHSTKAFDASKKLTAEAEKIKTLLQYSPSSSTNSQPWHFIVASTE GKARVAKS	62
FRaseI	M-----THPIIHDLENRYTSKKYDPSKQVQEDLAVLLEALRLSASSINSQPWKFIVIESDAKQRMHDS	65
FrxA	M---DREQVVALQHQRFAAKKYDPNRRISQKDWEALVEVGRAPSSIGLEPWWMLLKNERMKEDLKPM	66
NOX	MEATLPVLDAKTAALKRRSIRRYRK-DEVP EGLLEIIEAALRAPSANWLQPWIRVVVRDPATKRALREA	69
Consensus	hhhhhhhhh hh hhhhhhhhhhhh eeeee hhhhhhhhhh	
Conserv	6666 6 659	
PnrB	TAEGYAYNTQKLLDASHVIVFCTRTEMT EHLNAVLQDEAADGRFRDEQARAGQNQRRHYVNLHRFDQK	133
NEC	AAGTYVFNERKMLDASHVVVFCAKTAMD DAWLERVVDQDEADGRFNTPEAKAANHKGRTYFADMHRVDLK	132
FRaseI	FANMHQFNQPHIKACSHVILFANKLSYTRDDYDVVLSKAVADKRIT EEQKEAAAFASFKEV--ELNCDENG	133
FrxA	AWG---ALFGLLEGASHFVYILARKGVTYDSYVKKVMHEVKRKYDITNSRFAQI IKNFQENDMKLNSE	132
NOX	AF-----GQAHVEEAPVVLVLYADLEDAHLHDEVIHPGVQ-----GERREAQKAIQRAFAAMGQE	126
Consensus	hhhhhh hhhhh eeeee hhhhhhhhhhhhhh hhhhhhhhhhhhhhhhhh hh	
Conserv	9 9 95 6 6 6565 6569 9 5 5 5 6 96	
PnrB	DVQHWMEKQTYLALGTALLGAAAHGLDATPIEGFDSKVLDAELGLR----ERGFTSVVILSLGYRSEADF	199
NEC	DDDQWMAKQVYLVNNGVNFLLGVGAMGLDVAVPIEGFDAA ILDEEFGLK---EKGFTSLVVVPVGHHSVEDF	198
FRaseI	EHKAWTKPQAYLALGNALHTLARLNIDSTTMEGIDPELLSEIFADE---LKGYCHVALAIGYHHPSED	199
FrxA	SLFDWASKQTYIQMANMMMAAAMLGIDSCPIEGYDQEKVEAYLEEKGYLNTAEFGVSVMACFGYRNQEI-	201
NOX	ARKAWASGQSYILLGYLLLLLEAYGLGSPMLGFDPERVRAILGL----PSRAAIPALVALGYPAEEG-	190
Consensus	hhhhhhhhhhhhhhhhhhhhhhh eeee hhhhhhh eeeeeeee	
Conserv	66 9 5 5	
PnrB	-NAGLNKSRPLPASQVFTFL-	217
NEC	-NATLPKSRPLSTIVTEC-	216
FRaseI	YNASLPKSRKAFEDVITIL-	218
FrxA	----TPKTRWKTEVIYEVIE	217
NOX	----YPSHRLPLERVVLR-	205
Consensus	hhhhhh	

A3.7 Alignment of group III protein sequences using Clustal-W2

Frp	-----MNNTIETILAHRSIRKFTAVPITDEQR	27
NfsA	-----MTPTIELICGHR SIRHFTDEPISEAQR	27
NitA	-----MNNTIDTMKNHRSIRQYLDKEVPNDLI	27
pnrA	MSLQDEALKAWQARYGEPANLPAADTVIAQMLQHRSVRAYSDDLVDQML	24
NitB	-----MIDLLKTRRSIRKYKNKEIEKEKV	30
NbzA	-----MPTSPFIDDLIRDRRAKRGFLDQVPSIEMV	
	: : * * : :	
Frp	QTIIQAGLAASSSSMLQVVSIVRVT--DSEKRNELAQFAGNQAYVESAAE	75
NfsA	EAIINSARATSSSSFLQCSSIIIRIT--DKALREELVTLTGGQKHVAQAAE	75
NitA	DEIVKSAQAMPNSINGQQTSVIVVR--DKKKREKLAELVGNQEVYAKAPV	75
pnrA	SWAIAAAQSASTSSNLQAWSVLAVR--DRERLARLARLSGNQRHVEQAPL	98
NitB	DTLLKAALLAPTSMGKRSWEFIAVT--NKNLISELSTARMKGSQFLKGAP	72
NbzA	KDILSAAKYAPSSSNTQPWCYVITGEARERITTAVEAYRAAPEGLPPE	80
	: : . * : :	
Frp	FLVFCIDYQR----HATINPDVQADFTLTLIGAVDSGIMQNCLLAAE	120
NfsA	FWVFCADFNR----HLQICPDAQLGLAEQLLVGVDTAMMAQNALIAAE	120
NitA	FLVFMDFYRTYLAGKTKGLKQVIHEDIEGILAGSVDVGIALGASVVAEE	125
pnrA	FLVVLVDWSRLRLRLARTLQAPTAGIDYLESYTVGVVDAALAAQNAALAFE	148
NitB	LVIVVVENPE-----ATDAYIEDGAIASITLIQVTAHSMGLGSCWCHVR	115
NbzA	YPPFPQLHEP----YATRFNSFRGQLGDALGIPRSDITLRRRDRVERQFR	126

Frp	SMGLGGVYIGGLRNSAAQVDELLG---LPENSALVLFGMCLGHPD--QNPE	165
NfsA	SLGLGGVYIGGLRNNIEAVTKLLK---LPQHVLPFLGCLGWPA--DNPD	165
NitA	SLGLGTVPIGGIRKMPEEVIIEILG---LPKYTFPMVGLVVGYP--DESH	170
pnrA	AQGLGIVYIGGMRNHPEAMSEELG---LPNDTFAVFGMCGVGHPPAQPAE	195
NitB	NRDR-----IDYDTTEAFIKTLN---IPENLKVCEMLGIGYPD---EE	153
NbzA	FFDAPVGLIFTMDRRLEWASFCYGCFLQNIMLAAKGRGLDTCTQVFSM	176
	: : :	
Frp	VKPRLPAHVVVHENQYQELNDD--IQSYDQTMQAYASRTSNQKLTWS	213
NfsA	LKPRLPASILVHENSYQPLDKGA--LAQYDEQLAEYYLTRGSNNRRDTWS	213
NitA	KKPRVPFESFRHNESYDIKAVEDS-INVYDEQMNKYLKEIGRAEKEINWS	219
pnrA	IKPRLAQSIVLHRRERYEATEAEAVSVAAYDRRMSDFQHRQREN--SWS	243
NitB	RKAYTDEDLALNKLHYDKYSK-----	174
NbzA	QHPVLRTELNLPPDDQMVVAGMSLG---WADNSLPENQMSISKMELEEFT	222
	:	
Frp	QEVTKLAGES----RPHILPYLNSKGLAKR	240
NfsA	DHIRRTIIKES----RPFILDYLHKQGWATR	240
NitA	TFTSTIYQSVY----YPKVKGAINKQGLTK	246
pnrA	SQAVERVKGADSLSGRHRLRDALNTLGFGLR	274
NitB	-----	
NbzA	TFVHE-----	227

A3.8 Alignment of group III protein sequences using T-Coffee

NitA	M-----NN---TIDTMKNHRSIRQYLDKEVNPNDLIDEIVKSA	
NitB	M-----IDLLKTRRSIRKYKKEIEKEKVDTLKAA	
NfsA	M-----TP---TIELICGHR SIRHFTDEPISEAQREAIINSA	
Frp	M-----NN---TITILAHRSIRKFTAVPITDEQRQTIIQAG	
nbzA	M-----PTSPPFIDDLIRRRRAKRGFLDQPVSIEMVKDILSAA	
PnrA	MSLQDEALKAWQARYGEPANLPAADT---VIAQMLQHRSVRAYSDDLVPDEQMLSWAIAAA	
	* : : * * : : : : :	
NitA	QAMPNSINGQOTSIVIVRDKKKREKLAE-----	
NitB	LLAPTSMGKRSWEIFIAVTNKNLISELSTARK-----	
NfsA	RATSSSFLQCSSIIIRITDKALREELVTL-----	
Frp	LAASSSSMLQVVSIVRVTDSKRNELAQF-----	
nbzA	KYAPSSNTQPWRCYVITGE-ARERITTAAVEAYRAAPEGLPPEYPPFPQLHEPYATRF	
PnrA	QSASTSSNLQAWSVLAVRDRERLARLARL-----	
	..* : : : :	
NitA	-----VGNQEYVAKAPVFLVFMDFYR TYLAGEKTGLKQVI	
NitB	-----MGSQ-FLKGAPLIVVVEN-----	
NfsA	-----TGGQKHVAQAAEFWVFCADFNRHLQICP-----DAQ	
Frp	-----AGNQAYVESAAEFLVFCIDYQRHATINP-----DVQ	
nbzA	NSFRGQLGDALGIPRSDITLRRRDVERQFRFFDAPVGLIFTMD-----	
PnrA	-----SGNQRHVEQAPLFLVWLVDWSRLRRLARTLQAPTAG	
	* * : : :	
NitA	HEDIEGILAGSVDVGIALGASVVAESLGLGTVPIGGIRK-----MPEE-VIEILGLP	
NitB	---PEATDAYIEDGAIASITLIQVTAHSMGLGSCWCH-VRNRDRIDYDTTEAFIKTLNINP	
NfsA	LGLAEQLLLGVVDTAMMAQNALIAAESLGLGGVYIGGLRN-----NIEA-VTKLLKLP	
Frp	ADFTELTLIGAVDSGIMAQNCLLAAESMGLGGVYIGGLRN-----SAAQ-VDELLGLP	
nbzA	---RRLEWASFCYGCFLQNIMLAAKGRGLDTCTQV-----FWSMQHPVLRTELNLNLP	
PnrA	IDYLESYTVGVVDAALAAQNAALAFEAQGLGIVYIGGMRN-----HPEA-MSEELGLP	
	: : . . * * : : * * :	
NitA	KYTFPMVGLVVGYP--ADESHKKPRVPFESFRHNESYDIKAVED-SINVYDEQMNKYLKE	
NitB	ENLKVCEMLGIGYP--DEERKAY-----TDEDL-----	
NfsA	QHVLPFLGCLGWPA--ADNPDLKPRLPASILVHENSYQPLDKG-ALAQYDEQLAEYYL	
Frp	ENSAVLFGMCLGHP--DQNEVKPRLPAAHVVVHENQYQE-LNLD-DIQSYDQTMQAYAS	
nbzA	DDQMVVAGMSLGWA--DNSL-----PENQM-----	
PnrA	NDTFAVFGMCGVGHPPAQPAEIKPRLAQSIVLHRRERYEATEAEAVSVAAYDRRMSDFQHR	
	: : * : : : : :	
NitA	IGRAEKEINWSTFTSTIYQS---V-YPKVKGAINKQG--LK---TK	
NitB	-----ALNKLHYDKY---SK	
NfsA	RGSNNRRDTWSDHIRRTIIK---E-SRPFILDYLHKQGWATR---WA---TR	
Frp	RTSNQKLTWSQEVTKLAG---E-SRPHILPYLNSKGL--LA---KR	
nbzA	-----SISKMELEEFTTFVHE	
PnrA	QQREN--SWSSQAVERVKGADSLSGRHRLRDALNTLG--FG---LR	
	: : :	

NQ01	EEKNKKFGLSVGHHLGKSIPTDNQIKARK	273
NQ02	-----AHWHFGQ-----	232
ywrO	-----	
YieF	-----	

A3.11 Alignment of group IV protein sequences using T-Coffee

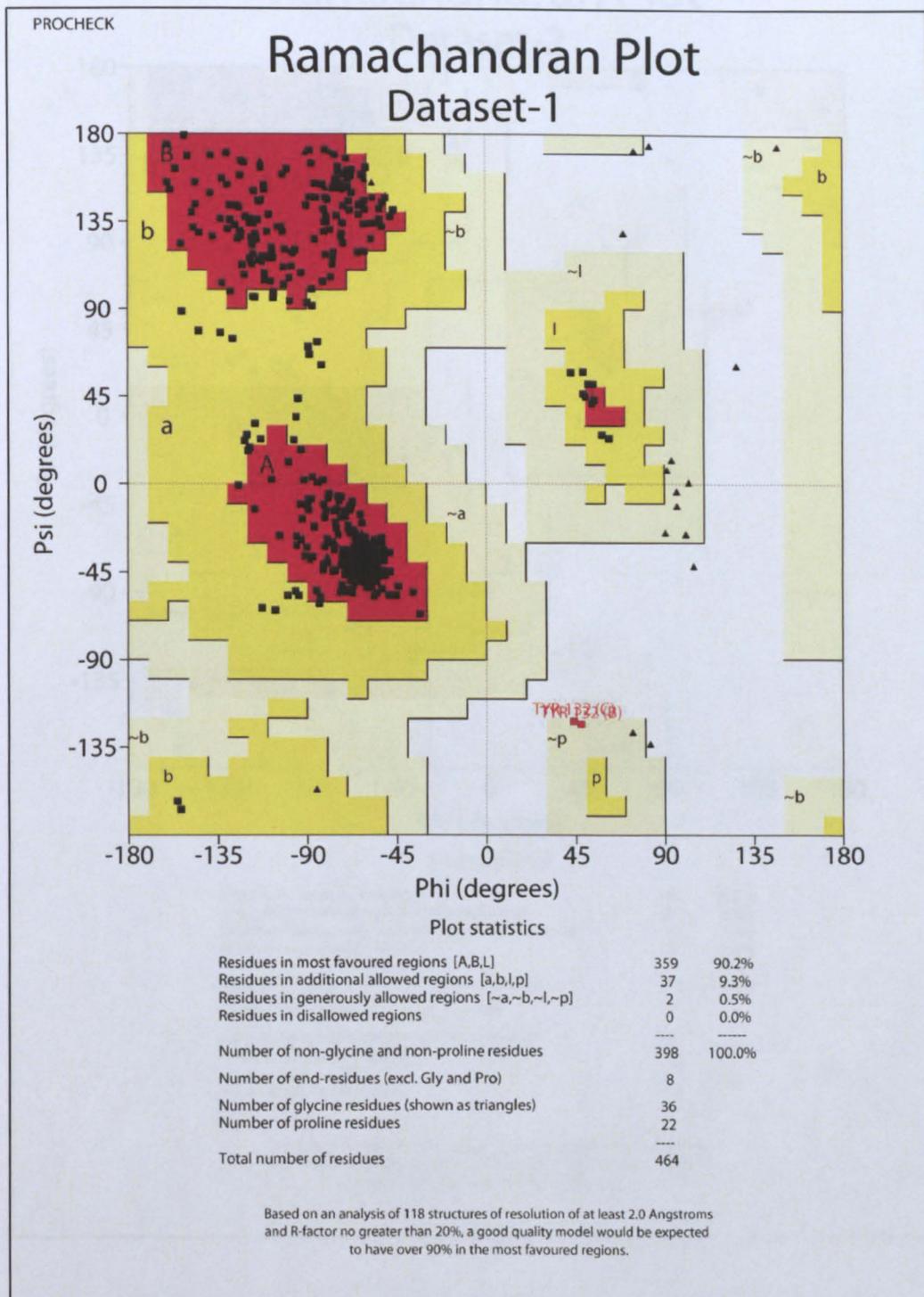
NQ01	---VGRRALIVLA--HSERTSFNYAMKEAAAAALKKKGWEVVESDLYAMNFPNII SRKDI
NQ02	GAMAGKKVLIVYA--HQEPKSFNGSLKNVADELRSQGCTVTVSDLYAMNFEPRATDKDI
YieF	---MSEKLQVVTLLGSLRKGSFNGMVARTL----PKIAPASME-----VNALPSIADI
ywrO	---M--KVLVLA FHPNMEQSVVNRADF-----TLKDAPGITLRDLQY-----
	: : : * :
NQ01	TGKLDKDPANFQYPAESVLAYKEGHLSPDIVAEQKKLEAADLVIFQFPLQWFGVPAILKGW
NQ02	TGTLNSPEVFNYSVETHEA-KQRSLASDITDEQKKVREADLVIFQFPLYWFSVPAILKGW
YieF	-----PLYDADVQQEEGFPATVEALAEQIRQADGVVIVTPEYNYVSPGGLKNA
ywrO	-----EYP-----DEAIDVEKEQKLCHEHDRIVFQFPLYWYSSPPLLKKW
	: : : * : : * : * * *
NQ01	F---ERVFIGEFAYTYAAMYDKGPFRRSKKAVLSITT---GGSGMSYSLQGIHG-DMNVI
NQ02	M---DRVLCQGFAPFDIPGFYDSGLLQKGLALLSVTT---GGTAEMYTKTG VNG-DSRYF
YieF	IDWLSRL-----PDQPLAGKPVLIQTSSMGVIGGARCQYHLR-----QI
ywrO	L---DHVLLYGWAYGT---NGTALRGKEFMVAVSA---GAPEEAYQAGGSNHYAISEL
	: : : : * : : : * * *
NQ01	L---WPIQSGILHFCGF---QVLEPQLTYSIGHTPADARIQILEGWKKRLENIWDETPLY
NQ02	L---WPLQHGTLHFCGF---KVLAPQISFAPETIASEEERKGMVAAWSQRLQTIWKEEP--
YieF	LVFLDAMVMNKPEFMGGVIGNKVDPPQTGEVIDQGTLDHLTGQLTAFGEFIQ-----
ywrO	L---RPFQ-ATSNFIGT---TYLPPYVIFYQAGTAGKSELAEGATQYRE-----
	* : * * : *
NQ01	FAPSSFLDLNFQAGFLMKKEVQDEEKNKKFGLSVGHHLGKSIPTDNQIKARK
NQ02	-----IPCTAHWHFGQ-----
YieF	-----RVKI-----
ywrO	-----HVLKSF-----

A3.12 Alignment of group IV protein sequences using PROMALS

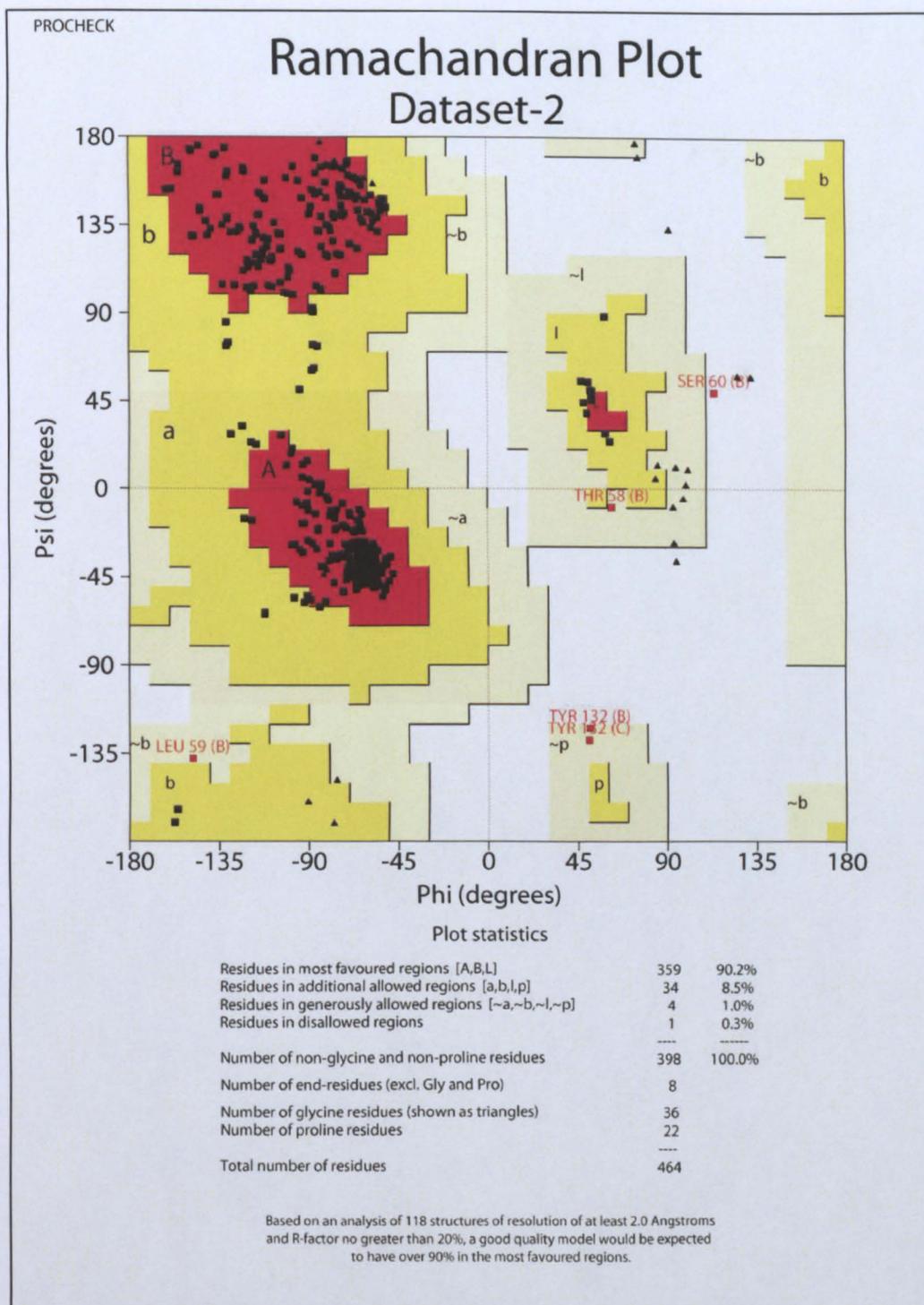
Conserv	55 5 5 5 5 555	
ywrO	--M---KVLVLA FHPNMEQ---SVVNRADFATLKDAPGITLRDLQYQYEPDEAI-----	45
NQ02	GAMAG-KKVLIVYAHQEPKSFNGSLKNVADELRSQGCTVTVSDLYAMNFEPRATDKDITGTLNSPEVFN	69
NQ01	---VGRRALIVLAHSERTSFNYAMKEAAAAALKKKGWEVVESDLYAMNFPNII SRKDI	66
YieF	---MSEKLQVVTLLGSLRKGSFNGMVARTLPKIAPASMEVNALPSIADIPLYDAD-----	52
consensus	eeeeee hhhhhhhhhhhhhhh eeeee hhh hhhhhh	
Conserv	55 555 59 5555595555559 99 5 5 5	
ywrO	-----DVEKEQKLCHEHDRIVFQFPLYWYSSPPLLKKWLDHVLLYGWAYGTN---GTA	95
NQ02	YGVETHEAYKQRSLASDITDEQKKVREADLVIFQFPLYWFSVPAILKGMWDRVLCQGFAPFDIPGFYDSGL	139
NQ01	YPAESVLAYKEGHLSPDIVAEQKKLEAADLVIFQFPLQWFGVPAILKGFERFVIFEFAYTYAAMYDKGP	136
YieF	-----VQEEGFPATVEALAEQIRQADGVVIVTPEYNYVSPGGLKNAIDWLSRLP-----DQP	105
consensus	hhhhhhh hhhhhhhhhhh eeeee hhhhhhhhhhh	
Conserv	5 59 5 55 5 55 5 5 5 5	
ywrO	LRGKEFMVAVSA GAPEEAYQAGGSNHYAISELLRPFQATSNFIGT TYLPPYVIFYQAGT-----AG	155
NQ02	LQKGLALLSVTTGGTAEMYTKTG VNGDSRYFLWPLQHGTLHFCGFKVLAPQISFAPEI-----AS	199
NQ01	FRSKKAVLSITGGSGMSYSLQGIHGDMNVILWP IQSGILHFCGFQVLEPQLTYSIGH-----TP	196
YieF	LAKGPLVLIQTSSMGVIG-----GARCQYHLRQILVFLDAMVMNKPEFMGGVIGNKVDPPQTGEVI	164
consensus	eeeeeee hhh hhh hhhhhhhhhhhhhhh eeeee	
Conserv		
ywrO	KSELAEGATQYREHVLKSF-----	174
NQ02	EEERKGMVAAWSQRLQTIWKEEIPCTAHWHFGQ-----	233
NQ01	ADARIQILEGWKKRLENIWDETPLYFAPSSFLDLNFQAGFLMKKEVQDEEKNKKFGLSVGHHLGKSIPTD	266
YieF	DQGTLDHLTGQLTAFGEFIQRVKI-----	188
consensus	hhhhhhhhhhhhhhhhhhhh	
Conserv		
ywrO	-----	
NQ02	-----	
NQ01	NQIKARK	273
YieF	-----	
consensus		

A4. Ramachandran plots

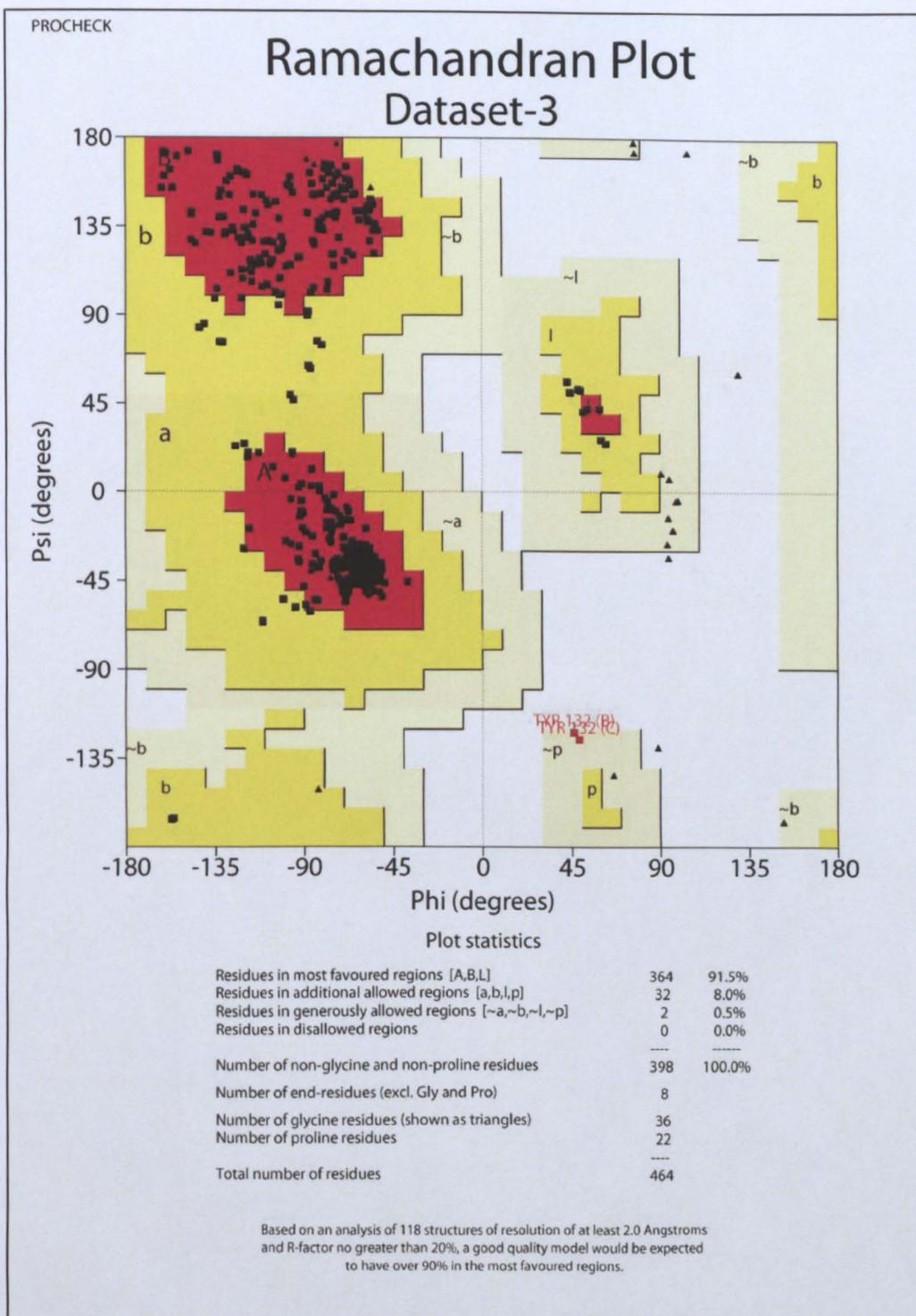
A4.1 Ramachandran plot of dataset one



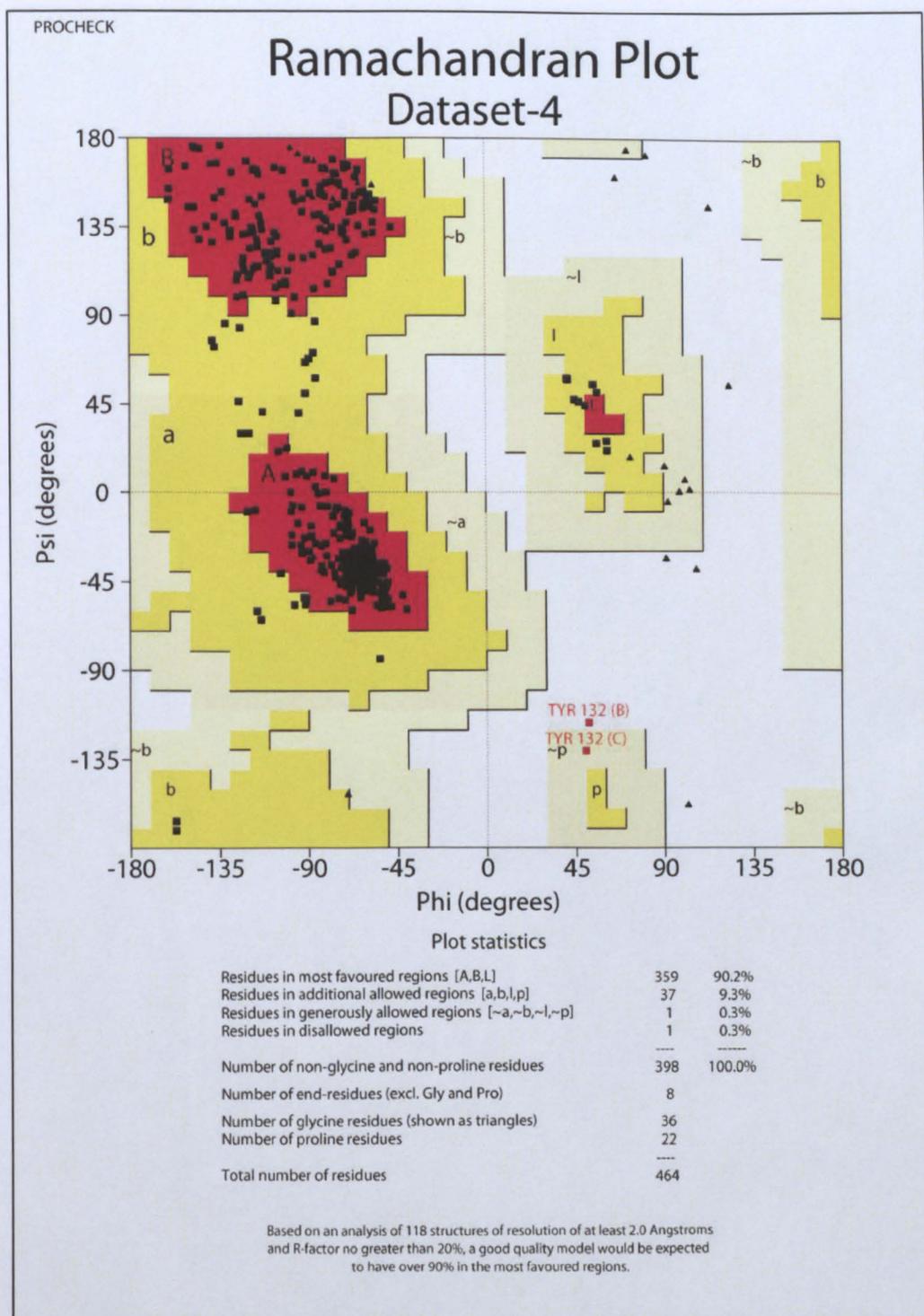
A4.2 Ramachandran plot of dataset two



A4.3 Ramachandran plot of dataset three



A4.4 Ramachandran plot of dataset four



A4.5 Ramachandran plot of dataset five

