



<https://theses.gla.ac.uk/>

Theses Digitisation:

<https://www.gla.ac.uk/myglasgow/research/enlighten/theses/digitisation/>

This is a digitised version of the original print thesis.

Copyright and moral rights for this work are retained by the author

A copy can be downloaded for personal non-commercial research or study,
without prior permission or charge

This work cannot be reproduced or quoted extensively from without first
obtaining permission in writing from the author

The content must not be changed in any way or sold commercially in any
format or medium without the formal permission of the author

When referring to this work, full bibliographic details including the author,
title, awarding institution and date of the thesis must be given

Enlighten: Theses

<https://theses.gla.ac.uk/>
research-enlighten@glasgow.ac.uk

**Structure based design of type II Dehydroquinase
inhibitors against *Mycobacterium tuberculosis***

David Alexander Robinson

Division of Biochemistry and Molecular Biology

&

Protein Crystallography Group

University of Glasgow

Glasgow, G12 8QQ

Scotland, UK

**This thesis is presented in submission for the degree of Doctor of Philosophy in
the Faculty of Biomedical and Life Sciences**

September 2003

© David Robinson 2003

ProQuest Number: 10646763

All rights reserved

INFORMATION TO ALL USERS

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

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



ProQuest 10646763

Published by ProQuest LLC (2017). Copyright of the Dissertation is held by the Author.

All rights reserved.

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

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

GLASGOW
UNIVERSITY
LIBRARY:

13296
COPY 1

Summary

There is currently an alarming increase in the world-wide incidence of tuberculosis (TB), the disease caused due to infection by *Mycobacterium tuberculosis*. There is no easy treatment for TB which currently requires a regime of four drugs administered over a six month time course. The rise in strains of *M. tuberculosis* resistant to one or more of these frontline anti-tubercular drugs, is a strong incentive to develop novel anti-tubercular agents.

The shikimate pathway has been identified as a potential target for the development of new antimicrobial drugs and knockout strains of *M. tuberculosis* lacking shikimate kinase have been shown to be non-viable. In this study the third enzyme in the pathway, type II dehydroquinase is considered an attractive target as many pathogenic bacteria possess the type II enzyme whereas many bacteria that inhabit the human gut possess the type I form. Structure based ligand design was used to identify potential inhibitors of the enzyme type II 3-dehydroquinase dehydratase (DHQase) from *M. tuberculosis* (MTDHQase). This methodology allows the identification of non-substrate like inhibitors that have solubility profiles more amenable to development as potential drugs. The crystal structures of MTDHQase available in house, were analysed to generate Pharmacophore models, which were subsequently used to screen a virtual compound library, to generate a list of potential inhibitors. This process was repeated for the type II DHQases from *Streptomyces coelicolor* (SCDHQase) and *Helicobacter pylori* (HPDHQase) and the hit lists for each enzyme compared. The hits unique to MTDHQase were screened *in vitro*, identifying six compounds that were active inhibitors.

In tandem the cloning and structural analysis of the type II 3-dehydroquinase from *Helocobacter pylori* was undertaken as a series of inhibitors of this enzyme obtained from a traditional high throughput screen (HTS) against a compound library was available from GlaxoSmithKline.

The structure of HPDHQase in complex with the transition state analogue 2,3 anhydro quinic acid was determined to 3.1Å resolution. While the structure of HPDHQase in complex with inhibitor AH9095, an HPDHQase specific non-substrate-like inhibitor, was determined to 1.5Å resolution. Both structures were solved using the molecular replacement method.

Comparison of the structures of HPDHQase with those of SCDHQase and MTDHQase provided significant insight into the factors that affect ligand specificity. The complex of HPDHQase and AH9095 provided the first crystal structure of a type II 3-dehydroquinase with a non-substrate like ligand. This structure allowed the identification of new, unexploited areas of the type II 3-dehydroquinase active site which may be utilised to develop specificity and potency of inhibitors.

The present assay for 3-dehydroquinase activity is based upon UV absorbance at 234nm from the product 3-dehydroshikimate. Detection of UV absorbance at this wavelength is not possible when the assay is carried out in plastic 96 well plates, therefore not amenable for high throughput methods. Development of a simple colorimetric assay for 3-dehydroquinase activity was investigated to permit high throughput testing of compounds. By coupling the dehydroquinase step to dehydroshikimate dehydratase, the third enzyme in the quinate degradation pathway, the product protochaechuate can be detected by a strong colour change in the

presence of ferric chloride or sodium molybdate. The amount of protocatechuate detectable by this method, in 96 well plate format, was investigated and expression studies were carried out upon the coupling enzyme dehydroshikimate dehydratase from *Acinetobacter coalceticus*.

Contents

	Page
Title Page	i
Summary	ii-iv
List of Contents	v-xii
List of Figures	xiii-xviii
List of Tables	xix-xx
Abbreviations	xxi-xxii
Declaration	xxiii
Acknowledgements	xxiv-xxv
Chapter 1. Introduction	1
1.1 Overview	1
1.2 Tuberculosis – from inhalation to infection	3
1.3.1 Current drug regimen	4
1.3.2 Modes of action of current drugs	5
1.4 Current targets in TB drug development	6
1.5 The Shikimate Pathway	7
1.5.1 The Shikimate Pathway as a drug target	10
1.6 3-Dehydroquinate Dehydratase	11
1.6.1 Two types of Dehydroquinase	12
1.6.2 Crystal structures of type I and type II Dehydroquinases	13
1.6.3 Type I Dehydroquinase	15
1.6.4 Type II Dehydroquinase	16
1.6.5 Selective inhibition of Dehydroquinase	21
1.7 The drug discovery process	22
1.7.1 Target selection and validation	22
1.7.2 Hit identification to lead optimisation	24
1.7.3 Toxicology and clinical trials	25
1.8 The impact of structural biology on the drug discovery process	26
1.8.1 Structure based drug design	27
1.9 Aims	28

Chapter 2. Materials and Methods	31
2.1.1 Chemicals	31
2.1.2 Proteins and Enzymes	31
2.1.3 Oligonucleotides	31
2.1.4 Chromatography Media	31
2.2 Media and Supplements	32
2.2.1 Media for Bacterial Growth	32
2.2.2 Antibiotics	32
2.2.3 Isopropyl- β -D-thiogalactopyranoside (IPTG)	33
2.3 General Molecular Biology Methods	33
2.4 Bacterial strains and cloning vectors	34
2.4.1 Storage of Bacterial Strains	34
2.4.2 Bacterial strains	34
2.4.3 Growth of Plasmid containing <i>E. coli</i>	34
2.4.4 Growth of cells and isolation of plasmids	34
2.4.5 Transformation protocol	35
2.5 Manipulations of DNA	37
2.5.1 Restriction digests	37
2.5.2 Gel electrophoresis	37
2.5.3 Gel Purification of DNA fragments	37
2.5.4 DNA Ligation	37
2.6 Cloning by Polymerase Chain Reaction (PCR)	38
2.6.1 Primers for PCR	38
2.6.2 PCR reactions	41
2.6.3 Cloning Methods	42
2.6.3.1 Blunt Cloning	42
2.6.3.2 TOPO [®] Cloning	42
2.7 Over-expression of recombinant type II Dehydroquinases	43
2.8 Purification	43
2.8.1 Purification by Affinity Chromatography	44
2.8.1.1 MBP affinity chromatography	44
2.8.1.2 Ni-NTA affinity chromatography	44
2.8.2 Ion exchange chromatography	45

2.8.3.	Size exclusion chromatography	45
2.9	Protein quantification and analysis	46
2.9.1	SDS-PAGE	46
2.9.2	Dynamic Light Scattering	46
2.9.3	Protein concentration calculation using Absorbance at 280nm (A_{280})	47
2.10	Standard Dehydroquinase Assay	48
2.11	Inhibitors of type II Dehydroquinase	48
2.11.1	Synthesis of GR122160X	50
2.12	Crystallisation of biological macromolecules	50
2.12.1	Crystallisation by sitting drop vapour diffusion	51
2.12.2	Crystallisation Screens	51
Chapter 3 Wet lab results		53
3.1	Cloning, expression and purification of type II DHQases as Maltose Binding Protein Fusion	53
3.1.1	PCR amplification and sub-cloning of <i>aroQ</i> from <i>H. pylori</i>	54
3.1.2	Identification of positive pMAL-haQ clones	55
3.1.3	Over-expression of <i>H. pylori</i> and <i>M. tuberculosis</i> DHQase as a MBP fusion	56
3.1.4	Purification of MBP-DHQase fusion peptides	57
3.1.5	Factor Xa Cleavage of MBP-DHQase fusion peptides	58
3.1.6	Separation of DHQase from MBP	59
3.2	Crystallisation of type II DHQases expressed as MBP fusion peptides	61
3.2.1	Crystallisation of MTDHQase	62
3.2.2	Crystallisation of HPDHQase	63
3.2.3	Reasons for poor diffraction	64
3.3	Cloning, expression and purification of type II Dehydroquinases with an N-terminal poly-histidine tag	65
3.3.1	Sub-cloning of <i>M. tuberculosis aroD</i> and <i>H. pylori aroQ</i> into pT7-Blue3	65

3.3.2	Sub-cloning of <i>M. tuberculosis aroD</i> and <i>H. pylori aroQ</i> into pET-15b	66
3.3.3	Expression studies upon 6xHis tagged DHQase	66
3.3.	Purification of 6xHis tagged DHQases	68
3.4	Crystallisation of 6xHis tagged DHQases	69
3.4.1	Crystallisation of 6xHis-MTDHQase	69
3.4.2	Crystallisation of 6xHis-HPDHQase	69
3.4.2.1	Crystallisation of HPDHQase and FA1	70
3.4.2.2	Crystallisation of HPDHQase and AH9095	71
3.4.2.3	Crystallisation of HPDHQase and GR122110X	72
3.4.2.4	Crystallisation of HPDHQase and Phosphate	73
3.4.3	Soaking experiments with HPDHQase	74
3.5	Discussion	75
Chapter 4 Collection and Processing of X-ray diffraction data		76
4.1	Protein Crystallography	76
4.1.1	Scattering of X-rays by an atom	76
4.1.2	Bragg's Law	77
4.1.3	The Ewald Construction	78
4.2	Collection of X-ray diffraction data by the Rotation Method	79
4.2.1	The Rotation Method	80
4.2.2	Crystal Mosaicity	82
4.3	Data Collection Strategy	83
4.4	Collection of X-ray diffraction data from <i>Helicobacter pylori</i> type II Dehydroquinase + AH9095	84
4.4.1	Initial data collection of HPDHQase + AH9095 using a rotating anode source	84
4.4.2	Determination of data collection strategy for HPDHQase + AH9095	85
4.4.3	Processing of data - HPDHQ + AH9095	87
4.5	High resolution data collection – HPDHQase + AH9095	89
4.5.1	Synchrotron radiation sources	90
4.5.2	Advantages of synchrotron radiation	90

4.5.3	High resolution data collection – HPDHQase + AH9095	91
4.6	Collection of diffraction data from other HPDHQase crystals	94
4.6.1	HPDHQase + GR78752X Soak	94
4.6.2	HPDHQase + FA1	95
4.6.3	HPDHQase + GR122110	98
4.7	Structure Factor Equation	100
4.8	Conclusion	101

Chapter 5 Structure solution of *H.pylori* DHQase by

Molecular Replacement	102	
5.1	Molecular Replacement	102
5.1.1	The Rotation Function	104
5.1.2	The Translation Function	105
5.1.3	Multidimensional Molecular Replacement	105
5.1.4	Molecular Replacement by Evolutionary Search	106
5.2	Attempts to solve HPDHQase structure using AmoRe	107
5.3	Structure solution of HPDHQase using EPMR	109
5.4	Removal of phase bias by Density Modification	110
5.5	Structure solution of HPDHQase + FA1 by molecular replacement	111
5.6	Conclusion	113

Chapter 6 Refinement and Model Building of HPDHQase + AH9095

6.1	Refinement of Macromolecular Models	115
6.2	Electron Density Maps	117
6.3	Manual Alteration of the Model	118
6.4	Refinement and model building of HPDHQase + AH9095	119
6.5	Refinement and model building of HPDHQase + AH9095 with high resolution data	120
6.5.1	Automated addition of solvent molecules	122
6.5.2	Restrained Anisotropic B-factor Refinement	122
6.6	Quality of the final HPDHQase + AH9095 model	123
6.7	Conclusion	128

Chapter 7 Structural analysis of the type II Dehydroquinase from <i>Helicobacter pylori</i>	130
7.1 Crystal Structure of type II DHQase from <i>Helicobacter pylori</i>	130
7.2.1 A comparison of the monomer folds of the type II DHQases from <i>Helicobacter pylori</i> and <i>Mycobacterium tuberculosis</i>	133
7.2.2 Similarity of HPDQase monomer to other type II DHQase monomers	135
7.3 Type II dehydroquinase from <i>Helicobacter pylori</i> – Quaternary Structure	136
7.3.1 The trimer interface	136
7.3.2 The dimer interface	140
7.3.3 Contribution from 6x His tag	141
7.4 Ligand binding of AH9095	143
7.5 Implications of ligand binding mode	145
7.6 Rationalisation of the specific inhibition by AH9095 of HPDQase and not MTDQase	148
7.7 Crystal Structure of HPDQase + 2,3-anhydroquinic acid (FA1)	151
7.7.1 Refinement and Model Building of HPDQase + FA1	151
7.7.2 Structure of HPDQase + FA1	152
7.8 Discussion	156
7.8.1 Design of specific inhibitors of type II DHQase from <i>M. tuberculosis</i> based upon AH9095	158
Chapter 8 Structure Based Design of <i>Mycobacterium tuberculosis</i> type II Dehydroquinase inhibitors	160
8.1 Introduction	160
8.2 Structural validation of selectivity shown by Transition State Analogues	161
8.2.1 Type II DHQase ligand binding	164
8.2.2 Specificity of 3-hydroxyimino quinic acid (FA6) towards MTDQase	166
8.2.3 Considerations for design of Pharmacophores based upon DHQase/ligand complexes	168

8.3	Generation of Pharmacophore models for structure based design	171
8.3.1	Definition of the Active Site	171
8.3.2	Generation of a ligand interaction map using LUDI	174
8.3.3	Generation of multiple Pharmacophore models	174
8.4	Database searching using Catalyst	177
8.5	Assay to determine inhibition of MTDHQase by <i>in silico</i> screen hits	178
8.5.1	Activity assay – first round	178
8.5.2	Activity assay- second round	186
8.6	Predicted binding modes of active compounds	187
8.7	Discussion	191
Chapter 9 Development of a novel assay to follow activity of Type II DHQase		192
9.1	Need for a new DHQase assay	192
9.2	Dehydroshikimate dehydratase	194
9.3	Absorbance spectra of 3-dehydroquininate and PCA	195
9.4	Detection of PCA using Fe(III) Cl ₃ as an indicator	196
9.5	Optimisation of Fe(III) Cl ₃ /PCA concentrations required to detect PCA in 96 – well plate format	197
9.6	Detection of PCA using Na ₂ MoO ₃ as an indicator	199
9.7	Optimisation of Na ₂ MoO ₃ /PCA concentrations required to detect PCA in 96 – well plate format	200
9.8	Cloning of Dehydroshikimate dehydratase	202
9.9.1	Amplification of <i>quiC</i> by PCR	202
9.9.2	TOPO [®] cloning of <i>quiC</i>	203
9.9.3	Subcloning of <i>quiC</i> into pET-15b	204
9.9.4	Expression of <i>quiC</i> in <i>E. coli</i> BL21	205
9.9.5	Analysis of <i>quiC</i> expression by SDS PAGE	207
9.10	Conclusion	208
Chapter 10 Conclusion and General Discussion		209

References	213
Appendix A	228
Appendix B	240

Figures

		Page
Figure 1.1	Electron micrograph of a single <i>M. tuberculosis</i> bacillus	1
Figure 1.2	Estimated world wide incidence of TB.	2
Figure 1.3	Frontline drugs used in the treatment of tuberculosis	4
Figure 1.4	Schematic diagram of the mycobacterial cell wall.	5
Figure 1.5	Schematic diagram of the enzymatic steps that form the shikimate pathways	9
Figure 1.6	Conversion of 3-dehydroquinate to 3-dehydroshikimate, the reaction catalysed by DHQase	11
Figure 1.7	Dehydroquinase is active in the shikimate and quinate pathway	11
Figure 1.8	The enzyme folds of type I and II Dehydroquinases shown in ribbon representation.	14
Figure 1.9	Proposed reaction mechanism for type I DHQase	15
Figure 1.10	Proposed reaction mechanism for type II DHQase	16
Figure 1.11	A sequence alignment of type II DHQases from representative prokaryotic and eukaryotic organisms.	18
Figure 1.12	SCDHQase with 2,3-anhydroquinic acid bound in the active site.	20
Figure 1.13	Reaction mechanism for type II DHQase	21
Figure 1.14	General 'pipeline' diagram illustrating the stages of the drug discovery process.	22
Figure 1.15	Flow chart illustrating the structure based process to be followed in this thesis.	30
Figure 2.1	Formation of conjugated system in the conversion of 3-dehydroquinate to 3-dehydroshikimate	48
Figure 2.2	Oxidation of 5-methyl-[1,3,4]thiadiazole-2-thiol to form GR122160X	50

Figure 3.1	Agarose gel showing result of PCR amplification of <i>H. pylori aroQ</i> .	54
Figure 3.2	SDS PAGE time course of MBP-MTDHQase expression after induction with 1mM IPTG	56
Figure 3.3	SDS PAGE of MBP-HPDHQase purification.	57
Figure 3.4	SDS PAGE showing results of MBP-MTDHQase cleavage with Factor Xa.	58
Figure 3.5a	Separation of MBP/MTDHQase by ion exchange chromatography monitored by SDS PAGE,	59
Figure 3.5b	Separation of MBP/MTDHQase by size exclusion chromatography and heat shocking monitored by SDS PAGE.	60
Figure 3.6	SDS PAGE monitoring separation by ion exchange chromatography of MBP/HPDHQase,	60
Figure 3.7	Crystal of MTDHQase + FA4 in 15% (w/v) PEG 8K, 0.2M NaCl, 0.1M Tris-HCl pH 8.8	62
Figure 3.8	Crystal of HPDHQase + FA1 grown in 14% (w/v) PEG 4K, 0.2M MgCl ₂ , 0.1M Tris-HCl pH 8.8	63
Figure 3.9	Monitoring of expression of 6xHis- DHQases by SDS PAGE.	67
Figure 3.10	Monitoring purification of 6xHis-HPDHQase by SDS PAGE.	68
Figure 3.11	Crystal of HPDHQase + FA1 grown in 30% 1,4 butandiol, 0.1M Na Acetate pH 4.5	70
Figure 3.12	Crystal of HPDHQase + AH9095 grown in to 1.0M Na Acetate, 0.1M Imidazole-HCl pH 8.2	71
Figure 3.13	Crystals of HPDHQase + GR122110X grown in 20% 1,4 butandiol, 0.1M Na Acetate pH 4.5	72
Figure 3.14	Crystals of HPDHQase + 10mM Ammonium Phosphate grown in 1M Ammonium Acetate, 0.1M Tris-HCl pH 7.5	73

Figure 4.1	Electromagnetic radiation scattered by a single atom	77
Figure 4.2	Scattering from a periodic arrangement of atoms produces a characteristic diffraction pattern	77
Figure 4.3	The Bragg construction describing reflection of X-rays from planes of atoms within a crystal.	78
Figure 4.4	The Ewald construction	79
Figure 4.5	Rotation of the reciprocal lattice with respect to the Ewald sphere.	80
Figure 4.6	Rotation of the Ewald sphere around an axis normal to the plane of the paper	81
Figure 4.7	Illustration of a macromolecular crystal with high mosaic spread	82
Figure 4.8	A typical image of the diffraction pattern measured from a crystal of HPDHQase + AH9095 using the rotation method with a rotating anode source.	85
Figure 4.9	Spot profiles for and overloaded reflection and a well defined reflection.	86
Figure 4.10	Spot integration using Denzo	87
Figure 4.11	A typical diffraction pattern from HPDHQase + AH9095 collected at SRS beamline 9.6 using the ADSC Quantum-4r detector	92
Figure 4.12	A typical diffraction image from a HPDHQase crystal soaked in GR78752X recorded on beamline 9.5 at SRS using the MAR-165 CCD detector	94
Figure 4.13	A typical diffraction image from a HPDHQase + FA1 crystal recorded on beamline 9.5 at SRS using the MAR-165 CCD detector.	96
Figure 4.14a	A typical diffraction image from a HPDHQase + GR122110 crystal recorded on beamline 9.6 at SRS using the ADSC Quantum 4r CCD detector.	98
Figure 4.14b	Diffraction spots with small d-spacing collected from a crystal of HPDHQase + GR122110.	99

Figure 5.1	Showing two tri-atomic molecules related by a two-fold rotation axis and the resulting Patterson function	104
Figure 5.2	Outline of the procedure carried out by EPMR for molecular replacement.	107
Figure 5.3	Graph showing Correlation Coefficient and R-factor for 10 cycles of evolutionary searching using EPMR.	109
Figure 5.4	Packing of monomers within the F23 unit cell of HPDHQase generated from the MR solution.	110
Figure 5.5	2Fo-Fc electron density map calculated after density modification	111
Figure 5.6	Crystal packing of HPDHQase + FA1 within the P3 ₁ unit cell.	113
Figure 6.1	2Fo-Fc electron density map of HPDHQase + AH9095 centred upon active site region.	120
Figure 6.2	2Fo-Fc electron density map around AH9095	124
Figure 6.3	2Fo-Fc electron density map around strand 3 of HPDHQase	124
Figure 6.4	Ramachandran plot generated by PROCHECK showing the main chain torsion angles of HPDHQase + AH9095.	126
Figure 7.1	Ribbon representation of HPDHQase monomer.	131
Figure 7.2	HPDHQase dodecamer viewed down the crystallographic 3-fold axis and down the crystallographic 2-fold axis	132
Figure 7.3	Stereo pair diagrams of C α traces of HPDHQase monomers superimposed upon MTDHQase, SCDHQase and BSDHQase	134
Figure 7.4	Trimer and dimer interfaces of HPDHQase	137
Figure 7.5	HPDHQase trimer and dimer interfaces	138
Figure 7.6	Interactions formed by the 6xHis tag and linker	142
Figure 7.6	Structure of ligand AH9095	143
Figure 7.7	Diagram illustrating AH9095 bound in the active site region of HPDHQase.	143

Figure 7.8	Active site of HPDHQase with AH9095 bound and dehydroquinone modelled into binding pocket.	146
Figure 7.9	Interaction between Arg15, Glu92* and Asp67* in MTDHQase	149
Figure 7.10	Stereo surface representations of HPDHQase and MTDHQase with AH9095 bound in the active site	150
Figure 7.11	Quaternary structure of HPDHQase + FA1	153
Figure 7.12	Monomer fold of HPDHQase + FA1.	153
Figure 7.13	Stereo pair diagram showing a detailed view of the HPDHQase active site with FA1 bound.	155
Figure 7.14	General template for inhibitors of type II DHQases based upon AH9095.	157
Figure 8.1	Detailed view of MTDHQase active site with the inhibitorFA1 bound.	165
Figure 8.2	Active site of SCDHQase with FA6 modelled into active site based upon superposition of SCDHQ+FA1 and MTDHQ+FA1 complexes	167
Figure 8.3	Sequence alignment of type II DHQases.	169
Figure 8.4	Ribbon diagram of an SCDHQase monomer illustrating the key areas for Pharmacophore design.	170
Figure 8.5	Structure based design process using the Accelrys software packages CeriusII and Catalyst	172
Figure 8.6a	Definition of the active site of MTDHQase.	173
Figure 8.6b	Definition of a sphere of radius 10Å around the centre of the MTDHQase active site.	173
Figure 8.7a	Interaction map generated by LUDI for MTDHQase.	175
Figure 8.7b	Centroids of interaction clusters in MTDHQase.	175
Figure 8.8	A Pharmacophore model generated for MTDHQase	176
Figure 8.9	%V/V _{max} measured for 30 compounds against MTDHQase.	186
Figure 8.10	Proposed binding modes of MTDHQase inhibitors.	190

Figure 9.1	Conversion of 3-dehydroquinone to 3-dehydroshikimate by dehydroquinase (DHQase)	193
Figure 9.2	Absorbance spectra in the range 200-800nm of 3-dehydroquinone, 3-dehydroshikimate and PCA	195
Figure 9.3	Absorbance spectra in the range 200 – 800nm for Fe(III) Cl ₃ , Fe(III) Cl ₃ + 3-dehydroquinone, Fe(III) Cl ₃ + 3-dehydroshikimate and Fe(III) Cl ₃ + PCA	196
Figure 9.4	Expansion of absorbance spectra in the range 400 – 800nm for Fe(III) Cl ₃ , Fe(III) Cl ₃ + 3-dehydroquinone, Fe(III) Cl ₃ + 3-dehydroshikimate and Fe(III) Cl ₃ + PCA	196
Figure 9.5	96-well plate illustrating the colour change observed upon addition of the Fe(III) Cl ₃ to PCA at various concentrations	198
Figure 9.6	Plot of A ₅₉₀ vs PCA concentration using various concentrations of Fe(III) Cl ₃	198
Figure 9.7	Absorbance spectra in the range 200 – 800nm for Na ₂ MoO ₃ , Na ₂ MoO ₃ + 3-dehydroquinone, Na ₂ MoO ₃ + 3-dehydroshikimate and Na ₂ MoO ₃ +PCA	199
Figure 9.8	Expansion of absorbance spectra in the range 350 – 600nm for Na ₂ MoO ₃ , Na ₂ MoO ₃ + 3-dehydroquinone, Na ₂ MoO ₃ + 3-dehydroshikimate and Na ₂ MoO ₃ + PCA	199
Figure 9.9	96-well plate illustrating the colour change observed upon addition of Na ₂ MoO ₃ to PCA at various concentrations	201
Figure 9.10	Plot of A ₄₀₅ vs PCA concentration using various concentrations of Na ₂ MoO ₃	201
Figure 9.11	Gel electrophoresis of <i>quiC</i> PCR.	203
Figure 9.12	SDS PAGE analysis of <i>quiC</i> expression.	207

Tables

	PAGE	
Table 1.1	References and PDB codes for structures shown in figure 1.5	8
Table 2.1.	Bacterial strains used for general cloning purposes	36
Table 2.2.	Plasmids used for general cloning	36
Table 2.3.1.	Oligonucleotides used for molecular cloning – PCR primers	39
Table 2.3.2.	Oligonucleotides used for molecular cloning - Custom sequencing primers	40
Table 2.4.	Inhibitors of type II DHQase used in this thesis	49
Table 2.5.	Sparse matrix crystallisation screen	52
Table 2.6.	Ammonium sulphate grid screen	52
Table 4.1.	Statistics for HPDHQase + AH9095 data collected using in-house rotating anode source	89
Table 4.2.	Reflection conditions for the space group <i>F23</i>	92
Table 4.3.	Statistics for HPDHQase + AH9095 data collected on SRS beamline 9.6	93
Table 4.4.	Data collection statistics for HPDHQ + GR78752 soak	95
Table 4.5.	Data collection statistics for HPDHQ + FA1	96
Table 4.6.	Scaling statistics for HPDHQase + FA1	97
Table 4.7.	Data collection statistics for HPDHQ + GR122110	99
Table 6.1.	Geometric indicators of the quality of the final HPDHQase + AH9095 model	127
Table 6.2.	B-factor statistics for final HPDHQase + AH9095 model	127

Table 6.3.	Refinement procedure for HPDHQase + AH9095	128
Table 7.1.	RMS deviation between atomic positions of HPDHQase with other type II DHQase structures	135
Table 7.2.	H-bond interactions at the AB trimer interface of HPDHQase	139
Table 7.3.	H-bond interactions at the dimer interface of HPDHQase	140
Table 7.4.	H-bond interactions between AH9095 and HPDHQase	145
Table 7.5.	Refinement statistics for HPDHQase + FA1.	152
Table 7.6.	H-bond interactions between FA1 and HPDHQase	155
Table 7.7.	Potential MTDHQase specific ligands based upon AH9095	159
Table 8.1.	Inhibition statistics for a series of transition state analogues against different DHQases	162
Table 8.2.	Protein-ligand hydrogen-bond distances and angles for the MTDHQase - FA1 complex	165
Table 8.3.	Number of hits identified by Catalyst searching Maybridge '99 database using type II DHQase pharmacophore models	177
Table 8.4.	Compounds identified by structure based design that were not available for purchase	179
Table 8.5.	Compounds identified by structure based design to be inhibitors of MTDHQase	185
Table 8.6.	Second round assay results for top six compounds against MTDHQase	188

Abbreviations

ADP	Adenosine 5'-diphosphate
ANDHQase	3-dehydroquininate dehydratase from <i>Aspergillus nidulans</i>
ATP	Adenosine 5'-triphosphate
BSA	Bovine serum albumin
BSDHQase	3-dehydroquininate dehydratase from <i>Bacillus subtilis</i>
CCD	Charge-coupled device
DESY	Deutsche Elektronen Synchrotron, Hamburg
DAHP	3-Dehydroxy-D-arbino heptulose 7-phosphate
DHQ	3-dehydroquininate
DHQase	3-dehydroquininate dehydratase
DHS	3-dehydroshikimate
DHSase	3-dehydroshikimate dehydratase
DMSO	Dimethyl sulphoxide
E4P	Erythrose 4-phosphate
EDTA	Ethylenediaminetetracetic acid
EMB	Ethambutol
EPSP	5-enoylpyruvate shikimate 3-phosphate
ESRF	European Synchrotron Radiation Facility, Grenoble
FA1	2,3 anhydro quinic acid
FA3	3-fluoro-2,3 anhydro quinic acid
FA4	3,3 difluoro quinic acid
FA6	3 hydroxyimino quinic acid
GSK	GlaxoSmithKline
HPDQHase	3-dehydroquininate dehydratase from <i>Helicobacter pylori</i>
INH	Isoniazid
MAD	Multi-wavelength anomalous dispersion
MBP	Maltose Binding Protein
MPD	Methyl pentane diol

MR	Molecular replacement
MTDHQase	3-dehydroquininate dehydratase from <i>Mycobacterium tuberculosis</i>
Ni-NTA	Nickel nitrotriacetic acid
PCR	Polymerase Chain Reaction
PBS	Phosphate buffered saline
PCA	Protocatechuate
PEG	Polyethylene glycol
PEP	Phosphoenol pyruvate
PZA	Pyrizanamide
RMP	Rifampicin
SAD	Single wavelength anomalous dispersion
SBD	Structure Based Design
SCDHQase	3-dehydroquininate dehydratase from <i>Streptomyces coelicolor</i>
SDS	Sodium dodecyl sulphate
SDS PAGE	Sodium dodecyl sulphate polyacrylamide gel electrophoresis
SRS	Synchrotron Radiation Source, Daresbury
TB	Tuberculosis
TIM	Triose phosphate isomerase
TRIS	Tris(hydroxymethyl)aminomethane

Declaration

I declare that the work recorded in this thesis is entirely my own, unless otherwise stated and that it is of my own composition. No part of this work has been submitted for any other degree

Acknowledgements

I would like to thank my co-supervisors Professor John Coggins and Dr. Adrian Laphorn from the University of Glasgow, and Dr. Peter Chalk from GlaxoSmithKilne, Stevenage, UK for their guidance and patience. I would also like to thank Dr. Ken Duncan from GlaxoSmithKline / Action TB for funding the project.

Special thanks to Professor Nick Price for guidance not only in the ways of science but also written English.

Thanks to Chris Abell and members of his lab at Cambridge for the kind gifts of inhibitor compounds and their generous hospitality. I would also like to thank Iain Hunter of the University of Strathclyde for allowing me to carry out experiments within his lab.

Thanks to all the members of the Coggins lab, especially (Dr) Bernhard Lohkamp, Dr Lewis DB Evans and Mr. John Greene. Thanks to all the members of the Glasgow Protein Crystallography Group who I have worked with throughout the past years, especially Paul Emsley, Alan Riboldi-Tunncliffe, Colin Bent, Claire Martin, Ally McEwan, Kirsty Stewart and Aleks Roszak. Thanks also to Gordon Barr, Graeme Tate and Stuart Mackay.

Thanks to all at (soon to be ex) Millennium Pharmaceuticals R&D Ltd in Cambridge for help and abuse over the past year especially Dina Fotinou and Martin Fisher. Thanks to Mark Stevens, and Mel Wood for giving me competition for getting the thesis in on time, and thanks to the Millennium social club including Neil Hunt, Fran Neal, Kyla Thompson, Kerry 'box of matches and a can of petrol' Jenkins, Rob Judkins, Jez Major, Julia Atkin and Giles Pergl-Wilson.

Special thanks to all those who have helped and listened/switched off through my many crises; Andrea and Blair Johnston, Nicky Meenan, Derek Blair, Craig Spence, Craig Egdell and David Donnelly.

Special thanks to Gwyndaf Evans for persuading me that I know some crystallography really, then going out to the pub when I should have been writing about it.

I am indebted to Dr. Samantha Rutherford for helping me through the rigours of not only my PhD, but all aspects of life.

Finally I would like to thank my parents and my sister Alison for their constant support and love.

Chapter 1 – Introduction

1.1 Overview

The organism *Mycobacterium tuberculosis* (Figure 1.1) is the causative agent of the disease tuberculosis (TB), one of the most common infectious diseases known to man. Current statistics estimate that around 32% of the world's population (1.86 billion people) are infected with *M. tuberculosis* of which approximately 8 million people will develop active TB. Of those suffering from active TB, 2 million people will die (WHO, 2003).

Figures from the World Health Organisation (WHO) show that the areas where TB is most prevalent are sub-Saharan Africa, the former Soviet Union and the far east (Figure 1.2), however in the developed world cases of TB are on the rise.



Figure 1.1: Electron micrograph of a single *M. tuberculosis* bacillus

Despite increasing numbers of multi drug resistant strains of *M. tuberculosis* no new anti-tubercular agents have been developed in over 30 years. As we begin to understand the cellular, molecular and structural biology of *M. tuberculosis* the search for novel anti-tubercular agents has accelerated in response to the growing re-emergence of the 'white plague'.

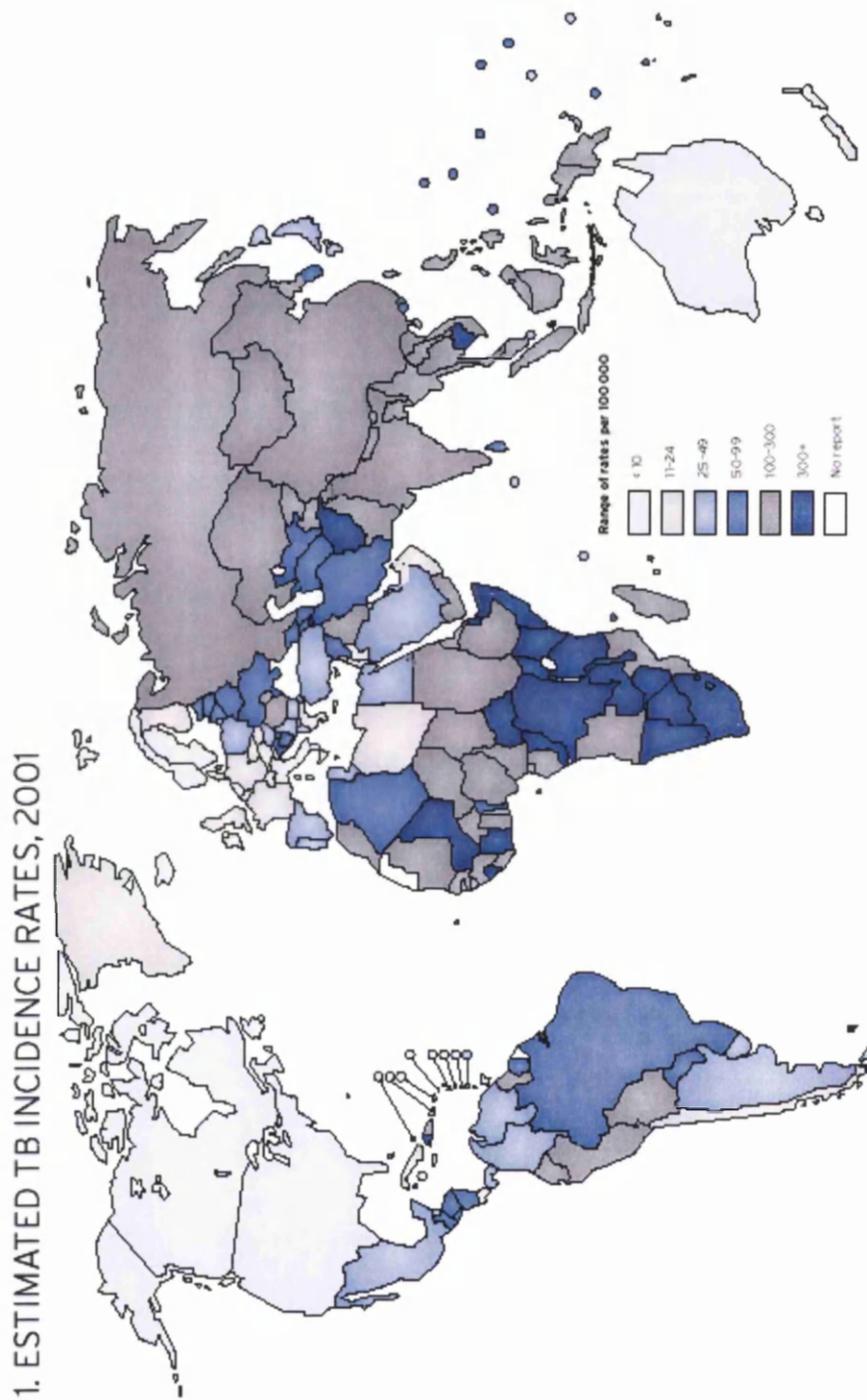


Figure 1.2: Estimated world wide incidence of TB. Figure from WHO (2003)

1.2 Tuberculosis – from inhalation to infection

M. tuberculosis passes from host to host via inhalation of aerosols containing the bacilli. In most cases the bacilli are ingested by alveolar phagocytic macrophages and are killed or grow intracellularly in small lesions called tubercles. At this point of infection, a disease state referred to as primary tuberculosis, the host shows few obvious symptoms relating to the infection. Despite the actions of the primary immune response, some viable *M. tuberculosis* bacilli persist within the host from which the disease can develop.

When the infection activates, the bacilli spread from the initial point of infection via the blood or lymphatic system to infect other areas of the body. At this point the disease is referred to as post-primary tuberculosis. Despite a large number of bacilli being killed, the number of phagocytes killed is also significant, resulting in solid necrosis in the lung in which the bacilli can survive. At this point the infection can be arrested by the immune system although a large amount of tissue damage will remain. If the necrosis expands to form a cavity within the lung, large numbers of bacilli will multiply which can then be transmitted to the outside world via coughing. As the disease progresses further the solid necrosis may liquefy, resulting in a rich medium for the proliferation of the bacilli.

The symptoms of TB; weakness, fever, coughing and severe chest pain are a result of the pathological and inflammatory response. Bloody sputum is observed when blood vessels within the lungs become eroded and burst. In cases where patients are immuno-compromised, such as AIDS patients, the onset of disease is much more rapid, usually resulting in fatality.

1.3.1 Current drug regimen

The current standard chemotherapeutic regimen for TB involves a combination of 4 drugs (Figure 1.3) over a time period of six months split into 2 phases. The first phase, which runs for 2 months involves the administration of isoniazid (INH), rifampicin (RMP), pyrazanimide (PZA) and ethambutol (EMB), is followed by 4 months of INH and RMP only (Global Alliance for TB drug development, 2001).

RMP shows poor bioavailability in some multi-drug formulations and is prohibitively expensive for many developing countries. Therefore in some areas RMP is replaced by thiacetozone.

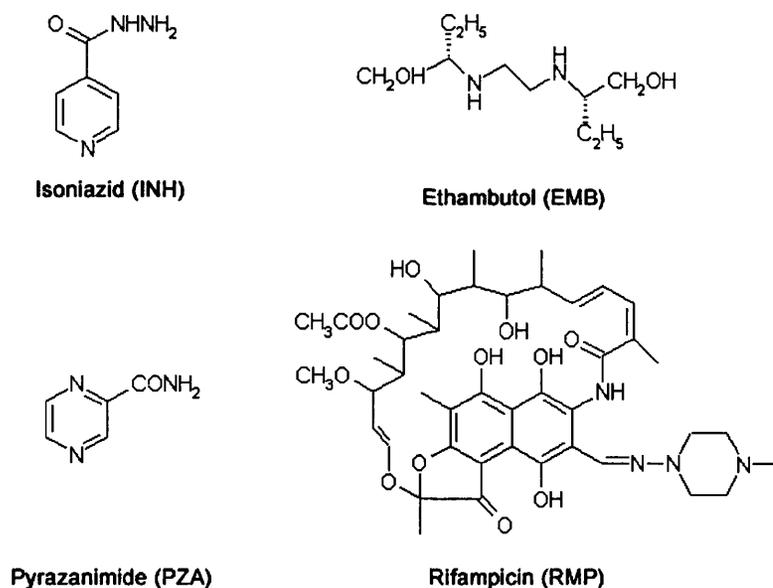


Figure 1.3: Frontline drugs used in the treatment of tuberculosis

1.3.2 Modes of action of current drugs

Both INH and EMB have been shown to inhibit the biosynthesis of the unique mycobacterial cell wall, which is rich in mycolic acids (Figure 1.4). INH is a prodrug activated by a catalase-peroxidase, *katG*, and its target is an enzyme involved in mycolic acid biosynthesis, encoded by *inhA*. Most INH-resistant *M. tuberculosis* isolates have mutations in *katG* as opposed to *inhA* (Slayden and Barry, 2000). EMB appears to hi-jack the construction of arabinogalactan, a major constituent of the cell wall, by introducing a rogue arabinose moiety into the molecule (Lee *et al.*, 1995).

PZA is also a prodrug, activated by conversion to pyrazanoic acid by a hydrolytic pyrazamidase encoded by *pncA*. Clinical isolates from PZA resistant patients often carry mutations in the *pncA* (Scorpio and Zhang, 1996). PZA has been proposed to have many specific and non-specific effects upon the bacilli.

RMP is a transcription inhibitor which has been shown to act upon the β -subunit of RNA polymerase encoded by *rpoB* (Cole, 1996).

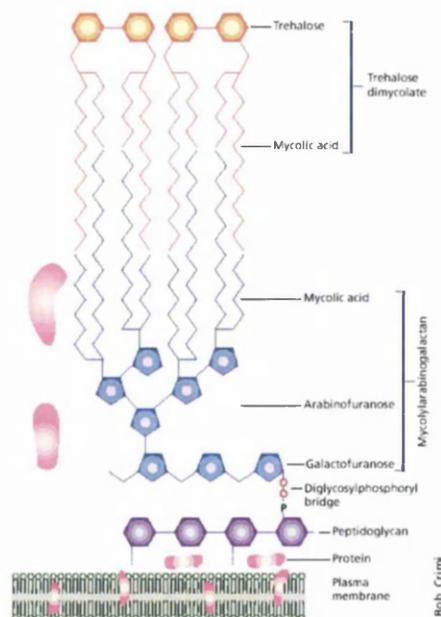


Figure 1.4: Schematic diagram of the mycobacterial cell wall. (From Tonge, 2000)

1.3.3 Multidrug-resistance in *M. tuberculosis*

The re-emergence of TB as global killer has been accompanied by rising numbers of clinical strains resistant to at least one of the four frontline anti-TB drugs. In the year 2000 an estimated 3.2% (273 000) of new TB cases were classed as multi-drug resistant (MDR), i.e. resistant to both INH and RMP (Dye *et al.*, 2002). The prevalence of MDR TB is much greater in cases where the patient has been treated previously for TB infection. However, 70% of the cases of MDR TB are found in only 10 countries. The cost of treating MDR TB is many times greater than the cost of treating susceptible TB, prohibitively so in many of the developing countries where the disease is most prevalent.

The increase in MDR TB has been attributed to patients not taking the full course of drugs prescribed, resulting in ‘selection’ of stronger mutant strains of *M. tuberculosis*. To tackle this problem, directly observed therapy short course (DOTS), where patients are supervised by medical staff while taking the drugs, has been introduced in many developing countries.

1.4 Current targets in TB drug development

As opposed to modifying current anti-tubercular drugs to combat resistance, current TB drug development programmes are focussed upon novel targets. This strategy may be a ‘long road’ to the development of a drug, however, the generation of novel agents may prevent a rapid rise in resistance towards the drug. The sequencing of the complete genome of *M. tuberculosis* strain H37Rv (Cole *et al.*, 1998) has allowed the identification of novel targets for anti-tubercular drug development.

The structure determination of antigen 85c (Ronning *et al.*, 2000), an enzyme involved in the biosynthesis of the mycobacterial cell wall again highlighted that the

unique nature of the cell wall provides an attractive drug target. Recent studies by Kuo and co-workers (Kuo *et al.*, 2003) have targeted the FAS-II fatty acid synthesis pathway, responsible for the production of pre-cursors for the cell wall, in the search for anti-tubercular agents which resulted in the identification of active compounds.

Current TB drugs are effective only against multiplying bacteria, however the ability of bacilli to enter a persistent state of infection where growth is minimal renders these drugs ineffective. While in the persistent state, *M. tuberculosis* bacilli downshift their metabolism and begin to degrade long chain fatty acids, utilising the glyoxylate shunt pathway (Bishai, 2000). McKinney and co-workers (McKinney *et al.*, 2000) demonstrated that the enzyme isocitrate lyase (ICL) is essential for the survival of persistent bacilli. Structure determination of ICL (Sharma *et al.*, 2000) has aided progress towards inhibition of this enzyme.

Studies by Parish and co-workers (Parish and Stoker, 2002) have shown that disruption of the shikimate pathway in *M. tuberculosis* results in the generation of non-viable strains of the organism.

1.5 Shikimate Pathway

The shikimate pathway, sometimes referred to as the common aromatic amino acid biosynthesis pathway, is a seven step pathway starting from phosphoenol pyruvate (PEP) and D-erythrose-4-phosphate (E4P) leading to the production of chorismate (Figure 1.5). Chorismate is a precursor for the biosynthesis of aromatic amino acids, folic acid, ubiquinone and other essential aromatic compounds (Haslam, 1993; Bentley, 1990). The shikimate pathway is present in bacteria, fungi, protozoa and plants but not mammals (Bentley, 1990; Roberts *et al.*, 1998), presenting an ideal target for anti-infective agents (Payne *et al.*, 2000). A schematic diagram of the

pathway showing representative crystal structures of each enzyme is shown in figure 1.5.

In bacteria each step is catalysed by a single enzyme transcribed from a single gene, whereas in higher plant species the activities of dehydroquinase and shikimate dehydrogenase are coupled together on a bi-functional polypeptide (Deka *et al.*, 1994). In filamentous fungi such as *Aspergillus nidulans* a large multi-enzyme complex referred to as AROM incorporates the activities of dehydroquinase synthase, dehydroquinase, shikimate dehydrogenase, shikimate kinase and EPSP synthase (Coggins *et al.*, 1987).

Structures of DAHP synthase (Shumilin *et al.*, 1999), Dehydroquinase synthase (Carpenter *et al.*, 1998), type I and type II Dehydroquinase (Gourley *et al.*, 1999), Shikimate dehydrogenase (Michel *et al.*, 2003), Shikimate kinase (Krell *et al.*, 1998) and EPSP synthase (Stallings *et al.*, 1991) have been determined by X-ray crystallography.

Number	Structure	PDB I.D.	Reference
1	DAHP synthase	1QR7	Shumilin <i>et al.</i> , 1999
2	Dehydroquinase synthase	1DQS	Carpenter <i>et al.</i> , 1998
3	Type I Dehydroquinase	1QFE	Gourley <i>et al.</i> , 1999
4	Type II Dehydroquinase	2DHQ	Gourley <i>et al.</i> , 1999
5	Shikimate dehydrogenase	1NYT	Michel <i>et al.</i> , 2003
6	Shikimate kinase	1SHK	Krell <i>et al.</i> , 1998
7	EPSP synthase	1EPS	Stallings <i>et al.</i> , 1991

Table 1.1 References and PDB codes for structures of shikimate pathway enzymes shown in figure 1.5

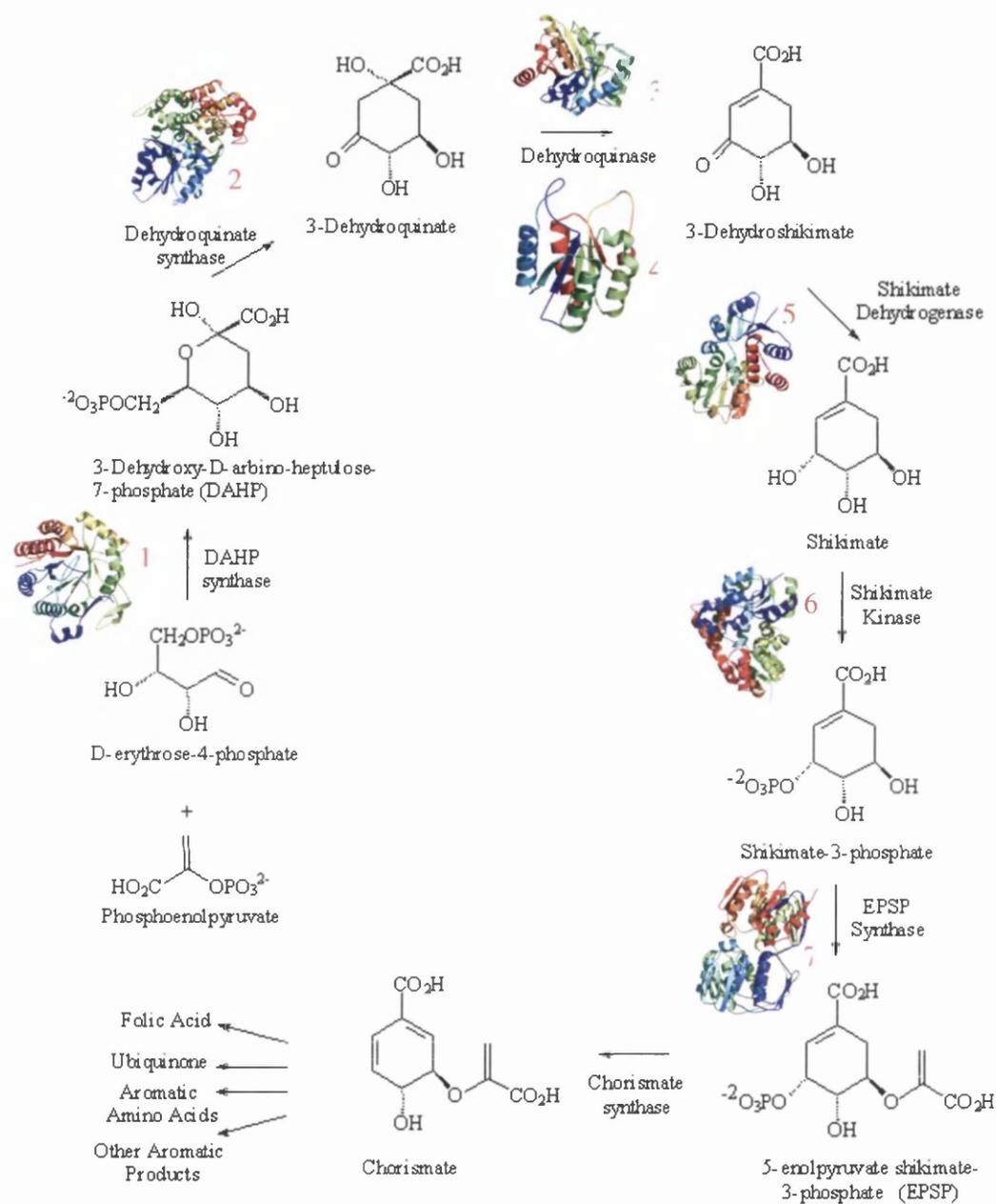


Figure 1.5: Schematic diagram of the enzymatic steps that form the shikimate pathway.

Representative crystal structures of each enzyme are shown where available. For key to crystal structures see table 1.1 on previous page.

1.5.1 The Shikimate Pathway as a drug target

The history of the shikimate pathway as a drug target dates back to research by the Monsanto Corporation in the 1970's where the compound N-phosphonomethyl glycine (Glyphosate™) was shown to act upon the shikimate pathway (Jaworski, 1972). Glyphosate went on to become a billion-dollar herbicide. The enzyme inhibited by glyphosate was shown to be 5-enolpyruvylshikimate acid 3-phosphate synthase (EPSP synthase) (Steinrucken and Amrhein, 1980), the penultimate step in the shikimate pathway. Compounds that inhibit shikimate dehydrogenase (Bailie *et al.*, 1972) and chorismate synthase (Davies *et al.*, 1994) have also been demonstrated to show herbicidal and anti-microbial activity.

The use of *aro* deficient mutants of infectious microorganisms as attenuated live vaccines has also been investigated. Hoiseth and Stocker (1981) demonstrated that disrupting the *aroA* gene, encoding EPSP synthase, of *Salmonella typhimurium* resulted in strains that were not only auxotrophic for the aromatic amino acids tryptophan, tyrosine and phenylalanine but non-virulent and effective as live vaccines. Mice inoculated with *S. typhimurium aroA*⁻ strains that expressed antigens from streptococcal species were shown to produce an immune response after challenge with *Streptococcus pyogenes*, suggesting that *aro*⁻ strains could be used as a vehicle for vaccination (Poirier *et al.*, 1988). *Aro*⁻ strains of *Yersinia enterocolitica* (Bowe *et al.*, 1989) and *Bacillus anthracis* (Ivins *et al.*, 1990) have also been shown to be effective as attenuated live vaccines.

In this study the enzyme 3-dehydroquinate dehydratase (dehydroquinase, DHQase) was studied to develop novel anti-mycobacterial compounds.

Two structurally and mechanistically unrelated forms of DHQase have been characterised (Gourley *et al.*, 1999). The type I DHQase is associated exclusively with the shikimate pathway, whereas the type II DHQase is involved with both the shikimate and the quinate pathways.

1.6.1 Two types of Dehydroquinase

Biophysical characterisation of the biosynthetic DHQases from *E. coli* (type I) and the catabolic DHQase for *A. nidulans* (type II) revealed striking differences between the two forms (Kleanthous *et al.*, 1992). The type I DHQase was shown to be dimeric and of approximate molecular weight 56kDa in solution whereas the type II enzyme formed a dodecamer in solution with approximate molecular mass 190kDa.

The type II enzyme was shown to be more stable to thermal and chemical denaturation. Using differential scanning calorimetry (DSC) three discrete transitions were observed before the enzyme completely denatured above 82°C, whereas the type I enzyme denatured in a single transition at 57°C. Price *et al.* (1999) demonstrated that the three transitions observed in the thermal denaturation of the type II DHQase resulted from the dissociation of the dodecamer units into trimer units, then into monomers before completely unfolding.

The reaction mechanisms of the two types of DHQase were shown by NMR studies to proceed via different stereochemical paths (Harris *et al.*, 1993). It was shown that the type I DHQase reaction proceeded via a *syn* elimination whereas the type II DHQase reaction was shown to proceed via an *anti* elimination.

1.6.2 Crystal structures of type I and type II Dehydroquinases

The structures of the type I DHQase from *S. typhi* and the type II DHQase from *M. tuberculosis* were elucidated by X-ray crystallography (Gourley *et al.*, 1999). The type I DHQase monomer was shown to have an α/β^8 barrel fold, exemplified by the structure of triose phosphate isomerase (TIM) (Figure 1.8a), whereas the type II DHQase monomer was shown to possess an α/β fold consisting of a central 5 stranded anti-parallel β -sheet flanked by 4 α -helices (Figure 1.8c). The 21345 strand order of the central sheet is similar to that of flavodoxin. The type I dimer is a simple side-by-side dimer formed via interactions between α -helices from both monomers (Figure 1.8b) whereas the quaternary structure of the type II enzyme is dodecameric, as predicted from biophysical studies, the dodecamer possessing unusual 23 tetrahedral symmetry creating a tetramer of trimers (Figure 1.8d). Each monomer of type II DHQase therefore interacts with other subunits at two interfaces; one which contributes to forming the trimer unit (referred to as the trimer interface), and the other which forms the interaction between two trimer units (the dimer interface).

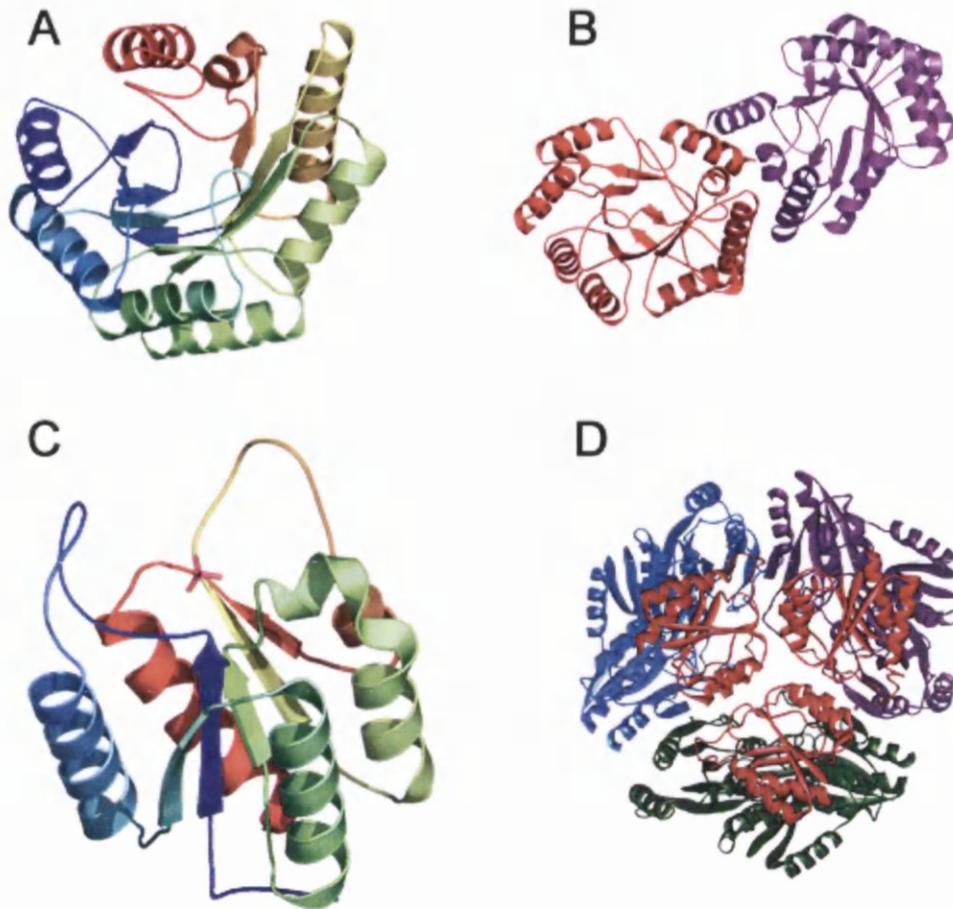


Figure 1.8: The enzyme folds of type I and II Dehydroquinases shown in ribbon representation.

A) type I DHQase monomer fold, B) type I DHQase quaternary structure, C) type II DHQase monomer fold, D) type II DHQase quaternary structure. In A and C the ribbons are coloured by residue number from N terminus (Blue) to C terminus (Red). In B each chain is coloured separately and in D each trimer unit is coloured separately.

1.6.3 Type I Dehydroquinase

The type I DHQase from *E. coli* (Duncan *et al.*, 1986) has been studied extensively as the model for type I DHQase activity, however many more type I DHQases have been isolated including the enzymes from *Salmonella typhi* (Servos *et al.*, 1991) and *Shigella dysenteriae* (Walker and Verma, 1997).

The reaction mechanism of type I DHQase proceeds via formation of a Schiff-base intermediate as shown in figure 1.9 (Butler *et al.*, 1974). The residue that forms the enzyme-substrate bond was suggested to be an invariant lysine identified by borohydride reduction (Chaudhuri *et al.*, 1991) and NMR (Leech *et al.*, 1995). The crystal structure of the *S. typhi* type I DHQase was determined as the enzyme-substrate complex confirming Lys 170 as the active residue (Gourley *et al.*, 1999). His 143 was postulated to act as the proton donor to complete the reaction, however no structures of type I enzymes have been determined with this residue in the appropriate orientation for catalysis (Deka *et al.*, 1992; Leech *et al.*, 1998).

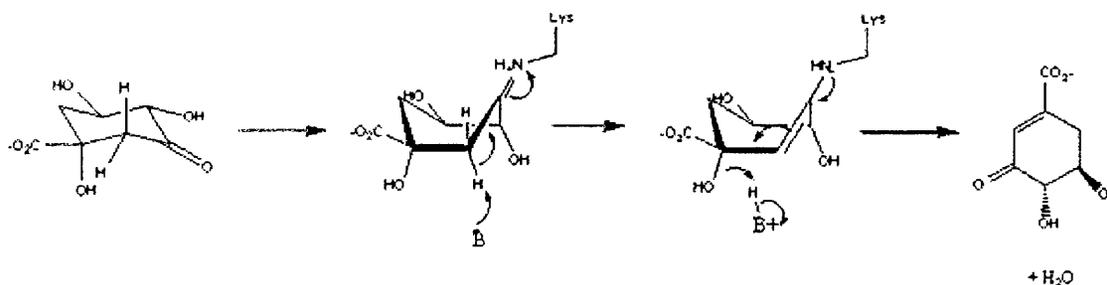


Figure 1.9 Proposed reaction mechanism for type I DHQase (Gourley *et al.*, 1999)

1.6.4 Type II Dehydroquinase

Type II dehydroquinases have been identified, isolated and characterised from many species including *M. tuberculosis* (Garbe *et al.*, 1991), *Streptomyces coelicolor* (White *et al.*, 1990), *Helicobacter pylori* (Bottomley *et al.*, 1996) and *Aspergillus nidulans* (Beri *et al.*, 1990). Crystal structures of the type II DHQases from *M. tuberculosis* (Gourley *et al.*, 1999) (PDB I.D. 2DHQ), *S. coelicolor* (Roszak *et al.*, 2002) (PDB I.D. 1GU0) and *Bacillus subtilis* (PDB I.D. 1GQO) have been elucidated previously.

Analysis of the kinetic parameters of the type II DHQases previously isolated have shown that there are two broad classes of enzyme (Evans *et al.*, 2002). The type II DHQases from *S. coelicolor* (SCDHQase) and *A. nidulans* (ANDHQase) have high k_{cat} values of $\sim 100\text{-}1000\text{ s}^{-1}$, whereas the enzymes from *M. tuberculosis* (MTDHQase) and *H. pylori* (HPDHQase) have low values for $k_{cat} \sim 10\text{ s}^{-1}$.

The reaction catalysed by type II DHQase was proposed to proceed via an E_1CB mechanism (Harris *et al.*, 1993). Kinetic isotope studies (Harris *et al.*, 1996a) provided further evidence for this reaction proceeding via an enolate intermediate.

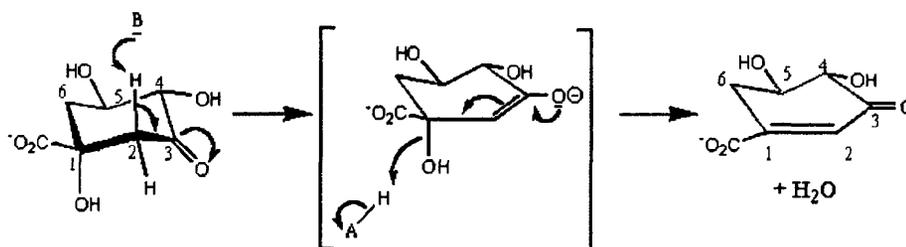


Figure 1.10: Proposed reaction mechanism for type II DHQase

Chemical modification experiments (Krell *et al.*, 1996) identified two key residues required for catalysis, a hyper-reactive arginine residue and a tyrosine residue. This strongly suggested that the hyper-reactive arginine residue would act as the general base in the initial step of the reaction.

The first crystal structure of MTDHQase (Gourley *et al.*, 1999) revealed the position of the hyper-reactive arginine on a flexible loop between strand β 1 and helix α 1. The essential tyrosine residue also identified by Krell and co-workers was also on this loop however this section of polypeptide was disordered in the crystal structure. From this structure, the location of residues conserved throughout all type II DHQases were identified (Figure 1.11), and many were found to be at the C-terminal ends of the β -strands, or on loops near to the β -strands. The identification of a Glu-His pair near the proposed active site suggested that these residues removed the proton in the initial step of the reaction, with the intermediate stabilised by arginine residues (Gourley *et al.*, 1999; Price *et al.*, 1999).

The structure determination of SCDHQase in complex with phosphate ions and a transition state analogue (Roszak *et al.*, 2002) provided a solution to the reaction mechanism. The transition state analogue 2,3-anhydro quinic acid was identified within the area of the proposed active site, however it was orientated very differently from the proposed model. The Glu-His pair implicated in the proton abstraction step were found to coordinate the C5 hydroxyl of the ligand whereas the C2 carbon, the site of the proton abstraction, was located near the flexible loop between strand β -1 and helix α -1.

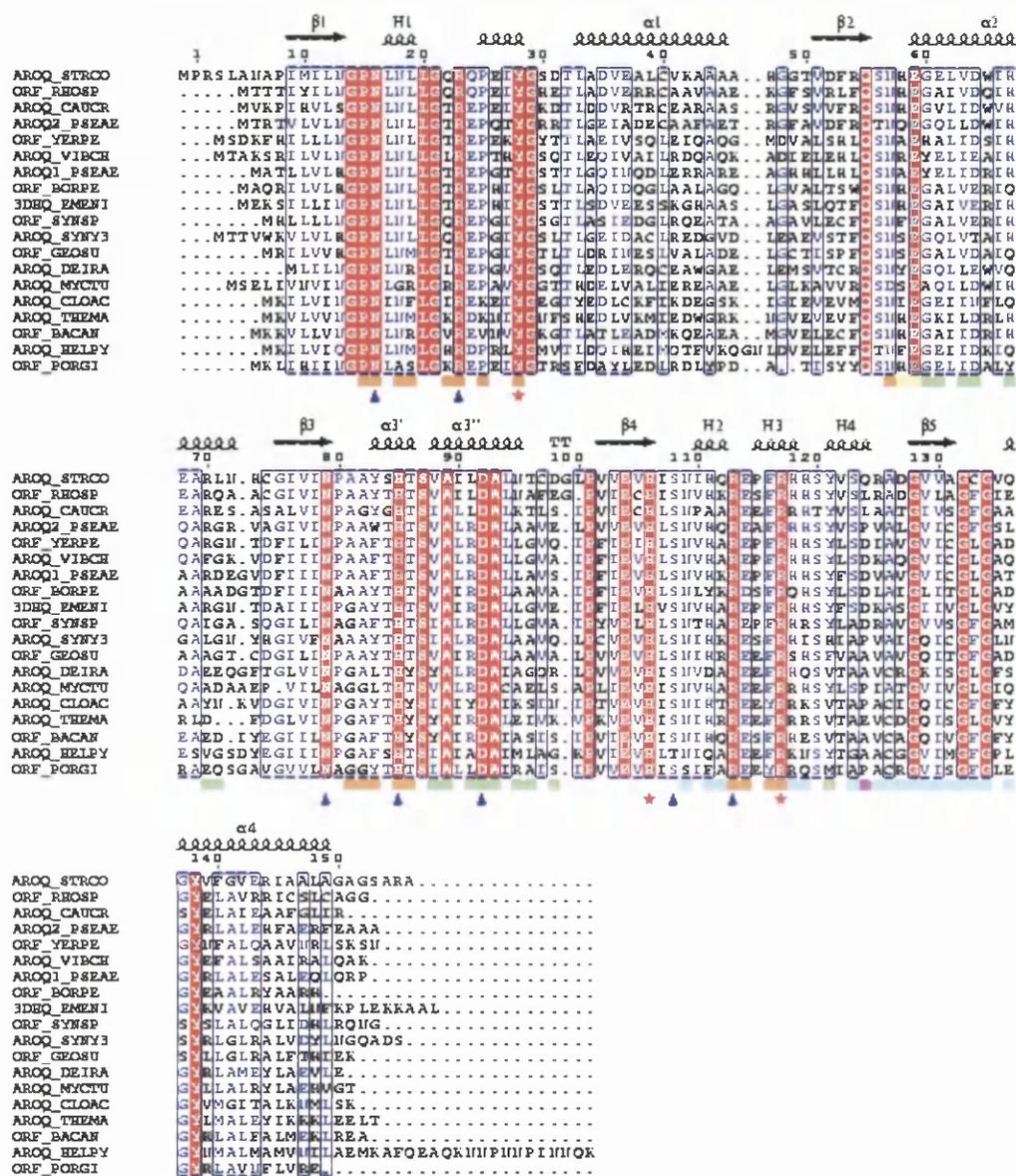


Figure 1.11: A sequence alignment of type II DHQases from representative prokaryotic and eukaryotic organisms. Alpha helices and beta strands are represented as helices and arrows respectively, and beta turns are marked with TT. The first row under the aligned sequences highlights those amino acid residues buried at the dimer interface (cyan), and the trimer interface, surface 1 (green) and surface 2 (orange). Some amino acid residues are buried at both the dimer and trimer interfaces (magenta) or both trimer interfaces (yellow). Key residues in the active site are marked with triangle in blue, those responsible for catalysis are marked with star in red. The sequence alignment was created using the following sequences (Organism,

Genebank/Swissprot accession in parentheses): AROQ_STRCO (*S. coelicolor*, P15474), AROQ_CAUCR (*Caulobacter crescentus*, AAK23857), AROQ_SYNY3 (*Synechocystis* sp. P73367), 3DHQ_EMENI (*Emericella nidulans*, P05147), AROQ1_PSEAE (*Pseudomonas aeruginosa* AF010322), AROQ2_PSEAE (*P. aeruginosa*, AAG03634), AROQ_VIBCH (*Vibrio cholerae*, AAF93471), AROQ_MYTUB (*M. tuberculosis*, P36918), AROQ_CLOAC (*Clostridium acetobutylicum*, AAK78875), AROQ_HELPHY (*Helicobacter pylori*, Q48255), AROQ_THEMA (*Thermotoga maritima*, Q9WYI4), AROQ_DEIRA (*Deinococcus radiodurans*, P54517). Open reading frames (ORF) from the incomplete genome sequences were obtained from The Institute for Genomic Research website at "<http://www.tigr.org>" for PORGI (*Porphyromonas gingivalis*), GEOSU (*Geobacter sulfurreducens*) and BACAN (*Bacillus anthracis*), from the Sanger Centre Sequencing Project website at "<http://www.sanger.ac.uk/Projects/>" BORPE (*Bordetella pertussis*) and YERPE (*Yersinia pestis*), and Preliminary sequence data was obtained from The DOE Joint Genome Institute (JGI) at "http://www.jgi.doe.gov/JGI_microbial/html/index.html" for RHOSP (*Rhodobacter sphaeroides*) and SYNSP (*Synechococcus* sp.).

The lid domain not visible in the MTDHQase structure was fully resolved in all the SCDHQase structures determined, thus allowing the locations of the essential arginine residue (Arg23) and tyrosine residues (Tyr28) to be determined. Figure 1.12 shows the location of the transition state analogue 2,3-anhydro quinic acid in the active site of SCDHQase. Key conserved residues required for catalysis are shown and labelled for clarity.

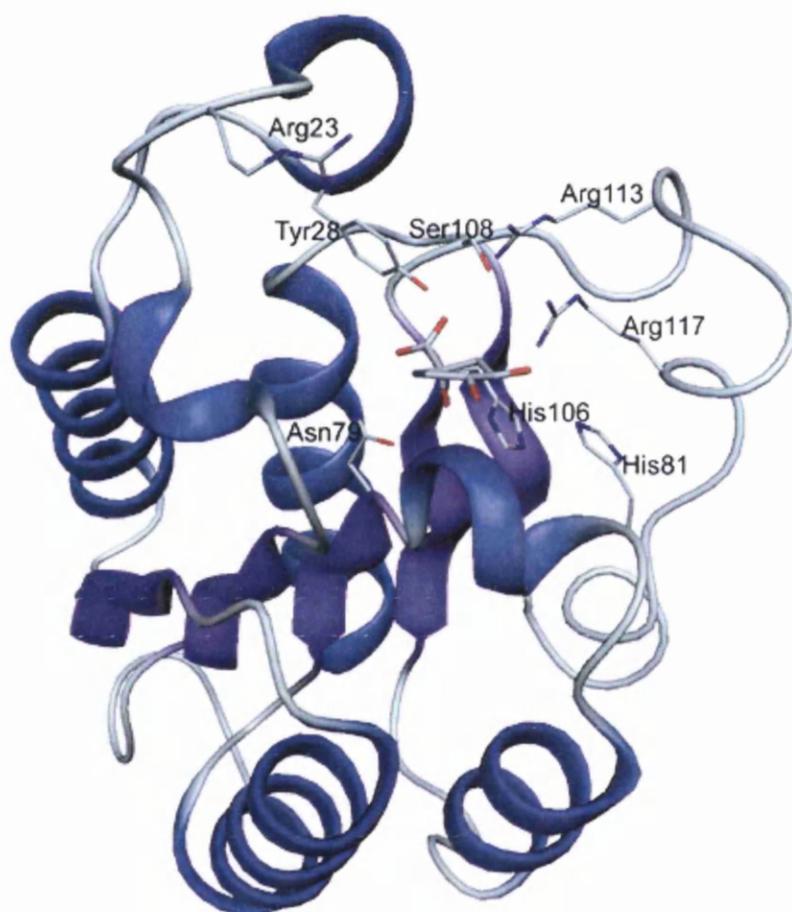


Figure 1.12: SCDHQase with 2,3-anhydroquinic acid bound in the active site. Key residues are shown in stick representation and labelled for clarity.

From the SCDHQase structures determined, Roszak *et al.* (2002) proposed the following reaction mechanism for type II DHQase (Figure 1.13)

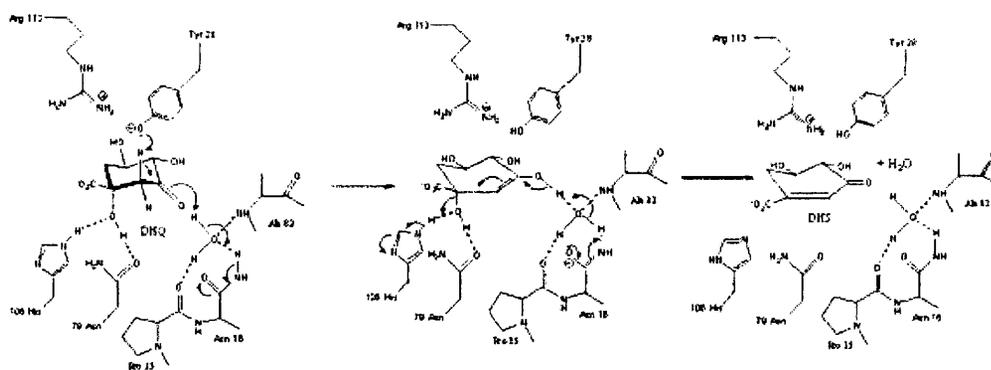


Figure 1.13: Reaction mechanism for type II DHQase (From Roszak *et al.*, 2002)

The reaction is proposed to proceed as follows. The residue Tyr28 is deprotonated by virtue of its basic micro-environment in close proximity to Arg113. The pro-*S* proton is removed from the C2 carbon of dehydroquinone, with an enol being formed at the C3 carbonyl group, stabilised by a bound water molecule and Asn16. This transition state subsequently collapses, eliminating water with the loss of the C1 hydroxyl group.

1.6.5 Selective inhibition of 3-Dehydroquinone dehydratase.

In order to develop a successful anti-mycobacterial agent targeting DHQase, the compound must not only be potent, but must also be selective towards MTDHQase. Work carried out using transition state analogues to determine the differing stereochemical paths of the type I and type II DHQase reactions illustrated that selective inhibition of type II DHQases over type I DHQase was possible (Harris *et al.*, 1996b; Bello *et al.*, 2000; Frederickson *et al.*, 1999). These findings make the type II DHQases attractive targets for drug design as many pathogenic bacteria including *M. tuberculosis* and *H. pylori* possess type II DHQases, whereas organisms that inhabit the human gut such as *E. coli* possess only the type I enzyme. Therefore, any therapy targeting type II DHQase would reduce unwanted side-effects resulting from disruption of the fauna of the human gut.

Further studies developing transition state analogues to probe the active site of type II DHQases resulted in the generation of compounds that showed selectivity between MTDHQase and SCDHQase of over 2 orders of magnitude (Frederickson *et al.*, 1999, 2002).

These results suggest that compounds may be identified that show high potency and high selectivity towards the type II DHQase from *M. tuberculosis*.

1.7 The drug discovery process

The development of a novel drug is a long and arduous path, which on average takes 8-10 years from target selection to final approval for sale. Each pharmaceutical and biotechnology company approaches the development process in a different manner (Ratti and Trist, 2001). Figure 1.14 shows the general 'pipeline' illustrating the discovery process.

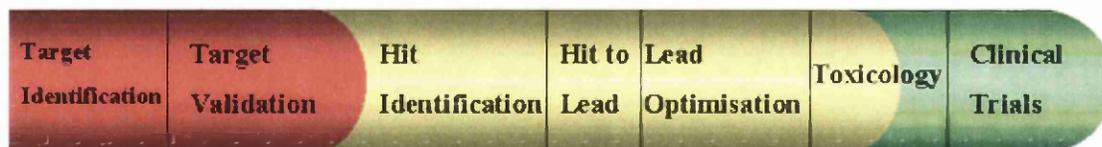


Figure 1.14 General 'pipeline' diagram illustrating the stages of the drug discovery process.

1.7.1 Target Selection and Validation

The drug development target can be identified via many processes. In many historical cases, compounds based upon natural products that showed the desired biological activity were developed. In some cases, such as the mode of action of INH towards *M. tuberculosis* (developed in 1952), the target that the molecule acts upon was not identified until decades after the drug went on the market (Quemard *et al.*, 1995).

With the recent explosion in genome technologies, many drug discovery programmes use genome based target identification processes. The use of micro-array, or gene-chip, technology combined with proteomics techniques allows comparison of the transcription and translation of genes between 'normal' specimens and 'diseased' specimens. It is therefore possible to identify genes that are up or down regulated during the disease state of interest.

The validation of target involves determining whether inhibition (or stimulation) of target does indeed confer a change in phenotype of the specimen. In many cases animal models are used for this process. The validation process involves the development of initial *in-vitro* and *in-vivo* assays to follow the action of compounds upon the target using generic inhibitors of the protein class in question. At this stage the tractability of the target is also assessed. A target may be identified that appears to provide a 'silver-bullet' cure for a disease, however if there is no feasible way to analyse the inhibition of the target, then the target is less attractive for the development process.

1.7.2 Hit identification to lead optimisation

Once a target has been identified and suitably validated for drug discovery, chemical matter must then be discovered that shows some potency towards the target; the hit identification process. In the current drug discovery environment, high throughput screening (HTS) is most commonly used to identify compounds that show some degree of activity against the target of choice.

In HTS an entire compound collection (typically between five hundred thousand and two million compounds) is run through a simple assay to determine the activity of each compound towards the target. The HTS process only became feasible within the last two decades as automation was developed to cope with such high numbers of sensitive assays. HTS assays are performed in multi-well plates (96, 384 or 1536 well formats) and involve little manual intervention as all the solution dispensing and analysis is highly automated.

Despite the obvious advantage that HTS allows screening of all available compounds from a collection, there are drawbacks with the methodology. Firstly, it is a very expensive process, and secondly there are issues with the stability of compounds used in the screen. As all the compounds are dissolved in a universal solvent (e.g. DMSO) and are stored for sometimes significant periods of time, the chemical matter may have degraded and therefore may give false results (negative or positive).

Once hits have been identified, the types of compound that show activity are analysed and re-tested in a more accurate assay. Libraries of compounds based around the hits are then designed and tested. This determines which compounds are synthetically accessible, how much diversity can be added around the basic core of

the molecule and ultimately identifies the most potent compounds that will form the lead series for drug discovery.

In general one or two lead series are carried forward, with one or two back up series. A series is a set of compounds based upon a specific core template that can be derivitised and to generate a chemically diverse set of compounds.

In the lead optimisation stage, the lead series is investigated in greater detail. Examination of the structure activity relationships (SAR) between the molecules and their target allows optimisation of the lead series. This involves modification of areas of the molecule around the core template. Assays carried out on these modified molecules identify which areas of the molecule can be functionalised to increase potency, increase selectivity, reduce toxicity, alter solubility etc. Once a molecule enters lead optimisation, many types of pharmacological assay are performed to monitor as many properties of the compounds as possible. Toxicity and solubility problems can be identified at this stage, preventing further investment in molecules that are likely to fail later in the drug discovery process. The DMPK (Drug Metabolism Pharmacokinetics) and ADMET (Adsorption, Distribution, Metabolism, Excretion and Toxicity) characteristics are assessed at this stage.

1.7.3 Toxicology and clinical trials

Once a compound that shows the correct potency, selectivity, DMPK and ADMET characteristics has been developed, full toxicological studies have to be carried out. Some of these toxicological studies are run as part of phase I clinical trials, where healthy patients take the drug to allow assessment of the pharmacokinetics. Phase II trials are carried out upon a small group of patients to allow assessment of dose levels before full scale phase III trials are entered. Phase II

clinical trials are arguably the most expensive section of the drug discovery process, with each trial costing up to 300 million US dollars. In phase III studies large samples of patients are studied and the effect of the novel compound is compared to the best available drug for the chosen disease area. Placebo trials are also carried out as a control where patients receive a treatment that has no effect. Phase III trials are generally carried out using 'double blind' methodology, where neither the patient nor the doctor are aware if the preparation is the active formulation or the placebo.

If the phase III trial is successful and the drug is approved by the Food and Drug Administration (FDA) the new drug can be marketed.

1.8 The impact of structural biology on the drug discovery process

As a result of the recent advances in protein structure determination by X-ray crystallography and NMR, the use of structural information has impacted greatly at all stages of the drug discovery process from target validation through lead optimisation. The 'traditional' position for structural biology in the drug discovery process was around the hit to lead step, where structure based drug design (or rational drug design) was implemented. Currently, the availability of a crystal structure of a target can impact on the target validation stage as this allows structure based procedures to be implemented in the process, hence making the target more attractive for development.

The advent of high throughput crystallography has allowed the development of structure based lead development as implemented by Astex Technology Ltd, Cambridge, UK (Blundell *et al*, 2001). In this methodology small molecule fragments are soaked into crystals of the target protein in order to identify high

affinity binding sites within the active site. From these fragments entire molecules can be constructed that show high potency.

Structural biology plays a major role during the lead optimisation stage in the development pipeline. As the SAR for a target is investigated, the availability of many crystal structures of the target in complex with compounds allows the rationalisation of the SAR results. This not only guides the lead optimisation towards probing areas of the active site not fully investigated, but may also prevent costly (in time and money) investigations into altering areas of the molecule that will not improve the potency or selectivity.

1.8.1 Structure based drug design

Structure based drug design is in general an iterative process whereby firstly the structure of a drug target is determined by X-ray crystallography or NMR spectroscopy, preferably in complex with a ligand. A detailed analysis of the ligand binding site is then carried out using computational methods to either suggest alterations to the ligand that will improve its binding affinity, or to analyse potential electrostatic or steric interactions within the binding site allowing the de novo generation of potential high affinity ligands. The modified ligands are subsequently synthesised and assayed for activity towards the target protein and structures obtained of the ligands in complex with the target. The cycle begins again with analysis of the new ligand – target complexes and the procedure is followed until a high affinity ligand is generated.

The first example of structure based ligand design reported was the design of a compound to mimic diphosphoglycerate, the allosteric effector of haemoglobin (Beddell *et al.*, 1976). However, the first drug to be successfully designed using

structural information was the antihypertensive captopril, which inhibits the carboxypeptidase angiotensin-converting enzyme (ACE) (Cushman *et al.*, 1977). In this case an initial lead molecule had been identified from snake venom, however using the crystal structure of the related protein carboxypeptidase A (Lipscomb *et al.*, 1968) as a guide, the lead molecule was modified to produce the blockbuster drug. The structure of the captopril target human ACE-I has only recently been determined (Natesh *et al.*, 2003).

The most famous success of structure based design is in the development of the peptidomimetic HIV-protease inhibitors nelfinavir, saquinavir, ritonavir, indinavir and lopinavir (Greer *et al.*, 1994). Structure based design has also played a major role facilitating the design of inhibitors targeting sialidases (Taylor 1996). The anti-influenza drug zanamivir (Relenza) (Varghese, 1999) was developed with major input from structural biology.

1.9 Aims

The primary aim of the work undertaken in this thesis was to identify novel selective inhibitors of type II 3-dehydroquinase from *Mycobacterium tuberculosis* using structure based ligand design methods. Large differences in inhibitor affinities and kinetic parameters have been observed between different type II 3-dehydroquinases. A key aim was therefore to determine the 3D structure of the type II 3-dehydroquinase from *Helicobacter pylori*. This structure has been used to aid the structure based ligand design, providing insight into the differing affinities for ligands shown by the various type II 3-dehydroquinases.

A secondary aim of the project was to develop a novel assay for 3-dehydroquinase activity that is amenable to high throughput screening.

The specific objectives were therefore:

- 1) Over-express and purify type II 3-dehydroquinases from *M. tuberculosis* and *H. pylori*.
- 2) Solve the structure of type II 3-dehydroquinase from *H. pylori* by X-ray crystallography in complex with a specific inhibitor identified from High Throughput Screening (HTS)
- 3) Analyse the available crystal structures of type II 3-dehydroquinases from *M. tuberculosis*, *Streptomyces coelicolor* and *H. pylori* to determine factors affecting ligand specificity.
- 4) Create pharmacophore models for type II 3-dehydroquinase from *M. tuberculosis* against which a compound library will be searched to identify potential inhibitors.
- 5) Screen compounds identified from the *in silico* screen for activity against type II 3-dehydroquinase from *M. tuberculosis*.
- 6) Develop a high throughput assay for 3-dehydroquinase activity.

Figure 1.15 illustrates the process for structure based ligand design used in this thesis. The diagonal boxes illustrate at which stage of the process the studies upon each enzyme are at the outset of the project.

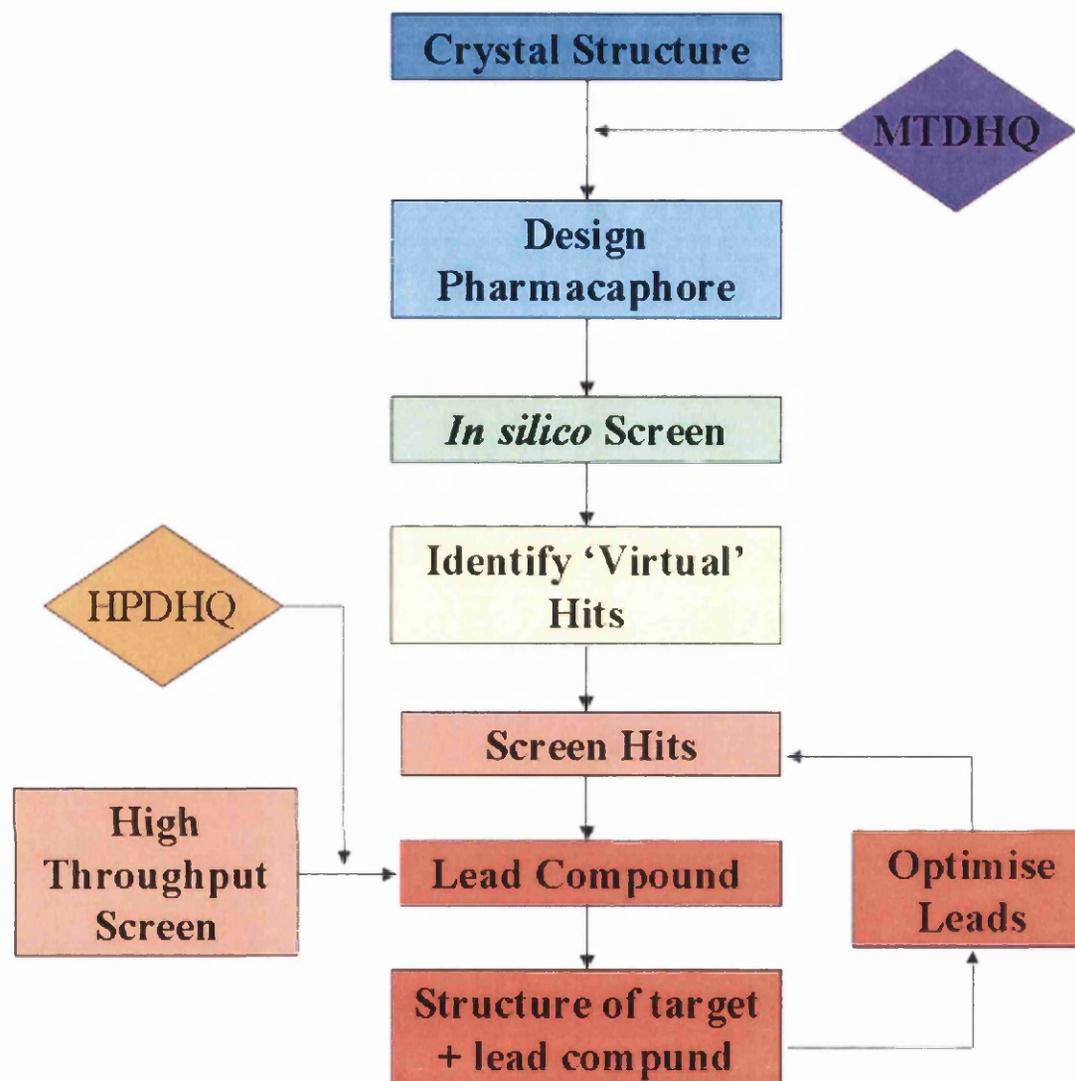


Figure 1.15: Flow chart illustrating the structure based process to followed in this thesis.

Diagonal boxes show where each enzyme entered the process.

Chapter 2 – Materials and Methods

2.1.1 Chemicals

Chemicals and biochemicals unless otherwise stated were analytical grade or the highest available. Unless otherwise stated all chemicals were obtained from the Sigma Aldrich Company Ltd., England or Fisher Scientific, UK. Agarose was obtained from GIBCO BRL Life Technologies, Scotland. EDTA free Protease Inhibitor cocktail tablets were obtained from Boehringer Mannheim, UK.

2.1.2 Proteins and Enzymes

Restriction enzymes and their appropriate buffers were obtained from Promega Corporation, England. Pfu DNA polymerase was obtained from Promega Corporation, England, VENT DNA polymerase was obtained from New England Biolabs Incorporated, England, Taq DNA polymerase was obtained from Sigma Aldrich Company Ltd., England and from Invitrogen, UK.

2.1.3 Oligonucleotides

Oligonucleotides were obtained from either GIBCO BRL Life Technologies, Scotland, Cruachem Ltd., Scotland or MWG Biotech, Germany.

2.1.4 Chromatography Media

Amylose resin was obtained from New England Biolabs Ltd., England. Q-Sepharose was obtained from Amersham Pharmacia, England and Ni-NTA resin purchased from Qiagen, England. Amersham Pharmacia XK-16/20 columns were

used to pack all media. Pre-packed Superdex-200 and Mono-Q columns were obtained from Amersham Pharmacia, England.

2.2 Media and Supplements

2.1 Media for Bacterial Growth

E. coli strains BL21(DE3)pLysS and DH5 α were routinely cultured in Miller's Luria Bertani Broth (LB) obtained from Sigma Aldrich Company Ltd., England. Solid media was prepared by addition of 1.2% bacto agar (Difco, USA) to liquid LB.

2.2.2 Antibiotics

Antibiotics were used at the following concentrations in all experiments.

Ampicillin - 100 μ g/ml

Tetracycline – 12.5 μ g/ml

Chloramphenicol - 34 μ g/ml

Kanamycin - 25 μ g/ml

i) Ampicillin

A 100mg/ml stock solution was prepared using sterile water, and the solution filter sterilised by passage through a 0.22 μ m filter. 1ml aliquots were stored at -20°C. Growth media was allowed to cool to 55-50°C before addition of Ampicillin to a final concentration of 100 μ g/ml

ii) Tetracycline

A stock solution of 12.5mg/ml was prepared in bulk ethanol and stored at -20°C in the dark. Growth media was allowed to cool to 55-50°C before addition of antibiotic to a final concentration of 12.5µg/ml

iii) Chloramphenicol

A stock solution of 34mg/ml was prepared in bulk ethanol and stored at -20°C. Growth media was allowed to cool to 55-50°C before addition of antibiotic to a final concentration of 34µg/ml

iv) Kanamycin

A stock solution of Kanamycin was prepared at 50mg/ml in sterile water. Solution was filter sterilised by passage through a 0.22µm filter and stored at room temperature. Growth media was allowed to cool to 55-50°C before addition of antibiotic to a final concentration of 25µg/ml.

All media and plates were stored at 4°C and used within 4 weeks of preparation.

2.2.3 Isopropyl-β-D-thiogalactopyranoside (IPTG)

A stock solution was prepared at 1M concentration and filter sterilised by passage through a 0.22µm filter. 1ml aliquots were stored at -20°C

2.3 General Molecular Biology Methods

General molecular biology methods were carried out as described in Sambrook *et. al.*,(1989)

2.4 Bacterial strains and cloning vectors

2.4.1 Storage of Bacterial Strains

Bacterial strains were stored as glycerol stocks. These stocks were prepared by addition of 60% (v/v) glycerol/water to bacterial cultures to a final concentration of 30% (v/v). Stocks were stored at -80°C. Cultures were temporarily stored on agar plates at 4°C for up to 4 weeks.

2.4.2 Bacterial strains

The bacterial strains and plasmids used in this thesis are listed in tables 2.1 and 2.3.

2.4.3 Growth of Plasmid containing *E. coli*

E. coli cells containing plasmids of interest were grown on LB media supplemented with the appropriate antibiotic to select for cells containing the correct plasmid.

2.4.4 Growth of cells and isolation of plasmids

Plasmid DNA was isolated by overnight growth of a 3ml LB culture (with appropriate antibiotics), inoculated with a single colony of the correct strain. Cells were grown overnight at 37°C with shaking. Plasmid isolation was performed either using the method of Birnboim and Dolly (Sambrook *et al.*, 1989) or using the Qiaprep Spin mini-prep kit.

2.4.5 Transformation protocol

Unless otherwise instructed by manufacturers protocols the following transformation methodology was used. Chemically competent cells were allowed to thaw on ice for 30 minutes prior to use. 2 μ l of plasmid DNA was added to pre-cooled eppendorf tube and 20 μ l of competent cells added with gentle mixing. The cells were left on ice for 30 minutes before heat shocking at 42°C for 40 seconds. The cells were again placed on ice for 2 minutes before addition of 80 μ l of SOC medium (LB medium + 2.5mM KCl, 10mM MgCl₂, 20mM Glucose) (Invitrogen) warmed to 37°C. Cells were incubated at 37°C with shaking for 60 minutes before being plated onto LB agar plates.

Bacterial strain	Genotype	Reference
<i>Escherichia coli</i> DH5 α	F ϕ 80 <i>lacZ</i> Δ M15 <i>recA1 endA1 gyrA96 thi-1 hsdR17</i> (r _k ⁻ m _k ⁻) <i>supE44 relA1 deoR</i> Δ (<i>lacZY-argF</i>)U169	Hanahan, 1985
<i>Escherichia coli</i> BL21(DE3)pLysS	F <i>ompT hsdS_{BRB}⁻,mB⁻ dcm gal</i> λ DE3 pLysS Cm ^r	Studier and Moffat, 1986
<i>Escherichia coli</i> NovaBlue Singles	<i>EndA1 hsdR17</i> (r _{K12} ⁻ M _{K12} ⁺) <i>supE44 thi-1 recA1 gyrA96 relA1 lac F'</i> [<i>proA⁺B⁺ lacI^q</i> <i>Z</i> Δ M15::Tn10 (T _c ^R)]	Novagen

Table 2.1 Bacterial Strains used for general cloning purposes

Plasmid	Antibiotic resistance	Reference
pMAL-c2x	Ampicillin	MBP fusion expression plasmid New England Biolabs
p-T7Blue-3	Ampicillin/Kanamycin	Blunt cloning vector Novagen
pET-15b	Ampicillin	N-term 6xHis tag expression plasmid Novagen
pCR [®] 2.1- TOPO [®]	Ampicillin/Kanamycin	TOPO cloning vector Invitrogen

Table 2.2 Plasmids used for general cloning

2.5 Manipulations of DNA

2.5.1 Restriction digests

Digests contained 3-10 μ g of DNA with 1-2 units of each restriction endonuclease in the appropriate reaction buffer. Digests were carried out at 37°C for one hour.

2.5.2 Gel electrophoresis

Gel electrophoresis of DNA samples was performed using 1% agarose (w/v) (Gibco-BRL) gel in 1xTAE buffer (Sigma) containing approximately 0.5 μ g/ml ethidium bromide. Samples were loaded in 5x Nucleic Acid Electrophoresis Stain (Biorad). Standard molecular weight markers, such as 1kb ladder (Promega), molecular weight markers XIV and VI (Boehringer), were used on all gels to identify the size of separated bands. Gels were run at room temperature in the Mini-Sub GT (Biorad) gel tank at 80mA until the sample stain neared the end of the gels. DNA bands were visualised by viewing under UV light.

2.5.3 Gel Purification of DNA fragments

DNA fragments were purified by electrophoresis on 1% low melting point agarose gels. Bands of interest were excised from the gel and DNA purified using either QIAquick or QIEx II gel extraction kits (Qiagen)

2.5.4 DNA Ligation

Ligation reactions were carried out using 1 μ l T4 DNA ligase (Promega) and 1 μ l of the appropriate buffer along with various ratios of vector:insert DNA (1:1, 2:1, 1:2) to a final volume of 20 μ l. Reactions were carried out at 16°C for 16 hours.

2.6 Cloning by Polymerase Chain Reaction (PCR)

2.6.1 Primers for PCR

PCR primers were designed to incorporate restriction sites to facilitate easier cloning of the amplified DNA fragments into new vector systems. Primer oligonucleotides were ordered from MWG Biotech, Germany, and stored as 100 μ M stock solutions in water. A list of all primers used is given in Table 2.3.

Oligonucleotide	Sequence	Usage
DHPF	5'-T ATG CGT <u>GGA TCC</u> ATG AAA ATT TTA GTG ATT CAA GGG CCT-3'	Generates 5' <i>Bam</i> HI site for cloning <i>Hpy-aroQ</i> into pMAL-c2x
DHPR	5'-TCA GGC <u>AAG CTT</u> TTA TTT TTG ATT GTT AAT GGG-3'	Generates 3' <i>Hind</i> III site for cloning <i>Hpy-aroQ</i> into pMAL-c2x
MTNDE5	5'-GGC AGC GGC <u>CAT ATG AGC GAA CTG ATC GTG AAC GTG</u> -3'	Generates 5' <i>Nde</i> I for cloning <i>Mtb-aroD</i> into pET-15b
MTBAM3	5'-AGC GGG ACG CGT <u>GGA TCC</u> TCA CGT CCC GAC ATG CTC AGC-3'	Generates 3' <i>Bam</i> HI site for cloning <i>Mtb-aroD</i> into pET-15b
HPN5	5'-C GGA ATT <u>CAT ATG AAA ATT TTA GTG ATT CAA GGG CC</u> -3'	Generates 5' <i>Nde</i> I for cloning <i>Hpy-aroQ</i> into pET-15b
HPB3	5'-GCG CCA CCG <u>GGA TCC</u> TTA TTT TTG ATT GTT AAT GGG-3'	Generates 3' <i>Bam</i> HI site for cloning <i>Hpy-aroQ</i> into pET-15b
QCNDF	5'-GGT AGCCAC GGT <u>CAT ATG AAA TTA ACT TCT TTA CGC GTA</u> -3'	Generates 5' <i>Nde</i> I for cloning <i>qui-C</i> into pET-15b
QCBMR	5'-CCT TGC TAT AAC <u>GGA TCC</u> TTA TCT CAT TTT CAC TGC AGA ACT-3'	Generates 3' <i>Bam</i> HI site for cloning <i>qui-C</i> into pET-15b

Table 2.3.1 Oligonucleotides used for molecular cloning - PCR primers

Oligonucleotide	Sequence	Usage
HPSEQ1	5'-AAT AAC AAC AAC CTC GGG ATC-3'	Sequencing pMAL-haq from 5' region of pMAL-c2x vector
HPSEQ2	5'-CGA CGT TGT AAA ACG ACG GCC-3'	Sequencing pMAL-haq from 3' region of pMAL-c2x vector

Table 2.3.2 Oligonucleotides used for molecular cloning – Custom sequencing primers

2.6.2 PCR reactions

PCR reactions were carried out using Pfu DNA polymerase (Promega) unless otherwise stated. Reactions were carried out in 0.2ml PCR tubes, using a Biorad Gene Cyclor. Primer mixes consisted of the primers required for PCR at a concentration of 10 μ M. As DNA from *M. tuberculosis* has a high G/C content, DMSO was added to the PCR reaction mixes to a final concentration of 10% (v/v). The PCR reaction mixes were prepared as follows:

- 5 μ l 10x Pfu reaction buffer (200mM Tris-HCl pH8.8, 100mM KCl, 100mM (NH₄)₂SO₄, 200mM MgSO₄, 1% Triton-X100, 1mg/ml BSA)
- 1 μ l dNTP mix (dATP, dCTP, dGTP and dTTP, 10mM stock of each)
- 5 μ l Primer mix (Both Oligonucleotide primers at 10 μ M concentration)
- 2 μ l Template DNA
- 1-3 μ l MgSO₄ (100mM stock)
- 5 μ l DMSO
- Water to 49 μ l

The mix was incubated at 94°C for 5 minutes before 1 μ l of polymerase was added to the reaction.

The following temperature profile was used for 30 cycles in the gene cyclor:

30 seconds @ 94°C

30 seconds @ 65°C – annealing temperature of primers

60 seconds @ 72°C – extension of amplified strand

A final elongation step of 5 minutes at 72°C was performed at the end of the amplification reaction to ensure that all amplified material was full length. Primer annealing temperatures were those determined theoretically by the supplier.

2.6.3 Cloning Methods

2.6.3.1 Blunt Cloning

Cloning of PCR product into an acceptor vector by blunt ended ligation was carried out using the Perfectly Blunt[®] Cloning kit from Novagen. Blunt ended cloning allows the use of high fidelity proof reading polymerises in the initial PCR amplification, therefore reducing the possibilities of PCR errors. The PCR product is treated with 'End conversion mix', which removes any overhanging bases and phosphorylates the DNA at the 5' termini. The acceptor vector is purchased pre-linearised and de-phosphorylated, allowing ligation of the insert and vector using T4-DNA ligase. The resulting vector is then transformed into NovaBlue Singles competent cells, and plated out onto solid media.

2.7.2 TOPO[®] Cloning

TOPO cloning is a method by which PCR products can be directly inserted into a plasmid vector in the absence of DNA ligase. In this project the TOPO TA kit from Invitrogen was used. The PCR is carried out with *Taq* polymerase, which adds a single deoxyadenosine to the 3' end of the DNA strand. The linearised vector supplied has a single deoxythymidine residue at the 3' end, thus allowing the ligation of the PCR product to the vector.

The linearised vector is bound to molecules of topoisomerase I via the 3' hydroxyls, which are covalently bound to Tyr-274 of the enzyme. This phosphotyrosyl bond is attacked by the 5' hydroxyl of the PCR product, releasing the topoisomerase.

The PCR amplification is carried out with *Taq* polymerase and the product gel purified. The amplified gene is then incubated with the TOPO-vector at room

temperature allowing the ligation to take place, with the resulting vector transformed into chemically competent *E. coli* DH5 α and plated out onto solid media

2.7 Over-expression of recombinant type II Dehydroquinases

Four 25ml cultures of LB + ampicillin (100 μ g/ml) were inoculated with either a freshly picked colony from an agar plate or directly from a glycerol stock. These were grown overnight with shaking at 37°C and used to inoculate 4x 500ml LB + ampicillin (100 μ g./ml) cultures. The 500ml cultures were grown at 37°C with shaking until the OD₆₀₀ was between 0.4 and 0.6. IPTG was added to a final concentration of 0.8mM to induce expression of the target protein, and the cultures grown for a further 4 hours. Cells were harvested by centrifugation at 4000g for 20 minutes (Sigma 4K 15 centrifuge, 11150 rotor, 4x 13350 buckets). Cells were resuspended in phosphate buffered saline (PBS) and collected in 50ml culture tubes. Cells were again pelleted by centrifugation at 4000g for 20 minutes (Sigma 4K 15 centrifuge, 12690 rotor) and stored at -20°C.

Cells were broken by passage through a French pressure cell (2x950psi). Cell debris was removed by centrifugation 8000g 60 minutes (Sigma 4K 15, 12690 rotor) and the crude lysate passed through a sterile 0.22 μ m filter. The resultant supernatant was removed and used in subsequent purification steps.

2.8 Purification

All type II DHQases studied in this thesis were purified by liquid chromatography using the Fast Protein Liquid Chromatography (FPLC) system (Amersham Pharmacia). The first step in purification was affinity chromatography

with further purification achieved by ion exchange chromatography or size exclusion chromatography

2.8.1 Purification by Affinity Chromatography

2.8.1.1 MBP affinity chromatography

The MBP domain of the MBP-DHQase fusion peptides show high affinity towards Amylose resin (New England Biolabs), which mimics the natural substrate maltose. The crude lysate is passed through the amylose column, immobilising the MBP-DHQase fusion.

The column was equilibrated with buffer A (50mM Tris-HCl pH 7.8, 100mM NaCl, 1mM EDTA, 1mM DTT), before the crude lysate, also in buffer A, was passed through the column, which was washed until A_{280} reached the baseline level. The fusion peptide was eluted with buffer A + 10mM maltose. The maltose competes with the amylose for binding to the MBP, releasing the MBP-DHQase fusion.

2.8.1.2 Ni-NTA affinity chromatography

Poly-histidine tags show high affinity for nickel ions, therefore affinity chromatography can be carried out by passage of the sample through a nickel charged metal chelate column. The media used was Ni-NTA Superflow (Qiagen). Cell pastes were re-suspended in buffer A (20mM Tris-HCl pH8.0, 300mM NaCl (to prevent non-specific charged interactions), 20mM Imidazole), lysed and the supernatant isolated. The column was equilibrated with buffer A and the lysate was passed through and washed until the A_{280} of the flow through reached baseline level. A second wash step was carried out with buffer B (20mM Tris-HCl pH 8.0, 300mM

NaCl, 75mM Imidazole) before elution of the tagged DHQase with buffer C (20mM Tris-HCl, 300mM NaCl, 200mM Imidazole).

As the project progressed it became apparent the high imidazole concentration made the protein unstable, therefore 3ml fractions of the eluted protein were collected in tubes containing 3ml 10mM Tris-HCl pH7.0, 1mM EDTA. The EDTA was present to prevent aggregation of protein around Ni ions flushed from the column matrix.

2.8.2 Ion exchange chromatography

As a second purification step, to remove MBP contaminants from DHQase expressed as MBP-DHQase fusion peptides, or to further purify DHQase expressed with an N-terminal 6x His tag, ion exchange chromatography was carried out using Q-Sepharose anion exchange resin (Amersham Pharmacia). The protein was dialysed into buffer A (50mM Tris-HCl pH 7.5, 100mM NaCl, 1mM EDTA) before being applied to the column (previously equilibrated with buffer A). The column was washed with buffer A until A_{280} reached baseline level. A linear gradient was set up between buffer A and buffer B (50mM Tris-HCl pH 7.5, 500mM NaCl, 1mM EDTA) to run at 1.5 ml/min for 60 minutes. The pure DHQase was eluted from the column near the mid-point of this gradient around 250mM NaCl.

2.8.3. Size exclusion chromatography

To separate DHQase from MBP contaminants, size exclusion chromatography was carried out using a Superdex-200 column (Amersham Pharmacia). The column was equilibrated with 150ml 50mM Tris-HCl pH 7.8, 100mM NaCl, 1mM EDTA, before a 1ml concentrated sample of DHQase/MBP

mixture was loaded. The column was washed with 150ml of equilibration buffer, at a flow rate of 0.5ml/min. Peaks were observed for DHQase, MBP and MBP-DHQase fusion.

2.9 Protein quantification and analysis

2.9.1 SDS-PAGE

Sodium dodecylsulphate poly acrylamide gel electrophoresis (SDS-PAGE) is a method by which proteins can be separated on the basis of mass, allowing rapid analysis of protein solutions to determine the presence, or purity, of a target protein. The Hoefer Scientific Mighty Small system was used for all SDS-PAGE. The system was assembled according to the manufacturers instructions and 15% resolving gel solution and stacking gel solutions prepared (Sambrook et al., 1989).

Analysis of the electrophoretic profiles was performed by staining the gel with Coomassie Brilliant Blue (CBB R-250, Sigma Aldrich Company Ltd., England). The gel was incubated in 100 ml staining solution (0.1% CBB-R250, 50% (v/v) ethanol) for 20 minutes, before being rinsed with water. The background stain was removed by immersion of the gel in 200ml water heated to around 80°C for 10 minutes.

2.9.2 Dynamic Light Scattering

To determine the fraction of poly dispersity within a protein sample, dynamic light scattering was performed using the Dynapro-801 system (Protein Solutions). The protein sample (200-400µl) is passed into the sample chamber through a 0.1µm sterile filter. In the chamber the sample is illuminated with a laser, and the scattered light detected by a photo diode array. The software calculates a hydrodynamic radius (R_H) for the sample, which is used to estimate the molecular weight of the protein

based upon a globular model. The instrument also indicates if the sample is mono or poly-disperse by mono-modal curve fit analysis.

2.9.3 Protein concentration calculation using Absorbance at 280nm (A_{280})

The concentration of a protein sample in solution can be calculated using the Beer-Lambert equation (Equation 2.1)

$$A = \epsilon cl$$

Equation 2.1: Beer-Lambert law facilitating the calculation of the concentration of a protein (c), when the absorbance (A), sample path length (l) in cm and molar extinction coefficient (ϵ) are known

Protein molecules absorb radiation in the near UV region (~280nm) due to presence of tryptophan, tyrosine, and to a lesser extent cysteine and phenylalanine residues. Calculation of the molar extinction coefficient (ϵ) at 280 nm is based upon the numbers of these residues within the protein (Equation 2.2) (Pace *et al.*, 1995). This allows a quantitative measurement to be made of the protein concentration.

$$\epsilon = \frac{[1280nY + 5960nW + 120nC]}{Mwt}$$

Equation 2.2: Calculation of molar extinction coefficient (ϵ , $M^{-1} \text{ cm}^{-1}$) at 280nm when the number of tyrosine (nY), tryptophan (nW), cysteine (nC) residues and the molecular weight (Mwt) are known

All spectroscopic measurements were carried out using a Perkin Elmer Lambda 800 UV/vis spectrophotometer.

2.10 Standard Dehydroquinase Assay

To determine the activity of a DHQase sample, a simple standard assay can be carried out (White, 1990). The natural substrate 3-dehydroquinate is added to the protein sample, and the rate of conversion of this to 3-dehydroshikimate is followed by the increase in absorbance at 234nm. 3-dehydroshikimate absorbs light at 234nm due to the conjugated system between C1 and the C3 carbonyl (red in figure 2.1).

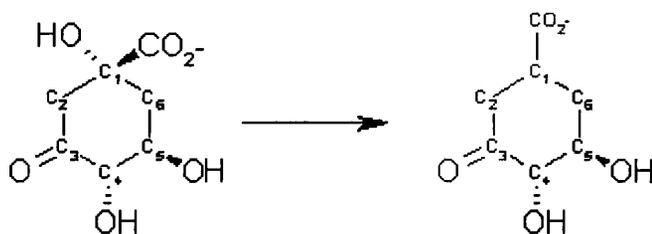


Figure 2.1: Formation of conjugated system in the conversion of 3-dehydroquinate to 3-dehydroshikimate

The standard assay is performed in 1ml quartz cuvettes (Helma), using 50mM Tris-Acetate buffer pH 7.0, and 20 μ g of substrate. Assays were carried out using a Jasco V560 spectrophotometer, and data analysed with Origin 5.0 (MicroCal Corp.)

2.11 Inhibitors of type II Dehydroquinase

The inhibitors of type II dehydroquinase used in this project are shown in table 2.4. Compounds FA1, FA3 and FA4 were obtained from Dr. Chris Abell, University of Cambridge. AH9095, GR78752X and GR122160X were identified from an in-house high throughput screen (HTS) at GSK.

Inhibitor	Structure	Name	Reference	Source
FA1		2,3 anhydro quinic acid	Frederickson <i>et al.</i> , 1999	Dr. C. Abell University of Cambridge
FA3		3 fluoro, 2,3 anhydro quinic acid	Frederickson <i>et al.</i> , 1999	Dr. C. Abell University of Cambridge
FA4		3,3 difluoro quinic acid	Personal Communication	Dr. C. Abell University of Cambridge
AH9095		9-Oxo-8a,10a-dihydro-9H-xanthen-2-sulfonic acid (1H-tetrazol-5-yl)-amide	GSK	GSK
GR78752X		2-Phenylsulfanyl-isoindole-1,3-dione	GSK	Sigma Aldrich Cat 51,114-5
GR122160X		di-5-Methyl-[1,3,4]thiadiazole-2-thiol	GSK	Synthesised (see 2.1.1.1)

Table 2.4 Inhibitors of type II Dehydroquinase used in this thesis

2.11.1 Synthesis of GR122160X

The compound GR122160X was identified from high throughput screen (HTS) against HPDHQase, carried out at GSK. No sample was available from the in-house compound library, or from any chemical company. The monomer 5-Methyl-[1,3,4]thiadiazole-2-thiol was available from Sigma Aldrich (cat 18,857-3) and formation of a di-sulphide link between two thiol groups was carried out. Many reviews of thiol chemistry are available (Procter 2001) and the procedure of Noureldin et al. (1998) was identified as a simple route to form the required di-sulphide (Figure 2.2).

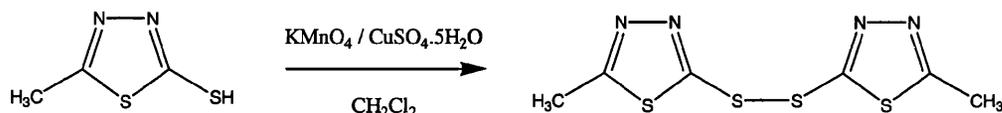


Figure 2.2: Oxidation of 5-Methyl-[1,3,4]thiadiazole-2-thiol to form GR122160X

A 1:1 (w/w) mixture of KMnO_4 and $\text{CuSO}_4 \cdot 5\text{H}_2\text{O}$ was used as the oxidant, and was prepared by grinding until homogenous. 0.8g of the oxidant and 2.5mMoles of the thiol were added 20ml DCM in 50ml round bottomed flask. The heterogeneous mixture was stirred at room temperature overnight, and the contents of the flask filtered through celite (Fluka) and washed with DCM (2x10ml) and ethanol (10ml). The product was isolated by evaporation under reduced pressure, and the product analysed by mass spectroscopy. A 78.8% yield was obtained as determined by mass.

2.12 Crystallisation of biological macromolecules

There are many current texts and publications available upon the methodology of crystallisation of biological macromolecules (McPherson, 1999, Ducruix and Giege 1992, Bergfors 1999).

2.12.1 Crystallisation by sitting drop vapour diffusion

All crystallisations were performed by the sitting drop vapour diffusion method using 24-well Cryschem Plate™ (Hampton Research Ltd.) at room temperature unless otherwise stated. The reservoir chamber was filled with 800µl of the precipitant solution, and a 1µl aliquot of the concentrated protein solution added to the crystallisation well. A 1µl aliquot of the precipitant solution was then added to the protein drop, thus reducing the respective concentrations of the solutions by half. The plates were sealed with Crystal Clear sealing tape (Hampton Research Ltd.). This ensured that the crystallisation drops were within a sealed system, allowing slow equilibration with the excess reservoir solution by vapour diffusion, bringing the drop to the point of supersaturation. Crystal trays were checked for crystals 2 days, 4 days, 7 days and 14 days after they were prepared.

2.12.2 Crystallisation Screens

Various commercial and in-house sparse matrix crystallisation screens were used in this project (Table 2.5). Sparse matrix screens sample a wide portion of crystallisation space, therefore identifying precipitants, buffers and additives, at approximate concentrations and pH that may facilitate crystal growth. From these initial conditions, variation of precipitant, buffer and additive concentrations, and pH, allow finer sampling of crystallisation space to identify conditions that yield diffraction quality crystals. In addition to sparse matrix screening, a grid screen, sampling ammonium sulphate concentration and various pHs, was also employed (Table 2.6)

Screen	No. Conditions	Source/Reference
M Screen	114	In-house
N Screen	60	In-house
Magic 50	50	Jancarik and Kim, 1991
PEG Footprint Screen	24	Stura, 1991
Crystal Screen I	48	Hampton Research Ltd.
Crystal Screen II	48	Hampton Research Ltd.
Wizard I	48	Emerald Biostructures
Wizard II	48	Emerald Biostructures
Cryo I	48	Emerald Biostructures
Cryo II	48	Emerald Biostructures

Table 2.5: Sparse matrix crystallisation screens

	2.0M AS	1.8M AS	1.6M AS	1.4M AS	1.2M AS	1.0M AS
0.1M Acetate pH 5.5						
0.1M MOPS pH 6.5						
0.1M HEPES pH 7.5						
0.1M Tris-HCl pH 8.5						

Table 2.6: Ammonium Sulphate Grid Screen

Chapter 3- Wet lab results

3.1 Cloning, expression and purification of type II DHQases as Maltose Binding Protein Fusion

The published purification strategy for type II DHQases involved five steps which were uneconomical and time consuming, therefore a new strategy utilising affinity chromatography was investigated. To over-express, and purify, type II DHQase, the pMAL expression system from New England Biolabs was used initially. This system involves expression of the target protein with an N-terminal Maltose Binding Protein (MBP) tag (Maina et al., 1988). This tag can act as a folding leader for proteins that fail to adopt native conformations under over-expression conditions (Kapust and Waugh, 1999), and, for our purposes as an affinity tag for purification. Located in the linker region between the MBP and the target protein is a Factor Xa protease cleavage site for removal of the MBP tag.

The *aroQ* gene that encodes the type II DHQase in *H. pylori* had previously been cloned from genomic DNA (Bottomley et al., 1996), and subcloned into pUC-18 as a *HindIII* - *SacI* insert. These restriction sites were unsuitable for direct sub-cloning of the gene into the pMAL-c2x vector as the restriction sites were incompatible, therefore *H. pylori aroQ* was amplified from the pUC-18 vector by the Polymerase Chain Reaction, with the addition of a 5' *BamHI* and a 3' *Hind III* restriction site.

The *aroD* gene from *M. tuberculosis* had previously been cloned into pMAL-c2x by our collaborators at GSK.

3.1.1 PCR amplification and sub-cloning of *aroQ* from *H. pylori*

Using the *H. pylori_aroQ*-pUC18 vector as template DNA, *aroQ* was amplified using 30 cycles of PCR as outlined in section 2-6, with the oligonucleotide primers DHPF and DHPR (Table 2.3), using the proofreading polymerase Pfu (Promega). Analysis of the PCR reaction was carried out on a 1% agarose gel stained with ethidium bromide, using the BVI DNA ladder (Boehringer) as a marker.

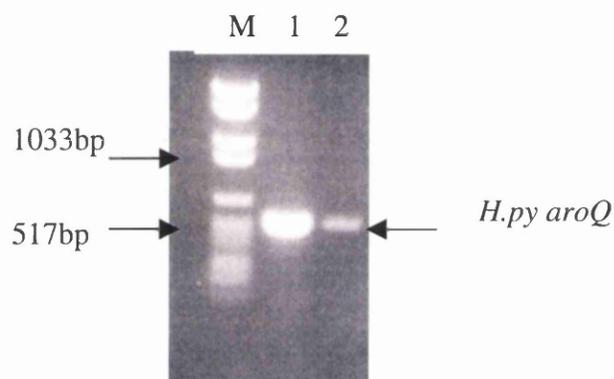


Figure 2.1: Agarose gel showing result of PCR amplification of *H. pylori aroQ*. M -marker BVI, 1 – 10 μ l PCR product, 2- 10 μ l PCR Product 1:10 dilution

From the gel it can be seen that the PCR produced a single section of DNA slightly greater than 517bp in length. The expected length of the PCR product is 530bp. The PCR product was washed using the QIAquick kit (Qiagen) using the manufacturer instructions, and eluted in 50 μ l of water.

Both the clean PCR product and pMAL-c2x were double digested for 4 hours at 37 $^{\circ}$ C using *Bam*HI and *Hind*III (Promega) and the products isolated by gel purification using the Qiaquick gel extraction kit (Qiagen). *H. pylori aroQ* was cloned into pMAL-c2x by ligation with T4 DNA ligase (Promega) to generate the vector pMAL-haQ. The ligation reactions were incubated at 16 $^{\circ}$ C overnight. The vector pMAL-haQ was transformed into chemically competent *E. coli* DH5 α cells

(Life Technologies) and plated onto LB-Agar plates containing 100µg/ml ampicillin. The plates were incubated overnight at 37°C and the number of colonies noted.

3.1.2 Identification of positive pMAL-haQ clones

Colonies were picked from the LB-Agar plates using clean sterile pipette tips and dotted onto fresh LB-Agar/Ampicillin plates then the remaining material on the tip was solubilised in 10µl sterile water and heated to 99°C for 10 minutes. 1µl of the heated mixture was used as template DNA for PCR reactions carried out as described previously. The product of each PCR reaction was run on a 1% agarose gel stained with ethidium bromide. Successful clones were identified as those that gave PCR products corresponding in size to *H. pylori aroQ*.

Successful clones were grown overnight at 37°C in 5ml LB/Ampicillin and plasmid DNA extracted using the Qia-quick mini-prep kit (Qiagen). Oligonucleotides were designed for DNA sequencing reactions incorporating the 5' and 3' ends of *H. pylori aroQ* and the respective flanking sequences in the pMAL-c2x sequence (Table 2.3). All successful clones were shown by sequencing to be in the correct reading frame for expression and no errors had been incorporated in the PCR process.

3.1.3 Over-expression of *H. pylori* and *M. tuberculosis* DHQase as a MBP fusion

The vectors pMAL-haq and pMAL-mtd were transformed into chemically competent *E.coli* BL21 DE3 (pLysS) cells (Invitrogen) for over-expression. To maximise the yield of target protein per preparation, expression studies were carried out to determine the optimal incubation time following induction with IPTG. A time course was followed using induced cells, and non-induced cells as a control.

Typically the time course ran, from the point of inoculation of the culture with IPTG, for three hours. The levels of protein expression were monitored by SDS PAGE (Fig. 2.2)

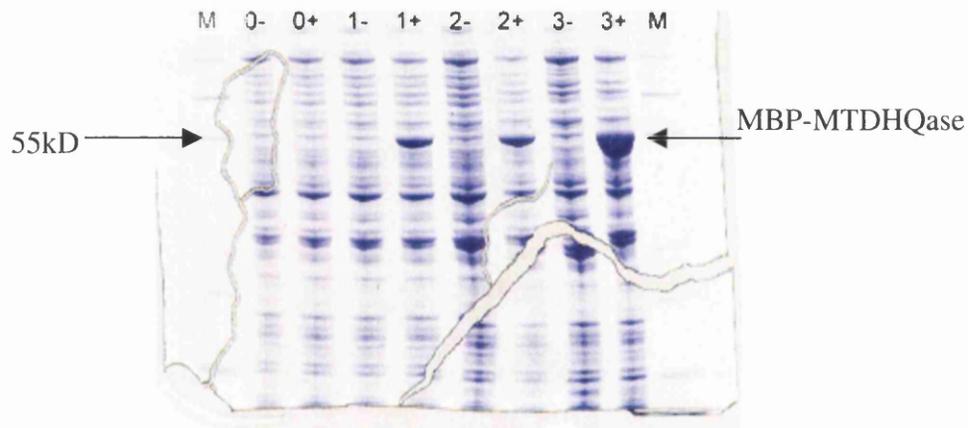


Figure 2.2 SDS PAGE time course of MBP-MTDHase expression after induction with 1mM IPTG, M- molecular weight marker, 0 – pre induction, 1 – 1 hour after induction, 2- 2 hours after induction, 3 – 3 hours after induction, - control, + IPTG added. All lanes have 10 μ l of sample loaded

The pairs of tracks on the gel shown in figure 2.2 show the level of protein expression over a 3 hour time course. Samples induced with IPTG are indicated (+ or -) with the corresponding incubation time in hours

3.1.4 Purification of MBP-DHQase fusion peptides

Following over-expression, each MBP-DHQase fusion was purified using a 50ml amylose affinity column as described in section 2.8.1 to yield protein typically greater than 95% pure as monitored by SDS PAGE (Figure 2.3)

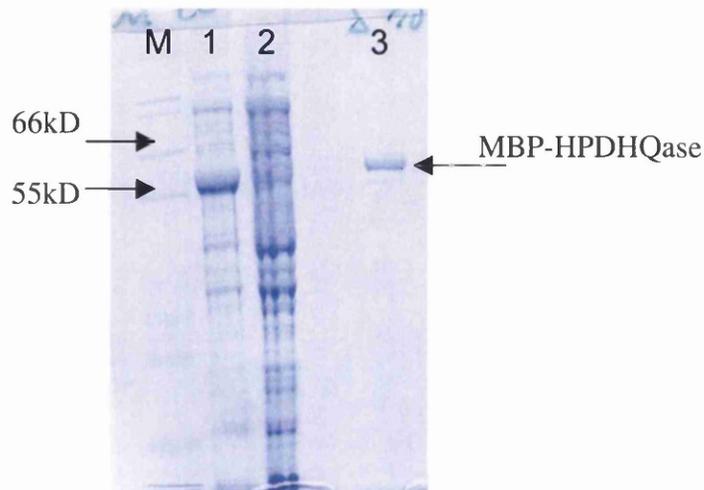


Figure 2.3 SDS PAGE of MBP-HPDHQase purification. M- molecular weight marker, 1-crude lysate, 2- column wash, 3- maltose eluted fraction MBP-HPDHQase. Lanes 1-3 have 10 μ l of sample loaded

Parallel experiments were carried out analysing expression of the *M. tuberculosis* construct (results not shown).

3.1.5 Factor Xa Cleavage of MBP-DHQase fusion peptides

To determine the optimal ratio of fusion peptide to factor Xa required for efficient cleavage of the MBP-DHQase fusion peptides, different ratios of protein:protease were incubated overnight at 4°C using MBP-MTDHase. The results were analysed by SDS PAGE (Figure 2.4).

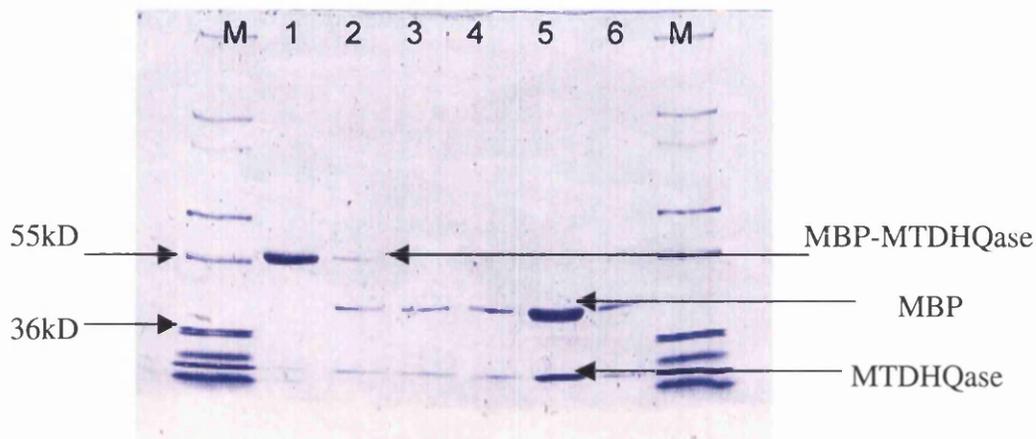


Figure 2.4 SDS PAGE showing results of MBP-MTDHase cleavage with Factor Xa. M- molecular weight marker, (1) MBP-MTDHase fusion, (2) 1000:1 peptide:protease, (3) 500:1 peptide:protease, (4) 100:1 peptide:protease, (5) 50:1 peptide:protease, (6) 10:1 peptide:protease. All lanes have 10µl sample loaded

The experiment shows that a 50:1 protein:protease ratio was optimal for overnight cleavage at 4°C

3.1.6 Separation of DHQase from MBP

The manufacturer's protocol for separation of the target protein from MBP following cleavage suggests re-loading the mixture of proteins onto the Amylose resin, allowing the MBP to bind to the column and the target protein to flow through. When this was attempted with the MTDHQase/MBP mixture the flow through fractions contained both proteins. The most likely reason for this was that the MBP fusion peptide was eluted from the amylose resin using maltose, which has been shown to be difficult to remove from MBP (Silhavy et al., 1975). The retention of maltose by the MBP prevents binding of MBP to the Amylose resin. To overcome this problem three purification strategies were investigated. Firstly separation by size exclusion chromatography, secondly ion exchange chromatography and finally heat shocking the mixture to take advantage of the heat stability of type II DHQase. The results were analysed by SDS PAGE (Figure 3.5).

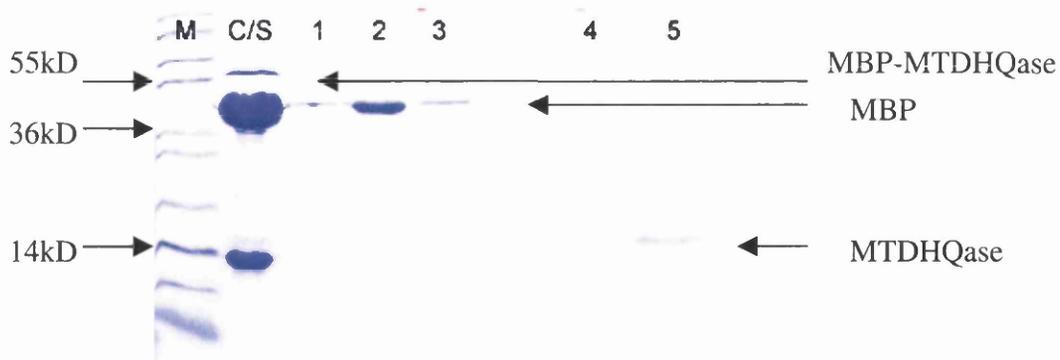


Figure 3.5a Separation of MBP/MTDHQase by ion exchange chromatography monitored by SDS PAGE, M- Molecular weight marker, C/S crude sample, 1-5 ion exchange fractions. All lanes have 10 μ l of sample loaded

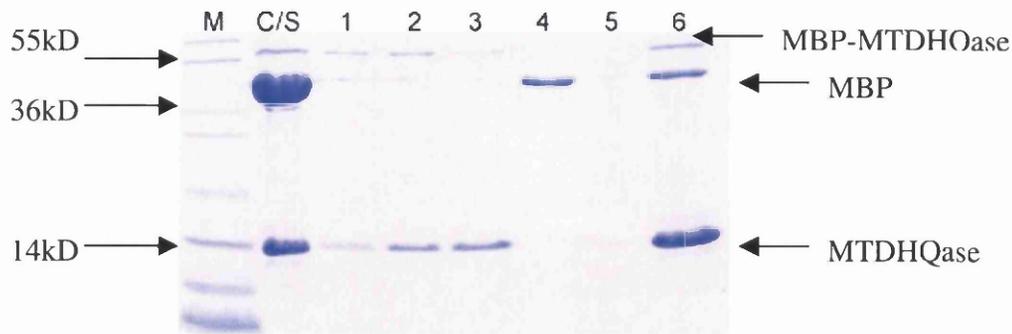


Figure 3.5b Separation of MBP/MTDHQase by size exclusion chromatography and heat shocking monitored by SDS PAGE. M- molecular weight marker, C/S – crude sample, 1-5 – size exclusion fractions, 6- Heat shock soluble fraction. All lanes have 10 μ l sample loading

These experiments indicated that ion exchange chromatography using a mono-Q anion exchange column proved the best method of purification of the type II DHQases. The results of a typical purification of are shown below (Figure 3.6)

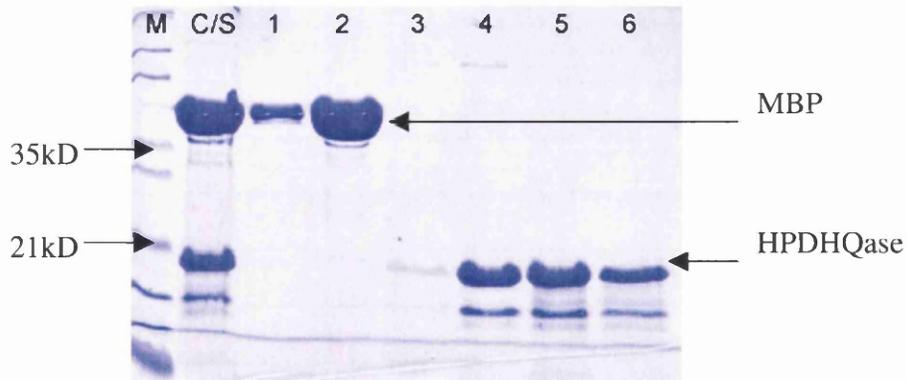


Figure 3.6 SDS PAGE monitoring separation by ion exchange chromatography of MBP/HPDHQase, M- molecular weight marker, C/S- crude sample, Fractions 1-2 MBP, Fractions 3-6 HPDHQase. All lanes have 10 μ l sample loaded

A small band approximately 4kD smaller than the main 19kD band corresponding to HPDHQase was observed on SDS PAGE. This is likely to be a

product of the protease cleavage of the MBP-HPDHQase fusion peptide, although no factor Xa cleavage site was identified in the protein sequence.

3.2 Crystallisation of type II DHQases expressed as MBP fusion peptides.

All crystallisation experiments were performed by the 'sitting drop' vapour diffusion method. Initial screening for crystallisation conditions was performed using a variety of in-house sparse matrix screens (Jancarik and Kim, 1991). Optimisation of initial conditions was achieved by varying precipitant concentration, pH, additives and protein concentration as described in section 2.12. Co-crystallisation experiments with transition state analogues (Table 2.4) were carried out by incubating protein with 100mM inhibitor in a 25:1 (v/v) ratio at room temperature for 30 mins before setting up the crystallisation experiment.

3.2.1 Crystallisation of MTDHQase

Initial crystallisation experiments used protein concentrated to 10mg/ml, however this led to the formation of protein 'skins' on top of the crystallisation drops. This indicated that this protein concentration was too high and subsequent crystallisations use protein concentrations of 7-8 mg/ml. Sparse matrix screening yielded small crystals of MTDHQase/FA3 and MTDHQase/FA4 complexes in 20% (w/v) PEG 8K, 0.2M NaCl, 0.1M Tris-HCL pH 8.5. These conditions were optimised and larger crystals were obtained in 15% (w/v) PEG 8K, 0.2M NaCl, 0.1M Tris-HCl pH 8.8 (Figure 3.7). These crystals were taken to the DESY synchrotron (Hamburg), however no diffraction was observed.

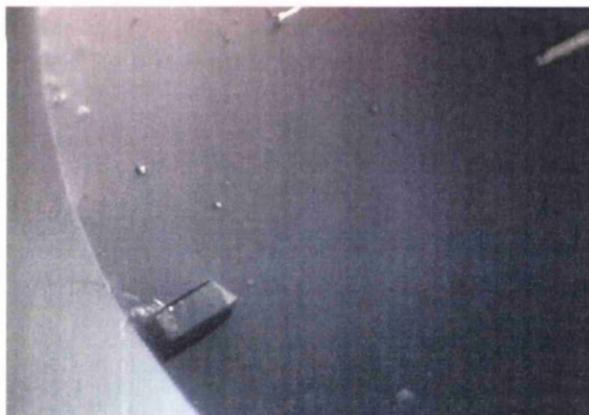


Figure 3.7 Crystal of MTDHQase + FA4 in 15% (w/v) PEG 8K, 0.2M NaCl, 0.1M Tris-HCl pH

8.8

3.2.2 Crystallisation of HPD HQase

Initial crystallisation experiments used HPD HQase concentrated to 10mg/ml. Initial crystals of HPD HQase/FA1 complex were obtained in 20% (w/v) PEG 4K, 0.2M MgCl₂, 0.1M Tris-HCl pH 8.5. These conditions were optimised to yield diffraction quality crystals in 14% (w/v) PEG 4K, 0.2M MgCl₂, 0.1M Tris-HCl pH 8.8 (Figure 3.8). The crystals were octahedral in morphology, with dimensions 0.05mm x 0.05mm x 0.05mm. These crystals were found to diffract to 3.2Å on beamline BW-7B (DESY, Hamburg)



Figure 2.8 Crystal of HPD HQase + FA1 grown in 14% (w/v) PEG 4K, 0.2M MgCl₂, 0.1M Tris-HCl pH 8.8

3.2.3 Reasons for poor diffraction

Crystals grown using protein expressed via the MBP fusion system proved difficult to grow and showed poor or no diffraction when tested at a synchrotron source. This was almost certainly due to the poor quality of the protein preparation (Section 3.1.6). In particular, when cleavage of the fusion peptides had not gone to 100% completion, a fraction of the DHQase still carried the MBP tag. As type II DHQase is dodecameric, it is possible that dodecamers may exist with some monomers still tagged, while others have been cleaved successfully.

During crystallogenesis rogue tags will poison the crystal growth, resulting in small crystals of poor quality. In the case of HPDQHase a significant degradation product was observed on SDS PAGE that was not removable by chromatographic methods. This impurity may also contribute to the poor quality of the crystals.

From this point, expression of DHQases as MBP fusions was abandoned, and a new expression and purification strategy adopted.

3.3 Cloning, expression and purification of type II Dehydroquinases with an N-terminal poly-histidine tag.

Expression and purification of type II DHQases using the MBP fusion expression system had proven to be problematic, and did not result in the growth of diffraction quality crystals. However, affinity chromatography methods were still applied to ease purification of type II DHQases using an N-terminal poly-histidine tag. The vector pET-15b (Invitrogen) was used to express type II DHQase with a 6x His tag, followed by a thrombin protease cleavage site, joined by a linker region to the N-terminal of the protein.

3.3.1 Sub-cloning of *M. tuberculosis aroD* and *H. pylori aroQ* into pT7-Blue3

The multiple cloning site of pET-15b contains only three restriction sites in the following order (5'-3') *Nde I*, *Xho I* and *Bam HI*. These sites were not compatible with the restriction sites in *pMAL-mtd* and *pMAL-hpq* therefore new restriction sites had to be engineered by PCR. Primer oligonucleotides were designed to incorporate a 5' *Nde I* site and a 3' *Bam HI* site on *M. tuberculosis aroD* and *H. pylori aroQ* (Table 2.3). PCR was carried out using these primers, Taq polymerase, with *pMAL-mtd* and *pMAL-hpq* as template DNA to amplify each gene.

The PCR products were purified by gel electrophoresis and purified using the QIEX II kit (Qiagen) and cloned into p-T7-Blue3 (Novagen) by blunt end ligation (Section 2.6.3) using the 'Perfectly Blunt' Cloning kit (Novagen) to generate the vectors p-Blue-mtd and p-Blue-hpq. The products from the ligation reactions were transformed into Novablue Singles Competent Cells (Novagen) and plated out on

Agar plates containing 100µg/ml Ampicillin + IPTG + X-gal. This allowed identification of successful transformants by blue/white screening.

Three colonies were picked and grown overnight in 5ml LB + 100µg/ml ampicillin, and plasmid DNA prepared by the method of Birndoin and Dolly (1979). Positive clones were identified by restriction digests and were sequenced, to confirm that no mutations had accumulated during the PCR process.

3.3.2 Sub-cloning of *M. tuberculosis aroD* and *H. pylori aroQ* into pET-15b

The vector pBlue-mtd and pBlue-hpq were digested with *Nde I* and the linear vector DNA isolated by gel electrophoresis. The linear fragments were digested with *Bam HI*, releasing the 0.5kb insert from the vector DNA, which were isolated by gel electrophoresis. The vector pEE-65 (pET-15b with a 1.1kb insert) was obtained from Dr. E. Ellis (University of Strathclyde) and digested in a similar manner, with the 5.7kb pET-15b vector DNA purified.

Ligation reactions were carried out and the resultant constructs transformed into *E. coli* DH5α and plated out on agar plates + 100µg/ml ampicillin. Ten colonies were picked and 5ml cultures grown at 37°C overnight in LB + 100µg/ml ampicillin. Plasmid DNA was isolated using the QISpin kit (Qiagen) and restriction digest analysis performed with *Nde I* and *Bam HI*.

3.3.3 Expression studies upon 6xHis tagged DHQase

To optimise the overexpression of target protein from each preparation, expression studies were carried out to identify the optimal concentration of IPTG required for maximal expression. The vectors MpET (encoding 6xHis – MTDHQase)

and HpET (encoding 6xHis- HPDHQase) were transformed in to *E.coli* BL21 (plysS) cells and colonies propagated at 37°C overnight in 2ml of LB + 100µg/ml ampicillin + 34µg/ml chloramphenicol. 200µl of each overnight culture were used to inoculate 4x2ml LB + 100µg/ml ampicillin, which were grown at 37°C until OD₆₀₀ reached 0.6. The four cultures were inoculated with 0mM, 0.4mM, 0.8mM and 1.0mM IPTG respectively and grown for a further 3 hours. Total cell protein from the crude cell extract of these cultures were analysed by SDS PAGE (Figure 3.9)

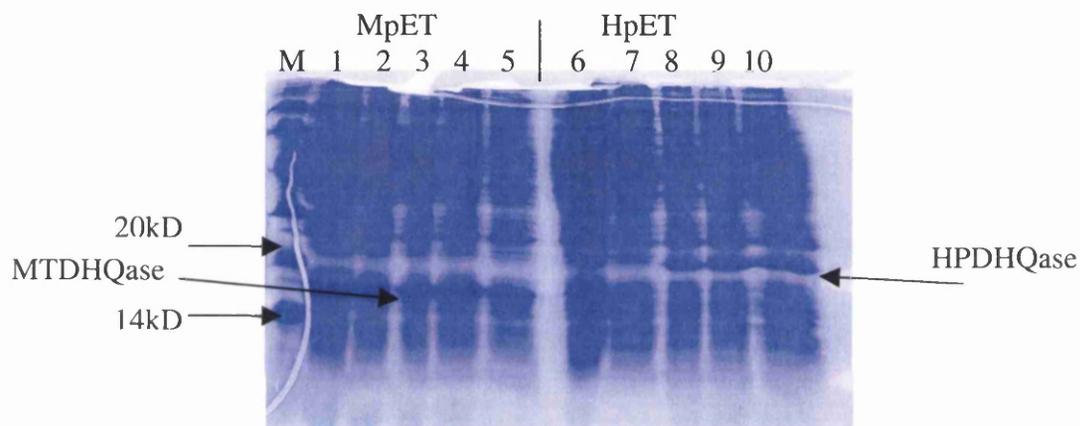


Figure 3.9 Monitoring of expression of 6xHis- DHQases by SDS PAGE. M- molecular weight marker, 1-2 MpET no induction, 3-5 3 hours after induction with 0.4mM, 0.8mM and 1mM IPTG respectively, 6-7 HpET no induction, 8-10 3 hours after induction with 0.4mM, 0.8mM and 1mM IPTG respectively. All lanes have 5µl of sample loaded

From this experiment it can be seen that maximal expression occurred after induction with 0.8mM IPTG.

3.3.4 Purification of 6xHis tagged DHQases

Following over-expression, 6xhis tagged DHQases were purified by affinity chromatography using a 5ml Ni-NTA column as described in section 2.8.1 to yield protein of purity greater than 99%. Ion exchange chromatography (section 2.8.2) using a 10ml Q-sepharose column was used to further purify and concentrate the protein samples. Protein purification was monitored by SDS PAGE (Figure 3.10)

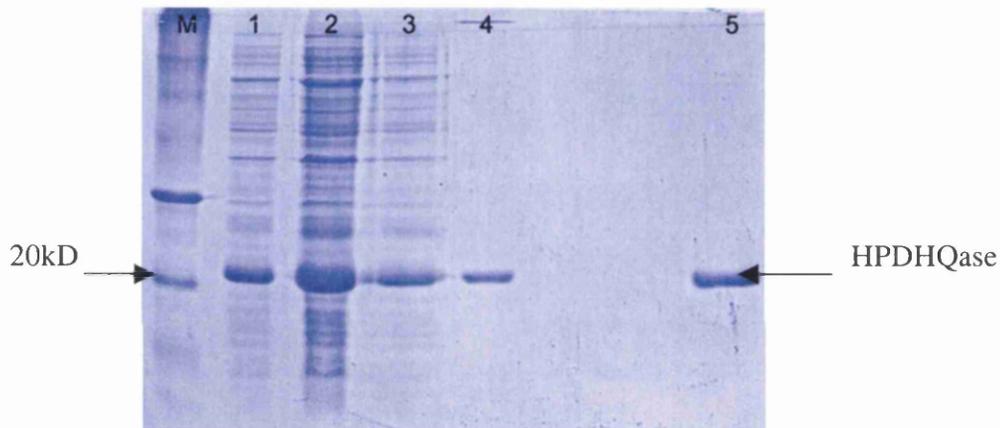


Figure 3.10 Monitoring purification of 6xHis-HPDHQase by SDS PAGE. M- molecular weight marker, 2- cell lysate, 3- Ni-NTA wash1, 4- Ni-NTA wash2, 4- Ni-NTA product, 5- Q-sepharose peak fraction. All lanes have 10 μ l of sample loaded

3.4 Crystallisation of 6xHis tagged DHQases

The problems encountered after cleaving the MBP affinity tag from previous DHQase constructs, namely separating the target protein from the tag and non-specific cleavage of the target protein, were deemed to be detrimental to the growth of diffraction quality crystals. Therefore no attempt was made to remove the 6xHis tag from the target protein before crystallisation experiments.

3.4.1 Crystallisation of 6xHis-MTDHQase

Samples of 6xHis-MTDHQase were concentrated using centricon centrifugal concentrators (Amicon) to 7mg/ml and crystallisation experiments carried out using the sitting drop vapour diffusion method. In-house sparse matrix screens and the commercial crystallisation screens Crystal screen I, Crystal screen II, Wizard I, Wizard II, Cryo I and Cryo II (Emerald Biostructures) were used, however no promising crystals were observed from these screens

3.4.2 Crystallisation of 6xHis-HPDHQase

Co-crystallisation experiments were carried out with transition state analogues FA1, FA3, FA4 and GSK inhibitors AH9095, GR78752X and GR122110X. The transition state analogues were water soluble however as they are acidic compounds the pH of the inhibitor solution had to be adjusted. The GSK inhibitors were insoluble in aqueous solution therefore were dissolved in 100% DMSO. Protein was concentrated to 10mg/ml in 10mM Tris-HCl pH 7.5, and shaken overnight at 4°C in the presence of 50mM inhibitor. Crystallisation experiments using the sitting drop vapour diffusion method were set up using in-house sparse

matrix screens and the commercial crystallisation screens Crystal screen I, Crystal screen II, Wizard I, Wizard II, Cryo I and Cryo II (Emerald Biostructures).

3.4.2.1 Crystallisation of HPDHQase and FA1

Initial sparse matrix screening provided several initial crystallisation conditions for this complex and two conditions were optimised to provide diffraction quality crystals. Initial crystals grown in 20% MPD, 0.1M Citrate pH4.5 were tested at BM ID14.4 (ESRF, Grenoble) and showed diffraction to 3.5Å resolution. Larger crystals (0.2mm x 0.1mm x 0.1mm) were obtained following optimisation of conditions to 16% MPD, 0.1M Citrate pH4.2, however no increase in resolution was obtained.

A second crystal form was identified grown in 30% 1,4 butandiol, 0.1M Na Acetate pH 4.5 (Figure 3.11). These crystals diffracted to 3.1Å at beamline 9.5 (SRS, Daresbury) and a complete data set was collected.



Figure 3.11 Crystal of HPDHQase + FA1 grown in 30% 1,4 butandiol, 0.1M Na Acetate pH 4.5

3.4.2.2 Crystallisation of HPDHQase and AH9095

Small octahedral crystals were observed in wells containing 1M Na Acetate, 0.1M Imidazole-HCl pH 8.0 during initial screening. This condition was optimised to 1.0M Na Acetate, 0.1M Imidazole-HCl pH 8.2 to yield crystals of dimensions 0.3mm x 0.3 mm x 0.3mm after two days growth (Figure 3.12). These crystals diffracted to 2.5Å on the in-house rotating anode source (Nonius FR591) and to 1.55Å at beamline 9.6 (SRS, Daresbury). Complete data sets were collected using both sources.

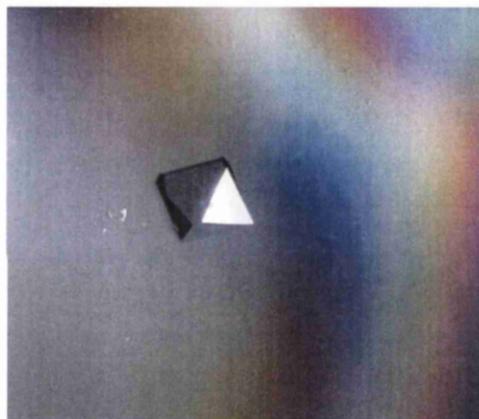


Figure 3.12 Crystal of HPDHQase + AH9095 grown in to 1.0M Na Acetate, 0.1M Imidazole-HCl pH 8.2

3.4.2.3 Crystallisation of HPDHQase and GR122110X

After initial screening, crystals were observed in wells containing 20% 1,4 butandiol, 0.1M Na Acetate pH 4.5 (Figure 3.13). These crystals diffracted to 2.9Å resolution on beamline 9.6 (SRS, Daresbury) and a complete data set collected. Despite attempts to optimise the crystallisation conditions no improvement in diffraction was observed.

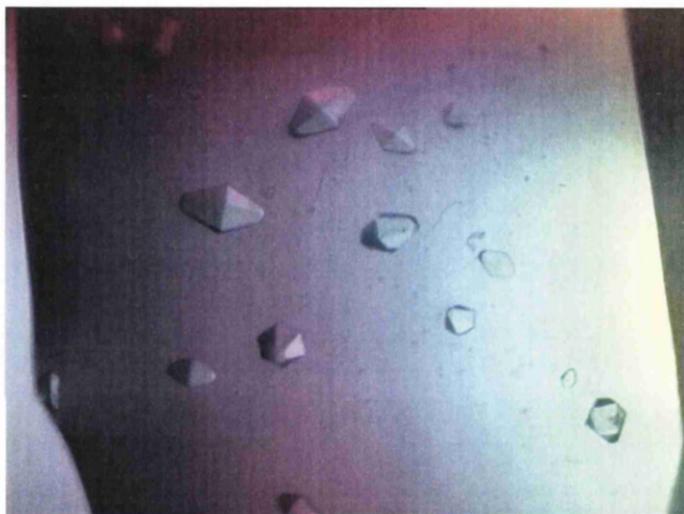


Figure 2.13 Crystals of HPDHQase + GR122110X grown in 20% 1,4 butandiol, 0.1M Na Acetate pH 4.5

3.4.2.4 Crystallisation of HPDHQase and Phosphate

Phosphate ions have been shown to inhibit type II DHQase (Kleanthous *et al.*, 1992), and have been observed bound in the active site of MTDHQase and *S. coelicolor* DHQase (Evans *et al.*, 2002). Sparse matrix screening was carried out upon HPDHQase in the presence of 10mM Ammonium phosphate. Octahedral crystals of dimensions 0.1mm x 0.1mm x 0.1mm were observed in conditions containing 1M ammonium acetate, 0.1M Tris-HCl pH 7.5 (Figure 3.14) after two weeks. These crystals showed no diffraction on the in-house rotating anode source. A second crystal form was identified in conditions containing 30% Isopropanol, 0.2M ammonium acetate, 0.1M Tris-HCl pH 7.5 however no diffraction was observed using the in-house rotating anode source.

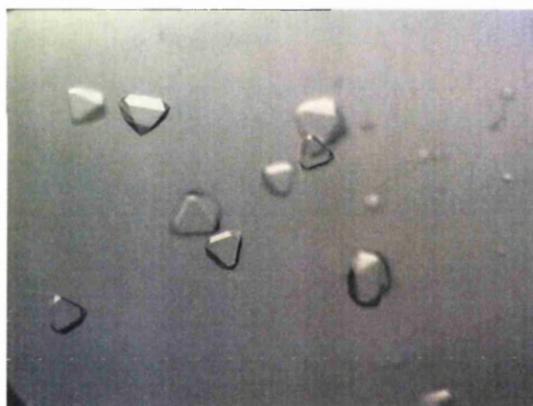


Figure 3.14 Crystals of HPDHQase + 10mM Ammonium Phosphate grown in 1M Ammonium Acetate, 0.1M Tris-HCl pH 7.5

3.4.3 Soaking experiments with HPDHQase

The best crystals obtained were those of HPDHQase complexed with AH9095. Soaking experiments were attempted to exchange this ligand for other ligands within the crystal. Large numbers of crystals of HPDHQase + AH9095 were prepared. Drops containing the corresponding crystallisation conditions were prepared with high concentrations of GR78752X, GR122160X or FA1 instead of the HPDHQase/AH9095 mixture, and left to equilibrate. After 48 hours, crystals of HPDHQase + AH9095 were transferred to the pre-equilibrated drops and incubated for a further 48 hours before diffraction experiments were carried out.

GR78752X was only soluble in DMSO, and the presence of high concentrations of DMSO resulted in destruction of the crystals. Therefore solid GR78752X was added to drops of pre-equilibrated artificial mother liquor prior to addition of the crystal. A crystal treated in this manner diffracted to 2.5Å on beamline 9.5 (SRS, Daresbury) and a complete data set collected. However, after the data were processed and the structure phased it was apparent that AH9095 had not been competed out of the active site.

Soaking experiments using GR122110X and FA1 resulted in complete loss of diffraction from the crystals. This suggests that these compounds probably displaced AH9095, but this also results in disrupting the crystal contacts, hence disturbing the order of the crystals.

3.5 Discussion

Samples of MTDHQase and HPD HQase over-expressed and purified using the pMAL expression system and amylose resin affinity chromatography were obtained in high yield, however crystals grown using these samples were of poor quality. This can be attributed to problems with obtaining mono-disperse samples of the proteins, a result of non-complete cleavage of the MBP fusion peptide, and subsequent incomplete separation of the cleaved MDP from the DHQase.

Diffraction quality crystals of HPD HQase were obtained with protein that was over-expressed using the pET-15b expression vector, and subsequently purified using Ni-NTA affinity chromatography. The best crystals were obtained when the protein was co-crystallised with the ligand AH9095, however crystals grown in the presence of other ligands were of diffraction quality. No crystals of MTDHQase expressed and purified by this protocol were obtained.

Despite repeated attempts, the crystallisation conditions for HPD HQase published by Kwak *et al.* (2001) were non-reproducible. It is noted that in our hands we used HPD HQase with an N-terminal 6xHis tag plus linker, whereas Kwak *et al.* purified and crystallised native protein.

Chapter 4 – Collection and Processing of X-ray diffraction data

4.1 Protein Crystallography

Since the pioneering work of Bernal and Crowfoot upon the collection of an X-ray diffraction pattern from a crystal of pepsin (Bernal and Crowfoot 1934) and the determination of the structure of myoglobin in 1957 (Kendrew *et al.*, 1958), the field of protein crystallography has become one of the most powerful tools in modern biology. The determination of the precise 3D structure of biological macromolecules not only provides topological information, but can also provide unprecedented insight into enzymatic regulation and function.

Despite the development of crystallographic methods and innovations in computational techniques, the underlying theory of protein structure determination has remained unchanged since the birth of the field. Many excellent texts have been published that describe crystallographic theory (Blundell and Johnson, 1976; Drenth, 1994; McRee, 1999), therefore in this thesis only the basic concepts and methodologies that relate directly to the work carried out will be described.

4.1.1 Scattering of X-rays by an atom

When atoms are placed in a beam of electromagnetic radiation, the atoms begin to oscillate, thus becoming a source of radiation themselves, scattering the incident beam. Electrons within an atom oscillate at the same frequency as the incident radiation and therefore emit radiation at the same frequency as the incident radiation (Figure 4.1).



Figure 4.1: Electromagnetic radiation scattered by a single atom

If the atoms are in a periodic arrangement (such as in a crystal), the scattered waves interfere with each other, resulting in constructive and destructive interference depending upon the phase of the waves (Figure 4.2).

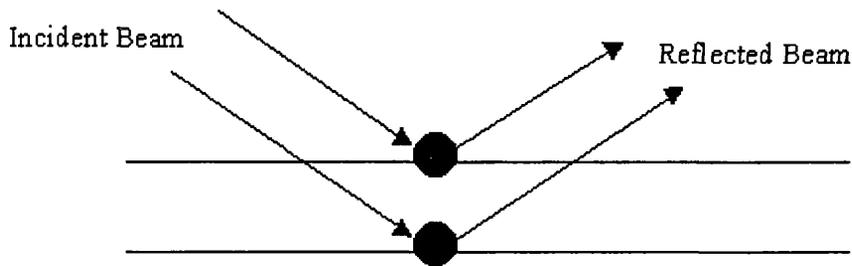


Figure 4.2: Scattering from a periodic arrangement of atoms produces a characteristic diffraction pattern

4.1.2 Bragg's Law

If we visualise a crystal as a series of planes of atoms, we can consider diffraction arising from the reflection of the incident beam by planes within the crystal.

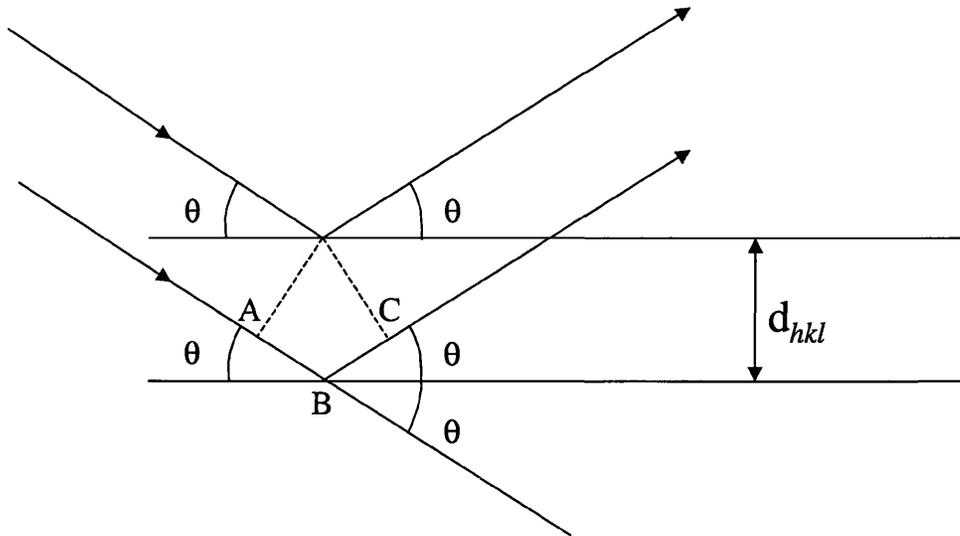


Figure 4.3 The Bragg construction describing reflection of X-rays from planes of atoms within a crystal.

Figure 4.3 illustrates the Bragg's theory of the reflection of X-rays by 'reflecting planes' of atoms within a crystal lattice. The path difference between waves scattered by atoms from adjacent lattice planes of spacing d is given by

$$(AB + BC) = (d\sin\theta + d\sin\theta) = 2d\sin\theta$$

For constructive interference this path distance must equal an integer number of wavelengths of the incident X-rays.

$$n\lambda = 2d\sin\theta$$

where n is an integer and λ the wavelength of the incident radiation.

4.1.3 The Ewald Construction

Despite describing the condition required for diffraction of X-rays by a crystal, the Bragg construction does not illustrate the geometric conditions required for collection of all structure factors from a diffracting crystal. The Ewald construction, or synthesis, is a geometrical formulation of Bragg's law describing the interaction between the reciprocal lattice and a 'limiting sphere of reflection' (Figure

4.4). The radiation of wavelength λ is represented by a sphere of radius $1/\lambda$ centred upon the X-ray beam. The crystal is represented by the reciprocal lattice, with its origin at the point on the Ewald sphere where the direct beam leaves it. The relationship between the Ewald construction and the Bragg construction is as follows; each reciprocal lattice point lies at the end of a vector perpendicular to a plane within the crystal. Any reciprocal lattice point that lies upon the surface of the Ewald sphere fulfils the criteria for Bragg diffraction.

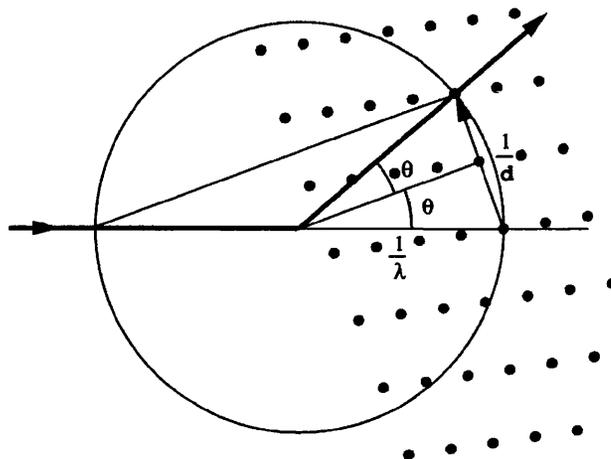


Figure 4.4 The Ewald construction; The Ewald sphere is shown as a circle of radius $1/\lambda$ centered upon the origin of the reciprocal lattice. The reciprocal lattice is shown as green dots (Figure from Dauter , 1999)

4.2 Collection of X-ray diffraction data by the Rotation Method

At any particular crystal orientation, only a few lattice points will be in a diffracting position, Therefore, to collect a full set of diffraction data either the Ewald sphere must move (by changing the wavelength of the incident radiation), or the reciprocal lattice must move (by rotating the crystal). By rotating the crystal, successive lattice points are brought to the surface of the Ewald sphere, into diffracting position. This method is known as the rotation method, and is the method

most commonly used in the routine collection of X-ray diffraction data from crystals of macromolecules.

4.2.1 The Rotation Method

In the rotation method the crystal is rotated, therefore the reciprocal lattice is rotated, such that lattice points are brought to a position where they satisfy the criteria for Bragg reflection of monochromatic radiation. Figure 4.5 illustrates how rotation of the lattice results in different lattice points intersecting the Ewald sphere.

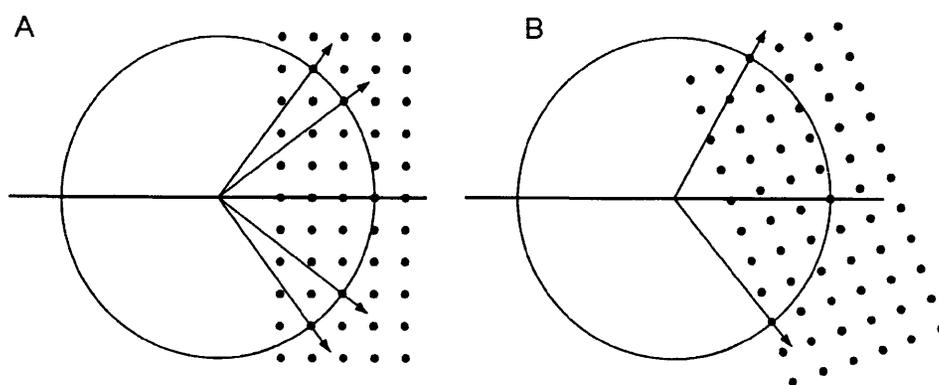


Figure 4.5 Rotation of the reciprocal lattice with respect to the Ewald sphere. Panel A shows the diffracted X-rays from one orientation, Panel B shows the diffracted beams generated when we rotate the reciprocal lattice. (Figure from Dauter, 1999)

For visualisation purposes it is easier to represent rotation of the crystal during data collection by moving the Ewald sphere and keeping the lattice stationary. Figure 4.6 shows the movement of the Ewald sphere around an axis in the plane of the page. This shows that an area of reciprocal space is missed by the rotation of the Ewald sphere; and area known as the blind region. To fully record these reflections, the crystal must be re-orientated such that this area of reciprocal space is traversed by the Ewald sphere.

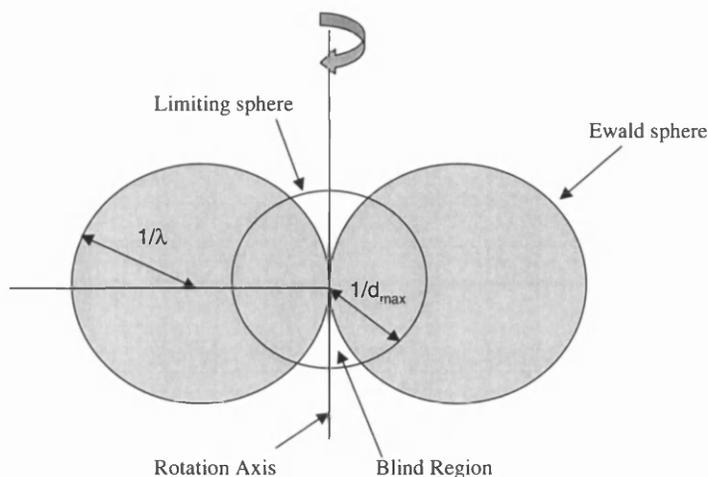


Figure 4.6; Rotation of the Ewald sphere around an axis normal to the plane of the paper illustrating the area of reciprocal space traversed and the blind region

The entire reciprocal lattice cannot be sampled in a single rotation as this would generate too many spots on the image to allow accurate indexing. Therefore the rotation is split into a series of oscillations, typically $0.5^\circ - 2^\circ$, after which the diffraction pattern is recorded. The choice of oscillation angle is critical for the collection of a high quality data set. If the oscillation is too large, too many spots will appear on the image, resulting in spot overlap. If the range is too small, only a few reflections will be recorded per image, with the majority of reflections being partially recorded over several images. The theoretical optimum oscillation angle, $\Delta\phi_{\max}$, can be calculated by considering the separation between adjacent spots, a^* , the maximum resolution, d_{\max} , and the maximum oscillation angle which a particular spot is observable, ϕ_r , in the following manner

$$\Delta\phi_{\max} = a^* / d_{\max} - \phi_r$$

The variable ϕ_r is dependent upon several factors, notably the mosaicity of the crystal and characteristics of the incident X-ray beam.

4.2.2 Crystal Mosaicity

To determine the optimum oscillation range, the degree of mosaicity, or mosaic spread, of the crystal has to be taken into consideration. Crystals are not perfect and do not consist of a uniform lattice. Instead they exist as a series of mosaic blocks, which do not merge precisely with neighbouring blocks (Figure 4.7)

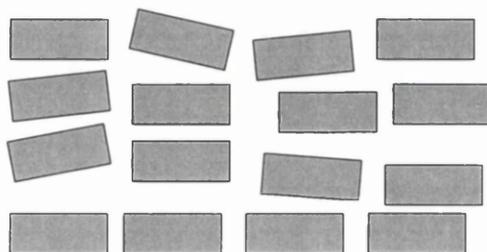


Figure 4.7 Illustration of a macromolecular crystal with high mosaic spread

The result of this feature is that reciprocal lattice points become smeared, i.e. the perfect lattice point would consist of an infinitesimal point in space, however in reality they are smudged, covering a larger area of space. If now consider the Ewald sphere passing through the reciprocal lattice, the larger lattice points will remain in the diffracting position, hence may be spread over many images.

4.3 Data Collection Strategy

The first stage in collecting a high quality X-ray diffraction data set is to determine features of the crystal relating to the data collection. Determination of;

- Unit Cell Dimensions
- Laue Class
- Approximate mosaicity
- Diffraction limit of the crystal

These factors relate to the strategy employed in the data collection, and can be approximately determined from initial processing of a few initial images.

Consideration of the unit cell dimensions and mosaicity influence the oscillation angle employed in the data collection. If there is a particularly long unit cell length, this results in diffraction spots overlapping each other, possibly too close to resolve. By reducing the oscillation angle (fine slicing), less spots will appear on the image reducing spot overlap. Similar considerations are made for highly mosaic samples.

The diffraction limit significantly influences the strategy for collecting a complete data set. The intensity of diffraction spots drops off as the Bragg angle increases, i.e. high resolution data is inherently weaker than low resolution data. If high resolution data is to be collected, exposure times may increase resulting in overloaded reflections at lower resolution. To collect high resolution data, unless the detector has the capacity to move off axis, the detector will be close to the sample. If the detector is very close to the sample, the spots near the edge of the image will be well resolved whereas those close to the centre of the image will overlap. Due to the considerations above if high resolution data collection using a single rotation axis is to be collected, a high resolution data set is collected and processed omitting poorly measured low resolution data. A second pass through reciprocal

space with a greater crystal to detector distance is collected to record the low resolution data.

The oscillation range required to collect the unique portion of x-ray diffraction data is determined by the space group of the sample. If the crystal is in a low symmetry space group, greater than 180° of data must be sampled. Conversely if the crystal is of a high symmetry space group less than 30° data needs to be collected.

4.4 Collection of X-ray diffraction data from *Helicobacter pylori* type II

Dehydroquinase + AH9095

In section 3.4.4.2 the growth of large crystals of HPDHQase + AH9095 was described. In this section, the data collection strategy used to obtain a complete data set is described.

4.4.1 Initial data collection of HPDHQase + AH9095 using a rotating anode source

A large single crystal of HPDHQase + AH9095 was taken from a well using a 0.2mm cryoloop (Hampton) and immersed in mother liquor + 20% glycerol to act as a cryoprotectant. The crystal was then frozen in a stream of gaseous dry nitrogen at 100K using an Oxford Systems Cryostream and orientated into the centre of the X-ray beam. The rotating anode source was a Nonius FR591, with a copper target generating Cu K α radiation (wavelength 1.5418Å) equipped with focussing mirrors. The goniometer had a single ϕ rotation axis and the detector was a MAC Science DIP 2000 double image plate. The crystal to detector distance was adjusted to 140mm, resulting in a d_{\max} at the edge of the image plate of 2.5Å. The crystal was

exposed for 20 minutes per image, with an oscillation angle of 1° . Figure 4.8 shows a typical image.

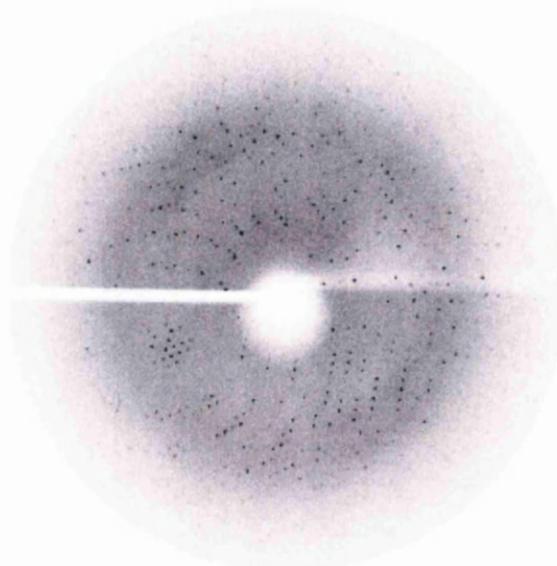


Figure 4.8 A typical image of the diffraction pattern measured from a crystal of HPDHQase + AH9095 using the rotation method with a rotating anode source.

4.4.2 Determination of data collection strategy for HPDHQase + AH9095

A single initial image was processed using DENZO from the HKL suite (Otwinowski and Minor, 1997). The first step in the process is to identify the well defined reflections on the image and attempt to index them. This generates initial values for the unit cell dimensions and the orientation of the crystal in relation to the incident beam.

The well defined reflections are picked as opposed to the strongest reflections as the indexing routine requires very accurate spot positions. At the centre of a strong reflection the detector may become overloaded, therefore many pixels will have the same intensity. Overloaded spots have poor profiles, therefore the exact centre of the spot cannot be accurately determined. If we choose a slightly weaker spot, the intensity will not be as great at the centre of the spot, giving rise to a more accurate spot profile, therefore the centre point will be easily determined (Figure 4.9).

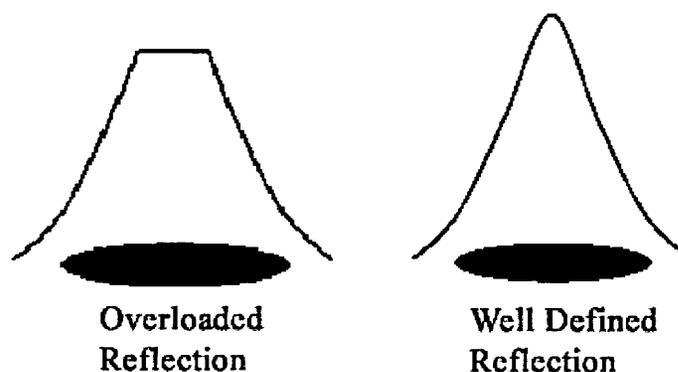


Figure 4.9: Spot profiles for and overloaded reflection and a well defined reflection.

To determine which Bravais lattice type the crystal belongs to, a cell with the initial unit cell dimensions is generated. This cell is then forced to obey the constraints of each Bravais lattice and a deformation matrix (a measure of the force required) is calculated for each lattice type. The lattice types with corresponding unit cell dimensions and deformation matrices are displayed in order of symmetry. The lattice type with the highest symmetry, with the lowest deformation matrix is the correct solution. Using these hypothesised values, a data collection strategy can be formulated. Determination of the space group can only be determined once the data has been scaled by analysis of the merging R- factor (R_{merge}) statistics (section 4.4.3)

For HPDHQase + AH9095, an F centred cubic unit cell was the suggested Bravais lattice with unit cell dimension $a = b = c = 130.5\text{\AA}$ $\alpha = \beta = \gamma = 90^\circ$. For this high symmetry space group very little unique data is required (7.5°) for a complete data set, therefore 30° of data were collected (30 images, 1° oscillation) to give an overall multiplicity of 3.4.

4.4.3 Processing of data - HPDHQ + AH9095

The 30 images collected were processed with the HKL suite. The data processing routine is complex and only a brief overview of the methods implemented in the HKL suite will be described. The initial stages are performed with the program DENZO. From the predicted crystal characteristics (unit cell dimensions, space group, mosaicity, crystal orientation) the program predicts the position of reflections on the image. The intensity of these regions is measured, a process referred to as spot integration.

Denzo performs 2D spot integration by mapping the profile of the diffraction spot. A box is created around the predicted spot position, within which areas are defined as spot, safety and background (Figure 4.9). The average intensity across the background area is calculated and subtracted from the intensity within the spot region. The safety region acts as a buffer between spot and background to prevent rogue spots which are smeared into the background region affecting the intensity calculation.

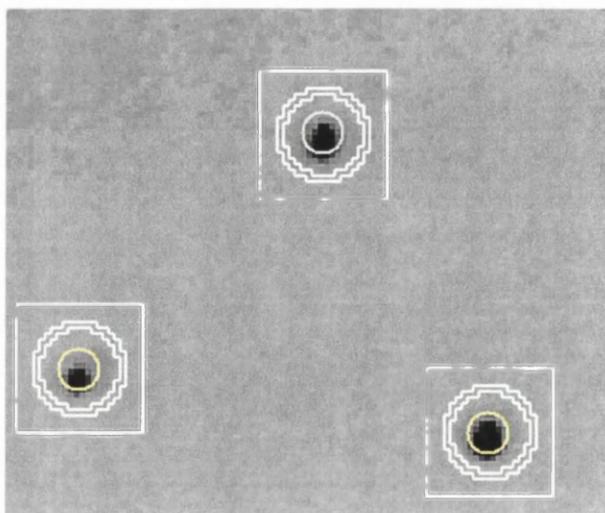


Figure 7.9 Spot integration using Denzo. The predicted spot position is indicated by the coloured circles. The area within the central white circle is the spot area, the area between the two white circles the safety region. The remaining area within the box is defined as background.

Profile fitting is then carried out upon the spot intensity profiles. The profiles of spots within a defined radius are analysed to create a predicted spot profile. The observed and predicted spot profiles are then combined, the spot positions defined and an accurate spot intensities calculated.

The subsequent processing is carried out with the program Scalepack. Due to differences in crystal and beam characteristics throughout the data collection, a scale factor has to be applied to the data to take these differences into account. Two corrections are applied to the integrated spot intensities as the observed spot intensity I_o to estimate the desired intensity measurement I , the Lorentz-Polarisation corrections.

$$I = \frac{I_o}{LP}$$

Diffraction images collected using the rotation method show a distorted view of the reciprocal lattice due to the diffraction geometry. Reciprocal lattice points that lie on, or near, the rotation axis will spend longer traversing through the Ewald sphere than reciprocal lattice points far from the rotation axis. This is corrected by the Lorentz kinematical factor (L). The polarisation correction (P) relates to the polarisation content of the X-ray beam as it travels from source to detection.

Scalepack also presents statistics about how complete a data set is, what the average intensity of the reflections above noise level ($I / \sigma I$) and an R-factor (R_{merge} or R_{sym}) that describes the quality of the data. The definition of R_{merge} is shown below

$$R_{merge} = \frac{\sum I - \langle I \rangle}{\sum \langle I \rangle}$$

The data collection statistics for HPDHQ + AH9095 are shown in table 4.1

HPDHQ + AH9095	In-house data collection
Wavelength	1.5418Å
Detector	DIP 2000
Crystal to detector distance	140mm
Resolution range	30 – 2.53Å
Completeness	99%
Completeness Highest resolution shell	97%
Total number of reflections	20,693
Unique Reflections	6,274
Average Redundancy	3.4
I / σ I	11.6
I / σ I Highest resolution shell	2.25
R _{merge}	9.9%
R _{merge} Highest resolution shell	51.1%

Table 4.1 Statistics for HPDHQase + AH9095 data collected using in-house rotating anode source

4.5 High resolution data collection – HPDHQase + AH9095

Crystals of HPDHQase + AH9095 similar to those used for in-house experiments were taken to the Synchrotron Radiation Source (SRS) at Daresbury, UK, for diffraction experiments with synchrotron radiation.

4.5.1 Synchrotron radiation sources

A synchrotron circulates charged particles (positrons or electrons) around a 'ring' at nearly the speed of light. The particles are accelerated using a linear accelerator (LINAC) and injected in the synchrotron storage ring. The orbit of the particles is controlled by magnets (~1 Tesla, ESRF), which deflect the path of the particles. When the particles change direction, they emit electromagnetic radiation at a tangent to the path deflection. This radiation can be harnessed for many experimental purposes, including X-ray crystallography. X-rays can be generated from the particle beam either at the bending magnets (which are used to control the orbit) or at insertion devices such as wigglers and undulators.

4.5.2 Advantages of Synchrotron radiation

There are many advantages associated with diffraction experiments using synchrotron radiation as opposed to $\text{CuK}\alpha$ radiation from a rotating anode source. The characteristics of the X-ray beam generated at a synchrotron source are ideal for the collection of diffraction data from macromolecular crystals. The intensity of the beam generated from a synchrotron is at least 2 orders of magnitude greater than that generated from a rotating anode. For modern synchrotrons (e.g. ESRF Grenoble FR, APS Argonne USA, Spring-8 JP) the radiation generated can be up to 7 orders of magnitude more intense. This results in much shorter exposure times for the collection of diffraction images (< 10 secs per image - beamline ID14-4, ESRF) and the observation of higher resolution reflections, too weak to be observed with less intense radiation. Synchrotron radiation is highly focused onto the sample and the beam itself has very low divergence, resulting in small, well defined diffraction spots.

The wavelength of the radiation focussed upon the sample is typically around 1.0Å, however many synchrotron beamlines can be 'tuned', allowing a wide spectrum of wavelengths to be selected. This attribute is used in the collection of multiple wavelength anomalous dispersion (MAD) and single wavelength anomalous dispersion (SAD) data to obtain experimental phasing information. The shorter exposure times coupled with the shorter wavelength of the incident radiation results in less background scatter, therefore a better signal to noise ratio for the data collection.

Data was collected upon crystals of HPDHQase + AH9095 at beamline 9.6 at SRS Daresbury. The radiation is generated by a wiggler insertion device and the resultant beam has a wavelength of 0.87Å. The beamline is equipped with an ADSC quantum 4r CCD detector with an active area of 188mm x 188mm.

4.5.3 High resolution data collection statistics – HPDHQase + AH9095

A single crystal of HPDHQase + AH9095 was looped from a well and transferred briefly to an artificial mother liquor + 20% (v/v) glycerol to act as a cryoprotectant. The crystal was the flash frozen in stream of dry gaseous nitrogen cooled to 100K. Initial 1 minute exposures with an oscillation angle of 1° were collected and diffraction spots were observed to ~1.5Å. A typical image is shown in Figure 4.10.

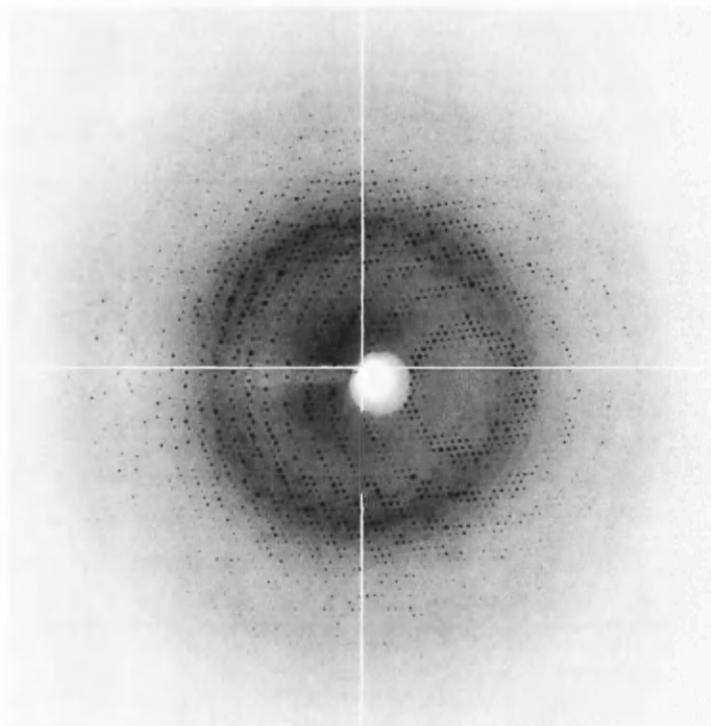


Figure 4.10 A typical diffraction pattern from HPDHQase + AH9095 collected at SRS beamline 9.6 using the ADSC Quantum-4r detector

Initial processing of the images showed the crystal to be isomorphous with the crystal used for in-house data collection, therefore the data collection strategy was identical. Due to the large number of reflections that are systematically absent in $F23$ (table 4.2) indexing of all the reflections on the image was possible without any problems of spot overlap.

hkl	$h+k, h+l, k+l = 2n$
$0kl$	$k, l = 2n$
hhl	$h+l=2n$
$h00$	$h = 2n$

Table 7.2 Reflection conditions for the space group $F23$. h , k and l are cyclically permeable

Despite the exposure time being optimal for collection of high resolution data, this did not result in overloaded reflections at low resolution, therefore there

was no need for high and low resolution passes. The data collection consisted of 120 images of 1° phi oscillation, with 1 minute exposures. Final data collection statistics are shown in table 4.3.

HPDHQ + AH9095	SRS 9.6 data collection
Wavelength	0.87Å
Detector	ADSC Q4r CCD
Crystal to detector distance	163mm
Resolution range	50 – 1.5Å
Completeness	100%
Completeness Highest resolution shell	100%
Total number of reflections	334,701
Unique Reflections	27,302
Average Redundancy	12.2
$I / \sigma I$	59.1
$I / \sigma I$ Highest resolution shell	2.01
R_{merge}	4.4%
R_{merge} Highest resolution shell	80.8%

Table 4.3 Statistics for HPDHQase + AH9095 data collected on SRS beamline 9.6

The data set collected at SRS 9.6 was of good quality despite the very high R_{merge} in the highest resolution shell. Highly redundant data improves the $I / \sigma I$ (signal to noise), however R_{merge} is also influenced by the redundancy of the data, with highly redundant data increasing the R_{merge} . Data is usually cut off where 50% of the reflections within the resolution shell have $I / \sigma I = 3$, however data beyond this resolution can still be useful and is therefore included.

4.6 Collection of diffraction data from other HPDHQase crystals

4.6.1 HPDHQase + GR78752X Soak

Data were collected from a crystal of HPDHQase soaked in a solution of GR78752X (section 3.4.3). The data were collected at beamline 9.5 at SRS (Daresbury). The crystal diffracted weakly to 2.4Å resolution and the space group was F23. A typical diffraction image is shown in Figure 4.11

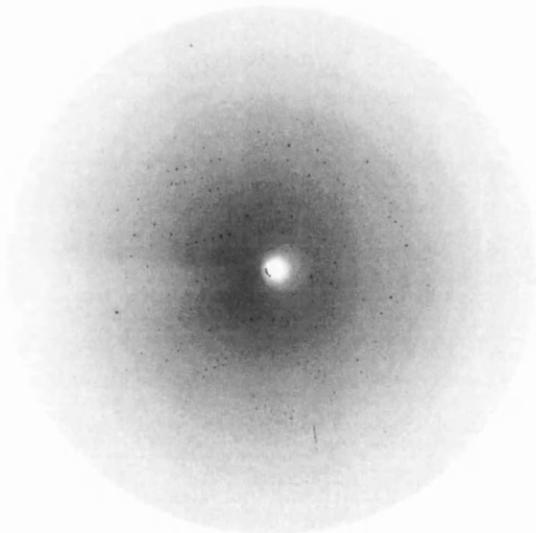


Figure 4.11 A typical diffraction image from a HPDHQase crystal soaked in GR78752X recorded on beamline 9.5 at SRS (Daresbury) using the MAR-165 CCD detector.

Due to the low resolution of the diffraction and the large number of systematic absences in the space group F23 the data was collected in a single pass. 56 frames were collected with an oscillation angle of 1°. Data collection and processing statistics are shown in table 4.4.

HPDHQ + GR78752 Soak	SRS 9.5 data collection
Space Group	F23
Unit Cell	a = b = c = 130.53Å
Wavelength	1.0Å
Detector	MAR 165 CCD
Crystal to detector distance	170mm
Resolution range	50 – 2.4Å
Completeness	99.4%
Completeness Highest resolution shell	99.5%
Total number of reflections	38,721
Unique Reflections	7,489
Average Redundancy	5.17
I / σ I	11.05
I / σ I Highest resolution shell	1.52
R _{merge}	12.1%
R _{merge} Highest resolution shell	79.0%

Table 4.4 Data collection statistics for HPDHQ + GR78752 soak.

4.6.2 HPDHQase + FA1

Data were collected from crystals of HPDHQase + FA1, prepared as described in section 3.4.2.1. The data were collected at beamline 9.5 at SRS (Daresbury). Diffraction spots were observed to 3.1Å resolution and initial data processing suggested the crystals belonged to a primitive trigonal space group. A typical diffraction image is shown in figure 4.12. 130 frames were collected with an oscillation angle of 1°. Data collection and processing statistics are shown in table 4.5.

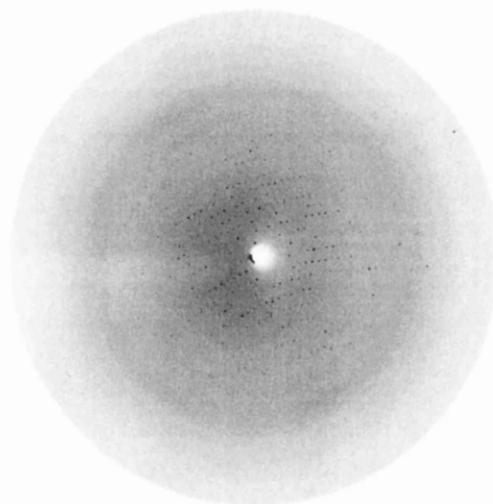


Figure 4.12 A typical diffraction image from a HPDHQase + FA1 crystal recorded on beamline 9.5 at SRS (Daresbury) using the MAR-165 CCD detector.

HPDHQ + FA1	SRS 9.5 data collection
Space Group	$P3_1$
Unit Cell	$a = b = 103.86\text{\AA}$ $c = 217.53\text{\AA}$
Wavelength	1.0\AA
Detector	MAR 165 CCD
Crystal to detector distance	190mm
Resolution range	$50 - 3.3\text{\AA}$
Completeness	92.4%
Completeness Highest resolution shell	82.2%
Total number of reflections	124,789
Unique Reflections	37,103
Average Redundancy	3.36
$I / \sigma I$	9.64
$I / \sigma I$ Highest resolution shell	1.47
R_{merge}	8.0%
R_{merge} Highest resolution shell	44.8%

Table 4.5 Data collection statistics for HPDHQ + FA1

As this data indexed in a trigonal system there are 5 possible Laue groups that the crystal could belong to. To begin the space group determination the data were indexed in $P3_1$, the subsequently scaled in space groups representative of each Laue group (Table 4.6).

Space Group	Laue Group	$R_{\text{merge}} (\%)$	Chi^2
$P3_1$	-3	9.5	0.86
$P3_12$	-3m1	10.3	0.85
$P3_121$	-31m	10.3	0.85
$P6_1$	6/m	10.3	0.85
$P6_122$	6/mmm	10.6	0.85

Table 4.6 Scaling statistics for HPDHQase + FA1 data processed in the 5 possible Laue groups that obey a primitive trigonal Bravais lattice.

There does not appear to be significant rise in R_{merge} as that data is scaled in space groups with higher symmetry, however analysis of the systematic absences showed that the general reflection condition $(0,0,3n)$ was obeyed indicating that the crystal system was trigonal as opposed to hexagonal (where the condition $(0,0,6n)$ would be satisfied). The data were therefore scaled in space groups belonging to both $-3m1$ and $-31m$ Laue groups and molecular replacement attempted (Section 5.5). No molecular replacement solution was found in any $-3m1$ or $-31m$ space group, therefore the data was re-processed in a lower symmetry space group $P3_1$. In this case a molecular replacement was found indicating that $P3_1$ was indeed the correct space group.

4.6.3 HPDHQase + GR122110

Data were collected from crystals of HPDHQase + GR122110, prepared as described in section 3.4.2.3. The data were collected at beamline 9.6 at SRS (Daresbury). Diffraction spots were observed to 2.9Å resolution and initial data processing suggested the crystals belonged to a rhombohedral space group R32. A typical diffraction image is shown in figure 4.13a. The presence of spots along one direction with very fine d-spacing (Figure 4.13b) suggested that one unit cell axis was very long. The calculated unit cell dimensions for this crystal were $a = b = 182.95\text{Å}$ $c = 658.83\text{Å}$.

Due to the long unit cell axis, an oscillation angle of 0.5° was used per frame and 120 frames were collected. Data collection statistics are shown in table 4.7

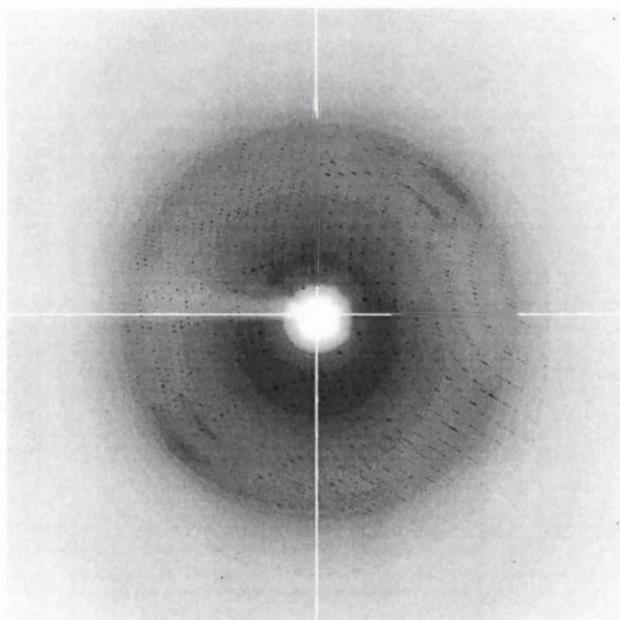


Figure 4.13a A typical diffraction image from a HPDHQase + GR122110 crystal recorded on beamline 9.6 at SRS (Daresbury) using the ADSC Quantum 4r CCD detector.

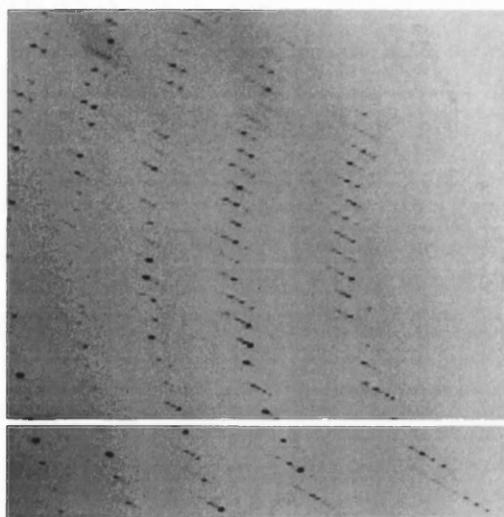


Figure 4.12b Diffraction spots with small d-spacing collected from a crystal of HPDHQase + GR122110. The small d-spacing relates to the long length of one of the unit cell axes.

HPDHQ + GR122160	SRS 9.6 data collection
Space Group	R32
Unit Cell	$a = b = 182.95\text{\AA}$ $c = 658.834\text{\AA}$ $\gamma = 120^\circ$
Wavelength	0.87\AA
Detector	ADSC Q4r CCD
Crystal to detector distance	240mm
Resolution range	$30 - 3.1\text{\AA}$
Completeness	99.4%
Completeness Highest resolution shell	99.9%
Total number of reflections	262,992
Unique Reflections	76,880
Average Redundancy	3.42
$I / \sigma I$	8.46
$I / \sigma I$ Highest resolution shell	1.58
R_{merge}	10.2%
R_{merge} Highest resolution shell	45.3%

Table 4.7 Data collection statistics for HPDHQ + GR122110

4.7 Structure Factor Equation

The information obtained from X-ray diffraction data collection is the intensities of Bragg reflections and their Miller indices hkl . Each Bragg reflection has a corresponding structure factor (F_{hkl}), which describes the amplitude and phase of the reflection. The standard notation of the structure factor is shown below

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

As each Bragg reflection can be represented as a wave, the structure factor can be described in terms of an amplitude $|F_{hkl}|$ and phase α

$$F_{hkl} = |F_{hkl}| \exp(2\pi i \alpha_{hkl})$$

From the X-ray diffraction data we have recorded the reflection intensities (I_{hkl}) which are related to the square of the structure factor intensities as shown. The figure k is a scaling factor.

$$|F_{hkl}| = k \sqrt{I_{hkl}}$$

The conversion of Bragg reflection intensities to structure factor amplitudes was carried out using the CCP4 program Truncate (CCP4, 1994).

Unfortunately no estimation of the phase of the structure factor can be obtained from observed X-ray diffraction data. A complete estimation of the structure factor can't be calculated directly from a single data set, unless there are atoms present that do not obey classical scattering of X-rays. This phenomenon is anomalous scattering which facilitates phase estimation by Multiple wavelength Anomalous Dispersion (MAD) or Single wavelength Anomalous Dispersion (SAD), however as neither of these techniques were carried out in the work in this thesis they will not be described.

If an analogous protein structure has been determined previously, the molecular replacement technique (Rossmann and Blow, 1962), can be applied to gain initial phase information

4.8 Conclusion

Complete sets of X-ray diffraction data have been collected for HPDQHase + AH9095 using an in-house rotating anode X-ray source and a synchrotron radiation source. The in-house data set was complete to 2.5Å resolution and the synchrotron data set complete to 1.5Å resolution. The observed reflection intensities were converted to structure factor amplitudes and the data sets prepared for phase estimation by molecular replacement.

Complete data sets were also collected from a crystal of HPDQHase soaked in GR75872 to 2.4Å resolution and HPDQHase + FA1 to 3.3Å resolution, both of which could be taken forward for phasing by molecular replacement. A data set from a crystal of HPDQHase + GR122110 was collected to 3.1Å, however the very large unit cell dimensions calculated for this crystal would make phasing a refinement problematic therefore this data was taken no further.

Chapter 5 – Structure solution of *H.pylori* DHQase by Molecular Replacement

In order to gain initial phase information for structure refinement a variety of techniques is currently routinely performed. If no analogous structure information is available, methods such as multiple isomorphous replacement (MIR), multi-wavelength anomalous dispersion (MAD) and in some cases direct methods, can be applied to obtain initial phase information. In the case of HPDHQase, the crystal structures of the type II DHQases from *M. tuberculosis* and *S. coelicolor* have been solved previously, therefore the molecular replacement (MR) technique was applied for structure determination. Many molecular replacement packages are currently available such as AMoRe (Navazza, 1994), MOLREP (Vagin and Teplyakov, 1997), BEAST (Read, 2001) and EPMR (Kissinger *et al.*, 1999). As each program treats the MR process differently, an overview of the methods applied in this thesis will be given.

5.1 Molecular Replacement

The molecular replacement method, first proposed by Rossmann and Blow in 1962, was first developed to utilise non-crystallographic symmetry (NCS) to reduce phase uncertainties in *ab-initio* structure determination. However, as the number of structures in the protein data bank (PDB) has grown, molecular replacement has been used to gain initial phase information using the structure of a homologous protein as search model. The position of the molecular model within the crystal is determined by the rotation (**R**) and the translation (**T**) that move the model from a reference initial position (\mathbf{r}'), to the current position (\mathbf{r}).

$$r = Rr' + T$$

Equation 1: Position r after rotation (R) and translation (T) from reference point r'

Most MR packages are based upon the properties of the Patterson function, which requires no phase information.

$$P_{(u, v, w)} = \frac{1}{V} \sum_h \sum_k \sum_l |F_{hkl}|^2 \cos 2\pi(hu + kv + lw)$$

Equation 2: Patterson function. $P_{(u,v,w)}$ is a contour map in which peaks correspond to interatomic vectors

As the molecular replacement problem has six search variables (three rotational and three translational), to compute all six variables simultaneously, despite modern computing speeds, is not feasible. Therefore most MR packages split the search into a rotation search, followed by a translation search, and the Patterson function too can be split into parts that are sensitive to these parameters.

1. Self-Patterson vectors - Short intramolecular vectors between atoms of the same molecule
2. Cross-Patterson vectors – Vectors arising from intermolecular atoms

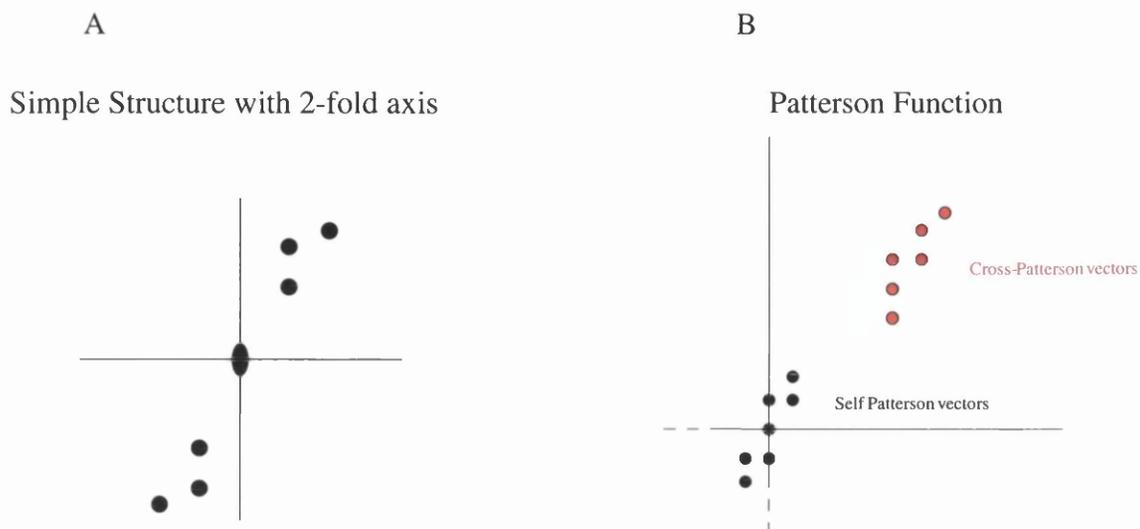


Figure 5.1: (A)- two tri-atomic molecules related by a two-fold rotation axis and (B) the resulting Patterson function showing self-vectors (black) and cross vectors (red)

5.1.1 The Rotation Function

For MR to be successful, the rotation applied to the search model must be as accurate as possible, increasing the likelihood of a successful translation search. The rotation function should find the orientation of the search model that results in the maximal overlap of the Patterson self-vectors as these are sensitive to the orientation of the model, but not to interactions with other molecules (i.e. position within the crystal). As it is only the correlation between the self Patterson vectors that are of interest, only a shell of Patterson space is analysed omitting the cross Patterson vectors and the Patterson origin peak. For validation of the rotation function, a correlation coefficient is calculated between the search model and the target structure.

Many current MR packages such as AMoRe utilise the Crowther fast rotation function (Crowther, 1972) which approximates Pattersons by spherical harmonics,

allowing the rotation function to be calculated by FFT. Packages such as AMoRe and MOLREP take the top ranking solutions from the rotation function and proceed to perform the translation search.

5.1.2 The Translation Function

Having determined the correct rotation of the search model, the search model can then be translated within the target unit cell, taking into account symmetry elements. Patterson methods can again be used, this time analysing the overlap of the cross Patterson vectors, however many current MR packages calculate an R-factor between the search model and its symmetry mates as they are translated around the unit cell.

$$R = \frac{\sum_h \left| |F_h(obs)| - k |F_h(calc)| \right|}{\sum_h |F_h(obs)|}$$

Equation 5.3: R factor calculated between observed and calculated structure factor amplitudes for the translation function, k is a scaling factor

The solutions found with the lowest R-factors are considered to be the top ranking results. The final stage in MR is to perform rigid body refinement upon the best solutions and calculate a final R factor for that solution.

5.1.3 Multidimensional Molecular replacement

As stated previously, traditional approaches to molecular replacement have been split into a rotation search followed by a translation search. Although this

approach results in greatly reduced computation times, there are various disadvantages associated with the methodology. Firstly, there is a very low signal to noise ratio observed for the critical rotation function, and secondly, a successful translation search is heavily reliant upon an accurate rotation function. Six dimensional searches involve large amounts of calculations and until recently were seen to be unfeasible (Sherriff *et al.*, 1999). However, with advances in computational strategy such as genetic algorithms and evolutionary programming, six dimensional searches have been successfully implemented in molecular replacement packages (Kissinger *et al.*, 1999).

5.1.4 Molecular Replacement by Evolutionary Search

The program EPMR (Kissinger *et al.*, 1999) uses an evolutionary algorithm to perform rapid six dimensional molecular replacement searches. In evolutionary algorithms, the global optimum of a function is found through the iterative optimisation of a population of initially random trial solutions. As each cycle progresses, each solution is ranked and the top solutions retained to serve as 'parents' of the next generation. Random variations are applied to the 'parent' solutions to generate a new population and the process repeated iteratively. In the case of molecular replacement by evolutionary searching, an initial population of molecular replacement solutions is generated, assigning a random orientation and position to each member of the population. A correlation coefficient is calculated for each solution and compared with a small number of other randomly chosen solutions. The number of these 'competitions' that a particular solution wins is used to rank the solutions and to determine which solutions 'survive' to be parents in the next round.

These solutions enter the next stage unmodified, but are also used to generate the new population by applying random variations of orientation and position. The general method applied by EPMR is as follows.

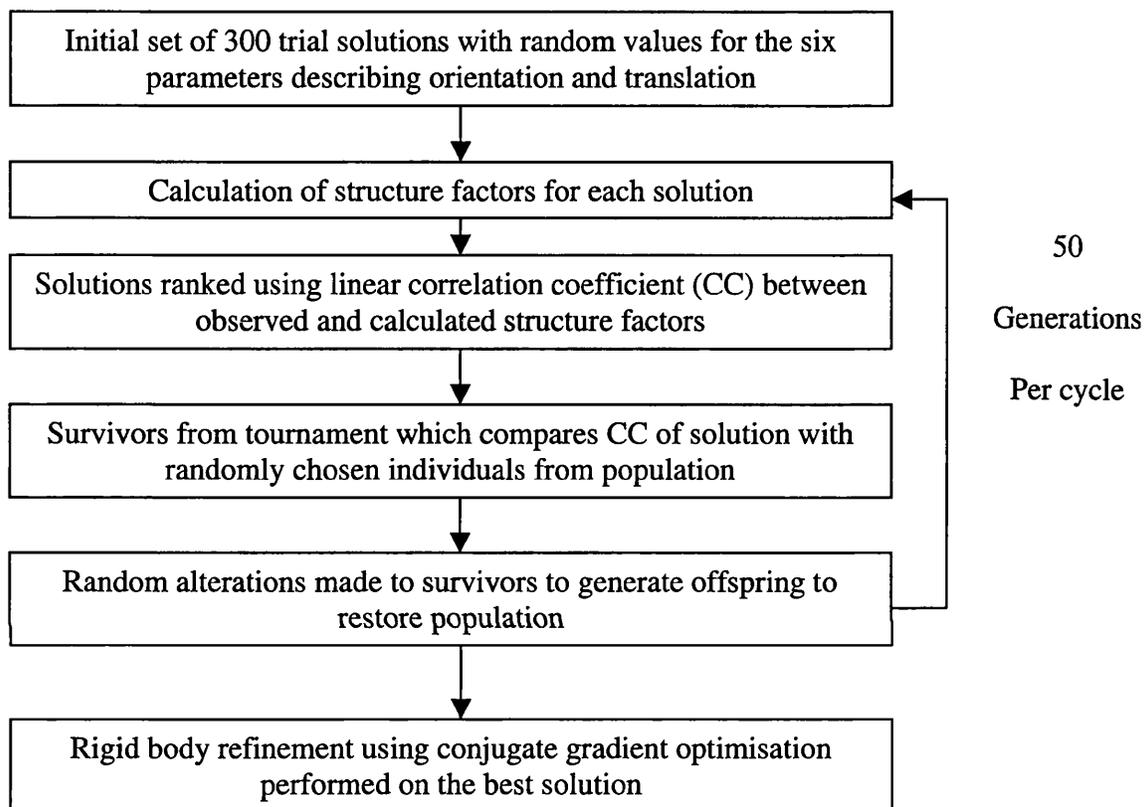


Figure 5.2: Outline of the procedure carried out by EPMR for molecular replacement.

5.2 Attempts to solve HPDHQase structure using AMoRe

In previous work upon complexes of MTDHQase, molecular replacement using AMoRe had been successful in structure solution. However, when AMoRe was used upon the data from HPDHQase + AH9095, no obvious solutions were found. A variety of search models was used, including main chain traces of MTDHQase, SCDHQase and type II DHQase from *Bacillus subtilis* (BSDHQase). The main

problem appeared to be that no significant peaks were observed in the rotation function, resulting in a poor translation search (table 5.1)

Model	% Sequence identity	CC for highest peak in Rotation function	No of peaks >50% maximum peak height	Best solution CC	Best solution R (%)
MTDHQ	33.5	17.8	32	28.5	67.4
SCDHQ	34.1	12.8	53	24.7	64.1
BSDHQ	37.5	13.3	46	25.4	67.9
MT_MAIN	-	19.6	34	33.2	66.7
SC_MAIN	-	19.5	34	38.1	65.9
BS_MAIN	-	15.4	48	34.7	64.8

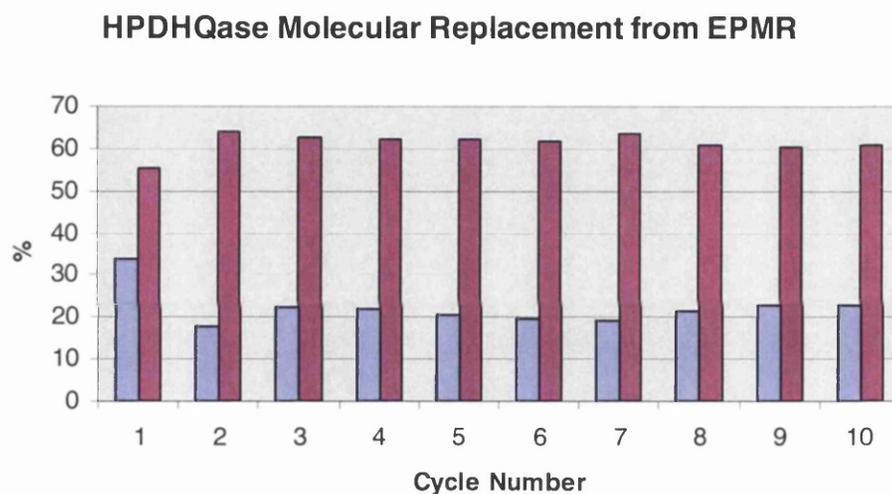
Table 5.1: Summary of results from attempts to solve HPDHQase by molecular replacement with AMoRe

One reason for the failure of AMoRe to find a correct solution for the rotation function may occur due to the high symmetry of the HPDHQase crystals. As outlined previously, many MR packages compare self-Patterson vectors to analyse the success or failure of a rotation solution. In a space group such as $F23$ there is a great deal of overlap between the areas of Patterson space occupied by self-vectors and cross-vectors, as the unit cell is tightly packed with extensive crystal contacts in all directions. This may result in problems partitioning self and cross vectors for rotation validation.

To address this issue the evolutionary search program EPMR was used for the molecular replacement.

5.3 Structure solution of HPDHQase using EPMR

As ongoing attempts to solve the HPDHQase structure by conventional molecular replacement methodologies, EPMR was used in parallel. After initial experimentation with the program, a poly-alanine model based upon the MTDHQase structure was used as the search model. A population size of 400 was generated, with a minimum intermolecular distance (as calculated from the centre of mass of the model) of 20\AA input. Initially the program was set to run for 100 cycles, but was terminated after 10.



Graph 5.1: Graph showing Correlation Coefficient (blue) and R-factor (maroon) for 10 cycles of evolutionary searching using EPMR.

The first solution from EPMR had a correlation coefficient of 33.8% and an R factor of 55.4%, better than any solution thus far generated by AMoRe.

To determine the validity of the molecular replacement solution, the packing of monomers within the unit cell generated from the model was analysed with Quanta (Accelrys) to identify any steric clashes (Figure 5.3). No steric clashes were apparent within the unit cell suggesting the MR solution was correct.

This procedure does not rely upon the atomic coordinates of a model, therefore the modified phases will have less bias towards the molecular replacement model.

The modified phases produced from DM were used to calculate a 2Fo-Fc electron density map (Figure 5.4).

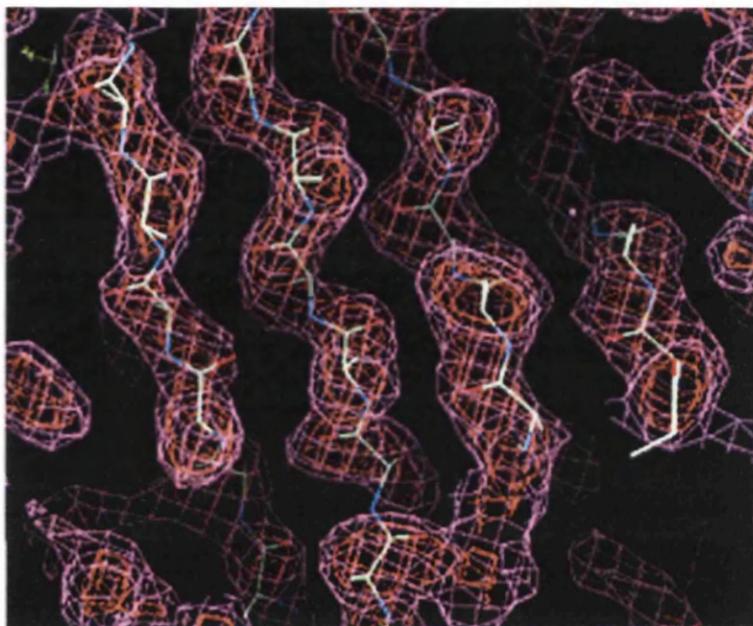


Figure 5.4: 2Fo-Fc map contoured at 1 (purple) and 2 (orange) sigma, calculated using modified phases from DM. MTDHQase poly-alanine model is also shown.

The maps generated after rigid body refinement and density modification were of good quality and indicated that the molecular replacement solution was correct.

5.5 Structure solution of HPDQase + FA1 by molecular replacement

Molecular replacement was carried out to phase the data collected upon HPDQase + FA1 described in section 4.6.2. The initial data processing suggested that the crystal belonged to a space group within the $-31m$ or $-3m1$ Laue groups,

space groups with 6 symmetry operators. Estimation of the content of the asymmetric unit was carried out by calculation of the Matthews coefficient (V_M) (Matthews, 1968) which is derived below where V_{cell} is the volume of the unit cell, MW is the molecular weight of the protein and Z is the number of symmetry operators in the unit cell.

$$V_M = \frac{V_{cell}}{MW \cdot Z}$$

Assuming a protein density of 1.35g/cm^3 , the V_M should be between 1.66 and 4, which corresponds to a solvent content of the crystal between 30% and 75% respectively

For 4 molecules per asymmetric unit, resulting in 24 molecules per unit cell, the $V_M = 4.1$ (69.77% solvent), and for 6 molecules per asymmetric unit $V_M = 2.33$ (54.66% solvent).

Molecular replacement was attempted with AMoRe and EPMR, using a monomer and trimer of HPDHDQase as the search model, however no solution was identified with either program.

To solve the molecular replacement the data was re-processed in the lower symmetry space group $P3_1$ which has only 3 equivalent positions within the unit cell. The calculated V_M in this space group for a complete HPDHDQase dodecamer was 2.8 corresponding to 56.4% solvent. Molecular replacement was carried out with AMoRe using the HPDHDQase trimer as the search model. 12 peaks of identical height were found in the rotation search corresponding to the 12 trimers in unit cell. No other peaks were found above 50% of the maximum peak height. The final MR solution had a correlation coefficient of 69.9% and an R-factor of 34.4%.

The crystal packing within the unit cell was analysed with Quanta (Figure 5.5) which showed no steric clashes. The MR solution formed a complete dodecamer within the asymmetric unit.

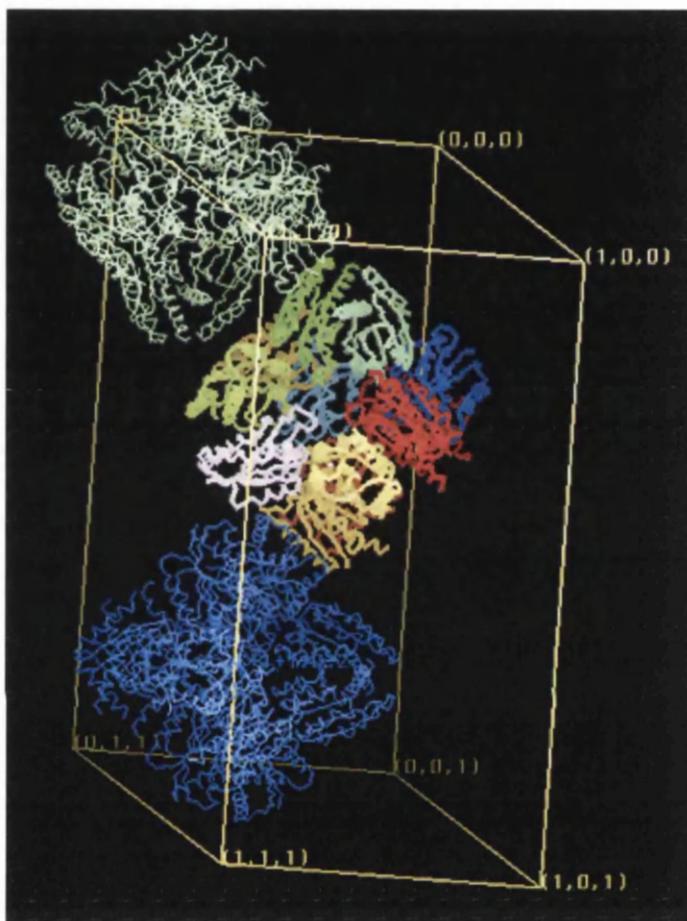


Figure 5.5 Crystal packing of HPDHQase + FA1 within the $P3_1$ unit cell.

5.6 Conclusion

In this chapter the basic principles of the molecular replacement technique have been outlined and shown how they were used in the structure solution of the type II DHQase from *H. pylori*.

Attempts to solve the molecular replacement using the program AMoRe were unsuccessful, despite the use of several type II DHQases as search models. A possible explanation for the failure of the program was the requirement of an

accurate solution of the rotation search. The high symmetry space group and close packing of the crystals would result in a complex Patterson function, hindering the rotation search.

A good molecular replacement solution was found using the program EPMR which performs a 6D search through space using genetic algorithms, with a poly-alanine model of type II DHQase from *M. tuberculosis* used as a search model. The packing of monomers within the unit cell was analysed and no obvious steric clashes were observed suggesting the MR solution was correct. SIGMAA weighted phases were calculated from the poly alanine MR solution which were modified with the program DM, and initial 2Fo-Fc maps calculated. The good quality of the maps suggested that the molecular replacement solution was correct.

Despite problems in accurately determining the space group of the HPDHQase + FA1 crystal a correct molecular replacement solution was found using AMoRe. Initial data processing had suggested that the space group was of higher symmetry than the correct solution, however this can be accounted for by the large amount of non crystallographic symmetry within the asymmetric unit.

Chapter 6 - Refinement and Model Building of HPDHQase + AH9095

Two high quality data sets have been collected upon crystals of type II DHQase from *H. pylori* in the presence of the inhibitor AH9095. A molecular replacement solution was found using the program EPMR using a poly-alanine model of the type II DHQase from *M. tuberculosis* as a search model. Refinement must now be carried out upon the model.

6.1 Refinement of Macromolecular Models

In the refinement of macromolecular models, the model coordinates are altered slightly and theoretical structure factors calculated for the new model. Currently, many methods are used to perform model refinement, namely least squares, maximum likelihood, energy minimisation and molecular dynamics. The program Refmac (Murshodov *et al.*, 1997) which applies maximum likelihood methods was used for refinement of the HPDHQase + AH9095 model. The quality of the model is assessed by comparing new theoretical data, calculated from the new model, with the observed data. This relationship is monitored with an R-factor (Equation 6.1)

$$R = \frac{\sum_{hkl} \left| |F_{obs}| - k|F_{calc}| \right|}{\sum_{hkl} |F_{obs}|}$$

Equation 6.1: The crystallographic R-factor where $|F_{obs}|$ are the observed structure factor amplitudes, $|F_{calc}|$ is the calculated structure factor amplitude and k is a scale factor

Unfortunately the crystallographic R-factor is highly susceptible to over-fitting where the R-factor can be reduced by fitting more parameters e.g. water molecules to structures. Therefore another indicator is required to get an unbiased assessment of the fit of the model to the data. To this end a proportion of the structure factors (typically 10%) are omitted from refinement. These structure factors are referred to as the R_{free} set, while the remainder used in refinement are the R_{work} , are used to calculate the R_{free} factor (equation 6.2) (Brunger, 1992). As the R_{free} set is not refined, it provides an independent assessment of the agreement between the model and the observed data. During the refinement process careful analysis of the relationship between the R_{work} and R_{free} indicates the quality of the model. Typically the R_{free} value is 3-5 % higher than the R_{work} .

$$R_{\text{free}} = \frac{\sum_{hkl \subset T} \left| |F_{\text{obs}}| - k |F_{\text{calc}}| \right|}{\sum_{hkl \subset T} |F_{\text{obs}}|}$$

Equation 6.2: The crystallographic R-factor where $hkl \subset T$ are the R_{free} structure factors, $|F_{\text{obs}}|$ are the observed structure factor amplitudes, $|F_{\text{calc}}|$ is the calculated structure factor amplitude and k is a scale factor

After each refinement step, new values for F_{calc} (calculated structure factor amplitudes for the model) and PHI_{calc} (calculated phases) are generated. Weighted values for F and PHI are also calculated which are used to generate electron density maps using the program FFT.

6.2 Electron Density Maps

The aim of an X-ray diffraction experiment is ultimately the generation of an electron density map. The equation for the calculation of an electron density map, $\rho(xyz)$, is shown below (Equation 6.3) where V is the volume of the unit cell, $|F_{hkl}|$ is the structure factor amplitude and α_{hkl} the associated phase.

$$\rho(xyz) = \frac{1}{V} \sum_{hkl} |F(hkl)| \cos[2\pi(hx + ky + lz) - \alpha(hkl)]$$

Equation 6.3: Equation for the calculation of an electron density map

This map (an F_{obs} map) is not entirely useful at this stage as the calculated phases result in strong electron density for parts of the structure already phased while the electron density for the missing parts of the structure is much weaker resulting in a map that is biased towards the model. A much more useful map to use at this stage is a $2F_o-F_c$ map. In the calculation of this map the component derived from the observed data (F_o) is amplified to remove some of the bias. In some cases a $3F_o-2F_c$ map may be used to enhance this effect further. The equation for the calculation of a $2F_o-F_c$ map is shown below (Equation 6.4).

$$\rho(xyz) = \frac{1}{V} \sum_{hkl} (2|F_{obs}| - |F_{calc}|) \exp[-2\pi i(hx + ky + lz) + i\alpha_{calc}]$$

Equation 6.4: Equation for the calculation of a $2F_o-F_c$ electron density map

The $2F_o-F_c$ map is still biased towards the current model, therefore a F_o-F_c , or difference, map is generally used in tandem with the $2F_o-F_c$ map is utilised in the later stages of refinement. The difference map, $\Delta\rho(xyz)$, uses the difference between observed and calculated structure factor amplitudes, $\Delta|F(hkl)|$, and calculated phases α (Equation 6.5)

$$\Delta\rho(xyz) = \frac{1}{V} \sum_{hkl} \Delta|F(hkl)|_{iso} \cos[2\pi(hx + ky + lz) - \alpha_{\phi}(hkl)]$$

Equation 6.5: Equation for calculation of a Fo-Fc electron density map.

Sections of the difference map of positive magnitude indicate areas where there is scattering matter (i.e. atoms) not present in the model, that are in the crystal. Conversely, areas of negative magnitude indicate areas of the model that are not present in the crystal. Analysis of the model, how it fits to the 2Fo-Fc map and how it relates to features in the Fo-Fc map allows us to adjust the model to fit the density better. In early stages of refinement, when the fit of the model to the data is poor the difference map is very noisy and yields little information, however in the later stages where the model fits well to the data strong peaks and troughs are observed in the map, aiding model building.

6.3 Manual Alteration of the Model

In most cases, manual alteration of the model is required after inspection of the electron density maps. Most computer packages cannot perform the kind of visual analysis that the human crystallographer can, therefore alteration of side chain conformations, modelling of ligands etc. is carried out manually. The module X-AUTOFIT within the Quanta package (Accelrys) was used for manual inspection of electron density maps and manual alteration of the model. Quanta not only allows visualisation of the model and electron density maps, but also can perform local refinements upon areas of the model, and can also be used to dock ligands into areas of suitable electron density.

6.4 Refinement and model building of HPDHQase + AH9095.

Initial refinement of the HPDHQase + AH9095 structure was carried out as follows. A full atomic model of an MTDHQase monomer was superimposed upon the poly-alanine molecular replacement solution. Using the program Quanta, residues that were conserved between MTDHQase and HPDHQase were retained in the model, and non-glycine residues that were not conserved were cut back to alanines. Refmac 5 from the CCP4 suite of programs was used to refine the model against the 2.5Å data collected in-house. 3 rounds with 6 cycles of refinement were applied to give a final model with $R_{\text{work}} = 33.9\%$ and $R_{\text{free}} = 40.1\%$.

2Fo-Fc and Fo-Fc electron density maps were calculated with the program FFT and analysed using Quanta. From the 2Fo-Fc map 90% of the non-glycine residues had their side chains modelled successfully. A second round of refinement with Refmac5 was carried applying 4 rounds of 8 cycles to give a model with final statistics of $R_{\text{work}} = 29.1\%$ and $R_{\text{free}} = 36.7\%$. Inspection of the resulting 2Fo-Fc map indicated the presence of significant density in the active site region to indicate the presence of a bound ligand (Figure 6.1)

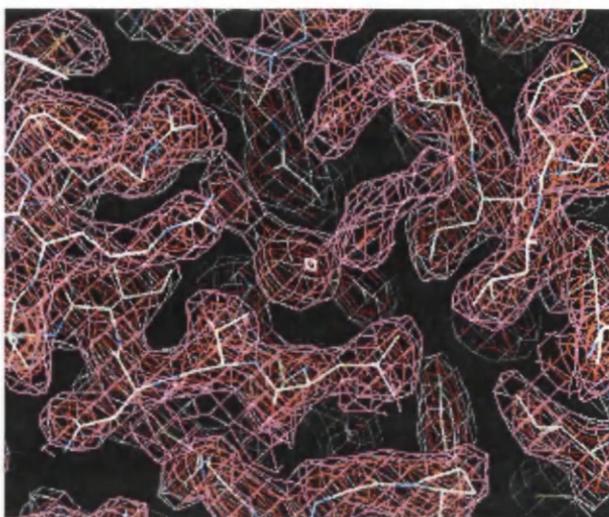


Figure 6.1: 2Fo-Fc electron density map contoured at 1 (purple) and 2 (orange) sigma of HPDHQase + AH9095 centred upon active site region. HPDHQase model is shown in stick representation.

At this stage of the refinement, the model started at residue Met 3 and was complete until residue 17, the start of the flexible lid domain. This region of structure had not been visible in previous MTDHQase structures, but was modelled successfully in SCDHQase and BSDHQase. The model continued from residue 26 to residue 66 where again no density was visible between residues 66 and 70. This section of the structure is a 6 residue loop connecting helix 2 to strand 3, which contains more residues in HPDHQase than in other type II DHQases. The model was extended to residue 144, the C-terminus of the MTDHQase model, however density was visible for an extended terminal α helix.

6.5 Refinement and model building of HPDHQase + AH9095 with high resolution data

At this point in the refinement, the high resolution data became available as described in section 4.5. After processing of the higher resolution data it was

apparent that the two crystals upon which the in-house and synchrotron experiments were carried out were isomorphous with each other. The model was now refined against the high resolution data that extended to 1.55Å resolution. To allow direct comparison of the refinement against the low resolution and high resolution data, the R_{free} set from the low resolution data set was copied to the high resolution data set and extended to 1.55Å.

A further 3 rounds of refinement and model building were carried out with manual adjustment of amino acid side chains and the addition of 25 water molecules using the X-SOLVATE module in Quanta. By this stage, the model statistics were $R_{\text{work}} = 23.7\%$ and $R_{\text{free}} = 26.2\%$. The flexible lid domain between residues 18 and 26 was still not visible, but the C-terminal helix was extended to residue 149.

A model for the ligand AH9095 was created and submitted to energy minimisation using the program Insight II (Accelrys) and successfully docked into the corresponding Fo-Fc electron density in the active site of the model using the program X-LIGAND (Accelrys) within Quanta. There was also Fo-Fc density within the active site corresponding to the presence of a glycerol molecule. A coordinate file for a glycerol molecule was obtained from a complex SCDHQase deposited in the PDB (PDB ID 1GU1) and the glycerol molecule modelled into the active site in a similar manner.

Further rounds of refinement and manual model correction were carried out to give final model statistics of $R_{\text{work}} = 20.0\%$ and $R_{\text{free}} = 22.7\%$. Density was visible at the N-terminus corresponding to the linker region and the first residues of the 6xHis tag, allowing the modelling of 8 extra residues in this region. Attempts to model the lid domain were unsuccessful, however the C-terminus was extended to residue 157.

6.5.1 Automated addition of solvent molecules

The addition of ordered solvent molecules to a protein structure improves the accuracy of the model as these molecules scatter to the same extent as the protein itself. Identifying the location of the ordered water molecules and placing atoms at those positions can be tedious and time consuming, therefore an automated procedure can be employed.

The program ARP (Lamzin and Wilson, 1993) was designed to refine high resolution macromolecular structures by modelling the protein with a series of dummy atoms, assessing the validity of the atoms and iteratively removing 'bad' atoms and placing new ones. This procedure is ideal for the identification of ordered solvent molecules in a protein structure. 2 cycles of Refmac in conjunction with ARP were run and 196 water molecules added to the model. The validity of these water molecules was assessed using Quanta, and 24 of these water molecules were discarded as they did not fit the observed electron density.

6.5.2 Restrained Anisotropic B-factor Refinement

When X-ray diffraction data is collected to greater than 1.2\AA resolution, we can stop treating atoms as perfect spheres, and begin to model the thermal motion of the atoms more accurately. The thermal motion of an atom is indicated by its B-factor. When dealing with very low resolution structures, the B-factors for groups of atom are kept the same. As we refine using higher resolution data we give each individual atom an individual B-factor, although adjacent atoms must have similar B-factors. This is referred to as isotropic B-factor refinement. At high resolution we can model the thermal motion of the atom by assigning not only an overall B-factor but

we can describe the directional motion of the atom with a series of 6 other values.

This process is anisotropic B-factor refinement.

This process increases the number of parameters in relation to the number of observations hence high resolution data is required. However, within Refmac, restrained anisotropic B-factor refinement can be implemented at resolutions less than 1.2Å. In this process, more restraints are placed upon the model, effectively increasing the number of observations, allowing efficient refinement (Murshudov *et al.*, 1999).

In the final stages of the refinement of HPDHQase + AH9095, anisotropic B-factor refinement was carried out as implemented in Refmac. The final model statistics were $R_{\text{work}} = 15.6\%$, $R_{\text{free}} = 19.6\%$.

6.6 Quality of the final HPDHQase + AH9095 model

Electron density maps calculated from then final HPDHQase + AH9095 model were of excellent quality. The density around the AH9095 ligand (Figure 6.2) is typical of the maps. Figure 6.3 shows a region of one of the central strands of the structure. Atomic positions can be clearly identified.

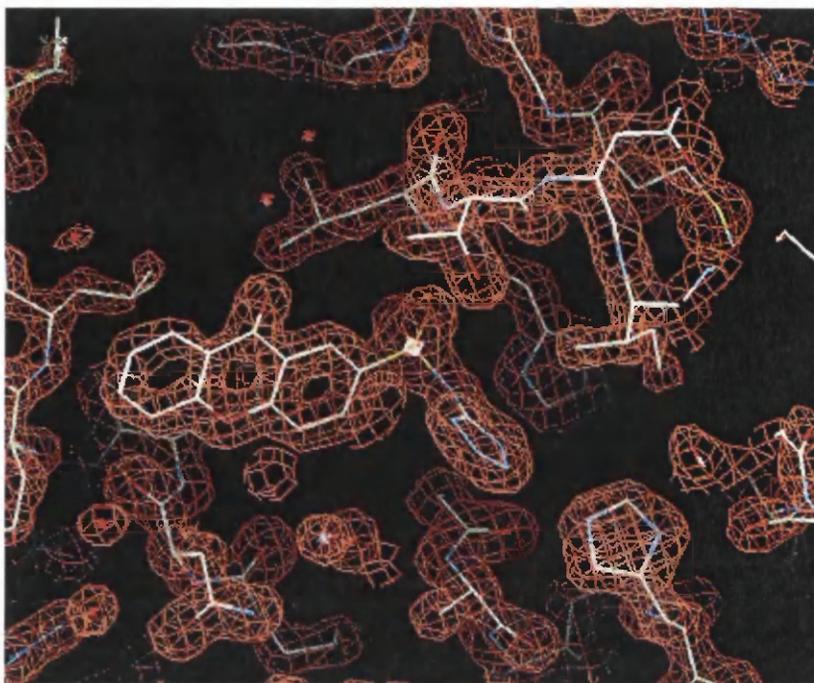


Figure 6.2: 2Fo-Fc electron density map contoured at 4 sigma around AH9095

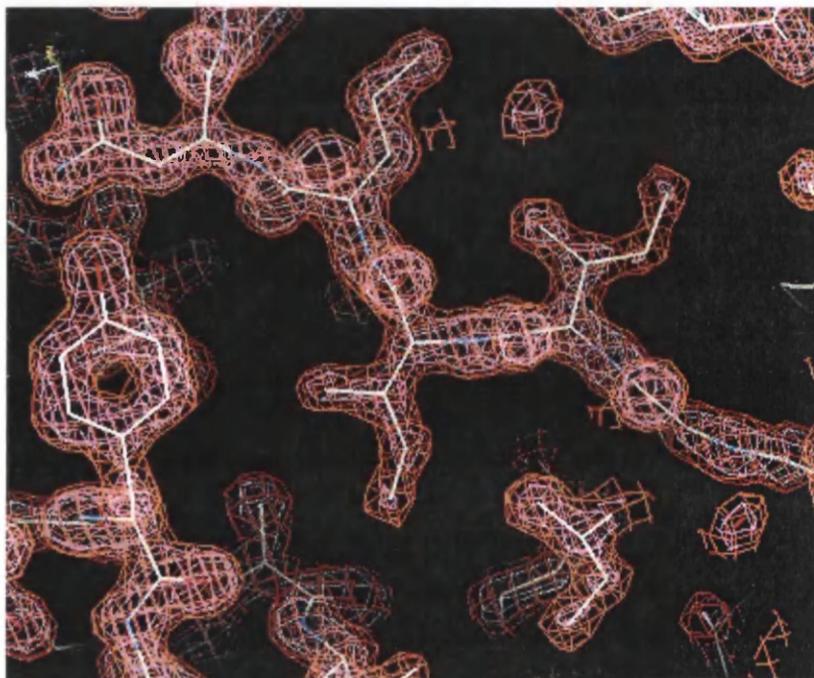


Figure 6.3: 2Fo-Fc electron density map contoured at 4 sigma around strand 3 of HPDHQase

Typically in the refinement of a small molecule structure determined by X-ray crystallography is only considered 'finished' when the model fits the observed data, with an R-factor of around 5%. In proteins the situation is more complex due to the poor resolution of the data and the inherent level of disorder within the structure. Analysis of the R-factors and model geometry provide a guideline towards how the model fits the data, but how realistic the arrangement of atoms is. There are no rules governing what R-factors should be expected for a model determined from data of a specific resolution. Validation in this respect is carried out by comparison of our model the R-factors of existing structures in the PDB of a similar resolution (Kleywegt and Jones, 2002)

The programs PROCHECK (Laskowski *et al.*, 1993) and WHAT IF (Vriend, 1990) perform automated analysis of a model and identify areas of the models that may require further attention with respect deviation from expected values. PROCHECK was used to analyse the final HPDHQase + AH9095 model, and plot the positions of each residue on a Ramachandran plot. 117 (90%) of the residues were in the core allowed regions whereas 9 (6.9%) residues were in the additionally allowed regions and 1 (0.8%) residue in the generously allowed region. 3 residues (2.3%), His 2, Ser 70 and Tyr 72, were found to lie in the 'disallowed' region of the Ramachandran plot. Upon inspection of the electron density in these areas of the structure it was apparent that the current model fitted the electron density well. His 2 may be in a twisted conformation as there are contacts between the N-terminal linker region, an artefact of the 6xHis tag, with the C-terminal residues. Residues 70 and 72 are located on a loop between helix 2 and strand 3. Removal of this section of the model followed by a round refinement resulted in electron density maps suggesting

that the 'disallowed' conformation of the loop was indeed correct. The Ramachandran plot is shown below (Figure 6.4)

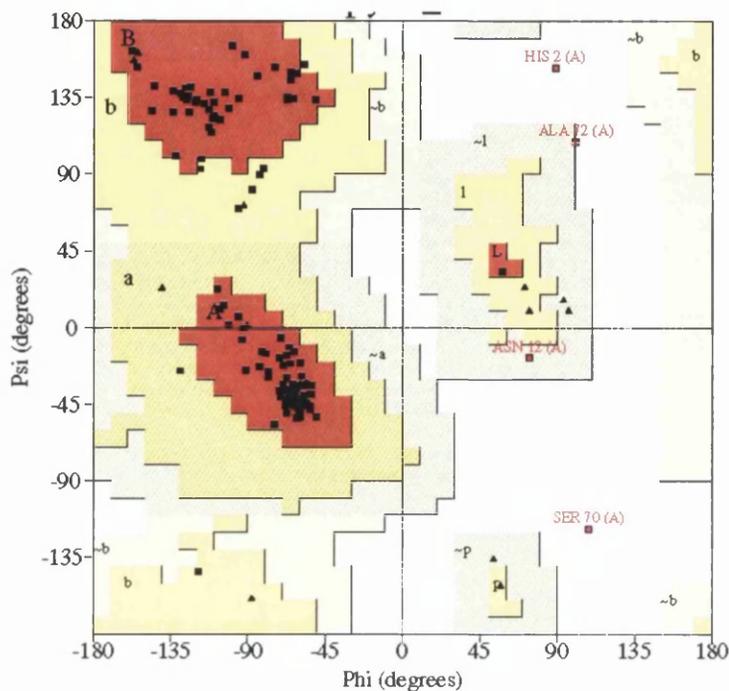


Figure 6.4: Ramachandran plot generated by PROCHECK showing the main chain torsion angles of HPDHase + AH9095. Core allowed regions are shaded red, allowed regions yellow, generously allowed regions pale yellow and disallowed regions white.

The overall geometry of the final model is assessed by analysis of the RMS deviation of features such as bond lengths, bond angles and the deviation from planarity of planar groups. The diffraction precision index (DPI) as defined by Cruickshank (Cruickshank, 1999) provides an estimation of the model accuracy by taking into account the number of data, number of parameters, crystallographic R-factor and the resolution of the observed data. A Summary of these indicators is shown in table 6.1.

Analysis of the B (temperature) factors of atoms in the model also give an indication of the quality of the model, although again this measure is not only dependant upon the resolution of the observed data but can be affected by the

refinement program used. A summary of the B-factors for HPDHQase + AH9095 is shown in table 6.2.

Indicator	Value	Target
RMS bond length (Å)	0.014	0.022
RMS bond angle (Å)	1.597	1.985
RMS planarity (Å)	0.006	0.020
Cruickshank DPI	0.0929	-

Table 6.1: Geometric indicators of the quality of the final HPDHQase + AH9095 model

Chain	Description	No. of atoms	Average B- factor
A	HPDHQase monomer	1166	23.2
A	AH9095	24	15.6
A	Heteroatoms excl. AH9095	17	29.4
W	Water molecules	177	45.6
ALL	Total Model	1384	26.0

Table 6.2: B-factor statistics for final HPDHQase + AH9095 model.

6.7 Conclusion

A summary of the refinement procedures and associated model statistics are shown in table 6.1

Refinement Cycle	Description	R _{work} (%)	R _{free} (%)
0	MTDHQase poly alanine model	49.1	52.2
1-2	Addition of HPDHQase side chains	29.1	36.7
3-5	High resolution data added Water molecules added	23.7	26.2
6-8	Ligands modelled	20.0	22.7
9-10	Automated addition of water molecules (ARP)	17.4	20.6
11-14	Anisotropic B-factor refinement	15.6	19.6

Table 6.1: Refinement procedure for HPDHQase + AH9095.

The initial MTDHQase poly-alanine model, used to find a molecular replacement solution, was modified to include side chains conserved between MTDHQase and HPDHQase and refined using Refmac5. HPDHQase side chains that were not modelled previously were added to fit the electron density maps generated after each round of refinement. The high resolution data was used for refinement at this stage and water molecules were added with X-SOLVE and the model refined for 3 more cycles. There was electron density corresponding to the ligand AH9095 in the active site region into which a model was successfully docked using X-LIGAND. A glycerol molecule was also modelled in this way. After 3

rounds of refinement and model correction, coordinated solvent molecules were automatically added using ARP/wARP. Finally, anisotropic B-factor refinement was carried out to produce a model with final crystallographic statistics of $R = 15.6\%$ and $R\text{-free} = 19.6\%$.

At this stage, no more cycles of refinement and manual model adjustment resulted in a decrease of the R-factors therefore the refinement process was stopped. The final model was of good quality as evaluated by PROCHECK, and can now be used for analysis and modelling.

Chapter 7 – Structural analysis of the type II Dehydroquinase from *Helicobacter pylori*

7.1 Crystal Structure of type II DHQase from *Helicobacter pylori*

The final refined model of the type II DHQase from *H. pylori* consists of a single monomer of 162 residues forming the flavodoxin type α/β fold observed in other type II DHQases. A ribbon cartoon of HPDQase is shown in figure 7.1

The catalytic lid domain is disordered, between residue 18 and residue 26, as are the final 14 residues at the C-terminus. The first 10 residues (numbered -7 to 2) at the N-terminus of the model are part of the 6x His tag and linker region used to purify the protein. The C-terminal helix is much longer than that observed in any other type II DHQase, extending 12 residues (3 1/2 turns) beyond the main body of the monomer. A single molecule of the inhibitor AH9095 is bound in the active site.

The complete dodecamer is formed by crystal symmetry and is shown in Figure 7.2

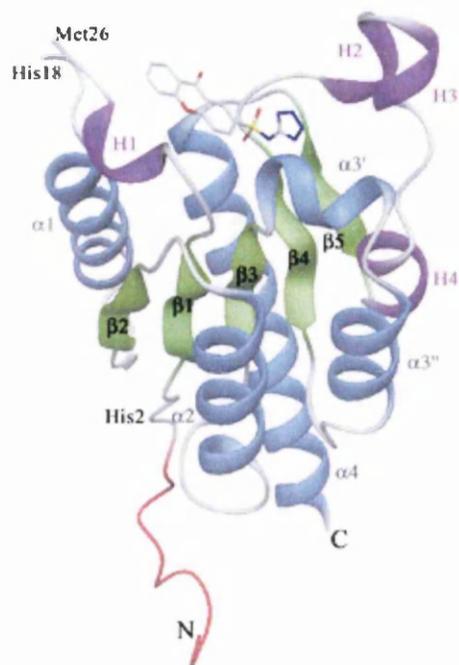
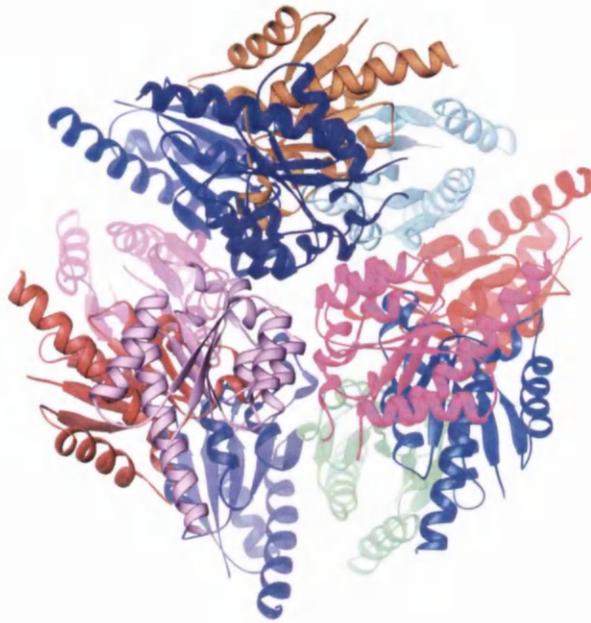


Figure 7.1: Ribbon representation of HPDHQase monomer. β -strands are coloured green, α -helices blue and 3_{10} helices purple. The N-terminal extension corresponding to the 6xHis tag is coloured red and the ligand AH9095 is shown in stick representation

A



B

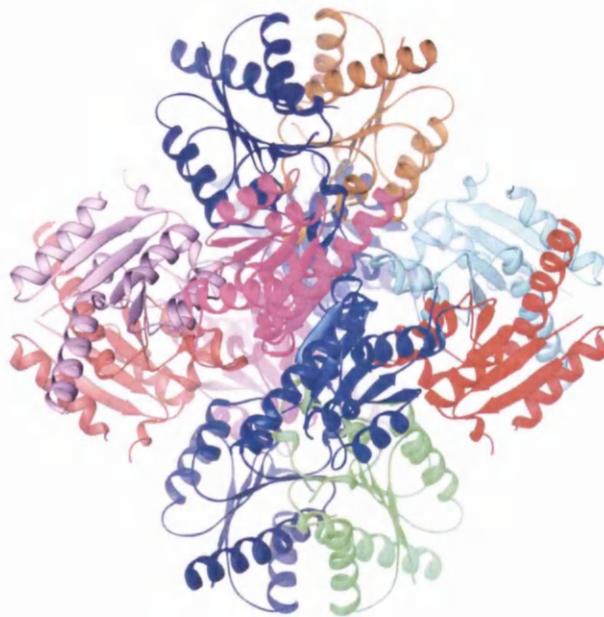


Figure 7.2 HPDHQase dodecamer viewed A- down the crystallographic 3-fold axis, B- down the crystallographic 2-fold axis. Each monomer is coloured separately. Figure created with RIBBONS (Carson, 1991).

7.2.1 A comparison of the monomer folds of the type II DHQases from *Helicobacter pylori* and *Mycobacterium tuberculosis*

The HPDHQase monomer superimposes upon the apo MTDHQase monomer (Gourley *et al.*, 1999 - PDB I.D. 2DHQ) with an RMSD of 1.05Å between the positions of the main chain atoms and an RMSD of 1.01Å between the positions of the C α atoms only (Section 7.2.2). All superimpositions were calculated using LSQKB from the CCP4 suite of programs.

The central β -strands of the monomers overlay accurately, however the presence of additional residues within the loop regions of HPDHQase affects the orientation of the loops and helices (Figure 7.3A). The additional residues Gly43 and Asn44 located in the loop region between helix α 1 and strand β 2 not only extend the loop but also re-orientate helix α 1, moving the N-terminal end of the helix away from the core of the monomer. The result of this movement is a 10° tilt of helix α 1 in relation to the position of helix α 1 in MTDHQase.

Residue Glu71, located in the loop between helix α 2 and strand β 3 is present in HPDHQase but not MTDHQase. This addition does not alter the relative position of secondary structure features but does extend the loop, making this region of the structure more flexible. As a result residues Ser70 and Tyr72 adopt unfavourable main chain conformations. These two sequence related differences between HPDHQase and MTDHQase account for the major structural differences between the models.

Helix H4 in HPDHQase lies 0.4Å closer to the core of the monomer than the corresponding helix in MTDHQase, however this region may move as a result of the presence and nature of a bound ligand.

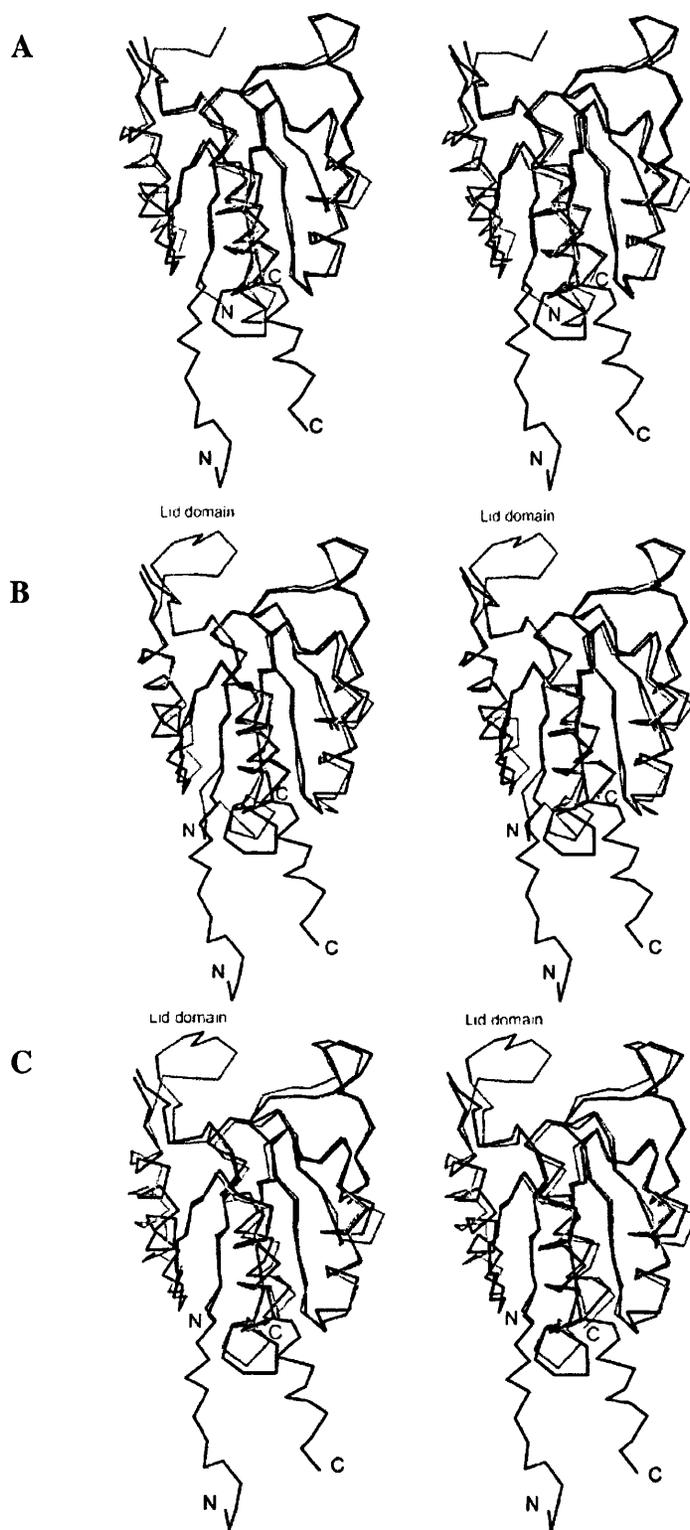


Figure 7.3: Stereo pair diagrams of C α traces of HPDQase monomers superimposed upon A – MTDQase (2DHQ), B-SCDQase (1GU0) and C- BSDQase (1GQO). HPDQase is shown in black, other DHQases grey. Figure created with SETOR (Evans, 1993)

The C-terminal helix ($\alpha 4$) is much longer in HPDHQase than in MTDHQase, extending 3 turns beyond the C-terminus of MTDHQase.

7.2.2 Similarity of HPDHQase monomer to other type II DHQase monomers

The HPDHQase monomer was superimposed upon MTDHQase, SCDHQase and BSDHQase monomers using LSQKAB from the CCP4 suite of programs. RMS deviations between atomic positions of each structure were calculated (Table 7.1)

DHQase	MTDHQase	SCDHQase	BSDHQase	HPDHQase	Main Chain RMSD (Å)
MTDHQase		0.94	1.05	1.05	
SCDHQase	39.6		0.97	0.95	
BSDHQase	42.7	45.9		0.96	
HPDHQase	33.5	34.1	37.1		
% Sequence Identity					

Table 7.1: RMS deviations between atomic positions of HPDHQase with other type II DHQase structures

Initial phasing of the HPDHQase structure was achieved by molecular replacement using a $C\alpha$ model of MTDHQase as a search model. Despite sharing the lowest sequence identity and the highest main chain RMSD, MTDHQase and HPDHQase have the most similar $C\alpha$ positions of all the type II DHQases.

7.3 Type II dehydroquinase from *Helicobacter pylori* – Quaternary Structure

The HPDHQase + AH9095 crystals contained a single monomer in the asymmetric unit therefore the dodecameric quaternary structure is defined by crystallographic symmetry. The dodecamer has tetrahedral 23 symmetry, an arrangement seen rarely in protein structure (Gourley *et al.*, 1999). Each monomer interacts with neighbouring subunits along the 3-fold and 2-fold crystallographic axes. Illustrations of the trimer and dimer interfaces are shown in Figures 7.4 and 7.5

7.3.1 Trimer interface

The minimal catalytic unit of type II DHQase is the disc shaped trimer (Price, *et al.* 1999), concave on one face and convex on the other, with the convex faces packing into the core of the dodecamer. With respect to a single monomer we can consider the trimer interface consisting of two surfaces, surface 1 packs against monomer B and surface 2 interacting with monomer C (Figures 4a and 5a).

Surface 1 measures approximately 605\AA^2 and is formed by residues from helices H1, $\alpha 3'$, H2 and H3. These secondary structure elements are located near the C-terminal end of the central β sheet. Surface 2 buries 577\AA^2 of surface and is formed by residues that are located on helices $\alpha 2$ and $\alpha 3''$. Surface 1 of monomer A interacts with surface 2 of monomer B as described for the structure of SCDHQase (Roszak *et al.*, 2002). The interaction between the two surfaces consists of 7 H-bonds and 1 salt bridge. A list of the hydrogen bonds between monomers A and B is shown in table 7.2. The residues listed from monomer A are from surface 1, and the residues listed from monomer B are from surface 2.

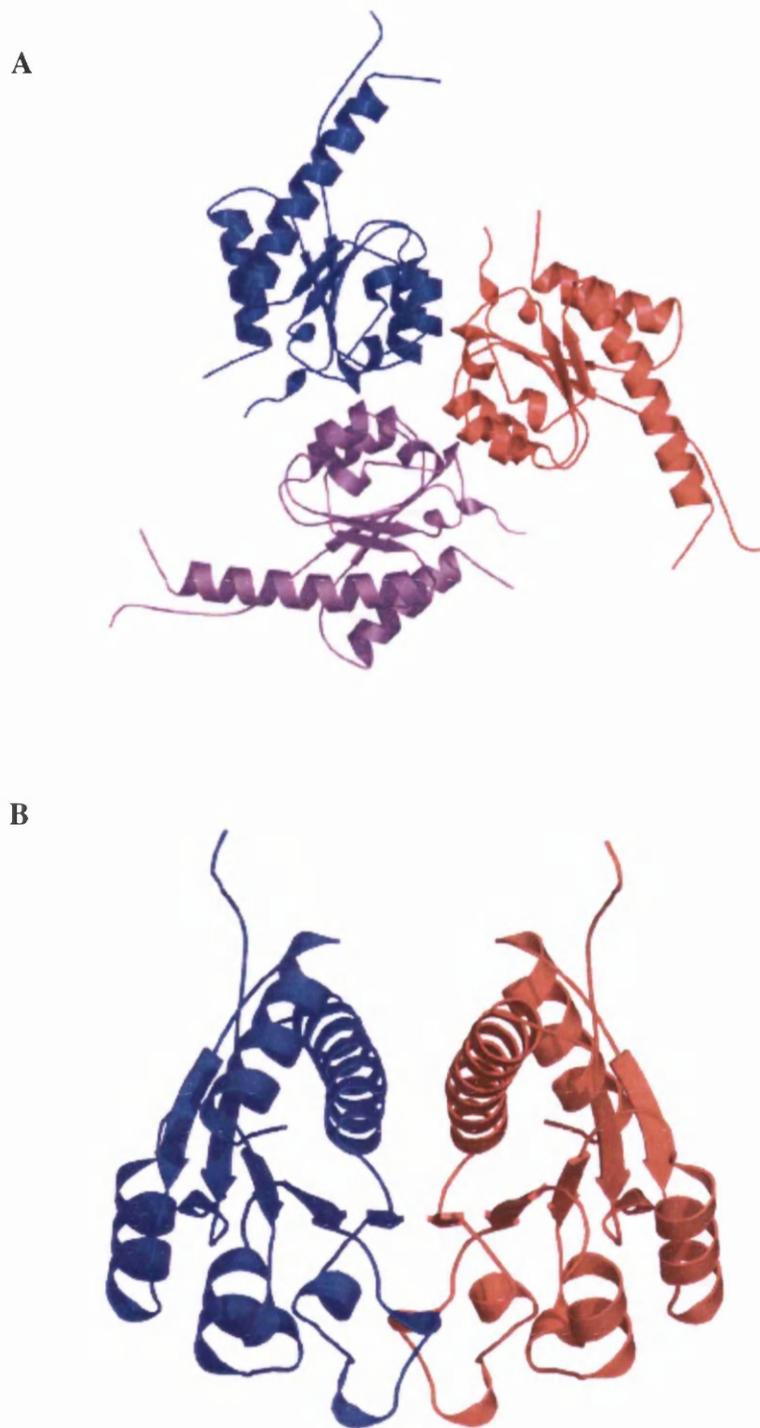


Figure 7.4: Trimer and dimer interfaces of HPDHQase: A) the arrangements of monomers around the trimer interface. B) the dimer interface, formed by strand $\beta 5$ from each participating subunit lying antiparallel to each other, resulting in the extension of the central β -sheet between subunits. Figure created with PyMOL (DeLano, 2002)

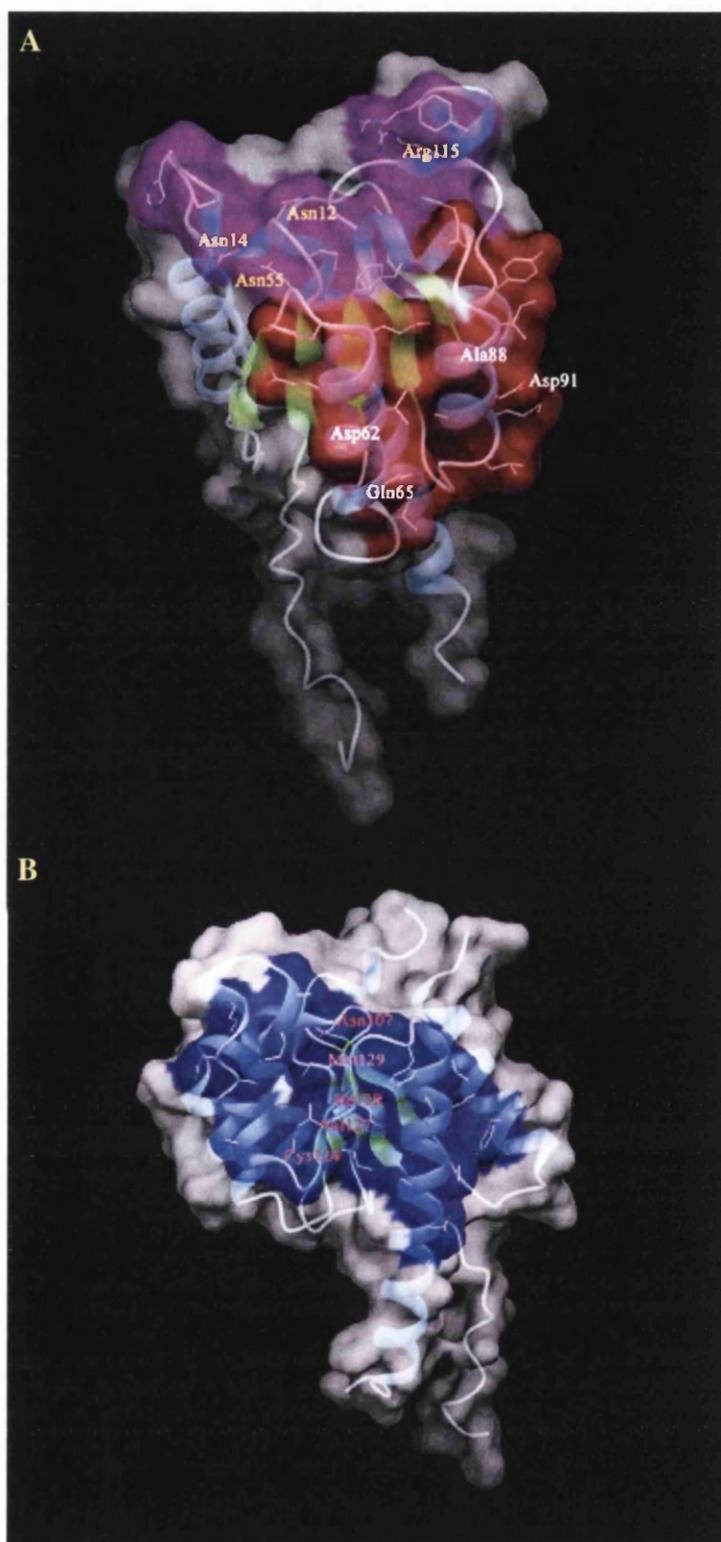


Figure 7.5: HPDHQase trimer and dimer interfaces: Molecule A shows location of trimer surfaces 1 (magenta) and 2 (red). Molecule B shows the dimer surface coloured blue. Residues that form the surfaces are shown, residues that form H-bond interactions are labelled. Figure created using DINO (2002)

Two interactions at the trimer interface are invariant throughout all type II DHQases, the interaction between ND2 of Asn 12 with the backbone O atom of Ala 88, and the salt bridge between Asp91 and Arg115. Asn 55 and Asp 62 are strongly conserved throughout the type II DHQases, though residue 65 is less conserved. Residue 65 is predominately a histidine residue which would only facilitate formation of a single H-bond with Asn14. The presence of a glutamine residue allows the formation of two H-bonds with Asn14, increasing the strength of the trimer interaction.

Monomer A	Location	Monomer B	Location	Bond Length (Å)
Asn12 ND2	β 1	Ala88 O	α 3''	3.11
Asn14 OD1	H1	Gln65 NE2	α 2	2.99
Asn14 ND2	H1	Gln65 OE1	α 2	2.94
Asn55 ND2	β 2	Asp62 OD2	α 2	2.84
Asn55 ND2	β 2	Gln65 OE1	α 2	3.05
Arg115 NH1	H3	Asp91 OD2	α 3''	2.90
Arg115 NH2	H3	Asp91 OD1	α 3''	2.89

Table 7.2 H-bond interactions at the AB trimer interface of HPDHQase

7.3.2 The Dimer interface

The dimer interface is formed between monomers from adjacent trimeric units. Strand $\beta 5$ of each subunit lies anti-parallel with strand $\beta 5$ of the neighbouring subunit, extending the central β -sheet between both monomers (Figure 7.5). The nature of this surface is predominately hydrophobic, with H-bond interactions formed between main chain atoms from residues Val127 and Met109. Additional H-bonds are formed between highly conserved residue Asn107, located adjacent to the active site, with main chain atoms of Gly121 and Cys124. Each monomer buries 910 \AA^2 at this surface. Illustrations of the dimer surface are shown in Figures 7.4b and 7.5b, and a summary of the H-bonds present at the dimer interface is shown in table 7.3.

Monomer A	Monomer B	Bond Length (\AA)
Asn107 ND2	Gly121 O	3.03
Asn107 ND2	Cys124 O	2.93
Gly121 O	Asn107 ND2	3.03
Cys124 O	Asn107 ND2	2.93
Val127 N	Met129 O	2.92
Val127 O	Met129 N	2.98
Met129 N	Val127 O	2.98
Met129 O	Val127 N	2.92

Table 7.3 H-bond interactions at the dimer interface of HPDHQase

7.3.3 Contribution from 6x His tag

No attempt had been made to cleave the 6x His tag from the N-terminus of the HPDHQase protein before crystallisation as removal of the MBP tag from the previous pMAL constructs had proved to be detrimental to the protein. The extra residues consist of an N terminal Met followed by a spacer, then the 6x His residues followed by a spacer, then the thrombin cleavage sequence. A further spacer lies between the cleavage site and the native N-terminal Met residue. Thus the following 21 residues are added at the N-terminus; MGSS**HHHHHHH**SGLVPRGSH where the 6xHis residues are shown in bold and the thrombin cleavage site is underlined. In the HPDHQase/AH9095 crystal structure density was visible for 2 histidine residues and the linker region of the 6x His tag. Analysis of the structure revealed that the His tag and linker region form crystal contacts with subunits from other dodecamers within the unit cell.

The His tag linker region lies across the dimer interface of neighbouring subunits (Figure 7.6) forming 3 H-bonds, 2 to one subunit and one to the other. With one subunit interactions are formed between the main chain carbonyl oxygen atoms of Gly0 and Leu-4 with the side chain hydroxyl of Thr38 and the main chain nitrogen of Leu134 respectively, whereas atom NH1 of Arg-1 forms an H-bond with OE1 of Glu149 from the other subunit.

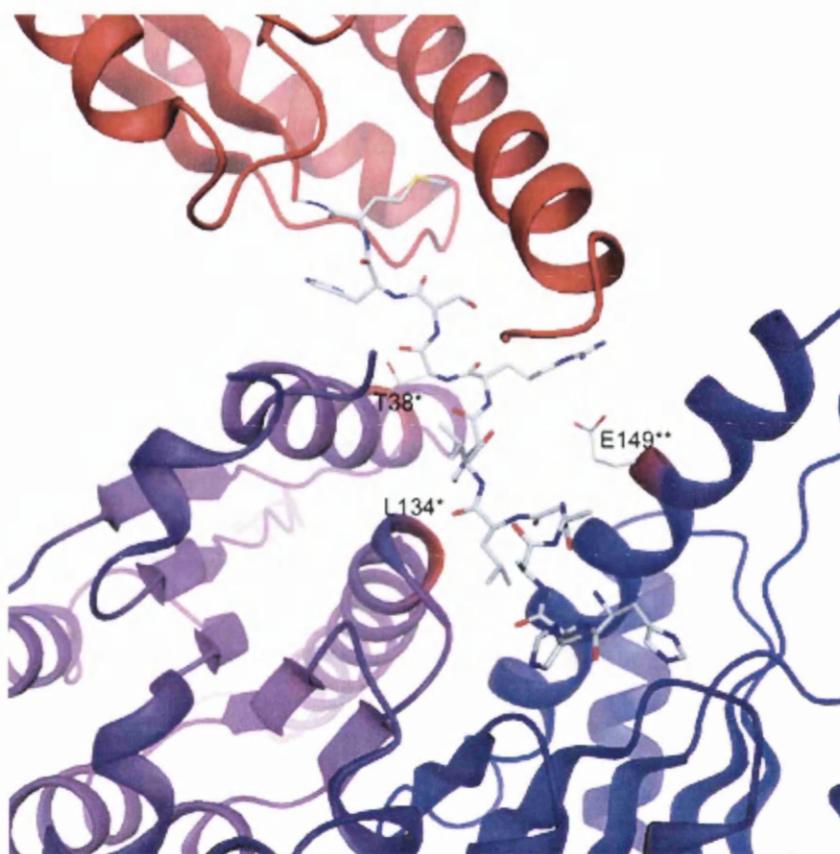


Figure 7.6: Interactions formed by the 6xHis tag and linker. N-terminal his tag (red monomer) lies across the dimer interface of neighbouring subunits (blue and purple); His tag is shown in stick representation, residues from the dimer that interact with the his tag are coloured red.

Figure composed using DINO (2002)

7.4 Ligand Binding of AH9095

The ligand AH9095 is much larger than any ligand previously crystallised with a type II DHQase, therefore the binding mode was of great interest. As the ligand is large we shall analyse the protein ligand interactions splitting the ligand into 3 segments, the tetrazole moiety, the sulphonamide moiety and the poly-cyclic xanthene moiety (Figure 7.6). Figure 7.7 shows a detailed diagram of the HPDHQase active site with AH9095 bound.

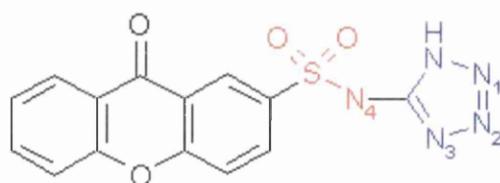


Figure 7.6: Ligand AH9095; tetrazole moiety (blue), sulphonamide (red) and xanthene (black). Numbering of Nitrogen atoms are shown in subscript

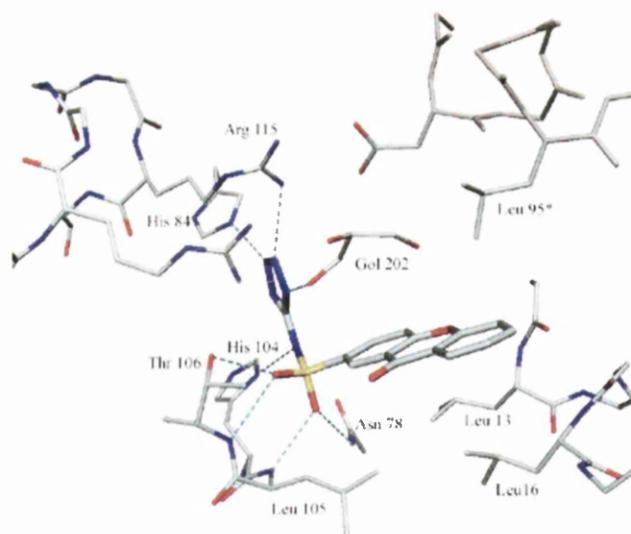


Figure 7.7: Diagram illustrating AH9095 bound in the active site region of HPDHQase. Key residues are labelled and H-bonds shown in cyan

The tetrazole moiety occupies the area of the active site where the main body of the substrate dehydroquinone has been shown to bind in SCDHQase (Roszak *et al.*, 2002). The tetrazole ring is smaller than dehydroquinone and lies perpendicular to the plane of dehydroquinone. H-bonds are formed between the conserved active site residues Arg115 and His84 to atoms N1 and N2 of AH9095 respectively. Both these residues interact with the C5 hydroxyl of dehydroquinone, a group that has been shown to be critical for activity (Harris, 1996b). An additional H-bond is formed from atom N to a bound glycerol molecule from the cryo-protectant solution.

The sulphonamide group of AH9095 forms 5 H-bonds to HPDHQase. Atom O1 forms H-bonds with Asn78 ND2 and the backbone N atom of Leu105, with atom O2 forming H-bonds with atoms OG1 and N of Thr106. Another H-bond is formed between atom N4 and atom ND1 of His104. These interactions are analogous to the those observed between polyanions with other type II DHQases (Evans *et al.*, 2002) and the interactions between the C1 hydroxyl and carboxylate of the substrate with type II DHQases (Roszak *et al.*, 2002)

The xanthene moiety lies in the entrance to the active site, packing on one face against a flat hydrophobic surface formed by the side chains of Leu13, Leu16 and Leu105. The other face is open to solvent at the trimer interface, packing against Leu95 of the neighbouring subunit. The carbonyl group on ring 2 of the xanthene moiety does not interact with any protein residues. A glycerol molecule from the cryo-protectant packs against the solvent accessible face of AH9095, occupying the rest of the active site.

A list of all the H-bond interactions between AH9095 and HPDHQase is shown in table 7.4.

Protein	Ligand	Section of Ligand	Bond Length (Å)	Bond Angle (°)
Asn78 ND2	O1	Sulphonamide	2.90	168.1
Leu105 N	O1	Sulphonamide	2.81	156.4
Thr106 N	O2	Sulphonamide	3.03	161.0
Thr106 OG1	O2	Sulphonamide	2.74	107.0
His104 ND1	N4	Sulphonamide	2.77	164.3
Arg115 NH1	N1	Tetrazole	2.99	136.3
His84 NE2	N2	Tetrazole	2.99	157.3

Table 7.4: H-bond interactions between AH9095 and HPDHQase

7.5 Implications of Ligand Binding Mode

This is the first X-ray structure of a type II DHQase in complex with a non ‘substrate like’ inhibitor. This is highly significant, as if we are to develop inhibitors of type II DHQase as anti-infective agents, as the highly functionalised and synthetically challenging transition state analogues previously developed to date (Frederickson *et al.*, 1999, 2002) may not be viable as potential drugs.

Kleanthous *et al.* (1992) showed that phosphate ions acted as competitive inhibitors of type II DHQases and Evans *et al.* (2002) expanded this work to characterise the inhibition of type II DHQases by phosphate and sulphate ions. The crystal structures of SCDHQase and MTDHQase in complex with polyanions described in this work indicate that inhibition is achieved as the polyanions occupy the same position in the active site as the C1 position of dehydroquinone. The binding mode of AH9095 clearly indicates that this is indeed the case, with 5 of the 7 hydrogen bonds formed by the ligand with the protein mimicking those formed by the C1 functionalities of dehydroquinone. The sulphonamide oxygen atoms map the

positions of the C1 carboxylate, with the NH group forming similar contacts as the C1 hydroxyl.

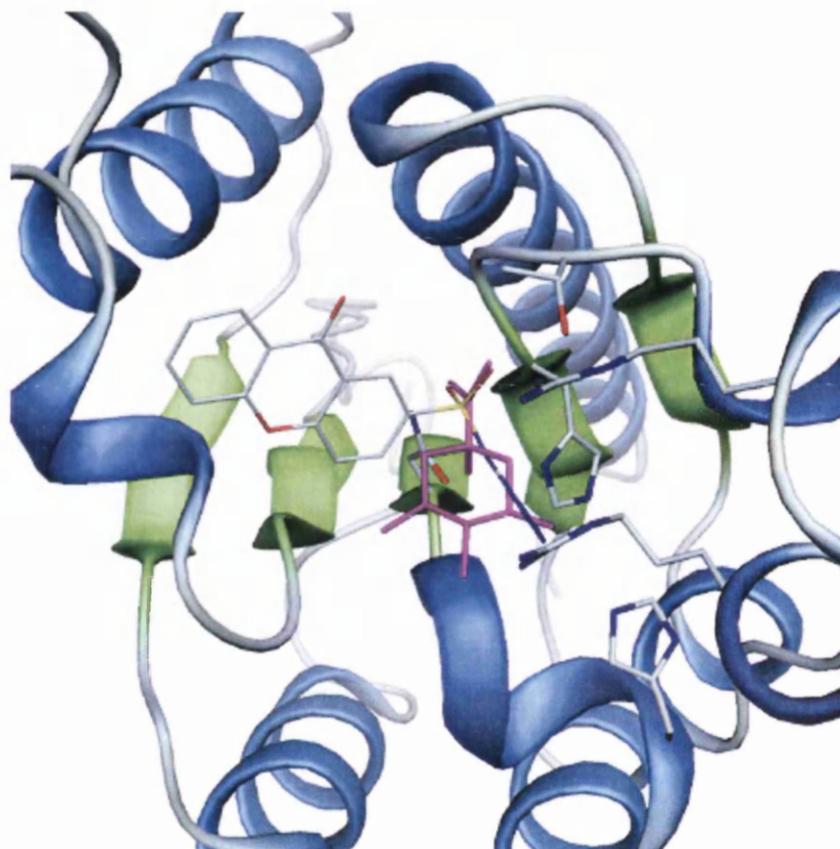


Figure 7.8: Active site of HPDHQase with AH9095 bound and dehydroquinone modelled into binding pocket. Dehydroquinone is coloured magenta and based upon the structure of the inhibitor FA6.

The sulphonamide moiety mimics the C1 functionality of dehydroquinone and the tetrazole moiety lies perpendicular to the plane of the dehydroquinone.

A recent patent filed by Arrow Therapeutics also identifies bisulphonamides as potent inhibitors of type II DHQase and Dehydroquinone synthase (International publication number WO 01/28537; Madge *et al.*, 2001). Clearly this moiety is of great the importance in the generation of high affinity ligands.

Although the sulphonamide is the most important moiety of the inhibitor, the tetrazole moiety plays a role extending the ligand into the active site, creating additional interactions with the protein. The addition of these interactions raises the affinity of the ligand above the level of inhibition achieved by a simple polyanion.

The xanthene moiety occupies the solvent channel leading to the entrance of the active site, packing against an almost perfectly planar hydrophobic surface on one side of HPDHQase. The hydrophobic interaction of this group with the active site entrance may lead to tighter ligand binding by holding the sulphonamide within the substrate cavity. This 'plugging' of the solvent channel may also hinder displacement of the sulphonamide by substrate.

7.6 Rationalisation of the specific inhibition by AH9095 of HPDHQase and not MTDHQase

High throughput screening data from Glaxo SmithKline indicated that AH9095 inhibits HPDHQase effectively but shows no activity against MTDHQase, therefore we should be able to rationalise this specificity using structural information.

The H-bond interactions between AH9095 and HPDHQase are formed with conserved residues within the active site. The sulphonamide and tetrazole moieties are smaller than the substrate, therefore it is unlikely that any steric considerations hinder binding. If we consider the surface of the active site that AH9095 packs against, leucine 13 and leucine16 are highly conserved residues while leucine 105 is conserved as a hydrophobic residue.

Consideration of the monomer in isolation does not explain the difference in affinity of AH9095 for MTDHQase and HPDHQase. Consideration of the surface contributed by subunit B which closes off the active site however, provides an explanation. Residue 95* shows no conservation within type II DHQases despite being located at the trimer interface. This residue is generally a small hydrophobic residue such as the leucine in HPDHQase, however in MTDHQase, the corresponding residue is a glutamic acid, Glu 92*. In MTDHQase, Glu92* forms an H-bond with Arg15 at the trimer interface. Arg15 also forms a salt bridge with Asp67* (Figure 7.9), therefore the negative charge of Glu92* is not fully neutralised. This results in a charged patch on the surface in close proximity to where the hydrophobic moiety of AH9095 would lie (Figure 7.10). This would provide a repulsive force towards the hydrophobic moiety of the ligand, dramatically reducing the binding affinity.

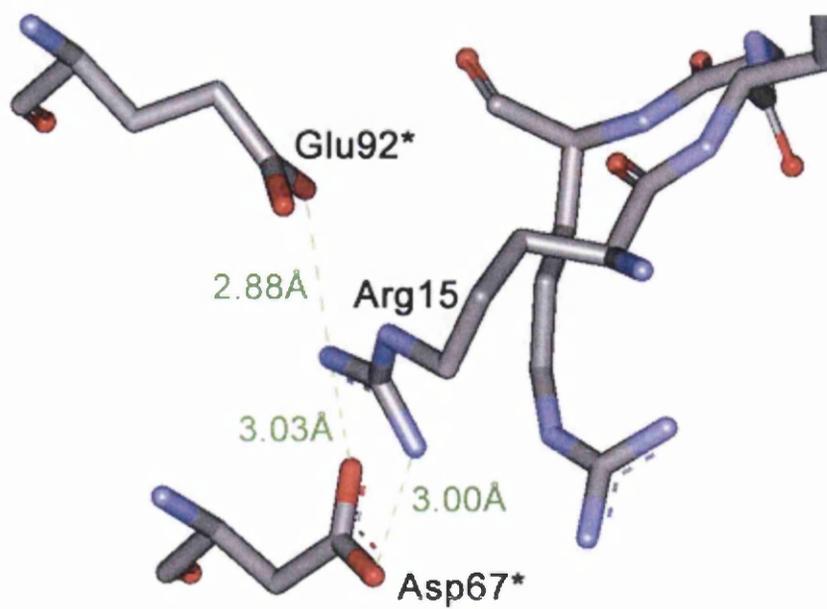
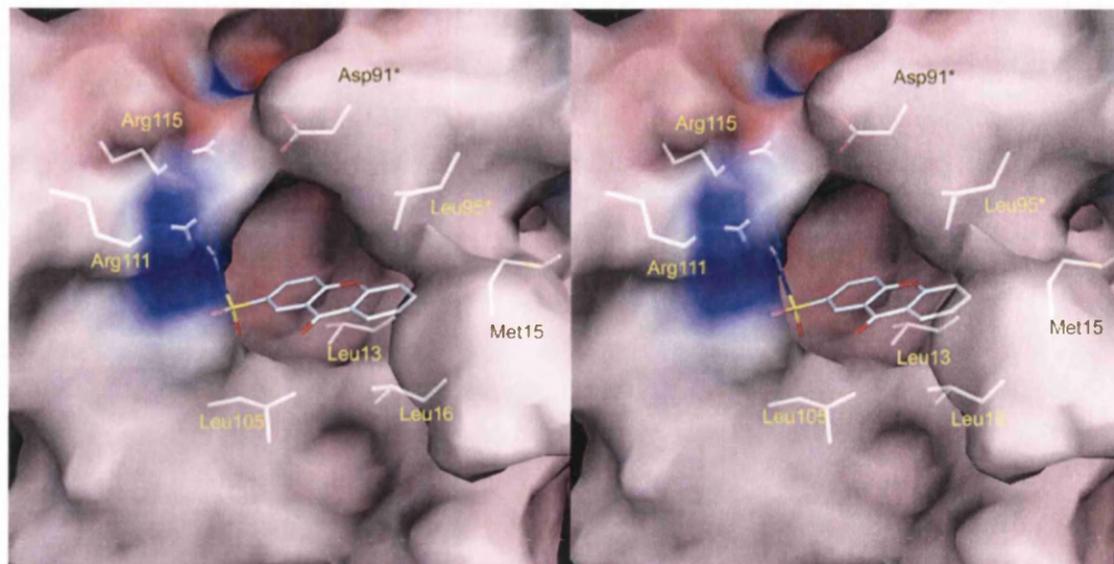


Figure 7.9: Interaction between Arg15, Glu92* and Asp67* in MTDHQase. H-bonds $<3.1\text{\AA}$ in length are shown as green dotted lines.

A



B

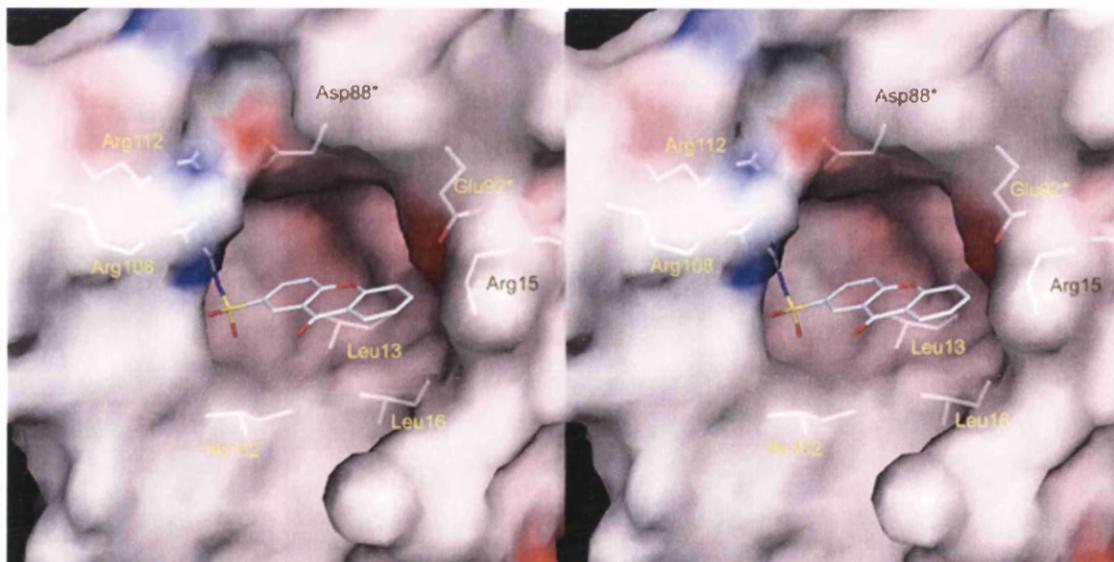


Figure 7.10: Stereo surface representations of HPDHQase and MTDHQase with AH9095 bound in the active site: A) HPDHQase/AH9095 complex; B) MTDHQase with AH9095 modelled into the active site. The surface is coloured highlighting patches of positive potential blue, negative potential red and neutral white. Key residues are shown in stick representation and labelled. The surfaces were calculated with GRASP (Nichols *et al.*, 1991), figure created with DINO (2002)

7.7 Crystal Structure of HPDHQase + 2,3-anhydroquinic acid (FA1)

7.7.1 Refinement and Model Building of HPDHQase + FA1

The model generated from the correct molecular replacement solution was refined using REFMAC from the CCP4 suite of programs. Initial rigid body refinement was carried out upon the MR model, with each of the 12 chains defined as a rigid group. The R_{work} and R_{free} after this stage were 33.6% and 33.9% respectively.

Restrained refinement was carried out for 2 cycles, applying strict NCS restraints upon each chain of the dodecamer, and the model altered manually using X-AUTOFIT within Quanta (Accelrys). Inspection of the 2Fo-Fc and Fo-Fc maps showed visible density within the active site regions corresponding to the presence of the ligand FA1. The R_{work} and R_{free} at this point in refinement were 28.0% and 32.8% respectively.

To allow docking of the ligand within the active site, the signal to noise ratio of the difference density was increased by averaging the density from each of the 12 monomers upon chain A. The NCS operators relating chain A to the other chains within the asymmetric unit were calculated using LSQKAB (CCP4, 1994), and a mask corresponding to the position of chain A was created using NCSMASK (CCP4, 1994). The program MAPROT (CCP4, 1994) was then used to superimpose the density from each monomer onto chain A using the density mask. A model for the ligand FA1 was obtained from the SCDHQase + FA1 complex (PDB I.D. 1GU1) and was successfully docked into the averaged Fo-Fc density within the active site of chain A. The NCS matrices calculated previously were inverted and used to transform the ligand from chain A into the active site regions of the other monomers.

A further 2 cycles of refinement were carried out using REFMAC, again applying strict NCS restraints. Due to the poor resolution of the diffraction data, the B value for the solvent contribution was fixed to 200 to aid the scaling of the Fobs to the Fcalc. TLS parameters were also included in refinement with each chain described as a single group. The final refinement statistics are shown in table 7.5

Refinement Statistics	HPDHQase + FA1
Resolution Range (Å)	30.0 – 3.2
R _{work}	26.8%
R _{free}	29.4%
RMS bond length deviation (Å)	0.032
RMS bond angle deviation (°)	2.592
Mean B factor (Å) main chain	28.89
Mean B factor (Å) side chain	30.06
Mean B factor (Å) heteroatom	35.28
Average Intersubunit RMS deviation (Å)	0.081

Table 7.5: Refinement statistics for HPDHQase + FA1

7.7.2 Structure of HPDHQase + FA1

The crystal structure of HPDHQase + FA1 consists of one complete dodecamer found in the asymmetric unit. The complete dodecamer is shown in figure 7.11. The monomer fold is similar to the structure of HPDHQase + AH9095, however due to the lower resolution of the diffraction data, density was only visible from residue His 2 at the N-terminus and no interpretable density was observed beyond residue Phe 153 at the C terminus. No interpretable density was observed in the lid domain between residues 19-25. The monomer fold of HPDHQase + FA1 is shown in figure 7.12

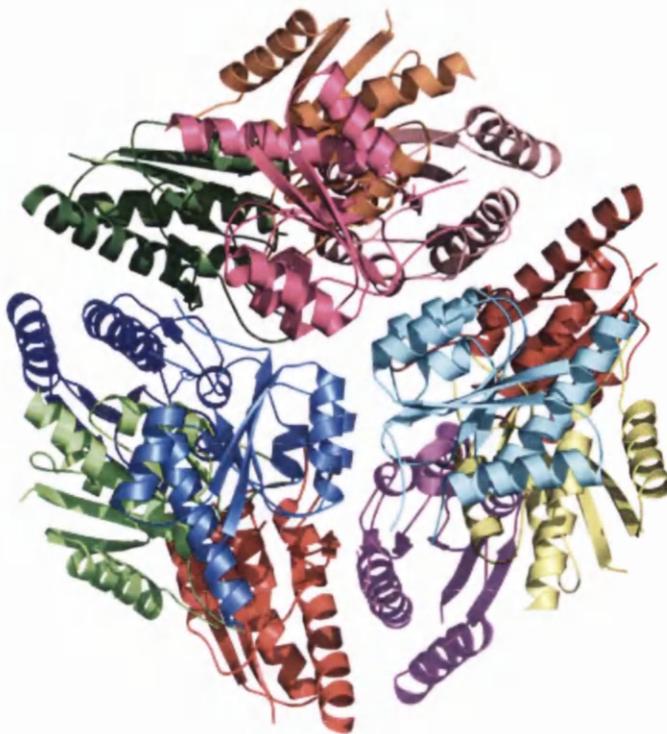


Figure 7.11: Quaternary structure of HPDHDQase + FA1. The dodecamer is viewed down the 3-fold NCS axis and each chain is coloured individually

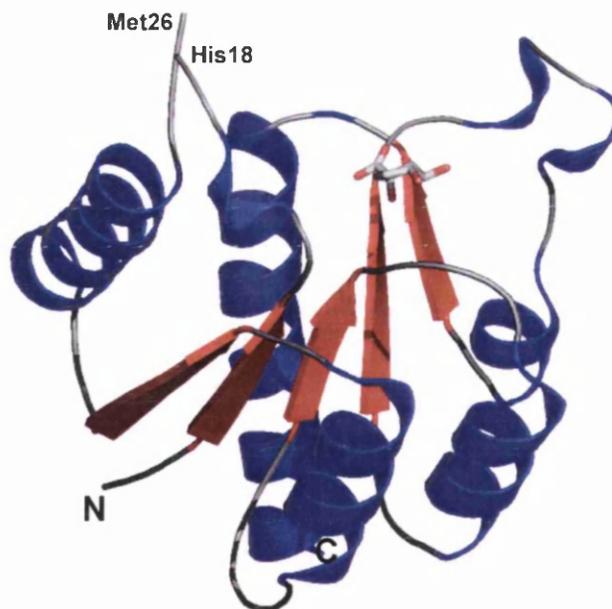


Figure 7.12: Monomer fold of HPDHDQase + FA1. α -helices are coloured blue and β -strands are coloured red. The ligand FA1 is shown in stick representation within the active site

7.7.3 Binding mode of ligand FA1 to HPDHQase

HPDHQase interacts with the ligand FA1 in a similar manner to the previously reported SCDHQase + FA1 complex (Roszak *et al.*, 2002). The carboxylate substituent at the C1 position of the ligand forms 4 H-bonds to the protein. Atom O1 interacts with the backbone N atom and the OG1 atom of Thr106 while atom O2 forms H-bonds with the backbone N atom of Leu105 and ND2 of Asn78. The C1 hydroxyl group forms 2 H-bonds to the protein, with atom ND2 of His 104 and with atom OD1 of Asn78. Of these 6 H-bonds formed between the protein and the C1 substituents of the ligand, 4 H-bonds are formed with residues in the H-X-S/T-N motif (residues 104-107) common to type II DHQases. Additional H-bonds are formed between the C5 hydroxyl group of the ligand with NH1 of Arg115 and NE2 of His84.

The interaction between the C4 hydroxyl of FA1 with the active site Asp residue observed in the SCDHQase/FA1 and MTDHQase/FA1 complexes does not appear to be as strong in the HPDHQase/FA1 complex. The distance between the O4 atom with the OD2 atom of Asp91* is 3.28Å which is outside the distance criteria defining a strong H-bond. A list of the H-bonds formed between the ligand FA1 and HPDHQase are shown in table 7.6 and stereo pair diagram of the interactions between FA1 and HPDHQase is shown in Figure 7.13

Protein	Ligand	Distance (Å)	Angle (°)
Thr106 N	O1	2.82	146.3
Thr 106 OG1	O1	2.38	137.8
Leu105 N	O2	2.90	173.2
Asn 78 ND2	O2	2.62	157.2
His 104 ND1	O3	3.05	156.6
Asn 78 OD1	O3	2.96	112.9
Arg 115 NH1	O5	2.93	138.4
His 84 NE2	O5	3.05	158.3

Table 7.6: H-bond interactions between FA1 and HPDHQase

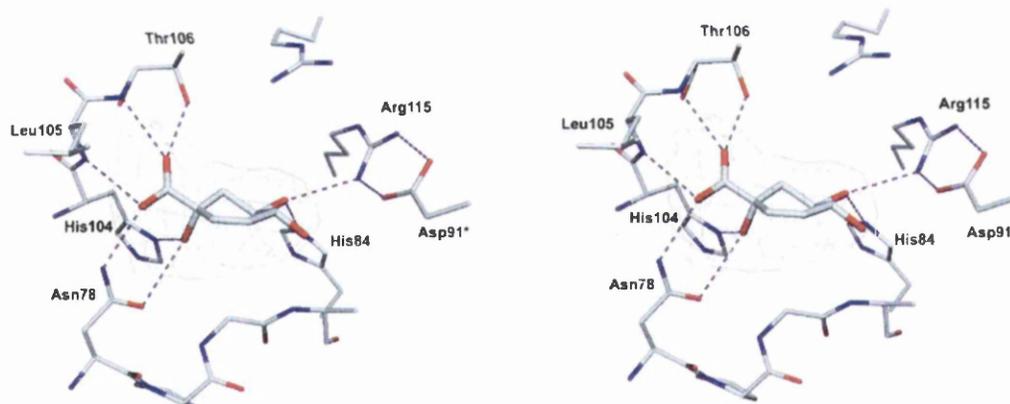


Figure 7.13: Stereo pair diagram showing a detailed view of the HPDHQase active site with FA1 bound. The 12-fold averaged omit map electron density is shown in green. H-bonds <math>< 3.3\text{\AA}</math> in length are shown as magenta dashed lines.

7.8 Discussion

The HPDHQase + AH9095 complex provides the first crystal structure of the type II DHQase from *H. pylori*. The overall monomer fold is a α/β flavodoxin type fold as described for other type II DHQases. The N-terminus is extended due to the addition of a 6x His tag and linker region employed in the protein purification. The amino acid sequence of HPDHQase is 162 residues in length with an extended C-terminal region. This C-terminal region results in the extension of the terminal helix α_4 which extends below the bulk of the monomer fold.

The ligand AH9095 is the first non-‘substrate-like’ inhibitor to be crystallised with a type II DHQase. The sulphonamide moiety of the ligand mimics the C1 carbon of dehydroquinone forming 5 of the 6 ligand-protein H-bonds. The tetrazole moiety occupies the substrate binding pocket forming a single H-bond to the protein, and the xanthene moiety lies in the entrance to the active site packing on one side against a hydrophobic surface. This shows that only similarity to the C1 position of dehydroquinone is essential for ligand affinity.

The ligand AH9095 has been shown to be active against HPDHQase, but not active against MTDHQase. The sulphonamide and tetrazole moieties are smaller than the substrate therefore it is unlikely that they play a specific role in the selectivity observed. The pocket occupied by the xanthene moiety in HPDHQase is hydrophobic, whereas in MTDHQase the glutamic acid at residue 95 confers a charge upon this surface, potentially hindering the binding of AH9095.

These observations may pave the way for rational design of specific type II DHQase inhibitors. Previously designed type II DHQase inhibitors have been small highly functionalised ‘substrate-like’ compounds that are not only synthetically

challenging, but are very hydrophilic and therefore predicted to be very susceptible to first pass metabolism (Lipinski *et al.*, 1997). Despite being invaluable for the elucidation of the enzyme mechanism, they would make very poor anti-infective agents. The addition of a large lipophilic group to a sulphonamide derivative will improve the solubility characteristics of an inhibitor, making it more 'drug-like'.

We can derive a general template for new inhibitors of type II DHQases based upon AH9095 shown in Figure 7.14.

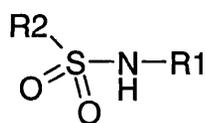


Figure 7.14: General template for inhibitors of type II DHQases based upon AH9095.

Group R1, on the amino side of the sulphonamide, could be used to probe the substrate binding pocket, thus should be small and potentially functionalised. Optimisation of this moiety could increase affinity of a ligand towards type II DHQases but not confer specificity.

Group R2 should be large and essentially hydrophobic in character. Modification of this group can be used to probe the entrance channel of the active site, allowing the specificity of a ligand to be optimised. This group may also be altered to change the solubility profile of a ligand without affecting the affinity towards type II DHQases.

7.8.1 Design of specific inhibitors of type II DHQase from *M. tuberculosis* based upon AH9095

The ligand AH9095 shows specificity towards HPDHQase as opposed to MTDHQase due to differences within the surfaces at the mouth of the active site. The MTDHQase active site has a negative charge at one end due to the presence of a non-conserved glutamate residue whereas the HPDHQase active site is totally hydrophobic. The modification of the R2 group of AH9095 to incorporate a H-bond donor to interact with this glutamate residue may increase the specificity, and possibility the affinity, of a modified ligand towards MTDHQase. The R1 group shall remain a tetrazole as the H-bond interactions between this moiety and conserved active site residues should retain some of the affinity shown by AH9095.

As there only appears to be space for a tri-cyclic moiety within this area of the active site, the xanthene moiety could be reduced to a chromene skeleton. From this chromene skeleton, amino groups can be attached via linkers, or tri-cyclic moieties incorporating amino groups can be designed. Ligands designed along these lines are shown in table 7.7.

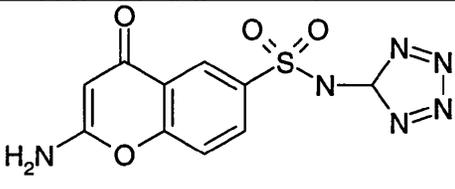
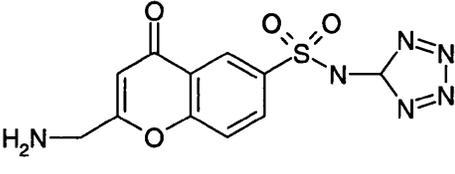
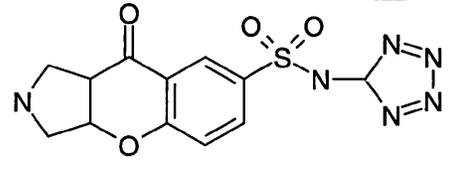
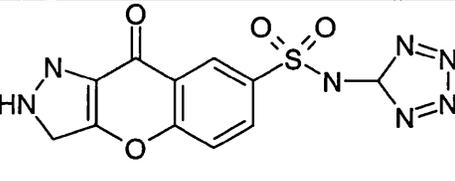
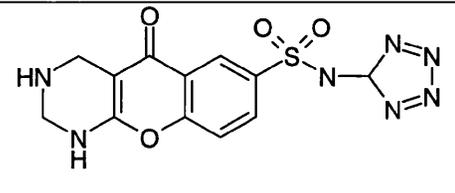
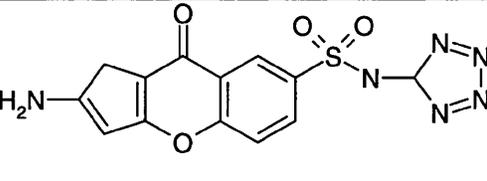
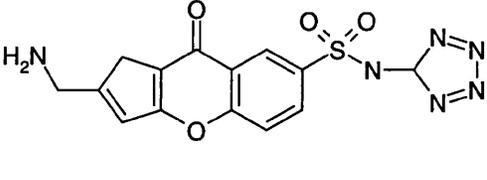
Ligand	Chromene derivative
	2-amino
	2-methylamino
	Pyrrolidine
	Pyrazole
	Piperidine
	amino-cyclopentadiene
	methylamino-cyclopentadiene

Table 7.7: Potential MTDHQase specific ligands based upon AH9095

Chapter 8 – Structure Based Design of *Mycobacterium tuberculosis* type II

Dehydroquinase inhibitors

8.1 Introduction

As more protein structures are elucidated, by methods such as protein crystallography and NMR, structural information can be used to drive the generation of highly specific enzyme inhibitors, as opposed to the traditional methodology of high throughput screening (HTS) and exhaustive structure-activity studies (Blundell, 1996). Structure based design (SBD) has become an essential part of the modern drug discovery process, and has played an important role in the generation of new HIV protease inhibitors amprenavir (Agenerase) and nelfinavir (Viracept) (Greer *et al.*, 1994), the flu drug zanamivir (Relenza) (Varghese, 1999) and rationalisation of the action of the protein kinase inhibitor imatinib (Gleevec) (Schindler *et al.*, 2000)

Many experimental strategies can be employed in SBD, depending upon the amount of information available at the outset of the process. The central requirements are however an accurate 3D structure of the target molecule, and the design of pharmacophore models, hypotheses which describe the various potential interactions that enable ligand binding within the protein target. An excellent review of advances in the SBD process has been published by Klebe (2000).

In this project we have several sources of information. Transition state analogues of type II DHQase, used initially to probe the enzyme mechanism, have been developed and their activity characterised against type II DHQases from *M. tuberculosis*, *S. coelicolor* and *A. nidulans*. Crystal structures are available of type II DHQase from *M. tuberculosis* (Gourley *et al.*, 1999; Evans *et al.*, 2002) and *S. coelicolor* (Roszak *et al.*, 2002) with various transition state analogues bound in the

active site. The crystal structure of *H. pylori* type II DHQase has been determined in this study and HTS data for this enzyme provided by Glaxo Smithkline.

8.2 Structural validation of selectivity shown by Transition State Analogues

Transition state analogues have long been used as enzyme inhibitors, notably in the field of protease inhibition (Radzicka and Wolfenden 1995). Fluoro-shikimate analogues have been shown to elicit antibacterial activity (Davies *et al.*, 1994) and transition state analogues been shown to inhibit dehydroquinase synthase (Montchamp and Frost 1997).

Initial studies probing the mechanisms of type I and type II DHQases developed transition state analogues which showed some efficacy and some selectivity between type I and type II DHQases (Harris *et al.*, 1996b; Bello *et al.* 1998, 2000; Parker *et al.* 2000). Frederickson and co-workers (Frederickson *et al.* 1999, 2002) progressed this work developing selective inhibitors of type II DHQases via a series of compounds that mimicked the proposed transition state for type II DHQase. These studies showed that selectivity between different type II DHQases could be obtained based upon then nature and geometry of the C3 group of the inhibitor. A summary of the inhibition statistics is shown in table 8.1.

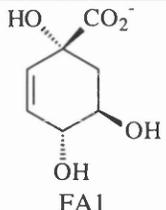
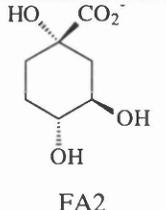
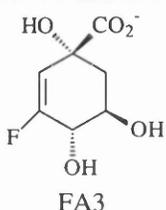
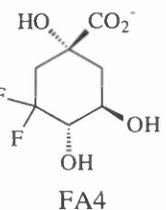
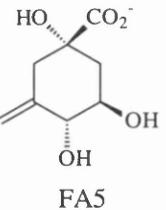
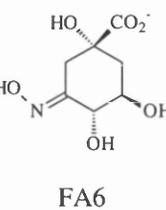
	Type I <i>S. typhi</i>	Type II <i>A. nidulans</i>	Type II <i>M. tuberculosis</i>	Type II <i>S. coelicolor</i>
 FA1	3 ± 1 mM	$60 \pm 10\mu\text{M}$	$200 \pm 20\mu\text{M}$	$30 \pm 10\mu\text{M}$
 FA2	4.5 ± 0.5 mM	1.5 ± 0.2 mM	1.2 ± 0.2 mM	0.6 ± 0.2 mM
 FA3	1.5 ± 0.5 mM	$50 \pm 5\mu\text{M}$	$10 \pm 2\mu\text{M}$	$15 \pm 2\mu\text{M}$
 FA4	8 ± 1 mM	0.6 ± 0.1 mM	0.7 ± 0.1 mM	0.7 ± 0.1 mM
 FA5	> 25 mM	2.2 ± 0.1 mM	0.7 ± 0.1 mM	2.5 ± 0.5 mM
 FA6	> 25 mM	$15 \pm 1\mu\text{M}$	$20 \pm 2\mu\text{M}$	$500 \pm 200\mu\text{M}$

Table 8.1 Inhibition statistics for a series of transition state analogues against different DHQases. Figure courtesy of Dr. Martyn Frederickson (University of Cambridge)

This series investigated the activity of compounds with sp^2 geometry at the C3 position (FA1 and FA2), and the simulation of the proposed charged C3 enolate by an oxime substituent (FA5 and FA6). A striking feature was the difference in K_i observed for compounds FA1 and FA6 with respect to MTDHQase and SCDHQase. The compound 2,3-anhydroquinic acid (FA1) has a K_i of $30\mu\text{M}$ with respect to SCDHQase, whereas for MTDHQase the K_i is $200\mu\text{M}$, nearly an order of magnitude greater. In contrast, the compound 3-hydroxyimino quinic acid (FA6) showed greater affinity towards MTDHQase ($K_i = 20\mu\text{M}$) compared with SCDHQase ($K_i = 500\mu\text{M}$), an order of magnitude greater. This work suggested that sp^2 geometry at the C3 position of the ligand was indeed very important for selective inhibition of type II DHQases. The observation that FA6 is a better inhibitor of type II DHQase than FA5 showed that mimicking the charge of the transition state enolate increased efficacy. However, the low K_i of FA6 with respect to SCDHQase was puzzling.

The 3D structures of native SCDHQase, SCDHQase + FA1, MTDHQase, MTDHQase + FA1, MTDHQase +FA6 have been determined previously (Gourley *et al.* 1999, Roszak *et al.* 2002, Robinson *et al.*, manuscript in preparation). Attempts to co-crystallise SCDHQase with FA6 were unsuccessful due to the very low affinity of the ligand. From these structures, the binding mode of the ligands will be analysed and the basis for the selectivity of FA6 towards MTDHQase will be rationalised. This information will be used in the design of a structure based design experiment.

8.2.1 Type II DHQase ligand binding

To design high quality pharmacophore models, the important protein ligand interactions have to be identified. High resolution crystal structures of FA1 bound to MTDHQase, SCDHQase and a low resolution structure of FA1 bound to HPDHQase (Section 7.7) have been determined. A diagram of the MTDHQase/FA1 complex is shown in figure 8.1 and a list of protein – ligand interactions is given in table 8.2. From this data it is apparent that the interactions between the C1 substituents of the ligand with the H-X-S/T (101-103) segment and Asn75 form the basis of the ligand recognition. Clearly these interactions must be included in any pharmacophore model.

The interactions between the C4 hydroxyl with Asp88* and the interactions between the C5 hydroxyl with His81 and Arg112 are also important as they would orientate the ligand as to fill the substrate binding pocket. The observation from Harris *et al.* (1996b) that 5-deoxy-3-dehydroquinone is a poor substrate for type II DHQases, and 4,5-dideoxy-3-dehydroquinone is a poor inhibitor of type II DHQases suggest that the interactions fulfilled by the C5 and C4 hydroxyl groups not only orientate the substrate into the correct position for catalysis but are critical for binding.

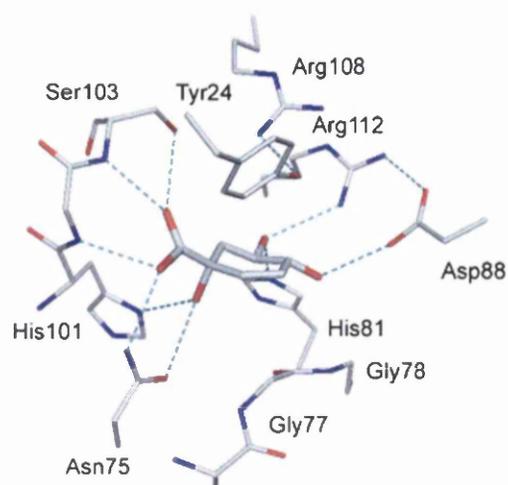


Figure 8.1: Detailed view of MTDHQase active site with the inhibitor FA1 bound. Hydrogen bonds <math>< 3.2\text{\AA}</math> in length are shown in blue

Protein	Ligand	Bond Length (\AA)	Bond Angle ($^\circ$)
Ser 103 OG	O1	2.97	97.1
Ser 103 N	O1	2.79	166.3
Ile 102 N	O2	2.89	158.1
Asn 75 ND2	O2	3.03	163.4
Asn 75 OD1	O3	3.12	124.8
His 101 ND1	O3	2.87	176.4
His 81 NE2	O5	2.92	170.7
Arg 112 NH1	O5	2.94	147.9
Asp 88* OD2	O4	2.38	167.5

Table 8.2 Protein – ligand hydrogen-bond distances and angles for the MTDHQase - FA1 complex

8.2.2 Specificity of 3-hydroxyimino quinic acid (FA6) towards MTDHQase

The ligand 3-hydroxyimino quinic acid (FA6) inhibits MTDHQase with a K_i over an order of magnitude smaller than the equivalent K_i for SCDHQase. Using the crystal structures available to structural basis for this selectivity, the SCDHQase/FA1 complex and the MTDHQase/FA6 complex were superimposed using LSQKAB from the CCP4 suite of programs.

The comparison of ligand binding modes between structures was not simple, not only due to the differences in sequence between the two enzymes, but also due to ligand induced conformational changes that became apparent during the analysis. The minimal catalytic unit of the type II DHQases has been shown to be the trimer (Price *et al.*, 1999). This is due to each active site being closed off by a neighbouring monomer. As both subunits contribute to ligand binding the trimer unit was identified as the template for comparison. Due to the ligand induced movement of the loops around the active site, the 3 central β strands (strands 2, 3, and 4) of each monomer were used as the reference residues for superimposition.

The ligand FA6 was modelled into the active site of SCDHQase using the position of FA1 as a reference. The position of the C1 carbonyl and hydroxyl groups were matched and the C4 and C5 hydroxyl groups were orientated to the best fit due to the difference in ring conformation of the two inhibitors. Figure 8.2 shows the position of FA6 modelled into the SCDHQase active site.

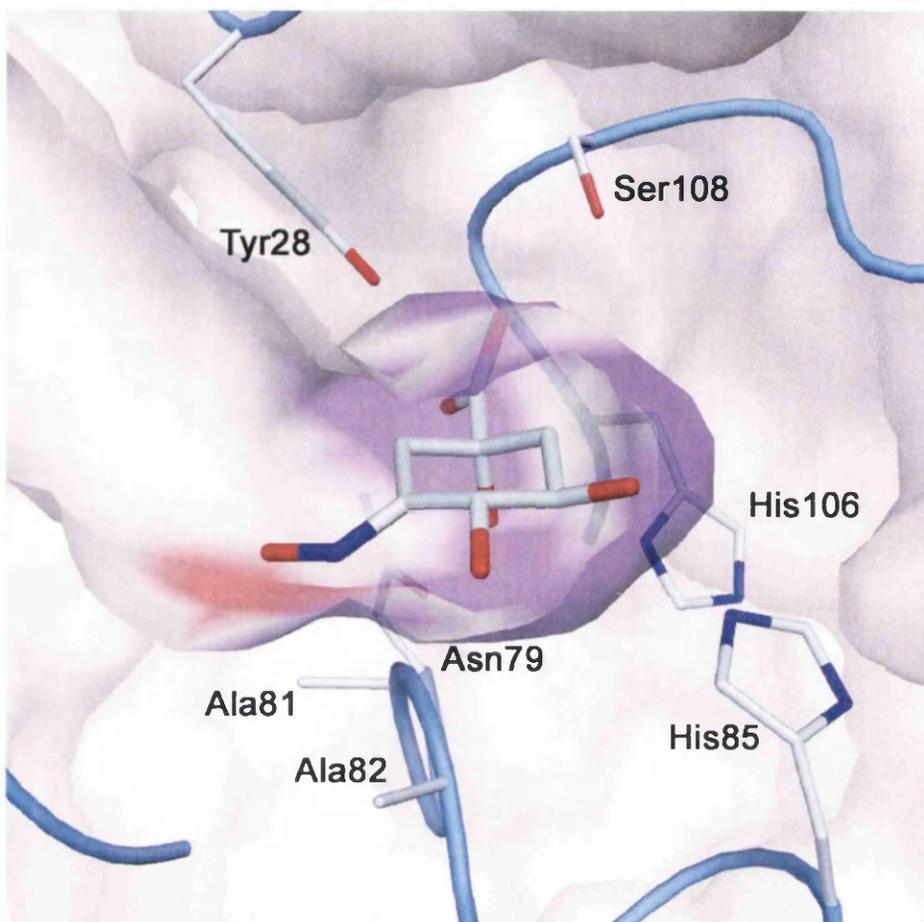


Figure 8.2 Active site of SCDHQase with FA6 modelled into active site based upon superposition of SCDHQ+FA1 and MTDHQ+FA1 complexes. The surface is coloured to show proximity of the ligand to the surface. Areas shaded purple show interactions $<3.2\text{\AA}$ but $>1.5\text{\AA}$, areas shaded red show interactions $<1.5\text{\AA}$, the proposed clash between FA6 and Ala81

The superimposition clearly shows that a steric clash would occur between the hydroxyl group of the oxime and the side chain of Ala81 on the floor of the active site, hence perturbing binding of the ligand. The residues that form the floor of the active site are not conserved throughout type II DHQases, though they are always glycine or alanine residues. MTDHQase has two glycine residues at this position whereas SCDHQase has two alanine residues. The nature of the first residue appears to affect the binding of FA6 as the catabolic type II DHQase from *Aspergillus nidulans* (ANDHQase) has a glycine at the first position and an alanine at the second (Figure 8.3). The K_i for FA6 with respect to ANDHQase is $15\mu\text{M}$, compared to $20\mu\text{M}$ for MTDHQase and $500\mu\text{M}$ for SCDHQase.

This shows that selectivity between type II DHQases can be obtained by consideration of steric effects within the active site as opposed to charge interactions.

8.2.3 Considerations for design of pharmacophores based upon DHQase/ligand complexes

In conclusion, from the DHQase/ligand complexes analysed, pharmacophores must take into account interactions satisfied by the C1 functionalities of the substrate and must consider that steric effects may influence specificity of resultant ligands. Pharmacophore models must therefore have a cluster of H-bond acceptor sites that mimic the C1 carbonyl and hydroxyl groups. Figure 8.4 shows the key areas of the structure that must be considered in pharmacophore design. The H-X-S/T segment that coordinates the C1 carbonyl is coloured red, and the floor of the active site coloured green.

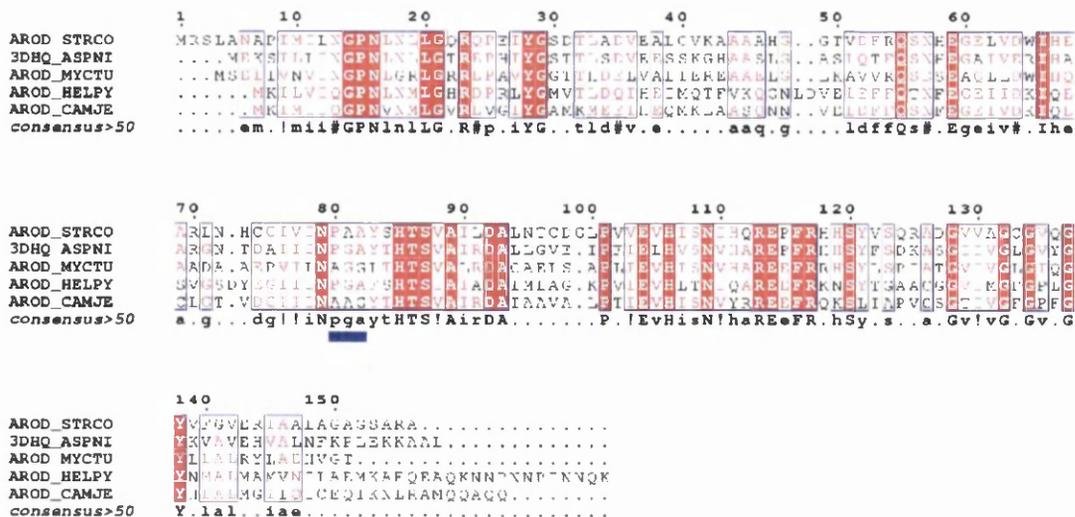


Figure 8.3: Sequence alignment of type II DHQases. The band below the alignment indicates residues that form the floor of the active site. Sequences shown are AROD_STRCO (SCDHQase), 3DHQ_ASPNI (ANDHQase), AROD_MYCTU (MTDHQase), AROD_CAMJE (CJDHQase) and AROD_HELPA (HPDHQase)

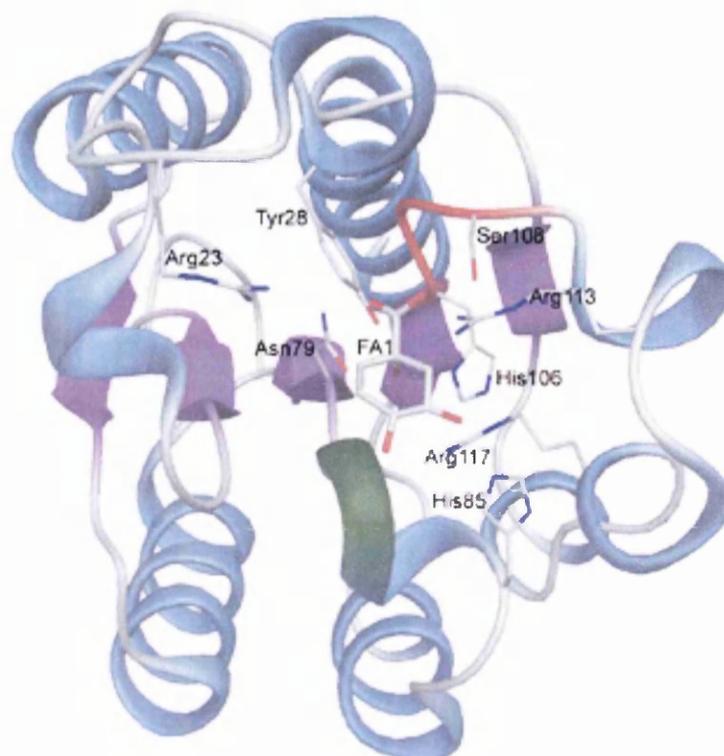


Figure 8.4: Ribbon diagram of an SCDHQase monomer illustrating the key areas for pharmacophore design. Conserved residues within the active site are shown and labelled. The H-X-S/T loop (106-108) is highlighted red and the floor of the active site is (81-82) highlighted green

8.3 Generation of Pharmacophore models for structure based design

To increase the possibility of identifying a novel small molecule that acts a selective inhibitor of the type II DHQase from *M. tuberculosis*, the SBD protocol was carried out for not only MTDHQase but also SCDHQase and HPDHQase. Pharmacophore models were constructed for MTDHQase, SCDHQase and HPDHQase in an identical manner.

The SBD process was carried out using the LUDI module within the Cerius II package and the CATALYST package, both from Accelrys (Cambridge, UK). Figure 8.5 shows a flow diagram of the SBD process using this software.

8.3.1 Definition of the Active Site

The first stage in the generation of a pharmacophore model is the definition of the 3D area of the protein structure that the pharmacophore will sample. In the case of type II DHQase this was trivial as the active site is well defined and all structures had the ligand FA1 bound in the active site.

The centre of the active site was defined as the centre of mass of the ligand FA1 (Figure 8.6a). From this point a sphere of radius 10Å was selected as the area of structure to be analysed to construct the pharmacophores (Figure 8.6b).

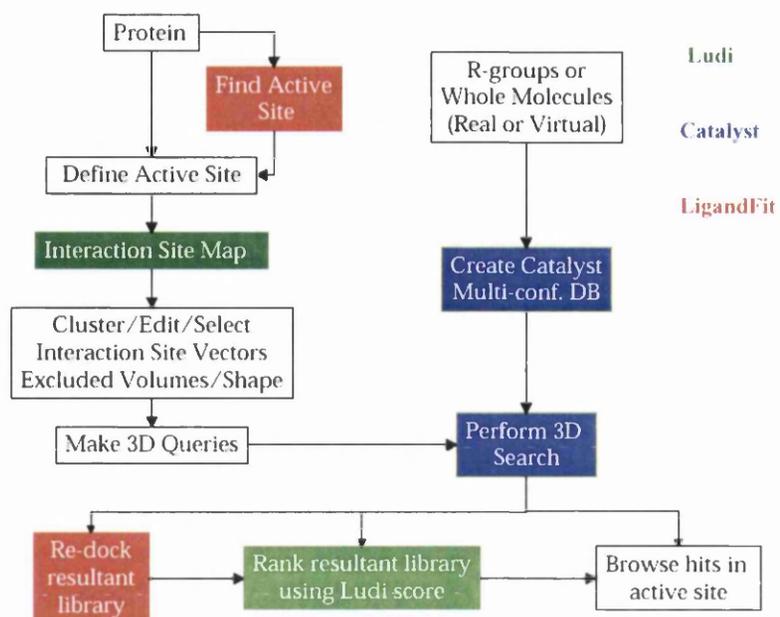


Figure 8.5: Structure based design process using the Accelrys software packages CeriusII and Catalyst

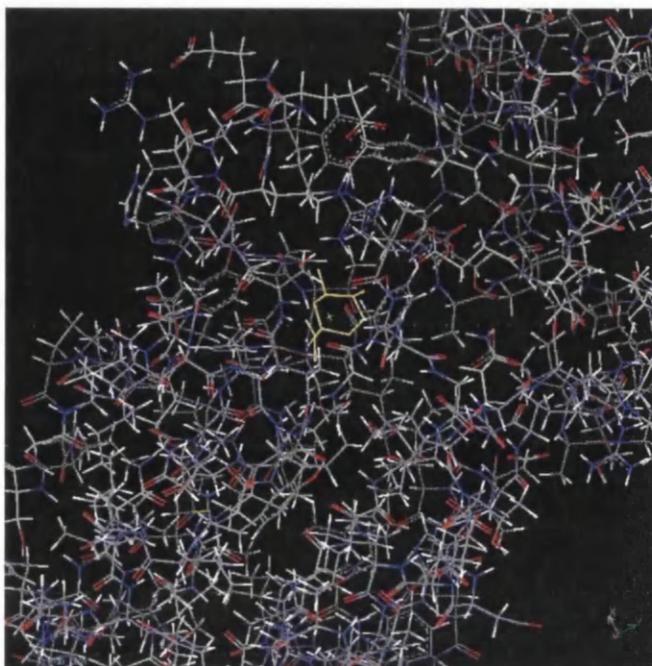


Figure 8.6a: Definition of the active site of MTDHQase. The ligand FA1 is highlighted in yellow

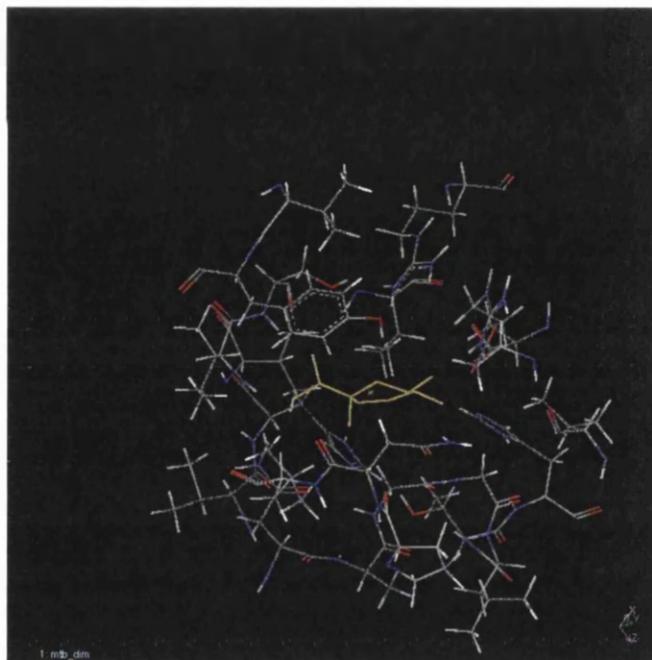


Figure 8.6b: Definition of a sphere of radius 10Å around the centre of the MTDHQase active site. The ligand FA1 is highlighted in yellow

8.3.2 Generation of a ligand interaction map using LUDI

The next stage in the pharmacophore design is to probe the defined active site for sites of potential protein-ligand interactions. This was performed using the LUDI module within CeriusII. This procedure identifies vectors of potential H-bond interactions within the active site, defining the sites as either H-bond donor, H-bond acceptor. Lipophilic sites are also identified. This results in an array of potential interactions around functional groups within the protein (Figure 8.7a) i.e around a backbone carbonyl, an array of H-bond donor sites will be identified.

As modelling each potential interaction would be very complex, the interactions are clustered around each protein functionality. The centroid of each cluster is then identified and used to describe the interaction. Figure 8.7b shows the centroids of the H-bond acceptor sites (red and white) and H-bond donor sites (Blue and white) within the active site of MTDHQase. The ligand FA1 (yellow) is shown to illustrate how the potential interactions map to the real interactions between the protein and the ligand. For each DHQase 7 H-bond acceptors sites, 3 H-bond donor and 4 lipophilic sites were identified.

8.3.4 Generation of multiple pharmacophore models

In order to maximise the number of hits from the *in silico* screen, a series of pharmacophore models are created using different combinations of the potential interactions. Each pharmacophore consists of 5 features, with at least a single H-bond acceptor, H-bond donor and lipophilic site. However as we know that the 2 H-bond acceptor sites corresponding to the C1 carbonyl are essential for ligand binding these interactions are conserved throughout all the pharmacophores. In total, 12 combinations of interactions were constructed for each DHQase.

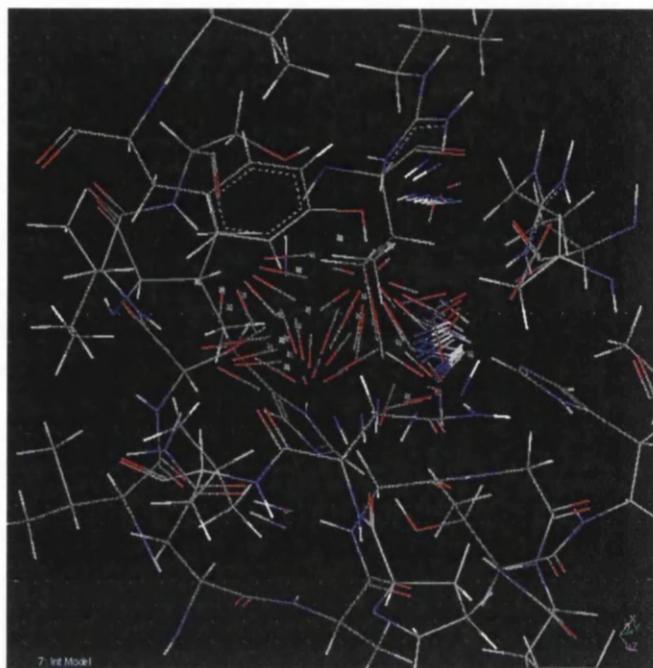


Figure 8.7a: Interaction map generated by LUDI for MTDHQase. H-bond acceptor sites are red and white bars, H-bond donor sites blue and white bars and lipophilic sites grey.

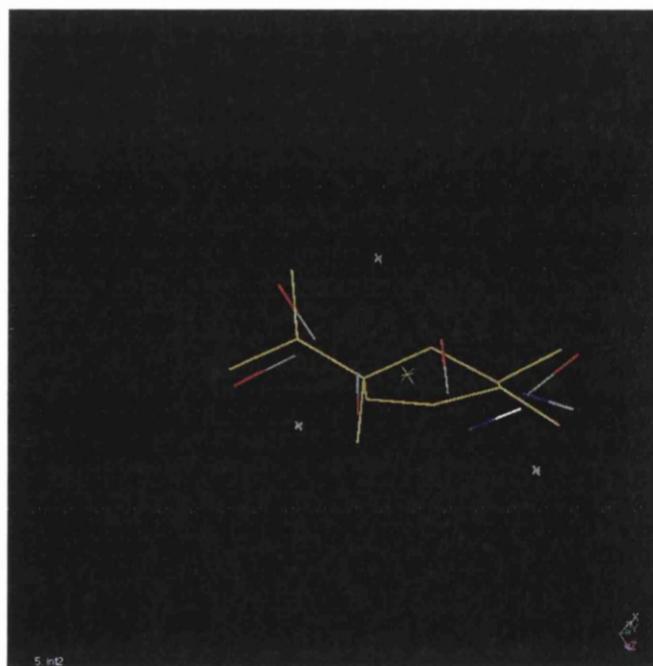


Figure 8.7b: Centroids of interaction clusters in MTDHQase. The ligand FA1 is shown in yellow to illustrate how the potential interactions map the observed protein ligand interactions.

As it has been demonstrated that the specificity shown by FA6 towards MTDHQase is a result of steric effects within the active site, the pharmacophores were designed to accommodate steric constraints. A series of 'exclusion spheres' were generated mapping the van der Waals surfaces of the protein atoms, and incorporated into the pharmacophore models.

An example of a pharmacophore model generated from MTDHQase is shown in figure 8.8. The H-bond acceptor sites are green spheres, H-bond donor sites pink spheres and lipophilic sites blue spheres. The exclusion spheres are shown as grey spheres.

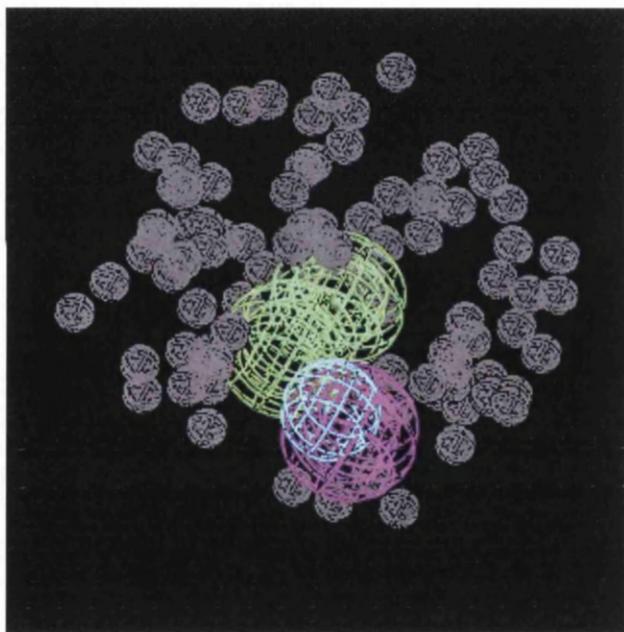


Figure 8.8: A pharmacophore model generated for MTDHQase

8.4 Database searching using Catalyst

The pharmacophore models generated for each type II DHQase will now be used to search a compound database using the program Catalyst (Accelrys). Catalyst searches multiple-conformation databases of compounds identifying hits that satisfy pharmacophore models. Multiple conformation databases are used as there is no guarantee that a ligand will bind to a protein in the lowest energy conformation, easing steric constraints.

The database searched was the Maybridge '99 database. This database was selected as it contains quite a diverse range of compounds which are all commercially available. This database contains 55,273 compounds.

Catalyst was run in 'fast' mode, which rejects any compounds that collide with the exclusion spheres in the pharmacophore model. This took into consideration steric effects upon ligand specificity.

Maybridge '99 was searched by Catalyst using each of the 12 pharmacophore models generated for each type II DHQase as templates. The numbers of hits found by Catalyst for each type II DHQase are shown in table 8.3

DHQase	No. Hits
MTDHQase	111
SCDHQase	881
HPDHQase	386

Table 8.3: Number of hits identified by Catalyst searching Maybridge '99 database using type II DHQase pharmacophore models

Of the 111 hits identified for MTDHQase, 36 were unique hits not identified for SCDHQase or HPD HQase. The 36 MTDHQase unique compounds are shown in Table 8.4.

8.5 Assay to determine inhibition of MTDHQase by *in silico* screen hits

The compounds identified as potential selective MTDHQase inhibitors from the SBD experiment were obtained from Maybridge. 30 of the 36 hits were available for purchase and 1mg of each available compound was obtained. The compounds were dissolved in 100% DMSO at a concentration of 10mM and stored at 4°C. Table 8.4 shows the compounds that were not available.

8.5.1 Activity assay – first round

Compounds were tested for inhibition of MTDHQase using the standard DHQase assay (Section 2.10). The assays were carried out in 1ml quartz cuvettes with a path length of 1cm. The assay conditions were 0.5µg MTDHQase, 2mM dehydroquinone and 100µM inhibitor. Table 8.5 and figure 8.9 show the % inhibition for each compound under these assay conditions.

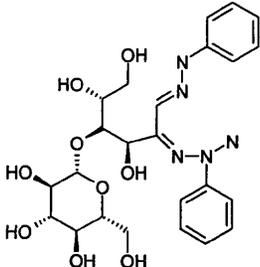
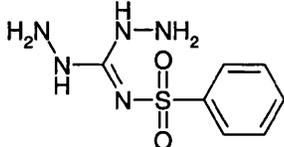
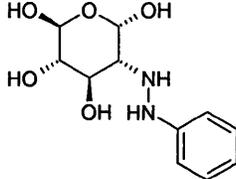
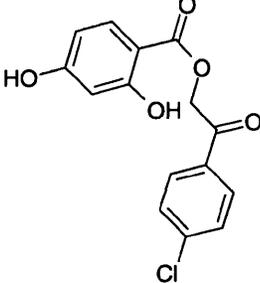
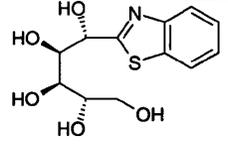
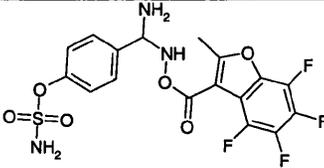
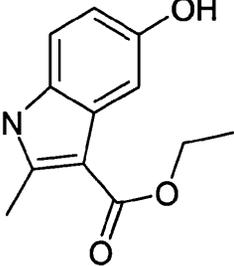
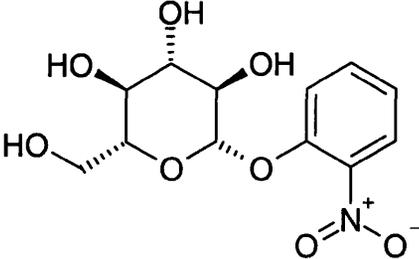
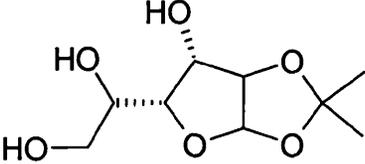
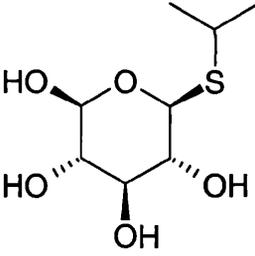
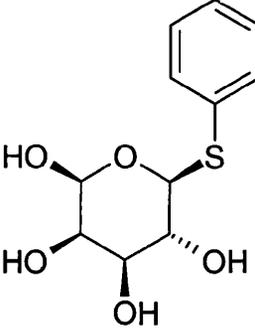
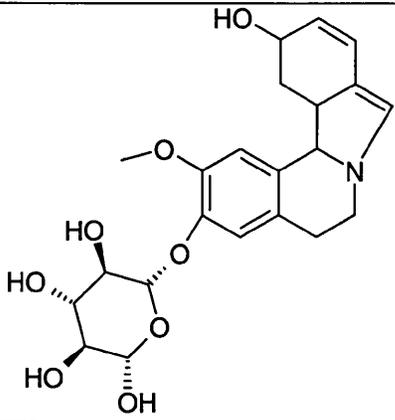
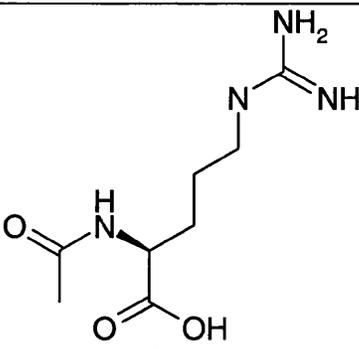
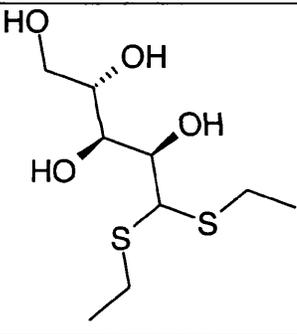
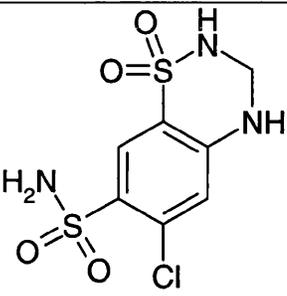
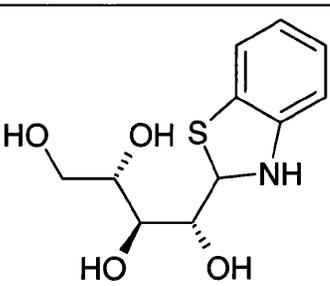
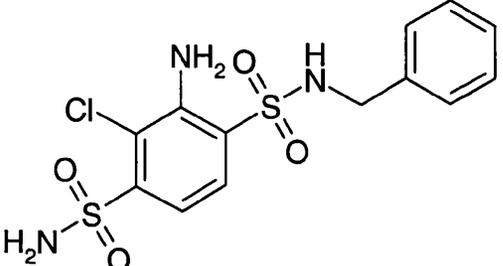
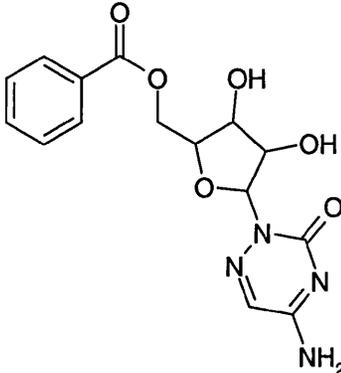
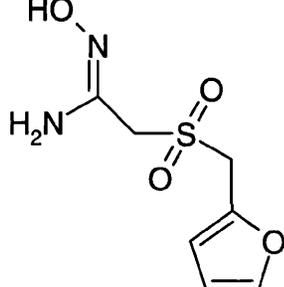
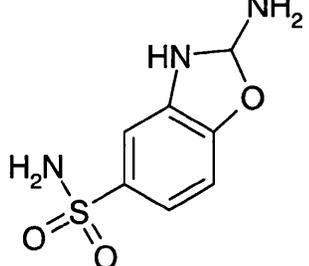
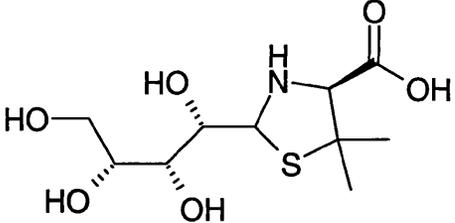
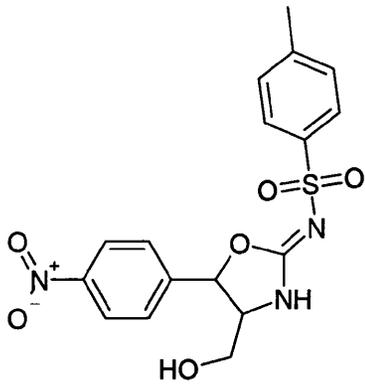
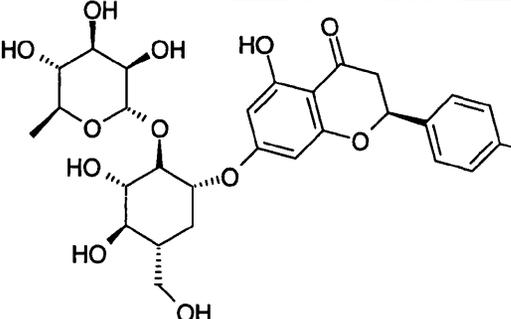
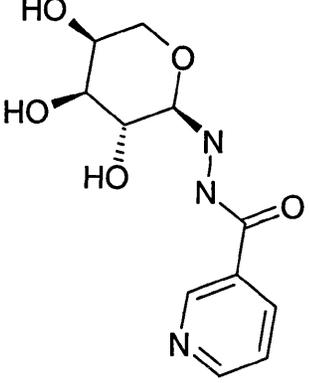
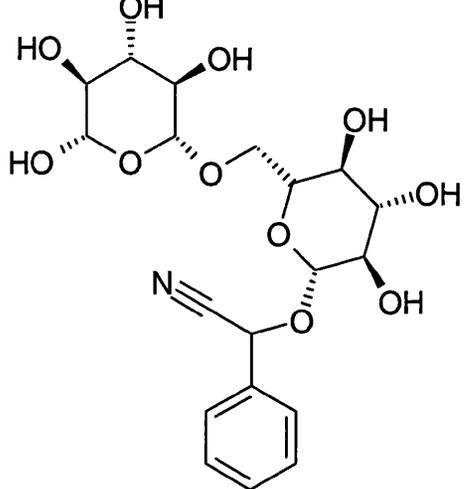
Structure	Maybridge Catalogue No.
	BTB 12004
	CD 03087
	JFD 00804
	RDR 01608
	RJC 01330
	SEW 02056

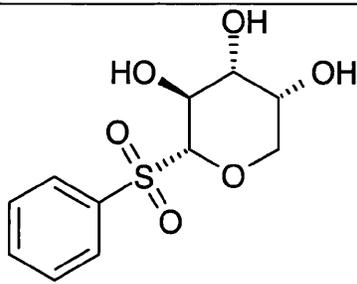
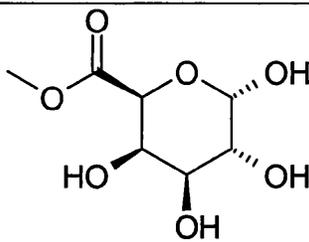
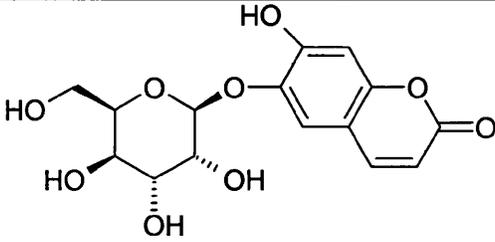
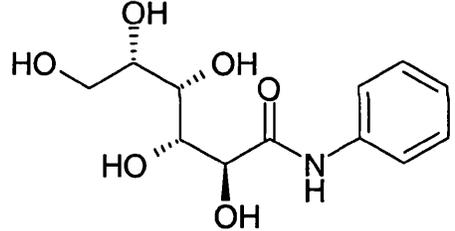
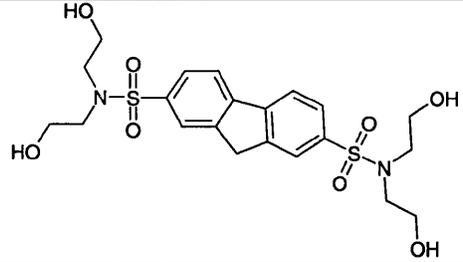
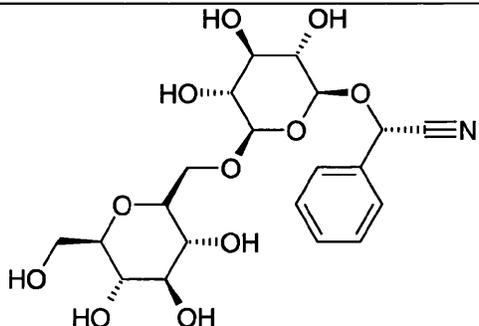
Table 8.4 Compounds identified by structure based design that were not available for purchase

Compound Number	Structure	Maybridge Catalogue No.	% Vo/Vmax
1		BTB 01488	67
2		BTB 11975	88
3		BTB 11978	94
4		BTB 11985	88
5		BTB 11990	97

6		BTB 12776	67
7		BTB 13543	79
8		JFD 00282	85
9		JFD 00715	82
10		JFD 00891	73

11	 <chem>Nc1cc(Cl)cc(S(=O)(=O)NCc2ccccc2)c1S(=O)(=O)N</chem>	JFD 01751	52
12	 <chem>Nc1nc2c(ncn2C3OC(O)C(O)C3OC(=O)c4ccccc4)nc1</chem>	JFD 02279	55
13	 <chem>NC(=N)CCS(=O)(=O)CC1=CC=CO1</chem>	KM 03112	70
14	 <chem>Nc1nc2c(ncn2C3=CC=C(C=C3)S(=O)(=O)N)nc1</chem>	NRB 01507	64
15	 <chem>CC1(S)NC(C1)C(=O)O[C@@H]2C[C@@H](O)[C@H](O)CO2</chem>	NRB 03989	94

16		NRB 04410	58
17		NRB 04426	76
18		NRB 04521	52
19		NRB 04549	88

20		RJC 00283	100
21		RJC 00828	88
22		RJC 00925	82
23		RJC 01515	64
24		RJC 01776	94
25		RJC 02360	94

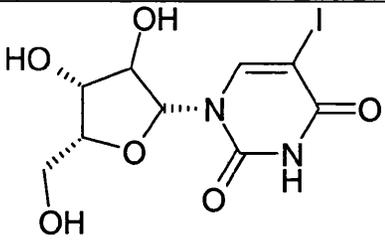
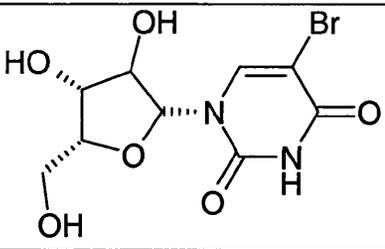
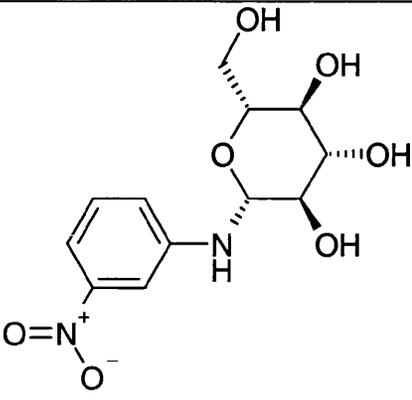
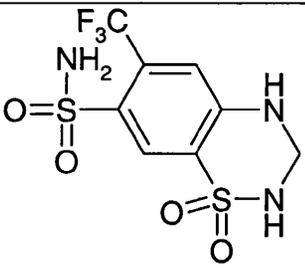
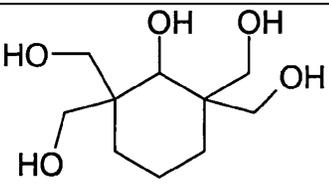
26		RJC 03215	73
27		RJC 03234	88
28		RJC 03292	17.6
29		SB 01887	95
30		XBX 00089	100

Table 8.5: Compounds identified by structure based design to be inhibitors of MTDHQase.

Inhibition statistics from the first round of assays are shown

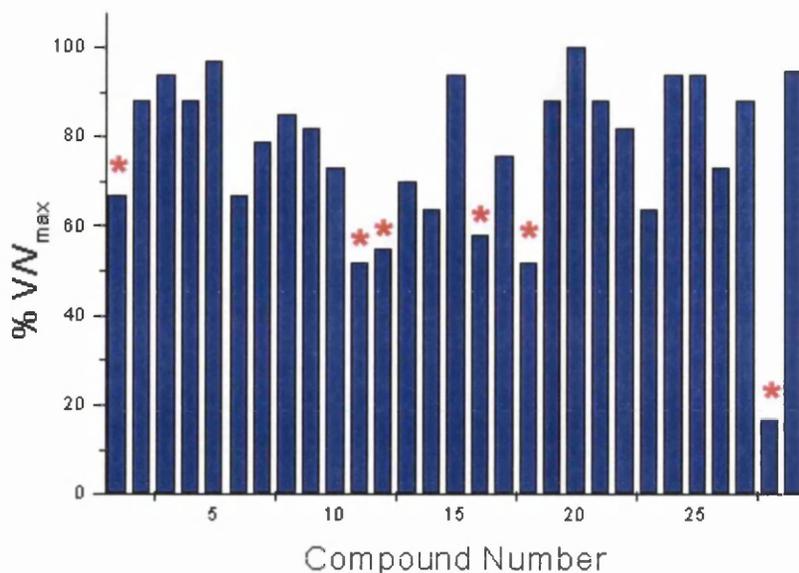


Figure 8.9: % V/V_{max} measured for 30 compounds against MTDHQase. Those marked with a star were taken forward into the next round of assays.

8.5.2 Activity assay- Second round

Six compounds that showed the highest inhibition of MTDHQase were taken forward into the second round of activity assays. In this assay each result was measured in triplicate against 2 concentrations of substrate, 500 μ M and 100 μ M. The inhibitor concentration was 100 μ M and the protein concentration 0.5 μ g/ml. Results are shown in table 8.6.

The results show that five of the compounds have similar activity against MTDHQase in the presence of 100 μ M and 500 μ M dehydroquinone. Compound 16 appeared to form a precipitate in the presence of 500 μ M dehydroquinone therefore results were not recorded. Compound 28 showed the best inhibition reducing MTDHQase activity by greater than 50% in the presence of 500 μ M dehydroquinone.

8.6 Predicted binding modes of active compounds

As Catalyst performs fitting of molecules into an active site, the predicted poses of the 6 most active compounds were analysed. Figure 8.10 shows these compounds within the active site of MTDHQase.

Compounds 11 and 16 contain sulphone moieties similar to the HPDQase specific ligand AH9095, however the predicted binding orientation of these compounds do not position the sulphone moiety such that it is fully coordinated by protein atoms. The coordination of the sulphone groups appears to be restricted by steric constraints upon the ligands, thereby not allowing this group to bind in the same orientation as the sulphate polyanion.

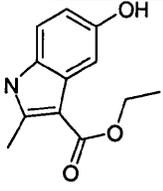
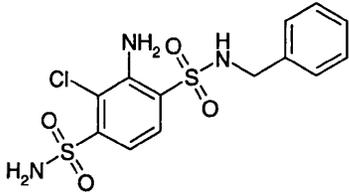
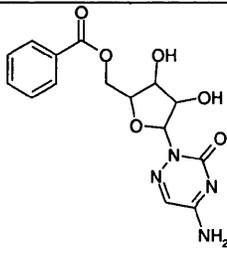
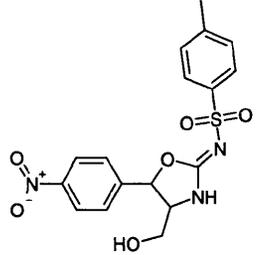
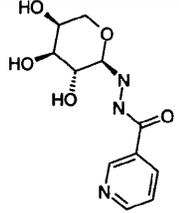
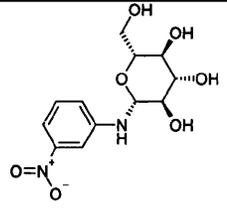
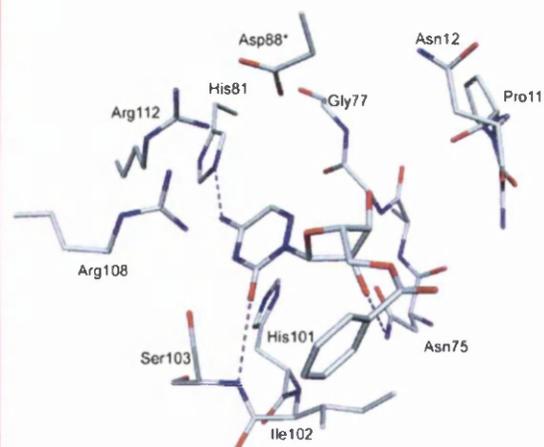
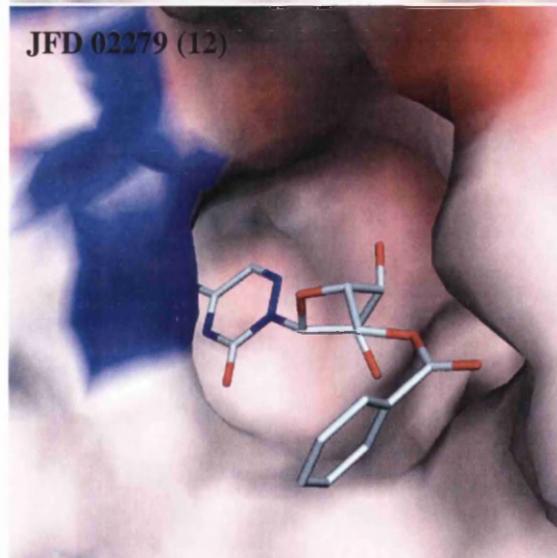
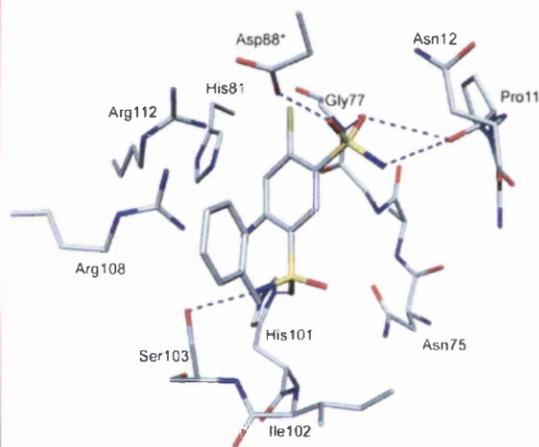
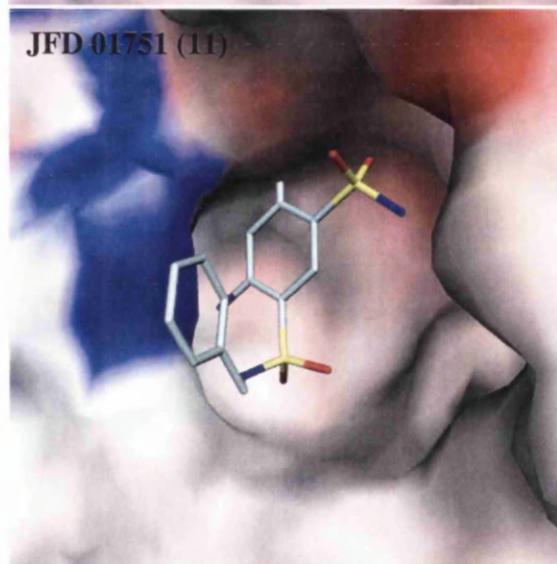
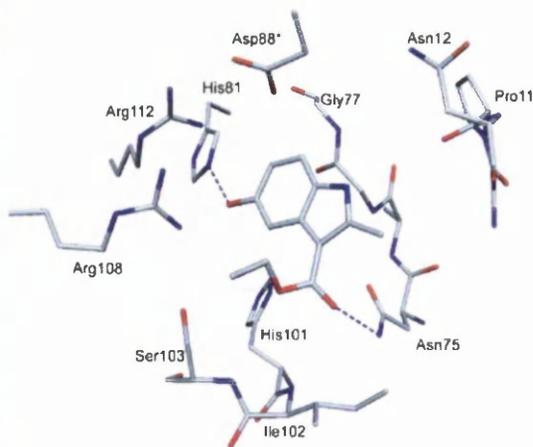
Compound Number	Structure	% V/V _{max}	
		500μM DHQate	100μM DHQate
1		62	58
11		62	60
12		68	72
16		79	N/A
18		75	65
28		41	46

Table 8.6: Second round assay results for top six compounds against MTDHQase



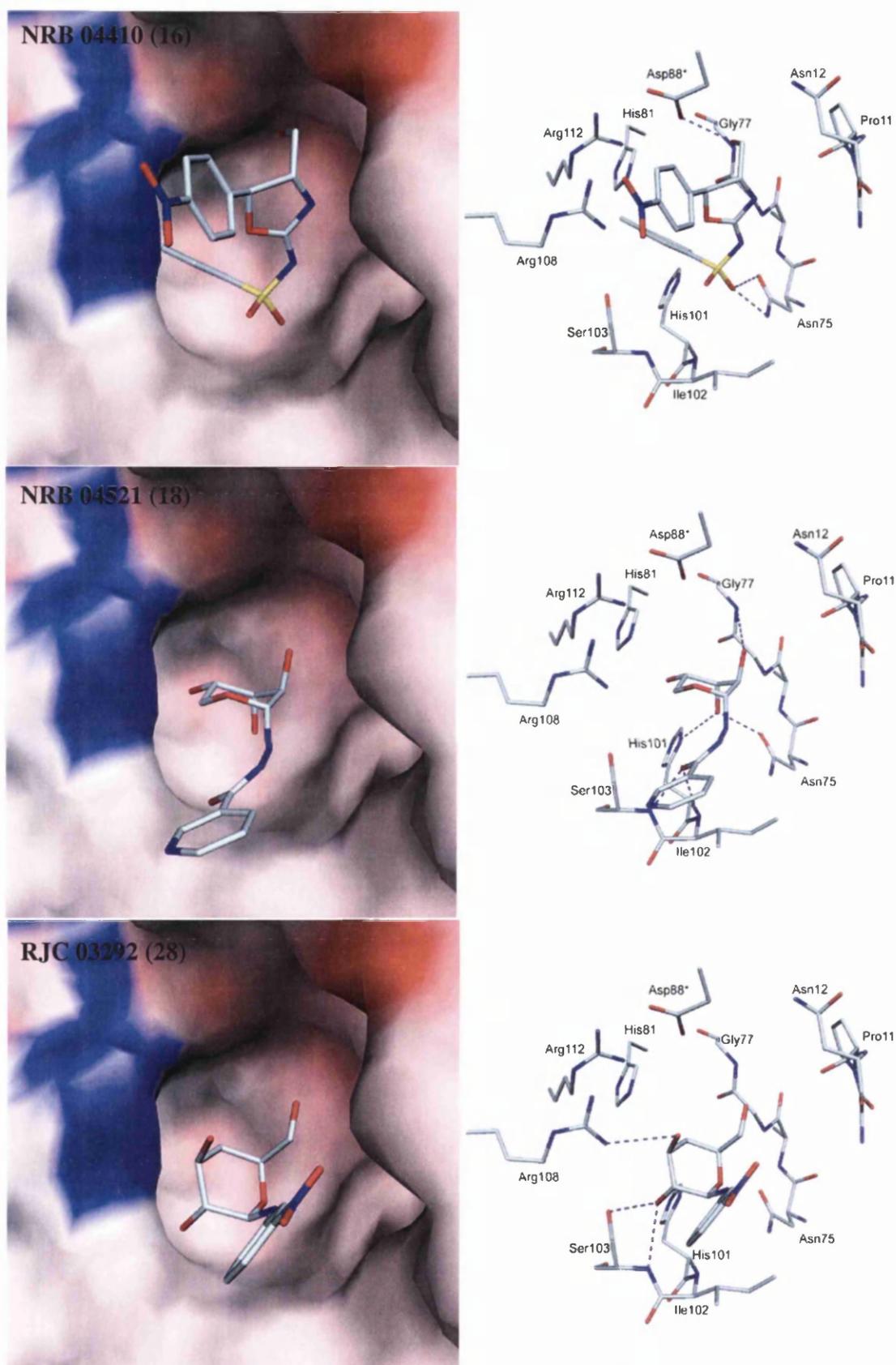


Figure 8.10: Proposed binding modes of MTDHQase inhibitors. Left panels show a surface representation of the enzyme coloured by potential. Right panels show detailed views of the active site. H-bonds between the ligands and the protein are shown as magenta dotted lines. See figure 8.1 for substrate binding mode

Compound 28 appears to bind tightly to Ser103 and Arg108, occupying the substrate binding cavity more effectively than the other ligands. As Ser103 is critical for substrate binding this suggests a reason why this compound is the most effective MTDHQase inhibitor identified.

8.7 Discussion

The use of structure based design has identified a selection of compounds that show activity towards MTDHQase. Pharmacophore models were designed to represent the H-bond interactions formed by transition state analogues and the protein. Analysis of DHQase/transition state analogue crystal structures suggested that target specificity could be obtained by consideration of steric constraints within the active site, therefore pharmacophore models incorporated exclusion volumes.

The six compounds that show the greatest affinity towards MTDHQase are quite diverse, sharing few functional features. Two of the compounds contain sulphone groups similar to the HPDHQase specific inhibitor AH9095, however the predicted binding mode of these compounds does not result in full coordination of the sulphone moiety as observed in HPDHQase/AH9095.

Future study should involve analysis of key features of the active compounds to generate new pharmacophore models that should enable identification of compounds with greater efficacy. The pharmacophores in this study only occupied a radius of 10Å around the centre of the substrate binding cavity. As shown in chapter 7, the specificity of AH9095 towards HPDHQase as opposed to MTDHQase is due to differences in topology and charge in the entrance to the active site. Therefore, by extending the pharmacophore models to incorporate this area of the active site may result in the identification of more potent MTDHQase specific inhibitors.

Chapter 9 – Development of a novel assay to follow activity of type II DHQase

9.1 Need for a new DHQase assay

As outlined in section 2.10, the standard assay for DHQase activity follows conversion of 3-dehydroquininate to form 3-dehydroshikimate. An increase in absorbance at 234nm is observed as the product has a conjugated bond system not present in the substrate. This assay is robust, but does require high quality substrate and is carried out using quartz cuvettes and a high quality spectrophotometer. This assay is suitable to determine kinetic characteristics of DHQases and to study in depth the inhibitory properties of potential ligands, but to achieve higher throughput, analysing larger number of compounds, the use of cuvettes is severely time limiting. Many standard assays are routinely carried out using 96-well plates, however due to the low wavelength of the absorbance in this assay, the signal is lost as the plastic plates also absorb light in this region of the spectrum. High throughput screens have been carried out on DHQases by coupling the assay with that of shikimate dehydrogenase (SDH), the next step in the pathway. Shikimate dehydrogenase requires NADPH as a cofactor, and the conversion of NADPH to NADP⁺ can be followed by an increase in absorbance at 340nm.

There are problems associated with coupled assays with respect to identifying inhibitors of a specific enzyme as the compounds may elicit effects upon the coupling enzyme, generating false positives. This is a significant problem with coupling enzymes that utilise a co-factor, as many compounds in the screen may block the co-factor binding site. In the quest to develop a new assay for DHQase, the conversion of 3-dehydroshikimate to protocatechuate (PCA) by 3-dehydroshikimate

dehydratase (DHSase), an enzyme from the quinate pathway, was investigated

(Figure 1).

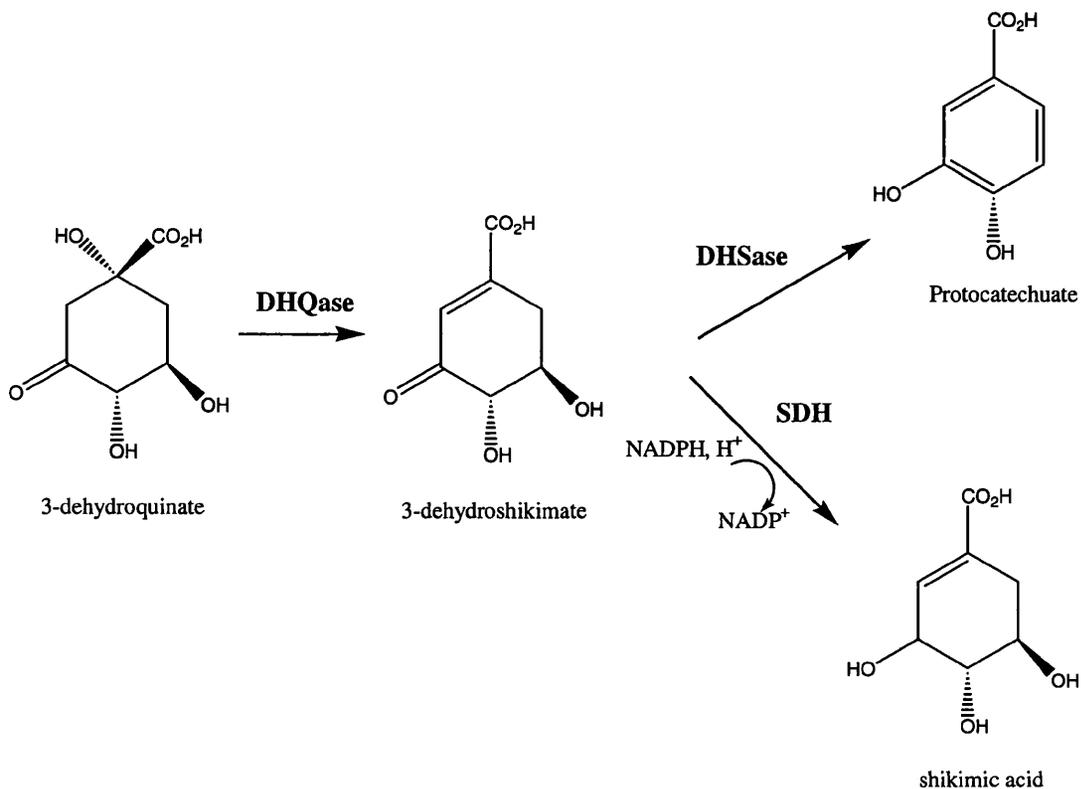


Figure 9.1: Conversion of 3-dehydroquininate to 3-dehydroshikimate by dehydroquinase

(DHQase) and its subsequent transformation to shikimic acid, catalysed by shikimate

dehydrogenase (SDH), and to protocatechuate, catalysed by dehydroshikimate dehydratase

(DHSase)

PCA generation has been used to quantify activity in the quinate pathway of fungi as PCA forms a strongly coloured complex in the presence of iron (III) chloride (Partridge *et al.*, 1972; Grant *et al.*, 1988), which can be observed on solid growth media.

9.2 Dehydroshikimate dehydratase

Dehydroshikimate dehydratase (DHSase) is the third enzyme in the quinate pathway, which converts quinic acid to protocatechuate for use as a carbon source. In *Neurospora crassa* (Case and Giles, 1976) and *Acinetobacter calcoaceticus* (Elsemore and Ornston, 1995) this pathway is controlled by a tightly linked cluster of four genes (*qa 1-4* in *N. crassa* and *quiBCXA* in *A. calcoaceticus*), three of which encode for the individual enzymes of the pathway. In *Aspergillus nidulans* (Grant *et al.*, 1988) there are six genes in the cluster (*qutCDBAER*) consisting of two regulatory genes, three enzymes and a permease. The first step in the pathway converting quinate to 3-dehydroquinate is catalysed by a NAD dependent quinate dehydrogenase. The second step is the DHQase reaction converting 3-dehydroquinate to 3-dehydroshikimate, followed by production of protocatechuate by DHSase.

DHSase from *N. crassa* has been purified and characterised (Stroman *et al.* 1978), to show that the enzyme is active as a monomer of mass 37kDa. It was also noted that DHSase was sensitive to thermal denaturation, though this was stabilised by addition of Mg^{2+} ions. The gene *quiC* encoding the DHSase from *A. calcoaceticus* has also been cloned (Elsemore and Ornston, 1995). Expression of this gene in *E. coli* induced production of PCA.

Two compounds were identified that form coloured complexes in the presence of PCA. The classical indicator, $Fe(III) Cl_3$, forms a complex with PCA that absorbs around 500 – 580nm, forming a dark blue solution. The observation that $Na_2 MoO_3$ forms coloured complexes in the presence of vicinal di-hydroxy phenols was also investigated. The complex between PCA and $Na_2 MoO_3$ is coloured dark yellow, absorbing light between 350 and 450nm. By varying the PCA and indicator

concentration in a 96 well plate, the minimal amount of indicator and PCA required to detect a signal was investigated. The spectroscopic reading of the 96-well plates was carried out using a Genosys plate reader (Tecan).

9.3 Absorbance spectra of 3-dehydroquinate and PCA

Before any experiments with indicators were designed, absorbance spectra of the substrate, intermediate and product of the assay were recorded to determine if they conferred any intrinsic background absorbance in the visible region (Figure 9.2). All experiments used 50mM Tris-acetate pH 7.0, 2.5mM MgCl₂ as the assay buffer.

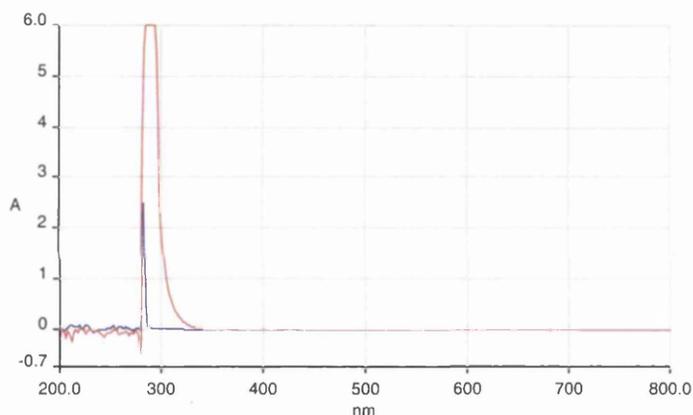


Figure 9.2: Absorbance spectra in the range 200-800nm of 3-dehydroquinate (Blue) and PCA (Red)

This shows that neither substrate nor product absorb light in the visible region.

9.4 Detection of PCA using Fe(III) Cl₃ as an indicator

The absorbance spectra in the range 200 – 800nm was recorded for 1mM Fe(III) Cl₃, Fe(III) Cl₃ + 1mM 3-dehydroquinone, Fe(III) Cl₃ + 1mM 3-dehydroshikimate and Fe(III) Cl₃ + 1mM PCA (Figure 9.3, Figure 9.4)

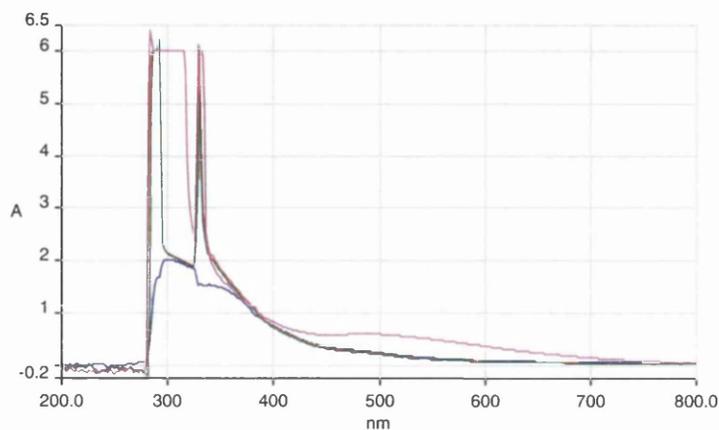


Figure 9.3: Absorbance spectra in the range 200 – 800nm for 1mM Fe(III) Cl₃ (Blue), Fe(III) Cl₃ + 1mM 3-dehydroquinone (Red), Fe(III) Cl₃ + 1mM 3-dehydroshikimate (Green) and Fe(III) Cl₃ + 1mM PCA (Pink)

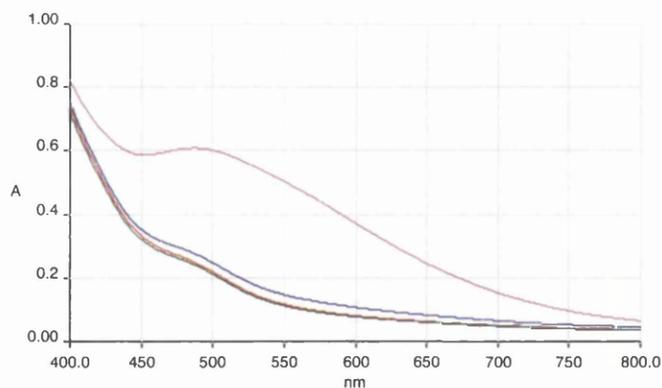


Figure 9.4: Expansion of absorbance spectra in the range 400 – 800nm for 1mM Fe(III) Cl₃ (Blue), Fe(III) Cl₃ + 1mM 3-dehydroquinone (Red), Fe(III) Cl₃ + 1mM 3-dehydroshikimate (Green) and Fe(III) Cl₃ + 1mM PCA (Pink)

This experiment shows that Fe(III) Cl₃ + PCA absorbs light in the range 460 – 600nm

9.5 Optimisation of Fe(III) Cl₃ /PCA concentrations required to detect PCA in

96 – well plate format

The previous experiment to show the colour change upon the formation of the Fe(III) Cl₃ /PCA complex was carried out using 1cm path length quartz cuvettes. In this experiment the assay was transferred to 96-well plate format, using a range of Fe(III) Cl₃ and PCA concentrations. The range of concentrations used were, for Fe(III) Cl₃ 1μM – 5mM, and for PCA 0.05μM – 5mM (Figure 8.5). The absorbance was measured at 590nm, and the results plotted using Origin 5.0 (Figure 9.6).

From Figure 8.6 it can be seen that using 5mM Fe(III) Cl₃, 50μM is the minimum concentration of PCA that can be detected as an increase in A₅₉₀. Using 1mM Fe(III) Cl₃, 100μM PCA may be the limit of detection, although there is a spurious result in the experiment at that value.

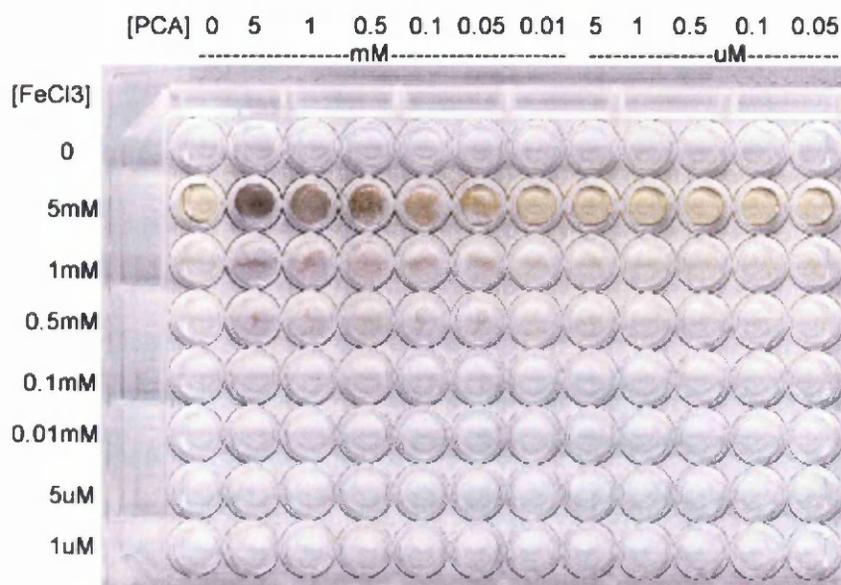


Figure 9.5: 96-well plate illustrating the colour change observed upon addition of the Fe(III) Cl_3 to PCA at various concentrations

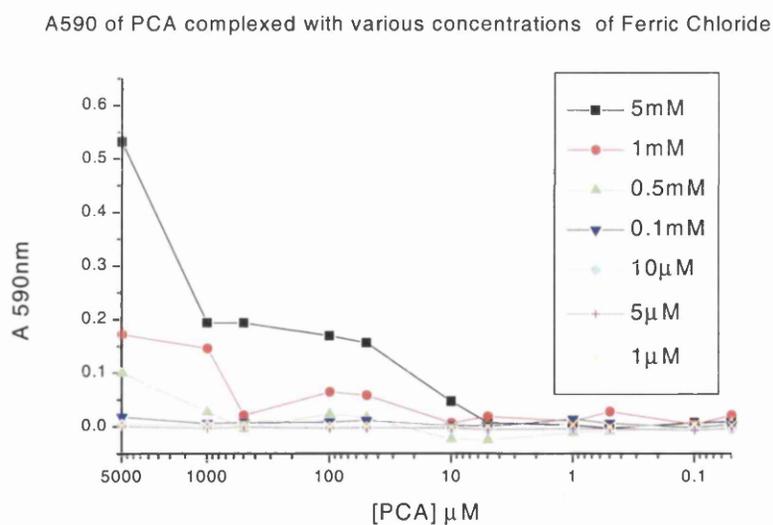


Figure 9.6: Plot of A_{590} vs PCA concentration using various concentrations of Fe(III) Cl_3

9.6 Detection of PCA using Na_2MoO_3 as an indicator

In the presence of vicinal di- (or tri) hydroxy phenols, solutions of Na_2MoO_3 form complexes that are coloured dark yellow, absorbing light around 400 – 450nm. PCA is a vicinal di-hydroxy phenol; therefore the use of Na_2MoO_3 as an indicator for the presence of PCA was investigated (Figure 9.7, 9.8).

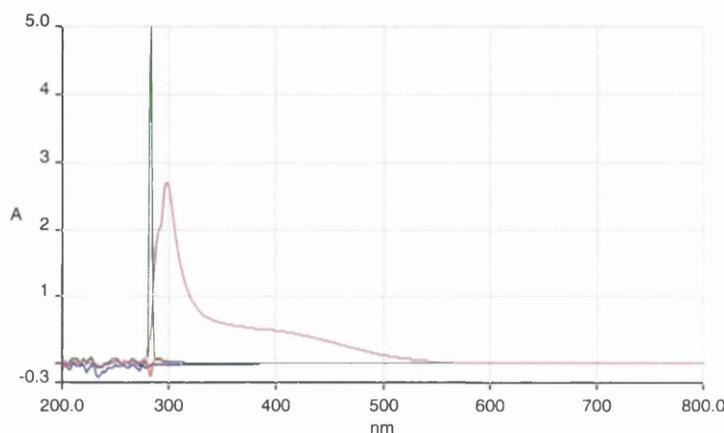


Figure 9.7: Absorbance spectra in the range 200 – 800nm for 1mM Na_2MoO_3 (Blue), Na_2MoO_3 + 1mM 3-dehydroquininate (Red), Na_2MoO_3 + 1mM 3-dehydroshikimate (Green) and Na_2MoO_3 + 1mM PCA (Pink)

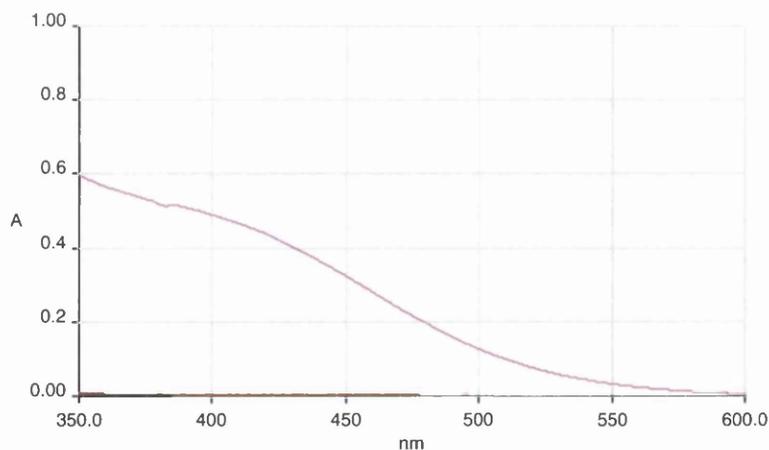


Figure 9.8: Expansion of absorbance spectra in the range 350 – 600nm for 1mM Na_2MoO_3 (Blue), Na_2MoO_3 + 1mM 3-dehydroquininate (Red), Na_2MoO_3 + 1mM 3-dehydroshikimate (Green) and Na_2MoO_3 + 1mM PCA (Pink)

Figures 9.7 and 9.8 show that PCA and Na_2MoO_3 form a complex that absorbs strongly in the range 350- 450nm.

9.7 Optimisation of Na_2MoO_3 /PCA concentrations required to detect PCA in 96 – well plate format

The previous experiment illustrated that using 1cm path length quartz cuvettes, PCA can be detected by the formation of a complex with Na_2MoO_3 that absorbs strongly between 350 and 450 nm. This assay was then transferred to 96-well plate format and the signal to noise ratio determined using various concentrations of PCA and Na_2MoO_3 . Figure 9.9 shows the colour change observed on the plate, and figure 9.10 shows plot of absorbance at 405nm as a function of PCA concentration, using different concentrations of Na_2MoO_3 .

From figure 9.10 it can be seen that there is a large absorbance at 405nm in the presence of 5mM PCA and 5mM Na_2MoO_3 , however this absorbance drops rapidly to noise level as the concentrations of PCA and Na_2MoO_3 are reduced. The minimal detectable levels of PCA are 0.5mM using 5mM Na_2MoO_3 as an indicator, and 1mM PCA using 1mM Na_2MoO_3 as an indicator.

This level of detection is not as sensitive as the classical Fe(III) Cl_3 method.

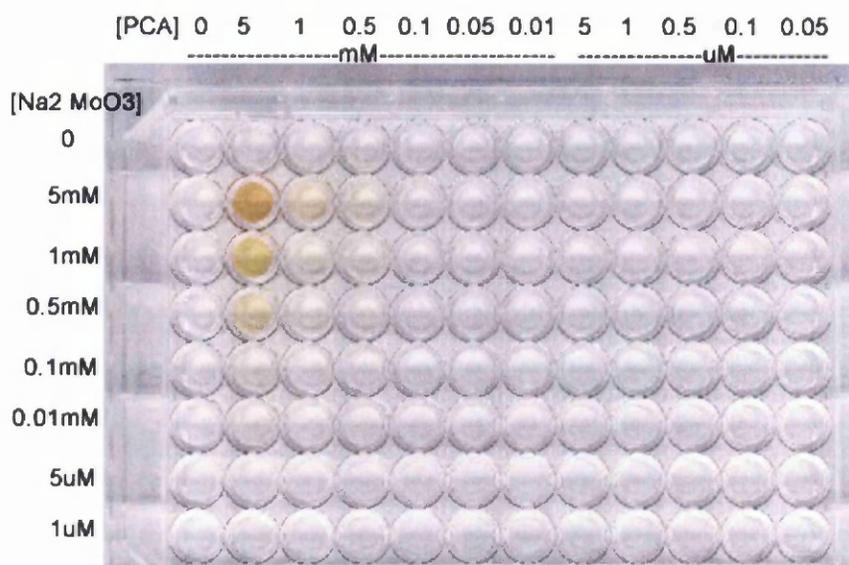


Figure 9.9: 96-well plate illustrating the colour change observed upon addition of $\text{Na}_2 \text{MoO}_3$ to PCA at various concentrations

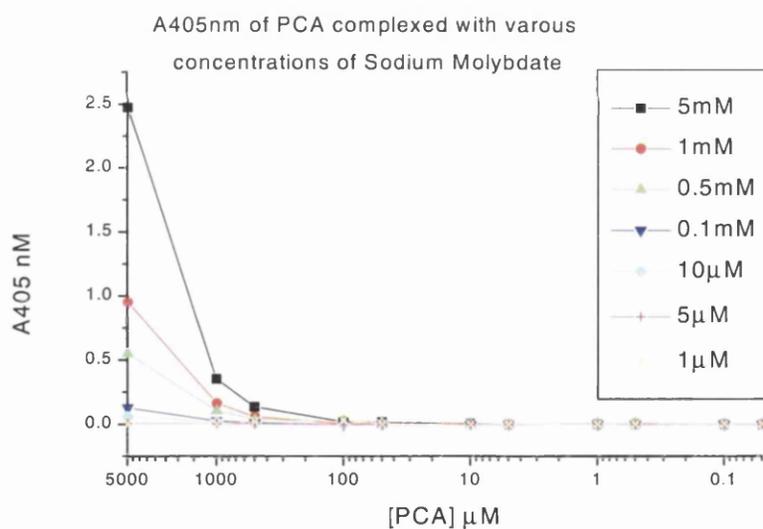


Figure 9.10: Plot of A_{405} vs PCA concentration using various concentrations of $\text{Na}_2 \text{MoO}_3$

9.8 Cloning of Dehydroshikimate dehydratase

A purification method for purification of DHSase for *N. crassa* has been established previously (Stroman *et al.* 1978), however, large-scale growth of the fungi is problematic, and the 7 step purification very long. Elsemore and Ornston cloned a 2.0kb *Hind III* fragment of *A. calcoaceticus* DNA that included the *quiC* gene encoding DHSase into the p-Bluescript vector pBSK (Stratagene). DHSase activity was detected in *E.coli* DH5 α cells expressing the resultant plasmid pZR564 (Elsemore and Ornston, 1995).

A colony of *E coli* DH5 α cells containing plasmid pZR564 was obtained from Prof. L.N. Ornston (Dept. of Biology, Yale University, New Haven, Connecticut) and the plasmid propagated and purified. To ease expression and purification of DHSase, the *quiC* gene was sub-cloned into pET-15b (Novagen) to express the protein with an N-terminal 6xHis tag. The restriction sites present in the pBSK vector were incompatible with the *Nde I* and *Bam HI* sites in pET-15b, therefore PCR was required to add appropriate restriction sites onto the *quiC* gene.

9.9.1 Amplification of *quiC* by PCR

Oligonucleotide primers were designed to amplify the 1.49kb *quiC* gene with the appropriate 5' *Nde I* and 3' *Bam HI* restriction sites for cloning into pET-15b. PCR was carried out as outlined in section 2.6 using *Taq* DNA polymerase (Invitrogen), and pZR564 as template DNA. The results were analysed by gel electrophoresis (Figure 9.11) and PCR products of the correct size purified by gel electrophoresis.

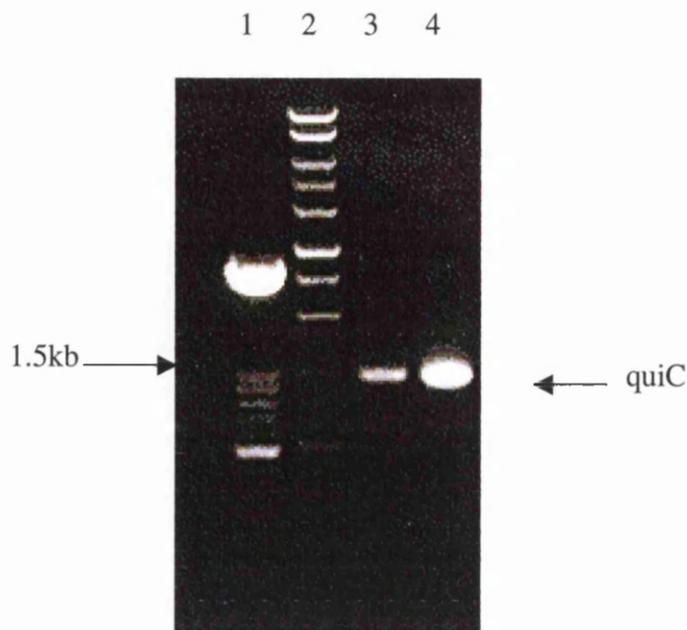


Figure 9.11: Gel electrophoresis of *quiC* PCR; 1, Boehringer XIV marker 2, Promega 1kb ladder 3, PCR product 1 μ l 4, PCR product 5 μ l

9.9.2 TOPO[®] cloning of *quiC*

To allow the PCR product to be sequenced, and to ease the digestion of the *quiC* gene by restriction endonucleases, the PCR product was inserted into the pCR[®]2.1 TOPO[®] vector using the TOPO TA kit (section 2.7.2) to generate the vector pCR*quiC*. The vector pCR*quiC* was transformed into chemically competent *E. coli* DH5 α and plated out on the LB-Agar + 100 μ g/ml ampicillin plates. After incubation overnight at 37 $^{\circ}$ C, 10 colonies had grown on the LB-Agar + 100 μ g/ml ampicillin plates. Each colony was picked from the plate and used to inoculate 2ml LB + 100 μ g/ml ampicillin which was incubated overnight at 37 $^{\circ}$ C with shaking. Plasmid DNA was extracted from the 2ml cultures using the Qiaquick Spin kit (Qiagen)

Positive clones were analysed by digestion with the restriction endonuclease *Eco RI*. The pCR[®]2.1 TOPO[®] vector has two *Eco RI* sites, one upstream of the

cloning and site and another downstream of the cloning site. When the vector is digested with *Eco RI*, two bands, one corresponding to the length of the vector and the other corresponding to the length of the cloned sequence, can be separated by gel electrophoresis. All 10 clones were shown to contain an insert sequence of the correct length by this method. One positive clone, pCRquiC1 was sequenced to check for errors resulting from the PCR process and it was shown to contain no errors. This clone was used for all resulting manipulations.

9.9.3 Subcloning of *quiC* into pET-15b

The expression vector pET-15b had been used to successfully over-express type II DHQase, therefore it was chosen for the over-expression of DHSase.

The vectors pCRquiC1 and HpET were propagated using *E.coli* DH5 α in 5ml cultures of LB + the appropriate antibiotic and plasmid DNA isolated using the Qiaquick Spin kit (Qiagen). 30 μ l of each vector was double digested using the restriction endonucleases *NdeI* and *BamHI* (Promega) in Promega buffer D. Buffer D was chosen as it is the optimal buffer for *NdeI* activity and *BamHI* has 75% activity in this buffer system. *NdeI* has little or no activity in other buffer systems. The resulting DNA fragments were analysed by gel electrophoresis, and bands corresponding to pET-15b and *quiC* were extracted using the Quiex II kit (Qiagen).

Ligation of the pET-15b and *quiC* fragments was carried out overnight at 16 $^{\circ}$ C using T4 DNA Ligase (Promega) as outlined in section 2.5.4 to generate the vector pET*quiC*. The ligation products were transformed into chemically competent *E. coli* DH5 α cells and plated out on LB-Agar + 100 μ g/ml ampicillin plates. The plates

were incubated overnight at 37°C, after which time 25 colonies were visible. 10 colonies were picked and grown overnight with shaking at 37°C in 2ml LB + 100µg/ml ampicillin and vector DNA isolated using the Qiaquick Spin kit. Clones were analysed by restriction digests with *NdeI* and *BamHI* and all 10 clones were shown to be positive by gel electrophoresis.

9.9.4 Expression of *QuiC* in *E. coli* BL21

The vector pET*quiC* was transformed into chemically competent *E. coli* BL21 (pLysS S) and plated out on LB-Agar + 100µg/ml ampicillin plates. The plates were incubated overnight at 37°C. A single colony was picked from the plate and used to inoculate 5ml LB + 100µg/ml ampicillin which was incubated at 37°C with shaking overnight.

1ml of the overnight culture was removed; the cells were pelleted by centrifugation and stored on ice for future analysis by SDS PAGE. 2ml samples were used to inoculate 2 x 25ml LB + 100µg/ml ampicillin cultures which were grown with shaking at 37°C until the OD₆₀₀ was between 0.4 and 0.6. IPTG to a final concentration of 1mM was added to one culture and water added to the other culture to act as a control. Both cultures were incubated with shaking at 37°C for 3 hours.

After 3 hours 1 ml samples were removed from each culture and the cells pelleted by centrifugation and stored on ice. The cell pellet from the induced culture was re-suspended in 500µl of cytobuster reagent (Invitrogen) and incubated at room temperature for 30 minutes to lyse the cells. A 20 µl aliquot was removed for SDS PAGE analysis before the remainder of the sample was centrifuged (13000 rpm, 5 minutes) to separate the soluble and insoluble components. The soluble fraction was removed and both soluble and insoluble fractions were stored on ice.

The remainder of the 25ml sample induced with IPTG was centrifuged to isolate the cell pellet (4000g, 20 minutes). The cell pellet was then resuspended in 20mM Tris-HCl pH 8.0, 300mM NaCl, 20mM imidazole, 2.5mM MgCl₂. The cells were lysed by passage through a French press (2 x 950 psi) and solid debris removed by centrifugation (8000g, 60 minutes). The supernatant was loaded onto a pre-equilibrated Ni-NTA superflow column (Qiagen) and then flow through collected. The column was washed with 20mM Tris-HCl pH 8.0, 300mM NaCl, 75mM imidazole, 2.5mM MgCl₂ and the flow through collected. Bound protein was eluted by washing the column with 20mM Tris-HCl pH 8.0, 300mM NaCl, 200mM imidazole, 2.5mM MgCl₂ and the peak fraction collected.

9.9.5 Analysis of *QuiC* expression by SDS PAGE

The following samples were loaded onto a 12.5% SDS polyacrylamide gel for analysis; 1- Low molecular weight marker, 2 - overnight culture total cell protein, 3 - un-induced sample total cell protein, 4 - induced sample total cell protein, 5 - induced sample insoluble fraction, 6 - induced sample soluble fraction, 7 – Ni-NTA flow through, 8 – Ni-NTA wash flow through, 9 – Ni-NTA eluted fraction. The stained gel is shown in Figure 9.12

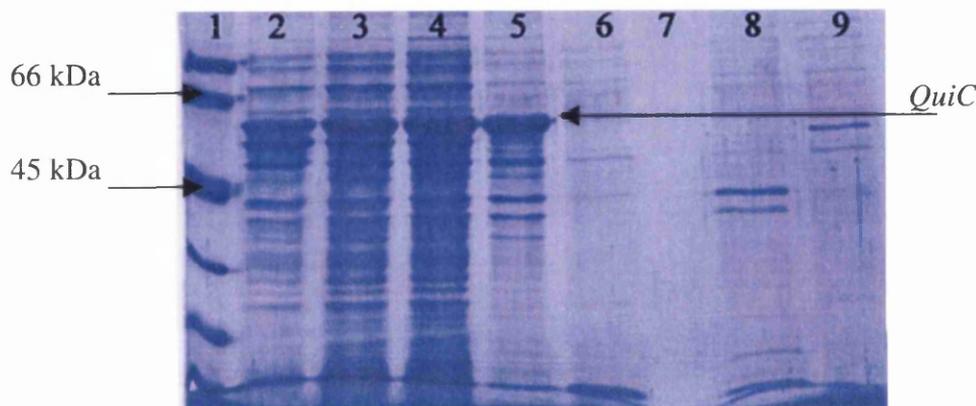


Figure 9.12: SDS PAGE analysis of *QuiC* expression.

The SDS PAGE analysis shows that *QuiC* is heavily over-expressed in *E. coli* BL21 cells, even without the addition of IPTG, however most of the expressed protein appears in the insoluble fraction. A small band appears to be enriched in the eluant from the Ni-NTA column, however if this is indeed a small amount of soluble *QuiC*, the yield is very low.

Stroman *et al.* (1978) showed that the DHSase from *N. crassa* was thermally unstable and required Mg^{2+} ions to stabilise it in solution. Expression of DHSase at lower temperature (20°C) may increase the yield of soluble protein. A similar experiment varying the concentration of Mg^{2+} ions present in the media and purification buffers may also result in production of soluble protein.

9.10 Conclusion

There is a need for a DHQase activity assay that is amenable for high throughput screening methods is great. Coupling DHQase activity to DHSase to generate protocatechuate allows easy detection by a strong colour change in the presence of Fe(III) ions. The intensity of the colour change allows detection of sub mM quantities of PCA in 96 well plate format.

The gene for the DHSase for *A. calcoaceticus* was obtained from Prof L.N. Ornston (Yale University, New Haven, CT) and subcloned into the pET-15b vector for over expression in *E. coli*. The protein produced from this expression system was shown to be insoluble, however modification of the expression protocol and buffer contents may yield soluble protein, allowing full development of this assay.

The use of coupling the activity of dehydroquinase and DHQase to DHSase, generating concentrations PCA detectable using Fe(III) ions has recently been outlined in a recent patent filed by Arrow Therapeutics (US patent 6,406,878).

Chapter 10 – Conclusions

The rise in the incidence of the human disease tuberculosis (TB) caused by infection with the pathogen *Mycobacterium tuberculosis* is extremely worrying. This fact coupled with the rise in strains of *M. tuberculosis* that are resistant to at least one of the current frontline anti-tubercular drugs has hastened the need for novel TB drugs. The Action TB initiative, funded by GlaxoSmithKline, and the Global Alliance for TB drug development (<http://www.tballiance.org>) (Global Alliance for TB drug development, 2001) are committed to research and development to identify novel drugs active against *M. tuberculosis*. Enzymes of the shikimate pathway have become attractive drug targets and the use of structural information upon potential drug targets will be of prime importance in the development of novel drugs. This is reflected in the recent structure of TB shikimate kinase (Gu *et al.*, 2002) and the large number of groups who have and signed up to study the enzymes of the pathway in the TB structural genomics consortium (<http://www.doe-mbi.ucla.edu/TB/>). In addition, structures of almost all the enzymes of the pathway have now been solved including structures of shikimate dehydrogenase and related iso-enzymes by non-profit organisations, structural genomics organisations and commercial organisations (Michel *et al.*, 2003; Benach *et al.*, 2003; Ye *et al.*, 2003; Padyana and Burley, 2003; commentary in Vogan, 2003).

The primary result from the studies presented in this thesis is the identification of a novel, non substrate-like inhibitor of type II 3-dehydroquinase from *M. tuberculosis* (MTDQase) by structure based ligand design methods. After analysis of the available structures of MTDQase determined by X-ray crystallography, pharmacophore models were generated which describe potential electrostatic, steric and hydrophobic

interactions within the active site. These models were used to screen a virtual compound library *in silico* to generate a list of compounds that potentially satisfy the criteria of the pharmacophore models, hence should inhibit MTDHQase *in vitro*. Similar screens were carried out against the same database using pharmacophores derived from crystal structures of type II 3-dehydroquinases from *Streptomyces coelicolor* (SCDHQase) and *Helicobacter pylori* (HPDHQase). These searches generated 36 hits for compounds that were unique hits for MTDHQase and the testing of these compounds in an *in vitro* assay resulted in the identification of 6 compounds that showed inhibitory activity.

Future study upon these compounds should involve characterisation of their inhibition profiles against other type II DHQases. The structure determination of MTDHQase in complex with these compounds will progress this work to the next stage, using structural information to optimise ligand potency.

This example of structure based ligand design was carried out on a small scale yet active compounds were identified illustrating proof of concept of the experiment design. Currently, structural information is highly sought after for drug targets in the pharmaceutical industry with most major pharmaceutical and biotech companies either having teams of structural biologists and computational chemists, or have contracts with companies that specialise in protein structure determination.

The major advantage of using a structure based approach was to search for non-substrate-like inhibitors that are more 'drug-like'. Most current inhibitors of type II 3-dehydroquinase are transition state analogues, hence are small, hydrophilic and extensively functionalised, hence not ideal for development as drugs.

The structure determination of HPDHQase in complex with the ligand AH9095 represents the first structure of a type II DHQase in complex with a non-substrate like inhibitor. AH9095 was identified by high throughput screening methods and was shown to be a specific HPDHQase inhibitor, showing no potency against MTDHQase. Comparison of the HPDHQase and MTDHQase structures suggested that the presence of a negatively charged patch on the surface of MTDHQase near the entrance to the active site may hinder the binding of ligands with large polar moieties such as the xanthene moiety of AH9095. This area of the active site has not been investigated in previous studies attempting to design specific type II DHQase inhibitors, therefore probing this area of the active site may provide an opportunity to build specificity into more potent generic type II DHQase inhibitors. A recent study by Toscano and co-workers (Toscano *et al.*, 2003) probed a hydrophilic pocket in the active site of SCDHQase (identified by a glycerol fortuitously bound to the structure) with only limited success.

The recent publication of HPDHQase at 2.5Å resolution will help to provide further insight towards generation of novel inhibitors of type II DHQase (Lee *et al.*, 2003 PDB I.D. 1J2Y). After re-assessment of the experimental data submitted to the PDB, it is apparent that the electron density observed within the active site corresponds to a molecule of citrate, which they report to have used as precipitant in the crystallisation media (Kwak *et al.*, 2001) as opposed to the reported ligand, the substrate 3-dehydroquininate. Hence there are three structures of HPDHQase available with three different types of ligand; one rational design (FA1), one identified from HTS (AH9095) and one identified via serendipity (citrate). Citrate appears to bind tightly in the active

site of HPDHQase, therefore full investigation of inhibition of type II DHQases by citrate and related molecules should be carried out.

This study has shown that novel enzyme inhibitors can be identified by the structure based ligand design methodologies. Further analysis of these, and related compounds may yield inhibitors that are more potent and selective than the current first generation inhibitors. In addition novel non substrate like inhibitors identified may have better solubility profiles necessary for further drug development leading to a successful therapeutic agent.

References

- CCP4. The CCP4 Suite - Programs for Protein Crystallography. *Acta Crystallographica Section D-Biological Crystallography* (1994),**50**:760-763.
- Baillie AC, Corbett JR, Dowsett JR, McCloskey P. Inhibitors of shikimate dehydrogenase as potential herbicides. *Pesticide Science* (1972),**3**:113-120.
- Beddell CR, Goodford PJ, Norrington FE. Compounds designed to fit a site of known structure in human haemoglobin. *Br. J. Pharmacol.* (1976),**57**:201-209.
- Bello CG, Harris JM, Manthey MK, Coggins JR, Abell C. Irreversible inhibition of type I dehydroquinase by substrates for type II dehydroquinase. *Bioorg Med Chem Lett* (2000),**10**(5):407-9.
- Bello CG, Manthey MK, Harris JM, Hawkins A, Coggins JR, Abell C. Synthesis of 2-Bromo- and 2-Fluoro-3-dehydroshikimate Acids and 2-Bromo- and 2-Fluoroshikimic Acids using synthetic and enzymatic approaches. *Journal of Organic Chemistry* (1998),**63**:1591 - 1597.
- Benach J, Lee I, Edstrom W, Kuzin AP, Chiang Y, Acton TB, et al. The 2.3A Crystal structure of the shikimate 5-dehydrogenase orthologue YdiB from *Escherichia coli* suggests a novel catalytic environment for an NAD-dependent dehydrogenase. *J Biol Chem* (2003),**278**:19176-19182.
- Bentley R. The shikimate pathway--a metabolic tree with many branches. *Crit Rev Biochem Mol Biol* (1990),**25**(5):307-84.
- Bergfors TM. Protein Crystallisation: International University Line; 1999.
- Beri RK, Grant S, Roberts CF, Smith M, Hawkins AR. Selective overexpression of the QUTE gene encoding catabolic 3- dehydroquinase in multicopy transformants of *Aspergillus nidulans*. *Biochem J* (1990),**265**(2):337-42.

-
- Bernal JD, Crowfoot D. X-ray photographs of crystalline pepsin. *Nature* (1934),**133**:794-795.
- Bishai W. Lipid lunch for persistent pathogen. *Nature* (2000),**406**(6797):683-5.
- Blundell TL. Structure-based Drug Design. *Nature* (1996),**384**:S23-S26.
- Blundell TL, Jhoti H, Abell C. High-Throughput Crystallography for Lead Discovery in Drug Design. *Nature Reviews Drug Discovery* (2001),**1**:45-54.
- Blundell TL, Johnson LN. Protein Crystallography. London: Academic Press; 1976.
- Bottomley JR, Clayton CL, Chalk PA, Kleanthous C. Cloning, sequencing, expression, purification and preliminary characterization of a type II dehydroquinase from *Helicobacter pylori*. *Biochem J* (1996),**319**(Pt 2):559-65.
- Bowe EF, O'Gaora P, Maskell D, Cafferkey M, Dougan G. Virulence, persistence and immunogenicity of *Yersinia enterocolitica* O:8 aroA mutants. *Infection and Immunity* (1989),**57**:3234-3236.
- Brunger AT. Free R-Value - a Novel Statistical Quantity for Assessing the Accuracy of Crystal-Structures. *Nature* (1992),**355**(6359):472-475.
- Butler JR, Alworth WL, Nugent MJ. Mechanism of Dehydroquinase Catalyzed Dehydration I. Formation of a Schiff Base Intermediate. *Journal of the American Chemical Society* (1974),**95**(5):1617-1618.
- Carpenter EP, Hawkins AR, Frost JW, Brown KA. Structure of dehydroquinase synthase reveals an active site capable of multistep catalysis. *Nature* (1998),**394**(6690):299-302.
- Carson M. Ribbons 2.0. *Journal of Applied Crystallography* (1991),**24**:958-&.

Case ME, Giles NH. Gene order in the qa gene cluster of *Neurospora crassa*. *Mol Gen Genet* (1976),**147**(1):83-9.

Chaudhuri S, Duncan K, Graham LD, Coggins JR. Identification of the active-site lysine residues of two biosynthetic 3- dehydroquinases. *Biochem J* (1991),**275**(Pt 1):1-6.

Coggins JR, Boocock MR, Chaudhuri S, Lambert JM, Lumsden J, Nimmo GA, et al. The arom multifunctional enzyme from *Neurospora crassa*. *Methods Enzymol* (1987),**142**:325-41.

Cole ST. Rifamycin resistance in mycobacteria. *Res Microbiol* (1996),**147**:48-52.

Cole ST, Brosch R, Parkhill J, al e, Barel BG. Dechiphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence. *Nature* (1998),**393**:537-544.

Cowtan K. DM: An automated procedure for phase improvement by density modification. Joint CCP4 and ESF-EACBM Newsletter on Protein Crystallography 1994:34-38.

Crowther RA. In: Rossmann MG, editor. The Molecular Replacement Method. New York: Grodon & Breach; 1972. p. 173-178.

Cruickshank DWJ. Remarks about protein structure precision. *Acta Crystallographica Section D-Biological Crystallography* (1999),**55**:583-601.

Cushman DW, Cheung HS, Sabo EF, Ondetti MA. Design of potent competitive inhibitors of angiotensin-converting enzyme. Carboxyalkanoyl and mercaptoalkanoyl amino acids. *Biochemistry* (1977),**16**:5484-5491.

Dauter Z. Data-collection strategies. *Acta Crystallogr D Biol Crystallogr* (1999),**55**(10):1703-1717.

Davies GM, Barrett-Bee KJ, Jude DA, Lehan M, Nichols WW, Pinder PE, et al. (6S)-6-fluoroshikimic acid, an antibacterial agent acting on the aromatic biosynthetic pathway. *Antimicrob Agents Chemother* (1994),**38**(2):403-6.

Deka RK, Anton IA, Dunbar B, Coggins JR. The characterisation of the shikimate pathway enzyme dehydroquinase from *Pisum sativum*. *FEBS Lett* (1994),**349**(3):397-402.

Deka RK, Kleanthous C, Coggins JR. Identification of the essential histidine residue at the active site of *Escherichia coli* dehydroquinase. *J Biol Chem* (1992),**267**(31):22237-42.

DeLano WL. The PyMOL Molecular Graphics System. San Carlos, CA, USA: DeLano Scientific; 2002.

DINO: Visualizing Structural Biology (2002), <http://www.dino3d.org>

Drenth J. Principles of Protein X-ray Crystallography. 2nd ed. New York: Springer-Verlag; 1999.

Ducruix A, Giege R. Crystallization of Nucleic Acids and Proteins: A Practical Approach: Oxford University Press; 1992.

Duncan K, Chaudhuri S, Campbell MS, Coggins JR. The overexpression and complete amino acid sequence of *Escherichia coli* 3-dehydroquinase. *Biochem J* (1986),**238**(2):475-83.

Dye C, Williams BG, Espinal MA, Raviglione MC. Erasing the World's Slow Stain: Strategies to beat Multidrug-resistant Tuberculosis. *Science* (2002),**295**:2042-2046.

Elsemore DA, Ornston LN. Unusual ancestry of dehydratases associated with quinate catabolism in *Acinetobacter calcoaceticus*. *J Bacteriol* (1995),**177**(20):5971-8.

- Evans LDB, Roszak AW, Noble LN, Robinson DA, Matthews JL, Chalk PA, et al. Specificity of substrate recognition by type II dehydroquinases as revealed by the binding of polyanions. *FEBS Lett* (2002),**530**:24-30.
- Evans SV. Setor - Hardware-Lighted 3-Dimensional Solid Model Representations of Macromolecules. *Journal of Molecular Graphics* (1993),**11**(2):134 et seq.
- Frederickson M, Coggins JR, Abell C. Vinyl fluoride as an isoelectric replacement for an enolate ion: Inhibition of type II dehydroquinase. *Chemical Communications* (2002):1886-1887.
- Frederickson M, Parker EJ, Hawkins A, Coggins JR, Abell C. Selective inhibition of type II Dehydroquinases. *Journal of Organic Chemistry* (1999),**64**:2612 - 26134.
- Garbe T, Servos S, Hawkins A, Dimitriadis G, Young D, Dougan G, et al. The *Mycobacterium tuberculosis* shikimate pathway genes: evolutionary relationship between biosynthetic and catabolic 3-dehydroquinases. *Mol Gen Genet* (1991),**228**(3):385-92.
- Global Alliance for TB Drug Development . Tuberculosis. Scientific blueprint for tuberculosis drug development. *Tuberculosis (Edinb)* (2001),**81**(Suppl 1):1-52.
- Gourley DG, Shrive AK, Polikarpov I, Krell T, Coggins JR, Hawkins AR, et al. The two types of 3-dehydroquinase have distinct structures but catalyze the same overall reaction. *Nat Struct Biol* (1999),**6**(6):521-5.
- Grant S, Roberts CF, Lamb H, Stout M, Hawkins AR. Genetic regulation of the quinic acid utilization (QUT) gene cluster in *Aspergillus nidulans*. *J Gen Microbiol* (1988),**134**(Pt 2):347-58.
- Greer J, Erickson JW, Baldwin JJ, Varney MD. Application of the three-dimensional structure of protein target molecules in structure based drug design. *Journal of Medicinal Chemistry* (1994),**37**:1035-1054.

- Gu Y, Reshetnikova L, Li Y, Wu Y, Yan H, Singh S, Ji X. Crystal structure of shikimate kinase from *Mycobacterium tuberculosis* reveals the dynamic role of the LID domain in catalysis. *J. Mol. Biol.* (2002), **319**: 779-789
- Hanahan D. DNA Cloning: Oxford IRL Press; 1985.
- Harris JM, Kleanthous C, Coggins JR, Hawkins A, Abell C. Different mechanistic and stereochemical courses for the reactions catalysed by type I and type II Dehydroquinases. *Journal of the Chemical Society., Chemical Communications* (1993),**1993**:1080 - 1081.
- Harris JM, Gonzalez-Bello C, Kleanthous C, Hawkins AR, Coggins JR, Abell C. Evidence from kinetic isotope studies for an enolate intermediate in the mechanism of type II dehydroquinases. *Biochem J* (1996a),**319**(Pt 2):333-6.
- Harris JM, Watkins WJ, Hawkins A, Coggins JR, Abell C. Comparison of the substrate specificity of type I and type II dehydroquinases with 5-deoxy and 4,5-dideoxy-dehydroquinic acid. *Journal of the Chemical Society., Perkin Transactions* (1996b),**1996**:2371 - 2377.
- Haslam E. Shikimic acid: Metabolism and Metabolites. Chichester. UK: J. Wiley and Sons.; 1993.
- Hoiseth SK, Stocker BA. Aromatic-dependent *Salmonella typhimurium* are non-virulent and effective live vaccines. *Nature* (1981),**241**:238-239.
- Ivins BE, Welkos SL, Knudson GB, Little SF. Immunization against anthrax with aromatic compound-dependant (Aro-) mutants of *Bacillus anthracis* and with recombinant strains of *Bacillus subtilis* that produce anthrax protective antigen. *Infection and Immunity* (1990),**58**:303-308.
- Jancarik J, Kim SH. Sparse-Matrix Sampling - a Screening Method for Crystallization of Proteins. *Journal of Applied Crystallography* (1991),**24**:409-411.

Jaworski EG. Mode of action of N-phosphomethyl glycine: inhibitor of aromatic amino acid biosynthesis. *J. Agric. Food Chem.* (1972),**20**:1195-1198.

Kapust RB, D.S. W. *Escherichia coli* maltose-binding protein is uncommonly effective at promoting the solubility of polypeptides to which it is fused. *Protein Sci* (1999),**8**:1668-1674.

Kendrew JC, Bodo G, Dintzis HM, Parrish RG, Wyckoff H, Phillips DC. A three-dimensional model of the myoglobin molecule obtained by X-ray analysis. *Nature* (1958),**181**:662-666.

Kissinger CR, Gehlhaar DK, Fogel DB. Rapid automated molecular replacement by evolutionary search. *Acta Crystallographica Section D-Biological Crystallography* (1999),**55**:484-491.

Kleanthous C, Deka R, Davis K, Kelly SM, Cooper A, Harding SE, et al. A comparison of the enzymological and biophysical properties of two distinct classes of dehydroquinase enzymes. *Biochem J* (1992),**282**(Pt 3):687-95.

Klebe G. Recent developments in structure-based drug design. *Journal of Molecular Medicine* (2000),**78**:269-281.

Kleywegt GJ, Jones TA. Homo crystallographicus - Quo vadis? *Structure* (2002),**10**:465-472.

Krell T, Coggins JR, Laphorn AJ. The three-dimensional structure of shikimate kinase. *J Mol Biol* (1998),**278**(5):983-97.

Krell T, Horsburgh MJ, Cooper A, Kelly SM, Coggins JR. Localization of the active site of type II dehydroquinases. Identification of a common arginine-containing motif in the two classes of dehydroquinases. *J Biol Chem* (1996),**271**(40):24492-7.

- Kuo MR, Morbidoni HR, Alland D, Sneddon SF, Gourlie BB, Staveski MM, et al. Targeting tuberculosis and malaria through inhibition of enoyl reductase: compound activity and structural data. *J Biol Chem* (2003),**278**:20851-20859.
- Kwak JE, Lee JY, Han BW, Moon JJ, Sohn SH, Suh SW. Crystallization and preliminary X-ray crystallographic analysis of type II dehydroquinase from *Helicobacter pylori*. *Acta Crystallogr D Biol Crystallogr* (2001),**57**(Pt 2):279-80.
- Lamzin VS, Wilson KS. Automated Refinement of Protein Models. *Acta Crystallographica Section D-Biological Crystallography* (1993),**49**(Pt1):129-147.
- Laskowski RA, Macarthur MW, Moss DS, Thornton JM. Procheck - a Program to Check the Stereochemical Quality of Protein Structures. *Journal of Applied Crystallography* (1993),**26**(Pt2):283-291.
- Lee BI, Kwak JE, Suh SW. Crystal structure of the type II 3-dehydroquinase from *Helicobacter pylori*. *Proteins: Structure, Function and Genetics* (2003),**51**:616-617.
- Lee RE, Mikusova K, Brennan PJ, Besra GS. Synthesis of the mycobacterial arabinose donor beta-D-arabinofuranosyl-1-monophosphoryl-deccaprenol, development of a basic arabinosyl-transferase assay, and identification of ethambutol as an arbinose-transferase inhibitor. *Journal of the American Chemical Society* (1995),**117**:11829-11832.
- Leech AP, Boetzel R, McDonald C, Shrive AK, Moore GR, Coggins JR, et al. Re-evaluating the role of His-143 in the mechanism of type I dehydroquinase from *Escherichia coli* using two-dimensional ¹H,¹³C NMR. *J Biol Chem* (1998),**273**(16):9602-7.
- Leech AP, James R, Coggins JR, Kleanthous C. Mutagenesis of active site residues in type I dehydroquinase from *Escherichia coli*. Stalled catalysis in a histidine to alanine mutant. *J Biol Chem* (1995),**270**(43):25827-36.

Lipinski CA, Lombardo F, Dominy BW, Feeney PJ. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Advanced Drug Delivery Reviews* (1997),**23**:3-25.

Lipscomb WN, Hartsuch JA, Reche GN, Quioco FA, Bethge PH, Ludwig ML, et al. The structure of carboxypeptidase A. VII. The 2Å resolution studies of the enzyme and of its complex with glycylytyrosine and mechanistic deductions. *Brookhaven Symp Biol* (1968),**1**:24-90.

Madge D, Wishart G, Dolaman M, Maunder P, inventors; Arrow Therapeutics Ltd., assignee. Bissulfonamides as inhibitors of the dehydroquinase synthase enzyme AroB and the Type II dehydroquinase enzyme AroQ. UK. 2001.

Maina CV, Riggs PD, Grandea AG, Slatko BE, Moran LS, Tagliamonte JA, et al. An *Escherichia coli* vector to express and purify foreign proteins by fusion to and separation from maltose-binding protein. *Gene* (1988),**74**:365-373.

Matthews BW. Solvent content of protein crystals. *J Mol Biol* (1968),**33**:491-497.

McKinney JD, Honer zu Bentrup K, Munoz-Elias EJ, Miczak A, Chen B, Chan WT, et al. Persistence of *Mycobacterium tuberculosis* in macrophages and mice requires the glyoxylate shunt enzyme isocitrate lyase. *Nature* (2000),**406**(6797):735-8.

McPherson A. Crystallization of Biological Macromolecules: Cold Spring Harbor Laboratory; 1999.

McRee DE, David PR. Practical Protein Crystallography. 2nd ed. New York: Academic Press; 1999.

Michel G, Roszak AW, Sauve V, Maclean J, Matte A, Coggins JR, et al. Structures of Shikimate Dehydrogenase AroE and its paralog YdiB. *J Biol Chem* (2003),**278**:19463-19472.

Montchamp JL, Frost JW. Cyclohexenyl and cyclohexylidene inhibitors of 3-dehydroquinase synthase. Active site interactions relevant to enzyme mechanism and inhibitor design. *Journal of the American Chemical Society* (1997),**119**:7645-7653.

Murshudov GN, Vagin AA, Dodson EJ. Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallographica Section D-Biological Crystallography* (1997),**53**(Pt3):240-255.

Murshudov GN, Vagin AA, Lebedev A, Wilson KS, Dodson EJ. Efficient anisotropic refinement of macromolecular structures using FFT. *Acta Crystallographica Section D-Biological Crystallography* (1999),**55**(Pt1):247-255.

Natesh R, Schwager SLU, Sturrock ED, Acharya KG. Crystal structure of human angiotensin-converting enzyme lisinoprol complex. *Nature* (2003),**421**:551-554.

Navaza J. AMoRe - an Automated Package for Molecular Replacement. *Acta Crystallographica Section A* (1994),**50**:157-163.

Nicholls A, Sharp KA, Honig B. Protein Folding and Association - Insights from the Interfacial and Thermodynamic Properties of Hydrocarbons. *Proteins-Structure Function and Genetics* (1991),**11**(4):281-296.

Nourelidin NA, Caldwell M, Hendry J, Lee DG. Heterogeneous permanganate oxidation of thiols. *Synthesis* (1998):1587-1589.

Otwinowski Z, Minor W. Processing of X-ray diffraction data collected in oscillation mode. *Methods in Enzymology* (1997),**276**:307-326.

Padyana AK, Burley SK. Crystal structure of shikimate 5-dehydrogenase (SDH) bound to NADP: insights into function and evolution. *Structure* (2003),**11**:1105-1113.

Pace CN, Vajdos F, Fee L, Grimsley G, Gray T. How to measure and predict the molar absorption coefficient of a protein. *Protein Science*, 1995, **4**:2411-2423.

- Parish T, Stoker NG. The common aromatic amino acid biosynthesis pathway is essential in *Mycobacterium tuberculosis*. *Microbiology* (2002),**158**:3069-3077.
- Parker EJ, Gonzalez Bello C, Coggins JR, Hawkins AR, Abell C. Mechanistic studies on type I and type II dehydroquinase with (6R)- and (6S)-6-fluoro-3-dehydroquinic acids. *Bioorg Med Chem Lett* (2000),**10**(3):231-4.
- Partridge CW, Case ME, Giles NH. Direct induction in wild-type *Neurospora crassa* of mutants (qa-1 c) constitutive for the catabolism of quinate and shikimate. *Genetics* (1972),**72**(3):411-7.
- Payne DJ, Wallis NG, Gentry DR, M R. The impact of genomics on novel antibacterial targets. *Current Opinion in Drug Discovery and Development* (2000),**3**:177 - 190.
- Poirier TP, Kehoe MA, Beachey EH. Protective immunity exoked by oral administration of attenuated *aroA Salmonella typhimurium* expressing cloned streptococcal M protein. *Journal of Expertimental Medicine* (1988),**168**:25-32.
- Price NC, Boam DJ, Kelly SM, Duncan D, Krell T, Gourley DG, et al. The folding and assembly of the dodecameric type II dehydroquinases. *Biochem J* (1999),**338**(Pt 1):195-202.
- Procter DJ. The synthesis of thiols, selenols, sulfides, selenides, sulfoxides, selenoxides, sulfones and selenones. *Journal of the Chemical Society., Perkin Transactions* (2001):335-354.
- Quemard A, Sacchettini JC, Dessen A, Vilcheze C, Bittman R, Jacobs WR, Jr., et al. Enzymatic characterization of the target for isoniazid in *Mycobacterium tuberculosis*. *Biochemistry* (1995),**34**(26):8235-41.
- Radzicka A, Wolfenden R. Transition state and multisubstrate analog inhibitors. *Methods Enzymol* (1995),**249**:284.

- Ratti E, Trist D. The continuing evolution of the drug discover process in the pharmaceutical industry. *Farmaco* (2001),**56**:13-19.
- Read RJ. Pushing the boundaries of molecular replacement with maximum likelihood. *Acta Crystallogr D Biol Crystallogr* (2001),**57**(10).
- Roberts F, Roberts CW, Johnson JJ, Kyle DE, Krell T, Coggins JR, et al. Evidence for the shikimate pathway in apicomplexan parasites. *Nature* (1998),**393**(6687):801-5.
- Ronning DR, Klabunde T, Besra GS, Vissa VD, Belisle JT, Sacchettini JC. Crystal structure of the secreted form of antigen 85c reveals potential targets for mycobacterial drugs and vaccines. *Nat Struct Biol* (2000),**2**:141-146.
- Rossmann MG, Blow DM. The detection of sub-units within the crystallographic asymmetric unit. *Acta Crystallographica* (1962),**15**:24-31.
- Roszak AW, Robinson DA, Krell T, Hunter IS, Frederickson M, Abell C, et al. The Structure and Mechanism of the Type II Dehydroquinase from *Streptomyces coelicolor*. *Structure* (2002),**10**(4):493-503.
- Sambrook J, Fritsch EF, Maniatis T. *Molecular Cloning; a Laboratory Manual*. 2nd ed: Cold Spring Harbor Laboratory; 1989.
- Schindler T, Bornmann W, Pellicena P, Miller WT, Clarkson B, Kuriyan J. Structural mechanism for STI-571 inhibition of abelson tyrosine kinase. *Science* (2000),**289**:1938-1942.
- Scorpio A, Zhang Y. Mutations in *pncA* encoding pyrazinamidase/nicotinamidase, cause resistance to the antituberculosis drug pyrazinamide in the tubercle bacillus. *Nat Med* (1996),**2**:662-667.
- Servos S, Chatfield S, Hone D, Levine M, Dimitriadis G, Pickard D, et al. Molecular cloning and characterization of the *aroD* gene encoding 3- dehydroquinase from *Salmonella typhi*. *J Gen Microbiol* (1991),**137**(Pt 1):147-52.

- Sharma V, Sharma S, Hoener zu Bentrup K, McKinney JD, Russell DG, Jacobs WR, Jr., et al. Structure of isocitrate lyase, a persistence factor of *Mycobacterium tuberculosis*. *Nat Struct Biol* (2000),**7**(8):663-8.
- Sheriff S, Klei HE, Davis ME. Implementation of a six-dimensional search using the AMoRe translation function for difficult molecular-replacement problems. *Journal of Applied Crystallography* (1999),**32**:98-101.
- Shumilin IA, Kretsinger RH, Baurele R. Crystal structure of phenylalanine regulated 3-Deoxy-D-Arabino-Heptulosonate-7-phosphate synthase from *Escherichia coli*. *Structure* (1999),**7**:865-875.
- Silhavy TJ, Szmelcman S, Boos W, Schwartz M. On the Significance of the Retention of Ligand by Protein. *Proc Natl Acad Sci U S A* (1975),**72**:2120 - 2124.
- Simmons CP, Dunstan SJ, Tachedijan M, Krywult J, Hodgson ALM, Strugnell RA. Vaccine potential of attenuated mutants of *Corynebacterium pseudotuberculosis* in sheep. *Infection and Immunity* (1998),**66**:474-479.
- Slayden RA, Barry CE. The genetics and biochemistry of isoniazid resistance in *Mycobacterium tuberculosis*. *Microbes Infect* (2000),**2**:659-669.
- Stallings WC, et al. Structure and topological symmetry of the glyphosate target 5-enolpyruvylshikimate-3-phosphate synthase : a distinctive protein fold. *Proc Natl Acad Sci U S A* (1991),**88**:5046 - 5050.
- Steinrucken HC, Amrhein N. The herbicide glyphosate is a potent inhibitor of 5-enolpyruvylshikimate acid 3-phosphate synthase. *Biochem. Biophys. Res. Commun.* (1980),**94**:1207-1212.
- Stroman P, Reinert WR, Giles NH. Purification and characterization of 3-dehydroshikimate dehydratase, an enzyme in the inducible quinic acid catabolic pathway of *Neurospora crassa*. *J Biol Chem* (1978),**253**(13):4593-8.

Studier FW, Moffat BA. Use of Bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes. *J Mol Biol* (1986),**189**:113-130.

Stura E. Strategy in protein crystallisation. *Journal of Crystal Growth* (1991),**110**:1-12.

Taylor G. Sialidases: structures, biological significance and therapeutic potential. *Curr Opin Struct Biol* (1996),**6**:830-837.

Tonge PJ. Another brick in the wall. *Nat Struct Biol* (2000),**7**(2):94-6.

Toscano MD, Frederickson M, Evans DP, Coggins JR, Abell C, Bello CG. Design, synthesis and evaluation of bifunctional inhibitors of type II dehydroquinase. *Organic and Biomolecular Chemistry* (2003),**1**:2075-2083.

Vagin AA, Teplyakov A. MOLREP: an automatic program for molecular replacement. *Journal of Applied Crystallography* (1997),**30**:1022-1025.

Varghese JN. Development of neuraminidase inhibitors as anti-influenza drugs. *Drug Development Research* (1999),**46**:176-196.

Vogan E. Shikimate dehydrogenase structure reveals novel fold. *Structure* (2003),**11**:902-903.

Vriend G. WHAT IF: a molecular modelling and drug design program. *Journal of Molecular Graphics & Modelling* (1990),**8**:52-56.

Walker JC, Verma NK. Cloning and characterisation of the *aroA* and *aroD* genes of *Shigella dysenteriae* type 1. *Microbiol Immunol* (1997),**41**(10):809-13.

Wang BC. Resolution of Phase Ambiguity in Macromolecular Crystallography. *Methods in Enzymology* (1985),**115**:90-112.

White PJ, Young J, Hunter IS, Nimmo HG, Coggins JR. The purification and characterization of 3-dehydroquinase from *Streptomyces coelicolor*. *Biochem J* (1990),**265**(3):735-8.

WHO. Global Tuberculosis Control: Surveillance, Planning, Financing. Geneva: World Health Organization; 2003. Report No.: WHO/CDS/TB/2003.316.

Ye S, von Delft F, Brooun A, Knuth MW, Swanson RV, McRee DE. The crystal structure of Shikimate Dehydrogenase (AroE) reveals a unique NADPH binding mode. *J Bacteriol* (2003),**185**:4144-4151.

Zhang KYJ, Cowtan K, Main P. Combining constraints for electron-density modification. *Methods in Enzymology* (1997),**277**:53-64.