

IDENTIFICATION OF INHIBITORS OF HUMAN AND *PLASMODIUM*
FALCIPARUM OROTATE PHOSPHORIBOSYLTRANSFERASE AND
OROTIDINE 5'-MONOPHOSPHATE DECARBOXYLASE, AND A
NOVEL HYBRID SCREENING TECHNIQUE

MICHAEL JOEL ROACH

BBiotech (Honours)



Flinders
UNIVERSITY

School of Biological Sciences, Faculty of Science and Engineering

Thesis presented for the degree of Doctor of Philosophy

Contents

Contents.....	i
Summary.....	v
Declaration.....	vii
Acknowledgments.....	viii
Abbreviations.....	ix
List of Tables.....	xi
List of Figures.....	xii
1. Introduction.....	1
1.1. Malaria.....	1
1.1.1. <i>Plasmodium falciparum</i> and Antimalarial Drug Resistance.....	1
1.2. <i>De novo</i> Synthesis of Pyrimidines.....	2
1.2.1. Pyrimidine Salvage Pathway.....	5
1.2.2. Druggability of the Pyrimidine <i>de novo</i> Biosynthesis Pathway.....	6
1.2.3. Orotate Phosphoribosyltransferase.....	8
1.2.4. Oritidine 5'-Monophosphate Decarboxylase.....	11
1.2.5. Uridine Monophosphate Synthase.....	15
1.3. Homology Modelling of Proteins.....	16
1.3.1. Homology Modelling by Satisfaction of Spatial Restraints (UCSF Modeller).....	18
1.4. Drug Discovery: Early Pipeline.....	19
1.5. Drug Discovery: Virtual Screening.....	20
1.5.1. Structure-Based Virtual Screening.....	22
1.5.1.1. Flexible, Anchor-Grow Docking (UCSF Dock).....	22
1.5.1.2. Rigid, Exhaustive Docking (OpenEye FRED).....	24
1.5.2. Ligand-Based Virtual Screening.....	25
1.5.2.1. Jaccard Index (Tanimoto Similarity).....	26
1.5.2.2. Ligand Gaussian Shape Matching (OpenEye ROCS).....	26
1.5.2.3. Electrostatic Similarity (OpenEye EON).....	27
1.5.2.4. 2D Fingerprint Similarity (MOLPRINT 2D).....	28
1.5.3. Validation of a Virtual Screening Method.....	29
1.6. Drug Discovery: Hit to Lead Strategies.....	30
1.7. Aims and Objectives of this Study.....	31
2. Materials and Methods.....	34
2.1. Materials.....	34
2.2. <i>E. coli</i> Strains, Plasmids, and Growth Conditions.....	35
2.2.1. <i>E. coli</i> Strains and Plasmids.....	35
2.2.2. Growth Conditions.....	36
2.3. Molecular Biology and Cloning Techniques.....	37
2.3.1. Plasmid Purification.....	37
2.3.2. DNA Quantification.....	37
2.3.3. PCR Amplification.....	37
2.3.3.1. High-Fidelity PCR.....	37
2.3.3.2. Analytical PCR.....	38
2.3.4. Purification of Restriction Enzyme Digests and PCR Amplicons.....	38
2.3.5. Restriction Enzyme Digests for cloning.....	38
2.3.6. DNA Ligations.....	38
2.3.7. Preparation of Competent <i>E. coli</i> Cells.....	39
2.3.8. Heat-shock Transformation.....	39
2.3.9. Electroporation Transformation of <i>E. coli</i>	40
2.3.9.1. Preparation of Electrocompetent Cells.....	40
2.3.9.2. Electroporation.....	40
2.3.10. Colony PCR screening.....	40
2.3.11. DNA Sequencing.....	41
2.3.12. Agarose Gel Electrophoresis.....	41

2.4. Recombinant Protein Techniques.....	42
2.4.1. Recombinant Protein Expression in <i>E. coli</i>	42
2.4.2. Harvesting of <i>E. coli</i> Cells	42
2.4.3. Lysing of <i>E. coli</i> Cells.....	42
2.4.4. Clearing of Cell Lysate	43
2.4.5. Nickel-Affinity Chromatography.....	43
2.4.6. Sodium Dodecyl Sulphate Polyacrylamide Gel Electrophoresis (SDS- PAGE)	43
2.4.6.1. Coomassie Staining	44
2.4.6.2. Silver Staining	44
2.4.7. Western Blotting.....	45
2.4.7.1. Membrane Transfer	45
2.4.7.2. Blotting and Detection	45
2.4.8. Protein Estimations.....	46
2.4.9. Protein Crystallography and X-Ray Diffraction	46
2.5. Biochemical Assays and Kinetics	48
2.5.1. Preparation of Compounds for Screening.....	48
2.5.2. OPRTase Biochemical Assay	48
2.5.3. ODCase Biochemical Assay	48
2.5.4. Reaction Progress Kinetic Analysis of ODCase Inhibitors	49
2.6. List of Software	51
2.6.1. Protein Structure Viewing and Imaging	51
2.6.2. Sequence Alignments	51
2.6.3. Comparative Modelling of Protein.....	51
2.6.4. Structure-Based Screening.....	51
2.6.5. Ligand-based Screening	52
3. Cloning, Expression and Purification of Recombinant Enzymes	53
3.1. Introduction	53
3.2. Results.....	54
3.2.1. Cloning of <i>H. sapiens</i> UMPS, OPRTase and ODCase into pET30a	54
3.2.2. Expression and Purification of Recombinant Enzymes	57
3.2.2.1. PfOPRTase.....	58
3.2.2.2. PfODCase	62
3.2.2.3. HsOPRTase, HsODCase and HsUMPS.....	65
3.2.3. Specific activities of <i>P. falciparum</i> and <i>H. sapiens</i> OPRTase and ODCase.....	71
3.2.4. Substrate Kinetics of PfODCase and HsODCase	71
3.2.5. Crystallisation Trials	73
3.3. Discussion	75
3.3.1. Expression and Purification of Recombinant Enzymes	75
3.3.1.1. PfOPRTase Expression and Purification	75
3.3.1.2. PfODCase Expression and Purification.....	80
3.3.1.3. HsOPRTase and HsODCase Expression and Purification	80
3.3.1.4. HsUMPS Expression and Purification.....	82
3.3.2. Specific Activities of OPRTases and ODCases.....	86
3.3.3. Substrate kinetics of PfODCase and HsODCase	86
3.3.4. Macromolecular Crystallography of PfOPRTase.....	87
4. Virtual Screening of Human and <i>P. falciparum</i> OPRTase and ODCase	89
4.1. Introduction	89
4.2. Methods.....	90
4.2.1. Homology Modelling of PfOPRTase	90
4.2.1.1. Identification of Template Structure and Homologous Sequences	90
4.2.1.2. Multiple Sequence Alignment	91
4.2.1.3. Generation and Selection of Homology Model.....	91
4.2.2. <i>In silico</i> Screening of <i>P. falciparum</i> and <i>H. sapiens</i> OPRTase and ODCase.....	92
4.2.2.1. File preparation	93
4.2.2.2. Optimisation and validation for docking.....	95

4.2.2.3. Three phase screening	95
4.2.3. Selection of Available Compounds for Biochemical Assays	96
4.3. Results	97
4.3.1. Homology Model of <i>PfOPRTase</i>	97
4.3.2. Comparison of Homology Model to Crystal Structure	101
4.3.3. Validation of Receptor Structures	104
4.3.4. Structure-Based <i>in silico</i> Screening	110
4.3.5. Final Selection of Compounds	113
4.4. Discussion	118
4.4.1. Homology Model of <i>PfOPRTase</i>	118
4.4.2. Validation of Receptor Structures	120
4.4.3. Structure Based <i>in silico</i> Screening	121
4.4.4. Final Selection of Compounds	123
5. Identification and Characterisation of Inhibitors of Human and <i>P. falciparum</i> OPRTase and ODCase	124
5.1. Introduction	124
5.2. Results: Initial Biochemical Inhibition Screen of Compounds	125
5.2.1. Inhibitors of <i>HsOPRTase</i>	127
5.2.2. Inhibitors of <i>PfOPRTase</i>	129
5.2.3. Specific Inhibition of <i>PfOPRTase</i> and <i>HsOPRTase</i>	131
5.2.4. Inhibitors of <i>HsODCase</i>	133
5.2.5. Inhibitors of <i>PfODCase</i>	135
5.2.6. Specific Inhibition of <i>PfODCase</i> and <i>HsODCase</i>	137
5.3. Results: Hit Expansion and Beginnings of a Structure Activity Relationship	139
5.3.1. 2D Fingerprint Rescreen	139
5.3.2. Inhibition Screen of 2D Fingerprint Hits Against <i>PfODCase</i> and <i>HsODCase</i>	141
5.4. Results: <i>PfODCase</i> and <i>HsODCase</i> Michaelis-Menten Kinetics	144
5.4.1. Inhibition Kinetics of Compounds Against <i>PfODCase</i> and <i>HsODCase</i>	144
5.4.2. Kinetic Characterisation of Compound 4049-0191 as an Alternative Substrate of <i>HsODCase</i>	162
5.5. Discussion	166
5.5.1. Initial Inhibition Screen of OPRTases and ODCases	166
5.5.2. 2D Fingerprint Screening for Analogues of Hits and Inhibition Assays	167
5.5.3. Inhibition Kinetics of Compounds Against <i>PfODCase</i> and <i>HsODCase</i>	168
5.5.4. Kinetic Characterisation of Compound 4049-0191 as an Alternative Substrate of <i>HsODCase</i>	173
6. Development and Validation of a Novel Hybrid Screening Method	175
6.1. Introduction	175
6.2. Development of the Hybrid Screening Protocol	177
6.3. Methods: Hybrid Screening and Analysis	182
6.3.1. Selection of Pharmaceutically Relevant Protein Targets, Actives and Decoys	182
6.3.2. Hybrid Screening: Structure-Based Component	185
6.3.2.1. Probe Library Design	185
6.3.2.2. Preparation and Validation of Docking Parameters and Files	187
6.3.2.3. Probe Docking	187
6.3.2.4. Clustering of Docked Probe Conformations	188
6.3.2.5. Cluster Selection	189
6.3.3. Hybrid screening: Ligand-based Component	190
6.3.3.1. Aligning the Known Active and Decoys	190
6.3.3.2. Individual Pharmacophore Descriptor Scoring	190
6.3.3.3. Active Site Occupancy Scoring	191
6.3.3.4. Ligand-Enzyme Overlap Penalty	191
6.3.3.5. Electrostatic Similarity	192
6.3.3.6. Compilation of Final Arbitrary Score	192
6.3.4. Ligand-Based Screening Control: MOLPRINT 2D	192
6.3.5. Structure-Based Screening Control: UCSF Dock	193

6.3.6. Data Analysis.....	193
6.3.6.1. Receiver Operating Characteristic Curves.....	193
6.3.6.2. Hit Diversity Analysis.....	194
6.3.6.3. Enrichment Values.....	194
6.4. Results:.....	195
6.4.1. ROC Curves.....	195
6.4.2. Area Under the Curve Values for ROC Curves.....	201
6.4.3. Enrichment Values.....	203
6.4.4. Hit Diversity.....	205
6.5. Discussion.....	208
7. Discussion.....	215
7.1. Stability of <i>PfOPRTase</i> and <i>HsUMPS</i> Gene in <i>E. coli</i>	215
7.2. OPRTase Drug Discovery Pipeline.....	215
7.3. ODCases as Drug Targets.....	218
7.4. New Insights into the Structural and Catalytic Properties of the ODCase Active Site.....	220
7.5. Applications of Novel Hybrid Screening Protocol.....	223
7.6. Future Directions.....	227
Appendices.....	229
Appendix 8.1: High-Fidelity PCR Step Lengths and Temperatures for the Gene <i>H. sapiens</i> UMPS, and its domains OPRTase and ODCase.....	229
Appendix 8.2: Analytical PCR Step Lengths and Temperatures for the Gene <i>H. sapiens</i> UMPS, and its domains OPRTase and ODCase.....	230
Appendix 8.3: Chromatography Buffers A and B for the Recombinant Proteins <i>HsUMPS</i> , <i>HsOPRTase</i> , <i>HsODCase</i> , <i>PfOPRTase</i> and <i>PfODCase</i>	231
Appendix 8.4: Example of Reaction Progress Kinetic Analysis.....	232
Appendix 8.5: Michaelis-Menten Kinetics Model.....	233
Appendix 8.6: Michaelis-Menten Competitive Inhibition Model.....	234
Appendix 8.7: Michaelis-Menten Noncompetitive Inhibition Model.....	235
Appendix 8.8: Michaelis-Menten Uncompetitive Inhibition Model.....	236
Appendix 8.9: Michaelis-Menten Mixed Inhibition Model.....	237
Appendix 8.10: Translated <i>HsUMPS</i> Protein Sequence.....	238
Appendix 8.11: pET30a Expression Vector.....	239
Appendix 8.12: Sequencing of <i>HsUMPS</i> in pET30a- <i>HsUMPS</i>	240
Appendix 8.13: Sequencing of <i>HsOPRTase</i> in pET30a- <i>HsOPRT</i>	243
Appendix 8.14: Sequencing of <i>HsODCase</i> in pET30a- <i>HsODC</i>	245
Appendix 8.15: 'convertpir.pl' Script for Alignment File Conversion.....	247
Appendix 8.16: <i>PfOPRTase</i> Homology Modelling Alignment File.....	250
Appendix 8.17: Modified 'model-default.py' Script for Homology Modelling of <i>PfOPRTase</i>	252
Appendix 8.18: Default High-Stringency Docking Parameters.....	253
Appendix 8.19: Changes to Grid File Generation Parameters for <i>PfOPRTase</i> , <i>PfODCase</i> , and <i>HsOPRTase</i>	254
Appendix 8.20: Very-Low Stringency Docking Parameters.....	255
Appendix 8.21: Low-Stringency Docking Parameters.....	256
Appendix 8.22: Python Clustering Script.....	257
Appendix 8.23: Hybrid Screening Arbitrary Score: Descriptor Multipliers and Formulae.....	260
Appendix 8.24: Dock Score to .csv Script.....	261
Appendix 8.25: Modified .mol2 to .tab Conversion Script.....	262
References.....	263

Summary

The enzymes of the *de novo* pyrimidine biosynthetic pathway are ideal drug targets for the treatment of malarial infections as, unlike the human host, the parasite is solely reliant on this pathway for survival. Furthermore, this pathway is associated with, and therefore, a drug target for a variety of diseases including cancer, autoimmune disorders and viral infections. Hence, this study focussed on the use of bioinformatics approaches to identify potential lead compounds against human and *Plasmodium* enzymes from this pathway.

The *Plasmodium falciparum de novo* biosynthesis enzymes orotate phosphoribosyltransferase (OPRTase) and orotidine 5'-monophosphate decarboxylase (ODCase) were recombinantly expressed and purified. The expression and purification of *P. falciparum* OPRTase (*PfOPRTase*) was optimised in this study to an extent which enabled biochemical assay and x-ray crystallography experiments to be performed on the enzyme. Constructs were made for the recombinant expression of the bifunctional human homologue Uridine Monophosphate Synthase (*HsUMPS*) as well as its individual OPRTase (*HsOPRTase*) and ODCase (*HsODCase*) domains. These were successfully expressed and purified.

A 3-D homology model for the structure of *PfOPRTase* was generated. The *PfOPRTase* homology model and crystal structures for *P. falciparum* ODCase (*PfODCase*), *HsOPRTase* and *HsODCase* were screened virtually by docking against approximately 1 million drug-like compounds. The hits were analysed and 19 novel, diverse compounds were selected for inhibition assays.

Recombinant *HsOPRTase*, *HsODCase*, *PfOPRTase* and *PfODCase* were used in biochemical inhibition assays to characterise the 19 compounds that were identified by virtual screening for inhibition. For *PfOPRTase* and *HsOPRTase*, nine

compounds showed inhibition against one or both enzymes and five were specific for one of the homologues. These inhibitors would be suitable for further development with hit to lead (H2L) drug discovery experiments.

Eight very promising inhibitors for *Pf*ODCase and *Hs*ODCase were identified which led to some small scale 'hit expansion' H2L experiments, and further biochemical evaluation. Seven compounds successfully underwent further inhibition kinetic characterisation. One inhibitor had good potency ($K_i^{\text{app}}/\alpha K_i^{\text{app}} \sim 50 \mu\text{M}$) and six inhibitors were specific for one of the homologues. Four of the inhibitors (including the most potent inhibitor) satisfy the criteria for drug-likeness and bioavailability.

An alternative substrate for *Hs*ODCase was discovered and characterised. The discovery of this alternative substrate and the unexpected modes of inhibition of the identified ODCase inhibitors allowed for new insights into the structure and catalytic mechanism of this enzyme. Interconnectivity between the ODCase dimer active sites is likely.

Finally, a novel method was developed for performing virtual screening on an enzyme target. This novel method is a 'hybrid' of structure- and ligand-based methods. The method requires knowledge of only a single known binder and a crystal structure. It is estimated to be approximately 60-fold faster than a typical structure-based approach. It outperformed a typical structure-based approach with 25 % better mean bias towards actives, 45 % better mean enrichment at 1 %, whilst maintaining similar hit diversity. It performed similarly compared to a ligand-based approach for bias towards actives and enrichment at 1 % (but with far greater consistency) and 30 % better hit diversity.

Declaration

I certify that this thesis does not incorporate without acknowledgment any material previously submitted for a degree or diploma in any university; and that to the best of my knowledge and belief it does not contain any material previously published or written by another person except where due reference is made in the text.

No part of this thesis may be reproduced without the permission of the author.

Michael J. Roach

School of Biological Sciences, Flinders University

Acknowledgments

I would like to thank my supervisor, Ian Menz. Your help and support over the years was valuable beyond measure. Thank you also to my associate supervisor Melissa Brown.

To my assessors, thank you in advance for reviewing my thesis. Also, I apologise; the thesis ended up being much larger than I was anticipating.

Thank you to Flinders University for giving me the opportunity to complete this higher degree by research.

To the many people in Biological Sciences that have helped me over the years you have my deepest gratitude. Quite often the most support we receive is from fellow students passing on techniques and training one another for using equipment. Special thanks to James, Anjum, and Vivek for their support and friendship, and thank you to the other past and present friends in the Menz lab—Petrol, Wendy, Raj, Sunil, Dan, Eiman, Immy, Mel. Thank you also to Drew Sutton, Melissa Pitman, David Rudd, Uwe Stroehler, and anyone else that helped me with various experiments and pieces of equipment.

I'd like to also thank my family; I would not have been able to finish this project without their love and support. Thank you to my mum, Meredith; my dad, Adrian; my brother, Daniel and his family; and my sister Nicole. Thank you especially to my partner Samantha, and her family.

Abbreviations

Excludes common abbreviations, abbreviations for units, buffers and some compounds, and Journal of Biological Chemistry-listed abbreviations that do not need to be defined.

ACT	Artemisinin Combinational Therapy
ADME	Absorption, Distribution, Metabolism, and Excretion
AGRF	Australian Genome Research Facility
ALR2	Aldose Reductase
AmpC	AmpC beta lactamase
ATCase	Aspartate transcarbamoylase
AUC	Area Under the Curve
COX-2	Cyclooxygenase 2
CPK	Corey, Pauling, and Koltun (CPK colouring)
CPU	Central Processing Unit
CV	Column Volumes
DHFR	Dihydrofolate reductase
DHOase	Dihydroorotase
DHODase	Dihydroorotate dehydrogenase
DMSO	Dimethyl sulfoxide
DUD	Directory of Useful Decoys
EGFr	Epidermal growth factor receptor kinase
FGFr1	Fibroblast growth factor receptor kinase
FPLC	Fast Protein Liquid Chromatography
FRED	Fast Rigid Exhaustive Docking (a screening program)
FXa	Factor Xa
GPU	Graphics Processing Unit
GST	Glutathione S-transferase
H-bond	hydrogen bond
H2L	Hit to Lead
HIVPR	HIV protease
<i>Hs</i> in <i>HsUMPS</i> , <i>HsOPRTase</i> , etc.	<i>Homo sapiens</i>
HTS	High-Throughput Screening
IPTG	Isopropyl β -D-1-thiogalactopyranoside
LB	Lysogeny broth (also known as Luria-Bertani medium)
LogP	Octanol-Water Partition Coefficient
MD	Molecular Dynamics
NA	Neuraminidase
NEB	New England Biolabs
NCBI	National Center for Biotechnology Information

OD	Optical Density
ODCase	Orotidine 5'-monophosphate decarboxylase
OMP	Orotidine monophosphate
OPRTase	Orotate phosphoribosyltransferase
P38	P38 mitogen activated protein kinase
PB	Poisson-Boltzmann
PDB	Protein Data Bank
<i>Pf</i> in <i>Pf</i> OPRTase, <i>Pf</i> ODCase, etc.	<i>Plasmodium falciparum</i>
PMSF	Phenylmethanesulfonylfluoride
PRPP	Phosphoribosyl pyrophosphate
PSA	Polar Surface Area
QM/MM	Quantum Mechanics/Molecular Mechanics
QSAR	Quantitative Structure-Activity Relationship
RE	Restriction Enzyme
RMSD	Root-mean-square deviation of atomic positions
ROC	Receiver Operating Characteristic
ROCS	Rapid Overlay of Chemical Structures (a screening program)
SAR	Structure-Activity Relationship
SRC	Tyrosine kinase SRC
UCSF	University of California, San Francisco
UMPS	UMP Synthase
UPRTase	Uracil phosphoribosyltransferase
WHO	World Health Organisation

List of Tables

TABLE 2.1: EXPRESSION VECTORS, HELPER PLASMIDS AND ANTIBIOTIC SELECTION.....	35
TABLE 3.1: PRIMERS USED IN PRODUCING EXPRESSION VECTORS OF H. SAPIENS UMPS, OPRTASE AND ODCASE	56
TABLE 3.2: ENZYME SPECIFIC EXPRESSION AND PURIFICATION CONDITIONS.....	57
TABLE 5.1: INHIBITOR ANALOGUES SELECTED FOR INHIBITION ASSAYS FROM 2D FINGERPRINT SCREENING.....	142
TABLE 5.2: MICHAELIS-MENTEN KINETICS FOR <i>PFODCASE</i> IN THE PRESENCE OF COMPOUND C337-0223.....	159
TABLE 5.3: SUMMARY OF INHIBITION KINETICS.....	160
TABLE 5.4: CHEMICAL PROPERTIES OF HIT COMPOUNDS.....	161
TABLE 6.1: PHARMACEUTICALLY RELEVANT PROTEIN TARGETS.....	183
TABLE 6.2: AREA UNDER THE CURVE (AUC) VALUES FOR ROC CURVES FROM SECTION 6.4.1. .	202
TABLE 6.3: ENRICHMENT VALUES AT 1% FOR ALL SCREENING METHODS.....	204
TABLE 6.4: NUMBER OF UNIQUE CLUSTERS IDENTIFIED BY SUBSET 1.0 AT A TANIMOTO CUTOFF OF 0.7.....	206
TABLE 6.5: TOTAL UNIQUE CLUSTERS (AT A TANIMOTO CUTOFF OF 0.7) FOR EACH TARGET'S COMBINED ACTIVE, DUD DECOYS, AND GLIDE DECOY COMPOUNDS, AND, THE DIVERSITY INDEX.....	207

List of Figures

FIGURE 1.1: PYRIMIDINE DE NOVO BIOSYNTHESIS PATHWAY.	3
FIGURE 1.2: GENERAL INTERCONVERSION PATHWAY (NON-EXHAUSTIVE) OF PYRIMIDINES.	4
FIGURE 1.3: CRYSTAL STRUCTURE OF HUMAN OPRTASE (PDB ID: 2WNS).	10
FIGURE 1.4: CRYSTAL STRUCTURE OF HUMAN ODCASE. A) RIBBON STRUCTURE (PDB ID: 2QCD).	13
FIGURE 1.5: POPULAR THEORY FOR STEPWISE DIRECT DECARBOXYLATION REACTION MECHANISM FOR ODCASE.	14
FIGURE 3.1: USE OF PRIMERS FOR PRODUCING H. SAPIENS OPRTASE, ODCASE, AND UMPS AMPLICONS FROM CDNA BY PCR.	56
FIGURE 3.2: A TYPICAL ELUTION PROFILE OF RECOMBINANT <i>PFOPRTASE</i> PURIFIED BY NICKEL AFFINITY CHROMATOGRAPHY.	60
FIGURE 3.3: SDS-PAGE GELS OF NICKEL-AFFINITY PURIFIED RECOMBINANT <i>PFOPRTASE</i>	61
FIGURE 3.4: FPLC CHROMATOGRAM OF ELUTION OF RECOMBINANT <i>PFODCASE</i> FROM NICKEL AFFINITY CHROMATOGRAPHY.	63
FIGURE 3.5: COOMASSIE STAINED SDS-PAGE GELS OF NICKEL-AFFINITY PURIFIED RECOMBINANT <i>PFODCASE</i>	64
FIGURE 3.6: FPLC CHROMATOGRAM OF ELUTION OF RECOMBINANT <i>HSODCASE</i> FROM NICKEL AFFINITY CHROMATOGRAPHY.	66
FIGURE 3.7: FPLC CHROMATOGRAM OF ELUTION OF RECOMBINANT <i>HSOPRTASE</i> FROM NICKEL AFFINITY CHROMATOGRAPHY.	67
FIGURE 3.8: FPLC CHROMATOGRAM OF ELUTION OF RECOMBINANT <i>HSUMPS</i> FROM NICKEL AFFINITY CHROMATOGRAPHY.	68
FIGURE 3.9: COOMASSIE STAINED SDS-PAGE GELS OF NICKEL-AFFINITY PURIFIED RECOMBINANT <i>HSODCASE</i> AND <i>HSOPRTASE</i>	69
FIGURE 3.10: COOMASSIE STAINED SDS-PAGE GELS OF NICKEL-AFFINITY PURIFIED RECOMBINANT <i>HSUMPS</i>	70
FIGURE 3.11: MICHAELIS-MENTEN KINETICS OF <i>PFODCASE</i> AND <i>HSODCASE</i> WITH THE SUBSTRATE OMP.	72
FIGURE 3.12: FIGURE FROM KRUNGKRAI ET AL. (2004A):.	76
FIGURE 3.13: DNA SECONDARY STRUCTURE PREDICTION OF <i>PFOPRTASE</i>	79
FIGURE 3.14: DNA SECONDARY STRUCTURE PREDICTION OF <i>HSUMPS</i> DOMAIN LINKER.	85
FIGURE 4.1: FILE PREPARATION OF RECEPTOR AND LIGAND STRUCTURE FILES FOR DOCKING WITH UCSF DOCK 6.	94
FIGURE 4.2: MULTIPLE SEQUENCE ALIGNMENT OF OPRTASES FOR USE WITH GENERATING A HOMOLGY MODEL OF <i>PFOPRTASE</i>	100
FIGURE 4.3: PROCHECK SUMMARY FOR BEST <i>PFOPRTASE</i> HOMOLGY MODEL.	101
FIGURE 4.4: SUPERPOSITION (SECTION 4.3.2) OF HOMOLGY MODEL AND CRYSTAL STRUCTURE OF <i>PFOPRTASE</i>	103
FIGURE 4.5: VALIDATION OF <i>PFOPRTASE</i> RECEPTOR STRUCTURE.	106
FIGURE 4.6: VALIDATION OF <i>HSOPRTASE</i> RECEPTOR STRUCTURE.	107
FIGURE 4.7: VALIDATION OF <i>PFODCASE</i> RECEPTOR STRUCTURE.	108
FIGURE 4.8: VALIDATION OF <i>HSODCASE</i> RECEPTOR STRUCTURE.	109

FIGURE 4.9: GRID SCORES OF THE TOP COMPOUNDS FROM STRUCTURE BASED SCREENING AGAINST OPRTASES.....	111
FIGURE 4.10: GRID SCORES OF THE TOP COMPOUNDS FROM STRUCTURE BASED SCREENING AGAINST ODCASES.	112
FIGURE 4.11.A: FINAL SELECTION OF POTENTIAL PFOPRTASE AND HSOPRTASE BINDERS: COMPOUNDS 1–8.	115
FIGURE 4.11.B: FINAL SELECTION OF POTENTIAL PFODCASE BINDERS: COMPOUNDS 9–16.....	116
FIGURE 4.11.C: FINAL SELECTION OF POTENTIAL OF HSODCASE BINDERS: COMPOUNDS 17–20.....	117
FIGURE 5.1: COMPOUNDS IDENTIFIED BY <i>IN SILICO</i> SCREENING AND OBTAINED FROM CHEMDIV (SAN DIEGO, USA)..	126
FIGURE 5.2: EFFECT OF COMPOUNDS IDENTIFIED BY DOCKING ON HSOPRTASE ACTIVITY.	128
FIGURE 5.3: EFFECT OF COMPOUNDS IDENTIFIED BY DOCKING ON PFOPRTASE ACTIVITY.....	130
FIGURE 5.4: SELECTIVITY OF THE MOST POTENT INHIBITORS OF PFOPRTASE AND HSOPRTASE.	132
FIGURE 5.5: EFFECT OF COMPOUNDS IDENTIFIED BY DOCKING ON HSODCASE ACTIVITY.....	134
FIGURE 5.6: EFFECT OF COMPOUNDS IDENTIFIED BY DOCKING ON PFODCASE ACTIVITY.	136
FIGURE 5.7: SELECTIVITY OF THE MOST POTENT INHIBITORS OF PFOPRTASE AND HSOPRTASE*.	138
FIGURE 5.8: EFFECT OF COMPOUNDS IDENTIFIED BY 2D FINGERPRINT SCREENING ON PFODCASE AND HSODCASE. ...	143
FIGURE 5.9: UNCOMPETITIVE INHIBITION OF PFODCASE BY COMPOUND 8008-2619.	148
FIGURE 5.10: UNCOMPETITIVE INHIBITION OF PFODCASE BY COMPOUND 4470-0385.	149
FIGURE 5.11: UNCOMPETITIVE INHIBITION OF PFODCASE BY COMPOUND 4470-0386.	150
FIGURE 5.12: MIXED INHIBITION OF PFODCASE BY COMPOUND C197-0379.....	151
FIGURE 5.13: UNCOMPETITIVE INHIBITION OF HSODCASE BY COMPOUND 4470-0385.	152
FIGURE 5.14: UNCOMPETITIVE INHIBITION OF HSODCASE BY COMPOUND 4470-0386.	153
FIGURE 5.15: UNCOMPETITIVE INHIBITION OF HSODCASE BY COMPOUND C197-0379.....	154
FIGURE 5.16: MIXED INHIBITION OF HSODCASE BY COMPOUND 7009-0959.	155
FIGURE 5.17: MIXED INHIBITION OF HSODCASE BY COMPOUND C563-0380.	156
FIGURE 5.18: COMPETITIVE INHIBITION OF HSODCASE BY COMPOUND C337-0223.	157
FIGURE 5.19: MICHAELIS-MENTEN KINETICS OF PFODCASE IN THE PRESENCE OF COMPOUND C337-0223.	158
FIGURE 5.20: MICHAELIS-MENTEN KINETICS OF PFODCASE WITH THE SUBSTRATE 4049-0191.	164
FIGURE 5.21: 1) POPULAR THEORY FOR STEPWISE DIRECT DECARBOXYLATION REACTION MECHANISM FOR ODCASE. 2) POSSIBLE ANALOGOUS REACTION WITH COMPOUND 4049-0191.....	165
FIGURE 5.22: MOLECULAR SCAFFOLDS (BEMIS-MURCKO FRAMEWORKS) FOR A) 4047-0385 AND B) 8008-2619.	167
FIGURE 6.1: ORIGINAL IDEA FOR HYBRID SCREENING METHOD.	177
FIGURE 6.2: REVISED CONCEPT FOR HYBRID SCREENING METHOD.....	178
FIGURE 6.3: FINAL CONCEPT AND WORKFLOW FOR HYBRID SCREENING METHOD.	180
FIGURE 6.4: BASIC CONCEPT FOR PERFORMING THE UNIQUE SCREENING AND SCORING.....	181
FIGURE 6.5: PROBE MOLECULES USED FOR CHARACTERISATION OF MOLECULAR BINDING SITES OF PHARMACEUTICALLY RELEVANT PROTEIN TARGETS.	186
FIGURE 6.6 (A–D): COMPARISON OF SCREENING METHODS FOR AMPC, ALR2, COX-2 AND DHFR.....	198
FIGURE 6.6 (E–H): COMPARISON OF SCREENING METHODS FOR EGFR, FGFR1, FXA, AND HIVPR.	199
FIGURE 6.6 (I–L): COMPARISON OF SCREENING METHODS FOR NA, P38, SRC AND THROMBIN.	200

1. Introduction

1.1. Malaria

The causative agent of malaria is the *Plasmodium* protozoan parasite which is transmitted by female *Anopheles* mosquitos. There are an estimated 1.2 billion people at high risk of infection. In 2013, there were an estimated 198 million cases of malaria globally and an estimated 584 000 deaths resulting from infection (WHO, 2014). There are currently five species of *Plasmodium* known to infect humans: *P. falciparum*, *P. vivax*, *P. ovale*, *P. malariae*, and *P. knowlesi*. Of the five species *P. falciparum* is the most deadly.

Infection of humans by the parasite occurs with the transmission of sporozoites via the bite of an *Anopheles* mosquito. The sporozoites infect hepatocytes and undergo asexual reproduction to produce merozoites. Merozoites infect red blood cells and undergo further asexual reproduction to produce more merozoites. This blood schizogony is responsible for the symptoms of malaria. Some merozoites develop into gametes or gametocytes which are transmitted back to *Anopheles* mosquitoes where the sexual stage of the parasite's lifecycle begins. A more detailed outline can be found in Schlagenhauf-Lawlor (2008).

1.1.1. *Plasmodium falciparum* and Antimalarial Drug Resistance

Resistance against all classes of antimalarial drugs except for the artemisinins is widespread (White, 2004). Signs of possible early artemisinin resistance had appeared several years ago in western Cambodia (Dondorp et al., 2009) and has now been detected in Cambodia, the Lao People's Democratic Republic, Myanmar, Thailand and Vietnam (WHO, 2014). The treatment for diagnosed, uncomplicated cases of malaria as recommended by the World Health Organisation (WHO) is an Artemisinin Combination Therapy (ACT) (WHO, 2010). ACTs are the current best method for treating malaria while minimising the likelihood of drug resistance emerging (Lin et al., 2010). As ACTs are the only reliable treatment for malaria the consequences of drug resistance

emerging for this treatment would be catastrophic. As such, there is a pressing need for new antimalarials in the event that drug resistance to ACTs emerges and becomes endemic. The WHO 'Guidelines for the Treatment of Malaria' and annual global malaria reports outline the state of drug resistance and current best practice for treating and preventing malaria, and combating proliferation of drug resistant strains (WHO, 2010, WHO, 2014).

1.2. *De novo* Synthesis of Pyrimidines

Pyrimidines are compounds that are used in cells for the production of RNA and DNA. The *de novo* pyrimidine biosynthesis pathway is the ubiquitous pathway responsible for the synthesis of pyrimidines. The five step pathway creates uridine monophosphate (UMP) from aspartate, carbamoyl phosphate and phosphoribosyl pyrophosphate (PRPP). The steps are outlined in Figure 1.1. Aspartate and carbamoyl phosphate are joined by aspartate transcarbamoylase (ATCase). The ring structure is formed by dihydroorotase (DHOase) to form dihydroorotate. Dihydroorotate dehydrogenase (DHODase) removes two hydrogens to form orotate. Orotate phosphoribosyltransferase (OPRTase) attaches the phosphoribosyl group (from PRPP) to orotate to form orotidine monophosphate (OMP). Decarboxylation of OMP by orotidine 5'-monophosphate decarboxylase (ODCase) yields UMP (Nelson et al., 2008).

Uridine 5'-triphosphate (UTP) is created from UMP and cytidine 5'-triphosphate (CTP) is interconverted from UTP. Ribonucleotide reductases generate the deoxyribonucleotides dUDP and dCDP (the dUTP and dCTP precursors) from UDP and CDP, respectively. The deoxyribonucleotide dTMP (dTTP precursor) is derived from either dUDP or dCDP as shown in Figure 1.2 (Nelson et al., 2008).

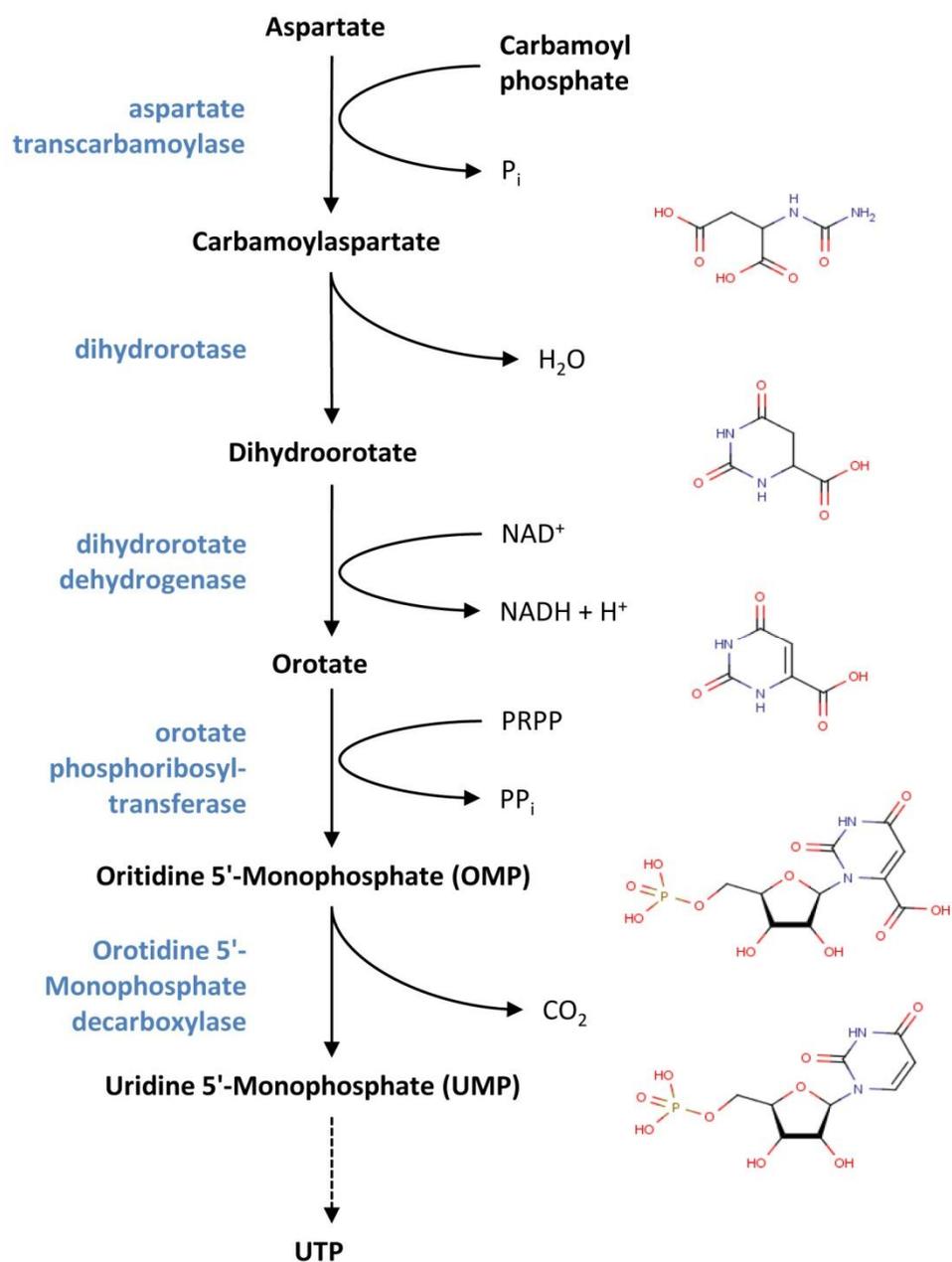


Figure 1.1: Pyrimidine *de novo* biosynthesis pathway. Carbamoyl phosphate formation and the individual steps for converting UMP to UTP are not shown. All compound structures are shown with implicit hydrogens except for heteroatoms, and using CPK colouring (Corey and Pauling, 1953, Koltun, 1965).

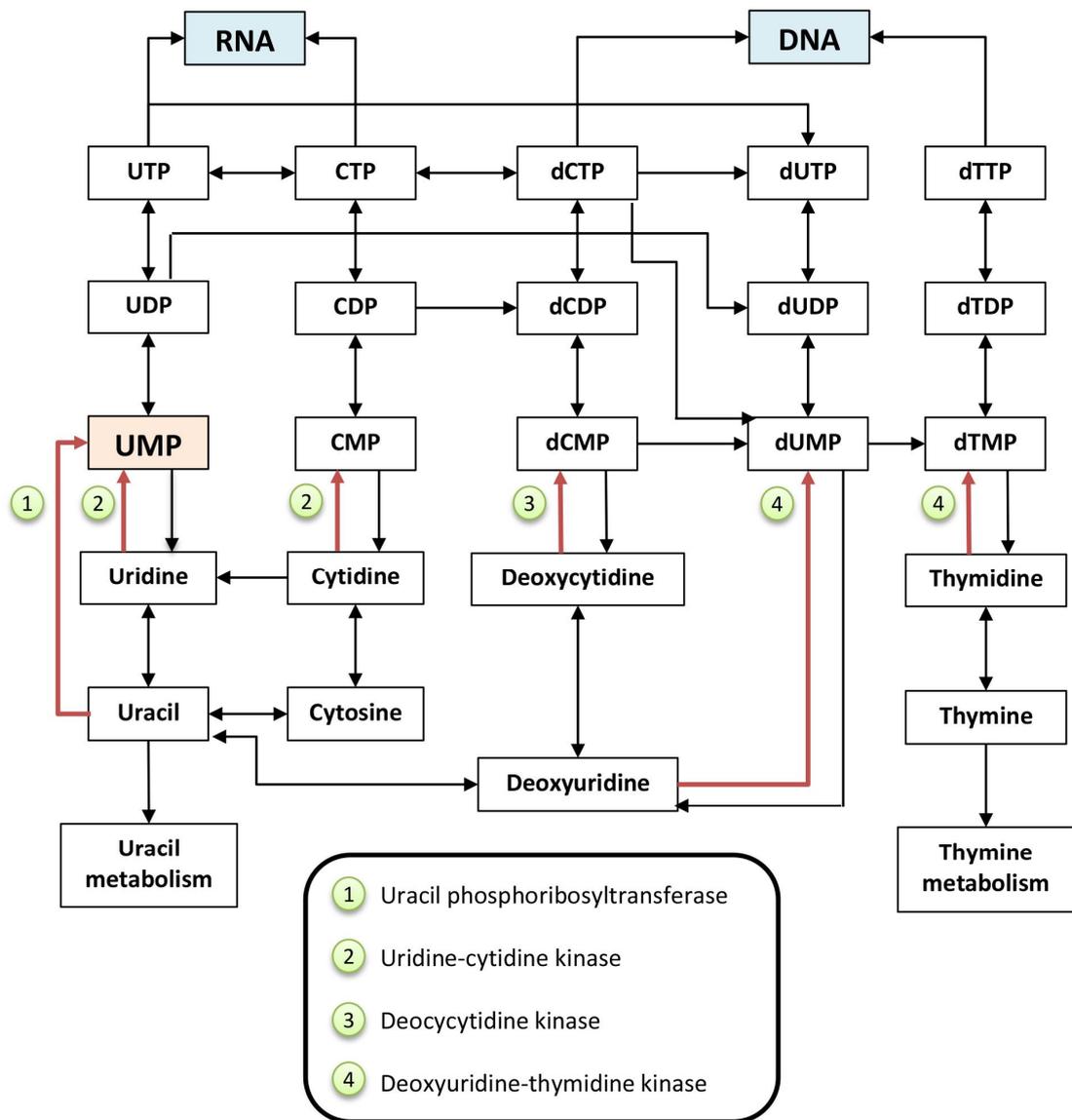


Figure 1.2: General interconversion pathway (non-exhaustive) of pyrimidines.

The product of *de novo* pyrimidine biosynthesis, UMP, is highlighted. Some redundant interconversion enzymes are missing in *P. falciparum* as well as most of the enzymes for interconverting and metabolising the free bases. The four key salvage enzymes that are missing in *P. falciparum* are shown in the figure key and the steps are shown in red in the figure. The figure is derived from pathway maps from Kanehisa Laboratories (2014a)—complete pathways are available for *Homo sapiens*, *P. falciparum*, and many other organisms including nearly all the pathogenic species mentioned in Section 1.2.2. A complete list of the enzymes expressed in *P. falciparum* is also available at PlasmoDB (Aurrecochea et al., 2009).

1.2.1. Pyrimidine Salvage Pathway

Cells are in a constant state of using and releasing free nucleotide bases during the natural production and degradation of RNA and DNA. Catabolism of purines ultimately yields uric acid. Pyrimidines are ultimately catabolised to urea, CO₂ and H₂O. Thymine is broken down by a series of steps to yield NH₃ and succinyl-CoA (a citric acid cycle intermediate). Cytidine and uridine are broken down into uracil. Uracil can be further degraded to β-alanine, malonyl-CoA, CO₂ and NH₃. Production of NH₃ from nucleotide degradation promotes the production of urea via the urea cycle (Nelson et al., 2008). β-alanine is a precursor for several important cellular molecules (Kanehisa Laboratories, 2014b).

A salvage pathway exists to recycle free nucleotide bases. Following the degradation of RNA the free pyrimidine bases uridine and cytidine are released. Cytidine is irreversibly converted to uridine. Uridine can be reversibly converted to uracil. Cytidine or uridine can be recycled to yield CMP or UMP, respectively by uridine-cytidine kinase. Uracil can be either degraded or recycled by uracil phosphoribosyltransferase (UPRTase). UPRTase catalyses a reaction similar to that of OPRTase by attaching a phosphoribosyl group to uracil to yield UMP. As mentioned, UMP is the product of the biosynthesis pathway and can be converted to any pyrimidine as needed. Deoxycytidine is recycled to form dCMP by deoxycytidine kinase. Deoxyuridine and thymidine are recycled to form dUMP and dTMP by deoxyuridine-thymidine kinase.

The importance of nucleotide salvage in humans is highlighted by Lesh-Nyhan syndrome. This defect in the purine salvage pathway (non-functional hypoxanthine-guanine phosphoribosyltransferase) results in uric acid build-up and gout-like damage, and damage to the central nervous system (Nelson et al., 2008). Defects in the pyrimidine salvage pathway are less severe due to the higher solubility of the by-products. Pyrimidine salvage is however especially important during cell division.

Deoxyuridine-thymidine kinase is overexpressed prior to cell division (Bello, 1974, Littlefield, 1966). Orotic aciduria is a condition characterised by a defective UMPS which results in the build-up of orotic acid and leads to physical and mental retardation (due to lack of availability of nucleotides during cell division) (Webster DR et al., 1995, Winkler and Suttle, 1988). It can be treated with uridine and/or cytidine to allow for production of UMP and/or CMP by the salvage enzyme uridine-cytidine kinase. The elevated uridine levels also help to inhibit *de novo* pyrimidine biosynthesis and alleviate high orotate levels (Yazaki et al., 1987).

1.2.2. Druggability of the Pyrimidine *de novo* Biosynthesis Pathway

The *de novo* pyrimidine biosynthesis enzymes shown in Figure 1.1 are all expressed in *P. falciparum* (Aurrecoechea et al., 2009). This pathway has been identified as a potential drug target for malaria (Queen et al., 1990, Rathod et al., 1989). This is because the *Plasmodium* parasite is entirely dependent on this pathway for its supply of pyrimidines for RNA and DNA synthesis, whereas mammalian cells are not as they contain pyrimidine salvage enzymes. Inhibition of the *de novo* pathway has a cytotoxic effect on the *Plasmodium* parasite whereas it only has a cytostatic effect on mammalian cells (as mentioned in Section 1.2.1). As such, specificity for the *P. falciparum* enzymes over the human homologues, while beneficial, is not required. Side-effects of pyrimidine *de novo* biosynthesis inhibition (such as from leflunomide) can be quite unpleasant (Burst and Teschner, 2010); however they could be alleviated with supplementation of a pyrimidine salvage enzyme precursor such as uridine. Uridine supplementation is already used in the treatment of deficient *de novo* pyrimidine pathway in people (Nyhan, 2005).

The pyrimidine *de novo* biosynthesis pathway has been identified as a potential target for treatment of a wide range of other conditions. This pathway is a known potential target for other Apicomplexans such as *Toxoplasma gondii* (toxoplasmosis) (Asai et al., 1983, Schwartzman and Pfefferkorn, 1981) and *Babesia rodhaini* (babesiosis) (Holland

et al., 1983). It has also been identified as a potential target for other protozoan parasites such as *Trypanosoma brucei* (sleeping sickness) (Coustou et al., 2006), and *Trypanosoma cruzi* (Chagas disease) (Hashimoto et al., 2012).

The pathway has also been identified as a target in the treatment of cancer. Inhibitors of this pathway have shown activity against mouse and rat lymphocytic leukaemia cells (Anderson et al., 1989, Bismuth et al., 1982) and human leukaemia cells (Bhalla and Grant, 1987). OPRTase expression levels in tumour cells have been shown to correlate with sensitivity to 5-fluorouracil (an anticancer drug) (Sakamoto et al., 2007). More recently, leflunomide (an inhibitor of DHODase) has shown activity against tumours in a rat model (Zhu et al., 2013). Inhibition of human dihydroorotate dehydrogenase has also been demonstrated to have anti-viral activity against a broad range of viruses (Bonavia et al., 2011, Hoffmann et al., 2011, Marschall et al., 2013, Qing et al., 2010, Smee et al., 2012, Zhang et al., 2012)

Leflunomide is itself an immunosuppressive drug used in the treatment of rheumatoid and psoriatic arthritis. It is also undergoing clinical trials for the treatment of the autoimmune disease lupus (Wu et al., 2013), types of vasculitis (Sanders and Harisdangkul, 2002, Unizony et al., 2013), and a range of inflammatory diseases (Dai et al., 2011, Haibel et al., 2005, Panselinas and Judson, 2012, Pirildar, 2003, Roy, 2007).

There has also been a lot of work on the *de novo* pyrimidine pathway as a target for novel antibacterial drugs. Some examples include demonstrated activity of inhibitors of this pathway against *Helicobacter pylori* (Copeland et al., 2000), *Escherichia coli*, *Enterococcus faecalis* (Marcinkeviciene et al., 2000) and *Mycobacterium tuberculosis* (Breda et al., 2012, Kantardjieff et al., 2005). Antibacterial drug resistance is an ever present concern, and one that has been around for a long time. Despite the pathway being ubiquitous, many of these papers suggest that there are significant structural differences between the bacterial and human homologues to allow for specificity of potential drugs.

1.2.3. Orotate Phosphoribosyltransferase

OPRTase catalyses the formation of OMP (and pyrophosphate) from orotate and PRPP. The reaction is reversible and is Mg^{2+} dependent. In some higher eukaryotes, OPRTase and ODCase exist as a bi-functional enzyme called UMP synthase (UMPS); they otherwise exist as homo-dimers. There is evidence to suggest that the *P. falciparum* OPRTase (*PfOPRTase*) dimer forms a hetero-tetramer with a dimer of *P. falciparum* ODCase (*PfODCase*) (Krungkrai et al., 2005, Krungkrai et al., 2004b). Most bacterial OPRTases are approximately 24 kDa in mass. *PfOPRTase* (*P. falciparum* 3D7) is 33 kDa and contains a unique, 66-amino acid N-terminal insert of low complexity that is not seen in prokaryote or other eukaryote OPRTases (Gardner et al., 2002). These inserts are well documented in *P. falciparum* enzymes (Williams et al., 2007). Their specific function is currently unknown, although in the case of *PfOPRTase* and *PfODCase* their function may be involved with protein-protein interaction (Imprasittichail et al., 2014).

OPRTase from *P. falciparum* and *H. sapiens* has been expressed and purified in the past. The work of Krungkrai et al. (2004a) cloned the gene into a Zero Blunt TOPO (Invitrogen) vector for expression in *E. coli*. The method outlined for expression and purification resulted in a specific activity of 4–5 $\mu\text{mol}\cdot\text{min}^{-1}\cdot\text{mg}^{-1}$ and a 60-fold purification. Yablonski et al. (1996) used a baculovirus expression system to produce human OPRTase (*HsOPRTase*) of UMPS. Suchi et al. (1997) outlines an analytical method for investigating cases of orotic aciduria that involved expression of *H. sapiens* UMPS (*HsUMPS*) in *E. coli*. The structure for *HsOPRTase* also lists *E. coli* as the expression host (Moche et al., 2009).

The human OPRTase domain (when cloned and expressed by itself) has a K_m of $7.1 \pm 0.27 \mu\text{M}$ for orotate (Yablonski et al., 1996) (the K_m for PRPP was not reported by Yablonski et al. (1996)). Recombinantly-expressed *PfOPRTase* (in *E. coli*) has a K_m of $18.2 \pm 0.9 \mu\text{M}$ for orotate and $28.6 \pm 1.3 \mu\text{M}$ for PRPP (Krungkrai et al., 2004a). *PfOPRTase* purified from the parasite has a K_m of $5.6 \pm 0.8 \mu\text{M}$ for orotate and $11.3 \pm$

1.0 μM for PRPP (Krungskrai et al., 2004b). The kinetic parameters from different sources can vary considerably. The K_m values for PRPP and orotate were $18.2 \pm 4.5 \mu\text{M}$ and $18.7 \pm 3.4 \mu\text{M}$ respectively for *Salmonella typhimurium* (Wang et al., 1999b), $62 \mu\text{M}$ and $32 \mu\text{M}$ respectively for *S. cerevisiae* (Umezu et al., 1971) and $40 \mu\text{M}$ and $30 \mu\text{M}$ respectively for *E. coli* (Shimosaka et al., 1985).

Structures are available from the Protein Data Bank (PDB) for OPRTases from human (PDB ID: 2WNS), *P. falciparum* (PDB ID: 4FYM, only recently released in 2013), *Saccharomyces cerevisiae* (PDB ID: 2PRY, 2PRZ, 2PS1), and a number of prokaryotes. The main structure of OPRTases consists of a core of five parallel beta sheets surrounded by seven alpha helices (Figure 1.3). The active site contains a beta hairpin with a large and highly conserved flexible loop that is involved with catalysis. The N-terminal insert in *P. falciparum* codes for an extra alpha helix and an elongation of the N-terminal conserved alpha helix.

The OPRTase reaction follows a random sequential mechanism for *S. typhimurium* (Bhatia et al., 1990, Wang et al., 1999a) and a unique Theorell-Chance mechanism for *S. cerevisiae* (McClard et al., 2006). It begins with the binding of orotate and a PRPP- Mg^{2+} complex (Bhatia and Grubmeyer, 1993). The highly conserved flexible loop in the active site region adopts a closed conformation whereby the substrates are enclosed in the active site and a highly conserved lysine on the flexible loop hydrogen bonds with Mg^{2+} (Henriksen et al., 1996). The reaction involves a classical oxocarbonium ion transition state with the flexible loop involved with protecting the transition state from solvent and hydrolysis (Bhatia et al., 1990, Goitein et al., 1978, Henriksen et al., 1996, Tao et al., 1996).

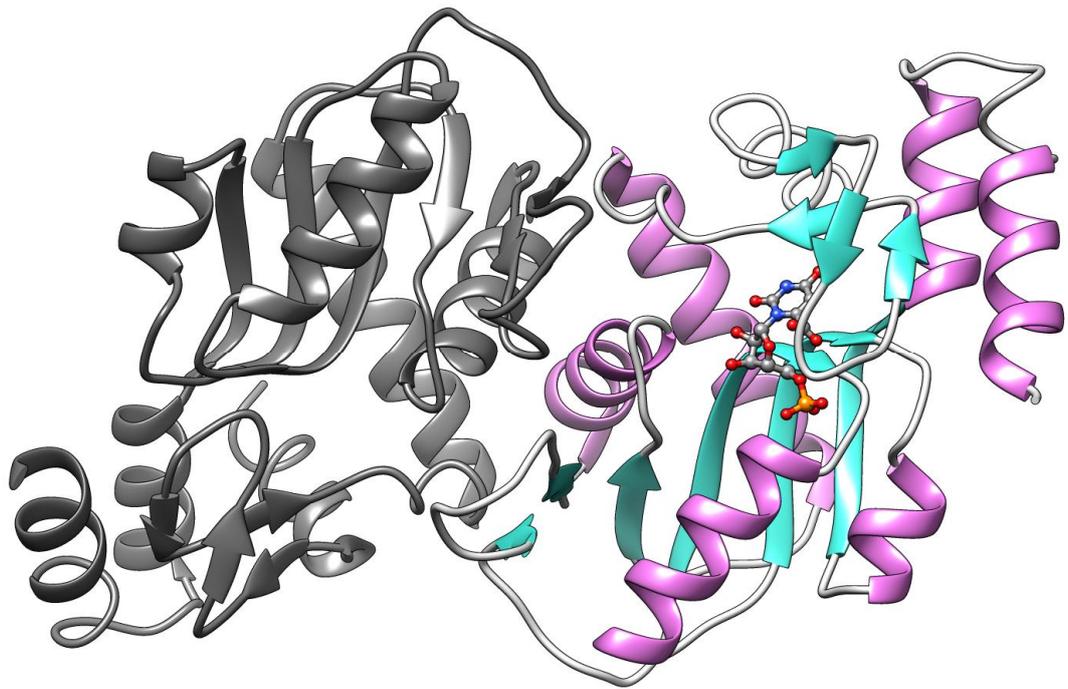


Figure 1.3: Crystal structure of human OPRTase (PDB ID: 2WNS). The dimer is shown as ribbon with chain A coloured dark grey. Chain B is coloured as: alpha-helices, purple; beta-sheets, blue; loops, white. OMP is shown as ball-and-stick in the chain B active site, coloured by CPK.

1.2.4. Orotidine 5'-Monophosphate Decarboxylase

ODCase is responsible for the decarboxylation of orotidine to yield UMP. It is the final step in *de novo* pyrimidine biosynthesis. It is remarkable in that it exhibits a rate enhancement of approximately 10^{17} -fold compared to the spontaneous decarboxylation of OMP (which has a half-life of 78 million years) without the use of metals or other co-factors (Miller and Wolfenden, 2002).

As mentioned, like OPRTase it forms a homodimer, has been shown in the case of *P. falciparum* to form a hetero-tetramer with OPRTase, and in some higher eukaryotes exists as a bifunctional protein with OPRTase. The structure of ODCase consists of a core of seven parallel beta sheets in a beta barrel surrounded by alpha helices. The active site is situated at one end of the beta barrel (Figure 1.4) (Wu and Pai, 2002, Langley et al., 2008, Wittmann et al., 2008, Heinrich et al., 2009).

ODCases seem to readily express in *E. coli*. The ODCase domain for human UMPS (*HsODCase*) has been expressed and purified numerous times in various crystallography studies (Bello et al., 2009, Heinrich et al., 2009, Lewis et al., 2011, Meza-Avina et al., 2010, Purohit et al., 2012, Wittmann et al., 2008). The works of Menz et al. (2002) and Krungkrai et al. (2005) independently outline methods to express *PfODCase* in *E. coli* and purify it to apparent homogeneity.

The human ODCase domain (when cloned and expressed by itself) has a K_m of 295 ± 18 nM for OMP (Yablonski et al., 1996). The *PfODCase* extracted and purified from the parasite has a K_m of 3.2 ± 0.4 μ M (Krungkrai et al., 2004b) and the recombinant *PfODCase* (expressed and purified in *E. coli*) has a K_m of 13.4 ± 1.2 μ M (Krungkrai et al., 2005).

The catalytic mechanism for ODCase had remained controversial for some time. Many papers, including one recently published, are drawing a consensus for a step-wise direct decarboxylation mechanism involving an anionic intermediate, followed by

protonation by an active site lysine residue (Vardi-Kilshtain et al., 2013). The carboxyl is removed to produce CO₂ and a carbanion at C6. Proton transfer occurs from an active site lysine residue (Lys314 in *HsODCase*) to C6 to complete the reaction (Heinrich et al., 2009, Tsang et al., 2012).

The work of Wu et al. (2000) first suggested the direct decarboxylation mechanism. The theory states the use of 'electrostatic stress' whereby the phosphate and ribose are stabilised in the substrate-enzyme complex but the interaction between the orotate ring and ODCase is very destabilising as the C6 carboxyl (as annotated in Figure 1.5) is in a repulsive state with an active site aspartate. There are two lysine and two aspartate residues in an alternating sequence in the active site that create this electrostatic destabilising stress (Figure 1.4). Stabilisation of transition states has alternatively been proposed as the main driving force of catalysis (Warshel et al., 2000). Miller and Wolfenden (2002) describe how the isolation of a transition state from solvent (together with the extremely long half-life of the uncatalysed reaction in solvent) could allow for such a dramatic rate enhancement for ODCases. More recent work by Vardi-Kilshtain et al. (2013) also shows that the decarboxylation is likely to be tightly coupled with the forming of an ion pair between the positively-charged active site lysine residue and the developing anion at C6.

There are currently many structures available for ODCases from numerous organisms. These include 16 structures for *P. falciparum* ODCase and 38 structures for the ODCase domain of human UMPS. Structures are also available for *Plasmodium yoelii*, *Plasmodium berghei*, and *P. vivax* (Vedadi et al., 2007) and other pathogens including *Mycobacterium gastri* (Orita et al., 2010), *T. gondii* (Minasov et al., 2013), *Vibrio cholera* (Halavaty et al., 2010), *Campylobacter jejuni* (Halavaty et al., 2011).

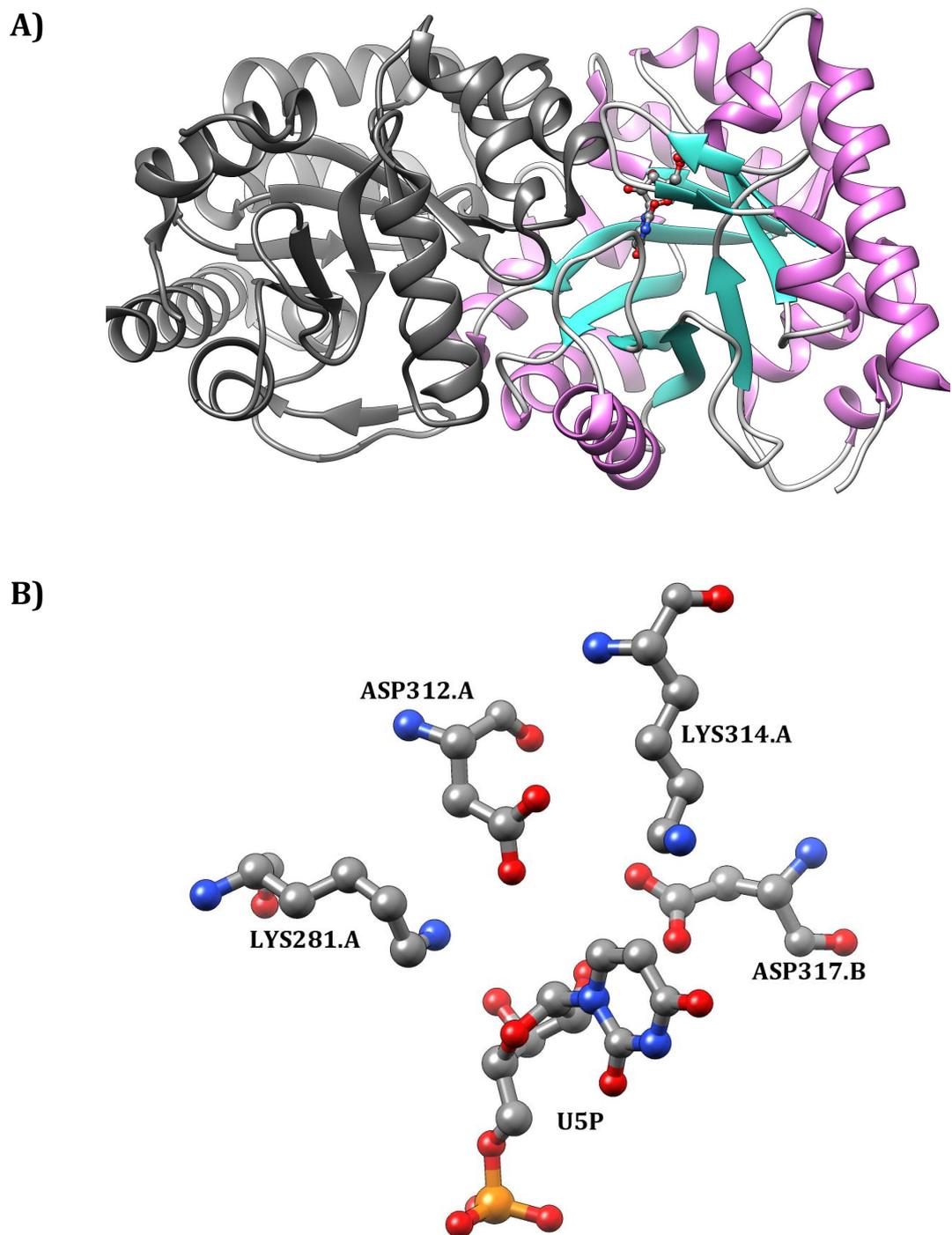


Figure 1.4: Crystal structure of human ODCase. A) Ribbon structure (PDB ID: 2QCD). The dimer is shown as ribbon with chain A coloured dark grey. Chain B is coloured as: alpha-helices, purple; beta-sheets, blue; loops, white. The product UMP is shown as ball-and-stick in the chain B active site, coloured by CPK. **B) Active site lysine and aspartate residues (PDB ID: 2V30).** The catalytically relevant active site two lysine and two aspartate residues are shown with the co-crystallised ligand U5P (ball-and-stick, CPK coloured). In the case of OMP the carboxyl group would be protruding towards ASP312.

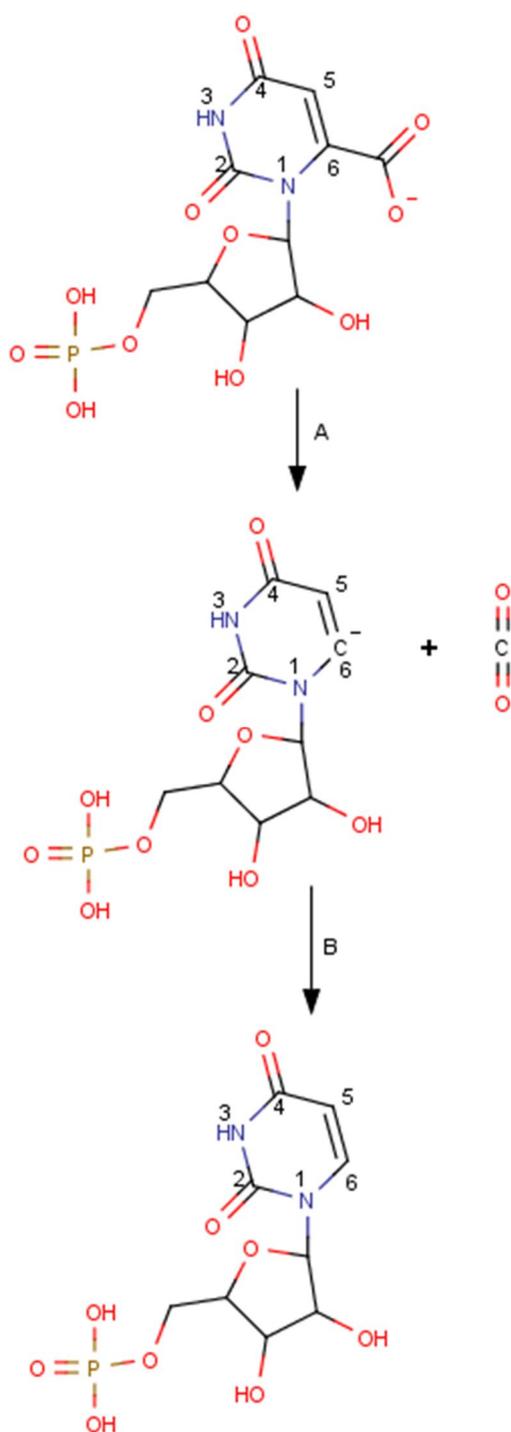


Figure 1.5: Popular theory for stepwise direct decarboxylation reaction mechanism for ODCase. Reaction Step A) the carboxyl group is removed from carbon C6 to produce CO_2 and the anionic intermediate. **Reaction Step B)** the C6 carbon is protonated by the active site lysine residue.

1.2.5. Uridine Monophosphate Synthase

UMP synthase is a bifunctional protein containing the fifth and sixth enzymes of the pyrimidine biosynthesis pathway—OPRTase and ODCase. The fusion event of these two proteins has occurred several times. UMPSs in higher eukaryotes are orthologs with the OPRTase domain situated at the N-terminal end and ODCase the C-terminal end of the hybrid protein. However, the reverse configuration has been shown to have occurred independently for some streptophiles, kinetoplastids, and possibly cyanobacteria (Makiuchi et al., 2007).

As mentioned, defective function of either the OPRTase or ODCase domains of UMP synthase is the cause of type I orotic aciduria. The disease causes a deficiency in pyrimidines for RNA and DNA synthesis and the excessive excretion of orotate in the urine, resulting in anaemia and retarded growth (Webster DR et al., 1995).

Analysis of the crystal structure suggests that substrate channelling could occur from the OPRTase to the ODCase active sites for various UMPS enzymes, and for the multienzyme complex between *Pf*OPRTase and *Pf*ODCase (Pragobpol et al., 1984, Traut, 1989, Wittmann et al., 2008, Kanchanaphum and Krungkrai, 2009), however no definitive experimental evidence exists to support that this channelling actually occurs.

The binding affinities of the OPRTase and ODCase domains are also higher for the recombinant, bifunctional human UMPS than when the domains are cloned, expressed and purified individually. The K_m for the OPRTase domain of recombinant UMPS is $2.1 \pm 0.12 \mu\text{M}$ for orotate compared to $7.1 \pm 0.27 \mu\text{M}$ for the domain by itself. The K_m for the ODCase domain of recombinant UMPS is $230 \pm 8.7 \text{ nM}$ for OMP compared to $295 \pm 18 \text{ nM}$ for the domain by itself (Yablonski et al., 1996). The multienzyme complex between *Pf*OPRTase and *Pf*ODCase shows a similar trait. The K_m for *Pf*OPRTase in the complex is $5.6 \pm 0.8 \mu\text{M}$ and $11.3 \pm 1.0 \mu\text{M}$ for orotate and PRPP, respectively (compared to $18.2 \pm 0.9 \mu\text{M}$ and $28.6 \pm 1.3 \mu\text{M}$ for orotate and PRPP, respectively for the monofunctional recombinant *Pf*OPRTase). *Pf*ODCase in the complex has a K_m of

3.2 ± 0.4 μM for OMP compared to 13.4 ± 1.2 μM for the monofunctional recombinant *Pf*ODCase (Krungskrai et al., 2005).

There are currently many structures available for the human ODCase and one for the OPRTase domains of UMPS. However, there are no structures available for the whole bifunctional protein for any of the related higher eukaryotes. The tertiary structure of this enzyme family is of interest for both human and *P. falciparum* enzymes (the structures of UMP synthases may give insights into how the *P. falciparum* enzymes form a heterotetramer). As such, a crystal structure of the human (or any closely related eukaryote) UMPS is desirable.

A structure is available for the *Leishmania donovani* (a kinetoplastid) UMP synthase. The structure shows a tetramer configuration (α - β)₄ when taking into account higher order symmetry in the crystal structure. Interestingly, no significant contacts occur between the OPRTase and ODCase domains of the same chain, but there are interactions with domain-domain interfaces between the other subunits (both α : α and α : β interfaces) (French et al., 2011). The domain order is the reversed configuration for *L. donovani* as the result of an evolutionary event independent to UMP synthases in higher eukaryotes. As such, the arrangement of human UMPS cannot be derived or inferred from this structure.

It does however show support for the notion that a substrate channelling mechanism is involved with the UMPS. Electrostatic potential mapping of the structures shows a positive pathway linking the active sites, similar to the example given for electrostatic channelling described in Miles et al. (1999). This fits well with the substrate OMP having a net negative charge (no isoelectric point).

1.3. Homology Modelling of Proteins

The pool of experimentally solved protein structures is ever growing. Advancements in macromolecular crystallography are improving the quality and number of solved

protein structures, and allowing structures to be solved for more and more difficult targets. The pool of potential targets however is still orders of magnitude greater and growing with each genome that is sequenced. Targets such as membrane proteins are still notoriously difficult and expensive to crystallise. Even solving easier target proteins can be very time consuming. As such homology modelling is every bit, if not more relevant today than it was 20 years ago.

There are two main categories for protein structure prediction, *ab initio* (where a model is built from the sequence alone) and homology modelling (also known as comparative modelling, where the sequence is built using a template structure as a reference). Only homology modelling will be covered here.

Homology modelling is a computational technique to predict the structure of a protein by modelling it on a closely-related, homologous protein for which there is a known structure. It offers the ability to produce a structure of a target protein in a matter of hours, at no cost. Homology models are used for a number of applications including virtual screening in drug discovery and investigations into regions of importance for further experimental analysis.

This method for modelling protein structures works on the premise that protein structures are more highly conserved than the sequence. Proteins of a particular family will be divergent from a common ancestor. While the sequence may differ significantly, there is enough redundancy that the basic fold will usually be the same. Generally a model can be made if the sequence identity between two proteins is 30 % or greater (Rost, 1999). Less than 25 % identity makes it difficult to establish common ancestry from the sequence alone (Chung and Subbiah, 1996).

General steps involved with homology modelling are: identification of related template structure(s), alignment of the target and template sequences, building the model, model optimisation, and model evaluation (Pitman and Menz, 2006).

1.3.1. Homology Modelling by Satisfaction of Spatial Restraints (UCSF Modeller)

Modeller is a homology modelling program that models proteins by satisfaction of spatial restraints (Sali and Blundell, 1993, Fiser et al., 2000). An alignment of the target sequence (the sequence to be modelled) and the template structure sequence(s) (the structure(s) to be used as a 3D template during modelling) is provided. Modeller generates a series of spatial restraints based on the template structures and the alignment. The restraints take into account a number of factors such as residue type, likeness to the corresponding template residue and local sequence similarity. The restraints are weighted. For instance, a residue that is identical to the corresponding residue in the template structure and is in a highly conserved region will have far greater restraints than one that is different to the template residue and in a poorly conserved region. These input restraints dictate where each residue should be and in what conformation. The model is built in the likeness of the template structures and then optimised to violate as few restraints as possible across the whole model.

Modeller itself requires the following to generate a homology model: alignment file, template structure(s), and a python script file. The alignment file is an .ali file, this is essentially an alignment file in PIR format with included fields that inform modeller which structure belongs to which sequence in the alignment and where to start and stop reading from the template structure file. The template structures are simply .pdb files as downloaded from protein data bank (Berman et al., 2000). The script is a python script file that runs modeller, it calls which parameters and files to use, number of models to generate, what post modelling optimisation to perform and some other optional parameters.

1.4. Drug Discovery: Early Pipeline

There are many steps in the drug discovery pipeline, from an initial screen to clinical trials and post approval review. This project will focus on only the first few steps: initial screening and hit generation, hit to lead and lead selection, and lead development.

Modern drug discovery begins with the screening of an identified target enzyme. For some time this was performed with an extremely expensive and time consuming process known as High-throughput Screening (HTS) where hundreds of thousands of compounds are tested biochemically (Kenny et al., 1998). Virtual screening however has replaced it in many modern drug discovery pipelines and remains the favoured alternative to HTS for identifying bioactive compounds. Compounds that show some inhibition (or whatever the desired effect) in HTS are termed 'hits'. In virtual screening the top scoring compounds are tested biochemically and any found to be bioactive are termed 'hits' (Tanrikulu et al., 2013).

The lead compound is the compound that has been chosen for development into a drug candidate. Screening will often yield multiple hits. The lead compound is the hit that is chosen for further development to eventually become a drug candidate ready for clinical trials. There are many aspects to consider when choosing a lead compound. As such, screening will not always yield hit compounds that are deemed suitable as a lead compound. Most importantly, hit compounds need to satisfy a number of chemical properties before they are considered for lead compound selection.

Lipinski's rule-of-five (Lipinski et al., 2001) should absolutely be satisfied. This rule states the molecular properties for a drug-like compound: no more than 5 hydrogen bond (H-bond) donors (total number of hydrogen atoms bonded to oxygen or nitrogen atoms); no more than 10 H-bond acceptors (total number of nitrogen and oxygen atoms); a molecular mass less than 500 Daltons; and an octanol-water partition coefficient (logP) no greater than 5. However, as groups are added to a lead compound during optimisation the molecular weight, lipophilicity and hydrogen bonders tend to

increase (Oprea et al., 2001). An important consideration for potential lead compounds is to ensure there is 'room' for optimisation. Congreve et al. (2003) suggest a 'rule of three' (variant of Lipinski's 'rule of five') for lead-likeness consisting of: molecular mass less than 300 Daltons; no more than 3 H-bond donors; no more than 3 H-bond acceptors; no more than 3 rotatable bonds; and a logP not greater than 3. Veber et al. (2002) suggest two rules for bio-availability for a compound: 10 or fewer rotatable bonds and a polar surface area equal to or less than 140 Å². The other main consideration is the potency of the hit compound. There are a number of strategies to improve hit compounds' inhibition and chemical properties through minor alterations prior to lead compound selection. These are known as hit to lead (H2L) strategies.

Lead development itself involves extensive modifications and assessment/reassessment of the compound to improve potency, specificity and pharmacokinetics. It is generally very time consuming and intensive. As mentioned, the lead compound will usually change significantly during development. These changes are designed to be a balance between increasing the binding affinity of the compound to the target, and improving its pharmacological properties.

Sections 1.5.1 and 1.5.2 outline virtual screening and describe the methods and programs used in this project. Section 1.5.3 describes how new virtual screening methods are typically validated. Section 1.6 outlines some H2L strategies.

1.5. Drug Discovery: Virtual Screening

The chemical space for drug-like compounds has been estimated to be as many as 10⁶⁰ compounds (Kirkpatrick and Ellis, 2004). The number of known compounds in corporate collections and public databases has been estimated at 100 million compounds (Reymond and Awale, 2012). With such a large chemical space it is necessary to use computational methods to predict which compounds are bioactive prior to screening biochemically.

Virtual screening (or *in silico* screening) is a computational method used in drug discovery to rapidly assess compound libraries and identify the compounds that are likely to bind to a drug target. In this way it replaces high-throughput screening as only the high scoring compounds are tested biochemically. The time and money saved compared to high-throughput screening is enormous.

There are two main methods for performing virtual screening: ligand- and structure-based (McInnes, 2007). Ligand-based methods screen ligands for particular properties that are known to be required for bioactivity. Structure-based methods on the other hand will orient the compound in the target's binding site and calculate an energy score (change in Gibbs free energy from Van Der Waal forces, hydrogen bonding, electrostatic interactions, desolvation) (Krüger and Evers, 2010, Drwal and Griffith, 2013).

Ligand-based screening is performed in a fraction of the time of structure-based screening. Structure-based screening usually requires the use of a high-performance computing cluster and hundreds or even thousands of hours of Central Processing Unit (CPU) wall time to screen a large compound database. Most ligand-based methods can screen a library of the same size on a typical personal computer in a matter of hours or days. This is due to the difference in complexity of the calculations taking place to screen and score the compounds. Structure-based screening is not without its advantages. Ligand-based screening requires the knowledge of multiple known binders and screens for similar compounds. It can be quite limiting due to the bias towards finding compounds that are similar to the pool of already known ligands. Structure-based methods require only the knowledge of the target structure and are more likely to identify novel compound scaffolds (the scaffold of a compound is essentially the compound without the side-chains) and interactions. Thus structure-based screening will typically return a more diverse range of compounds.

1.5.1. Structure-Based Virtual Screening

There are a large number of docking programs available for use with virtual screening; two were used in this project. The programs used in this study have very different methods for orientating the ligand. Outlined here are the principles behind the docking programs, the requirements for use and the output returned from them.

1.5.1.1. Flexible, Anchor-Grow Docking (UCSF Dock)

UCSF Dock 6 (Moustakas et al., 2006) is a structure-based virtual screening program and the primary docking program used in this project. The program will flexibly dock a ligand into a conformationally rigid receptor structure and calculate an energy score based on non-covalent interactions. It has been used recently to identify novel inhibitors of Bcl-XL (a major antiapoptotic protein) with the most potent inhibitor having an IC₅₀ value of $0.49 \pm 0.12 \mu\text{M}$ (Park et al., 2013).

Dock uses target spheres with an 'anchor-grow' algorithm for ligand positioning. Scoring is calculated using receptor grids (like most docking programs). The 'receptor' is simply the target enzyme. A 'grid' file is a file that represents the structure of the receptor as a 3D grid or matrix. There are multiple grid files of the receptor for things such as atom volumes and electrostatic potential. Using grids is far more efficient for docking and scoring than using atom coordinate files. The inputs required for Dock are: the ligand library file to be screened, grid files of the receptor target, target spheres file and the docking parameters file.

The target spheres essentially act as a negative image of an active site by representing possible positions in which ligand atoms can be situated. Spheres are generated for an enzyme target's molecular surface and clusters of spheres identify binding pockets. A docking run would typically be performed for one sphere cluster (one binding site). It is up to the user to select the appropriate sphere cluster that corresponds to the binding site of interest (or to only generate spheres for the desired binding site).

The anchor-grow method for positioning a ligand works as follows. Firstly an 'anchor' is identified in the ligand to be docked. This is ideally the largest part of the ligand that is rigid. The anchor is randomly placed so that at least one atom is positioned in a target sphere. The rest of the ligand is then 'grown'. This is achieved by adding and rotating each rigid section of the ligand until the entire ligand is positioned. Each anchor position and grow step is checked to make sure it does not clash with the target enzyme. The positioned ligand is then scored and the process repeated until no more positions are possible or until the (user defined) maximum number of poses to attempt is reached. Each anchor position yields multiple possible poses, each primary grow step can yield multiple further poses (if more than one grow step is required) and so on. As highly flexible ligands increase the degrees of freedom for this anchor-grow algorithm, the computation time increases exponentially with ligand flexibility.

The target grid consists of a number of files but generally it represents the receptor binding site in the format of a 3D matrix or grid (hence the name). Each grid reference is paired with its contents. Representing the active site in this way allows the program to perform calculations (volumes, clashes, energy scores) far more efficiently than if it were to draw each atom and bond from a coordinate file (such as a .pdb file). As mentioned, there are multiple grid files, each with specialised roles such as detecting atom clashes, calculating electrostatic scores, van der Waals scores and hydrogen bonding.

The parameters file contains every aspect of the docking that is customisable and generally instructs Dock what it is docking and how it's docking and scoring it.

Parameters include: the ligand library to score, flexible or rigid docking, number of poses to attempt, number of grow steps, clustering of poses and clustering method, minimisation of poses and anchor/grow steps, which scoring algorithms to use, output ranking and hit-list size.

The default scoring function for Dock is the Grid score. This is calculated using implicit solvent and is comprised of a Van Der Waals and an electrostatic component. The AMBER molecular mechanics force fields and Poisson-Boltzmann (PB) electrostatic potentials (Fogolari et al., 2002) are precalculated for the receptor and saved to grid files for use with scoring. The scoring itself uses the Lennard-Jones potential for approximating interactions between atoms (the attractive and repulsive exponents, and dielectric factors are specified during grid file generation) (Lang et al., 2015). While this model is extremely fast it is not the most accurate model for estimating free energy changes. The docking to a rigid receptor however is a far greater limitation (as it is for most docking programs) as it often prevents known actives from being able to be oriented into the target active site. The rigid body assumption is ideally only suitable where the ligand binding does not change the active site conformation dramatically, or where it is intended for a ligand to stabilise a particular conformation (Moustakas et al., 2006).

1.5.1.2. Rigid, Exhaustive Docking (OpenEye FRED)

A different approach to generating conformations of a ligand in a binding site is to use an exhaustive approach rather than using target spheres to guide the positioning. OpenEye FRED (Fast Rigid Exhaustive Docking) (McGann, 2011) is a structure-based screening program that performs its ligand orientating in such a way. It relies on OpenEye OMEGA (Hawkins et al., 2010) to generate multiple bioactive conformations of ligands prior to screening and then rigidly docks the conformers to the rigid-body receptor binding site. This increases the total number of structures to dock but actually reduces computation time compared to flexible docking. FRED has been used recently by Brus et al. (2014) to identify a novel inhibitor with an IC_{50} of 21.3 nM for BChE—a therapeutic target for Alzheimer's disease.

The exhaustive approach to generating conformers works as follows. OpenEye FRED automatically characterises and prepares an active site for docking. A user-defined site

box will define the active site and grid files are generated. Automated probe docking will then generate an inner and outer 'contour'. These contours essentially define the shape of an active site and are used in filtering poses.

Docking of a ligand conformer begins with generation of a pose ensemble. As mentioned, exhaustive docking occurs to generate poses of every orientation at every position in the active site. The ligand is rotated with approximately 1 Å Root-mean-square deviation of atomic positions (RMSD) between subsequent poses. Each ligand rotation is positioned at the beginning of the grid and moved along the x , y , and z axes in 1 Å steps to generate the pose ensemble. The contours are then used in filtering the pose ensemble to remove poses that do not complement the active site shape well enough. The remaining poses are scored and ranked.

The default 'chemgauss' scoring function for FRED utilises a receptor grid for scoring ligand poses. The function is comprised of a simple Gaussian-smoothed potential for scoring non-covalent lipophilic and metal-protein interactions, and an explicit hydrogen bond score with a flexibility penalty. Like Dock's Grid score, the chemgauss scoring function is fast but not the most accurate for estimating free energy changes. The rigid-body receptor assumption means that FRED suffers the same limitations as Dock. The rigid-ligand assumption is an additional limitation; while it greatly increases docking speed, it is reliant on OMEGA to accurately predict the bioactive conformations of the ligands.

1.5.2. Ligand-Based Virtual Screening

There is far more diversity in methodologies for ligand-based than there are structure-based screening methods. Outlined here are the three main methods (and programs) used in this project as well as a definition of Tanimoto similarity—an almost ubiquitous scoring measure among ligand-based screening methodologies.

1.5.2.1. Jaccard Index (Tanimoto Similarity)

Many ligand-based screening methods use some form of the Jaccard Index (also referred to as Tanimoto similarity). This statistic is commonly used for measuring similarity and diversity of two sets of data. It is defined as the size of the intercept divided by the size of the union for those two data sets (Jaccard, 1901).

$$J(A, B) = \frac{|A \cap B|}{|A \cup B|} \quad 0 \leq J(A, B) \leq 1$$

In ligand-based screening methods it is commonly used for comparing the ligand that is being screened (dataset B) to the query ligand or model (dataset A). The way the ligands are represented, and intercepts and unions are measured differs between programs. For example some programs will represent atoms as static volumes and measure the overlap between the two volume datasets.

1.5.2.2. Ligand Gaussian Shape Matching (OpenEye ROCS)

OpenEye ROCS (Rapid Overlay of Chemical Structures) (Grant et al., 1996) is a ligand-based screening program that uses a Gaussian function to generate volumes to represent a ligand's shape and the key chemical features (rings, ionisable groups, H-bonding atoms and hydrophobic groups). This is similar to a pharmacophore which is a 3D ensemble of chemical features responsible for molecular recognition, except that ROCS also adds a shape matching component to the screening model. The representative volumes are superimposed on the screening model (also referred to as the 'query'), volume overlaps of ligand shape and the chemical features (also referred to as pharmacophore descriptors) are measured, and the Tanimoto similarities are returned. Like OpenEye FRED, ROCS screens rigid bioactive conformers of ligands (generated by OpenEye OMEGA). Recently, Sun et al. (2014) report using ROCS to identify novel inhibitors of Hsp90 (therapeutic target for cancer) the most potent of which has an IC_{50} of $0.10 \pm 0.01 \mu\text{M}$.

The screening itself is carried out as follows. The centres of mass of the ligand and query are aligned. The ligand is rotated in a solid-body optimisation process that maximises the volume overlap. The Tanimoto similarities for the 'shape' (essentially the heavy atom volume overlap) and 'color' (the pharmacophore component) are calculated using Gaussian functions rather than solid-sphere functions. The use of Gaussian functions greatly reduces computation time and increases hit diversity compared to solid-sphere functions. A Tanimoto 'combo' score is the default primary score and is the combination of the Tanimoto 'shape' and 'colour' scores.

The inputs required are simply the ligands to be screened and the query. Query models can be generated from a single known binder, a set of aligned known binders, or created manually. Customised parameters can also be incorporated for screening.

The use of Gaussian volumes makes ROCS somewhat insensitive to ligand-scaffolds while still selecting for a particular ligand shape. This allows for scaffold hopping while retaining shape complementarity. The limitations to this type of scoring are similar to other pharmacophores, i.e. the quality of the model is largely dependent on known binders. There is also virtually no scope for calculating free energy changes upon ligand binding.

1.5.2.3. Electrostatic Similarity (OpenEye EON)

OpenEye EON (Muchmore et al., 2006) is a specialised program that compares the electrostatic potential map of two small molecules. It can be used to complement a virtual screening run by taking ligands that have been pre-aligned and scoring their electrostatic similarity to a known binder. OpenEye EON does not orient compounds and can only take input poses and compare them with the query molecule as is. As it is a specialised tool designed to complement virtual screening it does not align the ligands, rather the aligned and scored poses from a virtual screen (such as a ROCS screen) are used as the input database to be screened by EON.

To run EON a query ligand is provided, as well as the pre-aligned ligands to be screened. EON uses PB electrostatics to calculate electrostatic potential maps and returns the Tanimoto similarity.

1.5.2.4. 2D Fingerprint Similarity (MOLPRINT 2D)

MOLPRINT 2D (Bender et al., 2004) is a 2D fingerprint (ligand-based) virtual screening program. It is a ligand comparison tool; it compares the 2D fingerprints of a query ligand and a target ligand and returns the Tanimoto similarity. It has been used recently in a H2L study for the identification of high-affinity leads for the β -1 adrenergic receptor (Christopher et al., 2013).

2D fingerprint screening is an extremely simple and fast ligand-based screening method. Screening with 2D fingerprints however is extremely sensitive to compound scaffolds, in that it will generally return the compounds with the most similar scaffolds as the query compound. It can be very useful for finding compounds that are structural analogues of a promising hit compound, but has virtually no potential for scaffold hopping.

There are many different methods for representing ligands as 2D fingerprints. The method used for MOLPRINT 2D involves heavy atom type counts in layers around each heavy atom in a ligand (Bender et al., 2003). This is different to the more popular chemical hashed fingerprints (also referred to as bitstring fingerprints) used by many 2D fingerprint comparison programs such as SUBSET 1.0 used in Chapter 6 for assessing hit diversity. With bitstring fingerprints, short linear patterns and branching points of a compound are converted to a bitstring (a sequence of zeros and ones) using a hashing method (Daylight, 2011)).

Running MOLPRINT 2D involves the ligands to be screened and the query ligand being converted to 2D fingerprints. A script then compares the ligands to the query and a Tanimoto similarity score is returned for each.

The main limitations, other than the high sensitivity to ligand scaffolds, is that this type of screening can be very inconsistent. The quality of a screen can be influenced by the parameters used to convert ligand structures to 2D fingerprints, size of query molecule compared to actives, and methods can differ greatly in performance depending on the enzyme target (Duan et al., 2010, Scior et al., 2012).

1.5.3. Validation of a Virtual Screening Method

Developing a new virtual screening method requires validation to test if a method will score active compounds over inactive ones. This is most commonly achieved by retrospective virtual screening. Performing this validation requires a set of protein targets for which there are multiple known binders (a large number of diverse active compounds is highly desired) and a set of decoy compounds. The purpose of the decoy compounds is to attempt to outscore the active compounds. The decoys could be a generic set of highly diverse drug-like compounds, or they could be a set of compounds specifically selected for their similarity to known actives for a particular target.

The UCSF Directory of Useful Decoys (UCSF DUD) is a database of 40 protein targets with a combined total of 2950 known active compounds (Huang et al., 2006). For each known active compound there are 36 decoy compounds with similar chemical properties but dissimilar topologies. This database has been used in an extremely large number of virtual screening method validation studies (von Korff et al., 2009, Brylinski, 2013, Ge et al., 2013, Sastry et al., 2013, Wei and Hamza, 2013, Awale and Reymond, 2014, Kalászi et al., 2014) including validation of OpenEye FRED and HYBRID (McGann, 2011), and UCSF Dock (Cross et al., 2009). The other type of decoy library that can be used is a generic decoy library such as the Schrodinger GLIDE 1k drug-like ligand decoys set (Friesner et al., 2004, Halgren et al., 2004). Generic decoy libraries do not have compounds that are tailored to a particular target or group of related enzymes but rather rely on having a small number of highly diverse, drug-like compounds.

The protein targets should generally be diverse with regards to binding site shape, size, and electrostatic and hydrogen bonding potential. This is to assess how consistent the screening method is over different types of binding sites. Some validation studies have taken a high-throughput approach by selecting a very large number of protein targets. Salam et al. (2009) tests a screening method against 30 protein targets, Cross et al. (2012) uses a set of 81 targets, and a number of studies have simply used the entire UCSF DUD dataset (Arciniega and Lange, 2014, Awale and Reymond, 2014). Test sets generally do not need to be so large; some recent studies have used as few as 6 to 10 protein targets (Loving et al., 2009, Dixit and Verkhivker, 2012, Wei and Hamza, 2013, Kalászi et al., 2014, Wang et al., 2014).

Finally, the relative success or failure of a screening method can be measured in a number of different ways. There are several elements that are usually analysed: early enrichment, overall bias, hit diversity, consistency and quite often CPU wall time. No method does exceptionally well in all areas and there is usually a large trade-off between these measures. As such, the desired outcomes of a screening method need to be clearly defined. Is the method designed to be fast or accurate? Is a high early enrichment or high hit diversity the goal? Is it supposed to be a consistent, Jack-of-all-trades screening method? Once the aims are established the measures can then be used to compare to other screening methods to determine whether or not it would be better and in what circumstances.

1.6. Drug Discovery: Hit to Lead Strategies

After hit discovery comes a step referred to as 'Hit to Lead' (H2L). The success or failure of a potential drug candidate is heavily influenced by lead compound selection. The cost of developing and modifying a lead compound, and subsequently subjecting it to clinical trials, is also extremely expensive. As such, lead selection is an extremely important step and a large range of confirmed hits from which to choose the lead is highly desired. The lead needs to be: potent enough to work on the target, have the

right properties to make it bioavailable to its target, patentable, and safe for use as a drug. A lot of effort is therefore placed on developing and expanding the pool of hit compounds into lead candidates to allow for the best selection possible.

H2L strategies generally involve minor steps to modify hit compounds in some way to increase potency or improve chemical features. Potency is usually measured biochemically with a series of assays to determine the half maximal inhibitory concentration (IC_{50}), dissociation constant (K_d) or inhibition constant (K_i).

Improvement of chemical features involves changing chemical properties to make the hit compound more drug-like or lead-like, or improve *in vitro* ADME properties (absorption, distribution, metabolism, and excretion) (Keseru and Makara, 2006).

Keseru and Makara (2006) list some popular H2L strategies that include: isosteric replacement (where chemical groups are replaced with similar ones), hit evolution (where chemical groups are added), hit fragmentation (large hits are broken down into smaller ones and undergo fragment-based techniques), fragment linking (fragment hits that occupy neighbouring areas of the active site are bonded together), fragment self-assembly (which is the same as fragment linking but the protein facilitates the bonding using click chemistry (Kolb et al., 2001)), fragment expansion (analogous to hit evolution—chemical groups are added), and hit expansion (ligand-based screening software is used to find structural analogues of hits). Such methods have improved the activity of hit compounds from the high micromolar and even low millimolar range (293 μ M–2.7 mM) to the low micromolar and even nanomolar range (13 μ M–22 nM) (Hajduk et al., 1999, Erlanson et al., 2000, Wang et al., 2002, Liu et al., 2003, Szczepankiewicz et al., 2003, Oltersdorf et al., 2005).

1.7. Aims and Objectives of this Study

There has been limited research into identifying inhibitors of *Pf*OPRTase and *Pf*ODCase and this has predominantly involved the testing of compounds that were structurally analogous to products and substrates. Given the absence of novel scaffolds for these

targets there is a need to undertake comprehensive screening against a large drug-like database. Specific inhibitors for the *P. falciparum* enzymes over the human homologues are highly desired. However, inhibition of the human homologues and inhibition of these enzymes in other organisms has applications in a range of other diseases due to the broad druggability of the pathway. There is therefore a need for more experimentally-solved structural information for these targets, specifically the whole, bifunctional UMPS of higher eukaryotes. Structural information for the *P. falciparum* multienzyme complex between *PfOPRTase* and *PfODCase* is also highly desired. The evidence of substrate channelling and kinetic benefits for the OPRTase-ODCase complexes in various organisms is compelling but not completely understood; features such as dimer interfaces and areas key to substrate channelling may be viable alternative druggable binding sites.

In silico screening is an ever-evolving field and a lot of research has gone into addressing the limitations of structure- and ligand-based methods; old programs are always being updated with the latest screening and scoring algorithms. While modest progress has been made with such structure- and ligand-based methods there has not been a fundamental, game-changing advancement that has truly bridged the gap between the two antithetic approaches to virtual screening.

The main objective of this project was to identify novel, specific inhibitors of *PfOPRTase* and *PfODCase* in the interests of malaria drug discovery. Given the broad druggability of the *de novo* pyrimidine biosynthesis pathway another main objective was included of finding inhibitors of the human homologues in the interests of drug discovery for a range of diseases (outlined in 1.2.2). Secondary to identification of inhibitors was an investigation into the structures of *PfOPRTase* and *HsUMPS* for its application in future drug discovery research. A final main objective was included after the *in silico* screening was performed to develop a method of performing virtual screening that specifically addressed the speed limitation of structure-based screening (which was a

significant limitation in Chapter 4), and the hit diversity and prior knowledge limitation of ligand-based screening. This method could then be used for future virtual screening, including, but not limited to, OPRTases and ODCases.

Chapter 3 outlines the expression and purification of *Pf*OPRTase, *Pf*ODCase, *Hs*OPRTase and *Hs*ODCase for use in biochemical assays for the identification of inhibitors.

Chapter 3 also outlines the crystallography conditions that were tested for *Pf*OPRTase and a method for producing recombinant *Hs*UMPS for use in crystallography studies.

Chapter 4 outlines the virtual screening that was performed on these enzymes; this includes homology modelling of *Pf*OPRTase as the experimentally solved structure was not available at the time. Chapter 5 outlines the assay validation and inhibition assays for these enzymes, and the kinetic characterisation of the ODCase inhibitors. Chapter 6 describes the early development, and the current implementation of a hybrid method of performing virtual screening. The method is validated with a suite of pharmaceutically relevant targets against their known binders and decoy compounds. Chapter 6 also compares the method to a simple ligand-based screening program and a structure-based method.

2. Materials and Methods

2.1. Materials

Tryptone, yeast extract, and bacteriological agar were supplied by Oxoid Ltd. (Basingstoke, England). Ampicillin and kanamycin were obtained from F. Hoffmann-La Roche Ltd. (Basel, Switzerland). Chloramphenicol was obtained from Boehringer Ingelheim (Ingelheim am Rhein, Germany). IPTG was supplied by Fisher Biotech (Wembley, Australia). Ammonium persulphate, Bradford reagent, brilliant blue R, MOPS, PMSF, streptomycin, and tris base were obtained from Sigma-Aldrich (St. Louis, United States).

Acrylamide and SDS were obtained from Amresco (Solon, USA). Agarose, DNA purification kits, DNTPs, Gel/PCR Clean-Up kits, and T4 DNA Ligase were obtained from Promega (Fitchburg, USA). Protein and DNA size markers, *Taq* and Vent_R[®] polymerases, and restriction enzymes were obtained from New England Biolabs[®]*inc.* (Massachusetts, USA). All chromatography resins and columns were obtained from GE Healthcare (Little Chalfont, UK).

X-ray crystallography screens NeXtal PEGs, NeXtal PEGs II, Cryo, and Classics screens were obtained from Qiagen Pty Ltd-Australia (Victoria, Australia). X-ray crystallography screens Crystal Screen and Crystal Screen 2 and crystallography plastic ware and supplies were obtained from Hampton Research (California, USA).

All other chemicals were obtained at reagent grade or better from Chem-Supply Pty Ltd (Adelaide, Australia). Buffers were all produced in the laboratory.

2.2. *E. coli* Strains, Plasmids, and Growth Conditions

2.2.1. *E. coli* Strains and Plasmids

The *E. coli* strains DH5 α (Taylor et al., 1993), DH10B (Grant et al., 1990) and PMC103 (Doherty et al., 1993) were used for cloning experiments and BL21 (DE3) (Studier and Moffatt, 1986) for recombinant expression of proteins. The expression vectors and their respective helper plasmids and antibiotic selection that were used in this study are shown in Table 2.1.

Table 2.1: Expression vectors, helper plasmids and antibiotic selection.

Expression plasmid	Antibiotic selection	Helper plasmid (used for expression)	Helper plasmid antibiotic selection
pET3a- <i>Pf</i> OPRT (Kuehn, 2003)	Ampicillin	pRIG (Baca and Hol, 2000)	Chloramphenicol
pET3a- <i>Pf</i> ODC (Menz et al., 2002)	Ampicillin	pIMICO (Cinquin et al., 2001)	Chloramphenicol
pET30a- <i>Hs</i> OPRT	Kanamycin	pLysS (Studier, 1991)	Chloramphenicol
pET30a- <i>Hs</i> ODC			
pET30a- <i>Hs</i> UMPS			
<i>E. coli</i> Strain		Role	Antibiotic selection
DH5 α		Cloning (human genes)	-
DH10B		Cloning (human genes)	-
PMC103		Cloning (<i>P. falciparum</i> genes)	Streptomycin
BL21 (DE3)		Recombinant expression	-

2.2.2. Growth Conditions

Lysogeny broth (LB) was used for all *E. coli* liquid cultures, unless otherwise stated, consisting of 10.0 g tryptone, 5.0 g yeast extract and 5.0 g sodium chloride per litre of deionised water. The pH was adjusted to 7.5 with sodium hydroxide (Lennox, 1955). LB-agar plates were prepared using LB with 1.5 % (w/v) bacteriological agar. Sterilised 20 % (w/v) glucose was added to a final concentration of 0.1 % (w/v) to the LB prior to inoculation and to the LB-agar media prior to pouring. If antibiotics were needed, these were added to LB before inoculation and to LB-agar prior to pouring. Chloramphenicol and streptomycin were used at a final concentration of 50 $\mu\text{g}\cdot\text{mL}^{-1}$, kanamycin was used at a final concentration of 30 $\mu\text{g}\cdot\text{mL}^{-1}$ and ampicillin was used at a final concentration of 150 $\mu\text{g}\cdot\text{mL}^{-1}$. Liquid cultures were grown at 37 °C in Erlenmeyer flasks with agitation (150–180 rpm) either overnight (16–20 hours) or until the OD_{600 nm} was 0.6–0.8. Inoculated LB-agar plates were incubated overnight (16–20 hours) at 37 °C.

2.3. Molecular Biology and Cloning Techniques

2.3.1. Plasmid Purification

Plasmid DNA was isolated and purified using Promega Wizard® Plus SV Minipreps DNA Purification System. The plasmid purification protocol followed the manufacturer's 'centrifuge protocol' instructions. For all plasmid samples, 10 mL of overnight liquid culture was used, except when purifying pET3a-*PfOPRT* (the pET3a plasmid in which the *P. falciparum* OPRTase gene was cloned). The pET3a-*PfOPRT* plasmid exhibited a consistently low copy number, such that 50 mL of overnight culture had to be used when performing the plasmid purification. As the kit was only designed for 1–10 mL of culture the buffer volumes were scaled up accordingly when purifying from 50 mL of cell culture.

2.3.2. DNA Quantification

DNA concentrations were estimated spectrophotometrically using a Thermo Scientific (Waltham, USA) NanoDrop 1000 spectrophotometer. The 'NanoDrop' was blanked with molecular-grade water and 2 µL of sample was loaded. The calculated concentration of DNA was returned, as were the DNA purity ratios of absorbance at 260 and 280 nm, and, at 260 and 230 nm. A DNA sample was considered contaminant free if the 260/280 absorbance ratio was 1.65–1.9 and the 260/230 absorbance ratio was 1.8–2.2.

2.3.3. PCR Amplification

All PCR amplifications were carried out in a PerkinElmer (Massachusetts, USA) GeneAmp 2400 thermocycler.

2.3.3.1. High-Fidelity PCR

High-fidelity PCRs were carried out in 25 µL reactions consisting of 0.4 U of Vent polymerase, Thermopol buffer to 1 × concentration, 40 µM of dNTPs, 1.6 µM oligo

primers and approximately 4–8 ng·μL⁻¹ of template DNA. The thermocycler was set to 25 cycles. Temperatures and step lengths for each gene are tabled in Appendix 8.1.

2.3.3.2. Analytical PCR

Analytical PCRs were carried out in 25 μL reactions consisting of 0.6 U of *Taq* Polymerase, Thermopol buffer to 1 × concentration, 40 μM of dNTPs, 1.6 μM oligo primers and, an unquantified amount of template DNA for colony screen PCR (2.3.10), or approximately 4–8 ng·μL⁻¹ of template DNA where the quantity was known. The thermocycler was set to 35 cycles. Temperatures and step lengths for each gene are tabled in Appendix 8.2.

2.3.4. Purification of Restriction Enzyme Digests and PCR Amplicons

Restriction enzyme (RE) digests and PCR reactions were purified using Promega Wizard® SV Gel and PCR Clean-Up System. PCR purification followed the manufacturer's instructions for 'PCR clean-up Protocol'.

2.3.5. Restriction Enzyme Digests for cloning

Restriction enzyme double-digests were carried out on *H. sapiens* OPRTase, ODCase, and UMPS high-fidelity PCR amplicons, and purified pET30a vectors using the type II REs *Not* I and *Nco* I. The 20 μL reactions contained 5 U each of *Not* I and *Nco* I, 1 mg of either PCR amplicon or pET30a plasmid DNA, and the manufacturer's buffer. The reactions were incubated at 37 °C for approximately 4 hours and terminated by incubation at approximately 60 °C for 3–5 minutes. Reactions were then purified as per Section 2.3.4.

2.3.6. DNA Ligations

Promega T4 DNA Ligase was used for DNA ligation reactions. The reaction contained 200–400 ng of vector DNA, insert DNA to a vector:insert molar ratio of 1:3, ligase buffer to 1 × concentration, and nuclease-free water to adjust to a final volume of 20 μL. The

reaction was incubated overnight at 4–8 °C. The mass of insert DNA was calculated using the formula:

$$\frac{ng\ of\ vector\ \times\ kb\ size\ of\ insert}{kb\ size\ of\ vector} \times molar\ ratio\ of\ \frac{insert}{vector} = ng\ of\ insert$$

2.3.7. Preparation of Competent *E. coli* Cells

E. coli cells were made heat-shock competent for plasmid transformation. A 10 mL liquid culture was inoculated with a single colony grown on an LB-agar plate and the culture was incubated overnight at 37 °C. A 50 mL culture was inoculated with 1 mL of the overnight liquid culture. The 50 mL shaking culture was grown until OD_{600 nm} was 0.6–0.8 (as per Section 2.2.2) and harvested by centrifugation at 4000 × *g* for 10 minutes at 4 °C. The cells were re-suspended in 20 mL of ice-cold 0.1 M CaCl₂ and incubated on ice for approximately 30 minutes. The cells were then harvested again by centrifugation at 4000 × *g* for 10 minutes and re-suspended in 2 mL of ice-cold 0.1 M CaCl₂. The heat-shock competent cells were kept refrigerated at 4 °C for up to one week for use with heat-shock transformation.

2.3.8. Heat-shock Transformation

Heat-shock transformation of competent *E. coli* cells with plasmid DNA was carried out using cells produced as per Section 2.3.7 and plasmid isolated as per Section 2.3.1. Approximately 100 ng of plasmid DNA was added to 50 µL of heat-shock competent cells in a microfuge tube and gently mixed. This was incubated on ice for 30 minutes. The cells were then heat-shocked via incubation in a dry heating block at 42 °C for 90 seconds after which they were immediately placed back on ice. An aliquot of 50 µL of LB was added to the cells which were then incubated without agitation at 37 °C for approximately 45 minutes. The transformed cells were then plated on antibiotic selective LB-agar plates and grown overnight as per Section 2.2.2.

2.3.9. Electroporation Transformation of *E. coli*

2.3.9.1. Preparation of Electrocompetent Cells

Electrocompetent *E. coli* DH10B cells had been prepared according to Sambrook et al. (1989). A 1 L LB culture was inoculated with 10 mL of the overnight liquid culture. The culture was grown until OD_{600 nm} was 0.4–0.6 (as per Section 2.2.2) and harvested by centrifugation at 4000 × *g* for 10 minutes at 4 °C (split into four 250 mL bottles). The four pellets were each resuspended in 200 mL of ice cold Milli-Q water (Merck Millipore, Darmstadt, Germany) and centrifuged a second time using the same conditions. The four pellets were then each resuspended in 100 mL of ice cold Milli-Q water and centrifuged a third time using the same conditions. The four pellets were combined by resuspension in 80 mL of ice cold 10 % glycerol and centrifuged a fourth time using the same conditions. The supernatant was carefully aspirated. The pellets were resuspended in 2 mL of ice cold 10 % glycerol. The suspension was aliquoted, 50 µL aliquots were made and snap frozen in liquid nitrogen.

2.3.9.2. Electroporation

E. coli cells were electroporated using a method similar to Sambrook et al. (1989). First, 40 µL of electrocompetent cells were incubated with 25 ng of plasmid on ice for 30–60 seconds and pipetted into the electroporation chamber. The Bethesda Research Laboratories Cell-Porator was set according to the manufacturer's instructions for *E. coli* for a 1 mm chamber gap size. The cells were immediately removed following electroporation, 50 µL of LB added, and the cells were incubated without agitation at 37 °C for 1 hour. The cells were then grown on selective LB-agar plates as per Section 2.2.2.

2.3.10. Colony PCR screening

Colonies of freshly transformed cells were routinely checked by PCR to confirm they contained the gene of interest. To prepare the DNA template sample for the PCR reaction 20 µL of sterilised Milli-Q water was dispensed into a microfuge tube. A yellow

micropipette tip (suitable for the 2–20 μL and 20–200 μL micropipettes), while attached to the 2–20 μL pipette, was used to 'pluck' part of a bacterial colony. This was then pipette mixed into the 20 μL of sterilised Milli-Q water by drawing up and dispensing the 20 μL approximately 10 times. The 20 μL of suspended cells were incubated at 100 °C for 5 minutes. This lysed the cells and helped to denature any proteins in the sample. A 2 μL aliquot of this preparation was then used as the template for a 25 μL DNA reaction for 35 PCR cycles (Section 2.3.3.2).

2.3.11. DNA Sequencing

All sequencing was performed by the Australian Genome Research Facility's (AGRF) (agrif.org.au) routine sequencing of purified DNA service. For each expression vector to be sequenced, two reactions using a flanking primer were prepared; one with a forward primer and one with the reverse complement primer. Primers for the T7 RNA polymerase promoter and terminator region were used to ensure the entire gene would be sequenced. To each reaction, 1500 ng of plasmid was added with 9.6 pmol of either the forward or reverse primer as per AGRF's sample submission requirements.

2.3.12. Agarose Gel Electrophoresis

All DNA samples were separated on a 1 % (w/v) agarose gel with TAE electrophoresis buffer (40 mM Tris-acetate pH 8.3, 1 mM EDTA). Samples were prepared by the addition of 6 \times DNA loading Buffer (60 % (v/v) Glycerol, 2.5 $\text{mg}\cdot\text{mL}^{-1}$ Bromophenol Blue). Where DNA concentration had been quantified, approximately 100–200 ng of DNA was loaded. Either Promega or New England Biolabs 1 kb DNA ladder was added for sizing of DNA bands. Gels were set to run at a constant voltage of 86 V and were stopped when the dye front had migrated approximately 80–90 % of the length of the gel.

2.4. Recombinant Protein Techniques

2.4.1. Recombinant Protein Expression in *E. coli*

Recombinant protein expression was carried out in *E. coli* BL21 (DE3) host cells using the Novagen pET expression system (Merck Millipore, Darmstadt, Germany) (all vectors were created from either pET3a or pET30a). Positive colonies were grown in 2×TY medium (16 g·L⁻¹ tryptone, 10 g·L⁻¹ yeast extract, 5 g·L⁻¹ NaCl) (Sambrook et al., 1989) + 0.01 % (w/v) glucose, until OD_{600 nm} was 0.5–0.8 and induced with 0.5 mM IPTG.

Trial expressions were carried out in either 50 mL or 200 mL cultures. Large-scale expressions were carried out in 4 L batches in a BioFlo® 110 New Brunswick Scientific bioreactor (Eppendorf-New Brunswick, Enfield, USA). Sterile air was added at approximately 7 L·min⁻¹ throughout growth and expression. Dissolved oxygen was maintained at 95 % by the adjustment of agitation (150–750 rpm). The pH was maintained at 7.0 with the addition of ammonium hydroxide. Induction was carried out at 37 °C for 1 hour for the 50 mL and 200 mL trial expressions, and at either 18 °C or 22 °C for 16–18 hours for the 4 L large-scale expressions.

2.4.2. Harvesting of *E. coli* Cells

Cells were cooled to 4 °C after the induction period and pelleted by centrifugation at 4000 × *g* for 10 minutes at 4 °C. The cells were then resuspended in Buffer A (50 mM of either Tris-Cl or Sodium phosphate pH 8.0, 150 mM NaCl—Appendix 8.3) at a ratio of 50 mL of Buffer A per litre of culture harvested. Re-suspended cells were stored frozen at –20 °C until needed.

2.4.3. Lysing of *E. coli* Cells

Harvested *E. coli* cells were lysed using an Avestin (Ottawa, Canada) Emulsiflex C-5 homogeniser. The input pressure was adjusted to 80–100 PSI and the cells were passed once with no homogenising pressure to ensure that the cells were homogeneously

suspended. The cells were then passed through twice at a homogenising pressure of 5 000–10 000 PSI. PMSF was added to a final concentration of 1 mM just prior to lysing. A change in colour and viscosity was observed consistent with lysed *E. coli* cells. The lysate was placed immediately back on ice.

2.4.4. Clearing of Cell Lysate

E. coli cell lysate was cleared using centrifugation and filtration. Lysate was first centrifuged at $4000 \times g$ for 10 minutes at 4 °C. The centrifuge-supernatant was then ultra-centrifuged at $150\,000 \times g$ for 1 hour at 4 °C. The ultracentrifuge supernatant was then filtered using Sartorius Minisart High-Flow syringe filters of 0.20 µm pore size. The cleared lysate was then placed back on ice and Buffer B (Buffer A + 500 mM Imidazole, Appendix 8.3) was added to obtain a final concentration of imidazole of 10 mM unless stated otherwise.

2.4.5. Nickel-Affinity Chromatography

Recombinant protein was purified from cleared cell lysate using nickel affinity chromatography. A 5 mL column packed with GE Healthcare Ni Sepharose 6 Fast Flow resin was used on a GE Healthcare ÄKTA explorer 100 Fast Protein Liquid Chromatography (FPLC) system. The column was equilibrated with 5 Column Volumes (CV) of Buffer A (Appendix 8.3). The cleared lysate was then loaded and washed with 40 CV of Buffer A + Buffer B to an imidazole concentration of 40–50 mM. The recombinant protein was eluted with 5 CV of Buffer B. All chromatography procedures were carried out with a flow rate of $1 \text{ mL}\cdot\text{min}^{-1}$ at 4 °C.

2.4.6. Sodium Dodecyl Sulphate Polyacrylamide Gel Electrophoresis (SDS-PAGE)

Tris-HCl buffered SDS-PAGE gradient gels (7–20 %) were used. SDS-PAGE gels were precast as a stack of 10 gels at a time. The 20 % solution consisted of 14 mL of 40 % (w/v) acrylamide, 7 mL of 1.5 M Tris-HCl (pH 8.8), 280 µL of 10 % (w/v) SDS, and

6.72 mL of distilled water. The 7 % solution consisted of 4.9 mL of 40 % (w/v) acrylamide, 7 mL of 1.5 M Tris-HCl (pH 8.8), 280 μ L of 10 % (w/v) SDS, and 15.82 mL of distilled water. The 4 % stacking solution consisted of 4.5 mL of 40 % (w/v) acrylamide, 10.5 mL of 0.5 M Tris-HCl (pH 8.8), 450 μ L of 10 % (w/v) SDS and 29.55 mL of distilled water.

The gradient gels were poured using a Hoefer® (Holliston, USA) Mighty Small multiple gel caster following the manufacturer's instructions.

2.4.6.1. Coomassie Staining

SDS-PAGE gels for Coomassie staining were incubated with gentle agitation at room temperature for 2–8 hours in Coomassie stain solution (40 % methanol, 10 % glacial acetic acid, 0.1 % (w/v) Coomassie Brilliant Blue R-250). Gels were then destained at room temperature with gentle agitation in destain solution (40 % methanol, 10 % glacial acetic acid) until the gel background was clear. The destain solution was replaced as necessary.

2.4.6.2. Silver Staining

Silver staining was carried out using a method similar to that outlined in Ogut and Jin (2000). Silver staining of SDS-PAGE gels consisted of incubating the gel in a series of solutions. Each wash step was carried out at room temperature with gentle agitation. The solutions used were: Solution 1 (50 % methanol, 12 % (w/v) trichloroacetic acid, 2 % (w/v) CuCl_2), Solution 2 (10 % ethanol, 5 % glacial acetic acid), Solution 3 (0.01 % (w/v) potassium permanganate), Solution 4 (10 % ethanol), Solution 5 (0.1 % (w/v) AgNO_3), and Solution 6 (2 % (w/v) K_2CO_3 , 0.04 % formaldehyde). The wash steps were as follows: Solution 1 for 5 minutes, Solution 2 for 5 minutes, Solution 3 for 5 minutes, Solution 2 for 1 min, Solution 4 for 5 minutes, Water for 5 minutes, Solution 5 for 5 minutes, Solution 6 for 5–10 minutes (developing step: during this step the bands appear, the time varies depending on the desired intensity of the bands), Water for 10

minutes (This step stops the developing of the gel bands, water was changed 2–3 times during this step).

2.4.7. Western Blotting

2.4.7.1. Membrane Transfer

The transfers for western blots were carried out using a Bio-Rad Trans-blot® SD Semi-Dry Transfer Cell. Millipore Immobilon-P, 0.45 µm, PVDF transfer membranes were used. The membrane, SDS-PAGE gel, and 10 × gel blotting papers were soaked for 30 minutes in Semi-Dry Transfer Buffer (20 % Methanol, 25 mM Tris base, 152 mM Glycine, pH 8.0). The transfer was then arranged in the cell in the following way: anode base, 5 × blotting papers, PVDF membrane, SDS-PAGE gel, 5 × blotting papers, cathode top plate. Transfers were carried out at maximum 22 V and at 0.8 mA·cm⁻² (of SDS-PAGE gel) for 1.5 hours.

2.4.7.2. Blotting and Detection

Blocking buffer was prepared by dissolving 50 g·L⁻¹ of skim milk powder in TBST (25 mM Tris-HCl pH 7.4, 140 mM NaCl, 2.7 mM KCl, 0.1 % Tween-20) and centrifuging at 4000 × *g* for 20 minutes to remove any undissolved particulate material and stored at 4 °C. TBST and blocking buffer were prepared fresh and used no later than one week after being prepared.

Membranes were blocked overnight at 4 °C in 10 mL of blocking buffer with gentle rocking. Subsequent steps were carried out at room temperature with gentle agitation. Membranes were washed five times with 10 mL of TBST for 5 minutes per wash. They were then incubated with 10 mL of TBST containing 1:2 000 Rockland (Limerick, USA) rabbit Anti 6× His tag primary antibody for 1 hour at room temperature. Membranes were then washed five times with 10 mL of blocking buffer for 5 minutes per wash. They were then incubated in 10 mL of blocking buffer containing 1:5 000 Rockland goat anti rabbit horseradish peroxidase-conjugated secondary antibody for 1 hour at room

temperature. Membranes were then washed five times with TBST for 5 minutes per wash.

Blotted membranes were treated with Thermo Scientific SuperSignal® West Pico Chemiluminescent Substrate using the manufacturer's instructions. Membranes were then exposed to Kodak X-Omat™ K XK-1 Diagnostic Film for 30 minutes, and developed using the manufacturer's instructions and solutions.

2.4.8. Protein Estimations

Protein samples for biochemical assays were estimated using the method of Bradford (1976) using a dilution series of bovine serum albumin (consisting of 0, 0.2, 0.4, 0.6, 0.8 and 1 mg·mL⁻¹) and Sigma-Aldrich's premixed Bradford Reagent.

Concentrated protein samples for crystallisation screens were estimated using a Thermo Scientific NanoDrop 1000. The 'Protein A280' module was used. The 'NanoDrop' was blanked with the same buffer that the protein was in and 2 µL of sample was loaded. The calculated protein concentration in mg·mL⁻¹ was returned as well as the ratio of sample absorbance at 260 and 280 nm.

2.4.9. Protein Crystallography and X-Ray Diffraction

Samples of *PfOPRTase* at a concentration of approximately 5 mg·mL⁻¹ were subject to screening for crystallisation conditions. An Art Robbins Instruments Phoenix Liquid Handling System (Sunnyvale, USA) was used to prepare 96-well sitting-drop screening plates with 500 nL droplets of 1:1 protein sample to crystallisation solution. The following Screens were used: NeXtal's PEGs, PEGs II, Cryo, Classics and Hampton Research's Crystal Screen and Crystal Screen 2. Crystallisation plates were incubated at 20 °C. After preparation the plates were observed for protein crystals at 1 hour, 24 hours, 48 hours, 1 week, and weekly thereafter using a dissecting microscope with polarising filter.

Conditions that yielded possible protein crystals were scaled up to 24-well sitting-drop screening plates with 2 μ L droplets of 1:1 protein sample to crystallisation solution.

The diffraction properties of the crystals were analysed using the High-throughput PX and Micro Crystallography beamlines at the Australian Synchrotron, Melbourne, Australia.

2.5. Biochemical Assays and Kinetics

2.5.1. Preparation of Compounds for Screening

Compounds identified by *in silico* screening that required biochemical screening for inhibition, were obtained as a dry powder or salt from ChemDiv (San Diego, USA). Compounds were dissolved in dimethyl sulfoxide (DMSO) to yield stock solutions at a concentration of 100 mM. For compounds that did not fully dissolve, more DMSO was added to yield 50 mM stocks and 25 mM stocks where necessary. Further dilutions of the compounds were made with DMSO when necessary for the purposes of pipetting small amounts for use in assays.

2.5.2. OPRTase Biochemical Assay

A modified method of the spectrophotometric activity assay previously described (Han et al., 1995, Yablonski et al., 1996) was used to assay *Pf*OPRTase and *Hs*OPRTase for activity. The assay consisted of 25 mM Tris-HCl (pH 8.0), 2.5 mM MgCl₂, 1 mM PRPP and 0.25 mM orotate. The reaction was carried out in 96-well UV-plates at 37 °C, initiated with the addition of PRPP and measured spectrophotometrically following the consumption of orotate as a decrease in absorbance at 295 nm. For *Pf*OPRTase, 0.75 µg of purified protein was used in the assay. For *Hs*OPRTase, 1.38 µg of purified protein was used. DMSO was added to the controls to achieve the same % (v/v) as the inhibition assays. All assays were performed in duplicate.

Specific activities were calculated by applying Beer-Lambert's law with the extinction coefficient for orotate at 295 nm of 3670 mol·L⁻¹·cm⁻¹ (Yablonski et al., 1996).

2.5.3. ODCase Biochemical Assay

The method for performing the ODCase activity assays was similar to Yablonski et al. (1996). The assay consisted of 25 mM MOPS (pH 7.0), 100 mM NaCl and 0.39 mM OMP. An assay volume of 250 µL was used. The assay was carried out in 96-well UV-plates at 30 °C, initiated with the addition of OMP and measured spectrophotometrically

following the consumption of OMP as a decrease in absorbance at 295 nm. For *Pf*ODCase, 0.7 ng of protein was used in the activity assays. For *Hs*ODCase, 0.6 ng of protein was used. All assays were performed in duplicate.

The extinction coefficient for the difference between OMP and UMP at 295 nm was calculated by allowing assays to run to completion and comparing the absorbances:

$$\varepsilon = \frac{A_{OMP} - A_{UMP}}{[S]}$$

A_{OMP} : the absorbance at 295 nm of the assay without the enzyme

A_{UMP} : the absorbance at 295 nm of the assay that has run to completion

[S]: the known concentration of OMP in the assay

ε : the extinction coefficient

Under standard conditions the millimolar extinction coefficient for the difference between OMP and UMP at 295 nm and 20 °C was calculated to be 770 mol·L⁻¹·cm⁻¹. DMSO was added to the controls to achieve the same % (v/v) as the inhibition assays.

2.5.4. Reaction Progress Kinetic Analysis of ODCase Inhibitors

Due to the low K_m that the ODCases exhibit, OMP kinetic analysis could not be conducted spectrophotometrically by calculating initial reaction rates at a number of different substrate concentrations. A method is described for performing assays using ¹⁴C radiolabelled OMP (Shostak and Jones, 1992, Yablonski et al., 1996), however as radiolabelled OMP or orotate was not commercially available at the time of the study an alternative assay was required.

A simple method for reaction progress kinetic analysis was instead used (Blackmond, 2005). With this method, a reaction is allowed to run to completion and the velocities were derived from an arbitrary function fit to the data. Assays for reaction progress kinetic analysis of ODCases were as detailed in Section 2.5.3 but with a starting OMP concentration of 100 μM. Assays were performed in triplicate with no inhibitor and two

different concentrations of inhibitor. Inhibitor concentrations were chosen based on the following: capturing of velocities below and above estimated 50 % inhibition, availability of compound, and minimisation of noise at $OD_{295\text{ nm}}$.

The assays were measured in a BMG Labtech FLUOstar Omega spectrophotometer (Ortenberg, Germany) using well-mode kinetics. Parameters were optimised to get the maximum number of readings (1000) over 15 minutes (75 'flashes' per reading, 0.78 s interval time). The raw data was imported into GraphPad Prism (graphpad.com) for curve fitting. A 3rd, 4th, or 5th order polynomial, or a 1st order exponential decay function was fitted (whichever fit the data best). A baseline was drawn at the negative plateau. Velocities were derived as tangents from the fitted curve at various points along the curve; the substrate concentrations for each point were calculated from the difference in absorbance to the baseline using Beer-Lambert's law. An example of this can be seen in Appendix 8.4. Derived velocities from the three curves were analysed using GraphPad Prisms' Michaelis-Menten kinetics model. All inhibition mechanisms (competitive, non-competitive, uncompetitive, and mixed-mode inhibition) were applied to assess which model best fit the data (Appendices 8.5–8.9). The best fitting inhibition models are shown in the results; where uncompetitive and mixed-mode inhibition models fit similarly well the uncompetitive inhibition model was used.

As this method for deriving kinetic data from progress curves does not use true initial rates, the kinetic constants (K_m) and inhibition constants (K_i or αK_i) are not likely to be accurate. Instead apparent K_m (K_m^{app}) and apparent $K_i/\alpha K_i$ ($K_i^{\text{app}}/\alpha K_i^{\text{app}}$) are reported.

2.6. List of Software

2.6.1. Protein Structure Viewing and Imaging

UCSF Chimera (v1.5.3) (cgl.ucsf.edu/chimera/) (Pettersen et al., 2004): PDB structure files were viewed and edited using UCSF Chimera. All structure files used for ligand docking were formatted using UCSF Chimera prior to docking with UCSF Dock.

OpenEye VIDA (v4.1.1) (eyesopen.com): OpenEye VIDA was used for visual analysis of small molecule screening results produced with the OpenEye bioinformatics development suite programs: ROCS, FRED and EON. It was also used for formatting and exporting screening results for manual rescoring in the hybrid screening approach.

2.6.2. Sequence Alignments

ClustalX (v2.0.3) (Larkin et al., 2007): Multiple sequence alignments were performed using ClustalX. The generated multiple sequence alignments were further used in comparative protein modelling using Modeller 9v3.

2.6.3. Comparative Modelling of Protein

Modeller 9v3 (salilab.org/modeller/) (Fiser et al., 2000, Sali and Blundell, 1993): Modeller was used for comparative (homology) modelling of protein structures.

CCP4 (v6.1.13): PROCHECK (ccp4.ac.uk) (Laskowski et al., 1993): PROCHECK (part of the CCP4 suite of molecular modelling applications) was used to assess the stereochemistry of homology models generated by Modeller 9v3.

2.6.4. Structure-Based Screening

UCSF Dock 6.2 (dock.compbio.ucsf.edu/) (Moustakas et al., 2006): Dock was the default molecular docking application. Dock was used for high throughput docking (Chapter 4) as well as probe docking of enzyme active sites for the hybrid screening (Chapter 6). It was also used to rescore probe molecule poses from OpenEye FRED.

OpenEye FRED (v2.2.5) (eyesopen.com) (McGann, 2011): FRED (Fast Rigid Exhaustive Docking) is OpenEye's flagship molecular docking program and was used for probe docking of a water molecule into binding sites for hybrid screening.

2.6.5. Ligand-based Screening

OpenEye ROCS/vROCS (v3.1.2) (eyesopen.com) (Grant et al., 1996): ROCS is OpenEye's flagship ligand-based screening program. vROCS is a user interface to ROCS and is also used to prepare query files (or models) for use with ROCS. It was used to generate and screen ligand-based models for the hybrid screening.

OpenEye OMEGA (v2.4.3) (eyesopen.com) (Hawkins et al., 2010): OMEGA is a file preparation utility for molecular compound libraries. It generates multi-conformers of ligands prior to screening with OpenEye ROCS, FRED and EON (which do not flexibly screen ligands). It was used to prepare all compound libraries that were used by these programs.

OpenEye EON (v2.1.0) (eyesopen.com) (Muchmore et al., 2006): EON was used to screen compounds for electrostatic similarity as part of the hybrid screening protocol.

MOLPRINT 2D (v1.2) (molprint.com) (Bender et al., 2004): MOLPRINT 2D was used for hit expansion in Chapter 5 and as a ligand-based screening control in Chapter 6.

3. Cloning, Expression and Purification of Recombinant Enzymes

3.1. Introduction

As outlined in Section 1.7 the project aims included identifying novel inhibitors of the human and *P. falciparum* OPRTase and ODCase, and attempting to obtain structural information about *Pf*OPRTase and *Hs*UMPS. Purified active enzyme was required primarily for use in inhibitions assays and x-ray crystallography.

As no experimentally solved structure was available for *Pf*OPRTase at the time the project commenced, an attempt was made to find crystallisation conditions that would yield x-ray diffracting crystals for this protein. While many structures were available for *Hs*OPRTase and *Hs*ODCase, no structure of the whole, bi-functional human UMPS was available. Therefore another aim was to clone and express human UMPS in *E. coli*, so that the resulting protein could also be utilised for crystallography, with the intent of determining the 3D structure of human UMPS.

Previous work had developed the expression vector and outlined a method to express and purify *Pf*ODCase to homogeneity with the inclusion of a step to remove the poly-histidine-tag (Menz et al., 2002). For *Pf*OPRTase two expression vectors were available (full length with, and a truncated version without the native 66 amino acid N-Terminal insert) and a method had been developed which resulted in a low level expression of *Pf*OPRTase (with the insert) that was confirmed by SDS-PAGE and Western-Blot analysis (attempts to express the enzyme without the insert had been unsuccessful and was not attempted in this study) (Kuehn, 2003, Van Ngyuen, 2005). There was a need for the fermentation, expression and purification conditions to be optimised to obtain both yield and purity suitable for use in assays and crystal screens. Published work also outlined a similar recombinant method for expressing and purifying *Pf*OPRTase and *Pf*ODCase (Krungkrai et al., 2004a, Krungkrai et al., 2005, Krungkrai et al., 2004b).

Expression vectors for the *H. sapiens* OPRTase and ODCase domains of UMPS were not readily available and therefore needed to be constructed. The *Hs*OPRTase and *Hs*ODCase domains of *Hs*UMPS are known to readily express in a correctly folded and active form in *E. coli*; crystal structures available list *E. coli* as the expression host, as do several papers (Wittmann et al., 2008, Bello et al., 2009, Heinrich et al., 2009, Moche et al., 2009, Meza-Avina et al., 2010, Lewis et al., 2011, Purohit et al., 2012). Published literature includes methods to extract and purify UMPS from human placenta (Livingstone and Jones, 1987), expression and immunoaffinity purification using baculovirus expression system (Han et al., 1995), and active protein was expressed (but not purified beyond cell lysate) in Suchi et al. (1997). Based on this evidence it was decided that expression in *E. coli* and purification by nickel chromatography would be most appropriate for the recombinant production of the *H. sapiens* enzymes.

3.2. Results

3.2.1. Cloning of *H. sapiens* UMPS, OPRTase and ODCase into pET30a

The cDNA of the *H. sapiens* UMPS gene was purchased from GeneCopoeia (Rockville, USA). The cDNA supplied was cloned in the pOTB7 DNA vector. The vector was recovered from storage medium following the manufacturer's instructions and transformed into *E. coli* DH10B cells by electroporation (Section 2.3.9). Transformed cells were grown as per Section 2.2.2 and stocks of these transformed cells were kept at -80 °C. A plasmid purification (Section 2.3.1) was performed to provide DNA that was used as the template in High-Fidelity PCR.

Four primers were designed to amplify the *H. sapiens* UMPS as well as the individual OPRTase and ODCase domains. Domain boundaries used for OPRTase and ODCase primers were taken from published literature (Wittmann et al., 2008) and previous crystal structures (Moche et al., 2007b, Moche et al., 2007a). The translated protein sequence with identified domains is shown in Appendix 8.10. The primers, the enzymes they amplify, and the incorporated restriction sites are shown in Table 3.1 and

Figure 3.1. All primers were designed with Oligo Analyser 2.0 (sg.idtdna.com) to ensure similar melting temperatures and that no significant hairpins or dimers were predicted to occur. The forward primers were checked to ensure the genes would be in-frame with the N-terminal tags and cleavage sites on the pET30a vector when translated (pET30a vector shown in Appendix 8.11).

High-Fidelity PCR (Section 2.3.3.1) was used to amplify *HsUMPS*, and the OPRTase and ODCase domains. The resulting amplicons were purified as per Section 2.3.4. A plasmid purification was performed on a culture of DH10B pET30a cells. A double-digest was performed with *Not* I and *Nco* I (Section 2.3.5) for the inserts (PCR amplicons) and the vector (isolated pET30a plasmid). The inserts were ligated into the vector (Section 2.3.6) to yield expression vectors for *H. sapiens* UMPS and the OPRTase and ODCase domains of UMPS, referred to as pET30a-*HsUMPS*, pET30a-*HsOPRT* and pET30a-*HsODC* respectively.

The vectors were sent for routine DNA sequencing (Section 2.3.11) using T7 promoter and terminator primers. Analysis of the resulting sequence confirmed that the sequence of the expression vectors was consistent with the vector design strategy (Sequencing results in Appendices 8.12–8.14).

Table 3.1: Primers used in producing expression vectors of *H. sapiens* UMPS, OPRTase and ODCase. The restriction endonuclease cleavage sites are shown in blue and orange for *Nco* I and *Not* I respectively.

Primer	Sequence	Amplicon(s)	Restriction Site
UMPS_F	5'- GTA TCC ATG GCG GTC GCT CGT GCA G -3'	UMPS & OPRTase	<i>Nco</i> I
UMPS_B	5'- CAA GGC GGC CGC TCA AAC ACC AA TCT ACT -3'	UMPS & ODCase	<i>Not</i> I
OPRT_B	5'- GTA GGC GGC CGC TCA AGA ACC ATT ATG ATT -3'	OPRTase	<i>Not</i> I
ODC_F	5'- GCC ATG GAA CTC AGC TTC GGT GC -3'	ODCase	<i>Nco</i> I

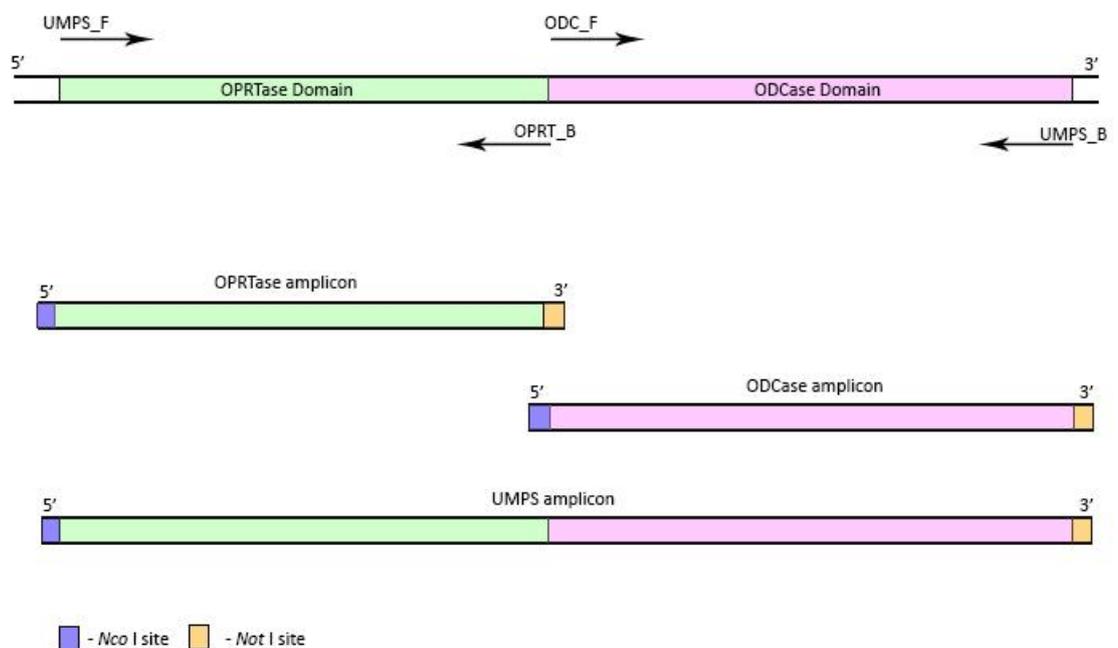


Figure 3.1: Use of primers for producing *H. sapiens* OPRTase, ODCase, and UMPS amplicons from cDNA by PCR.

3.2.2. Expression and Purification of Recombinant Enzymes

Transformation was carried out in *E. coli* BL21(DE3)pRIG cells for *PfOPRTase*; *E. coli* BL21(DE3)pIMICO cells for *PfODCase*; and *E. coli* BL21(DE3)pLysS cells for *HsOPRTase*, *HsODCase*, and *HsUMPS*. Cells were transformed (Section 2.3.8) and colonies were screened via colony screen PCR (Section 2.3.10). Cells were grown and recombinant expression carried out (Section 2.4.1), cells harvested and cleared (Sections 2.4.2 and 2.4.3), and the cleared lysate (Section 2.4.4) was subjected to nickel affinity chromatography (Section 2.4.5). Table 3.2 shows the specific details for expression and purification of the different enzymes. Biochemical activity assays were carried out and specific activities calculated following Sections 2.5.2 and 2.5.3.

Table 3.2: Enzyme specific expression and purification conditions.

Enzyme	Fermenter OD_{600 nm} at Induction	Induction Temp. (°C)	Induction length (hours)	Lysate loaded onto column (mL)	Imidazole wash conc. (mM)
<i>PfOPRTase</i>	0.5	18	18-20	200	50
<i>PfODCase</i>	0.8	22	16	50	50
<i>HsOPRTase</i>	0.8	22	16	50	50
<i>HsODCase</i>	0.8	22	16	50	50
<i>HsUMPS</i>	0.8	18	18-20	100	40

3.2.2.1. *PfOPRTase*

The growth rate of *E. coli* BL21(DE3) cells when transformed with pET3a-*PfOPRT* was significantly slower than for the same cell line transformed with any of the other recombinant expression plasmids. The generation time (as estimated by OD_{600 nm}) was approximately 30 minutes compared to approximately 20 minutes before transformation or when transformed with the other expression plasmids. As a consequence, the overnight plate cultures yielded very small colonies. Occasionally a colony would be significantly larger than the others. In these cases colony PCR screening revealed that the *PfOPRTase* gene was no longer present, suggesting that a recombination event had occurred which resulted in the AT-rich *P. falciparum* gene being removed from the expression plasmid.

The FPLC chromatogram (Figure 3.2) shows the expected shift in the OD_{280 nm} baseline that coincides with the increase in the concentration of imidazole during elution (comparing fractions 0–8 with fractions 12–20). This occurs as imidazole absorbs light at 280 nm. The protein during elution can be seen as a small peak of approximately 50 milli Absorbance Units (mAU) from fractions 8–10. The Coomassie stained SDS-PAGE of the fractions purified by Nickel affinity chromatography shows a single band of approximately 35 kDa (Figure 3.2). This is consistent with the presence of *PfOPRTase* which has a predicted molecular weight of 35.9 kDa (including the 6×His tag). Peak fractions from the Nickel affinity column were pooled and concentrated using a Sartorius Vivaspin 20 Centrifugal Concentrator (Göttingen, Germany) to approximately 1 mL for use with activity and inhibition assays. Bradford assays (Section 2.4.8) were used to determine protein concentrations. The batch that was used for biochemical assays had a protein concentration of 0.15 mg·mL⁻¹; this sample was aliquoted and stored at –80 °C. The batches that were used for crystallography (Section 3.2.5) were concentrated to approximately 200 µL which typically yielded a protein concentration of 4–5 mg·mL⁻¹. Expression levels were generally poor and varied from batch to batch.

The total yield of protein was always far less than it was for the other enzymes produced in this project.

Coomassie staining (Section 2.4.6.1) of a typical purified sample (Figure 3.3) shows a greater than 90 % purity as estimated using Gel Analyzer (gelanalyzer.com). A more sensitive silver staining (Section 2.4.6.2) shows the presence of some contaminating proteins with molecular weights of approximately 70, 25, 15, 10, and 5 kDa. It should be noted that purity is being estimated by the relative proportion of the target compared to the total protein as determined by Coomassie stained SDS-PAGE rather than as a fold increase in specific activity, which is impossible to determine due to the presence of endogenous *E. coli* OPRTase.

Further purification was attempted by cleavage of the N-terminal histidine tag. The enzyme was treated with thrombin following the manufacturer's instructions. It was then passed through a nickel chromatography column and the flow-through (containing the cleaved *Pf*OPRTase) and imidazole eluate (containing uncleaved *Pf*OPRTase) were collected and subject to SDS-PAGE, Coomassie stain (Section 2.4.6), and western blot (Section 2.4.7) (method details and data not shown). While this step did improve purity the yield was far too low for use with assays or other experiments. Further purification was also attempted using size exclusion chromatography. This was attempted on a GE Healthcare HighLoad® 16/60 Superdex® 200 column (methods details and data not shown). This typically resulted in a very minor improvement in purity by the removal of some of the lower molecular weight proteins that can be seen in Figure 3.3. The higher molecular weight contaminating proteins were still present and the total yield had dropped by approximately 20–30 %. As such the concentrated nickel fractions were used for all subsequent experiments.

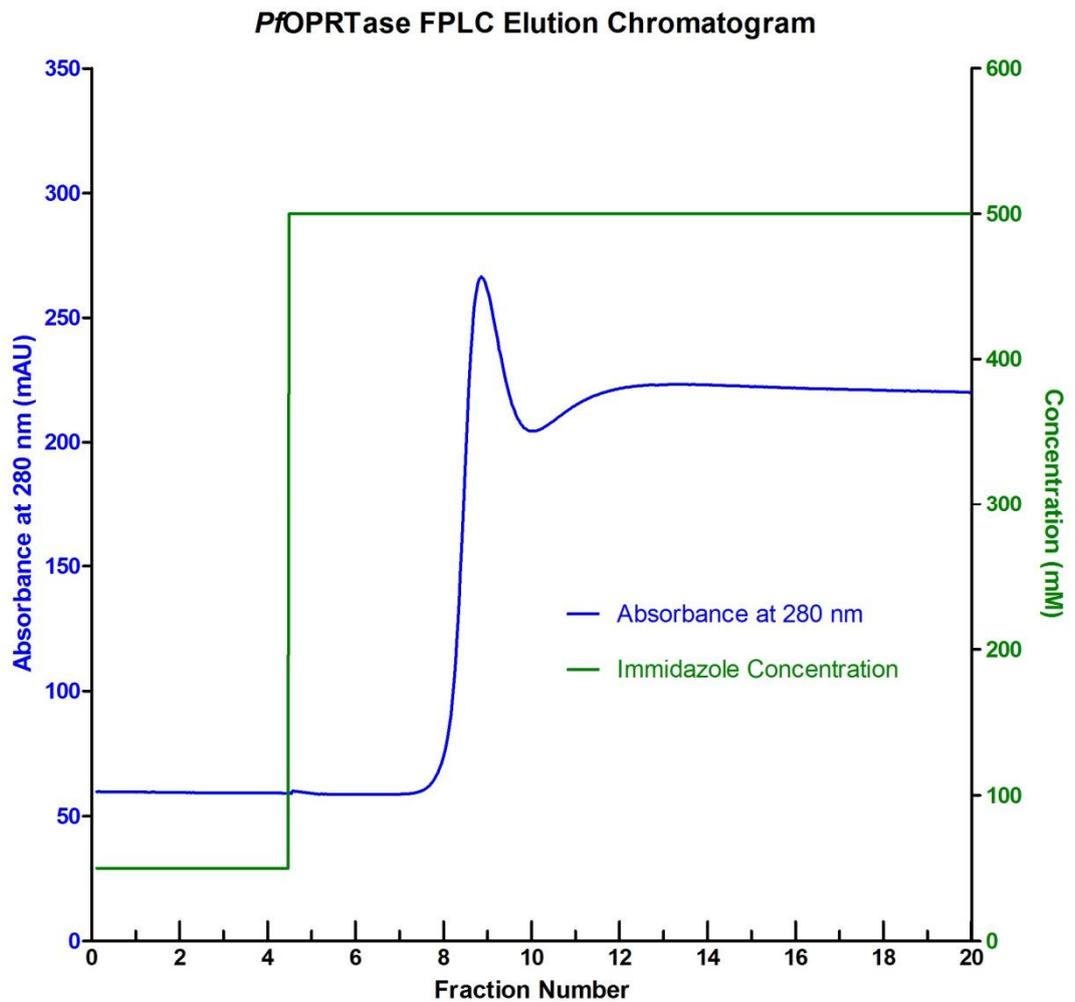


Figure 3.2: A typical elution profile of recombinant *Pf*OPRTase purified by Nickel affinity chromatography. The absorbance at 280 nm is shown in blue. The concentration of imidazole in the elution buffer is shown in green. Chromatography was carried out as detailed in Sections 2.4.5 and 3.2.2. The sample loading and 40 CV wash steps are not shown. The elution was collected in 2 mL fractions.

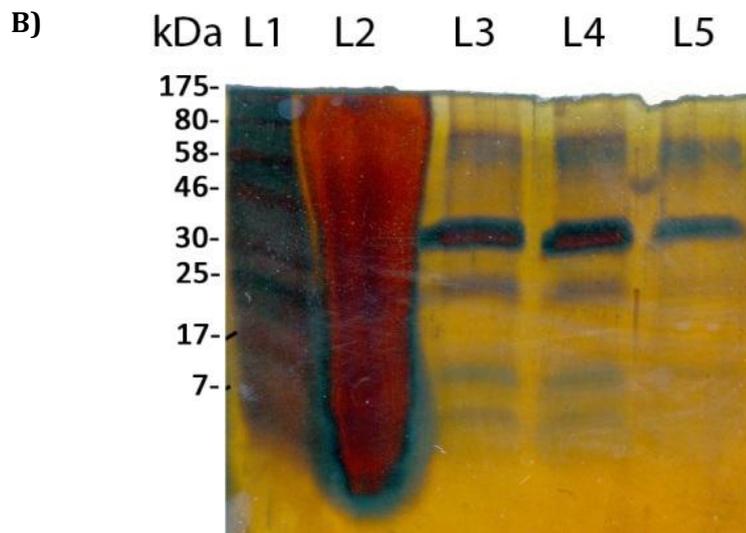
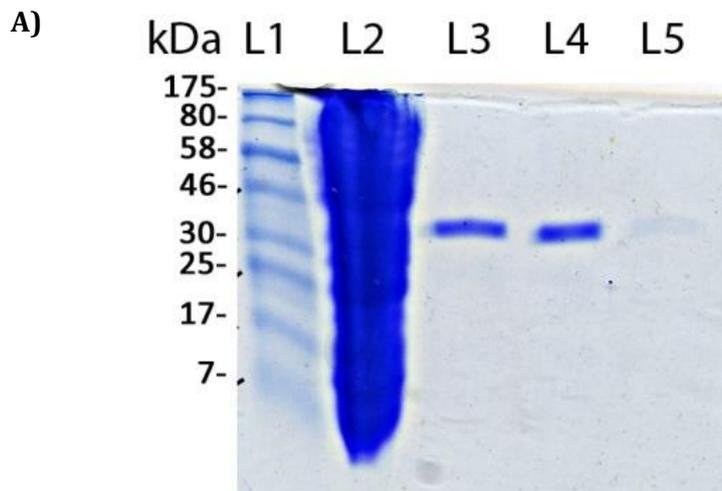


Figure 3.3: SDS-PAGE gels of Nickel-affinity purified recombinant PfOPRTase.

Samples were resolved on 5–17 % acrylamide gradient gels (Section 2.4.6) **A)**

Coomassie stained gel (Section 2.4.6.1) of nickel chromatography fractions: Lane 1 New England Biolabs (NEB) Prestained Broad Range (7–175 kDa) protein ladder used (discontinued product); Lane 2, 5 μ L sample of cleared cell lysate prior to nickel affinity chromatography; Lanes 3–5, 30 μ L samples of fractions 8–10 (Figure 3.2). **B)** Gel from

Figure 3.2.A subject to silver staining (Section 2.4.6.2).

3.2.2.2. *PfODCase*

Recombinant *PfODCase* was expressed and purified (as outlined in Section 3.2.2).

Previous in-house work (Tam Tam, 2007) had reliably and consistently expressed and purified the enzyme. Recombinant *PfODCase* exhibited a relatively high level of expression. As such only a single batch was required to provide the enzyme needed for assays.

The FPLC chromatogram (Figure 3.4) from the nickel affinity chromatography shows a significant peak in $OD_{280\text{ nm}}$ of approximately 500 mAU during elution indicative of a high concentration of protein. Fractions 12–18 were pooled. The protein concentration was measured to be $0.168\text{ mg}\cdot\text{mL}^{-1}$ (Section 2.4.8). SDS-PAGE with Coomassie staining showed a very large band at approximately 38 kDa (Figure 3.5). This is consistent with the recombinant *PfODCase* with a predicted size of 40.1 kDa (including the poly-histidine tag). Figure 3.4 shows the presence of other proteins in the sample with approximate sizes of 150, 80, 60, 30, and 20 kDa. An estimated purity of 85–95 % was achieved as determined using gel analyser. Cleavage of the N-terminal His-tag was not necessary for biochemical assays and no crystallography experiments were planned for this enzyme. The sample was aliquoted and stored at $-80\text{ }^{\circ}\text{C}$ for later use with assays.

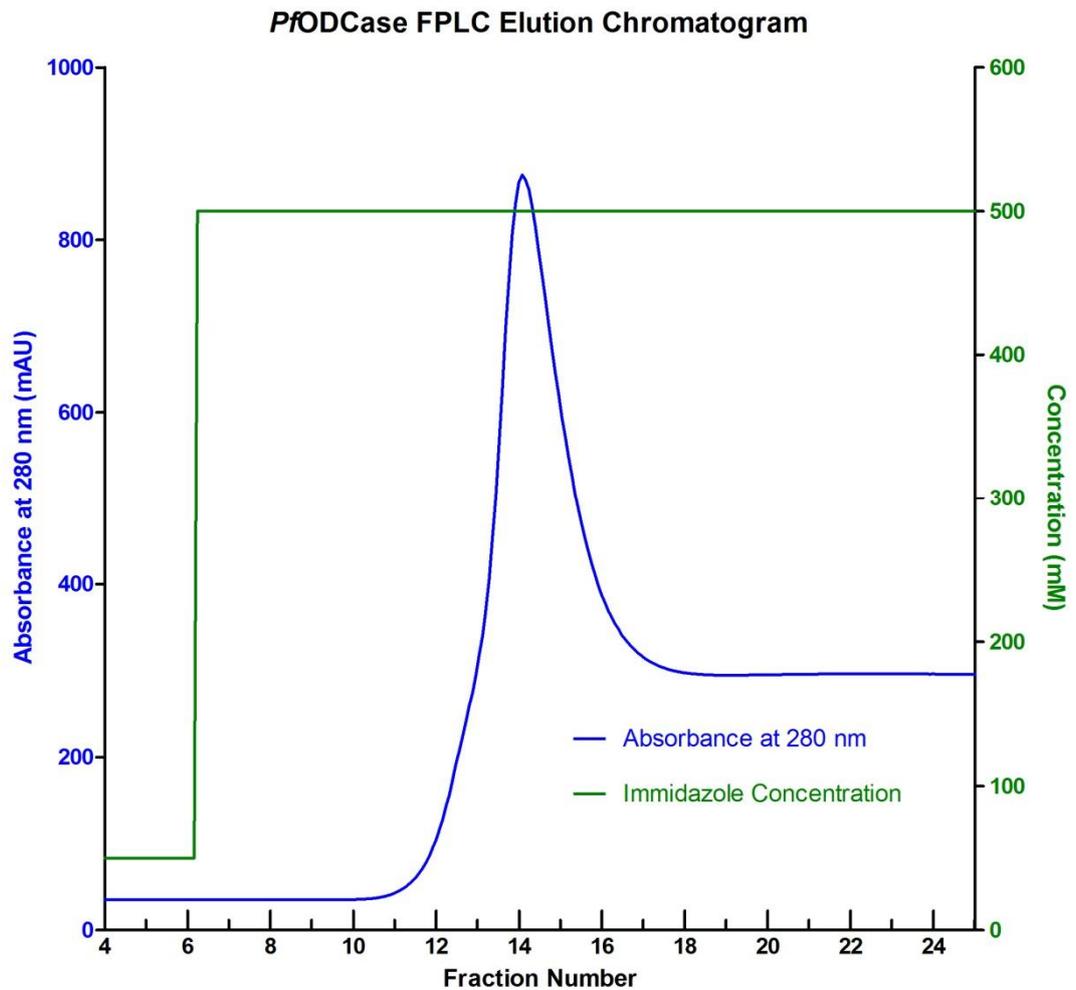


Figure 3.4: FPLC Chromatogram of elution of recombinant *Pf*ODCase from Nickel affinity chromatography. The absorbance at 280 nm is shown in blue. The concentration of imidazole in the elution buffer is shown in green. Chromatography was carried out as detailed in Section 2.4.5 and 3.2.2. The sample loading and 40 CV wash steps are not shown. The elution was collected in 2 mL fractions.

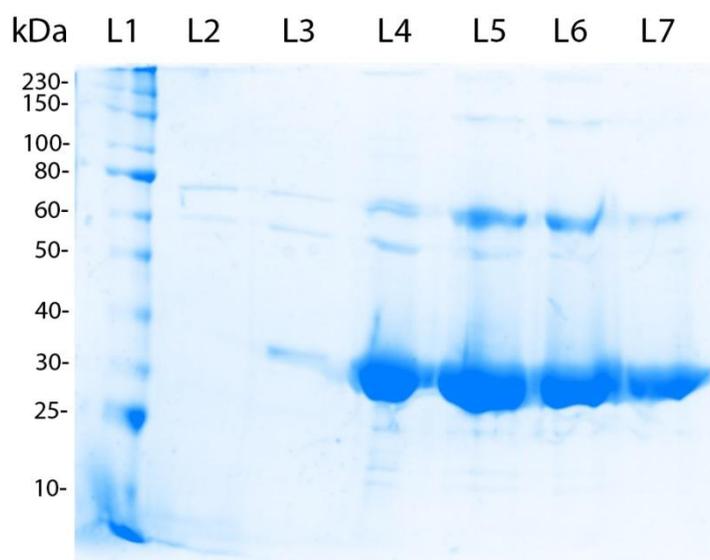


Figure 3.5: Coomassie stained SDS-PAGE gels of Nickel-affinity purified recombinant *PfODCase*. Samples were resolved on 5–17 % acrylamide gradient gels (Section 2.4.6). NEB Prestained Broad Range (10–230 kDa) protein ladder used (cat# P7710S). Lane 1, protein ladder; lanes 2–7, 30 μ L samples of nickel affinity fractions corresponding to fractions 11–16 in Figure 3.4.

3.2.2.3. *HsOPRTase*, *HsODCase* and *HsUMPS*

The human OPRTase, ODCase and UMPS were expressed and purified several times in various trial expressions. The batches that were used for the biochemical assays are shown here and represent typical expressions and purifications for these enzymes.

The enzymes were expressed and purified as outlined in Section 3.2.2. The FPLC chromatograms from the nickel chromatography for *HsODCase* and *HsOPRTase* showed extremely large OD_{280 nm} peaks of approximately 1500 mAU during the elution step (Figures 3.6 and 3.7) indicative of high protein concentration. The FPLC chromatogram for *HsUMPS* showed a much smaller peak of approximately 100 mAU at fractions 33–36 (Figure 3.8). The OD_{280 nm} elution peaks scale exactly with the chromatograms for *PfOPRTase* and *PfODCase* (Figures 3.2 and 3.4) and serve as a comparison for relative levels of expression. While the individual domains of *HsUMPS*—*HsOPRTase* and *HsODCase*—both showed extremely high levels of expression, *HsUMPS* exhibited a low level of expression that was only slightly higher than *PfOPRTase*.

Analysis by Coomassie stained SDS-PAGE of the nickel chromatography fractions of *HsOPRTase* and *HsODCase* (Figure 3.9) showed extremely intense bands at approximately 28 kDa and 33 kDa. This is consistent with the expected sizes of 27.8 and 32.8 kDa (including the N-Terminal tags) for *PfOPRTase* and *PfODCase* respectively. The purity as estimated with Gel Analyzer was found to be 95 and 99 % for *HsOPRTase* and *HsODCase* respectively.

Analysis by Coomassie stained SDS-PAGE of *HsUMPS* (Figure 3.10) shows a band at approximately 55 kDa (consistent with the predicted size of 57 kDa for *HsUMPS* with the N-Terminal tag) and Gel Analyzer estimates purity at 85 %. Total yield was approximately 8–10 times that of *PfOPRTase*.

Following Nickel chromatography, the peak fractions 18–24 of *HsOPRTase* were pooled as were fractions 4–6 for *HsODCase*. Bradford assays of the pooled fractions reveal the

protein concentrations to be approximately 0.69 and 2.91 mg·mL⁻¹ for *HsOPRTase* and *HsODCase* respectively. The samples were aliquoted and stored at -80 °C for use with biochemical assays.

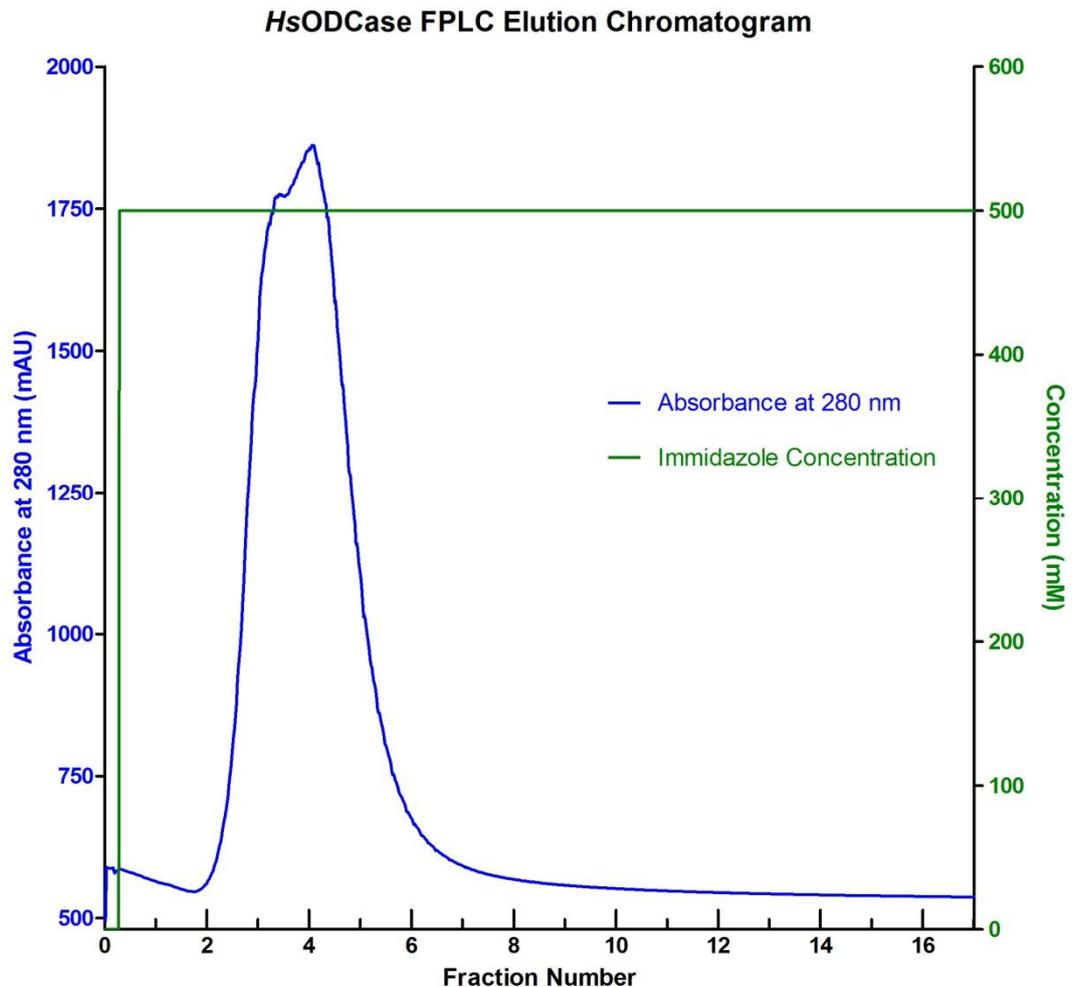


Figure 3.6: FPLC Chromatogram of elution of recombinant *HsODCase* from Nickel affinity chromatography. The absorbance at 280 nm is shown in blue. The concentration of imidazole in the elution buffer is shown in green. Chromatography was carried out as detailed in Sections 2.4.5 and 3.2.2. The sample loading and 40 CV wash steps are not shown. The elution was collected in 2 mL fractions.

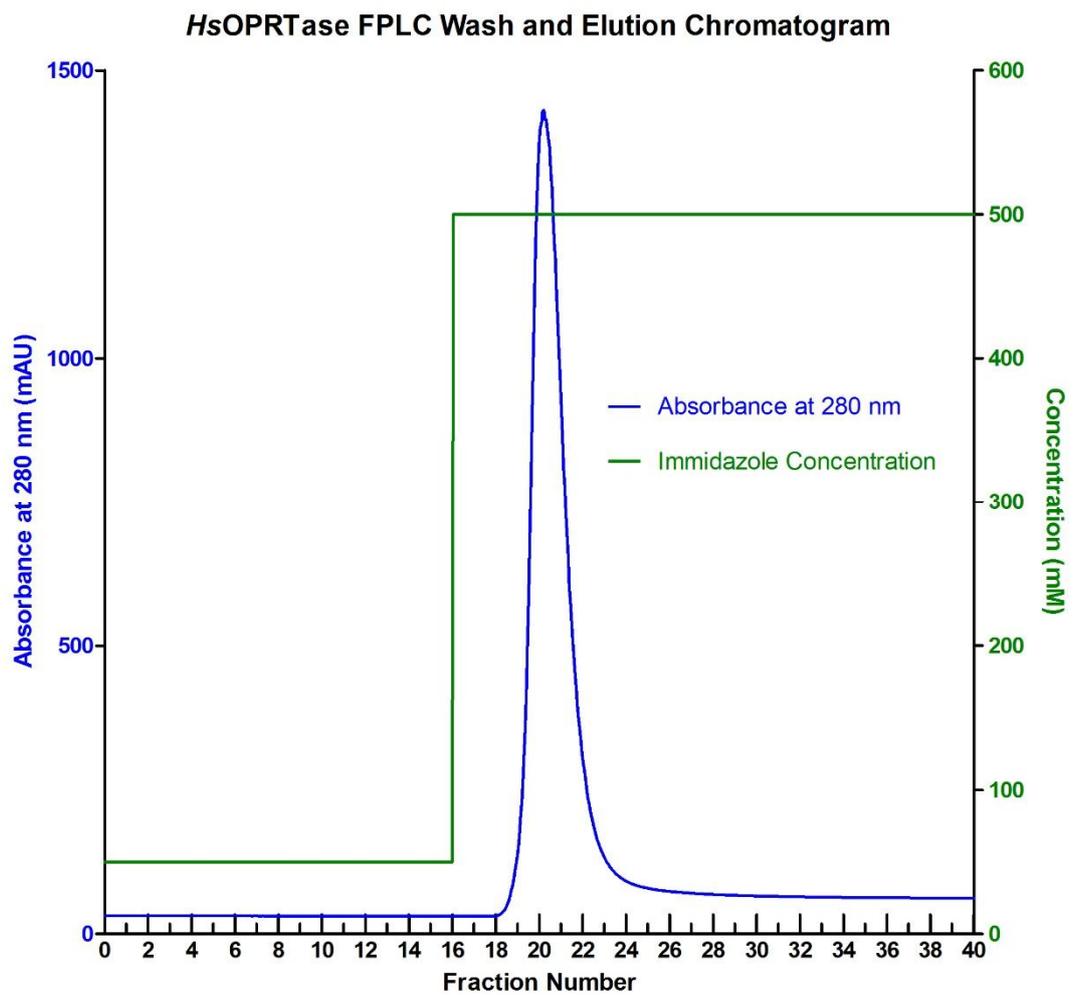


Figure 3.7: FPLC Chromatogram of elution of recombinant *HsOPRTase* from Nickel affinity chromatography. The absorbance at 280 nm is shown in blue. The concentration of imidazole in the elution buffer is shown in green. Chromatography was carried out as detailed in Sections 2.4.5 and 3.2.2. The sample loading and most of the 40 CV wash step are not shown. The elution was collected in 2 mL fractions.

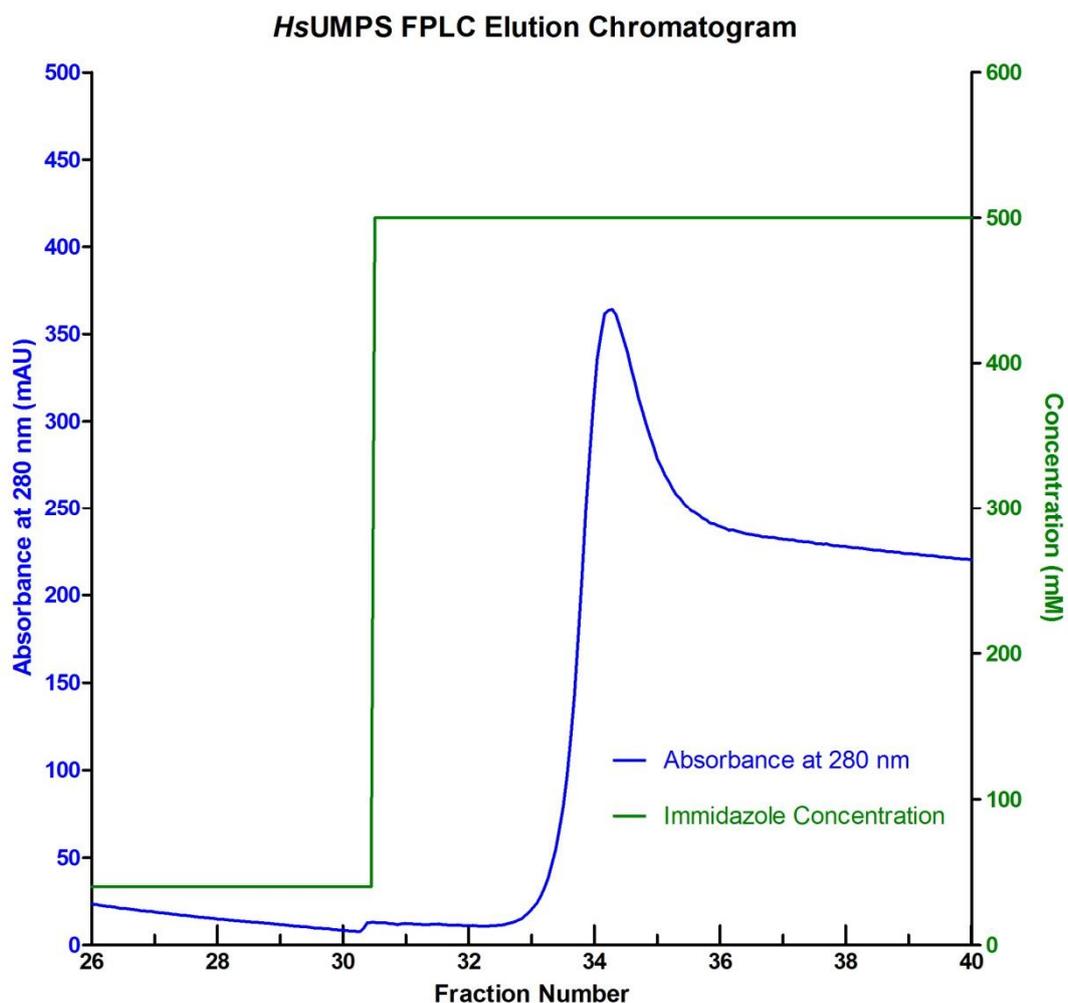


Figure 3.8: FPLC Chromatogram of elution of recombinant *HsUMPS* from Nickel affinity chromatography. The absorbance at 280 nm is shown in blue. The concentration of imidazole in the elution buffer is shown in green. Chromatography was carried out as detailed in Sections 2.4.5 and 3.2.2. The sample loading and 40 CV wash steps are not shown. The elution was collected in 2 mL fractions.

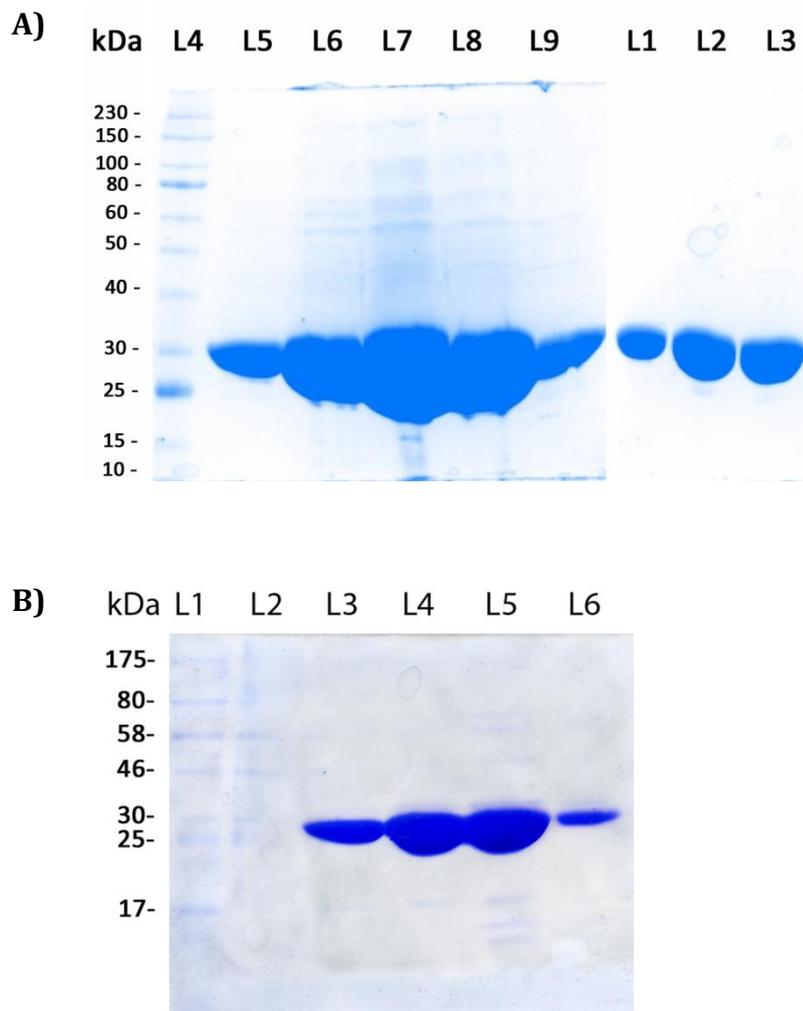


Figure 3.9: Coomassie stained SDS-PAGE gels of Nickel-affinity purified recombinant *HsODCase* and *HsOPRTase*. Samples were resolved on 5–17 % acrylamide gradient gels (Section 2.4.6). **A)** *HsODCase* Nickel fractions: lane 4, NEB Prestained Broad Range (10–230 kDa) protein ladder used (cat# P7710S); lanes 5–9, 30 μ L samples of nickel fractions corresponding to fractions 2–6 (Figure 3.6); lanes 1–3, 5 μ L samples of fractions 3–5 (Figure 3.6). **B)** *HsOPRTase* Nickel fractions: lane 1, NEB Prestained Broad Range (7–175 kDa) protein ladder used (discontinued product); lanes 3–6, 20 μ L samples of nickel fractions corresponding to fractions 19–22 (Figure 3.7).

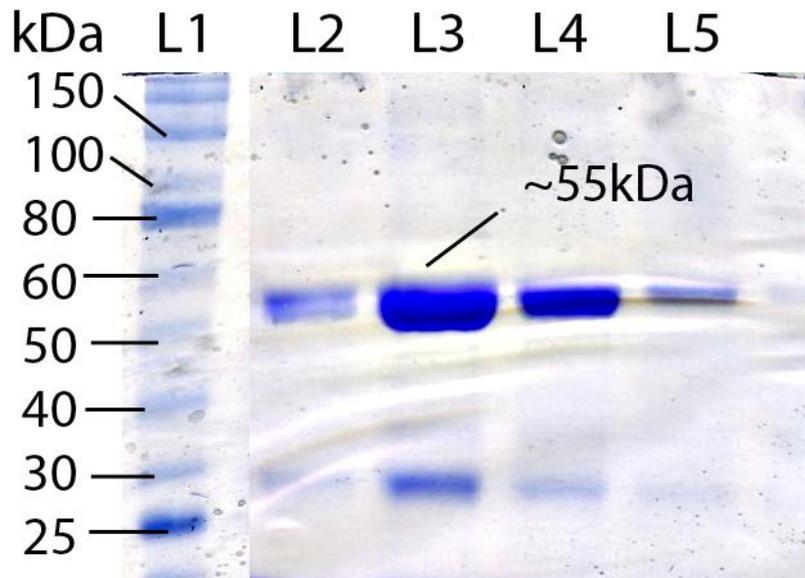


Figure 3.10: Coomassie stained SDS-PAGE gels of Nickel-affinity purified recombinant *HsUMPS*. Samples were resolved on 5–17 % acrylamide gradient gels (Section 2.4.6). NEB Prestained Broad Range (10–230 kDa) protein ladder used (cat# P7710S). Lane 1, protein ladder; Lanes 2–5, 30 μ L samples of nickel affinity fractions corresponding to fractions 33 to 36 (Figure 3.8).

3.2.3. Specific activities of *P. falciparum* and *H. sapiens* OPRTase and ODCase

*Pf*OPRTase had a specific activity of 3–4 $\mu\text{mol}\cdot\text{min}^{-1}\cdot\text{mg}^{-1}$. *Hs*OPRTase had a specific activity of 0.8–1.6 $\mu\text{mol}\cdot\text{min}^{-1}\cdot\text{mg}^{-1}$. *Pf*ODCase had a specific activity of 2.7–6.8 $\mu\text{mol}\cdot\text{min}^{-1}\cdot\text{mg}^{-1}$. *Hs*ODCase had a specific activity of 6.4–9.8 $\mu\text{mol}\cdot\text{min}^{-1}\cdot\text{mg}^{-1}$. The recombinant enzymes had been aliquoted and stored at $-80\text{ }^{\circ}\text{C}$ and stored at $4\text{ }^{\circ}\text{C}$ between assays. The enzymes showed a decrease in specific activity of approximately 10–40 % per day when stored at $4\text{ }^{\circ}\text{C}$ (data not shown, the ODCases in particular appeared to be less stable). As such all assays conducted in Chapter 5 were normalised against control assays and aliquots were regularly replaced with new ones from $-80\text{ }^{\circ}\text{C}$ stocks.

3.2.4. Substrate Kinetics of *Pf*ODCase and *Hs*ODCase

Kinetics were performed on the human and malarial ODCases using reaction progress kinetic analysis. Kinetic assays were performed as outlined in Section 2.5.4. The reaction rates at different substrate concentrations were derived from the assay curves and this data were then fitted to the Michaelis-Menten equation (Figure 3.11). The K_m^{app} values were determined to be approximately $9.19 \pm 1.78\ \mu\text{M}$ ($R^2 = 0.931$) and $3.95 \pm 0.38\ \mu\text{M}$ ($R^2 = 0.976$) for *Pf*ODCase and *Hs*ODCase respectively. The V_{max} values were found to be $12.1 \pm 0.86\ \mu\text{mol}\cdot\text{min}^{-1}\cdot\text{mg}^{-1}$ and $14.4 \pm 0.47\ \mu\text{mol}\cdot\text{min}^{-1}\cdot\text{mg}^{-1}$ for *Pf*ODCase and *Hs*ODCase respectively.

HsODCase and *PfODCase* Michaelis-Menten Kinetics

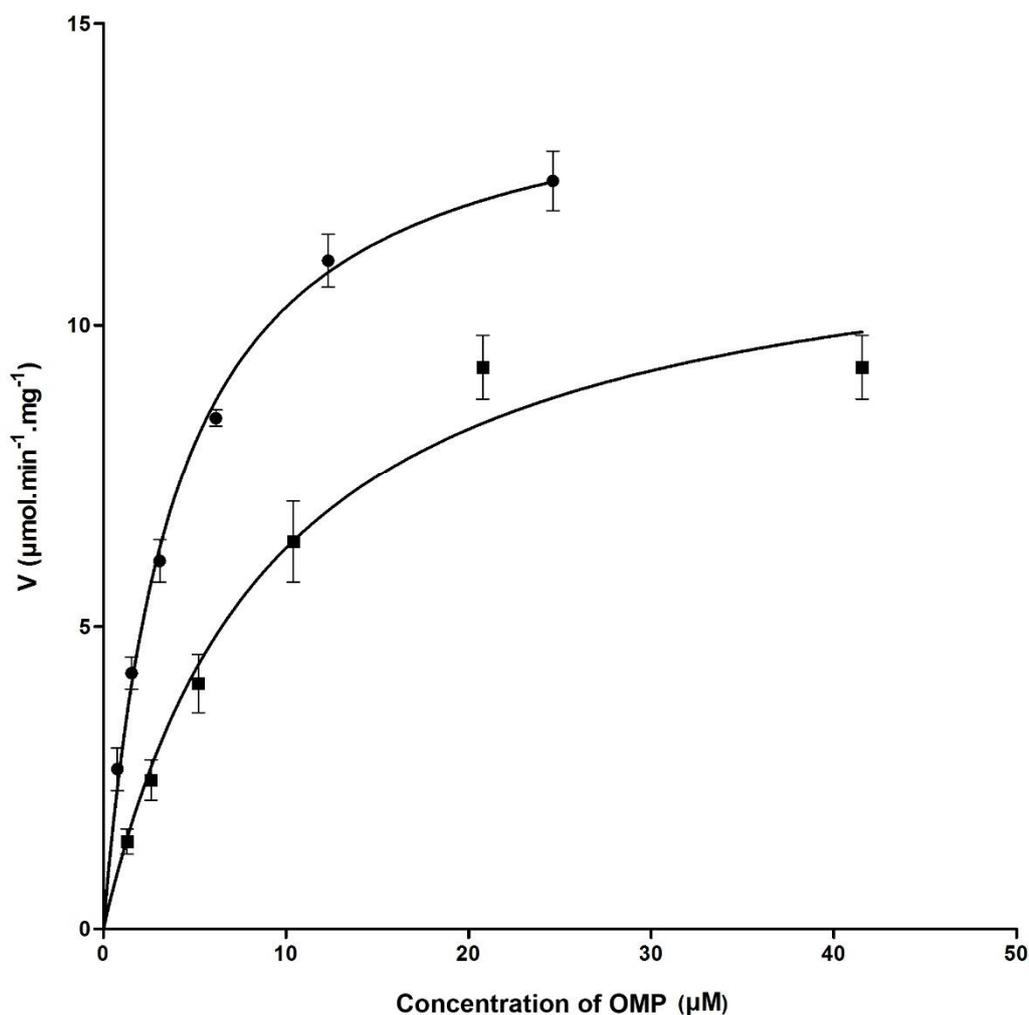


Figure 3.11: Michaelis-Menten kinetics of *PfODCase* and *HsODCase* with the substrate OMP. *PfODCase* (■) and *HsODCase* (●) activity were measured as detailed in Section 2.5.4, at a constant temperature of 30 °C. Rates are means \pm SEM ($n = 3$). The sets of data were fit to the Michaelis-Menten equation (Appendix 8.5) with an R^2 value of 0.931 and 0.976 for *PfODCase* and *HsODCase* respectively. The estimated K_m^{app} values were $9.19 \pm 1.78 \mu\text{M}$ (*PfODCase*) and $3.95 \pm 0.38 \mu\text{M}$ (*HsODCase*). The estimated V_{max} values were $12.1 \pm 0.86 \mu\text{mol}\cdot\text{min}^{-1}\cdot\text{mg}^{-1}$ (*PfODCase*) and $14.4 \pm 0.47 \mu\text{mol}\cdot\text{min}^{-1}\cdot\text{mg}^{-1}$ (*HsODCase*).

3.2.5. Crystallisation Trials

The crystal screens trialled initially were the Hampton Research screens 1 and 2, and the NeXtal PEGs in 96 well crystal screening plates. Crystal growth was assessed using polarising light for birefringence. Of these screens, there were four Hampton Research 1 and 2 screens that showed promising crystals. These were P37 (0.1 M HEPES pH 7.5, 0.8 M potassium dihydrogen phosphate), Q47 (0.1 M tri-sodium citrate dihydrate pH 5.6, 0.2 M potassium sodium tartrate hexahydrate, 2 M ammonium sulphate), Q45 (0.1 M MES pH 6.5, 0.01 M cobaltous chloride hexahydrate, 1.8 M ammonium sulphate) and Q31 (0.1 M Tris-cl pH 8.5, 0.2 M magnesium chloride hexahydrate, 3.4 M 1,6 hexanediol) (data not shown).

Condition Q31 showed a relatively large (approximately $10 \times 5 \mu\text{m}$) rectangular (possibly tetragonal or monoclinic) crystal that was noticeable within minutes of preparing the screening plate and did not grow noticeably larger thereafter. Hence, this condition was replicated with larger drops of $2 \mu\text{L}$ in 12-well plates. However, these larger scale experiments failed to produce crystals.

Condition Q45 showed at first very small crystals. These grew over the course of seven days to form unevenly shaped crystals, although there was no apparent symmetry to the crystals at all. Upscaling to 12-well plates with $2 \mu\text{L}$ drops reproduced this crystal formation for multiple replicates. The size of these “crystals” after seven days were approximately 5 to $10 \mu\text{m}$ with still no clear symmetry. One of the upscaled replicates showed a very different crystal system. A single large crystal had formed after seven days. It was approximately $20 \times 10 \mu\text{m}$ (length \times width/depth) and appeared to be a hexagonal crystal system.

Conditions P37 and Q47 both showed some small square crystals in the 96-well plate screen. Upscaling to 12-well plates with $2 \mu\text{L}$ drops resulted in several more crystals for each screen.

X-ray diffraction carried out at the Australian Synchrotron returned no usable diffraction data (data not shown). The large, rectangular crystal from Q31 showed diffraction patterns with a unit cell size of approximately $5 \times 10 \times 20 \text{ \AA}$ —too small to be *PfOPRTase* (higher order symmetry would not be occurring). The small, asymmetric crystals from condition Q45 showed diffraction resembling that of powder diffraction patterns. The large, hexagonal crystal from the same screen showed no diffraction at all. Crystals from P37 and Q47 showed diffraction patterns identifying them as salt crystals.

Subsequent crystal screens of NeXtal PEGsII, Cryo and Classics showed some conditions that yielded potential protein crystals. In particular were NeXtal Cryo conditions A1 (0.085 M Sodium acetate pH 4.6, 0.085 M Cobalt chloride, 0.85 M 1,6-Hexanediol, 15 % (w/v) Glycerol), B1 (0.085 M TRIS pH 8.5, 17 % (w/v) Ethanol, 15 % (w/v) Glycerol), and B6 (0.09 M Sodium cacodylate pH 6.5, 0.18 M Magnesium acetate, 27 % (w/v) MPD, 10 % (w/v) Glycerol). These conditions however were not examined further. Most conditions simply yielded protein precipitant immediately on formation of the drop.

3.3. Discussion

3.3.1. Expression and Purification of Recombinant Enzymes

3.3.1.1. *Pf*OPRTase Expression and Purification

Krungskrai et al. (2004a), in an independently developed method, expressed and purified recombinant *Pf*OPRTase with a very similar protocol to that outlined here. Their study involved genomic parasite DNA being amplified and cloned into a pET15b expression vector (using lac operon/T7 RNA polymerase expression system), induction with IPTG early in the growth phase, at approximately OD_{600 nm} 0.5, and Nickel affinity purification. Thrombin cleavage was used in this study to purify to near homogeneity with size exclusion chromatography to remove the thrombin. The SDS-PAGE gel from Krungskrai et al. (2004a) (Figure 3.12) shows the nickel fraction (lane 2) with a much lower ratio of recombinant *Pf*OPRTase to contaminant protein to that obtained here (Figure 3.3). This is likely attributed to the fact that Krungskrai et al. (2004a) used a much lower imidazole concentration (20 mM) in the wash step of the nickel affinity chromatography compared to here (50 mM). In Figure 3.12 the final, purified fraction (lane 3) that was subject to thrombin cleavage and size-exclusion chromatography had some visible contaminant protein bands—a similar level of purity to that obtained in this study. The recombinant *Pf*OPRTase sample expressed and purified here was therefore deemed suitable for use in inhibition assays.

The level of purity was not optimal for crystallography as an enzyme purified to homogeneity has a much greater chance of forming protein crystals than one with protein contaminants (HR, 2001). There is a chance of protein crystals forming however and given the value at the time of having an experimentally solved structure of *Pf*OPRTase an attempt was subsequently made to crystallise the enzyme.

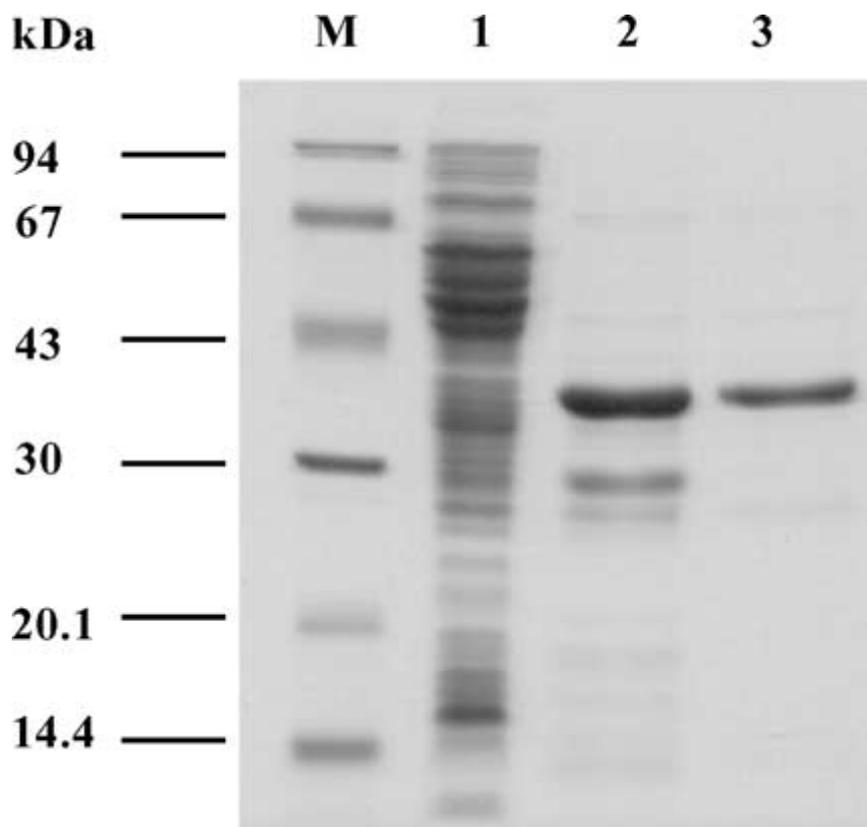


Figure 3.12: Figure from Krungkrai et al. (2004a): “Purification and SDS-PAGE analysis of the *pfOPRT* enzyme. Three purification steps of the recombinant enzyme were subjected to SDS-PAGE on a 12% acrylamide gel and visualized by staining with Coomassie Blue dye. Lane 1, the supernatant (30 μ g protein) obtained after centrifugation of the sonicated *E. coli* expressing cells; lane 2, Ni²⁺-NTA affinity chromatography step (10 μ g protein), a dominant 33-kDa protein is observed; lane 3, gel-filtration chromatography step (5 μ g protein), a single 33-kDa band is shown. The molecular mass markers are indicated in kilodalton (lane M).”

A higher yield was achieved in Krungkrai et al. (2004a). They report a yield of 1.2–1.5 mg of purified, recombinant *PfOPRTase* from 1 L of bacterial culture. The yield obtained here was approximately 0.1–0.2 mg per litre of bacterial culture. It is unlikely that the pET15b expression vector is 10 times more efficient than the pET3a vector at expressing the enzyme. This is more likely due to the difference in either the batch size or imidazole wash concentration (or both). The paper implicitly states a batch size of 1 L whereas here, two 4 L fermenters were inoculated from a single 50 mL starter culture. It was observed that inoculating at $OD_{600\text{ nm}}$ greater than 0.5 resulted in a lower rather than a higher yield. It is possible that there is also diminishing, if any, returns from scaling up beyond a batch size of 1 L. The gene is known to be unstable in *E. coli*. Krungkrai et al. (2004a) specifically states the requirement of freshly transformed cells for inoculating the expression culture. This step was also found to be necessary in this study as transformed cells recovered on solid media were not stable for more than 2 days (data not shown).

In this study the molecular cloning was carried out in *E. coli* PMC103: a cell line designed to cope with palindromic sequences in genes (Doherty et al., 1993), however expression was carried out in BL21(DE3) cells which lack the same host mutations for optimal tolerances for palindromic DNA secondary structure. The likely cause of the instability in *PfOPRTase* can be seen in Figure 3.13. This figure shows secondary structure predicted for single stranded DNA (which occurs during replication) of the *PfOPRTase* gene. Several stems contribute to a large hairpin including a 15 bp long sequence. Doherty et al. (1993) state that a palindrome sequence of 30 or more bases creates this DNA instability so it stands to reason that a 15 base pair-long stem would have the same effect.

This type of DNA secondary structure is most likely the cause of the gene instability for *PfOPRTase*. The large number of generations it takes to get from a single colony to a 4 L fermenter with an $OD_{600\text{ nm}}$ of 0.5 likely results in far fewer cells containing an intact

copy of the *PfOPRTase* gene. Greater efficiency would therefore be achieved by reducing the cell culture size and up scaling by means of increasing the number of cultures grown and pooling the cultures during cell harvesting. Another method to address the gene instability would be to re-engineer and normalise the gene by changing codons to remove the palindromic sequences but keep the translated product the same (balancing the AT:GC content and removing rare codons would also further improve gene stability and expression efficiency and remove the need for the pRIG helper plasmid). There are a number of companies that offer gene synthesis services such as Life Technologies (Thermo Fisher Scientific, Waltham, USA). Alternatively, expression in a baculovirus or a yeast system would likely improve the gene stability and protein yield (Victor et al., 2010, Wagner et al., 2013).

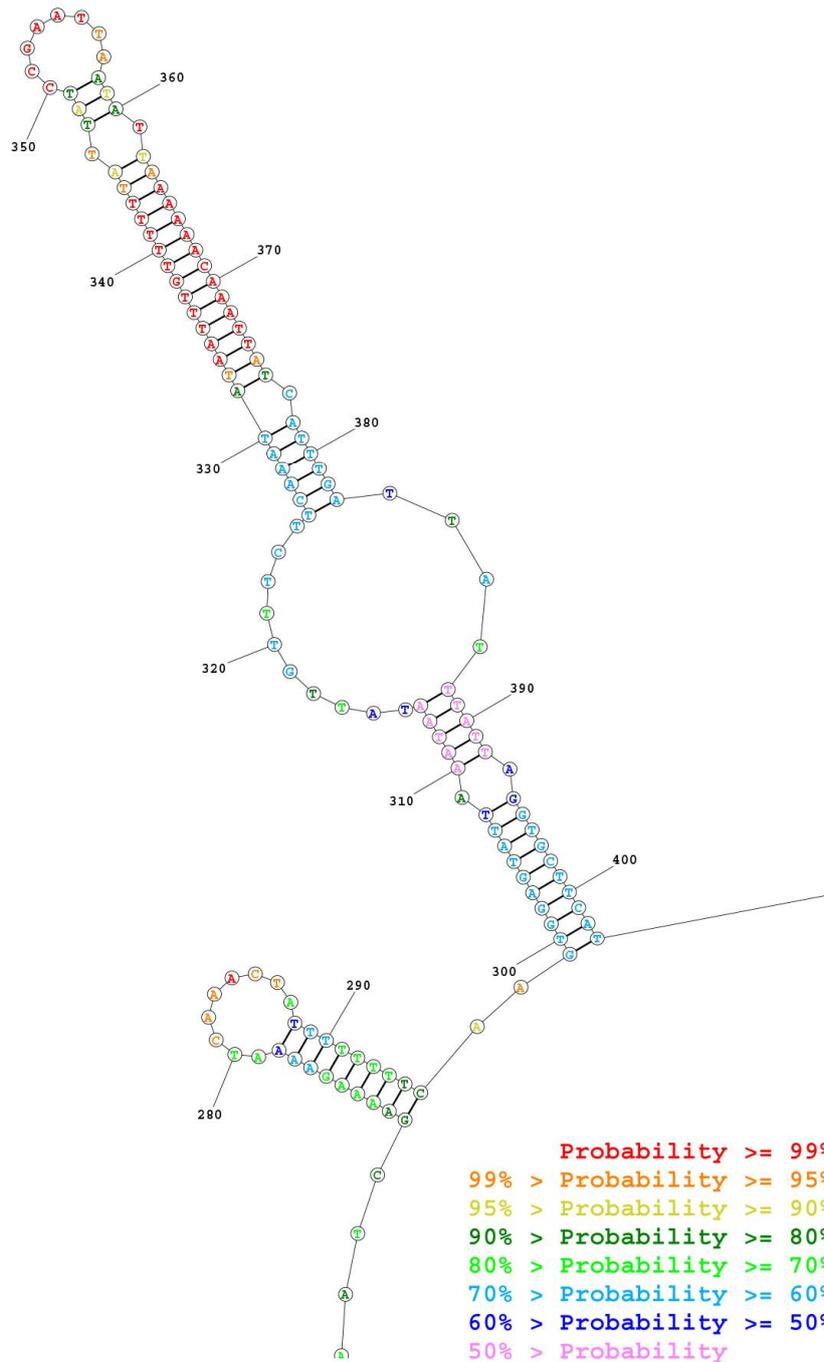


Figure 3.13: DNA secondary structure prediction of *PfoPRTase*. Secondary structure predicted using ‘RNAstructure’ web server (Reuter and Mathews, 2010). Bases 270–400 are shown here. Several palindromic sections form stem-loops/hairpins. In particular bases 331–345 and 362–376 form a 15 bp long stem. Probability is calculated for RNA or single stranded DNA.

3.3.1.2. *Pf*ODCase Expression and Purification

Recombinant *Pf*ODCase was purified to a degree that was suitable for accurate kinetic analysis. The contaminating proteins that were present in purified *Pf*OPRTase were also present for the purified *Pf*ODCase, however the purity was far greater as there was a much higher ratio of recombinant enzyme to contaminants. This is likely attributed to the higher level of expression impacting on the nickel-affinity chromatography step where recombinant enzyme was outcompeting contaminants for interactions with immobilised nickel ions.

*Pf*ODCase was first recombinantly expressed and purified by Menz et al. (2002). The same method was used here (minus the histidine tag cleavage) and in Langley et al. (2008). It is very similar to other reported methods. Krungkrai et al. (2005) and Tokuoka et al. (2008) both independently produced an expression vector using the pTrcHis-TOPO plasmid (Invitrogen) and expressed without the use of the pMICO helper plasmid. The pTrcHis-TOPO vector utilises a trc and lac operon, induction with 1 mM IPTG, and T7 RNA polymerase transcription. They purified in a similar way, nickel affinity chromatography with cleavage of the histidine-tag using enterokinase and subsequent gel filtration. They report a yield of approximately 3 mg per mL of cell culture, compared to approximately 1.5 mg per mL of culture here. It is likely the pTrcHis-TOPO expression vector is more efficient than the pET3a and pMICO combination. The difference in yield is unlikely to be occurring for the same reason as with *Pf*OPRTase (diminishing returns on scaling up beyond a cell culture size of 1 L) as Tokuoka et al. (2008) implicitly report a batch size of 10 L. In any case the yield obtained from using the pET3a + pMICO system was more than adequate for all experiments conducted here, in Menz et al. (2002) and Langley et al. (2008).

3.3.1.3. *Hs*OPRTase and *Hs*ODCase Expression and Purification

Both *Hs*ODCase and *Hs*OPRTase appear to be easily expressed in *E. coli* and purified to homogeneity in published work. *Hs*ODCase was produced for a number of crystal

structures in a method similar to that employed here (Wittmann and Rudolph, 2007, Wittmann et al., 2008, Heinrich et al., 2009, Meza-Avina et al., 2010, Purohit et al., 2012). A pETM-30 vector was used containing a poly-histidine tag, Glutathione S-transferase (GST) tag, TEV protease cleavage site, lac operon/T7 RNA transcription system, and IPTG induction at 30 °C overnight. Purification involved utilisation of the GST tag with a glutathione Sepharose column (as opposed to Nickel affinity using a poly-histidine tag), removal of the tag with TEV protease, and size-exclusion chromatography. The use of a helper plasmid was not employed in these papers.

The only available structure for *HsOPRTase* lists *E. coli* as the expression host (Moche et al., 2009). Zhang et al. (2009) also expressed *HsOPRTase* in *E. coli*. In Zhang et al. (2009) a pDEST14 vector was used (lac operon and T7 promoter), cloned in BL21-AI cells and induced overnight with 0.02 % L-arabinose at 37 °C. Purification involved Nickel affinity chromatography (albeit with much lower wash length of 5 CV and lower imidazole wash concentration of 20 mM), removal of the poly-histidine tag, and subsequent size-exclusion chromatography. Yield and purification were not reported, however it is assumed that purification to apparent homogeneity was achieved.

It may be beneficial in the future to attempt crystallisation of these proteins for use in x-ray crystallography. Were a potent inhibitor to be discovered, it would be beneficial to have a protein crystal ready for crystal soaking with the inhibitor, or conditions known to produce crystals for co-crystallising with a ligand (and subsequent x-ray diffraction collection) to determine a crystal structure of the enzyme with the bound inhibitor. While there is no shortage of ODCase structures there remain very few OPRTase structures, especially with ligands bound. Furthermore crystal structures that are co-crystallised with a lead compound would greatly aid in lead development.

The extra steps to remove the histidine tag and GST tag could be employed for purification of *HsODCase* and *HsOPRTase* for future crystallography experiments. This would involve cleavage of the histidine/GST tag with either thrombin or enterokinase

(followed by another nickel chromatography step where the flow-through is collected, and either size exclusion chromatography or affinity chromatography to remove the thrombin or enterokinase).

3.3.1.4. *HsUMPS* Expression and Purification

The main purpose to producing this vector was to allow for future work on solving the structure of *HsUMPS*. There is currently no known structure available for the whole human UMPS bi-functional enzyme. While the structural information of the bi-functional complex would be valuable in itself, it may also give insights into how the *P. falciparum* OPRTase and ODCase could form a hetero-tetramer, as has been shown to occur (Krungkrai et al., 2005). A high yield is very desirable for macromolecular crystallography but more important is purification to apparent homogeneity.

The UMPS enzyme had been studied in rat cell models (Kanalas and Suttle, 1984), purified from human placenta (Livingstone and Jones, 1987), and more recently, expressed in a baculovirus system (Han et al., 1995) and in *E. coli* (Suchi et al., 1997). Purification from tissue is limited in yield and purification. While Livingstone and Jones (1987) report a method for purification of *HsUMPS* to homogeneity from tissue it is unlikely to yield enough purified enzyme to be useful for macromolecular crystallography. The baculovirus expression system is an excellent system for producing large amounts of protein with proper post-translational modifications. Expression in *E. coli* however is more advantageous in terms of time, cost, and scalability. As such expression by baculovirus in insect cells is often a last resort if expression in *E. coli* is not possible. Han et al. (1995) report that post-translational modifications occur in *HsUMPS* (not achievable by expressing in *E. coli*). The reported modifications are the removal of the N-Terminal methionine and acetylation of the penultimate alanine. Suchi et al. (1997) used expression in *pyrE* and *pyrF* deficient *E. coli* as an analytical means for studying UMPS in patients suffering from hereditary orotic aciduria. In this paper the gene was not over-expressed and purified, rather the

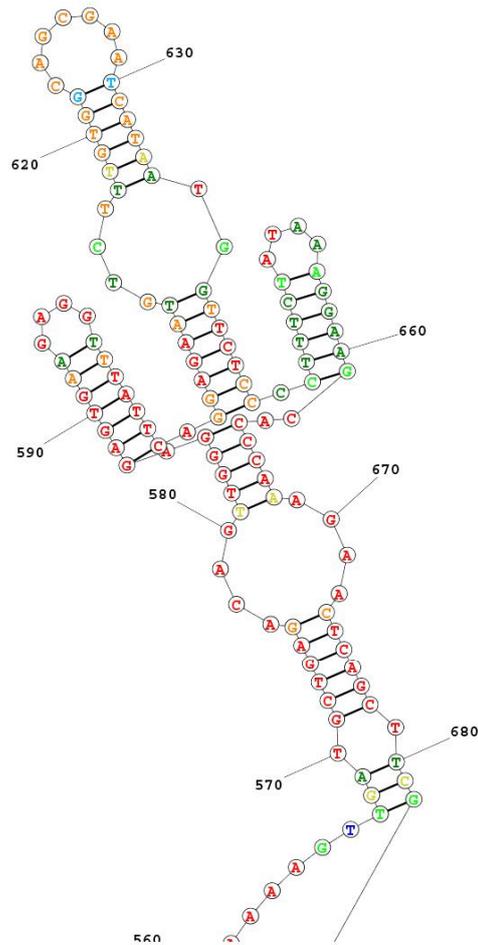
presence of a functional UMPS was tested by growing the *pyrE* or *pyrF* deficient, *HsUMPS*-transformed *E. coli* without uridine supplementation. What the paper does show is that the post-translational modifications are not necessary for activity of *HsUMPS*. Expression in *E. coli* for the purpose of macromolecular crystallography and biochemical assays should therefore be suitable for this gene, though there may be minor structural and functional differences between the post-translational modified and non-modified forms.

Early trial purification of *HsUMPS* with nickel affinity chromatography showed it was being eluted off the column at 50 mM of imidazole during the wash step. A smaller wash concentration was later used and appeared to improve the yield. The total yield was still very small compared to the individual OPRTase and ODCase domains of *HsUMPS*. The yield was approximately eight times greater than that of *PfOPRTase* (approximately twice the yield from a quarter the cell culture). Obtaining enough *PfOPRTase* for crystallography was very difficult but possible. As such the focus on future crystallography work on *HsUMPS* does not necessarily need to focus on improving yield. No attempts were made to optimise the fermentation and induction conditions so this remains an avenue of approach for improving yield in future work.

It is possible the protein folding of the recombinant *HsUMPS* was such that the histidine tag was only partly accessible resulting in a reduced column binding affinity. It is also possible that the enzyme does not transcribe or translate as readily as the individual OPRTase or ODCase domains, possibly due to a similar gene stability issue similar to what was seen for *PfOPRTase*. For this to be affecting the UMPS gene but not the ODCase or OPRTase domains, the cause would likely be in the domain linker region.

Figure 3.14 shows the single-stranded DNA secondary structure prediction for the domain linker region and neighbouring bases for the UMPS gene. Here there is a cluster of high probability stem loops forming a crucible-like structure whereas the rest of the gene predominantly contains only relatively short (4–10 bases) and isolated stems

with much lower predicted probability of occurrence. Crucible-like features are, like stem-loops, known to cause stability issues in *E. coli* (Kogo et al., 2007). These features are structurally analogous to Holliday junctions which commonly mediate processes such as DNA repair and recombination in a range of organisms (Stahl, 1994). The low level of expression of *HsUMPS* (relative to *HsOPRTase* and *HsODCase*) is likely caused by this secondary DNA/RNA feature. This can be confirmed by engineering a new construct with different codons to remove the inverted repeats but maintaining the translated enzyme. This would be beneficial in the interests of producing recombinant *HsUMPS* as it should significantly increase expression levels of this gene in *E. coli* BL21(DE3) cells. This feature also forms a potential avenue of research into the evolution of the bifunctional *HsUMPS* by a Holliday junction-mediated recombination event, or gene expression and regulation of the gene *in vivo*. Alternatively, to simply address the yield and post-translational modifications, a yeast expression system would be an ideal alternative to a baculovirus system in terms of cost and scalability (Midgett and Madden, 2007).



Probability >= 99%
 99% > **Probability** >= 95%
 95% > **Probability** >= 90%
 90% > **Probability** >= 80%
 80% > **Probability** >= 70%
 70% > **Probability** >= 60%
 60% > **Probability** >= 50%
 50% > **Probability**

Figure 3.14: DNA secondary structure prediction of *HsUMPS* domain linker.

Secondary structure predicted using 'RNAstructure' web server (Reuter and Mathews, 2010). Bases 560–682 (domain linker region and neighbouring bases) shown here. Several short palindromic sections form a crucible-like structure of high probability. Probability is calculated for RNA or single stranded DNA.

3.3.2. Specific Activities of OPRTases and ODCases

The specific activity of *Pf*OPRTase of 3–4 $\mu\text{mol}\cdot\text{min}^{-1}\cdot\text{mg}^{-1}$ protein was very similar to the 4–5 $\mu\text{mol}\cdot\text{min}^{-1}\cdot\text{mg}^{-1}$ that was achieved by Krungkrai et al. (2004a). While the expression level of *Hs*OPRTase far exceeded that of *Pf*OPRTase the specific activity was actually lower. Given the observed stability of the recombinant OPRTases, if the expression, purification, and assays were performed within a much shorter timeframe it is likely that the specific activity levels would be higher and more in line with that of *Pf*OPRTase. The nickel affinity fractions for both *Pf*OPRTase and *Hs*OPRTase had been stored at 4 °C for 2–3 days prior to aliquoting and storing at –80 °C, this is likely a cause for a reduced specific activity compared to the literature. Enzyme folding, presence of the poly-histidine tag and lack of post-translational modification may also have contributed to the decreased activity or stability of *Hs*OPRTase.

The specific activity achieved here for *Pf*ODCase of 2.7–6.8 $\mu\text{mol}\cdot\text{min}^{-1}\cdot\text{mg}^{-1}$ was similar to the 9–12 $\mu\text{mol}\cdot\text{min}^{-1}\cdot\text{mg}^{-1}$ quoted by Krungkrai et al. (2005). The enzyme was not purified to apparent homogeneity here whereas it was by Krungkrai et al. (2005). The specific activity of *Hs*ODCase of 6.4–9.8 $\mu\text{mol}\cdot\text{min}^{-1}\cdot\text{mg}^{-1}$ was similar to that of *Pf*ODCase here and by Krungkrai et al. (2005). It was unsurprisingly orders of magnitude higher than the ODCase specific activity of 52 $\text{nmol}\cdot\text{hour}^{-1}\cdot\text{mg}^{-1}$ reported by Suchi et al. (1997) for cell extracts containing recombinant *Hs*UMPS rather than a purified sample.

3.3.3. Substrate kinetics of *Pf*ODCase and *Hs*ODCase

The calculated K_m^{app} for *Pf*ODCase of $8.9 \pm 1.15 \mu\text{M}$ correlates well with the K_m values quoted by Krungkrai et al. (2005). In that paper a K_m of $13.4 \pm 1.2 \mu\text{M}$ was found for the recombinant *Pf*ODCase (similar to the value found here) and $3.2 \pm 0.4 \mu\text{M}$ for the enzyme when purified from the parasite. Recombinant *Hs*OPRTase in this study had a K_m^{app} of $15.95 \pm 2.28 \mu\text{M}$. This is nearly 900-fold greater than the $295 \pm 18 \text{ nM}$ found for the baculovirus-expressed enzyme produced in Yablonski et al. (1996).

In both these cases the ODCases recombinantly expressed in *E. coli* had significantly lower binding affinities (higher K_m values) than enzymes either purified from the original organism or expressed in eukaryotic cells. This suggests that post-translational modification is the most likely cause for improved binding affinities of eukaryotic ODCases. This is supported by Krungkrai et al. (2005) where the K_m for *Pf*ODCase purified from the parasite was lower than that expressed in *E. coli*. The calculated K_m^{app} for *Hs*ODCase here when compared to baculovirus-expressed *Hs*ODCase in Yablonski et al. (1996) suggests a similar occurrence for the human homologue, although the difference is far greater. The presence of the N-Terminal tag did not result in a large difference between the recombinantly-expressed *Pf*ODCases produced here compared to Krungkrai et al. (2005). It may however be the cause of the extremely large difference in binding affinities of the *Hs*ODCases produced here compared to Yablonski et al. (1996). The radio assay used in Yablonski et al. (1996) would be needed for determining accurate K_m values.

3.3.4. Macromolecular Crystallography of *Pf*OPRTase

It remains unclear what the $10 \times 5 \mu\text{m}$ crystal from Hampton Research Crystal Screen Q31 was (with the $5 \times 10 \times 20 \text{ \AA}$ unit cell). It is unlikely that it would have been another protein (the enzyme sample was not completely pure) as it's not likely that there would have been enough to form such a large crystal. The other possibilities are that either the *Pf*OPRTase itself was degraded or cleaved at a specific spot such that part of the protein formed a crystal (also unlikely) or that one of the condition components—perhaps the 1,6 hexanediol—formed a crystal with a relatively large unit cell for a small molecule or salt. Without any indication as to what the crystal was and given it was unlikely to be useful in any case no attempt was made to solve an electron density map, or stain the crystal with Coomassie blue to confirm whether or not it was protein.

The Hampton Research crystal screen condition Q45 showed two distinctly different crystals. It was unfortunate that the large hexagonal crystal failed to diffract. The

smaller, asymmetric 'crystals' showed an interesting result, albeit not useful in solving the *PfOPRTase* structure. The cobalt(II) hexahydrate was clearly working as a molecular glue as the crystals were a purple colour. The asymmetry in the crystal and the powder diffraction patterns indicate that it did not possess a uniform crystal lattice. Rather it may be a slightly disordered but repeating pattern of protein. This may have started randomly as an aggregation event and built up over the course of several days from that.

The crystal structure 4FYM that was only released very recently used a condition containing 20 % (w/v) PEG-3350 and 0.18 M Potassium Sulphate pH 7.3 and was formed using hanging drop vapour diffusion at 4 °C. This condition is almost identical to the PEGs suite condition number 82 (0.2 M Potassium Sulphate, 20 % (w/v) PEG-3350). This condition—like most of the PEGs conditions—had simply yielded precipitated protein and did not develop from there. Were *PfOPRTase* to be screened against the PEGs suite again the protein concentration should be reduced, and if possible, the protein sample should be more purified.

4. Virtual Screening of Human and *P. falciparum* OPRTase and ODCase

4.1. Introduction

Currently no other known, successful *in silico* screening attempts have led to novel inhibitors of *Pf*OPRTase other than our own previous, unpublished work (Roach, 2007). Known OPRTase inhibitors include substrate and product analogues such as pyrazofurin (Scott et al., 1986) and several identified in Krungkrai et al. (2004a), and transition-state mimics and analogues (Witte et al., 2006, Zhang et al., 2013).

Most discoveries of ODCase inhibitors appear to be limited to analogues of known binders including uridine and UMP derivatives (Bello et al., 2007, Heinrich et al., 2009, Crandall et al., 2013), and other nucleoside 5'-monophosphate analogues (Langley et al., 2008, Purohit et al., 2012). Takashima et al. (2012) performed *in silico* screening on *Pf*ODCase and discovered 12 novel inhibitors. Most of these however had very high polar surface areas (PSA) and would be unsuitable for use as lead compounds.

While there is a modest pool of known inhibitors available for OPRTases and ODCases the scaffolds used are mostly analogues of either products or substrates. This scenario results in a greater chance of interference with other nucleoside binding enzymes. The discovery of a novel scaffold that binds one of these targets would also greatly increase flexibility in lead optimisation and understanding of the structure-activity relationship of the targets. Furthermore, novel scaffolds offer a greater chance of securing intellectual property rights.

A typical, structure-based screening method was used to identify potential competitive inhibitors of *P. falciparum* OPRTase, ODCase and the *H. sapiens* OPRTase and ODCase domains of UMPS. Experimentally solved structures for all targets except *Pf*OPRTase were available at the time of screening. Orthologous OPRTase structures were used to generate a homology model of *Pf*OPRTase. Structure-based screening was carried out with UCSF Dock v6.2 (Moustakas et al., 2006) on the eResearchSA Hydra cluster

(ersa.edu.au). The ChemDiv vendor subset of the UCSF ZINC database (Irwin and Shoichet, 2005) was screened.

4.2. Methods

4.2.1. Homology Modelling of *Pf*OPRTase

It was necessary to generate a homology model of *P. falciparum* OPRTase as an experimentally solved structure was not available at the time. A model had been generated previously using the *E. coli* structure (PDB ID: 1ORO) and successfully used to identify inhibitors (Roach, 2007). However, at the commencement of this study structures had subsequently been released of the *S. cerevisiae* OPRTase (Gonzalez-Segura et al., 2007) and the human OPRT domain of UMP synthase (Moche et al., 2009). Given the low sequence identity between the *P. falciparum* and *E. coli* OPRTases and that either of these new template structures were likely to be more suitable for use as a template for a homology model, it was decided that a new homology model needed to be generated.

4.2.1.1. Identification of Template Structures and Homologous Sequences

At the time of producing the homology model of *Pf*OPRTase there were several template structures available. The structures 2PRY for *S. cerevisiae* OPRTase and 2WNS for the *Homo sapiens* OPRT domain of UMPS were used to construct the homology model as they had sequence homology, above the 30% level generally accepted as suitable for homology modelling. The resolutions for 2PYR and 2WNS were 2.35 Å and 1.90 Å respectively. At these resolutions most features in the crystal structures should be accurately modelled with very few errors making them ideal for use as templates. The protein sequences for the templates were downloaded from the PDB (rcsb.org) for use with UCSF Modeller (Sali and Blundell, 1993, Fiser et al., 2000). The protein sequence for *Pf*OPRTase was obtained from the National Center for Biotechnology Information (NCBI, ncbi.nlm.nih.gov) (gi|20804383). The sequence identities for the *Pf*OPRTase against the *Hs*OPRTase and *S. cerevisiae* OPRTase was found to be

approximately 40 % and 24 % respectively (calculated as the identical residues over the average length of sequences, not taking into account gaps in the sequence). This level of sequence identity is ideal for homology modelling (Pitman and Menz, 2006).

Protein BLAST-P searches (Altschul et al., 1990) were performed on the *PfOPRTase*, *S. cerevisiae* OPRTase and *HsOPRTase* sequences to identify related OPRTase sequences. BLAST searches were run using the NCBI webservice (ncbi.nlm.nih.gov), with default algorithm parameters, screening the non-redundant protein sequences database. A total of 22 related sequences were obtained for use in the multiple sequence alignment to ensure that the target and template sequences were aligned by the conserved regions.

4.2.1.2. Multiple Sequence Alignment

Sequences from Section 4.2.1.1 were aligned using ClustalX v2.0.12 (Larkin et al., 2007) using default parameters for a complete multiple sequence alignment. The alignment was exported in .pir format. The convertpir.pl script (Appendix 8.15) was used to convert the .pir alignment file to the .ali format used by modeller.

4.2.1.3. Generation and Selection of Homology Model

The .ali file was manually modified in a text editor to include the start and end amino acids from the .pdb files, and to incorporate the ligand OMP from 2WNS into the homology model; this is shown in Appendix 8.16. The model-default.py example script from the modeller package was modified (Appendix 8.17) to use the input .ali alignment file (from Section 4.2.1.2) and the .pdb file that was downloaded from the PDB. All heterologous atoms except for OMP in the structure 2WNS were removed. The script generated 10 models of the target; multiple models were generated as there is a degree of variability with mapping the target to the template scaffold and this usually ensures at least one useable model.

The first 66 amino acids were removed from each of the models as there was no template structure for this region as this N-terminal extension is unique to *Pf*OPRTase. The models were then assessed for stereochemistry using PROCHECK (Laskowski et al., 1993) (from CCP4 v6.2.1 modelling package). PROCHECK uses stereochemical parameters from Morris et al. (1992) and Engh and Huber (1991) to check dihedral angles (ϕ , ψ , ω , χ 1–3), Ramachandran distribution, and main chain and side chain bond angles and lengths. The model with the best overall stereochemistry (> 95 % of residues in allowed region of a Ramachandran plot, fewest bad contacts, and fewest bond angle and length violations) was checked to ensure that none of the contact or bond errors occurred in conserved regions. The model was then further validated by docking and then used for virtual screening.

4.2.2. *In silico* Screening of *P. falciparum* and *H. sapiens* OPRTase and ODCase.

Structures were chosen as receptors for screening based on the resolution and whether or not they were co-crystallised with ligands (which ensure that the active sites are in a conformation which is able to bind potential inhibitors). The structure 2WNS was selected as the receptor for screening *Hs*OPRTase as it was the only structure available. Fortunately this structure met the above criteria being solved to a high resolution of 1.9 Å and being co-crystallised with the enzyme's product. The structures 2Q8Z (Bello et al., 2008) (1.8 Å, co-crystallised with a UMP analogue) and 2QCD (Wittmann et al., 2008) (2.03 Å, co-crystallised with UMP) were selected as the receptor structures for *Pf*ODCase and *Hs*ODCase respectively, both also meeting the above requirements. Structure files were prepared for docking and validated by docking of the co-crystallised ligand (or other known binder) prior to being used in screening for novel inhibitors.

4.2.2.1. File Preparation

Figure 4.1 outlines the programs and utilities used, and the files generated and modified in preparation for docking. The ligand coordinate files in .pdb format were prepared using UCSF Chimera (Pettersen et al., 2004). Protonation state was assessed, hydrogen atoms were added, and partial atomic charges were assigned according to Gasteiger-Marsili (Gasteiger and Marsili, 1980) and the file was converted to .mol2 format. The receptor structures were prepared for docking using Chimera's 'dockprep' tool, which also adds partial atomic charges and hydrogen atoms, as well as removing solvent, non-complexed ions and models incomplete side chains.

A .dms surface file of the receptor structure's active site was required by the Dock utility 'Sphgen' for generating the .sph spheres file. The spheres file is used by Dock during ligand orientation. The surface was calculated in Chimera for the active site residues and the .dms file was generated using Chimera's 'Write DMS' tool. Spheres were generated by Sphgen using default parameters.

The Dock utility 'Showbox' was used to generate the box file required for grid file generation. The box was generated for each .sph file to encompass the spheres with an additional 5 Å in all 6 directions ($\pm x, y, z$). The Dock utility 'Grid' was used to generate the grid files using default parameters unless otherwise stated.

Program/ Utility	Input	Output
Chimera	receptor.pdb (A) ligand.pdb (B)	receptor.mol2 (A) ligand.mol2 (B) active_site.dms
Sphgen	active_site.dms	active_site.sph (C)
Showbox	active_site.sph (C)	box.pdb (D)
Grid	receptor.mol2 (A) box.pdb (D)	grid.cnt (A/D) grid.nrg (A/D) grid.bmp (A/D)
Dock	ligand.mol2 (B) active_site.sph (C) grid.cnt (A/D) grid.nrg (A/D) grid.bmp (A/D)	docked_ligand.mol2 (E)

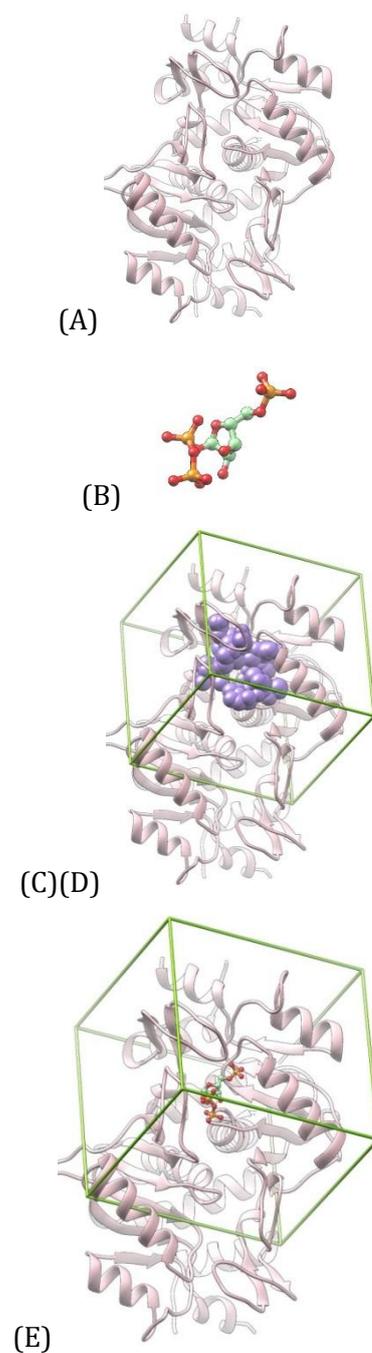


Figure 4.1: File preparation of receptor and ligand structure files for Docking with UCSF Dock 6. Pictures on right show (A) the receptor structure (B) ligand to be docked to the receptor (C) spheres used by dock (D) Box used by Grid and (E) the docked ligand (with box shown).

4.2.2.2. Optimisation and Validation for Docking

The receptor grid files and docking parameters were considered valid if the orientation of the co-crystallised ligand could be replicated by docking using the optimised high-stringency docking parameters (Appendix 8.18). Some initial high-stringency parameters (not shown) were used to attempt validation. Optimisation involved modification of the Grid input parameter file (and subsequent re-generation of the grid files) and small adjustments to the docking parameters. Adjustments that were made typically involved increasing the grid resolution by dropping the grid spacing from the default 0.3 Å to 0.2 Å (changes for each target tabled in Appendix 8.19). The docking parameters were optimised by way of a slight relaxation of the clash-overlap distance and maximum number of 'bumps' during ligand orientation, and by reducing the minimum anchor size to 2. This became the optimised high-stringency docking parameters for all targets.

4.2.2.3. Three Phase Screening

In silico screening of the ChemDiv subset of the UCSF ZINC database was carried out over three phases: very low, low and high stringency screening. The screening was carried out in three phases to minimise CPU wall time by eliminating compounds in the first two phases that were highly unlikely to bind. The very low and low stringency screens reduced the size of the compound library from approximately 1 million to 100 000 and from 100 000 to 10 000 respectively.

Very low-stringency and low-stringency docking parameters are shown in Appendices 8.20 and 8.21. The very low stringency settings involved only 50 attempted orientations. The low stringency parameters increased the attempted orientations to 150, more rigorous clustering and included ligand minimisation. High stringency screening involved a greatly increased number of attempted orientations of 500, included active chemical matching, and energy minimisation of the ligand, anchor and flexible growth.

4.2.3. Selection of Available Compounds for Biochemical Assays

Approximately the top 200 hits from the high stringency screen were visually analysed for their suitability as a potential inhibitor. Compounds were loaded into Chimera with the receptor structure using the 'viewdock' utility to display the hits. Compounds were removed from the potential candidate list if they appeared to adopt a stereochemically unfavourable or impossible orientation (occasional error with clashing ligand atoms etc.), bound partially or completely outside of the active site, or contained unrealistically high electrostatic scores (this was another error that occurred occasionally). Usually the electrostatic scores were several times greater than the VDW score but anything with the electrostatic score accounting for > 70 % of the Grid score was eliminated.

The UCSF ZINC website at the time termed some compounds as 'yuck' compounds. This was in later versions of ZINC clarified as compounds predicted to interfere with biochemical assays. The hitlist was further revised by removal of these identified problem compounds, or compounds that had large numbers of phosphate groups or other groups likely to result in non-specific protein binding.

Final selection of compounds on the ChemDiv website (chemdiv.com) involved eliminating duplicates (duplicate UCSF ZINC entries for one ChemDiv compound) and compounds that were not commercially available at the time. Compounds were prioritised based on the Dock Grid score and their diversity. Between four and eight compounds were selected for each of the four targets *HsOPRTase*, *HsODCase*, *PfOPRTase* and *PfODCase*.

4.3. Results

4.3.1. Homology Model of *Pf*OPRTase

The sequence alignment (Figure 4.2) showed homologous and conserved regions between the sequences. N-terminal extensions were present in many of the sequences from the *Plasmodium* species as well as *T. gondii*. There was however little to no homology between these extensions. Another small insert was present from approximately position 210–230 in only the *Plasmodium* and *T. gondii* OPRTases.

Ten models were generated and checked with PROCHECK. Of the ten homology models generated two were of a significantly higher quality than the others. Eight of the models had over 40 bad contacts between residues and most had a significant number of bad bond angles/lengths and residues outside of allowed regions on a Ramachandran plot (data not shown). As a comparison, analysis of the template structure 2WNS with PROCHECK identifies 46 bad contacts (exclusively involving heterologous atoms, mainly water molecules) and no residues in disallowed regions of a Ramachandran plot (data not shown).

The PROCHECK summary for the best model is shown in Figure 4.3. The model had 89.7 % of residues in most favoured region of a Ramachandran plot. This signifies a reasonable quality model (90 % is listed in the PROCHECK documentation as the baseline for a good model). Some residues were identified as outside of allowed regions on a Ramachandran plot (Figure 4.3 lines 5 and 7), or were identified to contain atoms that clash with atoms on neighbouring residues (bad contacts—line 13), or identified to have incorrect bond angles/lengths (lines 14 and 19). These residues were analysed visually in Chimera to determine if they were located in the active site or in highly conserved regions of the protein. The residues with disallowed Ramachandran angles were found to be located outside of the active site. The bad contacts also occurred outside of the active site and mostly outside of the conserved regions; however, clashing atoms were exclusively between Oxygen atoms, (mostly on main chain

peptides) with distances between the atoms of 2.5–2.6 Å. The “Find Clashes/Contacts” tool in Chimera confirms the bad contacts to be outside of the active site and not involving the main-chain atoms of highly conserved residues (data not shown). The final validation for the homology model was by docking the template structure’s co-crystalised ligand (Section 4.3.3).

```

*          20          *          40          *          60
P. berghei: OPRtase      : -----*-----*-----*-----*----- : -
P. chabaudi: OPRtase    : -----MEEHN--KEAHHISEEELHKKYNELCKKIELGK----- : 31
P. knowlesi: OPRtase    : -----MEEEKSTKQNTNYLSDEDLHKKYIHLRECIELEK----- : 33
P. vivax: OPRtase       : -----MLNRMERGEKSINQNTNYLSDEDLHKKYTQLRECIELEK----- : 37
P. falciparum: OPRtase  : -----MTTIKENEFLLCDEEIKYKSFVHLKDKICEERKKKELVN : 37
A. gossypii: HP         : -----MP----- : 2
K. lactis: HP           : -----MPT----- : 3
Z. rouxii: HP           : -----MPA----- : 3
C. glabrata: HP         : -----MP----- : 2
S. cerevisiae: OPRtase  : -----MP----- : 2
L. thermotolerans: HP   : -----MPA----- : 3
V. polyspora: HP       : -----MPA----- : 3
S. cerevisiae S288c: OPRtase : -----MSASTT----- : 6
C. albicans: OPRtase    : -----MLPFFSSHVEQHC----- : 13
L. elongisporus: OPRtase : -----MSSSSPQ----- : 7
S. stipitidis: OPRtase  : -----MES----- : 3
T. gondii: OPRtase      : -----MSSGNLSVSRRTLRKVLGPPGET----- : 25
H. sapiens: UMPS        : ----- : -
P. abelii: UMPS         : ----- : -
M. musculus: UMPS       : ----- : -
R. norvegicus: UMPS     : ----- : -
B. taurus: UMPS         : ----- : -
G. gallus: UMPS         : ----- : -
D. rerio: UMPS          : ----- : -

*          80          *          100         *          120
P. berghei: OPRtase      : -----*-----*-----*-----*----- : -
P. chabaudi: OPRtase    : -----AHENSDEIKEMKKLLDTLLIKYKALIFGNFVLLSKKTSHY : 72
P. knowlesi: OPRtase    : -----DDMDNCHVKEMKKLLDTALLKYKALKFGDFILKSKRKS : 74
P. vivax: OPRtase       : -----DDQ-NSHVKEMKSLDTALLKYNAIKFGDFILKSKRKS : 77
P. falciparum: OPRtase  : NNIDNVNFNDDDDNNYDDGNSYSSYIKEMKKLLKVLLKYKALK : 97
A. gossypii: HP         : -----VLEDYQKNFIDLAVESQSLRFGDFTLKSKRKS : 37
K. lactis: HP           : -----ALEDYQKNFLELAIESKALRFGDFTLKSKRVSE : 38
Z. rouxii: HP           : -----VLEDYQKNFLELAIESQALRFGDFTLKSKRKS : 38
C. glabrata: HP         : -----LEDYQKNFLELAIESQALKFGSFTLKSQRQSE : 36
S. cerevisiae: OPRtase  : -----IMLEDYQKNFLELAIECQALRFGSEKLLKSCRE : 38
L. thermotolerans: HP   : -----VLEDYQKNFLELAIESQALKFGFTLKSQRQSE : 38
V. polyspora: HP       : -----VLEDYQRNFLELAIESQALKFGSFTLKSQRQSE : 38
S. cerevisiae S288c: OPRtase : -----SLEEYQKTFLEGLCKALRFGSEKLNQRQSE : 41
C. albicans: OPRtase    : -----YKHQTTMASYKSSFTLALDSQALKFGFTLKSQR : 54
L. elongisporus: OPRtase : -----SYQANFQLALNCQALKFGFTLKSQRSE : 39
S. stipitidis: OPRtase  : -----YQTSFLESALESNALKFGSFTLKSQRSE : 34
T. gondii: OPRtase      : -----DEYQGRQLEDELQKALRLAYNHGALKFGSEKLLK : 66
H. sapiens: UMPS        : -----MAVARAALGPLVTGLYDVQAFKFGDFVLLKSGLS : 37
P. abelii: UMPS         : -----MAAVGAALGPLVTGLYDVQAFKFGDFVLLKSGLS : 37
M. musculus: UMPS       : -----MEVASQALGPLVTGLYDVQAFKFGDFVLLKSGLS : 37
R. norvegicus: UMPS     : -----MEVARQALGPLVTGLYDVQAFKFGDFVLLKSGLS : 37
B. taurus: UMPS         : -----MAAADALLGSLVTGLYDVQAFKFGDFVLLKSGLS : 37
G. gallus: UMPS         : -----MAAGGAGRVAALLLEARVRFKFGDFVLRSCIAS : 35
D. rerio: UMPS          : -----MDDTSLDSLTKLLHDIQAVKFGDFVLLKSGLS : 35

```

```

*          140          *          160          *          180
P. berghei: OPRtase : -----ISANIVSFLISNLLSKNIAFDYLLFGASYKGIPIVSTSHFLLNTNK--F : 48
P. chabaudi: OPRtase : VSTG-FLNNAISSNIVSFLISNLLSKNLSFDYLLFGASYKGIPIVSTSHFLLNTNK--F : 129
P. knowlesi: OPRtase : FSSG-VLNNIVSAHIIISFLISHLILKEKIPFDYLLFGASYKGIPIVSTSHFLFOSNK--F : 131
P. vivax: OPRtase : FSSG-VLNNIVSAHIIISFLISHLILKEKIPFDYLLFGASYKGIPIVSTSHFLFRSNK--F : 134
P. falciparum: OPRtase : FSSG-VLNNIVSSNIIIFLLSLELILKKNLSFDYLLFGASYKGIPIVSTSHFLFESK--Y : 154
A. gossypii: HP : FNLG-FNTGKLLSNLATAYAIAIIQSDLKFDVIFGPAYKGIPLAAVCAKLAIEGGTKF : 96
K. lactis: HP : FNLG-FNTGKLLSNLATAYAIAIIQSELKFDVIFGPAYKGIPLAAVCAKLAIEGGSKF : 97
Z. rouxii: HP : FNLG-FNTGKLLSNLATAYAIAIIQSDIKFDVIFGPAYKGIPLAAVCAKLAIEGGSKF : 97
C. glabrata: HP : FNLG-FNTGKLLSNLATAYAIAIIQSDLKFDVIFGPAYKGIPLAAVCAKLAIEGGSKF : 95
S. cerevisiae: OPRtase : FNLG-FNTGKLLSNLATAYAIAIIQSDLKFDVIFGPAYKGIPLAAVCAKLAIEGGSKF : 97
L. thermotolerans: HP : FNLG-QFNTGKLLSNLATAYAIAIIQSDLKFDVIFGPAYKGIPLAAVCAKLAIEGGTKF : 97
V. polyspora: HP : FNLG-FNTGKLLSNLATAYAIAIIHSDIKFDVIFGPAYKGIPLAAVCAKLAIEGGTKY : 97
S. cerevisiae S288c: OPRtase : FNLG-FNSGKLLANLATAYAATAIIQSELKFDVIFGPAYKGIPLAAVCAKLAIEGGTKF : 100
C. albicans: OPRtase : FNLG-FNTGKLLSNLATSYAEAIIASGLKFDVIFGPAYKGIPLAAVCAKLAELDPINY : 113
L. elongisporus: OPRtase : FNLG-FNTGALLSNLATSYAQAIVASGIEFDVIFGPAYKGIPLAAVCAKLAELAPEKY : 98
S. stipitis: OPRtase : FNLG-FNTGKLLSNLATSYAKAIIKSGLEFVIFGPAYKGIPLAAVCAKLAELDPHY : 93
T. gondii: OPRtase : FNAG-FVNGQAELISKAYATEIVRSGVDFVIFGPAYKGIPIVACTAMSLNRVYS--- : 122
H. sapiens: UMPS : IDLRGVSRPRLSQVADILFQTAQNAGISFDVTCGGVFPYALPLATVICSSTNQ----- : 90
P. abelii: UMPS : IDLRGVSRPRLSQVADTLFQTAQNAGISFDVTCGGVFPYALPLATVICSSTNQ----- : 90
M. musculus: UMPS : IDLRGVSRPRLSQVADILFQTAQNAGISFDVTCGGVFPYALPLATVICSSTNQ----- : 90
R. norvegicus: UMPS : IDLRGVSRPRLSQVADILFQTAQNAGISFDVTCGGVFPYALPLATVICSSTNQ----- : 90
B. taurus: UMPS : IDLRGVSRPRLSQVADTLFQTAQNAGISFDVTCGGVFPYALPLATVICSSTNQ----- : 90
G. gallus: UMPS : VDLRGLASRPSLRVADLDFQTAQNAGISFDVTCGGVFPYALPLATVICSSTNQ----- : 88
D. rerio: UMPS : EDLRGVSRPRLSQVADTLFQTAQNAGISFDVTCGGVFPYALPLATVICSSTNQ----- : 88

```

```

*          200          *          220          *          240
P. berghei: OPRtase : HNFYLYDRKPKKKEYGDKTILVGNIKESSQDCV--INSCNPQFEK-KKKVLIIDVDFTCG : 105
P. chabaudi: OPRtase : HNFYLYDRKPKKKEYGDKTILVGNLEEN---H--IGSAQVEKKTDKKVVLIIDVDFTCG : 183
P. knowlesi: OPRtase : SNFYLYDRKPKKKEYGDKTILVGNLDEEFNG----DVHNAKDEKNEKVVLIIDVDFTCG : 186
P. vivax: OPRtase : ANFYLYDRKPKKKEYGDKTILVGNLDEAVNG----DVHNAKAEASEKVVLIIDVDFTCG : 189
P. falciparum: OPRtase : SNFYLYDRKPKKKEYGDKTILVGNLDDDDKDIINLKKKTNNQDEEKKNI LIIDVDFTCG : 214
A. gossypii: HP : QDQYAFNRKPAKHEGEGGIVGAS--LNDQ-----RILIIDVDFTCG : 137
K. lactis: HP : QDQYAFNRKPAKHEGEGGIVGAS--LKDQ-----KILIIDVDFTCG : 138
Z. rouxii: HP : QDQYAFNRKPAKHEGEGGIVGAS--LEGK-----RILIIDVDFTCG : 138
C. glabrata: HP : QNQYAFNRKPAKHEGEGGIVGAS--LEGK-----KILIIDVDFTCG : 136
S. cerevisiae: OPRtase : QNQYAFNRKPAKHEGEGGIVGAS--LENK-----RILIIDVDFTCG : 138
L. thermotolerans: HP : QNQYAFNRKPAKHEGEGGIVGAS--LEEQ-----RVLIIDVDFTCG : 138
V. polyspora: HP : QNQYAFNRKPAKHEGEGGIVGAS--LENK-----RILIIDVDFTCG : 138
S. cerevisiae S288c: OPRtase : QGQYAFNRKPAKHEGEGGIVGAS--LEDK-----RVLIIDVDFTCG : 141
C. albicans: OPRtase : GDQYAFNRKPAKHEGEGGIVGAS--LKDQ-----KILIIDVDFTCG : 154
L. elongisporus: OPRtase : GNDQYAFNRKPAKHEGEGGIVGAS--LQNK-----RILIIDVDFTCG : 139
S. stipitis: OPRtase : GNLQYAFNRKPAKHEGEGGIVGAS--LKDQ-----KILIIDVDFTCG : 134
T. gondii: OPRtase : QPAPFIYDRKPKKKEYGDKTILVGNLQLEPKQI----PGTTDKYRPARVLLLDVDFTCG : 177
H. sapiens: UMPS : --PMLIRRKPAKHEGEGGIVGAS--LQNK-----TCLVIDVDFTCG : 129
P. abelii: UMPS : --PMLIRRKPAKHEGEGGIVGAS--LQNK-----TCLVIDVDFTCG : 129
M. musculus: UMPS : --PMLIRRKPAKHEGEGGIVGAS--LQNK-----TCLVIDVDFTCG : 129
R. norvegicus: UMPS : --PMLIRRKPAKHEGEGGIVGAS--LQNK-----TCLVIDVDFTCG : 129
B. taurus: UMPS : --PMLIRRKPAKHEGEGGIVGAS--LQNK-----TCLVIDVDFTCG : 129
G. gallus: UMPS : --PMLIRRKPAKHEGEGGIVGAS--LQNK-----TCLVIDVDFTCG : 127
D. rerio: UMPS : --YPMILIRRKPAKHEGEGGIVGAS--LQNK-----RCLVIDVDFTCG : 127

```

```

*          260          *          280          *          300
P. berghei: OPRtase : TALTEIFNKKKAYEY-LQVACIVLLNRFNEHEIN-ENNEKLYFKDLFEQKYNLEIYTT--- : 160
P. chabaudi: OPRtase : TALTEIFNKKKAFDY-LEIVACIVLLNRFNEHEIN-EKNEKLYFKDLFEQKYNLEIYTT--- : 238
P. knowlesi: OPRtase : TALTEIMNKKLSYPN-LKVVAFIVLLNRFNEYELN-EHNEKLYFKDLFEQKYNLEIYTT--- : 241
P. vivax: OPRtase : TALTEIINKKLSYPN-LRVVALIVLLNRFNEYELN-EQNEKLYFKDLFEQKYNLEIYTT--- : 244
P. falciparum: OPRtase : TALTEILAKLKTYEH-LKVVAFIVLLNRFNEYEIN-ENNQKLYFKDIFEKRYGIVLNY--- : 269
A. gossypii: HP : TAINAEFEITAKTAGG--KVVGTIHALDROEIVDT-TSKEGATATQSVSKRYDIPVLS--- : 191
K. lactis: HP : TAINAEFEITAKEEG--KVVGTIHALDROEVVNT-EDTEVSATQSVSKRYDIPVLS--- : 192
Z. rouxii: HP : TAINAEFEITANKG--HTVGTIHALDROEIVIST-EDREGSATQSVSKRYDIPVLS--- : 192
C. glabrata: HP : TAINAEFEITAAANQG--QVVGSIHALDROEVVSPDSTDSATQAVSKRYDIPVLS--- : 191
S. cerevisiae: OPRtase : TAINAEFEITSNARG--QVVGSIHALDROEVVSTDD-KEGATATQSVSKRYDIPVLS--- : 192
L. thermotolerans: HP : TAINAEFEITAAQNG--NVVGTIVSLDROEVLST-ESKEPISATQAVTQKHGIPVLS--- : 192
V. polyspora: HP : TAINAEFEITAKENG--QVVGSIHALDROEIVNS-EDRDGATATQSVSKRYDIPVLS--- : 192
S. cerevisiae S288c: OPRtase : TAINAEFEITIAQGG--RVVGTIHALDROEVIHE-SDPERTSATQSVSKRYDIPVLS--- : 195
C. albicans: OPRtase : TAINAEFEITIGNEK--QVVGSIHALDROEITAT---DPTKSATQAVSERYGIPVLS--- : 206
L. elongisporus: OPRtase : TAINAEFEITIAQEG--ECVGTIHALDROEITVQ---DPTKSATKAVSERYGIPVLA--- : 191
S. stipitis: OPRtase : TAINAEFEITIAAEG--NVVGTIHALDROEITAN---SDT-SATQAVSERYGIPVLS--- : 185
T. gondii: OPRtase : TATRGNMKLLQELQHTVEVAGIFVLLDROERVSE---DADISAAEQLEETKYTKVFS--- : 231
H. sapiens: UMPS : SSVLETAEVLEQKEGL--KVTDAIVLLDRECGGKDKLQAAGRLHNSVTLSEMLEEQQK : 187
P. abelii: UMPS : SSVLETAEVLEQKEGL--KVTDAIVLLDRECGGKDKLQAAGRLHNSVTLSEMLEEQQK : 187
M. musculus: UMPS : ASVLETAEVLEQKEGL--KVTDAIVLLDRECGGKDKLQAAGRLHNSVTLSEMLEEQQK : 187
R. norvegicus: UMPS : ASVLETAEVLEQKEGL--KVTDAIVLLDRECGGKDKLQAAGRLHNSVTLSEMLEEQQK : 187
B. taurus: UMPS : SSVLETAEVLEQKEGL--KVTDAIVLLDRECGGRDNLQAAGRLHNSVTLSEMLEEQQK : 187
G. gallus: UMPS : SSVLETAEVLRKEGL--KVTDAIVLLDRECGGKARLEEHGRLHNSVTLSEMLEEQQK : 185
D. rerio: UMPS : SSVLETAEVLEQKEGL--KVTDAIVLLDRECGGSTRLDAGSLHNSVTLSEMLEEQQK : 185

```

```

*           320           *           340           *           360
P. berghei: OPRTase : ----- : -
P. chabaudi: OPRTase : -----VLSYNDDISHLIK-- : 251
P. knowlesi: OPRTase : -----LISYNEDLEPMG-- : 254
P. vivax: OPRTase : -----LISYHEDLEPMG-- : 257
P. falciparum: OPRTase : -----LISYKDDIQSVI--- : 281
A. gossypii: HP : -----VSLADLIAYDGR : 206
K. lactis: HP : -----VNLSNLSYDGR : 207
Z. rouxii: HP : -----VSLSHINYEGR : 207
C. glabrata: HP : -----VLSLHVINF DGR : 206
S. cerevisiae: OPRTase : -----VSLIHITYEGR : 207
L. thermotolerans: HP : -----VTLADISYDGR : 207
V. polyspora: HP : -----VTLTDITYEGK : 207
S. cerevisiae S288c: OPRT : -----VSLTQVQFMGNRL : 210
C. albicans: OPRTase : -----VNLGEVSIINGK : 221
L. elongisporus: OPRTase : -----VNLKEVSI SGK : 206
S. stipitis: OPRTase : -----VSLSDVEISHKL : 200
T. gondii: OPRTase : -----VNLMDLLFEEL : 246
H. sapiens: UMPS : KVDAETVGRVKRFIQENVFAANHNGSPLSIKEAPKELSFSGARAEPRIHVASKLRLM : 247
P. abelii: UMPS : KIDAETVGRVKRFIQENVFAANHNGSPLSIKEAPKELSFSGARAEPRIHVASKLRLM : 247
M. musculus: UMPS : KIDADMVGRVKRFIQENVFAANHNGLPPEKKACKELSFSGARAEVPGTHVASKLRLM : 247
R. norvegicus: UMPS : KIDAEVGRVKRFIQENVFTAANHNGVPPPEKKACKELSFSGARAEVGVHPVASKLRLM : 247
B. taurus: UMPS : KINAETVERVKRFIQENAFVAANPNDSLPSVKKEPKELSFSGARAEVPGTHVAAKLRLM : 247
G. gallus: UMPS : EVDVEMVEKVNFIQGNVFEFGARNG-FTPVKRVCKELSFSGARAEVGVHPVAARLRLM : 244
D. rerio: UMPS : RIDTDTAQSVKRFVQDNNTYILKKNNG-SSAAKKSCELSYGARAGVPTHPVAARLRLM : 244

*           380           *           400           *           420
P. berghei: OPRTase : ----- : -
P. chabaudi: OPRTase : ----- : -
P. knowlesi: OPRTase : ----- : -
P. vivax: OPRTase : ----- : -
P. falciparum: OPRTase : ----- : -
A. gossypii: HP : SP-----ADKRRVDEYRQTYGA----- : 223
K. lactis: HP : SV-----EEREKVEQYRQTYGASA----- : 226
Z. rouxii: HP : SA-----DEKKEVEYRNTYGV----- : 224
C. glabrata: HP : TA-----EEKKVEDYLQTYGA----- : 223
S. cerevisiae: OPRTase : TA-----EEKKVEQYLQTYGASA----- : 226
L. thermotolerans: HP : PA-----EEKKVEEYLQIYGV----- : 224
V. polyspora: HP : SP-----EEKEKVEYRQNFV----- : 224
S. cerevisiae S288c: OPRT : SP-----EQKSAVENYRKAYGI----- : 227
C. albicans: OPRTase : ND-----EDLKSVEQYRSKYGA----- : 238
L. elongisporus: OPRTase : KD-----EDLKSVENYRSQYGA----- : 223
S. stipitis: OPRTase : TE-----EQLASIKKEYRQYSPKN----- : 219
T. gondii: OPRTase : SSGDHLEPRETLQKARNVILEYRRKYGVTDDELPTDVM----- : 283
H. sapiens: UMPS : QKKETNLCLSADVSLARELQLADALGPSICMLKTHVDILNDFTLDVMKELITLAKRHEF : 307
P. abelii: UMPS : QKKETNLCLSADVSEARELQLADALGPSICMLKTHVDILNDFTLDVMKELITLAKRHEF : 307
M. musculus: UMPS : QKKETNLCLSADVSEARELQLADALGPSICMLKTHVDILNDFTLDVMKELITLAKRHEF : 307
R. norvegicus: UMPS : QKKETNLCLSADVSEARELQLADALGPSICMLKTHVDILNDFTLDVMKELITLAKRHEF : 307
B. taurus: UMPS : QKKETNLCLSADVSEARELQLADALGPSICMLKTHVDILNDFTLDVMKELITLAKRHEF : 307
G. gallus: UMPS : EKKQTNLCLSADVTSKELQLAASLGPSICILKTHVDILNDFTEVVKELRALADQHEF : 304
D. rerio: UMPS : EDKKTNLCLSADVTHSKELLDIAVTLGPLICVLKTHVDILEDFTADVASNKELAKKHF : 304

```

Figure 4.2: Multiple sequence alignment of OPRTases for use with generating a Homology Model of PfOPRTase. The last approximately 300 residues of the sequence alignment is omitted as it contains only the ODCase domains of the UMPSs. Entries are listed as organism and protein (HP = Hypothetical Protein). NCBI protein accessions (ncbi.nlm.nih.gov/protein/) for entries from top to bottom are as follows: XP_668850.1, XP_744110.1, XP_002259686, XP_001613829, BAB92089.1, NP_985147.1, XP_453152.1, XP_002498586, XP_447993.1, 2PRY_A, XP_002552406, XP_001643041, P30402, XP_718318.1, XP_001527683, XP_001386033, XP_002365160, CAG33068.1, NP_001126858, NP_033497.1, NP_001020573.1, AAI12873, NP_001026431.1, NP_956468.2

```

1  +-----<<< P R O C H E C K       S U M M A R Y >>>-----+
2  |
3  | E:/CCP4/FILES/pf3.pdb      2.0                                402 residues |
4  |
5  *| Ramachandran plot:      89.7% core      7.1% allow      1.6% gener      1.6% disall |
6  |
7  *| All Ramachandrans:      16 labelled residues (out of 398)
8  +| Chil-chi2 plots:        8 labelled residues (out of 300)
9  |
10 | Main-chain params:       6 better      0 inside      0 worse
11 | Side-chain params:      5 better      0 inside      0 worse
12 |
13 *| Residue properties: Max.deviation:      4.0                      Bad contacts:  22 |
14 *|                        Bond len/angle:  9.0      Morris et al class:  1  1  2 |
15 +|      2 cis-peptides
16 | G-factors      Dihedrals: -0.15  Covalent: -0.38      Overall: -0.22 |
17 |
18 | M/c bond lengths: 98.9% within limits  1.1% highlighted
19 *| M/c bond angles:  89.0% within limits  11.0% highlighted      2 off graph |
20 | Planar groups:    100.0% within limits  0.0% highlighted
21 |
22 +-----+
23 + May be worth investigating further.  * Worth investigating further.

```

Figure 4.3: PROCHECK summary for best *Pf*OPRTase homology model. Line numbers are included. An asterisk appears for a particular property where errors occur in the model that may be significant.

4.3.2. Comparison of Homology Model to Crystal Structure

A crystal structure of *Pf*OPRTase was released only relatively recently (Rathod and Kumar, 2013) solved to a resolution of 2.60 Å. A superposition of the homology model to chain C of the crystal structure was performed in Chimera using the matchmaker tool. The tool forms an automated superposition using the best-aligning pair of chains between the two structures (Needleman-Wunsch algorithm (Needleman and Wunsch, 1970) with the BLOSUM-62 matrix (Henikoff and Henikoff, 1992)). The RMSD between 4FYM chain C and the homology model Chain A was calculated to be 2.58 Å (using UCSF Chimera's matchmaker tool: best aligning pair of chains with default parameters and pruning turned off. RMSD calculated for atoms present in both structures). The superposition shows some significant differences in short segments of the alpha-carbon chain (Figure 4.4.A and Figure 4.4.B). A beta hairpin exists between GLU 243 and TYR 254 that was not in the model (Figure 4.4.A, top left). Because of this hairpin the model was misaligned for the neighbouring alpha-helix from PHE 255 to VAL 263 and the model contained a small loop from PRO 266 that was not in the crystal structure

(Figure 4.4.A, middle). The 66 amino acid extension at the N-terminus of *Pf*OPRTase was omitted from the homology model. The insert forms a large helix hairpin in the crystal structure (Figure 4.4.A, right side).

There were only minor differences in the positions of active site side chain atoms for the majority of active site residues. There was however, a significant difference in the placement of the active site flexible loop (GLY 84 to ASN 95, data not shown) for the crystal structure monomers. Superposition of the crystal structure monomers (using the same method) revealed there were at minimum two different conformations of this loop in the apo crystal structure. Two chains show an open conformation, with one chain showing a closed conformation (chain C) and the other chains having only a few of the loop residues modelled. Chain G is partially modelled but potentially shows a third conformation. The homology model adopted a conformation of this flexible loop most closely resembling chain C (closed conformation). Figure 4.4.B (right side) shows the displacement of the flexible loop from the crystal structure and the misalignment of these residues.

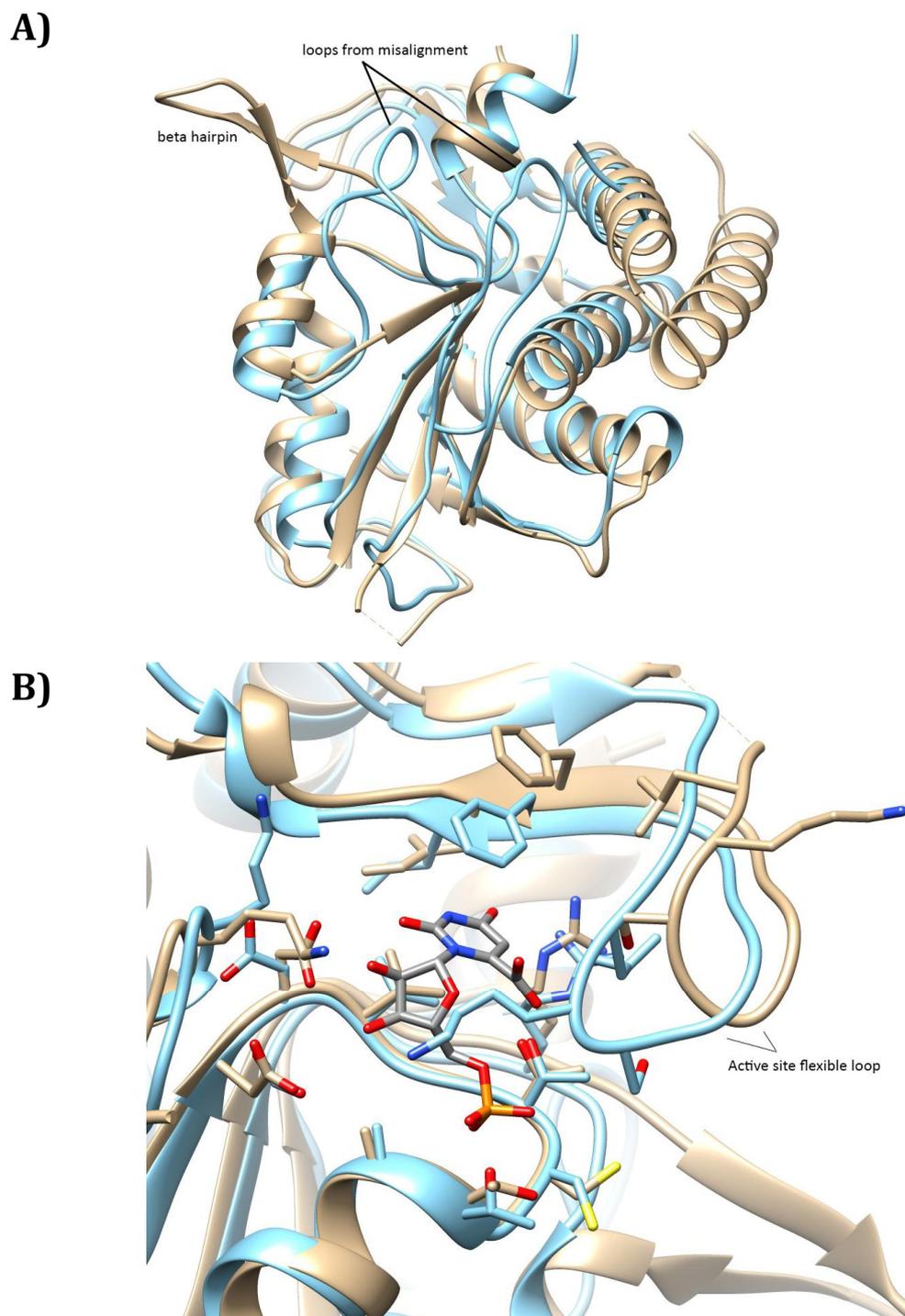


Figure 4.4: Superposition (Section 4.3.2) of homology model and crystal structure of *PfOPRTase*. Enzymes depicted as ribbon view with the crystal structure coloured brown and homology model light blue. Non-carbon atoms are coloured by CPK and carbon atoms coloured to match the ribbon. **A)** Superposition of crystal structure (PDB ID: 4FYM) Chain C and homology model chain A. Ribbon view, Atoms hidden. **B)** The superposition of A with active site atoms and aligned OMP product shown.

4.3.3. Validation of Receptor Structures

Structures were considered valid if they docked the ligand in a similar orientation to that observed in the crystal structure. A docked orientation was considered correct if the docked ligand's atoms were less than 2 Å RMSD from the co-crystallised ligand orientation. The substrate orotate was able to be docked into the *Pf*OPRTase homology model in the correct orientation however docking of the other substrate PRPP did not result in the correct orientation. A new file containing the homology model and docked orotate was made and new grid files generated for this structure. Docking to the homology model in the presence of orotate resulted in the correct orientation of PRPP. Figure 4.5.A shows the crystal structure of one of the two template structures used to generate the homology model with the co-crystallised substrates; this was used as a comparison to determine correct orientation of substrates docked in the homology model. Figure 4.5.B shows the homology model with the docked orotate and PRPP. The orientation of orotate is very similar to that in the crystal structure and there is only a small difference between the docked and the co-crystallised orientations of PRPP.

OMP was not able to be docked in the correct orientation for *Hs*OPRTase despite several attempts. However, orotate and PRPP were both docked in the correct orientations with several slight variations of docked PRPP orientation (Figure 4.6.B, highest scoring orientation shown only). PRPP was not docked to the enzyme with orotate pre-docked (unlike for docking to *Pf*OPRTase homology model) and as a consequence there is some overlap between the docked position of PRPP and orotate as shown in Figure 4.6.B. Given the flexibility displayed by PRPP in binding orientations this was not considered to be a limitation of the homology model. Comparing with Figure 4.6.A, the orotate ring and phosphoribose group of OMP are in similar positions and orientations to that of the docked orotate and the phosphoribose group of the docked PRPP. The approximate 180 degree rotation of the ring group of the docked PRPP, relative to the co-crystallised conformation, was not considered an issue as the

correct rotation can be achieved without significant displacement of the other ligand atoms.

Figures 4.7 and 4.8 show the comparisons between the co-crystallised and docked ligand orientations for *HsODCase* and *PfODCase* respectively. In both cases the docked ligand orientations were almost identical to that of the co-crystallised ligand orientations.

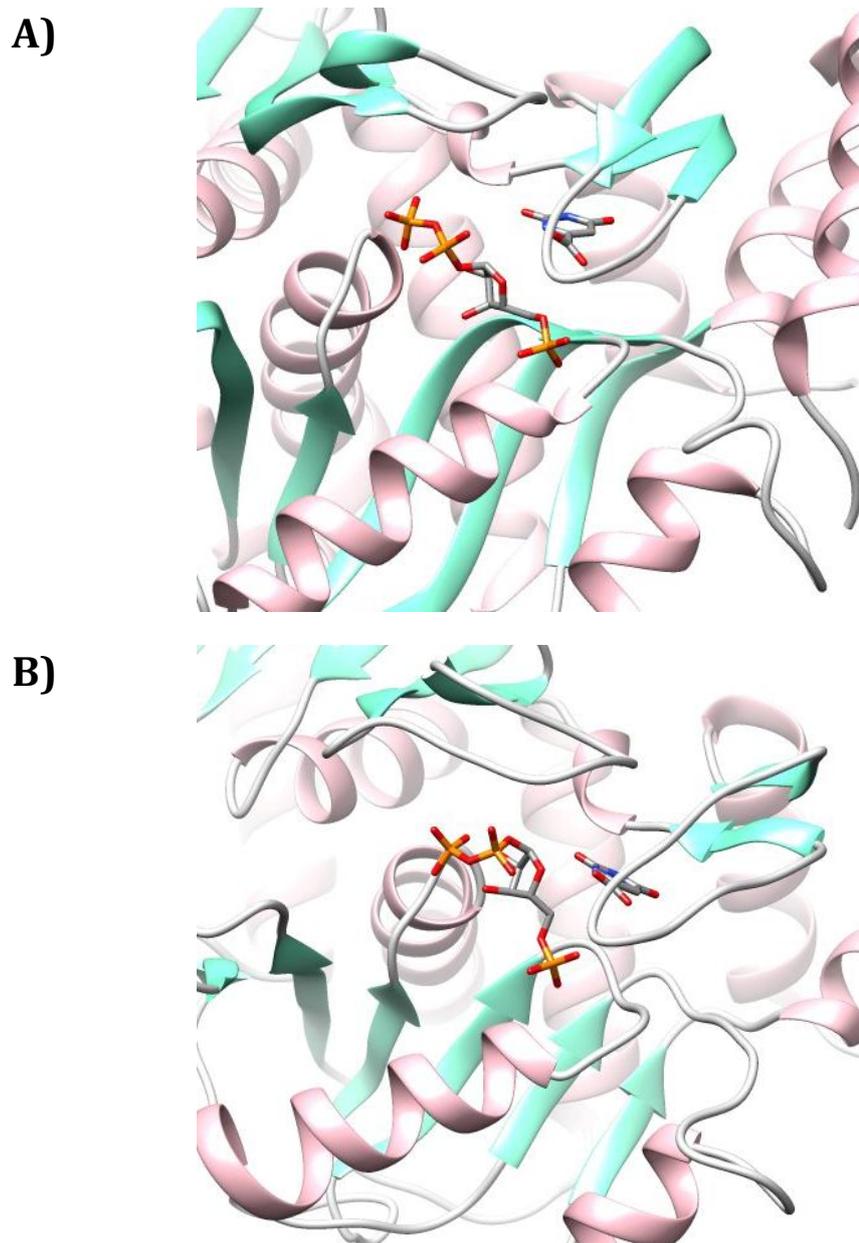


Figure 4.5: Validation of *Pf*OPRTase receptor structure. Images of ribbon-view enzyme structures (coloured by secondary structure) used for *in silico* screening comparing the orientations of co-crystallised ligands with docked ligand orientations. **A)** *S. cerevisiae* OPRTase Crystal structure 2PS1 co-crystallised with orotate and PRPP. **B)** *Pf*OPRTase homology model with orotate and PRPP docked with Grid scores of -21.0 and -121.9 respectively.

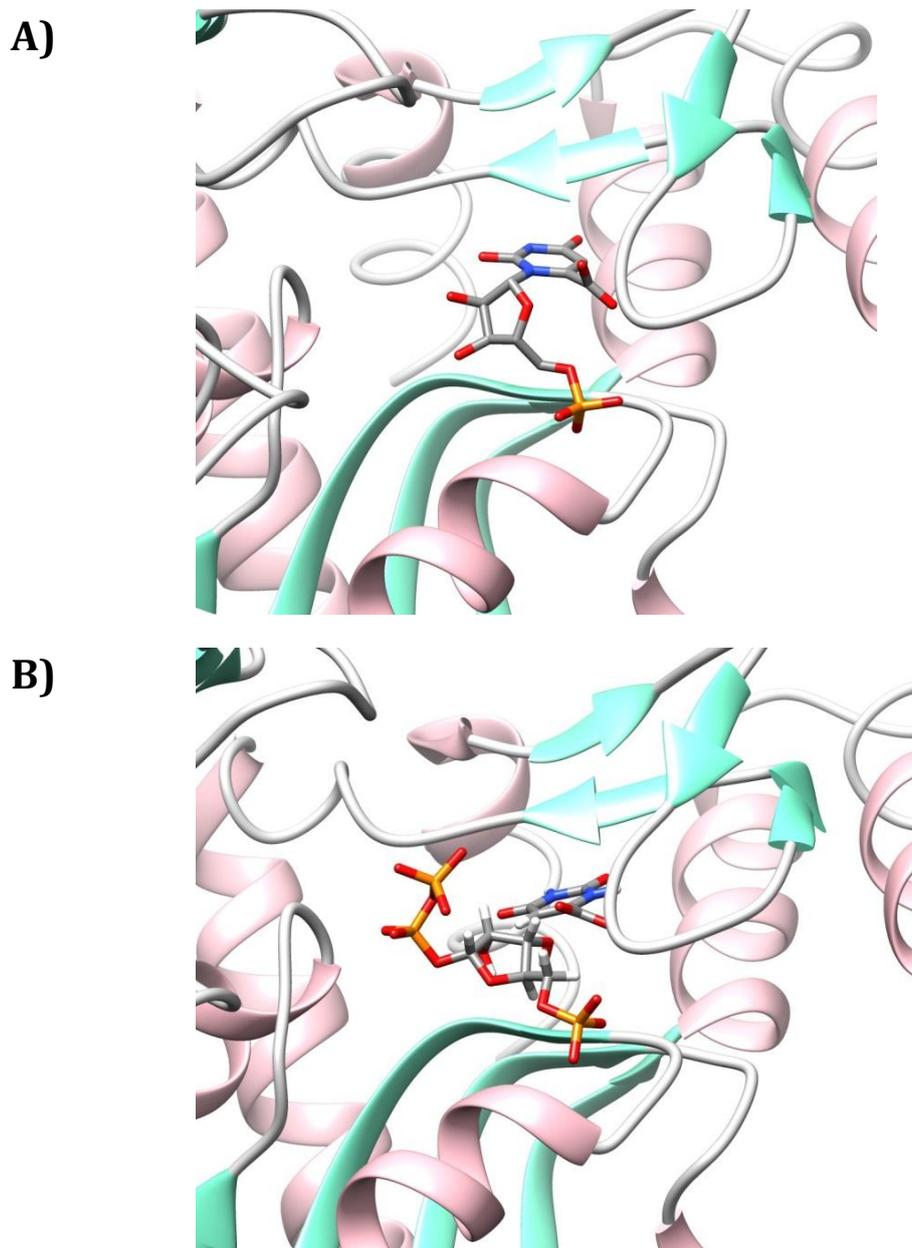


Figure 4.6: Validation of *HsOPRTase* receptor structure. Images of ribbon-view enzyme structures (coloured by secondary structure) used for *in silico* screening comparing the orientations of co-crystallised ligands with docked ligand orientations. **A)** *HsOPRTase* crystal structure 2WNS co-crystallised with the product OMP. **B)** *HsOPRTase* crystal structure with orotate and PRPP docked with Grid scores of -31.0 and -90.2 respectively.

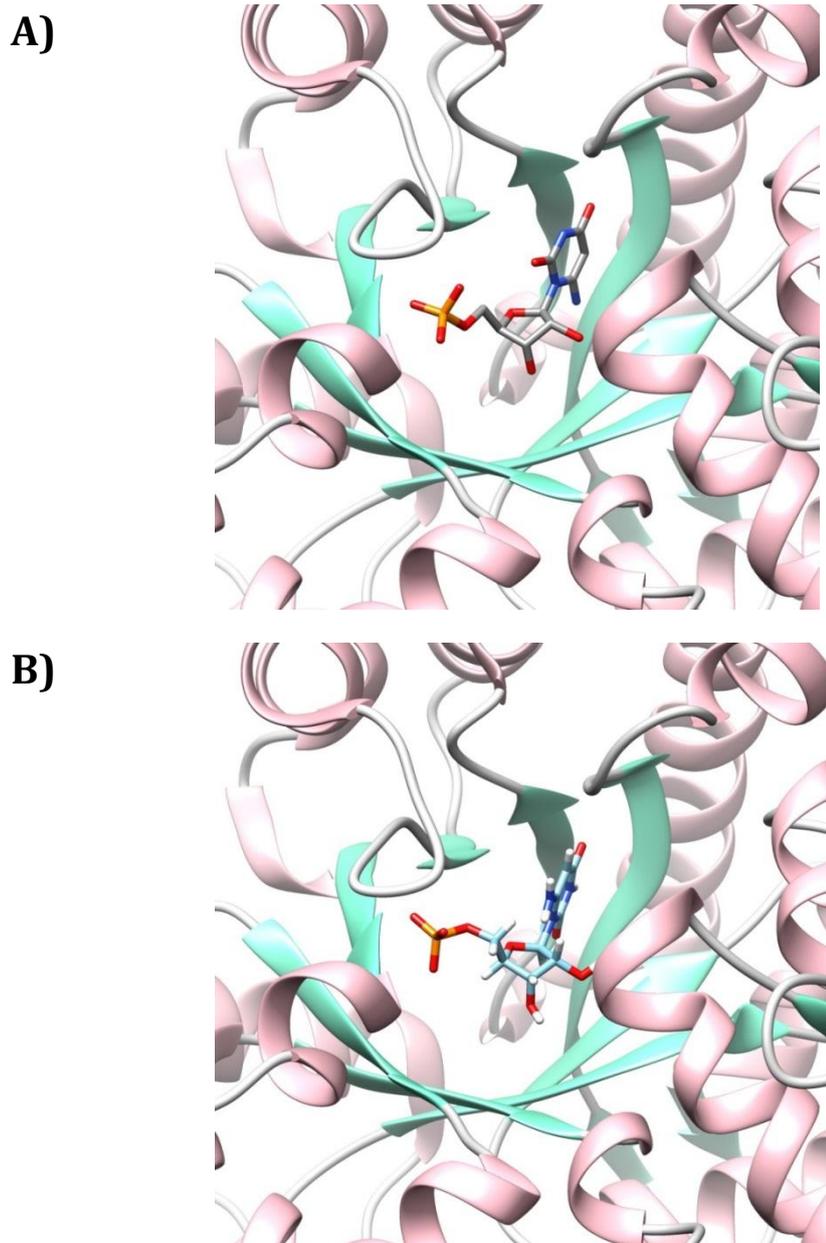


Figure 4.7: Validation of *PfODCase* receptor structure. Images of ribbon-view enzyme structures (coloured by secondary structure) used for *in silico* screening comparing the orientations of co-crystallised ligands with docked ligand orientations. **A)** *PfODCase* Crystal structure 2Q8Z with co-crystallised ligand (6-aminouridine 5'-monophosphate). **B)** *PfODCase* Crystal structure with the ligand docked with a -65.7 Grid score.

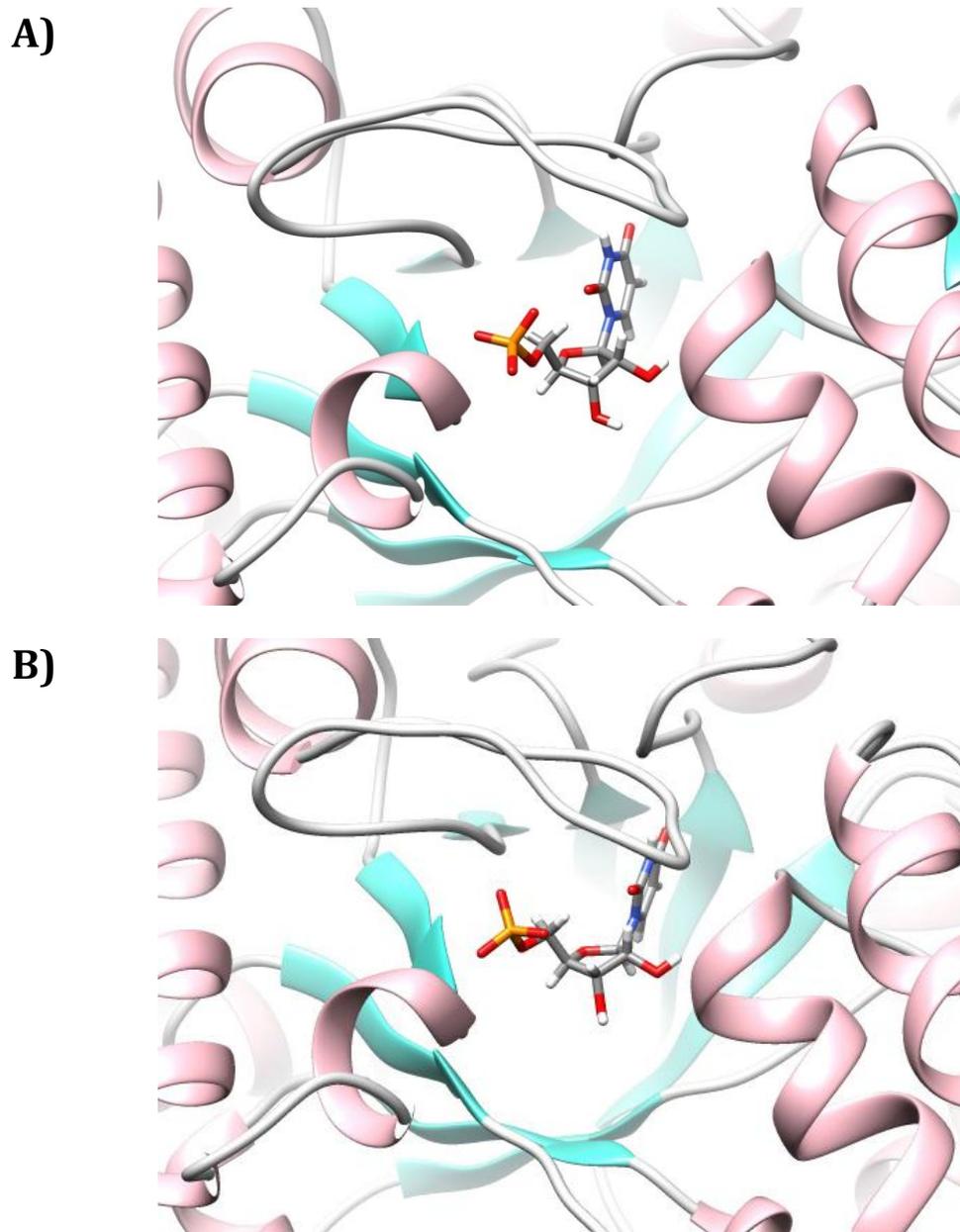


Figure 4.8: Validation of *HsODCase* receptor structure. Images of ribbon-view enzyme structures (coloured by secondary structure) used for *in silico* screening comparing the orientations of co-crystallised ligands with docked ligand orientations. **A)** *HsODCase* crystal structure 2QCD co-crystallised with UMP. **B)** *HsODCase* crystal structure with docked UMP ligand with a -78.3 Grid score.

4.3.4. Structure-Based *in silico* Screening

The three phase, structure-based *in silico* screening was completed in approximately one month per target enzyme (total wall time not calculated). The Grid scores for the top scoring compounds docked to *PfOPRTase*, *PfODCase*, *HsOPRTase* and *HsODCase* that were considered suitable for further evaluation (as described in 4.2.3) are shown in Figure 4.9 and Figure 4.10. A large number of compounds for all four screens were eliminated for various reasons. These included problem compounds identified by ZINC, compounds that either bound partially or completely outside of the active-site or erroneously with atom clashes (both assessed visually in Chimera; these occurrences are unmistakable), and compounds with extremely high electrostatic scores.

For the *P. falciparum* targets there was a much lower variation in Grid scores between the top hits compared to the human targets. *HsODCase* hit Grid scores (Figure 4.10.B) were both the lowest of the four targets and the most varied. For the *P. falciparum* targets where the Grid scores were far less varied, a stronger emphasis was placed on other aspects when choosing which compounds to test in biochemical assays. This included ligand rigidity (number of rotatable bonds), scaffold similarity to known binders, active site coverage, and electro-static score (hydrophobic compounds tend to be better lead candidates).

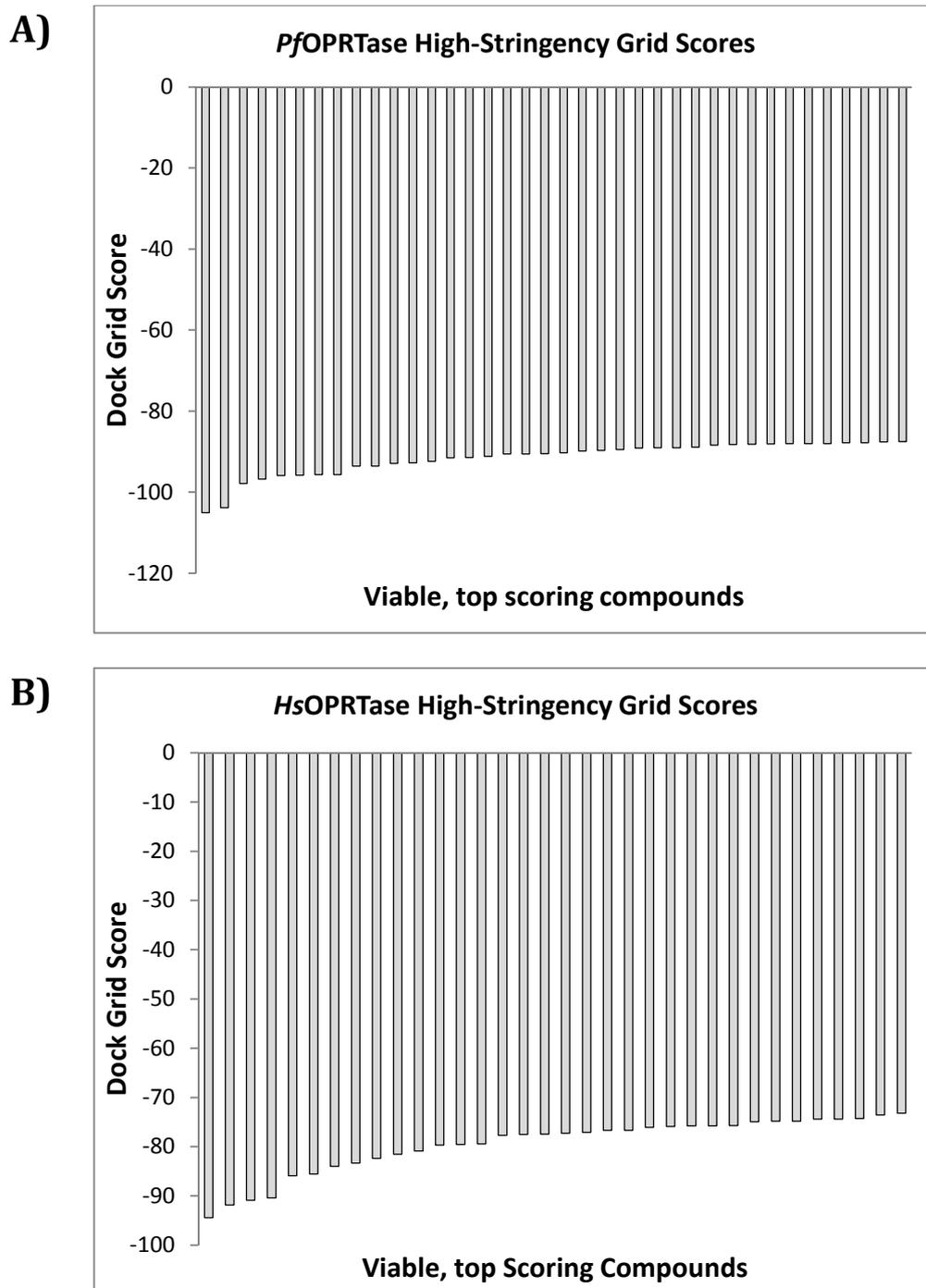


Figure 4.9: Grid scores of the top compounds from structure based screening against OPRTases. Column graphs of top Grid scores for top scoring compounds after filtering from the final screening phase (high-stringency screening) for **A)** *PfOPRTase* **B)** *HsOPRTase*.

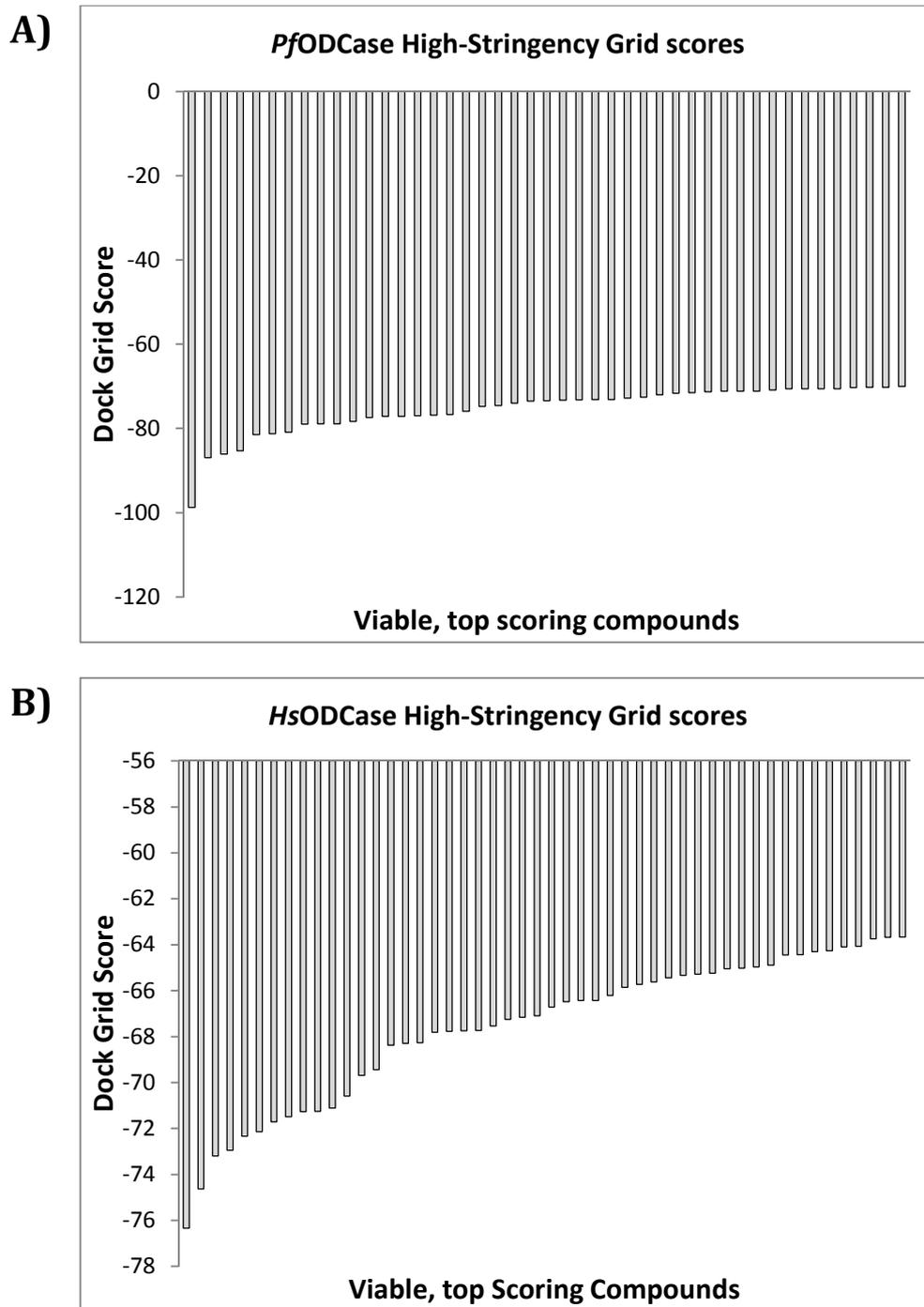


Figure 4.10: Grid scores of the top compounds from structure based screening against ODCases. Column graphs of top Grid scores for top scoring compounds after filtering from the final screening phase (high-stringency screening) for **A)** *Pf*ODCase and **B)** *Hs*ODCase.

4.3.5. Final Selection of Compounds

While the Grid score was the primary parameter for ranking compounds, final selection for evaluation in biochemical assays involved a range of other criteria. Figure 4.11 shows the final selected compounds, their target, Grid score, and reason or reasons for selection. Many top scoring hits were eliminated as they were structurally very similar to a previously selected compound. Compounds were eliminated because they had too many highly polar groups (such as phosphate) that are likely to result in non-specific binding (no more than 3 phosphate and/or ionic groups; no greater than Lipinski's rule for H-bond donors, acceptors, and polar surface area). In most instances the compounds were selected based on their Grid score and occupancy of the active site, and to maximise for diversity (assessed visually, but limited to eliminating compounds that were obviously very similar to another higher scoring compound). Finally, some compounds were not chosen as they were not available from the supplier at the time the biochemical assays were conducted.

For each of *Pf*OPRTase and *Hs*OPRTase four compounds were selected. In most cases the compounds were predicted to completely occupy the active site and have ring or H-bonding groups in similar positions to those groups on known binders. The compound ZINC15163251 (Figure 4.11.A top-middle) had a large number of rotatable bonds. This is not normally desired but as it had a very high Grid score it was selected anyway. Compounds ZINC00317635 and ZINC00167151 (Figure 4.11.A middle-middle and bottom-middle) were both chosen as they had very high Grid scores for relatively small compounds. For these three cases the compounds would be ideal lead-candidates as they could allow for modification of the compound to maximise binding affinity. Optimisation would involve increasing rigidity for ZINC15163251 and increasing the size of ZINC00317635 and ZINC00167151 by adding groups to complement the active site.

Screening of *Pf*ODCase returned a large number of diverse hits and as such eight compounds were chosen (Figure 4.11.B). ZINC22108888 was a top scoring compound for *Hs*ODCase as well as *Pf*ODCase. Common features in most of the hits were an aromatic ring group deep in the active site where the orotate ring on OMP would bind and a highly polar group (such as a phosphate or a carboxyl) near the opening of the active site where the phosphate group on OMP would be positioned.

For *Hs*ODCase four compounds were selected. Three of the four compounds (ZINC01782151, ZINC22108888, and ZINC03269047) had an aromatic ring group in the orotate pocket, as was seen in the *Pf*ODCase hits. Unlike the *Pf*ODCase hits there were hydrophobic methyl groups (ZINC22108888, ZINC03269047) or halides (ZINC01782151) attached to the aromatic rings on these compounds. ZINC22108888—the compound that was a top hit for both ODCase targets—was the only compound in the *Pf*ODCase hits that contained a hydrophobic group on the aromatic ring binding in the orotate binding pocket. The ring group for this compound was not positioned as deeply in the orotate binding pocket for *Pf*OPRTase as it was for *Hs*OPRTase. All of this might suggest a subtle difference in the conformations of the active sites that could be exploited for specific binding of a compound to one enzyme over the other.

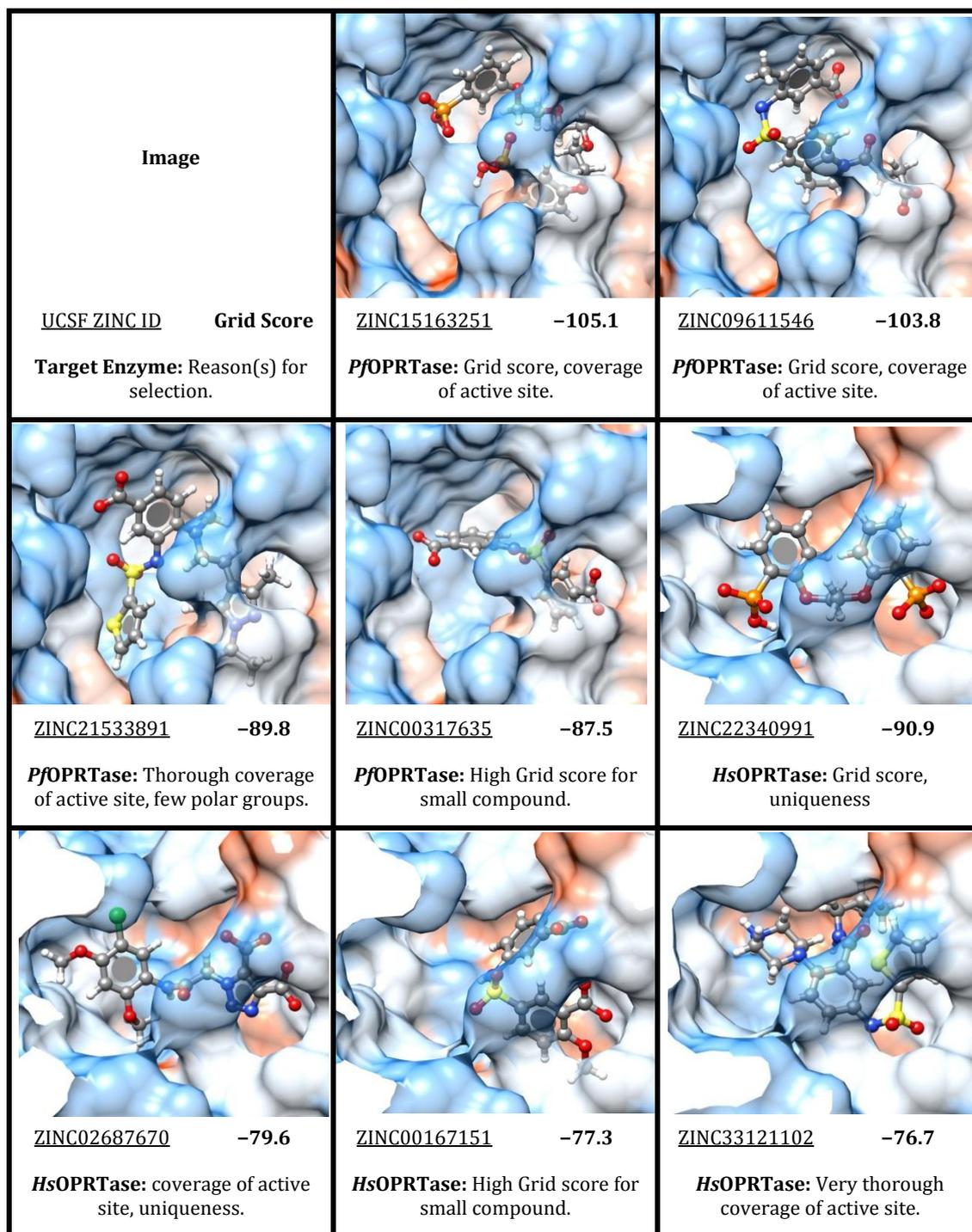


Figure 4.11.A: Final selection of potential *PfOPRTase* and *HsOPRTase* binders: Compounds 1–8. Enzymes are rendered as 70 % transparent surface view (hydrophobic, orange; hydrophilic, blue). Ligands are shown as ball and stick and coloured by CPK.

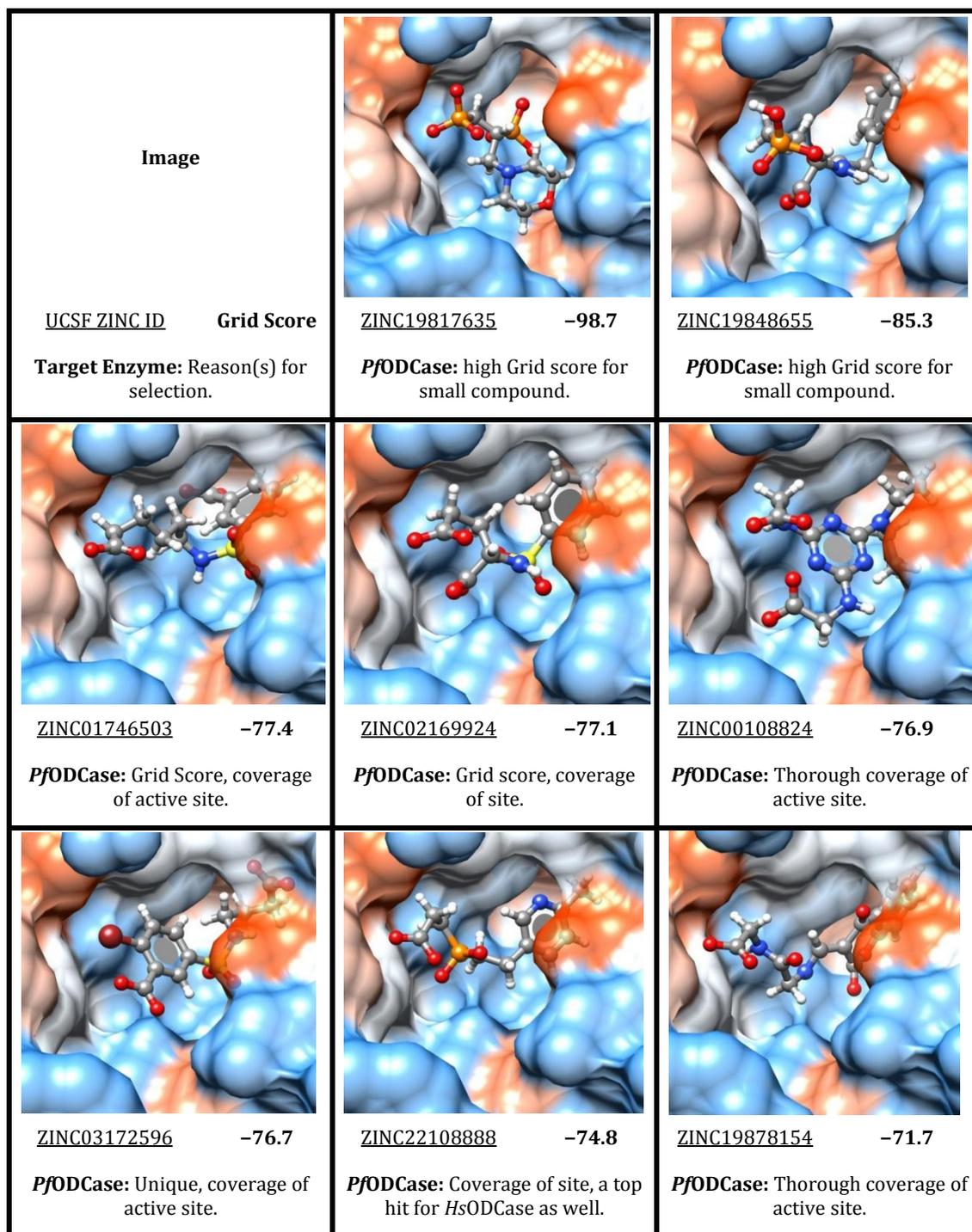


Figure 4.11.B: Final selection of potential *PfODCase* binders: Compounds 9–16.

Enzymes are rendered as 50% transparent surface view (hydrophobic, orange; hydrophilic, blue). Some residues are not drawn in order to provide a clear view of the active site. Ligands are shown as ball and stick and coloured by CPK.

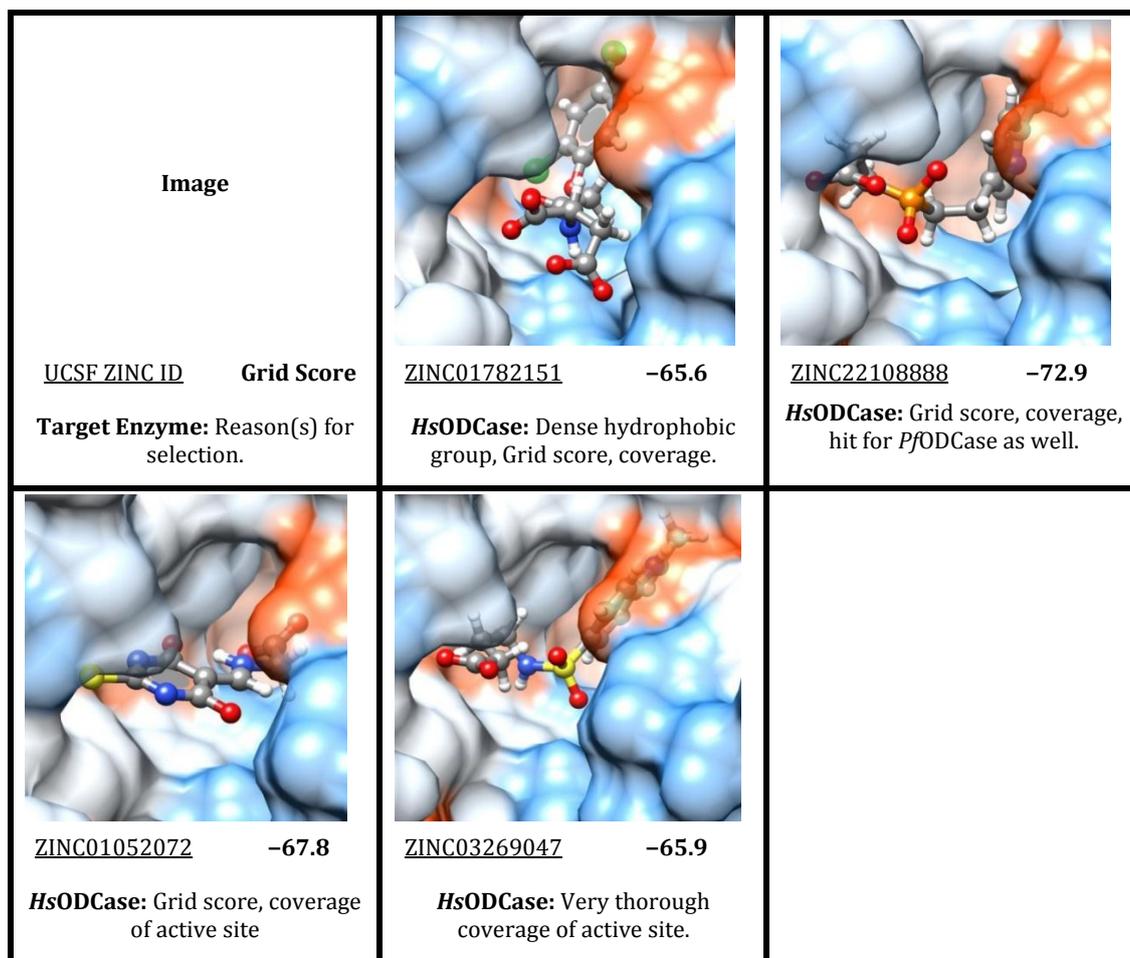


Figure 4.11.C: Final selection of potential of *HsODCase* binders: Compounds 17–20. Enzymes are rendered as 50% transparent surface view (hydrophobic, orange; hydrophilic, blue). Some residues are not drawn in order to provide a clear view of the active site. Ligands are shown as ball and stick and coloured by CPK.

4.4. Discussion

4.4.1. Homology Model of PfOPRTase

The Homology model was generated with the ligands of the crystal structure 2WNS positioned in the active site. This had the advantage of helping to ensure that the active site was in a conformation in which it was able to bind the known binders. This approach is analogous to choosing crystal structures co-crystallised with ligands for the other three targets. A comparison with previous projects where this approach had not been taken (Haynes, 2008, Roach, 2007) highlighted that this step appeared to be advantageous (data not shown). It is likely that none of the models generated would have been suitable for use in docking were it not for this step.

Comparison to the crystal structure (Section 4.3.2) showed the homology model to be very similar to the crystal structure of the enzyme that was subsequently solved with a RMSD of 2.58 Å (especially when taking into account the crystal structure's resolution of 2.60 Å). Minor errors in the alignment were evident from the comparison to the crystal structure, some of which may have been impossible to avoid due to the features not being present in the template structures, such as the misalignment arising from the unique beta hairpin between GLU 243 and TYR 254. These structural features may be unique to *Plasmodium* and would need to be in a template structure to be incorporated into the model.

Misalignment errors of highly conserved regions are generally avoided by using an alignment with a diverse range of homologous sequences (Pitman and Menz, 2006).

The method used here included a step to find other homologous sequences and include them in the alignment to help ensure the conserved regions are identified and aligned.

The related enzymes that were used in this alignment were the result of BlastP searches against the target and template sequences. This was very quick and simple to perform but meant that the homologous sequences were closely related to the template or target sequence. It may be better to pick sequences with equally distant homology

between the target and template sequences. For instance, instead of matching two Animalia with two Protozoa sequences it is likely the conserved regions would be better aligned by matching the Animalia and Protozoa sequences together with a Fungi and a Plantae sequence.

There was a significant difference in the active site flexible loop. The results from the *in silico* screening are still considered valid for several reasons. The loop in the *P. falciparum* crystal structure has shown several different conformations. The structure was not derived from a crystallised enzyme complexed with a known binder and it is likely there may be yet another conformation that the flexible loop assumes when bound to its substrates, product, intermediates, or inhibitors. It is known that this loop conformation changes significantly during binding and catalysis. The active site lysine on the flexible loop (seen in Figure 4.4) is facing away from the active site in the crystal structure where it is known to play a role in binding the substrate (Henriksen et al., 1996). The homology model was also validated by docking. A new crystal structure of *P. falciparum* with a co-crystallised ligand would best show the active conformation of this flexible loop in the *P. falciparum* OPRTase but Molecular Dynamics (MD) simulations may make this apo structure more useful in future structure-based screening projects.

While no published *in silico* screening results using a homology model of *Pf*OPRTase could be found, a homology model was used in the work of Zhang et al. (2010). In this work a homology model was used with Fourier transform infrared spectrometry to characterise the hydrogen bonding network with orotate. The model was generated by SWISS-MODEL (Biasini et al., 2014) with the *S. cerevisiae* OPRTase crystal structure 2PS1 (Gonzalez-Segura et al., 2007) as the template. Other homology models for *Pf*OPRTase were also available on the SWISS-MODEL repository. However, automatically generated models were not considered for use with *in silico* screening here as they tend to be less accurate (Mosimann et al., 1995, Bates et al., 1997).

4.4.2. Validation of Receptor Structures

For receptor structure and docking parameter validation the ligands had to be docked in the correct orientation. The correct orientation was considered to be a replication of the orientation seen in the crystal structure with co-crystallised ligand. There were some minor differences in orientations as mentioned but these were not considered significant.

The OMP substrate has two defining features: the orotate ring and the phosphoribose group. Whilst PRPP is comprised of the phosphoribose group and the pyrophosphate group. For the OPRTase receptor structures the orotate, phosphoribose and pyrophosphate groups were positioned in the same binding pockets as the co-crystallised ligands with only minor rotational differences. The only exception to this was the 180 degree rotation of the phosphoribose ring group for the docked PRPP ligand to the *HsOPRTase* receptor (Figure 4.6). The structure was still considered acceptable as the groups were in their correct binding pockets and the ring could easily be rotated without a significant re-positioning of the phosphate or pyrophosphate groups.

It is not clear why OMP could not be docked to the *HsOPRTase* receptor, especially given that the enzyme was co-crystallised with OMP, and also that PRPP and orotate were both successfully docked. This may simply be a nuance of Dock v6.2 or of the anchor-grow method of orientating the ligand. Different versions of Dock or another docking program may be able to successfully dock OMP to this receptor structure. The reason for the failure to dock OMP might otherwise stem from the lack of enzyme flexibility during docking. If the active site is in a very tightly-bound state to OMP then it may be that a very minor orientation difference between the crystal orientation and the predicted Dock orientation could trigger an atom-clash event where the orientation is rejected by Dock. This is usually overcome (and was attempted here) by relaxing the restraints around atomic clashes during ligand orientation.

Whatever the reason for the failure to dock OMP, there is almost certainly a significant potential false negative rate arising in the virtual screening results. Without biochemically testing all ~1 million compounds the false negative rate will remain unknown. Failure to dock rates of known actives to structures—even when validated with a known binder—can be quite substantial. This is discussed in Chapter 6 where failure to dock rates of known actives ranged from approximately 25 % to greater than 99 %. However, false negatives can still be identified through hit expansion (H2L strategy) as long as a structural analogue of the false negative that is an inhibitor is identified during docking and tested biochemically.

While some false negatives are unavoidable with this type of screening it does highlight the importance of the initial selection of the target structure to minimise the impact. Validation of multiple structures with multiple known binders would aid in receptor structure selection. Where CPU wall time is not an issue, multiple screens on different receptor structures (with slightly different conformations) of the same target would also help address the issue. Flexible receptor docking is currently a ‘holy grail’ for structure based screening. While most docking programs include methods for limited receptor flexibility (induced fit) some offer complete flexibility (such as Flipdock (Zhao and Sanner, 2007) and MORDOR (mondale.ucsf.edu/index_mordor.html)). Unfortunately it is still considered a ‘holy grail’ of docking as the CPU wall time required increases exponentially as the extent of receptor flexibility is increased.

4.4.3. Structure Based *in silico* Screening

In some instances the known binder, during validation, was not able to be docked into the active site in the correct orientation (but docked nonetheless) under the phase one screening parameters (data not shown). This may lead to an increase in false negatives during the first screening phase.

The compounds that were not likely to bind (would not fit in the active site) had been eliminated by failure to dock in the first phase as was intended. The remaining

compounds, particularly for the ODCase screens, were very few in number and the second phase was completed in a fraction of the time (data not shown). The docking parameters could have been made more stringent in the second phase. The maximum conformations to attempt could be increased greatly to ensure the best orientation for each compound leading into the third phase screen. The third phase screen also wouldn't necessarily need to re-orient the ligands if the second phase screen had attempted enough orientations. Instead the CPU wall time could be focused on performing energy minimisation on the top several orientations for the top scoring ligands (from the second phase screen) and re-scoring.

As mentioned, purging of ligands from the third phase screening involved eliminating compounds if they were not completely positioned within the active site, as well as those that were in an unrealistic pose (usually clashing ligand atoms, an error that occurs occasionally), or contained groups or properties likely to lead to non-specific binding. The top approximately 100 compounds formed a viable hit list from which the compounds would be selected for biochemical inhibition assays.

Compounds containing groups likely to lead to non-specific binding have also been a problem in previous in-house projects (Haynes, 2008, Roach, 2007, Stenson, 2011). When screening completely unrelated enzymes in the past (an older homology model of *PfOPRTase*, an earlier screen of *PfODCase* and an alternative NADH dehydrogenase), among the top of hits for both screens were several compounds that outscored nearly all others. The most notorious of these compounds were simply aromatic rings with 3–5 phosphate groups and thus likely to interact with nearly any enzyme. The library that was used in these previous projects had a larger number of these problem compounds and this meant that there were fewer viable hits among the top scoring compounds (data not shown). 'Frequent hitters' and compounds with reactive species are well-documented problems with library preparation and hit filtering (Roche et al., 2001, Rishton, 2003, Merkwirth et al., 2004, Brown et al., 2006, Crisman et al., 2007, Baell and Holloway, 2010, Schorpp et al., 2013, Curpan et al., 2014, Nissink and Blackburn, 2014).

While the ChemDiv subset of the UCSF ZINC library contained for the most part drug-like compounds, there were still a number of problem compounds that were eliminated during compound selection. The use of a database filter (such as OpenEye FILTER (Ertl et al., 2000)) to remove these compounds prior to screening would reduce overall CPU wall time and save time from having to manually purge them. Bologa et al. (2006) recommends removing problem compounds prior to screening (partly to reduce screening time but also for the purpose of tailoring a library to a particular enzyme). The cost associated includes a significant portion of time spent in eliminating unwanted compounds. Depending on how well the filtering program can be optimised to remove unwanted compounds as well as compounds unlikely to bind, it may eliminate the need for multiple phase screening.

4.4.4. Final Selection of Compounds

Final selection of the compounds was a combination of Grid score, availability, uniqueness, and maximising for diversity. As the four targets were closely related the intention from the outset was to screen all compounds against all targets. The assumption was that a compound that was predicted to bind to one of the targets had a reasonable chance of binding to any of the other closely related targets.

The *Pf*ODCase screen yielded by far the most diverse and unique compounds. As such there were more compounds selected from this hit list than the others. Four compounds were selected each for the other targets to complete the 19 compounds for which the project had budgeted.

An improvement to this selection method in the future could be to cross reference the Grid scores of hit compounds from the *P. falciparum* enzyme with the human homologues. The compounds with the greatest difference in scores between the two enzymes would be the best candidates for specific binders. The drawback to this is that the compounds selected would not necessarily be the outright top scoring compounds. This would likely result in a reduced true-positive hit rate and/or weaker inhibitors.

5. Identification and Characterisation of Inhibitors of Human and *P. falciparum* OPRTase and ODCase

5.1. Introduction

Chapter 4 identified a number of compounds that might be potential inhibitors of *P. falciparum* and *H. sapiens* OPRTase and ODCase. Nineteen compounds were purchased from the supplier ChemDiv and had to be tested experimentally to determine whether or not they inhibited the target enzymes. Confirmed inhibitors had to be characterised further to more accurately assess their binding affinities.

A typical method for biochemical testing of compounds that are identified by *in silico* screening involves a high-throughput approach. This is because the number of compounds being tested can be in the hundreds. Therefore biochemical assays are typically conducted without replicates and at a single concentration of the compound. This is usually performed spectrophotometrically in large scale format (96-well or 384-well plates for instance). Assays are later replicated if they show inhibition by the compound to confirm the result. Compounds that are identified as inhibitors are then characterised further by determination of the half maximal inhibitory concentration (IC_{50}), dissociation constant (K_d) or inhibition constant (K_i).

Several problems usually occur when assaying large numbers of compounds. The solubility can vary greatly between compounds. Some compounds will absorb at the same wavelength as the product or substrate that is being measured. The initial screening can lead to false negatives if the concentration of the compound is too low. Solubility issues are usually mitigated by dissolving all compounds in DMSO and including DMSO in the control assay to the same percentage. Spectrophotometric interference and false negatives are typically managed by screening compounds at a set concentration that is low enough that most compounds will not absorb significant amounts of light but high enough for the weaker inhibitors to be identified.

The 19 compounds that were purchased (Chapter 4) were assayed to determine if they were inhibitors of any of the four enzymes. Activity assays were conducted at a single concentration of each compound. Due to the relatively low number of compounds being tested, all of these assays were performed in duplicate. A small scale H2L strategy known as 'hit expansion' (Section 1.6) was conducted to expand the pool of inhibitors and these were also screened at a single concentration. Confirmed inhibitors were characterised further to determine the inhibition constants and modes of inhibition.

5.2. Results: Initial Biochemical Inhibition Screen of Compounds

Compounds that were selected and purchased from ChemDiv are shown in Figure 5.1. These were prepared following the protocol outlined in Section 2.5.1 and inhibition assays were carried out as detailed in Sections 2.5.2 and 2.5.3. For OPRTase the compounds were screened initially at a single concentration that ranged from 0.05–2 mM depending on the compound. For ODCase the compounds were screened at a single concentration that ranged from 0.05–5 mM depending on the compound. Some compounds had to be screened at lower concentrations in order to keep the total absorbance of the assay within the operating range of the spectrophotometer. The concentrations of other compounds were adjusted (usually increased) in subsequent screening assays to maximise the chance of identifying weak binders. This is detailed in Sections 5.2.1–5.2.4.

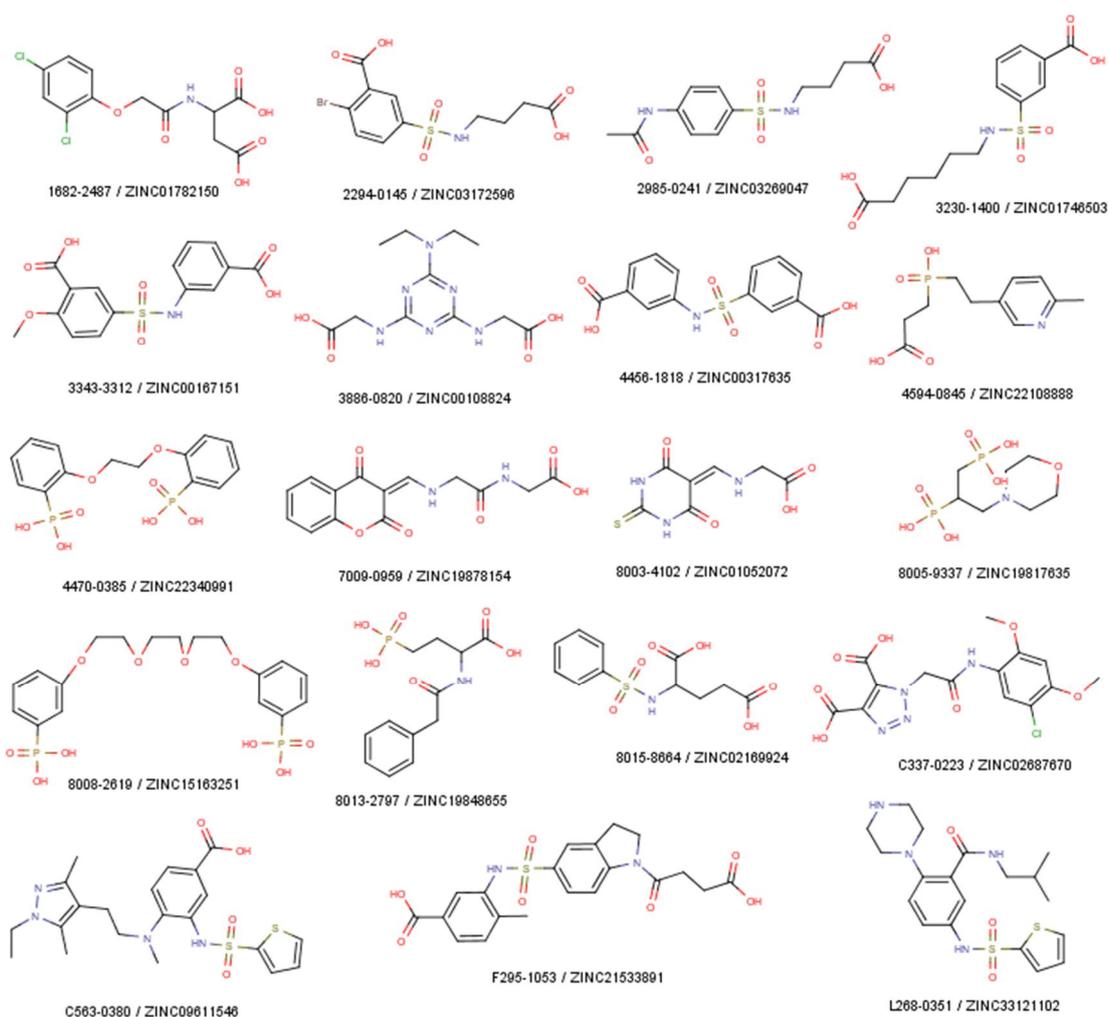


Figure 5.1: Compounds identified by *in silico* screening and obtained from ChemDiv (San Diego, USA). Compounds that were identified as potential binders of the human and *P. falciparum* OPRTase and ODCase enzymes were identified in Chapter 4. The ChemDiv catalogue number and UCSF ZINC ID for each compound are shown beneath the structures of each compound.

5.2.1. Inhibitors of *HsOPRTase*

Human OPRTase was assayed with each of the 19 compounds identified in Chapter 4 at a concentration of 0.5 mM. It was hoped that this would be a low enough concentration that spectrophotometric interference would be minimal but still high enough to identify weaker inhibitors. Compounds 7009-0959, C337-0223, and C565-0380 yielded an absorbance baseline that exceeded the limits of the spectrophotometer. After some dilutions it was found that 50 μ M was a low enough concentration to not interfere with the absorbance reading and these compounds were therefore assayed at this concentration.

Figure 5.2 shows the initial inhibition screen of the compounds against *HsOPRTase*. A number of compounds inhibited human OPRTase. The most promising were compounds 7009-0959 and C565-0380 which showed approximately 25 % inhibition at 50 μ M. While this wasn't the highest level of inhibition seen among the compounds, it was a significant level of inhibition at a much lower concentration of the compounds. This would likely translate to a much lower inhibition constant. Compounds 3230-1400 and F295-1053 showed approximately 35 % inhibition at 0.5 mM. Compounds 2294-0145, 8013-2797, 2985-0241, and 4594-0845 showed modest inhibition of approximately 25 % also at 0.5 mM.

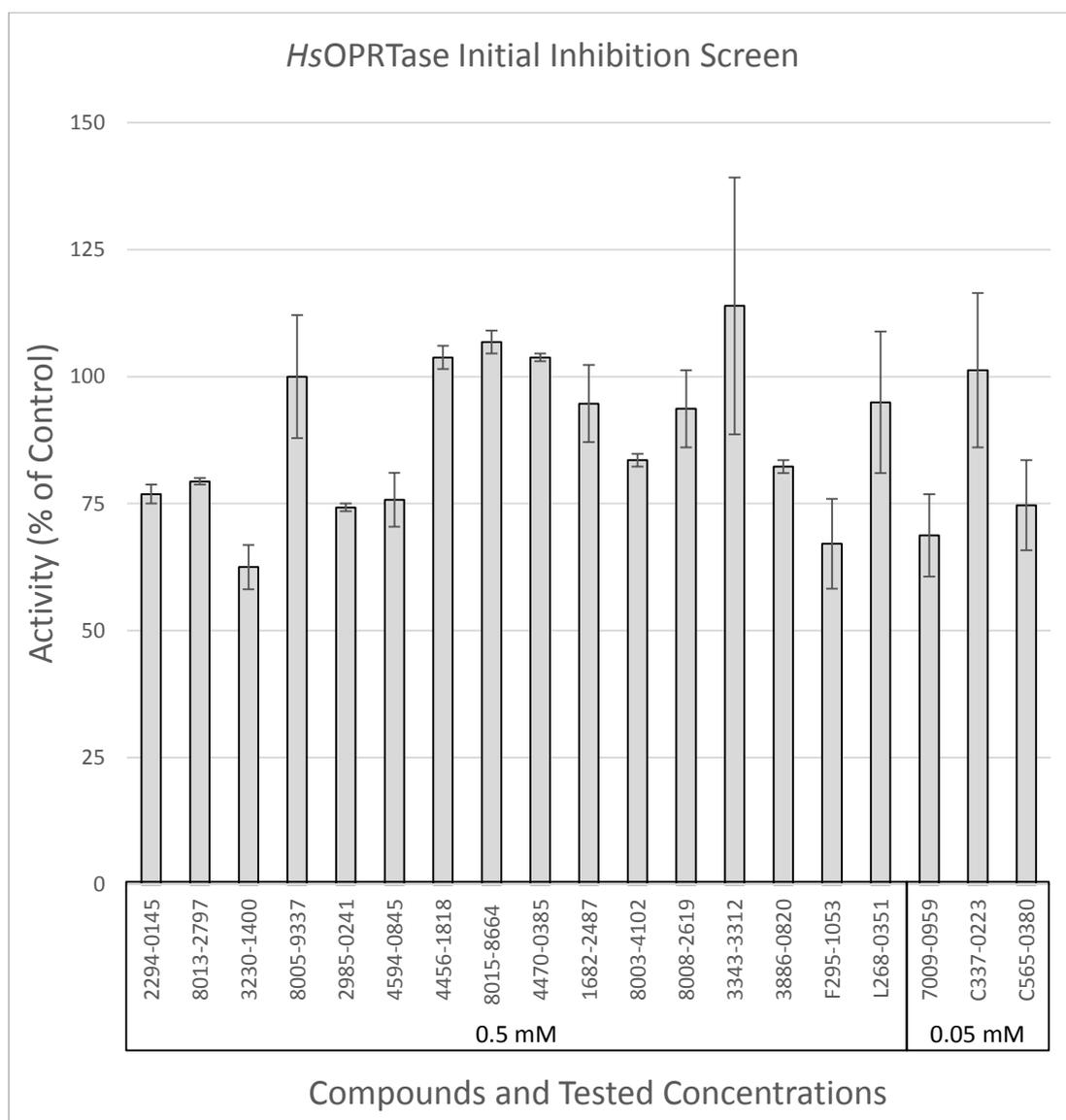


Figure 5.2: Effect of compounds identified by docking on *HsOPRTase* activity.

Inhibition assays were carried out as detailed in Section 2.5.2 at a constant temperature of 25 °C with 1 mM PRPP and 0.25 mM orotate. All rates are means \pm SD (n=2) and are expressed relative to the control rate determined in the absence of the compound. The final concentration of the compounds in the assay is indicated.

5.2.2. Inhibitors of *Pf*OPRTase

Screening of the compounds at 0.5 mM in *Hs*OPRTase revealed that most of the compounds had minimal interference at this concentration. A trial was conducted with several of the compounds to determine if the screening concentration could be increased to ensure that weak inhibitors were being identified. Several of the compounds with low base absorbance at 295 nm were screened at 2 mM (instead of 0.5 mM). The interference was minimal for all of these compounds except for compound F295-1053. This had a base absorbance that was at the upper limit of the spectrophotometer's optimal operating range ($OD_{295\text{ nm}}$ 3–4). When this compound was rescreened at 1 mM the noise in the absorbance readings was greatly improved.

Figure 5.3 shows the initial inhibition screen of the compounds against *Pf*OPRTase.

Compound L268-0351 showed the highest inhibition of approximately 50 % at 0.5 mM. Compounds 8008-2619 and 4470-0385 showed approximately 30 % inhibition also at 0.5 mM. At 2 mM compounds 2985-0241 and 8015-8664 showed approximately 25 % inhibition and compound 4594-0845 approximately 40 % inhibition.

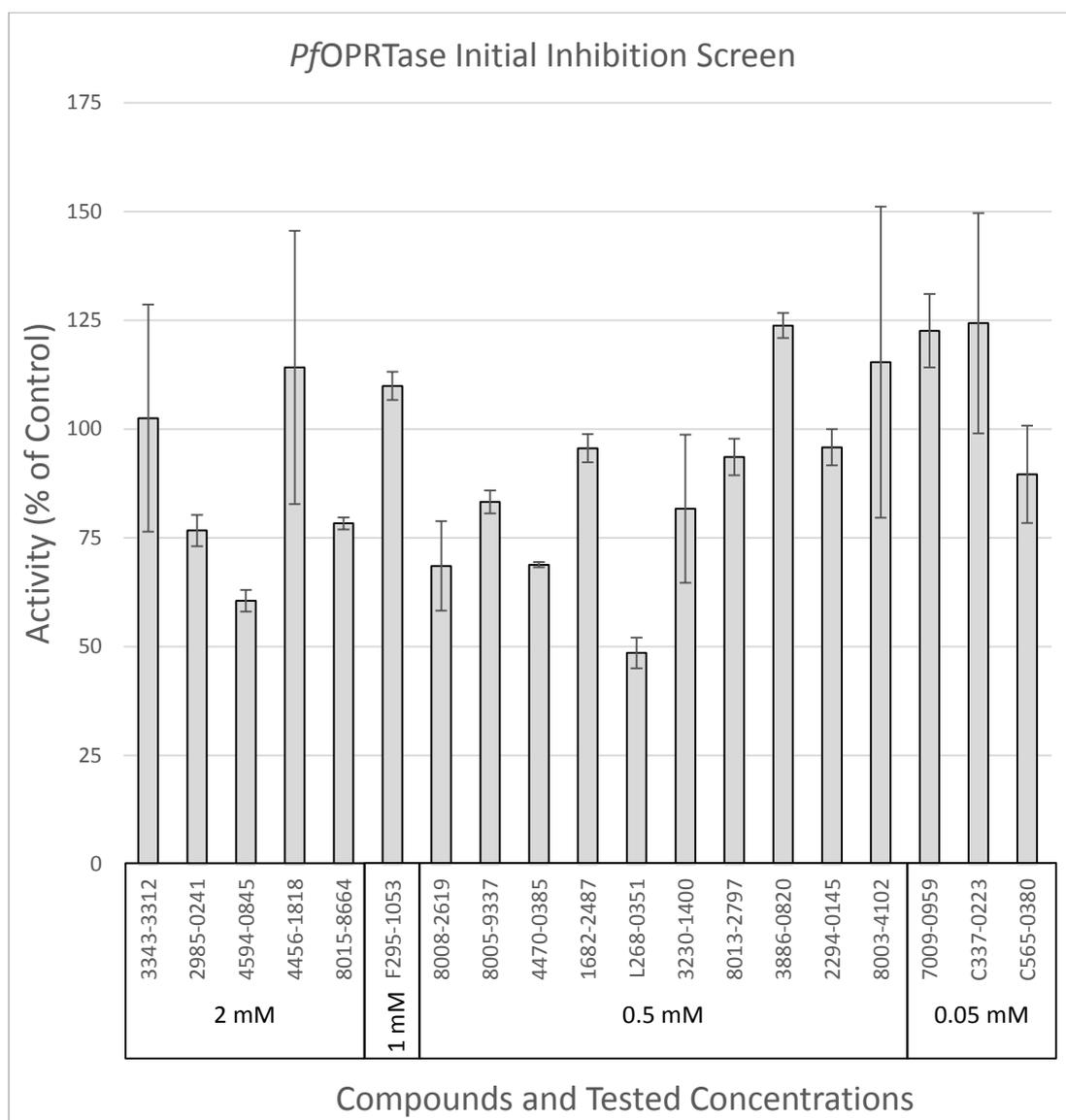


Figure 5.3: Effect of compounds identified by docking on *Pf*OPRTase activity.

Inhibition assays were carried out as detailed in Section 2.5.2 at a constant temperature of 25 °C with 1 mM PRPP and 0.25 mM orotate. All rates are means \pm SD (n=2) and are expressed relative to the control rate determined in the absence of the compound. The final concentration of the compounds in the assay is indicated.

5.2.3. Specific Inhibition of *Pf*OPRTase and *Hs*OPRTase

Figure 5.4 shows the activities of the best inhibitors for both *Hs*OPRTase and *Pf*OPRTase allowing the species specificity to be assessed. As activities are measured at a single concentration of inhibitor and substrate only the conclusions that can be drawn from the data are limited. Some of the compounds were also screened against the human and malarial enzymes at different concentrations which in retrospect creates another variable when assessing specificity. Nevertheless some degree of specific inhibition of *Pf*OPRTase is seen in compounds 268-0351, 8008-2619, and 4470-0385. Compound 4594-0845 showed slightly more inhibition for *Pf*OPRTase but was screened at four times the concentration and is unlikely to be specific. Compounds F294-1053 and 7009-0959 both showed specificity for *Hs*OPRTase; F295-1053 showed specificity for the human OPRTase despite being screened at twice the concentration used for the malarial enzyme.

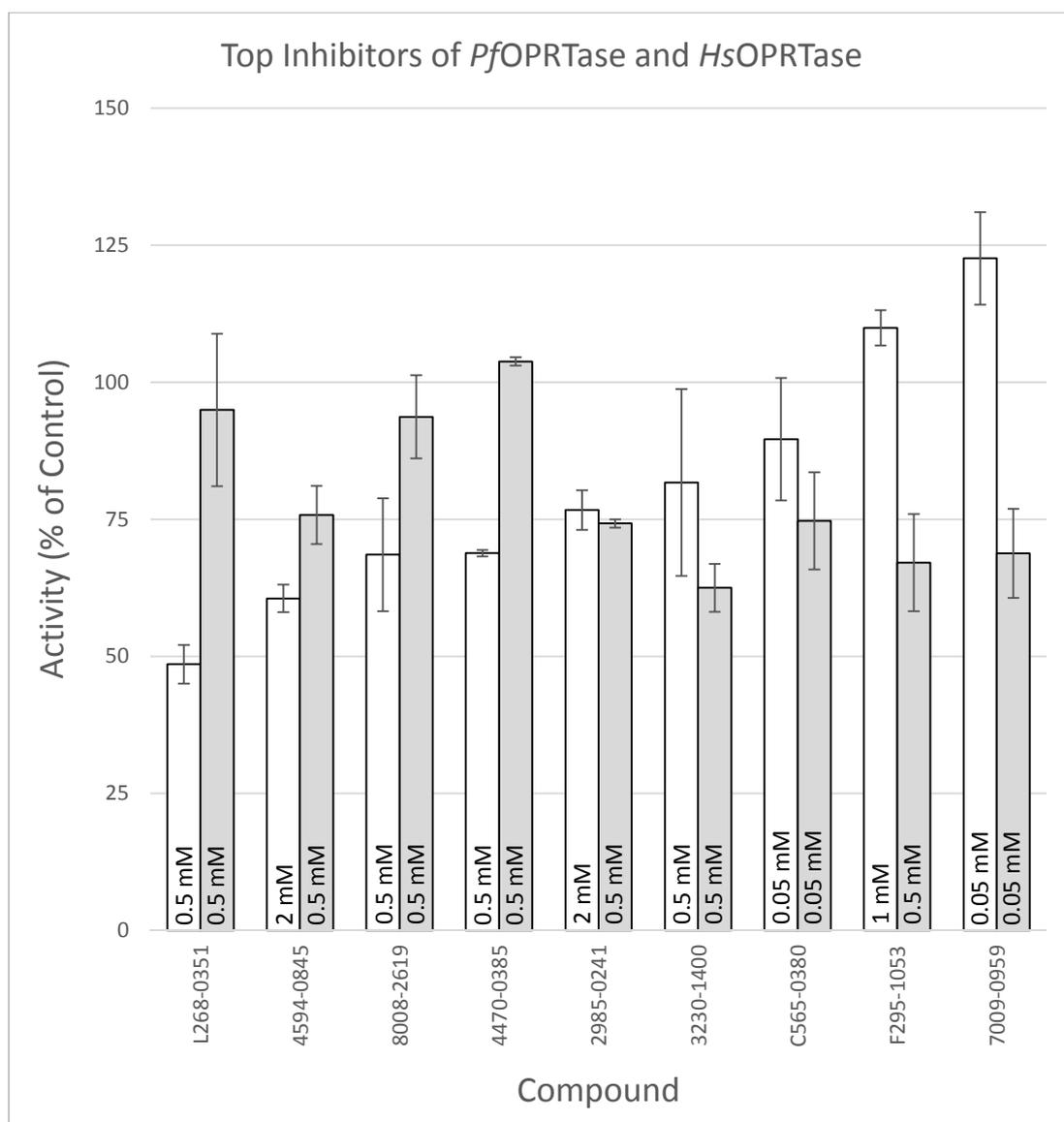


Figure 5.4: Selectivity of the most potent inhibitors of *PfOPRTase* and *HsOPRTase*. Relative inhibition of *PfOPRTase* (□) and *HsOPRTase* (■) by the selected compounds. All rates are means \pm SD (n=2) and are expressed relative to the control rate determined in the absence of the compound for each enzyme. The final concentration of the compounds in the assay is indicated.

5.2.4. Inhibitors of *HsODCase*

Screening of the compounds against *PfOPRTase* revealed that most compounds could be increased in concentration without introducing spectrophotometric interference. For screening of the compounds against *HsODCase* the concentrations of all the compounds were increased to 5 mM with the exceptions of compounds F295-1053, C337-0223, 7009-0959, and C563-0380 (as it was already known that these compounds couldn't be screened at a higher concentration than they were against *PfOPRTase*). Compounds 8003-4102, L2687-0351, 8013, 3230-1400, 1682-2487, 8005-9337, and 4456-1818 all exceeded the spectrophotometer's limit at 5 mM and were rescreened at 1 mM.

Figure 5.5 shows the initial inhibition screen of the compounds against *HsODCase*. At 5 mM compounds 8008-2619 and 4470-0385 showed complete inhibition of *HsODCase*, and compounds 2985-0241 and 2294-0145 showed approximately 35 and 25 % inhibition respectively. Compounds C563-0380, 7009-0959 and C337-0223 were very promising; these compounds showed inhibition of approximately 25–35 % at only 50 μ M. Compounds 4456-1818 and 1682-2487 showed approximately 30 and 25 % inhibition respectively, at 1 mM.

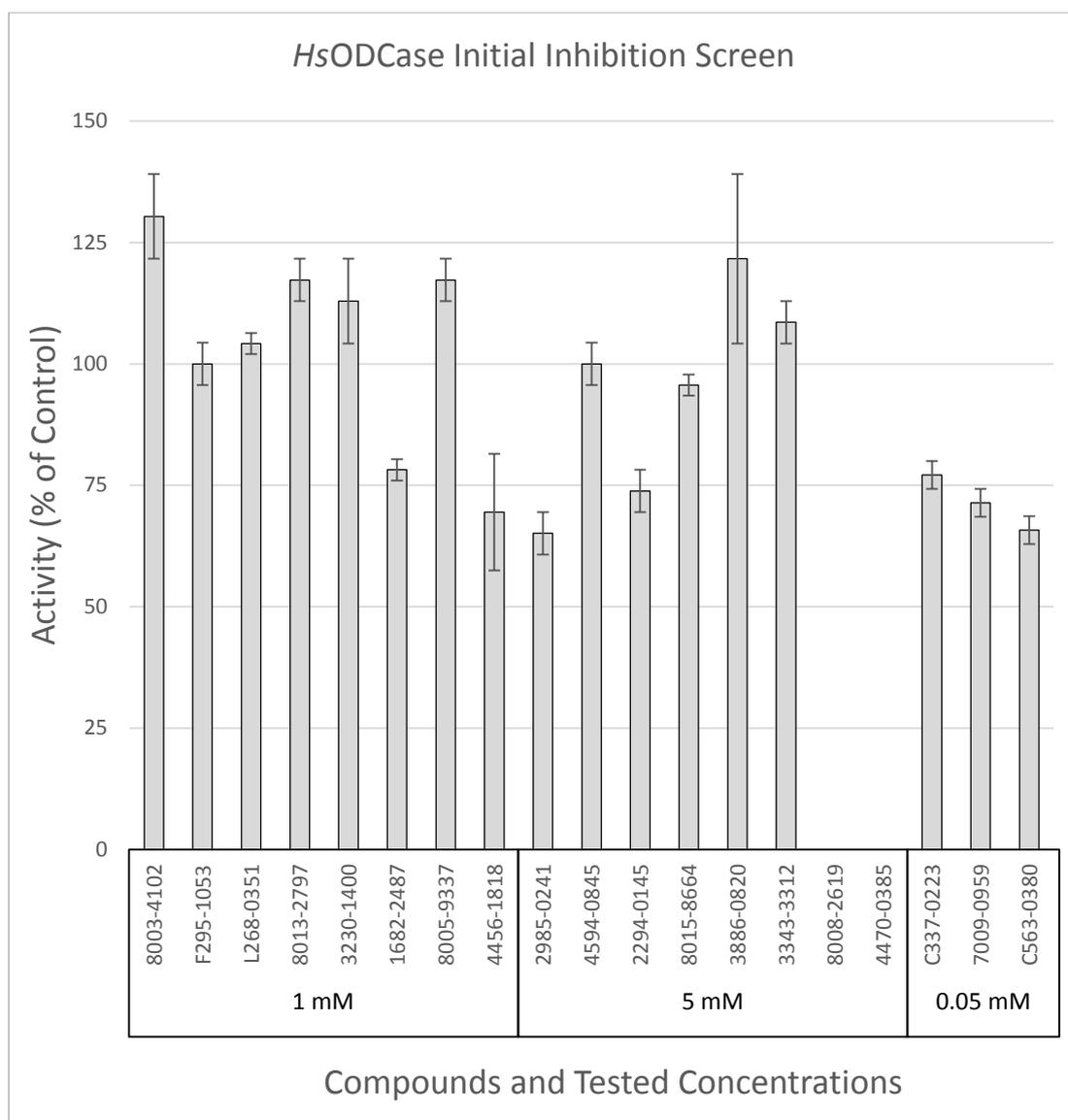


Figure 5.5: Effect of compounds identified by docking on *HsODCase* activity.

Inhibition assays were carried out as detailed in Section 2.5.3 at a constant temperature of 30 °C with 0.39 mM OMP. All rates are means \pm SD (n=2) and are expressed relative to the control rate determined in the absence of the compound. The final concentration of the compounds in the assay is indicated.

5.2.5. Inhibitors of *Pf*ODCase

For screening of the compounds against *Pf*ODCase the same concentrations were used as they were for screening of *Hs*ODCase (Section 5.2.4). Figure 5.6 shows the initial inhibition screen of the compounds against *Pf*ODCase. At 5 mM, compounds 8008-2619 and 4470-0385 showed complete inhibition, and compounds 2985-0241 and 2294-0145 both showed approximately 30 % inhibition. Compound C337-0223 showed approximately 20 % inhibition at 50 μ M. At 1 mM, compounds 1682-2487, F295-1053, and 8005-9337 showed approximately 35 %, 20 % and 15 % inhibition respectively.

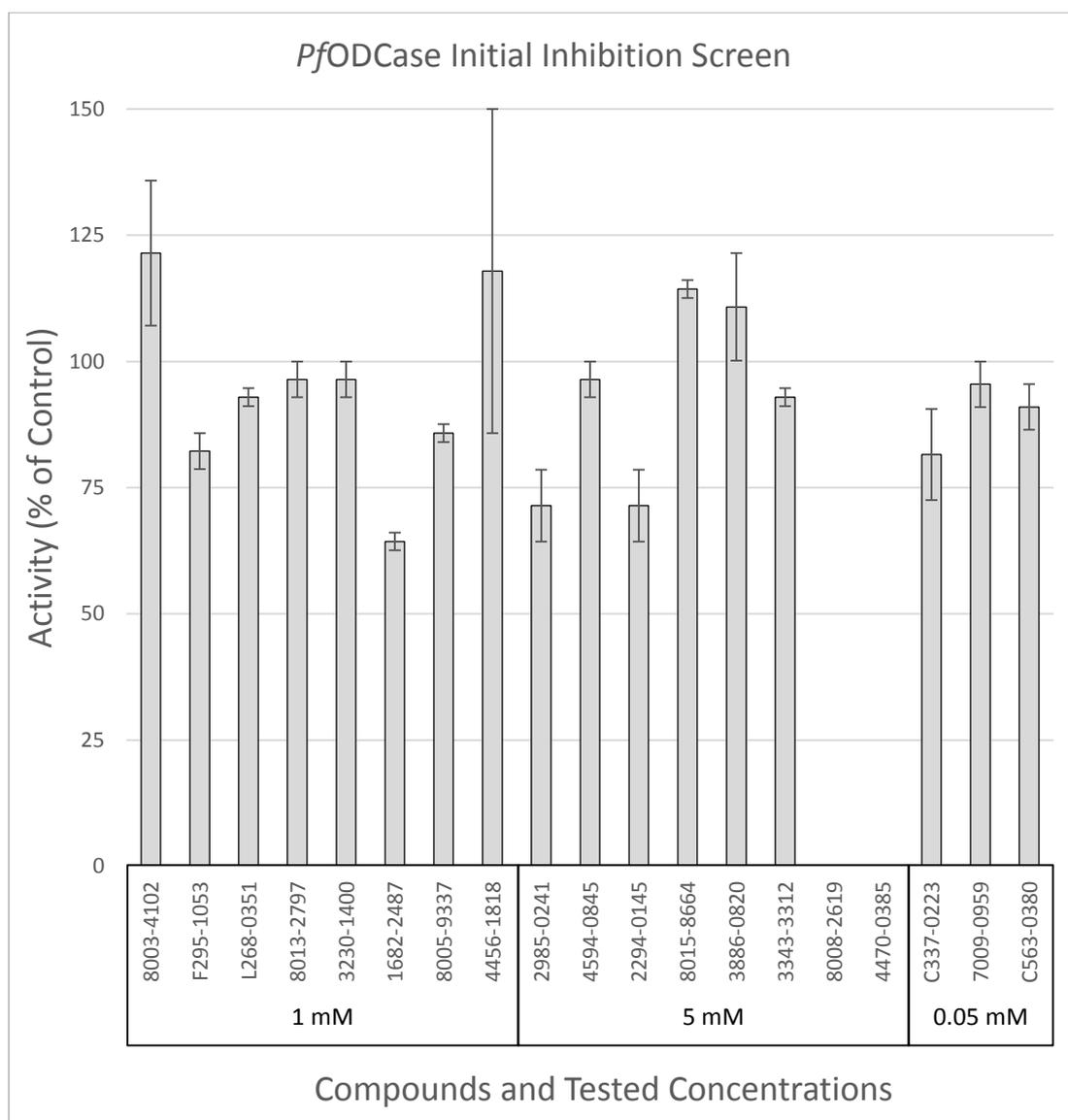


Figure 5.6: Effect of compounds identified by docking on *PfODCase* activity.

Inhibition assays were carried out as detailed in Section 2.5.3 at a constant temperature of 30 °C with 0.39 mM OMP. All rates are means \pm SD (n=2) and are expressed relative to the control rate determined in the absence of the compound. The final concentration of the compounds in the assay is indicated.

5.2.6. Specific Inhibition of *Pf*ODCase and *Hs*ODCase

Figure 5.7 shows the activities of the best inhibitors of *Pf*ODCase and *Hs*ODCase, except for compounds 8008-2619 and 4470-0385. The compounds were screened at the same concentration for each enzyme making identification of specificity for the human or malarial enzyme much easier, relative to the initial screen for the OPRTases. Again, as activities are measured at a single concentration of inhibitor and substrate only, the conclusions that can be drawn from the data are limited. Compounds 1682-2487, 2985-0241, 2294-0145 and C337-0223 showed no significant specificity for one enzyme over the other. Compounds F295-1053 and 8005-9337 both showed some degree of specificity for the human ODCase. Compounds C563-0380, 7009-0959 and 4456-1818 showed promising specificity for the *P. falciparum* ODCase.

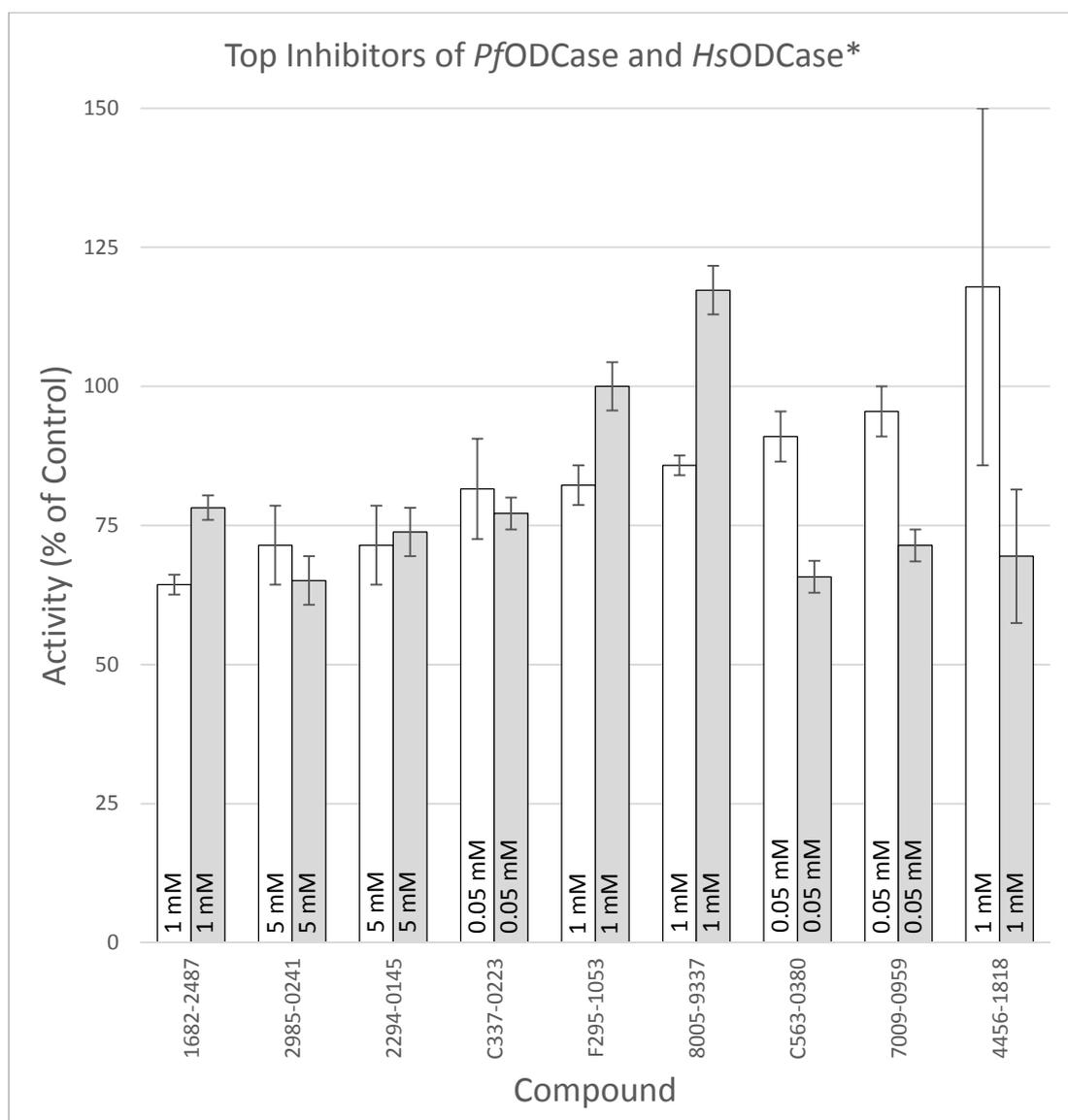


Figure 5.7: Selectivity of the most potent inhibitors of *PfOPRTase* and *HsOPRTase.** Relative inhibition of *PfODCase* (□) and *HsODCase* (■) by the selected compounds. All rates are means \pm SD (n=2) and are expressed relative to the control rate determined in the absence of the compound for each enzyme. The final concentration of the compounds in the assay is indicated.

*Excluding compounds 4470-0385 and 8008-2619

5.3. Results: Hit Expansion and Beginnings of a Structure Activity

Relationship

Screening of the compounds against the ODCases revealed several promising inhibitors and these compounds were planned to be characterised further to determine the inhibition kinetics. Prior to this however a small scale H2L strategy, hit expansion, was conducted. The principle behind this strategy is that if a compound binds to a target enzyme, it is likely that analogues of that compound will also bind. The structural differences between analogues and the resulting differences in binding affinities are used to build a Structure-Activity Relationship (SAR). A very fast and effective approach to identifying analogues of a compound is through the use of a 2D fingerprint screen.

5.3.1. 2D Fingerprint Rescreen

As outlined in Section 1.5.2.4, a 2D fingerprint search will take a compound library and return the Tanimoto similarity of each compound to a single query molecule. The 2D molecular fingerprint screens were performed with MOLPRINT 2D v1.2 (Bender et al., 2004). The ChemDiv subset of the UCSF ZINC database (Irwin and Shoichet, 2005) was used for the screening. This was the same library that was screened by docking originally in Chapter 4. Compounds 4456-1818, 4470-0385, and C337-0223 (Figure 5.1) were the query molecules.

Compound C337-0223 was chosen as it exhibited clear inhibition against *Hs*ODCase (and potentially against *Pf*ODCase as well) at a relatively low concentration. Compound 4470-0385 was chosen for its complete inhibition of both ODCase enzymes in the initial screen (Figures 5.5 and 5.6). Compound 8008-2619 was not chosen as it was deemed to be far too flexible (12 rotatable bonds) to be a promising scaffold. Compound 4470-0385 is also a fairly flexible compound (7 rotatable bonds) but far less so than compound 8008-2619. The best way forward with highly flexible compounds is generally to confirm the conformation it adopts in the active site and work towards

modifying the compound to reduce the number of rotatable bonds and increase its rigidity while maintaining its bioactive conformation (Luque, 2010).

Compound 4456-1818 was an unlikely candidate chosen for a number of reasons. Firstly it was an inhibitor (albeit a very weak inhibitor). More importantly it was somewhat of a consensus structure for the compounds with confirmed activity against the ODCases. It had the two separate ring groups that were present in a large number of compounds that showed inhibition, including 4470-0385 and C337-0223. It also contained the sulfone-amide group that was present in a large number of confirmed compounds—a feature that 4470-0385 and C337-0223 lacked. Finally it had carboxyl groups. The electronegative groups were seen in all of the confirmed compounds; their role would likely be complementing the PRPP binding site or the carboxyl-group pocket on the orotate ring binding site. Hence while 4456-1818 was a weak inhibitor it was still a promising scaffold.

The molecular fingerprints for the queries and the ChemDiv library were calculated using the utility 'mol22aefp' using default parameters. The fingerprint library was then compared to each query using the 'tanimoto.pl' script as per MOLPRINT 2D's instructions. The output consisted of a list of the input compounds with associated Tanimoto similarity to the query structure.

A large number of compounds were returned with varying Tanimoto similarity scores for each query molecule. Three compounds were selected from the screen against compound 4456-1818, and one compound selected for each of compounds 4470-0385 and C337-0223. These selected compounds are shown in Table 5.1. The selection of these new compounds for biochemical assays is discussed in Section 5.5.2.

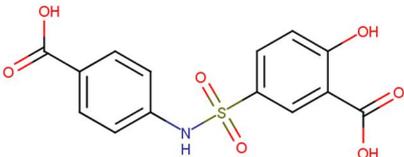
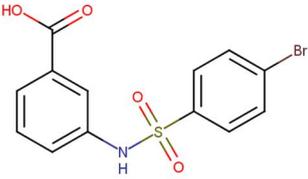
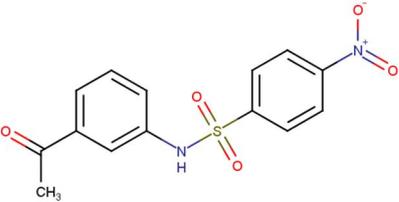
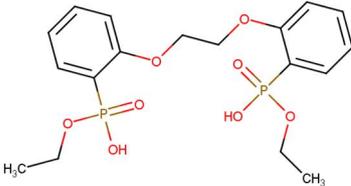
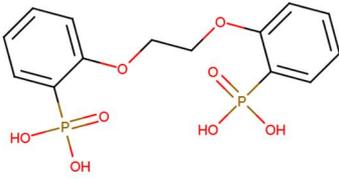
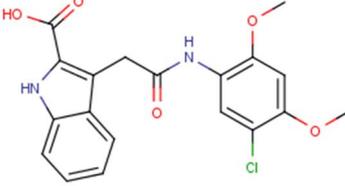
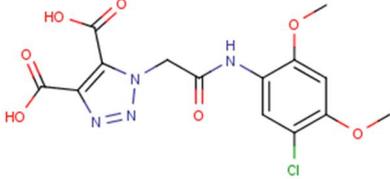
5.3.2. Inhibition Screen of 2D Fingerprint Hits against *PfODCase* and *HsODCase*

Absorbances at 295 nm were measured for all the compounds at 0.5 mM to first determine whether they would absorb strongly and interfere with the accuracy of the spectrophotometer readings. The concentrations of compounds C197-0379 and 4049-0191 had to be decreased to 0.2 and 0.4 mM respectively to ensure the base absorbance at 295 nm did not exceed 3 absorbance units.

The compounds were screened following the protocol outlined in Section 2.5.3. The percent activity in the presence of various concentrations of the compounds, relative to control in the absence of compound is shown in Figure 5.8 for *PfODCase* and *HsODCase*.

Of the hits from the 4456-1818 2D fingerprint screen only K783-0416 showed any clear activity with weak inhibition against *PfODCase* of 16 % inhibition at 0.5 mM (Figure 5.8). Compound 4049-0191 showed a large apparent increase in activity of approximately 45 % for *HsODCase*. Both 4470-0386 and C197-0379 showed clear inhibition against *HsODCase* and *PfODCase*.

Table 5.1: Inhibitor analogues selected for inhibition assays from 2D Fingerprint screening. The ChemDiv catalogue number and Tanimoto similarity is shown for each hit, juxtaposed with the query compound.

Hit Compound Image ChemDiv cat# Tanimoto similarity	Query Compound Image ChemDiv cat#
 <p data-bbox="564 658 707 723">1494-0562 0.611</p>	 <p data-bbox="1123 981 1265 1010">4456-1818</p>
 <p data-bbox="564 949 707 1014">K783-0416 0.571</p>	
 <p data-bbox="564 1267 707 1332">4049-0191 0.55</p>	
 <p data-bbox="564 1570 707 1635">4047-0386 0.375</p>	 <p data-bbox="1123 1585 1265 1615">4047-0385</p>
 <p data-bbox="564 1861 707 1926">C197-0379 0.379</p>	 <p data-bbox="1123 1877 1265 1906">C337-0223</p>

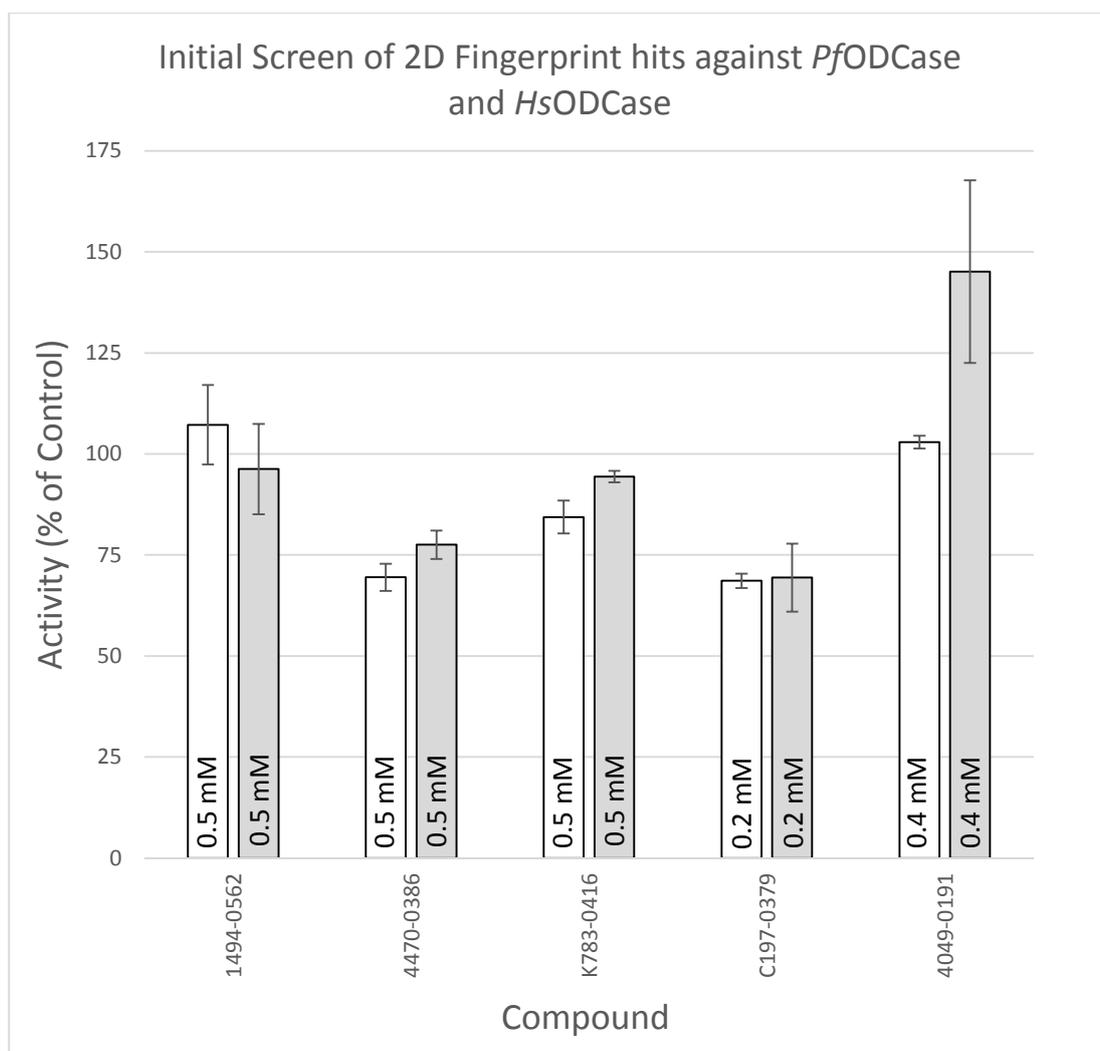


Figure 5.8: Effect of compounds identified by 2D Fingerprint screening on *PfODCase* and *HsODCase*. Relative inhibition of *PfODCase* (□) and *HsODCase* (■) by the compounds. Inhibition assays were carried out as detailed in Section 2.5.3 at a constant temperature of 30 °C. All rates are means \pm SD (n=2) and are expressed relative to the control rate determined in the absence of the compound for each enzyme. The final concentration of the compounds in the assay is indicated.

5.4. Results: *PfODCase* and *HsODCase* Michaelis-Menten Kinetics

Michaelis-Menten kinetics was originally attempted spectrophotometrically for *PfODCase* and *HsODCase* using initial velocities at various concentrations of substrate. Assays were performed at concentrations of 50–390 μM of substrate, however no significant difference was observed for the initial velocities (data not shown). As mentioned before, due to the low K_m for the substrate relative to the spectrophotometric detection limit of the substrate, accurate determination of the kinetic parameters would require the published method using ^{14}C radio-labelled substrate (Yablonski et al., 1996). Reaction progress kinetic analysis (Blackmond, 2005), in which the rate of reaction is measured as the substrate concentration is reduced as the reaction progresses, was instead used here to give K_m^{app} values. Inhibition kinetics were performed on *PfODCase* and *HsODCase* for compounds 8008-2619, 4470-0385, C337-0223, 7009-0959, C563-0380, 4470-0386, and C197-0379 as detailed in Section 2.5.4.

5.4.1. Inhibition Kinetics of Compounds against *PfODCase* and *HsODCase*

The inhibition constants for each of the promising compounds, identified by the initial biochemical assays, were determined by fitting the activity data obtained at a range of substrate and inhibitor concentrations to the Michaelis-Menten equations for competitive, uncompetitive or mixed inhibition. The equations used by GraphPad Prism as well as graphs illustrating idealised data for each model are shown in Appendices 8.6–8.9.

For *PfODCase*, compounds 8008-2619 (Figure 5.9), 4470-0385 (Figure 5.10) and 4470-0386 (Figure 5.11) were found to best fit the equation for uncompetitive inhibition with αK_i^{app} values of 380.8 ± 49.5 , 249.5 ± 29.2 , and 178.1 ± 21.0 μM respectively. Compound C197-0379 (Figure 5.12) was found to best fit the equation for mixed inhibition with a K_i^{app} of 49.3 ± 14.5 μM and α value of 2.81 ± 1.85 . For *HsODCase*, compounds 4470-0385 (Figure 5.13), 4470-0386 (Figure 5.14), and

C197-0379 (Figure 5.15) were found to best fit the equation for uncompetitive inhibition with αK_i^{app} values of 1956 ± 489 , 543.7 ± 69.2 and $44.21 \pm 5.15 \mu\text{M}$ respectively. Compounds 7009-0959 (Figure 5.16) and C563-0380 (Figure 5.17) were found to best fit the equation for mixed inhibition with K_i^{app} values of 172 ± 42.1 and $86.7 \pm 23.3 \mu\text{M}$ respectively and α values of 1.25 ± 0.42 and 6.34 ± 3.70 respectively. Compound C337-0223 (Figure 5.18) was found to best fit the model for competitive inhibition with a K_i^{app} value of $118 \pm 22.6 \mu\text{M}$.

Compound 8008-2619 did not yield enough inhibition against *HsODCase* for kinetic modelling at 500 and 1500 μM (data not shown). Compounds 7009-0959 and C563-0380 were not characterised for *PfODCase* as they did not show significant inhibition in the initial screen (Section 5.6.2).

Compound C337-0223 had shown possible inhibition in the initial screen (Section 5.6.2) against *PfODCase* but showed very unusual data (did not fit any inhibition model) for the kinetic characterisation (Figure 5.19 and Table 5.2). The K_m^{app} for OMP increased with increasing concentration of C337-0223. The V_{max} was higher in the presence of C337-0223 (but slightly lower at 250 μM than it was at 150 μM). Control assays on C337-0223 without OMP and another without the enzyme confirmed that it was not an alternative substrate, nor was it reacting with OMP and the enzyme in the initial inhibition screen for this compound showed the enzyme to be stable for the duration of the assay (data not shown).

The inhibition data is summarised in Table 5.3. Table 5.4 shows the chemical properties of the compounds with respect to the rules of drug and lead likeness. Three of the compounds (8008-2619, 4470-0386 and C337-0223) showed Lipinski or bioavailability violations. These would need to be addressed during H2L or lead development if they were chosen for further development. The other four compounds (C197-0379, 4470-0385, C563-0380 and 7009-0959) show promising drug-like properties with no Lipinski or bioavailability violations.

Uncompetitive modes of inhibition are highly unexpected as the *in silico* screening was intended to identify competitive inhibitors of the enzymes. The data for two of the uncompetitive inhibitors were converted to double reciprocal Lineweaver-Burk plots to confirm visually that an uncompetitive inhibition model could be fit (Lineweaver and Burk, 1934) (data not shown), however the kinetic data reported here were calculated using GraphPad Prism's more accurate algorithms. Compounds 7009-0959 and C563-0380 showed mixed inhibition against *HsODCase*, as did C197-0379 against *PfODCase*. Compound 7009-0959 had an α value of 1.25—close to 1—indicating it is very near a non-competitive mode of inhibition. While mixed inhibition is not that unusual it is still expected that it would lean closer to competitive inhibition rather than being an almost non-competitive model.

Compounds 4470-0385 and 4470-0386 showed 8-fold and 3-fold higher affinity respectively for the *PfODCase* over the *HsODCase*. Compound 8008-2619 showed a higher affinity for *PfODCase* over *HsODCase* (as mentioned the same concentrations of inhibitor were used for both, but only *PfODCase* was significantly inhibited). While the inhibition constants for these compounds weren't the most potent of the identified inhibitors they do show some level of specificity for the malarial enzyme. Compounds 7009-0959 and C563-0380 showed significant inhibition against *HsODCase* but hadn't inhibited *PfODCase* in the initial screen (Section 5.6.2).

The kinetic assays proved a little problematic in fitting of some of the kinetic curves. This occurred mostly where levels of inhibition were generally low, such as for compound 4470-0385 against *HsODCase* (Figure 5.4), or where there was a large degree of variation between replicates and/or inhibitor concentrations (usually due to spectrophotometric interference) such as with compound C337-0223 (Figures 5.5 and 5.6). This resulted in some low global R^2 values.

In many cases the inhibition constant had a large standard deviation (> 10 % of the value), but was generally not significant enough to hinder comparing inhibition

constants. This occurred where the inhibitor concentrations yielded similar levels of inhibition leading to a larger degree of extrapolation from the kinetic model, such as the case with compound C197-0379 (Figures 5.12 and 5.16).

PfODCase Inhibition by 8008-2619

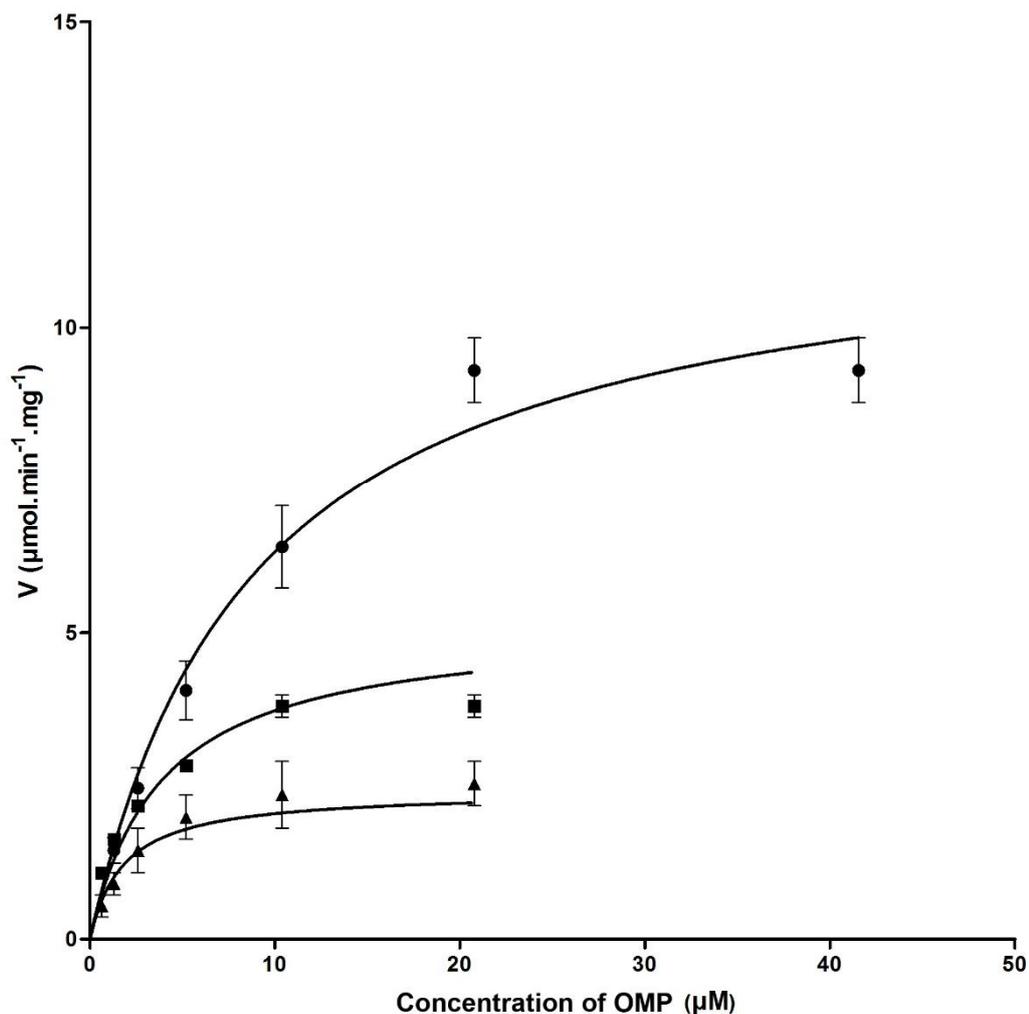


Figure 5.9: Uncompetitive inhibition of *PfODCase* by compound 8008-2619.

PfODCase activity was measured as detailed in Section 2.5.4, at a constant temperature of 30 °C, at three concentrations of inhibitor (0 μM (●), 500 μM (■), and 1500 μM (▲)). Rates are means \pm SEM ($n = 3$). The data were found to best fit the Michaelis-Menten equation for uncompetitive inhibition (Appendix 8.8) with a global R^2 fit of 0.95. The estimated αK_i^{app} was $380.8 \pm 49.5 \mu\text{M}$. Kinetic data is also summarised in Table 5.3.

PfODCase Inhibition by 4470-0385

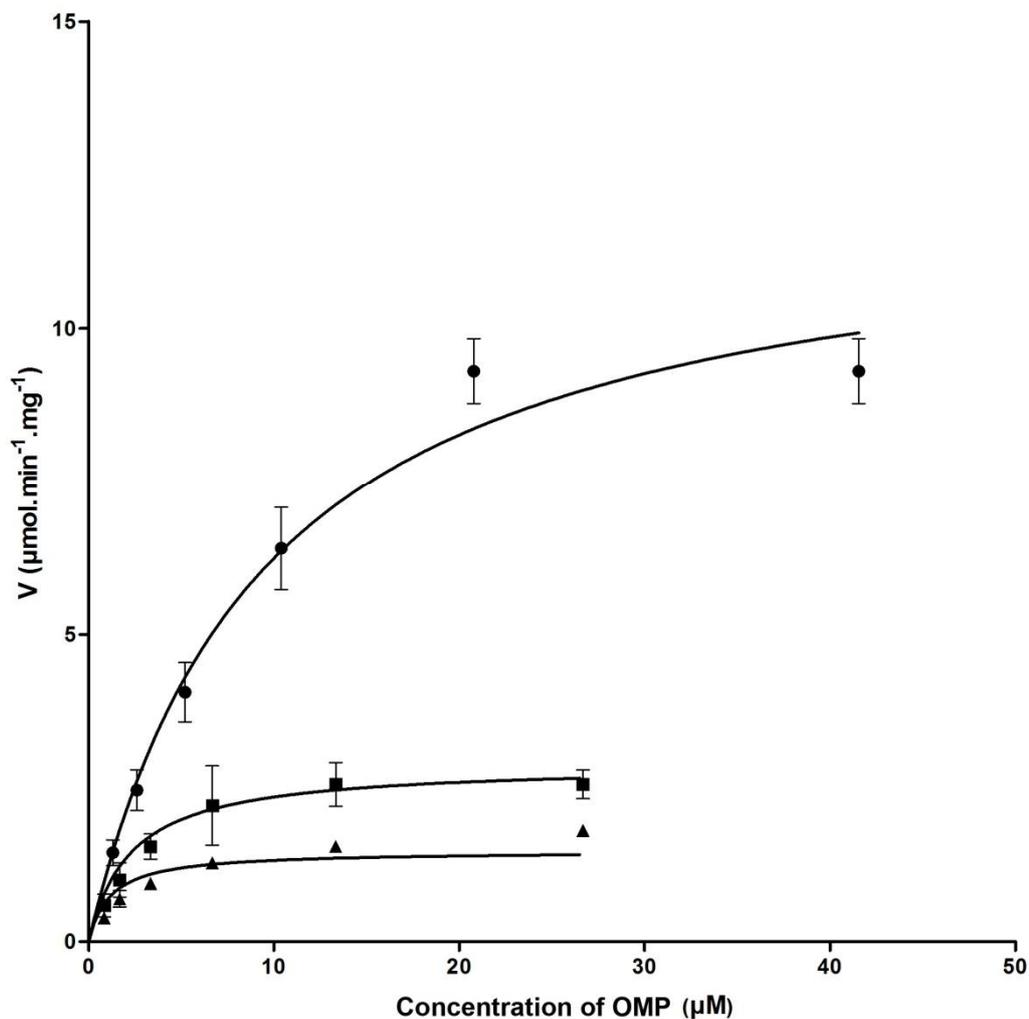


Figure 5.10: Uncompetitive inhibition of *PfODCase* by compound 4470-0385.

PfODCase activity was measured as detailed in Section 2.5.4, at a constant temperature of 30 °C, at three concentrations of inhibitor (0 μM (●), 800 μM (■), and 1800 μM (▲)). Rates are means ± SEM (n = 3). The data were found to best fit the Michaelis-Menten equation for uncompetitive inhibition (Appendix 8.8) with a global R² fit of 0.96. The estimated αK_i^{app} was 247.5 ± 29.2 μM. Kinetic data is also summarised in Table 5.3.

*Pf*ODCase Inhibition by 4470-0386

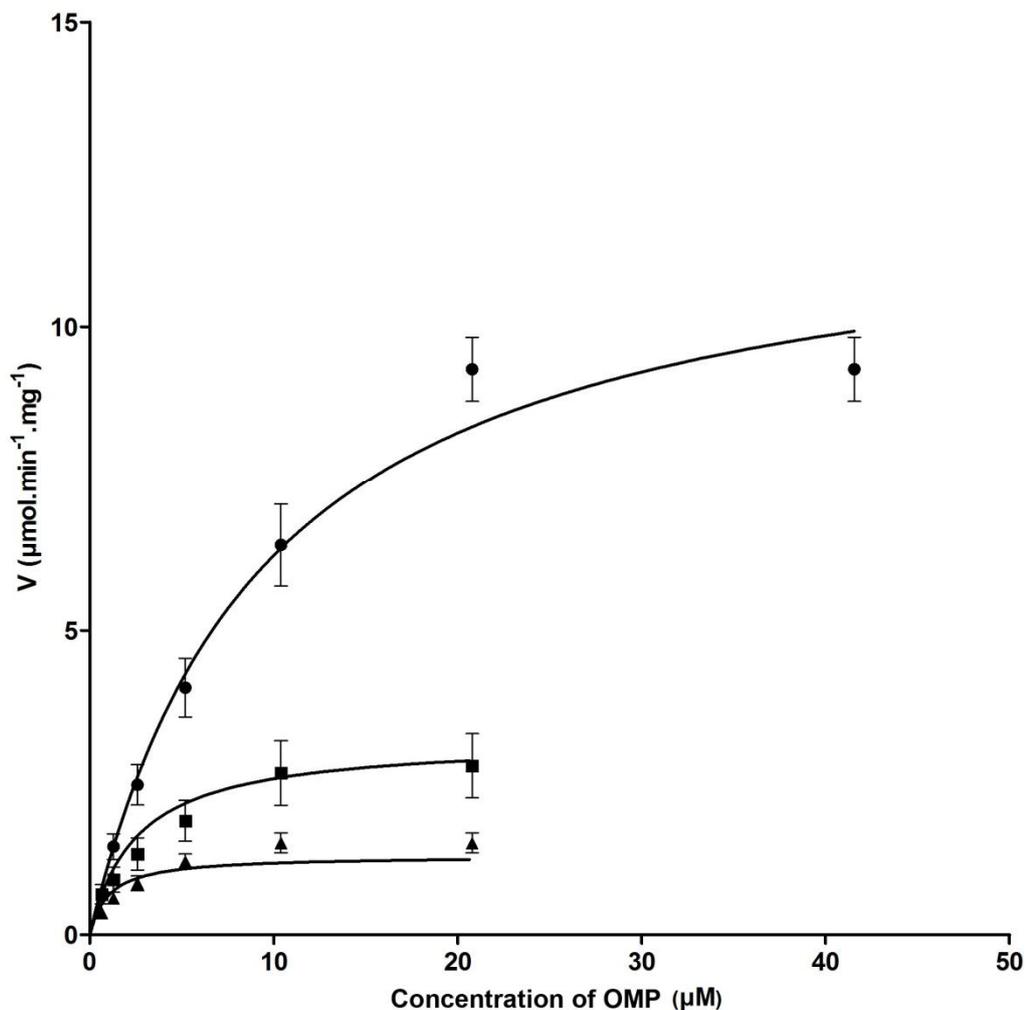


Figure 5.11: Uncompetitive inhibition of *Pf*ODCase by compound 4470-0386.

*Pf*ODCase activity was measured as detailed in Section 2.5.4, at a constant temperature of 30 °C, at three concentrations of inhibitor (0 μM (●), 500 μM (■), and 1500 μM (▲)). Rates are means ± SEM (n = 3). The data were found to best fit the Michaelis-Menten equation for uncompetitive inhibition (Appendix 8.8) with a global R² fit of 0.95. The estimated αK_i^{app} was 178.1 ± 21.0 μM. Kinetic data is also summarised in Table 5.3.

*Pf*ODCase Inhibition by C197-0379

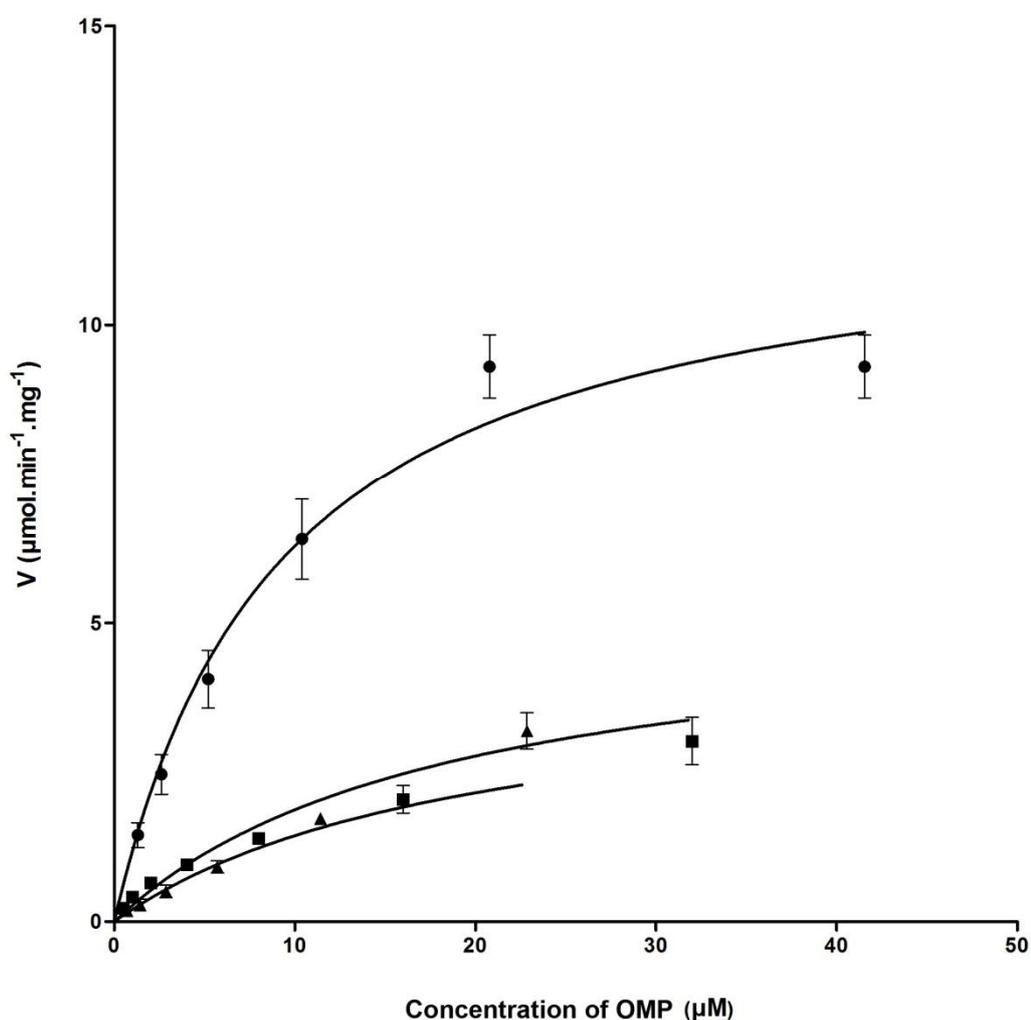


Figure 5.12: Mixed inhibition of *Pf*ODCase by compound C197-0379. *Pf*ODCase activity was measured as detailed in Section 2.5.4, at a constant temperature of 30 °C, at three concentrations of inhibitor (0 μM (●), 175 μM (■), and 250 μM (▲)). Rates are means \pm SEM ($n = 3$). The data were found to best fit the Michaelis-Menten equation for mixed inhibition (Appendix 8.9) with a global R^2 fit of 0.96. The estimated K_i^{app} was $49.3 \pm 14.5 \mu\text{M}$ and the estimated α value was 2.81 ± 1.85 . Kinetic data is also summarised in Table 5.3.

HsODCase Inhibition by 4470-0385

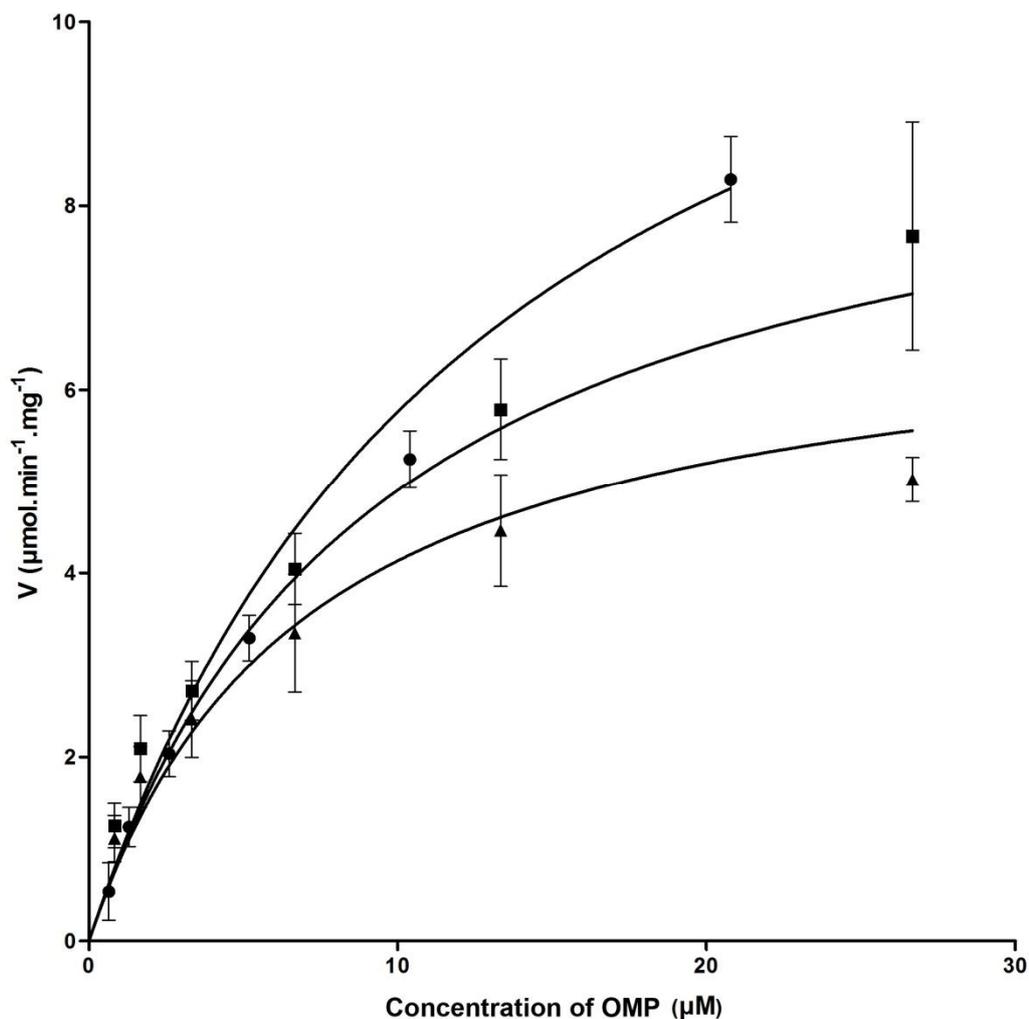


Figure 5.13: Uncompetitive inhibition of HsODCase by compound 4470-0385.

HsODCase activity was measured as detailed in Section 2.5.4, at a constant temperature of 30 °C, at three concentrations of inhibitor (0 μM (●), 800 μM (■), and 1800 μM (▲)). Rates are means \pm SEM (n = 3). The data were found to best fit the Michaelis-Menten equation for uncompetitive inhibition (Appendix 8.8) with a global R^2 fit of 0.89. The estimated αK_i^{app} was $1956 \pm 489 \mu\text{M}$. Kinetic data is also summarised in Table 5.3.

HsODCase Inhibition by 4470-0386

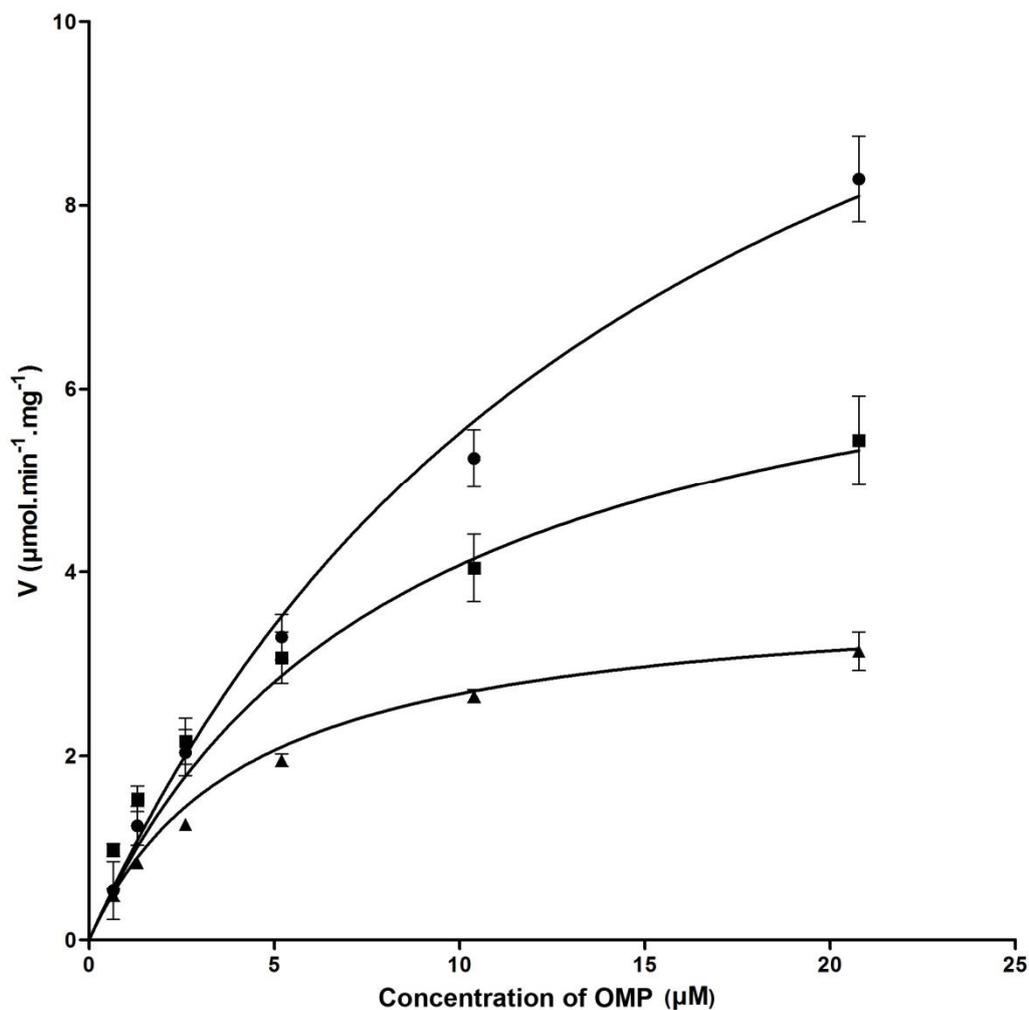


Figure 5.14: Uncompetitive inhibition of HsODCase by compound 4470-0386.

HsODCase activity was measured as detailed in Section 2.5.4, at a constant temperature of 30 °C, at three concentrations of inhibitor (0 μM (●), 500 μM (■), and 1500 μM (▲)). Rates are means \pm SEM (n = 3). The data were found to best fit the Michaelis-Menten equation for uncompetitive inhibition (Appendix 8.8) with a global R^2 fit of 0.95. The estimated αK_i^{app} was $543.7 \pm 69.2 \mu\text{M}$. Kinetic data is also summarised in Table 5.3.

HsODCase Inhibition by C197-0379

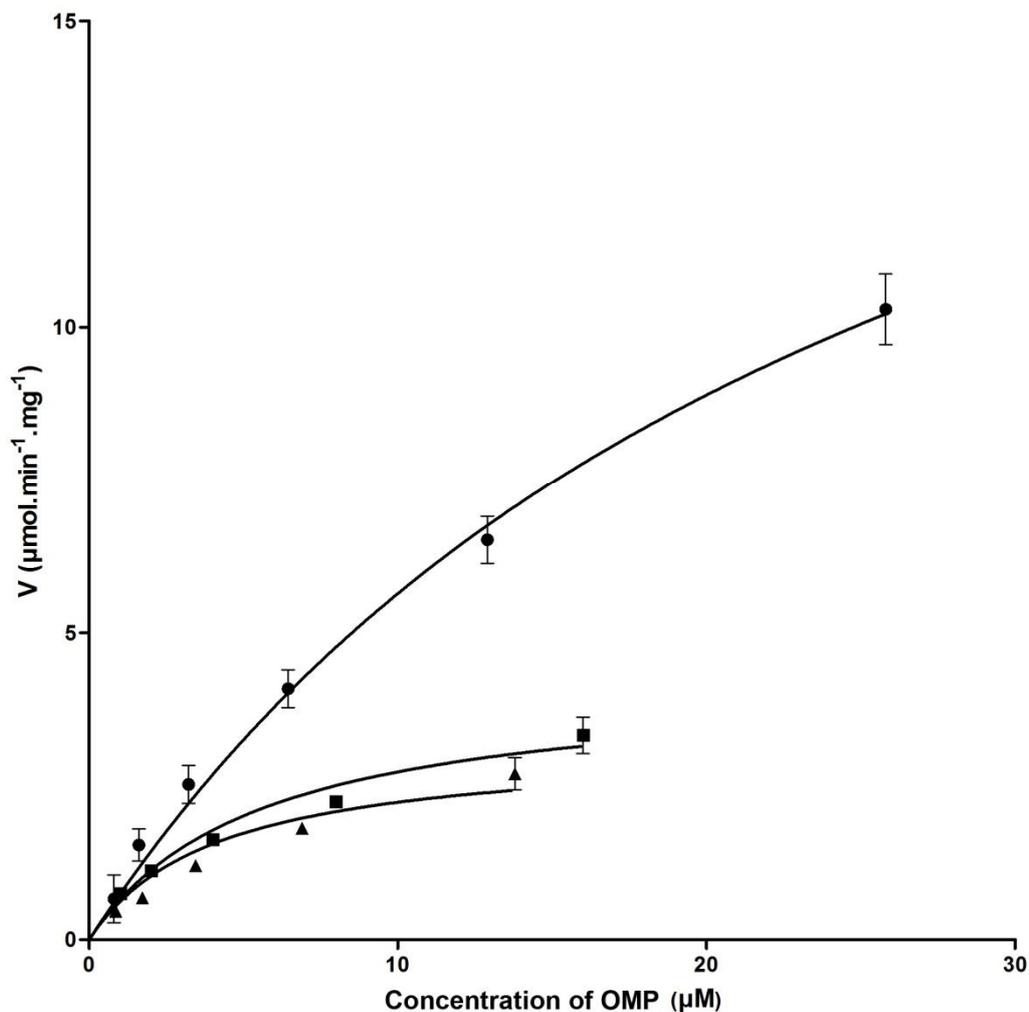


Figure 5.15: Uncompetitive inhibition of *HsODCase* by compound C197-0379.

PfODCase activity was measured as detailed in Section 2.5.4, at a constant temperature of 30 °C, at three concentrations of inhibitor (0 μM (●), 175 μM (■), and 250 μM (▲)). Rates are means \pm SEM ($n = 3$). The data were found to best fit the Michaelis-Menten equation for uncompetitive inhibition (Appendix 8.8) with a global R^2 fit of 0.97. The estimated αK_i^{app} was $44.21 \pm 5.15 \mu\text{M}$. Kinetic data is also summarised in Table 5.3.

HsODCase Inhibition by 7009-0959

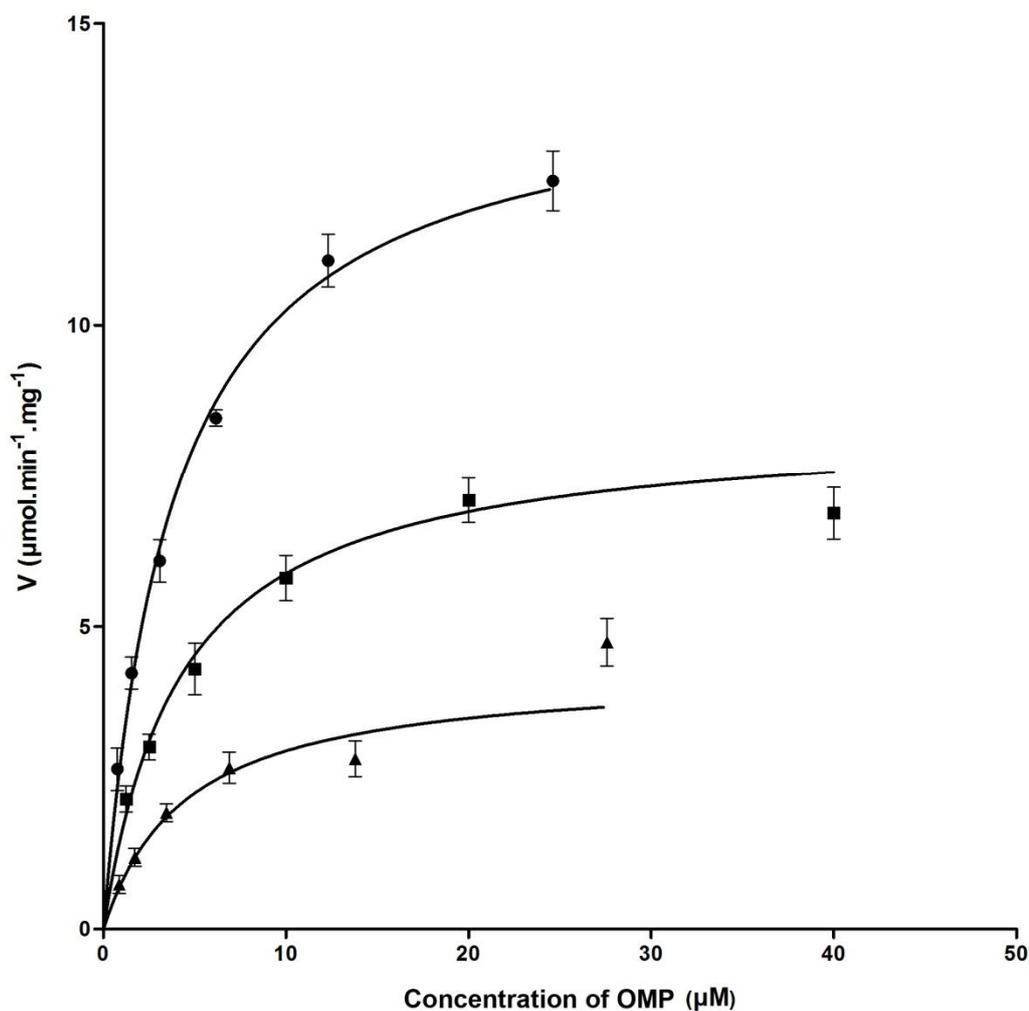


Figure 5.16: Mixed inhibition of *HsODCase* by compound 7009-0959. *HsODCase* activity was measured as detailed in Section 2.5.4, at a constant temperature of 30 °C, at three concentrations of inhibitor (0 μM (●), 150 μM (■), and 500 μM (▲)). Rates are means ± SEM (n = 3). The data were found to best fit the Michaelis-Menten equation for mixed inhibition (Appendix 8.9) with a global R² fit of 0.97. The estimated K_i^{app} was 172 ± 42.1 μM and the estimated α value was 1.25 ± 0.42 . Kinetic data is also summarised in Table 5.3.

HsODCase Inhibition by C563-0380

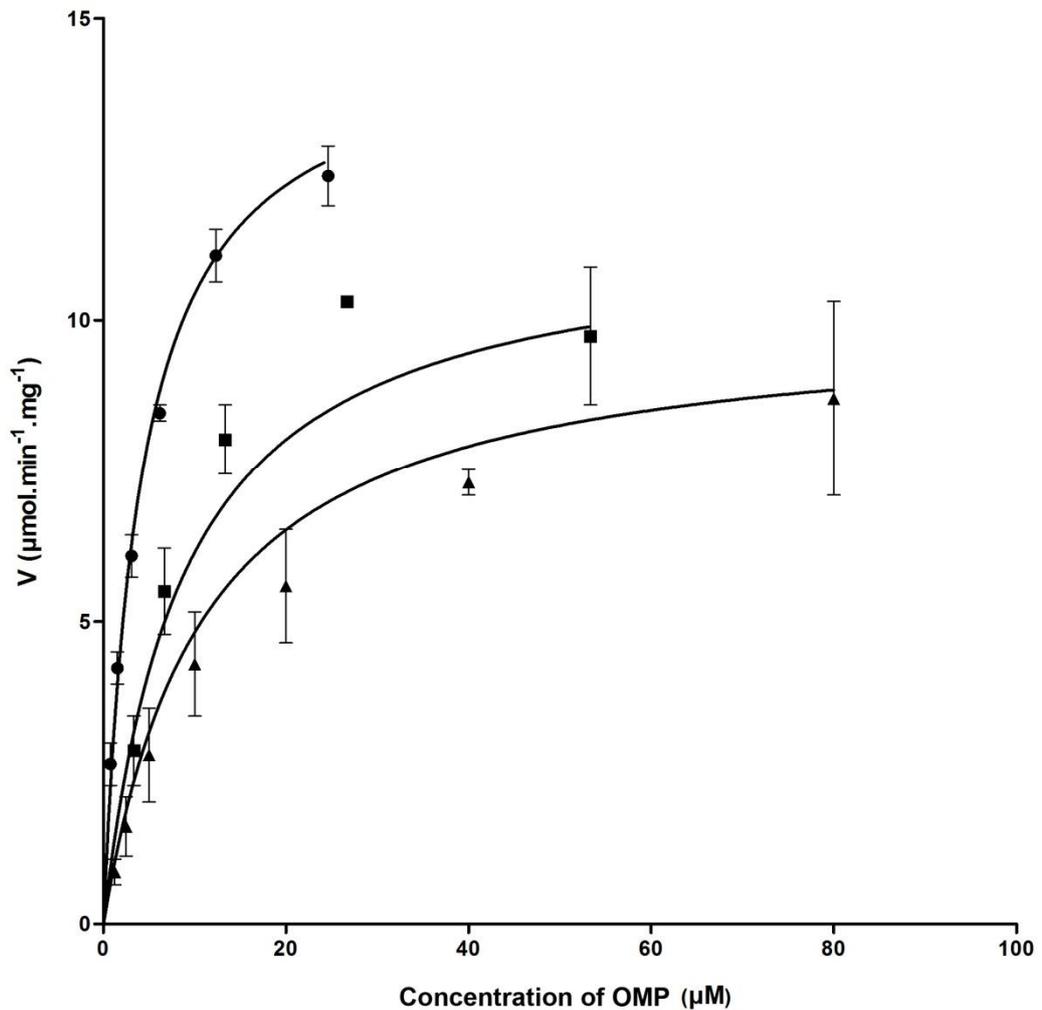


Figure 5.17: Mixed inhibition of *HsODCase* by compound C563-0380. *HsODCase* activity was measured as detailed in Section 2.5.4, at a constant temperature of 30 °C, at three concentrations of inhibitor (0 μM (●), 150 μM (■), and 250 μM (▲)). Rates are means \pm SEM (n = 3). The data were found to best fit the Michaelis-Menten equation for mixed inhibition (Appendix 8.9) with a global R² fit of 0.97. The estimated K_i^{app} was $86.7 \pm 23.3 \mu\text{M}$ and the estimated α value was 6.34 ± 3.70 . Kinetic data is also summarised in Table 5.3.

HsODCase Inhibition by C337-0223

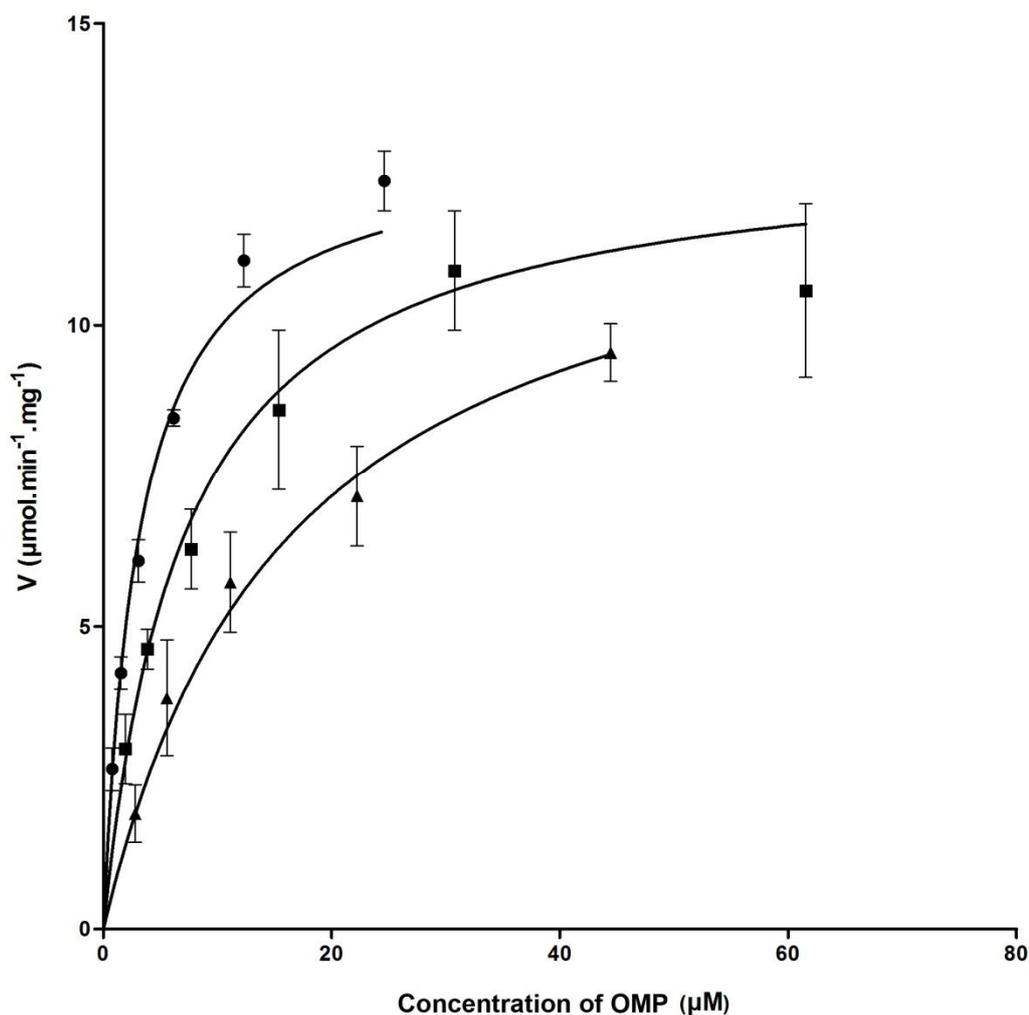


Figure 5.18: Competitive inhibition of HsODCase by compound C337-0223.

HsODCase activity was measured as detailed in Section 2.5.4, at a constant temperature of 30 °C, at three concentrations of inhibitor (0 μM (●), 150 μM (■), and 500 μM (▲)). Rates are means ± SEM (n = 3). The data were found to best fit the Michaelis-Menten equation for competitive inhibition (Appendix 8.6) with a global R² fit of 0.88. The estimated K_i^{app} was 118 ± 22.6 μM. Kinetic data is also summarised in Table 5.3.

*Pf*ODCase Kinetics with C337-0223

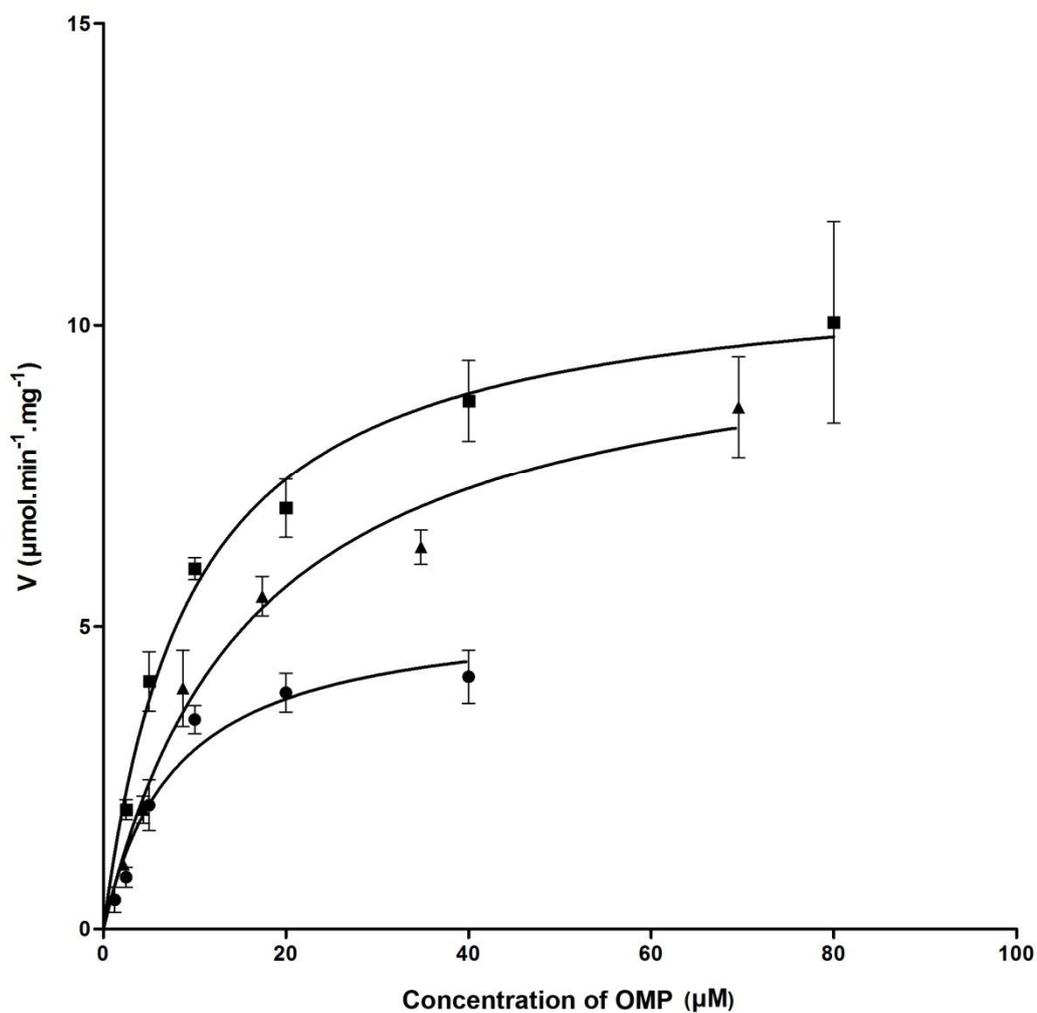


Figure 5.19: Michaelis-Menten kinetics of *Pf*ODCase in the presence of compound C337-0223. *Pf*ODCase activity was measured as detailed in Section 2.5.4, at a constant temperature of 30 °C, at three concentrations of inhibitor (0 μM (●), 150 μM (■), and 250 μM (▲)). Rates are means \pm SEM ($n = 3$). The data were fit to the Michaelis-Menten equation (Appendix 8.5). Kinetic data is summarised below in Table 5.2.

Table 5.2: Michaelis-Menten kinetics for *Pf*ODCase in the presence of compound

C337-0223. K_m^{app} and V_{max} values from curves in Figure 5.19.

	C337-0223 0 μM	C337-0223 150 μM	C337-0223 250 μM
K_m^{app} (μM)	7.86 \pm 2.16	9.51 \pm 2.38	16.19 \pm 3.50
V_{max} ($\mu\text{mol}\cdot\text{min}^{-1}\cdot\text{mg}^{-1}$)	14.23 \pm 1.40	29.49 \pm 2.27	27.5 \pm 2.22
R^2	0.87	0.82	0.92

Table 5.3: Summary of Inhibition Kinetics. Table summarises the results depicted in Figures 5.9–5.18. Where the mode of inhibition is uncompetitive the inhibition constant αK_i^{app} is shown. Where the mode of inhibition is mixed the inhibition constant K_i^{app} as well as the α value are shown.

Compound	Enzyme	Mode of Inhibition	Global Fit R ²	K_i^{app} or αK_i^{app} (μM)	α
8008-2619	<i>Pf</i> ODCase	Uncompetitive	0.95	380.8 ± 49.5	-
4470-0385	<i>Pf</i> ODCase	Uncompetitive	0.96	249.5 ± 29.2	-
	<i>Hs</i> ODCase	Uncompetitive	0.89	1956 ± 489	-
C337-0223	<i>Hs</i> ODCase	Competitive	0.88	118 ± 22.6	-
7009-0959	<i>Hs</i> ODCase	Mixed	0.97	172 ± 42.1	1.25 ± 0.42
C563-0380	<i>Hs</i> ODCase	Mixed	0.91	86.7 ± 23.3	6.34 ± 3.70
4470-0386	<i>Pf</i> ODCase	Uncompetitive	0.95	178.1 ± 21.0	-
	<i>Hs</i> ODCase	Uncompetitive	0.95	543.7 ± 69.2	-
C197-0379	<i>Pf</i> ODCase	Mixed	0.96	49.3 ± 14.5	2.81 ± 1.85
	<i>Hs</i> ODCase	Uncompetitive	0.97	44.21 ± 5.15	-

Table 5.4: Chemical properties of hit compounds. Calculations performed by webservice at molinspiration.com. 'Rule of three' violations are highlighted **light grey**, Lipinski 'rule of five' violations and bioavailability violations highlighted **dark grey** (rules outlined in Section 1.4).

Compound	Log <i>P</i>	PSA (Å ²)	Molecular Weight (Da)	H-bond Donors	H-bond Acceptors	Rotatable Bonds
C197-0379	3.23	100	403	3	7	7
8008-2619	0.931	152	448	4	10	12
4470-0385	1.21	134	374	4	8	7
4470-0386	2.51	112	430	2	8	11
C337-0223	0.93	153	385	3	11	7
C563-0380	3.01	105	467	2	8	9
7009-0959	-1.09	126	304	3	8	5

5.4.2. Kinetic Characterisation of Compound 4049-0191 as an Alternative Substrate of *HsODCase*

Further control assays were conducted for compound 4049-0191 to determine if the compound was an activator or an alternative substrate. The controls consisted of the assay buffer with: the compound by itself (to ensure the compound was not simply decaying in the assay), the compound and the substrate (to ensure no cross reaction between the two), and the compound and the enzyme (no substrate, to confirm activity occurring between the enzyme and the compound).

It was confirmed that the compound was indeed behaving as an alternative substrate (data not shown). In the presence of the enzyme there is a drop in OD_{295 nm}. This was confirmed to be the case in the absence of the substrate OMP which eliminates direct reaction with OMP as a possibility. No reaction occurred in the absence of the enzyme. Increasing the enzyme concentration increased the rate of reaction by a direct correlation indicating an enzyme-dependent rate of reaction and eliminating the possibility that the compound was simply reacting with a reaction buffer component.

A wavelength scan was conducted on an assay consisting of the assay mix with the compound but without enzyme, and a similar assay but with 10 times the amount of enzyme used in the normal assay and incubated for approximately 5 minutes. The λ_{\max} (the greatest *difference* between the absorbance of the compound and its product) was found to be 250 nm (data not shown). This was used for the assays for kinetic characterisation of the compound as alternative substrate.

Michaelis-Menten kinetics was performed on the compound using reaction progress kinetic analysis (Section 2.5.4). Several conditions of the assay were changed to accommodate the new possible substrate. The amount of enzyme was increased to 10 times that used for the regular assays; this was to give a similar rate to the assay with the regular substrate. The assays were carried out with 200 μM of compound 4049-019; this was the maximum possible concentration before exceeding the limits of

the spectrophotometer at the λ_{\max} . The assays were performed at 250 nm (λ_{\max}). The extinction coefficient was calculated for the compound using the method described in Section 2.5.3. No inhibitor was included in the assays so a standard Michaelis-Menten kinetics curve was fit to the data using GraphPad Prism.

Figure 5.20 shows the apparent Michaelis-Menten constant (K_m^{app}) for compound 4049-0191 for *HsODCase* was $678.4 \pm 348.8 \mu\text{M}$ —approximately 170-fold greater than for OMP (which had a K_m^{app} of $3.95 \pm 0.38 \mu\text{M}$). The V_{\max} value of $11.59 \pm 4.9 \mu\text{mol}\cdot\text{mg}^{-1}\cdot\text{min}^{-1}$ is fairly similar to that of OMP ($14.4 \pm 0.47 \mu\text{mol}\cdot\text{mg}^{-1}\cdot\text{min}^{-1}$).

The standard deviations for the K_m^{app} and V_{\max} values are very large. The assay conditions did not approach V_{\max} and the Michaelis-Menten curve is not close to reaching its plateau (Figure 5.20). Thus there is a large degree of extrapolation needed to derive V_{\max} and K_m^{app} which leads to the high standard deviation for these values despite the R^2 value being relatively high at 0.967. Figure 5.21 shows a possible analogous reaction for compound 4049-0191 as an alternative substrate for *HsODCase*.

Mass spectrometry experiments were also conducted months after the kinetic experiments (data not shown). The samples that were run were an assay without enzyme, and an assay with the enzyme that had been incubated for approximately 10 minutes. The samples were subject to MALDI-TOF and LDI-TOF mass spectrometry. The data was inconclusive.

HsODCase: 4049-0191 as Alternative Substrate

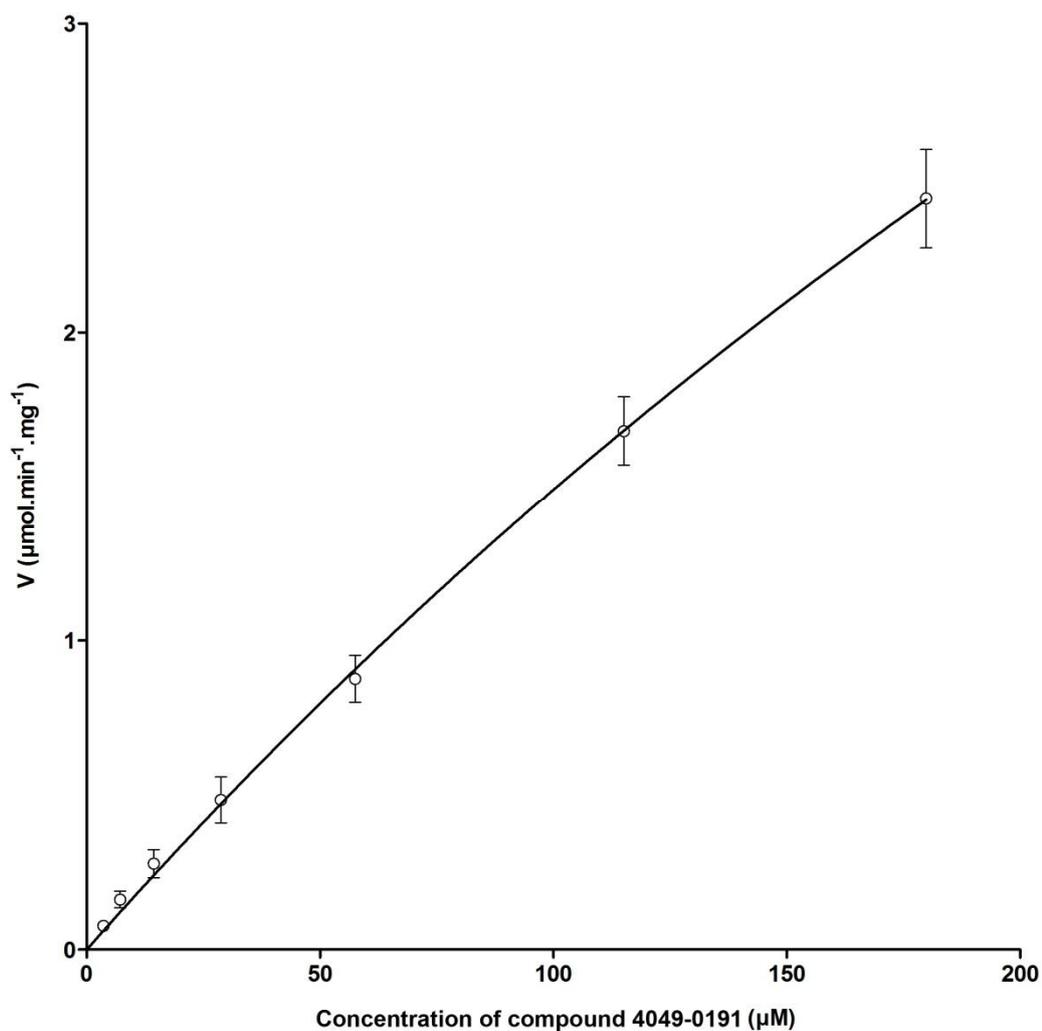


Figure 5.20: Michaelis-Menten kinetics of *PfODCase* with the substrate

4049-0191. *PfODCase* activity was measured as detailed in Sections 2.5.4 and 5.4.2, at a constant temperature of 30 °C. Rates are means \pm SEM ($n = 3$). The data were found to fit the Michaelis-Menten equation (Appendix 8.5) with an R^2 value of 0.967. The estimated K_m^{app} was 678.4 ± 348.8 μM . The estimated V_{max} was 11.59 ± 4.9 $\mu\text{mol}\cdot\text{mg}^{-1}\cdot\text{min}^{-1}$.

5.5. Discussion

5.5.1. Initial Inhibition Screen of OPRTases and ODCases

The compounds that had some level of activity against the OPRTases had several common features. Aside from all the compounds containing one or more planar ring structure (all the screened compounds had this), L268-0351 was the only compound without a carboxyl or phosphate group. There otherwise seemed to be a bias towards the presence of a highly electronegative chemical group in the *in silico* and assay results. This is not surprising as the substrate PRPP—a small compound with three phosphate groups—is extremely electronegative. Another common feature was the presence of a sulfone-amide group branching from a planar ring. This sulfone-amide group was present in half of the compounds that showed some activity against the OPRTases (the group itself was present in approximately 40 % of all the ligands that were screened). In the *in silico* screening results the docked orientations for most of these compounds had the sulfone-amide group positioned roughly where the ribose ring group of the substrate PRPP and product OMP is positioned. It is likely that this group is favoured in the docking and inhibition assay results as the angle induced in the scaffold by the sulfone-amide group and the electrostatic potential is similar to that of the ribose ring group in PRPP, OMP, and UMP.

The ODCase inhibition screen showed similar results. Several of the compounds that showed activity against the ODCases contained the sulfone-amide group. Half of the compounds contained two separate planar ring groups including the two compounds that showed complete inhibition in the initial screen. Highly electronegative groups were also present in most of the compounds showing activity against the ODCases, again including the two compounds with complete inhibition which both had two phosphate groups.

The two compounds that showed complete inhibition were somewhat unique from the rest of the compounds that were screened. They contained the aromatic rings and

phosphate groups that many of the others had. However, the scaffold itself consists only of the two rings connected by highly flexible linker atoms (Figure 5.22). It is most likely that the ring and phosphate groups contribute most greatly to the binding affinity while the flexible scaffold allows the best positioning of these groups. Excessively flexible compounds themselves are not desirable in a drug candidate. The reduction in the entropy increase upon binding leads to a lower Gibbs free energy compared to a rigid ligand that is pre-shaped to the binding site. Likewise modifications to highly flexible inhibitors to increase rigidity and pre-shape them to the binding site tend to significantly increase binding affinity (Velazquez-Campoy et al., 2000).

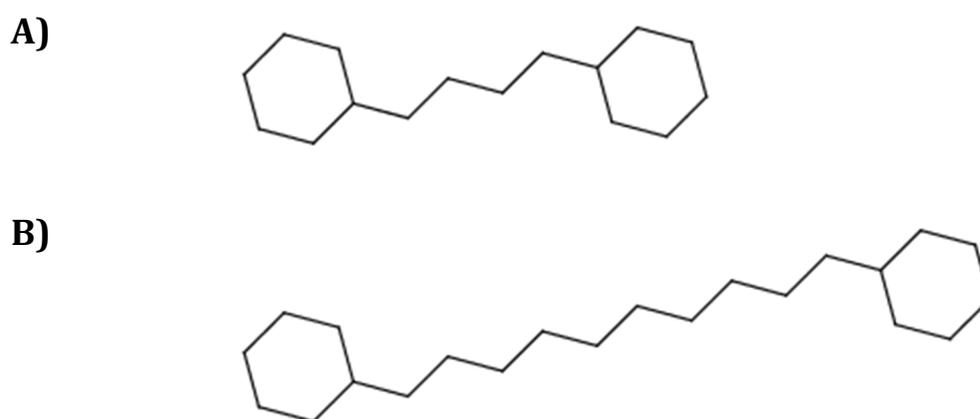


Figure 5.22: Molecular scaffolds (Bemis-Murcko frameworks) for A) 4047-0385 and B) 8008-2619. Calculated at chemicalize.org.

5.5.2. 2D Fingerprint Screening for Analogues of Hits and Inhibition Assays

As mentioned, the opportunity presented itself to search for analogues of confirmed inhibitors of the ODCases. Compounds were used as the query molecule for 2D fingerprint screening of the ChemDiv library and the top scoring compounds would be theoretically analogues of (or very similar to) these compounds.

The screen using 4456-1818 as the query showed three interesting analogues (as can be seen in Table 5.1). The compound 1494-0562 had a large number of polar chemical groups for its size, K783-0416 only had the one carboxyl group and had a dense hydrophobic group (a bromide) on the other, and 4049-0191 also had fewer polar groups as well as a nitro-group in place of a carboxyl group. As compound 4456-1818 was the weakest of the three inhibitors it was prudent to explore more chemical space to maximise the chance of finding a new binder.

Compound 4047-0386 was the visual standout hit from the screen of 4047-0385 (Table 5.1). It was only different for the inclusion of the two ethyl groups attached to the two phosphate groups. It was thought that the extra molecular mass would increase the Van Der Waal forces between the compound and active site. It was also hoped that there was enough room in the active site to accommodate the additional groups or enough flexibility in the ligand to conform to the active site.

Compound C197-0379 was chosen as it was almost identical to C337-0223 (Table 5.1) but for two differences: the removal of a carboxyl group and the inclusion of an extra ring structure in its place. This meant that the volume that the compound occupied would be very similar but the molecular mass was significantly greater. It also meant that the polar surface area and electronegativity was greatly reduced. This allowed for the comparison of similar compounds with dissimilar polarity/electronegativity, with the less polar/more hydrophobic compound yielding a much lower K_i^{app} ($118 \pm 22.6 \mu\text{M}$ for C337-0223 compared to $44.21 \pm 5.15 \mu\text{M}$ for C197-0379).

5.5.3. Inhibition Kinetics of Compounds against *Pf*ODCase and *Hs*ODCase

The progress curves allowed for kinetic characterisation of the ODCases without having to resort to a radiolabeled substrate which would have been both time consuming and costly. It did however have its disadvantages. The assays often went from substrate-saturated, to substrate-limiting rate of reaction, to substrate-exhausted in a short amount of time. The result was that the assays often had a very sharp curve between

the two linear segments which made curve-fitting difficult; this can be seen in Appendix 8.4. This also meant that the bulk of the velocities were derived from a small part of the curve. This occasionally resulted in poor R^2 values (< 0.95) for some of the kinetic curve fits. The assays with inhibitor had curves spanning a longer period of time, but this did not always result in better R^2 values as it was sometimes offset by a greater level of noise from spectrophotometric interference by the inhibitor.

Compounds 8008-2619, 4470-0385, and 4470-0386 showed varying specificity for the *PfODCase* over the human homologue. As was mentioned in Section 1.2.2, specificity whilst not essential is highly desirable in a drug candidate. Compound 8008-2619 had a surprisingly low αK_i^{app} when considering the level of flexibility of the ligand.

Optimisation of this compound to increase rigidity and make its shape complementary to the active site would no doubt increase the binding affinity by many orders of magnitude (as well as increasing bioavailability by fixing the rotatable bond violation for this parameter). To accomplish this, the compound would first need to be co-crystallised with the enzyme to confirm its bound conformation.

Compound 4470-0385 and 4470-0386 showed strong specificity but more interestingly the extra ethyl groups in 4470-0386 correlate with an increase in binding affinity of 1.4-fold and 3.6-fold for *PfODCase* and *HsODCase* respectively. This shows that there is scope for the compounds to be entropically optimised (by way of desolvation of hydrophobic groups). The difference in binding affinity was also much greater in *HsODCase* which suggests that the human homologue may have a higher affinity for hydrophobic interactions. Compound 4470-0386 has a high molecular weight that violates the 'rule of three' for lead likeness as well as too many rotatable bonds (bioavailability violation). This would make lead optimisation on this compound quite limited as only a few more atoms could be added and they would likely need to be spent adding some extra rigidity to the compound. The extra mass and rigidity should however further increase binding affinity.

Compounds 7009-0959 and C563-0380 showed specific inhibition for *HsODCase* (compound C337-0223 also showed specific inhibition for *HsODCase* but not necessarily specific recognition). While there are no immediate applications for compounds that inhibit *HsODCase* whilst *not* inhibiting the parasite homologue the information may be useful in building and comparing SARs to guide ligand optimisation when increasing specificity of existing lead *PfODCase* inhibitors.

As was also mentioned in Section 1.2.2, inhibitors of *HsODCase* have applications for a range of diseases. Compounds C337-0223, 7009-0959, and C565-0380 all inhibited the human *ODCase* well with K_i^{app} values from 88–172 μM . Of these three compounds C337-0223 would be the most problematic lead as it has two Lipinski violations (H-bond acceptors and PSA) which would need to be addressed if it were to be selected as a lead compound. Compounds 7009-0959 and C563-0380 have ‘rule of three’ violations only, which may or may not make them limiting during lead optimisation. Compound 7009-0959 has the lowest molecular weight of all the hit compounds at 304 Da. This allows for plenty of room for expansion of the ligand with the addition of chemical groups and expansion of the scaffold. The binding affinity is also very high considering that it has a significantly lower molecular weight than the other two compounds.

Compound C197-0379 is the most promising lead compound from this study for three reasons. First, it is by far the most potent inhibitor with considerably low kinetic constants for an unmodified screening hit compound. Second, it has only minor ‘rule of three’ violations allowing for a reasonable scope of lead optimisation. Third, it inhibits both the human and *P. falciparum* *ODCase* almost equally well. While this last point is not ideal for a malarial drug, it does greatly amplify the potential applications for this compound beyond just the treatment of malaria (if it were chosen as a lead for the treatment of malaria it could still be modified to specifically bind the malarial *ODCase*). As outlined in Section 1.2.2 inhibitors of the *de novo* biosynthesis pathway in human

cells have a large number of applications. More importantly given that the compound has shown almost no specificity for either of these enzymes it stands to reason that it would also likely be an inhibitor of at least some of the bacterial and protozoan targets outlined in Section 1.2.2.

Compounds C197-0379 and 4470-0386 are analogues of C337-0223 and 4470-0385 respectively, which were found from 2D fingerprint screening. This sort of re-screening was a very small scale 'hit expansion' H2L method (hit expansion outlined in Section 1.6). This popular method is used to find and biochemically test a large number of analogues of initial hits. The extra hits result in new information that can help build a Quantitative Structure-Activity Relationship (QSAR). QSARs are used to find new inhibitors, identify the best lead compound, and even assist with lead development (Aguiar-Pulido et al., 2013, Cruz-Monteagudo et al., 2012).

During lead development compounds undergo chemical group additions and substitutions to attempt to elicit improved binding affinity or chemical properties. Compound 4470-0386 is identical to compound 4470-0385 with the addition of two ethyl groups; this is an example of a possible 'hit evolution' step (Section 1.6) that might be performed during lead development. Compounds C197-0379 and C337-0223 are an example of a possible 'isosteric replacement' step (Section 1.6) where the substitution of the carboxyl (on C337-0223) for a fused ring (to yield C197-0379) shows an improvement in binding (lower K_i) and improved chemical properties (as C337-0223 has rule-of-five violations but C197-0379 does not).

The *in silico* screening was designed to find competitive inhibitors of the target enzymes. Instead many of the compounds showed a mode of action that was uncompetitive, mixed and non-competitive. There was only one case of competitive inhibition (compound C337-0223 against *HsODCase*). The *ODCases* have no apparent allosteric sites where the inhibitors could bind (and it is extremely unlikely that most of the inhibitors were found by happenstance). Neither is the active site large enough to

fit both the substrate and any one of the inhibitors. The active sites themselves are only approximately 15 Å apart with the active sites being comprised of residues from both monomers. The key active site lysine residue responsible for protonation of the C6 carbonyl (as labelled in Figure 5.21) is only three residues from an aspartate involved in hydrogen bonding to the ribose on the OMP in the other active site. It is therefore possible that there is a degree of interconnectivity between the active sites.

The non-competitive inhibitor suggests that one active site on the dimer can be blocked from binding the substrate by the inhibitor binding in the other active site. The competitive inhibitor suggests that binding of the substrate (or the competitive inhibitor) doesn't prevent binding in the other active site (although this was already known from the crystal structures). The uncompetitive inhibitors however suggest that binding of the substrate affects the other active site in some way (perhaps inducing a more stabilised or enclosed conformation). It also suggests that a conformation change is either involved with catalysis or product release (the reaction is reversible to a degree (Vardi-Kilshtain et al., 2013)) which is blocked by the presence of the uncompetitive inhibitor.

Catalysis requires the presence of the key two lysine and two aspartate residues in a specific conformation (shown in Figure 1.4b) around the carboxyl group of OMP (to either destabilise the ground state or stabilise the transition state, and to protonate the C6 carbanion). It's likely the conformation change induced by the presence of the uncompetitive inhibitors disrupts the specific arrangement of these four residues, thus preventing catalysis.

Also of interest, compound C337-0223 gave an apparent *increase* in V_{max} (and K_m^{app}) for *PfODCase*. The increase in apparent V_{max} shows that it is acting as an allosteric activator by improving the speed of catalysis. It may be that catalysis is accelerated upon binding of this compound to the enzyme-substrate complex or that it binds prior to the

substrate. Either way it does further show significant conformation change in one active site upon ligand binding in the other.

Surface plasmon resonance or isothermal titration calorimetry could be used to characterise and compare the conformational changes upon binding of substrate and inhibitors, as well as confirming the conformation changes that occur during catalysis and product release as reviewed in Miller and Wolfenden (2002). MD simulations might also be used to determine structural changes that might occur upon binding of substrate and inhibitor, catalysis, and release of product, particularly of the active site lysine and aspartate residues.

5.5.4. Kinetic Characterisation of Compound 4049-0191 as an Alternative Substrate of *HsODCase*

The main limitation to accurately measuring the K_m^{app} and V_{max} for the enzyme with compound 4049-0191 had been due to the low binding affinity of the compound and the high base absorbance at 250 nm. The wavelength scan used to identify λ_{max} (data not shown) indicated that it would be possible to use a wavelength with a slightly lower base absorbance (but still with a significant difference between the substrate and the product) to allow a higher substrate concentration to be used (however, probably not enough to approach V_{max}). Alternatively a non-spectrophotometric assay could be used.

Compound 4049-0191 does not have a carboxyl group (*HsODCase* is a decarboxylase). It does however have a nitro group attached to a planar ring structure, similar to the carboxyl on the orotate ring. While mass spectrometry experiments need to be conducted to confirm that the nitro group is being removed, the proposed theory in this study is in line with the direct decarboxylation mechanism theory for the catalytic mechanism of *ODCase* (as outlined in Section 1.2.4), as it could form a very similar reaction shown in Figure 5.21 (same as Figure 1.5 but juxtaposed with the proposed analogous reaction). The nitro group would provide a similar electrostatic potential that is needed for local destabilisation of the planar ring. The C6 anion can still form in

conjunction with ion pairing with the positively charged active site Lysine, followed by protonation of C6. The stabilising forces of the OMP-ODCase complex are achieved in large part by electrostatic complementarity and hydrogen bonding by the phosphoribose moiety and pyrimidine ring carbonyl groups (Hu et al., 2008, Langley et al., 2008, Tokuoka et al., 2008). The sulfone-amide bridge in 4049-0191 may be involved in similar interactions with ODCase but it is most likely that the bulk of stability comes from Van Der Waal forces and desolvation of the hydrophobic components of the compound. The main problem with this theory is the lack of similarity between the OMP and the compound, and the position of the nitro group. It is not clear how the compound would be oriented in the active site or how similar in position and angle the nitro group would have to be to the carboxyl for the reaction to occur.

Docking followed by Quantum Mechanics-Molecular Mechanics (QM/MM) simulations could shed some light on the issue. The difference in K_m^{app} values between 4049-0191 and OMP likely arises from a reduced overall binding affinity of 4049-0191 compared to OMP. The nitro group may also not be as locally-destabilising as the carboxyl, which would affect V_{max} and could affect the K_m^{app} value.

6. Development and Validation of a Novel Hybrid Screening Method

6.1. Introduction

Virtual screening has become an essential component of lead drug discovery (Nicholls, 2008). While there are a large number of programs available to perform virtual screening they all fall into one of two categories: structure- or ligand-based. Structure-based methods involve molecular docking to a binding site followed by a scoring function to estimate the binding affinity. This is far more CPU intensive than ligand-based methods (von Korff et al., 2009) and requires parallel computing infrastructure for screening on any meaningful scale. The accuracy of docking to a rigid receptor is also incredibly limited to the receptor binding site's conformation. Ligand-based methods such as screening with a pharmacophore model are much faster and can be performed on a single CPU in a matter of hours. Ligand-based methods are however generally limited to targets for which there are multiple known binders. They also have a reduced scope for identifying structurally diverse compounds compared to structure-based screening as the method involves searching for compounds that are similar to the ligands on which the pharmacophore model is based (Salam et al., 2009, Scior et al., 2012). Scior et al. (2012) highlights the stringency of queries and feature weighting as significant pitfalls of pharmacophores, where pharmacophores with too many descriptors yield extremely low structural diversity in hit compounds and 'fuzzy' models yield an extremely high false positive rate.

Hybrid and parallel methods have been attempted in the past with mixed success. There are studies where pharmacophores have been produced from probe docking (Arnold et al., 2004) or fragment docking (Loving et al., 2009), and numerous methods are described (Carlson et al., 2000, Chen and Lai, 2006, Barillari et al., 2008, Cross et al., 2012) and software is available (accelrys.com/products/discovery-studio/) for generating structure-derived pharmacophores. While this enables the generation of a pharmacophore model based only on a protein structure there is no experimental data

to identify features that are known to be explicitly required for recognition of a ligand to a binding site. Nor does it eliminate the large trade-off between specificity and sensitivity that is endemic with the way in which screening with pharmacophores is carried out, i.e. treating every descriptor in a pharmacophore as essential rather than simply beneficial to binding affinity. Parallel screening, where two independent approaches are used on the same target, has seen a growing popularity to combat inconsistency in screening methods depending on the receptor target (Tan et al., 2008, Swann et al., 2011, Svensson et al., 2012, Drwal and Griffith, 2013, Houston and Walkinshaw, 2013).

The purpose of combining structure-based and ligand-based methods for *in silico* screening was to eliminate the limitations of each approach. The aims were to create a hybrid screening method that is similar in speed to ligand-based methods, requires only the information that would be needed of a structure-based method, maintains the higher hit diversity that structure-based screening methods have compared to ligand-based methods, and has a greater consistency of results (without having to conduct parallel screening). The success of the screening methods is determined by its ability to consistently rank active molecules (actives) over decoy molecules (decoys). The two measures used here for this are early enrichment and Area Under the Curve (AUC) values of Receiver Operating Characteristic (ROC) curves. Early enrichment is a measure of the fraction of actives ranking in the top few percent of the top scoring compounds. Actives need to score in the top few percent of hits for a screening method as these are typically the only compounds that are further evaluated by inhibition assays. AUC values are a measure of the overall bias towards actives or decoys. A method that yields poor early enrichment values but high AUC values would still be very useful as a pre-screening/filtering method. Screening speed is measured in CPU wall-time. A method for analysing hit diversity of a group of compounds is described in this chapter. A docking screen (structure-based) and 2D-Fingerprint screen (simple ligand-based) were also performed on the well characterised targets during validation

and the results were analysed in the same way. Hence the results of the hybrid screening method could be compared to both a structure- and ligand-based approach.

6.2. Development of the Hybrid Screening Protocol

Briefly, the original aim of the hybrid method was to conduct fragment docking on a target enzyme, use the information generated by docking to derive information about the active site (such as chemical group 'hot spots') and use this to generate a ligand-based screening model (Figure 6.1).

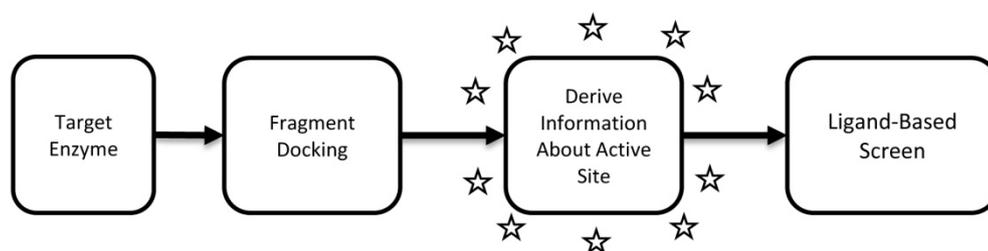


Figure 6.1: Original idea for hybrid screening method.

This approach had been undertaken to some degree in a previous study (Loving et al., 2009) however with poor enrichment values. In Loving et al. (2009) 15 of the top scoring docked fragment-like compounds were clustered and used to generate a pharmacophore. Here it was planned to use all of the docked fragment poses and cluster the key chemical groups (such as rings, hydrogen bonding atoms etc.) and use them to generate a pharmacophore model. This was originally attempted with the UCSF ZINC 'fragments' subset. This library was filtered to only include compounds that were less than a molecular weight of 100 Da. This resulted in approximately 8 000 small fragments. The problem that was encountered after docking was conducted was that the vast majority of the compounds were docking to the same small area of the active site and hence the active site was not being properly characterised.

The fragment docking approach was therefore dropped in favour of docking a small library of 'probe' molecules (described in more detail in Section 6.3.2.1). Essentially, the molecules were either ring structures, or a scaffold with a single chemical group. Only a small number of probe molecules would be docked, but thousands of orientations of these molecules would be generated. This would force UCSF Dock to thoroughly characterise the active site with the probe molecules. This revised approach is illustrated in Figure 6.2.

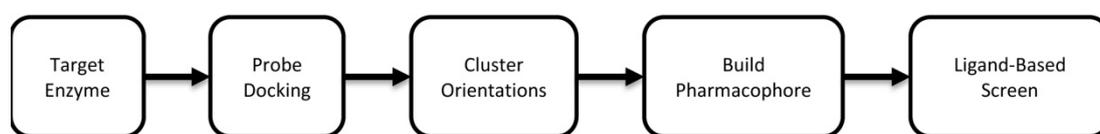


Figure 6.2: Revised concept for hybrid screening method.

The probe docking and clustering appear to work well (described in more detail in Sections 6.3.2.3–6.4.2.5). OpenEye ROCS was chosen as the ligand-based screening program as the interface was conducive to making custom models from the type of data that was generated, as well as the use of ligand 'shape' score that is not present in many ligand-based screening programs. Some trial-and-error attempts were made to generate a ligand-based screening model in OpenEye ROCS. Initial models attempted to include pharmacophore descriptors to represent all of the clusters identified by the probe docking. This resulted in a very 'fuzzy' model and the ROC curves that resulted when using these models were generally very poor. A few models were generated that 'cherry picked' descriptors and added weighting to some. These models resulted in better ROC curves but introduced an enormous degree of subjectivity and user input that would not be able to be included in a generalised screening method. It was recognised that the problem was with ROCS treating every descriptor as being essential for molecular recognition. Hence the large amount of information generated by the

probe docking could not all be included in a traditional ligand-based screening model. The method needed further optimisation to either: predict which features were essential, or, perform the ligand screening in a novel way that didn't treat every descriptor as essential. The latter appeared more desirable as it would use all the probe docking information and potentially result in more structural diversity in the hits.

The final method (Figure 6.3), briefly, was to generate descriptors for all the identified clusters, pre-align the compounds to the active site, individually score the aligned compounds to every descriptor in ROCS, and compile the Tanimoto scores (with weighting from the average docking Grid scores) to create a completely arbitrary score for ranking compounds. The scoring is simplified and illustrated in Figure 6.4.

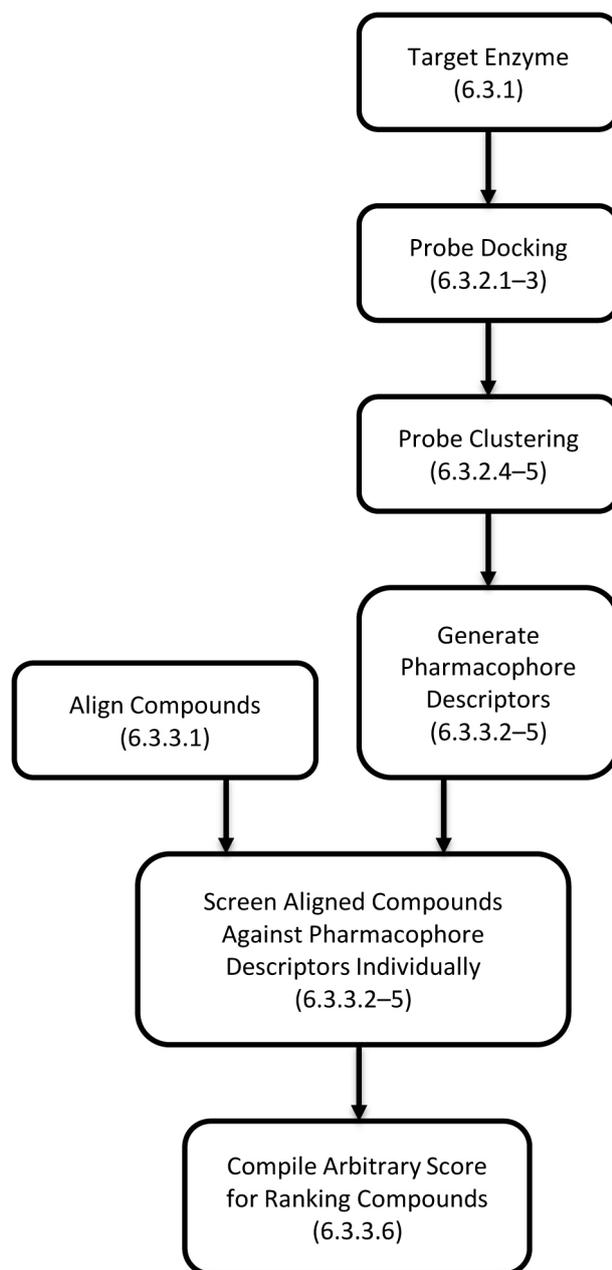


Figure 6.3: Final concept and workflow for hybrid screening method. The sections in this chapter relating to each step are indicated.

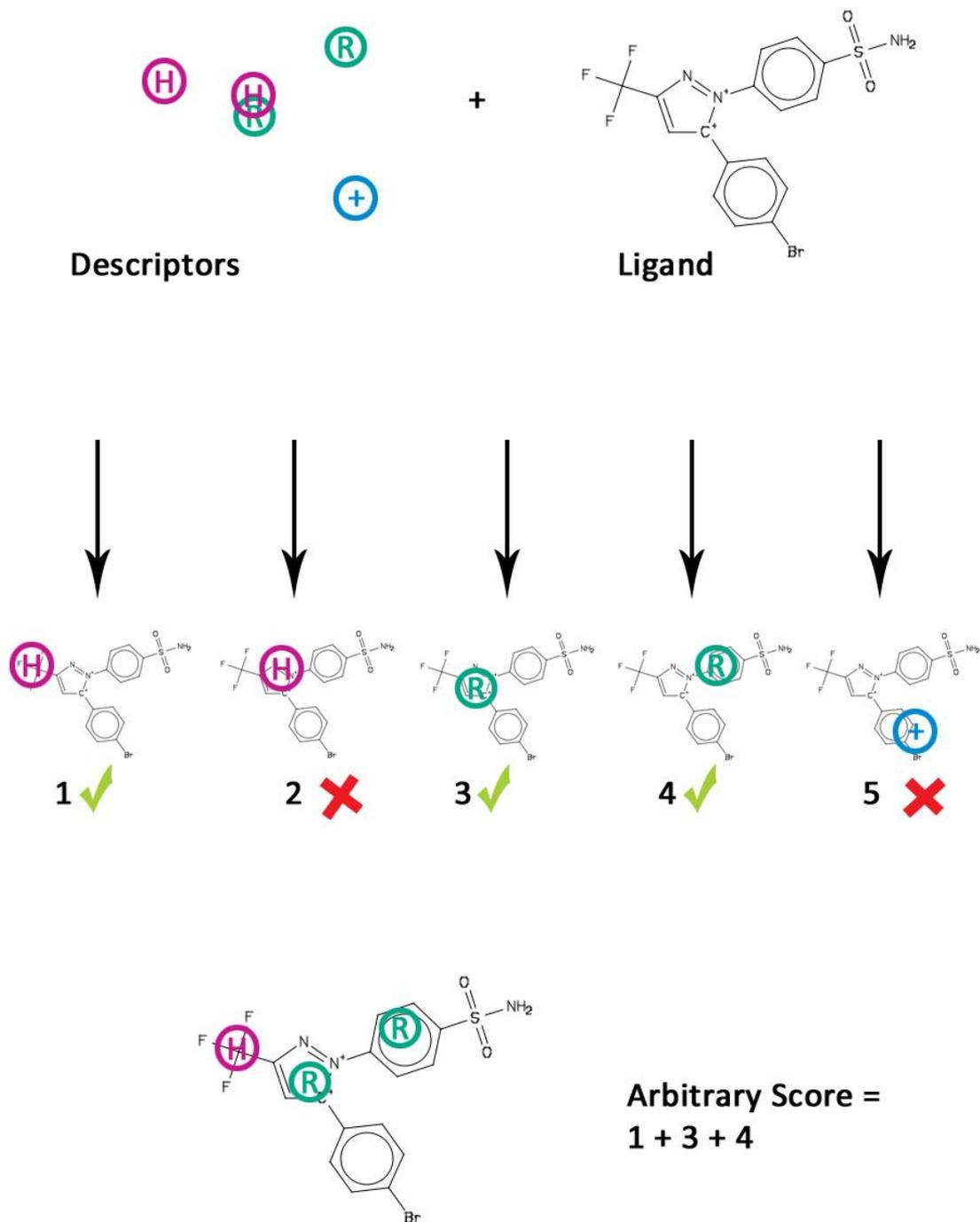


Figure 6.4: Basic concept for performing the unique screening and scoring. The descriptors are individually screened against the aligned compounds. The scores are compiled to make an arbitrary score. Example descriptors are **(H)**, hydrophobic group; **(R)**, ring group; and **(+)**, positive ionisable group.

6.3. Methods: Hybrid Screening and Analysis

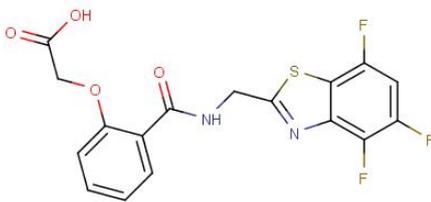
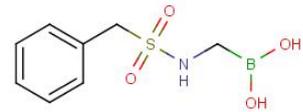
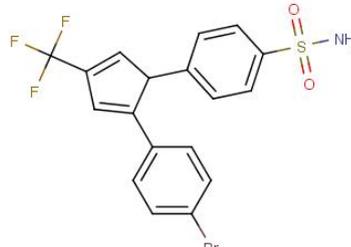
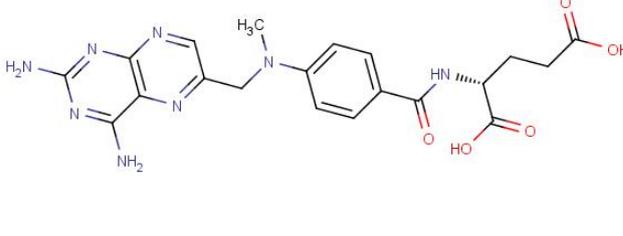
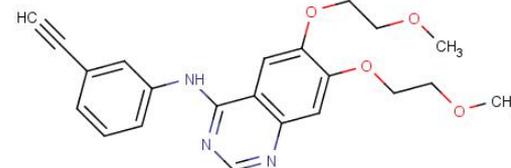
6.3.1. Selection of Pharmaceutically Relevant Protein Targets, Actives and Decoys

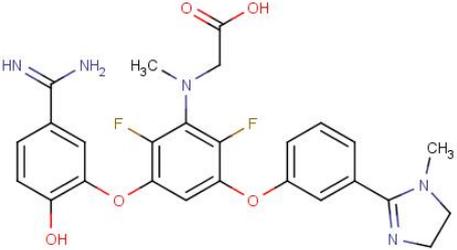
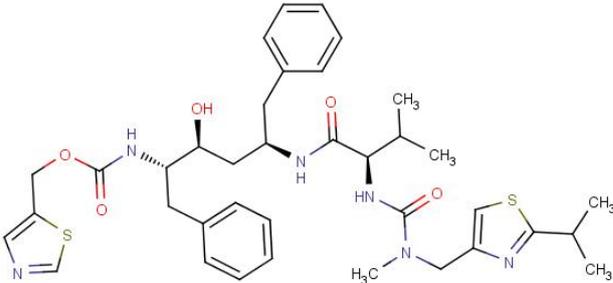
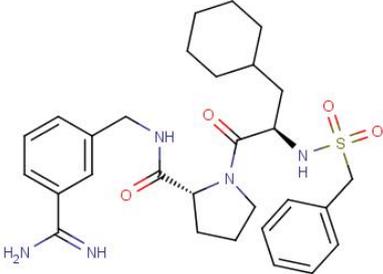
A set of diverse protein active sites for which there are multiple known binders was required in order to assess the quality of the hybrid screening method. Enzymes were selected to ensure that there was diversity in the active site characteristics such as size, hydrophobicity, accessible functional groups that carry a charge and accessible functional groups with hydrogen bond donor or acceptor potential. Crystal structures that were co-crystallised with a known binder were required over apo structures to ensure that the active sites were in a conformation that would favour ligand binding, and for aligning of the ligands to the co-crystallised ligand. Based on these criteria, Table 6.1 shows the enzymes that were selected (and their abbreviations), the PDB ID code for the structures of these enzymes that were obtained from the Protein Data Bank (Berman et al., 2000), and the ligands with which they had been co-crystallised.

For each of the protein targets the UCSF DUD (Huang et al., 2006) was used to obtain active and decoy compounds in a ready-to-dock .mol2 format. The UCSF DUD active data sets contain many known binders for each of the target enzymes. The UCSF DUD decoy sets contain, for each active, 36 decoys with similar physical properties (e.g. molecular weight, calculated LogP) but dissimilar topology.

The Schrödinger GLIDE (Friesner et al., 2004) 400MW decoy set was also used as an alternative decoy set. This is a generic decoy set of approximately 1 000 drug-like compounds with an average molecular weight of 400 Da and is optimised for high diversity.

Table 6.1: Pharmaceutically relevant protein targets.

Target Protein Abbreviation PDB ID code	Co-crystallised ligand
Aldose Reductase ALR2 1T40	
AmpC beta lactamase AmpC 3O86	
Cyclooxygenase 2 COX-2 6COX	
Dihydrofolate reductase DHFR 3DFR	
Epidermal growth factor receptor kinase EGFr 1M17	
Fibroblast growth factor receptor kinase FGFr1 2FGI	

<p>Factor Xa FXa 1FJS</p>	 <p>The structure shows a central benzene ring with a hydroxyl group at the para position and an amide group at the other para position. This central ring is connected via ether linkages to two other benzene rings. One of these outer rings has a methyl group and a carboxylic acid group at the para position. The other outer ring is connected to a 1-methyl-1H-imidazole ring.</p>
<p>HIV protease HIVPR 1HXW</p>	 <p>The structure is a complex molecule featuring a central chiral carbon atom bonded to a hydroxyl group, a benzyl group, and two amide groups. One amide group is attached to a thiazole ring, and the other is attached to a 2-isopropyl-5-methylthiazole ring. There is also a methyl group attached to the central carbon.</p>
<p>Neuraminidase NA 1IVD</p>	 <p>The structure shows a benzene ring with a hydroxyl group at the para position, a carboxylic acid group at the other para position, and a nitro group at the meta position. There is also an amide group attached to the benzene ring.</p>
<p>P38 mitogen activated protein kinase P38 1YWR</p>	 <p>The structure features a central pyridine ring substituted with a methyl group, a benzyl group, and a piperidine ring. It also has a fluorophenyl group and a methyl group attached to the pyridine ring.</p>
<p>Tyrosine kinase SRC SRC 10PK</p>	 <p>The structure shows a central pyridine ring substituted with a methyl group, a hydroxyl group, and a benzyl group. It also has a chlorophenyl group and a hydroxyl group attached to the pyridine ring.</p>
<p>Thrombin Thrombin 3RM2</p>	 <p>The structure is a complex molecule featuring a central chiral carbon atom bonded to a hydroxyl group, a benzyl group, and two amide groups. One amide group is attached to a piperidine ring, and the other is attached to a benzene ring. There is also a methyl group attached to the central carbon.</p>

6.3.2. Hybrid Screening: Structure-Based Component

6.3.2.1. Probe Library Design

The probe molecules used for characterisation of the receptor binding site are shown in Figure 6.5. These probe molecules were selected specifically to mimic the functional groups that OpenEye ROCS (Section 1.5.2.2) recognises and represents as pharmacophore descriptors for screening.

The probe molecules are spread into three libraries referred to here as the 'Rings', 'Hydrophobes', and 'Ions' probe libraries. There was originally a library for probing hydrogen bonders (these were mainly simple scaffolds with -OH chemical groups); however Dock would combine Van Der Waal contacts with hydrogen bond contacts in its VDW score, and this meant that probes couldn't be ranked on hydrogen bonding affinity. Instead, a single water molecule was used as the molecular probe to characterise hydrogen bonding potential of the active site (not shown in Figure 6.5) such that the majority of the Dock VDW score would come from any hydrogen bonding interactions rather than Van Der Waal contacts.

The Rings probe library consists of several aromatic, cyclic organic compounds (although ROCS also classes non-aromatic cyclic and even non-planar cyclic groups under the 'Ring' descriptor); both polar and non-polar molecules are included. The Hydrophobes probe library consists of several non-polar, organic compounds of similar molecular weight, but varying size and shape. The Ions probe library consists of several simple scaffolds with either a carboxylate to represent an anionic functional group or a protonated amine to represent a cationic functional group.

The probe compounds were generated and exported in .pdf format using an online structure file generator (<http://cactus.nci.nih.gov/translate/>). The probe compounds in .pdf format were then prepared for docking and converted into .mol2 format as previously described (Section 4.2.2.1).

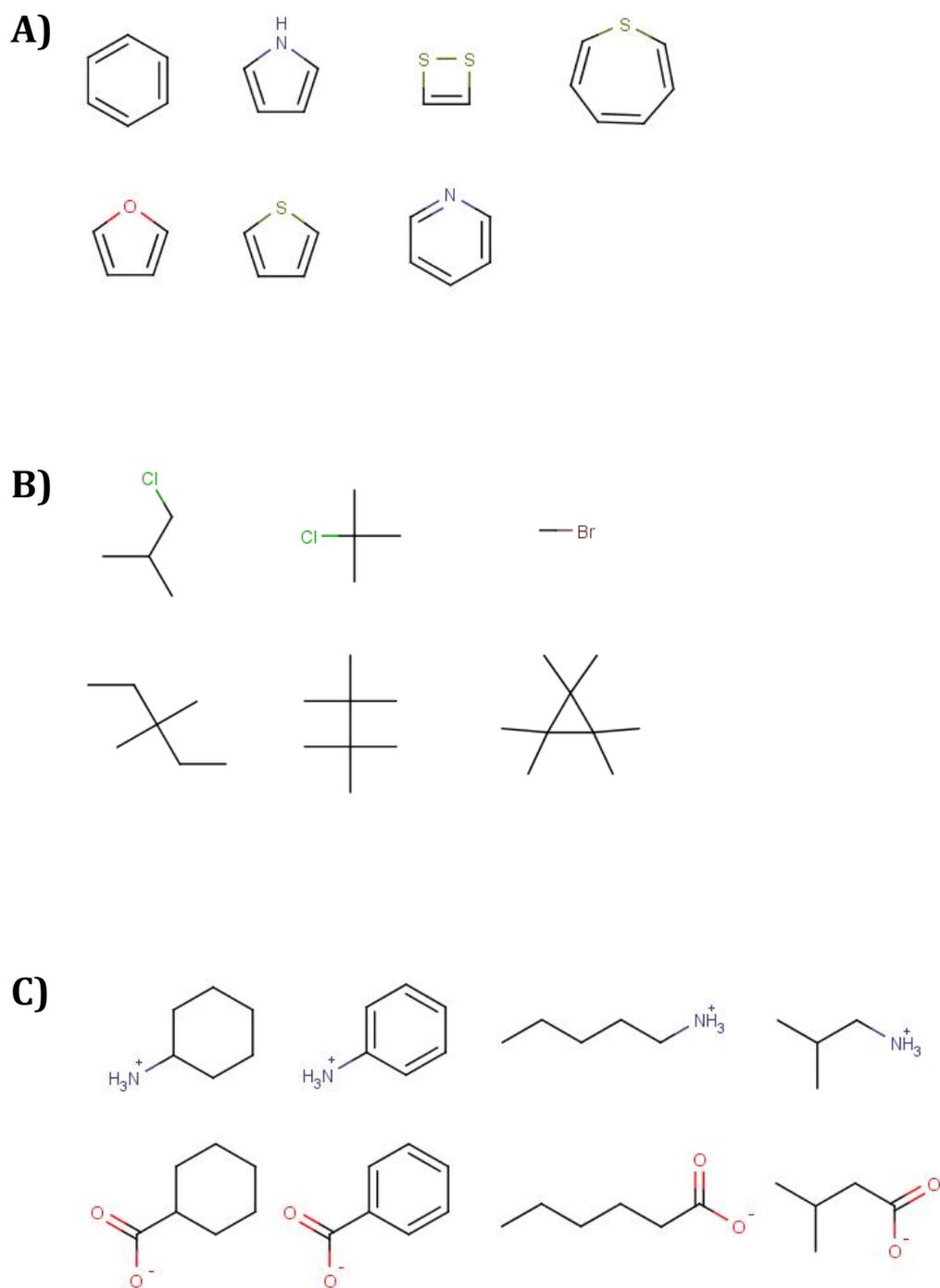


Figure 6.5: Probe molecules used for characterisation of molecular binding sites of pharmaceutically relevant protein targets. A) Rings probe library molecules. B) Hydrophobes probe library molecules. C) Ions probe library molecules. A water molecule (used to probe for hydrogen bonds) is not shown.

6.3.2.2. Preparation and Validation of Docking Parameters and Files

Receptor active sites and probe libraries were prepared for docking using UCSF Chimera (Pettersen et al., 2004) as previously described (Section 4.2.2.1). Spheres were generated by writing the molecular surface of the aligned, bound ligand to a .dms file using UCSF Chimera's 'write DMS' tool. The DMS file was then used by UCSF Dock's 'SPHGEN' utility to generate the sphere files. SPHGEN was run to generate spheres on the inside of the DMS surface, using default parameters. Grid files were generated using the method previously described (Section 4.2.2.1). Validation of the active sites by docking of the known binders was carried out as described previously (Section 4.2.2.2).

Probe docking with a water molecule was also carried out using OpenEye FRED (McGann, 2011). The Receptor structures in .pdf format were loaded into FRED Receptor, and protein chains and ligands were designated. The site boxes were automatically generated for the ligands (minor manual adjustments were made if part of the active site was not within the box). Shape was calculated using the 'high quality' setting, the 'inner contour' was disabled and the 'outer contour' was increased to approximately 75 %. Trial docking was then performed using the bound ligands and the prepared receptor structures were saved in OpenEye's .oeb.gz compressed binary format.

The active and decoy libraries were converted to multi-conformer structure databases using OpenEye OMEGA (Hawkins et al., 2010) for use with OpenEye ROCS and EON (Muchmore et al., 2006) (multi-conformer structure databases are used as OpenEye screening programs treat the ligands as rigid structures). The multi-conformer databases were saved as OpenEye's .oeb.gz compressed binary format.

6.3.2.3. Probe Docking

Each of the protein targets was docked with the probe molecule libraries using UCSF Dock. The validated, high stringency docking parameters were used for probe docking with some modification; the maximum number of orientations to attempt and score

was increased to 10 000 and the top 2 000 scored orientations per molecule were saved. This ensured very thorough coverage of the active site for each probe molecule.

The water molecule was docked using OpenEye FRED (as UCSF Dock was unable to properly orient the molecule) and re-scored with UCSF Dock. Docking with FRED was carried out using default parameters with the following changes: maximum number of conformations was increased to 10 000, maximum number of scored conformations was increased to 10 000, and the output was saved in .mol2 format. Rescoring with UCSF Dock simply involved a docking run using the default scoring function and with 'ligand orientation' switched off.

6.3.2.4. Clustering of Docked Probe Conformations

The docked probe orientations were output from Dock as one .mol2 file for each probe library containing all 2 000 orientations of each of the probe molecules in that library. Both the Rings and the Hydrophobes docked probe orientations were subject to clustering without modification.

UCSF Chimera was used to modify the .mol2 output files for the Ions and Hydrogen Bond docked probe orientations (the water molecule orientations rescored with Dock). For the Ions, all atoms except for oxygen and nitrogen atoms were removed and the nitrogen and oxygen atoms were saved as separate .mol2 files referred to hereafter as 'Cations' and 'Anions' respectively. This had the same effect as retaining only the carboxyl and amine groups but reduced the time required to perform the clustering. For the hydrogen bond docked probe orientations the .mol2 file was opened along with the .mol2 target enzyme file used for docking. Hydrogen bonds were then calculated between the probe orientations and the target enzyme and all atoms were removed except for the probe molecule atoms that were involved in hydrogen bonding. These hydrogen and oxygen atoms were then saved to separate .mol2 files referred to hereafter as "Donors" and "Acceptors" respectively.

The Rings, Hydrophobes, Anions, Cations, Donors and Acceptors were clustered using the Python clustering script shown in Appendix 8.22. The clustering method involved grouping probe orientations into groups within a pre-defined radius (2 Å was used here). The centres of mass of each probe orientation were calculated and saved to a clustering array along with that orientation's Grid score. The centroid for all the orientations in the array was calculated. Probe orientations were removed from the clustering array one by one (the farthest one from the centroid) to a reserve array and the new clustering array centroid re-calculated each time until all remaining probe orientation centres of mass were within the pre-defined (2 Å) radius of the array centroid. These probe orientations were output as a cluster. The x, y, z coordinates of the cluster's centroid were saved, along with the mean Dock score and number of probe orientations in the cluster. The reserve array then becomes the new clustering array and the cycle repeated. The process is repeated identifying as many clusters as necessary until no more probe orientations remain.

6.3.2.5. Cluster Selection

Clusters were used in generating descriptors in the ligand-based screening component of the hybrid screening approach if they had a significant number of orientations representing that cluster, and if they had a mean Grid score below the cut-off value. Rings, Hydrophobes, Cations and Anions clusters that were comprised of 50 or fewer probe orientations were not used. For the Donors and Acceptors, clusters that were comprised of five or fewer probe orientations were not used. The cut-off values for the mean Grid scores for the clusters varied for the different libraries. The cut-off values were -7.0 for the Rings and Hydrophobes, -15.0 for the Cations and Anions, and -1.5 for hydrogen bond Donors and Acceptors.

6.3.3. Hybrid Screening: Ligand-Based Component

6.3.3.1. Aligning the Known Active and Decoys

For the individual scoring to be carried out (Sections 6.3.3.2–6.3.3.5) the compounds to be screened had to already be pre-aligned to target the active site. For each target enzyme, an initial ROCS query was used for orientating the compounds. The co-crystallised ligand from the target's .pdb file was imported into ROCS (Section 1.5.2.2) and the automatically-generated descriptors were removed. The centroids of clusters that were flagged for use as descriptors were imported. Cluster centroids were kept if they matched features on the ligand (for instance, if a 'rings' cluster was in the same position as a ring group on the ligand then it was retained). The representative descriptors for the retained clusters were included in the ROCS query. These descriptors comprised the query's 'color' descriptor score. The ligand itself comprised the query's shape Tanimoto scoring.

A ROCS screen was performed on the active and decoy multiconformer databases against these initial pharmacophores, using default parameters with the following changes: 'maximum conformers' was set to 1 and 'hitlist size' increased such that all compounds would be scored. Essentially, this orients each compound in the active site and returns only the single highest scoring orientation.

These aligned databases were saved in .oeb.gz format for use with individual descriptor screening, occupancy and atom clash scoring, and electrostatic similarity scoring (Sections 6.3.3.2–6.3.3.5).

6.3.3.2. Individual Pharmacophore Descriptor Scoring

The Tanimoto 'color' component of OpenEye ROCS's score was utilised for scoring of clustered probes. A ROCS query was created for each of the identified clusters. The queries consisted of a single descriptor representative of the cluster, placed at the centroid. The pre-aligned actives and decoys were then screened against each query

using ROCS. The screening was carried out using default parameters with the following changes: the screening was set to 'score only' (such that incoming poses were simply scored and output) and the hitlist size was increased such that all compounds would be scored and saved. The results from each screen were saved in .oeb.gz format and opened in OpenEye VIDA (eyesopen.com/vida). VIDA was then used to export the Tanimoto 'color' scores for each compound in .csv spreadsheet format. The spreadsheets were used for compiling the arbitrary score in Section 6.3.3.6.

6.3.3.3. Active Site Occupancy Scoring

The Tanimoto shape component of OpenEye ROCS's score was utilised for analysing active site occupancy and ligand-enzyme overlap. A ROCS query was generated for each target using the co-crystallised ligands. The ligands were imported into the manual query builder, all automatically-generated descriptors were removed (such that only the atoms and shape contour remained) and the query was saved. The aligned actives and decoys (Section 6.3.3.1) were then scored against these queries (without reorientating), saved and exported to .csv format as previously described (Section 6.3.3.2).

6.3.3.4. Ligand-Enzyme Overlap Penalty

The .pdb enzyme structure files were prepared in UCSF Chimera. The active site atoms (excluding hydrogen atoms if present) within 5 Å of any known inhibitor atom were saved to a separate .pdb file. The number of atoms in this file was noted. This was used to represent the active site for the ligand-enzyme overlap penalty. A Tanimoto score greater than zero would indicate that there is an atom clash. Using both the Tanimoto score and the known number of atoms used in the query the extent of the clashing atoms could be approximated.

A ROCS query was made for each enzyme active site using these newly created .pdb files. The file was imported into the manual query builder, all automatically-generated descriptors were removed (such that only the atoms and shape contour remained) and

the query saved. The aligned actives and decoys (Section 6.3.3.1) were then screened against these queries, saved and exported to .csv format as previously described (Section 6.3.3.2).

6.3.3.5. Electrostatic Similarity

OpenEye EON was used to perform an electrostatic similarity search on the aligned actives and decoys from Section 6.3.3.1 against the co-crystallised ligand. The EON search was performed using default parameters with the following options: ‘-scdbase’ was flagged and ‘-besthits’ was set to zero to ensure all compounds were scored and output, and ‘-scoreonly’ was flagged to prevent orientating of the already aligned conformations. The output was saved to .oeb.gz format and exported to .csv format as previously described (Section 6.3.3.2).

6.3.3.6. Compilation of Final Arbitrary Score

The arbitrary score for a compound was the summation of the descriptor, shape, overlap penalty and electrostatic similarity scores. The descriptor scores were calculated as the product of the ROCS Tanimoto score, the cluster’s mean Grid score from docking (Section 6.3.2.3–6.3.2.5), and a multiplier. The shape score and overlap penalty were calculated from the Tanimoto score a multiplier and the previous descriptor scores. The electrostatic score was the product of the EON ‘combo’ score and a multiplier. Multipliers were used to keep the relative contributions for the various intermediate scores similar to ensure they didn’t over or under contribute in the final score. The multipliers and formulae are outlined in Appendix 8.23.

6.3.4. Ligand-Based Screening Control: MOLPRINT 2D

As a comparison for the hybrid screening method a 2D molecular fingerprint search was performed with MOLPRINT 2D v1.2 (Bender et al., 2004). For each of the pharmaceutically relevant protein targets (Section 6.3.1), the active and decoy datasets were screened against the co-crystallised ligand as the query structure. The molecular

fingerprints for the queries, actives and decoys were calculated and the screening performed as described in Section 5.3.1.

6.3.5. Structure-Based Screening Control: UCSF Dock

Each of the protein targets were docked with the UCSF DUD actives and decoys. The validated, high stringency docking parameters were used for docking (Section 6.3.2.2). The Grid scores were read from the .mol2 output file and written to a .csv spreadsheet for the purposes of ranking using the script shown in Appendix 8.24.

Upon producing the UCSF DUD database, Huang et al. (2006) also performed docking screens using Dock v3.5.54 on the targets against the DUD actives and decoys. These energy scores were downloaded and used in this study to produce ROC curves, AUC values, enrichments, and hit diversity analysis (described in Section 6.3.6).

6.3.6. Data Analysis

6.3.6.1. Receiver Operating Characteristic Curves

ROC curves were generated for each screening method for the pharmaceutically relevant targets. These were generated by plotting the fraction of actives found versus the fraction of decoys found (true positive rate versus false positive rate at all thresholds) (Jain and Nicholls, 2008).

AUC values were generated for the ROC curves as the summation of the areas under the curve between neighbouring x data points. The same principle is described in Gagnon and Peterson (1998) and used by GraphPad Prism. The method used here is as follows:

$$AUC = \sum_{k=1}^n (x_k - x_{k-1}) \left(\frac{y_k + y_{k-1}}{2} \right)$$

n = Number of actives plus decoys

x_k and x_{k-1} : The value of x at index k and at $k-1$

y_k and y_{k-1} : The value of y at index k and at $k-1$

6.3.6.2. Hit Diversity Analysis

Hit diversity was determined by 2D fingerprint clustering analysis on the highest scoring hits as well as the pooled actives and decoys for each target. For the hybrid screening approach (Sections 6.3.2 and 6.3.3) and 2D molecular fingerprint screening (Section 6.3.4) the hit lists for the UCSF DUD actives and decoys and the Schrödinger GLIDE decoys were pooled (and re-sorted by screening score).

The top 10 % of hits for each target, for each of the three screening methods (Sections 6.3.2–6.3.3, 6.3.4 and 6.3.5) as well as the pooled actives and decoys libraries, were converted to bitstring fingerprints using the CACTVS chemical information toolkit (Ihlenfeldt et al., 1994). The modified .tcl script used is shown in Appendix 8.25. The number of unique clusters was then calculated using SUBSET 1.0 (cactus.nci.nih.gov/subset/ referred to in Voigt et al. (2001)) at a Tanimoto cutoff level of 0.7.

A diversity index was calculated for the combined active and decoy libraries for each target as a means of assessing the ‘diversity density’ of the libraries. It is the number of representatives at a given Tanimoto cutoff divided by the total number of compounds that were clustered.

6.3.6.3. Enrichment Values

Enrichment values were calculated for each of the three screening methods (Sections 6.3.2–6.3.3, 6.3.4 and 6.3.5). Enrichment is expressed as the true positive rate divided by the false positive rate at a given threshold. Enrichment was calculated as the fraction of actives scored divided by fraction of decoys scored. This was calculated at a threshold of 1 % of decoys scored.

6.4. Results:

6.4.1. ROC Curves

The ROC curves are a useful tool for assessing and comparing screening methods at a glance. The shape of the curve will intimate the overall bias it has for actives over decoys, whether or not a screening method has good early enrichment, and whether or not actives are being scored in clusters.

The ROC curves are shown in Figure 6.6 for the following: The hybrid screening method using the UCSF DUD decoys, The hybrid screening method using the Glide decoys, The 2D Fingerprint screening method (MOLPRINT 2D) using both the UCSF DUD and Glide decoys, and the in-house docking (Dock v6.2, Section 6.3.5) using the UCSF DUD decoys. The ROC curves have a striped red baseline to show a 'no bias' model towards actives or decoys, this baseline represents a trace based on random chance. Traces that are above this baseline indicate a method that scores actives more favourably than decoys. This is more accurately measured by the AUC value where the baseline has a value of 0.5 and a perfect model would have a value of 1. For EGFr it is clear that the 2D fingerprint has a near-perfect plot (Figure 6.6.E), however the AUC value is needed for more accurately comparing traces with similar overall bias but different shapes (such as the 2D fingerprint and DUD docking traces for NA, Figure 6.6.I).

Early enrichment is indicated on the ROC curves as a sharp vertical increase before any significant horizontal progression. The extent of early enrichment is indicated by the size of the initial vertical step. For EGFr (Figure 6.6.E), early enrichment is evident for all the traces; however the enrichment values themselves differ greatly (Section 6.4.3). The 2D fingerprint trace has a much larger initial vertical step than the others and hence a much better early enrichment. For Thrombin (Figure 6.6.L) the 2D fingerprint trace shows a general trend for actives over decoys but the early enrichment is very poor as there is no vertical step in the first part of the trace unlike the hybrid traces.

The ROC curves can often show more than an enrichment value at a particular cutoff. The 2D fingerprint screen of ALR2 (Figure 6.6.B) shows a very sharp vertical jump after a small subset of decoy compounds. However as that small number of decoys is more than 1 % of the entire decoy compounds the enrichment at 1 % is 0 whereas at say 5 % it would be considerably high.

The ROC curves can indicate that a method may be ranking clusters of similar compounds together. For COX-2, FGFr1, NA and SRC (Figures 6.6.C, 6.6.F, 6.6.I and 6.6.K respectively) the 2D fingerprint traces show multiple steep vertical and horizontal jumps throughout the trace compared to the smooth transition of the hybrid and docking screening traces. This is most clear for SRC where there is the large initial vertical step (excellent early enrichment) and another three vertical steps. This shows that there were four clusters of actives with similar scores for SRC using 2D fingerprinting.

The quality of the ROC curve depends largely on the number of actives and decoys being scored. For several targets (COX-2, FXa, HIVPR: Figures 6.6.C, 6.6.G and 6.6.H respectively) large portions of the compounds failed to return a score for 2D fingerprinting, docking or both. The effect this had on the shape of the ROC curve is quite substantial compared to the smooth curve that traces the scores of thousands of compounds for the hybrid screening. For HIVPR (Figure 6.6.H) the hybrid ROC curves for the UCSF DUD decoys and Schrödinger GLIDE decoys both show a smooth transition. A very small amount of early enrichment is evident and there is a clear and significant bias towards actives over decoys. For the docking ROC curve it appears that there is a bias towards actives over decoys. The 2D fingerprint ROC curve appears to favour decoys over actives. However, for these two ROC curves there are sharp vertical and horizontal transitions. This is because only a fraction of the entire active and decoy compounds are represented. Approximately 60 % of the compounds returned a score

for the 2D fingerprint (including only four actives), and only approximately 2 % of the compounds were successfully docked (including only six actives).

The ROC curves show that there was a large degree of variation for the 2D fingerprint screening and the Docking. The 2D Fingerprint screening clearly had the ROC curves with both the strongest and second-to-weakest bias for actives over decoys (Figure 6.6.E, EGFr and Figure 6.6.H, HIVPR respectively; this is also shown by their AUC values, Table 6.2). The Docking curve for P38 (Figure 6.6.I) had the ROC curve with the smallest bias towards actives over decoys that was seen in this study, however the ROC curves for NA (Figure 6.6.I) shows that Dock v6.2 greatly outperformed the other methods for early enrichment as the initial vertical jump is far greater than it is for the other methods (also reflected in Section 6.4.3)

The hybrid screening consistently showed a bias for actives over decoys, even for the targets for which the hybrid method was outperformed by 2D fingerprinting and docking (DHFR, SRC, P38: Figures 6.6.D, 6.6.K and 6.6.J respectively). Reasonable early enrichment was seen for all the targets except DHFR, and the best early enrichment for COX-2 and FGFr1 (Figures 6.6.C and 6.6.F respectively) were achieved by the hybrid screening (also shown in Table 6.3).

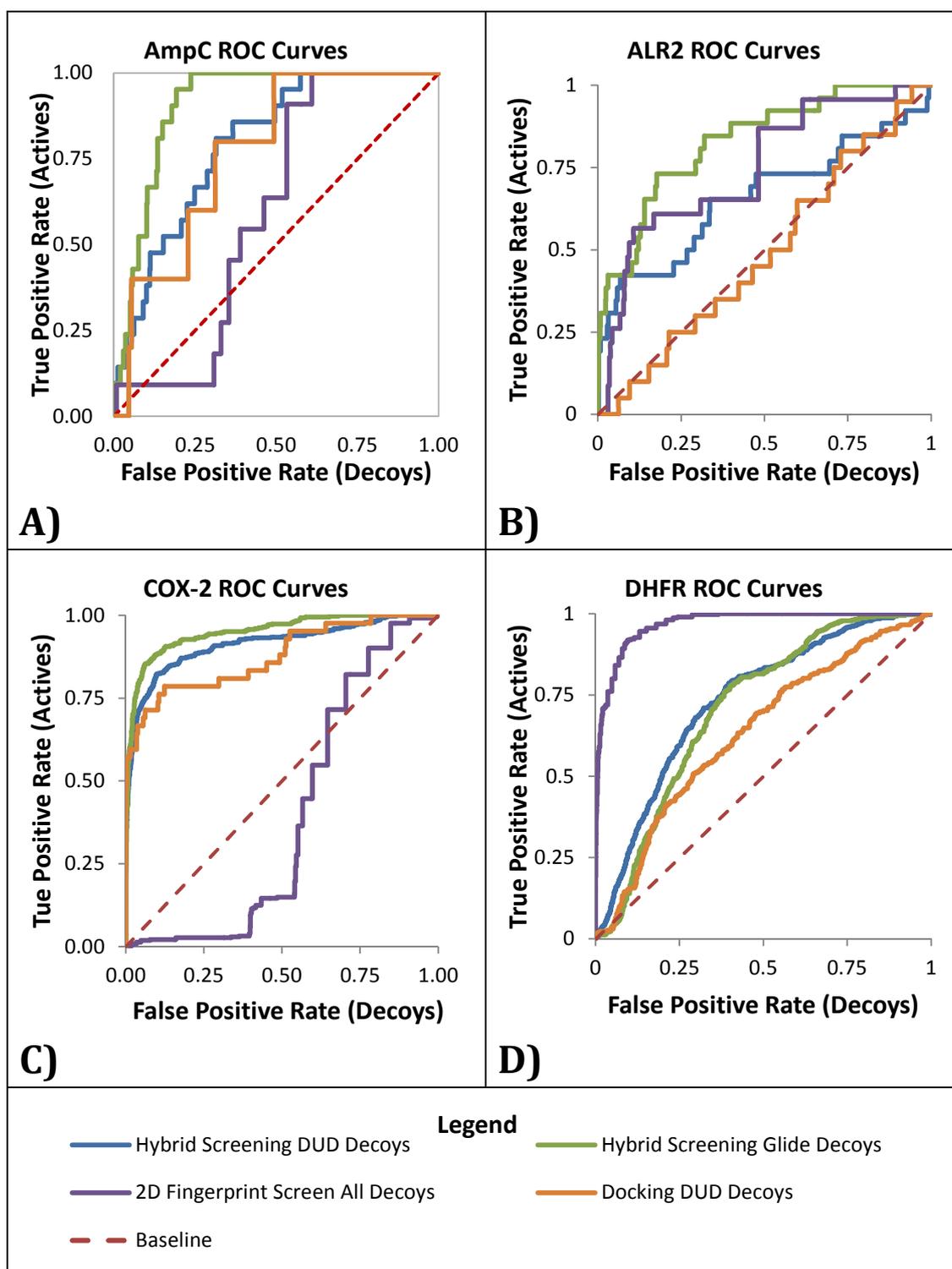


Figure 6.6 (A-D): Comparison of screening methods for AmpC, ALR2, COX-2 and DHFR. ROC curves are shown for each target for the hybrid screening method (Sections 6.3.2 and 6.3.3), Dock v6.2 (Section 6.3.4), and MOLPRINT 2D (Section 6.3.5).

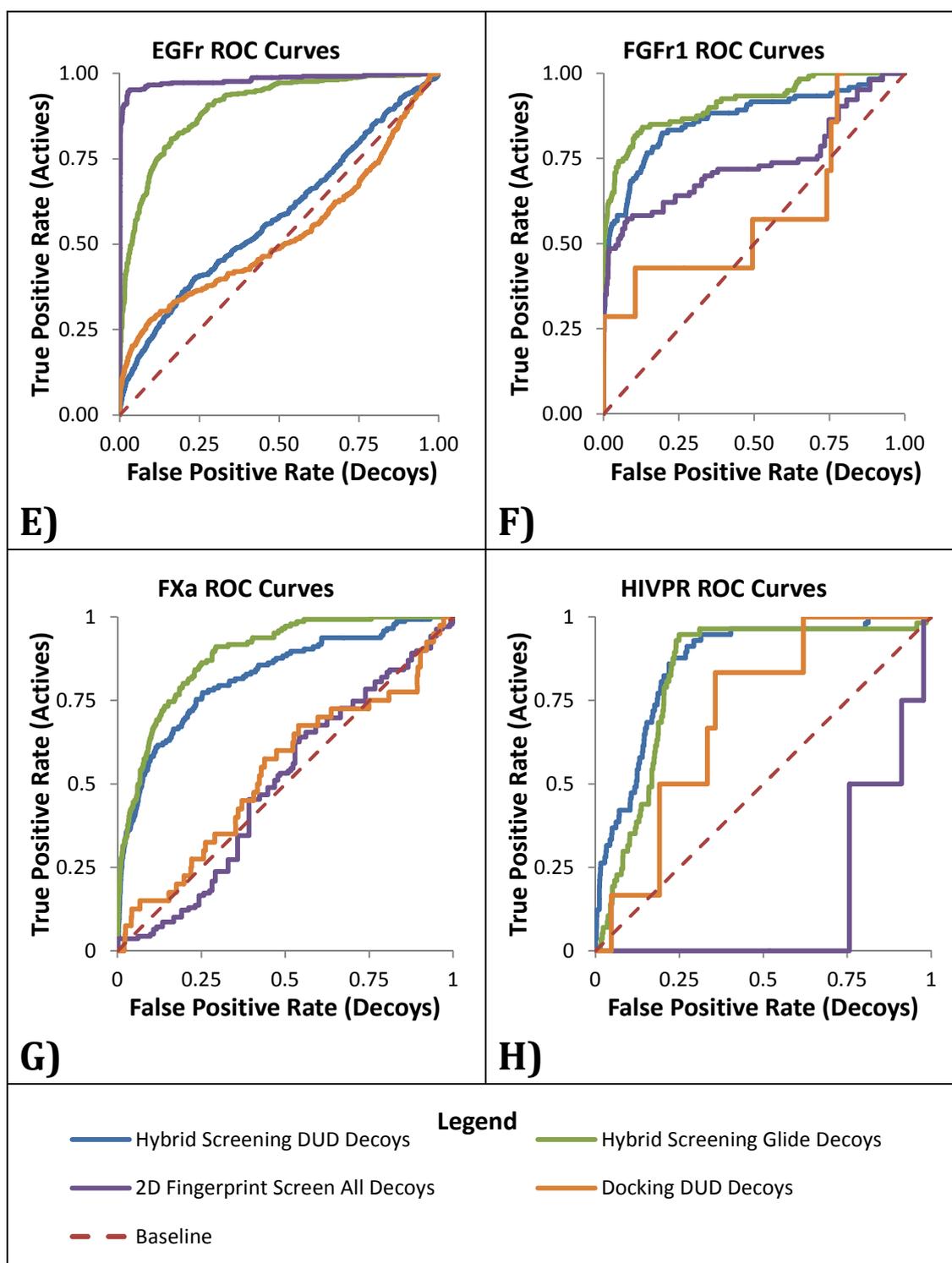


Figure 6.6 (E-H): Comparison of screening methods for EGFr, FGFr1, FXa, and HIVPR. ROC curves are shown for each target for the hybrid screening method (Sections 6.3.2 and 6.3.3), Dock v6.2 (Section 6.3.4), and MOLPRINT 2D (Section 6.3.5).

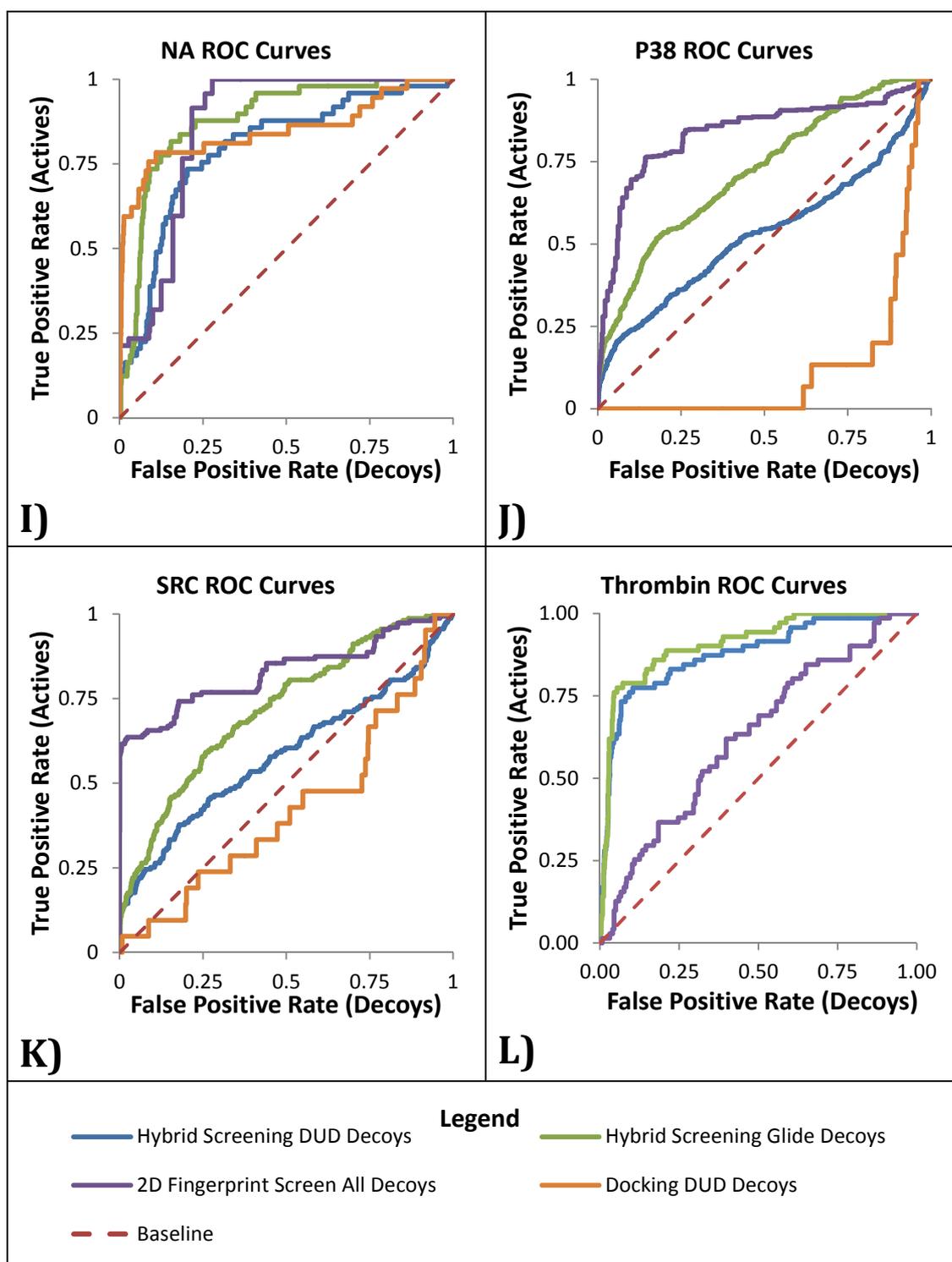


Figure 6.6 (I-L): Comparison of screening methods for NA, P38, SRC and Thrombin. ROC curves are shown for each target for the hybrid screening method (Sections 6.3.2 and 6.3.3), Dock v6.2 (Section 6.3.4), and MOLPRINT 2D (Section 6.3.5), with the exception of the Dock v6.2 ROC curve for Thrombin.

6.4.2. Area Under the Curve Values for ROC Curves

Comparing the AUC values for the hybrid screening (all decoys) and the 2D fingerprint screening (Table 6.2), the hybrid screening method outscored the 2D fingerprint for six of the twelve targets. For three of the twelve targets the 2D fingerprint showed a bias towards decoys over actives (AUC values less than 0.5) whereas the hybrid method shows a bias towards actives for every target. The hybrid screening had better mean and median AUC values than the 2D fingerprint screening and a smaller standard deviation.

Comparing the hybrid screening (DUD decoys) with the in-house docking (Dock v6.2), the hybrid screening outscored the docking for every target except NA. Two of the targets for the docking returned AUC values less than 0.5 indicating a bias towards decoys. It should be noted that the AUC values for the docking results were calculated only on compounds that returned a score.

Comparison of the hybrid screening (DUD decoys) to the structure-based screen data from Huang et al. (2006) (Dock v3.5.54) the hybrid method outscored the docking for every target except DHFR. More importantly the Dock v3.5.54 screen returned AUC values less than 0.5 for four of the targets. While there was considerable difference between the two docking data sets, the mean, median and standard deviation values between the two were similar. The hybrid screening had significantly higher mean and median AUC values than either docking set as well as a smaller standard deviation.

Comparing the hybrid screening results using the UCSF DUD and the Glide decoy libraries, for ten of the twelve targets the AUC values were higher using the Glide decoy library indicating that there was a stronger bias towards actives when using the Glide decoys. The difference was most significant for AmpC, EGFr, P38 and SRC. The mean and median AUC values were larger when using the GLIDE decoys and standard deviation smaller. This indicates that the UCSF DUD decoys are more challenging

decoys where ROC curves and AUC values are the principle measure of a screening method's success.

Table 6.2: Area Under the Curve (AUC) values for ROC Curves from Section 6.4.1.

Best scores for each target enzyme for DUD decoys and All decoys are shown in green, AUC values less than 0.5 shown red. Decoys are colour coded for ease of comparison.

Target Enzyme	Hybrid Screening			MOLPRINT 2D	DOCK v6.2	DOCK v3.5.54*
	Glide Decoys	All Decoys	DUD Decoys	All Decoys	DUD Decoys	DUD Decoys
AmpC	0.908	0.889	0.797	0.599	0.773	0.591
ALR2	0.831	0.745	0.659	0.747	0.490	0.626
COX-2	0.950	0.919	0.914	0.393	0.876	0.826
DHFR	0.710	0.759	0.733	0.969	0.634	0.833
EGFr	0.901	0.597	0.583	0.981	0.530	0.563
FGFr1	0.913	0.888	0.867	0.738	0.619	0.189
FXa	0.886	0.843	0.821	0.499	0.533	0.705
HIVPR	0.830	0.826	0.863	0.149**	0.710	0.438
NA	0.882	0.826	0.796	0.863	0.849	0.685
P38	0.719	0.549	0.531	0.838	0.123	0.499
SRC	0.717	0.600	0.581	0.829	0.423	0.432
Thrombin	0.910	0.899	0.878	0.627	NA	0.501
MEAN	0.846	0.778	0.752	0.744	0.596	0.574
MEDIAN	0.884	0.826	0.797	0.788	0.619	0.577
STD DEV (% of Mean)	9.70	16.0	16.8	23.5	34.4	30.2

*Data from Huang et al. (2006)

**Very small sample size

6.4.3. Enrichment Values

The enrichment values at 1 % for the screens are shown in Table 6.3. Comparing the hybrid screening (all decoys) to the 2D fingerprint screen, the hybrid screen outscored the 2D fingerprint for seven of the twelve targets. Two of the targets failed to show any enrichment at 1 % for the 2D fingerprint, whereas all of the targets showed at least some enrichment at 1 % for the hybrid screen. The mean enrichment was higher for the 2D fingerprint however this is skewed by the few targets that scored extremely highly. The median enrichment at 1 % was better for the hybrid screening and the standard deviation significantly better indicating more consistent results were obtained with the hybrid screening.

Comparing the hybrid screening (DUD decoys) to both the docking screens, the hybrid screen outscored both docking screens for nine of the twelve targets and returned at least some enrichment for every target. The in-house docking (Dock v6.2) failed to generate any enrichment at 1 % for three of the targets (AmpC, ALR2 and P38) and the docking from Huang et al. (2006) (Dock v3.5.54) failed to show enrichment at 1 % for one target—FGFr1. The mean and median values were higher and the standard deviation was lower for the hybrid screening method.

There were some considerable differences in enrichments at 1 % between the in-house Dock v6.2 and the Dock v3.5.54 screens. The Dock v3.5.54 docking showed better mean, median and standard deviation than the in-house docking and outscored the in-house docking for six of the twelve targets (not including Thrombin). It should be noted that calculating accurate enrichment values for Thrombin with the in-house (Dock v6.2) docking was not possible. Only one active and two decoy compounds were successfully docked despite the docking parameters being validated using the co-crystallised ligand. Several attempts were made to relax the docking conditions (such as increasing the allowed 'bumps' during growth, increased orientations, and reduced atom clash distances) with no improvement in the number of actives or decoys docked.

Comparing the DUD and GLIDE decoys libraries with the hybrid screening, higher enrichment values at 1 % were seen in six of the twelve cases for the GLIDE decoys, and no difference was seen for one of the targets. The GLIDE decoys yielded a slightly higher mean, but slightly lower median and a worse standard deviation. Both decoy libraries performed similarly in their ability to assess the quality of the screening method.

Table 6.3: Enrichment values at 1% for all screening methods. Best scores for each target enzyme for DUD decoys and All Decoys are shown in green. Decoys are colour coded for ease of comparison.

Target Enzyme	Hybrid Screening			2D Fingerprint	DOCK v6.2	DOCK v3.5.54*
	Glide Decoys	All Decoys	DUD Decoys	All Decoys	DUD Decoys	DUD Decoys
AmpC	9.52	14.3	14.3	4.76	0	38.5
ALR2	30.8	23.1	23.1	0	0	9.52
COX-2	55.9	50.9	55.4	0.650	7.98	25.1
DHFR	0.973	1.46	1.50	63.0	1.46	29.8
EGFr	30.7	6.53	6.11	90.7	10.1	4.21
FGFr1	56.7	44.2	45.8	41.7	1.67	0
FXa	50.7	24.0	43.8	3.42	4.11	6.85
HIVPR	0	5.26	12.3	0	8.77	1.61
NA	12.2	12.2	12.2	20.4	40.8	14.3
P38	13.0	8.81	8.81	16.1	0	1.32
SRC	12.6	13.2	13.2	58.5	0.630	0.629
Thrombin	14.1	21.1	16.9	1.39	NA	8.33
MEAN	23.9	18.8	21.1	25.1	6.87	11.7
MEDIAN	13.6	13.8	13.8	10.4	1.67	7.59
STD DEV (% of Mean)	82.6	77.8	79.1	118	165	105

*Data from Huang et al. (2006)

6.4.4. Hit Diversity

The diversity at a Tanimoto cutoff of 0.7 is shown in Table 6.4 as the number of unique clusters of compounds found by each screening method in the top 10 % of the hits. The total pooled active and decoys compound libraries for each target were analysed to show the total number of clusters, and hence diversity, in those libraries (Table 6.5).

Comparing the hybrid screening (all decoys) to the 2D Fingerprint screen, the hybrid screen yielded more unique clusters for ten of the twelve targets and an equal number for another. The mean and median values were significantly higher for the hybrid screening and the standard deviations were similar. As diversity was calculated using 2D fingerprints it is expected that MOLPRINT 2D would yield lower diversity.

For the in-house docking (Dock v6.2) of COX-2, FXa and HIVPR, the number of unique clusters was calculated on all of the scored actives and decoys (which totalled less than 10 % of the attempted compounds) and is likely to have significantly affected the scores. Comparing the hybrid screening (DUD decoys) with both the in-house docking (Dock v6.2) and the UCSF DUD docking (Dock v3.5.54) the hybrid screening identified more unique clusters for six of the twelve targets than either of the docking screens (and vice versa). The hybrid screening and in house docking returned the equal highest mean number of clusters and the in-house docking had the highest median score. The standard deviation of the diversity values for each of the targets, for the docking from Huang et al. (2006) was the lowest indicating a more consistent level of diversity. However, the mean, median and standard deviations between the hybrid screening and both docking screens were all very similar.

The diversity of the target's active and decoy libraries (Table 6.5) shows that there is considerable variation in the diversity between these different targets. COX-2, DHFR, EGFR and P38 all have a large number of unique clusters in the library, however these libraries were considerably larger (10,000–17,000 compounds) than the others (1,800–7000 compounds). This gives them a low diversity 'density' as measured shown

by the diversity index (number of unique clusters divided by total number of compounds).

Table 6.4: Number of unique clusters identified by SUBSET 1.0 at a Tanimoto cutoff of 0.7. Clusters calculated for the pooled actives and decoys for each target and top 10 % of hits for each target and screening method. Best scores for each target enzyme for DUD decoys and All Decoys are shown in green. Decoys are colour coded for ease of comparison.

Target Enzyme	Hybrid Screening		2D Fingerprint	DOCK v6.2	DOCK v3.5.54*
	ALL Decoys	DUD Decoys	ALL Decoys	DUD Decoys	DUD Decoys
AmpC	52	27	46	19	20
ALR2	60	31	46	36	25
COX-2	96	97	73	67**	80
DHFR	76	71	39	61	61
EGFr	71	71	71	95	74
FGFr1	35	26	46	37	34
FXa	48	38	20	48**	38
HIVPR	68	25	41	13**	17
NA	56	33	43	36	42
P38	70	61	34	62	55
SRC	62	49	45	44	30
Thrombin	41	36	23	NA	26
MEAN	61.3	47.1	43.9	47.1	41.8
MEDIAN	61	37	44	44	36
STD DEV (% of MEAN)	15.9	22.0	15.1	22.2	20.2

*Data from Huang et al. (2006)

**Incomplete Data

Table 6.5: Total unique clusters (at a Tanimoto cutoff of 0.7) for each target's combined active, DUD decoys, and Glide decoy compounds, and, the Diversity Index. The total number of compounds is shown.

Target Enzyme	Unique clusters	Total number of compounds	Diversity index
AmpC	212	1807	0.117
ALR2	247	2021	0.122
COX-2	344	14715	0.023
DHFR	265	9777	0.027
EGFr	307	17471	0.018
FGFr1	210	5670	0.037
FXa	216	6891	0.031
HIVPR	181	3100	0.058
NA	300	2923	0.103
P38	250	10595	0.024
SRC	225	7478	0.030
Thrombin	192	3528	0.054

6.5. Discussion

The aim of creating the hybrid screening method was to replace a structure-based approach (such as docking) with a far less computationally expensive method as a hit discovery step. The advantages of a structure-based approach (compared to a ligand based approach) that were to be retained were a higher hit diversity and not requiring knowledge of multiple known binders. The advantage of a ligand-based approach to be incorporated was the intrinsic insensitivity to active-site conformations and greatly reduced CPU wall time. Greater consistency of results was also hoped to be gained.

The docking data included here from the original UCSF DUD paper (Huang et al., 2006) shows some significant differences to the in-house docking. All of the target structures used in this study except for DHFR and EGFR were different to those used in the original DUD paper and in the paper it is explained that the site box was greatly increased beyond what would typically be used. This meant that compounds may have been docked to the surface of the enzyme rather than docked in the active site, resulting in no failure to dock/score. This together with the different version of Dock and slightly different docking parameters most likely accounts for the significant difference in docking results seen here. This is not unusual as many papers comparing docking programs observe significant differences depending on program and parameters used (Cross et al., 2009, von Korff et al., 2009) and McGann (2012) reports an improvement in screening with their docking program that uses multiple crystal structures for one target protein compared to a single target.

The hybrid screening method outperformed the docking for both early enrichment and general bias towards actives over decoys. The hit diversity was similar between the docking and hybrid methods. While the aim was to have a screening method that performs at least similarly to docking, the hybrid screening actually worked better than docking for the tested targets.

For the docking itself, seven of the twelve targets failed to dock most of the actives and a significant portion of decoys to the target binding site. The added advantage for the hybrid screening method was that all compounds were scored. This, with the better AUC values and enrichment scores, would indicate a reduced sensitivity to the active site conformation compared to traditional docking.

While 2D fingerprinting is a very simple ligand-based approach it is arguably one of the best where there is very little diversity in the known actives for a target (or where hit expansion is desired). This seems to be the case for DHFR and EGFr where nearly all of the actives score above all of the decoys, and for COX-2 and NA where only a small subset of decoys score above the majority of the actives (albeit enough to significantly affect the enrichment values in Table 6.3). This is exemplified in Figure 6.6 in the ROC curves (most notably for COX-2, FGFr1 and NA) where there are steep steps in the 2D fingerprint curve due to subsets of similar actives returning very similar scores.

2D Fingerprinting showed some of the lowest AUC values (targets HIVPR, AmpC and FXa) as well as the second highest failure to score rate (target HIVPR). The inconsistent results support the notion that 2D fingerprinting is not the best screening method where there is a high degree of diversity in the actives. This poses a problem as it's not always clear prior to hit discovery if a target is capable of binding a diverse or a narrow set of compounds and hence whether or not 2D fingerprint screening is the best approach to take.

The number of unique clusters were generally much higher for the hybrid method than they were for 2D fingerprinting and would indicate, together with the more consistent AUC and enrichment values, that the hybrid screening is not limited to simply finding analogues of the known binder. Interestingly the 2D fingerprinting actually outscored the Hybrid for unique clusters for one of the targets (FGFr1) and scored equally as well for another (EGFr). As 2D fingerprint clustering was the method used for determining hit diversity this was not an expected result. It is possibly caused by differences in the

way MOLPRINT2D matches ligands compared to SUBSET 1.0's bitstring fingerprint matching; one program might not be matching ligands as well as it should. Alternatively the hybrid model for EGFr could simply be unusually selective for a few particular scaffolds.

There was considerable difference between results obtained using the DUD and Glide decoys for the hybrid screening. The DUD decoys consistently yielded lower AUC values. The Glide decoys yielded higher enrichment values for seven of the targets and a slightly higher mean enrichment at 1 %. As mentioned, the DUD decoys are generated based on the DUD actives, where for every active there are 36 decoys with similar chemical properties but dissimilar topology. The Glide decoys are a small set of 1 000 compounds picked to maximise for diversity. The small total number of compounds compared to some of the DUD decoys sets (EGFr's DUD decoy set is nearly 16 000) means that only a few decoy compounds need to score above most of the actives to significantly affect enrichment, thus it was anticipated that the Glide decoys would yield lower enrichments. The DUD decoys did appear to be more challenging for ROC curves. However, they were also slightly more challenging for enrichment values.

While it was mentioned that the hybrid screening had a reduced sensitivity to the active site conformation there were still some limitations. The probe docking was able in most cases to characterise different pockets in the binding site but in the case of DHFR, the crystal structure co-crystallised with methotrexate did not appear to have a conformation that would allow it to bind a large subset of the actives. The subset that was not able to bind had a large fused ring/hydrophobic group in a semi-enclosed area of the active site. Likewise there was limited scope for the probe docking to identify this pocket and highlights an occasional problem related to structure-based methods' sensitivity to active site conformations. Possible solutions to overcome this limitation would include flexible receptor probe docking or molecular dynamics simulations to generate a number of active site conformations.

While the active site conformation was a problem for DHFR and likely the other targets, another limitation of the hybrid method could be its intrinsic inability to identify and flag crucial elements for active site recognition. The fused aromatic ring group is present in all of DHFR's currently known actives but received normal weighting in the hybrid screening. This would have severely affected AUC and enrichment values in the retrospective screen. Contrast this with COX-2, which also had several elements required for recognition (and subsequently a low level of diversity in actives) the hybrid screening performed well, not due to flagging necessary components, but due to not flagging many other optional ones. It is not clear whether or not the fused ring group is actually required for recognition to DHFR; a limitation with this type of retrospective screen is often there is not a great deal of diversity in the known active compounds (Good and Oprea, 2008). While improvements could be made in predicting features and optimising weightings, attempting to incorporate hard limits or necessary features could also lead to models becoming pigeonholed for particular subsets of compounds—a current limitation of typical ligand-based approaches.

For the hybrid screening, only a single pose of each compound was subject to rescoring. This was mainly to simplify the process of compiling scores. This introduces a level of dependency on the known binder that is not ideal, in that the compound is essentially aligned based largely on how it matches to the known binder. It is quite likely that many compounds would bind better to the active site in a different orientation to the one used. One way around this would be a shape-only match to the active site outputting a number of orientations per compound. As the rescoring takes very little CPU wall time this would not significantly impact the total screening time and RMSD clustering could be used to keep the number of orientations manageable.

ROCS has a very simple implementation of its descriptors. For instance the 'rings' descriptor can be anything from planar-aromatic to non-planar and doesn't distinguish between polar and hydrophobic. Often, clusters of rings from probe docking were all on

a similar plane, but this could not be incorporated into the ROCS descriptor. If ROCS were expanded to identify more features like this, more groups and scaffolds could be incorporated into the hybrid screening method for potentially much greater screening accuracy. This would undoubtedly lead to a larger number of descriptors for rescoring but as this step takes very little CPU wall time the impact would be minimal.

A somewhat unique method for clustering was chosen for this work. While there are many different types of clustering algorithms available they all require some input of information. Clustering by hierarchical cluster analysis requires a distance cut-off input or cluster number input and clusters that are identified will change based on this distance. This type of clustering was used in Loving et al. (2009) and Salam et al. (2009) in generating pharmacophores. Centroid-based clustering (*k*-means clustering) also typically clusters given a cluster number input (Lloyd, 1982). The problem with these clustering methods is that the initial number of clusters is not known and even a near guess could give very inconsistent results depending on the probes used and active site size. Distribution-based clustering was considered too user-intensive for use with a standardised procedure. Density-based cluster analysis identifies clusters of similar densities. This poses several issues. Some docking programs (like OpenEye FRED) generate an evenly dense grid of ligand poses during docking. For these docking programs, neighbouring binding pockets using this clustering method could easily be misidentified as a single cluster. Likewise where UCSF Dock is used the density of ligand orientations differs greatly across the active site. While this may seem ideal there is a chance that pockets with an uneven distribution of ligand poses are misidentified as multiple clusters. Neighbouring pockets all with extremely high or low ligand orientation density might also be misidentified as a single cluster.

Instead of clustering into a set number of clusters or by density, the assumed information we could use when clustering involved the binding pocket size. The size of the probes and binding pockets for pharmacophore features are well known and so a

clustering method was developed to group data points into clusters of a particular size. This meant that neighbouring clusters would not be identified as a single cluster and a single binding pocket would not be split into multiple clusters. It did mean that once all the main clusters were identified there were a lot of left-over probe orientations that were returned as a number of clusters that contained as little as a single probe orientation but these were easily identified and removed.

The main advantage to ligand-based approaches is the speed with which it can be carried out. This comes down to CPU wall time and specifically the compounds processed per second. Docking can vary considerably depending on a number of factors: number of attempted orientations, energy minimisation, scoring method, anchor and growth parameters, active chemical matching during orientation, clustering of orientations, size and number of rotatable bonds of the ligand, and size and shape of the receptor binding site.

Docking of the DUD actives and decoys for the targets ranged between approximately 1–30 seconds per compound with most processed within 1–10 seconds. For the initial alignment of compounds (Section 6.3.3.1) using ROCS, approximately 20–40 compounds were processed per second. For the rescoring (Section 6.3.3.2–6.3.3.4) approximately 250–500 compounds were processed per second.

At an average of 5 seconds per compound, docking would take approximately 58 days of CPU wall time to screen 1 million compounds compared to approximately 24 hours using the hybrid method (assuming average initial alignment of 30 compounds per second, rescoring at 375 compounds per second for 20 descriptors). Generating multi-conformers of the compound library with OMEGA would take approximately 8 days of CPU wall time but once generated can be re-used for multiple targets. This indicates a nearly 60-fold reduction in CPU wall time for the hybrid method compared to docking and is a sufficient reduction to remove the necessity for a high performance computing cluster.

Furthermore, FastROCS—a build of ROCS on high performance Graphics Processing Unit (GPU)-accelerated computing—has shown the ability to process between 400 000 and 1.4 million compounds per second depending on the GPU used (not published, eyesopen.com/fastrocs).

7. Discussion

7.1. Stability of *PfOPRTase* and *HsUMPS* Gene in *E. coli*

Large quantities of recombinant protein are needed to carry out drug discovery.

Recombinant protein is used for biochemical assays and structural determination experiments that are required to facilitate hit discovery through to lead development.

The *PfOPRTase* gene was confirmed to be unstable in *E. coli* BL21(DE3) but not *E. coli* PMC103. Krungkrai et al. (2004a) alludes to this instability with explicit mention of fresh transformations and early induction but does not suggest a cause. This instability results in extremely poor levels of expression which significantly impacts drug discovery. The instability was identified in the current study to most likely be caused by a stem-loop that is present in the gene as *E. coli* PMC103 is engineered to be stable with these features (Doherty et al., 1993) but *E. coli* BL21(DE3) is not.

HsUMPS exhibited extremely poor levels of expression compared to the individually expressed domains *HsOPRTase* and *HsODCase*. A cruciform-like feature (RNA/DNA secondary structure feature) was identified in the domain linker region of *HsUMPS* that is not present in the individual domains. This is likely to be responsible for the heavily reduced expression efficiency as these features can be cleaved by endonucleases in *E. coli* (Taylor and Smith, 1990, Iwasaki et al., 1991).

These insights allow the genes to be re-engineered (altering the codon usage to change the DNA sequence but not the translated sequence) to be stable in *E. coli* BL21(DE3) to greatly expedite drug discovery for these targets.

7.2. OPRTase Drug Discovery Pipeline

Crystal structures are used in drug discovery for a number of applications. One major application is in structure-based virtual screening for hit discovery. Another very important application is in structure-guided lead development which involves co-crystallising the lead compound with the target enzyme. Crystal structures are also

required for the development of some 3D QSARs. For *HsOPRTase* the structure 2WNS (Moche et al., 2009) provided a suitable platform for virtual screening in this study. The high quality of the crystal structure, the fact that it was co-crystallised with a ligand, and the high expression and purity achieved are all promising for this target during future lead development as there is a reasonable chance that the conditions could also be used for co-crystallisation with a lead compound.

For *PfOPRTase* however an experimentally solved structure was not available at the time of performing the virtual screening. The *PfOPRTase* crystal structure that was only recently released is an apo structure with a resolution of 2.60 Å. The active site flexible loop adopts several conformations, all of which have more 'open' conformations that do not reflect the closed ligand-bound conformation described in Henriksen et al. (1996) when compared to other ligand-OPRTase structures. The *PfOPRTase* crystal structure would not be ideal for virtual screening without refinement such as MD simulations to obtain a predicted ligand-bound conformation of the active site flexible loop. It is therefore desirable to test this crystallisation condition for producing a ligand-*PfOPRTase* co-crystal or finding a condition that can. If this is not possible then structure-guided lead development of *PfOPRTase* inhibitors could not occur; lead development would be limited to the use of SARs/QSARs. The crystallography work performed in this study did not result in diffraction quality crystals. It did however demonstrate that a sample of *PfOPRTase* at a purity of 90–95 % (measured by Coomassie stained SDS-PAGE) can still produce protein crystals. It also identified several conditions that are worth further investigation.

A homology model was generated for *PfOPRTase* and validated by docking (prior to the release of the crystal structure). This was used in virtual screening by docking and led to the identification of several hit compounds for *PfOPRTase*. Comparison of the homology model to the recently released crystal structure of *PfOPRTase* showed the model to be reasonably accurate (RMSD 2.58 Å). Docking validation confirmed that the

active site was likely to be in the correct ligand-bound conformation. Superposition of the homology model and the crystal structure monomers confirmed that the flexible loop in the active site was significantly different in the homology model compared to the crystal structures. The successful identification of inhibitors conducted in Chapter 5 demonstrates the usefulness of homology models and docking validation in structure-based hit discovery. The homology model itself could also be used in future screening experiments until MD experiments on the crystal structure can produce a ligand-bound active site conformation.

Prior to this study no known successful virtual screening studies were performed on either the *P. falciparum* or *H. sapiens* OPRTase. A handful of structural analogues of the products and substrates were characterised (Krungkrai et al., 2004a, Scott et al., 1986, Witte et al., 2006, Zhu et al., 2013) but this alone would provide limited data for use in a SAR. Nine of the 19 compounds (47 %) that were tested in this study resulted in 75 % inhibition or better, in the micromolar range, against at least one of the OPRTases. The Soichet Laboratory at UCSF have performed a considerable number of virtual screening projects. As a comparison, for their numerous studies they consider a hit rate of 5 % to be low (Powers et al., 2002) and consider one screening project with hit rate of 35 % (Doman et al., 2002) to be high. They biochemically test up to hundreds of compounds from virtual screening experiments and as such used a stringent definition of a hit for these studies— IC_{50} less than 100 μ M. As outlined in Section 1.6 compounds with activity in the high micromolar range, and in some cases low millimolar range can still be successfully developed with H2L strategies.

Hit rates are extremely difficult to compare as the binding affinities are largely target-dependent (Drwal and Griffith, 2013). In any case it is more important for virtual screening to identify as many structurally diverse hits as possible rather than identifying potent inhibitors as binding affinity is optimised in H2L and lead-development stages (Scior et al., 2012).

The hit rates in this study are promising for the OPRTases as drug candidates, considering only 19 compounds were biochemically tested here compared with 56 and 365 compounds in Powers et al. (2002) and Doman et al. (2002) respectively. It provides justification for further virtual screening experiments on a larger compound library but more importantly a higher throughput biochemical screening of virtual screening hits.

There was significant structural diversity among the hit compounds in this study and even some overlap with some compounds also inhibiting the ODCases. Specificity was seen in many of the compounds; in particular L268-0351 and 4470-0385 which showed specificity for *Pf*OPRTase and compound 7009-0959 which showed high potency and specificity for *Hs*OPRTase. Structural diversity and specificity are both important factors in hit discovery for their roles later in H2L strategies, lead selection and development. These two factors contribute greatly in generating SARs and understanding what chemical traits will create a more potent and specific inhibitor.

7.3. ODCases as Drug Targets

Due to the unique nature of the enzyme's catalytic mechanism, this protein has been studied extensively. As such there are a large number of structures available for this well researched enzyme. ODCases are generally well expressed in recombinant systems and easily purified. The method for producing *Pf*ODCase described in Menz et al. (2002) was replicated without issue and the construct for *Hs*ODCase produced a large amount of pure active protein with little optimisation required. The standard spectrophotometric activity assays are easy to perform and the reaction progress kinetic analysis assay developed in this study allows kinetic characterisation to also be rapidly and cheaply performed on these enzymes.

The large amount of structural data provides an ideal platform for virtual screening experiments as well as H2L and lead development methods that rely on protein-ligand crystallography. The nature of the assays allows a large throughput approach to be

taken to screening the virtual hits during hit discovery, and rapid kinetic characterisation of compounds in later stages. It is for these reasons that the ODCases are extremely good drug targets.

Studies have identified structural analogues of ribonucleotides for inhibitors of ODCases (Bello et al., 2007, Langley et al., 2008, Heinrich et al., 2009, Purohit et al., 2012, Crandall et al., 2013). One study performed virtual screening, identifying 14 inhibitors (Takashima et al., 2012). Unfortunately most of these compounds are simple scaffolds surrounded by highly polar chemical groups. For instance one compound is a benzene ring with four nitro groups and a carboxyl group. As such there is a lack of structurally unique drug-like inhibitors and scaffolds for the ODCases. Any ligand-based screening models and SARs produced from the previously discovered compounds would be limiting.

The ODCases also had a reasonable hit rate of confirmed inhibitors from virtual screening with eleven of the 19 compounds tested showing approximately 20 % inhibition or better against at least one of the ODCases. If using the Soichet Laboratory's stringent definition of a hit as outlined in Section 7.2 (IC_{50} less than 100 μ M) the hit rate could be as high as 5 and 15 % for *Pf*ODCase and *Hs*ODCase respectively with three of the 19 compounds possibly having IC_{50} values less than 100 μ M against one or both ODCases. However, comparison of the inhibition constants for the ODCase inhibitors (Section 5.4.1) showed final hit rates (including the 2D fingerprint hits) of 4 and 8 % (for *Pf*ODCase and *Hs*ODCase respectively) for compounds with inhibition constants under 100 μ M, and 8 and 16 % (for *Pf*ODCase and *Hs*ODCase respectively) for compounds with inhibition constants under 200 μ M.

Specific inhibitors for each ODCase were identified, several compounds appeared to be quite potent and two showed complete inhibition at 5 mM. There was also a high level of diversity among the compounds, none of which were structural analogues of ribonucleotides. The small scale H2L hit expansion identified a further two inhibitors

and an alternative substrate. The success of this last step provides justification for a large scale hit expansion to be performed on the confirmed hits. This has the benefit of increasing the list of potential lead compounds and producing the data set that is needed for a higher quality SAR/QSAR.

Kinetic characterisation quantified the specificity and potency of confirmed inhibitors of ODCase. In particular, two compounds—8008-2619 and 4470-0386—were highly specific for the *Pf*ODCase and several were highly specific for *Hs*ODCase. This data is important in SARs/QSARs during lead development for predicting chemical features that confer specificity. All of the compounds that underwent kinetic characterisation had kinetic constants in the range that would be suitable for H2L development. This represents a significant advancement in the search for novel ODCase inhibitors.

Compound C197-0379 was the most potent inhibitor, showing moderate inhibition for both *Plasmodium* and human ODCases. This non-specific binder had no Lipinski or bioavailability violations. The potency and drug-likeness currently make it the best lead candidate. This compound could form the scaffold for new drugs to treat malaria, as well as the various applications to blocking the human *de novo* pyrimidine biosynthesis pathway such as autoimmune diseases, some viruses, and some forms of cancer that were outlined in Section 1.2.2. The non-specific inhibition suggests that it may also show activity against other protozoan and bacterial diseases—also listed in Section 1.2.2—by inhibiting ODCase in these organisms.

7.4. New Insights into the Structural and Catalytic Properties of the ODCase Active Site

Langley et al. (2008) describes some features in the ODCase crystal structures that could be a means for cooperativity in the catalytic mechanism of ODCase. In short, the study mentions that structural rearrangements of the $\beta\alpha 5$ -loop (at the dimer interface) in one monomer are mirrored in the other. Langley et al. (2008) suggests that binding of OMP to an open 'loop-out' active site creates a closing 'loop-in' movement with a

conformational change in the other monomer. The study also points to the Miller and Wolfenden (2002) review which notes significant differences in binding affinities of competitive inhibitors despite 'remarkable similarity' in the crystal structures and Langley et al. (2008) suggests that there may be active site conformations yet to be observed by crystallography.

Cooperativity has not been reported in other papers and the substrate kinetics performed in this study do not show any significant degree of cooperativity occurring (the inhibition kinetics do however show interconnectivity of the active sites). The Michaelis-Menten constant for *Pf*ODCase produced in this study correlates well with Krungkrai et al. (2004a). The K_m^{app} for the *E. coli* expressed *Hs*ODCase here was similar to the *Pf*ODCase but significantly higher than the K_m for eukaryotically expressed *Hs*ODCase produced by Yablonski et al. (1996). The N-terminal His-GST tag may be partly responsible for the reduced substrate binding affinity. The N-terminus is at the opposite end of the monomer to the active site and would not be directly affecting the active site, rather it would have to be acting in an allosteric fashion by subtly affecting the overall conformation of the enzyme.

The modes of inhibition of many of the inhibitors identified in Chapter 5 provide new insights into the structure and catalytic function of ODCase. The virtual screening carried out in Chapter 4 was designed to predict competitive inhibitors of the *P. falciparum* and *H. sapiens* ODCases. However only one of the compounds best fit the Michaelis-Menten model for competitive inhibition. The majority of compounds best fit the model for uncompetitive inhibition and several for mixed-mode inhibition. One of the mixed-mode inhibitors had an alpha value that indicated it was almost purely a noncompetitive inhibitor. There is no allosteric site on ODCase large enough to fit these inhibitors without a significant overall conformational change of the enzyme. It is also unlikely that multiple predicted competitive inhibitors would bind to an allosteric site by chance. It was therefore deduced that the compounds must be binding to one of the

two active sites on the homodimer and acting in an allosteric fashion through subtle conformation changes via interconnectivity between the two active sites. The proximity of the active sites is certainly conducive to conformation change in one active site affecting the conformation of the other. It would also support the observations by Langley et al. (2008) that led to the cooperativity hypothesis.

The noncompetitive inhibitor shows that inhibitor binding to one active site induces a conformation change that blocks binding (or catalysis) of OMP in the other. More significantly, the uncompetitive inhibitors confirm that OMP binding to one active site does induce a conformation change in the other active site (which enables binding of the inhibitor). This therefore leads to either: a conformation change occurs in the non-catalysing active site during catalysis (which is blocked by the presence of the uncompetitive inhibitor), or that binding of the uncompetitive inhibitor affects the conformation of the catalysing-active site that blocks catalysis. This confirms the interconnectivity (but not specifically cooperativity) proposed by Langley et al. (2008) between the ODCase dimer active sites and suggests that the active sites adopt different conformations during binding and possibly catalysis of OMP. It also suggests that an alternating rather than a synchronised mode of binding/catalysis may be occurring. X-ray crystallography of the inhibitor bound structures would need to be conducted to confirm that no allosteric site exists and to better visualise the induced conformation changes. Co-crystallisation of the ODCase with both OMP and an uncompetitive inhibitor would help in identifying the catalytic mechanism of ODCase and visualising the conformation changes that occur.

There have been studies that have identified alternative substrates for ODCase by producing various OMP analogues by chemical substitutions (Shostak and Jones, 1992) with the intention of understanding more about the reaction mechanism of the enzyme. To date no known alternative substrates exist for OMP that are not ribonucleotide analogues, other than compound 4049-0191 identified in this study. While the results

need to be confirmed with mass spectrometry, the results could hold great significance to understanding the reaction mechanism of ODCase. Most of the work to date suggests that the decarboxylation occurs through a local electrostatically-induced destabilisation of the carboxyl group and a stabilisation of the transition state. The destabilising forces are offset by a strong hydrogen bond network between the phosphoribose moiety of OMP and the active site which produces a favourable Gibbs free energy of binding (Goryanova et al., 2011, Amyes et al., 2012). Binding of the phosphoribose has also been explicitly mentioned as being an important factor in the conformational change needed for catalysis (Desai et al., 2012). Compound 4049-0191 contains no phosphate or ribose groups however and likely cannot mimic the enthalpy-driven binding of the ribonucleotide moiety of OMP. Instead this alternative substrate demonstrates a more entropy-driven binding mode that still induces the catalytic active conformation of the ODCase active site, although it has a much lower binding affinity (observed in the far greater K_m^{app} for ODCase and compound 4049-0191).

It was theorised that the nitro-group of 4049-0191 was being removed in the reaction with ODCase. The atomic charges of the oxygen atoms in the nitro group are very similar to those in a carboxyl group which lends itself to the local electrostatic destabilisation catalytic mechanism. The removal of the group to produce a stable gas and the subsequent protonation of the planar ring anionic carbon can also occur in an analogous fashion to that in the decarboxylation reaction of OMP. The orientation of the ligand and conformation of the active site residues however is not known. Elucidating this information could be key to confirming the reaction mechanism of ODCase and conformational changes that occur during binding and catalysis. It would also shed some light on the interconnectivity of the active sites.

7.5. Applications of the Novel Hybrid Screening Protocol

The hybrid screening method developed in this study was designed to be used against any protein target but the intention was to develop the method to overcome the

shortcomings of docking. There were significant drawbacks with regards to the docking carried out in Chapter 4 (despite the success in identifying inhibitors in Chapter 5). The screening took a long time (a typical amount of time for the program that was used (Moustakas et al., 2006)). The hit expansion strategy employed in Chapter 5 identified inhibitors that were not flagged in the docking. This confirmed that false negatives were occurring in the original docking screen. This is to be expected to a certain degree. However, Chapter 6 illustrated that it can be quite extensive for certain targets with some protein targets having most or nearly all of the known active compounds failing to dock or scoring lower than most decoys. This trend was also seen in the data from Huang et al. (2006) with AUC values as low as 0.189.

The hybrid method requires knowledge of only a single known binder and a crystal structure, as is the case with docking. The method has been shown to be much faster than docking. It achieves, on average, a better bias towards actives, greater early enrichment and more consistent results. The method has been shown to perform similarly to docking and greatly outperforms a simple ligand-based approach in terms of hit-diversity.

Many virtual screening approaches claiming to be a hybrid method are simply parallel or sequential screening approaches (Tan et al., 2008, Swann et al., 2011, Svensson et al., 2012). True hybrid approaches include structure-based pharmacophores (Carlson et al., 2000, Chen and Lai, 2006, Barillari et al., 2008, Cross et al., 2012, Loving et al., 2009, Salam et al., 2009) and ligand-guided docking (McGann, 2012, Pinto et al., 2011), each with their own merits. The method described here involves structure-derived pharmacophore descriptors (which in itself is not particularly novel) followed by the novel, non-penalising scoring methodology. The novel scoring method was specifically designed to address the significant flaw with pharmacophores. Specifically, the large tradeoff that exists between hit diversity and false positive rate (Scior et al., 2012). This is not completely addressed using structure-derived pharmacophores. The hybrid

method presented in this study performs the bulk of the virtual screening using ligand-based software to significantly reduce CPU wall time—an issue that remains with ligand-guided docking programs.

The rescreening of the human and *P. falciparum* OPRTases and ODCases can now occur using the hybrid screening method. As the hybrid screening is much faster than docking it creates an opportunity to perform virtual screening on a much larger compound library such as the Available Chemicals Directory and larger subsets of the UCSF ZINC database. This will allow a more thorough coverage of the chemical space for a shorter or similar amount of CPU wall time. Rescreening of the OPRTases and ODCases with the hybrid method should continue to yield diverse hits despite moving away from using the purely structure-based docking approach.

Another possibility for rescreening of the OPRTases and ODCases is to perform parallel screening. The hybrid screening would occur on the ChemDiv subset of the UCSF ZINC library as it did for docking in Chapter 4. Compounds that score highly in both the old docking results and the new hybrid results would then be screened in biochemical assays. This approach typically results in a higher hit rate of biochemically screened compounds but would not address the high rate of false negatives that can sometimes be seen in docking results (Drwal and Griffith, 2013, Houston and Walkinshaw, 2013). The hybrid screening was developed partly to address limitations when screening the OPRTases and ODCases. However this method is not limited to screening these enzymes and could be used to find inhibitors of some other recently identified therapeutic targets.

The hybrid screening results were more consistent than the docking or 2D fingerprint screening. The hybrid screening greatly outperformed both the 2D fingerprint screening and docking for several target enzymes. In particular screening against Thrombin and Factor Xa (two serine proteases), HIV protease (an aspartyl protease), and Fibroblast growth factor receptor kinase (a tyrosine kinase). The hybrid method

also outperformed docking for the other kinases tested in Chapter 6—Epidermal Growth Factor receptor kinase (EGFr), P38 mitogen activated protein kinase (P38), and Tyrosine kinase SRC (SRC). This is very promising for the application of this hybrid screening method in drug discovery as proteases and kinases make up a significant portion of therapeutically relevant target proteins.

Serine proteases are already popular targets for the treatment of various diseases such as HIV and Hepatitis C. They have also been suggested as potential targets in malaria (Alam, 2014). The HtrA family of serine proteases have been identified as potential targets for cancer (Chien et al., 2009) and pathogenic bacteria (Skorko-Glonek et al., 2013). Another serine protease, TMPRSS4, has also been identified as a potential target for cancer (de Aberasturi and Calvo, 2015). Aspartyl proteases are an important target for HIV with the famous example of HIV protease used in Chapter 6. Plasmepsin V is another aspartyl protease identified as being a potential therapeutic target for malaria (Boddey et al., 2010). A family of Secreted Aspartic Proteases have also been identified as antifungal targets for *Candida albicans* (Goldman et al., 1995). Protein kinase inhibitors are extremely important in the pharmaceutical industry as anticancer and antitumor drugs (Anafi et al., 1993, Druker et al., 1996, Meydan et al., 1996, Strawn et al., 1996). Recently identified tyrosine kinase anticancer targets include MERTK (Schlegel et al., 2013) and ROS receptor tyrosine kinase (El-Deeb et al., 2011). The results described in Chapter 6 suggest that these relatively new targets for a range of significant diseases and infections would be excellent candidates for virtual screening using the hybrid method presented here.

7.6. Future Directions

The information obtained in Chapter 3 will allow the *Pf*OPRTase and *Hs*UMPS vectors to be reengineered for better stability and expression in *E. coli*. This will greatly expedite crystallographic work on these enzymes. Crystal structures of these two enzymes in particular are of great importance. There is no solved structure showing the ligand-bound conformation of *Pf*OPRTase. Very little is known about the quaternary structures of the bifunctional human UMPS and the heterotetramer that forms between *P. falciparum* OPRTase and ODCase.

Virtual screening in Chapter 4 identified a number of potential inhibitors of *P. falciparum* and human OPRTase and ODCase. Nineteen of these compounds were tested biochemically. The majority of the potential inhibitors were not screened and a larger scale biochemical screening of these compounds should result in a larger pool of confirmed inhibitors of the OPRTases and ODCases when taking into account the reasonably high hit rates seen in the Chapter 5 results.

The identification of inhibitors of *P. falciparum* and *H. sapiens* OPRTase and ODCase in Chapter 5 has expanded the avenues of research into finding a lead compound for treating malaria, cancer, and a range of other diseases. Now that there is a pool of confirmed active compounds for these targets a large scale 'hit expansion' H2L strategy can be conducted. The pool of confirmed inhibitors will also allow SAR models to be produced for the OPRTases and ODCases to aid hit expansion as well as other H2L strategies such as isosteric replacement and hit evolution.

Chapter 5 identified some unexpected modes of inhibition. The results suggest a conformational interconnectivity between the active sites on the ODCase dimer. Following on from this would entail crystallographic experiments or MD simulations of the enzyme-inhibitor and even enzyme-inhibitor-substrate complexes to visualise this interconnectivity such that it may be exploited in the later lead development stage of

the drug discovery pipeline. Enzyme-inhibitor structures will also aid H2L strategies through the use of 3D-QSARs as well as structure-guided lead development.

Mass spectrometry experiments should be conducted on compound 4049-0191 to confirm that it is the nitro group that is being removed. This can then be followed up with crystallographic work or MD simulations to visualise the orientation and active site conformations. Understanding how this compound acts as an alternative substrate may be key to finally solving the reaction mechanism of ODCase that numerous crystal structures and QM/MM simulations thus far have not been able to fully explain.

A novel hybrid screening method was described in Chapter 6. This new method can be used to rescreen the OPRTases and ODCases on a much larger library. The results can also be used as part of a parallel screening approach with the Chapter 4 results. The hybrid screening method was designed to work on any protein target. It seems to perform particularly well against kinases and proteases; it could be used against a very large number of therapeutically relevant proteases and kinases such as those outlined in Section 7.5. It was however more consistent than docking and 2D fingerprint screening overall and is not limited to any particular targets.

Appendices

Appendix 8.1: High-Fidelity PCR Step Lengths and Temperatures for the Gene *H. sapiens* UMPS, and its domains OPRTase and ODCase.

Step	<i>H. sapiens</i> UMPS		<i>H. sapiens</i> OPRT		<i>H. sapiens</i> ODC	
	Step Temperature (°C)	Step Length	Step Temperature (°C)	Step Length	Step Temperature (°C)	Step Length
1: Initialisation	94	3 min	94	3 min	94	3 min
2: Denaturation	94	30 s	94	30 s	94	30 s
3: Annealing	63	30 s	63	30 s	63	30 s
4: Elongation	72	90 s	72	60 s	72	60 s
Number of cycles of steps 2 to 4	25 cycles		25 cycles		25 cycles	
5: Final Elongation	72	8 min	72	8 min	72	8 min
6: Hold	8	∞	8	∞	8	∞

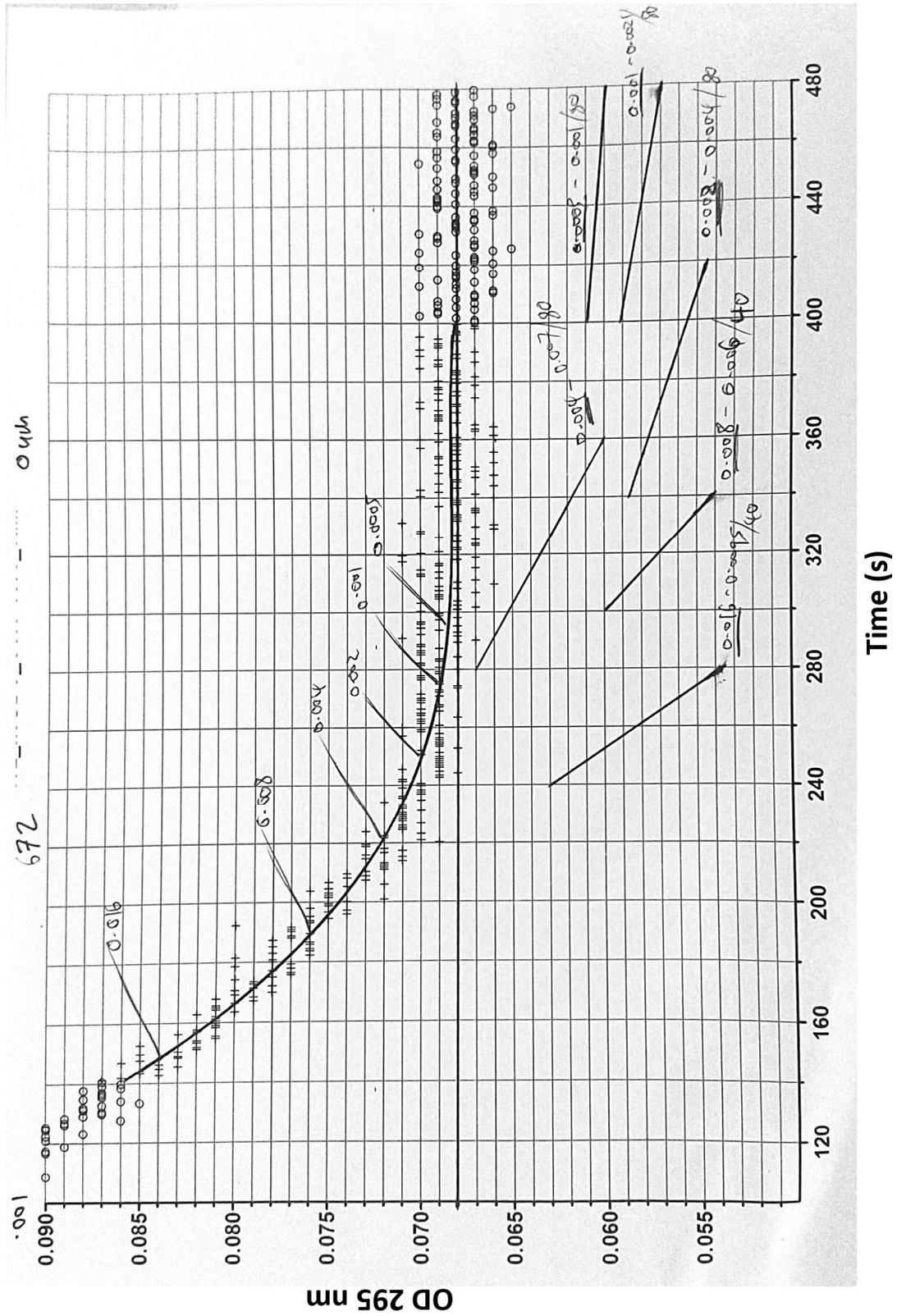
Appendix 8.2: Analytical PCR Step Lengths and Temperatures for the Gene *H. sapiens* UMPS, and its domains OPRTase and ODCase.

Step	<i>H. sapiens</i> UMPS		<i>H. sapiens</i> OPRT		<i>H. sapiens</i> ODC	
	Step Temperature (°C)	Step Length	Step Temperature (°C)	Step Length	Step Temperature (°C)	Step Length
1: Initialisation	94	3 min	94	3 min	94	3 min
2: Denaturation	94	30 s	94	30 s	94	30 s
3: Annealing	63	30 s	63	30 s	63	30 s
4: Elongation	72	90 s	72	60 s	72	60 s
Number of cycles of steps 2 to 4	35 cycles		35 cycles		35 cycles	
5: Final Elongation	72	8 min	72	8 min	72	8 min
6: Hold	8	∞	8	∞	8	∞

Appendix 8.3: Chromatography Buffers A and B for the Recombinant Proteins *HsUMPS*, *HsOPRTase*, *HsODCase*, *PfOPRTase* and *PfODCase*.

Protein	Buffer A	Buffer B
<i>HsUMPS</i> , <i>HsODCase</i> , and <i>PfODCase</i>	50 mM Tris-HCl (pH 8.0) 300 mM NaCl	50 mM Tris-HCl (pH 8.0) 300 mM NaCl 500 mM Imidazole
<i>HsOPRTase</i> and <i>PfOPRTase</i>	50 mM Sodium Phosphate (pH 8.0) 300 mM NaCl	50 mM Sodium Phosphate (pH 8.0) 300 mM NaCl 500 mM Imidazole

Appendix 8.4: Example of Reaction Progress Kinetic Analysis

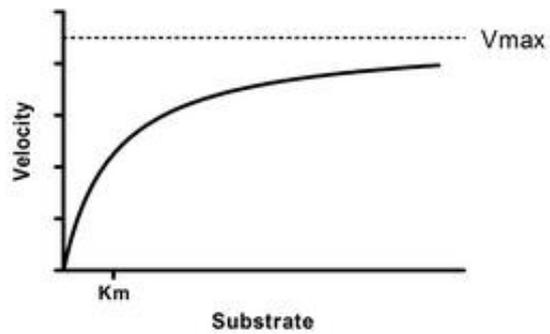


Appendix 8.5: Michaelis-Menten Kinetics Model

$$y = \frac{V_{max} \times x}{(K_m + x)}$$

V_{max} : The maximum rate of the reaction (same units as y)

K_m : The Michaelis-Menten constant (same units as x)



Appendix 8.6: Michaelis-Menten Competitive Inhibition Model

$$K_{mApp} = K_m \times \left(\frac{1 + [I]}{K_i} \right)$$

$$y = \frac{V_{max} \times x}{(K_{mApp} + x)}$$

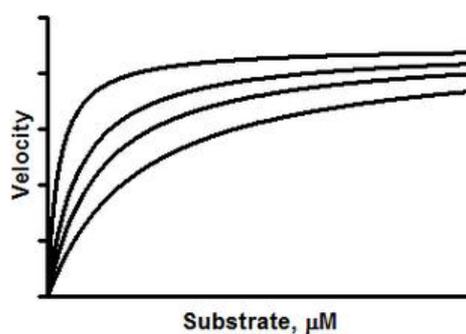
V_{max} : The maximum rate of the reaction (same units as y)

K_m : The Michaelis-Menten constant (same units as x)

K_i : The inhibition constant (same units as I)

I : Concentration of inhibitor (non-shared value, varies for each curve)

K_{mApp} : The K_m observed for a particular curve (non-shared value)



Appendix 8.7: Michaelis-Menten Noncompetitive Inhibition Model

$$V_{maxApp} = \frac{V_{max}}{\left(\frac{1 + [I]}{K_i}\right)}$$

$$y = \frac{V_{maxObs} \times x}{(K_m + x)}$$

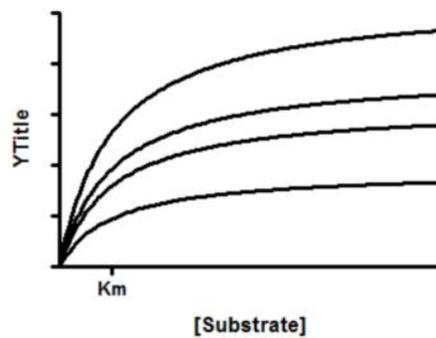
V_{max} : The maximum rate of the reaction (same units as y)

K_m : The Michaelis-Menten constant (same units as x)

K_i : The inhibition constant (same units as I)

I : Concentration of inhibitor (non-shared value, varies for each curve)

V_{maxApp} : The V_{max} observed for a particular curve (non-shared value)



Appendix 8.8: Michaelis-Menten Uncompetitive Inhibition Model

$$V_{maxApp} = \frac{V_{max}}{\left(\frac{1 + [I]}{aK_i}\right)}$$

$$K_mApp = \frac{K_m}{\left(\frac{1 + [I]}{aK_i}\right)}$$

$$y = \frac{V_{maxApp} \times x}{(K_mApp + x)}$$

V_{max} : The maximum rate of the reaction (same units as y)

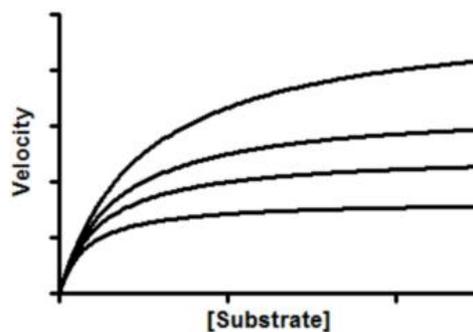
K_m : The Michaelis-Menten constant (same units as x)

I : Concentration of inhibitor (non-shared value, varies for each curve)

V_{maxApp} : The V_{max} observed for a particular curve (non-shared value)

K_mApp : The K_m observed for a particular curve (non-shared value)

aK_i : The inhibition constant (same units as I). It is the product of K_i (which is very high as uncompetitive binders don't bind the enzyme) and a (which is very low). This inhibition constant is also known as K'_i .



Appendix 8.9: Michaelis-Menten Mixed Inhibition Model

$$V_{maxApp} = \frac{V_{max}}{\left(\frac{1 + [I]}{\alpha \times K_i}\right)}$$

$$K_{mApp} = \frac{K_m \times \left(\frac{1 + [I]}{K_i}\right)}{\left(\frac{1 + [I]}{\alpha \times K_i}\right)}$$

$$y = \frac{V_{maxApp} \times x}{(K_{mApp} + x)}$$

V_{max} : The maximum rate of the reaction (same units as y)

K_m : The Michaelis-Menten constant (same units as x)

K_i : The inhibition constant (same units as I)

I : Concentration of inhibitor (non-shared value, varies for each curve)

V_{maxApp} : The V_{max} observed for a particular curve (non-shared value)

K_{mApp} : The K_m observed for a particular curve (non-shared value)

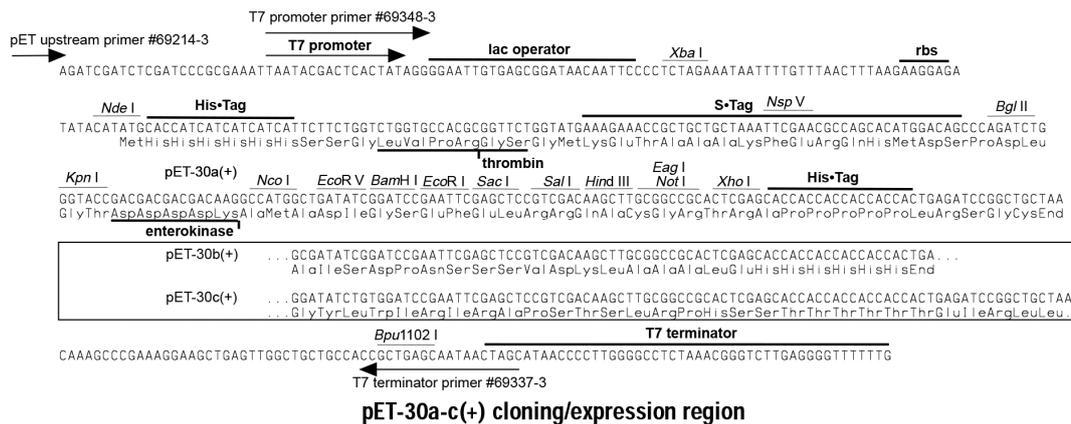
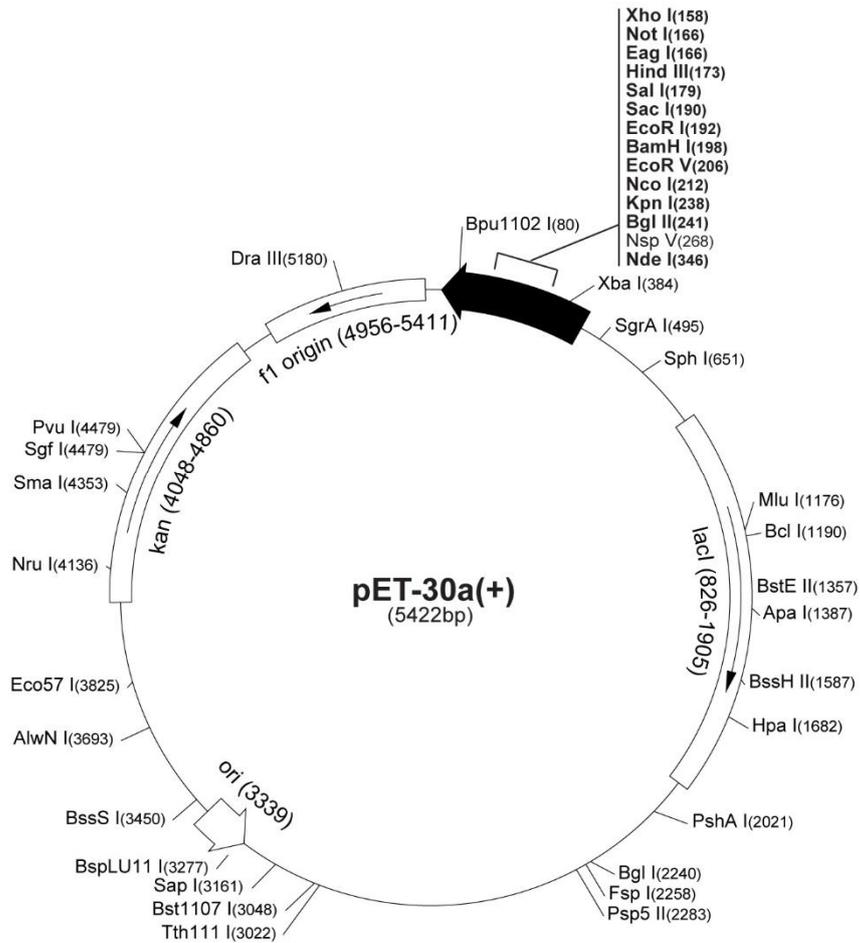
α : The alpha value determines the mechanism. Specifically the value determines the degree to which binding of the inhibitor changes the affinity of the enzyme for the substrate. Where $\alpha = 1$, the inhibitor does not alter binding of the substrate to the enzyme (identical to noncompetitive inhibition). Where α is extremely large, binding of the inhibitor prevents binding of the substrate (becomes identical to competitive inhibition). Where α is extremely small binding of the inhibitor enhances binding of the substrate to the enzyme (identical to uncompetitive inhibition).

Appendix 8.10: Translated *HsUMPS* Protein Sequence

1 MAVARAALGPLVTGLYDVQAFKFGDFVLKSGLSSPIYIDLRGIVSRPRLLSQVADILFQT
61 AQNAGISFDTVCGVPYTALPLATVICSTNQIPMLIRRKETKDYGTKRLVEGTINPGETCL
121 IIEDVVTSGSSVLETVEVLQKEGLKVTDIVLLDREQGGKDKLQAHGIRLHSVCTLSKML
181 EILEQQKKVDAETVGRVKRFIQENVFVAANHNGS **PLSIKEAPK**ELSFGARAELPRIHPVA
241 SKLLRLMQKKEETNLCLSADVSLARELLQLADALGPSICMLKTHVDILNDFTLDMKELIT
301 LAKCHEFLIFEDRKFADIGNTVKKQYEGGIFKIASWADLVNAHVVPVPGSGVVKGLQEVGLP
361 LHRGCLLIAEMSSTGSLATGDYTRAAVRMAEEHSEFVVGFISSRVSMKPEFLHLTPGVQ
421 LEAGGDNLGQQYNPQEVIGKRGSDIIIVGRGIISAADRLEAAEMYRKAWEAYLSRLGV

Amino acid sequence of human UMPS showing: **light grey**, OPRTase domain; **dark grey**, ODCase domain; and **outlined**, domain linker.

Appendix 8.11: pET30a Expression Vector



Appendix 8.12: Sequencing of *HsUMPS* in pET30a-*HsUMPS*

Query	190	ATGGCGGTCGCTCGTGCAGCTTTGGGGCCATTGGTGACGGGTCTGTACGACGTGCAGGCT	249
Sbjct	107	ATGGCGGTCGCTCGTGCAGCTTTGGGGCCATTGGTGACGGGTCTGTACGACGTGCAGGCT	166
Query	250	TTCAAGTTTGGGGACTTCGTGCTGAAGAGCGGGCTTCCTCCCCATCTACATCGATCTG	309
Sbjct	167	TTCAAGTTTGGGGACTTCGTGCTGAAGAGCGGGCTTCCTCCCCATCTACATCGATCTG	226
Query	310	CGGGGCATCGTGTCTCGACCGCTCTTCTGAGTCAGGTTGCAGATATTTTATTCCAACT	369
Sbjct	227	CGGGGCATCGTGTCTCGACCGCTCTTCTGAGTCAGGTTGCAGATATTTTATTCCAACT	286
Query	370	GCCCAAAATGCAGGCATCAGTTTTGACACCGTGTGGAGTGCCTTATACAGCTTTGCCA	429
Sbjct	287	GCCCAAAATGCAGGCATCAGTTTTGACACCGTGTGGAGTGCCTTATACAGCTTTGCCA	346
Query	430	TTGGCTACAGTTATCTGTCAACCAATCAAATCCAATGCTTATTAGAAGGAAAGAAACA	489
Sbjct	347	TTGGCTACAGTTATCTGTCAACCAATCAAATCCAATGCTTATTAGAAGGAAAGAAACA	406
Query	490	AAGGATTATGGAATAAGCGTCTGTAGAAAGAACTATTAATCCAGGAGAAACCTGTTTA	549
Sbjct	407	AAGGATTATGGAATAAGCGTCTGTAGAAAGAACTATTAATCCAGGAGAAACCTGTTTA	466
Query	550	ATCATTGAAGATGTTGTCACCAGTGGATCTAGTGTGTTTGGAACTGTTGAGGTTCTTCAG	609
Sbjct	467	ATCATTGAAGATGTTGTCACCAGTGGATCTAGTGTGTTTGGAACTGTTGAGGTTCTTCAG	526
Query	610	AAGGAGGGCTTGAAGTCACTGATGCCATAGTGTGTTGGACAGAGAGCAGGGAGGCAAG	669
Sbjct	527	AAGGAGGGCTTGAAGTCACTGATGCCATAGTGTGTTGGACAGAGAGCAGGGAGGCAAG	586
Query	670	GACAAGTTGCAGGCGCACGGATCCGCCTCCACTCAGTGTGTACATTGTCCAAAATGCTG	729
Sbjct	587	GACAAGTTGCAGGCGCACGGATCCGCCTCCACTCAGTGTGTACATTGTCCAAAATGCTG	646
Query	730	GAGATTCTCGAGCAGCAGAAAAAAGTTGATGCTGAGACAGTTGGGAGAGTGAAGAGGTTT	789
Sbjct	647	GAGATTCTCGAGCAGCAGAAAAAAGTTGATGCTGAGACAGTTGGGAGAGTGAAGAGGTTT	706
Query	790	ATTCACGAGAATGTCTTTGTGGCAGCGAATCATAATGGTTCTCCCCTTCTATAAAGGAA	849
Sbjct	707	ATTCACGAGAATGTCTTTGTGGCAGCGAATCATAATGGTTCTCCCCTTCTATAAAGGAA	766
Query	850	GCACCCAAAGAACTCAGCTTCGGTGCACGTGCAGAGCTGCCAGGATCCACCCAGTTGCA	909
Sbjct	767	GCACCCAAAGAACTCAGCTTCGGTGCACGTGCAGAGCTGCCAGGATCCACCCAGTTGCA	826
Query	910	TCGAAGCTTCTCAGGCTTATGCAAAAAGAAGGAGACCAATCTGTGTCTATCTGCTGATGTT	969
Sbjct	827	TCGAAGCTTCTCAGGCTTATGCAAAAAGAAGGAGACCAATCTGTGTCTATCTGCTGATGTT	886
Query	970	TCACTGGCCAGAGAGCTGTTGCAGCTAGCAGATGCTTTA GACCTAGTATCTGCATGCTG	1028
Sbjct	887	TCACTGGCCAGAGAGCTGTTGCAGCTAGCAGATGCTTTA GACCTAGTATCTGCATGCTG	946
Query	1029	AAGACTCATGTAGATATTTGA TGATTT ACTCT GATGTGATGAAGGAGTTGATAACT	1085
Sbjct	947	AAGACTCATGTAGATATTTGA TGATTT ACTCT GGATGTGATGAAGGAGTTGATAACT	1006
Query	1086	CTGGCAAA TGCCATG GT C TGATATTTGAAGACCG AGGT GCAGAT TAG AACT	1136
Sbjct	1007	CTGGCAAA TGCCATG GT TCT TGATATTTGAAGACCG AAGTT GCAGAT ATAGGAAAC	1066
Query	1137	A AGTGAAA GCAGTATGAG AG TATCTT AA TAGCTT CTGGGC GATCTAGTA	1188
Sbjct	1067	A AGTGAAAA GCAGTATGAG AG TATCTT AAAA TAGCTT CTGGGC GATCTAGTA	1126

```

Query 1189  --TGCTCAC 1195
          |||
Sbjct 1127  AATGCTCAC 1135

```

A) BlastN search of the results of pET30a-*HsUMPS* routine sequencing using T7 promoter primer (Zhang et al., 2000). Mismatches (occurring due to low signal to noise ratio at end of sequencing results) highlighted **dark grey.**

```

Query 74      CTCAAACACCAAGTCTACTCAAATACGCTTCCCAAGCAGCTTTTCTGTACATCTCTGCTG 133
          |||
Sbjct 1550    CTCAAACACCAAGTCTACTCAAATACGCTTCCCAAGCAGCTTTTCTGTACATCTCTGCTG 1491

Query 134     CTTCAGACGATCAGCTGCTGAGATTATGCCACGACCTACAATGATGATATCGGAACCTC 193
          |||
Sbjct 1490    CTTCAGACGATCAGCTGCTGAGATTATGCCACGACCTACAATGATGATATCGGAACCTC 1431

Query 194     GTTTGCCAATAACTTCTTGTGGGCTATTGTACTGTTGGCCAAGATTATCTCCTCTGCTT 253
          |||
Sbjct 1430    GTTTGCCAATAACTTCTTGTGGGCTATTGTACTGTTGGCCAAGATTATCTCCTCTGCTT 1371

Query 254     CCAACTGAACTCCTGGAGTCAAGTGAAGAAATTCTGGTTTCATGCTTACTCGGGAGCCAG 313
          |||
Sbjct 1370    CCAACTGAACTCCTGGAGTCAAGTGAAGAAATTCTGGTTTCATGCTTACTCGGGAGCCAG 1311

Query 314     AAATAAAACCAACAACAATAATTCAGAGTGTCTCCTCAGCCATTCTAACCGTGCTCTAGTGT 373
          |||
Sbjct 1310    AAATAAAACCAACAACAATAATTCAGAGTGTCTCCTCAGCCATTCTAACCGTGCTCTAGTGT 1251

Query 374     AGTCCCCAGTGGCCAGGGAGCCGGTGGAGCTCATTTCCGCAATAAGGAGGCACCCCCGAT 433
          |||
Sbjct 1250    AGTCCCCAGTGGCCAGGGAGCCGGTGGAGCTCATTTCCGCAATAAGGAGGCACCCCCGAT 1191

Query 434     GCAAAGGCAGGCCCACTTCTTGAGGCTTTTCACTCCTGAGCCTGGCACCACGTGAG 493
          |||
Sbjct 1190    GCAAAGGCAGGCCCACTTCTTGAGGCTTTTCACTCCTGAGCCTGGCACCACGTGAG 1131

Query 494     CATTTACTAGATCTGCCAGGAAGCTATTTTAAAGATACCTCCTTCATACTGCTTTTTTCA 553
          |||
Sbjct 1130    CATTTACTAGATCTGCCAGGAAGCTATTTTAAAGATACCTCCTTCATACTGCTTTTTTCA 1071

Query 554     CTGTGTTTCTATATCTGCAAACCTCCGGTCTTCAAATATCAAGAAGCTCATGGCATTGTTG 613
          |||
Sbjct 1070    CTGTGTTTCTATATCTGCAAACCTCCGGTCTTCAAATATCAAGAAGCTCATGGCATTGTTG 1011

Query 614     CCAGAGTTATCAACTCCTTCATCACATCCAGAGTAAAATCATTCAAAATATCTACATGAG 673
          |||
Sbjct 1010    CCAGAGTTATCAACTCCTTCATCACATCCAGAGTAAAATCATTCAAAATATCTACATGAG 951

Query 674     TCTTCAGCATGCAGATACTAGGTCTTAAAGCATCTGCTAGCTGCAACAGCTCTCTGGCCA 733
          |||
Sbjct 950      TCTTCAGCATGCAGATACTAGGTCTTAAAGCATCTGCTAGCTGCAACAGCTCTCTGGCCA 891

Query 734     GTGAAACATCAGCAGATAGACACAGATTGGTCTCCTTCTTTTGCATAAGCCTGAGAAGCT 793
          |||
Sbjct 890      GTGAAACATCAGCAGATAGACACAGATTGGTCTCCTTCTTTTGCATAAGCCTGAGAAGCT 831

Query 794     TCGATGCAACTGGGTGGATCCTGGGCAGCTCTGCACGTGCACCGAAGCTGAGTTCTTTGG 853
          |||
Sbjct 830      TCGATGCAACTGGGTGGATCCTGGGCAGCTCTGCACGTGCACCGAAGCTGAGTTCTTTGG 771

Query 854     GTGCTTCCTTTATAGAAAGGGGAGAACCATTATGATTCGCTGCCACAAAGACATTCTCCT 913
          |||
Sbjct 770      GTGCTTCCTTTATAGAAAGGGGAGAACCATTATGATTCGCTGCCACAAAGACATTCTCCT 711

Query 914     GAATAAACCTCTTCACTCTCCCAACTGTCTCAGCATCAACTTTTTTCTGCTGCTCGAGAA 973
          |||
Sbjct 710      GAATAAACCTCTTCACTCTCCCAACTGTCTCAGCATCAACTTTTTTCTGCTGCTCGAGAA 651

Query 974     TCTCCAGCATTTTGGACA-TGTACACACTGAGTGGAGGCGGATCCCGTGCCTGCA-CT 1031
          |||
Sbjct 650      TCTCCAGCATTTTGGACAATGTACACACTGAGTGGAGGCGGATCCCGTGCCTGCAACT 591

```



```

Query 387 ATTAATAGTTCCTTCTACAAGACGCTTAGTTCATAATCCTTTGTTTCTTTCCTTCTAAT 446
          |||
Sbjct 448 ATTAATAGTTCCTTCTACAAGACGCTTAGTTCATAATCCTTTGTTTCTTTCCTTCTAAT 389

Query 447 AAGCATTGGAATTTGATTGGTTGAACAGATAACTGTAGCCAATGGCAAAGCTGTATAAGG 506
          |||
Sbjct 388 AAGCATTGGAATTTGATTGGTTGAACAGATAACTGTAGCCAATGGCAAAGCTGTATAAGG 329

Query 507 CACTCCACACACGGTGTCAAACTGATGCCTGCATTTTGGGCAGTTTGGGAATAAAAATATC 566
          |||
Sbjct 328 CACTCCACACACGGTGTCAAACTGATGCCTGCATTTTGGGCAGTTTGGGAATAAAAATATC 269

Query 567 TGCAACCTGACTCAGAAGACGCGGTTCGAGACACGATGCCCGCAGATCGATGTAGATGGG 626
          |||
Sbjct 268 TGCAACCTGACTCAGAAGACGCGGTTCGAGACACGATGCCCGCAGATCGATGTAGATGGG 209

Query 627 GGAGGAAAGCCCGCTTTCAGCACGAAGTCCCCAACTTGAAAGCCTGCACGTCGTACAG 686
          |||
Sbjct 208 GGAGGAAAGCCCGCTTTCAGCACGAAGTCCCCAACTTGAAAGCCTGCACGTCGTACAG 149

Query 687 ACCCGTCACCAATGGCCCCGAAAGCTGCACGAGCGACCGCC 727
          |||
Sbjct 148 ACCCGTCACCAATGGCCCCGAAAGCTGCACGAGCGACCGCC 109

```

B) BlastN search of the results of pET30a-*HsOPRTase* routine sequencing using T7 terminator primer. Mismatches (occurring due to low signal to noise ratio at end of sequencing results) highlighted dark grey.

BlastN of routine sequencing results of pET30a-*HsUMPS*. “Query” is the sequencing result, “Sbjct” is the *HsUMPS* nucleotide sequence (NCBI Sequence Reference: NM_000373.3). The sequencing data shows intact sequence of *HsOPRTase* domain of *HsUMPS* in the cloning vector. The data covers the entire coding region of *HsOPRTase* with an overlap of 426 bases. **A)** Sequencing intensity = 68 (“Signals less than 700 may be affected by background and adjacent sequence. Signals over 6000 may produce overloaded/poor reads”: agfr.org.au). Bases = 657. Intact sequence is confirmed for bases 107 (start codon) to 556. **B)** Sequencing intensity = 103. Bases = 714. Intact sequence is confirmed for bases 130 to 748 (end of gene).

Appendix 8.14: Sequencing of *HsODCase* in pET30a-*HsODC*

Query	198	GAACTCAGCTTCGGTGCACGTGCAGAGCTGCCAGGATCCACCCAGTTGCATCGAAGCTT	257
Sbjct	776	GAACTCAGCTTCGGTGCACGTGCAGAGCTGCCAGGATCCACCCAGTTGCATCGAAGCTT	835
Query	258	CTCAGGCTTATGCAAAGAAGGAGACCAATCTGTGTCTATCTGCTGATGTTTCACTGGCC	317
Sbjct	836	CTCAGGCTTATGCAAAGAAGGAGACCAATCTGTGTCTATCTGCTGATGTTTCACTGGCC	895
Query	318	AGAGAGCTGTTGCAGCTAGCAGATGCTTTAGGACCTAGTATCTGCATGCTGAAGACTCAT	377
Sbjct	896	AGAGAGCTGTTGCAGCTAGCAGATGCTTTAGGACCTAGTATCTGCATGCTGAAGACTCAT	955
Query	378	GTAGATATTTTGAATGATTTTACTCTGGATGTGATGAAGGAGTTGATAACTCTGGCAAAA	437
Sbjct	956	GTAGATATTTTGAATGATTTTACTCTGGATGTGATGAAGGAGTTGATAACTCTGGCAAAA	1015
Query	438	TGCCATGAGTTCCTTGATATTTGAAGACCGGAAGTTTGCAGATATAGGAAACACAGTGAAA	497
Sbjct	1016	TGCCATGAGTTCCTTGATATTTGAAGACCGGAAGTTTGCAGATATAGGAAACACAGTGAAA	1075
Query	498	AAGCAGTATGAAGGAGGTATCTTTAAAAATAGCTTCCTGGGCAGATCTAGTAAATGCTCAC	557
Sbjct	1076	AAGCAGTATGAAGGAGGTATCTTTAAAAATAGCTTCCTGGGCAGATCTAGTAAATGCTCAC	1135
Query	558	GTGGTGCCAGGCTCAGGAGTTGTGAAAGGCCTGCAAGAAGTGGGCCTGCCTTTGCATCGG	617
Sbjct	1136	GTGGTGCCAGGCTCAGGAGTTGTGAAAGGCCTGCAAGAAGTGGGCCTGCCTTTGCATCGG	1195
Query	618	GGGTGCCTCCTTATTGCGGAAATGAGCTCCACCGGCTCCCTGGCCACTGGGGACTTACACT	677
Sbjct	1196	GGGTGCCTCCTTATTGCGGAAATGAGCTCCACCGGCTCCCTGGCCACTGGGGACTTACACT	1255
Query	678	AGAGCA ^o CGGTTAGAATGGCTGAGGAGCACTCTGAATTTGTTGTTGGTTTTATTCTGGC	737
Sbjct	1256	AGAGCA ^o CGGTTAGAATGGCTGAGGAGCACTCTGAATTTGTTGTTGGTTTTATTCTGGC	1315
Query	738	TCCCAGTAAGCATG ^o AAAACAGAATTTCTTCAC ^o TTGACTCCAGGAGTTCAGTTGGAAG	797
Sbjct	1316	TCCCAGTAAGCATG ^o AAAACAGAATTTCTTCAC ^o TTGACTCCAGGAGTTCAGTTGGAAG	1373
Query	798	CAGGAGGAGATAATCTTGGCCAACAG ^o TACCATAGCCAC ^o AGAGGTTATT ^o TG ^o CAAACG	856
Sbjct	1374	CAGGAGGAGATAATCTTGGCCAACAG ^o TACAATAGCCAC ^o AGAAGTTATT ^o TG ^o CAAACG	1431
Query	857	AGGGT ^o CCGATATCATCATTGTAGGTCGTGGCAT ^o TATCTCAGCAGCTGATCGTCTGGAAG	916
Sbjct	1432	AGGTT ^o CCGATATCATCATTGTAGGTCGTGGCAT ^o TATCTCAGCAGCTGATCGTCTGGAAG	1490
Query	917	CAGCAAAGATGTAC ^o GAAAAGCTGCTTGGGAAGCGTATTT ^o GAGTAA ^o ACTTGGTGG ^o TTTG	975
Sbjct	1491	CAGCAAAGATGTAC ^o GAAAAGCTGCTTGGGAAGCGTATTT ^o GAGTAA ^o ACTTGGTGG ^o TTTG	1548

A) BlastN search of the results of pET30a-*HsODC* routine sequencing using T7 promoter primer. Mismatches that show correct peaks in the sequencing chromatogram are outlined. Mismatches (occurring due to low signal to noise ratio at end of sequencing results) highlighted **dark grey**.

Query	83	CTCAAACACCAAGTCTACTCAAATACGCTTCCCAAGCAGCTTTTCTGTACATCTCTGCTG	142
Sbjct	1550	CTCAAACACCAAGTCTACTCAAATACGCTTCCCAAGCAGCTTTTCTGTACATCTCTGCTG	1491
Query	143	CTTCCAGACGATCAGCTGCTGAGATTATGCCACGACCTACAATGATGATATCGGAACCTC	202
Sbjct	1490	CTTCCAGACGATCAGCTGCTGAGATTATGCCACGACCTACAATGATGATATCGGAACCTC	1431
Query	203	GTTTGCCAATAACTTCTTGTGGGCTATTGTACTGTTGGCCAAGATTATCTCCTCTGCTT	262
Sbjct	1430	GTTTGCCAATAACTTCTTGTGGGCTATTGTACTGTTGGCCAAGATTATCTCCTCTGCTT	1371

Query	263	CCAACTGAACTCCTGGAGTCAAGTGAAGAAATTCTGGTTTCATGCTTACTCGGGAGCCAG	322
Sbjct	1370	CCAACTGAACTCCTGGAGTCAAGTGAAGAAATTCTGGTTTCATGCTTACTCGGGAGCCAG	1311
Query	323	AAATAAAACCAACAACAATTCAGAGTGTCTCCAGCCATTCTAACCGTGTCTAGTGT	382
Sbjct	1310	AAATAAAACCAACAACAATTCAGAGTGTCTCCAGCCATTCTAACCGTGTCTAGTGT	1251
Query	383	AGTCCCCAGTGGCCAGGGAGCCGGTGGAGCTCATTTCCGCAATAAGGAGGCACCCCGAT	442
Sbjct	1250	AGTCCCCAGTGGCCAGGGAGCCGGTGGAGCTCATTTCCGCAATAAGGAGGCACCCCGAT	1191
Query	443	GCAAAGGCAGGCCACTTCTTGAGGCCTTTCACAACCTCTGAGCCTGGCACCACGTGAG	502
Sbjct	1190	GCAAAGGCAGGCCACTTCTTGAGGCCTTTCACAACCTCTGAGCCTGGCACCACGTGAG	1131
Query	503	CATTTACTAGATCTGCCAGGAAGCTATTTTAAAGATACCTCCTTCATACTGCTTTTTCA	562
Sbjct	1130	CATTTACTAGATCTGCCAGGAAGCTATTTTAAAGATACCTCCTTCATACTGCTTTTTCA	1071
Query	563	CTGTGTTTCCTATATCTGCAAACCTCCGGTCTTCAAATATCAAGAACTCATGGCATTG	622
Sbjct	1070	CTGTGTTTCCTATATCTGCAAACCTCCGGTCTTCAAATATCAAGAACTCATGGCATTG	1011
Query	623	CCAGAGTTATCAACTCCTTCATCACATCCAGAGTAAAATCATTCAAAATATCTACATGAG	682
Sbjct	1010	CCAGAGTTATCAACTCCTTCATCACATCCAGAGTAAAATCATTCAAAATATCTACATGAG	951
Query	683	TCTTCAGCATGCAGATACTAGGTCTTAAAGCATCTGCTAGCTGCAACAGCTCTCTGGCCA	742
Sbjct	950	TCTTCAGCATGCAGATACTAGGTCTTAAAGCATCTGCTAGCTGCAACAGCTCTCTGGCCA	891
Query	743	GTGAAACATCAGCAGATAGACACAGATTGGTCTCCTTCTTTTGCATAAGCCTGAGAAGCT	802
Sbjct	890	GTGAAACATCAGCAGATAGACACAGATTGGTCTCCTTCTTTTGCATAAGCCTGAGAAGCT	831
Query	803	TCGATGCAACTGGGTGGATCCTGGGCAGCTCTGCACGTGCACCGAAGCTGAGTTC	857
Sbjct	830	TCGATGCAACTGGGTGGATCCTGGGCAGCTCTGCACGTGCACCGAAGCTGAGTTC	776

B) BlastN search of the results of pET30a-HsODCase routine sequencing using T7 terminator primer. No Mismatches to display.

BlastN of routine sequencing results of pET30a-HsODCase. “Query” is the sequencing result, “Sbjct” is the *HsUMPS* nucleotide sequence (NCBI Sequence Reference: NM_000373.3). The sequencing data shows intact sequence of *HsODCase* region of *HsUMPS* in the cloning vector. The data covers the entire coding region of *HsODCase* with an overlap of 554 bases. **A)** Sequencing intensity = 104 (“Signals less than 700 may be affected by background and adjacent sequence. Signals over 6000 may produce overloaded/poor reads”: agfr.org.au). Bases = 771. Intact sequence is confirmed for bases 776 (start *ODCase* region) to 1330 of *HsUMPS*. **B)** Sequencing intensity = 139. Bases = 859. Intact sequence is confirmed for bases 776 to 1550 (end of gene) of *HsUMPS*.

Appendix 8.15: 'convertpir.pl' Script for Alignment File Conversion

```
#!/usr/sbin/perl
# Document: perl script 'convertpir' (v3.0)
# Author: Diana Kolbe (creation -> v2.0)
# Revisor: Will Ford (v2.0 -> v 3.0)
# Description: This script take a multiple alignment in standard NBRF/PIR format.
# Multiple sequences are output as either template(s) or target(s). A
# second output file is the script file to be used by modeller4 to create
# a model based on these sequences.
# Note: This program will not prevent you from entering more than one target
sequence.
# If you specify more than one target, the .top file will be incorrect and will
# not run. This is to allow for a feature that should be implemented soon.
# Note: If you get "Command not found" when trying to run this program:
# 1. Make sure permissions are set to allow execution ("chmod +x convertpir")
# 2. The first line must match the location of perl on your local system.
# For us, that is /usr/bin/local/perl. /usr/bin/perl is also common
# You can check this with "which perl"
# Note: Some manual editing may be required before modeller can use the output files.
# 1. The sequence in the alignment file must exactly match the sequence in the ATOM
# portion of the pdb file. I suggest modifying the sequence in the alignment
file
# before running this conversion, because standard PIR format is somewhat easier
# to read.
# 2. Unless you are reading the entire pdb file, you will need to add residue
numbers
# and chain IDs to the modeller file. This can only be done after the
conversion.
# Replace the four periods in the description line with the first residue
number,
# first chain ID, last residue number, last chain ID, in that order.
# Consult http://guitar.rockefeller.edu/modeller/manual/node66.html for details.
# Query for input file
print "What is the name of your input file (default is alignment.pir)?\n";
$infile = <STDIN>;
chomp($infile);
$infile = 'alignment.pir' if ($infile eq "");
open(INFILE,"$infile") or die "Could not open $infile!";

# Query for output file; will overwrite if it already exists
print "What is the output alignment file (default: alignment.ali)?\n";
chomp ($alignfile = <STDIN>);
$alignfile = 'alignment.ali' if ($alignfile eq "");
open(ALIGNFILE,">$alignfile") or die "Could not open $alignfile!";
print "Hi";
$count = 0;
while ($newseq = <INFILE>) {
    if ($newseq =~ /^>P1;/) { #first line of next sequence
        $newseq .= <INFILE>;
        $sequences[$count] = { 'header' => $newseq };
        #Store header information (anonymous hash)
        $newseq = '';
        while ($newseq !~ /\*/) { #Concatenate sequence until end reached <-
revised/corrected - Will Ford
            $newseq .= <INFILE>;
        }
        $newseq =~ s/\r//g; #<- added to remove DOS style <CR>'s - Will Ford
        $newseq =~ s/[.~]/-/g; #Substitutue . or ~ with -
        $newseq =~ s/ //g; #Remove spaces
        ${$sequences[$count]}{'sequence'} = $newseq;
        #Store sequence

        $count++;
    }
}
close INFILE;

@templates = (); #Clear lists
```

```

@targets = ();

# Display all sequences found, query for which one to be output
print STDOUT "The following sequences were found.  Add which one to the output?\n";
for ($i=0; $i<$count; $i++) {
    print STDOUT $i+1, ". ", ${$sequences[$i]}{'header'};
}
print STDOUT "\nChoose by entering the number.  When finished, just press
RETURN.\n\n";
chomp ($select = <STDIN>);

while ($select) {
    $select--;
        #Adjust for 0 to n-1 subscripts (instead of 1 to n)

    if ($select < $count and $select >= 0) {
        print STDOUT "You have selected\n", ${$sequences[$select]}{'header'};
        print STDOUT "Is this sequence a\n1\. Template or a\n2\. Target\n";
        chomp ($tt = <STDIN>);

        if ($tt == '1') {
            #If it is a template
            # Query for pdb code
            print STDOUT "What is the protein code for this sequence (ie, lad3)?\n";
            print STDOUT "Important: this must exactly match the name of the pdb file
(ie, lad3.pdb)\n";
            chomp (${sequences[$select]}{'code'} = <STDIN>);

            # Query for type of structure information available
            print STDOUT "\nWhat data is available for this sequence? (Choose menu
number)\n";
            print STDOUT "1. Crystal Structure\n2. NMR Structure\n3. Model\n";
            print STDOUT "4. Structure (any type)\n";
            chomp($type = <STDIN>);
            if ($type == 1) {$type = 'structureX'};
            elsif ($type == 2) {$type = 'structureN'};
            elsif ($type == 3) {$type = 'structureM'};
            else {$type = 'structure'};
            ${sequences[$select]}{'type'} = $type;

            # Copy pointer to this sequence info to list of templates
            push @templates, $sequences[$select];
        }
        else {
            #If it is a target
            # Query for protein label
            print STDOUT "What is the label for this sequence?\n";
            chomp (${sequences[$select]}{'code'} = <STDIN>);

            # Automatically label it as sequence only (no structure data available)
            ${sequences[$select]}{'type'} = 'sequence';

            # Copy pointer to this sequence info to list of targets
            push @targets, $sequences[$select];
        }

        # Remove the sequence just processed from the list of available sequences
        splice @sequences, $select, 1;
        $count--;
    }
    else {
        print STDOUT "That selection is invalid.  Please try again.\n";
    }

    # Menu of available sequences
    print STDOUT "\nThe following sequences were found.  Add which one to the
output?\n";
    for ($i=0; $i<$count; $i++) {
        print STDOUT $i+1, ". ", ${$sequences[$i]}{'header'};
    }
    print STDOUT "\nChoose by entering the number.  When finished, just press
RETURN.\n\n";
    chomp ($select = <STDIN>);
}

```

```

    }

# Output sequence file
foreach $sequence (@templates,@targets) {
    print ALIGNFILE ">P1;", $$sequence{'code'}, "\n", $$sequence{'type'};
    print ALIGNFILE "\:", $$sequence{'code'}, "\:\.\:\.\:\.\:\.\:\:\:\:\n";
    print ALIGNFILE $$sequence{'sequence'}, "\n";
}
close ALIGNFILE;

# Query for starting and ending model numbers
print STDOUT "How many models do you want to produce?\n";
chomp($endno = <STDIN>);
print STDOUT "What is the number of the first model to produce (usually 1)?\n";
chomp($startno = <STDIN>);
$endno = $startno + $endno - 1;

foreach $target (@targets) {
    print STDOUT "\nYou have selected ", $$target{'code'}, " as a target.\n";
    print STDOUT "What is the name of the output TOP script file (default is
", $$target{'code'}, ".top).\n";
    chomp($topfile = <STDIN>);
    $topfile = $$target{'code'} . ".top" if (!$topfile);
    open (TOPFILE, ">$topfile") or die "Could not open $topfile!\n";

# Output modeller script (top) file
print TOPFILE "INCLUDE      # Include the predefined TOP routines\n\n";
print TOPFILE "SET ALNFILE = '$alignfile' # Alignment filename\n";
print TOPFILE "SET KNOWNNS = ";
foreach $template (@templates) {
    print TOPFILE "' ', $$template{'code'}, ' ' ";
}
print TOPFILE "\t\t# Code of the template\n";
print TOPFILE "SET SEQUENCE = ', $$target{'code'}, ' '\n";
print TOPFILE "SET STARTING_MODEL = $startno\n";
print TOPFILE "SET ENDING_MODEL = $endno\n\n";
print TOPFILE "SET DEVIATION = 4.0 # Amount of randomization between
models\n\n";
print TOPFILE "SET LIBRARY_SCHEDULE = 1 # thorough VTF schedule\n";
print TOPFILE "SET MAX_VAR_ITERATIONS = 300\n\n";
print TOPFILE "SET FINAL_MALIGN3D = 1 # MALIGN3D and write superposed\n #
templates & models\n";
print TOPFILE "SET MD_LEVEL = 'refine_1' # thorough MD annealing\n\n";
print TOPFILE "CALL ROUTINE = 'model'\n";

close TOPFILE;
}

```

convertpir.pl Perl script file. This script was used for converting PIR format alignments from ClustalX to .ali format for use with Modeller. Script also outputs .TOP input files for running modeller which are now obsolete.

Appendix 8.17: Modified 'model-default.py' Script for Homology Modelling of PfOPRTase

```
# Homology modelling by the automodel class
from modeller import *           # Load standard Modeller classes
from modeller.automodel import * # Load the automodel class

log.verbose() # request verbose output
env = environ() # create a new MODELLER environment to build this model in

# directories for input atom files
env.io.atom_files_directory = ['.']

a = automodel(env,
              alnfile = 'alignment.ali',      # alignment filename
               knowns  = ('2WNS', '2PRY'),    # codes of the templates
               sequence = 'pfop')            # code of the target
a.starting_model= 1 # index of the first model
a.ending_model  = 10 # index of the last model
                 # (determines how many models to calculate)
a.make()          # do the actual homology modelling
```

Modified version of model-default.py example python script. This script was used for running Modeller to generate the homology models. The template codes and target code were changed to match the .pdb files and target sequence names in the alignment file.

Appendix 8.18: Default High-Stringency Docking Parameters

ligand_atom_file	LIGANDS.mo12
limit_max_ligands	no
skip_molecule	no
read_mol_solvation	no
calculate_rmsd	no
orient_ligand	yes
automated_matching	yes
receptor_site_file	SPHERES.sphgen
max_orientations	500
critical_points	no
chemical_matching	yes
chem_match_tbl	chem_match.tbl
use_ligand_spheres	no
flexible_ligand	yes
min_anchor_size	2
pruning_use_clustering	yes
pruning_max_orients	100
pruning_clustering_cutoff	100
use_internal_energy	no
use_clash_overlap	yes
clash_overlap	0.5
bump_filter	yes
bump_grid_prefix	GRIDPREFIX
max_bumps_anchor	12
max_bumps_growth	12
score_molecules	yes
contact_score_primary	no
contact_score_secondary	no
grid_score_primary	yes
grid_score_secondary	no
grid_score_rep_rad_scale	1
grid_score_vdw_scale	1
grid_score_es_scale	1
grid_score_grid_prefix	GRIDPREFIX
dock3.5_score_secondary	no
continuous_score_secondary	no
gbsa_zou_score_secondary	no
gbsa_hawkins_score_secondary	no
amber_score_secondary	no
minimize_ligand	yes
minimize_anchor	yes
minimize_flexible_growth	yes
use_advanced_simplex_parameters	no
simplex_max_cycles	1
simplex_score_converge	0.1
simplex_cycle_converge	1.0
simplex_trans_step	1.0
simplex_rot_step	0.1
simplex_tors_step	10.0
simplex_anchor_max_iterations	500
simplex_grow_max_iterations	500
simplex_final_min	no
simplex_random_seed	0
atom_model	united
vdw_defn_file	vdw.defn
flex_defn_file	flex.defn
flex_drive_file	flex_drive.tbl
chem_defn_file	chem.defn
ligand_outfile_prefix	OUTPUT.mo12
write_orientations	no
num_scored_conformers_written	1
rank_ligands	yes
max_ranked_ligands	100

Optimised high-stringency docking input parameters file. **LIGANDS.mo12** = the compound database file to screen. **SPHERES.sphgen** = the spheres file. **GRIDPREFIX** = the grid file prefix (not the names of the grid files themselves). **OUTPUT.mo12** = the top scoring compounds, oriented in the active site with Dock's VDW, ES and Grid scores.

**Appendix 8.19: Changes to Grid File Generation Parameters for
PfOPRTase, *PfODCase*, and *HsOPRTase***

Target	GRID input parameter changes from default settings
<i>PfOPRTase</i>	grid_spacing = 0.2 bump_overlap = 0.6
<i>PfODCase</i>	grid_spacing = 0.2 bump_overlap = 0.6
<i>HsOPRTase</i>	grid_spacing = 0.2 bump_overlap = 0.45

Appendix 8.20: Very-Low Stringency Docking Parameters

ligand_atom_file	LIGANDS.mol2
limit_max_ligands	no
skip_molecule	no
read_mol_solvation	no
calculate_rmsd	no
orient_ligand	yes
automated_matching	yes
receptor_site_file	SPHERES.sphgen
max_orientations	50
critical_points	no
chemical_matching	no
use_ligand_spheres	no
flexible_ligand	yes
min_anchor_size	2
pruning_use_clustering	no
pruning_max_orients	50
pruning_orient_score_cutoff	25.0
pruning_max_conformers	50
pruning_conformer_score_cutoff	25.0
use_internal_energy	no
use_clash_overlap	yes
clash_overlap	0.5
bump_filter	yes
bump_grid_prefix	GRIDPREFIX
max_bumps_anchor	2
max_bumps_growth	2
score_molecules	yes
contact_score_primary	no
contact_score_secondary	no
grid_score_primary	yes
grid_score_secondary	no
grid_score_rep_rad_scale	1
grid_score_vdw_scale	1
grid_score_es_scale	1
grid_score_grid_prefix	GRIDPREFIX
dock3.5_score_secondary	no
continuous_score_secondary	no
gbsa_zou_score_secondary	no
gbsa_hawkins_score_secondary	no
amber_score_secondary	no
minimize_ligand	no
atom_model	united
vdw_defn_file	vdw.defn
flex_defn_file	flex.defn
flex_drive_file	flex_drive.tbl
ligand_outfile_prefix	OUTPUT.mol2
write_orientations	no
num_scored_conformers_written	1
rank_ligands	yes
max_ranked_ligands	2000

Very low-stringency docking input parameters file. **LIGANDS.mol2** = the compound database file to screen. **SPHERES.sphgen** = the spheres file. **GRIDPREFIX** = the grid file prefix (not the names of the grid files themselves). **OUTPUT.mol2** = the top scoring compounds, oriented in the active site with Dock's VDW, ES and Grid scores.

Appendix 8.21: Low-Stringency Docking Parameters

ligand_atom_file	LIGANDS.mol2
limit_max_ligands	no
skip_molecule	no
read_mol_solvation	no
calculate_rmsd	no
orient_ligand	yes
automated_matching	yes
receptor_site_file	SPHERES.sphgen
max_orientations	150
critical_points	no
chemical_matching	no
use_ligand_spheres	no
flexible_ligand	yes
min_anchor_size	2
pruning_use_clustering	yes
pruning_max_orients	100
pruning_clustering_cutoff	100
use_internal_energy	yes
internal_energy_att_exp	6
internal_energy_rep_exp	12
internal_energy_dielectric	4.0
use_clash_overlap	yes
clash_overlap	0.5
bump_filter	yes
bump_grid_prefix	GRIDPREFIX
max_bumps_anchor	2
max_bumps_growth	2
score_molecules	yes
contact_score_primary	no
contact_score_secondary	no
grid_score_primary	yes
grid_score_secondary	no
grid_score_rep_rad_scale	1
grid_score_vdw_scale	1
grid_score_es_scale	1
grid_score_grid_prefix	GRIDPREFIX
dock3.5_score_secondary	no
continuous_score_secondary	no
gbsa_zou_score_secondary	no
gbsa_hawkins_score_secondary	no
amber_score_secondary	no
minimize_ligand	yes
minimize_anchor	no
minimize_flexible_growth	no
use_advanced_simplex_parameters	no
simplex_max_iterations	100
simplex_max_cycles	1
simplex_score_converge	0.1
simplex_cycle_converge	1.0
simplex_trans_step	1.0
simplex_rot_step	0.1
simplex_tors_step	10.0
simplex_final_min	no
simplex_random_seed	0
atom_model	united
vdw_defn_file	vdw.defn
flex_defn_file	flex.defn
flex_drive_file	flex_drive.tbl
ligand_outfile_prefix	OUTPUT.mol2
write_orientations	no
num_scored_conformers_written	1
rank_ligands	yes
max_ranked_ligands	100

Low-stringency docking input parameters file. **LIGANDS.mol2** = the compound database file to screen. **SPHERES.sphgen** = the spheres file. **GRIDPREFIX** = the grid file prefix (not the names of the grid files themselves). **OUTPUT.mol2** = the top scoring compounds, oriented in the active site with Dock's VDW, ES and Grid scores.

Appendix 8.22: Python Clustering Script

```
# Author: Michael Roach
import os
import re
from operator import itemgetter
from math import sqrt

def XyzGrab(atomblock):
    p = n.replace(" ", " ")
    p = p.replace(" ", " ")
    p = p.replace(" ", " ")
    p = p.replace(" ", " ")
    q = p.split(" ")
    x = q[3]
    y = q[4]
    z = q[5]
    return [x,y,z]

def Mean(numlist):
    if len(numlist) == 0:
        return float('nan')

    floatNums = [float(x) for x in numlist]
    return sum(floatNums) / len(numlist)

def MeanXYZ(mainlist):
    avx = []
    avy = []
    avz = []
    for m in mainList:
        avx.append(m[-3])
        avy.append(m[-2])
        avz.append(m[-1])

    ax = Mean(avx)
    ay = Mean(avy)
    az = Mean(avz)
    avcoords = [ax, ay, az]
    return avcoords

def MeanGrid(mainlist):
    grids = []
    for n in mainlist:
        grids.append(n[0])
    avgrid = Mean(grids)
    return avgrid

def DistDiff(xyz1, xyz2):
    dist = sqrt((xyz2[0]- xyz1[0])**2 + (xyz2[1]- xyz1[1])**2 + (xyz2[2]-
xyz1[2])**2)
    return dist

# Finds and prints the files in this script's directory
print "\n\nFiles in current directory"
dirList=os.listdir('.')
for fname in dirList:
    print fname

# prompts for file for processing
print "Type filename for clustering."
filename = raw_input("> ")

print "Clustering radius? [2]"
cluster_radius = float(raw_input("> "))

# appends .mol2 file extension if needed
if not '.mol2' in filename:
    filename = "%s.mol2" % filename
```

```

# opens file and reads contents to mol2
mol2 = open(filename).read()

# splits mol2 file into separate entries and stores as list
# in mol2list and removes any entries that are simply \n
d = '0 ROOT'
mol2list = mol2.split(d)
for n in mol2list:
    if n.replace("\n","") == "":
        mol2list.pop(mol2list.index(n))

mainList = []

for n in mol2list: # for each molecule entry isolate xyz block
    o = n.split("@<TRIPOS>ATOM")
    m = o[1]
    p = m.split("@<TRIPOS>BOND")
    atom = []
    for n in p[0].split("\n"):
        if len(n) > 1:
            a = XyzGrab(n)
            atom.append(a)
    mainList.append(atom)

# mainList ordered as [molecules[atoms[xyz]]]

gridscores = [] # this creates a list of gridscores
for n in mol2list:
    s1 = n.split("Grid Score:")
    s2 = s1[1].split("\n")
    s3 = float(s2[0])
    gridscores.append(s3)

for m in mainList: # this block adds the average x,y,z coords for each molecule
    avx = []
    avy = []
    avz = []
    for a in m:
        avx.append(a[0])
        avy.append(a[1])
        avz.append(a[2])
    m.append(Mean(avx))
    m.append(Mean(avy))
    m.append(Mean(avz))
    m.insert(0, gridscores[mainList.index(m)]) # this appends the grid scores
                                                # to their molecules

# NO MAINLIST SORTING ABOVE THIS POINT

avgxyz = MeanXYZ(mainList) # works out initial dist diffs for molecules
for m in mainList:
    mxyz = [m[-3],m[-2],m[-1]]
    x = DistDiff(mxyz,avgxyz)
    m.insert(-3,x)

mainList.sort(key=itemgetter(-4)) # sorts mainlist by dist from average

output = []
while len(mainList) > 1:
    next_cluster = []

    while mainList[-1][-4] >= cluster_radius:
        next_cluster.append(mainList.pop())
        av_xyz = MeanXYZ(mainList)

    for n in mainList:
        xyz = [n[-3],n[-2],n[-1]]

```

```

        dist = DistDiff(av_xyz, xyz)
        n.pop(-4)
        n.insert(-3,dist)

    mainList.sort(key=itemgetter(-4))

no_in_cluster = len(mainList)
av = MeanXYZ(mainList)
average_score = MeanGrid(mainList)
return_info = [no_in_cluster, average_score, av[0], av[1], av[2]]
output.append(return_info)

mainList = next_cluster

print "\n%s clusters identified" % len(output)
print "Output file name?"
outfile = raw_input("> ")
out = open(outfile, 'w')
output.sort(key=itemgetter(1))

for m in output:
    out.write("No._in_Cluster: %s Average_Grid_Score: %s Average_X,Y,Z: %s %s %s \n"
% (m[0], round(m[1],3), round(m[2],3), round(m[3],3), round(m[4],3)))

    x = "%s" % round(m[2],1)
    y = "%s" % round(m[3],1)
    z = "%s" % round(m[4],1)
    while len(x)<5:
        x = " %s" % x
    while len(y)<5:
        y = " %s" % y
    while len(z)<5:
        z = " %s" % z
    pdb_temp = "%s_%s.pdb" % (outfile, output.index(m))
    pdb = open(pdb_temp, 'w')
    pdb.write("HETATM    1  C1  <O>      1      %s00 %s00 %s00  1.00  0.00          C"
% (x,y,z))
    pdb.close()

```

Python Clustering Script. This was the 'mol2cluster.py' python script used to cluster docked probe molecule orientations in the hybrid screening method.

Appendix 8.23: Hybrid Screening Arbitrary Score: Descriptor Multipliers and Formulae

Descriptor type/Score	Multiplier	Descriptor/Score formula
H-bond donor/acceptor that is H-bonding to a side-chain O, N or H	5	$D = T \times M \times G$
H-bond donor/acceptor that is H-bonding to a α -carbon-chain O, N or H	2.5	$D = T \times M \times G$
Hydrophobe	1	$D = T \times M \times G$
Ring	1	$D = T \times M \times G$
Ionisable	1	$D = T \times M \times G$
Electrostatic	2	$D = E \times M$
Total Descriptors	-	$Td = \sum D$
Shape Match	0.75	$S = T \times (M \times Td)$
Overlap Penalty	0.75	$O = (T \times A) \times \left(\frac{(M \times Td)}{10} \right)$
Total Shape	-	$Ts = S - O$
Arbitrary Score	-	$As = Td + Ts$

D = Descriptor score

T = Tanimoto score

M = Multiplier

G = Mean Grid score for particular descriptor

E = EON Tanimoto 'ES_Combo' score

Td = Total descriptor score (sum of the descriptor scores)

S = Shape score

O = Overlap penalty

A = Number of atoms present in overlap penalty query file

Ts = Total shape score (shape score minus overlap penalty)

As = Arbitrary score

Appendix 8.24: Dock Score to .csv Script

```
# Author: Michael Roach
#!/usr/bin/perl
use warnings;
use strict;
my @zincid;
my @files = glob("*.mol2");
print "\nThis script will take the dock scores from all the .mol2
files in this folder and output them all into a single .csv
spreadsheet file. Make sure all the files have been scored the
same way.\n";
print "\n.mol2 files to be processed:\n\n";
for (@files)
{
    print "$_\n";
}
print "\nPlease specify the output .csv file name (don't
include the file extension):";
my $outfile = <STDIN>;
chomp $outfile;
$outfile = "$outfile.csv";
$/ = "0 ROOT";
for (@files)
{
    open FILE, $_;
    while (<FILE>)
    {
        if ( m/Name:\t\t(.*)TRIPOS.MOLECULE/s)
        {
            push @zinzincid, $1 if defined $1;
        }
    }
}
pop @zincid;
for (@zincid)
{
    s/#{10}//,/g; s/\s//g;
    s/\n/,/g; s/@</\n/g;
    s/:/,/g; s/GB\SAScore/GB\SA Score/g;
}
print @zincid;
open OUT,"> $outfile" or die $!;
print OUT @zincid;
exit;
```

score_script.pl. This script was used to output the dock scores from all compounds in a .mol2 file to a .csv spreadsheet file.

Appendix 8.25: Modified .mol2 to .tab Conversion Script

```
#!/bin/sh
#\
exec csts -f "$0" ${1+"$@"}

set infile [lindex $argv 0]
set outfile [open "bitstring.tab" w]

prop setparam E_SCREEN extended 0
set record 0

set fh [molfile open $infile r hydrogens add]

molfile loop $fh ehandle {
    incr record
    if {$record > 10} break
    set screen [ens get $ehandle E_SCREEN]
    set nsc [ens get $ehandle E_NAME]
    puts $outfile [format "%s\t%s" $nsc $screen]
}
```

Modified version of the get_screen_addH.tcl script. Script was used for generating a .tab binary 2D-fingerprint file from a .mol2 file using the CACTVS toolkit (Ihlenfeldt et al., 1994).

References

- AGUIAR-PULIDO, V., GESTAL, M., CRUZ-MONTEAGUDO, M., RABUNAL, J. R., DORADO, J. & MUNTEANU, C. R. 2013. Evolutionary computation and QSAR research. *Curr Comput Aided Drug Des*, 9, 206-25.
- ALAM, A. 2014. Serine Proteases of Malaria Parasite Plasmodium falciparum: Potential as Antimalarial Drug Targets. *Interdisciplinary Perspectives on Infectious Diseases*, 2014, 7.
- ALTSCHUL, S. F., GISH, W., MILLER, W., MYERS, E. W. & LIPMAN, D. J. 1990. Basic local alignment search tool. *J Mol Biol*, 215, 403-10.
- AMYES, T. L., MING, S. A., GOLDMAN, L. M., WOOD, B. M., DESAI, B. J., GERLT, J. A. & RICHARD, J. P. 2012. Orotidine 5'-monophosphate decarboxylase: transition state stabilization from remote protein-phosphodianion interactions. *Biochemistry*, 51, 4630-2.
- ANAFI, M., GAZIT, A., ZEHAZI, A., BEN-NERIAH, Y. & LEVITZKI, A. 1993. Tyrphostin-induced inhibition of p210bcr-abl tyrosine kinase activity induces K562 to differentiate. *Blood*, 82, 3524-9.
- ANDERSON, L. W., STRONG, J. M. & CYSYK, R. L. 1989. Cellular pharmacology of DUP-785, a new anticancer agent. *Cancer Commun*, 1, 381-7.
- ARCINIEGA, M. & LANGE, O. F. 2014. Improvement of Virtual Screening Results by Docking Data Feature Analysis. *Journal of Chemical Information and Modeling*, 54, 1401-1411.
- ARNOLD, J. R., BURDICK, K. W., PEGG, S. C., TOBA, S., LAMB, M. L. & KUNTZ, I. D. 2004. SitePrint: three-dimensional pharmacophore descriptors derived from protein binding sites for family based active site analysis, classification, and drug design. *J Chem Inf Comput Sci*, 44, 2190-8.
- ASAI, T., O'SULLIVAN, W. J., KOBAYASHI, M., GERO, A. M., YOKOGAWA, M. & TATIBANA, M. 1983. Enzymes of the de novo pyrimidine biosynthetic pathway in *Toxoplasma gondii*. *Mol Biochem Parasitol*, 7, 89-100.
- AURRECOECHEA, C., BRESTELLI, J., BRUNK, B. P., DOMMER, J., FISCHER, S., GAJRIA, B., GAO, X., GINGLE, A., GRANT, G., HARB, O. S., HEIGES, M., INNAMORATO, F., IODICE, J., KISSINGER, J. C., KRAEMER, E., LI, W., MILLER, J. A., NAYAK, V., PENNINGTON, C., PINNEY, D. F., ROOS, D. S., ROSS, C., STOECKERT, C. J., JR., TREATMAN, C. & WANG, H. 2009. PlasmoDB: a functional genomic database for malaria parasites. *Nucleic Acids Res*, 37, D539-43.
- AWALE, M. & REYMOND, J.-L. 2014. Atom Pair 2D-Fingerprints Perceive 3D-Molecular Shape and Pharmacophores for Very Fast Virtual Screening of ZINC and GDB-17. *Journal of Chemical Information and Modeling*.
- BACA, A. M. & HOL, W. G. 2000. Overcoming codon bias: a method for high-level overexpression of Plasmodium and other AT-rich parasite genes in *Escherichia coli*. *Int J Parasitol*, 30, 113-8.
- BAELL, J. B. & HOLLOWAY, G. A. 2010. New substructure filters for removal of pan assay interference compounds (PAINS) from screening libraries and for their exclusion in bioassays. *J Med Chem*, 53, 2719-40.
- BARILLARI, C., MARCOU, G. & ROGNAN, D. 2008. Hot-spots-guided receptor-based pharmacophores (HS-Pharm): a knowledge-based approach to identify ligand-anchoring atoms in protein cavities and prioritize structure-based pharmacophores. *J Chem Inf Model*, 48, 1396-410.
- BATES, P. A., JACKSON, R. M. & STERNBERG, M. J. 1997. Model building by comparison: a combination of expert knowledge and computer automation. *Proteins, Suppl* 1, 59-67.

- BELLO, A. M., KONFORTE, D., PODUCH, E., FURLONGER, C., WEI, L., LIU, Y., LEWIS, M., PAI, E. F., PAIGE, C. J. & KOTRA, L. P. 2009. Structure-activity relationships of orotidine-5'-monophosphate decarboxylase inhibitors as anticancer agents. *J Med Chem*, 52, 1648-58.
- BELLO, A. M., PODUCH, E., FUJIHASHI, M., AMANI, M., LI, Y., CRANDALL, I., HUI, R., LEE, P. I., KAIN, K. C., PAI, E. F. & KOTRA, L. P. 2007. A potent, covalent inhibitor of orotidine 5'-monophosphate decarboxylase with antimalarial activity. *J Med Chem*, 50, 915-21.
- BELLO, A. M., PODUCH, E., LIU, Y., WEI, L., CRANDALL, I., WANG, X., DYANAND, C., KAIN, K. C., PAI, E. F. & KOTRA, L. P. 2008. Structure-activity relationships of C6-uridine derivatives targeting plasmodia orotidine monophosphate decarboxylase. *J Med Chem*, 51, 439-48.
- BELLO, L. J. 1974. Regulation of thymidine kinase synthesis in human cells. *Exp Cell Res*, 89, 263-74.
- BENDER, A., MUSSA, H. Y., GLEN, R. C. & REILING, S. 2003. Molecular Similarity Searching Using Atom Environments, Information-Based Feature Selection, and a Naïve Bayesian Classifier. *Journal of Chemical Information and Computer Sciences*, 44, 170-178.
- BENDER, A., MUSSA, H. Y., GLEN, R. C. & REILING, S. 2004. Similarity searching of chemical databases using atom environment descriptors (MOLPRINT 2D): evaluation of performance. *J Chem Inf Comput Sci*, 44, 1708-18.
- BERMAN, H. M., WESTBROOK, J., FENG, Z., GILLILAND, G., BHAT, T. N., WEISSIG, H., SHINDYALOV, I. N. & BOURNE, P. E. 2000. The Protein Data Bank. *Nucleic Acids Res*, 28, 235-42.
- BHALLA, K. & GRANT, S. 1987. Effect of deoxycytidine on the in vitro response of human leukemia cells to inhibitors of de novo pyrimidine biosynthesis. *Cancer Chemother Pharmacol*, 19, 226-32.
- BHATIA, M. B. & GRUBMEYER, C. 1993. The role of divalent magnesium in activating the reaction catalyzed by orotate phosphoribosyltransferase. *Arch Biochem Biophys*, 303, 321-5.
- BHATIA, M. B., VINITSKY, A. & GRUBMEYER, C. 1990. Kinetic mechanism of orotate phosphoribosyltransferase from *Salmonella typhimurium*. *Biochemistry*, 29, 10480-7.
- BIASINI, M., BIENERT, S., WATERHOUSE, A., ARNOLD, K., STUDER, G., SCHMIDT, T., KIEFER, F., CASSARINO, T. G., BERTONI, M., BORDOLI, L. & SCHWEDE, T. 2014. SWISS-MODEL: modelling protein tertiary and quaternary structure using evolutionary information. *Nucleic Acids Research*, 42, W252-W258.
- BISMUTH, G., THUILLIER, L., PERIGNON, J. L. & CARTIER, P. H. 1982. Uridine as the only alternative to pyrimidine de novo synthesis in rat T lymphocytes. *FEBS Lett*, 148, 135-9.
- BLACKMOND, D. G. 2005. Reaction progress kinetic analysis: a powerful methodology for mechanistic studies of complex catalytic reactions. *Angew Chem Int Ed Engl*, 44, 4302-20.
- BODDEY, J. A., HODDER, A. N., GUNTHER, S., GILSON, P. R., PATSIOURAS, H., KAPP, E. A., PEARCE, J. A., DE KONING-WARD, T. F., SIMPSON, R. J., CRABB, B. S. & COWMAN, A. F. 2010. An aspartyl protease directs malaria effector proteins to the host cell. *Nature*, 463, 627-31.
- BOLOGA, C., OLAH, M. & OPREA, T. 2006. Chemical Database Preparation for Compound Acquisition or Virtual Screening. In: LARSON, R. (ed.) *Bioinformatics and Drug Discovery*. Humana Press.
- BONAVIA, A., FRANTI, M., PUSATERI KEANEY, E., KUHEN, K., SEEPERSAUD, M., RADETICH, B., SHAO, J., HONDA, A., DEWHURST, J., BALABANIS, K., MONROE, J., WOLFF, K., OSBORNE, C., LANIERI, L., HOFFMASTER, K., AMIN, J., MARKOVITS, J., BROOME, M., SKUBA, E., CORNELLA-TARACIDO, I., JOBERTY, G., BOUWMEESTER, T., HAMANN, L.,

- TALLARICO, J. A., TOMMASI, R., COMPTON, T. & BUSHELL, S. M. 2011. Identification of broad-spectrum antiviral compounds and assessment of the druggability of their target for efficacy against respiratory syncytial virus (RSV). *Proc Natl Acad Sci U S A*, 108, 6739-44.
- BRADFORD, M. M. 1976. A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal Biochem*, 72, 248-54.
- BREDA, A., MACHADO, P., ROSADO, L. A., SOUTO, A. A., SANTOS, D. S. & BASSO, L. A. 2012. Pyrimidin-2(1H)-ones based inhibitors of Mycobacterium tuberculosis orotate phosphoribosyltransferase. *Eur J Med Chem*, 54, 113-22.
- BROWN, N., ZEHENDER, H., AZZAOU, K., SCHUFFENHAUER, A., MAYR, L. M. & JACOBY, E. 2006. A cheminformatics analysis of hit lists obtained from high-throughput affinity-selection screening. *J Biomol Screen*, 11, 123-30.
- BRUS, B., KOŠAK, U., TURK, S., PIŠLAR, A., COQUELLE, N., KOS, J., STOJAN, J., COLLETIER, J.-P. & GOBEC, S. 2014. Discovery, Biological Evaluation, and Crystal Structure of a Novel Nanomolar Selective Butyrylcholinesterase Inhibitor. *Journal of Medicinal Chemistry*, 57, 8167-8179.
- BRYLINSKI, M. 2013. Nonlinear Scoring Functions for Similarity-Based Ligand Docking and Binding Affinity Prediction. *Journal of Chemical Information and Modeling*, 53, 3097-3112.
- BURST, V. & TESCHNER, S. 2010. Leflunomide: a drug with a potential beyond rheumatology. *Immunotherapy*, 2, 637+.
- CARLSON, H. A., MASUKAWA, K. M., RUBINS, K., BUSHMAN, F. D., JORGENSEN, W. L., LINS, R. D., BRIGGS, J. M. & MCCAMMON, J. A. 2000. Developing a dynamic pharmacophore model for HIV-1 integrase. *J Med Chem*, 43, 2100-14.
- CHEN, J. & LAI, L. 2006. Pocket v.2: further developments on receptor-based pharmacophore modeling. *J Chem Inf Model*, 46, 2684-91.
- CHIEN, J., CAMPIONI, M., SHRIDHAR, V. & BALDI, A. 2009. HtrA serine proteases as potential therapeutic targets in cancer. *Curr Cancer Drug Targets*, 9, 451-68.
- CHRISTOPHER, J. A., BROWN, J., DORE, A. S., ERREY, J. C., KOGLIN, M., MARSHALL, F. H., MYSZKA, D. G., RICH, R. L., TATE, C. G., TEHAN, B., WARNE, T. & CONGREVE, M. 2013. Biophysical fragment screening of the beta1-adrenergic receptor: identification of high affinity arylpiperazine leads using structure-based drug design. *J Med Chem*, 56, 3446-55.
- CHUNG, S. Y. & SUBBIAH, S. 1996. A structural explanation for the twilight zone of protein sequence homology. *Structure*, 4, 1123-7.
- CINQUIN, O., CHRISTOPHERSON, R. I. & MENZ, R. I. 2001. A hybrid plasmid for expression of toxic malarial proteins in Escherichia coli. *Mol Biochem Parasitol*, 117, 245-7.
- CONGREVE, M., CARR, R., MURRAY, C. & JHOTI, H. 2003. A 'rule of three' for fragment-based lead discovery? *Drug Discov Today*, 8, 876-7.
- COPELAND, R. A., MARCINKEVICIENE, J., HAQUE, T. S., KOPCHO, L. M., JIANG, W., WANG, K., ECRET, L. D., SIZEMORE, C., AMSLER, K. A., FOSTER, L., TADESSE, S., COMBS, A. P., STERN, A. M., TRAINOR, G. L., SLEE, A., ROGERS, M. J. & HOBBS, F. 2000. Helicobacter pylori-selective antibacterials based on inhibition of pyrimidine biosynthesis. *J Biol Chem*, 275, 33373-8.
- COREY, R. B. & PAULING, L. 1953. Molecular Models of Amino Acids, Peptides, and Proteins. *Review of Scientific Instruments*, 24, 621-627.
- COUSTOU, V., BIRAN, M., BESTEIRO, S., RIVIERE, L., BALTZ, T., FRANCONI, J. M. & BRINGAUD, F. 2006. Fumarate is an essential intermediary metabolite produced by the procyclic Trypanosoma brucei. *J Biol Chem*, 281, 26832-46.
- CRANDALL, I. E., WASILEWSKI, E., BELLO, A. M., MOHMMED, A., MALHOTRA, P., PAI, E. F., KAIN, K. C. & KOTRA, L. P. 2013. Antimalarial activities of 6-iodouridine and its prodrugs and potential for combination therapy. *J Med Chem*, 56, 2348-58.

- CRISMAN, T. J., PARKER, C. N., JENKINS, J. L., SCHEIBER, J., THOMA, M., KANG, Z. B., KIM, R., BENDER, A., NETTLES, J. H., DAVIES, J. W. & GLICK, M. 2007. Understanding false positives in reporter gene assays: in silico chemogenomics approaches to prioritize cell-based HTS data. *J Chem Inf Model*, 47, 1319-27.
- CROSS, J. B., THOMPSON, D. C., RAI, B. K., BABER, J. C., FAN, K. Y., HU, Y. & HUMBLET, C. 2009. Comparison of several molecular docking programs: pose prediction and virtual screening accuracy. *J Chem Inf Model*, 49, 1455-74.
- CROSS, S., ORTUSO, F., BARONI, M., COSTA, G., DISTINTO, S., MORACA, F., ALCARO, S. & CRUCIANI, G. 2012. GRID-Based Three-Dimensional Pharmacophores II: PharmBench, a Benchmark Data Set for Evaluating Pharmacophore Elucidation Methods. *Journal of Chemical Information and Modeling*, 52, 2599-2608.
- CRUZ-MONTEAGUDO, M., CORDEIRO, M. N., TEJERA, E., DOMINGUEZ, E. R. & BORGES, F. 2012. Desirability-based multi-objective QSAR in drug discovery. *Mini Rev Med Chem*, 12, 920-35.
- CURPAN, R., AVRAM, S., VIANELLO, R. & BOLOGA, C. 2014. Exploring the biological promiscuity of high-throughput screening hits through DFT calculations. *Bioorg Med Chem*, 22, 2461-8.
- DAI, L., WEI, X. N., ZHENG, D. H., MO, Y. Q., PESSLER, F. & ZHANG, B. Y. 2011. Effective treatment of Kimura's disease with leflunomide in combination with glucocorticoids. *Clin Rheumatol*, 30, 859-65.
- DAYLIGHT. 2011. *Daylight Theory Manual: 6. Fingerprints - Screening and Similarity* [Online]. Daylight Chemical Information Systems, Inc. Available: <http://www.daylight.com/dayhtml/doc/theory/theory.finger.html>.
- DE ABERASTURI, A. L. & CALVO, A. 2015. TMPRSS4: an emerging potential therapeutic target in cancer. *Br J Cancer*, 112, 4-8.
- DESAI, B. J., WOOD, B. M., FEDOROV, A. A., FEDOROV, E. V., GORYANOVA, B., AMYES, T. L., RICHARD, J. P., ALMO, S. C. & GERLT, J. A. 2012. Conformational changes in orotidine 5'-monophosphate decarboxylase: a structure-based explanation for how the 5'-phosphate group activates the enzyme. *Biochemistry*, 51, 8665-78.
- DIXIT, A. & VERKHIVKER, G. M. 2012. Integrating Ligand-Based and Protein-Centric Virtual Screening of Kinase Inhibitors Using Ensembles of Multiple Protein Kinase Genes and Conformations. *Journal of Chemical Information and Modeling*, 52, 2501-2515.
- DOHERTY, J. P., LINDEMAN, R., TRENT, R. J., GRAHAM, M. W. & WOODCOCK, D. M. 1993. Escherichia coli host strains SURE and SRB fail to preserve a palindrome cloned in lambda phage: improved alternate host strains. *Gene*, 124, 29-35.
- DOMAN, T. N., MCGOVERN, S. L., WITHERBEE, B. J., KASTEN, T. P., KURUMBAIL, R., STALLINGS, W. C., CONNOLLY, D. T. & SHOICHET, B. K. 2002. Molecular docking and high-throughput screening for novel inhibitors of protein tyrosine phosphatase-1B. *J Med Chem*, 45, 2213-21.
- DONDORP, A. M., NOSTEN, F., YI, P., DAS, D., PHYO, A. P., TARNING, J., LWIN, K. M., ARIEY, F., HANPITHAKPONG, W., LEE, S. J., RINGWALD, P., SILAMUT, K., IMWONG, M., CHOTIVANICH, K., LIM, P., HERDMAN, T., AN, S. S., YEUNG, S., SINGHASIVANON, P., DAY, N. P. J., LINDEGARDH, N., SOCHEAT, D. & WHITE, N. J. 2009. Artemisinin Resistance in Plasmodium falciparum Malaria. *New England Journal of Medicine*, 361, 455-467.
- DRUKER, B. J., TAMURA, S., BUCHDUNGER, E., OHNO, S., SEGAL, G. M., FANNING, S., ZIMMERMANN, J. & LYDON, N. B. 1996. Effects of a selective inhibitor of the Abl tyrosine kinase on the growth of Bcr-Abl positive cells. *Nat Med*, 2, 561-6.
- DRWAL, M. N. & GRIFFITH, R. 2013. Combination of ligand- and structure-based methods in virtual screening. *Drug Discov Today Technol*, 10, e395-401.
- DUAN, J., DIXON, S. L., LOWRIE, J. F. & SHERMAN, W. 2010. Analysis and comparison of 2D fingerprints: Insights into database screening performance using eight fingerprint methods. *Journal of Molecular Graphics and Modelling*, 29, 157-170.

- EL-DEEB, I. M., YOO, K. H. & LEE, S. H. 2011. ROS receptor tyrosine kinase: a new potential target for anticancer drugs. *Med Res Rev*, 31, 794-818.
- ENGH, R. A. & HUBER, R. 1991. Accurate bond and angle parameters for X-ray protein structure refinement. *Acta Cryst.*, A47, 392-400.
- ERLANSON, D. A., BRAISTED, A. C., RAPHAEL, D. R., RANDAL, M., STROUD, R. M., GORDON, E. M. & WELLS, J. A. 2000. Site-directed ligand discovery. *Proceedings of the National Academy of Sciences*, 97, 9367-9372.
- ERTL, P., ROHDE, B. & SELZER, P. 2000. Fast calculation of molecular polar surface area as a sum of fragment-based contributions and its application to the prediction of drug transport properties. *J Med Chem*, 43, 3714-7.
- FISER, A., DO, R. K. & SALI, A. 2000. Modeling of loops in protein structures. *Protein Sci*, 9, 1753-73.
- FOGOLARI, F., BRIGO, A. & MOLINARI, H. 2002. The Poisson-Boltzmann equation for biomolecular electrostatics: a tool for structural biology. *J Mol Recognit*, 15, 377-92.
- FRENCH, J. B., YATES, P. A., SOYSA, D. R., BOITZ, J. M., CARTER, N. S., CHANG, B., ULLMAN, B. & EALICK, S. E. 2011. The *Leishmania donovani* UMP synthase is essential for promastigote viability and has an unusual tetrameric structure that exhibits substrate-controlled oligomerization. *J Biol Chem*, 286, 20930-41.
- FRIESNER, R. A., BANKS, J. L., MURPHY, R. B., HALGREN, T. A., KLICIC, J. J., MAINZ, D. T., REPASKY, M. P., KNOLL, E. H., SHELLEY, M., PERRY, J. K., SHAW, D. E., FRANCIS, P. & SHENKIN, P. S. 2004. Glide: A New Approach for Rapid, Accurate Docking and Scoring. 1. Method and Assessment of Docking Accuracy. *Journal of Medicinal Chemistry*, 47, 1739-1749.
- GAGNON, R. & PETERSON, J. 1998. Estimation of Confidence Intervals for Area Under the Curve from Destructively Obtained Pharmacokinetic Data. *Journal of Pharmacokinetics and Biopharmaceutics*, 26, 87-102.
- GARDNER, M. J., HALL, N., FUNG, E., WHITE, O., BERRIMAN, M., HYMAN, R. W., CARLTON, J. M., PAIN, A., NELSON, K. E., BOWMAN, S., PAULSEN, I. T., JAMES, K., EISEN, J. A., RUTHERFORD, K., SALZBERG, S. L., CRAIG, A., KYES, S., CHAN, M. S., NENE, V., SHALLOM, S. J., SUH, B., PETERSON, J., ANGIUOLI, S., PERTEA, M., ALLEN, J., SELENGUT, J., HAFT, D., MATHER, M. W., VAIDYA, A. B., MARTIN, D. M., FAIRLAMB, A. H., FRAUNHOLZ, M. J., ROOS, D. S., RALPH, S. A., MCFADDEN, G. I., CUMMINGS, L. M., SUBRAMANIAN, G. M., MUNGALL, C., VENTER, J. C., CARUCCI, D. J., HOFFMAN, S. L., NEWBOLD, C., DAVIS, R. W., FRASER, C. M. & BARRELL, B. 2002. Genome sequence of the human malaria parasite *Plasmodium falciparum*. *Nature*, 419, 498-511.
- GASTEIGER, J. & MARSILI, M. 1980. Iterative partial equalization of orbital electronegativity—a rapid access to atomic charges. *Tetrahedron*, 36, 3219-3228.
- GE, H., WANG, Y., LI, C., CHEN, N., XIE, Y., XU, M., HE, Y., GU, X., WU, R., GU, Q., ZENG, L. & XU, J. 2013. Molecular Dynamics-Based Virtual Screening: Accelerating the Drug Discovery Process by High-Performance Computing. *Journal of Chemical Information and Modeling*, 53, 2757-2764.
- GOITEIN, R. K., CHELSKY, D. & PARSONS, S. M. 1978. Primary 14C and alpha secondary 3H substrate kinetic isotope effects for some phosphoribosyltransferases. *Journal of Biological Chemistry*, 253, 2963-2971.
- GOLDMAN, R. C., FROST, D. J., CAPOBIANCO, J. O., KADAM, S., RASMUSSEN, R. R. & ABAD-ZAPATERO, C. 1995. Antifungal drug targets: *Candida* secreted aspartyl protease and fungal wall beta-glucan synthesis. *Infect Agents Dis*, 4, 228-47.
- GONZALEZ-SEGURA, L., WITTE, J. F., MCCLARD, R. W. & HURLEY, T. D. 2007. Ternary complex formation and induced asymmetry in orotate phosphoribosyltransferase. *Biochemistry*, 46, 14075-86.
- GOOD, A. & OPREA, T. 2008. Optimization of CAMD techniques 3. Virtual screening enrichment studies: a help or hindrance in tool selection? *Journal of Computer-Aided Molecular Design*, 22, 169-178.

- GORYANOVA, B., AMYES, T. L., GERLT, J. A. & RICHARD, J. P. 2011. OMP decarboxylase: phosphodianion binding energy is used to stabilize a vinyl carbanion intermediate. *J Am Chem Soc*, 133, 6545-8.
- GRANT, J. A., GALLARDO, M. A. & PICKUP, B. T. 1996. A fast method of molecular shape comparison: A simple application of a Gaussian description of molecular shape. *Journal of Computational Chemistry*, 17, 1653-1666.
- GRANT, S. G., JESSEE, J., BLOOM, F. R. & HANAHAN, D. 1990. Differential plasmid rescue from transgenic mouse DNAs into Escherichia coli methylation-restriction mutants. *Proc Natl Acad Sci U S A*, 87, 4645-9.
- HAIBEL, H., RUDWALEIT, M., BRAUN, J. & SIEPER, J. 2005. Six months open label trial of leflunomide in active ankylosing spondylitis. *Ann Rheum Dis*, 64, 124-6.
- HAJDUK, P. J., DINGES, J., SCHKERYANTZ, J. M., JANOWICK, D., KAMINSKI, M., TUFANO, M., AUGERI, D. J., PETROS, A., NIENABER, V., ZHONG, P., HAMMOND, R., COEN, M., BEUTEL, B., KATZ, L. & FESIK, S. W. 1999. Novel inhibitors of Erm methyltransferases from NMR and parallel synthesis. *J Med Chem*, 42, 3852-9.
- HALAVATY, A. S., MINASOV, G., SHUVALOVA, L., DUBROVSKA, I., WINSOR, J., PAPAIZISI, L. & ANDERSON, W. F. 2011. 1.8 Angstrom resolution crystal structure of orotidine 5'-phosphate decarboxylase (pyrF) from Campylobacter jejuni subsp. jejuni NCTC 11168. PDB ID: 3RU6. Center for Structural Genomics of Infectious Diseases. UNPUBLISHED WORK.
- HALAVATY, A. S., SHUVALOVA, L., MINASOV, G., DUBROVSKA, I., WINSOR, J., GLASS, E. M., PETERSON, S. N. & ANDERSON, W. F. 2010. 1.77 Angstrom resolution crystal structure of orotidine 5'-phosphate decarboxylase from Vibrio cholerae O1 biovar eltor str. N16961. PDB ID: 3LDV. Center for Structural Genomics of Infectious Diseases. UNPUBLISHED WORK.
- HALGREN, T. A., MURPHY, R. B., FRIESNER, R. A., BEARD, H. S., FRYE, L. L., POLLARD, W. T. & BANKS, J. L. 2004. Glide: a new approach for rapid, accurate docking and scoring. 2. Enrichment factors in database screening. *J Med Chem*, 47, 1750-9.
- HAN, B. D., LIVINGSTONE, L. R., PASEK, D. A., YABLONSKI, M. J. & JONES, M. E. 1995. Human uridine monophosphate synthase: baculovirus expression, immunoaffinity column purification and characterization of the acetylated amino terminus. *Biochemistry*, 34, 10835-43.
- HASHIMOTO, M., MORALES, J., FUKAI, Y., SUZUKI, S., TAKAMIYA, S., TSUBOUCHI, A., INOUE, S., INOUE, M., KITA, K., HARADA, S., TANAKA, A., AOKI, T. & NARA, T. 2012. Critical importance of the de novo pyrimidine biosynthesis pathway for Trypanosoma cruzi growth in the mammalian host cell cytoplasm. *Biochem Biophys Res Commun*, 417, 1002-6.
- HAWKINS, P. C. D., SKILLMAN, A. G., WARREN, G. L., ELLINGSON, B. A. & STAHL, M. T. 2010. Conformer Generation with OMEGA: Algorithm and Validation Using High Quality Structures from the Protein Databank and Cambridge Structural Database. *Journal of Chemical Information and Modeling*, 50, 572-584.
- HAYNES, M. 2008. *In silico screening for novel Plasmodium falciparum Orotidine 5'-monophosphate decarboxylase inhibitors*. Bachelor of Science (Honours), Flinders University.
- HEINRICH, D., DIEDERICHSEN, U. & RUDOLPH, M. G. 2009. Lys314 is a nucleophile in non-classical reactions of orotidine-5'-monophosphate decarboxylase. *Chemistry*, 15, 6619-25.
- HENIKOFF, S. & HENIKOFF, J. G. 1992. Amino acid substitution matrices from protein blocks. *Proc Natl Acad Sci U S A*, 89, 10915-9.
- HENRIKSEN, A., AGHAJARI, N., JENSEN, K. F. & GAJHEDE, M. 1996. A flexible loop at the dimer interface is a part of the active site of the adjacent monomer of Escherichia coli orotate phosphoribosyltransferase. *Biochemistry*, 35, 3803-9.

- HOFFMANN, H. H., KUNZ, A., SIMON, V. A., PALESE, P. & SHAW, M. L. 2011. Broad-spectrum antiviral that interferes with de novo pyrimidine biosynthesis. *Proc Natl Acad Sci U S A*, 108, 5777-82.
- HOLLAND, J. W., GERO, A. M. & O'SULLIVAN, W. J. 1983. Enzymes of de novo pyrimidine biosynthesis in *Babesia rodhaini*. *J Protozool*, 30, 36-40.
- HOUSTON, D. R. & WALKINSHAW, M. D. 2013. Consensus Docking: Improving the Reliability of Docking in a Virtual Screening Context. *Journal of Chemical Information and Modeling*, 53, 384-390.
- HR. 2001. Hampton Research: Crystal Growth 101: Preliminary Sample Preparation. Available: https://hamptonresearch.com/documents/growth_101/1.pdf.
- HU, H., BOONE, A. & YANG, W. 2008. Mechanism of OMP decarboxylation in orotidine 5'-monophosphate decarboxylase. *J Am Chem Soc*, 130, 14493-503.
- HUANG, N., SHOICHET, B. K. & IRWIN, J. J. 2006. Benchmarking Sets for Molecular Docking. *Journal of Medicinal Chemistry*, 49, 6789-6801.
- IHLENFELDT, W. D., TAKAHASHI, Y., ABE, H. & SASAKI, S. 1994. Computation and Management of Chemical Properties in CACTVS: An extensible Networked Approach toward Modularity and Flexibility. *J. Chem. Inf. Comp. Sci.*, 34, 109-116.
- IMPRASITTICHAIL, W., ROYTRAKUL, S., KRUNGKRAI, S. R. & KRUNGKRAI, J. 2014. A unique insertion of low complexity amino acid sequence underlies protein-protein interaction in human malaria parasite orotate phosphoribosyltransferase and orotidine 5'-monophosphate decarboxylase. *Asian Pacific Journal of Tropical Medicine*, 7, 184-192.
- IRWIN, J. J. & SHOICHET, B. K. 2005. ZINC--a free database of commercially available compounds for virtual screening. *J Chem Inf Model*, 45, 177-82.
- IWASAKI, H., TAKAHAGI, M., SHIBA, T., NAKATA, A. & SHINAGAWA, H. 1991. Escherichia coli RuvC protein is an endonuclease that resolves the Holliday structure. *Embo j*, 10, 4381-9.
- JACCARD, P. 1901. Étude comparative de la distribution florale dans une portion des Alpes et des Jura. *Bulletin del la Société Vaudoise des Sciences Naturelles*, 37, 547-579.
- JAIN, A. N. & NICHOLLS, A. 2008. Recommendations for evaluation of computational methods. *J Comput Aided Mol Des*, 22, 133-9.
- KALÁSZI, A., SZISZ, D., IMRE, G. & POLGÁR, T. 2014. Screen3D: A Novel Fully Flexible High-Throughput Shape-Similarity Search Method. *Journal of Chemical Information and Modeling*, 54, 1036-1049.
- KANALAS, J. J. & SUTTLE, D. P. 1984. Amplification of the UMP synthase gene and enzyme overproduction in pyrazofurin-resistant rat hepatoma cells. Molecular cloning of a cDNA for UMP synthase. *J Biol Chem*, 259, 1848-53.
- KANCHANAPHUM, P. & KRUNGKRAI, J. 2009. Kinetic benefits and thermal stability of orotate phosphoribosyltransferase and orotidine 5'-monophosphate decarboxylase enzyme complex in human malaria parasite Plasmodium falciparum. *Biochemical and Biophysical Research Communications*, 390, 337-341.
- KANEHISA LABORATORIES. 2014a. *Kyoto Encyclopedia of Genes and Genomes: Pyrimidine Metabolism Pathway Map* [Online]. Available: www.genome.jp/kegg/pathway/map/map00240.html.
- KANEHISA LABORATORIES. 2014b. *Kyoto Encyclopedia of Genes and Genomes: β -alanine Metabolism Pathway Map* [Online]. Available: www.genome.jp/kegg/pathway/map/map00410.html.
- KANTARDJIEFF, K. A., VASQUEZ, C., CASTRO, P., WARFEL, N. M., RHO, B. S., LEKIN, T., KIM, C. Y., SEGELKE, B. W., TERWILLIGER, T. C. & RUPP, B. 2005. Structure of pyrR (Rv1379) from *Mycobacterium tuberculosis*: a persistence gene and protein drug target. *Acta Crystallogr D Biol Crystallogr*, 61, 355-64.
- KENNY, B. A., BUSHFIELD, M., PARRY-SMITH, D. J., FOGARTY, S. & TREHERNE, J. M. 1998. The application of high-throughput screening to novel lead discovery. *Prog Drug Res*, 51, 245-69.

- KESERU, G. M. & MAKARA, G. M. 2006. Hit discovery and hit-to-lead approaches. *Drug Discov Today*, 11, 741-8.
- KIRKPATRICK, P. & ELLIS, C. 2004. Chemical space. *Nature*, 432, 823-823.
- KOGO, H., INAGAKI, H., OHYE, T., KATO, T., EMANUEL, B. S. & KURAHASHI, H. 2007. Cruciform extrusion propensity of human translocation-mediating palindromic AT-rich repeats. *Nucleic Acids Research*, 35, 1198-1208.
- KOLB, H. C., FINN, M. G. & SHARPLESS, K. B. 2001. Click Chemistry: Diverse Chemical Function from a Few Good Reactions. *Angew Chem Int Ed Engl*, 40, 2004-2021.
- KOLTUN, W. L. 1965. Space filling atomic units and connectors for molecular models. Google Patents.
- KRÜGER, D. M. & EVERS, A. 2010. Comparison of Structure- and Ligand-Based Virtual Screening Protocols Considering Hit List Complementarity and Enrichment Factors. *ChemMedChem*, 5, 148-158.
- KRUNGKRAI, S. R., AOKI, S., PALACPAC, N. M. Q., SATO, D., MITAMURA, T., KRUNGKRAI, J. & HORII, T. 2004a. Human malaria parasite orotate phosphoribosyltransferase: functional expression, characterization of kinetic reaction mechanism and inhibition profile. *Molecular and Biochemical Parasitology*, 134, 245-255.
- KRUNGKRAI, S. R., DELFRAINO, B. J., SMILEY, J. A., PRAPUNWATTANA, P., MITAMURA, T., HORII, T. & KRUNGKRAI, J. 2005. A novel enzyme complex of orotate phosphoribosyltransferase and orotidine 5'-monophosphate decarboxylase in human malaria parasite *Plasmodium falciparum*: physical association, kinetics, and inhibition characterization. *Biochemistry*, 44, 1643-52.
- KRUNGKRAI, S. R., PRAPUNWATTANA, P., HORII, T. & KRUNGKRAI, J. 2004b. Orotate phosphoribosyltransferase and orotidine 5'-monophosphate decarboxylase exist as multienzyme complex in human malaria parasite *Plasmodium falciparum*. *Biochemical and Biophysical Research Communications*, 318, 1012-1018.
- KUEHN, D. 2003. *Cloning and Heterologous Overexpression of de novo Pyrimidine Biosynthetic Enzymes from Plasmodium falciparum*. BSc (Honours), Flinders University.
- LANG, P. T., MOUSTAKAS, D., BROZELL, S., CARRASCAL, N., MUKHERJEE, S., BALIUS, T., ALLEN, W. J., HOLDEN, P., PEGG, S., RAHA, K., SHIVAKUMAR, D., RIZZO, R., CASE, D., SHOICHET, B. & KUNTZ, I. 2015. *DOCK 6.7 Users Manual* [Online]. Available: http://dock.compbio.ucsf.edu/DOCK_6/dock6_manual.htm.
- LANGLEY, D. B., SHOJAEI, M., CHAN, C., LOK, H. C., MACKAY, J. P., TRAUT, T. W., GUSS, J. M. & CHRISTOPHERSON, R. I. 2008. Structure and inhibition of orotidine 5'-monophosphate decarboxylase from *Plasmodium falciparum*. *Biochemistry*, 47, 3842-54.
- LARKIN, M. A., BLACKSHIELDS, G., BROWN, N. P., CHENNA, R., MCGETTIGAN, P. A., MCWILLIAM, H., VALENTIN, F., WALLACE, I. M., WILM, A., LOPEZ, R., THOMPSON, J. D., GIBSON, T. J. & HIGGINS, D. G. 2007. Clustal W and Clustal X version 2.0. *Bioinformatics*, 23, 2947-8.
- LASKOWSKI, R. A., MACARTHUR, M. W., MOSS, D. S. & THORNTON, J. M. 1993. PROCHECK: a program to check the stereochemical quality of protein structures. *Journal of Applied Crystallography*, 26, 283-291.
- LENNOX, E. S. 1955. Transduction of linked genetic characters of the host by bacteriophage P1. *Virology*, 1, 190-206.
- LEWIS, M., MEZA-AVINA, M. E., WEI, L., CRANDALL, I. E., BELLO, A. M., PODUCH, E., LIU, Y., PAIGE, C. J., KAIN, K. C., PAI, E. F. & KOTRA, L. P. 2011. Novel interactions of fluorinated nucleotide derivatives targeting orotidine 5'-monophosphate decarboxylase. *J Med Chem*, 54, 2891-901.
- LIN, J., JULIANO, J. & WONGSRICHANALAI, C. 2010. Drug-Resistant Malaria: The Era of ACT. *Current Infectious Disease Reports*, 12, 165-173.
- LINEWEAVER, H. & BURK, D. 1934. The Determination of Enzyme Dissociation Constants. *Journal of the American Chemical Society*, 56, 658-666.

- LIPINSKI, C. A., LOMBARDO, F., DOMINY, B. W. & FEENEY, P. J. 2001. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv Drug Deliv Rev*, 46, 3-26.
- LITTLEFIELD, J. W. 1966. The periodic synthesis of thymidine kinase in mouse fibroblasts. *Biochim Biophys Acta*, 114, 398-403.
- LIU, G., XIN, Z., LIANG, H., ABAD-ZAPATERO, C., HAJDUK, P. J., JANOWICK, D. A., SZCZEPANKIEWICZ, B. G., PEI, Z., HUTCHINS, C. W., BALLARON, S. J., STASHKO, M. A., LUBBEN, T. H., BERG, C. E., RONDINONE, C. M., TREVILLYAN, J. M. & JIROUSEK, M. R. 2003. Selective protein tyrosine phosphatase 1B inhibitors: targeting the second phosphotyrosine binding site with non-carboxylic acid-containing ligands. *J Med Chem*, 46, 3437-40.
- LIVINGSTONE, L. R. & JONES, M. E. 1987. The purification and preliminary characterization of UMP synthase from human placenta. *J Biol Chem*, 262, 15726-33.
- LLOYD, S. 1982. Least squares quantization in PCM. *Information Theory, IEEE Transactions on*, 28, 129-137.
- LOVING, K., SALAM, N. & SHERMAN, W. 2009. Energetic analysis of fragment docking and application to structure-based pharmacophore hypothesis generation. *Journal of Computer-Aided Molecular Design*, 23, 541-554.
- LUQUE, I. 2010. Biophysics of Protein-Protein Interactions: 2.4 Thermodynamically Driven Drug Design. *Protein Surface Recognition*. John Wiley & Sons, Ltd.
- MAKIUCHI, T., NARA, T., ANNOURA, T., HASHIMOTO, T. & AOKI, T. 2007. Occurrence of multiple, independent gene fusion events for the fifth and sixth enzymes of pyrimidine biosynthesis in different eukaryotic groups. *Gene*, 394, 78-86.
- MARCINKEVICIENE, J., ROGERS, M. J., KOPCHO, L., JIANG, W., WANG, K., MURPHY, D. J., LIPPY, J., LINK, S., CHUNG, T. D., HOBBS, F., HAQUE, T., TRAINOR, G. L., SLEE, A., STERN, A. M. & COPELAND, R. A. 2000. Selective inhibition of bacterial dihydroorotate dehydrogenases by thiadiazolidinediones. *Biochem Pharmacol*, 60, 339-42.
- MARSCHALL, M., NIEMANN, I., KOSULIN, K., BOOTZ, A., WAGNER, S., DOBNER, T., HERZ, T., KRAMER, B., LEBAN, J., VITT, D., STAMMINGER, T., HUTTERER, C. & STROBL, S. 2013. Assessment of drug candidates for broad-spectrum antiviral therapy targeting cellular pyrimidine biosynthesis. *Antiviral Res*, 100, 640-8.
- MCCLARD, R. W., HOLETS, E. A., MACKINNON, A. L. & WITTE, J. F. 2006. Half-of-sites binding of orotidine 5'-phosphate and alpha-D-5-phosphorylribose 1-diphosphate to orotate phosphoribosyltransferase from *Saccharomyces cerevisiae* supports a novel variant of the Theorell-Chance mechanism with alternating site catalysis. *Biochemistry*, 45, 5330-42.
- MCGANN, M. 2011. FRED pose prediction and virtual screening accuracy. *J Chem Inf Model*, 51, 578-96.
- MCGANN, M. 2012. FRED and HYBRID docking performance on standardized datasets. *J Comput Aided Mol Des*, 26, 897-906.
- MCINNES, C. 2007. Virtual screening strategies in drug discovery. *Curr Opin Chem Biol*, 11, 494-502.
- MENZ, R. I., CINQUIN, O. & CHRISTOPHERSON, R. I. 2002. The identification, cloning and functional expression of the gene encoding orotidine 5'-monophosphate (OMP) decarboxylase from *Plasmodium falciparum*. *Ann Trop Med Parasitol*, 96, 469-76.
- MERKWIRTH, C., MAUSER, H., SCHULZ-GASCH, T., ROCHE, O., STAHL, M. & LENGAUER, T. 2004. Ensemble methods for classification in cheminformatics. *J Chem Inf Comput Sci*, 44, 1971-8.
- MEYDAN, N., GRUNBERGER, T., DADI, H., SHAHAR, M., ARPAIA, E., LAPIDOT, Z., LEEDER, J. S., FREEDMAN, M., COHEN, A., GAZIT, A., LEVITZKI, A. & ROIFMAN, C. M. 1996. Inhibition of acute lymphoblastic leukaemia by a Jak-2 inhibitor. *Nature*, 379, 645-8.

- MEZA-AVINA, M. E., WEI, L., LIU, Y., PODUCH, E., BELLO, A. M., MISHRA, R. K., PAI, E. F. & KOTRA, L. P. 2010. Structural determinants for the inhibitory ligands of orotidine-5'-monophosphate decarboxylase. *Bioorg Med Chem*, 18, 4032-41.
- MIDGETT, C. R. & MADDEN, D. R. 2007. Breaking the bottleneck: eukaryotic membrane protein expression for high-resolution structural studies. *J Struct Biol*, 160, 265-74.
- MILES, E. W., RHEE, S. & DAVIES, D. R. 1999. The Molecular Basis of Substrate Channeling. *Journal of Biological Chemistry*, 274, 12193-12196.
- MILLER, B. G. & WOLFENDEN, R. 2002. Catalytic proficiency: the unusual case of OMP decarboxylase. *Annu Rev Biochem*, 71, 847-85.
- MINASOV, G., WAWRZAK, Z., RUAN, J., NGO, H., SHUVALOVA, L., DUBROVSKA, I. & ANDERSON, W. F. 2013. 2.75 Angstrom Resolution Crystal Structure of Putative Orotidine-monophosphate-decarboxylase from *Toxoplasma gondii*. PDB ID: 4MJZ. Center for Structural Genomics of Infectious Diseases. UNPUBLISHED WORK.
- MOCHE, M., FLODIN, S., NYMAN, T., STENMARK, P. & NORDLUND, P. 2007a. *The Crystal Structure of Human Orotidine-5'-Decarboxylase Domain of Human Uridine Monophosphate Synthetase (Umps)* [Online].
- MOCHE, M., OGG, D., ARROWSMITH, C., BERGLUND, H., BUSAM, R., COLLINS, R., DAHLGREN, L. G., EDWARDS, A., FLODIN, S., FLORES, A., GRASLUND, S., HAMMARSTROM, M., HALLBERG, B. M., HOLMBERG-SCHIAVONE, L., JOHANSSON, I., KALLAS, A., KARLBERG, T., KOTENYOVA, T., LEHTIO, L., NYMAN, T., PERSSON, C., SAGEMARK, J., STENMARK, P., SUNDSTROM, M., THORSELL, A. G., VAN-DEN-BERG, S., WEIGELT, J., WELIN, M. & NORDLUND, P. 2007b. The Crystal Structure of Human Orotidine 5'-Decarboxylase Domain of Human Uridine Monophosphate Synthetase (Umps). PDB ID: 2JGY. UNPUBLISHED WORK.
- MOCHE, M., ROOS, A., ARROWSMITH, C. H., BERGLUND, H., BOUNTRA, C., COLLINS, R., EDWARDS, A. M., FLODIN, S., FLORES, A., GRASLUND, S., HAMMARSTROM, M., JOHANSSON, A., JOHANSSON, I., KARLBERG, T., KOTYENOVA, T., KOTZCH, A., NIELSEN, T. K., NYMAN, T., PERSSON, C., SAGEMARK, J., SCHUELER, H., SCHUTZ, P., SIPONEN, M. I., SVENSSON, L., THORSELL, A. G., TRESAUGUES, L., VANDENBERG, S., WEIGELT, J., WELIN, M., WISNIEWSKA, M. & NORDLUND, P. 2009. Human Orotate Phosphoribosyltransferase (Oprtase) Domain of Uridine 5'-Monophosphate Synthase (Umps) in Complex with its Substrate Orotidine 5'-Monophosphate (Omp). PDB ID: 2WNS. UNPUBLISHED WORK.
- MORRIS, A., MACARTHUR, M., HUTCHINSON, E. & THORNTON, J. 1992. Stereochemical quality of protein structure coordinates. *Proteins 12: 345-64 (1992)*.
- MOSIMANN, S., MELESHKO, R. & JAMES, M. N. 1995. A critical assessment of comparative molecular modeling of tertiary structures of proteins. *Proteins*, 23, 301-17.
- MOUSTAKAS, D. T., LANG, P. T., PEGG, S., PETTERSEN, E., KUNTZ, I. D., BROOIJMANS, N. & RIZZO, R. C. 2006. Development and validation of a modular, extensible docking program: DOCK 5. *J Comput Aided Mol Des*, 20, 601-19.
- MUCHMORE, S. W., SOUERS, A. J. & AKRITOPOULOU-ZANZE, I. 2006. The Use of Three-Dimensional Shape and Electrostatic Similarity Searching in the Identification of a Melanin-Concentrating Hormone Receptor 1 Antagonist. *Chemical Biology & Drug Design*, 67, 174-176.
- NEEDLEMAN, S. B. & WUNSCH, C. D. 1970. A general method applicable to the search for similarities in the amino acid sequence of two proteins. *J Mol Biol*, 48, 443-53.
- NELSON, D. L., L., L. A. & M., C. M. 2008. *Lehninger principles of biochemistry*, New York, W.H. Freeman.
- NICHOLLS, A. 2008. What do we know and when do we know it? *Journal of Computer-Aided Molecular Design*, 22, 239-255.
- NISSINK, W. & BLACKBURN, S. 2014. Quantification of frequent-hitter behavior based on historical high-throughput screening data. *Future Med Chem*, 6, 1113-26.
- NYHAN, W. L. 2005. Disorders of purine and pyrimidine metabolism. *Mol Genet Metab*, 86, 25-33.

- OGUT, O. & JIN, J.-P. 2000. Cooperative Interaction between Developmentally Regulated Troponin T and Tropomyosin Isoforms in the Absence of F-actin. *Journal of Biological Chemistry*, 275, 26089-26095.
- OLTERS DORF, T., ELMORE, S. W., SHOEMAKER, A. R., ARMSTRONG, R. C., AUGERI, D. J., BELLI, B. A., BRUNCKO, M., DECKWERTH, T. L., DINGES, J., HAJDUK, P. J., JOSEPH, M. K., KITADA, S., KORSMEYER, S. J., KUNZER, A. R., LETAI, A., LI, C., MITTEN, M. J., NETTESHEIM, D. G., NG, S., NIMMER, P. M., O'CONNOR, J. M., OLEKSIJEW, A., PETROS, A. M., REED, J. C., SHEN, W., TAHIR, S. K., THOMPSON, C. B., TOMASELLI, K. J., WANG, B., WENDT, M. D., ZHANG, H., FESIK, S. W. & ROSENBERG, S. H. 2005. An inhibitor of Bcl-2 family proteins induces regression of solid tumours. *Nature*, 435, 677-81.
- OPREA, T. I., DAVIS, A. M., TEAGUE, S. J. & LEESON, P. D. 2001. Is there a difference between leads and drugs? A historical perspective. *J Chem Inf Comput Sci*, 41, 1308-15.
- ORITA, I., KITA, A., YURIMOTO, H., KATO, N., SAKAI, Y. & MIKI, K. 2010. Crystal structure of 3-hexulose-6-phosphate synthase, a member of the orotidine 5'-monophosphate decarboxylase suprafamily. *Proteins*, 78, 3488-92.
- PANSELINAS, E. & JUDSON, M. A. 2012. Acute pulmonary exacerbations of sarcoidosis. *Chest*, 142, 827-36.
- PARK, D., MAGIS, A. T., LI, R., OWONIKOKO, T. K., SICA, G. L., SUN, S. Y., RAMALINGAM, S. S., KHURI, F. R., CURRAN, W. J. & DENG, X. 2013. Novel small-molecule inhibitors of Bcl-XL to treat lung cancer. *Cancer Res*, 73, 5485-96.
- PETTERSEN, E. F., GODDARD, T. D., HUANG, C. C., COUCH, G. S., GREENBLATT, D. M., MENG, E. C. & FERRIN, T. E. 2004. UCSF Chimera--a visualization system for exploratory research and analysis. *J Comput Chem*, 25, 1605-12.
- PINTO, M., ORZAEZ MDEL, M., DELGADO-SOLER, L., PEREZ, J. J. & RUBIO-MARTINEZ, J. 2011. Rational design of new class of BH3-mimetics as inhibitors of the Bcl-xL protein. *J Chem Inf Model*, 51, 1249-58.
- PIRILDAR, T. 2003. Treatment of adult-onset Still's disease with leflunomide and chloroquine combination in two patients. *Clin Rheumatol*, 22, 157.
- PITMAN, M. R. & MENZ, R. I. 2006. Methods for protein homology modelling. In: DILIP K. ARORA, R. M. B. & GAUTAM, B. S. (eds.) *Applied Mycology and Biotechnology*. Elsevier.
- POWERS, R. A., MORANDI, F. & SHOICHET, B. K. 2002. Structure-based discovery of a novel, noncovalent inhibitor of AmpC beta-lactamase. *Structure*, 10, 1013-23.
- PRAGOBPOL, S., GERO, A. M., LEE, C. S. & O'SULLIVAN, W. J. 1984. Orotate phosphoribosyltransferase and orotidylate decarboxylase from *Crithidia luciliae*: subcellular location of the enzymes and a study of substrate channeling. *Arch Biochem Biophys*, 230, 285-93.
- PUROHIT, M. K., PODUCH, E., WEI, L. W., CRANDALL, I. E., TO, T., KAIN, K. C., PAI, E. F. & KOTRA, L. P. 2012. Novel cytidine-based orotidine-5'-monophosphate decarboxylase inhibitors with an unusual twist. *J Med Chem*, 55, 9988-97.
- QING, M., ZOU, G., WANG, Q. Y., XU, H. Y., DONG, H., YUAN, Z. & SHI, P. Y. 2010. Characterization of dengue virus resistance to brequinar in cell culture. *Antimicrob Agents Chemother*, 54, 3686-95.
- QUEEN, S. A., JAGT, D. L. & REYES, P. 1990. In vitro susceptibilities of *Plasmodium falciparum* to compounds which inhibit nucleotide metabolism. *Antimicrob Agents Chemother*, 34, 1393-8.
- RATHOD, P. K., KHATRI, A., HUBBERT, T. & MILHOUS, W. K. 1989. Selective activity of 5-fluoroorotic acid against *Plasmodium falciparum* in vitro. *Antimicrob Agents Chemother*, 33, 1090-4.
- RATHOD, P. K. & KUMAR, S. 2013. Crystal structure of *Plasmodium falciparum* orotate phosphoribosyltransferase. PDB ID: 4FYM. UNPUBLISHED WORK.

- REUTER, J. S. & MATHEWS, D. H. 2010. RNAstructure: software for RNA secondary structure prediction and analysis. *BMC Bioinformatics*, 11, 129.
- REYMOND, J. L. & AWALE, M. 2012. Exploring chemical space for drug discovery using the chemical universe database. *ACS Chem Neurosci*, 3, 649-57.
- RISHTON, G. M. 2003. Nonleadlikeness and leadlikeness in biochemical screening. *Drug Discov Today*, 8, 86-96.
- ROACH, M. 2007. *In silico Screening for structural inhibitors of Plasmodium falciparum Orotate Phosphoribosyltransferase*. (Honours), Flinders University.
- ROCHE, O., SCHNEIDER, P., ZUEGGE, J., GUBA, W., KANSY, M., ALANINE, A., BLEICHER, K., DANIEL, F., GUTKNECHT, E.-M., ROGERS-EVANS, M., NEIDHART, W., STALDER, H., DILLON, M., SJÖGREN, E., FOTOUHI, N., GILLESPIE, P., GOODNOW, R., HARRIS, W., JONES, P., TANIGUCHI, M., TSUJII, S., VON DER SAAL, W., ZIMMERMANN, G. & SCHNEIDER, G. 2001. Development of a Virtual Screening Method for Identification of "Frequent Hitters" in Compound Libraries. *Journal of Medicinal Chemistry*, 45, 137-142.
- ROST, B. 1999. Twilight zone of protein sequence alignments. *Protein Eng*, 12, 85-94.
- ROY, M. 2007. Early clinical experience with leflunomide in uveitis. *Can J Ophthalmol*, 42, 634.
- SAKAMOTO, E., NAGASE, H., KOBUNAI, T., OIE, S., OKA, T., FUKUSHIMA, M. & OKA, T. 2007. Orotate phosphoribosyltransferase expression level in tumors is a potential determinant of the efficacy of 5-fluorouracil. *Biochem Biophys Res Commun*, 363, 216-22.
- SALAM, N. K., NUTI, R. & SHERMAN, W. 2009. Novel Method for Generating Structure-Based Pharmacophores Using Energetic Analysis. *Journal of Chemical Information and Modeling*, 49, 2356-2368.
- SALI, A. & BLUNDELL, T. L. 1993. Comparative protein modelling by satisfaction of spatial restraints. *J Mol Biol*, 234, 779-815.
- SAMBROOK, J., FRITSCH, E. F. & MANIATIS, T. 1989. *Molecular Cloning: A Laboratory Manual*. 2nd ed, Plainview, N.Y., Cold Spring Harbor Laboratory Press.
- SANDERS, S. & HARISDANGKUL, V. 2002. Leflunomide for the treatment of rheumatoid arthritis and autoimmunity. *Am J Med Sci*, 323, 190-3.
- SASTRY, G. M., INAKOLLU, V. S. S. & SHERMAN, W. 2013. Boosting Virtual Screening Enrichments with Data Fusion: Coalescing Hits from Two-Dimensional Fingerprints, Shape, and Docking. *Journal of Chemical Information and Modeling*, 53, 1531-1542.
- SCHLAGENHAUF-LAWLOR, P. 2008. *Travelers' Malaria, 2nd ed., Chapter 6: The Parasite*, PMPH-USA.
- SCHLEGEL, J., SAMBADE, M. J., SATHER, S., MOSCHOS, S. J., TAN, A.-C., WINGES, A., DERYCKERE, D., CARSON, C. C., TREMBATH, D. G., TENTLER, J. J., ECKHARDT, S. G., KUAN, P.-F., HAMILTON, R. L., DUNCAN, L. M., MILLER, C. R., NIKOLAISHVILI-FEINBERG, N., MIDKIFF, B. R., LIU, J., ZHANG, W., YANG, C., WANG, X., FRYE, S. V., EARP, H. S., SHIELDS, J. M. & GRAHAM, D. K. 2013. MERTK receptor tyrosine kinase is a therapeutic target in melanoma. *The Journal of Clinical Investigation*, 123, 2257-2267.
- SCHORPP, K., ROTHENAIIGNER, I., SALMINA, E., REINSHAGEN, J., LOW, T., BRENKE, J. K., GOPALAKRISHNAN, J., TETKO, I. V., GUL, S. & HADIAN, K. 2013. Identification of Small-Molecule Frequent Hitters from AlphaScreen High-Throughput Screens. *J Biomol Screen*, 19, 715-726.
- SCHWARTZMAN, J. D. & PFEFFERKORN, E. R. 1981. Pyrimidine synthesis by intracellular *Toxoplasma gondii*. *J Parasitol*, 67, 150-8.
- SCIOR, T., BENDER, A., TRESADERN, G., MEDINA-FRANCO, J. L., MARTÍNEZ-MAYORGA, K., LANGER, T., CUANALO-CONTRERAS, K. & AGRAFIOTIS, D. K. 2012. Recognizing Pitfalls in Virtual Screening: A Critical Review. *Journal of Chemical Information and Modeling*, 52, 867-881.

- SCOTT, H. V., GERO, A. M. & O'SULLIVAN, W. J. 1986. In vitro inhibition of Plasmodium falciparum by pyrazofurin, an inhibitor of pyrimidine biosynthesis de novo. *Mol Biochem Parasitol*, 18, 3-15.
- SHIMOSAKA, M., FUKUDA, Y., MURATA, K. & KIMURA, A. 1985. Purification and Properties of Orotate Phosphoribosyltransferases from Escherichia coli K-12, and Its Derivative Purine-Sensitive Mutant. *Journal of Biochemistry*, 98, 1689-1697.
- SHOSTAK, K. & JONES, M. E. 1992. Orotidylate decarboxylase: insights into the catalytic mechanism from substrate specificity studies. *Biochemistry*, 31, 12155-61.
- SKORKO-GLONEK, J., ZURAWA-JANICKA, D., KOPER, T., JARZAB, M., FIGAJ, D., GLAZA, P. & LIPINSKA, B. 2013. HtrA protease family as therapeutic targets. *Curr Pharm Des*, 19, 977-1009.
- SMEE, D. F., HURST, B. L. & DAY, C. W. 2012. D282, a non-nucleoside inhibitor of influenza virus infection that interferes with de novo pyrimidine biosynthesis. *Antivir Chem Chemother*, 22, 263-72.
- STAHL, F. W. 1994. The Holliday junction on its thirtieth anniversary. *Genetics*, 138, 241-6.
- STENSON, C. 2011. *In silico Screening for Functional Inhibitors of Plasmodium falciparum Alternative NADH Dehydrogenase and cloning of the Plasmodium falciparum Alternative NADH Dehydrogenase*. Bachelor of Science (Honours), Flinders University.
- STRAWN, L. M., MCMAHON, G., APP, H., SCHRECK, R., KUCHLER, W. R., LONGHI, M. P., HUI, T. H., TANG, C., LEVITZKI, A., GAZIT, A., CHEN, I., KERI, G., ORFI, L., RISAU, W., FLAMME, I., ULLRICH, A., HIRTH, K. P. & SHAWVER, L. K. 1996. Flk-1 as a target for tumor growth inhibition. *Cancer Res*, 56, 3540-5.
- STUDIER, F. W. 1991. Use of bacteriophage T7 lysozyme to improve an inducible T7 expression system. *J Mol Biol*, 219, 37-44.
- STUDIER, F. W. & MOFFATT, B. A. 1986. Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes. *J Mol Biol*, 189, 113-30.
- SUCHI, M., MIZUNO, H., KAWAI, Y., TSUBOI, T., SUMI, S., OKAJIMA, K., HODGSON, M. E., OGAWA, H. & WADA, Y. 1997. Molecular cloning of the human UMP synthase gene and characterization of point mutations in two hereditary orotic aciduria families. *Am J Hum Genet*, 60, 525-39.
- SUN, H. P., JIA, J. M., JIANG, F., XU, X. L., LIU, F., GUO, X. K., CHERFAOUI, B., HUANG, H. Z., PAN, Y. & YOU, Q. D. 2014. Identification and optimization of novel Hsp90 inhibitors with tetrahydropyrido[4,3-d]pyrimidines core through shape-based screening. *Eur J Med Chem*, 79, 399-412.
- SVENSSON, F., KARLEN, A. & SKOLD, C. 2012. Virtual screening data fusion using both structure- and ligand-based methods. *J Chem Inf Model*, 52, 225-32.
- SWANN, S. L., BROWN, S. P., MUCHMORE, S. W., PATEL, H., MERTA, P., LOCKLEAR, J. & HAJDUK, P. J. 2011. A unified, probabilistic framework for structure- and ligand-based virtual screening. *J Med Chem*, 54, 1223-32.
- SZCZEPANKIEWICZ, B. G., LIU, G., HAJDUK, P. J., ABAD-ZAPATERO, C., PEI, Z., XIN, Z., LUBBEN, T. H., TREVILLYAN, J. M., STASHKO, M. A., BALLARON, S. J., LIANG, H., HUANG, F., HUTCHINS, C. W., FESIK, S. W. & JIROUSEK, M. R. 2003. Discovery of a potent, selective protein tyrosine phosphatase 1B inhibitor using a linked-fragment strategy. *J Am Chem Soc*, 125, 4087-96.
- TAKASHIMA, Y., MIZOHATA, E., KRUNGKRAI, S. R., FUKUNISHI, Y., KINOSHITA, T., SAKATA, T., MATSUMURA, H., KRUNGKRAI, J., HORII, T. & INOUE, T. 2012. The in silico screening and X-ray structure analysis of the inhibitor complex of Plasmodium falciparum orotidine 5'-monophosphate decarboxylase. *Journal of Biochemistry*, 152, 133-8.
- TAM TAM, S. 2007. *An in silico approach for the discovery of Plasmodium falciparum OMP decarboxylase inhibitors*. Masters Degree of Biotechnology (Masters), Flinders University.

- TAN, L., GEPPERT, H., SISAY, M. T., GUTSCHOW, M. & BAJORATH, J. 2008. Integrating structure- and ligand-based virtual screening: comparison of individual, parallel, and fused molecular docking and similarity search calculations on multiple targets. *ChemMedChem*, 3, 1566-71.
- TANRIKULU, Y., KRUGER, B. & PROSCHAK, E. 2013. The holistic integration of virtual screening in drug discovery. *Drug Discov Today*, 18, 358-64.
- TAO, W., GRUBMEYER, C. & BLANCHARD, J. S. 1996. Transition state structure of Salmonella typhimurium orotate phosphoribosyltransferase. *Biochemistry*, 35, 14-21.
- TAYLOR, A. F. & SMITH, G. R. 1990. Action of RecBCD enzyme on cruciform DNA. *J Mol Biol*, 211, 117-34.
- TAYLOR, R. G., WALKER, D. C. & MCINNES, R. R. 1993. E. coli host strains significantly affect the quality of small scale plasmid DNA preparations used for sequencing. *Nucleic Acids Res*, 21, 1677-8.
- TOKUOKA, K., KUSAKARI, Y., KRUNGKRAI, S. R., MATSUMURA, H., KAI, Y., KRUNGKRAI, J., HORII, T. & INOUE, T. 2008. Structural basis for the decarboxylation of orotidine 5'-monophosphate (OMP) by Plasmodium falciparum OMP decarboxylase. *J Biochem*, 143, 69-78.
- TRAUT, T. W. 1989. Uridine-5'-phosphate synthase: Evidence for substrate cycling involving this bifunctional protein. *Archives of Biochemistry and Biophysics*, 268, 108-115.
- TSANG, W. Y., WOOD, B. M., WONG, F. M., WU, W., GERLT, J. A., AMYES, T. L. & RICHARD, J. P. 2012. Proton transfer from C-6 of uridine 5'-monophosphate catalyzed by orotidine 5'-monophosphate decarboxylase: formation and stability of a vinyl carbanion intermediate and the effect of a 5-fluoro substituent. *J Am Chem Soc*, 134, 14580-94.
- UMEZU, K., AMAYA, T., YOSHIMOTO, A. & TOMITA, K. 1971. Purification and properties of orotidine-5'-phosphate pyrophosphorylase and orotidine-5'-phosphate decarboxylase from baker's yeast. *J Biochem*, 70, 249-62.
- UNIZONY, S., STONE, J. H. & STONE, J. R. 2013. New treatment strategies in large-vessel vasculitis. *Curr Opin Rheumatol*, 25, 3-9.
- VAN NGYUEN, S. 2005. *Characterisation of de novo pyrimidine biosynthesis enzymes from Plasmodium falciparum*. (Masters), Flinders University.
- VARDI-KILSHAIN, A., DORON, D. & MAJOR, D. T. 2013. Quantum and classical simulations of orotidine monophosphate decarboxylase: support for a direct decarboxylation mechanism. *Biochemistry*, 52, 4382-90.
- VEBER, D. F., JOHNSON, S. R., CHENG, H. Y., SMITH, B. R., WARD, K. W. & KOPPLE, K. D. 2002. Molecular properties that influence the oral bioavailability of drug candidates. *J Med Chem*, 45, 2615-23.
- VEDADI, M., LEW, J., ARTZ, J., AMANI, M., ZHAO, Y., DONG, A., WASNEY, G. A., GAO, M., HILLS, T., BROKX, S., QIU, W., SHARMA, S., DIASSITI, A., ALAM, Z., MELONE, M., MULICHAK, A., WERNIMONT, A., BRAY, J., LOPPNAU, P., PLOTNIKOVA, O., NEWBERRY, K., SUNDARARAJAN, E., HOUSTON, S., WALKER, J., TEMPEL, W., BOCHKAREV, A., KOZIERADZKI, I., EDWARDS, A., ARROWSMITH, C., ROOS, D., KAIN, K. & HUI, R. 2007. Genome-scale protein expression and structural biology of Plasmodium falciparum and related Apicomplexan organisms. *Mol Biochem Parasitol*, 151, 100-10.
- VELAZQUEZ-CAMPOY, A., TODD, M. J. & FREIRE, E. 2000. HIV-1 protease inhibitors: enthalpic versus entropic optimization of the binding affinity. *Biochemistry*, 39, 2201-7.
- VICTOR, M. E., BENGTTSSON, A., ANDERSEN, G., BENGTTSSON, D., LUSINGU, J. P., VESTERGAARD, L. S., ARNOT, D. E., THEANDER, T. G., JOERGENSEN, L. & JENSEN, A. T. 2010. Insect cells are superior to Escherichia coli in producing malaria proteins inducing IgG targeting PfEMP1 on infected erythrocytes. *Malar J*, 9, 325.

- VOIGT, J. H., BIENFAIT, B., WANG, S. & NICKLAUS, M. C. 2001. Comparison of the NCI open database with seven large chemical structural databases. *J Chem Inf Comput Sci*, 41, 702-12.
- VON KORFF, M., FREYSS, J. & SANDER, T. 2009. Comparison of ligand- and structure-based virtual screening on the DUD data set. *J Chem Inf Model*, 49, 209-31.
- WAGNER, J. C., GOLDFLESS, S. J., GANESAN, S. M., LEE, M. C., FIDOCK, D. A. & NILES, J. C. 2013. An integrated strategy for efficient vector construction and multi-gene expression in *Plasmodium falciparum*. *Malar J*, 12, 373.
- WANG, B., BUCHMAN, C. D., LI, L., HURLEY, T. D. & MEROUEH, S. O. 2014. Enrichment of Chemical Libraries Docked to Protein Conformational Ensembles and Application to Aldehyde Dehydrogenase 2. *Journal of Chemical Information and Modeling*.
- WANG, G. P., CAHILL, S. M., LIU, X., GIRVIN, M. E. & GRUBMEYER, C. 1999a. Motional dynamics of the catalytic loop in OMP synthase. *Biochemistry*, 38, 284-95.
- WANG, G. P., LUNDEGAARD, C., JENSEN, K. F. & GRUBMEYER, C. 1999b. Kinetic mechanism of OMP synthase: a slow physical step following group transfer limits catalytic rate. *Biochemistry*, 38, 275-83.
- WANG, X., CHOE, Y., CRAIK, C. S. & ELLMAN, J. A. 2002. Design and synthesis of novel inhibitors of gelatinase B. *Bioorg Med Chem Lett*, 12, 2201-4.
- WARSHEL, A., STRAJBL, M., VILLA, J. & FLORIAN, J. 2000. Remarkable rate enhancement of orotidine 5'-monophosphate decarboxylase is due to transition-state stabilization rather than to ground-state destabilization. *Biochemistry*, 39, 14728-38.
- WEBSTER DR, BECROFT DM & DP, S. 1995. *Hereditary orotic aciduria and other disorders of pyrimidine metabolism In: Scriver CR, Beaudet AL, Sly WS, Valle D (eds) The metabolic basis of inherited disease, 7th ed., McGraw-Hill, New York.*
- WEI, N.-N. & HAMZA, A. 2013. SABRE: Ligand/Structure-Based Virtual Screening Approach Using Consensus Molecular-Shape Pattern Recognition. *Journal of Chemical Information and Modeling*, 54, 338-346.
- WHITE, N. J. 2004. Antimalarial drug resistance. *J Clin Invest*, 113, 1084-92.
- WHO. 2010. *World Health Organisation: Guidelines for the Treatment of Malaria, Second ed.* [Online]. World Health Organization. Available: www.who.int.
- WHO. 2014. *World Health Organisation: World Malaria Report 2014* [Online]. World Health Organisation. Available: www.who.int.
- WILLIAMS, M., LOUW, A. I. & BIRKHOLTZ, L. M. 2007. Deletion mutagenesis of large areas in *Plasmodium falciparum* genes: a comparative study. *Malar J*, 6, 64.
- WINKLER, J. K. & SUTTLE, D. P. 1988. Analysis of UMP synthase gene and mRNA structure in hereditary orotic aciduria fibroblasts. *Am J Hum Genet*, 43, 86-94.
- WITTE, J. F., BRAY, K. E., THORNBURG, C. K. & MCCLARD, R. W. 2006. 'Irreversible' slow-onset inhibition of orotate phosphoribosyltransferase by an amidrazone phosphate transition-state mimic. *Bioorg Med Chem Lett*, 16, 6112-5.
- WITTMANN, J. G., HEINRICH, D., GASOW, K., FREY, A., DIEDERICHSEN, U. & RUDOLPH, M. G. 2008. Structures of the Human Orotidine-5'-Monophosphate Decarboxylase Support a Covalent Mechanism and Provide a Framework for Drug Design. *Structure*, 16, 82-92.
- WITTMANN, J. G. & RUDOLPH, M. G. 2007. Pseudo-merohedral twinning in monoclinic crystals of human orotidine-5'-monophosphate decarboxylase. *Acta Crystallogr D Biol Crystallogr*, 63, 744-9.
- WU, G. C., XU, X. D., HUANG, Q. & WU, H. 2013. Leflunomide: friend or foe for systemic lupus erythematosus? *Rheumatol Int*, 33, 273-6.
- WU, N., MO, Y., GAO, J. & PAI, E. F. 2000. Electrostatic stress in catalysis: structure and mechanism of the enzyme orotidine monophosphate decarboxylase. *Proc Natl Acad Sci U S A*, 97, 2017-22.
- WU, N. & PAI, E. F. 2002. Crystal structures of inhibitor complexes reveal an alternate binding mode in orotidine-5'-monophosphate decarboxylase. *J Biol Chem*, 277, 28080-7.

- YABLONSKI, M. J., PASEK, D. A., HAN, B. D., JONES, M. E. & TRAUT, T. W. 1996. Intrinsic activity and stability of bifunctional human UMP synthase and its two separate catalytic domains, orotate phosphoribosyltransferase and orotidine-5'-phosphate decarboxylase. *J Biol Chem*, 271, 10704-8.
- YAZAKI, M., OKAJIMA, K., SUCHI, M., MORISHITA, H. & WADA, Y. 1987. Increase of protein synthesis by uridine supplement in lectin-stimulated peripheral blood lymphocytes and EB virus-transformed B cell line of hereditary orotic aciduria type I. *Tohoku J Exp Med*, 153, 189-95.
- ZHANG, L., DAS, P., SCHMOLKE, M., MANICASSAMY, B., WANG, Y., DENG, X., CAI, L., TU, B. P., FORST, C. V., ROTH, M. G., LEVY, D. E., GARCIA-SASTRE, A., DE BRABANDER, J., PHILLIPS, M. A. & FONTOURA, B. M. 2012. Inhibition of pyrimidine synthesis reverses viral virulence factor-mediated block of mRNA nuclear export. *J Cell Biol*, 196, 315-26.
- ZHANG, Y., DENG, H. & SCHRAMM, V. L. 2010. Leaving group activation and pyrophosphate ionic state at the catalytic site of Plasmodium falciparum orotate phosphoribosyltransferase. *J Am Chem Soc*, 132, 17023-31.
- ZHANG, Y., EVANS, G. B., CLINCH, K., CRUMP, D. R., HARRIS, L. D., FRÖHLICH, R. F. G., TYLER, P. C., HAZLETON, K. Z., CASSERA, M. B. & SCHRAMM, V. L. 2013. Transition State Analogues of Plasmodium falciparum and Human Orotate Phosphoribosyltransferases. *Journal of Biological Chemistry*, 288, 34746-34754.
- ZHANG, Y., LUO, M. & SCHRAMM, V. L. 2009. Transition States of Plasmodium falciparum and Human Orotate Phosphoribosyltransferases. *Journal of the American Chemical Society*, 131, 4685-4694.
- ZHANG, Z., SCHWARTZ, S., WAGNER, L. & MILLER, W. 2000. A greedy algorithm for aligning DNA sequences. *J Comput Biol*, 7, 203-14.
- ZHAO, Y. & SANNER, M. F. 2007. FLIPDock: Docking flexible ligands into flexible receptors. *Proteins: Structure, Function, and Bioinformatics*, 68, 726-737.
- ZHU, S., YAN, X., XIANG, Z., DING, H. F. & CUI, H. 2013. Leflunomide reduces proliferation and induces apoptosis in neuroblastoma cells in vitro and in vivo. *PLoS One*, 8, e71555.